Internet use and job market sentiment: An early assessment of COVID-19 pandemic shock across the EU

Marcin Wolski*, Patricia Wruuck

European Investment Bank
98-100 Boulevard Konrad Adenauer
L-2950 Luxembourg

Abstract

This paper investigates to what extent the widespread of internet access cushioned the job market sentiment during the outbreak of the COVID-19 pandemic. The empirical strategy applies a novel regional data set, constructed from the Google Trends reports for the EU in the first months of 2020. The findings suggest that while the overall impact of the pandemic spurred the online interest in unemployment-related topics, the effects were less severe in regions with more frequent internet use and where internet is more commonly related to business or civic activities. This cushioning effect is, however, short-lived. By the time the contagion reached 100 infections in a country, the difference between regions with frequent and infrequent internet access disappears.

Keywords: COVID-19, jobs, unemployment, digital, internet

JEL: J60, J63, R11, L86

*Corresponding author. Views presented in the paper are those of the authors only and do not necessarily represent the views of the European Investment Bank (EIB).

Email address: m.wolski@eib.org (Patricia Wruuck)

1. Introduction

Stringent containment measures, imposed to slow down the spread of the coronavirus (COVID-19 or C19), were quickly reflected in short-term declines in economic activity and surges in public interventions across major economies (Gurría, 2020). The early data suggest that the pace and magnitude of the shock were unprecedented, as exemplified by business and consumer confidence indicators plummeting across the globe, with some subcomponents reaching all-time lows. Furthermore, March 2020 represented the month with the highest number of downgrades by rating agencies in the last twenty years. Fitch Ratings put 83% of industry and structured finance asset performance outlooks, and all sovereign sector outlooks, at negative as of the end of March 2020 (Fitch Ratings, 2020).

Firms across regions have faced significant challenges from the worsening operating environment and economic shutdowns. The quick rise in unemployment and jobless claims in major developed and emerging markets have been, however, uneven across regions and sectors. While the demand has been shattered for the vast majority, the collapse was relatively less severe, or even ginned up, in sectors including utilities, telecommunications, food and healthcare (Fitch Ratings, 2020).

Aside from demand considerations, the coronavirus crisis has changed Europeans' social and working habits, with many employees exercising their job-related duties remotely during the lockdown. The widespread of the internet and communication technologies have been the necessary requirements to allow for such a prompt readjustment in the way people work and live (EIB, 2019). In this respect, digital and easy-to-digitalise jobs could have been swiftly moved to home offices, while person-to-person tasks have been either suspended or made redundant instead.

While many firms allowed employees to work from home, the protection offered by teleworking could be at most partial, due to indirect and spill-over effects through supply-chain linkages and re-prioritisation of expenditures. On a broader scope, more digitialised regions can be better technologically equipped, and therefore faster, in their response to a pandemic shock. This, in fact, can boost business confidence and correlate with job resilience indicators. Having pointed this out, this paper investigates to what extent internet access, as a proxy for digitalisation, shielded the jobs against the outburst of the pandemic crisis across the European Union (EU).

The question raises two main identification challenges. Firstly, the rapidness of the pandemic shock makes it difficult to find representative data sources. European unemployment statistics, either survey-based or related to claims on unemployment benefits, usually come at monthly frequency, and with substantial delay. Secondly, the relation between digitalisation and job resilience can be blurred by (often unobservable) common factors, like shock exposure, cultural considerations, work ethic or levels of technological progress. To address these issues, I turn to an alternative data source, i.e. a web-based unemployment indicator, which is available at daily frequency and at a high level of regional disaggregation.

More specifically, I rely on Google Trends data for the unemployment topic since the beginning of January until the end of March 2020. Topic indices offer two advantages over the standard query-based web indicators. Firstly, they categorize keywords, queries and page contents, in a consistent and comparable way across the jurisdictions. Secondly, they control for language differences. The Trends data are available for 353 regions across 28 Member States. I complement them by the daily situation reports from the European Centre for Disease Prevention and Control (ECDC). Concerning digitalisation, I put under the microscope the internet segment, with the data on regional internet use delivered by the Eurostat.

