**PNAS Template for Main Manuscript**

This PNAS template for the Main Manuscript may be used to organize your main text source file. The template is intended to provide a clearly organized PDF to facilitate the review process. Further information is available in our [Author Center](https://www.pnas.org/authors/submitting-your-manuscript#article-types).

**Using the template**

Paste the appropriate text from your manuscript into the relevant section of the template. You may maintain the template formatting, or reapply styles after pasting your text into the template.

Figures should be placed on separate pages with legends set immediately below each figure. Table titles should be set immediately above each table.

References cited in the main text should be included in a separate reference list at the end of this file. Examples of the PNAS citation style are included below.

**Notes about submission**

The following items are required on the title page: **Title, Author Line, Author Affiliations, Corresponding Author information**. Each of these items **must** be provided on the title page for us to proceed with processing your paper.

You are not required to adhere to the section order outlined below. For example, you may combine your Results and Discussion or use alternate section headings. Materials and Methods should be included after the Results and Discussion in most cases. If your paper does not include the standard section headings, please provide a brief explanation in the **Comments to Editorial Staff** field of the submission form.

You may include subheadings within the standard headings listed below. You may also include line numbers.

**Submitting your main manuscript**

Delete this first page, and then save your completed main text file as a PDF for submission, following instructions available [here](http://www.pnas.org/site/authors/procedures.xhtml#preparation).

*Updated September 2020*



**Main Manuscript for**

Which national factors are most influential in the spread of COVID-19?

Hakyong Kim1†, Catherine Apio2†, Yeonghyeon Ko3,4†, Kyulhee Han2, Taewan Goo2, Gyujin Heo2, Taehyun Kim3, Hyewon Chung5, Doeun Lee2, Jisun Lim6, Taesung Park3\*

1Department of Industrial Engineering, Seoul National University, Seoul 08826, Republic of Korea; gkrdyd111@snu.ac.kr (H.K.)

2Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul 08826, Republic of Korea; hgj0106@snu.ac.kr (G.H.); 2019-20240@snu.ac.kr (C.A.); hgh1031@snu.ac.kr (K.H.); gootec92@snu.ac.kr (T.G.); 2corin417@snu.ac.kr (D.L.)

3Department of Statistics, Seoul National University, Seoul 08826, Republic of Korea; newstellar@snu.ac.kr (Y.K.); qeeqee10@snu.ac.kr (T.K.); taesungp@stats.snu.ac.kr (T.P.)

4Department of Archeology and Art History, Seoul National University, Seoul 08826, Republic of Korea; newstellar@snu.ac.kr (Y.K.)

5Department of Chemistry, Seoul National University, Seoul 08826, Republic of Korea; hyewon.chung@snu.ac.kr (H.C.)

6The Research Institute of Basic Sciences, Seoul National University, Seoul 08826, Republic of Korea; [swanjslim@gmail.com](mailto:swanjslim@gmail.com) (J.L.)

†These authors contributed equally to this work as co-first authors.

\*Corresponding author, Tel: 82-2-880-8924, Fax: 82-2-883-6144

**Email:**  tspark@snu.ac.kr (T.P)

PNAS strongly encourages authors to supply an [ORCID identifier](https://orcid.org/) for each author. Do not include ORCIDs in the manuscript file; individual authors must link their ORCID account to their PNAS account at [www.pnascentral.org](http://www.pnascentral.org). For proper authentication, authors must provide their ORCID at submission and are not permitted to add ORCIDs on proofs.

**Author Contributions:**

Conceptualization: T.P.

Methodology: T.P., H.K., Y.K.

Investigation: H.K., Y.K., K.H., H.C., T.G., A.C., D.L., J.L., T.K.

Visualization: H.K., K.H., Y.K., A.C.

Supervision: T.P.

Data Curation: K.H., G.H., T.G., H.K., Y.K., D.L.

Writing—original draft: A.C., H.K., Y.K.

Writing—review & editing: T.P., C.A., G.H., J.L.

Project Administration: T.P.

Funding Acquisition: T.P.

**Competing Interest Statement:** The authors declare no competing interests.

**Classification:** Paste the major and minor classification here. Dual classifications are permitted, but cannot be within the same major classification.

**Keywords:** Paste the keywords here. There should be at least three and no more than five.

**This PDF file includes:**

Main Text

Figures 1 to 7

**Abstract**The outbreak of the novel COVID-19 declared a global pandemic by WHO, is the most serious public health threat seen in respiratory virus since the 1918 H1N1 influenza pandemic. It is surprising the total number of COVID-19 confirmed cases and the number of deaths vary greatly across countries. Such great variations are caused by age population, health conditions, travel, economy and environmental factors. Here, we investigated which national factors (life expectancy, average annual temperature, aging index, human development index, percentage of malnourished people in the population, extreme poverty, economic ability, health policy, population, age distributions, etc.) influence the spread of COVID-19 through systematic statistical analysis. First, we employed segmented growth-curve models (Logistic and Gompertz) to model the cumulative confirmed cases for 134 countries from January 1, 2020 to August 31, 2020. Thus, each country’s COVID-19 spread pattern was summarized into three growth-curve model parameters. Second, we investigated the relationship of selected 31 national factors (from KOSIS and *Our World in Data*) to these growth curve model parameters. Our analysis showed that the parameter related to the maximum number of predicted cumulative confirmed cases was greatly influenced by the total population size, as expected. The other parameter related to the rate of spread of COVID-19 is influenced by aging index, cardiovascular death rate, extreme poverty, median age, percentage of population aged 65, 70 and older and so forth. We hope with consideration of a country’s resources and population dynamics; our results will help in making informed decisions that make the most impact against similar infectious diseases.

**Significance Statement**Through statistical analysis using segmented growth curve models, the spread of COVID-19 was found to be significantly influenced by population, age distribution, malnourishment, extreme poverty, health, life expectancy, movement, and environment.

**Main Text**

**Introduction**

The novel corona virus disease 2019 (COVID-19), a highly transferable viral disease is a respiratory illness caused by novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) which has person-to-person contact as the main route of transmission and causes flu like symptoms and in severe cases death (1,2). The spread of COVID-19 became a global threat and the World Health Organization (WHO) declared it a global pandemic on March 11, 2020 (3). The public health threat it represents is the most severe seen in respiratory viruses since the 1918 H1N1 influenza pandemic (4) with a total of 104,904,529 confirmed cases and 2,278471 deaths worldwide, as of February 4, 2021 (5).

Many factors can influence the epidemiological characteristics and contribute to the increased mortality rate of COVID 19 (6). Therefore, understanding the potential affecting factors involved in the outspread of COVID-19 will be of great significance in containing the spread of the epidemic and was the focus of many studies (6). There are several theories suggesting the impact of environmental factors like, temperature and humidity along with international travel and lack of proper social consciousness for isolation as causes of the global spread of COVID 19 (1). A few investigations have also considered social aspects potentially associated with the spread of COVID-19, such as population density, metropolitan population, intra-provincial traffic, and national lockdown, indicating that the social distancing measures have been successful in reducing new cases (7-10).

Furthermore, several studies have investigated the impact of weather on the COVID-19 transmission, with special attention to temperature and humidity, indicating that temperature is inversely related to COVID-19 incidence (11). Moreover, each 1 °C increase in temperature has been associated with decreases in daily new cases at different extents. Significant studies on the impacts of climatic predictors on COVID-19 transmission were also conducted in China, the United States, and Europe (12-13).

A look at history tells that pandemics and epidemics have consistently and significantly affected human lives, and governments have continually tried to find ways of slowing down the spread of these diseases, for example, quarantines were employed during the Ebola outbreak in West Africa (14-15). The objective of this study was to determine the relationship between potential national factors like life expectancy, average annual temperature, aging index, human development index, percentage of malnourished people in the population, extreme poverty, economic ability, health policy, population, age distributions, etc., and the spread of COVID-19.

Here, we first applied segmented growth-curve models (Logistic and Gompertz Models) to cumulative confirmed cases of 134 countries. As the spread of COVID-19 prolongs, several countries experienced more than one wave (a wave implies a rising number of sick individuals, a defined peak, and then a decline (70)) of the pandemic so that the cumulative confirmed cases could not be analyzed by conventional growth curve models, since cumulative cases cannot adopt the sigmoid curve. So, the study period of each country was divided into several segments; time periods corresponding to a specific wave, using the segmentation algorithm (see Materials and Methods, **Fig. S1**) which systematically partitions COVID-19 cumulative confirmed cases for each country into several segments of times series cases corresponding to a specific wave, which can then be modeled by the conventional growth curve models into a sigmoid curve (for example; **Fig. 1**).

The segmented growth curve models (GCMs) summarizes the spread of COVID-19 into sets of three parameters , and , where is the maximum number of predicted cumulative confirmed cases, is the time when we start to see a rise in the number of confirmed cases and is the rate of spread of COVID-19. Thus, each country’s COVID-19 spread pattern was summarized into three GCM parameters. Then, a regression model was employed to investigate the relationship between 31 selected national factors from *Our World in Data* (44) and Korean Statistical Information Services (KOSIS) (45) like life expectancy, average annual temperature, aging index, human development index, percentage of malnourished people in the population, extreme poverty, etc., (**Table S1**) and the spread of COVID-19, using the above parameters estimated from the segmented growth curve models. Significant relationship provides evidence that these variables may influence the spread of the novel SARS-COV-2 virus across the globe.

