INSTITUTE OF DEVELOPING ECONOMIES

IDE Discussion Papers are preliminary materials circulated
to stimulate discussions and critical comments

IDE DISCUSSION PAPER No. 793

A Gravity Solution to the Puzzling Low Number of COVID-19 Deaths in East Asia

Satoru KUMAGAI*

June 2020

Abstract

The number of deaths from the coronavirus disease 2019 (COVID-19) differs dramatically by country. The surprisingly low number of deaths from COVID-19 in several East Asian countries is a puzzle and we propose a different explanation for the variation in the number of deaths by country, the geographical distance from an epicenter of COVID-19. Combined with the gross domestic product for each country, we can explain the country variation in the number of deaths for more than 100 countries surprisingly well. By introducing these control variables, we can also correctly estimate the impacts of policy response and social/economic conditions. Adjusted by the gravity control variables, the difference in the number of deaths in European and American countries and East Asian countries becomes much smaller, from a factor of several hundred to a single digit, proposing a simple but powerful solution to the puzzle.

Keywords: COVID-19, gravity model, number of deaths
JEL classification: I10, C10, F10

---

* Director, Economic Geography Studies Group, Development Studies Center, IDE (satoru_kumagai@ide.go.jp)

The Institute of Developing Economies (IDE) is a semigovernmental, nonpartisan, nonprofit research institute, founded in 1958. The Institute merged with the Japan External Trade Organization (JETRO) on July 1, 1998. The Institute conducts basic and comprehensive studies on economic and related affairs in all developing countries and regions, including Asia, the Middle East, Africa, Latin America, Oceania, and Eastern Europe.

INSTITUTE OF DEVELOPING ECONOMIES (IDE), JETRO
3-2-2, WAKABA, MIHAMA-KU, CHIBA-SHI
CHIBA 261-8545, JAPAN

A Gravity Solution to the Puzzling Low Number of COVID-19 Deaths in East Asia

Satoru KUMAGAI

Abstract

The number of deaths from the coronavirus disease 2019 (COVID-19) differs dramatically by country, which makes it difficult for policymakers to effectively tackle this new infectious disease. The surprisingly low number of deaths from COVID-19 in several East Asian countries is a puzzle that has elicited various hypotheses. In this paper, we propose a different explanation for the variation in the number of deaths by country. It is the geographical distance from an epicenter of COVID-19. Combined with the gross domestic product for each country, we can explain the country variation in the number of deaths for more than 100 countries surprisingly well. By introducing these control variables, we can also correctly estimate the impacts of policy response and social/economic conditions. Adjusted by the gravity control variables, the difference in the number of deaths in European and American countries and East Asian countries becomes much smaller, from a factor of several hundred to a single digit, proposing a simple but powerful solution to the puzzle.

Keywords: COVID-19, gravity model, number of deaths
JEL Classification: I10, C10, F10

Introduction

Since early January 2020, the novel coronavirus disease 2019 (COVID-19) has infected more than 6 million people and caused more than 371,000 deaths all over the world, according to the World Health Organization (WHO) figures as of June 1, 2020. Still, the infection continues to spread. The spread of COVID-19 will almost certainly affect the world economy. However, the lack of scientific knowledge to explain the considerable differences in the number of deaths by country prevent policymakers from formulating proper policy measures to contain the virus with minimal economic costs.

Indeed, the number of deaths from COVID-19 differs significantly by country. As of June 1, 2020, total deaths per 1 million population was highest among European and

American countries. Spain has 622 deaths per 1 million; the United Kingdom (UK) has 579; Italy 553; France 429; and the United States (US) has 314. On the contrary, the number is meager in East Asian countries. Vietnam has 0 per 1 million; Thailand has 1; Korea has 5; and Japan has 7. These minimal numbers in East Asia have puzzled many researchers.

This variation in the number of deaths is so considerable that various assumptions explaining this puzzle have appeared on blogs and in newspapers, as well as in academic papers. Not all of them come to sound conclusions, however. For instance, Dowd et al. (2020) insists that the age structure of the people infected is essential in analyzing the transmission and mortality rates from COVID-19. The authors assert that the higher number of deaths in Italy is explained by a higher population of persons older than 65 years, while the impact in Korea is minimal because the initially infected populations were mainly young.

Miller et al. (2020) showed that countries without a mandatory Bacillus Calmette-Guerin (BCG) tuberculosis vaccination policy have been more severely affected by COVID-19. Akiyama and Ishida (2020) found that the Japanese strain of BCG has a higher effect of reducing the infection compared with other strains of BCG. In contrast, other researchers deny the effectiveness of BCG vaccinations on COVID-19 (Asahara 2020; Hensen et al., 2020). So far, the actual effects of BCG vaccination on the transmission and mortality rates from COVID-19 are controversial.

Some others insist that the Asian strains of COVID-19 are less harmful than the European strains, and there is a hypothesis that generic differences among races would affect the different mortality rates from COVID-19, but neither hypothesis is plausible at this moment.

In this paper, we introduce a new explanatory variable, the distances from an estimated epicenter, combined with country's gross domestic product (GDP). This approach is similar to the gravity model often used in the estimation of bilateral trade values in international economic literature. We found that this equation alone explains more than ninety percent of the country variation in the number of deaths from COVID-19, which provides an answer to the puzzle of COVID-19 death rates.

