medRxiv preprint doi: https://doi.org/10.1101/2020.07.12.20151860; this version posted July 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Implied susceptible population size: modelling and projecting Covid19 dynamics via the SIR equations

Isaac Meilijson
School of Mathematical Sciences
Raymond and Beverly Sackler Faculty of Exact Sciences
Tel Aviv University, Tel Aviv 6997801, Israel
E-mail: isaco@tauex.tau.ac.il

July 12, 2020

Abstract

A notion of implied susceptible population size is introduced in the context of the SIR differential equations in Epidemiology. This notion is applied to the analysis and projection of Covid19 2020, illustrated on the data of Germany, Israel and USA.

1 Introduction

The SIR (Susceptibles, Infected, Removed) model introduced by Kermack & McKendrick in 1927 ([6]) for the progress of an epidemic describes the interdependence between the cumulative number $X(t)$ of affected cases, the cumulative number $R(t)$ of removed cases (dead or recovered) and the ensuing current number of infected cases $I(t) = X(t) - R(t)$. The version of this model to be applied in the current report conforms with the usual assumption $dR(t) = \gamma I(t)dt$ that the number of new removed cases constitute a fixed proportion of the currently infected cases, and postulates the usual assumption that the number of new affected cases is proportional to the number $K - X(t)$ of susceptible cases and (less commonly, after Grenfell, Bjørnstad & Filkenstädt ([4])) to a fractional power $I(t)^\alpha$ $(0 \leq \alpha < 1)$ of the number of currently infected cases. In other words,

$$dX(t) = \beta(I(t))^\alpha(K - X(t))dt \tag{1}$$

The companion publication by Alon and the author ([1]) to the current report describes methods to estimate the parameters α, β, γ and K. It also describes a regression pre-processing procedure that modifies minimally the infected and removed components of the empirical data (X_1, I_1, R_1), $(X_2, I_2, R_2), \cdots, (X_n, I_n, R_n)$, so that the proportionality of new removals and currently infected cases will better hold. This procedure ignores equation (1).

The data analyzed in the current report has been submitted to this pre-processing procedure, and maximum likelihood estimation based on the SIR differential equations has led to the estimation of the four parameters in a period considered stationary, days 60 to 120 in the monitoring of Covid19 2020. The working paradigm is that the three parameters α, β, γ hold more generally. Under this assumption, equation (1) is rewritten as

$$K(t) = X(t) + \frac{\Delta(X)}{\beta(Y(t))^\alpha} \tag{2}$$

to be interpreted as the *implied susceptible population size* ISPS. Letting $\Delta(X)$ be a smooth version of $X(t+1) - X(t)$, K(t) is evaluated daily to provide an idea of the current target population of the epidemic. Due to the persistent missing data on weekends, the smooth version is taken throughout to be a moving average on the week around the date in question.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.12.20151860; this version posted July 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

The notion of implied susceptible population size ISPS mimics implied volatility in Finance. The Black & Scholes ([3]) formula for option pricing does not quite hold when the asset estimated volatility is substituted, but it is such a convenient language tool that reverse engineering as in (2) is applied, and implied volatility is defined as the value for which formula and empirical price are the same. This value changes with the maturity date and the moneyness, and this non-constant shape is termed "volatility smile". This is a word we will not export from Finance into Epidemiology.

A word on the parameter α. Imagine a small-world network topology composed of sparsely connected islands that are internally densely connected. Once an island is affected, contagion will apply locally. Much as in the Harris contact process ([5]), if the affected cases are considered a growing ball that transmits the disease by its perimeter, $\alpha = \frac{1}{2}$ models the intra-island spread force. The "ball" could have some fractal dimension rather than a solid area. [1] presents an argument that on infinite populations, $\alpha = 1$ yields exponential growth of X, while $\alpha < 1$ gives rise to linear growth of X, with $I(t)$ increasing asymptotically to a finite constant.

Without claiming to delve deeply into Epidemiology, it is suggested that isolation measures achieve stationarity by cutting contagion across islands. Relaxing isolation measures may have the effect of allowing for more contagion between islands, and this is what ISPS is intended to reflect.