**Results  
  
Growth curve models predict the spread of COVID-19 across countries**

In this analysis, we adapted and applied two GCMs: Logistic and Gompertz models. Since the countries experienced more than one wave of the pandemic as of 31 August 2020, segmented GCMs were used to fit each wave independently, with each wave corresponding to a segment. Therefore, these models summarized the spread patterns of COVID-19 cumulative confirmed cases of 134 countries (see Materials and Methods).i.e. three parameters (, , ) for countries with one wave (and therefore, one segment) and six parameters (, , ; , , ) for countries with two waves (two segments) of the pandemic. Here, the differences between parameters estimated from Logistic and Gompertz models among the countries are discussed (**Table S2** and **Table S3**). Fig. 1 shows the differences between the parameter values estimated from GCMs among the countries. The x-axis represents the parameter related to the number of maximum predicted cumulative confirmed cases (), while the y-axis represents the parameter related to the rate of spread of COVID-19 ).

Parameter estimation shows that Philippines, India and Brazil having the highest numbers of maximum predicted cumulative confirmed cases in the first segment of the pandemic using Logistic (**Fig. 2A**) while India and Zambia using Gompertz (**Fig. 2B**). In the second segment of the pandemic, USA has the highest number of maximum cumulative confirmed cases using both GCMs (**Fig. 2C** and **Fig 2D**). Therefore, by August 31, 2020, USA is the country with the greatest number of cumulative confirmed cases in the world. All the other remaining countries had not notably large differences in their numbers of maximum predicted cumulative confirmed cases in both GCMs.

However, we observed a somewhat large differences in the rate of spread of COVID-19 values, among the countries. In the first segment of the pandemic, Djibouti, Malawi and New Zealand had the highest rate of spread while Sweden had the lowest rate of spread ) of COVID-19 among their populations (**Fig. 2A** and **Fig 2B**), in both models. In the second segment, Democratic Republic of Congo, Montenegro and Cote d’Ivoire (Ivory Coast) had the highest rate of spread while Iceland, Finland, UK, Nepal, Australia and Japan had the lowest rate of spread of COVID-19 in their populations, in both models (**Fig. 2C** and **Fig 2D**).

Also, we observe that countries with the greatest numbers of predicted maximum cumulative confirmed cases have the smallest rate of spread and vice versa, in both models and segments (see **Fig. S7** of v. The correlation analysis (Pearson’s correlation) to determine the relationship between the parameters across the two models and segments (**Fig. S3- S6**) confirms that the parameters have similar interpretation across models and segments but a strong negative correlation (-0.5 and -0.55 for Logistic, -0.66 and 0.7 for Gompertz) between and parameters (**Fig. S5** and **Fig. S6**) was observed. This may explain the relationship observed between the numbers of maximum predicted cumulative confirmed cases and the rate of spread of COVID-19.

Furthermore, since the first day of the analysis period was set to the date when the number of cumulative confirmed cases exceeded 50 for each country (see Methods), the population scale among countries was not considered. So, the time when we start to see a rise in the number of confirmed cases (β) does not produce consistent results between segments and models, as the other parameters do, though its interpretation is the same between the models. Therefore, its results and any analysis concerning it was not a focus in our study, and its results were relegated to the Supplementary Materials for those interested. In addition, shows minimal correlation (-0.089 for Logistic and 0.15 for Gompertz) between the two segments and with other parameters (e.g. -0.069 and 0.19 for Logistic, 0.077 and 0.10 for Gompertz) in the same model, but it shows a strong positive correlation between the models (0.88 and 0.95).

**The relationship between national factors variables and the spread of COVID-19**

Regression model was employed to investigate the relationship of selected national factors (**Table S1**) reasonably assumed to be related to COVID-19 and the spread of COVID-19, using the number of maximum predicted cumulative confirmed cases, and the rate of spread of the pandemic, estimated from the segmented GCMs. The 31 national factors included developmental (called World Development Indicators by World Bank (16)) and non-developmental variables related to population, age distributions, health, environment and so forth.

The objective of our analysis was to determine whether these factors influence the spread of COVID-19. From the 1st segment to the 2nd segment in each growth curve model, our focus was on whether 1) the differences in the size of the estimated coefficients, and 2) the estimated coefficients are statistically significant between two models and two segments. We used a 5% significance level in this analysis. Statistically significant results provide evidence for the possibility of these factors influencing the spread of COVID-19.

For the number of maximum predicted cumulative confirmed cases (), several variables turned out to be significant such as population, annual precipitation, pharmaceutical sales and imports to GDP ratio (**Fig. 3A** and **Fig. 3C**). But population was the only variable that was outstandingly significant in both segments (1, 2) and models (Logistic, Gompertz). The rate of spread of COVID-19 ) is significantly related to 19 national indicator variables. For example, age-related variables like aging index, share of population aged 65 and older, share of population aged 70 and older, median age and life expectancy, health related variables like life cardiovascular death rate, share of female and male smokers in the population and percentage of malnourished people in the population, hospital beds per thousand, extreme poverty and human development index, cultural variables like of international travelers from a country and number of foreign visitors to a country, and environmental factors like average annual temperature (**Fig. 3B** and **Fig. 3D**).

In addition, a relationship between the size of coefficient values (of the relationship between national indicator variable and GCM parameter) and significance of a national indicator variables was observed whereby significant variables generally have larger coefficient values than non-significant variables (**Fig. 4**, **Fig. S9** and **Fig. S10**). Our results provide evidence of the influence of these significant national indicator variables like population, aging index, median age, cardiovascular death rate, extreme poverty, annual precipitation, number of foreign visitors and international travelers, etc., on the spread of COVID-19 across the globe. Also, we rarely observe a change in signs of the coefficients of the significant variables between models.

The number of maximum predicted cumulative confirmed cases is significantly influenced by only population in both the two GCMs and segments of each model (**Fig. S11**). The countries with the highest value of maximum predicted cumulative confirmed cases (India, USA, Brazil, Philippines and Zambia) have the highest population sizes in the world as of August 31, 2020 (22). In addition, USA, India and Brazil respectively are the hardest hit countries with the COVID-19 pandemic in the world (23, 24) showing a relationship between population sizes and the number of confirmed cases. i.e. the spread of COVID-19. High population may bring about congestion of people and higher rate of person-to-person contacts among the people in public places. However, other population dynamic factors may bring about this observation.

The rate of spread of COVID-19 is influenced by 16 significant variables in the Gompertz model, and 10 significant variables in the Logistic model (**Fig. 5**). Age related variables. i.e. aging index, median age, percentage of the population aged 65, 70 and older and life expectancy are significant in both models and segments. Aging is linked mainly with deteriorating immune system (25) and other common conditions like hearing loss, cataracts and refractive errors, back and neck pain and osteoarthritis, chronic obstructive pulmonary disease, diabetes, depression, dementia and so forth, where several of these conditions can be experienced at the same time (26-27). The risk for severe illness with COVID-19 increases with age, with older adults are at a greater risk of requiring hospitalization and dying of COVID-19 if diagnosed than younger people due to already a deteriorating immune system, pre-existing conditions and underlying medical problems (cardiovascular disease, diabetes, chronic respiratory disease, and cancer) that also makes them prone to newer infections (28-31). This includes other variables like cardiovascular death rate (54) and the percentage of female and male smokers in the population.

One-in-five (20%) adults in the world smoke tobacco (72) and is one of the world’s largest health problems. Active smoking and a history of smoking (cigarettes, waterpipes, bidis, cigars, heated tobacco products) may be vulnerable to contracting COVID-19 and has been linked to increased severity of COVID-19 illness due to the health complications it wrecks mainly on the immune system, especially on the lungs (epithelial cells) which is primary site of target of SARS-COV-2 (32-34). Also, the act of smoking involves contact of fingers (and possibly contaminated cigarettes) with the lips, which increases the possibility of transmission of viruses from hand to mouth. Smoking waterpipes, also known as shisha or hookah, often involves the sharing of mouth pieces and hoses, which could facilitate the transmission of the COVID-19 virus in communal and social settings (33). It is reported that Montenegro has 46% smoking prevalence, a country with second highest rate of spread of acovid-19 in the second segment of analysis (72), while OECD member countries have a prevalence of 23.50% as of 2016. African countries have one of lowest smoking prevalence in world (72).

Extreme poverty impairs rapid response of the government to newer pandemics or even other disasters, leaving its people highly susceptible to the infections. It influences a governments preparedness to deal with disasters (new pandemics included) and interferes with health system response like drugs, protective gears, information campaign and the inability of poor health systems to handle newer pandemics. Malnutrition increases one's susceptibility to and severity of infections and is thus a major component of illness and death from disease. The risk of death is directly correlated with the degree of malnutrition (55-57). Malnutrition is consequently the most important risk factor for the burden of disease in developing countries. Malnutrition continues to be a major public health problem throughout the developing world, particularly in southern Asia and sub-Saharan Africa (35-36).

