



Article

# Which National Factors Are Most Influential in the Spread of COVID-19?

Hakyong Kim <sup>1,†</sup>, Catherine Apio <sup>2,†</sup>, Yeonghyeon Ko <sup>3,4,†</sup>, Kyulhee Han <sup>2</sup>, Taewan Goo <sup>2</sup>, Gyu Jin Heo <sup>2</sup>, Taehyun Kim <sup>3</sup>, Hye Won Chung <sup>5</sup>, Doeun Lee <sup>2</sup>, Jisun Lim <sup>6</sup> and Taesung Park <sup>3,\*</sup>

- <sup>1</sup> Department of Industrial Engineering, Seoul National University, Seoul 08826, Korea; gkrdyd111@snu.ac.kr  
<sup>2</sup> Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul 08826, Korea; 2019-20240@snu.ac.kr (C.A.); hgh1031@snu.ac.kr (K.H.); gootec92@snu.ac.kr (T.G.); hgi0106@snu.ac.kr (G.H.); 2corin417@snu.ac.kr (D.L.)  
<sup>3</sup> Department of Statistics, Seoul National University, Seoul 08826, Korea; newstellar@snu.ac.kr (Y.K.); qeeqee10@snu.ac.kr (T.K.)  
<sup>4</sup> Department of Archeology and Art History, Seoul National University, Seoul 08826, Korea  
<sup>5</sup> Department of Chemistry, Seoul National University, Seoul 08826, Korea; hyewon.chung@snu.ac.kr  
<sup>6</sup> The Research Institute of Basic Sciences, Seoul National University, Seoul 08826, Korea; swanjslim@gmail.com  
\* Correspondence: tspark@stats.snu.ac.kr or tspark@snu.ac.kr; Tel.: +82-2-880-8924; Fax: +82-2-883-6144  
† These authors contributed equally to this work as co-first authors.

**Abstract:** The outbreak of the novel COVID-19, declared a global pandemic by WHO, is the most serious public health threat seen in terms of respiratory viruses since the 1918 H1N1 influenza pandemic. It is surprising that the total number of COVID-19 confirmed cases and the number of deaths has varied 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, aging index, human development index, percentage of malnourished people in the population, extreme poverty, economic ability, health policy, population, age distributions, etc.) influenced the spread of COVID-19 through systematic statistical analysis. First, we employed segmented growth curve models (GCMs) to model the cumulative confirmed cases for 134 countries from 1 January to 31 August 2020 (logistic and Gompertz). Thus, each country's COVID-19 spread pattern was summarized into three growth-curve model parameters. Secondly, we investigated the relationship of selected 31 national factors (from KOSIS and *Our World in Data*) to these GCM parameters. Our analysis showed that with time, the parameters were influenced by different factors; for example, 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 was influenced by aging index, cardiovascular death rate, extreme poverty, median age, percentage of population aged 65 or 70 and older, and so forth. We hope that with their consideration of a country's resources and population dynamics that our results will help in making informed decisions with the most impact against similar infectious diseases.

**Keywords:** COVID-19; SARS-CoV-2; growth curve models; pandemic



**Citation:** Kim, H.; Apio, C.; Ko, Y.; Han, K.; Goo, T.; Heo, G.; Kim, T.; Chung, H.W.; Lee, D.; Lim, J.; et al. Which National Factors Are Most Influential in the Spread of COVID-19? *Int. J. Environ. Res. Public Health* **2021**, *18*, 7592. <https://doi.org/10.3390/ijerph18147592>

Academic Editor: Luenda E. Charles

Received: 27 May 2021

Accepted: 10 July 2021

Published: 16 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The novel coronavirus 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 that has been seen in respiratory viruses since the 1918 H1N1 influenza pandemic [4], with a total of 104,904,529 confirmed cases and 2,278,471 deaths worldwide, as of 4 February 2021 [5].

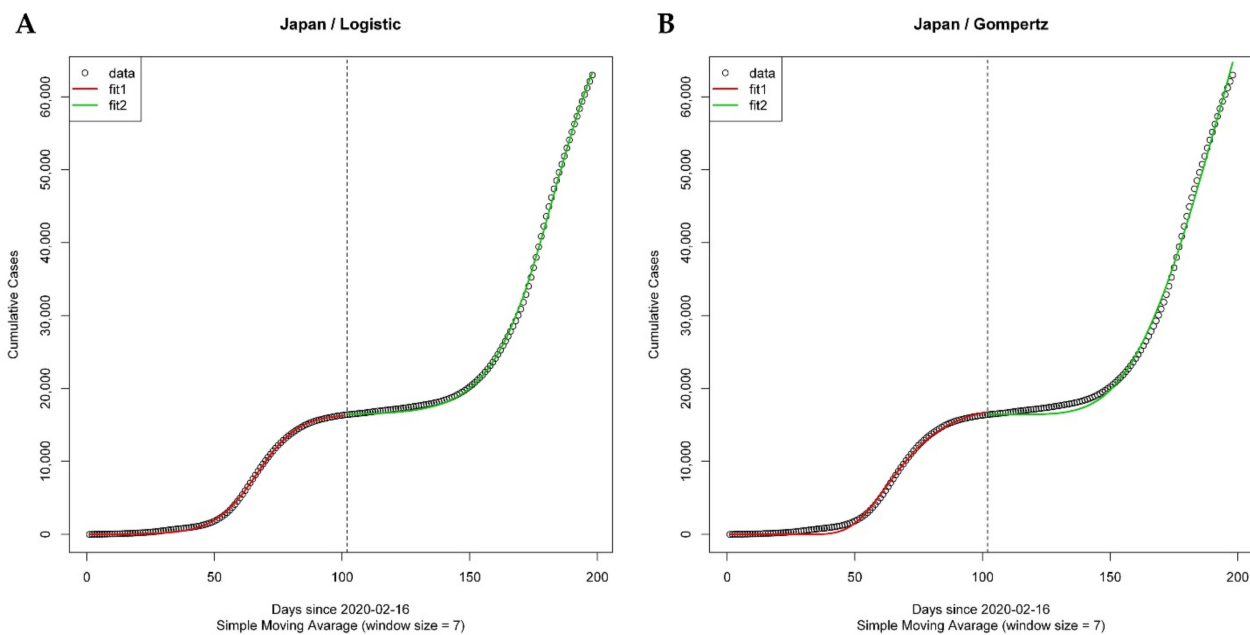
Many factors can influence the epidemiological characteristics and contribute to the increased mortality rate of COVID-19 [6–10]. Therefore, understanding the potential affecting factors involved in the outspread of COVID-19 will be of great significance in containing the spread of the epidemic and was the focus of many studies [6]. Several theories are suggesting the impact of environmental factors such as 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 being paid to temperature and humidity, indicating that temperature is inversely related to COVID-19 incidence [11]. Moreover, each 1 °C increase in temperature was found to be 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 us that pandemics and epidemics have consistently and significantly affected human lives, and that 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 such as life expectancy, average annual temperature, aging index, human development index, percentage of malnourished people in the population, extreme poverty, economic ability, health policy, population, and age distributions on the spread of COVID-19.

Here, we first applied segmented growth-curve models (logistic and Gompertz models) to the cumulative confirmed cases of 213 countries. Next, we applied the segmentation algorithm to divide the cumulative curve of COVID-19 cases into several segments of time series cases corresponding to a specific segment, which can then be modeled by the conventional growth curve models into a sigmoid curve. As the spread of COVID-19 has been prolonged, several countries experienced more than one wave (a wave implies a rising number of sick individuals, a defined peak, and then a decline [16] 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. Thus, the study period of each country was divided into several segments—time periods corresponding to a specific wave using the segmentation algorithm (Figure S1) that 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, Figure 1).

The segmented growth curve models (GCMs) summarize the spread of COVID-19 into sets of three parameters  $\alpha$ ,  $\beta$ , and  $\gamma$ , where  $\alpha$  is the maximum number of predicted cumulative confirmed cases,  $\beta$  is the time when we start to see a rise in the number of confirmed cases, and  $\gamma$  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* [17] and Korean Statistical Information Services (KOSIS) [18] such as life expectancy, average annual temperature, aging index, human development index, percentage of malnourished people in the population, and extreme poverty (Table S1), and the spread of COVID-19, using the above parameters estimated from the segmented growth curve models. The significant relationship provides evidence that these variables may influence the spread of the novel SARS-CoV-2 virus across the globe.



**Figure 1.** Cumulative confirmed cases divided into two segments using the segmentation algorithm. Japan is the typical country with two waves (red: first segment and green: second segment). **(A)** Epidemic segmented growth curve of COVID-19 fitted by the logistic model. Estimated parameters were  $(\alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2) = (16549.7588, 7.8244, 0.1231, 58877.8030, 6.7353, \text{ and } 0.0859)$ , respectively). **(B)** Epidemic segmented growth curve of COVID-19 fitted by the Gompertz model. Estimated parameters were  $(\alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2) = (17,531.5024, 89.4251, 0.0760, 98,622.3568, 15.6933, \text{ and } 0.0325)$ , respectively).

## 2. Materials and Methods

### 2.1. 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 [19–21]. ECDC is an EU agency aimed at strengthening Europe’s defenses against infectious diseases. Negative confirmed cases were corrected to 0, regarding it as abnormal data. Since data of cases on an international conveyance in Japan were included in a country list, we removed it. The data consisting of 213 countries from 1 January 2020 to 31 August 2020 were used in downstream analysis.

Data smoothing was used to remove noise from a dataset, allowing important patterns to stand out. Thereafter, daily confirmed case data were smoothed by simple moving average (1) to reduce the effect of outliers and (2) to 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, we set the window size to 7, and simple moving average (SMA) was used before model fitting as shown below

$$SMA = \frac{p_M + p_{M-1} + \dots + p_{M-(n-1)}}{n} = \frac{1}{n} \sum_{i=0}^{n-1} p_{M-i},$$

where  $p$  is the number of confirmed cases.

