Journal of Statistical Planning and Inference 140 (2010) 1874–1888

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Negative binomial time series models based on expectation thinning operators

Rong Zhu [a,*], Harry Joe [b]

[a] Department of Mathematics and Statistics, McMaster University, Canada
[b] Department of Statistics, University of British Columbia, Canada

ARTICLE INFO

Article history:
Received 9 September 2008
Received in revised form
28 December 2009
Accepted 15 January 2010
Available online 22 January 2010

Keywords:
Negative binomial time series
Autoregressive
Binomial thinning
Self-generalizability
Generalized discrete self-decomposability
Inversion of characteristic function

ABSTRACT

The study of count data time series has been active in the past decade, mainly in theory and model construction. There are different ways to construct time series models with a geometric autocorrelation function, and a given univariate margin such as negative binomial. In this paper, we investigate negative binomial time series models based on the binomial thinning and two other expectation thinning operators, and show how they differ in conditional variance or heteroscedasticity. Since the model construction is in terms of probability generating functions, typically, the relevant conditional probability mass functions do not have explicit forms. In order to do simulations, likelihood inference, graphical diagnostics and prediction, we use a numerical method for inversion of characteristic functions. We illustrate the numerical methods and compare the various negative binomial time series models for a real data example.

1. Introduction

Time series models for Poisson and overdispersion Poisson distributions have been studied since Phatarfod and Mardia (1973), McKenzie (1986, 1988); see McKenzie (2003) for an overview of this area of research. With count data that are overdispersed relative to the Poisson distribution, the negative binomial distribution is commonly used. Negative binomial time series are included in, for example, McKenzie (1986), Al-Osh and Aly (1992), Aly and Bouzar (1994, 2005), Joe (1996), Latour (1998), Zhu and Joe (2003). In this paper, we compare different negative binomial time series models based on expectation thinning operators, and show how they differ in properties of conditional heteroscedasticity. We also demonstrate computational techniques for inference and simulation of these models which are specified via probability generating functions. The main ideas extend to other marginal distributions that are overdispersed relative to the Poisson distribution.

A negative binomial random variable X can have several stochastic representations of the form

$$X \stackrel{d}{=} (\alpha)_K \circledast X + \varepsilon(\alpha) = \sum_{i=1}^{X} K_i(\alpha) + \varepsilon(\alpha), \tag{1.1}$$

where $\{(\alpha)_K : 0 < \alpha < 1\}$ is a family of (random) expectation thinning operators (defined in Section 2) based on integer-valued random variables $\{K(\alpha)\}$, and $\varepsilon(\alpha)$ is an integer random variable independent of X and the operator. The family of operators $\{(\alpha)_K\}$ satisfies a closure property that we call *self-generalizability*. At the end of Section 2, we give prior references

* Corresponding author.
E-mail address: rzhu@math.mcmaster.ca (R. Zhu).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2010.01.031

for these operators and their uses in time series models, as well as an explanation of how we obtained them in a different way.

The intuition behind (1.1) comes from time series where $(\alpha)_K \circledast X$ is interpreted as a random operator acting on the previous observation, and $\varepsilon(\alpha)$ is an innovation or a random component independent of the previous observation. From any representation of the form (1.1), we can define a stationary negative binomial time series model that has geometric autocorrelation function. There are differing amounts of conditional heteroscedasticity depending on the choice of operator $\{(\alpha)_K\}$. The stationary time series can be written as

$$X_t = (\alpha)_K \circledast X_{t-1} + \varepsilon_t(\alpha), \quad t = 2, 3, \ldots. \tag{1.2}$$

A consequence of the self-generalizability property is that for integer $i>0$, (X_t, X_{t+i}) has a bivariate distribution with dependence parameter α^i.

Background theory of the class of operators are given in Section 2 and the definitions of the stationary Markov negative binomial time series are given in Section 3. The proofs that the time series models are well-defined depend on a theorem which is proved in an Appendix. Computing details needed for likelihood inference and simulation are given in Section 4. A data example is given in Section 5. Section 6 consists of some discussion of further research.

The key contribution of this paper is to show that likelihood inference and simulation of count time series models are feasible for models with closed form probability generating functions but not closed form probability mass functions. Another main contribution is the presentation of specific families of self-generalized operators that can be used for negative binomial time series models, and showing how different operators affect the conditional heteroscedasticity.

Notation and abbreviations used throughout the remainder of the paper are: $\mathcal{N}_0$ for the set of non-negative integers, $\stackrel{d}{=}$ is used for equality in distribution for random variables; iid for independent and identically distributed; G_Z for the probability generating function (pgf) of an integer-valued random variable Z. By convention, a summation over a null set is defined as 0.

2. Expectation thinning operators

In this section, we go through a sequence of definitions in order to make Eq. (1.1) well defined. The definitions are for the concepts of compounding operation, self-generalized distribution family, expectation thinning operator, and generalized discrete self-decomposability. We also show the closure of the stochastic representation of (1.2) under iteration.

For a non-negative integer random variable X, αX for $0<\alpha<1$ is not always an integer, so that it is replaced by an operator $(\alpha)_K \circledast X = \sum_{i=1}^{X} K_i(\alpha)$, where $K_i(\alpha)$ are independent and identically distributed integer-valued random variables. The $K_i(\alpha)$ are Bernoulli (α) random variables in the case of binomial thinning. As the simplest branching operator, the binomial thinning operator has been used by many statistical researchers to construct various count data time series models; recent papers include Aly and Bouzar (2008), Bu and McCabe (2008), and Weiß (2008a). Aly and Bouzar (1994) introduced a two-parameter family of branching operators that include binomial thinning at one boundary; Zhu and Joe (2003) independently introduced the same family with a different parametrization, using a different derivation.

In order to derive time series models with some nice properties, we require $\{K(\alpha): 0 \le \alpha \le 1\}$ to satisfy a closure property which we call *self-generalizability*. It will result in the following.

(a) The conditional expectation is linear:

$$\mathrm{E}\,[(\alpha)_K \circledast X | X = x] = \alpha x.$$

That is, the thinning operator $(\alpha)_K \circledast$ is like a random multiplication by α.

(b) There is closure in form of conditional distribution: conditional distribution of X_{t+j} given X_t is the same parametric family for different j's with dependence parameter that depends on j. This then means that if there are missing values in the discrete-time series, the log-likelihood is not any harder to write down for maximum likelihood estimation.

(c) The time series can be considered as a realization of a continuous time stochastic process that is observed at equally-spaced time points.

Definition 2.1 (*Compounding operation*). Let K, X be random variables with support on $\mathcal{N}_0$. Let K_i, $i \ge 1$, be iid copies of K that are independent of X. Define $K \circledast X \stackrel{d}{=} \sum_{j=1}^{X} K_j$. The pgf of $K \circledast X$ is

$$G_{K\circledast X}(z) = \mathrm{E}\left(z^{\sum_{j=1}^{X} K_j}\right) = \mathrm{E}\left[\mathrm{E}\left(z^{\sum_{j=1}^{X} K_j} \middle| X\right)\right] = \mathrm{E}\left[G_K^X(z)\right] = G_X(G_K(z)).$$

So we could equivalently define $K \circledast X$ as a non-negative integer-valued random variable with pgf $G_X(G_K(z))$.

We note the following associative law for compounding operators as this property will be used later.

Theorem 2.1 (*Associative law*). *Let K_1, K_2 be two non-negative integer random variables acting as operators. Let X correspondingly have support on $\mathcal{N}_0$. Then*

$$K_1 \circledast (K_2 \circledast X) \overset{d}{=} (K_1 \circledast K_2) \circledast X.$$

Proof. Both sides of the equation have the same pgf $G_X[G_{K_2}(G_{K_1}(z)]$. □

In order to get desirable closure properties and a range of dependence from independence to perfect dependence for our time series models, we consider families of random variables $\{K(\alpha) : 0 \le \alpha \le 1\}$ with the property in the next definition. Families of random variables with this property lead to operators which extend binomial thinning for non-negative integers. Results that we need for families of self-generalized random variables are given in Appendix A.1.

Definition 2.2 (*Self-generalized distribution family*). Let $\{F(x;\alpha) : \alpha \in [0,1]\}$, be a family of distinct cumulative distribution functions (cdf) with support on $\mathcal{N}_0$. Let $K(\alpha)$ be a generic random variable with cdf $F(x;\alpha)$ and pgf $G_K(z;\alpha) = \mathrm{E}[z^{K(\alpha)}]$. Then the distribution family $\{F(x;\alpha)\}$ is said to be *self-generalized* with respect to parameter α if

$$G_K(G_K(z;\alpha);\alpha') = G_K(z;\alpha\alpha'), \quad \forall\ \alpha,\alpha' \in [0,1]. \tag{2.1}$$

From Definition 2.1, $\{K(\alpha); 0 \le \alpha \le 1\}$ satisfy the closure property:

$$K(\alpha) \circledast K(\alpha') \overset{d}{=} K(\alpha\alpha'), \quad 0 \le \alpha, \ \ \alpha' \le 1.$$

Next we define a family of expectation thinning operators as compounding operators with a family of self-generalized random variables $\{K(\alpha); 0 \le \alpha \le 1\}$.

