



## OPEN ACCESS

## EDITED BY

Han-Ying Liang,  
Tongji University, China

## REVIEWED BY

Ferra Yanuar,  
Andalas University, Indonesia  
Fuxia Cheng,  
Illinois State University, United States

## \*CORRESPONDENCE

Ahmed M. Gad  
✉ ahmed.gad@feeps.edu.eg

## SPECIALTY SECTION

This article was submitted to  
Statistics and Probability,  
a section of the journal  
Frontiers in Applied Mathematics and Statistics

RECEIVED 24 November 2022

ACCEPTED 03 February 2023

PUBLISHED 24 February 2023

## CITATION

Ibrahim WIM, Gad AM and Abd-Rabou A-ES  
(2023) A mixed effects changepoint quantile  
regression model for longitudinal data with  
application on COVID-19 data.  
*Front. Appl. Math. Stat.* 9:1106958.  
doi: 10.3389/fams.2023.1106958

## COPYRIGHT

© 2023 Ibrahim, Gad and Abd-Rabou. This is an  
open-access article distributed under the terms  
of the [Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction  
in other forums is permitted, provided the  
original author(s) and the copyright owner(s)  
are credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted which  
does not comply with these terms.

# A mixed effects changepoint quantile regression model for longitudinal data with application on COVID-19 data

Wafaa I. M. Ibrahim, Ahmed M. Gad\* and  
Abd-Elnaser S. Abd-Rabou

Statistics Department, Faculty of Economics and Political Science, Cairo University, Giza, Egypt

**Introduction:** Longitudinal individual response profiles could exhibit a mixture of two or more phases of increase or decrease in trend throughout the follow-up period, with one or more unknown transition points (changepoints). The detection and estimation of these changepoints is crucial. Most of the proposed statistical methods for detecting and estimating changepoints in literature rely on distributional assumptions that may not hold. In this case, a good alternative is to use a robust approach; the quantile regression model. There are methods in the literature to deal with quantile regression models with a changepoint. These methods ignore the within-subject dependence of longitudinal data.

**Methods:** We propose a mixed effects quantile regression model with changepoints to account for dependence structure in the longitudinal data. Fixed effects parameters, in addition to the location of the changepoint, are estimated using the profile estimation method. The stochastic approximation EM algorithm is proposed to estimate the fixed effects parameters exploiting the link between an asymmetric Laplace distribution and the quantile regression. In addition, the location of the changepoint is estimated using the usual optimization methods.

**Results and discussion:** A simulation study shows that the proposed estimation and inferential procedures perform reasonably well in finite samples. The practical use of the proposed model is illustrated using COVID-19 data. The data focus on the effect of global economic and health factors on the monthly death rate due to COVID-19 from 1 April 2020 to 30th April 2021. The results show a positive effect on the monthly number of patients with COVID-19 in intensive care units (ICUs) for both 0.5th and 0.8th quantiles of new monthly deaths per million. The stringency index, hospital beds, and diabetes prevalence have no significant effect on both 0.5th and 0.8th quantiles of new monthly deaths per million.

## KEYWORDS

longitudinal data, change point model, COVID-19, mixed effects quantile regression, bent line quantile regression, missing data

## 1. Introduction

Longitudinal studies play a prominent role in the health, social, and behavioral sciences, and many other disciplines. A response variable, of the same individual, is measured repeatedly over time, or under different conditions. The main aim of longitudinal studies is to study the change in the response variable over time. Due to the nature of the longitudinal data, in which a set of measurements for each subject is observed over time, there will be a correlation within the subject, i.e., measurements for each subject are correlated. Hence, special statistical analysis methods are needed for longitudinal data to accommodate the potential patterns of correlation. Ignoring such correlation may lead to invalid statistical inferences [1].

The longitudinal individual response profiles could exhibit a mixture of two or more phases of increase or decrease in trend, throughout the follow-up period. This could be at one or more unknown transition points, usually referred to as breakpoints or change points. Such change points are quite common in public health, medical, and many other disciplines. Change point models are useful to determine when changes have taken place and to use one model for the whole data. Change point models with one change point and two linear phases are most used especially in biological data [2]. Recently, there has been an increased interest in the application of change point models to longitudinal data. In the Bayesian framework, see for example, Ghosh and Vaida [3], Yang and Gao [2], and McLain and Albert [4]. Xing and Ying [5] proposed a semi-parametric change point regression model for longitudinal data. Lai and Albert [6] proposed a linear mixed effects modeling framework for multiple change points in longitudinal Gaussian data.

Most of the proposed techniques to fit longitudinal data with change points rely upon distributional assumptions, such as normality. These distributional assumptions may not generally hold. On the contrary, in some applications, the relationship between the response and covariates at the tails, rather than the center of the distribution is of main interest [7]. The quantile regression model is a good alternative when the distributional assumptions do not hold. The quantile regression does not require distributional assumptions. The quantile regression fits the conditional quantiles of the response variable given a set of covariates. The main advantage of quantile regression is its ability to provide a more complete picture of the conditional distribution of the response variable given the covariates. The quantile regression is particularly useful when upper or lower (or any other) quantiles are of interest. It is more flexible for modeling data with heterogeneous conditional distributions. Moreover, the quantile regression is robust to outliers in the response variable. There is research in the literature on using quantile regression or median regression in a longitudinal data context. Some of these works of literature employ the marginal models, see for example, Gad and Ibrahim [1], Jung [8], and Wang and Fygenon [9]. Others employed mixed effect quantile regression, see for example, Galarza et al. [10], Yirga et al. [11], and Tian et al. [12].

Li et al. [7] extended the quantile regression model with a change point, introduced by Li et al. [13], to accommodate longitudinal data. These models ignore the within-subject dependence of longitudinal data [7, 14]. Incorporating random effects in these models is a remedy to accommodate the within-subject dependence. Yu and Moyeed [15] used the connection between an asymmetric Laplace distribution (ALD) and the quantile regression model to incorporate random effects in the model. Liu and Bottai [16] developed a likelihood-based inferential approach for estimating parameters of mixed effects quantile regression models. They assume an ALD for the errors and the multivariate Laplace distribution (MLD) for the random effects. They use the MCEM algorithm for estimation and inference. Abdelwahab et al. [17] proposed a CUSUM test for testing the existence of change points in the quantile regression model for longitudinal datasets.

There are different stochastic algorithms to estimate the parameters for mixed effects quantile models rather than MCEM, such as the stochastic approximation EM (SAEM) algorithm [18]. The SAEM algorithm approximates the E-step of the EM algorithm, by splitting the E-step, into a simulation step and an approximation step. The SAEM algorithm has been proven to be more computationally efficient than the classical MCEM algorithm. This is because of the reusing of simulations from one cycle to another within the smoothing phase of the algorithm. Meza et al. [18] stated that the SAEM algorithm converges in a small number of simulations unlike the MCEM algorithm, which needs a large number of simulations.

The aim of this article is to extend the quantile regression model with a change point to accommodate the within-subject dependence of the longitudinal data. This is tackled *via* incorporating random effects into the model. A likelihood-based inferential approach is developed by assuming an ALD for the errors and the MLD distribution for the random effects. The advantage of using the multivariate Laplace distribution is to accommodate any number of possible outliers. Moreover, it can handle heavy-tailed distributions. The stochastic approximation EM (SAEM) algorithm is proposed and developed to obtain the MLE of the model parameters. The location of the change point is estimated using optimization methods. The SAEM algorithm exploits the link between an asymmetric Laplace distribution (ALD) and the quantile regression. The proposed techniques are evaluated using a simulation study. Furthermore, the proposed techniques are illustrated by real data. The rest of this article is organized as follows: Section Proposed model and estimation procedure presents the proposed quantile regression model and the estimation procedure. Simulation studies are conducted to assess the performance of the proposed techniques and to study the goodness-of-fit for the proposed model. Section Simulation study presents the simulation results. Section Application: COVID-19 applies the proposed model to COVID-19 data. We test whether there is a threshold effect (change point) in the relationship between the HDI and new monthly deaths per million. Finally, Section Conclusion is devoted to the concluding remarks and future study.

