Bocconi

BOCCONI UNIVERSITY

Università Commerciale Luigi Bocconi

Assignment: Simulation and Modelling

Rocco Gazzaneo (3040302)
Alessandro Caruso (3054408)

Date final version: March 21, 2021

Abstract

The Sars-CoV-2 pandemic had catastrophic effects on the Italian territory and especially on the Lombardy region. Different studies have been conducted on its spread to compensate for the limited knowledge on COVID-19. This paper provides the development of a SEIRD, a deterministic compartmental model that simulates the spread of COVID-19 in 12 cities of the Lombardy Region. Bergamo daily infections are closely examined by augmenting a SIR model with age structures. The model estimates are further refined with the implementation of preventive measures such as total lockdown, social distancing, and protective masks, employing deterministic and stochastic approaches.

Keywords: SEIRD, SIR models; stochastic simulation; compartmental models

1 Introduction

At the end of December 2019, China has witnessed the spread of an unrecognized virus which would have changed the sort of the world. The first case was identified in a seafood market in Wuhan, the capital of the Hubei province, which has been recognized as the first hotspot of the coronavirus disease caused by severe acute respiratory syndrome coronavirus 2, SARS-CoV-2[1]. Most countries around the world have been affected by COVID-19, with major outbreaks in the United States, Russia, Brazil and Western Europe. The WHO (World Health Organization) declared the pandemic on 11th March of 2020, and in most countries strict measures have been applied.

Regarding the Italian situation, the first appearance of the virus was recorded as of 31st January 2020 when two Chinese tourists tested positive in Rome. The first clusters of cases were identified in Lombardy on the 15th of February, especially in the province of Lodi (Codogno) where a man was diagnosed as the first Italian case [2]. Besides, the first Italian death was registered on the 21st of February in the region of Veneto when a resident of the municipality of Vo' died of pneumonia due to SARS-CoV-2 infection. To contain the epidemics, the Italian government took drastic measures such as the full lockdown on March 11th 2020 to reduce the mobility of citizens and the possibility for the virus to further spread. Different studies have been conducted in order to predict the spread of the virus and to suggest the best prevention measures for governments to adopt. For instance, regarding Italy, an adjusted time-dependent SIRD model has been used to predict the behavior of the epidemic, as described in Ferrari et al. (2020). Another study instead is proposed by Giordano et al. (2020), where a SIDARTHE model was used, which discriminates between infected individuals depending on whether they have been diagnosed and on the severity of their symptoms. Also, Guzzetta et al. (2020), have assessed the impact of an uncontrolled epidemic on the healthcare system for Lombardy.

Here, we provide an analysis of 31,503 confirmed cases reported in Lombardy, which corresponds to the period from 31st January to the beginning of May 2020. Twelve provinces have been analyzed by applying a SEIRD model which has been calibrated through MSE optimization techniques. Then the analysis focuses on Bergamo, which was one of the most affected provinces by the virus, for which a simple SIR model with age structure (5 years range) has been used. Lastly, stochasticity has been introduced in the recovery process and the infection transmission process to include the previously described restrictive measures adopted by the Italian government and make the model more realistic.

[1] en.wikipedia.org/wiki/COVID-19

[2] www.corriere.it/cronache/20_febbraio_21/coronavirus-italia-cosa-sappiamo-casi-codogno-lombardia-27fa736c-548b-11ea-9196-da7d305401b7.shtml

2 Data

The dataset provides time-series data for infections from SARS-CoV-2 for twelve among the largest cities in the Italian region of Lombardy during the first epidemic wave of March 2020. Data has been gathered for 92 days. No more than 5 daily cases per day are registered for all provinces until the 50th day after which the epidemic wake intakes, showing its peak at the 68th-69th day and then plummeting down to zero at the end of the time series. This behavior is homogeneous across all provinces. For each province, the total number of residents has been extracted from the Italian National Institute of Statistics (ISTAT) website[3], altogether with the division in age ranges for the province of Bergamo.

