HAL
open science

The Predictive Distributions of Thinning-based Count Processes

Yang Lu

To cite this version:

Yang Lu. The Predictive Distributions of Thinning-based Count Processes. Scandinavian Journal of Statistics, In press, 10.1111/sjos.12438 . hal-02419018

HAL Id: hal-02419018

https://hal.science/hal-02419018

Submitted on 19 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The Predictive Distributions of Thinning-based Count Processes

Yang Lu*

December 1, 2019

Abstract: This paper shows that the term structure of conditional, or predictive distributions allows for closed form expression in a large family of (possibly higher-order, or infinite order) thinning-based count processes such as INAR(p), INARCH(p), NBAR(p), and INGARCH(1,1). Such predictive distributions are currently often deemed intractable by the literature and existing approximation methods are usually time consuming and induce approximation errors. In this paper, we propose a Taylor's expansion algorithm for these predictive distributions, which is both *exact* and fast. Through extensive simulation exercises, we demonstrate its advantages with respect to existing methods in terms of the computational gain and/or precision.

Keywords: compound autoregressive process, probabilistic forecast of counts, matrix arithmetic.

Acknowledgement: I thank the two referees and the Associate Editor for their insightful comments and suggestions. I thank the ACPR chair in Regulation and Systemic Risk, the CNRS (délégation grant), the Society of Actuaries (2018 individual grant) as well as the ANR (Labex MME-DII grant) for financial support.

*University of Paris 13.

1 Introduction

Nonnegative integer-valued count process models are widely used in domains such as marketing [Böckenholt (1998)], economics [Blundell et al. (1999); Brännäs and Hellström (2001); Harris and McCabe (2018)], finance [Heinen and Rengifo (2007); Bien et al. (2011); Kirchner (2017)] and insurance [Gouriéroux and Jasiak (2004)]. The benchmark model, introduced by McKenzie (1985), Al-Osh and Alzaid (1987) in the first-order case (called INAR(1)) postulates that:

$$X_t = \alpha \circ X_{t-1} + \epsilon_t, \qquad \forall t, \tag{1}$$

where the thinning operator is defined as follows: conditionally on X_{t-1}, variable $\alpha \circ X_{t-1}$ has the distribution $\mathcal{B}(X_{t-1}, \alpha)$, that is the binomial distribution with probability parameter α and size parameter X_{t-1}. Moreover, this variable is conditionally independent from the i.i.d. sequence of innovation (ϵ_t), which follows Poisson distribution.

Since this seminal work, extensions have flourished. Du and Li (1991) introduce higher-order INAR(p) models satisfying:

$$X_t = \sum_{i=1}^{p} \alpha_i \circ X_{t-i} + \epsilon_t, \qquad \forall t, \tag{2}$$

where given all the past observations $\underline{X_{t-1}} = \{X_{t-1}, X_{t-2}, ...\}$, variables $\alpha_i \circ X_{t-i}$ are mutually independent with Binomial distribution $\mathcal{B}(X_{t-i}, \alpha_i)$, and are independent from ϵ_t, which is an i.i.d. sequence, and the distribution of the latter is not necessarily Poisson. In particular, if (ϵ_t) is Poisson (resp. negative binomial) distributed, then we say that (X_t) follows the Poisson-INAR(p) [resp. NB-INAR(p)] process. Note that Alzaid and Al-Osh (1990) has also proposed an alternative INAR(p) specification. This latter is less tractable and even the conditional expectation formula is highly complicated [see Jung and Tremayne (2006) for its expression in the special case where $p = 2$]. Thus this model is not considered in our paper. This choice is consistent with the majority of recent papers [see e.g. Pedeli et al. (2015) and references therein].

Meanwhile, it has also been proposed to relax the binomial distributional assumption of the thinning operator [see e.g. Latour (1998)]. Roughly speaking, these models trace their origin to the applied probability literature, in which they are called Galton-Watson process with immigration, or branching process, see e.g. Venkataraman and Nanthi (1982). They are special

cases of model:

$$X_t = \sum_{i=1}^{p} Z_i * X_{t-i} + \epsilon_t, \tag{3}$$

where integer p could be finite or infinite; the operator $*$ is defined as follows. Given $\underline{X_{t-1}}$, each $Z_i * X_{t-i}$ is the sum of X_{t-i} i.i.d. copies of count variable Z_i. In other words, equation (3) does not include random coefficient count processes in which the different copies of Z_i are dependent [see e.g. Zheng et al. (2006)]. This distribution of Z_i usually depends on some unknown parameters and is not necessarily Bernoulli, which is the case in model (2). Moreover variables $Z_i * X_{t-i}$ are mutually independent, and are independent from the i.i.d. innovation sequence ϵ_t. Prominent examples of special cases of model (3) include, among others:

- Ristić et al. (2009) and Gouriéroux and Lu (2019) consider first order models ($p = 1$), and assume that Z_1 is negative binomial (NB) distributed. Then they make special assumptions on the distribution of ϵ_t to ensure that process (X_t) is stationary with NB stationary distribution.

- Rydberg and Shephard (2000); Weiß (2010) consider the INARCH(1) model in which the conditional distribution of X_t given X_{t-1} is Poisson $\mathcal{P}(\alpha X_{t-1} + \beta)$, with parameters $\alpha, \beta > 0$. This corresponds to model (3) with $p = 1$, and both Z_1 and ϵ_t are Poisson distributed with parameters α and β, respectively.

- Zhu and Joe (2010) also focuses on first-order models but assume that Z_1 and ϵ_t follow other non-NB and non-Binomial distributions.

- Finally, a few models can be written in the form (3) with $p = \infty$. For instance, Kirchner (2016) introduces a Poisson-INARCH(∞) model (although the author terms this model INAR(∞), in his framework all the variables Z_i, as well as ϵ_t, follow Poisson distributions. Thus his model can also be called Poisson-INARCH(∞)), whereas the Poisson-INGARCH(1,1) model of Ferland et al. (2006) can also be viewed as a special case of Poisson-INARCH(∞) model (see section 3.3 for a more detailed analysis).

In this paper we call models of type (3) thinning-based [see Weiß (2008); Ristić et al. (2009); Zhu and Joe (2010); Scotto et al. (2015); Kirchner (2016)]. Note that some papers in the literature also use the term "generalized thinning". We choose to avoid this latter terminology

since we want to emphasize the our framework does not apply to models with random coefficient thinning. Family (3) is significantly larger than (2), but up to now the literature has only been able to handle first-order models, with quite restrictive parametric assumptions on Z_i's and ϵ_t. One explanation is that, up to now extensions (2) and (3) have the serious downside that the term structure of predictive probability mass functions (p.m.f.'s) is believed to be intractable, rendering estimation and/or forecasting cumbersome. This difficulty is especially acute when the order p, and/or the horizon of forecasting h increases. To our knowledge, the term structure of p.m.f.'s alows for closed form expression only in two special first-order models. These are the initial Poisson-INAR(1) model (1), as well as the negative binomial autoregressive process (NBAR(1)) introduced by Gouriéroux and Lu (2019). Although the literature has proposed some approximation methods [see e.g. Pedeli et al. (2015) for likelihood-based estimation and Jung and Tremayne (2006); McCabe et al. (2011); Bisaglia and Canale (2016) for forecasting], they either induce significant approximation errors, or are still computationally too costly.

This paper solves the above open difficulty for model (3). We show that the term structure of conditional p.m.f.'s in model (3) has closed form expression at *any* horizon, and hence significantly improve the existing estimation and forecasting procedures. The method is general in the sense that it does not require the distributions of Z_i's and ϵ_t to belong to the classical families such as Bernoulli, Poisson or negative binomial. Rather, it is applicable so long as some mild assumption (see Assumption 1) on the log p.g.f. of these variables is satisfied. Our method explores the Compound Autoregressive (CaR) property of process (3), and the natural link between the probability generating function (p.g.f.) and the corresponding p.m.f. of a count distribution. This approach has the distinctive advantage that the algorithm only involves low-dimensional matrix operations, whose computational cost is insensitive to p and remains very low even for large p.

The paper is organized as follows. Section 2 computes the Taylor's expansion of the one-step-ahead conditional p.g.f. and deduces the corresponding p.m.f., which facilitates, among others, the likelihood-based estimation. Section 3 derives the multiple-step-ahead predictive p.m.f.'s for finite-order models, before extending the result to some infinite order models such as Poisson-INGARCH(1,1). Section 4 compares our approach and existing methods, first in terms of maximum likelihood estimation, then in terms of higher-horizon forecasting. In particular,

for the estmation we show that depending on model parameters, our model is generically either comparable, or faster than the other, exact method of Bu et al. (2008). For the forecasting, we show that our method is more precise and much less time consuming than competing methods proposed by McCabe et al. (2011) and Jung and Tremayne (2006). Section 5 concludes. Technical details and proofs are gathered in Appendices.

2 Likelihood-based estimation

Conducting maximum likelihood (ML) estimation requires the evaluation of the conditional p.m.f. $p(X_t|\underline{X_{t-1}})$. In model (3), by standard convolution we get [see e.g. Drost et al. (2009)]:

$$p(X_t|\underline{X_{t-1}}) = \sum_{n_1+n_2+\cdots+n_p+n_{p+1}=X_t} \mathbb{P}[\epsilon_t = n_{p+1}] \prod_{i=1}^{p} \mathbb{P}[Z_i * X_{t-i} = n_i|X_{t-i}]. \tag{4}$$

For instance, in INAR(p) models, Z_i's are Bernoulli distributed and we have $\mathbb{P}[Z_i * X_{t-i} = n_i|X_{t-i}] = \binom{X_{t-i}}{n_i}\alpha_i^{n_i}(1-\alpha_i)^{X_{t-i}-n_i}$. Except in a few special cases such as INARCH(p) where each Z_i and ϵ_t follows Poisson distribution, the above $p-$dimensional summation cannot be further simplified, and becomes increasingly cumbersome for large p. Bu et al. (2008) point out that, this computation can be partially simplified by using the fact that the convolution of $p+1$ variables is the convolution between the first variable and that of the p remaining ones, which can be further decomposed recursively. However for large p, the recursive method remains computationally heavy [see also Pedeli et al. (2015) for discussions]. While there exist convenient and consistent moment based estimators [see e.g. Al-Osh and Alzaid (1987)], they typically suffer from significant efficiency loss [see e.g. Bu et al. (2008)]. Recently, Pedeli et al. (2015) propose a saddle-point approximation of $p(X_t|\underline{X_{t-1}})$ for ML estimation. Although this method is faster than the brute force convolution, it still has some drawbacks. First, the approximated p.m.f. $\tilde{p}(\cdot|\underline{X_{t-1}})$ does not sum up to one across $\mathbb{N} = \{0, 1, 2, ...\}$. Second, the method only applies to INAR(p) models. Third, this approximation itself lacks closed form expression and its computation requires numerical inversion of some nonlinear functions. This leads to extra computational complexity, as well as approximation error. For instance, Pedeli et al. (2015) report that for certain parameter values, the relative error of the likelihood function can be as

large as 2 to 5 percent. As a consequence, this approximate ML estimator is inconsistent and simulation results show that in finite sample, the bias can be significantly larger than that of the exact ML estimator (see their Table 1).

