JOURNAL OF APPLIED STATISTICS
2023, VOL. 50, NO. 2, 352–369
https://doi.org/10.1080/02664763.2021.1993798

OPEN ACCESS Check for updates

# A first-order binomial-mixed Poisson integer-valued autoregressive model with serially dependent innovations

Zezhun Chen, Angelos Dassios and George Tzougas[†]

Department of Statistics, London School of Economics, London, UK

**ABSTRACT**

Motivated by the extended Poisson INAR(1), which allows innovations to be serially dependent, we develop a new family of binomial-mixed Poisson INAR(1) (BMP INAR(1)) processes by adding a mixed Poisson component to the innovations of the classical Poisson INAR(1) process. Due to the flexibility of the mixed Poisson component, the model includes a large class of INAR(1) processes with different transition probabilities. Moreover, it can capture some overdispersion features coming from the data while keeping the innovations serially dependent. We discuss its statistical properties, stationarity conditions and transition probabilities for different mixing densities (Exponential, Lindley). Then, we derive the maximum likelihood estimation method and its asymptotic properties for this model. Finally, we demonstrate our approach using a real data example of iceberg count data from a financial system.

**ARTICLE HISTORY**
Received 14 April 2021
Accepted 7 October 2021

**KEYWORDS**
Count data time series; binomial-mixed Poisson INAR(1) models; mixed Poisson distribution; overdispersion; maximum likelihood estimation

## 1. Introduction

Modelling the integer-valued count time series has attracted a lot of attention over the last few years in a plethora of different scientific fields such as the social sciences, healthcare, insurance, economics and the financial industry. The standard ARMA model will inevitably introduce real-valued results, and so is not appropriate for modelling this type of data. As a result, many alternative classes of integer-valued time series models have been introduced and explored in the applied statistical literature. The Integer-valued autoregressive process of order one, abbreviated as INAR(1), was proposed by McKenzie [8] and Al-Osh and Alzaid [1] as a counterpart to the Gaussian AR(1) model for Poisson counts. This model was derived by manipulating the operation between coefficients and variables, as well as the innovation term, in such a way that the values are always integers. The relationship of coefficients and variables is defined as $\alpha \circ X_t = \sum_{i=1}^{k} V_i$ such that $V_i$ are i.i.d Bernoulli random variables with parameter $\alpha$ and $^\circ$ denotes the binomial thinning operator. The binomial thinning is very easy to interpret, and binomial INAR(1) has the same

**CONTACT** Zezhun Chen z.chen58@lse.ac.uk, czz0328@outlook.com Department of Statistics, London School of Economics, London WC2A 2AE, UK

[†]Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh EH14 4AS, UK

autocorrelation structure as the standard AR(1) model and hence can be applied to fit the count data. For a general review, please see [11,12].

Later on, in order to accommodate different features exhibited by count data, for example, under-dispersion, overdispersion, probability of observing zero and different dependent structures, many research studies introduced alternative thinning operators or varied the distribution of $V_i$ for different needs. The case where $V_i$ are i.i.d geometric random variables is analysed by Ristić *et al.* [10], which is called NGINAR(1). Kirchner [7] introduced reproduction operators so that $V_i$ are i.i.d Poisson random variables to explore the relationship between Hawkes process and integer-valued time series. For further variation, random coefficients thinning is introduced so that $V_i$ are i.i.d Bernoulli with the parameter $\alpha$ being a random variable. This type of thinning operator was proposed by McKenzie [8,9] and Zheng *et al.* [14]; they applied this to a generalized INAR(1) model. In particular, to accommodate the overdispersion feature, one way is to change the thinning operators from binomial to other types as discussed above. Another way is to replace the innovation distribution with some other overdispersed distribution; for example, see [2]. A third approach would be to keep the structure of binomial INAR(1) but to allow the innovation terms to be serially dependent; see [13].

In this study, motivated by Weiß [13], we develop a new family of binomial-mixed Poisson INAR(1) (BMP INAR(1)) processes by adding a mixed Poisson component to the innovations term of the classical Poisson INAR(1) process. The proposed class of BMP INAR(1) processes is ideally suited for modelling heterogeneity in count time series data since, due to the mixed Poisson component which we introduce herein, it includes many members with different transition probabilities that can adequately capture different levels of overdispersion in the data while keeping the innovation as independent Poisson.

The rest paper is organized as follows. Section 2 defines the Binomial mixed Poisson INAR(1) model by adding a mixed Poisson component in the Poisson INAR(1) model. Statistical properties and the stationarity condition are derived in Section 3. Section 4 derives the distribution of the mixed Poisson component based on two different mixing density functions from the exponential family, namely the Exponential and Lindley distributions. In Section 5, maximum likelihood estimation is discussed as well as its asymptotic properties for the estimators. In Section 6, the model is fitted to financial data (iceberg count) and discuss numerical results. Finally, concluding remarks are provided in Section 7.

## 2. Construction of binomial mixed Poisson INAR(1)

In [13], the classical Poisson INAR(1) was extended by allowing the innovations $\varepsilon$ to depend on the current state of the model $X_t$ such that $\varepsilon_t \sim \text{Po}(aX_{t-1} + b)$, where $a$ and $b$ are some positive constants. The innovation with this definition is separable in the sense that $\varepsilon_t = a * X_{t-1} + \epsilon_t$, where $a * X_{t-1} = \sum_{i=1}^{X_{t-1}} U_i$, with $U_i \overset{\text{i.i.d}}{\sim} \text{Po}(a)$ and $\epsilon_t \sim \text{Po}(b)$. To introduce further heterogeneity while maintaining serially dependent innovations structure in this model, we extend this by allowing $U_i$ to be a mixed Poisson random variable.

Starting from a Poisson random variable $U$ with parameter $\theta$, we may obtain a large class of random variables by allowing $\theta$ to be another random variable which follows some classes of density function $g(\theta \mid \varphi)$ where $\varphi$ can be a scalar or a vector; see Karlis [6].

The random variable $U$ follows a Mixed Poisson distribution with $g$ as a mixing density. The distribution function of $U$ is defined as

$$P(U=u)=\int_0^\infty \frac{e^{-\theta_i}\theta_i^u}{u!}g(\theta\mid\varphi)\,d\theta. \tag{1}$$

We now construct our model.

**Definition 2.1:** The Binomial-Mixed Poisson integer-valued Autoregressive model (BMP INAR(1)) is defined by the following equations:

$$\begin{aligned} X_{t+1} &= p_1\circ X_t+\varepsilon_{t+1}\\ &= p_1\circ X_t+\varphi *_g X_t+Z_{t+1},\\ p_1\circ X_t &= \sum_{k=1}^{X_t}V_k,\quad \varphi *_g X_t=\sum_{i=1}^{X_t}U_i,\\ P(U_i=x) &= \int_0^\infty \frac{e^{-\theta_i}\theta_i^x}{x!}g(\theta_i\mid\varphi)\,d\theta_i, \end{aligned} \tag{2}$$

where

- ° is a binomial thinning operator such that $V_i$ are i.i.d Bernoulli random variables with parameter $p_1\in[0,1]$;
- $\{Z_t\}_{t=1,2,\ldots}$ are i.i.d Poisson random variables with rate $\lambda_1>0$;
- $*_g$ is a reproduction operator such that $U_i$ are independent Mixed Poisson distributed with mixing density function $g(\theta_i\mid\varphi)$;
- $*_g$ and ° are independent of each other so that $U_i$ and $V_k$ are independent of each other.

