**COVID-19- Italy, Germany and Turkey Data Analysis   
and New Case Prediction Model for Turkey**

Dr. İlker Met, Ayfer Erkoç, Himmet Aksoy, Hacer Özge Özer

**Summary**

Although it is first seen in China, more than 4.5 months, Covid-19 infection affected nearly whole countries negatively from economical to the psychological dimension. Apart from economical and psychological effects, workforce planning was also an important problem for the organizations. Due to social distancing and quarantine processes; organizations couldn’t find the necessary staff for their needs. In moments of uncertainty and concern, it was not only about what leaders of organizations did, but how they reacted was also important. There was understandable anxiety among everyone as Covid-19 has been spreaded across the world. Governments and health organizations began to prioritize public health, companies used various strategies to help reduce the health risk of their employees and their customers. But main question was; how long this uncertainty will continue?

In this study, we tried to show an analytical approach, which we used in our organization, can be used to estimate the uncertainty. This model can be used for taking necessary precautions at the right time in such extraordinary situations. We used the polynomial quadratic model and the SIR model for estimation. Our findings show that these models can be used for estimating the end of the infection process. We hope that human being will never see such a disease again, but if it happens the model might help the organizations to grasp the uncertainty.

**Key words:** Moving Averages, SIR Model, Polynomial Model, Modeling of Disease; Cubic Curve, Work Force Planning, Uncertainty Estimation.

1. **Introduction**

First in China (Hubei-Wuhan) in mid-November 2019, an infection began with suspicious cases. It has been identified the first half of December as Covid-19 (SARS-CoV2 Infection). In the world, more than 4.5 months, infection actively affected all the aspects from economical to psychological dimensions negatively. Since January, the epidemic manifested itself in three stages: First, a China-based supply shock affected global trade, causing disruption in supply chains, then financial markets began to unravel when investors realized that recession was inevitable, and now; China, Europe and the USA are facing a serious demand shock that negatively affects both consumption and investment expenditures [1].

Most of firms have been giving full time service to their customers under those uncertain conditions. That uncertainty not only has increased work burden on firms, it has also decreased the motivation on employees. Under those circumstances firms should take precautions for the health of employees and customers while they should provide the sustainability of work. So planning the number employee on leave and on work has become very critical.

Main difficulty was estimating the future. How long this uncertainty will continue was the main question. In this study we built an analytical model to grasp uncertainty.

1. **Modelling**

For the diseases, it is not always easy to build a model. Because parameters like hospital capacity, the age distribution of the population, elderly population rate, and population with chronic disease and the number of health personnel are all different from each countries. However, when we look at “daily new case” data, it can be easily seen the rate of disease spread exponentially and particularly trends show more or less similar spreading rate in countries.

For these reasons, the spread and recovery times of the disease have been modelled by using the daily number of new cases in Italy, Germany and Turkey countries which are similar in terms of populations.

In this paper, Moving Averages, SIR Model, Quadratic Polynomial Model are used and an estimation model for new cases prediction is formed.

* 1. **Moving Averages**

A moving average is defined as an average of fixed number of items in the time series which move through the series by dropping the top items of the previous averaged group and adding the next in each successive average [2]. The period selected depends on the type of movement of interest, such as short, intermediate, or long-term. Countries should flattening the moving averages curve to talk about the down trend.

**Figure 1. Three Days Moving Averages for Italy, Germany and Turkey\***

|  |  |  |
| --- | --- | --- |
| **Italy** | **Germany** | **Turkey** |
|  |  |  |

\*The data for each countries include from the date of first case seen to 13 April 2020 [3].

As seen from Figure 1. As of 13 April 2020, Italy and Germany are started the flattening curve trend, Turkey is about to first peak. Weekly moving averages in Germany and Italy started to decrease the trend relative to the previous week. It is possible to mention that the decreasing trend has started in both countries unless the trend deteriorates.

