Open Journal of Statistics, 2026, 16(2), 89-106
https://www.scirp.org/journal/ojs
ISSN Online: 2161-7198
ISSN Print: 2161-718X

Nonstationary Autoregressive Modeling of Time Series Count Data with Covariates: Addressing Seasonality in Branching Negative Binomial Models

Bakary Traore*, Ibrahima Sory Mamikouny Camara, Alpha Oumar Baldé

Département Mathématiques, Université de N'Zérékoré (UZ), N'Zérékoré, Guinea
Email: *btraore850@gmail.com

How to cite this paper: Traore, B., Camara, I.S.M. and Baldé, A.O. (2026) Nonstationary Autoregressive Modeling of Time Series Count Data with Covariates: Addressing Seasonality in Branching Negative Binomial Models. *Open Journal of Statistics*, **16**, 89-106.
https://doi.org/10.4236/ojs.2026.162005

Received: June 12, 2025
Accepted: April 7, 2026
Published: April 10, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

Abstract

Various models for time series of count data account for discreteness, overdispersion and serial dependence. In addition to these, accounting for covariates incorparation pattern are the complexities that arise while dealing with data which involve seasonality aspects. Specifying a model that can handle such kind of time series of count data is very important in several real-life application. However in this paper, we present a non-stationary autoregressive model where covariates information are incorporated in the Branching Negative Binomial (bNB) autoregressive model in order to assess the seasonality in the process of time series event. A simulation study is done to evaluate how well the proposed strategy performs, and inference is based on maximum likelihood estimate. The model is used to analyse a real-world dataset, which is an infectious disease with covariates information, including temperature and rainfall.

Keywords

Non-Stationary Time Series, Covariates, Branching Process, Time Series of Count Data, Overdispersion

1. Introduction

Various models for time series of count data account for discreteness, overdispersion and serial dependence. Among those applications that require these aspects, road accident counts (see e.g. [1]), sandstorm counts (see e.g. [2]), stock market application (see e.g. [3]; [4]), crime analysis (see e.g. [5]; [6]), among others. Several models have been presented by scholars to handle the analysis of such kind of data. Those models include Zero truncated Poison INAR, INARGH, NGINAR, GLM, Zero-modified geometric INAR, bNB, INAR etc., (see [7]-[10]). In addition, complexities arise while specifying a modeling strategy that is able to deal with count time series data which are not stationary. Some of the nonstationary data involve trend, seasonal, cyclical patterns. It is common to encounter time series of counts that are observed together with covariate information. These kinds of data commonly arise in fields of application such as epidemiology, accident analysis, ecology, forest fires among others. A large number of studies revealed that, these nonstationary aspects are due to some natural phenomena in time series event process such as temperature, rainfall, whether, humidity, etc. In particular, weather variables may affect accident counts (see e.g. [11]), environmental factors influence species counts recorded over time (see e.g. [12]), occurrence of wildfires over time is influenced by meteorological variables (see e.g. [13]), etc. In such case non-stationarity needs to be incorporated into the modeling framework as well as mechanisms to capture dependence and other distributional characteristics. There are various modeling strategies that have been proposed to deal with the issues arising when handling time series of counts observed with covariate information.

Studying the effect of weather conditions on daily crash counts using a discrete time-series model is proposed in this paper [14]. The model was applied to daily car crash data, metereological data and traffic exposure data from the Netherlands to analyze the risk impact of weather conditions on the observed counts. According to the authors of [15], count data time series may have non-stationarity from trends or covariates. While modelling count data time series with Markov processes based on binomial thinning, they proposed an extension of stationary time series based on binomial thinning such that the univariate marginal distributions are always in the same parametric family, such as negative binomial. In order to conduct a statistical analysis of data collected by the Massachusetts Water Resources Authority (MWRA) between 1996 and 2002 to evaluate the effects of court-mandated improvements in sewage treatment, [16] proposed a nonstationary negative binomial time series with time-dependent covariates. Utilizing financial and economic variables as exogenous covariates, [17] developed a class of Poisson autoregressive models with exogenous covariates (PARX) that can be used to model and forecast monthly corporate defaults in the US from 1982 to 2011 using these variables as time series attributes. The results showed the effectiveness of their model in capturing the time series dynamics of corporate defaults, including the well-known clustering of default counts in the data. Furthermore, they discovered that while generally speaking, current defaults do affect the likelihood that other firms will experience default in the future, in recent years, economic and financial factors at the macro level have been able to explain a significant portion of the correlation of US firm defaults over time. [18] presented parameter time series models with covariates for integer Z. The used real data applications on accidents and financial returns are given. [19] presented a Bayesian analysis of a time series of counts to assess its dependence on an explanatory variable. As application, they analyse the incidence of an infectious disease with an explanatory variable which is he number of grams of antibiotic (third generation) cephalosporin. Statistically, they realized that there is significant relationship between disease occurrence and use of the antibiotic, lagged by three months. A popular approach entails developing regression models by introducing covariates in the marginal distribution, dependence component or both. Typical examples where the parameters are models as a function of covariates include Integer Autoregressive and Generalized Linear Modeling frameworks. However, we extend the Branching Negative Binomial Autoregressive model presented by the authors in [10] to a NonStationary model, where covariate information is incorporated to the dependence parameter.

The rest of the paper is organized as follows; Section 2 outlines the approach adopted in developing the non-stationary branching negative binomial (bNB) autoregressive model as well the procedure for inference, Section 3 provides the results for the simulation of stationary bNB autoregressive model, Section 4 presents the results from the application of the model to real-life data, Section 5 performs the goodness-of-fit tests of models. Finally, Section 6 concludes the work.

2. Methods and Materials

The existing stationary model presented in [10] is defined by

$$Y_t = N_{t-1} + E_t \tag{1}$$

where

$$N_{t-1} \sim Bin\left(Y_{t-1}, \frac{\rho\pi}{1-\rho+\rho\pi}\right) \text{ and } E_t \sim NB\left(m + N_{t-1}, \frac{\pi}{1-\rho+\rho\pi}\right)$$

They denote the model as $bNB(m,\pi,\rho)$.

The branching concept is that, since N_{t-1} is generated from Y_{t-1}, which is considered as previous observation. The generating process follows binomial law with probability $\frac{\rho\pi}{1-\rho+\rho\rho}$. In terms of dependence parameter ρ, it's the rate to determine how many offsprings are generated from the previous Y_{t-1} which also specifies the dependence in the autoregressive process from Y_t.

The marginal distribution of the arising process is Negative Binomial because the error component E_t has a Negative Binomial distribution.

In this framework, we extend the model (Equation 1) proposed in [10] by incorporating covariates information in the dependence parameter to obtain a non-stationary model capable to handle seasonal pattern in times series of count data.

2.1. Model Specification

Incorporating covariates into the dependence parameter, we considered the logit function in order to preserve the nature of the parameter ($\rho \in [0,1]$) from the stationary model. Hence, it will be a time varying function, denoted by ρ_t; and it follows

$$\rho_t = \frac{1}{1+\exp\left(-X^{\mathrm{T}}\beta\right)} \tag{2}$$

where $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}, \quad X = \begin{pmatrix} 1 \\ X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}$

where $\beta_1, \beta_2, \cdots, \beta_p$ are regression coefficients and $x_1, x_2, \cdots, x_p$ the covariates.

