ORIGINAL ARTICLE

Change-point analysis through integer-valued autoregressive process with application to some COVID-19 data

Subhankar Chattopadhyay[1] | **Raju Maiti[2]** | **Samarjit Das[2]** | **Atanu Biswas[1]**

[1]Applied Statistics Unit, Indian Statistical Institute, Kolkata, India

[2]Economic Research Unit, Indian Statistical Institute, Kolkata, India

Correspondence
Atanu Biswas, Applied Statistics Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata 700 108, India.
Email: atanu@isical.ac.in

Abstract

In this article, we consider the problem of change-point analysis for the count time series data through an integer-valued autoregressive process of order 1 (INAR(1)) with time-varying covariates. These types of features we observe in many real-life scenarios especially in the COVID-19 data sets, where the number of active cases over time starts falling and then again increases. In order to capture those features, we use Poisson INAR(1) process with a time-varying smoothing covariate. By using such model, we can model both the components in the active cases at time-point t namely, (i) number of nonrecovery cases from the previous time-point and (ii) number of new cases at time-point t. We study some theoretical properties of the proposed model along with forecasting. Some simulation studies are performed to study the effectiveness of the proposed method. Finally, we analyze two COVID-19 data sets and compare our proposed model with another PINAR(1) process which has time-varying covariate but no change-point, to demonstrate the overall performance of our proposed model.

KEYWORDS

active cases, change-point, COVID-19, INAR(1) process, Poisson distribution, smoothing function, time-varying covariates

1 | INTRODUCTION

Time series of count data have been widely studied during the last three decades or so due to its increased relevance toward various fields of science. There are several ways to model count time series data. For example, McKenzie (1985, 1986) and Al-Osh and Alzaid (1987) introduced a class of stationary integer-valued autoregressive (INAR) time series process based on binomial thinning operator. This process was further studied and generalized by Alzaid and Al-Osh (1990), Jin-Guan and Yuan (1991), Freeland and McCabe (2004), Ristić, Bakouch, and Nastić (2009), Jazi, Jones, and Lai (2012), Schweer and Weiß, (2014), Maiti, Biswas, and Das (2015) and many more. In particular, McKenzie (1986) introduced the integer-valued AR(1) or INAR(1) models with geometric and negative binomial marginals when the data are overdispersed. McKenzie (1985) and Al-Osh and Alzaid (1987) developed an INAR(1) process with Poisson marginals, well known as PINAR(1) process which is very popular due to its simple form. The INAR(1) process was further extended to a more general INAR(p) process by Alzaid and Al-Osh (1990) and Jin-Guan and Yuan (1991). Ristić et al. (2009) and Schweer and Weiß (2014) proposed a new INAR(1) process based on negative binomial thinning operator which can also handle the overdispersion problem. Jazi et al. (2012) and Maiti et al. (2015) studied zero-inflated PINAR(1) (ZIPINAR(1)) processes for zero-inflated count data. Apart from these thinning-based INAR processes, Cameron and Trivedi (1986) and Fokianos (2011) studied some regression-based time series models to model count time series data.

In this article, we employ the INAR process to model the data of COVID-19 active cases which is an example of count time series data. In an INAR process there are two components at time-point t namely, (i) nonrecovery cases from the previous time-point (survival part) and (ii) new cases coming in the process at time-point t (innovation terms). These INAR processes are mainly stationary since the innovation terms involve no time-varying covariate, that is, the new cases coming in the process are not time-dependent. But in real-life scenarios like the COVID-19 data sets, we can find that the rapid change in the number of infected cases makes the innovation terms time-dependent. Besides this time-varying nature of the innovation terms, we also notice some change-points in these data sets. In the current scenario of COVID-19 pandemic, we are seeing mainly two types of curves for daily new cases reported in different parts of the world, which are (i) the curve, at first, began to increase exponentially, but after major steps like "nationwide lockdowns," "social distancing" measures, a massive number of testing, and so on taken by the respective authorities in different countries, the curve started decreasing, and (ii) the curve which came down, started to rise again as the respective authorities began to ease those measures in some parts of the world. The curves of daily active cases are also changing in the same way in those parts of the world. Hence we can spot one change-point (upward to downward) for the curve described in Case (i) and two change-points (upward to downward and then downward to upward) for the curve in Case (ii). In this article, we try to develop a PINAR process based on binomial thinning operator for count time series data like the COVID-19 data where we model the innovation terms through some time-varying covariates and smoothing change-point function without changing the survival part.

PINAR process, introduced by McKenzie (1985) and Al-Osh and Alzaid (1987), is very popular due to its simple form and has a wide application in modeling count time series data. But this PINAR process based on binomial thinning operator is not capable of handling the count time series data which has both change-points and time-varying innovation terms. Hence we introduce a new suitable PINAR model which is able to tackle both these features which can be found in the

COVID-19 data sets. To incorporate the change-points in our proposed PINAR model, the innovation terms are modeled with a smoothing version (see Smooth maximum, n.d.) of time-varying covariate which consists of the change-points. The idea to capture the change-points in the innovation terms through time-varying smoothing covariate is inspired by Chan and Tong (1986), Hansen, (2000) and Fong, Huang, Gilbert, and Permar (2017) whose works are mainly based on continuous data. We use this smoothing version of time-varying components in our proposed model to catch the changing curvatures in the data of daily active cases. The effectiveness of the proposed model for both the studies of one change-point and two change-points is reviewed later by simulation study and the analysis of two COVID-19 data sets. We compare our proposed model with another PINAR model which has time-varying covariate but no change-point, to illustrate the overall performance of the proposed model.

The rest of the article is organized as follows. Section 2 discusses two real COVID-19 data sets. Section 3 describes our proposed model along with a brief illustration of the INAR(1) process. We provide the distributional forms of our proposed model and the h-step ahead forecasting distribution in Sections 4 and 5, respectively. In Section 6, we talk about the estimation method for our proposed model. Some extensive simulation studies for our proposed model are provided in Section 7. In Section 8, we analyze the COVID-19 data sets. Finally, some conclusions are drawn in Section 9. All the proofs of the theoretical results are provided in the Appendix.

2 | MOTIVATING DATA EXAMPLES: COVID-19 DATA

The world is now facing the biggest global health crisis in the name of COVID-19 pandemic unlike any in recent times. The outbreak was first identified in Wuhan, China, in early December 2019. The World Health Organization declared the outbreak a Public Health Emergency of International Concern on January 30 and a pandemic on March 11.

To restrict the spread of this virus in early stages, heavy measures have been implemented in different parts of the world by the respective authorities like "nationwide lockdowns," "rapid testing" process, strict "social distancing," using of masks and sanitizers in public places, and so on. Hence in certain parts of the world, the situation of COVID-19 has improved, and the lockdown has been eased in those parts. During that period, there were also some Gulf evacuations took place in different countries, especially in India. Therefore the "community transmission" has started in those parts of the world due to the highly infectious nature of this virus, and the number of infected cases began to pile up again. For further discussion in this regard, we explore two real COVID-19 data sets in Sections 2.1 and 2.2.

2.1 | COVID-19 data of Italy

This data set is an example of Case (i) described in Section 1. We can only see one change-point in the data of active cases of Italy and hence the study will be based on one change-point analysis. The data (see Worldometer, n.d.) are collected from February 15 to June 6 (total 113 days). Though the first case in this country was detected back in January 2020, the cases started to increase rapidly from the beginning of March. After continuous measurements taken by the authorities, the curve of active cases has started to come down. As of June 6, 2020, the total number of confirmed cases was more than 234k, and the number of deaths was more than 33.8k. The active number of cases was more than 35 000. Figure 1 displays the data of new daily cases, the data of
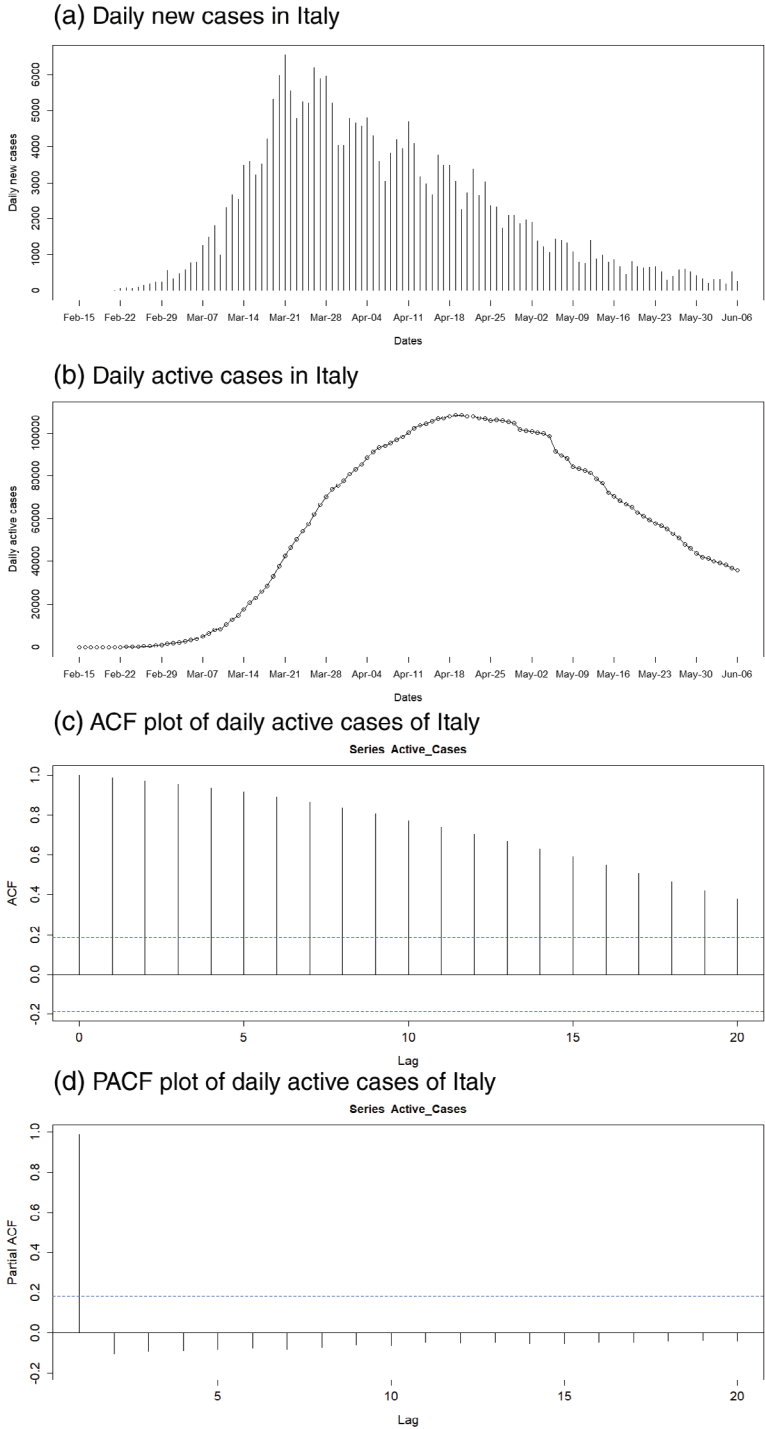
FIGURE 1 COVID-19 data of Italy

daily active cases, and the autocorrelation function (ACF) and partial ACF (PACF) plots of daily active cases. From the ACF and PACF plots, it seems that the data have a good fit for the AR(1) process.

2.2 | COVID-19 data of Kerala

This data set is an example of Case (ii). Here we observe two change-points in the data of daily active cases and hence the study will be based on two change-point analysis. In Kerala, the first case was also detected back in January 2020, but the cases started to pile up from mid-March. Due to heavy measurements taken by the state government of Kerala, the curve of active cases came down, but from mid-May, the cases again started to rise when the Gulf evacuees began to come into the state. The data for Kerala (see GoK Dashboard, n.d) are collected from March 9 to June 6 (total 90 days). More than 1800 cases and total of 15 deaths were reported in Kerala as of June 6, 2020, and the active number of cases was more than 1000. Figure 2 displays the data of new daily cases, the data of daily active cases, and the ACF and PACF plots of daily active cases. It seems that the data have a good fit for the AR(1) process.

3 | MODEL

In this section, we develop a new model based on the integer-valued AR(1) process to capture the change-points in the count time series data sets like the COVID-19 data sets of Italy and Kerala.