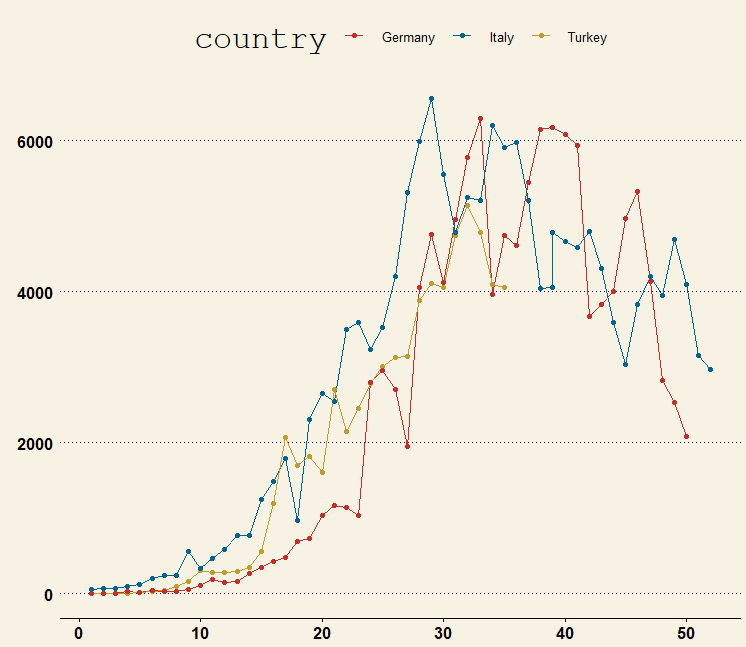
**Table 1. 7 Days Moving Averages Table for Italy, Germany and Turkey**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Date** | **Italy** | | **Germany** | | **Turkey** | |
| **New Case** | **Roll\_Mean** | **New Case** | **Roll\_Mean** | **New Case** | **Roll\_Mean** |
| 27.03.2020 | 5.909 | 5.401 | 5.780 | 4.946 | 2.069 | 1.328 |
| 28.03.2020 | 5.974 | 5.230 | 6.294 | 4.925 | 1.704 | 1.665 |
| 29.03.2020 | 5.217 | 5.169 | 3.965 | 5.116 | 1.815 | 1.892 |
| 30.03.2020 | 4.050 | 4.950 | 4.751 | 5.287 | 1.610 | 2.072 |
| 31.03.2020 | 4.053 | 4.761 | 4.615 | 5.344 | 2.704 | 2.174 |
| 01.04.2020 | 4.782 | 4.594 | 5.453 | 5.313 | 2.148 | 2.361 |
| 02.04.2020 | 4.668 | ***4.465*** | ***6.156*** | ***5.595*** | 2.456 | 2.550 |
| 03.04.2020 | 4.585 | ***4.401*** | ***6.174*** | ***5.441*** | 2.786 | 2.770 |
| 04.04.2020 | 4.805 | ***4.256*** | ***6.082*** | ***5.330*** | 3.013 | 2.939 |
| 05.04.2020 | 4.316 | ***4.121*** | ***5.936*** | ***5.123*** | 3.135 | 3.221 |
| 06.04.2020 | 3.599 | ***4.054*** | ***3.677*** | ***4.954*** | 3.148 | 3.449 |
| 07.04.2020 | 3.039 | ***3.964*** | ***3.834*** | ***4.832*** | 3.892 | 3.729 |
| 08.04.2020 | 3.836 | ***3.948*** | ***4.003*** | ***4.554*** | 4.117 | 4.033 |
| 09.04.2020 | 4.204 | ***3.916*** | ***4.974*** | ***4.109*** | 4.056 | 4.269 |
| 10.04.2020 | 3.951 | ***3.852*** | ***5.323*** | ***3.946*** | 4.747 | 4.404 |
| 11.04.2020 | 4.694 | - | 4.133 | - | 5.138 | - |
| 12.04.2020 | 4.092 | - | 2.821 | - | 4.789 | - |
| 13.04.2020 | 3.153 | - | 2.537 | - | 4.093 | - |

\*Data obtained from John Hopkins University-Covid-19 database.

Table 1 shows that, in the last two week (between 01.04.2020-13.04.2020) the decrease in the seven days moving averages for Italy and Germany confirms the down trend. Turkey still exhibits situation close to peaking [3].

**Figure 2. Daily New Cases for Italy, Germany and Turkey**



As we can see from Figure 2, there is a similar trend for three countries. Italy is on a higher trend than the other two countries.

