Article

Likelihood Inference for Generalized Integer Autoregressive Time Series Models

Harry Joe

Department of Statistics, University of British Columbia, Vancouver, BC V6T 1Z4, Canada; Harry.Joe@ubc.ca

check for updates

Abstract: For modeling count time series data, one class of models is generalized integer autoregressive of order p based on thinning operators. It is shown how numerical maximum likelihood estimation is possible by inverting the probability generating function of the conditional distribution of an observation given the past p observations. Two data examples are included and show that thinning operators based on compounding can substantially improve the model fit compared with the commonly used binomial thinning operator.

Keywords: count time series; binomial thinning; thinning operators; compounding operation; self-generalized property

1. Introduction

Modeling of count times series has been an active area of research (see Davis et al. 2015 and Weiß 2018) for a few decades. There are applications in business and econometrics, health and medical studies; the models can be applied in general situations for count response variables with time dependence and covariates.

One class of count time series models is based on thinning operators, which are replacements of the multiplication operator to maintain support on the non-negative integers. The theory for count time series modeling with thinning operators has two approaches: (a) models with a specified stationary univariate margin combined with operators that can yield Markov times series models with a lag 1 serial correlation between 0 to 1; (b) models based on thinning operators applied to p previous observations and a distribution for the innovation, so that in the stationary setting, the autocorrelation functions have similar properties to the Gaussian case (except only positive serial correlations are possible).

The general model of type (b) is generalized integer autoregressive of order p, denoted as GINAR(p). For GINAR(p), mainly the binomial thinning operator has been used, whereas there are many thinning operators proposed for models of type (a).

For models of type (a), there is an extension to include covariates if the distribution of the innovation and the univariate margin are in the same infinitely divisible family, and then the convolution parameter can be a function of covariates. For models of type (b), the extension to include covariates typically requires the parameters of the innovation distribution to be function of the covariates.

For type (a), estimation can usually proceed with numerical maximum likelihood. For type (b), the estimation method is typically method of moments or conditional least squares. The disadvantage of the latter two estimation methods is that the predictive distributions for the next observations given the recent observations cannot be obtained.

In this paper, we use some numerical techniques from the literature for models of type (a) for parameter estimation in GINAR(p) models. The numerical technique applies when there is closed form for the probability generating function (pgf) of the conditional distribution of the next observation

given the past because the pgf can be efficiently numerically inverted to get the conditional probability mass function for likelihood calculations.

The remainder of the paper is organized as follows. Section 2 describes the background and notation for count times series models and thinning operators. Section 3 presents the formulation of GINAR(p). Section 4 presents some probabilistic properties of GINAR(p) in the stationary setting. Section 5 presents the details to show how the likelihood can be numerically computed, and includes some simulation results. Section 6 presents two data examples that show how generalized thinning operators lead to better models than binomial thinning. Section 7 presents the concluding discussion.

2. Background for Thinning Operators

This section presents the background and notation for count data and thinning operators.

Let $\{Y_t : t = 1, 2, \ldots\}$ be a count time series (a sequence of dependent random variables), where $Y_t \in \mathcal{N}_0 = \{0, 1, 2 \ldots, \}$. The realized data will be denoted as $\{y_t : t = 1, 2, \ldots, n\}$ for a sample of size n. If there are trend/seasonal terms or covariates at time t, they are incorporated into a vector $\mathbf{x}_t$. We mainly consider the case of small counts, with zeros and overdispersion relative to Poisson.

For early research in Markov count time series with given margins, some references are McKenzie (1986, 1987), Al-Osh and Aly (1992), Alzaid and Al-Osh (1993). Letting α denote the non-negative lag 1 serial correlation, the integer autoregressive model of order 1 has the form:

$$Y_t = R_t(Y_{t-1}; \alpha) + \epsilon_t(\alpha), \quad t = 1, 2, \ldots, \ 0 \le \alpha \le 1. \tag{1}$$

Y_t has a (parametric) distribution such as Poisson, negative binomial (NB) or generalized Poisson. The innovation random variables are $\epsilon_t(\alpha) \in \mathcal{N}_0$ for $t = 1, 2, \ldots$, and the serially dependent component of the model consists of random variables $R_t(y; \alpha) \in \mathcal{N}_0$ such that

$$\mathrm{E}\left[R_t(Y_{t-1}; \alpha) | Y_{t-1} = y\right] = \mathrm{E}\left[R_t(y; \alpha)\right] = \alpha y, \quad y \in \mathcal{N}_0, \ 0 \le \alpha \le 1, \tag{2}$$

and $Y_t \overset{d}{=} R_t(Y_{t-1}; \alpha) + \epsilon_t(\alpha)$ for all $0 \le \alpha \le 1$. Please note that to preserve the space of $\mathcal{N}_0$ (non-negative integers), (2) must hold in expectation: $R_t(y; \alpha) \equiv \alpha y$ is possible only for real-valued or non-negative reals.

Some examples include the following. (a) $Y_t \sim$ Poisson and $R_t(y; \alpha) \sim$ Binomial$(y; \alpha)$ for binomial thinning; (b) $Y_t \sim$ negative binomial and $R_t(y; \alpha)$ has a beta-binomial distribution; (c) $Y_t \sim$ generalized Poisson and $R_t(y; \alpha)$ has a quasi-binomial distribution.

For an extension to include covariates, if Y_t, $R_t(Y_{t-1}; \alpha)$ and $\epsilon_t(\alpha)$ are in the same convolution-closed family for all $0 < \alpha < 1$ with a convolution parameter ϑ, then ϑ can be a function of the covariates. However, with this approach, the extension to Markov of higher order is not simple. Markov of order 2 is tractable, but not order 3 or more; see Joe (1996) and Section 8.4.4 of Joe (1997).

A broad class of generalized thinning operators with the property of (2) can be obtained with a family of compounding operators based on a family of random variables $\{K(\alpha) : 0 \le \alpha \le 1\}$, where $K(\alpha) \in \mathcal{N}_0$ and $\mathrm{E}\left[K(\alpha)\right] = \alpha$.

Definition 1 (Compounding operator based on random variable K)**.** *Let K be a non-negative integer random variable and let y be a non-negative integer. Then K as a compounding operator is denoted as $K \circledast y$, where*

$$K \circledast y \overset{d}{=} \sum_{i=1}^{y} K_i$$

with K_i being independent replicates of K. In the above, $K \circledast y = 0$ if $y = 0$ with the conventional meaning of a null sum.

