STATISTICS IN TRANSITION new series, September 2022
Vol. 23, No. 3, pp. 147–165, DOI 10.2478/stattrans-2022-0035
Received – 24.12.2020; accepted – 08.04.2022

Modified exponential time series model with prediction of total COVID-19 cases in Belgium, Czech Republic, Poland and Switzerland

Wachirapond Permpoonsinsup[1], Rapin Sunthornwat[2]

ABSTRACT

The coronavirus (COVID-19) pandemic affected every country worldwide. In particular, outbreaks in Belgium, the Czech Republic, Poland and Switzerland entered the second wave and was exponentially increasing between July and November, 2020. The aims of the study are: to estimate the compound growth rate, to develop a modified exponential time-series model compared with the hyperbolic time-series model, and to estimate the optimal parameters for the models based on the exponential least-squares, three selected points, partial-sums methods, and the hyperbolic least-squares for the daily COVID-19 cases in Belgium, the Czech Republic, Poland and Switzerland. The speed and spreading power of COVID-19 infections were obtained by using derivative and root-mean-squared methods, respectively. The results show that the exponential least-squares method was the most suitable for the parameter estimation. The compound growth rate of COVID-19 infection was the highest in Switzerland, and the speed and spreading power of COVID-19 infection were the highest in Poland between July and November, 2020.

Key words: COVID-19, modified exponential time-series model, method of parameter estimation, compound growth rate.

1. Introduction

Since the end of 2019, the coronavirus disease 2019 (COVID-19) outbreak caused by the SARS-Cov-2 virus, which started in Wuhan of Hubei Province, China, has spread throughout the world. The outbreak in Europe has entered the second wave with increasing numbers of COVID-19 cases in many countries. As of November 12, 2020,

[1] Industrial Technology and Innovation Management Program, Faculty of Science and Technology, Pathumwan Institute of Technology, Bangkok, Thailand. E-mail: wachirapond@pit.ac.th.
ORCID: https://orcid.org/0000-0003-4033-493X.

[2] Industrial Technology and Innovation Management Program, Faculty of Science and Technology, Pathumwan Institute of Technology, Bangkok, Thailand. E-mail: rapin@pit.ac.th, Corresponding Author.
ORCID: https://orcid.org/0000-0001-8981-5107.

there were 12,914,903 cases and 306,504 deaths in Europe (Worldometer, 2020). Especially, the second-wave COVID-19 outbreaks in Belgium, the Czech Republic, Poland, and Switzerland rapidly spread with what looks like an exponential function; at the same time point, there were 507,475 cases and 13,561 deaths in Belgium, 438,805 cases and 5,570 deaths in the Czech Republic, 641,496 cases and 9,080 deaths in Poland, and 243,472 cases and 3,113 deaths in Switzerland. Measures imposed by the governments of these countries, such as lockdown policies, face mask-wearing in public areas, encouraging hand washing, avoiding public areas, and prohibiting people from assembling were launched to control and protect the population from the spread of COVID-19. In addition, physical and social distancing have remained in practice in many areas, while learning from home for students and working from home have become necessary policies. Thus, the spread of COVID-19 has gone mainly unchallenged due to a lack of medical equipment and personnel to fight the pandemic.

Forecasting the number of COVID-19 cases (the number of people contracting the disease) is essential for planning the necessary provisions for medical treatment (hospital beds, ventilators, personal protective equipment, etc.). Many models for forecasting COVID-19 cases comprising time-series data have been studied. For example, linear regression analysis, machine learning, vector support regression machine, and autoregressive integrated moving average (ARIMA) models based on the linear relationship between the time variable and dependent variables have been popular for establishing models for forecasting COVID-19 cases. Forecasting new cases and new deaths from COVID-19 in Ethiopia was investigated by Argaru (2020). Linear regression analysis of COVID-19 data comprising new cases, deaths, the number of days, and recoveries in May and June to estimate the parameters for a forecasting model has been reported. The relationship between COVID-19 data and time has been analysed by using Pearson's correlation analysis. The results show that there is a correlation between new COVID-19 cases and deaths, while the number of days and new recoveries were significant to the new deaths. For the COVID-19 outbreak in Henan province, China, linear regression analysis was adopted to estimate parameters for constructing a forecasting model and to study the relationship between the number of people from Wuhan who had travelled and the number of cases in 18 cities in Henan province. The results show a statistically significant linear correlation between the number of people traveling from Wuhan and the number of cases (Cheng, 2020).

The COVID-19 outbreak in India has caused socioeconomic recession and mounting deaths. Both multiple and linear regression analyses have been adopted to predict the number of deaths and to study correlations in the COVID-19 data from India, with the ability of the developed predictive model being based on autoregression (Ghosal et al., 2020). The dependent variable (the number of active cases) in the forecasting model was correlated with the independent variables (the number of cases, deceased, and recovered). In a study of public health responses in various countries by Rath, Tripathy and Tripathy (2020), the performances of forecasting models were compared based on the coefficient of determination and correlation. Correlations between intervention scores, daily new cases, and doubling time were significant for identifying epidemiological changes in the spread of COVID-19. In research involving linear and polynomial regression analysis for predicting the COVID-19 fatality rate in Nigeria by Suleiman et al. (2020), the results reveal that the polynomial regression model is suitable for predicting the COVID-19 fatality rate in this particular country.

