

# A Brief Analysis of COVID-19 Confirmed Cases for Eight Countries

Selim Temizer<sup>1</sup>

April 4, 2020

## Contents

1. Software and Data .....	1
2. Countries .....	1
3. Analysis.....	1
3.1 Event A .....	2
3.2 Event B .....	3
3.3 Event C .....	7
4. Discussion .....	10
Disclaimer.....	11
Appendix .....	11

## 1. Software and Data

All figures and tables in this document were generated using *COVID-19 Status*<sup>2</sup> and *Microsoft Excel* applications. The data that was used for the analyses contained confirmed case counts from January 22, 2020 to April 3, 2020, inclusive. Excel data and analyses carried on using Excel are available as an **Appendix** to this document in electronic format (as an Excel file).

## 2. Countries

The following countries are focused on in this document:

- Countries with number of confirmed cases above 50,000 as of April 3, 2020 (excluding China, which experienced the phases and events analyzed here earlier than other countries). These countries are *United States* (US), *Italy*, *Spain*, *Germany*, *France* and *Iran*.
- *Turkey*, which I am a citizen of.
- *Kazakhstan*, where I currently work.

## 3. Analysis

Basically, 3 events will be analyzed:

- *Event A*: Number of confirmed cases as of April 3, 2020.
- *Event B*: The period after the 100<sup>th</sup> case has been observed in each country.
- *Event C*: The period after a population-normalized metric (which is the period after the 1<sup>st</sup> case per one million people) has been observed in each country.

The events for each country are listed in **Table 1**.

---

<sup>1</sup> <http://selimtemizer.com>

<sup>2</sup> <http://selimtemizer.com/software/covid19>

**Table 1.** Events.

	<b>A: Cases</b>	<b>B: Cases <math>\geq</math> 100</b>		<b>C: Cases <math>\geq</math> 1 per Million People</b>	
	<b>Count</b>	<b>Date</b>	<b>Days since B*</b>	<b>Date</b>	<b>Days since C*</b>
<b>US</b>	275586	03/03/2020	32	03/07/2020	28
<b>Italy</b>	119827	02/23/2020	41	02/22/2020	42
<b>Spain</b>	119199	03/02/2020	33	03/01/2020	34
<b>Germany</b>	91159	03/01/2020	34	03/01/2020	34
<b>France</b>	65202	02/29/2020	35	02/29/2020	35
<b>Iran</b>	53183	02/26/2020	38	02/25/2020	39
<b>Turkey</b>	20921	03/19/2020	16	03/18/2020	17
<b>Kazakhstan</b>	464	03/26/2020	9	03/17/2020	18

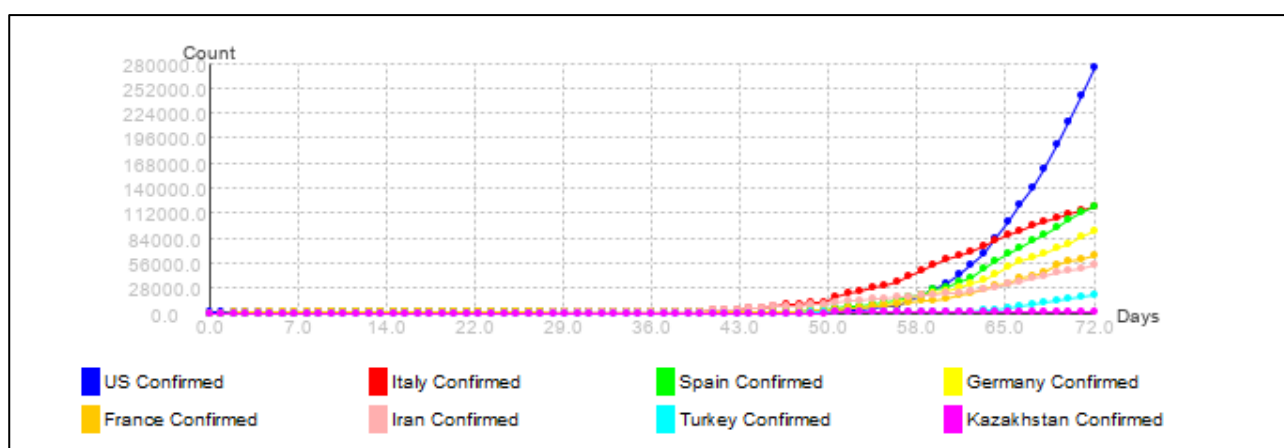
\*: Days since the event until April 3, 2020, inclusive.

It can be noticed that the first six countries at the top of the table have been dealing with large numbers of confirmed cases for events B and C for at least 10 more days than Turkey and Kazakhstan. Therefore, it might be necessary to carry out the analyses in this document again later, so that all countries could be compared more fairly in light of more data.

