ORIGINAL PAPER

Statistics for COVID-19 Pandemic Data

Study of the trend pattern of COVID-19 using spline-based time series model: a Bayesian paradigm

Jitendra Kumar[1] · Varun Agiwal[2] · Chun Yip Yau[3]

Abstract

A vast majority of the countries are under economic and health crises due to the current epidemic of coronavirus disease 2019 (COVID-19). The present study analyzes the COVID-19 using time series, an essential gizmo for knowing the enlargement of infection and its changing behavior, especially the trending model. We consider an autoregressive model with a non-linear time trend component that approximately converts into the linear trend using the spline function. The spline function splits the series of COVID-19 into different piecewise segments between respective knots in the form of various growth stages and fits the linear time trend. First, we obtain the number of knots with their locations in the COVID-19 series to identify the transmission stages of COVID-19 infection. Then, the estimation of the model parameters is obtained under the Bayesian setup for the best-fitted model. The results advocate that the proposed model appropriately determines the location of knots based on different transmission stages and know the current transmission situation of the COVID-19 pandemic in a country.

Keywords COVID-19 · Linear and non-linear trend · Spline function, autoregressive time series model · Bayesian inference

Mathematics Subject classifications 62F15 · 37M10 · 65C05

✉ Varun Agiwal
  varunagiwal.stats@gmail.com

  Jitendra Kumar
  vjitendrav@gmail.com

  Chun Yip Yau
  cyyau@sta.cuhk.edu.hk

[1] Department of Statistics, Central University of Rajasthan, Ajmer, India

[2] Department of Community Medicine, Jawaharlal Nehru Medical College, Ajmer, India

[3] Department of Statistics, Chinese University of Hong Kong, Hong Kong, People's Republic of China

1 Introduction

The 2019 novel coronavirus (COVID-19) is getting a lot of attention, because it is a new kind of pandemic disease that affects everywhere in the universe. Million of people have died from this disease, and million of confirmed cases are recorded worldwide because of the nonexistence of antiviral drugs and vaccines. Researchers develop various methodologies to analyze and control the spread of COVID-19 and predict the future perspective of coronavirus cases. Jiang et al. (2020) established the time series and kinetic model for infectious diseases, and predicted the trend and short-term prediction of the transmission of COVID-19. Al-Rousan and AL-Najjar (2020) analyzed the consequence of various factors such as sex, region, infection reasons, and birth year on recovered and deceased cases of the South Korea region. The results found that sex, region, and infection reasons affected both recovered and dead cases, while birth year only involved in deadly patients. Gondauri et al. (2020) considered the chain-binomial type Bailey's model for studying and analyzing the correlation between the total volumes of COVID-19 virus spread and recovery for the different countries. Most of the studies investigate the growth of COVID-19 cases based on various regression and time series models, because these models are frequently applied to examine the development or trend of any disease.

Under COVID-19 pandemic situation, series of daily recorded cases having non-linear characteristics such as shifting behavior, non-stationary, etc., shows a non-linear trend pattern due to enter in different stages of transmission. This non-linear trend may be casted by a piecewise time series model with a high order of polynomial-time way. The spline function is the alternative to deal with such a time trend polynomial in a piecewise form. It is analyzed period wise discontinuity by fitting a polynomial of a high order and join at knots. Knots are the points when there are sudden up and down in the series, and the result is a piecewise smooth time function. Eubank (1999) observed that the smoothest piecewise polynomial is a spline function that holds a segmented nature at present. Still, Hurley et al. (2006) called splines as continuous and smooth lines or curves function. Morton et al. (2009) considered a smoothing spline function to analyze the trend of generalized additive models with correlated errors and applied to data from a chemical process and to stream salinity measurements. Montoril et al. (2014) studied the estimation of the functional-coefficient regression model by splines with autoregressive errors and showed the convergence rates of the proposed estimator. Qiao et al. (2015) looked at a B-spline model on the durability of changes in the frequency signal over time. Conrad et al. (2017) modeled the forced expiratory volume 1 (FEV1) data from cystic fibrosis (CF) and chronic obstructive pulmonary disease (COPD) using median regression splines. Osmani et al. (2019) used the B-spline and kernel methods to estimate the model coefficients and showed the application for psoriasis patient's data.

In this paper, we study the trend pattern of the COVID-19 series using an autoregressive model with a trend approximated by a linear spline function. Identification of the number of knots and their locations are obtained using Bayes

factor and posterior probability, respectively. We use appropriate priors for model parameters to determine the posterior distribution and find the conditional posterior distribution for making inferences about the parameters. We apply the Metropolis–Hastings (M–H) algorithm within the Gibbs sampler to generate posterior samples from the conditional posterior distribution and get the Bayesian estimation for unknown parameters. The number and location of knots within a country explain the stages of transmission and the time points for achieving. Thus, this study gives an overview of the present trend of daily recorded COVID-19 cases and provides the current transmission stage of the most affected countries.

