Noncausal counting processes: A queuing perspective*

Christian Gouriéroux

University of Toronto, Toulouse School of Economics and CREST
e-mail: gouriero@ensae.fr

Yang Lu

Concorodia University, Montreal, QC, Canada
e-mail: yang.lu@concordia.ca

Abstract: We introduce noncausal counting processes, defined by time-reversing an INAR(1) process, a non-INAR(1) Markov affine counting process, or a random coefficient INAR(1) [RCINAR(1)] process. The noncausal processes are shown to be generically time irreversible and their calendar time dynamic properties are unreplicable by existing causal models. In particular, they allow for locally bubble-like explosion, while at the same time preserving stationarity. Many of these processes have also closed form calendar time conditional predictive distribution, and allow for a simple queuing interpretation, similar as their causal counterparts.

Contents

1. Introduction

Recently, real-valued linear noncausal processes have raised much interest thanks
to their ability to capture irreversible[1], bubble-like phenomena, that are widely
observed in financial applications [see e.g. [26, 20]]. However, these models are
not suitable for (low) counting processes. To motivate the need for introducing
"bubbles" for counting 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]

To understand and predict (micro-)floods, such high-frequency, high-resolution
rainfall data is much more useful than traditional ones, which are often either too
temporally (daily or monthly) aggregated, or spatially aggregated across vari-
ous sites are hence real-valued. We see that the above hourly rainfall amount
curve features cycles of "bubbles", or cloudbursts, characterized by a period of
steady increase of the rainfall intensity analogous to the accumulation phase of
a financial bubble, and then followed by a sharp decrease, which spells the end
of the cloudburst, that is the analog of the burst of the bubble. In the afore-
mentioned literature on real-valued time series, it has been shown that unlike
standard ARMA type processes, a new class of noncausal processes can capture

---

[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/
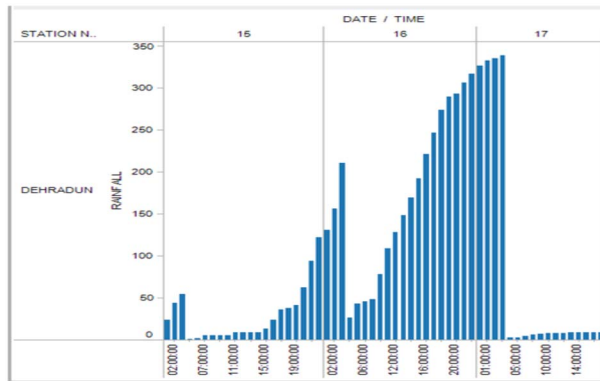playing-uttarakhand-rainfall-data

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

the bubble phenomenon very well. A similar problem exists for count time series, since the path of a standard, say, INAR(1) process [see equation (1) below for its definition] 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 introduce the notion of noncausality to the counting process literature. To fix the idea, let us first consider Markov, causal 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 [30], [1]]. 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 [33, 29, 38]]. The INAR(1) is the most famous counting 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. For instance, in [35, 18], the $Z_{i,t+1}$'s have geometric distribution and ϵ_{t+1} is negative binomial, whereas in the INARCH(1) model [[41]], the $Z_{i,t+1}$'s are Poisson[3]. These processes are also called Galton-Watson process in the probability lit-

---

[3]See also the review of [40].

erature [see e.g. [25]], where 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, who can independently give rise to 0, 1, 2... descendants. The major appeal of the INAR(1) model a) and its extension b) is the special form of the conditional probability generating function (p.g.f.) of X_{t+1} given $\underline{X_t}$, which characterizes the transition distribution. Indeed, it is conveniently given by:

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

which 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 [9]]. In the counting process context, [28] shows that they have tractable marginal and predictive distributions at any horizons.

c) Alternatively, one can also extend the INAR(1) model a) by relaxing the i.i.d. assumption of the $Z_{i,t+1}$'s and allow them to be only conditionally i.i.d. given a stochastic probability parameter p_{t+1}, which itself is an i.i.d. sequence that is independent of ϵ_{t+1}, as well as of past observations $\underline{X_t}$. This model is called random coefficient INAR(1), or RCINAR(1) [see e.g. [44]]. This model has a similar queuing interpretation as the INAR(1) process, but with stochastic probability of leaving for current customers.

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 equal to 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.

Model (1) is called causal, since X_{t+1} does not depend on the values of the future shocks $\epsilon_{t+2}, \epsilon_{t+3}, ...$, or future latent factors $Z_{i,t+2}, Z_{i,t+3}$, i varying. The noncausal counting process introduced in this paper can be defined simply as a Markov process whose reverse time dynamics satisfies (1). Our main contributions are the following ones. First, we characterize all time reversible counting processes satisfying model (1). From a practical modeling perspective, these examples are to be discarded since they cannot capture the (time) asymmetry of the bubble dynamics. Second, we show that when process (1) is time irreversible, its noncausal version is indeed suitable for capturing bubbles. In particular, when an affine process [including INAR(1)] is time reversed, the resulting noncausal process has necessarily a non-affine dynamics. We then derive in closed form this conditional 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 steady increase of the bubble). Thirdly, we show that for noncausal INAR(1), an alternative, equivalent specification is based on the aforementioned queuing interpretation, but with a different set of distributional assumptions. This new queuing model is specified through the departure cohorts, rather than the arrival cohorts. Under

this noncausal specification, the number of arrivals are typically heteroscedastic across time and dependent on the current population size.

The rest of the paper is organized as follows. Section 2 studies time-reversed Markov counting process, including their time (ir-)reversibility, as well as the calendar time predictive distribution. Section 3 relates the causal and noncausal INAR(1) model through a same, infinite server queuing system. Section 4 concludes. Proofs, as well as some results for noncausal RCINAR(1) processes are gathered in Appendices.

2. Time reversing a Markov counting process

Let us now analyze irreversible noncausal processes and their calendar time dynamics. This provides new families of Markov dynamics for counting process, that are easy to analyze under closed-form, and are appropriate to capture bubble phenomena.

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

$$\forall t, \qquad X_t = \sum_{i=1}^{X_{t+1}} Z_{t,i} + \tilde{\epsilon}_{t+1}, \tag{3}$$

where shock $\tilde{\epsilon}_{t+1}$ 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 and i vary, 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, process (X_t) is called noncausal RCINAR(1), if in model (3) the $Z_{t,i}$'s are conditionally i.i.d. and Bernoulli $\mathcal{B}(1, p_t)$, where p_t is itself i.i.d.

2.1. Time reversibility

The noncausal process (3) can be interpreted as a counting process (1) observed in reverse time. Let us derive some general properties of such processes. First, we note that the Markov and stationarity properties are usually preserved under time reversal.

Lemma 1 (See Section 1 of [6]). *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 (3)? This is equivalent to the time

reversibility of the process by the following lemma:

Lemma 2 (See Proposition 3 of [19]). *If the noncausal affine process* (3) *is also affine in calendar time, and is weakly ergodic[4] in both time directions, then it is time reversible.*

Let us now characterize the set of all affine time reversible processes. This result is important, since, once we exclude these reversible examples, other noncausal affine process will have non-affine calendar time dynamics.

Proposition 1 (Characterization of reversible affine counting processes). *Suppose a Markov counting process is affine and its support is $\mathbb{N}$. Then it is time reversible if and only if it is:*

- *Either the Poisson-INAR(1) process;*
- *or an affine process in which $Z_{i,t}$'s and ϵ_t have p.g.f.:*

$$\mathbb{E}[u^{Z_{i,t}}] = \frac{1 + \theta_2(1-u)}{1 + \theta_1(1-u)}, \qquad \mathbb{E}[u^{\epsilon_t}] = \frac{1}{[1 + \theta_1(1-u)]^\delta}, \quad \forall u \in [0, 1 + \frac{1}{\theta_1}), \ (4)$$

respectively, where parameters $\delta > 0$, $\theta_1 > 0$, $\theta_2 > -1$, and satisfy $0 < \theta_1 - \theta_2 < 1$. This process has the NB stationary distribution with p.g.f.:

$$\mathbb{E}[u^{X_t}] = \frac{1}{[1 + \frac{\theta_1}{1-(\theta_1-\theta_2)}(1-u)]^\delta}, \qquad \forall u \in [0, 1 + \frac{1 - \theta_1 + \theta_2}{\theta_1}).$$

Proof. See Appendix A.1. $\qquad\qquad\square$

The second reversible process above is introduced by [3]. The first equality in equation (4) defines a count distribution when parameter θ_2 is either positive, or between -1, 0. If it is positive, then this distribution is a zero-inflated geometric distribution since:

$$\frac{1 + \theta_2(1-u)}{1 + \theta_1(1-u)} = p + (1-p)\frac{1}{1 + \theta_1(1-u)}, \qquad \forall u \in [0, 1],$$

where probability parameter p is given by $p = 1 - \frac{\theta_2}{\theta_1}$. Moreover, in the special case where $\theta_2 = 0$ and $\theta_1 \in [0, 1[$, this process becomes the NBAR(1) process of [18], with geometric $Z_{i,t}$'s. By [18], Prop. 2, the $h-$step-ahead conditional p.g.f. of a NBAR(1) is: $\mathbb{E}[u^{X_{t+h}}|X_t] = \frac{[1 + \theta_1 \frac{1-\theta_1^h}{1-\theta_1}(1-u)]^{X_t}}{[1 + \theta_1 \frac{1-\theta_1^{h-1}}{1-\theta_1}(1-u)]^{\delta + X_t}}$. Thus a temporally aggregated NBAR(1) process, that is $(Y_t) = (X_{ht})$ where process (X_t) is NBAR(1), belongs also to affine family (4).

