https://doi.org/10.48048/tis.2022.3171

Bayesian Inference for the Negative Binomial-Quasi Lindley Model for Time Series Count Data on the COVID-19 Pandemic

Sirinapa Aryuyuen and Unchalee Tonggumnead*

Department of Mathematics and Computer Science, Faculty of Science and Technology, Rajamangala University of Technology Thanyaburi, Pathum Thani 12110, Thailand

(*Corresponding author's e-mail: unchalee_t@rmutt.ac.th)

Received: 13 March 2022, Revised: 26 April 2022, Accepted: 13 May 2022, Published: 1 November 2022

Abstract

In statistical models, the generalized linear model (GLMs) plays a role in studying to describe a response variable as a function of 1 or more predictor variables. Computational methods and mixed distributions are frequently used to build predictive models to perform time-to-event data analysis. To develop a statistical model so that the model can make predictions appropriately and accurately, it starts with developing a suitable distribution for the nature of the actual data. This paper proposes a new mixed negative binomial distribution for count data with over-dispersion, the so-called negative binomial-quasi Lindley (NB-QL) distribution. A new GLMs framework for the NB-QL model to build the time series count data model is introduced, and its application is carried out based on the actual data sets of the COVID-19 epidemic in Thailand. The models are related to GLMs as they are linear relationships between outcome variables and covariates. Where the response variable was in the form of time series count data under the exponential family distribution function, with the random components and link functions. In this study, we study the factors that affect the number of COVID-19 death cases in Thailand and provide the predictive modeling of the number of the COVID-19 death cases from 1 January 2020 to 31 December 2020, for which this data set has the observed sample of 366 days. In contrast, a model with an NB-QL distribution and NB has approached the uniform. Based on the deviance, DIC, p_D and the probability integral transform histogram, we can see that the proposed model is also suitable for forecasting the number of the COVID-19 death cases daily in Thailand, indicating that the NB-QL time series model was another efficient alternative to modeling count data that has an over-dispersion problem. According to the NB-QL time series model about the number of the COVID-19 death cases daily in Thailand, it is indicated that the average number of daily COVID-19 deaths is influenced by the number of the COVID-19 death cases in the previous 3 days. The average number of COVID-19 death cases in Thailand is also influenced by the previous 2 days. At the same time, the number of infected cases daily in Thailand is influenced by the number of the COVID-19 death cases daily. In addition, there are also the components interventions of internal covariate effects due to the data, as there was a surge in the number of the COVID-19 death cases daily in Thailand at the time between $73 \le t \le 143$ and $t \ge 352$.

Keywords: Bayesian inference, Negative binomial-quasi Lindley distribution, Time series count data, Poisson regression model, Negative binomial regression model

Introduction

In statistical models, linear models play a role in describing a continuous response variable as a function of 1 or more predictor variables. Linear regression is a statistical method used to create a linear model, in which the model describes the relationship between a response variable and 1 or more predictor variables. Using a linear regression may violate the normality assumption when we consider the data with a response variable in count data such as $N_0 = \{0,1,2,3,\ldots\}$. Hence, all the classical statistical tests would fail to evaluate the model. However, the generalized linear model (GLMs) can be used instead of specifying the suitable distribution. The GLMs generalizes linear regression by allowing the linear model to be related to the response variable via a link function and by allowing the magnitude of the variance of each measurement to be a function of its predicted value [1]. The interest in count models, in which the response variable is count data, has been increased in the last decade. The primary modeling alternative for count data has traditionally been a Poisson regression model, in which the response variable Y has been

distributed as a Poisson distribution with a parameter λ, denoted by $Y \sim \text{Pois}(\lambda)$. Its probability mass function (pmf) is

$$f(y;\lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!}, \quad y = 0,1,2,..., \text{ and } \lambda > 0. \tag{1}$$

The Poisson regression model is the best choice if the mean and the variance of the response variable are closer to each other [2]. But often, in practice, the variance is larger than the mean, so the Poisson distribution is not appropriate for this situation. An alternative is a negative binomial (NB) distribution for count modeling. The NB distribution is a mixed Poisson-gamma distribution, which was first derived by [3]. The variation of this parameter can account for a variance of the data that is higher than the mean (over-dispersion). The NB regression model is widely used and has been proved to fit well for the count data with over-dispersion. Therefore, the NB distribution was more appropriate than the Poisson distribution [4-7]. However, the NB distribution is proper for count data, presenting over-dispersion without necessarily being heavy-tailed; notably, heavy-tailed distributions tend to over-dispersion [8]. According to the over-dispersion and heavy tail troubles, the Poisson and NB distributions may not be appropriate to describe this type of data. As a solution, thus, we can use other distributions that do not have this restriction, such as some mixed distributions. The mixed NB distributions are 1 mixture of distributions. Many mixed NB distributions are introduced, such as the NB-Lindley [9], NB-generalized exponential [10], NB-gamma [11], NB-Sushila [12], and NB-generalized Lindley [13], etc.

In this paper, we firstly proposed a new mixture NB distribution to be a flexible alternative to analyze count data with over-dispersion. The new distribution is a mix of the NB and QL distributions; a name is Negative Binomial-Quasi Lindley, abbreviated NB-QL distribution. Furthermore, in this study, the response variable is the number of coronavirus disease 2019 (COVID-19) death cases in Thailand. The COVID-19 pandemic, also known as the coronavirus pandemic, is an ongoing global pandemic of the COVID-19 caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The novel virus was first identified from an outbreak in the Chinese city of Wuhan in December 2019, and attempts to contain it there failed, allowing it to spread across the globe. The World Health Organization (WHO) declared a Public Health Emergency of International Concern on 30 January 2020 and a pandemic on 11 March 2020. As of 4 January 2022, the pandemic had caused more than 292 million cases and 5.45 million deaths, making it one of the deadliest in history. According to the report of daily confirmed cases of COVID-19 from the Department of Disease Control (2021) in Thailand [14], the COVID-19 epidemic in Thailand is changing every day. These data are the numbers of the COVID-19 cases daily from 1 January 2020 to 31 December 2020, comprising 366 days in this study. There was the number of the COVID-19 death cases as 0 people of 324 days, 1 person of 26 days, 2 people of 8 days, 3 people of 7 days, and 4 people of 1 days. The mean and variance of the number of the COVID-19 death cases are 0.18 and 0.38, respectively. Since the variance is greater than the mean, this data set has an over-dispersion problem.

However, the number of the COVID-19 cases daily from 1 January 2020 to 31 December 2020 is time series data. Creating a model using the generalized linear model in terms of regression is not appropriate. This is because the regression model is subject to the condition that each observation must be independent of the others. An alternative to time series data is the count time series following generalized linear models, which are flexible towards serial correlation, deterministic and stochastic seasonality, deterministic trends, and covariates. Recently, there has been a growing interest in an advanced class of time series models known as integer-valued generalized autoregressive heteroscedastic (INGARCH) models. These offer an alternative to the integer-valued time series models and account for the over-dispersion, non-negativity and temporal correlation [15-17]. Ferland *et al*. [16] introduced the Poisson INGARCH model, Zhu [18] extended that with the NB conditional distribution to address the over-dispersion issue. Although the NB distribution is suitable for statistical data with over-dispersion, in some instances there may be a very high probability that no events of interest will occur. This makes the problem of over-dispersion more severe, resulting in an improper NB distribution. Thus, a new mixture NB distribution serves as a flexible alternative to analyze count data with over-dispersion.

