A Mixed Effects Changepoint Quantile Regression Model for Longitudinal Data

Ahmed Gad[1], Wafaa I. M. Ibrahim[2], and Abdelnaser S. Abdrabou[2]

[1]Cairo University
[2]Affiliation not available

August 29, 2022

Abstract

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, with one or more unknown transition points usually referred to as breakpoints or change points. The existence of such unknown point disturbs the sample characteristics, so the detection and estimation of these points is crucial. Most of the proposed statistical methods in literature, for detecting and estimating change points, assume distributional assumption that may not hold. A good alternative in this case is to use a robust approach which is the quantile regression model. There are trials in the literature to deal with quantile regression models with a change point. These trials ignore the within subject dependence of longitudinal data. In this paper we propose a mixed effect quantile regression model with a change point to account for dependence structure in the longitudinal data. Fixed effect parameters, in addition to the location of the change point, are estimated using profile estimation method. The stochastic approximation EM algorithm is proposed to estimate the fixed effect parameters exploiting the link between asymptotic Laplace distribution and the quantile regression. In addition, the location of the change point is estimated using the usual optimization methods. 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 a COVID-19 data. The data focus on the effect of global economic and health factors on the monthly death rate due to COVID-19 during from the 1st of April 2020 till the 31st of April 2021.

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, Egypt

**Corresponding author:* ahmed.gad@feps.edu.eg

Abstract

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, with one or more unknown transition points usually referred to as breakpoints or change points. The existence of such unknown point disturbs the sample characteristics, so the detection and estimation of these points is crucial. Most of the proposed statistical methods in literature, for detecting and estimating change points, assume distributional assumption that may not hold. A good alternative in this case is to use a robust approach which is the quantile regression model. There are trials in the literature to deal with quantile regression models with a change point. These trials ignore the within subject dependence of longitudinal data.

In this paper we propose a mixed effect quantile regression model with a change point to account for dependence structure in the longitudinal data. Fixed effect parameters, in addition to the location of the change point, are estimated using profile estimation method. The stochastic approximation EM algorithm is proposed to estimate the fixed effect parameters exploiting the link between asymptotic Laplace distribution and the quantile regression. In addition, the location of the change point is estimated using the usual optimization methods. 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 a COVID-19 data. The data focus on the effect of global economic and health factors on the monthly death rate due to COVID-19 during from the 1st of April 2020 till the 31st of April 2021.

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. The measurements of the same subject tend to be 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 is could be at one or more unknown transition points, usually referred to as breakpoints or changepoints. Such changepoints

are very 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 whole data. Change point models with one change point and two linear phases are most used specially in biological data [2]. Recently, there has been an increased interest of application of change point models to longitudinal data. In Bayesian framework, see for example, Ghosh and Vaida [3], Yang and Gao [2], McLain, and Albert [4]. Xing and Ying [5] propose a semi-parametric changepoint regression model for longitudinal data. Lai and Albert [6] propose 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 the normality. These distributional assumptions may not generally hold. On the other hand, in some applications, the relationship between the response and covariates at the tails, rather than the center of the distribution are of main interest [7]. The quantile regression model is a good alternative when the distributional assumptions are violated. The quantile regression does not require distributional assumption. The quantile regression fits the conditional quantiles of the response variable given a set of covariates. The main advantage of the 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. Also, the quantile regression is robust to outliers in the response variable. There are trials in literature to use the quantile regression or the median regression in longitudinal data context. Some of these trials employ the marginal models, see for example, Jung [8]; Wang & Fygenson [9]; and Gad & Ibrahim [1].

Li et al. [7]extend the quantile regression model with a change point, introduced by Li et al. [10], to accommodate longitudinal data. These models ignore the within subject dependence [7], [11]. Incorporating random effects in these models is a remedy to accommodate the within subject dependence. Yu and Moyeed [12] use the connection between the asymmetric Laplace distribution (ALD) and quantile regression model to incorporate random effects in the model. Liu and Bottai [13] develop a likelihood based inferential approach for estimating parameters of mixed effects quantile regression models. They assume an ALD for the errors and multivariate Laplace distribution (MLD) for the random effects. They use MCEM algorithm for estimation and inference.

There are different stochastic algorithms to estimate the parameters for mixed effect quantile models rather than MCEM, such as stochastic approximation EM (SAEM) algorithm [14]. The SEM algorithm approximates the E-step of the EM algorithm, by splitting the E-step, into a simulation step and an integration step. The SAEM algorithm has been proved to be more computationally efficient than the classical MCEM algorithm. This is because reusing of simulations from one cycle to the another within the smoothing phase of the algorithm. Meza, et al. [14] state that the SAEM algorithm converges in small number of simulations unlike the MCEM algorithm, which needs large number of simulations.

The aim of this article is to extend the quantile regression model with change point to accommodate the within subject dependence. This is tackled via incorporating random effects to the model. A likelihood based inferential approach is developed by assuming an ALD for the errors and a MLD distribution for the random effects. The advantage of using the multivariate Laplace distribution is to accommodate any possible outliers. Also, it can handle heavy-tailed distributions. A stochastic approximation EM (SAEM) algorithm is proposed to obtain the maximum likelihood estimates of the fixed effects.

The proposed techniques are evaluated using a simulation study. Also, the proposed techniques are illustrated by a real data. The aim of the data is to the study global economic and health factors that affect the monthly death rate due to COVID-19, during from the 1st of April 2020 till the 31st of April 2021. As reported by 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 the 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 to1, where 1 indicates higher human development in the three dimensions.

The rest of this article is organized as follows. Section 2 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 goodness of fit for the proposed model. The simulation results are presented in Section 3. The proposed model is applied to COVID-19 data in Section 4. We test if there is a threshold effect (change-point) in the relationship between the HDI and new monthly deaths per million. Finally, Section 5 is devoted to concluding remarks and future work.

