**A comparative study of social-economic factors on the transmission dynamics of coronavirus disease (COVID-19).**

Hakyong Kim1, Yeonghyeon Ko2,3, Catherine Apio4, Taewan Goo4, Gyujin Huh3, Kyulhee Han4, Taehyun Kim3, Hyewon Chung5, Doeun Lee4, Taesung Park3\*

1Department of Industrial Engineering, Seoul National University, Seoul 08826, Republic of Korea

2Department of Archeology and Art History, Seoul National University, Seoul 08826, Republic of Korea

3Department of Statistics, Seoul National University, Seoul 08826, Republic of Korea

4Interdisciplinary Program in Bioinformatics, Department of Statistics, Seoul National University, Seoul 08826, Republic of Korea

5Department of Chemistry, Seoul National University, Seoul 08826, Republic of Korea

\*Corresponding author, Tel: 82-2-880-8924, Fax: 82-2-883-6144, E-mail: taesungp@stats.snu.ac.kr

**Abstract**

The outbreak of the novel coronavirus, COVID-19, which was declared a global pandemic by WHO, is most serious public health threat seen in the respiratory virus since the 1918 H1N1 influenza pandemic. Social contact critically determines the spread of the infection and, in the absence of vaccine or effective treatment, the rapid spread of this disease elicited a wide range of large-scale social distancing measures from different governments across the globe as the most effective means of mitigating COVID-19. Here, we employed growth-curve models (Logistic, Bertalanffy and Gompertz models) for cumulative confirmed cases available daily from forty countries around the world to analyze the relationship between these models' summary statistics and country's development status using factors like life expectancy, GDP, urbanization, etc on the spread of COVID-19.We investigated the relationship between the transmission dynamics of the pandemic and the development status of a country as an indicator of selected variables like GDP, urbanization, life expectancy, etc. Also, we tested incorporated governments’ intervention (school closure, workplace closing, close public transport, income support, contact tracing, etc.) in mitigation of COVID-19 using the Generalized Estimation Equation (GEE) model. Our analysis found that a country’s population, social-economic and healthcare infrastructures have both weak and strong significant relationships with transmission dynamics of COVID-19 for all the indicator variables used. Developed countries tend to take long time before we see an increase in the number of daily confirmed cases but after that, the expansion rate is rapid among the population. Out of the 17 policies analyzed, closure of public transport, international travel controls and contact tracing are the only policies found to have significant impact in the control of numbers of daily confirmed cases.

**Keywords**: SARS-COV-2, COVID-19, growth-curve based models, Logistic, Gompertz, Bertalanffy

**Introduction**

A number of unexplained pneumonia cases which were successively discovered in Wuhan, Hubei, China, in early December 2019 and have been confirmed to be severe acute respiratory syndrome caused by a novel coronavirus 2 (SARS-CoV-2), has spread rapidly across the globe [1]. The spread of coronavirus disease 2019 (COVID-19) has become a global threat and the World Health Organization (WHO) declared COVID-19 a global pandemic on March 11, 2020 [2] and the public health threat it represents is the most serious seen in a respiratory virus since the 1918 H1N1 influenza pandemic [3]. As of December 22, 2020, there were a total of 77, 804, 369 confirmed cases and 1, 7111, 177 deaths from COVID-19 worldwide [4].

A look at history tells that pandemics and epidemics have consistently and significantly affected human history, and governments have continually implemented a variety of policies in their response like quarantines during Ebola outbreak [5-6]. The COVID-19 pandemic also elicited a wide range of unprecedented policies such as closing schools, travel restrictions, bans on public gatherings, stay at home orders, closure of public transportation, etc. from societies and governments around the world designed to mitigation and suppress the pandemic situation, as inferred from past epidemics [7-8]. The suppression of social contact in workplaces, schools and other public spheres is the target of such measures [9-10]. Specifically, governments adopted these policies hoping they will reduce the amount of person-to-person contact in the population. In theory, reducing the frequency of contact means that there will be fewer opportunities for the virus to pass from one person to the next [11].

The effectiveness of implemented government policies in the alleviation of COVID-19 pandemic has already been demonstrated in a lot of published papers [12-15]. All these studies prove the importance of government policies implemented in the control of the pandemic by governments. Also, evidence from microsimulation models suggests that these interventions will decrease the size of the epidemic and redistribute the number of cases over time [11,16], reducing the risk that local health care systems will be overwhelmed by surges in demand for health services [11]. However, no analysis has been done to determine whether relationships exist between a country’s development status and their ability in containment of the COVID-19 pandemic. For we believe that a country’s development status will have an impact on the rate of transmission of the SARS-COV-2 virus among the population and the policies to be implemented during a pandemic, hoping these would provide some insights for policy makers and governments on a probable course of action during a pandemic depending on the country’s development status.

In this article, we present growth-curve based models (Logistic, Bertalanffy and Gompertz Models) to study the relationship between a country’s development rank index using selected indicator variables like GDP, trade, life expectancy, urbanization, etc., and the transmission dynamics of COVID-19 using the daily confirmed cases. Our objective to find social and economic variables (indices) which are highly related with COVID-19 confirmation tendencies. Confirmation tendency is summarized into several statistical parameters from growth-curve based models which have interpretable meanings. We use simple linear regression between parameters and time-independent variable (social, economic variables).

1. **Materials and Methods**
   1. **COVID-19 Dataset**

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 [17-19]. 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 31 August 2020 was used in downstream analysis.

