

RESEARCH ARTICLE

Monitoring COVID-19 contagion growth

Arianna Agosto¹ | Alexandra Campmas²  | Paolo Giudici¹ | Andrea Renda³¹University of Pavia, Pavia, Italy²University of Bordeaux - LAREFI, Pessac, France³CEPS, Brussels, Belgium**Correspondence**Paolo Giudici, University of Pavia, Corso Str. Nuova, 65, 27100 Pavia PV, Italy.
Email: paolo.giudici@unipv.it**Funding information**

Horizon 2020 Framework Programme, Grant/Award Number: N°101016233 - H2020-SC1-PHE_CORONAVIRUS-2020-2-RT

We present a statistical model that can be employed to monitor the time evolution of the COVID-19 contagion curve and the associated reproduction rate. The model is a Poisson autoregression of the daily new observed cases and dynamically adapt its estimates to explain the evolution of contagion in terms of a short-term and long-term dependence of case counts, allowing for a comparative evaluation of health policy measures. We have applied the model to 2020 data from the countries most hit by the virus. Our empirical findings show that the proposed model describes the evolution of contagion dynamics and determines whether contagion growth can be affected by health policies. Based on our findings, we can draw two health policy conclusions that can be useful for all countries in the world. First, policy measures aimed at reducing contagion are very useful when contagion is at its peak to reduce the reproduction rate. Second, the contagion curve should be accurately monitored over time to apply policy measures that are cost-effective.

KEYWORDS

contagion models, COVID-19, Poisson autoregressive models, reproduction number

1 | BACKGROUND

The spread of the COVID-19 virus (see e.g. 17) at the beginning of 2020 caught many governments by surprise and unveiled a widespread lack of pandemic preparedness at the global and national level. Before the start of vaccination, at the end of the year, governments have inevitably focused on so-called non-pharmaceutical interventions (NPIs), such as generalized lockdowns or targeted lockdowns and quarantines for infected individuals, in the attempt to reduce the reproduction rate of the virus and thereby limit the number of new cases by preventing human-to-human transmission (so-called “suppression” strategy). In some countries, strong policy measures to contain the virus, such as generalized lockdowns, have led to very significant costs for the local economy, in the form of both a supply side and a demand-side shock. Other countries have, at least initially, considered milder measures (such as mask wearing and targeted quarantine, after contact tracing), which entail allowing a partial continuation of economic activity, in the hope to mitigate the overall cost for society.

In the current circumstances, with an incomplete information about several aspects of the virus (ie, the role of different risk factors, the dynamics of transmission and especially the role of asymptomatic transmission), governments operate under significant uncertainty. Against this background, contagion data from countries are a precious source of information to build data-driven models that can support policy making with scientific evidence. Comparative analysis of the spread of the virus in different countries is essential to start drawing lessons on which policy mixes proved more

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Statistics in Medicine* published by John Wiley & Sons Ltd.

effective, and which ones did not achieve the desired results. To the contrary, research on past pandemics is not particularly useful in this case, since the rate of contagion and the mortality rates of COVID-19 are different compared with HIV, SARS, H1N1, or Ebola.

Early attempts to model the contagion curve of the COVID-19 include Reference 1, which predicted that the outbreak would peak 126 to 147 days (around 4 months) after the start of person-to-person transmission in England and Wales, at a time in which the virus had been found in just 25 countries; and Reference 2, which combines a stochastic transmission model with data from four datasets on cases of COVID-19 originated in Wuhan to estimate how transmission varied over time, and calculate the probability that newly introduced cases might generate outbreaks in other areas. Reference 3 modified an individual-based simulation model developed to support pandemic influenza planning to explore scenarios for COVID-19 in the United Kingdom. In Reference 4, researchers have used early data from the spread of the disease in Wuhan to nowcast and forecast the evolution of the contagion in the region and throughout China, estimating a basic reproductive number for 2019-nCoV of 2.68 and an epidemic doubling time of 6.4 days.

Particularly relevant papers for our work are References 5,6 which, while mathematically expressing the current practices in the modeling on the spread of diseases, draw policy making suggestions. We follow the same line of research, combining mathematical rigor with attention to drawing results that can be useful for policy makers. Specifically, our contribution is a new statistical model for disease spread which, by taking dependence between contagion counts into account, can better capture the contagion curve dynamics and, thus, can draw further light on the understanding of its evolution and possible changes under the action of new policy measures.

Before presenting our model, it is important to recall that the main reference to estimate epidemic contagion is the susceptible infected recovered (SIR) compartmental model (see, eg, References 5-7), which is based on the determination of the reproduction rate (number) R_t of the virus, calculated as:

$$R_t = \frac{f * (1 - a) * E(T)}{h}, \quad (1)$$

which depends on four unknown parameters: f , the infection rate (the probability of an individual to be infected); $E(T)$, the mean generation time (the average serial time interval between infection and subsequent transmission); h , the confirmation rate (the probability of detecting the infected case, typically with a test); and a , the quarantine rate (the probability of isolating the infected case).

Compartmental models such as the SIR are very useful, as they are based on biologic assumptions about disease transmission. However, when many aspects of the disease and its transmission are unknown, as is the case for COVID-19, they hardly satisfy the underlying assumptions. The problem can be solved with the help of data-driven models which, once enough data is accumulated, can replace SIR parameters with their statistical estimates. For example, in the context of COVID-19 policy evaluation, Reference 5 estimates f from consecutive new contagion data and considers a range of alternative a and h fixed values, leading to alternative policy effects. The same author sets $E(T)$ at a fixed value of about one week, in line with References 8,9.

