# Why is COVID-19 Mortality in Lombardy so High? Evidence from the Simulation of a SEIHCR Model*

Carlo A. Favero
This Version, April 10th 2020

**Abstract**

The standard SEIR model based on a parameterization consistent with the international evidence cannot explain the very high COVID-19 related mortality in Lombardy. This paper proposes an extension of the standard SEIR model that is capable of solving the puzzle. The SEIR model features exogenous mortality: once Susceptible individuals become first Exposed, and then Infected, they succumb with a given probability. The extended model inlcudes an Hospitlization process and the possibility that Hospitalized patients, who need to resort to Intensive Care Unit, cannot find availability because the ICU is saturated. This Constraint creates an additional increase in mortality, which is endogenous to the diffusion of the disease. The SEIHCR (H stands for Hospitalization and C stands for Constraint) is capable of explaining the dynamics of COVID-19 related mortality in Lombardy with a paramerization consistent with the international evidence.

*Carlo Favero: Dept. of Finance, IGIER-Bocconi, and CEPR. carlo.favero@unibocconi.it. I thank the editor, Charles Wyplosz and an anonymous referee for their useful comments. I also thank Yakov Amihud,Francesco Corielli, Francesco Giavazzi, Luigi Guiso, Andrea Ichino, Marco Olivari, Fausto Panunzi, Franco Peracchi, Carlo Pinardi, Aldo Rustichini, Guido and Marco Tabellini for many useful comments and discussion. Giovanni Favero brought my attention to SEIR models. I worked on this project full-time during the Milan lockdown. The fact that I was stuck at home while my daughter Vittoria, who is a young doctor, was everyday on the front line gave me an enormous motivation. I dedicate this effort of mine to the hospital personnel in Lombardy, they deserve it fully.

# 1 Introduction

COVID-19 related mortality in Lombardy is way above COVID-related mortality everywhere else. Using data made available by the Civil Protection, https://github.com/pcm-dpc/COVID-19 , Figure 1 illustrates the ratio of deaths (morti) to total cases (casi_totali) observed daily over the period February 24th 2020-April 8th 2020 which shows a reported lethality growing rather steadily from an initial 4 per cent to a value reaching 16 per cent at the end of the sample (9722 Fatalities for 53414 total observed cases).

**INSERT FIGURES 1 HERE**

This pattern of lethality cannot be replicated by standard SEIR model, which has been successfully applied to the analysis of COVID-19 diffusion in China [Wu, et. al, Kucharski et. al.]. The second line in the graph reports the pattern of the ratio of fatalities to the sum of Exposed, Recovered and Removed as Fatalities generated by a SEIR model with an internationally consistent CFR of 0.0138 ( see, for example, Verity et al(2020)) which is nowhere near the observed data.

Two possible explanations can be considered for a discrepancy between observed data and model simulated data: either the model is wrong or the data are wrong. As matter of fact, the data can be wrong the total observed cases observe do not include patients with mild symptoms, which were not hospitalized and were therefore not tested.

In this paper we explore the possibility that the standard model misses an important dimension that is instead reflected in the data.

Figure 2 reports daily fatalities in Lombardy and Veneto, nearby region with about half of the population (4.9 millions inhabitants versus about 10 millions in Lombardy), which witnessed a remarkably lower number of deaths per day. Figure 3 illustrates another interesting feature of the data from Lombardy and Veneto: the share of hospitalized patients in ICU has been much higher in Veneto than in Lombardy, hinting at the possibility of an important mismatch between the demand of ICU beds and their supply in Lombardy.

**INSERT FIGURES 2-3 HERE**

The mismatch between data and model prediction in Lombardy and the heterogenous pattern of mortality observed in Lombardy and Veneto can be related to a specific feature of the SEIR model: mortality is exogenously given by a constant parameter. In the SEIR specification Infectious patients are divided in three groups, Mild, Severe and Fatal,

and the destiny of each patient is written at the time their are exposed to the disease. After some heterogenous duration, Mild and Severe patients inevitably recover, while fatal patients inevitably die. What if what happened in Lombardy can be described as follows: infectious patients are still divided in the three standard groups, but the Severe do not inevitably recover. In fact, some of them need assistance in Intensive Care Unit and if there are no available ICU positions, then their status changes from Severe to Fatal . There is therefore a time-varying endogenous component of mortality that cannot be captured by the standard SEIR model.

