Social Distancing Following the SARS-Cov-2 Outbreak

Tyler Atkinson, Jim Dolmas, Christoffer Koch, Evan Koenig, Karel Mertens, Anthony Murphy and Kei-Mu Yi
Social Distancing Following the SARS-Cov-2 Outbreak*

Tyler Atkinson, Jim Dolmas, Christoffer Koch, Evan Koenig, Karel Mertens†, Anthony Murphy and Kei-Mu Yi

May 18, 2020

Abstract

We develop a Social Distancing Index (SDI) based on a range of mobility metrics from Safegraph geolocation data, and validate the index with mobility data from Google and Unacast. We construct SDIs at the county, MSA, state and nationwide level, and link these measures to indicators of economic activity. According to our measures, the bulk of social distancing occurred during the week of March 15 and simultaneously across the U.S. At the national peak of social distancing in early April, localities that engaged in 10% more social distancing than average saw an additional 0.6% of their populations claiming unemployment insurance, an additional 2.8 pp reduction in small businesses employment, an additional 2.6 pp increase in small business closures, and an additional 3.2 pp reduction in new-business applications. A gradual and broad-based reduction in social distancing started in the third week of April.

Keywords: Social Distancing Index, Location Data, Covid 19, SARS-Cov-2, Social Distancing and Economic Activity

JEL Classifications: E66, C15, R11, R19

---

*All authors are at the Federal Reserve Bank of Dallas. The views in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Dallas or the Federal Reserve System. Thanks to Lia Mertens for research assistance.

†Contact: Karel Mertens, mertens.karel@gmail.com, tel: +(214) 922-6000.
1 Introduction

The COVID-19 pandemic has inflicted a heavy toll, including on economic activity, which declined sharply beginning in mid-March. A key driver of the slowdown was an increase in social distance—literally, the physical distance between people—in order to mitigate the spread of COVID-19. Many businesses sharply curtailed, or even ceased, operations owing to government-mandated closures and stay-at-home orders, concerns for the health of their workers, or a lack of business, as consumers avoided social interaction. An appropriate real-time index of social distancing, then, might provide valuable real-time insight into the economic impact of the pandemic.

To that end, we develop an index of social distancing based on geolocation data made available by the firm SafeGraph. Our Social Distancing Index (SDI), described in more detail below, spiked dramatically in mid-March, coinciding with the steep drop in economic activity. More recently, the SDI started to decline prior to the relaxation of stay-at-home orders. The drop suggests economic activity may have bottomed out and could soon improve.

SafeGraph has made a number of series available to researchers through their Social Distancing Metrics database, which contains aggregated, anonymized, privacy-safe data on a range of spatial behaviors of mobile devices. No single indicator in the SafeGraph dataset adequately captures all aspects of social distancing; moreover, each series is noisy and subject to some idiosyncrasies. Our SDI, therefore, combines the information in several different variables—described in more detail below—each of which is measured daily, at the county level, and relative to its weekday average over January-February.

We apply Principal Component Analysis to these indicators, at the county level, to produce a set of county-level SDIs. We then aggregate the county-level SDIs to the metropolitan statistical area (MSA), state, and national levels.

The behavior of our national SDI is consistent with patterns evident in other mobility metrics that have become available since the start of the crisis. The national SDI is also tightly related to several high-frequency measures of aggregate economic activity, as are our local-level indexes in cross-section.

The remainder of the paper is organized as follows. We first review some related literature. Then, Section 2 provides details of our SDI’s construction. Section 3 gives an overview of the SDI’s behavior in the past few months, and Section 4 compares the SDI to alternative mobility metrics. We discuss the SDI’s relationship to economic activity in Section 5. Section 6 concludes.

1.1 Related literature

A number of recent papers have used SafeGraph data or other mobility metrics to gauge the extent of social distancing.

Mongey, Pilossoph, and Weinberg (2020) dissect different occupations in terms of the ability to work from home and physical proximity to others at work. Looking across counties, they find that a lower prevalence of work-from-home jobs correlates with smaller declines in ‘staying-at-home’ as measured using SafeGraph data.
Maloney and Taskin (2020) use Google mobility data to identify the determinants of social distancing during the 2020 COVID-19 outbreak. Much of the decrease in mobility, they find, is voluntary and driven by the number of COVID-19 cases. Government interventions, such as closing nonessential business, sheltering in place, and school closings, are also effective, although with a total contribution dwarfed by voluntary distancing. Likewise, Gupta et al. (2020) use mobility and social interaction data from SafeGraph and PlaceIQ to identify the role of local government interventions in explaining observed patterns of social distancing.

