Modeling epidemic growth curves using nonlinear rational polynomial equations: an application to Brazil's Covid-19 data

ABSTRACT

This paper reports a broad study using epidemic-related counting data of COVID-19 disease caused by the novel coronavirus (SARS-CoV-2). The considered dataset refers to Brazil's daily and accumulated counts of reported cases and deaths in a fixed period (from January 22 to June 16, 2020). For the data analysis, it has been adopted a nonlinear rational polynomial function to model the mentioned counts assuming Gaussian errors. The least-squares method was applied to fit the proposed model. We have noticed that the curves are still increasing after June 16, with no evidence of peak being reached or decreasing behavior in the period for new reported cases and confirmed deaths by the disease. The obtained results are consistent and highlight the adopted model's capability to accurately predict the behavior of Brazil's COVID-19 growth curve in the observed time-frame.

INTRODUCTION

An epidemic could be defined as the community occurrence of a group of illnesses of similar nature derived from a common or a propagated source (GORDIS, 2014). When an epidemic grows on a worldwide scale, then it is called a pandemic. In human history, epidemic diseases have killed millions of people at different times at varying speeds. Viruses have caused some of the worst registered epidemics, as smallpox (1896 to 1980), Spanish flu (1918 to 1919), measles (until 1963), malaria (since 1980), AIDS (since 1981), and dengue (over the last thirty years) (World Health Organization, 2020b). In 2009, cases of Influenza A (H1N1) were registered worldwide, and an outbreak of Ebola occurred between 2013 to 2016 in West Africa. Influenza A, which is genetically close to the Spanish flu, has caused approximately 284,000 deaths (DAWOOD et al., 2012), while Ebola caused “only” 11,300 deaths (World Health Organization, 2016). According to (Ministério da Saúde do Brasil, 2016), Influenza A has victimized approximately 1,800 people at its peak in Brazil.

Over the last few years, Brazil’s major epidemic diseases were caused by Chikungunya (2013) and Zika (2015) viruses. These diseases had produced a vast international commotion as the number of new positive cases grew up considerably precisely when the country was about to host two of the most prominent sports events worldwide (the 2014 FIFA World Cup and the 2016 Summer Olympics Games). The main concern relied on the fact that such events had the potential to attract a massive number of tourists and that the Zika virus affects pregnant women, causing microcephaly and other equally severe congenital malformations in children (BURATTINI, 2016).

Despite all the battles won against previous epidemics and pandemics, the world is currently experiencing one of the most significant challenges of the 21st century, the pandemic caused by the novel coronavirus. The COVID-19 disease is a viral infection of disruptive nature spreading quickly around the world since January 2020. The virus has emerged late in 2019 in the city of Wuhan, Hubei Province of China. According to the Situation Report 80 from the World Health Organization (WHO), the number of global cases reached 8,043,487 people until June 16, 2020, among which 439,487 died with COVID-19. The incidence of new cases had increased and is increasing exponentially in many countries around the globe.

When the epidemic started (late 2019), the WHO was notified about a series of pneumonia cases detected in Wuhan by unknown causes. Later, however, the coronavirus was identified as the causative virus by Chinese authorities on January 07, 2020 (World Health Organization, 2020a). On January 30, 2020, following the Emergency Committee’s recommendations, the WHO declared that the outbreak constitutes a Public Health Emergency of International Concern (PHEIC). Since December 2019, the pandemic of COVID-19 has spread to new areas and has increased considerably in the already affected areas. In the Situation Report 1 (January 21, 2020), there were only 282 confirmed cases in the region covering China, Japan, the Republic of Korea, and Thailand. However, in Situation Report 77
(April 06, 2020), there were 1,210,956 confirmed cases and 67,594 deaths worldwide, being South Sudan, the last country to register cases of COVID-19.

The disease, initially called COVID-19 (Coronavirus Disease 2019), started to have the causative virus classified as SARS-CoV-2 by the International Virus Taxonomy Committee (CHEN; LIU; GUO, 2020). To track the virus, the WHO has updated the Laboratory Testing Strategy at March 21, 2020 (LAN et al., 2020) according to the different transmission scenarios: countries with no cases; countries with one or more cases (sporadic cases); countries experiencing a series of cases related to geographic location or common exposure (a group of cases); and countries experiencing massive outbreaks or sustained and pervasive local transmission (community transmission).

