Hindawi

Research Article

Information Approach for Change Point Analysis of EGGAPE Distribution and Application to COVID-19 Data

Mutua Kilai [1], **Gichuhi A. Waititu,**[2] **Wanjoya A. Kibira,**[2] **Ramy Aldallal,**[3] **M. E. Bakr** [4]
Yusra A. Tashkandy,[4] **and Fathy H. Riad**[5,6]

[1]*Department of Mathematics, Pan African Insitute of Basic Science, Technology and Innovation, Nairobi, Kenya*
[2]*Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya*
[3]*Department of Accounting, College of Business Administration in Hawtat Bani Tamim, Prince Sattam bin Abdulaziz University, Hawtat Bani Tamim, Saudi Arabia*
[4]*Department of Statistics and Operations Research, Faculty of Science, King Saud University, P. O. Box 2455, Riyadh 11451, Saudi Arabia*
[5]*Mathematics Department, College of Science, Jouf University, P. O. Box 2014, Sakaka, Saudi Arabia*
[6]*Department of Mathematics, Faculty of Science, Minia University, Minia 61519, Egypt*

Correspondence should be addressed to Mutua Kilai; kilaimutua@gmail.com

The exponentiated generalized Gull alpha power exponential distribution is an extension of the exponential distribution that can model data characterized by various shapes of the hazard function. However, change point problem has not been studied for this distribution. In this study, the change point detection of the parameters of the exponentiated generalized Gull alpha power exponential distribution is studied using the modified information criterion. In addition, the binary segmentation procedure is used to identify multiple change point locations. The assumption is that all the parameters of the EGGAPE distributions are considered changeable. Simulation study is conducted to illustrate the power of the modified information criterion in detecting change point in the parameters with different sample sizes. Three applications related to COVID-19 data are used to demonstrate the applicability of the MIC in detecting change point in real life scenario.

1. Introduction

The concept of change point is important in statistical analysis since it helps to identify the points at which a time series' distribution changes. Change point Analysis is of great interest in real-life phenomena such as in health science, finance, and survival analysis. The contents of change point inference include the aspect of determining the existence of a change point and then estimating the number and the positions of the change points. Since the inception of the concept of change point analysis, many studies have been conducted. Sen and Srivastava [1, 2] studied and came up with a statistic for detecting change in the mean of variables that were characterized by a normal distribution and derived the asymptotic, exact distribution. Change point in a

binomial probability model, the power of likelihood ratio, and cumulative sum tests were investigated in [3]. In the context of distributions, Ngunkeng and Ning [4] did a study on change point problem for generalized lambda distribution. For an exponential distribution characterized by repeated values, change point identification was carried out in [5]. Hassan et al. [6] obtained weighted power Lomax distribution and its length biased version. Shrahili et al. [7] discussed the alpha power moment exponential model with applications to biomedical science. Chen and Arjun [8] did an extensive study on the statistical change point in the parametric context and gave applications related to the fields of finance, medicine, and genetics. Change point detection utilizing the information approach considering regular models was explored in [9]. Arellano-Valle et al. [10]

considered the problem of change point for the Skew normal distribution using the Bayesian approach. Alghamdi et al. [11] studied the Rayleigh Lomax distribution and used the information approach to identify potential change points in the parameters. Almetwally [12] did an extensive study incorporating odd Weibull inverse Topp–Leone distribution with applications to COVID-19 data. Recently, Ratnasingam [13] did an extensive study incorporating modified information approach and confidence distribution skew normal distribution.

In the area of statistics, the unknown time point when observations follow different distributions before and after the point is described as change point. Following the definition of change point, the description follows. Let $Z_1, Z_2, \ldots, Z_n$ be a series of independent random variables with CDF given as $G_1, G_2, \ldots, G_n$, respectively. The change point problem entails testing the null hypothesis as follows:

$$H_0: G_1 = G_2 = \cdots = G_n. \tag{1}$$

Against the alternative,

$$H_a: G_1 = \cdots = G_{k_1} \neq G_{k_1+1} = \cdots = G_{k_q} \neq G_{k_q+1} = \cdots = G_n, \tag{2}$$

where $1 < k_2 < k_2 < \cdots < k_q < n$ denotes the change points number and $k_1, k_2, \ldots, k_q$ are the unknown locations of the change points that are to be estimated. Of importance, if $G_1, G_2, \ldots, G_n$ are from the same parametric family, the change point problem turns out to be a test of the null hypothesis of the parameters of the population $\phi_i, i = 1, 2, \ldots, n$, stipulated as

$$H_0: \phi_1 = \phi_2 = \cdots = \phi_n = \phi, \tag{3}$$

versus the alternative,

$$H_1: \phi_1 = \cdots = \phi_{k_1} \neq \phi_{k_1+1} = \cdots = \phi_{k_2} \neq \phi_{k_2+1}$$
$$= \cdots = \phi_{k_q-1} \neq \phi_{k_q} = \cdots = \phi_n. \tag{4}$$

The Exponentiated Generalized Gull Alpha Power Exponential (EGGAPE) distribution is a recently developed distribution in [14]. It is flexible enough and can take various shapes of the hazard functions depending on the values of the shape parameters. The probability distribution of the EGGAPE distribution is given by

$$g(z) = \frac{ab\alpha^{e^{-\lambda z}}\lambda e^{-\lambda z} - ab\alpha^{e^{-\lambda z}}\log(\alpha)\lambda\left(1 - e^{-\lambda z}\right)e^{-\lambda z}}{\left(1 - \alpha\left(1 - e^{-\lambda z}\right)/\alpha^{1-e^{-\lambda z}}\right)^{1-a}\left[1 - \left(1 - \alpha\left(1 - e^{-\lambda z}\right)/\alpha^{1-e^{-\lambda z}}\right)^a\right]^{1-b}}, \tag{5}$$

where $\alpha, a, b > 0$ are the shape parameter and λ is the scale parameter. The case, where $a = b = \alpha = 1$, is the exponential distribution, the case, where $\alpha = 1$ and $a = 1$, is the exponentiated exponential distribution, the case, where $\alpha = 1$, is the exponentiated generalized exponential distribution. Several authors have investigated the change point problem for several distributions. ElSherpienyAlmetwally [15] introduced exponentiated generalized alpha power family. Jandhyala et al. [16] came up with change-point

methodology that was used to identify changes in the parameters of the two-parameter Weibull distribution. The statistic they developed was the likelihood ratio test that was used to detected unknown changes in parameters, and the change points were located. Almongy et al. [17] discussed likelihood function for multicomponent stress-strength model under power Lomax distribution. Hafez et al. [18] studied likelihood of single and multiple ramp progressive stress with binomial removal. The application of the model was on temperature data.

