JOURNAL OF APPLIED STATISTICS
2024, VOL. 51, NO. 7, 1227–1250
https://doi.org/10.1080/02664763.2023.2194582

Taylor & Francis
Taylor & Francis Group

The balanced discrete Burr–Hatke model and mixing INAR(1) process: properties, estimation, forecasting and COVID-19 applications

Seyedeh Mahbubeh Hoseini Baladezaei, Einolah Deiri and Ezzatallah Baloui Jamkhaneh

Department of Statistics, Qaemshahr Branch, Islamic Azad University, Qaemshahr, Iran

ABSTRACT

The main concern of this paper is providing a flexible discrete model that captures every kind of dispersion (equi-, over- and under-dispersion). Based on the balanced discretization method, a new discrete version of Burr–Hatke distribution is introduced with the partial moment-preserving property. Some statistical properties of the new distribution are introduced, and the applicability of proposed model is evaluated by considering counting series. A new integer-valued autoregressive (INAR) process based on the mixing Pegram and binomial thinning operators with discrete Burr–Hatke innovations is introduced, which can model contagious data properly. The different estimation approaches of parameters of the new process are provided and compared through the Monte Carlo simulation scheme. The performance of the proposed process is evaluated by four data sets of the daily death counts of the COVID-19 in Austria, Switzerland, Nigeria and Slovenia in comparison with some competitor INAR(1) models, along with the Pearson residual analysis of the assessing model. The goodness of fit measures affirm the adequacy of the proposed process in modeling all COVID-19 data sets. The fundamental prediction procedures are considered for new process by classic, modified Sieve bootstrap and Bayesian forecasting methods for all COVID-19 data sets, which is concluded that the Bayesian forecasting approach provides more reliable results.

ARTICLE HISTORY
Received 10 March 2022
Accepted 14 March 2023

KEYWORDS
Balanced discrete; Bayesian forecasting; Burr–Hatke; mixing Pegram and thinning operators; modified sieve bootstrap

MATHEMATICS SUBJECT CLASSIFICATIONS
60E05; 60-02; 62P10

1. Introduction

Modeling the counting events has attained much attention in recent years, due to the broad applicability of these models, in various branch of sciences. The counting variables are observed in real-life phenomena, such as the number of damaging accidents in insurance, number of loss-making companies in economics, number of earthquakes in geology, number of cancer cells in medicine, and so on. The continuous probability models cannot provide adequate fitting for some discrete data sets. Therefore, the continuous distributions are replaced by discrete distributions to model count data, such as binomial, Poisson, geometric, and so on. Unfortunately, classic discrete distributions cannot cover all kinds of

CONTACT Einolah Deiri deiri53@gmail.com Department of Statistics, Qaemshahr Branch, Islamic Azad University, Qaemshahr, Iran

data. For instance, the binomial distribution has restricted domain, the Poisson distribution cannot be applied for over- or under-dispersed data, the geometric distribution cannot handle data sets with non-constant hazard rate function. Hence, some new discrete distributions are provided based on the classic version of distributions through various discretization methods.

One of the widely used procedures is the survival discretization method. In this method, the survival function is preserved among the continuous and discrete versions of the distribution. Some of the former discrete models, based on the survival discretization method, are discrete Normal [23], discrete Rayleigh [24], discrete Gamma [8] and discrete type-II half-Logistic Exponential [1].

Hagmark [15] introduced the mean-preserving discretization method with any interval domain, any theoretically possible mean-variance pair and different shapes that mean function is fixed between continuous and discrete analogous. Recently, Tovissodé *et al.* [29] introduced a new discretization approach based on partial moment preserving called the balanced discretization (BD) method, with simple stochastic representation and various types of dispersion. They introduced the balanced discrete Gamma distribution as a sub-model.

Lately, Al-Babtain *et al.* [3] introduced the natural discrete Lindley distribution, which is represented as a mixture of geometric and negative binomial distributions. Based on three practical data sets, they showed that their distribution outperformed various traditional discrete distributions. The flexible discrete model with one parameter introduced by Eliwa and El-Morshedy [12] has different shaped hazard rate functions (HRF), including increasing, decreasing or unimodal-shaped.

Due to the complex nature and some unique properties of the real-life phenomena, such as skewness, dispersion, inflation or deflation, and monotone or unimodal failure rate, the former discrete distributions are not convenient to capture and model various aspects of some real data. Hence, we introduce a new version of Burr–Hatke (BH) distribution under the balanced discretization method.

The discrete version of the BH model has been recently proposed by El-Morshedy *et al.* [11] to model count events exhibiting huge over-dispersion with varied skewness features. The discrete BH distribution of El-Morshedy *et al.* [11] can only be used for over-dispersed data, and the applicability of the new distribution for time series data is not checked. El-Morshedy *et al.* [11] have used the survival discretization method to obtain the discrete BH distribution, while in the present work, we use the balanced discretization method of Tovissodé *et al.* [29]. The new flexible discrete version of BH distribution handles all kinds of dispersion. We check the performance of the new distribution by considering the INAR(1) process for the count time series. The estimation, real data analysis and forecasting of the proposed process are provided comprehensively.

Since the prevalence of the COVID-19 pandemic, there has been ongoing research from different medical, statistics, and societal scholars to confine the count of infected, deaths, or hospital admissions due to this virus and its variants. While the medical researchers have been particularly focusing on the development of vaccines, the statisticians and data scientists have been toiling hard to provide efficient analyses and forecasts of the different mentioned variables.

Several models of the COVID-19 data are provided in recent articles. For instance, we cite nonlinear growth models for COVID-19 in Iraq [2], Kermack–McKendrick-type

compartmental (deterministic) epidemic of COVID-19 model in a homogeneously mixed population [14], modeling with flexible extension log–logistic tangent distribution in Somalia [19], autoregressive time series modeling and forecasting of cases of COVID-19 based on two-piece scale mixture normal distributions [18]. Recently, Pourreza *et al.* [21] provided a family of Gamma-generated distribution with attention to Gamma–Weibull distribution and represented the applicability of new distribution fitting COVID-19 data set from Saarland.

In particular, since the COVID-19 variables are collected on a sequential time basis, the research emphasis has been on exploring classical time series approaches. One of the reputable models for the counting time series is the INAR model that was applied in manifold branches of science. The first-order integer-valued autoregressive (INAR(1)) process is introduced based on the thinning operator, which is extended in the last years in different aspects, for instance, some new thinning operators, the random coefficient thinning, different marginal and innovation distributions. As the pioneer of INAR(1) models, we cite Poisson INAR(1) model [5], geometric and negative binomial INAR(1) [6] and generalized Poisson model [7]. Some new articles of INAR(1) modeling for COVID-19 data sets are investigated by Shamma and Mohammadpour [25], Triacca and Triacca [30] and Chattopadhyay *et al.* [9].

The main purpose of this paper is devoted to introducing a new discrete distribution with outstanding and flexible features. On the other side, we provide a pliant INAR(1) model based on discrete BH innovations and mixing Pegram and binomial thinning operators, introduced by Khoo *et al.* [17]. The applicability of the proposed model is exposed on some COVID-19 data sets. The new INAR(1) process stands out as a major competitor to some count models as it yields far more superior fitting criteria in modeling COVID-19 data sets.

The structure of this paper is organized as follows. The basic definition and features of the BD method and, accordingly, a new discrete model named balanced discrete Burr–Hatke distribution are provided in Section 2. In Section 3, a new INAR(1) process based on the mixing Pegram and binomial thinning operator is introduced with the balanced discrete Burr–Hatke innovations, along with several statistical properties of the proposed process. In Section 4, the parameters of the new process are estimated by the conditional maximum likelihood, Bayesian, modified conditional least square and Yule–Walker estimation methods. The Monte Carlo simulation comparison among different estimation techniques is accomplished in Section 5 to distinguish the efficient estimation method. Finally in Section 6, the applicability of the proposed process is investigated using four medical count data, which demonstrates the appropriateness of our model in comparison with some relevant INAR(1) models. Several forecasting methods containing the classic, modified Sieve Bootstrap and Bayesian methods are considered for all COVID-19 data sets, which recommended the Bayesian forecasting method to be applied for future research.

2. The BDBH distribution description

The balanced discretization method is introduced by Tovissodé *et al.* [29], with special work on balanced discrete Gamma distribution. We concentrate on the balanced discrete

strategy and introduce a new discrete distribution based on the BH baseline distribution, named balanced discrete BH (BDBH) model.

In the next section, we provide the basic definition of the balanced discrete method to define a discrete version of the continuous distribution [29] which preserves the partial moments of the base continuous model containing expectation.

2.1. The balanced discrete method and BDBH distribution

Definition 2.1 (Balanced discretization): The random variable Z is distributed as the balanced discrete counterpart denoted BD(β) of the continuous distribution CD(β) if it has the following stochastic representation [29]:

$$\begin{aligned} Z \mid U = u, Y = y &\overset{d}{=} v + u, \\ U \mid Y = y &\sim \text{Ber}(r), \\ Y &\sim \text{CD}(\beta), \end{aligned}$$

where Ber(.) denotes the Bernoulli distribution, $v = [y]$ that $[y]$ denotes the integer part of any real y and $r = y - v$.

Denote the probability density function (PDF), cumulative distribution function (CDF) of the continuous random variable Y as $g_Y(., \beta)$ and $G_Y(., \beta)$, respectively. The probability mass function (PMF) and CDF of BD distribution are defined as follows:

$$\begin{aligned} f_Z(z, \beta) &= (z-1)G_Y(z-1, \beta) - 2z\, G_Y(z, \beta) + (z+1)G_Y(z+1, \beta), \\ &\quad + E_Y(1, z-1 \mid \beta) - E_Y(1, z \mid \beta), \quad z \in \mathbb{Z} \end{aligned} \tag{1}$$

$$F_Z(z, \beta) = G_Y(z, \beta) + (z+1)H_Y(z, \beta) - E_Y(1, z \mid \beta), \tag{2}$$

where $H_Y(z, \beta) = G_Y(z+1, \beta) - G_Y(z, \beta)$ and $E_Y(n, z \mid \beta)$ stands the nth partial moment as below

$$E_Y(n, z \mid \beta) = \int_z^{z+1} y^n g_Y(y, \beta)\, \mathrm{d}y.$$

Corollary 2.2: *The expectation and variance of BD distribution are represented as below* [29]

$$E(Z) = E(Y), \quad Var(Z) = Var(Y) + \zeta_0(\beta),$$

where $\zeta_0(\beta) = \sum_{i=-\infty}^{\infty} \zeta_0(i, \beta)$ *with*

$$\zeta_0(i, \beta) = (2i+1)E_Y(1, i \mid \beta) - E_Y(2, i \mid \beta) - i(i+1)H_Y(i, \beta).$$

In the following, we consider the BH distribution as the CD counterpart in the balanced discrete method and introduce a new flexible discrete BH distribution. El-Morshedy *et al.* [11] provided the PDF, CDF and hazard rate function (HRF) of BH distribution,

respectively as

$$g_Y(y,\beta) = \frac{1+\beta(y+1)}{(y+1)^2}\,\mathrm{e}^{-\beta y}, \quad y \geq 0\ \beta > 0,$$
$$G_Y(y,\beta) = 1 - \frac{\mathrm{e}^{-\beta y}}{y+1},$$
$$\mathrm{HRF}_Y(y,\beta) = \frac{g_Y(y,\beta)}{1-G_Y(y,\beta)} = \frac{1}{y+1} + \beta.$$

We provide the quantile function of the BH distribution, derived from the CDF, as below

$$\mathrm{QF}_Y(p,\beta) = -1 - \frac{1}{\beta} W\left(\frac{-\beta\,\mathrm{e}^{\beta}}{1-p}\right), \quad 0 < p < 1, \tag{3}$$

where $W(.)$ is the Lambert function, and it is defined to be the function satisfying $p = W(z)\,\mathrm{e}^{W(z)}$.