Let us now show that the irreversible, noncausal dynamics defined by (10)

---

[4]The (Markov) process (X_t) is weakly ergodic if the conditional distribution of X_{t+h} given X_t converges to the stationary distribution when h goes to infinity.

has quite unique properties that distinguish them from the standard causal models.

2.2. 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 Poisson, i.e., (X_t) is irreversible. Since in practice X_{t+1} is only observed after X_t, the noncausal representation (3) is not directly usable for the forecasting of X_{t+1} given observation of X_t. Let us derive the corresponding calendar time one-step-ahead conditional p.m.f. $P(x_{t+1}|x_t) := \mathbb{P}(X_{t+1} = x_{t+1}|X_t = x_t)$. By the Bayes formula, we have:

$$P(x_{t+1}|x_t) = \frac{P(x_{t+1}, x_t)}{P(x_t)} = \frac{P(x_{t+1})}{P(x_t)} P(x_t|x_{t+1}), \tag{5}$$

where $P(x_t), P(x_{t+1})$ denote the stationary p.m.f. of the process (X_t) evaluated at values x_t and x_{t+1}, respectively, and $P(x_t|x_{t+1}) := \mathbb{P}(X_t = x_t|X_{t+1} = x_{t+1})$ is the one-step-backward conditional p.m.f. By simple convolution we have:

$$P(x_t|x_{t+1}) = \sum_{i=0}^{\min(x_t, x_{t+1})} \binom{x_{t+1}}{i} p^i (1-p)^{x_{t+1}-i} \mathbb{P}[\tilde{\epsilon}_{t+1} = x_t - i]. \tag{6}$$

Thus, so long as the p.m.f.'s of both $(\tilde{\epsilon}_t)$ and (X_t) have closed form, so does the predictive distribution $P(x_{t+1}|x_t)$. Moreover, this result can be extended in two important ways. First, we can also deduce close form expression for the predictive distribution $P(x_{t+h}|x_t)$, for any horizon h larger than 1. Indeed, the Bayes' formula can also be used on the joint distribution of (X_t, X_{t+h}), and the h-step-backward (resp. forward) predictive distribution of a noncausal (resp. causal) INAR(1) process is known to allow for closed form expressions [see [28]]. Second, instead of focusing on INAR(1) process, we can extend this result to affine Markov counting processes, whose transition distribution and stationary distribution are also generically available in closed form [see [28]].

Let us now provide two examples of noncausal, irreversible INAR(1) processes.

2.2.1. Example 1: Noncausal geometric INAR(1)

We first consider the noncausal analog of geometric INAR(1) process [see [31]]. This process has marginally the geometric distribution with p.g.f.: $\mathbb{E}[u^{X_t}] = \frac{1}{1+\beta(1-u)}$, with mean $\beta \in]0, \infty[$, or alternatively probability parameter $\frac{1}{1+\beta}$, which is not necessarily equal to p. Then the p.g.f. of $\tilde{\epsilon}_t$ is given by:

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

This is a zero-inflated geometric distribution with p.m.f.:

$$\mathbb{P}[\tilde{\epsilon}_t = n] = \frac{\beta^n}{(1+\beta)^{n+1}}(1-p) + p\mathbb{1}_{n=0}, \qquad \forall n \in \mathbb{N}. \tag{7}$$

Figure 2 displays a path of a noncausal geometric INAR(1). Such a simulated path is easily obtained by simulating backwards a causal INAR(1) process. The parameter values are set equal to $p = 0.85$ and $\beta = 3$.
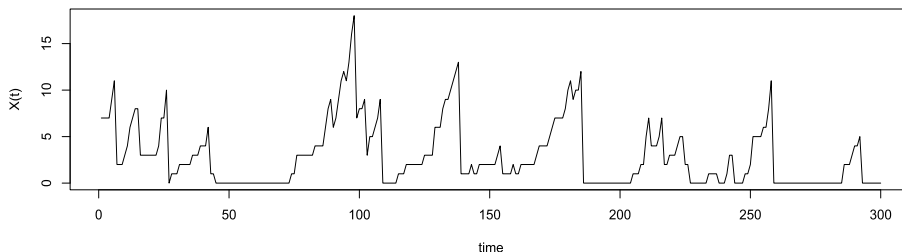

FIG 2. *Path of a noncausal geometric INAR(1) process. We see several bubble episodes with abrupt burst of the bubble.*

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

$$P(x_t|x_{t+1}) = p\mathbb{1}_{x_t \leq x_{t+1}}\binom{x_{t+1}}{x_t}p^{x_t}(1-p)^{x_{t+1}-x_t}$$
$$+ (1-p)\sum_{n=0}^{\min(x_t,x_{t+1})}\binom{x_{t+1}}{n}p^n(1-p)^{x_{t+1}-n}\frac{\beta^{x_t-n}}{(\beta+1)^{x_t-n+1}}.$$

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

$$P(x_{t+1}|x_t) = \frac{P(x_t|x_{t+1})P(x_{t+1})}{P(x_t)} = P(x_t|x_{t+1})\frac{\beta^{x_{t+1}-x_t}}{(\beta+1)^{x_{t+1}-x_t}}.$$

Figure 3 plots the conditional p.m.f. $\mathbb{P}[X_{t+1} = \cdot|X_t = i]$ of the above non-causal 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. Note that both processes have the same stationary distribution.

In both panels, the two conditional p.m.f.'s coincide at argument $x = i$. This is a direct consequence of the Bayes formula (6) and of the stationarity of the process. 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 expansion of the bubble. On the other hand it has also a larger probability of hitting zero, which corresponds to the collapse of the bubble. Figure 4 displays the corresponding conditional expectations 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.
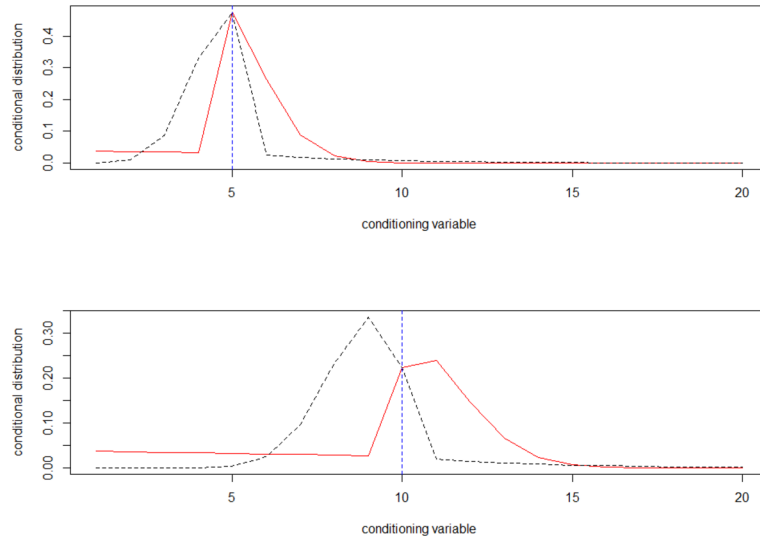
FIG 3. *Comparison between the conditional p.m.f. in a noncausal and causal geometric INAR(1) process. In the upper panel we set $i = 5$ and in the lower panel $i = 10$. In both panels, the red full (resp. black dashed) line represents the curve of the noncausal (resp. causal) conditional p.m.f.. In both figures we have also added a vertical blue line crossing the x axis at i.*
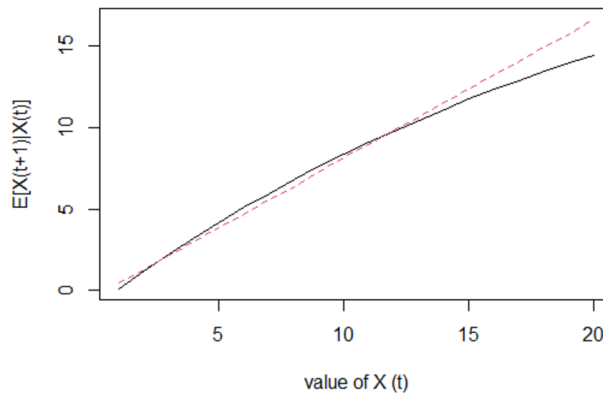
FIG 4. *Conditional expectations in a noncausal and causal geometric INAR(1) process. Black full line: noncausal model, red dashed line: causal model.*

In reverse time, the conditional expectation $\mathbb{E}[X_t|X_{t+1} = i] = pi + \mathbb{E}[\epsilon_t]$ is affine in i, but this affine property no longer holds for the causal expectation $\mathbb{E}[X_{t+1}|X_t = i]$. This is expected, since process (X_t) is irreversible. We 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.