As we will see shortly, the stationarity condition for this model is simply $p_1+\mu_g<1$ where $\mu_g$ is the first moment of $U_i$. When it comes to interpretation, this model can be seen as the evolution of a population where the binomial part indicates the survivors from the previous period, the mixed Poisson part is the total offspring and the innovation part indicates immigrants. Obviously, this model is a Markov Chain and its transition probability can be found easily once we know the mixing density $g(\theta\mid\varphi)$. The probability mass function of $Y_{t+1}=\varphi *_g X_t$ is given by

$$P(Y_{t+1}=y\mid X_t=n)=\mathbb{E}\left[\frac{e^{-\sum_{i=1}^n\theta_i}(\sum_{i=1}^n\theta_i)^y}{y!}\right], \tag{3}$$

where the expectation is taken over $\theta_1,\theta_2,\ldots,\theta_n$. In order to evaluate the expectation explicitly, it would be desirable that the random variables $\theta_i$ have an 'additivity' property such that density (or probability mass function) of the sum $\sum_{i=1}^n\theta_i$ is either itself with different parameters or can be written in a closed form. Many members of the exponential family have this kind of property. In general, we let $g(x\mid\varphi)$ be of an exponential family form such that

$$g(x \mid \varphi) = h(x) \exp\{\eta(\varphi)T(x) + \xi(\varphi)\}. \tag{4}$$

Denote the density of the sum $S_n = \sum_{i=1}^{n} \theta_i$ as $g_n(s \mid \varphi)$, where $\theta_i$ are i.i.d random variables with density $g(\theta \mid \varphi)$. The expectation above can be expressed as

$$P(Y_{t+1} = y \mid X_t = n) = \int_{\mathbb{R}^+} \frac{e^{-s} s^y}{y!} g_n(s \mid \varphi)\, ds. \tag{5}$$

The density $g_n(s \mid \varphi)$ is explicitly known in many cases, for example, it can be an Inverse Gaussian, Exponential, Gamma, Geometric, Bernoulli or Lindley. For the sake of parsimony, we use distributions with a single parameter. In other words, we assume that $\varphi$ is scalar. Note that, if we let $g(\theta \mid \varphi) = \delta_\varphi(\theta)$ – a Dirac delta function concentrating at $\varphi$, the model will recover to the Extended Poisson INAR(1) in [13].

## 3. Statistical properties of BMP INAR(1)

### *3.1. Moments and correlation structure*

We first need to derive the moments of $U_i$.

**Lemma 3.1:** *The first moment and second central moment of $U_i$ with density $g(x \mid \varphi)$ are given by*

$$\mathbb{E}[U_i] = \mu_g, \quad \mathrm{Var}(U_i) = \mu_g + \sigma_g^2, \tag{6}$$

*where $\mu_g = \mathbb{E}_g[\theta_i] = \int_R x g(x \mid \varphi)\, dx$ and $\sigma_g^2 = Var_g(\theta_i)$.*

***Proof:*** By the conditional expectation argument

$$\begin{aligned}
\mathbb{E}[U_i] &= \mathbb{E}_g[\mathbb{E}[U_i \mid \theta_i]] = \mathbb{E}_g[\theta_i] = \mu_g, \\
\mathbb{E}[U_i^2] &= \mathbb{E}_g[\mathbb{E}[U_i^2 \mid \theta_i]] = \mathbb{E}_g[\theta_i^2 + \theta_i], \\
\mathrm{Var}(U_i) &= \mathbb{E}[U_i^2] - (\mathbb{E}[U_i])^2 = \sigma_g^2 + \mu_g.
\end{aligned}$$

■

**Proposition 3.1:** *Assume $p_1 + \mu_g < 1$. The stationary moments of $X_t$ is given by*

$$\begin{aligned}
\mathbb{E}[X_t] &= \mu_x = \frac{\lambda_1}{1 - p_1 - \mu_g}, \\
\mathrm{Var}(X_t) &= \sigma_x^2 = \mu_x \frac{1 - p_1^2 + \sigma_g^2}{1 - (p_1 + \mu_g)^2}, \\
\mathrm{Cov}(X_t, X_{t-k}) &= \gamma(k) = (p_1 + \mu_g)^k \sigma_x^2.
\end{aligned} \tag{7}$$

***Proof:*** For the first moment, we have

$$\begin{aligned}\mathbb{E}[X_t] &= \mathbb{E}[p_1 \circ X_{t-1}] + \mathbb{E}[\varphi *_g X_{t-1}] + \mathbb{E}[Z_t],\\ \mu_x &= p_1\mu_x + \mu_g\mu_x + \lambda_1,\\ \mu_x &= \frac{\lambda_1}{1 - p_1 - \mu_g}.\end{aligned}$$

Since the operators ∘ and $*_g$ are independent of each other, for the second central moment, we have

$$\begin{aligned}\mathrm{Var}(X_t) &= \mathrm{Var}(p_1 \circ X_{t-1} + \varphi *_g X_{t-1}) + \mathrm{Var}(Z_t)\\ &= \mathrm{Var}\left(\mathbb{E}\left[\sum_{i=1}^{X_{t-1}}(V_i + U_i) \mid X_{t-1}\right]\right) + \mathbb{E}\left[\mathrm{Var}\left(\sum_{i=1}^{X_{t-1}}(V_i + U_i) \mid X_{t-1}\right)\right] + \lambda_1\\ &= (p_1 + \mu_g)^2\sigma_x^2 + (p_1(1 - p_1) + \sigma_g^2 + \mu_g)\mu_x + \lambda_1,\\ \sigma_x^2 &= \mu_x \frac{1 - p_1^2 + \sigma_g^2}{1 - (p_1 + \mu_g)^2}.\end{aligned}$$

Let $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \dots,)$ be the $\sigma$-algebra generated by the model $X_t$ up to time t, the covariance of the model is given by

$$\mathrm{Cov}(X_t, X_{t-k}) = \mathrm{Cov}(p_1 \circ X_{t-1}, X_{t-k}) + \mathrm{Cov}(\varphi *_g X_{t-1}, X_{t-k}) + \mathrm{Cov}(Z_t, X_{t-k}).$$

Again by using conditional expectations, we have

$$\begin{aligned}\mathrm{Cov}(p_1 \circ X_{t-1}, X_{t-k}) &= \mathrm{Cov}(\mathbb{E}[p_1 \circ X_{t-1} \mid \mathcal{F}_{t-1}], \mathbb{E}[X_{t-k} \mid \mathcal{F}_{t-1}])\\ &\quad + E[\mathrm{Cov}(p_1 \circ X_{t-1}, X_{t-k} \mid \mathcal{F}_{t-1})]\\ &= \mathrm{Cov}(p_1 X_{t-1}, X_{t-k}) + \mathbb{E}\left[\mathrm{Cov}\left(\sum_{i=1}^{X_{t-1}} V_i, X_{t-k} \mid \mathcal{F}_{t-1}\right)\right]\\ &= p_1\gamma(k-1) + 0.\end{aligned}$$

Obviously, $\mathrm{Cov}(X_t, X_{t-k}) = \gamma(k) = (p_1 + \mu_g)\gamma(k-1) = (p_1 + \mu_g)^k\gamma(0)$. ■

From the results above, it is clear that this model follows the same correlation structure as that of standard AR(1) model. Furthermore, unlike equal-dispersed Poisson INAR(1), BMP INAR(1) is in general an overdispersed model with Fisher index of dispersion

$$\mathrm{FI}_x = \frac{\sigma_x^2}{\mu_x} = 1 + \frac{\mu_g^2 + 2p_1\mu_g + \sigma_g^2}{1 - (p_1 + \mu_g)^2}. \tag{8}$$

### 3.2. Existence of stationary solution

**Proposition 3.2:** *Given that $P(U_i = 0) > 0$ and $p_1 + \mu_g < 1$ the following infinite sequence:*

$$f_i(\theta) = (1 - p_1 + p_1 f_{i-1}(\theta))\Phi_u(f_{i-1}(\theta)), \quad i \geq 1,$$
$$f_0(\theta) = \theta, \quad \theta \in [0, 1], \tag{9}$$

