mathematics

MDPI

Article

EWMA Control Chart Integrated with Time Series Models for COVID-19 Surveillance

Chen-Rui Hsu and Hsiuying Wang *

Institute of Statistics, National Yang Ming Chiao Tung University, Hsinchu 300093, Taiwan; frankdark000@gmail.com

* Correspondence: wang@stat.nycu.edu.tw

Abstract: The global outbreak of coronavirus disease 2019 (COVID-19) has posed a severe threat to public health and caused widespread socioeconomic disruptions in the past several years. While the pandemic has subsided, it is essential to explore effective disease surveillance tools to aid in controlling future pandemics. Several studies have proposed methods to capture the epidemic trend and forecast new daily confirmed cases. In this study, we propose the use of exponentially weighted moving average (EWMA) control charts integrated with time series models to monitor the number of daily new confirmed cases of COVID-19. The conventional EWMA control chart directly monitors the number of daily new confirmed cases. The proposed methods, however, monitor the residuals of time series models fitted to these data. In this study, two time series models—the auto-regressive integrated moving average (ARIMA) model and the vector auto-regressive moving average (VARMA) model—are considered. The results are compared with those of the conventional EWMA control chart using three datasets from India, Malaysia, and Thailand. The findings demonstrate that the proposed method can detect disease outbreak signals earlier than conventional control charts.

Keywords: average run length; COVID-19; control charts; residuals; time series model

MSC: 62M10

check for updates

Academic Editors: Manuel Alberto M. Ferreira and Li-pang Chen

Received: 2 December 2024
Revised: 22 December 2024
Accepted: 30 December 2024
Published: 30 December 2024

Citation: Hsu, C.-R.; Wang, H. EWMA Control Chart Integrated with Time Series Models for COVID-19 Surveillance. *Mathematics* **2025**, *13*, 115. https://doi.org/10.3390/math13010115

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The emergence of the coronavirus disease 2019 (COVID-19) pandemic has brought unprecedented challenges and disruption to the world. The first COVID-19 case was found in Wuhan in December 2019 [1]. The initial outbreak of this infectious disease profoundly affected societies, triggering widespread health concerns, economic uncertainties, and social unrest [2,3]. As the virus rapidly spreads throughout the world, its devastating consequences prompt a critical need for innovative methodologies to understand and manage the evolving crisis. Obtaining an estimate of the anticipated upcoming cases is crucial, as it helps healthcare systems and relevant authorities to prepare for the forecasting of daily new confirmed cases [4–6].

Time series analysis is a useful tool for understanding the underlying causes of trends or systemic patterns over time in epidemiology applications, particularly in forecasting COVID-19 dynamics [7–9]. Widely applied in traditional financial statistics, time series models, such as auto-regressive integrated moving average (ARIMA) and vector auto-regressive moving average (VARMA), have found extensive use in predicting confirmed cases of COVID-19 [10]. This is attributed to their ability to capture temporal dependencies

within data, providing a more nuanced understanding of trends over time compared to conventional models [11].

In addition to using time series models to forecast COVID-19 cases, control charts can also be used to monitor COVID-19 dynamics [12–14]. The utility of control charts in industrial applications is well established, providing a systematic means of identifying variations and ensuring quality control [15,16]. In the context of epidemiology, the exponentially weighted moving average (EWMA) control chart is instrumental in monitoring biological data that are series-correlated, especially where COVID-19 confirmed cases are highly correlated with previous cases [17]. Different from the prior research that primarily focused on the use of ARIMA and VARMA models for the prediction of COVID-19 [18], this study aims to integrate the ARIMA and VARMA models with EWMA control chart to develop residual-based EWMA control charts (RECCs) for monitoring daily COVID-19 confirmed cases. Residual-based charts use residuals from the time series model as the data being monitored. The conventional EWMA control chart monitors the observed data instead of residuals. RECCs are compared with the conventional EWMA control chart in terms of average run length (ARL) criterion. The results show that RECCs outperform the conventional observation-based EWMA control chart in three datasets, including data from India, Malaysia, and Thailand.

2. Materials and Methods

2.1. Data

This study utilizes data obtained from the website "Our World in Data" (https://ourworldindata.org/) for analysis. The data are provided as a Supplementary File (Table S1). The aim is to predict the second major wave of COVID-19 early by monitoring the number of daily confirmed cases. Data from three countries—India, Thailand, and Malaysia—are included in this study. We investigate the data to identify the second major wave in these countries. The data from up to one or half a month before the second wave are used as historical data to construct control charts.

The historical data cover the timeframe from January 2020 to one or half a month before the onset of the second wave in each respective country. The period of the historical data and the significant events regarded as another wave are provided in Table 1.

Table 1. Data information.

Country	Period	Significant Event
India	30 January 2020, to 28 February 2021	On 24 March 2021, a key announcement revealed a viral mutation, marking a critical point in the outbreak.
		By 29 March, confirmed cases had surpassed 12 million, underscoring growing severity.
Malaysia	3 January 2020 to 31 January 2022	On 12 February 2022, Malaysia surpassed three million confirmed cases.
		By 16 February, daily cases hit a record high, prompting an investigation.
		On 23 February, daily cases exceeded 30,000 for the first time.
Thailand	3 January 2020 to 30 March 2021	From 14 to 16 April 2021, Thailand tightened preventive measures and control over major transportation, marking a key moment in efforts to curb the virus.

2.2. Time Series Model

Time series analysis is a powerful statistical method used for analyzing a sequence of data points collected over an interval of time, uncovering underlying patterns and trends in the data [19]. Compared to the traditional model that does not account for the time-dependent structure of data, the significant advantage of time series models lies in their ability to capture complex temporal dependencies within data. In this study, we employ time series modeling to analyze COVID-19 data, concentrating on two prominent time series models: the auto-regressive integrated moving average (ARIMA) model and the vector auto-regressive moving average (VARMA) model [20,21].

The mathematical expression of the ARIMA model denoted as ARIMA (p, d, q) is given as follows:

$$v_t = c + \sum_{i=1}^{p} \phi_i v_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t$$

where v_t denotes the dth difference of data, c is a constant, ϕ_i's are the parameters of the auto-regressive (AR) part of the model, θ_i's are the parameters of the moving average (MA) part of the model, and ϵ_t is an error at time t.

ARIMAX is an extension of the traditional ARIMA model that incorporates additional variables, referred to as exogenous variables [22]. The mathematical expression of the ARIMAX model is given as follows:

$$v_t = c + \sum_{i=1}^{p} \phi_i v_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \sum_{i=1}^{m} \beta_i x_{t,i} + \epsilon_t \quad (1)$$

where β_i's are the coefficients of the regressors and $x_{t,i}$'s are the exogenous variables of the time series model.

