Which Factors Influence the Number of COVID-19 Cases and Deaths in the European Union?

Petronella Képes

PhD Student, Budapest University of Technology and Economics, Hungary, kepespetronella@kgt.bme.hu

ABSTRACT: The COVID-19 pandemic influenced the whole year of 2020, and it is still a determinative factor in everyday decisions. The schools and workplaces - where possible - moved to people's homes, changing the way humans used to operate in the last century. It affects almost every country on Earth, while this is only the first wave of the pandemic. Not even the researchers know when it will end, what the solution will be, or whether a vaccine will be found. The research focuses on this pandemic, limited to the European Union's territory, but also includes the following four countries: Iceland, Norway, Switzerland, and the United Kingdom. The countries in this region had different approaches and results regarding the coronavirus. This paper aims to define the most critical social and economic determinants (e.g., GDP, Healthcare expenditure, median age, Hofstede's dimensions), which have the highest impact on the number of confirmed COVID-19 Cases and Deaths. The paper highlights these factors by using the Ordinary Least Squares (OLS) regression model, creating the equations that define the number of cases and deaths.

KEYWORDS: COVID-19, OLS regression, European Union, Healthcare, Hofstede

Introduction

The World Health Organization was informed about the first case of “viral pneumonia” by Chinese authorities on 31st December 2019. This case was followed by significant growth in patient numbers in the upcoming days. On 9th January 2020 was reported that a novel coronavirus causes the outbreak, today called COVID-19 (WHO 2020a).

On 13th January, the first COVID-19 case was confirmed outside of China, in Thailand. There are 59 cases in China and 1 in Thailand.

On 14th January, WHO presented: there is a possibility for human-to-human transmission.

On 21st January, the first confirmed American coronavirus case was reported.

On 24th January, three confirmed cases were reported from France; all of the patients traveled from Wuhan. There are 897 cases: 882 in China, 2 in Japan, 3 in Singapore, 2 in South Korea, 1 in Taiwan, 4 in Thailand, 2 in Vietnam, and 1 in the US.

On 30th January, the WHO declared a public health emergency of international concern (PHEIC) because of the coronavirus, but still not recommended limits in international travel or trade. At this moment, there were 7,734 confirmed cases in China and 98 cases in 18 other countries, including 8 cases of human-to-human transmission in four countries (Germany, Japan, Vietnam, USA) (WHO 2020b).

On 11th March they categorized the COVID-19 as a pandemic. Till this day 118,319 people got infected (80,955 in China) and 4,292 people lost their lives (3,162 in China).

Until 1st July, there were 10,357,662 cases and 508,055 deaths because of the coronavirus, not to mention the deaths connected to the lockdowns and that the doctors stopped seeing patients.

In the analyzed countries, which include the European Union, Iceland, Norway, Switzerland, and the United Kingdom on 1st July, there were 1,639,858 cases and 178,804 deaths. The results imply a 10.9% death ratio, but the countries have different results in fighting the COVID-19. The research tries to find the important social and economic factors which can explain these differences (Our World in Data 2020a,b).

COVID-19 in Europe

The first case of COVID-19 in Europe was reported on 24th January. Since that date, the countries tried different approaches to defeat the virus. The following table (Table 1) presents the countries with the date of their first COVID-19 case, the date of the first national recommendations, and the localized or national lockdown date.

Table 1. Countries with their first cases and dates of applied governmental interventions