With the family $\{K(\alpha) : 0 \leq \alpha \leq 1\}$ of compounding random variables acting as thinning operators, in (1) let

$$R_t(y; \alpha) = \sum_{i=1}^{y} K_{ti}(\alpha),$$

where $K_{ti}(\alpha)$ are independent replicates of $K(\alpha)$ and $K(\alpha)$ has pgf $G_K(s; \alpha) = \mathrm{E}\left[s^{K(\alpha)}\right]$ for $0 \leq s \leq 1$. With the $\circledast$ notation, the Markov stationary count time series model can be written as

$$Y_t = K_t(\alpha) \circledast Y_{t-1} + \epsilon_t(\alpha) = \sum_{i=1}^{Y_{t-1}} K_{ti}(\alpha) + \epsilon_t(\alpha), \quad t = 1, 2, \ldots \tag{3}$$

For the expectation thinning requirement, $\alpha = 0$ implies $K(\alpha) \equiv 0$ and $\epsilon_t = Y_t$ in (3) for an independent and identically distributed sequence, and $\alpha = 1$ implies $K(\alpha) \equiv 1$ and $\epsilon_t = 0$ in (3) for perfect dependence.

Please note that if K has pgf G_K and Y has pgf G_Y, then $K \circledast y$ has pgf G_K^y and $K \circledast Y$ has pgf $G_Y \circ G_K$.

With the family $\{K(\alpha)\}$ of compounding operators, a subclass with the self-generalized property has some special properties. Self-generalization is defined next; properties based on this concept are in Zhu and Joe (2003, 2010a).

Definition 2 (Self-generalized family $\{K(\alpha) : 0 \leq \alpha \leq 1\}$ with pgf $G_K(\cdot; \alpha)$). *$\{K(\alpha) : 0 \leq \alpha \leq 1\}$ satisfies the property of self-generalization if*

$$K(\alpha) \circledast K(\alpha') = \sum_{i=1}^{K(\alpha')} K_i(\alpha) \overset{d}{=} K(\alpha \alpha') \text{ or } G_K[G_K(s; \alpha); \alpha'] = G_K(s; \alpha \alpha') \text{ for } 0 < \alpha, \alpha' < 1 \text{ and } 0 \leq s \leq 1. \tag{4}$$

This self-generalized property implies the following.

(a)　One can embed (3) into a continuous-time Markov process.
(b)　$\mathrm{Var}[K(\alpha)] = c\alpha(1 - \alpha)$ for some constant $c \geq 1$.
(c)　If (3) can hold in distribution for all $0 < \alpha < 1$, then the marginal distribution F_Y is infinitely divisible and said to be *generalized discrete self-decomposable* (GDSD).

There are three known families of self-generalized operators that are quite tractable. These are summarized in the next definition.

Definition 3 (Three families of self-generalized thinning operators).

(I1)　*(binomial thinning)* $G_K(s; \alpha) = (1 - \alpha) + \alpha s$ *with* $\mathrm{Var}[K(\alpha)] = \alpha(1 - \alpha)$.
(I2)　$G_K(s; \alpha; \gamma) = \frac{(1-\alpha)+(\alpha-\gamma)s}{(1-\alpha\gamma)-(1-\alpha)\gamma s}, 0 \leq \gamma \leq 1$, *with* $\mathrm{Var}[K(\alpha)] = \alpha(1 - \alpha)(1 + \gamma)/(1 - \gamma)$. *Please note that $\gamma = 0$ implies $G_K(z; \alpha) = (1 - \alpha) + \alpha s$.*
(I3)　$G_K(s; \alpha; \gamma) = \gamma^{-1}[1 + \gamma - (1 + \gamma - \gamma s)^\alpha], 0 \leq \gamma$, *with* $\mathrm{Var}[K(\alpha)] = \alpha(1 - \alpha)(1 + \gamma)$. *Please note that $\gamma \to 0$ implies $G_K(s; \alpha) = (1 - \alpha) + \alpha s$.*

The I2 and I3 families have an additional parameter γ besides $\alpha \in [0, 1]$ to allow different degrees of conditional heteroscedasticity. The I2 and I3 families include binomial thinning when $\gamma \to 0^+$. The second operator family I2 has been used in different parametrizations; see Aly and Bouzar (1994, 2019).

In the next section, we use these three classes of self-generalized compounding operators for GINAR(p) models in which compounding acts (independently) on each of the previous p observations. Please note that model (3) has no simple extension to Markov of higher order if a univariate stationary margin such as negative binomial is desired.

3. GINAR(p): Generalized Integer Autoregressive of Order $p \geq 1$

In this section, we define INAR(p) and GINAR(p) count times series models, and summarize estimation methods that have appeared in the literature. Also, it is indicated why the generalized form is more interpretable.

The INAR(p) model-based binomial thinning operators (denoted with $*$), is defined in Du and Li (1991), as:

$$Y_t = \sum_{j=1}^{p} \alpha_j * Y_{t-j} + \epsilon_t;$$

the innovation random variables ϵ_t have a distribution such as Poisson or negative binomial (NB), and they can have parameters that depend on covariates.

For an extension beyond binomial thinning, the GINAR(p) model as defined in Gauthier and Latour (1994) and Joe (2015) is as follows:

$$Y_t = \sum_{j=1}^{p} K_t(\alpha_j) \circledast Y_{t-j} + \epsilon_t = \sum_{j=1}^{p} \sum_{i=1}^{Y_{t-j}} K_{tji}(\alpha_j) + \epsilon_t, \tag{5}$$

where $0 \leq \alpha_j \leq 1$ for $j = 1, \ldots, p$ and the $K_{tji}(\alpha_j)$ are independent over t, j and i, and ϵ_t is the innovation at time t.

Model (5) could also be defined if the thinning operators are based on other non-compounding thinning operators that appear in other constructions of count time series models. However as shown below, the feasibility of numerical maximum likelihood depends on the use of compounding operators with closed form pgfs.

This GINAR(p) model is defined in Section 3.2 of Weiß (2018) but there is no discussion of good choices of generalized thinning operators and no applications in subsequent sections of the book.

Binomial thinning in the INAR(p) model with $p \geq 2$ does NOT have survivor-immigration interpretation (a random fraction of current counted units continue to next time point (page 18 of Weiß (2018))). GINAR(p) is more interpretable, with a unit count at time t branching into (or contributing) 0, 1 or more counts at time $t + 1$, $t + 2$, etc.; this is referred to as branching with immigration on page 19 of Weiß (2018).