The VARMA model constitutes a robust extension of the ARIMA model which has been developed to handle multivariate time series data. The mathematical expression of the VARMA(p, q) model is given as follows:

$$\mathbf{Y}_t = c + \sum_{i=1}^{p} \Phi_i \mathbf{Y}_{t-i} + \sum_{i=1}^{q} \Theta_i \boldsymbol{\epsilon}_{t-i} + \boldsymbol{\epsilon}_t$$

where Y_t denotes the h-dimensional data at time t, Φ_i's are har coefficient matrices for the AR part with lag i, Θ_j's are har coefficient matrices for the MA part with lag i, and ϵ_t is the h-dimensional error vector at time t.

The Akaike information criterion (AIC) can be used to evaluate statistical models [23]:

$$\mathrm{AIC} = -2\log(\hat{L}) + 2k$$

where $\hat{L}$ is the maximized value of the likelihood function for the model and k is the number of model parameters. A revised form of AIC known as the corrected Akaike information criterion (AICC) has been proposed that is particularly useful for small sample sizes [24]:

$$\mathrm{AICC} = \mathrm{AIC} + \frac{2k(k+1)}{n-k-1}$$

where n is the sample size. By accounting for the connection between sample size and model parameters, AICC strives to provide a more accurate assessment of model quality in situations with limited data.

2.3. Control Chart

The ARIMAX and VARMA models are integrated with the control chart to monitor the COVID-19 data in this study. The EWMA control chart statistic is the exponentially weighted moving average at time i:

$$z_i = \lambda w_i + (1-\lambda) z_{i-1} \tag{2}$$

where w_i is the data at time i, $z_0 = w_1$, and λ is a constant with $0 < \lambda \leq 1$.

The Equation (2) can be rewritten:

$$z_i = \lambda \sum_{j=0}^{i-1} (1-\lambda)^j w_{i-j} + (1-\lambda)^i z_0$$

The upper control limit (UCL) of an EWMA control chart for monitoring z_i is as follows:

$$z_0 + L\sigma\sqrt{\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2i}]} \tag{3}$$

where L is the factor of the control chart and σ is the standard deviation of the data [25].

When i increases, the term $(1-\lambda)^{2i}$ in (3) decreases. As a result, (3) tends to

$$z_0 + L\sigma\sqrt{\frac{\lambda}{2-\lambda}} \tag{4}$$

as i goes to infinity. In this study, we use (4) as an upper control limit, and L is selected to satisfy certain conditions.

The EWMA control chart is used to monitor COVID-19 outbreaks [26,27]. To the best of our knowledge, the residual-based EWMA control charts integrated with time series models have not been used to monitor COVID-19 data.

The ARL criterion is commonly used to evaluate the performance of a control chart, which is the average number of samples taken before an out-of-control signal is given. The notations ARL_0 and ARL_1 denote the values of ARL when the process is under control and out of control, respectively. When using the ARL criterion to compare the performance of different types of control charts, there is a need to conduct a simulation study to find factors satisfying the same ARL_0 for each control chart, and then compare their ARL_1s for out-of-control cases. The control chart with the smallest ARL_1 is regarded as the preferable control chart. However, in this study, we compare the control charts using the real dataset. As a result, we use a simulation study to construct control charts such that they have the same ARL_0 and then we evaluate their performances by the examination of which chart can detect the signal of the second wave earlier. Because, in this case, we can compare the results using real data, we do not need to conduct a simulation study to compare their ARL_1's.

2.4. The Proposed Methods

To apply the EWMA control chart to monitor the number of daily confirmed cases of COVID-19 data, the daily observed data can be directly used to construct EWMA charts. However, in our analysis, the EWMA control chart based on the daily observed data cannot efficiently detect an out-of-control signal. Therefore, we propose constructing EWMA control charts for the residuals obtained from the ARIMAX or VARMA models to monitor COVID-19 data. In this study, RECCs are compared with the conventional EWMA control chart constructed using observations rather than residuals, referred to as observation-based EWMA control charts (OECCs) in this study.

Let O_t, $t = 1, \cdots, m$ be the daily observation data which are used as historical data. Let $\hat{O}_t, t = h, \cdots, m (h \geq 1)$ be the predicted data obtained from ARIMAX or VARMA. Note that h may not be 1 because calculating $\hat{O}_t$ needs initial data. Let

$$R_t = (O_t - \hat{O}_t) \tag{5}$$

denote the residuals. The OECC is the EWMA control chart constructed based on the data O_t, where $t = 1, \cdots, m$. The OECC constructed is used to monitor future data O_t, $t = m+1, \cdots$. The RECC based on the ARIMA model is the EWMA control chart constructed based on the data R_t, $t = h, \cdots, m$, which are the residuals obtained from the ARIMA model. That is, we use O_t, $t = 1, \cdots, m$ to fit an ARIMAX model, and calculate R_t, $i = h, \cdots, m$. Then, we construct RECC to monitor R_t, $t = m+1, \cdots$. The RECC based on the VARMA model is constructed similarly, except the ARIMAX model is replaced with the VARMA model.

In this study, we only need to consider an upper control limit (UCL) and not a lower control limit, as our primary concern is whether the number of confirmed cases is large rather than small. Therefore, the evaluation of these charts focuses solely on the UCL. These charts are compared under different ARL_0 conditions. The method first involves setting an ARL_0. For the OECC chart, to find a UCL to satisfy this specified ARL_0, we conduct a simulation study by randomly choosing data from historical data $O_t, t = 1, \cdots, m$. Then, we can find an L in (3) such that the ARL is close to the specified ARL_0. For the RECC chart, we follow the same steps as the OECC chart to find the value of L, except for choosing data from historical data $R_t, t = 1, \cdots, m$. Based on these constructed charts, we compare the ARL_1 of these charts.

Procedure 1. The implementation of the RECC method based on the ARIMAX model.

Step 1. Fit the one-dimension observation data O_t, and select the exogenous variables data $x_{t,i}$'s with the ARIMAX model based on AIC or AICC criteria.

Step 2. Calculate the residuals R_t based on the fitted ARIMAX model.

Step 3. Set an ARL_0 and determine a parameter value λ for the EWMA control chart.

Step 4. Find a factor L for the EWMA control chart by resampling the data from the residuals obtained in Step 2 to achieve the ARL_0 determined in Step 3. Then, the EWMA control chart with this obtained L is the proposed RECC based on the ARIMAX model.

The steps of performing RECC based on the VARMA model are similar to those outlined in Procedure 1, except for replacing one-dimension observation data and the ARIMAX model with multiple-dimension data and the VARMA model, respectively.

3. Results

3.1. India

By investigating the history and data, we identified several notable occurrences as marking the beginning of the second wave in India (Table 1). On 24 March, a pivotal announcement was made about the detection of a viral mutation. By 29 March, the cumulative number of confirmed cases had surpassed 12 million, underscoring the gravity of the escalating situation. We regard these events as second waves and assess the effectiveness of the proposed method in predicting them.