2.1 The Taylor's expansion based algorithm

Our solution is based on the property that the one-step-ahead conditional p.g.f. of model (3) satisfies:

$$\mathbb{E}[u^{X_t}|\underline{X_{t-1}}] = \exp\Big[\sum_{i=1}^{p} X_{t-i}\log(\mathbb{E}[u^{Z_i}]) + \log(\mathbb{E}[u^{\epsilon_t}])\Big], \tag{5}$$

which is exponential affine in $\underline{X_{t-1}}$. Such processes (X_t) are called ($p-$th order) Compound Autoregressive (CaR(p)) processes [see Darolles et al. (2006)] and by Proposition 7 of Darolles et al. (2006), we can first deduce that such a process is strictly stationary if and only if: $\sum_{i=1}^{p}\mathbb{E}[Z_i] < 1$. Then we remark that this conditional p.g.f. is linked to the conditional p.m.f. through:

$$\mathbb{E}[u^{X_t}|\underline{X_{t-1}}] = \sum_{k=0}^{\infty}\mathbb{P}[X_t = k|\underline{X_{t-1}}]u^k. \tag{6}$$

Thus to get $\mathbb{P}[X_t = k|\underline{X_{t-1}}]$ for all $k = 0, 1, ..., n$, where n is a fixed integer, it suffices to compute the $n-$th order Taylor's expansion at $u = 0$ for the corresponding conditional p.g.f. To this end we make the following assumption:

Assumption 1. In model (3), the log p.g.f. of variables $Z_i, i = 1, ..., p$, and ϵ_t all have tractable Taylor's expansions:

$$g_i(u) := \log\mathbb{E}[u^{Z_i}] = \sum_{k=0}^{\infty} c_{i,k}u^k, \qquad \forall i = 1, ..., p, \qquad \forall u, \tag{7}$$

$$g_0(u) := \log\mathbb{E}[u^{\epsilon_t}] = \sum_{k=0}^{\infty} c_{0,k}u^k, \qquad \forall u. \tag{8}$$

By "tractable", we require the coefficients $c_{i,k}$ and $c_{0,k}$, where $k = 0, ...,$ and $i = 1, ..., p$, to be analytic functions of the parameters characterizing the distributions of Z_i's and of (ϵ_t). This assumption is quite mild and is satisfied by many standard count distributions. For instance, distributions in Examples 1-4 below have tractable p.g.f. and are infinitely differentiable at $u = 0$, such that *Assumption* 1 automatically holds.

Example 1 (Bernoulli and Poisson distributions)**.** If X follows Bernoulli distribution $\mathcal{B}(1,\alpha)$, then:

$$\begin{aligned}\log \mathbb{E}[u^X] &= \log(\alpha u + 1 - \alpha) = \log(1-\alpha) + \log(1 + \frac{\alpha}{1-\alpha}u) \\ &= \log(1-\alpha) + \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k}\Big(\frac{\alpha}{1-\alpha}\Big)^k u^k.\end{aligned}$$

On the other hand, if X follows Poisson distribution $\mathcal{P}(\lambda)$, then:

$$\log \mathbb{E}[u^X] = -\lambda + \lambda u.$$

Thus, in particular, *Assumption* 1 is satisfied by the Poisson-INAR(p) model, that is equation (2) with Poisson distributed innovation (ϵ_t). In this case, the coefficients in the Taylor's expansions (7) and (8) are given by:

$$\begin{aligned}&c_{0,0} = -\lambda, \qquad c_{0,1} = \lambda, \qquad \text{and } c_{0,j} = 0, \qquad \forall j \geq 2. \\ &c_{i,0} = \log(1-\alpha_i), \qquad \text{and } c_{i,k} = (-1)^k\Big(\frac{\alpha_i}{1-\alpha_i}\Big)^k, \qquad \forall k \geq 1, i = 1, ..., p.\end{aligned}$$

Example 2 (NB distribution (including geometric distribution), see Ristić et al. (2009); Gouriéroux and Lu (2019))**.** If X follows NB distribution with probability parameter $q \in (0,1)$ and size parameter r [in the following we will use the notation $X \sim \mathcal{NB}(r,q)$], then:

$$\mathbb{E}[u^X] = \frac{(1-q)^r}{(1-qu)^r}, \qquad \forall u \in [0, \frac{1}{q}), \tag{9}$$

$$\text{hence} \qquad \log\Big(\mathbb{E}[u^X]\Big) = r\log(1-q) - r\log(1-qu) = r\log(1-q) + r\sum_{j=1}^{\infty}\frac{q^j}{j}u^j.$$

Example 3 (Compound Poisson distributions, see Schweer and Weiß (2014))**.** We say that variable X follows compound Poisson distribution if it has the representation:

$$X =_{(d)} \sum_{j=1}^{N} Z_j,$$

where symbol $=_{(d)}$ denotes equality in distribution; variable N is Poisson $\mathcal{P}(\lambda)$ distributed, and

is independent from the sequence $(Z_j)_j$, where $(Z_j)_j$ is an i.i.d. sequence of count variables. It is easily checked that the log p.g.f. of X is:

$$\log\Big(\mathbb{E}[u^X]\Big) = \lambda\Big(\mathbb{E}[u^{Z_1}] - 1\Big) = \lambda\Big(\sum_{k=0}^{\infty}\mathbb{P}[Z_1 = k]u^k - 1\Big).$$

Thus $\log\Big(\mathbb{E}[u^X]\Big)$ has tractable Taylor's expansion, so long as the p.m.f. of Z_1 is tractable.

Example 4 (Distributions defined through the p.g.f.)**.** Zhu and Joe (2010) propose several first-order, thinning-based count processes in which Z_1 has simple p.g.f. but complicated p.m.f. For instance, the non central NB distribution has the p.g.f.:

$$\mathbb{E}[u^X] = \frac{(1-\alpha) + (\alpha-\gamma)u}{1-\alpha\gamma-(1-\alpha)\gamma u}, \qquad \forall u \in [0, \frac{1-\alpha\gamma}{(1-\alpha)\gamma}),$$

where parameters α, γ lie between 0 and 1. For this distribution, the log p.g.f. is easily expanded:

$$\begin{aligned}\log\Big(\mathbb{E}[u^X]\Big) &= \log\frac{1-\alpha}{1-\alpha\gamma} + \log(1+\frac{\alpha-\gamma}{1-\alpha}u) - \log[1-\frac{(1-\alpha)\gamma}{1-\alpha\gamma}u] \\ &= \log\frac{1-\alpha}{1-\alpha\gamma} + \sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k}\big(\frac{\alpha-\gamma}{1-\alpha}\big)^k u^k + \sum_{j=1}^{\infty}\big[\frac{(1-\alpha)\gamma}{1-\alpha\gamma}\big]^k u^k.\end{aligned}$$

Another example is the Hermite distribution, which is used by Fernández-Fontelo et al. (2017) in INAR models as the distribution of the innovation ϵ_t. Such a distribution has a log. p.g.f.:

$$\log\mathbb{E}[u^{\epsilon_t}] = a_1(u-1) + a_2(u^2-1),$$

where parameters $a_1, a_2 \geq 0$.

Finally, the majority of other distributions considered in the count process literature who do not have simple p.g.f. (or log p.g.f.) usually have simple p.m.f. For instance, Heinen and Rengifo (2007) introduce the double Poisson distribution, Jazi et al. (2012) consider zero-inflated mixture distributions. In Appendix 3 we show how the log p.g.f. can be easily expanded at any given order, so long as the p.m.f. is tractable.

Under *Assumption* 1, we can rewrite equation (5) into:

$$\mathbb{E}[u^{X_t}|\underline{X_{t-1}}] \tag{10}$$

$$= \exp\Big[c_{0,0} + \sum_{i=1}^{p} X_{t-i}c_{i,0}\Big] \exp\Big[\sum_{j=1}^{n}(c_{0,j} + \sum_{i=1}^{p} X_{t-i}c_{i,j})u^j + O(u^{n+1})\Big] \tag{11}$$

$$= \exp(A_0)\exp\Big[\sum_{j=1}^{n} A_j u^j\Big] + O(u^{n+1}) \tag{12}$$

$$= \exp(A_0)\sum_{k=0}^{n}\frac{1}{k!}\Big[\sum_{j=1}^{n} A_j u^j\Big]^k + O(u^{n+1}), \tag{13}$$

where coefficients A_j are given by:

$$A_j = c_{0,j} + \sum_{i=1}^{p} X_{t-i}c_{i,j}, \qquad \forall j = 0, ..., n. \tag{14}$$

From equation (10) to (11), we have expanded the log p.g.f. of Z_i's and ϵ_t; from equation (12) to (13) we have expanded the exponential function.

Thus to deduce the coefficient in front of the term u^n in expansion (13), it suffices to compute recursively the $n+1$ first terms of the polynomial $\Big[\sum_{j=1}^{n} A_j u^i\Big]^k$ for each k. While this can be achieved using a symbolic calculation package such as Mathematica (see Gordy (2002) for a similar application to credit risk), we suggest below a matrix-based algorithm that is simpler and more suitable for statistical packages.

Proposition 1. *For any coefficients $\gamma_i, i = 0, ..., n$, the $(n+1)$ first coefficients of polynomial $\Big(\sum_{j=0}^{n}\gamma_j u^j\Big)^k$, where $k = 0, ...n$, are given by the column vector:*

$$\begin{bmatrix} \gamma_0 & 0 & 0 & \cdots & 0 \\ \gamma_1 & \gamma_0 & 0 & \cdots & 0 \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \gamma_n & \gamma_{n-1} & \cdots & \gamma_1 & \gamma_0 \end{bmatrix}^k \begin{bmatrix} 1 \\ 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix}, \qquad \forall k = 0, ..., n. \tag{15}$$

The proof is obvious and omitted. In equation (13) we have $\gamma_0 = 0$, thus the square matrix

above, called the Toeplitz matrix, becomes triangular inferior. Thus we get:

$$\begin{bmatrix} \mathbb{P}[X_t = 0|\underline{X_{t-1}}] \\ \mathbb{P}[X_t = 1|\underline{X_{t-1}}] \\ \cdots \\ \mathbb{P}[X_t = n-1|\underline{X_{t-1}}] \\ \mathbb{P}[X_t = n|\underline{X_{t-1}}] \end{bmatrix} = \exp(A_0) \sum_{k=0}^{n} \frac{1}{k!} \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ A_1 & 0 & 0 & \cdots & 0 \\ A_2 & A_1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ A_n & A_{n-1} & \cdots & A_1 & 0 \end{bmatrix}^k \begin{bmatrix} 1 \\ 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix}. \tag{16}$$

Note that although the above algorithm involves a Taylor's expansion (of the conditional p.g.f.), it provides the *exact* value of the conditional p.m.f, because of the identity (6) (see also Example 5 for further discussions). In other words, if we were to compute, say, $\mathbb{P}[X_t = 5|\underline{X_{t-1}}]$, then an expansion of order 5 or 20 would lead to the same result.

On the contrary to formula (4), which becomes increasingly cumbersome when p increases, the computational cost of (15) is independent of p. Indeed, it only involves square matrices and vectors of dimension $(n+1)$, where n does not depend on p. The order p only impacts the computation of the scalars $A_0, ..., A_n$, whose cost is negligible. Thus this method is extremely convenient, especially for higher-order models.