Number of international travelers and foreign visitors increases the chance of spreading and catching the SARS-COV-2 virus among the population (37) mainly due to importation and exportation of cases leading to many domestic travel restrictions and flight suspensions between countries (38-39). Accelerated by human migration, exported COVID-19 cases have been reported in various regions of the world, including Europe, Asia, North America, and Oceania (58). National competitiveness that covers areas such as economic performance, government efficiency, corporate efficiency, and infrastructure, influencing the rate of spread of COVID-19 may involve all the above-mentioned areas for example, government efficiency in the response to disaster may determine the overall outcome of the situation. South Korea’s response to COVID-19 especially in the early stages of the pandemic, has been widely praised and encouraged to be emulated, around the globe, which shows the importance of national competitiveness in response to COVID-19 (59-60). Although climate factors may have influenced the rate of spread of COVID-19, they may have had a smaller effect size compared to the other significant factors. As a result, climate factors did not turn out to be consistently significant across models and segments (only 1 model, 1 segment). A recent review has addressed the role of climate change in the emergence and re-emergence of infectious diseases worldwide, indicating that temperature is an important environmental condition determining the success of infectious agents (40).

**Discussion**

In this study, we investigated the relationship of 31 national factors from KOSIS and *Our World in Data*, on the spread of COVID-19 in 134 countries. First, we modelled the spread of COVID-19 using segmented Logistic and Gompertz models and then we investigated the influence of national factors variables on the spread of COVID-19. We observed that some factors were significant in both GCMs or the two segments for each model while others were significant in only one model or segment, which implies a change in segments. We believe that although the curves from GCMs can describe similar behavior in some phases of growth, one of the most important differences is that the Gompertz process is asymmetric, whereas the Logistic curve is a symmetric process, explaining the differences observed in the results of the two models. Therefore, using a given growth curve model can have a substantial impact on forecasting (22). By building two models and analyzing the results (**Fig. 5**), we concluded that our findings provide reasonable proof that the significant variables influence the spread of COVID-19.

We observed that the number of maximum predicted cumulative confirmed cases is significantly influenced by only one factor while the rate of spread of COVID-19 is influenced by seven factors, in both the two GCMs and segments of each model (**Fig. 5**). This makes the rate of spread of the pandemic the most influenced aspect of the spread of COVID-19 among countries, among the two parameters

Also, we found out that the number of maximum predicted cumulative confirmed cases () did not vary much across countries (though we observed a few outliers e.g. USA, India, Brazil, Philippines, and Zambia), while the rate of spread of COVID-19 () varied greatly across countries. We observed that is only mainly influenced by population (**Fig. S11**) while is significantly influenced by many variables (**Fig. 5**). This may explain the differences observed in the rate of spread of COVID-19 among countries than in the number of maximum predicted cumulative confirmed cases. It was seen that different variables influenced the spread of COVID-19 at different segments of the pandemic.

We see the influence of population size on the spread of COVID-19. Among the hardest hit countries with the COVID-19 pandemic in the world, USA, India and Brazil have the largest populations in the world. Some countries with the highest number of maximum predicted cumulative confirmed cases (Zambia, India, Brazil and Philippines) and the highest rate of spread of COVID-19 (Democratic Republic of Congo and Malawi), have a large percentage of their population living in extreme poverty (61-62), malnourished (63-64), but with the youngest populations (especially African countries) in the world (65). Also, in the first segment, Iceland, South Korea, China, New Zealand and Australia which had high rate of spread of COVID-19, are characterized with older populations, longer life expectancy, higher GDP per capita, higher cardiovascular death rate, large percentage of population that smoke daily (71), better health systems and little to no malnutrition (65, 66-67, 68-69). Clearly, we observe the influence of these variables on the spread of COVID-19 (16). However, most of these countries in addition to Japan, UK, Italy, Germany and United Arabs Emirates (despite have the above listed characteristics), also have the lowest rate of spread of COVID-19 in the second segment of the pandemic (**Fig. 2**). This can be due to the influence of government implemented policies like “lock-downs” in response to the spread of COVID-19.

However, there are some limitations in our analysis. For example, a key limitation of this analysis is that though we modeled the spread of COVID-19 for 134 countries, the GCMs still produced some missing parameter values (10 countries in Logistic; 14 countries in Gompertz) between the segments and models for some countries mainly due to failure of convergence (**Table S2** and **Table S3**), which may have affected comparison and therefore, interpretation of the results. Also, we would only fit the model up to August 31, 2020 because beyond that more than two segments would have to be modeled as currently many countries are experiencing their third wave or beginning their fourth wave of the pandemic, which was challenging to the Segmentation Algorithm. In future, we hope to improve on this algorithm and then be able to study the other waves of the pandemic and solve the problem of failure of convergence in the models.

Also, COVID-19 which is a contact-transmissible infectious disease and is said to spread through the population via direct contact between individuals (2, 17-18) as the main route of transmission, elicited a wide range of control measures from each country, aimed at reducing the amount of mixing in the population. These government implemented polices have already been shown to mitigate and suppress the pandemic (19-20). It was determined that highly effective contact tracing and case isolation is enough to control a new outbreak of COVID-19 within three months in most scenarios (21). Although it is very important to include these policies or to model their effects, in our analysis since these policies may have influenced the results observed from the segmented GCMs. However, our analysis could not since our approach cannot handle time dependent variables like the containment policies. Therefore, we could not control for this bias in our analysis, as some may argue about that. In future, we hope to consider the impact of government implemented policies on the spread of COVID-19 in our analysis using other models.

Furthermore, we also hope to repeat this analysis using COVID-19 number of cumulative death cases soon. The number of death cases are just as important as confirmed cases in the understanding of influential factors and epidemiological characteristics of COVID-19, for we believe that COVID-19 death cases will provide more insight as they may be more related to age distributions and health related variables.

In conclusion, a lot is yet not known about the clinical and epidemiological characteristics of COVID-19, such as individual risk factors for contracting the virus and infections from asymptotic cases. However, from the above discussions, our findings show the relationship between age distributions, life expectancy, malnutrition, extreme poverty, cardiovascular death rate, smoking and population size and the spread of COVID-19. We hope these studies will provide important information for policy makers and governments in making informed scientific decisions while considering a country economy, population dynamics, climate and health system that would likely make the most impact in future prevention works against similar infectious diseases.

**Materials and Methods**

**ECDC COVID-19 Data**

The COVID-19 data of daily confirmed cases and deaths can easily be downloaded from the European Centre for Disease Prevention and Control (ECDC) website (41-43). ECDC is an EU agency aimed at strengthening Europe's defenses against infectious diseases. The core functions cover a wide spectrum of activities: surveillance, epidemic intelligence, response, scientific advice, microbiology, preparedness, public health training, international relations, health communication, and the scientific journal *Eurosurveillance*. Negative confirmed cases were corrected to 0 regarding it as an abnormal data. Since cases on an international conveyance in Japan was included in country list, we remove it. The data consisting of 213 countries from January 1, 2020 to August 31, 2020 was used in downstream analysis.

Data smoothing is used to remove noise from a data set, allowing important patterns to stand out. Thereafter, daily confirmed case data was smoothed by simple moving the average; 1) to reduce the effect of outliers and 2) remove the weekly periodicity observed in the data. There were several outliers that showed greater or smaller abnormalities, which made it difficult to fit the statistical model. In addition, weekly periodicity was observed in the daily confirmed case data for many countries. Although we tried to present numerically through autocorrelation function, the trend had randomness giving a limit to the analysis. Therefore, considering the period of 7 days, the window size was set to 7 and simple moving average (SMA) was used before model fitting as shown below;

where *p* is the number of confirmed cases.

**National factors**

Time-independent national indicator variables (**Table S1**) datasets are publicly available datasets easily obtained from *Our World in Data* website (44) and Korean Statistical Information Services (KOSIS) (45). *Our World in Data* website provides data about Research and data to make progress against the world’s largest problems like poverty, disease, hunger, climate change, war, existential risks, etc. It mainly focuses on: the large problems that continue to confront us for centuries or much longer and the long-lasting, forceful changes that gradually reshape our world. From this website, we obtained 15 time-independent social and economic variables we believe are related to COVID-19 such as population, population density, median age, aged 65 over, aged 70 over, GDP per capita, extreme poverty, cardiovascular death rate, diabetes prevalence, female smoker, male smoker, handwashing facilities, hospital beds per thousand, life expectancy, human development index, etc. (46).

The Korean Statistical Information Service (KOSIS) (45) website contains the national statistical database which offers a full range of major domestic, international and North Korean statistics, produced by over 120 statistical agencies covering more than 500 subject matters as well as the latest data on international finance and economy from international organizations (i.e. IMF, World Bank, OECD). From the 26 variables, 13 were selected which we believed would be related to the spread of COVID-19. These variables are measured over a period of several years. Therefore, we selected the year with the minimum number of missing values between 2016-2019, re-scaled by division with standard errors of the variables**.**

**Analysis of spread of COVID-19 using GCMs**

Under this analysis, growth curve models (GCMs), Logistic model and Gompertz model were employed to model the transmission of COVID-19 using the cumulative confirmed cases for each country. These growth models are commonly used to explore risk factors, predict the probability of occurrence of a certain disease, factors that control and affect growth, and extinction laws of the population respectively (47).The models take the following forms respectively;

Logistic model

(1)

where is the cumulative confirmed cases, is the maximum number of predicted cumulative confirmed cases, *b* is the time when we start to see a rise in the number of confirmed cases, *c* is the increase rate of number of confirmed cases, *t* is the number of days since the first case occurrence, is the time when the first case occurred.

Gompertz model

(2)

where is the cumulative confirmed cases, is the maximum number of predicted cumulative confirmed cases, *b* is the time when we start to see a rise in the number of confirmed cases, *c* is the increase rate of number of confirmed cases, *t* is the number of days since first case. is the time when the first case occurred.