In the classical solution to the GLMs, the regression coefficients are usually estimated by maximizing the nonlinear log-likelihood. The Newton-Raphson method can be applied to iteratively find the maximum likelihood estimation (MLE) of regression coefficients [5]. However, the MLE provides only a point estimate which may not be robust or may even fail to converge when the sample size is small or when the dispersion parameter is much larger than the mean. Moreover, it does not consider prior information, which may be helpful in the case of missing observations. As an alternative, Bayesian inference can account for

prior expert knowledge on variables of interest, especially in a small sample size setting. It provides a sample of estimators, which may be helpful for the uncertainty analysis [19]. The estimating of the parameters in the regression models with the Bayesian approach is proposed for some mixed NB distributions. For the parameter estimation method, some researchers were predisposed to the Bayesian approach over the MLE in recent times [19-21]. Some studies have used the Bayesian method to estimate parameters in the GLMs for Poisson and NB regression [13]. The practical advantages of the Bayesian approach are its flexibility and generality, as this allows it to cope with complex problems [21,22].

This paper provides a new mixed negative binomial distribution for time series count data with over-dispersion, and the Bayesian approach is the method used to estimate the parameters of the proposed model. We will apply the GLMs framework to build the time series count data, while the proposed model is constructed for the time series count data of the COVID-19 death cases in Thailand from 1 January 2020 to 31 December 2020. The number of deaths is the total number of deaths that the Ministry of Health has reported daily. Finally, the discussion and conclusion are presented.

Materials and methods

Generalized linear model for time series count

Let Y_t be a time series of count data, and $\mathbf{X_t}$ be a time-varying *r*-dimensional covariate vector, say $\mathbf{X_t} = (X_{t,1}, X_{t,2}, ..., X_{t,k})^T$ for $t = 1, 2, 3, \ldots$. The conditional mean, $\mathrm{E}(Y_t \mid F_{t-1})$ of the count time series is assumed to be identical to the sequence of mean process μ_t, such that $\mathrm{E}(Y_t \mid F_{t-1}) = \mu_t$. Denoted by F_{t-1} is the history of the joint process $\{Y_t, \mu_t, \mathbf{X_{t+1}}\}$ with the part process up to time t including the covariate information function at time $t+1$. There is a link function $\{g : \mathbf{R}^+ \to \mathbf{R}\}$ and transformation function $\{\tilde{g} : \mathbf{N}_0 \to \mathbf{R}\}$ where $\mathbf{N}_0 = \{0, 1, 2, 3, \ldots\}$, $\mathbf{R}^+$ and $\mathbf{R}$ as a positive real number and a real number respectively. The count time series following generalized linear models [23] is given by:

$$g(\mu_t) = \beta_0 + \sum_{k=1}^{p} \beta_k \tilde{g}(Y_{t-ik}) + \sum_{l=1}^{q} \alpha_l g(\mu_{t-jl}) + \boldsymbol{\eta}^T \mathbf{X_t} \tag{2}$$

Model (2) is a convenient and flexible model class for serial correlation, deterministic and stochastic seasonality, deterministic trends, and covariates. For GLMs, $\nu_t = g(\mu_t)$ is a linear predictor, defining P as $P = \{i_1, i_2, ..., i_p\}$ and i as integer $0 < i_1, i_2, ..., i_p < \infty$, $p \in \mathbf{N}_0$, which is relevant to observations at the previous period. It is possible to regress the observed lag $Y_{t-i_1}, Y_{t-i_2}, ..., Y_{t-i_p}$, defining Q as $Q = \{j_1, j_2, ..., j_q\}$, and j as integer $0 < j_1, j_2, ..., j_q < \infty$, $q \in \mathbf{N}_0$ which is related to the conditional mean at the previous j_q period. It is the regressor variable on the lag for the conditional mean $\mu_{t-j_1}, \mu_{t-j_2}, ..., \mu_{t-j_q}$. From model (2), consider the situation where g and $\tilde{g}$ equal the identity. Namely $g(x) = \tilde{g}(x) = x$, for $P = (1, 2, ..., p)$, $Q = (1, 2, ..., q)$ and $\boldsymbol{\eta} = 0$ [15-17]. Model (2) becomes:

$$g(\mu_t) = \beta_0 + \sum_{k=1}^{p} \beta_k Y_{t-k} + \sum_{l=1}^{q} \alpha_l g(\mu_{t-l}). \tag{3}$$

The parameter space of model (2) with covariates is given by:

$$\boldsymbol{\Theta} = \left\{ \theta \in \mathbf{R}^{1+p+q+r}; \beta_0 > 0, \beta_1, ..., \beta_p, \alpha_1, ..., \alpha_q, \eta_1, \eta_2, ..., \eta_r \geq 0, \sum_{k=1}^{p} \beta_k + \sum_{l=1}^{q} \alpha_l < 1 \right\}. \tag{4}$$

The log-linear of model (3) with covariates is given by:

$$\boldsymbol{\Theta} = \left\{ \theta \in \mathbf{R}^{1+p+q+r}; \beta_0 > 0, |\beta_1|, ..., |\beta_p|, |\alpha_1|, ..., |\alpha_q|, < 1, \left| \sum_{k=1}^{p} \beta_k + \sum_{l=1}^{q} \alpha_l \right| < 1 \right\}. \tag{5}$$

From Eq. (2), $\sum_{k=1}^{p}\beta_k \tilde{g}(Y_{t\text{-}ik})$ is the term that indicates the average of Y_t is influenced by the number of $Y_{t\text{-}ik}$, where the variable Y_t represents the number of covid death cases daily at time t, $Y_{t\text{-}ik}$ represents of the number covid deaths cases daily at the previous time $t-ik$, $\sum_{l=1}^{q}\alpha_l g(\mu_{t\text{-}jl})$ is the term that indicates the average of Y_t is influenced by the conditional mean $\mathrm{E}(Y_t \mid F_{t\text{-}1})$ at the previous time $t-jl$, $\boldsymbol{\eta}^T\mathbf{X_t}$ is the term that indicates that the average of Y_t is influenced by the covariates, where the covariates in this study include the new COVID-19 infection cases daily in Thailand from 1 January 2020 to 31 December 2020 and the intervention variable. While β_0, β_k, α_l, and $\boldsymbol{\eta}^T$ are unknown parameters.

The Poisson model for time series count data

From model (3), assume Y_t be a random variable distributed as the Poisson distribution. This model is a special case of model (2) called a linear model of order p and q. These models are also known as INGARCH, or autoregressive conditional Poisson (ACP) models. Consider again for model (2) when defining the logarithmic link function $g(x)=\log(x)$ and $\tilde{g}(x)=\log(x+1)$. Then, we obtain a log-linear model of order p and q. Let $\nu_t=\log(\mu_t)$ for $P=(1,2,...,p)$, $Q=(1,2,...,q)$ and $\eta=0$. Model (2) becomes [24]:

$$\nu_t=\log(\mu_t)=\beta_0+\sum_{k=1}^{p}\beta_k\log(Y_{t-k}+1)+\sum_{l=1}^{q}\alpha_l\log\nu_{t-1}. \tag{6}$$

For model (2), if Y_t has the Poisson distribution, if $\{Y_t \mid F_{t-1};\mu_t\}\sim \mathrm{Pois}(\mu_t)$, then the Poisson model for the time series count data is:

$$\mathrm{P}(Y_t \mid F_{t-1};\mu_t)=\frac{\mu_t^{y}e^{-\mu_t}}{y!},\quad y=0,1,2,... \text{ and } \mu_t>0. \tag{7}$$

The varaince and mean of $\{Y_t \mid F_{t-1};\mu_t\}$ are $\mathrm{V}(Y_t \mid F_{t-1};\mu_t)=\mathrm{E}(Y_t \mid F_{t-1};\mu_t)=\mu_t$. Because in the case of the Poisson response model the conditional mean is identical to the conditional variance. If the conditional variance is greater than the conditional mean, or if the response contains a large number of zeroes, which is often referred to as over-dispersion, the Poisson response is not suitable for modeling. In this case, the NB distribution can deal with the problem, since it does not require a conditional mean identical to the conditional variance. It rather allows variance to be larger than the mean.