Country	First coronavirus case	First national recommendation	Localized lockdown applied	National lockdown applied
Austria	February 25	March 13	-	March 16
Belgium	February 4	March 12	-	March 17
Bulgaria	March 8	March 13	March 16	March 18
Croatia	February 25	-	-	March 19
Cyprus	March 9	March 15	-	March 24
Czechia	March 1	-	-	March 16
Denmark	February 27	March 13	-	March 18
Estonia	February 27	March 13	-	March 30
Finland	January 29	March 16	March 28	-
France	January 24	March 8	-	March 17
Germany	January 27	March 18	March 20	March 23
Greece	February 26	March 11	March 17	March 23
Hungary	March 4	March 11	-	-
Iceland	February 28	March 15	-	-
Ireland	February 28	March 12	-	March 28
Italy	January 31	-	February 21	March 12
Latvia	March 2	March 13	-	-
Lithuania	February 28	March 13	-	March 24
Luxembourg	February 29	-	-	March 15
Malta	March 7	March 11	March 28	-
Netherlands	February 27	March 15	-	March 23
Norway	February 26	March 12	March 16	-
Poland	March 4	March 12	-	March 24
Portugal	March 2	March 12	-	March 19
Romania	February 26	March 12	-	March 23

Slovakia	March 6	March 9	-	March 15
Slovenia	March 5	March 12	-	March 16
Spain	January 31	March 9	-	March 14
Sweden	January 31	March 11	-	-
Switzerland	February 25	-	-	March 13
United Kingdom	January 31	March 22	-	March 23

Source: Based on the BBC.com and devex.com data

Based on the most robust governmental intervention applied, three groups can be created. The first group contains 24 countries, which initiated national lockdowns. The second group belongs to 3 countries (Finland, Malta, Norway), where only localized lockdowns were applied. The countries in the last group (Hungary, Iceland, Latvia, Sweden) operated only with national recommendations. However, there are also differences in this group, because, for example, Hungary closed the schools and universities, while in Sweden, no such action was put into effect.

Materials and methods

In order to find influential factors, the method of OLS regression was used. Based on the lecture notes of Shalabh, a regression model presents the connection between a dependent and one or more independent variables. A regression model contains the following: dependent or study variable, a parameter called intercept term (which is constant), the error component (which presents the failure of the data to lie on a straight line and with this the difference between the true and observed value of the dependent variable) and the regression coefficients (which give the slope parameter of every independent variable.

$$y = \beta_0 + \beta_1 X + \varepsilon \quad (1)$$
$$Var(y) = \sigma^2 \quad (2)$$

Where y is the dependent variable, β_0 is the constant, β_1 is the regression coefficient, X is the independent variable, and ε represents the error. We assume that ε is independent and identically distributed random variable, with a constant variance σ^2.

When β_0, β_1, and σ^2 are known, the model is complete. The values can be determined after, for example, using the Ordinary Least Squares method.

The OLS method estimates the relationship between the dependent and independent variables by "minimizing the sum of the squares in the difference between the observed and predicted values of the dependent variable" (Encyclopedia.com, 2020).

The dependent variables are either the Confirmed COVID-19 Cases or the Confirmed COVID-19 Deaths in the later presented models. Based on the type of data, cross-sectional regression can be created, which means that the data collected, so the dependent and independent variables are associated with the same period or point in time. The variables' presented values were collected on 1st July 2020, and every result is based on the previously mentioned 31 country observations. The data is based on the following statistics: CIA Factbook data for each country (CIA Factbook 2020a - 2020ae), World Bank data for each country (World Bank n.d.), Eurostat (n.d.), Hofstede country comparison (Hofstede Insights n.d.), and the WHO (Global Health data repository 2020). During the analysis, the Gretl statistical software was used.

Data for the models

The following data was used for each country in order to create the OLS regressions:

- Economic factors
  - GDP per capita
  - Healthcare expenditures in GDP percent
  - Healthcare expenditures per capita
- Social factors
  - Population
  - Population density for km^2
  - Dependency ratio: It contrasts the ratio of youths and the elderly to the working-age group (ages 15-64). The increasing proportion of elderly and youths increases the value of the dependency ratio.
  - Median age
  - Sex ratio: the number of males to females
  - Hofstede's cultural dimensions
    - Power Distance: how the less powerful members accept, that the power is distributed unequally.
    - Individualism: the degree of interdependence between the members of society.
    - Masculinity: the motivation behind the actions is working to be the best.
    - Uncertainty avoidance: how threatened the citizens are by unknown situations.
    - Long term orientation: how the society links to its past.
    - Indulgence: how much people try to control their desires and impulses.
- Healthcare factors
  - Obesity in the percentage of the population
  - Curative care beds per capita
  - Physicians per capita
  - Hospital beds per capita
  - Number of confirmed COVID-19 Cases
  - Number of confirmed COVID-19 Deaths
  - Number of confirmed COVID-19 cases per capita
  - Number of confirmed COVID-19 deaths per capita
- Environmental factors
  - Air pollution: A higher level of pollution indicates increasing mortality from respiratory infections and diseases, lung cancer, and selected cardiovascular diseases.