2. The Proposed Model and Estimation Procedure

2.1. The model

Li et al. [7] extend the quantile regression model with a change point, introduced by Li et al [10], to accommodate longitudinal data. However, this model ignores the within subject dependence Sha [11] and Li et al. [7]. We suggest extending this model to account for the within subject dependence by adding random effects, to capture the dependence structure of longitudinal data. The proposed model can be written as:

$$y_{ij} = \alpha_\tau + \left(\beta_{1,\tau} I\{x_{ij} \le t_\tau\} + \beta_{2,\tau} I\{x_{ij} \ge t_\tau\}\right)(x_{ij} - t_\tau) + s_{ij}^T \gamma_\tau + z_{ij}^T U_i + \varepsilon_{\tau,ij}, \quad (1)$$

for $j = 1, \dots\dots, m_i$, $i = 1, \dots\dots, n$, where 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_τ and s_{ij} is a q-dimensional vector of linear covariates with constant slopes. Also, 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; the τ^{th} conditional quantile of $\varepsilon_{\tau,ij}$ given x_{ij} , z_{ij}, and s_{ij} is 0, and $\varepsilon_{\tau,ij}$ is assumed to be independently distributed as 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

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 $\theta_\tau = \left(\alpha_\tau, \beta_{1,\tau}, \beta_{2,\tau}, t_\tau, \gamma_\tau^T\right)^T$ and $\eta_\tau = \left(\alpha_\tau, \beta_{1,\tau}, \beta_{2,\tau}, \gamma_\tau^T\right)^T$.

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 [15]. The distribution of the random effects U_i in model (1) is assumed to be a symmetric multivariate Laplace distribution with zero-mean as

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

where Σ is a $P \times P$ covariance matrix, $\nu = \frac{2-P}{2}$ and $K_\nu(u)$ is the modified Bessel function of the third kind which is given by

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

2.2. Estimation and inference

The estimates $\widehat{\theta_n}(\tau)$ are obtained by minimizing the objective function:

$$Q_{n,\tau}(\theta) = \sum_{ij}\rho_\tau(y_{ij} - g\,(L_{ij};\theta\,)),$$

where $\rho_\tau(u) = u\big(\tau - I(u \le 0)\big)$ is the quantile loss function, and $L_{ij} = \left(1, x_{ij}, z_{ij}^T, s_{ij}^T\right)^T$. However, due to the presence of change point, the objective function $Q_{n,\tau}(\theta)$ is non-convex [7]. Hence, the estimates $\widehat{\theta_n}(\tau)$ are obtained via profile estimation. The stochastic approximation EM algorithm is used to estimate the fixed effect parameters η_τ. The estimates can be obtained using the link between ALD and the quantile regression. In addition, the location of the change point is estimated using the optimization methods.

At fixed t, the profile estimator η given by

$$\widehat{\eta_{n,\tau}}(t) = \arg\min_{\eta\in R^{3+q}} Q_{n,\tau}(\eta, t). \tag{2}$$

We propose the stochastic approximation EM algorithm (SAEM) to estimate $\eta_\tau(t)$. An estimator of the change point *t* is given by

$$\widehat{t_{n,\tau}} = \arg\min_{t\in(a,b)\cap\left(X_{n(2)},\ X_{n(n-1)}\right)} Q_{n,\tau}\left(\widehat{\eta_{n,\tau}}(t), t\right), \tag{3}$$

where *a* and *b* are two constants such that t_τ is thought to be in the interval *(a, b)*, usually determined graphically, and $X_{n(2)}$ and $X_{n(n-1)}$ are the 2nd and (n−1)th order statistics of $X_{ij}'s$, respectively. Then $\widehat{\theta_n}(\tau)$ is obtained $\widehat{\eta_{n,\tau}}\left(\widehat{t_{n,\tau}}\right)$, and $\widehat{t_{n,\tau}}$. The optimization method that is used to minimize $Q_{n,\tau}(\widehat{\eta_{n,\tau}}(t), t)$ is a combination of golden section search and successive parabolic interpolation, implemented by the function "*optimize*" in R package as suggested by Li et al. [7].

2.2.1. The SAEM algorithm to estimate $\boldsymbol{\eta_\tau(t)}$

Galarza, et al. [16] use the SAEM algorithm to develop a likelihood-based approach to fit the quantile regression model, for continuous longitudinal data, using ALD distribution. They assume that the distribution for the random effects is multivariate Gaussian. In this article, we assume that the random effects follow an ALD. At fixed *t*, minimizing the loss function in Eq. (2) is the same as maximizing the ALD likelihood function. Likelihood based inferential approach is developed to estimate $\widehat{\eta_{n,\tau}}(t)$ in Eq. (2) by using the connection between the ALD distribution and the quantile regression [12]. This is done by assuming an ALD for the errors and a multivariate distribution for the random effects. At fixed *t*, the conditional the density function of $y_{ij}|U_i$ can be written as

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

where

$$\mu_{ij} = \alpha_\tau + \left(\beta_{1,\tau} I\{x_{ij} \le t_\tau\} + \beta_{2,\tau} I\{x_{ij} \ge t_\tau\}\right)\left(x_{ij} - t_\tau\right) + s_{ij}^T \gamma_\tau + z_{ij}^T U_i$$

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

Let

$$f(Y_i|U_i, X_i, S_i; \eta_\tau(t), \sigma) = \prod_{j=1}^{m_i} f\left(y_{ij}\middle|U_i, x_{ij}, s_{ij}; \eta_\tau(t), \sigma\right),$$

be the density for the i^{th} subject conditional on the random effect U_i, where

$$Y_i = [y_{i1} \quad y_{i2} \ldots . \quad y_{im_i}]^T, X_i = [x_{i1} \quad x_{i2} \ldots . \quad x_{im_i}]^T,$$

and

$$S_i = [s_{i1} \quad s_{i2} \ldots . \quad s_{im_i}]^T.$$

The complete data density of (Y_i, U_i), for $i = 1, 2, \ldots \ldots \ldots, m_i$, is then given by

$$f(Y_i, U_i|\, X_i, S_i; \omega) = f(Y_i|U_i, X_i, S_i; \eta_\tau(t), \sigma).\ f(U_i|\, X_i, S_i; \Sigma).$$

$$f(Y_i, U_i|\, X_i, S_i; \omega) = f(Y_i|U_i, X_i, S_i; \eta_\tau(t), \sigma).f(U_i|\, \Sigma).$$

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 $Y = (Y_1, Y_2, \ldots \ldots Y_n)$, $X = (X_1, X_2, \ldots \ldots X_n)$, $S = (S_1, S_2, \ldots \ldots S_n)$, and $U = (U_1, U_2, \ldots \ldots U_n)$, the joint density of (Y, U) based on the *n* subjects is given by

$$f(Y, U|\, X, S; \omega) = \prod_{i=1}^{n} f(Y_i|U_i, X_i, S_i; \eta_\tau(t), \sigma).f(U_i|\, \Sigma)\,. \qquad (4)$$

The maximum likelihood estimates for the parameter ω is obtained by maximizing the marginal density $f(Y|\omega)$, which is obtained by integrating out the random effect U in Eq. (4). That is $L(\omega; Y) = \int (f(Y|U;\ \omega)\,.f(U;\ \Sigma)\ d\ U$. 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 as unobserved (missing values).