The derived non-stationary bNB model is given in (Equation 3) with parameters $\theta = (m, \pi, \beta)$.

$$Y_t = N_{t-1} + E_t \tag{3}$$

where

$$N_{t-1} \sim Bin\left(Y_{t-1}, \frac{\rho_t \pi}{1-\rho_t+\rho_t\pi}\right) \text{ and } E_t \sim NB\left(m+N_{t-1}, \frac{\pi}{1-\rho_t+\rho_t\pi}\right)$$

We denote the obtained model as $bNB(m, \pi, \rho_t)$ for ρ_t defines in (Equation 2).

In this process (Equation 1), distribution of the two components are given bellow;

$$P(N_{t-1} = k) = \binom{Y_{t-1}}{k}\left(\frac{\rho_t\pi}{1-\rho_t+\rho_t\pi}\right)^k \left(1-\frac{\rho_t\pi}{1-\rho_t+\rho_t\pi}\right)^{Y_{t-1}-k} \tag{4}$$

$$P(E_t = r) = \binom{(m+N_{t-1})+r-1}{r}\left(\frac{\pi}{1-\rho_t+\rho_t\pi}\right)^{(m+N_{t-1})} \times \left(1-\frac{\pi}{1-\rho_t+\rho_t\pi}\right)^r \tag{5}$$

2.2. Model Estimation

2.2.1. Likelihood Function

Let $T = \{t_0 < t_1 < \cdots < t_n\}$ be an increasing set of times and $y = \{y_0, y_1, \cdots, y_n\} \subset \mathbb{N}$ an arbitrary set of non-negative integers with covariates $X = \{X_1, X_2, \cdots, X_n\} \in \mathbb{R}$. Then the joint pmf for $Y \sim bNB(m, \pi, \rho_t)$ is the product of the marginal and the conditionals.

$$P\left[Y = y \middle| X, m, \pi, \beta\right] = p(y_o)\prod_{t=1}^{n} P\left(y_t \middle| x_t, y_{t-1}\right) \tag{6}$$

where $m > 0$, $\pi \in [0,1]$, and $\beta \in \mathbb{R}^{N+1}$.

Conditional or transition probability mass function (pmf)

$$\begin{aligned} p\left(y_t \middle| y_{t-1}\right) &= P\left[Y_t = y_t \middle| Y_{t-1} = y_{t-1}\right] \\ &= \sum_{N_{t-1}=0}^{\infty}\binom{y_{t-1}}{N_{t-1}}\left(\frac{\rho_t\pi}{1-\rho_t+\rho_t\pi}\right)^{N_{t-1}} \times \left(\frac{1-\rho_t}{1-\rho_t+\rho_t\pi}\right)^{y_{t-1}-N_{t-1}} \\ &\quad \times \binom{m+N_{t-1}+E_t-1}{E_t}\left(\frac{\pi}{1-\rho_t+\rho_t\pi}\right)^{m+N_{t-1}} \times \left(\frac{(1-\rho_t)(1-\pi)}{1-\rho_t+\pi}\right)^{E_t} \end{aligned}$$

$$= \frac{y_{t-1}!\Gamma(m+y_{t-1})\pi^m (1-\pi)^{y_{t-1}} (1-\rho_t)^{y_t+y_{t-1}}}{(1-\rho_t+\rho_t\pi)^{m+y_{t-1}+y_t}}$$
$$\times \sum_{N_{t-1}=0}^{\infty} \left\{ \frac{1}{N_{t-1}!(y_t-N_{t-1})!(y_{t-1}-N_{t-1})!\Gamma(m+N_{t-1})} \right.$$
$$\left. \times \left[\frac{\rho_t\pi^2}{(1-\rho_t)^2(1-\pi)} \right]^{N_{t-1}} \right\} \tag{7}$$
$$= \frac{\Gamma(m+y_t)}{\Gamma(m)y_t!} \frac{\pi^m(1-\rho_t)^{y_t+y_{t-1}}(1-\pi)^{y_t}}{(1-\rho_t+\rho_t\pi)^{m+y_t+y_{t-1}}}$$
$$\times {}_2F_1(-y_{t-1},-y_t;m;z)$$

where $z = \frac{\rho_t\pi^2}{(1-\rho_t)^2(1-\pi)}$ and ${}_2F_1(a,b;c;k)$ is Gauss' hypergeometric function with parameters a,b,c and k.

Marginal distribution

The marginal distribution for y_o is:

$$p(y_o) = \frac{\Gamma(m+y_o)}{\Gamma(m)y_o!}\pi^m(1-\pi)^{y_o} \tag{8}$$

Substituting (Equation 7) and (Equation 8) in (Equation 6), the likelihood is defined as follow

$$P[Y=y,X|m,\pi,\beta]$$
$$= \frac{\Gamma(m+y_o)}{\Gamma(m)y_o!}\pi^m(1-\pi)^{y_o} \times \prod_{t=1}^{n} \left\{ \frac{\Gamma(m+y_t)}{\Gamma(m)y_t!} \frac{\pi^m(1-\rho_t)^{y_{t-1}+y_t}(1-\pi)^{y_t}}{(1-\rho_t+\rho_t\pi)^{m+y_{t-1}+y_t}} \right.$$
$$\left. \times_2 F_1(-y_{t-1},-y_t;m;z) \right\}$$
$$= \prod_{t=0}^{n} \left\{ \frac{\Gamma(m+y_t)\pi^m}{\Gamma(m)y_t!} \right\} \times \frac{(1-\pi)^{\sum_{t=0}^{n} y_t}(1-\rho_t)^{\sum_{t=1}^{n}(y_{t-1}+y_t)}}{(1-\rho_t+\rho_t\pi)^{nm+\sum_{t=1}^{n}(y_{t-1}+y_t)}} \tag{9}$$
$$\times \prod_{t=1}^{n} {}_2F_1(-y_{t-1},-y_t;m;z)$$
$$= \prod_{t=0}^{n} \left\{ \frac{\Gamma(m+y_t)\pi^m}{\Gamma(m)y_t!} \right\} \prod_{t=1}^{n} {}_2F_1(-y_{t-1},-y_t;m;z)$$
$$\times \frac{(1-\pi)^{y_+}(1-\rho_t)^{2y_+-y_o-y_n}}{(1-\rho_t+\rho_t\pi)^{nm+2y_+-y_o-y_n}}$$

where $y_+ = \sum_{t=0}^{n} y_t$, $\rho_t = \frac{1}{1+\exp(-X^{\mathrm{T}}\beta)}$, $z = \frac{\rho_t\pi^2}{(1-\rho_t)^2(1-\pi)}$

2.2.2. Log-Likelihood Function

Taking log (natural logarithm) of the previous equation (Equation 9), we obtain the log-likelihood (Equation 10)

$$
\begin{aligned}
&\log L\left(m,\pi,\beta \middle| Y=y\right) \\
&=\sum_{t=0}^{n}\log\Gamma\left(m+y_t\right)+\left(n+1\right)m\log\pi+\sum_{t=1}^{n}\log{}_2F_1\left(-y_{t-1},-y_t;m;z\right) \\
&\quad-\left(n+1\right)\log\Gamma\left(m\right)+y_+\log\left(1-\pi\right)+\left(2y_+-y_o-y_n\right)\log\left(1-\rho_t\right) \\
&\quad-\left(nm+2y_+-y_o-y_n\right)\log\left(1-\rho_t+\rho_t\pi\right)
\end{aligned} \tag{10}
$$

The parameters to be estimated are the marginal, the dependence and regression parameters. The parameter vector is $\theta=\left(m,\pi,\beta_0,\cdots,\beta_p\right)$, which are obtained by optimizing the log-likelihood function using the optim() routine in R software.