### 2.2. National Factors

Time-independent national factor (Table S1) datasets are publicly available datasets that are easily obtained from the *Our World in Data* website [22] and the Korean Statistical Information Services (KOSIS) [18]. The *Our World in Data* website provides data about

research and data to make progress against the world's largest problems such as poverty, disease, hunger, climate change, war, and existential risks. It mainly focuses on the large problems that continue to confront us for centuries or much longer, as well as the long-lasting forceful changes that gradually reshape our world. From this website, we obtained 15 time-independent social and economic factors assumed to be related to COVID-19 in the literature, such as population, population density, median age, being aged 65 or over, being aged 70 or over, GDP per capita, extreme poverty, cardiovascular death rate, diabetes prevalence, female smoker, male smoker, handwashing facilities, hospital beds per thousand people, life expectancy, and human development index [23–29].

The Korean Statistical Information Service (KOSIS) [18] 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 assumed to be related to the spread of COVID-19. These variables were measured for several years. Therefore, we selected the year with the minimum number of missing values between 2016 and 2019, re-scaled by division with standard errors of the variables.

### 2.3. Analysis of the Spread of COVID-19 Using GCMs

Under this analysis, the 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 and predict the probability of occurrence of a certain disease, investigate factors that control and affect growth, and extinction laws of the population [30]. The models take the following forms:

#### 2.4. Logistic Model

$$Q_t = \frac{\alpha}{1 + e^{\beta - \gamma(t - t_0)}} \quad (1)$$

where  $Q_t$  is the cumulative confirmed cases,  $\alpha$  is the maximum number of predicted cumulative confirmed cases,  $\beta$  is the time when we start to see a rise in the number of confirmed cases,  $\gamma$  is the increase rate of number of confirmed cases,  $t$  is the number of days since the first case occurrence, and  $t_0$  is the time when the first case occurred.

#### 2.5. Gompertz Model

$$Q_t = \alpha e^{-\beta e^{-\gamma(t - t_0)}} \quad (2)$$

where  $Q_t$  is the cumulative confirmed cases,  $\alpha$  is the maximum number of predicted cumulative confirmed cases,  $\beta$  is the time when we start to see a rise in the number of confirmed cases,  $\gamma$  is the increase rate of number of confirmed cases,  $t$  is the number of days since the first case, and  $t_0$  is the time when the first case occurred.

#### 2.6. Segmentation Algorithm

As the COVID-19 situation continues, fitting a growth curve model on daily confirmed cases over a long period of time has become impossible as it no longer takes on an s-curve (i.e. sigmoid function). To fit the above growth curve models, there is a need to divide the study period of countries experiencing more than one wave [31] (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). Thus, we applied the segmentation algorithm, which can systematically divide study periods into several segments (or waves) for each country (Figure 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 that splits the consecutive two segments in a time series dataset. To better see trends, we smoothed out the irregular roughness of the graph of daily confirmed cases. However, 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) [32–34] with Gaussian kernel to smoothen the daily new confirmed cases as demonstrated in Figure 2 using South Korea’s daily confirmed cases as an example. For the convenience of notation, let  $Y_t$  be the  $t$ -th daily new confirmed cases from data, and  $\hat{f}(t)$  be the estimated  $t$ -th daily new confirmed cases using above NWE since 1 January 2020.

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

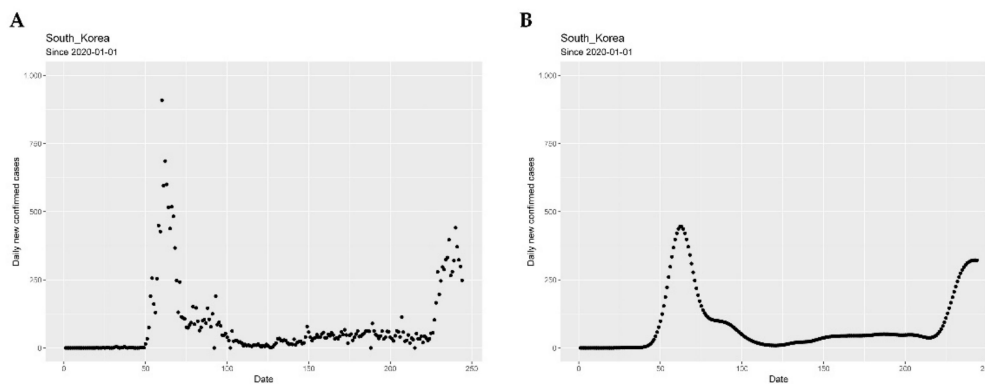
$$\Delta\hat{f}(t) = 0, \Delta^2\hat{f}(t) < 0 \quad (3)$$

where  $\Delta\hat{f}(t) = \hat{f}(t+1) - \hat{f}(t)$  and  $\Delta^2\hat{f}(t) = \Delta\hat{f}(t+1) - \Delta\hat{f}(t)$ .

For discontinuity and small variances of  $\hat{f}(t)$ , we used following condition:

$$\Delta\hat{f}(t)\Delta\hat{f}(t+1) \leq 0, \Delta^2\hat{f}(t) < -c \cdot \operatorname{argmax}_{t \in T} |\Delta^2\hat{f}(t)| \quad (4)$$

where  $c \in (0, 1)$  is sensitivity level and  $T$  is the set of time indices from 1 January 2020 to 31 August 2020. In addition, 3 additional conditions ((a) exclusion of small peaks, (b) resolution criteria, and (c) exclusion of peaks that are vibrations on increasing trend) were used in peak detection to enhance robustness. After all the peaks were found, breakpoints (Algorithm 2; Figure S1) were selected either as timestamps that 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). Figure S2 visualizes the segmentation process. Blue line represents the peak, and dotted sky-blue line represents breakpoint. In the first plot, the black solid line represents  $\hat{f}(t)$  and the black dotted line represents  $Y_t$ . The second plot represents cumulative confirmed cases of  $Y_t$  (black dotted line),  $\hat{f}(t)$  (black solid line). The third and fourth plots are graphs of  $\Delta\hat{f}(t)$ ,  $\Delta^2\hat{f}(t)$ . In the fourth plot, the green dotted line represents sensitivity level. If  $\Delta^2\hat{f}(t)$  is above the upper green dotted line,  $\hat{f}(t)$  is concave. On the other hand, if  $\Delta^2\hat{f}(t)$  is below the lower green dotted line,  $\hat{f}(t)$  is convex. Within the third and fourth plots, Equation (4) can be validated. The segmentation algorithm was successfully applied to 134 countries from the 213 countries in the ECDC dataset, which met  $\max_{t \in T} (\hat{f}(t)) \geq 50$ . If  $\max_{t \in T} (\hat{f}(t))$  is too small, segmentation algorithm would be difficult to apply due to small variances in  $\Delta^2\hat{f}(t)$ .



**Figure 2.** Daily new confirmed cases before (A) and after smoothing using Nadaraya–Watson kernel regression (B).

### 2.7. 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 did not preserve continuity at breakpoints, but this did not matter since the objective of our analysis was to condense daily new confirmed cases into several parameters ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) of the growth curves and not to accurately predict daily new confirmed cases. Equations (5) and (6) below are the segmented logistic and Gompertz models, respectively.

$$Q(t) = \sum_{i=1}^n \left( \frac{\alpha_i}{1 + e^{\beta_i - \gamma_i t}} + q_i \right) I_{Seg_i}(t) \tag{5}$$

$$Q(t) = \sum_{i=1}^n (\alpha_i e^{-\beta_i} e^{-\gamma_i t} + q_i) I_{Seg_i}(t) \tag{6}$$

where,  $q_i (i \geq 2)$  is the number of cumulative cases at  $(i - 1)^{th}$  breakpoint, and  $I_{Seg_i}(t)$  is indicator function where  $Seg_i$  is the set of indices of  $i^{th}$  segment and  $q_1 = 0$ .

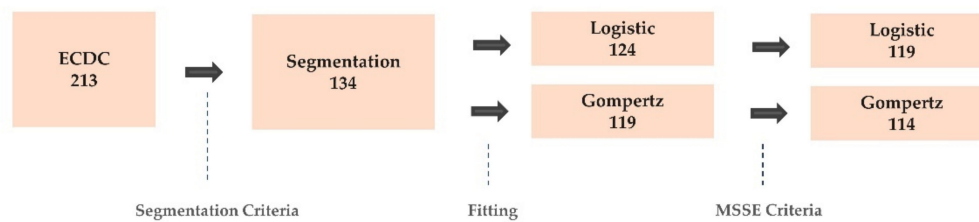
In this analysis, we only considered first and second segments, since most countries have 1 or 2 segments (1 segment: 62, 2 segments: 65, 3 segments: 7). The number of countries with three segments was very small, making the comparison analysis insignificant to use in the regression analysis. For countries with more than 2 segments, the analysis period was, therefore, cut off at the second breakpoint. For countries with 2 segments, segmented growth curve model then would produce 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 models. To filter out poorly fitted countries, we excluded countries whose MSSE (mean squared scaled error) was higher than 0.4, as defined below:

$$MSSE = \frac{1}{N} \sum_{t=1}^N \left( \frac{Y_t - \hat{Y}_t}{\bar{Y}} \right)^2 \tag{7}$$

where  $Y_t$  is the daily new confirmed cases,  $\hat{Y}_t$  is the predicted value for  $Y_t$  by segmented logistic and Gompertz models, and  $\bar{Y}$  is the mean of  $Y_t$  for  $t = 1, \dots, N$ .

MSSE is a more suitable measure compared with MSE (mean squared error) or MAPE (mean absolute percentage error) because the MSE does not consider scales of population among each country, while MAPE overestimates its error when the number of daily new confirmed cases, and  $Y_t$  is small. Among the 134 countries, 124 countries were fitted for the segmented logistic model and 119 countries for the 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 and segmented Gompertz models. Therefore, a total of 119 countries were used in the segmented logistic model, and 114 countries for the segmented Gompertz model (Figure 3).