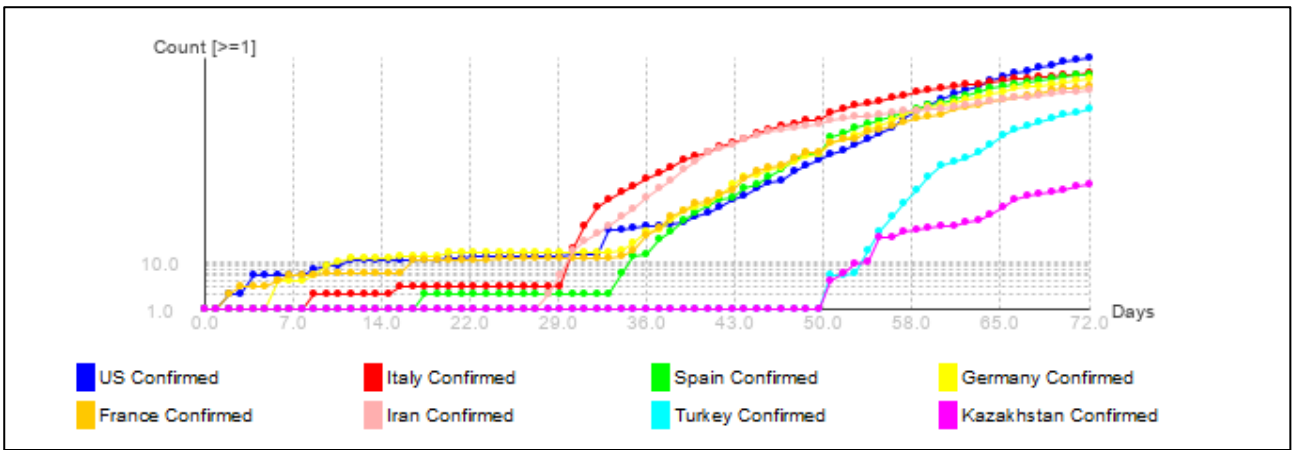
### 3.1 Event A

The number of confirmed cases, starting on January 22, 2020 up until April 3, 2020 are shown in **Figure A1** and **Figure A2** (logarithmic scale).

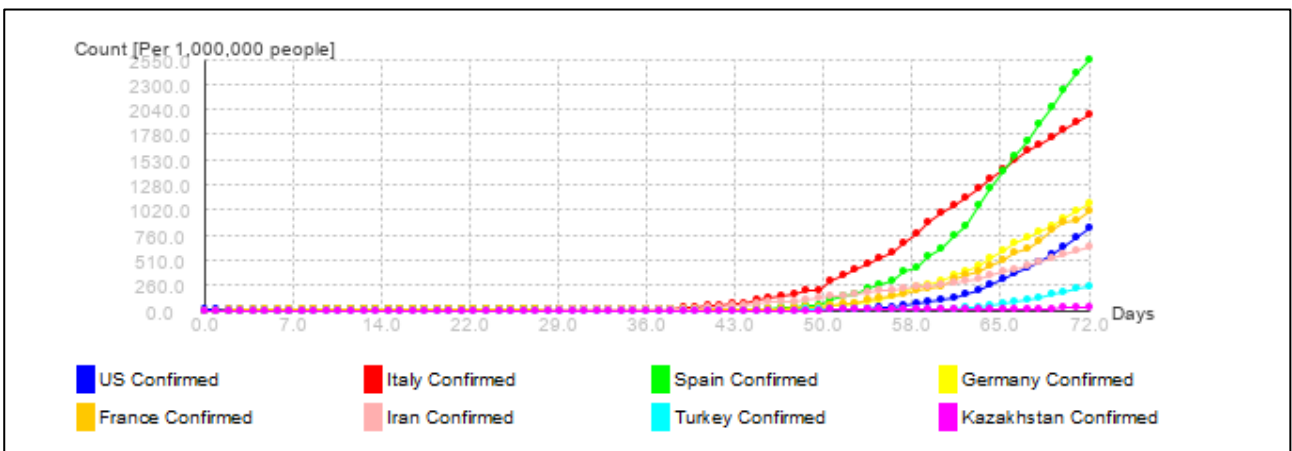
If we take the population sizes of each country into account and try to normalize the time series by looking at *number of confirmed cases per 1 million people*, we get the plots in **Figure A3** and **Figure A4** (logarithmic scale).



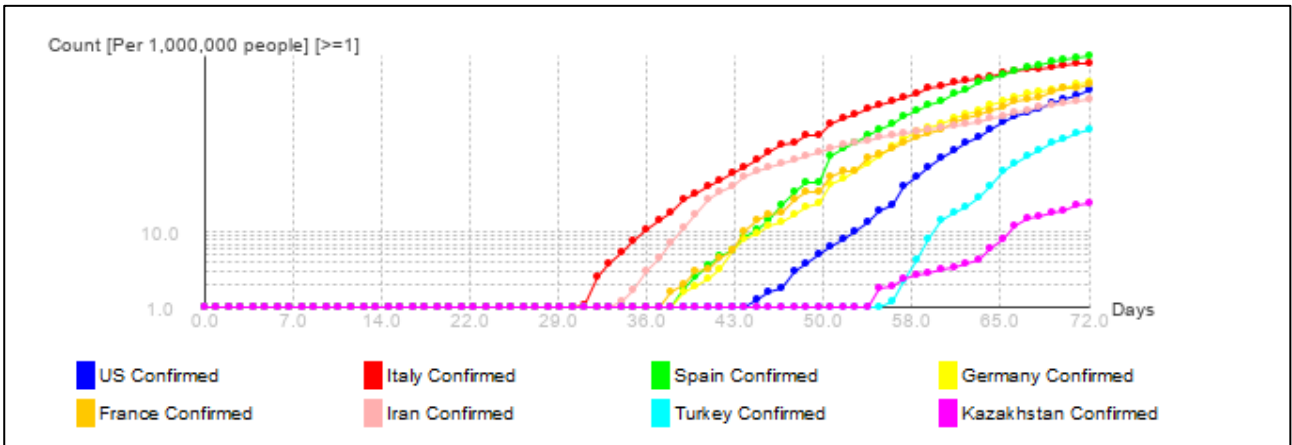
**Figure A1.** Confirmed cases from January 22, 2020 to April 3, 2020, inclusive.