One referee rightly points out that the computational cost of our algorithm is still (inevitably) increasing in n. Thus, if the values of the counts we deal with are very large, then this method can also become cumbersome. This downside is nevertheless partially mitigated by the fact that the successive powers of the Toeplitz matrix are still Toeplitz and thus can be computed very efficiently. More precisely, we can check that:

$$\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ \gamma_1 & 0 & 0 & \cdots & 0 \\ \gamma_2 & \gamma_1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \gamma_n & \gamma_{n-1} & \cdots & \gamma_1 & 0 \end{bmatrix}^k = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ \gamma_{1,k} & 0 & 0 & \cdots & 0 \\ \gamma_{2,k} & \gamma_{1,k} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \gamma_{n,k} & \gamma_{n-1,k} & \cdots & \gamma_1 & 0 \end{bmatrix} \tag{17}$$

for any integer k, where the sequence $(\gamma_{i,k})$ satisfies the recursion:

$$\begin{aligned}
\gamma_{1,k+1} &= 0 \\
\gamma_{2,k+1} &= \gamma_{1,k}\gamma_1 \\
\cdots &= \cdots \\
\gamma_{n,k+1} &= \gamma_{n-1,k}\gamma_1 + \gamma_{n-2,k}\gamma_2 + \cdots + \gamma_{1,k}\gamma_{n-1}.
\end{aligned}$$

In other words, the multiplication between Toeplitz matrices involves at most the computation of n coefficients for a total of $O(\frac{1}{2}n^2)$ operations, where the symbol $O(\cdot)$ should be understood as being equivalent to $\frac{1}{2}n^2$ when n goes to infinity. . As a comparison, ordinary matrix multiplications usually requires a cost of $O(n^3)$ operations. As a consequence, in total, computing the conditional p.m.f. of X_t given $\underline{X_{t-1}}$ involves roughly $O(\frac{1}{2}X_t^3)$ operations since we have to raise the Toeplitz matrix up to power X_t.

To summarize, the Taylor's expansion approach makes the ML estimation feasible for higher-order models without inducing bias. Alternatively, it also facilitates Bayesian inference [see McCabe and Martin (2005); Bisaglia and Canale (2016)] of the same models, the ease of which usually depending (among others) on the tractability of the likelihood function.

2.2 Examples

As an illustration, let us derive the conditional p.m.f. for some popular specifications of the distributions of Z_i and ϵ_t.

Example 5 (INARCH(p) model)**.** In this example we assume that, in equation (3), variables $Z_1, ..., Z_p$ and ϵ_t follow Poisson distributions with parameters $\lambda_1, ..., \lambda_p$ and λ_0, respectively. This can be viewed as the higher-order extension of the (Poisson-)INARCH(1) model studied by Rydberg and Shephard (2000); Weiß (2010). For this model, the one-step-ahead conditional distribution is Poisson with parameter:

$$\lambda_t := \lambda_0 + \sum_{i=1}^{p} X_{t-i}\lambda_i.$$

While it is possible to directly deduce the conditional p.m.f.:

$$\mathbb{P}[X_t = k|\underline{X_{t-1}}] = e^{-\lambda_t}\frac{\lambda_t^k}{k!}, \tag{18}$$

to support our claim that equation (16) provides an exact expression [see the comment below (16)], let us now check that we can recover equation (18) from (16). We first remark that for this model, $A_0, A_1, ..., A_n$ defined in (14) are given by: $A_0 = -\lambda_0 - \sum_{i=1}^p X_{t-i}\lambda_i = -\lambda_t$, $A_1 = \lambda_t$, and $A_2 = A_3 = \cdots = 0$. Then by some simple algebra, we can rewrite the RHS of equation (16) into:

$$\exp(-\lambda_t)\sum_{k=0}^n \frac{\lambda_t^k}{k} M \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}^k \begin{bmatrix} 1 \\ 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix} = \exp(-\lambda_t) \begin{bmatrix} 1 \\ \lambda_t \\ \frac{\lambda_t^2}{2!} \\ \cdots \\ \frac{\lambda_t^n}{n!} \end{bmatrix}.$$

Thus equation (16) provides indeed the exact expression of the conditional p.m.f.

Example 6 (Poisson-INAR(p) model)**.** Let us now consider the benchmark Poisson-INAR(p) model (2), under the assumption that ϵ_t is Poisson $\mathcal{P}(\lambda)$ distributed. The conditional p.g.f. is:

$$\begin{aligned} \mathbb{E}[u^{X_t}|\underline{X_{t-1}}] &= \exp\Big[\sum_{i=1}^p X_{t-i}\log(\alpha_i u + 1 - \alpha_i) + \lambda(u-1)\Big] \\ &= \exp(A_0)\sum_{k=0}^n \frac{1}{k!}\Big[\sum_{j=1}^n A_j u^j\Big]^k + O(u^{n+1}), \\ \text{with } A_0 &= -\lambda + \sum_{i=1}^p X_{t-i}\log(1-\alpha_i), \qquad A_1 = \lambda + \sum_{i=1}^p X_{t-i}\frac{\alpha_i}{1-\alpha_i}, \\ \text{and } A_j &= \frac{(-1)^{j-1}}{j}\sum_{i=1}^p X_{t-i}(\frac{\alpha_i}{1-\alpha_i})^j, \qquad \forall j = 2, ..., n. \end{aligned}$$

Example 7 (NB-INAR(p) model)**.** Let us now give the corresponding formulas, when the innovation term in the INAR(p) model (2) follows the $\mathcal{NB}(r, q)$ distribution instead of the Poisson

distribution. By (5), the conditional p.g.f. is:

$$\mathbb{E}[u^{X_t}|\underline{X_{t-1}}] = \exp\Big[\sum_{i=1}^{p} X_{t-i}\log(\alpha_i u + 1 - \alpha_i) + r\log(1-q) - r\log(1-qu)\Big]$$
$$= \exp\Big[r\log(1-q) + \sum_{i=1}^{p} X_{t-i}\log(1-\alpha_i)\Big]\sum_{k=0}^{n}\frac{1}{k!}\Big[\sum_{j=1}^{n} A_j u^j\Big]^k + O(u^{n+1}),$$

where $A_0 = r\log(1-q) + \sum_{i=1}^{p} X_{t-i}\log(1-\alpha_i)$,

$$A_j = \frac{rq^j}{j} + \frac{(-1)^{j-1}}{j}\sum_{i=1}^{p} X_{t-i}(\frac{\alpha_i}{1-\alpha_i})^j, \qquad \forall j = 1, ..., n.$$

Example 8 (NBAR(p) model)**.** Let us finally assume that in equation (3), variables ϵ_t and $Z_i, i = 1, ..., p$ follow $\mathcal{NB}(r_0, q_0)$ and $\mathcal{NB}(r_i, q_i)$ distributions, respectively. This model can be viewed as a higher-order extension of the Negative Binomial Autoregressive [NBAR(1)] process considered in Ristić et al. (2009); Gouriéroux and Lu (2019) and thus will be called NBAR(p). Its conditional p.g.f. is:

$$\mathbb{E}[u^{X_t}|\underline{X_{t-1}}] = \exp\Big[\sum_{i=1}^{p} X_{t-i}r_i[\log(1-q_i) - \log(1-q_i u)] + r_0[\log(1-q_0) - \log(1-q_0 u)]\Big]$$
$$= \exp(A_0)\sum_{k=0}^{n}\frac{1}{k!}\Big[\sum_{j=1}^{n} A_j u^j\Big]^k + O(u^{n+1}),$$

with $A_0 = r_0\log(1-q_0) + \sum_{i=1}^{p} X_{t-i}r_i\log(1-q_i)$,

$$A_j = \frac{r_0 q_0^j}{j} + \sum_{i=1}^{p} X_{t-i}\frac{r_i q_i^j}{j}, \qquad \forall j = 1, ..., n.$$

Finally, in all the above four examples, once the coefficients A_j are computed, it suffices to insert them into (16) to get the vector of conditional p.m.f.

3 Multi-step forecasting

One of the major challenges of count process models is that even though the predictive means $\mathbb{E}[X_{t+h}|\underline{X_{t-1}}]$ are often straightforward to compute, they are not informative enough to characterize the predictive count distributions. Moreover, McCabe et al. (2011) argue that since the

predictive means are non-integer, they are incompatible with the discrete sample space. A proper, probabilistic forecast thus necessitates the multiple-step-ahead conditional p.m.f. of X_{t+h} given $\underline{X_{t-1}}$:

$$p_h(X_{t+h}|\underline{X_{t-1}}) = \sum_{X_{t+h-1}=0}^{\infty} \sum_{X_{t+h-2}=0}^{\infty} \cdots \sum_{X_t=0}^{\infty} p(X_{t+h}|\underline{X_{t+h-1}})p(X_{t+h-1}|\underline{X_{t+h-2}}) \cdots p(X_t|\underline{X_{t-1}}). \tag{19}$$

This is a $h-$dimensional, *infinite* summation, and thus is extremely cumbersome for any $h \geq 1$, even if the previous section has provided a simpler method of computing the individual terms of the form $p(X_t|\underline{X_{t-1}})$.

So far there are two major solutions in the literature. Jung and Tremayne (2006) propose to approximate $p_h(\cdot|\underline{X_{t-1}})$ via simulations of future trajectories, but this latter is also highly computationally intensive. Alternatively, Bu and McCabe (2008); McCabe et al. (2011) approximate the probability of the process taking values larger than some integer threshold τ by 0. Then the p-dimensional vector $Y_t = (X_t, X_{t-1}, ..., X_{t-p+1})$ is approximated by a first-order, $(\tau+1)^p$ states Markov chain, and the approximate $p_h(\cdot|\underline{X_{t-1}})$ satisfies a recursive formula (in h) involving the $(\tau+1)^p \times (\tau+1)^p$ transition matrix Π of the chain (Y_t). Its drawbacks is that, when τ and/or p are not small enough, the dimension of Π is too high for the computation to be feasible. This necessity to consider, for some applications, a larger threshold has been highlighted by Zhu and Joe (2009) and references therein.

3.1 The algorithm

Let us now propose a simpler algorithm using the same, Taylor's expansion approach. As in Section 2, we first derive the conditional p.g.f. $\mathbb{E}[u^{X_{t+h}}|\underline{X_{t-1}}]$. One nice property of the CaR count processes is that such higher-order p.g.f. is still exponential affine. In other words, the following proposition is a higher-order extension of equation (4).