## 2. Proposed model and estimation procedure

### 2.1. Model

Li et al. [7] extended the quantile regression model with a change point, introduced by Li et al. [13], to accommodate longitudinal data. However, this model ignores the within-subject dependence in Li et al. [7] and Sha [14]. We suggest extending this model to account for the within-subject dependence by adding random effects, to capture the dependence structure of longitudinal data. Hence, the authors modify the Li et al. [7] model by adding a random effect term ( $z_{ij}^T U_i$ ). The modified model is as follows:

$$y_{ij} = \alpha_{\tau} + (\beta_{1,\tau} I\{x_{ij} \leq t_{\tau}\} + \beta_{2,\tau} I\{x_{ij} \geq t_{\tau}\})(x_{ij} - t_{\tau}) + z_{ij}^T \gamma_{\tau} + z_{ij}^T U_i + \varepsilon_{\tau,ij}, \quad (1)$$

for  $j = 1, \dots, m_i$ , and  $i = 1, \dots, n$ , where  $I\{a\}$  is an indicator function,  $I\{a\} = 1$  if  $a$  is true, and 0 otherwise. The  $y_{ij}$  is the response variable of subject  $i$  at time point  $j$ ,  $x_{ij}$  is the covariate whose slope changes at an unknown change point  $t_\tau$ , and  $s_{ij}$  is a  $q$ -dimensional vector of linear covariates with constant slopes. Moreover,  $z_{ij}$  is a  $p \times 1$  subset of  $s_{ij}$  with random effects;  $U_i$  is a  $p \times 1$  vector of random regression coefficients;  $\tau^{th}$  conditional quantile of  $\varepsilon_{\tau,ij}$  gives  $x_{ij}, z_{ij}$ , and  $s_{ij}$  is 0, and  $\varepsilon_{\tau,ij}$  is assumed to be independently distributed as an asymmetric Laplace distribution (ALD). The random regression coefficients  $U_i$ , which account for the correlation among observations, are assumed to be mutually independent and to follow the multivariate Laplace distribution (MLD). Independence between  $U_i$  and  $\varepsilon_{\tau,ij}$ , and between the random regression coefficients  $U_i$  and the explanatory variables  $x_{ij}, s_{ij}$ , are assumed. All the model parameters may be expressed as  $\theta_\tau = (\eta_\tau, t_\tau)^T$  and  $\eta_\tau = (\alpha_\tau, \beta_{1,\tau}, \beta_{2,\tau}, \gamma_\tau^T)^T$ .

In practice, the normality assumption of the random effects may be violated for many reasons, such as outliers, contaminated data, and heavy-tailed distributions. The multivariate Laplace distribution is a good robust alternative in this case [19]. The distribution of the random effects  $U_i$  in model (1) is assumed to be a symmetric-multivariate Laplace distribution with zero mean as follows [19]:

$$f(U_i | \Sigma) = \frac{2}{(2\pi)^{\frac{p}{2}} |\Sigma|^{-\frac{1}{2}}} \left( \frac{U_i^T \Sigma^{-1} U_i}{2} \right)^{\frac{p}{2}} \times K_\nu \sqrt{2(U_i^T \Sigma^{-1} U_i)}, \tag{2}$$

Where  $\Sigma$  is a  $P \times P$  covariance matrix, and  $\nu = \frac{2-p}{2}$  and  $K_\nu(u)$  is the modified Bessel function of the third kind, which is given in Equation (3) as follows [19]:

$$K_\nu(u) = \frac{1}{2} \left( \frac{u}{2} \right)^\nu \int_1^\infty t^{-\lambda-1} \exp\left(-t - \frac{u^2}{4t}\right) dt, u > 0. \tag{3}$$

## 2.2. Estimation and inference

The estimates  $\hat{\theta}_n(\tau)$  are obtained by minimizing the objective function as discussed in Li et al. [7], which is given in Equation (4) as follows:

$$Q_{n,\tau}(\theta) = \sum_{ij} \rho_\tau(y_{ij} - \mu_{ij}), \tag{4}$$

Where  $\rho_\tau(u) = u(\tau - I(u \leq 0))$  is the quantile loss function and  $\mu_{ij} = \alpha_\tau + (\beta_{1,\tau} I\{x_{ij} \leq t_\tau\} + \beta_{2,\tau} I\{x_{ij} \geq t_\tau\})(x_{ij} - t_\tau) + s_{ij}^T \gamma_\tau + z_{ij}^T U_i$ . However, due to the presence of a change point, the objective function  $Q_{n,\tau}(\theta)$  is non-convex [7]. Hence, the estimates  $\hat{\theta}_n(\tau)$  are obtained via profile estimation. The stochastic approximation EM algorithm is used to estimate the fixed effects parameters  $\eta_\tau$ . The estimates can be obtained using the link between an ALD and the quantile regression as suggested by Yu and Moyed [15]. In addition, the location of the change point is estimated using optimization methods.

### 2.2.1. SAEM algorithm to estimate $\eta_\tau(t)$

At fixed  $t$ , the initial value of  $t$  is determined graphically, the profile estimator  $\eta$  is given in equation (5) as discussed by Li et al. as follows [7]:

$$\hat{\eta}_{n,\tau}(t) = \arg Q_{n,\tau}(\eta, t) \tag{5}$$

We propose the stochastic approximation EM algorithm (SAEM) to estimate  $\eta_\tau(t)$ . Galarza et al. [20] used the SAEM algorithm to develop a likelihood-based approach to fit the quantile regression model for continuous longitudinal data using an ALD distribution. They assume that the distribution for the random effects is the multivariate Gaussian. In this article, we assume that the random effects follow an ALD. At fixed  $t$ , minimizing the loss function in Equation (5) is the same as maximizing an ALD likelihood function. A likelihood-based inferential approach is developed to estimate  $\hat{\eta}_{n,\tau}(t)$  in Equation (2) by using the connection between an ALD distribution and the quantile regression [15]. This is done by assuming an ALD for the errors and the multivariate Laplace distribution for the random effects. At fixed  $t$ , the conditional density function of  $y_{ij}|U_i$  can be written as follows [16]:

$$f(y_{ij} | U_i, x_{ij}, S_{ij}; \eta_\tau(t), \sigma) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{-\rho_\tau\left(\frac{y_{ij} - \mu_{ij}}{\sigma}\right)\right\} \tag{6}$$

where

$$\mu_{ij} = \alpha_\tau + (\beta_{1,\tau} I\{x_{ij} \leq t_\tau\} + \beta_{2,\tau} I\{x_{ij} \geq t_\tau\})(x_{ij} - t_\tau) + s_{ij}^T \gamma_\tau + z_{ij}^T U_i \tag{7}$$

is a linear predictor of the  $\tau^{th}$  quantile function at fixed  $t$ . The  $\tau$  is assumed to be fixed and known.

Let  $f(Y_i | U_i, X_i, S_i; \eta_\tau(t), \sigma) = \prod_{j=1}^{m_i} f(y_{ij} | U_i, x_{ij}, s_{ij}; \eta_\tau(t), \sigma)$  be the density for the  $i^{th}$  subject conditional on the random effect  $U_i$ , where  $Y_i = [y_{i1} \& y_{i2} \dots \& y_{im_i}]^T$ ,  $X_i = [x_{i1} \& x_{i2} \dots \& x_{im_i}]^T$ , and  $S_i = [s_{i1} \& s_{i2} \dots \& s_{im_i}]^T$ . The complete data density of  $(Y_i, U_i)$  for  $i = 1, 2, \dots, m_i$ , can be written as follows:

$$f(Y_i, U_i | X_i, S_i; \omega) = f(Y_i | U_i, X_i, S_i; \eta_\tau(t), \sigma) \cdot f(U_i | X_i, S_i; \Sigma) \tag{8}$$

As  $U_i$  and the explanatory variables  $X_i, S_i$  are assumed to be independent,  $f(U_i | \Sigma)$  is the density of  $U_i$ , and  $\omega = (\eta_\tau(t), \sigma, \Sigma)$  is the set of parameters of interest. If we let  $Y = (Y_1, Y_2, \dots, Y_n)$ ,  $X = (X_1, X_2, \dots, X_n)$ ,  $S = (S_1, S_2, \dots, S_n)$ , and  $U = (U_1, U_2, \dots, U_n)$ , the joint density of  $(Y, U)$  based on the  $n$  subjects can be derived as follows:

$$f(Y, U | X, S; \omega) = \prod_{i=1}^n f(Y_i | U_i, X_i, S_i; \eta_\tau(t), \sigma) \cdot f(U_i | \Sigma) \tag{9}$$

The maximum likelihood estimates for the parameter  $\omega$  are obtained by maximizing the marginal density  $f(Y|\omega)$ , which is obtained by integrating out the random effect  $U$  in Equation (9), that is,  $L(\omega; Y) = \int f(Y|U; \omega) \cdot f(U; \Sigma) dU$  [16]. In many cases, this integral has no closed form. Hence, the SAEM algorithm is proposed to maximize this function. Within this algorithm, the random effects are considered unobserved (missing values).

The three steps of the SAEM are as follows:

### 2.2.1.1. Simulation step

In the simulation step, at the  $(s+1)$ th step, a sample of size  $l_{(s+1)}$  is generated from the conditional distribution  $f(U_i|Y_i; \omega^s)$ , i.e.,

$$U_{ik}^{s+1} \sim f(U_i|Y_i; \omega^s) \text{ for } k = 1, 2, 3, \dots, l_{(s+1)}.$$

The conditional distribution does not have a standard form. Thus, the Metropolis–Hastings algorithm is adopted. The iterations are as follows [21]:

1. Initialize the parameters  $\omega^s = (\eta_\tau^s(t), \sigma^s, \Sigma^s)$  at  $s = 0$ . The initial values of  $(\eta_\tau^s(t), \sigma^s)$  are calculated using the fixed quantile regression model. While the initial values of  $\Sigma^s$  are calculated using moment methods for the simulated random effects.
2. For each subject, independently draw a sample  $\{U_{ik}^s; k = 1, \dots, l_{(s+1)}\}$  from the conditional distribution  $f(U_i|Y_i; \omega^s)$  using the Metropolis–Hastings algorithm. The proposal distribution is the density of the random effects  $f(U_i)$ , while  $f(U_i|Y_i; \omega^s)$  is the target distribution that takes the following form as suggested in equation (10) [21]:

$$f(U_i|Y_i; \omega^s) \propto f(Y_i|U_i, X_i, S_i; \eta_\tau^s(t), \sigma) \cdot f(U_i|\Sigma^s). \quad (10)$$

The choice of the proposal distribution is essential for the convergence of the Metropolis–Hastings algorithm. Different choices of the proposal covariance matrix lead to different results. If the variability is very small, then all moves will be accepted. However, the chain will not mix well. On the contrary, if the variability is very large, then most proposed moves will be rejected; consequently, the chain will not move. A simple solution to this problem is to calculate the acceptance rate (the fraction of proposed moves that is accepted) and choose the value of the standard deviation so that the acceptance rate is far from 0 and far from 1 [22].