In other studies, linear regression analysis was employed by Melik-Huseynov et al. (2020) to estimate new cases of COVID-19. Simple regression analysis was applied by Losif et al. (2020) to study the incidence of correlation between the COVID-19 peak or plateau and air traffic volume. Forecasting models based on polynomial regression were investigated by Ekum and Ogunsanya (2020) to forecast new cases of COVID-19; their results show that the cubic polynomial regression model performed better than other polynomial regressions. Calculating the fatality rate based on linear regression analysis and comparing its efficacy among countries affected by the COVID-19 pandemic was conducted by Hoseinpour et al. (2020). Support vector regression as a predictive model for the duration of spread and analysis of growth and transmission rates was used to evaluate the correlation between COVID-19 outbreaks and weather conditions by Yadav, Perumal and Srinivas (2020). Machine learning was applied as a forecasting model based on linear regression, least absolute shrinkage and selection operator, support vector machine, and exponential smoothing for the number of COVID-19 patients, new infection cases, deaths, and recoveries by Rustam et al. (2020); their results proved that exponential smoothing offered the best performance. Linear regression and support vector machine analyses have been used for predicting the number of COVID-19 cases to aid decision-making by the government in India (Likhesh et al., 2020). Prediction models and comparison between the susceptible-exposed/infectious-recovered model and regression analysis were used to predict the number of COVID-19 cases in India by Pandey et al. (2020).

ARIMA models have often been applied to COVID-19 time-series data. An ARIMA model and regression analysis were used to estimate the mortality rate of COVID-19 by Chaurasia and Pal (2020). Forecasting COVID-19 time-series data based on an ARIMA model in the US, Brazil, India, Russia, and Spain was investigated by Sahai et al. (2020). An ARIMA model was created to forecast new cases and deaths from COVID-19 time-series data by Yang et al. (2020). An ARIMA model for short-term prediction was developed by Fang, Wang and Pan (2020) to predict COVID-19 cases, deaths, and recoveries in Russia. An ARIMA model was applied by Benvenuto et al. (2020) to a COVID-19 time-series dataset from the Johns Hopkins database for forecasting the trend and incidence of COVID-19 outbreak. Singh et al. (2020)

developed an ARIMA model for predicting confirmed cases, deaths, and recoveries with a spatial map showing the intensity of each criterion. Moreover, they used the Akaike information criterion to validate the ARIMA model.

Because linear regression analysis, linear machine learning, linear support vector regression, and ARIMA model are dependent on the linear combination of time as the independent variable to predict dependent variables, our aim was to develop a modified exponential time-series model compared with hyperbolic time-series model that is nonlinear and uses the exponential growth rate to forecast the number of COVID-19 cases increasing rapidly each day in Belgium, the Czech Republic, Poland, and Switzerland. Herein, the accuracy and validation of the developed model are reported, while its ability to predict the speed and spreading power of the daily COVID-19 cases are illustrated.

2. Methods

2.1. Derivation of the modified exponential time-series model

A time series is a sequence of observations taken sequentially in time (George et al., 2015). The number of COVID-19 cases per day is an example of a time series. It can be represented by modelling its curve as the solution of a differential equation with time as the independent variable. In this research, a modified exponential curve is adopted to analyse and forecast the daily COVID-19 cases as follows.

Let $y(t)$ be the number of total daily COVID-19 cases at time t. Differential equation which represents the speed of COVID-19 cases and is solved into the modified exponential curve $y(t)$ can be derived as follows:

$$\frac{dy}{dt} = b\ln(c)c^t;\ y(0) = a + b;\ a, b, c \in \Re \tag{1}$$

Taking integral both sides of Equation (1), the result becomes

$$\int \frac{dy}{b\ln(c)} = \int c^t dt$$

$y(t) = bc^t + Kb\ln(c)$ where K is an arbitrary constant.

With initial condition $y(0) = a + b$, the K value can be carried out as

$$K = \frac{a}{b\ln(c)}$$

Therefore, a modified exponential curve is

$$y(t) = a + bc^t \tag{2}$$

where a, b, c are the parameters.

2.2. Compound growth rate and parameter estimations with the algorithm

Let $\{y(t)\}_{t=1}^{n}$ be a time series of the number of COVID-19 cases. The dependent variable y , which is related to independent variable t has the modified exponential relationship to t as in Equation (2). In this research, the exponential least-squares method, three selected points, partial-sums method, and the hyperbolic least-squares method were employed for the estimation of parameters as follows.

2.2.1. The exponential least-squares method

The exponential least-squares method is to seek an approximating function that best fits the data points (Kharab and Guenther, 2012). It is based on the sum of squares error (SSE) defined by

$$SSE = (y - \hat{y})^2 \tag{3}$$

where y is an actual value of time series and $\hat{y}$ is a forecasted value of time series.

The partial derivative is taken into both sides of Equation (3). Then, it is determined to be zero for evaluating the parameters a, b, c based on the minimum of the sum of squares error.

$$\frac{\partial}{\partial a} SSE = \frac{\partial}{\partial a}(y - \hat{y})^2 = 0,$$
$$\frac{\partial}{\partial b} SSE = \frac{\partial}{\partial b}(y - \hat{y})^2 = 0,$$
$$\frac{\partial}{\partial c} SSE = \frac{\partial}{\partial c}(y - \hat{y})^2 = 0.$$

In addition, the compound growth rate of the time series with initial value y_0 is defined as

$$y(t) = y_0(1 + r)^t \tag{4}$$

Rewriting Equation (2), the result becomes

$$Y(t) = A + Bt$$

where $A = \ln(a) + \ln(b) = \ln(ab)$, $B = \ln(c)$, $Y = \ln(y)$.

An estimate of the compound growth rate r based on the least squares estimation for estimation of parameters A and B is given by

$$\hat{B} = \frac{\dfrac{\sum_{t=1}^{n} t^2 y(t)}{n} - \dfrac{\sum_{t=1}^{n} t y(t) \sum_{t=1}^{n} t}{n}}{\sum_{t=1}^{n} t^2 - \dfrac{\left(\sum_{t=1}^{n} t\right)^2}{n}} \quad \text{and} \quad \hat{A} = \bar{Y} - \hat{B}\bar{t}$$

The estimator of parameter b is given by

$$\hat{c} = \exp(\hat{B}) = 1 + \hat{r}$$

where $\hat{r}$ is the estimator of the compound growth rate.

