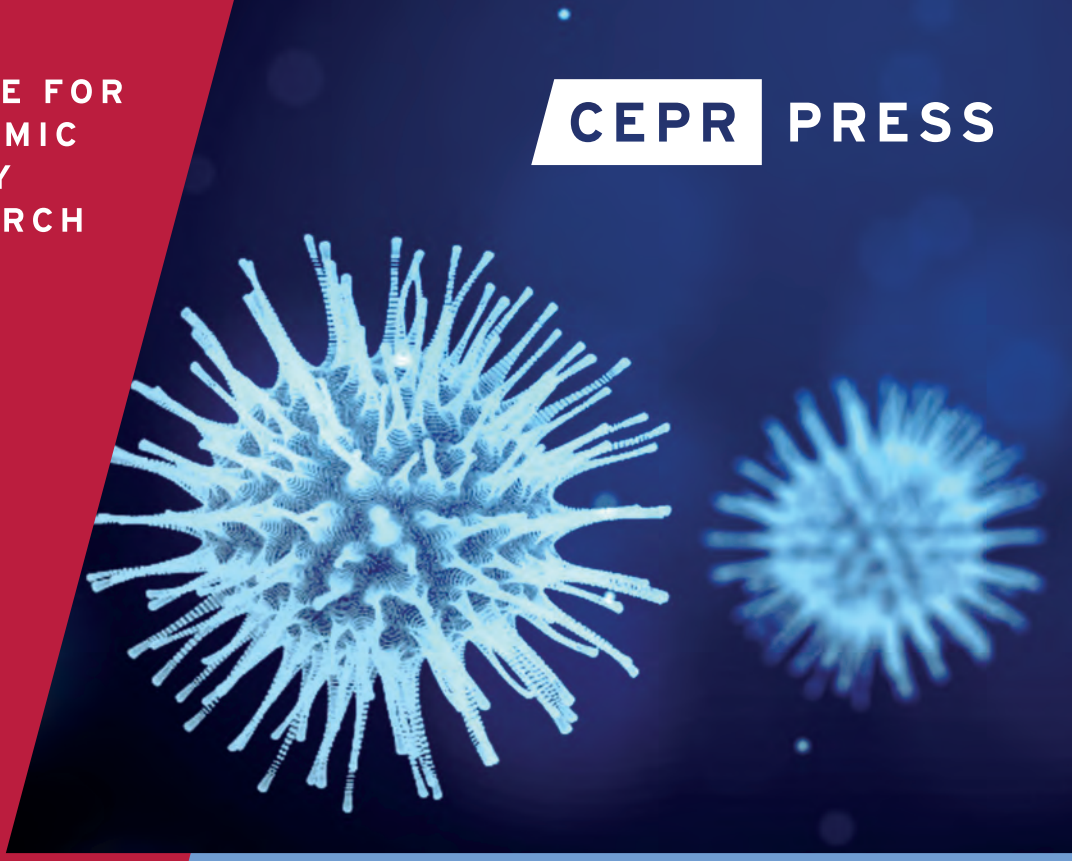
CENTRE FOR ECONOMIC POLICY RESEARCH

CEPR PRESS



COVID ECONOMICS

VETTED AND REAL-TIME PAPERS

ISSUE 43
21 AUGUST 2020

Covid Economics
Vetted and Real-Time Papers

Covid Economics, Vetted and Real-Time Papers, from CEPR, brings together formal investigations on the economic issues emanating from the Covid outbreak, based on explicit theory and/or empirical evidence, to improve the knowledge base.

Founder: Beatrice Weder di Mauro, President of CEPR
Editor: Charles Wyplosz, Graduate Institute Geneva and CEPR

Contact: Submissions should be made at https://portal.cepr.org/call-papers-covid-economics. Other queries should be sent to covidecon@cepr.org.

The Centre for Economic Policy Research (CEPR)

The Centre for Economic Policy Research (CEPR) is a network of over 1,500 research economists based mostly in European universities. The Centre's goal is twofold: to promote world-class research, and to get the policy-relevant results into the hands of key decision-makers. CEPR's guiding principle is 'Research excellence with policy relevance'. A registered charity since it was founded in 1983, CEPR is independent of all public and private interest groups. It takes no institutional stand on economic policy matters and its core funding comes from its Institutional Members and sales of publications. Because it draws on such a large network of researchers, its output reflects a broad spectrum of individual viewpoints as well as perspectives drawn from civil society. CEPR research may include views on policy, but the Trustees of the Centre do not give prior review to its publications. The opinions expressed in this report are those of the authors and not those of CEPR.

Editorial Board

Ethics

Covid Economics will feature high quality analyses of economic aspects of the health crisis. However, the pandemic also raises a number of complex ethical issues. Economists tend to think about trade-offs, in this case lives vs. costs, patient selection at a time of scarcity, and more. In the spirit of academic freedom, neither the Editors of *Covid Economics* nor CEPR take a stand on these issues and therefore do not bear any responsibility for views expressed in the articles.

Submission to professional journals

The following journals have indicated that they will accept submissions of papers featured in *Covid Economics* because they are working papers. Most expect revised versions. This list will be updated regularly.

American Economic Review
American Economic Review, Applied Economics
American Economic Review, Insights
American Economic Review, Economic Policy
American Economic Review, Macroeconomics
American Economic Review, Microeconomics
American Journal of Health Economics
Canadian Journal of Economics
*Econometrica**
Economic Journal
Economics of Disasters and Climate Change
International Economic Review
Journal of Development Economics
*Journal of Econometrics**
Journal of Economic Growth
Journal of Economic Theory
*Journal of the European Economic Association**
Journal of Finance
Journal of Financial Economics
Journal of International Economics
*Journal of Labor Economics**
Journal of Monetary Economics
Journal of Public Economics
Journal of Public Finance and Public Choice
Journal of Political Economy
Journal of Population Economics
*Quarterly Journal of Economics**
Review of Economics and Statistics
*Review of Economic Studies**
Review of Financial Studies

(*) Must be a significantly revised and extended version of the paper featured in *Covid Economics*.

Covid Economics
Vetted and Real-Time Papers

Issue 43, 21 August 2020

Contents

Covid Economics Issue 43, 21 August 2020

The short-run effect of COVID-19 on new marketing endeavors: Evidence from EUIPO's trademark applications

Kyriakos Drivas[1]

Date submitted: 14 August 2020; Date accepted: 18 August 2020

This study compares this year's trend of EUIPO's trademark applications during May and the first twenty days of June to 2019. There are four important findings. First, overall trademark applications appear to be at the same level to last year. Second, there is significant heterogeneity by country. While many countries are at the same levels to last year, China is an outlier increasing its filings dramatically compared to 2019. Certain countries also experience a sharp decrease including Canada and Brazil. Third, the presence of entrants is higher in 2020 compared to the same period of 2019. Finally, there are some clear winners and losers in terms of business activity. Overall, service-related endeavors are less frequent compared to last year while certain product-related initiatives have experienced a significant increase. This study urges dissemination of trademark applications, across Offices, in bulk to facilitate empirical work. Unlike patent applications, which take eighteen months to become known, information on trademark applications is disclosed relatively quickly. Since they can approximate real-time business expectations of demand and are related to innovation and firm value, they can provide us with significant insights in the short-run until more data become available.

Covid Economics 43, 21 August 2020: 1-18

1 Department of Economics, University of Piraeus.

Covid Economics 43, 21 August 2020: 1-18

1. Introduction

Recently, the European Union Intellectual Property Office (EUIPO) published a report on trademark filings until June of 2020. Comparing the trend of filings between the first semesters of 2019 and 2020, one would not be able to infer that the latter semester is that of a pandemic (Figure 1). Even for the last month of the sample, it appears that new trademark applications increase relatively to June of last year.

Trademark applications can capture meaningful economic activity. Scholars have shown that trademarks are correlated with innovation activity both at the firm and regional level (Mendonça *et al.,* 2004; Flikkema *et al.*, 2019; Drivas 2020). They can be used to protect innovations that otherwise cannot satisfy the patentability requirements and/or as a means to protect marketing activities aimed to capture rents from patentable technological inventions (Zhou *et al.,* 2016).

Similarly to patents, they have also been shown to contribute to a multitude of firm outcomes including, but not limited to, market value (Sandner and Block 2011), ability to attract venture capital (Block *et al.,* 2014; Hoenen *et al.,* 2014), organizational survival (Giarratana and Fosfuri, 2007; Srinivasan *et al.,* 2008), employment growth (Link and Scott 2012) and post-IPO performance (Chemmanur *et al.,* 2018); for a literature review see Castaldi (2020). At the regional level, scholars have also shown that they contribute to a region's resilience (Filippetti *et al.,* 2019).

At a more practical level, trademark applicants, to register their application need to provide evidence that the trademark is used commercially.[1] This is a directive followed in many Offices around the world ensuring that trademarks are attached to actual commercial activity. For our purposes, this directive implies that for these new trademark applications, entities have, at least to an extent, planned marketing activities for a new or improved product/service. Coupled with application fees, filing for a trademark application reveals cost and effort exerted from the applicant's side.

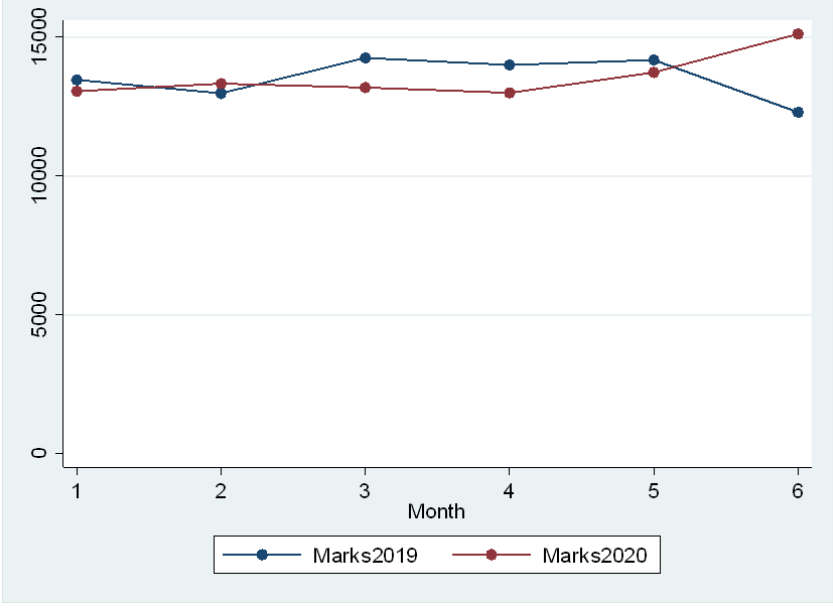
Given the aforementioned discussed importance of EUIPO trademark applications, it is encouraging to observe that trademark applications did not plummet in the aftermath of the first wave of COVID-19. If anything, applications appear to withstand and even increase during the last month of June. This study is motivated by this stylized fact and examines whether there is heterogeneity masked under this trend. Three major findings emerge:

[1] Guidelines, Part C, Opposition, Section 6, Proof of Use, paragraph 1.1

1) There is significant heterogeneity by country. There are many countries that have increased their trademark filings compared to May-June of last year with China leading the increase.
2) On average, new firms appear to be picking up the pace as there are more filings by new entrants compared to 2019. On the contrary, established firms file for fewer applications compared to last year.
3) There is also significant heterogeneity by class with some clear winners and losers. Classes in pharmaceutical and medical equipment, and home equipment including gymnastics appear to increase compared to last year. Service-related endeavours on the other hand have had the largest decrease.

This study urges the rapid dissemination of trademark application in bulk to facilitate empirical work. Most Offices by now disseminate this information digitally via the TMView platform.[2] While taking into account any legal restrictions regarding sensitive information and resource constraints, trademark applications can approximate real-time business expectations of demand and can provide significant insights until more data become available.

Figure 1. Trademark application trends (January-June) based on EUIPO's Report Data. All countries.



Notes: Marks2019 and Marks2020 stand for trademark applications filed in a given month in 2019 and 2020 respectively.
Source: EUIPO. "EUIPO Statistics in European Union Trade Marks: 1996-01 to 2020-06 Evolution" 2020

[2] https://www.tmdn.org/tmview/#/tmview

Covid Economics 43, 21 August 2020: 1-18

Covid Economics 43, 21 August 2020: 1-18

2. Data Description and Representativeness

During July 15th and July 30th, 2020, we downloaded information on trademark applications filed since 2015 from EUIPO's open data portal.[3] Bibliographic information on trademark applications is located in xml files. Information on trademark applicants is located in separate xml files. After a tedious rendering process of all the xml files a final dataset was compiled for the all the trademark applications filed during January 1, 2019 and June 20, 2020. While data for the last ten days of June were available, they were lacking applicant information which was crucial for this study. We denote this main dataset of this study, the "Rendered" dataset. For a trademark application to enter our sample it needs to satisfy the following criteria:

i) It has complete information on application date, NICE classification(s), priority date(s) (if any), Applicant information (applicant id and country location).

ii) The trademark application stems from a country that has filed at least one trademark in every single month since January of 2018. This way we ensure that firms in the country under consideration are engaging in continuous business activity within the EU.

Overall, we acquire complete information for 280,986 trademark applications stemming from 55 countries (28 EU member countries plus 27 more including US and CN). The frequency of trademark applications for 2019 by country is displayed in Table A1 of the Appendix.

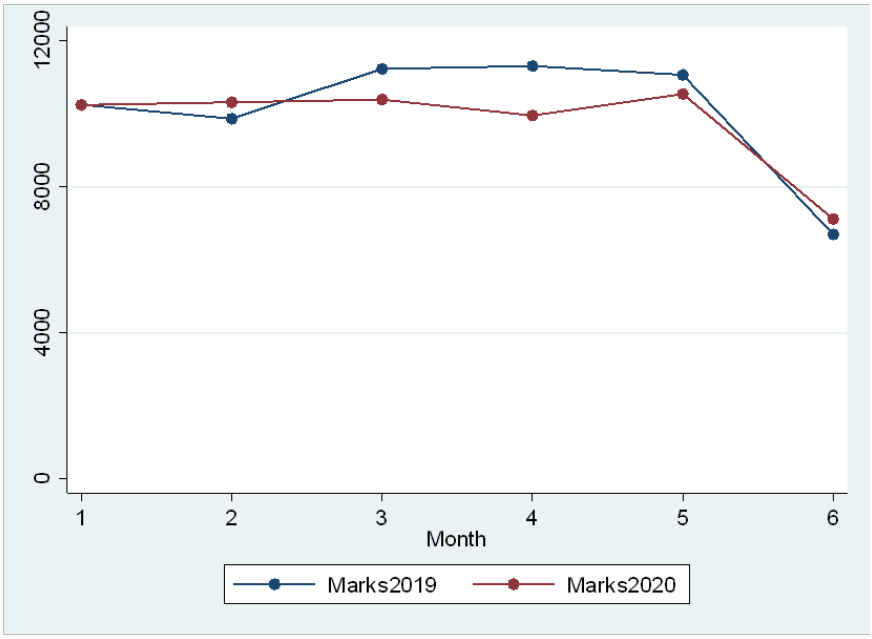
The next step is to check for the representativeness of the data that were collected over the entire population. To this end, we employ data from the recent EUIPO report (2020); note that any statistics used from this report, we refer to them as figures from the "Report" dataset.

Via simple calculations, the Report dataset shows that these 55 countries have filed 158,223 applications during 2019 and account for 98.6% of all trademark applications. From our Rendered data, these 55 countries have filed 122,439 applications and account for 95.9% of all trademark applications. This difference can be attributed to a multitude of reasons including: i) in case we observe more than one trademark applicant, we only consider the first thereby not double counting a trademark application in case it has more than one applicants, ii) in many occasions we excluded trademark applications that were missing applicant location information. This could be because of mismatching between the applicant ids and/or lack of applicant id information.

[3] https://euipo.europa.eu/ohimportal/en/open-data

Second, we compare the trends between these two datasets. Figure 2 shows the trend of trademark applications during 2019 and 2020 until the first twenty days of June. By comparing Figures 1 and 2, we see that the trends between the Report and Rendered data are very similar up until May. While there is an obvious decrease in trademark applications from the Rendered data, since there is a third of June missing, the 2020 filings exceed the 2019 filings consistent with Figure 1.

Figure 2. Trademark application trends (January-first twenty days of June) based on Rendered data and the 55 countries.



Notes: Marks2019 and Marks2020 stand for trademark applications filed in a given month in 2019 and 2020 respectively.

Finally, we compare the shares by country between the Report data and the Rendered data. In other words, we calculate for each country its share of patent applications over the total number of applications for 2019 and 2020 respectively. For both years, the correlations between the two shares are 0.99.[4]

At this point it is interesting to observe that application rates in 2020 are decreasing in March and April and then begin to increase in May and June relatively to 2019. This difference

[4] We also calculate the deviation in shares for 2019 and 2020 by country. In additional analysis available upon request, we find that for most countries the deviations are either small or close to zero indicating that the sample is representative of the population of data. In case a country is under-represented or over-presented it does so in both years (e.g. AU); therefore, our findings should not be biased due to this mis-representation. For a few countries where the deviation is large any results discussed in the study highlight explicitly such deviation.

Covid Economics 43, 21 August 2020: 1-18

in application rates could be due to the pandemic. Firms may have postponed their applications during these early months, where for many EU countries, COVID-19 peaked. Starting May, these entities are optimistic of future demand thereby filing for an increased number of trademark applications. While different from our study, such behaviour seems to be consistent with the stock market reaction of COVID-19 (Capelle-Blancard and Desroziers 2020).

3. Empirical Findings

3.1. Trends by Country

Our interest is on the last two months of the observed period: May and June of 2020. Note that whenever we refer to June, by default we refer to the first twenty days of June. To account for seasonality, we compare trademark filings of these two months with the same months of 2019. We complement these data with data on COVID-related cases and deaths as well as social distancing measures imposed by each country by Our World in Data COVID-19 dataset.[5] Our measure of interest is the following:

$$\frac{TrademarkApplications_{i,MayJune,2020} - TrademarkApplications_{i,MayJune,2019}}{TrademarkApplications_{i,MayJune,2019}} x100$$

Where $TrademarkApplications_{i,MayJune,2020}$ is the number of trademark applications filed by country i during May and first twenty days of June 2020. A similar definition applies for $TrademarkApplications_{i,MayJune,2019}$.[6]

Figure 3 graphs this measure. China appears to be an outlier increasing substantially its trademark applications compared to 2019. Two not-so-close seconds are Italy and Germany while many European countries are close to zero indicating moderate decrease from last year. Certain countries however face clear negative trends such as Brazil and Canada.[7] We should note that back-of-the-envelope calculations based on the EUIPO report corroborate these findings. For instance, China has already filed in the first semester of 2020, 66.4% of the total

[5] https://ourworldindata.org/coronavirus-source-data

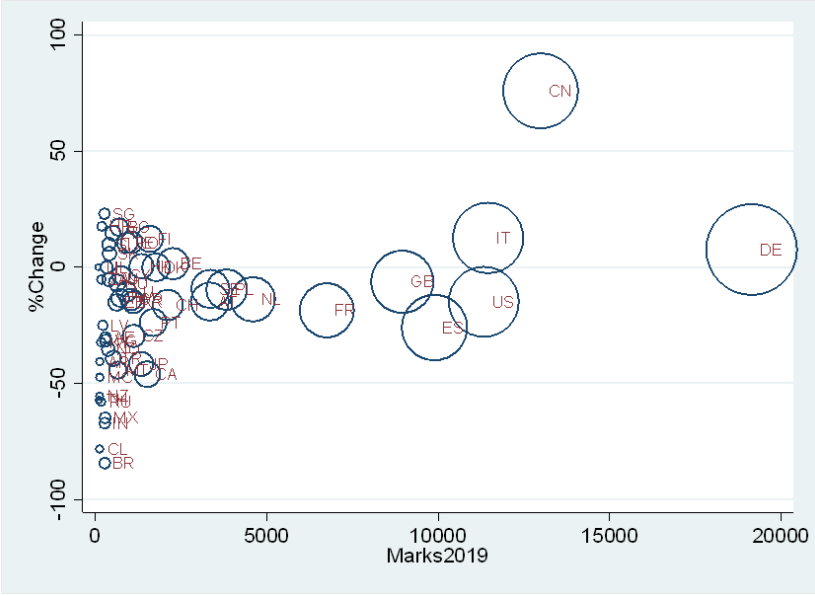
[6] In additional analysis, available upon request, we drew the same Figures by examining either May or June filings separately. Results are qualitatively similar.

[7] In additional analysis, we exclude trademark applications that claim apriority date more than ten days earlier than the filing date at the EUIPO. Approximately, thirteen thousand applications were dropped. Counterpart of Figure 3 is displayed in Figure A2 of the Appendix. Results are qualitatively similar.

Covid Economics 43, 21 August 2020: 1-18

trademark applications it filed in 2019 while Italy 54.8%. Conversely, these figures for Brazil and Canada are 32.4% and 40.2% respectively. [8]

Figure 3. Change in trademark applications between May-June 2020 and May-June 2019.



Notes: The y-axis is calculated as: $(TrademarkApplications_{i,MayJune,2020} - TrademarkApplications_{i,MayJune,2019})/TrademarkApplications_{i,MayJune,2019}$ where i is the country. We only consider applications that were filed in May and the first twenty days of June of each year. Note that this measure is multiplied by a 100. The size of the bubble is associated with the number of trademark applications country i has filed in 2019. The x-axis is also the the number of trademark applications country i has filed in 2019.

It is interesting to also examine whether these changes are correlated with COVID-related deaths and social distancing. Based on the Our World in Data COVID-19 dataset, we calculate the total deaths per capita, multiplied by a thousand, as of April 30 as well as the average stringency index as of the same date. With respect to deaths there is no clear correlation (ρ=0.007). With respect to stringency, there is a weak positive correlation (ρ=0.2). [9] In Figure 4, we plot the latter relationship for a more detailed inspection but no clear pattern emerges. For instance, both Poland and Sweden were resilient in their trademark applications with vastly

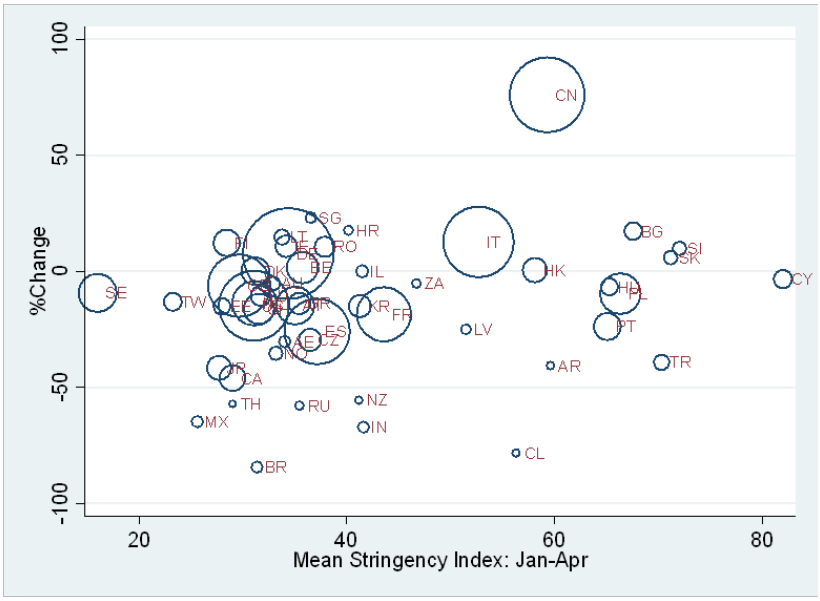
[8] Japan is a special case in our data. It is under-represented in the Rendered data compared to the Report data. However, any bias should be small because it is under-represented equally in both years. If we were to expand the sample dating from January, we find similar shares for this country both in the Report and Rendered data. However, in May and June, trademark applications of Japan have had a sharp decrease. Whether this trend will continue, it remains to be seen.

[9] For deaths, data are missing for HK while for the stringency index for KY, LI, MC, MT and VG.

Covid Economics 43, 21 August 2020: 1-18

different social distancing measures while China with some of the strongest measures experienced a dramatic increase.

Figure 4. Change in trademark applications comapred to social distancing measres.



Notes: The y-axis is calculated as: $(TrademarkApplications_{i,MayJune,2020} - TrademarkApplications_{i,MayJune,2019})/TrademarkApplications_{i,MayJune,t}$ where i is the country. We only consider applications that were filed in May and the first twenty days of June of each year. Note that this measure is multiplied by a 100. The size of the bubble is associated with the number of trademark applications country i has filed in 2019. The x-axis is the average Stringency index based on the Our World in Data COVID-19 dataset measure as of April 30th.

3.2. Trends by Firm Type

We decompose trademark applications based on whether they were filed by a new entrant or an established firm. A trademark application in a given month is categorized as an application by an entrant if the applicant is observed for the first time since January 1, 2015. We should note that we base these computations based on the EUIPO's applicant id variable. It is of course plausible that certain firms may have more than one id. To alleviate some of this bias, we restrict that an entrant should not apply within the first month of observation more than five trademark applications. This way we ensure that this is not an established firm filing for trademark applications for the first time under a different id. For Figure 5 our measures of interest are on the following shares:

$$EntrantApplications_{m,2020}/TotalApplications_{m,2020}$$

$$EntrantApplications_{m,2019}/TotalApplications_{m,2019}$$

Covid Economics 43, 21 August 2020: 1-18

CEPR PRESS

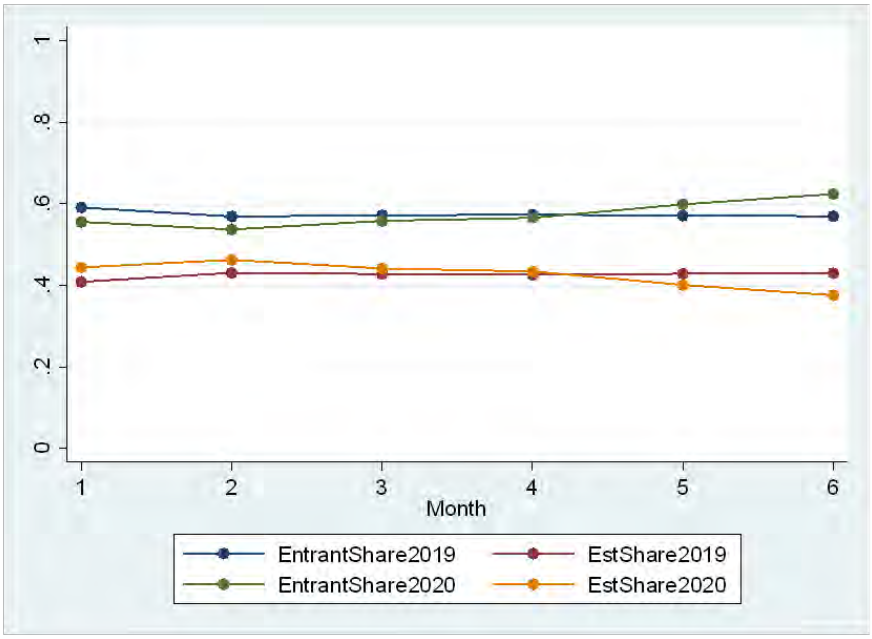
$EstablishedApplications_{m,2020}/TotalApplications_{m,2020}$

$EstablishedApplications_{m,2019}/TotalApplications_{m,2019}$

Where $EntrantApplications_{m,2020}$ is the number of trademark applications by entrants at month m in 2020, $TotalApplications_{m,2020}$ is the total number of trademark applications at month m in 2020, $EstablishedApplications_{m,2020}$ is the number of trademark applications by established firms at month m in 2020. Similar definitions apply to the rest of the variables.

As can be seen. The share of applications by entrants increases over the last two months compared to 2019. The opposite occurs for established firms. The difference is not big, but it does seem to indicate that new firms enter the EU market with new offerings. The major question however is whether this trend will continue over the next months.

Figure 5. Shares of entrants and established firms for 2019 and 2020.



Notes: *EntrantShare2019* = $EntrantApplications_{m,2019}/TotalApplications_{m,2019}$ *EstShare2019*=$EstablishedApplications_{m,2019}/TotalApplications_{m,2019}$. Similarly, for *EntrantShare2020* and *EstShare2020*.

It is interesting to examine whether there is heterogeneity by country. We focus on applications by entrants and calculate the following measure:

Covid Economics 43, 21 August 2020: 1-18

$$\frac{\frac{EntrantApplications_{i,MayJune,2020}}{TotalApplications_{i,MayJune,2020}} - \frac{EntrantApplications_{i,MayJune,2019}}{TotalApplications_{i,MayJune,2019}}}{\frac{EntrantApplications_{i,MayJune,2019}}{TotalApplications_{i,MayJune,2019}}} x100$$

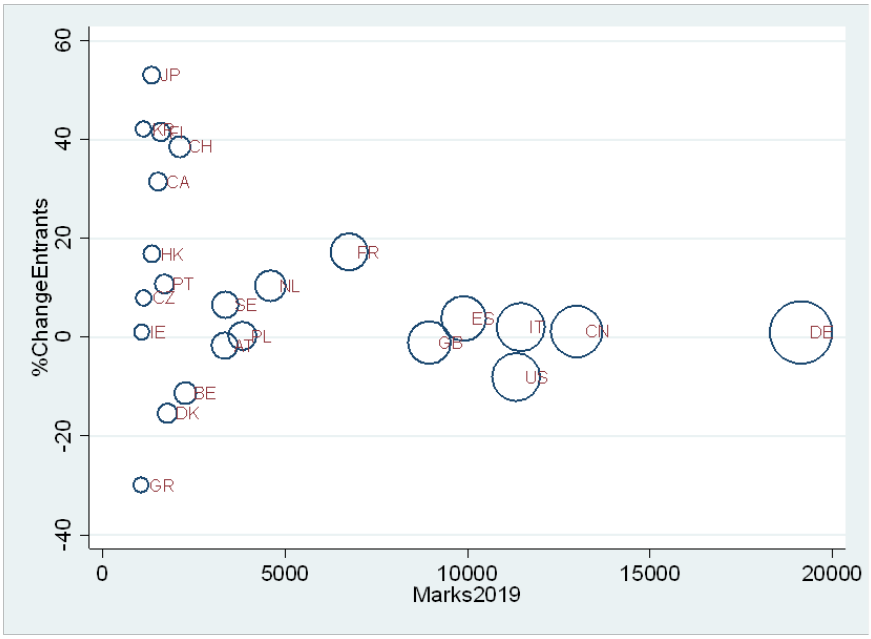
Where $EntrantApplications_{i,MayJune,2020}$ is the number of applications filed by entrants of country *i* during May-June of 2020. $TotalApplications_{i,MayJune,2020}$ is the total number of applications filed country *i* during May-June of 2020. A similar definition applies to the rest of the variables. This measure captures the change in the share of applications by entrants by country between 2019 and 2020.

Figure 6 presents this measure over the number of applications filed by each country in 2019. Note that we display countries with more than 1,000 trademark applications filed during 2019. Most of the larger countries' share change is close to zero. For smaller country applicants however there appears to be heterogeneity. These results should be viewed with caution as the number of applications by country is relatively small. However, they are still informative on which firms in each country appear to be picking up the pace after the first wave of the pandemic. For instance, Greece filed for approximately 120 applications during May-June of 2020. A few established firms, primarily in the agrifood sector, filed for applications in bulk during this season with a single firm filing for seven applications. For France entrants appear to be picking up the pace and file for more applications this year.

Covid Economics 43, 21 August 2020: 1-18

Covid Economics 43, 21 August 2020: 1-18

Figure 6. Change in the share of applications by entrants over May-June by country.



Notes: The variable in the y-axis is calculated as:

$$\frac{\frac{EntrantApplications_{i,MayJune,2020}}{TotalApplications_{i,MayJune,2020}} - \frac{EntrantApplications_{i,MayJune,2019}}{TotalApplications_{i,MayJune,2019}}}{\frac{EntrantApplications_{i,MayJune,2019}}{TotalApplications_{i,MayJune,2019}}} \text{x100}$$

The size of the bubble is associated with the number of trademark applications country *i* has filed in 2019. The x-axis is also the the number of trademark applications country *i* has filed in 2019.

3.3. Trends by Nice Class

The trademark system has a feature that makes it attractive for sectoral analysis. For a trademark application to be registered it needs to disclose at least one Nice class. The Nice classification system categorizes the entire business spectrum of products and services in 45 classes.[10] For a trademark applicant to claim that embedded figure, logo or any other differentiating feature in a particular product/service area she needs to have explicitly claimed the particular Nice class. To register each class, the applicant needs to show that she will be using it in commerce in each said class. Classes 1-34 pertain to products while classes 35-45 to services. For an example of a sectoral analysis based on Nice classes see Mendonça (2014).

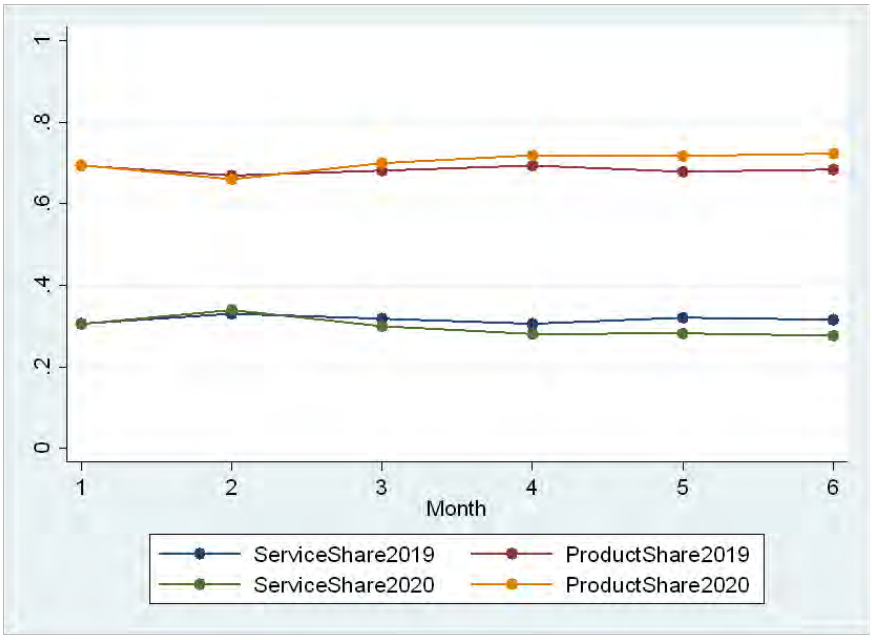
We first examine the share of service-related classes filed in 2020 compared to last year. To compute such share, we first need to count the instances where each Nice class is disclosed. Note that in our 2019-2020 sample, the average trademark application discloses 2.56 Nice

[10] http://www.wipo.int/classifications/nice/en/

classes. Therefore, we weigh each trademark application based on the number of classes it has. For instance, if a trademark application discloses only Nice class n_1, then we count 1 for class n_1. However, if trademark application discloses Nice classes n_1, n_2 and n_3, then for class n_1 we count 0.33. After this weighting scheme, we add all these counts for service-related classes and divide over the total number of trademark applications. We denote this share as *ServiceShare* for a given month and year. Similarly, we compute a *ProductShare*. Note that via the weighting scheme *ServiceShare*+*ProductShare*=1.

Figure 7 compares these shares by month for 2019 and 2020. As can be seen, there is small increase in the *ProductShare* in 2020 and an analogous decrease in the *ServiceShare*. This is intuitive as many services were hit hard by the COVID pandemic while a number of firms have introduced new products related to the implications of the pandemic and social distancing.

Figure 7. Compare shares of service-related and product-related classes for 2019 and 2020.



Notes: *ServiceShare2019* is the share of service Nice classes in a given month in 2019 while *ServiceShare2020* is the share of service Nice classes in a given month in 2020. Similar definitions apply for *ProductShare2019* and *ProductShare2020*.

To generate a better picture, we compare the share of each Nice Class separately and how that changed between May-June of 2019 to May-June of 2020. The difference can be stated as:

Covid Economics 43, 21 August 2020: 1-18

$$\frac{WApplications_{n,MayJune,2020}}{TotalApplications_{MayJune,2020}} - \frac{WApplications_{n,MayJune,2019}}{TotalApplications_{MayJune,2019}}$$

Where $WApplications_{n,MayJune,2020}$ is the number of applications (weighted) that disclose class *n* filed during May-June of 2020. $TotalApplications_{MayJune,2020}$ is the total number of applications filed during May-June of 2020. A similar definition applies to the rest of the variables.

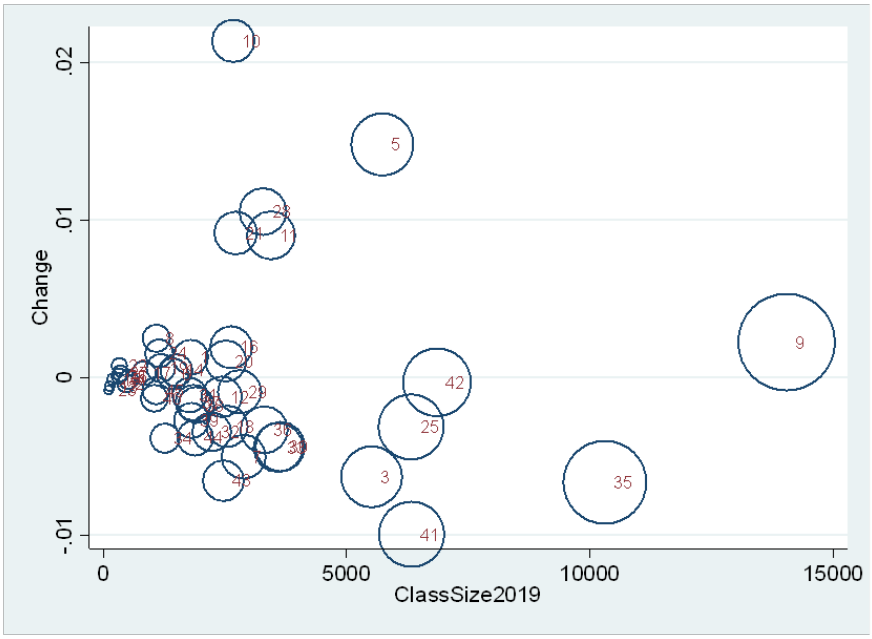
Figure 8 shows this change for each class. The size of the bubble displays the relative size of each class calculated as the total number of (weighted) applications that disclose class *n* filed during 2019. While for most classes the change is close to zero, there are some clear winners and losers. Classes 5 and 10 that pertain pharmaceutical products and medical equipment have unsurprisingly the largest increase.[11] On the contrary, class 41 which is related to entertainment services has experienced the largest decrease.[12] Interestingly class 3 which pertains to cleaning and cosmetic products has experienced a decrease.[13]

Covid Economics 43, 21 August 2020: 1-18

[11] Class 5: Pharmaceuticals, medical and veterinary preparations; sanitary preparations for medical purposes; dietetic food and substances adapted for medical or veterinary use, food for babies; dietary supplements for humans and animals; plasters, materials for dressings; material for stopping teeth, dental wax; disinfectants; preparations for destroying vermin; fungicides, herbicides.
Class 10: Surgical, medical, dental and veterinary apparatus and instruments; artificial limbs, eyes and teeth; orthopaedic articles; suture materials; therapeutic and assistive devices adapted for the disabled; massage apparatus; apparatus, devices and articles for nursing infants; sexual activity apparatus, devices and articles.
[12] Class 41: Education; providing of training; entertainment; sporting and cultural activities
[13] Class 3: Bleaching preparations and other substances for laundry use; cleaning, polishing, scouring and abrasive preparations; non-medicated soaps; perfumery, essential oils, non-medicated cosmetics, non-medicated hair lotions; non-medicated dentifrices

Covid Economics 43, 21 August 2020: 1-18

Figure 8. Change in the Nice classes claimed from May-June 2019 to May-June 2020.



Notes: The y-axis Change variable is calculated as $\frac{WApplications_{n,MayJune,2020}}{TotalApplications_{MayJune,2020}} - \frac{WApplications_{n,MayJune,2019}}{TotalApplications_{MayJune,2019}}$ where $WApplications_{n,MayJune,2020}$ is the number of applications (weighted) that disclose class n filed during May-June of 2020. $TotalApplications_{MayJune,2020}$ is the total number of applications filed during May-June of 2020. A similar definition applies to the rest of the variables. The size of the bubble and the x-axis considers the frequency class n is disclosed during 2019.

Overall, this analysis at the Nice class level reveals interesting patterns in the short-run after the pandemic. Whether some of these trends will last and perhaps induce firms to seek new opportunities is difficult to predict. Nonetheless, the patterns thus far show that firms focus on certain product endeavours shying away from service-related endeavours that at this point in time are related with higher uncertainty.[14]

4. Conclusion

Trademark applications have been shown to contribute to firm value and regional resilience. They can approximate innovative activity via a variety of mechanisms. Further, they are linked to actual commercial activity as prescribed in their institutional origins. Importantly, in a time where fast dissemination of data is crucial, trademark applications can provide significant insights. Unlike a patent application that requires 18 months to be published,

[14] In additional analysis we consider trademark applications by entrants and established firms separately. There is some heterogeneity with class 3 as an interesting case. While entrants have increased their offerings in this class, established firms have decreased them. One could only speculate this different trend; for established firms it could be the case that they focus on existing products of high demand without the need to introduce different versions of the same product.

information on trademark applications is disclosed much sooner. To this end, acquiring bulk information from other Offices around the world can provide significant insights on the effects of COVID-19 on economic activity. While by no means the tell-all of business activity, they can be helpful until more data become available.

Motivated by the EUIPO's report on an upward trend in trademark applications in 2020 compared to the same season of 2019, this study rendered the bulk of trademark application data. Whether the three major findings that we observe will persist as we enter in the Fall where scientists predict a second wave of the pandemic remains to be seen.

References

Block JH, De Vries G, Schumann JH, et al. (2014a) Trademarks and venture capital valuation. *Journal of Business Venturing* 29(4): 525–542.

Capelle-Blancard, G and A Desroziers (2020), "The stock market is not the economy? Insights from the COVID-19 crisis", Covid Economics: Vetted and Real-Time Papers, CEPR.

Castaldi, C. (2020). All the great things you can do with trademark data: Taking stock and looking ahead. *Strategic Organization*, *18*(3), 472-484.

Chemmanur, T. J., Rajaiya, H., Tian, X., & Yu, Q. (2018). Trademarks in Entrepreneurial Finance. Available at SSRN 3183879.

Drivas, K. (2020). The Role of Technology and Relatedness in Regional Trademark Activity. *Regional Studies*, Forthcoming in special issue *"Trademarks in space"*.

Filippetti, A., Gkotsis, P., Vezzani, A., & Zinilli, A. (2019). How to survive an economic crisis? Lessons from the innovation profiles of EU regions.

Giarratana MS and Fosfuri A (2007) Product strategies and survival in Schumpeterian environments: Evidence from the US security software industry. *Organization Studies* 28(6): 909–929.

EUIPO 2020. "EUIPO Statistics in European Union Trade Marks 1996-01 to 2020-06 Evolution." Available at: https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/contentPdfs/about_euipo/the_office/statistics-of-european-union-trade-marks_en.pdf
Accessed on August 1, 2020.

Hoenen, S., Kolympiris, C., Schoenmakers, W., & Kalaitzandonakes, N. (2014). The diminishing signaling value of patents between early rounds of venture capital financing. *Research Policy*, 43(6), 956-989.

Link AN and Scott JT (2012) Employment growth from the small business innovation research program. *Small Business Economics* 39(2): 265–287.

Mendonça, S., Pereira, T. S., & Godinho, M. M. (2004). Trademarks as an indicator of innovation and industrial change. *Research Policy, 33*(9), 1385-1404.

Covid Economics 43, 21 August 2020: 1-18

Mendonça, S. (2014). National adaptive advantages: Soft innovation and marketing capabilities in periods of crisis and change Structural Change, Competitiveness and Industrial Policy (pp. 149-166): Routledge.

Flikkema, M., Castaldi, C., de Man, A.-P., & Seip, M. (2019). Trademarks' relatedness to product and service innovation: A branding strategy approach. *Research Policy*.