This modelization is quite appealing from a mathematical viewpoint, as it exogenously fixes three unknown parameters and lets only the fourth (ie, the infection rate) to be estimated from the available data. For the infection rate, the most common estimation model is the exponential growth model (see for instance, Reference 7). The latter assumes that the expected number of infected cases at time t , $E(F_t) = \lambda(t)$, is a linear function of time:

$$\log(\lambda_t) = \kappa + \gamma * t + \epsilon, \quad (2)$$

where $\kappa \in \mathbb{R}$ is the intercept, $\gamma \in \mathbb{R}$ the regression coefficient, t represents the number of time periods since the start of the epidemic (t_0), and ϵ is an error term that follows a standard Gaussian distribution.

Once an estimate for the intercept $\hat{\kappa}$ and for the regression coefficient $\hat{\gamma}$ are obtained, $\log(\hat{\lambda}_t)$ can be estimated at any point in time, as $\log(\hat{\lambda}_t) = \hat{\kappa} + \hat{\gamma} * t$. The infection rate f can then be estimated as an instantaneous rate by $f_t = \frac{\hat{\lambda}_t}{\hat{\lambda}_{t-1}}$. Replacing f with f_t in Equation (1), we obtain a time dependent estimate of R_t as:

$$R_t = \frac{f_t * (1 - a) * E(T)}{h}. \quad (3)$$

Epidemiologically, the higher R_t , the higher the number of people that will be infected and, eventually, will be hospitalized in severe conditions, or will die. A value of R_t below 1 indicates that the epidemic is under control and is leading to an upper bound of cases.

The exponential growth model has the advantage of being a parsimonious model: it is based on two parameters, κ and γ , simple to interpret, and on the assumptions of fixed values for the quarantine rate a and the confirmation rate h . On the other hand, it has the disadvantage of being fit “globally,” on the whole distribution, with a limited account of “local” shocks (eg, daily changes in the contagion counts that depend on specific issues, such as changes in people’s behavior and/or periodic changes in policy measures). In addition, it assumes that a and h are fixed along the sample period, which is a rather restrictive assumption.

This shortfall of the SIR methodology motivates the development of a dynamic model, which uses all the available observations up to t to monitor the contagion evolution, also allowing for a and h to be time-varying.

To this aim, in this article, we propose a methodology that belongs to the class of time series models, whose advantages in epidemic modeling have been recently described by Reference 10. We show that our proposed model is able to take changes in contagion dynamics into account, not only in the short-term as in the related works of References 11,12, reflecting people’s behavior but, also, in the-long term, reflecting changes in policy testing (which affect h) and containment measures (which affect a).

The article is organized as follows. Section 2 introduces the methodology. Section 3 proposes a direct application of the model to four countries that have been dramatically hit by the virus in 2020. Section 4 concludes by outlining some implications for policymakers.

2 | METHODOLOGY

To develop a model for the contagion growth of the COVID-19 that properly accounts for time variations of the epidemic counts, we consider the weekly counts of new infections for each country, as in Equation (2). However, differently from Equation (2), which uses a simple regression over time, we consider an autoregressive model that can take both short-term and long-term dependence on past disease counts into account. A similar model has been introduced by Reference 13 in the context of financial contagion. Specifically, the authors introduce a Poisson autoregressive model to the monthly counts of corporate defaults, to model the time dynamics of the bankruptcies observed during financial crisis periods and to assess to what extent they can be explained by commercial and monetary channels through which companies “infect” each other, from an economic point of view.

Here we extend the model to the epidemiologic contagion, enriching SIR models with the capability of quickly adapting to time variations of contagion counts, thus becoming more reliable tools for disease monitoring and control. Doing so, we extend the SIR model with an autoregressive approach to estimating λ_t which, differently from Equation 2, depends on time-varying parameters. The use of time-varying parameters in the modeling of disease counts was introduced by Reference 11 and extended by Reference 12, by means of generalized linear models. We extend their approach by means of an autoregressive model that contains both a short- and a long-term time dependence of counts.

Formally, resorting to the log-linear version of the Poisson autoregression, introduced by Reference 14, we assume that the statistical distribution of new cases at a time t , conditional on the information up to $t - 1$, follows a Poisson law, with a log-linear autoregressive intensity, as follows:

$$Y_t | \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t)$$
$$\log(\lambda_t) = \omega + \alpha \log(1 + y_{t-1}) + \beta \log(\lambda_{t-1}), \quad (4)$$

where $y \in \mathbb{N}$, $\omega \in \mathbb{R}$, $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}$. Note that the inclusion of $\log(1 + y_{t-1})$, rather than $\log(y_{t-1})$, allows to deal with zero values.

In the model, ω is the intercept term, whereas α and β express the dependence of the expected number of new infections, λ_t , on the past counts of new infections.

The sum of the α and β parameters expresses the degree of dependence of the process on its history (its persistence). In other terms, the closer $\alpha + \beta$ to 1, the higher the persistence of shocks in the counting process. Indeed, the value of $\alpha + \beta$ determines the sign and the magnitude of the autocorrelation of the count series. When the sum of α and β is positive, the count time series shows a positive autocorrelation at lag 1. The closer the sum to 1, the slower the decrease of autocorrelation after the first lag. The level of persistence, $\alpha + \beta$, also affects the unconditional mean of the process, that, for a given value of ω , increases with it. Reference 14 show that, as long as $|\alpha + \beta| < 1$, Y_t is stationary and the model parameters can be consistently estimated through the maximum likelihood method.