In statistical modeling, the maximum likelihood estimation (MLE) method is commonly used to estimate the parameters of a probability distribution based on a given set of data. The MLE method involves finding the values of the parameters that maximize the log-likelihood function, which is the logarithm of the probability of observing the data given the parameter values.

However, in this case it is not easy to differentiate the log-likelihood function with respect to every parameter because of the presence of special functions such as the Gauss hypergeometric function. The derivative of the log-likelihood function with respect to a parameter may not have a closed-form solution, making it challenging to find the parameter value that maximizes the log-likelihood function.

Furthermore, the presence of the Gamma function in the expression of the derivative of the log-likelihood function can make it difficult to obtain a closed-form solution for the derivative of the log-likelihood function with respect to the parameter. The derivative of the Gamma function involves the digamma function, which does not have a closed-form solution in general.

Therefore, when there is a difficulty in deriving the log-likelihood function with respect to every parameter because of the Gauss hypergeometric function in the expression and its derivative has difficulty exhibiting a parameter because of the Gamma function, we can use numerical optimization techniques such as the optim() function in R software to estimate the parameters that maximize the log-likelihood function. The optim() function uses numerical methods to find the minimum or maximum of a given function, which in this case is the negative log-likelihood function.

By using the optim() function in R, we can estimate the parameter values that maximize the log-likelihood function, even when closed-form solutions for the derivative of the log-likelihood function with respect to the parameter are not available. This allows us to perform statistical inference and make predictions based on the estimated parameter values.

In the next section, the R software package is used in the simulation process to estimate parameters of the non-stationary bNB models.

3. Simulation Study

In this simulation study, our target is bear on one covariate variable where

$X = (1, X_1)$ and $\beta = (\beta_0, \beta_1)$. Base on this restriction, the Monte Carlo experiments have been conducted several time to evaluate the performance of the Branching Negative Binomial estimators for the parameters. The experiment entails generating n time series of size N from the models and then estimating the parameter vector $\theta = (m, \pi, \beta_0, \beta_1)$. For each combination ($\theta$, N), we compute the mean, bias, and the mean-squared error (MSE) given by,

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_i,\ Bias(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n}(\hat{\theta}_i - \theta) \text{ and } MSE(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n}(\hat{\theta}_i - \theta)^2$$

where $\hat{\theta}_i$ is the estimated parameter vector values from the i^{th} simulated series.

Using the software package R, we generate by Monte Carlo replicates n = 1000 random samples of length N = 250, N = 500 and N = 1000 from the non-stationary bNB, then calculate the mean, bias and mean-squared error of the estimator.

Simulation Results

To assess the performance of the proposed approach of incorporating covariates in terms of estimation, we focused on the case in which the covariates are introduced in the dependence parameter as it is more promising in terms of accommodating non-stationary patterns in both the marginal distribution m and the dependence parameter, ρ. We take motivation from epidemiology and biostatistics where harmonic functions are commonly employed within regression models as standard tools for capturing seasonal patterns (Equation 11) where we consider sinusoidal term $X = \sin(2\pi t/12)$ as covariates information in this simulation. Then the corresponding time dependence correlation parameter is given by

$$\rho_t = \frac{1}{1 + \exp(-\beta_0 - \beta_1 \sin(2\pi t/12))} \quad (11)$$

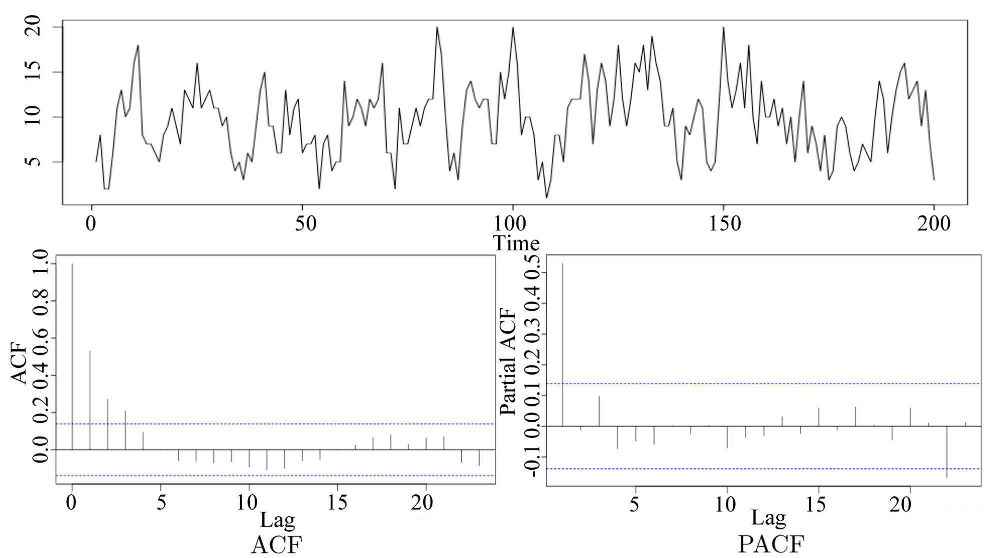


Figure 1. Time series plot and autocorrelation functions from a sample size of 200 simulated data with parameter $m = 15$, $\pi = 0.6$, $\beta_0 = 0.7$ and $\beta_1 = 0.8$.

Figure 1 shows plots of time series data generated from the nonstationary bNB model and the corresponding autocorrelation functions. Particularly, we display seasonal effects by considering parameters $m = 15$, $\pi = 0.6$, $\beta_0 = 0.7$ and $\beta_1 = 0.8$.

Table 1. Summary statistics for the estimator for different parameter values $\theta = (m, \pi, \beta_0, \beta_1)$ and different sample sizes N. These statistics are obtained from 1000 Monte Carlo replications of the developed model.