Therefore, an estimate of the compound growth rate is given by

$$\hat{r} = \hat{c} - 1 \tag{5}$$

Also, Student's T-test is a statistic for significant test of the compound growth rate given by

$$T = \frac{\hat{B}}{SE(\hat{B})}; \; df = n - 2. \tag{6}$$

The decision of the significance of the compound growth rate is dependent on the comparison between the calculated value of $|T|$ with the critical value of T or on the consideration of p-value.

2.2.2. The three selected points method

The estimation of parameters of the modified exponential curve is represented by the three selected points method (Das and Chakrabarty, 2017). Three points of the time series coordinates $(t_1^*, y_1^*), (t_2^*, y_2^*), (t_3^*, y_3^*)$ along the time series $\{y(t)\}_{t=1}^{n}$ are selected to estimate parameters a, b, c.

$$y_1^* = a + bc^{t_1^*} \tag{7}$$

$$y_2^* = a + bc^{t_2^*} \tag{8}$$

$$y_3^* = a + bc^{t_3^*} \tag{9}$$

where $h = t_2^* - t_1^* = t_3^* - t_2^*$.

By the algebraic way, the system of Equations (7)-(9) is solved for the estimation of parameters $\hat{a}, \hat{b}, \hat{c}$ as

$$\hat{c} = \left[\frac{y_3^* - y_2^*}{y_2^* - y_1^*}\right]^{\frac{1}{h}}$$

$$\hat{b} = \frac{y_3^* - y_2^*}{\hat{c}^{t_2^*}(\hat{c}^h - 1)} = \frac{y_2^* - y_1^*}{\hat{c}^{t_1^*}(\hat{c}^h - 1)}$$

$$\hat{a} = y_3^* - \hat{b}\hat{c}^{t_3^*} = y_2^* - \hat{b}\hat{c}^{t_2^*} = y_1^* - \hat{b}\hat{c}^{t_1^*}$$

2.2.3. The partial sums method

The partial sums method (Ikaya et al., 2005) is based on the partition of the time series data into three categories with equal length n points. The dependent variable is $y = \{y_1, y_2, y_3, ..., y_n; y_{n+1}, y_{n+2}, y_{n+3}, ..., y_{2n}; y_{2n+1}, y_{2n+2}, y_{2n+3}, ..., y_{3n}\}$ and the time independent variable is

$$t = \{1, 2, 3, ..., n; n+1, n+2, n+3, ..., 2n; 2n+1, 2n+2, 2n+3, ..., 3n\}.$$

Let S_1, S_2, S_3 be the partial sums of the partitions of the dependent variable y. Thus,

$$S_1 = \sum_{t=1}^{n} y_t = an + \frac{bc(c^n - 1)}{c - 1} \tag{10}$$

$$S_2 = \sum_{t=n+1}^{2n} y_t = an + \frac{bc^{n+1}(c^n - 1)}{c - 1} \tag{11}$$

$$S_3 = \sum_{t=2n+1}^{3n} y_t = an + \frac{bc^{2n+1}(c^n - 1)}{c - 1} \tag{12}$$

The algebraic way is adopted to carry out Equations (10)-(12) for the estimation of parameters $\hat{a}, \hat{b}, \hat{c}$ as

$$\hat{c} = \left[\frac{S_3 - S_2}{S_2 - S_1} \right]^{\frac{1}{n}}$$

$$\hat{b} = \frac{(S_2 - S_1)(\hat{c} - 1)}{\hat{c}} \left[\frac{S_3 - S_2}{S_2 - S_1} - 1 \right]^{-2}$$

$$\hat{a} = \frac{S_1(\hat{c} - 1) - \hat{b}\hat{c}(\hat{c}^n - 1)}{n(\hat{c} - 1)} = \frac{S_2(\hat{c} - 1) - \hat{b}\hat{c}^{n+1}(c^n - 1)}{n(\hat{c} - 1)} = \frac{S_3(\hat{c} - 1) - \hat{b}\hat{c}^{2n+1}(\hat{c}^n - 1)}{n(\hat{c} - 1)}$$

2.2.4. The hyperbolic least-squares method

The hyperbolic least-squares method (Kharab and Guenther, 2012) is a nonlinear model estimation. It is the fitting given observations with hyperbolic time-series model, which is given as

$$y(t) = a + \frac{b}{t}$$

Setting $Y(t) = y(t)$, $\alpha = a$, $\beta = b$, and $T = \frac{1}{t}$, the hyperbolic time-series model can be transformed as

$$Y(T) = \alpha + \beta T$$

Then, the least-squares method is applied to the estimation of parameters a and b.

2.2.5. Statistics for the accuracy and validation of the time-series model and spreading power

In this section, the accuracy and validation of the time series model are measured. The spreading power is also measured. The measurement of validation of the time series model is evaluated by the Root Mean Squared Percentage Error ($RMSPE$) as

$$RMSPE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} \left(\frac{y(t) - \hat{y}(t)}{y(t)} \right)^2} \tag{13}$$

where $y(t)$ is an actual value of y and $\hat{y}(t)$ is a forecasted value of y.

For accuracy of the time series model, the coefficient of determination (R^2) is defined as

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{t=1}^{n}\left[y(t) - \hat{y}(t)\right]^2}{\sum_{t=1}^{n}\left[y(t) - \bar{y}\right]^2} \tag{14}$$

where RSS is the Sum of Squares of Residuals,

TSS is the Total Sum of Squares.