For a given level of persistence, the values of α and β define the process dynamics. The α component represents the short-term dependence on the previous time point. The higher α , the higher the unconditional variance of the process. The β component expresses instead the long-term dependence on all past values of the observed process. The inclusion of the β component is analogous to moving from an autoregressive conditional heteroscedasticity (ARCH)¹⁵ to a generalized autoregressive conditional heteroscedasticity (GARCH)¹⁶ model in Gaussian processes and allows to capture long memory effects, particularly important in the disease contagion context, as they may relate to changes in government policies.

To further help interpret parameters and understand their role in the process, it is useful to express the log-intensity process as a function of all the past observations:

$$\log(\lambda_t) = \omega \frac{1 - \beta^t}{1 - \beta} + \beta^t \log(\lambda_0) + \alpha \sum_{i=0}^{t-1} \beta^i \log(1 + y_{t-i-1}), \quad (5)$$

where both λ_0 and y_0 are assumed to be fixed.

From Equation (5), one may note that the long-run tendency of the process is determined by β , which reflects the dependency of the current counts on the whole history of contagion, whereas α determines the dependency of the same counts on the latest value, which acts as a temporary correction. Indeed, for a given α , the higher the β parameter, the slower the decrease of the influence of remote changes on the new cases. And, for a given β , the higher the α parameter, the higher the chance of important differences between consecutive counts.

Once estimated values for the intercept and the autoregressive coefficients $\hat{\alpha}$ and $\hat{\beta}$ are obtained, the prediction of the new infection counts in t , $\hat{\lambda}_t$, is obtained and can be inserted in Equation (3), to get an estimate of the reproduction rate, R_t , which is time dependent.

A problem in the application of Equation (3) is that only the infection rate f , but also h and a , vary considerably over time and cannot be considered “exogenous” to the infection process. For example, a depends on how many tests are carried out, and this depends on the pressure on the health system: during “stress” times (with high levels of contagion), tests can be performed only on symptomatic hospitalized patients; during “calm” times they can be performed more widely, and on asymptomatic individuals as well. Similarly, h very much depends on the perception of the contagion levels: during “stress” times individuals are likely to be more compliant with policy measures and stay in quarantine, than during “calm” times, leading to a higher value of h .

We address this problem assuming that the Poisson distribution of Y_t “endogenously” includes the measurement error that is determined by the variation of a and f over time, so that:

$$R_t = f_t * E(T). \quad (6)$$

Note that if we further assume that $E(T) = 1$ week, without loss of generality, and in line with the research findings in the epidemiological literature, the estimate of R_t simplifies to the observed weekly rate of contagion:

$$R_t = \frac{\hat{\lambda}_t}{\hat{\lambda}_{t-1}}.$$

We remark that our assumption is in line with the literature on time series models for epidemic counts (as, eg, Reference 11 and with our proposed autoregressive model, that include both a short-term component, which explains a , and a long-term component, which explains h).

3 | APPLICATION

This section illustrates the application of our model to real data. We remark that, although the analysis presented in this article refers to data available up to a certain fixed time, it is continuously updated by means of the application of the open software R to the latest data with an infographics that is available at: <https://www.ceps.eu/ceps-publications/monitoring-covid-19-contagion-growth-in-europe/>. We also remark that, while in the article we focus on four countries where the virus has spread widely: Brazil, Italy, the United States, and the United Kingdom, the infographics presents results for several other countries.

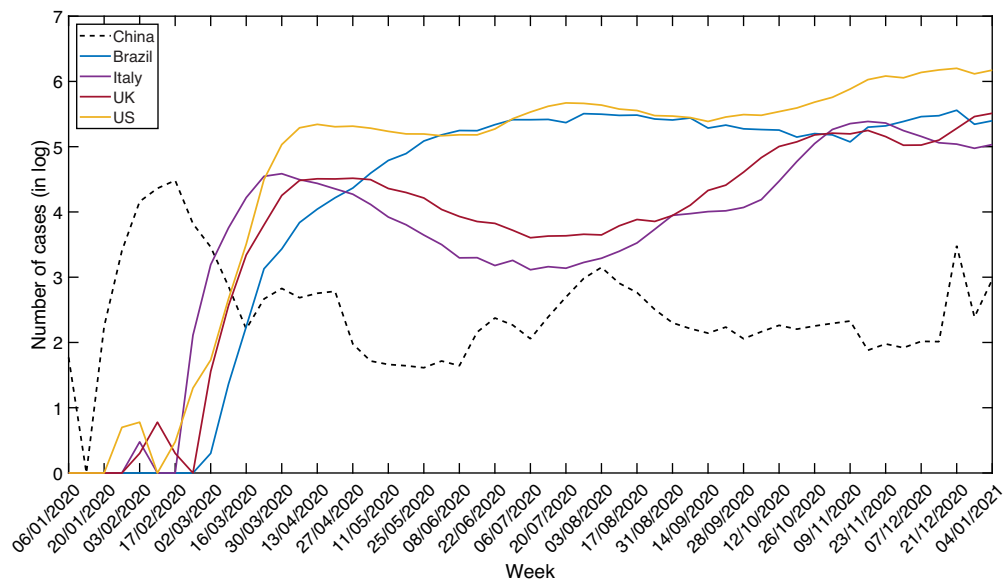


FIGURE 1 Observed new weekly infection counts in 2020, for Brazil, China, Italy, the United Kingdom, and the United States [Colour figure can be viewed at wileyonlinelibrary.com]

3.1 | Data

We have applied our model to the data retrieved from the European Centre for Disease Prevention and Control,¹⁷ which are available on a weekly basis. Although epidemiologists and pathologists are suggesting that isolated cases of COVID-19 might have been present several weeks and possibly months before what was originally believed to be the first cases, we start our analysis with the week that begins after the first recorded cases of local transmission in 2020, in line with our “data-driven” approach: 9 March for Italy; 16 March for Brazil; 23 March for the United Kingdom, and 30 March for the United States, up to 4 January, 2021 for all countries. According to the chosen sample data, Figure 1 presents the observed evolution of weekly new infections for the world’s countries currently most hit by the virus, as well as for China, the country where the outbreak started.