**Segmentation Algorithm**

As the COVID-19 situation prolongs, fitting a growth curve model on daily confirmed cases over long period of time becomes impossible as it no longer takes on s-curve (i.e. sigmoid function). To fit the above growth curve models, there is a need to divide study period of countries experiencing more than one wave (48) (a wave implies a rising number of sick individuals, a defined peak, and then a decline) of the pandemic into several segments (the time during which cumulative confirmed cases follow the s-curve). So, we applied segmentation algorithm which can systematically divide study periods into several segments (or waves) for each country **(Fig. 1).**

Segmentation is a method of finding peaks and breakpoints, where a peak is the timestamp at which daily new confirmed case is highest in a segment and breakpoint is the timestamp which splits the consecutive two segments in a time series dataset. To better see trends, we smooth out the irregular roughness of the graph of daily confirmed cases. But daily new confirmed cases have high randomness arising from 1) the fact that daily new confirmed cases have a periodicity of seven days (due to differences in daily new confirmed cases between weekends and weekdays) and 2) measure errors of one day. Therefore, we applied the Nadaraya-Watson kernel regression Estimator (NWE) (49-51) with Gaussian kernel to smoothen the daily new confirmed cases as demonstrated in Fig. 6 using South Korea’s daily confirmed cases as an example. For the convenience of notation, let be the -th daily new confirmed cases from data, be the estimated -th daily new confirmed cases using above NWE since January 1, 2020.

Peak detection (**Algorithm 1**; **Fig. S1**) utilizes the first and second derivative test to find local maxima on convex function. has convexity when is around peak due to the nature of epidemic dynamics. Considering daily new confirmed cases being discrete time series data, we find the location where the first difference is zero and second difference is negative (Since is not differentiable, we used difference operator instead of derivative):

, (3)

where and .

And for discontinuity and small variances of , we used following condition:

, (4)

where is sensitivity level and is set of time indices from January 1, 2020 to August 31, 2020 (**Fig. 2**). And 3 additional conditions ((a) Exclusion of small peaks, (b) Resolution criteria and (c) Exclusion of peaks which are vibrations on increasing trend) are used in peak detection to enhance robustness. After all the peaks are found, breakpoints (**Algorithm 2; Fig. S1**) are selected either as timestamps which have the smallest daily new confirmed cases between two consecutive peaks or the timestamp where the cumulative confirmed case of the last segment saturates (that is the last stage of the s-curve of last segment*).* **Fig. S2** visualizes the segmentation process. Blue line represents the peak and dotted sky-blue line represents breakpoint. In the 1st plot, black solid line represents and black dotted line represents . The 2nd plot represents cumulative confirmed cases of (black dotted line), (black solid line). 3rd, 4th plots are graphs of . In 4th plot, green dotted line represents sensitivity level. If is above the upper green dotted line, is concave. On the other hand, if is below the lower green dotted line, is convex. With 3rd, 4th plot, Equation (5) can be validated. The segmentation algorithm was successfully applied to 134 countries from the 213 countries in the ECDC dataset which met . If is too small, segmentation algorithm would be difficult to apply due to small variances in .

**Segmented growth curve models**

Segmented growth curve models (segmented Logistic model and segmented Gompertz model) fit the above-mentioned growth curve models ((1) and (2)) for each segment independently. These new models do not preserve continuity at breakpoints, but this does not matter since the objective of our analysis is to condense daily new confirmed cases into several parameters () of the growth curves, not to accurately predict daily new confirmed cases.

(5)

(6)

where, is the number of cumulative cases at breakpoint, is indicator function where is the set of indices of segment and .

In this analysis, we considered 1st and 2nd segments only since most countries have 1 or 2 segments (1 segment: 62, 2 segments: 65, 3 segments: 7) The number of countries with three segments were very few making comparison analysis insignificant to use in the regression analysis. For countries with more than 2 segments, the analysis period was therefore cut off at the 2nd breakpoint. For countries with 2 segments, segmented growth curve model then produces two sets of parameters one set from each segment.

After the segmentation algorithm was applied to 134 countries, these countries were fitted to segmented Logistic and Gompertz model respectively. To filter out poorly fitted countries, we exclude countries whose MSSE (Mean Squared Scaled Error) is higher than 0.4 as defined below:

MSSE = (7)

where is the daily new confirmed cases, is the predicted value for by segmented Logistic and Gompertz Model, and is the mean of for t = 1, …, N.

MSSE is a more suitable measure compared with MSE (Mean Squared Error) or MAPE (Mean Absolute Percentage Error) because the MSE doesn’t consider scales of population among each country, while MAPE overestimates its error when the number of daily new confirmed cases, is small. Among the 134 countries, 124 countries were fitted for segmented Logistic model and 119 countries for segmented Gompertz model. Among the fitted countries, 5 countries were excluded due to failure of meeting the MSSE criteria of 0.4 for segmented Logistic, segmented Gompertz models respectively. Therefore, a total of 119 countries were used in the segmented Logistic model, and 114 countries in the segmented Gompertz model, as shown in **Fig. 7**. In addition, correlation analysis for segmented Logistic and Gompertz models with the log-scaled of parameters are performed to determine the similarity between parameters of the two models (see **Fig. S3** - **Fig. S6**).

**Regression model**

The above segmented growth curve models summarize the spread of the pandemic into three parameters (, , ) for countries with one segment and into six parameters (, , , , , ) for countries with two segments. Each of the parameters from the two segmented GCMs was regressed against the national indicator variables shown in Table 1 as follows;

(8)

where is one of segmented GCM parameters () for model (Logistic), (Gompertz), segment and country . and  are regression coefficients, and is the time-independent variable of country . F- statistic is performed to test the significance of for each time-independent variable with the aim of finding out which the variables have significant relationship with *y*, a measure of the spread dynamics of COVID-19 for a country.Paste your materials and methods section here.

**Acknowledgments**

**Funding:**

Science & Technology Policy Institute of Korea (No. 0409-20200222)

**Author contributions:**

Conceptualization: T.P.

Methodology: T.P., H.K., Y.K.

Investigation: H.K., Y.K., K.H., H.C., T.G., A.C., D.L., J.L., T.K

Visualization: H.K., K.H., Y.K., A.C.

Supervision: T.P.

Data Curation: K.H., G.H., T.G., H.K., Y.K., D.L.

Writing—original draft: A.C., H.K., Y.K.

Writing—review & editing: T.P., C.A. and G.H., J.L.

Project Administration: T.P.

Funding Acquisition: T.P.

**Competing interests:**

The authors declare no competing interests.