Note that Ferguson et al. (2020) when providing estimates of the simulated effect on US and UK of an unmitigated epidemic featuring a basic reproduction number $R_o$ of 2.4 clearly state "... In total, in an unmitigated epidemic, we would predict approximately 510,000 deaths in GB and 2.2 million in the US, **not accounting for the potential negative effects of health systems being overwhelmed on mortality**...."

This paper extends the SEIR model to a SEIHCR that follows patients in their pattern of Hospitalization and endogeneize the possible Constraint of ICU availability and its impact on lethality.

The model is calibrated to the data from Lombardy to illustrate that it is capable to replicate the observed pattern of lethality in Lombardy with a parametrization fully in line with the international evidence.

## 2 The SEIHCR Model: Description

The SEIHCR model is a system of differential equations for the dynamics of a virus across different groups of the population.

The exact specification of the equations is reported in the Appendix, Figure 4 reports the Dependency Graph of the Model, while this section describes its structure and fundamental elements.

**INSERT FIGURE 4 HERE**

The model allows to simulate the dynamics of the virus diffusion starting from an initial period in which the total **Population** ($N_t$) of N individuals is divided in 1 **Infectious** ($I_t$) and N-1 **Susceptible** ($S_t$)**.** In each period (day) some Susceptible become **Exposed** ($E_t$)**,** their number is determined by the basic reproduction number $R_0$,that determines the number of secondary infections each infected individual produces, by the probability with which Susceptible meets Infectious, $\left(\frac{I_{t-1}}{N_{t-1}}\right)$, and by the average duration of the period in which a patient is infectious

$T_{\text{inf}}$. Exposed after an incubation period of length $T_{inc}$, become Infectious. The outflows from Susceptible is the inflows into Exposed in each period, and the outflows from Exposed is the inflows into Infectious. Infectious falls into three groups: those with mild symptoms ($MILD_t$), those with severe symptoms ($SEV_t$), and those with fatal symptoms ($FAT_t$). The allocation to these groups is controlled by three probabilities: $\left(1-p^{sev}-p^{fat}\right)$, $p^{sev}$, $p^{fat}$. Patients with mild symptoms recover after a recovery period, $T_{srec}$. Patients with severe and fatal are **Hospitalized.** All patients with symptoms that require hospitalization, are hospitalized after a period of average duration $T_{shosp}$, some hospitalized patients require intensive care unit with probability $p^{ic}$. Patient with fatal symptoms succumb notwithstanding hospitalization, even in intensive care, after the mean duration from the onset of symptoms to death, $T_{sd}$. Patient with severe symptoms either recover or become fatal. The recovered, with a mean duration of from the onset of symptoms to hospital discharge of $T_{shd}$ , are those who do need intensive care unit and those who need intensive care unit and find a place. The patients with severe symptoms that need ICU and do not find availability are **Constrained** and become fatalities. At the end of each period the population decreases because of the fatalities, while the stock of recovered grows as a consequence from the new additions of recovered with mild and severe symptoms.

After calibration, we shall compare model simulated data with observed data from Lombardy to assess the potential of the model and its explanation of the mortality in Lombardy.

.

# 3 The SEIHCR Model: a Calibration to data from Lombardy

Model simulation requires numerical values for all the relevant parameters. Given the availability of a sample of sufficient size of reliable data, parameters can be estimated (see Cereda et al.(2020)). The daily data made available on Lombardy by Protezione Civile cover a short sample of about forty observations and are affected by a change in regime. In fact, on March 8 2020 a full lockdown was legislated for the region and the entire country.

Moreover, estimation of the model in a conventional sense, that is, deriving parameter values by fitting equations using time series data, is not possible because there are no data on Exposed, Infectious and Recovered. In fact, there are no patients are not tracked and those with Severe symptoms are recorded as positive only if they are tested and with a lag. This lag depends on the duration of the time elapsed between

symptoms and hospitalization, between hospitalization and testing, and between the moment in which the swab is taken and the results become available. Obtaining estimates of the Infectious would require ongoing random testing of the population, which has not happened so far in Lombardy. Similarly, the Recovered are underestimated because the only Recovered observed are those Recovered from hospital. In principle estimates of the total recovered could be obtained by ongoing random sampling of tests for serum antibodies in response to the coronavirus, however such tests are currently under development but not yet available.

Measurement error in the data has strong implications for the estimation of the crucial model parameters and for the design of optimal policies for them (see Stock(2020)).

These considerations led to the design and implementation of a calibration strategy based on the exploitation of the data on hospitalization, that we regard as the observed variable with the closest distance from the theoretical variables in the model.