This paper is structured as follows. First, we proposed a simple model to explain the cumulative number of deaths from COVID-19, then we showed that the difference in the initial number of people infected matters. This number can be approximated by the distance from an estimated epicenter and country GDPs. Subsequently, we tried to estimate the impacts of other policy measures and social/economic conditions for each country, followed by the discussion on the difference of the number of deaths from COVID-19 between European and American countries and East Asian countries. We conclude the paper by summarizing the results from the estimation and its implications.

1. Model

At the early stage of a pandemic, we can approximate the number of deaths from the pandemic as the following exponential function:[1]

$$D_i(t) = \delta(X_i) I_i(0) e^{g(X_i)t} \quad \ldots (1)$$

where $D_i(t)$ is the number of cumulative deaths from COVID-19 for country *i* at time *t*. $I_i(0)$ is the initial number of people infected in country *i*. $g(X_i)$ is a function to determine the growth rate of the number of people infected. $\delta(X_i)$ is a function to determine the mortality rate from COVID-19. X_i is a vector of policy variables and social/economic conditions for country *i*.

By taking the log and evaluating it at a specific time *T*, Equation (1) is transformed into:

$$log(D_i(T)) = log(I_i(0)) + T \cdot g(X_i) + \log(\delta(X_i)) \quad \ldots (2)$$

Most of the previous literature seems to concentrate implicitly on $g(X_i)$ and $\delta(X_i)$ when discussing the country differences in $D_i(t)$. So far, the argument seems to concentrate on what policy measures/social-economic conditions determine the growth rate of infection and the mortality rate for different countries.

However, it is not plausible to believe that $g(X_i)$ and $\delta(X_i)$ are more than 100 times different by country. For the 2009 pandemic Influenza A H1N1 virus, the estimated case fatality ratio (CFR), defined as the ratio of the number of persons who died from the

[1] Equation (1) can be derived from a standard SIRD model, which is commonly used to model the spread of infectious disease

disease to the number of confirmed cases, for the age group 18–64 years old differs by country but the difference is less than ten times (Dawood et al., 2012). For the same 2009 pandemic, Biggerstaff et al. (2015) reported the interquartile range of the reproduction number as 1.30–1.70 from 57 studies.

Thus, we have no choice but to look at the differences in $I_i(0)$, the initial number of persons infected, as a source of the subsequent differences in the number of deaths. For most countries, it is very plausible to assume that a large part of the number of people infected at an early stage of the epidemic, especially before lockdown, was imported from other countries that already had a large number of people infected. This difference in the early imported cases leads to a significant difference in $I_i(0)$ by country.

2. A Gravity Approach

If the differences in $I_i(0)$ among countries are the size of a single-digit order, then the differences in $g(X_i)$ and $\delta(X_i)$ would dominate in the long run. However, if the differences in $I_i(0)$ are substantial, then they would dominate the number of deaths from COVID-19 for a long-time.

The main problem in estimating Equation (2) is that the actual number of the initial number of persons infected for each country was not observed, and we cannot proceed directly to the estimation. To address this problem, we need to estimate the numbers for each country.

COVID-19 is only transmitted through the movement of people. Here we assume that the infected persons at an early stage of the epidemic are imported into each country from an estimated epicenter (countries/regions with a large number of infected persons already there). We also assume that an initial number of infected persons brought into each country is inversely proportional to the distance from the epicenter and proportional to the GDP of a partner country.

This idea is based on the gravity equation, a famous empirical model in international economics that estimate the bilateral value of international trade, formulated as follows:

$$X_{ij} = \frac{(Y_i)^{\alpha} \cdot (Y_j)^{\beta}}{(K_{ij})^{\gamma}} \quad \dots (3)$$

where X_{ij} is the bilateral trade value between countries *i* and *j*, and Y_i and Y_j are GDPs for countries *i* and *j*, respectively. K_{ij} is the geographical distance between countries *i* and *j*. This equation means that the volume of bilateral trade is proportional to the sizes of GDPs of trading pairs, while inversely proportional to the distance between the two countries. By taking the log, Equation (3) becomes

$$log(X_{ij}) = \alpha log(Y_i) + \beta log(Y_j) - \gamma log(K_{ij}) \quad \ldots \quad (4)$$

By estimating Equation (4), we can get the coefficients α, β, and γ; the elasticities of trade value to GDP for country *I;* GDP for country *j;* and the distance between countries *i* and *j*, respectively. These elasticities are all reasonably close to unity in the previous literature. The gravity equation is known as one of the most robust models in empirical studies of international trade.

Here, we formulate the initial number of persons infected as follows:

$$log(I_i(0)) = \alpha + \beta log(Y_i) - \gamma log(K_i[lat, lng]) \quad \ldots \quad (5)$$

where $I_i(0)$ is the initial number of persons infected in country *i*. Y_i is GDP for country *i*. $K_i[lat, lng]$ is the distance between an epicenter located at the geographical coordinates (latitude, longitude) somewhere on the earth and the capital city of country *i*.