The theoretical or technical contribution of this report is the introduction of implied susceptible population size ISPS. The main goal is to illustrate on the Covid19 2020 data of Germany, Israel and USA, the potential uses of ISPS in measuring and projecting disease progress. Hopefully, proper quantitative assessments will prove useful in pandemic management.

2 Scenario analysis

Attach to the ISPS function its logarithmic daily rate of growth function "Delta Log K" $DLK(t) = log(\frac{K(t)}{K(t-1)})$. Under stationary behavior, this function should in principle be zero and in practice it should oscillate around zero. Figures 13, 8 and 3 display DLK in Germany, Israel and USA respectively. During the (presumably stationary) training period from day 60 to day 120, DLK oscillates around zero, progressively stable.

In Germany this mode persists to the last monitoring date, day 165, with a perturbation around day 150, that led to the re-adjustment of ISPS from 200000 to 220000 observed in Figure 14.

In Israel DLK is essentially zero between days 105 and 120, after which isolation measures were lifted, sharply increases and "stabilizes" around 0.015 until day 140 and then experiences a new height around 0.03 until the end of the monitoring period.

In USA DLK never quite stabilizes but tends to show mean reversion towards zero until day 110, then oscillates around 0.005 until day 140, sharply increases to its new metastable height 0.02.

It should be stressed that a period of constancy of DLK signifies exponential growth of the target population. This is apparent in Figures 9 and 4.

There is no basis on which to predict the near-future behavior of DLK. But it is possible to analyze the consequences of particular scenarios.

The graphs on the right of Figures 10 and 5 display the ultimate optimistic scenario - DLK drops instantly to zero and stays there, meaning that ISPS freezes at its current value. This scenario would see the total number of affected cases in USA, 2.7M at day 165, increase to 4.5M. The number of infected cases 1.7M will reach a maximum of 2.1M at day 187 (the basic reproduction number R_0 crosses from above to below 1) and then decrease towards zero. The parallel scenario in Israel would see the total number of affected cases 27000 at day 165, increase to 42000. The number of infected cases 10000 at day 165 will reach a maximum of 14000 at day 175 (again, the basic reproduction number R_0 crosses from above to below 1) and then decrease towards zero. This date 175 is tomorrow at the writing of this report, when the statistics are more grim than they were a week ago. These numbers provide unrealistic lower bounds to disease progress.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.12.20151860; this version posted July 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

The nature of scenarios to be investigated is as follows. A rate of exponential decay is fixed, and DLK is assumed to decay at this rate to zero, $DLK(t+1) = \exp(-1/N)DLK(t)$. N will be fixed throughout at 20 for illustration purposes. The initial value at the final date $F = 165$ of the monitoring period is set at $\max(DLK(F), DLK(F-1)\exp(-1/N), DLK(F-2)\exp(-2/N), DLK(F-3)\exp(-3/N), \cdots)$, going back, say, for one week.

These scenarios, displayed as the bottom functions in the graphs on the left of Figures 10 and 5, should be perceived as markedly optimistic, although these are not lower bounds, since the ultimate optimistic scenario is "possible".

The analysis proceeds under increasingly pessimistic scenarios, all decaying exponentially at rate $\exp(-1/N)$ and all sharing the definition of the value at the final date of monitoring. Under these scenarios DLK stays constant for a number (window) of days and then decays exponentially.

The middle and top functions in the graphs on the left of Figures 10 and 5 show the result of windows 15 and 30. The number of affected cases in Israel, below 17000 in the training period and 27000 in day 165, would grow to anywhere between 65000 and 120000 under the three scenarios. The number of affected cases in the USA, 2.7M in day 165, would grow to anywhere between 6.5M and 12M under the three scenarios.

Although removed cases are not split in the current report into dead and recovered, to help assess this issue, the proportion of dead among removed cases is reported for every country under study.