Our procedure allowed to select the dating of the lockdown in the sample of simulated data and to estimate the impact of the lockdown on the basic reproduction number $R_0$. $R_0$ was set initially at 2.2 in line with the international evidence, reflected in the baseline parameterization in the epidemic calculator available online (https://gabgoh.github.io/COVID/index.html).

Also all the other parameters that determines the transmission dynamics and the clinical dynamics were chosen in line with the international evidence. The calibration of these parameters is summarized in Table 1.

**INSERT TABLE 1 HERE**

To set the value of $R_0$ after the intervention the model was simulated first in a pre-lockdown scenario, when the capacity constraint in terms of ICU was still irrelevant, with an initial population of 10 millions, and all duration parameters set in line with the international evidence. The lockdown was then dated in the model simulated data to match the number of hospitalized patients observed on the March 8 2020. This procedure dates March 8th as day 95 of our 730 (2 years) of simulated data. Having dated the lockdown, the post lockdown $R_0$ was calibrated to match the observed number of hospitalized patients two weeks after the lockdown. This procedure delivered a $R_0$ post lockdown of 0.95.

Finally, the probability with which of an hospitalized patient needs intensive care was calibrated at 0.2. As Figure 2-3 suggest, this is the value around which the share of hospitalized patients in ICU stabilizes over time in Veneto, the nearby region that recorded a much lower daily deaths than Lombardy.

## 4 The SEIHCR Model: Simulations

The model has been simulated under two scenarios. In the baseline scenario the IC capacity is set to 400 beds before the lockdown (the maximum of the observed utilization) and to the observed number of occupied ICU beds reported by Protezione civile that grows constantly to reach a total size of 1500, with a batch of 200 new units released simultaneously toward the end of the sample (this jump in the series is generated by the availability of the new Fiera Hospital in Milan). In the alternative scenario the capacity constraint has been removed, making counterfactually available to Lombardy a number of ICU beds equal to 20 per cent of hospitalized patients.

The most interesting results from the model extension is reported in Figure 5. Figure 5 clearly illustrates that the capacity constraint is essential to replicate the pattern of mortality observed in Lombardia. The model with the ICU constraint imposed generates a number of simulated daily deaths very close to the number of observed daily deaths in Lombardy. When the constraint is counterfactually removed, the simulated data provide a very close match for the pattern of daily deaths observed in Veneto and are nowhere near to the daily deaths observed in Lombardy.

**INSERT FIGURE 5 HERE**

The results of these simulations point to a crucial role for the expansion of ICU capacity to save lives in Lombardy and to the importance of the NPI policy implemented in Venteto. These policies were capable of keeping the diffusion of the disease under control and therefore the hospitalization rate much lower than that observed in Lombardy .

Figure 6 reports the pattern of model simulated total recovered, model simulated hospital recovered and tracked patients recovered (guariti).

**INSERT FIGURE 6 HERE**

The model based variables that tracks well the observed recovered patient is the patients recovered from hospital while the effective number of recoveries is much higher because of the relevance of patients with mild symptoms which were not tested. However, the model simulated number of total recovered patients at the end of May 2020 is of about half a million, which is five per cent of the total population in Lombardy.

Finally Figure 7.1 and 7.2 report the pattern of model based exposed and observed exposed (total cases- death-fatalities), looking at their level and their daily changes.

**INSERT FIGURE 7.1-7.2 HERE**

The figure shows that the observed daily exposed are also in line with the model prediction. In fact, while Susceptible becomes Exposed instantaneously in the model, their observation in the data requires testing which is implemented only some time after hospitalization. The model based pattern of the change in Exposed, which again is followed with a lag by the actual data, might be of help in designing optimal interpolant for reduced form data based prediction of the dynamics of the virus.(Peracchi(2020))

## 5 Conclusions and Policy Implications

A SEIHCR Model used for simulation of the fatality of the COVID disease in Lombardy is capable of explaining the high mortality rate observed in this region.

The main innovation of the model is the endogeneization of the fatality rate, that becomes higher than the CFR of 0.0138 when the demand of ICU beds exceeds their available supply.

Data simulated from the model with endogenous mortality can explain the high number of deaths observed in Lombardy and the striking difference in observed COVID mortality bewteen Lombardy and Veneto.

The impact of different NPI approaches to reduce hospitalization is amplified when mortality increases because of excess demand of ICU beds.

The cost of not implementing NPI interventions aimed at "flattening the curve" becomes higher when the curve becomes "more humpy" as a consequence of the strain on healthcare and Intensive Unit Care capacity.

