Afrika Statistika

**Afrika Statistika**
Vol. 19 (1), 2024, pages 3797 - 3813.
DOI: https://dx.doi.org/10.16929/as/2024.3797.328

ISSN 2316-090X

# A New Extension of Thinning-Based Random Coefficient Integer-Valued Autoregressive Models for Count Data

**Marcel Ilboudo** [1,*], **Ouindllassida Jean-Etienne Ouédrago** [1] **and Tégawendé Martin Kaboré** [1]

[1] L@MIA laboratory, Université Norbert Zongo

Received on December 14, 2024; Accepted on January 19, 2025; Published online on March 14, 2025



**Abstract.** The aims of this paper is to deal with overdispersed integer time series. Thus, new thinning called Discrete Lindley Distribution (*DLD*) thinning was build. Based on the proposed thinning operator, a new Random Coefficient Integer-Valued Autoregressive (*RCINAR*) model is formulated and its statistical properties were established. Parameters estimation were carried out via the Yule-Walker, conditional least squares and conditional maximum likelihood methods. Properties of proposed estimators were studies through Monte Carlo simulation. Then, to compare the proposed process to existing models, a real data set is used. Finally, the model is used to fit weekly data of meningitis cases in Burkina Faso.

**Key words:** integer auto-regressive process; random coefficient integer-valued autoregressive; Discrete Lindley Distribution; innovation and thinning; Poisson distribution.
**AMS 2010 Mathematics Subject Classification Objects :** 62M10; 62F10; 60J05.

---

(*) Corresponding author: Marcel Ilboudo(imarcelilboudo@yahoo.fr)
Ouindllassida Jean-Etienne Ouédrago: ouedraogoetienne@yahoo.fr
Tégawendé Martin Kaboré: tzorghoe@gmail.com

**Résumé** (Abstract in French) Le présent article propose un modèle permettant de modéliser les series temporelles à valeurs entières et surdispersées. Pour ce faire, un nouvel type d'amincissement dénommé amincissement Distribution discrète de Lindley Distribution a été construit. Utilisant ce nouveau amincissement, un nouveau modèle de serie temporelles à valeurs entières et à coefficient variable a été formulé et ses propriétés statistiques ont été établies. Les paramètres ont été estimés par les méthodes de Yule-Walker, des moindres carrés conditionnelles et du maximum de vraissemblance conditiontionnelle. Les propriétés statistiques des estimateurs ont été étudiées à travers une simulation de Monte-Carlo. Pour comparer le modèle proposé aux modèles déja existants, des données rélles ont été utilisées. Enfin, le modèle proposé est utilisé pour modéliser les cas hebdomadaires de la méningite dans une région au Burkina Faso.

**Presentations of authors.**

**Marcel ILBOUDO**, M.Sc., is preparing a Ph.D thesis at: Doctoral Studies, L@MIA, Université Norbert Zongo, Koudougou, BURKINA FASO, under the supervision of Ouindllassida Jean-Etienne Ouédraogo.

**Ouindllassida Jean-Etienne Ouédraogo**, Ph.D., is a professor of Statistics and Mathematics at Université Norbert Zongo, Koudougou, BURKINA FASO.

**Tégawendé Martin Kaboré**, M.Sc., is preparing a Ph.D thesis at: Doctoral Studies, L@MIA, Université Norbert ZONGO, Koudougou, BURKINA FASO, under the supervision of Ouindllassida Jean-Etienne Ouédraogo.

## 1. Introduction

Integer time series are widely used to model count data such as the number of patients still present in a hospital, the number of tourists present in a territory, the number of births recorded monthly or annually, etc. Existing classical models such as *ARMA* and *ARIMA* models often do not give accurate estimates when dealing with integer numbers. To obtain better accuracy in modeling discrete time series Mohamed A Al-Osh *et al.* (1987) and Ed McKenzie (1985) proposed the first-order integer auto-regressive process [*INAR(1)*] defined by equation (1):

$$X_t = \beta \circ X_{t-1} + \varepsilon_t, \tag{1}$$

with $\beta$ a constant over time and " $\circ$ " is the binomial thinning operator proposed by Fred W Steutel *et al.* (1979) and define by equation (2):

$$\beta \circ X_{t-1} = \begin{cases} \Sigma_{i=1}^{X_{t-1}} Y_i & , \; X_{t-1} > 0 \\ 0 & , \; X_{t-1} = 0 \end{cases} , \tag{2}$$

where $Y_i$ are a sequence of independent and identically distributed binomial random variables with distribution $B(1, \beta)$. The error term $\varepsilon_t$ also called innovation is a sequence i.i.d of non-negative integer value random variable. This error term denotes the immigration of the process in the time interval $[t-1, t]$ with mean $\mu_\varepsilon$ and finite variance $\sigma_\varepsilon^2$.

Over the years, the generalized form of the model was studied by Harry Joe (1996), Dulce Gomes *et al.* (2009), Haitao Zheng *et al.* (2007) . The model thus obtained takes into account the fact that the coefficient $\beta$ is a random variable unlike the *INAR(1)* model where it is constant. It is more careful to consider the time-varying $\beta$ parameter in some situations. This is the case in the healthcare sector, where patient care conditions can be improved, particularly financial, technical and climatic conditions. These factors could have a considerable impact on the duration of a patient's recovery. The model which takes into account the change of parameters over time is called the first-order random coefficient integer-valued auto-regressive (*RCINAR)* process (see Haitao Zheng *et al.* (2007) and Christian H Weiß (2008) ).

As shown in Harry Joe (1996), if $\{X_t\}_{t\in\mathbb{Z}}$ is a non-negative integer random variable generated according to the *RCINAR(1)* model, then there exists a sequence of variables $\{\beta_t\}_{t\in\mathbb{Z}}$ i.i.d defined on $\mathbb{R}_+$ with finite variance and a non-negative discrete random variable $\varepsilon_t$ such that $\{X_t\}_{t\in\mathbb{Z}}$ verifying:

$$\begin{cases} X_t = \beta_t \circ^G X_{t-1} + \varepsilon_t \\ \beta_t \circ^G X_{t-1} | X_{t-1}, \beta_t \sim G(\mu, \sigma^2) \end{cases}, \tag{3}$$

where $G$ denotes a discrete random variable with distribution such that $\mu = \beta_t X_{t-1}$ and $\sigma^2 = \delta_t X_{t-1}$ are respectively the expectation and the variance of $(\beta_t \circ^G X_{t-1} | X_{t-1}, \beta_t)$ with $\delta_t = f(\beta_t, X_{t-1})$.