Sandner PG and Block J (2011) The market value of R&D, patents, and trademarks. *Research Policy* 40(7): 969–985.

Srinivasan R, Lilien GL and Rangaswamy A (2008) Survival of high tech firms: The effects of diversity of product–market portfolios, patents, and trademarks. *International Journal of Research in Marketing* 25(2): 119–128.

Zhou H, Sandner PG, Martinelli SL, et al. (2016) Patents, trademarks, and their complementarity in venture capital funding. *Technovation* 47: 14–22.

Covid Economics 43, 21 August 2020: 1-18

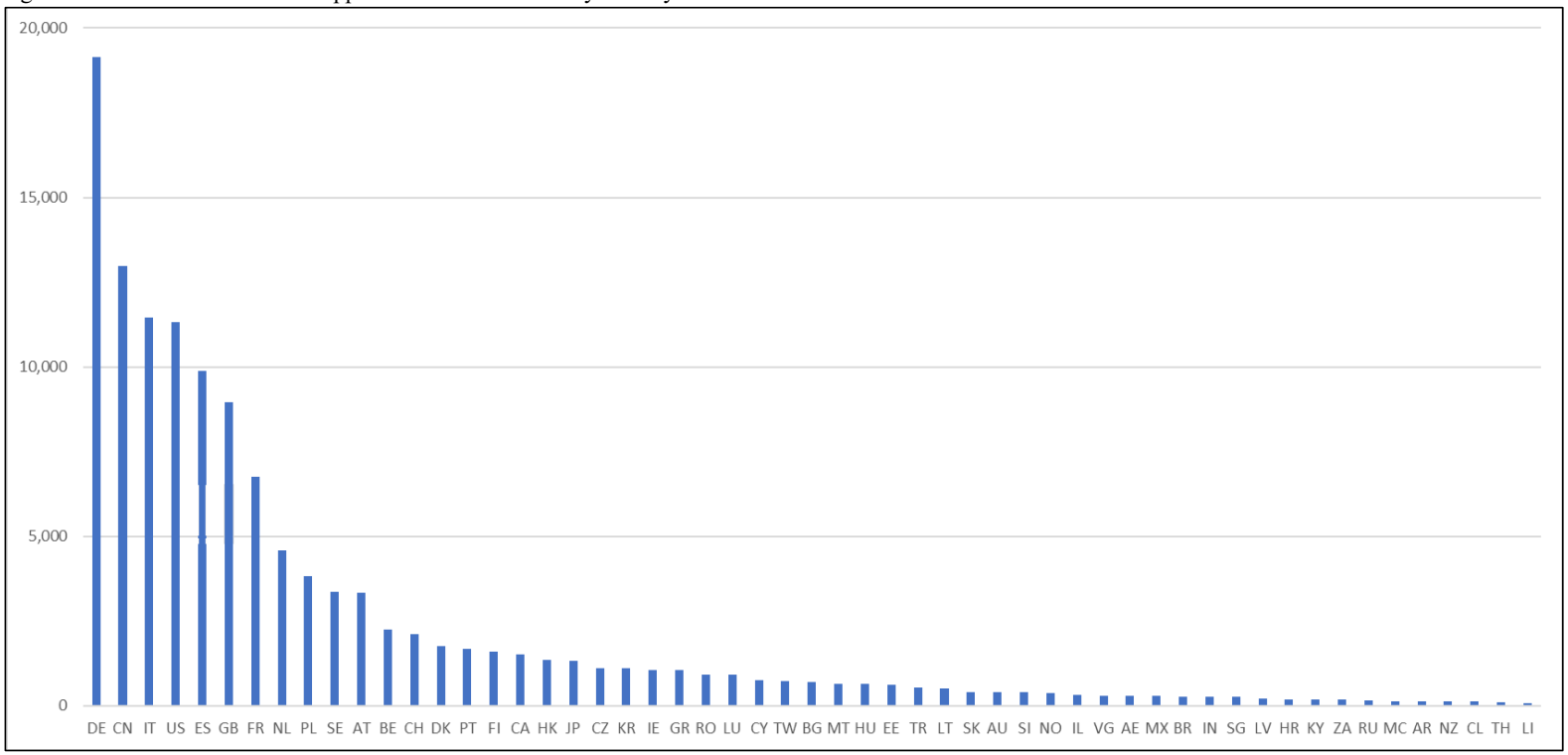
CEPR PRESS

17

Covid Economics 43, 21 August 2020: 1-18

Appendix

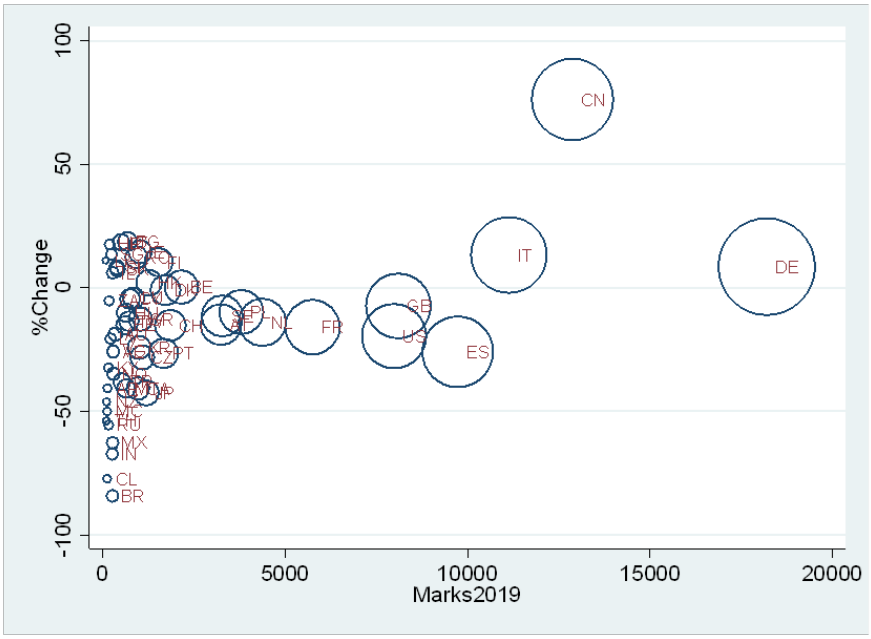
Figure A1. Number of trademark applications filed in 2009 by country.



18

Covid Economics 43, 21 August 2020: 1-18

Figure A2. Counterpart of Figure 3. Exclude trademark applications with a priority date more than ten days earlier than the filing date.



Notes: The y-axis is calculated as: $(TrademarkApplications_{i,MayJune,2020} - TrademarkApplications_{i,MayJune,2019})/TrademarkApplications_{i,MayJune,2019}$ where i is the country. We only consider applications that were filed in May and the first twenty days of June of each year and they did not disclose a priority data ten days earlier than the EUIPO filing date. Note that this measure is multiplied by a 100. The size of the bubble is associated with the number of trademark applications country i has filed in 2019. The x-axis is also the number of trademark applications country i has filed in 2019.

Covid Economics Issue 43, 21 August 2020

Estimating the prevalence of the COVID-19 infection, with an application to Italy[1]

Franco Peracchi[2] and Daniele Terlizzese[3]

Date submitted: 11 August 2020; Date accepted: 18 August 2020

Knowing the prevalence of the COVID-19 infection in a population of interest, and how it changes over time and across space, is of fundamental importance for public health. Unfortunately, the fraction of cases who turn out to be positive in a test provides a distorted picture of the prevalence of the infection because the tested cases are not a random sample of the population. Since random testing of the population is costly and complicated to carry out, in this note we show how to use the available information, in conjunction with credible assumptions about unknown quantities, to obtain a range of plausible values for the prevalence of the infection. We discuss the difference between two alternative measures of prevalence and argue that one of the two is much harder to pin down with the data currently available. We apply our method to the Italian data.

Covid Economics 43, 21 August 2020: 19-41

1 We thank Giulia Faggio, Luigi Guiso, Chuck Manski, Francesca Molinari, and Antonia Stazi for helpful comments. We also thank the press office of Dipartimento della Protezione Civile and officers at the Italian Ministry of Health for detailed information about the COVID-19 data.

2 Professor of Econometrics, University of Rome Tor Vergata; Fellow, Einaudi Institute for Economics and Finance (EIEF); Professor of the Practice, Georgetown University.

3 Director, Einaudi Institute for Economics and Finance (EIEF).

Covid Economics 43, 21 August 2020: 19-41

1 Introduction

Knowing the prevalence of the COVID-19 infection in a population of interest, and how it changes over time and across space, is of fundamental importance for public health. Unfortunately, the fraction of people who are confirmed positive to COVID-19, reported daily in different countries by the agencies monitoring the evolution of the pandemics, provides a biased picture of the prevalence of the infection, since people tested for the presence of the virus are not a random sample of the population. In fact, tests are mainly done on subjects with severe enough symptoms or high likelihood of a contact with infected cases, and the pool of those tested grossly underrepresents the asymptomatic and those with only mild symptoms. As the latter are a powerful vehicle of contagion, the lack of information on their prevalence is a serious hindrance to the design of effective containment policies.

This is of course well known (see, e.g., Li et al., 2020 and Stock, 2020), and it is understandable that, in the heat of the crisis, attention is focused on saving lives rather than having a statistically accurate representation of the phenomenon. However, the correct measurement of the prevalence of the infection is starting to be perceived as an essential input in planning a gradual removal of lockdown policies. For example, the Italian national statistical agency (Istituto Nazionale di Statistica, or Istat) just launched a relatively large survey (with a target sample size of 150 thousands individuals) to ascertain, through serological testing, the proportion of the population that has been infected by the virus. But these surveys are costly, complicated to carry out, take time to complete and may be plagued by (non random) refusal to participate. It is therefore of some value to come up with easy to implement, though necessarily less accurate, estimates of the prevalence of the infection in the population based on readily available data.

Indeed, there are two somewhat different concepts of prevalence. One is point prevalence, namely the fraction of people in the population who, *at a given point in time*, are infected; they might have contracted the infection in a previous date, but are still infected at the point in time we are considering. The other is period prevalence, namely the fraction of people in the population who have been infected during a given period; in particular, if the period is from the beginning of the pandemic to the current moment, period prevalence measures the fraction of people in the population who were *ever* infected, even though at the present date might no longer be – because recovered or dead. Point prevalence gives information on the evolution of the contagion since it measures the fraction of people who are still infectious. It is therefore more useful to monitor the risk of transmitting the infection. Period prevalence is instead more useful to monitor the risk of being infected – since it readily gives the fraction of people susceptible of being infected – and for reporting purposes, for example to correctly compute the case fatality ratio or to assess

Covid Economics 43, 21 August 2020: 19-41

the productivity of the resources deployed to fight the infection. Depending on the nature of the tests conducted, different surveys would measure one or the other concept of prevalence: roughly speaking, a survey based on nasopharyngeal swabs, tested for the presence of the virus, aims at measuring point prevalence, while a survey based on serological tests, that detect the antibodies produced in response to the infection, aims at measuring period prevalence.

In this short paper we present an approach to obtain estimates of point prevalence based on readily available data and minimal, transparent, and easy to interpret assumptions. A recent paper by Manski and Molinari (2020) has a similar goal and approach but focuses on period prevalence. Since both concepts of prevalence are interesting and useful, the two papers nicely complement each other. We then apply our approach to data from Italy, both for the country as a whole and for its different regions.

Estimates of the point or period prevalence have also been produced in the context of structural or semi-structural models of the epidemic (see for example Flaxman et al., 2020 and Vollmer et al., 2020). The main advantage of these models is their stylized description of some of the mechanisms underlying the epidemic. However, the resulting estimates typically require strong and untested parametric assumptions, and the complexity of these models makes the mapping between the assumptions and the estimates more opaque.

The paper proceeds as follows. Section 2 outlines our approach. Section 3 offers a brief description of the data we use. Section 4 presents our empirical results. Section 5 clarifies in a simplified setting the links between point and period prevalence, and Section 6 briefly concludes.

2 Framework

To avoid complicating the analysis, we assume that each person can have at most one episode of infection (after which she either recovers or dies), that is infectious as long as she is infected, and that the recovered are immune and can no longer transmit the infection. These assumptions, though not strictly true, are close enough to what is known about the COVID-19 pandemic. For the sake of simplicity, we also neglect the effect of deaths and new births on the population size, which we take as constant. Time is measured in days.

2.1 Notation

Let I_t be a binary random variable that is equal to one if a person is infected on day t, and is equal to zero otherwise. An infected person on day t may have become infected before t, and some of those who became infected before t may have recovered (or died) by day t; for them $I_t = 0$ would hold. The quantity that we want to estimate is $\Pr(I_t = 1)$, which we interpret as the population fraction of individuals who on day t are infected, and therefore can infect others.

Covid Economics 43, 21 August 2020: 19-41

Also, let T_t be a binary random variable that is equal to one if a person has been first tested on day t for the presence of the virus through a nasopharyngeal swab, and is equal to zero otherwise. Notice that, if $T_t = 1$, then $T_{t+k} = 0$ for all $k \geq 1$. Further, $T_t = 0$ if either a person was first tested at some previous date or was never tested. We interpret $\Pr(T_t = 1)$ as the population fraction of individuals who were first tested on day t.

Finally, let P_t be a binary random variable that is equal to one if a person has first received a positive test result on day t, and is equal to zero otherwise. Hence, if $P_t = 1$, then $P_{t+k} = 0$ for all $k \geq 1$. Further, $P_t = 0$ if either a person was first tested on day t with a negative test result, or was not first tested on that day, i.e., was first tested on some previous day or was never tested.

It is convenient to assume that people are tested at most once, and that test results are available the same day when the swabs are taken. If follows from these two assumptions that $P_t = 1$ implies $T_t = 1$, that is, those who first receive a positive test result on day t must have been first tested on that day.[1] Although neither assumption is exactly true, we will discuss in Section (3) how to transform the data to get a reasonable approximation.

2.2 The estimating equation

To simplify the notation, in this section we drop the time subscript t. By the Law of Total Probability, we can write

$$\Pr(I = 1) = \Pr(I = 1|T = 1)\ \Pr(T = 1) + \Pr(I = 1|T = 0)(1 - \Pr(T = 1)). \tag{1}$$

Since $\Pr(T = 1)$ is directly measured in the data, to estimate $\Pr(I = 1)$ we need information about $\Pr(I = 1|T = 1)$ and $\Pr(I = 1|T = 0)$. Let us consider these two probabilities separately.

As for $\Pr(I = 1|T = 1)$, we can exploit the available information on the operational properties of the tests conducted on nasopharyngeal swabs, namely their Type-I and Type-II error probabilities. There seems to be widespread consensus that these tests have a probability of Type-I error – the probability of a false positive – that is very close to zero.[2] We therefore assume that

$$\Pr(P = 1|I = 0, T = 1) = 0. \tag{2}$$

This assumption has two important implications. The first is that

$$\Pr(I = 1|P = 1) = 1, \tag{3}$$

[1] The assumption that nobody is tested more than once rules out the case of a person who is first tested on day t, resulting negative, and then retested at a later date, resulting positive for the first time. This case would contradict the implication claimed in the text.

[2] In the Italian case, for a period at the beginning of the pandemic, all swabs found positive by regional labs were retested by the Istituto Superiore di Sanità (the Italian equivalent of the National Institute of Health) and always confirmed. This reinforces the assumption of zero probability of a false positive.

Covid Economics 43, 21 August 2020: 19-41

which follows directly from Bayes Law and the already established implication that $T = 1$ if $P = 1$. The other is that

$$\Pr(I = 1|T = 1) = \frac{\Pr(P = 1|T = 1)}{1 - \beta}, \tag{4}$$

where $\beta = \Pr(P = 0|I = 1, T = 1)$ is the probability of Type-II error – the probability of a false negative. To see this, notice that

$$\begin{aligned}
\frac{\Pr(P = 1|T = 1)}{1 - \beta} &= \frac{\Pr(P = 1|T = 1)}{\Pr(P = 1|I = 1, T = 1)} \\
&= \frac{\Pr(P = 1, T = 1)}{\Pr(T = 1)} \times \frac{\Pr(I = 1, T = 1)}{\Pr(I = 1, T = 1, P = 1)} \\
&= \frac{\Pr(P = 1)}{\Pr(T = 1)} \times \frac{\Pr(I = 1, T = 1)}{\Pr(I = 1, P = 1)} \\
&= \frac{\Pr(I = 1|T = 1)}{\Pr(I = 1|P = 1)} \\
&= \Pr(I = 1|T = 1),
\end{aligned}$$

where the first equality uses the definition of β, the second and the fourth follow from the rules of conditional probabilities, the third from the fact that $T = 1$ if $P = 1$, which implies that $\Pr(P = 1, T = 1) = \Pr(P = 1)$ and $\Pr(I = 1, T = 1, P = 1) = \Pr(I = 1, P = 1)$, and the last from (3).

Equation (4) is important because it expresses $\Pr(I = 1|T = 1)$ as a function of quantities for which we either have direct measures, namely $\Pr(P = 1|T = 1)$, or we can make an educated guess based on medical knowledge, namely β. Although estimating the probability of a false negative is not easy, there is a general consensus in the medical profession that it is not negligible and largely reflects issues with specimen collection; specifically, the sample might have been collected too early or too late, might be contaminated, or might have been stored for too long.[3] A review of the available health literature suggests a range of values for β between .02 and .4, though a narrower range between .1 and .3 is more often quoted.[4] To the extent that β reflects practical issues in the implementation of the test, one might expect some time variability due to learning. However, there is little evidence of this and we will assume that β is constant.

As for $\Pr(I = 1|T = 0)$, we rely on information about the testing process and assume that

$$\Pr(I = 1|T = 0) \leq \Pr(I = 1|T = 1). \tag{5}$$

This is because the subjects with $T = 0$ on day t are those who either have not been tested yet or have been tested at an earlier date. Because nasopharyngeal swabs are mainly taken from subjects

[3] https://asm.org/Articles/2020/April/False-Negatives-and-Reinfections-the-Challenges-of.

[4] See for example Watson et al., 2020. Also see https://www.medpagetoday.com/infectiousdisease/covid19/86047, https://www.healthline.com/health-news/false-negatives-covid19-tests-symptoms-assume-you-have-illness, and https://theconversation.com/coronavirus-tests-are-pretty-accurate-but-far-from-perfect-136671.

who have visible symptoms, or are suspected of exposure to the infection, those who have not been tested are clearly less likely to be infected as of day t; moreover, among those who have been tested at an earlier date, some will have recovered and are no longer infected. Therefore, it seems reasonable to assume that subjects with $T_t = 0$ are on average less likely to be infected than those with $T_t = 1$.[5] Thus, we define

$$\lambda = \frac{\Pr(I = 1|T = 0)}{\Pr(I = 1|T = 1)},$$

and assume that λ ranges between 0 and 1. If people are tested at random, the infection rate is the same among the tested and the untested, and therefore $\lambda = 1$. If the tested sample is instead biased towards the symptomatic (or, more generally, towards those with higher infection risk), then $\lambda < 1$.

Putting it all together we can rewrite (1) as

$$\Pr(I = 1) = \frac{\Pr(P = 1|T = 1)}{1 - \beta} \left[\Pr(T = 1) + \lambda \left(1 - \Pr(T = 1)\right)\right]. \tag{6}$$

This equation is our basis for estimating $\Pr(I = 1)$ as a function of λ and β given knowledge of two other quantities that we can measure in the data, namely $\Pr(P = 1|T = 1)$ and $\Pr(T = 1)$.

2.3 The anatomy of λ

As already mentioned, we can bring information obtained from medical expertise and testing practice to bear on β. The natural question is whether something similar can be done about λ. To do this, it is useful to derive an expression for λ in terms of quantities that are potentially observable.

We distinguish between symptomatic ($S = 1$) and asymptomatic ($S = 0$) cases, where the former are those who show some of the symptoms associated with the infection and the latter show none of them, and let $\Pr(I = 1|S = s)$, $s = 0, 1$, be the fraction of the infected among the two types. It is reasonable to assume that the infection rate is higher among the symptomatic, so we posit

$$\mu = \frac{\Pr(I = 1|S = 1)}{\Pr(I = 1|S = 0)} \geq 1.$$

Let $n = \Pr(T = 1)$ be the fraction of the population that is tested, $\pi = \Pr(S = 1)$ the fraction of the population that is symptomatic, and $p = \Pr(S = 1|T = 1)$ the fraction of the symptomatic among the tested. Therefore $\gamma = p/\pi$ is a measure of the bias implicit in the testing protocol, which we know targets disproportionally symptomatic subjects. This implies that $\gamma \geq 1$.

By the Law of Total Probability, we can write the infection rate among the tested cases as

$$\Pr(I = 1|T = 1) = \Pr(I = 1|S = 1, T = 1)\, p + \Pr(I = 1|S = 0, T = 1)(1 - p).$$

[5] Assumption (5) is similar to the monotonic testing assumption in Manski and Molinari (2020), though we refer here to point prevalence while they refer to period prevalence.

Covid Economics 43, 21 August 2020: 19-41

Covid Economics 43, 21 August 2020: 19-41

We know that the bias in the testing protocol is mainly due to the higher proportion of symptomatic case been tested; to obtain a sharp characterization, we then make the extreme assumption that the bias is entirely due to the oversampling of the symptomatic subjects. Specifically, we assume that, conditional on symptomatology, the infection rate is the same in the tested sample and the general population.[6] Formally, we assume

$$\Pr(I=1|S=s,T=1)=\Pr(I=1|S=s),\quad s=0,1. \tag{7}$$

Under this assumption

$$\begin{aligned}\Pr(I=1|T=1)&=\Pr(I=1|S=1)\,p+\Pr(I=1|S=0)(1-p)\\&=[(\mu-1)p+1]\,\Pr(I=1|S=0),\end{aligned}$$

where we used the definition of μ. Among the untested, the infection rate is instead

$$\begin{aligned}\Pr(I=1|T=0)&=\frac{\Pr(I=1)-\Pr(I=1,T=1)}{\Pr(T=0)}\\&=\frac{\Pr(I=1|S=1)\,\pi+\Pr(I=1|S=0)(1-\pi)-\Pr(I=1|T=1)\,\Pr(T=1)}{\Pr(T=0)}\\&=\frac{(\mu-1)(\pi-np)+1-n}{1-n}\,\Pr(I=1|S=0).\end{aligned}$$

Combining these two results, using $p=\gamma\pi$, we obtain

$$\lambda=\frac{(\mu-1)(1-n\gamma)\pi+1-n}{(1-n)[\gamma\pi(\mu-1)+1]}. \tag{8}$$

This equation shows that λ can be represented in terms of four quantities – the share of the population tested (n), the share of the symptomatic in the population (π), the ratio of the infection rates among the symptomatic and the asymptomatic (μ), and the bias of the tested sample towards the symptomatic (γ) – on some of which health authorities may have updated and detailed information.[7] This knowledge could in turn be exploited to narrow the range of possible values for λ and therefore, through (6), the range of possible values of $\Pr(I=1)$.

We illustrate by showing how the results of the near-complete testing conducted in a small town in Northern Italy can be used to this effect. Specifically, we take advantage of the data from Vo', the municipality of the Veneto region with the first COVID-19 death in Italy (on February 21, 2020), that conducted (twice) an almost complete testing of its population.[8] These data allow us to

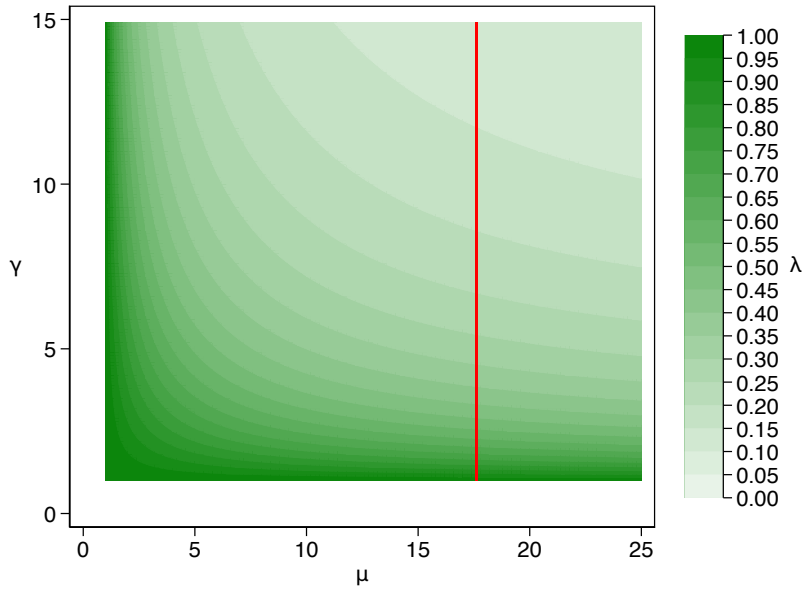
[6] While it is likely true that, even among the symptomatic cases, the doctors administering the test might further discriminate, targeting those with more severe or more numerous symptoms, we believe that our assumption captures the bulk of the bias.

[7] Note that μ and γ both measure relative differences: how much more infected are the symptomatic vs. the asymptomatic, and how much higher is the fraction of symptomatic in the tested sample than in the general population. As such, they are arguably easier to assess, and possibly to calibrate with available evidence.

[8] The data from Vo' contain the results of two separate surveys carried out between February 21 and March 10, 2020 (covering 2,812 and 2,343 subjects respectively, corresponding to 86% and 72% of the eligible population of the municipality), and have been downloaded from the supplementary material of Lavezzo et al. (2020).

Covid Economics 43, 21 August 2020: 19-41

Figure 1: Contour plot of λ as a function of γ and μ for $n = .0004$ and $\pi = .067$.



Notes: The contour plot is based on equation (8) in the main text. The red vertical line corresponds to $\mu = 17.9$, the value observed in Vo' on late February.

compute the population fraction of symptomatic (people with fever or cough or other symptoms like respiratory difficulties or diarrhoea) and asymptomatic cases (people without any of the symptoms related to COVID-19) – these were .067 and .933, respectively; we then calibrate π at .067. For n we take the value .0004 (the testing rate in Italy on June 20, 2020; see Section 3). Given these two values, from equation (8) we can express λ as a function of γ and μ.

Figure 1 shows the contour plot of this function, for $1 \leq \gamma \leq 1/\pi$ (which ensures that $p \leq 1$) and $1 \leq \mu \leq 25$. The color intensity ranges from dark green in the lower left corner, for values of λ close to one (the value we would observe with random sampling) arising from a combination of values of γ and μ close to one, to light green in the upper right corner, for values of λ close to zero arising from a combination of high values of γ and μ.

In the case of Vo', where the fraction of positive cases is .232 among the symptomatic and .0129 among the asymptomatic, we observe $\mu = 17.9$. For this value of μ, λ would vary between 0.11, when the fraction of symptomatic in the sample is 15 times the fraction in the population, and .65, when the sample contains twice as much symptomatic subjects as in the population (i.e., $\gamma = 2$),

a value that we take as the lower bound of the bias intensity.

We do not know whether the value of μ observed in the data from Vo' between late February and early March 2020 is representative of the relative prevalence of COVID-19 among symptomatic and asymptomatic subjects in different regions of Italy or at different times into the evolution of the pandemic. Nevertheless, we think it provides a useful benchmark. More importantly, we think that the health authorities might have a better sense of the size of the bias implicit in the testing protocol, and equation (8) might then be used to narrow down the values for λ.

3 Data

We confine attention to Italy, breaking down the analysis to the regional level to account for the large geographical heterogeneity. We use two data sources: Istat for the Italian population at the onset of COVID-19 and the Italian Department of Civil Protection (Dipartimento della Protezione Civile, or DPC) for daily summaries of the epidemic.

The Istat data are estimates of the resident population as of January 1, 2019, broken down by region (Istituto Nazionale di Statistica, 2019).

The DPC data are the product of the data collection effort coordinated by the Italian integrated COVID-19 surveillance system (Riccardo et al., 2020) and consist of daily time-series at various levels of geographic aggregation. For our purposes, the most important series are: the total number of detected cases ("totale casi"), the number of new detected cases ("nuovi positivi"), equal to the daily change in the total number of detected cases, the total number of swabs ("tamponi"), and the total number of tested cases ("casi testati"). The first three series are available at the national and regional level from February 24, 2020, the last only from April 19.

A number of remarks on the DPC series are in order. First, the series of new detected cases gives the number of subjects who *first* test positive on day t – i.e., the number of subjects for whom $P_t = 1$ – irrespective of whether they became infected on day t or earlier.

Secondly, the number of swabs overstates the number of people actually tested because certain subjects are tested repeatedly. In addition to health personnel and other people employed in critical services, these include people discharged from a hospital who need to test negative at least twice before being sent back home. Besides duplications, the timing between swabs and tests is not fully aligned. In particular, the change in the number of swabs between day $t-1$ and day t actually records the number of tests results obtained on day t. Due to delays in processing and reporting, these include test results from swabs taken before t and excludes swabs taken in t but not yet processed, or for which test results are not yet available.

Thirdly, the series of tested cases records the total number of subjects from whom a swab was taken, thus eliminating the duplications contained in the swabs series. To estimate $\Pr(P = 1|T =$

Covid Economics 43, 21 August 2020: 19-41

Covid Economics 43, 21 August 2020: 19-41

1), it is therefore more appropriate to take the ratio between detected and tested cases, rather than the ratio between detected and swabs. We think that the advantage of excluding the duplications outweighs some drawbacks associated with the series of tested cases, namely the shorter time span and two issues that arise because of the delays in test processing and the retesting rules. The first issue is timing misalignments between swabs and tests, which we just discussed. We address this issue by taking centered moving averages of daily changes of both detected and tested cases. The second issue instead arises because the series of new tested cases excludes subjects who, after being tested negative, are retested at a later date. This, although consistent with our assumption that no subject is tested more than once, leads to overstate the prevalence of the infection among the tested. We have no fix for this issue, but we conjecture that, given the tight constraints on testing capacity, the number of subjects retested after a negative result is relatively small, except possibly for some health personnel and workers in other critical services.

A final warning concerning the data is that the series of tested cases in one of the regions (Lazio) has a clear break on April 24, most likely because of an initial reporting error. We will therefore consider the data starting from April 25.

4 Results

To produce estimates of $\Pr(I = 1)$ based on equation (6), we need to compute the population fraction of new tested cases, $\Pr(T = 1)$, and the population fraction of new positive among the new tested cases, $\Pr(P = 1|T = 1)$. We compute these two fractions using the population estimates from Istat and the DPC data starting from April 25, 2020. For the reason explained in Section 3, we present results for a 7-day centered moving average, though results for 3-day and 5-day moving averages are very similar. A 7-day moving average considerably reduces the sample period but has the advantage of removing the day-of-week effects that are clearly present in the data.

Table 1 presents the observed values of $\Pr(T = 1)$ and $\Pr(P = 1|T = 1)$ as of June 20, 2020. Only a small fraction of the population is newly tested in a given day (.04 percent on average), with substantial variation across regions, from .01 percent in Campania to .08 percent in the Northern regions of Emilia-Romagna, Friuli Venezia Giulia and Trentino-Alto Adige. The fraction of new positives among the newly tested is .78 percent on average, with much larger variation across regions, from less than .005 percent in Basilicata to 2.7 percent in Lombardia, reflecting substantial regional differences in both the intensity of the epidemic and the bias in testing. The correlation between $\Pr(P = 1|T = 1)$ and $\Pr(T = 1)$ is positive but very weak (less than .10).

To produce estimates of $\Pr(I = 1)$, we also need to assign values to β and λ.[9] For the former,

[9] Following Manski and Molinari (2020) we do not provide measures of statistical precision because we are unsure what type of sampling process would be reasonable to assume for our data.

Covid Economics 43, 21 August 2020: 19-41

Table 1: Observed values of $\Pr(T=1)$ and $\Pr(P=1|T=1)$, June 20, 2020.

Region	$\Pr(T=1)$	$\Pr(P=1\|T=1)$
Abruzzo	.0005	.0007
Basilicata	.0005	.0000
Calabria	.0004	.0022
Campania	.0001	.0041
Emilia-Romagna	.0008	.0063
Friuli Venezia Giulia	.0008	.0011
Lazio	.0004	.0038
Liguria	.0004	.0106
Lombardia	.0006	.0273
Marche	.0004	.0027
Molise	.0007	.0035
Piemonte	.0004	.0151
Puglia	.0003	.0014
Sardegna	.0005	.0007
Sicilia	.0003	.0008
Toscana	.0004	.0026
Trentino-Alto Adige	.0008	.0069
Umbria	.0005	.0003
Valle d'Aosta	.0006	.0035
Veneto	.0005	.0018
Italy	.0004	.0078

Notes: 7-day centered moving averages of daily changes of tested cases and new detected cases. Source: DPC and Istat.

as already mentioned, the available information suggests a relatively narrow range between .10 and .30, which we conservatively broaden to the range between .01 and .50. For the latter, we have little a priori information. We showed in Section 2.3 that, using data from a small town in the Veneto region, λ could plausibly be narrowed down to the range $[.10, .65]$. Again conservatively, we report results for the wider range $[.01, .99]$.

Table 2 presents our estimates of $\Pr(I=1)$ for Italy as a whole, as of June 20, 2020. The values in red correspond to $.10 \leq \lambda \leq .65$ and $.10 \leq \beta \leq .30$ and range between .10 percent when $(\lambda, \beta) = (.10, .10)$ and .70 percent when $(\lambda, \beta) = (.65, .30)$. Given the Italian population of 60.4 million, this corresponds to a range between 60 and 423 thousands infected people as of June 20, 2020.

Table 3 shows our estimates, for Italy as a whole and separately by region, for three illustrative pairs of values of (λ, β), namely $(.25, .10)$, $(.50, .20)$, and $(.75, .30)$. Figure 2 instead shows the value of $\Pr(I=1)$ by region at a point in time, namely June 20, 2020, for $(\lambda, \beta) = (.50, .20)$. The estimates reveal a clear North-South gradient, with point prevalence near or above 1 percent in some Northern regions (Piedmont and Lombardy) but well below .05 percent in Umbria and several Southern regions. These large differences reflect the fact that, while $\Pr(T=1)$ and $\Pr(T=0)$ do not vary much, regional variation in $\Pr(P=1|T=1)$ is huge.

Covid Economics 43, 21 August 2020: 19-41

Table 2: Estimates of $\Pr(I = 1)$ for different values of λ and β, June 20, 2020.

	β										
λ	.01	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
.01	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
.05	.000	.000	.000	.000	.000	.001	.001	.001	.001	.001	.001
.10	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.002
.15	.001	.001	.001	.001	.001	.002	.002	.002	.002	.002	.002
.20	.002	.002	.002	.002	.002	.002	.002	.002	.003	.003	.003
.25	.002	.002	.002	.002	.002	.003	.003	.003	.003	.004	.004
.30	.002	.002	.003	.003	.003	.003	.003	.004	.004	.004	.005
.35	.003	.003	.003	.003	.003	.004	.004	.004	.005	.005	.005
.40	.003	.003	.003	.004	.004	.004	.004	.005	.005	.006	.006
.45	.004	.004	.004	.004	.004	.005	.005	.005	.006	.006	.007
.50	.004	.004	.004	.005	.005	.005	.006	.006	.007	.007	.008
.55	.004	.005	.005	.005	.005	.006	.006	.007	.007	.008	.009
.60	.005	.005	.005	.006	.006	.006	.007	.007	.008	.009	.009
.65	.005	.005	.006	.006	.006	.007	.007	.008	.008	.009	.010
.70	.006	.006	.006	.006	.007	.007	.008	.008	.009	.010	.011
.75	.006	.006	.007	.007	.007	.008	.008	.009	.010	.011	.012
.80	.006	.007	.007	.007	.008	.008	.009	.010	.010	.011	.013
.85	.007	.007	.007	.008	.008	.009	.010	.010	.011	.012	.013
.90	.007	.007	.008	.008	.009	.009	.010	.011	.012	.013	.014
.95	.008	.008	.008	.009	.009	.010	.011	.011	.012	.014	.015
.99	.008	.008	.009	.009	.010	.010	.011	.012	.013	.014	.015

Notes: Values obtained from equation (6) in the main text. The values in red correspond to $.10 \leq \lambda \leq .65$ and $.10 \leq \beta \leq .30$.

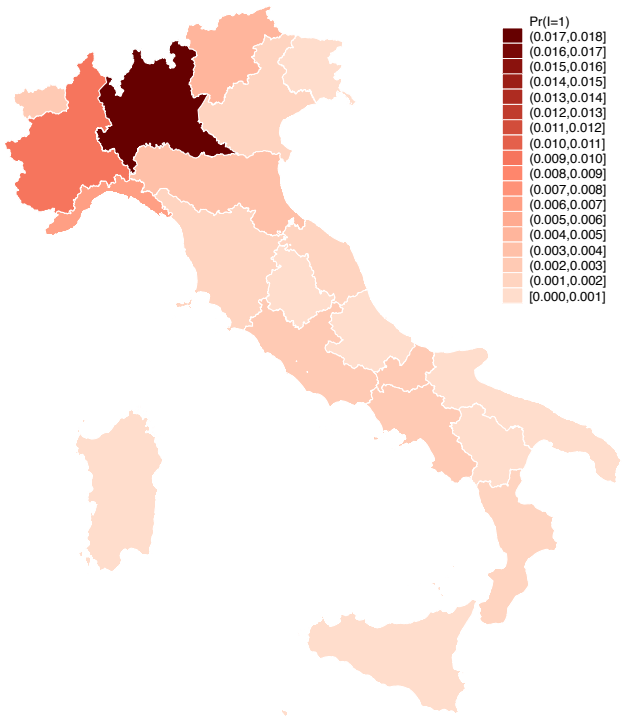
Table 3: Estimates of $\Pr(I = 1)$ by region, for different values of (λ, β), June 20, 2020.

	(λ, β)		
Region	$(.25, .10)$	$(.50, .20)$	$(.75, .30)$
Abruzzo	.000	.000	.001
Basilicata	.000	.000	.000
Calabria	.001	.001	.002
Campania	.001	.003	.004
Emilia-Romagna	.002	.004	.007
Friuli Venezia Giulia	.000	.001	.001
Lazio	.001	.002	.004
Liguria	.003	.007	.011
Lombardia	.008	.017	.029
Marche	.001	.002	.003
Molise	.001	.002	.004
Piemonte	.004	.009	.016
Puglia	.000	.001	.001
Sardegna	.000	.000	.001
Sicilia	.000	.000	.001
Toscana	.001	.002	.003
Trentino-Alto Adige	.002	.004	.007
Umbria	.000	.000	.000
Valle d'Aosta	.001	.002	.004
Veneto	.001	.001	.002
Italy	.002	.005	.008

Notes: Values obtained from equation (6) in the main text.

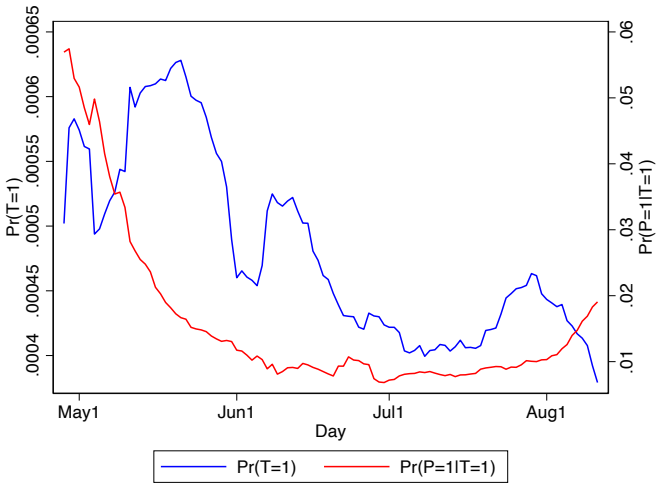
Covid Economics 43, 21 August 2020: 19-41

Figure 2: Estimates of $\Pr(I = 1)$ by region for $\lambda = .50$ and $\beta = .20$, June 20, 2020.



Notes: The figure corresponds to the middle column of Table 3.

Figure 3: Observed values of $\Pr(T=1)$ and $\Pr(P=1|T=1)$ over time.



Notes: 7-day centered moving averages of daily changes of tested cases and new detected cases. Source: DPC and Istat. The scale on the left refers to $\Pr(T=1)$, the one on the right to $\Pr(P=1|T=1)$.

Figure 3 shows the observed values of $\Pr(T=1)$ and $\Pr(P=1|T=1)$ over time. Both probabilities fall, but the decline in $\Pr(P=1|T=1)$ is much stronger. This decline is likely due to, on the one hand, a fall in the intensity of the epidemic and, on the other, a reduction in the bias implicit in the testing criteria. The latter is because the falling number of daily new cases resulting from the containment policies created more room to take swabs from less severe cases and to conduct a more intense contact tracing, again leading to more swabs taken from asymptomatic subjects.

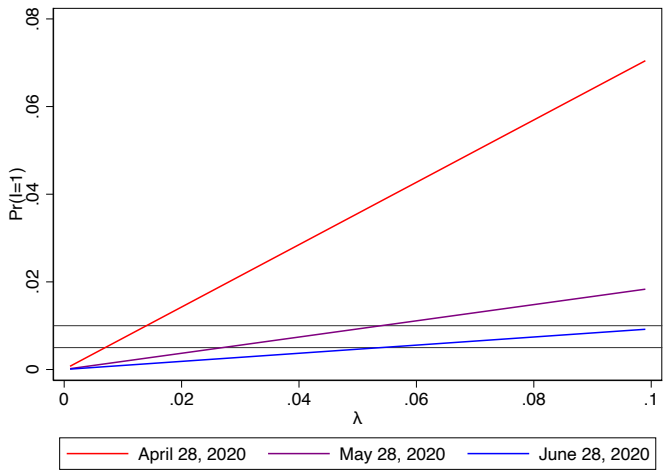
Since we are not able to quantitatively assess the change in the testing bias, we refrain from showing time changes in our estimates. Rather, we show in Figure 4 how $\Pr(I=1)$ increases with λ, for a given value of β, on three different dates. This helps to gauge the extent to which a fall in the testing bias, i.e. an increase in λ, would have offset the observed reduction of prevalence among the tested cases ($\Pr(P=1|T=1)$), reflected in the clockwise rotation of the three lines. The two horizontal lines show, for example, that prevalence would have remained constant at about 1 percent (.5 percent) if λ had increased from about .15 to 1 (from about .05 to .55).

Covid Economics 43, 21 August 2020: 19-41

CEPR PRESS

Covid Economics 43, 21 August 2020: 19-41

Figure 4: Effect of λ on $\Pr(I = 1)$ for $\beta = .20$.



Notes: The different curves show how the estimate of $\Pr(I = 1)$ would vary as λ varies, for a fixed β, at different dates. The horizontal lines are drawn at .01 and .005.

5 Point vs. period prevalence

As mentioned in the Introduction, we have a similar goal as Manski and Molinari (2020): offering an estimate of the prevalence of the infection based on readily available data and minimal and transparent assumption. The main difference is that we aim to provide estimates of *point* prevalence, while they aim to estimate *period* prevalence. An estimate of period prevalence is in the making in some countries, including Italy, that are organizing a random serological testing of the population. As both concepts of prevalence are interesting and play a different role, we explore here the link between them in order to understand what can be learned about period prevalence from estimates of point prevalence and, more generally, from the data currently available.

