Nonlinear models: a case of the COVID-19 confirmed rates in top 8 worst affected countries

Serdar Neslihanoglu

Received: 30 December 2020 / Accepted: 22 May 2021 / Published online: 7 June 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract Over the last 9 months, the most prominent global health threat has been COVID-19. It first appeared in Wuhan, China, and then rapidly spread throughout the world. Since no treatment or preventative strategy has been identified until this time, millions of people across the world have been seriously affected by COVID-19. The modelling and prediction of confirmed COVID-19 cases have been given much attention by government policymakers for the purpose of combating it more effectively. For this purpose, the modelling and prediction performances of the linear model (LM), generalized additive model (GAM) and the time-varying linear model (Tv-LM) via Kalman filter are compared. This has never yet been undertaken in the literature. This comparative analysis also evaluates the linear relationship between the confirmed cases of COVID-19 in individual countries with the world. The analysis is implemented using daily COVID-19 confirmed rates of the top 8 most heavily affected countries and that of the world between 11 March and 21 December 2020 and 14-day forward predictions. The empirical findings show that the Tv-LM outperforms others in terms of model fit and predictability, suggesting that the relationship between each country's rates with the world's should be locally linear, not globally linear.

Keywords Confirmed rate · COVID-19 · Generalized additive model · Kalman filter · Time-varying linearity

1 Introduction

The Coronavirus (COVID-19) pandemic of 2019 is caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), which was formerly known as the 2019 novel Coronavirus (2019-nCoV) [17]. COVID-19, which was firstly detected in Wuhan, China, on 31 December 2019, was reported as being a global pandemic by the World Health Organization (WHO) on 11 March 2020 [21]. COVID-19, which is transmitted by means of respiratory droplets and contact routes, can lead to death in people of all ages, particularly those with chronic illness or the elderly. As of 20 December 2020, the WHO reported that there were, globally, over 75 million confirmed cases and over 1.6 million deaths [22]. No valid, well-tested treatments and preventative strategy against COVID-19 have been discovered by the date of the submission of this article. COVID-19 has spread rapidly worldwide in a very short amount of time. It has become a major threat to the economic, environmental and social development of all countries. To control and prevent the rapid spread of COVID-19 until the production of effective diagnostic kits and well-tested medications, governments have encouraged preventative strategies which have been primarily based on those first adopted by the Chinese government, such as: quarantine; lockdown

S. Neslihanoglu (✉)
Department of Statistics, Faculty of Science and Letters,
Eskisehir Osmangazi University, Meselik Yerleskesi,
26480 Eskisehir, Turkey
e-mail: sneslihanoglu@ogu.edu.tr

of internal and external borders; social distancing; and the isolation of infected populations. After around 13–14 days of implementing these kinds of strategies, a decreased trend of confirmed COVID-19 cases was discerned [15].

COVID-19 spread rapidly from China across the rest of the world. It can be useful for governments to utilize epidemiological, mathematical and statistical models in order to design better strategies by which to control the spread of COVID-19 and to more precisely manage some public sectors, including that of health and economy (e.g. extending the personnel, bed and equipment capacity of hospitals and supporting companies economically for not bailing out during the COVID-19 pandemic). Due to COVID-19's rapid growth in such a short amount of time, numerous studies in international scientific and medical journals have focused on an epidemiological analysis of COVID-19 (such as clinical, laboratory and imaging features; reproductive, transmission and fatality rates; and epidemic trends) in order to better cope with its temporary and long-term impacts on the economy, the environment and social development in general [1–4, 11, 15, 17, 18, 23].