The three steps of the SAEM are as follow.

- **Simulation step**

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

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

The conditional distribution does not have a standard form. Thus, the Metropolis-Hastings algorithm is adopted. The iterations are as follows [17].

1. Initialize the parameters $\omega^s = (\eta_\tau^s(t), \sigma^s, \Sigma^s)$ at s=0.
2. For each subject, independently draw a sample $\{U_i^s{}_k$: k=1,…., $l_{(s+1)}\}$ from the conditional distribution $f(U_i|Y_i;\ \omega^s)$ using 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

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

The choice of the proposal distribution is essential for 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 other hand, 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 are accepted) and choose the value of the standard deviation so that the acceptance rate is far from 0 and far from 1 [18].

- **Integration step**

The integration step involves approximating the Q-function. At the (s+1)$^{\text{th}}$ iteration, the Q-function approximated as:

$$Q\left(\omega\middle|\omega^{(s+1)}\right) = (1-\varphi_s)Q\left(\omega\middle|\omega^{(s)}\right) + \varphi_s \frac{1}{l_{s+1}} \sum_{k=1}^{l_{s+1}} L\left(\omega;\ Y_i, U_{ik}^{(s+1)}\right)$$

$$= Q\left(\omega\middle|\omega^{(s)}\right) + \varphi_s \left\{\frac{1}{l_{s+1}} \sum_{k=1}^{l_{s+1}} L\left(\omega;\ Y_i, U_{ik}^{(s+1)}\right) - Q\left(\omega\middle|\omega^{(s)}\right)\right\}, \qquad (5)$$

where φ_s is a smoothness parameter which is a decreasing sequence of positive numbers such that $\sum_{t=1}^{\infty} \varphi_t \to \infty$, and $\sum_{t=1}^{\infty} {\varphi_t}^2 < \infty$, $L\left(\omega;\ Y_i, U_i^{(s+1)}\right)$ is the pseudo log-likelihood for the i^{th} subject at (s+1) step. The pseudo log-likelihood takes the following form

$$L(\omega;\ Y, U) = \log \prod_{i=1}^{n} f(Y_i|U_i, X_i, S_i; \eta_\tau(t), \sigma).f(U_i|\ \Sigma)$$

$$= \sum_{i=1}^{n} \log\left(\ f(Y_i|U_i, X_i, S_i; \eta_\tau(t), \sigma).f(U_i|\ \Sigma)\right)$$

$$= \sum_{i=1}^{n} \log\left(\ f(Y_i|U_i, X_i, S_i; \eta_\tau(t), \sigma)\right) + \sum_{i=1}^{n} \log\left(\ f(U_i|\ \Sigma)\right).$$

- **The maximization step**

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

The above steps are repeated until convergence. The value of the smoothing parameter φ_t governs the convergence of the estimates. If the smoothing parameter φ_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 other hand, when the smoothing parameter φ_t is different from 1, the algorithm will converge slowly (almost sure convergence) to the ML solution [16].

Galarza, et al. [16] suggest the following choice of the smoothing parameter:

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

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 results in a Markov chain where the mean of the chain observations can be a satisfactory estimate, after removing a burn-in period [16]. 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 uses some or all previous simulations, not only the current simulation of the missing data. This 'memory' property is set by the smoothing parameter φ_t, and this unlike the traditional EM algorithm and its variants [16].

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 a possible to choose these constants, such that the convergence of the estimates for all the parameters can be monitored. Also, it is possible to monitor the difference (relative difference) between two successive evaluations of the log-likelihood $l(\omega|\mathrm{Y_i})$, given by $\|l(\omega^s|\mathrm{Y_i}) - l(\omega^{s+1}|\mathrm{Y_i})\|$ or $\left\|\frac{l\left(\omega^{s+1}\middle|\mathrm{Y_i}\right)}{l(\omega^s|\mathrm{Y_i})} - 1\right\|$ respectively. Also, the Akaike information criteria (AIC) can be calculated from the final estimated log-likelihood to evaluate the model fit.

2.2.2 Standard errors and confidence intervals

The construction of confidence intervals, calculating standard errors, and calculating P-values of the parameters is usually based on the asymptotic normality of the maximum likelihood estimator (MLE). There are trials in the literature that study the asymptotic theory for quantile regression, but the development of convenient inference procedures has been still challenging.

This is because the asymptotic covariance matrix of quantile estimates involves the unknown error density function, which cannot be estimated reliably [13].

In our case, the error term has been set to be ALD, and for a given τ , the mode of ALD is located at the τ_{th} quantile of residuals. Maximizing the likelihood may lead to unbiased point estimate, but in some cases, the error may not be distributed as ALD. Also, the density function might not be differentiable with respect to parameters. There are some alternatives, that provide inference for quantile regression with longitudinal data, such as the rank score test proposed by Wang and Fygenson [9], and the block bootstrap method which has been applied in Buchinsky [19] and Lipsitz, et al. [20]. We consider the block bootstrap method to construct the confidence intervals for β_τ. The bootstrap method has been widely used in applications of quantile regression. To retain the dependent structure in a longitudinal data, independent subjects are assumed and the xy-pairs from each subject $\{(Yi, Xi)\}$, for $i = 1, \ldots, n$ are treated as basic resampling units. We sample from the original data, with replacement B times [1]. The practical question about choosing the number of replications B was addressed by [21].

3. Simulation Study

The aim of this simulation study is twofold. The first is to assess the performance of the proposed techniques and to compare its performance with the method of Li, et. Al. [7], when the error follows a symmetric distribution. The second is to test the performance of the proposed techniques when the errors follow a skewed distribution.

We consider the following linear mixed change point model:

$$y_{ij} = \alpha_\tau + \left(\beta_{1,\tau} I\{x_{ij} \le t_\tau\} + \beta_{2,\tau} I\{x_{ij} \ge t_\tau\}\right)(x_{ij} - t_\tau) + s_{ij}^T \gamma_\tau + z_{ij}^T U_i + \varepsilon_{\tau,ij}, \quad (6)$$

for $j = 1,2,3,4,5, \; i = 1, \ldots \ldots, n.$

The goal is to estimate the fixed effects parameters β, and the location of the change point for a grid of percentiles p ={0.25, 0.50, 0.75}.