The original estimation method in Du and Li (1991) is Yule-Walker or method of moments. An approximate likelihood inference method based on the saddlepoint approximation is used in Pedeli et al. (2015), and Lu (2018) has an implementation of maximum likelihood by getting the probability mass function from the pgf by differentiation and Taylor expansion. In Pedeli et al. (2015), only the binomial thinning operator is used and a NB innovation with a fixed convolution parameter is assumed. That is, the convolution parameter is fixed (whose reciprocal was called the dispersion parameter) at different values and then estimates of the remaining parameters use the saddlepoint approximation. Lu (2018) assumes a Poisson innovation, and the approach extends to NB innovations but may not be practical for thinning operators with support on all non-negative integers.

By inverting the pgf using the numerical integral in Davies (1973), the likelihood can be evaluated to high precision whenever the conditional distribution of $[Y_t | Y_{t-1} = y_{t-1}, \ldots, Y_{t-p} = y_{t-p}]$ has a numerically tractable pgf. The pgf of this conditional distribution is

$$\prod_{j=1}^{p} [G_K(s; \alpha_j)]^{y_{t-j}} G_\epsilon(s).$$

This pgf is tractable for thinning operators I1, I2 and I3, combined with Poisson or NB binomial distributions for innovations. Some details are provided in Section 5. In Zhu and Joe (2010b), the numerical technique for the likelihood was used for some Markov time series models of order 1 with NB marginal distributions.

An advantage of the likelihood method over method of moments and least squares is that different distributions for ϵ_t can be used in a sensitivity analysis, and prediction intervals of Y_{t+1} given $y_t, y_{t-1}, \ldots, y_{t-p}$ are possible. Covariates can be accommodated into the parameters of F_{ϵ_t}.

4. Probabilistic Properties and Numerical Techniques

In this section, probabilistic properties and numerical techniques for GINAR(p) are given.

The method for inverting the conditional pgf is summarized in Section 4.1 and it can be used for simulation of GINAR(p) as shown in Section 4.2. In Section 4.3, an algorithm is given for obtaining the autocorrelation function of (5), assuming stationarity. Section 4.4 summarizes the validation of the numerical methods.

4.1. Conditional Distributions

In this subsection, we explain how to compute the probability mass function (pmf) $\Pr(Y_t = y | Y_{t-1} = y_{t-1}, \ldots, Y_{t-p} = y_{t-p})$ for GINAR(p) in (5).

There are two approaches indicated below.

(a) Let the pmf for extended binomial distribution be

$$f(z; \alpha_1, \ldots, \alpha_p; G_K) = \Pr\left(\sum_{j=1}^{p} K_t(\alpha_j) \circledast y_{t-j} = z\right) \tag{6}$$

for $z = 0, 1, 2 \ldots$; this is obtained by inverting $\prod_{j=1}^{p} G_K^{y_{t-j}}(s; \alpha_j)$, which is the pgf of $\sum_{j=1}^{p} K_t(\alpha_j) \circledast y_{t-j}$. Please note that (6) is a binomial distribution when the thinning operator is binomial thinning and $\alpha_1 = \ldots = \alpha_p$. Let $f_\epsilon(\cdot)$ be the pmf of the innovation random variable. Then

$$f_{Y_t|Y_{t-1},\ldots,Y_{t-p}}(z) = \Pr(Y_t = y | Y_{t-1} = y_{t-1}, \ldots, Y_{t-p} = y_{t-p}) = \sum_{z=0}^{y} f(z; \alpha_1, \ldots, \alpha_p; G_K) f_\epsilon(y - z) \tag{7}$$

(b) The conditional pmf (7) is obtained by inverting $G_\epsilon(s) \prod_{j=1}^{p} G_K^{y_{t-j}}(s; \alpha_j)$, which is the pgf of $\sum_{j=1}^{p} K_t(\alpha_j) \circledast y_{t-j} + \epsilon_t$.

Approach (a) can be used if the innovation random variable has simple form for the pmf but not the pgf. In the next result that shows the inversion, W is either $\sum_{j=1}^{p} K_t(\alpha_j) \circledast y_{t-j}$ or $\sum_{j=1}^{p} K_t(\alpha_j) \circledast y_{t-j} + \epsilon_t$.

Let W be a random variable with support in $\mathcal{N}_0$. The characteristic function $\varphi_W(t) = E(e^{itW}) = G_W(e^{it})$ of W can be inverted via the algorithm of Davies (1973). Let

$$a(w) := \tfrac{1}{2} - (2\pi)^{-1} \int_{-\pi}^{\pi} \mathrm{Re}\left(\frac{\varphi_W(u)e^{-iuy}}{1 - e^{-iu}}\right) du \text{ with } \Pr(W < w) = a(w).$$

The function $a(w)$ is straightforward to evaluate via numerical quadrature. The cumulative distribution function (cdf) and pmf of W are

$$
\begin{aligned}
F_W(w) &= a(w+1), \quad w = 1, 2, \ldots \tag{8} \\
f_W(0) &= a(1), \ f_W(w) = a(w+1) - a(w), \quad w = 1, 2, \ldots. \tag{9}
\end{aligned}
$$

4.2. Simulating GINAR(p)

For simulating from GINAR(p), one can start with something approximate for the first p observations and then generate remaining observations from the conditional distributions in the preceding subsection. If one ignores the initial burn-in period, then rest of the resulting time series will be close to stationary, assuming that the parameters $\alpha_1, \ldots, \alpha_p$ satisfy the condition for stationary (as given in the next subsection).

Suppose one has the cdf for $\Pr(Y_t = z | Y_{t-1} = y_{t-1}, \ldots, Y_{t-p} = y_{t-p})$ as in the previous subsection. Then the usual technique can be used for simulating a discrete random variable. Below is an algorithm to generate a sequence of length $n > p$.

- Simulate y_1 so that it has the theoretical stationary mean and variance.
- Simulate $y_2, \ldots, y_p$ to try to match the lag 1 to lag $p - 1$ serial correlations (approximately).
- For $i = p + 1, \ldots, n$:
- 　 obtain the cdf $F_{cond}(z) \leftarrow F_{Y_t | Y_{t-1}, \ldots, Y_{t-p}}(z | y_{t-1}, \ldots, y_{t-p})$ for z from 0 to, say, the integer closest to
$$E(Y_t | Y_{t-1} = y_{t-1}, \ldots, Y_{t-p} = y_{t-p}) + 5\sqrt{\mathrm{Var}(Y_t | Y_{t-1} = y_{t-1}, \ldots, Y_{t-p} = y_{t-p})}.$$
- 　 generate a random number r in $(0, 1)$
- 　 assign $y_i \leftarrow \min\{z : F_{cond}(z) \geq r\}$.
- End of for loop

4.3. Moments under Stationarity

In this subsection, some results are derived and summarized under the assumption that model (5) is in a stationary state. It is assumed that the sequence $\{Y_t\}$ has finite variance σ_Y^2 and mean μ_Y. We use the property that $\{K(\alpha) : 0 \leq \alpha \leq 1\}$ is a family of self-generalized random variables with $E[K(\alpha)] = \alpha$ and $\mathrm{Var}[K(\alpha)] = c\alpha(1 - \alpha)$ for a constant $c \geq 1$. Du and Li (1991) has most of these results for the special case of binomial thinning. The condition for stationarity is that $0 \leq \sum_{j=1}^{p} \alpha_j < 1$.