Jarušková [19] did a study to test the presence of change point using the log-likelihood statistic in a three-parameter Weibull distribution. Ratnasingam[20] proposed a procedure that was built on the MIC and the confidence distribution in a three-parameter Weibull distribution for detecting and estimating changes. To identify and find changes in the parameters of a four-parameter EGGAPE distribution concurrently, we present a methodology based on the information approach, specifically modified information approach, and Schwarz information approach. The proposed method can be applied to a variety of parametric distributions as long as the regularity and Wald requirements are met.

The following is how the rest of the study is organized. In Section 2, we look at approaches based on the MIC and SIC for detecting simultaneous changes in all parameters. In Section 3, simulations for the scenarios will be run with a variety of parameter and sample size variations in order to examine the test's power. Section 4 shows how the algorithm was applied to three COVID-19 datasets to demonstrate change point detection. The results and areas for additional research are presented in Section 5.

2. Methodology

2.1. Information Approach. This section presents the methodology applied to detect the possible change points. The modified information approach and the Schwarz information criterion are discussed. Change point problem generally usually involves the estimation of parameters and testing of hypothesis. To be more specific, the null hypothesis tested is that there is no change point against that there exists at least one change point which is the alternative hypothesis. The use of model selection criteria is one of the most popular methods for change point detection. The Schwarz information criterion (SIC) was developed in [21]. As pointed out in [9], the SIC technique does not take into account the model's complexity, which might lead to redundancy in the parameter space. To address this shortcoming, Chen et al. [9] came up with the MIC technique that adjusted SIC penalty's term to reflect the contributions of the change points' locations to model complexity. Let $y_1, y_2, \ldots, y_n$ be a sample chosen at random from a density function. The following is the SIC criterion:

$$\text{SIC}(n) = -2\log L\left(\widehat{\Phi_q}\right) + k\log(n), \tag{6}$$

where $L(\widehat{\Phi_k})$ is defined as the likelihood function of the model, n is the sample size, and q is the number of

parameters in the model. In this, we denote Φ_B and Φ_A to be the parameters before and after the change point. The symbol k denotes the unknown change point location. When at least one change point is present, the SIC is as follows:

$$\text{SIC}(k) = -2l\left(\widehat{\Phi_B}(k), \widehat{\Phi_A}(k), k\right) + \{2q + 1\}\log(n), \qquad (7)$$

where $1 \leq k < n$. Equation (7) does not treat the change position as a parameter, which could result in redundancy in the parameter space if the change happens near the end or beginning of the data. The MIC under the null hypothesis is given as

$$\text{MIC}(n) = -2\ln L(\widehat{\Phi}) + \dim(\widehat{\Phi})\log(n), \qquad (8)$$

where $\widehat{\Phi}$ maximizes the log likelihood $\log L(\widehat{\Phi})$. The MIC criterion under the alternative hypothesis is defined as follows:

$$\text{MIC}(k) = -2\ln L(\widehat{\Phi}) + \left[2\dim(\widehat{\Phi}) + \left(\frac{2k}{n} - 1\right)^2\right]\log(n) \, 1 \leq k < n, \qquad (9)$$

where $1 \leq k < n$.

If

$$\text{MIC}(n) > \min_{1 \leq k < n} \text{MIC}(k), \qquad (10)$$

choose the model that has a change point, and the location of the specific change point is estimated by $\widehat{\kappa}$ such that

$$\text{MIC}(\widehat{\kappa}) = \min_{1 \leq k < n} \text{MIC}(k). \qquad (11)$$

The MIC test statistic, which is used to determine the statistical significance of a change point, is defined as follows:

$$Q_n = \text{MIC}(n) - \min_{1 \leq k < n} \text{MIC}(k) + \dim(\Phi)\log(n). \qquad (12)$$

Chen et al. [9] showed that, as $n \longrightarrow \infty$,

$$Q_n \longrightarrow \chi_d^2, \qquad (13)$$

in distribution under the null hypothesis. The SIC test statistic is given as

$$P_n = \text{SIC}(n) - \min_{1 \leq k < n} \text{SIC}(k) + \dim(\phi)\log(n). \qquad (14)$$

The asymptotic distribution of the statistic in (14) is the type I extreme value distribution.

2.2. MIC and SIC Detection Approach for EGGAPE Distribution.

In this section, the study focuses on change point problem using SIC and MIC approaches to detect changes in parameters of the EGGAPE distribution defined in Equation (1). Let $Z_1, Z_2, \ldots, Z_n$ be a sequence of independently random variables from the EGGAPE distribution with scale parameter λ and shape parameters α, a, and b. The null hypothesis is

$$\begin{aligned} H_0 : \lambda_1 &= \lambda_2 = \cdots = \lambda_n = \lambda \\ \alpha_1 &= \alpha_2 = \cdots = \alpha_n = \alpha \\ a_1 &= a_2 = \cdots = a_n = a \\ b_1 &= b_2 = \cdots = b_n = b, \end{aligned} \qquad (15)$$

versus

$$\begin{aligned} H_1 : \lambda_1 &= \cdots \lambda_k = \lambda_k^* \neq \lambda_{k+1} = \cdots = \lambda_n = \lambda^{**} \\ \alpha_1 &= \cdots \alpha_k = \alpha_k^* \neq \alpha_{k+1} = \cdots = \alpha_n = \alpha^{**} \\ a_1 &= \cdots a_k = a_k^* \neq a_{k+1} = \cdots = a_n = a^{**} \\ b_1 &= \cdots b_k = b_k^* \neq b_{k+1} = \cdots = b_n = b^{**}, \end{aligned} \qquad (16)$$

where $1 < k < n$ denotes the unknown location to be estimated.

For H_0, the SIC and MIC are defined as

$$\text{SIC}(n) = \text{MIC}(n) = -2\sum_{i=1}^{n} \log\left(f\left(z_i, \widehat{\lambda}, \widehat{\alpha}, \widehat{a}, \widehat{b}\right)\right) + 4\log(n), \qquad (17)$$

where $\widehat{\lambda}, \widehat{\alpha}, \widehat{a}, \widehat{b}$ are MLEs of scale parameter λ and shape parameters α, a, b, respectively, fitted to whole dataset.