**Figure A2.** Confirmed cases from January 22, 2020 to April 3, 2020, inclusive (logarithmic scale).



**Figure A3.** Confirmed cases per 1 million people.

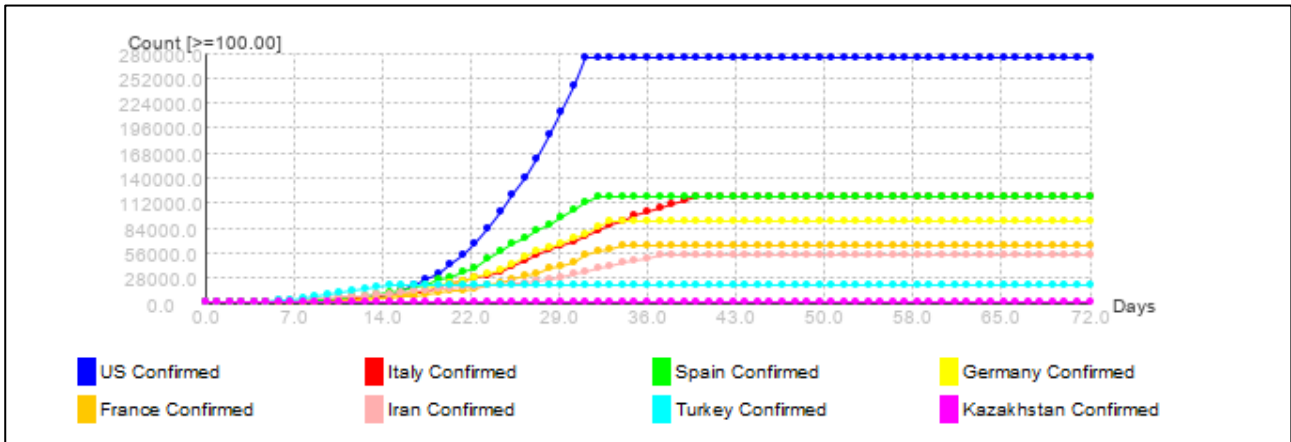


**Figure A4.** Confirmed cases per 1 million people (logarithmic scale).

### 3.2 Event B

The spread of the virus in different countries did not start all at the same time. Some countries started observing cases earlier than others. In order to have a more similar and common starting point for all eight countries (for a fairer comparison), we drop the days with less than 100 confirmed cases from the time series in **Figure A1**. The result is shown in **Figure B1**. Note that when we shift the time

series towards left, some entries on the right become empty, and in **Figure B1**, the empty entries are just filled with a duplicate of the most recent value (this is why all time series end with a horizontal section).



**Figure B1.** The period after the 100<sup>th</sup> case has been observed.

For each country, if we fit an *exponential growth curve* in the form  $y = C_0(1 + r)^n$  where

- $C_0$  is the initial number of cases (after the 100<sup>th</sup> case),
- $r$  is the growth rate (that we would like to find out), and
- $n$  is the number of days since Event B,