Also, Data smoothing uses an algorithm to remove noise from a data set, allowing important patterns to easily be discovered. Thereafter, daily confirmed case data was smoothed by simple moving the 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 abnormality greater or smaller, 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 also had randomness, so there was a limit to the analysis. Therefore, considering the period of 7 days, the window size of 7 and simple moving average (SMA) was used before model fitting as shown below;

where *p* is the number of confirmed cases.

* 1. **Time-independent Variable Dataset**

Time-independent variable (Table 1) datasets are publicly available datasets easily obtained from *Our World in Data* website [20] and KOSIS [21]. *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.

KOSIS [21] is a central government organization for statistics in South Korea with aims at prompting services of overall planning and coordination of national statistics, establishment of statistical standards, production & distribution of various economic and social statistics, processing & management of statistical information and provision of various statistical data. We 15 variables like Public social welfare 2018, national competiveness 2020, pharmaceutical sales, trade, travel, aging index etc., related to a country’s development status. These variables are measured over a period of several years, so we selected the year with the minimum number of missing values within 2016-2019 and then standardized between themselves. The selected variables are described more in

|  |  |  |  |
| --- | --- | --- | --- |
| **Summary of time-independent variables** | | | |
| **Source** | **Category** | **Variable name** | **Variable information** |
| **Our World in Data** | **National accounts** | population | Population in 2020 |
| population\_density | Number of people divided by land area, measured in square kilometers, most recent year available |
| gdp\_per\_capita | Gross domestic product at purchasing power parity (constant 2011 international dollars), most recent year available |
| **Age** | median\_age | Median age of the population, UN projection for 2020 |
| aged\_65\_older | Share of the population that is 65 years and older, most recent year available |
| age\_70\_older | Share of the population that is 70 years and older in 2015 |
| life\_expectancy | Life expectancy at birth in 2019 |
| **Health, Society, And Welfare** | cardiovasc\_death\_rate | Death rate from cardiovascular disease in 2017 (annual number of deaths per 100,000 people) |
| diabetes\_prevalence | Diabetes prevalence (% of population aged 20 to 79) in 2017 |
| female\_smokers | Share of women who smoke, most recent year available |
| male\_smokers | Share of men who smoke, most recent year available |
| handwashing\_facilities | Share of the population with basic handwashing facilities on premises, most recent year available |
| hospital\_beds\_per\_thousand | Hospital beds per 1,000 people, most recent year available since 2010 |
| extreme\_poverty | Share of the population living in extreme poverty, most recent year available since 2010 |
| human\_development\_index | Summary measure of average achievement in key dimensions of human development: a long and healthy life, being knowledgeable and have a decent standard of living |
| **KOSIS** | **Trade, International Balance** | Trade Dependency (2016) |  |
| **Transportation, telecommunications** | Internet Usage (2017) |  |
| **Price, household** | Savings rate as a percentage of GDP (2016) |  |
| Participation rate of economic activity (2016) |  |
| **Environment** | Average annual temperature (1961-1990) |  |
| Annual precipitation (1961-1990) |  |
| **National accounts** | 1-person GDP per capita (2016) |  |
| GDP Growth (2018) |  |
| Gross National Income (2017) |  |
| **Education, Culture, Science** | Number of international travelers (2017) |  |
| Number of foreign visitors (2017) |  |
| Human Development Index (2017) |  |
| National Competitiveness (2019) |  |
| **Health, Society, And Welfare** | Infant vaccination rate (2017) |  |
| Percentage of malnourished population (2016) |  |

Appendix 1.

**Table 1**: List of the time-independent variables obtained from the KOSTAT and *Our World in Data* websites

**3. Methods**

3.1. Growth Curve Models for modelling the spread dynamics of COVID-19

Under this analysis, four statistical models, Logistic model, Bertalanffy model and Gompertz model were used for fitting analysis on the COVID-19 daily 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 [22].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.

Bertalanffy 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 the first case, is the time when the first case occurred.

Gompertz Model

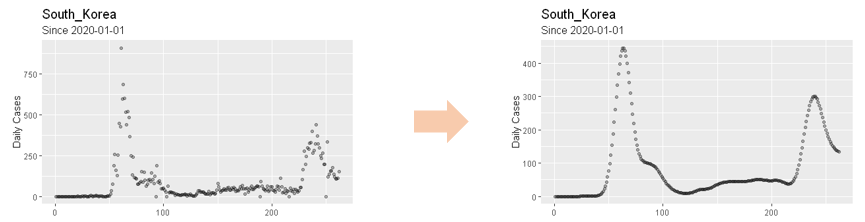
(3)

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.

3.2. Segmentation Algorithm

As the COVID-19 situation prolongs, fitting an growth curve model on daily confirmed cases over long period of time becomes impossible as it no longer takes on sigmoid curve. To fit the above growth curve models, there is a need to divide study period of countries experiencing more than one wave of the pandemic into segments (the time period where cumulative confirmed cases has the s-curve). So we applied segmentation method which can systematically divide study periods for each country into segments.

Segmentation is the method of finding peaks (the timestamp at which daily new confirmed cases is highest for each segment) and breakpoints (the timestamp which splits the adjacent two segments) during study period for each country. So it consists of peak detection algorithm and breakpoint detection algorithm. To detect peaks and breakpoints, graph of daily new confirmed cases must be smooth. But daily new confirmed cases have high randomness arising from 1) the fact that daily new confirmed cases has 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 [23-25] with Gaussian kernel and window size 7(different from SMA) to smoothen the daily new confirmed cases as demonstrated in Figure 1 using South Korea’s daily confirmed cases as an example. For the convenience of notation, let be the daily new confirmed cases from data, estimated daily new confirmed cases using above estimator respectively since December 31, 2019.