The log likelihood function under H_0 is

$$l = 2n\log(ab\lambda) + \log(\log(\alpha)) - 2\lambda\sum_{i=0}^{n} z_i + \sum_{i=1}^{n}\log\left(1 - e^{-\lambda z_i}\right) + 2(a-1)$$

$$\sum_{i=1}^{n}\log\left(1 - \frac{\alpha\left(1 - e^{-\lambda z_i}\right)}{\alpha^{1-e^{-\lambda z_i}}}\right) + 2(b-1)\sum_{i=1}^{n}\log\left(1 - \left(\frac{\alpha\left(1 - e^{-\lambda z_i}\right)}{\alpha^{1-e^{-\lambda z_i}}}\right)^a\right). \qquad (18)$$

To obtain the MLEs of a, b, α, λ, then we let

$$w = 2(a-1)\sum_{i=1}^{n}\log\left(1 - \frac{\alpha\left(1 - e^{-\lambda z_i}\right)}{\alpha^{1-e^{-\lambda z_i}}}\right) \qquad (19)$$

and

$$g = 2(b-1)\sum_{i=1}^{n}\log\left(1 - \left(\frac{\alpha\left(1 - e^{-\lambda z_i}\right)}{\alpha^{1-e^{-\lambda z_i}}}\right)^a\right) \qquad (20)$$

so that $w_a' = \partial y/\partial a, w_\alpha' \partial y/\partial \alpha =, w_\lambda' = \partial y/\partial \lambda, m_a' = \partial m/\partial a, m_b' = \partial m/\partial b, m_\alpha' = \partial m/\partial \alpha$, and $m_\lambda' = \partial m/\partial \lambda$:

$$g_b' = 2 \sum_{i=1}^{n} \log\left(1 - \left(\frac{\alpha\left(1 - e^{-\lambda z_i}\right)}{\alpha^{1-e^{-\lambda z_i}}}\right)^a\right),$$

$$g_a' = 2 \sum_{i=1}^{n} \log\left(1 - \frac{\alpha\left(1 - e^{-\lambda z_i}\right)}{\alpha^{1-e^{-\lambda z_i}}}\right),$$

$$g_a' = \sum_{i=1}^{n} -\frac{\left(\alpha\left(1 - e^{-\lambda z_i}\right)/\alpha^{1-e^{-\lambda z_i}}\right)^a \log\left(\alpha\left(1 - e^{-\lambda z_i}\right)/\alpha^{1-e^{-\lambda z_i}}\right)}{\left(1 - \alpha\left(1 - e^{-\lambda z_1}\right)/\alpha^{1-e^{-\lambda z_i}}\right)^a},$$

$$g_\alpha' = 2(a-1) \sum_{i=1}^{n} \frac{\left(1 - e^{-\lambda z_i}\right)\left(1 - \alpha\left(1 - e^{-\lambda z_i}\right)\alpha^{-1}\right)}{\left(1 - \alpha\left(1 - e^{-\lambda z_i}\right)/\alpha^{1-e^{-\lambda z_i}}\right)\alpha^{1-e^{-\lambda z_i}}},$$

$$g_\lambda' = 2(a-1) \sum_{i=1}^{n} \frac{\left(\alpha z_i e^{-\lambda z_i}\right)\left(1 - \log(\alpha)\left(1 - e^{-\lambda z_i}\right)\right.}{\left(1 - \alpha\left(1 - e^{-\lambda z_i}\right)/\alpha^{1-e^{-\lambda z_i}}\right)\alpha^{1-e^{-\lambda z_i}}},$$

$$g_\alpha' = 2(b-1)a\alpha \sum_{i=1}^{n} -\frac{\left(\alpha\left(1 - e^{-\lambda z_i}\right)/\alpha^{1-e^{-\lambda z_i}}\right)^{a-1} \times 1 - e^{-\lambda z_i}\left(1 - \left(1 - e^{-\lambda z_i}\right)\alpha^{e^{-\lambda z_i - 2}}\right)}{\left(1 - \alpha\left(1 - e^{-\lambda z_i}\right)/\alpha^{1-e^{-\lambda z_i}}\right)^a},$$

$$g_\lambda' = 2(b-1)a \sum_{i=1}^{n} -\frac{\left(\alpha\left(1 - e^{-\lambda z_i}\right)/\alpha^{1-e^{-\lambda z_i}}\right)^{a-1} \times \alpha z_i e^{-\lambda z_i}\left(1 - \alpha\left(1 - e^{-\lambda z_i}\log(\alpha)\right)\right)\alpha^{e^{-\lambda z_i - 1}}}{\left(1 - \alpha\left(1 - e^{-\lambda z_i}\right)/\alpha^{1-e^{-\lambda z_i}}\right)^a}.$$

$$(21)$$

Partial derivatives of the log-likelihood function with respect to a, b, α, λ and equating them to zero are given:

$$\frac{\partial l}{\partial \alpha} = \frac{1}{\alpha \log \alpha} + y_\alpha' + m_\alpha', \qquad (22)$$

$$\frac{\partial l}{\partial a} = \frac{2n}{a} + y_a' + m_a', \qquad (23)$$

$$\frac{\partial l}{\partial b} = \frac{2n}{b} + y_b' + m_b', \qquad (24)$$

$$\frac{\partial l}{\partial \lambda} = \frac{2n}{\lambda} - 2 \sum_{i=1}^{n} z_i + y_\lambda' + m_\lambda' + \sum_{i}^{n} \frac{z_i e^{-z_i}}{1 - e^{-\lambda z_i}}. \qquad (25)$$

The parameter estimates for a, b, α, λ are obtained by equating equations (22)–(25) to zero and solving the system of nonlinear equations.

Under H_1, the SIC and MIC are defined, respectively, as

$$\mathrm{SIC}(k) = -2 \sum_{i=1}^{k} \log\left(f\left(z_i, \widehat{\lambda}^*, \widehat{\alpha}^*, \widehat{a}^*, \widehat{b}^*\right)\right) - 2 \sum_{i=k+1}^{n} \log\left(f\left(z_i, \widehat{\lambda}^{**}, \widehat{\alpha}^{**}, \widehat{a}^{**}, \widehat{b}^{**}\right)\right) + 8 \log(n),$$

$$\mathrm{MIC}(k) = -2 \sum_{i=1}^{k} \log\left(f\left(z_i, \widehat{\lambda}^*, \widehat{\alpha}^*, \widehat{a}^*, \widehat{b}^*\right)\right) - 2 \sum_{i=k+1}^{n} \log\left(f\left(z_i, \widehat{\lambda}^{**}, \widehat{\alpha}^{**}, \widehat{a}^{**}, \widehat{b}^{**}\right)\right) + \left[8 + \left(\frac{2k}{n} - 1\right)^2\right] \log(n),$$

$$(26)$$

where $\widehat{\alpha}^*, \widehat{\lambda}^*, \widehat{a}^*$, and $\widehat{b}^*$ are the MLEs of α, λ, a, and b, respectively, fitted to the first segment of data and $\widehat{\alpha}^{**}, \widehat{\lambda}^{**}, \widehat{a}^{**}$, and $\widehat{b}^{**}$ are the MLEs of α, λ, a, and b, respectively, fitted to the second segment of data.