N	True	m	π	β_0	β_1	m	π	β_0	β_1
	value	**5**	**0.6**	**0.7**	**0.8**	**9**	**0.4**	**0.7**	**0.8**
250	**Mean**	4.1815	0.7363	0.8087	0.9935	7.6554	0.3429	1.2013	1.2368
	Bias	0.8184	−0.1363	−0.1087	0.1935	1.3446	0.0370	−0.5013	−0.4368
	MSE	1.3186	0.0323	2.9049	1.8290	3.2177	0.0091	3.0909	0.6041
500	**Mean**	4.4659	0.5512	0.6475	0.7201	8.6720	0.4380	0.6081	0.9928
	Bias	0.5340	0.0487	0.0524	0.0799	0.3279	−0.0380	0.0918	−0.1928
	MSE	0.6013	0.0045	1.9056	0.0100	2.5721	0.0025	0.0133	0.0378
1000	**Mean**	5.4009	0.6278	0.7227	0.8090	8.8609	0.3998	0.7503	0.7119
	Bias	−0.4009	−0.0278	−0.0227	0.0090	0.1391	0.0001	−0.0503	0.0881
	MSE	0.3939	0.0018	0.0101	0.0019	0.3703	0.0024	0.0059	0.0002
		15	**0.6**	**0.7**	**0.8**	**9**	**0.7**	**0.7**	**0.8**
250	**Mean**	16.5319	0.7894	0.9432	1.0074	8.2435	0.6001	1.0011	1.0918
	Bias	−1.5319	−0.1894	−0.2432	0.2074	0.7565	0.0999	−0.3011	−0.2918
	MSE	17.4271	0.0377	6.9168	1.5804	24.5055	0.0128	2.0281	1.4447
500	**Mean**	13.8483	0.5788	0.6450	0.7065	8.4458	0.6614	0.6478	0.9225
	Bias	1.1517	0.0212	0.0549	0.0935	0.5542	0.0386	0.0521	−0.1225
	MSE	12.4995	0.0071	0.0034	0.7118	8.0693	0.0100	0.0073	0.2582
1000	**Mean**	14.4685	0.6074	0.6645	0.7651	9.0618	0.6677	0.6907	0.7415
	Bias	0.5314	−0.0074	0.0354	0.0349	−0.0618	0.0322	−0.0093	0.0585
	MSE	9.4144	0.0040	0.0015	0.0118	1.6766	0.0088	0.0054	0.0146
		25	**0.6**	**0.7**	**0.8**	**9**	**0.9**	**0.7**	**0.8**
250	**Mean**	27.1642	0.5721	1.2894	1.1879	7.0774	0.8243	0.6385	0.9203
	Bias	−2.1642	0.0278	−0.5894	−0.3879	1.9226	0.0756	0.0614	−0.1203
	MSE	60.9869	0.0035	0.0139	0.6961	5.9608	0.0102	0.0109	0.8631
500	**Mean**	26.3776	0.5732	0.4264	0.7033	8.2728	0.8456	0.6683	0.6875
	Bias	−1.3776	0.0268	0.2736	0.0967	0.7271	0.0544	0.0317	0.1124
	MSE	52.0151	0.0026	0.0066	0.0077	3.6435	0.0044	0.0063	0.0774
1000	**Mean**	25.4659	0.6229	0.6206	0.7237	9.1964	0.8564	0.7042	0.7974
	Bias	−0.4659	−0.0229	0.0793	0.0763	−0.1964	0.0435	−0.0042	0.0025
	MSE	12.3419	0.0008	0.0044	0.0073	1.7976	0.0023	0.0036	0.0150

Table 2. Summary statistics for the estimator for different parameter values $\theta = (m, \pi, \beta_0, \beta_1)$ and different sample length of the series N. These statistics are obtained from 1000 Monte Carlo replications of the non-stationary developed model by varying regression parameters.

N	True	m	π	β_0	β_1	m	π	β_0	β_1
	value	**5**	**0.6**	**1**	**1**	**5**	**0.6**	**0.5**	**1**
250	**Mean**	5.8119	0.6654	0.7984	1.4729	6.2676	0.6428	0.9018	1.2708
	Bias	−0.8119	−0.0654	0.2016	−0.4729	−1.2676	−0.0428	−0.4018	−0.2708
	MSE	6.6358	0.0061	3.4765	5.7167	4.8348	0.0061	6.9498	13.8970
500	**Mean**	4.5136	0.6160	1.0958	1.3729	6.0625	0.6387	0.7070	0.7603
	Bias	0.4863	−0.0160	−0.0958	−0.3729	−1.0625	−0.0387	−0.2070	0.2397
	MSE	6.1471	0.0054	2.4043	2.3425	3.2988	0.0056	5.9714	4.7617
1000	**Mean**	5.3241	0.6119	1.0543	1.1611	5.6854	0.6042	0.3083	0.8388
	Bias	−0.3241	−0.0119	−0.0543	−0.1611	−0.6854	−0.0043	0.1917	0.1611
	MSE	2.3754	0.0049	1.1717	2.1101	2.4359	0.0039	2.1392	2.9154
		5	**0.6**	**1.5**	**1**	**5**	**0.6**	**0.5**	**1.5**
250	**Mean**	5.5564	0.6162	1.1633	0.5722	6.0423	0.6482	0.9887	2.4237
	Bias	−0.5564	−0.0162	0.3367	0.4278	−1.0423	−0.0482	−0.4887	−0.9237
	MSE	4.1024	0.0208	15.3164	8.8735	4.5660	0.0128	12.1446	9.8083
500	**Mean**	5.5288	0.5914	1.4095	1.2854	5.5250	0.5850	0.9780	1.7037
	Bias	−0.5288	0.0086	0.0905	−0.2854	−0.5250	0.0142	−0.4780	−0.2037
	MSE	3.6067	0.0079	7.7356	5.5361	3.1858	0.0070	5.8914	3.3913
1000	**Mean**	5.3571	0.6041	1.5397	1.1176	5.2954	0.5893	0.7162	1.5644
	Bias	−0.3571	−0.0041	−0.0397	−0.1176	−0.2954	0.0107	−0.2162	−0.0644
	MSE	2.6119	0.0062	4.0269	1.0375	2.6133	0.0067	2.7755	2.5546
		5	**0.6**	**2**	**1**	**5**	**0.6**	**0.5**	**2**
250	**Mean**	6.3143	0.5692	1.6288	1.7211	6.4382	0.6199	0.8669	2.2124
	Bias	−1.3143	0.0307	0.3711	−0.7211	−1.4382	−0.0199	−0.3669	−0.2124
	MSE	9.1454	0.0251	22.1182	10.6959	6.4997	0.0167	17.1612	21.8346
500	**Mean**	5.5721	0.5797	2.3411	1.2618	5.6294	0.5856	0.8481	2.1570
	Bias	−0.5721	0.0203	−0.3411	−0.2618	−0.6294	0.0143	−0.3481	−0.1570
	MSE	8.4854	0.0236	12.1728	7.2615	3.6445	0.0068	14.2013	6.0848
1000	**Mean**	5.1711	0.6039	1.9794	0.8534	4.7419	0.6104	0.5570	2.0005
	Bias	−0.1711	−0.0039	0.0205	0.1466	0.2581	−0.0104	−0.0570	−0.0005
	MSE	4.0622	0.0098	4.2468	3.4275	2.6811	0.0041	1.4188	2.3650

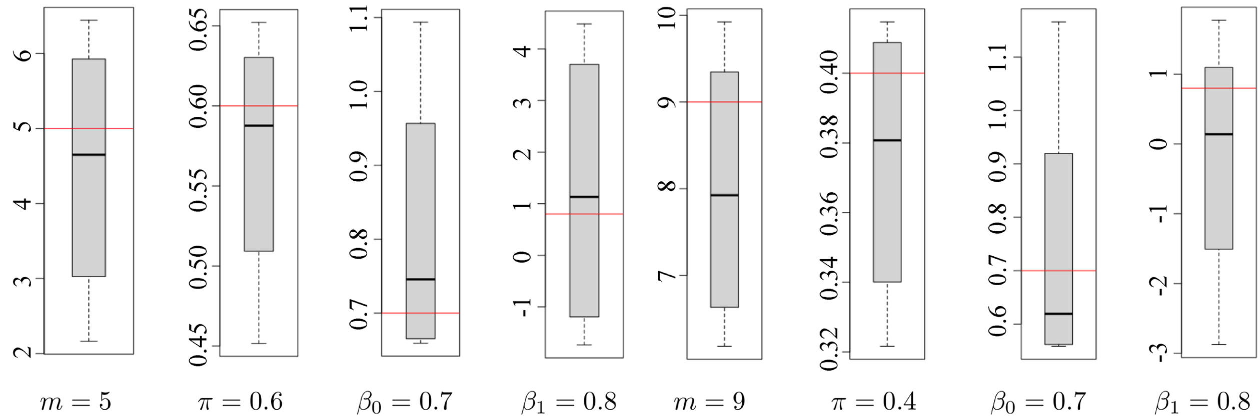


Figure 2. Boxplots with results from the simulation experiment for $N = 250$.