By substituting Equation (5) into Equation (2), we get the model.

$$log(D_i(T)) = \alpha + \beta log(Y_i) - \gamma log(K_i[lat, lng]) + \ T \cdot g(X_i) + \log(\gamma(X_i)) \ldots \quad (6)$$

In this study, we conducted cross-country regression on the cumulative number of deaths on April 30, 2020, so that the time dimension is fixed. In addition to the gravity variables Y_i and $K_i[lat, lng]$, we added several social/economic variables corresponding to X_i in $g(X_i)$ and $\gamma(X_i)$ and estimated the model as follows:

$$\log(D_i) = \alpha + \beta log(Y_i) - \gamma \log(K_i[lat, lng]) + \delta \ bcg.all_i + \theta_1 \log(gdppc_i) + \theta_2 log(pop.65_i) + \theta_{31} pop.urban_i + \theta_{32}(pop.urban_i)^2 + u \ \ldots \quad (7)$$

where D_i is the cumulative number of deaths for country *i* as of April 30, 2020, provided by Eurostat (2020) for more than 200 countries. We selected the data for 142 countries, dropping the countries with less than one million population and countries where other data is not available. Y_i is GDP for country *i* in 2018 from the World Development Indicator Online (WDIO) provided by the World Bank. $K_i[lat, lng]$ is the distance between an epicenter represented by latitude (*lat*) and longitude (*lng*), and the capital city of country *i*, *bcg.all* is the dummy variable that equals 1 if country *i* conducts mandatory BCG vaccinations and equals 0 otherwise—compiled from Zwerling et al. (2011). *pop.65* is the share of the population older than 65 years old and *pop.urban* is the share of the population that lives in cities with more than one million population, both variables are taken from WDIO. For *pop.urban*, a quadratic term is also included, because a negative impact of urban crowding on COVID-19 infection/deaths is expected to peak at some level.

We also tried to estimate the impacts of different strains of BCG vaccines. An argument has been made that specific strains of BCG vaccines are more effective on COVID-19 than others. We estimated the following equation in addition to Equation (7).

$$\log(D_i) = \alpha + \beta log(Y_i) - \gamma \log(K_i[lat, lng]) + \delta_1 \ bcg.jp_i + \delta_2 bcg.ru_i + \delta_3 bcg.dk_i + \delta_4 bcg.oth_i + \theta_1 \log(gdppc_i) + \theta_2 log(pop.65_i) + \theta_{31} pop.urban_i + \theta_{32}(pop.urban_i)^2 + u \ \ldots (8)$$

where *bcg.jp, bcg.ru*, *bcg.dk and bcg.oth* are the dummy variables that equal 1 if country *i* adopts BCG strains from Japan, Russia, Demark, and other/mixed strains under mandatory vaccination policy(and all of the dummy variables are 0 if country *i* does not conduct mandatory BCG vaccinations), compiled from Ritz and Curtis (2009).[2] The summary statistics of variables are shown in Table 1.

<<Insert Table 1 Here>>

[2] Although the BCG strain used in Malaysia is not provided by Ritz and Curtis (2009), we identified it as the Japanese strain, from Hooia and Athiyah (1994).

The remaining procedure determines the exact geographical coordinate of the epicenter of the COVID-19 infection. The first candidate for the epicenter was the average of the latitudes and longitudes of capital cities of countries weighted by the number of deaths. However, we dismissed this option because composing an explanatory variable from explained variables cannot exclude the endogeneity problem.

We solved this problem empirically. We repeatedly estimated the gravity equations (7) and (8) through Poisson regression, changing longitude and latitude by 1 degree around the world. Then we selected the coordinate with the highest log-likelihood and used it as the epicenter of the model. This approach is similar to the threshold regression approach developed by Hansen (2000).

3. Estimation Results

Table 2 shows the results of the estimation.[3] Model 1 includes only gravity variables., A point at the center of the Atlantic Ocean (coordinates 28.0°N, 22.0°W) was selected as the epicenter. All coefficients are statistically significant at the 1% level. Pseudo-R^2 is 0.912, which is surprisingly high. Model 1 is interpreted the elasticity of the number of initially infected persons to country GDP is very near to unity, while the elasticity of distance from the epicenter is significantly higher than unity, −2.420. This higher elasticity becomes most consistent with the elasticity that appeared in previous literature that estimated the gravity model for the number of bilateral travelers,[4] as other explanatory variables were included.

<<Insert Table 2 Here>>

[3] As Model (0), we only include the log of the distance from the epicenter as an explanatory variable. You may think this recursive method to determine the center of gravity for the dependent variable with geographical coordinates always leads to a better-fit gravity equation. However, the elasticity to the distance is −3.810, too large for the gravity equation. The pseudo-R^2 remains low, 0.153, and the selected epicenter, (51.0N, 47.0W) located at the Labrador Sea is challenging to interpret.

[4] Keum (2008) shows that the elasticity of the number of outbound and inbound travelers to the distance between the country and Korea is −1.92 and −1.38, respectively.

Model 2 shows the estimation results of Equation (7) without including gravity variables. The coefficient for *gdppc* is positive, meaning that the higher the income level, the higher the number of deaths from COVID-19 the country has. Considering that per capita income is highly correlated with the general goodness of public health, this result is not plausible and is likely to be biased. The coefficient for *pop.65* is positive and statistically significant. This result is plausible considering that the mortality rate from COVID-19 for elderly persons is much higher than for other generations. The coefficient for the degree of urbanization was interpreted as a higher population share in an urban area of more than one million leads to a higher number of deaths, but this peaks at the 40% share. When compared with a country where *pop.urban* is 10%, the country where *pop.urban* is 30% is expected to have 11 times more deaths from COVID-19. This seems to be an overestimation. The coefficient for *bcg.all* is positive and statistically significant at 1% level. A country with compulsory BCG vaccination is expected to have a reduced number of deaths from COVID-19 by 86%, compared with a country without compulsory vaccination.