The historical data used for the India case consist of approximately 396 days, spanning from 30 January 2020 to 28 February 2021. The monitoring period is from 1 March 2021 to 31 March 2021. To construct RECC based on the ARIMAX model, we identified possible interventional events and selected them as exogenous variables based on the AIC or AICC criteria. The exogenous variables selected for the ARIMAX are the three variables "lockdown", "unlock", and "vaccine" presented in Table 2. In the VARMA model, three-

dimensional data are used, including daily COVID-19 confirmed case numbers, daily COVID-19 testing numbers, and daily new death counts. In this study, we mainly focus on monitoring the daily COVID-19 confirmed case number. The other two variables in the VARMA model are used to increase the accuracy of model fitting.

Table 2. Interventional events in India.

Interventional Event	Time Interval
Lockdown	25 March 2020~31 May 2020
Unlock	8 June 2020~28 February 2021
Vaccine	16 January 2021~31 March 2021

The best ARIMAX and VARMA models under the AICC criterion for the India data are ARIMAX (2, 2, 2) and VARMA (1, 0), respectively. Therefore, we compare three control charts: OECC, RECC based on the ARIMAX (2, 2, 2) model, and RECC based on the VARMA (1, 0) model. In this case, λ in (4) is selected to be 0.2. Several scenarios are considered to compare these three charts. This includes the cases where ARL_0 is 10, 20, 50, 100, 200, and 400.

Tables 3–5 provide UCLs for the different specified ARL_0 values of the three charts, respectively. Figures 1–3 display the results of the three control charts for ARL_0 values of around 50, 100, and 200. Figure 1 shows that OECC is too conservative to detect the out-of-control signal. RECC based on ARIMAX (2, 2, 2) in India is shown in Figure 2, which signals around March 18. Figure 3 reveals that RECC based on VARMA model signals around March 13. In the India case, both RECC methods perform better than OECC along with all specified ARL_0 values. This outcome shows the shortcomings of the regular OECC technique. It lagged in detecting signals and, most importantly, completely missed the warning signs that a serious second wave of the epidemic was coming.

Table 3. The UCLs and signal dates of OECC for India.

Methodology	ARL_0	UCL	The First Signal Date
Observation	9.9	73,785.180	No detection
	19.8	83,356.369	No detection
	49.5	89,830.176	No detection
	99	91,364.915	No detection
	198	92,481.088	No detection
	396	92,564.801	No detection

Table 4. The UCLs and signal dates of ARIMA-RECC for India.

Methodology	ARL_0	UCL	The First Signal Date
ARIMA (2, 2, 2)	9.9	1037.954	14 March 2021
	19.8	1438.835	17 March 2021
	49.5	1819.166	17 March 2021
	99	2412.036	17 March 2021
	198	2591.015	18 March 2021
	396	2702.877	18 March 2021

Table 5. UCLs and signal dates of VARMA-RECC for India.

Methodology	ARL_0	UCL	The First Signal Date
VARMA (1, 0)	9.9	2765.219	11 March 2021
	19.8	3942.714	12 March 2021
	49.5	4832.036	13 March 2021
	99	5304.451	13 March 2021
	198	5540.659	13 March 2021
	396	5776.866	13 March 2021

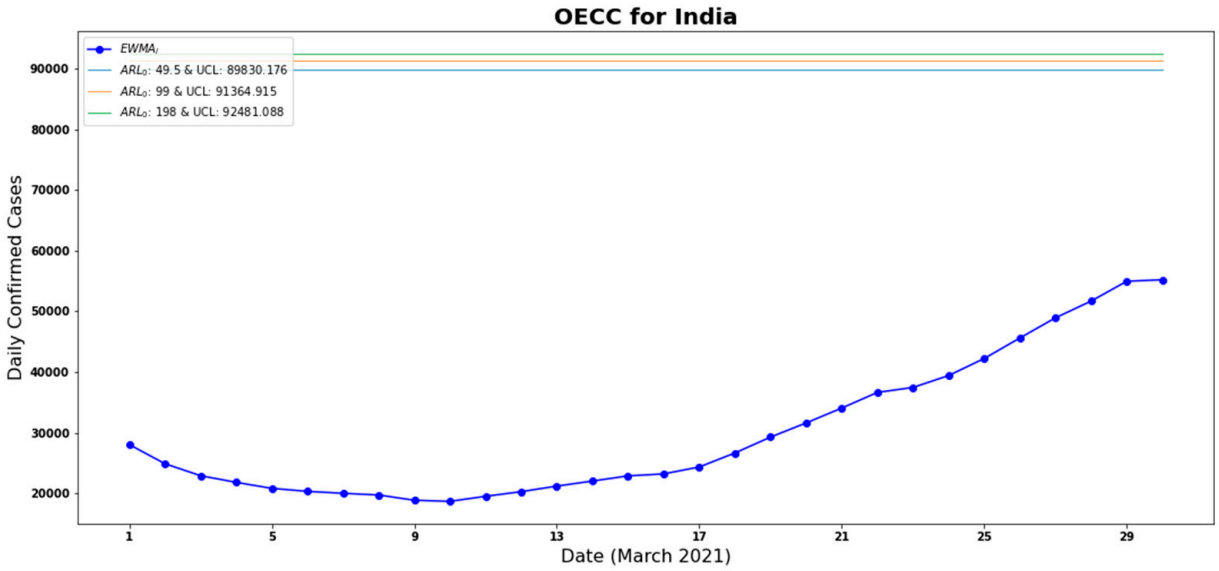


Figure 1. OECC for Indian data.

