Investigating country-level activity and mobility patterns and their interdependencies on COVID-19 outcomes

Atanas Apostolov[a], Jimi B. Oke[a]

[a] *Department of Civil and Environmental Engineering, University of Massachusetts Amherst, MA 01003, United States*

Abstract

In this paper we examine the effects of COVID-19 on mobility, segmented by transportation type, as well as social activity such as retail and recreation, workplaces and residential, and their interdependencies. Using time series data from 63 countries across five continents, we investigate patterns in activity and mobility trends from January through July. Our clustering approach yields four clusters of countries with similar behavior. We plan to estimate vector autoregression models to analyze the relationship between activity changes, mobility changes and COVID-19 growth. We will also fit cluster-level models and a global model which includes exogenous variables (such as interventions). Our findings yield insights into how various activities have been impacted by COVID-19 and vice-versa. We expect that our model outcomes will guide policymakers to adopt appropriate measures to mitigate and safely recover from the ongoing pandemic, as well as future ones.

Keywords: COVID-19, mobility, vector autoregression

1 Introduction

This section sets up the motivation and significance of the work presented in this paper. Key questions:

1. *What is the ongoing and potential future impact of COVID-19 across various countries, particularly on mobility and activity patterns? What are the implications (known and unknown)?*

2. *Briefly what has been done in terms of modeling mobility impacts?*

3. *Where are the current gaps in our knowledge?*

4. *(Final paragraph) What is the summary of this paper's contributions? What is the expected significance of our results?*

COVID-19 has had a severe impact on mobility and social activities since initial reports of the disease emerged in late 2019. By May 2020 most countries had implemented social distancing measures and imposed strict travel restrictions to curtail the spread of the virus both within their borders and internationally. According to [1], the international spread of the disease can be partially explained by risk perception, mobility behavior and social media influences using a global vector autoregression (GVAR) modeling framework[2]. [3] confirm that human mobility is key to understanding the transmission pathways. Using a mobility dataset of over 500,000 flights and over 100 million passengers, the authors develop a dynamic network model using time-dependent border restrictions as exogenous variables. [4] studied the patterns in 6 activity responses across 108 countries but this was an exploratory effort in comparison with government policies and it included no modeling attempt to explain COVID-19 outcomes.

Beyond these global studies, several city and country-specific simulations and models have been estimated to explain how activities, mobility (particularly transit) impact COVID-19 outcomes. [5] used an activity-based epidemiological framework to study how COVID-19 is propagated across various daily activities in car-dependent cities typically found in the US and Canada. They quantified the effect of the transit network on COVID-19 spread and demonstrated the importance of work and home activities in the early and latter stages of the epidemic. [6] developed a generalized linear model to explain COVID-19 growth based on overall mobility changes at the county level in the US. Inter-county movements were factored in but impacts by mode or activity were not analyzed. Given China's and then Italy's prominence in the earlier stages of the pandemic, several studies have examined their mobility patterns [7, 1, 8, 9]. While these have been important for country-wide epidemiological studies, they have not provided knowledge on activity-based impacts.

Therefore, there remains an urgent need to understand the interdependencies between COVID-19 outcomes and human mobility and activity patterns. Statistically significant models can yield insights to enable policymakers contain and recover from the ongoing pandemic. Furthermore, these models can guide future decisions in containing the spread of future epidemics. In this paper, we conduct a global scale country-level analysis of COVID-19 infections, mobility and activity patterns. First, we cluster the countries to discover patterns in mobility and activity changes. Second, we estimate cluster-representative models using the vector autoregression (VAR) framework, allowing for time lags across all endogenous variables, and thus including dynamic interdependencies among them. These models explain which activities impact COVID-19 outcomes and vice versa, providing potential pathways for understanding behavioral responses to the spread of the pandemic and a greater understanding of the disparities and similarities across nationwide outcomes.

1.1 Significance

1.2 Related work

Here, 3-4 paragraphs summarizing efforts to analyze activity-mobility patterns (i) historically, (ii) in context of COVID-19, and then (iii) brief history and application of VAR modeling.

