Noncausal Count Processes: A Queuing Perspective

Christian Gouriéroux[*] and Yang Lu[†]

First version, September 2019

Abstract

We introduce noncausal processes to the count time series literature. These processes are defined by time-reversing an INAR(1) process, a non-INAR(1) Markov affine count process, or a random coefficient INAR(1) [RCINAR(1)] process. We show that for INAR(1) and RCINAR(1), the causal process and its noncausal counterpart are closely related through a same queuing system with different stochastic specifications. The noncausal processes we introduce are generically time irreversible and have some unique calendar time dynamic properties that are unreplicable by existing causal models. In particular they allow for locally bubble-like explosion, while at the same time being stationary. These processes have closed form calendar time conditional probability mass function, which facilitates nonlinear forecasting.

Keywords: Bubble, Discrete Stable Distribution, Noncausal Process, Infinite Server Queue, Time Reversibility.

Acknowledgement: The first author gratefully acknowledges the financial support of the ACPR/Risk Foundation Chair: Regulation and Systemic Risk, and the ERC DYSMOIA. The second author thanks CNRS and the Labex MME-DII for délégation grants, as well as the Society of Actuaries for an individual grant.

[*] University of Toronto, Toulouse School of Economics and CREST. Email: gouriero@ensae.fr

[†] University of Paris 13, CEPN and THEMA. Email: yang.lu@univ-paris13.fr

1 Introduction

Recently, real-valued linear noncausal processes have raised interest in the econometric literature thanks to their ability of capturing irreversible[1], bubble-like phenomena, that are widely observed in financial applications [see e.g. Lanne and Saikkonen (2013); Gouriéroux and Zakoïan (2017)]. However, these models are not suitable for (low) count processes. To motivate the need for introducing “bubbles" for count processes, we display in Figure 1 the hourly amount of rainfall, measured by a rain gauge as multiples of mm in an Indian city.[2]

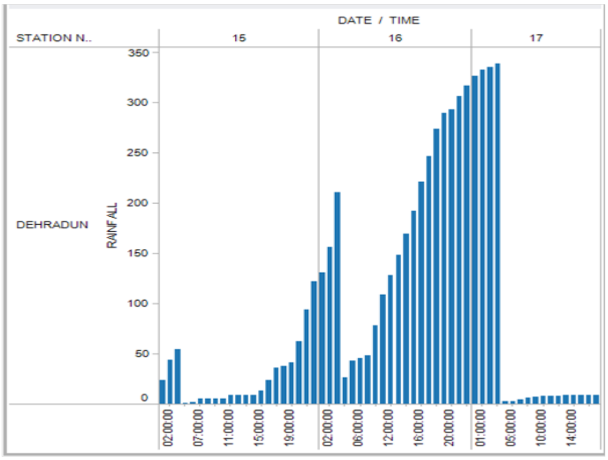

Figure 1: Hourly rainfall on June 15th, 16th and 17th in 2013 in Dehradun, Uttarakhand, India. This extreme rainfall resulted in the 2013 North India floods, with a casualty of more than 5000 killed.

This kind of rainfall data is very important for the prediction of floods. We can see that the amount process features cycles of “bubbles", or cloudbusts, characterized by a period of continuous increase of the rainfall intensity analogous to the accummulation phase of a financial bubble, followed by a sharp decrease, which spells the end of the cloudbust, that is the analog of the burst of the bubble. In the aforementioned literature on continuously valued time series, it has been shown that unlike standard autoregressive type processes, a new class of noncausal processes can capture the bubble phenomenon very well. A similar problem exists for count

[1]A process is time reversible (resp. irreversible) if and only if its dynamics is the same (resp. different) in both time directions. For a Markov process (X_t), this condition is equivalent to the symmetry (resp. asymmetry) of the distribution of (X_t, X_{t+1}).

[2]Thus these rainfall amounts are integer-valued. This figure is downloaded from the website of the Indian Water Portal: `https://www.indiawaterportal.org/articles/playing-uttarakhand-rainfall-data`

time series, since the path of a standard, say, INAR(1) process usually only features occasional abrupt positive jump followed by steady decrease, which is opposite to the phenomenon observed in Figure 1. The aim of our paper is to formally introduce the notion of noncausality to the count process literature. As the first step in this direction, we focus on Markov processes of the type:

$$X_{t+1} = \sum_{i=1}^{X_t} Z_{i,t+1} + \epsilon_{t+1}, \qquad \forall t, \tag{1}$$

where latent counts $Z_{i,t+1}$'s are independent of $\underline{X_t} = \{X_t, X_{t-1}, ...\}$, whereas the shocks ϵ_{t+1}'s are i.i.d., independent of the $\underline{X_t}$ and $Z_{i,t+1}$'s. The decomposition (1) leads to different dynamics depending on the distributions selected for the $Z_{i,t+1}$'s and for ϵ_{t+1}:

a) If the $Z_{i,t+1}$'s are i.i.d. Bernoulli variables, then process (X_t) is integer-autoregressive of order 1 [INAR(1)], [see McKenzie (1985), Al-Osh and Alzaid (1987)]. In the original model of McKenzie (1985), it is also assumed that the shocks ϵ_t's follow a Poisson distribution, and Schweer (2015) shows that process (X_t) is time-reversible, if and only if ϵ_t is Poisson distributed. An INAR(1) process can be viewed as the decomposition of the total number of customers X_{t+1} into the sum of new arrivals ϵ_{t+1} and staying old customers $\sum_{i=1}^{X_t} Z_{i,t+1}$ [see McCabe and Martin (2005); Schweer and Wichelhaus (2015)]. The INAR(1) is the most famous count process model and a special attention will be paid in the paper to this family, as well as its queuing interpretation.

b) The counts $Z_{i,t+1}$'s can also be i.i.d. only, but non binary. This larger family includes, among others, the negative binomial autoregressive process [NBAR(1), see Gouriéroux and Lu (2019a)], in which the $Z_{i,t+1}$'s have geometric distribution and ϵ_{t+1} is negative binomial with the same probability parameter, and the INARCH(1) [Weiss (2010)], in which both the $Z_{i,t+1}$'s and ϵ_{t+1} are Poisson[3]. These processes are also called Galton-Watson process in the probability literature [see e.g. Klebaner and Sagitov (2002)] and the term $(\sum_{i=1}^{X_t} Z_{i,t+1})$ can be interpreted as the number of descendants from the X_t existing customers, each of which can independently give rise to 0, 1, 2... descendants.

c) Alternatively, one can also extend the INAR(1) model *a*) by relaxing the i.i.d. assumption

[3]See also a complete list of existing affine count processes in Lu (2018).

of $Z_{i,t+1}$'s and assume that they are only conditionally i.i.d. given a stochastic probability parameter α_{t+1}, which itself is an i.i.d. sequence that is independent of ϵ_{t+1}, as well as past observations $\underline{X_t}$. This model is called random coefficient INAR(1), or RCINAR(1) [see e.g. Zheng et al. (2007)]. This model has a similar queuing interpretation as the INAR(1) process, but the probability of leaving of current customers α_{t+1} is time-varying instead of being constant.

Equation (1) defines a dynamic factor model with a random number of factors ϵ_{t+1}, $Z_{1,t+1}$,..., and $Z_{X_t,t+1}$ for date t. Since at each date $t+1$, this number, i.e. X_t+1, is strictly larger than the number of observables, i.e. 1, neither shocks ϵ_{t+1}, nor variables $Z_{i,t+1}$ can be deterministically recovered (except for dates with $X_t = 0$). Nevertheless, even if the underlying factors are not recoverable, their distributions can be identified from the observation of (X_t) only.

The major appeal of the INAR(1) model *a*) and its affine extension *b*) is that the conditional probability generating function (p.g.f.) of X_{t+1} given $\underline{X_t}$, which characterizes the transition distribution, is conveniently given by:

$$\mathbb{E}[u^{X_{t+1}}|\underline{X_t}] = \Big(\mathbb{E}[u^{Z_{i,t+1}}]\Big)^{X_t}\mathbb{E}[u^{\epsilon_{t+1}}]. \tag{2}$$

This conditional p.g.f. is an exponential affine function of the conditioning variable X_t. In particular (X_t) is a Markov process with respect to the filtration $\underline{X_t}$. Such Markov processes are called affine (or compound autoregressive) [see Darolles et al. (2006)]. In the count process context, Lu (2018) shows that they have tractable marginal and predictive distributions at any horizons.

Model (1) is called causal, since X_{t+1} does not depend on the values of the future errors $\epsilon_{t+2}, \epsilon_{t+3}, ...,$ or future latent factors $Z_{i,t+2}, Z_{i,t+3},$ i varying. The noncausal count process introduced in this paper can be defined simply as a Markov process whose reverse time dynamics is causal affine. Our main contributions are the following. First, we show that for noncausal INAR(1) and noncausal RCINAR(1), an alternative, equivalent specification is based on the aforementioned queuing interpretation, but with a different set of distributional assumptions from the causal queuing interpretation. In particular, under the noncausal specification, the number of arrivals are typically heteroscedastic across time and are dependent of the current population size. Second, we characterize all the time reversible count processes satisfying model (1). For INAR(1) models *a*) only the Poisson-INAR(1) process is reversible; as for non-INAR(1), affine models *b*), only the NBAR(1) process is reversible; among RCINAR(1) models *c*), the only reversible process is another (marginally) negtive binomial process introduced by Joe (1996). These examples are to be discarded from a modelling point of view since they are unable to capture the (time) asymmetry of the bubble dynamics. Thirdly, we show that when process (1) is time irrevrsible, its noncausal version is indeed suitable for capturing bubbles. In particular, we show that when an affine [including INAR(1)] process is time reversed, the resulting noncausal process has necessarily a non-affine dynamics. We then derive the closed form formula of this conditional predictive p.m.f., and through various examples we demonstrate that it often features multiple modes, with one near zero (corresponding to the burst of the bubble) and another one away from zero (corresponding to the continuous increase of the bubble).

The rest of the paper is organized as follows. Section 2 studies the general, deterministic properties of a queuing system with an infinite capacity. Section 3 explores the stochastic specification of the queuing system such that the queue length has a dynamics compatible with model (1). Section 4 derives the general properties of a time-reversed count process such as time reversibility. We also give examples of noncausal INAR(1) processes allowing for closed form conditional p.m.f. Section 5 derives the queueing interpretation of the noncausal INAR(1) and noncausal RCINAR(1) processes, based on the queue introduced in Section 2. Section 6 concludes. Proofs and technical algorithms are gathered in Appendices.

2 A queuing system

In this section we extend the queuing system associated with the INAR(1) process [Pickands and Stine (1997); McCabe and Martin (2005); Schweer and Wichelhaus (2015)] to a more general framework. We temporarily leave the laws of various count variables unspecified, and focus on the deterministic relationships between these variables. We do not assume model (1) in this section.

2.1 A Lexis diagram

Imagine a bar with an infinite capacity.[4] It has existed since the infinite past and will keep indefinitely open. Time is discrete and $\mathbb{Z}-$valued. This double-sidedness will be essential later on when we study the reverse time property of the queuing system. In particular, the date $t = 0$ is not the origin of the bar, which is open since the infinite past. Moreover, the choice of this time origin is not important, as we will focus on stationary queuing systems.

When a customer arrives at date t, she is immediately served and counts as a customer. At the earliest she can leave the bar at date $t + 1$, and she will no longer be a customer at the date of departure. Hence, the duration spent by each customer[5] at the bar is at least equal to one. We are interested in the counts of customers with various arrival and/or departure dates, as well as the total number of customers at each date. To this end, for each couple $(s, t) \in \mathbb{Z}^2$ such that $s < t$, we first denote by $\eta_{s,t}$ the number of customers arriving at date s and leaving at date t. It is important to emphasize that this definition prioritizes neither the two dates s, t, nor the two time directions. Indeed, for a given s, the sequence $(\eta_{s,t})_t$ is indexed in calendar time t, which increases from $s + 1$ to $+\infty$; on the other hand, given t, the sequence $(\eta_{s,t})_s$ is indexed in reverse time, s decreasing from $t - 1$ to $-\infty$. Let us now report these count variables in the following Lexis diagram:

$$\begin{array}{cccccc}
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \eta_{t-1,t} & \eta_{t-1,t+1} & \eta_{t-1,t+2} & \eta_{t-1,t+3} & \cdots \\
 & 0 & \eta_{t,t+1} & \eta_{t,t+2} & \eta_{t,t+3} & \cdots \\
 & & 0 & \eta_{t+1,t+2} & \eta_{t+1,t+3} & \cdots \\
 & & & 0 & \cdots & \cdots
\end{array} \tag{3}$$

This is an infinite, upper triangular matrix, in which variable $\eta_{s,t}$ appears on the $s-$th row and the t-th column.[6] In particular, only terms strictly above the diagonal $s = t$ can be non zero due to the constraint $s < t$. In the next subsection we will see that, under reasonable assumptions, this matrix is sparse.

[4]We could also follow the queuing literature and use the terminology of a queue with an infinity of servers. However, we think that, compared to “queue", the terminology “bar" can better convey the idea that customers in the bar do not wait to be served, whereas in many stochastic queue models, customers are served sequentially by order of arrival.

[5]In the queuing literature, this duration is sometimes called service time.

[6]The direction of the s axis is southwards and the direction of the t axis is eastwards.

2.2 Analyzing the arrival and departure cohorts

The doubly indexed sequence $(\eta_{s,t})$ will now serve as the building block of other count sequences allowing to follow the movements of customers by either arrival, or departure cohort.

2.2.1 The arrival cohorts

For a given date s and each posterior date $t \geq s$, we denote by $\epsilon_s(t)$ the number of customers arriving at date s and staying at least until date t, with the convention that $\epsilon_s(s) = \epsilon_s$ is the number of customers arriving at date s regardless of their departure date. Here index s is prioritized, whereas integer t appears merely as an argument of the $s-$indexed process. We call index s the **arrival cohort** of these individuals, and ϵ_s the (initial) size of this cohort. Clearly, the sequence $\epsilon_s(t)$, where $t \geq s$, is nonincreasing in t, and we have:

$$0 \leq \eta_{s,t} = \epsilon_s(t-1) - \epsilon_s(t), \qquad \forall t \geq s+1. \tag{4}$$

Thus this sequence converges to a nonnegative integer limit, and $\eta_{s,t}$ is necessarily zero for large t. Hence, in each row of the Lexis diagram (3), there are at most a finite number of positive terms. Throughout the paper we assume that:

Assumption 1. For each given cohort s, sequence $\epsilon_s(t)$ is equal to zero for sufficiently large t.