- For each instant $t$, the random variable $(\beta_t \circ^G X_{t-1} | \beta_t, X_{t-1})$ is independent of $X_{t-1-k}$, $\beta_{t-k}$ and $\varepsilon_{t-k}$, for all $k \in \mathbb{N}^*$;
- The random variable $X_t$, given $\beta_t$, is independent from $\varepsilon_{t+k}$ and $\beta_{t+k}$ for all $k > 0$.

Several examples of thinnings proposed in the literature can be seen as a particular case of the generalize *RCINAR* process. We can cite in particular Harry Joe (1996), Dulce Gomes *et al.* (2009), Manuel G Scotto *et al.* (2015) , Christian H Weiß (2008) and Robert C Jung *et al.* (2005) who proposed:

(i) binomial thinning when $G(\mu, \sigma^2) = B(X_{t-1}, \beta)$,

(ii) negative Binomial thinning when $G(\mu, \sigma^2) = NB(X_{t-1}, \frac{1}{1+\beta})$, (iii) Poisson thinning when $G(\mu, \sigma^2) = P(\beta X_{t-1})$, (iv) Geometric thinning when $G(\mu, \sigma^2)$ denotes a geometric distribution with parameter $\frac{1}{\beta X_{t-1}}$.

The main objective of this paper is to propose a new *RCINAR(1)* model with Poisson innovations based on a new thinning operator called DLD thinning. This new model denoted by *DLD-RCINAR(1)* is suitable for modeling non-negative integer valued time series with over-dispersion. The Discrete Lindley Distribution which probability mass function (*pmf*) is given by the equation (4) is obtained by discretization of the continuous Lindley distribution (see B Abebe *et al.* (2018)).

$$P(X=x)=(1-e^{-\theta})^2(1+x)e^{-\theta x}, x\in\mathbb{N} \tag{4}$$

The outline of the paper is as follow: In section (2), we present the new process with *DLD* thinning. Some estimation methods are presented in section (3). Section (4) is devoted to results of simulation. Comparaison of the behavior of the proposed model with existing models is carried out through data-sets in the literature. This is detailed in section (5). Illustration with real data of meningitis cases complete this section.

## 2. *RCINAR(1)* process with *DLD* thinning

The aim of this section is to present the new thinning operator based on the Lindley's Discrete Distribution.

### *2.1. DLD-RCINAR(1) model*

**Definition 1.** DLD thinning
Let $X$ be a discrete non-negative random variable and $\beta$ be a continuous random variable with support on $\mathbb{R}_+$. If the random variable $\beta\circ X$ is such that $(\beta\circ X|\beta, X)$ is $DLD(\theta)$ then, $\beta\circ X$ is called $DLD$ thinning and the notation " $\circ$ " denotes $DLD$ thinning operator.

After the above definition, some properties related to the moments of *DLD* thinning can be given as follows:

**Proposition 1.** *Let $X$ be a discrete non-negative random variable and $\beta\circ X$ obtained from $X$ by the $DLD$ thinning with parameter $\theta=\ln(\frac{2+\beta X}{\beta X})$. Then, the expectation and variance of the random variable $(\beta\circ X|\beta, X)$ are respectively, $\mu=\beta X$ and $\sigma^2=\frac{\beta(2+\beta X)X}{2}$.*

Proof:
If $T=(\beta\circ X|\beta, X)\sim DLD(\ln(\frac{2+\beta X}{\beta X}))$
Then, $P(T=k)=4(1+k)\frac{(\beta X)^k}{(2+\beta X)^{k+2}}, k=0,1,2,...$

$$\begin{aligned}
E(T) &= \sum_{k=0}^{\infty} 4k(1+k)\frac{(\beta X)^k}{(2+\beta X)^{k+2}} \\
&= \frac{4\beta X}{(2+\beta X)^3}\sum_{k=0}^{\infty} k(1+k)\frac{(\beta X)^{k-1}}{(2+\beta X)^{k-1}} \\
&= \frac{4\beta X}{(2+\beta X)^3}\frac{d^2}{dK^2}\left(\sum_{k=0}^{\infty} K^{k+1}\right) \qquad , K = \frac{\beta X}{2+\beta X} \\
&= \frac{4\beta X}{(2+\beta X)^3}\frac{d^2}{dK^2}\left(\frac{K}{1-K}\right) \\
&= \frac{4\beta X}{(2+\beta X)^3}\left(\frac{2}{(1-K)^3}\right) \\
&= \frac{4\beta X}{(2+\beta X)^3}\frac{2(2+\beta X)^3}{8} = \beta X
\end{aligned} \tag{5}$$

$$\begin{aligned}
E(T^2) &= \sum_{k=0}^{\infty} 4k^2(1+k)\frac{(\beta X)^k}{(2+\beta X)^{k+2}} \\
&= \sum_{k=0}^{\infty} 4k(1+k)\frac{(\beta X)^k}{(2+\beta X)^{k+2}} + \sum_{k=0}^{\infty} 4k(k-1)(1+k)\frac{(\beta X)^k}{(2+\beta X)^{k+2}} \\
&= \frac{4\beta X}{(2+\beta X)^3}\sum_{k=0}^{\infty} k(1+k)\frac{(\beta X)^{k-1}}{(2+\beta X)^{k-1}} + \frac{4(\beta X)^2}{(2+\beta X)^4}\sum_{k=0}^{\infty} k(k-1)(1+k)\frac{(\beta X)^{k-2}}{(2+\beta X)^{k-2}} \\
&= \frac{4\beta X}{(2+\beta X)^3}\frac{d^2}{dK^2}\left(\sum_{k=0}^{\infty} K^{k+1}\right) + \frac{4(\beta X)^2}{(2+\beta X)^4}\frac{d^3}{dK^3}\left(\sum_{k=0}^{\infty} K^{k+1}\right) \\
&= \frac{4\beta X}{(2+\beta X)^3}\frac{d^2}{dK^2}\left(\frac{K}{K-1}\right) + \frac{4(\beta X)^2}{(2+\beta X)^4}\frac{d^3}{dK^3}\left(\frac{K}{1-K}\right) \\
&= \frac{4\beta X}{(2+\beta X)^3}\left(\frac{2}{(1-K)^3}\right) + \frac{4(\beta X)^2}{(2+\beta X)^4}\left(\frac{6}{(1-K)^4}\right) \\
&= \beta X + \frac{3(\beta X)^2}{2}
\end{aligned} \tag{6}$$

Hence,

$$\begin{aligned}
V(T) &= \beta X + \frac{3(\beta X)^2}{2} - (\beta X)^2 \\
&= \beta X + \frac{(\beta X)^2}{2} = \frac{\beta(2+\beta X)X}{2}
\end{aligned} \tag{7}$$

**Definition 2.** *DLD-RCINAR(1)* process
Let $\{X_t\}_{t\in\mathbb{Z}}$ be a discrete positive-valued stochastic process. If $X_t$ is generated according to the *DLD-RCINAR(1)* model then there exists a sequence of random variables $\{\beta_t\}_{t\in\mathbb{Z}}$ i.i.d and uniformly distributed on $[0, 2\phi]$ with $E(\beta_t) = \phi$ and a discrete non-negative random variable $\varepsilon_t \sim Poisson(\lambda)$ such that $\{X_t\}_{t\in\mathbb{Z}}$ checks:

$$X_t = \begin{cases} \beta_t \circ X_{t-1} + \varepsilon_t & , \; X_{t-1} > 0 \\ \varepsilon_t & , \; X_{t-1} = 0 \end{cases} , \tag{8}$$

where " $\circ$ " denotes DLD thinning.