Including internet-based search results in near-term studies on unemployment is not new to the literature. While they seem to deliver substantial timing advantage over the survey-based indicators (Smith, 2016), they appear to correlate with hard unemployment indicators only during major economic downturns. Choi & Varian (2012) argue, for instance, that when taking into account the full sample between 2004 and mid-2011, including the Google Trend data in the unemployment forecasts reduces the fit by 5.95 per cent, compared to the baseline model. When focusing solely on the recession periods, however, search-engine data improve the fit by 13.6 per cent. In this respect, the dynamics of the search interest in early 2020 seems to highly overlap with early unemployment indicators, like the surging number of people registered as unemployed. While, the relation is not one-to-one, it supports the empirical rationale of this study.

Controlling for unobservable region-level, country-weekday and country-month fixed effects, I find that regions with higher internet use recorded, on average, a lower spike in unemployment-related Google search activity as a result of COVID-19 pandemic shock. This cushioning effect appears to be particularly strong when the internet use is related to business and civic activities, with no effects stemming from the use of social networks. Nevertheless, by the time the COVID-19 contagion reaches more than 100 infections per country, the difference between regions with frequent and infrequent internet access disappears. The results are confirmed by a variety of alternative model and data specifications.

This paper is organized as follows. Section 2 explains the data collection process and the empirical strategy. Section 3 describes the results, with concluding remarks presented in Section 4.

2. Data

As a starting point, I consider daily search interest in the unemployment topic as reported by Google Trends for individual EU regions, excluding cities.[1] The values represent search interest over time, and they are standardized to a scale between 0 and 100. A value of 100 is the peak popularity for the term. A

[1] The data were extracted from `https://trends.google.com` reports. The reports were downloaded on 2 April 2020 for the previous 90 days, hence covering the time span from 3 January until 31 March 2020. The Google Trends code for the unemployment topic is `/m/07s_c`.

value of 50 means that the term is half as popular as during the peak. A score of 0 means there was not enough data for this term.

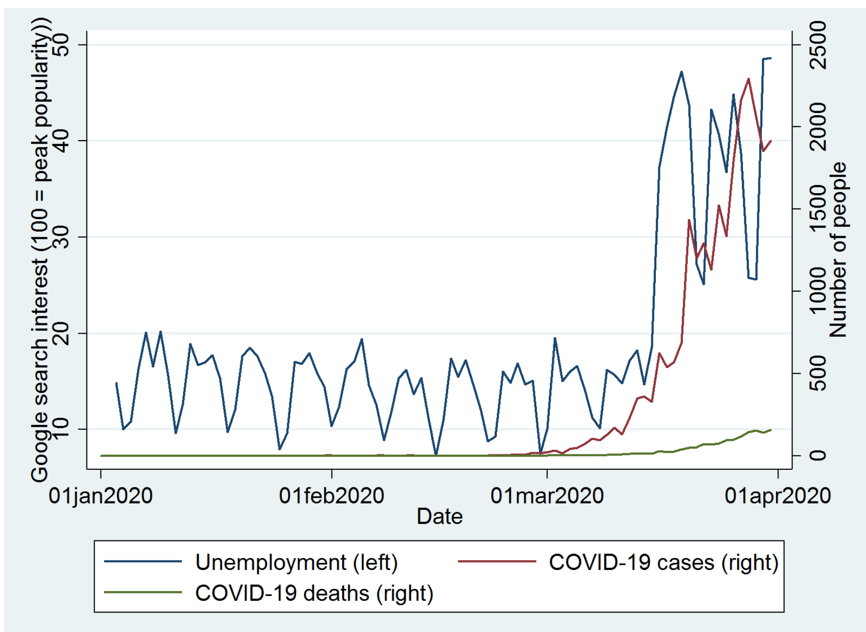
While there is no clearly-defined relation between search results and true unemployment levels, there are several advantages of using the search engine results over the standard labour market statistics. As the search interest data come with 36 hour delay, they offers nearly real-time overview of the market sentiment. Moreover, the statistics are available at more granular regional level, offering extra variation in the indicator across space. Last but not least, the online-related unemployment indices have been widely used to improve the accuracy of labour market models. Choi & Varian (2012) argue that simple seasonal auto-regressive models that include relevant Google Trends variables can outperform the standard models by 5 to 20 per cent. For instance, using the Trends information to estimate the claims of unemployment benefits during the recession periods in the US reduces the mean absolute error from 3.98 per cent to 3.44 per cent, an improvement of 13.6 per cent. These results are further supported by Pan (2019), who finds that internet-based unemployment indices in the US are positively associated with countercyclical labour market measures, including unemployment and layoff rates.