*where $\Phi_u(\theta)$ is the probability generating function (p.g.f) of $U_i$, has a limit $\lim_{i\to\infty} f_i(\theta) = 1$*

***Proof:*** Define the increment of the sequence

$$\begin{aligned} f_i(\theta) - f_{i-1}(\theta) &= (1 - p_1 + p_1 f_{i-1}(\theta))\Phi_u(f_{i-1}(\theta)) - f_{i-1}(\theta) \\ &= (1 - p_1 + p_1 x)\Phi_u(x) - x \quad x = f_i(\theta) \\ &=: Q(x). \end{aligned}$$

By the definition of p.g.f, $x \in [0, 1]$, the monotonicity of this function is shown by its first and second derivatives

$$\begin{aligned} Q'(x) &= p_1\Phi_u(x) + (1 - p_1 + p_1 x)\Phi_u(x) - 1, \\ Q''(x) &= 2p_1\Phi'_u(x) + (1 - p_1 + p_1 x)\Phi''_u(x). \end{aligned}$$

By the definition of p.g.f, $\Phi'(x) \geq 0$ and $\Phi''(x) \geq 0$. So $Q''(x) \geq 0$, which implies $Q'(x)$ is non-decreasing function. Then we have

$$Q'(x) \leq Q'(1) = p_1 + \mu_g - 1 < 0.$$

Notice that $Q(0) = (1 - p_1)P(U_i = 0) > 0,\ Q(1) = 0$. Hence we can conclude that $Q$ is a monotonic decreasing function ranging from 0 to $Q(0)$. In order words, for any $i = 1, \ldots,$ and $\theta \in [0, 1]$, the sequence $f_i(\theta) = f_{i-1}(\theta) + Q(f_{i-1}(\theta))$ is increasing with respect to $i$. Finally, $\lim_{i\to\infty} f_i(\theta) = 1$. ■

**Proposition 3.3:** *Let $X_t$ be the BMP INAR(1) model defined in Definition 2.1. If the condition $P(U_i) > 0$ and $p_1 + \mu_g < 1$ holds, then the process $X_t$ has a proper stationary distribution and $X_t$ is an ergodic Markov Chain. The stationary distribution is $\Phi_x(\theta) = \prod_{i=0}^{\infty} \Phi_z(f_i(\theta))$.*

***Proof:*** Denote the p.g.f of $X_n$ and the innovation $Z_n$ as $\Phi_{X_n}(\theta)$ and $\Phi_z(\theta)$respectively, then $\Phi_{X_n}(\theta)$ can be expressed as following product:

$$\begin{aligned} \Phi_{X_n}(\theta) &= \mathbb{E}[\mathbb{E}[\theta^{X_n}|X_{n-1}]|X_0] \\ &= \mathbb{E}[\mathbb{E}[\theta^{p_1 \circ X_{n-1} + \varphi *_g X_{n-1} + Z_t}|X_{n-1}]|X_0] \\ &= \mathbb{E}[f_1(\theta)^{X_{n-1}}|X_0]\Phi_z(f_0(\theta)) \end{aligned}$$

$$= \vdots \qquad \vdots$$

$$= \mathbb{E}[f_n(\theta)^{X_0}] \prod_{i=0}^{n-1} \Phi_z(f_i(\theta)).$$

To show the existence of the limiting distribution is equivalent to show the limit of the product as n goes to infinity is something other than 0, which means that we have to show that the series

$$LP_n = \log \Phi_{X_n}(\theta) = \log \mathbb{E}[f_n(\theta)^{X_0}] + \sum_{i=0}^{n-1} \log \Phi_z(f_i(\theta))$$

is convergent as $n \to \infty$. The convergence of the infinite series $\sum_{i=0}^{\infty} \log \Phi_z(f_i(\theta))$ can be shown by the ratio test

$$\begin{aligned}
&\lim_{i\to\infty} \left| \frac{\log \Phi_z(f_i(\theta))}{\log \Phi_z(f_{i-1}(\theta))} \right| \\
&\quad = \lim_{x\to 1} \frac{\log \Phi_z((1-p_1+p_1x)\Phi_u(x))}{\log \Phi_z(x)} \\
&\quad = \lim_{x\to 1} \frac{\Phi_z(x)}{\Phi_z'(x)} \frac{\Phi_z'((1-p_1+p_1x)\Phi_u(x))}{\Phi_z((1-p_1+p_1x)\Phi_u(x))} (p_1\Phi_u(x) + (1-p_1+p_1x)\Phi_u'(x)) \\
&\quad = p_1 + \mu_g < 1. \qquad (10)
\end{aligned}$$

Hence $\lim_{n\to\infty} LP_n > -\infty$, from which we can infer that $\lim_{n\to\infty} \Phi_{X_n}(\theta) > 0$ exists and the limiting distribution of $X_n$ exists. Furthermore, by the construction of $X_n$, the chain is defined on a countable state space $\mathcal{S} = \{0, 1, 2, \ldots\}$. The positivity of transition probability $\mathcal{P}(X_n = j \mid X_{n-1} = i) > 0$, $\forall i, j \in \mathcal{S}$ implies that $X_n$ is irreducible and aperiodic. Hence the limiting distribution $\Phi_x(\theta) = \lim_{n\to\infty} \Phi_{X_n}(\theta)$ is the unique stationary distribution for $X_n$. ■

In general, $P(U_i = 0) = \int_{\mathbb{R}^+} \mathrm{e}^{-\theta} g(\theta \mid \varphi)\, \mathrm{d}\theta > 0$ as long as $g(\theta \mid \varphi) > 0$, so we just need to ensure the existence of the first moment to achieve the stationarity of $X_n$. The infinite product $\Phi_x(\theta) = \prod_{i=0}^{\infty} \Phi_z(f_i(\theta))$ is the p.g.f of the stationary distribution, which also satisfies

$$\Phi_x(\theta) = \Phi_x((1-p_1+p_1\theta)\Phi_u(\theta))\Phi_z(\theta). \qquad (11)$$

## 4. Distribution function of the mixed Poisson component

In order to apply maximum likelihood estimation for the statistical inference of this model, we need to derive the distribution of $Y_{t+1} = \varphi *_g X_t$ according to different density functions $g$. As mentioned before, we focus on the density $g$ coming from the exponential family. For expository purposes, we will derive the distribution of $Y_{t+1}$ based on exponential and Lindley densities.

### 4.1. Mixed by exponential density

If $g(\theta \mid \varphi) = \frac{1}{\varphi} e^{-\frac{1}{\varphi}\theta}$, then the distribution of $U_i$ is given by

$$P(U_i = x) = \int_0^\infty \frac{e^{-\theta_i}\theta_i^x}{x!} \frac{1}{\varphi} e^{-\frac{1}{\theta_i}x} \, d\theta_i$$
$$= \frac{1}{\varphi x!} \int_0^\infty e^{-(1+\frac{1}{\varphi})\theta_i}\theta_i^x \, d\theta_i$$
$$= \left(\frac{1}{1+\varphi}\right)\left(\frac{\varphi}{1+\varphi}\right)^x, \quad x = 0, 1, \ldots, \tag{12}$$

which is a geometric distribution with parameter $\frac{\varphi}{1+\varphi}$. Then, the distribution function $f_\varphi(m, X_t)$ of $\varphi *_g X_t$ as well as its first and second derivatives are given by

$$f_\varphi(m, X_t) = C_{m+X_t-1}^m \left(\frac{1}{1+\varphi}\right)^{X_t}\left(\frac{\varphi}{1+\varphi}\right)^m,$$
$$\frac{\partial f_\varphi(m, X_t)}{\partial \varphi} = \left(\frac{m}{\varphi(1-\varphi)} - \frac{X_t}{1+\varphi}\right) f_\varphi(m, X_t),$$
$$\frac{\partial^2 f_\varphi(m, X_t)}{\partial (\varphi)^2} = \left(\left(\frac{m}{\varphi(1-\varphi)} - \frac{X_t}{1+\varphi}\right)^2 + \frac{X_t}{(1+\varphi)^2} - \frac{m(1+2\varphi)}{\varphi^2(1+\varphi)^2}\right) f_\varphi(m, X_t). \tag{13}$$