2 Data

We use transportation data provided by Apple[1] along with Community Mobility Reports compiled by Google[2]. The new confirmed cases global data set is provided by the Johns Hopkins University COVID-19 Data initiative[3]. Apple's data set focuses on transportation types (walking, driving, transit) whereas Google tracks movement trends by region and segments them into categories such as residential, workplace and transit. Our integrated data set captures changes in requests for directions from January 22 to July 14 as well as aggregated location data for a variety of categorized places. The data shows us the change in visitors from February 15 to July

[1] `https://www.apple.com/covid19/mobility`

[2] `https://www.google.com/covid19/mobility/`

[3] `https://github.com/CSSEGISandData/COVID-19`

12 compared to a baseline of days (January 3 - February 6, 2020). By tracking user requests of their Maps application, Apple contributes a snapshot of incidental trips as opposed to actual foot traffic which is the approach taken by Google for their Community Mobility Reports. The sources are summarized in Table 1 and the variables are described in Table 2. The integrated dataset (along with the code and plots generated for this paper) is publicly available on Github.[4]

Table 1: Summary of data sources

Source	**Overview**
Johns Hopkins University	New COVID cases and COVID-related deaths (global time series summary)
Google Community Mobility Reports	Change in traffic compared to a baseline day (median value over period from Jan 3 to Feb 6, 2020)
World Data Indicators	Economic indicators (growth, GDP)
Government Measures	ACAPS Government Measures Dataset

Table 2: Summary of variables

Variable	**Description**
cases	New COVID cases tracked by JHU
deaths	COVID-related deaths tracked by JHU
grocery	Grocery and pharmacy activity percent change from baseline
parks	Parks percent change from baseline
transit	Transit stations activity percent change from baseline
work	Workplaces activity percent change from baseline
home	Residential activity percent change from baseline
retail	Retail and recreation activity percent change from baseline
pop	Population
gdppc	GDP Per Capita
growth	Growth
pop_den	Population Density
pop_urb	Urban Population
pop0_14	Population Ages 0 to 14
workpop	Working population (Ages 15-64)
pub_health	Government Measures - Public Health
soc_econ	Government Measures - Socio-economic
soc_dist	Government Measures - Social Distancing
mov_rest	Government Measures - Movement Restrictions
lockdown	Government Measures - Lockdown
human	Government Measures - Humanitarian Exemption

2.1 Aggregation Procedure

Four of the data sets had to be aggregated in order to standardize the GVAR model inputs.

[4] `https://github.com/narslab/covid-analysis/tree/master/mobility`

1. New Cases and Deaths. Both new cases and death statistics are provided by JHU and follow the same format. First we sum up province data into a country total by checking each row if the 'province/state' field is NA. If that is not the case, we add all rows for the entire series. Then we incorporate the 'ISO' based on 3-digit alpha codes. Finally, we apply the melt function to the newly created DataFrame and use 'pivot table', with iso and activity set as index.

2. Google Activity. We retain a subset of the original data containing national stats without the subregional breakdown. Afterwards, we rename activity types for conventional purposes (i.e. retail and recreation as retail). Again, we apply melt and 'pivot table' with ISO and activity set as index. 2-digit country codes are replaced with the 3-digit ISO convention.

3. Government Interventions. We create a new DataFrame with 'MultiIndex' for Category and ISO. We ensure the ISO code conforms to the 3-digit convention.

3 Methods

First, we investigate patterns across the activity and mobility variables by clustering the countries across these nine dimensions. We use dynamic time warping and hierarchical agglomerative clustering to obtain the clusters. These yield preliminary insights into the trends at the country level. Finally, we intend to estimate vector autoregression models at the country and at the cluster level in order to quantify how changes in activity and mobility trends influence COVID-19 outcomes. In particular, we expect results as well on interdependencies between the mobility and activity variables.

3.1 Dynamic time warping and clustering

Not sure if we need to perform any exploratory analysis in this manuscript, but these methods might be useful either initially or in analyzing model results The dynamic time warping (DTW) algorithm [10] computes the optimal distance between pairwise time series. Any hierarchical clustering procedure requires a dissimilarity matrix as an input, which DTW provides. First, we compute the Euclidean distance matrix for each country across the dimensions of the endogenous variables. Second, we apply DTW to find the optimal matching between the countries based on their pairwise distance matrices across the dimensions observed. The symmetric dissimilarity matrix is then used an input to the clustering algorithm.

We use the Ward method [11], which is a hierarchical agglomerative clustering (HAC) approach, to group the countries. The dendogram is shown in Figure 1.

3.2 Dimensionality reduction of deterministic variables

Perhaps PCA or EFA.