The first and second moments of BH distribution are obtained as

$$E(Y) = -\mathrm{e}^{\beta}\mathrm{Ei}(-\beta),$$
$$E(Y^2) = 2\left(\frac{1}{\beta} + \mathrm{e}^{\beta}\mathrm{Ei}(-\beta)\right),$$

where Ei(.) is the exponential integral function as $\mathrm{Ei}(x) = \int_{-\infty}^{x} \frac{\mathrm{e}^t}{t}\,\mathrm{d}t$. We compute the first partial moment of BH distribution as below

$$\begin{aligned} E_Y(1,z\,|\,\beta) &= \int_z^{z+1} y\,g_Y(y,\beta)\,\mathrm{d}y \\ &= \int_z^{z+1} \frac{\mathrm{e}^{-\beta y}}{y+1}\,\mathrm{d}y + \beta\int_z^{z+1} \mathrm{e}^{-\beta y}\,\mathrm{d}y + \int_z^{z+1} \mathrm{e}^{-\beta y}\left(\frac{-\beta}{y+1} - \frac{1}{(y+1)^2}\right)\mathrm{d}y \\ &= \mathrm{e}^{\beta}\left[\mathrm{Ei}(-\beta(z+2)) - \mathrm{Ei}(-\beta(z+1))\right] \\ &\quad + \left(1 - \frac{1}{z+1}\right)\mathrm{e}^{-\beta z} - \left(1 - \frac{1}{z+2}\right)\mathrm{e}^{-\beta(z+1)}. \end{aligned} \tag{4}$$

Regarding (1) and (2) and first partial moment (4), the PMF and CDF of BDBH distribution are represented as

$$\begin{aligned} f_Z(z,\beta) &= (z-1)\left(1 - \frac{\mathrm{e}^{-\beta(z-1)}}{z}\right) - 2z\left(1 - \frac{\mathrm{e}^{-\beta z}}{z+1}\right) + (z+1)\left(1 - \frac{\mathrm{e}^{-\beta(z+1)}}{z+2}\right) \\ &\quad + \mathrm{e}^{\beta}\left[2\mathrm{Ei}(-\beta(z+1)) - \mathrm{Ei}(-\beta z) - \mathrm{Ei}(-\beta(z+2))\right] + \left(1 - \frac{1}{z}\right)\mathrm{e}^{-\beta(z-1)} \end{aligned}$$

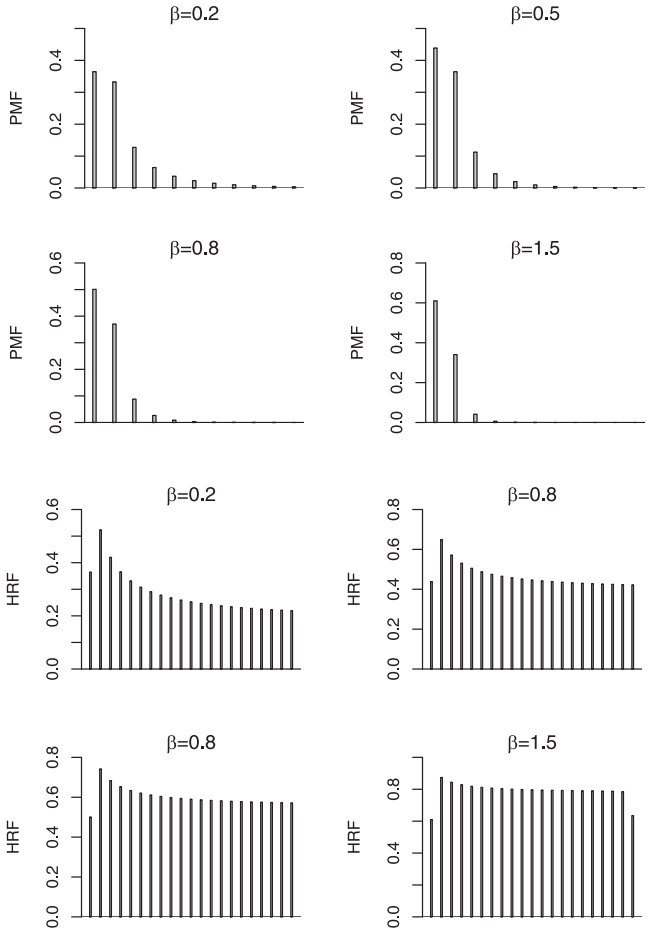

Figure 1. The PMF and HRF plots of BDBH distribution for different values of β.

$$+\left(1-\frac{1}{z+2}\right)\mathrm{e}^{-\beta(z+1)}-2\left(1-\frac{1}{z+1}\right)\mathrm{e}^{-\beta z}$$
$$=\begin{cases}1-\mathrm{e}^{\beta}\left(\mathrm{Ei}(-2\beta)-\mathrm{Ei}(-\beta)\right), & z=0\\ -\mathrm{e}^{\beta}\left(\mathrm{Ei}(-\beta z)-2\mathrm{Ei}(-\beta(z+1))+\mathrm{Ei}(-\beta(z+2))\right), & z=1,2,\ldots\end{cases}, \tag{5}$$

$$F_Z(z,\beta)=1-\mathrm{e}^{\beta}\left(\mathrm{Ei}(-\beta(z+2))-\mathrm{Ei}(-\beta(z+1))\right), \tag{6}$$

where, since $y \geq 0$, so $G_Y(-1,\beta)=0$ and $E_Y(1,-1\,|\,\beta)=0$.

The PMF and HRF plots of BDBH distribution are respectively depicted in Figure 1, for different values of β. Figure 1 shows that the PMF of BDBH is a decreasing function with respect to Z, where the HRF is a unimodal (inverted bath-tub shaped) function.

Corollary 2.3: *The expectation and variance of the BDBH distribution are represented as*

$$\mu_Z=E(Z)=-\mathrm{e}^{\beta}\mathrm{Ei}(-\beta), \tag{7}$$

Table 1. Some statistical index of the BDBH distribution for different values of the parameter β.

Measures	$\beta = 0.2$	$\beta = 0.5$	$\beta = 0.8$	$\beta = 1$	$\beta = 1.5$	$\beta = 2$
Mean	1.49335	0.922911	0.691245	0.596347	0.448257	0.361329
Variance	4.94434	1.461721	0.796825	0.607188	0.386821	0.292569
FDI	3.31091	1.583817	1.152738	1.018178	0.862945	0.809704

$$\sigma_Z^2 = Var(Z) = \frac{2}{\beta} + \mathrm{e}^{\beta}\,\mathrm{Ei}(-\beta)\left(2 - \mathrm{e}^{\beta}\,\mathrm{Ei}(-\beta)\right) + \zeta_0(\beta). \tag{8}$$

***Proof*:** Based on Corollary 2.2, the expectation of the BDBH distribution is obvious. The second partial moment of BDBH is computed as

$$E_Y(2, z \mid \beta) = \mathrm{e}^{-\beta z}\left(\frac{2}{\beta} + \frac{z^2}{z+1}\right) - \mathrm{e}^{-\beta(z+1)}\left(\frac{2}{\beta} + \frac{(z+1)^2}{z+2}\right) - 2\,\mathrm{e}^{\beta}\left[\mathrm{Ei}(-\beta(z+2)) - \mathrm{Ei}(-\beta(z+1))\right].$$

Subsequently, by elementary computation, it is concluded that

$$\zeta_0(i, \beta) = (2i+3)\,\mathrm{e}^{\beta}\left[\mathrm{Ei}(-\beta(i+2)) - \mathrm{Ei}(-\beta(i+1))\right] - \frac{2}{\beta}\left(\mathrm{e}^{-\beta i} - \mathrm{e}^{-\beta(i+1)}\right).$$

Hence, the variance (8) is induced by substituting the second partial moments and $\zeta_0(i, \beta)$ in Corollary 2.2. ■

The fisher dispersion index (FDI) is defined as variance to mean ratio, indicating whether a certain distribution is suitable for under(over)-dispersed or equi-dispersed data sets. If FDI$<$ 1, the distribution is under-dispersed, FDI$>$ 1 is over-dispersed and if FDI$=$ 1 is equi-dispersed. The mean, variance and FDI measures are provided in Table 1 for different values of the parameter β.

It can be seen in Table 1 that by increasing the values of parameter β, the mean, variance and FDI of BDBH distribution are decreased. For $\beta < 1$, the FDI measure is greater than 1, which indicates the over-dispersion of BDBH distribution, for $\beta > 1$, the FDI measure is less than 1, which indicates the under-dispersion of BDBH distribution and β near to one leads to equi-dispersion distribution. Different values of FDI measure represent the applicability of BDBH distribution for all kinds of dispersion including over-, under- and equi-dispersions.

3. The BDBH-MINAR(1) process characterization

In this section, we introduce a new mixing INAR(1) model based on BDBH innovations. The mixing Pegram and binomial thinning operator was considered by Khoo *et al.* [17].

Definition 3.1 (Binomial thinning operator): Consider a sequence of independent Bernoulli random variables $\{T_i\}$ with the parameter α, the binomial thinning operator is

generated by counting series $\{T_i\}$ as below

$$\alpha \circ X = \sum_{i=1}^{X} T_i, \quad 0 < \alpha < 1,$$

where X is a non-negative integer-valued random variable and independent of $\{T_i\}$.