Definition 2.3 (*Expectation thinning*). Let $\{K(\alpha); 0 \le \alpha \le 1\}$ be a family of self-generalized random variables with support on $\mathcal{N}_0$ and finite mean. The compounding operation restricted with $\{K(\alpha); 0 \le \alpha \le 1\}$, $K(\alpha) \circledast X = \sum_{i=1}^{X} K_i(\alpha)$ defined in Definition 2.1, is called the expectation thinning if $\mathrm{E}[K(\alpha)] \le 1$ for all $0 < \alpha < 1$.

For X with support on $\mathcal{N}_0$ and finite mean, the above definition implies that $\mathrm{E}[K(\alpha) \circledast X] = \mathrm{E}[K(\alpha)]\mathrm{E}[X] \le \mathrm{E}[X]$. For a similar notation with the constant multiplier and binomial thinning operators, we further denote:

$$(\alpha)_K \circledast X \overset{def}{=} K(\alpha) \circledast X. \tag{2.2}$$

With these expectation thinning operators, we are led to a time series of the form (1.2). However the time series in (1.2) is well-defined as a stationary time series only if X_t is in the generalized discrete self-decomposable (GDSD) classes defined in Definition 2.4.

Before we define GDSD, we list some parametric families of self-generalized distributions that we use for expectation thinning operators.

Example 2.1.

- **(I1)**: Binomial thinning, $K(\alpha) \sim \mathrm{Bernoulli}(\alpha)$, $0 \le \alpha \le 1$, with $G_K(s;\alpha) = (1-\alpha) + \alpha s$.
- **(I2)**: $G_K(s;\alpha) = ((1-\alpha)+(\alpha-\gamma)s)/((1-\alpha\gamma)-(1-\alpha)\gamma s)$, $0 \le \alpha \le 1$ and $0 < \gamma \le 1$ is fixed. When $\gamma = 0$, this becomes the family in **(I1)**. When $\gamma = 1$, the pgf will be 1, degenerating to distribution with mass 1 at the point zero.
- **(I3)**: $G_K(s;\alpha) = \delta^{-1}[1+\delta-(1+\delta-\delta s)^{\alpha}]$, $0 \le \alpha \le 1$, $\delta \ge 0$. Since the limit as $\delta \to 0$ is $1-\alpha+\alpha s$, the lower boundary leads to the family in **(I1)**. This family of pgf's can be obtained from Laplace transform family LTJ in the Appendix of Joe (1997) by adding additional mass at 0. When $\delta \to \infty$, the limiting pgf will be 1, degenerating to a distribution with mass 1 at the point zero.

In the above examples, the families of self-generalized random variables $K(\alpha)$ in **I2** and **I3** are indexed by another parameter γ or δ. The roles of the other parameter will be seen more clearly from the results in the next section.

I1 leads to binomial thinning often denoted as $\alpha * X$, $\alpha \circ X$ or $\alpha \odot X$ $(= \sum_{i=1}^{X} I_i(\alpha))$ which is always less than or equal to X. The symbol "$\circledast$" combines these and looks close to the constant multiplier symbol "$\bullet$", thus, we adopt "$\circledast$" to represent a large class of expectation thinning operations.

I2 is defined in the form of Zhu and Joe (2003). Aly and Bouzar (1994) have a different parametrization of

$$G(z;\beta,\eta) = 1-\beta\frac{(1-z)}{(1-(1-\beta)\eta z)}, \quad 0 < \beta < 1, \ \ 0 \le \eta \le 1$$

and this matches **I2** if $\gamma = \eta$ and $\alpha = \beta/[1-(1-\beta)\gamma]$. This parametrization is not in the form of self-generalizability; that is, $G(G(z;\beta;\eta);\beta',\eta) \ne G(z;\beta\beta',\eta)$. In their Example 3.4, Van Harn et al. (1982) have the **I2** family with a different parameter than γ (Aly and Bouzar, 2005, reparametrized this example in their (3.16)). Thus, the discovery of **I2** has come from different paths.

Note that **I1** and **I2** naturally exist in birth–death processes; see Zhu and Joe (2003). **I3** is obtained by generalizing the key properties of **I1** and **I2**. It is known (from above cited references) that for **I2**, there is a stochastic representation

$K(\alpha)=ZI$ where I,Z are independent random variables with I being Bernoulli(α) and Z being Geometric($(1-\gamma)/(1-\alpha\gamma)$) on the positive integers. This representation can be useful for simulation of $K(\alpha)$. For **I3**, the theory in Section 4 is used for simulation of $K(\alpha)$.

The class of possible margins for stationary times series of the form (1.2) depends on the self-generalized family $\{K(\alpha)\}$. For binomial thinning, the class is called discrete self-decomposable (DSD), and we generalize this in the next definition.

Definition 2.4 (*Generalized discrete self-decomposable, GDSD*). Let $\{K(\alpha)\}$ be a family of self-generalized random variables. Suppose X is a non-negative integer random variable with cdf F_X and pgf G_X. Also suppose that, for each $0\le\alpha\le1$, there exists a non-negative random variable $\varepsilon(\alpha)$ such that

$$X \stackrel{d}{=} K(\alpha)\circledast X+\varepsilon(\alpha)=(\alpha)_K\circledast X+\varepsilon(\alpha),$$

where $\varepsilon(\alpha)$ is independent of X and the random operator. Then F_X or X is said to be generalized discrete self-decomposable **(GDSD)** with respect to $\{K(\alpha)\}$. The pgf of $K(\alpha)\circledast X$ is $G_X(G_K(z;\alpha))$ so equivalently, for X to be GDSD, for every $\alpha\in[0,1]$, $G_X(z)/G_X(G_K(z;\alpha))$ must be a pgf and this would be the pgf of $\varepsilon(\alpha)$.

When referring to one of the specific families of self-generalized operators in Example 2.1, we show the label of the family, such as GDSD(**I2**(γ)).

Definition 2.4 implies that the innovation $\varepsilon_t(\alpha)$ in the stationary model (1.2) is determined by marginal distribution. Thus, to specify a model of the form (1.2), it suffices to specify an appropriate marginal distribution instead of innovation. This is more direct in modelling time series because we usually have an idea of the marginal distribution from the histogram of the data.

Since the innovation $\varepsilon_t(\alpha)$ is non-negative, GDSD implies expectation thinning so that both sides of (1.2) have the same expectation, when this exists. The expectation thinning operation comes intuitively from the current countable units (in $\mathcal{N}_0$) in a dynamic system. Each countable unit may be absent, present or split into more than one new unit in the next time point.

In the remainder of this section, we give some general properties of (1.2). We derive the autocorrelation function, and conditional mean and variance in the case of existing second moments. Several models of form (1.2) with different expectation thinning operations can have the same marginal distribution such as negative binomial. We would like to differentiate the models as a guideline in the choice of an appropriate model.

Note that the marginal distribution of (1.2) may not have mean or variance. The results below are for those models with marginal distributions and $\{K(\alpha)\}$ having finite second moments, and satisfying $\mathrm{E}[K(\alpha)]=\alpha$.

Suppose in (1.2), X_t has finite mean A and variance V. Taking expectation and variance on both sides of (1.2), and using results in Appendix A.1, lead to

$$\begin{aligned}A&=\mathrm{E}[K(\alpha)]\cdot A+\mathrm{E}[\varepsilon_t(\alpha)]=\alpha A+\mathrm{E}[\varepsilon_t(\alpha)],\\ V&=\mathrm{Var}(\mathrm{E}[(\alpha)_K\circledast X_{t-1}|X_{t-1}])+\mathrm{E}(\mathrm{Var}[(\alpha)_K\circledast X_{t-1}|X_{t-1}])+\mathrm{Var}[\varepsilon_t(\alpha)]\\ &=\mathrm{Var}(X_{t-1}\alpha)+\mathrm{E}(X_{t-1}\mathrm{Var}[K(\alpha)])+\mathrm{Var}[\varepsilon_t(\alpha)]\\ &=V\cdot\alpha^2+A\cdot\mathrm{Var}[K(\alpha)]+\mathrm{Var}[\varepsilon_t(\alpha)].\end{aligned}$$

Thus, the expectation and variance of the innovation are

$$\mathrm{E}[\varepsilon_t(\alpha)]=A(1-\alpha),\quad \mathrm{Var}[\varepsilon_t(\alpha)]=V\cdot(1-\alpha^2)-A\cdot\mathrm{Var}[K(\alpha)].$$

Conditioned on X_{t-1}, the stochastic representation of (1.2) becomes

$$[X_t|X_{t-1}=x] \stackrel{d}{=} (\alpha)_K\circledast x+\varepsilon_t(\alpha). \tag{2.3}$$

Thus, the conditional mean and variance are

$$\mathrm{E}[X_t|X_{t-1}=x]=\mathrm{E}[(\alpha)_K\circledast x]+\mathrm{E}[\varepsilon_t(\alpha)]=A+(x-A)\cdot\alpha, \tag{2.4}$$

$$\mathrm{Var}[X_t|X_{t-1}=x]=\mathrm{Var}[(\alpha)_K\circledast x]+\mathrm{Var}[\varepsilon_t(\alpha)]=V\cdot(1-\alpha^2)+(x-A)\cdot\mathrm{Var}[K(\alpha)]. \tag{2.5}$$

These show that the conditional mean and conditional variance depend on the marginal mean and variance, as well as the mean and variance of the self-generalized random variable $K(\alpha)$. Also, both the conditional mean and variance are linear in the conditional value x. Hence if we consider a plot of x_t versus x_{t-1} for a realization of (1.2), it would have a funnel shape which gets wider as x_{t-1} increases. The degree of widening of the funnel is governed by $\mathrm{Var}[K(\alpha))]$. Different families of

Table 1
Variances of non-negative integer self-generalized random variables $K(\alpha)$.