Here we use the INAR(1) process (proposed by McKenzie, 1985 and Al-Osh & Alzaid, 1987) consisting of binomial thinning operator (introduced by Steutel & van Harn, 1979), to develop our proposed model for change-point analysis, which is given by

$$Y_t = \alpha \circ Y_{t-1} + \varepsilon_t, \tag{1}$$

where Y_t denotes the number of daily active cases at time-point t and ε_t represents daily new cases reported at time-point t. We assume that ε_t follows Poisson(λ_t) where λ_t is assumed to have the following form:

$$\lambda_t = \exp\left(\beta_0 + \beta_1 \frac{(t - t_{ch})\exp(\delta_n(t - t_{ch}))}{1 + \exp(\delta_n(t - t_{ch}))} + \beta_2 t\right), \tag{2}$$

where the tuning parameter $\delta_n(> 0)$ helps to capture the changing curvature of the data. t_{ch} denotes the change-point in the data. The change-point t_{ch} can be easily estimated from the data. The above model is defined only for one change-point. $\boldsymbol{\beta}_1 = (\alpha, \beta_0, \beta_1, \beta_2)$ are the regression parameters, and β_1 is the regression coefficient which associates with the time-varying covariate consisting of the change-point. However, we can easily extend the model for more than one change-point. For two change-points, only the form of λ_t will change and the functional form is given by

$$\lambda_t = \exp\left(\beta_0 + \beta_1 \frac{(t - t_{ch1})\exp(\delta_n(t - t_{ch1}))}{1 + \exp(\delta_n(t - t_{ch1}))} + \beta_2 \frac{(t - t_{ch2})\exp(\delta_n(t - t_{ch2}))}{1 + \exp(\delta_n(t - t_{ch2}))} + \beta_3 t\right), \tag{3}$$
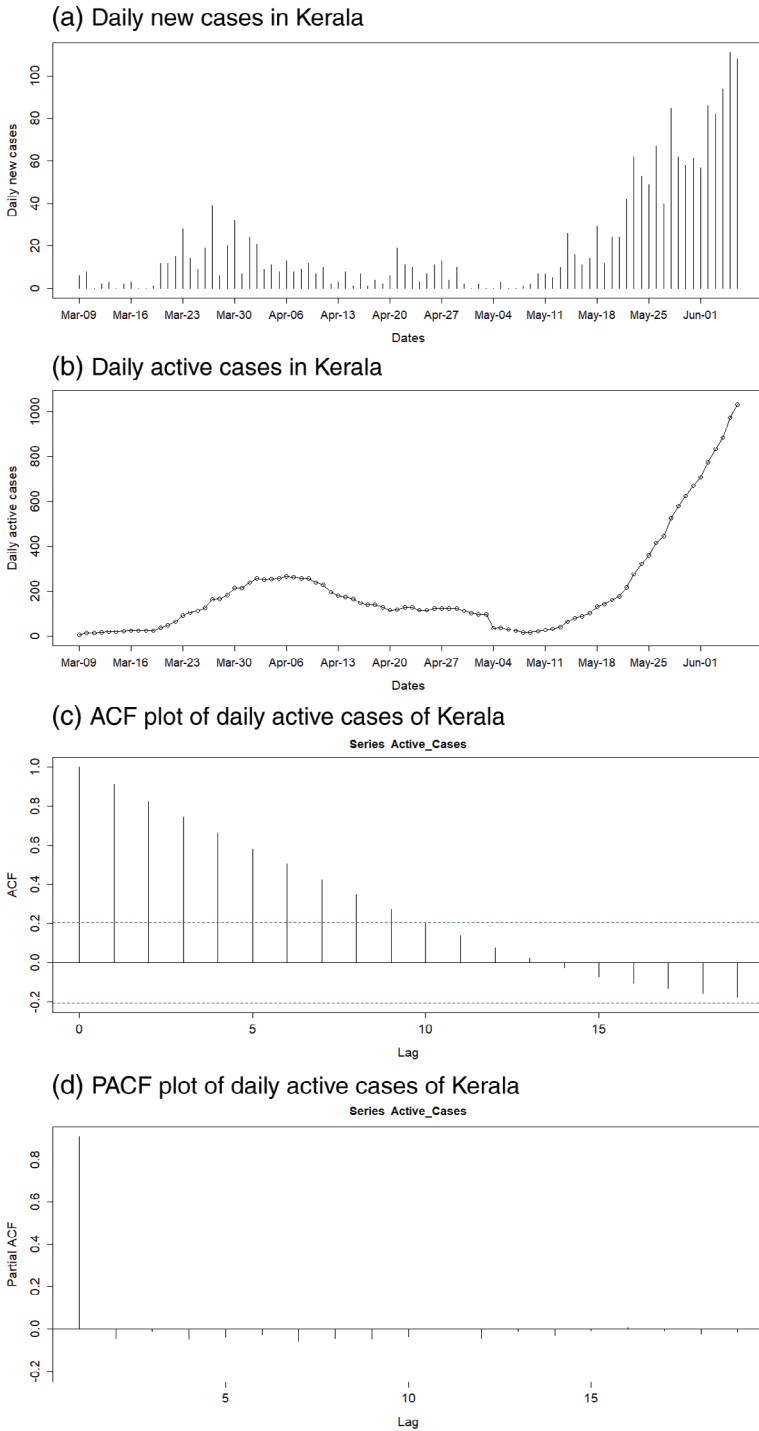
FIGURE 2 COVID-19 data of Kerala

where t_{ch1} and t_{ch2} are the two change-points in the data. Here $\boldsymbol{\beta}_2 = (\alpha, \beta_0, \beta_1, \beta_2, \beta_3)$ are the regression parameters, and β_1 and β_2 are the regression coefficients which associate with the time-varying covariates consisting of the change-points. In the subsequent section, we provide a general idea about our proposed model.

Remark 1. The use of one tuning parameter in data with one change-point can be widened for more than one change-point like using two different tuning parameters δ_{1n} and δ_{2n} for data with two change-points. But in our proposed process for two change-points, we put only δ_n instead of δ_{1n} and δ_{2n}, mainly because using δ_n reduces the computational difficulty and simplifies the form of the proposed model.

3.1 | Idea behind the model

The idea behind the form of λ_t, discussed in Equations (2) and (3), comes from the threshold regression model setup (see Chan & Tong, 1986; Fong et al. 2017; Hansen, 2000). From the concept of the segmented model in the threshold regression setup, we can write the form of $\log(\lambda_t)$ for one change-point as $\log(\lambda_t) = \beta_0 + \beta_1(t - t_{ch})_+ + \beta_2 t$, where $(t - t_{ch})_+ = (t - t_{ch})$ for $t > t_{ch}$ and $(t - t_{ch})_+ = 0$ for $t \leq t_{ch}$. In this segmented form of $\log(\lambda_t)$, we get to see sharp change (upward to downward or downward to upward) in the curve of daily new cases and hence in the curve of daily active cases. But in real-life scenarios like the COVID-19 data, we do not get to see sharp changes; most of the time we notice changing curvature(s) in these data sets. So we try to capture those changing curvature(s) in the data of daily active cases by modeling the data of daily new cases (innovation terms) in the proposed model through some time-varying covariates and smoothing change-point functions. Moreover, the function $(t - t_{ch})_+$ is not differentiable at t_{ch}. So we replace $(t - t_{ch})_+$ (for $\delta_n > 0$) by a smooth differentiable maximum function (see Smooth maximum, n.d.), which is given by

$$(t - t_{ch})_+ \approx \frac{0 \times \exp(0 \times \delta_n) + (t - t_{ch})\exp(\delta_n(t - t_{ch}))}{\exp(0 \times \delta_n) + \exp(\delta_n(t - t_{ch}))}.$$

Hence the functional form of λ_t for one change-point is given by

$$\log(\lambda_t) = \beta_0 + \beta_1 \frac{(t - t_{ch})\exp(\delta_n(t - t_{ch}))}{1 + \exp(\delta_n(t - t_{ch}))} + \beta_2 t \quad \text{for } \delta_n > 0,$$

that is,

$$\lambda_t = \exp\left(\beta_0 + \beta_1 \frac{(t - t_{ch})\exp(\delta_n(t - t_{ch}))}{1 + \exp(\delta_n(t - t_{ch}))} + \beta_2 t\right) \quad \text{for } \delta_n > 0.$$

In the similar way, we can find the functional form of λ_t for two change-points, which is given by

$$\lambda_t = \exp\left(\beta_0 + \beta_1 \frac{(t - t_{ch1})\exp(\delta_n(t - t_{ch1}))}{1 + \exp(\delta_n(t - t_{ch1}))} + \beta_2 \frac{(t - t_{ch2})\exp(\delta_n(t - t_{ch2}))}{1 + \exp(\delta_n(t - t_{ch2}))} + \beta_3 t\right) \quad \text{for } \delta_n > 0.$$

3.2 | Conditions on β_i's

The changing behaviors of these data sets depend on some conditions on β_i's. We try to provide those conditions through the form of the segmented model of threshold regression setup for both sets of β_i's (($\beta_0, \beta_1, \beta_2$) in Equation (2), and ($\beta_0, \beta_1, \beta_2, \beta_3$) in Equation (3)) which enable our proposed model to capture the change-point(s). The required conditions for both the studies of one change-point and two change-points are given below.

(i) In the segmented form of $\log(\lambda_t)$ for one change-point analysis, we model $\log(\lambda_t)$ as $\log(\lambda_t) = \beta_0 + \beta_2 t$ for $t \le t_{ch}$ and $\beta_0 + \beta_1(t - t_{ch}) + \beta_2 t$ for $t > t_{ch}$. So the derivatives of $\log(\lambda_t)$ are β_2 for $t \le t_{ch}$ and $(\beta_1 + \beta_2)$ for $t > t_{ch}$ in this segmented model setup for one change-point. So for $\beta_2 > 0$ and $(\beta_1 + \beta_2) < 0$, $\log(\lambda_t)$ increases when $t \le t_{ch}$ and decreases when $t > t_{ch}$, that is, λ_t increases when $t \le t_{ch}$ and decreases when $t > t_{ch}$. Hence the change-point is t_{ch}th time-point in the data of daily new cases. So for the count time series data of one change-point, the condition: $\{\beta_2 > 0, (\beta_1 + \beta_2) < 0\}$ must hold.

(ii) Similarly for the study of two change-points, we model $\log(\lambda_t)$ as $\log(\lambda_t) = \beta_0 + \beta_3 t$ for $t \le t_{ch1}$, $\beta_0 + \beta_1(t - t_{ch1}) + \beta_3 t$ for $t_{ch1} < t \le t_{ch2}$ and $\beta_0 + \beta_1(t - t_{ch1}) + \beta_2(t - t_{ch2}) + \beta_3 t$ for $t > t_{ch2}$. Hence the derivatives of $\log(\lambda_t)$ are β_3 for $t \le t_{ch1}$, $(\beta_1 + \beta_3)$ for $t_{ch1} < t \le t_{ch2}$ and $(\beta_1 + \beta_2 + \beta_3)$ for $t > t_{ch2}$ in the segmented model for two change-points. So for $\beta_3 > 0$, $(\beta_1 + \beta_3) < 0$ and $(\beta_1 + \beta_2 + \beta_3) > 0$, $\log(\lambda_t)$ increases when $t \le t_{ch1}$, decreases when $t_{ch1} < t \le t_{ch2}$ and again increases when $t > t_{ch2}$, that is, λ_t increases when $t \le t_{ch1}$, decreases when $t_{ch1} < t \le t_{ch2}$ and again increases when $t > t_{ch2}$. Here the two change-points are t_{ch1}th and t_{ch2}th time-points in the data of daily new cases. So the condition: $\{\beta_3 > 0, (\beta_1 + \beta_3) < 0, (\beta_1 + \beta_2 + \beta_3) > 0\}$ must hold for the count time series data containing two change-points.

3.3 | Choices of the tuning parameter δ_n

The tuning parameter of our proposed model, δ_n, helps to capture the changing curvature(s) in the data. Here $\delta_n > 0$. To compute the optimal value of the tuning parameter δ_n from the data, we consider a grid search method (see Chakraborty, Laber, & Zhao, 2013, James, Witten, Hastie, & Tibshirani, 2013). In this method, we use a goodness-of-fit measure based on which the optimal value of δ_n is calculated. The idea of δ_n comes from the concept of Smooth maximum (n.d.) and so as the value of δ_n increases the changing curvature becomes sharper. We show this property in Figures 3 and 4 where we can clearly see as the values of δ_n shift from 0.05 to 1; the changing curvatures become sharper for both the studies of one change-point and two change-points. We also add the nonsmoothing version (no use of δ_n) of the generated data, that is, the segmented data.