- $E[K(\alpha) \circledast Y | Y = y] = yE[K(\alpha)] = \alpha y$; $E[K(\alpha) \circledast Y] = \alpha \mu_Y$.
- For stationary GINAR(p), $\mu_Y = \mu_Y \sum_{j=1}^{p} \alpha_j + \mu_\epsilon$ or

$$\mu_Y = \mu_\epsilon \Big/ \Big[1 - \sum_{j=1}^{p} \alpha_j \Big].$$

- $\mathrm{Var}[K(\alpha) \circledast Y | Y = y] = y\mathrm{Var}[K(\alpha)] = c\alpha(1 - \alpha)y$.
- $\mathrm{Var}[K(\alpha) \circledast Y] = c\alpha(1 - \alpha)\mu_Y + \alpha^2 \sigma_Y^2$ and $E[(K(\alpha) \circledast Y)^2] = c\alpha(1 - \alpha)\mu_Y + \alpha^2 E(Y^2)$.
- Let Y_a, Y_b be two (distinct) dependent counts, with independent thinning operations at the same time or at different times: $\mathrm{Cov}[K_{t_1}(\alpha_j) \circledast Y_a, K_{t_2}(\alpha_m) \circledast Y_b \mid Y_a = y_a, Y_b = y_b] = 0$, and $\mathrm{Cov}[K_t(\alpha_j) \circledast Y_a, Y_b \mid Y_a = y_a, Y_b = y_b] = 0$.
- Let Y_a, Y_b be two (distinct) dependent counts, with independent thinning operations at the same time or at different times. From the preceding items,

$$
\begin{aligned}
\mathrm{Cov}[K_t(\alpha_j) \circledast Y_a, Y_b] &= \mathrm{Cov}[E(K_t(\alpha_j) \circledast Y_a | Y_a, Y_b), Y_b] = \mathrm{Cov}(\alpha_j Y_a, Y_b) = \alpha_j \mathrm{Cov}(Y_a, Y_b); \\
\mathrm{Cov}(K_{t_1}(\alpha_j) \circledast Y_a, K_{t_2}(\alpha_m) \circledast Y_b) &= \mathrm{Cov}[E(K_{t_1}(\alpha_j) \circledast Y_a | Y_a, Y_b), E(K_{t_2}(\alpha_m) \circledast Y_b | Y_a, Y_b)] \\
&= \mathrm{Cov}(\alpha_j Y_a, \alpha_m Y_b) = \alpha_j \alpha_m \mathrm{Cov}(Y_a, Y_b).
\end{aligned}
$$

- For the case of $a = b$ and write $Y_a = Y_b = Y$, then

$$\mathrm{Cov}[K_t(\alpha) \circledast Y, Y] = \mathrm{Cov}[E(K_t(\alpha) \circledast Y | Y), Y] = \mathrm{Cov}(\alpha Y, Y) = \alpha \mathrm{Var}(Y).$$

With $t_1 \neq t_2$,

$$
\begin{aligned}
\mathrm{Cov}(K_{t_1}(\alpha_j) \circledast Y, K_{t_2}(\alpha_m) \circledast Y) &= \mathrm{Cov}[E(K_{t_1}(\alpha_j) \circledast Y | Y), E(K_{t_2}(\alpha_m) \circledast Y | Y)] \\
&= \mathrm{Cov}(\alpha_j Y, \alpha_m Y) = \alpha_j \alpha_m \mathrm{Var}(Y).
\end{aligned}
$$

- $\mathrm{Cov}(Y_t, \epsilon_t) = \mathrm{Var}(\epsilon_t) = \sigma_\epsilon^2$ since ϵ_t is independent of past observations.

From the above, we can develop recursion equations for autocovariances γ_h or autocorrelations ρ_h for $h \in \mathcal{N}_0$. These are the same as for Gaussian AR(p) for lags $h \geq p$.

Variance calculations under stationarity are as follows.

$$
Y_t = \sum_{j=1}^{p} K_t(\alpha_j) \circledast Y_{t-j} + \epsilon_t
$$

$$
\mathrm{Var}\,(Y_t) = \sum_{j=1}^{p} \mathrm{Var}\,[K_t(\alpha_j) \circledast Y_{t-j}] + 2 \sum_{1 \le j < m \le p} \mathrm{Cov}\,[K_t(\alpha_j) \circledast Y_{t-j}, K_t(\alpha_m) \circledast Y_{t-m}] + \mathrm{Var}\,(\epsilon_t)
$$

$$
\gamma_0 = c\mu_Y \sum_{j=1}^{p} \alpha_j(1-\alpha_j) + \gamma_0 \sum_{j=1}^{p} \alpha_j^2 + 2 \sum_{1 \le j < m \le p} \alpha_j \alpha_m \gamma_{m-j} + \sigma_\epsilon^2
$$

$$
(1 - \sum_{j=1}^{p} \alpha_j^2)\gamma_0 = c\mu_Y \sum_{j=1}^{p} \alpha_j(1-\alpha_j) + 2 \sum_{1 \le j < m \le p} \alpha_j \alpha_m \gamma_{m-j} + \sigma_\epsilon^2
$$

$$
\gamma_h = \sum_{j=1}^{p} \alpha_j \mathrm{Cov}\,(Y_{t-j}, Y_{t-h}) = \sum_{j=1}^{p} \alpha_j \gamma_{|h-j|}, \quad h \ge 1.
$$

