BIS

BIS Bulletin

No 8

Identifying regions at risk with Google Trends: the impact of Covid-19 on US labour markets

Sebastian Doerr and Leonardo Gambacorta

21 April 2020

BIS Bulletins are written by staff members of the Bank for International Settlements, and from time to time by other economists, and are published by the Bank. The papers are on subjects of topical interest and are technical in character. The views expressed in them are those of their authors and not necessarily the views of the BIS. The authors are grateful to Giulio Cornelli for excellent analysis and research assistance, and to Louisa Wagner for administrative support.

The editor of the BIS Bulletin series is Hyun Song Shin.

This publication is available on the BIS website (www.bis.org).

ISSN: 2708-0420 (online)
ISBN: 978-92-9259-371-1 (online)

Sebastian Doerr
sebastian.doerr@bis.org

Leonardo Gambacorta
leonardo.gambacorta@bis.org

Identifying regions at risk with Google Trends: the impact of Covid-19 on US labour markets

Key takeaways

- *Information on local labour markets and Google searches can be used to construct a measure of the vulnerability of employment in different regions of the United States to the Covid-19 shock.*
- *Regional exposure to Covid-19 varies significantly, ranging from a low of 2% to a high of 98% of total local employment.*
- *We test for the usefulness of the Covid-19 exposure measure by showing that areas with higher exposure report more Google search queries related to the pandemic and unemployment benefits.*

1. Introduction

Covid-19 has hit the global economy hard, with some sectors affected particularly badly: activity in transport, leisure and retail industries has collapsed. Since the number of employees in these sectors varies across regions, the impact of Covid-19 on employment will show regional disparity. A better understanding of the regional impact of Covid-19 could help policymakers improve the design and effectiveness of policy measures.

This Bulletin uses US data on local industry-level employment before the outbreak, combined with current data on Google searches, to identify regions more likely to be affected by Covid-19. We first construct a measure of *Covid-19 employment exposure* (*CV19 exposure* for short) for each Designated Marketing Area (DMA). DMAs comprise multiple counties where the population can receive similar types of media content, such as television offerings or newspapers.[1] *CV 19 exposure* reflects the share of local employment in sectors that are most affected by Covid-19. Exposed regions are likely to suffer a decline in economic activity over the next quarters.

We then validate the accuracy of the *CV19 exposure* by comparing it with data on Google searches for terms related to Covid-19 and its economic impact. Specifically, we obtain information on the relative frequency of searches for "corona", as well as for "unemployment benefits" and "EDD" (Employment Development Department) for each DMA. The search term "corona" captures general interest in the virus, as well as its potential impact on health or the local economy. Searches for unemployment benefits are

[1] DMAs are widely used in audience measurements, which are compiled in the United States by Nielsen Media Research. DMAs comprise the vast majority of US counties (see Wikipedia, "Media market", accessed 1 April 2020). The average DMA contains 25 counties.

likely to capture the actual impact of Covid-19 on local labour markets. We show that areas with higher exposure report significantly more search queries for the virus and unemployment benefits.

The strong correlations of *CV19 exposure* and recent Google searches suggest that indicators based on local employment in hard-hit industries can help identify regions that will be most affected by the virus economically. Based on these findings, the note also constructs the exposure of each US county to Covid-19 and provides a more granular picture of areas at risk from the disease. Better information on the regional impact of the pandemic could help policymakers design effective responses, taking into account heterogeneity in the effects of the crisis.

2. Regional exposure to Covid-19

Covid-19 has affected some sectors more than others. At the top of the list of the hardest-hit industries are transportation, employment services, leisure and hospitality, and travel agencies. These industries are directly affected by containment measures such as social distancing, travel banks or lockdowns. Another sector that is suffering is mining / oil and gas; the pandemic triggered a sudden stop in capital flows to emerging market economies and a decline in demand, thereby drastically reducing oil prices.[2] We use local employment in these most affected industries to measure the regional impact of the pandemic for 209 DMAs. *Covid-19 employment exposure (CV19 exposure)* measures the share of jobs that are under imminent threat of being eliminated due to Covid-19 in each DMA. An exposure of zero indicates that no local jobs are at risk, while higher values indicate that more jobs are at risk. For example, a value of 0.25 would imply that 25% of local employees work within the most affected industries.[3]

On aggregate, 45 million out of a total of 125 million jobs are at risk. For the average DMA, 34% of jobs are exposed to the Covid-19 shock. Graph 1 (top panel) shows a map of DMAs and their respective *CV19 exposure*. Darker colours indicate higher exposure. DMAs in white have 19−30% exposure, while DMAs in dark red have at least 37% of their employment at risk. DMA exposure varies significantly across the United States, ranging from a low of 19% to a high of 54% of total employment. Areas with the highest exposure are, among others, Las Vegas (NV), Atlantic City (NJ) and Midland (TX).

