Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company’s public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Brief paper

Closed-form expressions and nonparametric estimation of COVID-19 infection rate

Mauro Bisiacco\textsuperscript{a}, Gianluigi Pillonetto\textsuperscript{a,\textasterm}, Claudio Cobelli \textsuperscript{a,b}

\textsuperscript{a} Department of Information Engineering, University of Padova, Padova, Italy
\textsuperscript{b} Consiglio Superiore di Sanità, Italian Ministry of Health, Italy

A R T I C L E   I N F O

Article history:
Received 18 February 2021
Received in revised form 5 August 2021
Accepted 13 January 2022
Available online 2 April 2022

Keywords:
Compartmental models
COVID-19
Infection rate
System identification
Nonparametric estimation
Epidemic spread

A B S T R A C T

Quantitative assessment of the infection rate of a virus is key to monitor the evolution of an epidemic. However, such variable is not accessible to direct measurement and its estimation requires the solution of a difficult inverse problem. In particular, being the result not only of biological but also of social factors, the transmission dynamics can vary significantly in time. This makes questionable the use of parametric models which could be unable to capture their full complexity. In this paper we exploit compartmental models which include important COVID-19 peculiarities (like the presence of asymptomatic individuals) and allow the infection rate to assume any continuous-time profile. We show that these models are universal, i.e. capable to reproduce exactly any epidemic evolution, and extract from them closed-form expressions of the infection rate time-course. Building upon such expressions, we then design a regularized estimator able to reconstruct COVID-19 transmission dynamics in continuous-time. Using real data collected in Italy, our technique proves to be an useful tool to monitor COVID-19 transmission dynamics and to predict and assess the effect of lockdown restrictions.

© 2022 Elsevier Ltd. All rights reserved.

1. Introduction

Hundreds of countries and territories around the world have been affected by the diffusion of COVID-19 (Velavan & Meyer, 2020; Wittkowski, 2020). Even if many efforts are now devoted to the administration of a vaccine, important tools to contain the pandemic, also in view of possible variants of the virus, are social distancing measures coupled with the use of precautions like masks, massive testing and tracing approach, or more severe restrictions like lockdown’s setting (Lavezzo, Franchin, Giavarella, & al., 2020). A better understanding of COVID-19 dynamics would be fundamental to monitor, and possibly increase, the effectiveness of such actions. Estimation of lockdown’s effect on people’s behavior is in fact key to inform health-care decisions on emergency management.

The knowledge of the infection rate of a virus is key to obtain the goal described above. It permits to monitor the epidemic evolution by assessing the transmissibility and contagiousness of pathogens agents. Unfortunately, such epidemiological variable is not accessible to direct measurement and its estimation is far from trivial. In fact, the infection rate depends not only on the biological characteristics of the virus but also on all those factors which influence the human behavior and the contact rate among people. Hence, it can change significantly in time due to social organization, risk perception, use of masks, seasonality (people tend to use less precautions during the summer season and to stay mostly in enclosed spaces when coming back to work).

An important class of models to describe COVID-19 dynamics relies on the so called compartmental models where the population is assumed to be well-mixed and divided into categories. The most popular description is the SIR which includes three compartments containing susceptible (S), infected (I) and removed (R) individuals (Kernack & McEndrick, 1927). Since its inception, many SIR variants have been then proposed in the literature to describe even more complex dynamics, like e.g. the SEIR model where the class of exposed (E), i.e. people who are host for infectious but cannot yet transmit the disease, is also included (Bootsma & Ferguson, 2007; Capasso & Serio, 2007; Korobeinikov & Maini, 2005; Liu, Hethcote, & Levin, 1987). Additional differential equations have been included also to describe how people can react to knowledge of infections and risk of death as well as to account for the increasing of vaccination rate and its consequence on the susceptible fraction of individuals (Buonomo, d’Onofrio, & Lacitignola, 2008; Funk, Gilad, Watkins, & Jansen, 2009; Kiss, Cassell, Recker, & Simon, 2010; Samanta

https://doi.org/10.1016/j.automatica.2022.110265
0005-1098/© 2022 Elsevier Ltd. All rights reserved.
and Sledge (2020), Gatto et al. (2020), Lavezzo et al. (2020) to Sontag, 2021; Wang, Chen, & Quin, 2020).

virus (Bi & Beck, 2021; Lavezzo et al., 2020; Sadeghi, Greene, & Chattopadhyay, 2014; Yu, Lin, Chiu, & He, 2017). For what M. Bisiacco, G. Pillonetto and C. Cobelli Automatica 140 (2022) 110265

so that values smaller than one imply that the total number of infected is decreasing.

In Appendix A we will also study an even more complex variant of this model where two infection rates \( a_1(t) \), \( a_2(t) \) describe separately the interaction between \( S(t) \) and \( I_1(t), I_2(t) \), respectively.