Definition 3.2 (Pegram operator): The Pegram operator mixes U and V with the weights ϕ and $1 - \phi$ as

$$W = (\phi, U) * (1 - \phi, V), \quad 0 < \phi < 1,$$

where U and V are two independent discrete random variables. The corresponding marginal probability function of the random variable W is provided as

$$P(W = j) = \phi P(U = j) + (1 - \phi)P(V = j).$$

Definition 3.3 (BDBH-MINAR(1) process): We introduce a new stationary mixing INAR(1) process $\{X_t\}$ by the following recursive equation:

$$X_t = (\phi, \alpha \circ X_{t-1} + Z_t) * (1 - \phi, Z_t), \quad \alpha, \phi < 1,\ t \geq 1, \tag{9}$$

where '$(\circ, *)$' stands the binomial thinning and Pegram operators, $\{Z_t\}$ be a sequence of BDBH random variables with the parameter β and given X_{t-1}, the random variables $\alpha \circ X_{t-1}$ and Z_t are independent of each other. We shall abbreviate the mixing model name as BDBH-MINAR(1).

The applicability of the BDBH-MINAR(1) process in modeling real phenomena can be expressed based on the epidemic of contagious diseases such as COVID-19, as follows. Consider the sequence X_t as the counts of infected patients during the time interval $(t - 1, t]$, which can be inferred in two scenarios. In the first scenario, due to the pandemic, with probability $1 - \phi$, every infected person is quarantined and cannot transmit the virus among other people and new cases are only the immigrants from abroad (Z_t). In the other scenario, with probability ϕ, the quarantine does not hold perfectly and in addition to immigrant patients, the indigenous patients at time $t-1$ denoted as X_{t-1} are transmitted the COVID-19 with probability α, causing new cases.

Proposition 3.4: *The BDBH-MINAR*(1) *process defined* (9) *can also be considered as a random coefficient model as below*

$$X_t = \alpha_t \circ X_{t-1} + Z_t, \quad t \geq 1,$$

where $\alpha_t = \begin{cases} \alpha & \text{w.p. } \phi \\ 0 & \text{w.p. } 1-\phi \end{cases}$, $\quad 0 < \phi < 1.$

Based on (9), the one-step transition probabilities are

$$P_{ij} = P\left(X_t = j \,|\, X_{t-1} = i\right) = \phi \sum_{k=0}^{i} \binom{i}{k} \alpha^k (1 - \alpha)^{i-k} f_Z(j - k, \beta) + (1 - \phi) f_Z(j, \beta). \tag{10}$$

where $f_Z(., \beta)$ is defined in (5).

Lemma 3.5: *The process* $\{X_t\}$ *is an irreducible, aperiodic and positive recurrent Markov chain, hence* $\{X_t\}$ *is ergodic and stationary.*

***Proof*:** The proof is routine and is omitted. ■

Proposition 3.6: *Suppose* $\{X_t\}$ *is a stationary process defined by* (9)*, then for* $0 < \alpha, \phi < 1$ *and* $t \geq 1$,

(i) *The conditional expectation is*

$$E\left(X_t \mid X_{t-k}\right) = (\alpha\phi)^k X_{t-k} + \frac{1-(\alpha\phi)^k}{1-\alpha\phi}\mu_Z. \tag{11}$$

When $k \to \infty$, *implies* $\lim_{k\to\infty} E(X_t \mid X_{t-k}) = \dfrac{\mu_Z}{1-\alpha\phi}$, *which is the unconditional expectation of the process.*

(ii) *The conditional variance is*

$$Var(X_t \mid X_{t-1}) = \alpha^2\phi(1-\phi)X_{t-1}^2 + \alpha\phi(1-\alpha)X_{t-1} + \sigma_Z^2, \tag{12}$$

and

$$\begin{aligned} Var\left(X_t \mid X_{t-k}\right) = {} & \alpha^{2k}\phi^k(1-\phi^k)X_{t-k}^2 + 2(\alpha\phi)^k\mu_Z X_{t-k}\left(\sum_{i=0}^{k-1}\alpha^i\left(1-\phi^i\right)\right) \\ & + (\alpha\phi)^k(1-\alpha)X_{t-k}\sum_{i=0}^{k-1}\alpha^i + E(Z_t^2)\sum_{i=0}^{k-1}\left(\alpha^2\phi\right)^i \\ & + \alpha\phi(1-\alpha)\mu_Z\sum_{i=0}^{k-1}(\alpha\phi)^i\sum_{j=0}^{i}\alpha^j \\ & + \mu_Z^2\left(2\alpha\phi\sum_{i=0}^{k-1}(\alpha\phi)^i\sum_{j=0}^{i}\alpha^j - \left(\sum_{i=0}^{k-1}(\alpha\phi)^i\right)^2\right), \end{aligned}$$

where σ_Z^2 *is the variance of* Z_t *and was computed in Corollary 2.2,*

$$\begin{aligned} \lim_{k\to\infty} Var\left(X_t \mid X_{t-k}\right) = {} & \frac{E(Z_t^2)}{1-\alpha^2\phi} + \frac{\alpha\phi(1-\alpha)\mu_Z}{(1-\alpha^2\phi)(1-\alpha\phi)} \\ & + \frac{\mu_Z^2\left(2\alpha\phi(1-\alpha\phi)+\alpha^2\phi-1\right)}{(1-\alpha^2\phi)(1-\alpha\phi)^2}, \end{aligned}$$

which is the unconditional variance.

(iii) *The autocorrelation function of the process* $\{X_t\}$ *is represented as*

$$\rho(k) = Corr(X_t, X_{t-k}) = (\alpha\phi)^k.$$

Proof: (i) The conditional expectation of the process is obtained as follows:

$$E(X_t \mid X_{t-1}) = \alpha\phi X_{t-1} + \mu_Z,$$
$$E(X_t \mid X_{t-2}) = E(E(X_t \mid X_{t-1}) \mid X_{t-2}) = (\alpha\phi)^2 X_{t-2} + (1+\alpha\phi)\mu_Z.$$

By induction, we can conclude that

$$E(X_t \mid X_{t-k}) = (\alpha\phi)^k X_{t-k} + \frac{(1-(\alpha\phi)^k)\mu_Z}{1-(\alpha\phi)},$$

which is a linear function in X_t.

(ii) The second-order conditional expectation of the BDBH-MINAR(1) process is computed as

$$E(X_t^2 \mid X_{t-1}) = \alpha^2\phi X_{t-1}^2 + \alpha\phi(1-\alpha)X_{t-1} + 2\alpha\phi\mu_Z X_{t-1} + E(Z_t^2),$$

and

$$\begin{aligned} E(X_t^2 \mid X_{t-2}) = {} & \alpha^4\phi^2 X_{t-2}^2 + \alpha^2\phi^2(1-\alpha^2)X_{t-2} + 2\alpha^2\phi^2(1+\alpha)\mu_Z X_{t-2} \\ & + (1+\alpha^2\phi)E(Z_t^2) + \alpha\phi(1-\alpha)\mu_Z + 2\alpha\phi\mu_Z^2, \end{aligned}$$

subsequently,

$$\begin{aligned} E(X_t^2 \mid X_{t-3}) = {} & \alpha^6\phi^3 X_{t-3}^2 + \alpha^3\phi^3(1-\alpha)(1+\alpha+\alpha^2)X_{t-3} \\ & + 2\alpha^3\phi^3(1+\alpha+\alpha^2)\mu_Z X_{t-3} \\ & + (1+\alpha^2\phi+\alpha^4\phi^2)E(Z_t^2) + \alpha\phi(1-\alpha)\left(1+\alpha\phi+\alpha^2\phi\right)\mu_Z \\ & + 2\alpha\phi\left(1+\alpha\phi+\alpha^2\phi\right)\mu_Z^2. \end{aligned}$$

By induction, we can conclude that

$$\begin{aligned} E(X_t \mid X_{t-k}) = {} & \alpha^{2k}\phi^k X_{t-k}^2 + 2(\alpha\phi)^k\mu_Z X_{t-k}\sum_{i=0}^{k-1}\alpha^i + (\alpha\phi)^k(1-\alpha)X_{t-k}\sum_{i=0}^{k-1}\alpha^i \\ & + E(Z_t^2)\sum_{i=0}^{k-1}\left(\alpha^2\phi\right)^i + \alpha\phi(1-\alpha)\mu_Z\sum_{i=0}^{k-1}(\alpha\phi)^i\sum_{j=0}^{i}\alpha^j \\ & + 2\alpha\phi\mu_Z^2\sum_{i=0}^{k-1}(\alpha\phi)^i\sum_{j=0}^{i}\alpha^j. \end{aligned}$$

By the variance definition and some elementary calculations, the proof is completed.

(iii) The proof is routine and eliminated.

■

4. The estimation of the parameters

This section is dedicated to several estimation approaches, including the conditional maximum likelihood, modified conditional least square and Yule–Walker estimation methods of the BDBH-MINAR(1) model.

4.1. Conditional maximum likelihood estimation

The conditional maximum likelihood (CML) estimates of the parameters of the BDBH-MINAR(1) model are obtained through log-likelihood maximization, with respect to parameters $\delta = (\alpha, \phi, \beta)$. The log-likelihood function can be written as

$$\ell(\delta) = \log L(X_2, \ldots, X_n \mid \delta) = \sum_{t=2}^{n} \log P\left(X_t = j \mid X_{t-1} = i\right),$$

where $P(X_t = j \mid X_{t-1} = i)$ is the transition probability given by (10). The CML estimate of the unknown parameters are numerically obtained by maximizing the log-likelihood function using statistical package 'R'. We consider the modified conditional least squares estimates as initial values of the parameters in optimization commands.

4.2. Bayesian estimation

In the Bayesian methodology, we consider that both X_t and parameters are random. We consider the Gamma priors for parameter α, ϕ and Beta prior for parameter β as below

$$\alpha \sim Gamma(a_1, b_1), \quad \phi \sim Gamma(a_2, b_2), \quad \beta \sim Beta(a_3, b_3).$$

Also assume that α, ϕ and β are independent and $a_1, a_2, a_3, b_1, b_2, b_3 > 0$. The posterior distribution of the parameters (α, ϕ, β) can be written as

$$\begin{aligned} P(\alpha, \phi, \beta \mid X_t) &\propto L(X_t, \alpha, \phi, \beta \mid X_1)\, P(\alpha, \phi, \beta) \\ &\propto L(X_t, \alpha, \phi, \beta \mid X_1)\, \alpha^{a_1 - 1} (1-\alpha)^{b_1 - 1} \phi^{a_2 - 1} (1-\phi)^{b_2 - 1} \beta^{a_3 - 1} e^{-b_3 \beta}, \end{aligned} \tag{13}$$

where $L(X_t, \alpha, \phi, \beta \mid X_1)$ is the conditional likelihood function.

Based on the adaptive rejection Metropolis sampling (ARMS) within Gibbs methodology, we generate samples $((\alpha_1, \phi_1, \beta_1), \ldots, (\alpha_m, \phi_m, \beta_m))$ by the full conditional distribution (13), which can be performed by 'arms(.)' command in 'dlm' package in 'R'. Then the Bayesian estimates are obtained as $\hat{\alpha}_B = \frac{1}{n} \sum_{i=1}^{m} \alpha_i$, $\hat{\phi}_B = \frac{1}{n} \sum_{i=1}^{m} \phi_i$ and $\hat{\beta}_B = \frac{1}{n} \sum_{i=1}^{m} \beta_i$.