In other words, there are no indefinitely staying customers. Then equation (4) can be equivalently rewritten into:

$$\epsilon_s(t) = \eta_{s,t+1} + \eta_{s,t+2} + \eta_{s,t+3} + \cdots, \qquad \forall t \geq s, \tag{5}$$

that is, those arriving at date s and staying until date t will ultimately leave at one of the following dates: $t+1$, $t+2$, ... In particular, taking $s = t$, the above equality yields:

$$\epsilon_s = \eta_{s,s+1} + \eta_{s,s+2} + \eta_{s,s+3} + \cdots. \tag{6}$$

Equations (5) and (6) are easy to interpret using the Lexis diagram (3). Equation (6) says that variables on the $s-$th row sum up to ϵ_s. Moreover, if we truncate this summation and keep only terms on the right hand side (RHS) of $\eta_{s,t}$ in the Lexis diagram, excluding $\eta_{s,t}$, we get (5) by equation $\epsilon_s(t)$.

2.2.2 The departure cohorts

Let us now consider the departure time t of the customers. More precisely, for a given date t and each $s \leq t$, we denote by $\tilde{\epsilon}_t(s)$ the number of customers leaving at date t and arriving before date s. This sequence is nondecreasing in s, and we have:

$$0 \leq \eta_{s,t} = \tilde{\epsilon}_t(s+1) - \tilde{\epsilon}_t(s), \qquad \forall s \leq t-1. \tag{7}$$

Thus each column of the Lexis diagram contains also a finite number of non zero terms and similarly as Assumption 1, we assume that:

Assumption 2. For each given t, sequence $\tilde{\epsilon}_s(t)$ goes to 0 when s goes back to $-\infty$.

In other words, no customers have been in the bar since the infinite past. Thus, for small s, both $\tilde{\epsilon}_t(s)$ (and hence also $\eta_s(t)$) are equal to 0 and we have:

$$\tilde{\epsilon}_t(s) = \eta_{s-1,t} + \eta_{s-2,t} + \eta_{s-3,t} + \cdots, \qquad \forall s \leq t. \tag{8}$$

In terms of the Lexis diagram, $\tilde{\epsilon}_t(s)$ is obtained by summing all the terms above $\eta_{s,t}$. In particular, when $s = t$, we get:

$$\tilde{\epsilon}_t := \tilde{\epsilon}_t(t) = \eta_{t-1,t} + \eta_{t-2,t} + \eta_{t-3,t} + \cdots, \tag{9}$$

which is the total number of customers leaving at date t, or the size of the **departure cohort** t. It is equal to the sum of all elements on the $t-$th column of the Lexis diagram.

2.2.3 An arrival/departure duality

Let us now relate definitions (7), (8), (9) to equations (4), (5), (6). Instead of observing customers' movements in calendar time (CT), alternatively we can observe them in reverse time (RT), that is by looking at the time-reversed video.[7] In other words, for any $t \in \mathbb{Z}$, the picture of the bar taken at CT date t is observed at RT date $-t$. Therefore:

[7] Such a duality has previously been considered in the (continuous time) queuing literature by Foster (1959).

- When a customer arrives at CT date $s \in \mathbb{Z}$, s is also the earliest CT date when she is at the bar. When the video is reversed, $-s$ becomes her last RT date at the bar, hence she leaves the bar at RT date $-s+1$;
- Similarly, when the customer leaves at CT date $t \in \mathbb{Z}$, the last CT date when she is at the bar is $t-1$; when the time is reversed, $-t+1$ becomes her first RT date at the bar. In other words she arrives at the bar at RT date $-t+1$.

To summarize, we have the following duality table:

	Arrival date	Departure date	Order of the two event dates
In calendar time	s	t	$s<t$
In reverse time	$-t+1$	$-s+1$	$-t+1<-s+1$

Table 1: Correspondence between calendar/reverse time and arrival/departure.

Reversing the time direction is equivalent to interchanging the arrivals and departures, taking the opposite then shifting[8] by +1 all the dates. As a consequence, counts defined in Section 2.2.2, which concern the different CT departure cohorts, correspond to RT arrival cohorts. This explains the similarity between equations (7), (8), (9) on the one hand, and equations (4), (5), (6) on the other hand.

2.3 Analyzing the current customer counts

It remains to count the current number of customers X_t, regardless of their arrival or departure cohort. We have:

$$X_t = \epsilon_t + \epsilon_{t-1}(t) + \epsilon_{t-2}(t) + \cdots, \tag{10}$$

by decomposing by their arrival cohort t, $t-1$, $t-2$.... In the Lexis diagram, ϵ_t is the sum of all the entries in the t-th row that are right to $\eta_{t,t+1}$ (including this latter count); $\epsilon_{t-1}(t)$ is the sum of all the entries in the $(t-1)$-th row that are right to $\eta_{t-1,t+1}$ (including this latter count), and so on. Thus we have:

$$X_t = \sum_{s \leq t} \sum_{t<\tau} \eta_{s,\tau}, \tag{11}$$

[8]This shift is due to the fact that when an individual arrives at date t, she is at the bar at date t. On the other hand, if she leaves at date t, then by definition she is no longer at the bar at date t.

where the double summation is with respect to both arrival time s with $s \leq t$, and departure time τ with $\tau > t$. In other words X_t is the sum of all the entries in the northeast of $\eta_{t,t+1}$, that is:

$$X_t = \sum_{(i,j) \in A_t} \eta_{i,j}, \qquad \text{where } A_t = \{(i,j) \in \mathbb{Z}^2, i \geq t, j \geq t+1\}.$$

Figure 2 below illustrates the region A_t in the Lexis diagram.

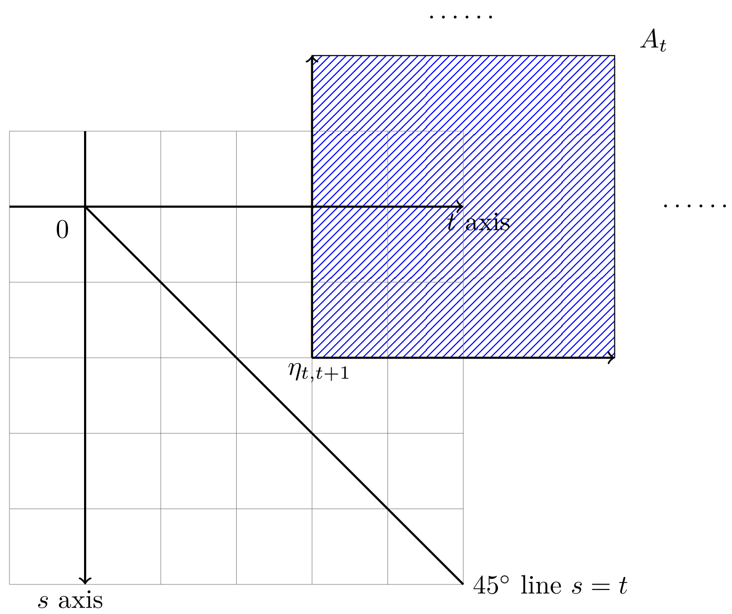

Figure 2: The infinite triangular area filled with northeast lines defines region A_t. Count X_t is the sum of all variables in the Lexis diagram on region A_t (borders included). When time t increases, this region moves jointly eastwards and southwards.

Similarly, by counting the current customers according to their departure dates, we get:

$$X_t = \tilde{\epsilon}_{t+1} + \tilde{\epsilon}_{t+1}(t+2) + \tilde{\epsilon}_{t+1}(t+3) + \cdots \tag{12}$$

The corresponding Lexis diagram interpretation is that the double summation in (11) can be alternatively conducted by exchanging the orders of summation, that is,

$$
\begin{aligned}
X_t &= \sum_{s \leq t} \sum_{t < \tau} \eta_{s,\tau} && (13)\\
&= \sum_{t < \tau} \sum_{s \leq t} \eta_{s,\tau} && (14)\\
&= \sum_{t < \tau} \tilde{\epsilon}_{\tau}(t+1)\\
&= \tilde{\epsilon}_{t+1} + \tilde{\epsilon}_{t+1}(t+2) + \tilde{\epsilon}_{t+1}(t+3) + \cdots , && (15)
\end{aligned}
$$

which is the RHS of equation (12). Thus, instead of first summing terms on the same rows [in (13)] and then aggregating these partial sums, we can alternatively first sum terms on the same columns [(14)].

2.4 Linking the departure and arrival population sizes

Finally, there also exists a relationship between the sizes of arrival cohorts (ϵ_t), those of departure cohorts ($\tilde{\epsilon}_t$), as well as (X_t):

$$\tilde{\epsilon}_{t+1} + X_{t+1} = X_t + \epsilon_{t+1}, \qquad (16)$$

that is, at date $t+1$, the counts of leaving customers, $\tilde{\epsilon}_{t+1}$, plus those currently in the bar, X_{t+1}, is equal to the previous customer count X_t plus the new arrivals ϵ_{t+1}. The interpretation in terms of the Lexis diagram is that the integer points covered by A_t and A_{t+1} differ in two ways [see Figure 1]. First, A_t covers all the points on the $(t+1)$-th column whereas A_{t+1} covers neither of them; Meanwhile, A_{t+1} covers all the points on the $(t+1)$-th row, but not A_t.

2.5 Summary

To summarize, we have introduced four count sequences $(\eta_{s,t})_{s,t}$, $(\epsilon_s(t))_{s,t}$, $(\tilde{\epsilon}_t(s))_{s,t}$, $(X_t)_t$. They are linked as follows:

- There is a one-to-one correspondence between $(\eta_{s,t})_{s,t}$ and $(\epsilon_s(t))_{s,t}$ [see eq. (4), (5)].
- There is a one-to-one correspondence between $(\eta_{s,t})_{s,t}$ and $(\tilde{\epsilon}_t(s))_{s,t}$ [see eq. (7), (8)].

- $(X_t)_t$ is a linear combination of either $(\epsilon_s(t))_{s,t}$, or $(\tilde{\epsilon}_t(s))_{s,t}$, or $(\eta_{s,t})_{s,t}$ [see eq. (10), (12)].
- The sizes of departure cohorts $(\tilde{\epsilon}_t)$, arrival cohorts (ϵ_t) and the count of current customers (X_t) are linked through (16). For a given (X_t), this link between departure cohort sizes $(\tilde{\epsilon}_t)$ and arrival cohort sizes (ϵ_t) is one-to-one.

In particular, as in model (1), it is impossible to recover the other underlying latent counts from the knowledge of $(X_t)_t$ alone.

3 Stochastic properties of the queuing system

The different counts in the queuing system, their interpretations and their interrelationships have been explored in Section 2. Let us from now on assume that only (X_t) is observed. This leads to another latent factor model for (X_t) that is *a priori* different from model (1), with a new set of unrecoverable latent counts $(\epsilon_s(t))$ [or equivalently $(\tilde{\epsilon}_t(s))$, or $(\eta_{s,t})$]. Nevertheless, under distributional assumptions on these latent factors, the queuing factor model will constrain the distribution of the observed process (X_t). In this section we characterize the distributions of the underlying latent counts $(\epsilon_s(t))$ leading to a dynamics for (X_t) that is compatible with the INAR(1) model.[9]

3.1 The Markov property

Since model (1) is Markov, we will focus on stochastic specifications for the latent $(\epsilon_s(t))$ under which the resulting process (X_t) defined by equation (10) is Markov, that is:

$$\ell(X_t|\underline{X_{t-1}}) = \ell(X_t|X_{t-1}). \tag{17}$$

Let us define $\epsilon_\bullet(t-1)$ as the information set of population sizes of various arrival cohorts at date $t-1$:

$$\epsilon_\bullet(t-1) = \{\epsilon_s(t-1), s = t-1, t-2, ...\}.$$

[9] By the arrival/departure duality, such an analysis can also be conducted for the specifications of the latent counts $(\tilde{\epsilon}_t(s))$. This is omitted due to space constraints.

The information set of all realized arrival counts up to time $t-1$ is: $\underline{\epsilon_\bullet(t-1)} := \{\epsilon_s(\tau), s \leq \tau \leq t-1\} = \{\eta_s(\tau), s \leq \tau \leq t-1\}$. These information sets are ordered by:

$$X_{t-1} \quad \subset \quad (\underline{X_{t-1}}, \epsilon_\bullet(t-1)) \quad \subset \quad \underline{\epsilon_\bullet(t-1)}.$$

Therefore, we have:

Lemma 1. *A sufficient condition for (X_t) to be Markov is that:*

$$\ell(X_t|\underline{\epsilon_\bullet(t-1)}) = \ell(X_t|X_{t-1}). \tag{18}$$

This lemma is intuitive and says that, if increasing the information set from X_{t-1} to $\underline{\epsilon_\bullet(t-1)}$ does not improve the prediction of X_t, then increasing to an intermediate information set $\underline{X_{t-1}}$ should not provide extra useful information either. Let us now look for conditions for equation (18) to hold. Since by equation (11), X_t is the sum of $\epsilon_s(t)$, s varying, it will be convenient to assume:

Assumption 3. For two different arrival cohorts $s_1 \neq s_2$, the processes of population sizes of different cohorts $(\epsilon_{s_1}(s_1+\tau))_{\tau\geq 0}$ and $(\epsilon_{s_2}(s_2+\tau))_{\tau\geq 0}$ are i.i.d.

Under this assumption, $\epsilon_s(t)$ is independent of, and has the same distribution as $\epsilon_{s+1}(t+1)$. Note, however, that within the same arrival cohort s, the process $(\epsilon_s(t))_{t>s}$ is generically serially dependent and nonstationary (since this sequence is nonincreasing and goes to zero at infinity). Next, we also assume that:

Assumption 4. Within each cohort s, the sequence of sizes $(\epsilon_s(t))_t$ is itself Markov, that is,

$$\ell_{s,t}\Big(\epsilon_s(t)|\epsilon_s(t-1), \epsilon_s(t-2), \cdots\Big) = \ell_{s,t}\Big(\epsilon_s(t)|\epsilon_s(t-1)\Big), \qquad \forall s, t. \tag{19}$$

In other words the future size of an arrival cohort s at time t depends only on its current size, but not the previous oes. We use the double subscript s,t in $\ell_{s,t}(\cdot|\cdots)$ since by definition the conditional distribution of $\epsilon_s(t)$ given $\epsilon_s(t-1)$ has the support $\{0, 1, ..., \epsilon_s(t-1)\}$, and is thus time-inhomogeneous. Moreover, by Assumption 3, $\ell_{s,t}(\cdot|\epsilon_s(t-1))$ as a measurable function depends on s,t and $\epsilon_s(t-1)$ only through $\epsilon_s(t-1)$ and the time lag $t-s$.

When all the $\epsilon_s(t)$ have finite mean[10], we also have, by Assumption 3:

$$\mathbb{E}[X_t] = \sum_{s=-\infty}^{t} \mathbb{E}[\epsilon_s(t)] = \sum_{\tau=0}^{\infty} \mathbb{E}[\epsilon_t(t+\tau)]. \tag{20}$$

Thus in order for X_t to be almost surely finite, not only the population of a cohort $(\epsilon_t(t+\tau))_\tau$ should decrease to zero when the “age" of the population τ goes to infinity (that is Assumption 1), but the convergence speed should be sufficiently fast.