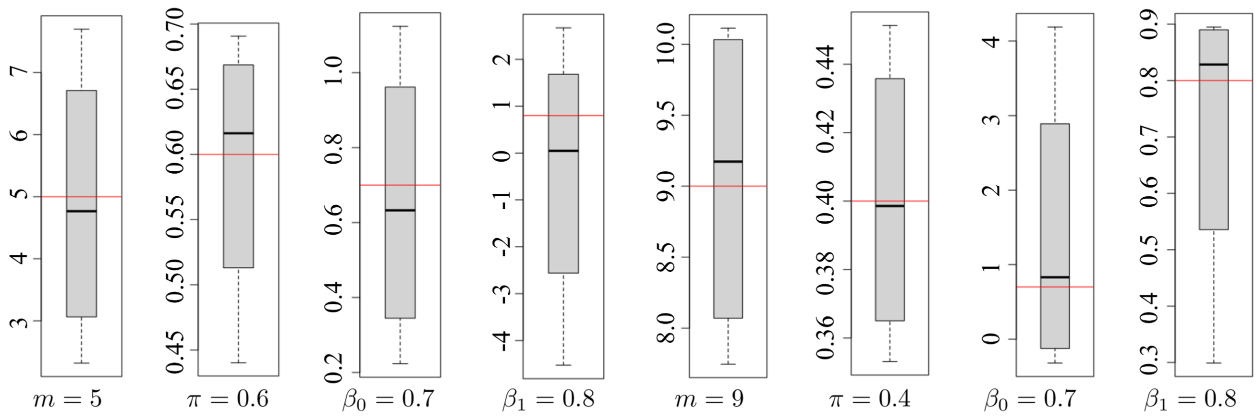


Figure 3. Boxplots with results from the simulation experiment for $N = 1000$.

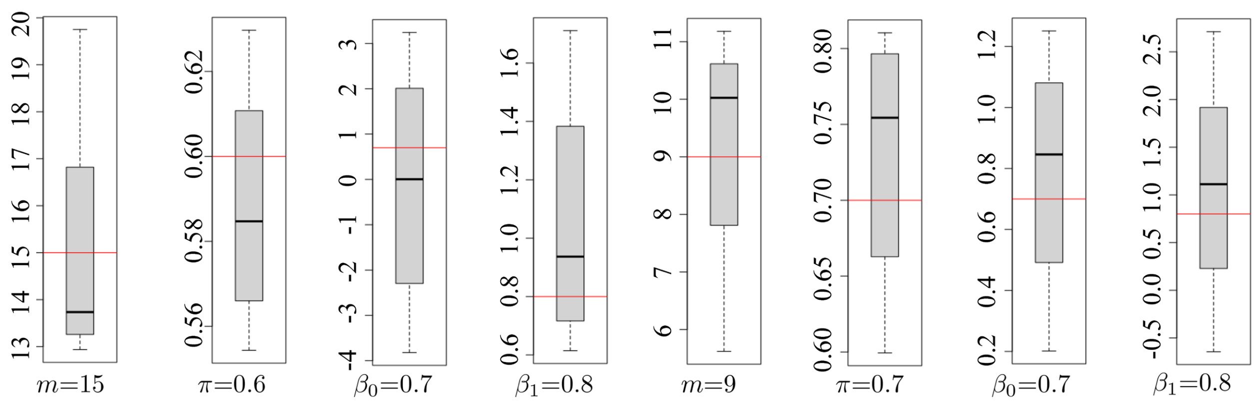


Figure 4. Boxplots with results from the simulation experiment for $N = 250$.

In **Table 1** and **Table 2**, we report the mean, bias and mean-squared error with parameters $\theta = (m, \pi, \beta_0, \beta_1)$ obtained using 1000 Monte Carlo replications of time series of length 250, 500 and 1000. As expected, the results indicate that the biases decrease as the sample size increases. We realized that in this scenario, when the parameter m varies by increasing, its mean square error increases while for rest of parameters decrease; and its bias increase while for the rest of parameters decrease. Also, when parameter π varies by increasing, its bias decreases with mean square errors of β_0 and β_1. These results are also illustrated in the boxplots via **Figure 2**, **Figure 3**, **Figure 4**, **Figure 5**, **Figure 6** and **Figure 7**. All these show that there is an improvement in the results as the sample size increases.

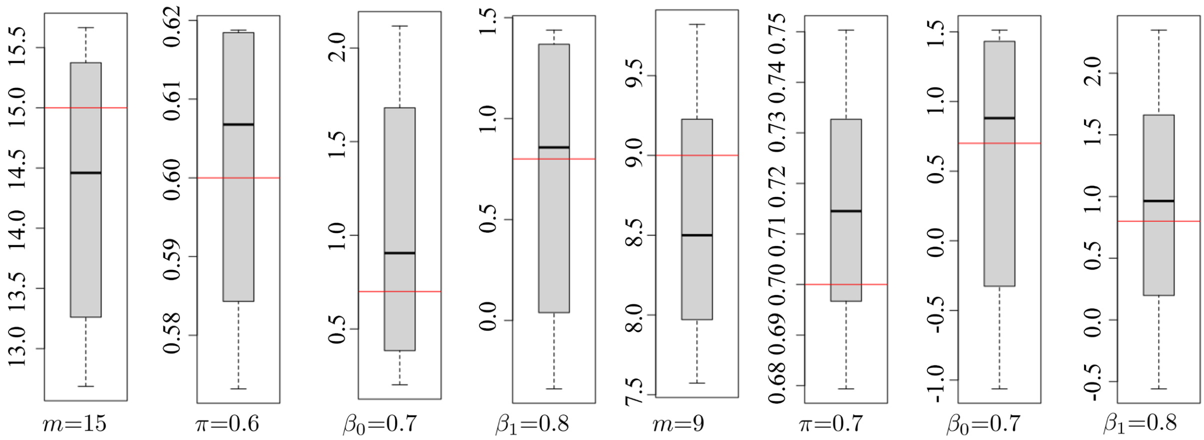


Figure 5. Boxplots with results from the simulation experiment for $N = 1000$.