Model 3 is the full estimation of Equation (7), including gravity variables. The selected epicenter is at coordinates (16.0°N, 29.0°W). The coefficient for *gdppc* now becomes negative and statistically significant, as expected. A 10% increase in GDP per capita reduces the number of deaths by 6%. The coefficient for *pop.65* is positive and statistically significant, but the coefficient becomes smaller than that of Model 2. The signs for the coefficients of *pop.urban* are the same as Model 2 but become smaller. The interpretation of coefficients is that compared with the country that has *pop.urban* 10%, the country with *pop.urban* 30% (peak) is expected to have 2.5 times more deaths from COVID-19. This seems more plausible than Model 1. The coefficient for *bcg.all* is positive and statistically significant at the 1% level. A country with compulsory vaccination of BCG is expected to have a reduced number of deaths from COVID-19 by 54%, compared with a country without compulsory vaccination.

Models 4 and 5 estimate Equation (8), which shows the impacts of different strains of BCG vaccinations. Without gravity controls (Model 4), the dummy variables for all the strains of BCG are negative and statistically insignificant at the 1% level, but the impacts seem to be too large, reducing the number of deaths by 95% for the Japanese strain, for example. With gravity controls (Model 5), the Japanese and Russian strains of

BCG are expected to reduce the number of deaths from COVID-19 by 75% and 69%, respectively, while the Demark strain reduces the number by 48%.

4. Discussion

We plotted the log-likelihoods for repeated estimations of the Model 5 for all the coordinates on the globe (Figure 1). The thicker red represents the higher log-likelihood, while the blue regions show the coordinates with positive slope for the coefficient of distance, naturally excluded for the candidates of the epicenter. The highest log-likelihood is for the coordinates 14.0N, 29.0W, and we selected this point as the epicenter for Model 5.

<<Insert Figure 1 Here>>

The epicenter is in the middle of the Atlantic Ocean because it was interpreted that the source of the current spread of COVID-19 is not a single country but a group of the countries within which people's movement is very active. Our estimation implies that Europe and American countries are the epicenter as a group.

The epicenter in this study is determined to be in the middle of the North Atlantic Ocean instead of in China. This is because as by late January 2020 most countries in East Asia were very conscious of allowing access to infected persons from China. Conversely, Western Europe and the US were generally not aware of the risk of COVID-19 because China is distant from these countries. However, a specific region in Northern Italy has a special relationship with Wuhan, China, the epicenter from January to February 2020, due to the garment industry. Clearly COVID-19 was imported to Northern Italy at the early stage of the epidemic (Caccavo, 2020). Then, infected persons were supposed to have spread through the network of free movement of people in the EU and then crossed the Atlantic to the US without notice. Our estimation result implies that the COVID-19 has spread from the North Atlantic countries to the world.

Figure 2 shows the trail of the centroid of deaths from COVID-19 for a week from the date labeled above the dot in the map. The centroid is the average of latitude and longitude of the capital city of each country (for China, it is Wuhan City, instead of the capital), weighted by the number of deaths for the last seven days. The centroid of

COVID-19 was in China, until the end of February 2020. Then the centroid started to shift westward and reached Italy in late March 2020. Then it shifted to the North Atlantic after mid-April 2020.

<<Insert Figure 2 Here>>

Wuhan City, China, was locked down on January 23, 2020, while the lockdown for all the regions of Italy was introduced on March 10, 2020, and Spain followed on March 17, 2020. The US restricted the travel from the EU on March 13, 2020, then from the UK and Ireland on March 17, 2020. It was not easy to determine which date was the early stage of the pandemic, as mentioned in this paper, but until mid-March 2020, the movement of the people in the North Atlantic countries was not rigorously restricted. The movement of people might have continued after the lockdown policy was introduced, and infected persons were still spreading the virus through movement. In this case, the gravity equation introduced here explains the situation well.

There are some possible other explanations for this gravity relationship between the number of deaths and the distances from a point in the North Atlantic. For instance, some other factors, such as culture and the way of life, are related to the infection/mortality rates of COVID-19, and these Euro-American factors spread geographically following the gravity equation. However, considering a very high R^2 of the model estimated here and the fact that the elasticity of deaths to the distance is not very different from the elasticity of the number of travelers noted in previous literature, it is plausible to think that this gravity equation is likely to be associated with the movement of the people from Europe and the US to the world.

We now propose a solution to the puzzle of the minimal number of deaths in East Asia. Figure 3 shows the relationship between numbers of deaths and the distances from the epicenter, adjusted by the size of GDP according to Model 1 shown in Table 2. The negative relationship between the numbers of deaths and the distances from the epicenter is apparent.

<<Insert Figure 3 Here>>

We interpret this to mean that a large part of this minimal number of deaths for East Asia results from the difference in the initial number of infected persons imported, associated with the distance from the epicenter of COVID-19 and the differences in country GDP. Table 3 shows the number of deaths for North Atlantic countries as of April 30, 2030, adjusted by the distance and GDP of Thailand, according to the coefficients of Model 3 in Table 2. If the UK was located at Thailand's place, the number of deaths from COVID-19 would significantly decline from 26,097 to 1,798. In addition to that, if the size of the GDP for the UK is the same as for Thailand, the number of deaths further decreases to 237.