The Root Mean Square (RMS) (Jones, 2019) is measured as the spreading power of the COVID-19 cases time series. The RMS can be defined as

$$RMS = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left[y(t)\right]^2} \tag{15}$$

2.2.6. The algorithm for evaluating the parameters, estimating the derivative, root-mean-square (RMS), RMS percentage error (RMSPE), and estimating the compound growth rate

In this section, the algorithm for this research is demonstrated.

Algorithm

Input: total COVID-19 cases

$y \leftarrow$ total COVID-19 cases

$n \leftarrow$ length(y)

$t \leftarrow 1:n$

$Para_{ExpoLeast}, RMSPE_{ExpoLeast}, R^2_{ExpoLeast} \leftarrow$ Expoleastsquare(modifiedexpo, t, y)

$Para_{Three}, RMSPE_{Three}, R^2_{Three} \leftarrow$ Threepoints(modifiedexpo, t, y)

$Para_{Partial}, RMSPE_{Partial}, R^2_{Partial} \leftarrow$ Partialsums(modifiedexpo, t, y)

$Para_{HyperLeast}, RMSPE_{HyperLeast}, R^2_{HyperLeast} \leftarrow$ Hyperleastsquare(hypebolic, t, y)

The optimal estimate parameter is based on the minimum of $RMSPE$ and the maximum of R^2

$del \leftarrow 1$

For $t \leftarrow 2:n-1$

$dy(t) \leftarrow (y(t+1) - y(t-1))/2/del$

End

$SS, SS_1, SS_2, SS_3, SS_4 \leftarrow 0$

For $i \leftarrow 1:n$

$SS \leftarrow SS + (y(i))^2$

$SS_1 \leftarrow SS_1 + t^2 y(t)$

$SS_2 \leftarrow SS_2 + ty(t)$

$SS_3 \leftarrow SS_3 + t$

$SS_4 \leftarrow SS_4 + t^2$

End

$RMS \leftarrow sqrt(\frac{1}{n} SS)$

$$\hat{B} = \frac{\frac{SS_1}{n} - \frac{(SS_2)(SS_3)}{n}}{SS_4 - \frac{(SS_3)^2}{n}}$$

$\hat{c} \leftarrow \exp(\hat{B})$

$\hat{r} \leftarrow \hat{c} - 1$

Output: estimate parameters, estimate derivative, *RMS*, *RMSPE*, $\hat{r}$

2.3. Data collection

The sampled countries, Belgium, the Czech Republic, Poland, and Switzerland, are selected for investigation because the spreading of COVID-19 in these countries is severe outbreak at the second wave in the manner of exponential outbreak. The total COVID-19 cases were only collected to model in the first stage of COVID-19 outbreak because the first stage is exponentially increasing. Then, the total COVID cases will pass the inflation point and will be converged to the carrying capacity (Areepong and Sunthornwat, 2021). For selected four countries, the outbreak in the first stage of the second wave started in the second wave between July and November, 2020. The data for this research are the number of daily total COVID-19 cases in Belgium, the Czech Republic, Poland, and Switzerland. The data is collected at the Worldometers website (Worldometer, 2020). This website reveals the real time data about world population, government and economics, society and media, environment, food, water, energy, health, as well as COVID-19 statistics. The duration of time for collecting data for making the forecasting model is dependent on the severity in each country. The time range for collection of data in Belgium is from July 15, 2020 ($t = 0$) to November 3, 2020 ($t = 111$). The time range for collection of da ta in Czech Republic is from August 23, 2020 ($t = 0$) to November 3, 2020 ($t = 72$). The time range for collection of data in Poland is from September 1, 2020 ($t = 0$) to November 3, 2020 ($t = 63$). The time range for collection of data in Switzerland is from September 2, 2020 ($t = 0$) to November 3, 2020 ($t = 62$). For the out of sample data, the time range is extended to 5 days from the last day of the data for making the forecasting model.

3. Results

The results of this research concern estimating the parameters and forecasting using the models, as well as the compound growth rate and spreading power of COVID-19 in Belgium, the Czech Republic, Poland, and Switzerland.

3.1. The parameters and forecasting models for the number of daily COVID-19 cases in Belgium

Here, we present the estimated parameters and forecasting models for the daily COVID-19 cases in Belgium. The estimated parameters evaluated by each method are as follows: $\hat{a} = 72207.129$, $\hat{b} = 316.821$, and $\hat{c} = 1.066$ by using the exponential least-squares method; $\hat{a} = 62757.544$, $\hat{b} = 114.456$, and $\hat{c} = 1.086$ by using the three selected points method; $\hat{a} = 67327.213$, $\hat{b} = 754.656$, and $\hat{c} = 1.058$ by using the partial-sums method; and $\hat{a} = 138260.53$ and $\hat{b} = -188328.83$ by using the hyperbolic least-squares method. The forecasting models based on the three methods and estimating the parameters of the daily COVID-19 cases in Belgium are shown in Figure 1.