The negative binomial model for time series count data

Let Y be a random variable distributed as the NB distribution with parameters r and m, denoted by $Y\sim \mathrm{NB}(r,m)$. Its pmf is:

$$f(y;r,m)=\frac{\Gamma(r+y)}{\Gamma(r)\Gamma(y+1)}m^r(1-m)^y,\quad y=0,1,2,...,\ r>0,\ \text{and}\ 0<m<r. \tag{8}$$

From the pmf in (8), alternatively, we can parameterize m in the term of r as $m=r/(\mu+r)$ for μ as the mean response variable and r as the reciprocal (or inverse of a dispersion parameter $\phi:\phi=1/r$). The traditional NB distribution [6] can be rewritten in terms of its pmf as follows,

$$f(y;r,\mu)=\frac{\Gamma(r+y)}{\Gamma(r)\Gamma(y+1)}\left(\frac{r}{\mu+r}\right)^r\left(\frac{\mu}{\mu+r}\right)^y,\quad y=0,1,2,..., \tag{9}$$

where $\mu > 0,\ r > 0$ and $\Gamma(.)$ is a complete gamma function.

For model (1), if Y_t is distributed as the NB distribution, $\{Y_t \mid F_{t-1}; \mu_t, r\} \sim \text{NB}(\mu_t, r)$, where the NB distribution is parametrized in terms of its mean with an additional dispersion parameter

$$F(Y_t \mid F_{t-1}; \mu_t, r) = \frac{\Gamma(r+y)}{\Gamma(y+1)\Gamma(r)}\left(\frac{r}{r+\mu_t}\right)^r\left(\frac{\mu_t}{r+\mu_t}\right)^y,\ \ y = 0,1,2,\ldots. \tag{10}$$

Its mean and variance are respectively:

$$\text{E}\left(Y_t \mid F_{t-1}; \mu_t, r\right) = \mu_t \text{ and } \text{V}\left(Y_t \mid F_{t-1}; \mu_t, r\right) = \mu_t + \frac{\mu_t^2}{r}. \tag{11}$$

This research consists of 3 parts: 1) A new distribution was developed in the form of a mixed negative binomial distribution; 2) The properties of the developed mixed negative binomial distribution were studied, and 3) A mixed negative binomial time series model in the form of count time series generalized linear models was created, and then compare the performance of the purpose model with the Poisson and NB models, namely determining that the response variables of GLMs flamework in Eq. (2) are represented by Poisson and NB distribution.

Results and discussion

A new mixed negative binomial model for time series count data

In this section, we proposed a new mixed negative binomial distribution, the so-called negative binomial-quasi Lindley (NB-QL) distribution. The NB-QL distribution is obtained by mixing between the NB distribution [3] and the quasi Lindley (QL) distribution [25]. Firstly, we introduce the QL distribution as follows.

Let λ be a random variable distributed as the QL distribution with parameters a and b, which is denoted by $\lambda \sim \text{QL}(a,b)$. The probability density function (pdf) of the QL distribution is:

$$g(\lambda; a, b) = \frac{a(b+a\lambda)}{b+1}e^{-a\lambda},\ \ \lambda > 0,\ \ a > 0 \text{ and } b > 0. \tag{12}$$

The moment generating function (mgf) of λ is

$$M_\lambda(t; a, b) = \text{E}(e^{t\lambda}) = \frac{1}{b+1}\left(\frac{ab}{a-t} - \frac{t}{a} + 1\right);\ t > 0. \tag{13}$$

If $Y \mid \lambda \sim \text{NB}(r, m = e^{-\lambda})$ with the pmf as (8) and

$$f(y_t \mid \lambda) = \frac{\Gamma(y+r)}{\Gamma(y+1)\Gamma(r)}e^{-\lambda r}\left(1-e^{-\lambda}\right)^y = \frac{\Gamma(y+r)}{\Gamma(y+1)\Gamma(r)}\sum_{j=0}^{y}\binom{y}{j}(-1)^j e^{-\lambda(r+j)}, \tag{14}$$

where $\lambda \sim \text{QL}(a,b)$ with the pdf in (12), then the pmf of Y can be obtained by

$$\begin{aligned} f(y; r, a, b) &= \int_0^\infty f(y \mid \lambda) g(\lambda; a, b) d\lambda = \frac{\Gamma(y+r)}{\Gamma(y+1)\Gamma(r)}\sum_{j=0}^{y}\binom{y}{j}(-1)^j \int_0^\infty e^{-\lambda(r+j)} g(\lambda; a, b) d\lambda \\ &= \frac{\Gamma(y+r)}{\Gamma(y+1)\Gamma(r)}\sum_{j=0}^{y}\binom{y}{j}(-1)^j M_\lambda[-(r+j); a, b]. \end{aligned} \tag{15}$$

By replacing the mgf of λ in (13) as in (15) with $t = -(r+j)$, we have the pmf of Y as follows:

$$f(y;r,a,b)=\frac{\Gamma(y+r)}{\Gamma(y+1)\Gamma(r)}\sum_{j=0}^{y}\binom{y}{j}\frac{(-1)^j}{b+1}\left(\frac{ab}{a+r+j}+\frac{r+j}{a}+1\right),\ y=0,1,2,.... \tag{16}$$

Therefore, we obtain a random variable Y distributed as the NB-QL distribution with parameters r, a and b denoted by $Y\sim\text{NB-QL}(r,a,b)$. Some pmf plots of the NB-QLdistribution are shown in **Figure 1**. The NB-QL distribution includes 2 submodels: 1) If $b=a$, then the NB-QL distribution reduces to the negative binomial-Lindley distribution [9]. 2) If $b=0$, then the NB-QL distribution reduces to the negative binomial-gamma distribution [11].

The negative binomial-quasi Lindley model for time series count data

For model (2), let Y_t be a random variable distributed as the NB-QL distribution with a vector parameter of $\mathbf{\Lambda}=\{\mu_t,r,a,b\}$, i.e., $\{Y_t\mid F_{t-1};\mathbf{\Lambda}\}\sim\text{NB-QL}(\mathbf{\Lambda})$, we have the NB-QL model for time series count data as follows:

$$\begin{aligned}\text{P}\left(Y_t\mid F_{t-1};\mathbf{\Lambda}\right)&=\int_0^{\infty}\text{NB}(y;\lambda\mu_t,r)\text{QL}(\lambda;a,b)d\lambda,\\&=\frac{\Gamma(y+r)}{\Gamma(y+1)\Gamma(r)}\int_0^{\infty}\left(\frac{r}{\lambda\mu_t+r}\right)^r\left(\frac{\mu_t}{\lambda\mu_t+r}\right)^y\frac{a(b+a\lambda)e^{-a\lambda}}{b+1}d\lambda,\end{aligned} \tag{17}$$

Its mean and variance are respectively:

$$\text{E}\left(Y_t\mid F_{t-1};\mathbf{\Lambda}\right)=\mu_t\text{E}(\lambda)\ \text{and} \tag{18}$$

$$\text{V}\left(Y_t\mid F_{t-1};\mathbf{\Lambda}\right)=\text{E}(Y_t\mid F_{t-1})+\mu_t^2\left(\frac{1+r}{r}\right)\text{E}(\lambda^2)-\text{E}^2(Y_t\mid F_{t-1}) \tag{19}$$

where $\text{E}(\lambda)$ and $\text{E}(\lambda^2)$ are the first and second moments of the QL distribution as follows:

$$\text{E}(\lambda)=\frac{1}{a}\left(\frac{b+2}{b+1}\right)\ \text{and}\ \text{E}(\lambda^2)=\frac{1}{a^2}\left(\frac{b+3}{b+1}\right) \tag{20}$$

In this paper, the vector of unknown parameters $\mathbf{\Lambda}=\{\mu_t,r,a,b\}$, can be customarily estimated using the Bayesian approach, which allows the consideration of prior information for parameter estimation.