3 Methods & Results

3.1 SEIRD model

To understand and forecast the trend of the process, a SEIRD compartmental model has been trained separately on all cities for the period precedent to the peak of cases. The choice of the model is supported by the intrinsic properties of the SARS-CoV-2 pandemic whereby a susceptible individual (S) that contracts the virus undergoes a short incubation period (E) before becoming infectious (I) and later recovered (R) or in the fatal case Dead (D). According to different studies such as in (Adrielle dos Santos et al., 2021), reinfection is possible for people who produce a weak antibody response against SARS-CoV-2 as they may be likely to be infected again by the virus in the future, also given the degree of the mutability registered at the end of 2020. Considering that the analysis is addressed to the first 90 days of the outbreak in Italy, and given that in the period between February and May 2020 no cases of reinfection have been officially registered, the SEIRD was considered the appropriate model. Furthermore, as the short period (max 74 days), both a homogeneous (birth rate = mortality rate) SEIRD model with demography and a simpler version of the SEIRD, which only takes into account death due to the SARS-CoV-2 disease, have been tested. Considering that the difference between the results obtained was negligible, the latter has been chosen as less complex for the calibration procedure. The basic continuous-time differential equation formulation of a SEIRD epidemic process in a homogenous population takes the following form:

$$
\begin{aligned}
\frac{dS}{dt} &= \frac{\beta I S}{N} \\
\frac{dE}{dt} &= \frac{\beta I S}{N} - aE \\
\frac{dI}{dt} &= aE - (\gamma + cv)I \\
\frac{dR}{dt} &= \gamma I \\
\frac{dD}{dt} &= cvI \\
S + E + I &+ R + D = N
\end{aligned}
\tag{1}
$$

with basic reproduction number as:

$$
R_0 = \frac{\beta}{\gamma + cv}. \tag{2}
$$

For each of the provinces, a SEIRD model was simulated and calibrated using as a training set the period where the daily cases are increasing (slightly before each peak). According to the CDC[4], the Centers for Disease Control and Prevention, adults with mild to moderate COVID-19 remain infectious no longer than 10 days after symptoms. Therefore, the parameter d, the average duration of infectiousness has been fixed to 7 days. The covid mortality rate in the region of Lombardy has been computed and it accounts for 0.04. Consequently, γ, the average rate of recovery, which formula is $\frac{1}{d} - cv$ and accounts for 0.10.

[3] `www.istat.it`

[4] `https://www.cdc.gov/coronavirus/2019-ncov`

The model has been fitted using the least-squares estimation (LSE), which involves minimizing the root mean squared error:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(I_{daily}-\hat{I}_{pred})^2} \quad (3)$$

Furthermore, it has been calibrated using Maximum Likelihood Estimation (MLE) embedded in an MCMC where a Poisson distribution has been assumed for the model outcomes, firstly with likelihood-function-free methods and then using Bayesian approximates, as shown in figure 1 and 2 in the Appendix. Here we decide to show the results with the LSE calibration method because of its simplicity and ease of interpretability. As is shown in figure 1 and 2, which has been limited to show the provinces of Bergamo and Milano, the model correctly simulates the increase of the daily cases with a relatively low mean squared error.

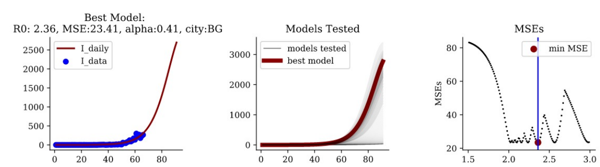

Figure 1: Calibration of Bergamo province using RMSE

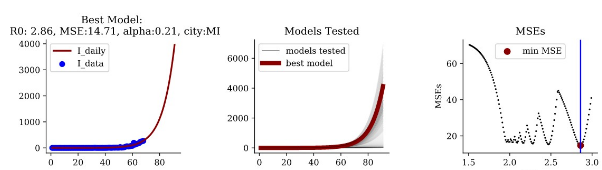

Figure 2: Calibration of Milan province using RMSE

It can be seen that in both provinces the best model delivers and R0 which is greater than 1, respectively 2.86 for Milan and 2.36 for Bergamo. These results are expected as the calibration was conducted on the first stage of the infection for each city. Given that the R0 measures how transferable a disease is, an $R0 > 1$ would lead to an exponential spread of the infection. This is why we should not expect the rest of the simulated curve to resemble the pattern we observe in the rest of the data, as it is shown in Figure 3 for the province of Bergamo.

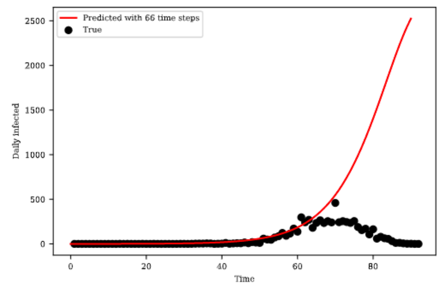

Figure 3: Prediction of cases if epidemics left uncontrolled

3.2 Age-stratified SIR

At a later stage, the analysis focus was restricted to the city of Bergamo only. This choice is motivated by the unusual pattern of infected cases and by the relatively high MSE (23.62) compared to other cities. The latter allows for a better understanding of model improvements during parameter calibrations.