To provide guidance for COVID-19 policymakers, there have been numerous recent studies in the literature on the model fitting and predicting of reported COVID-19 cases, such as confirmed, recovered and death, especially with relation to the linear model (LM) and its specifications. For example, Gupta et al. [6] have analysed the outbreak of COVID-19 in India for the period between 30 January 2020 and 30 March 2020. It predicted the number of COVID-19 cases for the next 2 weeks using the SEIR (susceptible–exposed–infectious–recovered) and multiple linear models. In addition, Ghosal et al. [5] analysed the top 15 most heavily affected countries predicting the number of deaths for the 5th and 6th weeks of the COVID-19 epidemic in India using both multiple and linear models based on the total number of infected, active, confirmed and death/fatality cases rates (CFR). The nonlinear extension of the LM via the generalized additive model (GAM) was investigated by Zareie et al. [24], who modelled the time-dependent transmission, recovery and death rates of COVID-19 in Iran based on the Chinese data between 22 January to 24 March 2020 in order to effectively predict the number of patients for the next month in Iran. The time-varying parameter extension of the LM (Tv-LM) in the form of the state space model via Bayesian methodology was investigated by

Kobayashi et al. [9], who focused on the real-time data of Japan from 1 March to 22 April 2020 by adopting the susceptible–infected–removed (SIR) model. To sum up, most of these analyses mainly focused on the model fitting and prediction of COVID-19 cases by using sophisticated epidemiological models, such as SEIR and SIR, and statistical modelling tools, such as the Bayesian methodology. None of these have yet provided benchmark models for assessing the relationship between the confirmed cases of COVID-19 in individual countries and that of the world in terms of comparing the LM with its GAM and Tv-LM via Kalman filter [8] extensions in the in- and out-of-sample procedures. The Kalman filter, which is also thought to be one of the most preferred algorithms for modelling time-varying linearity due to the accuracy of the results which it yields [13, 14], has yet to be undertaken for measuring time-varying linearity in the COVID-19 literature.

This research mainly focuses on closing the aforementioned gaps by conducting a comparative analysis which explains the time-varying linear relationship between the confirmed rate of COVID-19 for individual countries and the world. For this purpose, the performance comparison of the LM via ordinary least squares (OLS) with its two extensions, including the generalized additive model (GAM) and the time-varying parameter of the LM (Tv-LM) via the Kalman filter, is performed while in-model fitting and predicting the confirmed rates of COVID-19, which has yet to be undertaken in the literature. The daily COVID-19 confirmed case data of the world and the top 8 most heavily affected countries, including Brazil, France, Germany, Italy, Russia, Spain, the UK and the USA, were obtained based on a continuous listing from 11 March to 21 December 2020. Forecasts were also based on 14-day forward predictions of the confirmed COVID-19 rates in those countries with a 1- and 14-day ahead rolling window forecast while using the models which are being compared. The aforementioned models' performance is compared using the mean square error (MSE), the mean absolute error (MAE) and graphical summaries.

The rest of this research paper is organized as follows. Section 2 summarizes the characteristics of the research data. Section 3 details the statistical methodology of the models being compared. Section 4 specifies the empirical outcomes and graphical summaries from the aforementioned models' comparison, while also presenting further results from the best model fit

and forecasting model. Finally, Sect. 5 details the conclusion of this research.

2 Data description

The daily cumulative number of reported confirmed COVID-19 cases dated from 11 March to 21 December 2020 across 8 countries, including Brazil, France, Germany, Italy, Russia, Spain, the UK and the USA and the world, was taken from Krispin [10] based on the Johns Hopkins University Center for Systems Science and Engineering (JHU CCSE) and 27 other COVID-19 databases. The main criteria of selection which were used were based on those being the most heavily affected in terms of total confirmed cases as of the end of May 2020 (the date being at the end of first partial national lockdown and quarantine for the COVID-19 pandemic in many countries), and having continuous listings of epidemic data for all countries and the world from 11 March (the date COVID-19 was reported as being a global pandemic by the World Health Organization (WHO) [21]) to 21 December 2020. The daily COVID-19 confirmed rate, R_{it} , is defined as follows:

$$R_{it} = \frac{\text{Accumulated confirmed COVID-19 cases}_{it}}{\text{Accumulated confirmed COVID-19 cases}_{i,t-1}}, \quad (1)$$

for $t = 2, \dots, T$ and $i = 1, \dots, 8$ for each country and $i = w$ for the world. The descriptive statistics and time series plots of COVID-19 confirmed rate for each country and the world are shown in Table 1 and Fig. 1, respectively.