3.4 | Estimation of change-point(s)

To estimate the change-point(s), we take the difference between every two consecutive observations (i.e., $\Delta_t = Y_t - Y_{t-1}$) and consider the sign of those differences denoted by $S_t = \text{sign}(\Delta_t)$ where $S_t = +$ if $\Delta_t > 0$ and $-$ otherwise. For a data set with one change-point, the sequence $\{S_t\}$ should give us two runs: (1) run of $+$, and (2) run of $-$ (see Wald & Wolfowitz, 1940). Depending on the increasing or decreasing curve of Y_t, the run of $+$ and the run of $-$ will be interchanged. For example, if the original time series plot of Y_t is bell-shaped (i.e., initially the observations are
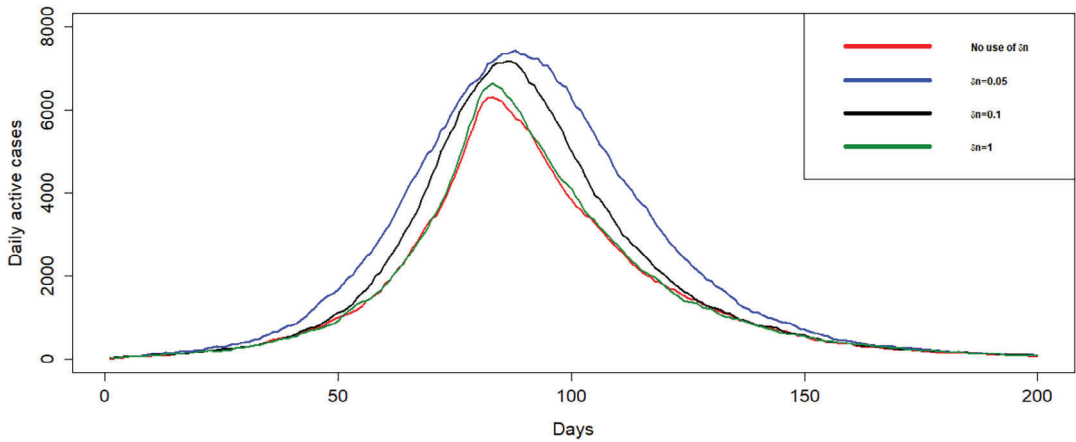
FIGURE 3   The changing curvatures for one change-point study for $\delta_n = 0.05, 0.1, 1$ along with segmented data (no use of δ_n)
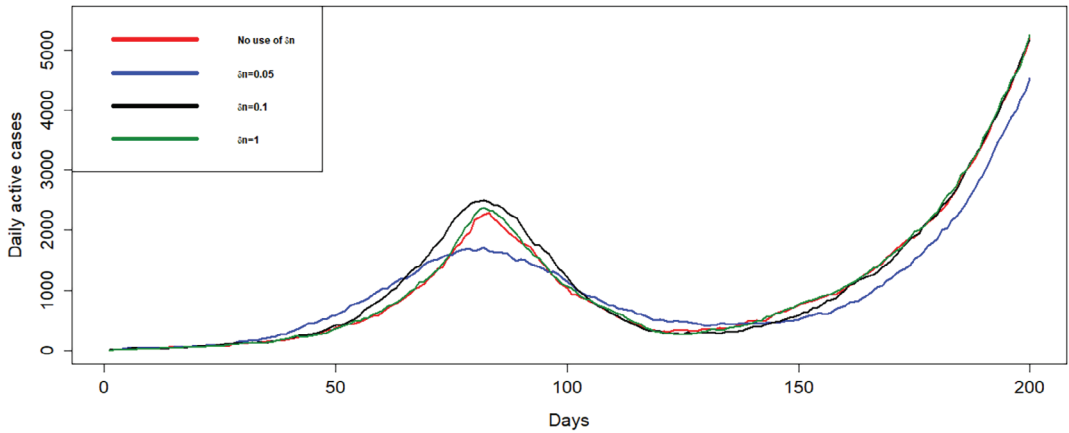

FIGURE 4   The changing curvatures for two change-point study for $\delta_n = 0.05, 0.1, 1$ along with segmented data (no use of δ_n)

increasing and then after a certain time-point (say, t_{ch}) the observations are decreasing), we will have a run of + first and then after the time-point t_{ch} we will have a run of −. The time-point at which the first run of + ends gives us an estimate of the original change-point t_{ch}. However, in real scenarios, time series data with one change-point may not be smooth and often there are random fluctuations present in the data. As a result, there might be many small runs of + and − which make the above estimation procedure difficult to locate the true change-point. Hence we employ a presmoothing approach before implementing the above run-based point estimation. That is, instead of working with the actual time series data, we make the data smooth by implementing some standard statistical approaches like m-point moving average, or through a pth degree polynomial function.

For the time series data with two change-points (say, t_{ch1} and t_{ch2}), the sequence $\{S_t\}$ should produce three runs: (1) run of +, (2) run of −, and (3) again run of +. Here the run of + and

the run of $-$ will be interchanged twice, that is, a run of $+$ for the increasing curvature, then a run of $-$ for the decreasing curvature and another run of $+$ when cases again begin to rise (another increasing curvature). The time-point at which the first run ends gives us an estimate of the first change-point t_{ch1} and the time-point at which the second run ends provides an estimate of the second change-point t_{ch2}. However, like the case of one change-point, here also time series data sets are nonsmooth and hence the implementation of presmoothing approaches like m-point moving average, or through pth degree polynomial function is required.

Later, in Section 7.2, we perform a simulation study where we estimate the true change-point(s) and provide 95% confidence interval(s) (CI(s)) based on normal approximation. And we study the large sample properties by varying the sample size.

4 | DISTRIBUTIONAL PROPERTIES

In this section, we study the conditional and the marginal distributions of the proposed model.

4.1 | Conditional distribution

Under our proposed setup, the conditional distribution of Y_t given Y_{t-1} and $\mathcal{X}_t$ (the set of all covariates up to time-point t including smooth time-varying and simple time-varying covariates up to time-point t) can be derived as

$$p(j|i) = P(Y_t = j|Y_{t-1} = i, \mathcal{X}_t) = \sum_{k=0}^{\min(i,j)} \binom{i}{k} \alpha^k (1-\alpha)^{i-k} \exp(-\lambda_t)\lambda_t^{j-k}[(j-k)!]^{-1}I_{[(j-k)=(0,1,\dots)]}, \quad (4)$$

where $I_{(\cdot)}$ is the indicator function. This is the probability of going from state i to state j in a single step. The conditional mean and variance can be given as $E(Y_t|Y_{t-1}, \mathcal{X}_t) = \alpha Y_{t-1} + \lambda_t$, and $V(Y_t|Y_{t-1}, \mathcal{X}_t) = \alpha(1-\alpha)Y_{t-1} + \lambda_t$, respectively.

4.2 | Marginal distribution

Since the marginal distribution of Y_t is difficult to obtain, we find the partial marginal distribution of Y_t given $\mathcal{X}_t$ for $t > 1$, henceforth it is called the marginal distribution. Here we derive the probability generating function (PGF) of Y_n given $\mathcal{X}_n$.

The derivation is valid for $t > 1$ and hence we assume that given $\mathcal{X}_1$, the marginal distribution of Y_1 is Poisson(λ_1). The reason behind this assumption can be given as follows. We know the elements which enter the system in the interval $(t-1, t]$ are the innovation term at time-point t (ε_t). Now for $t = 1$, the interval is $(0, 1]$, and there is no previous existing interval in the system. So in the interval $(0, 1]$, the elements which enter the system can be seen as the first count process Y_1. Hence we can assume $Y_1|\mathcal{X}_1 \sim$ Poisson(λ_1).

Theorem 1. *Under the assumptions that $Y_1|\mathcal{X}_1 \sim$ Poisson(λ_1) and $\varepsilon_n|\mathcal{X}_n \sim$ Poisson(λ_n), we can show that the PGF of $Y_n|\mathcal{X}_n$ is*

$$\Phi_{Y_n|\mathcal{X}_n}(s) = \exp\left[-(\alpha^{n-1}\lambda_1 + \alpha^{n-2}\lambda_2 + \dots + \lambda_n)(1-s)\right], \quad (5)$$

that is, Y_n given $\mathcal{X}_n$, follows Poisson distribution with mean $\left(\alpha^{n-1}\lambda_1 + \alpha^{n-2}\lambda_2 + \ldots + \lambda_n\right)$.

Proof. The derivation of this result is presented in Appendix A. ∎

Here we can also use a recursive formula as an alternative way to derive the marginal distribution, which is given by

$$
\begin{aligned}
p_{t,j} &= P(Y_t = j | \mathcal{X}_t) \\
&= \sum_{i=0}^{\infty} p(j|i) P(Y_{t-1} = i | \mathcal{X}_t) \\
&= \sum_{i=0}^{\infty} \sum_{k=0}^{\min(i,j)} \binom{i}{k} \alpha^k (1-\alpha)^{i-k} \exp(-\lambda_t) \lambda_t^{j-k} [(j-k)!]^{-1} P(Y_{t-1} = i | \mathcal{X}_t) I_{[(j-k)=(0,1,\ldots.)]} \\
&= \sum_{i=0}^{\infty} \sum_{k=0}^{\min(i,j)} \binom{i}{k} \alpha^k (1-\alpha)^{i-k} \exp(-\lambda_t) \lambda_t^{j-k} [(j-k)!]^{-1} p_{t-1,i} \; I_{[(j-k)=(0,1,\ldots.)]},
\end{aligned}
\tag{6}
$$

where $I_{(\cdot)}$ is the indicator function.

Here the marginal mean and the marginal variance are given by

$$
E(Y_n | \mathcal{X}_n) = \alpha^{n-1}\lambda_1 + \alpha^{n-2}\lambda_2 + \ldots + \lambda_n,
\tag{7}
$$

and

$$
V(Y_n | \mathcal{X}_n) = \alpha^{n-1}\lambda_1 + \alpha^{n-2}\lambda_2 + \ldots + \lambda_n.
\tag{8}
$$

Theorem 2. *Under the above setup, the autocovariance function (ACVF) of Y_t given $\mathcal{X}_{t+h}$ using the equation $Y_{t+h} = \alpha^h \circ Y_t + \sum_{i=1}^{h} \alpha^{h-i} \circ \varepsilon_{t+i}$ can be derived as*

$$
\gamma_y(h) = \mathrm{Cov}(Y_t, Y_{t+h} | \mathcal{X}_{t+h}) = \alpha^h \left(\alpha^{t-1}\lambda_1 + \alpha^{t-2}\lambda_2 + \ldots + \lambda_t \right).
$$

Proof. The derivation of this result is presented in Appendix B. ∎

Hence for $h \neq 0$, the ACF can be derived as follows:

$$
\begin{aligned}
\rho_y(h) &= \frac{\mathrm{Cov}(Y_t, Y_{t+h} | \mathcal{X}_{t+h})}{\sqrt{\mathrm{Var}(Y_t | \mathcal{X}_{t+h}) \, \mathrm{Var}(Y_{t+h} | \mathcal{X}_{t+h})}} \\
&= \frac{\alpha^h \left(\alpha^{t-1}\lambda_1 + \alpha^{t-2}\lambda_2 + \ldots + \lambda_t \right)}{\sqrt{\left(\alpha^{t-1}\lambda_1 + \ldots + \lambda_t\right)\left(\alpha^{t+h-1}\lambda_1 + \ldots + \lambda_{t+h}\right)}} \\
&= \alpha^h \sqrt{\left(\frac{\alpha^{t-1}\lambda_1 + \ldots + \lambda_t}{\alpha^{t+h-1}\lambda_1 + \ldots + \lambda_{t+h}} \right)}.
\end{aligned}
$$

It can be seen that the above expression decays exponentially to 0 as h goes to ∞ for $\alpha \in (0, 1)$ and the restricted β_i's discussed in Section 3.2.

5 | FORECASTING

5.1 | h-Step ahead forecasting distribution

To find the h-step ahead forecasting distribution, we use the following recursive method:

$$
\begin{aligned}
Y_{n+h} &= \alpha \circ Y_{n+h-1} + \varepsilon_{n+h} \\
&= \alpha \circ \{\alpha \circ Y_{n+h-2} + \varepsilon_{n+h-1}\} + \varepsilon_{n+h} \\
&\cdots \\
&= \alpha^h \circ Y_n + \sum_{i=1}^{h} \alpha^{h-i} \circ \varepsilon_{n+i}.
\end{aligned}
$$

Thus the h-step ahead conditional mean and conditional variance can be given as

$$
E(Y_{n+h}|Y_n, \mathcal{X}_{n+h}) = \alpha^h Y_n + \sum_{i=1}^{h} \alpha^{h-i} \lambda_{n+i},
$$

and

$$
V(Y_{n+h}|Y_n, \mathcal{X}_{n+h}) = \alpha^h \left(1 - \alpha^h\right) Y_n + \sum_{i=1}^{h} \alpha^{h-i} \lambda_{n+i}.
$$

The h-step ahead forecasting distribution of PINAR(1) process was derived by Freeland and McCabe (2004) using the binomial thinning operator discussed by Al-Osh and Alzaid (1987) and it turned out to be a convolution of binomial and Poisson distributions. Here we can calculate the conditional PGF of Y_{n+h} given Y_n and $\mathcal{X}_{n+h}$ and then derive the forecasting distribution using this.

Theorem 3. *The conditional PGF of Y_{n+h} given Y_n and $\mathcal{X}_{n+h}$ can be shown as*

$$
\Phi_{Y_{n+h}|Y_n, \mathcal{X}_{n+h}}(s) = \left(1 - \alpha^h + \alpha^h s\right)^{Y_n} \left[\exp\left(-\sum_{i=1}^{h} \lambda_{n+i} \, \alpha^{h-i}(1-s)\right)\right].
$$

Proof. The derivation of this result is presented in Appendix C. ∎

Corollary 1. *From the above result, we can say that the h-step ahead prediction distribution of Y_{n+h} given Y_n and $\mathcal{X}_{n+h}$ is a convolution of $Bin\left(Y_n, \alpha^h\right)$ and some random variable Z_{n+h} having the PGF of the form $\left[\exp\left(-\sum_{i=1}^{h} \lambda_{n+i} \, \alpha^{h-i}(1-s)\right)\right]$. Therefore Z_{n+h} follows Poisson distribution with mean $\left(\sum_{i=1}^{h} \lambda_{n+i} \, \alpha^{h-i}\right)$.*

Thus, the prediction distribution can be presented as

$$Y_{n+h}|Y_n, \mathcal{X}_{n+h} \overset{d}{=} \text{Bin}(Y_n, \alpha^h) * Z_{n+h},$$

where " $*$ *" is called the convolution between two distributions.*

Theorem 4. *Using Corollary 1, the h-step ahead forecasting distribution of Y_{n+h} given Y_n and $\mathcal{X}_{n+h}$ can be derived as*

$$p_h(j_1|j_2) = \sum_{q=0}^{\min(j_1, j_2)} \binom{j_2}{q} \alpha^{qh} (1 - \alpha^h)^{j_2 - q} \exp(-\lambda(h)) \frac{\lambda(h)^{j_1 - q}}{(j_1 - q)!} I_{(0,1,\dots)}(j_1 - q), \tag{9}$$

where $I_{(\cdot)}$ is the indicator function, $\lambda(h) = \left(\sum_{i=1}^{h} \alpha^{h-i} \lambda_{n+i} \right)$, and $p_h(k_1|k_2) = P(Y_{n+h} = k_1|Y_n = k_2, \mathcal{X}_{n+h})$.