Example 1 (A conditionally binomial specification)**.** Let us now give an example of specification compatible with condition (20). If there exists a constant probability parameter α such that for each $s < t$, the conditional distribution of $\epsilon_s(t)$ given $\epsilon_s(t-1)$ is binomial $\mathcal{B}(\epsilon_s(t-1), \alpha)$, then we have $\mathbb{E}[\epsilon_s(t)|\epsilon_s(t-1)] = \alpha\epsilon_s(t-1)$, and by iteration $\mathbb{E}[\epsilon_s(t)] = \alpha^{t-s}\mathbb{E}[\epsilon_t]$, hence equation (20) is satisfied. It will be shown in the next lemma that this specification is the only specification of the queuing system satisfying Assumptions 3, 4, as well as equation (18).

Lemma 2. *If the queuing system is such that both Assumptions 3, 4 and equation* (18) *are satisfied, then, for any $s < t$, the conditional distribution of $\epsilon_s(t)$ given $\epsilon_s(t-1)$ is binomial $\mathcal{B}(\epsilon_s(t-1), \alpha)$.*

Proof. See Appendix A.1. □

3.2 A queuing interpretation of the INAR(1) process

The following proposition is a direct consequence of the above Lemma 2. It says that in the above Markov queuing system, the dynamics of the customer count process (X_t) is indeed INAR(1).

Proposition 1 (Arrival cohort-disaggregated representation)**.** *The solution to the causal INAR(1) model* (1) *has the disaggregate queuing representation* (10)*, that is:*

$$X_t = \epsilon_t + \epsilon_{t-1}(t) + \cdots,$$

where the joint distribution of $(\epsilon_s(t))$ is as follows:

[10]This finite mean assumption is made for expository purpose. In the following sections we will see examples of stationary INAR(1) processes with $\mathbb{E}[\epsilon_t] = \infty$.

1. *For a given cohort s, sequence $(\epsilon_s(t))$ is Markov and the conditional distribution of $\epsilon_s(t+1)$ given $\epsilon_s(t)$ is binomial with a fixed parameter $\alpha \in [0,1[$. That is,*

$$\epsilon_s(t+1) = \sum_{i=1}^{\epsilon_s(t)} Z_{i,s,t+1}, \qquad \forall t \geq s, \tag{21}$$

where $Z_{i,s,t+1}$ are i.i.d. Bernoulli with parameter α and are independent of $\epsilon_s(t), \epsilon_s(t-1), \cdots$

2. *The sequence (ϵ_s) is i.i.d., and sequences $(\epsilon_s(t))_{t\geq s}$ are independent for different s.*

Proof. See Appendix A.2. □

Let us recall that the causal INAR(1) model (1) satisfies: $\mathbb{E}[X_t|X_{t-1}] = \alpha X_{t-1} + \mathbb{E}[\epsilon_t]$, provided that $\mathbb{E}[\epsilon_t]$ is finite. This is exactly the conditional expectation formula of a linear AR(1) model, which also has an equivalent linear MA(∞) interpretation. Formula (10) can be viewed as the (nonlinear) MA(∞) representation of X_t. since for fixed t, the terms $\epsilon_s(t)$ are independent across s and conditionally follow binomial distribution with probability α^{t-s} and size parameter ϵ_t given ϵ_s. In particular we have: $\mathbb{E}[\epsilon_s(t)] = \alpha^{t-s}\mathbb{E}[\epsilon_t]$, so long as $\mathbb{E}[\epsilon_t]$ is finite.

We can also check that the collection of $Z_{i,s,t+1}$'s in equation (21), where i and s vary, is equal to the collection of $Z_{i,t+1}$'s in equation (1). Indeed, in both cases these variables represent the binary decisions of X_t individuals of whether or not to stay at time $t+1$, depending on whether the X_t individuals are disaggregated by arrival cohort.[11]

Remark 1. The above representation shares similarities with the integer-valued moving average (INMA), or INARMA models [see e.g. Al-Osh and Alzaid (1988); Jørgensen and Song (1998); Enciso-Mora et al. (2009); Brännäs and Quoreshi (2010) for discrete time processes and Wheatley et al. (2018) for a continuous time analog]. For instance, an INMA(q) process, where q is an integer, has the representation:

$$X_t = \epsilon_t + \sum_{i=1}^{q} \sum_{j=1}^{\epsilon_{t-i}} Z_{i,j,t}, \tag{22}$$

[11] Note that the same index in equations (21) and (1) does not necessarily correspond to the same "individual". Indeed, the underlying queuing system of the INAR(1) process assumes that customers of the same cohort are exchangeable.

where $Z_{i,j,t}$ are mutually independent when i, j, t vary, Bernoulli distributed with parameter $\beta_i \in [0, 1]$, and are independent of $X_{t-1}, X_{t-2}, ...$

The main difference between representations (11) and (22) is that, in the former case, although each component $\epsilon_{t-i}(t)$ has the same distribution as $\alpha^i \circ \epsilon_{t-i}$, these terms are *dependent* within the same cohort, that is when both t and i vary while keeping $t - i$ constant.

The causal INAR(1) process is specified through the joint distribution of the cohort-by-cohort population sizes $(\epsilon_s(t))_{s,t}$. Then the distribution of departure counts $(\tilde{\epsilon}_t(s))_{s,t}$ is implicitly derived through $(\eta_{s,t})$. What properties possesses the departure process? Can they also be i.i.d.? The following proposition says that this is generically not the case:

Proposition 2 (The dynamics of the departure cohort size)**.** *Let us consider the INAR(1) process* (1) *along with its corresponding queue defined in Proposition 1. Then the arrival process* (ϵ_t) *and the departure process* $(\tilde{\epsilon}_t)$ *are both i.i.d., if and only if* ϵ_t *(resp.* $\tilde{\epsilon}_t$*) is Poisson* $\mathcal{P}(\lambda)$ *distributed. Moreover, in this case,* $(\tilde{\epsilon}_t)$ *is also i.i.d. and Poisson* $\mathcal{P}(\lambda)$ *distributed.*

This result complements Thm 1 in Schweer (2015), which says that process (X_t) is time reversible if and only if variable ϵ_t is Poisson $\mathcal{P}(\lambda)$ distributed.[12] The "if" part of Proposition 2 is easily illustrated using the Lexis diagram. If ϵ_t is $\mathcal{P}(\lambda)$ distributed, then by the property of the Poisson distribution, variables $\eta_t(t+1)$, $\eta_t(t+2)$,... in equation (6) are independent and Poisson distributed with parameters $\alpha\lambda$, $\alpha^2\lambda$,...respectively. Thus $\tilde{\epsilon}_t$ defined in (9) is still Poisson.

Proof. See Appendix A.3. □

3.3 A queuing interpretation of the RCINAR(1) process

The queuing system of the RCINAR(1) process is very similar to that of the INAR(1) process, except that Assumption 3 is replaced by the assumption that the population counts $\epsilon_s(t)$, s varying are conditionally independent given $(\epsilon_{s,t-1})_s$ and a time-varying factor α_t. More precisely, we have the following result:

Proposition 3 (A queuing interpretation of the RCINAR(1) model)**.** *The solution to the causal, RCINAR(1) model has the disaggregate queuing representation* (10)*, where the joint distribution of* $(\epsilon_s(t))$ *is as follows:*

[12]Schweer's result focuses on time reversibility whereas ours concerns the independence of the errors.

1. *For a given s, sequence* $(\epsilon_s(t))$ *is Markov and the conditional distribution of* $\epsilon_s(t+1)$ *given* $\epsilon_s(t)$ *is conditionally binomial given the time-varying parameter* $\alpha_t \in [0,1[$*, and* α_t *is i.i.d.*

2. *For any* $s_1 \neq s_2$*, variables* $\epsilon_{s_1,t+1}$ *and* $\epsilon_{s_2,t+1}$ *are independent given* $\epsilon_{s_1,t}$*,* $\epsilon_{s_2,t}$ *and* α_t*.*

3. *The sequence* $(\epsilon_s)_s$ *is i.i.d.*

The proof is obvious and omitted.

3.4 Markov affine count process

In the above subsections we have focused on INAR(1) and RCINAR(1) processes, defined by recursive equation (1) with i.i.d. Bernoulli variables $Z_{j,t+1}$'s. What happens in the affine case, where the $Z_{j,t+1}$'s are i.i.d., but with a possibly non-binomial count distribution? In this case process (X_t) generically no longer has the queuing interpretation. Indeed, we can still analogously define the sequence $\epsilon_s(t)$ through the recursion $\epsilon_s(t) = \sum_{j=1}^{\epsilon_s(t-1)} Z_{j,s,t}$, for a fixed s, and disaggregate X_t into:

$$X_t = \epsilon_t + \epsilon_{t-1}(t) + \epsilon_{t-2}(t) + \cdots ,$$

which is the MA(∞) representation of process (X_t). Nevertheless, sequence $(\epsilon_s(t))_t$ is no longer nonincreasing, since the $Z_{j,s,t}$ can be larger than 1. Thus we can define neither the count sequence $\eta_s(t)$ through equation (4), nor the departure counts $(\tilde{\epsilon}_t)$.

4 Time reversing a Markov count process

Let us now analyze irreversible noncausal processes and their calendar time dynamics. This provides new families of Markov dynamics for count process, that are easy to analyze under closed-form. In particular we consider the noncausal geometric and the noncausal discrete stable INAR(1) processes. We also explain why some of these dynamics are appropriate to capture bubble phenomena.

4.1 The definition

Let us now give the definition of a noncausal count process.

Definition 1. *The process (X_t) is noncausal affine, if it has the representation:*

$$\forall t, \qquad X_t = \sum_{i=1}^{X_{t+1}} Z_{t,i} + \tilde{\epsilon}_{t+1}, \tag{23}$$

where[13] *$\tilde{\epsilon}_{t+1}$ is independent of $\overline{X_{t+1}} = \{X_{t+1}, X_{t+2}, ...\}$ and is i.i.d. across t; variables $Z_{t,i}$'s are i.i.d. when t varies, and are independent of $\overline{X_{t+1}}$ and $\tilde{\epsilon}_{t+1}$. In particular, if $Z_{t,i}$'s are Bernoulli distributed, we say that (X_t) is noncausal INAR(1).*

Alternatively, if $Z_{t,i}$'s are only i.i.d. and Bernoulli conditional on a commun probability parameter α_t which is itself i.i.d., then process (X_t) is called noncausal RCINAR(1).

4.2 Time reversibility

The noncausal process (23) can be interpreted as a count process observed in reverse time. Let us derive some general properties of such processes. First, we will see that the Markov and stationarity properties are usually preserved by time revrsal.

Lemma 3 (See Section 1 of Cambanis and Fakhre-Zakeri (1995))**.** *If process (X_t) is stationary and Markov in calendar time, then it is also stationary and Markov in reverse time.*

Thus the noncausal INAR(1) process is also Markov. But what is the form of its calendar time conditional distribution of X_{t+1} given X_t? For instance, if the reverse time dynamics is affine (including INAR(1)), could the calendar time dynamics be also affine, that is, can process (X_t) have both a representation of type (1) and another representation of type (23)? To answer this question we have to first discard the obvious case of time reversible processes for which the answer is affirmative. By Thm 1 in Schweer and Wichelhaus (2015), the only time reversible INAR(1) process is Poisson-INAR(1). By Darolles et al. (2006), Section 5, the only non-INAR(1) Markov affine process that is reversible is the NBAR(1) process formally introduced by Gouriéroux and Lu (2019a). Then we have the answer to the above question:

Lemma 4 (See Proposition 3 of Gouriéroux and Lu (2019b))**.** *If the noncausal affine process (23) is also affine in calendar time, and is weakly ergodic in both time directions, then it is time*

[13] Note that we could have indexed the error term $\tilde{\epsilon}_t$ instead of $\tilde{\epsilon}_{t+1}$. However we prefer the second indexation, since it will be become clear later that this error term has the interpretation of the number of customers leaving at date $t+1$.

reversible.

Thus, except the two reversible cases, model (23) leads to a genuinely new, non-affine Markov dynamics for count processes. Let us now show that such dynamics has quite unique properties that distinguish them from the standard causal count processes of the literature.

Compared to the INAR(1) family, the RCINAR(1) family is substantially larger, but less tractable[14]. Nevertheless, we have the following result concerning the reversibility:

Proposition 4. *If the sequence (α_t) is i.i.d. and independent of (ϵ_t), then the RCINAR(1) process is time reversible if and only if (α_t) is beta $\mathcal{B}(p, q)$ distributed, for some $p, q \in (0, \infty)$, and $(\tilde{\epsilon}_t)$ is negative binomial distributed $NB(\theta, q)$, that is, the degree of freedom parameter is equal to q, with p.g.f. $\mathbb{E}[u^{\tilde{\epsilon}_t}] = \frac{(1-\theta)^q}{(1-\theta u)^q}$.*

Proof. See Appendix A.4. □

When $(\tilde{\epsilon}_t)$ is $NB(\theta, q)$, we can check that the marginal distribution of (X_t) is $NB(\theta, p+q)$. In this case we can easily check that the departure process ϵ_t is also i.i.d. $NB(\theta, q)$. This process is first introduced by Joe (1996) and applied to real data by Silva et al. (2019). It is non affine, and is hence different from the NBAR(1) process introduced in Gouriéroux and Lu (2019a). Nevertheless, if we keep the ratio of $\frac{p}{p+q} = \alpha_0$ fixed and let $p+q$ go to infinity, while at the same time keeping the product $q\theta = \lambda_0$ constant, then in the limiting case the beta (resp.) distribution reduces to the Dirac mass at α_0 (resp. λ_0). Hence we recover the Poisson-INAR(1) process as the limiting process.

4.3 Calendar time dynamics of a noncausal INAR(1)

From now on we will focus on noncausal INAR(1) processes for which (ϵ_t) is not Poissonian, i.e., (X_t) is irreversible. Since in practice X_{t+1} is only observed after X_t, the noncausal representation (23) is not directly usable for the forecasting of X_{t+1} given observation of X_t. Let us now derive the corresponding calendar time one-step-ahead conditional p.m.f. $\ell(X_{t+1}|X_t)$. By the Bayes

[14] Indeed, these processes are non-affine, and the (joint or conditional) Laplace transforms of the process is generically complicated.

formula we have:

$$\ell(X_{t+1}|X_t) = \frac{\ell(X_{t+1}, X_t)}{\ell(X_t)} = \frac{\ell(X_{t+1})}{\ell(X_t)}\ell(X_t|X_{t+1}), \tag{24}$$

where $\ell(X_t), \ell(X_{t+1})$ denote the stationary p.m.f. of the process (X_t) evaluated at X_t and X_{t+1}, respectively, and $\ell(X_t|X_{t+1})$ is, by simple convolution:

$$\ell(X_t|X_{t+1}) = \sum_{i=0}^{\min(X_t, X_{t+1})} \binom{X_{t+1}}{i} \alpha^i (1-\alpha)^{X_{t+1}-i} \mathbb{P}[\tilde{\epsilon}_{t+1} = X_t - i]. \tag{25}$$

Thus the predictive distribution $\ell(X_{t+1}|X_t)$ has closed form, whenever the p.m.f.'s of both $(\tilde{\epsilon}_t)$ and (X_t) do. In subsections 4.4-4.6 below, we check that these conditions are satisfied by a large range of noncausal INAR(1) models.