Table 1 depicts some significant information. The mean of the COVID-19 confirmed rate of 8 countries and the world is positive and over 1 during that period, suggesting that the COVID-19 pandemic was still progressing during that period. Within the 8 selected countries, the largest mean of the rate of confirmed COVID-19 cases was 1.049 in Brazil, while the lowest mean of the confirmed COVID-19 rate was 1.018 in Italy. Also, the mean of the world's confirmed COVID-19 rate was 1.023, which is smaller than that of the 8 selected countries, with the exception of Italy. This means that the daily change in reported confirmed COVID-19 cases of the 8 selected countries, excluding Italy, was much higher than that in the world during that period. The maximum confirmed COVID-19 rate of Brazil was 2.904, which was higher than that of the other countries and the world, suggesting that there was a higher chance of there being a reported confirmed COVID-19 case than 1-day before in Brazil during that period than for any of the other countries or the world. The minimum COVID-19 confirmed rate is not equal to 1 for Russia (1.005), the USA (1.004) and the world (1.007), meaning that these countries had reported COVID-19 patients on each day of that period.

As can be seen in Fig. 1, the time series plots of COVID-19 confirmed rates for the 8 selected countries and the world show that they are not stationary and depict a decreasing trend after around 15 March 2020, while after around 15 August 2020, the COVID-19 confirmed rate of European Union countries (France, Germany, Italy and Spain) displays an increasing trend compared to the others when compared to the previous COVID-19 pandemic period. That may have been due

Table 1 Descriptive statistics of COVID-19 confirmed rate of 8 countries and the world

Country	Minimum	1st quartile	Median	Mean	3rd quartile	Maximum
World	1.007	1.010	1.015	1.023	1.020	1.131
Brazil	1.000	1.006	1.015	1.049	1.053	2.904
France	1.000	1.003	1.008	1.027	1.022	1.645
Germany	1.000	1.003	1.007	1.025	1.022	1.769
Italy	1.000	1.002	1.006	1.018	1.021	1.213
Russia	1.005	1.008	1.011	1.045	1.029	1.607
Spain	1.000	1.001	1.007	1.026	1.019	2.298
UK	1.000	1.004	1.011	1.027	1.024	1.376
USA	1.004	1.009	1.014	1.037	1.019	1.490

Fig. 1 The time series plots of 8 countries and the world COVID-19 confirmed rates

to the quarantine, lockdown of internal and external borders and social distancing strategies which governments obliged, as discussed by [15]. Overall, the differences in behaviour of the COVID-19 confirmed rate in all countries and the world have been observed due to their having different COVID-19 epidemic exposure dates, as well as their having different social, cultural and technological developments.

Due to the appeal of the non-stationarity of the COVID-19 confirmed rate, the logarithmic transformation of the difference in the daily COVID-19 confirmed rate (with some suitable transformations having stationary series [19]) is applied to the data; then the stationary COVID-19 confirmed rate series of the 8 selected countries and the world are obtained and used throughout this research's analysis. The time series plots of the stationary COVID-19 confirmed rate which have been omitted for the sake of brevity are available upon request.

3 Methodology

The statistical methodology of the compared models, including the linear model (LM), generalized additive model (GAM) and time-varying linear model (Tv-LM), is discussed here.

Linear model (LM)

The linear model (LM) is defined as follows:

$$R_{it} = \alpha_i + \beta_i R_{wt} + \varepsilon_{it}. \quad (2)$$

Here, R_{it} and R_{wt} are the COVID-19 confirmed rates of the country i 's ($i = 1, \dots, 8$) and the world at time t ($t = 1, \dots, T$), respectively. ε_{it} are the residuals, $\varepsilon_{it} \sim N(0, H_i)$ with $E(\varepsilon_{it}\varepsilon_{kt})=0$, for $i \neq k$ and $E(\varepsilon_{it}\varepsilon_{i,t+j})=0$, for $j > 0$. α_i and β_i are the regression intercept and slopes, respectively. The unknown parameters of LM are estimated by using the ordinary least squares (OLS) outlined by Shumway and Stoffer [19].