Furthermore, the number of people belonging to each age group in the population is required to simulate a real-case scenario, which is downloaded from ISTAT [5]. Each age population is then divided by the total population (obtaining relative terms) and then multiplied by the provided number of susceptibles in Bergamo. The outcome is a vector containing an estimate of the number of susceptibles per each age group which is the input vector of susceptibles S0 fed to the model.

As required a simple SIR model is run to explain the pattern of Bergamo infections across age groups. The initialization of the infected people forces the authors to choose the age group in

[5] www.data.istat.it/bergamo

which the first infection occurred. As the first case of coronavirus in Bergamo city corresponds to an 83 years old resident, then the last age group has been chosen[6].

The model does not allow for individuals to switch age groups given the restricted period under analysis and the negligible impact it has on the model outcome.

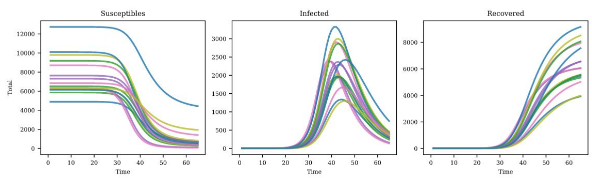

Figure 4: Behavior of susceptibles, infected and recovered per each group

3.2.1 Calibration and results

At a later stage, it is necessary to calibrate the model, which amounts to find the optimal value of q, the infection-specific parameter, that minimizes the RMSE. The optimal value for q is obtained by looping among many sensible estimates of R0 ranging from (1, 3), solving for the corresponding value of q, by multiplying to the maximum eigenvalue of the contact matrix, and using it to find the predicted daily infected that are the lowest in mean squared difference compared to the true data.

The optimal value of q is 0.0124 which corresponds to a value of R0 equal to 1.81, as shown in the following graph.

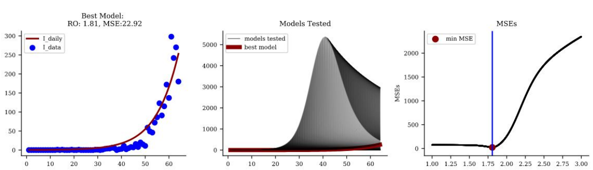

Figure 5: Model output after calibration

[6]https://www.ilpost.it/2021/02/19/bergamo-coronavirus-un-anno-dopo/

3.2.2 The most infected age group

To calculate which age group is infected the most, after the parameters have been calibrated, the cumulative sums of the daily cases for each age group has been computed as it is shown in the next figure.

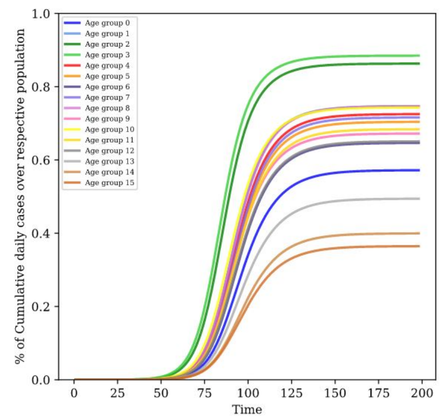

Figure 6: The third age group is the most infected

Thus, each of the cumulative sums is divided by the initial population of susceptibles per age group to get the proportion of infected per age group. **The most infected age group is 15-19** and the second most infected age group is 10-14.

3.3 Improving the model

In this paper, two different strategies have been adopted to analyze the impact of the restrictive measures implemented in Lombardy. Both of them consider the same restrictive measures which have been implemented by the Italian government in sequence: school closing, quarantine and social distancing/adoption of masks. In both strategies, the corresponding parameters are tuned dynamically to influence the behavior of the epidemic wave.

The two approaches differ by the presence of stochasticity in the model.

3.4 The deterministic approach

The idea behind the implementation of the restrictive measures in a deterministic approach is similar to the previous implementations with one crucial difference: now, instead of training one model for all the data points, different models would be sequentially trained depending on how many restrictive measures are considered.

In particular, if the three restrictions are decided to take place at times 57, 62 74, four models have been trained.