3. Simulation

In this section, simulations are carried out to assess the test's power in two scenarios: when there is a change point and when there is not. First, we conduct the simulation for SIC and MIC when there is change point.

3.1. Simulation Study: When There Is Change Point. In this section, change point problem of the scale and the shape parameters of the EGGAPE distribution were studied. To be able to calculate the statistic $\mathrm{SIC}(n), \mathrm{MIC}(n), \mathrm{SIC}(k)$, and

MIC (k), the bbmle package developed in [22] was used to fit a dataset with EGGAPE distribution since the first derivatives of the $\log(f(z_i; \widehat{a}, \widehat{b}, \widehat{\alpha}, \widehat{\lambda}))$, $\log(f(z_i; \widehat{a^*}, \widehat{b^*}, \widehat{\alpha^*}, \widehat{\lambda^*}))$, and $\log(\widehat{a^{**}}, \widehat{b^{**}}, \widehat{\alpha^{**}}, \widehat{\lambda^{**}})$

We conduct simulations 1000 times under the EGGAPE (a, b, α, λ) with different values of the shape parameter a, b, α and the scale parameter λ. The test statistic T_n and S_n are calculated and compared to the critical values corresponding to the significant level 0.05.

After rejecting, the null hypothesis, the powers of SIC and MIC with different sample sizes $n = 100, 200, 300, 400$, and different change locations are shown in Tables 1–4. The EGGAPE parameters are changing as $a^* = 0.9, a^{**} = 1.2$, $b^* = 0.8, b^{**} = 0.9$, $\alpha^* = 0.2, \alpha^{**} = 0.4$, and $\lambda^* = 0.5$, $\lambda^{**} = 0.8$.

The purpose of the simulation power test is in order to verify the accuracy of detecting the change point at different locations. As indicated in Tables 1–4, the power increases as the change point location moves to the middle of the data. It can be observed that the MIC has high powers to detect the change point compared to SIC.

Compared to the power of the traditional SIC, MIC has a higher value for the EGGAPE distribution, as shown in Figures 1. It is clear that MIC has a higher power when the change point location k is in the middle of the dataset. This is because of the penalty term in MIC $(2k/n - 1)^2$ which is different from the traditional value of 1 in SIC.

If the location of the change point is found at the start of the data and the end of the data, as $k \longrightarrow 1$ or $k \longrightarrow n$ and $n \longrightarrow \infty$,

$$\left(\frac{2k}{n} - 1\right)^2 \longrightarrow 1. \tag{27}$$

This is very close to SIC. However, when the change point is in the middle of the dataset, as $k \longrightarrow n/2$,

$$\left(\frac{2k}{2} - 1\right)^2 \longrightarrow 0. \tag{28}$$

Then, this quadratic term $(2k/2 - 1)^2$ will be canceled. When the change point is exactly the middle term $(2k/2 - 1)^2 = 0$ and the penalty term of MIC will be $\log n$ smaller than that of SIC. It is easier to reject the null hypothesis and detect a change in the data when the information criterion gets smaller.

The main difference between SIC and MIC is that MIC has a higher power than SIC to detect the change when the changes happen in the middle of the dataset, as displayed in Figures 1 and 2.

The following conclusions can be made with respect to the simulation study when there is change point:

(i) As change point location approaches middle of data, the power of the test increases

(ii) When the difference between parameters increases, the power of the test increases

TABLE 1: Power comparison between SIC and MIC as $n = 100$.

Criteria	$k = 5$	$k = 10$	$k = 20$	$k = 50$	$k = 70$	$k = 80$	$k = 90$
SIC	0.086	0.216	0.468	0.703	0.530	0.332	0.081
MIC	0.238	0.497	0.817	0.974	0.920	0.734	0.353

TABLE 2: Power comparison between SIC and MIC as $n = 200$.

Criteria	$k = 5$	$k = 10$	$k = 40$	$k = 100$	$k = 150$	$k = 170$	$k = 190$
SIC	0.052	0.175	0.866	0.979	0.858	0.543	0.038
MIC	0.229	0.497	0.986	1.000	0.997	0.923	0.286

TABLE 3: Power comparison between SIC and MIC as $n = 300$.

Criteria	$k = 5$	$k = 70$	$k = 100$	$k = 150$	$k = 170$	$k = 250$	$k = 280$
SIC	0.056	0.995	1.000	1.000	1.000	0.885	0.211
MIC	0.214	1.000	1.000	1.000	1.000	0.996	0.695

(iii) As sample size increases, the power of the test also increases

(iv) Since the MIC has a higher power than the SIC, then we use the MIC in the application of the real data in detecting a change points

3.2. Simulation Study: When There Is No Change Point. In this section, we conduct a simulation study to investigate the power of the test when there is no change point in the parameters of the distribution. EGGAPE parameters are not changing as $a^* = 0.9, a^{**} = 0.9$, $b^* = 0.8, b^{**} = 0.8$, $\alpha^* = 0.2, \alpha^{**} = 0.2$, and $\lambda^* = 0.5, \lambda^{**} = 0.5$.

We conduct simulations 1000 times under $EGGAPE(a, b, \alpha, \lambda)$ with different values of the shape parameter a, b, α and the scale parameter λ. The results for both the SIC and the MIC are given in Tables 5 and 6 when $n = 100$ and $n = 200$, respectively.

The plots of the figures when there is no change point are given in Figures 3 and 4.

The following conclusions can be made with respect to the simulation study when there is no change point. For a sample size of $n = 100$ and $n = 200$,

(i) The power of the tests are low for both SIC and MIC

(ii) Comparing the power of the test, when there is change point and when there is no change point, the power of the test is higher when there is change point, signifying that the test correctly identifies change point.

4. Application to COVID-19 Data

In this section, we introduced application to COVID-19 data for Italy, UK, and Mexico. More studies analyzed COVID-19 data [17, 23–30].

4.1. Italy COVID-19 Data. This section explains the change point analysis of COVID-19 death rates data in Italy for a

TABLE 4: Power comparison between SIC and MIC as $n = 400$.