Generalized additive model (GAM)

The generalized additive model (GAM), developed by Hastie and Tibshirani [7], is a nonlinear extension of

the linear model (Eq. 2) which retains the smooth functions of covariates. These generate shapes which are estimated based on data rather than being specifically inputted by the researcher, thereby allowing one to assess the nonlinear relationship with the response. The GAM function is defined as follows:

$$R_{it} = \alpha_i + f_i(R_{wt}) + \varepsilon_{it}. \quad (3)$$

Here, $\varepsilon_{it} \sim N(0, H_i)$ with $E(\varepsilon_{it}\varepsilon_{kt})=0$, for $i \neq k$, and $E(\varepsilon_{it}\varepsilon_{i,t+j})=0$, for $j > 0$, and $f_i(R_{wt})$ is a smooth function of R_{wt} . The parameter estimation procedure of the GAM is outlined by Wood [20]. Here, the *mgcv* package in R software [16] is used in the GAM analysis.

Time-varying linear model (Tv-LM)

The time-varying parameter extension of the linear model (Eq. 2) is called the time-varying linear model (Tv-LM) in the mean reverting specification form of the state space model via Kalman filter [14] algorithm. This allows one to assess the instability and to estimate the time-varying parameters of the Tv-LM. The state space form of the Tv-LM is designed to become an observation equation by defining it as follows:

$$R_{it} = \alpha_i + \beta_{it} R_{wt} + \varepsilon_{it}. \quad (4)$$

Here, α_i and β_{it} ($i = 1, \dots, 8$) are the regression intercept and the regression slopes, respectively. ε_{it} are the residuals with $\varepsilon_{it} \sim N(0, H_i)$. The state equation of Tv-LM is defined as follows

$$\beta_{it} = \bar{\beta}_i + \phi_i(\beta_{i,t-1} - \bar{\beta}_i) + v_{it}, \quad (5)$$

where $v_{it} \sim N(0, Q_i)$ and priors

$$\beta_{i0} \sim N(\mu_{\beta_i}, \Sigma_{\beta_i}),$$

where the parameters of these distributions are computed from the data within the estimation algorithm.

$\bar{\beta}_i = \frac{1}{T} \sum_{t=1}^T \beta_{it}$. The error terms for the observation (ε_{it}) and state (v_{it}) equations are postulated to be mutually independent of each other and independent in time t ($t = 1, \dots, T$). Furthermore, they are normally distributed with a 0 mean and with variances H_i and Q_i , respectively. Also, ϕ_i quantifies the temporal autocorrelation in β_{it} which is estimated from the COVID-19 confirmed rate data. The unknown parameter of the Tv-LM is estimated by using the Kalman filter. The Kalman filter, which was firstly proposed by Kalman [8], is outlined in Neslihanoglu et al. [14] and

is a potent recursive algorithm used for estimating the unobserved components in the state space model.

The rolling window technique for forecasting is used in the out-of-sample forecasting procedure with the models aforementioned (see Fig. 2 for an illustration of the partitions of this technique).

T relates to the sample size, m denotes the rolling window period size (i.e. the number of consecutive observations per rolling window), and n signifies the size of the prediction period, with the size of m being set in terms of the corresponding sample size in order not to diminish the level of model accuracy. Throughout this forecasting procedure, the first rolling window contains observations for period 1 through m , the second rolling window contains observations for period 2 through $m + 1$, and so on. The partitions, however, can vary; namely, rather than rolling one observation ahead, one can roll fourteen observations for bi-weekly date [12].

The modelling (forecasting) accuracy performance of the aforementioned model is compared by using it with the mean absolute error (MAE) and the mean squared error (MSE). These are defined as follows:

$$\text{MAE} = \frac{1}{T} \sum_{t=1}^T |\hat{R}_{it,\text{model}} - R_{it}|, \quad (6)$$

$$\text{MSE} = \frac{1}{T} \sum_{t=1}^T (\hat{R}_{it,\text{model}} - R_{it})^2. \quad (7)$$

According to these criteria, the models with the lowest MSE and MAE values have a better performance at modelling (forecasting). Note that the statistical analysis of this research was made by using R software [16].