Proof. The derivation of this result is presented in Appendix D. ∎

5.2 | Descriptive measure of forecasting accuracy

Given an observed data set $\{Y_1, \dots, Y_n, Y_{n+1}, \dots, Y_{n+m}\}$ of size $(n + m)$, we partition the data into two sets. The training set containing the first n observations is used to estimate the parameters of the model and based on the rest of m observations called the test set, we define the following descriptive measure of forecasting accuracy. The h-step ahead predicted root mean squared error (denoted by PRMSE(h)) is defined as

$$\text{PRMSE}(h) = \sqrt{\frac{\sum_{t=n}^{n+m-h} \left(Y_{t+h} - \hat{Y}_{t+h}^{\text{me}} \right)^2}{m - h + 1}}, \tag{10}$$

where $\hat{Y}_{t+h}^{\text{me}}$ is the mean of the estimated h-step ahead forecasting distribution of Y_{t+h} given Y_t and $\mathcal{X}_{t+h}$ mentioned in Theorem 4. Intuitively, the PRMSE(h) should increase in h.

6 | ESTIMATION METHOD FOR THE MODEL PARAMETERS

6.1 | Conditional least squares estimation

Conditional least squares estimation is usually used for estimating the regression parameters of the model in the context of time series models. Freeland and McCabe (2004, 2005) used this approach for PINAR(1) process.

In order to implement the conditional least squares estimation method, we need to minimize the sum of squared deviation about the conditional expectation which is given as $Q^*(\boldsymbol{\beta}) = \sum_{t=2}^{n} [Y_t - E(Y_t|Y_{t-1}, \mathcal{X}_t)]^2$ instead of $Q(\boldsymbol{\beta}) = \sum_{t=2}^{n} [Y_t - E(Y_t|\mathcal{X}_t)]^2$ with respect to the regression parameters of the model, where $E(Y_t|Y_{t-1}, \mathcal{X}_t) = \alpha Y_{t-1} + \lambda_t$ and $\boldsymbol{\beta}$ is the vector for regression parameters. Here numerical methods are being employed to obtain the CLS estimates of the regression parameters of the model as there are no closed forms of the CLS estimators.

In the subsequent section, we have done an extensive simulation study for both the studies of one change-point and two change-points and from the simulation results, we have shown consistency of the CLS method.

Note 1. In maximum likelihood estimation, given a data set of size n, the likelihood function for the process is given by $L(\boldsymbol{\beta}) = p(Y_1, \dots, Y_n | \mathcal{X}_n) = p(Y_1 | \mathcal{X}_1) \prod_{t=2}^{n} p(Y_t | Y_{t-1}, \mathcal{X}_t)$. In order to obtain the MLE estimators, we maximize the log-likelihood function with respect to regression parameters, which can be written as $\log L(\boldsymbol{\beta}) = \log p(Y_1 | \mathcal{X}_1) + \sum_{t=2}^{n} \log p(Y_t | Y_{t-1}, \mathcal{X}_t)$.

Here $P(Y_t = j | Y_{t-1} = i, \mathcal{X}_t) = \sum_{k=0}^{\min(i,j)} \binom{i}{k} \alpha^k (1-\alpha)^{i-k} \exp(-\lambda_t) \lambda_t^{j-k} [(j-k)!]^{-1}$. In real-life scenarios like the COVID-19 data, the number of daily active cases at time-point t (represented by Y_t) and the number of daily new cases at time-point t (represented by λ_t) will often be large and hence in R programming language, we face difficulties to execute the MLE method because of the terms like λ_t^{j-k} ("j" is the number of daily active cases at time-point t and $k = \min(i,j)$ where "i" is the number of daily active cases at time-point $(t-1)$) involved in the likelihood function. So the estimation method which we have employed for data analysis is CLS method.

7 | SIMULATION STUDY

7.1 | General setup

In this section, we perform extensive simulation studies for (a) the estimation of change-point(s), (b) the estimation of model parameters, and (c) the forecasting performances of the proposed model. To perform the studies, we simulate data from (1) one change-point model and (2) two change-point model. The simulation studies are performed for varying sample sizes along with different choices of model parameters, tuning parameter, and change-points.

For the simulation studies regarding the analysis of one change-point (t_{ch}), $\mathcal{X}_n$ (the set of all covariates up to time-point n) is equal to $\{1, Z_1, P_1, Z_2, P_2, \dots, Z_n, P_n\}$ where $Z_t = \frac{(t-t_{ch})\exp(\delta_n(t-t_{ch}))}{1+\exp(\delta_n(t-t_{ch}))}$, which is the smooth time-varying component and $P_t = t$, which is the simple time-varying component. And for the simulation studies regarding the analysis of two change-points (t_{ch1} and t_{ch2}), $\mathcal{X}_n = \{1, Z_{11}, Z_{21}, Q_1, Z_{12}, Z_{22}, Q_2, \dots, Z_{1n}, Z_{2n}, Q_n\}$ where $Z_{1t} = \frac{(t-t_{ch1})\exp(\delta_n(t-t_{ch1}))}{1+\exp(\delta_n(t-t_{ch1}))}$, $Z_{2t} = \frac{(t-t_{ch2})\exp(\delta_n(t-t_{ch2}))}{1+\exp(\delta_n(t-t_{ch2}))}$, and $Q_t = t$; here Z_{1t}'s and Z_{2t}'s are the smooth time-varying components and Q_t's are the simple time-varying components. In the simulation studies, we use these components for each of the studies to generate data sets of varying sample sizes by the data-generating processes mentioned in Equation (2) for one change-point and Equation (3) for two change-points.

In the simulation study regarding forecasting performances, we compare our proposed model with the following model

$$Y_t = \alpha^* \circ Y_{t-1} + \varepsilon_t, \tag{11}$$

where Y_t denotes the daily number of active cases at time-point t and ε_t represents the daily number of new cases at time-point t. Here ε_t follows Poisson(λ_t^*) where λ_t^* is assumed to have the following form:

$$\lambda_t^* = \exp(\gamma_0 + \gamma_1 t).$$

This model involves no change-point. But the innovation terms depend on time-varying covariates.

7.2 | Results on change-point(s) estimation

Here we perform a simulation study in order to provide 95% CIs for the true change-points from the simulated data sets and examine the widths of those intervals with increasing sample size. The estimation method of change-point(s) is discussed in Section 3.4. In order to perform this simulation study, we simulate data from the proposed model with (1) one change-point (given in Equation (2)), and (2) two change-points (given in Equation 3)). Two sets of regression parameters are considered for each of the above two data-generating cases. Three different sample sizes (n) of 400, 450, and 500 are explored. Throughout the whole simulation study, we consider two different values of δ_n as 0.1 and 0.2. All the simulations results are based on 1000 Monte Carlo replications.

7.2.1 | Case 1: Analysis of one change-point

For one change-point simulation study, we assume the value of the true change-point t_{ch} to be $0.5n$ where n is the sample size of the data. The estimation method of the change-point is discussed in Section 3.4. Two sets of regression parameters used in the data-generating process are $\boldsymbol{\beta}_1 = (\alpha, \beta_0, \beta_1, \beta_2) = (0.4, 1.5, -0.08, 0.04)$ and $(0.7, -1, -0.09, 0.05)$. For each set of the regression parameters and the tuning parameter δ_n, we simulate the data using model (1) with λ_t given in Equation (2). Here for the data-generating method of one change-point, $\mathcal{X}_n$, set of all covariates up to time-point n, consists of both the smooth time-varying components and the simple time-varying components up to time-point n as described in Section 7.1, where n is the sample size of the simulated data set. The process is repeated for 1000 times and we report the 95% CIs in Tables 1 and 2 where we can see that as the sample sizes increase the widths of the CIs decrease.

7.2.2 | Case 2: Analysis of two change-points

For the simulation study of two change-points, the true change-points t_{ch1} and t_{ch2} are assumed to be $0.4n$ and $0.6n$, respectively. Two sets of values of the regression parameters

TABLE 1 95% confidence intervals (CIs) for the true change-point for different sample sizes for different values of δ_n where the true change-point is at $0.5n$th time-point and true $\boldsymbol{\beta}_1 = (0.4, 1.5, -0.08, 0.04)$

$\delta_n = 0.1$		
n	95% CI	Width
400	(196.6439, 203.3801)	6.7362
450	(222.5360, 227.3340)	4.7980
500	(248.1794, 251.7426)	3.5632
$\delta_n = 0.2$		
n	95% CI	Width
400	(197.8427, 202.1033)	4.2606
450	(223.4354, 226.6266)	3.1912
500	(248.9330, 251.0090)	2.0760

TABLE 2 95% confidence intervals (CIs) for the true change-point for different sample sizes for different values of δ_n where the true change-point is at $0.5n$th time-point, and true $\beta_1 = (0.7, -1, -0.09, 0.05)$

$\delta_n = 0.1$		
n	95% CI	Width
400	(197.9393, 205.0067)	7.0674
450	(224.1986, 228.8794)	4.6808
500	(249.8242, 252.8998)	3.0756
$\delta_n = 0.2$		
n	95% CI	Width
400	(198.4818, 203.1122)	4.6304
450	(224.1968, 227.2532)	3.0564
500	(249.5226, 251.7134)	2.1908

TABLE 3 95% confidence intervals (CIs) for the true change-points for different sample sizes for different values of δ_n where the true change-points are at $0.4n$th and $0.6n$th time-points, and true $\beta_2 = (0.5, 3.5, -0.1, 0.1, 0.05)$

$\delta_n = 0.1$				
n	95% CI for first change-point	Width	95% CI for second change-point	Width
400	(158.7577, 161.6283)	2.8706	(235.0461, 245.1539)	10.1078
450	(178.9962, 181.0778)	2.0816	(265.5051, 274.5129)	9.0078
500	(199.3109, 200.7131)	1.4022	(296.0952, 303.7808)	7.6856
$\delta_n = 0.2$				
n	95% CI for first change-point	Width	95% CI for second change-point	Width
400	(159.1228, 160.9412)	1.8184	(236.4025, 243.1755)	6.7730
450	(179.4911, 180.5349)	1.0438	(266.7304, 272.4936)	5.7632
500	(199.9124, 200.0876)	0.1752	(297.2184, 301.8216)	4.6032

used in the data-generating process are $\beta_2 = (\alpha, \beta_0, \beta_1, \beta_2, \beta_3) = (0.5, 3.5, -0.1, 0.1, 0.05)$ and $(0.7, 4.5, -0.08, 0.09, 0.04)$. For each set of the regression parameters and the tuning parameter δ_n, we simulate the data using model (1) with λ_t given in Equation (3). The estimation method of the change-point is discussed in Section 3.4. Here for the data-generating method of two change-points, $\mathcal{X}_n$, set of all covariates up to time-point n, consists of both the smooth time-varying components and the simple time-varying components up to time-point n as described in Section 7.1, where n is the sample size of the simulated data set. The process is repeated for 1000 times and the 95% CIs are reported in Tables 3 and 4. From the tables, we can see that as the sample sizes increase the widths of the CIs decrease.

TABLE 4 95% confidence intervals (CIs) for the true change-points for different sample sizes for different values of δ_n where the true change-points are at $0.4n$th and $0.6n$th time-points, and true $\beta_2 = (0.7, 4.5, -0.08, 0.09, 0.04)$

$\delta_n = 0.1$				
n	**95% CI for first change-point**	**Width**	**95% CI for second change-point**	**Width**
400	(158.5181, 162.2079)	3.6898	(233.8636, 244.1944)	10.3308
450	(178.6895, 181.8585)	3.1690	(264.5409, 273.4491)	8.9082
500	(198.8991, 201.3409)	2.4418	(294.9574, 302.7986)	7.8412

$\delta_n = 0.2$				
n	**95% CI for first change-point**	Width	**95% CI for second change-point**	**Width**
400	(158.8499, 161.4501)	2.6002	(235.8058, 242.6082)	6.8024
450	(179.0436, 181.0664)	2.0228	(266.2779, 272.0281)	5.7502
500	(199.3532, 200.7068)	1.3536	(296.5623, 301.5857)	5.0234

7.3 | Results on estimation of model parameters

Here we perform a simulation study to investigate the consistency of the estimation method used for the proposed model. In order to perform this simulation study, we simulate data from the proposed model with (1) one change-point (given in Equation (2)), and (2) two change-points (given in Equation (3)). Three sets of regression parameters are considered for each of the above two data-generating cases. Those values are mentioned in the subsequent sections. Three different sample sizes (n) of 100, 200, and 500 are explored. Throughout the whole simulation study, we consider three different values of δ_n as 0.1, 0.5, and 1. All the simulations results are based on 1000 Monte Carlo replications.