Note that $X_t$ will recover to the NGINAR(1) in [10] if we further let $p_1 = 0$. In general, the stationarity condition becomes $p_1 + \varphi < 1$ and the probability generating function of $X_t$ satisfies the equation

$$\Phi_x(\theta) = \Phi_x\left(\frac{1 - p_1 + p_1\theta}{1 + \varphi - \varphi\theta}\right)\Phi_z(\theta). \tag{14}$$

We will now relax the assumption of the innovation term being Poisson and let the marginal distribution of X be a geometric random variable with parameter $\frac{\alpha}{1+\alpha}, \alpha > 0$. Using the relationship of the p.g.f, we can infer the required distribution of $Z$.

**Proposition 4.1:** *If $p_1 > \varphi, \alpha > \varphi$ or $p_1 < \varphi, \alpha < \varphi$ and the distribution of $\{Z_t\}_{t=1,2,\ldots}$ follows a mixed geometric distribution such that*

$$Z_t = \begin{cases} \text{Geom}\left(\frac{\varphi}{1+\varphi}\right), & W.P. \quad \frac{(p_1-\varphi)\alpha}{\alpha-\varphi}, \\ \text{Geom}\left(\frac{\alpha}{1+\alpha}\right), & W.P. \quad 1 - \frac{(p_1-\varphi)\alpha}{\alpha-\varphi}, \end{cases} \tag{15}$$

*then the marginal distribution of X follows a* Geom$(\frac{\alpha}{1+\alpha})$ *distribution.*

***Proof:*** By utilizing equation (14), we assume the X has a geometric distribution such that $\Phi_x(\theta) = \frac{1}{1+\alpha-\alpha\theta}$. Then, the probability generating function of Z has the following form:

$$\Phi_z(\theta) = \frac{\Phi_x(\theta)}{\Phi_x(\frac{1-p_1+p_1\theta}{1+\varphi-\varphi\theta})}$$

$$= \frac{(1+\varphi-\varphi\theta)(1+\alpha)-\alpha(1-p_1+p_1\theta)}{(1+\alpha-\alpha\theta)(1+\varphi-\varphi\theta)}$$
$$= \frac{(p_1-\varphi)\alpha}{\alpha-\varphi}\frac{1}{1+\varphi-\varphi\theta}+\left(1-\frac{(p_1-\varphi)\alpha}{\alpha-\varphi}\right)\frac{1}{1+\alpha-\alpha\theta}. \tag{16}$$

■

## 4.2. Mixed by Lindley density

Suppose now the density $g(\theta\,|\,\varphi)=\frac{\varphi^2}{1+\varphi}(\theta+1)\,\mathrm{e}^{-\varphi\theta}$ is a Lindley density function. The distribution of $U_i$ is the so-called Poisson–Lindley distribution, see [6], which has the following probability mass function

$$\begin{aligned} P(U_i=x) &= \int_0^\infty \frac{\mathrm{e}^{-\theta_i}\theta_i^x}{x!}\frac{\varphi^2}{1+\varphi}(\theta_i+1)\,\mathrm{e}^{-\varphi\theta_i}\,\mathrm{d}\theta_i \\ &= \frac{\varphi^2}{(1+\varphi)x!}\left(\int_0^\infty \theta_i^{x+1}\,\mathrm{e}^{-(\varphi+1)\theta_i}\,\mathrm{d}\theta_i+\int_0^\infty \theta_i^{x}\,\mathrm{e}^{-(\varphi+1)\theta_i}\,\mathrm{d}\theta_i\right) \\ &= \frac{\varphi^2}{(1+\varphi)x!}\left(\frac{\Gamma(x+2)}{(1+\varphi)^{x+2}}+\frac{\Gamma(x+1)}{(1+\varphi)^{x+1}}\right) \\ &= \frac{\varphi^2(\varphi+2+x)}{(1+\varphi)^{x+3}}, \quad x=0,1,\ldots. \end{aligned} \tag{17}$$

Under this parameter setting, $\mathbb{E}[U_i]=\mu_g=\frac{\varphi+2}{\varphi(\varphi+1)}$ which makes the parameter $\varphi$ less interpretable. So we adopt the following parameter setting for the mixing density $g(\theta\,|\,\varphi)$

$$g(\theta\,|\,\varphi)=\frac{\tilde{\varphi}^2}{1+\tilde{\varphi}}(\theta+1)\,\mathrm{e}^{-\tilde{\varphi}\theta} \quad \tilde{\varphi}=\frac{1-\varphi+\Delta}{2\varphi} \quad \Delta=\sqrt{(\varphi-1)^2+8\varphi}. \tag{18}$$

Then, $\mu_g=\varphi$, $\sigma_g=\varphi^2-\frac{2}{(\tilde{\varphi}(1+\tilde{\varphi}))^2}$. On the other hand, the additivity of $U_i$ is not that clear. In order to evaluate the expectation (3), we need to find out the distribution of $S_n=\sum_{i=1}^n\theta_i$.

**Proposition 4.2:** *Suppose $\theta_i$ are i.i.d Lindley distributed. The density of the sum $S_n=\sum_{i=1}^n\theta_i$ is given by*

$$g_n(s\,|\,\varphi)=\left(\frac{\tilde{\varphi}^2}{1+\tilde{\varphi}}\right)^n\mathrm{e}^{-\tilde{\varphi}s}\sum_{k=0}^n\frac{C_n^k}{\Gamma(n+k)}s^{n+k-1}. \tag{19}$$

***Proof:*** We can prove this by inverting the Laplace transform. The Laplace transform of $\theta_i$ is

$$\begin{aligned} \mathbb{E}[\mathrm{e}^{-\nu\theta_i}] &= \int_0^\infty \frac{\tilde{\varphi}^2}{1+\tilde{\varphi}}(\theta_i+1)\,\mathrm{e}^{-(\nu+\tilde{\varphi})\theta_i}\,\mathrm{d}\theta_i \\ &= \frac{\tilde{\varphi}^2}{1+\tilde{\varphi}}\frac{\tilde{\varphi}+\nu+1}{(\tilde{\varphi}+\nu)^2}. \end{aligned}$$

Then the Laplace transform of $S_n$ is simply the product of $\mathbb{E}[e^{-\nu\theta_i}]$, which is

$$\mathbb{E}[e^{-\nu S_n}] = \left(\frac{\tilde{\varphi}^2}{1+\tilde{\varphi}}\right)^n \frac{(\tilde{\varphi}+\nu+1)^n}{(\tilde{\varphi}+\nu)^{2n}}.$$

Using a binomial expansion, we have

$$\begin{aligned}
\mathbb{E}[e^{-\nu S_n}] &= \left(\frac{\tilde{\varphi}^2}{1+\tilde{\varphi}}\right)^n \frac{1}{(\tilde{\varphi}+\nu)^{2n}} \sum_{k=0}^{n} C_n^k (\tilde{\varphi}+\nu)^k \\
&= \left(\frac{\tilde{\varphi}^2}{1+\tilde{\varphi}}\right)^n \frac{1}{(\tilde{\varphi}+\nu)^{n}} \sum_{k=0}^{n} C_n^{n-k} (\tilde{\varphi}+\nu)^{-(n-k)} \\
&= \left(\frac{\tilde{\varphi}^2}{1+\tilde{\varphi}}\right)^n \sum_{k=0}^{n} C_n^k (\tilde{\varphi}+\nu)^{-(n+k)} \\
&= \left(\frac{\tilde{\varphi}^2}{1+\tilde{\varphi}}\right)^n \sum_{k=0}^{n} \int_0^\infty \frac{C_n^k}{\Gamma(n+k)} s^{n+k-1}\, e^{-\tilde{\varphi}s}\, e^{-\nu s}\, ds \\
&= \int_0^\infty e^{-\nu s} \left(\frac{\tilde{\varphi}^2}{1+\tilde{\varphi}}\right)^n e^{-\tilde{\varphi}s} \sum_{k=0}^{n} \frac{C_n^k}{\Gamma(n+k)} s^{n+k-1}\, ds.
\end{aligned}$$

Obviously, the density function of $S_n$ is the integrand except $e^{-\nu s}$. ■

Then, the distribution of $Y_{t+1} = \theta *_g X_t$ is given by the following proposition.