3.3 Global vector autoregression model with time-varying weights

The global VAR model [2] is given by

$$x_{i,t} = \sum_{l=1}^{p_i} \Phi_{il} x_{i,t-l} + \Lambda_{i0} x^*_{i,t} + \sum_{l=1}^{q_i} \Lambda_{il} x^*_{i,t-l} + \varepsilon_{i,t} \tag{1}$$

where $x_{i,t}$ is a vector of $k_i \times 1$ endogenous variables and $i = 1, 2, \ldots, N$ is the number of buildings in the system. The matrix Φ_{il} has dimensions $k_i \times k_i$, while the matrices Λ_{io} and Λ_{i1} both have the dimensions $k_i \times k_i^*$. Each country i has k_i endogenous variables and k_i^* foreign (weakly exogenous) variables. Φ_{il} is a coefficient matrix of the cointegrated endogenous variables. Λ_{i0} is a coefficient matrix for the contemporanoues weakly exogenous variable $x_{i,t}^*$. Λ_{il} is the coefficient matrix of the q_i lagged foreign variables $x_{i,t-l}^*$, where

$$x_{i,t}^* = \overline{W}_i^{'} x_t. \tag{2}$$

The matrix $\overline{W}_i^{'}$ is the $k^* \times N$ matrix of building-specific weights. Stacking the domestic and foreign variables, we obtain

$$z_{i,t} = \left[x_{i,t}^{'}, x_{i,t}^{*'}\right]' = W_i x_t \tag{3}$$

where $W_i = E_i \overline{W}_i$ and E_i is a selection matrix We can then compactly write the model as

$$G_0 x_t = \sum_{l=1}^{p} G_l x_{t-l} + \varepsilon_t \tag{4}$$

where

$$G_l = \begin{bmatrix} A_{1,l} W_1 \\ A_{2,l} W_2 \\ \vdots \\ A_{N,l} W_N \end{bmatrix} \tag{5}$$

and $A_{i,l} = [\Phi_{1l} \Lambda_{2l}]$.

The global model can then be compactly written as

$$x_t = \sum_{l=1}^{p} G_0^{-1} G_l x_{t-l} + G_0^{-1} u_t \tag{6}$$

where x_t x_{t-l} and u_t are stacked vectors of length kN.

4 Results

Our clustering procedure based on the activity and mobility trends of 63 countries optimally yields four clusters (Figure 1). Clusters 1 and 2 are much smaller than the other two, with 9 and 11 members, respectively.

We perform a time series transformation of the community mobility data. Google clusters their baseline trends around zero which affects our log differencing. Therefore, we shift all of the movement indicators such as workplace, residential and transit by 100 which is also consistent with Apple's approach of keeping track of requests for directions.

There is no strong continental dominance in any one of the clusters. This indicates at least there is geographic heterogeneity in the distribution of these trends as observed by Google location and Apple navigation-based data. We note that these variables are only a partial observation of the true activity/mobility trends across these dimensions, given that only a select proportion in each country uses Google devices with location data turned on, or Apple devices for navigation. Further, these samples may be more representative of urban or suburban populations in some countries than in others. Nevertheless, some useful insights may still be gained.

We then plot the trends of activity changes as observed from Google location data, grouping these by cluster (**??**). Trends are shown by means of a generalized linear model estimated for each