COVID-19 is a respiratory disease that affects different people in different ways. Most people infected with the novel coronavirus will experience mild to moderate respiratory illness, recovering without requiring specific treatments (World Health Organization, 2020a). According to the WHO, the primary symptoms of COVID-19 are fever, tiredness, dry cough, shortness of breath, and sore throat. Very few people will report diarrhea, nausea, or runny nose. Older people and those with underlying medical conditions like cardiovascular disease, diabetes, chronic respiratory disease, and cancer are more likely to develop severe illness and death. The virus spreads primarily through droplets of saliva or discharge from the nose when an infected person coughs or sneezes. At this time, there are no specific vaccines or treatments for COVID-19. However, several ongoing clinical trials are being conducted to find potential treatments and drug repurposing opportunities.

Many researchers are putting great effort into the last months to understand the behavior of the novel coronavirus better. It can be seen in the hundreds of scientific works being released in a short time: Kandel et al. (2020), Pung et al. (2020), Chan et al. (2020), Huang et al. (2020), Wu et al. (2020), Lu et al. (2020), Chen, Liu e Guo (2020), Li et al. (2020), Lai et al. (2020), Lupia et al. (2020), Shereen et al. (2020), Chen, Liu e Guo (2020), Sohrabi et al. (2020), Han et al. (2020), Chen, Liu e Guo (2020), Wu et al. (2020), and Zhao et al. (2020). Most of these papers are related to the transmission of the virus, to the genomic characterization and the epidemiology of the disease, to new specific treatments, to clinical features of patients infected with the virus, to developments of vaccines, to decrease the lethality rate, especially for the elderly, and to the effects of the confinement to either minimize the spread of the disease and consequently not overburden the health systems.

Moreover, many recent studies indicates that the researches related to COVID-19 emerges from many different fields of study as viral origin and structure (LAN et al., 2020; SHANG et al., 2020; LAM et al., 2020), epidemiology (FERRETTI et al., 2020), preclinical research (KIM et al., 2020), diagnostic and serology (JU et al., 2020; WÖLFEL et al., 2020), and therapy and clinical trials (SHEN et al., 2020). Studies involving the forecasting for the COVID-19 worldwide pandemic can be found in Ribeiro et al. (2020), Boccaletti et al. (2020), Zhang, Ma e Wang (2020), Postnikov (2020), Chakraborty e Ghosh (2020), Ndaïrou et al. (2020), and Barmparis e Tsironis (2020).

According to Rafael et al. (2020), the slower the rate of progress of an epidemic, the longer its duration. In this case, however, health services will be
more responsive as they will not always be overburdened. For this reason, social structure interventions are essential and urgent measures to face epidemics of this magnitude. Social isolation is a restrictive measure for controlling the growth of the curve of COVID-19. Monitoring the behavior of the epidemic curve allows us to predict each region’s epidemiological scenario and so to anticipate public policies and specific assistance to cope with the progress of the disease. In addition to social detachment, the international experience has indicated essential strategies for containing the epidemiological curve’s growth.

Among those strategies, the ones that proved to be the most effective were: expanding the testing of suspected cases (with quick delivery of the results), identification of contaminated people (with immediate home isolation), and investments aiming to protect health professionals (World Health Organization, 2020a). It has been reported that South Korea has adopted very restrictive measures to control the spread of COVID-19. Fortunately, they are having great success in controlling the spreading of the disease. However, some countries, such as Italy, Spain, the United Kingdom, and the US, were slow to take restrictive measures, including social isolation. In this way, these countries already reached tens of thousands of deaths. There are still divergences about attempting to control Brazil's epidemiological curve, mainly because some authorities keep underestimating the problem’s gravity.