Farboodi, Jarosch, and Shimer (2020) build a SIR (Susceptible-Infected-Recovered) model of COVID-19 in which voluntary (though sub-optimal) social distancing plays an important role. They argue that their laissez-faire equilibrium accounts for the decline in social activity measured in SafeGraph data.

A number of papers look at the issue of compliance with mandatory social distancing requirements (or extent of voluntary distancing) and how that correlates with local characteristics such as income, education or political affiliation. Work in this vein includes Allcott et al. (2020), Barrios and Hochberg (2020), Painter and Qiu (2020), Brzezinski et al. (2020), and Engle, Stromme, and Zhou (2020).

Finally, Benzell, Collis, and Nicolaides (2020) and Goldfarb and Tucker (2020) both use location data to ask—with an aim to informing shutdown policy choices—which sorts of businesses generate the greatest amount of social interaction among their customers.

2 Construction of the Social Distancing Index

2.1 Geolocation Data

The social distancing indices are based on the Covid-19 Response Datasets provided by SafeGraph (2020). The Social Distancing Metrics dataset (Version 2) contains daily geolocation information for around 16 to 20 million mobile devices (Figure 1). Our sample starts on January 3, 2020. The data are continually updated by SafeGraph with a delay of a few days.

The data are generated using a panel of GPS pings from anonymous mobile devices. All variables are derived from data made available by SafeGraph at the census-block level of the home location of the device. The home location is defined as the usual nighttime location, determined for each mobile device over a 6-week period to a Geohash-7 granularity (153m x 153m). We aggregate the data from the census block to the county level. We retain only the data for the 50 US states and Washington DC, eliminating American Samoa, Virgin Islands, Guam, and Puerto Rico. We eliminate all counties for which there are missing data over one or more days in the estimation sample. We also omit all counties for which there are fewer than 100 mobile devices on any given day in the sample. This leaves more than 3000 counties in our baseline sample. Figure 1 plots device counts for counties with at least 1K, 5K, and 10K mobile devices on any given day. A large fraction of the observations come from a couple hundred high-population counties.

1See https://docs.safegraph.com/
2.2 Constituent Series of the Index

The Safegraph dataset contains a variety of mobility metrics, including counts of devices at home for each hour of the day, median time spent at home, bucketed counts by distance of trips, bucketed times spent away from home by travel distance, median distance traveled, etc.² The SDI is based on seven variables constructed from the raw data as observed from Jan 3, 2020 onwards. Those variables are chosen based on several criteria. The first criterion is that they show relative stability in the weeks prior to outbreak of SARS-Cov-2 in the U.S., such that they provide a reasonable baseline for normal mobility behaviors. The second criterion is that they show a clear change in pattern following the first case of suspected local transmission in United States on Feb. 26, 2020. The final criterion is that this change in pattern can reasonably be attributed to social distancing.

The constituent series are:

1. The fraction of devices leaving home in a day;
2. The fraction of devices away from home for 3-6 hours at a fixed location;
3. The fraction of devices away from home longer than 6 hours at a fixed location;
4. An adjusted average of daytime hours spent at home;
5. The fraction of devices taking trips longer than 16 kilometers;
6. The fraction of devices taking trips less than 2 kilometers; and
7. The average time spent at locations far from home.

The left panel in Figure 2 shows the total fraction of devices for which at least one trip away from home was recorded, as well as the fraction of devices that were not recorded as leaving the home location. Because of measurement problems, some devices fall in neither category, but fortunately that fraction is small and very stable over time. As the first input to

²For most of these variables, February 25 is a strong outlier for reasons that are not immediately clear to us. We simply omit this day from the sample and treat it as a missing observation.
We use the percentage point deviation in the fraction of devices relative to the average in the first 8 calendar weeks of 2020. Because most series display strong weekday seasonality, we calculate all deviations relative to weekday-specific averages. The right panel of Figure 2 shows an unweighted average across counties of the deviations in the fraction of devices leaving home, as well as the 80% percentile range.