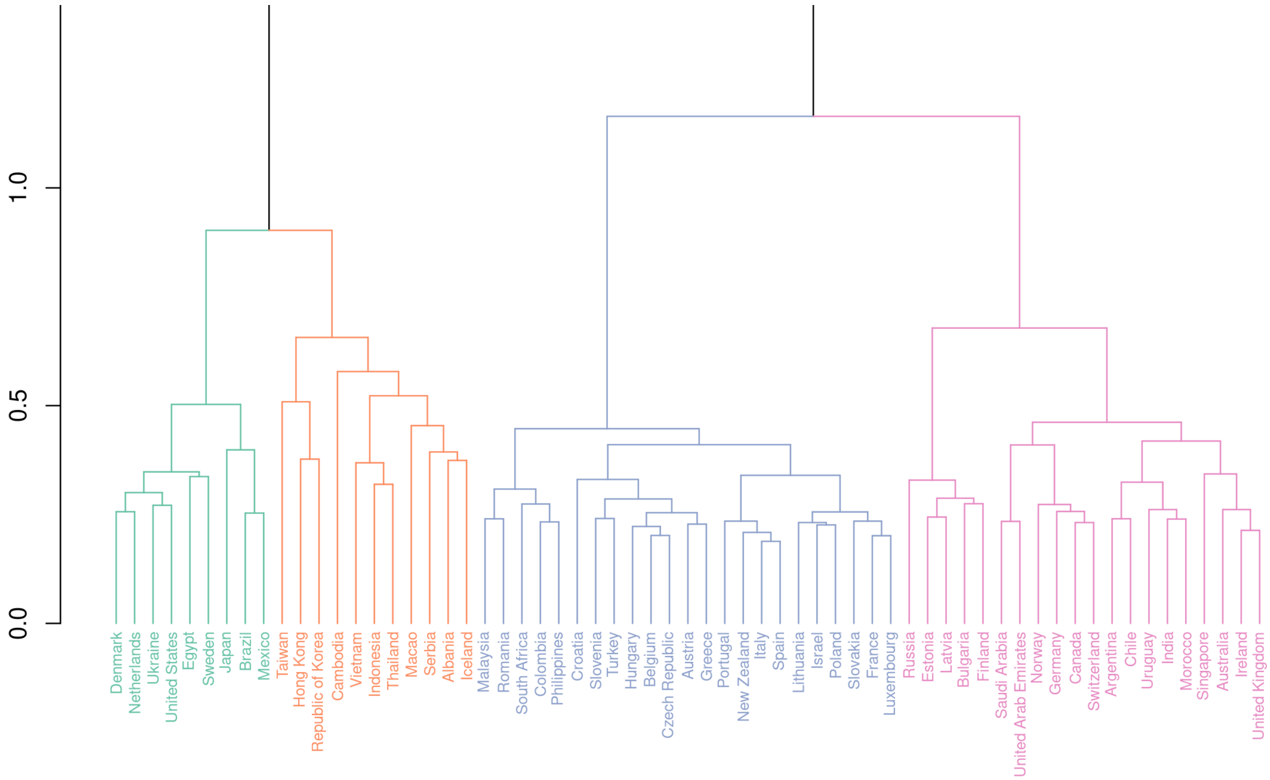

Figure 1: Dendrogram indicating four clusters of countries based on nine mobility and activity indicators

country. Except in Cluster 1, which consists notably of South Korea, Hong Kong, Taiwan and Vietnam (countries with some of the best COVID-19 outcomes), there was a significant downturn in work activities from mid-February to early April. This decrease is as great as 75 percentage points in Clusters 2 and 3. However, in Cluster 4 (e.g. United States, Brazil, Japan), it only reaches 50 percentage points. The change in work activities can be seen as a partial indicator of lockdown policies or behavioral responses to the pandemic in each of the countries. Transit activities similarly reduced in Clusters 2 and 3, but not as severely in Cluster 4. Notably, in Cluster 1, transit stop activities dipped the lowest in South Korea compared to the other countries. This outcome can be readily understood given that South Korea focused heavily on contact tracing to mitigate the spread of COVID-19 [12, 13].

Grocery shopping experienced a downturn in most countries. However, Clusters 1 and 4 show the smallest disruptions (a decline of about 20 percentage points). The same trend is observed for retail activities. However, in Cluster 4, this category had a more severe downturn (as many as 60 percentage points from the baseline), compared to grocery shopping. This could be explained by policies keeping grocery stores open as essential services, while other retail outlets were shut. Also, home activities (people staying in their places of residence) increased across all countries, peaking in April and then declining thereafter. The change from baseline was greatest in Cluster 3 (generally over 50 percentage points), indicating that countries in this cluster stayed at home the most. Furthermore, the peak for home activities occurs the latest in Cluster 3.

There are no discernible cluster-specific patterns for park activities for Clusters 1, 2 and 4 (as trends vary by country). However, in Cluster 3, many countries observed a decline of up to 75 percentage points through April, followed by a large uptick.

We show the daily confirmed COVID-19 cases at the country level by cluster in Figure 2 on a log-y scale. Most of the countries in Cluster 1 have experienced a decrease in COVID-19 cases. However, Serbia and Indonesia are outliers. Given that the countries in this cluster had the least disruption in activities compared to other clusters, other exogenous factors are probably responsible for the favorable COVID-19 outcomes. Clusters 2 and 3 were harder hit in the early weeks of the year, but indicate a downturn in COVID-19 infections from late March and early April. These clusters had the most severe decreases in work and transit activities (and highest increases in home activities) from the baseline. Clearly, there are few exceptions in each of these clusters, such as India in Cluster 2 and Colombia in Cluster 3. Finally, in Cluster 4, most of the countries have been on an upward trajectory in recent weeks (except for Denmark). In these countries, there is evidence of lockdowns being lifted too soon and perhaps less strict public health policies than elsewhere. In terms of activities, this is one cluster where home activities showed the least disruption, which is an indicator of the aforementioned trend.