Proposition 2. *For any integer $h \geq 0$, we have:*

$$\mathbb{E}[u^{X_{t+h-1}}|\underline{X_{t-1}}] = \exp\Big[B_{h,0}(u) + \sum_{i=1}^{p} B_{h,i}(u)X_{t-i}\Big], \tag{20}$$

where functional (in u) coefficients $B_{h,i}(u), i = 0, ..., p$, satisfy the recursive formula (in h):

$$B_{1,0}(u) = \log \mathbb{E}[u^{\epsilon_t}], \tag{21}$$

$$B_{1,i}(u) = \log \mathbb{E}[u^{Z_i}], \qquad \forall i = 1, ..., p, \tag{22}$$

$$B_{h+1,0}(u) = B(h, 0)(u) + g_0(e^{B_{h,1}(u)}), \tag{23}$$

$$B_{h+1,i}(u) = \mathbb{1}_{i<p} B_{h,i+1}(u) + g_i(e^{B_{h,1}(u)}), \qquad \forall i = 1, ..., p. \tag{24}$$

Proof. See Appendix 1. □

Although $B_{h,i}(u)$ is typically intractable for large h, their Taylor's expansions up to a finite order (at $u = 0$), and hence that of $\mathbb{E}[u^{X_{t+h}} | \underline{X_{t-1}}]$, can be easily obtained by recursion, and we will see later on that this is sufficient for probabilistic forecasting purpose. Indeed, given a large integer threshold τ and suppose that the $\tau-$th order Taylor's expansions of $B_{h,i}(u)$ have already been obtained for each $i = 0, ..., p$, then those of $B_{h+1,i}(u)$ are available through:

Proposition 3. *We have:*

$$B_{h+1,0}(u) = B_{h,0}(u) + c_{0,0} + \sum_{k=0}^{\tau} \frac{B_{h,1}^k(u)}{k!} \sum_{j=1}^{\tau} c_{0,j} j^k + O(u^{\tau+1}),$$

$$B_{h+1,i}(u) = \mathbb{1}_{i<p} B_{h,i+1}(u) + c_{i,0} + \sum_{k=0}^{\tau} \frac{B_{h,1}^k(u)}{k!} \sum_{j=1}^{\tau} c_{i,j} j^k + O(u^{\tau+1}), \qquad i = 1, ..., p.$$

Proof. See Appendix 2. □

Thus it suffices to apply Proposition 1 onceto get the $\tau-$th order Taylor's expansions of the successive powers of $B_{h,1}(u)$, to derive that of $B_{h+1,i}(u)$, $i = 0, ..., p$. More precisely, if we write the Taylor's expansion of $B(h, i)(u)$ as:

$$B_{h,i}(u) = \sum_{k=0}^{\tau} b_{h,i,k} u^k + O(u^{\tau+1}),$$

then for each $i = 0, ..., p$, the vector of coefficients $(b_{h,i,0}, b_{h,i,1}, \cdots, b_{h,i,\tau})'$ satisfies the recursion:

$$\begin{bmatrix} b_{h+1,i,0} \\ b_{h+1,i,1} \\ \cdots \\ b_{h+1,i,\tau} \end{bmatrix} = \mathbb{1}_{i=0} \begin{bmatrix} b_{h,0,0} \\ b_{h,0,1} \\ \cdots \\ b_{h,0,\tau} \end{bmatrix} + \mathbb{1}_{0<i<p} \begin{bmatrix} b_{h,i+1,0} \\ b_{h,i+1,1} \\ \cdots \\ b_{h,i+1,\tau} \end{bmatrix} + \begin{bmatrix} c_{i,0} \\ 0 \\ \cdots \\ 0 \end{bmatrix} + \sum_{k=0}^{\tau} \frac{1}{k!} \Big[\sum_{j=1}^{\tau} c_{i,j} j^k \Big] \begin{bmatrix} b_{h,i,0} & 0 & \cdots & 0 \\ b_{h,i,1} & b_{h,i,0} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ b_{h,i,\tau} & b_{h,i,\tau-1} & \cdots & b_{h,i,0} \end{bmatrix}^k \begin{bmatrix} 1 \\ 0 \\ \cdots \\ 0 \end{bmatrix}.$$

Finally, we can re-arrange equation (20) into:

$$\begin{aligned} \mathbb{E}[u^{X_{t+h+1}} | \underline{X_t}] &= \exp\Big[A_{h+1,0} + \sum_{j=1}^{\tau} A_{h+1,j} u^j + O(u^{\tau+1}) \Big] \\ &= \exp(A_{h+1,0}) \sum_{k=0}^{\tau} \frac{1}{k!} \big(\sum_{j=1}^{\tau} A_{h+1,j} u^j \big)^k + O(u^{\tau+1}), \end{aligned} \tag{25}$$

where each $A_{h+1,j}, j = 0...\tau$ is linear in $\underline{X_{t-1}}$:

$$A_{h+1,j} = b_{h+1,0,j} + \sum_{i=1}^{p} b_{h,i,j} X_{t-i}. \tag{26}$$

Thus the Taylor's expansion of the RHS of equation (25) can be obtained by applying Proposition 1 again and the values of $\mathbb{P}[X_{t+h+1} = j | \underline{X_{t-1}}]$, $j = 0, ...\tau$, are obtained by coefficient matching:

$$\begin{bmatrix} p_{h+1}(0|\underline{X_{t-1}}) \\ p_{h+1}(1|\underline{X_{t-1}}) \\ \cdots \\ p_{h+1}(\tau-1|\underline{X_{t-1}}) \\ p_{h+1}(\tau|\underline{X_{t-1}}) \end{bmatrix} = \sum_{k=0}^{\tau} \frac{\exp(A_{h+1,0})}{k!} \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ A_{h+1,1} & 0 & 0 & \cdots & 0 \\ A_{h+1,2} & A_{h+1,1} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ A_{h+1,\tau} & A_{h+1,\tau-1} & \cdots & A_{h+1,1} & 0 \end{bmatrix}^k \begin{bmatrix} 1 \\ 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix}. \tag{27}$$

In terms of the computational effort, the Taylor's expansions of $B_{h,i}(u)$ do not depend on the values of $\underline{X_t}$ and thus only need to be computed once in an online forecasting framework. They necessitate only the multiplication between matrices and vectors of dimension $(\tau + 1)$, instead of $(\tau + 1)^p$ in the Markov chain approach. Thus the method is expected to be much faster than the Markov chain approach, especially for large p. Note also that although the computation of

the Taylor's expansion of $B_{h,i}$ is recursive, that of the $h-$step$-$ahead conditional p.m.f. does not necessitate a recursion. That is, if one is interested in the forecasting performance at a specific horizon H [see e.g. Chevillon (2007)], then she can use the above formula directly once $B_{H,i}, i = 0, ..., p$ are obtained, without computing the conditional p.m.f. at shorter horizons.

3.2 Examples

Let us now give the explicit forecasting formula in the four examples considered in Section 2.2.

Example 9 (INARCH(p) continued.)**.** We first consider the INARCH(p) model introduced in Example 5. We have:

$$
\begin{aligned}
B_{1,0}(u) &= \lambda_0(u-1), \\
B_{1,i}(u) &= \lambda_i(u-1), \qquad \forall i = 1, ..., p, \\
B_{h+1,0}(u) &= B(h,0)(u) + \lambda_0[e^{B_{h,1}(u)} - 1], \qquad (28) \\
B_{h+1,i}(u) &= \mathbb{1}_{i<p} B_{h,i+1}(u) + \lambda_i(e^{B_{h,1}(u)} - 1), \qquad \forall i = 1, ..., p. \qquad (29)
\end{aligned}
$$

By expanding (28) and (29) at $u = 0$, we get:

$$
\begin{bmatrix} b_{h+1,0,0} \\ b_{h+1,0,1} \\ \cdots \\ b_{h+1,0,\tau} \end{bmatrix} = \mathbb{1}_{i=0} \begin{bmatrix} b_{h,0,0} \\ b_{h,0,1} \\ \cdots \\ b_{h,0,\tau} \end{bmatrix} + \begin{bmatrix} -\lambda_0 \\ 0 \\ \cdots \\ 0 \end{bmatrix} + \lambda_0 \sum_{k=0}^{\tau} \frac{1}{k!} \begin{bmatrix} b_{h,i,0} & 0 & \cdots & 0 \\ b_{h,i,1} & b_{h,i,0} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ b_{h,i,\tau} & b_{h,i,\tau-1} & \cdots & b_{h,i,0} \end{bmatrix}^k \begin{bmatrix} 1 \\ 0 \\ \cdots \\ 0 \end{bmatrix},
$$

$$
\begin{bmatrix} b_{h+1,i,0} \\ b_{h+1,i,1} \\ \cdots \\ b_{h+1,i,\tau} \end{bmatrix} = \mathbb{1}_{0<i<p} \begin{bmatrix} b_{h,i+1,0} \\ b_{h,i+1,1} \\ \cdots \\ b_{h,i+1,\tau} \end{bmatrix} + \begin{bmatrix} -\lambda_i \\ 0 \\ \cdots \\ 0 \end{bmatrix} + \sum_{k=0}^{\tau} \frac{\lambda_i}{k!} \begin{bmatrix} b_{h,i,0} & 0 & \cdots & 0 \\ b_{h,i,1} & b_{h,i,0} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ b_{h,i,\tau} & b_{h,i,\tau-1} & \cdots & b_{h,i,0} \end{bmatrix}^k \begin{bmatrix} 1 \\ 0 \\ \cdots \\ 0 \end{bmatrix}.
$$

Example 10 (Poisson-INAR(p) continued)**.** Let us now consider the Poisson-INAR(p) model

introduced in Example 6. First, the recursion for functions $B_{h,i}$ is:

$$
\begin{aligned}
B_{1,0}(u) &= \lambda(u-1), \\
B_{1,i}(u) &= \log(\alpha_i u + 1 - \alpha_i), \qquad \forall i = 1, ..., p, \\
B_{h+1,0}(u) &= B_{h,0}(u) + \lambda[e^{B_{h,1}(u)} - 1], \\
B_{h+1,i}(u) &= \mathbb{1}_{i<p} B_{h,i+1}(u) + \log(\alpha_i e^{B_{h,1}(u)} + 1 - \alpha_i), \qquad \forall i = 1, ..., p.
\end{aligned}
$$

their Taylor's expansions at zero are:

$$
\begin{bmatrix} b_{h+1,0,0} \\ b_{h+1,0,1} \\ \cdots \\ b_{h+1,0,\tau} \end{bmatrix} = \mathbb{1}_{i=0} \begin{bmatrix} b_{h,0,0} \\ b_{h,0,1} \\ \cdots \\ b_{h,0,\tau} \end{bmatrix} + \begin{bmatrix} -\lambda \\ 0 \\ \cdots \\ 0 \end{bmatrix} + \lambda \sum_{k=0}^{\tau} \frac{1}{k!} \begin{bmatrix} b_{h,i,0} & 0 & \cdots & 0 \\ b_{h,i,1} & b_{h,i,0} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ b_{h,i,\tau} & b_{h,i,\tau-1} & \cdots & b_{h,i,0} \end{bmatrix}^k \begin{bmatrix} 1 \\ 0 \\ \cdots \\ 0 \end{bmatrix},
$$

$$
\begin{bmatrix} b_{h+1,i,0} \\ b_{h+1,i,1} \\ \cdots \\ b_{h+1,i,\tau} \end{bmatrix} = \mathbb{1}_{0<i<p} \begin{bmatrix} b_{h,i+1,0} \\ b_{h,i+1,1} \\ \cdots \\ b_{h,i+1,\tau} \end{bmatrix} + \begin{bmatrix} \log(1-\alpha_i) \\ 0 \\ \cdots \\ 0 \end{bmatrix} + \sum_{k=0}^{\tau} \frac{1}{k!} \Big[\sum_{j=1}^{\tau} \frac{(-1)^{j-1}\alpha_i^j}{(1-\alpha_i)^j} j^{k-1} \Big] \begin{bmatrix} b_{h,i,0} & 0 & \cdots & 0 \\ b_{h,i,1} & b_{h,i,0} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ b_{h,i,\tau} & b_{h,i,\tau-1} & \cdots & b_{h,i,0} \end{bmatrix}^k \begin{bmatrix} 1 \\ 0 \\ \cdots \\ 0 \end{bmatrix}
$$