2 Model specification

A time series model is popularly known to regulate the trend pattern for the series of coronavirus (COVID-19). The COVID-19 transmission is mainly people to people contact and attains four stages: stage-1 (imported cases), stage-2 (local transmission), stage-3 (community transmission), and stage-4 (transmission out of control) as per the infection trend. Trend/growth patterns of the daily COVID-19 recorded cases are not always linear due to attain different transmission stages. Hence, this reaches a non-linear form, because more variability observes in the growth rate of COVID-19 infected cases. Thus, rather than using a linear process, a non-linear model is more suitable to study the trend scenario. The present paper expresses the non-linear trend using the spline function. Spline function is a piecewise polynomial segment, join together at knots based on shifts-or-quantal-jumps. The knot is a common fusion point that occurs when pattern behaviors at different intervals are changed. This change is also seen in the COVID-19 series based on different transmission stages, because the infected person's rate within a country depends upon various steps taken by the government and administration. Thus, knots are useful to determine a country is achieved which stage of transmission. For example, the GDP of India during demonetization goes down. Therefore, the GDP series pattern transforms at different periods, such as before, after, and in-between demonetization. Hence, there is a need to determine the exact location of the point (knots) where the series structure is changed. The detailed discussion to select the location of knots is discussed in Denison (1998), Biller (2000), and Ülker and Arslan (2009). Recently, this model is discussed by Kumar et al. (2020) for testing the unit root hypothesis in the presence of spline function through posterior odds ratio and applied in monthly import series of ASEAN Regional Forum (ARF) countries. The complete detail about this model is well described by Kumar et al. (2020). Here, we only write the key expression of the model. Let $\{y_t: t=1, 2, …, T\}$ is a time series from the model given in Eq. (1):

$$y_t = (1 - \rho)\phi + \rho y_{t-1} + \beta\, t + \sum_{i=1}^{r} \psi_i \big[s_i(t) - \rho s_i(t-1)\big] + \varepsilon_t, \qquad (1)$$

where $\phi = \delta_0 + \frac{\rho}{1-\rho}\delta$ and $\beta = (1 - \rho)\delta$.

In Eq. (1), ρ is the autoregressive coefficient, δ_0 is the intercept term, δ is the trend coefficient, r is the number of knots that contain the location of knots $t_1, t_2, \ldots, t_r$, and ψ_i is the coefficient of the ith knot. The ε_t's are *i.i.d.* normally distributed random variables with mean zero and unknown variance τ^{-1} and $s_i(t)$ is a spline function, describes as a linear polynomial form

$$s_i(t) = (t - t_i)^+ = \begin{cases} t - t_i & \text{if} \quad t > t_i \\ 0 & \text{if} \quad t \le t_i \end{cases}.$$

In the existing literature, researchers do the modeling of the COVID-19 series based on various regression and time series models but ignore the irregular behavior of daily conformed cases, because most countries take necessary steps to control the spread of COVID-19. These steps change the growth of COVID-19 infected patients in an up and down manner. As a result, the trend pattern is not linear form, and there is an occurrence of sudden jumping phenomena in the series at different stages of transmission. Thus, there is a need to apply other non-linear piecewise models that provide better results based on various transmission stages. The proposed model is very suitable to analyze the non-linear trend pattern of the COVID-19 series, because this model splits the series into a linear format at their knot locations to observe the infection growth using the stages of transmission (Liu, 2009, Lusa & Ahlin, 2020). In matrix notations, the model is marked as

$$y = \rho y_{-1} + Z(\rho)\gamma + S(\rho)\psi + \varepsilon, \tag{2}$$

where

$$y = \left(y_1 \ y_2 \ \ldots \ y_T \right)', \quad y_{-1} = \left(y_0 \ y_1 \ \ldots \ y_{T-1} \right)',$$

$$\xi_T = (1 \ 2 \ldots T)', \quad l_T = \left(1 \ 1 \ \cdots \ 1 \right)',$$

$$Z(\rho) = \left((1 - \rho)l_T \ \xi_T \right), \quad S(\rho) = (I - \rho L)\Gamma,$$

$$\psi = \left(\psi_1 \ \psi_2 \ \ldots \ \psi_r \right)', \quad \varepsilon = \left(\varepsilon_1 \ \varepsilon_2 \ \ldots \ \varepsilon_T \right)',$$

$$\Gamma = \begin{bmatrix} s_1(1) & s_2(1) & \cdots & s_r(1) \\ \vdots & \vdots & \ddots & \vdots \\ s_1(t_1) & s_2(t_1) & \cdots & s_r(t_1) \\ s_1(t_1 + 1) & s_2(t_1 + 1) & \cdots & s_r(t_1 + 1) \\ \vdots & \vdots & \ddots & \vdots \\ s_1(t_2) & s_1(t_2) & \cdots & s_r(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ s_1(T) & s_2(T) & \cdots & s_r(T) \end{bmatrix}, \quad \gamma = \begin{pmatrix} \phi \\ \beta \end{pmatrix},$$

and L is $T \times T$ matrix with the entire $(i+1)$th row and the ith column elements equal to 1, and the remaining elements are equal to 0 and I is $T \times T$ identity matrix.

The main objective is to study the tendency of daily COVID-19 confirmed cases by fitting this model in a piecewise form and understand the spread of infection in terms of different transmission stages. For this, the number of knots with their locations is determined using posterior probability in each country COVID-19 series to know the transmission stages. These study countries are the USA, India, Brazil, Russia, South Africa, and Peru. Then, Bayesian estimators of the model parameters are derived using the conditional posterior distribution and obtain the estimated values using numerical techniques.