### 2.2. Statistical properties of the model

Let $\{X_t\}_{t\in\mathbb{Z}}$ be a discrete non-negative integer valued random variable generated according to the *DLD-RCINAR(1)* process. Then, the next properties are verified.

1. $\rho(k) = \phi^k$, with $\phi = E(\beta_t)$
2. $E(X_t|X_{t-1}) = \phi X_{t-1} + \lambda$ , $V(X_t|X_{t-1}) = \phi X_{t-1} + \phi^2 X_{t-1}^2 + \lambda$
3. $P(X_t = i|X_{t-1} = j) = \frac{e^{-\lambda}}{1+\phi j} \sum_{k=0}^{i} \frac{\lambda^{i-k}}{(i-k)!} [\frac{\phi j}{1+\phi j}]^k$
4. $E(X_t) = \frac{\lambda}{1-\phi}$ , $V(X_t) = \frac{\lambda(1-\phi)+\phi^2\lambda^2}{(1-\phi)^2(1-\frac{4\phi^2}{3})}$

The proofs of these properties can be found in the appendix.

Through the above properties, it can be seen that the *DLD-RCINAR(1)* process has a linear first order auto-regressive structure. Furthermore its correlation structure is analogous to the **INAR(1)** and the *AR(1)* processes. The major difference with the previous processes lies in the fact that the conditional variable $X_t|X_{t-1}$ has quadratic conditional variance. Another important result is given below.

**Proposition 2.** *If $(X_t)_{t\in\mathbb{Z}}$ is a process generated according to the DLD-RCINAR(1) process, then the random variable $T_t = (\beta_t \circ X_{t-1}|X_{t-1})$ follows a geometric distribution with parameter $p = \frac{1}{1+\phi X_{t-1}}$.*

Proof:

$$\begin{aligned} P(\beta_t \circ X_t = k|X_{t-1} = j) &= \int_0^{2\phi} 4(1+k)\frac{(\beta j)^k}{2\phi(2+\beta j)^{k+2}} d\beta \\ &= \frac{1}{j\phi}\left[ (\frac{\beta j}{2+\beta j})^{k+1} \right]_0^{2\phi} \\ &= \frac{1}{j\phi}\left(\frac{2\phi j}{2+2\phi j}\right)^{k+1} = \frac{(\phi j)^k}{(1+\phi j)^{k+1}} \end{aligned} \tag{9}$$

Based on the above result, simulating the proposed process can be carried out through geometric distribution with parameter $p = \frac{1}{1+\phi x_{t-1}}$ given $X_{t-1} = x_{t-1}$. Weakly stationarity of the proposed process is ensured under the following conditions.

**Proposition 3.** *Stationarity conditions*
*Let $\{\beta_t\}_{t\in\mathbb{Z}}$ be a sequence of random variables iid defined on $\mathbb{R}^*{}_+$. The DLD-RCINAR(1) model is stationary in the weak sense if $\phi < \frac{\sqrt{3}}{2}$.*

Proof:
According to Dulce Gomes *et al.* (2009) and Alain Latour (1998) the sufficient condition for a *RCINAR(1)* process to be stationary is that $E(\beta_t^2) < 1$.
Let $\beta_t \sim U_{[0,2\phi]}$
hence ,

$$\begin{aligned} E(\beta_t^2) &= \frac{\phi^2}{3} + \phi^2 \\ &= \frac{4\phi^2}{3} \end{aligned} \tag{10}$$

Thus, $E(\beta_t^2) < 1 \longrightarrow \phi < \frac{\sqrt{3}}{2}$

**Proposition 4.** *The DLD-RCINAR(1) model is suitable for modeling over-dispersed time series.*

Proof:
Let $\{X_t\}_{t\in\mathbb{Z}}$ be a discrete, non-negative integer valued random variable generated according to the *DLD-RCINAR(1)* process. By noting $ID$ the Fisher dispersion index, one has:

$$\begin{aligned} ID(X_t) &= \frac{V(X_t)}{E(X_t)} \\ &= \frac{\phi(1-\phi) + \phi^2\lambda + (1-\phi)^2}{(1-\phi)(1-\frac{4\phi^2}{3})} \end{aligned} \tag{11}$$

$$\begin{aligned} ID(X_t) - 1 &= \frac{\phi(1-\phi) + \phi^2\lambda + (1-\phi)^2 - (1-\phi)(1-\frac{4\phi^2}{3})}{(1-\phi)(1-\frac{4\phi^2}{3})} \\ &= \frac{\phi^2\lambda + \frac{4\phi^2}{3}(1-\phi)}{(1-\phi)(1-\frac{4\phi^2}{3})} \end{aligned} \tag{12}$$

Under the proposition (3) and for all strictly positive $\lambda$, $ID(X_t) > 1$.

One of the advantages of our model is the possibility of simulating the data in two different ways. The first option suppose that the conditional variable $(\beta_t \circ X_{t-1}|\beta_t, X_{t-1})$ follows a DLD distribution and for the second option we use the proposition (2) assuming the random variable $(\beta_t \circ X_{t-1}|X_{t-1})$ is a geometric random variable.