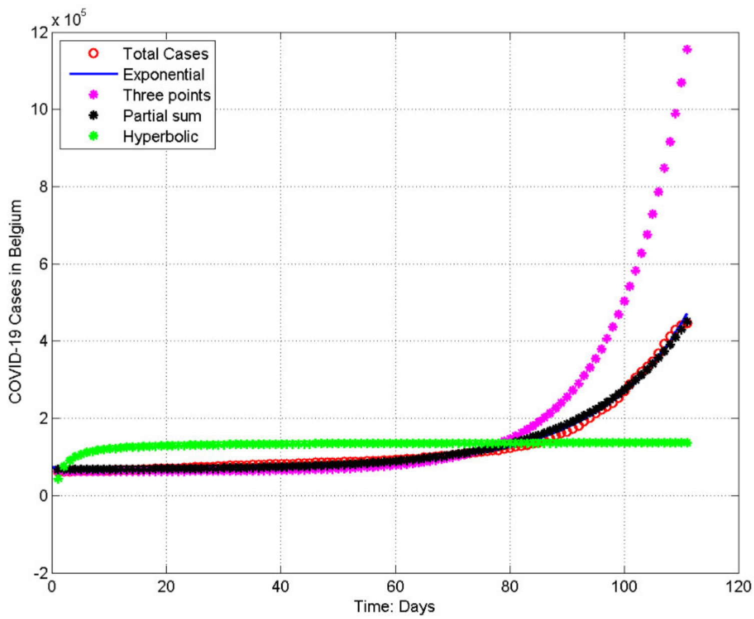

Figure 1. Estimation of the daily COVID-19 cases in Belgium

3.2. The parameters and forecasting models for the daily COVID-19 cases in the Czech Republic

Here, we present the estimated parameters and forecasting models for the daily COVID-19 cases in the Czech Republic. The estimated parameters evaluated by each method are as follows: $\hat{a} = 9077.568$, $\hat{b} = 9261.801$, and $\hat{c} = 1.053$ by using the least-squares method; $\hat{a} = 21701.501$, $\hat{b} = 221.499$, and $\hat{c} = 1.125$ by using the three selected points method; $\hat{a} = 3408.564$, $\hat{b} = 13198.966$, and $\hat{c} = 1.047$ by using the partial-sums method; and $\hat{a} = 125497.33$ and $\hat{b} = -235555.95$ by using the hyperbolic least-squares method. The forecasting models based on the three methods and estimates of the parameters for the daily COVID-19 cases in the Czech Republic are shown in Figure 2.

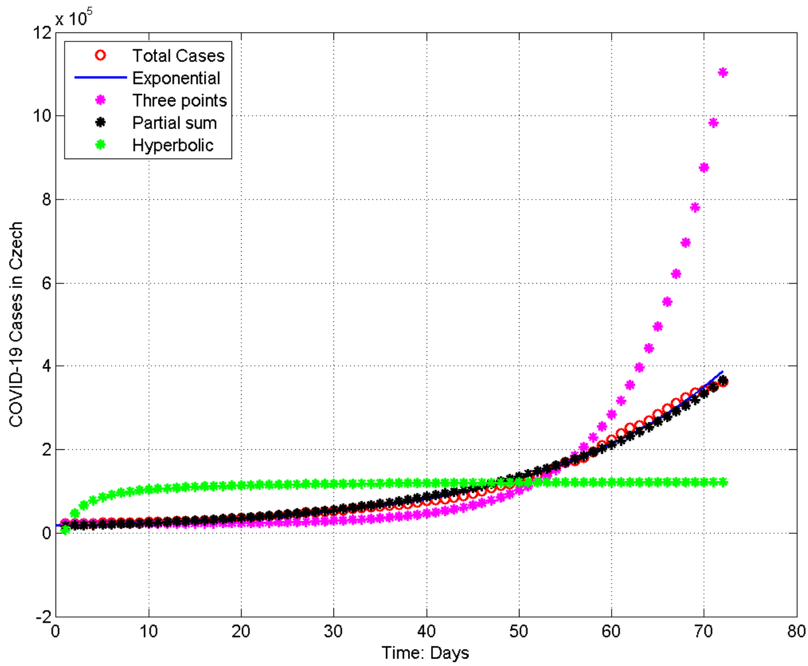

Figure 2. Estimation of the daily COVID-19 cases in the Czech Republic

3.3. The parameters and forecasting models for the daily COVID-19 cases in Poland

Here, we present the estimated parameters and forecasting models for the daily COVID-19 cases in Poland. The estimate parameters evaluated by the three methods are as follows: $\hat{a} = 64655.708$, $\hat{b} = 3192.084$, and $\hat{c} = 1.078$ by using the least-squares method; $\hat{a} = 47097.000$, $\hat{b} = 20825.000$, and $\hat{c} = 1.029$ by using the three selected points method; $\hat{a} = 67958.474$, $\hat{b} = 1649.921$, and $\hat{c} = 1.099$ by using the partial-sums method; and $\hat{a} = 157216.97$ and $\hat{b} = -198139.63$ by using the hyperbolic least-squares method. The forecasting models based on the three methods and the estimated parameters for the daily COVID-19 cases in Poland are shown in Figure 3.