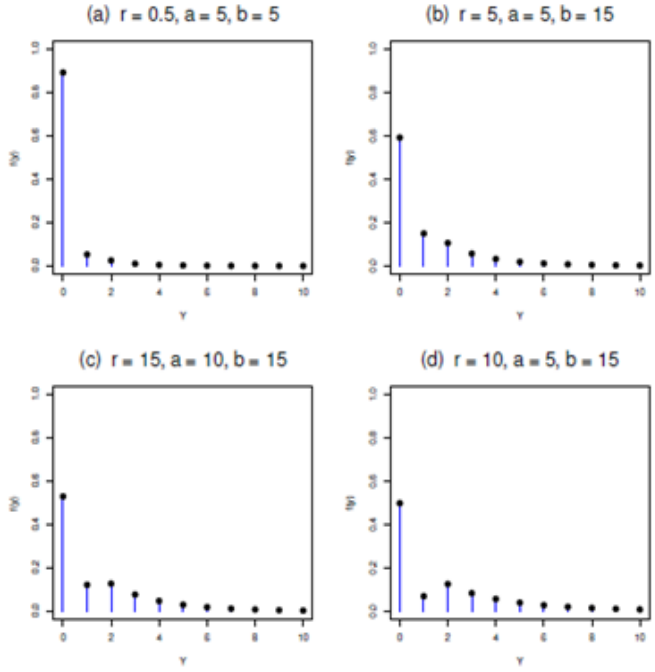


Figure 1 The pmf plots of Y for $Y \sim \text{NB-QL}(r, a, b)$ with some specified parameters r, a and b.

Bayesian inference for the negative binomial-quasi Lindley model for time series count data

In this article, we implement the Bayesian approach using the Markov Chain Monte Carlo (MCMC) technique for the NB-QL model for time series count data, which involves the Bayesian hierarchical model, prior distributions, and joint posterior density. Some statistical programs for Bayesian analysis are available for parameter estimation, such as WinBUGS, OpenBUGS, and JAGS, which are the most popular. The techniques of Bayesian inference may be extended to hierarchical Bayesian analysis. Numerous researchers have shown interest in the study of the hierarchical Bayesian modeling approach relying on MCMC techniques as referred to in [22] and [26]. Since the likelihood function of the NB-QL regression model in (17) is not a closed-form, e.g.,

$$\mathrm{L}\left(Y_t \mid F_{t-1}; \boldsymbol{\Lambda}\right) = \prod_{t=1}^{n} \frac{\Gamma(y_t + r)}{\Gamma(y_t + 1)\Gamma(r)} \int_0^{\infty} \left(\frac{r}{\lambda\mu_t + r}\right)^r \left(\frac{\mu_t}{\lambda\mu_t + r}\right)^{y_t} \frac{a(b + a\lambda)e^{-a\lambda}}{b+1} d\lambda. \tag{21}$$

It can be executed using the representation of the hierarchical model implicit both in the integral and the definition of the QL distribution. Since the QL distribution is mixing between the exponential (Exp) distribution with scale parameter a, denoted by $\text{Exp}(a)$ and the Gamma (Gam) distribution with the shape parameter 2 and the scale parameter b, denoted by $\text{Gam}(2, b)$, the pdf of the QL distribution as in (12) can therefore be written as [25]:

$$\lambda \sim \frac{b}{1+b}\text{Exp}(a) + \frac{1}{1+b}\text{Gam}(2, a). \tag{22}$$

The NB-QL distribution is conditional upon the unobserved site-specific frailty term λ, which describes the additional heterogeneity [27]. Consequently, the hierarchical framework can be represented as:

$$\mathrm{P}(y_t \mid F_{t-1}; \mu_t, r \mid \lambda) = \text{NB}(y_t; \lambda\mu_t, r) \text{ and } \lambda \sim \text{QL}(a, b) \tag{23}$$

In Bayesian inference, the prior distribution plays a defining role in the estimation of the unknown parameters in any distribution. In this study, all unknown parameters r, a, b and parameter space $\boldsymbol{\Lambda}$ as in Eqs. (2) or (3) are considered. Assume the parameters of NB-QL regression model with parameters r, a and b are distributed as the gamma distribution, and $\boldsymbol{\Theta}$ is distributed as the normal distribution. They are mutually independently distributed in each parameter, and the joint prior distribution of all unknown parameters is as follows:

$$r \sim \text{Gam}(\gamma_r, \theta_r),\ \ a \sim \text{Gam}(\gamma_a, \theta_b),\ \ b \sim \text{Gam}(\gamma_b, \theta_b) \text{ and } \boldsymbol{\Theta} \sim \text{Normal}(b_0, S_\beta), \tag{24}$$

where the positive real values of $\gamma_r, \theta_r, \gamma_a, \theta_a, \gamma_b, \theta_b, b_0$ and S_β are known or fixed. Suppose that b_0 is a $(k+1)\times 1$ hyper-parameter vector and S_β is a $(k+1)\times(k+1)$ known non-negative specific matrix. Each parameter is supposed to be independently distributed, and the joint prior distribution of all unknown parameters can be written as:

$$\pi(\boldsymbol{\Lambda}) = \pi(r)\pi(a)\pi(b)\pi(\boldsymbol{\Theta}). \tag{25}$$

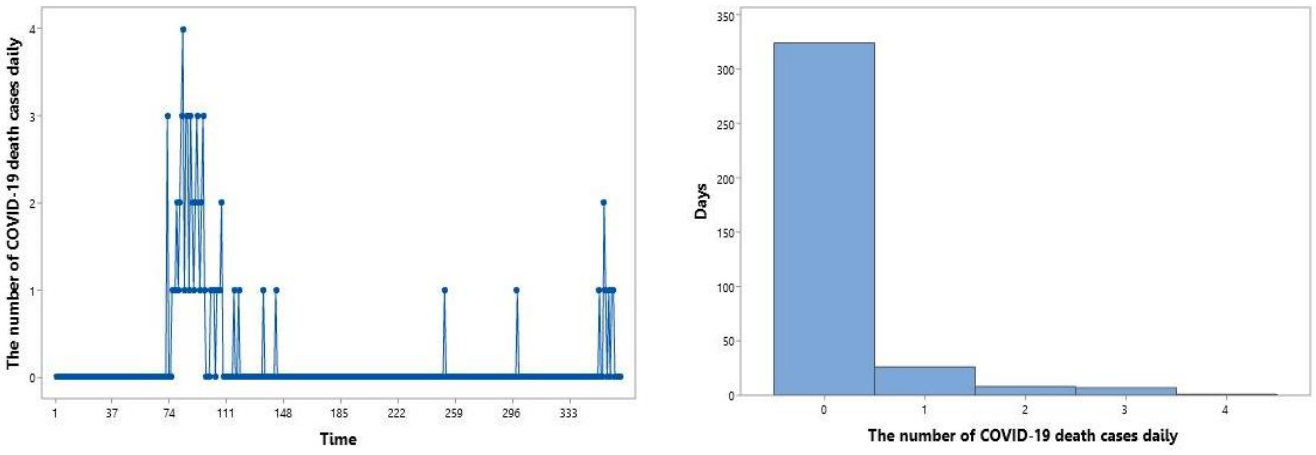
From the likelihood function in (21) and the prior distribution in (25), we derive the posterior distribution as follows:

$$\pi(\boldsymbol{\Lambda} \mid X) \propto \prod_{i=1}^{n} f(y_i \mid x_i^T, \boldsymbol{\Lambda})\pi(r)\pi(a)\pi(b)\pi(\boldsymbol{\Theta}). \tag{26}$$