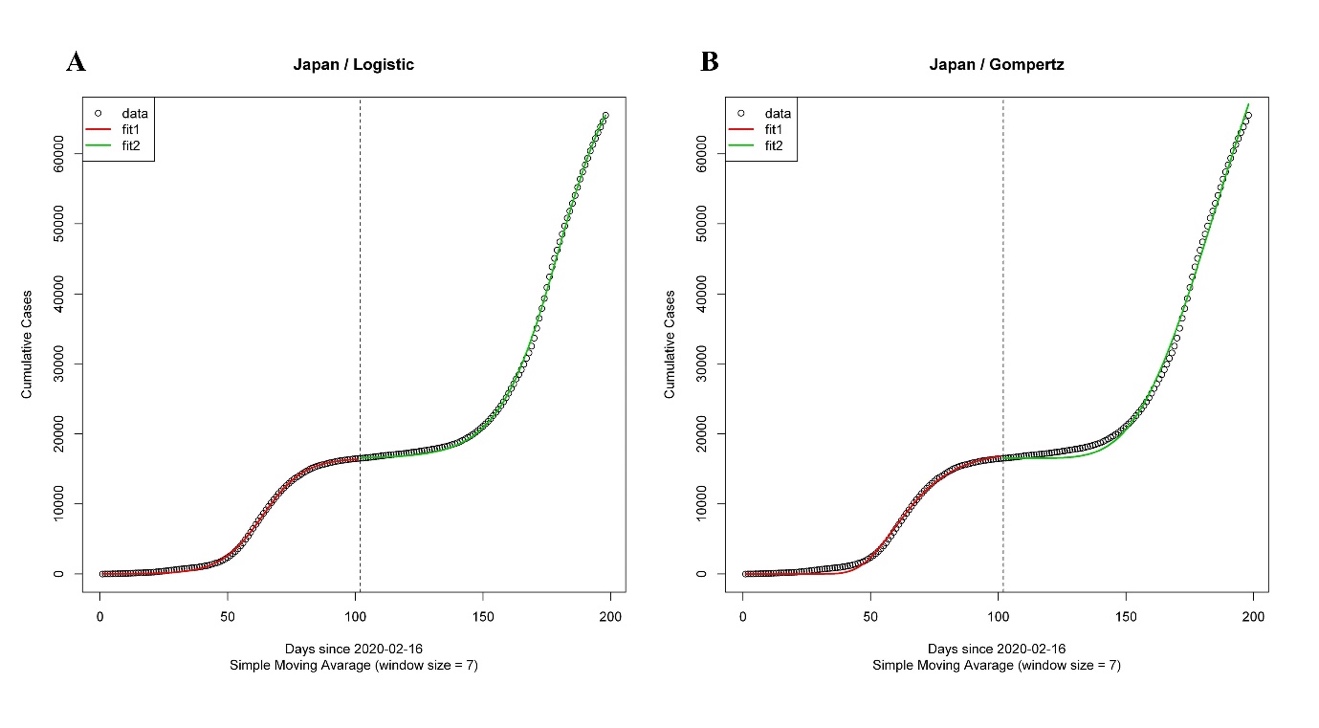
**Data and materials availability:**

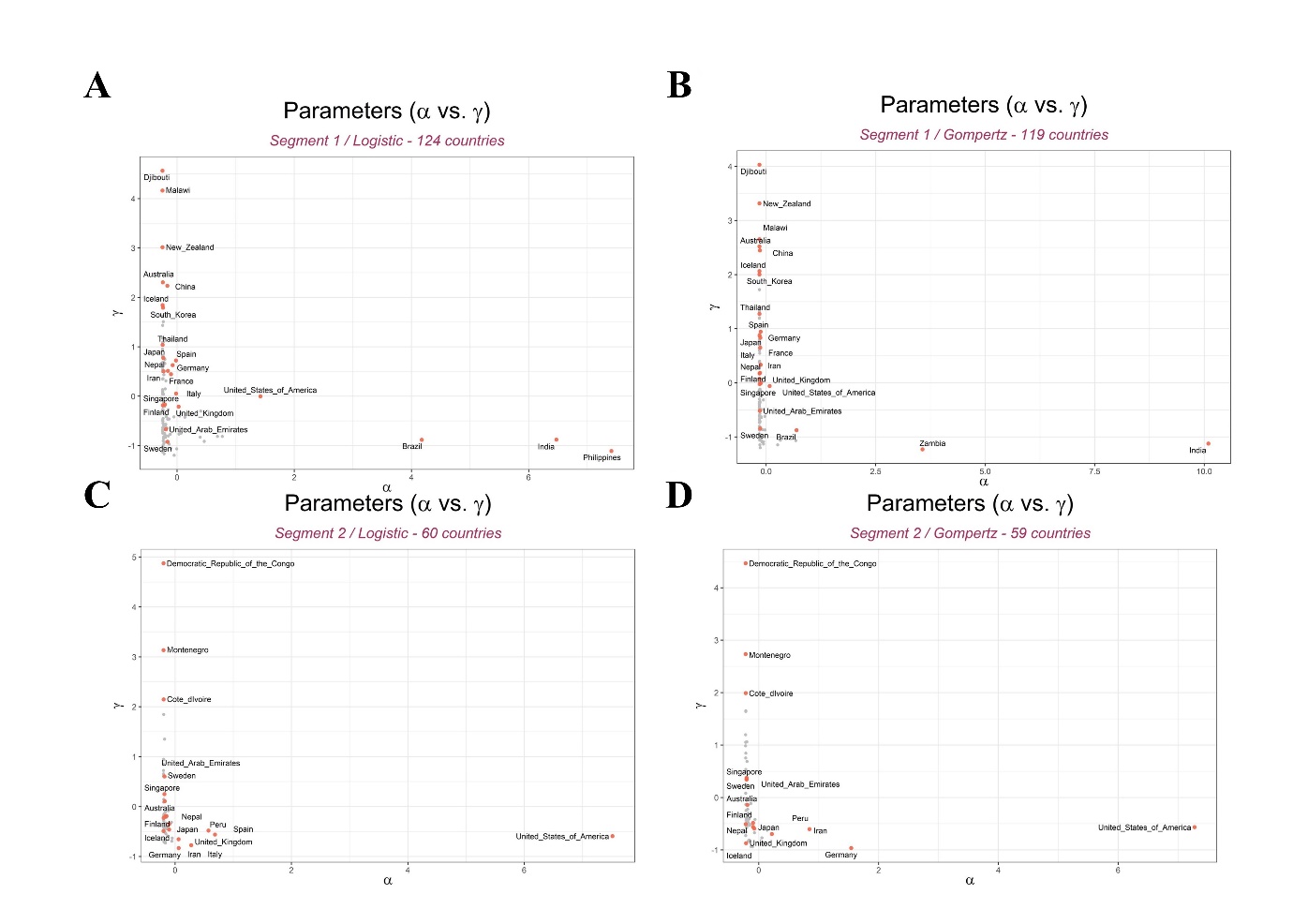
Publicly available datasets were analyzed in this study. These datasets can be found at the data links provided in the references. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. All additional data used are available from the authors.