**Figure 1**: Daily new confirmed cases before and after Nadaraya-Watson kernel regression.

Peak detection (Algorithm 1) utilizes local maximum condition on convex function. However, considering daily new confirmed cases being discrete time series data, local maximum condition is substituted with

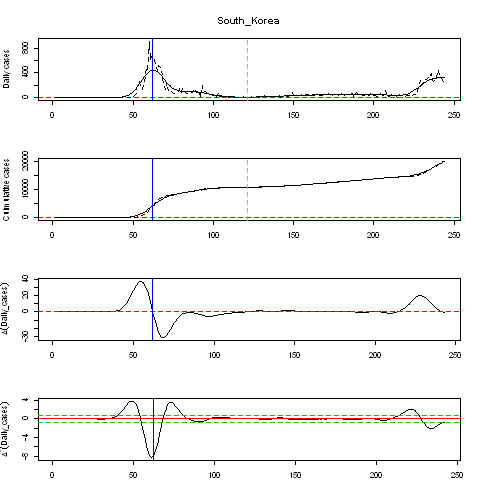
, (4)

where and .

And since Equation(4) is not applicable due to discontinuity and small variances of , we use following condition

, (5)

where is sensitivity level and is set of time indices from January 1, 2020 to 31 August 2020 (Figure 2). And 3 additional conditions ((a) Exclusion of small peaks, (b) Resolution criteria and (c) Exclusioin of peaks which are vibrations on increasing trend) are used in peak detection to enhance robustness. After finding all the peaks, breakpoints (Algorithm 2) are selected either as timestamps which have the smallest daily new confirmed cases between two peaks or the timestamp where the cumulative confirmed case of the last segment saturates (that is the s-curve of last segment is at the last stage*).*(Figure 2) consists of 4 plots. 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 . 3rd, 4th plot are graphs of . In 4th plot, green dotted line represents sensitivity level.



**Figure 2** : on South Korea’s daily new confirmed cases.

The Segmentation method was applied to 134 countries among 213 countries which met . If is too small, segmentation algorithm would be difficult to apply due to small variances in . The Segmentation algorithm is explained in details in the Supplementary materials.

3.3. Segmented Growth Curve models

Segmented Growth Curve models (Segmented Logistic model, Segmented Bertalanffy model and Segmented Gompertz model) fit the above mentioned Growth curve models ((1), (2) and (3)) 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.

(7)

(8)

(9)

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 consider only 1st and 2nd segment since most countries have 1 or 2 segment(1 segment : 62, 2 segments : 65, 3 segments : 7) and the number of countries which has 3rd segment are very few so that it is insignificant to use in Linear regression. For countries with more than 2 segments, our analysis period was cut off at the 2nd breakpoint. Therefore, for countries more than 2 waves, segmented growth curve model produces two sets of parameters for each segment.

Also, correlation analysis for segmented Logistic and Gompertz models with a breakpoint derived based on the Peak and Breakpoint Detection Algorithms was implemented between the log-scaled of parameters and the segments, determine the similarity of these parameters across the two models.

3.4. Simple Linear regression (SLR) model

The above 3 growth curve models summarize the pandemic situation into three parameters(, , ) for countries with 1 segment and into six parameters(, , , , , ) for countries with more than 1 segment. Each of these parameters from three model are then regressed against the time-independent variables shown in Table 1. The model is as below;

(10)

where y is a segmented growth curve parameter (a, b or c) for model segment and country , and are regression coefficients, and is the time-independent variable. F- statistic is performed to test the significance of for each major time-independent variable with the aim of finding out if the variables have any significant relationship with y, a measure of the spread dynamics of COVID-19 for a country.

1. **Results**
   1. The Interpretation of correlation coefficients between the log-scaled parameters of the Growth Curve Models

With Segmentation Algorithm and MSSE(the criterion for filtering not properly fitted countries was set at 0.4.) described in the above section, 124 countries are fitted and 5 countries are excluded among 134 countries for Logistic Models, so the number of properly fitted countries is 119. On the other hand, for Gompertz Models, 119 and 5 countries are included and excluded respectively, which means that 114 countries are remained.

Before performing the simple linear regression for the coefficients and the time-independent variables, we tried to interpret the correlation coefficients by both models and each segment. The points we focused on are whether the tendency of coefficient variation is maintained in both Logistic and Gompertz models, and whether the coefficients , , are correlated in the same segment.

For Logistic Models (Supplementary Figure 2), correlation analysis between two parameters and , the maximum number of predicted cumulative confirmed cases in the 1st segment and 2nd segment respectively, have a strong positive correlation of 0.7. On the other hand, it was found that there was little correlation between the parameters and (-0.089) and in and (-0.32). While for Gompertz Models (Supplementary Figure 3), similar with Logistic Model, two parameters and , the maximum number of predicted cumulative confirmed cases in the 1st segment and 2nd segment respectively, have a strong positive correlation (0.57) while there was no correlation between the coefficients and (0.15), also in and (-0.34).

The parameters and have little correlation (-0.069 and 0.077 in the 1st segment, -0.27 and -0.31 in the 2nd segment for Logistic and Gompertz Models) that is negligible when analyzing correlation coefficients. However, in the 1st segment, the correlation coefficients between parameters and for both Logistic(-0.52) and Gompertz(-0.66) models show that the relationship between the cumulative confirmed cases and the slope of the model has a negative correlation. It is same with the parameters and (-0.55 for Logistic Model and -0.7 for Gompertz Model) in the 2nd segment (Supplementary Figure 4 and 5).