For example, for GINAR(3),

$$
\begin{aligned}
\gamma_0 &= \{c\mu_Y \sum_{j=1}^{3} \alpha_j(1-\alpha_j) + \sigma_\epsilon^2\} + \gamma_0 \sum_{j=1}^{3} \alpha_j^2 + 2[\alpha_1\alpha_2\gamma_1 + \alpha_2\alpha_3\gamma_1 + \alpha_1\alpha_3\gamma_2] \\
&=: a_0 + b_0\gamma_0 + 2[\alpha_1\alpha_2\gamma_1 + \alpha_2\alpha_3\gamma_1 + \alpha_1\alpha_3\gamma_2] \\
\gamma_1 &= \mathrm{Cov}\,(Y_t, Y_{t-1}) = \alpha_1\gamma_0 + \alpha_2\gamma_1 + \alpha_3\gamma_2 \\
\gamma_2 &= \mathrm{Cov}\,(Y_t, Y_{t-2}) = \alpha_1\gamma_1 + \alpha_2\gamma_0 + \alpha_3\gamma_1 = \alpha_2\gamma_0 + (\alpha_1 + \alpha_3)\gamma_1 \\
\gamma_1 &= \alpha_1\gamma_0 + \alpha_3\alpha_2\gamma_0 + \alpha_2\gamma_1 + \alpha_3(\alpha_1 + \alpha_3)\gamma_1 \\
\gamma_1 &= (\alpha_1 + \alpha_2\alpha_3)\gamma_0 / [1 - \alpha_2 - \alpha_3(\alpha_1 + \alpha_3)] =: \rho_1\gamma_0 \\
\gamma_2 &= [\alpha_2 + (\alpha_1 + \alpha_3)\rho_1]\gamma_0 =: \rho_2\gamma_0 \\
\gamma_0 &= a_0 + b_0\gamma_0 + 2(\alpha_1\alpha_2 + \alpha_2\alpha_3)\rho_1\gamma_0 + 2\alpha_1\alpha_3\rho_2\gamma_0 \\
\gamma_0 &= a_0 / [1 - b_0 - 2(\alpha_1\alpha_2 + \alpha_2\alpha_3)\rho_1 - 2\alpha_1\alpha_3\rho_2].
\end{aligned}
$$

In general, the first p serial correlations can be obtained by solving a linear system in the equations for $\rho_1, \ldots, \rho_p$, and then $\rho_{p+1}, \rho_{p+2}, \ldots$ can obtained by recursion.

Next is the algorithm for GINAR(p) for computing $\rho_1, \ldots, \rho_p$, and $\gamma_0 = \sigma_Y^2$, given inputs of $\alpha_1, \ldots, \alpha_p$ The higher order serial correlations then are obtained via:

$$
\rho_h = \sum_{j=1}^{p} \alpha_j \rho_{h-j}, \quad h > p.
$$

- Initialize a $p \times p$ matrix M to 0.
- For $j_1 \in \{1, \ldots, p\}$
- $M_{j_1, j_1} \leftarrow 1$
- for $j_2 \in \{1, \ldots, p\}$
- $h \leftarrow |j_1 - j_2|$
- if $(h > 0)\ M_{j_1, h} \leftarrow M_{j_1, h} - \alpha_{j_2}$
- end of loop for j_2
- end of loop for j_1.
- Solve

$$
M \begin{pmatrix} \rho_1 \\ \vdots \\ \rho_p \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{pmatrix}.
$$

- Let $a_0 = c\mu_Y \sum_{j=1}^{p} \alpha_j(1 - \alpha_j) + \sigma_\epsilon^2$, $b_0 = \sum_{j=1}^{p} \alpha_j^2$. Then

$$\gamma_0 = a_0 \left/ \left(1 - b_0 - 2 \sum_{1 \le j_1 < j_2 \le p} \alpha_{j_1} \alpha_{j_2} \rho_{j_2 - j_1} \right) \right.$$

is the stationary variance.

Please note that the stationary autocorrelation function (acf) depends only on $\alpha_1, \ldots, \alpha_p$ and not on the distribution of the innovation random variables. The acf also does not depend on the family $\{K(\alpha)\}$ of thinning operations, but the stationary mean and variance are affected by $\{K(\alpha)\}$ and the distribution of the innovations.

4.4. Validation

With the simulation method in Section 4.2 and numerical maximum likelihood based on (7), we simulated many time series of length in the thousands under different stationary parameter settings for $(\alpha_1, \ldots, \alpha_p)$ and different parameters for the Poisson or NB innovation. Some representative simulation results are summarized in the next section.

The sample acf's are close to the theoretical acf's in Section 4.3 and the maximum likelihood estimates are close to the "true" parameters when considering sampling variability. We also checked that as p increases and $\sum_{j=1}^{p} \alpha_j$ gets closer to 1, the serial correlations require longer lags before they are closer to 0.

5. Likelihood and Numerical Implementation

In this section, we summarize the log-likelihoods that can be used for model (5), where the innovation random variables can be independent and identically distributed with a parametric distribution or they can depend parametrically on covariates.

To compare models with different autoregressive order using Akaike information criterion (AIC) values, we use the (conditional) likelihood as the product of the conditional densities starting at an index i_{start} which is larger than p_{max} (the maximum autoregressive order that will be considered). With a large sample size, the maximum likelihood estimates are not sensitive to the few initial conditional probabilities so that one could fit GINAR(p) starting from the conditional probability for Y_{p+1} to get parameter estimates and then omit a few conditional probability terms at the beginning so that i_{start} is common for different p. This form of likelihood implies that we do not have to determine the distribution of the first few observations under the assumption of stationarity. Actually, we then do not have to assume that the time series starts in a stationary state, and covariates can be included. The likelihood is:

$$L = \prod_{i=i_{start}}^{n} f_{Y_t | Y_{t-1}, \ldots, Y_{t-p}}(y_t | y_{t-1}, \ldots, y_{t-p}; \alpha_1, \ldots, \alpha_p, \gamma, \boldsymbol{\theta}_{innov}), \tag{10}$$

where $\alpha_1, \ldots \alpha_p \in (0,1)$, with $0 \le \alpha_1 + \cdots + \alpha_p < 1$, are the autoregressive parameters, γ is the conditional heteroscedatic parameter in Definition 3 if I2 or I3 thinning is used, and $\boldsymbol{\theta}_{innov}$ is the vector of parameters for the innovation random variables. If there are covariates, then $\boldsymbol{\theta}_{innov}$ consists of (regression) parameters linking the covariates to parameters of the innovation distribution. The conditional density is obtained via the numerical technique in Section 4.1. We refer to the parameter vector maximizing (10) or minimizing the negative log conditional likelihood as the conditional maximum likelihood (CML) estimate.

For example, $\boldsymbol{\theta}_{innov} = \lambda > 0$ for Poisson innovations, $\boldsymbol{\theta}_{innov} = (\vartheta, \xi)$ for NB innovations with convolution parameter ϑ, mean $\vartheta\xi$ and variance $\vartheta\xi(1 + \xi)$. With covariate vector $\mathbf{x}_t$ at time t, $\boldsymbol{\theta}_{innov} = (\beta_0, \boldsymbol{\beta})$ for Poisson innovations with mean $\mu(\mathbf{x}_t) = \exp\{\beta_0 + \boldsymbol{\beta}^T \mathbf{x}_t\}$ and $\boldsymbol{\theta}_{innov} = (\beta_0, \boldsymbol{\beta}, \xi)$ for NB innovations with mean $\mu(\mathbf{x}_t) = \exp\{\beta_0 + \boldsymbol{\beta}^T \mathbf{x}_t\}$, $\vartheta_t = \mu(\mathbf{x}_t)/\xi$ with constant overdispersion.