2.2.2. Example 2: Noncausal discrete stable INAR(1) [DS-INAR(1)]

Let us now consider another noncausal INAR(1) model, in which the distribution of $(\tilde{\epsilon}_t)$ is discrete stable $\mathcal{DS}(\beta, \nu)$ [see e.g. [39, 7]][5], with p.g.f.:

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

where scale parameter $\beta > 0$, and shape parameter $\nu \in]0, 1[$.[6] The corresponding p.m.f. is:

$$\mathbb{P}[\tilde{\epsilon}_t = k] = (-1)^k \sum_{j=0}^{\infty} \binom{j\nu}{k} \frac{(-\beta)^j}{j!}, \qquad \forall k \in \mathbb{N}. \tag{8}$$

This distribution is such that the mean $\mathbb{E}[\tilde{\epsilon}_t]$ is infinite [see [7]], and thus the resulting model is suitable for heavy-tailed counting process data. Figure 5 plots a trajectory of a noncausal DS-INAR(1) process with $\nu = 0.5$, $\beta = 0.05$, and $p = 0.5$.
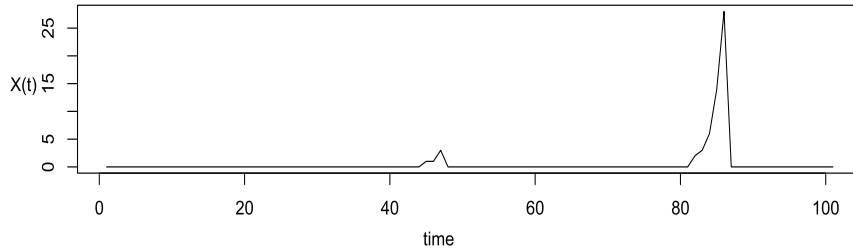
FIG 5. *Simulated path of a noncausal DS-INAR(1) process. We can see two bubble periods.*

Let us first derive the stationary distribution of this process. We have:

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

By iterating we get:

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

which corresponds to the $\mathcal{DS}(\frac{\beta}{1-p^{\nu}}, \nu)$ distribution. Thus by equation (8), the marginal p.m.f. is:

$$\mathbb{P}[X_t = k] = (-1)^k \sum_{j=0}^{\infty} \binom{j\nu}{k} \frac{1}{j!} \left(\frac{-\beta}{1-p^{\nu}}\right)^j, \qquad \forall k \in \mathbb{N}. \tag{9}$$

---

[5]The discrete stable distribution can be interpreted as a Poisson distribution with stochastic, alpha-stable distribution distributed parameter.

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

Let us now plot the analog of Figure 3. Using equations (6), (8) and (9), we can compute the conditional p.m.f. $P(x_{t+1}|X_t = i)$. Since a discrete stable distribution is heavy-tailed, we will consider two large values for the conditioning observation $i = 10$, and 25. These conditional p.m.f.'s are plotted in Figure 6.
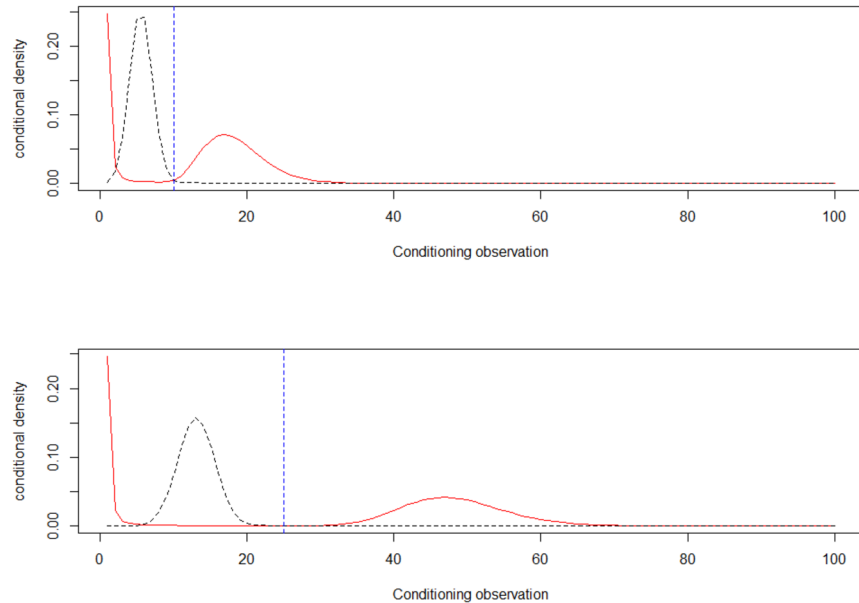
FIG 6. *Comparison between the direct time conditional p.m.f. $\mathbb{P}[X_{t+1} = \cdot|X_t = i]$ (red full line) and the reverse time counterpart, i.e. $\mathbb{P}[X_t = \cdot|X_{t+1} = i]$ (black dashed line) in a noncausal DS-INAR(1) process. Upper panel: the conditioning observation is $i = 10$; 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 3. In particular, they have two and one local modes, respectively. 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.

2.3. *Other examples*

Besides the two examples given in Section 2.2, 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 [28]]. In this subsection, 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 shocks.

2.3.1. The stationary distribution

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

$$\mathbb{E}[u^{X_t}] = \prod_{i=0}^{\infty} g(p^i u + 1 - p^i) = \exp\Big[\sum_{i=0}^{\infty} h(p^i u + 1 - p^i)\Big], \qquad \forall u \in [0,1], \quad (10)$$

where h is the log p.g.f. of $\tilde{\epsilon}_t$. Thus, if h has a simple Taylor's expansion everywhere[7] in $[0,1]$, we can expand each term $h(p^i u + 1 - p^i)$ around $u = 0$ and obtain the Taylor's expansion of $\log \mathbb{E}[u^{X_t}]$:

$$\mathbb{E}[u^{X_t}] = \exp\Big[\sum_{i=0}^{\infty} \sum_{k=0}^{\infty} h^{(k)}(1 - p^i)\frac{(p^i u)^k}{k!}\Big]$$

$$= \exp\Big[c_0 + c_1 u + c_2 u^2 + \cdots + c_n u^n + O(u^{n+1})\Big], \qquad (11)$$

where the 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 2.

$$
\begin{bmatrix}
\mathbb{P}[X_t = 0] \\
\mathbb{P}[X_t = 1] \\
\cdots \\
\mathbb{P}[X_t = n-1] \\
\mathbb{P}[X_t = n]
\end{bmatrix}
= \exp(c_0) \sum_{j=0}^{n} \frac{1}{j!}
\begin{bmatrix}
0 & 0 & 0 & \cdots & 0 \\
c_1 & 0 & 0 & \cdots & 0 \\
c_2 & c_1 & 0 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
c_n & c_{n-1} & \cdots & c_1 & 0
\end{bmatrix}^{j}
\begin{bmatrix}
1 \\
0 \\
0 \\
\cdots \\
0
\end{bmatrix}. \quad (12)
$$

Proof. See Appendix A.2. □

Although the above algorithm involves a Taylor's expansion, it is an *exact* formula, as it is obtained by matching two Taylor's expansions.

Finally, although equations (10) and (11) are derived for INAR(1) processes, the case of general affine counting processes with non-binary $Z_{i,t+1}$'s is similar [see [28] for more details].

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

2.3.2. Negative binomial (NB) shock

Let us assume that $\tilde{\epsilon}_t$ is NB distributed, with p.g.f.:

$$\mathbb{E}[u^{\tilde{\epsilon}_t}] = \frac{(1 - p_0)^\nu}{(1 - p_0 u)^\nu}, \qquad \forall u \in [0, \frac{1}{p_0})$$

---

[7] A large family of distributions satisfying this condition is the compound Poisson distributions, or equivalently the infinitely divisible count distributions. [37] have shown that for these compound Poisson distributions, the p.m.f. of the stationary distribution can also be obtained through a recursive equation.

where parameter $p_0 \in]0,1[$. Then the p.g.f. of the stationary distribution of the process is:

$$\mathbb{E}[u^{X_t}] = \exp\Big(\sum_{i=0}^{\infty} \nu[\log(1-p_0) - \log(1-p_0+p_0p^i - p_0p^i u)]\Big)$$

$$= \exp\Big(\nu\sum_{i=0}^{\infty}[\log(1-p_0) - \log(1-p_0+p_0p^i) - \log(1 - \frac{p_0p^i}{1-p_0+p_0p^i}u)]\Big)$$

$$= \exp\Big(\nu\sum_{i=0}^{\infty}[\log(1-p_0) - \log(1-p_0+p_0p^i)] - \nu\sum_{i=0}^{\infty}\sum_{k=1}^{\infty}[\frac{1}{k}\frac{p_0p^i}{1-p_0+p_0p^i}]^k u^k\Big)$$

$$= \exp\Big(\nu\sum_{i=0}^{\infty}[\log(1-p_0) - \log(1-p_0+p_0p^i)] - \nu\sum_{k=1}^{\infty}\sum_{i=0}^{\infty}[\frac{1}{k}\frac{p_0p^i}{1-p_0+p_0p^i}]^k u^k\Big).$$

Thus we have:

$$c_0 = \nu\sum_{i=0}^{\infty}[\log(1-p_0) - \log(1-p_0+p_0p^i)],$$

$$c_k = -\nu\sum_{i=0}^{\infty}\Big(\frac{1}{k}\frac{p_0p^i}{1-p_0+p_0p^i}\Big)^k, \qquad \forall k \geq 1.$$

2.3.3. Poisson-inverse Gaussian (PIG) shock

The negative binomial distribution can be viewed as a mixture of Poisson distributions with a gamma mixing distribution. [43] introduced an alternative, mixed Poisson distribution with inverse Gaussian mixing distribution. The resulting distribution is called PIG[8] and is better suited for heavy-tailed count distributions. The causal version of the PIG INAR(1) process has been introduced by [4]. Below we plot a simulated trajectory of the noncausal PIG INAR(1) process, in which the inverse Gaussian distribution has mean 0.5 and dispersion 10.