- In the period 0-57, the SIR model developed for the previous section 2 has been applied, using the best q parameter that was minimizing the MSE (0.0135).
- In the period 57-62, school closing restrictions take place, which means that a different model has been trained, tuning new parameters specific to this time range. In particular, it has been introduced a **scaling factor** that partially modifies the upper-left box of length 5 of the contact matrix, under the assumption that only the first 5 youngest age classes are affected by the first restriction. The optimal scaling factor is 0.48, **This reduces the MSE by half, shrinking it to 14.44**. The new estimate of R0 goes from 1.98 to 1.87.
- In the period 62-74, a new model has been trained, but the previous q and scaling factor remained unchanged for the following periods. Since the restriction that will take place is quarantine, the measure adopted is to shrink uniformly the contact matrix by multiplying it with a number between 0 and 1 which is tuned to minimize the RMSE. The optimal new scaling factor parameter is 0.78, which decreases the estimate of R0 to 1.81. This reduction makes sense because it is known that the R0 positively depends on the value of q and the maximum eigenvalues of the contact matrix. Therefore, leaving q constant, if the latter (p(C)) decreases, the estimate of R0 would decrease as well.
- For the last epidemic stage, the value of q has been tuned, which is equivalent to uniformly scale the contact matrix, but it differs from a theoretical perspective since the adoption of protective masks and social distancing reduce the rate at which individuals get infected rather than reducing the contact rate.

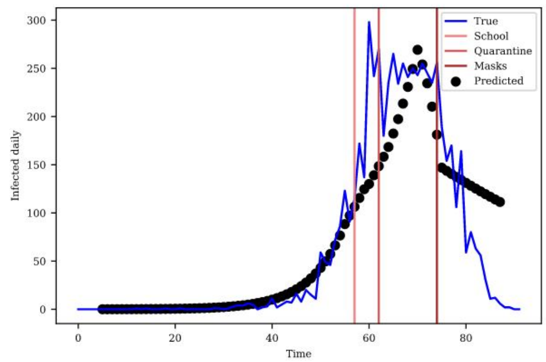

Figure 7: Ensemble of deterministic models to enforce restrictive measures

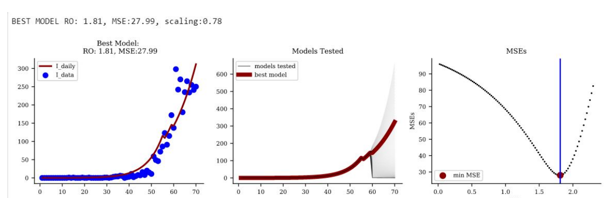

Figure 8: Tuning scaling factor for mask preventive measures

3.5 The stochastic approach

Next, a stochastic approach has been adopted. This would be able to incorporate uncertainty in the parameters, that would be disregarded by a completely deterministic model. The idea is to treat the transmission rate as a Bernoulli random variable and to sample at each iteration the number of people that get infected in each age group.

The procedure has different outcomes at each simulation as the number of people that get infected varies across iterations. The outcome of 50 simulations with recovery rate 1/8, 100 days, and 10 steps per day outputs the following results for which only the last age group has been selected.

At this stage, the stochasticity of the model can be used to implement the restrictive measures in the Lombardy region described as follows

- **School closing**: For this measure, the first 5 elements of the transmission rate vector are multiplied by a predefined constant (0.2) that shrinks the probability that the age groups 0-20 will get infected.
- **Quarantine**: For this approach, the full event rate vector is again multiplied by a number between 0 and 1 that simulates the reduced contact across age groups during the quarantine.
- **Social distancing / Masks**: Finally, the parameter to be modified is q, that is infection-specific parameter has been reduced due to preventive measures.

In the next figure, the results after the application of restrictive measures are shown, only for one age group (75+ age range).

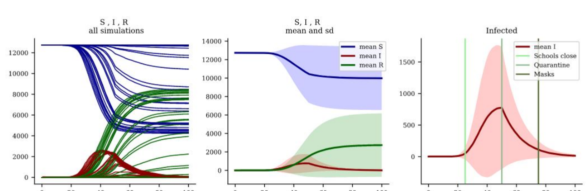

Figure 9: Application of restrictive measures

The following figures show that the last age group experienced much more infections than the third age group both because of the fragility of the elderly class but mainly after the shrinkage of the first 5 components of lambda after 25 iterations.