**Figure 3.** Differences in the number of countries across segmentation, growth curve models, and MSSE criteria. For the segmented logistic model, 124 countries were fitted, and for the segmented Gompertz model, 119 countries were fitted. To check (validate) the goodness of fit of the above 2 models, we employed MSSE (mean squared scaled error) criteria. For each of the two models, 5 countries showed high MSSE, and therefore those 5 countries were excluded in the subsequent analysis (Aruba, Equatorial Guinea, Kyrgyzstan, Rwanda, and Thailand for the segmented logistic model, and China, Equatorial Guinea, Kyrgyzstan, Rwanda, and Zambia for the segmented Gompertz model).

In addition, correlation analysis for segmented logistic and Gompertz models with the log-scaled of parameters was performed to determine the similarity between parameters of the two models (see Figures S3–S6).

### 2.8. Regression Model

The above segmented growth curve models summarize the spread of the pandemic into three parameters ( $\alpha_1, \beta_1, \gamma_1$ ) for countries with one segment, and into six parameters ( $\alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2$ ) for countries with two segments. Each of the parameters from the two segmented GCMs was regressed against the national factors shown in Figure 1 as follows:

$$y_j^{ik} = \theta_0^{ik} + \theta_1^{ik} x_j + \varepsilon \quad (8)$$

where  $y_j^{ik}$  is one of the segmented GCM parameters ( $\alpha, \beta, \gamma$ ) for model  $i = 1$  (logistic), 2 (Gompertz), segment  $k = 1, 2$  and country  $j$ .  $\theta_0$  and  $\theta_1$  are regression coefficients, and  $x_j$  is the national factor of country  $j$ . F-statistic was performed to test the significance of  $\theta_1$  for each national factor to find out which variables had a significant relationship with  $y$ , a measure of the spread dynamics of COVID-19 for a country.

## 3. Results

### 3.1. Growth Curve Models Predicted 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, i.e., three parameters ( $\alpha_1, \beta_1, \gamma_1$ ) for countries with one wave (and therefore, one segment), and six parameters ( $\alpha_1, \beta_1, \gamma_1; \alpha_2, \beta_2, \gamma_2$ ) for countries with two waves (two segments). Here, the differences between parameters estimated from logistic and Gompertz models among the countries are discussed (Tables S2 and S3). Figure 4 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 ( $\alpha$ ), while the  $y$ -axis represents the parameter related to the rate of spread of COVID-19 ( $\gamma$ ).

Parameter estimation showed that the Philippines, India, and Brazil had the highest numbers of maximum predicted cumulative confirmed cases in the first segment of the pandemic using the logistic model (Figure 4A), while India and Zambia were shown to have the highest numbers using the Gompertz model (Figure 4B). In the second segment of the pandemic, the USA had the highest number of maximum cumulative confirmed cases using both GCMs (Figure 4C,D). Therefore, by 31 August 2020, the USA was the country with the greatest number of cumulative confirmed cases in the world. All the other

remaining countries did not have notably large differences in their numbers of maximum predicted cumulative confirmed cases in both GCMs.

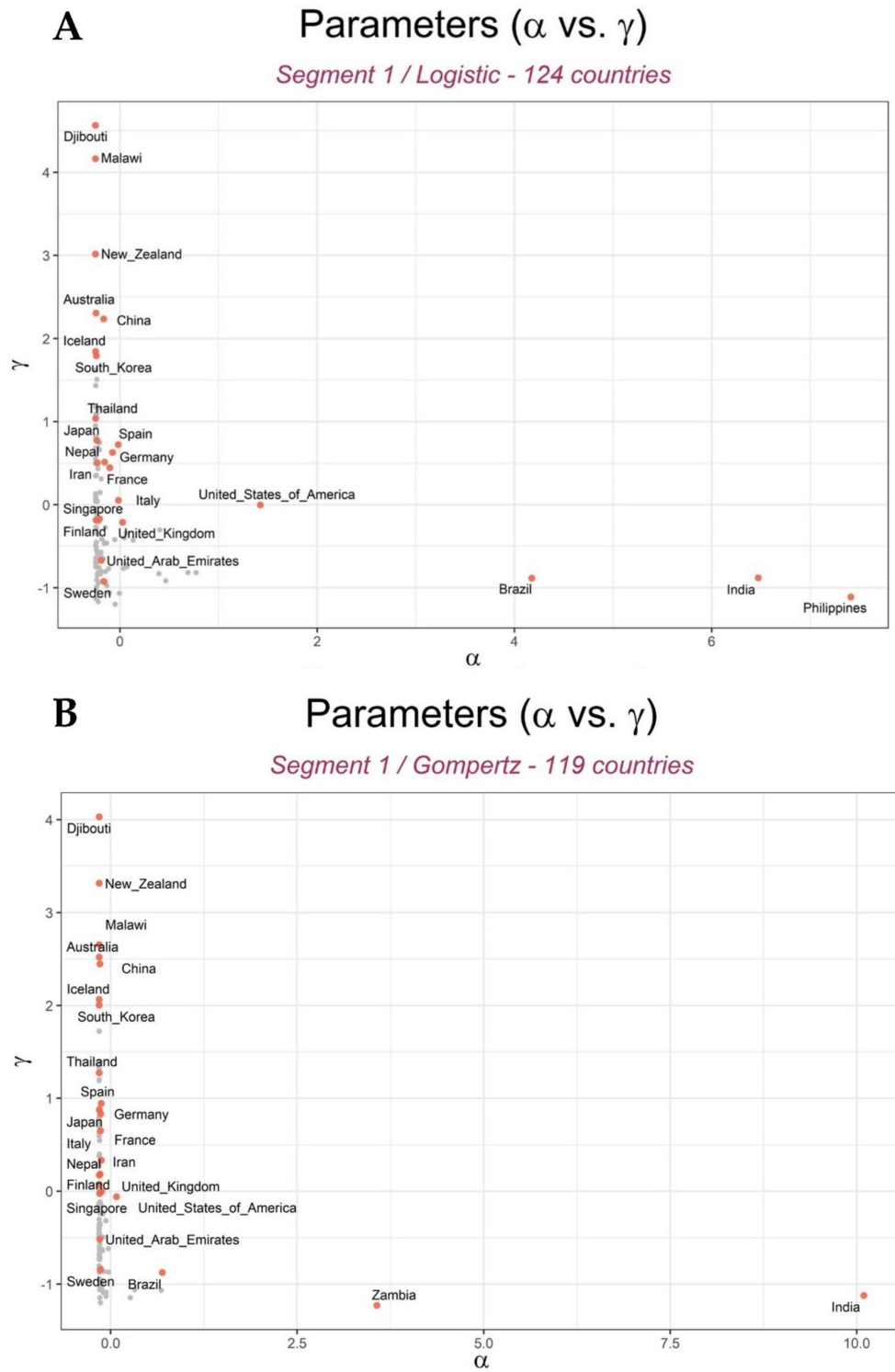
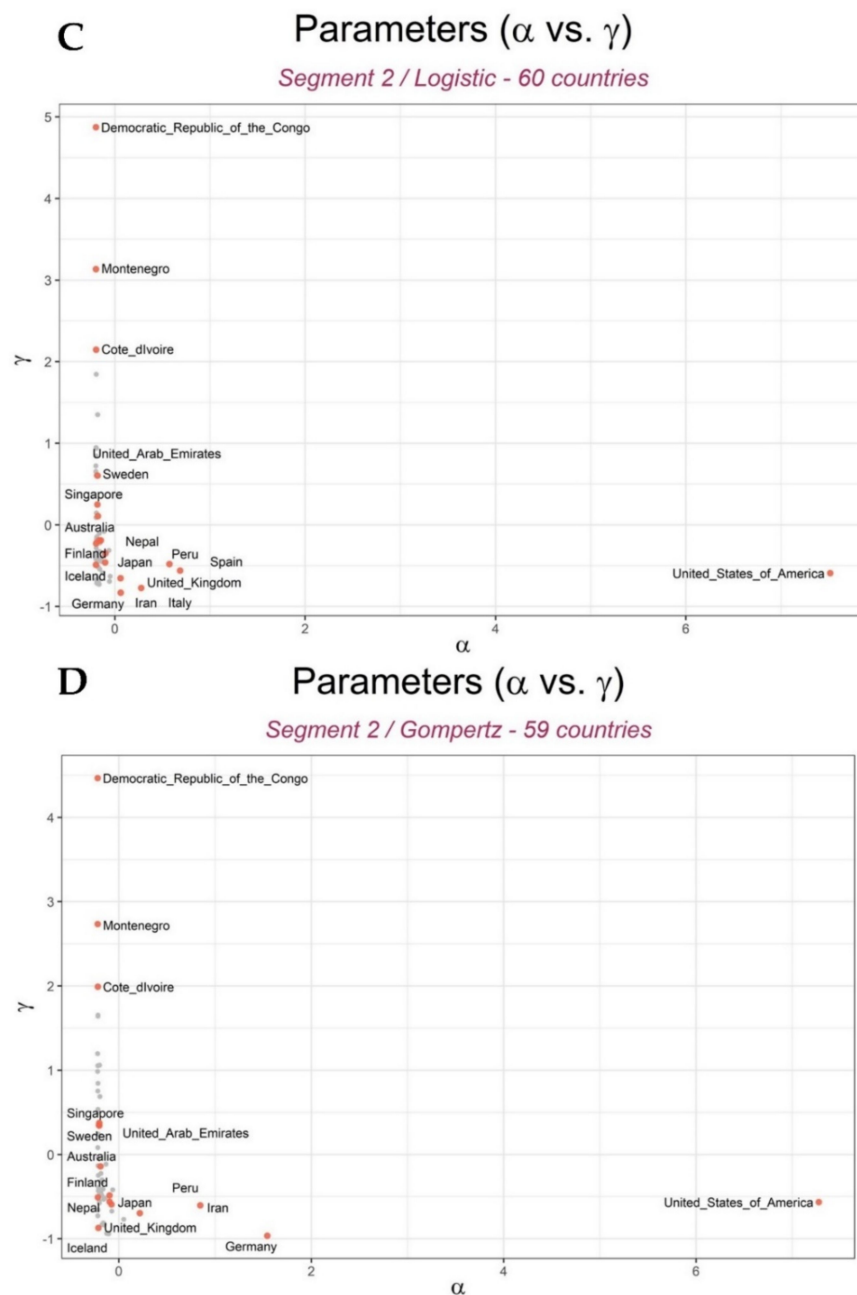


Figure 4. Cont.





