scientific reports

Check for updates

OPEN

Modeling and forecasting the early evolution of the Covid-19 pandemic in Brazil

Saulo B. Bastos[1] & Daniel O. Cajueiro[1,2,3✉]

We model and forecast the early evolution of the COVID-19 pandemic in Brazil using Brazilian recent data from February 25, 2020 to March 30, 2020. This early period accounts for unawareness of the epidemiological characteristics of the disease in a new territory, sub-notification of the real numbers of infected people and the timely introduction of social distancing policies to flatten the spread of the disease. We use two variations of the SIR model and we include a parameter that comprises the effects of social distancing measures. Short and long term forecasts show that the social distancing policy imposed by the government is able to flatten the pattern of infection of the COVID-19. However, our results also show that if this policy does not last enough time, it is only able to shift the peak of infection into the future keeping the value of the peak in almost the same value. Furthermore, our long term simulations forecast the optimal date to end the policy. Finally, we show that the proportion of asymptomatic individuals affects the amplitude of the peak of symptomatic infected, suggesting that it is important to test the population.

The world has seen an ongoing pandemic of COVID-19 (coronavirus 2) caused by severe acute respiratory syndrome SARS-CoV-2. According to the World Health Organization (WHO)[1], although most people infected with it will present mild respiratory symptoms, or no signs of the disease, and recover without needing special treatment, older people, and those with severe medical conditions like diabetes, cardiovascular disease, or chronic respiratory disease may develop serious illness. While the COVID-19 outbreak was first identified in Wuhan, Hubei, China, in December 2019, we could only confirm the first case in Brazil on February 25, 2020. The first known patient in Brasil was a 61-year-old man from São Paulo who had returned from Lombardy (Italy) and tested positive for the virus. Since then, we may confirm 4579 cases and 159 deaths (March 30, 2020) in roughly the entire Brazilian territory. Like in the rest of the world[2], the Brazilian government response to the pandemic has been the introduction of measures to ensure social distancing, such as schools closure, restricting commerce, banning public events and home office.

We use the Brazilian recent data from February 25, 2020 to March 30, 2020 to model and forecast the evolution of the COVID-19 pandemic. Our study focuses on the early period of the pandemics that accounts for unawareness of the epidemiological characteristics of the disease in a new territory, sub-notification of the real numbers of infected people and the timely introduction of social distancing policies to flatten the spread of the disease. This work has had the practical appeal for providing preliminary estimates of Covid-19 epidemiological parameters and the duration of the social distancing policy in Brazil.

The computational modeling of infectious diseases comprises a large collection of models[3–5]. In order to model the evolution of the Covid-19 in Brazil we modify two versions of the the Susceptible-Infected-Recovered (SIR) model[6] to consider the effects of social distancing measures in the evolution of the disease. The SIR model describes the spread of a disease in a population split into three non-intersecting classes: susceptible (S) are individuals who are healthy but can contract the disease; Infected (I) are individuals who are sick; Recovered (I) are individuals who recovered from the disease. Due to the evolution of the disease, the size of each of these classes change over time and the total population size N is the sum of these classes

$$N(t) = S(t) + I(t) + R(t). \tag{1}$$

[1]Department de Economics, FACE, Universidade de Brasília (UnB), Campus Universitário Darcy Ribeiro, Brasília 70910-900, Brazil. [2]Nacional Institute of Science and Technology for Complex Systems (INCT-SC), Universidade de Brasília (UnB), Brasília, Brazil. [3]Machine Learning Laboratory for Finance and Organizations (LAMFO), FACE-Universidade de Brasília (UnB), Campus Universitário Darcy Ribeiro, Brasília 70910-900, Brazil. ✉email: danielcajueiro@unb.br

nature research

Let β be the average number of contacts that are sufficient for transmission of a person per unit of time t. Then $\beta I/N$ is the average number of contacts that are sufficient for transmission with infective individuals per unit of time of one susceptible and $(\beta I/N)S$ is the number of new cases per unit of time due to the S susceptible individuals. Furthermore, let γ be the recovery rate, which is the rate that infected individuals recover or die, leaving the infected class, at constant per capita probability per unit of time.

Based on these definitions, we can write the SIR model as

$$
\begin{aligned}
\frac{dS}{dt} &= -\frac{\beta IS}{N} \\
\frac{dI}{dt} &= \frac{\beta IS}{N} - \gamma I \quad \text{[SIR]}. \\
\frac{dR}{dt} &= \gamma I
\end{aligned}
\tag{2}
$$

It is worth mentioning that we can also evaluate the number of recovered individuals from Eq. (1) using also the number of susceptible and infected individuals, since in this version of the SIR model (Eq. 2) the population is constant. Actually, since we are modeling a short term pandemic, we do not consider the demographic effects and we assume that an individual does not contract the disease twice. We do not implement this model, we only included it for the sake of reference.

We actually want to estimate the fraction of people that die from the disease. Then we include a probability ρ of an individual in the class I dying from infection before recovering[4]. In this case, we get the following set of equations

$$
\begin{aligned}
\frac{dS}{dt} &= -\frac{\beta IS}{N} \\
\frac{dI}{dt} &= \frac{\beta IS}{N} - \gamma I - \frac{\rho}{1-\rho}\gamma I = \frac{\beta IS}{N} - \frac{\gamma I}{1-\rho} \\
\frac{dR}{dt} &= \gamma I \\
\frac{dD}{dt} &= \frac{\rho}{1-\rho}\gamma I
\end{aligned}
\quad \text{[SIRD]},
\tag{3}
$$

where $\frac{\rho}{1-\rho}\gamma I$ is the number of people in the population that die due to the disease per unity of time and D is the number of people that die due to the disease. Note that in this case the number of individuals in the population reduces due to the infection according to $\frac{dN}{dt} = -\frac{\rho}{1-\rho}\gamma I$. For the ease of reference, we call this model "**SIRD**" (Susceptible-Infected-Recovered-Dead) model.