The p.g.f. of the shock distribution is:

$$\mathbb{E}[u^{\tilde{\epsilon}_t}] = \exp\Big[\frac{\mu}{\beta}(1 - \sqrt{1+2\beta-2\beta u})\Big], \qquad \forall u \in [0, 1 + \frac{1}{2\beta}].$$

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

$$\mathbb{E}[u^{X_t}] = \exp\Big(\frac{\mu}{\beta}\sum_{i=0}^{\infty}[1 - \sqrt{1+2\beta-2\beta(p^i u + 1 - p^i)}]\Big)$$

$$= \exp\Big(\frac{\mu}{\beta}\sum_{i=0}^{\infty}[1 - \sqrt{1+2\beta p^i}\sqrt{1 - \frac{2\beta p^i}{1+2\beta p^i}u}]\Big)$$

$$= \exp\Big(\frac{\mu}{\beta}\sum_{i=0}^{\infty}\Big[1 - \sqrt{1+2\beta p^i} - \sqrt{1+2\beta p^i}\sum_{k=1}^{\infty}\binom{k}{-\frac{1}{2}}\Big(\frac{2\beta p^i}{1+2\beta p^i}\Big)^k u^k\Big]\Big),$$

---

[8]See also [45] for an extension called Generalized PIG.
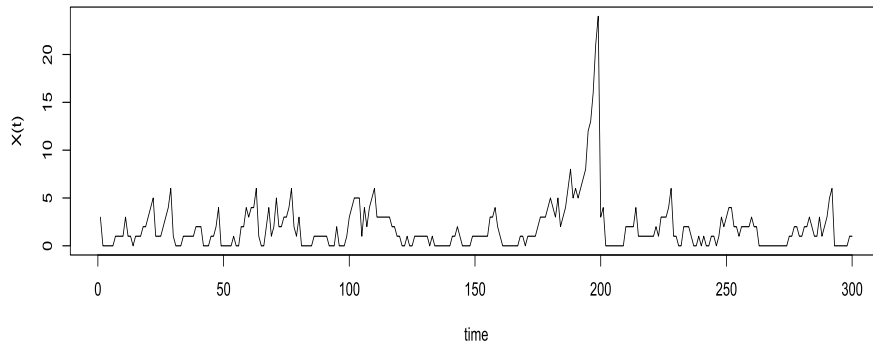
FIG 7. *Simulated trajectory of a noncausal PIG INAR(1) process.*

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

$$c_0 = \frac{\mu}{\beta} \sum_{i=0}^{\infty} (1 - \sqrt{1 + 2\beta p^i}),$$

$$c_k = -\frac{\mu}{\beta} \sum_{i=0}^{\infty} \sqrt{1 + 2\beta p^i} \binom{k}{-\frac{1}{2}} \left(\frac{2\beta p^i}{1 + 2\beta p^i}\right)^k, \qquad \forall k \geq 1.$$

2.4. Extremal behavior

Let us now show that the dynamics of a noncausal INAR(1) process can be totally different from its causal counterpart, especially for large current values of X_t. For illustration purpose, we re-focus on the noncausal DS-INAR(1) process introduced in Section 2.2.2.

Since the discrete stable distribution has an infinite mean, the noncausal conditional expectation $\mathbb{E}[X_t|X_{t+1}]$ does not exist. This might explain why the causal, DS-INAR(1) process has never been proposed in the literature. Interestingly, the calendar time dynamics of a noncausal DS-INAR(1) has thin tail as shown in the next proposition.

Proposition 3. *For fixed X_t, the conditional p.m.f. is asymptotically equal to:*

$$\mathbb{P}[X_{t+h} = x_{t+h}|X_t = x_t] = O\left[\frac{(1 - p^h)^{x_{t+h}}}{x_{t+h}^{\nu+1-x_t}}\right], \qquad \textit{when } x_{t+h} \textit{ goes to infinity,}$$

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

Proof. See Appendix A.3. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

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. The values are computed numerically from those of the conditional p.m.f., by truncating the identity:

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

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.

conditional expectation
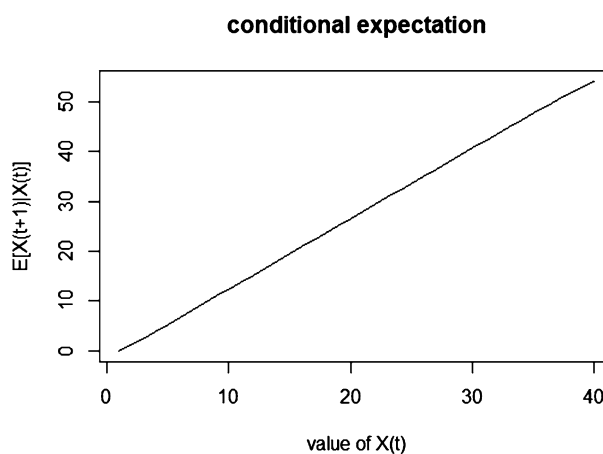
FIG 8. *Causal conditional expectation as a function of the conditioning variable in the non-causal DS-INAR(1) model.*

The conditional expectation seems to increase nearly linearly in i, with a slope that is larger than 1. This impression is confirmed by Proposition 4, which is the discrete analog of a result for (real-valued) noncausal linear alpha-stable AR(1) processes [see [6, 20]], for which the convergence in equation (13) is simply replaced by an equality.

Proposition 4. *For the noncausal DS-INAR(1) process, the conditional expectation is such that:*

$$\frac{\mathbb{E}[X_{t+h}|X_t = i]}{i} \to (p^{\nu-1})^h, \qquad \forall h \geq 1, \tag{13}$$

when i goes to infinity.

Proof. See Appendix A.4. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \square$

Thus, for large values of X_t, the conditional expectation is linear with an "asymptotic" autoregressive coefficient equal to $p^{\nu-1}$, which is *larger* than 1, since $\nu, p \in [0, 1[$. In particular this ratio is different from the autoregressive

coefficient p in reverse time, which is constant and *smaller* than 1. Therefore, 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 $p < 1$ and hence the corresponding limit of $\frac{\mathbb{E}[X_{t+1}|X_t=i]}{i}$ is $p < 1$.

As shown in Figure 6, 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 5. *For a given $\epsilon > 0$, when X_t goes to infinity,*

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

Proof. See Appendix A.5. □

Thus, when the process is currently in the bubble episode with X_t 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 p^ν, the bubble continues to grow at the geometric rate of $\frac{1}{p}$, whereas with fixed probability $1 - p^\nu$, the bubble collapses. This also echoes Proposition 4, which says that the conditional expectation $\mathbb{E}[X_{t+1}|X_t]$ is approximately equal to $p^{\nu-1}X_{t-1}$, that is $\frac{X_t}{p}$ times the probability p^ν.

Proposition 5 is linked to the recent literature on (linear) noncausal AR(1) processes, which is the real-valued analog of the noncausal DS-INAR(1) [[15], Prop. 2.2]. 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 [15] applies.

The presence of several local modes also implies that standard point forecasting tools for time series counts, based on the conditional expectation and the (single) conditional mode, only, [see [14]] can be misleading. Hence the importance of the closed form conditional p.m.f. derived in this paper. Whereas the marginal distribution of a (causal or noncausal) INAR(1) process is necessarily uni-modal [see [39]], the conditional distribution of a noncausal INAR(1) process is instead often multi-modal. This is the reason for bubble detection by bivariate pattern recognition developed in [17].

3. The queuing interpretation

Besides its tractability for estimation and (nonlinear) prediction purposes, another advantage of the causal INAR(1) process is its intuitive, queuing interpretation. In this section, we show that a duality between arrival and departure events leads to a similar queuing interpretation for the noncausal INAR(1) process.

3.1. A Lexis diagram

We temporarily leave the laws of count variables unspecified, and focus on the deterministic relationships between these variables. We do not assume model (1) in this section.

Imagine a bar with an infinite capacity.[9] It has existed since the infinite past and will stay indefinitely open. Time is discrete and $\mathbb{Z}-$valued. This double direction is 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. In other words, we have an infinite server queue. 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 at the bar (or service time) 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 denote by $\eta_{s,t}$ the number of customers arriving at date s and leaving at date t. 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 [see e.g. [10]]:

$$
\begin{array}{cccccc}
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \eta_{t-1,t} & \eta_{t-1,t+1} & \eta_{t-1,t+2} & \eta_{t-1,t+3} & \cdots \\
 & 0 & \eta_{t,t+1} & \eta_{t,t+2} & \eta_{t,t+3} & \cdots \\
 & & 0 & \eta_{t+1,t+2} & \eta_{t+1,t+3} & \cdots \\
 & & & 0 & \cdots & \cdots
\end{array}
\tag{14}
$$

This is an infinite, upper triangular matrix, in which variable $\eta_{s,t}$ appears on the $s-$th row and the t-th column.[10] In particular, only terms strictly above the diagonal $s = t$ can be non zero due to the constraint $s < t$.

3.2. 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.

---

[9]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.

[10]The direction of the $s-$axis is southwards and the direction of the $t-$axis is eastwards.