5.1 Relationship between the two concepts

Assume for simplicity that the infection lasts at most two periods, nobody dies because of it, and a fraction $1 - \delta$ of the *new* infected recovers after one period while the remaining fraction δ after two periods (the analysis would be qualitatively similar had we considered more complex patterns of recovery). Hence, at any time $t \geq 1$, the number of *currently* infected (i.e., still active carriers of the infection) F_t is

$$F_t = F_t^* + \delta F_{t-1}^*, \tag{9}$$

Covid Economics 43, 21 August 2020: 19-41

where F_t^* are the *new* infected at time t, with $F_0^* = 0$ just before the start of the epidemic. The number of people who *just* recovered is

$$V_t = (1-\delta)F_{t-1}^* + \delta F_{t-2}^*, \tag{10}$$

with $V_1 = 0$ and $V_2 = (1-\delta)F_1^*$, and the total number of people who are currently recovered from past infections is

$$U_t = \sum_{i=1}^{t} V_i = (1-\delta)F_{t-1}^* + \sum_{j=2}^{t-1} F_{t-j}^*. \tag{11}$$

Point prevalence at time t in a population of constant size N is

$$\alpha_t = \frac{F_t}{N} = \frac{F_t^* + \delta F_{t-1}^*}{N},$$

while *period* prevalence is

$$\eta_t = \frac{\sum_{i=1}^{t} F_i^*}{N}.$$

Note that point prevalence is obtained from the *currently* infected, who might have been first infected in previous periods, whereas period prevalence is constructed using the *new* infected in each period. It follows from (9) and (11) that $\sum_{i=1}^{t} F_i^* = F_t + U_t$. Therefore we have

$$\eta_t = \alpha_t + \frac{U_t}{N}. \tag{12}$$

Equation (12) clarifies the link between the two concepts in the population: starting from point prevalence, to assess period prevalence we need data on the current and past recovered. Equivalently, while point prevalence can be measured using just the data on the currently infected, to estimate period prevalence one also needs the current and past recovered.

We typically do not observe the population, but only a sample drawn from it; the people in the sample are subject to a virological test that reveals – possibly with noise – whether they are currently infected. To further explore the link between the two concepts of prevalence, assume that the sample is a random one, drawn without replacement. Though counterfactual, this provides a tractable and useful benchmark. We maintain the assumption of zero probability of Type-I error and positive probability β of Type-II error.

Thus suppose that, at each time t, a random sample of size n_t is tested for the presence of the virus. If O_t is the number of people who tested positive in the period-t sample, then

$$\mathbb{E}[O_t] = n_t \alpha_t (1-\beta).$$

Given a reliable estimate of β, we can construct an unbiased estimate of *point* prevalence as $\widehat{\alpha}_t = O_t/[n_t(1-\beta)]$.

Covid Economics 43, 21 August 2020: 19-41

5.2 Data requirement for estimating period prevalence

We could obtain an estimate of *period* prevalence if we also had data recording, in the random samples drawn each period, the number of people recovered from the infection contracted in any of the previous periods. In our simple example, at any date t the fraction of recovered in the population is U_t/N, and therefore the number of recovered in the period-t sample would be (on average) $n_t\, U_t/N$. Dividing by n_t would then give an unbiased estimate of U_t/N and therefore, from (12), an unbiased estimate of period prevalence. Notice however that to identify the recovered, the subjects in the sample should be tested with a *serological* test to detect *past* contacts with the virus.

This is not how the available data are constructed. We know the number of *future* recovered *among the currently infected*, as detected by the *virological* test. But we do not know whether the people who test negative at a point in time *had been infected in the past and are now recovered*, which is precisely the information we would need to compute period prevalence. Hence, even in the ideal case of random sampling, there is a wedge between the data available as by-product of surveillance and treatment protocols and the data that would be needed to construct an accurate estimate of period prevalence. Having a biased sample would only add an extra layer of difficulty, so this conclusion would hold *a fortiori* also in the realistic case of non-random testing.[10]

A wedge between needed and available data is also revealed by the estimate of period prevalence in Manski and Molinari (2020). In keeping with their notation, let C_t be a random variable equal to 1 if, as of day t, a subject has ever been infected and to 0 otherwise, and R_t a random variable equal to 1 if, as of day t, a subject has ever resulted positive to the virological test and to 0 otherwise, i.e. if a subject has never been tested or, if ever tested, resulted negative. With some abuse of notation, for the purposes of this section, also let T_t be a random variable equal to 1 if, as of day t, a subject has ever been tested and to 0 otherwise, i.e. if a subject has never been tested. We can then rewrite equation (10) in their paper, similarly to our equation (6),[11] as

$$\Pr(C_t = 1) = \frac{\Pr(R_t = 1 | T_t = 1)}{1 - \beta_t^*} \left[\Pr(T_t = 1) + \lambda_t^* \left(1 - \Pr(T_t = 1)\right)\right], \tag{13}$$

where $\beta_t^* = \Pr(R_t = 0 | C_t = 1, T_t = 1)$ and $\lambda_t^* = \Pr(C_t = 1 | T_t = 1) / \Pr(C_t = 1 | T_t = 0)$. Though superficially similar, β_t^* is not the same as β, the Type-II error probability of a virological test: while β refers to contemporaneous events (i.e., testing negative while infected), β_t^* refers to events that do not necessarily occur at the same point in time.

In general β_t^* is larger than β as the event associated with β_t^* includes, besides all false negatives up to time t, two types of subjects: (i) those tested on day $d < t$ who became infected on a later

[10] This problem resembles the difficulties faced when trying to estimate the unemployment rate in a population from the stock of currently unemployed people without having data on the flow of new unemployed.

[11] The equivalence between (13) and the estimating equation in Manski and Molinari (2020) is shown in Appendix A.

day $h \leq t$; and (ii) those tested on day $d < t$ who became infected on an earlier day $h < d$ and have fully recovered by day d. The subjects in (i) are not yet infected on the day they are tested, and therefore are correctly recorded as negative on that day, are no longer tested (under the assumption that nobody is tested twice), and yet by day t they have been infected. Similarly, the subjects in (ii) are no longer infected when they are tested, so the virological test correctly finds them as negative, were not tested when infected, and yet by day t they have been infected.

In the simple setting explored in this section, it is easy to verify that, for example,

$$\beta_2^* = \frac{\beta\left[F_1^*(n_1 + \delta n_2) + F_2^* n_2\right] + F_2^* n_1 + F_1^*(1-\delta)n_2}{(F_1^* + F_2^*)(n_1 + n_2)}, \tag{14}$$

which is equal to β only if $\beta = \delta = 1$, and is otherwise larger than β for the presence of terms corresponding to (i) and (ii), namely $F_2^* n_1$ and $F_1^*(1-\delta)n_2$ respectively. Appendix B providess the details.

Our key point is that, even in the ideal case of random virological testing, terms of this kind are inherently not observable. While information on β can be obtained from medical expertise and practice, moving from β to β_t^* involves quantities that are not observable with the data currently produced as a by-product of the surveillance and treatment protocols. Therefore, bounds on β_t^* based on bounds on β are unlikely to be tight due to the unobservable (at least with the data readily available) wedge between the two concepts.[12]

6 Conclusions

We showed in this paper how to use the data currently produced as a byproduct of the fight against the pandemic to obtain estimates of the point prevalence of the infection, an important gauge of its evolution.

Our estimates depend on a number of clearly identified features of the pandemic – the Type-I and Type-II error probabilities of the test used to detect the presence of the virus, the relative prevalence of the infection in subsets of the population, the relative size of these subsets in the sample of the tested subjects – about which it is in principle possible to gather background information. The latter, in turn, can be easily incorporated to narrow down the range of the estimate.

We also showed that another important statistic concerning the pandemic – the period prevalence of the infection – is much harder to pin down with the data currently available, and we pointed out which additional information would be needed to obtain such an estimate.

[12] As explained in Appendix A, Manski and Molinari (2020) impose bounds on $\Pr(C_t = 1|R_t = 0, T_t = 1)$ – which they interpret as (1 minus) the negative predicted value (NPV) of the test – rather than on $\Pr(R_t = 0|C_t = 1, T_t = 1)$. These two probabilities are linked by Bayes Law, and any difficulty in bounding one translates into a difficulty in bounding the other. In fact, the logical possibilities that create a difference between β_t^* and β are also responsible for a difference between $\Pr(C_t = 1|R_t = 0, T_t = 1)$ and (1 minus) the NPV of the test.

Covid Economics 43, 21 August 2020: 19-41

Finally, we applied our method to data from Italy, a country that experienced an early and massive outbreak of the epidemic and is now lifting the social distancing measures introduced to contain the contagion. Our estimates show that in some regions of that country the prevalence of the currently infected might be still large enough to suggest caution in removing all restrictions on mobility, unless an effective system to quickly identify the newly infected, trace their contacts and implement quarantine measures is in place.

We hope that the health authorities, at the national or regional level, by incorporating in our framework their more detailed and disaggregated information about the nature of the pandemic, can obtain a quickly updated gauge of the evolution of the infection, which will help them to contain it.

Covid Economics 43, 21 August 2020: 19-41

Covid Economics 43, 21 August 2020: 19-41

References

Flaxman, S., S. Mishra, A. Gandy, H. J. T. Unwin, et al. (2020). Estimating the number of infections and the impact of non- pharmaceutical interventions on COVID-19 in 11 European countries. *Nature*. https://doi.org/10.1038/s41586-020-2405-7.

Istituto Nazionale di Statistica (2019). Resident population as of January 1, 2019. Available at: https://www.istat.it/it/popolazione-e-famiglie/.

Lavezzo, E., E. Franchin, C. Ciavarella, G. Cuomo-Dannenburg, et al. (2020). Suppression of COVID-19 outbreak in the municipality of Vo', Italy. *Nature*. https://doi.org/10.1038/s41586-020-2488-1.

Li, R., S. Pei, B. Chen, Y. Song, et al. (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science 386*(6490), 489–493.

Manski, C. F. and F. Molinari (2020). Estimating the COVID-19 infection rate: Anatomy of an inference problem. *Journal of Econometrics*. Forthcoming.

Riccardo, F., M. Ajelli, X. D. Andrianou, A. Bella, et al. (2020). Epidemiological characteristics of COVID-19 cases in Italy and estimates of the reproductive numbers one month into the epidemic. Istituto Superiore di Sanità, Rome. Available at: https://www.medrxiv.org/content/10.1101/2020.04.08.20056861v1.

Stock, J. H. (2020). Data gaps and the policy response to the novel coronavirus. *Covid Economics: Vetted and Real–Time Papers 3*, 1–11.

Vollmer, M. A. C., S. Mishra, H. J. T. Unwin, A. Gandy, et al. (2020). Using mobility to estimate the transmission intensity of COVID-19 in Italy: A subnational analysis with future scenarios. Imperial College COVID-19 Response Team Report 20. Available at: https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2020-05-04-COVID19-Report-20.pdf.

Watson, J., P. F. Whiting, and J. E. Brush (2020). Interpreting a covid-19 test result. *British Medical Journal* (May 12, 2020). Available at: https://www.bmj.com/content/369/bmj.m1808.

Covid Economics 43, 21 August 2020: 19-41

A Equivalence between two formulations

With the notation introduced in Section 5, equation (10) in Manski and Molinari (2020) can be written

$$\begin{aligned} \Pr(R_t = 1) + B_* \Pr(R_t = 0|T_t = 1) \Pr(T_t = 1) \leq \\ \leq \Pr(C_t = 1) \leq \\ \leq \Pr(R_t = 1|T_t = 1) + B^* \Pr(R_t = 0|T_t = 1), \end{aligned} \tag{15}$$

where B_* and B^* are, respectively, the lower and upper bound on $\Pr(C_t = 1|R_t = 0, T_t = 1)$; the first expression in the chain of inequalities, which provides a lower bound on $\Pr(C_t = 1)$, obtains when $\Pr(C_t = 1|T_t = 0) = 0$, while the last expression, which provides an upper bound on $\Pr(C_t = 1)$, obtains when $\Pr(C_t = 1|T_t = 0) = \Pr(C_t = 1|T_t = 1)$.

We can rewrite (15) as

$$\begin{aligned} \Pr(C_t = 1) = [\Pr(T_t = 1) + \lambda_t^* (1 - \Pr(T_t = 1))] \times \\ \times [\Pr(R_t = 1|T_t = 1) + \Pr(C_t = 1|R_t = 0, T_t = 1) \Pr(R_t = 0|T_t = 1)] . \end{aligned} \tag{16}$$

Thus $\Pr(C_t = 1)$ is equal to the top term in (15) when $\Pr(C_t = 1|R_t = 0, T_t = 1) = B_*$ and $\lambda_t^* = 0$ – i.e., $\Pr(C_t = 1|T_t = 0) = 0$ – and to the bottom term when $\Pr(C_t = 1|R_t = 0, T_t = 1) = B^*$ and $\lambda_t^* = 1$ – i.e., $\Pr(C_t = 1|T_t = 0) = \Pr(C_t = 1|T_t = 1)$. Now,

$$\begin{aligned} \Pr(C_t = 1|R_t = 0, T_t = 1) \Pr(R_t = 0|T_t = 1) &= \frac{\Pr(C_t = 1, R_t = 0, T_t = 1)}{\Pr(T_t = 1)} \\ &= \frac{\Pr(C_t = 1, R_t = 0, T_t = 1)}{\Pr(C_t = 1, T_t = 1)} \frac{\Pr(C_t = 1, T_t = 1)}{\Pr(T_t = 1)} \\ &= \Pr(C_t = 1|T_t = 1) \beta_t^*, \end{aligned} \tag{17}$$

with β_t^* defined in Section 5. From the definition of λ_t^* and the Law of Total Probability, we get

$$\begin{aligned} \Pr(C_t = 1) &= \Pr(C_t = 1|T_t = 1) \Pr(T_t = 1) + \lambda_t^* \Pr(C_t = 1|T_t = 1) \Pr(T_t = 0) \\ &= \Pr(C_t = 1|T_t = 1) [\Pr(T_t = 1) + \lambda_t^* \Pr(T_t = 0)] . \end{aligned} \tag{18}$$

Plugging (17) and (18) into (16), simplifying, and solving out for $\Pr(C_t = 1|T_t = 1)$ yields

$$\Pr(C_t = 1|T_t = 1) = \frac{\Pr(R_t = 1|T_t = 1)}{1 - \beta_t^*}. \tag{19}$$

Substituting (19) into (18) then gives

$$\Pr(C_t = 1) = \frac{\Pr(R_t = 1|T_t = 1)}{1 - \beta_t^*} [\Pr(T_t = 1) + \lambda_t^* \Pr(T_t = 0)] ,$$

which is equation (13) in the main text. Note also that, under the assumption of zero Type-I error probability,

$$\Pr(C_t = 1 | R_t = 0, T_t = 1) = \frac{\beta_t^*}{\beta_t^* + \lambda_t^*}. \tag{20}$$

Hence, the bounds on $\Pr(C_t = 1 | R_t = 0, T_t = 1)$ readily translate into bounds on β_t^*.

B Decomposition under random sampling

Assume, as in Section 5, that (i) each person can be infected at most once, the infection lasts at most two periods, nobody dies because of it, and a fraction $1 - \delta$ of the *new* infected recovers after one period while the remaining δ after two periods; (ii) the sample of people tested in each period is random, drawn without replacement; and (iii) the virological test has zero probability of Type-I error and (constant) probability $\beta > 0$ of Type-II error. Using the notation introduced in that section, we can break down the population in each period t into the following groups:[13]

- sampled in period t, currently infected, testing negative: $n_t(F_t^* + \delta F_{t-1}^*)\beta/N$;
- sampled in period t, currently infected, testing positive: $n_t(F_t^* + \delta F_{t-1}^*)(1-\beta)/N$;
- sampled in period t, newly infected in period $t-1$, currently recovered, testing negative: $n_t(1-\delta)F_{t-1}^*/N$;
- sampled in period t, newly infected in period $r < t-1$, currently recovered, testing negative: $n_t F_r^*/N$;
- sampled in period t, not yet infected, testing negative: $n_t(N - \sum_{i=1}^t F_i^*)/N$;
- sampled in period $s < t$, currently infected: $n_s(F_t^* + \delta F_{t-1}^*)/N$;
- sampled in period $s < t$, newly infected in period $t-1$, currently recovered: $n_s(1-\delta)F_{t-1}^*/N$;
- sampled in period $s < t$, newly infected in period $r < t-1$, currently recovered: $n_s F_r^*/N$;
- sampled in period $s < t$, not yet infected: $n_s(N - \sum_{i=1}^t F_i^*)/N$;
- not previously sampled and currently infected: $(N - \sum_{i=1}^t n_i)(F_t^* + \delta F_{t-1}^*)/N$;
- not previously sampled, newly infected in period $t-1$, currently recovered: $(N - \sum_{i=1}^t n_i)(1-\delta)F_{t-1}^*/N$;
- not previously sampled, newly infected in period $r < t-1$, currently recovered: $(N - \sum_{i=1}^t n_i)F_r^*/N$;

[13] Some of the items will only make sense for $t > 1$ or $t > 2$.

Covid Economics 43, 21 August 2020: 19-41

Covid Economics 43, 21 August 2020: 19-41

- not yet sampled, not yet infected: $(N - \sum_{i=1}^{t} n_i)(N - \sum_{i=1}^{t} F_i^*)/N$.

Specializing for simplicity to the case $t = 2$, we can use this decomposition to compute $\beta_2^* = \Pr(R_2 = 0, C_2 = 1, T_2 = 1)/\Pr(C_2 = 1, T_2 = 1)$. On the numerator we must include all the people who never tested positive, were sampled in one of the two periods and were first infected in one of the two periods:

- sampled in period 1, newly infected in period 1, wrongly classified as negative in period 1, negative in period 2 because not tested again: $n_1 F_1^* \beta/N$;
- sampled in period 2, newly infected in period 2, wrongly classified as negative in period 1, negative in period 1 because in that period they were not tested: $n_2 F_2^* \beta/N$;
- sampled in period 2, newly infected in period 1, still infected in period 2, wrongly classified as negative in period 2, negative in period 1 because in that period they were not tested: $n_2 F_1^* \delta\beta/N$;
- sampled in period 2, newly infected in period 1, recovered by period 2, correctly classified as negative in period 2, negative in period 1 because in that period they were not tested: $n_2 F_1^* (1 - \delta)/N$;
- sampled in period 1, newly infected in period 2, correctly classified as negative when tested (in period 1), negative in period 2, because not tested again: $n_1 F_2^*/N$.

As to the denominator, we must include all the people who where first infected, in either period 1 or period 2, and were first tested, in either period 1 or period 2. These are, clearly, $(n_1 + n_2)(F_1^* + F_2^*)/N$. Putting all this together we obtain equation (14) in the main text.

The fragmented US: The impact of scattered lockdown policies on country-wide infections[1]

Jacek Rothert,[2] Ryan Brady[3] and Michael Insler[4]

Date submitted: 19 August 2020; Date accepted: 19 August 2020

Fragmented by policies, united by outcomes: This is the picture of the United States that emerges from our analysis of the spatial diffusion of Covid-19 and the scattered lock-down policies introduced by individual states to contain it. We first use spatial econometric techniques to document direct and indirect spillovers of new infections across county and state lines, as well as the impact of individual states' lock-down policies on infections in neighboring states. We find consistent statistical evidence that new cases diffuse across county lines, holding county level factors constant, and that the diffusion across counties was affected by the closure policies of adjacent states. Spatial impulse response functions reveal that the diffusion across counties is persistent for up to ten days after an increase in adjacent counties. We then develop a spatial version of the epidemiological SIR model where new infections arise from interactions between infected people in one state and susceptible people in the same or in neighboring states. We incorporate lock-down policies into our model and calibrate the model to match both the cumulative and the new infections across the 48 contiguous U.S. states and DC. Our results suggest that, had the states with the less restrictive social distancing measures tightened them by one level, the cumulative infections in other states would be about 5% smaller. In our spatial SIR model, the spatial containment policies such as border closures have a bigger impact on flattening the infection curve in the short-run than on the cumulative infections in the long-run.

Covid Economics 43, 21 August 2020: 42-94

1 The views expressed here are those of the authors and do not represent the views of the United States Naval Academy, the Department of Defense, or the Federal Government.

2 Associate Professor of Economics, United States Naval Academy.
3 Professor of Economics, United States Naval Academy.
4 Associate Professor of Economics, United States Naval Academy.

Covid Economics 43, 21 August 2020: 42-94

1 Introduction

In this paper we assess the spatial diffusion of Covid-19 in the United States and the effect that state-level lock-down policies have on that diffusion. Our analysis is motivated by the idea that, if there were substantial spillovers of new infections between states, then the uncoordinated responses at the state level may have exacerbated the outbreak of the disease. But, how large those inter-region spillovers were, or are, is uncertain—as is the extent to which the relatively lax policies of one state contributed to new infections in surrounding states. Indeed, as is now well-documented, the virus spread quickly throughout the United States, with notable variations in state-level policy to follow. By March 6, a majority of U.S. states had at least one confirmed case of the virus, and by March 17 the last state (West Virginia) reported its first case. While the Center for Disease Control (CDC) and and other federal entities issued guidance on appropriate measures to mitigate the spread of the virus, the final decisions regarding the timing and the extent of restrictions were made by individual states, and sometimes even counties. The first state-wide "shelter-in-place" order was issued in California on March 19, but ultimately only 24 additional states followed suit over the next two weeks. The compliance with social distancing measures also varied greatly across regions (Painter and Qiu, 2020; Simonov et al., 2020). The goal of this paper is to assess the impact of such a scattered policy response on the country-wide spread of the virus.

Our analysis follows a two-pronged approach. First, we employ both spatial econometric and time-series methods to measure the extent of the spatial correlation between regions in the U.S., and to understand the dynamics, or the persistence, of that spatial correlation over time. To that end, we estimate direct and indirect spatial spillovers from a variety of workhorse spatial models. For each model the dependent variable is the simple growth rate of county-level cases, and our primary covariate is a measure of the number of restrictions put in place in the state in which the county resides. We find consistent statistical evidence that not only did new cases diffuse across county lines, holding county level factors constant, but that the spatial diffusion across counties was affected by the closure policies of adjacent states. Next, to measure the temporal dynamics of

Covid Economics 43, 21 August 2020: 42-94

the spatial spillovers we generate "spatial impulse response functions" (IRFs), showing how long a particular county was affected by its neighbors' rate of new cases. We find that the spatial diffusion of new cases is statistically significant and persistent over time, for at least ten days over our forecast horizon. Considered together, our results from the spatial models and IRFs provide an informative picture on the nature of the spatial correlation of the Covid-19 phenomenon. Our empirical results suggest, too, that more stringent state-level restrictions are consistent with a decline in the growth rate of new cases at the county level.

Second, we develop a spatial version of the standard epidemiological SIR (Susceptible-Infected-Recovered) model based on Kermack and McKendrick (1927), which has been popularized in the economics literature by Atkeson (2020b). In our model individuals can be infected by people from their own states and from other states. Those inter-state contacts endogenously create a spatial diffusion of the infections, with the speed of such diffusion depending on the model parameters that measure the relative frequency of connections across state lines, potentially altered by social distancing measures. We calibrate the model parameters by minimizing the distance between the data and the model generated series. We then use the model to simulate the impact of lock-down policies implemented in the states with the most restrictive and most lax policies.[1] Our main results in that section are twofold. First, if the individual states had the ability to restrict the travel across their borders, infections would be smaller.[2] Specifically, cutting the value of the calibrated inter-state spillover parameter by 25% results in the reduction of country-wide infections by almost 40% in the first 3 months, and by almost 7% in the long-run. Second, if the states with the more lenient lock-down policies tightened them by one level, the cumulative cases in the remaining states would be reduced by 2% in the first three months, and by more than 5% over the 21-month period.

Our analysis contributes to a large and quickly growing literature on the economics of Covid-19. First, we expand the empirical literature that focuses on the spatial aspects of the outbreak. A few

[1] We discuss possible limitations arising due to the Lucas' critique in Section 4.3.

[2] Of course, we do not suggest that giving the states such ability would be desirable. We are merely evaluating its potential impact on the spread of infections across the whole country.

Covid Economics 43, 21 August 2020: 42-94

studies analyzed drivers of spatial heterogeneity in the scope or severity of the Covid-19 pandemic: Desmet and Wacziarg (2020) and Gerritse (2020) looked at US counties, Verwimp (2020) at Belgian municipalities, and Ginsburgh et al. (2020) at French regions. Very few papers seemed to focus on understanding the geographic spread of the infections. Kuchler et al. (2020) analyze the correlation between the growth in new cases and the degree of social connectedness with the Covid hotspots, using an aggregated data from Facebook. Cuñat and Zymek (2020) analyze geographical spread of the virus in the U.K. by incorporating individual's location and mobility decisions with the SIR model. The most closely related study is Eckardt et al. (2020), where the authors analyze how the border closures slowed down the spread of the virus. Our paper offers the first empirical attempt to estimate the extent of spatial diffusion of Covid-19 in the United States. Our main objective is to quantify the extent of inter-state spillovers and the impact of one state's containment measures on outcomes in surrounding states, with close attention paid both to containment measures and possible non-compliance with them.

Second, our results are important for the discussion of policy coordination. It is quite well known that in the presence of inter-state spillovers, an uncoordinated policy response may lead to sub-optimal outcomes. In the context of Covid-19, the discussion in this area has been mostly theoretical. Beck and Wagner (2020) provide a model of optimal international coordination, focusing on the timing of such coordination. Rothert (2020) uses a heterogeneous agents framework with rich and poor households from Michaud and Rothert (2018) to shows that in the presence of an uncoordinated response, a federal income-based redistribution can flatten the curve if the state governments cannot easily increase welfare spending. Our paper offers the first empirical insight into the actual magnitude of such inter-regional spillovers. Our findings suggest that those spillovers are substantial and therefore emphasize the importance of a coordinated policy response.

Third, the variation in state-level restrictions plays a key role in our analysis. This associates us with a number of papers that focus on the effectiveness of various social distancing measures or on the compliance with the official rules.[3] Painter and Qiu (2020) and Simonov et al. (2020)

[3]See e.g., Weber (2020), Pragyan Deb and Tawk (2020), Jinjarak et al. (2020).

Covid Economics 43, 21 August 2020: 42-94

show that compliance with social distancing rules in the U.S. is correlated with party affiliation, and with exposure to certain opinion-forming programs on Fox News. Briscese et al. (2020) show that the compliance can vary over time and that people can become "tired of" restrictions. In our analysis we use a measure of state-imposed restrictions and we also allow for imperfect compliance with them. Our results indicate that stricter social distancing measures introduced by individual states, and better compliance with them, limit the spread of the disease not only within those states, but also in the neighboring states. Conversely, the lack of such restrictions makes it harder for the state's neighbors to contain the virus.

Finally, following Atkeson (2020b), a number of papers have contributed to modelling the spread of the pandemic. The SIR model has become the standard in that literature with different papers suggesting different modifications, depending on the paper's focus.[4] The closest papers to ours are Bisin and Moro (2020) and Acemoglu et al. (2020). The former builds a theoretical framework that formalizes aspects such as local travel and changes in individuals' behavior, but their focus is on the local diffusion around the hot-spot of the outbreak. The latter develops a multi-group version of the SIR model where infection risks differ across population groups (e.g., nursing homes, schools, etc.) and allows for the transmission of infections between population subgroups. Our main contribution is to develop a spatial version of the benchmark SIR model that allows us to quantify the spillover effects of local infections as well as local lock-down policies on the spread of the virus in other parts of the country.

2 Covid-19 outbreaks and policy responses across the U.S.

We start by documenting some stylized facts about the time and spatial dimensions of the spread of Covid-19 and containment measures in the United States.

[4] A non-exhaustive list of examples includes Atkeson et al. (2020), Holden and Thornton (2020), McAdams (2020), Favero (2020), Berger et al. (2020), Hornstein (2020), or Ellison (2020).

Covid Economics 43, 21 August 2020: 42-94

2.1 Data

We utilize three data sources in this paper:

1. Daily county-level data on confirmed cases and deaths are from the Covid-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (Dong et al. (2020)).

2. Daily state-level data on business closures and mandated social distancing measures are from the Institute for Health Metrics and Evaluation (IHME). Specifically, we utilize IHME-compiled information on five such metrics including the date on which a state proceeded as follows: forbade mass gatherings, introduced an initial round of business closures, closed schools, closed all non-essential businesses, and adopted a stay-at-home order. For each day in our time series, we sum the number of currently-imposed restrictions within a state to generate a measure of government-imposed behavior restrictions at the state level. This metric is thus a count variable taking values 0 through 5, which we call r-score.

3. County-level data on socioeconomic, demographic, and geographic characteristics from the Bureau of Economic Analysis (BEA). In the analysis below we incorporate county-level information on population density for the year 2018 (BEA and U.S. Census Bureau), the share of the county-level population over age 60 for the year 2018 (U.S. Census Bureau), and partisan voting share from the 2016 presidential election.[5]

2.2 Covid-19 cases across time and space

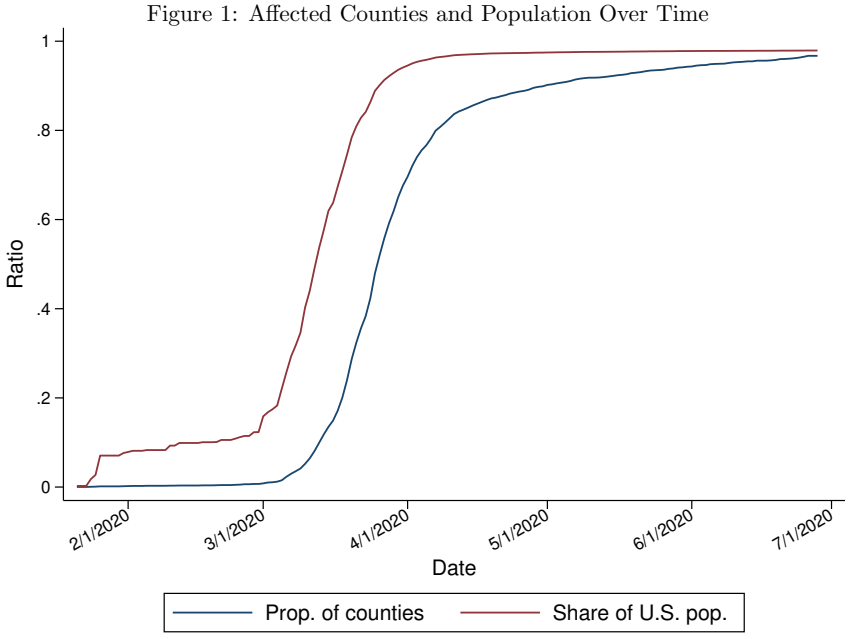
In this section we conduct some basic visual analysis to demonstrate the extent of the Covid-19 epidemic in the United States. We do so to provide a descriptive look at the dynamics of the inter-state and inter-county spillover of the virus.

Figure 1 displays the spread of Covid-19 across the whole country over time. The blue line

[5] Via Luis Sevillano on GitHub, but originally published in the New York Times and available here.

Covid Economics 43, 21 August 2020: 42-94

Figure 1: Affected Counties and Population Over Time

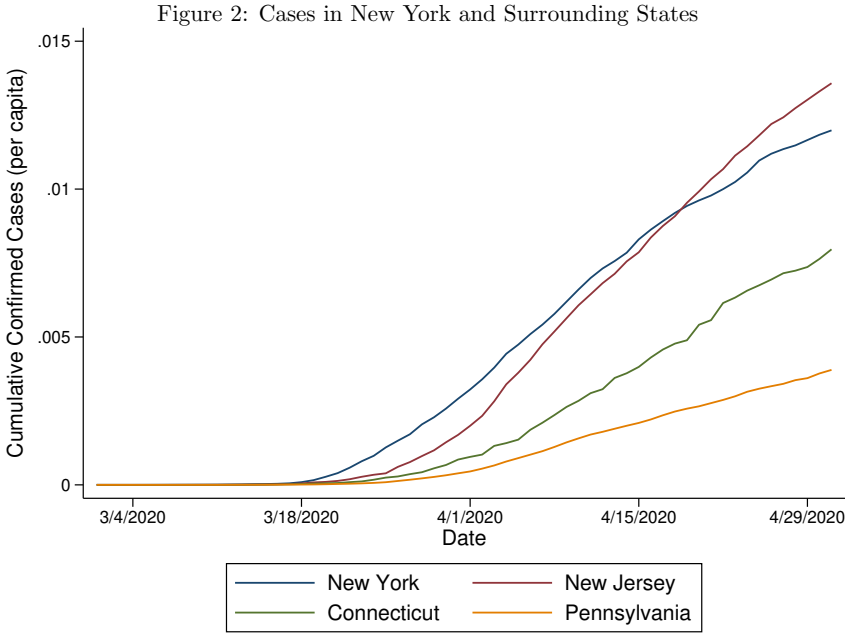


plots the proportion of counties with confirmed cases over time, demonstrating the breadth of the epidemic. The red line reveals its depth, displaying how the share of the population living in a county with at least one confirmed case quickly rose from close to zero to close to one during the month of March. It is clear that from March through June, Covid-19 transitioned from a fairly sparse outbreak to a widespread epidemic among counties along both the extensive and intensive margins.

New York was the first state in the U.S. to experience a very significant outbreak, with its epicenter in New York City (NYC). A simple inspection of the dynamics of case numbers in and around NYC indicates the importance of interstate spillovers (within the CT-NJ-NY-PA areas). Figure 2 shows a rapid expansion of per-capita case numbers in New York during the second half of March, followed by all other states surrounding the NYC metropolitan area in late March and

Covid Economics 43, 21 August 2020: 42-94

Figure 2: Cases in New York and Surrounding States



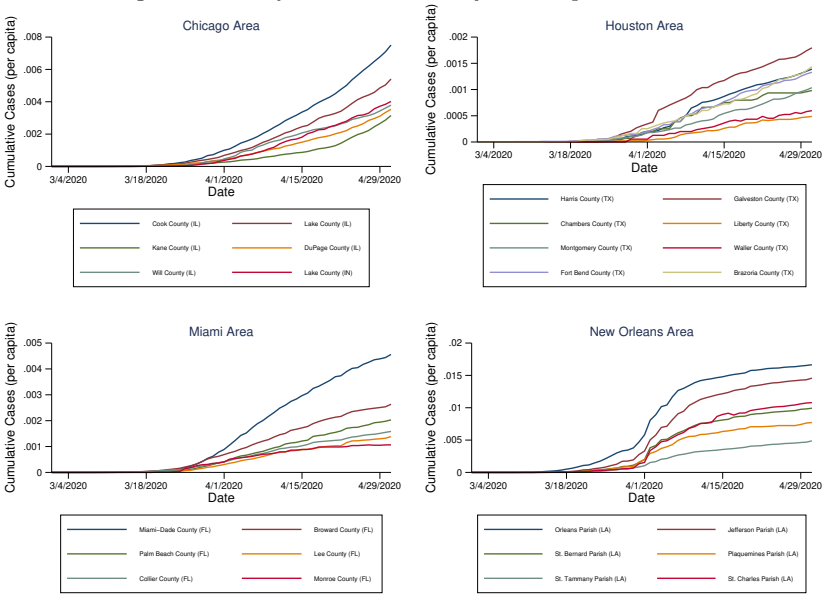
the first half of April. In the case of New Jersey, per-capita case counts caught up to—and began to outpace—those of New York in mid-April.

Closer inspection at the county level in other metropolitan areas reveals similar patterns. Figure 3 plots confirmed county-level cases per-capita over time for four particular metropolitan areas: Chicago, IL (top left); Houston, TX (top right); Miami, FL (bottom left); New Orleans, LA (bottom right). In each case, the county containing the urban center appears to trigger the area outbreak (respectively: Cook County, IL; Harris County, TX;[6] Miami-Dade County, FL; Orleans Parish, LA). In the case of Chicago, which lies on the Illinois-Indiana border, the spillover appears to extend to Lake County, IN. Together, these figures provide preliminary evidence of transmission dynamics in which Covid-19 cases emanate from major urban centers into the surrounding areas,

[6]Here the outbreak appears to stem from both Harris County (Houston) and Galveston County, which is also a fairly densely populated area.

Covid Economics 43, 21 August 2020: 42-94

Figure 3: County-level Cases in Major Metropolitan Areas



and across state lines.

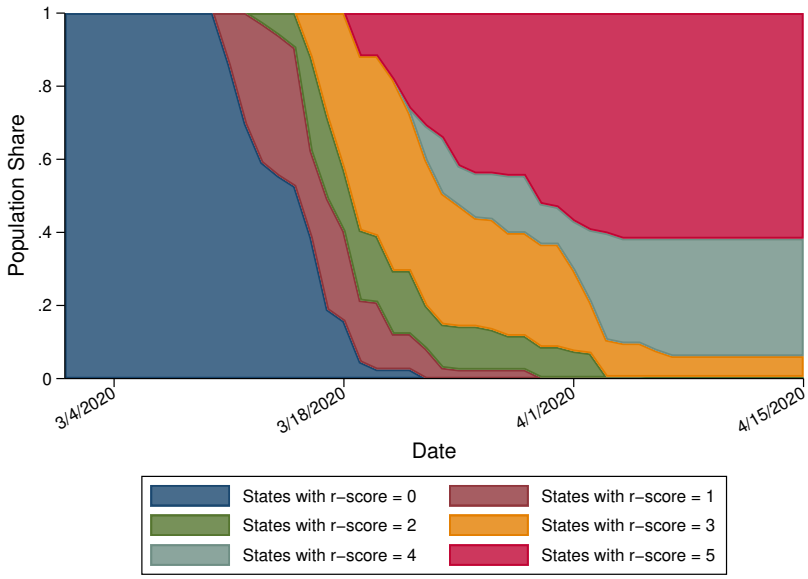
2.3 Containment measures across time and space

There is also clear evidence that during the initial shutdown of March and April, government-led containment measures varied geographically and over time. Figure 4 displays the transition from no official containment measures (early March) to universal adoption (of at least some measures) by all states by early April. The most action occurred between mid-March and early April. 60 percent of the U.S. population lived in a state with no containment measures on March 15; however, by April 1 nearly 60 percent of the population lived in a state that had adopted all five factors described by our r-score variable.

In the sections below, our spatial-econometric analysis and our spatial model and calibration

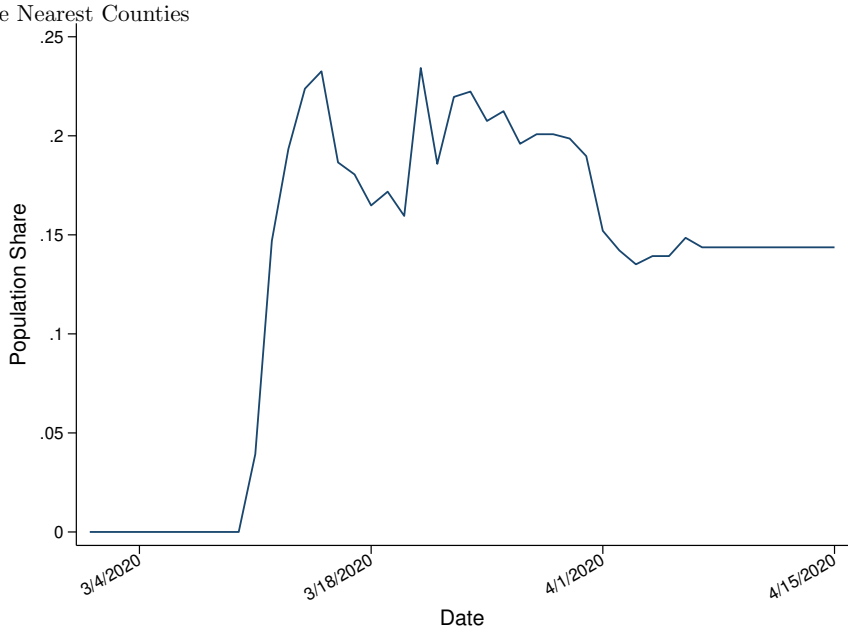
Covid Economics 43, 21 August 2020: 42-94

Figure 4: Containment Measures Over Time



Covid Economics 43, 21 August 2020: 42-94

Figure 5: Share of U.S. Population Living in a County with Different Containment Measures than its Five Nearest Counties



exercise seek to understand further the geographical spillovers across county and state lines. For such effects to be meaningful and identifiable, there must be substantial variation in government-led Covid-19 responses near state borders. Figure 5 displays the share of the U.S. population living in a county with a different level of containment measures (i.e., r-score value) than at least one of its five nearest counties, over time. During the containment action period of mid-March through early April, this share quickly rose to about 23 percent of the population, fluctuated between 15 and 23 percent for the next two weeks, and then settled at 15 percent for the remainder of April.

Covid Economics 43, 21 August 2020: 42-94

3 Estimating the Spatial Diffusion of Cases

In this section and the next we estimate the spatial characteristics of cases. First, we estimate the spillovers of cases from county to county in the United States using oft-used spatial econometric models. Second we then use a VAR-framework to estimate "spatial impulse response functions," which capture the time-series persistence of the spatial spillovers.

3.1 Estimation of Spatial Correlation

For the first exercise, we estimate the following standard spatial models: the Spatial Durbin Model (SDM), the SDM model with spatially correlated errors (SDM Error), the SAR Model, and the SLX model. We start with SDM which, in a panel setting, can be written as:

$$Y_t = \rho W Y_t + \beta X_t + \theta W X_t + \delta + \epsilon_t,$$

where $Y_t = Y_{1,t}, ..., Y_{N,t}$ is a $NT \times 1$ vector of the dependent variable; WY_t is the spatial lag term; X_t is an $NT \times r$ matrix of r exogenous variables, $\delta = \delta_1, ..., \delta_N$ is the vector of region-fixed effects; and WX_t is the spatial lag term for the exogenous variables, which here we refer to as the "exogenous spatial interaction" term (to distinguish this from the spatial lag variable).[7] The spatial term WY_t captures the direct effect of the spatial correlation in determining the dependent variable (capturing what otherwise might be an omitted variable). The SDM error model amends the typical SDM to include a spatially correlated error. The SAR model imposes a restriction that $\theta = 0$, and the SLX model imposes a restriction that $\rho = 0$.

For the models above, Y_t is the county level cases, and X_t is the r-score. We include a county-level fixed effect to control for county-level features such as population density, relative industry, employment shares, and so on (given the data sources, such variables are fixed over the time period). We report the baseline spatial results for each model, along with the direct and indirect spatial

[7] See Halleck Vega and Elhorst (2015) for a detailed discussion on the SLX model and the other spatial models employed in this paper.

CEPR PRESS

Covid Economics 43, 21 August 2020: 42-94

effects. We also consider additional specifications, including with the r-score lagged 14 days, and with the lag of the dependent variable included as a regressor.

Table 1: Baseline results for Spatial Models of New Cases (measured as a growth rate)

	SAR	SAR	SLX	SLX	SDM	SDM	SDM Error	SDM Error
r-score	-6.223***		4.280***		3.851***		4.938***	
	(0.278)		(1.138)		(1.130)		(1.096)	
r-score 14 day lag		-6.030***		2.691**		2.292**		3.192***
		(0.273)		(1.119)		(1.112)		(1.080)
W × cases	0.156***	0.152***			0.155***	0.151***	0.552***	0.544***
	(0.00295)	(0.00298)			(0.00295)	(0.00298)	(0.00497)	(0.00515)
W × r-score			-12.75***		-10.99***		-8.868***	
			(1.203)		(1.195)		(1.141)	
W × (r-score 14 day lag)				-10.71***		-9.077***		-6.938***
				(1.183)		(1.176)		(1.125)
W × error							-0.521***	-0.512***
							(0.00783)	(0.00803)
N	276408	276408	276408	276408	276408	276408	276408	276408

Notes: Spatial models estimated from April 1 through June 28, controlling for county-level fixed effects. Dynamic refers to the lag of the dependent variable included in the model. 14 day lag of r-score is the value from 14 days prior. The Wald test for spatial correlation is statistically significant in all models. Standard errors are in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$.

Tables 1 and 2 report the results for the four models. Table 2 displays results with a lag of the dependent variable included in each model; we refer to this as the dynamic version.[8] We report

[8]There are various versions of spatial models with temporal dynamics. Elhorst (2012) refers to a SAR model augmented with temporal lag of the dependent variable and a temporal lag of the spatial lag as the "time-space

Covid Economics 43, 21 August 2020: 42-94

Table 2: Dynamic Version of Spatial Models

	SAR	SAR	SLX	SLX	SDM	SDM	SDM Error	SDM Error
cases (t-1)	0.0493***	0.0488***	0.0544***	0.0536***	0.0493***	0.0488***	0.0500***	0.0495***
	(0.00190)	(0.00190)	(0.00191)	(0.00191)	(0.00190)	(0.00190)	(0.00171)	(0.00171)
r-score	-5.900***		3.183**		2.698**		4.007***	
	(0.364)		(1.344)		(1.336)		(1.297)	
r-score 14 day lag		-5.940***		2.629**		2.245**		3.078***
		(0.273)		(1.118)		(1.111)		(1.078)
W × cases	0.150***	0.148***			0.149***	0.147***	0.545***	0.540***
	(0.00298)	(0.00299)			(0.00298)	(0.00299)	(0.00496)	(0.00505)
W × r-score			-11.30***		-9.571***		-7.823***	
			(1.440)		(1.431)		(1.364)	
W × (r-score 14 day lag)				-10.50***		-8.929***		-6.714***
				(1.181)		(1.174)		(1.123)
W × error							-0.521***	-0.515***
							(0.00777)	(0.00785)
N	276408	276408	276408	276408	276408	276408	276408	276408

Notes: Spatial models estimated from April 1 through June 28, controlling for county-level fixed effects. Dynamic refers to the lag of the dependent variable included in the model. 14 day lag of r-score is the value from 14 days prior. The Wald test for spatial correlation is statistically significant in all models. Standard errors are in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$.