4.4 Noncausal geometric INAR(1)

We first consider the noncausal analog of geometric INAR(1) process introduced by McKenzie (1986). This process has a geometric stationary distribution, that is: $\mathbb{E}[u^{X_t}] = \frac{1}{1+\beta(1-u)}$, and the p.g.f. of $\tilde{\epsilon}_t$ is given by:

$$\mathbb{E}[u^{\tilde{\epsilon}_t}] = \frac{\mathbb{E}[u^{X_{t-1}}]}{\mathbb{E}[(\alpha u + 1 - \alpha)^{X_t}]} = \frac{1 + \beta\alpha(1-u)}{1+\beta(1-u)} = (1-\alpha)\frac{1}{1+\beta(1-u)} + \alpha, \qquad \forall u \in [0, 1].$$

In other words, $\tilde{\epsilon}_t$ has a two-component mixture distribution. The first component has weight $\alpha \in]0, 1[$, and is a point mass at zero. The other component is the geometric distribution with parameter $p = \frac{1}{1+\beta}$, with $\beta \in]0, \infty[$. Thus this is a zero-inflated geometric distribution and we can deduce the p.m.f. of $\tilde{\epsilon}_t$:

$$\mathbb{P}[\tilde{\epsilon}_t = n] = \frac{\beta^n}{(1+\beta)^{n+1}}(1-\alpha) + \alpha \mathbb{1}_{n=0}, \qquad \forall n \in \mathbb{N}. \tag{26}$$

Figure 3 plots the path of a noncausal geometric INAR(1). The parameter values are set equal to $\alpha = 0.85$ and $\beta = 3$. For this process, the theoretical marginal expectation is $\mathbb{E}[X_t] = \frac{p}{1-p} = 2.43$.

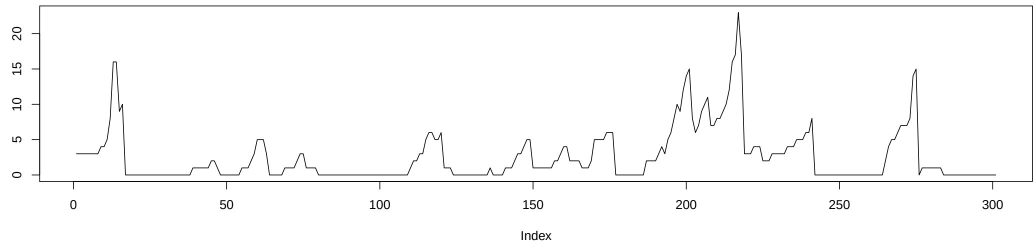

Figure 3: Path of a noncausal geometric INAR(1) process. We can see severl bubble periods with abrupt burst of the bubble.

The reverse time conditional p.m.f. is:

$$\ell(X_t|X_{t+1}) = \alpha \mathbb{1}_{X_t \leq X_{t+1}} \binom{X_{t+1}}{X_t} \alpha^{X_t} (1-\alpha)^{X_{t+1}-X_t} \\ + (1-\alpha) \sum_{n=0}^{\min(X_t, X_{t+1})} \binom{X_{t+1}}{n} \alpha^n (1-\alpha)^{X_{t+1}-n} \frac{\beta^{X_t-n}}{(\beta+1)^{X_t-n+1}}.$$

Thus the predictive p.m.f. has closed form:

$$\ell(X_{t+1}|X_t) = \frac{\ell(X_t|X_{t+1})\ell(X_{t+1})}{\ell(X_t)} = \ell(X_t|X_{t+1}) \frac{\beta^{X_{t+1}-X_t}}{(\beta+1)^{X_{t+1}-X_t}}.$$

Figure 4 plots the conditional p.m.f. $\mathbb{P}[X_{t+1} = \cdot | X_t = i]$ for the above noncausal INAR(1) process, for two different values of $i \in \{5, 10\}$. As a comparison, we also plot $\mathbb{P}[X_t = \cdot | X_{t+1} = i]$, which is equal to the conditional p.m.f. of the corresponding causal INAR(1) process.

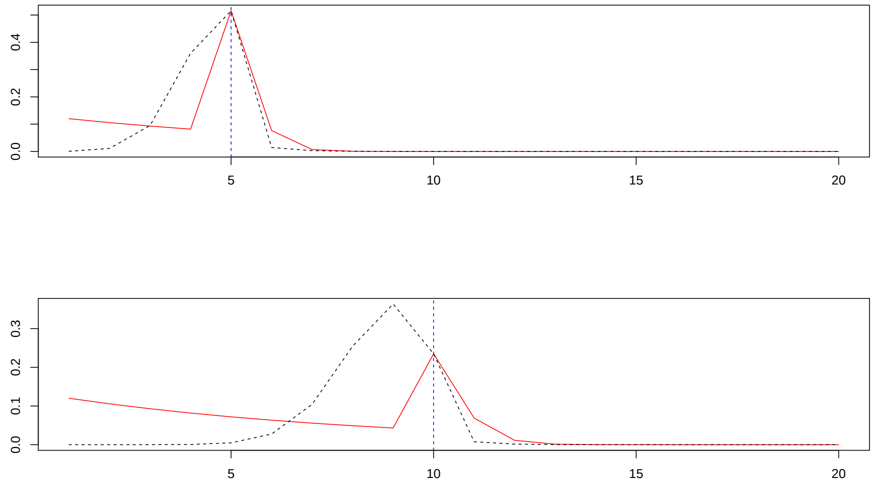

Figure 4: Comparison between the conditional p.m.f. in a noncausal and causal geometric INAR(1) process. In the upper panel we set $i = 5$ and in the lower panel $i = 10$. In both figures, the red full (resp. black dashed) line represents the curve of the noncausal (resp. causal) conditional p.m.f.. In both figures we have also added a vertical blue line crossing the x axis at i.

In both panels, the two conditional p.m.f.'s coincide at argument $x = i$. This is a direct consequence of the Bayes formula (25) and of the stationarity of the process. Moreover, the noncausal conditional p.m.f. assigns more weights immediately right to i, as well as towards zero. In other words, under the noncausal model, process (X_t) has a larger probability of growing larger and larger, that is the accumulation of the bubble. On the other hand it also has a larger probability of hitting zero, which corresponds to the collapse of the bubble. Figure 5 displays the corresponding conditional expectation in calendar time ($\mathbb{E}[X_{t+1}|X_t = i]$) and reverse time ($\mathbb{E}[X_t|X_{t+1} = i]$), as a function of i.

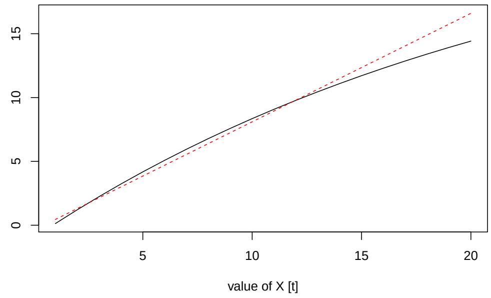

Figure 5: Comparison between the conditional expectation in a noncausal and causal discrete stable INAR(1) process. Black full line: noncausal model, red dashed line: causal model.

The reverse time conditional expectation $\mathbb{E}[X_t|X_{t+1} = i] = \alpha i + \mathbb{E}[\epsilon_t]$ is affine in i, but the affine property no longer holds for the causal expectation $\mathbb{E}[X_{t+1}|X_t = i]$.[15] This is expected since process (X_t) is irreversible. Moreover, the direct time expectation function is larger than the reverse time expectation function, when the value of i is neither too large, nor too small. We can also remark that the difference between the two conditional p.m.f.'s is much more important than that between the two conditional expectations, suggesting the limit of the conditional expectation in summarizing the nonlinear dynamics.

Finally, the causal, (marginally) geometric INAR(1) process has an obvious extension, that is the (marginally) negative binomial INAR(1) process of McKenzie (1987). Thus we can also extend our model to noncausal INAR(1) with negative binomial marginal distribution. This is straightforward and is omitted.

4.5 Noncausal discrete stable INAR(1)

In this subsection we consider another noncausal INAR(1) model, in which the distribution of $(\tilde{\epsilon}_t)$ is discrete stable $\mathcal{DS}(\beta, \nu)$ [see e.g. Steutel and Van Harn (1979); Christoph and Schreiber (1998)][16], with p.g.f.:

$$\mathbb{E}[u^{\tilde{\epsilon}_t}] = e^{-\beta(1-u)^\nu}, \qquad \forall u \in [0, 1],$$

[15]However, the noncausal INAR(1) and its causal counterpart share the same autocorrelation function since $\text{corr}(X_t, X_{t+1}) = \text{corr}(X_{t+1}, X_t)$.

[16]Despite their similar terminology, the discrete stable distributions do not belong to the family of α-stable distributions, which are continuously valued. However, Steutel and Van Harn (1979) interpret them as the discrete analog of the $\alpha-$stable distributions.

where scale parameter $\beta > 0$, and shape parameter $\nu \in]0, 1[$.[17] The corresponding p.m.f. is:

$$\mathbb{P}[\tilde{\epsilon}_t = k] = (-1)^k \sum_{j=0}^{\infty} \binom{j\nu}{k} \frac{(-\beta)^j}{j!}, \qquad \forall k \in \mathbb{N}. \tag{27}$$

The specificity of new noncausal process is that the mean $\mathbb{E}[\tilde{\epsilon}_t]$ is infinite [see Christoph and Schreiber (1998)], and is thus suitable for heavy-tailed count process data. Figure 6 plots the trajectory of a noncausal discrete stable INAR(1) process with $\nu = 0.5$, $\beta = 0.05$, and $\alpha = 0.5$.

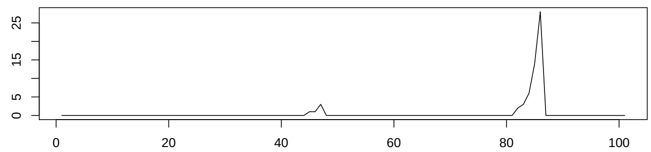

Figure 6: Simulated path of a noncausal discrete stable INAR(1) process. We can see two bubble periods.

Let us first derive the stationary distribution of this process by considering its p.g.f. We have:

$$\mathbb{E}[u^{X_t}] = \mathbb{E}[u^{\tilde{\epsilon}_t}]\mathbb{E}[(\alpha u + 1 - \alpha)^{X_{t+1}}] = \mathbb{E}[u^{\tilde{\epsilon}_t}]\mathbb{E}(\alpha u + 1 - \alpha)^{X_t}].$$

By iterating we get:

$$\mathbb{E}[u^{X_t}] = \prod_{h=0}^{\infty} e^{-\beta(1-\alpha^i u-(1-\alpha^i))^\nu} = \prod_{h=0}^{\infty} e^{-\beta\alpha^{\nu i}(1-u)^\nu} = e^{-\beta\frac{(1-u)^\nu}{1-\alpha^\nu}},$$

which corresponds to the discrete stable distribution $\mathcal{DS}(\frac{\beta}{1-\alpha^\gamma}, \nu)$. Thus by eq. (27), the marginal p.m.f. is:

$$\mathbb{P}[X_t = k] = (-1)^k \sum_{j=0}^{\infty} \binom{j\nu}{k} \frac{\big[-\beta/(1-\alpha^\nu)\big]^j}{j!}, \qquad \forall k \in \mathbb{N}. \tag{28}$$

Let us now plot the analog of Figure 4. Using equations (25), (27) and (28), we can compute the conditional p.m.f. $\ell(X_{t+1}|X_t = i)$. Since a discrete stable distribution is heavy-tailed, we

[17] In the limiting case where $\nu = 1$, we recover the Poisson distribution $\mathcal{P}(\beta)$; thus this case is omitted.

will consider two larger values for the conditioning observation $i = 10$, or 25. These conditional p.m.f.'s are plotted in Figure 6.

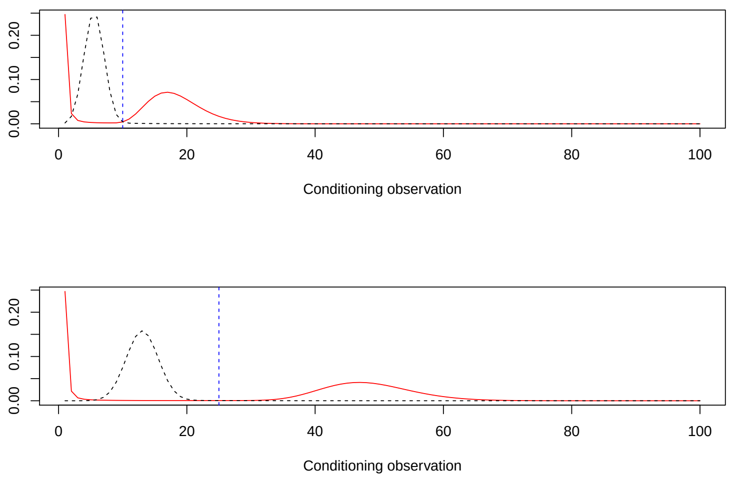

Figure 7: Comparison between the direct time conditional p.m.f. $\mathbb{P}[X_{t+1} = \cdot | X_t = i]$ (red full line) and the reverse time counterpart, i.e. $\mathbb{P}[X_t = \cdot | X_{t+1} = i]$ (black dashed line) in a noncausal discrete stable INAR(1) process. Upper panel: the conditioning observation is $i = 10$; lower panel: the conditioning observation is $i = 25$.

The comparison between the causal and noncausal conditional p.m.f.'s is quite similar as that for Figure 4. In particular, they have respectively two and one local modes. Indeed, for a large value of X_t, the causal conditional distribution assigns a negligible probability to the conditional probability $\mathbb{P}[X_{t+1} > X_t]$, whereas under the noncausal model, this probability is quite significant.

Since the discrete stable distribution has an infinite mean, the noncausal conditional expectation $\mathbb{E}[X_t | X_{t+1}]$ does not exist. However, the calendar time dynamics has thin tail as shown in the next proposition:

Proposition 5. *For fixed X_t, the conditional p.m.f. is asymptotically equal to:*

$$\ell(X_{t+h}|X_t) = O\Big(\frac{(1-\alpha^h)^{X_{t+h}}}{X_{t+h}^{\nu+1-X_t}}\Big), \qquad \text{when } X_{t+h} \text{ goes to infinity,}$$

and the conditional moment $\mathbb{E}[X_{t+h}^p | X_t]$ is finite for any power $p > 0$ and any horizon h.

Proof. See Appendix A.5. □

The finiteness of $\mathbb{E}[X_{t+h}|X_t]$ does not contradict the non existence of $\mathbb{E}[X_t]$. Roughly speaking, the causal prediction $\mathbb{E}[X_{t+h}|X_t]$ is finite for any fixed horizon h, but increases to infinity when h goes to infinity.