Because of the summations and numerical integrals, to gain computational speed in the numerical maximum likelihood, the negative log-likelihood can be coded in a high-level programming language such as Fortran 90. The numerical optimization could be done by interfacing to a statistical software such as R.

Table 1 reports on some representative simulation results for I1 and I2 thinning to show the accuracy of the numerical methods and how the computing time increases when there are increases in the (i) number of parameters, (ii) sample size and (iii) maximum count. Similar patterns occur for GINAR with I3 thinning. When the sample size increases by a factor of 4 (from 500 to 2000), the SDs of the parameter estimates decrease by a factor of 2 (as expected); the computing time increase by a factor of slightly more than 4 because the maximum count is larger for a sample size of 2000 and the pmf in (7) must be computed up to a larger value. The parameter vectors in Table 1 are based on maximum likelihood values when fitting these different models to the Ericcson data in Section 6.1.

Table 1. Simulation results for different parameter vectors that come from fits to the Ericcson data in Section 6.1. The likelihoods were coded in Fortran90 and the numerical optimization was done with a link to R using `nlm` as the implementation of the quasi-Newton method. The simulation sample size was 500, and the timings were based on a PC with Intel Core i7-6770HQ processor at 2.6 GHz.

	I1/NB, $p = 2$, $n = 500$, av = 0.85 min				
parameter	$\alpha_1 = 0.27$	$\alpha_2 = 0.15$		$\theta = 1.85$	$\xi = 3$
bias	-0.001	-0.003		0.06	-0.02
rmse	0.035	0.037		0.34	0.38
	I1/NB, $p = 2$, $n = 2000$, av = 3.7 min				
parameter	$\alpha_1 = 0.27$	$\alpha_2 = 0.15$		$\theta = 1.85$	$\xi = 3$
bias	0.000	-0.001		0.02	-0.02
rmse	0.018	0.018		0.17	0.20
	I2/Po, $p = 2$, $n = 500$, av = 1.2 min				
parameter	$\gamma = 0.7$	$\alpha_1 = 0.3$	$\alpha_2 = 0.2$		$\lambda = 4.5$
bias	0.001	0.000	-0.005		0.03
rmse	0.027	0.045	0.046		0.35
	I2/Po, $p = 2$, $n = 2000$, av = 5.0 min				
parameter	$\gamma = 0.7$	$\alpha_1 = 0.3$	$\alpha_2 = 0.2$		$\lambda = 4.5$
bias	-0.001	0.000	0.000		0.01
rmse	0.013	0.022	0.022		0.17
	I2/Po, $p = 3$, $n = 500$, av = 2.9 min				
parameter	$\gamma = 0.64$	$\alpha_1 = 0.27$	$\alpha_2 = 0.14$	$\alpha_3 = 0.20$	$\lambda = 4.5$
bias	0.003	-0.001	-0.005	-0.003	0.07
rmse	0.030	0.045	0.049	0.043	0.46
	I2/Po, $p = 3$, $n = 2000$, av = 12.2 min				
parameter	$\gamma = 0.64$	$\alpha_1 = 0.27$	$\alpha_2 = 0.14$	$\alpha_3 = 0.20$	$\lambda = 4.5$
bias	0.000	-0.001	-0.001	-0.001	0.02
rmse	0.014	0.022	0.024	0.022	0.23

6. Data Examples

In this section, we show results of fitting model (5) to two data examples that appeared in the literature. The use of operators with branching stochastic representation is more interpretable. So it is not surprising that, based on AIC, the I2 and I3 thinning operators in Definition 3 provide much better fits than binomial thinning.

Because the I2 and I3 thinning operators provide conditional heterscedasticity, it turns out that the innovation random variable need not be as overdispersed relative to Poisson as p gets larger. With the binomial thinning operator which each count from the previous p time points can contribute at most 1 to the next count observation, a NB innovation random variable provides a much better fit than Poisson.

6.1. Ericsson Transaction Data

The data set consists of the number of transactions per minute for the stock Ericsson B for business days and hours during 2 to 22 July in the year 2002. The sample size is $n = 460$. The original source is Brännäs and Quoresh (2010). This data set is also used in Fokianos et al. (2009) and in Examples 4.1.5 and 4.2.4 in Weiß (2018).

Different models were used by the previous authors: integer moving average INMA of order q with a large q and INGARCH(1,1) with Poisson and overdispersed Poisson distributions for the innovations. The empirical autocorrelation function (see Table 3) suggests that a low-order autoregressive model is not appropriate for these data (as indicated by Weiß (2018)). Here, for comparison, we consider stationary GINAR(p) with p increasing until there is no improvement in the log-likelihood and the highest order α parameter becomes close to 0.

Table 2 presents a summary of AIC values to compare GINAR(p) with binomial thinning and the I2, I3 thinning operators.

For thinning with I2 and I3, based on AIC values, the models with Poisson innovations are a little better than with NB innovations for $p = 3$ to 7 for I2 and $p = 4$ to 7 for I3. For binomial thinning, based on AIC values, the models with NB innovations are much better than with Poisson innovations.

The best models based on AIC values are GINAR(6) models with I2 and I3 thinning operators and Poisson-distributed innovations. They are quite an improvement on INAR(6) with binomial thinning and NB-distributed innovations. Based on the context of the data, thinning operations based on compounding with support on all of $\mathcal{N}_0$ instead of $\{0, 1\}$ are more reasonable.

The I2 and I3 thinning operators account for some conditional heteroscedasticity so that the use of the NB-distributed innovations leads to a flatter log-likelihood over the NB parameters. Hence Poisson distributed innovations are adequate to handle the marginal overdispersion.