Criteria	$k = 10$	$k = 50$	$k = 100$	$k = 150$	$k = 200$	$k = 280$	$k = 300$	$k = 380$
SIC	0.133	0.958	1.000	1.000	1.000	1.000	1.000	0.162
MIC	0.434	0.998	1.000	1.000	1.000	1.000	1.000	0.647
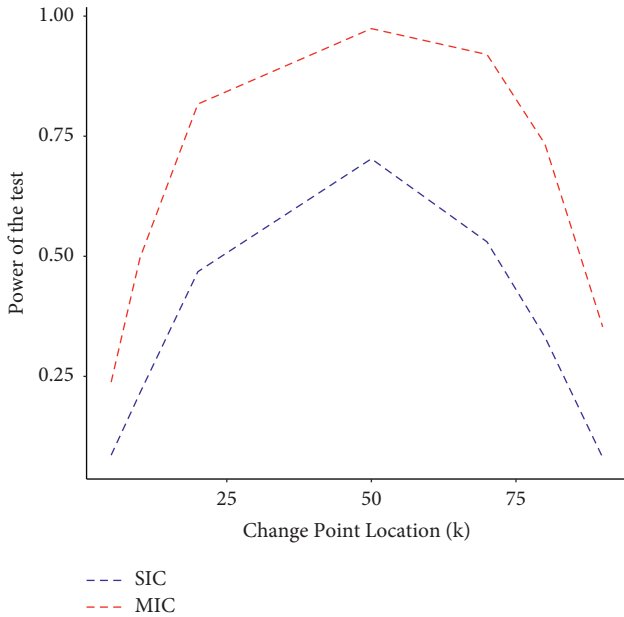
FIGURE 1: Power comparison for MIC and SIC with change point at different locations, $n = 100$.
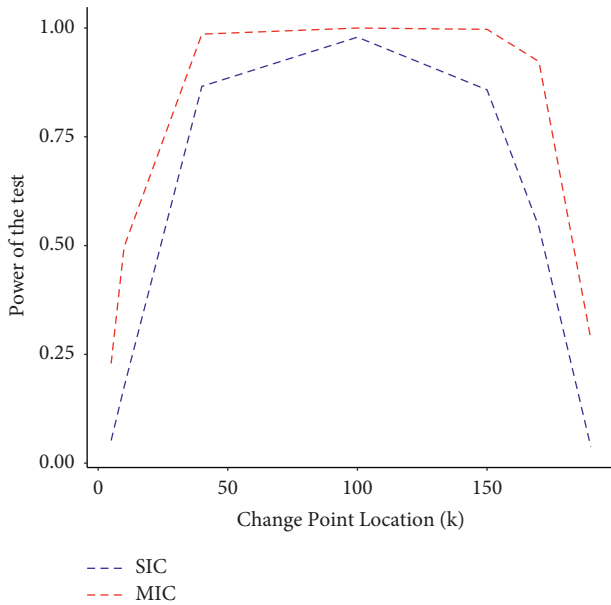

FIGURE 2: Power comparison for MIC and SIC with change point at different locations, $n = 200$.

TABLE 5: Power comparison between SIC and MIC as $n = 100$.

Criteria	$k = 5$	$k = 10$	$k = 20$	$k = 50$	$k = 70$	$k = 80$	$k = 90$
SIC	0.003	0.002	0.000	0.000	0.002	0.002	0.003
MIC	0.022	0.030	0.037	0.053	0.048	0.033	0.019

TABLE 6: Power comparison between SIC and MIC as $n = 200$.

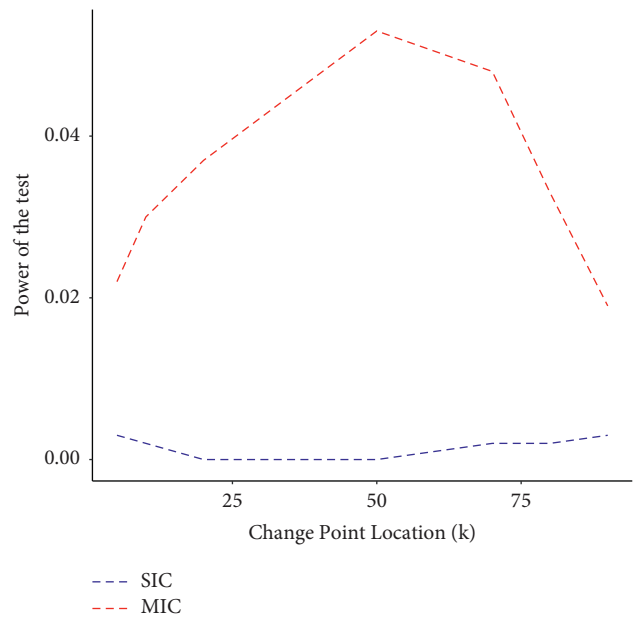| Criteria	$k = 5$	$k = 10$	$k = 40$	$k = 100$	$k = 150$	$k = 170$	$k = 190$
SIC	0.001	0.001	0.000	0.001	0.000	0.001	0.000
MIC	0.014	0.022	0.036	0.055	0.028	0.025	0.009

FIGURE 3: Power comparison for MIC and SIC with change point at different locations, $n = 100$.

A time-series plot of the dataset is displayed in Figure 5. The mortality rate is calculated as

$$\text{MortalityRate} = \frac{\text{deaths}}{\text{Population}} \times 10^6. \qquad (29)$$

All the parameters are considered changeable.

To identify a change point in the dataset of the Italy Mortality rates, we apply the test statistics defined in Equation (10). The results are displayed in Table 8.

From Table 8, H_0 is rejected and conclude that the change point exists at MIC (40) which equals the mortality rate of 5.073 and reflects the date of 2020-04-06. Based on the binary segmentation method, the dataset was separated into
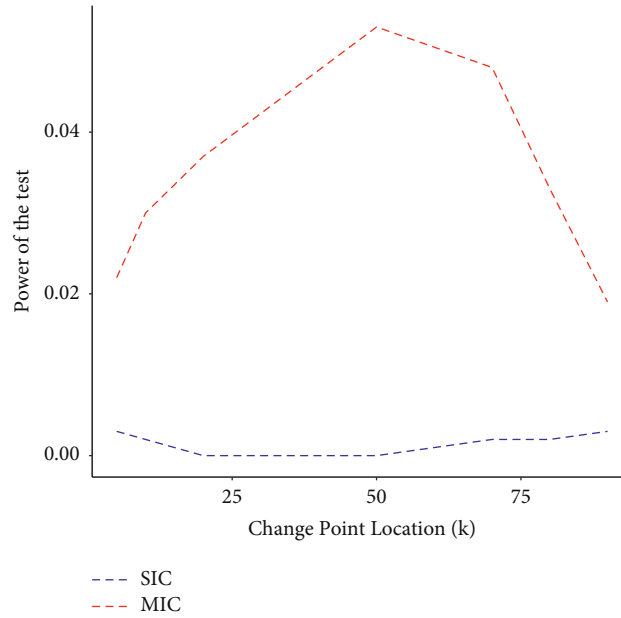
59-day period from February 2 to April 25, 2020. https://COVID-19.who.int/ was the source of the data. Table 7 gives the data.

Figure 4: Power comparison for MIC and SIC with change point at different locations, $n = 200$.

Table 7: Italy data.