Figure 1 shows that China has experienced one main wave of contagion, with an upward trend, a peak in February, a downward trend, and subsequent fluctuations. Some weeks later, Italy has experienced a similar cycle, but with a longer peak “plateau” in March, followed by a descent. The contagion curve for the United Kingdom is similar to that of Italy, with a shift in the outbreak, and a second wave that is not yet settled.

The contagion curve for the United States initially follows that of the United Kingdom, but then it departs with smaller downtrends and the presence of a summer wave. The contagion growth curve for Brazil shows a different path; it progresses toward a peak in July and then stays at that level throughout the year, with some fluctuations.

3.2 | Results

We present the results from the application of our proposed model in five main steps. In a first step, we present parameter estimates on the whole available sample, for each country under investigation. In a second step, we evaluate the predictive accuracy of the estimated model, comparing it with the classic exponential growth model outlined in Equation (2). In a third step, we evaluate whether the model estimates are robust, in which case they can be taken as the basis to derive policy actions. In a fourth step, we insert model estimates in Equation (3) and derive a dynamic estimate of R_t which allows to understand the point in time effects of policy measures. Finally, in a fifth step, we examine what could be the effects of alternative policy measures on R_t , in a scenario analysis.

3.2.1 | Estimation

In this section, we present the parameters, ω , α , and β , for each country estimated on the whole sample. The results are presented in Table 1.

TABLE 1 Parameter estimates

	Brazil		Italy		UK		USA	
	Estimate	SE (P-value)	Estimate	SE (P-value)	Estimate	SE (P-value)	Estimate	SE (P-value)
ω	2.540	0.362 (.000)	1.269	0.565 (.030)	0.255	0.519 (.626)	0.706	0.652 (.285)
α	0.664	0.119 (.000)	1.493	0.063 (.000)	1.582	0.093 (.000)	1.352	0.010 (.000)
β	0.132	0.103 (.205)	-0.599	0.053 (.000)	-0.595	0.064 (.000)	-0.400	0.095 (.000)

Source: Authors' elaboration.

TABLE 2 Out-of-sample predictive accuracy

Model	Brazil		Italy		UK		USA	
	RMSE	MPE	RMSE	MPE	RMSE	MPE	RMSE	MPE
PAR	88 988	13.09%	15 532	13.08%	100 688	25.55%	379029	24.10%
Exp.	281 549	91.12%	72 976	61.73%	172 562	29.67%	3 022 785	168.61%

Source: Authors' elaboration.

When first considering the estimates for the β parameter, Italy, and the United Kingdom have a high negative sign, consistently with what observed in Figure 1. For the United States, the parameter is also negative, but with a lower magnitude, again in line with the more limited downward trends found in Figure 1. Finally, as expected from Figure 1, for Brazil the parameter is not significant. These results may indicate that policy measures have been effective in reducing the contagion, in Italy and in the United Kingdom; to a lesser degree, in the United States, and have not been effective in Brazil.

The estimates for the α parameter, instead, indicate a positive dependence of consecutive counts, for the United Kingdom, Italy, and the United States, and, to a more limited extent, for Brazil. These results may indicate the action of the reproduction rate, but also the fact that test measurements vary over time.

Overall, the calculated model estimates help monitor the evolution and the dynamic of the virus, accounting for different factors, such as policy measures, testing procedures, and different country characteristics.

3.2.2 | Predictive accuracy

To evaluate the predictive accuracy of the model we compare our results with the ones obtained from the exponential growth model (see Equation (2)). We apply both models—estimated using data until the third week of 2020 (21/12/2020)—to make predictions of the weekly new cases for the subsequent four weeks, that is, up to the third week of 2021 (18/01/2021). The predictions are then compared in terms of out-of-sample predictive performance. To this purpose, we compute the value of root mean squared error (RMSE) and the absolute mean percentage error (MPE) for the two specifications. The predictive performance are reported in Table 2.

The results show that the PAR model always outperforms the exponential model. This finding is consistent with what obtained in Reference 13 and confirms the superiority of our proposed model with respect to an exponential one. This is particularly evident for the United States, for which the MPE is more than 1.5 times the observed value for the exponential model, against 0.25 times for the Poisson autoregressive model.

3.2.3 | Robustness

To evaluate robustness, we apply the model to each country data series repeatedly, using an increasing amount of data: we first fit the model over the first two weeks, then over the first three weeks, and so on. The plot of the resulting (α , β) estimates, from week 2 onward, indicates whether (and when) parameter estimates have converged, in which case the model can be taken as a valid statistical representation of the underlying contagion phenomena.