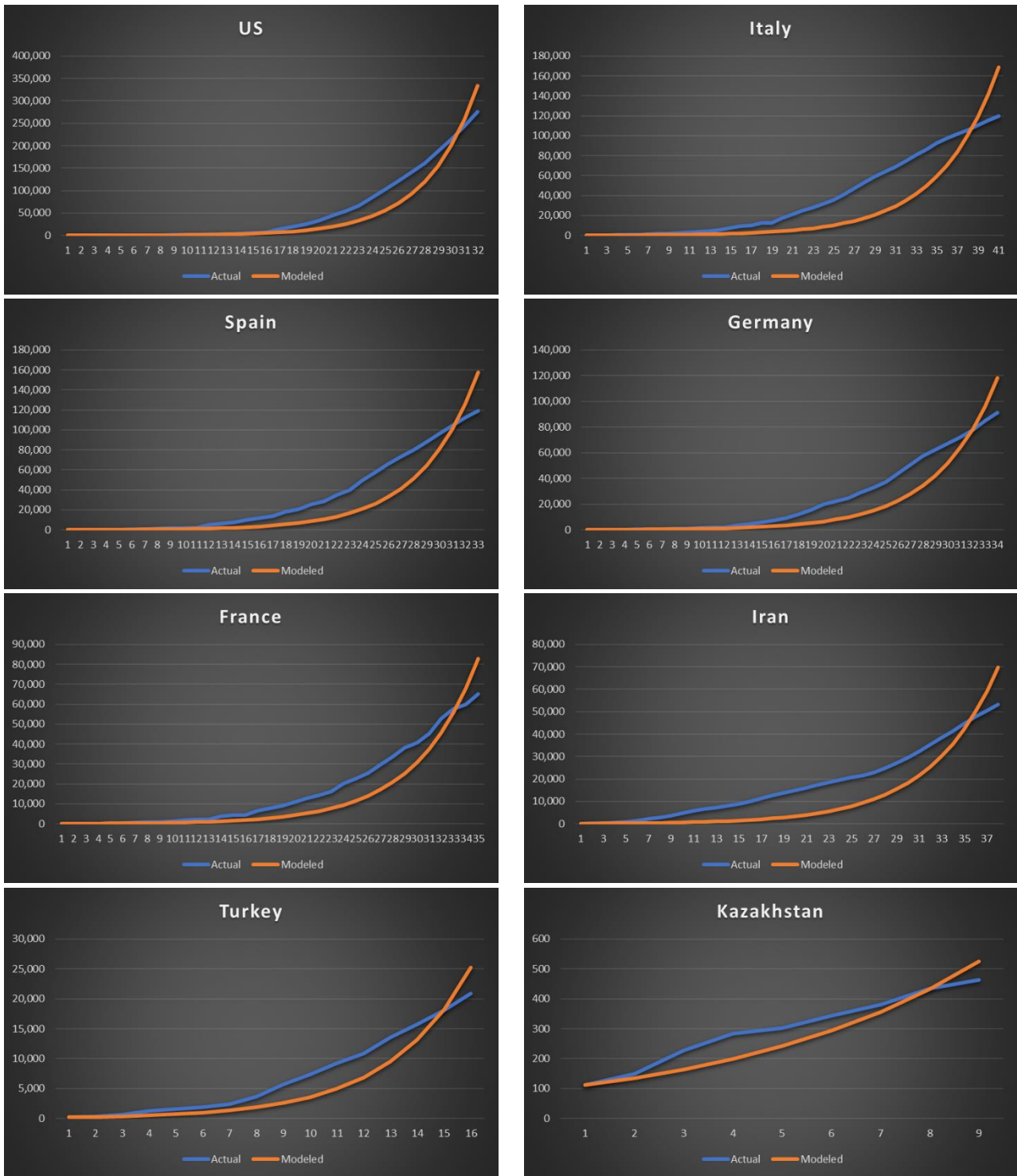
we get the growth rate values shown in **Table B1**.

**Table B1.** Modeled growth rates after 100<sup>th</sup> case.

	$C_0$	$r$
<b>US</b>	118	0.2923
<b>Italy</b>	155	0.1910
<b>Spain</b>	120	0.2516
<b>Germany</b>	130	0.2293
<b>France</b>	100	0.2185
<b>Iran</b>	139	0.1831
<b>Turkey</b>	192	0.3842
<b>Kazakhstan</b>	111	0.2147

The modeled exponential growth curves are displayed in **Figure B2**.

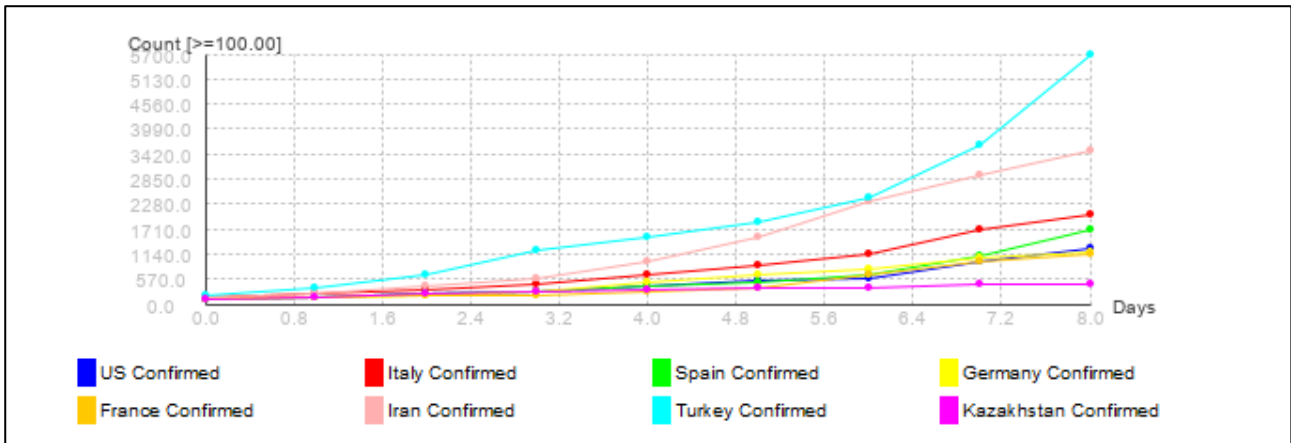
A rough interpretation of the growth rate is that for every single infection case, there were  $1 + r$  cases the next day (or we can say that a single infected person spread the virus to  $r$  other people every day). It should be noted that the modeled curves do not follow the actual data very well, therefore the modeled growth rates may not mean much and may not be very useful, except maybe for just pairwise comparisons.