Finally, the following measurement equation is introduced

\[ y(t) = \frac{I(t)}{H}. \]

where \( H \) is an unknown scalar. This parameter is necessary since the number of infected is never perfectly known during an epidemic. The observable measurements \( y(t) \) instead represent epidemiological measurements like e.g. the number of hospitalized or diagnosed infected. Data regarding the diagnosed infected have some limitations. They do not give an accurate information on how many subjects were affected by the virus exactly at a certain day, due to delays in the swabs processing. In addition, the amount of performed swabs and the criteria used to select people who are tested may vary in time. The number of hospitalized people and, even more, of patients in critical care, appear more reliable and informative on COVID-19 dynamics. For this reason, in our model the multiplier \( H \) formalizes the assumption, which appears statistically reasonable, that the number of true infected \( I(t) \) is proportional to the number of people in intensive care. Reconsidering (3), this assumption implies that \( y(t) < 1 \) that the number of people in intensive care is closing at instant \( t \).

3. Universal models and infection rate estimator

3.1. Universal models and infection rate closed-form expressions

We summarize the main findings (described in detail in Appendix A) which are instrumental for our future developments:

- the models enriched with a nonparametric description of the infection rate are extremely expressive. In fact, also the apparently simple time-varying SIR turns out to be universal. Borrowing the term universal from the machine learning literature (Hastie, Tibshirani, & Friedman, 2001; Micchelli, Xu, & Zhang, 2006), this means that, given any smooth and positive function \( y(t) \), there exists a time-varying SIR whose output is \( y(t) \);
- using the SAIR model, the infection rate admits the following closed-form expression:

\[
\dot{y}(t) = \frac{y(t) + b_2 y(t) + c - b_2}{b_1 + b_2} e^{-b_1 + c + \tau} + c - b_2 \int_0^t \frac{e^{-b_1 + c + \tau - \tau} y(\tau) d\tau}{H(1 + c)}
\]

where

\[
F(t) = y(t) - y(0)
\]

After summing up the second and third equation, one can also introduce

\[
\gamma_2(t) = \frac{a(t) S(t) (I_1(t) + I_2(t)) + b_1 I_1(t)}{(b_1 + b_2) I_1(t) + c I_2(t)}
\]

and values smaller than one now indicate that the total number of infected is decreasing.

established by Berti, Franco, Mohler, Short, and Sledge (2020), Gatto et al. (2020), Lavezzo et al. (2020), to better capture the complexity of the problem, in this work we use a SAIR model where (along the line of Pil-lonetto, Bisiacco, Palù, and Cobelli (2021) which considers the SIR model) \( a(t) \) can assume any temporal profile. Well-posedness of the problem will be then restored through a nonparametric regularized strategy which incorporates also knowledge on COVID-19 dynamics coming from a recent seroprevalence study performed in Italy.

The paper is organized as follows. In Section 2 we introduce the SAIR model of COVID-19 dynamics. In Section 3 we report some important properties of such models and describe our infection rate estimator. In Section 4 case studies regarding the Italian scenario are illustrated. Conclusions then end the paper while all the mathematical details are gathered in Appendix A.

2. SAIR model and nonparametric description of COVID-19 infection rate

The classical SIR model includes the classes \( S(t) \), \( I(t) \) and \( R(t) \) which evolve in time and contain, respectively, susceptible, infected and removed people. They are normalized, so that their sum is equal to one for any \( t \). In the SAIR, the class \( I(t) \) contains two kinds of infected people. The first ones are asymptomatic/paucisymptomatic who can directly recover with a rate \( b_2 \). The other ones move to the second class of infected \( I_2(t) \) with a rate \( b_1 \) and then recover with a rate \( c \). This leads to the following SAIR model

\[
\dot{S}(t) = -a(t) S(t) (I_1(t) + I_2(t))
\]

\[
\dot{I}_1(t) = a(t) S(t) (I_1(t) + I_2(t)) - (b_1 + b_2) I_1(t)
\]

\[
\dot{I}_2(t) = b_1 I_1(t) - c I_2(t)
\]

\[
\dot{R}(t) = b_2 I_2(t) + c I_2(t)
\]

where \( a(t) \) now describes the temporal evolution of the interaction between susceptible \( S(t) \) and the two classes \( I_1(t), I_2(t) \) of infected. For models including several classes of infected, a scalar reproduction number can still be defined in terms of the eigenvalue of a matrix which includes all their effects (van den Driessche & Watmough, 2002). In this work, we instead introduce different functions to monitor the evolution of different classes of infected. Such functions can be seen as generalized time-varying reproduction numbers. In particular, from the second differential equation one can define

\[
\gamma_1(t) = \frac{a(t) S(t) (I_1(t) + I_2(t))}{(b_1 + b_2) I_1(t)}
\]

so that values smaller than one imply that \( I_1(t) \), the class which feeds both \( I_2(t) \) and \( R(t) \), decreases over time. This thus gives crucial information e.g. to assess if restraints are effective, predicting that people in intensive care will be soon under control.
3.2. Infection rate estimator

The infection rate has to be reconstructed from real epidemiological data. For this purpose the SAIR will be adopted exploiting (4). However, since our model is universal, model complexity control is a fundamental issue.