### 2.2.1.2. Approximation step

The EM algorithm [23] evaluates the expected value of the Q function as  $Q(\theta|\theta^{(t)}) = E\{l(\theta|Y)|Y_{obs}\}$ , where  $l(\theta|Y)$  is the log-likelihood function. At the  $(s+1)$ th iteration, the SAEM algorithm

(the adopted algorithm) approximates the Q function as follows [10, 20]:

$$\begin{aligned} Q(\omega|\omega^{(s+1)}) &= (1 - \varphi_s) Q(\omega|\omega^{(s)}) \\ &+ \varphi_s \frac{1}{l_{s+1}} \sum_{k=1}^{l_{s+1}} L(\omega; Y_i, U_{ik}^{(s+1)}) \\ &= Q(\omega|\omega^{(s)}) + \varphi_s \left\{ \frac{1}{l_{s+1}} \sum_{k=1}^{l_{s+1}} L(\omega; Y_i, U_{ik}^{(s+1)}) \right. \\ &\left. - Q(\omega|\omega^{(s)}) \right\}, \end{aligned} \quad (11)$$

Where  $\varphi_s$  is a smoothness parameter which is a decreasing sequence of positive numbers such that  $\sum_{t=1}^{\infty} \varphi_t \rightarrow \infty$ ,  $\sum_{t=1}^{\infty} \varphi_t^2 < \infty$ , and  $L(\omega; Y_i, U_{ik}^{(s+1)})$  is the pseudo log-likelihood for the  $i^{th}$  subject at  $(s+1)$ th step. It can be easily seen that the pseudo log-likelihood takes the following form:

$$\begin{aligned} L(\omega; Y, U) &= \log \prod_{i=1}^n f(Y_i|U_i, X_i, S_i; \eta_\tau(t), \sigma) \cdot f(U_i|\Sigma) \\ &= \sum_{i=1}^n \log(f(Y_i|U_i, X_i, S_i; \eta_\tau(t), \sigma) \cdot f(U_i|\Sigma)) \\ &= \sum_{i=1}^n \log(f(Y_i|U_i, X_i, S_i; \eta_\tau(t), \sigma)) \\ &+ \sum_{i=1}^n \log(f(U_i|\Sigma)). \end{aligned} \quad (12)$$

### 2.2.1.3. Maximization step

In the maximization step,  $Q(\omega|\omega^{(s)})$  is maximized to update the parameter estimates.

The aforementioned steps are repeated until convergence. The value of the smoothing parameter  $\varphi_t$  governs the convergence of the estimates. If the smoothing parameter  $\varphi_t$  is equal to 1 for all iterations, then the SAEM algorithm will be equivalent to the MCEM algorithm. This is because the algorithm does not take any memory into consideration. In this case, the SAEM will converge quickly (convergence in distribution) to a neighborhood solution. On the contrary, when the smoothing parameter  $\varphi_t$  is different from 1, the algorithm will converge slowly (almost sure convergence) to the ML solution [20].

Galarza et al. [20] suggested the following choice of the smoothing parameter:

$$\varphi_t = \begin{cases} 1 & 1 \leq S \leq cW \\ \frac{1}{T-cW} & cW + 1 \leq S \leq W \end{cases},$$

Where  $W$  is the maximum number of Monte Carlo iterations, and  $c$  determines the percentage of initial iterations with no memory. It takes a value between 0 and 1, that is, the algorithm will have memory for all iterations if  $c = 0$ , and in this case, the algorithm will converge slowly to the ML estimates, and  $W$  needed to be large to achieve the ML estimates. However, if  $c = 1$ , the algorithm will have no memory and so will converge quickly to a neighborhood solution. In this case ( $c = 1$ ), the algorithm will result in a Markov chain where the mean of the chain observations

can be a satisfactory estimate, after removing a burn-in period [20]. A number between 0 and 1 ( $0 < c < 1$ ) will ensure an initial convergence, in distribution, to a solution neighborhood for the first cW iterations and an almost sure convergence for the rest of the iterations. Hence, this combination will lead to a fast algorithm with good estimates.

For the SAEM algorithm, the E-Step coincides with the MCEM algorithm, but a small number of simulations  $l$  (advised to be  $l \leq 20$ ) is necessary. This is feasible because the SAEM algorithm not only uses some or all previous simulations but also the current simulation of the missing data. This “memory” property is set by the smoothing parameter  $\phi_t$ , and this is unlike the traditional EM algorithm and its variants [20].

When implementing the SAEM algorithm, several settings must be fixed. These include the number of total iterations  $W$  and the cut point  $c$  that defines the starting of the smoothing step. However, choosing those parameters depend on the model and the data. A graphical approach is possible to choose these constants such that the convergence of the estimates for all the parameters can be monitored. Moreover, it is possible to monitor the difference (relative difference) between two successive evaluations of the log-likelihood  $l(\omega|Y)$  given by  $\|l(\omega^s|Y) - l(\omega^{s+1}|Y)\|$  or  $\left\| \frac{l(\omega^{s+1}|Y)}{l(\omega^s|Y)} - 1 \right\|$ , respectively. Furthermore, the Akaike information criteria (AIC) can be calculated from the final estimated log-likelihood to evaluate the model fit [10].

### 2.2.2. Estimation of the location of the change point

An estimator of the change point  $t$  is as follows [7]:

$$\hat{t}_{n,\tau} = \arg t \in (a, b) \cap (X_{n(2)}, X_{n(n-1)}) Q_{n,\tau}(\hat{\eta}_{n,\tau}(t), t), \quad (13)$$

where  $a$  and  $b$  are two constants such that  $t_\tau$  is thought to be in the interval  $(a, b)$ , usually determined graphically, and  $X_{n(2)}$  and  $X_{n(n-1)}$  are the second and  $(n-1)$ th order statistics of  $X_{ij}^s$ , respectively. Then  $\hat{\theta}_n(\tau)$  is obtained,  $\hat{\eta}_{n,\tau}(\hat{t}_{n,\tau})$ , and  $\hat{t}_{n,\tau}$ . The optimization method that is used to minimize  $Q_{n,\tau}(\hat{\eta}_{n,\tau}, t)$  is a combination of golden section search and successive parabolic interpolation implemented by the function “optimize” in the R package as suggested by Li et al. [7].

### 2.2.3. Standard errors and confidence intervals

There are studies on asymptotic theory for quantile regression in the literature, however, it has been difficult to come up with practical inference techniques [16]. Moreover, one of the main disadvantages of the SAEM algorithm is that the estimated standard errors are not automatically calculated. Hence, several methods have been proposed to handle this problem. For instance, the rank score test [9] and the block bootstrap method [24, 25]. The use of the bootstrap method with quantile regression is very common.

The block bootstrap method to construct the confidence intervals for model parameters is adopted in this article. When the mistakes in a model or the data are correlated, the block bootstrap is used. A simple case or residual resampling will not work in this

situation since the correlation in the data cannot be replicated. By resampling within data blocks, the block bootstrap attempts to replicate the correlation. The block bootstrap has primarily been employed with time-series data, but it can also be utilized with data that are correlated in space or between groups [26]. The number of bootstrap replications is chosen according to the suggestion of Andrews and Buchinsky [27].

## 3. Simulation study

The aim of this simulation study was 2-fold. First, to assess the performance of the proposed techniques and to compare its performance with the method of Li et al. [7], when the errors follow a symmetric distribution. Second, to test the performance of the proposed techniques when the errors follow a skewed distribution.

We consider the following linear mixed change point model:

$$y_{ij} = \gamma_{0,\tau} + (\beta_{1,\tau}I\{x_{ij} \leq t_\tau\} + \beta_{2,\tau}I\{x_{ij} \geq t_\tau\})(x_{ij} - t_\tau) + s_{ij}^T \gamma_\tau + z_{ij}^T U_i + \varepsilon_{\tau,ij}, \quad (14)$$

for  $j = 1, 2, 3, 4$ ,  $i = 1, \dots, n$ .

The goal was to estimate the fixed effects parameters  $\beta_\tau$  and  $\gamma_\tau$ , and the location of the change point for a grid of percentiles  $p = \{0.25, 0.50, 0.75\}$ .