4.2. The Relationship between socioeconomic indicators and parameters of the Growth Curve Models

- The Korean Statistical Information Service(KOSIS) and Our World in Data provide socioeconomic indicators measured over a year such as population density, the rate of malnutrition, etc. From KOSIS and Our World in Data, 18 and 15 indicators were chosen, respectively.

- We implemented simple linear regression between the above variables and the parameters of the Growth Curve Models (Logistic and Gompertz Models). The results are on the below tables. Under the significance level 0.05, the time-independent socioeconomic indicators such as Population, Median age, aged over 65 older, Aged over 70 older, GDP, Diabetes prevalence, Life expectancy, HDI (Human Development Index) for *Our World in Data*, and Malnutrition, Exportation, Importation, Total GDP, Travel In, Travel Out, Aging index, Birth rate for *KOSIS*.

- The indicators, turned out to be statistically significant under significance level 0.05, can be divided into two groups depending on whether the estimated regression coefficient of each indicator is negative or positive.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SLR results for the 1st Segment** | | | | | |
| **Indicator\Parameter** | | **Logistic Model** | | **Gompertz Model** | |
| **coefficient** | **p-value** | **coefficient** | **p-value** |
| Child Vaccination rate |  |  |  |  |  |
|  |  |  |  |  |
|  | 0.000814228 | 0.04736071 |  |  |
| Malnutrition rate |  |  |  |  |  |
|  | 0.094611423 | 0.0000818 | 0.683770701 | 0.010293602 |
|  | -0.001949621 | 0.004626816 | -0.001626774 | 0.001259812 |
| National Competitiveness  (2019) |  |  |  |  |  |
|  |  |  |  |  |
|  | -0.00088953 | 0.018372709 | -0.000691172 | 0.014797163 |
| Temperature |  |  |  |  |  |
|  | 0.00269697 | 0.016471182 |  |  |
|  |  |  |  |  |
| Rain |  | 416.3360187 | 0.02839313 |  |  |
|  | 0.000519428 | 0.032079459 |  |  |
|  |  |  |  |  |
| Pharmacy Sales  (2018) |  |  |  |  |  |
|  | 0.0000467 | 0.000448077 |  |  |
|  |  |  |  |  |
| Exportation  (2016) |  |  |  |  |  |
|  | -0.016452595 | 0.046720281 |  |  |
|  |  |  | 0.000333569 | 0.037698585 |
| Importation  (2016) |  |  |  |  |  |
|  | -0.010660886 | 0.015840189 |  |  |
|  |  |  |  |  |
| Travel out  (2017) |  |  |  |  |  |
|  |  |  |  |  |
|  | -0.001365408 | 0.042172691 |  |  |
| Aging Index  (2020) |  |  |  |  |  |
|  | -0.009629515 | 0.000670669 |  |  |
|  | 0.000226851 | 0.009358969 |  |  |
| GDP per P  (2019) |  |  |  |  |  |
|  | -0.0000149 | 0.027771485 |  |  |
|  | 0.000000636 | 0.00179224 | 0.000000586 | 0.0000138 |
| Birth Rate  (2020) |  |  |  |  |  |
|  |  |  |  |  |
|  |  |  | -0.006039077 | 0.023478317 |
| Population |  | 0.002625297 | 0.0000000513 | 0.054930982 | 4.07E-48 |
|  |  |  |  |  |
|  |  |  |  |  |
| Median Age |  |  |  |  |  |
|  | -0.056172963 | 0.000693256 |  |  |
|  | 0.001132705 | 0.023270785 | 0.00123353 | 0.000246605 |
| Aged 65 older |  |  |  |  |  |
|  | -0.0828007 | 0.000318613 |  |  |
|  | 0.001910233 | 0.005765253 | 0.001926844 | 0.0000324 |
| Aged 70 older |  |  |  |  |  |
|  | -0.1238193 | 0.00025543 |  |  |
|  | 0.00271752 | 0.00758984 | 0.00279911 | 0.0000404 |
| GDP per Capita |  |  |  |  |  |
|  | -0.0000148 | 0.041114537 |  |  |
|  |  |  | 0.000000418 | 0.00403012 |
| Extreme poverty |  |  |  |  |  |
|  | 0.025938272 | 0.005945381 | 0.269815086 | 0.003819147 |
|  |  |  |  |  |
| Cardiovascular Death |  |  |  |  |  |
|  |  |  |  |  |
|  | -0.0000937 | 0.011142145 | -0.0000694 | 0.006647072 |
| Female Smokers |  |  |  |  |  |
|  | -0.041890337 | 0.004397759 |  |  |
|  |  |  | 0.000913903 | 0.005661459 |
| Hospital Beds per Thousand |  |  |  |  |  |
|  | -0.178519644 | 0.005243062 |  |  |
|  |  |  | 0.003664155 | 0.006108424 |
| Life Expectancy |  |  |  |  |  |
|  | -0.046678931 | 0.025687166 |  |  |
|  | 0.001279172 | 0.040121585 | 0.001357435 | 0.001186577 |
| Human Development Index |  |  |  |  |  |
|  | -2.154178075 | 0.037197296 |  |  |
|  |  |  | 0.051780414 | 0.012390061 |

- The important changes from the 1st segment to the 2nd segment for each indicator we focused are (1) whether the estimated coefficients of the parameter *a* (the number of estimated total cases) and *c* (the estimated spread rate) is maintained, and (2) whether the estimated coefficients have newly become statistically significant for the 2nd segment.