* 1. **SIR Model**

The SIR model is one of the simplest compartmental models, and many models are derivatives of this basic form. A SIR model is an epidemiological model that computes the theoretical number of people infected with a contagious illness in a closed population over time [4]. The name of this class of models derives from the fact that they involve coupled equations relating the number of susceptible people S(t), number of people infected I(t), and number of people who have recovered R(t) [5],[6].

The SIR (susceptible-infected-removed) model, developed by Ronald Ross, William Hamer, and others in the early twentieth century [7], consists of a system of three coupled non-linear ordinary differential equations, which does not possess an explicit formula solution. However, simple tools from calculus allow people to extract a great deal of information about the solutions. Along the way to illustrate how this simple model helps to lay a theoretical foundation for public health interventions and how several cornerstones of public health

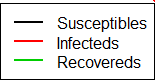
required a similar model to discover.

In this study cornerstones was used to predict to time when the Covid-19 reduce its impact on countries. After that Polynomial model was used to predict new cases in Turkey during the epidemic.

Models are created by R-programming packages [8] and the cumulative new case numbers announced in the first 25-30 days are used. The data of the first week are ignored in order not to cause contradictions in the model because of insufficient epidemic tests.

**Figure 3. SIR Model Graphs for Italy, Germany and Turkey \***

|  |  |  |
| --- | --- | --- |
| **Italy** | **Germany** | **Turkey** |
|  |  |  |



\* X-axis shows day passed during the first case, y-axis shows the days.

When Figure 3 is examined, it can be seen that:

* For Italy, between 55-60th days are represent the time for healing. First Covid-19 case is seen in January 31, 2020. The SIR Model for Italy indicates the May 1 to May 5 for recovery (the first ten day omitted from the data because of the deficient data)
* For Germany, between 55-60th days are represent the time for healing. First Covid-19 case is seen in Germany February 27, 2020 (The date of the first case seen in Germany is actually 27 January, however, since there was no big change in the number of cases until 27 February, so the start date is accepted 27 February). The SIR Model for Germany indicates the April 28 to May 3 for recovery (first seven days are omitted).
* For Turkey, between 45-50th days represents the time for healing. First Covid-19 case is seen in Turkey March 11, 2020. The SIR Model for Turkey indicates the April 27 to May 2 for recovery.

Although the start dates of the disease are different in all 3 countries, the models show that the disease will lost its power, at the end of April 2020 and the beginning of May 2020.

* 1. **Polynomial Model For Turkey New Cases Prediction**

Polynomial regression is one example of regression analysis using [basis functions](https://en.wikipedia.org/wiki/Basis_functions) to model a functional relationship between two quantities. The goal of polynomial regression is to model a non-linear relationship between the independent and dependent variables (technically, between the independent variable and the conditional mean of the dependent variable). This is similar to the goal of [nonparametric regression](https://en.wikipedia.org/wiki/Nonparametric_regression), which aims to capture non-linear regression relationships. Therefore, non-parametric regression approaches such as [smoothing](https://en.wikipedia.org/wiki/Smoothing) can be useful alternatives to polynomial regression. Some of these methods make use of a localized form of classical polynomial regression [[9]](https://en.wikipedia.org/wiki/Polynomial_regression#cite_note-6). An advantage of traditional polynomial regression is that the inferential framework of multiple regression can be used [10].

In this study, quadratic and cubic polynomial models are both examined. The β0 coefficient is rejected in the cubic model. Therefore, estimations are made by creating a quadratic model.

In the study, 26 days data used. The data used in model below:

days: 16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,34,36,37,38,39,40,41

cumulative cases : 47 , 98 , 191 , 359 , 670 , 947 , 1.236 , 1.529 , 1.872 , 2.433 , 3.629 , 5.698 , 7.402 , 9.217 , 10.827 , 13.531 , 15.679 , 18.757, 20.921, 23.934, 27.069, 30.217, 34.109, 38.226, 42.282, 47.029