Since, in the case of the COVID-19, there is a relevant percentage of the infected individuals that are asymptomatic, we split the class of infected individuals in symptomatic and asymptomatic[7–9]:

$$
\begin{aligned}
\frac{dS}{dt} &= -(\beta_A I_A + \beta_S I_S)\frac{S}{N} \\
\frac{dI_A}{dt} &= (1-p)(\beta_A I_A + \beta_S I_S)\frac{S}{N} - (\gamma_A)I_A \\
\frac{dI_S}{dt} &= p(\beta_A I_A + \beta_S I_S)\frac{S}{N} - \frac{\gamma_S I_S}{1-\rho} \\
\frac{dR_A}{dt} &= \gamma_A I_A \\
\frac{dR_S}{dt} &= \gamma_S I_S \\
\frac{dD}{dt} &= \frac{\rho}{1-\rho}\gamma_S I_S
\end{aligned}
\quad \text{[SIRASD]},
\tag{4}
$$

where I_A is the number of asymptomatic individuals, I_S is the number of symptomatic individuals, R_A and R_S are the recovered individuals from the asymptomatic and symptomatic infection, respectively, and p is the proportion of individuals who develop symptoms. For ease of reference, we call this model "**SIRASD**" (Susceptible-Infected-Recovered for Asymptomatic-Symptomatic and Dead) model. Like the SIRD model, the condition that N is constant does not hold anymore and if we need to evaluate N over time, we need to integrate $\frac{dN}{dt} = -\frac{\rho}{1-\rho}\gamma_S I_S$.

In order to consider the effect of the social distancing policy, we modify the transmission factors of Eqs. (3) and (4) by multiplying them by a parameter $\psi \in [0,1]$, when the date belongs to the period of the implementation of government policy. Otherwise, we use $\psi = 1$. To be precise, we replace β in Eq. (3) by $\psi\beta$, β_A in Eq. (4) by $\psi\beta_A$ and β_S in Eq. (4) by $\psi\beta_S$. Note that doing this procedure we avoid the introduction and estimations of new "βs" and we may use ψ to evaluate the effectiveness of social distancing policy. In the end, we may measure the social distance as $1 - \psi$.

Our models provide estimates of the epidemiological parameters, that are consistent with the international literature, and good forecasts of the short-term Brazilian time series of infected individuals in Brazil. Furthermore, one of our models assesses the number of asymptomatic (or individuals with mild symptoms that do not
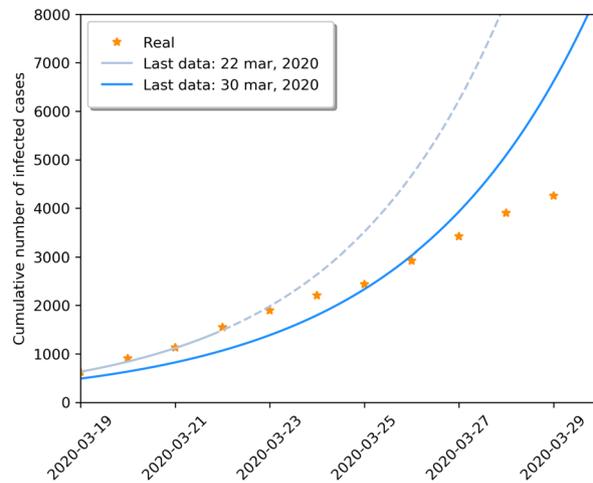
Figure 1. Estimations of the SIRD model for different final date points. The solid line corresponds to the last date which the model was estimated, and the dashed line are model predictions. We represent the real data as points.

look for the hospitals and are not being tested). We use these models to simulate long-term scenarios of the pandemics that depend on the level of engagement of the Brazilian social distancing policy. We show that: (1) the social distancing policy imposed by the government is able to flatten the pattern of contamination provided by the COVID-19; (2) there is an optimal date for abandoning the social distancing policy; (3) short-term social distancing policies only shift the peak of infection into the future keeping the value of the peak in almost the same value. (4) The proportion of asymptomatic individuals affects the amplitude of the peak of symptomatic infected, meaning that it is important to invest in testing the population, massively or by random sampling.

Our work relates to the recent interesting contributions[10–13] in the sense that all these works try to model the spread of the COVID-19 and to evaluate the countermeasures against this virus. However, our paper differs from these works in the following dimensions: (1) data: our work focuses in Brazilian data. This is an important characteristic since different countries may present different demographies and we know that the COVID-19 is riskier for older populations that appear with higher proportion in developed countries. Furthermore, the level of nutrition of the population of the country may affect the probability of contracting and developing the disease. The quality of data may vary from developed countries to underdeveloped ones and, in our paper, we do not use data from other countries to calibrate our models. (2) Model: we use variations of the SIR model mentioned above. One of the advantages of the SIR model is the simplicity and researchers have used this model in several successful attempts to model the spread of infectious diseases[14–17]. (3) Estimation: our paper estimates all the parameters based on a clear hierarchical procedure based on squared error minimization.