For the NB-QL model, the full conditional posterior distributions for each parameter of $\boldsymbol{\Lambda}$ derived from (25) are obtained as:

$$\pi(r \mid \mathbf{y}, \mathbf{X}, r, a, b) \propto \prod_{i=1}^{n} f(y_i \mid x_i^T, \boldsymbol{\Lambda})\pi(r), \pi(a \mid \mathbf{y}, \mathbf{X}, r, a, b) \propto \prod_{i=1}^{n} f(y_i \mid x_i^T, \boldsymbol{\Lambda})\pi(a),$$

$$\pi(a \mid \mathbf{y}, \mathbf{X}, r, a, b) \propto \prod_{i=1}^{n} f(y_i \mid x_i^T, \boldsymbol{\Lambda})\pi(b), \pi(\boldsymbol{\Theta} \mid \mathbf{y}, \mathbf{X}, r, a, b) \propto \prod_{i=1}^{n} f(y_i \mid x_i^T, \boldsymbol{\Lambda})\pi(\boldsymbol{\Theta}).$$



(a) (b)

Figure 2 (a) A plot between the number of COVID-19 death cases daily in Thailand (Y_t) and the time (t); (b) The observed frequency (days) of the number of people for COVID-19 deaths in Thailand.

Table 1 Summary of the actual data.

Variables	Minimum	Maximum	Median	Mean	Standard deviation	Variance
Y_t	0	4	0	0.18	0.58	0.34
X_1	0	745	5	25.11	68.87	4,743.08
I	0	1				

In this study, the model parameters of $\mathbf{\Lambda}$ can be estimated from the Bayesian method using the MCMC algorithm to produce the posterior inference for each parameter. Based on these prior densities, we generated 3 parallel independent MCMC chains for 30,000 iterations in each parameter, discarding the 1st 15,000 iterations as a burn-in for computation. In this paper, the expected posterior of parameters is calculated using the jags function in the R2jags package of the R language [28,29].

Generalized linear model for time series count data on the daily COVID-19 pandemic

Empirical data

The data used are the number of the COVID-19 cases daily in Thailand from 1 January 2020 to 31 December 2020, comprising 366 days of daily [14]. All random variables in this study are as follows, i.e., 1) Y_t is time series data of the number of COVID-19 death cases daily (unit: People); 2) X_1 is the number of a new COVID-19 infection cases (unit: People); 3) I_1 is the intervention variable for the time $73 \le t \le 143$ ($I_1 = 1$ for $73 \le t \le 143$, otherwise $I_1 = 0$); 4) I_2 is the intervention variable for the time $t \ge 352$ ($I_2 = 1$ for $t \ge 352$, otherwise $I_2 = 0$).

From the actual data, there were COVID-19 deaths recorded as 0 people for 324 days, 1 person for 26 days, 2 people for 8 days, 3 people for 7 days, and 4 people for 1 day (**Figure 2**(**b**)). The mean and variance of the number of people of the COVID-19 deaths (people) are 0.18 and 0.38, respectively (**Table 1**). Since the variance of the number of the COVID-19 deaths is greater than mean, this data set has an over-dispersion problem.

Stages of data analysis

Stages of analysis of the model are as follows:

1) Explore data to see a general data overview through a plot of the number of the COVID-19 death cases daily from 1 January 2020 to 31 December 2020.

2) Define the P and Q as initial values by forming an ACF and PACF plot, specifyingthe optimal combination based on the smallest AIC value.

3) Modeling with the NB-QL, Poisson, and NB distribution approaches by adding internal covariate effects, and estimating model parameters with the Bayesian approach.

4) Comparing the performance of the proposed model with the Poisson and NB distributions. The present study chose the Poisson and NB distribution to compare the NB-QL distribution because the 3 distributions were related. Namely, the NB-QL distribution was a mixed distribution between the QL and NB distributions. Meanwhile, the NB distribution is obtained by mixing the Poisson distribution and the gamma distribution.

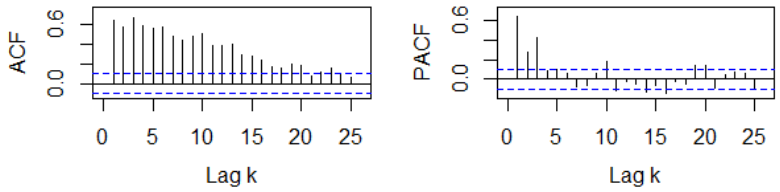


Figure 3 The plot between the number of COVID-19 death cases daily in Thailand (Y_t), ACF plot, and PACF plot of Y_t.

P and Q value determination

Before modeling the initial step, determine values of P and Q based on the partial autocorrelated function (PACF) and autocorrelated function (ACF) plot. The result is shown in **Figure 3**. So the values of P and Q to be tested are P = {1, 2, 3, 10, 11, 14, 16, 19, 20, 25} and Q = {1 to 23}, then from the values of P and Q it is possible to determine the optimal combination based on the smallest AIC value (Ahmad and Francq, 2016). For the numerous combinations listed of P and Q in **Figure 4**, the optimal combination is $P=3$ and $Q=2(AIC=322.30)$, and then these values are used in modeling.

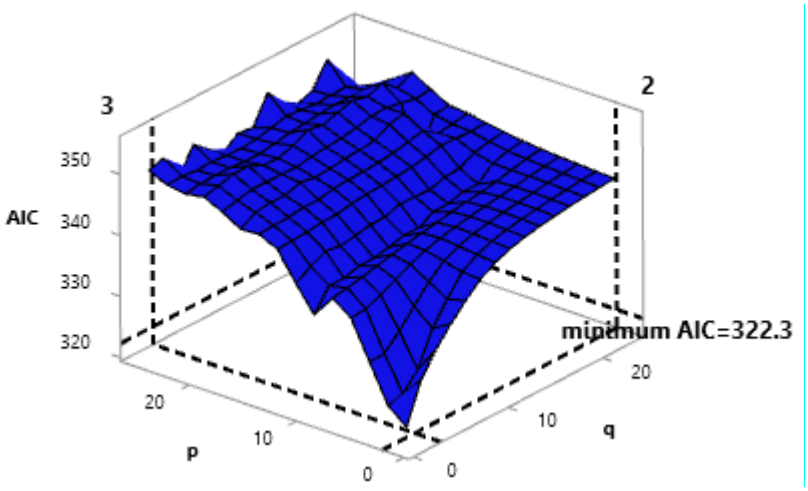


Figure 4 Specifying the optimal combination based on the smallest AIC value with the NB-QL distribution.

Criteria for model evaluation

In the model comparison stage, 3 criteria are considered for model comparisons:

1) The deviance is $D(\Omega)=[-2\log L(y\,|\,\mathbf{\Omega})]$ where $L(y\,|\,\mathbf{\Omega})$ is the likelihood function, and the conditional joint pdf of the observations is given unknown parameters.

2) The DIC is regarded as a generalization of Akaike's information criterion and the Bayesian information criterion, which is often and widely used as a goodness-of-fit measure when we use the Bayesian approach. The DIC is defined as $\mathrm{DIC}=\bar{D}(\mathbf{\Omega})+p_D$, for $\bar{D}(\mathbf{\Omega})=\mathrm{E}[-2\log L(y\,|\,\mathbf{\Omega})]$ and $p_D=Var[D(\mathbf{\Omega})]/2$, where the first term is the posterior mean of the deviance, and the 2nd term is an alternative measure of the effective number of parameters [29]. The DIC is beneficial to Bayesian model comparison problems where the posterior distributions have been obtained by MCMC simulation [30,31]. Therefore, deviance and DIC are statistics to compare the models. The model has the smallest value of deviance and DIC is the best model.