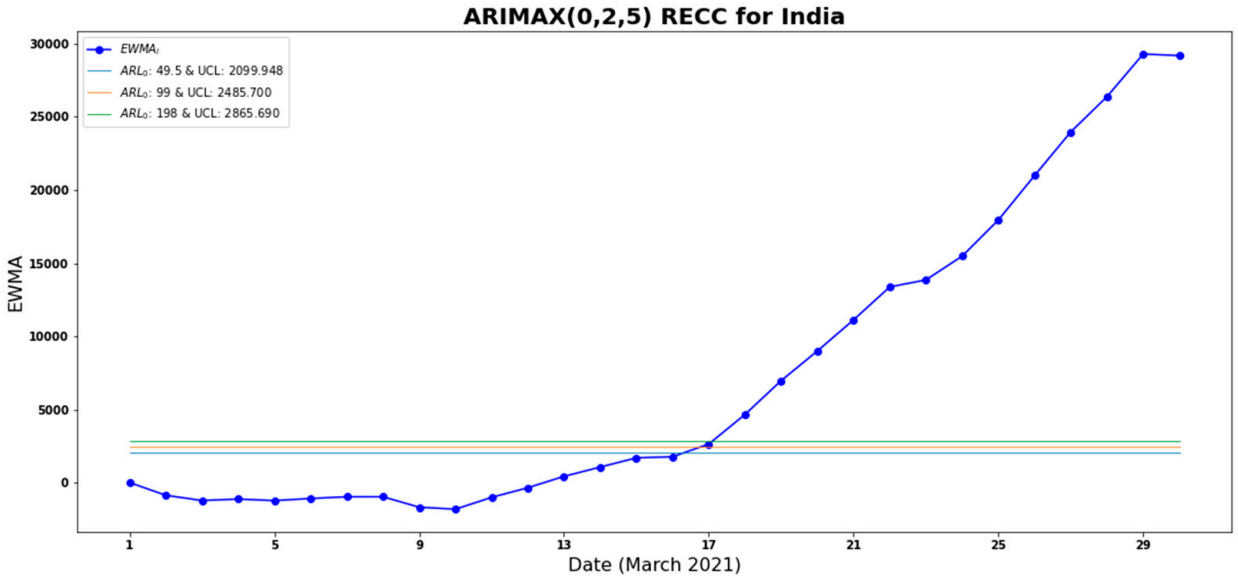


Figure 2. RECC based on the ARIMAX model for Indian data.

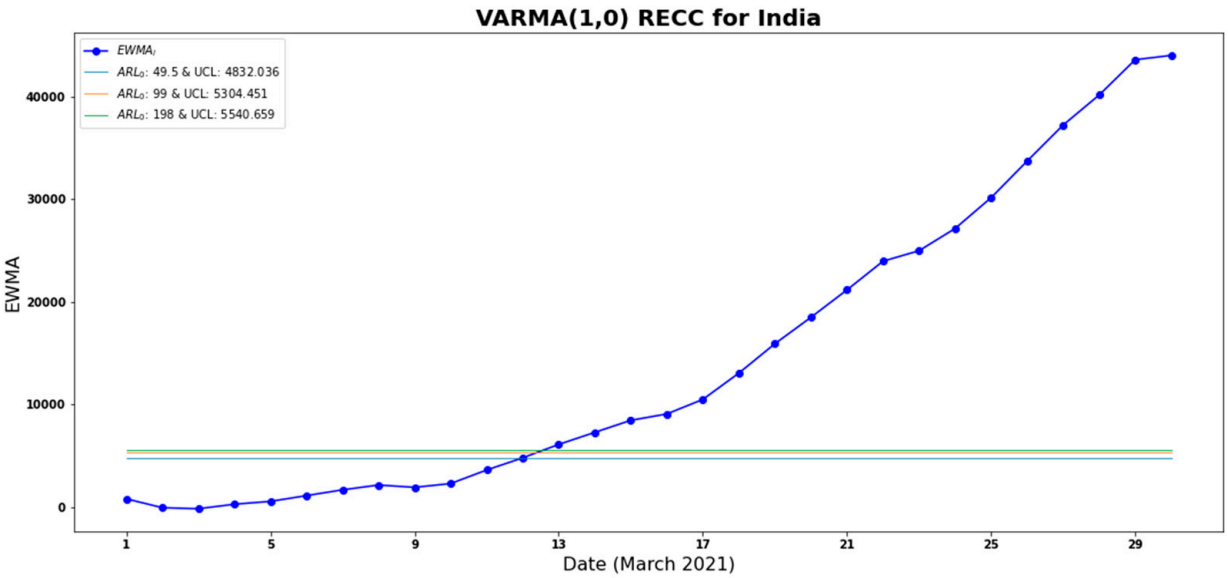


Figure 3. RECC based on the VARMA model for Indian data.

3.2. Malaysia

By analyzing the history and data, several pivotal events were identified in Malaysia as marking the beginning of the second wave of the COVID-19 pandemic in 2022 (Table 1). In particular, on 12 February, the nation crossed a significant threshold, surpassing 3 million total confirmed cases, signifying the escalating severity of the epidemic. Subsequently, on 16 February, Malaysia experienced a concerning surge in new daily cases, reaching a record high of 27,734. This surge intensified further on 23 February, as daily cases soared past 30,000 for the first time, peaking at 31,199. We consider these events to be second waves and evaluate the effectiveness of the proposed method in predicting the second wave.

The number of historical data points used for the Malaysia case was about 760, with the period assessed running from 3 January 2020 to 31 January 2022. The monitoring period was from 1 February 2021 to 2 March 2021. The exogenous variable used for the ARIMAX is "lockdown", with results presented in Table 6. In the VARMA model, the three-dimensional data used for the VARMA model are daily COVID-19 confirmed case numbers, daily COVID-19 testing numbers, and daily new death counts. As in the India case, we monitor the daily COVID-19 confirmed case number, and the other two variables are used to increase the accuracy of the model fitting. The best ARIMA and VARMA models under the AICC criterion for the Malaysia data are ARIMAX (2, 1, 2) and VARMA (1, 0), respectively. The scenarios cases of ARL_0 being 10, 20, 50, 100, 200, and 400 are considered to compare the three control charts.

Table 6. Interventional events in Malaysia.

Interventional Event	Time Interval
Lockdown	18 March 2020~18 February 2021 12 May 2021~28 June 2021

Tables 7–9 provide UCLs for the different specified ARL_0 values of OECCs and RECCs based on ARIMAX and VARMA models, respectively. Figures 4–6 display the results of the three control charts. Similar to the India case, the OECC is too conservative to detect the out-of-control signal (Figure 4). From Figures 5 and 6, it is evident that compared to the original OECC method, RECCs based on the ARIMAX and VARMA models yield better signal detection results for different ARL_0 settings. In these cases, signals are detected

between 2 and 5 February, which are 5 to 12 days earlier than the day detected using the OECC method. The reliability of the earlier detection of signals using RECCs presents a considerable advantage over the conventional OECC method.

Table 7. The UCLs and signal dates of OECCs for Malaysia.

Methodology	ARL_0	UCL	The First Signal Date
Observation	8.539	7012.608	7 February 2022
	50.667	19,026.456	16 February 2022
	152	20,949.153	17 February 2022
	190	21,281.184	17 February 2022
	380	21,478.690	17 February 2022
	760	21,677.600	17 February 2022

Table 8. The UCLs and signal dates of the ARIMAX-RECC model for Malaysia.

Methodology	ARL_0	UCL	The First Signal Date
ARIMA (2, 1, 2)	8.539	138.770	2 February 2022
	50.667	428.418	4 February 2022
	152	677.578	5 February 2022
	190	708.724	5 February 2022
	380	786.586	5 February 2022
	760	864.448	5 February 2022

Table 9. The UCLs and signal dates of the VARMA-RECC model for Malaysia.