If *CV19 exposure* measures the true impact of Covid-19 at the local level, the top panel in Graph 1 would allow policymakers to identify its regional effect and develop an appropriate response. The question is whether *CV19 exposure* is an accurate measure of the havoc wrought by Covid-19. In principle, it could be the case that pre-outbreak local employment shares provide an outdated snapshot of today's actual employment in each industry. There also remains uncertainty over the extent to which firms in the most affected industries have to shed workers. A validation exercise with auxiliary data from different sources could mitigate concerns about these potential inaccuracies.

3. Validating local *CV19 exposure* with Google Trends

Researchers are increasingly relying on Google search data to obtain a real-time picture of economic activity (Jun et al (2018)). Unsurprisingly, users also ask Google for information on the coronavirus. Graph 2 (left-hand panel) plots the relative frequency of Google searches for the terms "car" and "corona" from

[2] We follow the categorisation of industries proposed in Muro et al (2020) and Zandi (2020). Specifically, the industries included cover three-digit North American Industry Classification System (NAICS) codes: 211, 212, 213, 481, 483, 485, 487, 488, 541, 561, 611, 711−713, 721, 722 and 811−813. While mining industries are not suffering directly from the pandemic, the collapse in oil prices is generally linked to the Covid-19 outbreak. We report correlations when excluding the mining sector below.

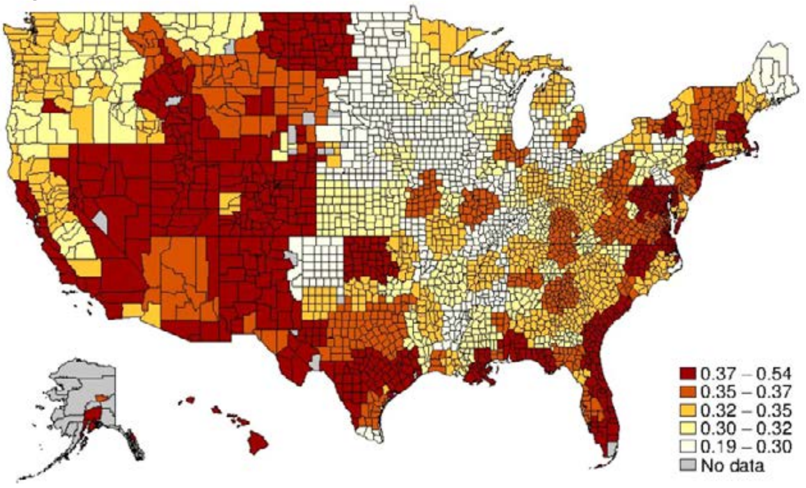
[3] County Business Patterns provide detailed employment data for each US county at the three-digit industry level in 2017. We compute DMA exposure as the sum of the share of employment in affected industries out of total employment in each DMA. Geographical exposure measures based on the shares of a given entity (eg an industry or bank) in a region are a common way to proxy local exposure to shocks. See eg Autor et al (2013) and Doerr (2019).

1 January to 7 April 2020 in the United States. "Car" is a term that exhibits little variation in search intensity over time. Instead, interest in corona sky-rocketed from mid-February onwards. The rise in corona-related searches anticipated the rise in active cases in the US, strongly suggesting that people are searching for corona to learn about the virus.