3 Bayesian estimation

For analysis purposes, a Bayesian approach is used to make inferences about the unknown parameter and draw a better inference. In the Bayesian approach, the posterior distribution is the product of the likelihood function and prior distribution. Here, the discrete uniform prior is assumed for the location of knots under consideration of all ordered sub-sequences of $(2, 3,…,T)$ of length r, i.e., $\pi(\mathbf{t}|r) = \frac{1}{T-1}C_r = \frac{r!(T-r-1)!}{(T-1)!}$ for any set $\mathbf{t}=(t_1,..., t_r)$ for given r. The remaining model parameters $(\rho, \psi, \gamma, \tau)$ consider similar prior distributions, as described in Kumar et al. (2020). The form of prior distribution of the parameters is

$$\pi(\tau) \propto \frac{1}{\tau}, \quad 0 < \tau < \infty$$

$$\pi(\rho) = \frac{1}{1-a}, \quad a < \rho < 1, \quad a > -1$$

$$\psi \sim \text{MN}\left(\psi_0, \frac{1}{\tau}\Omega^{-1}\right), \quad \psi, \psi_0 \in R, \quad \Omega > 0$$

$$\gamma \sim \text{MN}\left((1-\rho)\phi_0, \frac{1}{\tau}V(\rho)^{-1}\right) \text{ with } \phi_0 = \begin{pmatrix} y_0 \\ 0 \end{pmatrix}, \quad y_0 \in R$$

$$\text{and } V(\rho) = \begin{pmatrix} \frac{1+\rho}{1-\rho} & 0 \\ 0 & \frac{\vartheta}{(1-\rho)^2} \end{pmatrix}, \quad \vartheta > 0.$$

Then, the posterior specification for this model is

$$\pi(\Theta|y) \propto \frac{\tau^{\frac{T+r}{2}}|\Omega|^{\frac{1}{2}}|V(\rho)|^{\frac{1}{2}}}{{}^{T-1}C_r(2\pi)^{\frac{T+r}{2}+1}} \exp\left[-\frac{\tau}{2}\left\{(y - \rho y_{-1} - Z(\rho)\gamma - S(\rho)\psi)'(y - \rho y_{-1} - Z(\rho)\gamma - S(\rho)\psi)\right.\right.$$

$$\left.\left. + (\psi - \psi_0)'\Omega(\psi - \psi_0) + (\gamma - (1-\rho)\phi_0)'V(\rho)(\gamma - (1-\rho)\phi_0)\right\}\right].$$

$$(3)$$

For parameter estimation under the Bayesian setup, a loss function is used to select the best estimator based on the posterior distribution that minimizes the risk associated with each parameter. Here, we consider the squared error loss function (SELF) as a symmetric loss function. Under this loss function, the Bayesian estimator of a parameter is the posterior mean of the posterior distribution. Due to multiple integrals, it cannot solve without any computational method. Therefore, a computational approach such as the Markov chain Monte Carlo (MCMC) technique is applied for obtaining the value from the estimators. For that, we derive the conditional posterior distribution/ probability for the model parameters

$$\pi(\mathbf{t}|\mathbf{y}, r) = \frac{\pi(\mathbf{t}, \mathbf{y}|r)}{\pi(\mathbf{y}|r)}, \tag{4}$$

$$\pi(\rho|y, \psi, \gamma, \tau) = |V(\rho)|^{\frac{1}{2}} \exp\left[-\frac{\tau}{2}\left\{\left(y - \rho y_{-1} - Z(\rho)\gamma - S(\rho)\psi\right)'\left(y - \rho y_{-1} - Z(\rho)\gamma - S(\rho)\psi\right)\right.\right.$$
$$\left.\left. + \left(\gamma - (1-\rho)\phi_0\right)'V(\rho)\left(\gamma - (1-\rho)\phi_0\right)\right\}\right], \tag{5}$$

$$\gamma|y, \rho, \psi, \tau \sim \text{MN}\left(\frac{\left(Z'(\rho)\left(y - \rho y_{-1} - S(\rho)\psi\right) + (1-\rho)V(\rho)\phi_0\right)}{\left(Z'(\rho)Z(\rho) + V(\rho)\right)}, \frac{\left(Z'(\rho)Z(\rho) + V(\rho)\right)^{-1}}{\tau}\right), \tag{6}$$

$$\psi|y, \rho, \gamma, \tau \sim \text{MN}\left(\frac{\left(S'(\rho)\left(y - \rho y_{-1} - Z(\rho)\gamma\right) + \Omega\psi_0\right)}{S'(\rho)S(\rho) + \Omega}, \frac{\left(S'(\rho)S(\rho) + \Omega\right)^{-1}}{\tau}\right), \tag{7}$$

$$\tau|y, \rho, \gamma, \psi \sim \text{Gamma}\left(\frac{T+r}{2} + 1, \frac{\text{K}}{2}\right), \tag{8}$$

where

$$\pi(t, y|r) \propto \frac{1}{T^{-1}C_r} \int_a^1 \frac{(1+\rho)^{\frac{1}{2}}}{(1-\rho)^{\frac{3}{2}} |A(\rho)|^{\frac{1}{2}} |D(\rho)|^{\frac{1}{2}} [\xi(\rho)]^{\frac{T}{2}}} d\rho,$$

$$\pi(y|r) = \sum_{t_1} \cdots \sum_{t_r} \pi(t, y|r),$$

$$A(\rho) = S'(\rho)S(\rho) + \Omega,$$

$$B(\rho) = I - S(\rho)A^{-1}(\rho)S'(\rho),$$

$$C(\rho) = \left(Z'(\rho)B(\rho)(y - \rho y_{-1}) + (1+\rho)\phi_o - Z(\rho)S(\rho)A^{-1}(\rho)\Omega\psi_0\right),$$