Figure 8 plots $\mathbb{E}[X_{t+1}|X_t = i]$ as a function of i. Their values are computed numerically from those of the conditional p.m.f., by truncating the identity:

$$\mathbb{E}[X_{t+1}|X_t = i] = \sum_{n=0}^{\infty} n \Big\{ \mathbb{P}[X_{t+1} = n | X_t = i] \Big\},$$

at a large, finite order. On the contrary to Figure 4, we have not plotted the corresponding reverse time conditional expectation, as this latter is infinite.

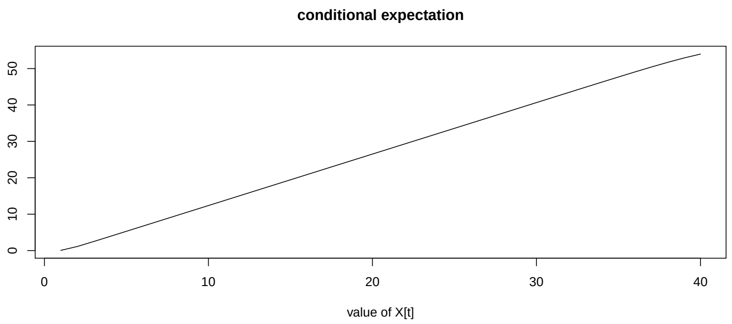

Figure 8: Noncausal conditional expectation

The conditional expectation seems to increase nearly linearly[18] in i, with a slope that is larger than 1. This impression is confirmed by Lemma 3, which is the discrete analog of a result for (continuously valued) noncausal linear $\alpha-$stable AR(1) processes [see Cambanis and Fakhre-Zakeri (1995); Gouriéroux and Zakoïan (2017)], for which the convergence in equation (29) is simply replaced by equality.

Proposition 6. *For the noncausal discrete stable INAR(1) process, the conditional expectation is such that:*

$$\frac{\mathbb{E}[X_{t+1}|X_t = i]}{i} \to \alpha^{\nu-1}, \tag{29}$$

[18] A closer numerical inspection reveals that the curve is not exactly linear.

when i goes to infinity.

Proof. See Appendix A.6. □

Thus for large values of X_t, the conditional expectation is linear with an "asymptotic" autoregressive coefficient equal to $\alpha^{\nu-1}$, which is *larger* than 1, since $\nu, \alpha \in [0,1[$. In particular this ratio is different from the autoregressive coefficient α in reverse time, which is constant and *smaller* than 1. As a consequence, conditional on X_t being large, X_{t+1} is in average even larger than X_t. Such a locally explosive feature cannot be replicated by a standard INAR(1) process, since in model (1) the autocorrelation coefficient is $\alpha < 1$ and hence the corresponding limit of $\frac{\mathbb{E}[X_{t+1}|X_t=i]}{i}$ is $\alpha < 1$.

As Figure 7 shows, the conditional distribution of X_{t+1} given X_t has two (local) modes, one near zero and the other larger than the current value X_t. A further investigation reveals that we have the following result:

Proposition 7. *For a given $\epsilon > 0$, when X_t goes to infinity,*

- *the probability $\mathbb{P}\Big[\frac{X_{t+1}}{X_t} \in (0, \epsilon) \mid X_t\Big]$ goes to $1 - \alpha^\nu$.*
- *the probability $\mathbb{P}\Big[\frac{X_{t+1}}{X_t} \in (\frac{1}{\alpha} - \epsilon, \frac{1}{\alpha} + \epsilon) \mid X_t\Big]$ goes to α^ν.*

Proof. See Appendix A.7. □

Thus when the process is currently in the bubble (i.e. X_t is large), the conditional distribution of the future scaled value $\frac{X_{t+1}}{X_t}$ given X_t is approximately binary. With a fixed probability α^ν, the bubble continues to grow at the geometric rate of $\frac{1}{\alpha}$; whereas with fixed probability $1 - \alpha^\nu$, the bubble collapses. This also echoes Proposition 6, which says that the conditional expectation $\mathbb{E}[X_{t+1}|X_t]$ is approximately equal to $\alpha^{\nu-1}X_{t-1}$, that is $\frac{X_t}{\alpha}$ times the probability α^ν.

Proposition 7 is again linked to the recent literature on (linear) noncausal AR(1) processes, which is the continuously valued analog of the noncausal discrete stable INAR(1). Fries (2018), Prop. 2.2, shows that the very same proposition holds for this continuous analog. Roughly speaking, this similarity is due to the fact that the joint distribution of (X_t, X_{t+1}) is within the domain of attraction of a certain bivariate $\alpha-$stable vector (Y_t, Y_{t+1}), where process (Y_t) is a noncausal $\alpha-$stable linear AR(1) process. Thus the joint distribution of (X_t, X_{t+1}) is asymptotically "close to" that of (Y_t, Y_{t+1}), for which the result of Fries (2018) applies.

The presence of more than one local modes also implies that standard forecasting tools for time series counts, such as the conditional expectation and the (single) conditional mode [see Freeland and McCabe (2004)] can be misleading. Hence the importance of the closed form conditional p.m.f. derived in this paper. The result can also be compared with the fact that the marginal distribution of a (causal or noncausal) INAR(1) process is necessarily uni-modal [see Steutel and Van Harn (1979)]. What we have seen is that a noncausal INAR(1) processes can have multi-modal conditional distribution, even though the the marginal distribution is uni-modal since infinitely divisible. This multi-modality cannot be uncovered by standard data visualization tools such as the empirical marginal histogram.

4.6 Other examples

Besides the examples given in Section 4.4 and 4.5, in general the stationary p.m.f. of a (causal or noncausal) affine process has no known parametric form. However, under mild assumptions, this p.m.f. remains computable through inexpensive, closed form matrix operations [see Lu (2018)]. In Appendix B, we briefly review and illustrate this method using two further examples of noncausal INAR(1) processes with respectively negative binomial and Poisson-Inverse Gaussian distributed innovations. See also Aly and Bouzar (1994) for examples of affine count processes with closed form stationary distributions.

5 A queuing interpretation of the noncausal count process

We now show that the noncausal INAR(1) can also be specified through the queuing system introduced in Section 2.

5.1 The (noncausal) specification of the queuing system for the noncausal INAR(1) process

Proposition 8. *The noncausal INAR(1) process has the disaggregate representation* (12)*, that is:*

$$X_t = \tilde{\epsilon}_{t+1} + \tilde{\epsilon}_{t+1}(t+2) + \cdots,$$

where the joint distribution of $\tilde{\epsilon}_t(s)$, $t \geq s$ is defined as follows:

1. *For a given t, the sequence $(\tilde{\epsilon}_t(s))_{s \leq t-1}$ is such that the terminal value $\tilde{\epsilon}_t(t)$ is equal to $\tilde{\epsilon}_t$, whereas previous values are defined backwardly and recursively by:*

$$\tilde{\epsilon}_t(s-1) = \sum_{i=1}^{\tilde{\epsilon}_t(s)} Z^*_{i,s,t}, \qquad \forall s \leq t-1, \tag{30}$$

   where $Z^*_{i,s,t}, i = 1, ...$ are i.i.d. Bernoulli $\mathcal{B}(\alpha)$, and are independent of $\tilde{\epsilon}_t(s)$.

2. *The departure counts $\tilde{\epsilon}_s$, $s \in \mathbb{Z}$, are i.i.d., and are independent of $Z^*_{i,s,t}$.*

Let us consider a statistician observing the reverse time video: she first observes the total number of departures $\tilde{\epsilon}_t$ of the entire departure cohort, then tries to “trace back" the time of arrival of these customers. Given that a customer is in the bar at a certain date $s < t$, she thinks that there is a probability of α (resp. $1-\alpha$) that the customer was (resp. was not) already at the bar at the previous date.

This proposition is the noncausal analog of Proposition 1, and equation (12) can be viewed as the MA(∞) representation of the noncausal process, that is the analog of (10). However, the two models differ fundamentally in terms of observability. Indeed, the variables on the RHS of (12) involve a departure date that is posterior to t; thus they are fundamentally unobservable at the present date t, even if all the movements of the queuing system are observed up to present.

5.2 The causal dynamics of the queuing system of the noncausal INAR(1) process

We now turn to the queuing system introduced in Section 2. In the causal INAR(1) model, customers count processes of $(\epsilon_s(t))_{t>s}$ are i.i.d. across different *arrival* cohorts s, and Proposition 2 says that the departure cohort size $\tilde{\epsilon}_t$ cannot be i.i.d., except for Poisson distributed (ϵ_t). By the arrival/departure duality, Proposition 2 also holds for the noncausal INAR(1) model. That is, since in the non-Poisson noncausal model, counts $\tilde{\epsilon}_t(s)$ are i.i.d., the resulting arrival counts ϵ_t cannot be i.i.d.[19] What else can be said about the arrival count process (ϵ_t) and the departure

[19] In the continuous time queuing literature, models with non i.i.d. arrival sequence have already been considered by various authors [see e.g. Fakinos (1984)]. But our model is fundamentally different from theirs in the sense that none of these previous models specify the queuing system in reverse time.

decisions of existing customers? Does the probability of leaving of each customer depend on the previous numbers of entries? To answer these questions, let us write:

$$X_{t+1} = (X_{t+1} - \epsilon_{t+1}) + \epsilon_{t+1},$$

where process (X_t) is noncausal INAR(1) and process (ϵ_t) is defined by (16). This representation is analogous to representation (1), where X_{t+1} is decomposed into the sum of the number of newcomers and the number of old customers who stay. By equation (16), $X_{t+1} - \epsilon_{t+1}$ is equal to $X_t - \tilde{\epsilon}_{t+1}$, which is nonnegative and non larger than X_t. Lemma 5 below summarizes the major differences of between the calendar time dynamics of a noncausal INAR(1) and the causal INAR(1) process:

Lemma 5. *For a noncausal INAR(1) process defined in Definition 2, the following statements are equivalent:*

i) *Process (X_t) is time reversible;*

ii) *Process (ϵ_t) defined by equation (16) is Poisson;*

iii) $X_{t+1} - \epsilon_{t+1}$ *is binomial given* X_t*;*

iv) ϵ_{t+1} *is independent of* X_t*;*

v) $X_{t+1} - \epsilon_{t+1}$ *and* ϵ_{t+1} *are independent given* X_t*.*

Proof. See Appendix A.8. □

As a reminder, for a causal INAR(1) model with arbitrarily distributed (ϵ_t), conditions *iii*), *iv*) and *v*) in Lemma 5 are all satisfied. On the other hand, this lemma says that in a non-Poisson, noncausal INAR(1) model, the calendar time dynamics of (X_t) is non thinning-based, the arrival size process (ϵ_{t+1}) is state-dependent[20] and is conditionally dependent of the number of staying old customers $X_{t+1} - \epsilon_{t+1}$. As a consequence, we have introduced a new, non-INAR family of dynamic count models.

[20]That is, ϵ_{t+1} depends on the previous customer count X_t.

5.3 The distribution of the arrival count

Although (ϵ_t) features serial dependence by Lemma 6, it has a time-invariant stationary distribution, with p.g.f.:

$$\begin{aligned}
\mathbb{E}[u^{\epsilon_{t+1}}] &= \mathbb{E}[u^{\tilde{\epsilon}_{t+1}+X_{t+1}-X_t}] && \text{(by equation (16))} \\
&= \mathbb{E}\Big\{u^{\sum_{i=1}^{X_{t+1}}(1-Z_{i,t})}\Big\} && \text{(by equation (23))} \\
&= \mathbb{E}\Big\{[(1-\alpha)u+\alpha]^{X_{t+1}}\Big\}, && (31)
\end{aligned}$$

by compounding the p.g.f. of the distribution of X_t and of the Bernoulli distribution $\mathcal{B}(1, 1-\alpha)$. Let us now derive the explicit formula for the two special models considered in Sections 4.4 and 4.5.

Example 2 (Noncausal geometric INAR(1) process, see Section 4.4)**.** If the noncausal INAR(1) process has the geometric marginal distribution, then $\mathbb{E}[u^{X_t}] = \frac{1}{1+\beta(1-u)}$, and equation (31) becomes:

$$\mathbb{E}[u^{\epsilon_t}] = \frac{1}{1+\beta[1-(1-\alpha)u-\alpha]} = \frac{1}{1+\beta(1-\alpha)(1-u)}.$$

Hence (ϵ_t) is geometric distributed and we have: $\mathbb{E}[\tilde{\epsilon}_t] = \beta(1-\alpha) = \mathbb{E}[\epsilon_t]$.

Example 3 (Noncausal discrete stable INAR(1) process, see Section 4.5)**.** Let us re-consider the noncausal discrete stable INAR(1) process. Since $\mathbb{E}[u^{X_t}] = e^{-\beta\frac{(1-u)^\nu}{1-\alpha^\nu}}$, equation (31) becomes:

$$\mathbb{E}[u^{\epsilon_t}] = e^{-\beta\frac{(1-\alpha)^\nu}{1-\alpha^\nu}(1-u)^\nu}.$$

Hence the distribution of (ϵ_t) is $\mathcal{DS}(\beta\frac{(1-\alpha)^\nu}{1-\alpha^\nu}, \nu)$. Since[21] $\frac{(1-\alpha)^\nu}{1-\alpha^\nu} \geq 1$, the stationary distribution of ϵ_t has a larger scale parameter than the distribution of $\tilde{\epsilon}_t$.

5.4 Extension to noncausal RCINAR(1) process

When the RCINAR(1) process is time irreversible, its reverse time dynamics is different from the calendar time dynamics. Then a noncausal RCINAR(1) is defined as a time-reversed RCINAR(1)

[21] In the limiting Poisson case, when $\nu = 1$, the inequality becomes an equality.

process, that is: Then we can give a similar queuing interpretation for the noncausal RCINAR(1), in a similar way as in Section 5.1 for noncausal INAR(1). Such an interpretation would be based on the stochastic specification of the sequence $(\tilde{\epsilon}_s(t))$ instead of $(\epsilon_s(t))$ and is omitted due to space constraint.

6 Conclusion

In this paper we have introduced the concept of noncausality for count processes. In particular we have studied in details the noncausal INAR(1) process and have shown that (resp. noncausal) INAR(1) process can be interpreted by an infinite server queue, in which the sizes of the arrival (resp. departure) cohorts are i.i.d. We have also seen that the resulting noncausal processes, which is generically time irreversible and non affine, can feature bubble-like dynamics characterized by some unique bi-modal conditional distribution. Moreover, the same analysis has been extended to both Markov affine processes and RCINAR(1) processes. Another important implication of the paper is that measures such as autocovariance should be used with care[22]. Indeed, these moments need not be finite, and even if this is the case, they do not characterize the serial dependence pattern. In particular, we have seen in the paper that a noncausal process and its causal counterpart share the same autocovariance function[23], but can have completely different nonlinear dynamics. Finally, the noncausal, Markov affine process is the simplest example of noncausal count process. In the paper we have briefly discussed the extension to RCINAR(1), but a large range of other extensions are also possible, including non-Markov, mixed causal-noncausal models. Such a process allow the bubble to burst gradually instead of abruptly, at a rate that is potentially different from the rate of explosion of the bubble[24]. These all await future research.