**Figure 4.** Differences in values of maximum predicted cumulative cases ( $\alpha$ ) and rate of spread of COVID-19 observed ( $\gamma$ ) among countries. (A,C) The variation of  $\alpha$  vs.  $\gamma$  using the logistic model for the first and second segments, respectively. (B,D) The variation of  $\alpha$  vs.  $\gamma$  using the Gompertz model for the first and second segments, respectively.

However, we observed 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 ( $\gamma$ ) of COVID-19 among their populations (Figure 4A,B), according to both models. In the second segment, the Democratic Republic of Congo, Montenegro, and Cote d’Ivoire (Ivory Coast) had the highest rate of spread, while Iceland, Finland, the UK, Nepal, Australia, and Japan had the lowest rate of spread of COVID-19 in their populations, according to both models (Figure 4C,D).

Moreover, we observed that countries with the greatest numbers of predicted maximum cumulative confirmed cases had the smallest rate of spread and vice versa, in both

models and segments (see Figure S7 of  $\log_{10}\alpha$  vs.  $\gamma$ ). The correlation analysis (Pearson's correlation) to determine the relationship between the parameters across the two models and segments (Figures S3–S6) confirmed that the parameters had similar interpretation across models and segments, but a noticeable negative correlation ( $-0.5$  and  $-0.55$  for logistic,  $-0.66$  and  $0.7$  for Gompertz) between  $\alpha$  and  $\gamma$  parameters (Figures S5 and 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, the population scale among countries was not considered. Thus, the time when we started to see a rise in the number of confirmed cases ( $\beta$ ) did not produce consistent results between segments and models as the other parameters did, although its interpretation was the same between the models. Therefore, its results and any analysis concerning it were not a focus in our study, and its results were relegated to the Supplementary Materials for those interested. In addition,  $\beta$  showed 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 showed a strong positive correlation between the models ( $0.88$  and  $0.95$ ).

### 3.2. The Relationship between National Factors 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,  $\alpha$ , and the rate of spread of the pandemic,  $\gamma$ , estimated from the segmented GCMs. The 31 national factors included developmental (called World Development Indicators by World Bank [35]) and non-developmental variables related to population, age distribution, health, and environment (Table S4).

The objective of our analysis was to determine whether these factors influence the spread of COVID-19. From the segments 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 were statistically significant between two models and two segments. We used a 5% significance level in this analysis. Statistically significant results provided evidence for the possibility of these factors influencing the spread of COVID-19.

For the number of maximum predicted cumulative confirmed cases ( $\alpha$ ), several national factors turned out to be significant, such as population, annual precipitation, pharmaceutical sales, and imports to GDP ratio (Figure 5A,C). However, 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 ( $\gamma$ ) was significantly related to 19 national factors. For example, age-related variables such as aging index, share of population aged 65 and older, share of population aged 70 and older, median age and life expectancy, health-related variables such as 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 such as international travelers from a country and number of foreign visitors to a country, and environmental factors such as average annual temperature (Figure 5B,D).

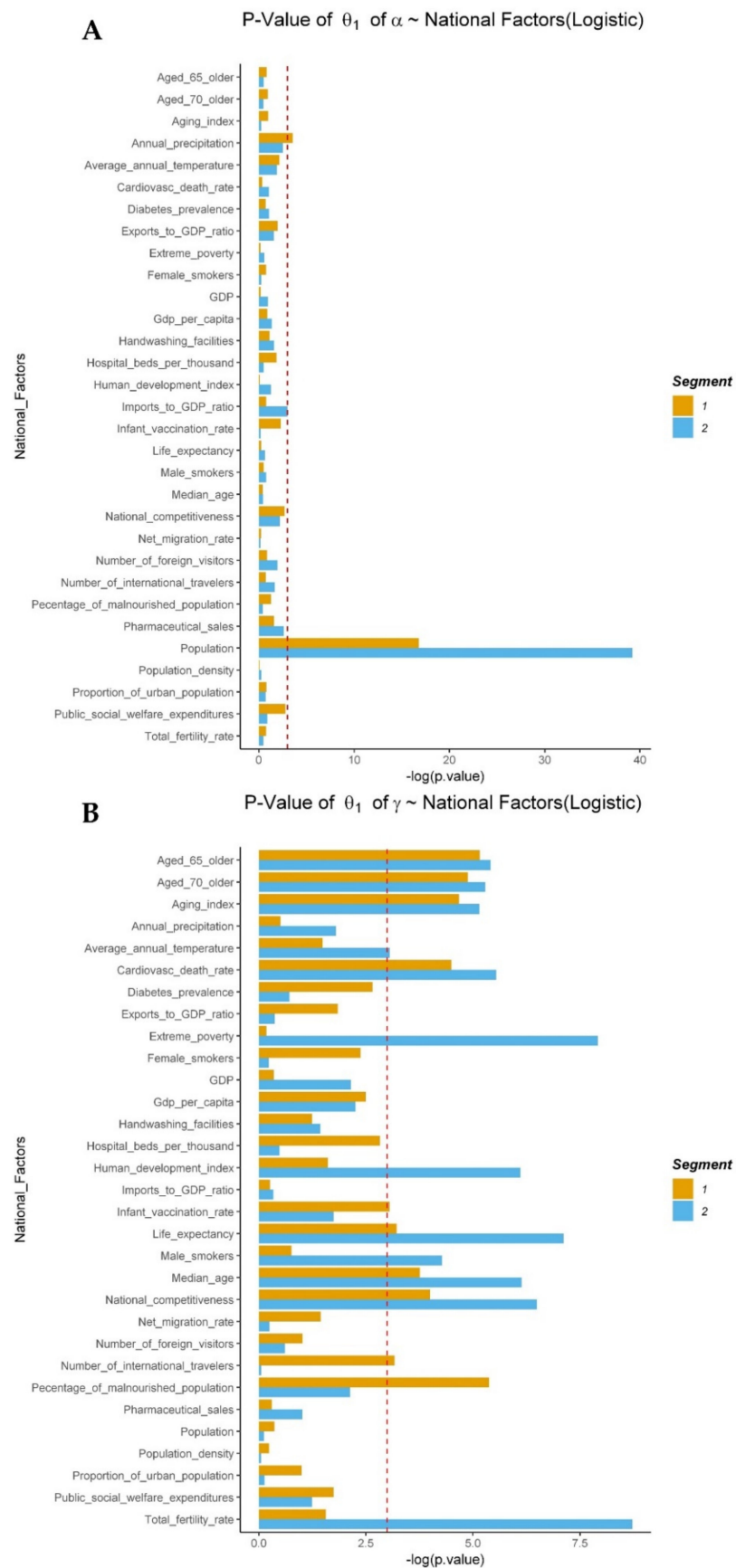
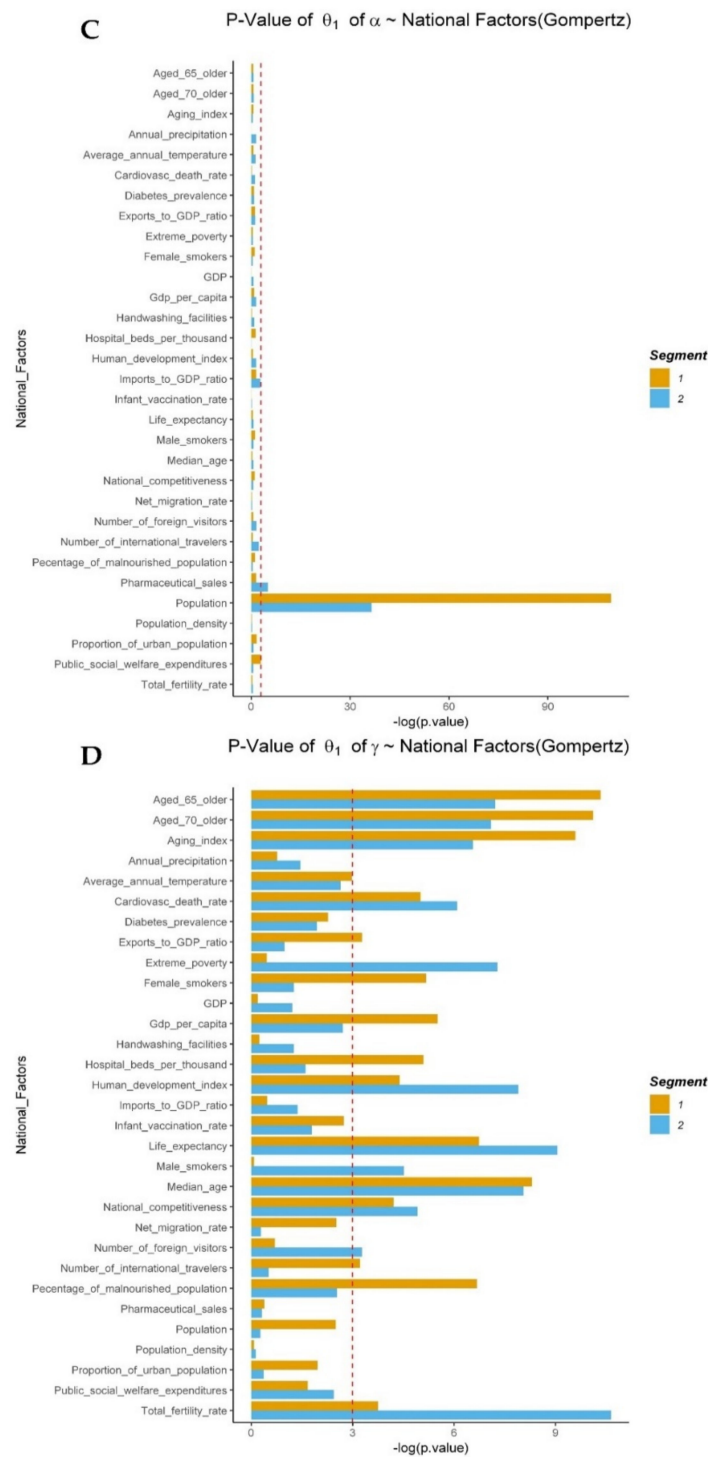


Figure 5. Cont.