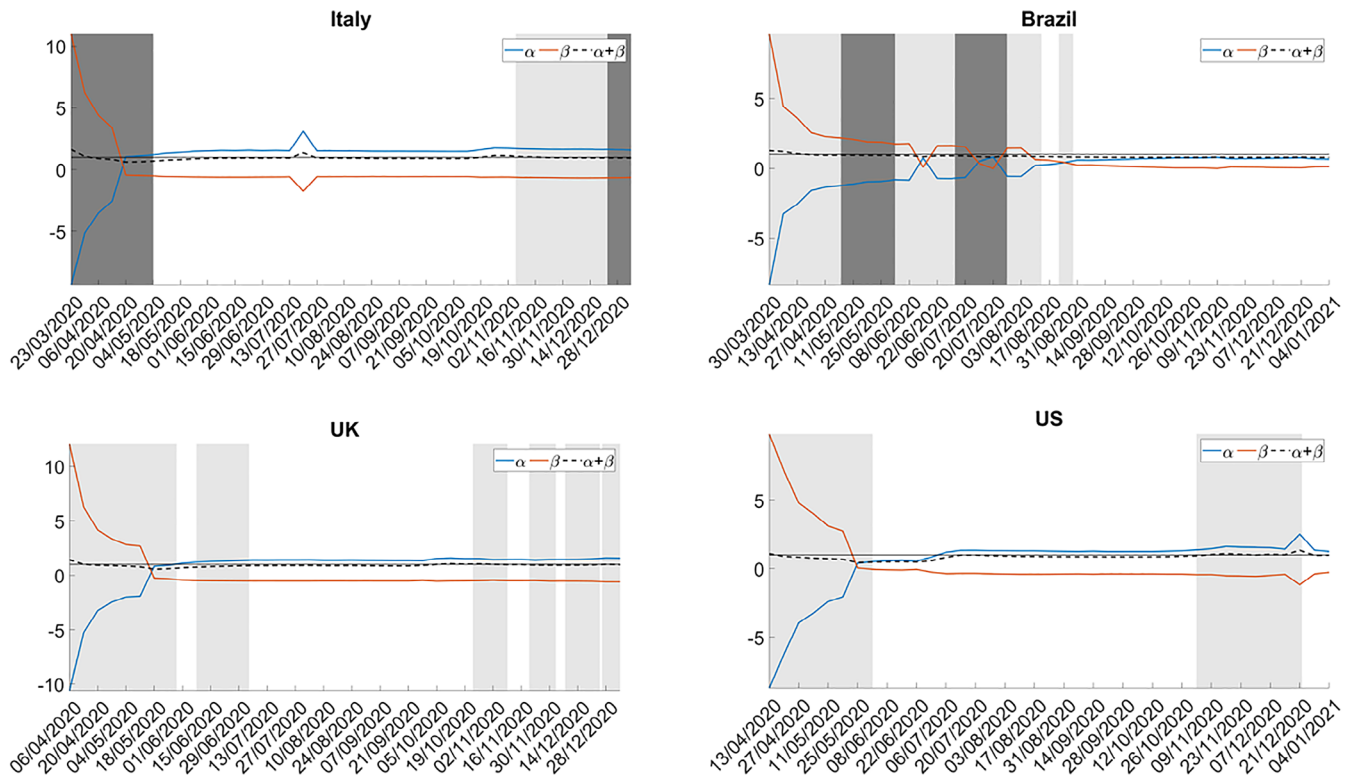


FIGURE 2 Evolution of the α , β and $\alpha + \beta$ parameters for European countries and the United States. Note that the light and dark gray areas correspond respectively to a government response stringency index ranging between 70% and 80% and a government response stringency index above 80%. The dashed line reports the sum of α and β [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.com)]

The dynamic representation of the parameter estimates can be enriched with the indication of the main policy measures adopted by the four countries. To this aim, we rely on the government response stringency index calculated by Reference 18 and add to the plot a light gray or a dark gray bar when the index signals a medium or a high stringency, respectively above 70% and 80%. Figure 2 provides information on the weekly estimated pair of coefficients (α , β), from week 3 in the sample onward*, along with the described gray bars.

From Figure 2, one may note that for Italy, the order between α and β is reverted and stabilized at the beginning of April, with a slight reopening in July. This proves that the model is stable, and can be used as basis for an evidence-based policy making. Note that the date of inversion of the parameters approximately corresponds to the beginning of the descending phase of the cycle in Figure 1 (ie, toward the end of March).

A similar comment applies to the United Kingdom, although the found evidences are delayed in time with respect to Italy.

Brazil and the United States show a lower degree of robustness. For Brazil, the gap between the two parameter estimates seems to have stabilized quite late, around September; for the United States, this does not seem to be yet the case. These results may depend on a wide variation of cases within a country, greater than for other countries.

When the curves in Figure 2 are read jointly with the policy stringency indications, further indications emerge. Note, for example, that the order between β and α has been reverted, indicating an improvement of the situation, when Italy and the United Kingdom have implemented, respectively, strong and mild policy measures. A similar phenomena occurs, although to a lesser extent, for the United States. On the other hand, the implementation of stringent measures in Brazil is not paired by a corresponding change in the curves. Last, for the United States, we cannot reach a firm conclusion.

It is important to mention that when $|\alpha + \beta| > 1$, as it occurs at the beginning of the period, the contagion process diverges and indicates that the estimates are not reliable. This suggests an important monitor check on data quality: when the persistence diverges, more data should be accumulated.

*Note that the values in Table 1 reflect the latest estimated parameters for each country in Figure 2, as they correspond to the values estimated over the entire sample, that is, up to the first week of 2021.