**Proposition 4.3:** *The probability mass function of $Y_{t+1} = \varphi *_g X_t$ as well as its derivatives are given by*

$$\begin{aligned}
f_\varphi(y,n) &= P(Y_{t+1}=y \mid X_t=n) = \left(\frac{\tilde{\varphi}^2}{1+\tilde{\varphi}}\right)^n \sum_{k=0}^{n} C_n^k C_{n+k+y-1}^y (1+\tilde{\varphi})^{-(n+k+y)}, \\
\frac{\partial f_\varphi(y,n)}{\partial\tilde{\varphi}} &= n\left(\frac{2}{\tilde{\varphi}} - \frac{1}{1+\tilde{\varphi}}\right) f_\varphi(y,n) - (y+1) f_\varphi(y+1,n), \\
\frac{\partial^2 f_\varphi(y,n)}{\partial\tilde{\varphi}^2} &= \left(n^2\left(\frac{2}{\tilde{\varphi}} - \frac{1}{1+\tilde{\varphi}}\right)^2 - n\left(\frac{2}{\tilde{\varphi}^2} - \frac{1}{(1+\tilde{\varphi})^2}\right)\right) f_\varphi(y,n) \\
&\quad - 2n(y+1)\left(\frac{2}{\tilde{\varphi}} - \frac{1}{1+\tilde{\varphi}}\right) f_\varphi(y+1,n) + (y+1)(y+2) f_\varphi(y+2,n), \\
\frac{\partial f_\varphi(y,n)}{\partial\varphi} &= \frac{\partial f_\varphi(y,n)}{\partial\tilde{\varphi}} \frac{\partial\tilde{\varphi}}{\partial\varphi}, \\
\frac{\partial^2 f_\varphi(y,n)}{\partial\varphi^2} &= \frac{\partial^2 f_\varphi(y,n)}{\partial\tilde{\varphi}^2}\left(\frac{\partial\tilde{\varphi}}{\partial\varphi}\right)^2 + \frac{\partial f_\varphi(y,n)}{\partial\tilde{\varphi}} \frac{\partial^2\tilde{\varphi}}{\partial\varphi^2},
\end{aligned} \tag{20}$$

*where*

$$\frac{\partial\tilde{\varphi}}{\partial\varphi} = -\frac{1}{2\varphi} + \frac{\varphi+3}{2\varphi\Delta} - \frac{1-\varphi+\Delta}{2\varphi^2},$$

$$\frac{\partial^2\tilde{\varphi}}{\partial\varphi^2} = \frac{1}{\varphi^2} + \frac{1}{2\varphi\Delta} + \frac{1-\varphi+\Delta}{\varphi^3} - \frac{(\varphi+3)^2}{2\varphi\Delta^3} - \frac{\varphi+3}{\varphi^2\Delta}.$$

***Proof:***

$$\begin{aligned}
P(Y_{t+1} = y \mid X_t = n) &= \mathbb{E}\left[\frac{e^{-\sum_{i=1}^{n}\theta_i}(\sum_{i=1}^{n}\theta_i)^y}{y!}\right] \\
&= \int_0^\infty \frac{e^{-s}s^y}{y!}\left(\frac{\varphi^2}{1+\varphi}\right)^n e^{-\varphi s}\sum_{k=0}^{n}\frac{C_n^k}{\Gamma(n+k)}s^{n+k-1}\,ds \\
&= \left(\frac{\varphi^2}{1+\varphi}\right)^n\sum_{k=0}^{n}C_n^k\frac{\Gamma(n+k+y)}{\Gamma(n+k)\Gamma(y+1)}(1+\varphi)^{-(n+k+y)} \\
&= \left(\frac{\varphi^2}{1+\varphi}\right)^n\sum_{k=0}^{n}C_n^kC_{n+k+y-1}^{y}(1+\varphi)^{-(n+k+y)}.
\end{aligned}$$

■

## 5. Maximum likelihood estimation and its asymptotic property

In general, the transition probability can be written down explicitly as

$$\begin{aligned}
P(X_{t+1} = i|X_t = j) &= \sum_{m=0}^{\min(i,j)} C_j^m p_1^m(1-p_1)^{j-m}P(Y_{t+1}+Z_{t+1} = i-m) \\
&= \sum_{m=0}^{\min(i,j)}\sum_{x=0}^{i-m} F_{p_1}(m,j)f_\varphi(x,j)F_{\lambda_1}(i-m-x), \\
F_{p_1}(m,j) &= C_j^m p_1^m(1-p_1)^{j-m}, \\
f_\varphi(x,j) &= \int_{\mathbb{R}^+}\frac{e^{-s}s^x}{x!}g_j(s\mid\varphi)\,ds, \\
F_{\lambda_1}(i-m-x) &= \frac{e^{-\lambda_1}\lambda_1^{i-m-x}}{(i-m-x)!}.
\end{aligned} \tag{21}$$

The log likelihood function is simply $\ell(p_1,\varphi,\alpha) = \sum_{t=0}^{n-1}\log P(X_{t+1}|X_t)$.

**Proposition 5.1:** *Suppose we have a random sample $\{X_1, X_2, \ldots, X_n\}$. Let $\mathbf{p} = (p_1, \varphi, \lambda_1)$ denote the parameters vector for the stationary BMP INAR(1) model. The maximum likelihood estimator $\hat{\mathbf{p}}$ has the following asymptotic distribution:*

$$\sqrt{n}(\hat{\mathbf{p}} - \mathbf{p}) \sim N(0, \mathbf{I}^{-1}), \tag{22}$$

*where*

$$\mathbf{H} = \begin{Bmatrix} \ell_{p_1p_1} & \ell_{p_1\varphi} & \ell_{p_1\lambda_1} \\ \ell_{\varphi p_1} & \ell_{\varphi\varphi} & \ell_{\varphi\lambda_1} \\ \ell_{\lambda_1p_1} & \ell_{p_1\varphi} & \ell_{\lambda_1\lambda_1} \end{Bmatrix}, \quad \mathbf{I} = -\mathbb{E}[\mathbf{H}], \tag{23}$$

$$
\begin{aligned}
\frac{\partial P(X_{t+1}|X_t)}{\partial p_1} &= \sum_{m=0}^{\min(X_{t+1},X_t)} \sum_{x=0}^{X_{t+1}-m} \frac{\partial F_{p_1}(m,X_t)}{\partial p_1} f_\varphi(x,X_t) F_{\lambda_1}(X_{t+1}-m-x), \\
\frac{\partial^2 P(X_{t+1}|X_t)}{\partial (p_1)^2} &= \sum_{m=0}^{\min(X_{t+1},X_t)} \sum_{x=0}^{X_{t+1}-m} \frac{\partial^2 F_{p_1}(m,X_t)}{\partial (p_1)^2} f_\varphi(x,X_t) F_{\lambda_1}(X_{t+1}-m-x), \\
\frac{\partial^2 P(X_{t+1}|X_t)}{\partial p_1 \partial \varphi} &= \sum_{m=0}^{\min(X_{t+1},X_t)} \sum_{x=0}^{X_{t+1}-m} \frac{\partial F_{p_1}(m,X_t)}{\partial p_1} \frac{\partial f_\varphi(x,X_t)}{\partial \varphi} F_{\lambda_1}(X_{t+1}-m-x), \\
\ell_{xy} &= \sum_{t=0}^{T-1} \frac{\partial^2 P(X_{t+1}|X_t)}{\partial x \partial y} \frac{1}{P(X_{t+1}|X_t)} - \frac{\partial P(X_{t+1}|X_t)}{\partial x} \frac{\partial P(X_{t+1}|X_t)}{\partial y} \frac{1}{P(X_{t+1}|X_t)^2},
\end{aligned} \tag{24}
$$