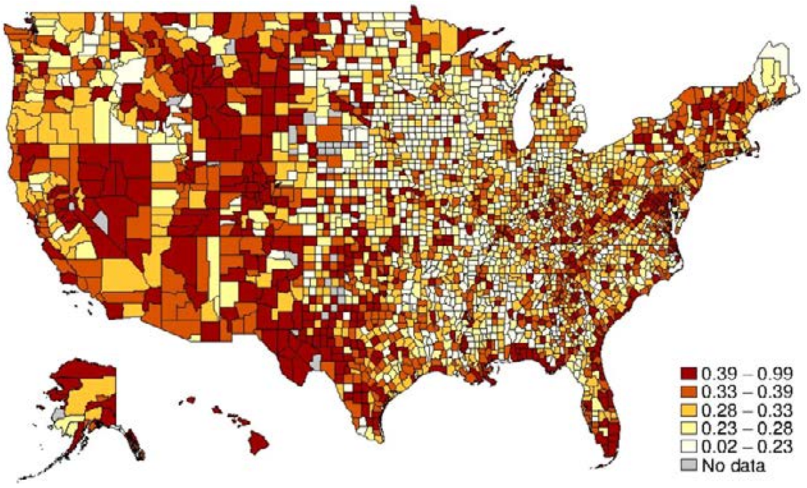
Employment exposure to Covid-19

In per cent | Graph 1

Designated Marketing Areas



Counties



Each panel shows the share of employment in industries that are likely to be the most affected by the Covid-19 shock. The top panel maps the Designated Marketing Area (DMA) level, the bottom panel the county level. Darker colours denote that a higher employment share is at risk, ie the areas that are more exposed to Covid-19.

Sources: County Business Patterns, 2017; authors' calculation.

Graph 2 (right-hand panel) shows that there has been an uptick in searches for "unemployment benefits" or "EDD" since the middle of March 2020. While searches for terms such as "lamp" show no change, searches for "unemployment benefits" increased more than eightfold in relative terms. The increase in searches coincides with media outlets reporting that companies are letting workers go.[4]

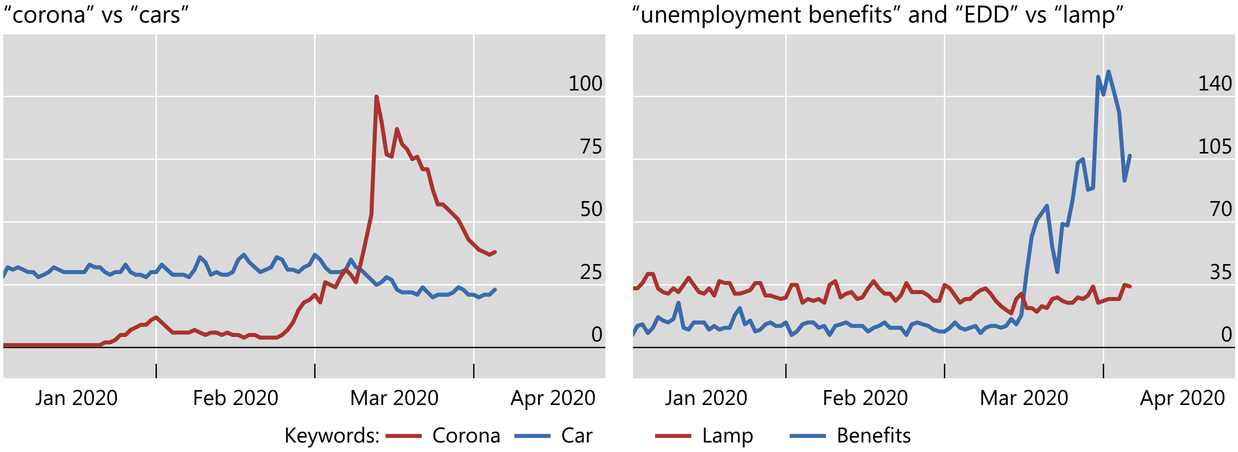
In addition to information over time, Google provides geographical information on search terms for each DMA. We construct two DMA-level measures based on the intensity of searches for corona (relative

4 See *Financial Times* (2020a,b).

to the control term "car") as well as "unemployment benefits" and/or "EDD" (relative to "lamp").[5] These two measures provide almost real-time information on the interest in Covid-19 and the state of local labour markets. Importantly, they allow us to validate whether *CV19 exposure* captures the virus's economic impact.

Google searches in 2020

Interest over time | Graph 2



Each panel shows the relative frequency of Google searches from 1 January to 7 April 2020. The panels show frequencies for the search terms in their headings ("EDD" = Employment Development Department).

Sources: Google Trends; author's calculations.

We find that DMAs with higher *CV19 exposure* also report more searches for "corona" and "unemployment benefits". Graph 3 (left-hand panel) plots exposure on the horizontal axis and searches for "corona" on the vertical axis for each DMA, together with a linear fit. There is a strong positive relationship between the two measures – in other words, DMAs that are more exposed to the corona shock also search more for information on the virus. Graph 3 (right-hand panel) paints a similar picture for searches on "unemployment benefits" or "EDD": in more-exposed areas, users also search more for information on unemployment benefits.[6] The strong correlation between exposure and searches for "unemployment benefits" is particularly striking: search volumes have increased only over the last few weeks.