Methodology	ARL_0	UCL	The First Signal Date
VARMA (1, 0)	8.539	148.876	4 February 2022
	50.667	342.676	5 February 2022
	152	520.972	5 February 2022
	190	528.724	5 February 2022
	380	559.733	5 February 2022
	760	652.757	5 February 2022

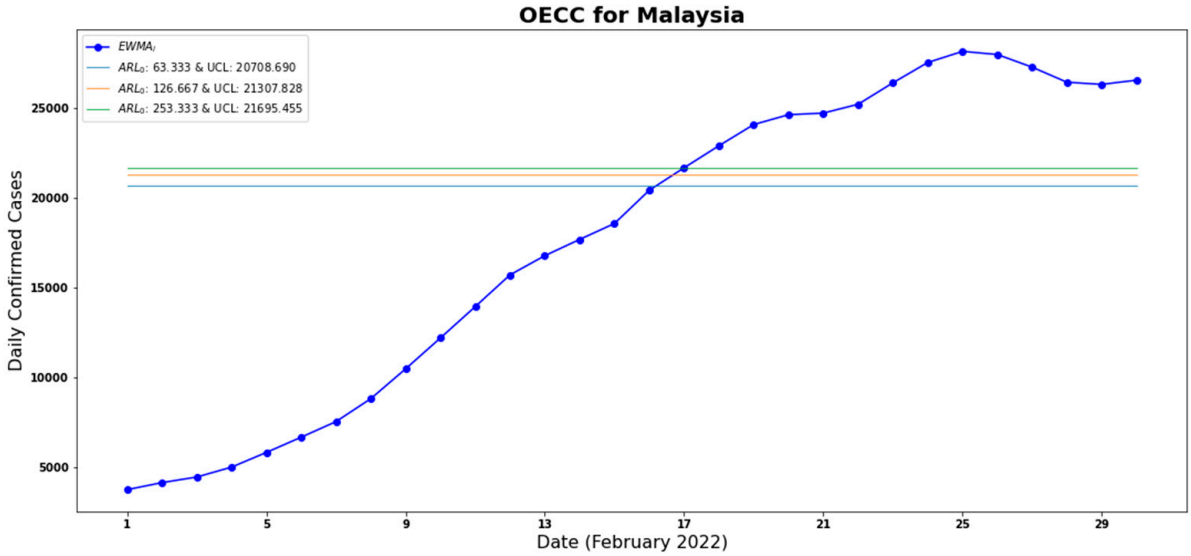


Figure 4. OECC for Malaysia data.

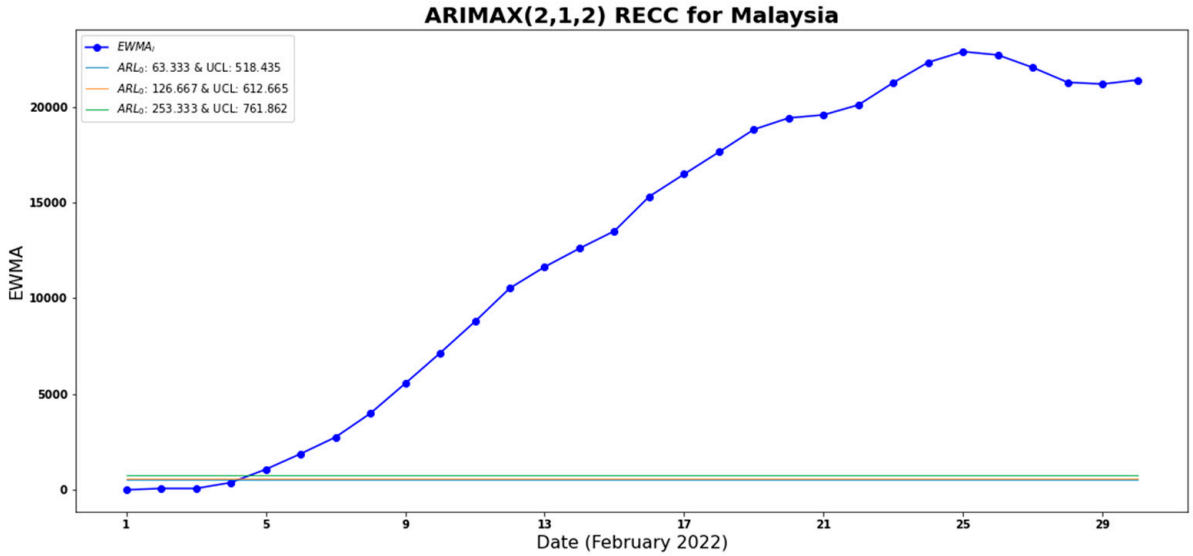


Figure 5. An RECC based on the ARIMA model for Malaysia data.

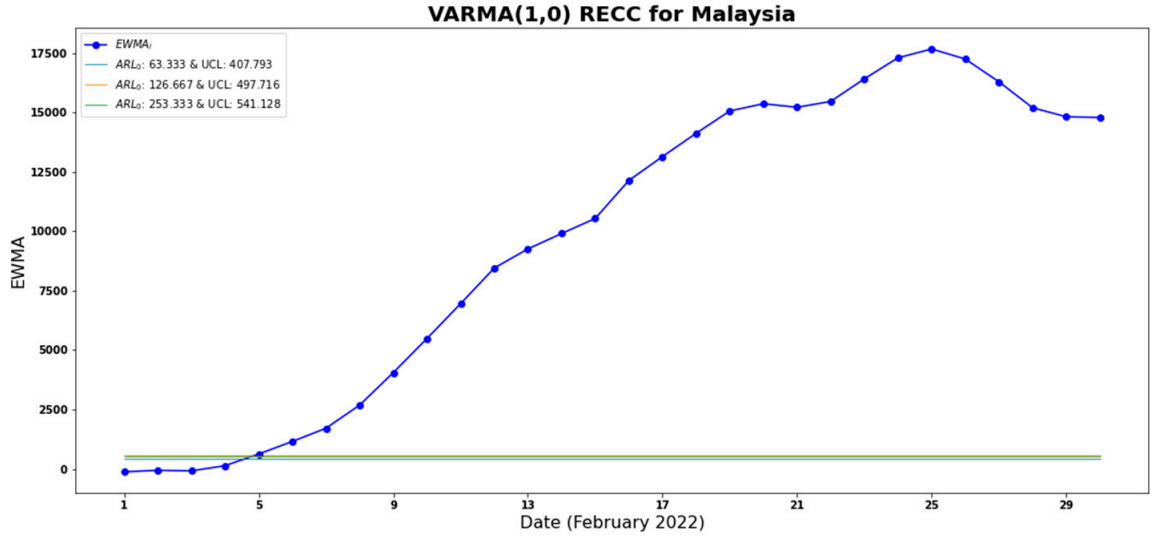


Figure 6. An RECC based on the VARMA model for Malaysia data.