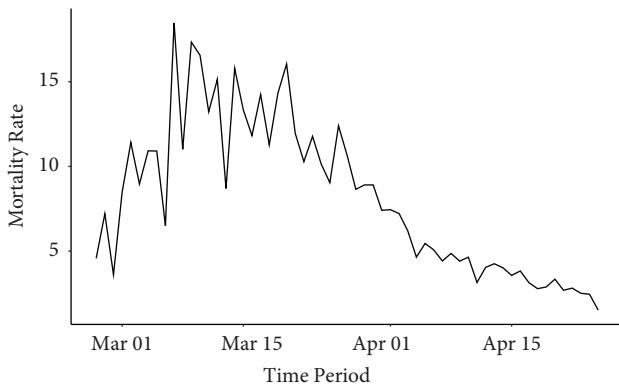|									
4.571	7.201	3.606	8.479	11.410	8.961	10.919	10.908	6.503	18.474
11.010	17.337	16.561	13.226	15.137	8.697	15.787	13.333	11.822	14.242
11.273	14.330	16.046	8.646	8.905	8.906	7.407	7.445	7.214	6.194
4.640	5.542	5.073	4.416	4.859	4.408	4.639	3.148	4.040	4.253
4.011	3.564	3.827	3.134	2.780	2.881	3.341	2.686	2.814	2.508
2.450	1.518	11.950	10.282	11.775	10.644	10.138	9.037	12.396	

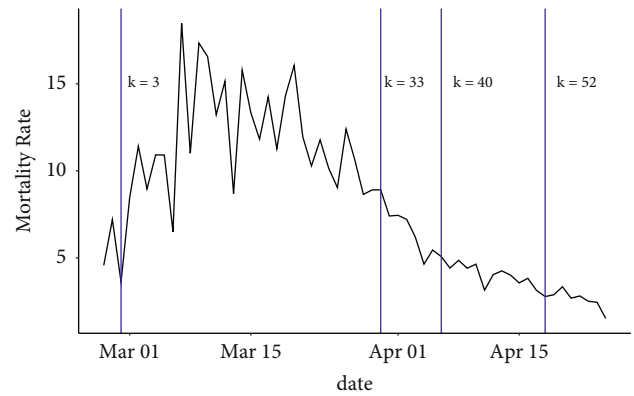Figure 5: Time-series plot for Italy COVID-19 mortality rate.

Figure 6: Change points location for Italy COVID-19 mortality rate.

Table 8: MIC test statistics for change point detection.

MIC (n)	Min MIC (k)	$\widehat{k}$
345.459	293.769	40
233.435	220.676	33
191.251	189.117	3
62.020	53.921	52

two parts. The first part is $(1 : 40)$ and the second part $(41 : 59)$. The second change point is successfully identified at $k = 33$ which reflects the mortality rate at 2020-03-30. Next we conduct the same procedure, and a change point is located at $k = 3$ corresponding to 2020-02-29. However, no further change points were located. Next, we analyze the

TABLE 9: Change point of Italy COVID-19 mortality rate data.

Change point location	Date	Possible reason
3	29-02-2020	Rapid spread of COVID-19 in Northern Italy
33	30-03-2020	Decline in number of new cases and deaths
40	06-04-2020	Extension of lock down until 13 april
52	18-04-2020	Extension of lock down until 5 may

TABLE 10: UK COVID-19 mortality rates' data.

|									
0.0587	0.0863	0.1165	0.1247	0.1277	0.1303	0.1652	0.2079	0.2395	0.2751
0.2845	0.2992	0.3188	0.3317	0.3446	0.3553	0.3622	0.3926	0.3926	0.4110
0.4633	0.4690	0.4954	0.5139	0.5696	0.5837	0.6197	0.6365	0.7096	0.7193
0.7444	0.8590	1.0438	1.0602	1.1305	1.1468	1.1533	1.2260	1.2707	1.3423
1.4149	1.5709	1.6017	1.6083	1.6324	1.6998	1.8164	1.8392	1.8721	1.9844
2.1360	2.3987	2.4153	2.5225	2.7087	2.7946	3.3609	3.3715	3.7840	3.9042
4.1969	4.3451	4.4627	4.6477	5.3664	5.4500	5.7522	6.4241	7.0657	7.4456
8.2307	9.6315	10.1870	11.1429	11.2019	11.4584				
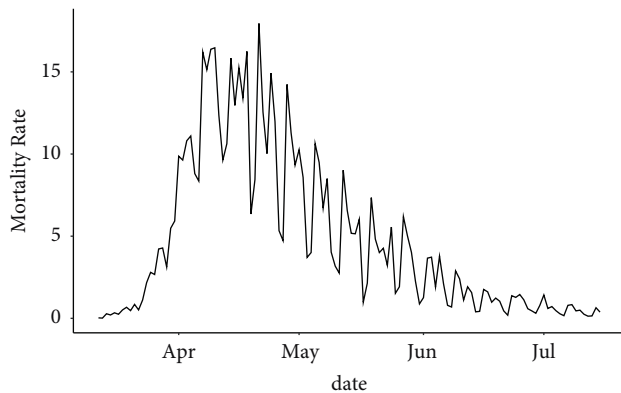
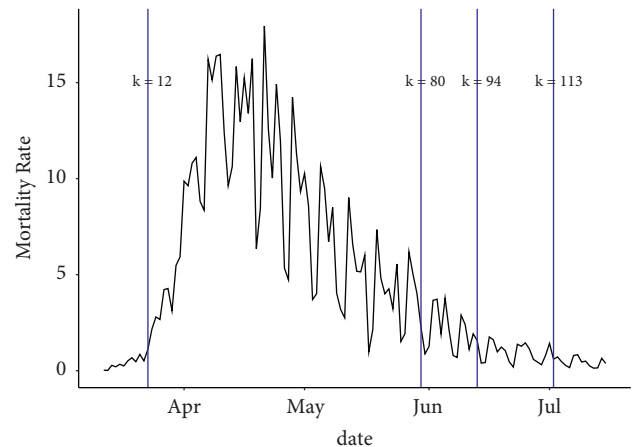FIGURE 7: Time-series plot for UK COVID-19 mortality rate.

FIGURE 8: Change points' location for UK COVID-19 mortality rate.

TABLE 11: MIC test statistics for change point detection.

MIC(n)	Min MIC(k)	$\hat{k}$
653.818	582.643	80
472.091	425.897	12
110.106	88.224	94
46.483	39.355	113

second segment (41 : 59) below. Thus, the change point occurs at 2020-04-18 located at $k = 52$. Next, we conduct the same procedure and no further change points were located.

Figure 6 represents the change points' location for Italy COVID-19 mortality rate. The possible reason for change point of Italy COVID-19 mortality data is displayed in Table 9.