To sum up, calculating the regional impact of Covid-19 on the basis of local employment shares appears to be a valid approach: it can help identify regions that will be most affected by the virus economically.[7] To provide an even more granular picture of the impact of Covid-19, we construct exposure for each US county. Graph 1 (bottom panel) shows a map of US counties and their respective exposure. In the average county, 32% of jobs are exposed to the Covid-19 shock, and county exposure

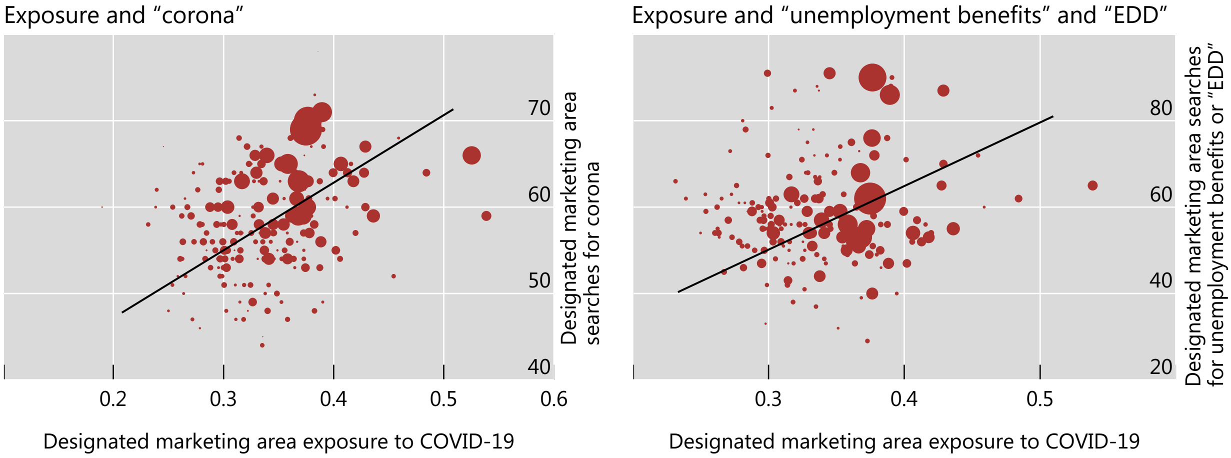
[5] Google Trends does not provide the actual number of searches. Instead, it gives the percentage of searches on one topic relative to another topic. For example, it could report that in Los Angeles (CA), out of total searches for "tacos" and "ramen", 80% of searches were for "ramen" and 20% for "tacos". To interpret any changes, one thus needs a comparator search term that does not fluctuate much over time. We choose "car" and "lamp", which exhibit little variation over time.

[6] Increasing the share of employment at risk in a DMA by 50% (ie increasing *CV19 exposure* from 0.25 to 0.75) is associated with a 21% increase in relative search frequency for "corona". An increase in exposure by 50% is associated with a 32% increase in relative search frequency for "unemployment benefits".

[7] Two industries among those at risk are indirectly affected by Covid-19: mining and transportation equipment manufacturing. Excluding the mining industry from the construction of *CV19 exposure* does not materially affect correlations shown in Graph 3. In the left-hand panel, regressions yield $\beta = 45.96$ with a t-value of 4.18; in the right-hand panel, $\gamma = 61.15$ with a t-value of 1.96. Including the transportation equipment manufacturing sector (NAICS 336, covering also motor vehicle manufacturing) slightly reduces the correlation of exposure with both search terms.

ranges from a low of 2% to a high of 98% of total employment. For example, counties in Texas are generally more exposed, in large part due to the importance of the local mining, oil and gas industry. Likewise, several counties in Florida and Hawaii have a high exposure, reflecting the importance of the local tourism industry.