Covid Economics 43, 21 August 2020: 42-94

baseline results shown in Tables 1 and 2 for purposes of comparison with "traditional" approaches to spatial estimation, of which these models represent. However, as emphasized in LeSage and Pace (2009) and elaborated upon in the spatial literature since, one should not use these "baseline" results to interpret the spatial effects. As explained in detail by Golgher and Voss (2016), one cannot interpret the coefficients as typical partial derivatives (see Elhorst (2010) or Elhorst (2014), for additional explanation). Instead, it is more appropriate to focus on the direct and indirect effects of these models.

Table 3 displays those direct and indirect spatial effects from estimating each model. The direct effect is the effect of a change in explanatory variable X, in county i, on the number of new Covid-19 cases in county i. The indirect effect is, instead, the effect of the change in explanatory variable X, in county j, on the average case levels in surrounding states. The indirect effect is the spatial spillover. We eschew discussing the details of how to derive these effects; instead, we refer the reader to Golgher and Voss (2016), LeSage and Pace (2009), or Elhorst (2010) for technical explanations.[9]

In all cases, we employ a contiguity matrix (row-normalized) since this version of the spatial-weighting matrix is the most common in the spatial literature though we considered other versions for robustness (such as an inverse-distance-based matrix), but do not report those results here for brevity. Tables 1 and 2 help underscore the spatial spread of county-level Covid-19 cases—all parameters are statistically significant across the various models. The Wald test for spatial correlation is statistically significant for each model (the Wald results are not shown explicitly in Tables 1 and 2). As noted, we eschew focusing on the parameter estimates from Tables 1 and 2 an instead focus on the direct and indirect effects reported in Table 3.

dynamic model." Pace et al. (1998) provide an example with their "STAR" model. Brady (2014) provides a brief overview of some of these models. See also Debarsy et al. (2012) for discussion.

[9]For a "crib-note" version of these effects, the direct effect for the SDM model shown above can be expressed as, $\left(\frac{3-\rho^2}{1-\rho^2}\right)\beta_k + \left(\frac{3\rho}{3(1-\rho^2)}\right)\theta_k$, which is a re-print from Elhorst (2010) using an example with three regions. And, the indirect effect (again, from Elhorst (2010)) is, $\left(\frac{3\rho-\rho^2}{3(1-\rho^2)}\right)\beta_k + \left(\frac{3+\rho}{3(1-\rho^2)}\right)\theta_k$. The point, here, is the direct and indirect effects are a combinations of the baseline model parameters.

Covid Economics 43, 21 August 2020: 42-94

Table 3: Direct and Indirect Spatial Effects

	SAR	SLX	SDM	SDM Error	SAR	SLX	SDM	SDM Error
	Baseline				*Baseline with 14 day lag of r-score*			
Direct								
r-score	-2.039	1.155	0.880	1.202	-6.055	2.691	2.055	2.523
	0.000	*0.300*	*0.415*	*0.232*	*0.000*	*0.016*	*0.058*	*0.013*
Indirect								
r-score	-0.370	-3.900	-3.606	-4.030	-1.051	-10.701	-10.036	-10.729
	0.000	*0.001*	*0.002*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
Total								
r-score	-2.409	-2.745	-2.727	-2.827	-7.106	-8.011	-7.981	-8.206
	0.000	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
	SAR	SLX	SDM	SDM Error	SAR	SLX	SDM	SDM Error
	Dynamic				*Dynamic with 14 day lag of r-score*			
Direct								
cases (t-1)	0.049	0.054	0.049	0.054	0.049	0.054	0.049	0.053
	0.000	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
r-score	-5.924	3.183	2.452	3.280	-5.963	2.629	2.018	2.436
	0.000	*0.018*	*0.060*	*0.007*	*0.000*	*0.019*	*0.063*	*0.016*
Indirect								
cases (t-1)	0.008	–	0.008	0.056	0.008	–	0.008	0.055
	0.000	–	*0.000*	*0.000*	*0.000*	–	*0.000*	*0.000*
r-score	-1.016	-11.294	-10.522	-11.656	-1.005	-10.493	-9.845	-10.336
	0.000	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
Total								
cases (t-1)	0.058	0.054	0.058	0.110	0.057	0.054	0.057	0.108
	0.000	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
r-score	-6.940	-8.111	-8.070	-8.376	-6.969	-7.865	-7.826	-7.900
	0.000	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*

Notes: Spatial models estimated from April 1 through June 28, controlling for county-level fixed effects. P-values are in italics. Dynamic refers to the lag of the dependent variable included in the model. 14 day lag of r-score is the value from 14 days prior.

Covid Economics 43, 21 August 2020: 42-94

As displayed in Table 3, the direct effect of the r-score—the "own" county effect—is negative in the SAR model, but positive in the other models. This may reveal these estimates are affected by endogeneity between the r-score and the change in cases each day—that is, state governments are responding to their own case counts in determining the lock-down measures).

However, the indirect effect—the spatial spillover—of a change in the r-score in surrounding counties is negative in each model (with either the 14-day lag of the score or the current value). Note, for this variable in particular, the spillover will come from adjacent counties in other states, since our r-score variable will be identical for counties within a state. For the dynamic versions, the results are similar with respect to the signs of the effects—though for the r-score estimates (direct and indirect) the magnitudes are affected by the including of the temporal lag of cases.

The results from the standard spatial models indicate the relevance of spillovers related to the spread of Covid-19. This is not necessarily surprising, but this exercise provides clear statistical evidence on the spatial relationships not only between the spread of cases across regions but also how that spread was affected by the closure policies of adjacent states.

To further understand the spatial characteristics of the growth rate of new cases across counties, in the next section we consider the spatial "diffusion" of new cases across counties—that is, the temporal response of county i's cases in response to a change to its neighbors' caseloads.

3.2 Estimation of Spatial Impulse Response Functions

To estimate the temporal diffusion of the spatial connection between counties, we follow Brady (2014), Holly et al. (2011), Kuethe and Pede (2011), and Pollakowski and Ray (1997), and use an autoregressive-based strategy to measure the dynamics of spatial spillovers. While each of those studies focused on housing market spillovers, the methods therein are easily applicable to the county-level data on Covid-19 cases. In the latter three papers, a region's dependent variable is modeled as a function of the same variable in surrounding regions and the impulse response functions calculated to measure the spatial spillovers are estimated from a VAR.

In this application, we focus on single-equation estimation since, as demonstrated in Auerbach

Covid Economics 43, 21 August 2020: 42-94

and Gorodnichenko (2012), and Brady (2011, 2014), one can easily generate impulse response functions in the single-equation setting using Jordà (2005) local linear projections technique (as opposed to employing a fully-specified VAR). The basic motivation is that impulse response estimates can be produced from a single equation by projecting the endogenous variables in a system onto their lags for each horizon, h. Specifically, to paraphrase Auerbach and Gorodnichenko (2012), if $j = 1, ..., h$, and you are estimating some dependent variable as a function of its own lag and some other factor, X, then the IRF is estimated from the sequence of regressions:

$$Y_{t+1} = \rho_1 Y_{t-1} + \beta_1 X_t + \epsilon_t$$

$$Y_{t+2} = \rho_2 Y_{t-1} + \beta_2 X_t + \epsilon_t$$

$$\vdots$$

$$Y_{t+h} = \rho_h Y_{t-1} + \beta_h X_t + \epsilon_t$$

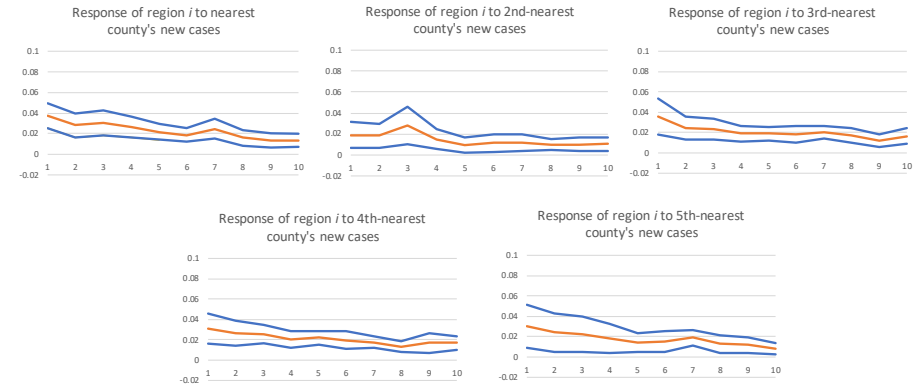
The impulse response estimates of the response of Y to a "shock" to X are, $IRF = \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_h$. This sequence of coefficient estimates approximates the impulse response coefficients you would recover from a VAR using recursive methods (under the null hypothesis as noted by Auerbach and Gorodnichenko (2012), which assumes the data generating processes are the same). However, the sequence estimated directly is not "tied" to the recursive structure of the VAR-generated impulse response functions—which is one reason Jordà (2005) mentions the the direct estimates are less subject to mis-specification.

Plagborg-Møller and Wolf (2019) provide a detailed discussion on the comparison between VAR-generated IRFs and the local projection approach; Rana and Shea (2015), Haug and Smith (2012) and Jordà et al. (2016, 2020) are other examples of the local projection technique. In the spatial-oriented literature, Brady (2011, 2014) estimate local projection-generated impulse response functions from SAR and dynamic SDM models for housing price data at the county level and state level in the U.S., respectively.

We first estimate county i's growth rate of new cases as a function of the same variable in its five nearest neighbors, and the $t - l$ lag of each, along with a county-level fixed effect. We define

Covid Economics 43, 21 August 2020: 42-94

Figure 6: Response of county i's new cases to a shock to new cases to cases its jth-nearest neighboring county



Notes: Estimated from single-equation autoregressive-distributed lag model with the time t and t-1 values of each independent variable (each neighbor), and including a county-level fixed effect. Nearest neighbor is measured by distance. New Cases measured as a simple growth rate.

"nearest neighbors" by distance, using geospatial coordinates to identify the closest neighboring county, the next closest, the third closest and so on up to the fifth closest neighbor.[10]

As an alternative, we eschew the fixed effect and instead include the following county and state level variables: the r-score for county i and the r-scores of its five-closest neighbors; the population density of county i and the population density of the five closest neighbors; the share of county i's population over the age of 60; and the share of voters in county i that voted Republican in the 2016 presidential election.

Figure 6 reports the IRFs of county i's response to new cases in its nearest five neighbors over

[10]This is similar to Pollakowski and Ray (1997) who include the lags of price changes in adjacent regions included as regressors in an equation for a change in housing prices across regions within the United States. Kuethe and Pede (2011) do something similar in their study of housing prices in the western part of the United States, as do Holly et al. (2011) in their analysis of housing prices in the United Kingdom. One difference from our estimation with Holly et al. (2011) is we do not specify a "dominant" region, since we wish to leave the possible spread "unconstrained" in the estimation.

Covid Economics 43, 21 August 2020: 42-94

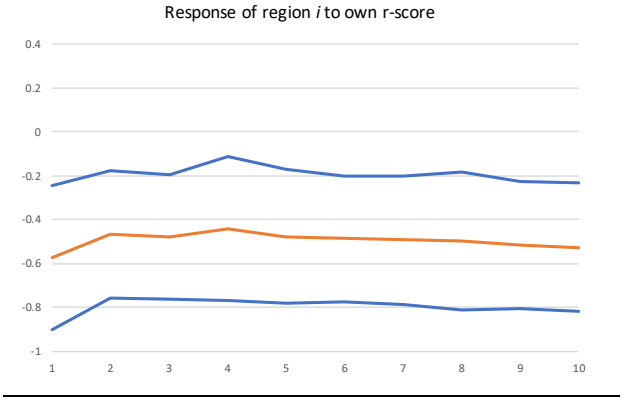
Figure 7: Response of county i's new cases to a shock to new cases to cases its jth-nearest neighboring county: county and state controls

Notes: Estimated using local projection technique from single-equation autoregressive-distributed lag model; specifications included county and state level controls. See text for details. Nearest neighbor is measured by distance. New cases measured as simple growth rate. Sample from March 31st through June 28th.

a forecast horizon of ten days (95 percent confidence intervals are constructed using Driscoll-Kraay standard errors). These results are from the specification that includes the county-level fixed effect. For each of the nearest five neighbors, the response of new cases in county i increases and this increase persists for the duration of the forecast horizon (with a gradual decay). The effect of each neighbor on county i's cases is about the same, which makes sense from a geographic perspective. Since the "nearest" is measured using latitude and longitude coordinates, in general the nearest neighbors will be those that surround the county in its own state or in bordering states. In other words, the IRFs reveal that spillovers, on average, appear to come from all sides of county i. With respect to the magnitudes, recall that "new cases" is measured as a simple day-to-day growth rates.

Figure 7 displays the same estimates, but for the specification with county-level and state-level covariates. The responses do not differ much compared to those displayed in Figure 6. With respect to the response of the new cases in county i to the other covariates, only for the r-score can we report any meaningful dynamics. The population density and the other shares-variables mentioned

Figure 8: Response of County New Cases to state-level r-score



Notes: See text for definition of r-score

above do not vary over time in our sample.

For the r-score, which is a state-level variable, we find that only the within-state r-score matters for the growth rate of new cases. Figure 7 displays this response. At each horizon the response is negative. It is worth noting, of course, that the dynamics of the r-score variable are "lumpy"—the r-score changes at discrete and infrequent intervals, with the frequency and the value depending on the state. Moreover, we find little response of the response of county i's new cases to the r-score of counties in surrounding states (therefore, we do not report these figures). This, too, may not be surprising since this effect will only record the relationship between "nearest neighbors" if those counties are in different states. We also consider a variation of this relationship and isolate only those counties that had a nearest neighbor in a state where the r-score differs by two or more. However, we find no statistically significant relationship in that case either.

On balance, the "spatial" IRFs support the notion that the number of new cases easily spread across counties and states and provide a "stylized" picture of that spatial diffusion—the spatial diffusion is statistically significant and persistent over time. The spatial IRFs provide an alternative

Covid Economics 43, 21 August 2020: 42-94

perspective from the "static" spatial models estimated in section 3. Considered together, however, both the traditional "off-the-shelf" spatial models and the spatial IRFs provide an informative picture on the nature of the spatial correlation of the Covid-19 phenomenon. As an added perspective, too, the estimation provides some insight into the effectiveness of the state-level closures—more stringent measures are consistent with a decline in the growth rate of new cases at the county level.

4 Spatial SIR and Counterfactual Experiments

The previous sections provide evidence of a substantial degree of inter-state spillovers of Covid-19 across regions. In this section, motivated by that evidence, we construct and calibrate a structural SIR model, similar to the one described in Atkeson (2020b), Eichenbaum et al. (2020), Glover et al. (2020), or Fernández-Villaverde and Jones (2020), but with a few modifications. Most importantly, the model allows for infections across state boundaries. We then use the model to evaluate the extent to which the presence of such spillovers contributed to the spread of the infections in the U.S. and within each state. We also evaluate how lock-down policies implemented in one state impact the rest of the country.

At this stage, our results need to be treated with some caution. The outbreak is still in its early phase, and our data covers only the first 5 months, with many unreported cases, and with a possibly large but unknown number of cases "imported" from other countries (the model only considers internal transmission). We abstract from a number of features that affect the spread of the disease over time, such as voluntary changes in people's behavior[11] or the change in infectiousness of people who are still infected. Additionally, as pointed out by Fernández-Villaverde and Jones (2020), the identification of parameters in the compartmental models such as the SIR model can be challenging (this is mostly discussed by Atkeson (2020a) in the context of estimating the fatality rate, which is not the central point of our analysis). We partially address this last issue by con-

[11] We account to some extent for the impact of changes in social distancing measures and the state-specific effectiveness of those.

Covid Economics 43, 21 August 2020: 42-94

sidering alternative specifications that differ with respect to the free parameters, and by exploring how those specifications affect different parts of the model fit. We believe our preliminary analysis can still provide useful insights into both the nature and the potential magnitude of inter-state spillovers.

4.1 The model

The model is an extension of the SIR model that allows us to account for the spatial diffusion of infections. We specify the model in discrete, rather than continuous time. In each period t, the initial population of region n is divided into four disjoint sets: Susceptible (S), Infected (I), Recovered (R), and Dead (D):

$$Pop_{n,0} = S_{n,t} + I_{n,t} + R_{n,t} + D_{n,t}$$

and population at time t is: $Pop_{n,t} = S_{n,t} + I_{n,t} + R_{n,t}$. The new infections in state n result from interactions between susceptible people S_n in that state, with infected people in potentially all other states $I_{n'}$, where $n' = 1, ..., N$. The new infections in state n at time t are given by:

$$I_{n,t}^{new} = \frac{S_{n,t}}{Pop_{n,t}} \cdot \sum_{n'} \rho(n', n) \cdot \sqrt{\beta_n \beta_{n'}} \cdot \sqrt{\kappa_{n,t} \kappa_{n',t}} \cdot I_{n',t}$$

In the expression above, the whole term $\rho(n', n) \cdot \sqrt{\beta_n \beta_{n'}} \cdot \sqrt{\kappa_{n,t} \kappa_{n',t}}$ describes the average number of close contacts that a person from state n has with a person from state n' in day t. The close contact is defined as one that would result in a transmission of a virus from an infected person to a healthy person. The new infections in state n then occur when an infected person from state n' — $I_{n',t}$ — comes in a close contact with a susceptible person from state n. The probability that a person we come in a close contact with is susceptible is $\frac{S_{n,t}}{Pop_{n,t}}$.

The parameter β_n measures the average number of distinct inter-personal contacts that any person in state n has on a regular day. We allow this parameter to vary across states, given the substantial heterogeneity in the fraction of people living in densely populated areas. We expect, of course, that a typical person in New York will have more distinct inter-personal contacts than

Covid Economics 43, 21 August 2020: 42-94

Covid Economics 43, 21 August 2020: 42-94

a person living in Montana. At this stage we assume β_n is constant over time. It is certainly possible that the typical number of inter-personal contacts will vary over time in each state, and it is quite likely that this variation will differ by state (for example, the value of β would likely plummet during the Spring Break in college towns but skyrocket in the nightclubs or bars in Florida). Given how specific this time variation would be to individual states, we have decided to assume it away, and only allow the model to have a cross-sectional variation in β, which yields 49 parameters to be calibrated. We also consider a simpler specification, with three rather than 49 parameters, where β_n is a polynomial function of the population density in region n: $\beta_n = b_0 + b_1 \log(density) + b_1 \log(density)^2$.

Next, $\kappa_{n,t}$ measures the degree to which the personal interactions are reduced by the implemented lock-down policies. The actual reduction in the personal interactions results from a combination of two factors: the official lock-down policies and their effectiveness in the particular region. That effectiveness (from the perspective of the model) can capture at least two important factors. The first factor is related to individuals' compliance and the region's enforcement of social distancing measures. The second factor is related to the fact that each social distancing measure (as recorded in our data) comes with exceptions. Those exceptions may be different in different states, or the same exception can have a different coverage in different states. In general, we should not expect the same restriction that we code as a particular value of the r-score variable to have an identical impact in each state. While we cannot speak to the reasons behind that heterogeneity, we can incorporate it in a straightforward fashion into our model. In order to do that we model $\kappa_{n,t}$ as follows:

$$\kappa_{n,t} = (1 - \xi_n) + \xi_n \cdot \sum_{i=0}^{5} \kappa^i \cdot 1_{\{i\}} \left(\text{r-score}_{n,t}\right)$$

where i is the value of the r-score variable (0 through 5), κ^i is the benchmark effect of restriction i in the region where restrictions are most effective, ξ_n is the relative effectiveness of restrictions in region n, and $1_{\{i\}}(\cdot)$ is a characteristic function of a singleton set with element i (essentially, $1_{\{i\}} \left(\text{r-score}_{n,t}\right)$ equals 1 if $\text{r-score}_{n,t} = i$ and 0 otherwise). We normalize $\xi_n = 1$ in one of the regions

(determined endogenously), and calibrate the 48 remaining values. We also normalize $\kappa^0 = 1$ ($i = 0$ corresponds to no restrictions). We also impose a restriction that $\kappa^{i+1} \leq \kappa^i$, so a tighter restriction would never lead to *more* contacts between people. Overall, this adds $48 + 5 = 53$ additional parameters to the calibration. We also consider a simpler specification where $\xi_n = 1$ in every region n.

Finally, $\rho(n', n)$ denotes the spillover parameter from state n' to n. We restrict the possible values for $\rho(n', n)$. First, we normalize $\rho(n, n) = 1$. Next, we set $\rho(n', n) = 0$ when two states n' and n are not adjacent and we require it to be positive (even if arbitrarily small) when they are. In that case, we set $\rho(n', n) = \rho \cdot \frac{1}{\sum_m 1_{\{x \in \mathbb{R}: x > 0\}}(\rho(n', m))}$, where $\rho > 0$ will be the parameter to be calibrated. In words, the spillover from state n' to state n is divided by the total number states that the state n' is adjacent to. We do that in order to ensure that if Virginia and Maryland were one state, the total spillover from DC would be the same as it is when they are two separate states.[12]

The full dynamics of the model are described by the following equations:

$$S_{n,t+1} = S_{n,t} - I^{new}_{n,t} \tag{4.1}$$

$$I_{n,t+1} = I_{n,t} - \pi_R \cdot I_{n,t} - \pi_D I_{n,t} + I^{new}_{n,t} \tag{4.2}$$

$$R_{n,t+1} = R_{n,t} + \pi_R \cdot I_{n,t} \tag{4.3}$$

$$D_{n,t+1} = D_{n,t} + \pi_D \cdot I_{n,t} \tag{4.4}$$

$$Pop_{n,t+1} = Pop_{n,t} - \pi_D \cdot I_{n,t} \tag{4.5}$$

where π_R is the daily recovery rate and π_D is the daily death rate. We set $\pi_R = 0.03267$ and $\pi_D = 0.00067$, so that the model implies a 2% mortality and a 30-day duration of an average infection.

[12] Allowing for $\rho(n', n)$ to have distinct value for each pair of states would yield $49 \times 24 = 1,176$ parameters to be calibrated if we assume symmetric spillovers, and double that if we do not.

Covid Economics 43, 21 August 2020: 42-94

Covid Economics 43, 21 August 2020: 42-94

4.2 Calibration, model Fit and Parameter Values

We calibrate the model by minimizing the sum of squared errors between the data and the model-generated series of both the cumulative and the new confirmed cases per-capita in each region and in the entire country. In the benchmark calibration our vector of parameters has 103 elements:[13]

$$\theta := [\rho, \beta_1, ..., \beta_{49}, \kappa_1, ..., \kappa_5, \xi_1, ..., \xi_{48}]$$

In our calibration we assume that the confirmed cases per-capita in each state lag the infections by 14 days and we start our analysis on February 1, 2020 under the assumption that the cumulative infections on that day corresponded to confirmed cases on February 14, 2020. Our two main outcome variables are then defined as:

$$y_{n,t} := \frac{I_{n,t-7}}{Pop_{n,0}} \quad \text{and} \quad Y_t := \frac{\sum_n I_{n,t-7}}{\sum_n Pop_{n,0}}, \qquad t = 15, 16, ...$$

The sum of squared errors between the model and the data is then calculated as:

$$SSE(\theta) = \sum_n \left(\sum_{t=15}^{T} \left(y_{n,t}^m(\theta) - y_{n,t}^d \right)^2 + \sum_{t=16}^{T} \left(\Delta y_{n,t}^m(\theta) - \Delta y_{n,t}^d \right)^2 \right) + \sum_{t=15}^{T} \left(Y_t^m(\theta) - Y_t^d \right)^2 + \sum_{t=16}^{T} \left(\Delta Y_t^m(\theta) - \Delta Y_t^d \right)^2$$

The vector of calibrated parameters $\hat{\theta}$ is then given as

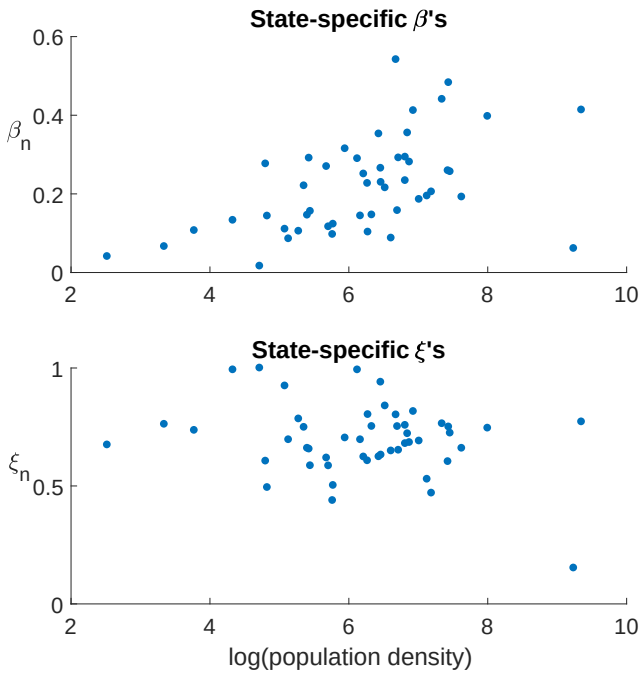
$$\hat{\theta} := \arg\min_{\theta} SSE(\theta)$$

The results of the calibration are reported in Table 4, which displays the overall fit of the model as well as the values of selected parameters, except for the individual regions' values of β_n and ξ_n. The latter are reported in the appendix. Here, we plot state-specific parameter values against each region's measure of population density in the two panels of Figure 9.[14] The parameter values

[13] The two alternative calibrations we consider have the following parameter vectors: (1) $\theta := [\rho, \beta_1, ..., \beta_{49}, \kappa_1, ..., \kappa_5]$ when we assume that $\xi_n = 1$ in each region n, and (2) $\theta := [\rho, b_0, b_1, b_2, \kappa_1, ..., \kappa_5]$ when we additionally assume that $\beta_n = b_0 + b_1 \log(density) + b_1 \log(density)^2$.

[14] We compute the population density for each state as a population-weighted average of the density in each county.

Figure 9: State-specific values of β_n and ξ_n vs. log of population density

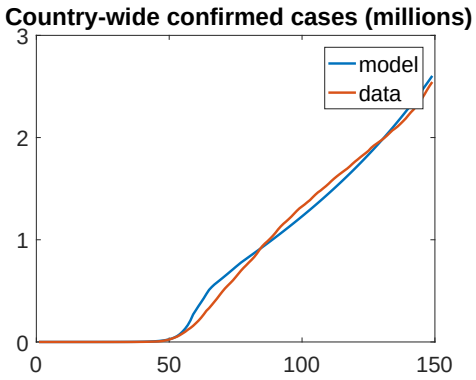


by themselves do not mean much, so we defer their discussion to Section 4.3 where we perform counterfactual simulations. There are two main takeaways here. First, the calibrated values of κ's are much smaller in our benchmark calibration (when we allow ξ_n to vary by state) than when we assume identical, perfect effectiveness of each restriction ($\xi_n = 1$ for all n). This is not surprising at all; higher values κ's in the first two columns of Table 4 reflect the fact that in the average state the effectiveness of that restriction measure is not perfect. Second, modeling β_n as a fixed effect rather than a simple polynomial function of the population density makes a huge difference in terms of the model's ability to account for the variation of Covid-19 cases across states.

While it may not seem so, our model has relatively few free parameters (even with 49 state-

Covid Economics 43, 21 August 2020: 42-94

Figure 10: Model Fit - all confirmed cases



specific values of β and 48 state-specific values of ξ), because we have over 7,000 observations (49 regions over 150+ days). Despite that, the model does a remarkably good job in replicating the data. Figure 10 plots the total number of confirmed cases observed in the data and generated by the model for the whole country. The time paths of confirmed cases per-capita for individual states and for DC are reported in the appendix. The R^2 between the cumulative infections for the whole country in model and in the data is 0.99. For individual states, the model accounts for 98% of the overall variation in cumulative infections, for 97% of the variation within states, and 99% of the variation between states. Naturally, the model does a poorer job in accounting for the dynamics of the new infections. For the whole country, it accounts for the 73% of the variation in the data. For individual states, it accounts for 40% of the total variation, 30% of the variation within states, but for the 97% of the variation between states.

4.3 Counter-factual simulations

Given the overall good fit of the structural model, we proceed with using the model to perform two counterfactual simulations. In all counterfactual simulations, we use our benchmark parametriza-

Covid Economics 43, 21 August 2020: 42-94

Covid Economics 43, 21 August 2020: 42-94

Table 4: Model Fit and Parameter Values

Model specification	$\beta_n = f(density_n)$; $\xi_n = 1$	β_n as fixed effects; $\xi_n = 1$	ξ_n and β_n as fixed effects
Parameters			
ρ	0.144	0.094	0.167
κ^1	0.519	0.999	0.829
κ^2	0.519	0.997	0.477
κ^3	0.518	0.593	0.377
κ^4	0.322	0.359	0.114
κ^5	0.297	0.277	0.023
b_0	-4.999	n.a.	n.a.
b_1	0.717	n.a.	n.a.
b_2	-0.029	n.a.	n.a.
Model Fit - levels			
country-wide	0.984	0.987	0.993
states - total	0.470	0.960	0.976
states - within	0.545	0.947	0.970
states - between	0.375	0.991	0.992
Model Fit - first differences			
country-wide	0.574	0.580	0.725
states - total	0.131	0.301	0.395
states - within	0.107	0.207	0.303
states - between	0.291	0.919	0.968

Notes: $f(density) \equiv b_0 + b_1 \log(density) + b_2 \log(density)^2$.

Covid Economics 43, 21 August 2020: 42-94

tion with state-specific values of β_n and ξ_n. Naturally, any counterfactual simulation of a calibrated or estimated model that does not explicitly model people's behavior has to address the Lucas' critique (Lucas, 1976). We want to point out that to some extent we already capture the differences across states in behavioral response to restrictions by calibrating a state-specific parameter ξ_n. While it captures more than that, it partially mitigates those concerns. Ideally, we would have ξ_n vary by the level of imposed restriction, but we are not able to identify that with current data. Given these considerations, our results should be interpreted as showing the impact of changes in restrictions under the assumption that compliance with them remains the same as it was before (i.e., not necessarily perfect).

Any counterfactual experiment is to some extent *ad hoc*. We present a few that we find the most interesting and informative. First, we investigate what would happen if states increased their maximum level of social distancing measures by 1 (of course, excluding the states where the max of the r-score variable equals 5). Second, we investigate what would happen if the states adopted their maximum level of social distancing 14 days earlier. Third, we investigate the role of the spillover parameter, by simulating the path of infections when its value is 25% lower. Finally, we combine the first with the last by simulating the impact of changing the restrictions in the environment where the spillover parameter is smaller (of course, using the simulated series from the third counterfactual as a benchmark).

The results of the counterfactual experiments are presented in Tables 5 and 6, and in Figures 11 through 14. In Figures 11-13 the first column indicates the impact on the whole country, the second column indicates the impact on the group of states where the counterfactual policy is implemented ("own effect"), and the third column indicates the impact on the group of remaining states ("spillover effect"). These two groups may be different for different counterfactuals, and in a few cases the second group is an empty set (so the spillover effect is not applicable). Tables 5 and 6 additionally separates these groups into two sub-columns - in each case the left sub-column shows the simulated effect at the end of our data sample (June 28, 2020) and the right

Covid Economics 43, 21 August 2020: 42-94

sub-column shows the simulated effect by December 31, 2021.[15] In all simulations we assume that social distancing measures which are in place on June 28, 2020, remain in place forever. One can interpret the difference between them as showing the short-term vs. the long-term effects. Table 5 shows percentage change, and Table 6 shows the change in the actual numbers of people.

Raising/lowering the max r-score by 1 The effects of increasing the maximum level of restrictions by 1 (or lowering it by 1 among states where max(r-score) = 5) are shown in rows 1, 3, 5, 7-9 of Tables 5 and 6 . The first thing we notice is that the own effect is an order of magnitude larger than the spillover effect. Second, with the exception of the counterfactual in row 9 (most restrictive states reduce their restrictions by 1), the spillover effect in the long-run is stronger than in the short-run. Third, even though the relative spillover effect is much smaller than the own effect, it can be quite sizeable. Among states where the max(r-score) = 4, the spillover effect in the long-run is -4.2%, and the combined spillover effect among states with max(r-score) < 5 is -2.2% in the short-run and -5.3% in the long-run (in this case, the spillover effect is the impact on states with max(r-score) = 5). While the percentages look small, as of 6/28/2020, the total number of confirmed cases among states with max(r-score) = 5 was 1.7 mln, so 2.2% corresponds to 37,000 people. The -5.3% in the long-run corresponds to reducing the cumulative confirmed cases (in states other those that raise their restrictions) by 4 million towards the end of December 2021.[16] The short-term effects are depicted in Figures 11, 12, and 13.

Adopting max r-score earlier Similarly to changing the level of restrictions, earlier adoption of the maximum level of restrictions matters most when done by states for which the maximum r-score is 5. The difference is that now the short-term effect is much stronger than the long-term effect. In the short-term, the own effect from early adoption for states with r-score = 5 is -84%, and the spillover effect is -36%. In the long-term, those effects are drastically diminished, to -2.3%

[15] Naturally, the long-term effects should be interpreted more cautiously.

[16] Long-run simulations in all scenarios assume that states keep their maximum level of restrictions until the end of the simulation sample.

Covid Economics 43, 21 August 2020: 42-94

Table 5: % change in the number of confirmed cases (from the benchmark)

Counterfactual (states; action)	country-wide		own effect		spillover	
	short-run	long-run	short-run	long-run	short-run	long-run
r-score = 2; max add 1	-0.19	-0.26	-36.33	-32.46	-0.03	-0.04
r-score = 2; set max earlier	-0.16	0.00	-25.21	-0.07	-0.04	0.00
r-score = 3; max add 1	-3.83	-5.49	-62.74	-51.64	-0.55	-0.98
r-score = 3; set max earlier	-1.77	0.00	-26.26	-0.02	-0.41	0.00
r-score = 4; max add 1	-12.56	-21.56	-43.22	-52.35	-1.94	-4.20
r-score = 4; set max earlier	-19.21	-0.26	-60.53	-0.60	-4.89	-0.07
r-score = 2,3,4; max add 1	-16.36	-27.43	-47.28	-53.80	-2.17	-5.32
r-score = 2,3,4; max add 1; ($\rho \downarrow$ by 25%)	-13.03	-28.26	-44.29	-55.64	-1.11	-5.03
r-score = 5; max less 1	92.25	40.02	125.91	67.62	18.91	7.11
r-score = 5; set max earlier	-69.35	-1.58	-84.60	-2.32	-36.13	-0.69
all; set max earlier	-85.32	-1.89	-85.32	-1.89	n.a.	n.a.
all; set max on 3/19/2020	-36.76	-0.55	-36.76	-0.55	n.a.	n.a.
all; $\rho \downarrow$ by 25%	-38.84	-6.45	-38.84	-6.45	n.a.	n.a.

Notes: On 3/19/2020 California was the first state that raised its social distancing measures to the level which corresponds to r-score = 5. "set max earlier" means that the state sets the r-score to the state's specific maximum value 14 days earlier than in the data. "max add 1" means that when the state's r-score reaches the maximum value, we increase that maximum value by 1 (for states where max(r-score) = 5 "max less 1" means that when such state's r-score reaches maximum, it is set to 4). Short-run refers to simulated outcomes on 6/28/2020. Long-run refers to simulated outcomes on 12/31/2021. In the long-run simulations we assumed that restrictions in place on 6/28/2020, remain in effect until the end of the simulation period.

Covid Economics 43, 21 August 2020: 42-94

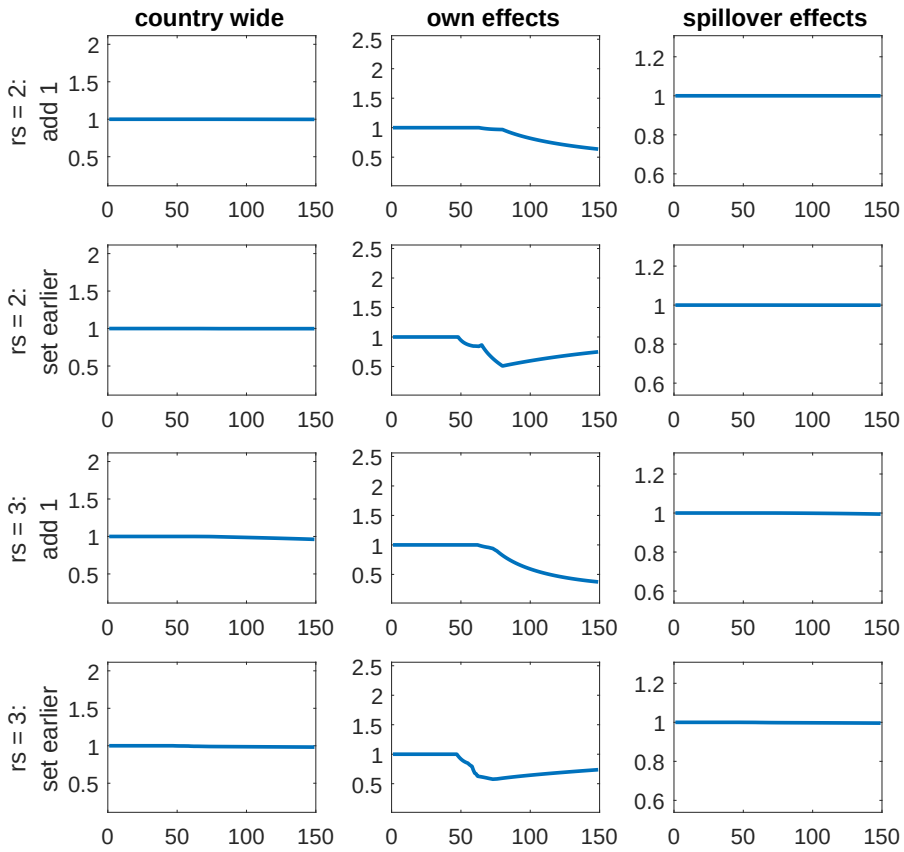
Table 6: change in the number of confirmed cases (from the benchmark) in thousands

Counterfactual (states; action)	country-wide		own effect		spillover	
	short-run	long-run	short-run	long-run	short-run	long-run
r-score = 2; max add 1	-5.1	-369.8	-4.4	-307.7	-0.7	-62.2
r-score = 2; set max earlier	-4.0	-0.9	-3.0	-0.7	-1.0	-0.2
r-score = 3; max add 1	-99.6	-7,810.8	-86.0	-6,534.7	-13.7	-1,276.1
r-score = 3; set max earlier	-46.0	-4.8	-36.0	-3.1	-10.0	-1.6
r-score = 4; max add 1	-327.0	-30,678.8	-289.5	-26,852.9	-37.5	-3,825.9
r-score = 4; set max earlier	-500.1	-373.6	-405.5	-307.5	-94.6	-66.2
r-score = 2,3,4; max add 1	-425.9	-39,029.8	-387.2	-34,914.4	-38.7	-4,115.5
r-score = 2,3,4; max add 1; ($\rho \downarrow$ by 25%)	-207.5	-37,609.2	-194.7	-33,983.2	-12.8	-3,626.0
r-score = 5; max less 1	2,401.7	56,940.0	2,246.8	52,325.5	154.9	4,614.5
r-score = 5; set max earlier	-1,805.6	-2,244.7	-1,509.7	-1,794.4	-295.9	-450.3
all; set max earlier	-2,221.3	-2,688.1	-2,221.3	-2,688.1	n.a.	n.a.
all; set max on 3/19/2020	-957.1	-780.6	-957.1	-780.6	n.a.	n.a.
all; $\rho \downarrow$ by 25%	-1,011.2	-9,177.2	-1,011.2	-9,177.2	n.a.	n.a.

Notes: See Table 5.