*where $x, y \in \{p_1, \varphi, \lambda_1\}$. The first and second derivatives of each distribution function is given by*

$$
\begin{aligned}
\frac{\partial F_{p_1}(m,X_t)}{\partial p_1} &= \frac{m - p_1 X_t}{p_1(1-p_1)} F_{p_1}(m,X_t), \\
\frac{\partial f_\varphi(m,X_t)}{\partial \varphi} &= \frac{\partial}{\partial \varphi} \int_{\mathbb{R}^+} \frac{e^{-s} s^x}{x!} g_{X_t}(s\,|\,\varphi)\, ds, \\
\frac{\partial F_{\lambda_1}(m)}{\partial \lambda_1} &= \left(\frac{m}{\lambda_1} - 1\right) F_{\lambda_1}(m) \\
\frac{\partial^2 F_{p_1}(m,X_t)}{\partial (p_1)^2} &= \left(\frac{m(m-1-(X_t-1)p_1)}{p_1^2(1-p_1)} - \frac{(X_t-m)(m-(X_t-1)p_1)}{p_1(1-p_1)^2}\right) F_{p_1}(m,X_t), \\
\frac{\partial^2 f_\varphi(m,X_t)}{\partial \varphi^2} &= \frac{\partial^2}{\partial \varphi^2} \int_{\mathbb{R}^+} \frac{e^{-s} s^x}{x!} g_{X_t}(s\,|\,\varphi)\, ds, \\
\frac{\partial^2 F_{\lambda_1}(x)}{\partial (\lambda_1)^2} &= \left(1 - \frac{2x}{\lambda_1} + \frac{x(x-1)}{\lambda_1^2}\right) F_{\lambda_1}(x).
\end{aligned}
$$

***Proof:*** From Proposition 3.3, we know that the $X_n$ is stationary and ergodic and its stationary distribution is characterized by the p.g.f $\Phi_x(\theta) = \prod_{i=0}^{\infty} \Phi_z(f_i(\theta))$. Then score functions and information matrix **I** are also stationary and ergodic. Then the proof for asymptotic normality is similar to the proof of Theorem 4 in Appendix A of [3]. ■

The expectation of information matrix **I** can be calculated numerically by finding out unconditional distribution $P(X_t)$ and joint distribution $P(X_{t-1}, X_t)$. However, this would be computational intensive when the sample size $n$ is large. In practice, since the process $X_t$ is stationary and ergodic, $\mathbf{I} \approx -\mathbf{H}$ when $n$ is large.

To verify the asymptotic normality of the maximum likelihood estimators, we conduct a Monte Carlo experiment. This experiment is based on 2000 replications. For each replication, a time series of BMP-INAR(1) with chosen mixing density, either Exponential or Lindley, of size $n = 100, 200, \ldots, 500$ is generated. The parameters are set as $p_1 = \varphi =$

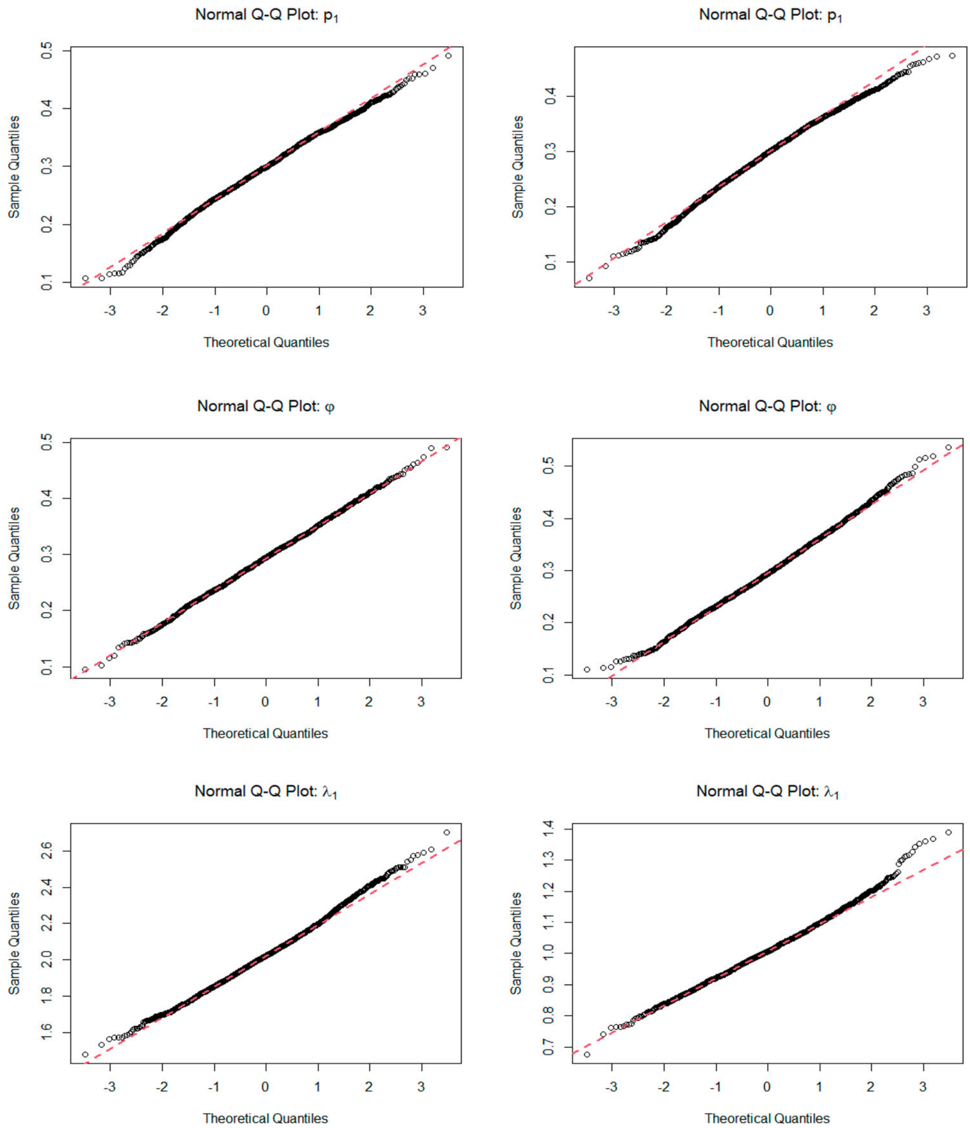


**Figure 1.** Quantile–Quantile plots for maximum likelihood estimators of BMP-INAR(1) model. The left panel shows plots for the Exponential mixing density, while the right panel depicts the plots for the Lindley mixing density.