3.3. Thailand

By analyzing historical records and data, several pivotal events in Thailand during the COVID-19 pandemic of 2020 were identified as the beginning of the second wave (Table 1). The traditional Thai New Year (Songkran Festival) is held from 13–14 April. Due to the escalating situation of COVID-19, the local government canceled large celebrations in an effort to reduce the risk of virus transmission during large gatherings. Despite these precautions, confirmed cases increased sharply on April 14, highlighting the challenges of controlling the spread of the virus during the holiday season. Subsequently, on 16 April, the government further escalated domestic epidemic prevention measures, including closing large public places and implementing stricter social distancing protocols. We identify these events as second waves and evaluate how effectively the proposed method predicts them.

The number of historical data points used for the Thailand case is about 454. The period examined ran from 3 January 2020 to 30 March 2021. The best ARIMAX and VARMA models under the AICC criterion for the Thailand data are ARIMAX (1, 1, 2) and VARMA (1, 0), respectively. The exogenous variables used in the ARIMAX model for the Thailand data are the “lockdown”, “bank close”, and “airline shutdown”. These are presented

in Table 10. Three variables, including daily COVID-19 confirmed case numbers, daily COVID-19 testing numbers, and daily new death counts, are considered in the VARMA model. The scenarios include cases of ARL_0 being 10, 20, 50, 100, 200, and 400, and these are considered to compare the three control charts.

Table 10. Interventional events in Thailand.

Interventional Event	Time Interval
Lockdown	22 March 2020~15 June 2020
Bank close	25 March 2020~29 March 2020
Airline shutdown	1 April 2020~30 April 2020

Tables 11–13 provide UCLs for different specified ARL_0 values for the three charts, respectively. Figures 7–9 display the results of the three control charts. Similar to the India and Malaysia cases, RECCs have better performances than OECCs for different ARL_0 settings. In these cases, signals are detected between 5 and 9 April, approximately 2 to 5 days earlier than the OECC method, which detects signals between 7 April and 14 April.

Table 11. The UCLs and signal dates of OECCs for Thailand.

Methodology	ARL_0	UCL	The First Signal Date
Observation	10.558	153.119	7 April 2021
	41.273	447.710	11 April 2021
	90.8	634.655	13 April 2021
	113.5	683.167	14 April 2021
	227	718.008	14 April 2021
	454	733.926	14 April 2021

Table 12. The UCLs and signal dates of the ARIMA-RECC model for Thailand.

Methodology	ARL_0	UCL	The First Signal Date
ARIMA(1, 1, 2)	10.558	7.804	5 April 2021
	41.273	82.588	7 April 2021
	90.8	111.967	8 April 2021
	113.5	113.303	8 April 2021
	227	129.328	8 April 2021
	454	149.359	9 April 2021

Table 13. The UCLs and signal dates of the VARMA-RECC model for Thailand.

Methodology	ARL_0	Modified UCL	First Signal Date
VARMA(1, 0)	10.558	9.091	5 April 2021
	41.273	91.182	8 April 2021
	90.8	101.144	8 April 2021
	113.5	102.568	8 April 2021
	227	121.069	9 April 2021
	454	128.185	9 April 2021

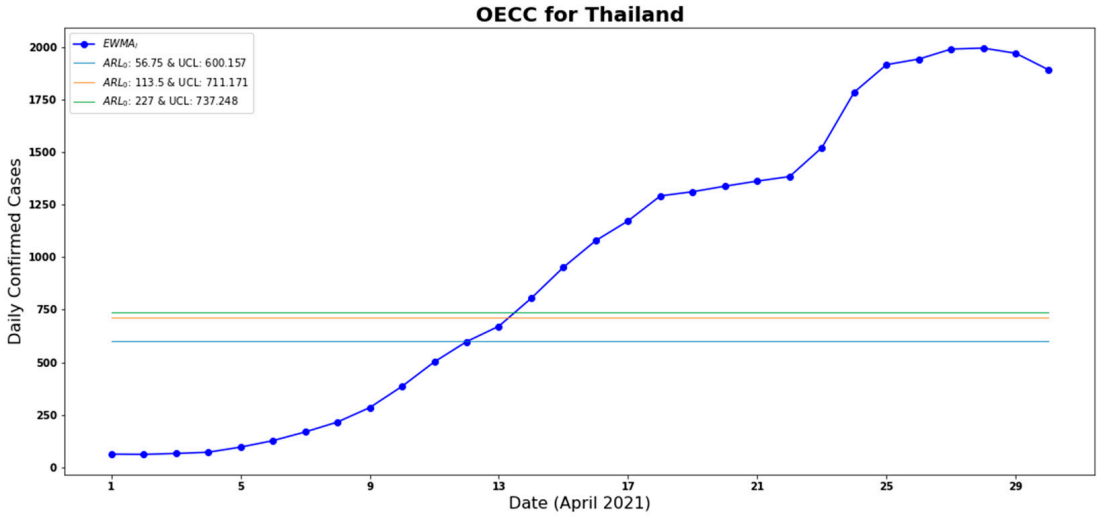


Figure 7. OECCs for Thailand data.

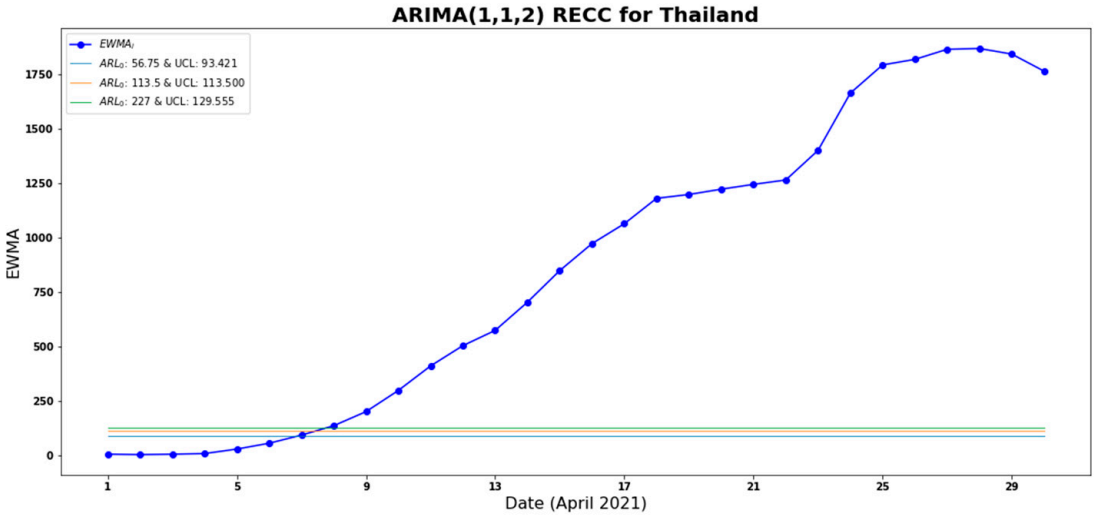


Figure 8. The RECC based on the ARIMAX model for Thailand data.