The model also shows that the number of actual recovered patients in Lombardy is much higher than the observed number of patients recovered from hospitalization. However, the estimate of model simulated recovered subjects by the end of May 2020 stands only half a million individuals (under a parameterization in which the lockdown has brought $R_0$ to 0.95).

The importance of the ICU constraint in increasing fatality should be taken into account in the design of exit strategies from the lockdown.

NPI measures that have brought $R_0$ close to one have greatly relieved the strain of healthcare capacity. In their removal strategy the benefits of a growing proportion of the immune indivduals in the population should be carefully weighted against the risk that the number of new infections saturates again healthcare capacity.

## 6 Appendix: The SEIRHC Model Specification

We report in this appendix the full model specification equation by equation. The model is made of 16 equations. It allows to simulate the dynamics of the virus diffusion starting from an initial period in which the total **Population** ($N_t$) of N individuals is divided in 1 **Infectious** ($I_t$) and N-1 **Susceptible** ($S_t$)**.** In each period (day) some Susceptible become **Exposed** ($E_t$)**,** their number is determined by the basic reproduction number $R_0$,that determines the number of secondary infections each infected individual produces, by the probability with which Susceptible meets Infectious, $\left(\frac{I_{t-1}}{N_{t-1}}\right)$, and by the average duration of the period in which a patient is infectious $T_{\text{inf}}$. Exposed after an incubation period of length $T_{inc}$, become Infectious. The outflows from Susceptible is the inflows into Exposed in each period, and the outflows from Exposed is the inflows into Infectious. Infectious falls into three groups: those with mild symptoms ($MILD_t$), those with severe symptoms ($SEV_t$), and those with fatal symptoms ($FAT_t$). The allocation to these groups is controlled by three probabilities: $\left(1-p^{sev}-p^{fat}\right)$, $p^{sev}$, $p^{fat}$. Patients with mild symptoms recover after a recovery period, $T_{srec}$. The daily change in **Mild** patients stock is determined by the share $\left(1-p^{sev}-p^{fat}\right)$ of the outflows from Infectious, and the outflows from the share of Mild who recover that depends on the average duration form symptoms to recovery for mild patients, $T_{srec}$ . Patients with severe and fatal symptoms require hospitalization, both these group are hospitalized after a period between developing symptoms and hospitalization of average duration $T_{shosp}$, hospitalized. The daily change in **Fatal** patients is determined by the share $p^{fat}$ of the outflows from Infectious and the outflows by the share of Fatal who are hospitalized. Patient with fatal symptoms succumb notwithstanding hospitalization, even in intensive care, after the mean duration from the onset of symptoms to death, $T_{sd}$. The daily change in **Severe** patients is determined by the share $p^{sev}$ of the outflows from Infectious and the outflows by the share of Severe who are hospitalized. Patients in hospital, independently from their initial status, require intensive care $p^{sev}$ unit with probability $p^{ic}$. Patient with severe symptoms either recover or become fatal. The recovered, with a mean duration of from the onset of symptoms to hospital discharge of $T_{shd}$ , are those who do need intensive care unit and those who need intensive care unit and find a place. The patients with severe symptoms that need ICU and do not find availability become fatal. At the end of each period the population decreases because of the fatalities, while the stock of recovered grows as a consequence from the new additions of recovered with mild and severe symptoms.