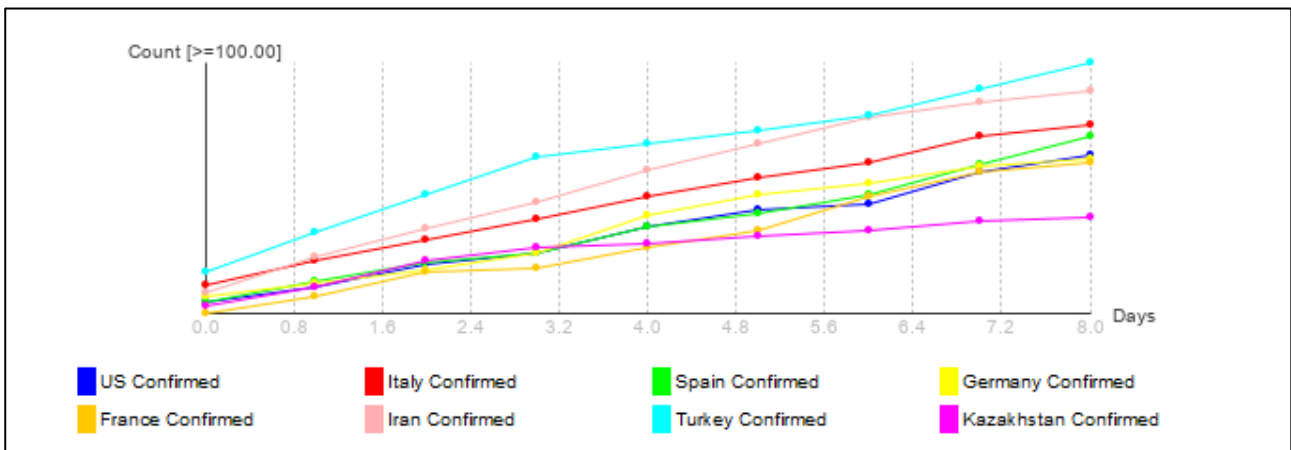
**Figure B2.** Modeled exponential growth curves.

As indicated in **Table 1**, United States observed Event B on March 3, 2020. Therefore, the left-shifted plot for US starts with that date in **Figure B1** (and has 32 *count* values indexed from 0 to 31, the remaining values in the horizontal section are simply the 32<sup>nd</sup> *count* value replicated). Italy has 41 *count* values in its new left-shifted time series, Spain has 33, and so on. Of all eight countries, Kazakhstan has the least number of *count* values (just 9), since it observed Event B much later than the other seven countries.

Now, let's zoom in on the first 9 days (for which there is valid data for all eight countries) and look at counts for each country in that 9-day period which starts with the 100<sup>th</sup> cases. **Figure B3** and **Figure B4** (logarithmic scale) show the counts for the first 9 days after Event B.

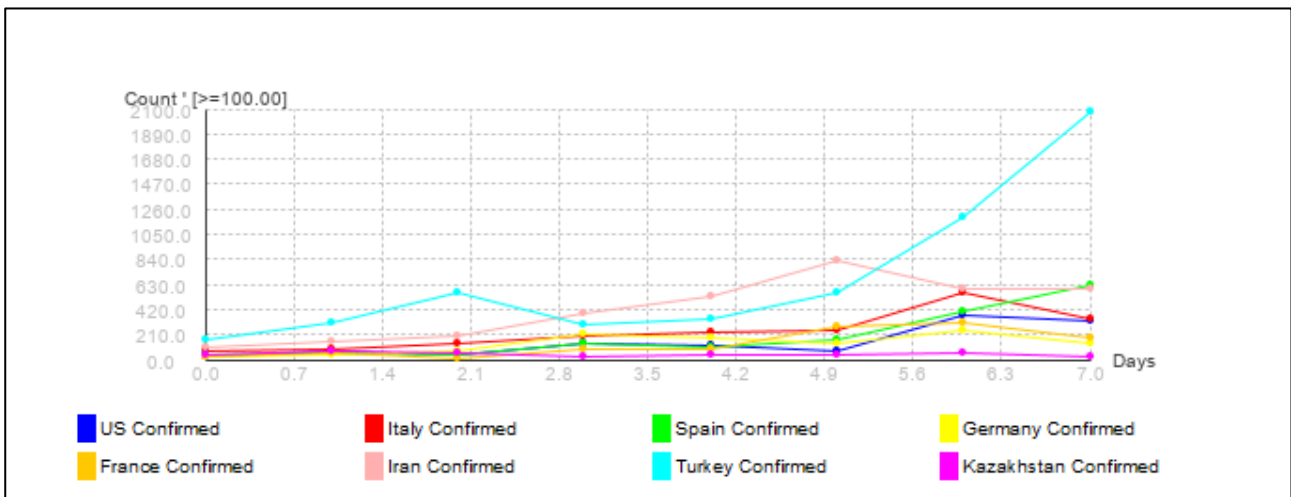


**Figure B3.** The counts for the first 9 days after Event B.



**Figure B4.** The counts for the first 9 days after Event B (logarithmic scale).

We can differentiate the time series to get the rate of change of counts and this *speed of the spread* plot is displayed in **Figure B5**.