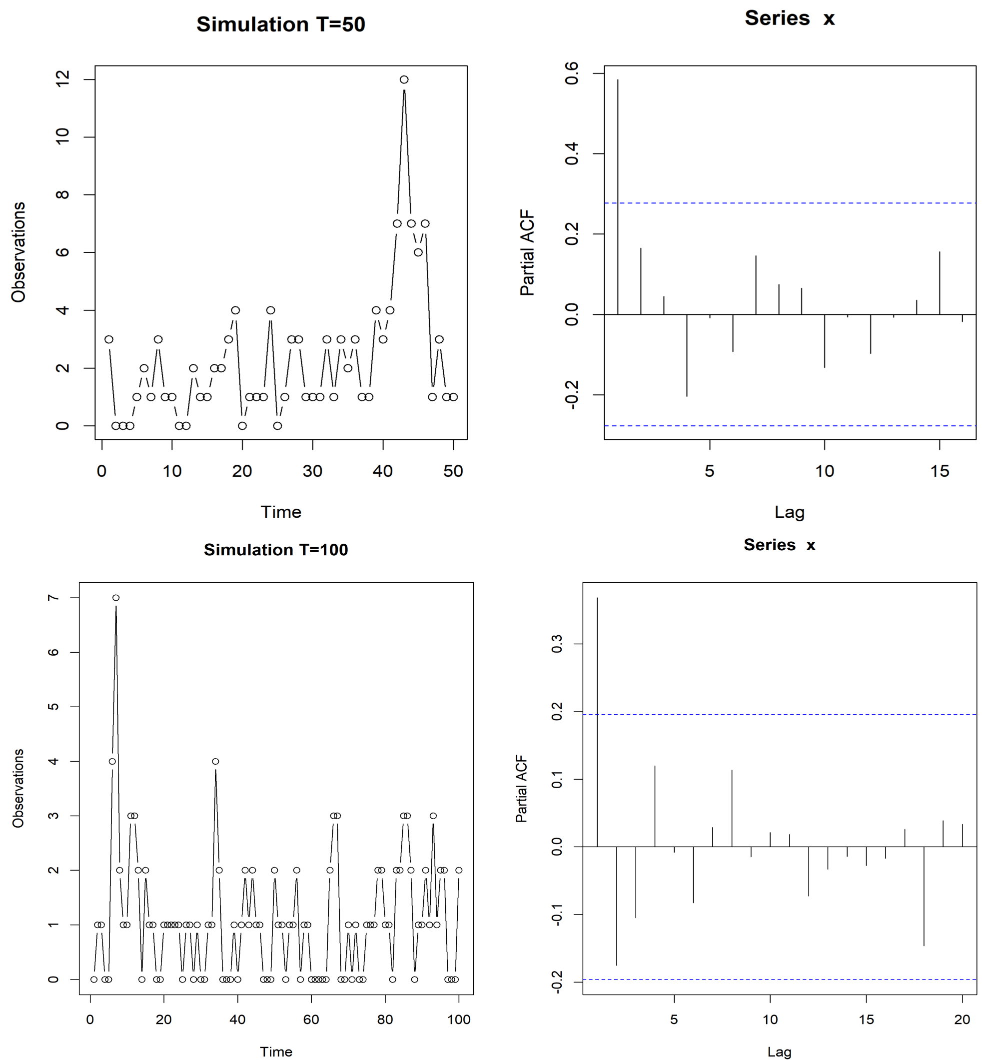


**Fig. 1.** Simulated series and its PACF graphs for T=50,100

The observations in Fig. 1) are generated according to the *DLD-RCINAR(1)* process whose parameters are $\lambda = 1$ and $\phi = 0.6$. Similar results are obtained by varying the sample size. Thus, samples of size T=50 and 100 are generated.

The graphs (1) suggest the stationarity of the process for both medium and large series. Furthermore the PACF indicates that this is an auto-regressive model of order 1.

After describing our model, we will discuss about the different methods for the parameters estimation in the following section.

## 3. Estimation of unknown parameters

The model parameters are estimated by the Yule-Walker estimation method, the conditional least squares method and the conditional maximum likelihood method.

Let $X = (X_1, X_2, \cdots, X_T)$ be a sample of size $T$ and $\theta = (\phi, \lambda)$ the vector of parameters.

### 3.1. Yule-Walker (YW) method

The Yule-Walker estimators are obtained by considering the auto-covariance and the first moment of $X_t$ defined by $Cov(X_t, X_{t-1}) = \phi$ and $E(X_t) = \frac{\lambda}{1-\phi}$, respectively. Using the random sample proposed before $\{X_t\}_{t=1}^{N}$, the Yule-Walker estimators $\hat{\theta}_{YW} = (\hat{\phi}_{YW}, \hat{\lambda}_{YW})^T$ of $\theta_{YW} = (\phi_{YW}, \lambda_{YW})^T$ are given by:

$$\begin{cases} \hat{\phi}_{YW} &= \frac{\sum_{t=2}^{T}(X_t-\bar{X})(X_{t-1}-\bar{X})}{\sum_{t=2}^{T}(X_t-\bar{X})^2} \\ \hat{\lambda}_{YW} &= (1-\hat{\phi}_{YW})\bar{X} \end{cases}, \bar{X} = \frac{1}{T}\sum_{t=1}^{T} X_t \tag{13}$$

### 3.2. Conditional Least Square (CLS) method

Parameters estimates by the conditional least squares method consists of minimizing $S(X, \theta) = Min \sum_{t=2}^{T}(X_t - E(X_t|X_{t-1}))^2$ with $\hat{\theta}_{CLS} = (\hat{\phi}_{CLS}, \hat{\lambda}_{CLS})$.

$$\begin{aligned} S(X,\theta) &= Min \sum_{t=2}^{T}[X_t - E(X_t|X_{t-1})]^2 \\ &= Min \sum_{t=2}^{T}(X_t - \phi X_{t-1} - \lambda)^2, \end{aligned} \tag{14}$$

hence,

$$\begin{aligned} &\begin{cases} \frac{\partial S(X,\theta)}{\partial \phi} &= 0 \\ \frac{\partial S(X,\theta)}{\partial \lambda} &= 0 \end{cases} \Longrightarrow \begin{cases} -2\sum_{t=2}^{T} X_{t-1}(X_t - \phi X_{t-1} - \lambda) &= 0 \\ -2\sum_{t=2}^{T}(X_t - \phi X_{t-1} - \lambda) &= 0 \end{cases} \\ &\Longrightarrow \begin{cases} \hat{\phi}_{CLS} &= \frac{\sum_{t=2}^{T} X_t \sum_{t=2}^{T} X_{t-1} - (T-1)\sum_{t=2}^{T} X_t X_{t-1}}{(\sum_{t=2}^{T} X_{t-1})^2 - (T-1)\sum_{t=2}^{T} X_{t-1}^2} \\ \hat{\lambda}_{CLS} &= \frac{1}{T-1}(\sum_{t=2}^{T} X_t - \hat{\phi}_{CLS}\sum_{t=2}^{T} X_{t-1}) \end{cases} \end{aligned} \tag{15}$$

The asymptotic properties of CLS estimators were studied in Lawrence A Klimko *et al.* (1978).