$$D(\rho) = Z'(\rho)B(\rho)Z(\rho) + V,$$

$$\xi(\rho) = (y - \rho y_{-1})'B(\rho)(y - \rho y_{-1}) + (1-\rho)^2\phi_0'V(\rho)\phi_0 - 2\left((y - \rho y_{-1})'S(\rho)A^{-1}(\rho)\Omega\Psi_o\right)$$
$$+ \psi_o'\Omega\psi_o - \psi_0'\Omega'A^{-1}(\rho)\Omega\psi_0 - C'(\rho)(D(\rho))^{-1}C(\rho),$$

$$\text{K} = \left(y - \rho y_{-1} - Z(\rho)\gamma - S(\rho)\psi\right)'\left(y - \rho y_{-1} - Z(\rho)\gamma - S(\rho)\psi\right) + (\psi - \psi_0)'\Omega(\psi - \psi_0)$$
$$+ \left(\gamma - (1-\rho)\phi_0\right)'V(\rho)\left(\gamma - (1-\rho)\phi_0\right).$$

The location of knots and autoregressive coefficient are not closed distribution forms. Hence, the M–H algorithm is applied to draw samples from the conditional posterior distribution. In contrast, the remaining parameters generate posterior

samples from the Gibbs sampler algorithm, because conditional posterior distribution is obtained in a close distribution. The step-by-step procedure for implementing this proposed method is given as:

Step 1. At the first iteration, start with an initial value of the parameters $\rho^{(0)}$, $\gamma^{(0)}$, $\psi^{(0)}$, and $\tau^{(0)}$.

Step 2. Generate posterior samples from the conditional posterior distribution of the model parameters given in Eqs. (5)–(8) and putting the estimated value of parameters in the conditional posterior distribution of the location of a knot.

Step 3. For the ith location of a knot, it is detected by considering every time point in the interval $(T_{i-1}+1, T-r+i-1)$ as a knot point and recording the probabilities that occur parallel to these time points.

Step 4. We find the ith location of a knot as a maximum of all probabilities corresponding to a single time point, i.e., $\hat{t}_i = \max \pi(\hat{t}_i | y, r)$.

Step 5. We get a vector of knot points at the first iteration $\hat{\mathbf{t}}^{(1)} = (\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_r)$.

Step 6. Repeat the process up to k iterations, make a sequence of parameters and location of knots. The average comes out to be the estimated value of the parameters and location of knots.

The number of knots is determined using the Bayes factor. The Bayes factor ($\text{BF}_{n,m}$) is the ratio of one versus another model/hypothesis, i.e., it defines by the posterior probability of n knots divided by m knots. For this model, $\text{BF}_{n,m}$ is expressed as

$$\text{BF}_{n,m} = \frac{\pi(y|r=n)}{\pi(y|r=m)} = \frac{\sum_{t_1} \cdots \sum_{t_n} \pi(t_i, y|n)}{\sum_{t_1} \cdots \sum_{t_m} \pi(t_i, y|m)}. \tag{9}$$

The procedure starts with the series has no knot and evaluates the evidence to support one or more knots. If there is significant evidence for supporting the existence of knots, then check whether there is one knot, two knots, or so on. Therefore, we aim to find a piece of strong evidence between the models/hypotheses before making a better decision about the number of knots. Kass and Raftery (1995) provided a rule of thumb for interpreting the magnitude of Bayes factor using the transformation $2\log_e(\text{BF}_{n,m})$ as defined in Table 1 and put on the same scale as the likelihood ratio.

Another approach is to find out the number of knots using an information criterion discussed by Kumar et al. (2020).

4 Modeling of COVID-19 series

We collect COVID-19 data from the World Health Organization's official daily reports (https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports). This report covers the total number of infected people due to this virus daily for every country. We model the series of some countries (USA, India, Brazil, Russia, South Africa, and Peru) that are most affected and determine the growth structure and current stage of transmission by fitting the proposed model.

The study starts from 500 outbreaks of coronavirus cases and up to the date on 1st September 2020 for the selected countries. We notice that confirmed cases of some

Table 1 Selection criterion based on Bayes factor

$2\log_e(BF_{i,j})$	$BF_{i,j}$	Evidence against the null hypothesis (H_0: $j = m$)
<2	1–3	Not worth more than a bare mention
≥ 2 and <6	3–20	Positive
≥ 6 and <10	20–150	Strong
≥ 10	>150	Very strong

countries like the USA, India, Brazil, and Peru are rising and increasing rapidly. In contrast, the coronavirus spread is slowed down in the remaining countries (Russia and South Africa). Based on the proposed methodology, the number of knots is determined using the Bayes factor to know the transmission stage of a country, and the results are recorded in Table 2.