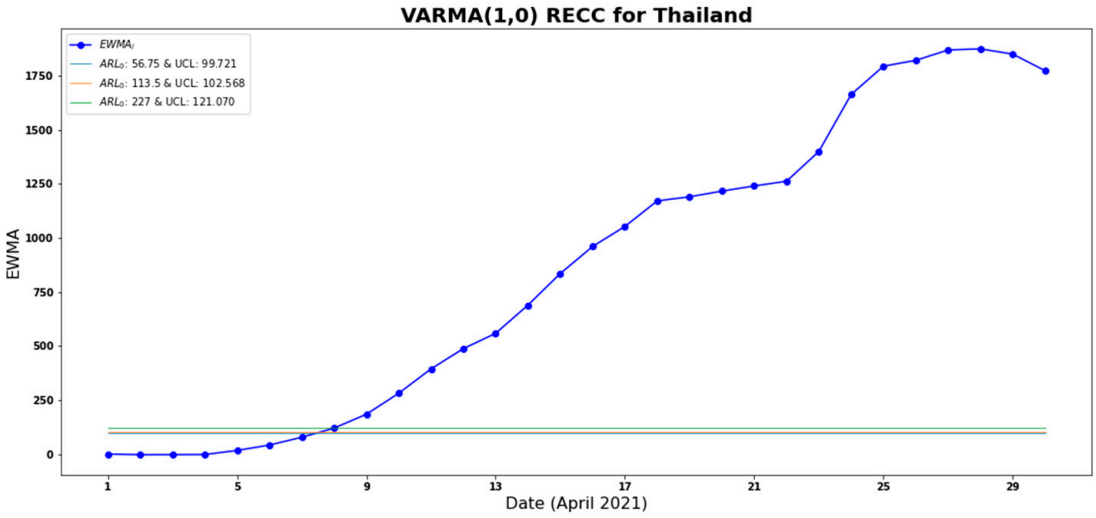


Figure 9. The RECC based on the VARMA model for Thailand data.

4. Discussion

This study uses the past data from three countries to validate the efficiency of the proposed methods for the real-time monitoring of COVID-19 daily confirmed cases. Although the COVID-19 pandemic has subsided, past data and experiences can serve as valuable resources to aid in public health planning for ongoing or future pandemics. Therefore, the success of established prediction methods in analyzing past data may provide useful tools to assist these countries in making accurate decisions during future pandemics.

The technique of combining control charts with time series models has been established and applied in data analysis in other applications [28,29]. This study first applies this method to COVID-19 data prediction. We innovatively applied EWMA control charts to the residuals of time series models to monitor daily confirmed COVID-19 cases. This approach enhances the detection of small shifts in trends, which is critical for early warnings during pandemic outbreaks. In addition to the models used in this study, other models that can handle nonlinear or non-stationary data have the potential to be incorporated with the control chart technique for COVID-19 data prediction. The advanced time series models or machine learning methods such as long short-term memory networks or the Prophet model can be studied further [30,31].

In these proposed methods, there are several parameters for determination. In the time series models, AICC is applied to select the models. In the EWMA chart, the parameter λ is usually selected to be small, such as 0.2, to ensure the current observation has a large weight compared with the past data. In addition to the value of $\lambda = 0.2$ used in the result section, we present two other cases, with values of $\lambda = 0.1$ and 0.3, in this section. Tables 14 and 15 provide the detection days of these two scenarios. The results demonstrate that RECCs can predict the second wave earlier than the OECC regardless of the different λ selected in some ranges.

Table 14. $\lambda = 0.1$ and comparison of corresponding results.

λ	Country	OECC First Signal Date	RECC (ARIMA) First Signal Date	RECC (VARMA) First Signal Date
0.1	India	No detection	18 March 2021	15 March 2021
	Malaysia	22 February 2022	5 February 2022	6 February 2022
	Thailand	15 April 2021	9 April 2021	9 April 2021

Table 15. $\lambda = 0.3$ and comparison of corresponding results.

λ	Country	OECC First Signal Date	RECC (ARIMA) First Signal Date	RECC (VARMA) First Signal Date
0.1	India	No detection	18 March 2021	12 March 2021
	Malaysia	16 February 2022	5 February 2022	5 February 2022
	Thailand	13 April 2021	8 April 2021	8 April 2021

While RECC and OECC can be applied to these three countries, this does not imply that they are universally applicable to all countries. Some countries face different critical situations that may require more advanced techniques for monitoring instead of relying solely on control charts and time series methods. Since the proposed methods are effective in these three countries, we believe they can be applied to many other countries, and expanding the application of these methods to other countries is part of our future work.

5. Conclusions

In this study, control charts integrated with a time series model are used to monitor the daily confirmed cases of COVID-19 data. Control charts are not only useful tools for monitoring the production line in industry but are also used to surveil disease count. Since the number of confirmed cases may depend on the cases from previous days, we incorporate a time series model to capture this dependency. The residuals obtained from time series models are used to construct control charts for monitoring the daily number of confirmed COVID-19 cases. Data from three countries—India, Malaysia, and Thailand—are analyzed to compare the proposed methods with conventional control charts.

The first proposed method uses daily confirmed cases, combined with several significant events during the COVID-19 pandemic, to fit an ARIMAX time series model. The second proposed method incorporates two additional variables—daily COVID-19 testing numbers and daily new death counts—to fit a three-dimensional VARMA time series model to monitor the daily number of confirmed COVID-19 cases. The proposed methods can detect the signal of COVID-19 outbreak earlier than the conventional EWMA control chart. This highlights the potential of the proposed methods for COVID-19 surveillance. Since COVID-19 seriously impacts United Nations Sustainable Development objectives such as health systems and economies, it is important to develop early outbreak detection methodologies to prevent global pandemics. The proposed methods can provide useful strategies to assist in the surveillance of COVID-19 and may reduce the risk of a major pandemic.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/math13010115/s1. Table S1: Data of three countries.

Author Contributions: Conceptualization, H.W.; methodology, C.-R.H. and H.W.; formal analysis, C.-R.H.; data curation, C.-R.H.; writing, H.W. and C.-R.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Ministry of Science and Technology 111–2118-M−A49−001−MY2, Taiwan.