**+ Interpretation**

**References**

1. Q Li, X Guan, P Wu, et al. Early transmission dynamics in Wuhan, China of the novel Coronavirus- infected pneumonia. N Engl J Med (2020) published online Jan 29. DOI:10.1056/NEJMoa2001316
2. WHO director-general’s opening remarks. <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19-11-march-2020>
3. <https://www.worldometers.info/coronavirus/>
4. Bootsma MCJ, Ferguson NM. The effect of public health measures on the 1918 influenza pandemic in U.S. cities. Proc Natl Acad Sci U S A 2007;104(18):7588–93.
5. Cheng, C., Barceló, J., Hartnett, A.S. et al. COVID-19 Government Response Event Dataset (CoronaNet v.1.0). Nat Hum Behav (2020). https://doi.org/10.1038/s41562-020-0909-7
6. Institute of Medicine (US) Forum on Microbial Threats. Ethical and Legal Considerations in Mitigating Pandemic Disease: Workshop Summary. Washington (DC): National Academies Press (US); 2007. 1, Learning from Pandemics Past. Available from: https://www.ncbi.nlm.nih.gov/books/NBK54171/
7. Hsiang, S., Allen, D., Annan-Phan, S. et al. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. Nature 584, 262–267 (2020). <https://doi.org/10.1038/s41586-020-2404-8>.
8. Ferguson, N. M. et al. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Technical Report (Imperial College London, 2020).
9. Zhang, Xiaolei, Renjun Ma, and Lin Wang. Predicting turning point, duration and attack rate of COVID-19 outbreaks in major Western countries. Chaos, Solitons & Fractals (2020): 109829.
10. Singh, Rajesh, and R. Adhikari. Age-structured impact of social distancing on the COVID-19 epidemic in India. arXiv preprint arXiv:2003.12055 (2020).
11. Davies, N.G., Klepac, P., Liu, Y.  et al.   Age-dependent effects in the transmission and control of COVID-19 epidemics.   Nat Med   26, 1205–1211 (2020). https://doi.org/10.1038/s41591-020-0962-9.
12. Hsiang, S., Allen, D., Annan-Phan, S. et al. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. Nature 584, 262–267 (2020). <https://doi.org/10.1038/s41586-020-2404-8>
13. Yun Qiu, Xi Chen and Wei Shi. Impacts of social and economic factors on the transmission of coronavirus disease (COVID-19) in China. medRxiv 2020.03.13.20035238. doi: <https://doi.org/10.1101/2020.03.13.20035238>
14. Liang XH, Tang X, Luo YT, Zhang M, Feng ZP. Effects of policies and containment measures on control of COVID-19 epidemic in Chongqing.  World J Clin Cases . 2020;8(14):2959-2976. doi:10.12998/wjcc.v8.i14.2959
15. Haug, N., Geyrhofer, L., Londei, A.  et al.   Ranking the effectiveness of worldwide COVID-19 government interventions.   Nat Hum Behav   4,   1303–1312 (2020). <https://doi.org/10.1038/s41562-020-01009-0>
16. Flaxman, S., Mishra, S., Gandy, A.  et al.   Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe.   Nature   584, 257–261 (2020). <https://doi.org/10.1038/s41586-020-2405-7>.
17. European Centre for Disease Prevention and Control. <https://www.ecdc.europa.eu/en/covid-19-pandemic>
18. <https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide>.
19. <https://opendata.ecdc.europe.eu/covid19/casedistribution/csv.>
20. <https://ourworldindata.org/>
21. <https://kostat.go.kr>
22. Lin Jia, Kewen Li, Yu Jiang, Xin Guo and Ting Zhao. Prediction and analysis of Coronavirus Disease 2019 (2019)
23. Nadaraya, E. A. (1964). "On Estimating Regression". Theory of Probability and Its Applications. 9 (1): 141–2. doi:10.1137/1109020.
24. Watson, G. S. (1964). "Smooth regression analysis". Sankhyā: The Indian Journal of Statistics, Series A. 26 (4): 359–372. JSTOR 25049340.
25. Bierens, Herman J. (1994). "The Nadaraya–Watson kernel regression function estimator". Topics in Advanced Econometrics. New York: Cambridge University Press. pp. 212–247. ISBN 0-521-41900-X.