**Figure 5.** *p*-values of coefficients of national factors with  $\alpha$  and  $\gamma$ . Population, annual precipitation, pharmaceutical sales, and imports to GDP ratio were statistically significant with  $\alpha$ , the number of maximum predicted confirmed cases (A,C). Age-related factors, population, percentage of malnourished population, life expectancy, temperature, etc. were significantly related with  $\gamma$ , rate of spread of COVID-19 (B,D) (see Figure S8 for the relationship of national factors with  $\beta$ ).

In addition, a relationship between the size of coefficient values (of the relationship between national factor and GCM parameter) and significance of national factors was observed, whereby significant variables generally had larger coefficient values than non-significant variables (Figure 6, Figures S9 and S10). Our results provide evidence of the influence of these significant national factors such as population, aging index, median

age, cardiovascular death rate, extreme poverty, annual precipitation, number of foreign visitors and international travelers, on the spread of COVID-19 across the globe. Moreover, we rarely observed a change in signs of the coefficients of the significant variables between models.

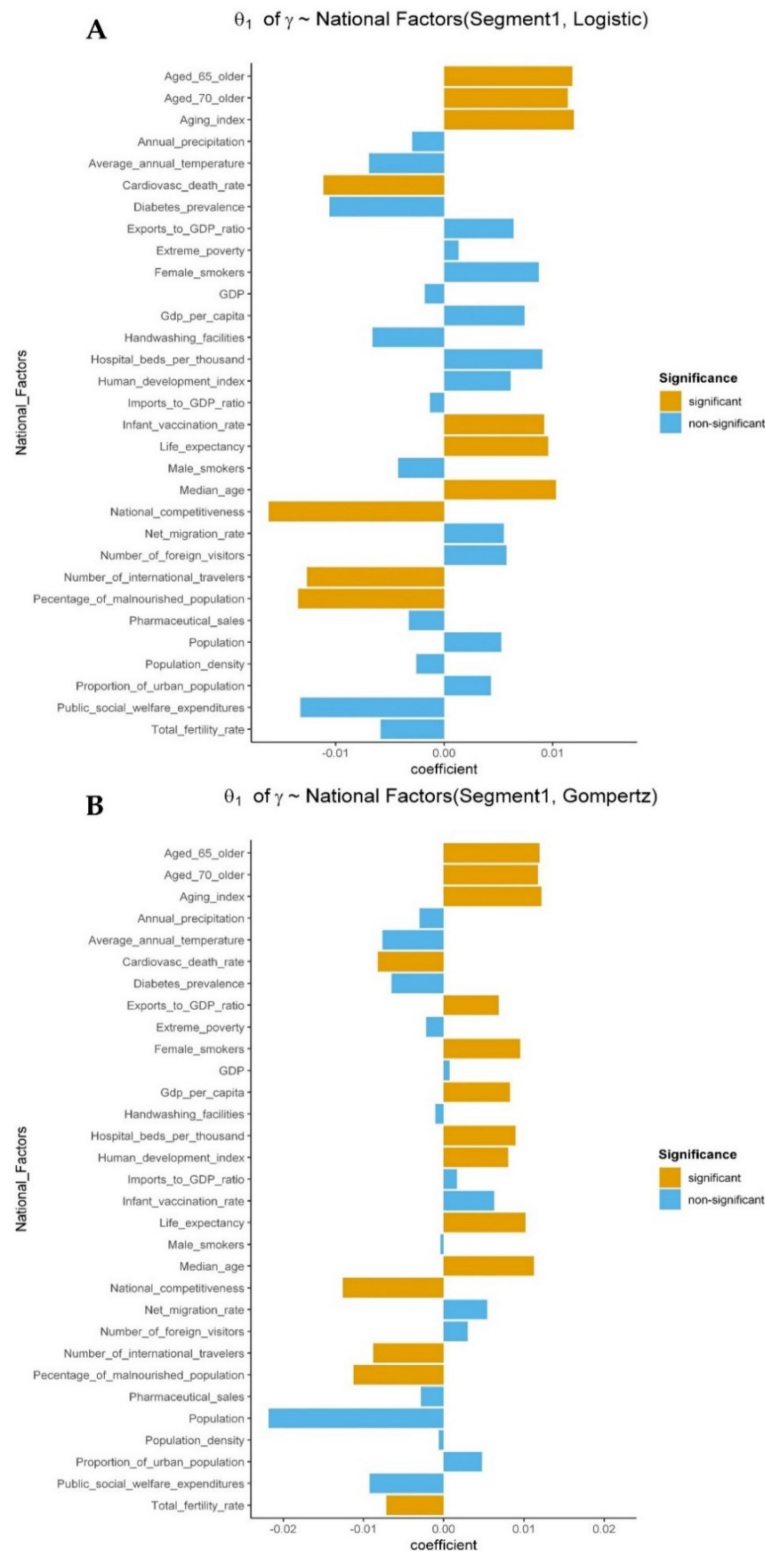
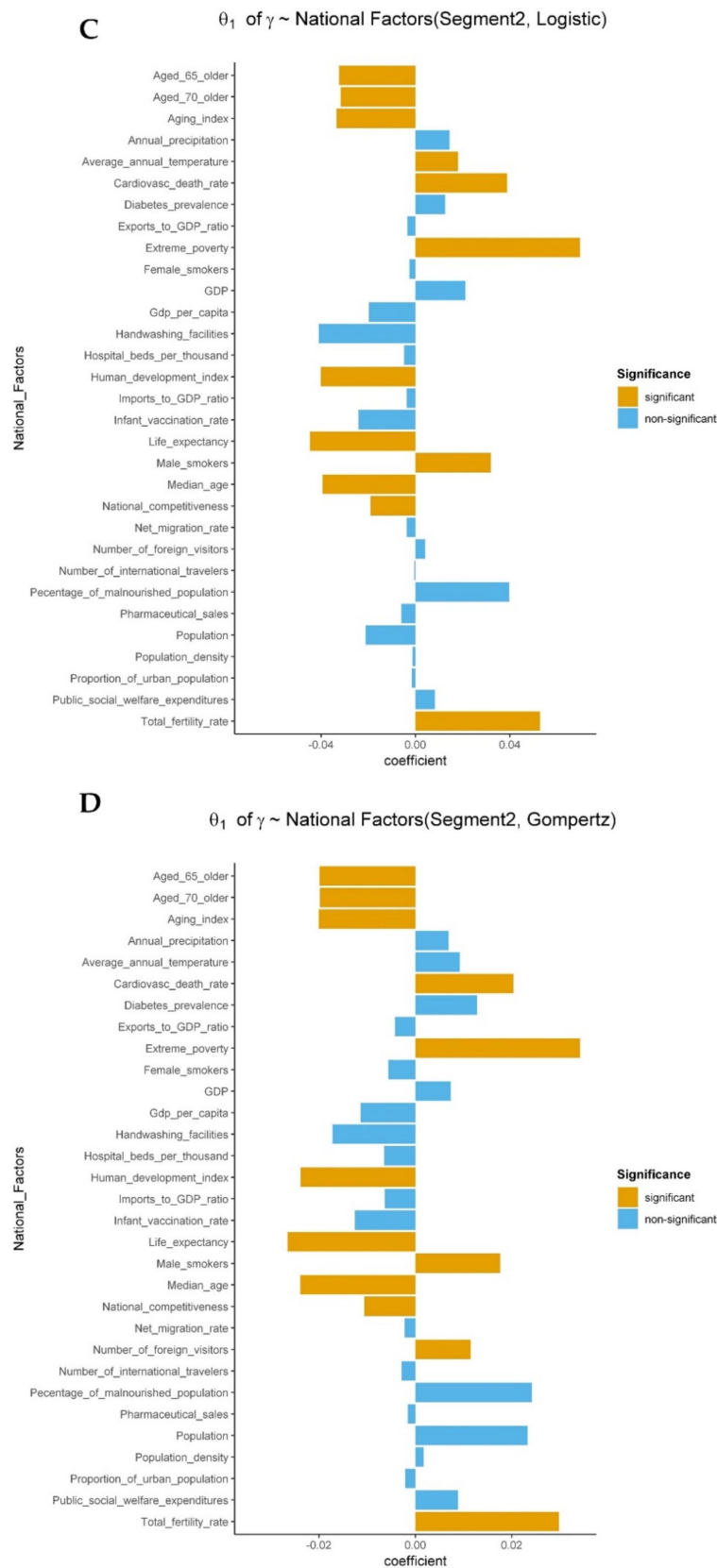
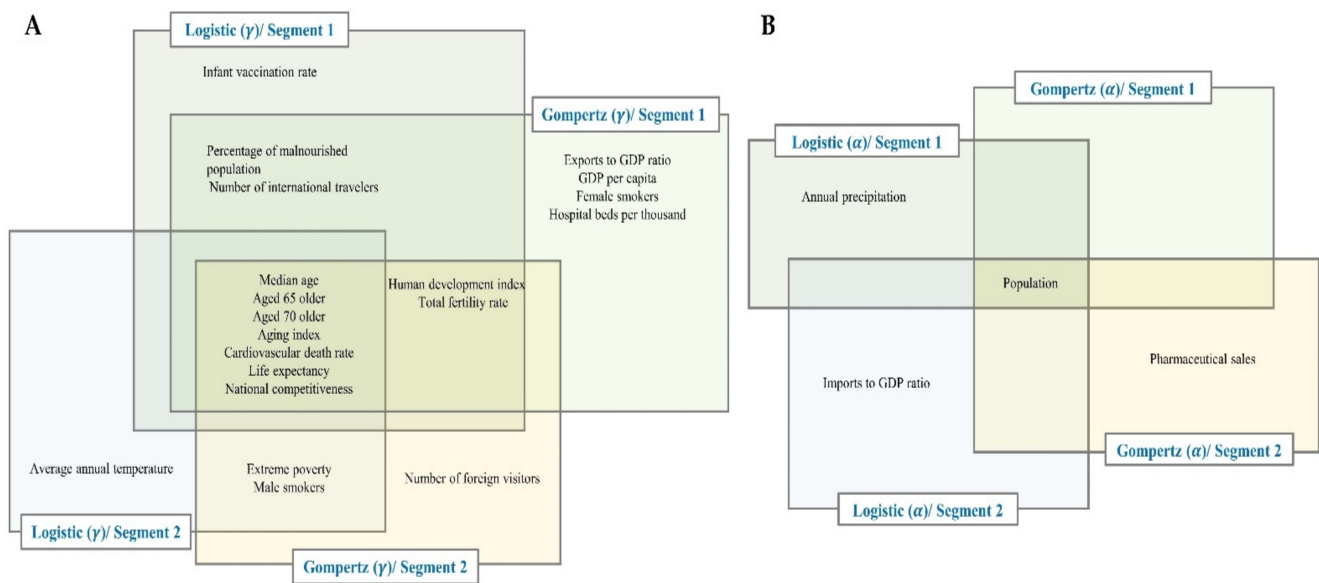