In this paper, we propose a nonlinear statistical model to describe the behavior of the daily and accumulated curves of the reported cases and deaths by COVID-19 in Brazil between January 22 to June 16, 2020. This approach could be a useful alternative to Richard’s model (HSIEH; LEE; CHANG, 2004) for epidemiology studies. The proposed model’s primary advantage is the dynamic of the iteration process and the use of polynomial functions.

This paper is organized as follows. Section 2 presents an overall description of the considered datasets and the proposed nonlinear model for the analysis of the growth curves. The obtained results and the pertinent discussions are presented in Section 3. General comments and concluding remarks are addressed in Section 4.

**MATERIAL AND METHODS**

**Datasets**

The Center of Systems Science and Engineering (CSSE) of Johns Hopkins University (Baltimore, US) collects and manages data related to the COVID-19 pandemic since January 22. The data involve the number of new daily positive cases and deaths. The purpose of monitoring the disease’s advance is to define new strategies to control the pandemic and make short-time predictions (for days or weeks). The datasets owned to CSSE were made available in a GitHub directory (<https://github.com/CSSEGISandData/COVID-19>). Our analysis has considered the number of daily and accumulated reported cases in Brazil between January 22 to June 16, 2020.

**Nonlinear statistical model**

Statistical modeling of daily and accumulated disease counting data can be considered under different approaches. From a probabilistic point of view, the
epidemic curves related to disease counting data could be modeled as a stochastic process (in the form of a counting process). Alternatively, one may consider using classical nonlinear models as those widely used to describe phenomena as population growth in ecology and demography or the individual body height or biomass (for growth analysis of subjects in physiology). The Richards and Gompertz models are among the most frequently used tools for the analysis of growth data. Standard inference methods to obtain point and interval estimates for the parameters of growth models are well discussed within the nonlinear modeling literature (BATES; WATTS, 1980; RATKOWSKY, 1983; BATES; WATTS, 1988; HAZEWINNEL, 2001; SEBER; WILD, 2003).

Our approach in this paper is based on the proposition of a nonlinear regression structure to describe epidemic curves’ growth. Conceptually, nonlinear and linear regression models are similar since the underlying methodology is based in relate a response variable \( y \) to a vector of covariates \( x = (x_1, ..., x_k)^T \). Nonlinear regression is featured because the prediction equation induces nonlinearity in one or more unknown parameters. Unlike linear regression, nonlinear models usually arise when there is some physical reason implying that the relationship between the response and the predictors follows a particular functional form. A nonlinear regression model has the general form

\[
y_i = f(x_i; \theta) + \varepsilon_i
\]  

(1)

where \( y_i \) the \( i \)-th observed response \( (i = 1, ..., n) \), \( x_i = (x_{i1}, ..., x_{ik})^T \), is the vector of covariates, \( \theta = (\theta_1, ..., \theta_p) \) is the vector of \( p \) unknown parameters, and \( \varepsilon_i \) is the stochastic error. The errors are usually assumed to be uncorrelated \( (\text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \text{ for all } i \neq j ) \) and to be Gaussian distributed with zero mean and constant variance \( \sigma^2 \). In order to describe the epidemic curves of COVID-19 incidence of new positive cases and deaths in Brazil, we have assumed a rational polynomial function indexed by the vector \( \theta = (\theta_1, \theta_2, \theta_3, \theta_4) \) for the nonlinear component of model (1), that is,

\[
y_i = \frac{\theta_1 + \theta_2 x_i}{1 + \theta_3 x_i + \theta_4 x_i^2} + \varepsilon_i
\]  

(2)

where the \( x_i \) sequential label of the day in which \( y_i \) was observed, starting on January 22. A typical problem when fitting a nonlinear model is finding good starting values. In this way, the primary advantage of rational function models is the possibility of obtaining starting values by using a linear least-squares fit (HAZEWINNEL, 2001).