Results

Data analysis. We use the real data provided by the Ministry of Health of Brazil from February 25, 2020 to March 30, 2020 in our estimations. If we change the final date of the period of estimation of the epidemiological parameters of the model, we note that there is a structural change in the data suggesting the effectiveness of the social distancing policy. It is worth mentioning that it is hard to know exactly when social distance measures took effect mostly because there is a variable incubation period of the virus given by a range from 2 to 10 days[18] and some initiatives of social distance measures (such as home office) started even before the official implementation of the social distancing policy. In fact, after March 23, 2020, we are able to see in the data three consecutive reductions in the first difference of the cumulative number of infections, so depending on the final date that is used for the estimation of the SIRD model, the estimated parameters cannot fit the real data anymore, as shown in Fig. 1. Thus, we define two estimation periods: (1) February 25, 2020–March 22, 2020, in which we estimate the epidemiological parameters of Eqs. (3) and (4); (2) March 23, 2020–March 30, 2020, in which we estimate the paramter ψ.

Regarding the estimation of the epidemiological parameters of Eqs. (3) and (4), we estimate all parameters of our model by minimizing the squared error of integrated variables and their real values[5,28]. We proceed in a hierarchical procedure. We start by estimating the parameters of the SIRD model, namely β, γ and ρ by minimizing the squared error

$$\min_{\beta,\gamma,\rho} \ \frac{1}{2}\left(\sum_t f\left([(I_t^{cum} - D_t) - (\hat{I}_t + \hat{R}_t)]^2\right) + f\left([D_t - \hat{D}_t]^2\right)\right), \tag{5}$$

where I_t^{cum} and D_t are the cumulative number of infected individuals and deaths, which are the real data provided by the Ministry of Health of Brazil, and $\hat{I}_t$, $\hat{R}_t$ and $\hat{D}_t$ are estimated values of the infected, recovered and deaths, respectively. We use the nonlinear function $f(z) = C^2 \log\left((g(z)/C)^2\right)$ to correct the exponential characteristic of the series so that the errors of the last values of the series do not dominate the minimization, where

Model	Parameters	Value	Other sources
SIRD	β	0.4417 (0.3695–0.6043)	–
SIRD	γ	0.1508 (0.0714–0.3295)	1/10 to 1/2[18]
SIRD	ρ	0.0292 (0.0100–0.0485)	0.049 released by WHO[19] in 2020-04-01, 0.028 in 2020-03-27 released by Brazilian Ministry of Health[20], 0.032 in 2020-03-29 released by Brazilian Ministry of Health[21] and 0.014 (0.009–0.021) in Wuhan[22]
SIRD	R_0	2.8421 (1.8142–4.9886)	3.8 (3.6-4.0)[12] and 2.68 (2.47–2.86)[23] in early stages of the disease in China. 2.76–3.25 in Italy[24] 2.28 (2.06–2.52)[25] for the passengers of the Diamond Princess cruise
SIRASD	β_S	0.4417 (0.3695–0.6043)	–
SIRASD	γ_S	0.1508 (0.0714–0.3295)	–
SIRASD	R_S	1.1807 (0.5281–1.8613)	–
SIRASD	β_A	0.4417 (0.4417–0.4417)	–
SIRASD	γ_A	0.1260 (0.1130–0.1445)	–
SIRASD	R_A	2.4209 (1.9143–2.7083)	–
SIRASD	R_0	3.6017 (2.5933–4.1529)	The same as above
SIRASD	ρ	0.0347 (0.0175–0.0527)	The same as above
SIRASD	p	0.3210 (0.2916–0.3736)	0.821 (0.798–0.845) for the passengers of the Diamond Princess Cruise[26] 0.692 (0.462–0.923) for the Japanese citizens evacuated from Wuhan[27].

Table 1. Estimated values of the epidemiological parameters. (1) In the SIRD model, $R_0 = \beta(1 - \rho)/\gamma$. In the SIRASD model, $R_A = (1 - p)\beta_A/\gamma_A$ and $R_S = p\beta_S(1 - \rho)/\gamma_S$ and $R_0 = R_A + R_S$. (2) Some parameters have not presented relevant variation in the significance level of this study. In these cases, the 90% interval includes only the value of the parameter.

$g(z) = \log(1 + z)$. Furthermore, we use the scaling parameter $C = 2$ to soft threshold between inliers and outliers. Using this procedure, we note that the estimated epidemiological parameters vary less among simulations with different random seeds.

After estimating the SIRD model, we proceed by estimating the SIRASD model. Note that we lack information on the number of asymptomatic individuals, since the clear recommendation of the Ministry of Health is to test for the virus only if one has moderate or severe symptoms. Otherwise, follow the "stay at home" policy, which recommends individuals with mild symptoms to stay at home and do not seek for medical attention. Furthermore, the mortality rate is evaluated mostly over the symptomatic ones, since the asymptomatic are in many cases not tested. Therefore, we suppose that $\beta_S = \beta, \gamma_S = \gamma$ and we keep the value of ρ. Using these parameters, and assuming that there is only one asymptomatic individual in the beginning of the simulation, we estimate the parameters β_A, γ_A and p in order to minimize the squared error

$$\min_{\beta_A,\gamma_A,p} \; \frac{1}{2}\left(\sum_t f\left([(I_t^{cum} - D_t) - (\hat{I}_{S,t} + \hat{R}_{S,t})]^2\right) + f\left([D_t - \hat{D}_t]^2\right)\right), \tag{6}$$

where I_t^{cum} and D_t are real data provided by the Ministry of Health of Brazil, the cumulative number of infected individuals and deaths, and $\hat{I}_{S,t}$, $\hat{R}_{S,t}$ and $\hat{D}_t$ are estimated values of the symptomatic infected and recovered individuals, and deaths.