$$
\begin{aligned}
\Delta S_t &= \left(-\frac{R_0}{T_{\inf}}\frac{I_{t-1}}{N_{t-1}}\right) S_{t-1} \\
\Delta E_t &= \left(\frac{R_0}{T_{\inf}}\frac{I_{t-1}}{N_{t-1}}\right) S_{t-1} - \left(\frac{1}{T_{inc}}\right) E_{t-1} \\
\Delta I_t &= \left(\frac{1}{T_{inc}}\right) E_{t-1} - \left(\frac{1}{T_{\inf}}\right) I_{t-1} \\
\Delta MILD_t &= p^{mild}\left(\frac{1}{T_{\inf}}\right) I_{t-1} - \left(\frac{1}{T_{srec}}\right) MILD_{t-1} \\
\Delta REC\_MILD_t &= \left(\frac{1}{T_{srec}}\right) MILD_{t-1} \\
\Delta SEV_t &= p^{sev}\left(\frac{1}{T_{\inf}}\right) I_{t-1} - \left(\frac{1}{T_{shosp}}\right) SEV_{t-1} \\
\Delta SEV\_H_t &= \left(\frac{1}{T_{shosp}}\right) SEV_{t-1} - \left(\frac{1}{T_{shd} - T_{\inf} - T_{shosp}}\right) SEV\_H_{t-1} - \Delta SEV\_FAT_t \\
\Delta SEV\_FAT_t &= I_t^{ICCS}\left(S_t\right)\left(p^{ic} SEV\_H_t - \left(ICC_t - p^{ic} FAT\_H\right)\right) \frac{SEV\_H_t}{HOSPITALIZED_t} \\
\Delta REC\_SEV_t &= \left(\frac{1}{T_{shd} - T_{\inf} - T_{shosp}}\right) SEV\_H_{t-1} \\
\Delta FAT_t &= p^{fat}\left(\frac{1}{T_{\inf}}\right) I_{t-1} - \left(\frac{1}{T_{shosp}}\right) FAT_{t-1} \\
\Delta FAT\_H_t &= \left(\frac{1}{T_{shosp}}\right) FAT_{t-1} - \left(\frac{1}{T_{sd} - T_{\inf} - T_{shosp}}\right) FAT\_H_{t-1} \\
\Delta REM\_FAT_t &= \left(\frac{1}{T_{sd} - T_{\inf} - T_{shosp}}\right) FAT\_H_{t-1} \\
HOSPITALIZED_t &= SEV\_H_t + FAT\_H_t \\
EFF\_FAT_t &= REM\_FAT_t + SEV\_FAT_t \\
RECOVERED_t &= REC\_MILD_t + REC\_SEV_t \\
\Delta N_t &= -\Delta EFF\_FAT_t \\
I_t^{ICCS}\left(ED_t\right) &= \left\{ \begin{array}{l} 1 \quad if \quad ED_t > 0 \\ 0, \quad otherwise \end{array} \right\} \\
ED_t &= p^{ic} SEV\_H_t + p^{ic} FAT\_H_t - ICC_t
\end{aligned}
$$

# 7 Table and Figures

Table1: Calibration of the SEIRHC model for Lombardy

| Parameter | Pre-lockdown | Post-Lockdown | Source PreL | Source PostL |
|---|---|---|---|---|
| $R_0$ | 2.2 | 0.95 | EC | Calibrated[1] |
| $T_{\text{inf}}$ | 2.9 days | 2.9 days | EC | EC |
| $T_{inc}$ | 5.2 days | 5.2 days | EC | EC |
| $T_{srec}$ | 11.1 days | 11.1 days | EC | EC |
| $T_{shosp}$ | 5 days | 5 days | EC | EC |
| $T_{sd}$ | 17.8 days | 17.8 days | Verity et al. | Verity et al. |
| $T_{shd}$ | 22.6 days | 22.6 days | Verity et al. | Verity et al. |
| $p^{fat}$ | 0.0138 | 0.0138 | Verity et al. | Verity et al. |
| $p^{sev}$ | 0.1 | 0.1 | IE | IE |
| $p^{ic}$ | 0.2 | 0.2 | PC | PC |

EC (Epidemic Calculator): https://gabgoh.github.io/COVID/index.html

IC (International Evidence) https://www.worldometers.info/coronavirus/#countries

PC (Protezione Civile data on Lombardy) https://github.com/pcm-dpc/COVID-19

[1]The value is chosen to generate a match between model simulated hospitalization and observed hospitalization in Lombardy two weeks after the lockdown date.

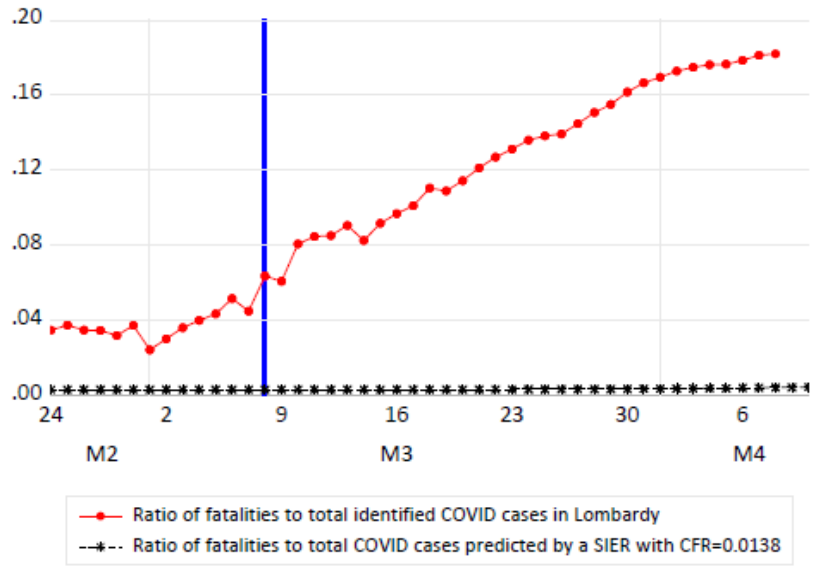


Figure 1: COVID observed fatality rate and SEIR model simulated fatality rate (daily data Febraury 24th-April 8th 2020)