**References**

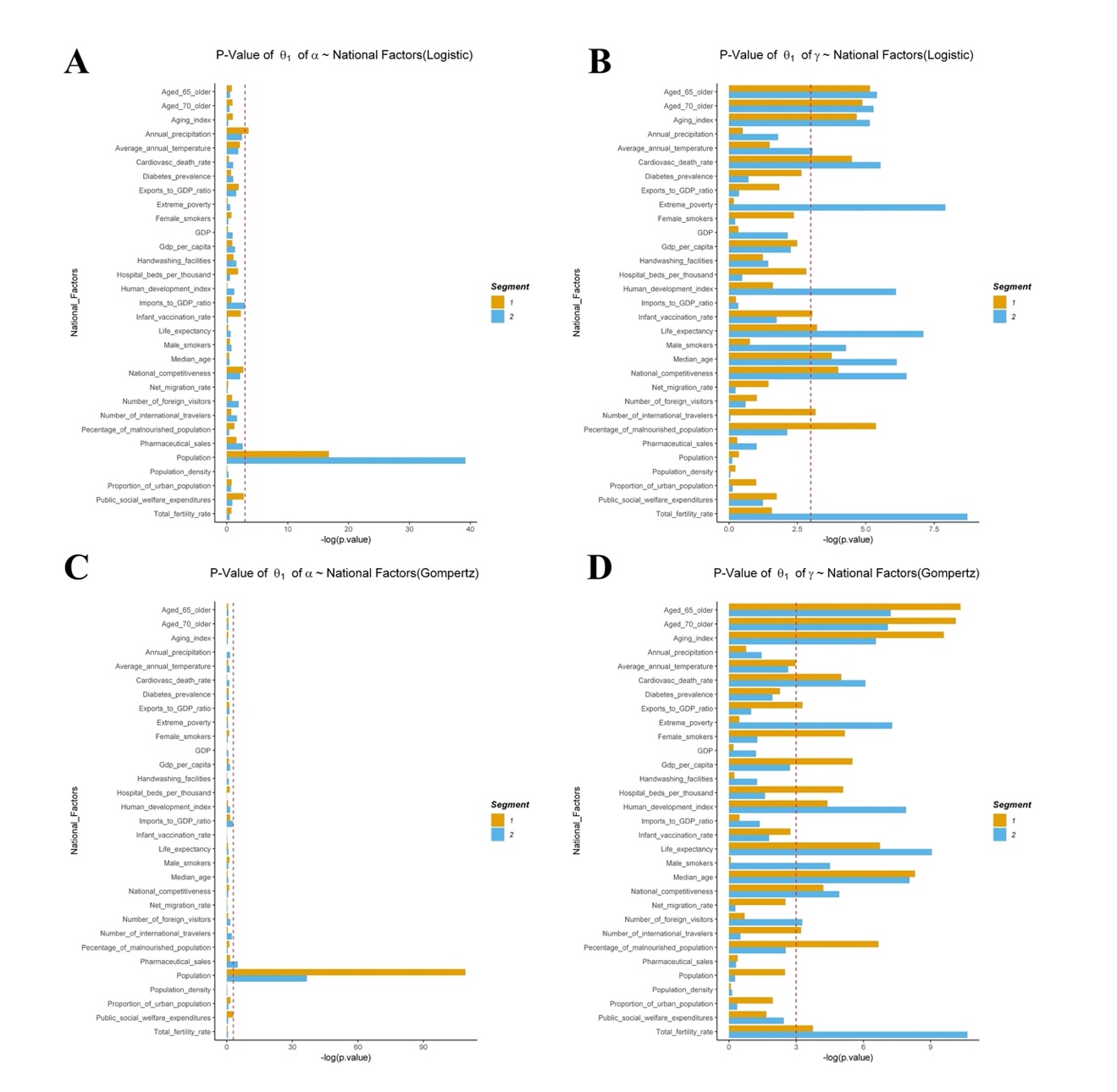
1. S. Lakshmi Priyadarsini, M. Suresh, Factors influencing the epidemiological characteristics of pandemic COVID 19: A TISM approach. International Journal of Healthcare Management 13:2, 89-98, DOI: 10.1080/20479700.2020.1755804 (2020).
2. Q. Li, X. Guan, P. Wu, et al., Early transmission dynamics in Wuhan, China of the novel Coronavirus- infected pneumonia. N Engl J Med, DOI:10.1056/NEJMoa2001316 (2020).
3. WHO director-general’s opening remarks. [https://www.who.int/dg/speeches/detail/whComing soono-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19-11-march-2020](https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19-11-march-2020)
4. Worldometer. <https://www.worldometers.info/coronavirus/> (Accessed Feb 08, 2021)
5. M. C. J. Bootsma, N. M. Ferguson, The effect of public health measures on the 1918 influenza pandemic in U.S. cities. Proc Natl Acad Sci USA 104(18):7588–93 (2007).
6. H. Xu, C. Yan, Q. Fu, K. Xiao, Y. Yu, D. Han, W. Wang, J. Cheng, Possible environmental effects on the spread of COVID-19 in China. Science of The Total Environment 731, 139211, ISSN 0048-9697. <https://doi.org/10.1016/j.scitotenv.2020.139211> (2020).
7. L. Y. K. Nakada, R. C. Urban, COVID-19 pandemic: environmental and social factors influencing the spread of SARS-CoV-2 in São Paulo, Brazil. Environ Sci Pollut Res, <https://doi.org/10.1007/s11356-020-10930-w> (2020).
8. M. Ahmadi, A. Sharifi, S. Dorosti, S. F. Ghoushchi, N. Ghanbari, Investigation of effective climatology parameters on COVID-19 outbreak in Iran. Sci Total Environ 729:138705. <https://doi.org/10.1016/j.scitotenv.2020.138705> (2020).
9. S. Hamidi, S. Sabouri, R. Ewing, Does density aggravate the COVID-19 pandemic? J Am Plan Assoc 1–15. <https://doi.org/10.1080/01944363.2020.1777891> (2020).
10. A. Tobías, Evaluation of the lockdowns for the SARS-CoV-2 epidemic in Italy and Spain after one month follow up. Sci Total Environ 725:138539. <https://doi.org/10.1016/j.scitotenv.2020.138539> (2020).
11. N. N. Harmooshi, K. Shirbandi, F. Rahim, Environmental concern regarding the effect of humidity and temperature on 2019-nCoV survival: fact or fiction. Environ Sci Pollut Res 27:36027–36036. <https://doi.org/10.1007/s11356-020-09733-w> (2020).
12. M. Pramanik, P. Udmale, P. Bisht, K. Chowdhury, S. Szabo, I. Pal, Climatic factors influence the spread of COVID-19 in Russia. International Journal of Environmental Health Research, DOI: 10.1080/09603123.2020.1793921 (2020).
13. D. N. Prata, W. Rodrigues, P. H. Bermejo, Temperature significantly changes COVID-19 transmission in (sub)tropical cities of Brazil. Sci Total Environ 729:138862, <https://doi.org/10.1016/j.scitotenv.2020.138862> (2020).
14. C. Cheng, J. Barceló, A. S. Hartnett, et al., COVID-19 Government Response Event Dataset (CoronaNet v.1.0). Nat Hum Behav <https://doi.org/10.1038/s41562-020-0909-7> (2020).
15. Institute of Medicine (US) Forum on Microbial Threats. Ethical and Legal Considerations in Mitigating Pandemic Disease: Workshop Summary. Washington (DC): National Academies Press (US), Learning from Pandemics Past. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK54171/> (2007).
16. The World Bank. DataBank. <https://databank.worldbank.org/home> (Accessed Feb 08, 2021)
17. J. Riou, C. L. Althaus, Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. Euro Surveill, 25, Article 2000058 (2020).
18. J. F. W. Chan, S. Yuan, K. H. Kok, et al., A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet 395, 514-523 (2020).
19. Zhang, Xiaolei, R. Ma, L. Wang, Predicting turning point, duration and attack rate of COVID-19 outbreaks in major Western countries. Chaos, Solitons & Fractals: 109829 (2020).
20. Singh, Rajesh, R. Adhikari, Age-structured impact of social distancing on the COVID-19 epidemic in India. arXiv preprint arXiv:2003.12055 (2020).
21. J. Hellewell, S. Abbott, A. Gimma, N. I. Bosse, C. I. Jarvis, T. W. Russell, J. D. Munday, A. J. Kucharski, W. J. Edmunds, F. Sun, et al., Feasibility of controlling covid-19 outbreaks by isolation of cases and contacts. The Lancet Global Health (2020).
22. P. H. Franses, A method to select between Gompertz and logistic trend curves. Technological Forecasting and Social Change 46, Issue 1, 45-49. ISSN 0040-1625, <https://doi.org/10.1016/0040-1625(94)90016-7> (1994).
23. Countries in the World by Population (2021). <https://www.worldometers.info/world-population/population-by-country/> (Accessed Feb 08, 2021)
24. Worldometer. <https://www.worldometers.info/coronavirus/#countries> (Accessed Feb 08, 2021)
25. Coronavirus Disease-19, Republic of Korea. <http://ncov.mohw.go.kr/en/> (Accessed Feb 08, 2021)
26. G. Alderton, Aging immunity. Science 369, Issue 6501, 264-266. DOI: 10.1126/science.369.6501.264-k
27. World Health Organization. <https://www.who.int/news-room/fact-sheets/detail/ageing-and-health> (Accessed Feb 08, 2021)
28. K. Hirokawa, M. Utsuyama, M. Kasai, C. Kurashima, Aging and Immunity. Pathology International, 42: 537-548. <https://doi.org/10.1111/j.1440-1827.1992.tb03103.x> (1992).
29. Centers for Disease Control and Prevention. COVID-19. <https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/older-adults.html> (Accessed Feb 08, 2021)
30. A. L. Mueller, M. S. McNamara, D. A. Sinclair, Why does COVID-19 disproportionately affect older people? Aging 12(10), 9959–9981. <https://doi.org/10.18632/aging.103344> (2020).
31. N. G. Davies, P. Klepac, Y. Liu, et al., Age-dependent effects in the transmission and control of COVID-19 epidemics. Nat Med 26, 1205–1211 <https://doi.org/10.1038/s41591-020-0962-9> (2020).
32. F. K. Ho, F. Petermann-Rocha, S. R. Gray, B.D. Jani, S. V. Katikireddi, C. L. Niedzwiedz, et al., Is older age associated with COVID-19 mortality in the absence of other risk factors? General population cohort study of 470,034 participants. PLoS ONE 15(11): e0241824. <https://doi.org/10.1371/journal.pone.0241824> (2020).
33. A. Gülsen, B. A. Yigitbas, B. Uslu, D. Drömann, O. Kilinc, The Effect of Smoking on COVID-19 Symptom Severity: Systematic Review and Meta-Analysis. Pulmonary Medicine 2020, Article ID 7590207, 11, <https://doi.org/10.1155/2020/7590207> (2020).
34. T. Wenzl, Smoking and COVID-19 - A review of studies suggesting a protective effect of smoking against COVID-19. EUR 30373 EN, Publications Office of the European Union, Luxembourg, ISBN 978-92-76-22062-6 (online), doi:10.2760/564217 (online), JRC121837 (2020).
35. O. Müller, M. Krawinkel, Malnutrition and health in developing countries. CMAJ 2;173(3):279-86. doi: 10.1503/cmaj.050342. PMID: 16076825; PMCID: PMC1180662 (2005).
36. J. Saunders, T. Smith, Malnutrition: causes and consequences. Clin Med (Lond) 0(6):624-7. doi: 10.7861/clinmedicine.10-6-624. PMID: 21413492; PMCID: PMC4951875 (2010).
37. Centers for Disease Control and Prevention. COVID-19. <https://www.cdc.gov/coronavirus/2019-ncov/travelers/travel-during-covid19.html> (Accessed Feb 08, 2021)
38. A. Anzai, T. Kobayashi, N. M. Linton, R. Kinoshita, K. Hayashi, A. Suzuki, Y. Yang, S. M. Jung, T. Miyama, A. R. Akhmetzhanov, H. Nishiura, Assessing the Impact of Reduced Travel on Exportation Dynamics of Novel Coronavirus Infection (COVID-19). Journal of Clinical Medicine 9(2):601. <https://doi.org/10.3390/jcm9020601> (2020).
39. A. F. Siegenfeld, Y. Bar-Yam, The impact of travel and timing in eliminating COVID-19. Commun Phys 3, 204, <https://doi.org/10.1038/s42005-020-00470-7> (2020).
40. A. El-Sayed, M. Kamel, Climatic changes and their role in emergence and re-emergence of diseases. Environ Sci Pollut Res 27:22336–22352. <https://doi.org/10.1007/s11356-020-08896-w> (2020).
41. European Centre for Disease Prevention and Control. <https://www.ecdc.europa.eu/en/covid-19-pandemic> (Accessed Feb 08, 2021)
42. European Centre for Disease Prevention and Control. <https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide>. (Accessed Feb 08, 2021)
43. European Centre for Disease Prevention and Control. <https://opendata.ecdc.europe.eu/covid19/casedistribution/csv.> (Accessed Feb 08, 2021)
44. Our World in Data. <https://ourworldindata.org/> (Accessed Feb 08, 2021)
45. KOrean Statistical Information Service. <http://kosis.kr/eng/> (Accessed Feb 08, 2021)
46. <https://github.com/owid/covid-19-data/tree/master/public/data> (Accessed Feb 08, 2021)
47. L. Jia, K. Li, Y. Jiang, X. Guo, T. Zhao, Prediction and analysis of Coronavirus Disease 2019 (2019).
48. The Conversation. [*https://theconversation.com/what-makes-a-wave-of-disease-an-epidemiologist-explains-141573*](https://theconversation.com/what-makes-a-wave-of-disease-an-epidemiologist-explains-141573)
49. E. A. Nadaraya, On Estimating Regression. Theory of Probability and Its Applications 9 (1): 141–2. doi:10.1137/1109020 (1964).
50. G. S. Watson, Smooth regression analysis, Sankhyā: The Indian Journal of Statistics, Series A. 26 (4): 359–372. JSTOR 25049340 (1964).
51. J. Bierens, Herman, The Nadaraya–Watson kernel regression function estimator. Topics in Advanced Econometrics. New York: Cambridge University Press. 212–247. ISBN 0-521-41900-X (1994).
52. Stier, Andrew, Berman, G. Marc, Bettencourt, Luis, COVID-19 Attack Rate Increases with City Size. Mansueto Institute for Urban Innovation Research Paper No. 19, Available at SSRN: <https://ssrn.com/abstract=3564464> (2020).
53. G. Mitacchione, M. Schiavone, A. Curnis, M. Arca, S. Antinori, A. Gasperetti, G. Mascioli, P. Severino, F. Sabato, M. M. Caracciolo, G. Arabia, L. D'Erasmo, M. Viecca, M. Mancone, M. Galli, G. B. Forleo, Impact of prior statin use on clinical outcomes in COVID-19 patients: data from tertiary referral hospitals during COVID-19 pandemic in Italy. Journal of Clinical Lipidology, ISSN 1933-2874, <https://doi.org/10.1016/j.jacl.2020.12.008> (2020).
54. J. V. Broeck, R. Eeckels, J. Vuylsteke, Influence of nutritional status on child mortality in rural Zaire. Lancet 341, 1491-5.CrossRefPubMedGoogle Scholar (1993).
55. W. D. Man, M. Weber, A. Palmer, G. Schneider, R. Wadda, S. Jaffar, et al., Nutritional status of children admitted to hospital with different diseases and its relationship to outcome in The Gambia, West Africa. Trop Med Int Health 3, 678-86 (1998).
56. I. D. Fernandez, J. H. Himes, M. D. Onis, Prevalence of nutritional wasting in populations: building explanatory models using secondary data. Bull World Health Organ 80, 282-91 (2002).
57. World Health Organization, Novel Coronavirus (2019-nCoV) Situation Report-21, <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200210-sitrep-21-ncov.pdf> (2020).
58. J. You, Lessons From South Korea’s Covid-19 Policy Response. The American Review of Public Administration 50(6-7), 801-808, doi:10.1177/0275074020943708 (2020).
59. V. Cha, D. Kim, A Timeline of South Korea’s Response to COVID-19, Center for Strategic and International Studies, <https://www.csis.org/analysis/timeline-south-koreas-response-covid-19> (2020).
60. The World Bank. <https://www.worldbank.org/en/topic/poverty/overview#:~:text=The%20global%20extreme%20poverty%20rate,%245.50%20a%20day%20in%202017>. (Accessed Feb 04, 2021)
61. Our World in Data. Global Extreme Poverty. <https://ourworldindata.org/extreme-poverty#:~:text=We%20can%20also%20see%20that,of%20important%20changes%20across%20time>. (Accessed Feb 04, 2021)
62. COVID-19. <https://www.unicef.org/press-releases/west-and-central-africa-more-15-million-cases-acute-malnutrition-expected-2020> (Accessed Feb 04, 2021).
63. K. Annan, Data can help to end malnutrition across Africa. Nature 555, Gale Academic OneFile. [Accessed 4 Feb. 2021] (2018).
64. World By Map. <https://www.citypopulation.de/en/world/bymap/youngpopulation/> (Accessed Feb 04, 2021).
65. Worldometer. Life Expectancy of the World Population. <https://www.worldometers.info/demographics/life-expectancy/> (Accessed Feb 04, 2021).
66. Worldometer. GDP by Country. <https://www.worldometers.info/gdp/gdp-by-country/> (Accessed Feb 04, 2021).
67. OECD/World Health Organization, “Mortality from cardiovascular disease”, in Health at a Glance: Asia/Pacific 2018: Measuring Progress towards Universal Health Coverage, OECD Publishing, Paris. DOI: <https://doi.org/10.1787/health_glance_ap-2018-11-en> (2018).
68. J. A. Finegold, P. Asaria, D. P. Francis, Mortality from ischaemic heart disease by country, region, and age: statistics from World Health Organisation and United Nations. International journal of cardiology, 168(2), 934–945. <https://doi.org/10.1016/j.ijcard.2012.10.046> (2013).
69. World Health Organization. Coronavirus disease (COVID-19): Tobacco. <https://www.who.int/westernpacific/news/q-a-detail/coronavirus-disease-covid-19-tobacco#:~:text=Tobacco%20smokers%20(cigarettes%2C,communal%20and%20social%20settings>. (Accessed Feb 08, 2021)
70. School of Public Health. University of Michigan. What Makes a “Wave” of Disease? An Epidemiologist Explains. <https://sph.umich.edu/pursuit/2020posts/what-makes-a-wave-of-disease.html> (Accessed Feb 08, 2021)
71. Percentage of population who smoke on a daily basis in selected countries as of 2017. <https://www.statista.com/statistics/236600/proportion-of-smokers-in-the-population-of-selected-countries/> (Accessed Feb 08, 2021)
72. Our World in Data. In which countries do people smoke the most? <https://ourworldindata.org/which-countries-smoke-most> (Accessed Feb 08, 2021)