(a) Levels

(b) Deviation From Jan/Feb

Figure 2: Fraction of Devices Leaving Home

*Source:* Authors’ calculations. Safegraph (2020) Social Distancing Metrics. The right panel shows the unweighted average and the 80% percentile range of the deviations from the Jan-Feb baseline in the fraction of devices leaving home.

Figure 3 plots the fraction of devices that record extended stays away from home at a fixed location. The first series shown is for stays between 3 and 6 hours, and the second is for stays of longer than 6 hours at the same non-home location. Such stays are consistent with part-time or full-time work behavior. However, we find only a weak cross-sectional correlation with other county-level indicators of workplace visits, such as those available from Google (see below). The extended variables therefore likely capture also a broad range of non-work behaviors, such as at time spent at school or at family and friends’ homes. We use deviations in both series relative to weekday-specific Jan/Feb averages as inputs for the SDI. To avoid clutter, the right panel in Figure 3 only shows the deviations for extended stays longer than 6 hours.

Figure 4 shows a measure of time spent at the home location during daytime. This measure is a ratio of two variables. In the numerator is the sum of the hourly counts of devices recorded at home between 6 AM and 10 PM. Some types of devices that show little movement in space do not send GPS information, such that there is underreporting of time spent at home. Moreover this measure of time spent at home, as well as related ones available (median time at home in minutes, median percent time at home) show a downward trend during January and February. It is unclear what causes these trends. If the fraction of devices that is recorded as being at the home location is the same during daytime and nighttime, and assuming the vast majority of devices spend nights at their home locations, an easy way to correct for these measurement problems is to divide by the sum of the hourly counts of devices recorded at home during the night hours (between 10 PM and 6 AM). The resulting ratio, shown in the left panel of Figure 4, is stable before mid-March. The left panel shows the percent deviation in this rate from the

---

3For March 8, we multiply the observations by 7/8 to adjust for daylight savings time.
Figure 3: Fraction of Devices with Extended Non-Home Stays in Fixed Locations

*Source:* Authors’ calculations. Safegraph (2020) Social Distancing Metrics. The right panel shows the unweighted average and 10th and 90th percentiles of deviations from the Jan-Feb baseline in the fraction of devices recording extended stays longer than 6 hours.

weekday-specific Jan-Feb averages. This series is the fourth input into the SDI.

Figure 4: Home Dwelling Time


Figure 5 shows the fraction of long and short distance trips. For each device, Safegraph computes the median distance for of all of the device’s daily trips. The series shown are the fraction of devices with median distance of either more than 16km or less than 2km away from home (the fraction of devices traveling with medium median distance is the complement of these series, and therefore contains no independent information). We use differences relative to weekday-specific Jan/Feb averages in both fractions as the fifth and sixth inputs to our index. To avoid clutter, the right panel in Figure 5 shows the unweighted average across counties for
long-distance trips only.

**Figure 5: Frequency of Long and Short Trips Away From Home**

*Source:* Authors’ calculations. Safegraph (2020) Social Distancing Metrics. The right panel shows the unweighted average and 10th and 90th percentiles of the deviation from the Jan-Feb baseline in the fraction of devices making long-distance trips.

The final input series is based on the median dwelling times at non-home locations according to distance. We use percent deviations in dwelling times at long distance locations, shown in the right panel of Figure 6, as the seventh input into the SDI. We do not include the series for the medium and short distance trips because they received little weight in the principal components analysis.

**Figure 6: Time Away From Home**

2.3 Principal Component Analysis

We hypothesize that the patterns of change in mobility that are evident in each of the Safegraph metrics can reasonably be attributed to a common underlying driving factor related to social distancing. The SDI is based on extracting this common factor by principal components analysis of the county-level time series. This means that, by construction, the SDI is a weighted average of the underlying series. The sample for the estimation of the weights runs from Jan 3 through April 6, 2020. Over this period, we pool all the county-level time series, normalize all seven metrics, and extract the first principal component. Table 1 provides the weights associated with the first principal component, as well as the total variance explained. The first principal component explains 60.49% of the overall variance in the seven county-level mobility metrics.