**Table 2.Cubic Polynomial Regression Model**

|  |
| --- |
| lm(formula = sample$case ~ poly(sample$day, 2, raw = TRUE)) |
|  |
| Residuals: |
| Min 1Q Median 3Q Max |
| -873.72 -432.87 -88.45 225.11 2581.10 |
|  |
| Coefficients: |
| Estimate Std. Error t value Pr(>|t|) |
| (Intercept) 36224.115 2216.821 16.34 <3.76e-14\*\*\* |
| poly(sample$day, 2, raw = TRUE)1 -3778.972 162.803 -23.21 < 2e-16 \*\*\* |
| poly(sample$day, 2, raw = TRUE)2 98.282 2.838 34.63 < 2e-16 \*\*\* |
| --- |
| Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1 |
|  |
| Residual standard error: 729.1 on 23 degrees of freedom |
| Multiple R-squared: 0.9978, Adjusted R-squared: 0.9976 |
| F-statistic: 5116 on 2 and 23 DF, p-value: < 2.2e-16 |

Since p-value <0.05 all the coefficients and model are mean full. Adjusted R2 is 0.99. Model's explanation rate is high.

The model is:

y = 36.224,115 -3.778,972 x + 98,282 x2

y= cumulative new case

x=day

**Table3. New Case Predictions for Turkey.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Date** | **Real Cases** | **Predicted Cases** | **x-3σ** | **x+3σ** |
| 02.04.2020 | 18.135 | 18.550 | 17.878 | 19.186 |
| 03.04.2020 | 20.921 | 21.356 | 20.665 | 21.973 |
| 04.04.2020 | 23.934 | 24.358 | 23.647 | 24.955 |
| 05.04.2020 | 27.069 | 27.557 | 26.822 | 28.130 |
| 06.04.2020 | 30.217 | 30.953 | 30.192 | 31.500 |
| 07.04.2020 | 34.109 | 34.545 | 33.756 | 35.064 |
| 08.04.2020 | 38.226 | 38.334 | 37.514 | 38.821 |
| 09.04.2020 | 42.282 | 42.319 | 41.466 | 42.773 |
| 10.04.2020 | 47.029 | 46.501 | 45.612 | 46.920 |
| 11.04.2020 | 52.167 | 50.880 | 49.952 | 51.260 |
| 12.04.2020 | 56.966 | 55.455 | 54.486 | 55.794 |
| 13.04.2020 | 61.049 | 60.226 | 59.215 | 60.523 |
| 14.04.2020 | 65.111 | 65.195 | 64.137 | 65.445 |
| 15.04.2020 |  | 70.359 | 69.254 | 70.562 |
| 16.04.2020 |  | 75.721 | 74.565 | 75.873 |
| 17.04.2020 |  | 81.279 | 80.070 | 81.378 |
| 18.04.2020 |  | 87.033 | 85.769 | 87.077 |
| 19.04.2020 |  | 92.984 | 91.662 | 92.970 |
| 20.04.2020 |  | 99.132 | 97.749 | 99.057 |
| 21.04.2020 |  | 105.476 | 104.031 | 105.339 |
| 22.04.2020 |  | 112.017 | 110.506 | 111.814 |
| 23.04.2020 |  | 118.754 | 117.176 | 118.484 |
| 24.04.2020 |  | 125.688 | 124.040 | 125.348 |
| 25.04.2020 |  | 132.818 | 131.098 | 132.405 |
| 26.04.2020 |  | 140.145 | 138.350 | 139.657 |
| 27.04.2020 |  | 147.669 | 145.796 | 147.104 |

The results are evaluated with the SIR model. SIR model is indicated the recovery date 27.04.2020 for Turkey. The week before the April 27 is estimated to enter the stable period. The rate of increase of the outbreak will decrease. So, it is predicted at the end of the epidemic process approximately 120.000-130.000 people suffer from Covid-19 in Turkey.

1. **Conclusion**

In this paper, we focused on data analyses and prediction modeling for 3 country; Italy, Germany and Turkey. Firstly moving averages analyses is examined. From this analyses we can say that Italy and Germany are started their down trend. Then with SIR model and quadratic polynomial model are studied for prediction. From the results of analysis and modeling, it is estimated that the effects of the virus will decrease in end of April to the first week of May for the mentioned countries.