From Table 2, we observe that the proposed model with $r = 3$ knots is superior to the model with a lesser number of knots for the series of USA, Brazil, and India, because the corresponding value of $2\log_e(BF_{n,m})$ is between 2 and 6 (positive evidence). It indicates a piece of robust evidence in favor of $r = 3$ against $r = 0$, 1, and 2 for the observed series. It concludes that these countries achieve the third stage of coronavirus, i.e., community transmission. The following countries Russia, South Africa, and Peru obtain the model with $r = 2$ knots, because strong and positive evidence in favor of two knots is recorded compared to $r = 0$, 1, and 3 knots. For these countries, a maximum of two knots is presented to fit the proposed model as these countries control the COVID-19 confirmed cases in the second stage of transmission. Hence, we observe that the first knot happens in the early days of transmission of the coronavirus, because most of the cases are reported based on travel history from the affected countries, whereas second and third knots are recorded based on local and community transmission within the country, respectively. Hence, the number of knots shows the stage of transmission of the COVID-19 for a particular country.

Once the suitable number of knots is determined, locations of the knots are found based on the conditional posterior probability given in Eq. (4). The values of the conditional posterior probability are fitted to the observed series of the confirmed cases and display in Fig. 1. The occurrence of the locations of knots is showed by vertical dash lines at which maximum probability is recorded at a particular interval.

Based on Fig. 1, the first joint point is selected by considering every time point in the interval $(2, T - r)$ as a knot location and record the probabilities parallel to these time points. The study finds the first-knot point $(\hat{t}_1)$ at the maximum of all probabilities corresponding to a single time point. Next, the second-knot location is also determined based on higher probability in the given interval, i.e., it records the maximum probability among the bunch of all probabilities corresponding to the time interval $(\hat{t}_1 + 1, \quad T - r + 1)$, denoted as $(\hat{t}_2)$. Similarly, the $(i+1)$th knot location is

Table 2 Determine the number of knots in the COVID-19 series based on Bayes factor

Country	Numerator ($r=n$) against denominator model ($r=m$)	$2\log_e(B_{n,m})$	Evidence against model with $r=m$
USA	$r=1$ against $r=0$	0.9998	Not worth more than a bare mention
	$r=2$ against $r=0$	1.1162	Not worth more than a bare mention
	$r=2$ against $r=1$	0.1164	Not worth more than a bare mention
	$r=2$ against $r=3$	1.0334	Not worth more than a bare mention
	$r=3$ against $r=0$	3.4807	Positive
	$r=3$ against $r=1$	2.4809	Positive
Brazil	$r=1$ against $r=0$	0.5149	Not worth more than a bare mention
	$r=2$ against $r=0$	1.0969	Not worth more than a bare mention
	$r=2$ against $r=1$	1.6118	Not worth more than a bare mention
	$r=2$ against $r=3$	0.1650	Not worth more than a bare mention
	$r=3$ against $r=0$	2.9319	Positive
	$r=3$ against $r=1$	3.4468	Positive
India	$r=1$ against $r=0$	0.3285	Not worth more than a bare mention
	$r=2$ against $r=0$	0.5217	Not worth more than a bare mention
	$r=2$ against $r=1$	0.8502	Not worth more than a bare mention
	$r=2$ against $r=3$	0.5833	Not worth more than a bare mention
	$r=3$ against $r=0$	5.1050	Positive
	$r=3$ against $r=1$	5.4335	Positive
Russia	$r=1$ against $r=0$	0.6469	Not worth more than a bare mention
	$r=2$ against $r=0$	4.2159	Positive
	$r=2$ against $r=1$	4.8628	Positive
	$r=2$ against $r=3$	3.5609	Positive
	$r=3$ against $r=0$	0.6550	Not worth more than a bare mention
	$r=3$ against $r=1$	1.3020	Not worth more than a bare mention
South Africa	$r=1$ against $r=0$	1.5424	Not worth more than
	$r=2$ against $r=0$	4.1505	Positive
	$r=2$ against $r=1$	2.6081	Positive
	$r=2$ against $r=3$	2.3199	Positive
	$r=3$ against $r=0$	1.8306	Not worth more than a bare mention
	$r=3$ against $r=1$	0.2882	Not worth more than a bare mention
Peru	$r=1$ against $r=0$	0.8057	Not worth more than a bare mention
	$r=2$ against $r=0$	3.3421	Positive
	$r=2$ against $r=1$	2.5364	Positive
	$r=2$ against $r=3$	2.0966	Positive
	$r=3$ against $r=0$	1.2454	Not worth more than a bare mention
	$r=3$ against $r=1$	0.4397	Not worth more than a bare mention

obtained based on the range $\left(\hat{t}_i + 1, \quad T - r + i\right)$ and get the time point equivalent to the maximum probability.