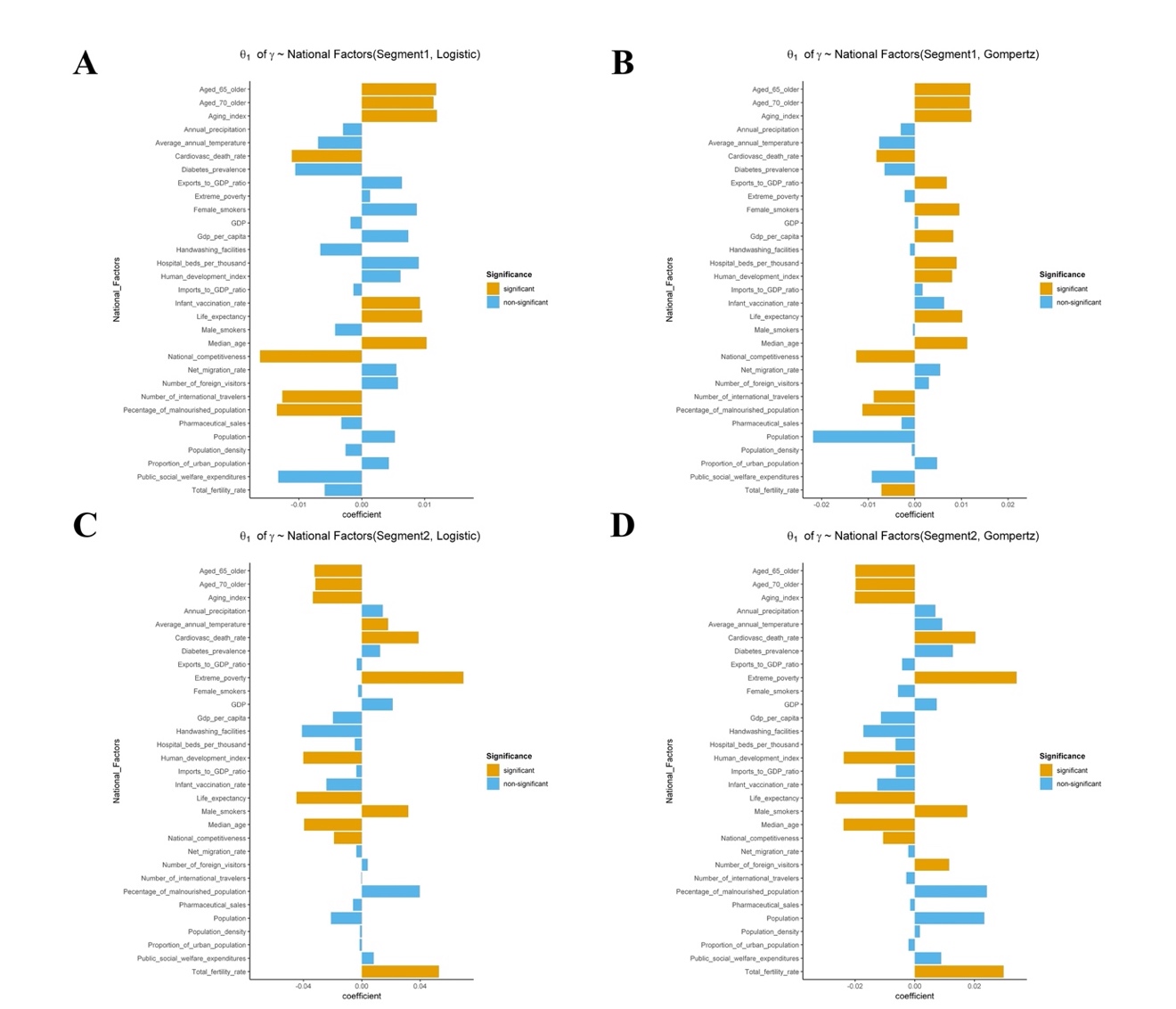
**Figures and Tables**

**Fig. 1. Cumulative confirmed cases divided into two segments using the Segmentation Algorithm.** Japan is the typical country with two waves (red: 1st segment and green: 2nd segment). **(A)** Epidemic segmented growth curve of COVID-19 modelled by Logistic model. Estimated parameters were ) = (16549.7588, 7.8244, 0.1231, 58877.8030, 6.7353, and 0.0859 respectively). (**B**) Epidemic segmented growth curve of COVID-19 modeled by Gompertz model. Estimated parameters were ) = (17531.5024, 89.4251, 0.0760, 98622.3568, 15.6933, and 0.0325 respectively).