**Supplementary Information**

1. **Pseudo-code of Segmentation Algorithm**

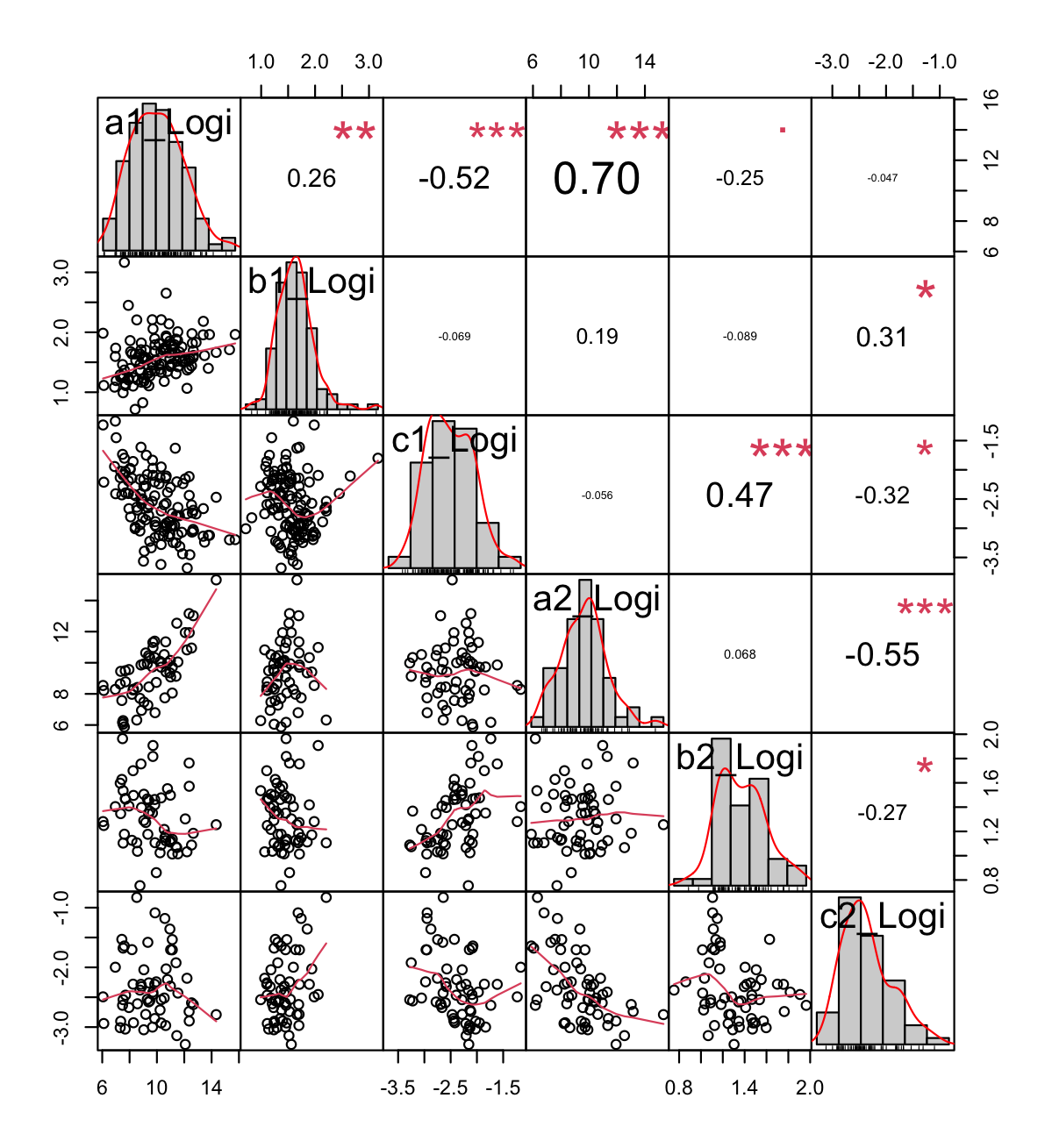
Here, we denote by .

|  |
| --- |
| **Algorithm 1** Peak Detection Algorithm |
| 1. (Initialization)  declare where  2. (Local maximum condition)  **while** **do**  **if** **then**    **end if**  **end while**  3. (Excluding small peaks)  **if** **then**    **end if**  4. (Resolution criteria)  **if** **then**  declare *p* = 1,  **while** *p* **do**  **if** **then**  **if** **then**    **else**    **end if**  **else**    **end if**  *p* *p+1*  **end if**  5. (Exclude peaks which are vibrations on increasing trend)  **if** **then**  declare ;  **if** **then**    **end if**  **end if** |