3.1. Simulation setting

The variable x_{ij} is simulated from the normal distribution with a mean of 5 and a standard deviation of 2. We simulated a 5 × 4 design matrix s_{ij} for the fixed effects γ_τ, where the first column corresponds to the intercept, the second column represent group variable that is the subjects are randomized on two groups. The third column represents a variable follows uniform distribution on (0,1). The fourth column represents a variable follows normal distribution with a mean of 3 and a standard deviation of 1. The matrix z_{ij}, that is associated with the random effects, is subset from matrix s_{ij} with 3 columns which are the intercept and columns number 3 and 4 in matrix s_{ij}.

The fixed effects parameters were chosen as γ_1 = 5.5, γ_2 = 4 and γ_3 = 2, $\gamma_4 = 3$. 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:

1- An ALD (*0, σ, p*), where *p* stands for 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 3×1 vector that is generated from multivariate asymmetric Laplace distribution with a mean of 0 and a variance-covariance matrix Σ. The matrix Σ is chosen to follows AR(1) structure with $\rho = 0.5$ and $\sigma = 0.8$. This structure gives the better acceptance rate for the Metropolis-Hasting algorithm.

We use *m=4* and *n = 40* and *100*. In addition, *l*= 20 (number of simulations), *W = 500* (the number of Monte-Carlo iterations) and *c = 0.2*. Note, the choice of *c* depends on the dataset, and the underlying model. We generate *1000* data samples for each scenario.

3.2. Simulation results

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

Figures (1) and (2) display samples from the trace plots of the estimates for n=40 and n=100, respectively. Figure (3) displays the visual monitoring of the relative difference between two successive evaluations of the log-likelihoods. Assume that the first100 iterations (which is 20% of *W*), as burn-in period, it is clear that all the estimates converge.

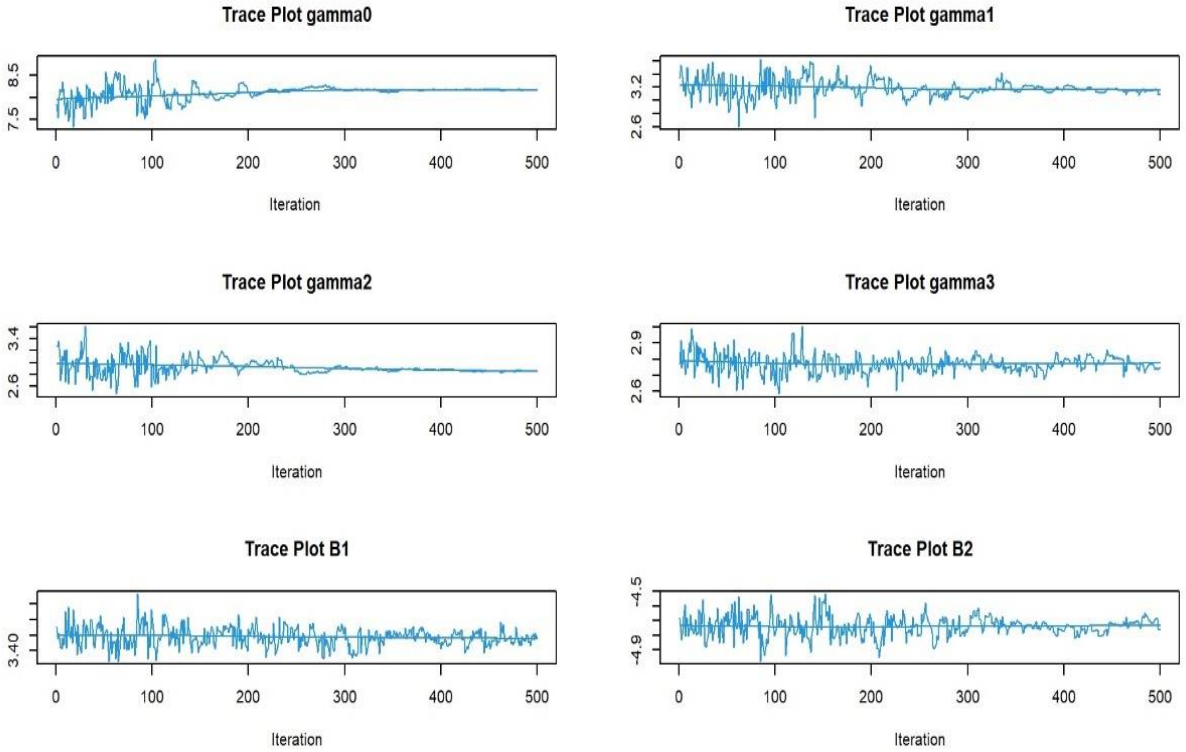

Figure (1): A Sample of the trace plot for the SAEM estimates (n= 40), $\varepsilon_{\tau,ij} \sim ALD$