3 Covid19 2020 in the United States

The MLE of α is 0.43 but it makes no difference to apply $\alpha = 0.5$ (see Figure 3). The values of the other parameters corresponding to the adopted value $\alpha = 0.5$ are $\beta = 2.3768 * 10^{-5}, \gamma = 0.0111$ and $K = 238219 = 1.53 * X(120)$.

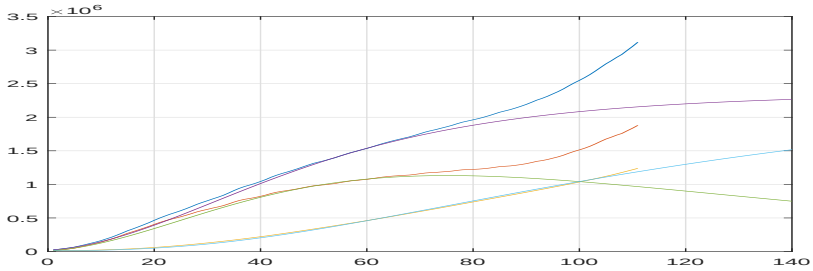

Figure 1: USA: Cumulative affected, currently infected, cumulative removed cases in the training period (first 60 days) and beyond, pre-processed data and SIR solution. Heavy disparity at later dates consistent with the non-stationarity displayed in Figure 3.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.12.20151860; this version posted July 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

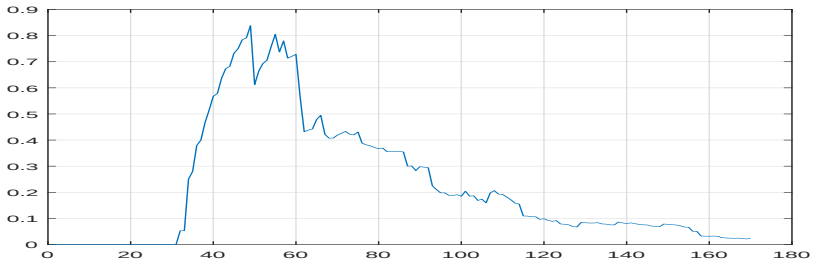

Figure 2: USA: Proportion of dead among removed (dead + recovered) cases. Decreased from 50% to 10% during learning period (days 60 to 120), below 10% and still decreasing thereafter.

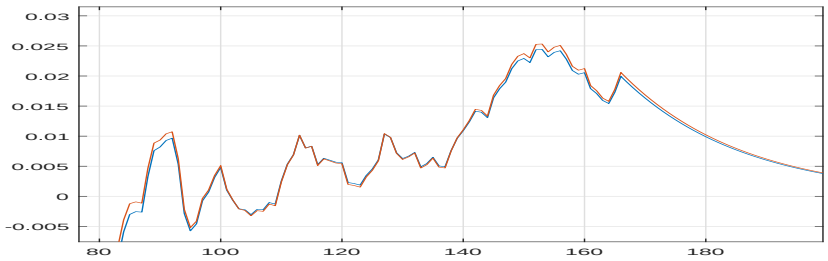

Figure 3: USA: Incremental logarithm of ISPS K. Comparison between MLE $\alpha = 0.43$ and $\alpha = 0.5$. The value $\alpha = 0.5$ is adopted.

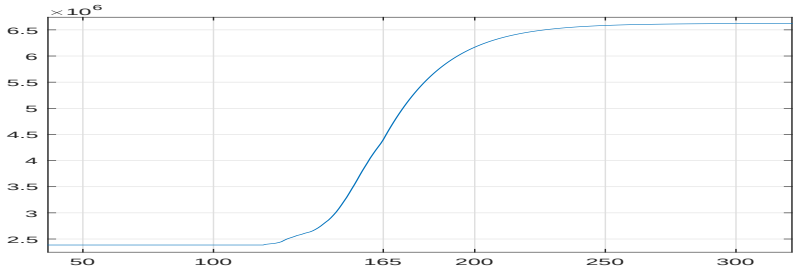

Figure 4: USA: ISPS K under the no-delay exponential decay scenario. The total number of affected cases until day 165 was 2.84M.