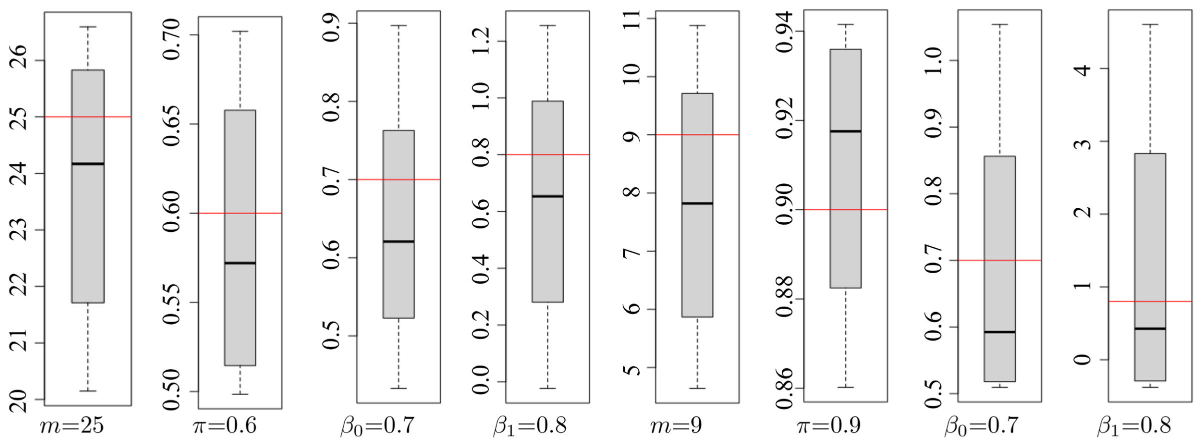


Figure 6. Boxplots with results from the simulation experiment for $N = 250$.

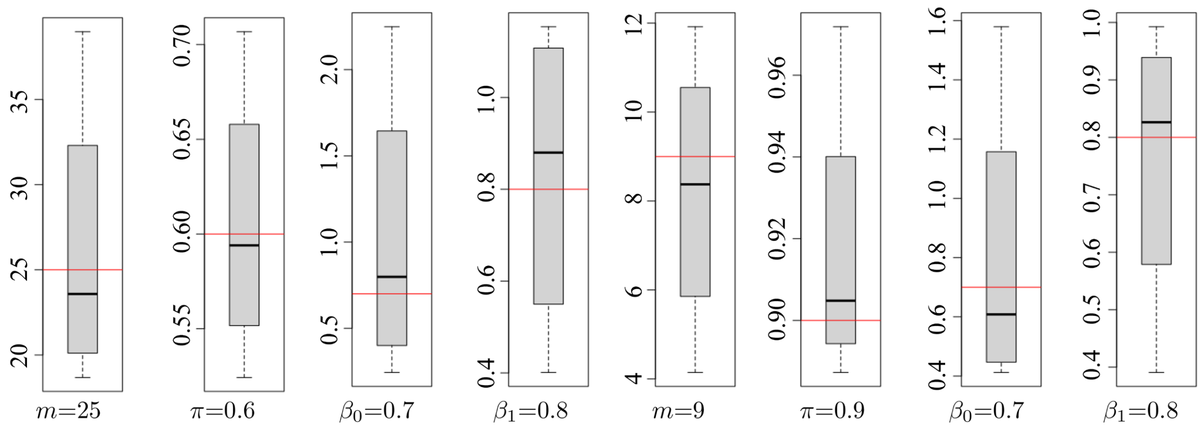


Figure 7. Boxplots with results from the simulation experiment for $N = 1000$.

The following figures are the boxplots of the simulation experiment by using parameters from Table 1, where the black line is the median and the red is true parameter.

4. Application to Real-Life Data

The purpose of this application is to assess : (1) the effect of temperature, and (2) the effect of rainfall on the Dengue fever count series. Therefore we applied the nonstationary Branching Negative Binomial model to the weekly data from Thailand over the period January 2006 to February 2013, which is 373 observations in total.

Table 3 is the summary statistic of the temperature data and Figure 8 is its time series plot. The temperature is higher in the middle of the year while lower between the end and the beginning of two consecutive years, respectively. In particular, higher from April to September, and lower from October to March.

Table 3. Summary statistics for the weekly mean temperature of Thailand from January 2006 to February 2013.

Var	Min	1st Quartile	Median	Mean	3rd Quartile	Max
7.15	17.87	25.61	27.54	26.85	28.50	33.11

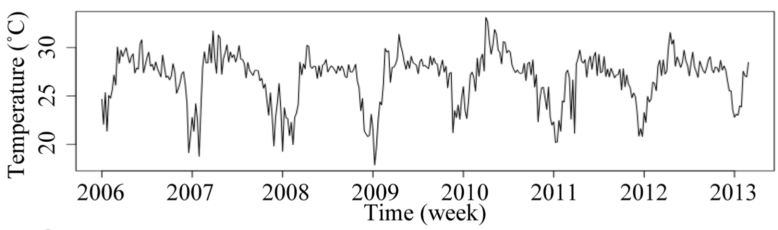


Figure 8. Weekly plot of Thailand temperature over the period of January 2006 to October 2013.

Here, Table 4 displays the summary statistic of the rainfall data and Figure 9 is its time series plot. The rainfall quantity seems to increase in the middle of the year while decreases between the end and the beginning of two consecutive years, respectively. In particular, higher from April to September, and lower from October to March.

Table 4. Summary statistics for the weekly rainfall mean of Thailand from January 2006 to February 2013.

Var	Min	1st Quartile	Median	Mean	3rd Quartile	Max
1640.14	0.00	0.00	7.20	26.29	36.10	244.80

Table 5 is the summarize statistic of the Dengue Fever Counts data and Figure 10 is its time series plot. The variance and the mean are respectively 17.68 and 3.20, which implies the existence of overdispersion in the data since the variance is greater than the mean. This is one of the requirements of our developed models. So far, in addition to the Seasonality pattern in the data (**Figure 10**), the partial autocorrelation functions (**Figure 11**) displayed that data is autoregressive of order one. Therefore our model is suitable to apply to this particular data.

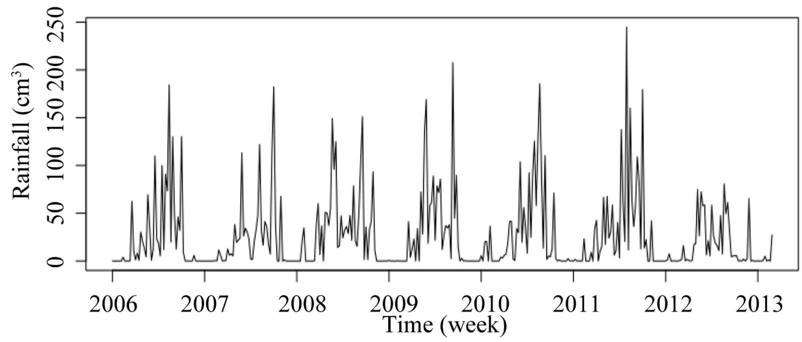


Figure 9. Weekly plot of Thailand rainfall volume over the period of January 2006 to February 2013.

Table 5. Summary statistics for the weekly mean temperature of Thailand from January 2006 to February 2013.