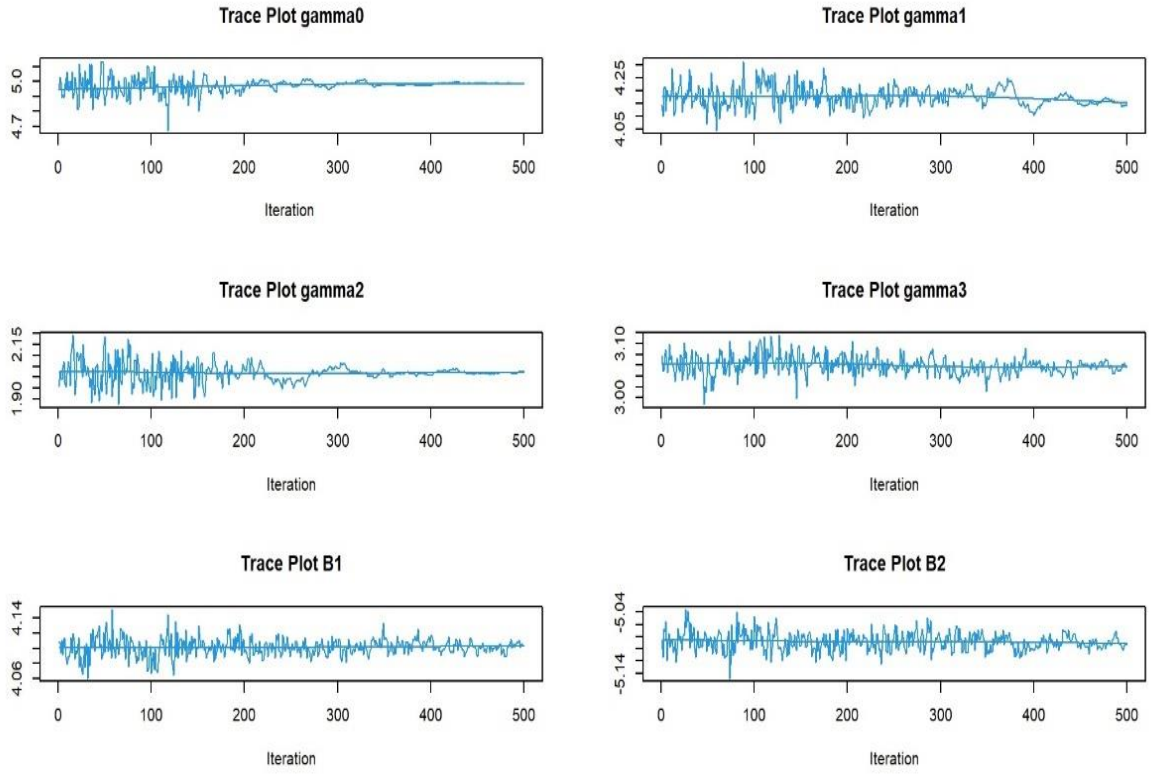

Figure (2): A Sample of the trace plot for the SAEM estimates (n = 100), $\varepsilon_{\tau,ij} \sim ALD$

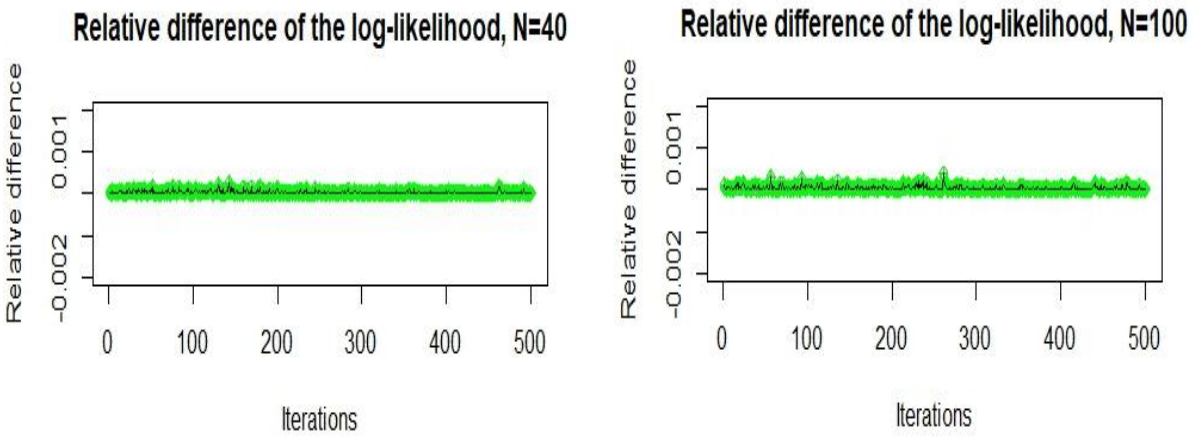

Figure (3): Sample of visual monitoring the difference (relative difference) between two successive evaluations of the log-likelihood, $\varepsilon_{\tau,ij} \sim ALD$

For all scenarios, the absolute relative bias (ARB) for each parameter over the 1000 replicates is obtained as:

$$\text{ARB} = \left|\frac{\text{estimated value}-\text{true value}}{\text{true value}}\right|,$$

and the standard error of each estimator is obtained as:

$$\widehat{SE}\left(\hat{\theta}\right) = \sqrt{\frac{1}{K}\sum_{k=1}^{1000}(\hat{\theta}_k - \hat{\theta}^*)^2},$$

where $\hat{\theta}_k$ is the K^{th} estimate of θ using the K^{th} sample and $\hat{\theta}^*$ is the average of the multiple estimates. That is,

$$\hat{\theta}^* = \frac{\sum_{k=1}^{k}\hat{\theta}_k}{k}$$

Table (1): Simulation results, the standard errors, and the relative bias of different models at n=40,100, $\varepsilon_{\tau,ij}$~ALD

		n=40, c=0.2				n=100, c=0.2			
		Proposed Method		Li.et al. (2015)		Proposed Method		Li.et al. (2015)	
Distribution	parameter	Relative bias%	S.E	Relative bias%	S.E	Relative bias%	S.E	Relative bias%	S.E
τ = 0.25	*γ0*	0.077	0.043	0.142	0.052	0.084	0.053	0.115	0.115
	γ1	0.053	0.027	0.041	0.026	0.032	0.039	0.036	0.036
	γ2	0.181	0.035	0.274	0.043	0.130	0.044	0.058	0.058
	γ3	0.024	0.011	0.086	0.012	0.003	0.010	0.000	0.000
	B1	0.008	0.008	0.005	0.010	0.012	0.012	0.022	0.022
	B2	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
	σ	0.322	0.019			0.375	0.024		
AIC		11603.900		11604.880		29134.340		29136.100	
τ = 0.5	*γ0*	0.121	0.165	0.142	0.188	0.078	0.043	0.122	0.052
	γ1	0.299	0.117	0.041	0.146	0.028	0.027	0.058	0.026
	γ2	0.389	0.125	0.274	0.177	0.065	0.035	0.032	0.043
	γ3	0.120	0.045	0.086	0.056	0.021	0.011	0.025	0.012
	B1	0.058	0.029	0.005	0.039	0.025	0.008	0.031	0.010
	B2	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
	σ	0.140	0.075			0.124	0.019		
AIC		11619.340		11620.040		29065.320		29066.240	
τ = 0.75	*γ0*	0.008	0.152	0.045	0.167	0.103	0.070	0.104	0.083
	γ1	0.016	0.151	0.139	0.176	0.043	0.039	0.030	0.041
	γ2	0.039	0.101	0.395	0.113	0.037	0.043	0.117	0.051
	γ3	0.005	0.033	0.007	0.039	0.113	0.014	0.134	0.015
	B1	0.033	0.035	0.037	0.040	0.019	0.014	0.010	0.019
	B2	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
	σ	0.354	0.074			0.398	0.026		
AIC		11663.790		11664.600		29164.130		29166.860	