The change points segmented the span into three segments. The first segment was between 29-02-2020 to 30-03-2020 characterized by high mortality rates. The second segment between 30-03-2020 to 06-04-2020 is characterized by a decline of mortality rates, and finally, the third segment

between 06-04-2020 to 18-04-2020 is characterized by a further decline in mortality rates.

4.2. Change Point Analysis for UK Data. This section describes the change point analysis of COVID-19 mortality rates data from the United Kingdom for a period of 76 days, from March 12 to July 15, 2020. https://COVID-19.who.int/ was the source of the data. The data are presented in Table 10.

A time-series plot of the dataset is displayed in Figure 7. The mortality rate is calculated as

$$\text{MortalityRate} = \frac{\text{deaths}}{\text{Population}} \times 10^6. \tag{30}$$

All the parameters are considered changeable.

To identify the change point in the dataset of the UK mortality rates, we apply the test statistics defined in Equation (10). The results are displayed in Table 11.

TABLE 12: Change point of UK COVID-19 mortality rate data.

Change point location	Date	Possible reason
12	23-03-2020	Increase in number of cases and hospitalized patients
80	30-05-2020	Decline in number of new deaths
94	13-06-2020	Reinforcement of lock down measures
113	02-07-2020	Compulsory wearing of masks

TABLE 13: Mexico COVID-19 mortality rates' data.

|										
8.826	6.105	10.383	7.267	13.220	6.015	10.855	6.122	10.685	5.242	7.630
14.604	7.903	6.327	9.391	14.962	4.730	3.215	16.498	11.665	9.284	12.878
6.656	3.440	5.854	8.813	10.043	7.260	5.985	4.424	4.344	5.143	9.935
7.840	9.550	6.968	6.370	3.537	3.286	10.158	8.108	6.697	7.151	6.560
2.988	3.336	6.814	8.325	7.854	8.551	3.228	3.499	3.751	7.486	6.625
6.140	4.909	4.661	1.867	2.838	5.392	12.042	8.696	6.412	3.395	1.815
3.327	5.406	6.182	4.949	4.089	3.359	2.070	3.298	5.317	5.442	4.557
4.292	2.500	6.535	4.648	4.697	4.120	3.922	3.219	1.402	2.438	3.257
3.632	3.233	3.027	2.352	1.205	2.077	3.778	3.218	2.926	2.601	2.065
1.041	1.800	3.029	2.058	2.326	2.506	1.923				
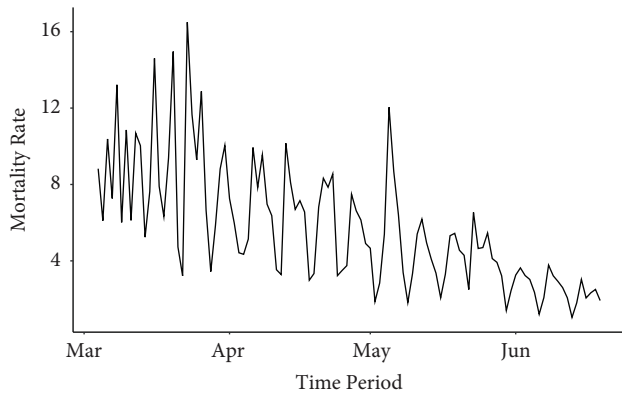
FIGURE 9: Time-series plot for Mexico COVID-19 mortality rate.

TABLE 14: MIC test statistics for change point detection.

MIC(n)	Min MIC(k)	$\hat{k}$
550.198	509.681	64
548.053	342.571	23
543.487	139.192	15
167.263	145.139	86

A visual display of the change point locations in the dataset is displayed in Figure 8.

The possible causes of the change point in the UK COVID-19 mortality rate data are given in Table 12.

4.3. Change Point Analysis for Mexico Data. This section explains the change point analysis of COVID-19 death rates for Mexico for a period of 108 days that is from 4 March to 19 June 2020. https://COVID-19.who.int/ was the data source. The data are given in Table 13.

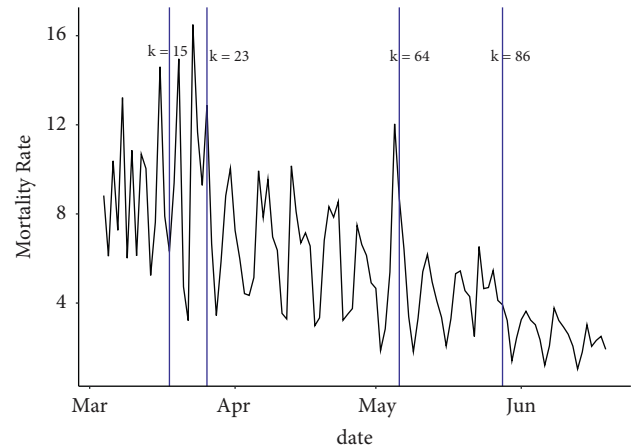A time-series plot of the dataset is displayed in Figure 9. The mortality rate is calculated as

FIGURE 10: Change points' location for Mexico COVID-19 mortality rate.

TABLE 15: Change point of Mexico COVID-19 mortality rate data.

Change point location	Date	Possible reason
15	18-03-2020	Surge in number of cases
23	26-03-2020	Decline in number of new cases and deaths
64	06-05-2020	Rise in number of deaths
86	28-05-2020	Compulsory wearing of masks

$$\text{MortalityRate} = \frac{\text{deaths}}{\text{Population}} \times 10^6. \qquad (31)$$

To detect the change point in the dataset of the Mexico Mortality rates, we apply the test statistics defined in Equation (10). The results are displayed in Table 14.

A visual display of the change point locations in the dataset is displayed in Figure 10.

The possible causes of the change point in the Mexico COVID-19 mortality rate data are given in Table 15.

5. Conclusions

Although the EGGAPE distribution is a more flexible distribution that may describe data with monotonic and nonmonotonic hazard shapes, few or no studies of the change point problem for such a distribution have been done. For this study, we present a change point detection method for a four parameter EGGAPE distribution based on the information approach specifically modified information criterion (MIC). All the parameters are considered changeable. The benefit of using MIC-based test is in order to avoid the complications of deriving the complicated asymptotic distributions of test statistic of likelihood ratio test and cumulative sum tests. In addition, we have applied the binary segmentation to detect multiple change points and their locations. In the simulation study for the power of the test, two scenarios were considered: a simulation study when there was a change point and a simulation study when there was no change point. When there was a change point, the power of the test was high, and when there was no change point, the power of the test was so small. The testing procedure is applied to three real datasets related to COVID-19 mortality rates in Italy, the United Kingdom, and Mexico. Multiple change points were successfully identified and their location identified. In this study, we have only considered a case where all the parameters are changing, for future work, study can be done when at least one of the parameters is not changing.