Vicente et al. (2015) use Google Trends data to nowcast and forecast the unemployment dynamics in Spain in the aftermath of the global financial crisis. They find that internet-based data brings 15 per cent improvement in forecasting accuracy. These findings are confirmed by González-Fernández & González-Velasco (2018), who find high correlation between relevant internet queries and unemployment in Spain. Last but not least, Naccarato et al. (2018) study the dynamics of youth unemployment in Italy and show that that the use of Google Trends information leads to an average decrease in the forecasting error.

The core variables measuring the spread of the COVID-19 pandemic include the number of newly identified cases and related deaths, as published by the ECDC. The series are available at the daily frequency for each of the Member States, and they begin when the first COVID-19 case was discovered in a country. Due to the exponential growth rate, in the empirical analysis I standardize these metrics by the log transformation, hence focusing only on positive values.

The relation between the series becomes apparent in Fig. 1. Moreover, the search results show a clear cyclical pattern, with less search activity over the weekends. The pace of the pandemic seems to accelerate rapidly in March, both in terms of newly reported cases and deaths, however, the first COVID-19 infections in the EU28 were reported already in the late January.

Figure 1: Newly diagnosed COVID-19 cases, related deaths and Google unemployment search trends.



Notes: The metrics are aggregated over the EU regions using the 2019 active population weights. Source: Google Trends, ECDC, Eurostat.

Finally, I consider several internet use indicators collected and published by the Eurostat. The main indicators comprise frequency of use and the type of internet-based activities. They describe the percentage of population that (i) use internet once a week, (ii) use internet daily, (iii) participate in social networks, (iv) use online banking, (v) use internet for selling goods or services, or (vi) use internet for civic or political purposes.

Before turning to the analysis, there are three data caveats that need to be pointed out. Firstly, with the exception of using internet for civic or political purposes, which correspond to 2017, the internet use values represent the year 2019. Secondly, the data for Poland is available at the NUTS1, whereas for all the other countries at the NUTS2 level. Thirdly, there is a geographical mismatch between the Google Trends and Eurostat regions. While the latter uses the official Eurostat nomenclature, the former relies on several vintages of the ISO 3166-1 standard, with non-standardized disaggregation rules. Many of the regions are straightforward to match, however to guarantee the EU-wide coverage, the remaining records are matched on the best-effort basis, by comparing the geographical borders of respective regions. Overall, Google regional representation determines the panel dimension in this study, comprising 353 regions.

The summary statistics are depicted in Table 1. The average Google Trends search interest for the topic of unemployment is 13.59 points,[2] however, as exemplified by Fig 1, it spiked towards the end of the sample. According to ECDC, as of the end of March 2020 380,906 people were infected by COVID-19, with nearly 26,082 deaths. The countries with the highest number of cases are Italy (101,739), Spain (85,195), Germany (61,913) and France (44,550). Regarding the internet use statistics, more than 81 per cent of people in the EU use internet at least once a week, and some 74 per cent access it on a daily basis. Social media and banking activities turn out to be nearly 4 times more popular than using internet for selling goods or for civic purposes. Overall, European regions show a substantial degree of variation in internet access, which I explore in greater detail in the next section.