$K(\alpha)$	Variance
I1	$\alpha(1-\alpha)$
I2	$\alpha(1-\alpha)(1+\gamma)/(1-\gamma)$
I3	$\alpha(1-\alpha)(1+\delta)$

self-generalizable random variables may have different variances, and will lead to a different degree of heteroscedasticity in the scatterplot of lag 1 pairs. For reference for subsequent sections, Table 1 summarizes the variances of the families of self-generalized random variables in Example 2.1.

The covariance of X_{t-1} and X_t, via (2.4), is

$$\mathrm{Cov}\,[X_{t-1},X_t] = \mathrm{E}\,\{\mathrm{E}\,[(X_{t-1}-A)\cdot(X_t-A)|X_{t-1}=x]\} = \alpha\cdot V$$

and hence $\mathrm{Corr}\,(X_t,X_{t-1})=\alpha$ when second moments exist. Since conditional on X_{t-k}, X_t has stochastic representation

$$X_t \overset{d}{=} (\alpha^k)_K \circledast X_{t-k} + \varepsilon_t'\,(\alpha^k)$$

the lag k autocorrelation is α^k for $k>1$. Thus, autocorrelation function $\rho(X_{t-k},X_t)=\mathrm{Corr}\,(X_{t-k},X_t)=\alpha^k(k\geq 1)$, is geometrically decreasing as lag increases.

Related to the conditional distribution of X_t given X_{t-1} is the innovation. The behavior of innovation also influences the behavior of the sequence, because conditional on $X_{t-1}=x$, X_t is the convolution of the dependent term $(\alpha)_K \circledast x$ and the innovation ε_t.

Historically, Steutel and van Harn (1979) first proposed a class of operations in an abstract way that generalize binomial thinning. These operations must satisfy five requirements Van Harn et al., 1982, (2.3a)–(2.3e). See also Steutel and van Harn (2004); they gave two specific examples of such operations which are both compounding operations, one with summands being Bernoulli rv's (corresponding to **I1**) and the other one with summands being zero-inflated geometric rv's (corresponding to **I2** with $\gamma=1/3$; see Example 8.2 on Page 303). However, in general, the explicit form of such defined operations is not available. Steutel and van Harn originally introduced the concept of F-self-decomposibility based on these operations in Van Harn et al. (1982). Later, Steutel et al. (1983), Van Harn and Steutel (1985, 1993) applied the F-self-decomposibility to construct integer-valued Markov processes. Applications of generalized Steutel and van Harn operators to integer-valued autoregressive time series of order p are given in Latour (1998, Section 4) and Aly and Bouzar (2005); in particular, Aly and Bouzar (2005) use a different parametrization of our **I2** family as a specific example for their general theory.

In the study of modelling unequally-spaced count data time series, we observed two special cases in the linear birth–death processes proposed by Kendall (1948); see Zhu (2002) and Zhu and Joe (2003). Starting from these two natural operators and specific continuous-time first-order Markov processes, we independently discover the necessary multiplicative property which we call self-generalizability in terms of explicit compounding operations. We have shown that our conditions on the compounding operation are equivalent to the requirements given by van Van Harn et al. (1982). Thus, GDSD is the same as F-self-decomposability.

Note that the survey paper Weiß (2008b) includes a list of references on thinning operations and count data time series models.

3. Negative binomial series

In this section, we summarize results for negative binomial and **I1–I3**, with that for **I3** being new. The negative binomial is a common distribution to use for count data that are overdispersed relative to the Poisson distribution. Other count distributions can also be GDSD but we will defer a general discussion to future research. In the next section, we show that likelihood inference for (1.2) is possible by numerical inversion of pgf's.

Using the operators in Example 2.1, we have some parametric families of negative binomial time series models with a geometric autocorrelation function, and our models include previously studied families at the boundary cases of one of the parameters.

Consider the negative binomial $\mathrm{NB}(\theta,p)$ distribution with pgf $G(z)=[p/(1-qz)]^\theta$ and probability mass function (pmf) $f(x)=\Gamma(\theta+x-1)p^\theta q^x/[\Gamma(\theta)x!]$, $x\in\mathcal{N}_0$, where $q=1-p$. This is well known to be DSD (namely GDSD(**I1**)) and the corresponding time series model with binomial thinning has been discussed in McKenzie (1986). It is shown in Zhu and Joe (2003) and Aly and Bouzar (1994) that the $\mathrm{NB}(\theta,p)$ distribution is GDSD(**I2**(γ)) for $0\leq\gamma\leq q$. For the boundary case GDSD(**I2**(q)), Al-Osh and Aly (1992) study this time series with a different stochastic representation than (1.2). Using Theorem A.2, the $\mathrm{NB}(\theta,p)$ distribution is GDSD(**I3**(δ)) for $0\leq\delta\leq q/p$. The proof of this result is given in the following.

Substituting the NB mean and variance into (2.5) leads to $\mathrm{Var}\,(X_t|X_{t-1}=x)=\theta qp^{-2}(1-\alpha^2)+(x-\theta qp^{-1})\mathrm{Var}\,[K(\alpha)]$, which is linearly increasing in x. The magnitude of the slope $\mathrm{Var}\,[K(\alpha)]$ affects the conditional heteroscedasticity. The model with most and least conditional heteroscedasticity respectively have operators **I2**(q) [$\mathrm{Var}\,[K(\alpha)]=\alpha(1-\alpha)(1+q)/p$] and **I1** (or **I2**(0) or **I3**(0)). There is increasing conditional heteroscedasticity as γ or δ increases for the time series models from **I2**(γ) and **I3**(δ). That is, for families of self-generalized random variables $K(\alpha)$ indexed by other parameters, the other parameter affect the range of conditional variances.

For convenience, we summarize these properties of the negative binomial time series models in Table 2.

Another property that can be compared for the different negative binomial time series models is reversibility. Reversibility holds if the conditional distribution of X_t given $X_{t-1}=x$ is the same as that of X_{t-1} given $X_t=x$, or if the joint pmf $f_{X_1,X_2}(x_1,x_2)$ is symmetric in its two arguments, or if the joint pgf G_{X_1,X_2} satisfies $G_{X_1,X_2}(z_1,z_2)=G_{X_1,X_2}(z_2,z_1)$. The

Table 2
Features of time series with NB (θ, p) margins.

K	Parameter range	Heteroscedasticity
I1	–	Least
I2	$\gamma \le 1-p$	Most
I3	$\delta \le (1-p)/p$	Medium

joint pgf is

$$G_{X_1,X_2}(z_1,z_2) = G_X(z_1\, G_K(z_2;\alpha))\, G_X(z_2)/G_X(G_K(z_2;\alpha))$$

so by substituting the negative binomial pgf and the pgf's given in Example 2.1, it can be shown that reversibility only holds for $\mathbf{I2}(q)$.

Negative binomial time series based on compounding operators that are not self-generalized can also be constructed; an example is given in Example 3.1 of Latour (1998). However in this case, not all of the properties mentioned in Section 2 are satisfied.

3.1. Proof for I3 for negative binomial

We show that NB(θ, p) is in GDSD$(\mathbf{I3}(\delta))$ for $\delta \le q/p$, using Theorem A.2; note that sometimes complex arguments must be used to establish that a function is a pgf.

Letting $c = \delta/(1+\delta)$, the condition is $0 \le c \le q$, and $H^*(z) = (z - c^{-1})[\log(1-cz) - \log(1-c)]$ where $0 \le c < 1$.

Since $G(z) = p^\theta/(1-qz)^\theta$, then $G'(z)/G(z) = \theta q/(1-qz)$, and

$$\begin{aligned} H^*(z)G'(z)/G(z) &= \frac{\theta q}{c}\frac{1-cz}{1-qz}[\log(1-c)-\log(1-cz)] = \frac{\theta q}{c}\frac{1-cz}{1-qz}\left[-\int_0^c \frac{1}{1-\lambda}d\lambda + z\int_0^c \frac{1}{1-\lambda z}d\lambda\right] \\ &= \frac{\theta q}{c}\int_0^c (1-cz)\left[\frac{1-q}{q-\lambda}\frac{1}{1-qz} - \frac{1-\lambda}{q-\lambda}\frac{1}{1-\lambda z}\right]\frac{1}{1-\lambda}d\lambda \\ &= \frac{\theta q}{c}\int_0^c \left\{-\frac{c}{q\lambda} + \frac{(1-q)(q-c)}{q(q-\lambda)}\frac{1}{1-qz} + \frac{(1-\lambda)(c-\lambda)}{\lambda(q-\lambda)}\frac{1}{1-\lambda z}\right\}\frac{1}{1-\lambda}d\lambda. \end{aligned}$$

Because $0 \le \lambda \le c \le q \le 1$, $(1-q)(q-c)/q(q-\lambda) \ge 0$ and $(1-\lambda)(c-\lambda)/\lambda(q-\lambda) \ge 0$. Thus, the coefficients of z^i for $i \ge 1$ in the series expansion of the integrand are all non-negative. Therefore, $1 + C \cdot H^*(z)G'(z)/G(z)$ is a pgf if $1 + C \cdot H^*(0)G'(0)/G(0) \ge 0$ or $C \le c/[-\theta q \log(1-c)]$. According to Theorem A.2, we have the conclusion.