Table 1 presents the epidemiological parameters of our model and some reference values. We also show other values for these epidemiological parameters obtained from other simulations in Table 5 in "Methods". Some of the lines of Table 1 deserve remarks. First, the basic reproductive number R_0 in both models are comparable to the values for China and Italy. Second, the death rate ρ is very close to the values disclosed by the Brazilian Ministry of Health and the average of international values. We point out that our estimation of the death rate uses data that presumes there are places in hospitals to treat patients with severe infections, that is the situation that is present in the data now. Depending on the government policy, we do not know whether this is true or not at the peak of infection. Third, the proportion of symptomatic individuals p is smaller than the international reference due to the Brazilian Ministry of Health policy "only test if you have strong symptoms". In fact, the same problem of underdiagnosis also seems to have happened in the early epidemics in China[29].

In the last step of the estimation procedure, in order to estimate the parameter ψ, we keep all model parameters as previously estimated and we also minimize the mean squared error using loss functions similar to the ones defined in Eqs. (5) and (6), depending on the case, in the period after March 23, 2020. Furthermore, in order to evaluate the effectiveness of the social distancing policy, we estimate a new value of ψ for each new point of the time series as shown in Table 2, where the column 2 shows the estimations of ψ for the SIRD model and column 4 shows estimations of the same parameter for the SIRASD model. Although there is a small gap between the values of ψ for different models (SIRD or SIRASD), both columns suggest that the social distance factor ψ is going down, meaning that more people are joining the government policy. According to the models, the transmission rate is reduced to approximately 62% of its original value. Table 2 also presents the effective reproductive number derived from the impact of ψ on the transmission factors.

Forecasts. Figures 2 and 3 present respectively the short-term forcasts of the SIRD and the SIRASD models, where the models incorporate the ψ factor in order to rescale the transmission factors (β, β_A and β_S) in the scenario with the social distancing policy imposed by the government. Note that Fig. 3 explicitly shows the propor-

Date	SIRD		SIRASD	
	ψ	R	ψ	R
03-23-2020	0.8182	2.325	0.8799	2.8968
03-24-2020	0.7471	2.123	0.7786	2.5633
03-25-2020	0.6639	1.887	0.6891	2.2685
03-26-2020	0.6526	1.854	0.6510	2.1433
03-27-2020	0.6464	1.837	0.6409	2.1100
03-28-2020	0.6421	1.825	0.6302	2.0747
03-29-2020	0.6356	1.806	0.6190	2.0379
03-30-2020	0.6254	1.777	0.6156	2.0267

Table 2. Estimated values of ψ for the SIRD and SIRASD models and the impact on the basic reproductive number R.
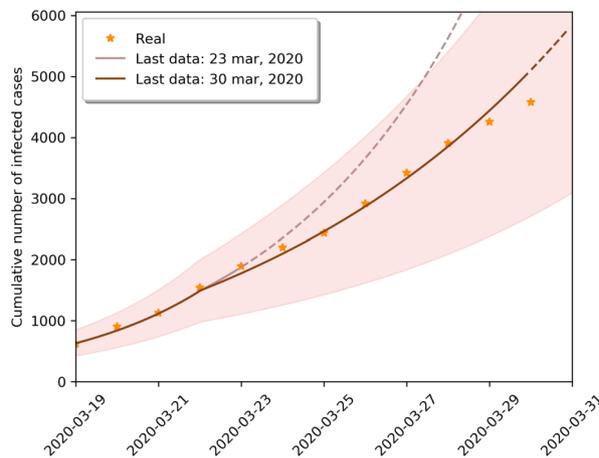
Figure 2. Short term forecast of the SIRD model taking into account government social distance measures. The solid line corresponds to the last date which the model was estimated, and the dashed line are model predictions. We show the evolution of the cumulative number of infected with 95% confidence interval. We represent the real data as points.
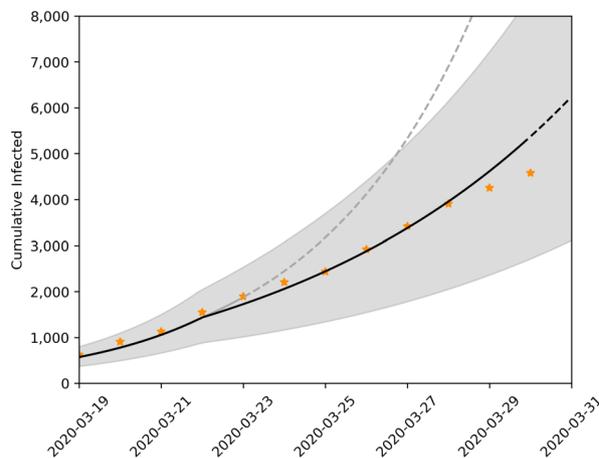
Figure 3. Short term forecast of the SIRASD model taking into account government social distance measures. The solid line corresponds to the last date which the model was estimated, and the dashed line are model predictions. We show the evolution of the infected (assymptomatic, symptomatic and both) with 95% confidence interval. We represent the real data as points.
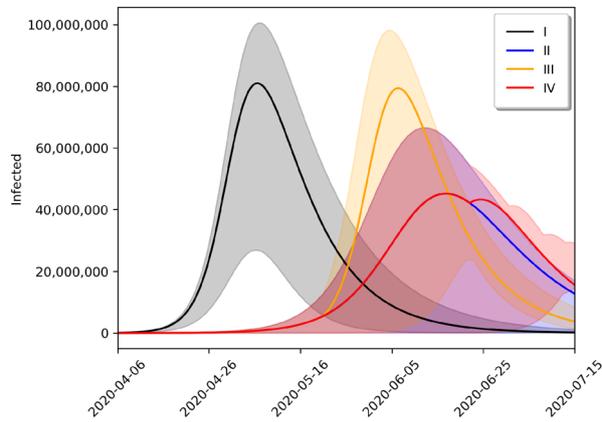
Figure 4. Long term forecasts of number of infected for different scenarios using the SIRD model. Black, blue, yellow and red lines represent scenarios I–IV, respectively.
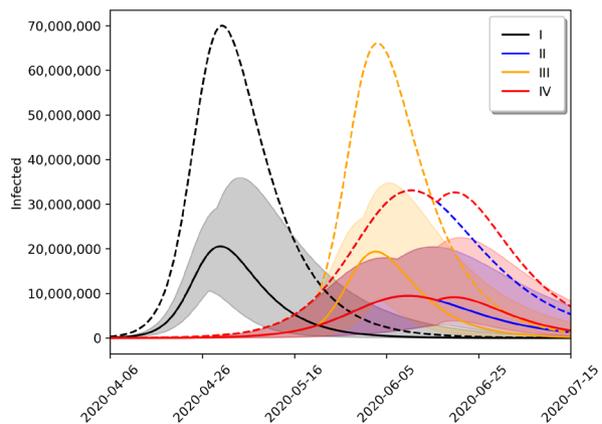