[2]While the 0-100 scale allows to refer to search interest in terms of percentage points, to avoid confusion against other metrics, I call them points throughout the paper.

Table 1: Summary statistics.

	Variation	Obs.	Mean	St. dev.	Min.	Max.
Google Trend	c-r-t	30,349	13.59	23.03	0.00	100.00
C19 cases (log)	c-t	823	3.76	2.30	0.00	9.06
C19 deaths (log)	c-t	334	2.29	2.03	0.00	6.88
Weekly	r	353	81.08	9.71	61.00	98.00
Daily	r	353	73.71	11.83	49.00	96.00
Social	r	353	58.35	9.54	34.00	82.00
Banking	r	353	50.38	26.86	4.00	95.00
Selling	r	332	14.97	9.77	1.00	42.00
Civic	r	334	13.11	6.55	3.00	40.00

Notes: Variation describes the level of original variation in data series as c = country, r = regional and t = time. Internet access categories describe the percentage of population that (i) use internet once a week (weekly), (ii) use internet daily (daily), (iii) participate in social networks (social), (iv) use online banking (banking), (v) use internet for selling goods or services (selling), or (vi) use internet for civic or political purposes (civic). Source: Google Trends, ECDC, Eurostat.

3. Methodology and results

The main identification strategy measures to which extent the country-wide shock, in the form of newly diagnosed COVID-19 cases, is reflected in within-country unemployment search trends, conditional on regional internet use metrics.

Under the baseline specification, I estimate the following fixed-effects model

$$\mathrm{SI}_{crt} = \beta_1 \ln(\mathrm{C19})_{ct-1} + \beta_2 \ln(\mathrm{C19})_{ct-1} \times \mathrm{I}_r + \mu_r + \nu_{cw} + \psi_{cm} + \varepsilon_{crt}, \qquad (1)$$

where SI is the search intensity, C19 is the COVID-19 variable and I is the measure of internet use, with c, r, t representing country, region and time dimensions, respectively. As ECDC updates the C19 cases typically in the afternoon each day, I consider the lagged values. To control for potential confounders and cyclicality, the model is saturated with a vector of region-specific fixed effects μ_r, country-weekday fixed effects ν_{cw} and country-month fixed effects ψ_{cm}. The main results are presented in Tables 2 and 3, for the number of newly identified C19 cases and C19-related deaths, respectively.

Table 2: Impact of newly identified COVID-19 cases on unemployment by internet access.

	(1) Weekly	(2) Daily	(3) Social	(4) Banking	(5) Selling	(6) Civic
C19	8.425***	9.778***	5.719***	7.390***	6.475***	6.745***
	(1.889)	(1.296)	(1.041)	(0.453)	(0.384)	(0.454)
C19 $\times$ I	-0.040*	-0.062***	-0.010	-0.043***	-0.072***	-0.111***
	(0.022)	(0.017)	(0.018)	(0.008)	(0.020)	(0.026)
Obs.	10,198	10,198	10,198	10,198	9,484	9,783
R2	0.313	0.314	0.312	0.316	0.328	0.314
Adj. R2	0.273	0.274	0.272	0.276	0.289	0.275

*Notes: Model specification is given in Eq. (1) with C19 variable describing the lagged number of newly identified COVID-19 cases (in logs). Internet access categories correspond to Table 1. Standard errors (in parentheses) are clustered at the region level. Significance codes: *** for 0.01, ** for 0.05 and * for 0.1 level.*

Table 3: Impact of COVID-19-related deaths on unemployment by internet access.