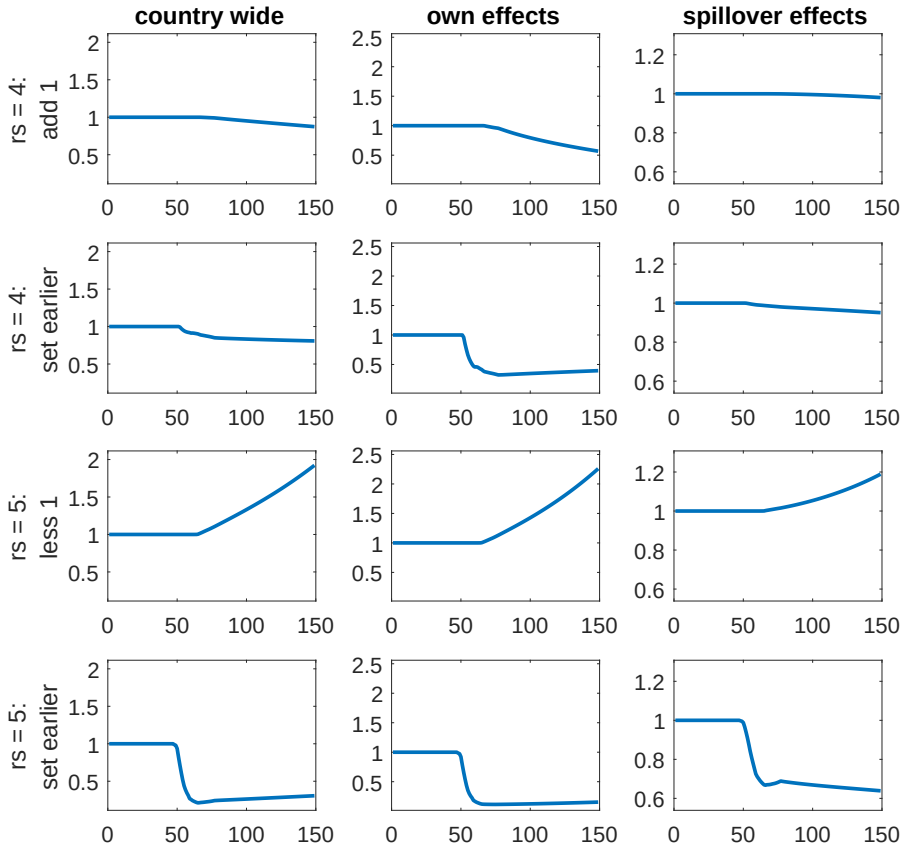
Covid Economics 43, 21 August 2020: 42-94

Figure 11: Changing restrictions in the more lax states (benchmark = 1)



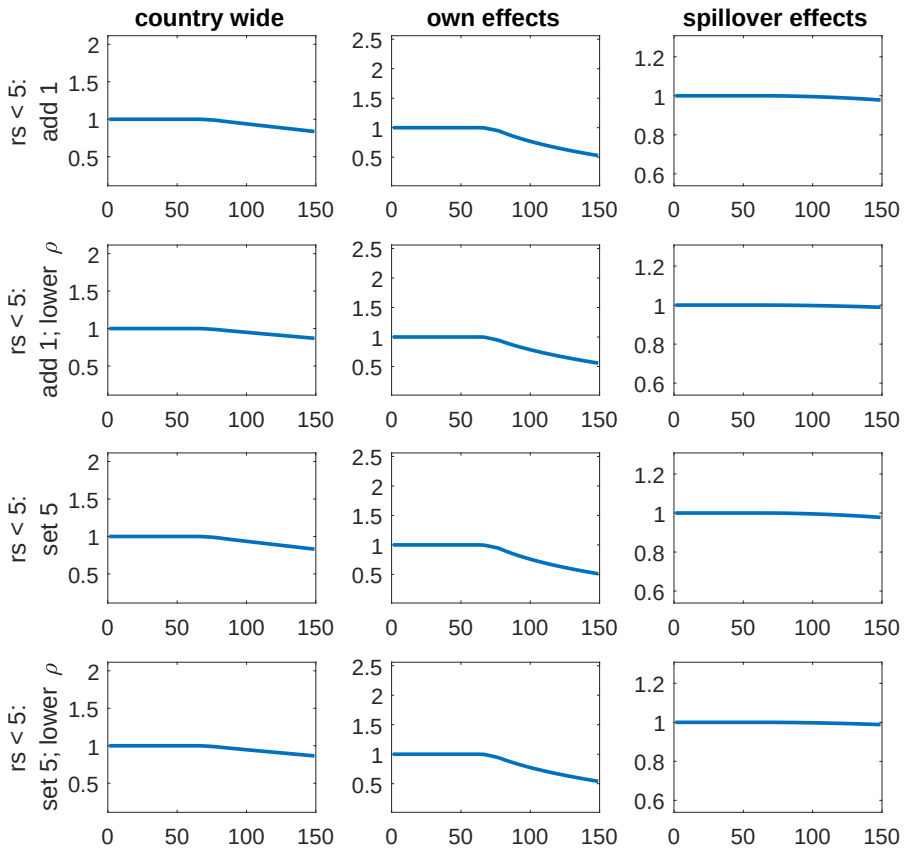
Covid Economics 43, 21 August 2020: 42-94

Figure 12: Changing restrictions in the more restrictive states (benchmark = 1)



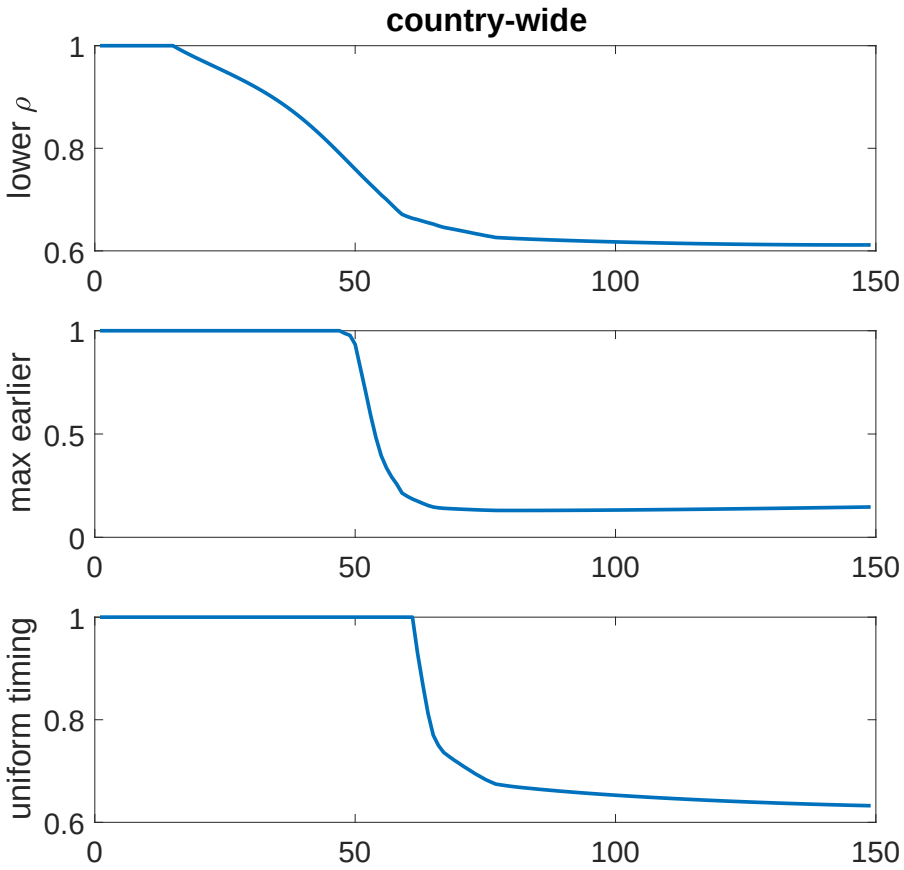
Covid Economics 43, 21 August 2020: 42-94

Figure 13: Changing restrictions in all but most restrictive states (benchmark = 1)



Covid Economics 43, 21 August 2020: 42-94

Figure 14: Lower spillovers, early restriction, and uniform timing (benchmark = 1)



Covid Economics 43, 21 August 2020: 42-94

and -0.7%, respectively. In other words, changing the timing of restrictions (without changing its ultimate maximum level), operates largely by flattening the infection curve rather than by reducing the total number of people that will eventually become infected. Among states that never reached the r-score of 5, we observe a similar pattern, but its magnitude is smaller. The short-term effects are depicted in Figures 11 and 12, and in the second panel of Figure 14.

Lower spillover The effect of reducing the value of the spillover parameter is shown in the top panel of Figure 14 and in the bottom row of Tables 5 and 6. Not surprisingly, if spillovers across states' borders are smaller, the cumulative infections decline. Interestingly, the long-term effect is disproportionately smaller (see Tables 5 and 6). In other words, if U.S. states had the ability to restrict travel between them (akin to border closures between countries in the Schengen Zone), the main epidemiological benefit would operate through the flattening of the infection curve.[17] The total number of confirmed cases in the long-run would still be smaller, but that decline would be nowhere near as large as the decline in the short-term. Finally, when the spillover parameter is smaller, the impact of changing the restrictions is diminished substantially in the short-run, but only marginally in the long-run (see Figure 13 and rows 7 and 8 in Tables 5 and 6).

Fragmented by policies, united by outcomes Finally, Table 7 shows how changing a policy within one group of states impacts that group as well as other groups. The key message from the table is quite clear - policy changes implemented by subsets of states have quantitatively significant effects on the rest of the country. In that sense, while a policy change may be local and uncoordinated, its impact is felt across the whole country.

[17] Eckardt et al. (2020) show that border closures between European regions significantly slowed down the spread of the virus.

Table 7: Fraction of people infected by 12/31/2020 (median within subsets of states)

	Subset of states		
	max(r-score) = 3	max(r-score) = 4	max(r-score) = 5
benchmark	0.61	0.38	0.33
r-score = 3: add 1	0.49	0.31	0.26
r-score = 4: add 1	0.42	0.17	0.28
r-score = 5: less 1	0.65	0.53	0.60
r-score = 2, 3, 4: add 1	0.30	0.12	0.22

5 Conclusion

In this paper we estimate the magnitude of inter-state diffusion of the Covid-19 infections. We find evidence that new cases diffuse across county lines, and that the spatial diffusion across counties is affected by the closure policies of adjacent states. Using a spatial version of the SIR model we find that tightening restrictions in states with the less restrictive policies could have reduced the infections in *other* states by more 2% in the first 3 months, and by more than 5% by the end of December 2021, corresponding to a reduction in the number of confirmed cases by 40,000 and 4 million, respectively.

The stylized picture of spatial diffusion provides consistent evidence of the effect of surrounding states' policies on a particular county's case load. The estimates from the spatial impulse response functions suggest that spatial correlation is significant for up to ten days (following a "shock" to new cases); while the estimates from traditional spatial models show the spatial correlation is significant between counties, and the r-score of counties in adjacent counties have a significant effect on the growth rate of a county's new cases.

The presence of inter-state spillovers significantly affected the rate of increase in the number of confirmed cases in the early stages of the outbreak. This result is important when we evaluate the

Covid Economics 43, 21 August 2020: 42-94

Covid Economics 43, 21 August 2020: 42-94

"performance" of different regions in battling the pandemic. A unique feature of the United States is that its federal government cannot compel individual states to simply close their borders nor mandate state-specific lock-down policies. This only emphasizes the importance of other tools that promote coordination between states' authorities and regular citizens. First, uniform and consistent messaging on precautionary measures such as masks, or encouraging the compliance with social distancing restrictions and discouraging unnecessary inter-state travel are examples of such tools that would impact individual behavior. Second, the evidence provided in the literature thus far (Piguillem and Shi, 2020; Berger et al., 2020) suggests that there are potentially huge benefits from implementing a country-wide testing system—aimed at reducing the delay between test and result—thus revealing virus hot-spots much sooner to potential travelers. Finally, the literature on fiscal federalism may offer some insights into the role the federal government can play when the jurisdictional boundaries do not overlap with the boundaries of regions affected by local policies.[18] Given that by the very nature of the problem any policy implemented or not implemented in response to a viral outbreak creates external effects on surrounding regions, we believe this is a very important area for further research.

References

Acemoglu, D., V. Chernozhukov, I. Werning, and M. D. Whinston (2020): "Optimal Targeted Lockdowns in a Multi-Group SIR Model," Working Paper 27102, National Bureau of Economic Research.

Atkeson, A. (2020a): "How Deadly Is COVID-19? Understanding The Difficulties With Estimation Of Its Fatality Rate," Working Paper 26965, National Bureau of Economic Research.

——— (2020b): "What Will Be the Economic Impact of COVID-19 in the US? Rough Estimates of Disease Scenarios," Working Paper 26867, National Bureau of Economic Research.

[18] See Oates (1999) for the literature review on that topic. Rothert (2020) discusses a few examples of federal fiscal tools that could impact local policies.

Atkeson, A., K. Kopecky, and T. Zha (2020): "Estimating and Forecasting Disease Scenarios for COVID-19 with an SIR Model," Working Paper 27335, National Bureau of Economic Research.

Auerbach, A. J. and Y. Gorodnichenko (2012): "Measuring the output responses to fiscal policy," *American Economic Journal: Economic Policy*, 4, 1–27.

Beck, T. and W. Wagner (2020): "National containment policies and international cooperation," *Covid Economics*, 8, 120–134.

Berger, D. W., K. F. Herkenhoff, and S. Mongey (2020): "An SEIR Infectious Disease Model with Testing and Conditional Quarantine," Working Paper 26901, National Bureau of Economic Research.

Bisin, A. and A. Moro (2020): "Learning Epidemiology by Doing: The Empirical Implications of a Spatial-SIR Model with Behavioral Responses," Working Paper 27590, National Bureau of Economic Research.

Brady, R. R. (2011): "Measuring the diffusion of housing prices across space and over time," *Journal of Applied Econometrics*, 26, 213–231.

——— (2014): "The spatial diffusion of regional housing prices across US states," *Regional Science and Urban Economics*, 46, 150–166.

Briscese, G., N. Lacetera, M. Macis, and M. Tonin (2020): "Compliance with COVID-19 Social-Distancing Measures in Italy: The Role of Expectations and Duration," Working Paper 26916, National Bureau of Economic Research.

Cuñat, A. and R. Zymek (2020): "The (structural) gravity of epidemics," *Covid Economics*, 17.

Debarsy, N., C. Ertur, and J. P. LeSage (2012): "Interpreting dynamic space–time panel data models," *Statistical Methodology*, 9, 158–171.

Covid Economics 43, 21 August 2020: 42-94

Desmet, K. and R. Wacziarg (2020): "Understanding Spatial Variation in COVID-19 across the United States," Working Paper 27329, National Bureau of Economic Research.

Dong, E., H. Du, and L. Gardner (2020): "An interactive web-based dashboard to track COVID-19 in real time," *Lancet Infectious Diseases*, 20, 533–534.

Eckardt, M., K. Kappner, and N. Wolfl (2020): "Covid-19 across European Regions: the Role of Border Controls," Working Paper DP15178, Center for Economic and Policy Research.

Eichenbaum, M. S., S. Rebelo, and M. Trabandt (2020): "The Macroeconomics of Epidemics," Working Paper 26882, National Bureau of Economic Research.

Elhorst, J. (2014): "Spatial Panel Models [in:] Fisher, MM, Nijkamp, P.(eds.) Handbook of Regional Science," .

Elhorst, J. P. (2010): "Applied spatial econometrics: raising the bar," *Spatial economic analysis*, 5, 9–28.

——— (2012): "Dynamic spatial panels: models, methods, and inferences," *Journal of geographical systems*, 14, 5–28.

Ellison, G. (2020): "Implications of Heterogeneous SIR Models for Analyses of COVID-19," Working Paper 27373, National Bureau of Economic Research.

Favero, C. (2020): "Why is Covid-19 mortality in Lombardy so high? Evidence from the simulation of a SEIHCR model," *Covid Economics*, 4.

Fernández-Villaverde, J. and C. I. Jones (2020): "Estimating and Simulating a SIRD Model of COVID-19 for Many Countries, States, and Cities," Working Paper 27128, National Bureau of Economic Research.

Gerritse, M. (2020): "Cities and COVID-19 infections: population density, transmission speeds and sheltering responses," *Covid Economics*, 37.

Covid Economics 43, 21 August 2020: 42-94

Ginsburgh, V., G. Magerman, and I. Natali (2020): "COVID-19 and the Role of Economic Conditions in French Regional Departments," Working Papers ECARES 2020-17, ULB – Universite Libre de Bruxelles.

Glover, A., J. Heathcote, D. Krueger, and J.-V. Rios-Rull (2020): "Health versus Wealth: On the Distributional Effects of Controlling a Pandemic," Working Paper DP14606, Center for Economic and Policy Research.

Golgher, A. B. and P. R. Voss (2016): "How to interpret the coefficients of spatial models: Spillovers, direct and indirect effects," *Spatial Demography*, 4, 175–205.

Halleck Vega, S. and J. P. Elhorst (2015): "The SLX model," *Journal of Regional Science*, 55, 339–363.

Haug, A. A. and C. Smith (2012): "Local linear impulse responses for a small open economy," *Oxford Bulletin of Economics and Statistics*, 74, 470–492.

Holden, R. and D. Thornton (2020): "The Stochastic Reproduction Rate of a Virus," *Covid Economics*, 41.

Holly, S., M. H. Pesaran, and T. Yamagata (2011): "The spatial and temporal diffusion of house prices in the UK," *Journal of Urban Economics*, 69, 2–23.

Hornstein, A. (2020): "Social distancing, quarantine, contact tracing and testing: Implications of an augmented SEIR model," *Covid Economics*, 18.

Jinjarak, Y., R. Ahmed, S. Nair-Desai, W. Xin, and J. Aizenman (2020): "Accounting for Global COVID-19 Diffusion Patterns, January-April 2020," Working Paper 27185, National Bureau of Economic Research.

Jordà, Ò. (2005): "Estimation and inference of impulse responses by local projections," *American Economic Review*, 95, 161–182.

Covid Economics 43, 21 August 2020: 42-94

Jordà, Ò., M. Schularick, and A. M. Taylor (2016): "Sovereigns versus banks: credit, crises, and consequences," *Journal of the European Economic Association*, 14, 45–79.

——— (2020): "The effects of quasi-random monetary experiments," *Journal of Monetary Economics*, 112, 22–40.

Kermack, W. O. and A. G. McKendrick (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, 700–721.

Kuchler, T., D. Russel, and J. Stroebel (2020): "The Geographic Spread of COVID-19 Correlates with the Structure of Social Networks as Measured by Facebook," Working Paper 26990, National Bureau of Economic Research.

Kuethe, T. H. and V. O. Pede (2011): "Regional housing price cycles: a spatio-temporal analysis using US state-level data," *Regional studies*, 45, 563–574.

LeSage, J. and R. Pace (2009): *Introduction to Spatial Econometrics*, Boca Raton: Taylor and Francis.

Lucas, R. J. (1976): "Econometric policy evaluation: A critique," *Carnegie-Rochester Conference Series on Public Policy*, 1, 19–46.

McAdams, D. (2020): "Nash SIR: An economic-epidemiological model of strategic behaviour during a viral epidemic," *Covid Economics*, 16.

Michaud, A. and J. Rothert (2018): "Redistributive fiscal policies and business cycles in emerging economies," *Journal of International Economics*, 112, 123 – 133.

Oates, W. E. (1999): "An Essay on Fiscal Federalism," *Journal of Economic Literature*, 37, 1120–1149.

Covid Economics 43, 21 August 2020: 42-94

Pace, R. K., R. Barry, J. M. Clapp, and M. Rodriquez (1998): "Spatiotemporal autoregressive models of neighborhood effects," *The Journal of Real Estate Finance and Economics*, 17, 15–33.

Painter, M. and T. Qiu (2020): "Political Beliefs affect Compliance with COVID-19 Social Distancing Orders," *Covid Economics*, 4.

Piguillem, F. and L. Shi (2020): "Optimal COVID-19 quarantine and testing policies," *Covid Economics*, 27.

Plagborg-Møller, M. and C. K. Wolf (2019): "Local projections and VARs estimate the same impulse responses," *mimeo*.

Pollakowski, H. O. and T. S. Ray (1997): "Housing price diffusion patterns at different aggregation levels: an examination of housing market efficiency," *Journal of Housing Research*, 107–124.

Pragyan Deb, Davide Furceri, J. D. O. and N. Tawk (2020): "The effect of containment measures on the COVID-19 pandemic," *Covid Economics*, 19.

Rana, G. A. and P. Shea (2015): "Estimating the causal relationship between foreclosures and unemployment during the great recession," *Economics Letters*, 134, 90–93.

Rothert, J. (2020): "Optimal federal redistribution during the uncoordinated response to a pandemic," Departmental working papers, United States Naval Academy Department of Economics.

Simonov, A., S. K. Sacher, J.-P. H. Dubé, and S. Biswas (2020): "The Persuasive Effect of Fox News: Non-Compliance with Social Distancing During the Covid-19 Pandemic," Working Paper 27237, National Bureau of Economic Research.

Verwimp, P. (2020): "The Spread of COVID-19 in Belgium: a Municipality-Level Analysis," Working Papers ECARES 2020-25, ULB – Universite Libre de Bruxelles.

Covid Economics 43, 21 August 2020: 42-94

Weber, E. (2020): "Which measures flattened the curve in Germany?" *Covid Economics*, 24.

Covid Economics 43, 21 August 2020: 42-94

Covid Economics 43, 21 August 2020: 42-94

A States' summary statistics and calibrated fixed effects

State	β_n	ξ_n	Density	Cases per 1 mln	max(r-score)	GOP vote
AL	0.29	0.66	227.8	7250.8	5	0.63
AR	0.09	0.70	169.4	6339.5	3	0.61
AZ	0.10	0.44	319.3	10313.0	3	0.50
CA	0.21	0.47	1326.5	5484.6	5	0.34
CO	0.29	0.68	909.7	5739.1	5	0.44
CT	0.23	0.76	908.9	12890.1	4	0.41
DC	0.06	0.15	10275.8	14588.8	5	0.04
DE	0.09	0.65	742.1	11576.0	5	0.42
FL	0.54	0.80	795.9	6616.4	4	0.48
GA	0.41	0.82	1021.6	6694.3	4	0.51
IA	0.16	0.59	231.8	9034.4	4	0.52
ID	0.11	0.78	196.0	3049.8	5	0.60
IL	0.26	0.60	1681.7	11150.1	5	0.38
IN	0.23	0.63	641.7	6884.6	5	0.57
KS	0.29	0.99	456.7	4814.0	3	0.57
KY	0.15	0.75	561.7	3464.3	4	0.63
LA	0.31	0.70	382.4	12065.5	5	0.58
MA	0.26	0.72	1735.3	15702.4	4	0.33
MD	0.48	0.75	1698.8	11147.3	5	0.34
ME	0.28	0.61	121.8	2380.4	5	0.45
MI	0.36	0.72	940.1	6991.2	5	0.47
MN	0.28	0.68	966.2	6320.8	4	0.45
MO	0.16	0.75	812.6	3450.8	4	0.57

Continued on next page

Covid Economics 43, 21 August 2020: 42-94

State	β_n	ξ_n	Density	Cases per 1 mln	max(r-score)	GOP vote
MS	0.14	0.49	124.7	8671.9	5	0.59
MT	0.07	0.76	28.3	812.4	5	0.56
NC	0.35	0.62	621.6	5988.1	5	0.51
ND	0.11	0.74	43.6	4598.2	2	0.64
NE	0.27	0.94	639.1	9796.5	3	0.59
NH	0.12	0.59	301.1	4235.3	5	0.47
NJ	0.40	0.75	2978.5	19350.6	5	0.40
NM	0.22	0.75	211.6	5635.6	4	0.42
NV	0.15	0.66	222.0	5664.1	5	0.45
NY	0.41	0.77	11497.0	20330.0	5	0.35
OH	0.29	0.65	825.9	4303.8	5	0.52
OK	0.14	0.70	476.9	3282.7	4	0.65
OR	0.10	0.80	531.6	1992.0	4	0.41
PA	0.19	0.66	2050.2	7010.6	5	0.49
RI	0.19	0.69	1110.4	14134.9	4	0.40
SC	0.27	0.62	292.8	6553.7	4	0.55
SD	0.13	0.99	76.0	7697.7	2	0.61
TN	0.23	0.61	527.9	5900.3	5	0.61
TX	0.19	0.53	1245.4	5306.1	4	0.51
UT	0.22	0.84	680.8	6692.3	3	0.46
VA	0.44	0.76	1545.9	7560.2	5	0.45
VT	0.02	1.00	111.8	1909.6	5	0.33
WA	0.25	0.62	499.7	4424.5	5	0.39
WI	0.12	0.50	322.2	4791.7	5	0.48

Continued on next page

CEPR PRESS

State	β_n	ξ_n	Density	Cases per 1 mln	max(r-score)	GOP vote
WV	0.11	0.92	160.8	1568.3	5	0.69
WY	0.04	0.67	12.4	2452.7	3	0.70

Covid Economics 43, 21 August 2020: 42-94

Covid Economics 43, 21 August 2020: 42-94

B SIR model

This appendix shows the model fit for individual 49 contiguous regions (48 states + DC) from the benchmark calibration of the model described in Section 4.

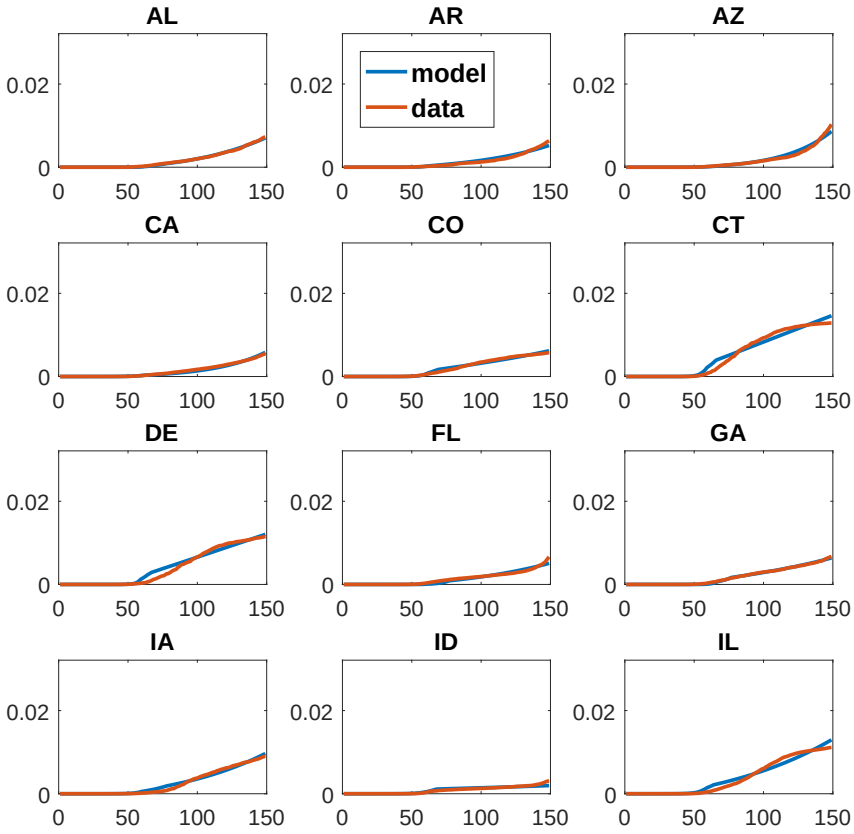


Figure 15: Model (blue) vs. data (red) - individual states

Covid Economics 43, 21 August 2020: 42-94

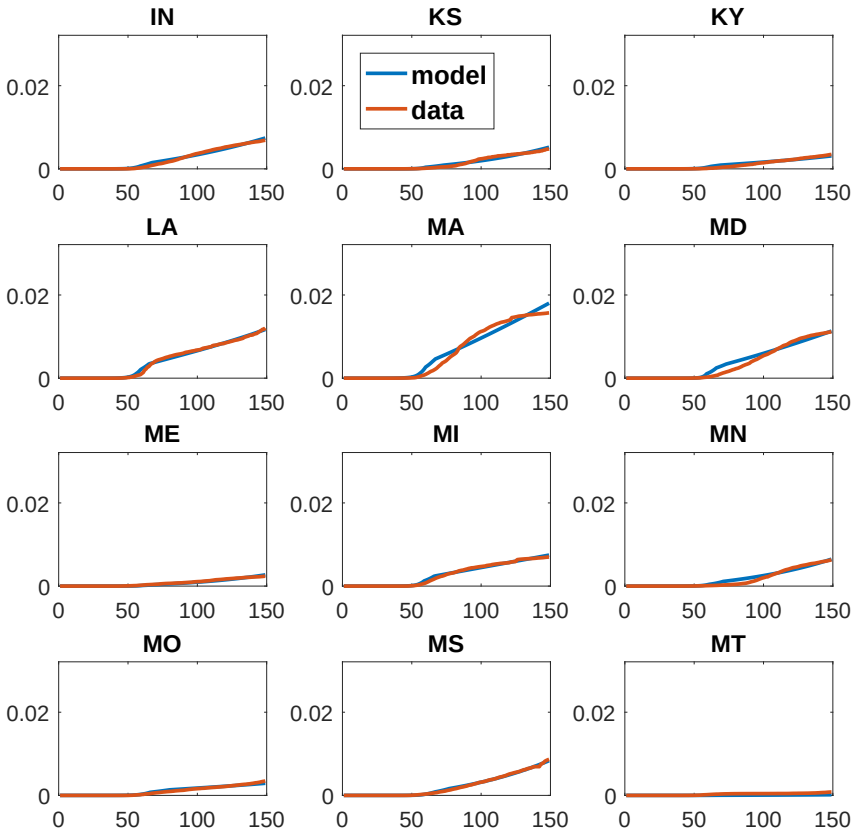


Figure 16: Model (blue) vs. data (red) - individual states

Covid Economics 43, 21 August 2020: 42-94

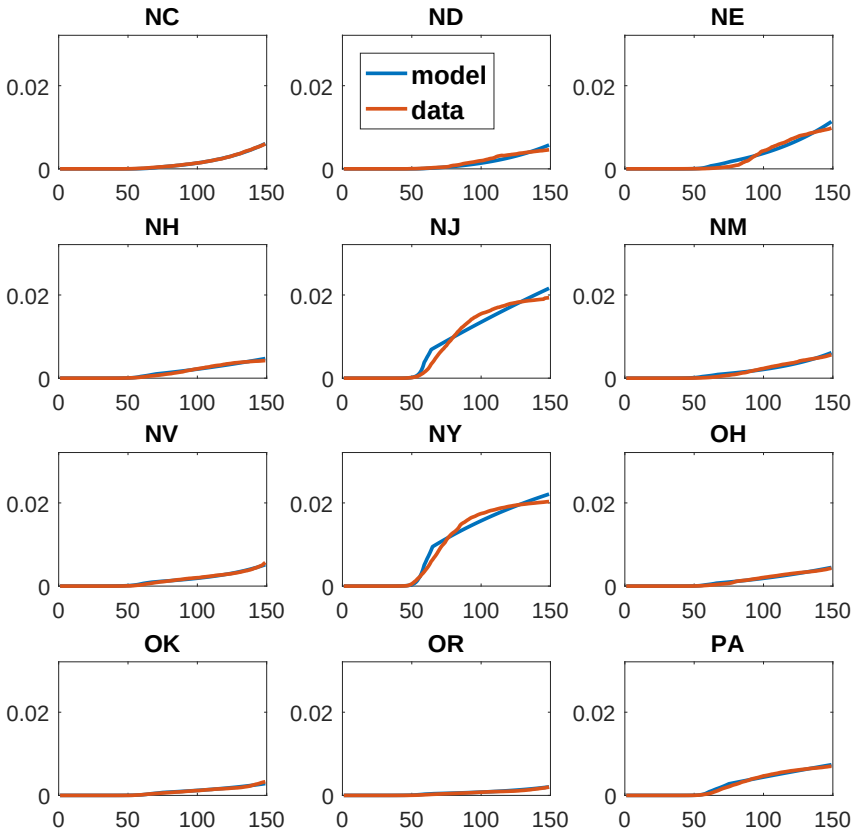


Figure 17: Model (blue) vs. data (red) - individual states

Covid Economics 43, 21 August 2020: 42-94

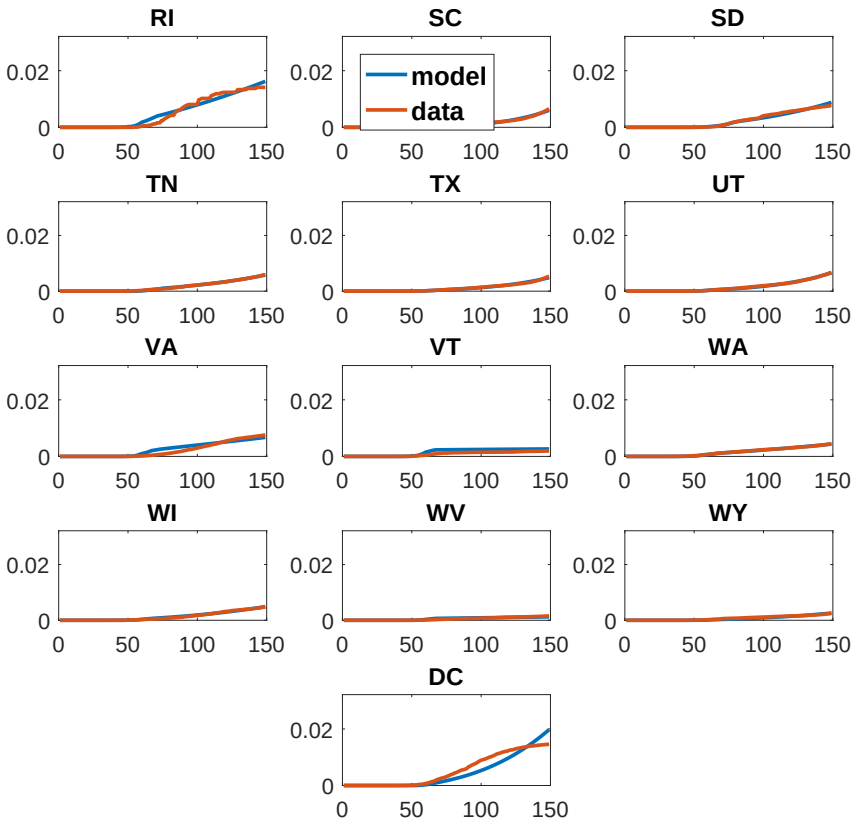


Figure 18: Model (blue) vs. data (red) - individual states

Covid Economics Issue 43, 21 August 2020

Reacting quickly and protecting jobs: The short-term impacts of the COVID-19 lockdown on the Greek labor market[1]

Gordon Betcherman,[2] Nicholas Giannakopoulos,[3] Ioannis Laliotis,[4] Ioanna Pantelaiou,[5] Mauro Testaverde[6] and Giannis Tzimas[7]

Date submitted: 7 August 2020; Date accepted: 18 August 2020

This paper uses administrative, survey, and online vacancy data to analyze the short-term labor market impacts of the COVID-19 lockdown in Greece. We find that flows into unemployment have not increased; in fact, separations were lower than would have been expected given trends in recent years. At the same time, employment was about 12 percent lower at the end of June than it would have been without the pandemic. The interrupted time series and difference-in-differences estimates indicate that this was due to a dramatic slowdown in hiring during months when job creation typically peaks in normal years, mostly in tourism. While we do not formally test the reasons for these patterns, our analysis suggests that the measures introduced to mitigate the effects of the crisis in Greece have played an important role. These measures prohibited layoffs in industries affected by the crisis and tied the major form of income support to the maintenance of employment relationships.

Covid Economics 43, 21 August 2020: 95-136

1 The authors would like to thank Athina Raftopoulou for insights and feedback while developing this paper. The authors would also like to acknowledge the contribution of Spyros Ntouroukis in collecting and organizing online job posting data.
2 Professor Emeritus, School of International Development and Global Studies, University of Ottawa and IZA.
3 Associate Professor, Department of Economics, University of Patras.
4 Lecturer, Department of Economics, City, University London.
5 PhD candidate, Department of International and European Economic Studies, Athens University of Economics and Business.
6 Senior Economist, World Bank.
7 Associate Professor, Department of Electrical and Computer Engineering, University of Peloponnese.

Reacting quickly and protecting jobs: The short-term impacts of the COVID-19 lockdown on the Greek labor market

1. Introduction

The COVID-19 pandemic has had dramatic consequences for economies and labor markets around the world. The pandemic has been unique in triggering both supply and demand shocks. To contain the spread of the virus, governments introduced various confinement and lockdown measures that shut down businesses and removed workers from their jobs. At the same time, there have been substantial declines in demand despite large government transfers to firms, workers, and households in many countries. According to the World Bank's June forecast, global GDP is expected to fall by 5% in 2020, with the decrease projected at 7% in the advanced economies (World Bank, 2020). The ILO estimates that the decrease in labor demand in the second quarter of 2020 compared to the last quarter of 2019 will be equivalent to the loss of about 400 million jobs worldwide (ILO, 2020c). The OECD expects that overall unemployment in member countries will be in double digits for the remainder of 2020, more than twice the 5.3% unemployment rate in the first quarter of the year (OECD, 2020).

While a dramatic reduction in labor demand has been a consequence of the pandemic everywhere, there has been considerable variation in how different labor markets have adjusted to the shock. Take the examples of two G-7 countries that have been hard-hit by the virus. In the United States, the unemployment rate more than tripled between February and May, with 14 million workers added to the unemployment rolls in this period. Italy, on the other hand, experienced a decline in unemployment, with almost 400,000 fewer unemployed workers in May compared to February. There, the adjustment to declining labor demand has occurred through labor force withdrawal. There are also important differences emerging between countries in terms of the types of jobs and workers most affected by the crisis.

In this paper, we analyze how the labor market in Greece has been affected during the early months of the pandemic and lockdown. Greece is a particularly interesting case, both because of its situation prior to the arrival of COVID-19 and because of how the pandemic has been handled. Greece entered the current crisis just as it seemed to be finally emerging from the protracted and deep recession the country had endured since the financial crisis began more than a decade earlier. After declining every year from 2008 to 2013, Greece's real GDP has experienced modest growth since 2014, with an annual increase of about 2% in 2018 and 2019. At the same time, labor market conditions, while still difficult, were improving. After peaking at over 27% in 2013, the unemployment rate had slowly but consistently declined to 14.3% in March of this year, the lowest monthly rate in a decade.

To this point, Greece has been relatively successful in holding the pandemic in check, although there has been an increase in cases during the summer. While many neighboring countries suffered rapid escalations in cases and deaths, and severe burdens on health care systems, for the most part Greece has been able to avoid difficulties of such magnitude thus far. As of mid-August, Greece had around 8,000

Covid Economics 43, 21 August 2020: 95-136

confirmed cases which, on a per capita basis, is roughly one-tenth of the rate in Spain and one-fifth of Italy's rate. Fewer than 250 deaths had been attributed to the virus, about one-twenty-fifth of the per capital rate in Italy and Spain. These numbers reflect a swift response of the Government after the first case was confirmed on February 26. After that, and within a relatively narrow time window, closures shut down public events, schools, workplaces, travel, and public gatherings, and finally a general stay-at-home order was implemented on March 23 (eventually lifted on May 4). The Government was also very active in introducing a series of measures to help employers and workers weather the economic consequences of the lockdown. These included various forms of tax and rent relief for businesses, unemployment benefit extensions, financial support and social insurance coverage for employees whose contracts have been suspended, financial support for the self-employed, and prohibitions on dismissals for businesses shut down by state order.

Nonetheless, as elsewhere, the pandemic and lockdown are having major economic impacts in Greece. The Hellenic Statistical Authority (ELSTAT) reports that on a seasonally adjusted basis, GDP declined by 1.6% in the first quarter compared to the fourth quarter of 2019.[1] In its Summer Forecast, the European Commission predicts that economic activity will decrease by 9% in 2020.[2] In this situation, a significant deterioration in labor market conditions would be expected.

The main empirical contribution we make in this paper is to describe in some detail how the Greek labor market has evolved in the first few months of the pandemic. Relying on a range of sources including administrative data, survey data, and data from online job posting sites, we document the drop in employment following the imposition of the lockdown and the subsequent flat employment trend through the first months that followed. Two things are particularly noteworthy about the patterns we have observed. First, while Greece did not experience the major declines in employment that some other OECD countries did in March, April, and May, these are normally months when employment growth is substantial in the heavily seasonal Greek economy. Second, employment in these months in 2020 differed from the story in previous years not because the lockdown fueled large numbers of separations but rather it choked off new hiring in what should have been expansionary months. In fact, compared to recent years, separations were comparable or even lower. We attribute this largely to the regulations and wage subsidies introduced by the Government that were designed to minimize job loss.

Indeed, one of the messages from our analysis is that policy choices help to explain how the labor market has responded differently to the pandemic shock in different OECD countries. A particularly relevant distinction is between countries, like Greece, that have primarily linked their support to the maintenance of the employment relationship through dismissal restrictions, wage subsidies, and short-term compensation and those countries that have largely let layoffs occur and supported workers through unemployment benefits and cash transfers. Highlighting this link between policy choices and labor market outcomes is a second contribution of this paper.

Covid Economics 43, 21 August 2020: 95-136

The remainder of this paper is organized into the following sections. In section 2, we review the early evidence on how the COVID-19 pandemic has affected labor markets in OECD countries. We note how a roughly similar shock has translated into different outcomes across countries and highlight the role of government mitigation policies in shaping those outcomes. Section 3 turns to the pandemic in Greece. It describes the measures introduced by the Government to control the spread and to compensate firms and workers, and it uses mobility data to document the evolution of the lockdown. In section 4, we describe the different data sources and methods used for our labor market analysis. That analysis is presented in section 5. It tracks employment and unemployment trends during the lockdown, with an emphasis on how the overall picture has been shaped by the dynamics of hirings and separations. Finally, in section 6, conclusions are presented as well as some key research questions moving forward to understand the impact of the pandemic on the Greek labor market.

Covid Economics 43, 21 August 2020: 95-136

2. Literature Review

2.1 Challenges related to data and measurement

The COVID-19 pandemic has affected entire economies and labor markets at a very fast pace, requiring policy makers to have access to up-to-date information to design suitable and timely policy responses. However, the disruptions caused by the pandemic and the speed with which the crisis has unfolded have shown that the methods and sources normally used to track labor market outcomes may have significant shortcomings in this context. For example, even gold standard surveys designed to provide up-to-date information, such as the Current Population Survey (CPS) in the United States, may not be sufficient to keep up with the fast spread of the virus and its disruptive impacts. Administrative data released with shorter time lags, as in the case of UI claims in the United States,[3], have also shown to have shortcomings that limit their effectiveness in timely and comprehensively informing policy responses during the pandemic (Cajner et al, 2020).

Furthermore, social distancing and other transmission prevention measures have affected several data-related activities around the world with potential impacts on data quality and reliability. ILO (2020b) shows that data collection, supervision, cleaning and analysis have been affected in several countries. Adjustments in survey instruments, data collection methods and weighting schemes have become necessary to address issues related to low response rates and non-random patterns in non-responses. Even with these adjustments, response rates have dropped in several cases. The United States Bureau of Labor Statistics documents that the response rate for the 2020 May Establishment Survey was 69%, compared to a 75% average between March 2019 and February 2020. The corresponding figure for the household survey in May 2020 was 67%, compared to the 82% average over the 12 months ending in February 2020 (BLS, 2020). Furthermore, focusing on the nature on non-responses in the March and April 2020 rounds of the CPS, Montenovo et al. (2020) show that the drops in responses in these two months were not random.

Covid Economics 43, 21 August 2020: 95-136

The specific disruptions emerging from the crisis have also implied that standard labor market definitions may not be sufficient to fully capture labor market dynamics under the pandemic. For example, given the different forms of mobility restrictions and social distancing measures currently in place, variations in unemployment may be misleading. In fact, in the COVID-19 era, slow increases in unemployment may co-exist with significant job losses. This is because non-employed people, despite being interested in working, might not be actively looking for a job as a result of restrictions on economic activities or the perceived risk of contracting the disease at work. As such, going beyond the analysis of employment, unemployment and labor force participation trends becomes important to fully understand labor market dynamics during the crisis (ILO, 2020b; Abraham, 2020; Hamermesh, 2020).

Several efforts have been made over the last months to address these challenges. Some researchers have complemented administrative and survey data with online vacancy data that record information in real-time and are available with short time lags (Kahn et al, 2020; Campello et al., 2020; Hensvick et al.; 2020). Kong and Prinz (2020) and Goldsmith-Pinkham and Sojourner (2020) use Google search data to predict UI claims in the United States with the objective to reduce the time lag with which this information becomes available. Other authors have leveraged on private sector data, specifically payroll data (Cajner et al., 2020), data from a time and scheduling software (Kurman et al., 2020) or data from daily purchases (Coibion et al., 2020). By combining these data with information from traditional data sources or augmenting these data with newly collected COVID-related information, they have been able to provide timely and detailed insights on the labor market impacts of the crisis.

Several other researchers have used newly collected data based on surveys specifically implemented to better understand the impacts of the crisis. In the United States, Bick and Blandin (2020) fielded a survey that follows a similar structure to the CPS but that generates more timely estimates. Brynjolfsson et al. (2020) used Google Consumer Surveys to collect two waves of survey data in April and May 2020. Bartik et al. (2020) focused on firms and collected data from approximately 5,800 businesses using an online survey. Online surveys were also used in the UK (Gardiner and Slaughter, 2020) and in Belgium (Baert at al. 2020) to collect data from workers, and in Denmark (Bennedsen et al., 2020) to collect data from firms. Finally, some researchers implemented multi-country online surveys. Adams-Prassl et al. (2020) covered Germany, the United States and the UK, while Belot et al. focused on China, Japan, Korea, the United States, UK, and Italy.

2.2 Evidence on employment impacts and the role of policies

Combining data from surveys, administrative and real-time sources, the ILO estimates large drops in employment due to the pandemic. To address the data and methodological challenges emerging from the crisis, the ILO developed what they refer to as a "nowcasting" model, which provides real-time statistical prediction based on a multiplicity of traditional and non-traditional data sources (ILOa, 2020). Based on this model, the ILO estimates that between April and June 2020, Europe alone experienced a decline in hours worked equivalent to 37 million full-time jobs compared to the last quarter of 2019. Projections for the second half of 2020 show that

even in the most optimistic scenario, hours worked would still be far from pre-COVID levels. The OECD projects unemployment in OECD countries to be at 11.5% in mid-2020, twice the level at the end of 2019. The projections for the rest of the year still show unemployment rates well above the pre-outbreak levels, with the most optimistic scenario suggesting levels comparable those recorded during the peak of the Global Financial Crisis (OECD, 2020).

These significant drops in hours worked are the result of different labor market adjustments in various countries. ILO (2020c) shows that working hour losses were not due to significant job losses in the UK and in Korea, as the vast majority of workers were still able to keep their jobs even if working fewer or no hours. As a result of this, unemployment was not greatly affected in these countries. The implications of reductions in hours worked were significantly different in Canada and the US. In Canada, almost half of the reduction in hours worked was due to people losing their jobs. In the United States, two-thirds of the decline in hours worked was due to people losing their jobs. Among those who lost their jobs, relatively more people became inactive in Canada, while the majority became unemployed in the United States. Furthermore, these disruptions did not equally affect all workers and segments of the economy. Estimates suggest that women, migrants, young people, informal workers, and specific vulnerable sectors and occupations were particularly hit by the crisis (ILOa, 2020; OECD, 2020).