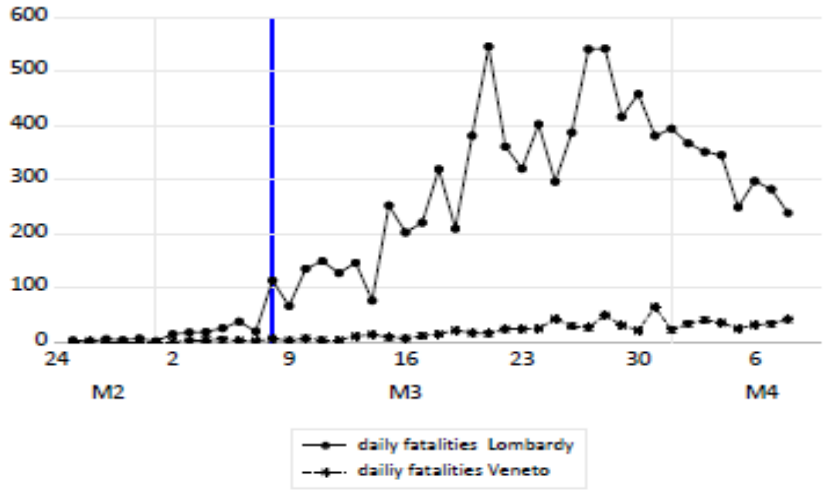


Figure 2: COVID in Lombardy and Veneto. Daily Fatalities

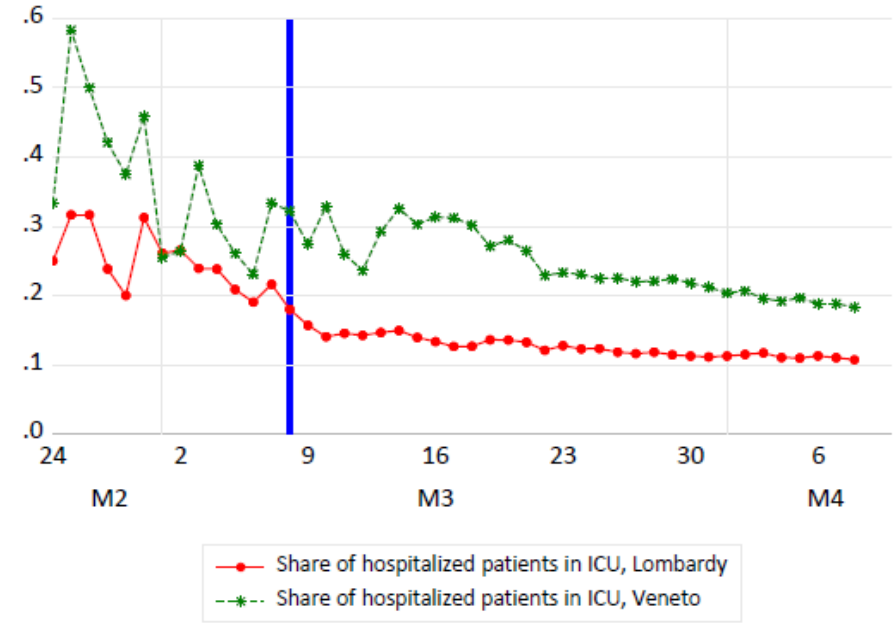


Figure 3: COVID in Lombardy and Veneto. Share of total COVID hospitalized patients in ICU