4. Computational details for simulation, likelihood inference and diagnostics

In Section 4.1, we show how simulation and likelihood inference can be performed for a model specified with closed form pgf's but not closed form pmf's. In Section 4.2, the conditional pmf's and pmf's of the innovations, computed using the technique of Section 4.1, are compared for the various NB time series models. In Section 4.3, a graphical diagnostic method is introduced for assessing the fit of the models.

4.1. Inversion of characteristic function for a non-negative integer-valued random variable

For time series models which derive from the theory of expectation thinning operators, conditional probability distributions have a simple form for the pgf but not pmf. In this section we show that simulation and likelihood inference for a model of the form (1.2) is numerically feasible, even if only closed form pgfs are available; the theory extends to other models of the form (1.2) where X is not negative binomial but has a closed form pgf. For the binomial thinning operator, a direct recursive algorithm is given in Zhu and Joe (2006), but this algorithm would be too time-consuming to implement for **I2** and **I3** because the random variables $\{K(\alpha)\}$ have support on all non-negative integers and not just $\{0,1\}$.

For a non-negative integer random variable Y with pgf $G_Y(z) = \mathrm{E}(z^Y)$, its characteristic function (chf) is $\varphi_Y(t) = \mathrm{E}(e^{itY}) = G_Y(e^{it})$. Davies (1973) has a result from which one can obtain a pmf for a non-negative integer-valued random variable based on the chf. The steps are the following.

(a) Let

$$a(y) \stackrel{def}{=} \frac{1}{2} - (2\pi)^{-1}\int_{-\pi}^{\pi} \mathrm{Re}\left(\frac{\varphi_Y(u)e^{-iuy}}{1-e^{-iu}}\right)du. \tag{4.1}$$

Then $a(y) = \Pr(Y < y)$ is the cdf.

(b) The pmf of non-negative integer-valued random variable Y is

$$f_Y(0)=a(1) \quad \text{and} \quad f_Y(y)=a(y+1)-a(y), \quad y=1,2,\ldots.$$

Numerically, (4.1) can be computed as the sum of numerical integrals with regions $(0,\pi)$ and $(-\pi,0)$. Applying this inversion method to some cases of NB(θ, p), we have found that the approximation error is smaller than 10^{-8} for both the cdf and pmf. Hence, we consider the algorithm to be adequate for applications.

For an observed time series $\{x_t: t=1,\ldots,n\}$, for application of (1.2), the likelihood is

$$f_{X_1}(x_1)\prod_{t=2}^{n} f_{X_t|X_{t-1}}(x_t|x_{t-1}) = f_{X_1}(x_1)\prod_{t=2}^{n} f_{X_2|X_1}(x_t|x_{t-1}). \tag{4.2}$$

The above algorithm is applied with $Y \equiv [X_t|X_{t-1}=x_{prev}]$, where $x_{prev}$is used generically for the previous observation. The pgf of Y is

$$G_Y(z)=G_K^{x_{prev}}(z;\alpha)\cdot\frac{G_X(z)}{G_X(G_K(z;\alpha))}. \tag{4.3}$$

Also for simulation of (1.2), the pgf (4.3) can be inverted, or if one could separately simulate the two components (i) innovation ε_t with pgf $G_X(z)/G_X(G_K(z;\alpha))$, (ii) the dependence on the previous observation $(\alpha)_K \circledast x_{prev}$ with pgf $G_K^{x_{prev}}(z;\alpha)$.

Given fixed parameters, θ, p, α, the likelihood in (4.2) can be computed. The negative logarithm of (4.2) can be inputted into a quasi-Newton numerical optimization method in order to get the maximum likelihood estimate (MLE) and an inverse Hessian of the negative log-likelihood. For example, if R (http://www.r-project.org) is used, the MLE can be obtained using either `optim()` or `nlm()`, and `integrate()` for the numerical integrations in (4.1). Based on large sample theory, the log-likelihood behaves like a quadratic near the MLE if the sample size is large enough and the parameter vector is not on the boundary of the parameter space. In practice, for uniqueness, it is checked that different starting points for the quasi-Newton iterations lead to the same convergent solution in the interior of the parameter space. The method of moments estimator (or something close to it) is usually a good starting point for quasi-Newton iterations.

For initial choices of values $(\theta_0, p_0, \alpha_0)$ of parameters for numerical maximum likelihood in three NB count data time series models, we use marginal moments to estimate the parameters of the marginal distribution and the serial dependence. Specifically, assuming $s^2/\overline{x}>1$,

$$p_0=\overline{x}/s^2, \quad \theta_0=\overline{x}p_0/(1-p_0),$$

where $\overline{x}$ and s^2 are the sample mean and variance of the observations, and α_0 is serial correlation of lag 1. For the additional parameter γ in **I2** and δ in **I3**, since $0<\gamma<(1-p)$ and $0<\delta<(1-p)/p$, we can set $\gamma_0=C_0(1-p_0)$ and $\delta_0=C_0(1-p_0)/p_0$, where C_0 is a number in (0, 1), say 0.5.

We simulate many sequences of various sizes such as 100, 200, 500, 1000 for each of NB count data time series model based on **I1**–**I3** separately to check numerical methods. The lag one scatter plots show strong association between the current observation and the previous observation, indicating the serial dependence. The autocorrelation function (ACF) plots show the geometrically decreasing pattern for the first few lags. Interestingly, there is always just one large positive value at lag one in the partial ACF plots, similar to Gaussian AR(1) models. Our experience shows that if simulating from **I2** or **I3** with γ or δ in the mid-range, one can easily estimate γ or δ at a boundary value, unless sample size is in the hundreds.

As to the asymptotic normality of the MLE, Billingsley (1961) established the fundamental results for discrete-time Markov processes using martingale theory. The distributions we are using are well behaved with derivatives satisfying the required conditions.

4.2. Graphical comparisons of NB time series models

In this subsection, we compare the pmf's for the three types of expectation thinnings. For illustration, we choose $\alpha=0.2$ and 0.8. Fig. 1 shows the probability masses of the random variable $K(\alpha)$ for **I2** (left column) and **I3** (right column) with γ and δ varying over their ranges. Note that $\gamma=0$ and $\delta=0$ lead to **I1**. The other boundary case in two families, when γ and δ attain their upper bounds, corresponds to the degenerate distribution with probability mass one at point zero. For a specific marginal distribution like negative binomial, these upper bounds can not be reached, because of constraints for GDSD. Although **I2** and **I3** allow their random variables $K(\alpha)$ to take integer values larger than one, essentially, the $K(\alpha)$' s take values 0 and 1 with large probabilities, and have small probabilities for values larger than 1. From Fig. 1, we can see that these two families behave very similarly. The difference is very small, but **I2** shows more variability than **I3**. However, further investigation shows that the probability differences among the summations of three types of thinnings will become large when the number of summands increases.

Probability masses of the innovations in the NB time series models based on **I1**–**I3** are of interest. They are shown in Fig. 2. The innovation probabilities have larger variability in **I2** than in **I3**.

Fig. 3 compares the conditional probability masses with the marginal NB(2.5, 0.4) probabilities for NB time series based on **I2** when $\gamma=0.3$ and $\alpha=0.5$. The difference between the marginal and conditional probabilities is very large. As the value of the previous observation increases, the curve shifts to right, implying the increase of conditional mean.

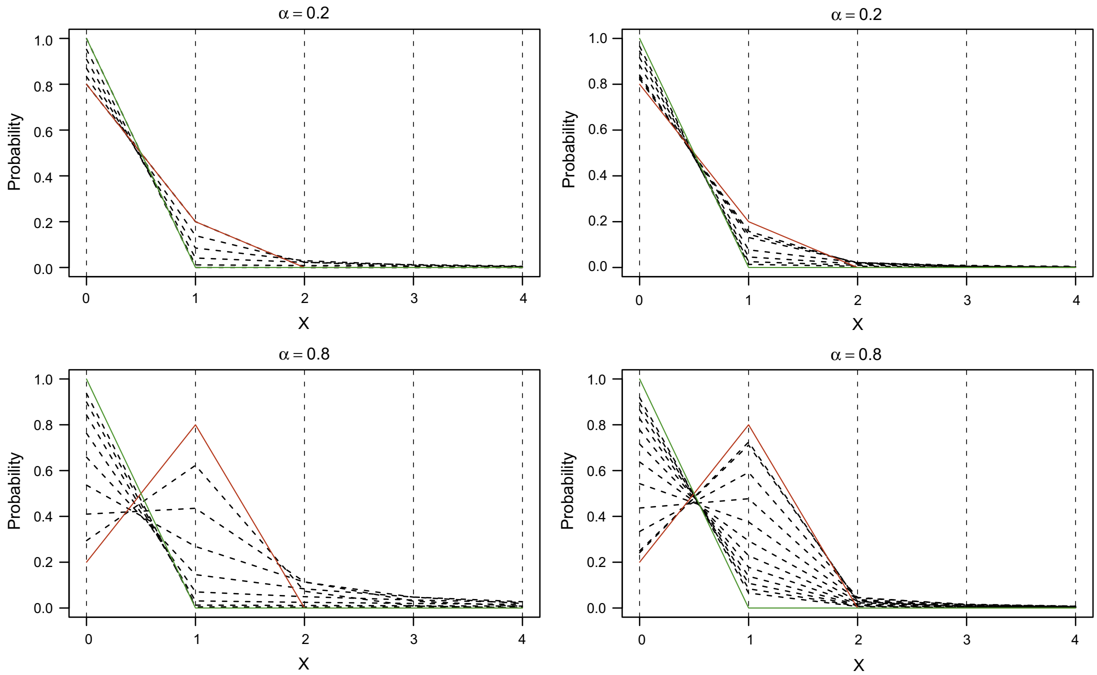

Fig. 1. Probability masses of the random variable $K(\alpha)$ for **I2** and **I3**. The serial dependence parameter $\boldsymbol{\alpha}$ is set as 0.2 (1st row) and 0.8 (2nd row). (1) Left column: **I2** with $\boldsymbol{\gamma}$ varying from 0 to 1. (2) Right column: **I3** with $\boldsymbol{\delta}$ varying from 0 to ∞. Two boundary cases are: $\gamma=0$ or $\delta=0$, degenerating to **I1** (red line) and $\boldsymbol{\gamma}=1$ or $\delta\to\infty$, degenerating to point zero (green line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