The locations of knots are displayed in Table 3 based on Fig. 1. Table 3 shows that most countries have occurred the location of first-knot during the lockdown
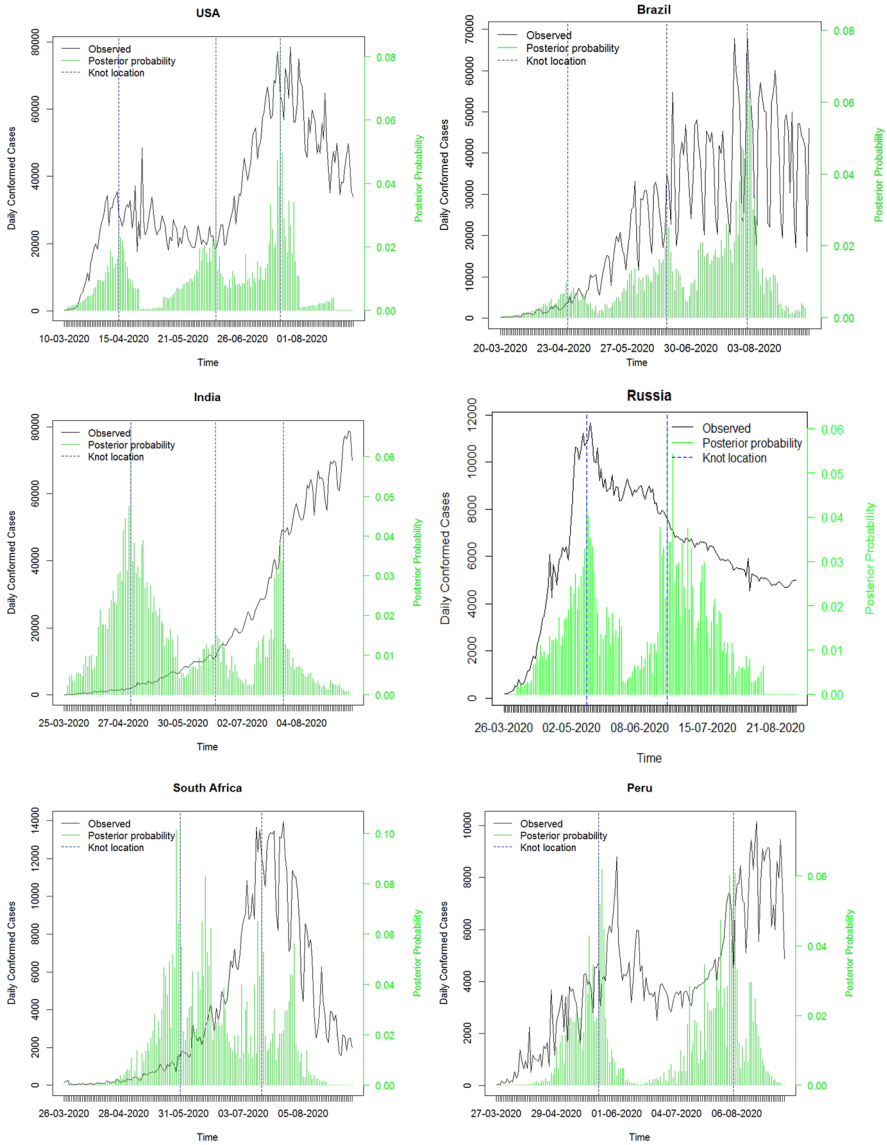
Fig. 1 Selection of knot location(s) based on the posterior probability

period, suspension of traveling, economic activities restrictions, etc. (Mbunge, 2020; Pai et al., 2020; Soni, 2021; Tang et al., 2020). During this period, most cases are reported based on the travel history of the infected person who is traveled from affected nations and imports the virus into the country.

The location of the second knot happens in the mid of June when the number of daily COVID-19 cases increases rapidly for India, Brazil, and USA (da Candido

Table 3 Locations of knots of the COVID-19 series for the selected countries

Country	No. of knots (r)	Locations of knots		
		$\hat{t}_1$	$\hat{t}_2$	$\hat{t}_3$
USA	3	12-April	10-June	19-July
Brazil	3	25-April	17-June	30-July
India	3	01-May	17-June	25-July
Russia	2	10-May	23-June	–
South Africa	2	29-May	13-July	–
Peru	2	22-May	04-August	–

et al., (2020); Saha & Chouhan, 2021; Zhang et al., 2020). At this period, some relaxations are given by the government to reduce the daily time of lockdown, permit economic activities, and reopen the public offices under some guidelines. Therefore, the COVID-19 virus may have affected nearby people who have direct contact with the infected person. It is still manageable to locate and provides prompt medical care to the infected person and control the spread of the virus. This situation is under control in the country of Russia, South Africa, and Peru (Garba et al., 2020; Stiegler & Bouchard 2020; Zemtsov & Baburin, 2020).

The location of the third knot is obtained in the period when the number of confirmed cases is rapidly increasing (more than 50,000 per day) for the series of India, Brazil, and USA (Lin et al., 2020; Ray & Subramanian, 2020). In this period, there is a community transmission in various cluster forms at multiple locations and not easily traceable. Based on the best-fitted model with appropriate locations, Bayesian estimated values of the model parameters for each country series are summarized in Table 4.

Table 4 concludes that variability is more in all country series, because all record a higher number of COVID-19 confirmed cases. The series of all countries are stationary based on the estimated value of the autoregressive coefficient (ρ), because it is under the stationarity condition. The positive (negative) value of the intercept term (δ_0) indicates the increment (decrement) of the daily confirmed cases when other variables are not there. The positive (negative) value of the trend coefficient (δ) tells the total increase (decrease) of confirmed cases expected to a unit change in the time (t). The daily confirmed cases shows an

Table 4 Estimated value of the best-selected model for each country COVID-19 series