Data Availability Statement: Data of three countries are provided in the Supplementary File Table S1.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Chen, Y.H.; Wang, H. Exploring Diversity of COVID-19 Based on Substitution Distance. *Infect. Drug Resist.* **2020**, *13*, 3887–3894. [CrossRef]
2. Ciotti, M.; Angeletti, S.; Minieri, M.; Giovannetti, M.; Benvenuto, D.; Pascarella, S.; Sagnelli, C.; Bianchi, M.; Bernardini, S.; Ciccozzi, M. COVID-19 outbreak: An overview. *Chemotherapy* **2020**, *64*, 215–223. [CrossRef] [PubMed]
3. Padhan, R.; Prabheesh, K. The economics of COVID-19 pandemic: A survey. *Econ. Anal. Policy* **2021**, *70*, 220–237. [CrossRef] [PubMed]
4. Rajgor, D.D.; Lee, M.H.; Archuleta, S.; Bagdasarian, N.; Quek, S.C. The many estimates of the COVID-19 case fatality rate. *Lancet Infect. Dis.* **2020**, *20*, 776–777. [CrossRef] [PubMed]
5. Noh, J.; Danuser, G. Estimation of the fraction of COVID-19 infected people in US states and countries worldwide. *PLoS ONE* **2021**, *16*, e0246772. [CrossRef] [PubMed]
6. Baud, D.; Qi, X.; Nielsen-Saines, K.; Musso, D.; Pomar, L.; Favre, G. Real estimates of mortality following COVID-19 infection. *Lancet Infect. Dis.* **2020**, *20*, 773. [CrossRef] [PubMed]
7. Kumar, R.; Jain, A.; Tripathi, A.K.; Tyagi, S. COVID-19 outbreak: An epidemic analysis using time series prediction model. In Proceedings of the 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 28–29 January 2021; pp. 1090–1094.
8. Ballı, S. Data analysis of COVID-19 pandemic and short-term cumulative case forecasting using machine learning time series methods. *Chaos Solitons Fractals* **2021**, *142*, 110512. [CrossRef]
9. Tomov, L.; Chervenkov, L.; Miteva, D.G.; Batselova, H.; Velikova, T. Applications of time series analysis in epidemiology: Literature review and our experience during COVID-19 pandemic. *World J. Clin. Cases* **2023**, *11*, 6974. [CrossRef] [PubMed]

10. Ceylan, Z. Estimation of COVID-19 prevalence in Italy, Spain, and France. *Sci. Total Environ.* **2020**, *729*, 138817. [CrossRef] [PubMed]
11. Singh, V.; Poonia, R.C.; Kumar, S.; Dass, P.; Agarwal, P.; Bhatnagar, V.; Raja, L. Prediction of COVID-19 corona virus pandemic based on time series data using support vector machine. *J. Discret. Math. Sci. Cryptogr.* **2020**, *23*, 1583–1597. [CrossRef]
12. Perla, R.J.; Provost, S.M.; Parry, G.J.; Little, K.; Provost, L.P. Understanding variation in reported COVID-19 deaths with a novel Shewhart chart application. *Int. J. Qual. Health Care* **2021**, *33*, mzaa069. [CrossRef] [PubMed]
13. Mandal, S.; Roychowdhury, T.; Bhattacharya, A. Pattern of genomic variation in SARS-CoV-2 (COVID-19) suggests restricted nonrandom changes: Analysis using Shewhart control charts. *J. Biosci.* **2021**, *46*, 11. [CrossRef] [PubMed]
14. Wang, H. Tolerance limits for mixture-of-normal distributions with application to COVID-19 data. *Wiley Interdiscip. Rev. Comput. Stat.* **2023**, *15*, e1611. [CrossRef]
15. Wang, H.; Huwang, L.; Yu, J.H. Multivariate control charts based on the James–Stein estimator. *Eur. J. Oper. Res.* **2015**, *246*, 119–127. [CrossRef]
16. Wang, H. Comparison of p control charts for low defective rate. *Comput. Stat. Data Anal.* **2009**, *53*, 4210–4220. [CrossRef]
17. Mahmood, Y.; Ishtiaq, S.; Khoo, M.B.; Teh, S.Y.; Khan, H. Monitoring of three-phase variations in the mortality of COVID-19 pandemic using control charts: Where does Pakistan stand? *Int. J. Qual. Health Care* **2021**, *33*, mzab062. [CrossRef] [PubMed]
18. Dawoud, I. Modeling Palestinian COVID-19 cumulative confirmed cases: A comparative study. *Infect. Dis. Model.* **2020**, *5*, 748–754. [CrossRef] [PubMed]
19. Hamilton, J.D. *Time Series Analysis*; Princeton University Press: Princeton, NJ, USA, 2020.
20. Brockwell, P.J.; Davis, R.A. *Time Series: Theory and Methods*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1991.
21. Brillinger, D.R. *Time Series: Data Analysis and Theory*; SIAM: Philadelphia, PA, USA, 2001.
22. Islam, F.; Imteaz, M.A. Use of teleconnections to predict Western Australian seasonal rainfall using ARIMAX model. *Hydrology* **2020**, *7*, 52. [CrossRef]
23. Portet, S. A primer on model selection using the Akaike Information Criterion. *Infect. Dis. Model.* **2020**, *5*, 111–128. [CrossRef]
24. Anderson, D.; Burnham, K. *Model Selection and Multi-Model Inference*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2004; Volume 63, p. 10.
25. Montgomery, D.C. *Statistical Quality Control*; Wiley New York: New York, NY, USA, 2009; Volume 7.
26. Yupaporn, A.; Rapin, S. EWMA control chart based on its first hitting time and coronavirus alert levels for monitoring symmetric COVID-19 cases. *Asian Pac. J. Trop. Med.* **2021**, *14*, 364–374. [CrossRef]
27. Waqas, M.; Xu, S.H.; Aslam, M.U.; Hussain, S.; Shahzad, K.; Masengo, G. Global contribution of statistical control charts to epidemiology monitoring: A 23-year analysis with optimized EWMA real-life application on COVID-19. *Medicine* **2024**, *103*, e38766. [CrossRef]
28. Phanthuna, P.; Areepong, Y. Detection sensitivity of a modified EWMA control chart with a time series model with fractionality and integration. *Emerg. Sci. J.* **2022**, *6*, 1134–1152. [CrossRef]
29. Bisgaard, S.; Kulahci, M. Quality quandaries: Using a time series model for process adjustment and control. *Qual. Eng.* **2007**, *20*, 134–141. [CrossRef]
30. Wen, X.; Li, W. Time series prediction based on LSTM-attention-LSTM model. *IEEE Access* **2023**, *11*, 48322–48331. [CrossRef]
31. Chaturvedi, S.; Rajasekar, E.; Natarajan, S.; McCullen, N. A comparative assessment of SARIMA, LSTM RNN and Fb Prophet models to forecast total and peak monthly energy demand for India. *Energy Policy* **2022**, *168*, 113097. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.