### 3.1. Simulation setting

We use time points ( $m$ ) = 4. Different sample sizes ( $n$ ) have been chosen, however, for the sake of parsimony and because the results are similar, we report the results of  $n = 40$  and 100. The variable  $x_{ij}$  is simulated from the normal distribution with a mean of 5 and a standard deviation of 2. The design matrix  $s_{ij}$ , which is associated with fixed effects  $\gamma_\tau$  consists of three columns. The first column represents a binary variable that describes the treatment group to which the subjects belong. The second column represents a uniform (0,1) variable. The third column represents a normal variable with a mean of 3 and a standard deviation of 1. The matrix  $z_{ij}$ , that is associated with the random effects, is a subset of the matrix  $s_{ij}$  with three columns that are the intercepts and column numbers 2 and 3 in the matrix  $s_{ij}$ .

The fixed effects parameters were chosen as  $\gamma_0 = 5.5$ ,  $\gamma_1 = 4$ ,  $\gamma_2 = 2$ , and  $\gamma_3 = 3$ . The location of the change point is fixed at 4.8,  $\beta_{1,\tau} = 4$ , and  $\beta_{2,\tau} = -5$ . The error terms  $\varepsilon_{\tau,ij}$  are generated independently from two different distributions as follows:

1. An ALD ( $0, \sigma, p$ ), where  $p$  stands for the respective percentile to be estimated and  $\sigma = 1$ . This represents a symmetric distribution.
2. A lognormal normal distribution with a mean of 0 and a variance  $\sigma^2 = 1$ . This represents a skewed distribution.

The random effects  $U_i$  is a  $3 \times 1$  vector that is generated from a multivariate-asymmetric Laplace distribution with a mean of 0 and a variance-covariance matrix  $\Sigma$ . The matrix  $\Sigma$  is chosen to follow

AR(1) structure with  $\rho = 0.5$  and  $\sigma = 0.8$ . This structure gives a better acceptance rate for the Metropolis–Hasting algorithm.

We use  $l = 20$  (number of simulations),  $W = 500$  (number of Monte Carlo iterations), and  $c = 0.2$ . Note that the choice of  $c$  depends on the dataset and the underlying model. We generate 1,000 data samples for each scenario.

### 3.2. Simulation results

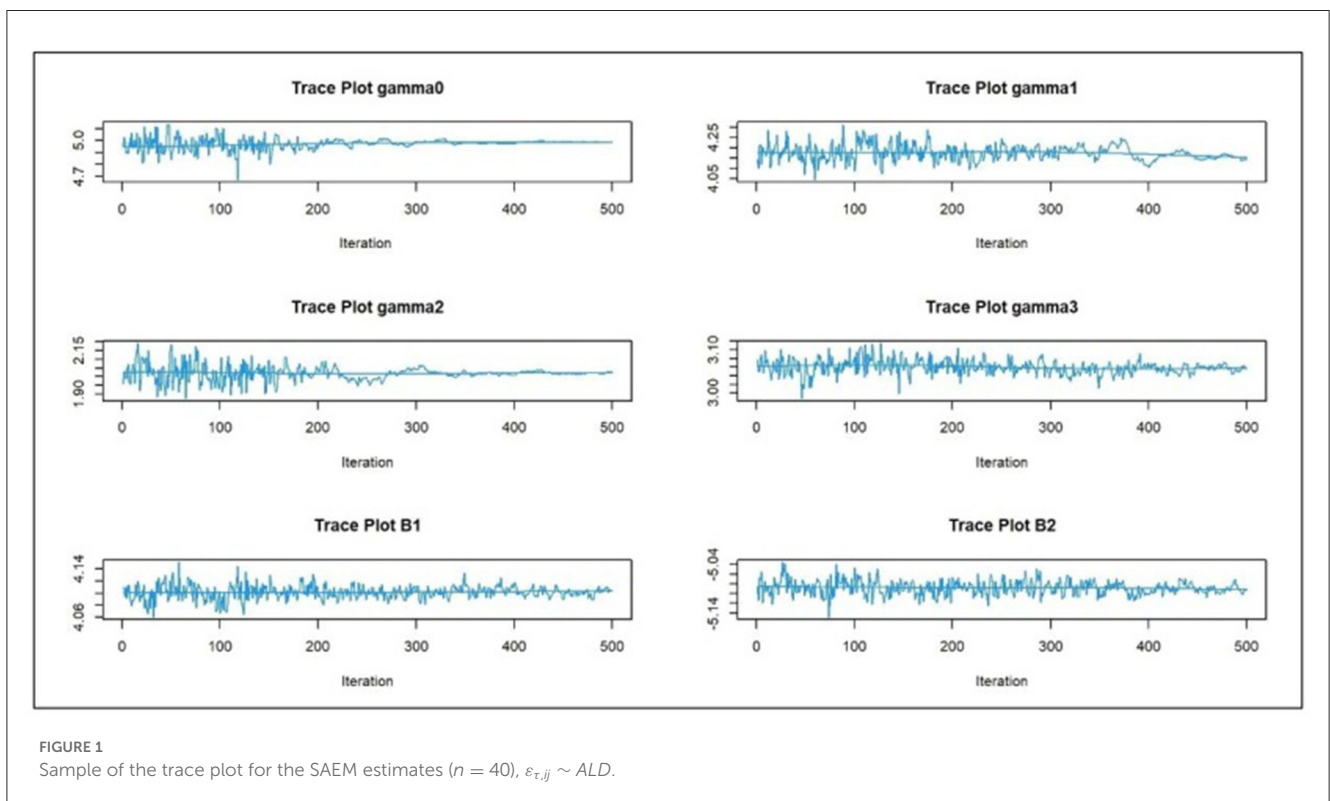
The convergence of the SAEM estimates is evaluated using the visual inspection *via* trace plot. Moreover, through monitoring the difference (relative difference) between two successive evaluations of the log-likelihood  $l(\omega|Y)$  given by  $\|l(\omega^s|Y) - l(\omega^{s+1}|Y)\|$  or  $\left\| \frac{l(\omega^{s+1}|Y)}{l(\omega^s|Y)} - 1 \right\|$ , respectively.

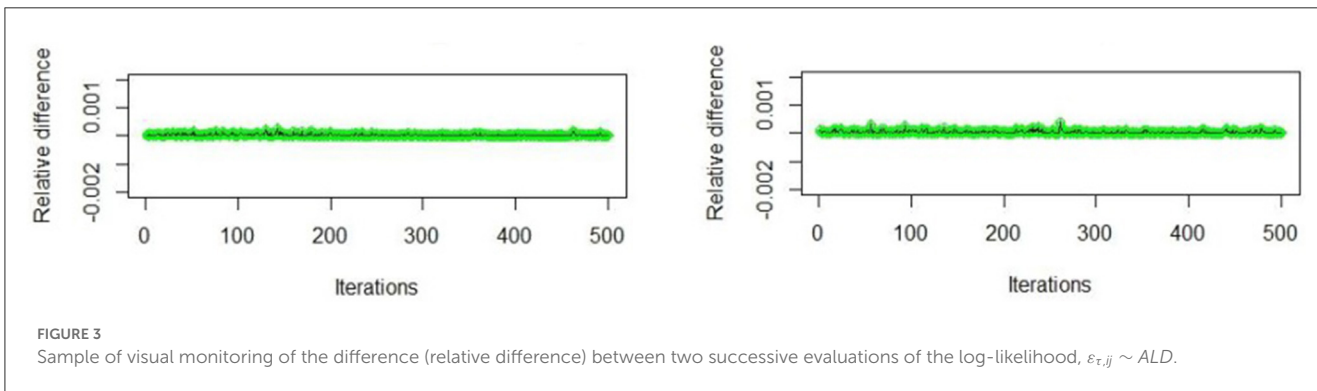
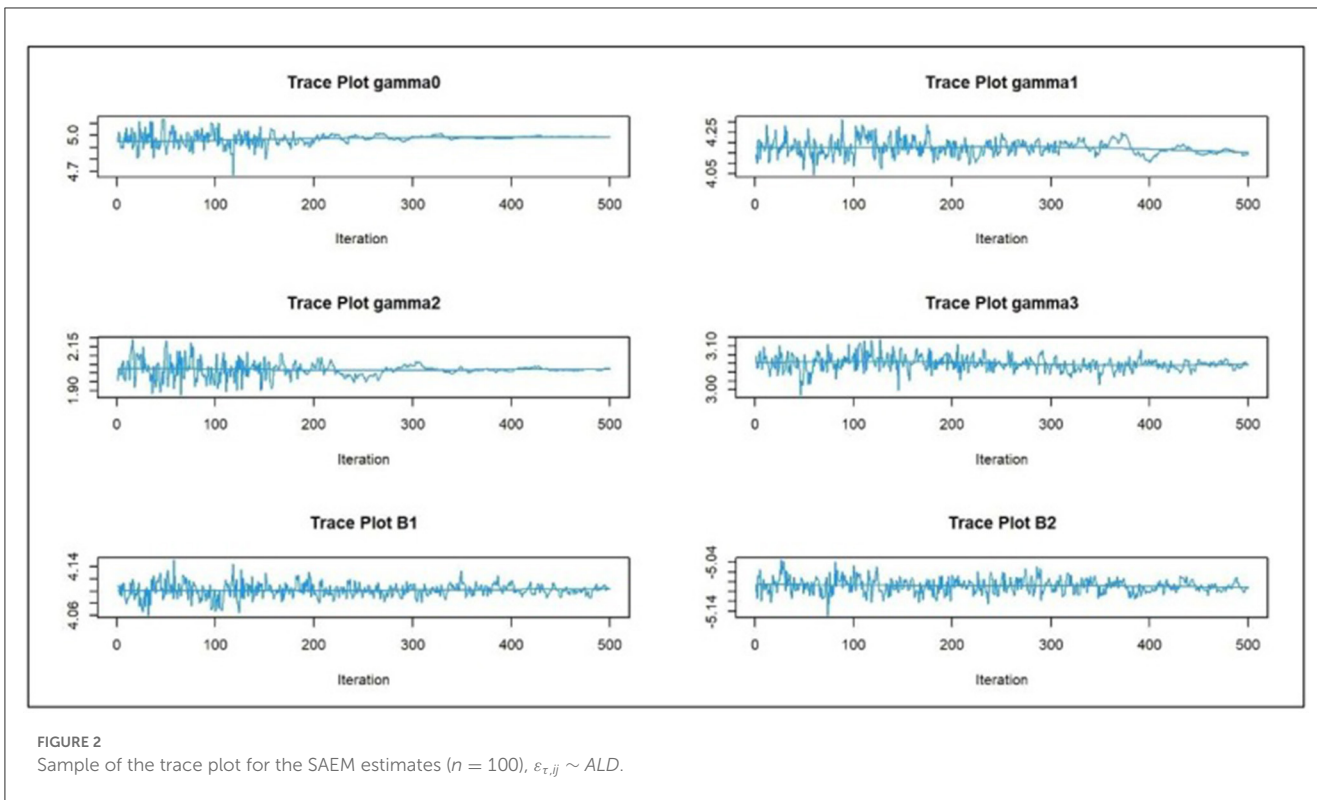
Figures 1, 2 display samples from the trace plots of the estimates for  $n = 40$  and  $n = 100$ , respectively. The trace plot indicates whether the estimates converge to the stationary distribution or not. The convergence is slow if the sequence has a long-term trend [28]. Figure 3 displays the visual monitoring of the relative difference between two successive evaluations of the log-likelihoods. Assume that the first 100 iterations (which is 20% of  $W$ ), as burn-in period, it is clear that all the estimates converge.