Influencing factors for the number of cases

There were many independent variables collected, between the variables (e.g., GDP and size of the population), the problem of multicollinearity appeared, so two different models were created.

Both models try to answer the following question: What influences the number of COVID-19 cases in Europe?

The first model contains the population's size and does not contain the previously happened COVID-19 deaths; the second model is the reverse of the first model.

The model containing the size of the population

After excluding the insignificant factors, Table 2 contains the result.

Table 2. Model of the cases with the population

Variable	Coefficient	Std. error	T-ratio	P-value	
const.	-50580.7	51209.2	-0.9877	0.3324	
Population	0.00358873	0.000227249	15.79	7.68e-15	***
Obesity	429465	211136	2.034	0.0523	*
Hospital bed per capita	-2.37397e+07	3.98930e+06	-5.951	2.80e-06	***
Long term orientation	1102.89	367.937	2.997	0.0059	***

Source: Gretl output

In the form of an equation:

*Number of COVID-19 Cases = -5.06e+04 + 0.00359*Population + 4.29e+05*Obesity - 2.37e+07*Hospital bed per capita + 1.10e+03*Long term orientation* (3)

From the results, the following statements can be determined. The number of cases will increase if the population, the ratio of obesity in the given population, or the long-term orientation level increases. The hospital beds per capita decreases the number of COVID-19 cases.

The R^2 = 0.9148 and the adjusted R^2 = 0.9017, so the model is suitable for global forecasting.

The value of the Akaike information criterion decreased during the exclusion of variables. Based on the Variance Inflation Factors (VIF) results, there is no collinearity problem present. The White-test statistic results are 14.349, with p-value 0.4240, which means there is no heteroskedasticity present in the model.

The model containing the number of COVID-19 deaths

After excluding the insignificant factors, Table 3 contains the result.

Table 3. Model of the cases with the number of COVID-19 deaths

Variable	Coefficient	Std. error	T-ratio	P-value	
const.	-453804	216860	-2.093	0.0467	**
Confirmed COVID-19 Deaths	7.14452	0.427655	16.71	4.49e-15	***
Dependency ratio	−2634.60	1245.07	−2.116	0.0445	**
Median age	9907.92	2613.28	3.791	0.0008	***
Sex ratio	239101	132100	1.810	0.0823	*
Air pollution	−3461.69	1369.25	−2.528	0.0182	**

Source: Gretl output

In the form of an equation:

*Number of COVID-19 Cases = -4.54e+05 + 7.14*Confirmed COVID-19 Deaths - 2.63e+03*Dependency ratio + 9.91e+03*Median age + 2.39e+05*Sex ratio - 3.46e+03*Air pollution* (4)

Based on these results, the previously happened COVID-19 Deaths, median age, and sex ratio increases the number of cases, while the dependency ratio and air pollution decreases. Interestingly, air pollution decreases the number of cases, but for example, people wear masks often in the highly polluted countries, that could be a reason behind the negative correlation between the number of cases and this variable.

The R^2 = 0.9304 and the adjusted R^2 = 0.9165, so the model is suitable for global forecasting.

The value of the Akaike information criterion decreased during the exclusion of variables. Based on the results of the VIF, there is no collinearity problem present. The White-test statistic results are 29.656, with p-value 0.0756, which means there is no heteroskedasticity present in the model.