Estimates for the parameters of a nonlinear regression model can be obtained by using iterative procedures based on optimization methods to minimize \( \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} [ y_i - f(x_i; \theta)]^2 \). A popular iterative technique to find the least-squares estimator of nonlinear models is the Gauss-Newton algorithm. This algorithm increments the working estimate of \( \theta \) at each iteration by an amount equal to the coefficients from a linear regression based on the current residual sand the current gradient matrix \( V \). If the function \( f \) in (1) is continuously differentiable in \( \theta \), then it can be linearized locally as

\[
f(x_i; \theta) = f(x_i; \theta_0) + V_0(\theta - \theta_0),
\]
where \( \theta_0 \) is the vector of initial values for the iterative procedure and \( V_0 \) is the \( n \times p \) gradient matrix whose elements are given by \( \frac{\partial f(x_i; \theta_0)}{\partial \theta_j} \) for \( j = 1, \ldots, p \). Such specification leads to the Gauss-Newton algorithm to obtaining updates for \( \theta \) as

\[
\theta_k = \theta_{k-1} + (V_0^T V_0)^{-1} V_0^T e, \quad k = 1, 2, \ldots
\]

where \( e = (\epsilon_1, \ldots, \epsilon_n)^T \) is the vector of working residuals. If the errors \( \epsilon_i \) are independent and normally distributed, then the Gauss-Newton algorithm is an application of the Fisher’s scoring method.

**RESULTS AND DISCUSSION**

Using data from daily and accumulated counts of reported cases and deaths caused by COVID-19 in Brazil between January 22 to June 16, 2020, we have fitted the regression model (2) assuming a Gaussian distribution with zero mean and constant variance \( \sigma^2 \) for the random errors \( (\epsilon_i) \). The statistical analysis was carried out in the R software (R Core Team, 2019). The `nlstools` package was used to obtain the least-squares estimates, and the function `confint` was used to compute the Profile Likelihood Confidence Intervals (PLCI) for the model parameters. The results for each fitted model are presented in Table 1.

### Table 1 – Summary of the fitted nonlinear models for daily and accumulated counts of reported cases and deaths caused by COVID-19 in Brazil between January 22 to June 16, 2020.

<table>
<thead>
<tr>
<th>Counts</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Std.Error</th>
<th>95% PLCI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lower</td>
<td>Upper</td>
</tr>
<tr>
<td>Daily</td>
<td>( \theta_1 )</td>
<td>-423.66550</td>
<td>268.34891</td>
<td>-977.67377</td>
</tr>
<tr>
<td></td>
<td>( \theta_2 )</td>
<td>9.83427</td>
<td>2.98697</td>
<td>4.47918</td>
</tr>
<tr>
<td></td>
<td>( \theta_3 )</td>
<td>-0.01421</td>
<td>0.00018</td>
<td>-0.01452</td>
</tr>
<tr>
<td></td>
<td>( \theta_4 )</td>
<td>0.00005</td>
<td>0.00001</td>
<td>0.00004</td>
</tr>
<tr>
<td>Accumulated</td>
<td>( \theta_1 )</td>
<td>-13075.37819</td>
<td>961.54491</td>
<td>-14900.20738</td>
</tr>
<tr>
<td></td>
<td>( \theta_2 )</td>
<td>247.46234</td>
<td>10.78733</td>
<td>227.82037</td>
</tr>
<tr>
<td></td>
<td>( \theta_3 )</td>
<td>-0.01322</td>
<td>0.00004</td>
<td>-0.01329</td>
</tr>
<tr>
<td></td>
<td>( \theta_4 )</td>
<td>0.00004</td>
<td>0.00001</td>
<td>0.00004</td>
</tr>
<tr>
<td>Daily</td>
<td>( \theta_1 )</td>
<td>-34.54866</td>
<td>14.14523</td>
<td>-62.42579</td>
</tr>
<tr>
<td></td>
<td>( \theta_2 )</td>
<td>0.73378</td>
<td>0.1752</td>
<td>0.43380</td>
</tr>
<tr>
<td></td>
<td>( \theta_3 )</td>
<td>-0.01492</td>
<td>0.00022</td>
<td>-0.01532</td>
</tr>
<tr>
<td></td>
<td>( \theta_4 )</td>
<td>0.00006</td>
<td>0.00001</td>
<td>0.00005</td>
</tr>
<tr>
<td>Accumulated</td>
<td>( \theta_1 )</td>
<td>-942.73939</td>
<td>80.38882</td>
<td>-1092.56250</td>
</tr>
<tr>
<td></td>
<td>( \theta_2 )</td>
<td>17.37437</td>
<td>0.94618</td>
<td>15.71019</td>
</tr>
<tr>
<td></td>
<td>( \theta_3 )</td>
<td>-0.01338</td>
<td>0.00006</td>
<td>-0.01348</td>
</tr>
<tr>
<td></td>
<td>( \theta_4 )</td>
<td>0.00005</td>
<td>0.00001</td>
<td>0.00004</td>
</tr>
</tbody>
</table>