Fig. 2 The figure of the rolling window technique

4 Results

4.1 In-sample model fit

The model fit performance (in-sample) of the linear model (LM), generalized additive model (GAM) and time-varying linear model (Tv-LM) outlined in Sect. 3 for each country's COVID-19 confirmed rate is discussed in this section. The model fit accuracy performance comparison of the compared models [using the MAE (Eq. 6) and MSE (Eq. 7)] is shown in Table 2.

When comparing the models in relation to MAE and MSE, the Tv-LM via Kalman filter (with the lowest MAE and MSE) outperforms both LM and GAM for all countries during the time period. It is worth noting that, according to the results, the Tv-LM performs better than the LM (with the highest MAE and MSE) for the 8 selected countries in relation to MAE and MSE by, on average, 74.9% (92.5%), while it improves the GAM for 8 countries in relation to MAE and MSE by, on average, 67.0% (88.3%). This suggests that the Tv-LM captures the volatility of each country's COVID-19 confirmed rate better than the others. Also, GAM improves LM for 8 countries in terms of MAE (MSE) by, on average, 23.9% (35.7%). To sum up, the Tv-LM via Kalman filter is the preferable model for each country's COVID-19 confirmed rate during the time period.

4.2 Out-of-sample forecasting

The out-of-sample forecasting adopted by the research utilized a rolling window technique to assess the performance of the aforementioned models. In the rolling window process, the length of the rolling window is 270 days (9 months) and the length of the prediction period is 14 days (2 weeks) in order to predict β_{it} by generating a 1- and 14-day ahead forecast for each country's COVID-19 confirmed rate. The MSE and MAE values between the actual and the predicted values of each country's COVID-19 confirmed rate are calculated in Table 3.

A comparison was made between the aforementioned models' 1- and 14-day forecasting performances when quantified over the COVID-19 confirmed rate for the 8 selected countries in relation to MAE (and MSE). It is apparent that the Tv-LM (with the lowest MAE and MSE) outperforms both the LM and GAM during that period. It can clearly be seen that Tv-LM performs bet-

Table 2 In-sample model fit comparison criteria

Criteria Model	MAE $\times 10^2$			MSE $\times 10^4$		
	LM	GAM	Tv-LM	LM	GAM	Tv-LM
Brazil	2.433	1.870	0.480	52.789	35.128	3.272
France	1.896	1.503	0.652	16.583	13.001	2.640
Germany	1.155	0.687	0.252	7.747	3.335	0.220
Italy	0.260	0.227	0.089	0.243	0.181	0.027
Russia	1.001	0.885	0.027	7.263	6.043	0.011
Spain	2.291	1.425	0.848	23.548	11.268	2.360
UK	0.396	0.327	0.147	1.007	0.491	0.082
USA	0.543	0.669	0.012	5.496	4.347	0.001
Average	1.247	0.949	0.313	14.334	9.224	1.077

Bold entries show the best modelling performance for countries with relation to their lowest MAE and MSE, respectively

Table 3 Out-of-sample forecasting comparison criteria

Criteria Model	MAE $\times 10^2$			MSE $\times 10^4$		
	LM	GAM	Tv-LM	LM	GAM	Tv-LM
Period	1-day ahead forecast					
Brazil	0.438	0.546	0.032	0.569	1.423	0.002
France	1.784	1.398	0.195	10.690	7.345	0.097
Germany	0.639	0.444	0.456	0.673	0.659	0.313
Italy	0.144	0.214	0.082	0.027	0.123	0.014
Russia	0.397	0.314	0.013	0.578	0.321	0.001
Spain	0.683	0.661	0.469	0.777	0.723	0.399
UK	0.247	0.405	0.162	0.089	0.404	0.048
USA	0.332	0.189	0.018	0.491	0.159	0.001
Average	0.583	0.521	0.178	1.737	1.395	0.109
Period	14-day ahead forecast					
Brazil	0.301	0.508	0.035	0.303	1.844	0.004
France	1.600	0.721	0.185	9.450	2.056	0.080
Germany	0.676	0.711	0.473	0.717	0.814	0.326
Italy	0.149	0.168	0.095	0.028	0.048	0.017
Russia	0.263	0.244	0.011	0.212	1.386	0.001
Spain	0.579	0.636	0.471	0.547	0.705	0.387
UK	0.177	0.264	0.108	0.052	0.212	0.022
USA	0.212	0.306	0.017	0.146	0.350	0.001
Average	0.495	0.445	0.174	1.432	0.927	0.105