Figure 5. Long term forecasts of number of infected for different scenarios using the SIRASD model. Black, blue, yellow and red lines represent scenarios I–IV, respectively. While solid lines represent the symptomatic infected individuals, dashed lines represent total infected individuals.

tion of unknown asymptomatic individuals that when added to the symptomatic individuals skew the total value of infected individuals upwards.

We also use the SIRD and SIRASD models to provide long term forecasts of the evolution of the COVID-19 pandemic in Brazil depending on the social distancing policy considered. While Fig. 4 shows the forecasts for the SIRD model, Fig. 5 shows the forecasts for the SIRASD model. In particular, we may note that while the SIRASD model predicts that the number of infected is higher than the estimates of the SIRD model, it also predicts a lower peak for the infected with symptoms, which are the ones that could require medical attention.

We explore four cenarios: (I) no measures of social distancing policy (black line); (II) current social distancing policy imposed by the government for an indefinite time (blue line); (III) 2-month social distancing policy imposed by the government (yellow line); and (IV) optimum limited time social distancing policy imposed by the government, so that the second infection peak is not greater than cenario II (red line). Scenario III suggests that policies based on short-term social distancing policy are not enough to constrain the evolution of the pandemic, that is, if social distancing policy measurements are released before the optimal time, a second peak should be experienced. The peaks and dates in which they occur are detailed in Table 3. In the case of Scenario IV, the last day of the social distancing policy is June 22, 2020 for the SIRD model and June 16, 2020 for the SIRASD model.

In addition to Fig. 5, we also present the evolution of the proportion of asymptomatic and symptomatic in Fig. 6. In this figure, we show the instant proportion [$I_{S,t}/(I_{A,t} + I_{S,t})$ for the symptomatic and $I_{A,t}/(I_{A,t} + I_{S,t})$ for the asymptomatic] and the cumulative proportion as well. Note that the proportion of individuals who develop symptoms, p in Eq. (4), alters the transmission rate, so it also affects the evolution of the number of asymptomatic and symptomatic individuals over time. So this plot estimates the evolution of this proportion. The last column of the last line of Table 1 shows that the proportion of asymptomatic may vary from 29 to 37%, but this value is not fixed and evolves over time[26]. Our estimates suggest that the proportion of cumulative asymptomatic is approximately 68% in March 30, 2020, which converged to $1 - p$ (with p given in Table 1); that may account for some individuals with mild symptoms that were not tested.

	SIRD		SIRASD			
	Infected (I)		Infected ($I_A + I_S$)		Symptomatic (I_S)	
Scenario	Peak (%)	Date	Peak (%)	Date	Peak (%)	Date
I (Black)	38.5	May 7	33.3	April 30	9.7	April 30
II (Blue)	21.4	June 17	15.7	June 10	4.4	June 10
III (Orange)	37.7	June 6	31.4	June 3	9.2	June 3
IV (Red)	21.4	June 17	15.7	June 10	4.4	June 10

Table 3. Peaks in each scenario and the dates of occurrence.