This estimation was very important for an organization, because work force management should be done and anxiety of the staff should be controlled by estimating the uncertainty. We have been used this model since the beginning of the Covid-19 cases in Turkey. We used Turkey’s advantage of late entry of the pandemic period. We deeply investigated the other countries to guess the future.

With the help of findings we can easily say that modeling is also important for organization’s planning and essential for crisis management. Although there is uncertainty, analyzing the existing data and following a model will shed light to the future. Informing the top management about critical dates and following the trends closely, we tried to manage the situation through the workforce planning, which were not easy without a plan. With this model we gave a full time service to our customers in line with the State’s economic support plan and tried to manage increased work burden on our branches. In order to minimize risk in branches and provide the sustainability of work we have planned the number employee on leave and on work with the help of findings of models. Above all, these models give us motivation and hope about the future.

The impact of the COVID-19 global pandemic has changed our daily lives, which has led to major changes in industries and also COVID-19 is still an unclear infectious disease. One way to trying to understand this unclear situation is through modeling. These models are varied, often confusing to interpret, and are not crystal balls, especially because the ideal data has not been available yet. Data transparency is crucial factor. The model can be used by the decision makers for allocating resources. But it is also our responsibility not to spread unverified news and to remain calm in these hard times. As Ghbreyesus said “We need to continue to be rational. Irrationality doesn’t help. We need to deal with the facts.”

We hope not to see such a disease again, but if we see it again; the findings in this study may help the others.

***References:***

1. <https://www.eulerhermes.com/tr_TR/ekonomik-arastirmalar/ekonomik-gorunum-raporlari/ekonomik-gorunum-covid-19-karantinadaki-ekonomiler.html>, Access date: 11.04.2020
2. Kumar Molugaram, G. Shanker Rao, in [Statistical Techniques for Transportation Engineering](https://www.sciencedirect.com/book/9780128115558), 2017
3. WHO. Coronavirus disease 2019 (COVID-19). Situation report 84. April 13, 2020. Geneva: World Health Organization, 2020.
4. [Weisstein, Eric W.](https://mathworld.wolfram.com/about/author.html) "SIR Model." From [MathWorld](https://mathworld.wolfram.com/)--A Wolfram Web Resource. <https://mathworld.wolfram.com/SIRModel.html>

# [Sk Shahid Nadim](https://arxiv.org/search/q-bio?searchtype=author&query=Nadim%2C+S+S), [Indrajit Ghosh](https://arxiv.org/search/q-bio?searchtype=author&query=Ghosh%2C+I), [Joydev Chattopadhyay](https://arxiv.org/search/q-bio?searchtype=author&query=Chattopadhyay%2C+J). “Short-term predictions and prevention strategies for COVID-2019: A model based study”, Populations and Evolution (q-bio.PE); Dynamical Systems (math.DS), 2020.

1. Linda JS Allen. An İntroduction to Stochastic Epidemic Models. In Mathematical Epidemiology, pages 81–130. Springer, 2008.
2. R.M. Anderson. Discussion: the Kermack-McKendrick Epidemic Threshold Theorem. Bulletin of Mathematical Biology, 53(1):1–32, 1991.
3. Soetaert K, Petzoldt T, Setzer RW ,“Solving Differential Equations in R: Package deSolve.” Journal of Statistical Software, **33**(9), 1–25. ISSN 1548-7660, 2010. Doi: [10.18637/jss.v033.i09](https://doi.org/10.18637/jss.v033.i09), <http://www.jstatsoft.org/v33/i09>.
4. M. Zhan, C. B. Dean, R. Routledge, P. Gallaugher, A. P. Farrell and H. Thorarensen .

Inference on Segmented Polynomial Models .*Biometrics* Vol. 52, No. 1 (Mar., 1996), pp. 321-327

1. *Yin-Wen Chang; Cho-Jui Hsieh; Kai-Wei Chang; Michael Ringgaard; Chih-Jen Lin (2010).* ["Training and testing low-degree polynomial data mappings via linear SVM"](http://jmlr.csail.mit.edu/papers/v11/chang10a.html)*.* [Journal of Machine Learning Research](https://en.wikipedia.org/wiki/Journal_of_Machine_Learning_Research)*. 11: 1471–1490.*