4.3. Modified conditional least square estimation

The modified conditional least squares (MCLS) estimators of the parameters $\delta_1 = (\rho, \mu_Z)$ are obtained by minimizing the following expression:

$$Q(\delta_1) = \sum_{t=2}^{n} (X_t - E(X_t \mid X_{t-1}))^2 = \sum_{t=2}^{n} \left(X_t - \rho X_{t-1} - \mu_Z\right)^2, \tag{14}$$

where we set $\rho = \alpha\phi$.

The MCLS estimators are given by

$$\hat{\rho}_{MCLS} = \frac{(n-1)\sum_{t=2}^{n} X_t X_{t-1} - \sum_{t=2}^{n} X_t \sum_{t=2}^{n} X_{t-1}}{(n-1)\sum_{t=2}^{n} X_{t-1}^2 - \left(\sum_{t=2}^{n} X_{t-1}\right)^2},$$

$$\hat{\mu}_{Z,MCLS} = \frac{\sum_{t=2}^{n} X_t - \hat{\rho}_{MCLS}\sum_{t=2}^{n} X_{t-1}}{n-1}.$$

The MCLS estimates of the parameter β are computed by the root of Equation (7), by putting $\hat{\mu}_{Z,MCLS}$. We apply the 'uniroot' command with lower and upper values equal to 0 and $\beta + 5$, respectively.

The estimation of the parameters α and ϕ is computed under the modified method proposed by Karlsen and Tjøstheim [16] by minimizing the following expression:

$$S(\alpha, \phi) = \sum_{t=2}^{n} \left(V_t - Var(X_t \mid X_{t-1})\right)^2, \tag{15}$$

where $V_t = (X_t - E(X_t \mid X_{t-1}))^2 = (X_t - \hat{\rho}_{MCLS}X_{t-1} - \hat{\mu}_{Z,MCLS})^2$ and $Var(X_t \mid X_{t-1})$ is defined in (12) with estimated values of the parameters (ρ, μ_Z). The parameter ϕ can be written with respect to α as $\phi = \frac{\rho}{\alpha}$, so we have

$$Var(X_t \mid X_{t-1}) = \hat{\rho}_{MCLS}(\alpha - \hat{\rho}_{MCLS})X_{t-1}^2 + \hat{\rho}_{MCLS}(1-\alpha)X_{t-1} + \hat{\sigma}_{Z,MCLS}^2,$$

where σ_Z^2 is a function of β, and the corresponding estimator $\hat{\sigma}_{Z,MCLS}^2$ is obtained based on $\hat{\beta}_{MCLS}$. Hence the MCLS estimation of the parameter α is achieved as below

$$\hat{\alpha}_{MCLS} = \frac{\sum_{t=2}^{n}(V_t + \hat{\rho}_{MCLS}^2 X_{t-1}^2 - \hat{\rho}_{MCLS}X_{t-1} - \hat{\sigma}_{Z,MCLS}^2)(X_{t-1}^2 - X_{t-1})}{\hat{\rho}_{MCLS}\sum_{t=2}^{n}(X_{t-1}^2 - X_{t-1})^2}.$$

Subsequently, the MCLS estimation of the parameter ϕ is concluded as $\hat{\phi}_{MCLS} = \frac{\hat{\rho}_{MCLS}}{\hat{\alpha}_{MCLS}}$.

4.4. Yule–Walker estimation

Consider the expectation $E(X_t) = \frac{\mu_Z}{1-\rho}$ and autocorrelation $Corr(X_t, X_{t-1}) = \rho$ of the process, so by using the sample mean and sample autocorrelation function, the YW estimates of the parameters (ρ, μ_Z) are obtained as

$$\hat{\rho}_{YW} = \frac{\sum_{t=2}^{n}(X_t - \overline{X})(X_{t-1} - \overline{X})}{\sum_{t=1}^{n}(X_t - \overline{X})^2},$$

$$\hat{\mu}_{Z,YW} = \overline{X}(1 - \hat{\rho}_{YW}).$$

Hence, the YW estimates of the parameter β are computed based on the YW estimation of μ_Z. To estimate parameters α and ϕ, we use the second expectation of the process as below

$$E(X_t^2) = \frac{\alpha\phi(1-\alpha)E(X_t) + 2\alpha\phi(1-\alpha\phi)E^2(X_t) + E(Z_t^2)}{1-\alpha^2\phi},$$

Let $\overline{X^2} = \frac{1}{n}\sum_{t=1}^{n} X_t^2$ and $\phi = \frac{\rho}{\alpha}$, then

$$\hat{\alpha}_{YW} = \frac{\overline{X^2} - \hat{\rho}_{YW}\overline{X} - 2\hat{\rho}_{YW}(1-\hat{\rho}_{YW})\overline{X}^2 - E(Z_t^2)}{\hat{\rho}_{YW}\left(\overline{X^2} - \overline{X}\right)}, \tag{16}$$

where, based on (7) and (8), we set

$$E(Z_t^2) = \frac{2}{\hat{\beta}_{YW}} + e^{\hat{\beta}_{YW}}\,\mathrm{Ei}(-\hat{\beta}_{YW})\left(2 - e^{\hat{\beta}_{YW}}\,\mathrm{Ei}(-\hat{\beta}_{YW})\right) + \zeta_0(\hat{\beta}_{YW}) + \hat{\mu}_{Z,YW}^2.$$

The YW estimation of the parameter ϕ is obtained as $\hat{\phi}_{YW} = \frac{\hat{\rho}_{YW}}{\hat{\alpha}_{YW}}$.

5. Simulation approach

In this section, the data generating process and efficiency comparison of estimations based on the simulation scheme of the BDBH-MINAR(1) model are discussed.

5.1. Data generating process of the BDBH-MINAR(1) model

Here, we provide the algorithm of data generating process of the BDBH-MINAR(1) model step by step:

Consider $X_0 = 0$ and for $t = 1, \ldots, n$,

(1) Generate a random sample from BDBH distribution with parameter β as below
   (a) Generate the random sample u_1 from Uniform $(0, 1)$ distribution
   (b) Based on (3), compute $y = \mathrm{QF}_{\mathrm{Y}}(u_1, \beta)$
   (c) Set $v = [y]$
   (d) Based on values y and v, generate the random sample u from Bernoulli distribution with parameter $y-v$
   (e) Set $Z_t = v + u$
(2) Generate the random sample $\alpha \circ X_{t-1}$ from Binomial distribution with parameters (X_{t-1}, α)
(3) Generate the random sample u_2 from Uniform $(0, 1)$
(4) If $u_2 < \phi$, set $X_t = \alpha \circ X_{t-1} + Z_t$, otherwise set $X_t = Z_t$.

Based on the Monte Carlo simulation, we compare the efficiency of estimates of the BDBH-MINAR(1) model, under different sample sizes $n = 100, 200, 500, 1000$, over $h = 1000$ iterations. Four different combinations of the parameters are considered as $(\alpha, \phi, \beta) = (0.3, 0.7, 0.2)$, $(0.7, 0.3, 0.4)$, $(0.3, 0.7, 1)$ and $(0.7, 0.3, 2)$.

Based on the four combinations of the parameters, the sample mean, variance and FDI, for $n = 500$ and $h = 100$, are represented in Table 2, where the sample mean, variance and FDI are computed based on the mean of each measure in $h = 100$ iterations.

Table 2. Sample measures of the BDBH-MINAR(1) process for different combinations (α, ϕ, β).

Measures	(0.3, 0.7, 0.2)	(0.7, 0.3, 0.4)	(0.3, 0.7, 1)	(0.7, 0.3, 2)
Mean	1.47	1.19	0.64	0.33
Variance	5.64	2.36	0.67	0.29
FDI	3.83	1.98	1.04	0.87

To evaluate the performance of the estimators, we utilize the root mean squared error (RMSE) measure. The results summarized in Table 3 represent the convergence of all estimates of the parameters to their actual values. Further, by increasing the sample size, the RMSE is gradually decreased. Among different estimation methods, the Bayes estimates have the preference over the CML, MCLS and YW estimation, since they have small RMSE for all parameters.

6. Real-life data application

In this section, we investigate the application of the BDBH-MINAR(1) process by using four daily COVID-19 death counts data from Robert Koch Institute: SurvStat@RKI 2.0 (https://survstat.rki.de) site, listed as:

- The first data set represents $n = 80$ daily counts of COVID-19 death, reported from Austria, from 7th June until 25th August in 2021.
- The second data set represents $n = 65$ daily counts of COVID-19 death, reported from Switzerland, from 4th June until 7th August in 2021.
- The third data set represents $n = 87$ daily counts of COVID-19 death, reported from Nigeria, from 27th October in 2021 until 21th January in 2022.
- The fourth data set represents $n = 70$ daily counts of COVID-19 death, reported from Slovenia, from 14th June until 22th August in 2021.

The sample path, autocorrelation function (ACF) and partial autocorrelation function (PACF) of four data series are displayed in Figure 2. The PACF plots suggest a first-order autoregressive model for all data sets.

Some statistical properties of actual data sets are reported in Table 4. Due to the results of Table 4, four clinical data series are empirically over-dispersed, and the stationarity of data is justified by using the Augmented Dickey-Fuller (ADF) test. The BDBH-MINAR(1) process covers all types of dispersion, including overdispersion, hence a convenient choice for the COVID-19 data sets.

We compare the BDBH-MINAR(1) model to some competitive INAR(1) models such as GINAR(1) [6], NGINAR(1) [22], NBRCINAR(1) [31], NBIINAR(1) [4], P-MPT(1) [17], MPDINAR(1) [28] and MPTSD(1) [26]. Also, we define an INAR(1) process by considering the discrete BH distribution [11] for innovations and binomial thinning operator, called DBH-INAR(1), and compare the DBH-INAR(1) process with the proposed BDBH-MINAR(1) model.

We reported the CML estimates, the goodness-of-fit (GOF) statistics as AIC, BIC, HQIC and CAIC for each relevant INAR(1) model. The results of the COVID-19 data series are represented in Tables 5–8. Regarding Tables 5–8, the values of the GOF and RMS are the

Table 3. Simulation results of estimates of parameters of the BDBH-MINAR(1) with RMSE in brackets.