First, the output has to be regularized. In fact, (4) requires the knowledge of the entire continuous-time system output but actually \( y(t) \) is known only in noisy and sampled form, over \( t = 1, 2, \ldots \) using day as time unit. The measurement noise affecting the intensive care data here describes both possible delays/mistakes in reporting the number of people in intensive care and inevitable modeling errors. Interestingly, as in the SIR case, (4) shows that when adopting the more complex SAIR model the infection rate depends only on \( y \) and \( y' \) (ill-conditioning increases with the derivative order). A smooth reconstruction of these two signals can then be obtained by a spline estimator (Bottega & Pillonetto, 2018; Wahba, 1990) implemented e.g. by a Kalman smoother (Aravkin, Burke, Ljung, Lozano, & Pillonetto, 2017) (or a Kalman filter to obtain an infection rate online estimator) where the regularization parameter is estimated from data using marginal likelihood optimization (Aravkin, Burke, Chiuso, & Pillonetto, 2012; Bell & Pillonetto, 2004; Pillonetto, Dinuzzo, Chen, De Nicolao, & Ljung, 2014; Pillonetto; Saccomani, 2006). This defines the first step of our estimator, as graphically illustrated in the left part of Fig. 1. As case studies, Fig. 2 displays the estimates of \( y(t) \) (left) and \( y'(t) \) (right) in Lombardy (top) and Italy (bottom).

Model complexity is regulated by introducing uncertainty intervals on the system parameters \( b_1, b_2, c \) and \( H \) supported from the literature (Gatto et al., 2020; Giordano et al., 2020; Lavezzi et al., 2020; Worldometer, 2021). We assume \( b_1 \in [0.14, 0.34] \), which roughly means that infected people show the first symptoms in an interval ranging from three to seven days, and \( b_2 \in [0.04, 0.14] \), so that asymptomatic people heal in an interval ranging from one to three weeks. We also let \( c \in [0.02, 0.14] \) which implies that symptomatic people recover in an interval ranging from one week to almost two months. The constraint \( c \leq b_2 \leq b_1 \) is also adopted.

For what regards the multiplier \( H \), which defines how many people are infected for each individual present in intensive care, its interval is set to [50, 200]. The lower bound 50 follows from simple considerations on the time-courses of diagnosed infected and people in intensive unit in Lombardy and Italy. The value \( H = 200 \) means that almost 3% of people were infected in Lombardy when the number of people in intensive care reached the peak of the first infection wave. In addition, with \( H = 200 \) and using values of \( b_1, b_2, c \) in the intervals specified above, the time-varying SAIR predicts a total (cumulative) number of infected which can reach 15%. Hence, the upper bound for \( H \) is robust since recent studies have assessed the presence of antibodies against SARS-CoV-2 in 7.5% of the population in Lombardy and 2.5% in Italy, e.g. see La Repubblica (Italian newspaper) (2020).

Then, as illustrated in the right part of Fig. 1, the inputs to the last block of our estimator are the smoothed profiles of \( y(t), y'(t) \) and the uncertainty intervals for \( b_1, b_2, c, H \). The closed-form expression in (4) is finally used to compute for any \( t \) the minimum and maximum value of \( a(t) \) compatible with the values which the model parameters can assume, obtaining lower and upper bounds of the infection rate. Hence, bounds around the reproduction numbers displayed in (2) and (3) become also immediately available.

4. Case studies: infection rate estimation in Lombardy and Italy

To test our procedure, we consider the Italian scenario, focusing in particular on the Lombardy region, the most affected by the outbreak. Results regarding the infection rate are in Fig. 3. It is worth noting that Italy has been the first country in Europe to set nationwide restrictions by introducing the lockdown on the whole territory on March 9, 2020. Restrictions have then been first further reinforced and, then, gradually relaxed. A second wave of infection affected the country after the summer season and forced the authorities to resume new restrictions in the whole territory in October. In particular, new restrictions were set around the mid of October, 2020, with also a new lockdown (milder than the first one) introduced in some regions including Lombardy on November 5, 2020.