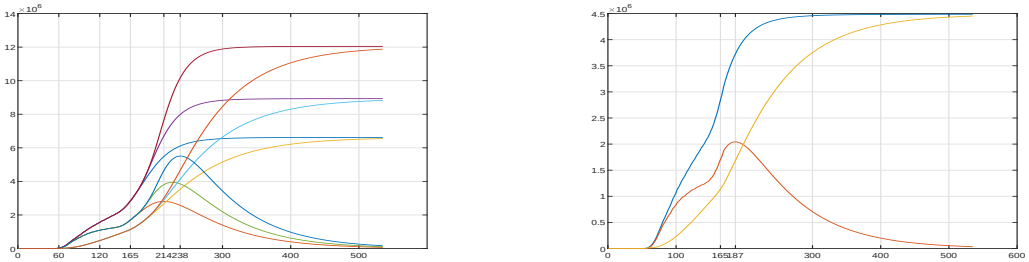

Figure 5: USA: Total affected, currently infected and total removed cases, data and projection. Right: the ISPS is frozen at its value in the last monitoring date. Left: exponential decay scenario at no delay and at 15 and 30 day delays.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.12.20151860; this version posted July 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

4 Covid19 2020 in Israel

The MLE of α is 0.54 but it makes no difference to apply $\alpha = 0.5$ (The profile loglikelihood is 0.65 higher than at the MLE). The values of the other parameters corresponding to the adopted value $\alpha = 0.5$ are $\beta = 9.314 * 10^{-4}, \gamma = 0.0310$ and $K = 16850 = 1.011 * X(120)$.

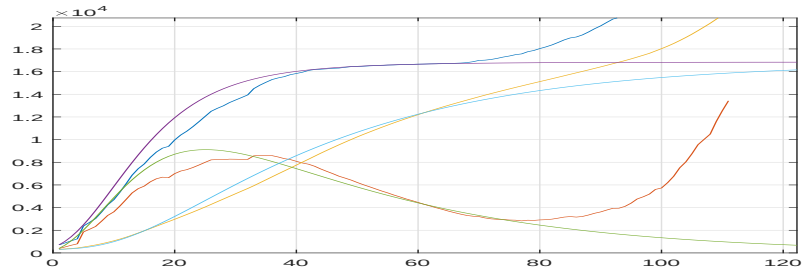

Figure 6: ISRAEL: Cumulative affected, currently infected, cumulative removed cases in the training period (first 60 days) and beyond, pre-processed data and SIR solution. Heavy disparity at later dates consistent with non-stationarity displayed in Figure 8.

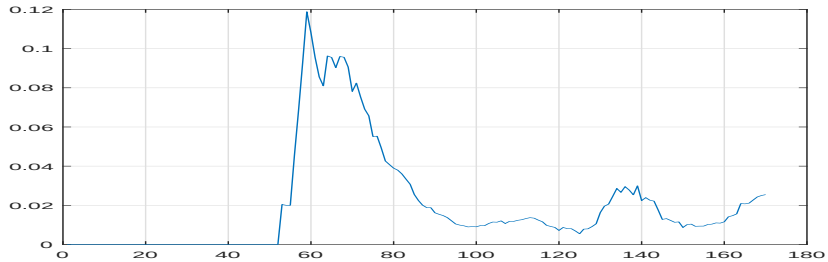

Figure 7: ISRAEL: Proportion of dead among removed (dead + recovered) cases. Decreased from 10% to 1% during the learning period (days 60 to 120), fluctuates around 1.5% thereafter.