Example 11 (NB-INAR(p) continued)**.** Let us now give the formula in the NB-INAR(p) model considered in Example 7. First, the recursion for functions $B_{h,i}$ is:

$$
\begin{aligned}
B_{1,0}(u) &= r_0 \log(1-q_0) - r_0 \log(1-q_0 u), \\
B_{1,i}(u) &= \log(\alpha_i u + 1 - \alpha_i), \qquad \forall i = 1, ..., p, \\
B_{h+1,0}(u) &= B_{h,0}(u) + r_0 \log(1-q_0) - r_0 \log(1 - q_0 e^{B_{h,1}(u)}), \\
B_{h+1,i}(u) &= \mathbb{1}_{i<p} B_{h,i+1}(u) + \log(\alpha_i e^{B_{h,1}(u)} + 1 - \alpha_i), \qquad \forall i = 1, ..., p.
\end{aligned}
$$

Their Taylor's expansions at zero are:

$$\begin{bmatrix} b_{h+1,0,0} \\ b_{h+1,0,1} \\ \cdots \\ b_{h+1,0,\tau} \end{bmatrix} = \mathbb{1}_{i=0} \begin{bmatrix} b_{h,0,0} \\ b_{h,0,1} \\ \cdots \\ b_{h,0,\tau} \end{bmatrix} + \begin{bmatrix} r_0 \log(1-q_0) \\ 0 \\ \cdots \\ 0 \end{bmatrix} + r_0 \sum_{k=0}^{\tau} \frac{1}{k!} \sum_{j=1}^{\tau} q_0^j j^{k-1} \begin{bmatrix} b_{h,i,0} & 0 & \cdots & 0 \\ b_{h,i,1} & b_{h,i,0} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ b_{h,i,\tau} & b_{h,i,\tau-1} & \cdots & b_{h,i,0} \end{bmatrix}^k \begin{bmatrix} 1 \\ 0 \\ \cdots \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} b_{h+1,i,0} \\ b_{h+1,i,1} \\ \cdots \\ b_{h+1,i,\tau} \end{bmatrix} = \mathbb{1}_{0<i<p} \begin{bmatrix} b_{h,i+1,0} \\ b_{h,i+1,1} \\ \cdots \\ b_{h,i+1,\tau} \end{bmatrix} + \begin{bmatrix} r_i \log(1-q_i) \\ 0 \\ \cdots \\ 0 \end{bmatrix}$$

$$+ \sum_{k=0}^{\tau} \frac{1}{k!} \Big[\sum_{j=1}^{\tau} \frac{(-1)^{j-1} \alpha_i^j}{(1-\alpha_i)^j} j^{k-1} \Big] \begin{bmatrix} b_{h,i,0} & 0 & \cdots & 0 \\ b_{h,i,1} & b_{h,i,0} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ b_{h,i,\tau} & b_{h,i,\tau-1} & \cdots & b_{h,i,0} \end{bmatrix}^k \begin{bmatrix} 1 \\ 0 \\ \cdots \\ 0 \end{bmatrix}.$$

Example 12 (NBAR(p) continued)**.** Finally, in the Poisson-INAR(p) model, the recursion for functions $B_{h,i}$ is:

$$\begin{aligned} B_{1,0}(u) &= r_0 \log(1-q_0) - r_0 \log(1-q_0 u), \\ B_{1,i}(u) &= r_i \log(1-q_i) - r_i \log(1-q_i u), \qquad \forall i = 1, ..., p, \\ B_{h+1,0}(u) &= B_{h,0}(u) + r_0 \log(1-q_0) - r_0 \log(1-q_0 e^{B_{h,1}(u)}), \\ B_{h+1,i}(u) &= \mathbb{1}_{i<p} B_{h,i+1}(u) + \log(\alpha_i e^{B_{h,1}(u)} + 1 - \alpha_i), \qquad \forall i = 1, ..., p., \end{aligned}$$

whereas their Taylor's expansion at zero are:

$$\begin{bmatrix} b_{h+1,0,0} \\ b_{h+1,0,1} \\ \cdots \\ b_{h+1,0,\tau} \end{bmatrix} = \mathbb{1}_{i=0} \begin{bmatrix} b_{h,0,0} \\ b_{h,0,1} \\ \cdots \\ b_{h,0,\tau} \end{bmatrix} + \begin{bmatrix} r_0 \log(1-q_0) \\ 0 \\ \cdots \\ 0 \end{bmatrix} + r_0 \sum_{k=0}^{\tau} \frac{1}{k!} \sum_{j=1}^{\tau} q_0^j j^{k-1} \begin{bmatrix} b_{h,i,0} & 0 & \cdots & 0 \\ b_{h,i,1} & b_{h,i,0} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ b_{h,i,\tau} & b_{h,i,\tau-1} & \cdots & b_{h,i,0} \end{bmatrix}^k \begin{bmatrix} 1 \\ 0 \\ \cdots \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} b_{h+1,i,0} \\ b_{h+1,i,1} \\ \cdots \\ b_{h+1,i,\tau} \end{bmatrix} = \mathbb{1}_{0<i<p} \begin{bmatrix} b_{h,i+1,0} \\ b_{h,i+1,1} \\ \cdots \\ b_{h,i+1,\tau} \end{bmatrix} + \begin{bmatrix} r_i \log(1-q_i) \\ 0 \\ \cdots \\ 0 \end{bmatrix} + \sum_{k=0}^{\tau} \frac{1}{k!} \Big[\sum_{j=1}^{\tau} r_i q_i^j j^{k-1} \Big] \begin{bmatrix} b_{h,i,0} & 0 & \cdots & 0 \\ b_{h,i,1} & b_{h,i,0} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ b_{h,i,\tau} & b_{h,i,\tau-1} & \cdots & b_{h,i,0} \end{bmatrix}^k \begin{bmatrix} 1 \\ 0 \\ \cdots \\ 0 \end{bmatrix}.$$

Finally, the conditional p.m.f. at horizon h is given by equation (27) with the same definition of coefficients $A_{h+1,0}$ as in (26).

3.3 The case of some infinite order models

Although our paper is mainly interested in models with finite order p, it can be straightforwardly extended to some infinite order, GARCH-type models. For instance, the Poisson-INGARCH(1,1) model of Ferland et al. (2006) postulates that:

$$X_t | \underline{X_{t-1}} \sim \mathcal{P}(\lambda_t), \qquad \text{where } \lambda_t = \rho\lambda_{t-1} + \lambda_0 + aX_{t-1}, \tag{30}$$

with $a > 0, \lambda_0 > 0$ and $\rho \in [0, 1[$. This model can be viewed as an infinite extension of model (3), since by the infinite divisibility of the Poisson distribution, we have the representation:

$$X_t = \sum_{i=1}^{\infty} Z_i * X_{t-i} + \epsilon_t \tag{31}$$

where $Z_i, i = 1, ...$ and ϵ_t follow Poisson distributions with parameters $a\rho^{i-1}$ and $\frac{\lambda_0}{1-\rho}$, respectively.

The above example can be extended by replacing the Poisson distribution by some other infinitely divisible distributions [see discussions in Gonçalves et al. (2015)]. For instance, Xu

et al. (2012) define a NB-INGARCH(1,1) process by:

$$X_t|\underline{X_{t-1}} \sim \mathcal{NB}(r_t, q), \qquad \text{where } r_t = \rho r_{t-1} + r_0 + aX_{t-1}, \tag{32}$$

and this process still has the infinite thinning representation (31) with Z_i and ϵ_t following $\mathcal{NB}(a\rho^{i-1}, q)$ and $\mathcal{NB}(\frac{r_0}{1-\rho}, q)$ distributions, respectively. Note that this definition is different from that given by Zhu (2011), which assumes instead that the conditional negative binomial distribution has a fixed degree of freedom parameter r but a time-varying probability parameter q_t. In particular, Zhu's model does not admit the thinning representation and its multi-step ahead conditional distribution is unknown.

The advantage of models of type (30) and (32) is that the one-step-ahead conditional distribution is tractable. Nevertheless, for horizon $h \geq 2$, the predictive distribution is (up to now) unknown. Let us now explain how the approach developed in our paper can be adapted to these models.

A first, universal solution to tackle infinite order models is to approximate an infinite order model by a finite order one, by truncating the summation in (31) at a large, deterministic order p [see Kirchner (2016) for a formal discussion of the Poisson-INGARCH(1,1) case]. Since the computational cost of our approach is rather insensitive to the lag p, this method would provide an excellent approximation, especially when the number of observations is relatively small and the infinite order model should anyway be truncated for estimation. The advantage of this method is that it also applies to any other infinite order thinning-based models with, for instance, Bernoulli distributed variables Z_i, for which the one-step-ahead conditional p.m.f. need not necessarily be simple [this has motivated the assumption of Poisson distribution on Z_i's in Kirchner (2016)].

However, given the popularity of INGARCH type models in the applied literature, it seems beneficial if an *exact* forecasting method could be derived. This is indeed the case and is due to the property that in both models (30) and (32), the augmented process (X_t, λ_t) is first-order Markov. As a consequence, for the Poisson-INGARCH(1,1) model, say, we have the following analogy of Proposition 2:

Proposition 4. *In the Poisson-INGARCH(1,1) model defined by* (30)*, the multi-step-ahead*

conditional p.g.f. is:

$$\mathbb{E}[u^{X_{t+h-1}}|\underline{X_{t-1}}] = \exp\Big[B_{h,1}(u)\lambda_t + B_{h,0}(u)\Big], \qquad \forall h \geq 1, \tag{33}$$

$$\textit{with } B_{1,1}(u) = u - 1, \qquad B_{1,0}(u) = 0, \tag{34}$$

$$B_{h+1,1}(u) = e^{aB_{h,1}(u)} - 1 + \rho B_{h,1}(u), \tag{35}$$

$$B_{h+1,0}(u) = B_{h,1}(u)\lambda_0 + B_{h,0}(u), \qquad \forall h \geq 0. \tag{36}$$

Proof. See Appendix 4. □

Using this lemma, we can mimic the procedure described in Section 3.1 to obtain recursively the Taylor's expansion of $B_{h,1}(u)$ and $B_{h,0}(u)$ at zero, and then expand the exponential function to recover the exact value of the $h-$step$-$ahead conditional p.m.f. Since the conditional p.g.f. given $\underline{X_{t-1}}$ is completely characterized by λ_t, the resulting computational cost is comparable to model (3) with $p = 1$.

Finally, Proposition 4 can also be adapted to the NB-INGARCH(1,1) model (32). Due to space constraint this analysis is relegated to Appendix 5.