7.3.1 | Case 1: Analysis of one change-point

For one change-point simulation study, we assume the value of the change-point t_{ch} to be $0.4n$ where n is the sample size of the data. Three sets of regression parameters used in the data-generating process are $\beta_1 = (\alpha, \beta_0, \beta_1, \beta_2) = (0.5, 0.1, -0.2, 0.02), (0.6, -0.2, -0.04, 0.02)$, and $(0.8, -0.1, -0.02, 0.01)$. For each set of the regression parameters and the tuning parameter δ_n, we simulate the data using model (1) with λ_t given in Equation (2). Then we estimate the regression parameters using CLS estimation method. Here for the data-generating method of one change-point, $\mathcal{X}_n$, set of all covariates up to time-point n, consists of both the smooth time-varying components and the simple time-varying components up to time-point n as described in Section 7.1, where n is the sample size of the simulated data set. The process is repeated for 1000 times and we report the mean estimates and mean squared errors (MSEs) of the regression parameters in Tables 5–7. From Tables 5 to 7, we can see that as the sample size increases MSE of the estimated regression parameters decreases. This empirically establishes the consistency of the CLS estimation.

TABLE 5 Mean estimates of the regression parameters $\boldsymbol{\beta}_1$ with their respective mean squared errors (MSEs) for different sample sizes and different values of δ_n where the true $\boldsymbol{\beta}_1 = (0.5, 0.1, -0.2, 0.02)$

$\delta_n = 0.1$				
n	$\hat{\alpha}$ (MSE)	$\hat{\beta}_0$ (MSE)	$\hat{\beta}_1$ (MSE)	$\hat{\beta}_2$ (MSE)
100	0.3840 (0.0317)	0.1896 (0.1905)	−0.2264 (0.0245)	0.0207 (0.0002)
200	0.4198 (0.0188)	0.1862 (0.0853)	−0.2024 (0.0016)	0.0205 (0.0000)
500	0.4503 (0.0089)	0.1522 (0.0284)	−0.1985 (0.0002)	0.0202 (0.0000)
$\delta_n = 0.5$				
n	$\hat{\alpha}$ (MSE)	$\hat{\beta}_0$ (MSE)	$\hat{\beta}_1$ (MSE)	$\hat{\beta}_2$ (MSE)
100	0.4007 (0.0290)	0.1596 (1.7079)	−0.3031 (0.2222)	0.0220 (0.0011)
200	0.4357 (0.0149)	0.1782 (0.0926)	−0.2333 (0.0302)	0.0203 (0.0000)
500	0.4595 (0.0076)	0.1536 (0.0333)	−0.1982 (0.0004)	0.0201 (0.0000)
$\delta_n = 1$				
n	$\hat{\alpha}$ (MSE)	$\hat{\beta}_0$ (MSE)	$\hat{\beta}_1$ (MSE)	$\hat{\beta}_2$ (MSE)
100	0.4015 (0.0290)	0.2021 (0.1847)	−0.3268 (0.0006)	0.0211 (0.0003)
200	0.4383 (0.0145)	0.1802 (0.0838)	−0.2412 (0.0006)	0.0205 (0.0001)
500	0.4602 (0.0071)	0.1466 (0.0311)	−0.1979 (0.0001)	0.0201 (0.0000)

TABLE 6 Mean estimates of the regression parameters $\boldsymbol{\beta}_1$ with their respective mean squared errors (MSEs) for different sample sizes and different values of δ_n where the true $\boldsymbol{\beta}_1 = (0.6, -0.2, -0.04, 0.02)$

$\delta_n = 0.1$				
n	$\hat{\alpha}$ (MSE)	$\hat{\beta}_0$ (MSE)	$\hat{\beta}_1$ (MSE)	$\hat{\beta}_2$ (MSE)
100	0.5341 (0.0131)	−0.1468 (0.2255)	−0.0427 (0.0005)	0.0218 (0.0002)
200	0.5600 (0.0064)	−0.1627 (0.0995)	−0.0409 (0.0000)	0.0206 (0.0000)
500	0.5701 (0.0040)	−0.1676 (0.0270)	−0.0401 (0.0000)	0.0202 (0.0000)
$\delta_n = 0.5$				
n	$\hat{\alpha}$ (MSE)	$\hat{\beta}_0$ (MSE)	$\hat{\beta}_1$ (MSE)	$\hat{\beta}_2$ (MSE)
100	0.5308 (0.0136)	−0.1345 (0.2083)	−0.0425 (0.0004)	0.0218 (0.0002)
200	0.5597 (0.0060)	−0.1507 (0.0871)	−0.0405 (0.0000)	0.0204 (0.0000)
500	0.5778 (0.0036)	−0.1884 (0.0257)	−0.0402 (0.0000)	0.0202 (0.0000)
$\delta_n = 1$				
n	$\hat{\alpha}$ (MSE)	$\hat{\beta}_0$ (MSE)	$\hat{\beta}_1$ (MSE)	$\hat{\beta}_2$(MSE)
100	0.5329 (0.0126)	−0.1276 (0.2027)	−0.0414 (0.0005)	0.0211 (0.0002)
200	0.5583 (0.0065)	−0.1511 (0.0857)	−0.0409 (0.0000)	0.0205 (0.0000)
500	0.5752 (0.0037)	−0.1776 (0.0240)	−0.0401 (0.0000)	0.0201 (0.0000)

TABLE 7 Mean estimates of the regression parameters β_1 with their respective mean squared errors (MSEs) for different sample sizes and different values of δ_n where the true $\beta_1 = (0.8, -0.1, -0.02, 0.01)$

$\delta_n = 0.1$				
n	$\hat{\alpha}$ (MSE)	$\hat{\beta}_0$ (MSE)	$\hat{\beta}_1$ (MSE)	$\hat{\beta}_2$ (MSE)
100	0.7190 (0.0126)	−0.0579 (0.3042)	−0.0262 (0.0026)	0.0134 (0.0004)
200	0.7586 (0.0045)	−0.0924 (0.1265)	−0.0204 (0.0002)	0.0105 (0.0001)
500	0.7765 (0.0018)	−0.1037 (0.0622)	−0.0200 (0.0000)	0.0100 (0.0000)
$\delta_n = 0.5$				
n	$\hat{\alpha}$ (MSE)	$\hat{\beta}_0$ (MSE)	$\hat{\beta}_1$ (MSE)	$\hat{\beta}_2$ (MSE)
100	0.7211 (0.0129)	−0.0614 (0.3907)	−0.0231 (0.0078)	0.0110 (0.0104)
200	0.7625 (0.0039)	−0.1007 (0.1208)	−0.0209 (0.0001)	0.0106 (0.0000)
500	0.7771 (0.0018)	−0.1011 (0.0654)	−0.0199 (0.0000)	0.0100 (0.0000)
$\delta_n = 1$				
n	$\hat{\alpha}$ (MSE)	$\hat{\beta}_0$ (MSE)	$\hat{\beta}_1$ (MSE)	$\hat{\beta}_2$ (MSE)
100	0.7234 (0.0116)	−0.0582 (0.2681)	−0.0240 (0.0006)	0.0135 (0.0003)
200	0.7602 (0.0043)	−0.0883 (0.1321)	−0.0207 (0.0006)	0.0103 (0.0001)
500	0.7769 (0.0020)	−0.1058 (0.0663)	−0.0199 (0.0001)	0.0098 (0.0000)

7.3.2 | Case 2: Analysis of two change-points

For two change-point simulation study, the change-points t_{ch1} and t_{ch2} are assumed to be $0.4n$ and $0.6n$, respectively. Three sets of values of the regression parameters used in the data-generating process are $\beta_2 = (\alpha, \beta_0, \beta_1, \beta_2, \beta_3) = (0.5, 0.1, -0.05, 0.04, 0.02)$, $(0.6, -0.2, -0.04, 0.03, 0.02)$, and $(0.8, -0.3, -0.02, 0.02, 0.01)$. For each set of the regression parameters and the tuning parameter δ_n, we simulate the data using model (1) with λ_t given in Equation (3). Then we estimate the regression parameters using CLS estimation method for a given simulated data. Here for the data-generating method of two change-points, $\mathcal{X}_n$, set of all covariates up to time-point n, consists of both the smooth time-varying components and the simple time-varying components up to time-point n as described in Section 7.1, where n is the sample size of the simulated data set. The process is repeated for 1000 times and the combined mean estimates and MSEs of the regression parameters are reported in Tables 8–10. We can see as the sample size increases MSE of the estimated regression parameters decreases. This establishes the consistency of the CLS estimation empirically.

7.4 | Results on forecasting performance

Another simulation study is done to study the h-step ahead forecasting performances of the proposed process for varying h, compared with the comparison method mentioned in Equation (11). For comparison, we consider the measure of forecasting criteria, namely PRMSE(h), defined in Equation (10). In order to perform this simulation study, we simulate data from the pro-

TABLE 8 Mean estimates of the regression parameters $\boldsymbol{\beta}_2$ with their respective mean squared errors (MSEs) for different sample sizes and different values of δ_n where the true $\boldsymbol{\beta}_2 = (0.5, 0.1, -0.05, 0.04, 0.02)$

$\delta_n = 0.1$					
n	$\hat{\alpha}$ (MSE)	$\hat{\beta}_0$ (MSE)	$\hat{\beta}_1$ (MSE)	$\hat{\beta}_2$ (MSE)	$\hat{\beta}_3$ (MSE)
100	0.4294 (0.0140)	0.1483 (0.1438)	−0.0531 (0.0008)	0.0415 (0.0007)	0.0218 (0.0002)
200	0.4522 (0.0072)	0.1455 (0.0627)	−0.0509 (0.0001)	0.0405 (0.0001)	0.0205 (0.0000)
500	0.4642 (0.0037)	0.1249 (0.0210)	−0.0502 (0.0000)	0.0400 (0.0000)	0.0202 (0.0000)
$\delta_n = 0.5$					
n	$\hat{\alpha}$ (MSE)	$\hat{\beta}_0$ (MSE)	$\hat{\beta}_1$ (MSE)	$\hat{\beta}_2$ (MSE)	$\hat{\beta}_3$ (MSE)
100	0.4222 (0.0155)	0.1727 (0.1312)	−0.0529 (0.0007)	0.0414 (0.0007)	0.0215 (0.0001)
200	0.4549 (0.0069)	0.1584 (0.0559)	−0.0502 (0.0001)	0.0398 (0.0001)	0.0202 (0.0000)
500	0.4572 (0.0042)	0.1512 (0.0243)	−0.0502 (0.0000)	0.0400 (0.0000)	0.0201 (0.0000)
$\delta_n = 1$					
n	$\hat{\alpha}$ (MSE)	$\hat{\beta}_0$ (MSE)	$\hat{\beta}_1$ (MSE)	$\hat{\beta}_2$ (MSE)	$\hat{\beta}_3$ (MSE)
100	0.4231 (0.0154)	0.1713 (0.1180)	−0.0512 (0.0007)	0.0394 (0.0007)	0.0212 (0.0001)
200	0.4549 (0.0069)	0.1284 (0.0634)	−0.0510 (0.0001)	0.0404 (0.0001)	0.0207 (0.0000)
500	0.4563 (0.0042)	0.1482 (0.0234)	−0.0502 (0.0000)	0.0400 (0.0000)	0.0202 (0.0000)

TABLE 9 Mean estimates of the regression parameters $\boldsymbol{\beta}_2$ with their respective mean squared errors (MSEs) for different sample sizes and different values of δ_n where the true $\boldsymbol{\beta}_2 = (0.6, -0.2, -0.04, 0.03, 0.02)$

$\delta_n = 0.1$					
n	$\hat{\alpha}$ (MSE)	$\hat{\beta}_0$ (MSE)	$\hat{\beta}_1$ (MSE)	$\hat{\beta}_2$ (MSE)	$\hat{\beta}_3$ (MSE)
100	0.5145 (0.0161)	−0.1179 (0.2584)	−0.0420 (0.0013)	0.0298 (0.0010)	0.0219 (0.0003)
200	0.5563 (0.0062)	−0.1405 (0.0818)	−0.0403 (0.0001)	0.0299 (0.0001)	0.0203 (0.0000)
500	0.5669 (0.0040)	−0.1677 (0.0419)	−0.0402 (0.0000)	0.0300 (0.0000)	0.0202 (0.0000)
$\delta_n = 0.5$					
n	$\hat{\alpha}$ (MSE)	$\hat{\beta}_0$ (MSE)	$\hat{\beta}_1$ (MSE)	$\hat{\beta}_2$ (MSE)	$\hat{\beta}_3$ (MSE)
100	0.5212 (0.0151)	−0.1173 (0.1854)	−0.0427 (0.0010)	0.0310 (0.0009)	0.0217 (0.0002)
200	0.5547 (0.0058)	−0.1381 (0.0738)	−0.0405 (0.0001)	0.0302 (0.0001)	0.0204 (0.0000)
500	0.5576 (0.0045)	−0.0985 (0.0462)	−0.0399 (0.0000)	0.0299 (0.0000)	0.0200 (0.0000)
$\delta_n = 1$					
n	$\hat{\alpha}$ (MSE)	$\hat{\beta}_0$ (MSE)	$\hat{\beta}_1$ (MSE)	$\hat{\beta}_2$ (MSE)	$\hat{\beta}_3$ (MSE)
100	0.5163 (0.0150)	−0.1113 (0.1854)	−0.0410 (0.0010)	0.0288 (0.0009)	0.0216 (0.0003)
200	0.5575 (0.0057)	−0.1376 (0.0738)	−0.0406 (0.0001)	0.0304 (0.0001)	0.0203 (0.0000)
500	0.5594 (0.0042)	−0.0948 (0.0462)	−0.0398 (0.0000)	0.0299 (0.0000)	0.0199 (0.0000)