<table>
<thead>
<tr>
<th>Mobility Variable</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Devices Leaving Home</td>
<td>−0.43</td>
</tr>
<tr>
<td>Devices With Non-Home Stays &gt; 6 Hours</td>
<td>−0.38</td>
</tr>
<tr>
<td>Devices With Non-Home Stays &gt; 3 and &lt; 6 Hours</td>
<td>−0.43</td>
</tr>
<tr>
<td>Daytime to Nightime Hours at Home</td>
<td>0.45</td>
</tr>
<tr>
<td>Short Distance Trips Away From Home (&lt; 2km)</td>
<td>0.40</td>
</tr>
<tr>
<td>Long Distance Trips (&gt; 16km)</td>
<td>−0.27</td>
</tr>
<tr>
<td>Time at Long Distance Locations</td>
<td>0.22</td>
</tr>
</tbody>
</table>

Total County-Level Variance Explained 60.49%

The nationwide SDI shown in the upper left panel of Figure 7 is the nationwide weighted average of the first principal component for each county, where we use device counts as the weights. Similarly, we construct MSA and state-level indices as averages weighted by device counts. From the national SDI, we subtract a constant and apply a scaling factor such the national SDI averages zero between Jan 3 and Mar 1, 2020, and averages 100 for the week starting April 5. We then apply the same scaling constant and factor for the indices at all regional levels. This means that the values of the index can be compared across geographies. A value of zero represents nationwide pre-Covid normal mobility behaviors, and a value of 100 represents the nationwide-average level of social distancing during the week of April 5, 2020. To date, that week represents the peak in social distancing according to our index.

3 Recent Social Distancing Index Behavior

Figure 7 plots the Social Distancing Index (SDI) for the nation as a whole and for different regions. As we discussed above, all indices are scaled so that the nationwide daily SDI averages zero for January and February and reaches 100 in the second week of April. That week represents, according to the national SDI, the peak of social distancing behavior to date.

---

4 Principal components analysis is a linear transformation of the data into components that are uncorrelated. The first principal component is the component with the greatest variance.

5 We checked that using county-level population as weights makes no material difference.
The SDI varied minimally in January through early March of 2020, reflecting the normal spatial behavior of cell phone users before the pandemic. The index began to increase beyond its normal range in the second week of March, and increased sharply in the week ending March 21. By the end of March, social distancing had leveled off, and over the first half of April it moved sideways. In the second half of April, the index began to decline, as more Americans began leaving home each day and started taking longer trips away from home. For the average of the week ended May 2, the index was 20 percent below the mid-April peak.

Each of the grey lines in panel (c) of Figure 7 is the SDI for an MSA. The tightness of the range in these local SDIs indicates that most metro areas increased their social distancing at approximately the same time and by roughly the same amount, despite differences in the timing and scale of government interventions and the local pace of COVID-19 infections. A
similar pattern holds for counties and states, and suggests that stay-home-orders only partially explain social distancing. This is corroborated by the SDI’s decline in the second half of April, before stay-at-home orders were lifted, as well as by the findings of several academic studies.\(^6\)

Appendix A shows the SDIs for the largest metropolitan areas in the US.

### 4 Comparison with Other Mobility Data

There are a number of other mobility datasets that have been used to quantify social distancing behaviors, such as Google’s Covid-19 Community Mobility Reports.\(^7\) The Google reports contain county-level data on movement trends across different location categories, including retail and recreation locations, groceries and pharmacies, parks, transit stations, workplaces, and places of residence. Each of the Google metrics measures the percentage point increase or decrease in visits to the given location over a 5-week period from Jan 3 to Feb 6, 2020. Figures 8a and 8b show two of the metrics at the national level, alongside the SDI constructed from Safegraph data. The nationwide Google metrics are constructed from county level data in the same way as the SDI, i.e. as weighted averages using device counts as weights. Figure 8a shows Google’s measure of time spent at home, whereas Figure 8b shows a measure of workplace visits.\(^8\) Both measures are very similar to the SDI. Specifically, each of the Google measures show a similar rapid rise in social distancing in the Week starting March 15, 2020, a peak in the second week of April, and a gradual decline afterwards.