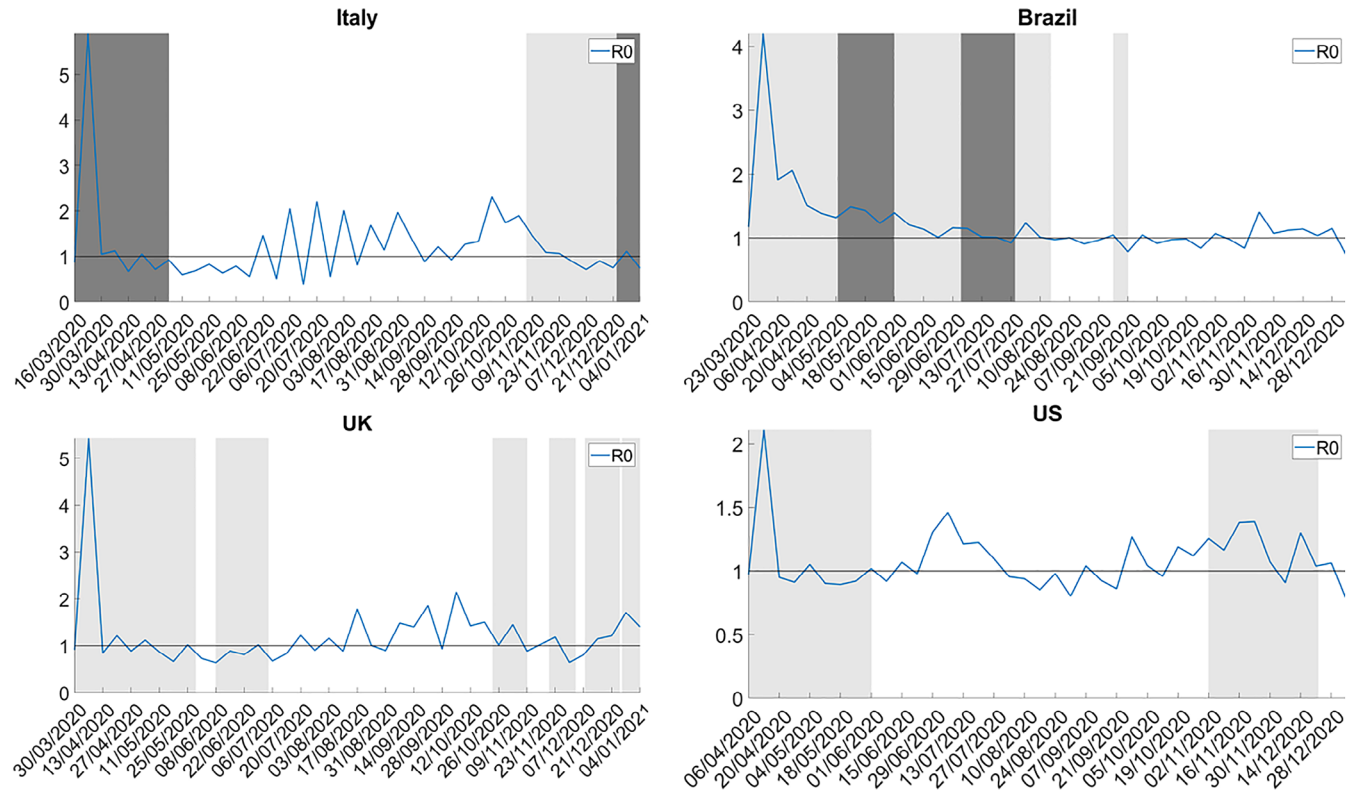


FIGURE 3 Evolution of the R_t parameter in Brazil, Italy, the United Kingdom, and the United States. Note that the light and dark gray areas correspond respectively to a government response stringency index ranging between 70% and 80% and a government response stringency index above 80% [Colour figure can be viewed at wileyonlinelibrary.com]

3.2.4 | Reproduction rates

Following what suggested in Equation (3) with mean generation time of one week (ie, $E(t)$ is set to one week as suggested in Reference 5), we now calculate R_t dynamically, for each of our time point. A reproduction rate estimates, dynamically calculated, can constitute a very important monitoring tool, which can be used to strengthen or to relax policy actions, over time. When the estimated R_t goes below the value of 1, for a sufficient length of time, actions can be relaxed. Conversely, when the estimated R_t shows persistent values greater than 1, actions could be strengthened. Figure 3 presents the results for each surveyed country and shows that, while the previous findings are confirmed, more conclusions can be drawn.

Specifically, Figure 3 illustrates a very high value for R_t at the beginning of the growth curve, for Brazil, the United Kingdom, and Italy. The United States have a lower initial value, consistent with a heterogeneous spread across the country. As a consequence of the adopted policy measures, Italy, the United Kingdom, and the United States managed to bring R_t down quite rapidly, differently from Brazil. During the subsequent waves of contagion, however, all countries have more difficulty in reducing R_t . This may be due to milder policies or to a more limited behavioral compliance to the measures, or both.

We remark the advantage of the dynamic estimates obtained with our model, with respect to a classic exponential model: when a new data point comes in a new estimated R_t , it can be added to the graph, without the need to refit the model on the whole sample.

3.2.5 | Scenario analysis

It is important to recall that our methodology to compute R_t assumes that the probability to be subject to quarantine as well as the probability of a case to be detected (respectively, the parameter a and h in Equation (1)) are retrieved along with the infection rate f from the data, through the estimated b parameter. In other words, the process followed by the

TABLE 3 Estimation of the overall R_t under the alternative scenarios

Stringent measures		
Country	Scenario 1a ($a = 0.6$ and $h = 0.4$)	Scenario 2a ($a = 0.6$ and $h = 0.6$)
Italy	0.761	0.509
UK	1.405	0.810
Mild measures		
Country	Scenario 1b ($a = 0.4$ and $h = 0.6$)	Scenario 2b ($a = 0.6$ and $h = 0.6$)
Brazil	0.733	0.693
USA	0.745	0.666

Note: The parameters a and h respectively denote the probability of an infected individual to be quarantined and the probability of a case to be detected.