### 3.3. Conditional Maximum Likelihood (CML) method

The likelihood function is given by the following relation:

$$L(X,\theta) = \prod_{t=2}^{T} P_\theta(X_t = x_t | X_{t-1} = x_{t-1}) \tag{16}$$

$$\begin{aligned} l(X,\theta) &= log(L(X,\theta)) \\ &= -\lambda T - \sum_{t=2}^{T} log(1+\phi x_{t-1}) + \sum_{t=2}^{T} log(\sum_{k=0}^{x_t} \frac{\lambda^{x_t-k}}{(x_t-k)!}[\frac{\phi x_{t-1}}{1+\phi x_{t-1}}]^k) \end{aligned} \tag{17}$$

The conditional maximum likelihood estimators $\hat{\theta}_{CML} = (\hat{\phi}_{CML}, \hat{\lambda}_{CML})$ are obtained by maximizing the log- likelihood $l(\theta)$.
Estimation of *CML* parameters based on a sample of data can be obtained using the Newton-Raphson algorithm Robert B Schnabel *et al.* (1985) with instruction "nlminb" in R software. The asymptotic properties of *CML* estimators were studied in Han Li *et al.* (2018).

The next section is devoted to the simulation and the comparison of the three estimation methods discribed previously.

## 4. Simulation studies

In this section, we will study the performances of the *YW*-estimators, CLS-estimators and *CML*-estimators.

### 4.1. Sampling disign

For generating data $\{X_t\}_{t=1}^{T}$ of model (8),we set the following various combinations of the model parameters $(\phi, \lambda)$: $(0.3, 1)$, $(0.3, 2)$ and $(0.5, 1)$.

We consider the sample size T=50,100,300,500,1000 and compute the empirical biases and mean squared errors (MSE) based on m=1000 replications for each parameter combination. The bias and the MSE of the estimators are given by: $Bias = \frac{1}{m}\sum_{k=1}^{m} \hat{\theta}_k - \theta$ and $MSE = \frac{1}{m}\sum_{k=1}^{m} (\hat{\theta}_k - \theta)^2$.

Using the Bias and the MSE of estimations, we make a comparison of the three methods performances (figures (2), (3) and (4)) and the estimators behavior illustrated through the figures (5) and (6).

### 4.2. Comparison of the three estimation methods

Through the simulation results given by figures (2), (3) and (4) we notice that for all three estimation methods, the bias and the MSE both decrease as the sample size

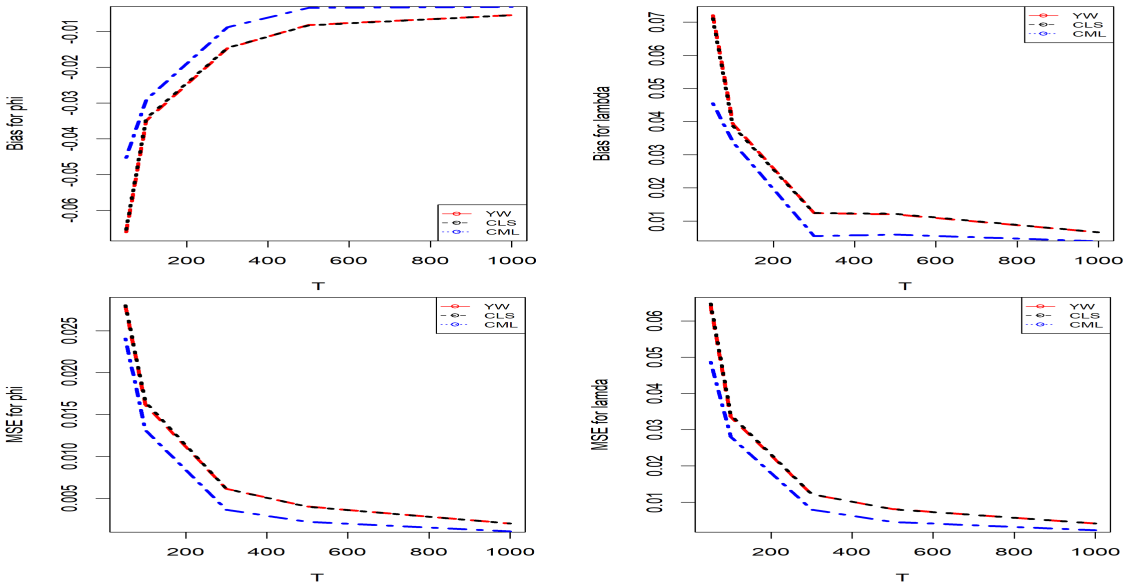


**Fig. 2.** Bias and MSE for different sample sizes with $\phi = 0.3$ and $\lambda = 1$

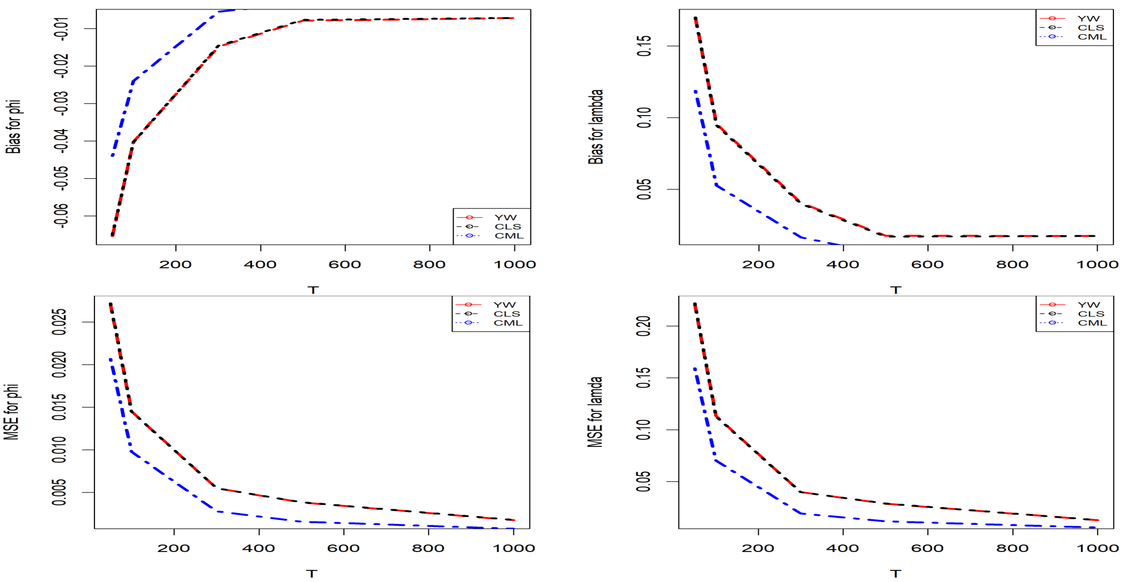


**Fig. 3.** Bias and MSE for different sample sizes with $\phi = 0.3$ and $\lambda = 2$

increases. In particular, the conditional Maximum Likelihood estimators converge faster than its competitors. Afterward, only the *CML* estimators will be considered.