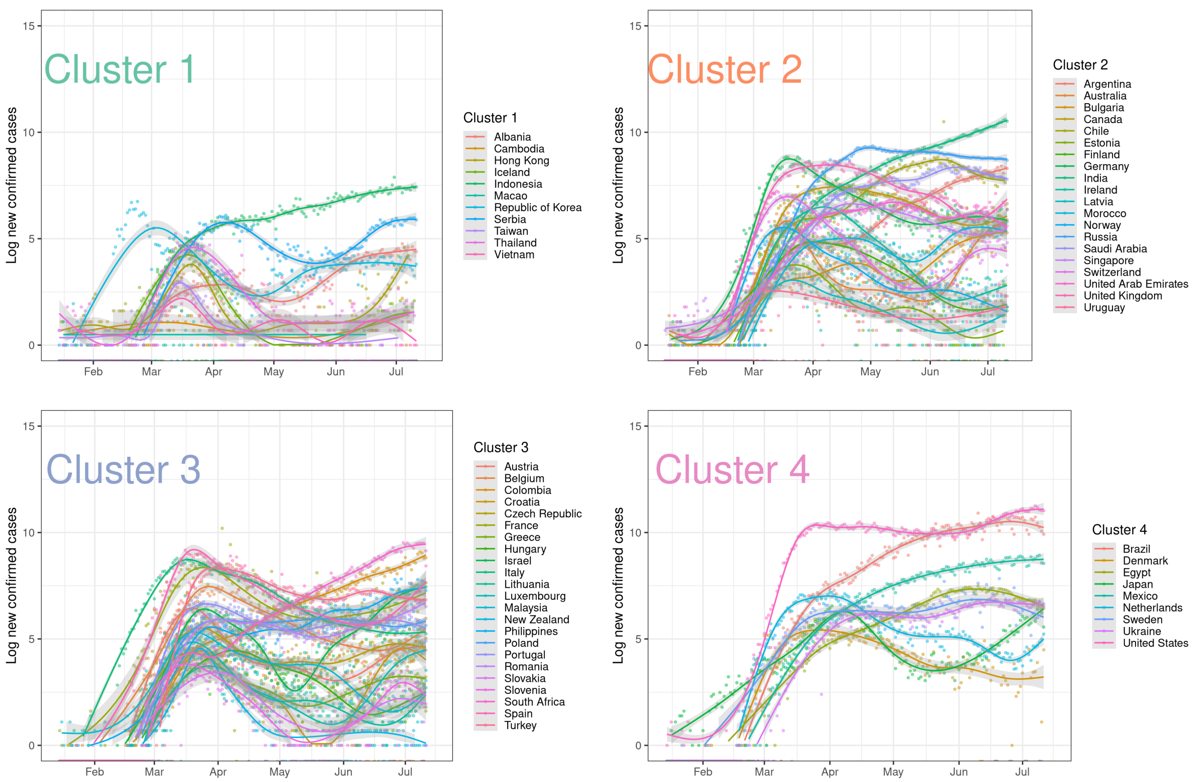

Figure 2: Daily confirmed COVID-19 cases for each of the clusters (log y-axis); Johns Hopkins dataset.

4.1 Expected results

We plan to estimate country-level vector autoregression models (VARs), cluster-level panel VARs and a global VAR. These models will enable us to explain how COVID-19 outcomes were impacted by changes across these activities. They would also allow us to analyze how changes in one dimension would affect the other (using impulse response functions). Other variables that can explain COVID-19 impacts (particularly interventions) will be included in these models as strictly exogenous variables. Air passenger travel flows will also be incorporated as weights on weakly exogenous (foreign) variables in the global VAR model. These would allow the inclusion of the effects of observations in other countries across these dimensions.

5 Discussion

6 Conclusion

We cluster the countries in our sample into four categories based on similar activity and mobility trends. Future work will focus on estimating VAR models for each of the countries, adding exogenous and weakly exogenous variables such as interventions and restrictions (social distancing orders, travel bans, etc.), as well as deterministic ones (country characteristics). Subsequently, cluster-level VARs and a global VAR will be estimated. These models will provide insights into the processes governing the changes in the endogenous variables (COVID-19 outcomes, activity and mobility trends), as well as potential forecasting capabilities. The insights gained from these models would enable policy and decision makers plan efficiently for recovery from the current pandemic, as well as better prepare for future ones, by tailoring interventions relevant to the behavioral profiles of their respective countries.