(a)   (b)
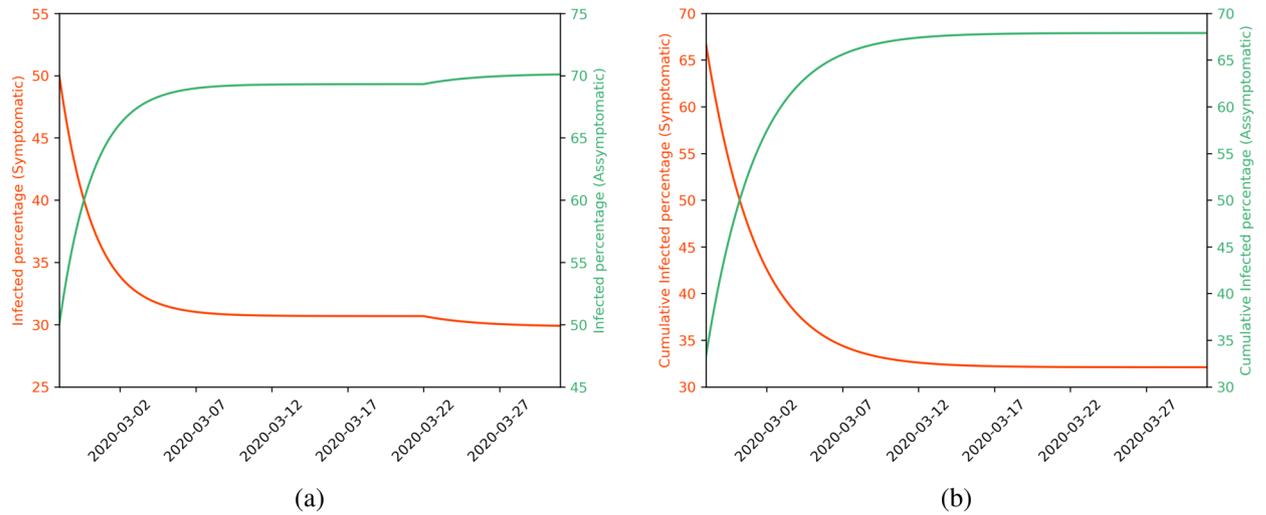
Figure 6. Proportions of asymptomatic and symptomatic over time using $I_{A,0} = 1$. We show the instant proportion of infected (left) and the cumulative number of infected (right). Approximately 70% are asymptomatic in March 30, 2020, which corresponds to 68% cumulatively or $(1 - p)$.
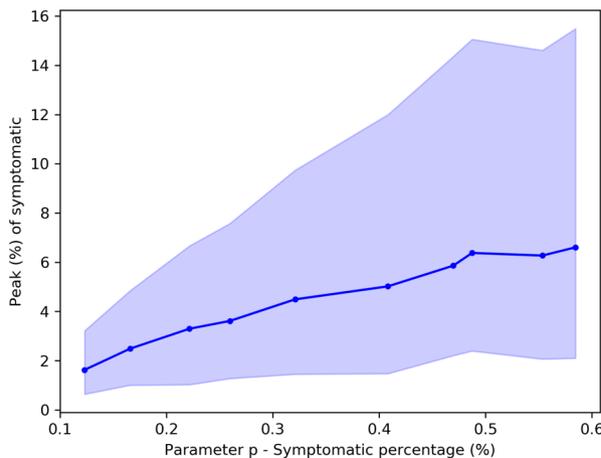
Figure 7. The effect of symptomatic percentage (parameter p) in the proportion of symptomatic in the peak.

Finally, it is worth considering that the SIRASD differential equations, presented in Eq. (4), need an initial condition for the number of asymptomatic individuals. If we find the parameters values (β_A, γ_A, p) by solving the optimization problem of Eq. (6) using different conditions, we get different results, that is, different peak values for the symptomatic individuals. If the proportion of asymptomatic individuals is larger, then this may be good news since it may represent less pressure for the health care system. But since we do not have enough tests to map the whole population, we need to work with hypotheses. Figure 7 shows the effect of different initial conditions in the symptomatic percentage and the peak value of symptomatic, that is, we vary the initial conditions, evaluate

the symptomatic proportion (parameter p in the SIRASD model), then calculate the peak value of symptomatic infected. So if we assume that the number of asymptomatic (symptomatic) individuals in data is larger (smaller) today, the number of asymptomatic (symptomatic) individuals will also be larger (smaller) in the time of the peak, leading to a smaller peak for the symptomatic.

Discussion

We use the Brazilian recent data from February 25, 2020 to March 30, 2020 to model and forecast the evolution of the COVID-19 pandemic in Brazil.

We estimate two variations of the SIR model using historical data and we find parameters that are in accordance with the international literature. We also introduce a factor ψ to account for the effect of the government social distancing measures. Our methodology is able to estimate the asymptomatic individuals, that may not be entirely present in data. Since the Brazilian government does not have enough tests for mass testing, this measure may provide some additional information. In fact, we show the relevance of the number of asymptomatic individuals, since the larger the number of asymptomatic individuals, the smaller the number hospital beds needed. The "stay at home" and "only test if you have strong symptoms" policies present contradictory effects in the disease control. While they avoid an increase in the number of infected people and the use of extra resources with people that present only mild symptoms, they reduce the amount of information about the real number of infected individuals. In particular, it explains the low value of the parameter that measures the proportion of individuals who present symptoms, since we count many individuals with mild symptoms as asymptomatic.

While our short-term forecasts are in great accordance with the data, our long-term forecasts may help us to discuss different types of social distancing policies. We also show that the social distancing policy imposed by the government is able to flatten the pattern of contamination provided by the COVID-19, but short-term policies are only able to shift the peak of infection into the future keeping the value of the peak in almost the same value. Furthermore, we define the idea of the optimal social distancing policy as the finite social distancing policy that the second peak that happens after stopping the policy is not larger than the first. Based on this definition, we provide an estimate of the optimal date to end the social distancing policy.