4 Comparison with existing methods using simulated data

4.1 Likelihood function

To illustrate the computational gain, we first implement the conditional likelihood function, that is the likelihood function given the first p initial values of the process, using both our method and the direct application of (4). We consider a Poisson-INAR(2) model, whose parameters and initial values are fixed as:

$$\alpha_1 = 0.2, \alpha_2 = 0.2, \lambda \in \{1, 5, 10\}, X_1 = X_2 = 1, \tag{37}$$

as well as a Poisson-INAR(5) model with parameters and initial values set as:

$$\alpha_1 = 0.2, \alpha_2 = 0.2, \alpha_3 = 0.1, \alpha_4 = 0.1, \alpha_5 = 0.1, \lambda \in \{1, 5, 10\}, X_1 = \cdots = X_5 = 1. \tag{38}$$

In other words, in either of the two cases where $p = 2$ and $p = 5$, we simulate three models who differ only in terms of the value of λ. This will allow us to investigate how our algorithm performs when the average value of the observations becomes larger. For each of the simulated processes, the sample size is set to $T = 500$. For illustration purpose, Figure 1 reports the empirical marginal p.m.f.'s of the two INAR(2) processes. As we can see, when $\lambda = 1$, the whole process takes small values (at most 6), whereas when $\lambda = 10$, we can observe a large proportion of "large" counts.

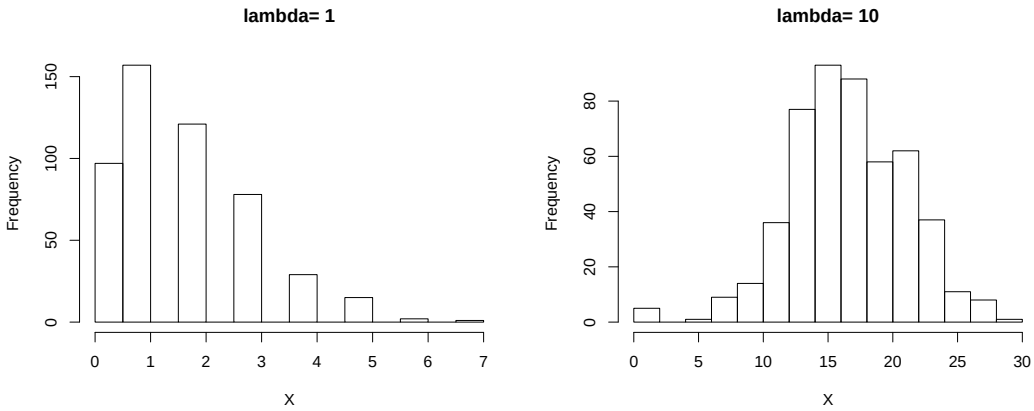

Figure 1: Empirical marginal p.m.f. of the simulated INAR(2) process with $\lambda = 1$ (left panel) and $\lambda = 10$ process (right panel).

Then we compute the likelihood function using our method to compare its speed with respect to the brute force convolution. For the latter approach we have adopted the recursive algorithm suggested by Bu et al. (2008), Thm 1. Both algorithms are implemented in R using the same laptop (intel i5, 3.0 GHz, 8GB RAM) and the program is available from the author upon request. Table 1 compares their respective CPU time. Note that we do not compare the likelihood values, since both methods provide the same, exact likelihood.

Value of p	$p=2$			$p=5$		
Value of λ	$\lambda=1$	$\lambda=5$	$\lambda=10$	$\lambda=1$	$\lambda=5$	$\lambda=10$
Simulated empirical mean	1.59	7.9	16.5	2.28	12.3	19.6
Simulated empirical variance	1.72	7.8	17.9	2.44	14.0	25.6
CPU of our method	0.006 s	0.015 s	0.033 s	0.007 s	0.02 s	0.05 s
CPU of convolution method	0.008 s	0.016 s	0.031 s	0.02 s	0.06s	0.11 s

Table 1: CPU time of the two methods for different values of p and λ. The time unit is second.

When order p is small (i.e. $p=2$), both methods are roughly equivalent and sufficiently fast. When p increases to 5, however, our method becomes more competitive. Indeed, its CPU time remains nearly unchanged, whereas the CPU of the direct method is increasing in n. As for the impact of λ (and hence the impact of the expectation of the process) on the CPU time, we can see that for both methods, when λ increases, the computation time increases. Nevertheless the speed of increase is (slightly) faster for our method than for the convolution method. An intuitive explanation of these comparison results is as follows. In Section 2, we have shown that computing the conditional p.m.f. $\ell(X_t|\underline{X_{t-1}})$ involves a complexity of $O(\frac{1}{2}X_t^3)$ using our method, or, roughly speaking, $O(\frac{1}{2}\lambda^3)$. On the other hand, Bu et al. (2008)'s method involves the convolution of $\alpha_1 \circ X_{t-1}$, ..., $\alpha_p \circ X_{t-p}$, and ϵ_t, by a $p-$step recursion. Each step involves the convolution of two integer-valued variables for $O(X_t^2)$ operations. Hence Bu et al. (2008)'s method has a complexity of $O(pX_t^2)$, or roughly $O(p\lambda^2)$. Thus when λ is fixed and p goes to infinity, our method becomes dominant; whereas when p is fixed but λ goes to infinity, the method of Bu et al. (2008) is more suitable. Note, however, that since count time series models are specifically designed for low counts, the case where λ is very large is somehow scarce and in that case, we might just as well approximate the count time series by a continuously valued time series and use standard continuous time series models.

4.2 Multi-step forecasting

We now illustrate how our method fares against the Markov chain approach of McCabe et al. (2011) and the Monte-Carlo approach [Jung and Tremayne (2006); Bisaglia and Canale (2016) in terms of both the computational time and precision, when it comes to higher-horizon forecasting. To this end we use the simulated processes (37). Given observations up to time T, we will compute

the conditional distribution of $X_{T+1}, ...,$ and X_{T+h} given $\underline{X_T}$.

Clearly, the quality of approximation of the Markov chain approach depends on the choice of the threshold τ. In the following we consider three such values: $\tau \in \{6, 10, 20\}$. As a comparison, by looking at the empirical p.m.f.'s of Figure 1, we can see that the probability that either of the two processes takes values larger than 6 is quite low.

Similarly, by the law of large numbers, the larger the number of replications N in the Monte-Carlo approach, the smaller the difference between the exact predictive distributions and their Monte-Carlo approximations. Thus we will consider three values of N, that are 5000, which is the number recommended by Jung and Tremayne (2006), as well as $N = 50000$ and $N = 5000000$.

Table 2 reports the CPU time and the conditional p.m.f.'s at horizons $h = 1, 5$ and 10 obtained from the three methods, given observations up to time T of the INAR(2) process, with $\lambda = 1$.

Method	Taylor's expansion, $\tau = 6$			Markov chain, $\tau = 6$			Monte-Carlo, $N = 5000$		
CPU time	0.003 s			0.008 s			0.6 s		
horizon h	$h = 1$	$h = 5$	$h = 10$	$h = 1$	$h = 5$	$h = 10$	$h = 1$	$h = 5$	$h = 10$
$p_h(0\|\underline{X_T})$	0.235	0.198	0.193	0.236	0.198	0.193	0.231	0.210	0.200
$p_h(1\|\underline{X_T})$	0.353	0.317	0.315	0.353	0.317	0.314	0.352	0.312	0.314
$p_h(2\|\underline{X_T})$	0.250	0.258	0.259	0.25	0.258	0.259	0.258	0.255	0.248
$p_h(6\|\underline{X_T})$	0.002	0.006	0.006	0.002	0.006	0.006	0.002	0.005	0.007
Method	Taylor's expansion, $\tau = 10$			Markov chain, $\tau = 10$			Monte-Carlo, $N = 50000$		
CPU time	0.005 s			0.06 s			6 s		
horizon h	$h = 1$	$h = 5$	$h = 10$	$h = 1$	$h = 5$	$h = 10$	$h = 1$	$h = 5$	$h = 10$
$p_h(0\|\underline{X_T})$	0.235	0.198	0.193	0.235	0.198	0.193	0.232	0.199	0.191
$p_h(1\|\underline{X_T})$	0.353	0.317	0.314	0.353	0.317	0.314	0.356	0.321	0.314
$p_h(2\|\underline{X_T})$	0.250	0.258	0.259	0.250	0.261	0.271	0.249	0.259	0.264
$p_h(6\|\underline{X_T})$	0.002	0.006	0.006	0.002	0.006	0.006	0.002	0.006	0.006
Method	Taylor's expansion, $\tau = 20$			Markov chain, $\tau = 20$			Monte-Carlo, $N = 500000$		
CPU time	0.005 s			1.2 s			60 s		
horizon h	$h = 1$	$h = 5$	$h = 10$	$h = 1$	$h = 5$	$h = 10$	$h = 1$	$h = 5$	$h = 10$
$p_h(0\|\underline{X_T})$	0.235	0.198	0.193	0.235	0.198	0.193	0.232	0.199	0.191
$p_h(1\|\underline{X_T})$	0.353	0.317	0.314	0.353	0.317	0.314	0.356	0.321	0.314
$p_h(2\|\underline{X_T})$	0.250	0.258	0.259	0.250	0.258	0.259	0.253	0.257	0.257
$p_h(6\|\underline{X_T})$	0.002	0.006	0.006	0.002	0.006	0.006	0.002	0.006	0.006

Table 2: Conditional p.m.f. at different horizons obtained by different methods for the Poisson-INAR(2) model. For expository purpose we only report their values at points $0, 1, 2,$ and 6.

The CPU times correspond to the calculation of $p_h(k|\underline{X_t})$ for all horizons $h = 1, 2, ..., 20$ *and* all arguments $k = 0, ..., \tau$. The Monte-Carlo approach is computationally the most intensive,with

a CPU time that is roughly proportional to the number of draws. Our approach, on the other hand, is the quickest, with a computational time that is slowly increasing in τ. As for the Markov chain approach, its CPU time is small when $\tau = 6$, but increases quite significantly in τ. Indeed, the Markov chain approach requires the computation of a matrix of dimension $(\tau+1)^p \times (\tau+1)^p$, whose computational cost is quadratic in τ for $p = 2$. Note that in this comparison, we have fixed the values of the parameters. In practice, it might also be preferable to take into account the uncertainty of the parameter forecast. In other words, if the simulation approach is used, one should first draw a large number of possible values of the parameter; then for each parameter value, a large number of draws of future path of the process are drawn. As a consequence, the computational cost would be even much larger than the 6 seconds reported in the table.

Let us now compare the p.m.f.'s obtained from the three approaches. As expected, for different values of τ, our method produces the same conditional p.m.f. at various horizons. For the Monte-Carlo approach, as N increases, the simulated conditional p.m.f.'s become closer and closer to the results obtained by our method, and for $N = 5000000$, both approaches provide roughly the same results. Finally, the p.m.f.'s obtained by the Markov chain approach are nearly (resp. exactly, up to a precision of 3 decimal places) the same as the Taylor's expansion. This is expected, since as τ increases, the quality of approximation in the Markov chain approach becomes better and better. Overall, in terms of precision, the Taylor's expansion and the Markov chain approaches strongly dominate the Monte-Carlo approach.