Bold entries show the best forecasting performance for countries with relation to the lowest MAE and MSE, respectively

ter than the LM (with the highest MAE and MSE) for the 8 selected countries in relation to MAE (MSE) by, on average, 69.5% (93.7%) for the 1-day forecast and

by, on average, 64.8% (92.7%) for the 14-day forecast while improving the GAM for the 8 selected countries in terms of MAE (MSE) for the 1-day forecast by, on

average, 65.8% (92.2%) and for the 14-day forecast by, on average, 60.9% (88.7%). This suggests that the Tv-LM captures the COVID-19 confirmed rate outliers of each country better than the others. Also, the GAM improves on LM in terms of MAE (MSE) for the 1-day forecast by, on average, 10.6% (19.7%) and for the 14-day forecast by, on average, 10.1% (35.3%). As a result, the Tv-LM via Kalman filter seems to be the preferable model for modelling both the 1- and 14-day COVID-19 confirmed rate forecasts for each country.

4.3 Graphical summary

The performance of the in-sample model fit of the above-mentioned models is discussed by using scatter plots representing the correlation between each country's COVID-19 confirmed rate and that of the world. The fitted model plots for the LM, GAM and Tv-LM via Kalman filter are represented in Fig. 3.

As can be seen from some of the key points from Fig. 3, the daily COVID-19 confirmed rate for each country (except for Italy) and the world is positively correlated, ranging from 2.15% (UK) to 75.27% (France). The Tv-LM affords a much better fit to the COVID-19 confirmed rate in comparison to the others because it better captures the time-varying association between the COVID-19 confirmed rates of each country and the world. In addition, the greatest strength of this study was the very highly reported adjusted R^2 by, on average, being equal to 0.9593. This means that there is a 95.93% variability in each country's COVID-19 confirmed rate which can be explained by the world COVID-19 confirmed rate for Tv-LM via the Kalman filter's in-sample procedure. This is actually a very high accuracy rate, which could be counted as proof of the Tv-LM via the Kalman filter's power of fit modelling. It is worth noting that there is a short-term volatility in the COVID-19 confirmed rate relationship which is not adequately captured by the GAM. Moreover, the LM might be the most suitable model for depicting the relationship between the confirmed COVID-19 rate of individual countries with the world without the extreme confirmed rate in the data. To sum up, the Tv-LM via Kalman filter appears to be the most appropriate model for the 8 selected countries in terms of the graphical summary of the COVID-19 confirmed rate model-fitted data.

4.4 Best model fit and forecasting

According to the best model fitting and forecasting performance of each country's COVID-19 confirmed rate, the Tv-LM via Kalman filter has been provided with its parameters by the in-sample procedure provided in Table 4. The table of parameters with the standard error for the LM and GAM models has not been included for brevity's sake. Nevertheless, one can request them from the author at any time.

The estimated values of $\hat{Q}_i$ are higher than $\hat{H}_i$ for each country during the time period, meaning that the volatility of the COVID-19 confirmed rate is captured by the state variance being greater than that of the variance displayed by the observations. Moreover, the autocorrelation in terms of time, signified by β_i , is demonstrated by $\hat{\phi}_i$ for each country. Seeing as they are equal or close to 0 suggests that the time-varying β_i changed rapidly due to a low autocorrelation. It is worth noting that the regression intercept $\hat{\alpha}_i$ is negative for France and Spain. This indicates that the actual COVID-19 confirmed rate is lower than the expected COVID-19 confirmed rate during that time period. Moreover, the mean value of the time-varying $\hat{\beta}_i$ is positive and over 1 for all countries except for Italy, Russia and the UK during the same period. This suggests that the COVID-19 confirmed rate of Italy, Russia and the UK is less volatile than that of the world, while the COVID-19 confirmed rate of the rest of the countries is more volatile than that of the world. According to the range of projected by the β_i values observed during the time period, France has a wider range of values, suggesting that the relationship between France and the world is less consistent on the whole than that of other countries due to the volatility of the COVID-19 confirmed rate, as evidenced in Fig. 1.