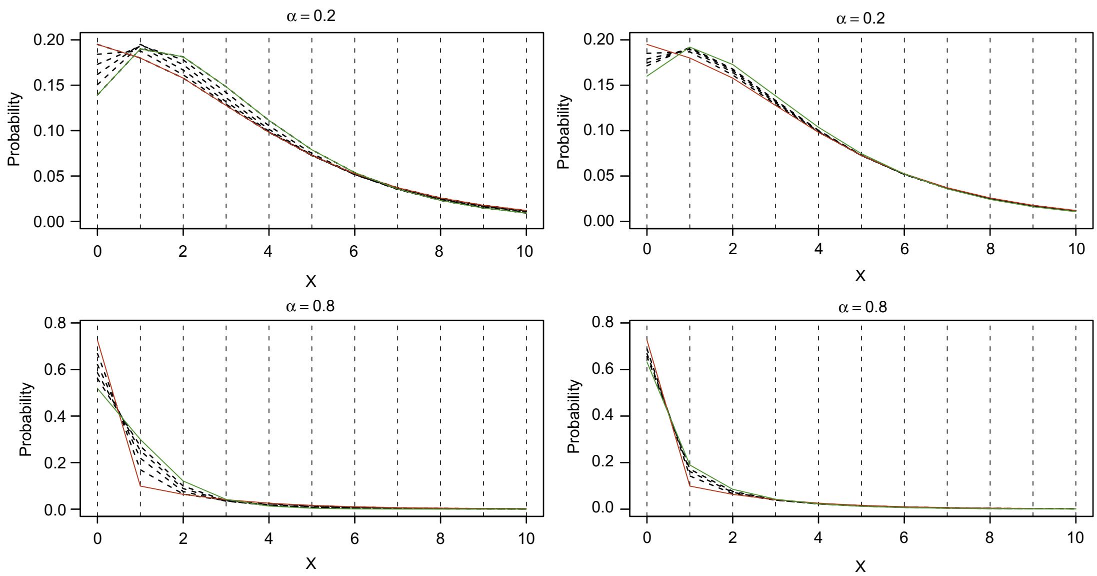

Fig. 2. Probability masses of the innovations of NB time series based on **I2** and **I3**. The marginal distribution is NB(2.5, 0.4), and the parameter $\boldsymbol{\alpha}$ takes on the values 0.2 (1st row) and 0.8 (2nd row). (1) Left column: **I2** with $\boldsymbol{\gamma}$ varying from 0 to $(1-0.4)$=0.6. (2) Right column: **I3** with $\boldsymbol{\delta}$ varying from 0 to $(1-0.4)/0.4$=1.5. Two boundary cases are: $\gamma=0$ or $\delta=0$, leading to **I1** (red line) and $\boldsymbol{\gamma}=1-p=0.6$ or $\delta=(1-p)/p=1.5$, the upper bound (green line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 4 compares the conditional probability masses for NB(2.5, 0.4) time series models based on **I1**, **I2** and **I3** when $\alpha=0.5$ and γ,δ vary over their ranges. The previous observation is set as 0 (first column), 2 (second column) and 10 (third column). Subplots in the third row indicate that the varying region of conditional probability profiles of models based on **I2**

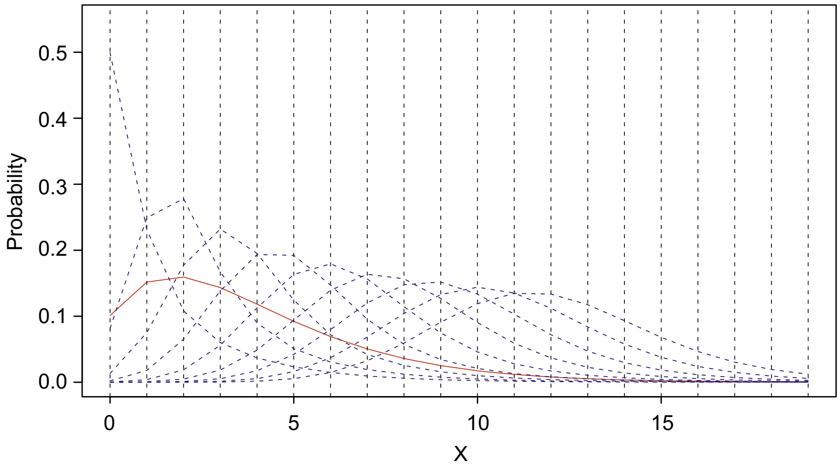

Fig. 3. Comparison of marginal and conditional probability masses for the NB time series model based on **I2**. The marginal distribution is NB(2.5, 0.4) (red solid line). $\gamma=0.3$ and $\alpha=0.5$. The conditional previous observation varies among {0, 2, 4, 6, 8, 10, 12, 14, 16}, corresponding to profiles with peaks moving from the left to the right (blue dotted lines). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

covers that of models based on **I3**. Therefore, NB time series models based on **I2** can accommodate for more conditional heteroscedasticity than **I1** and **I3**.

4.3. Diagnostics

In this subsection, we mention some graphical diagnostics of fitted models. We suggest the (jittered) QQ plot for transformed observations based on marginal and conditional cdf plotted against the standard uniform distribution. In general, let $F(\cdot)$ be the cdf of a discrete random variable Y with support on $\mathcal{N}_0$. Define the transform of Y through the following conditioning:

$$Z|Y=y\sim \mathrm{Uniform}(F(y-1),F(y)). \tag{4.4}$$

The marginal distribution of Z is the standard uniform. For a given observation y_t ($1\le t\le N$), its transformed observation is defined as z_t which is an outcome of random variable given in (4.4) with $y=y_t$. Note that the transformed observations vary from one realization to another. If the fitted model is close to the true model, then the transformed observations against their standard uniform quantiles of empirical distribution, i.e., points

$$(z_{(i:N)}, i/N), \quad i=1,\ldots,N$$

will scatter around the diagonal line. Here $z_{(i:N)}$ indicates the i th smallest value among the N z_t's. If the fitted model is far from the true model, then a departure from the diagonal line will appear.

In count data time series, each observation (except the first one) has dual roles: as a marginal observation and as a conditional observation (conditioned on the previous observation). Thus, we can transform the observations according to both the marginal cdf and conditional cdf. When the sample size is not large, in general, the sample of the latter looks closer to the diagonal line than that of the former. This diagnostic tool is applied to the data example in the next section in Fig. 7.

5. Data example

In this section, we compare the NB time series models based on **I1**–**I3** for some real data. We choose a data set for which from the context, one might expect a model based on **I2** or **I3** to be more reasonable than binomial thinning; binomial thinning should be better in a context where a contribution to a count (such as numbered injured) at time t has a probability between 0 and 1 of contributing at time $t+1$.

The Research Papers in Economics (RePEc) web site (http://repec.org/) provides online access statistics of working papers at http://logec.repec.org/. For each working paper, the access statistics include monthly counts of views of the abstract and downloads of the working paper. Graphic displays are also provided in the website. We select the one titled “Cycles of Learning in the Centipede Game” whose abstract views over the years seems to be a stationary count time series. The sequence started from August 1998 and we use the monthly data up to May 2008, so the length of the time series is 118.