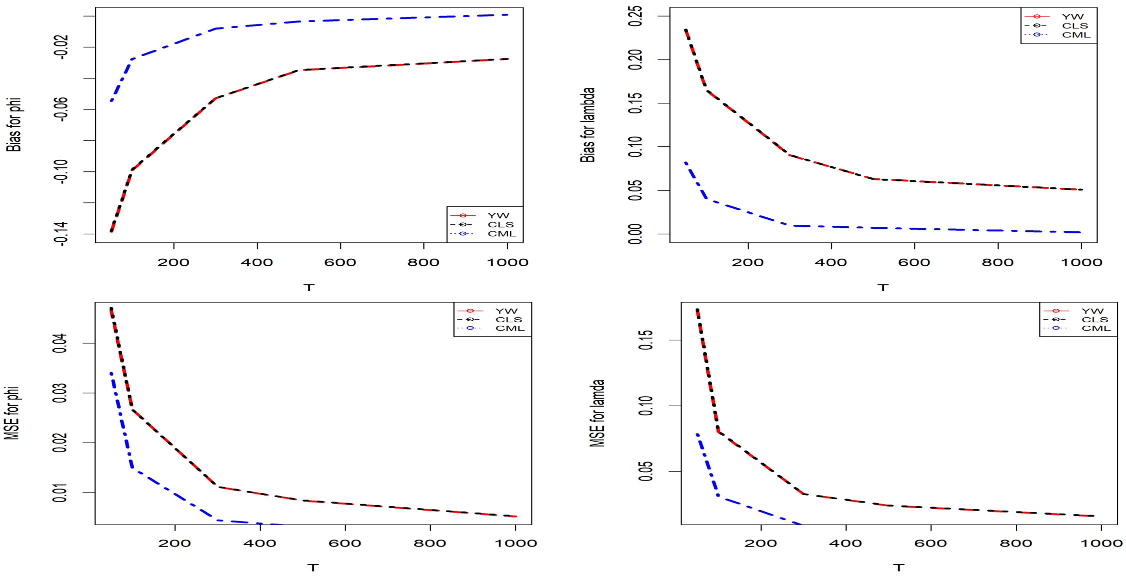


**Fig. 4.** Bias and MSE for different sample sizes with $\phi = 0.5$ and $\lambda = 1$

### *4.3. Behavior of CML estimators*

The aim is to analyze the behavior of estimators when they take on different values and for different sample sizes.

The analysis of the performance of the estimators shows that it improves as the sample size increases. Indeed, the bias and the MSE decrease as the sample size increases. Note that the bias for the parameters $\phi$ and $\lambda$ is negative and positive respectively. Proof that the *CML* method tends to underestimate the theoretical value of $\phi$ while the latter is overestimated.

In order to compare our model to some existing models, we use two data sets. The first is used to show the competitiveness of the proposed model and the second to applications to a real dataset in the case of Burkina Faso.

## 5. Application to real data

### *5.1. Soap sales data*

In this section, we consider the weekly sales series (in whole units) of particular soap products (figure (7)). These data are presented by Tito Lívio *et al.* (2018) and MM Gabr *et al.* (2024). The series has 242 observations whose mean and variance