Table 2. Ericsson transaction data: AIC values for (5) with three thinning operators and Poisson (Po) and negative binomial (NB) distributions for the innovation. When the autoregressive order reaches 7, there is no improvement in the log-likelihood and the last estimated α_j is close to 0. The AIC values are based on (10) with $i_{start} = 8$. The AIC values for the best models for each of I1,I2,I3 are boldfaced. For $p = 1$ and 2, the AIC values for I2/NB are 2694.6 and 2679.4 respectively, and for I3/NB they are 2690.5 and 2677.1 respectively. For larger p, there is enough conditional heterscedasticity from the thinning operators and NB innovations did not lead to improved AIC values over Poisson innovations.

p	I1/NB	I2/Po	I3/Po
1	2695.7	2702.9	2722.0
2	2682.5	2681.0	2686.0
3	2670.6	2663.9	2666.1
4	2662.7	2654.3	2654.9
5	2657.0	2648.1	2647.3
6	**2651.9**	**2641.1**	**2639.5**
7	2653.3	2642.6	2640.9

We next compare AIC values with other models that have been used for this data set. Table 4.4 and Example 4.2.4 of Weiß (2018) have values of maximized log conditional likelihoods at the CML estimates for INGARCH(1,1) models with Poisson, NB and generalized Poisson innovations. The use of NB and generalized Poisson distributions lead to much better fitting models with AIC values in the range 2662 to 2666. With a further adjustment for starting in the eighth observation in the conditional log-likelihood, the AIC values are comparable with those in Table 2.

With also compare with the Poisson autoregression or INGARCH models in Fokianos et al. (2009). The models are conditional Poisson with a latent mean process $\{\Lambda_t\}$, where $[Y_t | \Lambda_s = \lambda_s : s \leq t, Y_s = y_s : s < t] \sim \text{Poisson}(\lambda_t)$ with

$$\lambda_t = b_0 + b_1 \lambda_{t-1} + b_2 y_{t-1}$$

or

$$\lambda_t = (b_0 + b_1 \exp\{-\gamma \lambda_{t-1}^2\})\lambda_{t-1} + b_2 y_{t-1}.$$

As in Fokianos et al. (2009), we start the latent process with $\lambda_0 = 0$. There is some sensitivity to the starting value y_0 but we get similar maximum likelihood parameter estimates when starting y_0 at the sample mean. When evaluating AIC values with $i_{start} = 8$ are between 2847 and 2849 for these two models and correspond roughly to the AIC value for INGARCH(1,1) with Poisson innovations in Weiß (2018).

Table 3 has (a) CML estimates for the GINAR(6) models with I1, I2 and I3 thinning operators, and (b) model-based moments and acf's from these models to compare with the empirical values. The comparison with empirical values provides a simple goodness-of-fit procedure; in this case, it shows that the fit from GINAR(6) with binomial thinning is a worse fit.

Based on the parameter estimates in this table, and the last 6 values in the data series: $(y_{460}, y_{459}, y_{458}, y_{457}, y_{456}, y_{455}) = (3, 9, 29, 18, 20, 7)$, we can estimate the conditional mean $\hat{\mu}_{cond}$, conditional variance $\hat{\sigma}^2_{cond}$ and central 50% and 80% intervals from estimated pmf: $f_{Y_{n+1}|Y_n,\ldots,Y_{n-5}}(\cdot|y_n,\ldots,y_{n-5}; \hat{\alpha}_1,\ldots,\hat{\alpha}_6, \hat{\theta}_{innov})$. These are summarized below.

- I1 (binomial thinning): $\hat{\mu}_{cond} = 11.62$, $\hat{\sigma}^2_{cond} = 25.66$, $[7, 12]$ with probability content 0.52; $[5, 16]$ with probability content 0.82;
- I2: $\hat{\mu}_{cond} = 12.20$, $\hat{\sigma}^2_{cond} = 30.31$, $[7, 14]$ with probability content 0.56; $[5, 18]$ with probability content 0.82;
- I3: $\hat{\mu}_{cond} = 12.20$, $\hat{\sigma}^2_{cond} = 30.92$, $[7, 13]$ with probability content 0.51; $[5, 18]$ with probability content 0.82.

For binomial thinning, the point and interval predictions are smaller and shorter. Please note that these prediction intervals would not be possible with estimation based on conditional least squares or the method of moments.

Table 3. Ericsson transaction data: CML parameter estimates and corresponding SEs for GINAR(6) with I1, I2 and I3 thinning; also model-based summary statistics, to compare with empirical.

Parameter	I1/NB	I2/Po	I3/Po	
$\hat{\gamma}$		0.533 (0.036)	2.321 (0.333)	
$\hat{\alpha}_1$	0.172 (0.037)	0.187 (0.047)	0.194 (0.047)	
$\hat{\alpha}_2$	0.057 (0.037)	0.068 (0.048)	0.071 (0.047)	
$\hat{\alpha}_3$	0.086 (0.036)	0.109 (0.049)	0.109 (0.047)	
$\hat{\alpha}_4$	0.086 (0.037)	0.116 (0.048)	0.117 (0.047)	
$\hat{\alpha}_5$	0.093 (0.038)	0.104 (0.050)	0.109 (0.047)	
$\hat{\alpha}_6$	0.105 (0.038)	0.142 (0.048)	0.146 (0.047)	
$\hat{\vartheta}$	1.068 (0.125)			
$\hat{\zeta}$	3.717 (0.509)			
$\hat{\lambda}$		2.704 (0.538)	2.507 (0.531)	
Summary	**I1/NB**	**I2/Po**	**I3/Po**	**Empirical**
$\hat{\mu}_Y$	9.879	9.889	9.892	9.909
$\hat{\sigma}^2_Y$	27.460	30.070	31.707	32.837
$\hat{\rho}_1$	0.257	0.350	0.374	0.405
$\hat{\rho}_2$	0.176	0.278	0.302	0.340
$\hat{\rho}_3$	0.189	0.297	0.317	0.372
$\hat{\rho}_4$	0.193	0.305	0.326	0.377
$\hat{\rho}_5$	0.201	0.302	0.326	0.358
$\hat{\rho}_6$	0.205	0.321	0.343	0.352
$\hat{\rho}_7$	0.123	0.227	0.250	0.298

6.2. Meningococcal Disease Data

The data set comes from the German national surveillance system for notifiable diseases, administered by the Robert Koch Institute. The time series consists of weekly numbers of meningococcal disease cases in Germany for the years 2001–2006 and the sample size is $n = 312$. In Pedeli et al. (2015), INAR(p) models are fitted with approximate likelihood based on a saddlepoint approximation. There is a seasonal pattern over the year and sinusoidal terms were used as covariates in the mean parameter of the innovations as indicated in Section 5.

We fit several GINAR(p) models with I2 and I3 thinning in addition to binomial thinning, using the numerical techniques in the preceding sections. The primary sinusoidal terms $x_{t1} = \sin(2\pi t/52)$ and $x_{t2} = \cos(2\pi t/52)$. As indicated in Pedeli et al. (2015), the addition of additional harmonic terms $x_{t3} = \sin(4\pi t/52)$ and $x_{t4} = \cos(4\pi t/52)$ do not lead to improvements based on AIC; the estimates of the corresponding β regression parameters are at least 10 times smaller than those for x_{t1}, x_{t2}. The autoregressive order started at 1 and increased to a value p_{max} so that the last $\hat{\alpha}_j$ was close to 0 and the negative log-likelihood value was not improving.