Source: Authors' elaboration.

reproduction rate of positive cases in Figure 3 is “learning” all these three parameters together from the newly observed number of infected cases.

With the aim of assessing the effectiveness of alternative policy measures, we can consider the estimated value of b as a “baseline” current setting scenario, to which we can add new policy scenarios, and evaluate their effects. To this aim, we can assume, without loss of generality, that the estimated R_t is consistent with a baseline scenario such that $1 - a = h$. We can then evaluate how R_t could evolve from its latest value in Figure 3, under the effect of new policy measures which impact either a , b , or both.

The scenario values for the quarantine rate a can be related to the work carried out by Reference 5, who consider a policy measure “strong” when a exceeds a level that can be reached only applying harsh measures, such as generalized lockdowns, which have significant economic costs. Conversely, a policy measure is considered “mild” when a is below the same level, with lower costs for the economy. While Reference 5 sets for China the threshold at $a = 0.8$ we consider a more realistically achievable level of $a = 0.6$ in countries featuring stringent measures such as generalized quarantines (as it occurred in Italy during the first wave of the pandemic), and $a = 0.4$ in countries enacting “mild” measures such as localized lockdowns and targeted quarantines, based on contact tracing (eg, in the United States).

Concerning the confirmation rate h , we recall that it expresses the probability that an infected case is confirmed by a test as a positive case. Ideally, this rate should be equal to 1. However, there is an inherent risk that, for instance, tests are limited or too costly, that the test fails, or that it is applied to an individual with symptoms, that are those of the classic flu, and not of the COVID-19, with a waste of time and resources. While Reference 5 targets this rate at $h = 0.9$ in China, we assume a more realistic threshold of $h = 0.6$. Indeed, Reference 19 indicates that this assumption is reliable: they estimate h for Italy at different time points and show that it is rather low in the early stages of virus spread (ie, until the beginning of March) and, then, it increases afterward. Following the previous considerations, Table 3 summarizes two different scenarios that can be used as a complement of the dynamic estimations illustrated in Figure 3.

The scenarios in Table 3 can be classified in strong measures (scenarios 1a and 2a) and mild measures (scenarios 1b and 2b). Scenarios 1a and 1b are considered as the baseline scenarios, assuming $a = 0.4$ and $h = 0.6$ for countries adopting stringent measures and $a = 0.6$ and $h = 0.4$ for countries adopting “mild” measures. We then assume that all countries can strengthen their measures by a third. Thus, while in scenario 2a, h shifts to $h = 0.6$ for countries enacting stringent measures account, in scenario 2b, a shifts to 0.6 for countries that have adopted “mild” measures. Accordingly, the latest R_t values in Figure 3, in line with Equation (1), can be multiplied (divided) by the ratio between the shifted value and the assumed current value of $1 - a$ (h), equal to two third, to obtain new values for R_t .

Taking into account both an increased testing and a strengthening of policy interventions, the value of R_t becomes lower for all countries, proportionally to their starting point. Countries with an important R_t under Scenario 1 are more likely to further compress the reproduction rate of infected cases under Scenario 2. For all countries, heightening the measures would definitely help reduce the growth of newly infected cases.

Having said that further measures may be useful, their marginal benefit (as shown, eg, by the reduction of R_t) should be weighed against their costs. To this aim, we believe that our research should be complemented with research work that aims to measure the socioeconomic costs for controlling the COVID-19 pandemic, as in the recent work of Reference 20.

We finally would like to mention that while we have presented one possible scenario configuration, many others can be easily considered by means of application of Equation (1) as shown in,^{21,22} and can be extended in a Bayesian framework, as in Reference 23.

4 | CONCLUSIONS

Faced with a pandemic for which they were hardly prepared, all world's countries have adopted a plethora of rushed policy measures, ranging from mere reliance on the population's responsible behavior, to complete country lockdown coupled with relatively harsh enforcement measures. Countries have adopted different criteria for testing, a variety of technological means to track the spread of the virus, and different ways to report cases. All this is important in order to assess the effectiveness of the response to the pandemic in different countries.

This article proposes a model that allows for the identification of short-term and long-term components in data relative to COVID-19 contagions. Our intention is to propose a dynamic model, rather than a set of definitive results: as the timeframe becomes longer, the model predictions evolve, particularly as a result of the adopted policies. The results of the model can be updated on a weekly basis and can offer governments a basis for forecasting contagion peaks, monitoring the evolution of contagion, and plan policy choices for the "postlockdown" period. As a matter of fact, once the peak is reached, countries will start considering a mix of new policy measures, gradually lifting restrictions to relaunch the economy, such as the obligation to wear masks in public places, the development of apps for self-reporting of symptoms, or a widespread, randomized testing strategy. These measures, depending on the circumstances, may also trigger a return of the COVID-19 contagion, which can be detected by the proposed model. Indeed, the model can be used to monitor the spread of the virus in the postlockdown phase, enabling a comparative analysis of the effectiveness of alternative policy measures.

The model must be analyzed with a number of caveats in mind. First, the model relies on official data on reported contagion cases, which may represent only part of the total cases, both due to limited testing and also to the fact that part of the infected population (although apparently a small proportion) remains asymptomatic throughout the infection period. Second, countries have followed different approaches to testing and to containment measures, and some countries (or some subnational governments) have changed approach to testing and containment during the period analyzed. Although the model can capture these aspects into its autoregressive dynamics, more data, in time and space, could help to better explain the underlying causes behind the model estimates.