The infection rate profiles displayed in Fig. 3 appear realistic and informative. Before the lockdown (interval \([1, 9]\) on the \(x\)-axis) the infection rate value was around 0.5 both in Lombardy and Italy. Then, \( a(t) \) quickly decreased: this describes the rapid change in people’s behavior near and after the beginning of the lockdown. The curves then kept decreasing, a bit slower in Lombardy. The first days of April (after 30 on the \(x\)-axis) the transmission rate was below 0.1 and the epidemic was under control, reaching its lowest value a few days before the end of the lockdown. At the end of the restrictions, \( a(t) \) was inside the interval \([0.01, 0.06]\) both in Lombardy and in Italy.

After the end of the lockdown, on May, June and partly in July the situation was still under control, with the upper bound not exceeding 0.1. But the curve rapidly increased at the very end of July (153 on the \(x\)-axis), in particular in Lombardy. A peak was reached on August 29 in Italy, with the upper bound close to 0.2 and the lower bound larger than 0.1. In September the curve decreased, but in October the situation became critical again, especially in Lombardy where the upper bound assumed value 0.28 a few days before November. Then, the effect of the restrictions became evident. The infection rate decreased significantly in Lombardy and on the last day here considered, January 15, 2021, the upper bound decreased from 0.28 to 0.1. The imposed restrictions seemed more effective in Lombardy than in Italy, where the decrease is slower, and our estimator provides a precise quantification of this phenomenon.

Fig. 4 displays the bounds for the generalized reproduction number defined in (2). Its profile is similar to that followed by the infection rate. What is interesting to note is that during the summer season the maximum of \( \gamma(t) \) was obtained in August, around 1.25 in both Lombardy and Italy. In October the peaks reached the values 1.7 in Lombardy and 1.3 in Italy. Then, the restrictions have been effective, making \( \gamma(t) \) assume values smaller than the critical threshold 1. So, results suggest that the
number of infected $I_1(t)$ (which is the compartment containing asymptomatic people and feeding both $I_2(t)$ and $R(t)$) is currently decreasing. Our model thus suggests that the number of total infected $I(t) = I_1(t) + I_2(t)$ and, hence, the number of people in critical care, should also start decreasing. This is confirmed by Fig. 5 where the temporal profile of the reproduction number defined in (3) is displayed. The upper bound around $\gamma_2(t)$ became smaller than 1 on January, 2021, in both Lombardy and Italy.

5. Conclusions

We feel that this work represents a significant addition to the COVID-19 literature. Models like those described in Flaxman et al. (2020), Gatto et al. (2020), Giordano et al. (2020) have certainly a broader scope than recovering the infection rate. In principle their complexity can permit to describe more in depth system dynamics, e.g. spatial models can overcome the (approximate) homogeneity assumption underlying the compartmental
class here used. However, a limitation is that they exhibit a large number of parameters and their identification requires the introduction of significant and possibly delicate prior information. Our approach instead requires the user to specify just a few intervals where the SAIR parameters can assume their values. Also, identification of the currently used models of COVID-19 requires sophisticated numerical procedures, e.g., nonconvex optimization procedures which may undergo local minima or stochastic simulation techniques like MCMC which are powerful but may be difficult to implement, computationally demanding and subject to uncertain convergence (Gilks, Richardson, & Spiegelhalter, 1996).

Here we have shown that, also when asymptomatic people are included in the model, the infection rate admits a closed-form expression which depends only on the SAIR output and its first-order derivative. This allows to design an estimator which does not suffer of local minima and is extremely efficient. From the model

\[ \dot{S}(t) = -a(t) \mathcal{S}(t) \mathcal{I}(t), \quad \dot{I}(t) = a(t) \mathcal{S}(t) \mathcal{I}(t) - bI(t), \]

and

\[ y(t) = \frac{I(t)}{H}, \quad t \in [0, T] \]

which is in one-to-one correspondence with the infinite-dimensional parameter \( \theta = [H \ b \ I(0) \ S(0) \ a(t)]^T \), subject to the following constraints

\[ S(0) > 0, \ I(0) > 0, \ S(0) + I(0) \leq 1, \ b > 0, \ H > 0, \ a(t) > 0 \]

where \( a(\cdot) \in \mathcal{C}_0([0, T]) \) with \( \mathcal{C}_0([0, T]) \) to denote the space of continuous functions over the interval \([0, T]\). Here \( I(t), S(t) \) represent the positive state variables, while \( y(t) \) is the output here assumed to be perfectly known.