Another source of mobility data is Unacast’s Covid 19 Location Toolkit.\(^9\) Unacast publishes county-level data on average distance traveled, number of visits to nonessential locations, and a measure of the number of human encounters. The Unacast metrics represent percentage point changes in each of the metrics compared to a 4-week period from Feb 10 to Mar 8, 2020. Figures 8c and 8d show two of the metrics at the nationwide level, alongside the SDI constructed from Safegraph data. As before, the nationwide metrics are constructed from weighted averages using device counts as weights. Figure 8c shows the change in average distance traveled. Figure 8d shows the change in non-essential visits.\(^10\) Both measures again show a rapid rise in social distancing in the Week starting March 15, 2020, a peak in the second week of April, and a gradual decline afterwards.

The cross-sectional relationship between the SDI and the Google metrics is very strong. The first row of Figure 9 provides scatter plots and regression results for county-level averages for the week of April 5, 2020. The left panel shows that, during that week, counties with a 10 point higher score on the SDI experienced an additional 1.3% increase in time at home. The right panel shows that counties with a 10 point higher SDI value on average saw an additional 2.3% reduction in workplace visits. Both of these relationships are highly statistically significant, with t-statistics exceeding 30 in both cases. The second row of the figure shows that the SDIs are also strongly correlated with the Unacast metrics, although these relationships are somewhat noisier.

\(^6\)See, for example, Gupta et al. (2020), Farboodi, Jarosch, and Shimer (2020) or Brzezinski et al. (2020).
\(^7\)See [https://www.google.com/covid19/mobility/](https://www.google.com/covid19/mobility/).
\(^8\)The remaining Google metrics are reported in Appendix B.
\(^10\)The encounters measure is erratic for some large counties, and we choose not to report it.
Figure 8: Comparison with Other Mobility Data, National Time Series

Source: Authors’ calculations. Safegraph (2020) Social Distancing Metrics. Google (2020) Covid-19 Community Mobility Reports. Unacast (2020) Covid 19 Location Toolkit. The Google metrics represent percentage point increases or decreases in visits to the respective locations relative to the 5-week period between Jan 3 and Feb 6, 2020. The Unacast metrics represent percentage point decreases in average distance traveled and in the number of visits to nonessential locations relative to the 4-week period Feb 10 and Mar 8.
Figure 9: Comparison with Other Mobility Data, County Level Regressions, Week of April 5, 2020

5 Social Distancing and Economic Activity

The rapid switch towards social distancing in mid-March was a major impediment to economic activity. In this section, we quantify the relationship between the SDIs and several high frequency indicators of economic activity.

Figure 10a plots the national SDI along with the Lewis, Mertens, and Stock (2020) Weekly Economic Index (WEI). The WEI was developed to track the rapid economic developments associated with virus outbreak in the U.S., and summarizes the signal in 10 real time economic activity indicators collected mostly from private sources. The WEI is scaled to four-quarter GDP growth, and for the last week of April it indicates a level of GDP that is almost 12% below the level one year prior. Figure 10 shows that, according to the WEI, a strong and sudden decline in economic activity started in the week of March 15, 2020, precisely the week in which the SDI started its rapid rise. Whereas the SDI suggest a peak in social distancing occurred in the second week of April, to date the WEI suggest that economic activity in the U.S. has not yet bottomed, although the rate of decline has slowed markedly. Appendix C discusses the SDI in the context of the various constituent series of the WEI.

Figure 10: Social Distancing, Output and Unemployment


Figure 10b shows the relationship between the national SDI and indicators of small business activity and employment from the Homebase scheduling and time clock software used by 100,000+ local businesses and their hourly employees.11 Specifically, Figure 10b shows the percentage changes in hours worked by hourly employees, the number of hourly employees, and the number of businesses open relative to the baseline for the period Jan 4, 2020 – Jan 31, 2020. Similar to the WEI, each of these series shows a sharp decline in the second week of April, the week in which the SDI rises sharply. In contrast to the WEI, the indicators of small business activity show signs of a bottoming that is roughly coincident with the gradual decline in social distancing in late April.