Figure 6. Cont.



**Figure 6.** Coefficients of the relationship between national factors and  $\gamma$ .  $\theta_1$  is the coefficient of the relationship between a national factor and a GCM parameter. Significant national factors (orange) had large coefficients compared with non-significant factors (blue) in both Gompertz (A,C) and logistic models (B,D).

The number of maximum predicted cumulative confirmed cases is significantly influenced by only population in both the two GCMs and segments of each model (Figure 7B). The countries with the highest value of maximum predicted cumulative confirmed cases (India, the USA, Brazil, the Philippines, and Zambia) had the highest population sizes in the world as of 31 August 2020 [36]. In addition, the USA, India, and Brazil have been, in that order, the countries hardest hit by the COVID-19 pandemic worldwide [37,38], 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.



**Figure 7.** Significant national factors with number of maximum predicted cumulative confirmed cases ( $\alpha$ ) and rate of spread of COVID-19 ( $\gamma$ ). Median age, being aged 65 or older, being aged 70 or older, aging index, cardiovascular death rate, life expectancy, and national competitiveness were the only national factors that were found to be significant across the two models and segments (A). Population was significant across the two models and segments (B) (see Figure S11 for significant national factors with  $\beta$ ).

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 (Figure 7A). Age-related variables. i.e., aging index, median age, percentage of the population aged 65 or 70 and older, and life expectancy are significant in both models and segments. Aging is linked mainly with deteriorating immune system [39] and other common conditions such as hearing loss, cataracts and refractive errors, back and neck pain and osteoarthritis, chronic obstructive pulmonary disease, diabetes, depression, and dementia, wherein several of these conditions can be experienced at the same time [40,41]. The risk for severe illness with COVID-19 increases with age, with older adults being at a greater risk of requiring hospitalization and dying of COVID-19 when diagnosed in comparison with younger people. This is due to already 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 [22,42–44]. This includes other variables such as cardiovascular death rate [45] and the percentage of female and male smokers in the population.

One in five (20%) adults in the world smoke tobacco [46], being one of the world’s largest health problems. Active smoking and a history of smoking (cigarettes, waterpipes, bidis, cigars, heated tobacco products) may lead an individual to being vulnerable to contracting COVID-19, having been linked to increased severity of COVID-19 illness due

to the health complications, wrecking mainly the immune system, especially on the lungs (epithelial cells), which is a primary site of target of SARS-CoV-2 [47–49]. Moreover, 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 [48]. It is reported that Montenegro has 46% smoking prevalence, being a country with the second highest rate of spread of COVID-19 in the second segment of analysis [46], while OECD member countries were found to have a prevalence of 23.50% as of 2016. African countries have some of the lowest levels of smoking in the world [46].

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 government's preparedness to deal with disasters (new pandemics included) and interferes with health system response such as drugs, protective gear, 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 [50–52]. 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 [53,54].

Number of international travelers and foreign visitors increases the chance of spreading and catching the SARS-CoV-2 virus among the population [55], mainly due to importation and exportation of cases, leading to many domestic travel restrictions and flight suspensions between countries [39,56]. Accelerated by human migration, exported COVID-19 cases have been reported in various regions of the world, including Europe, Asia, North America, and Oceania [57]. 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, showing the importance of national competitiveness in response to COVID-19 [58,59]. 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 [60,61].

#### 4. 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 modeled the spread of COVID-19 using segmented logistic and Gompertz models, and then we investigated the influence of national factors 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, with one of the most important differences being 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 [36]. By building two models and analyzing the results (Figure 7), 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 was significantly influenced by only one factor, while the rate of spread of COVID-19 was influenced by seven factors, in both the two GCMs and segments of each model (Figure 5). This made the rate of spread of the pandemic the most influenced aspect of the spread of COVID-19 among countries among the two parameters. Moreover, we found out that the number of maximum predicted cumulative confirmed cases ( $\alpha$ ) did not vary much across countries (although we observed a few outliers, e.g., the USA, India, Brazil, the Philippines, and Zambia), while the rate of spread of COVID-19 ( $\gamma$ ) varied greatly across countries. We observed that  $\alpha$  was only mainly influenced by population (Figure 7B), while  $\gamma$  was significantly influenced by many variables (Figure 7A). This may explain the differences observed in the rate of spread of COVID-19 among countries in comparison with 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 saw the influence of population size on the spread of COVID-19. Among the hardest hit countries by the COVID-19 pandemic in the world, the USA, India, and Brazil are also among the countries with the largest populations in the world. Some countries with the highest number of maximum predicted cumulative confirmed cases (Zambia, India, Brazil, and the 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 [62,63] and in a malnourished state [64,65], as well as having the youngest populations (especially African countries) in the world [66]. Moreover, in the first segment, Iceland, South Korea, China, New Zealand, and Australia, which had a high rate of spread of COVID-19, are characterized as having older populations, longer life expectancy, higher GDP per capita, higher cardiovascular death rate, large percentage of population that smoke daily [67], better health systems, and little to no malnutrition [66,68–71]. Clearly, we observed the influence of these variables on the spread of COVID-19 [35]. However, most of these countries, in addition to Japan, the UK, Italy, Germany, and the United Arab Emirates (despite having the characteristics listed above), also had the lowest rate of spread of COVID-19 in the second segment of the pandemic (Figure 4). This could have been due to the influence of government-implemented policies such as “lockdowns” 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 although we modeled the spread of COVID-19 for 134 countries, the GCMs still produced some missing parameter values (10 countries in the logistic model, 14 countries in the Gompertz model) between the segments and models for some countries mainly due to failure of convergence (Tables S2 and S3), which may have affected comparison and therefore the interpretation of the results. Moreover, we could 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 the 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. While the relationship between several national factors and COVID-19 via regression has been studied at the univariate level, multivariate analysis/regression has not been performed to adjust for the influence of one factor on the association between another factor and COVID-19. We hope to perform this type of analysis too by including some sort of variable/feature selection, and then study the relationship between the selected feature set of national factors and COVID-19 at multivariate level.

Moreover, COVID-19, which is a contact-transmissible infectious disease and is said to spread through the population via direct contact between individuals [2,72,73] 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 [74,75]. These government-implemented policies have already been shown to mitigate and suppress the pandemic [76,77]. 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 [78]. However, it was 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 such as the containment policies. Therefore, we could not control for this bias in our analysis, as some may argue on this topic. In the future, we hope to consider the impact of government-implemented policies on the spread of COVID-19 in our analysis using other models.

In addition, the role of host genetics interaction and COVID-19 progression has gained a large amount of interest as one of the factors being proposed to influence the spread of COVID-19 [79]. For example, the difference in terms of incidence of COVID-19 observed between the northern and southern regions of Italy was attributed to genetics as being one of the factors causing this inhomogeneous distribution of cases [24]. However, we analyzed a country's COVID-19 pandemic situation instead of specific COVID-19-confirmed individuals. This made it difficult to include genetic information in the current models. However, provided that ethnic or ancestral difference data of each of the countries analyzed is available, we can indirectly analyze effects of genetics using the ethnical differences of a given country as another covariate in the GCMs. We hope to model the role of genetics in relation to the spread of COVID-19 in a future study.