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. The sequence $\epsilon_s(t)$, where $t \geq s$, is nonincreasing in t, and we have:

$$0 \leq \eta_{s,t} = \epsilon_s(t-1) - \epsilon_s(t), \qquad \forall t \geq s+1. \tag{15}$$

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 (14), 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 (15) can be equivalently rewritten into:

$$\epsilon_s(t) = \eta_{s,t+1} + \eta_{s,t+2} + \eta_{s,t+3} + \cdots, \qquad \forall t \geq s, \tag{16}$$

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:

$$\epsilon_s(s) := \epsilon_s = \eta_{s,s+1} + \eta_{s,s+2} + \eta_{s,s+3} + \cdots. \tag{17}$$

Equations (16) and (17) are easy to interpret using the Lexis diagram (14). Equation (17) 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 $\epsilon_s(t)$ by equation (16).

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:

$$0 \leq \eta_{s,t} = \tilde{\epsilon}_t(s+1) - \tilde{\epsilon}_t(s), \qquad \forall s \leq t-1. \tag{18}$$

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:

$$\tilde{\epsilon}_t(s) = \eta_{s-1,t} + \eta_{s-2,t} + \eta_{s-3,t} + \cdots, \qquad \forall s \leq t. \tag{19}$$

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:

$$\tilde{\epsilon}_t := \tilde{\epsilon}_t(t) = \eta_{t-1,t} + \eta_{t-2,t} + \eta_{t-3,t} + \cdots , \tag{20}$$

In the following we will denote by $\tilde{\epsilon}_s$ the RHS of the above equation, 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.

An arrival/departure duality. Let us now relate definitions (18), (19), (20) to equations (15), (16), (17). 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.[11] 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:

• 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:

TABLE 1
Correspondence between calendar/reverse time and arrival/departure.

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

Reversing the time direction is equivalent to interchanging the arrivals and departures, taking the opposite, then shifting 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 (18), (19), (20) on the one hand, and equations (15), (16), (17) on the other hand.

3.3. The current customer counts

It remains to count the current number of customers X_t. We have:

$$X_t = \epsilon_t + \epsilon_{t-1}(t) + \epsilon_{t-2}(t) + \cdots , \tag{21}$$

according to 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 $\eta_{t,t+1}$ (inclusively);

---

[11]Such a duality has previously been considered in the (continuous time) queuing literature by [13].

$\epsilon_{t-1}(t)$ is the sum of all the entries in the $(t-1)$-th row that are right to $\eta_{t-1,t+1}$ (inclusively), and so on. Thus we have:

$$X_t = \sum_{s \leq t} \sum_{t < \tau} \eta_{s,\tau}, \tag{22}$$

where the 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 $\eta_{t,t+1}$:

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

Figure 9 below illustrates the region A_t in the Lexis diagram.
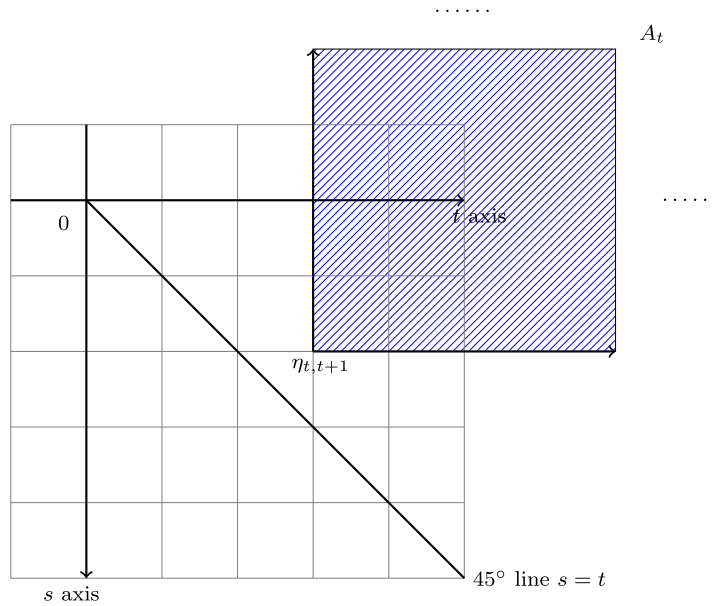
FIG 9. *The infinite rectangular area filled with northeast lines defines region A_t.*

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

$$X_t = \tilde{\epsilon}_{t+1} + \tilde{\epsilon}_{t+2}(t+1) + \tilde{\epsilon}_{t+3}(t+1) + \cdots \tag{23}$$

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

$$\tilde{\epsilon}_{t+1} + X_{t+1} = X_t + \epsilon_{t+1}, \tag{24}$$

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}.

Summary. The four count sequences $(\eta_{s,t})_{s,t}$, $(\epsilon_s(t))_{s,t}$, $(\tilde{\epsilon}_t(s))_{s,t}$, $(X_t)_t$ introduced so far are linked as follows:

- There is a one-to-one relationship between $(\eta_{s,t})_{s,t}$ and $(\epsilon_s(t))_{s,t}$ [see eq. (15), (16)].
- There is a one-to-one relationship between $(\eta_{s,t})_{s,t}$ and $(\tilde{\epsilon}_t(s))_{s,t}$ [see eq. (18), (19)].
- $(X_t)_t$ 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. (21), (23)].
- The sizes of departure cohorts $(\tilde{\epsilon}_t)$, arrival cohorts (ϵ_t) and the count of current customers (X_t) are linked through (24). 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.

3.4. 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 (21) is Markov, that is such that:

$$P(x_t|\underline{x_{t-1}}) = P(x_t|x_{t-1}), \tag{25}$$

where the symbol $P(\cdot)$ refers to the conditional probability mass function (p.m.f.), that is the conditional density with respect to the counting measure on $\mathbb{N}$. Let us define $\epsilon_\bullet(t-1)$ as the information set of population sizes of various arrival cohorts at date $t-1$:

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

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

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

Therefore, we have:

Lemma 3. *A sufficient condition for (X_t) to be Markov is that:*

$$P(x_t|\underline{\epsilon_\bullet(t-1)}) = P(x_t|x_{t-1}). \tag{26}$$

Intuitively, this lemma 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 (26) to hold. Since by (21), 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, $\epsilon_{s+1}(t+1)$. Next, we also assume that:

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

$$P_{s,t}\Big(\epsilon_s(t)|\epsilon_s(t-1), \epsilon_s(t-2), \cdots\Big) = P_{s,t}\Big(\epsilon_s(t)|\epsilon_s(t-1)\Big), \qquad \forall s, t. \qquad (27)$$

In other words the future size of an arrival cohort s at time t depends only on its current size, but not on the previous ones. We use the double subscript s, t in $P_{s,t}(\cdot|\cdots)$, since by definition the conditional p.m.f. 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, the measurable function $P_{s,t}(\cdot|\epsilon_s(t-1))$ depends on s, t and $\epsilon_s(t-1)$ only through $\epsilon_s(t-1)$ and the time lag $t - s$.

Lemma 4. *If the queuing system is such that both Assumptions 3, 4 and equation* (26) *are satisfied, then there exists $p \in]0, 1[$ such that for any $s < t$ and integer n, the conditional distribution of $\epsilon_s(t)$ given $\epsilon_s(t-1) = n$ is binomial $\mathcal{B}(n, p)$.*

Proof. See Appendix A.6. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

3.5. *Queuing interpretation of the INAR(1) process*

The following proposition is a direct consequence of Lemma 4 above. It says that INAR(1) process can be indeed embedded into the above Markov queuing system.

Proposition 6 (Arrival cohort-disaggregated representation)**.** *The solution to the causal INAR(1) model* (1) *has the disaggregate queuing representation* (21), *that is:*

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

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

1. *For a given cohort s, sequence $(\epsilon_s(t))$ is Markov and the conditional distribution of $\epsilon_s(t+1)$ given $\epsilon_s(t) = n$ is $\mathcal{B}(n, p)$ with a fixed parameter $p \in]0, 1[$. That is,*

$$\epsilon_s(t+1) = \sum_{i=1}^{\epsilon_s(t)} Z_{i,s,t+1}, \qquad \forall t \geq s, \qquad (28)$$

   where $Z_{i,s,t+1}$ are i.i.d. Bernoulli $\mathcal{B}(1, p)$ 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.7. □

Remark 1. The above representation shares similarities with the integer-valued moving average (INMA), or INARMA models [see e.g. [2, 24, 11, 5] for discrete time processes and [42] for a continuous time analog]. For instance, an INMA(n) process, where n is an integer, has the representation:

$$X_t = \epsilon_t + \sum_{i=1}^{n} \sum_{j=1}^{\epsilon_{t-i}} Z_{i,j,t}, \tag{29}$$

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

The main difference between representations (22) and (29) is that, in the former case, although each component $\epsilon_{t-i}(t)$ has the same distribution as $\sum_{j=1}^{\epsilon_{t-i}} Z_{i,j,t}$, these terms are *dependent* within the same cohort, that is when both t and i vary while keeping $t - i$ constant.

The RCINAR(1) process has a similar queuing interpretation as 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 $p_t \in]0, 1[$. This is provided in Online Appendix.

3.6. Queuing interpretation of the noncausal INAR(1) process

The queuing interpretation of the noncausal INAR(1) is obtained by using the arrival-depature duality.

Proposition 7. *The noncausal INAR(1) process has the disaggregate representation* (23), *that is:*

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

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 backward 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 $Z_{i,s,t}^*, i = 1, \dots$ are i.i.d. Bernoulli $\mathcal{B}(1, p)$, 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 p (resp. $1 - p$) that the customer was (resp. was not) already at the bar at the previous date.