[22]See e.g. Livsey et al. (2018); Jia et al. (2018) for recent developments on count processes with flexible autocorrelation function.

[23]This is due to the fact that the covariance operator is symmetric in the two arguments.

[24]See e.g. Fries and Zakoian (2019) for a continuous analog.

Appendix

A Proofs

A1 Proof of Lemma 2

We have:

$$\mathbb{E}[u^{X_t}|\underline{\epsilon_\bullet(t-1)}] = \mathbb{E}[u^{\epsilon_t+\epsilon_{t-1}(t)+\epsilon_{t-2}(t)+\cdots}|\epsilon_\bullet(t-1)]$$
$$= \mathbb{E}[u^{\epsilon_t}] \prod_{s=-\infty}^{t-1} \mathbb{E}[u^{\epsilon_s(t)}|\epsilon_s(t-1)], \tag{a.1}$$

where the first term $\mathbb{E}[u^{\epsilon_t}]$ is factored out, since by Assumption 3, ϵ_t is independent of $\underline{\epsilon_\bullet(t-1)}$, whereas the conditional p.g.f. of $\epsilon_{t-1}(t)+\epsilon_{t-2}(t)+\cdots$ becomes the product of different conditional p.g.f. by the conditional independence (see Assumption 4). By Assumption 3, the s-th element in the infinite product is a measurable function of $\epsilon_s(t-1), u$ and $t-s$ only, and by equation (10), these conditioning variables should also sum up to X_{t-1}. Thus, in order for $\ell(X_t|\underline{\epsilon_\bullet(t-1)})$ to depend on $\underline{\epsilon_\bullet(t-1)}$ through X_{t-1} only for all possible values of $\underline{\epsilon_\bullet(t-1)}$, there should exist a constant function of u, $g_1(u)$, say, as well as a function of u and $t-s$, $g_2(u, t-s)$, say, such that:

$$\mathbb{E}[u^{\epsilon_s(t)}|\epsilon_s(t-1)] = [g_1(u)]^{\epsilon_s(t-1)} g_2(u, t-s),$$

for any $s \leq t-1$. That is, $\mathbb{E}[u^{\epsilon_s(t)}|\epsilon_s(t-1)]$ should be an exponential affine function of $\epsilon_s(t-1)$ with a slope that can only depend on u, but not on $t-s$. Since the LHS of the above equation is a conditional p.g.f., functions g_1 and g_2 are constrained. Indeed,

- First by taking the special case where $\epsilon_s(t-1) = 0$ (in this case $\epsilon_s(t)$ is almost surely zero as well), we get $g_2(u, t-s) = 1$ is constant.
- Then we take the special case where $\epsilon_s(t-1) = 1$, and can conclude that $g_1(u)$ is necessarily the p.g.f. of a count distribution. Thus given $\epsilon_s(t-1)$, the conditional distribution of $\epsilon_s(t)$ is the convolution of $\epsilon_s(t-1)$ identical count distributions with p.g.f. $g(\cdot)$. But since the distribution of $\epsilon_s(t)$ is bounded by $\epsilon_s(t-1)$, function $g_1(\cdot)$ must be the p.g.f. of a Bernoulli

distribution, say, $\mathcal{B}(1,\alpha)$, i.e. $g_1(u)=\alpha u+1-\alpha$.

As a consequence, the conditional distribution of $\epsilon_s(t)$ given $\epsilon_s(t-1)$ is binomial $\mathcal{B}(\epsilon_s(t-1),\alpha)$.

A2 Proof of Proposition 1

By Lemma 2, equation (a.1) becomes:

$$\mathbb{E}[u^{X_t}|X_{t-1}]=\mathbb{E}[u^{\epsilon_t}]\prod_{s\leq t-1}(\alpha u+1-\alpha)^{\epsilon_s(t-1)}=\mathbb{E}[u^{\epsilon_t}](\alpha u+1-\alpha)^{X_{t-1}},$$

which is exactly equation (2) for Bernoulli distributed $Z_{i,t+1}$'s. This completes the proof.

A3 Proof of Proposition 2

First, since (X_t) is stationary, by equation (16), process (ϵ_t) is stationary too. Thus it suffices to show, for instance, that $\tilde{\epsilon}_t$ and $\tilde{\epsilon}_{t+1}$ are independent. Let us compute their joint p.g.f.. We have:

$$\begin{aligned}
\mathbb{E}[u^{\tilde{\epsilon}_t}v^{\tilde{\epsilon}_{t+1}}] &= \mathbb{E}[u^{X_t+\epsilon_{t+1}-X_{t+1}}v^{X_{t+1}+\epsilon_{t+2}-X_{t+2}}] \\
&= \mathbb{E}\Big[\mathbb{E}[u^{X_t+\epsilon_{t+1}-X_{t+1}}v^{X_{t+1}+\epsilon_{t+2}-X_{t+2}}|X_{t+1},\epsilon_{t+2},X_t,\epsilon_{t+1}]\Big] \\
&= \mathbb{E}\Big[u^{X_t+\epsilon_{t+1}-X_{t+1}}v^{X_{t+1}}\mathbb{E}[v^{-X_{t+2}+\epsilon_{t+2}}|X_{t+1},\epsilon_{t+2},X_t,\epsilon_{t+1}]\Big] \\
&= \mathbb{E}\Big[u^{X_t+\epsilon_{t+1}-X_{t+1}}v^{X_{t+1}}(\frac{\alpha}{v}+1-\alpha)^{X_{t+1}}\Big] \\
&= \mathbb{E}\Big[u^{X_t+\epsilon_{t+1}}(\frac{\alpha+(1-\alpha)v}{u})^{X_{t+1}}\Big] \\
&= \mathbb{E}\Big[u^{X_t+\epsilon_{t+1}}\mathbb{E}[(\frac{\alpha+(1-\alpha)v}{u})^{X_{t+1}}|X_t,\epsilon_{t+1}]\Big] \\
&= \mathbb{E}\Big[u^{X_t+\epsilon_{t+1}}[(\frac{\alpha+(1-\alpha)v}{u})\alpha+1-\alpha]^{X_t}(\frac{\alpha+(1-\alpha)v}{u})^{\epsilon_{t+1}}\Big] \\
&= \mathbb{E}\Big[[\alpha^2+\alpha(1-\alpha)v+(1-\alpha)u]^{X_t}\Big]\mathbb{E}\Big[[\alpha+(1-\alpha)v]^{\epsilon_{t+1}}\Big].
\end{aligned}$$

Thus the RHS is multiplicatively separable if and only if the marginal p.g.f. function $g(u):=\mathbb{E}[u^{X_t}]$ satisfies:

$$g(x+y-1)=g(x)g(y)/g(1),$$

where $x = \alpha^2 + \alpha(1-\alpha)v + 1 - \alpha$ and $y = \alpha + (1-\alpha)u$. We can check that under suitable regularity conditions (such as the continuity of g), the only solution of this functional equation is $g(x+1) = e^{\lambda x}$, for some constant λ. Thus we deduce that $\tilde{\epsilon}(t)$ and $\tilde{\epsilon}(t+1)$ are independent if and only if $g(u) = e^{\lambda(u-1)}$, that is if (X_t) is Poisson $\mathcal{P}(\lambda)$ distributed. Then by stationarity, the p.g.f. of ϵ_t is:

$$\mathbb{E}[u^{\epsilon_t}] = \frac{\mathbb{E}[u^{X_t}]}{\mathbb{E}[(\alpha u + 1 - \alpha)^{X_t}]} = e^{(1-\alpha)\lambda(u-1)},$$

and ϵ_t is also Poisson $\mathcal{P}((1-\alpha)\lambda)$ distributed.

A4 Proof of Proposition 4

We follow the proof of Schweer (2015) for INAR(1) processes. The “if" part of the proposition is straightforward and we focus on the “only if" part. Suppose that process (X_t) is RCINAR(1). Then the conditional distribution of the process is:

$$p(i|j) := \mathbb{P}[X_{t+1} = i | X_t = j] = \sum_{k=0}^{\min(i,j)} \binom{j}{k} \ell_0(i-k) \mathbb{E}[\alpha_t^k (1-\alpha_t)^{i-k}], \qquad \forall i, j \in \mathbb{N},$$

where function ℓ_0 is the p.m.f. of ϵ_t. Under reversibility, we have:

$$\mathbb{P}[(X_t, X_{t+1}, X_{t+2}, X_{t+3}) = (0, 1, i, 0)] = \mathbb{P}[(X_t, X_{t+1}, X_{t+2}, X_{t+3}) = (0, i, 1, 0)], \qquad \forall i \in \mathbb{N}, \tag{a.2}$$

or equivalently[25],

$$p(1|0)p(i|1)p(0|i) = p(i|0)p(1|i)p(0|1). \tag{a.3}$$

[25]We can check that $\mathbb{P}[\epsilon_t = 0] > 0$.

Equation (a.2) leads to:

$$
\begin{aligned}
&p(i|0) = \ell_0(i),\\
&p(0|i) = \ell_0(0)\mathbb{E}[(1-\alpha_t)^i],\\
&p(i|1) = \ell_0(i)\mathbb{E}[(1-\alpha_t)] + \ell_0(i-1)\mathbb{E}[\alpha_t],\\
&p(1|i) = \ell_0(1)\mathbb{E}[(1-\alpha_t)^i] + i\ell(0)\mathbb{E}[\alpha_t(1-\alpha_t)^{i-1}].
\end{aligned}
$$

We plug in these expressions into equation (a.3) to get:

$$
\begin{aligned}
&\ell_0(1)\Big[\ell_0(i)\mathbb{E}[(1-\alpha_t)] + \ell_0(i-1)\mathbb{E}[\alpha_t]\Big]\ell_0(0)\mathbb{E}[(1-\alpha_t)^i]\\
&= \ell_0(i)\Big[\ell_0(1)\mathbb{E}[(1-\alpha_t)^i] + i\ell(0)\mathbb{E}[\alpha(1-\alpha_t)^{i-1}]\Big]\ell_0(0)\mathbb{E}[(1-\alpha_t)],
\end{aligned}
$$

or equivalently:

$$
\ell_0(i) = \ell_0(i-1)\frac{\ell_0(1)\mathbb{E}[\alpha_t]\mathbb{E}[(1-\alpha_t)^i]}{i\ell_0(0)\mathbb{E}[\alpha_t(1-\alpha_t)^{i-1}]\mathbb{E}[(1-\alpha_t)]}. \tag{a.4}
$$

Similarly, by considering the paths $(0,2,i,0)$ and $(0,i,2,0)$, we get:

$$
p(2|0)p(i|2)p(0|i) = p(i|0)p(2|i)p(0|2), \tag{a.5}
$$

or equivalently:

$$
\begin{aligned}
&\ell_0(2)\Big[\ell_0(i)\mathbb{E}[(1-\alpha_t)^2] + 2\ell_0(i-1)\mathbb{E}[\alpha_t(1-\alpha_t)] + \ell_0(i-2)\mathbb{E}[\alpha_t^2]\Big]\ell_0(0)\mathbb{E}[(1-\alpha_t)^i]\\
&= \ell_0(i)\Big[\ell_0(2)\mathbb{E}[(1-\alpha_t)^i] + i\ell_0(1)\mathbb{E}[\alpha_t(1-\alpha_t)^{i-1}] + \frac{i(i-1)}{2}\mathbb{E}[\alpha_t^2(1-\alpha_t)^{i-2}]\Big]\ell_0(0)\mathbb{E}[(1-\alpha_t)^2].
\end{aligned}
$$

Let us now expand the parenthesis on both sides and check that the first terms (resp. the second terms) on both sides exactly cancel out due to equation (a.4). Then the equality between the third terms shows that:

$$
\frac{\mathbb{E}[(1-\alpha_t)^{i-1}]\mathbb{E}[\alpha_t^2(1-\alpha_t)^{i-2}]}{\mathbb{E}[\alpha_t(1-\alpha_t)^{i-1}]\mathbb{E}[\alpha_t(1-\alpha_t)^{i-2}]} \tag{a.6}
$$

is constant when i varies. Let us denote:

$$m_{i-1} = \mathbb{E}[(1-\alpha_t)^{i-1}], \qquad \forall i > 1.$$

Then we have:

$$\mathbb{E}[\alpha_t(1-\alpha_t)^{i-1}] = m_{i-1} - m_i;$$
$$\mathbb{E}[\alpha_t^2(1-\alpha_t)^{i-2}] = \mathbb{E}[(1-\alpha_t)^i] - \mathbb{E}[(1-\alpha_t)^{i-1}] + \mathbb{E}[\alpha_t(1-\alpha_t)^{i-1}] = m_i + m_{i-2} - 2m_{i-1},$$

and the term in equation (a.6) becomes:

$$\frac{(m_i + m_{i-2} - 2m_{i-1})m_{i-1}}{(m_{i-1} - m_i)(m_{i-2} - m_{i-1})} = \frac{m_{i-1}}{m_{i-2} - m_{i-1}} - 1 - \frac{m_i}{m_{i-1} - m_i},$$

which should be constant in i. In other words, the sequence $(\frac{m_i}{m_{i-1}-m_i})_i$ is an arithmetic sequence. If this sequence is constant, then the sequence of integer moments of $1-\alpha_t$ is geometric. Since this sequence of moments characterizes a distribution on $[0,1]$, we conclude that α_t follows a point mass distribution, that is, process (X_t) is indeed INAR(1). Otherwise, when the increment of this arithmetic sequence is strictly positive, we have $\frac{m_i}{m_{i-1}-m_i} = (i-1)c + d$ for positive constants c and d, or equivalently:

$$\frac{m_i}{m_{i-1}} = \frac{(i-1) + d/c}{(i-1) + d/c + 1/c} = \frac{q+i-1}{p+q+i-1}, \qquad \forall i \geq 1,$$

where $p = \frac{1}{c}, q = \frac{d}{c} \in [0, \infty[$, or by iteration $m_i = \frac{B(p,q+i)}{B(p,q)}$. In other words the sequence of integer moments of $1-\alpha_t$ coincides with those of a beta distribution $\mathcal{B}eta(q,p)$, thus we deduce that the stochastic probability α_t follows the beta distribution $\mathcal{B}eta(p,q)$. Let us finally get back to equation (a.4) and show that, in this case, the distribution of (ϵ_t) is negative binomial. Summing up this equation for $i = 1, 2, ...$ yields:

$$q\mu = (q+\mu)\frac{\ell_0(1)}{\ell_0(0)}, \tag{a.7}$$

where μ is the expectation of ϵ_t. Thus equation (a.4) becomes:

$$\ell_0(i) = \ell_0(i-1)\frac{q+i-1}{qi}\frac{\ell_0(1)}{\ell_0(0)} = \ell_0(i-1)\frac{q+i-1}{i}\frac{\mu}{\mu+q}, \qquad \forall i \geq 1. \tag{a.8}$$

By iteration we deduce that (ϵ_t) follows a NB distribution with degree of freedom parameter q (and with mean $\mu \in \mathbb{R}^+$).