5 Conclusion

This paper investigated the modelling and forecasting ability of the time-varying linearity in terms of the confirmed COVID-19 rates in order to better guide government policymakers to take the necessary measures against COVID-19 and to intervene early. The analysis was conducted utilizing data from the top 8 most heavily affected countries and the world. The study's empirical findings are in favour of the time-varying linear model (Tv-LM) via Kalman filter, which outper-

Fig. 3 The scatter plots of 8 countries and the world COVID-19 confirmed rates

Table 4 The parameter estimates of Tv-LM via Kalman filter for COVID-19 confirmed rates of 8 countries

Country	$\hat{Q}_i$	$\hat{H}_i$	$\hat{\phi}_i$	$\hat{\alpha}_i$	$\hat{\beta}_i$	Range($\hat{\beta}_{it}$)
Brazil	135.891 (40.696)	0.000 (0.000)	0.201 (0.029)	0.001 (0.000)	3.892 (5.500)	(1.100; 4.591)
France	33.678 (8.906)	0.000 (0.000)	0.520 (0.112)	− 0.001 (0.000)	4.071 (4.298)	(0.781; 6.603)
Germany	9.698 (1.801)	0.000 (0.000)	0.289 (0.047)	0.001 (0.000)	1.496 (0.553)	(− 0.224; 3.232)
Italy	2.063 (0.390)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.400 (0.053)	(− 0.256; 1.007)
Russia	138.424 (15.147)	0.000 (0.000)	0.024 (0.005)	0.000 (0.000)	0.593 (0.461)	(− 0.961; 2.358)
Spain	24.502 (6.537)	0.000 (0.000)	0.154 (0.041)	− 0.001 (0.000)	2.038 (1.439)	(0.741; 3.091)
UK	2.968 (0.695)	0.000 (0.000)	0.029 (0.011)	0.000 (0.000)	0.331 (0.061)	(− 0.201; 0.867)
USA	30.180 (2.794)	0.000 (0.000)	0.222 (0.002)	0.000 (0.000)	1.322 (0.576)	(− 1.107; 3.743)

Range displays the range of β_{it} . *Italic* numbers in parentheses denote the standard errors of Tv-LM parameter estimates via Kalman filter for COVID-19 confirmed rates of 8 countries.

forms linear model specifications, with the structural changes of the confirmed COVID-19 rate being better captured by the time-varying parameter with relation both to predictability and model fit, especially for countries with higher volatility than the world. This comparative analysis illustrates the time-varying linear relationship between the confirmed COVID-19 rate for individual countries and the world. This suggests that the linearity between the confirmed COVID-19 rates of each country with the world should be local instead of global. The proposed benefits of this procedure will be applicable for a large number of epidemiological applications in any infectious disease. It may also promote further cross-disciplinary research.

Availability of data and materials The data that support the findings of this study are openly available at <https://github.com/RamiKrispin/coronavirus/tree/master/data>.

Declarations

Conflicts of interest The author declares no conflict of interest.

References

1. Almeshal, A.M., Almazrouee, A.I., Alenizi, M.R., Alhajeri, S.N.: Forecasting the spread of COVID-19 in Kuwait using compartmental and logistic regression models. *Appl. Sci.* **10**, 3402 (2020)
2. Anastassopoulou, C., Russo, L., Tsakris, A., Siettos, C.: Data-based analysis, modelling and forecasting of the COVID-19 outbreak. *PLoS ONE* **15**(3), 1–21 (2020)
3. Bartoszek, K., Guidotti, E., Iacus, S.M., Okrój, M.: Are official confirmed cases and fatalities counts good enough to study the COVID-19 pandemic dynamics? A critical assessment through the case of Italy. *Nonlinear Dyn.* **101**(3), 1951–1979 (2020)
4. Gao, Y., Zhang, Z., Yao, W., Ying, Q., Long, C., Fu, X.: Forecasting the cumulative number of COVID-19 deaths in China: a Boltzmann function-based modeling study. *Infect. Control Hosp. Epidemiol.* **41**, 1–3 (2020)
5. Ghosal, S., Sengupta, S., Majumder, M., Sinha, B.: Linear regression analysis to predict the number of deaths in India due to SARS-CoV-2 at 6 weeks from day 0 (100 cases—March 14th 2020). *Diabet. Metab. Syndrome Clin. Res. Rev.* **14**(4), 311–315 (2020)
6. Gupta, R., Pandey, G., Chaudhary, P., Pal, K.: SEIR and regression model based COVID-19 outbreak predictions in India. *medRxiv* (2020)
7. Hastie, T., Tibshirani, R.: *Generalized Additive Models. Monographs on Statistics and Applied Probability*. Chapman and Hall, London (1990)