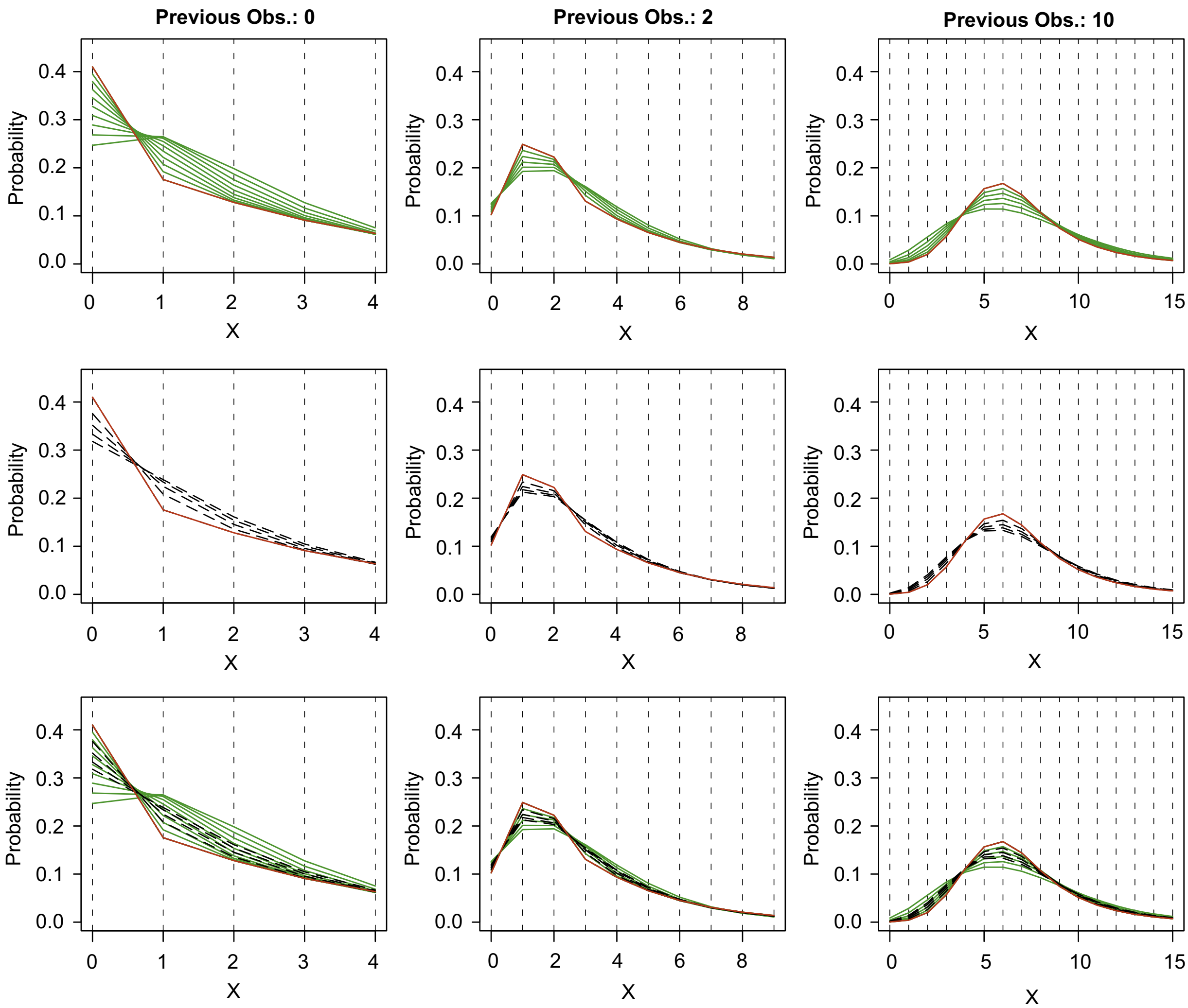

Fig. 4. Comparison of conditional probability masses/profiles for NB time series model based on **I1** (red solid lines), **I2** with γ from 0 to 0.6 (green lines) and **I3** with δ from 0 to 1.5 (black dashed lines). The marginal distribution is NB(2.5, 0.4), and $\alpha = 0.5$. Previous observations are 0, 2, 10 for columns 1, 2 and 3 respectively. The last row combines the first two rows for easier visual comparisons. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

For this sequence, the plot of monthly counts of abstract views against time is displayed in the top subplot of Fig. 5. The sample mean and variance of the monthly counts are 18.8 and 176.3 respectively. Thus, the ratio of variance to mean is 9.4, indicating relatively large overdispersion relative to Poisson.

The lag one scatterplot of monthly counts is shown in the bottom subplot of Fig. 5. An obvious linear association stands out. In addition, the lag one autocorrelation coefficient is 0.68, suggesting moderate serial dependence. The ACF and partial ACF plots are drawn in Fig. 6. The dotted lines in these two plots indicate the 95% asymptotic confidence intervals based on an assumption of Gaussian time series, and the bounds are smaller than 0.2. Here we use them as rough references to judge if some of autocorrelations and partial autocorrelations are non-zero. There is a geometric decrease for the first seven lags in the ACF plot (top subplot), and a large positive value at lag one in the partial ACF plot (bottom subplot). The geometric decrease is slow, implying that the value of true serial dependence is most likely larger than 0.5. These patterns suggest that a count data time series of the form of (1.2) is reasonable for this data set.

Since overdispersion exists, we use negative binomial as the marginal distribution for the count data time series, and fit all three types of expectation thinnings using maximum likelihood. For model choice, we use AIC $(= -2\times \log - \text{likelihood} + 2\times[\text{of model parameters}])$ as the criterion; the smaller the AIC, the better the fitted model.

When fitting NB time series models based on **I2** and **I3**, we have found that the parameters γ and δ almost reach their upper bounds. This suggests setting the additional parameter γ in **I2** and δ in **I3** to their upper bounds. In this situation, these two models will have four parameters instead of five. Thus, for model comparisons, with a table of log-likelihoods,

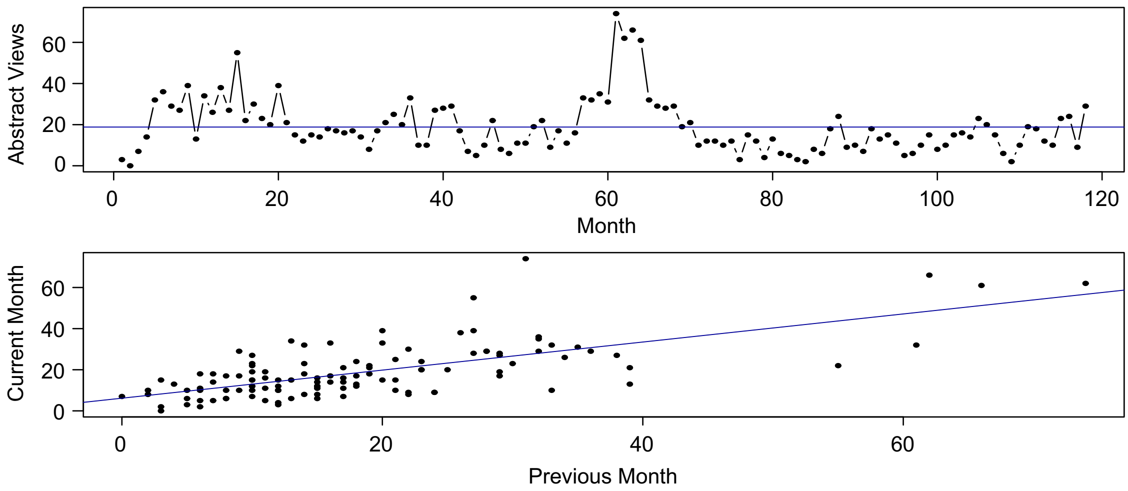

Fig. 5. Top: time series plot of the monthly counts of abstract views. The (blue) horizontal line indicates the sample mean. Bottom: lag one scatterplot of monthly counts of abstract views. The (blue) line is the fitted regression line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

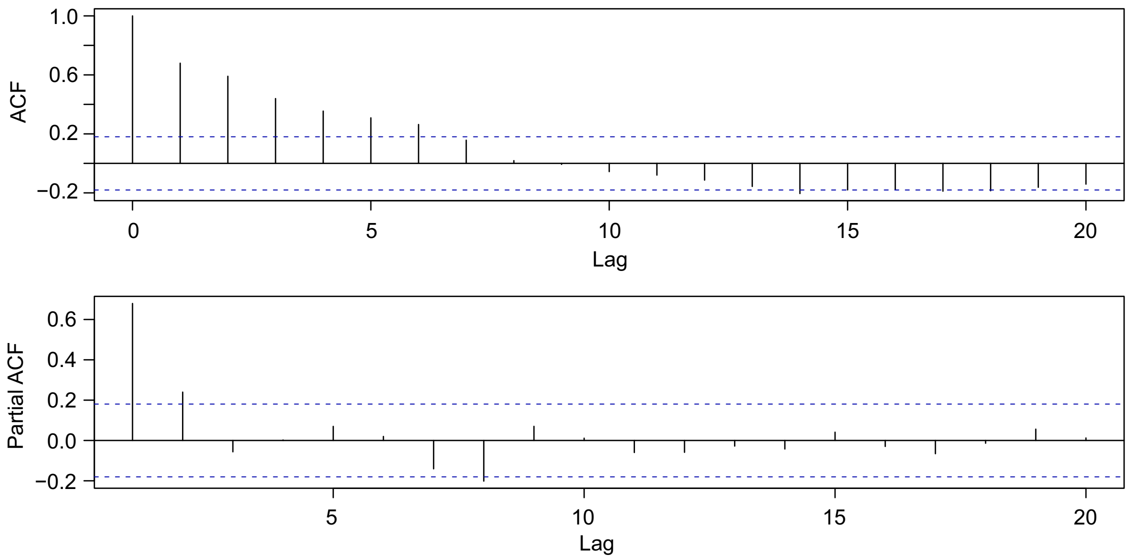

Fig. 6. The ACF and partial ACF plots of the monthly counts of abstract views.

one can easily compare the AIC of **I1** versus **I2** and **I3** by checking if the log-likelihoods for latter are at least larger by a value of 2.

The MLEs and log-likelihoods are summarized in Table 3. For **I1**, the results with (4.1) are the same as those from the algorithm given in Zhu and Joe (2006). Based on the log-likelihoods or the corresponding AIC values, the model with **I2** is the best, and the model with **I1** is the worst. Also the estimated α from **I2** is the closest to the serial correlation of lag one. Graphical diagnostics are displayed in Fig. 7 for NB time series model based on **I2** at its upper bound value of γ. These QQ plots show a good fit of the model.

To assess the model for 1-step predictions, we propose two methods: (1) rounded conditional mean and (2) conditional mode. Fig. 8 illustrates one-step predictions using both methods for the best fitted NB time series model based on **I2**. Roughly, both predictions follow the fluctuation of true observations over time. Prediction intervals can be constructed from the fitted conditional probability mass functions.