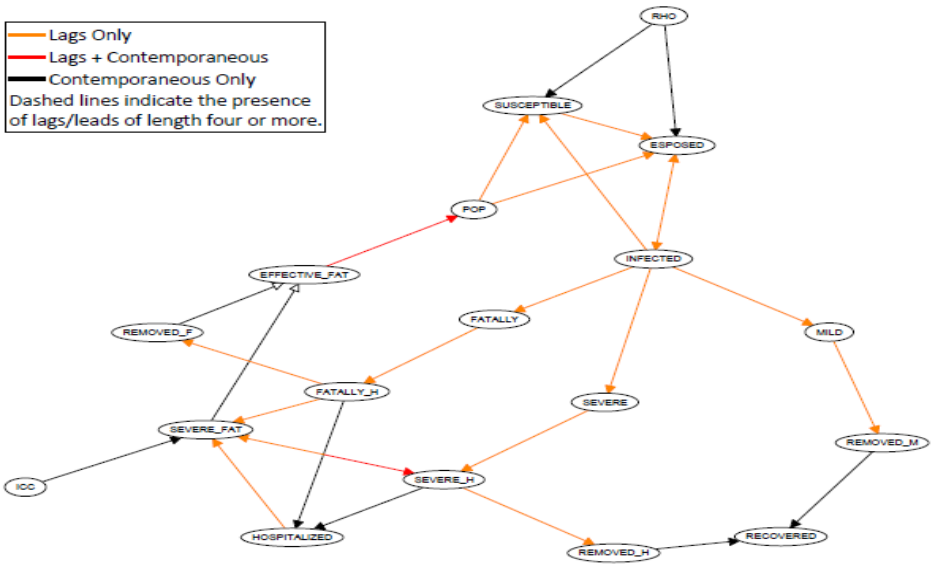


Figure 4: The SEIRHC model dependency graph

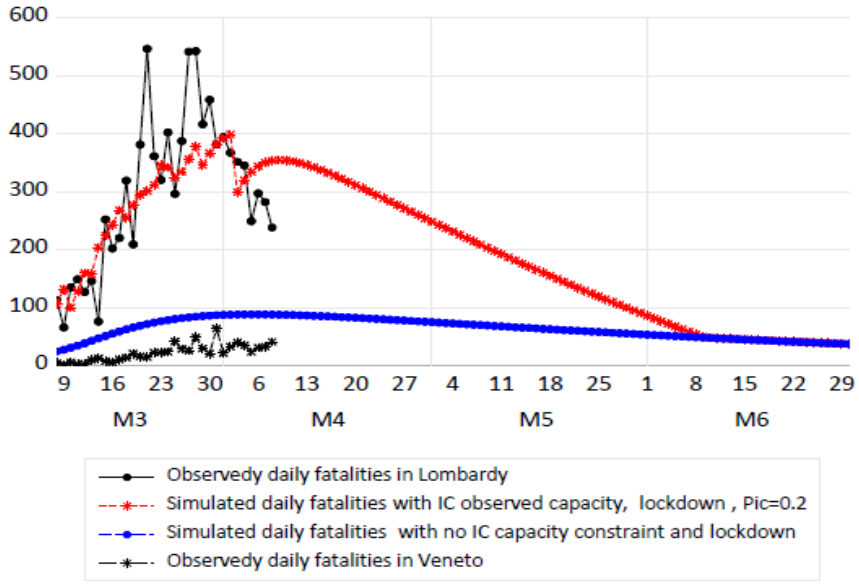


Figure 5: Actual and Simulated Fatalities in Lombardy and Veneto

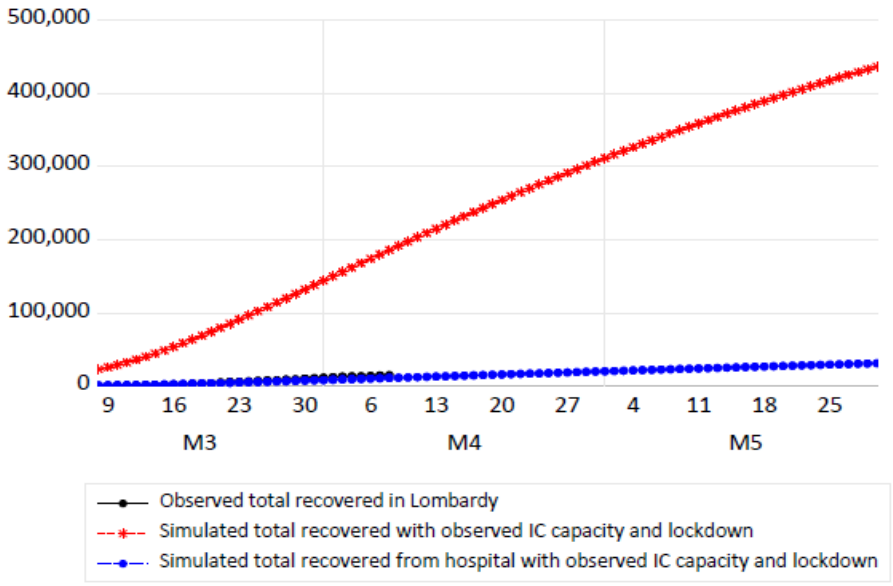


Figure 6: Model Simulated Recovered patients and observed recovered (guariti)