Table (2): Simulation results, the standard errors, and the relative bias for different models at n=40,100, $\varepsilon_{\tau,ij} \sim Lognormal$

		n=40, c=0.2				n=100, c=0.2			
		Proposed Method		Li.et al. method		Proposed Method		Li.et al. method	
Distribution	parameter	Relative bias%	S.E	Relative bias%	S.E	Relative bias%	S.E	Relative bias%	S.E
τ = 0.25	*γ0*	0.092	0.141	0.106	0.187	0.017	0.055	0.101	0.082
	γ1	0.167	0.083	0.198	0.103	0.102	0.058	0.089	0.064
	γ2	0.127	0.054	0.241	0.079	0.042	0.036	0.015	0.054
	γ3	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
	Change-point	0.0748	0.057	0.0741	0.076	0.0362	0.030	0.036	0.042
	σ	0.363	0.043			0.0811	0.025		
AIC		11647.47		11650.32		29150.83		29150.87	
τ = 0.5	*γ0*	0.204	0.131	0.267	0.137	0.029	0.060	0.033	0.070
	γ1	0.017	0.074	0.025	0.063	0.107	0.047	0.090	0.055
	γ2	0.078	0.071	0.126	0.103	0.009	0.025	0.157	0.039
	γ3	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
	Change-point	0.077	0.056	0.072	0.063	0.038	0.027	0.036	0.034
	σ	0.1054	0.043			0.249	0.022		
AIC		11663.5		11665.6		29124.52		29125.03	
τ = 0.75	*γ0*	0.044	0.177	0.051	0.198	0.049	0.066	0.046	0.096
	γ1	0.039	0.107	0.045	0.107	0.153	0.052	0.174	0.062
	γ2	0.186	0.058	0.164	0.081	0.002	0.053	0.094	0.075
	γ3	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
	Change-point	0.024	0.068	0.026	0.115	0.036	0.035	0.042	0.048
	σ	0.011	0.050			0.103	0.030		
AIC		11653.44		11654.07		29130.44		29139.75	

Table (1) and Table (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 symetric and skewed distributions. This is because the relative bias of all estimates using the proposed algorithm is relatively small for both the ALD and the lognormal distributions. The result is valid for sample sizes = 40 and 100.
2. It is clear that the RAB 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 ALD or lognormal distribution and sample sizes = 40, and 100. This mean that the proposed method is better than the algorithm proposed by Li, et al. [7].

3. The proposed method is more efficient than the the algorithm proposed by Li, et al. [7]. The standard errors of all estimators for the proposed algorithm is less than their counterparts of the algorithm proposed by Li, et al. [7].
4. All the AIC values for the proposed technique is less than that for the model estimaed by Li, et al. [7]. This means that the proposed algorithm outperforms the algorithm propsed by Li, et al. [7].

4. Application: COVID-19

The first diagnosed case of COVID-19 symptoms is dated back to December 2019 in Wuhan city, China. The COVID-19 affected the whole world, socially, economically, and even politically. There are 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 under focus of many studies.

Studies have shown that the environmental factors, in general, may affect the fast spread of the COVID-19. Also, studies tried to investigate the impact of economic factors on the virus transmission. A study on some Chinese cities during the period January, the 19th and February, the 29th of 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 [22].

The effect of demographic factors on spread of COVID-19 was studied. Khan, et al. [23] demonstrate that certain demographic attributes, such as the age distribution, the poverty ratio, the female smoker's percentage, the obesity level, and the average annual temperature of the country, are significantly associated with COVID-19 death rate distribution.

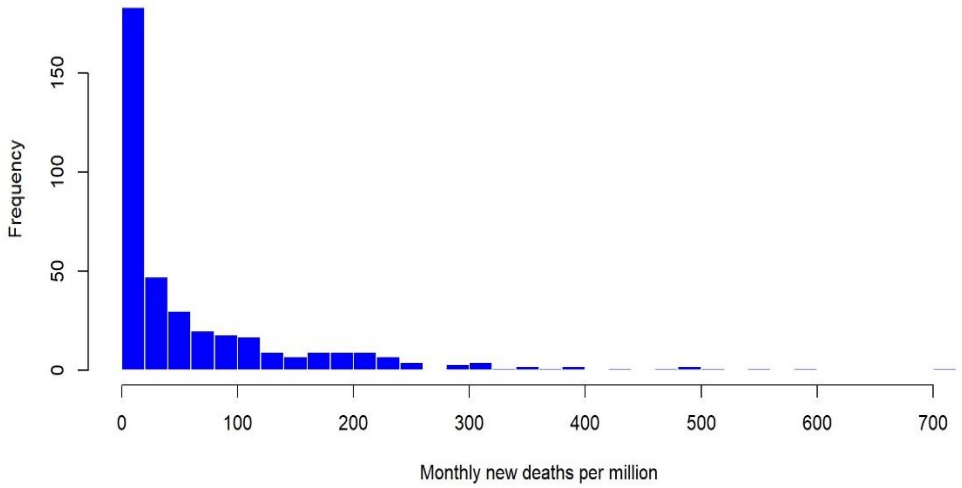

Figure (4): Histogram of the monthly new deaths per million

In this article we focus on global economic factors and health factors affecting 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 is skewed to the right. So, we need to focus on the factors that affect the lower and/or upper tails of the distribution of the new deaths. However, most of the proposed statistical methodology for describing longitudinal data with change point rely upon distributional assumptions that is not hold in this case.

4.1. Data

The used data about the COVID-19 new deaths were obtained from the Our World In Data online (https://ourworldindata.org/coronavirus) publication. This cite has become one of the world leading websites during the pandemic in 2020. The study focuses on the monthly data during the period starting from the 1st of April 2020 till the 31st of April 2021. The dependent variable is the new monthly deaths per million, while the independent variables are

1- ICU: the monthly number of COVID-19 patients in intensive care units (ICUs) on a given day per million.
2- Number of tests: the monthly tests conducted per new confirmed case of COVID-19.
3- Diabetes prevalence: Diabetes prevalence (% of population aged 20 to 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.
6- Stringency index (SI): The government response stringency index; it is a composite measure based on 9 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 to1 in which 1 indicate higher development countries in the 3 aspects.

The 30 countries incorporated in the study are:

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

These countries are chosen to represent different area of the world, and according to the availability of the data. It is well mentioning that there are some countries suffers from missing data specially in the variables monthly number of COVID-19 patients in intensive care units (ICUs), and monthly tests conducted per new confirmed case of COVID-19. These values are imputed using regression method. The variables that will be associated to 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.