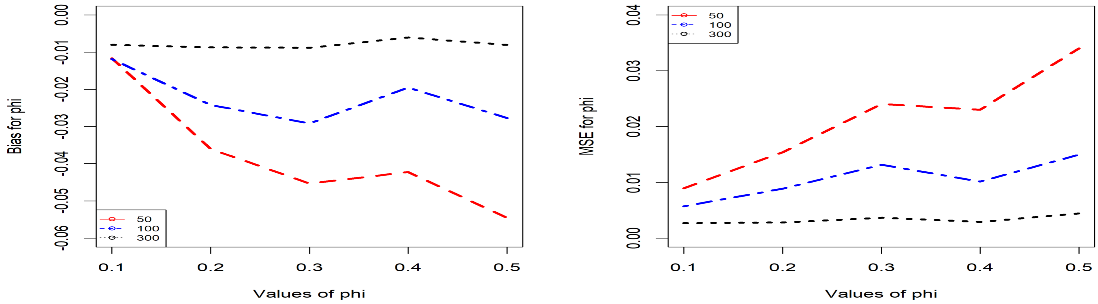


**Fig. 5.** Bias and MSE for different values of $\phi$

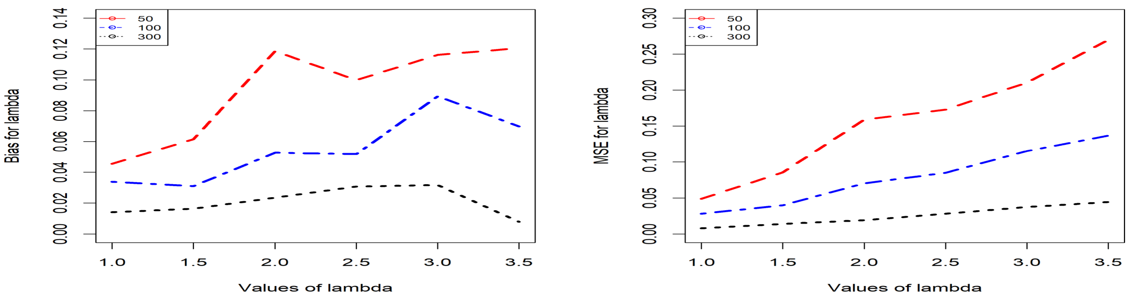


**Fig. 6.** Bias and MSE for different values of $\lambda$

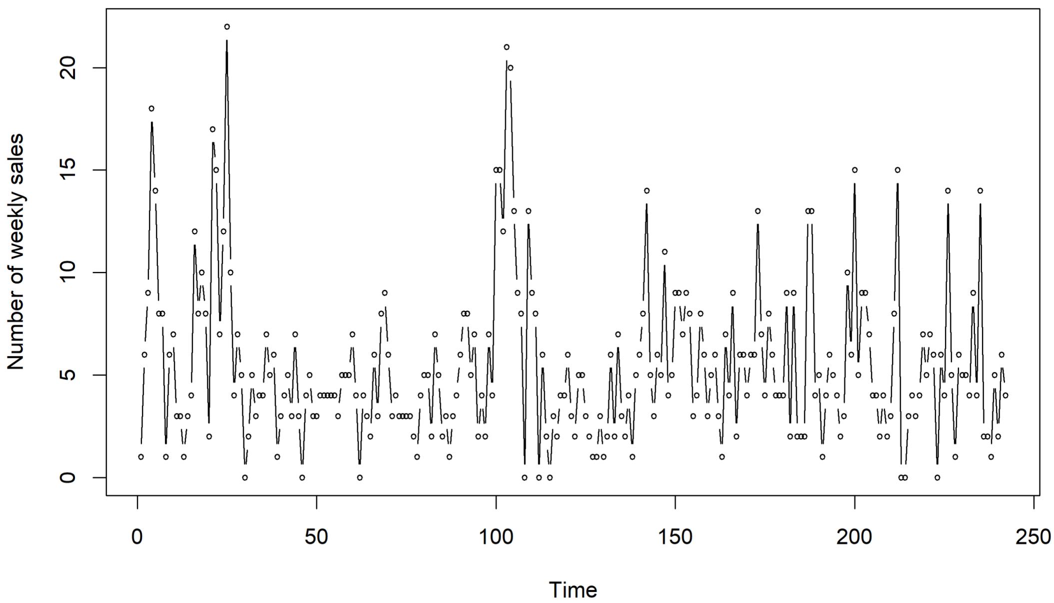


**Fig. 7.** Evolution of soap sales

| | Parameter 1 | Parameter 2 | AIC | BIC |
|---|---|---|---|---|
| ***DLD-RCINAR(1)*** ***($\phi$,$\lambda$)*** | ***0.4312*** | ***3.1737*** | ***1249.26*** | ***1256.24*** |
| *INARPL(1)* $(\alpha, \theta)$ | 0.3202 | 0.4533 | 1249.07 | 1256.05 |
| *PLINAR(1)* $(\alpha, \theta)$ | 0.3152 | 0,3516 | 1287.12 | 1294.10 |
| Poisson INAR(1) $(\alpha, \lambda)$ | 0.2340 | 4.1855 | 1363.81 | 1370.79 |
| NGINAR(1) $(\alpha, \mu)$ | 0.5770 | 4.6146 | 1296.43 | 1303.41 |

**Table 1.** Comparison of models

| | Observations | Mean | Variance | Index |
|---|---|---|---|---|
| Meningitis data | 260 | 5.4923 | 54.8918 | 9.9943 |

**Table 2.** Data description

are 5.44 and 15.40 respectively. It is therefore an over-dispersed series. The series is also stationary.

The *DLD-RCINAR(1)* model is compared to the *INARPL(1)* (Tito Lívio *et al.* (2018) ), *PLINAR(1)* (M Mohammadpour *et al.* (2018) ), *NGINAR(1)* (Miroslav M Risti´ c *et al.* (2009) ) models and with the Poisson *INAR(1)* model(Mohamed A Al-Osh *et al.* (1987)).

The best model is the one that minimizes the AIC and the BIC. We note that the proposed model presents substantially the same results as the *INARPL(1)* model (Table 1). These two models give better results than the other three models.

### *5.2. Meningitis data*

The data we used show the situation of meningitis cases (`https://www.who.int/fr/news-room/fact-sheets/detail/meningitis`) recorded per week in health facilities in the South-Central region of Burkina Faso from 2012 to 2016. The data was collected from the South-Central Regional Management of the Ministry of Health in 2017 (Epidemiological surveillance).

The PACF graph (figure (8)) indicates that a first-order auto-regressive model may be suitable for modeling data on meningitis cases in the South-Central region.

Results in Table 2 suggest that the data is over-dispersed. Furthermore, Dickey-Fuller stationarity test gives p-value = 0.01. Thus, time series is stationary. Parameter estimation was carried out using the conditional maximum likelihood (*CML*) method. The results are in Table 3.

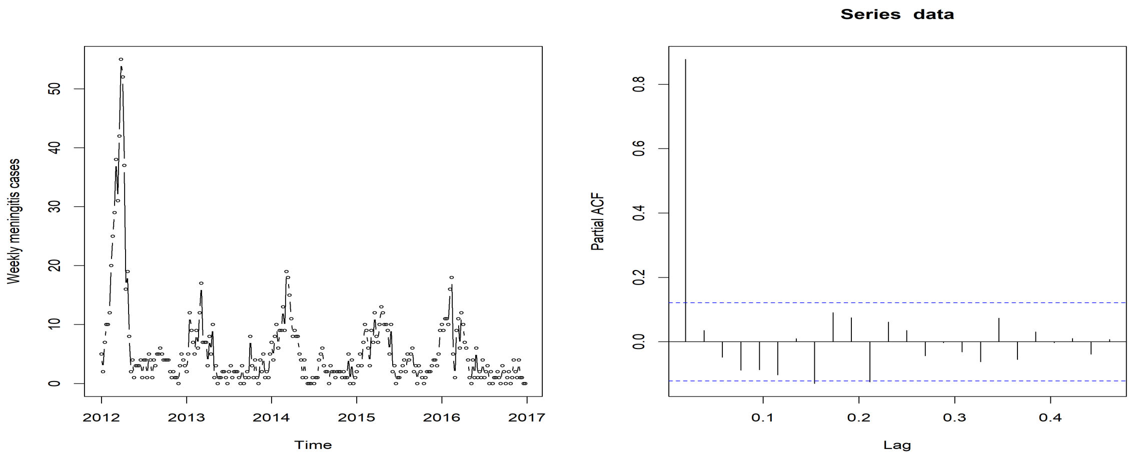


**Fig. 8.** Weekly cases of meningitis and its PACF graph

| $\phi$ | $\lambda$ | AIC | BIC |
|---|---|---|---|
| 0.6345 | 1.5422 | 1253.2730 | 1260.3944 |

**Table 3.** Parameters estimation

## 6. Conclusion

This article studies a new random coefficient integer-valued auto-regressive process by extending existing thinning operators. Classical estimations methods such as *YW*-estimators, modified **CLS**-estimators and *CML*-estimators of the model are used. In the simulation study, we compare these three methods. Its appear that the *CML* method is more efficient than others. This latter method was used to fit reals data-sets. The results of the proposed model are comparable to those of models usually used for the adjustment of over-dispersed series. Future of the proposed model is certainly to investigate on its applicability in other field like agriculture, tourism, education, etc. The challenge of proposing an efficient method for consistent forecasts is also current.

**Acknowledgments.** The authors wish to thank the referee and the associate editor for careful reading and constructive suggestions which led to this improved version of the paper.