A5 Proof of Proposition 5

Let us first prove the lemma for $h = 2$. By Christoph and Schreiber (1998), the tail of the discrete stable distribution satisfies:

$$\mathbb{P}[X_{t+1} = k] = O(\frac{1}{k^{\nu+1}}),$$

when k goes to infinity. On the other hand, for a given X_t and for large X_{t+1} such that $X_{t+1} > X_t$, the noncausal conditional p.m.f. is:

$$\begin{aligned}
\ell(X_t|X_{t+1}) &= \sum_{i=0}^{\min(X_t,X_{t+1})} \binom{X_{t+1}}{i}\alpha^i(1-\alpha)^{X_{t+1}-i}\ell(X_t-i) \\
&= \sum_{i=0}^{X_t} \binom{X_{t+1}}{i}\alpha^i(1-\alpha)^{X_{t+1}-i}\ell(X_t-i) \\
&= (1-\alpha)^{X_{t+1}} \sum_{i=0}^{X_t} \binom{X_{t+1}}{i}\Big(\frac{\alpha}{1-\alpha}\Big)^i \ell(X_t-i). \qquad (a.9)
\end{aligned}$$

For fixed X_t, the term $\ell(X_t - i)$ is uniformly bounded for varying i, whereas the term $\binom{X_{t+1}}{i}$ is bounded by:

$$\binom{X_{t+1}}{i} = \frac{X_{t+1}(X_{t+1}-1)\cdots(X_{t+1}-i+1)}{i!} \leq \frac{X_{t+1}^i}{i!}.$$

Thus the RHS of equation (a.9) is bounded by a polynomial of X_{t+1} of degree X_t, and hence $\ell(X_t|X_{t+1}) = O((1-\alpha)^{X_{t+1}} \frac{X_{t+1}^{X_t}}{X_t!})$, when X_{t+1} goes to infinity. Thus by the Bayes' formula:

$$\begin{aligned} \ell(X_{t+1}|X_t) &= \frac{\ell(X_t|X_{t+1})\ell(X_{t+1})}{\ell(X_t)} \\ &= O\Big(\frac{(1-\alpha)^{X_{t+1}}}{X_{t+1}^{\nu+1-X_t}}\Big). \end{aligned}$$

Finally, let us explain why the above result remains true for higher horizons h. For instance for $h = 2$, we have, by iterated expectation formula:

$$\begin{aligned} \mathbb{E}[u^{X_t}|X_{t+2}] &= \mathbb{E}[(\alpha u + 1 - \alpha)^{X_{t+1}}|X_{t+2}]\mathbb{E}[u^{\tilde{\epsilon}_t}] \\ &= (\alpha^2 u + 1 - \alpha)^{X_{t+2}}\mathbb{E}[(\alpha u + 1 - \alpha)^{\tilde{\epsilon}_t}]\mathbb{E}[u^{\tilde{\epsilon}_t}] \\ &= (\alpha^2 u + 1 - \alpha)^{X_{t+2}} e^{-\beta(1-\alpha)^\nu \frac{(1-u)^\nu}{1-\alpha^\nu}} e^{-\beta \frac{(1-u)^\nu}{1-\alpha^\nu}}. \end{aligned}$$

In other words, the joint distribution of (X_t, X_{t+2}) has the same parametric form as that of (X_t, X_{t+1}), but with a different probability parameter α^2. Thus the above result for $h = 1$ still applies to the case $h = 2$.

A6 Proof of Proposition 6

The joint p.g.f. of (X_t, X_{t+1}) is:

$$\begin{aligned} \mathcal{L}(u, v) &:= \mathbb{E}[u^{X_t} v^{X_{t+1}}] = \mathbb{E}\Big[\mathbb{E}[u^{X_t}|X_{t+1}]v^{X_{t+1}}\Big] \\ &= \exp\Big(-\beta(1-u)^\nu - \beta\frac{[1-(\alpha u + 1 - \alpha)v]^\nu}{1-\alpha^\nu}\Big), \qquad \forall u, v \in [0, 1[. \end{aligned} \tag{a.10}$$

Then we get, for each $u \in [0, 1[$:

$$\mathbb{E}[X_{t+1}u^{X_t}] = \frac{\partial \mathcal{L}(u, v)}{\partial v}|_{v=1} = e^{-\beta\frac{(1-u)^\nu}{1-\alpha^\nu}} \frac{\beta\nu}{1-\alpha^\nu}(\alpha u + 1 - \alpha)\alpha^{\nu-1}(1-u)^{\nu-1}, \forall u \in [0, 1]. \tag{a.11}$$

Similarly, by considering the marginal p.g.f. $\mathcal{L}(u) := \mathbb{E}[u^{X_t}] = e^{-\beta\frac{(1-u)^\nu}{1-\alpha^\nu}}$, we get:

$$\begin{aligned} \frac{\partial \mathcal{L}(u)}{\partial u} &= \mathbb{E}[X_t u^{X_t-1}], \\ \text{or } \mathbb{E}[\alpha^{\nu-1}X_t u^{X_t}] = \alpha^{\nu-1}u\frac{\partial \mathcal{L}(u)}{\partial u} &= e^{-\beta\frac{(1-u)^\nu}{1-\alpha^\nu}}\frac{\beta\nu}{1-\alpha^\nu}\alpha^{\nu-1}(1-u)^{\nu-1}u. \end{aligned} \tag{a.12}$$

On the other hand, we can expand the LHS of (a.11) and (a.12) and get:

$$\mathbb{E}[X_{t+1}u^{X_t}] = \sum_{i=0}^{\infty} u^i \mathbb{E}[X_{t+1}|X_t = i]\mathbb{P}[X_t = i], \tag{a.13}$$

$$\mathbb{E}[\alpha^{\nu-1}X_t u^{X_t}] = \sum_{i=0}^{\infty} u^i \alpha^{\nu-1} i\mathbb{P}[X_t = i]. \tag{a.14}$$

Thus, to show that $\mathbb{E}[X_{t+1}|X_t = i]$ and $\alpha^{\nu-1}i$ are asymptotically equivalent, it suffices to show that the coefficients of the Taylor's expansions in u at $u = 0$ are asymptotically equivalent in equations (a.13) and (a.14). Let us expand:

$$e^{-\beta\frac{(1-u)^\nu}{1-\alpha^\nu}}\frac{\beta\nu}{1-\alpha^\nu}\alpha^{\nu-1}(1-u)^{\nu-1} = \sum_{i=0}^{\infty} c_i u^i. \tag{a.15}$$

If we multiply the LHS of this above equation times by $(\alpha u + 1 - \alpha)$, we get the RHS of (a.11), and, if we multiply it by u, we get the RHS of (a.12). Thus the $i-$th coefficient of the expansion of $\mathbb{E}[X_{t+1}u^{X_t}]$ is $\alpha c_{i-1} + (1-\alpha)c_i$, whereas that of $\mathbb{E}[\alpha^{\nu-1}X_t u^{X_t}]$ is c_{i-1}. Thus it suffices to show that when i goes to infinity, we have: $\frac{c_i}{c_{i-1}} \to 1$. This latter condition is satisfied since the radius of convergence of expansion (a.15) is equal to 1.

A7 Proof of Proposition 7

Let us first show that the distribution of (X_t, X_{t+1}) is within the domain of attraction of a certain bivariate $\alpha-$stable vector (Y_t, Y_{t+1}), where process (Y_t) is a noncausal $\alpha-$stable linear AR(1) process. For a deterministic sequence a_n that goes to infinity when n goes to infinity (at a rate to be determined later), we consider the scaled sample average:

$$\Big(\frac{X_t^{(1)} + X_t^{(2)} + \cdots + X_t^{(n)}}{a_n}, \frac{X_{t+1}^{(1)} + X_{t+1}^{(2)} + \cdots + X_{t+1}^{(n)}}{a_n}\Big),$$

where $(X_t^{(i)}, X_{t+1}^{(i)}), i = 1, ...n$, are i.i.d. copies of (X_t, X_{t+1}). By applying equation (a.10), the Laplace transform of this vector is equal to:

$$\Big\{\mathbb{E}[e^{-\frac{u}{a_n}X_t - \frac{v}{a_n}X_{t+1}}]\Big\}^n = \exp\Big\{n\beta(1-e^{-u/a_n})^\nu - \frac{n\beta}{1-\alpha^\nu}[1-(\alpha e^{-u/a_n}+1-\alpha)e^{-v/a_n}]\Big\}, \quad \forall u, v.$$

Straightforward, but tedious algebra, shows that, for $a_n = n^{1/\nu}$, which goes to infinity in n, the RHS of the above equality converges to:

$$\exp\Big(-\beta u^\nu - \frac{\beta}{1-\alpha^\nu}(\alpha u + v)^\nu\Big). \tag{a.16}$$

Thus the joint distribution of (X_t, X_{t+1}) belongs to the domain of attraction of some $\alpha-$stable distribution with Laplace transform given by equation (a.16) [see e.g. Resnick and Greenwood (1979)]. The spectral measure of the limiting distribution has two point masses at $(1,0)$ and $(\frac{\alpha}{\sqrt{1+\alpha^2}}, \frac{1}{\sqrt{1+\alpha^2}})$.[26] Then we apply Thm 4 of Resnick and Greenwood (1979) and conclude that process (X_t) satisfies Proposition 7.

A8 Proof of Lemma 5

The equivalence between $i)$ and $ii)$ is a consequence of Schweer (2015).

Let us show that conditions $i)$ and $iii)$ are equivalent. Recall that $X_{t+1} - \epsilon_{t+1} = X_t - \tilde{\epsilon}_{t+1}$, and that $X_t - \tilde{\epsilon}_{t+1}$ and $\tilde{\epsilon}_{t+1}$ are independent by definition (23). Then using a characterization theorem of the Poisson distribution [see the Theorem in Moran (1952)], $X_{t+1} - \epsilon_{t+1} = X_t - \tilde{\epsilon}_{t+1}$ is Binomial conditional on X_t if and only if $X_t - \tilde{\epsilon}_{t+1}$ and $\tilde{\epsilon}_{t+1}$ are Poisson, or equivalently if and only if process (X_t) is Poisson INAR(1).

For the equivalence between conditions $i)$ and $iv)$, we have by definition:

$$\epsilon_{t+1} = \tilde{\epsilon}_{t+1} + X_{t+1} - X_t = X_{t+1} - \sum_{i=1}^{X_{t+1}} Z_{i,t}^*, \tag{a.17}$$

[26]Alternatively, we recognize that the limit (a.16) is the Laplace transform of (Y_t, Y_{t+1}), where (Y_t) is a linear noncausal $\alpha-$stable AR(1) process defined by:

$$Y_t = \alpha Y_{t+1} + \xi_t,$$

where ξ_t is i.i.d., independent of the future values $Y_{t+1}, ...$, and follows $\alpha-$stable distribution, with Laplace transform $\mathbb{E}[e^{-u\xi_t}] = \exp(-\beta u^\nu)$, for all $u > 0$. By Fries (2018), Thm 2.2, this latter process satisfies the proposition.

where the $Z^*_{i,t}$'s are independent of X_{t+1}. On the other hand, we have:

$$X_t = \tilde{\epsilon}_{t+1} + \sum_{i=1}^{X_{t+1}} Z^*_{i,t},$$

where $\tilde{\epsilon}_{t+1}$ is independent of X_{t+1}. Thus ϵ_{t+1} and X_t are independent if and only if $X_{t+1} - \sum_{i=1}^{X_{t+1}} Z^*_{i,t}$ and $\sum_{i=1}^{X_{t+1}} Z^*_{i,t}$ are independent. A simple calculation shows that the joint p.g.f. of these two variables is equal to:

$$\mathbb{E}\Big[u^{X_{t+1}-\sum_{i=1}^{X_{t+1}} Z^*_{i,t}} v^{\sum_{i=1}^{X_{t+1}} Z^*_{i,t}}\Big] = \mathbb{E}\Big[[\alpha v + (1-\alpha)u]^{X_t}\Big].$$

By a similar argument as in Proposition 2, this latter p.g.f. is multiplicatively separable if and only if the p.g.f. of X_t is Poisson.

Finally, let us prove the equivalence between conditions $i)$ and $iv)$ by using the calendar time conditional p.m.f. [eq. (24)]. $X_{t+1} - \epsilon_{t+1}$ and ϵ_{t+1} are independent, if and only if their joint p.g.f. is separable. Using equation (a.17), we can easily check that this p.g.f. is equal to:

$$\mathbb{E}[u^{X_{t+1}-\epsilon_{t+1}} v^{\epsilon_{t+1}}] = \mathbb{E}\Big[[(\alpha u + (1-\alpha)v]^{X_{t+1}}\Big].$$

By the same argument as in Appendix A3, this p.g.f. is separable if and only if X_{t+1} is Poisson.

B Conditional p.m.f. for general noncausal INAR(1) processes

B1 The stationary distribution

Let us first show under what conditions a general noncausal INAR(1) process has a computable p.g.f.. We have:

$$\mathbb{E}[u^{X_t}] = \prod_{i=0}^{\infty} g(\alpha^i u + 1 - \alpha^i) = \exp\Big[\sum_{i=0}^{\infty} h(\alpha^i u + 1 - \alpha^i)\Big], \qquad \forall u \in [0,1[, \qquad \text{(b.1)}$$

where h is the log p.g.f. of $\tilde{\epsilon_t}$. Thus, if h has a simple Taylor's expansion everywhere[27] in $[0,1]$, we can expand each term $h(\alpha^i u + 1 - \alpha^i)$ around $u = 0$ and obtain the Taylor's expansion of $\log \mathbb{E}[u^{X_t}]$:

$$\begin{aligned}
\mathbb{E}[u^{X_t}] &= \exp\Big[\sum_{i=0}^{\infty}\sum_{k=0}^{\infty} h^{(k)}(1-\alpha^i)\frac{(\alpha^i u)^k}{k!}\Big] \\
&= \exp\Big[c_0 + c_1 u + c_2 u^2 + \cdots + c_n u^n + O(u^{n+1})\Big], \quad \text{(b.2)}
\end{aligned}$$

where all coefficients $c_j, j = 0, ..., n$ can be computed in closed form and we choose an order n such that the probability of X_t being larger than n is negligible. Then we have:

Proposition 9.

$$\begin{bmatrix} \mathbb{P}[X_t = 0] \\ \mathbb{P}[X_t = 1] \\ \cdots \\ \mathbb{P}[X_t = n-1] \\ \mathbb{P}[X_t = n] \end{bmatrix} = \exp(c_0) \sum_{j=0}^{n} \frac{1}{j!} \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ c_1 & 0 & 0 & \cdots & 0 \\ c_2 & c_1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ c_n & c_{n-1} & \cdots & c_1 & 0 \end{bmatrix}^j \begin{bmatrix} 1 \\ 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix}. \quad \text{(b.3)}$$

Proof. Let us remark that for any count variable X_t, we have:

$$\mathbb{E}[u^{X_t}] = \sum_{n=0}^{\infty} \mathbb{P}[X_t = n] u^n.$$

In other words, the stationary p.m.f. of X_t can be obtained by Taylor-expanding the corresponding p.g.f. More precisely, by equation (b.2), we have:

$$\begin{aligned}
\mathbb{E}[u^{X_t}] &= \exp\Big[c_0 + c_1 u + \cdots + c_n u^n + O(u^{n+1})\Big] \quad \text{(b.4)} \\
&= \exp(c_0) \sum_{j=0}^{n} (c_1 u + \cdots + c_n u^n)^j + O(u^{n+1}). \quad \text{(b.5)}
\end{aligned}$$

Then we obtain recursively, for each j, the expansion of $(c_1 u + \cdots + c_n u^n)^j$ up to order n using

[27] A large family of distributions satisfying this condition is the compound Poisson distributions, or equivalently infinitely divisible count distributions. Schweer and Weiss (2014) have shown that for these compound Poisson distributions, the p.m.f. of the stationary distribution can also be obtained through a recursive equation.

the following Lemma due to Lu (2018):

Lemma 6. *For any coefficients $\gamma_i, i = 0, ..., n$, the $(n+1)$ first coefficients of polynomial $(\sum_{j=1}^{n} \gamma_j u^j)^k$, where $k = 0, ...n$, are given by the column vector:*

$$\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ \gamma_1 & 0 & 0 & \cdots & 0 \\ \gamma_2 & \gamma_1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \gamma_n & \gamma_{n-1} & \cdots & \gamma_1 & 0 \end{bmatrix}^k \begin{bmatrix} 1 \\ 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix}, \qquad \forall k = 0, ..., n. \tag{b.6}$$

By this lemma and coefficient matching, we get the p.m.f. of the stationary distribution, that is equation (b.3).