Influencing factors for the number of deaths

The question to be answered: Which factors influence the number of COVID-19 deaths in the European countries?

After excluding the insignificant factors, Table 4 contains the result.

Table 4. Model of the deaths

Variable	Coefficient	Std. error	T-ratio	P-value	
const.	−250133	84476.4	−2.961	0.0072	***
Individualism	220.446	110.264	1.999	0.0581	*
Masculinity	207.783	77.0948	2.695	0.0132	**
Indulgence	−406.840	153.339	−2.653	0.0145	**
Dependency ratio	1078.83	458.355	2.354	0.0279	**
Median age	1550.94	882.970	1.756	0.0929	*
Sex ratio	140247	62082.4	2.259	0.0341	**
Curative carebeds per capita	−3.84786e+06	1.80151e+06	−2.136	0.0441	**
Number of cases per capita	2.60144e+06	932876	2.789	0.0107	**

Source: Gretl output

In the form of an equation:

*Number of COVID-19 Deaths = -2.50e+05 + 220*Individualism + 208*Masculinity - 407*Indulgence + 1.08e+03*Dependency ratio + 1.55e+03*Median Age + 1.40e+05*Sex ratio - 3.85e+06*Curative care beds per capita + 2.60e+06*Number of cases per capita* (6)

As can be seen, the level of individualism and masculinity, dependency ratio, median age, sex ratio, and the number of COVID-19 cases per capita increases the number of deaths, while the level of indulgence and the number of curative care beds per capita decreases the COVID-19 deaths.

The $R^2 = 0.6282$ and the adjusted $R^2 = 0.4930$, which is not as high as in the previous models because it does not contain the GDP or the population size

The value of the Akaike information criterion decreased during the exclusion of variables. Based on the results of the VIF, there is no collinearity problem present. The White-test statistic results are 19.407, with p-value 0.2481, which means there is no heteroskedasticity present in the model.

Conclusion

In order to find the previously presented models, also many others were created. There is a common thing among all of the models, which was very interesting: the number of physicians or the number of physicians per capita never became significant during the calculations.

Another interesting fact is that social factors were more significant than any other type of factor in calculating the number of deaths. Based on the news, it is not a big surprise because people's attitude determines the level of social distancing and the level of wearing a mask.

As expected, based on the knowledge the physicians collected, the higher the median age or, the fewer hospital beds the countries have, the higher the number of COVID-19 cases and deaths.

Discussion

Based on the collected data, the following research can be completed, which considers the delay between the appearance of the first COVID-19 case and the governmental action. Besides, more broad research can be created involving the data of more countries on other continents too.

References

BBC.com April 7, 2020. "Coronavirus: The world in lockdown in maps and charts." Accessed July 26, 2020. https://www.bbc.com/news/world-52103747.

CIA Factbook 2020a. "Austria." Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/au.html.

CIA Factbook 2020b. "Belgium." Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/be.html.

CIA Factbook 2020c. "Bulgaria." Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/bu.html.

CIA Factbook 2020d. "Croatia." Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/hr.html.

CIA Factbook 2020e. "Cyprus." Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/cy.html.

CIA Factbook 2020f. "Czechia." Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/ez.html.

CIA Factbook 2020g. "Denmark." Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/da.html.

CIA Factbook 2020h. "Estonia." Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/en.html.

CIA Factbook 2020i. "Finland." Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/fi.html.

CIA Factbook 2020j. "France." Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/fr.html.

CIA Factbook 2020k. "Germany." Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/gm.html.

CIA Factbook 2020l. "Greece." Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/gr.html.

CIA Factbook 2020m. “Hungary.” Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/hu.html.

CIA Factbook 2020n. “Iceland.” Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/ic.html.

CIA Factbook 2020o. “Ireland.” Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/ei.html.

CIA Factbook 2020p. “Italy.” Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/it.html.

CIA Factbook 2020q. “Latvia.” Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/lg.html.

CIA Factbook 2020r. “Lithuania.” Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/lh.html.