**Figure B5.** Speed of the spread for the first 8 days after Event B.

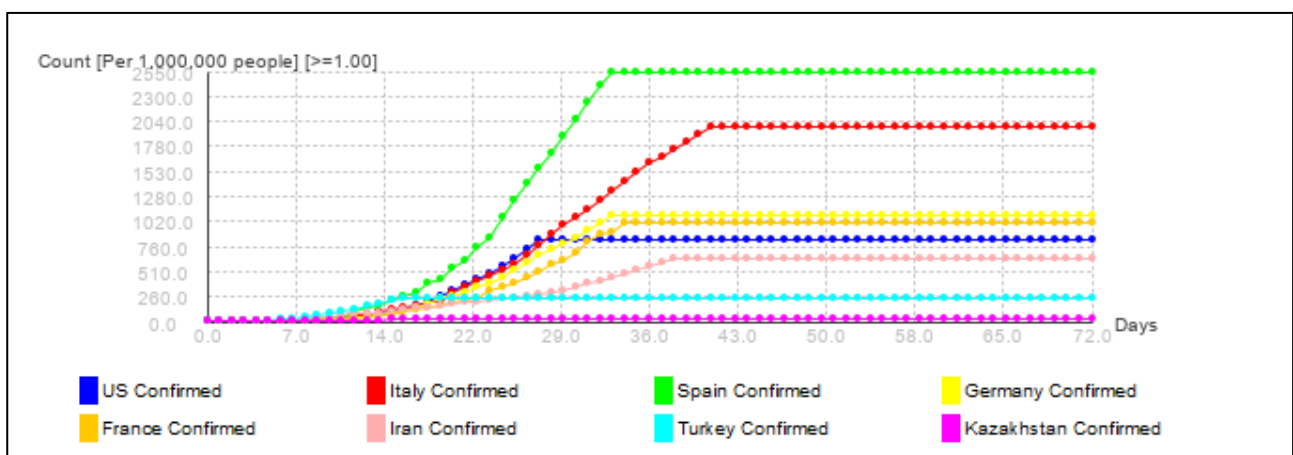
When interpreting the plots, the following points should be taken into consideration:

- Countries have different *population sizes*. It is normal to see larger counts in more crowded countries.
- Countries have different land area sizes (hence; *population densities*). Suppose that two countries, A and B, have the same population, but country A is much larger than country B in terms of land area. In this case, country A will have a lower population density (in terms of *people per square kilometer*) than country B, and it will, in principle, be advantageous in terms of lower spread rate, and possibly lower counts.
- Countries have different *population distributions*. Suppose that two countries, A and B, have the same population and the same land area sizes (hence, same population densities). However, citizens of country A are living all spread out uniformly in the country, but citizens of country B are all clumped up in just a handful of densely packed cities. In this case, country B will have the disadvantage that once cases start being seen in a city, it is likely that it will spread out in that city much quicker compared to how it would spread out in country A.
- Countries may have administered different *number of tests* to detect cases.

We can say that population distributions are never uniform, every country has different population densities for each of its cities, regions, etc., and when preparing this document, we did not have data on the number of tests administered. Therefore, normalization along those dimensions doesn't seem to be easy. However, we can try to normalize the results presented so far across *population sizes* to get a different perspective. This is what we will do in analyzing Event C in the following section.

### 3.3 Event C

Similar to Event B analysis, here we choose a common similar starting point for all eight time series, and this common point is *observing at least 1 case per 1 million people* (since the raw counts are normalized per 1 million people, the cases are now real numbers rather than integral values). This is equivalent to taking the time series in **Figure A3** and trimming them from the left until the starting values are greater than or equal to 1. The resulting plot is shown in **Figure C1**.