	Bayes			CML			MCLS			YW		
n	$\hat{\alpha}$	$\hat{\phi}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\phi}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\phi}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\phi}$	$\hat{\beta}$
						$(\alpha, \phi, \beta) = (0.3, 0.7, 0.2)$						
100	0.29116	0.6859	0.19144	0.32667	0.64854	0.21609	0.36098	0.65146	0.23273	0.38296	0.79511	0.24481
RMSE	(0.00737)	(0.05818)	(0.00312)	(0.02334)	(0.06084)	(0.00613)	(0.13212)	(0.22249)	(0.06714)	(0.19541)	(0.46111)	(0.07249)
200	0.30025	0.68913	0.19263	0.31791	0.67864	0.20957	0.34464	0.66847	0.22056	0.36721	0.75583	0.22405
RMSE	(0.00614)	(0.03174)	(0.00098)	(0.01066)	(0.04622)	(0.00532)	(0.11669)	(0.20028)	(0.05169)	(0.18036)	(0.44675)	(0.05252)
500	0.29585	0.69327	0.19441	0.30851	0.69055	0.19741	0.33121	0.68167	0.21216	0.33651	0.74066	0.22021
RMSE	(0.00429)	(0.01048)	(0.00076)	(0.00386)	(0.02581)	(0.00229)	(0.09729)	(0.18185)	(0.03283)	(0.12517)	(0.32927)	(0.03901)
1000	0.29908	0.69653	0.19884	0.30056	0.69996	0.19822	0.309162	0.69764	0.20583	0.31187	0.718397	0.20735
RMSE	(0.00156)	(0.00631)	(0.00052)	(0.00168)	(0.00743)	(0.00081)	(0.05892)	(0.11222)	(0.00815)	(0.07126)	(0.15086)	(0.00931)
						$(\alpha, \phi, \beta) = (0.7, 0.3, 0.4)$						
100	0.68492	0.29002	0.38923	0.73094	0.28978	0.42089	0.66736	0.36581	0.43648	0.79651	0.24781	0.46141
RMSE	(0.02619)	(0.00522)	(0.00493)	(0.04691)	(0.00831)	(0.00976)	(0.26843)	(0.15096)	(0.08705)	(0.35021)	(0.13567)	(0.10931)
200	0.71152	0.29218	0.40849	0.71894	0.29086	0.40431	0.67219	0.33481	0.43039	0.77942	0.25937	0.44783
RMSE	(0.00852)	(0.00346)	(0.00276)	(0.02121)	(0.00711)	(0.00607)	(0.22986)	(0.12431)	(0.06923)	(0.30477)	(0.12642)	(0.08629)
500	0.69203	0.29436	0.40541	0.70584	0.29826	0.40134	0.68973	0.31055	0.42726	0.73191	0.28173	0.43324
RMSE	(0.00517)	(0.00151)	(0.00081)	(0.00773)	(0.00261)	(0.00142)	(0.14185)	(0.08794)	(0.04763)	(0.21661)	(0.09397)	(0.04903)
1000	0.69558	0.29748	0.40017	0.70013	0.29959	0.40007	0.69901	0.30634	0.39678	0.70605	0.29122	0.41004
RMSE	(0.00239)	(0.00067)	(0.00042)	(0.00338)	(0.00129)	(0.00061)	(0.09212)	(0.05113)	(0.01883)	(0.16761)	(0.06975)	(0.02026)
						$(\alpha, \phi, \beta) = (0.3, 0.7, 1)$						
100	0.29194	0.71471	1.03683	0.28387	0.67384	0.96874	0.34583	0.65209	0.91105	0.26181	0.64391	0.92589
RMSE	(0.00952)	(0.04871)	(0.08493)	(0.02998)	(0.08856)	(0.12538)	(0.12887)	(0.29011)	(0.36081)	(0.10274)	(0.36871)	(0.35536)
200	0.29232	0.70706	1.03183	0.28633	0.67962	0.97008	0.27363	0.66553	0.94271	0.27769	0.65936	0.92721
RMSE	(0.00795)	(0.03656)	(0.07508)	(0.02901)	(0.07613)	(0.09974)	(0.10116)	(0.26984)	(0.34394)	(0.09063)	(0.31653)	(0.32387)
500	0.29486	0.69206	0.98784	0.29241	0.69124	0.98103	0.28288	0.68924	0.96118	0.28633	0.71912	0.95642
RMSE	(0.004596)	(0.00811)	(0.03554)	(0.01045)	(0.02193)	(0.06709)	(0.06179)	(0.17272)	(0.22224)	(0.07025)	(0.20682)	(0.25086)
1000	0.29744	0.69661	0.99232	0.29837	0.69374	0.98742	0.29274	0.69008	1.01392	0.28942	0.70699	0.97827
RMSE	(0.00299)	(0.00535)	(0.00852)	(0.00773)	(0.00972)	(0.02573)	(0.02206)	(0.08855)	(0.09292)	(0.02148)	(0.13907)	(0.10232)
						$(\alpha, \phi, \beta) = (0.7, 0.3, 2)$						
100	0.71231	0.29206	2.03728	0.66306	0.28743	1.94202	0.6332	0.31303	1.87444	0.76741	0.27821	1.86421
RMSE	(0.04445)	(0.00607)	(0.15738)	(0.05412)	(0.01334)	(0.23959)	(0.28759)	(0.14279)	(0.43284)	(0.25271)	(0.15035)	(0.44768)
200	0.68875	0.29497	2.02951	0.67781	0.28959	1.95782	0.65152	0.31017	2.09414	0.65994	0.28304	1.88629
RMSE	(0.04146)	(0.00449)	(0.12361)	(0.05048)	(0.01046)	(0.21807)	(0.26348)	(0.13705)	(0.41102)	(0.24006)	(0.13946)	(0.43741)
500	0.69039	0.29527	1.98771	0.68503	0.29109	1.96185	0.67026	0.28593	1.92196	0.73152	0.28429	1.90765
RMSE	(0.01324)	(0.00084)	(0.07029)	(0.02001)	(0.00763)	(0.14995)	(0.18066)	(0.09832)	(0.28616)	(0.17594)	(0.10697)	(0.31922)
1000	0.69455	0.30106	1.99197	0.69196	0.29477	1.97496	0.7198	0.28843	1.95811	0.68702	0.29091	1.94292
RMSE	(0.00878)	(0.00062)	(0.02091)	(0.01443)	(0.00455)	(0.08907)	(0.09027)	(0.06054)	(0.11159)	(0.08686)	(0.04031)	(0.14425)

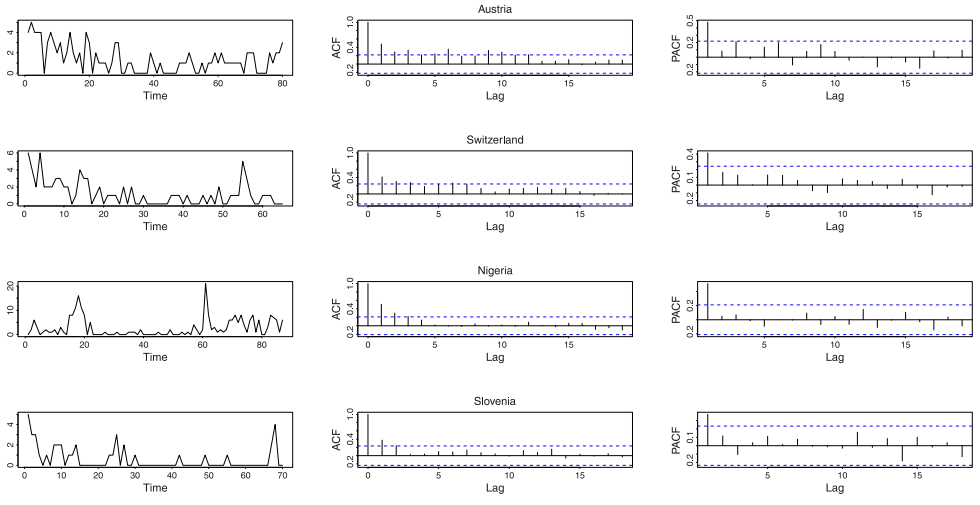

Figure 2. The sample path, ACF and PACF of the COVID-19 daily deaths sets.

Table 4. Statistical measures and *p*-values of Augmented Dickey–Fuller test for clinical data sets.

Data sets	Mean	Variance	Autocorrelation	FDI	*p*-value of ADF
Austria	1.362	1.727	0.477	1.268	0.043
Switzerland	1.215	2.14	0.415	1.761	0.015
Nigeria	2.747	14.77	0.511	5.377	0.039
Slovenia	0.571	1.147	0.375	2.007	0.042

Table 5. The CML estimates and some GOF measures for the COVID-19 daily deaths in Austria.

Model	CML	AIC	BIC	HQIC	CAIC
GINAR(1)	$\hat{p} = 0.53246,\ \hat{\alpha} = 0.42945$	246.64	251.41	248.55	248.96
DBH-INAR(1)	$\hat{\alpha} = 0.66037,\ \hat{\beta} = 0.53994$	255.34	260.11	257.25	255.55
NGINAR(1)	$\hat{p} = 6.44531,\ \hat{\alpha} = 1.45502$	241.37	246.14	243.28	243.69
NBRCINAR(1)	$\hat{n} = 4.25648,\ \hat{p} = 0.75411,\ \hat{\rho} = 0.48918$	235.34	242.49	239.21	237.87
NBIINAR(1)	$\hat{n} = 6.5813,\ \hat{p} = 8.98974,\ \hat{\rho} = 0.48112$	235.22	242.37	238.09	237.76
P-MPT(1)	$\hat{\alpha} = 0.77688,\ \hat{\phi} = 0.33379,\ \hat{\lambda} = 1.30933$	236.65	243.8	239.52	239.19
MPDINAR(1)	$\hat{\alpha} = 0.77003,\ \hat{\theta} = 4.40936e-06,\ \hat{\phi} = 0.34108,\ \hat{\mu} = 1.23296$	243.01	252.53	246.83	245.82
MPTSD(1)	$\hat{\phi} = 0.26994,\ \hat{\rho} = 0.73063, \hat{a} = 0.27367,\ \hat{b} = 1.13981$	238.35	247.88	242.17	238.89
BDBH-MINAR(1)	$\hat{\alpha} = 0.68834,\ \hat{\phi} = 0.79484,\ \hat{\beta} = 0.28659$	232.48	239.63	235.34	232.88

smallest for the BDBH-MINAR(1) model. Therefore, we can conclude that the BDBH-MINAR(1) model provides the best fitting among other competitive INAR(1) models.

6.1. The residual analysis of the clinical data sets

The adequacy of the BDBH-MINAR(1) process is evaluated by the residual analysis of all COVID-19 data sets. The Pearson residuals are defined as

$$e_t = \frac{X_t - E(X_t \mid X_{t-1})}{\sqrt{Var(X_t \mid X_{t-1})}},$$

Table 6. The CML estimates and some GOF measures for the COVID-19 daily deaths in Switzerland.