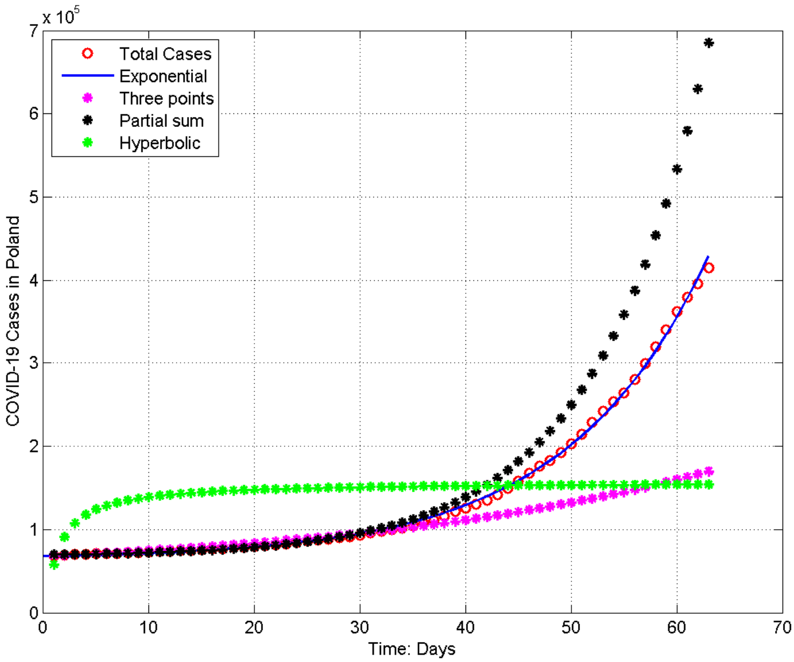

Figure 3. Estimation of the daily COVID-19 cases in Poland

3.4. Parameters and forecasting models for the daily COVID-19 cases in Switzerland

Here, we present the estimate parameters and forecasting models for the daily COVID-19 cases in Switzerland. The estimated parameters evaluated by each method are as follows: $\hat{a} = 44629.163$, $\hat{b} = 699.939$, and $\hat{c} = 1.089$ by using the least-squares method; $\hat{a} = 42258.5002$, $\hat{b} = 504.499$, and $\hat{c} = 1.119$ by using the three selected points method; $\hat{a} = 55456.458$, $\hat{b} = 901.478$, and $\hat{c} = 1.089$ by using the partial-sums method; and $\hat{a} = 77679.67$ and $\hat{b} = -75580.83$ by using the hyperbolic least-squares method.. The forecasting models based on the three methods and the estimated parameters for the daily COVID-19 cases in Switzerland are shown in Figure 4.

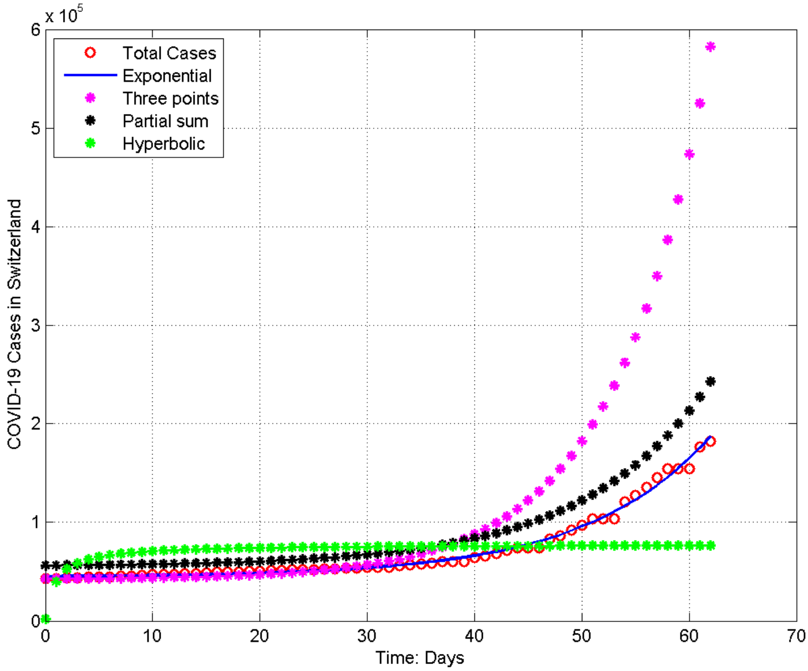

Figure 4. Estimation of the daily COVID-19 cases in Switzerland

3.5. Comparison of the spread of COVID-19 and appropriate parameters

The derivatives of the speed and spreading power of the daily COVID-19 cases in Belgium, the Czech Republic, Poland, and Switzerland are shown in Figures 5 and 6, respectively. In addition, the compound growth rate and forecasting model validation based on $RMSPE$ and R^2 values along with testing of the significance of the compound growth rate based on *p*-values are reported in Table 1. Moreover, forecasting of the daily COVID-19 cases for the out-of-sample data for November 4 – 8, 2020, and comparing the forecasted values with the actual values are summarized in Table 2. The results indicate that the increase in the speed and spreading power of the daily COVID-19 cases in Belgium was higher than for the other countries. Moreover, the compound growth rate for each country was statistically significant (p-value < 0.05). The exponential least-squares method provided the best fitting of the parameter estimations for the four countries, as indicated by the lowest $RMSPE$ and the highest R^2 values.