3) Probability Integral Transform (PIT) is a tool for assessing the probability calibration of the predictive distribution, which will follow a uniform distribution if the predictive distribution is correct [32]. For count data, Czado *et al*. [33] define a non-randomized PIT value by:

$$F_t(u\,|\,y)=\begin{cases}0 & ;\ u\le P_t(y-1)\\ \dfrac{u-P_t(y-1)}{P_t(y)-P_t(y-1)} & ;\ P_t(y-1)<u<P_t(y)\\ 1 & ;\ \mathrm{u}\ge P_t(y)\end{cases} \tag{27}$$

where $P_t(y)$ is the predictive distribution of the observed value y_t. The shape of the PIT histograms suggests the calibration accuracy of the predictive distribution. A convex shape indicates an under-dispersed predictive distribution, whereas concave histograms refer to over-dispersed predictive distributions.

Table 2 Parameter estimates and various statistics of the NB-QL, Poisson, and NB models.

Models	Parameters	Estimate (s.e.)	95 % CR.I	Deviance	DIC	p_D
NB-QL	β_0	–5.501 (1.502)	(–8.641, –2.803)	203.482	212.50	9.00
	β_3	0.514 (0.132)	(0.272, 0.780)			
	α_2	–2.884 (1.165)	(–5,135, –0.631)			
	η_1	0.013 (0.188)	(–0.385, 0.364)			
	η_2	4.828 (0.843)	(3.341, 6.673)			
	η_3	4.883 (1.367)	(2.245, 7.604)			
Poisson	β_0	–4.717 (0.922)	(–6.994, –3.142)	199.900	588.1	388.2
	β_3	0.457 (0.102)	(0.258, 0.652)			
	α_2	–2.623 (0.975)	(–4.643, –0.750)			
	η_1	0.022 (0.151)	(–0.304, 0.293)			
	η_2	4.800 (0.984)	(3.081, 7.195)			
	η_3	4.741 (1.366)	(2.036, 7.589)			
NB	β_0	–4.743 (0.856)	(–6.735, –3.334)	203.750	213.80	10.10
	β_3	0.511 (0.136)	(0.252, 0.787)			
	α_2	–2.913 (1.149)	(–5.195, –0.639)			
	η_1	0.013 (0.188)	(–0.380, 0.358)			
	η_2	4.868 (0.930)	(3.307, 7.910)			
	η_3	4.914 (1.438)	(2.226, 7.910)			

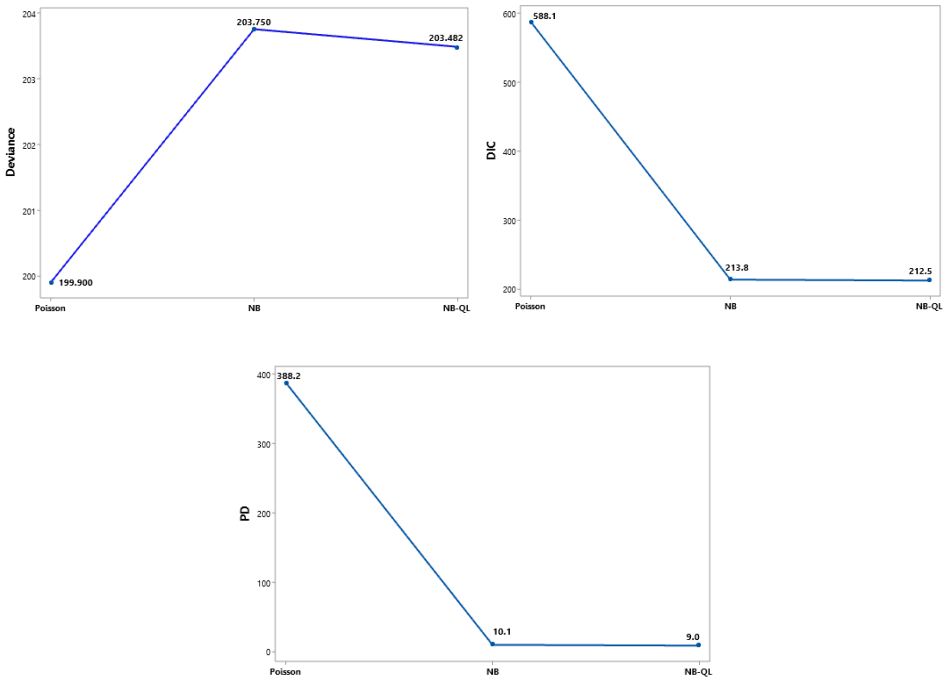


Figure 5 The plot of Deviance, DIC, and p_D of Poisson NB and NB-QL models.

Results of data analysis

In this section, the results of the data analysis provide an illustrative method of applying the generalized linear models of the time series count data to build the model derived for a new mixed NB

distribution: The NB-QL. This can be compared with some traditional approaches: The Poisson and NB distributions. The Bayesian approach is the method used to estimate the parameters of the count time series model. The posterior means, estimates, standard error (s.e.), 95 % credible intervals (Cr.I.) of each parameter, and statistics of the deviance, DIC, and p_D of the NB-QL, Poisson, and NB models, are presented in **Table 2**. The results of the parameter estimation of the NB-QL distribution in the generalized linear models of the time series count data revealed that the estimated parameters r, a and b are: $\hat{r} = 3.153 (s.e. = 1.377)$, $\hat{a} = 1.004$ (s.e. = 0.994) and $b = 1.015$ (s.e. = 1.023), respectively. The result of the estimation of the parameter models can be seen in **Table 2**. The parameter estimate of the NB-QL distribution is obtained as:

$$\mathrm{E}(Y_T \mid F_{t-1}) = \mu_i \mathrm{E}(\lambda) \text{ where } \mathrm{E}(\lambda) = \frac{1}{a}\left(\frac{b+2}{b+1}\right) = 1.490 .$$

According to **Table 2**, the number of COVID-19 death cases daily in Thailand with GLMs for the count data approach with the NB-QL distribution can be represented as:

$$\begin{aligned} \hat{Y} &= \mathrm{E}(\lambda)\exp\{\hat{\beta}_0 + \hat{\beta}_1 Y_{t-3} + \hat{\alpha}_1 \lambda_{t-2} + \hat{\eta}_1 Z + \hat{\eta}_2 I_1 + \hat{\eta}_3 I_2\} \\ &= 1.490\exp\{-5.501 + 0.514 Y_{t-3} - 2.884\lambda_{t-2} + 0.013Z + 4.828 I_1 + 4.883 I_2\}. \end{aligned} \quad (28)$$

The deviance, DIC, and p_D of the generalized linear models of the NB-QL model are 203.482, 212.50, and 9.00 respectively (**Table 2** and **Figure 5**). It can be described that with the time between $73 \le t \le 143$ and $t \ge 352$, there are internal covariate effects due to the data, as there was a surge in the number of the COVID-19 death cases daily in Thailand (Y_t). This model indicates that the average of Y_t is influenced by the number of Y_t in the previous 3 days. The average number of COVID-19 death cases daily in Thailand also influenced the previous 2 days. At the same time, the number of infected cases daily in Thailand $Z_1 = (X_1 - 25.11)/68.87$ is influenced by the number of COVID-19 death cases daily. Since there is a variable intervention in certain days, the intervention will be added with a specific coefficient as it is written in the model. For considering the traditional Poisson distribution, the parameter estimate of the Possion distribution is obtained as: $\mathrm{E}(Y_i \mid F_{t-1}) = \mu_i$. Consequently, the GLMs for the count data approach with the Possion distribution can be represented as:

$$\begin{aligned} \hat{Y} &= \exp\{\hat{\beta}_0 + \hat{\beta}_1 Y_{t-3} + \hat{\alpha}_1 \hat{\lambda}_{t-2} + \hat{\eta}_1 Z + \hat{\eta}_2 I_1 + \hat{\eta}_3 I_2\} \\ &= \exp\{-4.717 + 0.457 Y_{t-3} - 2.623\lambda_{t-2} + 0.022Z + 4.800 I_1 + 4.741 I_2\}. \end{aligned} \quad (29)$$