## 7. Appendix

$$
\begin{aligned}
V(X_t) &= \frac{E(\delta_t)E(X_t) + V(\beta_t)E^2(X_t) + \sigma^2_{\varepsilon_t}}{1-(V(\beta_t)+E^2(\beta_t))} = \frac{E(\frac{\beta_t(2+\beta_t X_{t-1})}{2})\mu_X + \frac{\phi^2}{3}\mu_X^2 + \lambda}{1-(\frac{\phi^2}{3}+\phi^2)} \\
&= \frac{\phi\mu_X + \frac{E(\beta_t^2)\mu_X^2}{2} + \frac{\phi^2}{3}\mu_X^2 + \lambda}{1-\frac{4\phi^2}{3}} = \frac{\phi\mu_X + \frac{2\phi^2\mu_X^2}{3} + \frac{\phi^2}{3}\mu_X^2 + \lambda}{1-\frac{4\phi^2}{3}} \\
&= \frac{\phi\mu_X + \phi^2\mu_X^2 + \lambda}{1-\frac{4\phi^2}{3}} = \frac{\phi\frac{\lambda}{1-\phi} + \phi^2\left(\frac{\lambda}{1-\phi}\right)^2 + \lambda}{1-\frac{4\phi^2}{3}} \\
&= \frac{\lambda\phi(1-\phi) + \phi^2\lambda^2 + \lambda(1-\phi)^2}{(1-\phi)^2(1-\frac{4\phi^2}{3})} = \frac{\lambda(1-\phi) + \phi^2\lambda^2}{(1-\phi)^2(1-\frac{4\phi^2}{3})}
\end{aligned} \tag{18}
$$

$$
\begin{aligned}
V(X_t|X_{t-1}) &= E\left(\beta\frac{2+\beta X_{t-1}}{2}\right)X_{t-1} + V(\beta_t)X_{t-1}^2 + V(\varepsilon_t) \\
&= (\phi + \frac{2\phi^2}{3}X_{t-1})X_{t-1} + \frac{\phi^2}{3}X_{t-1}^2 + \lambda \\
&= \phi X_{t-1} + \phi^2 X_{t-1}^2 + \lambda
\end{aligned} \tag{19}
$$

$$
\begin{aligned}
P(X_t = i|X_{t-1} = j) &= \sum_{k=0}^{i} \frac{\lambda^{i-k}e^{-\lambda}}{(i-k)!}\int_0^{2\phi} 4(1+k)\frac{(\beta j)^k}{2\phi(2+\beta j)^{k+2}}d\beta \\
&= \sum_{k=0}^{i} \frac{\lambda^{i-k}e^{-\lambda}}{j\phi(i-k)!}\left[(\frac{\beta j}{2+\beta j})^{k+1}\right]_0^{2\phi} \\
&= \sum_{k=0}^{i} \frac{\lambda^{i-k}e^{-\lambda}}{j\phi(i-k)!}\left(\frac{2\phi j}{2+2\phi j}\right)^{k+1} \\
&= \sum_{k=0}^{i} \frac{\lambda^{i-k}e^{-\lambda}}{(i-k)!}\frac{(\phi j)^k}{(1+\phi j)^{k+1}} = \frac{e^{-\lambda}}{1+\phi j}\sum_{k=0}^{i}\frac{\lambda^{i-k}}{(i-k)!}[\frac{\phi j}{1+\phi j}]^k
\end{aligned} \tag{20}
$$

## References


Alain Latour (1998). Existence and stochastic structure of a non-negative integer-valued autoregressive process. Journal of Time Series Analysis, 19(4):439–455.

B Abebe and RA Shanker(2018). Discrete lindley distribution with applications in biological sciences. *Biom. Biostat.* Int. J, 7(1):48–52.

Christian H Weiß(2008). Serial dependence and regression of poisson inarma models. *Journal of Statistical Planning and Inference*, 138(10):2975–2990.

Dulce Gomes and Luísa Canto Castro (2009). Generalized integer-valued random coefficient for a first order structure autoregressive (rcinar) process. *Journal of Statistical Planning and Inference*, 139(12):4088–4097.

Ed McKenzie(1985). Some simple models for discrete variate time series 1. *JAWRA Journal of the American Water Resources Association*, 21(4):645–650 .

Fred W Steutel and Klaas van Harn(1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability*, pages 893–899 .

Haitao Zheng, Ishwar V Basawa, and Somnath Datta(2007) . First-order random coefficient integer-valued autoregressive processes. *Journal of Statistical Planning and Inference*2007,137(1):212–229.

Han Li, Kai Yang, Shishun Zhao, and Dehui Wang (2018). First-order random coefficients integer-valued threshold autoregressive processes. *AStA Advances in Statistical Analysis*, 102:305–331.

Harry Joe (1996). Time series models with univariate margins in the convolution-closed infinitely divisible class. *Journal of Applied Probability*, 33(3):664–677.

Lawrence A Klimko and Paul I Nelson (1978). On conditional least squares estimation for stochastic processes. *The Annals of statistics*, pages 629–642.

M Mohammadpour, Hassan SBakouch, and MShirozhan(2018). Poisson–lindley inar (1) model with applications. *Brazilian Journal of Probability and Statistics*, 32(2):262–280.

Manuel G Scotto, Christian H Weiss, and Sónia Gouveia(2015). Thinning-based models in the analysis of integer-valued time series: a review. *Statistical Modelling*, 15(6):590–618.

Miroslav M Risti´ c, Hassan S Bakouch, and Aleksandar S Nasti´ c (2009). A new geometric first order integer-valued autoregressive (nginar (1)) process. *Journal of Statistical Planning and Inference*, 139(7):2218–2226.

MM Gabr, Hassan S Bakouch, and Shimaa M El-Hadidy (2024). One-misrecorded poisson inar (1) model via two random operators with application to crime and economics data. *Computational and Applied Mathematics*, 43(4):206.

Mohamed A Al-Osh and Aus A Alzaid(1987). First-order integer-valued autoregressive (inar (1)) process. *Journal of Time Series Analysis*, 8(3):261–275, 1987.

Robert B Schnabel, John E Koonatz, and Barry E Weiss(1985). A modular system of algorithms for unconstrained minimization. *ACM Transactions on Mathematical Software (TOMS)*, 11(4):419–440.

Robert C Jung, Gerd Ronning, and Andrew R Tremayne(2005). Estimation in conditional first order autoregression with discrete support. *Statistical papers*, 46:195–224.

Tito Lívio, N Mamode Khan, Marcelo Bourguignon, and Hassan S Bakouch(2018). An inar (1) model with poisson–lindley innovations. *Econ Bull*, 38(3):1505–1513 .