$0.3, \lambda_1 = 2$ for both mixing densities and they are estimated via the maximum likelihood method. The biases and standard errors of the estimated parameters are shown in Tables 1 and 2. We observe that the biases of the estimators are either reasonably small or decreasing with respect to the sample size $n$. And it is clear that the standard error is also decreasing with respect to $n$. Finally, in order to graphically inspect the distribution of estimators, normal quantile-quantile plots are provided in Figure 1.

**Table 1.** The bias of Maximum likelihood estimators of BMP-INAR(1) model with respect to different sample size $n$.

| | Bias($\hat{\mathbf{p}}$) | $n = 100$ | $n = 200$ | $n = 300$ | $n = 400$ | $n = 500$ |
|---|---|---|---|---|---|---|
| Exponential | $p_1$ | 0.0022 | −0.0021 | 0.0019 | −0.0003 | −0.0003 |
| | $\phi$ | −0.0284 | −0.0104 | −0.0110 | −0.0072 | −0.0059 |
| | $\lambda_1$ | 0.1089 | 0.0526 | 0.0384 | 0.0366 | 0.0279 |
| Lindley | $p_1$ | −0.0008 | 0.0004 | −0.0015 | −0.0020 | −0.0011 |
| | $\phi$ | −0.0209 | −0.0143 | −0.0085 | −0.0050 | −0.0039 |
| | $\lambda_1$ | 0.0387 | 0.0227 | 0.0141 | 0.0144 | 0.0101 |

**Table 2.** The standard error of Maximum likelihood estimators of BMP-INAR(1) model with respect to different sample size $n$.

| | S.E.($\hat{\mathbf{p}}$) | $n = 100$ | $n = 200$ | $n = 300$ | $n = 400$ | $n = 500$ |
|---|---|---|---|---|---|---|
| Exponential | $p_1$ | 0.1303 | 0.0965 | 0.0752 | 0.0663 | 0.0576 |
| | $\phi$ | 0.1384 | 0.0970 | 0.0783 | 0.0670 | 0.0581 |
| | $\lambda_1$ | 0.3982 | 0.2858 | 0.2276 | 0.2012 | 0.1764 |
| Lindley | $p_1$ | 0.1319 | 0.0991 | 0.0854 | 0.0711 | 0.0630 |
| | $\phi$ | 0.1432 | 0.1054 | 0.0880 | 0.0729 | 0.0661 |
| | $\lambda_1$ | 0.2050 | 0.1515 | 0.1166 | 0.0999 | 0.0911 |

**Table 3.** Descriptive statistics of iceberg count.

| Minimum | Maximum | Median | Mean | Variance | FI |
|---|---|---|---|---|---|
| 0 | 9 | 1 | 1.407 | 2.184 | 1.552 |

## 6. Real data example: iceberg order data

The iceberg order counts concern the Deutsche Telekom shares traded in the XETRA system of Deutsche Börse, and the concrete time series gives the number of iceberg orders (for the ask side) per 20 min for 32 consecutive trading days in the first quarter of 2004. The special feature of iceberg orders is that only a small part of the order (tip of the iceberg) is visible in the order book and the main part of the order is hidden. For detail description, please see the [4,5]. This dataset is also analysed in [13], where the Extended Poisson INAR(1) is applied to fit the data.

A table of descriptive statistics, a time series, as well as the ACF and PACF plots are shown in Table 3 and Figure 2. The variance of the iceberg count is higher than its mean, which indicates the data is overdispersed. The level of dispersion is described by the Fisher index of dispersion FI $> 1$. Evidence of the applicability of a first-order autoregressive model is indicated by the empirical ACF and PACF graphs. They illustrate a clear decay for ACF and cut-off at lag $= 1$ for PACF.

The likelihood function is constructed as in (21) with different $f_\varphi(x, j)$ (mixed by Exponential or Lindley). It is then maximized through 'optim' in R with 'method $=$ BFGS' (quasi-Newton method) while the standard deviations of MLEs are calculated through inverting the negative observed information matrix in Proposition 5.1 based on MLEs. To access the goodness of fit, we adopt the information criteria AIC and BIC as well as

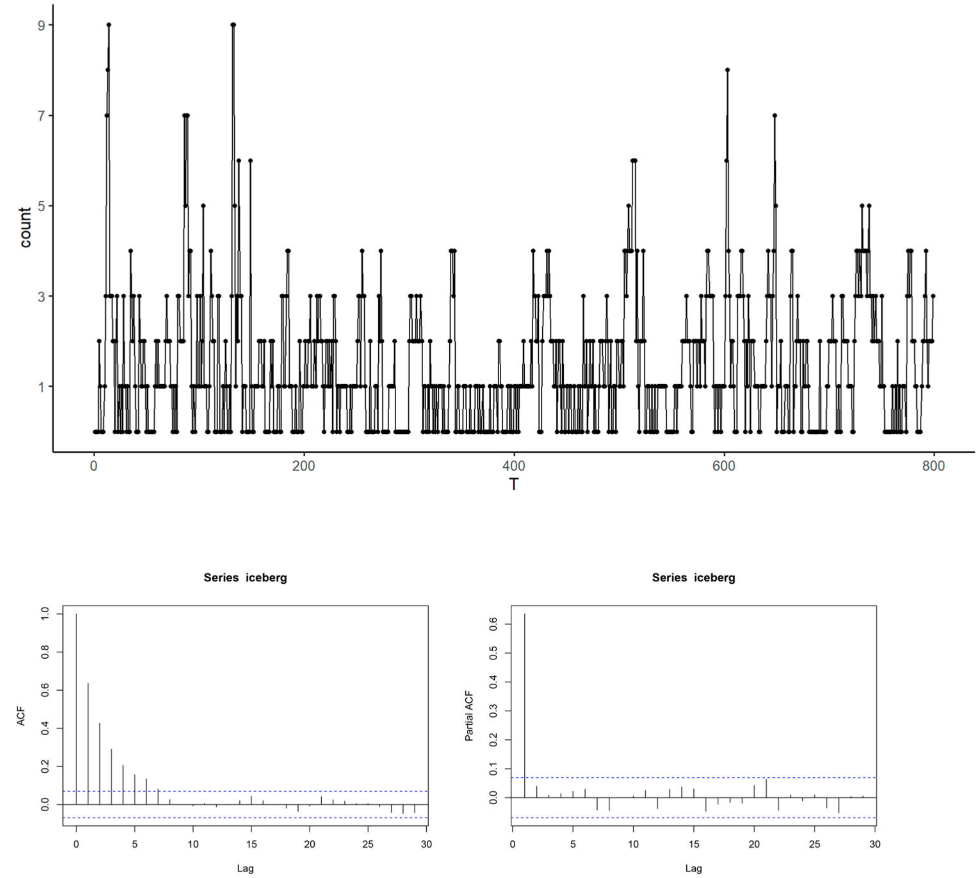


**Figure 2.** Time series plot of iceberg data and its empirical ACF PACF plots. The dash line lines are the 95% confident bands by assuming the series to be a white noise process.

the (standardized) Pearson residuals. If the model is correctly specified, Pearson residuals for BMP-INAR(1) are expected to have a mean and variance close to 0 and 1, respectively, with no significant autocorrelation. The Pearson residuals are calculated by the following formula:

$$e_t = \frac{x_t - \mathbb{E}[X_t \mid x_{t-1}]}{\sqrt{\mathrm{Var}(X_t \mid x_{t-1})}}, \tag{25}$$

where $x_t$ denotes the observed value.