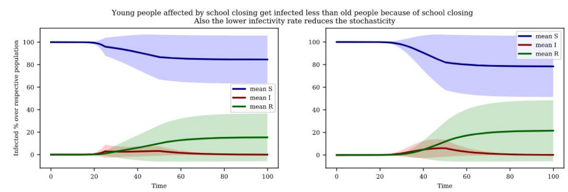

Figure 10: Response to school closing of old people compared to young people

4 Conclusion

In this paper, we firstly have proposed a SEIRD model to investigate the spread of COVID-19 in Italy. The results, which are obtained by fitting the data of twelve provinces of the Lombardy regions available from March 2020 until May 2020, show a good fit to the data with two different calibration procedures: least-squares estimation (LSE) and maximum-likelihood estimation (MLE). Secondly, a simple age-stratified SIR model has been proposed for the Bergamo province. Next, the model has been extended by adding those drastic restrictions that have been implemented in the Lombardy region by the Italian government during the same period. The results show how important those restrictions were in limiting COVID-19 transmission and what would have happened if the epidemic was left uncontrolled. Concerning the Italian epidemic evolution, further studies would be devoted to understanding the real impact of new Italian measures on the spread of the virus, such as the division of regions into different colors (yellow, orange, and red) depending on the value of Rt, together with the improvement of the calibration procedure.

References

Adrielle dos Santos, L., de GÃis Filho, P. G., & Silva, A. M. F. (2021). Recurrent covid-19 including evidence of reinfection and enhanced severity in thirty brazilian healthcare workers. *Journal of Infection*.

Ferrari, L., Gerardi, G., Manzi, G., & Micheletti, A. (2020). *Modelling provincial covid-19 epidemic data in italy using an adjusted time-dependent sird model.*

Giordano, G., Blanchini, F., Bruno, R., Colaneri, A., & Colaneri, M. (2020, 06). Modelling the covid-19 epidemic and implementation of population-wide interventions in italy. *Nature Medicine*, *26*, 1-6. doi: 10.1038/s41591-020-0883-7

Guzzetta, G., Poletti, P., Ajelli, M., Melegaro, A., Filippo Trentini, V. M., Cereda, D., ... Merler, S. (2020). *Potential short-term outcome of an uncontrolled covid-19 epidemic in lombardy, italy, february to march 2020.*

5 Appendix

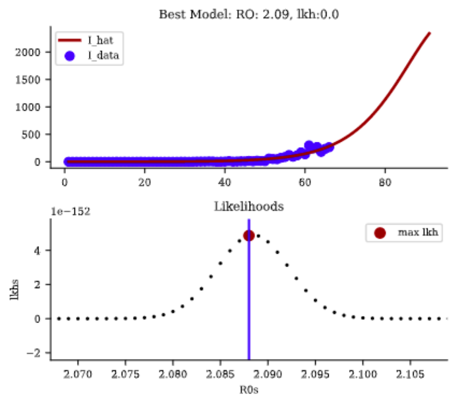

Figure 1: Best SEIRD model through MLE calibration for the province of Bergamo

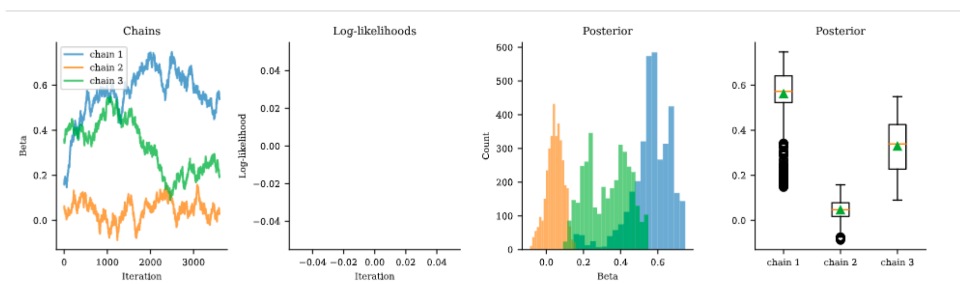

Figure 2: Calibration of Beta parameter in a SEIRD for the province of Bergamo

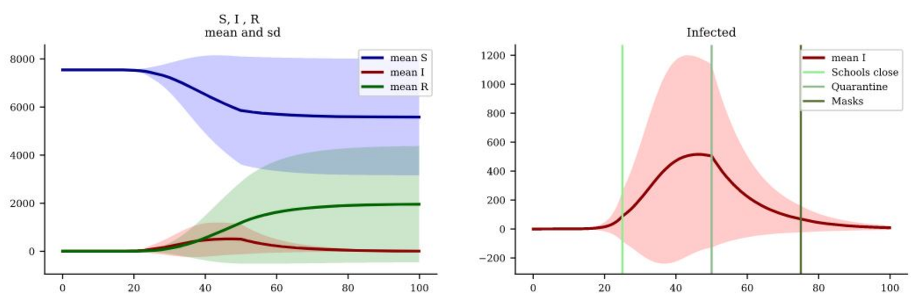

Figure 3: Stochastic model output by averaging age groups and implementing restrictive measures