Table (3): Descriptive Statistics of COVID-19 Data

	Minimum	Maximum	Mean	Std. Deviation
new monthly deaths per million	0	704	72	107

ICU	1	5882	1018	1217
Number of tests	83	33120	1906	3009
Diabetes prevalence	3	17	7	3
Hospital beds	0	8	3	2
Median age	20	48	35	9
SI	25	99	66	15
HDI	.485	.955	.79267	.128355

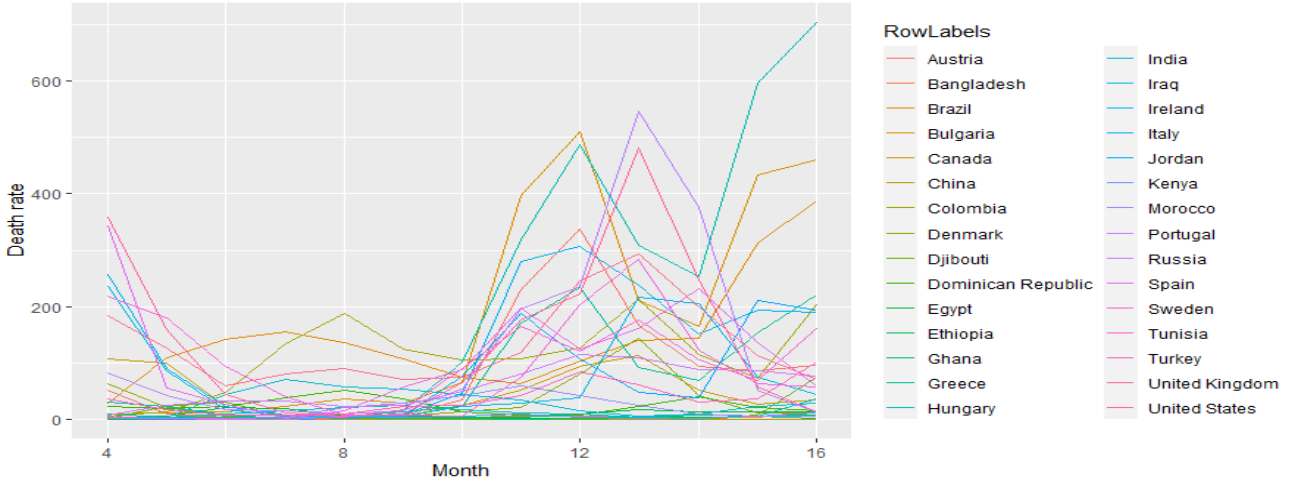

Figure (5): Spaghetti plot of new monthly deaths per million

4.2. Data analysis and results

The aim of this study is to know the global economic and health factors affecting monthly death rate per million due to COVID-19. 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 that there is a suspicion of having a threshold effect in relationship between the HDI and the new monthly deaths per million. This is obvious especially for mean regression and 0.8^{th} quantile 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 effect quantile regression with change point at 0.5^{th} quantile and at 0.8^{th} quantile.

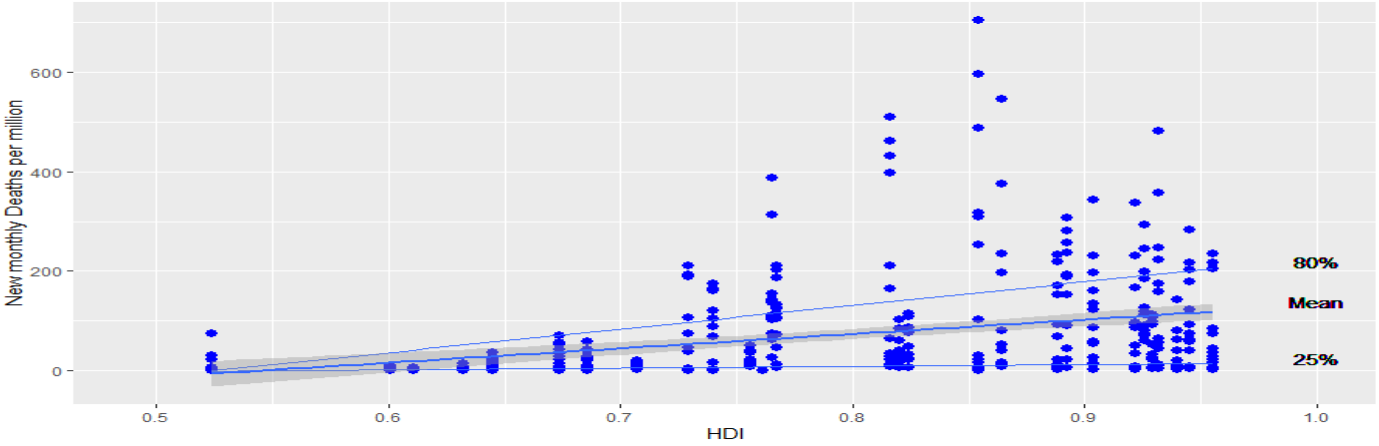

Figure (6): Scatter plot between HDI and monthly new deaths.

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	*γ0*	-346.03	-358.571	-284.866	-842.653	-855.085	-828.045
ICU	*γ1*	0.0302	0.026	0.031	0.105	0.100904	0.113117
Number of tests	*γ2*	-0.0062	-0.0066	-0.004	-0.007	-0.00767	-0.00511
SI	*γ3*	0.2042	-0.138	0.261	0.462	-0.10951	0.567555
Diabetes Prevalence	*γ4*	-0.0943	-1.795	0.868	-1.842	-2.24477	1.615969
Hospital beds	*γ 5*	-0.3533	-4.566	0.453	0.811	-4.10941	3.664168
Median age	*γ6*	-1.2929	-2.035	-0.848	-5.640	-6.26093	-4.95497
HDI	*B1*	459.41	438.673	468.887	1414.293	1408.687	1419.164
	B2	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

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

$$MDR = \alpha_\tau + \left(\beta_{1,\tau} I\{x_{ij} \le t_\tau\} + \beta_{2,\tau} I\{x_{ij} \ge t_\tau\}\right)(x_{ij} - t_\tau) + s_{ij}^T \gamma_\tau + z_{ij}^T U_i + \varepsilon_{\tau,ij}\,,$$
$$j = 1,2,3,..,13 \quad i = 1, \dots\dots, 30$$

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, stringency index (SI). The z_{ij}^T is a vector containing one's 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 change point at τ = 0.5, using the proposed algorithm and the algorithm of Li, et. al. [7] are summarized in Table (4). Also, the 95% confidence interval are obtained using block bootstrap method with number of replications equals 1000. From Table (4), depending on the confidence intervals, it is clear that there is a threshold effect in 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 0.5^{th} quantile of new monthly deaths per million due to COVID-19 is 482.41 where this effect declines after the location of the threshold to 108.48. Both effects are significant at 95% confidence. But for 0.8^{th} quantile we found that, prior the threshold value, the effect of the HDI on the 0.8^{th} 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 affect the 0.5^{th} quantile of new monthly deaths per million due to COVID-19 positively, before and after the threshold value. However, it affects the 0.8^{th} quantile of new monthly deaths per million due to COVID-19 before the threshold value positively and after the threshold value negatively.