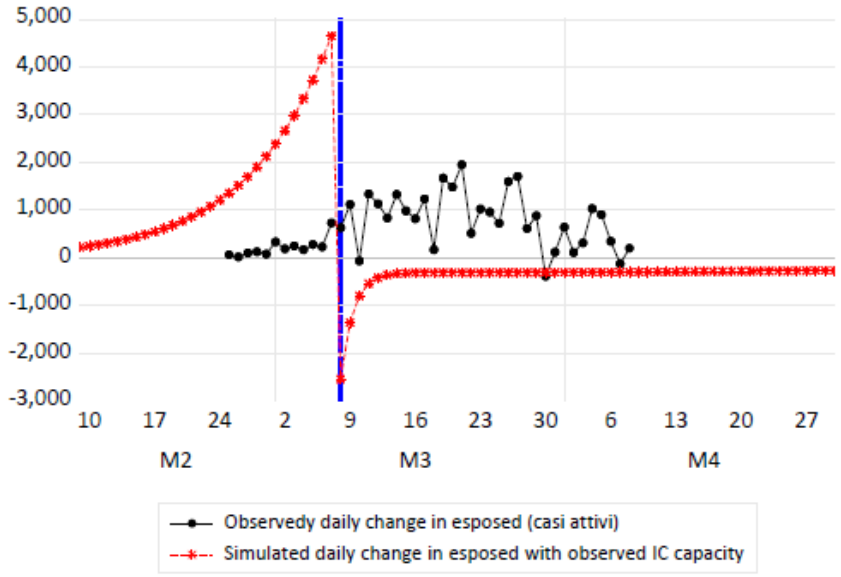


Figure 7.1: model simulated and observed exposed

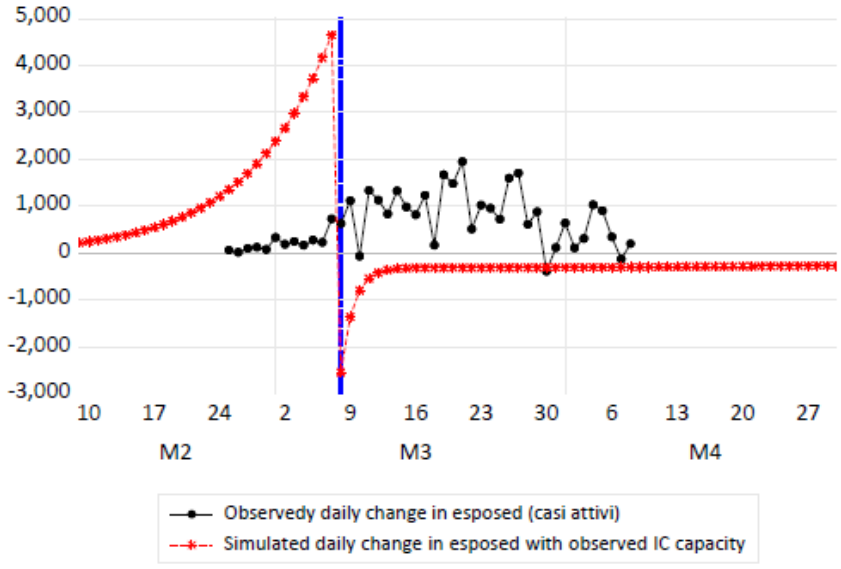


Figure 7.2: Model Simulated and Observed Change in Esposed

## 8 References

Allen, L.J.S. (2017). "A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis." Infectious Disease Modeling 2(2), 128-142.

Cereda et al.(2020) "The early phase of COVID-19 outbreak in Lombardy", mimeo

Ferguson et a.(2020) "Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand",

https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf

Goh Gabriel, the Epidemic Calculator, https://gabgoh.github.io/COVID/index.html

Kucharski et al. (2020) Analysis and projections of transmission dynamics of nCoV in Wuhan, https://cmmid.github.io/topics/covid19/current-patterns-transmission/wuhan-early-dynamics.html

Peracchi F.(2020), The Covid-19 pandemic in Italy, mimeo EIEF Rome

Stock J. (2020) Coronavirus Data Gaps and the Policy Response to the Novel Coronavirus, Harvard Mimeo

Verity et a.(2020), Estimates of the Severity of the COVID-19 disease, https://www.medrxiv.org/content/10.1101/2020.03.09.20033357v1

Wu., Leung and Leung (2020) Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study, VOLUME 395, ISSUE 10225, P689-697, The Lancet, https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30260-9/fulltext