To evaluate overall performance of a prediction method, and also judge the models based on prediction performance, we compute the average of absolute differences between observations and predictions for all three fitted models in Table 4. The average differences for rounded conditional mean method are always smaller than those for conditional mode method. This could imply that the rounded conditional mean method may be better than the conditional mode method, at least in

Table 3
MLEs and loglikelihoods for NB time series models based on **I1**, **I2** with upper bound $\gamma = 1-p$, and **I3** with upper bound $\delta = (1-p)/p$.

Margin	Expectation thinning	MLEs (SEs)	Log-likelihood
NB	**I1**	$\hat{\theta} = 3.460$ (0.545) $\hat{p} = 0.156$ (0.021) $\hat{\alpha} = 0.430$ (0.037)	−421.8
NB	Upper bound **I2**	$\hat{\theta} = 2.576$ (0.569) $\hat{p} = 0.121$ (0.028) $\hat{\alpha} = 0.678$ (0.072) $\hat{\gamma} = 0.878$	−414.3
NB	Upper bound **I3**	$\hat{\theta} = 2.870$ (0.559) $\hat{p} = 0.135$ (0.024) $\hat{\alpha} = 0.588$ (0.051) $\hat{\delta} = 6.407$	−415.8

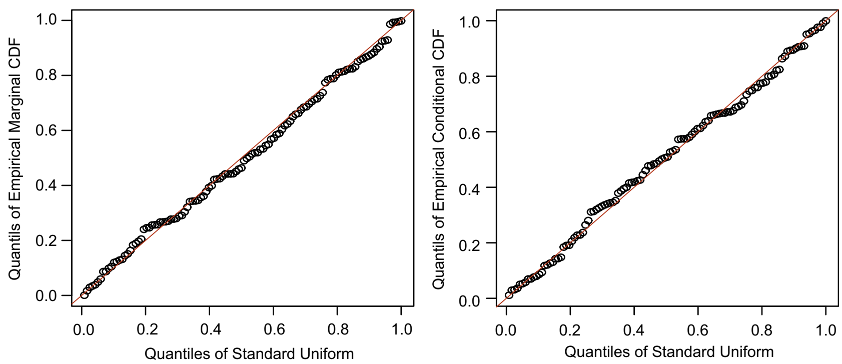

Fig. 7. Model is **I2** with upper bound value of γ. Left: QQ plot of empirical marginal cdf; right: QQ plot of empirical conditional cdf. Diagonal lines are included (in red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

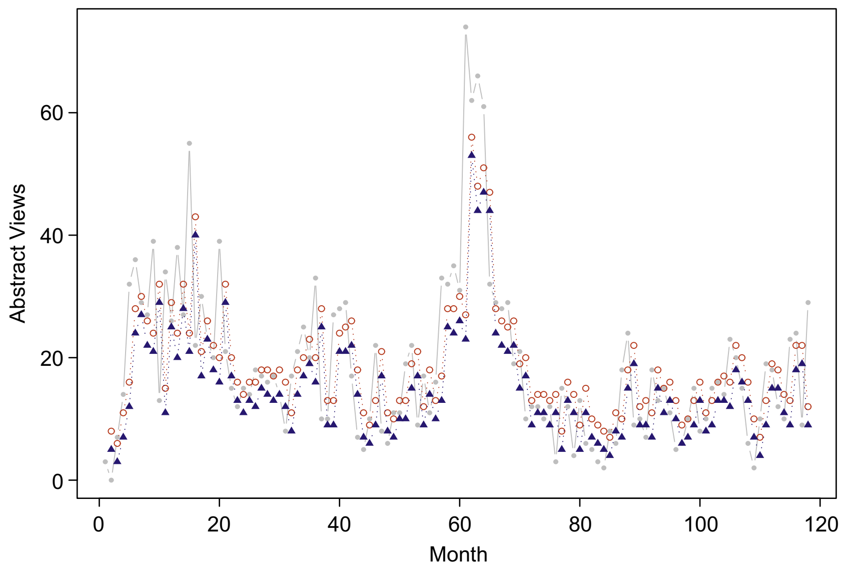

Fig. 8. Predictions using the mean and mode of the conditional distributions from the fitted NB time series model based on **I2** with upper bound γ value. For the predictions, the red circles and blue triangles indicate the rounded conditional mean and the conditional modes respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Average of absolute differences between observations and predictions for fitted NB time series models based on **I1**, **I2** *with upper bound* $\gamma = 1-p$, *and* **I3** *with upper bound* $\delta = (1-p)/p$.

Margin	Expectation thinning	Rounded conditional mean method	Conditional mode method
NB	**I1**	7.40	9.39
NB	Upper bound **I2**	6.86	7.39
NB	Upper bound **I3**	7.04	8.07

this example. For model fitting from the viewpoint of prediction error, **I2** is the best, **I3** is the second best, and **I1** is the worst.

Apart from this particular times series example, we have also downloaded several other series of monthly abstract views from the same web site, and fitted the NB time series models based on **I1**–**I3**. They all have the same modelling tendency, i.e., the parameter γ of **I2** and δ of **I3** tended to their upper bounds, and model **I2** was the best fit. This implies that the research pattern in the discipline represented by those working papers cannot be fitted by the model based on the simpler binomial thinning operator. Other expectation thinnings, which have no restriction on non-integer domain, seem to be better to describe the underlying research spread or conditional heteroscedasticity. Meanwhile, the attaining of the upper bound for γ and δ for both **I2** and **I3** respectively may imply that there is a large conditional heteroscedasticity in the knowledge spread process.

Joe (1996) has a NB time series model, within a class of models with margins that are convolution closed and infinitely divisible; it is based on a random dependence operator that does not have the form of (1.2) but can achieve a different form of conditional heteroscedasticity. For the time series in Fig. 5, the AIC value of the model in Joe (1996) is slightly worse than that of **I2**. However, for other similar count time series at the above-mentioned web site, sometimes (less than 50% of time), **I2** is not better based on AIC.

6. Discussion and further research

We summarized the theory of constructing count data time series models based on expectation thinning operators, and proposed a general numerical algorithm to calculate pmf's so that likelihood inference, graphical diagnostics and predictions are feasible. Data examples showed that binomial thinning may not be able to accommodate very well the conditional heteroscedasticity for the type of time series considered in Section 5. This numerical algorithm makes it possible to compare the MLE and other estimation methods such as the empirical characteristic function method (Rémillard and Theodorescu, 2001) via simulation.

Although we used the negative binomial distribution, there are other marginal distributions in the GDSD class which could be used for overdispersed Poisson count time series. Data analysis could proceed in a similar way as shown in Sections 4 and 5. The marginal distributions in GDSD classes are all infinitely divisible. The parameter constraints between specific marginal distribution and expectation thinning operators need to be established.

Although the new NB time series model based on **I3** was not a better fit than that based on **I2**, we still consider the new **I3** operator and other, to be discovered, expectation thinning operators to be important. For other potential margins for count data, the **I3** operator might lead to more conditional heteroscedasticity than **I2**. Even for a given marginal distribution such as NB, we do not know the theoretical upper bound of conditional heteroscedasticity that is possible among other expectation thinning operators. These topics can be investigated in future research.

There are many types of models other than the one based on expectation thinning for count data time series, and essentially they are first or higher order Markov processes. The developed numerical algorithm can be readily applied to these models for conditional pmf and other relevant pmf's which are then used for inference and simulation.

Acknowledgments

This research has been supported by NSERC Discovery Grants. We appreciate Dr. Nadjib Bouzar for his encouragement and discussion on this paper. In addition, we are grateful to the referees for valuable comments that have led to an improved presentation.

Appendix A

A.1. Properties of self-generalized distributions

We briefly state some properties of self-generalized distributions, which are needed for the results in Section 3 and Appendix A.2.

Theorem A.1. *Suppose* $\{K(\alpha)\}$ *is a family of self-generalized random variables with pgf* $G_K(\cdot;\alpha)$, *and finite mean. Let* $h(\alpha)=\mathrm{E}[K(\alpha)]$ *for* $0\le\alpha\le 1$. *Then* $h(\alpha)h(\alpha')=h(\alpha\alpha')$.

Proof. From (2.1), a stochastic representation is $K(\alpha\alpha')=\sum_{j=1}^{K(\alpha')}K_j(\alpha)$, where $K_j(\alpha)$ are iid copies of $K(\alpha)$, independent of $K(\alpha')$. The conclusion then follows from the identity for conditional expectation. □

We now assume that $\{K(\alpha):0\le\alpha\le 1\}$ is a family of self-generalized random variables leading to expectation thinning operators.

Property A.1. *Suppose* $h(\alpha)=\mathrm{E}[K(\alpha)]$ *is continuous in* α. *Then* $h(\alpha)=\alpha^r$, $r>0$.

Proof. This follows from the Cauchy functional equation. Note that $r=0$ is eliminated to avoid triviality, and $r<0$ is eliminated because of the expectation thinning requirement. □

Without loss of generality, we can assume $r=1$, by a reparametrization of $\{K(\alpha)\}$ (with $\alpha\to\alpha^{1/r}$). The families in Example 2.1 have been reparametrized if their initial parametrization did not have $r=1$.