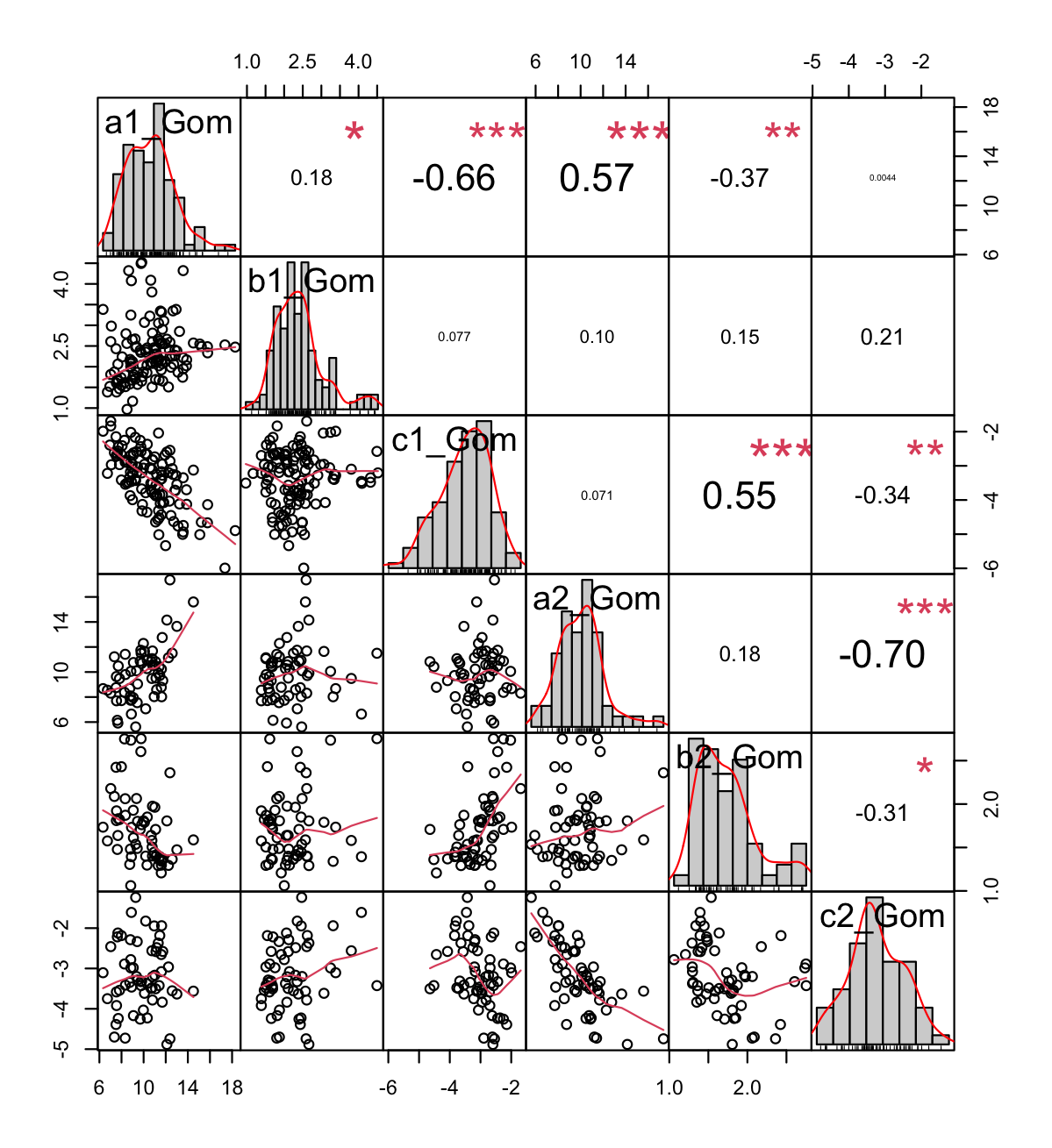
|  |
| --- |
| **Algorithm 2** Breakpoint Detection Algorithm |
| 1. (Initialization)  declare  2. (Between peaks)  **if** **then**  **for each** **do**  B;  **end do**  **end if**  3.(After the last peak)  declare U  **if** **then**  B ;  **end if** |

|  |
| --- |
| **Algorithm 2** Breakpoint Detection Algorithm |
| 1. (Initialization)  declare  2. (Between peaks)  **if** **then**  **for each** **do**  B;  **end do**  **end if**  3.(After the last peak)  declare U  **if** **then**  B ;  **end if** |

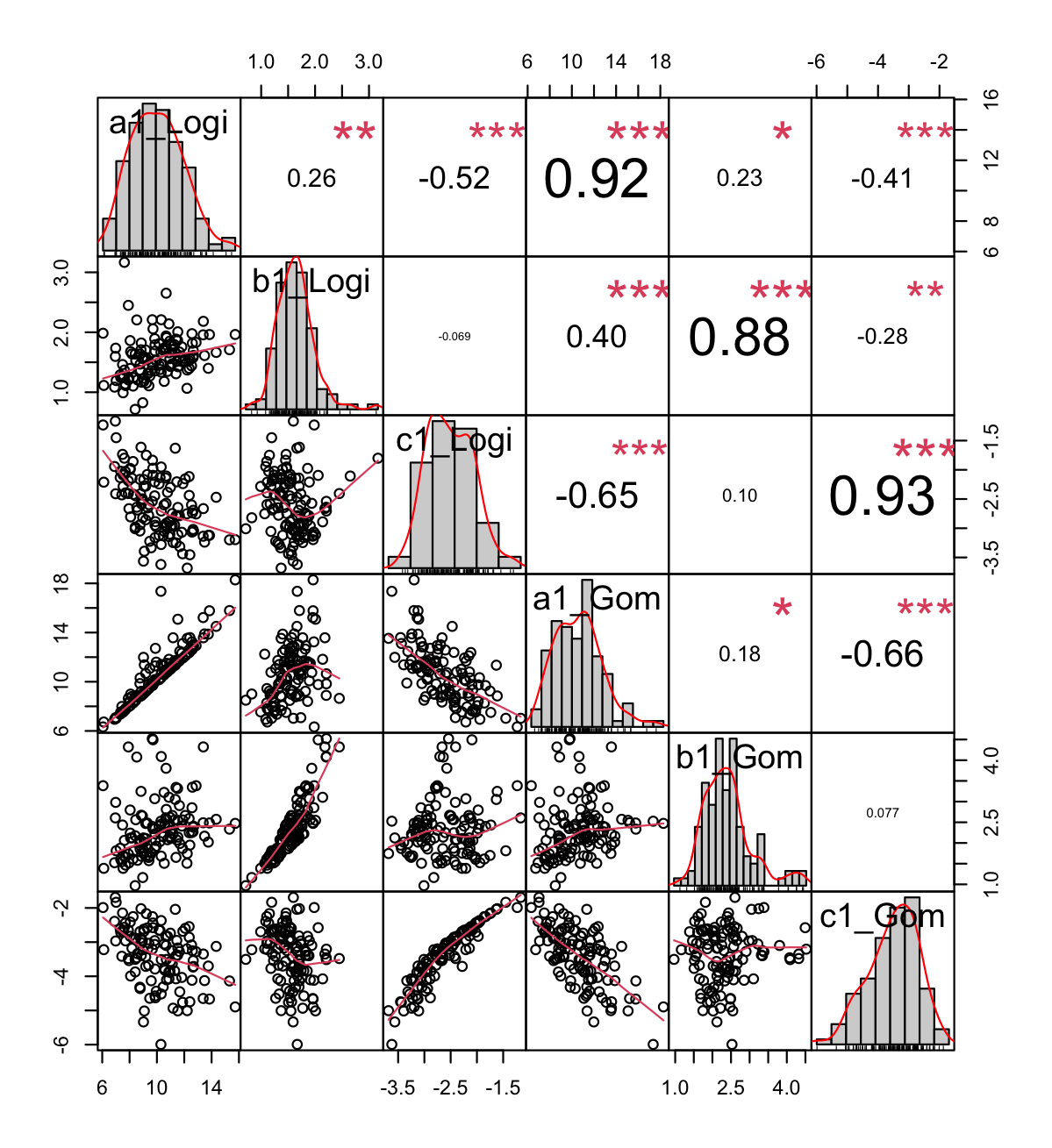
1. **Correlation analysis between Growth curve models**