This proposition is the noncausal analog of Proposition 6, and equation (23) can be viewed as the MA(∞) representation of the noncausal process, that is the analog of (21). However, the two models differ fundamentally in terms of observability. Indeed, the variables on the RHS of (23) 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.

The causal dynamics of the queuing system for a noncausal INAR(1) process In Section 2.2, we have derived general formulas to study the causal dynamics for a noncausal INAR(1) process. Let us now look at the queuing interpretation of this causal dynamics.

In the noncausal (resp. causal) INAR(1) process, the departure (resp. arrival) process $(\tilde{\epsilon}_t)$ [resp. (ϵ_t)] is i.i.d. Then the arrival (resp. departure) process is defined through equation (24). What properties possesses the arrival process? Can they also be i.i.d.? The following proposition says that this is generally not the case:

Proposition 8 (The dynamics of the arrival cohort size)**.** *In the noncausal INAR(1) process* (1)*, the arrival process* (ϵ_t) *is also i.i.d., if and only if* $\tilde{\epsilon}_t$ *is Poisson* $\mathcal{P}(\lambda)$ *distributed for some positive constant* λ*. In this case,* ϵ_t *is also Poisson* $\mathcal{P}(\lambda)$ *distributed.*

Similarly, in the causal INAR(1) process (1)*, the departure process* $(\tilde{\epsilon}_t)$ *is also i.i.d., if and only if* ϵ_t *(then also* $\tilde{\epsilon}_t$*) is Poisson* $\mathcal{P}(\lambda)$ *distributed for* $\lambda > 0$*.*

Proof. See Appendix A.8. $\square$

This result completes Thm 1 in [36], which focuses on the causal INAR(1), and says that (X_t) is truly time reversible if and only if variable (ϵ_t) is Poisson $\mathcal{P}(\lambda)$ distributed.[12] The "if" part of the proposition 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 (17) are independent and Poisson distributed with parameters $p\lambda$, $p^2\lambda$,..., respectively. Thus $\tilde{\epsilon}_t$ defined in (20) is still Poisson as the sum of independent Poisson variables.

Note that in the continuous time queuing literature, models with non i.i.d. arrival sequence have already been considered by various authors [see e.g. [12]]. Our model is fundamentally different from theirs since none of these previous models specify the queuing system in reverse time.

Besides the existence of serial correlation, what else can be said about the arrival counting process (ϵ_t) and the departure decisions of existing customers of a noncausal, non-Poissonian INAR(1) process? 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 (24). This representation is analogous to representation (1), where X_{t+1} is decom-

---

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

posed into the sum of the number of newcomers and the number of old customers who stay. By equation (24), $X_{t+1} - \epsilon_{t+1}$ is equal to $X_t - \tilde{\epsilon}_{t+1}$, which is nonnegative and non larger than X_t and is interpreted as the number of current customers who choose to stay at time t. Proposition 9 below summarizes the major differences between the calendar time dynamics of a noncausal INAR(1) and the causal INAR(1) process:

Proposition 9. *For a noncausal INAR(1) process defined in Definition 1, the following statements are equivalent:*

   i) *Process (X_t) is time reversible;*
  ii) *Process (ϵ_t) is Poisson;*
 iii) *The number of current customers who stay $X_{t+1} - \epsilon_{t+1}$ is binomial given X_t;*
 iv) *The number of arrivals 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.9. □

Thus on the contrary to a causal INAR(1) process, for which conditions $iii), iv)$ and $v)$ in Proposition 9 are all satisfied, 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[13] and is conditionally dependent of the number of staying, old customers, i.e. $X_{t+1} - \epsilon_{t+1}$. In other words, this result is the queuing version of Lemma 2.

The distribution of the arrival count Proposition 9 says that in general, the arrival cohort size (ϵ_t) cannot have Poisson stationary distribution. Let us now compute its p.g.f.:

$$
\begin{aligned}
\mathbb{E}[u^{\epsilon_{t+1}}] &= \mathbb{E}[u^{\tilde{\epsilon}_{t+1} + X_{t+1} - X_t}] && \text{(by equation (24))} \\
&= \mathbb{E}\left[u^{\sum_{i=1}^{X_{t+1}}(1 - Z_{i,t})}\right] && \text{(by equation (3))} \\
&= \mathbb{E}\left\{[(1-p)u + p]^{X_{t+1}}\right\}, && (31)
\end{aligned}
$$

by compounding the p.g.f. of the distribution of X_t and of the Bernoulli distribution $\mathcal{B}(1, 1 - p)$. Let us now derive the explicit formula for the two examples considered in Section 2.2.

Example 1 (Noncausal geometric INAR(1) cont.)**.** If (X_t) has geometric marginal distribution, i.e. $\mathbb{E}[u^{X_t}] = \frac{1}{1+\beta(1-u)}$, then equation (31) becomes:

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

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

---

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

Example 2 (Noncausal discrete stable INAR(1) cont.). Since $\mathbb{E}[u^{X_t}] = e^{-\beta\frac{(1-u)^\nu}{1-p^\nu}}$, equation (31) becomes:

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

Hence the distribution of (ϵ_t) is $\mathcal{DS}(\beta\frac{(1-p)^\nu}{1-p^\nu}, \nu)$. Since[14] $\frac{(1-p)^\nu}{1-p^\nu} \geq 1$, the stationary distribution of ϵ_t has a larger scale parameter than the distribution of $\tilde{\epsilon}_t$.

4. Conclusion

In this paper we have introduced the concept of noncausality for counting processes by time reversing a INAR(1), a Markov affine process, or a RCINAR(1). These noncausal processes are very often associated with a new, noncausal queuing interpretation. We have also seen that the resulting noncausal processes, which is generically time irreversible, can feature bubble-like calendar time dynamics characterized by some unique bi-modal conditional distribution. This approach is completely different from the current literature[15], which often focuses on the autocovariance of the process. This latter linear measure of serial dependence should be used with care, especially since a noncausal process and its causal counterpart share the same autocovariance function[16], but can have completely different nonlinear dynamics.

Our theoretical analysis of Markov noncausal processes is only the first, but a necessary step towards the empirical implementation of count models with noncausal features. While the parameter estimation of a noncausal INAR(1) process is rather straightforward (indeed, it suffices to estimate a standard causal INAR(1) model in reverse time), there are numerous other important issues that await further research. For instance, given that noncausal processes are different from their causal counterparts only when it is irreversible, can we design (possibly non-parametric) tests of reversibility on count time series data, before any meaning estimation is done? See e.g. [8] for similar tests applied to real-valued process. In a similar spirit, what is the consequence of mis-specifying the causality of the process, that is, estimating a causal INAR(1) process when the real data generating process is noncausal? See e.g. [21] for methods to distinguish between causal and noncausal models for real-valued processes. Finally, a more flexible, possible extension concerns mixed causal-noncausal INAR(1) models. Recently, a mixed causal-noncausal autoregressive process has been introduced by [16], as a natural extension of both the noncausal AR(1) model of [20] and the standard causal AR(1) model. In the context of count processes, such an extension would include the causal and noncausal INAR(1) as special cases and allow the bubble to burst gradually instead of abruptly, at a rate that is potentially different from its rate of explosion.

---

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

[15]See [27, 22] for recent developments on counting processes with flexible autocorrelation function.

[16]Indeed, the covariance operator is symmetric in the two arguments.

Appendix A: Proofs

A.1. Proof of Proposition 1

Let us follow the proof of [18], Section 5.3, which give the necessary condition for the joint and marginal p.g.f. of a general (not necessarily non-negative integer valued) time reversible affine process. More precisely, let us denote by $c(\cdot)$ the log-Laplace transform of process (X_t): $\mathbb{E}[e^{-sX_t}] = e^{c(s)}$. Then Proposition 10 of [18] shows that there exists constants β_0, β_1, and β_2 such that:

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2}c(s) = \beta_0 + \beta_1\frac{\mathrm{d}}{\mathrm{d}s}c(s) + \beta_2\Big[\frac{\mathrm{d}}{\mathrm{d}s}c(s)\Big]^2. \tag{a.32}$$

Then we have the following cases:

i) $\beta_1 = \beta_2 = 0$. Then, as in [9], we get a Gaussian process which is not a counting process.

ii) $\beta_1 \neq 0, \beta_2 = 0$. By solving the differential equation (a.32), we get: $c(s) = \delta_1 s + \delta_2(1 - e^{\delta_3 s})$. This is the log-Laplace transform of the Poisson distribution up to an affine transformation with intercept δ_1 and slope δ_3. Then by the same argument as in [9], we deduce that process (X_t) is Poisson-INAR(1).

iii) $\beta_1^2 - 4\beta_0\beta_2 = 0, \beta_2 \neq 0$. As in [9], we get a process with gamma stationary distribution.

iv) $\beta_2 \neq 0, \beta_1^2 - 4\beta_0\beta_2 > 0$. We get: $c(s) = \delta_1 s + \delta_2 \log[pe^{\delta_3 s} + 1 - p]$. Without loss of generality let us assume $\delta_3 = -1, \delta_1 = 0$. Then we distinguish two sub-cases:

- Either $\delta_2 > 0$, in which case we get the Bernoulli distribution. Thus the support of process (X_t) is not *all* the nonnegative integers.