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

## 5. Conclusions

Much is still unknown about the clinical and epidemiological characteristics of COVID-19, such as individual risk factors for contracting the virus and infections from asymptomatic 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 policymakers and governments in making informed scientific decisions while considering a country's economy, population dynamics, climate, and health system, which would likely have the most impact in future prevention works against similar infectious diseases.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/ijerph18147592/s1>, Figure S1. Pseudo-codes for the segmentation algorithm. Figure S2. Segmentation algorithm applied to South Korea's COVID-19 daily new confirmed cases. Figure S3. Comparison of logistic model (log-scaled) parameters  $\alpha$ ,  $\beta$ , and  $\gamma$  in the first and second segments. Figure S4. Comparison of Gompertz model (log-scaled) parameters  $\alpha$ ,  $\beta$ , and  $\gamma$  in the first and second segments. Figure S5. Comparison of logistic and Gompertz models (log-scaled) parameters  $\alpha$ ,  $\beta$ , and  $\gamma$  in the first segment. Figure S6. Comparison of logistic and Gompertz model (log-scaled) parameters  $\alpha$ ,  $\beta$ , and  $\gamma$  in the second segment. Figure S7. Variation of log<sub>10</sub> maximum predicted cumulative cases ( $\alpha$ ) and rate of spread of COVID-19 among countries ( $\gamma$ ). Figure S8.  $p$ -values of the relationship between national factors and  $\beta$ . Figure S9. Coefficients of the relationship between national factors and  $\alpha$ . Figure S10. Coefficients of the relationship between national factors and  $\beta$ . Figure S11. Significant national factors with  $\beta$ . Table S1. List of national factors. Table S2. Parameter values estimated from segmented logistic model. Table S3. Parameter values estimated from the segmented Gompertz model. Table S4. Results from the regression analysis of parameters of growth curve models and national factors. Supplementary text: Correlation analysis between the log-scaled parameters of the growth curve models.

**Author Contributions:** Conceptualization: T.P.; methodology: T.P., H.K., Y.K.; investigation: H.K., Y.K., K.H., H.W.C., T.G., C.A., D.L., J.L., T.K.; visualization: H.K., K.H., Y.K., C.A.; supervision: T.P.; data curation: K.H., G.H., T.G., H.K., Y.K., D.L.; writing—original draft: C.A., H.K., Y.K.; writing—

review and editing: T.P., C.A., G.H., J.L.; project administration: T.P.; funding acquisition: T.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Bio and Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (No. 2021M3E5E3081425).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Publicly available datasets were analyzed in this study. These datasets can be found at the data links provided in the references. All data that were required 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.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Priyadarsini, S.L.; Suresh, M. Factors influencing the epidemiological characteristics of pandemic COVID 19: A TISM approach. *Int. J. Healthc. Manag.* **2020**, *13*, 89–98. [CrossRef]
- Li, Q.; Guan, X.; Wu, P.; Wang, X.; Zhou, L.; Tong, Y.; Ren, R.; Leung, K.S.M.; Lau, E.H.Y.; Wong, J.Y.; et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N. Engl. J. Med.* **2020**, *382*, 1199–1207. [CrossRef] [PubMed]
- WHO Director-General’s Opening Remarks. Available online: <https://www.who.int/dg/speeches/detail/whComingsoono-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19--11-march-2020> (accessed on 8 February 2021).
- Worldometer. Available online: <https://www.worldometers.info/coronavirus/> (accessed on 8 February 2021).
- Bootsma, M.C.J.; Ferguson, N.M. The effect of public health measures on the 1918 influenza pandemic in U.S. cities. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 7588–7593. [CrossRef] [PubMed]
- Xu, H.; Yan, C.; Fu, Q.; Xiao, K.; Yu, Y.; Han, D.; Wang, W.; Cheng, J. Possible environmental effects on the spread of COVID-19 in China. *Sci. Total Environ.* **2020**, *731*, 139211. [CrossRef] [PubMed]
- Nakada, L.Y.K.; Urban, R.C. COVID-19 pandemic: Environmental and social factors influencing the spread of SARS-CoV-2 in São Paulo, Brazil. *Environ. Sci. Pollut. Res.* **2020**, 1–7. [CrossRef] [PubMed]
- Ahmadi, M.; Sharifi, A.; Dorosti, S.; Ghoushchi, S.J.; Ghanbari, N. Investigation of effective climatology parameters on COVID-19 outbreak in Iran. *Sci. Total Environ.* **2020**, *729*, 138705. [CrossRef]
- Hamidi, S.; Sabouri, S.; Ewing, R. Does Density Aggravate the COVID-19 Pandemic? *J. Am. Plan. Assoc.* **2020**, *86*, 495–509. [CrossRef]
- Tobías, A. Evaluation of the lockdowns for the SARS-CoV-2 epidemic in Italy and Spain after one month follow up. *Sci. Total Environ.* **2020**, *725*, 138539. [CrossRef]
- Harmooshi, N.N.; Shirbandi, K.; Rahim, F. Environmental concern regarding the effect of humidity and temperature on 2019-nCoV survival: Fact or fiction. *Environ. Sci. Pollut. Res.* **2020**, *27*, 36027–36036. [CrossRef]
- Pramanik, M.; Udmale, P.; Bisht, P.; Chowdhury, K.; Szabo, S.; Pal, I. Climatic factors influence the spread of COVID-19 in Russia. *Int. J. Environ. Health Res.* **2020**, 1–15. [CrossRef]
- Prata, D.N.; Rodrigues, W.; Bermejo, P.H.D.S. Temperature significantly changes COVID-19 transmission in (sub)tropical cities of Brazil. *Sci. Total Environ.* **2020**, *729*, 138862. [CrossRef]
- Cheng, C.; Barceló, J.; Hartnett, A.S.; Kubinec, R.; Messerschmidt, L. COVID-19 Government Response Event Dataset (CoronaNet v.1.0). *Nat. Hum. Behav.* **2020**, *4*, 756–768. [CrossRef]
- Institute of Medicine (US) Forum on Microbial Threats. Ethical and Legal Considerations in Mitigating Pandemic Disease: Workshop Summary. In *Learning from Pandemics Past*; National Academies Press: Washington, DC, USA, 2007. Available online: <https://www.ncbi.nlm.nih.gov/books/NBK54171/> (accessed on 8 February 2021).
- School of Public Health, University of Michigan. What Makes a “Wave” of Disease? An Epidemiologist Explains. Available online: <https://sph.umich.edu/pursuit/2020posts/what-makes-a-wave-of-disease.html> (accessed on 8 February 2021).
- Our World in Data. Research and Data to Make Progress against the World’s Largest Problems. Available online: <https://ourworldindata.org/> (accessed on 8 February 2021).
- Korean Statistical Information Service. Available online: <http://kosis.kr/eng/> (accessed on 8 February 2021).
- European Centre for Disease Prevention and Control. COVID-19—Situation Updates. Available online: <https://www.ecdc.europa.eu/en/covid-19-pandemic> (accessed on 8 February 2021).
- European Centre for Disease Prevention and Control. Download Historical Data (to 14 December 2020) on the Daily Number of New Reported COVID-19 Cases and Deaths Worldwide. Available online: <https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide> (accessed on 8 February 2021).

21. European Centre for Disease Prevention and Control. Available online: <https://opendata.ecdc.europa.eu/covid19/casedistribution/csv> (accessed on 8 February 2021).
22. Davies, N.G.; Klepac, P.; Liu, Y.; Prem, K.; Jit, M.; Pearson, C.A.B.; Quilty, B.J.; Kucharski, A.J.; Gibbs, H.; Clifford, S.; et al. Age-Dependent Effects in the Transmission and Control of COVID-19 Epidemics. *Nat. Med.* **2020**, *26*, 1205–1211. [[CrossRef](#)] [[PubMed](#)]
23. Data on COVID-19 (Coronavirus) by Our World in Data. Available online: <https://github.com/owid/covid-19-data/tree/master/public/data> (accessed on 8 February 2021).
24. Nioi, M.; Napoli, P.E.; Lobina, J.; Fossarello, M.; D’Aloja, E. COVID-19 and Italian Healthcare Workers From the Initial Sacrifice to the mRNA Vaccine: Pandemic Chrono-History, Epidemiological Data, Ethical Dilemmas, and Future Challenges. *Front. Public Health* **2021**, *8*, 591900. [[CrossRef](#)] [[PubMed](#)]
25. Aabed, K.; Lashin, M.M. An analytical study of the factors that influence COVID-19 spread. *Saudi J. Biol. Sci.* **2021**, *28*, 1177–1195. [[CrossRef](#)] [[PubMed](#)]
26. Farseev, A.; Chu-Farseeva, Y.Y.; Qi, Y.; Loo, D.B. Understanding Economic and Health Factors Impacting the Spread of COVID-19 Disease. *medRxiv* **2020**. [[CrossRef](#)]
27. Fortaleza, C.M.C.B.; Guimarães, R.B.; de Almeida, G.B.; Pronunciate, M.; Ferreira, C.P. Taking the inner route: Spatial and demographic factors affecting vulnerability to COVID-19 among 604 cities from inner São Paulo State, Brazil. *Epidemiol. Infect.* **2020**, *148*, e118. [[CrossRef](#)]
28. Bhadra, A.; Mukherjee, A.; Sarkar, K. Impact of population density on Covid-19 infected and mortality rate in India. *Model. Earth Syst. Environ.* **2021**, *7*, 623–629. [[CrossRef](#)]
29. Bargain, O.; Aminjonov, U. “Poverty and COVID-19 in Developing Countries”, Bordeaux Economics Working Papers 2020-08, Groupe de Recherche en Economie Théorique et Appliquée (GREThA). 2020. Available online: <https://ideas.repec.org/p/grt/bdxewp/2020-08.html> (accessed on 8 February 2021).
30. Jia, L.; Li, K.; Jiang, Y.; Guo, X. Prediction and analysis of coronavirus disease 2019. *arXiv* **2020**, arXiv:2003.05447.
31. The Conversation. What Makes a ‘Wave’ of Disease? An Epidemiologist Explains. Available online: <https://theconversation.com/what-makes-a-wave-of-disease-an-epidemiologist-explains-141573> (accessed on 8 February 2021).
32. Nadaraya, E.A. On Estimating Regression. *Theory Probab. Its Appl.* **1964**, *9*, 141–142. [[CrossRef](#)]
33. Watson, G.S. Smooth regression analysis. *Sankhyā Indian J. Stat. Ser. A* **1964**, *26*, 359–372.
34. Bierens, H.J. The Nadaraya—Watson kernel regression function estimator. In *Topics in Advanced Econometrics*; Cambridge University Press: New York, NY, USA, 1994; pp. 212–247. ISBN 0-521-41900-X.
35. The World Bank. Data Bank. Available online: <https://databank.worldbank.org/home> (accessed on 8 February 2021).
36. Franses, P.H. A method to select between Gompertz and logistic trend curves. *Technol. Forecast. Soc. Chang.* **1994**, *46*, 45–49. [[CrossRef](#)]
37. Countries in the World by Population. 2021. Available online: <https://www.worldometers.info/world-population/population-by-country/> (accessed on 8 February 2021).
38. Worldometer. Available online: <https://www.worldometers.info/coronavirus/#countries> (accessed on 8 February 2021).
39. Anzai, A.; Kobayashi, T.; Linton, N.M.; Kinoshita, R.; Hayashi, K.; Suzuki, A.; Yang, Y.; Jung, S.-M.; Miyama, T.; Akhmetzhanov, A.R.; et al. Assessing the Impact of Reduced Travel on Exportation Dynamics of Novel Coronavirus Infection (COVID-19). *J. Clin. Med.* **2020**, *9*, 601. [[CrossRef](#)] [[PubMed](#)]
40. Alderton, G. Aging immunity. *Science* **2020**, *369*, 264–266. [[CrossRef](#)]
41. World Health Organization. Ageing and Health. Available online: <https://www.who.int/news-room/fact-sheets/detail/ageing-and-health> (accessed on 8 February 2021).
42. Hirokawa, K.; Utsuyama, M.; Kasai, M.; Kurashima, C. Aging and Immunity. *Pathol. Int.* **2008**, *42*, 537–548. [[CrossRef](#)] [[PubMed](#)]
43. Centers for Disease Control and Prevention. COVID-19—Older Adults. Available online: <https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/older-adults.html> (accessed on 8 February 2021).
44. Mueller, A.L.; McNamara, M.; Sinclair, D.A. Why does COVID-19 disproportionately affect older people? *Aging* **2020**, *12*, 9959–9981. [[CrossRef](#)]
45. Broeck, J.V.D.; Eeckels, R.; Vuylsteke, J. Influence of nutritional status on child mortality in rural Zaire. *Lancet* **1993**, *341*, 1491–1495. [[CrossRef](#)]
46. Our World in Data. In Which Countries Do People Smoke the Most? Available online: <https://ourworldindata.org/which-countries-smoke-most> (accessed on 8 February 2021).
47. Ho, F.K.; Petermann-Rocha, F.; Gray, S.R.; Jani, B.D.; Katikireddi, S.V.; Niedzwiedz, C.L.; Foster, H.; Hastie, C.E.; Mackay, D.F.; Gill, J.M.R.; 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* **2020**, *15*, e241824. [[CrossRef](#)]
48. Gülsen, A.; Yigitbas, B.A.; Uslu, B.; Drömann, D.; Kilinc, O. The Effect of Smoking on COVID-19 Symptom Severity: Systematic Review and Meta-Analysis. *Pulm. Med.* **2020**, *2020*, 7590207. [[CrossRef](#)]
49. Wenzl, T. *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, 2020; p. JRC121837, ISBN 978-92-76-22062-6. [[CrossRef](#)]