Also we can see that there is a positive effect of monthly number of COVID-19 patients in intensive care units (ICUs) for both 0.5^{th} quantile and 0.8^{th} quantile of new monthly deaths per million. This effect is greater in case of 0.8^{th} quantile. There is negative effect of each of number

of tests, and median age of the population on both of 0.5^{th} quantile and 0.8^{th} quantile of new monthly deaths per million. However, their effect is greater in case of 0.8^{th} quantile. Each of the stringency index, hospital beds and diabetes prevalence have no significant effect on both of 0.5^{th} quantile and 0.8^{th} quantile of new monthly deaths per million due to COVID-19 .

Table (5): AIC values of the two models for COVID-19 data

	Proposed Method	Li-et al. (2015) method
tau= 0.5	24679.4	24690.74
tau= 0.8	11109.59	11121.28

Table (5) presents the AIC values for the proposed model and the Li, et. al [7] model at τ =0.5 and 0.8. The results show that the proposed model perform better than the model of Li, et al. [7].

5. Conclusion

In this article we propose a mixed effect 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 stochastic approximation EM algorithm to estimate the parameters using the link between asymptotic Laplace distribution and the quantile regression. In addition, the location of the change point is estimated using the 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 data about COVID-19. We found that there is a threshold effect in the relationship between the HDI and the 0.5^{th} quantile and, 0.8^{th} quantile of new monthly deaths per million. Also, the results show a positive effect of monthly number of COVID-19 patients in intensive care units (ICUs) for both the 0.5^{th} quantile and the 0.8^{th} quantile 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 the 0.5^{th} quantile and the 0.8^{th} quantile of new monthly deaths per million. The stringency index, hospital beds and diabetes prevalence have no significant effect on both the 0.5^{th} quantile and the 0.8^{th} quantile 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 of the researchers.

References

[1] Gad A. M. , Ibrahim W. I. M., "An adaptive linear regression approach for modeling heavy-tailed longitudinal data," Communications in Statistics - Simulation and Computation, vol. 49, no. 5, pp. 1181-1197, 2020.

[2] Yang L. , Gao S., "Bivariate random change point models for longitudinal outcomes," Statistics in medicine , vol. 32, pp. 1038-1053, 2012.

[3] Ghosh P. , Vaida F., "Random Changepoint modelling of HIV immunologic responses," Statistics in medicine , vol. 26, pp. 2074-2087, 2007.

[4] 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, vol. 70, pp. 1052-1060, 2014.

[5] Xing H. , Ying Z., "A semiparametric Change-Point Regression Model for Longitudinal Data," Journal of the American Statistical Association , vol. 107, pp. 1625-1637, 2012.

[6] Lai Y. , Albert P., "Identifying multiple change-points in a linear mixed effects model," Stat Med., vol. 33, pp. 1015-1028, 2014.

[7] 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 , vol. 71, pp. 625-635, 2015.

[8] Jung S., "Quasi-likelihood for median regression models.," Journal of American Statistical Association, vol. 91, pp. 251-257, 1996.

[9] Wang H. J. , Fygenson M., "Inference for censored quantile regression," Annals of Statistics., vol. 37, no. 2, pp. 756–781, 2009.

[10] Li C., Wei Y., Chappell R. , He X., "Bent Line Quantile Regression with application to an allometric study of land mammls' speed and mass," Biometrics , vol. 67, pp. 242-249, 2011.

[11] Sha N., "On Testing the Change-point in the Longitudinal Bent Line Quantile Regression Model," Graduate school of Arts and sciences, Columbia University, Coulmbia, 2011.

[12] Yu K. , Moyeed R., "Bayesian Quantile Regression," Statistics and Probability Letters, vol. 54, pp. 437-447, 2001.

[13] Liu Y. , Bottai M., "Mixed-eefects models for conditional quantiles with longitudinal data," International Journal of Biostatistics , vol. 5, p. Article 28, 2009.

[14] Meza C., Osorio F. , Cruz, R. D. l., "Estimation in nonlinear mixed-effects models using heavy-tailed distributions," Statistics and Computing, vol. 22, pp. 121-139, 2012.

[15] Kotz S., Kozubowski T. J. , Podgórski K., The Laplace Distribution and Generalizations. A Revisit with Applications to Communications, Economics, Engineering, and Finance., Basel: Birkhäuser, 2001, pp. 239-272.

[16] Galarza C. E., Lachos V. H., Bandyopadhyay D., "Quantile regression in linear mixed models: a stochastic approximation EM approach," Statistics and its Interface, vol. 10, no. 3, pp. 471-482, 2017.

[17] Abd Elwahab H. A., EL Kholy R. B. , Gad A. M., "Sensitivity analysis index for shared parameter models in longitudinal studies," Advances and Applications in Statistics., vol. 1, no. 57, pp. 1-20, 2019.

[18] Yaseen A. S. A. , Gad A. M., "A stochastic variant of the EM algorithm to fit mixed (discrete and continuous) longitudinal data with nonignorable missingness.," Communications in Statistics - Theory and Methods, vol. 49, no. 18, pp. 4446-4467, 2020.

[19] Buchinsky M., "Estimating the asymptotic covariance matrix for quantile regression models: a monte carlo study," Journal of Econometrics, no. 68, p. 303–338, 1995.

[20] Lipsitz R., Fitzmaurice G. M., Molenberghs G., Zhao L. P., "Quantile Regression Methods for Longitudinal Data with Drop-outs: Application to CD4 Cell Counts of Patients Infected with the Human Immunodeficiency Virus," Applied statistics, vol. 46, pp. 463-476, 1997.

[21] Andrews D. W. K. , Buchinsky M., "A three-step method for choosing the number of bootstrap repetitions," Econometrica, vol. 68, no. 1, pp. 23–52, 2000.

[22] Qiu Y., Chen X. and Shi W., "Impacts of social and economic factors on the transmission of coronavirus disease 2019 (COVID-19) in China," Journal of Population Economics , vol. 33, pp. 1127–1172, 2020.

[23] Khan W., Hussain A., Khan S. A., Al-Jumailey M., Nawaz R. and Liatsis P., "Analysing the impact of global demographic characteristics over the COVID-19 spread using class rule mining and pattern matching," Royal Society Open Science, vol. 8, no. 1, 2021. DOI: 0.1098/rsos.201823