<<Insert Table 3 Here>>

Figure 4 shows the actual and predicted number of deaths from COVID-19 as of April 30, 2020. The countries located below the diagonal lines managed to control the number of deaths from COVID-19 more than was predicted. Almost all the countries are located on the narrow band around the diagonal line, which means that there is no big puzzle remaining after controlling for the gravity condition and other essential social/economic variables.

<<Insert Figure 4 Here>>

We can see that some countries are doing better than other countries, but the countries that are doing well are not necessarily concentrated in East Asia. Among European countries, Austria, Norway, Finland and Greece are relatively doing well, while Italy, Belgium, and the Netherlands are not. Among East Asian countries, Thailand and New Zealand are doing well, while Indonesia and the Philippines are not.

The main policy implication of this gravity solution to the puzzle of COVID-19 deaths is that the movement of people at an early stage of the pandemic has a crucial impact on the subsequent impacts from the disease, although other policies and social/economic conditions also matter. Once a large number of infected persons silently spreads in a country at an early stage, it is exceedingly difficult to control the infection *ex post*.

Conclusions

In this paper, we proposed the hypothesis that the initial number of infected persons largely determines the subsequent number of deaths by country. By using a gravity equation, three-quarters of the country variation of the number of deaths from COVID-19 is successfully explained.

The estimation with gravity variables provides a solution to a puzzle in COVID-19 studies; that is, the minimal number of deaths in East Asian countries. When compared in raw numbers, the death toll in the North Atlantic countries is around 500 to 1,000 times larger than that in Thailand, but adjusted by gravity variables, the difference becomes much smaller, a factor of 1.7 to 13.3.

We also show that with these gravity variables, the impacts of various policy and social/economics variables can be properly estimated. Our estimations show that the higher the income and the smaller the share of the older persons a country has, the more deaths from COVID-19 are reduced. The relationship between the share of the urban population and the number of deaths is an inverse-U shape, which peaks if approximately forty percent of the population lives in the cities with more than 1 million population. In terms of the controversial argument that BCG vaccination reduces the deaths from COVID-19, we showed that the Japanese, Russian and Demark strains of BCG reduced deaths by two-thirds, by seventy percent, and by half, respectively. This is consistent with some the previous literature.

The gravity solution to the puzzle of the minimal number of deaths from COVID-19 in East Asia seems to be too simple, but it is not as unreasonable as assuming that the reproduction number and the mortality rate of COVID-19 differs significantly, sometimes by a factor of 100 to 1,000 among different countries. Although there are other possible explanations for this gravity model of the deaths from COVID-19, at this time it is plausible to relate this model to the differences in the initial number of infected persons imported from the epicenter.

COVID-19 is the largest menace to human beings in this century, and we need to control this disease by balancing the costs to the economy and the costs to human life. To formulate effective policy measures against COVID-19, it is essential to know what the determinants of deaths from COVID-19 are. So far, a considerable variation in the number of deaths by country presents a puzzle, and we have proposed a clue to solving it through the gravity model.

References

Akiyama, Y., and Ishida, T. (2020). Relationship between COVID-19 death toll doubling time and national BCG vaccination policy. *medRxiv*.

Asahara, M. (2020). The effect of BCG vaccination on COVID-19 examined by a statistical approach: no positive results from the Diamond Princess and cross-national differences previously reported by world-wide comparisons are flawed in several ways. *medRxiv*.

Biggerstaff, M., Cauchemez, S., Reed, C., Gambhir, M., & Finelli, L. (2014). Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature. *BMC infectious diseases*, *14*(1), 480.

Caccavo, D. (2020). Chinese and Italian COVID-19 outbreaks can be correctly described by a modified SIRD model. *medRxiv*.

Dawood, F. S., Iuliano, A. D., Reed, C., Meltzer, M. I., Shay, D. K., Cheng, P. Y., ... and

Eurostat (2020). *Download today's data on the geographic distribution of COVID-19 cases worldwide* (https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide).

Feikin, D. R. (2012). Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study. *The Lancet infectious diseases*, *12*(9), 687-695.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica*, *68*(3), 575-603.

Hensel, J., McGrail, D. J., McAndrews, K. M., Dowlatshahi, D., LeBleu, V. S., and Kalluri, R. (2020). Exercising caution in correlating COVID-19 incidence and mortality rates with BCG vaccination policies due to variable rates of SARS CoV-2 testing. *medRxiv*.

Hisaka, A.; Yoshioka, H.; Hatakeyama, H.; Sato, H.; Onouchi, Y.; and Anzai, N. (2020). Global Comparison of Changes in the Number of Test-Positive Cases and Deaths by Coronavirus Infection (COVID-19) in the World. *Preprints 2020*, 2020040374 (doi: 10.20944/preprints202004.0374.v1).

Hooi, L. N., and Athiyah, S. O. (1994). An outbreak of BCG related lymphadenitis in Malaysian infants. *Med J Malaysia*, *49*(4), 327-335.

Iwasaki, A., and Grubaugh, N. D. (2020). Why does Japan have so few cases of COVID19? *EMBO Molecular Medicine*.

Keum, K. (2008). Tourism flows and trade theory: a panel data analysis with the gravity model. *The Annals of Regional Science*, *44*(3), 541–557. http://doi.org/10.1007/s00168-008-0275-2

Miller, A., Reandelar, M. J., Fasciglione, K., Roumenova, V., Li, Y., and Otazu, G. H. (2020). Correlation between universal BCG vaccination policy and reduced morbidity and mortality for COVID-19: an epidemiological study. *MedRxiv*.