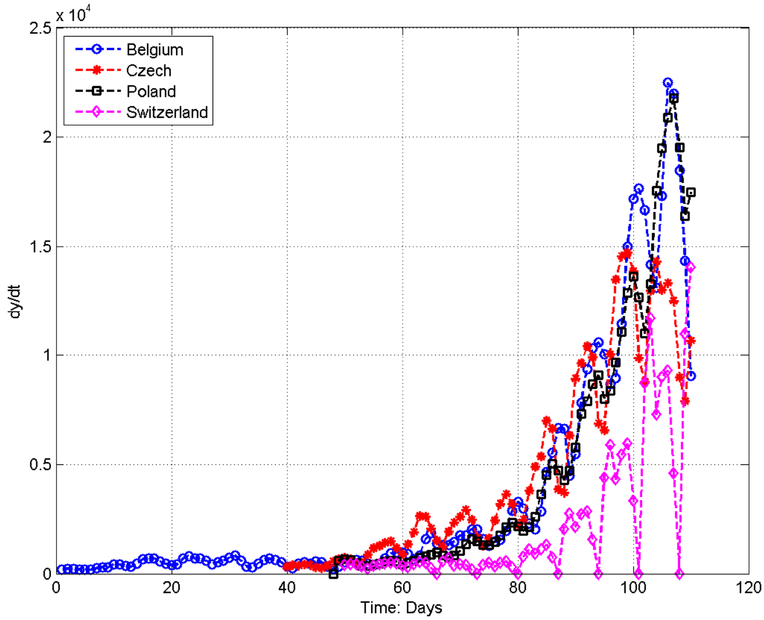

Figure 5. Estimated speed of the increase in daily COVID-19 cases

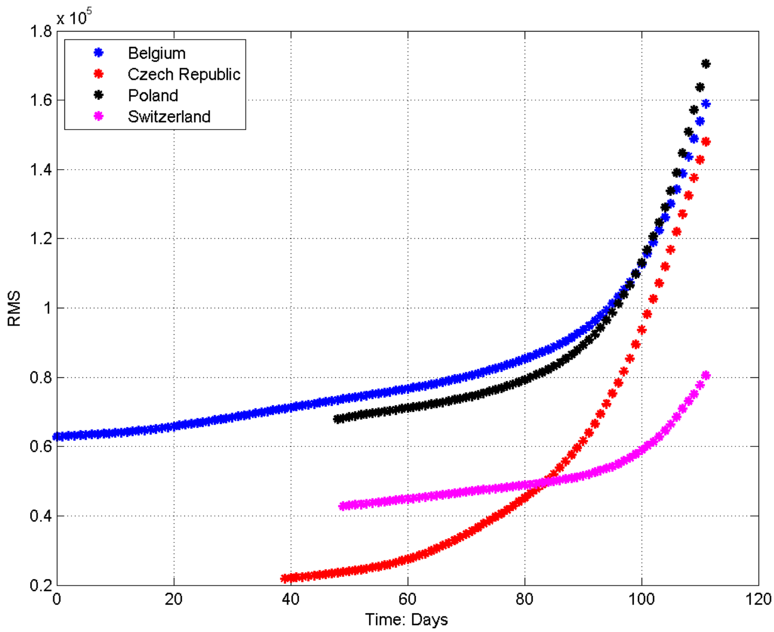

Figure 6. The spreading power for the daily COVID-19 cases

Table 1. The compound growth rate and forecasting model validation for the daily COVID-19 cases.

Country	Estimate B	Compound Growth Rate	Standard Error	t-test	p-value	*RMSPE*	R^2
Belgium	0.064	0.066	0.001	22.879	1.267e-43	0.456	0.982
						0.064	0.992
						0.063*	0.995*
						0.594	0.051
Czech Republic	0.052	0.053	3.613e-07	23.266	6.066e-35	0.543	0.867
						0.091	0.994
						0.058*	0.996*
						1.711	0.101
Poland	0.075	0.078	0.001	22.236	3.056e-31	0.248	0.904
						0.207	0.989
						0.020*	0.999*
						0.641	0.089
Switzer land	0.086	0.090	0.001	17.460	2.076e-25	0.735	0.978
						0.266	0.993
						0.034*	0.994*
						0.406	0.089

Note: * the best value. For each country, the top, middle, and bottom rows are for the models using the three points, partial-sums, exponential least-squares methods, and hyperbolic least-squares methods, respectively.

Table 2. Forecasting COVID-19 cases for the out-of-sample data using the least-squares method.

Country	Time					*RMSPE*	R^2
	Nov 4, 2020	Nov 5, 2020	Nov 6, 2020	Nov 7, 2020	Nov 8, 2020		
Belgium	453310	468213	479341	488044	494168	0.179	0.959
	497600.183	525854.870	555986.239	588118.939	622385.899		
Czech Republic	378717	391949	403497	411219	414827	0.134	0.943
	406855.462	427880.541	450016.929	473323.365	497861.694		
Poland	439536	466679	493765	521640	546425	0.061	0.997
	456977.627	487608.675	520631.283	556232.174	594612.651		
Switzerland	192376	202504	211913	211913	211913	0.137	0.735
	199713.233	213594.438	228718.115	245195.477	263147.688		

Note: For each country, the top line is the actual value and the bottom line is the forecasted value based on the exponential least-squares method.

4. Conclusions

In this research, we applied a modified exponential time series model to forecast daily COVID-19 cases. Belgium, the Czech Republic, Poland, and Switzerland were selected for this research because their curves for the daily COVID-19 cases in the second wave were exponentially increasing. Parameter estimation of the modified exponential time-series model was conducted by using the exponential least-squares method, the three selected points method, and the partial-sums methods. The hyperbolic least-squares time-series model, the other nonlinear model, which is a hyperbolic form, is applied to be compared with the previous models. The optimal forecasting model with the estimated parameters was selected based on having the lowest $RMSPE$ and the highest R^2. Moreover, the compound growth rate, speed, and spreading power of the daily COVID-19 cases were evaluated and compared. The findings show that the exponential least-squares method was the most appropriate method for parameter estimation for the modified exponential time-series model for the daily COVID-19 cases in all four countries. The compound growth rates were statistically significant for each country, with that of Switzerland being slightly higher than in the other countries. Moreover, the speed and spreading power of the daily COVID-19 cases in Belgium were higher than the other countries. When applying the optimal least-squares model to predict the daily COVID-19 cases from the out-of-sample data, the forecasted daily COVID-19 cases were in good agreement with the actual values. Changing the parameters of the modified exponential time-series model made the forecasting model less accurate. Hence, the modified exponential time-series model is suitable for short-term forecasting, and parameter estimation should be evaluated again if the accuracy of the forecasting model is reduced. However, the limitation of this research is that the modified exponential time-series model is effectively used for the first stage of the outbreak because the total COVID-19 cases will exponentially increase in the first stage of the outbreak. Consequently, the total COVID-19 cases will pass the inflation point and converge to the carrying capacity. Future research will encompass other variables related to the COVID-19 situation, such as the number of active cases and the number of deaths, to enable the authorities effectively control a COVID-19 outbreak and protect the population from it.

Acknowledgement

The authors wish to thank Professor Dr. Yupaporn Areepong, Department of Applied Statistics, King Mongkut's University of Technology North Bangkok, Thailand, for support during the conduct of this research.