TABLE 10 Mean estimates of the regression parameters $\boldsymbol{\beta}_2$ with their respective mean squared errors (MSEs) for different sample sizes and different values of δ_n where the true $\boldsymbol{\beta}_2 = (0.8, -0.3, -0.02, 0.02, 0.01)$

$\delta_n = 0.1$					
n	$\hat{\alpha}$ (MSE)	$\hat{\beta}_0$ (MSE)	$\hat{\beta}_1$ (MSE)	$\hat{\beta}_2$ (MSE)	$\hat{\beta}_3$ (MSE)
100	0.6923 (0.0183)	−0.1259 (0.4098)	−0.0265 (0.0019)	0.0200 (0.0021)	0.0150 (0.0004)
200	0.7489 (0.0053)	−0.1638 (0.1152)	−0.0206 (0.0001)	0.0195 (0.0001)	0.0108 (0.0000)
500	0.7786 (0.0015)	−0.2734 (0.0235)	−0.0204 (0.0000)	0.0201 (0.0000)	0.0103 (0.0000)
$\delta_n = 0.5$					
n	$\hat{\alpha}$ (MSE)	$\hat{\beta}_0$ (MSE)	$\hat{\beta}_1$ (MSE)	$\hat{\beta}_2$ (MSE)	$\hat{\beta}_3$ (MSE)
100	0.6951 (0.0178)	−0.0900 (0.2900)	−0.0241 (0.0013)	0.0196 (0.0012)	0.0137 (0.0003)
200	0.7500 (0.0050)	−0.1546 (0.1238)	−0.0202 (0.0001)	0.0195 (0.0001)	0.0105 (0.0000)
500	0.7808 (0.0013)	−0.2894 (0.0234)	−0.0205 (0.0000)	0.0201 (0.0000)	0.0101 (0.0000)
$\delta_n = 1$					
n	$\hat{\alpha}$ (MSE)	$\hat{\beta}_0$ (MSE)	$\hat{\beta}_1$ (MSE)	$\hat{\beta}_2$ (MSE)	$\hat{\beta}_3$ (MSE)
100	0.7013 (0.0161)	−0.1013 (0.3606)	−0.0232 (0.0017)	0.0182 (0.0019)	0.0135 (0.0004)
200	0.7501 (0.0052)	−0.1605 (0.1238)	−0.0211 (0.0002)	0.0205 (0.0001)	0.0106 (0.0000)
500	0.7809 (0.0014)	−0.2752 (0.0200)	−0.0204 (0.0000)	0.0201 (0.0000)	0.0103 (0.0000)

posed model with (1) one change-point (given in Equation (2)) and (2) two change-points (given in Equation (3)). Two sets of regression parameters are considered for each of the above two data-generating cases. Throughout the whole simulation study, we consider two different values of δ_n as 0.1 and 0.2. Each time we generate a total sample of size 100 of which a training set of size 85 is used to fit the two models considered for comparison and a test set of size 15 is considered to find PRMSE(h) for $h = 1, 2, 3$. This procedure is repeated for 100 times.

7.4.1 | Case 1: Analysis of one change-point

For one change-point simulation study, we assume the value of the change-point t_{ch} to be $0.3n$ where n is the sample size of the data. Two sets of regression parameters used in the data-generating process are $\boldsymbol{\beta}_1 = (\alpha, \beta_0, \beta_1, \beta_2) = (0.3, -0.8, -0.12, 0.09)$ and $(0.2, -0.5, -0.15, 0.1)$. For each set of the regression parameters and the tuning parameter δ_n, we simulate the data using model (1) with λ_t given in Equation (2). Here for the data-generating method of one change-point, $\mathcal{X}_n$, set of all covariates up to time-point n, consists of both the smooth time-varying components and the simple time-varying components up to time-point n as described in Section 7.1, where n is the sample size of the simulated data set. The process is repeated for 100 times and we report the h-step ahead forecasting performances for both the proposed model and the comparison model for $h = 1, 2, 3$ in Tables 11 and 12 where we see the average PRMSE(h) of the proposed process is relatively smaller than that of the comparison process. It is also observed from the tables that the measure seems to have an increasing pattern in h, and this coincides with the theoretical result mentioned in Section 5.2.

TABLE 11 Predicted root mean squared error (PRMSE(h)) values for varying h for different δ_n where the data-generating process is our proposed method of one change-point, and the true $\boldsymbol{\beta}_1 = (0.3, -0.8, -0.12, 0.09)$

$\delta_n = 0.1$		
h	Proposed model (PRMSE(h))	Comparison model (PRMSE(h))
1	1.2784	1.5581
2	1.3285	2.0101
3	1.3409	2.2905
$\delta_n = 0.2$		
h	Proposed model (PRMSE(h))	Comparison model (PRMSE(h))
1	1.2475	1.4388
2	1.3172	1.7750
3	1.3226	2.0305

TABLE 12 Predicted root mean squared error (PRMSE(h)) values for varying h for different δ_n where the data-generating process is our proposed method of one change-point, and the true $\boldsymbol{\beta}_1 = (0.2, -0.5, -0.15, 0.1)$

$\delta_n = 0.1$		
h	Proposed model (PRMSE(h))	Comparison model (PRMSE(h))
1	0.9143	1.2301
2	0.9193	1.6439
3	0.9221	1.9372
$\delta_n = 0.2$		
h	Proposed model (PRMSE(h))	Comparison model (PRMSE(h))
1	0.8870	1.1384
2	0.8898	1.4843
3	0.8922	1.7281

7.4.2 | Case 2: Analysis of two change-points

For the study of two change-points, the change-points t_{ch1} and t_{ch2} are assumed to be $0.2n$ and $0.6n$, respectively. Two sets of regression parameters used in the data-generating process are $\boldsymbol{\beta}_2 = (\alpha, \beta_0, \beta_1, \beta_2, \beta_3) = (0.3, -1.7, -0.3, 0.15, 0.2)$ and $(0.4, -0.8, -0.25, 0.15, 0.15)$. For each set of the regression parameters and the tuning parameter δ_n, we simulate the data using model (1) with λ_t given in Equation (3). Here for the data-generating method of two change-points, $\mathcal{X}_n$, set of all covariates up to time-point n, consists of both the smooth time-varying components and the simple time-varying components up to time-point n as described in Section 7.1, where n is the sample size of the simulated data set. The process is repeated for 100 times and we report the h-step ahead forecasting performances for both the proposed model and the comparison model for $h = 1, 2, 3$ in Tables 13 and 14 where we see the average PRMSE(h) of the proposed process is

TABLE 13 Predicted root mean squared error (PRMSE(h)) values for varying h for different δ_n where the data-generating process is our proposed method of two change-points, and the true $\boldsymbol{\beta}_2 = (0.3, -1.7, -0.3, 0.15, 0.2)$

$\delta_n = 0.1$		
h	Proposed model (PRMSE(h))	Comparison model (PRMSE(h))
1	1.1941	1.2209
2	1.2836	1.3087
3	1.3084	1.3088
$\delta_n = 0.2$		
h	Proposed model (PRMSE(h))	Comparison model (PRMSE(h))
1	1.1764	1.2360
2	1.2545	1.3146
3	1.2972	1.3215

TABLE 14 Predicted root mean squared error (PRMSE(h)) values for varying h for different δ_n where the data-generating process is our proposed method of two change-points, and the true $\boldsymbol{\beta}_2 = (0.4, -0.8, -0.25, 0.15, 0.15)$

$\delta_n = 0.1$		
h	Proposed model (PRMSE(h))	Comparison model (PRMSE(h))
1	1.1532	1.1551
2	1.2791	1.2953
3	1.2970	1.3138
$\delta_n = 0.2$		
h	Proposed model (PRMSE(h))	Comparison model (PRMSE(h))
1	1.2303	1.2378
2	1.3271	1.3631
3	1.3880	1.4039

relatively smaller than that of the comparison process. It is also observed that the measure seems to have an increasing pattern in h, and this coincides with the theoretical result mentioned in Section 5.2.

8 | DATA ANALYSIS

In this section, we consider two real data sets: (i) Italy COVID-19 data with total of 113 observations and (ii) Kerala COVID-19 data with total of 90 observations, to illustrate the usefulness of our proposed model. We compare our proposed model with the comparison model mentioned in Equation (11). For investigating the predictive performances of these two models, we take 100 observations of Italy data as the training set along with the remaining 13 observations as the test

set, and for Kerala data, we consider 78 observations as the training set along with the test set of remaining 12 observations.

8.1 | COVID-19 data of Italy

8.1.1 | Data fitting

In this section, we analyze the COVID-19 data of daily active cases of Italy (described in Section 2.1) through our proposed method. We also fit the comparison model (described in Equation (11)) to this data set. We consider 113 data points from February 15 to June 6. Here $\mathcal{X}_n$, the set of all time-varying covariates up to time-point n, contains both the smooth time-varying covariates which have the change-point and the simple time-varying covariates up to n time-points as described in Section 7.1. For this data set, $n = 113$. From the daily time series plot, we see that there is only one change-point during that period and hence we fit the proposed model with one change-point (given in Equation (2)).

For the proposed model with one change-point, the change-point t_{ch} for the COVID-19 data of Italy is estimated using the method described in Section 3.4 and the estimated point is the 36th time-point. This mostly implies that the number of cases increased up to March 21 (36 days from February 15) and after that, the number of daily new cases started decreasing gradually. In order to estimate the optimal δ_n, we consider a set of points in the interval [0.1, 10] with an increment of 0.1. For each of the δ_n in the set, we fit our one change-point model to the data. For every fit, we calculate the goodness-of-fit measure namely root mean squared error (RMSE). Then we consider the minimum value of this measure to obtain the optimal δ_n. $\delta_n = 0.1$ gives the minimum value of RMSE which is 949.85. Hence the estimated value of δ_n is 0.1.

For this data set, the estimates of the regression parameters of our proposed model by CLS method are $\hat{\boldsymbol{\beta}}_{1cls} = (\hat{\alpha}_{cls}, \hat{\beta}_{0cls}, \hat{\beta}_{1cls}, \hat{\beta}_{2cls}) = (0.9703, 3.5562, -0.1705, 0.1381)$, and that of the comparison model are $(\hat{\alpha}^*_{cls}, \hat{\gamma}_{0cls}, \hat{\gamma}_{1cls}) = (0.9925, 7.5794, -0.0156)$. The RMSE corresponding to our proposed model for the data set is 949.85, which is much lower compared with that for the comparison model, which is 1940.95. In Figure 5, we provide the plot of RMSEs against each of δ_n's in the set [0.1,10]. And in Figure 6, we give the plot of the original data along with the fitted data through both the comparison model and the proposed model.
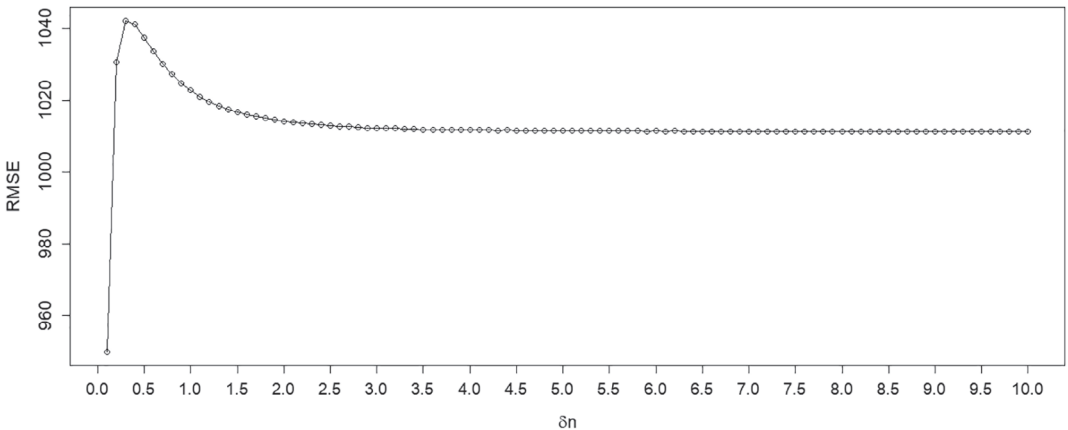

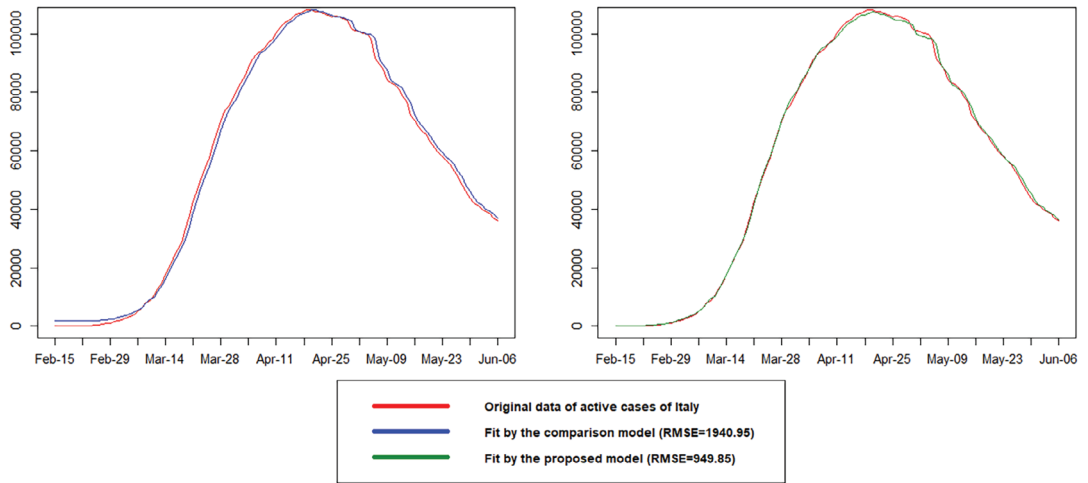**FIGURE 5** δ_n versus root mean squared error (RMSE) (for Italy)

FIGURE 6 Fitted data (active cases) by both the comparison model and the proposed model of one change-point study (for Italy)

In Figure 6, if we observe closely, we can find that the fitted data through our proposed model overlaps with major portions of the original data, but for the comparison method, we can distinguish between the original data and the fitted data in those major portions. The differences between the fitted data through our model and that through the comparison model seem to be small in Figure 6 since the magnitudes of observed data points are very high and hence the RMSEs help us here to see the differences between our proposed process and the comparison process easily rather than the plot. Overall, we can say the fit through our proposed model is good.