Source: The Authors (2020)
Figure 1 – Upper-panels: Model fit for the number of new positive cases (left-panel) and the number of deaths (right-panel) by COVID-19 in Brazil between January 22 to June 16, 2020. Lower-panels: Half-normal plots with simulated envelope for the ordered absolute standardized residuals (from the model based on the number of positive cases on the left and for the number of deaths on the right).

Source: The Authors (2020)
Figure 1 illustrates (upper-panels) the estimated growth curves of daily counts of reported cases and deaths caused by COVID-19 in Brazil between January 22 to June 16, 2020. The curves are still increasing after June 16, with no evidence of peak being reached or decreasing behavior in the period for new reported cases and confirmed deaths by the disease. These results highlight the adopted model’s capability to predict Brazil’s COVID-19 growth curve’s behavior in the observed period. We also present the Half-normal plots with simulated envelopes (lower-panels) for the estimated standardized residuals to check the model assumptions. Noticeably, the obtained fits were very satisfactory since all estimated residuals are lying within the simulated envelope. Similar results were observed for the accumulated counts of reported cases and deaths, with outcomes related to the estimated growth curve and the residual analysis presented in Figure 2.

Figure 2 –Upper-panels: Model fit for the accumulated number of new positive cases (left-panel) and the number of deaths (right-panel) by COVID-19 in Brazil between January 22 to June 16, 2020. Lower-panels: Half-normal plots with simulated envelope for the ordered absolute standardized residuals (from the model based on the accumulated number of positive cases on the left and for the accumulated number of deaths on the right).
Despite our study’s theoretical direction, our final message relies on reinforcing the importance of respecting social detachment during the quarantine time adopted by countries following the WHO directions and health authorities’ advice. For instance, several studies based on the COVID-19 cases in the US indicate that the virus’s degree of transmission is much higher than imagined: a single infected person can transmit the virus to approximately six people (NDIAYE; TENDENG; SECK, 2020). Since there is still no specific treatment for COVID-19, the Brazilian authorities must keep a strict stance towards social isolation and preservative actions. Otherwise, the country will keep registering a daily number of deaths similar to countries hugely affected by the disease, like Italy, France, Spain, and the US.

CONCLUDING REMARKS

This paper’s main goal was to propose a statistical model to understand the disease’s epidemic growth curve caused by the novel coronavirus in Brazil. For that, it has been considered a nonlinear regression structure based on a rational polynomial to describe the behavior of daily and accumulated numbers of reported cases and deaths by COVID-19 between January 22 to June 16, 2020, in the whole country. Our approach was based on adopting the day’s sequential label from which the observations were taken (beginning from January 22) as a covariate. The proposed methodology can be extended using other factors, such as the daily rate of social isolation. When available, the inclusion of additional covariates may provide more accurate model fits, whose underlying results may offer suggestions on how fast the virus will spread in the short-term period. The proposed nonlinear model has been shown to accommodate the nature of the incidence and deaths by the novel coronavirus in Brazil. Notwithstanding, the proposed methodology can also be applied either to COVID-19 data from other countries as toother disease’s epidemic growth curve.
ACKNOWLEDGEMENTS

The research of Jorge A. Achcar is supported by the Brazilian organization CNPq(301923/2019-1). Josmar Mazucheli gratefully acknowledges the Paraná Research Foundation (Grant: 064/2019 - UEM/Fundação Araucária). The research of Wesley Bertoli is supported by the Federal University of Technology – Paraná.

REFERENCES