CIA Factbook 2020s. “Luxembourg.” Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/lu.html.

CIA Factbook 2020t. “Malta.” Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/mt.html.

CIA Factbook 2020u. “Netherlands.” Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/nl.html.

CIA Factbook 2020v. “Norway.” Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/no.html.

CIA Factbook 2020w. “Poland.” Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/pl.html.

CIA Factbook 2020x. “Portugal.” Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/po.html

CIA Factbook 2020y. “Romania.” Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/ro.html

CIA Factbook 2020z. “Slovakia.” Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/lo.html

CIA Factbook 2020aa. “Slovenia.” Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/si.html

CIA Factbook 2020ab. “Spain.” Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/sp.html

CIA Factbook 2020ac. “Sweden.” Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/sw.html.

CIA Factbook 2020ad. “Switzerland.” Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/sz.html.

CIA Factbook 2020ae. “United Kingdom.” Accessed July 26, 2020. https://www.cia.gov/library/publications/the-world-factbook/geos/uk.html.

Devex.com 2020. “COVID-19 — a timeline of the coronavirus outbreak.” Accessed July 26, 2020. https://www.devex.com/news/covid-19-a-timeline-of-the-coronavirus-outbreak-96396

Encyclopedia.com 2020. “Ordinary Least Squares Regression.” Accessed July 26, 2020. https://www.encyclopedia.com/social-sciences/applied-and-social-sciences-magazines/ordinary-least-squares-regression.

Eurostat. n.d. “Curative care beds in hospitals.” Accessed July 26, 2020. https://ec.europa.eu/eurostat/tgm/table.do?tab=table&plugin=1&language=en&pcode=tps00168.

Eurostat. n.d. “Healthcare expenditure statistics.” Accessed July 26, 2020. https://ec.europa.eu/eurostat/statistics-explained/index.php/Healthcare_expenditure_statistics.

Hofstede Insights. n. d. “Compare countries.” Accessed July 26, 2020. https://www.hofstede-insights.com/product/compare-countries/.

Our World in Data. 2020a. “Coronavirus (COVID-19) Cases.” Accessed July 26, 2020. https://ourworldindata.org/covid-cases?country=AUT~BEL~BGR~HRV~CYP~CZE~DNK~EST~FIN~FRA~DEU~GRC~HUN~ISL~IRL~ITA~LVA~LTU~LUX~MLT~NLD~NOR~POL~PRT~ROU~SVK~SVN~ESP~SWE~CHE~GBR

Our World in Data. 2020b. “Coronavirus (COVID-19) Deaths.” Accessed July 26, 2020. https://ourworldindata.org/covid-deaths?country=AUT~BEL~BGR~HRV~CYP~CZE~DNK~EST~FIN~FRA~DEU~GRC~HUN~ISL~IRL~ITA~LVA~LTU~LUX~MLT~NLD~NOR~POL~PRT~ROU~SVK~SVN~ESP~SWE~CHE~GBR.

Shalabh n.d. “Econometrics, Chapter 2.” Online teaching material. Accessed July 26, 2020. http://home.iitk.ac.in/~shalab/econometrics/Chapter2-Econometrics-SimpleLinearRegressionAnalysis.pdf.

World Bank. n.d. “GDP (current US$).” Accessed July 26, 2020. https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=EU.

World Health Organization. June 30, 2020a. “Timeline of WHO’s response to COVID-19.” Accessed July 26, 2020. https://www.who.int/news-room/detail/29-06-2020-covidtimeline.
World Health Organization. January 30, 2020b. “WHO Director-General's statement on IHR Emergency Committee on Novel Coronavirus (2019-nCoV).” Accessed July 26, 2020. https://www.who.int/dg/speeches/detail/who-director-general-s-statement-on-ihr-emergency-committee-on-novel-coronavirus-(2019-ncov).
World Health Organization - Global Health data repository. April 4, 2020. “Exposure country average.” Accessed July 26, 2020. https://apps.who.int/gho/data/node.main.152?lang=en.