References

Areepong, Y., Sunthornwat, R., (2021). EWMA control chart based on its first hitting time and coronavirus alert levels for monitoring symmetric COVID-19 cases. *Asian Pacific Journal of Tropical Medicine*, 14(8), pp. 364–374.

Argaru, A. S., (2020). Linear regression model for predictions of COVID-19 New cases and new deaths based on May/June Data in Ethiopia, *Research Square*, doi:10.21203/rs.3.rs-61667/v1.

Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S. and Ciccozzi, M., (2020). Application of the ARIMA model on the COVID-2019 epidemic dataset, *Data in Brief*, 29(105340), pp. 1–4.

Chaurasia, V., Pal, S., (2020). COVID-19 pandemic: ARIMA and regression model-based worldwide death cases predictions, *SN Computer Science*, 1(288), pp. 1–12.

Cheng, Y., (2020). Linear regression analysis of COVID-19 outbreak and control in Henan province caused by the output population from Wuhan. *medrxiv*, doi: 10.1101/2020.05.03.20089193.

Das, B., Chakrabarty, D., (2017). Representation of numerical data by exponential curve. *Journal of Mathematics and Systems Sciences*, 13(1-2), pp. 1–6.

Ekum, M. and Ogunsanya, A., (2020). Application of Hierarchical Polynomial Regression Models to Predict Transmission of COVID-19 at Global Level. *International Journal of Clinical Biostatistics and Biometrics*, 6(1), pp. 1–18.

Fang, L., Wang, D. and Pan, G., (2020). Analysis and Estimation of COVID-19 Spreading in Russia Based on ARIMA Model. *SN Comprehensive Clinical Medicine*, doi: 10.1007/s42399-020-00555-y.

George, E. P. B., Gwilym M. J., Gregory, C. R. and Greta, M. L., (2015). *Time Series Analysis: Forecasting and Control.* 5th ed. Hoboken, New Jersey: John Wiley & Sons, Inc.

Ghosal, S., Sengupta, S., Majumder, M. and Sinha, B., (2020). 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). *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 4(4), pp.311–315.

Hoseinpour, D. A., Alizadeh, M., Derakhshan, P., Babazadeh, P. and Jahandideh, A., (2020). Understanding epidemic data and statistics: A case study of COVID-19. *Journal of Medical Virology*, 92(7), pp. 868–882.

Ikaya, K. K. Z., Balcióglu, M. S., Yolcu, H. I. and Karabag, K., (2005). The application of exponential method in the analysis of growth curve for Japanese quail. *European Poultry Science*, 69(5), pp. 193–198.

Jones, A. R., (2019). *Probability, statistics and other frightening stuff*. 2nd ed. Abingdon, UK: Routledge.

Kharab, A., Guenther, B., (2012). *An introduction to numerical methods: A MATLAB approach*. 3rd dd. Boca Raton, FL, USA: CRC Press.

Likhesh, K., Yogesh, P., Puja, K. and Vishal, J., (2020). Prediction of coronavirus Covid-19 cases using linear regression and support vector machine. *International Journal of Advanced Science and Technology*, 29(5s), pp.1911–1924.

Melik-Huseynov, D. V., Karyakin, N. N., Blagonravova, A. S., Klimko, V. I., Bavrina, A. P., Drugova, O. V., Saperkin, N. V. and Kovalishena, O. V., (2020). Regression models predicting the number of deaths from the new coronavirus infection. *Modern Technologies in Medicine*, 12(2), pp.6–13.

Pandey, G., Chaudhary, P., Gupta, R. and Pal, S., (2020). SEIR and regression model based COVID-19 outbreak predictions in India. *medrxiv*, doi: 10.1101/2020.04.01.20049825.

Rath, S., Tripathy, A. and Tripathy, A., (2020). Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(5), pp. 1467–1474.

Rustam, F., Reshi, A., Mehmood, A., Ullah, S., On, B., Aslam, W. and Choi, G. S., (2020). COVID-19 future forecasting using supervised machine learning models. *IEEE Access*, 8(2020), pp. 101489–101499.

Sahai, A. K., Rath, N., Sood, V. and Singh, M. P., (2020). ARIMA modelling and forecasting of COVID-19 in top five affected countries. *Diabetes and Metabolic Syndrome*, 14(5), pp. 1419–1427.

Singh, R. K., Rani, M., Bhagavathula, A. S., Sah, R., Rodriguez-Morales, A. J., Kalita, H., Nanda, C., Sharma, S., Sharma, Y. D., Rabaan, A. A., Rahmani, J. and Kumar, P., (2020). Prediction of the COVID-19 pandemic for the top 15 affected countries: Advanced autoregressive integrated moving average (ARIMA) model. *JMIR Public Health and Surveillance*, 6(2), pp. 1–10.

Suleiman, A. A., Suleiman, A., Abdullahi, U. A. and Suleiman, S. A., (2020). Estimation of the case fatality rate of COVID-19 epidemiological data in Nigeria using statistical regression analysis. *Biosafety and Health*, 3(1) pp. 4–7, doi:10.1016/j.bsheal.2020.09.003.

WORLDOMETER, (2020). *COVID-19 coronavirus pandemic*, [online]. Available at: <https://www.worldometers.info/coronavirus/country> [Accessed 12 November 2020].

Yadav, M., Perumal, M. and Srinivas, M., (2020). Analysis on novel coronavirus (COVID-19) using machine learning methods. *Chaos, Solitons and Fractals*, 139 (110050), pp. 1–12.

Yang, Q., Wang, J., Ma, H. and Wang, X., (2020). Research on COVID-19 based on ARIMA model—taking Hubei, China as an example to see the epidemic in Italy. *Journal of Infection and Public Health*, 13(10), pp.1415–1418.