That said, the dynamic estimates presented in this article suggest that the countries for which the first policy measures were taken to limit the spread of the contagion were relatively effective; however, subsequent measures have proved to be less effective and may require more fine-tuning of policies on the basis of the evidence found.

We finally remark that the article provides a base methodology that can be extended in several directions. In particular, an important direction of research could be to explicitly add in the model formulation policy covariates, once good-quality policy stringency variables are available. Another important extension is to take spatial dependence into account, and especially for larger countries, for which infection rates can vary greatly within the country. This can be possible applying the proposed model to subnational data, if available, including a component that models the spatial dependence of the contagion counts of a region on its neighbors' counts.


ACKNOWLEDGEMENTS

This research has received funding from the European Union's Horizon 2020 research and innovation program "PERISCOPE: Pan European Response to the Impacts of COVID-19 and future Pandemics and Epidemics," under the grant agreement N°101016233 - H2020-SC1-PHE_CORON-AVIRUS-2020-2-RTD. While the work is the result of the joint collaboration between the authors, Paolo Giudici, Arianna Agosto, Alexandra Campmas, and Andrea Renda wrote Sections 1, 2, 3, and 4, respectively.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article. Indeed, the data that supports the findings can directly be uploaded by means of the R-code provided as supplementary material for Review and Publication.

ORCID

Alexandra Campmas  <https://orcid.org/0000-0001-9108-1034>

REFERENCES

1. Danon L, Brooks-Pollock E, Bailey M and Keeling MJ. A spatial model of CoVID-19 transmission in England and Wales: early spread and peak timing. medRxiv; 2020.
2. Kucharski AJ, Timothy WR, Charlie D, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis.* 2020;20(5):553-558.
3. Imperial College COVID-19 Response Team. *Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID19 Mortality and Healthcare Demand. Technical Report 9.* Imperial College; 2020.
4. Waltersa CE, Mesléb M, Hall I. Modelling the global spread of diseases: a review of current practice and capability. *Epidemics.* 2018;25:1-8.
5. Gu C, Jiang W, Zhao T, and Zheng B. Mathematical recommendations to fight against COVID-19. SSRN; 2020.
6. Giordano G, Bianchini F, Bruno R, et al. Modelling the COVID-19 epidemic and implementation of population wide interventions in Italy. *Nat Med.* 2020;26:855-860.
7. Biggerstaff M, Cauchemez S, Reed C, Gambhir M, Finelli L. Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature. *BMC Infect Dis.* 2014;14:480.
8. Chen LM, Yuan Q, Song B, and Ma J. Transmission characteristics of the COVID-19 outbreak in China: a study driven by data. medRxiv; 2020.
9. Linton NM, Kobayashi T, Yang Y, et al. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. I. *J Clin Med.* 2020;9(2):538.
10. Harvey A, Kattuman P. Time series models based on growth curves with application to forecasting coronavirus. *Harvard Data Sci Rev.* 2020;4.
11. Held L, Hohle M, Hoffman M. A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Stat Modelling.* 2005;5:187-199.
12. Paul M, Held L. Predictive assessment of a non-linear effects for multivariate time series of infectious disease counts. *Stat Med.* 2011;30(10):1118-1136.
13. Agosto A, Cavaliere G, Kristensen D, Rahbek A. Modeling corporate defaults: poisson autoregressions with exogenous covariates (PARX). *J Empir Financ.* 2016;38(B):640-663.
14. Fokianos K, Tjøstheim D. Log-linear Poisson autoregression. *J Multivar Anal.* 2011;102:563-578.
15. Engle RF. Autoregressive conditional heteroscedasticity with estimates of the variance of UK. *Inflation Econom.* 1982;50:987-1008.
16. Engle RF, Bollerslev T. Modelling the persistence of conditional variances. *Econom Rev.* 1986;5(1):1-50.
17. European Center of Disease Control Download today's data on the geographic distribution of COVID-19 cases worldwide; 2020. <https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide>.
18. Hale T, Webster S, Petherick A, Phillips T, Kira B. Oxford COVID-19 government response tracker; 2020. <https://ourworldindata.org/grapher/covid-stringency-index?stackMode=absolute&country=BRA®ion=World>.
19. Brogi F, Guardabascio B. A model to estimate the effect of main COVID-19 containment measures in Italy. working paper; 2020.
20. Gros C, Valenti R, Valenti K, Gros D. Strategies for controlling the medical and socio-economic costs of the corona pandemic. *Proc Nat Acad Sci.* 2020.
21. Liboschik T, Fokianos K, Fried R. tscount: an R package for analysis of count time series following generalised linear models. *J Stat Softw.* 2017;82:5.
22. Agosto Arianna, Giudici Paolo. A Poisson Autoregressive Model to Understand COVID-19 Contagion Dynamics. *Risks.* 2020;8(3):77. <http://dx.doi.org/10.3390/risks8030077>.
23. Giudici Paolo, Mezzetti Maura, Muliere Pietro. Mixtures of products of Dirichlet processes for variable selection in survival analysis. *Journal of Statistical Planning and Inference.* 2003;111(1-2):101-115. [http://dx.doi.org/10.1016/s0378-3758\(02\)00291-4](http://dx.doi.org/10.1016/s0378-3758(02)00291-4).

How to cite this article: Agosto A, Campmas A, Giudici P, Renda A. Monitoring COVID-19 contagion growth. *Statistics in Medicine.* 2021;40:4150–4160. <https://doi.org/10.1002/sim.9020>