	(1) Weekly	(2) Daily	(3) Social	(4) Banking	(5) Selling	(6) Civic
C19	7.325*	5.686*	4.513***	5.380***	4.120***	5.083***
	(3.844)	(3.364)	(1.424)	(1.258)	(0.735)	(1.006)
C19 $\times$ I	-0.046	-0.029	-0.021	-0.034*	-0.031	-0.105*
	(0.045)	(0.043)	(0.028)	(0.021)	(0.033)	(0.061)
Obs.	4,418	4,418	4,418	4,418	4,103	4,272
R2	0.368	0.368	0.368	0.369	0.382	0.365
Adj. R2	0.292	0.292	0.292	0.293	0.307	0.291

*Notes: Model specification is given in Eq. (1) with C19 variable describing the lagged number of COVID-19-related deaths (in logs). Internet access categories correspond to Table 1. Standard errors (in parentheses) are clustered at the region level. Significance codes: *** for 0.01, ** for 0.05 and * for 0.1 level.*

Three main observations need to be pointed out. Firstly, while the country-wide C19 pandemic shock increases the unemployment-related online search activity, the transmission appears to be muted in regions with more frequent internet access, on average. Secondly, the type of internet activity matters. Using internet for social media does not support job resilience. To the contrary, activities related to banking, selling of goods and services, or using internet for civic purposes significantly correlate with lower threat of unemployment. Thirdly, at least in the short-term, the labour market seems to suffer more from the rapidness of contagion, rather than from its mortality, as exemplified by lower magnitude and statistical significance of corresponding coefficients in Table 3 compared to Table 2. This distinction can suggest that timing matters: an increase in C19-related deaths happens some time after a surge in newly identified cases.[3]

The results suggest that regions with broader internet use, and where internet is linked to business or civic activities, were more agile to adjust their working habits in the face of the pandemic. Overall, 100 newly diagnosed C19 cases would lead to a surge of 26.49 points in unemployment search activity in regions where 10 per cent of population use internet for selling goods or services, compared to 19.91 points increase in regions with 30 per cent internet use for the same purpose. That is a difference of 6.58 points or nearly 25 per cent, yet the availability of home offices or teleworking were not nearly enough to prevent the malaise from spreading to digital regions.

3.1. Robustness

While Model 1 allows to control for fixed common factors, the pandemic shock can spread unevenly within a country, hence biasing the main coefficients of interest. Therefore, in the last step, I consider several alternative model and sample specifications to address possibly time-varying confounders, verifying the robustness of the results. For clarity of exposition, I focus on the percentage of people that use internet for selling goods and services as regional differentiation indicator (Columns 5 in Tables 2 and 3).

Table 4 presents the robustness analysis. Column (1) considers a dynamic model specification with the first lag of the dependent variable. In Column (2) I exclude the capital regions, as possibly inflating the numbers by both being more digital and more exposed to the pandemic shock. Column (3) considers the observations from the moment when cumulatively 100 C19 cases have been identified in a country. Column (4) looks only at March 2020 and finally Column (5) excludes all the countries for which the regions are either inconsistent

[3] Indeed, the two measures are closely related, hence difficult to analyze through the prism of one model simultaneously. When they are jointly included in the horse-race specification, this dependence is revealed in either lack of statistical significance or opposite signs of the interaction coefficients.

between the Google and Eurostat classifications, or comprise single regions (Belgium, Cyprus, Estonia, Ireland, Latvia, Lithuania, Luxembourg, Malta, Poland, Portugal, Slovenia and the United Kingdom).

Table 4: Alternative model and sample specifications.

	(1)	(2)	(3)	(4)	(5)
SI (lag)	0.170***				
	(0.018)				
C19	5.471***	6.346***	6.174***	6.788***	7.367***
	(0.336)	(0.393)	(0.611)	(0.434)	(0.372)
C19 ×	-0.063***	-0.078***	0.001	-0.065***	-0.113***
I=Selling	(0.017)	(0.021)	(0.030)	(0.025)	(0.020)
Obs.	9,484	8,719	6,318	8,393	7,848
R2	0.347	0.311	0.337	0.324	0.332
Adj. R2	0.308	0.272	0.279	0.280	0.298