For all scenarios, the absolute relative bias (ARB) for each parameter over the 1,000 replicates is obtained as  $ARB = \left| \frac{\text{estimated value} - \text{true value}}{\text{true value}} \right|$ , and the standard error for each estimator is obtained as  $SE(\hat{\theta}) = \sqrt{\frac{1}{K} \sum_{k=1}^{1,000} (\hat{\theta}_k - \hat{\theta}^*)^2}$ , where  $\hat{\theta}_k$  is the  $K$ th estimate of  $\theta$  using the  $K$ th sample and  $\hat{\theta}^*$  is the average of the multiple estimates, that is,  $\hat{\theta}^* = \frac{\sum_{k=1}^k \hat{\theta}_k}{k}$ .

Tables 1, 2 show the results of the simulations. We compare the performance of the proposed algorithm with the algorithm proposed by Li et al. [7]. We can conclude the following:

1. The proposed estimators are asymptotically unbiased for symmetric and skewed distributions. This is because the relative absolute bias of all estimates using the proposed algorithm is relatively small, as all values of ARB are  $<0.2\%$  of the value of the parameters, for both an ALD and the lognormal distributions. The result is valid for sample sizes  $n = 40$  and  $100$ .
2. It is clear that the ARB associated with most of the parameter estimates, for the proposed algorithm, is less than those of the algorithm proposed by Li et al. [7], when the errors follow an ALD or the lognormal distribution and sample sizes  $n = 40$  and  $100$ . This means that the proposed method is better than the algorithm proposed by Li et al. [7].
3. The proposed method is more efficient than the algorithm proposed by Li et al. [7]. The standard errors of most of the estimators for the proposed algorithm are less than their counterparts of the algorithm proposed by Li et al. [7].
4. All the AIC values for the proposed technique are less than that for the model estimated by Li et al. [7]. This means that the proposed algorithm outperforms the algorithm proposed by Li et al. [7].
5. According to the median absolute deviation values, the same conclusion can be derived as the MAD values for the proposed algorithm are less than that of the Li et al. [7] algorithm.





### 4. Application: COVID-19

The first diagnosed case with COVID-19 symptoms is dated back to December 2019 in Wuhan city, China. COVID-19 affected the entire world, socially, economically, and even politically. There were more than 3 million deaths by the end of April 2021 as reported by the WHO. The factors that affect the spread of the disease, and the number of deaths, were the focus of many studies.

Studies have shown that environmental factors, in general, may affect the fast spread of COVID-19. Moreover, studies tried to investigate the impact of economic factors on virus transmission. A study on some Chinese cities during the period 19 January and 29 February of the year 2020 revealed that higher developed cities have high transmission rates. This may be due to the high economic activity that needs high social interactions [29].

The effect of demographic factors on the spread of COVID-19 was studied. Khan et al. [30] demonstrate that certain demographic attributes, such as the age distribution, the poverty ratio, the female smoker percentage, the obesity level, and the average annual temperature of the country, are significantly associated with COVID-19 death rate distribution.

In this article, we focus on global economic factors and health factors affecting the monthly death rate per million of COVID-19 to describe the spread and fatality of COVID-19 disease. We investigate whether there is a threshold effect (change point) in the relationship between the HDI and new monthly deaths per million. From Figure 4, we can conclude that the new deaths per million are skewed to the right. Thus, we need to focus on the factors that affect the lower and/or upper tails of the distribution of new deaths. However, most of the proposed statistical methodologies for describing longitudinal data with change points rely upon distributional assumptions that are not held in this case.

TABLE 1 Simulation results, the standard errors, and the relative bias for different models at  $n = 40$ , and  $100$ ,  $\epsilon_{\tau,ij} \sim ALD$ .

		$n = 40, c = 0.2$				$n = 100, c = 0.2$			
		Proposed method		Li et al. [7]		Method		Li et al. [7]	
Distribution	Parameter	Relative bias%	S.E	Relative bias%	S.E	Relative bias%	S.E	Relative bias%	S.E
$\tau = 0.25$	$\gamma_0$	0.077	0.043	0.142	0.052	0.084	0.053	0.115	0.115
	$\gamma_1$	0.053	0.027	0.041	0.026	0.032	0.039	0.036	0.036
	$\gamma_2$	0.181	0.035	0.274	0.043	0.130	0.044	0.058	0.058
	$\gamma_3$	0.024	0.011	0.086	0.012	0.003	0.010	0.000	0.000
	$B_1$	0.008	0.008	0.005	0.010	0.012	0.012	0.022	0.022
	$B_2$	0.040	0.009	0.045	0.009	0.016	0.012	0.030	0.030
	Change point	0.093	0.022	0.089	0.025	0.015	0.029	0.026	0.026
	$\sigma$	0.322	0.019			0.375	0.024		
AIC		11,603.900		11,604.880		29,134.340		29,136.100	
Median absolute deviation		1.43		3.19		2.55		3.76	
$\tau = 0.5$	$\gamma_0$	0.121	0.165	0.142	0.188	0.078	0.043	0.122	0.052
	$\gamma_1$	0.299	0.117	0.041	0.146	0.028	0.027	0.058	0.026
	$\gamma_2$	0.389	0.125	0.274	0.177	0.065	0.035	0.032	0.043
	$\gamma_3$	0.120	0.045	0.086	0.056	0.021	0.011	0.025	0.012
	$B_1$	0.058	0.029	0.005	0.039	0.025	0.008	0.031	0.010
	$B_2$	0.003	0.048	0.045	0.054	0.021	0.009	0.023	0.009
	Change point	0.006	0.088	0.089	0.110	0.006	0.022	0.015	0.025
	$\sigma$	0.140	0.075			0.124	0.019		
AIC		11,619.340		11,620.040		29,065.320		29,066.240	
Median absolute deviation		1.57		2.08		1.44		2.22	
$\tau = 0.75$	$\gamma_0$	0.008	0.152	0.045	0.167	0.103	0.070	0.104	0.083
	$\gamma_1$	0.016	0.151	0.139	0.176	0.043	0.039	0.030	0.041
	$\gamma_2$	0.039	0.101	0.395	0.113	0.037	0.043	0.117	0.051
	$\gamma_3$	0.005	0.033	0.007	0.039	0.113	0.014	0.134	0.015
	$B_1$	0.033	0.035	0.037	0.040	0.019	0.014	0.010	0.019
	$B_2$	0.064	0.039	0.105	0.048	0.007	0.013	0.013	0.014
	Change point	0.011	0.085	0.017	0.097	0.000	0.032	0.000	0.000
	$\sigma$	0.354	0.074			0.398	0.026		
AIC		11,663.790		11,664.600		29,164.130		29,166.860	
Median absolute deviation		2.28		2.33		2.46		3.97	

### 4.1. Data

The used data about the COVID-19 new deaths were obtained from the Our World In Data, an online (<https://ourworldindata.org/coronavirus>) publication. This cite has become one of the world’s leading websites during the pandemic in 2020. The study focuses on the monthly data during the period starting from 1 April 2020 to 30 April 2021. The dependent variable is the new monthly deaths per million while the independent variables are as follows:

- 1- ICU: the monthly number of patients with COVID-19 in intensive care units (ICUs) on a given day per million.
- 2- The number of tests: the monthly tests conducted per new confirmed case of COVID-19.
- 3- Diabetes prevalence: Diabetes prevalence (% of the population aged 20–79) in 2017.
- 4- Hospital beds: The number of hospital beds per 1,000 people, most recent year available since 2010.
- 5- Median age: The median age of the population; UN projection for 2020.