Var	Min	1st Quartile	Median	Mean	3rd Quartile	Max
17.68	0.00	0.00	2.00	3.20	5.00	21.00

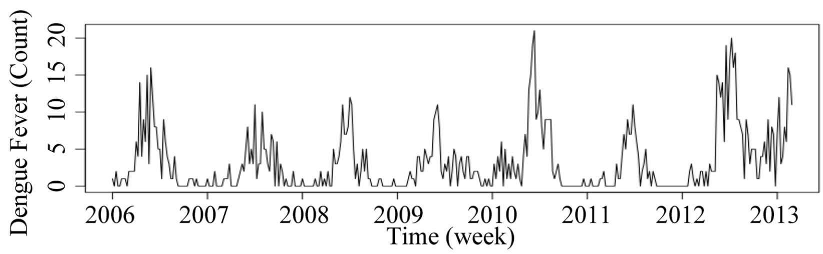
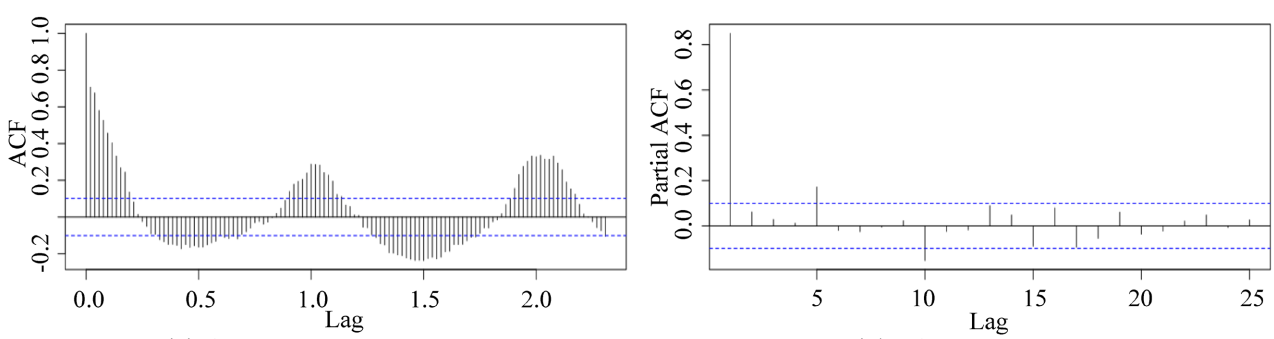


Figure 10. Time series plot of weekly Dengue Fever counts data from Thailand, over the period of January 2006 to February 2013.



(a) ACF plot of Dengue Fever (b) PACF plot of Dengue Fever

Figure 11. Plot of autocorrelation functions.

This application is to fit our developed model and the seasonality in Dengue

Fever counts data, where temperature is an effect factor (**Figure 12**). By fitting the Data to the nonstationary bNB model, we obtained the parameters in **Table 6**.

Table 6. Estimate parameters obtained by fitting Dengue Fever and temperature data in the nonstationary bNB model.

parameters	m	π	β_0	β_1
Estimate values	1.00	0.32	6.623	−0.29

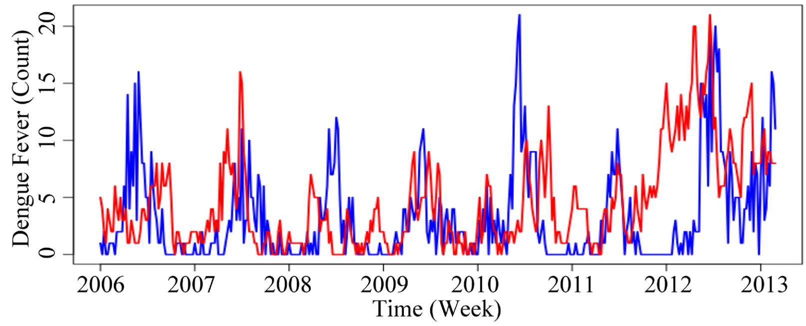


Figure 12. Time series plots of Dengue Fever observations in blue and its estimates values from the model in red when covariate is temperature.

This application is to check how best our developed model can capture the seasonality in Dengue Fever counts data, where Rainfall is also an effect factor (**Figure 13**). By fitting the Data to the nonstationary bNB model, we obtained the parameters and their corresponding Standard error in **Table 7**. The fitting plot of Dengue Fever process and its estimates.

Table 7. Estimate parameters obtained by fitting Dengue Fever and rainfall data in the nonstationary bNB model.

parameters	m	π	β_0	β_1
Estimate values	1.5	0.29	0.19	0.06
Standard error	0.101	0.024	0.026	0.004

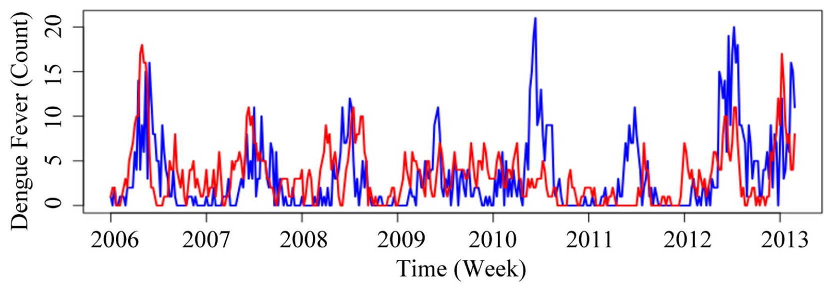


Figure 13. Time series plots of Dengue Fever observations in blue and its Estimates values from the model in red when covariate is rainfall.

5. Goodness-of-Fit Test

The chi-square test of goodness of fit is a statistical test used to determine if there is a significant difference between the observed data and the expected data. In the case of time series of count data model, the chi-square test can be used to assess whether the observe counts follow a specified distribution or if there are any significant deviations from the expected counts. To perform a chi-square test for goodness of fit on this study, we first need to define the expected distribution. This is done by fitting bNB models to the observed data in order to determine the parameters of the models, which are displayed in Table 6 and Table 7. Having the expected counts for each time point in the time series via estimate parameters, the observed counts and expected counts are compared using the chi-square test statistic as follows

$$\chi^2 = \sum_{t=0}^{n} \frac{\left(y_t - \hat{y}_t\right)^2}{\hat{y}_t} \tag{12}$$

where the $\sum$ is taken over all time points in the series. The test statistic follows a chi-square distribution with $(n-1)$ degrees of freedom, where n is the number of time points in the time series. To determine whether the observed counts significantly differ from the expected counts, the test statistic is compared to the critical value of the chi-square distribution at a specified significance level, $\alpha = 0.05$. In this case, the null hypothesis (the observed counts follow the nonstationary bNB model) is rejected if the test statistic is greater than the critical value and it can be conclude that there are significant deviations from the expected counts.

Non-Stationary Branching Negative Binomial Autoregressive Model

Formulation of the null and alternative hypotheses

- Null hypothesis H_0: Dengue Fever data follow the models
- Alternative hypothesis H_a: Dengue Fever data does not follow it

Test Statistic χ^2

```
Chi-squared test for given probabilities

data:  Dengue_Fever
X-squared = 393.83, df = 372, p-value = 0.2091
```

Figure 14. Chi-square goodness of fit test outcome of the non-stationary bNB model to Dengue Fever count data with covariable, temperature.

```
Chi-squared test for given probabilities

data:  Dengue_Fever
X-squared = 343.83, df = 372, p-value = 0.8498
```

Figure 15. Chi-square goodness of fit test outcome of the non-stationary bNB model to Dengue Fever count data with covariable, rainfall.

In this case, two chi-square statistics are determined, including the one where temperature is used as covariate and the one where rainfall is used as covariates, since they generate different expected data for Dengue Fever counts. Then Figure 14 and Figure 15 show the test outcomes from those two cases.

Critical Value

Knowing level of significance $\alpha = 5\%$ and the degree of freedom $df = n - 1 = 372$, qchisq $(1-\alpha, df)$ allowed us to get the critical value, which is $\chi^2 = 417.9735$, by use of R software.

As result, it's seeing in Figure 14 that the χ^2 value ($\chi = 393.83$) is less than the critical value ($\chi_0 = 417.9735$), which implies that the non-stationary bNB model fits well the Dengue Fever with covariate temperature.