Morley, C., Rosselló, J., and Santana-Gallego, M. (2014). Gravity models for tourism demand: theory and use. *Annals of Tourism Research*, 48, 1-10.

Ritz, N., and Curtis, N. (2009). Mapping the global use of different BCG vaccine strains. *Tuberculosis*, *89*(4), 248-251.

Zwerling, A., Behr, M. A., Verma, A., Brewer, T. F., Menzies, D., and Pai, M. (2011). The BCG World Atlas: a database of global BCG vaccination policies and practices. *PLoS medicine*, *8*(3).

Table 1: Summary statistics of variables used in the estimation

	Obs.	Min.	Mean	Median	Max.	S.D.
D	142	0	1598	43	60965	6646
Y	135	1	614	71	20500	2198
gdppc	135	0	15	7	83	20
pop.65	137	1	10	7	28	7
pop.urban	110	4	27	23	100	17
bcg.all	142	0	1	1	1	0
bcg.JP	142	0	0	0	1	0
bcg.RU	142	0	0	0	1	0
bcg.DK	142	0	0	0	1	0
bcg.OTH	142	0	1	1	1	0

(Source) Authors.

Table 2: Results of estimation

	(0)		(1)		(2)		(3)		(4)		(5)	
epicentre (lat, lng)	(51°N, 47°W)		(28°N, 22°W)		-		(16°N, 29°W)		-		(14°N, 29°W)	
(Intercept)	39.650	***	-4.609	***	0.439	***	-6.420	***	-0.339	***	-6.923	***
	(0.051)		(0.048)		(0.031)		(0.092)		(0.036)		(0.096)	
log(K[lat,lng])	-3.810	***	-2.420	***			-2.074	***			-2.001	***
	(0.006)		(0.006)				(0.011)				(0.011)	
log(Y)			1.188	***			1.046	***			1.058	***
			(0.002)				(0.003)				(0.003)	
log(gdppc)					0.156	***	-0.337	***	0.217	***	-0.423	***
					(0.005)		(0.007)		(0.006)		(0.008)	
log(pop.65)					1.614	***	1.119	***	1.800	***	1.096	***
					(0.011)		(0.014)		(0.012)		(0.014)	
pop.urban					0.222	***	0.144	***	0.222	***	0.146	***
					(0.001)		(0.001)		(0.001)		(0.001)	
$pop.urban^2$					-0.0028	***	-0.0024	***	-0.003	***	-0.0025	***
					(0.0000)		(0.0000)		(0.0000)		(0.0000)	
bcg.all					-1.987	***	-0.781	***				
					(0.008)		(0.011)					
bcg.JP									-2.894	***	-1.380	***
									(0.039)		(0.042)	
bcg.RU									-2.195	***	-1.096	***
									(0.016)		(0.019)	
bcg.DK									-2.352	***	-0.651	***
									(0.012)		(0.013)	
bcg.OTH									-1.318	***	-1.056	***
									(0.011)		(0.017)	
d.f.	140		132		100		98		97		95	
pseudo-R^2	0.153		0.912		0.448		0.972		0.450		0.974	

Numbers in brackets represent the standard error.

*** indicates 1% significance.

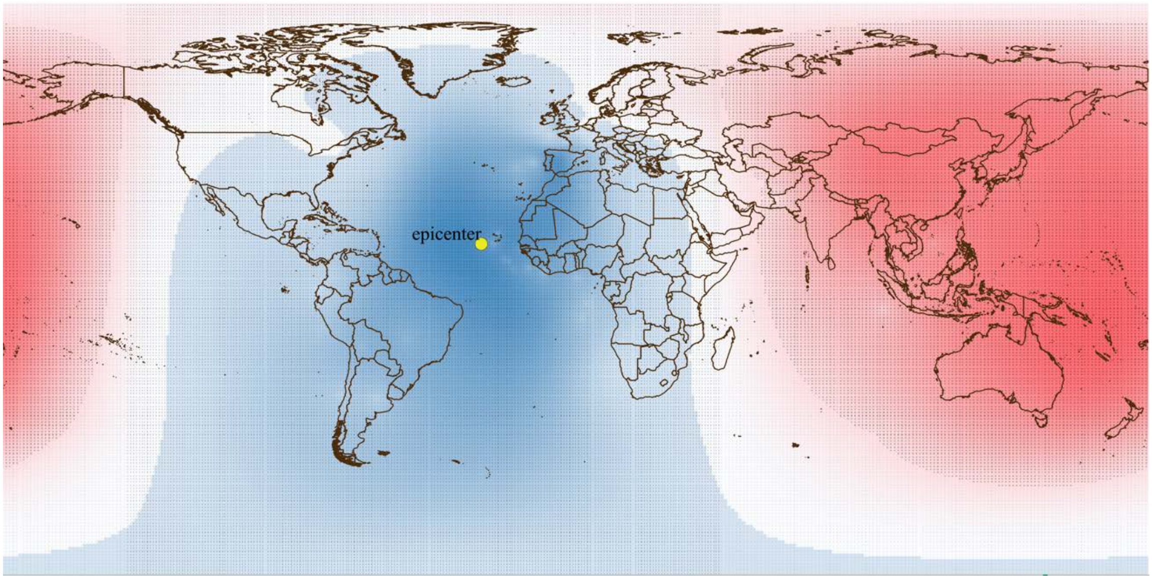
(Source) Authors.