The policies introduced to attenuate the disruptions caused by the pandemic have likely played a role in the way labor markets have responded in different countries. Gentilini et al. (2020) show that since the beginning of the outbreak, 200 countries implemented more than 1,000 social protection and employment measures to address the impacts of the crisis. While most of these interventions were cash transfer programs, several countries also introduced policies specifically focused on attenuating the labor market impacts of the crisis: 64 countries provided unemployment benefits, 53 social security subsidies, 69 wage subsidies, 24 labor market regulation adjustments and 10 shorter work time benefits.

Some clear patterns have emerged after these initial months of policies' implementation. A first group of countries has focused on policies and programs aimed at preserving existing employment relationships, often implemented through the provision of subsidies to reduce labor and other costs for employers and/or the introduction of measures to limit dismissals. New Zealand, Germany, Denmark, France, and Switzerland are all countries in which take-up rates in job retention schemes have been high. A second group of countries has focused on mitigating the impacts of the crisis on workers by expanding unemployment insurance systems. In the United States, Israel, Norway, Canada and Ireland, the unemployment insurance system has played an important role in response to the crisis. While projections, administrative data and surveys suggest that increases in unemployment have been minimal for the first group of countries, unemployment has increased significantly for the second group (Rothwell, 2020; OECD, 2020).

Covid Economics 43, 21 August 2020: 95-136

Covid Economics 43, 21 August 2020: 95-136

A large body of research in the last months has focused on better understanding the impacts of COVID on labor markets in specific countries, with particular attention to identifying groups severely affected by the crisis and jobs at risk. An increasing number of papers has also focused on the impacts of the pandemic on small firms, on the role of policies in shaping labor market dynamics, and on potential shock-induced changes in labor market behaviors such as job search.

Studies focused on the United States suggest that significant job losses were recorded in March and April, with some initial signs of recovery in May, which however seem to have slowed down by the end of June. Evidence from both standard surveys (Béland et al., 2020a; Cowan, 2020) and private sector data (Coibion et al., 2020; Cajner et al. 2020) points to unprecedented drops in employment, increases in unemployment and declines in labor force participation. Between February and April 2020, it was estimated that at least 20 million people lost their jobs (Coibion et al., 2020; Cajner et al. 2020). These patterns have been accompanied by significant declines in job vacancies posted by firms (Kahn et al., 2020; Campello et al., 2020). Estimates based on real-time population surveys document strong increases in employment and declines in unemployment during May and most of June. However, the current figures are still well below their pre-COVID levels (Bick and Blanding, 2020).

These findings are confirmed by studies focused on small businesses. Kurman et al. (2020) find that until mid-April employment in small businesses in the services sector dropped by 60%, equivalent to the loss of 18.2 million jobs. However, from mid-April to June more than half of the closed businesses reopened, resulting in 9.1 million additional jobs, mainly taken by previously furloughed workers. Bartik et. al (2020) find similar patterns and point to significant heterogeneity across sectors, with retail, arts and entertainment, personal services, food services, and hospitality reporting the largest declines. Using CPS data, Fairlie (2020) confirms that economic activities by small businesses significantly declined in April and only partially recovered in May.

Overall, the research so far seems to unanimously show that the workers hit the hardest by the crisis are women, young, low-educated (Béland et al., 2020a; Bick and Blanding, 2020; Cho and Winters, 2020; Cowan, 2020; Montenovo et al., 2020) or with an ethnic minority or migration background (Béland et al., 2020a; Borjas and Cassidy, 2020; Cho and Winters, 2020; Cowan, 2020; Fairlie et al., 2020; Montenovo et al., 2020). Nevertheless, some evidence suggests that during the initial stages of the crisis men might have been disproportionately affected (Béland et al., 2020a) and that some older workers might have chosen to go on early retirement (Coibion et al., 2020; Cowan, 2020).

These studies also explore whether social distancing measures have disproportionally impacted specific categories of workers. Findings from this research show that jobs that cannot be performed from home are at higher risk, while, jobs in workplaces classified as essential face lower risks (Béland et al., 2020a; Cajner et al., 2020; Montenovo et al., 2020). As an estimated 93% of workers around the world live in countries with some forms of workplace restrictions (ILOa, 2020), a large body of research across the world has focused on identifying vulnerable occupations, with a

particular focus on jobs that cannot be performed from home and in sectors that have been severely affected by the shutdown (Diengel and Neiman, 2020; Garrote-Sanchez et al., 2020; Hatayama et al., 2020; Hicks et al., 2020; Mongey et al. 2020; Pouliakas and Branka, 2020; Saltiel, 2020). ILO (2020a) shows that while only 7.9% of workers around the world worked from home before the crisis, almost 18% are in jobs or have access to the infrastructure that could allow them to work from home in the future. This research also shows that working from home is more feasible in high-income countries (23%) than in low-income countries (13%). Focusing on Greece, Pouliakas (2020) shows that more than one-third of jobs in the Greek labor market could be performed from home.

A number of studies have tried to identify the channels driving the observed employment impacts. Using data from the United States' Current Employment Statistics (CES), Brinca et al. (2020) try to disentangle the aggregate COVID shock in its demand and supply components. They observe that in April 2020 employment in the private sector was significantly lower than its historical average and estimated that more than 65% of this impact was due to labor supply shocks, i.e. inability of workers to perform their jobs. Kong and Prinz (2020) conclude that restaurant and bar limitations and non-essential business closures were the only transmission prevention measures that in the United States led to an increase in UI claims. Barrero et al. (2020) using forward-looking firm level data, find that the COVID-19 induced shocks lead to 3 new hires for every 10 layoffs. They also project that the total number of entire working days performed from home will triple after the end of the pandemic and that between 32% and 42% of all layoffs will be permanent.

Findings of research focused on the Canadian labor market are in line with the results in the United States. Using the Canadian Labor Force Survey up to April 2020, Béland et al. (2020b) document substantial increases in unemployment (approximately 5 percentage points) and drops in labor force participation (3.7 percentage points), hours worked (1.5 percentage points) and wages (0.4 percentage points). These impacts were less severe for essential workers or workers who can work remotely, while they were more pronounced for younger and less educated workers. Differently from the results for the United States, they do not find evidence of differential effects by gender or of disproportionate impacts of the crisis on labor market outcomes for migrants.

Adams-Prassl et al. (2020) compare the impacts of COVID-19 on jobs, earnings and hour worked in three countries that have introduced different policies in response to the pandemic, i.e. the UK, United States and Germany. They find substantial differences across and within countries. Job losses in the United States and the UK were substantially higher (18% and 15%) than in Germany (5%), a country with a well-established short-time work scheme. They point out that the UK also introduced a similar scheme, which however does not allow furloughed workers to do any work for their employers, thereby potentially discouraging firms from applying. Not surprisingly, the study also finds that furloughing was more prevalent in the UK (43%) than in the US (31%), a country that strongly relied on the expansion of unemployment benefits to respond to the crisis. The study also finds that in all countries, people who

Covid Economics 43, 21 August 2020: 95-136

Covid Economics 43, 21 August 2020: 95-136

can work from home are less likely to lose their jobs. This is also the case for people with permanent contracts, fixed hours and in salaried jobs. In the US and the UK, less educated workers and women were found to be more likely to lose their jobs during the pandemic. This is not the case in Germany. Based on a survey covering China, Japan, Korea, US, UK, and Italy, Belot et al (2020) also find that young people are severely affected by the crisis in these countries.

Focusing on Denmark, another country that introduced significant measures to encourage job retention, Bennedsen et al. (2020) provide additional evidence on the strong impact of these policies in helping firms keep their workers. Estimates presented in this study suggest that the policies introduced by the Danish Government contributed to a reduction in layoffs by 81,000 jobs and increase in furloughs by 285,000. Employment subsidies seem to have a stronger correlation with job retention, while the correlation is weaker for cost subsidies and the evidence for tax subsidies is mixed. The authors conclude that labor subsidies meet their objective of preserving employer-employee relationships, while the impact of the other policies is less clear.

Sweden is another example of a country that has leveraged on strong job retention interventions to respond to the crisis. Hensvick et al. (2020) study the impact of the COVID-19 crisis on job search using real-time data from the job board of the Swedish Public Employment Service. They find that between March and May employers posted 40% fewer vacancies. The drops were significant in sectors such as hotels and restaurants, and entertainment, as well as in occupations that are more difficult to perform from home. They also find that users reduced job-search intensity and seemed to have re-directed their searches to occupations that are more likely to be performed from home and more resilient to the crisis.

Alstadsæter et al. (2020) study the impacts of COVID-19 on layoffs in Norway, a country that has strongly relied on unemployment insurance benefits to mitigate the impacts of the crisis. Their analysis based on UI claims data shows that almost all layoffs up to April 19 were temporary. Even if accounting for only 10% of the total number of layoffs, permanent layoffs generated a 1.5 percentage point increase in unemployment, a significant month-to-month variation for Norway. They also show that layoffs affect populations that are already financially vulnerable (low-income, low-educated, immigrants) and are more common in jobs that require physical proximity, especially in the initial phases of the crisis. Similarly, they find that in the early stages of the crisis, the impacts were mostly felt by women and young workers, but as time passed men and older workers were also significantly impacted.

The Korean experience also provides interesting insights as the Government mainly relied on testing and tracing and less on lockdowns to contain the spread of the virus. Aum et al. (2020) find that a one per thousand increase in infections leads to an almost 3% decrease in local employment. They compare these effects to those in the UK and US, where lockdowns were introduced, and note that employment losses were almost double in these countries. Shedding light on the channels driving these results, the authors find that employment losses were mainly due to a slowdown in hiring by firms

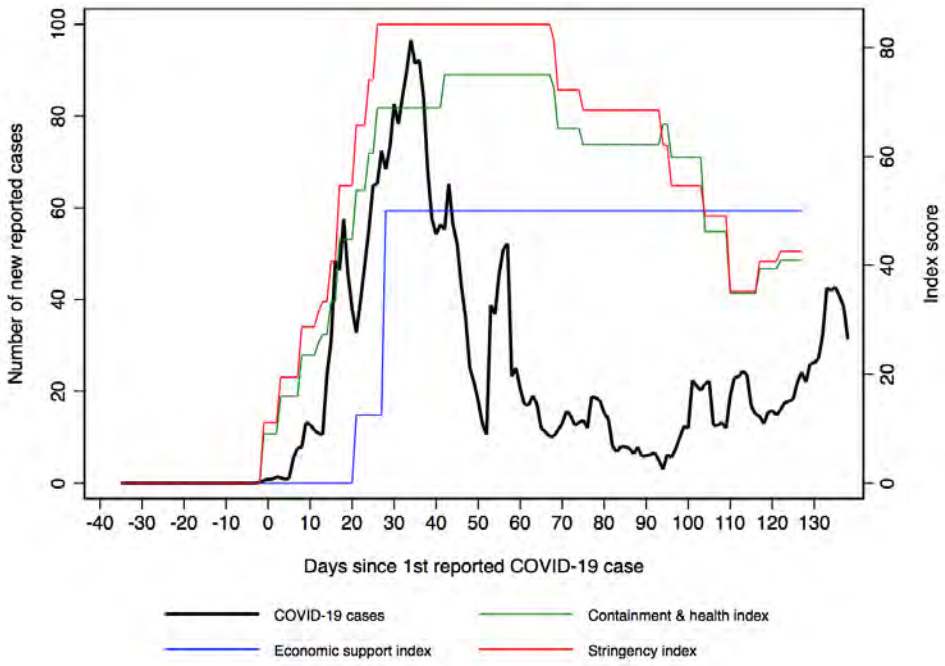
and to transition of workers out of the labor market rather than to unemployment. The authors note that at the time of publication, the Korean Government had not implemented any public furlough scheme. The paper also shows that employment losses were mainly experienced in small businesses (less than 30 employees), and that the workers with the highest probability to lose their jobs were less educated, young, employed in low-wage occupations, with temporary contracts, or self-employed. Men were also more affected than women. With the exception of the gender results, these results are similar to those found by other studies for the US and the UK.

3. COVID-19 in Greece: Evolution and measures taken

3.1 The spread of the virus and the lockdown measures

The first case of COVID-19 in Greece was confirmed on February 26. Figure 1 shows the trend in new cases from that date. Compared to many other countries in Europe, where cases and fatalities exploded quickly, the pandemic progressed slowly in Greece.

Figure 1. COVID-19 cases and public policy mitigation measures



Source: Johns Hopkins University; University of Oxford, Blavatnik School of Government.
Notes: Indices range from 0 to 100. The Containment & health index combines lockdown restrictions and closures with measures such as testing policy, contact tracing, short-term healthcare investment in healthcare, and investments in vaccine. The Economic support index records measures such income support and debt relief. The Stringency index records the strictness of lockdown-style policies that primarily restrict behavior and activities. The first confirmed COVID-19 case in Greece was reported on February 26, 2020.

Covid Economics 43, 21 August 2020: 95-136

The Government has been credited with reacting quickly to the pandemic, introducing various restrictions even when cases and fatalities were quite low. For example, on March 10, before most of Europe, schools and universities nationwide were ordered closed, when there were just 89 confirmed cases and no deaths. The first virus-related death in Greece was recorded on March 12. The outbreak peaked in early April when new cases were approaching 100 per day. By April 21, there were 2,401 confirmed cases; 150 new cases, all asymptomatic, were related to one refugee facility located in Northern Greece. As of early July, Greece had around 3,500 confirmed cases and slightly fewr than 200 deaths from the virus.

Even though the actual spread of the virus has been much lower than in most European countries, the impact on society and the economy has been substantial because of the strict lockdown measures. Figure 1 shows the rapid imposition of the lockdown, as shown by the steep rise in the Oxford/Blavatnik Stringency and Containment and Health Indices within the first three weeks after the initial recorded case.

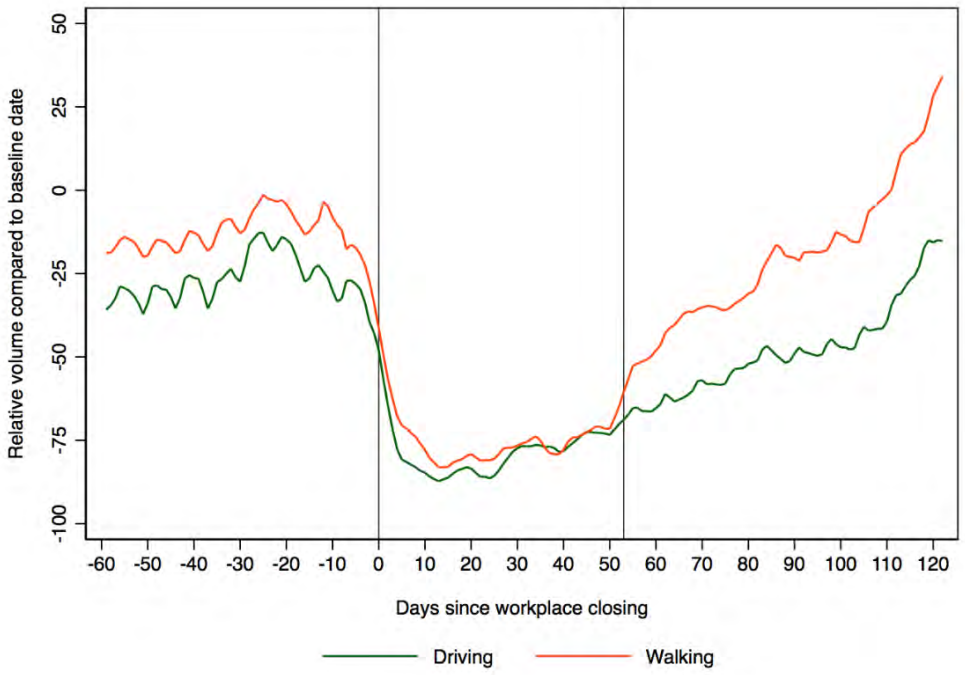
Some of the key measures put in place to slow the spread of the pandemic included cancellation of all carnival events (February 27), school closings (February 27 at a regional level and closed down nationally on March 10), closing of all non-essential workplaces (March 12-18), suspension of all public religious services (March 16), ban on gatherings of more than 10 people (March 19), internal and external travel restrictions (March 18-22), and finally, a general stay-at-home order (March 23), intensified by permanent roadblocks and checks of vehicles (April 8). The Government lifted the stay-at-home order on May 4, followed by the opening of schools, commercial activities, and workplaces (progressively from May 11, essentially completed by June 1).

These measures affected all aspects of everyday life. This effect can be visualized through mobility data provided by Apple and Google, which is sent from users' devices to these companies' maps services.[4] Using February 15 as the pre-pandemic baseline, Figure 2 presents Apple Maps data to illustrate how driving and walking in Greece started to decline after the first cases but then fell sharply as soon as the lockdown and workplace closing measures were implemented. For much of the period between March 12 (workplace closings initiated) and May 4 (end of the lockdown), driving and walking activity was well below 70% of February 15 levels. Towards the end of the lockdown period, mobility started to slowly pick up and this continued through May and June and approached pre-pandemic levels, at least in the case of walking.

Covid Economics 43, 21 August 2020: 95-136

Covid Economics 43, 21 August 2020: 95-136

Figure 2. Daily driving and walking activity

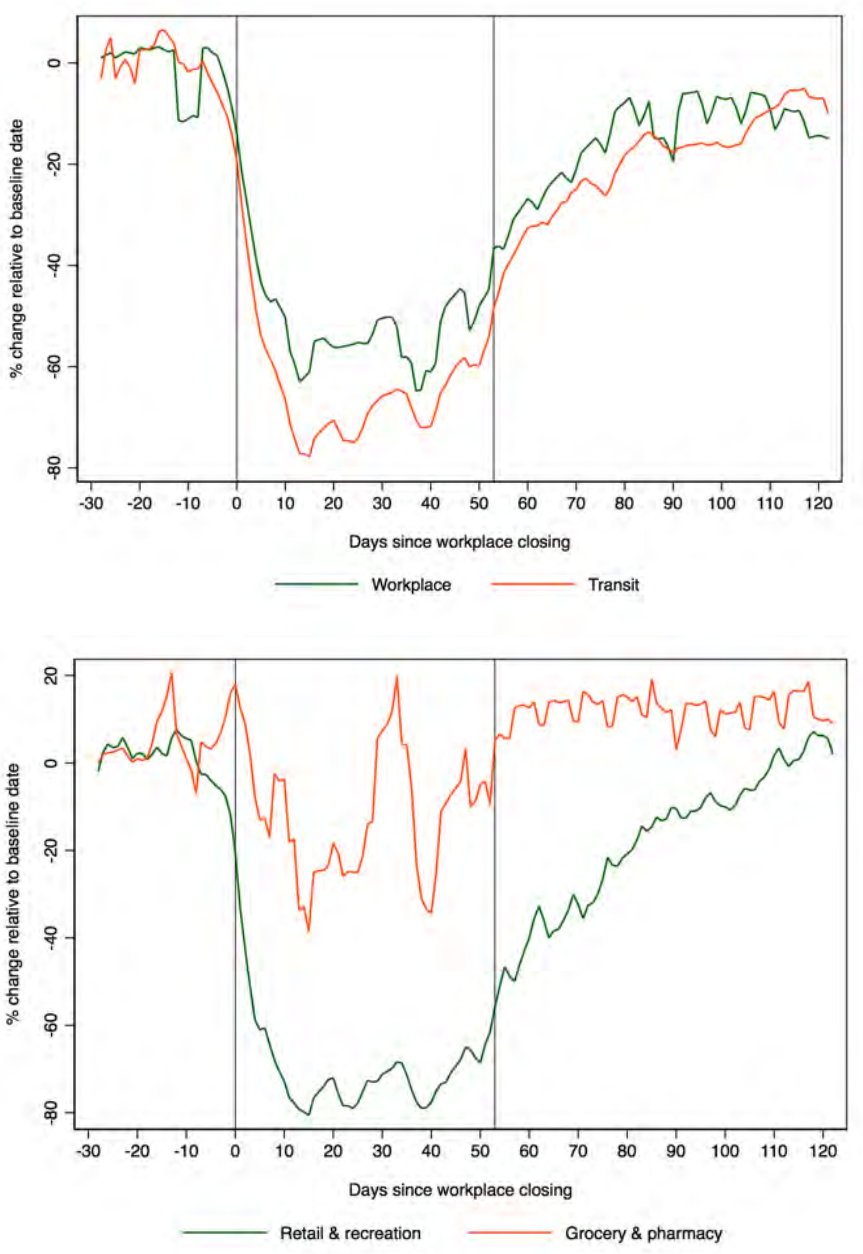


Source: Apple.
Notes: Vertical lines are set at workplace closing (March 12) and at the end of the lockdown (May 4). The baseline date is February 15.

The lockdown effects are also reflected in market-related activities. Google mobility reports provide data on trends in visits to various types of places, which can be compared to the pre-pandemic baseline (February 15). Figure 3 (top panel) shows that visits to workplaces and use of public transit declined by 50%-80% during the lockdown compared to pre-pre-pandemic levels. The data again show the return (albeit partial) towards the pre-pandemic baseline after the lockdown was lifted. A similar pattern is observed for non-essential shops (using retail and recreation as an example), while essential retail (grocery stores and pharmacies) were not nearly affected to the same extent (Figure 3, bottom panel).

Covid Economics 43, 21 August 2020: 95-136

Figure 3. Daily activity for selected indicators



Source: Google Community Mobility Reports.
Notes: Charts begin on 15 February 2020 (baseline date) after which Google data became available. Vertical lines are set at workplace closing (March 12) and at the end of the lockdown (May 4).

3.2 Economic impact and mitigation measures

As expected, restrictions imposed by the Government as well as the demand shock that affected most sectors resulted in a major economic slowdown with significant consequences for businesses and workers. GDP for the 1st quarter of 2020 decreased by 1.6% in comparison with the 4th quarter of 2019, while in comparison with the 1st quarter of 2019, the decline was 0.9%.[5] Various projections from national and international agencies estimate that GDP will shrink between 5.7% and 10% in 2020. A key factor will be declining exports, and especially tourism and shipping.[6]

To keep the economy afloat during the pandemic, the Greek government introduced a range of measures to support affected businesses and their employees. The government mobilized an immediate aid package amounting to 6.8 billion euros (or 3.5% of GDP) for March and April and legislated an additional package of 24 billion euros in May in order to stimulate the restart of the economy in the aftermath of the crisis.

The first legislative act to support businesses (March 11) was intended to provide firms with liquidity through the extension of tax and social contribution compliance deadlines, discounts on certified tax liabilities in case they were paid in due time, and suspension of debt payments. About 800,000 firms that had been financially affected in terms of a decline in their turnover or had ceased operation by state order were eligible for these and other benefits described below, on the condition of no layoffs.[7]

On March 18, the Ministries of Finance and Labor announced and then legislated a new package of measures. The key component was the provision of an 800-euro stipend (covering the period from March 15 to April 30, eventually extended through May) to workers whose contracts had been suspended because of the suspension of operations of their enterprise. In addition, for these workers, the Government covered all social insurance contributions and all tax payments were suspended for a period of four months. The same measure was applied to freelancers, self-employed, and individual business owners with up to 20 employees.[10] Enterprises whose operations had been mandatorily suspended and affected employees were asked to pay just 60% of their rents for March to May.

Overall, by early May, approximately 1.2 million employees and 550,000 self-employed and freelancers had benefited from this scheme. A one-off stipend of 600 euros in the form of a special training program was provided to specific professionals (economists/accountants, engineers, lawyers, doctors, teachers, and researchers) in April.[8] These occupations became eligible for the €800 financial support as of May. The budget allocation for the stipends for employees, freelancers/self-employed/individual businesses, and professionals is €2.36 billion, with an additional €1.36 billion for the social insurance payments.[9]

Additional funds were allocated to benefits for unemployed workers. On March 20, a measure was introduced to extend payments of the regular unemployment benefit, the long-term unemployment benefit, and the unemployment allowance for the self-employed by 2 months for those whose entitlement ended on March 31. The measure

Covid Economics 43, 21 August 2020: 95-136

was then extended to cover those whose entitlement ended at the end of April and at the end of May. In addition, a lump sum stipend of €400 was introduced for 155,000 long-term unemployed individuals, registered with the public employment agency (Hellenic Manpower Employment Organization (OAED)) from April 1, 2019 who were maintaining their status until April 16, 2020 and were not receiving any other benefit from the state. The budget allocation for these measures related to unemployment benefits is about €300 million.

In terms of numbers affected and financial commitment, the Government's mitigation measures have emphasized the preservation of employment in enterprises where operations were suspended. A key condition of the benefits provided to affected businesses was that they were obliged to maintain the same headcount. In fact, layoffs in designated industries were temporarily prohibited from March 18 until the restriction was lifted on June 16.

4. Data and methods

This section briefly presents the data sources, relevant indicators, and the methods that are used to assess labor market adjustments in Greece due to the COVID-19 pandemic and the relevant mitigation policies.

4.1. Labor market indicators

We report monthly estimates from the Labor Force Survey (from ELSTAT) on the labor force participation rate, employment to population ratio, and unemployment rate for the periods January-April 2019 and January-April 2020 to see differences in trends before and during the lockdown.[10]

We also use the LFS data to identify how other aspects of the Greek labor market have been affected by the COVID-19 crisis. In addition, to assess the degree of labor market slackness, we calculate the extended labor force indicator which is simply the active labor force plus the "potential additional labor force" (PALF), which takes into account persons seeking work but not immediately available and persons available for work but not seeking work.

4.2. Unemployment claims

As an indicator of the evolution of unemployment, we present data from OAED, the public employment agency, on the number of unemployment benefit recipients and the new claims for benefits covering the period from January 2017 to May 2020.

4.3. Labor market flows

Administered by the Ministry of Labor and Social Affairs (MoLSA), ERGANI is the national employment registry in Greece and covers all registered employers who contribute to the Social Security System. The ERGANI monthly reports provide daily information on labor market flows in the private sector, and we use the reports from January 2018 through June 2020. More specifically, the data we analyze covers new hires, overall and by type (full-time, part-time and shift work), and separations (lay-offs, quits and contract terminations). On a monthly frequency, labor market flows are

Covid Economics 43, 21 August 2020: 95-136

disaggregated by gender, age and region. In addition, we disaggregate these flows by occupation (2-digit) and sector of economic activity (2-digit) for the periods January-April 2019, and January-April 2020.[11]

Using ERGANI daily data, the change in daily flows since the onset of the pandemic and the government restriction on layoffs can be analyzed through a simple regression framework. More specifically, we adopt a single group interrupted time series analysis, in order to compare how outcomes change between the pre-pandemic period and two post-pandemic sub-periods, i.e., one since the onset of the pandemic and one since the government intervention to protect jobs:

$$y_t = a + b_1(c_t \times S_t) + b_2(c_t \times R_t) + W_t^d + c_t + u_t \quad (1)$$

where y is the daily number of hires (or separations), S is dummy for the period after the onset of the pandemic (26 February 2020), R is a dummy switched on after the implementation of layoff restrictions (18 March 2020), c a linear daily time trend, and W is a vector of day-of-week fixed effects (i.e. d=1,...,7); u is an error term. Model (1) is estimated with negative binomial regressions and for two different sample sizes: (a) one for 2020 only, and (b) one covering the total period (2018-2020) for which ERGANI daily data are available. In the latter case, models additionally control for month and year fixed effects. Under the assumption that pre-pandemic labor market flows would have prevailed in the absence of the pandemic and the government responses to it, this method offer an approach for identifying COVID-19-related impacts on daily labor market activity.

However, changing trends before and after the pandemic onset and the related government interventions could be driven by unobserved factors. To account for such unobservables, we follow Powdthavee et al. (2019) and Metcalfe et al. (2011) in constructing a "control group" of observations based on trends from earlier years unaffected by COVID-19. More specifically, we compare the size of weekly labor market flows during weeks after Greece was exposed to the virus ("treated" group) with weekly flows for previous years ("control") group.[12] The exposure period for hires begins after week 9 in 2020, corresponding to the first COVID-19 case, while, for separations, the exposure period is after week 12 when layoffs were restricted by the Government.

We use a difference-in-differences (DiD) approach, with the identifying assumption being that control and treated weeks move on parallel trends before the exposure period. This is tested visually as well as by including leads of the treatment for a sufficient number of weeks before (Autor, 2003; Cookson and Laliotis, 2017). The model is the following:

$$m_{it} = \alpha_0 + \alpha_1 Treat_i + \alpha_2 Post_t + \alpha_3 Treat_i \times Post_t + e_{it} \quad (2)$$

where m is the number of hires (or separations) in week t (t=1,...,21) for group i (i=0,1), Treat is a binary indicator equal to 1 if the week is observed in 2020 and 0 if observed before (average of weeks 2018-2019), Post is a dummy switched on during the

Covid Economics 43, 21 August 2020: 95-136

exposure period for both groups (week 9 or 12) and e is an error term. In this quasi-experimental setup, the coefficient of interest is the one associated with the interaction of treatment and exposure period indicators, i.e., α_3. As the lockdown was lifted on May 4, we restrict this analysis for weeks 1-21 so our DiD estimates are not affected by increased labor market activity due to relaxing restrictions. Under the pre-exposure parallel trends assumption, the estimated DiD coefficients will indicate the short-term labor market impacts of COVID-19. As in (1), the model is estimated using negative binomial regressions.

4.4. Online vacancies

Daily data on the number of vacancies posted online are extracted from the two most popular job search portals in Greece. In order to achieve a wide coverage of the market, we use Alexa's ranking, which is web traffic data-based metric.[13] An automated data acquisition mechanism was set up to scrape and store daily information on job postings from all selected portals. Extracted data were pre-processed (e.g., string cleaning, language detection, avoid multiple entries per job ad and harmonization of company name and sectoral affiliation) before being used for the analysis. Advanced machine learning techniques were employed for deduplication. After deduplication, a total of 17,812 job vacancies were collected. Most job vacancies cover occupations such as sales and purchasing agents and brokers, administrative and specialized secretaries, administration professionals, transport and storage laborers, and information and communications technology professionals. A combination of Natural Language Processing and Name Entity Extraction/Recognition methods was adopted to extract the core information from the job postings. Examples of job posting fields extracted through this process include: job title, job description, job category, location, job type, contract type, experience, qualification, employer, employer type, firm location, firm size. The extracted data cover the period from January 2020 through June 2020.

4.5. Job search and finding employment

One additional question we investigate is whether the lockdown and economic slowdown may have reduced job search activity. To answer this question, we use individual-level data from the quarterly LFS to estimate the probability that those not working in a specific quarter were actively searching for a job during that quarter. The estimation sample consists of jobless individuals observed in the first quarters of 2017-2020 who lost their jobs over the previous two years. We estimate the following:

$$\text{Prob}(U_i=1) = a + Y^{2020}\times[b^q Q^q_i + \gamma X_i] + \varepsilon_i \qquad (3)$$

where U indicates the i-th non-employed jobseeker, Y^{2020} is a dummy indicator which takes the value of 1 for the first quarter of 2020 and 0 for first quarters of earlier years, Q is a vector of quarter dummies (q=1,...,8) and X is a vector of observable individual characteristics, i.e. gender, age, country of birth, education, region and sector of economic activity in the individual's last job, as well as indicators for the reason they stopped working (i.e. laid-off, contract termination, and other reasons).

Covid Economics 43, 21 August 2020: 95-136

In addition, we estimate the probability of finding employment during the first quarter in each of 2017, 2018, 2019, and 2020 for those were out of employment prior to that quarter.

$$Prob(E_i=1) = \alpha + Y^{2020} \times [\beta^t T^t_i + \delta Z_i] + \eta_i \quad (4)$$

where E indicates the i-th individual entered into employment, Y^{2020} is a dummy indicator which takes the value of 1 for the first quarter of 2020 and 0 for first quarters of earlier years, T is a vector of year dummies (t=2018, 2019, 2020) and X is a vector of individual characteristics, i.e. gender, age, country of birth, education, and region. Both models in (3) and (4) use a probit link function.

5. Analysis of impacts on the labor market

5.1. Labor market indicators

Table 1 shows non-seasonally-adjusted monthly estimates of the main labor market indicators from the LFS since the onset of the pandemic. The February figures show the improvement in the Greek labor market during the period prior to the pandemic.[14] Most notably, the number of workers unemployed in February 2020 was 16.9% lower than in 2019, corresponding to a nearly three percentage point drop in the unemployment rate (from 19.8% to 17%). On the eve of the pandemic, unemployment had been falling compared to the previous year and there had been a modest increase in employment.

The data for March and April present the evolution of labor market conditions while the lockdown was in place. Normally these are months when seasonal factors lead to job creation and an improvement in labor indicators in Greece. In 2020, employment numbers rose very slightly in March and April, less than they had in 2019. The unemployment figures are interesting. Between February and March 2020, the number of unemployed actually decreased by 13.8%, larger than the decrease in 2019. However, as the lockdown continued in April, unemployment numbers rose by 9.8%. Yet unemployment in April 2020 was still 14.3% lower than it had been a year earlier.

Covid Economics 43, 21 August 2020: 95-136

Table 1. Main labor market indicators, February, March, and April, 2019 and 2020

	February	March	April	% change (monthly)	
	[1]	[2]	[3]	[2] vs [1]	[3] vs [2]
		2019			
[4] Employed	3758.9	3846.3	3884.3	2.3	1.0
[5] Unemployed	928.0	844.0	852.6	-9.1	1.0
[6] Inactive	3261.8	3254.4	3203.9	-0.2	-1.6
		2020			
[7] Employed	3779.2	3813.0	3839.3	0.9	0.7
[8] Unemployed	771.6	665.4	730.3	-13.8	9.8
[9] Inactive	3353.1	3423.1	3329.4	2.1	-2.7
		% change (annual)			
[7] vs [4]	0.5	-0.9	-1.2		
[8] vs [5]	-16.9	-21.2	-14.3		
[9] vs [6]	2.8	5.2	3.9		

Source: Labor Force Survey (EL.STAT.)
Notes: Seasonally unadjusted estimates for persons 15-74 years old.

In part, the falling unemployment rates reflect higher inactivity, as jobless workers were less likely to search for work. Table 1 confirms that more workers were inactive in February 2020 than one year earlier. The pandemic accentuated this trend: in March inactivity increased by 2.1%. However, labor force participation statistics need to be carefully interpreted while the lockdown is in place since some people who are counted as inactive might still have some attachment to the labor force. This is evident when we consider the "potential additional labor force" (PALF), which includes those "seeking work but who are not immediately available" and those "available for work and wanting to work but not currently seeking work" (Hornstein et al., 2014). According to LFS data, the size of the PALF increased by 40% in 2020Q1 compared to 2019Q1, and by 72% compared to 2019Q4.[15] This suggests that lockdown measures increased the underutilization of labor, with growing numbers awaiting recall, unable to look for jobs because of the lockdown, or discouraged by the lack of new job openings. As economic activity resumes, these marginally attached workers may be more likely to (re)join the labor force.

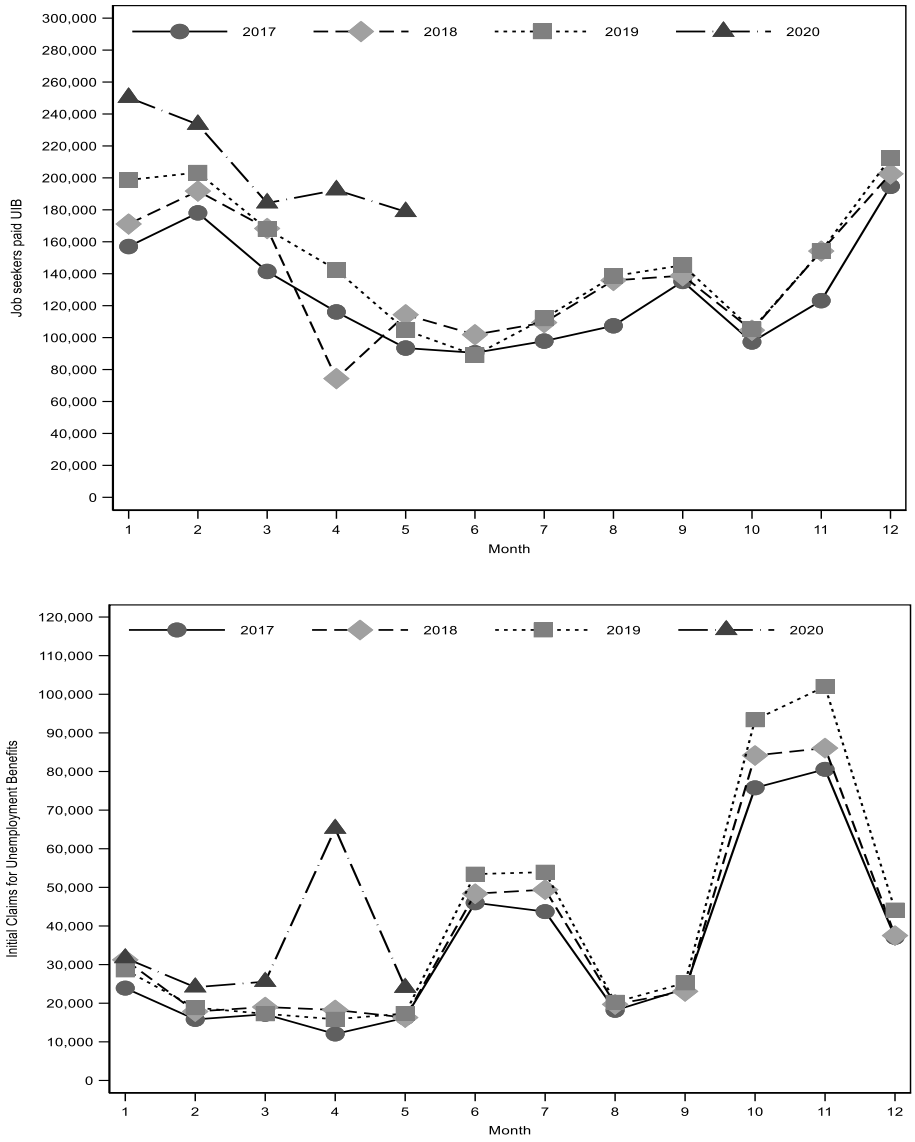
5.2 Unemployment claims

Figure 4 shows the monthly number of recipients of unemployment insurance benefits. The top panel shows overall beneficiaries, demonstrating a pattern which reflects the seasonal character of the Greek economy. The bottom panel shows that new claims for unemployment insurance benefits increased slightly in February and March 2020, and then tripled in April 2020. However, in May, initial claims moved back to a level and trend similar to that of earlier years.

Covid Economics 43, 21 August 2020: 95-136

Covid Economics 43, 21 August 2020: 95-136

Figure 4. Unemployment insurance benefit recipients and new monthly claimants



Source: OAED Monthly Reports. Authors' calculations.

5.3 Labor market flows

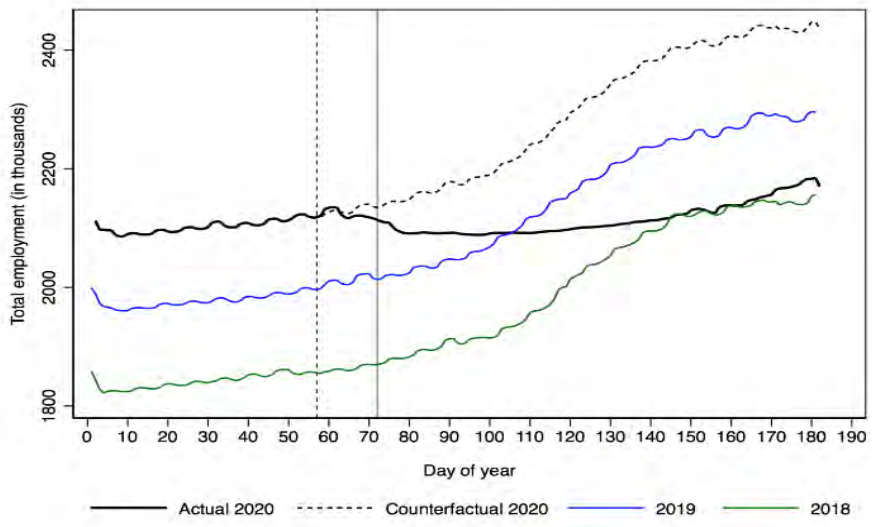
The employment registry, ERGANI, provides a unique data source since employment levels and flows in and out of employment can be tracked on a daily basis. Figure 5 presents the day-by-day employment for 2020, from the start of the year to the end of June. The first COVID case and the introduction of the workplace restrictions are marked with vertical lines on the chart so that the employment trend can be observed with reference to these key dates. In order to provide perspective on the 2020 numbers, a counterfactual trend line is included which estimates what the 2020

Covid Economics 43, 21 August 2020: 95-136

employment levels would have been if the daily 2018 and 2019 patterns had prevailed in 2020. The chart also shows the actual numbers for 2018 and 2019.

Figure 5 shows that employment started decreasing after COVID-19 appeared in Greece and this continued for a few days (about ten) during the lockdown period. At that point, employment levelled off and then gradually started increasing in May, when restrictions started to be relaxed. However, the employment impact is more striking when actual trends are compared with our best estimate of what the employment trajectory would have been in the absence of the pandemic and lockdown (i.e., based on 2018 and 2019 trends in daily changes).[16] This modest increase in employment corresponds to a period when job growth tends to be strong in Greece, because of seasonal factors primarily associated with the gearing up of the tourism industry. Comparing the actual employment to the counterfactual employment level results to a job deficit of 265,000 by end of June. This corresponds to a loss of 11.9% in total employment relative to a no-pandemic scenario.

Figure 5. Observed and counterfactual daily employment levels for 2020 and employment levels for 2018 and 2019

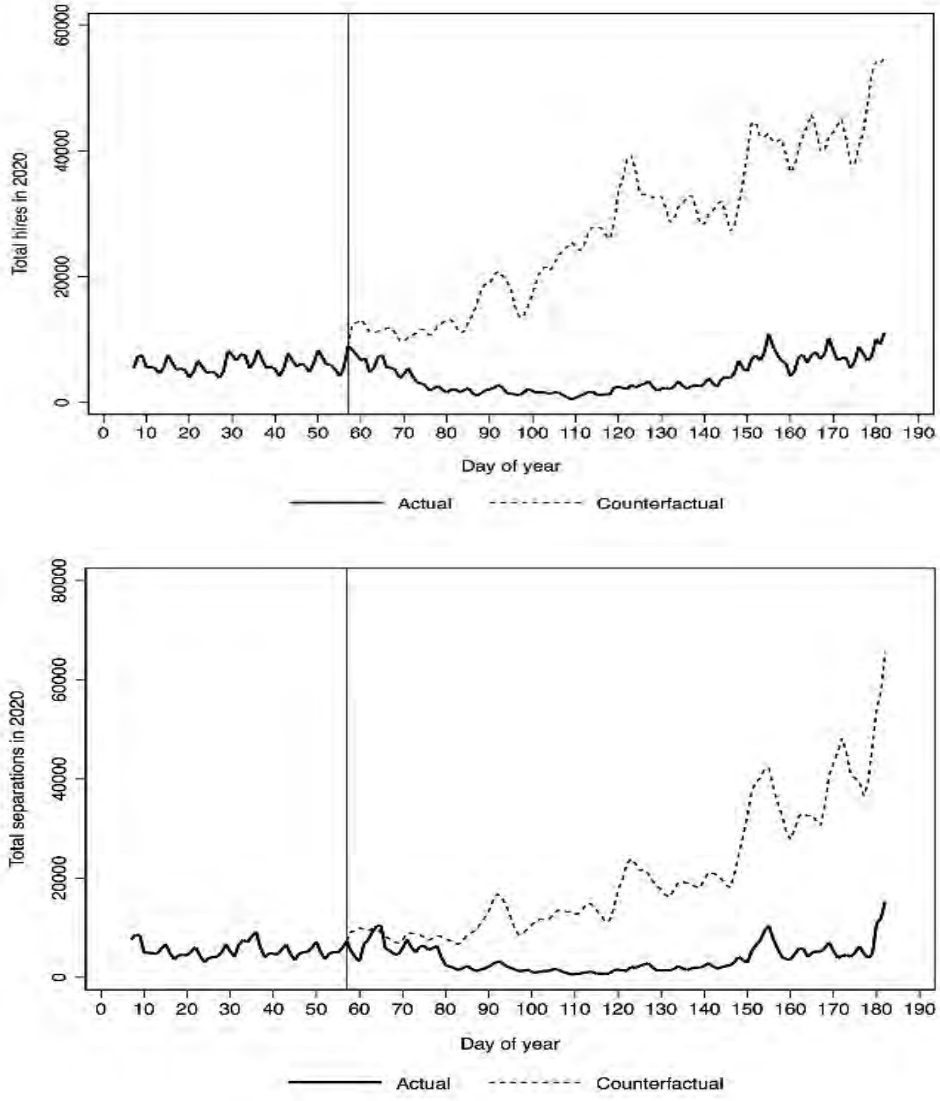


Source: ERGANI.
Notes: Vertical lines are set in the days when the first COVID-19 case was identified (26 February 2020) and when workplace restrictions were implemented (12 March 2020).