TABLE 2 Simulation results, the standard errors, and the relative bias for different models at  $n = 40$ , and  $100$ ,  $\varepsilon_{\tau,ij} \sim \text{Lognormal}$ .

Distribution	Parameter	$n = 40, c = 0.2$				$n = 100, c = 0.2$			
		Proposed method		Li et al. method		Proposed method		Li et al. method	
		Relative bias%	S.E	Relative bias%	S.E	Relative bias%	S.E	Relative bias%	S.E
$\tau = 0.25$	$\gamma_0$	0.092	0.141	0.106	0.187	0.017	0.055	0.101	0.082
	$\gamma_1$	0.167	0.083	0.198	0.103	0.102	0.058	0.089	0.064
	$\gamma_2$	0.127	0.054	0.241	0.079	0.042	0.036	0.015	0.054
	$\gamma_3$	0.043	0.028	0.041	0.035	0.098	0.009	0.151	0.012
	$B1$	0.004	0.025	0.051	0.035	0.0007	0.011	0.035	0.022
	$B2$	0.002	0.013	0.023	0.017	0.0001	0.009	0.006	0.018
	<i>Change-point</i>	0.0748	0.057	0.0741	0.076	0.0362	0.030	0.036	0.042
	$\sigma$	0.363	0.043			0.0811	0.025		
AIC		11,647.47		11,650.32		29,150.83		29,150.87	
Median absolute deviation		1.77		2.99		1.14		3.01	
$\tau = 0.5$	$\gamma_0$	0.204	0.131	0.267	0.137	0.029	0.060	0.033	0.070
	$\gamma_1$	0.017	0.074	0.025	0.063	0.107	0.047	0.090	0.055
	$\gamma_2$	0.078	0.071	0.126	0.103	0.009	0.025	0.157	0.039
	$\gamma_3$	0.008	0.025	0.022	0.029	0.004	0.011	0.014	0.015
	$B1$	0.001	0.022	0.0006	0.032	0.003	0.013	0.003	0.016
	$B2$	0.012	0.010	0.025	0.015	0.005	0.008	0.004	0.011
	<i>Change-point</i>	0.077	0.056	0.072	0.063	0.038	0.027	0.036	0.034
	$\sigma$	0.1054	0.043			0.249	0.022		
AIC		11,663.5		11,665.6		29,124.52		29,125.03	
Median absolute deviation		1.07		2.206		0.93		2.63	
$\tau = 0.75$	$\gamma_0$	0.044	0.177	0.051	0.198	0.049	0.066	0.046	0.096
	$\gamma_1$	0.039	0.107	0.045	0.107	0.153	0.052	0.174	0.062
	$\gamma_2$	0.186	0.058	0.164	0.081	0.002	0.053	0.094	0.075
	$\gamma_3$	0.061	0.02	0.072	0.033	0.143	0.014	0.121	0.021
	$B1$	0.008	0.029	0.014	0.030	0.011	0.014	0.004	0.018
	$B2$	0.007	0.018	0.014	0.242	0.005	0.011	0.002	0.016
	<i>Change-point</i>	0.024	0.068	0.026	0.115	0.036	0.035	0.042	0.048
	$\sigma$	0.011	0.050			0.103	0.030		
AIC		11,653.44		11,654.07		29,130.44		29,139.75	
Median absolute deviation		2.18		2.509		2.275		2.95	

6- Stringency index (SI): The government response stringency index; it is a composite measure based on nine response indicators, including school closures, workplace closures, and travel bans, rescaled to a value from 0 to 100 (100 is the strictest response).

7- HDI: A composite index measuring average achievement in three basic dimensions of human development; a long and healthy life, knowledge, and a decent standard of living values for 2019. It ranges from 0 to 1 in which 1 indicates higher development countries in the three aspects.

The 30 countries incorporated in the study are as follows:

- |                |                   |                 |                 |                  |
|----------------|-------------------|-----------------|-----------------|------------------|
| <i>Austria</i> | <i>Bangladesh</i> | <i>Brazil</i>   | <i>Bulgaria</i> | <i>Canada</i>    |
| <i>China</i>   | <i>Colombia</i>   | <i>Denmark</i>  | <i>Djibouti</i> | <i>Dominican</i> |
| <i>Egypt</i>   | <i>Ethiopia</i>   | <i>Ghana</i>    | <i>Greece</i>   | <i>Hungary</i>   |
| <i>India</i>   | <i>Iraq</i>       | <i>Ireland</i>  | <i>Italy</i>    | <i>Jordan</i>    |
| <i>Kenya</i>   | <i>Morocco</i>    | <i>Portugal</i> | <i>Russia</i>   | <i>Spain</i>     |
| <i>Sweden</i>  | <i>Tunisia</i>    | <i>Turkey</i>   | <i>UK</i>       | <i>USA</i>       |

These countries are chosen to represent different areas of the world according to the availability of the data. It is well mentioned

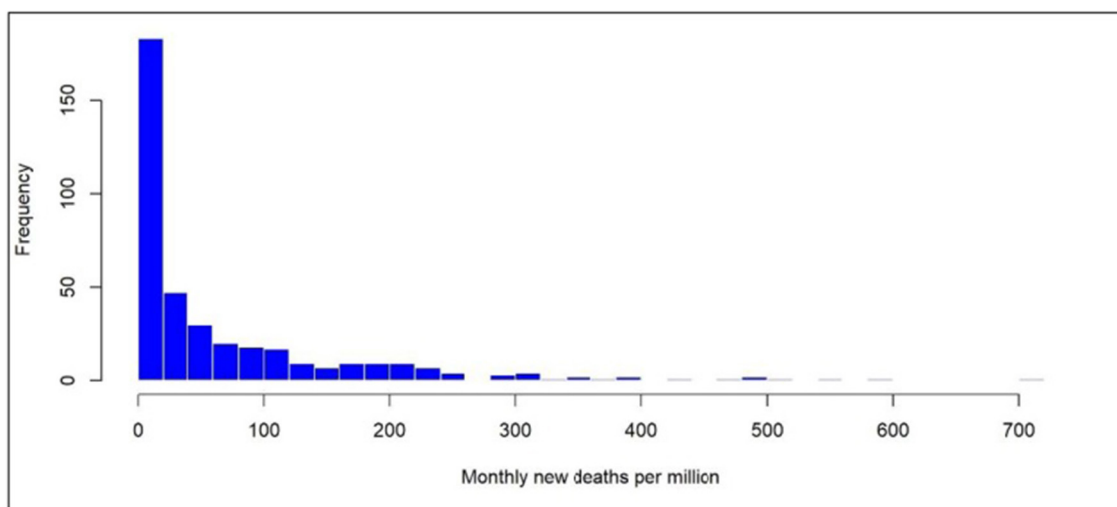


FIGURE 4 Histogram of the monthly new deaths per million.

TABLE 3 Descriptive statistics of COVID-19 data.

	Minimum	Maximum	Mean	Std. deviation
New monthly deaths per million	0	704	72	107
ICU	1	5,882	1,018	1,217
Number of tests	83	33,120	1,906	3,009
Diabetes prevalence	3	17	7	3
Hospital beds	0	8	3	2
Median age	20	48	35	9
SI	25	99	66	15
HDI	0.485	0.955	0.793	0.128

that there are some countries suffering from missing data, especially in the variables monthly number of patients with COVID-19 in intensive care units (ICUs), and monthly tests conducted per newly confirmed cases of COVID-19. These values are imputed using the regression method. The variables that will be associated with the random effects are stringency index (SI), diabetes prevalence, and median age in addition to the intercept.

Table 3 presents the descriptive statistics of the variables in the study. Figure 5 presents the profile plot of the death rate for all countries over time, from it we can conclude that death distribution for all countries is skewed.