Data Availability

The data used to support the findings of the study are available within the article.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

References

[1] A. Sen and M. S. Srivastava, "On tests for detecting change in mean," *Annals of Statistics*, vol. 3, no. 1, pp. 98–108, 1975.

[2] A. Sen and M. S. Srivastava, "Some one-sided tests for change in level," *Technometrics*, vol. 17, no. 1, pp. 61–64, 1975.

[3] K. J. Worsley, "On the likelihood ratio test for a shift in location of normal populations," *Journal of the American Statistical Association*, vol. 74, no. 366, pp. 365–367, 1979.

[4] G. Ngunkeng and W. Ning, "Information approach for the change-point detection in the skew normal distribution and its applications," *Sequential Analysis*, vol. 33, no. 4, pp. 475–490, 2014.

[5] A. Ramanayake and A. K. Gupta, "Testing for a change point in a sequence of exponential random variables with repeated values," *Journal of Statistical Computation and Simulation*, vol. 80, no. 2, pp. 191–199, 2010.

[6] A. S. Hassan, E. M. Almetwally, M. A. Khaleel, and H. F. Nagy, "Weighted power lomax distribution and its length biased version: properties and estimation based on censored samples," *Pakistan Journal of Statistics and Operation Research*, vol. 17, pp. 343–356, 2021.

[7] M. Shrahili, A. S. Hassan, E. M. Almetwally, A. Ben Ghorbal, and I. Elbatal, "Alpha power moment exponential model with applications to biomedical science," *Scientific Programming*, vol. 2022, Article ID 6897405, 14 pages, 2022.

[8] J. Chen, K. Arjun, and Gupta, *Parametric Statistical Change point Analysis: With Applications to Genetics, Medicine, and Finance*, 2012.

[9] J. Chen, A. K. Gupta, and J. Pan, "Information criterion and change point problem for regular models," *Sankhya: The Indian Journal of Statistics*, pp. 252–282, 2006.

[10] R. B. Arellano-Valle, L. M. Castro, and R. H. Loschi, "Change point detection in the skew-normal model parameters," *Communications in Statistics - Theory and Methods*, vol. 42, no. 4, pp. 603–618, 2013.

[11] A. Alghamdi, W. Ning, K. Arjun, and Gupta, "An information approach for the change point problem of the Rayleigh lomax distribution," *International Journal of Intelligent Technologies and Applied Statistics*, vol. 11, no. 4, pp. 233–254, 2018.

[12] E. M. Almetwally, "The odd weibull inverse topp–leone distribution with applications to covid-19 data," *Annals of Data Science*, vol. 9, no. 1, pp. 121–140, 2022.

[13] S. Ratnasingam and W. Ning, "Confidence distributions for skew normal change-point model based on modified information criterion," *Journal of Statistical Theory and Practice*, vol. 14, no. 3, pp. 42–21, 2020.

[14] M. Kilai, G. A. Waititu, W. A. Kibira, H. M. Alshanbari, and M. El-Morshedy, "A new generalization of gull alpha power family of distributions with application to modeling covid-19 mortality rates," *Results in Physics*, vol. 36, Article ID 105339, 2022.

[15] E. A. ElSherpieny and E. M. Almetwally, "The exponentiated generalized alpha power family of distribution: properties and applications," *Pakistan Journal of Statistics and Operation Research*, vol. 18, pp. 349–367, 2022.

[16] V. K. Jandhyala, S. B. Fotopoulos, and N. Evaggelopoulos, "Change-point methods for weibull models with applications to detection of trends in extreme temperatures," *Environmetrics*, vol. 10, no. 5, pp. 547–564, 1999.

[17] H. M. Almongy, E. M. Almetwally, H. M. Aljohani, A. S. Alghamdi, and E. H. Hafez, "A new extended Rayleigh distribution with applications of covid-19 data," *Results in Physics*, vol. 23, Article ID 104012, 2021.

[18] E. H. Hafez, R. Alharbi, E. M Almetwally, B. Alruwaili, A. M Gemeay, and F. H. Riad, "Single and multiple ramp progressive stress with binomial removal: practical application for industry," *Mathematical Problems in Engineering*, vol. 2022, Article ID 9558650, 21 pages, 2022.

[19] D. Jarušková, "Maximum log-likelihood ratio test for a change in three parameter weibull distribution," *Journal of Statistical Planning and Inference*, vol. 137, no. 6, pp. 1805–1815, 2007.

[20] S. Ratnasingam and W. Ning, "Modified information criterion for regular change point models based on confidence distribution," *Environmental and Ecological Statistics*, vol. 28, no. 2, pp. 303–322, 2021.

[21] G. Schwarz, "Estimating the Dimension of a Model," *The annals of statistics*, pp. 461–464, 1978.

[22] B. Ben and M. Ben Bolker, "Package 'bbmle,'" *Tools for General Maximum Likelihood Estimation*, vol. 641, 2017.

[23] E. M. Almetwally, R. Alharbi, D. Alnagar, and E. Hafez, "A new inverted topp-leone distribution: applications to the covid-19 mortality rate in two different countries," *Axioms*, vol. 10, no. 1, p. 25, 2021.

[24] H. Alsuhabi, I. Alkhairy, E. M. Almetwally et al., "A superior extension for the lomax distribution with application to covid-19 infections real data," *Alexandria Engineering Journal*, vol. 61, no. 12, Article ID 11077, 2022.

[25] E. M. Almetwally, D. A. Abdo, E. H. Hafez, Ta M. Jawa, N. Sayed-Ahmed, and H. M. Almongy, "The new discrete distribution with application to covid-19 data," *Results in Physics*, vol. 32, Article ID 104987, 2022.

[26] S. A. Gargoum and A. S. Gargoum, "Limiting mobility during covid-19, when and to what level? an international comparative study using change point analysis," *Journal of Transport & Health*, vol. 20, Article ID 101019, 2021.

[27] S. Ballı, "Data analysis of covid-19 pandemic and short-term cumulative case forecasting using machine learning time series methods," *Chaos, Solitons & Fractals*, vol. 142, Article ID 110512, 2021.

[28] R. Huang, M. Liu, and Y. Ding, "Spatial-temporal distribution of covid-19 in China and its prediction: a data-driven modeling analysis," *The Journal of Infection in Developing Countries*, vol. 14, no. 03, pp. 246–253, 2020.

[29] F. Jiang, Z. Zhao, and X. Shao, "Time series analysis of covid-19 infection curve: a change-point perspective," *Journal of Econometrics*, 2020.

[30] P. K. St, B. Lahiri, and R. Alvarado, "Multiple change point estimation of trends in covid-19 infections and deaths in India as compared with who regions," *Spatial Statistics*, vol. 49, Article ID 100538, 2022.