8.1.2 | Forecasting

To study the forecasting performance, we partition the data into two sets. As described earlier, the training set containing the first 100 observations is used to fit the models, and the test set with the remaining 13 observations, is used for finding the forecasting measure PRMSE for both models. For this setup, the estimates of the regression parameters of our proposed model by CLS method are $\hat{\boldsymbol{\beta}}_{1\text{cls}} = (\hat{\alpha}_{\text{cls}}, \ \hat{\beta}_{0\text{cls}}, \ \hat{\beta}_{1\text{cls}}, \ \hat{\beta}_{2\text{cls}}) = (0.9700, 3.6842, -0.1638, 0.1342)$, and that of the comparison model are $(\hat{\alpha}^*_{\text{cls}}, \ \hat{\gamma}_{0\text{cls}}, \ \hat{\gamma}_{1\text{cls}}) = (0.9930, 7.5369, -0.0123)$. For one-step ahead forecasting $(h = 1)$, PRMSEs for the proposed model and the comparison model are 1049.62 and 1903.99, respectively. For two-step ahead forecasting $(h = 2)$, PRMSEs for the proposed model and the comparison model are 1994.67 and 3788.10, respectively. So for both the one-step and two-step ahead forecasting results, our proposed model performs much better than the comparison model.

8.2 | COVID-19 data of Kerala

8.2.1 | Data fitting

Here we analyze the COVID-19 data of daily active cases of Kerala (see Section 2.2) through our proposed method. We also fit the comparison model (described in Equation (11)) to this data
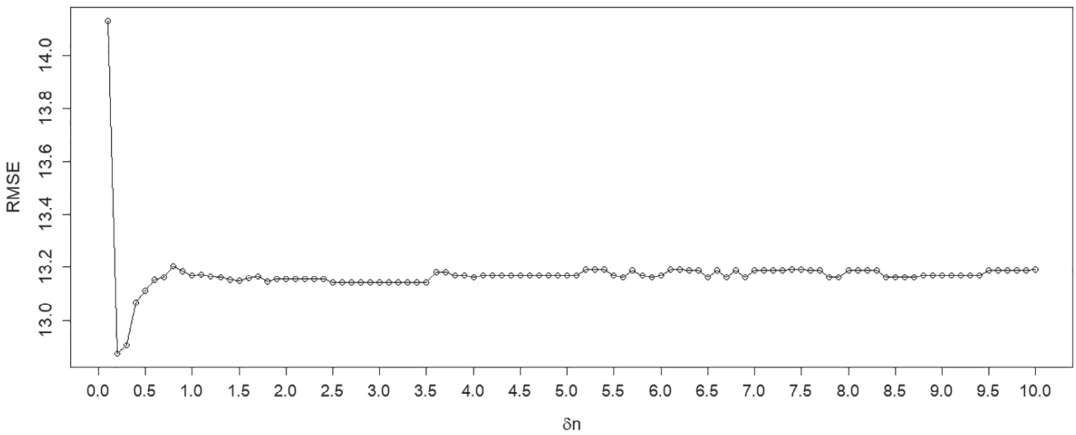
FIGURE 7 δ_n versus root mean squared error (RMSE) (for Kerala)

set. In this data set of Kerala, we consider 90 data points from March 9 to June 6. Here $\mathcal{X}_n$ consists of both the smooth time-varying covariates which have two change-points and the simple time-varying covariates up to n time-points as described in Section 7.1. For this data set, $n = 90$. From the daily time series plot, we notice that there are two change-points during that period and hence we fit the proposed model with two change-points (given in Equation (3)).

For the proposed model with two change-points, the change-points t_{ch1} and t_{ch2} are estimated using the method described in Section 3.4 and the estimated change-points are 19th and 54th time-points. This mostly implicates that the number of cases increased up to March 27 (19 days from March 9), then the number of daily new cases started decreasing gradually, but after May 1 (54 days from March 9) the cases again began to rise. In order to estimate the optimal δ_n for this data, we follow the same process as mentioned for the COVID-19 data of Italy. We find that $\delta_n = 0.2$ gives the minimum value of RMSE which is 12.88. Hence the estimated value of δ_n is 0.2.

For this data set, the estimates of the regression parameters of our proposed model by CLS method are $\hat{\boldsymbol{\beta}}_{2\text{cls}} = (\hat{\alpha}_{\text{cls}},\ \hat{\beta}_{0\text{cls}},\ \hat{\beta}_{1\text{cls}},\ \hat{\beta}_{2\text{cls}},\ \hat{\beta}_{3\text{cls}}) = (0.8700, -0.1156, -0.2352, 0.1269, 0.1956)$, and that of the comparison model are $(\hat{\alpha}^*_{\text{cls}},\ \hat{\gamma}_{0\text{cls}},\ \hat{\gamma}_{1\text{cls}}) = (0.9921, -2.8129, -0.0811)$. The RMSE corresponding to our proposed model for the data set is 12.88, which is much lower compared with that for the comparison model, which is 15.10. In Figure 7, we provide the plot of RMSEs against each of δ_n's in the set $[0.1, 10]$. And in Figure 8, we give the plot of the original data along with the fitted data through both the comparison model and the proposed model.

If we study Figure 8 closely, we see that the fitted data through our proposed model overlaps with major portions of the original data, whereas we can distinguish between the original data and the fitted data by the comparison model in those major portions. Here also the RMSEs help us to see the differences between the proposed method and the comparison method easily. So overall, we can say that the fit through our proposed model is good.

8.2.2 | Forecasting

To study the forecasting part, we partition the data into two sets. As described earlier, the training set, containing the first 78 observations, is used to fit the models, and the test set with the remaining 12 observations, is used for finding the forecasting measure PRMSE for both models.
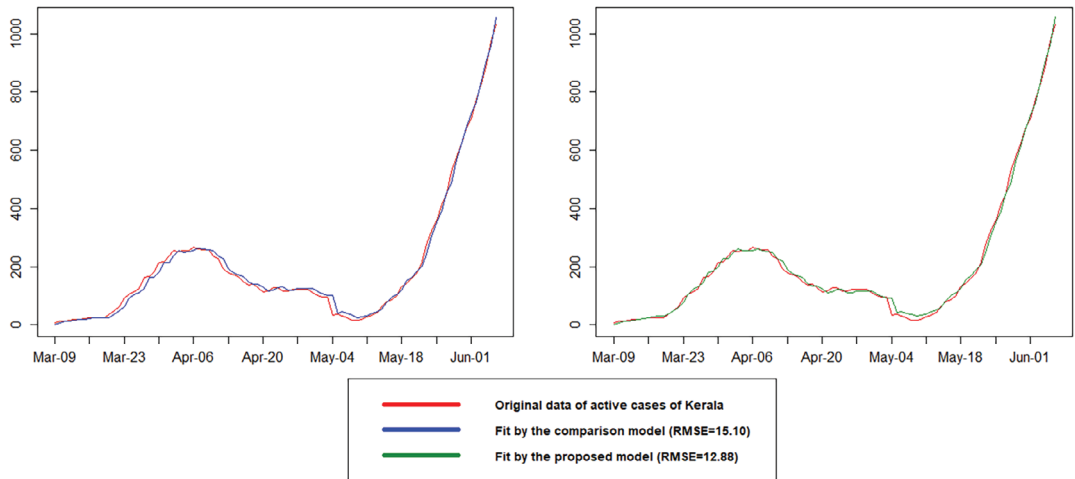
FIGURE 8 Fitted data (active cases) by both the comparison model and the proposed model of two change-point study (for Kerala)

For this setup, the estimates of the regression parameters of our proposed model by CLS method are $\hat{\boldsymbol{\beta}}_{2cls} = (\hat{\alpha}_{cls}, \hat{\beta}_{0cls}, \hat{\beta}_{1cls}, \hat{\beta}_{2cls}, \hat{\beta}_{3cls}) = (0.9021, -0.5960, -0.2897, 0.2164, 0.2171)$, and the estimates for the comparison model are $(\hat{\alpha^*}_{cls}, \hat{\gamma}_{0cls}, \hat{\gamma}_{1cls}) = (0.9947, -9.7129, 0.1759)$. For one-step ahead forecasting ($h = 1$), PRMSEs for the proposed model and the comparison model are 154.06 and 275.80, respectively. For two-step ahead forecasting ($h = 2$), PRMSEs for the proposed model and the comparison model are 186.99 and 348.68, respectively. So for both one-step and two-step ahead forecasting studies, our proposed model performs much better than the comparison model.

9 | CONCLUDING REMARKS

PINAR(1) process (introduced by McKenzie, 1985 and Al-Osh & Alzaid 1987) has received significant attention owing to its simplicity and is used widely in the field of count time series data. But this process is unable to model the count time series data like the COVID-19 data containing change-points and time-varying covariates. In this article, we have developed a new PINAR(1) model based on binomial thinning operator to handle the problem of change-point analysis through time-varying covariates. The development of our proposed model is inspired by Chan and Tong (1986); Hansen (2000), and Fong et al. (2017) who mainly worked on continuous data. We have used the concept of Smooth maximum (n.d.) in the proposed model to develop the smoothing change-point function which enables the model to capture the changing curvatures in the data. The key feature of our proposed model is its ability to accommodate both change-points and time-varying covariates. As described earlier, we can see these features in the COVID-19 data sets from which we have obtained the idea to develop our proposed model for both the studies of one change-point and two change-points. We have studied the distributional forms of our proposed model along with the h-step ahead forecasting distribution. Because of the difficulty in estimating the regression parameters through the maximum likelihood method, we have employed the CLS estimation method. We have performed an extensive simulation study to examine the CIs for true change-points for varying sample sizes and seen that as sample sizes increase widths of the CIs decrease. Regarding the estimation of parameters, the simulation results have shown consistency

of the CLS estimation method. From the data applications, we can see that our proposed model has led to much better performance over the comparison model with respect to standard statistical measure like RMSE. Our proposed model has also given a much better performance than the comparison model in the forecasting area with respect to the accuracy measure PRMSE in both the simulation study and the data analysis part. We can further extend our proposed model for more than two change-points in the same way as the model for one change-point analysis has been extended to that for two change-point analysis. Therefore we hope that our proposed model could be a viable choice for modeling these kinds of count time series data sets.

ACKNOWLEDGEMENTS

ORCID

Atanu Biswas https://orcid.org/0000-0001-5106-7684

REFERENCES

Al-Osh, M. A., & Alzaid, A. A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, *8*(3), 261–275.

Alzaid, A. A., & Al-Osh, M. (1990). An integer-valued *p*th-order autoregressive structure (INAR(*p*)) process. *Journal of Applied Probability*, *27*(2), 314–324.

Cameron, A. C., & Trivedi, P. K. (1986). Econometric models based on count data: Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics*, *1*(1), 29–53.

Chakraborty, B., Laber, E. B., & Zhao, Y. (2013). Inference for optimal dynamic treatment regimes using an adaptive *m*-out-of-*n* bootstrap scheme. *Biometrics*, *69*(3), 714–723.

Chan, K. S., & Tong, H. (1986). On estimating thresholds in autoregressive models. *Journal of Time Series Analysis*, *7*(3), 179–190.

Fokianos, K. (2011). Some recent progress in count time series. *A Journal of Theoretical and Applied Statistics*, *45*(1), 49–58.

Fong, Y., Huang, Y., Gilbert, P. B., & Permar, S. R. (2017). chngpt: threshold regression model estimation and inference. *BMC Bioinformatics*, *18*, 1–7.