11See https://joinhomebase.com/data/covid-19/.
Figure 11 provides cross-sectional evidence of the relationship between social distancing and economic activity. Figure 11a plots the state-level SDI for the week of April 5 against the sum of continued and initial claims for unemployment insurance, as a ratio of the state population, during that week. The relationship is statistically significant and has the expected sign. The estimated slope indicates that a 10% higher level in a state’s SDI relative to the national average is on average associated with an additional 0.6% of the state’s population claiming unemployment insurance. Figure 11b plots the state SDI for the week of April 5 against the change in the number of new-business applications relative to a year ago. The relationship is statistically significant. The estimated slope indicates that a 10% higher level in a state’s SDI relative to the national average is associated with a further 3.2 percentage points reduction in applications relative to a year ago.

The remaining panels of Figure 11 depict the relationship between the SDIs and the Homebase small business indicators for 50 of the larger metropolitan areas in the U.S. In each case, the relationship between the SDIs and the indicator is statistically significant at conventional levels, and has an economically meaningful magnitude. At the national peak of social distancing in early April, cities that engaged in 10% more social distancing relative to the national average saw 2.6% more small business closures, a 2.8% larger reduction in small businesses

employment, and a 2.4% fewer hours worked by hourly employees.

6 Concluding Remarks

The Dallas Fed Social Distancing Index captures what is arguably the primary driver of the large drop in economic activity experienced by the US economy in March and April of 2020, and its recent decline may bode a pickup in economic activity.

We expect social distancing behavior to decline further in the coming weeks and months as government restrictions are lifted. However, informed by the apparently voluntary nature of the SDI’s rise and and subsequent tapering, it is possible that social distancing remains elevated as long as the COVID-19 threat persists, regardless of official restrictions. As a result, economic activity is could remain depressed for some time relative to any non-COVID baseline. Of course there remain many uncertainties, chief among them the path of COVID-19 and the success or failure of efforts to develop treatments and vaccines. As these uncertainties resolve, our ability to update the SDI in near-real-time should provide a timely reading on the impact on social distancing and, consequently, on economic activity.

References


Painter, Marcus and Tian Qiu (2020). “Political beliefs affect compliance with covid-19 social distancing orders”. In: *CEPR Covid Economics 4*.


A Social Distancing in Major Metropolitan Areas

Figure 12 shows SDIs for some of the major metropolitan areas in the U.S. In each panel, the blue line represents the weighted average of all (200+) MSAs in our sample, where the weights are the number of mobile devices. The scale is the same as in Figure 7 and is such that the nationwide SDI averages zero for January and February and reaches 100 in the second week of April.

Figure 12: Social Distancing in Major Metropolitan Areas

Source: Authors’ calculations. Safegraph (2020) Social Distancing Metrics.
B Relationship with Other Google Mobility Metrics

Figure 13 shows national averages of the other mobility metrics from Google’s Covid-19 Community Mobility Reports, along with the national SDI. The first row shows that both changes in visits to transit stations and to retail and recreational locations align closely with the SDI. Visits to grocery and pharmacies first rise earlier in March before decreasing markedly before the end of the month, and visits to parks shows a similar pattern.

![Figure 13: Safegraph vs. Google Mobility, Other National Series](image)


All of the other Google metrics also correlate with SDI in the cross-section of counties, see Figure 14. For the week of April 5, counties with a higher SDI score see less visits to transit stations, retail and recreation locations, and groceries and pharmacies. Interestingly, in the cross-section there is a positive relationship between the SDI and foot traffic at parks.
Figure 14: Safegraph vs. Google, County Level Data, Week of April 5, 2020

C  Relationship with Other High Frequency Activity Indicators

Figures 15 plots the nationwide SDI together a number of other indicators of economic activity that are available for the nation as a whole and at the weekly frequency. Figure 15a shows that claims for unemployment insurance rose dramatically in the same week as the sharp rise in the SDI. Figure 15b depicts the Rasmussen consumer confidence index, which also moved coincidentally with the SDI. Same-store sales as measured by Redbook initially rose the week before the start of widespread social distancing, see Figure 15c, and remained relatively high through March, mostly likely reflecting stockpiling behavior. From the second week of April onward, however, sales are sharply down relative to the same period one year ago. Fuel sales to end users collapsed one week after the initial rise in social distancing, see Figure 15d.

Figure 15: Social Distancing and Economic Activity Indicators