****

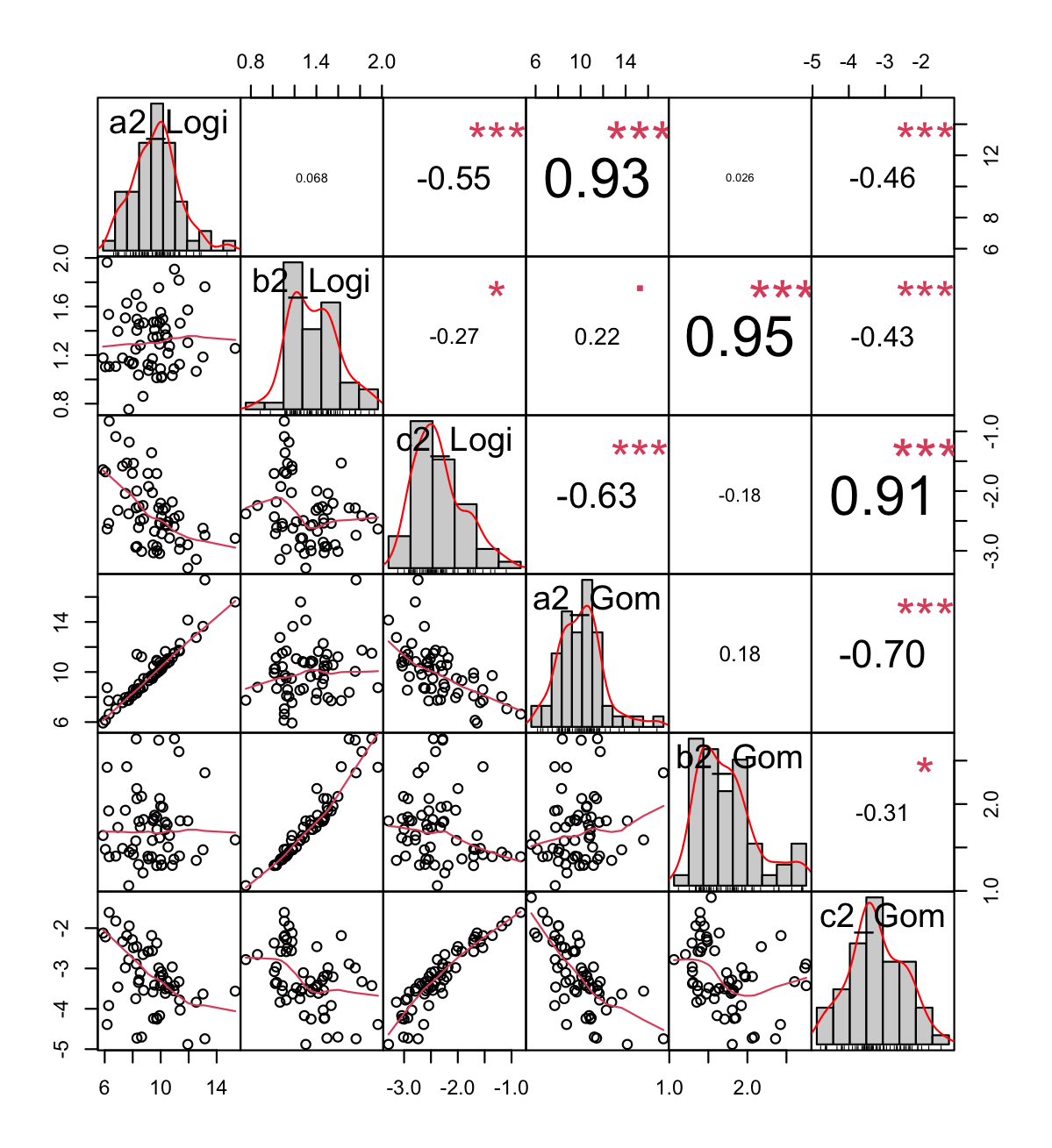
**Supplementary Figure 2.** Comparison of Logistic Model (log-scaled) parameters a, b, and c in the 1st and 2nd segments.

****

**Supplementary Figure 3**. Comparison of Gompertz Model (log-scaled) parameters a, b, and c in the 1st and 2nd segments.



**Supplementary Figure 4**. Comparison of Logistic and Gompertz Model (log-scaled) parameters a, b, and c in the 1st segment.



**Supplementary Figure 5**. Comparison of Logistic and Gompertz Model (log-scaled) parameters a, b, and c in the 2nd segment.