Freeland, R. K., & McCabe, B. (2005). Asymptotic properties of CLS estimators in the Poisson AR(1) model. *Statistics & Probability Letters*, *73*(2), 147–153.

Freeland, R. K., & McCabe, B. P. M. (2004). Forecasting discrete valued low count time series. *International Journal of Forecasting*, *20*(3), 427–434.

GoK Dashboard. (n.d.). Retrieved from https://dashboard.kerala.gov.in/index.php

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica*, *68*(3), 575–603.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. New York, NY: Springer-Verlag.

Jazi, M. A., Jones, G., & Lai, C.-D. (2012). First-order integer valued AR processes with zero inflated Poisson innovations. *Journal of Time Series Analysis*, *33*(6), 954–963.

Jin-Guan, D., & Yuan, L. (1991). The integer-valued autoregressive (INAR(*p*)) model. *Journal of Time Series Analysis*, *12*(2), 129–142.

Maiti, R., Biswas, A., & Das, S. (2015). Time series of zero-inflated counts and their coherent forecasting. *Journal of Forecasting*, *34*(8), 694–707.

McKenzie, E. (1985). Some simple models for discrete variate time series. *Journal of the American Water Resources Association*, *21*(4), 645–650.

McKenzie, E. (1986). Autoregressive moving-average processes with negative-binomial and geometric marginal distributions. *Advances in Applied Probability*, *18*(3), 679–705.

Ristić, M. M., Bakouch, H. S., & Nastić, A. S. (2009). A new geometric first-order integer-valued autoregressive (NGINAR(1)) process. *Journal of Statistical Planning and Inference*, *139*(7), 2218–2226.

Schweer, S., & Weiß, C. H. (2014). Compound Poisson INAR(1) processes: Stochastic properties and testing for overdispersion. *Computational Statistics & Data Analysis*, *77*, 267–284.

Smooth maximum. (n.d.). Retrieved from https://en.wikipedia.org/wiki/Smooth_maximum

Steutel, F. W., & van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability*, *7*(5), 893–899.

Wald, A., & Wolfowitz, J. (1940). On a test whether two samples are from the same population. *The Annals of Mathematical Statistics*, *11*(2), 147–162.

Worldometer. (n.d.). Retrieved from https://www.worldometers.info/coronavirus/country/italy/

APPENDIX A. PROOF OF THEOREM 1

We have

$$Y_n = \alpha \circ Y_{n-1} + \epsilon_n = \sum_{i=1}^{Y_{n-1}} N_{1i} + \epsilon_n,$$

where $\{N_{1i}\}$ is a sequence of iid Bernoulli(α) random variables, independent of Y_{n-1}, and we know that given $\mathcal{X}_n$, $\alpha \circ Y_{n-1}$, and ϵ_n are independent. So we can write the PGF of $Y_n | \mathcal{X}_n$ as

$$\Phi_{Y_n | \mathcal{X}_n}(s) = E\left[s^{\left(\sum_{i=1}^{Y_{n-1}} N_{1i} + \epsilon_n \right)} \middle| \mathcal{X}_n \right]$$

$$= E_{Y_{n-1} | \mathcal{X}_n} E\left[s^{\left(\sum_{i=1}^{Y_{n-1}} N_{1i} \right)} \middle| Y_{n-1}, \mathcal{X}_n \right] \Phi_{\epsilon_n | \mathcal{X}_n}(s)$$

$$= E_{Y_{n-1} | \mathcal{X}_n} (1 - \alpha + \alpha s)^{Y_{n-1}} \Phi_{\epsilon_n | \mathcal{X}_n}(s)$$

$$= \Phi_{Y_{n-1} | \mathcal{X}_n} (1 - \alpha + \alpha s) \Phi_{\epsilon_n | \mathcal{X}_n}(s).$$

So we have the recursive formula

$$\Phi_{Y_n | \mathcal{X}_n}(s) = \Phi_{Y_{n-1} | \mathcal{X}_n} (1 - \alpha + \alpha s) \Phi_{\epsilon_n | \mathcal{X}_n}(s). \tag{A1}$$

Putting $n = 2$ in Equation (A1), we have

$$\Phi_{Y_2 | \mathcal{X}_2}(s) = \Phi_{Y_1 | \mathcal{X}_2} (1 - \alpha + \alpha s) \Phi_{\epsilon_2 | \mathcal{X}_2}(s)$$

$$= \exp[-\lambda_1 (1 - (1 - \alpha + \alpha s)) - \lambda_2 (1 - s)] \quad [\text{Since } Y_1 | \mathcal{X}_2 \sim \text{Poisson}(\lambda_1)]$$

$$= \exp[-\lambda_1 \alpha (1 - s) - \lambda_2 (1 - s)]$$

$$= \exp[-(\alpha \lambda_1 + \lambda_2)(1 - s)].$$

Therefore

$$Y_2 | \mathcal{X}_2 \sim \text{Poisson}(\alpha \lambda_1 + \lambda_2).$$

Now putting $n = 3$ in Equation (A1), we have

$$
\begin{aligned}
\Phi_{Y_3|\mathcal{X}_3}(s) &= \Phi_{Y_2|\mathcal{X}_3}(1 - \alpha + \alpha s) \; \Phi_{\varepsilon_3|\mathcal{X}_3}(s) \\
&= \exp[-(\alpha\lambda_1 + \lambda_2)(1 - (1 - \alpha + \alpha s)) - \lambda_3(1 - s)] \\
&= \exp[-(\alpha^2\lambda_1 + \alpha\lambda_2)(1 - s) - \lambda_3(1 - s)] \\
&= \exp[-(\alpha^2\lambda_1 + \alpha\lambda_2 + \lambda_3)(1 - s)].
\end{aligned}
$$

Therefore

$$
Y_3|\mathcal{X}_3 \sim \text{Poisson}(\alpha^2\lambda_1 + \alpha\lambda_2 + \lambda_3).
$$

Now for $n = (k - 1)$, we assume that

$$
Y_{k-1}|\mathcal{X}_{k-1} \sim \text{Poisson}\left(\alpha^{k-2}\lambda_1 + \alpha^{k-3}\lambda_2 + \ldots + \lambda_{k-1}\right).
$$

For $n = k$,

$$
\begin{aligned}
\Phi_{Y_k|\mathcal{X}_k}(s) &= \Phi_{Y_{k-1}|\mathcal{X}_k}(1 - \alpha + \alpha s) \; \Phi_{\varepsilon_k|\mathcal{X}_k}(s) \\
&= \exp[-\left(\alpha^{k-2}\lambda_1 + \alpha^{k-3}\lambda_2 + \ldots + \lambda_{k-1}\right)(1 - (1 - \alpha + \alpha s)) - \lambda_k(1 - s)] \\
&= \exp[-\left(\alpha^{k-1}\lambda_1 + \alpha^{k-2}\lambda_2 + \ldots + \alpha\lambda_{k-1}\right)(1 - s) - \lambda_k(1 - s)] \\
&= \exp[-\left(\alpha^{k-1}\lambda_1 + \alpha^{k-2}\lambda_2 + \ldots + \lambda_k\right)(1 - s)].
\end{aligned}
$$

So,

$$
Y_k|\mathcal{X}_k \sim \text{Poisson}\left(\alpha^{k-1}\lambda_1 + \alpha^{k-2}\lambda_2 + \ldots + \lambda_k\right).
$$

This completes the proof.

APPENDIX B. PROOF OF THEOREM 2

The ACVF is given by

$$
\begin{aligned}
\gamma_y(h) = \text{Cov}(Y_t, Y_{t+h}|\mathcal{X}_{t+h}) &= E(Y_t \, Y_{t+h}|\mathcal{X}_{t+h}) - E(Y_t|\mathcal{X}_{t+h})E(Y_{t+h}|\mathcal{X}_{t+h}) \\
&= E(Y_t \, E(Y_{t+h}|Y_t, \mathcal{X}_{t+h})|\mathcal{X}_{t+h}) - E(Y_t|\mathcal{X}_{t+h})E(Y_{t+h}|\mathcal{X}_{t+h}) \\
&= E\left[Y_t\left(\alpha^h Y_t + \sum_{i=1}^{h}\alpha^{h-i}\lambda_{t+i}\right)\bigg|\mathcal{X}_{t+h}\right] - E(Y_t|\mathcal{X}_{t+h})E(Y_{t+h}|\mathcal{X}_{t+h}) \\
&= \alpha^h E\left(Y_t^2|\mathcal{X}_{t+h}\right) - E(Y_t|\mathcal{X}_{t+h}) \\
&\quad \left[\alpha^{t+h-1}\lambda_1 + \ldots + \lambda_{t+h} - \alpha^{h-1}\lambda_{t+1} - \ldots - \lambda_{t+h}\right] \\
&= \alpha^h\left[V(Y_t|\mathcal{X}_{t+h}) + E^2(Y_t|\mathcal{X}_{t+h})\right] - E(Y_t|\mathcal{X}_{t+h})\left(\alpha^{t+h-1}\lambda_1 + \ldots + \alpha^h\lambda_t\right) \\
&= \alpha^h\left[(\alpha^{t-1}\lambda_1 + \ldots + \lambda_t)(1 + \alpha^{t-1}\lambda_1 + \ldots + \lambda_t)\right] \\
&\quad - (\alpha^{t-1}\lambda_1 + \ldots + \lambda_t)\left(\alpha^{t+h-1}\lambda_1 + \ldots + \alpha^h\lambda_t\right) \\
&= \alpha^h\left(\alpha^{t-1}\lambda_1 + \alpha^{t-2}\lambda_2 + \ldots + \lambda_t\right).
\end{aligned}
$$

This completes the proof.

APPENDIX C. PROOF OF THEOREM 3

The conditional PGF is given by

$$\Phi_{Y_{n+h}|Y_n,\mathcal{X}_{n+h}}(s) = E\left(s^{\alpha^h \circ Y_n + \sum_{i=1}^h \alpha^{h-i} \circ \varepsilon_{n+i}}|Y_n, \mathcal{X}_{n+h}\right)$$

$$= E\left(s^{\alpha^h \circ Y_n}|Y_n, \mathcal{X}_{n+h}\right) \prod_{i=1}^h E\left(s^{\alpha^{h-i} \circ \varepsilon_{n+i}}|\mathcal{X}_{n+h}\right).$$

Since $\alpha^h \circ Y_n|Y_n, \mathcal{X}_{n+h} \sim \text{Bin}\left(Y_n, \alpha^h\right)$, we can write $E\left(s^{\alpha^h \circ Y_n}|Y_n, \mathcal{X}_{n+h}\right) = \left(1 - \alpha^h + \alpha^h s\right)^{Y_n}$.
Now,

$$\prod_{i=1}^h E\left(s^{\alpha^{h-i} \circ \varepsilon_{n+i}}|\mathcal{X}_{n+h}\right) = \prod_{i=1}^h E_{\varepsilon_{n+i}|\mathcal{X}_{n+h}} E\left(s^{\alpha^{h-i} \circ \varepsilon_{n+i}}|\varepsilon_{n+i}, \mathcal{X}_{n+h}\right)$$

$$= \prod_{i=1}^h E_{\varepsilon_{n+i}|\mathcal{X}_{n+h}} \left(1 - \alpha^{h-i} + \alpha^{h-i}s\right)^{\varepsilon_{n+i}}$$

$$= \prod_{i=1}^h \exp\left[-\lambda_{n+i}\left(1 - (1 - \alpha^{h-i} + \alpha^{h-i}s)\right)\right]$$

$$= \prod_{i=1}^h \exp\left[-\lambda_{n+i}\ \alpha^{h-i}(1-s)\right]$$

$$= \exp\left[-\sum_{i=1}^h \lambda_{n+i}\ \alpha^{h-i}(1-s)\right].$$

This completes the proof.

APPENDIX D. PROOF OF THEOREM 4

The h-step ahead forecasting distribution is given by

$$p_h(j_1|j_2) = P\left(\alpha^h \circ Y_n + \sum_{k=1}^h \alpha^{h-k} \circ \varepsilon_{n+k} = j_1|Y_n = j_2, \mathcal{X}_{n+h}\right)$$

$$= \sum_{q=0}^{\min(j_1,j_2)} P\left(\alpha^h \circ Y_n = q|Y_n = j_2, \mathcal{X}_{n+h}\right) P\left(\sum_{k=1}^h \alpha^{h-k} \circ \varepsilon_{n+k} = j_1 - q|\mathcal{X}_{n+h}\right)$$

$$= \sum_{q=0}^{\min(j_1,j_2)} \binom{j_2}{q} \alpha^{hq}\left(1 - \alpha^h\right)^{j_2-q} \exp\left[-\sum_{k=1}^h \alpha^{h-k}\lambda_{n+k}\right]\left(\sum_{k=1}^h \alpha^{h-k}\lambda_{n+k}\right)^{j_1-q} [(j_1 - q)!]^{-1}$$

$$I_{[(j_1-q)=(0,1,\dots)]}$$

$$= \sum_{q=0}^{\min(j_1,j_2)} \binom{j_2}{q} \left(\alpha^h\right)^q \left(1 - \alpha^h\right)^{j_2-q} \exp(-\lambda(h))\lambda(h)^{j_1-q}[(j_1 - q)!]^{-1} I_{[(j_1-q)=(0,1,\dots)]}.$$

This completes the proof.