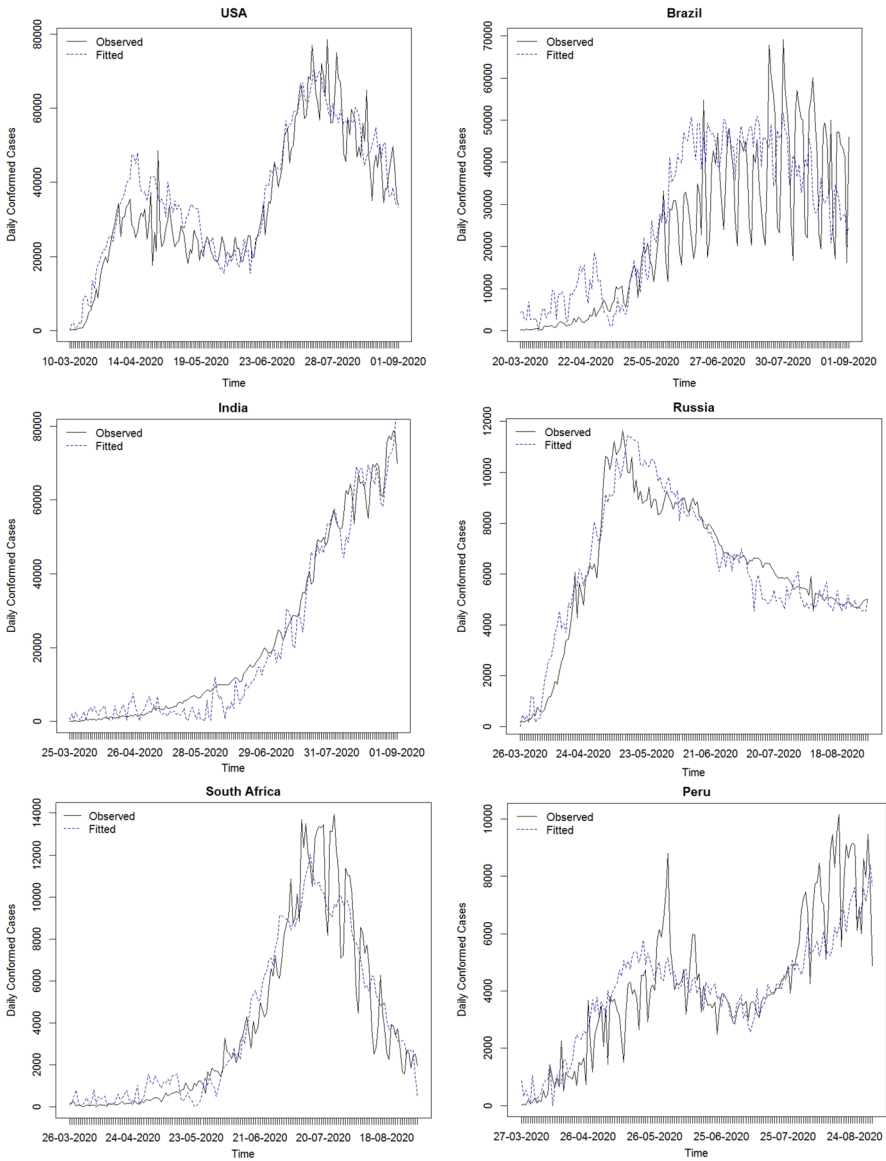| Country	ρ	δ_0	δ	ψ_1	ψ_2	ψ_3	τ^{-1}
USA	0.77	− 6673.93	1468.39	− 1862.71	1750.11	− 2108.48	1.62E−05
Brazil	0.33	− 421.14	− 321.52	1590.11	− 285.88	− 1270.18	1.86E−06
India	0.77	1401.47	− 121.48	315.07	862.39	− 126.32	1.07E−06
Russia	0.84	1975.01	− 182.19	469.47	850.57	–	1.03E−06
South Africa	0.91	99.77	− 6.67	298.03	− 496.26	–	4.10E−06
Peru	0.66	− 1311.29	181.21	− 268.11	236.05	–	5.25E−06

Fig. 2 Observed and fitted COVID-19 series for every country

increasing (decreasing) time trend pattern when a positive (negative) unit change is happened for the estimated value of the spline coefficient (ψ_i). Based on the estimated values of the parameter, the observed and fitted series is plotted in Fig. 2. From Fig. 2, we observe that estimated values of the model parameters give a better fit over the observed series.

5 Conclusion

Nowadays, COVID-19 pandemic is a severe challenge for the human to survive on earth. The COVID-19 has a wide range of consequences on human life worldwide, because million of people die due to coronavirus. Therefore, there is a need to study the growth of COVID-19 cases based on various predictive models. It is also available in the literature that transmission of the COVID-19 virus has four stages. Based on daily recorded cases, we observe that the structure of the COVID-19 series in various countries is not linear, because many reasons such as lockdown, infection modes, and poor health infrastructure are present to control or expand this disease. Thus, our paper deals with a non-linear time series model using the spline function that switches the non-linear trend component into the linear trend. It is analyzed based on different segments and fits the linear trend autoregressive model at each segment. Each segment shows the stage of COVID-19 transmission. Parameter estimators and the number of knots are determined under the Bayesian approach. The results conclude that the number of knots and their locations are useful to assess the transmission stage and the location of time for attaining this. Hence, the proposed methodology quickly analyzes the non-linear trend of the COVID-19 series using the spline function.

Availability of data and materials https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports and https://ourworldindata.org/coronavirus-source-data.

Code availability R-code is provided for the reviewers as per need.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Al-Rousan, N., & Al-Najjar, H. (2020). Data analysis of coronavirus COVID-19 epidemic in South Korea based on recovered and death cases. *Journal of Medical Virology, 92*(9), 1603–1608.

Biller, C. (2000). Adaptive Bayesian regression splines in semiparametric generalized linear models. *Journal of Computational and Graphical Statistics, 9*(1), 122–140.