## 4.2. Data analysis and results

The aim of the data was to study the global economic and health factors that affect the monthly death rate due to COVID-19 from

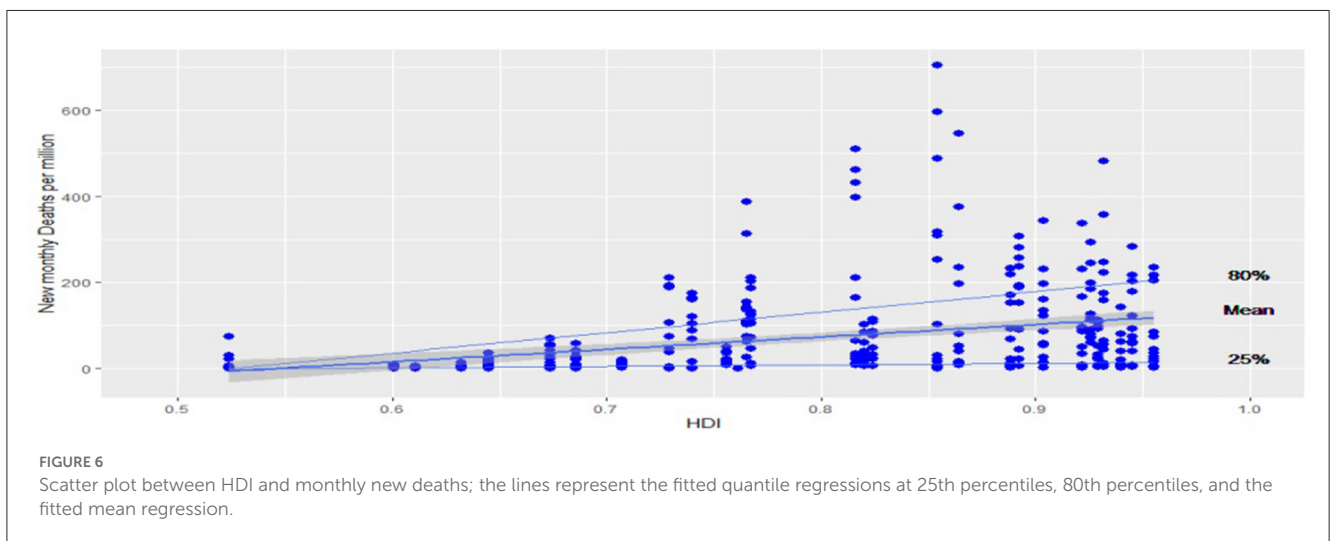
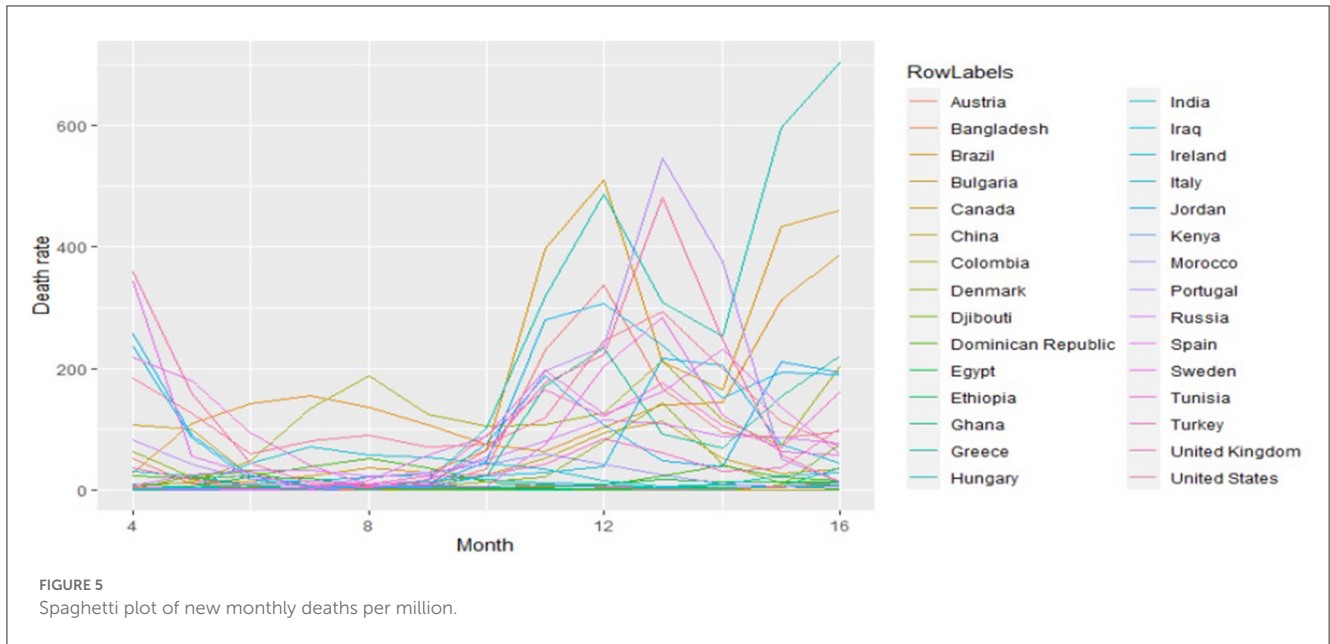
1 April 2020 to 30 April 2021. As reported by the WHO, there are more than 3 million deaths due to COVID-19 by the end of April 2021. Many studies tried to understand the factors behind the spread of the disease and the factors that affect the number of deaths throughout the world. This enables us to minimize the losses faced due to COVID-19. Different factors are considered, in this article, including a composite index measuring average achievement in three basic dimensions of human development: a long and healthy life, knowledge, and a decent standard of living (HDI). The HDI ranges from 0 to 1, where 1 indicates higher human development in three dimensions.

Figure 6 shows a scatter plot of the relationship between the HDI and the new deaths per million. From Figure 6, we can see that there is a suspicion of having a threshold effect on the relationship between the HDI and the new monthly deaths per million. This is obvious, especially for mean regression, 25th percentile regression, and 80th percentile regression. Hence, the location of the change point may be at HDI between 0.7 and 0.8. This leads us to estimate the mixed effects quantile regression with a change point at 0.5th quantile and at 0.8th quantile.

The proposed method was implemented at the quantiles  $\tau = 0.5$  and  $\tau = 0.8$  to test whether the HDI had any threshold effect on the monthly new deaths per million. The following model is fitted:

$$MDR = \gamma_{0,\tau} + (\beta_{1,\tau}I\{x_{ij} \leq t_\tau\} + \beta_{2,\tau}I\{x_{ij} \geq t_\tau\})(x_{ij} - t_\tau) + s_{ij}^T \gamma_\tau + z_{ij}^T U_i + \varepsilon_{\tau,ij}, j = 1, 2, 3, \dots, 13; i = 1, \dots, 30, (15)$$

Where  $x_{ij}$  is the HDI,  $s_{ij}^T$  is a vector containing the observation of the following variables: ICU, number of tests, diabetes prevalence, hospital beds, median age, and stringency index (SI). The  $z_{ij}^T$  is a vector containing ones to represent the intercept, stringency index (SI), diabetes prevalence, and median age. The estimates of the fixed effect of the model and the location of the change point at  $\tau = 0.5$ , using the proposed algorithm and the



algorithm of Li et. al. [7] are summarized in Table 4. Moreover, the 95% confidence interval is obtained using the block bootstrap method with the number of replications equal to 1,000. From Table 4, depending on the confidence intervals, it is clear that there is a threshold effect in the relationship between the HDI and the new monthly deaths per million at  $HDI = 0.79$  and  $0.816$ , respectively. The effect of the HDI on the 0.5th quantile of new monthly deaths per million due to COVID-19 is 459.41 where this effect declines to 128.68 after the threshold is reached. Both effects are significant at a 95% confidence level. But for the 0.8th quantile we found that, prior to the threshold value, the effect of the HDI on the 0.8th quantile of new monthly deaths per million due to COVID-19 is 1414.293. This effect declines and has a negative value (-122.95) after the location of the threshold. This means that increasing the HDI affects the 0.5th quantile of new monthly deaths per million due to COVID-19 positively,

before and after the threshold value. However, it affects the 0.8th quantile of new monthly deaths per million due to COVID-19 before the positive threshold value and after the negative threshold value.

Furthermore, we can see that there is a positive effect of the monthly number of patients with COVID-19 in intensive care units (ICUs) for both 0.5th and 0.8th quantiles of new monthly deaths per million. This effect is greater in the case of the 0.8th quantile. There is a negative effect of each of the number of tests, and the median age of the population on both 0.5th and 0.8th quantiles of new monthly deaths per million. However, their effect is greater in the case of the 0.8th quantile. Each of the stringency index, hospital beds, and diabetes prevalence has no significant effect on both 0.5th and 0.8th quantiles of new monthly deaths per million due to COVID-19.

TABLE 4 Estimation of the mixed effects quantile change point regression model for COVID-19 data.

	Coefficient	tau = 0.5			tau = 0.8		
		Estimate	95% CI		Estimate	95% CI	
Intercept	$\gamma_0$	-346.03	-358.571	-284.866	-842.653	-855.085	-828.045
ICU	$\gamma_1$	0.0302	0.026	0.031	0.105	0.100904	0.113117
Number of tests	$\gamma_2$	-0.0062	-0.0066	-0.004	-0.007	-0.00767	-0.00511
SI	$\gamma_3$	0.2042	-0.138	0.261	0.462	-0.10951	0.567555
Diabetes Prevalence	$\gamma_4$	-0.0943	-1.795	0.868	-1.842	-2.24477	1.615969
Hospital beds	$\gamma_5$	-0.3533	-4.566	0.453	0.811	-4.10941	3.664168
Median age	$\gamma_6$	-1.2929	-2.035	-0.848	-5.640	-6.26093	-4.95497
HDI	$B_1$	459.41	438.673	468.887	1,414.293	1,408.687	1,419.164
	$B_2$	128.68	128.622	128.677	-122.954	-128.65	-116.955
	Change point	0.79	0.783	0.842	0.816	0.813	0.817

TABLE 5 AIC and median absolute deviation values of the two models for COVID-19 data.