□

Although the above algorithm involves a Taylor's expansion, it is an **exact** formula for the marginal p.g.f. Indeed the Taylor's expansion of the p.g.f. of a count distribution provides the exact value of the p.m.f.

Moreover, compared to standard matrix multiplications, the successive powers of the Toeplitz matrix in the lemma can be computed much more quickly since:

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

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 updating step involves at most the computation of n coefficients.

Finally, although equations (b.1) and (b.2) are derived for INAR(1) processes, that is when $Z_{i,t+1}$'s are Bernoulli distributed in equation (23). The case of general affine count processes with non-binary $Z_{i,t+1}$'s is similar, but slightly more complicated [see Lu (2018) for more details].

B2 Examples

Let us now give the expression of $(c_i)_{i=0}^{\infty}$ for some examples of (causal or noncausal) INAR(1) processes.

Example 4 (Negative binomial (NB) error)**.** Let us assume that $\tilde{\epsilon}_t$ is NB distributed, with p.g.f.:

$$\mathbb{E}[u^{\tilde{\epsilon}_t}] = \frac{(1-p)^{\nu}}{(1-pu)^{\nu}},$$

where parameter $p \in]0,1[$. Then the p.g.f. of the stationary distribution of the process is:

$$\begin{aligned}
\mathbb{E}[u^{X_t}] &= \exp\Big(\sum_{i=0}^{\infty}\nu[\log(1-p) - \log(1-p+p\alpha^i - p\alpha^i u)]\Big) \\
&= \exp\Big(\nu\sum_{i=0}^{\infty}[\log(1-p) - \log(1-p+p\alpha^i) - \log(1-\frac{p\alpha^i}{1-p+p\alpha^i}u)]\Big) \\
&= \exp\Big(\nu\sum_{i=0}^{\infty}[\log(1-p) - \log(1-p+p\alpha^i)]\Big)\exp\Big(-\nu\sum_{i=0}^{\infty}\sum_{k=1}^{\infty}[\frac{1}{k}\frac{p\alpha^i}{1-p+p\alpha^i}]^k u^k\Big) \\
&= \exp\Big(\nu\sum_{i=0}^{\infty}[\log(1-p) - \log(1-p+p\alpha^i)]\Big)\exp\Big(-\nu\sum_{k=1}^{\infty}\sum_{i=0}^{\infty}[\frac{1}{k}\frac{p\alpha^i}{1-p+p\alpha^i}]^k u^k\Big).
\end{aligned}$$

Thus we have:

$$c_0 = \nu \sum_{i=0}^{\infty} [\log(1-p) - \log(1-p+p\alpha^i)],$$
$$c_k = -\nu \sum_{i=0}^{\infty} \Big(\frac{1}{k}\frac{p\alpha^i}{1-p+p\alpha^i}\Big)^k, \qquad \forall k \geq 1.$$

Example 5 (Poisson-inverse Gaussian (PIG) error)**.** The negative binomial distribution can be viewed as a mixture of Poisson distributions with a gamma mixing distribution. Willmot (1987) introduced an alternative, mixed Poisson distribution with inverse Gaussian mixing distribution. The resulting distribution is called PIG[28] and is better suited for highly over-dispersed count distribution. Below we plot a simulated trajectory of the noncausal PIG INAR(1) process.

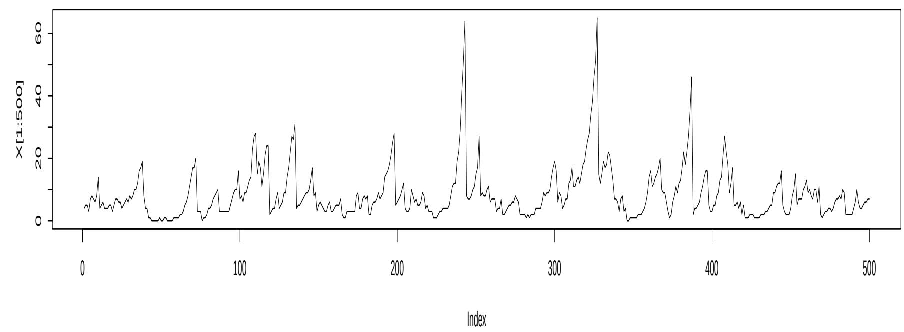

Figure B1: Simulated trajectory of a noncausal PIG INAR(1) process.

Let us now derive the p.m.f. of the stationary distribution. First, the p.g.f. of the error distribution is:

$$\mathbb{E}[u^{\epsilon_t}] = \exp\Big[\frac{\mu}{\beta}(1-\sqrt{1+2\beta-2\beta u})\Big].$$

[28]See also Zhu and Joe (2009) for an extension called Generalized PIG.

Hence the p.g.f. of the stationary distribution is given by:

$$
\begin{aligned}
\mathbb{E}[u^{X_t}] &= \exp\Big(\frac{\mu}{\beta}\sum_{i=0}^{\infty}[1-\sqrt{1+2\beta-2\beta(\alpha^i u+1-\alpha^i)}]\Big) \\
&= \exp\Big(\frac{\mu}{\beta}\sum_{i=0}^{\infty}[1-\sqrt{1+2\beta\alpha^i}\sqrt{1-\frac{2\beta\alpha^i}{1+2\beta\alpha^i}u}]\Big) \\
&= \exp\Big(\frac{\mu}{\beta}\sum_{i=0}^{\infty}(1-\sqrt{1+2\beta\alpha^i}\Big)\exp\Big[-\frac{\mu}{\beta}\sum_{i=0}^{\infty}\sqrt{1+2\beta\alpha^i}\sum_{k=1}^{\infty}\binom{k}{-\frac{1}{2}}(\frac{2\beta\alpha^i}{1+2\beta\alpha^i})^k u^k\Big],
\end{aligned}
$$

where the binomial coefficient $\binom{k}{-\frac{1}{2}}$ is defined by: $\binom{k}{-\frac{1}{2}} = \frac{\Gamma(\frac{1}{2})}{\Gamma(k+1)\Gamma(\frac{1}{2}-k)}, \forall k$, and $\Gamma(\cdot)$ is the Gamma function. Thus

$$
\begin{aligned}
c_0 &= \frac{\mu}{\beta}\sum_{i=0}^{\infty}(1-\sqrt{1+2\beta\alpha^i}), \\
c_k &= -\frac{\mu}{\beta}\sum_{i=0}^{\infty}\sqrt{1+2\beta\alpha^i}\binom{k}{-\frac{1}{2}}(\frac{2\beta\alpha^i}{1+2\beta\alpha^i})^k, \qquad \forall k \geq 1.
\end{aligned}
$$

Then the corresponding p.m.f. is obtained by expanding the exponential function and by coefficient matching.

B3 The corresponding conditional p.m.f.

Finally, once the marginal p.m.f. is computed, it suffices to apply Proposition 9 to compute the calendar time conditional p.m.f.

References

Al-Osh, M. and Alzaid, A. A. (1987). First-Order Integer-Valued Autoregressive (INAR(1)) Process. *Journal of Time Series Analysis*, 8(3):261–275.

Al-Osh, M. and Alzaid, A. A. (1988). Integer-Valued Moving Average (INMA) Process. *Statistical Papers*, 29(1):281–300.

Aly, E.-E. A. and Bouzar, N. (1994). Explicit Stationary Distributions for some Galton-Watson Processes with Immigration. *Stochastic Models*, 10(2):499–517.

Brännäs, K. and Quoreshi, A. (2010). Integer-Valued Moving Average Modelling of the Number of Transactions in Stocks. *Applied Financial Economics*, 20(18):1429–1440.

Cambanis, S. and Fakhre-Zakeri, I. (1995). On Prediction of Heavy-Tailed Autoregressive Sequences: Forward Versus Reversed Time. *Theory of Probability & Its Applications*, 39(2):217–233.

Christoph, G. and Schreiber, K. (1998). Discrete Stable Random Variables. *Statistics & Probability Letters*, 37(3):243–247.

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

Enciso-Mora, V., Neal, P., and Subba Rao, T. (2009). Efficient Order Selection Algorithms for Integer-Valued ARMA Processes. *Journal of Time Series Analysis*, 30(1):1–18.

Fakinos, D. (1984). The Infinite Server Queue with Arrivals Generated by a Non-Homogeneous Compound Poisson Process. *Journal of the Operational Research Society*, 35(5):439–445.

Foster, F. G. (1959). A Unified Theory for Stock, Storage and Queue Control. *Journal of the Operational Research Society*, 10(3):121–130.

Freeland, R. K. and McCabe, B. P. (2004). Forecasting Discrete Valued Low Count Time Series. *International Journal of Forecasting*, 20(3):427–434.

Fries, S. (2018). Conditional Moments of Anticipative α-Stable Markov Processes. *arXiv preprint arXiv:1805.05397*.

Fries, S. and Zakoian, J.-M. (2019). Mixed Causal-Noncausal AR Processes and the Modelling of Explosive Bubbles. *forthcoming Econometric Theory*, pages 1–37.

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

Gouriéroux, C. and Lu, Y. (2019b). Noncausal Affine Processes with Application to Derivative Pricing. *CREST DP*.

Gouriéroux, C. and Zakoïan, J.-M. (2017). Local Explosion Modelling by Non-Causal Process. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3):737–756.

Jia, Y., Kechagias, S., Livsey, J., Lund, R., and Pipiras, V. (2018). Latent Gaussian Count Time Series Modeling. *arXiv preprint arXiv:1811.00203*.

Joe, H. (1996). Time Series Models with Univariate Margins in the Convolution-Closed Infinitely Divisible Class. *Journal of Applied Probability*, 33(3):664–677.

Jørgensen, B. and Song, P. X.-K. (1998). Stationary Time Series Models with Exponential Dispersion Model Margins. *Journal of Applied Probability*, 35(1):78–92.

Klebaner, F. C. and Sagitov, S. (2002). The Age of a Galton-Watson Population with a Geometric Offspring Distribution. *Journal of Applied Probability*, 39(4):816–828.

Lanne, M. and Saikkonen, P. (2013). Noncausal Vector Autoregression. *Econometric Theory*, 29(3):447–481.

Livsey, J., Lund, R., Kechagias, S., and Pipiras, V. (2018). Multivariate Integer-Valued Time Series with Multivariate Flexible Autocovariances and their Application to Major Hurricane Counts. *Annals of Applied Statistics*, 12(1):408–431.

Lu, Y. (2018). The Term Structure of Predictive Distributions is Solvable for Thinning-Based Count Processes. *SSRN ejournal 3095219*.

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

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

McKenzie, E. (1986). Autoregressive Moving-Average Processes with Negative-Binomial and Geometric Marginal Distributions. *Advances in Applied Probability*, 18(3):679–705.

McKenzie, E. (1987). Innovation Distributions for Gamma and Negative Binomial Autoregressions. *Scandinavian Journal of Statistics*, 14(1):79–85.

Moran, P. (1952). A Characteristic Property of the Poisson Distribution. *Mathematical Proceedings of the Cambridge Philosophical Society*, 48(1):206–207.

Pickands, I. J. and Stine, R. A. (1997). Estimation for an M/G/∞ Queue with Incomplete Information. *Biometrika*, 84(2):295–308.

Resnick, S. and Greenwood, P. (1979). A Bivariate Stable Characterization and Domains of Attraction. *Journal of Multivariate Analysis*, 9(2):206–221.

Schweer, S. (2015). On the Time-Reversibility of Integer-Valued Autoregressive Processes of General Order. In *Stochastic Models, Statistics and Their Applications, eds. A., Steland, E., Rafajlowicz, K., Szajowski*, pages 169–177. Springer.

Schweer, S. and Weiss, C. H. (2014). Compound Poisson INAR (1) Processes: Stochastic Properties and Testing for Overdispersion. *Computational Statistics & Data Analysis*, 77:267–284.

Schweer, S. and Wichelhaus, C. (2015). Queueing Systems of INAR (1) Processes with Compound Poisson Arrivals. *Stochastic Models*, 31(4):618–635.

Silva, M. E., Pereira, I., and McCabe, B. (2019). Bayesian Outlier Detection in Non-Gaussian Autoregressive Time Series. *Journal of Time Series Analysis*, 40(5):631–648.

Steutel, F. and Van Harn, K. (1979). Discrete Analogues of Self-Decomposability and Stability. *Annals of Probability*, 7(5):893–899.

Weiss, C. H. (2010). INARCH (1) Processes: Higher-Order Moments and Jumps. *Statistics & Probability Letters*, 80(23-24):1771–1780.

Wheatley, S., Schatz, M., and Sornette, D. (2018). The ARMA Point Process and its Estimation. *arXiv preprint arXiv:1806.09948*.

Willmot, G. E. (1987). The Poisson-Inverse Gaussian Distribution as an Alternative to the Negative Binomial. *Scandinavian Actuarial Journal*, 1987(3-4):113–127.

Zheng, H., Basawa, I. V., and Datta, S. (2007). First-order Random Coefficient Integer-Valued Autoregressive Processes. *Journal of Statistical Planning and Inference*, 137(1):212–229.

Zhu, R. and Joe, H. (2009). Modelling Heavy-Tailed Count Data Using a Generalised Poisson-Inverse Gaussian Family. *Statistics & Probability Letters*, 79(15):1695–1703.