- Or $\delta_2 < 0$, in which case we get the negative binomial distribution. Then, using the similar calculus as in [9], we get the second reversible affine counting process mentioned in the Proposition.

v) $\beta_2 \neq 0, \beta_1^2 - 4\beta_0\beta_2 < 0$. As in [9], we get a process with continuous stationary distribution.

A.2. Proof of Proposition 2

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 (11), we have:

$$\mathbb{E}[u^{X_t}] = \exp\Big[c_0 + c_1 u + \cdots + c_n u^n + O(u^{n+1})\Big] \tag{a.33}$$

$$= \exp(c_0) \sum_{j=0}^{n} \frac{1}{j!} (c_1 u + \cdots + c_n u^n)^j + O(u^{n+1}). \qquad (a.34)$$

Then we obtain recursively, for each j, the expansion of $(c_1 u + \cdots + c_n u^n)^j$ up to order n using the following Lemma due to [28]:

Lemma 5. *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:*

$$\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ \gamma_1 & 0 & 0 & \cdots & 0 \\ \gamma_2 & \gamma_1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \gamma_n & \gamma_{n-1} & \cdots & \gamma_1 & 0 \end{bmatrix}^k \begin{bmatrix} 1 \\ 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix}, \qquad \forall k = 0, ..., n. \qquad (a.35)$$

By this lemma and coefficient matching, we get the p.m.f. of the stationary distribution, that is equation (12).

Remark 2. Note that, 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:

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

Thus the updating step involves at most the computation of n coefficients.

A.3. Proof of Proposition 3

Let us first prove the proposition for $h = 2$. By [7], the tail of the discrete stable distribution satisfies:

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

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:

$$P(x_t | x_{t+1}) = \sum_{i=0}^{\min(x_t, x_{t+1})} \binom{x_{t+1}}{i} p^i (1-p)^{x_{t+1}-i} P_{\tilde{\epsilon}}(x_t - i)$$

$$= \sum_{i=0}^{x_t} \binom{x_{t+1}}{i} p^i (1-p)^{x_{t+1}-i} P_{\tilde{\epsilon}}(x_t - i)$$

$$= (1-p)^{x_{t+1}} \sum_{i=0}^{x_t} \binom{x_{t+1}}{i} \left(\frac{p}{1-p}\right)^i P_{\tilde{\epsilon}}(x_t - i), \qquad \text{(a.36)}$$

where $P_{\tilde{\epsilon}}$ is the p.m.f. of $\tilde{\epsilon}$. For fixed x_t, the term $P_{\tilde{\epsilon}}(x_t - i)$ is uniformly bounded for varying i, whereas the term $\binom{x_{t+1}}{i}$ is bounded by:

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

Thus the RHS of equation (a.36) is bounded by a polynomial of x_{t+1} of degree x_t, and hence $P(x_t|x_{t+1}) = O\big[(1-p)^{x_{t+1}} \frac{x_{t+1}^{x_t}}{x_t!}\big]$, when x_{t+1} goes to infinity. Thus by the Bayes' formula:

$$P(x_{t+1}|x_t) = \frac{P(x_t|x_{t+1})P(x_{t+1})}{P(x_t)} = O\Big[\frac{(1-p)^{x_{t+1}}}{x_{t+1}^{\nu+1-x_t}}\Big].$$

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:

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

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 p^2. Thus the above result for $h = 1$ still applies to the case $h = 2$.

A.4. Proof of Proposition 4

Let us focus on the proof for $h = 1$. The formula for larger h is easily deduced by the iterative expectation theorem. The joint p.g.f. of (X_t, X_{t+1}) is:

$$\begin{aligned}
\mathcal{L}(u,v) &:= \mathbb{E}[u^{X_t} v^{X_{t+1}}] \\
&= \mathbb{E}\Big[\mathbb{E}[u^{X_t}|X_{t+1}] v^{X_{t+1}}\Big] \\
&= \exp\Big(-\beta(1-u)^\nu - \beta \frac{[1-(pu+1-p)v]^\nu}{1-p^\nu}\Big), \quad \forall u,v \in [0,1[. \quad \text{(a.37)}
\end{aligned}$$

Then we get, for all u between 0 and 1:

$$\begin{aligned}
\mathbb{E}[X_{t+1}u^{X_t}] &= \frac{\partial \mathcal{L}(u,v)}{\partial v}\Big|_{v=1} \\
&= e^{-\beta \frac{(1-u)^\nu}{1-p^\nu}} \frac{\beta\nu}{1-p^\nu} (pu+1-p)p^{\nu-1}(1-u)^{\nu-1}. \qquad \text{(a.38)}
\end{aligned}$$

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

$$\frac{\partial \mathcal{L}(u)}{\partial u} = \mathbb{E}[X_t u^{X_t-1}],$$

$$\text{or } \mathbb{E}[p^{\nu-1} X_t u^{X_t}] = p^{\nu-1} u \frac{\partial \mathcal{L}(u)}{\partial u} = e^{-\beta \frac{(1-u)^\nu}{1-p^\nu}} \frac{\beta \nu}{1-p^\nu} p^{\nu-1}(1-u)^{\nu-1}u. \quad \text{(a.39)}$$

On the other hand, we can expand the LHS of (a.38) and (a.39) and get:

$$\mathbb{E}[X_{t+1} u^{X_t}] = \sum_{i=0}^{\infty} u^i \mathbb{E}[X_{t+1}|X_t = i]\mathbb{P}[X_t = i], \quad \text{(a.40)}$$

$$\mathbb{E}[p^{\nu-1} X_t u^{X_t}] = \sum_{i=0}^{\infty} u^i p^{\nu-1} i \mathbb{P}[X_t = i]. \quad \text{(a.41)}$$

Thus, to show that $\mathbb{E}[X_{t+1}|X_t = i]$ and $p^{\nu-1}i$ 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.40) and (a.41). Let us expand:

$$e^{-\beta \frac{(1-u)^\nu}{1-p^\nu}} \frac{\beta \nu}{1-p^\nu} p^{\nu-1}(1-u)^{\nu-1} = \sum_{i=0}^{\infty} c_i u^i. \quad \text{(a.42)}$$

If we multiply the LHS of this above equation times by $(pu + 1 - p)$, we get the RHS of (a.38), and, if we multiply it by u, we get the RHS of (a.39). Thus the i−th coefficient of the expansion of $\mathbb{E}[X_{t+1} u^{X_t}]$ is $pc_{i-1} + (1-p)c_i$, whereas that of $\mathbb{E}[p^{\nu-1} X_t u^{X_t}]$ 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.42) is equal to 1.

A.5. Proof of Proposition 5

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.37), the Laplace transform of this vector is such that:

$$\Big\{ \mathbb{E}[e^{-\frac{u}{a_n} X_t - \frac{v}{a_n} X_{t+1}}] \Big\}^n$$

$$= \exp\left\{ n\beta(1 - e^{-u/a_n})^\nu - \frac{n\beta}{1 - p^\nu}[1 - (pe^{-u/a_n} + 1 - p)e^{-v/a_n}] \right\}, \quad \forall u, v.$$

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

$$\exp\left(-\beta u^\nu - \frac{\beta}{1 - p^\nu}(pu + v)^\nu \right). \tag{a.43}$$

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.43) [see e.g. [34]]. The spectral measure of the limiting distribution has two point masses at $(1, 0)$ and $(\frac{p}{\sqrt{1+p^2}}, \frac{1}{\sqrt{1+p^2}})$.[17] Then we apply Thm 4 of [34] and conclude that process (X_t) satisfies Proposition 5.

A.6. Proof of Lemma 4

We have:

$$\mathbb{E}[u^{X_t}|\underline{\epsilon_\bullet(t-1)}] = \mathbb{E}[u^{\epsilon_t + \epsilon_{t-1}(t) + \epsilon_{t-2}(t) + \cdots}|\epsilon_\bullet(t-1)]$$

$$= \mathbb{E}[u^{\epsilon_t}] \prod_{s=-\infty}^{t-1} \mathbb{E}[u^{\epsilon_s(t)}|\epsilon_s(t-1)], \tag{a.44}$$

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 conditional p.g.f.'s 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 (21), these conditioning variables should also sum up to X_{t-1}. Thus, in order for the conditional p.m.f. $P(X_t|\underline{\epsilon_\bullet(t-1)})$ to depend on $\underline{\epsilon_\bullet(t-1)}$ through X_{t-1} only, and 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 $s \le t - 1$. 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 further constrained:

---

[17]Alternatively, we recognize that the limit (a.43) is the Laplace transform of (Y_t, Y_{t+1}), where (Y_t) is a linear noncausal alpha-stable AR(1) process defined by:

$$Y_t = pY_{t+1} + \xi_t,$$

where $\xi'_t s$ are i.i.d., and follow alpha-stable distribution, with Laplace transform $\mathbb{E}[e^{-u\xi_t}] = \exp(-\beta u^\nu)$, for all $u > 0$. By [15], Thm 2.2, this latter process satisfies the proposition.

- First by choosing $\epsilon_s(t-1) = 0$ (in this case $\epsilon_s(t)$ is almost surely zero as well), we get $g_2(u, t-s) = 1$.
- Then we choose $\epsilon_s(t-1) = 1$, and conclude that $g_1(\cdot)$ 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 $\epsilon_s(t)$ is bounded by $\epsilon_s(t-1)$, function $g_1(\cdot)$ must be the p.g.f. of a Bernoulli distribution, say, $\mathcal{B}(1, p)$, i.e. $g_1(u) = pu + 1 - p$.