Bearing Figure 15, it's seeing that the χ^2 value ($\chi = 343.83$) is less than the critical value. This also proves that the non-stationary bNB model fits well the Dengue Fever count data when rainfall is the covariate.

6. Conclusions and Suggestions

The findings from both the simulation study and the analysis of the real-world dataset have significant implications for the proposed nonstationary autoregressive model for time series count data with covariates. Starting with the simulation study, it provides a controlled environment to evaluate the performance of the proposed modeling strategy. By simulating count time series data with known characteristics, researchers can assess how well the model captures the underlying seasonality and incorporates the covariate information. The simulation study allows for comparisons between the true values and the estimates obtained through maximum likelihood estimation. The results of the simulation study can provide insights into the accuracy and robustness of the proposed model in capturing the complex dynamics of count data with seasonality and covariate effects. Moving on to the analysis of the real-world dataset, which focuses on an infectious disease with covariates such as temperature and rainfall, the application of the proposed model provides insights into the relationship between the disease occurrence and the covariate variables. The findings can shed light on how weather factors impact the incidence of the infectious disease. By considering nonstationarity and incorporating the covariate information into the model, the analysis can provide a more comprehensive understanding of the factors influencing the disease occurrence and improve the accuracy of predictions or assessments. However, it is important to acknowledge the limitations of the study. One limitation may be the assumptions made in the proposed model. While the nonstationary autoregressive model with covariate incorporation offers an innovative approach, the assumptions underlying the model may not fully capture the complexities of the realworld data. Additionally, the generalizability of the findings might be limited to the specific context of the dataset used in the analysis. Different datasets with varying characteristics and applications may yield different results. Furthermore, the proposed model's performance depends on the quality and availability of the covariate information. In some cases, the covariates may not be accurately measured or may contain missing data, which can introduce uncertainties and affect the model's performance. It is crucial to address such limitations and explore techniques to handle missing or imperfect covariate data. Overall, the findings from both the simulation study and the analysis of the realworld dataset provide valuable insights into the proposed nonstationary autoregressive model's effectiveness in capturing seasonality and incorporating covariates in count time series data. These findings pave the way for further research and refinement of the modeling approach, considering potential limitations and exploring alternative methodologies to enhance the analysis of count data with covariate information and seasonality aspects.

Data Availability Statement

Data available on request from the authors

Conflicts of Interest

The authors declare no potential conflict of interests.

References

[1] Quddus, M.A. (2008) Time Series Count Data Models: An Empirical Application to Traffic Accidents. *Accident Analysis & Prevention*, **40**, 1732-1741. https://doi.org/10.1016/j.aap.2008.06.011

[2] Alqawba, M. and Diawara, N. (2021) Copula-Based Markov Zero-Inflated Count Time Series Models with Application. *Journal of Applied Statistics*, **48**, 786-803. https://doi.org/10.1080/02664763.2020.1748581

[3] Banerjee, D. (2014) Forecasting of Indian Stock Market Using Time-Series ARIMA Model. 2014 2*nd International Conference on Business and Information Management* (*ICBIM*), Durgapur, 9-11 January 2014, 131-135. https://doi.org/10.1109/icbim.2014.6970973

[4] Idrees, S.M., Alam, M.A. and Agarwal, P. (2019) A Prediction Approach for Stock Market Volatility Based on Time Series Data. *IEEE Access*, **7**, 17287-17298. https://doi.org/10.1109/access.2019.2895252

[5] Saridakis, G. (2004) Violent Crime in the United States of America: A Time-Series Analysis between 1960-2000. *European Journal of Law and Economics*, **18**, 203-221. https://doi.org/10.1023/b:ejle.0000045082.09601.b2

[6] Borowik, G., Wawrzyniak, Z.M. and Cichosz, P. (2018) Time Series Analysis for Crime Forecasting. 2018 26*th International Conference on Systems Engineering* (*ICSEng*), Sydney, 18-20 December 2018, 1-10. https://doi.org/10.1109/icseng.2018.8638179

[7] Jung, R.C. and Tremayne, A.R. (2011) Useful Models for Time Series of Counts or Simply Wrong Ones? *AStA Advances in Statistical Analysis*, **95**, 59-91. https://doi.org/10.1007/s10182-010-0139-9

[8] Jung, R.C., Kukuk, M. and Liesenfeld, R. (2006) Time Series of Count Data: Modeling, Estimation and Diagnostics. *Computational Statistics & Data Analysis*, **51**, 2350-2364. https://doi.org/10.1016/j.csda.2006.08.001

[9] Weiß, C.H. (2008) Thinning Operations for Modeling Time Series of Counts—A Survey. *AStA Advances in Statistical Analysis*, **92**, 319-341. https://doi.org/10.1007/s10182-008-0072-3

[10] Traore, B., Malenje, B.M. and Imboga, H. (2022) A First Order Stationary Branching Negative Binomial Autoregressive Model with Application. *Open Journal of Statistics*, **12**, 810-826. https://doi.org/10.4236/ojs.2022.126046

[11] Nogueira, A., Salvador, P., Valadas, R. and Pacheco, A. (2011) Markovian Modelling of Internet Traffic. In: *Lecture Notes in Computer Science*, Springer, 98-124. https://doi.org/10.1007/978-3-642-02742-0_5

[12] Ravishanker, N., Serhiyenko, V. and Willig, M.R. (2014) Hierarchical Dynamic Models for Multivariate Times Series of Counts. *Statistics and Its Interface*, **7**, 559-570. https://doi.org/10.4310/sii.2014.v7.n4.a11

[13] Joseph, M.B., Rossi, M.W., Mietkiewicz, N.P., Mahood, A.L., *et al.* (2019) Understanding and Predicting Extreme Wildfires in the Contiguous United States. *Ecological Applications*, **29**, e01898.

[14] Brijs, T., Karlis, D. and Wets, G. (2008) Studying the Effect of Weather Conditions on Daily Crash Counts Using a Discrete Time-Series Model. *Accident Analysis & Prevention*, **40**, 1180-1190. https://doi.org/10.1016/j.aap.2008.01.001

[15] Zhu, R. and Joe, H. (2006) Modelling Count Data Time Series with Markov Processes Based on Binomial Thinning. *Journal of Time Series Analysis*, **27**, 725-738. https://doi.org/10.1111/j.1467-9892.2006.00485.x

[16] Houseman, E.A., Coull, B.A. and Shine, J.P. (2006) A Nonstationary Negative Binomial Time Series with Time-Dependent Covariates: Enterococcus Counts in Boston Harbor. *Journal of the American Statistical Association*, **101**, 1365-1376. https://doi.org/10.1198/016214506000000627

[17] Agosto, A., Cavaliere, G., Kristensen, D. and Rahbek, A. (2016) Modeling Corporate Defaults: Poisson Autoregressions with Exogenous Covariates (PARX). *Journal of Empirical Finance*, **38**, 640-663. https://doi.org/10.1016/j.jempfin.2016.02.007

[18] Andersson, J. and Karlis, D. (2014) A Parametric Time Series Model with Covariates for Integers in Z. *Statistical Modelling*, **14**, 135-156. https://doi.org/10.1177/1471082x13504719

[19] Hay, J.L. and Pettitt, A.N. (2001) Bayesian Analysis of a Time Series of Counts with Covariates: An Application to the Control of an Infectious Disease. *Biostatistics*, **2**, 433-444.