Table 3: Number of Deaths for selected country adjusted by Thailand condition

		UK	France	Italy	Spain	US
(1)	# of deaths	26,097	24,087	27,682	24,543	60,966
	(Thailand=1.0)	483	446	513	455	1,129
	Distance (Thailand=1.0)	0.26	0.25	0.28	0.18	0.42
(2)	(1) adjusted by distance	3,276	2,908	3,948	1,696	16,051
	(Thailand=1.0)	61	54	73	31	297
	GDP (Thailand=1.0)	5.7	5.5	4.1	2.8	40.6
(3)	(2) adjusted by GDP	547	500	915	583	350
	(Thailand=1.0)	10.1	9.3	17.0	10.8	6.5

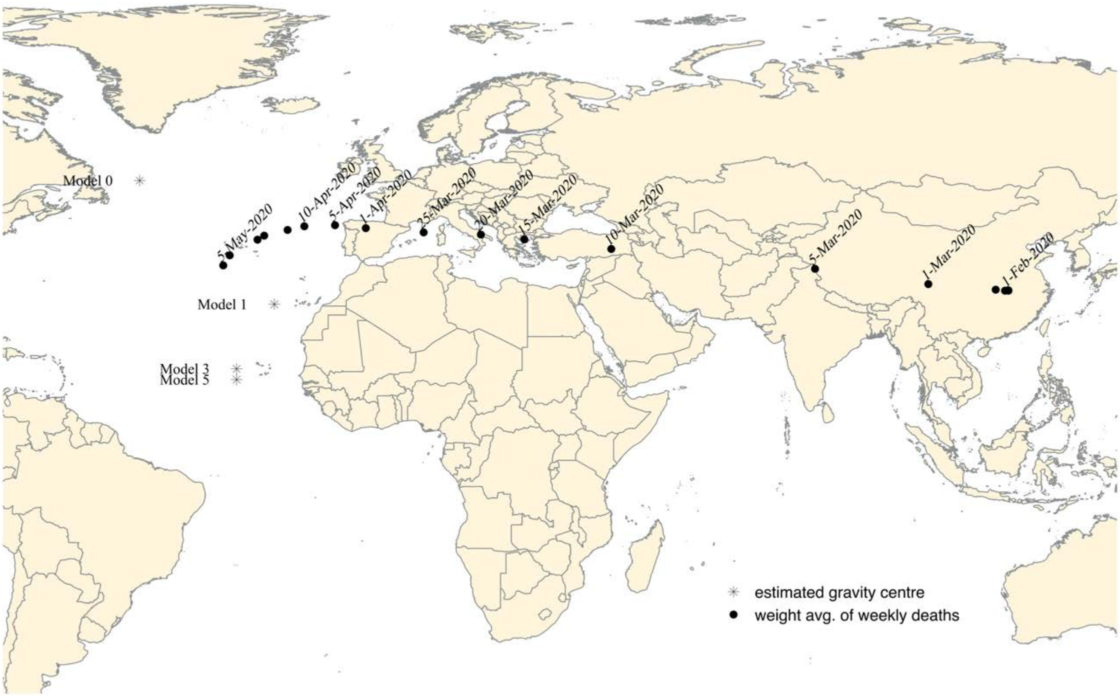
(Source)Authors.

Figure 1: Adjusted-R^2 of Model (5) for all the combinations of latitude and longitude



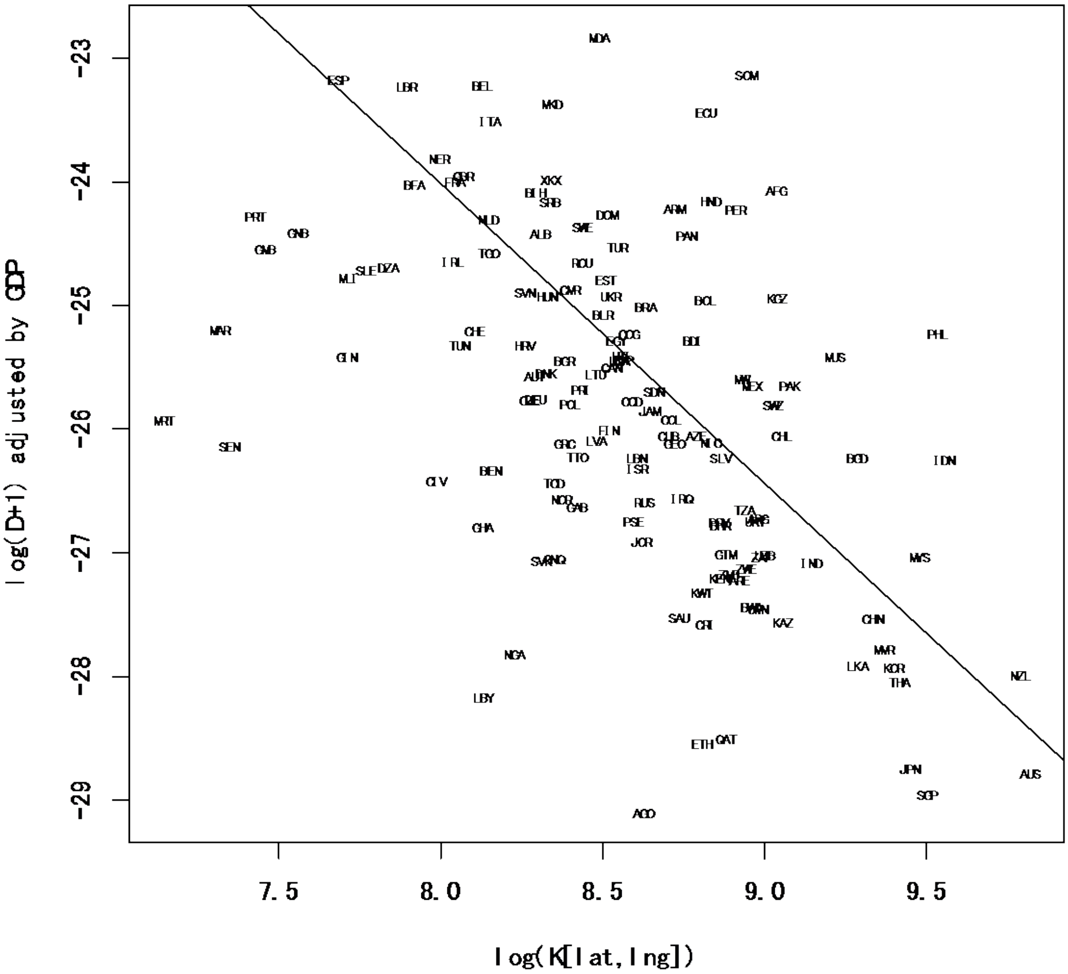
(Source) Authors.