The deviance, DIC, and p_D of the generalized linear models of the Possion model are 199.900, 588.1, and 388.2 respectively (**Table 2** and **Figure 5**). In the same way, the GLMs for the count data approach with the NB distribution can be represented as:

$$\begin{aligned} \hat{Y} &= \exp\{\hat{\beta}_0 + \hat{\beta}_1 Y_{t-3} + \hat{\alpha}_1 \hat{\lambda}_{t-2} + \hat{\eta}_1 Z + \hat{\eta}_2 I_1 + \hat{\eta}_3 I_2\} \\ &= \exp\{-4.743 + 0.511 Y_{t-3} - 2.913\lambda_{t-2} + 0.013Z + 4.868 I_1 + 4.914 I_2\}. \end{aligned} \quad (30)$$

The deviance, DIC, and p_D of the generalized linear models of the NB-QL model are 203.750, 213.80, and 10.10 respectively (**Table 2** and **Figure 5**).

Comparison of model performance

The next stage evaluates the 3 approaches' model performance based on deviance, DIC, p_D and the PIT histogram criteria. As regards the performance of the NB-QL, NB, and Poisson time series models from the criteria: deviance, DIC, and p_D, it is found that the NB-QL model has the highest efficiency when compared with the NB and Poisson models. When considering the PIT histogram criteria, the PIT histogram for the NB-QL, Poisson, and NB distributions can be seen in **Figure 6**.

Figure 6 shows that the Poisson model is still not good because it has not approached the uniform distribution form. In contrast, the NB-QL and NB models have approached the uniform. It means that the models with the NB-QL and NB are performing better compared to the Poisson approach. But it is seen from the BIC, DIC, p_D and the PIT histogram values that the model with the NB-QL distribution is the best model among the 3 model approaches. The goodness of the model is also demonstrated by the actual plot of the data comparison with the modeling results following the 3 approaches:

Figure 7 shows the data plot with the other 3 estimators. This plot shows that all the plot estimates have a reasonably good value because they follow the actual data pattern data. However, the blue plot, namely the estimation with the NB-QL model, has a value closer to the actual data plot, which is also substantial with the smallest BIC, DIC and p_D values.

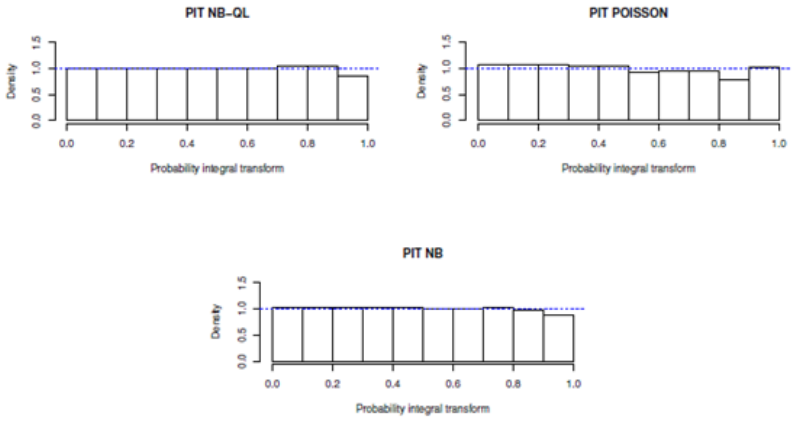


Figure 6 The PIT histogram with the NB-QL, Poisson, and NB models.

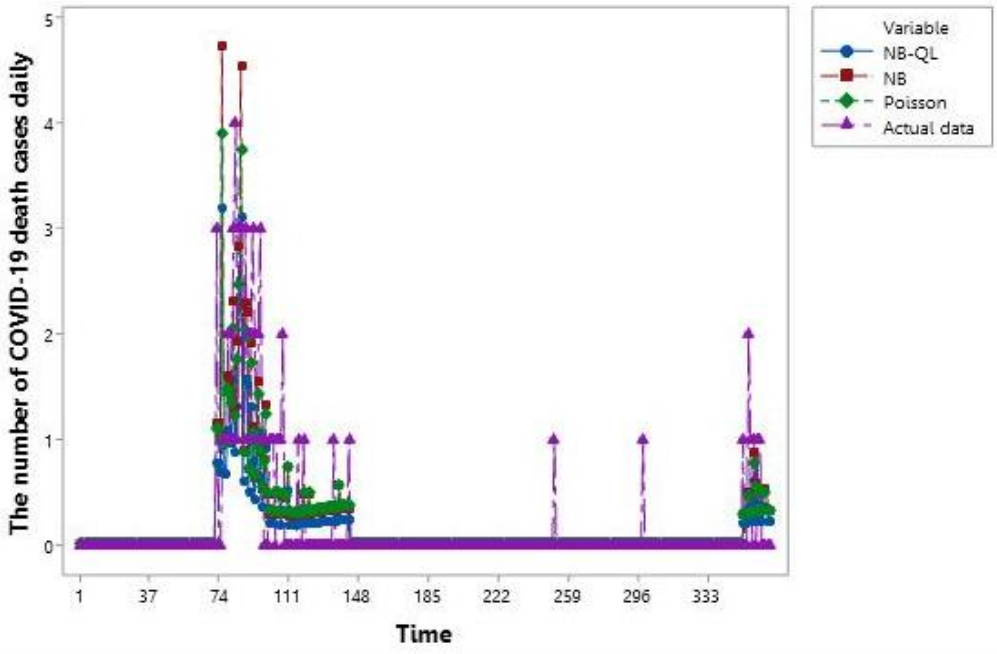


Figure 7 Comparison of actual data and models.

Conclusions

This work proposes the new mixed NB distribution, which is called the negative binomial-quasi Lindley (NB-QL) and applies the newly created distribution with a GLMs framework to build the time series count data model. Where the NB-QL and NB distributions, there are both similarities and differences. Namely, NB-QL and NB models are used for count data analysis. However, the NB-QL model can describe data better than the NB model for over-dispersed count data or count data with heavy-tailed since the NB-QL has flexible better than the NB model because the NB-QL model has 2 shaped parameters. In contrast, the NB model has 1 shaped parameter. In the application of the NB-QL model, we compare the accuracy of the proposed distribution with the Poisson model and an NB model using the actual data sets: The number of COVID-19 death cases daily in Thailand from 1 January 2020 to 31 December 2020, for which this data set has the observed sample of 366 days. This data set has over-dispersion problems, the resulting model is compared with some traditional models: The Poisson and NB models. The new findings show that the NB-QL model has the highest efficiency when compared with the NB and Poisson models. One of the reasons the NB-QL performs best is because the developed NB-QL distribution is more fitted to the actual data in the form of time series count data than the classical distribution. According to the results of the study, it is found overall that the NB-QL time series model was more suitable than the Poisson or NB models. The DIC and p_D of the new model were lower than that of the Poisson model at 63.87 and 97.68 %, respectively, while the deviance of the new model was slightly higher than that of the Poisson model at 1.76 %. When considering the comparison between the new model and the NB model, the deviance, DIC and p_D values of the NB-QL time series were lower than those of the NB model at 0.12, 0.61 and 10.89 %, respectively. When considering the model performance using the PIT histogram, the Poisson distribution model is still not good because it has not approached the uniform distribution form. In contrast, the model with the NB-QL distribution and NB has approached the uniform. Based on deviance, DIC, p_D and the PIT histogram, we can see that the proposed model is also suitable for forecasting the the number of COVID-19 death cases daily in Thailand, indicating that the NB-QL time series model was another efficient alternative to modeling for count data that has an over-dispersion problem. According to the NB-QL time series model concerning the number of COVID-19 death cases daily in Thailand from 1 January 2020 to 31 December 2020, this model indicates that the average of the number of COVID-19 death cases daily in Thailand is influenced by the number of COVID-19 death cases daily in Thailand in the previous 3 days. The average number of COVID-19 death cases daily in Thailand also influenced the previous 2 days. At the same time, the number of infected cases daily in Thailand is influenced by the number of COVID-19 death cases daily. In addition, there are also the components of interventions of internal covariate effects due to the data, as there was a surge in the number of COVID-19 death cases daily in Thailand at the time between $73 \le t \le 143$ and $t \ge 352$.