Next, let us illustrate that the advantage of our approach still exists for other thinning-type models satisfying (3). To this end we report a similar experiment for a NBAR(2) process. For illustration we assume that Z_1, Z_2 and ϵ_t in equation (3) all follow a special NB distribution: the geometric distribution, which has a degree of freedom parameter $r = 1$. Their probability parameters q_1, q_2 are set such that their respective expectations match the corresponding Bernoulli and Poisson parameters in equation (2), that is:

$$\frac{q_1}{1-q_1} = \frac{q_2}{1-q_2} = 0.2.$$

This makes the marginal mean of this NBAR(2) process to be the same as the above Poisson-INAR(2) process (and equal to $\frac{\mathbb{E}[\epsilon_t]}{1-\mathbb{E}[Z_1]-\mathbb{E}[Z_2]}$ under the stationarity condition $\mathbb{E}[Z_1]+\mathbb{E}[Z_2] < 1$),

hence facilitating the comparison between the two tables. The p.m.f.'s are reported in Table 3 and the conclusions we can draw are similar as for Table 2 and are thus omitted.

Method	Taylor's expansion, $\tau = 6$			Markov chain, $\tau = 6$			Monte-Carlo, $N = 5000$		
CPU time	0.003 s			0.008 s			0.6 s		
horizon h	$h = 1$	$h = 5$	$h = 10$	$h = 1$	$h = 5$	$h = 10$	$h = 1$	$h = 5$	$h = 10$
$p_h(0\|\underline{X_T})$	0.379	0.344	0.339	0.379	0.344	0.339	0.375	0.357	0.332
$p_h(1\|\underline{X_T})$	0.265	0.250	0.250	0.265	0.250	0.250	0.270	0.247	0.260
$p_h(2\|\underline{X_T})$	0.159	0.162	0.163	0.159	0.162	0.163	0.163	0.162	0.158
$p_h(6\|\underline{X_T})$	0.014	0.020	0.021	0.014	0.020	0.021	0.015	0.020	0.019
Method	Taylor's expansion, $\tau = 10$			Markov chain, $\tau = 10$			Monte-Carlo, $N = 50000$		
CPU time	0.005 s			0.06 s			0.6 s		
horizon h	$h = 1$	$h = 5$	$h = 10$	$h = 1$	$h = 5$	$h = 10$	$h = 1$	$h = 5$	$h = 10$
$p_h(0\|\underline{X_T})$	0.379	0.344	0.339	0.379	0.344	0.339	0.378	0.343	0.340
$p_h(1\|\underline{X_T})$	0.265	0.250	0.250	0.265	0.250	0.250	0.263	0.252	0.246
$p_h(2\|\underline{X_T})$	0.159	0.162	0.163	0.159	0.162	0.163	0.160	0.161	0.167
$p_h(6\|\underline{X_T})$	0.014	0.020	0.021	0.014	0.020	0.021	0.014	0.021	0.020
Method	Taylor's expansion, $\tau = 20$			Markov chain, $\tau = 20$			Monte-Carlo, $N = 500000$		
CPU time	0.005 s			1.2 s			60 s		
horizon h	$h = 1$	$h = 5$	$h = 10$	$h = 1$	$h = 5$	$h = 10$	$h = 1$	$h = 5$	$h = 10$
$p_h(0\|\underline{X_T})$	0.379	0.344	0.339	0.379	0.344	0.339	0.38	0.344	0.340
$p_h(1\|\underline{X_T})$	0.265	0.250	0.250	0.265	0.250	0.250	0.265	0.250	0.250
$p_h(2\|\underline{X_T})$	0.159	0.162	0.163	0.159	0.162	0.163	0.159	0.163	0.163
$p_h(6\|\underline{X_T})$	0.014	0.020	0.021	0.014	0.020	0.021	0.015	0.020	0.021

Table 3: Conditional p.m.f. at different horizons obtained by different methods for the NBAR(2) model. For expository purpose we only report their values at points $0, 1, 2$, and 6.

In the next exercise, we compare the computational cost of forecasting up to horizon 20 using different approaches, when we increase progressively the lag p of the Poisson-INAR(p) model, from $p = 2$ to 5. The individual models are obtained recursively and backwardly from the INAR(5), by dropping each time the last term in equation (3), while keeping the other parameters unchanged. We have not reported the comparison of the precision of the p.m.f.'s, as these results are rather similar across different values of p.

Method	Taylor's expansion			Markov chain			Monte-Carlo, $N = 50000$		
	$\tau = 6$	$\tau = 10$	$\tau = 20$	$\tau = 6$	$\tau = 10$	$\tau = 20$	$N = 5000$	$N = 50000$	$N = 500000$
$p = 2$	0.003	0.005	0.006	0.008	0.06	1.2	0.6	6	60
$p = 3$	0.004	0.006	0.006	0.20	0.4	5.2	0.8	9	58
$p = 4$	0.004	0.006	0.006	6.0	50	NA	0.11	10	90
$p = 5$	0.004	0.006	0.007	180	NA	NA	0.13	12	116

Table 4: CPU time for forecasting in a Poisson-INAR(5) model, using different approaches.

We can see that for the Markov chain approach, the CPU time is often missing when τ and p are large, indicating a failure of the computer program in these cases. This is mainly due to the necessity to compute a matrix of dimension $(\tau+1)^p \times (\tau+1)^p$. For instance, if $p = 5$, $\tau = 10$, the algorithm requires a flash memory of 25GB, which exceeds the capacity of a standard PC. As a consequence, while for small values of p, the Markov chain approach is an acceptable (but more costly and less precise) alternative to our approach, it becomes infeasible for higher-order models.

To conclude this section, Table 5 summarizes the above comparison in terms of three major criteria. We can see that our method combines the strength of its two competitors, while at the same time mitigating their downsides.

Method	Taylor's expansion	Markov chain	Monte-Carlo
Computation accuracy	Very High	High	Low
Sensitivity of CPU time w.r.t. threshold τ	Low	Very High	NA
Sensitivity of CPU time w.r.t. lag p	Low	Very High	High

Table 5: Comparison of the three approaches in terms of three major criteria. The sensitivity concerns the computational cost for forecasting.

Finally, note that while for expository purpose we have limited our comparison to models with order p non larger than 5, there is little doubt that for larger p, the advantage of our method becomes more and more important.

5 Conclusion

We have solved the open computational difficulty concerning the predictive distributions in thinning-based, higher-order count processes models. This family is quite large and includes, among others, Poisson-INARCH(p), Poisson-INAR(p), NB-INAR(p), NBAR(p), as well as some infinite order models such as Poisson-INGARCH(1,1). Existing methods usually suffer from one or several of the following downsides, such as:

- introducing significant approximation errors [see e.g. Pedeli et al. (2015)]
- being time consuming [see e.g. McCabe et al. (2011); Jung and Tremayne (2006)]
- being specific to the one-step-ahead predictive distribution [see Bu et al. (2008)]

- being specific to INAR(p) models [see Pedeli et al. (2015)]

Our approach is exact, generically faster than competing methods as demonstrated by simulations, and is unified in the sense that it can be used for forecasting at *any* horizons, for a large family of models. Our method is based on i) the simple relationship between the p.g.f. and p.m.f. for a count distribution; ii) the CaR property of these processes. Because of its reduced computational cost and wide applicability, our method significantly facilitates the practical implementation of many of these higher-order models.

Finally, is it possible to extend the result of this paper to bivariate count processes? Recently Darolles et al. (2019) show that for a *specific* bivariate extension of the INAR(p) models, this is indeed the case. However a systematic treatment of this question requires substantially more investigation and is left for future research.

References

Al-Osh, M. and Alzaid, A. A. (1987). First-order Integer-valued Autoregressive (INAR(1)) Process. *J. Time Series Anal.*, 8(3):261–275.

Alzaid, A. and Al-Osh, M. (1990). An Integer-valued pth-order Autoregressive Structure (INAR (p)) Process. *J. Appl. Probab.*, 27(2):314–324.

Bien, K., Nolte, I., and Pohlmeier, W. (2011). An Inflated Multivariate Integer Count Hurdle Model: An Application to Bid and Ask Quote Dynamics. *J. Appl. Econometrics*, 26(4):669–707.

Bisaglia, L. and Canale, A. (2016). Bayesian Nonparametric Forecasting for INAR Models. *Comput. Statist. Data Anal.*, 100:70–78.

Blundell, R., Griffith, R., and Van Reenen, J. (1999). Market Share, Market Value and Innovation in a Panel of British Manufacturing Firms. *Rev. Econ. Stud.*, 66(3):529–554.

Böckenholt, U. (1998). Mixed INAR (1) Poisson Regression Models: Analyzing Heterogeneity and Serial Dependencies in Longitudinal Count Data. *J. Econometrics*, 89(1-2):317–338.

Brännäs, K. and Hellström, J. (2001). Generalized integer-valued autoregression. *Econometric Rev.*, 20(4):425–443.

Bu, R. and McCabe, B. (2008). Model Selection, Estimation and Forecasting in INAR (p) Models: A Likelihood-based Markov Chain Approach. *International Journal of Forecasting*, 24(1):151–162.

Bu, R., McCabe, B., and Hadri, K. (2008). Maximum Likelihood Estimation of Higher-Order Integer-Valued Autoregressive Processes. *J. Time Series Anal.*, 29(6):973–994.

Chevillon, G. (2007). Direct multi-step estimation and forecasting. *Journal of Economic Surveys*, 21(4):746–785.

Darolles, S., Gouriéroux, C., and Jasiak, J. (2006). Structural Laplace Transform and Compound Autoregressive Models. *J. Time Series Anal.*, 27(4):477–503.

Darolles, S., Le Fol, G., Lu, Y., and Sun, R. (2019). Bivariate integer-autoregressive process with an application to mutual fund flows. *J. Multivariate Anal.*, 173:181–203.

Drost, F. C., Akker, R. v. d., and Werker, B. J. (2009). Efficient Estimation of Auto-Regression Parameters and Innovation Distributions for Semiparametric Integer-Valued AR (p) models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 71(2):467–485.

Du, J. and Li, Y. (1991). The Integer-Valued Autoregressive (INAR (p)) Model. *J. Time Series Anal.*, 12(2):129–142.

Ferland, R., Latour, A., and Oraichi, D. (2006). Integer-Valued GARCH Process. *J. Time Series Anal.*, 27(6):923–942.

Fernández-Fontelo, A., Fontdecaba, S., Alba, A., and Puig, P. (2017). Integer-valued AR processes with Hermite innovations and time-varying parameters: An application to bovine fallen stock surveillance at a local scale. *Stat. Model.*, 17(3):172–195.

Gonçalves, E., Mendes-Lopes, N., and Silva, F. (2015). Infinitely divisible distributions in integer-valued garch models. *J. Time Series Anal.*, 36(4):503–527.

Gordy, M. B. (2002). Saddlepoint Approximation of CreditRisk$^+$. *Journal of Banking & Finance*, 26(7):1335–1353.

Gouriéroux, C. and Jasiak, J. (2004). Heterogeneous INAR (1) Model with Application to Car Insurance. *Insurance Math. Econom.*, 34(2):177–192.

Gouriéroux, C. and Lu, Y. (2019). Negative Binomial Autoregressive Process with Stochastic Intensity. *J. Time Series Anal.*, 40(2):225–247.

Harris, D. and McCabe, B. (2018). Semiparametric independence testing for time series of counts and the role of the support. *Econometric Theory*, pages 1–35.

Heinen, A. and Rengifo, E. (2007). Multivariate Autoregressive Modeling of Time Series Count Data using Copulas. *Journal of Empirical Finance*, 14(4):564–583.