From \( S(t) + I(t) = -bI(t) < 0 \), it easily follows that \( I(t), S(t) \) evolve in the admissible region

\[ \mathcal{A} = \{ I(t) > 0, \ S(t) > 0, \ I(t) + S(t) \leq 1 \} \]

Appendix A

Universal models

To show that our model can perfectly fit any phenomenon connected with positive systems, we can just consider the simplest instance given by a SIR with the contact rate which can follow any (nonnegative) temporal profile. So, let us consider the SIR model

\[ \dot{S}(t) = -a(t) S(t) I(t), \quad \dot{I}(t) = a(t) S(t) I(t) - bI(t), \]

\[ y(t) = \frac{I(t)}{H}, \quad t \in [0, T] \]

which is in one-to-one correspondence with the infinite-dimensional parameter \( \theta = [H \ b \ I(0) \ S(0) \ a(t)]^T \), subject to the following constraints

\[ S(0) > 0, \ I(0) > 0, \ S(0) + I(0) \leq 1, \ b > 0, \ H > 0, \ a(t) > 0 \]

where \( a(\cdot) \in \mathcal{C}_0([0, T]) \) with \( \mathcal{C}_0([0, T]) \) to denote the space of continuous functions over the interval \([0, T]\). Here \( I(t), S(t) \) represent the positive state variables, while \( y(t) \) is the output here assumed to be perfectly known.

From \( S(t) + I(t) = -bI(t) < 0 \), it easily follows that \( I(t), S(t) \) evolve in the admissible region

\[ \mathcal{A} = \{ I(t) > 0, \ S(t) > 0, \ I(t) + S(t) \leq 1 \} \]
for any \( t \in [0, T] \) with \( l(t), S(t) \) which never assume null values since
\[
S(t) = S(0) e^{-\int_0^t a(t) y(t) \, dt} > 0, \quad l(t) = l(0) e^{\int_0^t a(t) \, dt} e^{-bt} > 0.
\]
This means that (6) represents a positive and stable system, where the usual meaning of infected and susceptible people fraction is preserved. Substitution of \( l(t) = H y(t) \) into the first and second equation leads to \( S(t) = S(0) e^{-H \int_0^t a(t) y(t) \, dt} \) and \( a(t) = \frac{b y(t) + \dot{y}(t)}{S(0) y(t)} \), which exploiting the previous relation, is equivalent to
\[
a(t) = \frac{b y(t) + \dot{y}(t)}{S(0) y(t)} e^{H \int_0^t a(t) y(t) \, dt}.
\] (8)

Hence, we need to investigate solvability of (8) in the whole \([0, T]\). This problem is far from trivial because such integral equation could exhibit divergent solutions. As we will see, the parameter \( H \) plays a crucial role in this respect. In fact, solvability can be ensured only for certain values of \( H \), as detailed in the following theorem.

**Theorem 1 (Solvability of (8)).** Given any \( S(0) \in (0, 1) \), there exist two real numbers \( b_0 > 0 \) and \( H_0 > 0 \) such that, for any \( b > b_0 \), (8) admits a positive continuous solution \( a(t) \) in \([0, T]\) for all \( H \) satisfying \( 0 < H < H_0 \).

**Proof.** Let \( b_0 = \max \left\{ -\min_{0 \leq t \leq T} \frac{\dot{y}(t)}{y(t)}, 0 \right\} \), so that \( b > b_0 \) implies the positiveness of both \( b \) and \( a(t) \). Consider the following complete metric space, equipped with the \( \ell_\infty \)-norm
\[
\mathcal{H}_R := \left\{ a \in \mathcal{C} \left[ 0, T \right], \text{ s.t. } a(t) \geq 0 \text{ and } ||a||_\infty \leq R \right\}
\]
From
\[
\frac{b y(t) + \dot{y}(t)}{S(0) y(t)} e^{H \int_0^t a(t) y(t) \, dt} \leq \frac{b}{S(0)} \left( 1 + \max_{0 \leq t \leq T} \frac{\dot{y}(t)}{y(t)} \right) e^{H \int_0^t a(t) y(t) \, dt} := \mu e^{H ||a||_\infty}
\]
for suitable \( \mu, \nu > 0 \) and defining
\[
\mathcal{F} : a \rightarrow c, \quad c(t) := \frac{b y(t) + \dot{y}(t)}{S(0) y(t)} e^{H \int_0^t a(t) y(t) \, dt};
\]
we obtain \( c(t) \geq 0 \) and \( ||c||_\infty \leq \mu e^{H R} \), if \( a \in \mathcal{H}_R \). If \( R := \mu e^R \) and \( H_1 := R^{-1} \), it follows that \( 0 < H < H_1 \) implies \( ||c||_\infty \leq \mu e^{-R} = R \), so that \( \mathcal{F} \) is a well-defined operator from \( \mathcal{H}_R \) into itself. Hence, (8) can be rewritten as
\[
a = \mathcal{F}(a), \quad a \in \mathcal{H}_R.
\]
Now, simple computations lead to
\[
||\mathcal{F}(a_1) - \mathcal{F}(a_2)||_\infty \leq RH \int_0^T |a_1(t) - a_2(t)| \, dt
\]
\[
\leq R ||a_1 - a_2||_\infty
\]
and, letting \( H_0 := H_1 \min(1, T^{-1}) \), one obtains that \( 0 < H < H_0 \) implies that \( a = \mathcal{F}(a) \), \( a \in \mathcal{H}_R \), satisfies, for a suitable \( \beta > 0 \), the following inequality
\[
||\mathcal{F}(a_1) - \mathcal{F}(a_2)||_\infty \leq \beta ||a_1 - a_2||_\infty, \quad \beta < 1.
\]
Exploiting the contraction theorem (Struthers & Potter, 2019) we obtain that a unique solution to the previous equation \( a \in \mathcal{H}_R \) exists and that it is positive everywhere.