The ACF plots of the Pearson residuals in Figure 3 indicate that the BMP-INAR(1) models are appropriate for fitting the iceberg data. The estimated parameters shown in Table 4 are significantly different from 0, which is indicated by their estimated standard deviation. Compared to the Dirac delta case, which is actually the Extended Poisson INAR(1) of [13], the other two cases do show some improvement with smaller AIC, BIC values and larger fitted Fisher index of dispersion $\hat{\mathrm{FI}}_x$ which, however, is slightly smaller than the empirical FI. On the other hand, it seems that there is little difference between the other

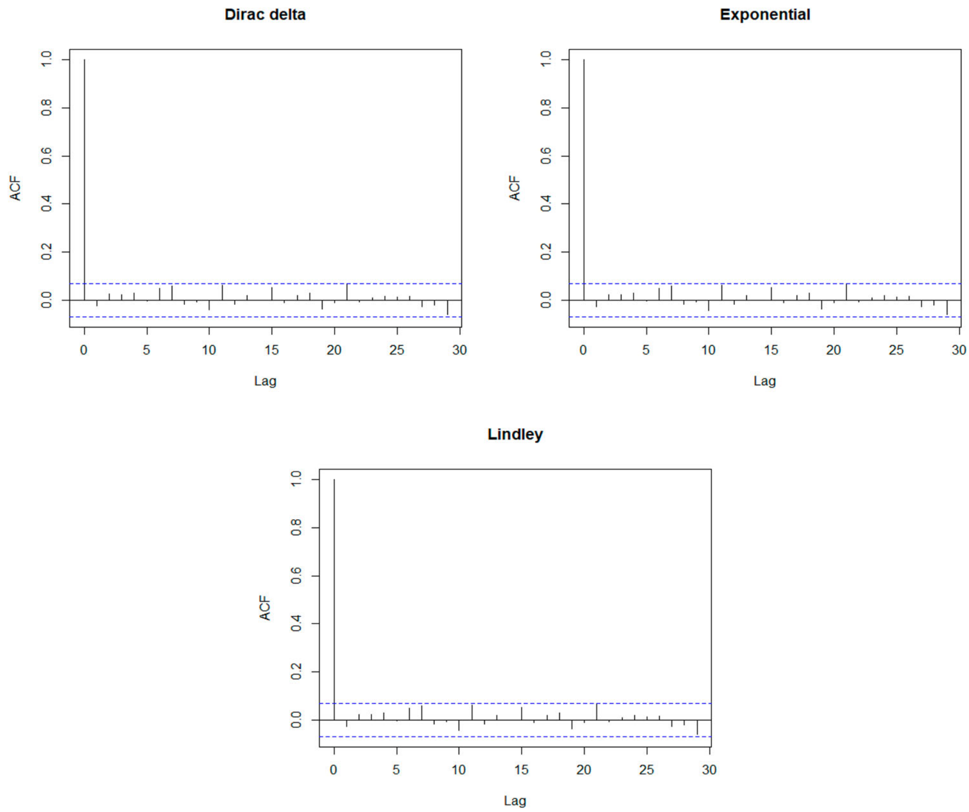


**Figure 3.** Autocorrelation of Standardized Pearson residuals for three different mixing densities.

**Table 4.** The results for the BMP INAR(1) model mixed by different density functions.

| | | | | | | Pearson residuals | | |
|---|---|---|---|---|---|---|---|---|
| Mixing density | $\hat{p}_1$ | $\hat{\varphi}$ | $\hat{\lambda}_1$ | AIC | BIC | Mean | Variance | $\widehat{\mathrm{FI}}_x$ |
| Dirac delta | 0.410 (0.058) | 0.188 (0.059) | 0.567 (0.040) | 2212 | 2226 | −0.001 | 1.159 | 1.295 |
| Exponential | 0.434 (0.044) | 0.167 (0.044) | 0.563 (0.040) | 2208 | 2222 | −0.002 | 1.154 | 1.315 |
| Lindley | 0.434 (0.043) | 0.167 (0.043) | 0.563 (0.040) | 2208 | 2222 | −0.002 | 1.154 | 1.314 |

Note: The results of Dirac delta case are from Table 2 of [13]. The estimated standard deviations for all models are in brackets.

two cases as they have very similar AIC and BIC values. This is due to the fact that the value of $\hat{\varphi}$ is identical for both densities. Finally, it should be noted that the variance of the Pearson residuals is visibly larger than 1. As it was previously mentioned, the exponential and Lindley mixing densities were considered for expository purposes. Therefore, since the proposed family of BMP INAR(1) models is quite general, another mixing distribution could potentially more efficiently capture the observed dispersion structure for this data.

Overall, the mixed Poison component in the BMP INAR(1) model efficiently captures the overdispersion in this type of financial data.

## 7. Concluding remarks

The BMP INAR(1) is an extension of the classical Poisson INAR(1) model obtained by adding an additional mixed Poisson component and hence it can capture the level of overdispersion coming from the data. The exponential family is a desired choice for the mixing density due to its 'additivity' property. The choice of the mixing density can control the dispersion level to some extent, although the BMP INAR(1) $X_t$ is always overdispersed in general. Furthermore, due to its simplicity, $X_t$ is actually a Markov chain and the maximum likelihood estimation method can be applied easily. The real data analysis shows that BMP INAR(1) can be a potential choice for modelling financial count data that exhibit standard AR(1) structure and overdispersion.

## Acknowledgments

The authors would like to thank the anonymous referee for their very helpful comments and suggestions which have significantly improved this article and would like to thank Prof Christian Wei for kindly sharing the financial count data.

## Disclosure statement

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

## ORCID

*Zezhun Chen* http://orcid.org/0000-0002-2523-9056

## References

[1] M. Al-Osh and A.A. Alzaid, *First-order integer-valued autoregressive (INAR (1)) process*, J. Time Ser. Anal. 8 (1987), pp. 261–275.
[2] M. Bourguignon, J. Rodrigues, and M. Santos-Neto, *Extended Poisson INAR (1) processes with equidispersion, underdispersion and overdispersion*, J. Appl. Stat. 46 (2019), pp. 101–118.
[3] R. Bu, B. McCabe, and K. Hadri, *Maximum likelihood estimation of higher-order integer-valued autoregressive processes*, J. Time Ser. Anal. 29 (2008), pp. 973–994.
[4] S. Frey and P. Sandås, *The impact of iceberg orders in limit order books*, in *AFA 2009 San Francisco Meetings Paper*, 2009.
[5] R.C. Jung and A. Tremayne, *Useful models for time series of counts or simply wrong ones?* Adv. Stat. Anal., 95 (2011), pp. 59–91.
[6] D. Karlis, *EM algorithm for mixed poisson and other discrete distributions*, ASTIN Bull. J. IAA 35 (2005), pp. 3–24.
[7] M. Kirchner, *Hawkes and INAR($\infty$) processes*, Stoch. Process. Appl. 126 (2016), pp. 2494–2525.
[8] E. McKenzie, *Some simple models for discrete variate time series*, J. Am. Water Resour. Assoc. 21 (1985), pp. 645–650.
[9] E. McKenzie, *Autoregressive moving-average processes with negative-binomial and geometric marginal distributions*, Adv. Appl. Probab. 18 (1986), pp. 679–705.
[10] 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.

[11] M.G. Scotto, C.H. Weiß, and S. Gouveia, *Thinning-based models in the analysis of integer-valued time series: A review*, Stat. Model. 15 (2015), pp. 590–618.
[12] C.H. Weiß, *Thinning operations for modeling time series of counts a survey*, Adv. Stat. Anal. 92 (2008), pp. 319–341.
[13] C.H. Weiß, *A poisson INAR(1) model with serially dependent innovations*, Metrika 78 (2015), pp. 829–851.
[14] H. Zheng, I.V. Basawa, and S. Datta, *First-order random coefficient integer-valued autoregressive processes*, J. Stat. Plan. Inference 137 (2007), pp. 212–229.