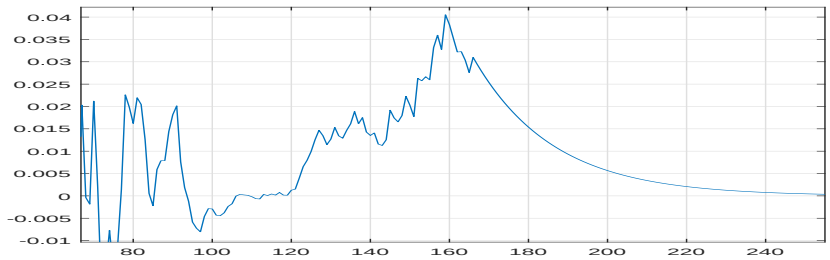

Figure 8: ISRAEL: Incremental logarithm of ISPS K.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.12.20151860; this version posted July 14, 2020. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

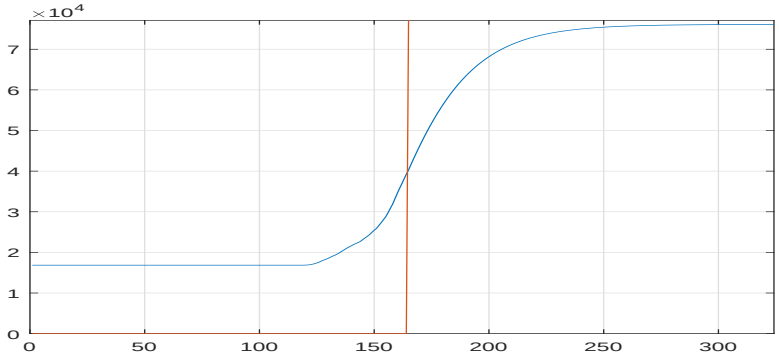

Figure 9: ISRAEL: ISPS K under the no-delay exponential decay scenario.

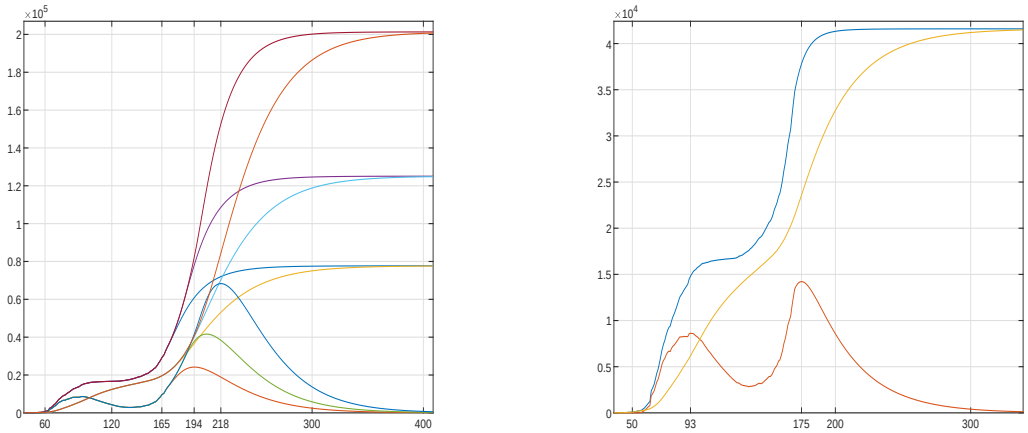

Figure 10: ISRAEL: Total affected, currently infected and total removed cases, pre-processed data and projection. Right: ISPS K is frozen at its value in the last monitoring date, day 165. Left: exponential decay scenario at no delay and at 15 and 30 day delays.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.12.20151860; this version posted July 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

5 Covid19 2020 in Germany

The MLE of α is 0.5. The values of the other parameters are $\beta = 1.9597 * 10^{-4}, \gamma = 0.0772$ and $K = 199890 = 1.12 * X(120)$.

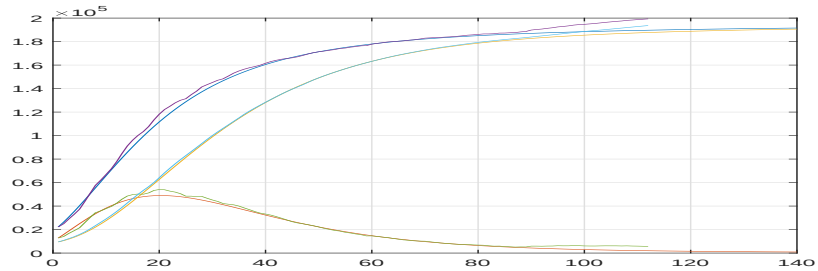

Figure 11: GERMANY: Cumulative affected, currently infected, cumulative removed cases in the training period (first 60 days) and beyond, pre-processed data and SIR solution. Small disparity at the end consistent with jump in Figure 14.