Model	CML	AIC	BIC	HQIC	CAIC
GINAR(1)	$\hat{p} = 0.52272,\ \hat{\alpha} = 0.33497$	194.65	199.01	196.37	197.05
DBH-INAR(1)	$\hat{\alpha} = 0.55696,\ \hat{\beta} = 0.55025$	200.76	205.11	202.48	201.02
NGINAR(1)	$\hat{p} = 1.31654,\ \hat{\alpha} = 0.61823$	189.91	194.25	191.62	192.31
NBRCINAR(1)	$\hat{n} = 1.86419,\ \hat{p} = 0.60321,\ \hat{\rho} = 0.38624$	192.35	198.88	194.93	195.02
NBIINAR(1)	$\hat{n} = 1.67351,\ \hat{p} = 2.39232,\ \hat{\rho} = 0.44401$	190.57	197.09	193.14	193.24
P-MPT(1)	$\hat{\alpha} = 0.70882,\ \hat{\phi} = 0.29357,\ \hat{\lambda} = 1.06696$	190.61	197.12	193.17	193.27
MPDINAR(1)	$\hat{\alpha} = 0.721022,\ \hat{\theta} = 0.05908, \hat{\phi} = 0.29221,\ \hat{\mu} = 1.01061$	187.85	196.55	191.28	190.87
MPTSD(1)	$\hat{\phi} = 0.36951,\ \hat{\rho} = 0.39702, \hat{a} = 0.34453,\ \hat{b} = 1.21351$	199.51	208.21	202.94	200.18
BDBH-MINAR(1)	$\hat{\alpha} = 0.52141,\ \hat{\phi} = 0.34537,\ \hat{\beta} = 0.27579$	183.61	190.13	186.18	184.11

Table 7. The CML estimates and some GOF measures for the COVID-19 daily deaths in Nigeria.

Model	CML	AIC	BIC	HQIC	CAIC
GINAR(1)	$\hat{p} = 0.73798,\ \hat{\alpha} = 0.25715$	365.04	369.97	367.02	367.33
DBH-INAR(1)	$\hat{\alpha} = 0.57511,\ \hat{\beta} = 0.85056$	457.64	462.57	459.62	457.83
NGINAR(1)	$\hat{p} = 2.82786,\ \hat{\alpha} = 0.42193$	358.55	363.48	360.54	360.84
NBRCINAR(1)	$\hat{n} = 0.75577,\ \hat{p} = 0.22631,\ \hat{\rho} = 0.33339$	361.22	368.61	364.21	363.79
NBIINAR(1)	$\hat{n} = 0.55525\ \hat{p} = 0.45711,\ \hat{\rho} = 0.55898$	356.97	364.37	359.95	359.46
P-MPT(1)	$\hat{\alpha} = 0.77523,\ \hat{\phi} = 0.29277,\ \hat{\lambda} = 0.32109$	494.15	501.55	497.13	496.64
MPDINAR(1)	$\hat{\alpha} = 0.52906,\ \hat{\theta} = 0.37554, \hat{\phi} = 0.26799,\ \hat{\mu} = 2.77464$	373.33	383.19	377.31	376.07
MPTSD(1)	$\hat{\phi} = 0.49001,\ \hat{\rho} = 0.23875, \hat{a} = 0.77546,\ \hat{b} = 2.58746$	388.31	398.17	392.28	388.79
BDBH-MINAR(1)	$\hat{\alpha} = 0.78274,\ \hat{\phi} = 0.64403,\ \hat{\beta} = 0.09167$	351.77	359.16	354.74	352.13

Table 8. The CML estimates and some GOF measures for the COVID-19 daily deaths in Slovenia.

Model	CML	AIC	BIC	HQIC	CAIC
GINAR(1)	$\hat{p} = 0.37212,\ \hat{\alpha} = 0.27939$	141.39	145.89	143.18	143.76
DBH-INAR(1)	$\hat{\alpha} = 0.42501,\ \hat{\beta} = 0.38262$	140.28	144.78	142.07	140.52
NGINAR(1)	$\hat{p} = 0.67834,\ \hat{\alpha} = 0.65699$	137.03	141.53	138.82	139.44
NBRCINAR(1)	$\hat{n} = 0.51506,\ \hat{p} = 0.47325,\ \hat{\rho} = 0.32685$	138.53	145.28	141.22	141.15
NBIINAR(1)	$\hat{n} = 0.43219,\ \hat{p} = 1.31108,\ \hat{\rho} = 0.46706$	137.77	144.52	140.45	140.39
P-MPT(1)	$\hat{\alpha} = 0.75842,\ \hat{\phi} = 0.35239,\ \hat{\lambda} = 0.49878$	136.46	143.21	139.14	139.07
MPDINAR(1)	$\hat{\alpha} = 0.765068,\ \hat{\theta} = 0.10762, \hat{\phi} = 0.28718,\ \hat{\mu} = 0.477602$	132.77	141.77	136.34	135.71
MPTSD(1)	$\hat{\phi} = 0.58524,\ \hat{\rho} = 0.04171, \hat{a} = 0.51171,\ \hat{b} = 0.95009$	141.58	150.57	145.15	142.19
BDBH-MINAR(1)	$\hat{\alpha} = 0.72094,\ \hat{\phi} = 0.28677,\ \hat{\beta} = 0.50124$	128.74	135.49	131.42	129.21

where $E(X_t \mid X_{t-1})$ and $Var(X_t \mid X_{t-1})$ are defined in (11) and (12), respectively, and parameters are substituted by their estimated counterparts in BDBH-MINAR(1) model. Figure 3 depicts the sample ACF of the Pearson residuals of four data series. Based on Figure 3, the residuals are non-correlated, and the findings are confirmed by p-values $(0.242, 0.667, 0.831, 0.214)$ of the Ljung–Box test. The cumulative periodogram plots in Figure 4 indicate the randomly distributed residuals and do not represent any specified trend.

Figure 5 shows the result of the parametric resampling method. We generate 5000 data sets of length $n = (80, 65, 87, 70)$ using the fitted BDBH-MINAR(1) model (with CML estimates of the parameters of each data set). Based on these bootstrap data sets, 5000 autocorrelation functions and for each fixed lag, 100(0.975)% and 100(0.025)% quantiles of the ACF are obtained as the bounds of an acceptance region. These bounds are shown as '+' symbols in the figure with the sample ACF presented by '•'. Based on Figure 5, all the sample autocorrelations appointed between the acceptable bounds and adequacy of the model are concluded.

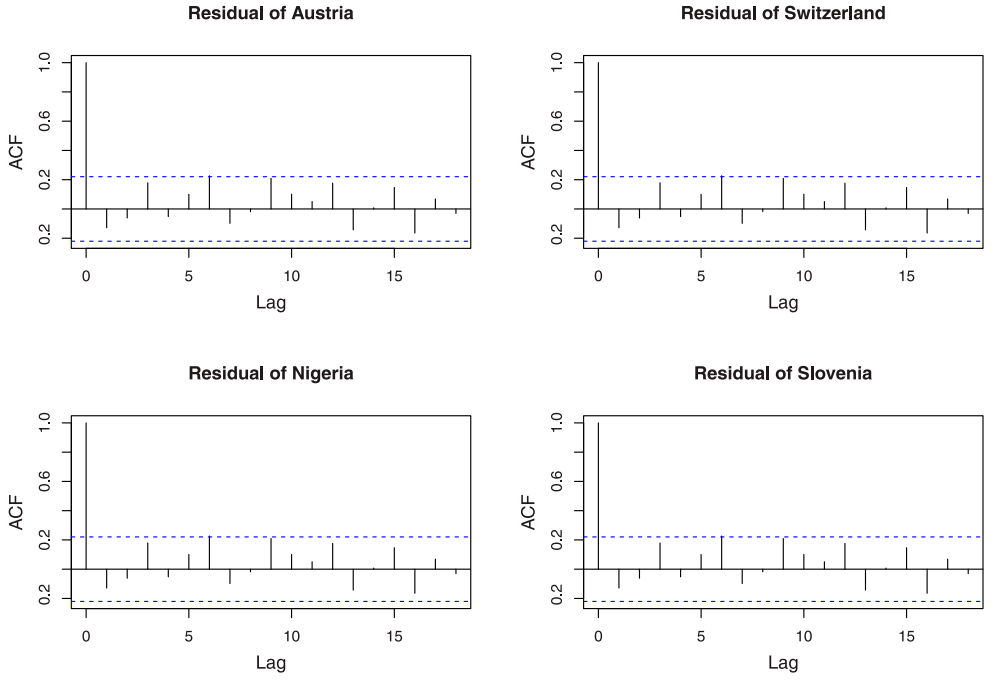

Figure 3. The Pearson residuals ACF for the COVID-19 daily deaths sets.

We evaluate the adequacy of fit and relative performance of the BDBH-MINAR(1) process within the competing models by scoring rules. The quality of probabilistic predictions is assessed by the scoring rules in decision analysis which allocate the numerical score based on the predictive distribution and the observed data. We consider the logarithmic, quadratic and spherical scores defined as $\text{logs}(P, x_t) = -\log P(x_t)$, $\text{qs}(P, x_t) = -2P(x_t) + \| P \|^2$ and $\text{sphs}(P, x_t) = -\dfrac{P(x_t)}{\| P \|^2}$, respectively, where $P(x_t)$ is the PMF of the predictive distribution at the observed count and $\| P \|^2 = \sum_{i=0}^{\infty} P(i)^2$. For more information, see Czado *et al.* [10].

Table 9 shows mean scores for competing models for the COVID-19 data series. The score diagnostic tools prefer the BDBH-MINAR(1) process among all of the competitive INAR models.

6.2. Forecasting methods

In this section, we provide the forecasting of considered data sets under the classical, modified Sieve bootstrap and Bayesian approaches to verify the appropriateness and predictability of the BDBH-MINAR(1) model.

The k-step ahead classical predictor of the BDBH-MINAR(1) model is represented as

$$\hat{X}_{t+k} = E\left(X_{t+k} \mid X_t\right) = (\alpha\phi)^k X_t + \frac{1-(\alpha\phi)^k}{1-\alpha\phi} E(Z_{t+k}),$$

where unknown parameters α, ϕ and $E(Z_{t+k})$ are substituted by the respective CML estimates.