The establishment of a data analysis model without considering the dispersion problem will lead to inaccurate parameter estimation. The common method is to use the NB distribution instead of the Poisson distribution [34]. According to a study by [35-37], it was found that the NB model was more appropriate than the Poisson model when there is over-dispersion of the data. In this study, applying the GLMs framework for time series count data to build the time series model derives a new mixed NB distribution: The NB-QL, whereby the results showed that the NB-QL time series model was more appropriate than the NB and Poisson models. This is because the Poisson distribution is a special case of the NB distribution. But if the variance is greater than the mean, the new mixed NB distribution in this study is an extremely effective alternative for modeling count data in the context of over-dispersion.

Acknowledgements

The authors gratefully acknowledge the participation in the Faculty of Science and Technology, Rajamangala University of Technology Thanyaburi, Thailand. We are also thankful to those who could not be mentioned here for their kindness and encouragement. And finally, the authors would like to thank the anonymous reviewers for their comments and suggestions.

References

[1] JA Nelder and RWM Wedderburn. Generalized linear models. *J. R. Stat. Soc. Ser. A* 1972; **135**, 370-84.

[2] AC Cameron and PK Trivedi. *Regression analysis of count data*. 2nd eds. Cambridge University Press, New York, 2013.

[3] M Greenwood and GU Yule. An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *J. R. Stat. Soc.* 1920; **83**, 255-79.

[4] W Gardner, EP Mulvey and EC Shaw. Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychol. Bull.* 1995; **118**, 392-404.

[5] JS Long and JS Long. *Regression models for categorical and limited dependent variables.* SAGE Publications, New York, 1997.

[6] AC Cameron and P Johansson. Count data regression using series expansions: With applications. *J. Appl. Econ*. 1997; **12**, 203-23.

[7] H He, W Tang, W Wang and P Crits-Christoph. Structural zeroes and zero-inflated models. *Shanghai Arch. Psychiatry* 2014; **26**, 236-42.

[8] Z Wang. One mixed negative binomial distribution with application. *J. Stat. Plan. Inference* 2011; **141**, 1153-60.

[9] H Zamani and N Ismail. Negative binomial-Lindley distribution and its application. *J. Math. Stat.* 2010; **6**, 4-9.

[10] S Aryuyuen and W Bodhisuwan. The negative binomial-generalized exponential (NB-GE) distribution. *Appl. Math. Sci*. 2013; **7**, 1093-105.

[11] Y Gençtürk and A Yiğiter. Modelling claim number using a new mixture model: Negative binomial gamma distribution. *J. Stat. Comput. Simul.* 2016; **86**, 1829-39.

[12] D Yamrubboon, W Bodhisuwan, C Pudprommarat and L Saothayanun. The negative binomial-Sushila distribution with application in count data analysis. *Thail. Stat.* 2017; **15**, 69-77.

[13] S Aryuyuen. Bayesian inference for the negative binomial-generalized Lindley regression model: Properties and applications. *Commun. Stat. Theory Methods* 2021. https://doi.org/10.1080/03610926.2021.1995434

[14] Department of Disease Control. Daily covid-19 report, Thailand information. Daily COVID-19 report, Available at: https://data.go.th/dataset/covid-19-daily, accessed January 2022.

[15] A Heinen. Modelling time series count data: An autoregressive conditional Poisson model, Available at: http://dx.doi.org/10.2139/ssrn.1117187, accessed May 2021.

[16] R Ferland, A Latour and D Oraichi. Integer-valued GARCH process. *J. Time Ser. Anal.* 2006; **27**, 923-42.

[17] K Fokianos, A Rahbek and D Tjøstheim. Poisson autoregression. *J. Am. Stat. Assoc.* 2009; **104**, 1430-9.

[18] F Zhu. A negative binomial integer-valued GARCH model. *J. Time Ser. Anal*. 2011; **32**, 54-67.

[19] S Fu. A hierarchical Bayesian approach to negative binomial regression. *Methods Appl. Anal.* 2015; **22**, 409-28.

[20] S Fu. Hierarchical Bayesian LASSO for a negative binomial regression. *J. Stat. Comput. Simul*. 2016; **86**, 2182-203.

[21] D Yamrubboon, A Thongteeraparp, W Bodhisuwan, K Jampachaisri and A Volodin. Bayesian inference for the negative binomial-Sushila linear model. *Lobachevskii J. Math.* 2019; **40**, 42-54.

[22] A Gelman, JB Carlin, HS Stern, DB Dunson, A Vehtari and DB Rubin. *Bayesian data analysis*. CRC Press, New York, 2013.

[23] T Liboschik, K Fokianos and R Fried. Tscount: An R package for analysis of count time series following generalized linear models. *J. Stat. Softw.* 2017; **82**, 1-51.

[24] K Fokianos and D Tjøstheim. Log-linear Poisson autoregression. *J. Multivar. Anal.* 2011; **102**, 563-78.

[25] R Shanker and A Mishra. A quasi Lindley distribution. *Afr. J. Math. Comput. Sci. Res.* 2013; **6**, 64-71.

[26] T Harris, JM Hilbe and JW Hardin. Modeling count data with generalized distributions. *Stata J.* 2014; **14**, 562-79.

[27] SR Geedipally, D Lord and SS Dhavala. The negative binomial-Lindley generalized linear model: Characteristics and application using crash data. *Accid. Anal. Prev.* 2012; **45**, 258-65.

[28] R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, Available at: https://www.R-project.org, accessed May 2021.
[29] Y Su and M Yajima. R2jags: Using R to Run 'JAGS'. R package version 0.7-1, Available at: https://CRAN.R-project.org/package=R2jags, accessed January 2022.
[30] D Lunn, C Jackson, N Best, A Thomas and D Spiegelhalter. *The bugs book: A Practical Introduction to Bayesian analysis*. Chapman Hall, London, 2013.
[31] DJ Spiegelhalter, NG Best, BP Carlin and AVD Linde. Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. C* 2002; **64**, 583-639.
[32] C Genest and J Neslehova. A primer on discrete copulas. *ASTIN Bull.* 2007; **37**, 475-515.
[33] C Czado, T Gneiting and L Held. Predictive model assessment for count data. *Biometrics* 2009; **65**, 1254-61.
[34] P McCullagh and JA Nelder. *Generalized linear models*. Routledge, Boca Raton, Florida, 2019.
[35] AL Byers, H Allore, TM Gill and PN Peduzzi. Application of negative binomial modeling for discrete outcomes: A case study in aging research. *J. Clin. Epidemiol*. 2003; **56**, 559-64.
[36] A Yesilova and A Yilmaz. The application of overdispersion and generalized estimating equations in repeated categorical data related to the sexual bahaviour traits of farm animals. *J. Appl. Sci*. 2007; **7**, 1762-7.
[37] DT Molla and B Muniswamy. Power of tests for overdispersion parameter in negative binomial regression model. *IOSR J. Math*. 2012; **1**, 29-36.