An important discussion is about the effectiveness of vertical containment policies, where only people at risk follow social distance policies. In these kinds of policies, the two fractions of the population, the one at risk and the other one, present very different behaviors. First, the dynamics of the population at risk behaves similarly to the case with social distancing measures, but with a higher death rate. Second, the dynamics of the population that is not at risk behaves similarly to the case without social distance measures but with a low death rate. Third, since the fraction of the population that is at risk is much smaller than the rest of the population, the number of infected of the total population behaves similarly to the case without control. In fact, the policy's effectiveness is not in reducing the number of infected, but in reducing the number of deaths by confining individuals at risk. It is worth mentioning that the effectiveness of these vertical containment polices depends strongly on the ability to separate the individuals at high risk from the individuals at low risk and on the number of vacancies in hospitals to treat the disease. We may extend our model to explore these type of scenarios and we leave for future work.

Finally, another interesting research path is to evaluate the economic side effects of pandemic control[30,31] and to propose measures to minimize these impacts[32].

Methods

The solution of the systems of differential equations.
We find the numerical solutions of Eqs. (3) and (4) through integration using the explicit Runge-Kutta method of order 5(4)[33]. While this method controls the error assuming accuracy of the fourth-order, it uses a fifth-order accurate formula to take the steps. We use the implementation "solve_ivp" of the scipy Python's library.

The solution of the systems of differential equations depends on the definition of initial conditions. We use $N_0 = 210147125$, that is the Brazilian population according to Brazilian Institute of Geography and Statistics (IBGE) which is the agency responsible for official collection of statistical, geographic, cartographic, geodetic and environmental information in Brazil, for both models. For the case of the SIRD model, we use $S_0 = N_0 - 1$ and $I_0 = 1$. For the case, SIRASD model, we use $I_{S0} = 1$ and $S0 = N_0 - I_{A0} - I_{S0}$. We use $I_{A0} = 1$ in all simulations of the paper but the simulations presented in Fig. 7, since we want to learn about the effect of I_{A0} in the proportion of symptomatic and asymptomatic individuals in the peak date.

The estimation procedure.
Our estimation procedure requires simultaneous integration of the differential equations (SIRD or SIRASD model depending on the case) and minimization of the loss functions [(5) or (6)] depending on the case for each time t. We minimize the loss functions using the method "optimize.least_squares" also from the scipy Python's library[34] using the cauchy loss with scaling parameter $C = 2$[35,36]. To minimize the impact of the initial point assumption and data incompleteness, we repeat the estimation procedure 100 times using random initial conditions, but we discarded estimations which did not converge. Since this is a difficult nonlinear problem we bound the parameters estimation region. In particular, we use the bounds presented in Table 4. To be clear, the fact that $\beta_S = \beta$ and $\gamma_S = \gamma$ is a consequence of our hierarchical estimation procedure previously described in "Results" section. Furthermore, $\beta_A \in [0, \beta_S]$ means that $\beta_A \leq \beta_S$[7], since the asymptomatic individuals do not have symptoms that may help the spread of the infection.

Finally it is worth mentioning that this estimation procedure is sensitive to the random seed used by the algorithm as an initial condition. In particular, depending on this seed, we have found different epidemiological parameters in different simulations of the SIRD model, as presented in Table 5. We have chosen the simulation results that provided the closest value of the median of the parameter γ, which is the one that used the random seed 7. We emphasize that although any of the presented epidemiological parameters could be a possible

Model	Parameter	Interval of initial conditions
SIRD	ρ	[0.01, 0.1]
SIRD	β	[1/10, 1/0.5]
SIRD	γ	[1/14, 1/2]
SIRASD	β_S	$\{\beta\}$
SIRASD	γ_S	$\{\gamma\}$
SIRASD	β_A	[1/10, β_S]
SIRASD	γ_A	[1/14, 1/2]
Both Models	ψ	[0, 1]

Table 4. Parameters estimation region.

Random seed	β	γ	ρ
7	0.441717	0.150876	0.0292182
511	0.432978	0.141411	0.0301936
1024	0.428903	0.136487	0.032915
90787	0.465757	0.176799	0.0273557
407850	0.449369	0.159107	0.029013
1905090	0.46357	0.174802	0.0277514

Table 5. Random seeds used in simulations and their respective estimated values of the epidemiological parameters.

estimation and we could use them in the main part of this paper, this choice does change the qualitative analysis and the conclusions of our paper.

The long term forecasts. The long term forecasts use the estimations presented in Table 1 and the integration of the systems of differential equations as described in the beginning of this section. We build the 95% confidence intervals of these curves randomizing the values of the parameters in the 95% confidence intervals presented in Table 1.

Data availability

The datasets and codes used to generate all the results of the current study are available in the Zenodo repository, https://zenodo.org/record/4041467#.X2joy4Zv88o.