Jazi, M. A., Jones, G., and Lai, C.-D. (2012). First-order Integer Valued AR Processes with Zero Inflated Poisson Innovations. *J. Time Series Anal.*, 33(6):954–963.

Jung, R. C. and Tremayne, A. (2006). Coherent Forecasting in Integer Time Series Models. *International Journal of Forecasting*, 22(2):223–238.

Kirchner, M. (2016). Hawkes and INAR (∞) Processes. *Stochastic Process. Appl.*, 126(8):2494–2525.

Kirchner, M. (2017). An Estimation Procedure for the Hawkes Process. *Quant. Finance*, 17(4):571–595.

Latour, A. (1998). Existence and stochastic structure of a non-negative integer-valued autoregressive process. *J. Time Series Anal.*, 19(4):439–455.

McCabe, B. and Martin, G. (2005). Bayesian Predictions of Low Count Time Series. *International Journal of Forecasting*, 21(2):315–330.

McCabe, B. P., Martin, G. M., and Harris, D. (2011). Efficient Probabilistic Forecasts for Counts. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 73(2):253–272.

McKenzie, E. (1985). Some Simple Models for Discrete Variate Time Series. *Journal of the American Water Resources Association*, 21(4):645–650.

Pedeli, X., Davison, A. C., and Fokianos, K. (2015). Likelihood Estimation for the INAR (p) Model by Saddlepoint Approximation. *J. Amer. Statist. Assoc.*, 110(511):1229–1238.

Ristić, M. M., Bakouch, H. S., and Nastić, A. S. (2009). A New Geometric First-order Integer-valued Autoregressive (NGINAR (1)) Process. *J. Statist. Plann. Inference*, 139(7):2218–2226.

Rydberg, T. H. and Shephard, N. (2000). A modelling framework for the prices and times of trades made on the new york stock exchange. *Nonlinear and nonstationary signal processing*, pages 217–246.

Schweer, S. and Weiß, C. H. (2014). Compound Poisson INAR (1) Processes: Stochastic Properties and Testing for Overdispersion. *Comput. Statist. Data Anal.*, 77:267–284.

Scotto, M. G., Weiß, C. H., and Gouveia, S. (2015). Thinning-Based Models in the Analysis of Integer-Valued Time Series: A Review. *Stat. Model.*, 15(6):590–618.

Venkataraman, K. and Nanthi, K. (1982). A limit theorem on a subcritical galton-watson process with immigration. *Ann. Probab.*, 10(4):1069–1074.

Weiß, C. H. (2008). Thinning operations for modeling time series of countsâa survey. *AStA Adv. Stat. Anal.*, 92(3):319.

Weiß, C. H. (2010). Inarch (1) processes: Higher-order moments and jumps. *Statist. Probab. Lett.*, 80(23-24):1771–1780.

Xu, H.-Y., Xie, M., Goh, T. N., and Fu, X. (2012). A model for integer-valued time series with conditional overdispersion. *Comput. Statist. Data Anal.*, 56(12):4229–4242.

Zheng, H., Basawa, I. V., and Datta, S. (2006). Inference for pth-order random coefficient integer-valued autoregressive processes. *J. Time Series Anal.*, 27(3):411–440.

Zhu, F. (2011). A negative binomial integer-valued garch model. *J. Time Series Anal.*, 32(1):54–67.

Zhu, R. and Joe, H. (2009). Modelling Heavy-Tailed Count Data Using a Generalised Poisson-inverse Gaussian Family. *Statist. Probab. Lett.*, 79(15):1695–1703.

Zhu, R. and Joe, H. (2010). Negative Binomial Time Series Models Based on Expectation Thinning Operators. *J. Statist. Plann. Inference*, 140(7):1874–1888.

Corresponding author:

Yang Lu, Department of Economics (CEPN, UMR CNRS), University of Paris 13.

99 Av. JB Clement, 93430 Villetaneuse, France.

Email: yang.lu@univ-paris13.fr

Appendix 1: Proof of Proposition 2

Let us proceed by induction. The case where $h = 1$ is a direct consequence of Equation (5). Assume that (20) holds for a certain horizon h, then:

$$\begin{aligned}
&\mathbb{E}[u^{X_{t+h}}|\underline{X_{t-1}}] = \mathbb{E}\Big[\mathbb{E}[u^{X_{t+h}}|\underline{X_t}] \mid \underline{X_{t-1}}\Big] \\
&= \mathbb{E}[\exp\Big[B_{h,0}(u) + \sum_{i=1}^{p} B_{h,i}(u)X_{t+1-i}\Big]|\underline{X_{t-1}}] \\
&= \exp\Big[B_{h,0}(u) + \sum_{i=2}^{p} B_{h,i}(u)X_{t+1-i}\Big]\mathbb{E}\Big[\exp\Big(B_{h,1}(u)X_t\Big)|\underline{X_{t-1}}\Big] \\
&= \exp\Big[B_{h,0}(u) + \sum_{i=2}^{p} B_{h,i}(u)X_{t+1-i} + g_0(e^{B_{h,1}(u)}) + \sum_{i=1}^{p} g_i(e^{B_{h,1}(u)})\Big].
\end{aligned} \tag{39}$$

Thus (20) holds for any h.

Appendix 2: Proof of Proposition 3

We have:

$$\begin{aligned}
B_{h+1,0}(u) &= B_{h,0}(u) + c_{0,0} + \sum_{j=1}^{\tau} c_{0,j} e^{jB_{h,1}(u)} + O(u^{\tau+1}) \\
&= B_{h,0}(u) + c_{0,0} + \sum_{j=1}^{\tau} c_{0,j} \sum_{k=0}^{\tau} \frac{B_{h,1}^k(u) j^k}{k!} + O(u^{\tau+1}) \\
&= B_{h,0}(u) + c_{0,0} + \sum_{k=0}^{\tau} \frac{B_{h,1}^k(u)}{k!} \sum_{j=1}^{\tau} c_{0,j} j^k + O(u^{\tau+1}),
\end{aligned}$$

and the proof of the second equality for $B_{h+1,i}(u)$ is similar and omitted.

Appendix 3: Expanding the log p.g.f. using the p.m.f.

In this section we discuss how the Taylor's expansion of the log p.g.f. can be obtained from the p.m.f. We have:

$$\begin{aligned}
\log\Big(\mathbb{E}[u^X]\Big) &= \log\Big[p_0 + p_1 u + p_2 u^2 + \cdots + p_\tau u^\tau + O(u^{\tau+1})\Big] \\
&= \log p_0 + \log\Big[1 + \frac{p_1}{p_0}u + \frac{p_2}{p_0}u^2 + \cdots + \frac{p_\tau}{p_0}u^\tau + O(u^{\tau+1})\Big] \\
&= \log p_0 + \sum_{k=1}^{n} \frac{(-1)^k}{k}\Big[\frac{p_1}{p_0}u + \frac{p_2}{p_0}u^2 + \cdots + \frac{p_\tau}{p_0}u^\tau\Big]^k + O(u^{\tau+1}), \qquad (40)
\end{aligned}$$

for a fixed integer n. Thus by applying Proposition 1, the vector of the first $n+1$ coefficients of the Taylor's expansion of $\log\Big(\mathbb{E}[u^X]\Big)$ are equal to:

$$\begin{bmatrix} \log p_0 \\ 0 \\ \cdots \\ 0 \\ 0 \end{bmatrix} + \sum_{k=1}^{\tau} \frac{(-1)^k}{k} \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ \frac{p_1}{p_0} & 0 & 0 & \cdots & 0 \\ \frac{p_2}{p_0} & \frac{p_1}{p_0} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{p_\tau}{p_0} & \frac{p_{\tau-1}}{p_0} & \cdots & \frac{p_1}{p_0} & 0 \end{bmatrix}^k \begin{bmatrix} 1 \\ 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix}.$$

Remark 1. Clearly, the above series expansion requires $p_0 = \mathbb{P}[X = 0]$ to be non zero. Let us discuss the case where this condition is not satisfied by either Z_i or ϵ_t in model (3).

- If either of the Z_i's is such that $Z_i \geq 1$ almost surely, then we have: $X_{t+1} \geq X_t$ almost surely for each t, with a nonzero probability that the inequality holds strictly. Thus the process is explosive, nonstationary. This is out of the scope of the present paper.

- If the distribution of ϵ_t is such that $\mathbb{P}[\epsilon_t = 0] = \cdots = \mathbb{P}[\epsilon_t = K-1] = 0$ and $\mathbb{P}[\epsilon_t = K] > 0$, then any random variable following this distribution admits the stochastic representation: $X = K + Z$, where Z is again a nonnegative count variable. Thus in this case it suffices to consider the process $(Z_t) = (X_t) - K$ instead of the initial process (X_t). Then the above

algorithm continues to be applicable.

Appendix 4: Proof of Proposition 4

We proceed by induction. The initial condition $h = 1$ is obvious. Let us assume that equations (35) and (36) are satisfied for a certain $h \geq 1$. Then we have:

$$\begin{aligned}
\mathbb{E}[u^{X_{t+h}}|\underline{X_{t-1}}] &= \mathbb{E}\Big[\mathbb{E}[u^{X_{t+h}}|\underline{X_t}] \mid \underline{X_{t-1}}\Big] \\
&= \mathbb{E}\Big[e^{B_{h,1}(u)\lambda_{t+1}+B_{h,0}(u)} \mid \underline{X_{t-1}}\Big] \\
&= \mathbb{E}\Big[e^{B_{h,1}(u)[\rho\lambda_t+\lambda_0+aX_t]+B_{h,0}(u)} \mid \underline{X_{t-1}}\Big] \\
&= e^{\rho B_{h,1}(u)\lambda_t+B_{h,1}(u)\lambda_0+B_{h,0}(u)}\mathbb{E}[(e^{aB_{h,1}(u)})^{X_t} \mid \underline{X_{t-1}}] \\
&= e^{\rho B_{h,1}(u)\lambda_t+B_{h,1}(u)\lambda_0+B_{h,0}(u)+\lambda_t(e^{aB_{h,1}(u)}-1)}.
\end{aligned}$$

Hence (35) and (36) are also satisfied for $h + 1$.

Appendix 5: The predictive distribution in NB-INGARCH(1,1) model

In this section we establish the analogue of Proposition 4 for NB-INGARCH(1,1) model defined by (32). We have:

Proposition 5. *In the NB-INGARCH(1,1) model defined by* (32)*, the multi-step-ahead conditional p.g.f. is:*

$$\begin{aligned}
\mathbb{E}[u^{X_{t+h-1}}|\underline{X_{t-1}}] &= \exp\Big[B_{h,1}(u)\lambda_t + B_{h,0}(u)\Big], \qquad \forall h \geq 1, \\
\textit{with } B_{1,1}(u) &= \log(\frac{1-q}{1-qu}), \qquad B_{1,0}(u) = 0, \\
B_{h+1,1}(u) &= \log(\frac{1-q}{1-qe^{aB_{h,1}(u)}}) + \rho B_{h,1}(u), \\
B_{h+1,0}(u) &= B_{h,1}(u)\lambda_0 + B_{h,0}(u), \qquad \forall h \geq 0.
\end{aligned}$$

The proof (by induction) is similar to the proof of Proposition 4 and is thus omitted.