		Proposed method	Li et al. [7] method
AIC		24,679.40	24,690.74
	tau = 0.5 tau = 0.8	11,109.59	11,121.28
Median absolute deviation	tau = 0.5	28.5	31.8
	tau = 0.8	40.4	42.9

Table 5 presents the AIC values for the proposed model and the Li et al. [7] model at  $\tau = 0.5$  and  $0.8$ . The results show that the proposed model performs better than the model of Li et al. [7].

## 5. Conclusion

In this article, we propose the mixed effects quantile regression with a change point model for longitudinal data by relaxing the independence assumption. The mixed effects are used to capture the dependence structure of the longitudinal data. We use the stochastic approximation EM algorithm to estimate the parameters using the link between an asymptotic Laplace distribution and the quantile regression. In addition, the location of the change point is estimated using optimization methods. Simulation studies are conducted to evaluate the proposed techniques. The simulation results show that the proposed techniques are better than those of Li et al. [7] in terms of the relative biases and the standard errors for symmetric and skewed distributions.

The proposed techniques are applied to a real dataset about COVID-19. We found that there is a threshold effect in the relationship between the HDI and the 0.5th and 0.8th quantiles of new monthly deaths per million. Moreover, the results show a positive effect on the monthly number of patients with COVID-19 in intensive care units (ICUs) for both 0.5th and 0.8th quantiles

of new monthly deaths per million. There is a negative effect of each of the number of COVID-19 tests and the median age of the population on both 0.5th and 0.8th quantiles of new monthly deaths per million. The stringency index, hospital beds, and diabetes prevalence have no significant effect on both 0.5th and 0.8th quantiles of new monthly deaths per million.

The proposed techniques are for complete longitudinal data and one change point. A new venue for future research is to extend the proposed techniques for longitudinal data with missing values. Another future research point is to modify the proposed methods to accommodate multiple change points. These points are under consideration by the researchers.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://ourworldindata.org/coronavirus>.

## Author contributions

AG has introduced the idea of the research and review the results, also help in editing the manuscript. A-EA-R has reviewed the idea and help in the change point part and help in editing the article. WI has written the codes, tried the necessary modifications, and the simulation and application part. All authors contributed to the article and approved the submitted version.

## Acknowledgments

The authors would like to thank the editor and reviewers for their constructive and helpful comments that significantly improved the article.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships

that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may

be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fams.2023.1106958/full#supplementary-material>

## References

- Gad AM, Ibrahim WIM. An adaptive linear regression approach for modeling heavy-tailed longitudinal data. *Commun. Stat. Simulat. Computat.* (2020) 49:1181–97. doi: 10.1080/03610918.2018.1491990
- Yang L, Gao S. Bivariate random change point models for longitudinal outcomes. *Stat. Med.* (2012) 32:1038–53. doi: 10.1002/sim.5557
- Ghosh P, Vaida F. Random Changepoint modeling of HIV immunologic responses. *Stat. Med.* (2007) 26:2074–87. doi: 10.1002/sim.2671
- Mclain A, Albert P. Modeling longitudinal data with a random change point and no time-zero: applications to inference and prediction of the labor curve *Biometrics.* (2014) 70:1052–60. doi: 10.1111/biom.12218
- Xing H, Ying Z. A semiparametric change-point regression model for longitudinal data. *J Am Statistic Assoc.* (2012) 107:1625–37. doi: 10.1080/01621459.2012.712425
- Lai Y, Albert P. Identifying multiple change-points in a linear mixed effects model. *Stat Med.* (2014) 33:1015–28. doi: 10.1002/sim.5996
- Li C, Dowling N, Chappell R. Quantile regression with a change-point model for longitudinal data: an application to the study of cognitive changes in preclinical alzheimer's disease. *Biometrics.* (2015) 71:625–35. doi: 10.1111/biom.12313
- Jung S. Quasi-likelihood for median regression models. *Am Stat Assoc.* (1996) 91, 251–7. doi: 10.1080/01621459.1996.10476683
- Wang HJ, Fyngenson M. Inference for censored quantile regression. *Annal Stat.* (2009) 37:756–81. doi: 10.1214/07-AOS564
- Li C, Wei Y, Chappell R, He X. Bent line quantile regression with application to an allometric study of land mammals' speed and mass. *Biometrics.* (2011) 67:242–9. doi: 10.1111/j.1541-0420.2010.01436.x
- Sha N. *On Testing the Change-point in the Longitudinal Bent Line Quantile Regression Model*, Graduate school of Arts and sciences, Columbia University, Columbia (2011).
- Yu K, Moyeed R. Bayesian quantile regression. *Statistics Probabil Lett.* (2001) 54:437–47. doi: 10.1016/S0167-7152(01)00124-9
- Liu Y, Bottai M. Mixed-effects models for conditional quantiles with longitudinal data. *Int J Biostatistic.* (2009) 5:28. doi: 10.2202/1557-4679.1186
- Meza C, Osorio F, I RCruz D. Estimation in nonlinear mixed-effects models using heavy-tailed distributions. *Stat Comput.* (2012) 22:121–39. doi: 10.1007/s11222-010-9212-1
- Kotz S, Kozubowski TJ, Podgórski K. *The Laplace Distribution and Generalizations. A Revisit with Applications to Communications, Economics, Engineering, and Finance.* Basel: Birkhäuser (2001).
- Galarza CE, Lachos VH, Bandyopadhyay D. Quantile regression in linear mixed models: a stochastic approximation EM approach. *Stat Interface.* (2017) 10:471–82. doi: 10.4310/SII.2017.v10.n3.a10
- Abd Elwahab HA, Kholy RB, Gad AM. Sensitivity analysis index for shared parameter models in longitudinal studies. *Adv Appl Stat.* (2019) 17:1–20. doi: 10.17654/AS05701,0001
- Yaseen ASA, Gad AM. A stochastic variant of the EM algorithm to fit mixed (discrete and continuous) longitudinal data with nonignorable missingness. *Commun Stat Theory Meth.* (2020) 498:4446–67. doi: 10.1080/03610926.2019.1601223
- Buchinsky M. Estimating the asymptotic covariance matrix for quantile regression models: a monte carlo study. *J Econometr.* (1995) 8:303–38. doi: 10.1016/0304-4076(94)01652-G
- Lipsitz R, Fitzmaurice GM, Molenberghs G, Zhao LP. Quantile regression methods for longitudinal data with drop-outs: application to CD4 cell counts of patients infected with the human immunodeficiency virus. *Appl Stat.* (1997) 46:463–76. doi: 10.1111/1467-9876.00084
- Andrews DWK, Buchinsky M. A three-step method for choosing the number of bootstrap repetitions. *Econometrica.* (2000) 68:23–52. doi: 10.1111/1468-0262.00092
- Qiu Y, Chen X, Shi W. Impacts of social and economic factors on the transmission of coronavirus disease 2019. (COVID-19) in China. *J Populat Econ.* (2020) 33:1127–72. doi: 10.1007/s00148-020-00778-2
- Khan W, Hussain A, Khan SA, Al-Jumaily M, Nawaz R, Liatsis P. Analyzing the impact of global demographic characteristics over the COVID-19 spread using class rule mining and pattern matching. *Royal Soc Open Sci.* (2021) 8:21. doi: 10.1098/rsos.201823
- Yirga AA, Melesse SF, Mwambi HG, Ayele DG. Additive quantile mixed effects modeling with application to longitudinal CD4 count data. *Scientific Rep.* (2021) 11:25. doi: 10.1038/s41598-021-97114-9
- Tian Y, Wang L, Tang M, Tian M. Weighted composite quantile regression for longitudinal mixed effects models with application to AIDS studies. *Commun Stat Simulat Computat.* (2021) 50:1837–53. doi: 10.1080/03610918.2019.1610440
- Galarza CE, Castro LM, Louzada F, Lachos VH. Quantile regression for nonlinear mixed effects models: a likelihood based perspective. *Stat Pap.* (2020) 1:1281–307. doi: 10.1007/s00362-018-0988-y
- Abdelwahab AS, Gad AM, Abdrabou AS. A CUSUM test for change point in quantile regression for longitudinal data, communications in statistics. *Simulat Applcat.* (2022) 22:p. To appear. doi: 10.1080/03610918.2022.2112600
- Politis DN, Romano JP. *Wolf Subsampling M*, Stanford, USA: Springer (1999).
- Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J Royal Stat Soc.* (1977) 39:1–38. doi: 10.1111/j.2517-6161.1977.tb01600.x
- Sinharay S. Assessing convergence of the Markov chain Monte Carlo algorithms: A review. ETS Research Report Series, A review. *ETS Res Rep Series.* (2003) 1:1–52. doi: 10.1002/j.2333-8504.2003.tb01899.x