Having established the solvability of (8), we can now rewrite it in terms of a differential equation. This will lead to the closed form solution for the \( a(t) \) contained in the SIR, as shown in the following theorem.

**Theorem 2 (Closed-form Expression of \( a(t) \) in the SIR).** If a solution of (8) exists, it is necessarily given by
\[
a(t) = \frac{b y(t) + \dot{y}(t)}{y(t) [S(0) - H b \int_0^t y(t) \, dt + y(t) - y(0)]}.
\] (9)

**Proof.** Letting \( w(t) := \int_0^t a(t) y(t) \, dt \) it follows that \( w(0) = 0, \quad \dot{w}(t) = a(t) y(t) \) and, hence, we obtain \( a(t) = \frac{b y(t) + \dot{y}(t)}{y(t) S(0) - H b \int_0^t y(t) \, dt + y(t) - y(0)} \). Substitution in (8) leads to
\[
\frac{b y(t) + \dot{y}(t)}{S(0)} e^{H w(t)}, \quad w(0) = 0, \quad a(t) = \frac{b y(t) + \dot{y}(t)}{y(t) S(0) - H b \int_0^t y(t) \, dt + y(t) - y(0)} e^{H w(t)}
\] (10)
where the first two equations define the solution (positive everywhere, except for \( t = 0 \)) of a differential equation, while the third one expresses \( a(t) > 0 \) in terms of this solution. Now, we provide the solution of (10). In fact, \( w(t) = g(t) e^{H w(t)} \), \( w(0) = 0 \) implies
\[
\frac{d}{dt} \left[ -\frac{1}{H} e^{-H w(t)} \right] = g(t) \Rightarrow e^{-H w(t)} = e^{-H w(0) - H \int_0^t g(t) \, dt}.
\]
Then, one obtains \( w(t) = -\frac{1}{H} \ln \left[ 1 - H \int_0^t g(t) \, dt \right] \). Hence, the solution of (10) is defined by
\[
\frac{b y(t) + \dot{y}(t)}{y(t) [S(0) - H b \int_0^t y(t) \, dt + y(t) - y(0)]} e^{H \left[ -\frac{1}{H} \ln \left[ 1 - H \int_0^t g(t) \, dt \right] \right]}
\]
and by
\[
a(t) = \frac{b y(t) + \dot{y}(t)}{y(t) [S(0) - H b \int_0^t y(t) \, dt + y(t) - y(0)]} e^{H \left[ -\frac{1}{H} \ln \left[ 1 - H \int_0^t g(t) \, dt \right] \right]}.
\]

**Theorem 3 (Existence and Uniqueness of the Solution).** The necessary and sufficient condition for \( y(t) \) to be generated by a SIR model given by (5) and (6) subject to the constraints in (7), with \( a(t) \) uniquely determined by (9), is the validity of the following constraints on the parameter \( \theta \)
\[
0 < l(0) = Hy(0) < 1, \quad 0 < S(0) \leq 1 - Hy(0),
\]
\[
b > \max \left( 0, -\min_{0 \leq t \leq T} \frac{\dot{y}(t)}{y(t)} \right),
\]
\[
0 < H < \frac{b}{b \int_0^T y(t) \, dt + y(T) - y(0)}.
\]

**Proof.** The first constraint on \( l(0) \) is trivial, while the second one depends on the conditions \( S(0) \geq 0 \) and \( l(0) + S(0) \leq 1 \). As for the third constraint, we already discussed its necessity for having \( b > 0, a(t) > 0 \), but now it also follows from (9). In fact, the denominator is positive for \( t \) close to zero and therefore must be positive everywhere, otherwise \( a(t) \) would diverge for some \( t \in [0, T] \). Furthermore, since \( \int_0^T b y(t) \frac{\dot{y}(t)}{y(t)} \, dt = b \int_0^T y(t) \frac{y(t) - y(0)}{y(t)} \) is monotone increasing in light of the previous constraint on \( b \), its maximum value is obtained for \( t = T \). To have the denominator of \( a(t) \) always positive, the constraint about \( H \) represents a necessary and sufficient condition. Finally, the solution is unique in view of the Lipschitz property of the function defining the differential equation in (10).