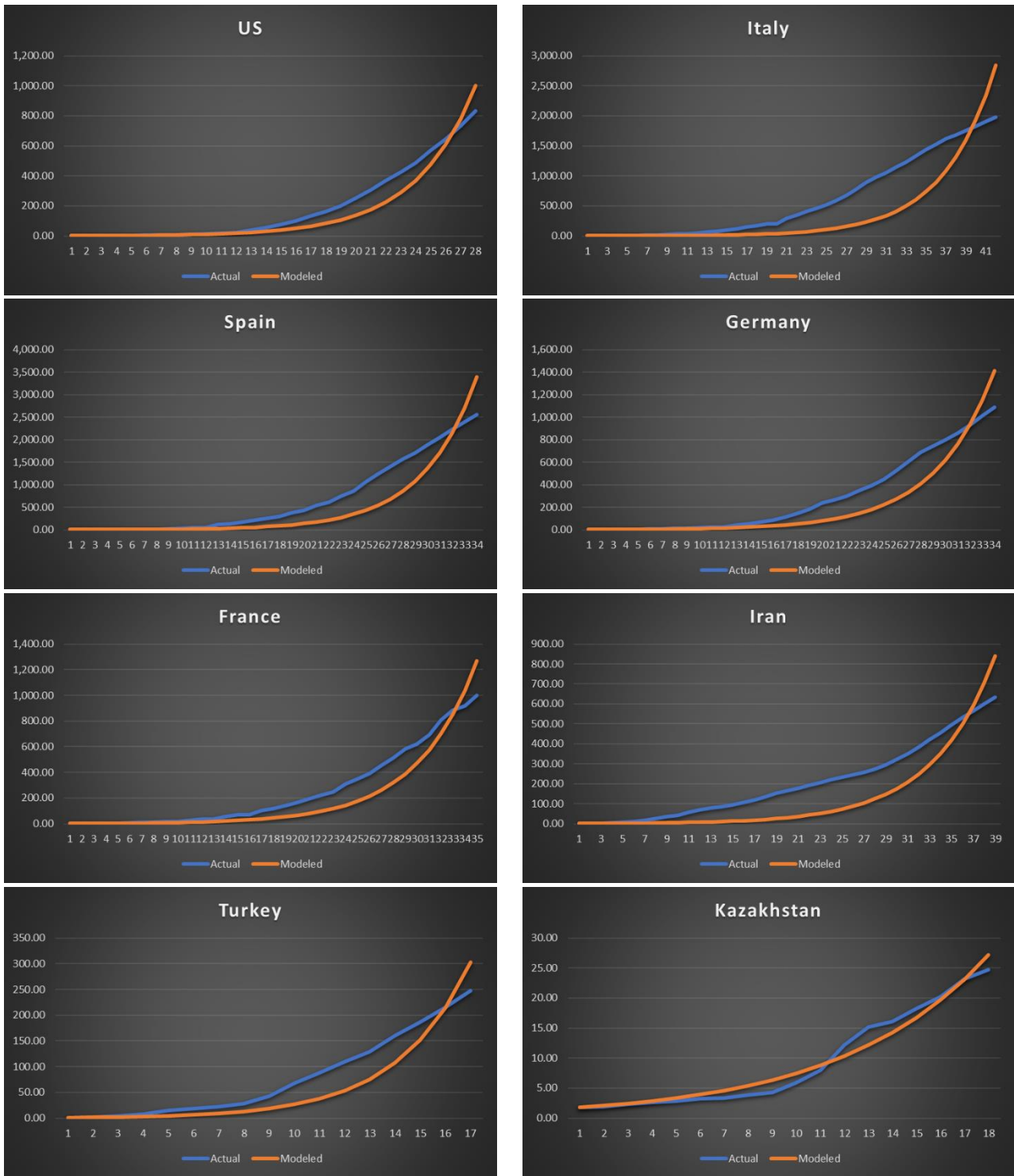


**Figure C1.** The period after the 1<sup>st</sup> case per 1 million people has been observed.

If we again fit exponential growth curves of the same form,  $y = C_0(1 + r)^n$ , as we did for before, we get the modeled growth rates in **Table C1**. The modeled curves are shown in **Figure C2**.

**Table C1.** Modeled growth rates after 1<sup>st</sup> case per 1 million people.

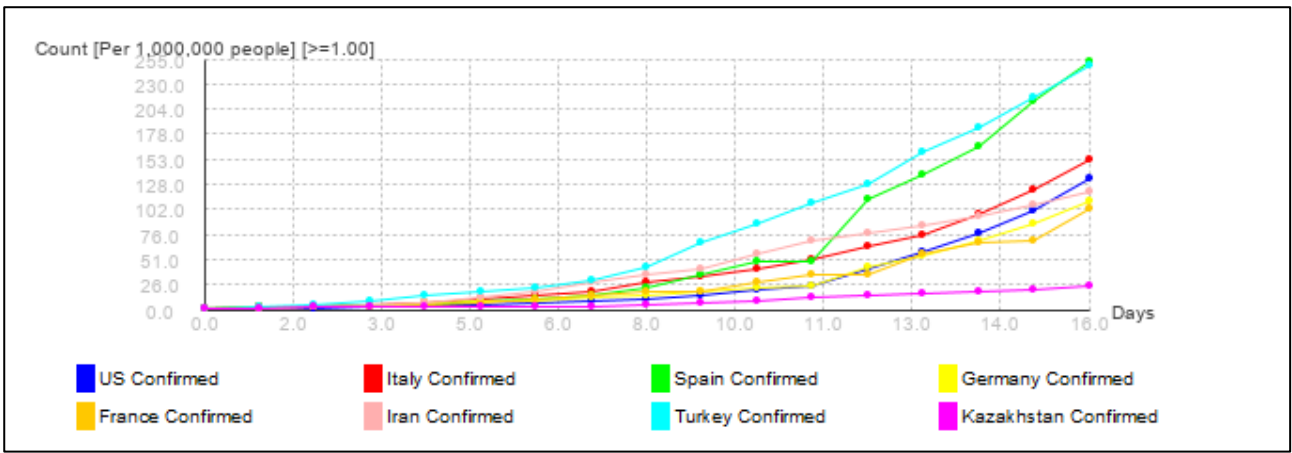
	$C_0$	$r$
<b>US</b>	1.21	0.2823
<b>Italy</b>	1.03	0.2133
<b>Spain</b>	1.80	0.2568
<b>Germany</b>	1.55	0.2293
<b>France</b>	1.53	0.2185
<b>Iran</b>	1.13	0.1900
<b>Turkey</b>	1.16	0.4159
<b>Kazakhstan</b>	1.76	0.1747