50. Man, W.D.C.; Man, W.D.C.; Weber, M.; Weber, M.; Palmer, A.; Palmer, A.; Schneider, G.; Wadda, R.; Jaffar, S.; Jaffar, S.; 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* **1998**, *3*, 678–686. [CrossRef] [PubMed]
51. Fernandez, I.D.; Himes, J.H.; De Onis, M. Prevalence of nutritional wasting in populations: Building explanatory models using secondary data. *Bull. World Health Organ.* **2002**, *80*, 282–291. [PubMed]
52. World Health Organization. Novel Coronavirus (2019-nCoV) Situation Report-21. 2020. Available online: <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200210-sitrep-21-ncov.pdf> (accessed on 8 February 2021).
53. Müller, O.; Krawinkel, M. Malnutrition and health in developing countries. *Can. Med. Assoc. J.* **2005**, *173*, 279–286. [CrossRef] [PubMed]
54. Saunders, J.; Smith, T. Malnutrition: Causes and consequences. *Clin. Med.* **2010**, *10*, 624–627. [CrossRef] [PubMed]
55. Centers for Disease Control and Prevention. Domestic Travel During COVID-19. Available online: <https://www.cdc.gov/coronavirus/2019-ncov/travelers/travel-during-covid19.html> (accessed on 8 February 2021).
56. Siegenfeld, A.F.; Bar-Yam, Y. The impact of travel and timing in eliminating COVID-19. *Commun. Phys.* **2020**, *3*, 204. [CrossRef]
57. You, J. Lessons From South Korea's Covid-19 Policy Response. *Am. Rev. Public Adm.* **2020**, *50*, 801–808. [CrossRef]
58. Cha, V.; Kim, D. A Timeline of South Korea's Response to COVID-19, Center for Strategic and International Studies. 2020. Available online: <https://www.csis.org/analysis/timeline-south-koreas-response-covid-19> (accessed on 8 February 2021).
59. The World Bank. Poverty—Overview. Available online: <https://www.worldbank.org/en/topic/poverty/overview#:~:text=The%20global%20extreme%20poverty%20rate,%245.50%20a%20day%20in%202017> (accessed on 4 February 2021).
60. El-Sayed, A.; Kamel, M. Climatic changes and their role in emergence and re-emergence of diseases. *Environ. Sci. Pollut. Res.* **2020**, *27*, 22336–22352. [CrossRef]
61. Rayan, R.A. Seasonal variation and COVID-19 infection pattern: A gap from evidence to reality. *Curr. Opin. Environ. Sci. Health* **2021**, *20*, 100238. [CrossRef]
62. Our World in Data. Global Extreme Poverty. Available online: <https://ourworldindata.org/extreme-poverty#:~:text=We%20can%20also%20see%20that,of%20important%20changes%20across%20time>. (accessed on 4 February 2021).
63. UNICEF. West and Central Africa: More than 15 Million Cases of Acute Malnutrition Expected in 2020. Available online: <https://www.unicef.org/press-releases/west-and-central-africa-more-15-million-cases-acute-malnutrition-expected-2020> (accessed on 4 February 2021).
64. Annan, K. Data can help to end malnutrition across Africa. *Nat. Cell Biol.* **2018**, *555*, 7. [CrossRef]
65. World by Map. Young Population. Available online: <https://www.citypopulation.de/en/world/bymap/youngpopulation/> (accessed on 4 February 2021).
66. Worldometer. Life Expectancy of the World Population. Available online: <https://www.worldometers.info/demographics/life-expectancy/> (accessed on 4 February 2021).
67. Percentage of Population Who Smoke on a Daily Basis in Selected Countries as of 2017. Available online: <https://www.statista.com/statistics/236600/proportion-of-smokers-in-the-population-of-selected-countries/> (accessed on 8 February 2021).
68. Worldometer. GDP by Country. Available online: <https://www.worldometers.info/gdp/gdp-by-country/> (accessed on 4 February 2021).
69. OECD/World Health Organization. Mortality from cardiovascular disease. In *Health at a Glance: Asia/Pacific 2018: Meas-Uring Progress Towards Universal Health Coverage*; OECD Publishing: Paris, France, 2018. [CrossRef]
70. Finegold, J.A.; Asaria, P.; Francis, D.P. Mortality from ischaemic heart disease by country, region, and age: Statistics from World Health Organisation and United Nations. *Int. J. Cardiol.* **2013**, *168*, 934–945. [CrossRef]
71. World Health Organization. Coronavirus Disease (COVID-19): Tobacco. Available online: [https://www.who.int/westernpacific/news/q-a-detail/coronavirus-disease-covid-19-tobacco#:~:text=Tobacco%20smokers%20\(cigarettes%2C,communal%20and%20social%20settings](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 on 8 February 2021).
72. Riou, J.; Althaus, C.L. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Eurosurveillance* **2020**, *25*, 2000058. [CrossRef]
73. Chan, J.F.W.; Yuan, S.; Kok, K.H.; To, K.K.W.; Chu, H.; Yang, J.; Xing, F.; Liu, J.; Yip, C.C.Y.; Poon, R.W.S.; 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* **2020**, *395*, 514–523. [CrossRef]
74. D'Aloja, E.; Finco, G.; Demontis, R.; Napoli, P.E.; Fossarello, M.; Nioi, M. COVID-19 and medical liability: Italy denies the shield to its heroes. *EClinicalMedicine* **2020**, *25*, 100470. [CrossRef]
75. Nioi, M.; Napoli, P.E.; Finco, G.; Demontis, R.; Fossarello, M.; d'Aloja, E. Fear of the COVID-19 and medical liability. Insights from a series of 130 consecutive medico-legal claims evaluated in a single institution during SARS-CoV-2-related pandemic. *Signa Vitae A J. Intensiv. Care Emerg. Med.* **2021**, *17*, 79–85. [CrossRef]
76. Zhang, X.; Ma, R.; Wang, L. Predicting turning point, duration and attack rate of COVID-19 outbreaks in major Western countries. *Chaos Solitons Fractals* **2020**, *135*, 109829. [CrossRef] [PubMed]
77. Singh, R.; Adhikari, R. Age-structured impact of social distancing on the COVID-19 epidemic in India. *arXiv* **2020**, arXiv:2003.12055.

- 
78. Hellewell, J.; Abbott, S.; Gimma, A.; Bosse, N.I.; Jarvis, C.I.; Russell, T.W.; Munday, J.D.; Kucharski, A.J.; Edmunds, W.J.; Sun, F.; et al. Feasibility of controlling covid-19 outbreaks by isolation of cases and contacts. *Lancet Glob. Health* **2020**, *8*, e488–e496. [[CrossRef](#)]
  79. Tanigawa, Y.; Rivas, M. Initial Review and Analysis of COVID-19 Host Genetics and Associated Phenotypes. *Preprints* **2020**, 2020030356. [[CrossRef](#)]