References

[1] G. Bonaccorsi, F. Pierri, M. Cinelli, A. Flori, A. Galeazzi, F. Porcelli, A. L. Schmidt, C. M. Valensise, A. Scala, W. Quattrociocchi, F. Pammolli, Economic and social consequences of human mobility restrictions under COVID-19, Proceedings of the National Academy of Sciences 117 (27) (2020) 15530–15535. `doi:10.1073/pnas.2007658117`.

[2] S. Dees, F. di Mauro, M. H. Pesaran, L. V. Smith, Exploring the international linkages of the euro area: A global VAR analysis, Journal of Applied Econometrics 22 (1) (2007) 1–38. `doi:10.1002/jae.932`.

[3] C. Zhang, L.-X. Qian, J.-Q. Hu, Pathways of the COVID-19 Pandemic with Human Mobility across Countries, medRxiv (2020) 2020.05.21.20108589`doi:10.1101/2020.05.21.20108589`.

[4] G. McKenzie, B. Adams, A country comparison of place-based activity response to COVID-19 policies, arXiv:2005.08738 [cs] (May 2020). `arXiv:2005.08738`.

[5] N. Kumar, J. B. Oke, B.-h. Nahmias-Biran, Activity-based contact network scaling and epidemic propagation in metropolitan areas, arXiv:2006.06039 [physics, q-bio] (Jun. 2020). `arXiv:2006.06039`.

[6] H. S. Badr, H. Du, M. Marshall, E. Dong, M. M. Squire, L. M. Gardner, Association between mobility patterns and COVID-19 transmission in the USA: A mathematical modelling study, The Lancet Infectious Diseases 0 (0) (Jul. 2020). `doi:10.1016/S1473-3099(20)30553-3`.

[7] M. U. G. Kraemer, C.-H. Yang, B. Gutierrez, C.-H. Wu, B. Klein, D. M. Pigott, O. C.-. D. W. Group†, L. du Plessis, N. R. Faria, R. Li, W. P. Hanage, J. S. Brownstein, M. Layan, A. Vespignani, H. Tian, C. Dye, O. G. Pybus, S. V. Scarpino, The effect of human mobility and control measures on the COVID-19 epidemic in China, Science 368 (6490) (2020) 493–497. `doi:10.1126/science.abb4218`.

[8] E. Pepe, P. Bajardi, L. Gauvin, F. Privitera, B. Lake, C. Cattuto, M. Tizzoni, COVID-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown, Scientific Data 7 (1) (2020) 230. `doi:10.1038/s41597-020-00575-2`.

[9] A. Cartenì, L. Di Francesco, M. Martino, How mobility habits influenced the spread of the COVID-19 pandemic: Results from the Italian case study, Science of The Total Environment 741 (2020) 140489. doi:10.1016/j.scitotenv.2020.140489.

[10] T. Giorgino, Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package, Journal of Statistical Software 031 (i07) (2009).

[11] F. Murtagh, P. Legendre, Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?, Journal of Classification 31 (3) (2014) 274–295. doi:10.1007/s00357-014-9161-z.

[12] S. Aum, S. Y. T. Lee, Y. Shin, COVID-19 Doesn't Need Lockdowns to Destroy Jobs: The Effect of Local Outbreaks in Korea, Tech. Rep. w27264, National Bureau of Economic Research, Cambridge, MA (May 2020). doi:10.3386/w27264.

[13] Y. J. Park, Y. J. Choe, O. Park, S. Y. Park, Y.-M. Kim, J. Kim, S. Kweon, Y. Woo, J. Gwack, S. S. Kim, J. Lee, J. Hyun, B. Ryu, Y. S. Jang, H. Kim, S. H. Shin, S. Yi, S. Lee, H. K. Kim, H. Lee, Y. Jin, E. Park, S. W. Choi, M. Kim, J. Song, S. W. Choi, D. Kim, B.-H. Jeon, H. Yoo, E. K. Jeong, Early Release - Contact Tracing during Coronavirus Disease Outbreak, South Korea, 2020 - Volume 26, Number 10—October 2020 - Emerging Infectious Diseases journal - CDCdoi:10.3201/eid2610.201315.