8. Kalman, R.E.: A new approach to linear filtering and prediction problems. *Trans. ASME J. Basic Eng.* **82**(Series D), 35–45 (1960)
9. Kobayashi, G., Sugawara, S., Tamae, H., Ozu, T.: Predicting infection of COVID-19 in Japan: state space modeling approach. 2004.13483 (2020)
10. Krispin, R.: Coronavirus package tasks tracker. <https://github.com/RamiKrispin/coronavirus> (2020). Accessed 22 Dec 2020
11. Maleki, M., Mahmoudi, M.R., Heydari, M.H., Pho, K.-H.: Modeling and forecasting the spread and death rate of coronavirus (Covid-19) in the world using time series models. *Chaos Solitons Fractals* **140**, 1151 (2020)
12. MathWorks: Rolling-window analysis of time-series models. <https://www.mathworks.com/help/econ/rolling-window-estimation-of-state-space-models.html> (2018). Accessed 20 Mar 2021
13. Mergner, S.: Applications of State Space Model in Finance. Universitätsverlag Göttingen (2009)
14. Neslihanoglu, S., Sogiakas, V., McColl, J.H., Lee, D.: Non-linearities in the CAPM: evidence from developed and emerging markets. *J. Forecast.* **36**(8), 867–897 (2017)
15. Pirouz, B., Haghshenas, S., Piro, P., Haghshenas, S.S.: Investigating a serious challenge in the sustainable development process: Analysis of confirmed cases of COVID-19 (new type of coronavirus) through a binary classification using artificial intelligence and regression analysis. *Sustainability* **12**, 2427 (2020)
16. R Core Team: R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2020)
17. Rodriguez-Morales, A.J., Cardona-Ospina, J.A., Gutiérrez-Ocampo, E., et al.: Clinical, laboratory and imaging features of Covid-19: a systematic review and meta-analysis. *Travel Med. Infect. Dis.* **34**, 1623 (2020)
18. Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J.M., Yan, P., Chowell, G.: Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. *Infect. Dis. Modell.* **5**, 256–263 (2020)
19. Shumway, R., Stoffer, D.: Time Series Analysis and Its Applications: With R Examples. Springer Texts in statistics. Springer, Berlin (2006)
20. Wood, S.: Generalized Additive Models: An Introduction with R Chapman and Hall/CRC Texts in Statistical Science Series. Chapman & Hall/CRC, Boca Raton (2006)
21. World Health Organization: Coronavirus diseases COVID-19 situation report-51 a. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf?sfvrsn=1ba62e57_10 (2020). Accessed 25 May 2020
22. World Health Organization: Coronavirus diseases COVID-19 situation report-Weekly epidemiological update—22 December 2020b. <https://www.who.int/publications/m/item/weekly-epidemiological-update---22-december-2020> (2020). Accessed 29 Dec 2020
23. Yonar, H., Yonar, A., Tekindal, M.A., Tekindal, M.: Modeling and forecasting for the number of cases of the COVID-19 pandemic with the curve estimation models, the Box–Jenkins and exponential smoothing methods. *EJMO* **4**(2), 160–165 (2020)
24. Zareie, B., Roshani, A., Mansournia, M.A., Rasouli, A., Moradi, G.: A model for COVID-19 prediction in Iran based on China parameters. *Arch. Iran Med.* **23**(4), 244–248 (2020)

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