Figure 2: The centroid of the deaths from COVID-19



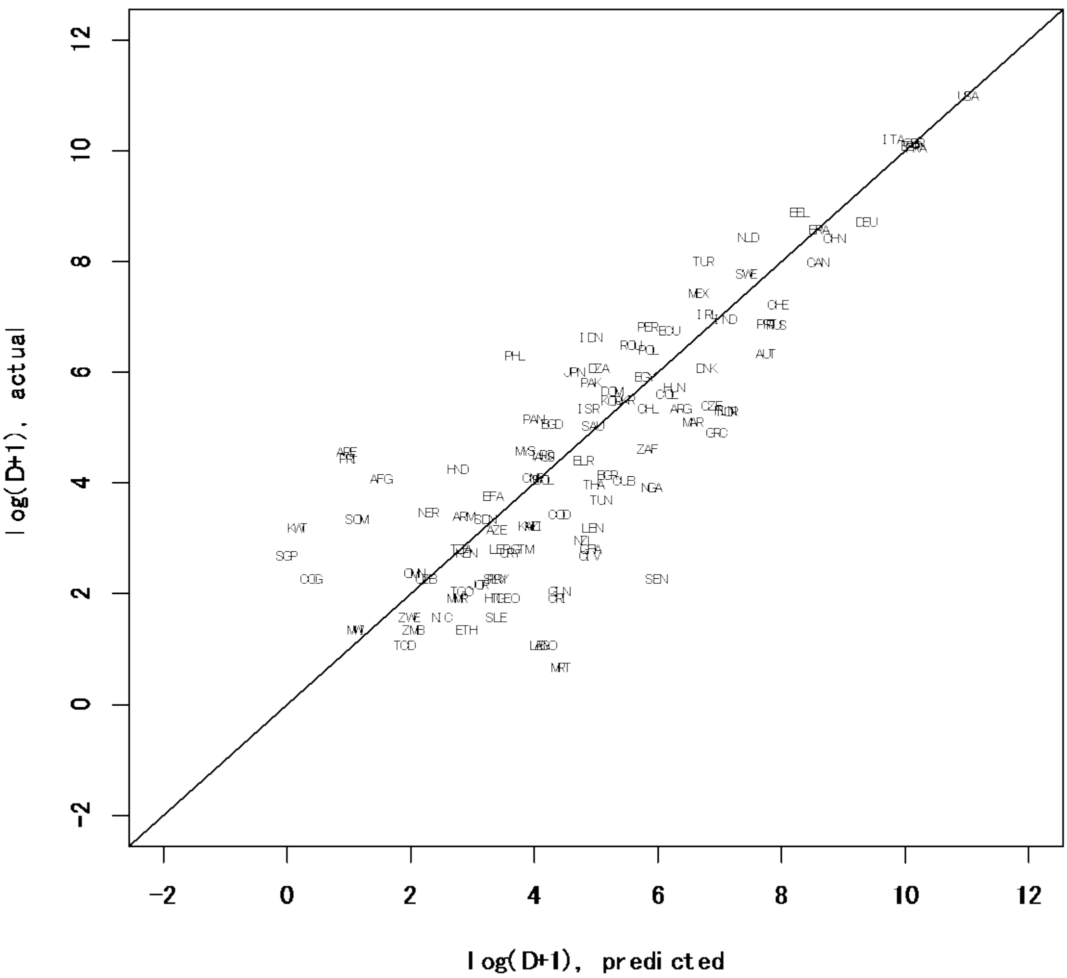
(Source) Composed by authors from the data provided by Eurostat (2020).

Figure 3: Relationship between the distance from the epicenter and the number of deaths from COVID-19



(Source) Authors.

Figure 4: Actual and predicted number of deaths from COVID-19 (as of April 30, 2020)



(Source)Authors.