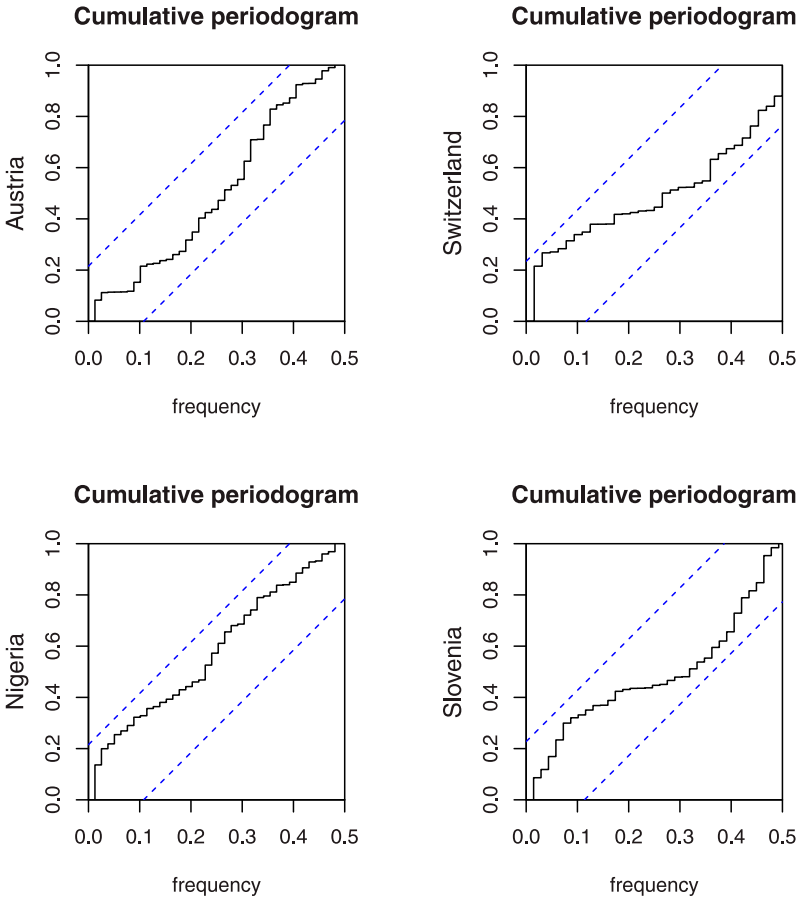

Figure 4. The Pearson residuals cumulative periodogram of for the COVID-19 daily deaths sets.

6.2.1. Modified Sieve bootstrap forecasting approach

The count time series has integer nature and the classical forecasts are not compatible with the nature of count data. The modified Sieve bootstrap method keeps the count data's integer character. Hence, we modified the bootstrap approach proposed by Pascual *et al.* [20] to apply in BDBH-MINAR(1) model as the following steps.

(1) The parameters (α, ϕ) are estimated based on the Yule–Walker (YW) estimation approach, due to its non-parametric structure.
(2) Compute residuals $\hat{Z}_t = X_t - \hat{\alpha}\hat{\phi}X_{t-1}$, for $t = 2, \ldots, n$.
(3) The empirical distribution of the modified residuals $\tilde{Z}_t$ are provided as $\hat{F}_{\tilde{Z}_t}(z) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}(\tilde{Z}_t < z)$, where I(.) is the indicator function, $\tilde{Z}_t = [\hat{Z}_t]$, and [.] shows the nearest integer value.
(4) The bootstrap series X_t^b is represented as

$$X_t^b = \left(\hat{\phi}, \hat{\alpha} \circ X_{t-1}^b + Z_t^b\right) * \left(1 - \hat{\phi}, Z_t^b\right), \quad b = 1, \ldots, B,$$

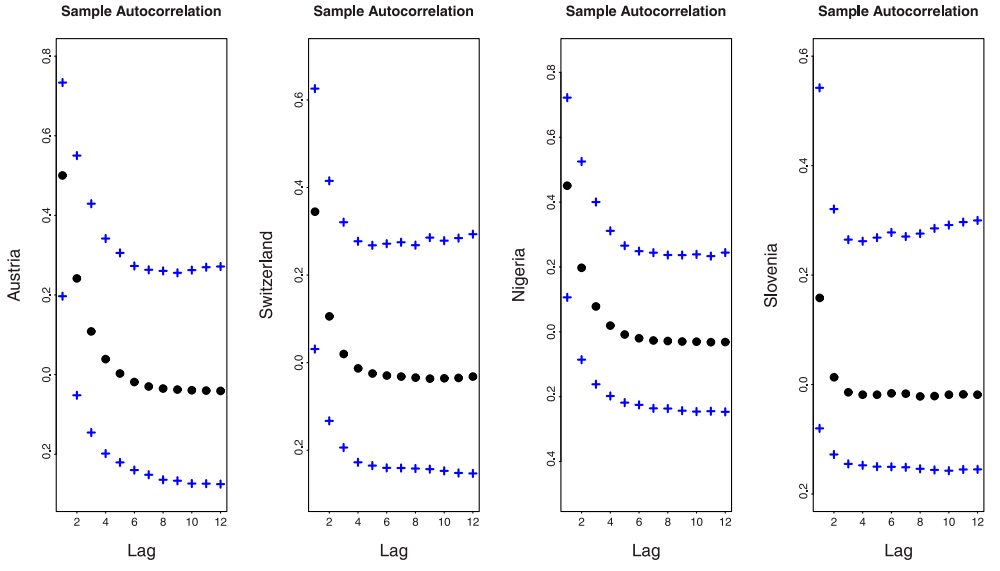

Figure 5. The bootstrap ACF and acceptance regions.

Table 9. Mean scores of COVID-19 data series.

	Austria			Switzerland			Nigeria			Slovenia		
Model	logs	qs	sphs	logs	qs	sphs	logs	qs	sphs	logs	qs	sphs
GINAR(1)	1.49	−0.23	−0.55	1.42	−0.13	−0.5	2.18	−0.16	−0.42	0.92	0.71	−0.39
DBH−INAR(1)	1.73	−0.37	−1.24	1.63	−0.36	−0.23	3.05	−0.19	−0.22	1.16	−0.56	−0.81
NGINAR(1)	1.47	0.15	−0.84	1.39	1.29	−0.32	2.17	0.71	−0.25	0.89	0.67	−0.24
NBRCINAR(1)	1.41	−0.56	−2.14	1.38	−0.59	−0.41	2.05	−0.44	−0.52	0.95	−1.19	−1.04
NBIINAR(1)	1.59	−0.31	−0.68	1.37	−0.38	−0.78	2.01	−0.22	−0.46	0.89	−0.55	−0.73
P−MPT(1)	1.44	−0.28	−0.53	2.51	0.04	−0.18	2.43	−0.11	−0.41	0.95	−0.53	−0.72
MPDINAR(1)	1.49	−0.27	−0.52	2.14	−0.04	−0.27	2.12	−0.19	−0.43	0.92	−0.51	−0.73
MPTSD(1)	1.41	−0.29	−0.53	1.82	−0.14	−0.38	2.18	−0.21	−0.44	0.88	−0.55	−0.74
BDBH−MINAR(1)	1.33	−0.64	−2.99	0.85	−1.71	−0.67	2.02	−0.58	−0.61	0.76	−1.52	−1.65

where B is the bootstrap sample size as $B = 500$, and Z_t^b is generated from the empirical distribution in step 3, for $t = 1, 2, \ldots, n$. The random samples Z_t^b from the empirical distribution are derived by the 'remp(.)' command in package 'EnvStats' of 'R' software.

(5) The YW estimation of the parameters $(\hat{\alpha}_{YW}, \hat{\phi}_{YW})$ is obtained by inserting the sample mean, variance, and solving the following equations:

$$\begin{cases} E(X_t) = \dfrac{E(Z_t)}{1-\alpha\phi} \\ Var(X_t) = \dfrac{E(Z_t^2)}{1-\alpha^2\phi} + \dfrac{\alpha\phi(1-\alpha)E(Z_t)}{(1-\alpha^2\phi)(1-\alpha\phi)} + \dfrac{E^2(Z_t)\left(2\alpha\phi(1-\alpha\phi)+\alpha^2\phi-1\right)}{(1-\alpha^2\phi)(1-\alpha\phi)^2} \end{cases}.$$

(6) The parameters (α, ϕ) are estimated based on the sample means $\hat{\alpha}^b = \frac{1}{B}\sum_{i=1}^{B}\hat{\alpha}_{i,YW}$ and $\hat{\phi}^b = \frac{1}{B}\sum_{i=1}^{B}\hat{\phi}_{i,YW}$.

(7) The future bootstrap observations are given by the following recursion:

$$\hat{X}_{t+1}^{b} = I_t \left(\hat{\alpha}^b \circ X_t^b + Z_{t+1}^b \right) + (1 - I_t) Z_{t+1}^b,$$

where $I_t = \begin{cases} 1 & \text{w.p. } \hat{\phi}^b \\ 0 & \text{w.p. } 1-\hat{\phi}^b \end{cases}$, and Z_{t+1}^b is generated from the empirical distribution in step 4.

6.2.2. Bayesian forecasting approach

The Bayesian predictive function of X_{t+h} given X_t is given by

$$f\left(x_{t+h} \mid x_t\right) \propto \int_{\alpha} \int_{\phi} \int_{\beta} f\left(x_{t+h} \mid x_t, \alpha, \phi, \beta\right) P\left(\alpha, \phi, \beta \mid x_t\right) \mathrm{d}\beta \, \mathrm{d}\phi \, \mathrm{d}\alpha.$$

The h-step conditional PMF, $f(x_{t+h} \mid x_t, \alpha, \phi, \beta)$ must be computed for the Bayesian predict. Since a closed-form expression for the h-step conditional PMF is only available for $h = 1$, we apply the simulation scheme given in Gorgi [13] to approximate $f(x_{t+h} \mid x_t, \alpha, \phi, \beta)$. Also, the algorithm given in Shirozhan and Mohammadpour [27] is used to estimate X_{t+h}, after some adaptation as the following steps.

(1) Compute the initial estimate of $\delta = (\alpha, \phi, \beta)$ by the sample $X_1, \ldots, X_t$ and set the CML estimates as the initial values of δ.
(2) The sample $(\delta_1, \ldots, \delta_m)$ is generated by the adaptive rejection Metropolis sampling (ARMS) within Gibbs methodology, which can be performed by 'arms(.)' command in 'dlm' package in 'R' software by the full conditional distribution (13).
(3) For $i = 1, \ldots, m$, $j = 1, \ldots, B$ and simulate

$$X_{t+h}^{(i,j)} = I_{i,t} \left(\alpha_i \circ X_{t+h-1}^{(i,j-1)} + Z_{t+h}^{(i,j)} \right) + \left(1 - I_{i,t}\right) Z_{t+h}^{(i,j)},$$

where $m = 100$, $B = 500$, $\alpha_i \circ X_{t+h-1}^{(i,j-1)}$ is generated from a binomial distribution with parameters $(\alpha_i, X_{t+h-1}^{(i,j-1)})$ and $Z_{t+h}^{(i,j)}$ is generated from BDBH distribution with parameter β_i and $I_{i,t} = \begin{cases} 1 & \text{w.p. } \phi_i \\ 0 & \text{w.p. } 1-\phi_i \end{cases}$.
(4) Compute an approximation of $f(X_{t+h} \mid X_t, \alpha_i, \phi_i, \beta_i)$ as $\hat{f}(X_{t+h} \mid X_t = x, \alpha_i, \phi_i, \beta_i) = \frac{n_x^h}{B}$, where n_x^h denotes the number of draws $X_{t+h}^{(i,j)}$, $j = 1, \ldots, B$, equal to x.
(5) For each i $(i = 1, \ldots, m)$ sample $X_{t+h,i}$ from $\hat{f}(X_{t+h} \mid X_t, \alpha_i, \phi_i, \beta_i)$, using the inverse transform method adapted to integer variables, that is,
   (a) sample u from uniform $(0, 1)$,
   (b) calculate the least integer valued v : $\sum_{j=0}^{v} \hat{f}(X_{t+h} = j \mid X_t, \alpha_i, \phi_i, \beta_i) \geq u$,
   (c) consider $X_{t+h,i} = v$.