Changes in employment levels are explained by trends in hires and separations. The impact on jobs we have observed in Greece is completely due to the effect of the pandemic on new hires. Figure 6 shows the daily progression in total new hires (top panel) and separations (bottom panel) and compares the actual to the counterfactual trends, calculated as before. The decline in new hires, compared to the counterfactual scenario, is apparent. This occurs for all types of hires, i.e., full-time, part-time and shift work (see Appendix Figure A.2).

Covid Economics 43, 21 August 2020: 95-136

Figure 6. Total daily hires and separations before and after the pandemic onset, with comparison to counterfactual scenario



Source: ERGANI. Authors' calculations.
Notes: Vertical line is set at the pandemic onset (February 26, 2020).

On the other hand, as the bottom panel of Figure 6 shows, there is no evidence that separations have increased because of the pandemic and lockdown. In fact, the actual number of separations is below what would have been expected if the 2018-19 trends had continued in 2020. This is true for layoffs, quits, and contract terminations (Appendix Figure A.1). Certainly, in the case of layoffs, this can be explained by the Government measures to protect jobs by prohibiting layoffs in affected industries and by tying income support to the maintenance of employment relationships. The reduction in quits is not surprising since one would expect fewer workers to leave their

jobs in a deteriorating labor market. It should be noticed that there is initial evidence of a small uptick in the number of separations at the end of the period, which may reflect the easing of the layoff restrictions.

Using monthly data on net labor market flows, we observe that the crisis has affected sectors and occupations differently. Table 2 presents the difference between 2020 and 2019 in the size of net flows (new hires minus separations) for January, February, March, and April by sector of economic activity. The results suggest that the accommodation and food sector was particularly affected by the crisis. The negative impacts were especially severe in March and April, the months in which tourism would normally be gearing up for the summer season. In March, accommodation and food services accounted for 52% of the 2020 net job decreases, relative to 2019, while by April, this share was 84%.

Table 2. Comparison of net job flows between 2019 and 2020 by sector

	Difference: Month 2020 – Month 2019			
NACE Rev. 2 sector of economic activity:	January	February	March	April
Agriculture, forestry etc.	402	8	-413	-127
Mining and quarrying	22	-75	-164	-58
Manufacturing	1151	83	-5014	-482
Electricity, gas, steam etc.	-419	829	-42	-248
Water supply; sewerage etc.	-187	-188	-53	-18
Construction	548	-383	-1472	292
Wholesale and retail trade	1009	592	-7310	-8119
Transportation and storage	500	371	-6556	-3572
Accommodation and food service	3865	3415	-43120	-84491
Information and communication	-80	-665	-1965	447
Financial and insurance activities	-485	-188	-415	126
Real estate activities	-17	-34	-405	-769
Professional, scientific and technical	-634	-1303	-3000	-389
Administrative and support service	-688	-515	-4896	-3435
Public administration, defense etc.	-941	-2612	235	-260
Education	299	-350	-1798	214
Human health and social work	179	-846	-943	2038
Arts, entertainment and recreation	573	-728	-4684	309
Other service, households and extra	-119	-295	-1623	-2114
Total additional jobs	4978	-2884	-83638	-100656

Source: ERGANI and National Institute of Labor and Human Resources (NILHR). Authors' calculations.

Table 3 presents a similar analysis by occupation. During March and April, the most affected group was workers employed in services and shop and market sales workers. This group accounted for a two-thirds of the total drop in net job flows in April. A large number of workers in this occupation in Greece are employed in the tourism sector.

Covid Economics 43, 21 August 2020: 95-136

Table 3. Comparison of net job flows between 2019 and 2020 by occupation

	Difference: Month 2020 – Month 2019			
NACE Rev. 2 sector of economic activity:	January	February	March	April
Legislators, senior officials and managers	-10	43	-641	-498
Professionals	418	-2890	-3951	990
Technicians and associate professionals	1970	-2018	-8894	-1362
Clerks	-951	-2508	-11418	-14781
Service workers and shop and market sales workers	1333	3456	-35567	-66066
Skilled agricultural and fishery workers	43	-89	-417	-388
Craft and related trades workers	-156	491	-3285	-1452
Plant and machine operators and assemblers	511	-11	-7577	-3452
Elementary occupations	1860	552	-13527	-16683
Total additional jobs	5018	-2974	-85277	-103692

Source: ERGANI and National Institute of Labor and Human Resources (NILHR). Authors' calculations.

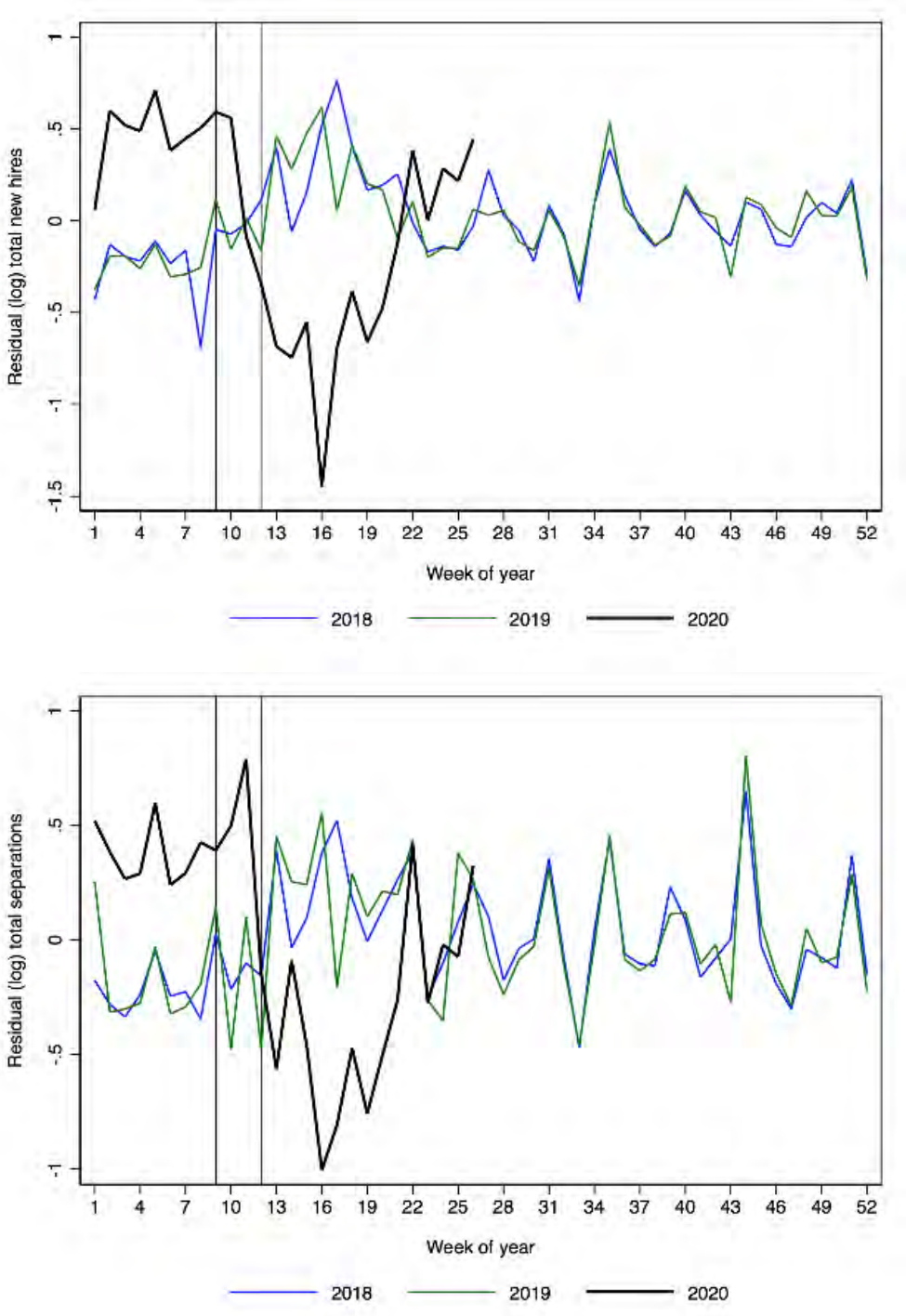
Covid Economics 43, 21 August 2020: 95-136

We now turn to our more formal analysis of the impact of the pandemic on labor market flows, first using interrupted time series and then difference-in-difference estimates. Regarding the former, we regress the log of total new hires (or separations) on a day-of-week, month, year fixed effects and a linear time trend, and then plot the mean residual by week and year. The results are shown in Figure 7; vertical lines are set at the weeks when the pandemic started and when restrictions on layoffs were implemented. The results confirm what we observed through the descriptive data analysis. There is a pronounced decline in new hires after the pandemic appeared and this decline started in those weeks during which new hires peaked in the pre-pandemic years. At the same time, separations were lower compared to the respective weeks of pre-pandemic "normal" years.

As the data show signs of over-dispersion, Equation (1) is also estimated using negative binomial regression. Under this specification, b_1 indicates the mean daily change on each outcome after the onset of the pandemic and b_2 denotes the effect of layoff restrictions. Estimation results are summarized in Table 4. The interrupted time series regression estimates confirm the graphical evidence. In 2020, new hires decrease by a significant 1.12%, on average, each day during the pandemic crisis. The effect is more sizeable for part-timers (1.38%) and those in shift work (2.09%), although full-timers are the biggest group in the labor market. Moreover, it seems that the labor market responded with a slight delay, in terms of full-time hires, after the onset of the pandemic, although coefficient estimates are everywhere negative. On the other hand, separations are significantly decreased relative to pre-pandemic days. This is true especially for firings which, as already noted, were restricted during the crisis.

Covid Economics 43, 21 August 2020: 95-136

Figure 7. Residual hires and separations by week and year



Source: ERGANI. Authors' calculations.

Table 4. Pandemic onset and layoff restriction effects on hires and separations: Interrupted time series estimates.

	Days since pandemic	Days since layoff restrictions	Days since pandemic	Days since layoff restrictions
Dependent variable:	[1]	[2]	[3]	[4]
Total new hires	-.0112***	-.0058**	.0002	-.0047**
	(.0029)	(.0023)	(.0020)	(.0019)
Full-time new hires	-.0071***	-.0055***	.0011	-.0049***
	(.0027)	(.0020)	(.0019)	(.0019)
Part-time new hires	-.0138***	-.0056**	-.0007	-.0044**
	(.0031)	(.0025)	(.0021)	(.0021)
Shift-work new hires	-.0209***	-.0079***	-.0011	-.0049*
	(.0038)	(.0031)	(.0028)	(.0028)
Total separations	-.0001	-.0070***	.0050***	-.0099***
	(.0030)	(.0022)	(.0017)	(.0016)
Quits	.0020	-.0087***	.0060***	-.0103***
	(.0028)	(.0020)	(.0016)	(.0015)
Firings	.0041	-.0120***	.0075***	-.0137***
	(.0031)	(.0020)	(.0016)	(.0014)
Contract terminations	-.0031	-.0028	.0024	-.0077***
	(.0037)	(.0031)	(.0028)	(.0025)
Day of week fixed effects	Yes	Yes	Yes	Yes
Month fixed effects	No	No	Yes	Yes
Year fixed effects	No	No	Yes	Yes
Daily time trend	Yes	Yes	Yes	Yes
Period covered	01 Jan 2020 – 30 Jun 2020	01 Jan 2020 – 30 Jun 2020	01 Jan 2018 – 30 Jun 2020	01 Jan 2018 – 30 Jun 2020
Observations	182	182	912	912

Source: ERGANI. Authors' calculations.
Notes: Negative binomial regression estimates. Robust standard errors in parentheses.
*** at 1%, ** at 5% and * at 1%.

Turning to the DiD estimates, this approach requires an identifying assumption that control and treated weeks move on parallel trends before the exposure period. To assess this, we have graphed weekly trends for total employment, hires, and separations for both time periods (see Appendix Figure A.2). In each case, the figures show that the trends move in parallel before the outbreak.

Table 5 presents the DiD results for cumulative employment, total new hires and total separations. Results for the last two variables are also presented by type of hire and separation. In all cases, the estimated DiD parameters are sizeable and highly significant confirming that the labor market impact of the pandemic has been quite severe, at least in the short-run. Cumulative employment and new hires (overall and by job type) in 2020 fell substantially after week 9, relative to the control group. For separations, the exposure period is set at week 12 when the government intervened to restrict layoffs. The associated DiD coefficients, overall and by separation type, are also sizeable and significant.

In addition to the graphical evidence confirming that employment, hires, and separations trended similarly before the pandemic onset and the layoff restrictions (Appendix Figure A.2), we also report results of an additional test. In this test outcomes are regressed on the set of controls already controlled for in Table 5, plus leads of the interaction term (Equation (2)) that range from one to five weeks before

Covid Economics 43, 21 August 2020: 95-136

the actual treatment takes place. The size and significance of those estimates will indicate how the series trended before the treatment period (Table 6).

Table 5. Pandemic and layoff restriction effects on labor market: Difference-in-differences results.

	DiD coefficient	Treatment group coefficient	Treatment period coefficient[1]
Dependent variable:	[1]	[2]	[3]
Cumulative employment	-.0602*** (.0092)	.0964*** (.0034)	.0127 (.0128)
Total new hires	-1.196*** (.1718)	.1138* (.0624)	.5483** (.2312)
Full-time hires	-1.097*** (.1594)	.1171 (.0724)	.6174*** (.2218)
Part-time hires	-1.264*** (.2009)	.1326** (.0651)	.4919* (.2524)
Shift work hires	-1.558*** (.2978)	.0432 (.1037)	.5835 (.3558)
Total separations	-1.330*** (.1650)	.1539 (.1113)	.1702 (.1590)
Firings	-1.407*** (.2076)	.1711 (.1101)	.1748 (.2052)
Quits	-1.236*** (.1576)	.1507* (.0812)	.1066 (.1462)
Contract terminations	-1.423*** (.2167)	.1481 (.1702)	.2550 (.2176)

Source: ERGANI. Authors' calculations.
Notes: Negative binomial regression estimates. Robust standard errors in parentheses. All models include a constant and a weekly linear trend. Sample size covers weeks 1-21 (lockdown lifting) and the effective observations are 42 in all models (21 weeks; 2 groups). [1] Treatment period for cumulative employment and hires is week 9 onwards (pandemic onset). Treatment period for separations runs from week 12 onwards (layoff restrictions).
*** at 1%, ** at 5% and * at 1%.

Table 6. Testing for parallel trends before the treatment period.

	-5 weeks lead	-4 weeks lead	-3 weeks lead	-2 weeks lead	-1 weeks lead	0 weeks lead
Dependent variable:	[1]	[2]	[3]	[4]	[5]	[6]
Cumulative employment	-.008*	-.002	-.005***	-.002	-.001	-.051***
	(.004)	(.001)	(.001)	(.001)	(.001)	(.008)
Total new hires	-.050	.372***	-.222***	.076***	-.001	-1.28***
	(.076)	(.021)	(.021)	(.021)	(.021)	(.157)
Full-time hires	-.120**	.486***	-.345***	.065***	-.086***	-1.05***
	(.052)	(.018)	(.018)	(.018)	(.018)	(.149)
Part-time hires	-.039	.289***	-.128***	.075***	.060**	-1.421***
	(.106)	(.023)	(.023)	(.023)	(.024)	(.192)
Shift work hires	.199*	.232***	-.048	.131***	.082**	-1.878***
	(.110)	(.034)	(.033)	(.034)	(.033)	(.288)
Total separations	-.148	.020	.197***	.025	.015	-1.437***
	(.126)	(.017)	(.017)	(.017)	(.017)	(.149)
Firings	-.141	-.010	.324***	.162***	.411***	-2.050***
	(.108)	(.020)	(.020)	(.020)	(.020)	(.179)
Quits	-.244***	-.032**	.215***	.190***	.094***	-1.472***
	(.090)	(.015)	(.015)	(.015)	(.015)	(.154)
Contract terminations	-.050	.075***	.143***	-.209***	-.422***	-.983***
	(.196)	(.023)	(.023)	(.023)	(.023)	(.174)

Source: ERGANI. Authors' calculations.
Notes: Negative binomial regression estimates. Robust standard errors in parentheses. All models include a constant and a weekly linear trend. Sample size covers weeks 1-21 (lockdown lifting) and the effective observations are 42 in all models (21 weeks; 2 groups). [1] Treatment period for cumulative employment and hires is week 9 onwards (pandemic onset). Treatment period for separations runs from week 12 onwards (layoff restrictions).
*** at 1%, ** at 5% and * at 1%.

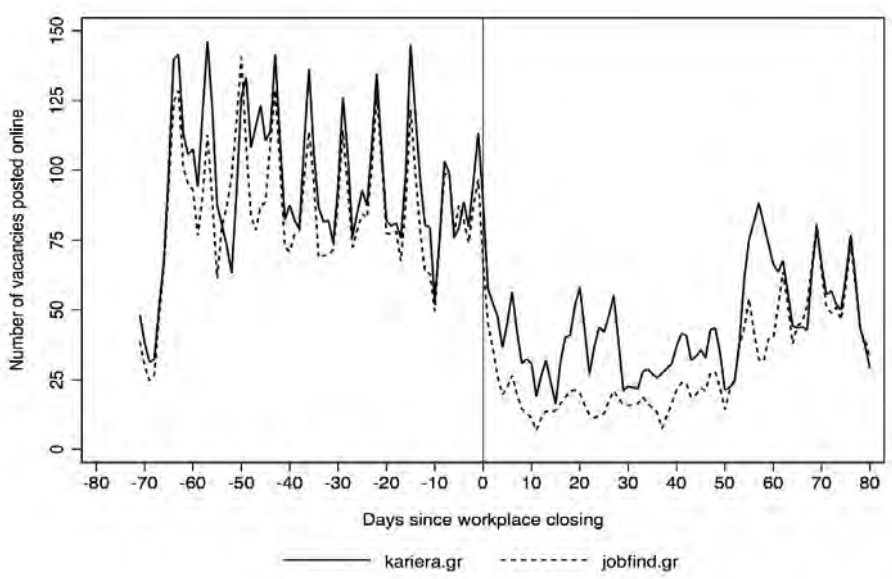
Covid Economics 43, 21 August 2020: 95-136

For cumulative employment, the results confirm the parallel trend hypothesis for treated and control groups before the pandemic. The estimated coefficients are not significant and remarkably low; the effect comes when specifying a zero weeks lead and it is comparable to the respective DiD estimate from Table 5. No clear patterns emerge when considering hires-related outcomes, the effects on the zero weeks lead coefficient are in line with the reported DiD estimates. When considering separations, there are some sizeable and positive effects taking place even three weeks before the actual treatment takes place. This is also in line with the graphical evidence (Figure 8), indicating that there might have been some anticipation of the pandemic in terms of labor market activity that induced the government to step in and restrict layoffs. Hence despite an upward tendency in separations, these were drastically reduced from week 12 onwards. Again, the coefficients when specifying a zero-week lead are comparable with the DiD estimates shown before.

5.4 Online vacancies

To further understand hiring dynamics during the crisis, we use daily data from two popular online job portals in Greece (www.kariera.gr and www.jobfind.gr). Following Hensvik et al. (2020), we measure the changes in labor demand by the average daily inflows of new vacancies. Although we do not have evidence to assess the representativeness or coverage of these data, the number of vacancies posted on them is consistent with the sharp decline in new hires reported in ERGANI. The results, summarized in Figure 8, show a steep decrease in new vacancies posted on both sites corresponding to the implementation of the workplace restrictions in March. There seems to be a slight but fluctuating increase in job postings in May as the restrictions were lifted but the number of new postings was still far below pre-pandemic levels.

Figure 8. Job vacancy postings, kariera.gr and jobfind.gr, January-May 2020



Source: kariera.gr and jobfind.gr

5.5 Job search and finding employment

We also look at the impact of the pandemic on job search behaviors and on the likelihood of finding employment. First, we use the LFS micro dataset to estimate Equation (2) and to assess the probability that those not employed are actively seeking work. The associated marginal effects are shown in Table 7.

Table 7. Job seeking during the COVID-19 pandemic

	2019-2020		2017-2020	
	[1]	[2]	[3]	[4]
Reasons for stop working				
Laid-off	-.128***	-.112***	-.156***	-.139***
	(.034)	(.035)	(.031)	(.031)
Contract termination	-.143***	-.144***	-.163***	-.158***
	(.023)	(.023)	(.022)	(.022)
Other	-.047	-.046	.018	.022
	(.045)	(.045)	(.043)	(.043)
Quarters since stop working				
0 (current quarter)	-	-.383***	-	-.467***
		(.079)		(.069)
1	-	-.135***	-	-.130***
		(.031)		(.031)
2	-	-.022	-	-.065*
		(.043)		(.039)
3	-	-.222***	-	-.210***
		(.062)		(.060)
4	-	-.070	-	-.134**
		(.065)		(.056)
5	-	-.148***	-	-.164***
		(.056)		(.054)
6	-	-.049	-	-.027
		(.057)		(.056)
7	-	-.054	-	.028
		(.056)		(.055)
Observations	3416		7637	

Source: EL.STAT, Quarterly Labor Force Survey-LFS (2017 – 2020, 1st quarter).
Notes: Reported estimates are average marginal effects drawn from a probit model (interaction effects model) and correspond to the post 2020Q1 period (additive effect). Based on the ILO definition of unemployment, the dependent variable takes the value of 1 if the individual is considered to be unemployed and 0 otherwise. Sample includes all individuals (aged 25-54) who have stopped working during the last 8 quarters (2 years). All models include controls for gender, age, country of birth, education, region and sector of economic activity of the last job. In parentheses, white heteroskedasticity corrected standard errors are reported. All estimates are weighted using the sampling weights provided by the EL.STAT.
*** at 1%, ** at 5% and * at 1%.

We observe (column 2) that, for those who had been laid-off, the probability of searching for work during the first quarter of 2020 was 11.2 percentage points lower than in the first quarter of 2019. For those who were not working because their last employment contract had been terminated, the drop in the probability of searching employment was even stronger, i.e. 14.4 percentage points lower. The largest drop was experienced by those who had lost their job within the current quarter (38.3 percentage points). This suggests that the slackness of the labor market in the first quarter of 2020 mostly affected the newly jobless individuals. These results are confirmed even when additional years of first quarters are added in the model (column 4).

Covid Economics 43, 21 August 2020: 95-136

We now turn to the estimation results of the probability of finding employment during the 1st quarter of 2020. Table 8 presents the estimated marginal effects of the probit model described in equation (3). We observe (column 2) that the probability of finding employment in the first quarter of 2020 was 4.6 percentage points lower than during the first quarter of 2019. When additional years are added (columns 3 and 4), we observe that the estimated marginal effect for the first quarter of 2020 compared to the corresponding quarter of 2017 is negative although only marginally significant. These results suggest that the slackness of the labor market in the first quarter of 2020 may well have contributed to slowing down the employment prospects of jobless individuals.

Table 8. Employment entry during the onset of COVID pandemic

	2019-2020		2017-2020	
Year start work	[1]	[2]	[3]	[4]
2020	-.045***	-.046***	-.017*	-.016
	(.010)	(.010)	(.009)	(.009)
2019	-	-	.028***	.029***
			(.010)	(.010)
2018	-	-	.001	.001
			(.009)	(.009)
Fixed effects	-	Yes	-	Yes
Observations	3802		7942	

Source: Hellenic Statistical Authority (EL.STAT.), Quarterly Labor Force Survey-LFS (2017 – 2020, 1st quarter).
Notes: Reported estimates are average marginal effects drawn from a probit model (the first year serves as the reference category, i.e., 2019 and 2017 in columns 1-2 and 3-4, respectively). Based on the ILO definition of employment, the dependent variable takes the value of 1 if the individual has start working in the current employer during the last 3 months and 0 otherwise. Sample includes all individuals (aged 15-54) who have start working any month during the last 8 quarters (2 years). In all models fixed effects include controls for gender, age, country of birth, education, region. In parentheses, white heteroskedasticity corrected standard errors are reported. All estimates are weighted using the sampling weights provided by the EL.STAT. *** at 1%, ** at 5% and * at 1%.

6. Conclusions

In many respects, Greece has been an interesting case for studying the COVID-19 pandemic and its impacts on employment. Although there has been a recent increase in cases, the virus has been controlled well when compared to other countries in Europe and elsewhere in the OECD. To a significant degree, this was due to a stringent lockdown quickly imposed by the Government after the first cases were confirmed in late February. Even without widespread contagion, though, the pandemic has had an important economic impact, with GDP expected to decrease as much as 10% in 2020.

The timing of the pandemic and lockdown is also an important part of the Greek story, in two ways. First, COVID-19 arrived at a point when the economy seemed to be finally on a sustainable growth path after the economic crisis that had persisted for the past decade. Second, the lockdown covered a period when the heavily seasonal Greek economy, quite reliant on tourism, would normally be gearing up and creating large numbers of jobs – something that could not happen in the spring of 2020. This is important to keep in mind in order to fully understand how the labor market has been affected by the COVID-19 crisis. A final characteristic of the Greek experience was the

Covid Economics 43, 21 August 2020: 95-136

Government decision to mitigate the economic consequences of the crisis by introducing regulatory and income support measures to maintain employment relationships. This has had an important effect on how the lockdown and the reduced labor demand have played out in the labor market.

We use administrative, survey, and online vacancy data to analyze how employment in Greece was affected during the first few months of the COVID-19 pandemic and lockdown. Our main findings are the following: First, in the early months of the lockdown, labor force participation and unemployment fell, while there was very little change in employment levels. Second, job search activity declined, both because of continued attachment to employers who had suspended operations and because of almost no hiring activity. Third, by the end of June, we estimate that (registered) employment was 11.9% less than it should have been, based on trends from the previous two years. Fourth, this lost employment was entirely due to the sharp decline in hiring activity in the first few months of the crisis. This is evident from both the administrative and online vacancy data. As noted above, the early months spanned a period when seasonal activities, especially related to tourism, would normally be expanding and this needs to be taken into consideration in assessing the impact of the crisis. Most of the "missing" jobs thus far in 2020 have been in accommodation and food, which reflects the pandemic's effect on tourism. Fifth, and somewhat unexpectedly, separations to the end of May were *lower* than would be predicted based on the trends from recent years. This almost certainly was due to the Government measures to protect existing employment relationships. This was done through a prohibition of layoffs in industries affected by the crisis and by tying the major form of income support to the maintenance of jobs.

To sum up, the analysis points to the important role that policy has played in determining how the Greek labor market has adjusted to the pandemic and lockdown. The measures put in place by the Government to mitigate the effects of the crisis on employers and workers emphasized job protection. The decreased labor demand, then, translated into a downturn in hiring rather than increases in separations that would lead to higher unemployment. In this respect, Greece has been similar to some other European countries that have adopted measures to avoid layoffs. This stands in contrast to some other countries, like the US and Canada, where unemployment rose quickly as policies emphasized income support more than job protection. Of course, it is still far too early to assess the efficacy of the different approaches. However, at least in the short run, the policy stance adopted by Greece and others to maintain employment relationships where possible seems to have had positive attributes.

We are still, of course, in the early days when it comes to understanding how COVID-19 is affecting labor markets in Greece and elsewhere. The analysis in this paper largely covers just the lockdown period and does not include analysis of what is happening to employment as Greece emerges from the lockdown and implements a "new normal" which may or may not include further lockdowns. So, there is an important research agenda going forward.

Covid Economics 43, 21 August 2020: 95-136

In the next stage of our research, we plan to assess three topics. First, an in-depth examination of how the impacts of the pandemic and lockdown were distributed across different types of workers, different occupations and industries, and different parts of the country. Second, an updated analysis of the labor market and employment relations adjustments as Greece emerges from the lockdown and as the mitigation measures are phased out. Third, the medium-term impacts of the mitigation strategies based on maintaining employment relationships rather than income support for workers who have lost their jobs. This last issue is particularly relevant for the ongoing debate in labor market policy about protecting jobs vs. protecting workers.

Covid Economics 43, 21 August 2020: 95-136

References

Abraham, K. (2020),"What is happening to unemployment in the post-Covid-19 labor market?", IZA World of Labor, available at: https://wol.iza.org/opinions/what-is-happening-to-unemployment-in-the-post-covid-19-labor-market

Adams-Prassl, A. Boneva, T., Golin, M., and Rauh, C. (2020), "Inequality in the Impact of the Coronavirus Shock: Evidence from Real Time Surveys," IZA Discussion Papers 13183, Institute of Labor Economics (IZA).

Alstadsæter, A., Bratsberg, B., Eielsen, G., Kopczuk, W., Markussen, S., Raaum, O., and Røed, K (2020), "The First Weeks of the Coronavirus Crisis: Who Got Hit, When and Why? Evidence from Norway", NBER Working Paper No. 27131, DOI: 10.3386/w27131.

Aum, S., Lee, S. Y. (Tim), and Shin, Y. (2020a), "COVID-19 Doesn't Need Lockdowns to Destroy Jobs: The Effect of Local Outbreaks in Korea", NBER Working Paper No. 27264, DOI: 10.3386/w27264.

Autor, D.H. (2003), Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing, Journal of Labor Economics, 21(1): pp. 1-42.

Baert, S., Lippens, L., Moens, E., Sterkens, P. and Weytjens, J., (2020), "How Do We Think the COVID-19 Crisis Will Affect Our Careers (If Any Remain)?", IZA Discussion Paper No 13164.

Barrero, J. M., Bloom, N., and Davis, S. J. (2020), "COVID-19 Is Also a Reallocation Shock"
NBER Working Paper No. 27137, DOI: 10.3386/w27137.

Bartik, A. W., Bertrand, M., Cullen, Z. B., Glaeser, E. L., Luca, M., and Stanton, C. T. (2020), "How Are Small Businesses Adjusting to COVID-19? Early Evidence from a Survey" NBER Working Paper No. 26989, DOI: 10.3386/w26989.

Béland, L. P., Brodeur, A., and Wright, T. (2020), "The Short-Term Economic Consequences of Covid-19: Exposure to Disease, Remote Work and Government Response", IZA discussion paper No 13159.

Béland, L. P., Brodeur, A., Mikola, D., and Wright, T. (2020), "The Short-Term Economic Consequences of Covid-19: Occupation Tasks and Mental Health in Canada", IZA Discussion Paper No 13254.

Belot, M., Choi, S.Tripodi, E. van den Broek-Altenburg, E.; Jamison, J.C. and Papageorge, N. W. (2020), "Unequal Consequences of COVID-19 across Age and Income: Representative Evidence from Six Countries", Covid Economics, 38: 196-217 (also IZA Discussion Paper No 13366).

Covid Economics 43, 21 August 2020: 95-136

Bennedsen, M, Larsen, B.,Schmutte, I.and Scur, D. (2020), "Preserving job matches during the COVID-19 pandemic: Firm-level evidence on the role of government aid", Covid Economics: Vetted and real-time papers No. 27.

Bick, A. and Blandin, A. (2020), "Real Time Labor Market Estimates During the 2020 Coronavirus Outbreak", Working Paper, available at: https://drive.google.com/file/d/1uQrlBJ_w4b2Fps6Zp9qgc-9pMnbC-mjq/view.

Brinca, P., Duarte J., Faria-e-Castro, M. (2020) "Measuring Sectoral Supply and Demand Shocks during COVID-19," Covid Economics, 20: 147-171 (also Federal Reserve Bank of St. Louis Working Papers 2020-011 D, DOI: 10.20955/wp.2020.011).

Borjas, G. J., and Cassidy, H. (2020), "The Adverse Effect of the COVID-19 Labor Market Shock on Immigrant Employment", NBER Working Paper No. 27243, DOI: 10.3386/w27243.

Bureau of Labor Statistics, U.S. Department of Labor (2020), "Frequently asked questions: The impact of the coronavirus (COVID-19) pandemic on The Employment Situation for May 2020", available at: https://www.bls.gov/cps/employment-situation-covid19-faq-may-2020.pdf

Brynjolfsson, E. and Horton, J., Ozimek, A., Rock D., Sharma, G., and TuYe, H.Y., (2020), "COVID-19 and Remote Work: An Early Look at US Data" NBER Working Papers No 27344.

Cajner, T., Dod Crane, L., Decker, R., Hamins-Puertolas, A., and Kurz, C. J. (2020), "Tracking Labor Market Developments during the COVID-19 Pandemic: A Preliminary Assessment", FEDS Working Paper No. 2020-030, DOI: 10.17016/FEDS.2020.030.

Campello, M., Kankanhalli, G., and Muthukrishnan, P. (2020), "Tracking labor market developments during the covid-19 pandemic: A preliminary assessment", Technical report, NBER Working Paper No 27208.

Cedefop (2018), "Mapping the landscape of online job vacancies. Background country report: Greece", available at: https://www.cedefop.europa.eu/files/rlmi_-_mapping_online_vacancies_greece.pdf.

Cho, S.J., Winters, J. V. (2020), "The Distributional Impacts of Early Employment Losses from COVID-19", IZA Discussion Paper No 13266.

Coibion, O., Gorodnichenko, Y. and Weber, M. (2020), "Labor Markets During the COVID-19 Crisis: A Preliminary View", Covid Economics, 21: 40-58 (also NBER Working Papers No 27017).

Cookson G. and Laliotis I. (2017), Promoting normal birth and reducing caesarean section rates: An evaluation of the Rapid Improvement Programme, Health Economics, 27(4): 675-679.

Covid Economics 43, 21 August 2020: 95-136

Cowan, B. W. (2020), "Short-run effects of COVID-19 on U.S. worker transitions" NBER Working Paper No 27315.

Dingel, J. I., and Neiman, B. (2020), "How Many Jobs Can be Done at Home?", Covid Economics, 1: 16-24 (also NBER Working Paper No. 26948, DOI: 10.3386/w26948).

Fairlie, R. W., Couch, K., and Xu, H. (2020), "The Impacts of COVID-19 on Minority Unemployment: First Evidence from April 2020 CPS Microdata", NBER Working Paper No. 27246, DOI: 10.3386/w27246.

Fairlie, R.W. (2020), "The Impact of COVID-19 on Small Business Owners: Continued Losses and the Partial Rebound in May 2020", NBER Working Paper No 27462.

Gardiner, L., Slaughter, H. (2020), "The effects of the coronavirus crisis on workers: Flash findings from the Resolution Foundation's coronavirus survey", Resolution Foundation, available at: https://www.resolutionfoundation.org/app/uploads/2020/05/The-effect-of-the-coronavirus-crisis-on-workers.pdf.

Garrote Sanchez, D., Gomez Parra, N., Ozden, C., Rijkers, B. (2020), "Which Jobs Are Most Vulnerable to COVID-19? What an Analysis of the European Union Reveals", Research and Policy Brief No. 34, World Bank, Washington, DC, available at: https://openknowledge.worldbank.org/handle/10986/33737.

Gentilini,U., Almenfi, M., Dale, P., Lopez, A. V., Mujica, I.V., Quintana. R., and Zafar, U. (2020), "Social Protection and Jobs Responses to COVID-19: A Real-Time Review of Country Measures (July 10, 2020)", COVID-19 Living Paper Washington, D.C. : World Bank Group, available at: http://documents1.worldbank.org/curated/en/454671594649637530/pdf/Social-Protection-and-Jobs-Responses-to-COID-19-A-Real-Time-Review-of-Country-Measures.pdf

Goldsmith-Pinkham, P., Sojourner, A. (2020), "Predicting Initial Unemployment Insurance Claims Using Google Trends", available at: https://paulgp.github.io/GoogleTrendsUINowcast/google_trends_UI.html.

Hamermesh, D. (2020), "Measuring employment and unemployment – Primer and predictions", IZA World of Labor, available at: https://wol.iza.org/opinions/measuring-employment-and-unemployment-primer-and-predictions

Hatayama, M., Viollaz, M. and Winkler, H. (2020a), "Work from home: which jobs?", Covid Economics: Vetted and Real-Time Papers, No. 19.

Hensvik L., Le Barbanchon T. and Rathelot, R. (2020) "Job Search During the COVID-19 Crisis", IZA Discussion Paper No. 13237.

Hicks, M.J., Faulk, D., and Devaraj, S. (2020), "Occupational Exposure to Social Distancing: A Preliminary Analysis using O*NET Data", Center for Business and

Covid Economics 43, 21 August 2020: 95-136

Economic Research, Ball University, Working Paper, available at: https://projects.cberdata.org/reports/SocialDistanceEffects-20200313.pdf.

Hornstein A., Kudlyak M. and Lange F. (2014), Measuring resource utilization in the labor market, Economic Quarterly, 100(1): pp. 1-21

ILO (2020a), "Working from home: estimating the worldwide potential", ILO Policy Brief, available at: https://www.ilo.org/wcmsp5/groups/public/---ed_protect/---protrav/---travail/documents/briefingnote/wcms_743447.pdf.

ILO (2020b), "COVID-19 impact on the collection of labour market statistics", ILOSTAT, available at: https://ilostat.ilo.org/topics/covid-19/covid-19-impact-on-labour-market-statistics/.

ILO (2020c), "ILO monitor: Covid-19 and the world of work",5th edition, available at: https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/documents/briefingnote/wcms_749399.pdf

Kahn, L., Lange, F., and Wiczer, D. (2020), "Labor Demand in the Time of COVID-19: Evidence from Vacancy Postings and UI Claims", NBER Working Paper No. 27061.

Kong, D., Prinz, D. (2020), "The impact of shutdown policies on unemployment during a pandemic", Covid Economics: Vetted and Real-Time Papers, No. 17.

Kurman, A.; Lale, E.; Ta, L. (2020). The Impact of COVID-19 on Small Business Employment and Hours: Real-Time Estimates With Homebase Data. Working Paper. http://www.andrekurmann.com/hb_covid

Metcalfe R., Powdthavee N. and Dolan P. (2011), Destruction and distress: Using a quasi-experiment to show the effects of September 11 attacks on mental well-being in the United Kingdom, Economic Journal, 121(550): pp. F81-F103.

Mongey, S., Pilossoph, L., and Weingberg, A. (2020), "Which workers bear the burden of social distancing measures?" Covid Economics, 12: 69-86 (also NBER Working Paper No 27085).

Montenovo, L, Jiang, X., Rojas, F. L., Schmutte, I. M., Simon, K. I., Weinberg B. A. and Wing, C. (2020), "Determinants of disparities in covid-19 job losses", NBER Working Paper No 27132.

OECD (2020), "OECD Economic Outlook", Volume 2020 Issue 1, OECD Publishing, Paris, available at: https://doi.org/10.1787/0d1d1e2e-en.

Pouliakas, K. (2020), "Working at Home in Greece: Unexplored Potential at Times of Social Distancing?", IZA Discussion Paper No 13408.

Pouliakas, K., Branka, J. (2020), "EU jobs at highest risk of Covid-19 social distancing: Is the pandemic exacerbating the labour market divide?", Luxembourg: Publications Office of the European Union, Cedefop Working Paper No 1.

Covid Economics 43, 21 August 2020: 95-136

Powdthavee N., Plangol A., Frijters A.C. and Clark A.E. (2019), Who got the Brexit blues? The effect of Brexit on subjective wellbeing in the UK, Economica, 86(343): pp. 471-494.

Rothwell, J. (2020), "The effects of COVID-19 on international labor markets: An update", Brookings, available at: https://www.brookings.edu/research/the-effects-of-covid-19-on-international-labor-markets-an-update/.

Saltiel, F. (2020), "Home working in developing countries", Covid Economics: Vetted and Real-Time Papers, No. 6.

World Bank (2020), "Global economic prospects", World Bank Group Flagship Report, June 2020.