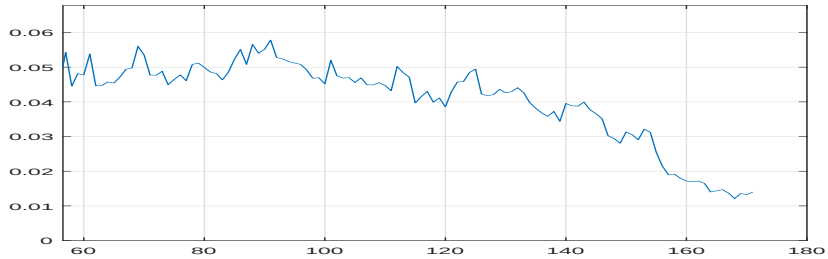

Figure 12: GERMANY: Proportion of dead among removed (dead + recovered) cases. Decreased from 50% to 10% during learning period (days 60 to 120), below 10% and still decreasing thereafter.

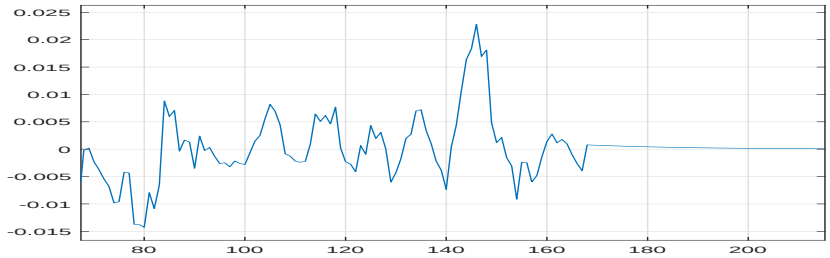

Figure 13: GERMANY: Incremental logarithm of ISPS K.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.12.20151860; this version posted July 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

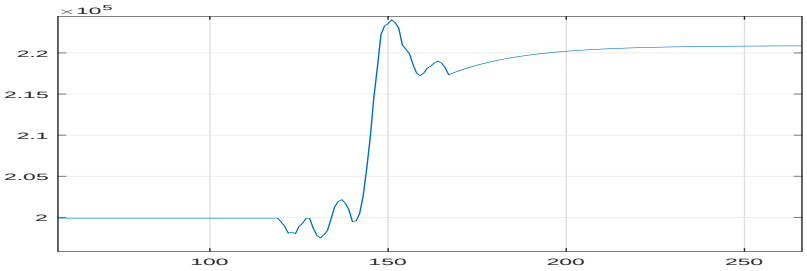

Figure 14: GERMANY: ISPS K under the no-delay exponential decay scenario.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.12.20151860; this version posted July 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Acknowledgements

Thanks are due to Nitay Alon for joint work and to Eytan Ruppin and Laura Sacerdote for helpful suggestions.

The data analyzed in this work is taken from the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University.b

References

[1] Alon, N. and Meilijson, I. (2020). Random time transformation analysis of Covid19 2020. MedRxiv.

[2] Bassan, B., Marcus, R., Meilijson, I. and Talpaz, H. (1997). Parameter erstimation in differential equations, using random time transformations. *Journal of the Italian Statistical Society*, **6**, 177–199.

[3] Black, F. and Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, **81 (3)**, 637-654.

[4] Grenfell, B.T., Bjørnstad, O.N. and Filkenstädt, B. A. (2002) Dynamics of Measles epidemics: scaling, noise, determinism and predictability with the TSIR model. *Ecological Monographs*, **72(2)**, 185-202.

[5] Harris, T.E. (1974). Contact interactions on a lattice. *Annals Probab*, **2 (6)**, 969-988.

[6] Kermack, W. O., McKendrick, A. G. (1927). A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society A*, **115 (772)**, 700-721.