Some final remarks are now in order. For any given positive and \( \Sigma \) function \( y(t) \), using (9) we can then build infinitely many SIR models which produce \( y(t) \) as output. Clearly, the previous constraints allow us to choose the parameters \( l(0), S(0), b, H \).
in infinitely many (admissible) ways. For any such choice, the existence and unicity of the corresponding infinite-dimensional parameter \( a(t) > 0 \) is guaranteed. Note also that, by defining \( H_{\text{MAX}} := \frac{b}{\bar{b}} \gamma(\delta e^{\gamma(t+\gamma t)} - \gamma) \), for \( H < H_{\text{MAX}} \), recalling that \( R(t) = \bar{b}t \), \( R(t) = \bar{b} \int_{-\infty}^{t} I(\tau) d\tau \), one has

\[
H_{\text{MAX}} = \frac{HS(0)}{R(T) - R(0) + I(T) - I(0)} = H \frac{1}{1 - \frac{S(t)}{S(0)}}
\]

and this clarifies why \( H \) needs to be upper-bounded. In fact, the more \( H \) is close to \( H_{\text{MAX}} \), the more the total infected people tend to saturate to 1, as \( S(T) \) tends to become zero. Another upper-bound on \( H \), implicitly taken into account in Theorem 3 by the constraint \( I(0) < 1 \), is finally given by \( H < y^{-1}(0) \).

Finally let us investigate the sensitivity of \( a(t) \) with respect to \( H \). (9) can be rewritten as follows \( a(t) = \left( \frac{\bar{b} + \gamma(t)}{\bar{b}} \right)^{\frac{1}{2}} \). Since \( I(t) = H y(t) \), if \( H \) increases, \( S(t) \) decreases. So, for small enough values of \( H \) (which correspond to small infectious spread), the value of \( S(t) \) will remain close to 1, making \( a(t) \) almost independent of \( H \). Instead, when \( H \) approaches the maximum value \( H_{\text{MAX}} \), very small variations in \( H \) can cause significant variations in \( a(t) \). This means that \( a(t) \) can be estimated more accurately in presence of small infections, an important point since the interest is in controlling the virus diffusion. All of these statements can be made more rigorous considering that a variation \( \Delta H \) of \( H \) leads to \( \frac{\Delta a(t)}{a(t)} \approx \frac{\Delta H}{H_{\text{MAX}} - H} \). Thus, if \( H = \mu H_{\text{MAX}} \), \( \Delta H = \epsilon H_{\text{MAX}} \), one has \( \frac{\Delta a(t)}{a(t)} \approx \frac{\epsilon}{1 - \epsilon} \). This points out that larger values of \( \epsilon \) imply a stronger dependence of \( \frac{\Delta a(t)}{a(t)} \) on \( \epsilon \).

**Closed-form expressions of the contact rate for the SAIR model**

Here, we derive desired closed-form expressions of the infection rate by considering a generalization of our models which contains the SIR and the SAIR as special cases. In particular, we introduce two different infection rates \( a_1(t) \) and \( a_2(t) \) which lead to the following enriched version of the SAIR model:

\[
\begin{align*}
\dot{S}(t) &= -a_1(t)S(t)I_1(t) - a_2(t)S(t)I_2(t) \\
\dot{I}_1(t) &= a_1(t)S(t)I_1(t) + a_2(t)S(t)I_2(t) - (b_1 + b_2)I_1(t) \\
\dot{I}_2(t) &= b_1I_1(t) - cI_2(t), \quad R(t) = b_2I_1(t) + cI_2(t) \\
y(t) &= \frac{\dot{I}_2(t) + \dot{I}_1(t)}{H}
\end{align*}
\]

Letting \( \bar{w}(t) := \int_{0}^{t} a_1(\tau) y(\tau) d\tau + \frac{1}{H} \int_{0}^{t} a_2(\tau) - a_1(\tau) I_2(\tau) d\tau \), so that \( \bar{w}(t) = a_1(t) y(t) + \frac{1}{H} \int_{0}^{t} a_2(\tau) - a_1(\tau) I_2(\tau) d\tau \), from the first model equation it easily follows that

\[
\dot{S} = -a_1S(I_1 + I_2) - (a_2 - a_1)S\bar{w} = -a_1SHy - (a_2 - a_1)S\bar{w}
\]

\[
\Rightarrow \frac{\dot{S}}{S} = -H \left( a_1 y + \frac{a_2 - a_1}{H} \right) \Rightarrow S(t) = S(0)e^{-Hy(t)}.
\]

Furthermore, the second and the third model equations can be rewritten as

\[
\frac{d}{dt}(I_1 + I_2) = -\dot{S} - b_2(I_1 + I_2) - (c - b_2)I_2
\]

\[
\dot{I}_2 = b_1(I_1 + I_2) - (b_1 + c)I_2.
\]