References
1. World Heath Organization. Coronavirus overview. https://www.who.int/health-topics/coronavirus (2020).
2. Adam, D. The simulations driving the world's response to covid-19. How epidemiologists rushed to model the coronavirus pandemic?. *Nature* **20**, 20 (2020).
3. Grassly, N. & Fraser, C. Mathematical models of infectious disease transmission. *Nat. Rev. Microbiol.* **6**, 477–487 (2008).
4. Keeling, M. J. & Rohani, P. Modeling infectious diseases in humans and animals (2011).
5. Brauer, F., Castillo-Chavez, C. & Feng, Z. Mathematical models in epidemiology (2019).
6. Kermack, W. O. & McKendrick, A. G. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. A* **115**, 700–721 (1927).
7. Robinson, M. & Stilianakis, N. I. A model for the emergence of drug resistance in the presence of asymptomatic infections. *Math. Biosci.* **243**, 163–177 (2013).
8. Arino, J., van-den Brauer, F., Driessche, P., Watmough, J. & Wu, J. How will country-based mitigation measures influence the course of the covid-19 epidemic?. *J. Theoret. Biol.* **253**, 118–130 (2008).
9. Longini, I. M. Jr., Halloran, M. E., Nizam, A. & Yang, Y. Containing pandemic influenza with antiviral agents. *Am. J. Epidemiol.* **159**, 623–633 (2004).
10. Kucharski, A. J. *et al.* Early dynamics of transmission and control of covid-19: A mathematical modelling study. *Lancet Infect. Dis.* **20**, 1–7 (2020).
11. Berger, D., Herkenhoff, K. & Mongey, S. *An Seir Infectious Disease Model with Testing and Conditional Quarantine* (Tech. Rep, Federal Reserve Bank of Minneapolis, 2020).
12. Read, J. M., Bridgen, J. R. E., Cummings, D. A. T., Ho, A. & Jewell, C. P. Novel coronavirus 2019-ncov: Early estimation of epidemiological parameters and epidemic predictions. *medRxiv*https://doi.org/10.1101/2020.01.23.20018549 (2020). https://www.medrxiv.org/content/early/2020/01/28/2020.01.23.20018549.full.pdf.
13. Walker, P. G. T. *et al.The Global Impact of Covid-19 and Strategies for Mitigation and Suppression* (Tech. Rep, Imperial College, 2020).
14. Shaman, J., Karspeck, A., Yang, W., Tamerius, J. & Lipsitch, M. Real-time influenza forecasts during the 2012–2013 season. *Nat. Commun.* **4**, 2837 (2013).

15. Berge, T., Lubuma, J.-S., Moremedi, G., Morris, N. & Kondera-Shava, R. A simple mathematical model for ebola in Africa. *J. Biol. Dyn.* **11**, 42–74 (2017).
16. Osthus, D., Hickmann, K. S., Caragea, P. C., Higdon, D. & Valle, S. Y. D. Forecasting seasonal influenza with a state-space sir model. *Ann. Appl. Stat.* **11**, 202–224 (2017).
17. Khaleque, A. & Sen, P. An empirical analysis of the ebola outbreak in west Africa. *Sci. Rep.* **7**, 42594 (2017).
18. World Heath Organization. Novel coronavirus(2019-ncov). situation report-7. Tech. Rep., World Heath Organization (2020).
19. World Heath Organization. Coronavirus disease 2019 (covid-19)situation report-72. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200401-sitrep-72-covid-19.pdf?sfvrsn=3dd8971b_2 (2020).
20. Brazilian Ministry of Health. Coronavirus panel. https://covid.saude.gov.br/ (2020). Accessed 27 Mar 2020.
21. Brazilian Ministry of Health. Coronavirus panel. https://covid.saude.gov.br/ (2020). Accessed 29 Mar 2020.
22. Wu, J. T. *et al.* Estimating clinical severity of covid-19 from the transmission dynamics in Wuhan, China. *Nat. Med.* **20**, 20 (2020).
23. Wu, J. T., Leung, K. & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in Wuhan, China: A modelling study. *Lancet* **395**, 689–697 (2020).
24. Remuzzi, A. & Remuzzi, G. Covid-19 and Italy: What next?. *Lancet* **20**, 1–4 (2020).
25. Zhang, S. *et al.* Estimation of the reproductive number of novel coronavirus (covid-19) and the probable outbreak size on the diamond princess cruise ship: A data-driven analysis. *Int. J. Infect. Dis.* **93**, 201–204 (2020).
26. Mizumoto, K., Kagaya, K., Zarebski, A. & Chowell, G. Estimating the asymptomatic proportion of coronavirus disease 2019 (covid-19) cases on board the diamond princess cruise ship. *Eurosurveillance* **25**, 2000180 (2020).
27. Nishiura, H. *et al.* Estimation of the asymptomatic ratio of novel coronavirus infections. *Forthcom. Int. J. Infect. Dis.* **20**, 20 (2020).
28. Bard, Y. *Nonlinear Parameter Estimation* (1974).
29. Nishiura, H. *et al.* The rate of underascertainment of novel coronavirus (2019-ncov) infection: Estimation using Japanese passengers data on evacuation flights. *J. Clin. Med.* **4**, 419 (2020).
30. Eichenbaum, M. S., Rebelo, S. & Trabandt, M. The macroeconomics of epidemics. Working Paper 26882, National Bureau of Economic Research (2020). https://doi.org/10.3386/w26882.
31. Gormsen, N. J. & Koijen, R. S. J. Coronavirus: Impact on stock prices and growth expectations. *Working Paper of the University of Chicago* 1–27 (2020).
32. Hone, T. *et al.* An empirical analysis of the ebola outbreak in west Africa. *Lancet* **7**, 1575–1583 (2019).
33. Dormand, J. R. & Prince, P. J. A family of embedded Runge-Kutta formulae. *J. Comput. Appl. Math.* **6**, 19–26 (1980).
34. Virtanen, P. *et al.* SciPy 1.0. Fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
35. Mayorov, N. Robust nonlinear regression in scipy (Tech, Rep, 2015).
36. Triggs, B., McLauchlan, P. F., Hartley, R. I. & Fitzgibbon, A. W. *IWVA 1999: Vision Algorithms: Theory and Practice, chap. Bundle Adjustment—A Modern Synthesis* (Springer, Berlin, 1999).

Acknowledgements

Author contributions

S.B.B. and D.O.C. designed the research, performed the research and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to D.O.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.