Property A.2. *Suppose X with support corresponding to* $\{K(\alpha)\}$ *has finite expectation. Assuming* $\mathrm{E}[K(\alpha)]=\alpha^r$ *with* $r=1$, *then* $\mathrm{E}[K(\alpha)\circledast X]=\mathrm{E}[K(\alpha)]\cdot\mathrm{E}[X]=\alpha\mathrm{E}[X]$.

Proof. This follows from the conditional expectation identity. □

Property A.3. $K(0)\equiv 0$ *and* $K(1)\equiv 1$. *Hence* $(0)_K\circledast X\overset{d}{=}0$ *and* $(1)_K\circledast X\overset{d}{=}X$.

Proof. $K(0)\equiv 0$ follows from the preceding property with $X\equiv 1$.

Because of the distinctness requirement for the family $\{F_K(\cdot;\alpha)\}$, non-zero random variables are included; thus $G_K(z;\alpha)$ is increasing in z for some $\alpha\in[0,1]$. With $\alpha=1$ in (2.1),

$$G_K(G_K(z;1);\alpha')=G_K(z;\alpha'),\quad 0<z<1.$$

Hence $G_K(z;1)=z$ because $G_K(\cdot;\alpha')$ are distinct for different α'. This implies that $K(1)\equiv 1$. □

A.2. Characterization results under expectation thinning

If X_t is GDSD with respect to $\{K(\alpha)\}$, then the distribution of the innovation in (1.2) depends on the distributions of X_t and $\{K(\alpha)\}$. Thus the focus is on the marginal distribution and not on the innovation when specifying a model of form (1.2). In this subsection, we give results that can be used to determine if a distribution is GDSD with respect to a particular family of expectation thinning operators $\{K(\alpha)\}$. This leads to the characterization of marginal distributions of (1.2) with respect to $\{K(\alpha)\}$. This result generalizes Theorem 5.1 of Zhu and Joe (2003) for the **I2** operator.

The main result in Theorem A.2 for $\mathcal{N}_0$ makes use of the following mixture operation: if $G(z;\beta)$ is a pgf for $\beta\in B$ and F is a distribution on B, then $\int_B G(z;\beta)\,dF(\beta)$ is a pgf.

Theorem A.2 (*Sufficiency*). *Let X be a non-negative integer random variable with pgf* $G_X(s)$. *Let* $K(\alpha)$ *be a family of self-generalized random variables with pgf* $G_K(\cdot;\alpha)$, *and let* $H^*(z)=\partial G_K(z;\alpha)/\partial\alpha|_{\alpha=1}$. *If* $1+C\cdot H^*(z)G_X'(z)/G_X(z)$ *is a pgf for some* $C>0$, *then X is GDSD with respect to* $\{K(\alpha)\}$.

Proof. Note that

$$H^*(G_K(z;\beta))=\frac{\partial}{\partial\alpha}[G_K(G_K(z;\beta),\alpha)]\Big|_{\alpha}=1=\frac{\partial}{\partial\alpha}[G_K(z;\alpha\beta)]\Big|_{\alpha=1}=\left(\frac{\partial}{\partial\gamma}[G_K(z;\gamma)]|_{\gamma=\beta}\right)\beta=\beta\left(\frac{\partial}{\partial\beta}[G_K(z;\beta)]\right).\tag{A.1}$$

Since $1+C\cdot H^*(z)G_X'(z)/G_X(z)$ is a pgf for some $C>0$, $1+C\cdot H^*(G_K(z;\beta))G_X'(G_K(z;\beta))/G_X(G_K(z;\beta))$ is also a pgf from the compounding operation (Definition 2.1). Applying the mixture operation,

$$g(z)=[-\log c]^{-1}\int_c^1\left(1+C\cdot\frac{H^*(G_K(z;\beta))G_X'(G_K(z;\beta))}{G_X(G_K(z;\beta))}\right)\frac{1}{\beta}d\beta$$

Table 5
Partial derivative H^* of pgf defined in Theorem A.2.

K	H^*
I1	$z-1$
I2	$(1-\gamma z)(z-1)/(1-\gamma)$
I3	$[z-(1+\delta)/\delta]\log(1+\delta-\delta z)$

is a pgf. By algebra, Eq. (A.1) and Property A.3, we obtain

$$g(z) = [-\log c]^{-1}\left\{ C\int_c^1 \frac{\frac{\partial G_K(z;\beta)}{\partial \beta} G_X'\,(G_K(z;\beta))}{G_X(G_K(z;\beta))}\,d\beta + \int_c^1 d(\log\beta)\right\} = [-\log c]^{-1} C \log\frac{G_X(z)}{G_X(G_K(z;c))} + 1.$$

Thus, $G_X(z)/G_X(G_K(z;c)) = \exp\{\log(c^{-1})C^{-1}[g(z)-1]\}$, the pgf of a compound Poisson distribution. Hence, X is GDSD. □

The condition in the preceding theorem is sufficient for GDSD. In fact, it is also necessary. But the converse result is not needed in Section 3.

To help in the checking of the condition for GDSD for other marginal distributions, Table 5 consists of the function H^* for the families in Example 2.1.

References

Al-Osh, M.A., Aly, E.A.A., 1992. First order autoregressive time series with negative binomial and geometric marginals. Comm. Statist. Theory Methods 21, 2483–2492.

Aly, E.A.A., Bouzar, N., 1994. Explicit stationary distributions for some Galton–Watson processes with immigration. Comm. Statist. Stoch. Models 10, 499–517.

Aly, E.A.A., Bouzar, N., 2005. Stationary solutions for integer-valued autoregressive processes. Int. J. Math. Math. Sci. 1, 1–18.

Aly, E.A.A., Bouzar, N., 2008. On a class of Z_+-valued autoregressive moving average (ARMA) processes. REVSTAT—Statist. J. 6, 101–121.

Billingsley, P., 1961. Statistical Inference for Markov Processes. In: Statistical Research Monographs, vol. II. The University of Chicago Press, Chicago, IL.

Bu, R., McCabe, B., 2008. Model selection estimation and forecasting in INAR(p) models: a likelihood-based Markov chain approach. Int. J. Forecast. 24, 151–162.

Davies, R.B., 1973. Numerical inversion of a characteristic function. Biometrika 60, 415–417.

Joe, H., 1996. Time series models with univariate margins in the convolution-closed infinitely divisible class. J. Appl. Probab. 33, 664–677.

Joe, H., 1997. Multivariate Models and Dependence Concepts. Chapman & Hall, London.

Kendall, D.G., 1948. On generalised birth and death processes. Ann. Math. Statist. 19, 1–15.

Latour, A., 1998. Existence and stochastic structure of a non-negative integer-valued autoregressive process. J. Time Ser. Anal. 19, 439–455.

McKenzie, E., 1986. Autoregressive moving-average processes with negative-binomial and geometric marginal distributions. Adv. in Appl. Probab. 18, 679–705.

McKenzie, E., 1988. Some ARMA models for dependent sequences of Poisson counts. Adv. in Appl. Probab. 20, 822–835.

McKenzie, E., 2003. Discrete variate time series. In: Shanbhag, D.N. (Ed.), Stochastic Processes: Modelling and Simulation. Handbook of Statistics, vol. 21. North-Holland, Amsterdam, pp. 573–606.

Phatarfod, R.M., Mardia, K.V., 1973. Some results for dams with Markovian inputs. J. Appl. Probab. 10, 166–180.

Rémillard, B., Theodorescu, R., 2001. Estimation based on the empirical characteristic function. In: Balakrishnan, N., et al. (Eds.), Asymptotic Methods in Probability and Statistics With Applications. Papers from the International Conference, St. Petersburg, Russia, June 24–28, 1998. Birkhauser, Boston, MA, pp. 435–449.

Steutel, F.W., van Harn, K., 1979. Discrete analogues of self-decomposability and stability. Ann. Probab. 7, 893–899.

Steutel, F.W., Vervaat, W., Wolfe, S.J., 1983. Integer-valued branching processes with immigration. Adv. in Appl. Probab. 15, 713–725.

Steutel, F.W., van Harn, K., 2004. Infinite Divisibility of Probability Distributions on the Real Line. Marcel Dekker, New York, Basel.

Van Harn, K., Steutel, F.W., Vervaat, W., 1982. Self-decomposable discrete distributions and branching processes. Z. Wahrschienlichkeitsth. 61, 97–118.

Van Harn, K., Steutel, F.W., 1985. Integer-valued self-similar processes. Comm. Statist. Stoch. Models 1, 191–208.

Van Harn, K., Steutel, F.W., 1993. Stability equations for processes with stationary independent increments using branching processes and Poisson mixtures. Stochastic Process. Appl. 45, 209–230.

Weiß, C.H., 2008a. Serial dependence and regression of Poisson INARMA models. J. Statist. Plann. Inference 138, 2975–2990.

Weiß, C.H., 2008b. Thinning operations for modeling time series counts—a survey. Adv. Statist. Anal. 92, 319–341.

Zhu, R., 2002. On continuous-time generalized AR(1) processes: models, statistical inference, and applications to non-normal time series. Ph.D. Thesis, Department of Statistics, University of British Columbia 〈http://www.collectionscanada.gc.ca/obj/s4/f2/dsk4/etd/NQ73263.PDF〉.

Zhu, R., Joe, H., 2003. A new type of discrete self-decomposability and its application to continuous-time Markov processes for modelling count data time series. Stoch. Models 19, 235–254.

Zhu, R., Joe, H., 2006. Modelling count data time series with Markov processes based on binomial thinning. J. Time Ser. Anal. 27, 725–738.