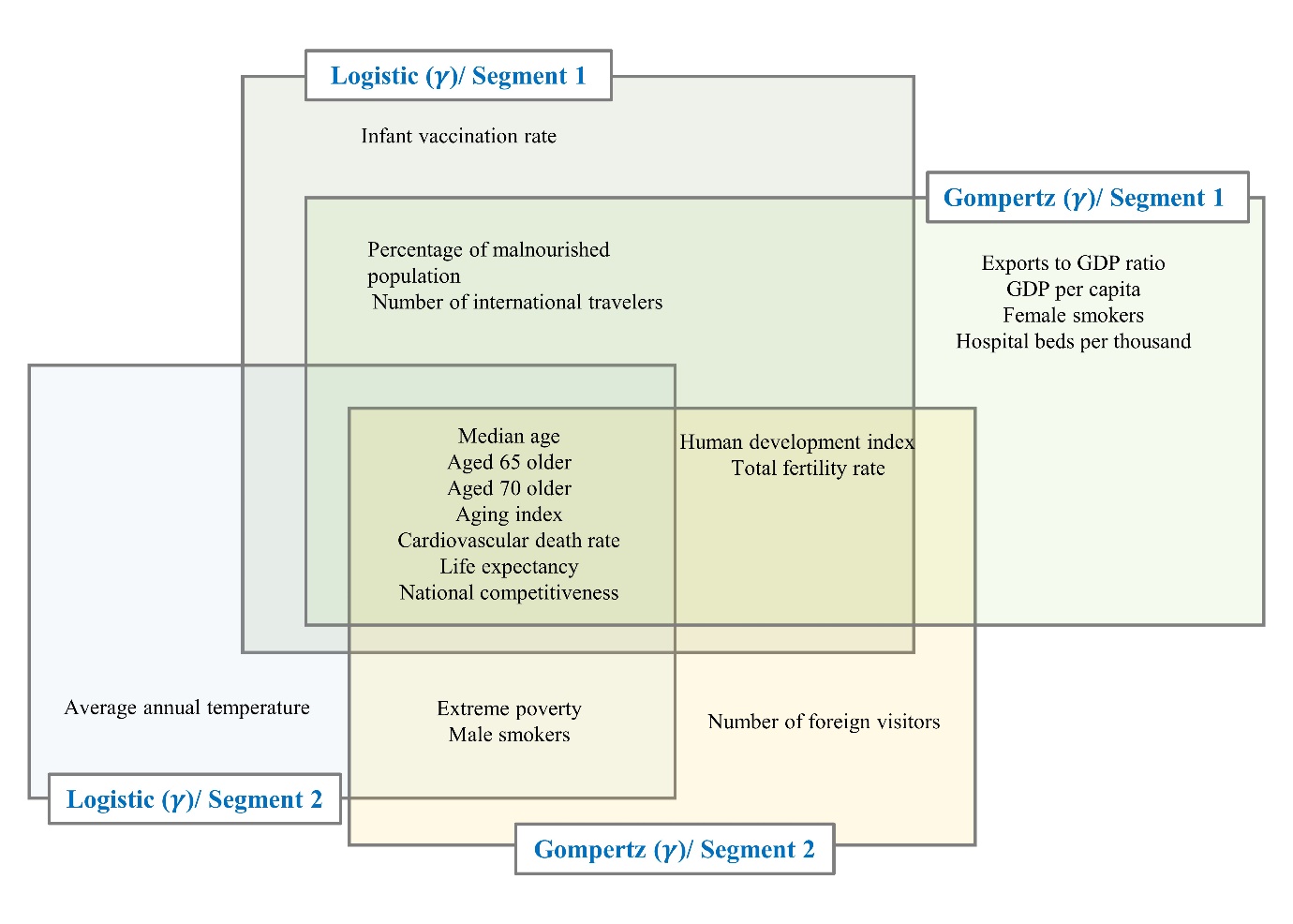


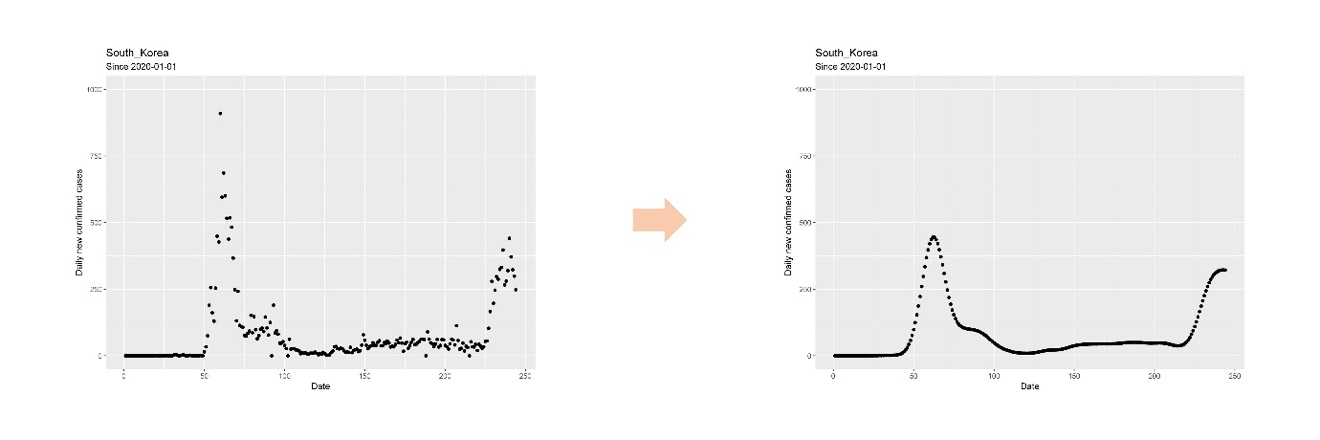
**Fig. 2. Differences in values of maximum predicted cumulative cases () and rate of spread of COVID-19 observed, among countries.**

  
**Fig. 3. *P*-values of coefficients of national factors with and .** Population, annual precipitation, pharmaceutical sales and imports to GDP ratio are statistically significant with , the number of maximum predicted confirmed cases (**Fig. 3A** and **Fig.3C**). Age related factors, population, percentage of malnourished population, life expectancy, temperature, etc. are significantly related with , rate of spread of COVID-19(**Fig.3B** and **Fig.3D**). (see **Fig.S8** for therelationship of national factors with )

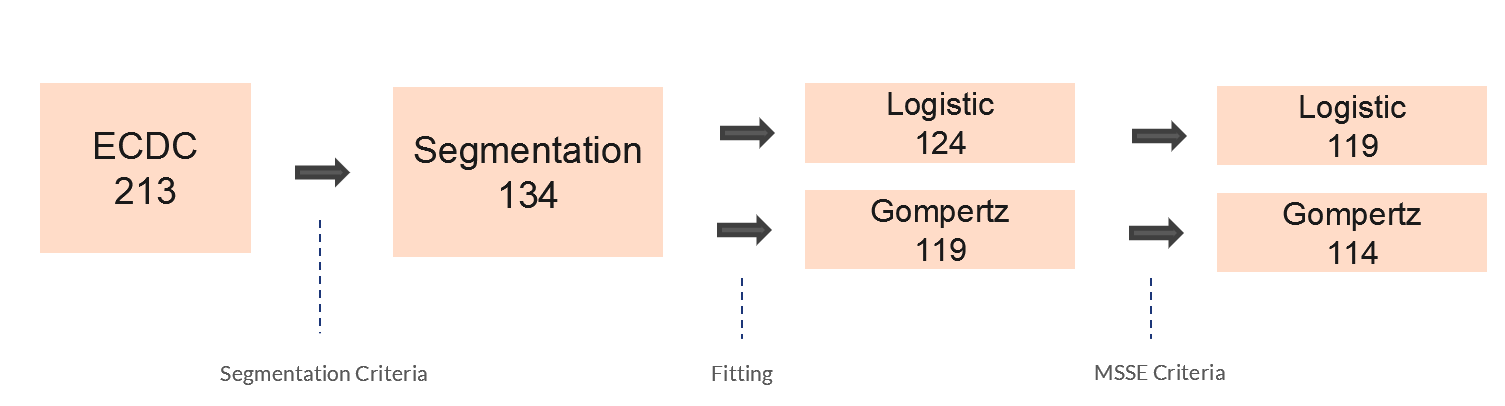


**Fig.4. Coefficients of the relationship between national factors and .** is the coefficient of the relationship between a national factor and a GCM parameter. Significant national factors (orange) have large coefficients compared with non-significant factors(blue) in both Gompertz (**Fig.4A** and **Fig.4C**) and Logistic models (**Fig.4B** and **Fig.4D**).

**Fig.5.** **Significant national indicator variables with rate of spread of COVID-19 (γ).** Median age, aged 65 older, aged 70 older aging index, cardiovascular death rate life expectancy and national competitiveness are the only national factors significant across the two models and segments. (see **Fig. S12** for significant national indicator variables with β)



**Fig.6. Daily new confirmed cases before and after smoothing using Nadaraya-Watson kernel regression**.



**Fig.7. Differences in the number countries across segmentation, growth curve models and MSSE criteria**. For segmented logistic model, 124 countries were fitted and for segmented Gompertz model, 119 countries were fitted. To check(validate) the goodness of fit of above 2 models, MSSE (Mean Squared Scaled Error) criteria was employed(used?). For each of two models, 5 countries showed high MSSE, so those 5 countries were excluded in the subsequent analysis (Aruba, Equatorial Guinea, Krygyzstan, Rwanda, Thailand for segmented logistic model and China, Equatorial Guinea, Kyrgyzstan, Rwanda, Zambia for segmented Gompertz model).