*Notes: Column (1) uses a dynamic model specification. Columns (2)-(5) include model specification as given in Eq. (1) with C19 variable describing the lagged number of newly identified COVID-19 cases (in logs) and the percentage of population using internet for selling goods or services (I=Selling). Column (2) excludes capital regions. Column (3) uses the sample with more than 100 identified cases. Column (4) uses a sample of March 2020. Column (5) excludes inconsistent or country-wide regions. Standard errors (in parentheses) are clustered at the region level. Significance codes: *** for 0.01, ** for 0.05 and * for 0.1 level.*

Table 4 broadly confirms the main findings. The lack of statistical significance in Column (3) may suggest that the shield with which internet protects jobs against the pandemic is rather short-lived. It seems that by the time 100 C19 cases have been identified in a country, the labour market sentiment is shattered to the same extent in regions with low and high internet use.[4]

4. Conclusions

This paper investigates to which extent the widespread use of internet cushioned the propagation of the COVID-19 pandemic shock to the labour market sentiment. To this end I utilize a web-based search interest data for the topic of unemployment in 353 European regions in the first three months of 2020.

The main results suggest that while the overall impact of COVID-19 spurred the online interest in unemployment-related topics, the effects were less severe in regions with more frequent internet use and where internet is more commonly related to business and civic activities. This cushioning effect is, however, short-lived. By the time the COVID-19 contagion reached more than 100 infections per country, the difference between regions with frequent and infrequent internet access disappears. It seems that the severe magnitude of the shock has not spared digital tasks either, affecting them indirectly through spill-over effects and supply chain linkages.

While this study focuses on short-term response to the recent pandemic shock, the longer-term response needs to be carefully evaluated as more data become available. In this respect, the regional internet use numbers can be further complemented by alternative digitalisation metrics, like the widespread of Internet of Things, Artificial Intelligence, and Big Data technology.

References

Choi, H., & Varian, H. (2012). Predicting the present with google trends. *Economic Record*, *88*, 2–9. doi:10.1111/j.1475-4932.2012.00809.x.

EIB (2019). *EIB Investment Report 2019/2020: accelerating Europe's transformation*. European Investment Bank.

[4] Similar pattern is observed when looking at different internet use and internet activity indicators.

Fitch Ratings (2020). Fitch Ratings Updates 2020 Sector Outlooks To Reflect Coronavirus Impact. URL: https://www.fitchratings.com/research/fund-asset-managers/fitch-ratings-updates-2020-sector-outlooks-to-reflect-coronavirus-impact-27-03-2020.

González-Fernández, M., & González-Velasco, C. (2018). Can google econometrics predict unemployment? evidence from spain. *Economics Letters*, *170*, 42 – 45. doi:https://doi.org/10.1016/j.econlet.2018.05.031.

Gurría, A. (2020). Statement for the G20 Videoconference Summit on COVID-19. URL: https://read.oecd-ilibrary.org/view/?ref=126_126445-5ofyod1xpv.

Naccarato, A., Falorsi, S., Loriga, S., & Pierini, A. (2018). Combining official and google trends data to forecast the italian youth unemployment rate. *Technological Forecasting and Social Change*, *130*, 114 – 122. doi:https://doi.org/10.1016/j.techfore.2017.11.022.

Pan, W.-F. (2019). Building sectoral job search indices for the united states. *Economics Letters*, *180*, 89 – 93. doi:https://doi.org/10.1016/j.econlet.2019.04.019.

Smith, P. (2016). Google's midas touch: Predicting uk unemployment with internet search data. *Journal of Forecasting*, *35*, 263–284. doi:10.1002/for.2391.

Vicente, M. R., López-Menéndez, A. J., & Pérez, R. (2015). Forecasting unemployment with internet search data: Does it help to improve predictions when job destruction is skyrocketing? *Technological Forecasting and Social Change*, *92*, 132 – 139. doi:https://doi.org/10.1016/j.techfore.2014.12.005.