Covid Economics 43, 21 August 2020: 95-136

Appendix

Table A.1: Labor market indicators, Greece, Q1 for 2018-2020

	Labor force participation (%)			Employment to population (%)			Unemployment (%)		
	2018Q1	2019Q1	2020Q1	2018Q1	2019Q1	2020Q1	2018Q1	2019Q1	2020Q1
Age 15 and over									
Total	51.62	51.78	50.61	40.68	41.83	42.21	21.19	19.21	16.21
Men	59.79	59.70	58.80	49.48	50.52	50.72	17.24	15.37	13.73
Women	44.01	44.42	43.01	32.48	33.75	34.69	26.18	24.02	19.35
Age group									
15-19	5.15	4.87	3.83	2.31	2.42	2.52	55.22	50.21	34.22
20-24	43.76	41.80	39.46	24.95	25.21	25.86	42.97	39.69	34.47
25-29	81.66	84.15	80.44	56.23	60.36	59.49	31.13	28.27	26.05
30-34	86.89	86.97	85.87	67.49	66.68	70.55	22.32	23.33	17.84
35-39	88.03	87.67	85.57	69.69	71.21	72.18	20.84	18.77	15.65
40-44	88.21	87.84	85.90	71.87	73.91	74.20	18.53	15.85	13.62
45-49	84.39	85.86	85.57	68.50	71.06	73.82	18.83	17.24	13.74
50-54	77.17	78.72	78.47	65.03	67.87	68.24	15.72	13.78	13.03
55-59	60.97	62.40	61.48	50.65	53.22	53.86	16.93	14.70	12.39
60-64	33.69	36.67	37.40	28.31	31.60	32.78	15.98	13.82	12.37
65-74	6.82	7.38	7.93	6.08	6.50	7.25	10.88	11.89	8.66
Country of birth (15 and over)									
Greece	50.50	50.51	49.53	40.31	41.51	42.08	20.18	17.82	15.03
Foreign	67.52	68.77	66.01	45.98	46.13	47.05	31.90	32.93	28.72
Region									
Eastern Macedonia & Thrace	49.21	50.62	48.95	41.13	42.12	40.79	16.43	16.79	16.66
Central Macedonia	50.70	50.68	49.25	39.58	40.40	39.76	21.94	20.28	19.25
Western Macedonia	50.24	50.24	45.87	36.14	36.63	36.98	28.06	27.08	19.39
Epirus	47.58	46.35	45.85	36.72	38.63	37.82	22.82	16.64	17.53
Thessaly	50.41	50.46	48.82	41.29	41.10	40.90	18.09	18.55	16.22
Ionian Islands	51.37	48.54	45.56	38.07	38.78	39.37	25.89	20.09	13.59
Western Greece	50.74	50.89	48.55	37.81	37.86	38.64	25.47	25.61	20.41
Central Greece	50.06	48.77	49.38	40.10	39.78	39.83	19.89	18.43	19.34
Attica	52.60	53.33	53.14	41.43	43.54	45.57	21.24	18.35	14.25
Peloponnese	52.80	51.38	51.84	44.39	44.48	46.26	15.92	13.44	10.76
Northern Aegean	54.24	54.99	53.66	40.96	43.86	44.93	24.48	20.24	16.27
Southern Aegean	55.51	55.93	46.10	42.14	40.91	39.63	24.08	26.85	14.02
Crete	53.40	53.82	52.97	42.76	44.92	43.94	19.92	16.54	17.05

Source: LFS, ELSTAT

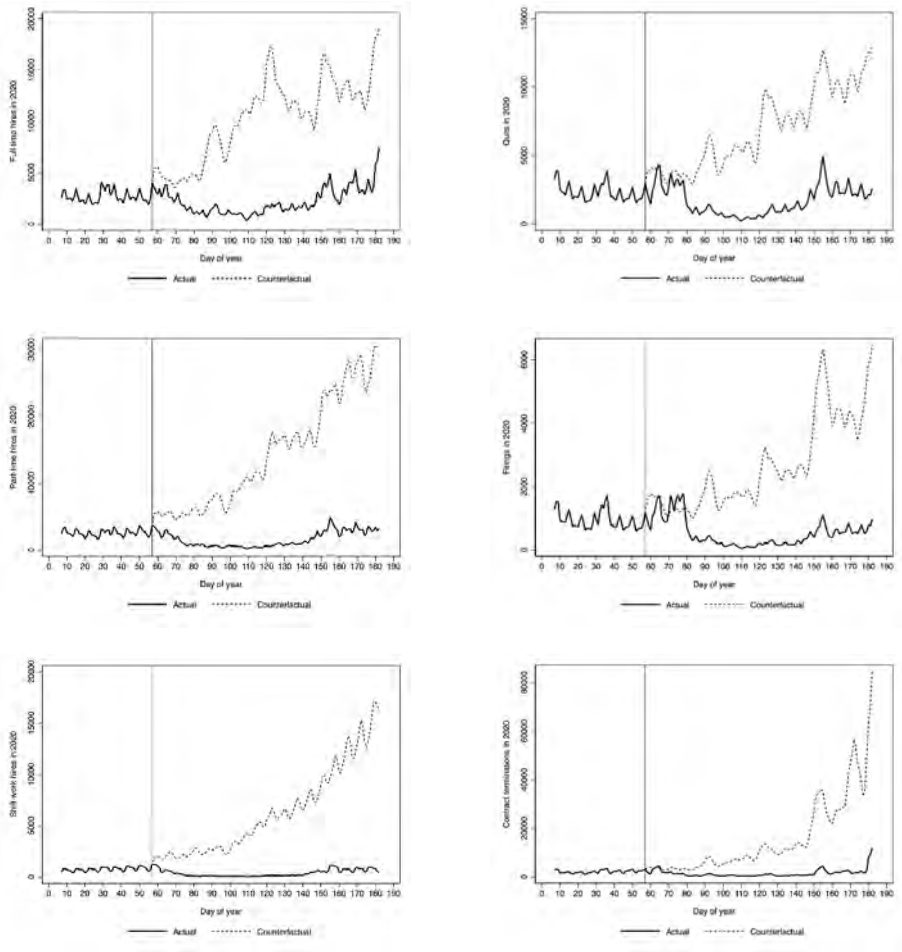
Covid Economics 43, 21 August 2020: 95-136

Covid Economics 43, 21 August 2020: 95-136

Figure A.1. Daily observed and counterfactual hires and separations by type before and after the pandemic onset

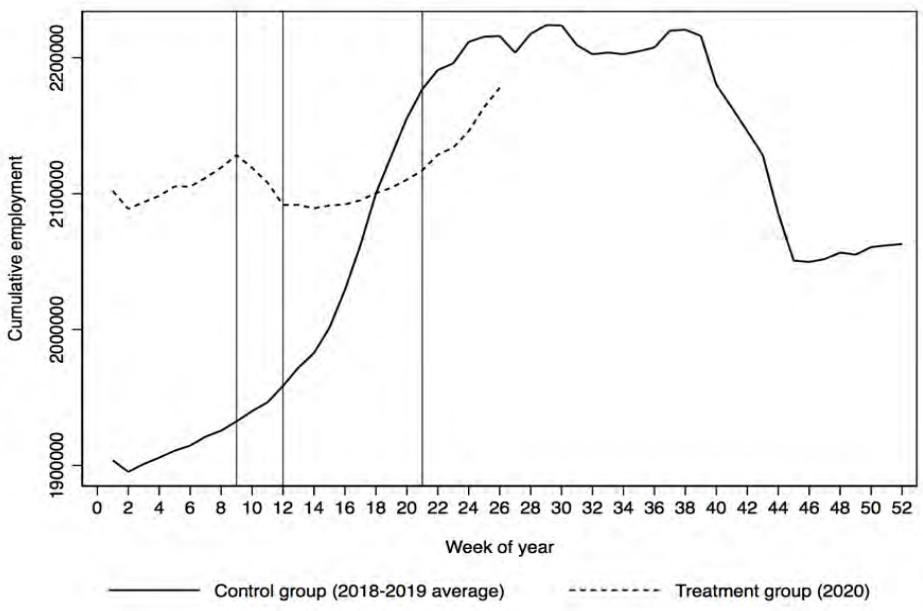
Number of new hires by type

Number of separations by type

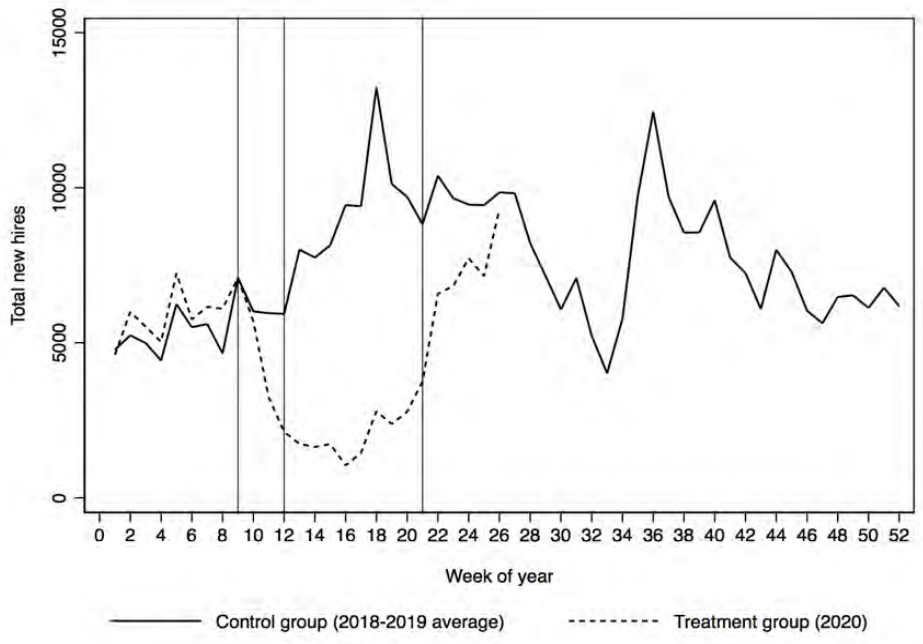


Source: ERGANI.
Notes: Vertical line is set at the pandemic onset (February 26, 2020).

Figure A.2. Trends in labor market flows for treatment and control groups.

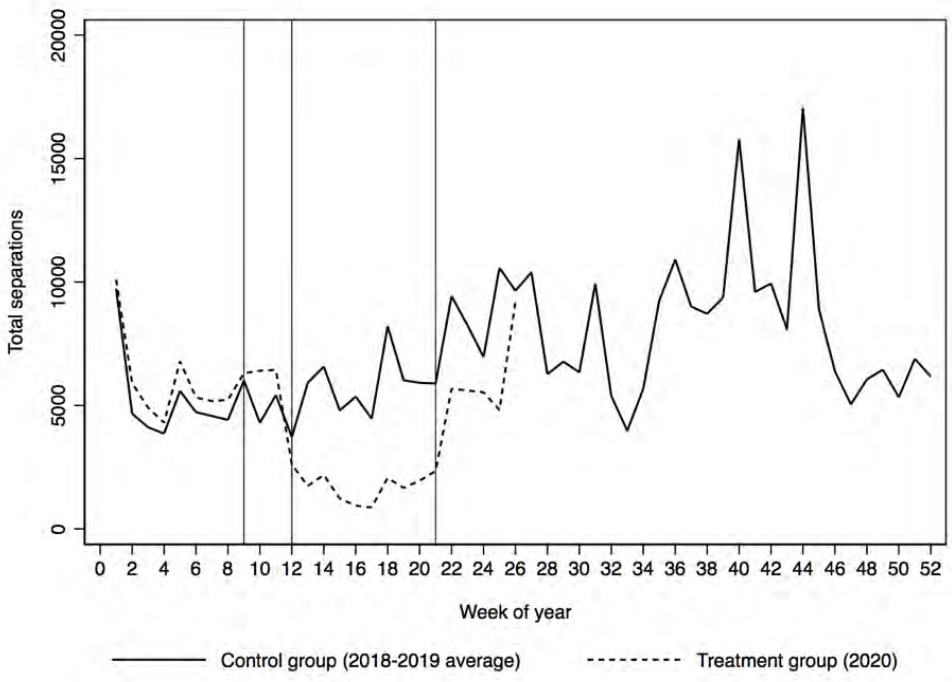


Panel A. Cumulative employment by week



Panel B. Total new hires

Covid Economics 43, 21 August 2020: 95-136



Panel C. Total separations

Source: ERGANI. Authors' calculations.
Notes: Vertical lines are set in weeks 9, 12 and 21 to indicate the pandemic onset, the layoffs restrictions and the end of our estimation sample (lockdown lifting), respectively.

For employment and new hires, there is a common trend up to week 9, and then there is a visible break in the trend for the treatment group (2020). For total separations, the trend for both groups is common up to week 11. Total weekly separations in 2020 are on a slightly higher level, compared to the control group, in weeks 9-11; however there is a sharp reduction after week 12 when the government restricted layoffs to protect the number of jobs. These observations hold even when looking within total new hires and total separations.

[1] The relevant ELSTAT press release can be found here.
[2] The Summer 2020 Interim forecast is here.
[3] Official estimates on UI claims are released 12 days after the end of the week they refer to.
[4] Apple data: www.apple.com/COVID19/mobility; Google data: www.google.com/COVID19/mobility/.
[5] See Table 2 here.
[6] See the relevant Centre of Planning and Economic Research report here.
[7] The list of eligible businesses was defined by virtue of Ministerial Decision by the Ministry of Finance (MoF), included business activities (sectors) per Code of Business Activity and was updated regularly during the health crisis. The most recent version of the list is provided on the MoF website.
[8] The eligibility criteria for the freelancers, self-employed and individual business owners who were entitled to such financial support were explained in Ministerial Decision 39162 ΕΞ 2020/ GG B' 1457/16.04.2020.
[9] Information regarding fiscal responses to the economic fallout from the coronavirus are provided by the Bruegel datasets; see here.

Covid Economics 43, 21 August 2020: 95-136

[10] It should be noted that the pandemic and mitigation measures affected the LFS data collection process, to some extent. From mid-March 2020 onwards, the LFS data collection switched from a blended style of personal and telephone interviews to solely telephone interviews. This decreased the response rate compared to previous months, especially in urban areas. The relevant ELSTAT press release is here.

[11] Due to data limitations in the ERGANI monthly reports, we extracted information on monthly labor market flows (January-April 2019 and January-April 2020) disaggregated by occupation (2-digit) and sector of economic activity (2-digit) from the National Institute of Labor and Human Resources (NILHR) website.

[12] Averaging over weeks for years before 2020 also smooths out any seasonal effects.

[13] The the two highest-ranked job portals according to Alexa (www.alexa.com) are kariera.gr and jobfind.gr. Alexa's traffic data take into account websites' unique visitors and page views. A recent assessment conducted by Cedefop also lists kariera.gr and jobfind.gr among the top private job portals in Greece based on different sources, including a study conducted by ELSTAT in 2017 and based on number of advertisements, number of monthly visitors and Alexa ranking (Cedefop, 2018).

[14] Disaggregated trends in the main labor market indicators for 2018Q1, 2019Q1, and 2020Q1 are in Appendix Table A.1.

[15] See ELSTAT Table 10 here.

[16] This is our identifying assumption of our interrupted time series analysis, i.e. Equation (1).

Covid Economics 43, 21 August 2020: 95-136

Days of Zipf and Covid? Looking for evidence of Zipf's Law in the infected Brazil[1]

Victor S. Comittit[2] and Claudio D. Shikida[3]

Date submitted: 11 August 2020; Date accepted: 18 August 2020

Does the ranking of Covid-19 cases by municipalities follow a Zipf's law (i.e. an estimated Pareto exponent of one)? This note tries to answer this question using daily data from Brazil for the Mar 30 - Aug 06 period. We used a Poisson Pseudo Maximum Likelihood (PPML) estimator for our estimates and the result is that the Pareto exponent of the ranking of Covid-19 cases is converging to the one obtained for the population ranking and seems to follow the pattern of the Zipf's Law. We did the same exercise using Italian regions' daily data which we use as a proxy to a long run equilibrium as the pandemic is apparently under control there. Contrary to Brazil, the Pareto exponent for the ranking for Covid-19 cases does not converge to the one estimated for the regular population. We try to advance some rationale for this contrasting behavior between them.

1 The authors thank Leonardo M. Monasterio (Enap), Erik A. Figueiredo (UFPB), Philipe Maciel and Elizabeth Anderson. Any errors that remain are our sole responsibility.

2 Professor, Federal Institute of Sudeste, Minas Gerais.

3 General Coordinator of Research, National School of Public Administration (Enap) and Professor of the Graduate Program in Organizations and Markets, Federal University of Pelotas.

Covid Economics 43, 21 August 2020: 137-149

Introduction

The year of 2020 has brought a heavy burden to the world: the Covid-19 pandemic. Luckily, the pandemic reached humanity in a period where the access to and the exchange of open data are almost costless for researchers.

Since the early days of the pandemic, many were collecting data about the disease. In this note, we make use of one of such databases in order to check if the confirmed cases of infected people by Covid-19 in Brazilian municipalities follow a specific and famous pattern known as Zipf's "law"[1]. What is interesting about this "law" is that, if you rank the data, the municipality with the most cases (the first one) will have twice the size of the second, three times the size of the third and so on.

As data have been collected daily and the number of observations increases with the spreading of the virus through Brazil' territory, we could estimate daily power-law equations. Among other things, the ranking of cases of Covid-19 by municipalities is the result of a mix of variables, including several public policies adopted by local, state and central governments. We are interested in the following question: how does the daily evolution of the estimated coefficient of the power law for the ranking of the number Covid-19's cases compare to the coefficient of the whole population (measured previously to the pandemic age)?

We do the same exercise to the Italian regions which we use as an (imperfect) *proxy* of a "long run equilibrium" as the pandemic seems to have been controlled by now in that country. The results are quite different, and we try to advance some rationale for this.

Is there a Zipf's law for the Covid-19 age's Brazil?

Zipf's law establishes an inverse relation between the size (denoted here by C) of a certain quantity and its rank (R) in a sample. Formally, this relation can be expressed as a Power-Law distribution, that is

$$R = AC^{-\alpha}, \qquad (1)$$

where A is a normalizing constant and α is called the Pareto exponent. Zipf's law is observed if the estimated value of α is equal to one. Suppose that Crepresents the population sizes of cities in a country; in this situation the Pareto exponent can be interpreted as a measure of population concentration: larger values of α are associated with countries of more equally distributed population. Conversely, lower values indicate that most of the population is concentrated in only

[1] See, for example, Adamic & Huberman (2002), Naldi (2003), Newman (2005), Saichev, Malevergne and Sornette (2010), Gomes-Lievano, Youn & Bettencourt (2012), Zaher & El-Maqd (2014), Aitchison Corradi and Latham (2016), Ausloos and Cerqueti (2016), Arshad, Hu and Ashraf (2017), Lestrade (2017).

a few cities. Power-Laws can be used to describe a wide range of phenomena such as the occurrence of words in a text, the magnitude of earthquakes or the number of citations received by scientific papers (Newman, 2005).

Several methods for estimating the Pareto Exponent can be found in the literature (see, for example, Zaher (2014)). A simple and popular approach is to define the following stochastic version of equation (1),

$$R_i = AC_i^{\alpha}\eta_i\,, \tag{2}$$

where η_i is an error term assumed to have a log-Normal distribution with $E(\eta_i\,|C_i\,) = 1$and $VAR(\eta_i|C_i)$=σ^2.This expression can be linearized by taking the logarithm on both sides. The resulting equation may be used to estimate the coefficient α via Ordinary Linear Regression (OLS) according to the model below,

$$log(R_i) = log(A) + \alpha log(C_i\,) + log(\eta_i). \tag{3}$$

Gabaix and Ioannides (2004) show that the OLS estimates for the Pareto Exponent are biased towards small samples. Gabaix and Ioannides (2011) propose to correct this bias by applying a shift of 0.5 to the 'rank' variable. Therefore, the correct model to be estimated will be:

$$log(R_i - 0.5) = log(A) + \alpha log(C_i\,) + log(\eta_i) \tag{4}$$

A major concern in the estimation of log-linear models is that, if the variance of the error term in Equation (2) is a function of the regressor C, this dependency will be carried over to the expected value of log(i) in Equation (4) -- thus violating one of the basic assumptions of linear regression models and leading to heteroscedasticity problems.

Santos and Tenreyro (2006) provide a lengthy discussion regarding this problem in the context of gravity models for international trade. They show by means of a Monte Carlo simulation that OLS estimates in constant-elasticity models are severely biased in the presence of heteroscedasticity. The authors suggest that these models should be estimated in their multiplicative forms and propose the use of a Poisson Pseudo Maximum Likelihood (PPML) estimator.

It is argued that this estimator is consistent even for non-count data as long as the conditional mean of the model is specified as $E[y_i|x\ \ ] = exp(x_i\beta)$, where y_i is the variable of interest and x is a set of explanatory variables. Santos and Tenreyro (2006) also present compelling evidence, in a variety of situations, that the proposed PPML estimator performs better than other methods in the estimation of log-linearized models.

Covid Economics 43, 21 August 2020: 137-149

Leitão et al. (2016) also address the problem of heterogeneous variability in the estimation of power-law relations and Aragón and Queiroz (2014) find evidence of heteroscedasticity in the estimation of Zipf's law for urban concentrations in Guatemala.[2] None of these authors, however, use the PPML estimator. So, to the best of our knowledge, this paper is the first to do so in the estimation of Zipf's Law.

Another common estimator for the Pareto exponent is the Maximum Likelihood Estimator (MLE) proposed by Hill (1975). This estimator, however, is the MLE only under the null hypothesis that the true distribution of the data is a Power-Law. Soo (2005) specifies the regression model with a quadratic term to test for non-linearities in the relation between rank and sizes. He concludes that, depending on the sign of the quadratic term, Hill's MLE can be negatively or positively biased. Munasinghe et al. (2020) arrives at the same conclusion using a Monte Carlo Simulation with samples generated from non-Pareto distributions. Since we have no evidence supporting the hypothesis that the distribution of Covid-19 cases is Pareto distributed, we chose to carry all the analysis in this work using the PPML estimator.

In terms of city population[3], according to our calculations, using the population of municipalities in Brazil, the Pareto exponent is 0.81[4], and using the OLS estimator with Gabaix correction, and 0.87 according to the PPML estimator. Would the estimation of α be the same for the number of confirmed cases of Covid-19? Using data from the Brasil.IO website, we obtained estimates of the Pareto exponent for the confirmed cases of Covid-19 in Brazil. The sample is composed by Brazilian municipalities and its size increases as the pandemics disseminate throughout the country. Our time period is from March 30 to August 06 (130 days).

Figure 1 below shows the evolution of the Pareto exponent according to the PPML estimator.

[2] Preliminary results of this paper using the OLS estimator with Gabaix correction also showed strong evidence of heteroscedasticity (according to the Breusch Pagan test) for all estimations of Zipf's law after April 6.

[3] We obtained data for 2019 estimated population of Brazilian municipalities from Brasil.IO. The original database is from IBGE, the Brazilian Institute of Geography and Statistics.

[4] Soo (2005) finds values for the Pareto exponent in Brazil of 1.13 for cities and 0.99 for urban agglomerations using data obtained from the 2000 census. Moura Jr & Ribeiro (2006) found higher coefficients (around 2).

Covid Economics 43, 21 August 2020: 137-149

Covid Economics 43, 21 August 2020: 137-149

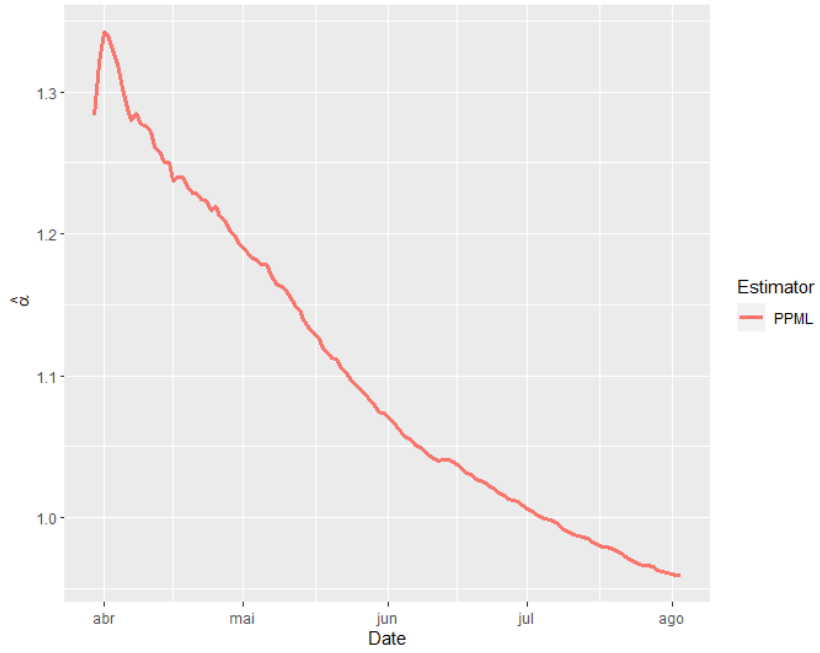


Figure 1 - Evolution of the Pareto Exponent during Brazil's Covid19 pandemic. Source: Brasil.IO (2020).

Our PPML estimates for the Pareto exponent show a decreasing behavior where last values seem to be near 0.95. Wang et al (2011) found that for H1N1 epidemics, also with daily data: as the epidemics spread, the evidence of a Zipf's law quickly vanishes. In contrast, our results show coefficients near one for approximately four months.

As the Brazilian crisis still persists, it is interesting to think about what the value of the Pareto exponent in a country would be if it had (apparently) passed the worst time of the pandemic. We could think of this country's $\hat{\alpha}$ at the final days of the sample as some type of "long-run equilibrium" of the pandemics. Italy is a good choice as it was one of the first countries affected. Contrary to China, there is virtually no concern about transparency with Italian data. Additionally, Italy and Brazil share some *cultural values*, which could be useful in our analysis[5].

Some institutional context is important here. In Brazil, first measures of social distancing were introduced by the federal government on March 20 (a short 7-day quarantine) with additional measures being introduced through April. In Italy, the state of emergency was issued on Jan 31, and the lockdown at the beginning of March (March 09), according to ACAPS (2020), but some

[5] See, for example, Ornelas (2020) for the importance of considering cultural aspects in the study of public policies that consider both economic and health aspects of the pandemics in the lockdown's design.

Covid Economics 43, 21 August 2020: 137-149

regions from the north of the country were closed starting February 02. Due to data availability, we used the R library *covid19italy* (Krispim (2020)) to obtain the evolution of the coefficient for the number of cases by region in Italy. The sample for this database is Feb 24 - Aug 06. The results are displayed in figure 2, below.

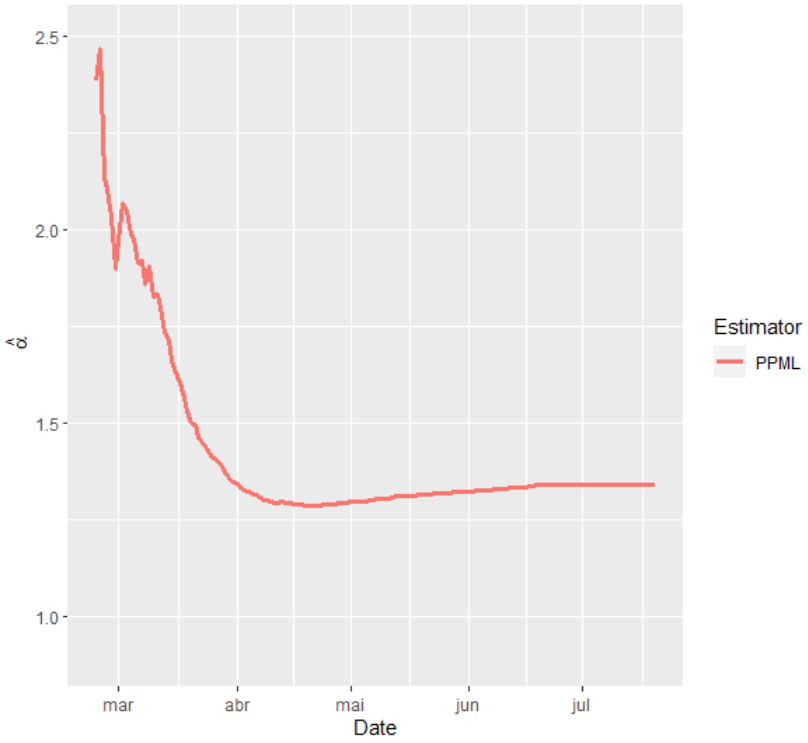


Figure 2 - Evolution of the Pareto Exponent during Italy's Covid19 pandemic. Source: Brasil.IO (2020).

The Pareto exponent for Covid-19 cases seems to stabilize around 1.34 at the end of the month[6], in a higher value than 0.79, the estimated Pareto exponent for the ranking of the population.

Why the different behavior between Italy and Brazil? We offer a speculative rationale for this. Maybe the Italian lockdown was able to minimize the spreading of the disease and this could have stopped the convergence of the Pareto exponents In the Brazilian case, the way the lockdown has been implemented was less uniform among all levels of government and the convergence between both Pareto exponents could be due to an unintended herd immunization.

[6] Our result seems to corroborate Alfano & Ercolano (2020). According to them: "(...) looking at the European case, its efficiency begins approximately 3 weeks after the lockdown and continues to reduce the number of COVID-19 infections up to 20 days later". [Alfano & Ercolano (2020), p.516]

Covid Economics 43, 21 August 2020: 137-149

Preliminary Conclusions

This note investigated the distribution of the Brazilian ranking for Covid-19 cases in a period of 130 days using a Poisson Pseudo Maximum Likelihood estimator. We found that the Brazilian Pareto exponent for the Covid-19 cases is decreasing and is near the theoretical value of the Pareto exponent for Zipf's law. Also, it seems to be near the one estimated for the 2019 population. The results go against Wang et al (2011), which found the Zipf's law is observed only at the early days of an epidemic scenario.

The second result we found is that, compared to a *proxy* for a "long run equilibrium", *i.e.*, an optimum low level of dissemination of the virus such as that found in Italy, Brazil seems to be far from it. Alfano & Ercolano (2020) found that the lag between the adoption of lockdown measures and its effect is approximately three weeks. Our results show that the, for Italy, values of the Pareto exponent stabilize around 1.34 according to this dynamic.

We speculate that this difference could be a result of the different success in the implementation of the lockdown's strategies by both countries. Maybe the Italian lockdown was able to minimize the spreading of the disease and this could have stopped the convergence of the Pareto exponents. However, in the Brazilian case, the convergence between both Pareto exponents could be due to an unintended herd immunization.

Bibliography

ACAPS. 2020. #COVID-19 Government Measures Dataset. https://www.acaps.org/covid19-government-measures-dataset (updated last on July 23).

Adamic, Lada A, and Bernardo A Huberman. 2002. "Zipf's Law and the Internet." Glottometrics 3: 143–50.

Aitchison, Laurence, Nicola Corradi, and Peter E. Latham. 2016. "Zipf's Law Arises Naturally When There Are Underlying, Unobserved Variables." PLoS Computational Biology 12(12): 1–32.

Alfano, V., & Ercolano, S. 2020. The Efficacy of Lockdown Against COVID 19 : A Cross Country Panel Analysis. Applied Health Economics and Health Policy, 18(4), 509–517. https://doi.org/10.1007/s40258-020-00596-3

Aragón, J.A.O, Queiroz, V.S. 2014. "The Zipf's law and the effects of free trade: The case of Guatemala," Economia, ANPEC - Associação Nacional dos Centros de Pós-Graduação em Economia [Brazilian Association of Graduate Programs in Economics], vol. 15(1), pages 82-99.

Arshad, Sidra, Shougeng Hu, and Badar Nadeem Ashraf. 2017. “Zipf’s Law and City Size Distribution: A Survey of the Literature and Future Research Agenda.” Physica A. https://doi.org/10.1016/j.physa.2017.10.005.

Ausloos, Marcel, and Roy Cerqueti. 2016. “A Universal Rank-Size Law.” PLoS ONE 11(11): 1–15.

Brasil.IO: boletins epidemiológicos da COVID-19 por município por dia, disponível em: https://brasil.io/dataset/covid19/ (last udpdated: July 31. Access in July 31).

Gabaix, Xavier, and Rustam Ibragimov. 2011. “Rank- 1/2: A Simple Way to Improve the Ols Estimation of Tail Exponents.” *Journal of Business and Economic Statistics* 29(1): 24–39.

Gabaix X, Ioannides Y. “The Evolution of city size distributions”. 2004. In: Handbook of Regional and Urban Economics. Vol. 4; Chapter 53. North-Holland ;. pp. 2341-2378.

Gomez-Lievano, A., Youn, H. J., & Bettencourt, L. M. A. (2012). “The statistics of urban scaling and their connection to Zipf’s law.” PLoS ONE, 7(7). https://doi.org/10.1371/journal.pone.0040393

Hill, B. M. 1975. "A simple approach to inference about the tail of a distribution", Ann. Statist.(3), 1163-1174.

Iorliam, Aamo, Anthony T S Ho, Santosh Tirunagari, and David Windridge. 2020. “Data Forensic Determination of the Accuracy of International COVID-19 Reporting: Using Zipf’s Law for Pandemic Investigation.” https://doi.org/10.20944/preprints202004.0531.v1.

Krispim, Rami. 2020. covid19italy: The 2019 Novel Coronavirus COVID-19 (2019-nCoV) Italy Dataset. https://cran.r-project.org/web/packages/covid19italy/index.html

Leitao, J.C. Miotto, J.M., Gerlach, M., Altmann, E.G.. 2016. “Is this scaling nonlinear?”, Royal Society Open Science 3, 150649.

Lestrade, Sander. 2017. “Unzipping Zipf ’s Law.” Plos One: 1–13.

Luckstead, Jeff, and Stephen Devadoss. 2014. “Do the World’s Largest Cities Follow Zipf’s and Gibrat’s Laws?” Economics Letters 125(2): 182–86. http://dx.doi.org/10.1016/j.econlet.2014.09.005.

Moura, N. J., & Ribeiro, M. B. (2006). “Zipf law for Brazilian cities”. Physica A: Statistical Mechanics and Its Applications, 367(December), 441–448. https://doi.org/10.1016/j.physa.2005.11.038.

Covid Economics 43, 21 August 2020: 137-149

Munasinghe, R., Kossinna, P., Jayasinghe, D. and Wijeratne, D. (2020). "Fast Tail Index Estimation for Power Law Distributions in R". https://arxiv.org/abs/2006.10308

Naldi, M. 2003. "Concentration Indices and Zipf's Law." Economics Letters 78(3): 329–34.

Newman, M.E.J. 2005. "Power laws, Pareto distributions and Zipf's law". Contemporary Physics 46(5): 323–351

Ornelas, E. (2020). Lockdown 101: Managing economic lockdowns in an epidemic. Revista Do Serviço Público, 1, 1-17. https://revista.enap.gov.br/index.php/RSP/article/view/4879

Saichev, Alexander, Yannick Malevergne, and Didier Sornette. 2010. Lecture Notes in Economics and Mathematical Systems Theory of Zipf's Law and Beyond. Springer Verlag.

Santos Silva, J.M.C. and Tenreyro, S. 2006. "The Log of Gravity", Review of Economics and Statistics 88(4):641-658

Soo, Kwook Tong. 2005. Zipf's Law for Cities: A Cross Country Investigation. Regional Science and Urban Economics, 35, p.239-263.

Wang, Lin et al. 2011. "Evolution of Scaling Emergence in Large-Scale Spatial Epidemic Spreading." PLoS ONE 6(7).

Zaher, H.M., El-Sheik, A.A. and El-Magd, N.A.T. 2014. "Estimation of Pareto Parameters Using a Fuzzy Least-Squares Method and Other Known Techniques with a Comparison". Journal of Advances in Mathematics and Computer Science. 4(14): 2067-2088.

Covid Economics 43, 21 August 2020: 137-149

Appendix - Brazilian Pareto's exponent (Gabaix and PPML)

Date	Number of municipalities	alpha_Gabaix	SE	Breusch Pagan test	alpha_PPML	SE
2020-03-30	359	0.7851***	0.0073	0.0510	1.2837***	0.0043
2020-03-31	404	0.7726***	0.0068	0.0004	1.3204***	0.0039
2020-04-01	440	0.7646***	0.0064	0.0010	1.3422***	0.0045
2020-04-02	484	0.7607***	0.0057	0.0173	1.3406***	0.0042
2020-04-03	537	0.7564***	0.0054	0.2717	1.3296***	0.0043
2020-04-04	576	0.7564***	0.0047	0.4294	1.32***	0.0038
2020-04-05	617	0.7466***	0.0046	0.0178	1.3021***	0.0033
2020-04-06	677	0.7497***	0.0044	0.0530	1.2885***	0.0031
2020-04-07	733	0.7406***	0.0041	0.0000	1.2802***	0.0030
2020-04-08	796	0.7347***	0.0039	0.0000	1.2849***	0.0030
2020-04-09	847	0.724***	0.0037	0.0000	1.2773***	0.0029
2020-04-10	891	0.7224***	0.0035	0.0000	1.2762***	0.0028
2020-04-11	938	0.7222***	0.0034	0.0000	1.2718***	0.0026
2020-04-12	977	0.7175***	0.0033	0.0000	1.2616***	0.0025
2020-04-13	1014	0.717***	0.0031	0.0000	1.2581***	0.0022
2020-04-14	1074	0.7133***	0.0030	0.0000	1.2504***	0.0021
2020-04-15	1145	0.7081***	0.0029	0.0000	1.2505***	0.0023
2020-04-16	1200	0.7026***	0.0028	0.0000	1.2373***	0.0023
2020-04-17	1260	0.6978***	0.0027	0.0000	1.2403***	0.0022
2020-04-18	1323	0.6935***	0.0027	0.0000	1.2399***	0.0023
2020-04-19	1389	0.6924***	0.0026	0.0000	1.2332***	0.0023
2020-04-20	1435	0.6926***	0.0025	0.0000	1.2289***	0.0022
2020-04-21	1476	0.6896***	0.0025	0.0000	1.2291***	0.0022
2020-04-22	1530	0.6878***	0.0024	0.0000	1.2245***	0.0023
2020-04-23	1596	0.6851***	0.0023	0.0000	1.2235***	0.0023
2020-04-24	1681	0.6844***	0.0023	0.0000	1.2168***	0.0022
2020-04-25	1745	0.6782***	0.0023	0.0000	1.2191***	0.0022
2020-04-26	1827	0.6729***	0.0022	0.0000	1.212***	0.0021
2020-04-27	1881	0.6691***	0.0022	0.0000	1.2099***	0.0022
2020-04-28	1953	0.6611***	0.0023	0.0000	1.2031***	0.0021
2020-04-29	2029	0.6552***	0.0023	0.0000	1.199***	0.0020
2020-04-30	2109	0.6494***	0.0023	0.0000	1.1937***	0.0020
2020-05-01	2165	0.6441***	0.0023	0.0000	1.1906***	0.0020

Covid Economics 43, 21 August 2020: 137-149

CEPR PRESS

2020-05-02	2239	0.6416***	0.0023	0.0000	1.1861***	0.0020
2020-05-03	2287	0.6387***	0.0023	0.0000	1.1825***	0.0020
2020-05-04	2345	0.6347***	0.0023	0.0000	1.181***	0.0019
2020-05-05	2440	0.6315***	0.0023	0.0000	1.1783***	0.0019
2020-05-06	2510	0.6274***	0.0023	0.0000	1.1786***	0.0019
2020-05-07	2589	0.6207***	0.0024	0.0000	1.1708***	0.0019
2020-05-08	2700	0.6152***	0.0024	0.0000	1.1644***	0.0018
2020-05-09	2764	0.6104***	0.0024	0.0000	1.1626***	0.0018
2020-05-10	2824	0.6082***	0.0024	0.0000	1.1614***	0.0018
2020-05-11	2877	0.6036***	0.0024	0.0000	1.1557***	0.0018
2020-05-12	2941	0.5994***	0.0025	0.0000	1.1503***	0.0018
2020-05-13	3015	0.5934***	0.0025	0.0000	1.1465***	0.0018
2020-05-14	3088	0.5865***	0.0025	0.0000	1.1397***	0.0018
2020-05-15	3189	0.5814***	0.0025	0.0000	1.1341***	0.0017
2020-05-16	3244	0.5758***	0.0026	0.0000	1.1308***	0.0017
2020-05-17	3271	0.5732***	0.0026	0.0000	1.1277***	0.0017
2020-05-18	3337	0.5674***	0.0025	0.0000	1.1193***	0.0018
2020-05-19	3410	0.563***	0.0026	0.0000	1.1168***	0.0017
2020-05-20	3491	0.5576***	0.0025	0.0000	1.113***	0.0018
2020-05-21	3557	0.5541***	0.0025	0.0000	1.1117***	0.0018
2020-05-22	3638	0.5491***	0.0025	0.0000	1.1053***	0.0018
2020-05-23	3712	0.5442***	0.0025	0.0000	1.1027***	0.0018
2020-05-24	3744	0.5411***	0.0025	0.0000	1.0978***	0.0018
2020-05-25	3773	0.5389***	0.0025	0.0000	1.0942***	0.0018
2020-05-26	3831	0.5352***	0.0025	0.0000	1.0909***	0.0018
2020-05-27	3876	0.5307***	0.0025	0.0000	1.0873***	0.0018
2020-05-28	3942	0.5262***	0.0025	0.0000	1.0838***	0.0018
2020-05-29	4011	0.523***	0.0025	0.0000	1.081***	0.0017
2020-05-30	4071	0.5152***	0.0025	0.0000	1.0748***	0.0018
2020-05-31	4110	0.5149***	0.0025	0.0000	1.0738***	0.0017
2020-06-01	4130	0.5136***	0.0025	0.0000	1.071***	0.0018
2020-06-02	4178	0.5105***	0.0025	0.0000	1.0682***	0.0017
2020-06-03	4222	0.5075***	0.0025	0.0000	1.0646***	0.0017
2020-06-04	4271	0.5042***	0.0025	0.0000	1.0601***	0.0017
2020-06-05	4336	0.5015***	0.0025	0.0000	1.0563***	0.0018
2020-06-06	4380	0.4997***	0.0026	0.0000	1.0545***	0.0017
2020-06-07	4415	0.4982***	0.0026	0.0000	1.0515***	0.0018

Covid Economics 43, 21 August 2020: 137-149

CEPR PRESS

2020-06-08	4434	0.4966***	0.0025	0.0000	1.0498***	0.0018
2020-06-09	4457	0.4953***	0.0026	0.0000	1.0468***	0.0018
2020-06-10	4501	0.4933***	0.0026	0.0000	1.0438***	0.0018
2020-06-11	4546	0.4919***	0.0026	0.0000	1.0418***	0.0018
2020-06-12	4563	0.4905***	0.0026	0.0000	1.0404***	0.0018
2020-06-13	4590	0.4889***	0.0026	0.0000	1.0406***	0.0017
2020-06-14	4607	0.4888***	0.0026	0.0000	1.0409***	0.0017
2020-06-15	4632	0.4856***	0.0026	0.0000	1.0394***	0.0017
2020-06-16	4674	0.4839***	0.0026	0.0000	1.0377***	0.0017
2020-06-17	4710	0.4824***	0.0026	0.0000	1.035***	0.0017
2020-06-18	4746	0.4814***	0.0026	0.0000	1.0315***	0.0017
2020-06-19	4785	0.4786***	0.0026	0.0000	1.0301***	0.0017
2020-06-20	4832	0.4776***	0.0026	0.0000	1.0272***	0.0017
2020-06-21	4848	0.477***	0.0026	0.0000	1.026***	0.0017
2020-06-22	4867	0.4765***	0.0026	0.0000	1.0246***	0.0017
2020-06-23	4897	0.476***	0.0026	0.0000	1.0222***	0.0017
2020-06-24	4918	0.4751***	0.0026	0.0000	1.0199***	0.0017
2020-06-25	4940	0.4757***	0.0026	0.0000	1.0179***	0.0017
2020-06-26	4988	0.4774***	0.0026	0.0000	1.0158***	0.0017
2020-06-27	5036	0.4768***	0.0027	0.0000	1.0135***	0.0017
2020-06-28	5059	0.4769***	0.0027	0.0000	1.0122***	0.0017
2020-06-29	5074	0.4767***	0.0027	0.0000	1.0113***	0.0017
2020-06-30	5104	0.4753***	0.0027	0.0000	1.0091***	0.0017
2020-07-01	5124	0.4754***	0.0027	0.0000	1.0062***	0.0017
2020-07-02	5141	0.4752***	0.0027	0.0000	1.0048***	0.0017
2020-07-03	5157	0.4765***	0.0027	0.0000	1.0023***	0.0017
2020-07-04	5180	0.4781***	0.0027	0.0000	1.0006***	0.0017
2020-07-05	5191	0.4782***	0.0027	0.0000	0.9997***	0.0016
2020-07-06	5210	0.4771***	0.0027	0.0000	0.9988***	0.0016
2020-07-07	5228	0.4775***	0.0027	0.0000	0.9968***	0.0016
2020-07-08	5246	0.4779***	0.0028	0.0000	0.9943***	0.0016
2020-07-09	5256	0.4788***	0.0028	0.0000	0.9917***	0.0016
2020-07-10	5273	0.4791***	0.0028	0.0000	0.9896***	0.0016
2020-07-11	5281	0.4793***	0.0028	0.0000	0.9878***	0.0016
2020-07-12	5288	0.4804***	0.0028	0.0000	0.9868***	0.0016
2020-07-13	5294	0.4819***	0.0028	0.0000	0.9861***	0.0016
2020-07-14	5306	0.4832***	0.0028	0.0000	0.9853***	0.0016

Covid Economics 43, 21 August 2020: 137-149

CEPR PRESS

2020-07-15	5319	0.485***	0.0028	0.0000	0.9832***	0.0016
2020-07-16	5335	0.4851***	0.0028	0.0000	0.981***	0.0016
2020-07-17	5338	0.4877***	0.0029	0.0000	0.9793***	0.0016
2020-07-18	5340	0.4897***	0.0029	0.0000	0.9788***	0.0016
2020-07-19	5349	0.4889***	0.0029	0.0000	0.9782***	0.0016
2020-07-20	5358	0.4891***	0.0029	0.0000	0.977***	0.0016
2020-07-21	5373	0.491***	0.0029	0.0000	0.9747***	0.0016
2020-07-22	5382	0.4929***	0.0029	0.0000	0.9723***	0.0016
2020-07-23	5393	0.4935***	0.0029	0.0000	0.9704***	0.0016
2020-07-24	5404	0.4931***	0.0029	0.0000	0.9686***	0.0016
2020-07-25	5413	0.4947***	0.0029	0.0000	0.9673***	0.0016
2020-07-26	5417	0.4946***	0.0029	0.0000	0.9666***	0.0016
2020-07-27	5419	0.4958***	0.0029	0.0000	0.9662***	0.0016
2020-07-28	5430	0.4952***	0.0030	0.0000	0.9649***	0.0016
2020-07-29	5436	0.4968***	0.0030	0.0000	0.9628***	0.0016
2020-07-30	5441	0.4976***	0.0030	0.0000	0.9617***	0.0016
2020-07-31	5447	0.4976***	0.0030	0.0000	0.961***	0.0016
2020-08-01	5454	0.4983***	0.0030	0.0000	0.9597***	0.0016
2020-08-02	5457	0.4979***	0.0030	0.0000	0.9592***	0.0016
2020-08-03	5468	0.4984***	0.0030	0.0000	0.9584***	0.0016
2020-08-04	5466	0.5005***	0.0030	0.0000	0.9568***	0.0016
2020-08-05	5469	0.501***	0.0030	0.0000	0.9559***	0.0016
2020-08-06	5475	0.5023***	0.0030	0.0000	0.955***	0.0016

Covid Economics 43, 21 August 2020: 137-149