Note that steps 3 and 4 have been applied for the approximation of $f(x_{t+h} \mid x_t, \alpha, \phi, \beta)$, where the other steps are used to estimate X_{t+h} from $f(x_{t+h} \mid x_t)$.

Based on sampling $X_{t+h,1}, \ldots, X_{t+h,m}$, the -step ahead predictor of X_{t+h} can be calculated from the sample mean $\hat{X}_{t+h} = \frac{1}{m} \sum_{i=1}^{m} X_{t+h,i}$.

Table 10. The one-step ahead predictions of COVID-19 data series.

Austria				Switzerland			
Actual data	Bayesian	Bootstrap	Classical	Actual data	Bayesian	Bootstrap	Classical
2	2.101	4	2.345	3	2.282	1	2.442
0	1.282	1	1.251	1	1.797	2	1.665
0	2.051	0	1.251	0	1.727	0	1.276
0	1.949	3	1.251	0	1.021	3	1.276
0	2.131	2	1.251	1	1.929	3	1.665
2	2.202	2	2.345	1	1.121	2	1.665
1	1.131	1	1.798	1	1.787	0	1.665
2	2.062	2	2.345	0	1.001	2	1.276
2	2.282	0	2.345	0	2.181	0	1.276
3	2.515	2	2.893	0	1.838	2	1.276
SMAPE	0.861	0.902	0.924		1.072	1.127	1.221

Table 11. The one-step ahead predictions of COVID-19 data series.

Nigeria				Slovenia			
Actual data	Bayesian	Bootstrap	Classical	Actual data	Bayesian	Bootstrap	Classical
1	1.373	2	2.588	0	0.591	1	0.921
6	4.818	4	5.109	0	0.484	1	0.921
0	1.161	0	2.084	0	1.313	0	0.921
0	2.575	6	2.084	0	1.303	2	0.921
3	2.767	2	3.596	0	1.151	3	0.921
8	6.818	5	6.117	0	1.272	1	0.921
7	5.474	5	5.613	2	1.898	0	1.335
6	5.878	3	5.109	4	4.121	1	1.748
1	2.828	2	2.588	0	0.434	2	0.921
6	4.949	1	5.109	0	1.404	2	0.921
SMAPE	0.618	0.701	0.692		1.608	1.709	1.718

6.3. Point prediction results

In Table 11, the classic, modified Sieve bootstrap and Bayesian forecasts of 10 last values of the COVID-19 data series are reported, for which we know the observed values. The symmetric mean absolute percent error (SMAPE) factors are provided to compare the forecast systems. The less SMAPE value leads to a better forecasting scheme. Based on Table 11, the SMAPE values of the Bayesian forecasting are smaller than two others, whereas the modified Sieve bootstrap predictors are integers compatible with the nature of actual data.

Conclusions

We introduce a new balanced discrete Burr–Hatke distribution which can be considered for data sets with all kinds of dispersion. The applicability of the new discrete distribution is evaluated in modeling counting time series. We introduce a flexible mixing INAR(1) model based on the Pegram and binomial thinning operators with BDBH innovations. The proposed mixing INAR(1) model is appropriate in modeling any type of dispersions of data sets. Based on the Monte Carlo simulation scheme, the efficiency of the Bayesian estimators is confirmed in comparison to the CML, MCLS and YW estimators. Compared with other relevant INAR(1) models, our model has the privilege to provide the best modeling of COVID-19 data sets. Based on four daily counts of the COVID-19 death, the residual

analysis of mixing INAR(1) process is provided to affirm the adequacy of fitness. Several forecasting approaches are considered for COVID-19 data sets, and outperforming the Bayesian forecasting method is demonstrated over classic and modified sieve Bootstrap methods.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

[1] M. Ahsan-ul-Haq, A. Babar, S. Hashmi, A.S. Alghamdi, and A.Z. Afify, *The discrete type-II half-logistic exponential distribution with applications to COVID-19 data*, Pakistan J. Stat. Oper. Res. 17 (2021), pp. 921–932.

[2] G.B. Al-Ani, *Statistical modeling of the novel COVID-19 epidemic in Iraq*, Epidemiol. Methods 10 (2021), Article ID 20200025.

[3] A.A. Al-Babtain, A.H.N. Ahmed, and A.Z. Afify, *A new discrete analog of the continuous Lindley distribution, with reliability applications*, Entropy 22 (2020), p. 603.

[4] M.A. Al-Osh and E.E.A.A. Aly, *First order autoregressive time series with negative binomial and geometric marginals*, Commun. Stat. Theory Methods 21 (1992), pp. 2483–2492.

[5] M.A. Al-Osh and A.A. Alzaid, *First-order integer-valued autoregressive (INAR(1)) process*, J. Time Ser. Anal. 8 (1987), pp. 261–275.

[6] A.A. Alzaid and M.A. Al-Osh, *First-order integer-valued autoregressive (INAR(1)) process: Distributional and regression properties*, Stat. Neerl. 42 (1988), pp. 53–61.

[7] A.A. Alzaid and M.A. Al-Osh, *Some autoregressive moving average processes with generalized Poisson marginal distributions*, Ann. Inst. Statist. Math. 45 (1993), pp. 223–232.

[8] S. Chakaraborty and D. Chakaraborty, *Discrete gamma distribution: Properties and parameter estimation*, Commun. Stat.-Theory Methods 41 (2012), pp. 3301–3324.

[9] S. Chattopadhyay, R. Maiti, S. Das, and A. Biswas, *Change-point analysis through INAR process with application to some COVID-19 data*, Stat. Neerl. 76 (2022), pp. 4–34.

[10] C. Czado, T. Gneiting, and L. Held, *Predictive model assessment for count data*, Biom. J. 65 (2009), pp. 1254–1261.

[11] M. El-Morshedy, M.S. Eliwa, and E. Altun, *Discrete Burr–Hatke distribution with properties, estimation methods and regression model*, IEEE Access. 8 (2020), pp. 74359–74370.

[12] M.S. Eliwa and M. El-Morshedy, *A one-parameter discrete distribution for over-dispersed data: Statistical and reliability properties with applications*, J. Appl. Stat. 49 (2022), pp. 2467–2487.

[13] P. Gorgi, *Integer-valued autoregressive models with survival probability driven by a stochastic recurrence equation*, J. Time Ser. Anal. 39 (2018), pp. 150–171.

[14] A.B. Gumel, E.A. Iboi, C.N. Ngonghala, and E.H. Elbasha, *A primer on using mathematics to understand COVID-19 dynamics: Modeling, analysis and simulations*, Infect. Dis. Model. 6 (2021), pp. 148–168.

[15] P.E. Hagmark, *On construction and simulation of count data models*, Math. Comput. Simul. 77 (2008), pp. 72–80.

[16] H. Karlsen and D. Tjøstheim, *Consistent estimates for the NEAR(2) and NLAR(2) time series models*, J. R. Stat. Soc. Series. B. Stat. Methodol. 50 (1988), pp. 313–320.

[17] W.C. Khoo, S.H. Ong, and A. Biswas, *Modeling time series of counts with a new class of INAR(1) model*, Stat. Pap. 58 (2017), pp. 393–416.

[18] M. Maleki, M.R. Mahmoudi, D. Wraith, and K.H. Pho, *Time series modelling to forecast the confirmed and recovered cases of COVID-19*, Travel. Med. Infect. Dis. 37 (2020), Article ID 101742.

[19] A.H. Muse, A.H. Tolba, E. Fayad, O.A. Abu Ali, M. Nagy, and M. Yusuf, *Modelling the COVID-19 mortality rate with a new versatile modification of the log-logistic distribution*, Comput. Intell. Neurosci. 2021 (2021), Article ID 8640794.

[20] L. Pascual, J. Romo, and E. Ruiz, *Bootstrap predictive inference for ARIMA processes*, J. Time Ser. Anal. 25 (2004), pp. 449–465.
[21] H. Pourreza, E. Baloui Jamkhaneh, and E. Deiri, *A family of gamma-generated distributions: Statistical properties and applications*, Stat. Methods. Med. Res. 30 (2021), pp. 1850–1873.
[22] M.M. Ristić, H.S. Bakouch, and A.S. Nastić, *A new geometric first-order integer-valued autoregressive (NGINAR(1)) process*, J. Stat. Plan. Inference. 139 (2009), pp. 2218–2226.
[23] D. Roy, *The discrete normal distribution*, Commun. Stat. Theory Methods 32 (2003), pp. 1871–1883.
[24] D. Roy, *Discrete Rayleigh distribution*, IEEE Trans. Reliab. 53 (2004), pp. 255–260.
[25] N. Shamma and M. Mohammadpour, *Alternative procedures in dependent counting INAR process with application on COVID-19*, Commun. Stat. Simul. Comput. (2021). https://doi.org/10.1080/03610918.2021.1987468
[26] M. Shirozhan and M. Mohammadpour, *An INAR(1) model based on the Pegram and thinning operators with serially dependent innovation*, Commun. Stat. Simul. Comput. 49 (2018), pp. 2617–2638.
[27] M. Shirozhan and M. Mohammadpour, *A new class of INAR(1) model for count time series*, J. Stat. Comput. Simul. 88 (2018), pp. 1348–1368.
[28] M. Shirozhan, M. Mohammadpour, and H.S. Bakouch, *A new geometric INAR(1) model with mixing Pegram and generalized binomial thinning operators*, Iran. J. Sci. Technol. Trans. A Sci. 43 (2019), pp. 1011–1020.
[29] C.F. Tovissodé, S.H. Honfo, J.T. Doumaté, and R. Glélé Kakaï, *On the discretization of continuous probability distributions using a probabilistic rounding mechanism*, Mathematics 9 (2021), p. 555.
[30] M. Triacca and U. Triacca, *Forecasting the number of confirmed new cases of COVID-19 in Italy for the period from 19 May to 2 June 2020*, Infect. Dis. Model. 6 (2021), pp. 362–369.
[31] C.H. Weiß, *Thinning operations for modeling time series of count – a survey*, AStA Adv. Stat. Anal. 92 (2008), pp. 319–341.