Hence, we obtain

\[
\frac{\dot{S}}{H} = - \left[ y + b_2y + \frac{c - b_2}{H} I_2 \right]
\]

and \( \dot{I}_2 = b_1Hy - (b_1 + c)I_2 \) which implies

\[
\dot{I}_2(t) = I_2(0)e^{-(b_1+c)t} + Hb_1 \int_{0}^{t} e^{-(b_1+c)(t-\tau)} y(\tau) d\tau.
\]

By substituting the expression of \( I_2(t) \) into the first equality we obtain

\[
\frac{\dot{S}}{H} = - \left[ y + b_2y + \frac{c - b_2}{H} I_2(0)e^{-(b_1+c)t} \right]
\]

\[
+ (c - b_2)b_1 \int_{0}^{t} e^{-(b_1+c)(t-\tau)} y(\tau) d\tau
\]

which leads to

\[
\begin{align*}
\dot{S}(0) &\left( 1 - e^{-Hy} \right) \\
&= \int_{0}^{t} \left[ y(x) + b_2y(x) + \frac{c - b_2}{H} I_2(0)e^{-(b_1+c)x} \right] \\
&+ (c - b_2)b_1 \int_{0}^{t} e^{-(b_1+c)(t-\tau)} y(\tau) d\tau \ dx
\end{align*}
\]

\[
= y(t) - y(0) + b_2 \int_{0}^{t} y(\tau) d\tau + \frac{[c - b_2]I_2(0)}{H(b_1 + c)} \left[ 1 - e^{-(b_1+c)t} \right] \\
+ (c - b_2)b_1 \int_{0}^{t} \int_{0}^{\tau} e^{-(b_1+c)(\tau-\tau')} y(\tau') d\tau' \ dx.
\]

It is useful to define

\[
F(t) := y(t) - y(0) + b_2 \int_{0}^{t} y(\tau) d\tau + \frac{[c - b_2]I_2(0)}{H(b_1 + c)} \left[ 1 - e^{-(b_1+c)t} \right]
\]

\[
+ (c - b_2)b_1 \int_{0}^{t} \int_{0}^{\tau} e^{-(b_1+c)(\tau-\tau')} y(\tau') d\tau' \ dx.
\]

which allows to express \( w(t) \) in terms of \( F(t) \)

\[
w(t) = -\frac{1}{H} \ln \left( \frac{1 - \frac{H}{S(0)} F(t)}{1 - e^{-Hy(t)}} \right) \Rightarrow \dot{w}(t) = \frac{\dot{F}(t)}{S(0) - HF(t)}
\]

so that, since \( \dot{w}(t) = a_1(t) y(t) + \frac{1}{H} \left[ a_2(t) - a_1(t) \right] I_2(t) \), the following relationship between \( a_1(t), a_2(t) \) must hold

\[
a_1(t) \left[ y(t) - \frac{I_2(t)}{H} \right] + a_2(t) \frac{I_2(t)}{H} = \frac{\dot{F}(t)}{S(0) - HF(t)}
\]

Summarizing, it holds that

\[
\dot{I}_2(t) = I_2(0)e^{-(b_1+c)t} + Hb_1 \int_{0}^{t} e^{-(b_1+c)(t-\tau)} y(\tau) d\tau
\]

\[
F(t) = y(t) - y(0) + b_2 \int_{0}^{t} y(\tau) d\tau + \frac{[c - b_2]I_2(0)}{H(b_1 + c)} \left[ 1 - e^{-(b_1+c)t} \right]
\]

\[
+ (c - b_2)b_1 \int_{0}^{t} \int_{0}^{\tau} e^{-(b_1+c)(\tau-\tau')} y(\tau') d\tau' \ dx
\]

\[
\dot{F}(t) = \dot{y}(t) + b_2y(t) + \frac{c - b_2}{H} I_2(0)e^{-(b_1+c)t}
\]

\[
+ (c - b_2)b_1 \int_{0}^{t} e^{-(b_1+c)(\tau-\tau')} y(\tau') d\tau'.
\]

This implies

\[
a_1(t) \left[ y(t) - \frac{I_2(t)}{H} \right] + a_2(t) \frac{I_2(t)}{H} = \frac{\dot{F}(t)}{S(0) - HF(t)}
\]

Such formula thus involves two different infection rates \( a_1(t), a_2(t) \) which describe virus transmission dynamics related to two different classes of infected: asymptomatic/paucysymptomatic and people who exhibit more severe symptoms. However, it also reveals that prior information (which is currently not available) would be needed to estimate/distinguish \( a_1(t), a_2(t) \) from the available epidemiological data. Such identifiability problem is overcome by considering the SAIR where \( a_1(t) = a_2(t) = a(t) \), so that

\[
a(t) = \frac{\dot{F}(t)}{y(t)[S(0) - HF(t)]}
\]

and this gives the closed-form expression for the contact rate in the SAIR model.