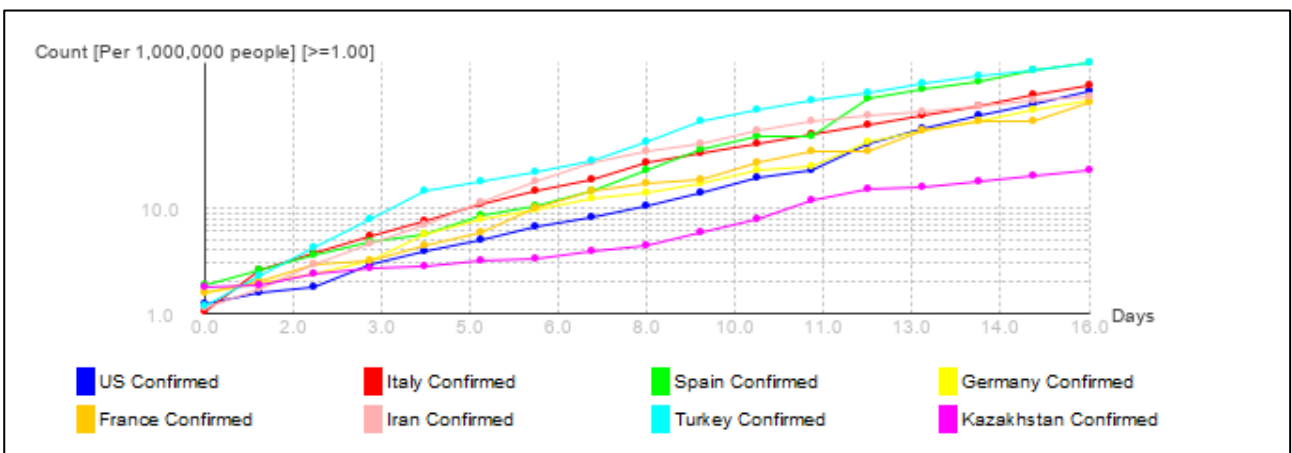
**Figure C2.** Modeled exponential growth curves.

Again, we observe that the modeled curves do not follow the actual data very well, and the modeled growth rates may not mean much and may not be very useful, except maybe for just pairwise comparisons.

Now, let's zoom in on the first 17 days of Event C (for which there is valid data for all eight countries). **Figure C3** and **Figure C4** (logarithmic scale) illustrate the situation.

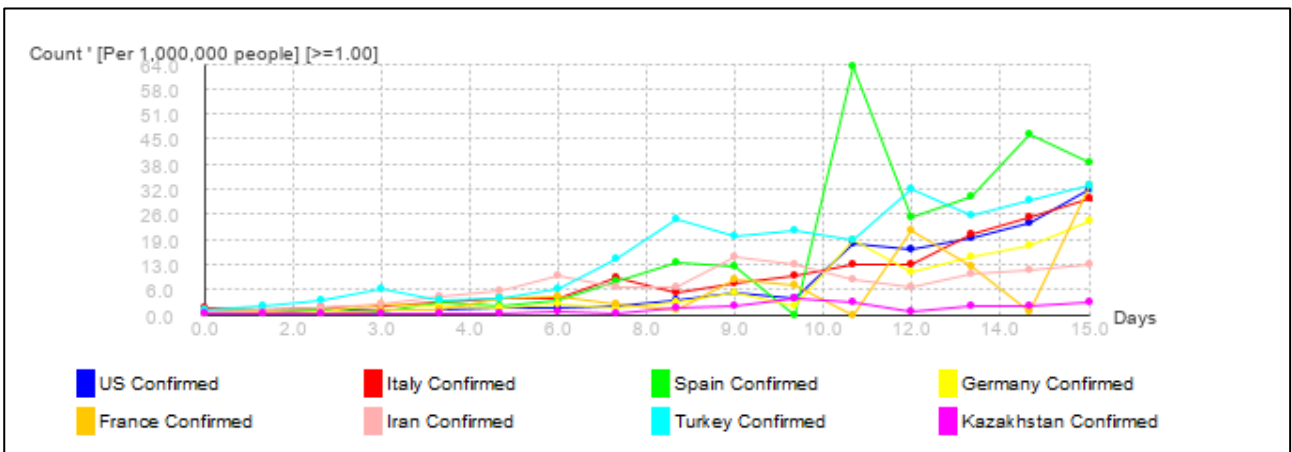


**Figure C3.** The counts for the first 17 days after Event C.



**Figure C4.** The counts for the first 17 days after Event C (logarithmic scale).

The rate of change of counts (*speed of the spread*) during the same period is shown in **Figure C5**.



**Figure C5.** Speed of the spread for the first 16 days after Event C.

#### 4. Discussion

The amount of data that the analyses in this document are based on is limited in quantity, and each country is continuously taking various measures against the pandemic, which is likely to have an

effect on changing the plots and numbers reported in this document.

It would be interesting and potentially useful to carry out the same analyses in this document using the recovery and death statistics.

### **Disclaimer**

The author of this document is a computer scientist, and he has absolutely no background in medicine. This analysis was prepared in the hope that it might be useful, but without any direct or implied warranty or claim of accuracy, reliability, or fitness for a particular purpose.

### **Appendix**

This document is accompanied by an appendix in electronic format (as an Excel file). If you haven't received it, you may download it here: <http://selimtemizer.com/software/covid19>