Designated marketing area Covid-19 exposure and Google searches — Graph 3



Each panel provides a scatter plot with linear fit for around 200 Designated Market Areas (DMAs). In each panel, DMA exposure to Covid-19 is on the horizontal axis. The left-hand panel plots the percentage of Google searches for "corona" (relative to "car") on the vertical axis. The DMA-level regression $\% \text{ corona}_{dma} = \beta \text{ exposure}_{dma} + \varepsilon_{dma}$ yields a coefficient of $\beta = 42.54$ and a *t*-value of 4.29 for the period 1 March to 7 April 2020. The right-hand panel plots the percentage of Google searches for benefits ("unemployment benefits" and/or "EDD", relative to "lamp") on the vertical axis. The DMA-level regression $\% \text{ benefits}_{dma} = \gamma \text{ exposure}_{cmd} + \nu_{dma}$ yields a coefficient of $\gamma = 64.16$ and a *t*-value of 2.17 for the period 1 March to 7 April 2020. All regressions are weighted by DMA total population. The right-hand panel excludes Washington DC.

Sources: County Business Patterns, 2017; Google Trends; authors' calculations.

4. Conclusions

Designing and implementing policy measures that stabilise markets and support economic activity is of the essence during the current pandemic. To effectively combat the havoc wrought by Covid-19, granular information on the impact of Covid-19 across regions is paramount. This note has shown that employment-based measures of regional exposure to Covid-19 are an accurate assessment of the pandemic's regional impact.

Employment-based exposure can be constructed for most advanced countries with relative ease. The combination of official statistics and real-time data from non-traditional sources could further enhance policymakers' understanding of the heterogeneous impact of the Covid-19 shock and their ability to develop adequate responses (Buckman et al (2020)). It could also help overcome limitations in official data, such as low-quality, limited coverage, or reporting lags that in some cases could be substantial.[8]

[8] While other data sources exist, Google has the advantage of being the leading search engine, with a worldwide market share of over 90% (*Business Insider* (2018)). It also makes its data publicly available (to a certain extent), eg in its Covid-19 Community Mobility Reports that "chart movement trends over time by geography, across different categories of places". Other indicators, such as unemployment statistics, often come with a lag. Due to the special type of shock that Covid-19 represents, data on credit card usage or electricity consumption could also be misleading: during quarantine, people shop exclusively online, pay with credit cards and make extensive use of home entertainment devices or home-office office equipment – in other words, credit card usage or electricity consumption could go up even if overall economic activity declines.

References

Autor, D, D Dorn and G Hanson (2013): "The China syndrome: local labor market effects of import competition in the United States", *American Economic Review*, vol 103, no 6, pp 2121–68.

Business Insider (2018): "How Google retains more than 90% of market share", 24 April.

Buckman, S, A Shapiro, M Sudhof and D Wilson (2020): "News sentiment in the time of COVID-19", Federal Reserve Bank of San Francisco, *Economic Letters*, 6 April.

Doerr, S (2019): "Unintended side effects: stress tests, entrepreneurship and innovation", *BIS Working Papers*, no 823, November.

Financial Times (2020a): "Mnuchin warns of 20% unemployment without federal action", 18 March.

——— (2020b): "US set to be inundated with jobless claims", 19 March.

Jun, S, H Yooab and S Choi (2018): "Ten years of research change using Google Trends: from the perspective of big data utilizations and applications", *Technological Forecasting and Social Change*, vol 130, pp 69–87.

Muro, M, R Maxim and J Whiton (2020): "The places a COVID-19 recession will likely hit hardest", Brookings, 17 March.

Zandi, M (2020): "COVID-19: a fiscal stimulus plan", *Moody's Analytics*.

Previous issues in this series

No 7 17 April 2020	Macroeconomic effects of Covid-19: an early review	Frédéric Boissay and Phurichai Rungcharoenkitkul
No 6 14 April 2020	The recent distress in corporate bond markets: cues from ETFs	Sirio Aramonte and Fernando Avalos
No 5 7 April 2020	Emerging market economy exchange rates and local currency bond markets amid the Covid-19 pandemic	Boris Hofmann, Ilhyock Shim and Hyun Song Shin
No 4 6 April 2020	The macroeconomic spillover effects of the pandemic on the global economy	Emanuel Kohlscheen, Benoît Mojon and Daniel Rees
No 3 3 April 2020	Covid-19, cash and the future of payments	Raphael Auer, Giulio Cornelli and Jon Frost
No 2 2 April 2020	Leverage and margin spirals in fixed income markets during the Covid-19 crisis	Andreas Schrimpf, Hyun Song Shin and Vladyslav Sushko
No 1 1 April 2020	Dollar funding costs during the Covid-19 crisis through the lens of the FX swap market	Stefan Avdjiev, Egemen Eren and Patrick McGuire

All issues are available on our website www.bis.org.