Table 4 has some AIC values, based on $i_{start} = 5$ in (10). For thinning with I2 and I3, based on AIC values, the models with Poisson innovations are a little better than those with NB innovations for $p = 2, 3, 4$. For binomial thinning, based on AIC values, the models with NB innovations are much better than those with Poisson innovations.

As in the Section 6.1, the I2 and I3 thinning operators account for some conditional heteroscedasticity so that the use of the NB-distributed innovations leads to a flatter log-likelihood over the NB parameters.

Overall, from Table 4, the GINAR(2) models with I2 or I3 thinning, Poisson-distributed innovations and x_{t1}, x_{t2} as covariates provide the best models. The best AIC values from GINAR(p) with I2 and I3 thinning are smaller than the best AIC values in Pedeli et al. (2015) (based on binomial thinning).

Table 4. Meningococcal disease data: AIC values for (5) with three thinning operators and Poisson (Po) and negative binomial (NB) distributions for the innovation. The AIC values are based on (10) with $i_{start} = 5$. The AIC values of the best models for each of I1,I2,I3 are boldfaced. The covariates are $x_{t1} = \sin(2\pi t/52)$ and $x_{t2} = \cos(2\pi t/52)$.

p	No Covariates			Covariates x_{t1}, x_{t2}		
	I1/NB	I2/Po	I3/Po	I1/NB	I2/Po	I3/Po
1	1766.5	1754.8	1758.5	1689.3	1684.8	1683.9
2	1738.5	1731.2	1730.0	1686.0	**1681.5**	**1681.9**
3	1726.6	1723.2	1721.6	**1684.5**	1683.5	1682.3
4	1728.7	1725.2	1723.6	1686.6	1685.9	1684.7

7. Discussion

We showed how GINAR(p) count time series models can be estimated using numerical maximum likelihood. This allows for sensitivity analysis to model assumptions that is not possible with the estimation methods of conditional least squares and the method of moments.

Future research includes the use of the thinning operators in Definition 3 in other applications, as well as derivations of other families of self-generalized random variables satisfying Definition 2.

Where researchers have been using binomial thinning in count time series models, we recommend the replacement with more general thinning operators that satisfy the self-generalized properties. These operators are more interpretable when the survivor interpretation for binomial thinning does not apply. The examples in this paper show that better fitting models can be obtained.

For likelihood computations, alternative methods can be considered. Although a greater coding effort would be needed, one could consider whether the method of Lu (2018), consisting of obtaining probabilities from derivatives of the pgf, is feasible for thinning operators with support on all non-negative integers.

Conflicts of Interest: The author declares no conflict of interest.

References

Al-Osh, Mohamed A., and Emad-Eldin A. A. Aly. 1992. First order autoregressive time-series with negative binomial and geometric marginals. *Communications in Statistics—Theory and Methods* 21: 2483–92. [CrossRef]

Aly, Emad-Eldin A. A., and Nadjib Bouzar. 1994. Explicit stationary distributions for some galton-watson processes with immigration. *Communications in Statistics—Stochastic Models* 10: 499–517. [CrossRef]

Aly, Emad-Eldin A. A., and Nadjib Bouzar. 2019. Expectation thinning operators based on linear fractional probability generating functions. *Journal of the Indian Society for Probability and Statistics* 20: 89–107. [CrossRef]

Alzaid, Abdulhamid A., and Mohamed A. Al-Osh. 1993. Some autoregressive moving average processes with generalized Poisson marginal distributions. *Annals of the Institute of Statistical Mathematics* 45: 223–32. [CrossRef]

Brännäs, Kurt, and A. M. M. Shahiduzzaman Quoreshi. 2010. Integer-valued moving average modelling of the number of transactions in stocks. *Applied Financial Economics* 20: 1429–40. [CrossRef]

Davies, Robert B. 1973. Numerical inversion of a characteristic function. *Biometrika* 60: 415–17. [CrossRef]

Davis, Richard A., Scott H. Holan, Robert Lund, and Nalini Ravishanker. 2015. *Handbook of Discrete-Valued Time Series*. Boca Raton: Chapman & Hall/CRC.

Du, Jin-Guan, and Yuan Li. 1991. The integer-valued autoregressive (INAR(p)) model. *Journal of Time Series Analysis* 12: 129–42.

Fokianos, Konstantinos, Anders Rahbek, and Dag Tjostheim. 2009. Poisson autoregression. *Journal of the American Statistical Association* 104: 1430–39. [CrossRef]

Gauthier, Geneviève, and Alain Latour. 1994. Convergence forte des estimateurs des paramètres d'un processus GENAR(p). *Ann. Sci. Math. Québec* 18: 49–71.

Joe, Harry. 1996. Time series models with univariate margins in the convolution-closed infinitely divisible class. *Journal of Applied Probability* 33: 664–77. [CrossRef]

Joe, Harry. 1997. *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

Joe, Harry. 2015. Markov count time series models with covariates. In *Handbook of Discrete-Valued Time Series*. Edited by Richard A. Davis, Scott H. Holan, Robert Lund and Nalini Ravishanker. Boca Raton: Chapman & Hall/CRC, chp. 2, pp. 29–49.

Lu, Yang. 2018. Probabilistic Forecasting in Higher-Order INAR(p) Models. MPRA Paper 83682, Munich: University Library of Munich.

McKenzie, Ed. 1986. Autoregressive moving-average processes with negative-binomial and geometric marginal distributions. *Advances in Applied Probability* 18: 679–705. [CrossRef]

McKenzie, Ed. 1987. Innovation distributions for gamma and negative binomial autoregressions. *Scandinavian Journal of Statistics* 14: 79–85.

Pedeli, Xanthi, Anthony C. Davison, and Konstantinos Fokianos. 2015. Likelihood estimation for the inar(p) model by saddlepoint approximation. *Journal of the American Statistical Association* 110: 1229–38. [CrossRef]

Weiß, Christian H. 2018. *An Introduction to Discrete-Valued Time Series*. Hoboken: Wiley.

Zhu, Rong, and Harry Joe. 2003. A new type of discrete self-decomposability and its application to continuous-time Markov processes for modeling count data time series. *Stochastic Models* 19: 235–54. [CrossRef]

Zhu, Rong, and Harry Joe. 2010a. Count data time series models based on expectation thinning. *Stochastic Models* 26: 431–62. [CrossRef]

Zhu, Rong, and Harry Joe. 2010b. Negative binomial time series models based on expectation thinning operators. *Journal of Statistical Planning and Inference* 140: 1874–88. [CrossRef]