Conrad, D. J., Bailey, B. A., Hardie, J. A., Bakke, P. S., Eagan, T. M., & Aarli, B. B. (2017). Median regression spline modeling of longitudinal FEV1 measurements in cystic fibrosis (CF) and chronic obstructive pulmonary disease (COPD) patients. *PLoS ONE, 12*(12), 1–13.

da Candido, D. S., Claro, I. M., de Jesus, J. G., de Souza, W. M., Moreira, F. R. R., Dellicour, S., & Faria, N. R. (2020). Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science, 369*, 1255–1260.

Denison, D., Mallick, B., & Smith, A. (1998). Automatic Bayesian curve fitting. *Journal of the Royal Statistical Society: Series B (statistical Methodology), 60*(2), 333–350.

Eubank, R. L. (1999). *Nonparametric regression and spline smoothing*. CRC Press.

Garba, S. M., Lubuma, J. M. S., & Tsanou, B. (2020). Modeling the transmission dynamics of the COVID-19 pandemic in South Africa. *Mathematical Biosciences, 328*, 108441.

Gondauri, D., Mikautadze, E., & Batiashvili, M. (2020). Research on COVID-19 virus spreading statistics based on the examples of the cases from different countries. *Electron Journal of General Medicine, 17*(4), em209.

Hurley, D., Hussey, J., McKeown, R., & Addy, C. (2006). An evaluation of splines in linear regression. In: SAS conference proceedings: SAS Users Group International 31(SUGI 31 Proceedings), Paper 147.

Jiang, X., Zhao, B., & Cao, J. (2020). Statistical analysis on COVID-19. *Biomedical Journal of Scientific and Technical Research, 26*(2), 19716–19727.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*, 773–795.

Kumar, J., Agiwal, V., Kumar, D., & Chaturvedi, A. (2020). Bayesian unit root test for AR(1) model with trend approximated by linear spline function. *Statistics, Optimization and Information Computing, 8*(2), 425–461.

Lin, Y. C., Chi, W. J., Lin, Y. T., & Lai, C. Y. (2020). The spatiotemporal estimation of the risk and the international transmission of COVID-19: A global perspective. *Scientific Reports, 10*(1), 1–12.

Liu, J. M. (2009). Non-linear time series modeling using spline-based nonparametric models. In: Proceedings of the 15th American conference on applied mathematics, pp. 183–189.

Lusa, L., & Ahlin, Č. (2020). Restricted cubic splines for modelling periodic data. *PLoS ONE, 15*(10), e0241364.

Mbunge, E. (2020). Effects of COVID-19 in South African health system and society: An explanatory study. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews, 14*(6), 1809–1814.

Montoril, M. H., Morettin, P. A., & Chiann, C. (2014). Spline estimation of functional coefficient regression models for time series with correlated errors. *Statistics and Probability Letters, 92*, 226–231.

Morton, R., Kang, E. L., & Henderson, B. L. (2009). Smoothing splines for trend estimation and prediction in time series. *Environmetrics, 20*(3), 249–259.

Osmani, F., Hajizadeh, E., & Mansouri, P. (2019). Kernel and regression spline smoothing techniques to estimate coefficient in rates model and its application in psoriasis. *Medical Journal of the Islamic Republic of Iran, 33*, 90.

Pai, C., Bhaskar, A., & Rawoot, V. (2020). Investigating the dynamics of COVID-19 pandemic in India under lockdown. *Chaos, Solitons and Fractals, 138*, 109988.

Qiao, B., Chen, X., Xue, X., Luo, X., & Liu, R. (2015). The application of cubic B-spline collocation method in impact force identification. *Mechanical Systems and Signal Processing, 64*, 413–427.

Ray, D., & Subramanian, S. (2020). India's lockdown: An interim report. *Indian Economic Review, 55*(1), 31–79.

Saha, J., & Chouhan, P. (2021). Lockdown and unlock for COVID-19 and its impact on residential mobility in India: An analysis of the COVID-19 Community Mobility Reports, 2020. *International Journal of Infectious Diseases, 104*, 382–389.

Soni, P (2021). Effects of COVID-19 lockdown phases in India: An atmospheric perspective. Environment, Development and Sustainability, pp. 1–12.

Stiegler, N., & Bouchard, J. P. (2020). South Africa: Challenges and successes of the COVID-19 lockdown. *Annales Médico-Psychologiques, Revue Psychiatrique, 178*(7), 695–698.

Tang, Y., Serdan, T. D., Masi, L. N., Tang, S., Gorjao, R., & Hirabara, S. M. (2020). Epidemiology of COVID-19 in Brazil: Using a mathematical model to estimate the outbreak peak and temporal evolution. *Emerging Microbes and Infections, 9*(1), 1453–1456.

Ülker, E., & Arslan, A. (2009). Automatic knot adjustment using an artificial immune system for B-spline curve approximation. *Information Sciences, 179*(10), 1483–1494.

Zemtsov, S. P., & Baburin, V. L. (2020). COVID-19: Spatial dynamics and diffusion factors across Russian regions. *Regional Research of Russia, 10*(3), 273–290.

Zhang, R., Li, Y., Zhang, A. L., Wang, Y., & Molina, M. J. (2020). Identifying airborne transmission as the dominant route for the spread of COVID-19. *Proceedings of the National Academy of Sciences, 117*(26), 14857–14863.

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