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

A.7. Proof of Proposition 6

By Lemma 4, equation (a.44) becomes:

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

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

A.8. Proof of Proposition 8

First, by the arrival-departure duality, it suffices to prove the proposition for the causal INAR(1) process. It suffices to show that $\tilde{\epsilon}_t$ and $\tilde{\epsilon}_{t+1}$ are independent, if and only if ϵ_t is Poisson distributed[18]. Let us compute their joint p.g.f.. We have:

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

---

[18]Indeed, Theorem 4.2 in [38] shows that if ϵ_t is Poisson, i.e. process (X_t) is Poisson-INAR(1) process, then the departure process $\tilde{\epsilon}_t$ is i.i.d.

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 $x = p^2 + p(1-p)v + 1 - p$ and $y = p + (1-p)u$. 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 positive[19] 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:

$$\mathbb{E}[u^{\epsilon_t}] = \frac{\mathbb{E}[u^{X_t}]}{\mathbb{E}[(pu + 1 - p)^{X_t}]} = e^{(1-p)\lambda(u-1)}, \forall u \in [0, 1 + \frac{1}{p}[$$

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

A.9. Proof of Proposition 9

The equivalence between $i)$ and $ii)$ is a consequence of [36].

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 (3). Then using a characterization theorem of the Poisson distribution [see the Theorem in [32]], $X_{t+1} - \epsilon_{t+1} = X_t - \tilde{\epsilon}_{t+1}$ 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:

$$\epsilon_{t+1} = \tilde{\epsilon}_{t+1} + X_{t+1} - X_t = X_{t+1} - \sum_{i=1}^{X_{t+1}} Z_{i,t}^*, \tag{a.45}$$

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:

$$\mathbb{E}\left[u^{X_{t+1} - \sum_{i=1}^{X_{t+1}} Z_{i,t}^*} v^{\sum_{i=1}^{X_{t+1}} Z_{i,t}^*}\right] = \mathbb{E}\left[[pv + (1-p)u]^{X_t}\right].$$

By a similar argument as in Proposition 8, this latter p.g.f. is multiplicatively separable if and only if the p.g.f. of X_t is Poisson.

---

[19]In the limiting case $\lambda = 0$, we have $\epsilon_t = 0$ almost surely, thus the only stationary solution of the model is the constant process 0.

Finally, let us prove the equivalence between conditions i) and iv) by using the calendar time conditional p.m.f. [eq. (5)]. $X_{t+1} - \epsilon_{t+1}$ and ϵ_{t+1} are independent, if and only if their joint p.g.f. is separable. Using equation (a.45), we can easily check that this p.g.f. is equal to:

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

By the same argument as in Appendix A.3, this p.g.f. is separable if and only if X_{t+1} is Poisson.

Online Appendix: Noncausal RCINAR(1) process

We have the following characterization concerning the reversibility of RCI-NAR(1) processes:

Proposition 10 (Characterization of time reversible RCINAR(1) processes)**.** *If the sequence (p_t) is i.i.d. and independent of (ϵ_t), then the RCINAR(1) process is time reversible if and only if (p_t) is $\mathcal{B}eta(\alpha, \beta)$ distributed, for some $\alpha, \beta \in (0, \infty)$, and (ϵ_t) is negative binomial $\mathcal{NB}(\beta, \theta)$ distributed with p.g.f. $\mathbb{E}[u^{\epsilon_t}] = \frac{(1-\theta)^\beta}{(1-\theta u)^\beta}$, where $\theta \in]0, 1[$.*

Proof. See Appendix OA.1. $\square$

When $(\tilde{\epsilon}_t)$ is $\mathcal{NB}(\beta, \theta)$, the marginal distribution of (X_t) is $\mathcal{NB}(\alpha + \beta, \theta)$, and the departure process $\tilde{\epsilon}_t$ is also i.i.d. $\mathcal{NB}(\beta, \theta)$. This process is introduced by [23] and applied to real data by [40]. If we let $\alpha + \beta$ go to infinity, while at the same time keeping the ratio $\frac{\alpha}{\alpha+\beta} = p_0$ and the mean of the $\mathcal{NB}(\beta, \theta)$ distribution $\beta \frac{\theta}{1-\theta} = \lambda_0$ constant, then in the limiting case the beta (resp. NB) distribution tends to the point mass at p_0 [resp. Poisson $\mathcal{P}(\lambda_0)$ distribution]. Hence we recover the Poisson-INAR(1) process as the limit.

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

Proposition 11 (A queuing interpretation of the RCINAR(1) model)**.** *The solution to the causal, RCINAR(1) model has the disaggregate queuing representation (21), where the joint distribution of $(\epsilon_s(t))$ is as follows:*

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 $p_t \in]0, 1[$, where (p_t) is i.i.d. for t varying.*
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 p_t.*

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

The proof is obvious and omitted.

Finally, the queuing interpretation of the RCINAR(1) process is obtained in a similar way, by reversing the roles of arrival and departure processes.

OA.1. Proof of Proposition 10

We follow the proof of [36] 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:

$$
\begin{aligned}
P(i|j) : &= \mathbb{P}[X_{t+1} = i | X_t = j] \\
&= \sum_{k=0}^{\min(i,j)} \binom{j}{k} P_\epsilon(i-k) \mathbb{E}[p_t^k (1-p_t)^{i-k}], \qquad \forall i, j \in \mathbb{N}, \qquad (a.46)
\end{aligned}
$$

where function P_ϵ is the p.m.f. of ϵ_t. Under reversibility, we have, for each $i \in \mathbb{N}$:

$$
\mathbb{P}[(X_t, X_{t+1}, X_{t+2}, X_{t+3}) = (0, 1, i, 0)] = \mathbb{P}[(X_t, X_{t+1}, X_{t+2}, X_{t+3}) = (0, i, 1, 0)], \tag{a.47}
$$

or equivalently[20],

$$
P(1|0)P(i|1)P(0|i) = P(i|0)P(1|i)P(0|1). \tag{a.48}
$$

Equation (a.46) leads to:

$$
\begin{aligned}
P(i|0) &= P_\epsilon(i), \\
P(0|i) &= P_\epsilon(0) \mathbb{E}[(1-p_t)^i], \\
P(i|1) &= P_\epsilon(i) \mathbb{E}[(1-p_t)] + P_\epsilon(i-1) \mathbb{E}[p_t], \\
P(1|i) &= P_\epsilon(1) \mathbb{E}[(1-p_t)^i] + i P_\epsilon(0) \mathbb{E}[p_t (1-p_t)^{i-1}].
\end{aligned}
$$

We plug in these expressions into equation (a.48) to get:

$$
\begin{aligned}
&P_\epsilon(1) \Big\{ P_\epsilon(i) \mathbb{E}[(1-p_t)] + P_\epsilon(i-1) \mathbb{E}[p_t] \Big\} P_\epsilon(0) \mathbb{E}[(1-p_t)^i] \\
&= p_\epsilon(i) \Big\{ P_\epsilon(1) \mathbb{E}[(1-p_t)^i] + i P_\epsilon(0) \mathbb{E}[p_t (1-p_t)^{i-1}] \Big\} P_\epsilon(0) \mathbb{E}[(1-p_t)],
\end{aligned}
$$

or equivalently:

$$
P_\epsilon(i) = P_\epsilon(i-1) \frac{P_\epsilon(1) \mathbb{E}[p_t] \mathbb{E}[(1-p_t)^i]}{i P_\epsilon(0) \mathbb{E}[p_t (1-p_t)^{i-1}] \mathbb{E}[(1-p_t)]}. \tag{a.49}
$$

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

$$
P(2|0)P(i|2)P(0|i) = P(i|0)P(2|i)P(0|2), \tag{a.50}
$$

---

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

or equivalently:

$$P_\epsilon(2)P_\epsilon(0)\mathbb{E}[(1-p_t)^i]\Big\{P_\epsilon(i)\mathbb{E}[(1-p_t)^2] + 2P_\epsilon(i-1)\mathbb{E}[p_t(1-p_t)]$$

$$+ P_\epsilon(i-2)\mathbb{E}[p_t^2]\Big\}$$

$$= P_\epsilon(i)P_\epsilon(0)\mathbb{E}[(1-p_t)^2]\Big\{P_\epsilon(2)\mathbb{E}[(1-p_t)^i] + iP_\epsilon(1)\mathbb{E}[p_t(1-p_t)^{i-1}]$$

$$+ \frac{i(i-1)}{2}P_\epsilon(0)\mathbb{E}[p_t^2(1-p_t)^{i-2}]\Big\}.$$

Let us now expand the parentheses on both sides and check that the first terms (resp. the second terms) on both sides exactly cancel out due to equation (a.49). Then the equality between the third terms shows that:

$$\frac{\mathbb{E}[(1-p_t)^{i-1}]\mathbb{E}[p_t^2(1-p_t)^{i-2}]}{\mathbb{E}[p_t(1-p_t)^{i-1}]\mathbb{E}[p_t(1-p_t)^{i-2}]} \qquad (a.51)$$

is constant when i varies. Let us denote:

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

Then we have:

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

and the term in equation (a.51) 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 $1 - p_t$ is geometric. Since this sequence of moments characterizes a distribution on $[0,1]$, we conclude that p_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:

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

where $\alpha = \frac{1}{c}, \beta = \frac{d}{c} \in [0,\infty[$, or by iteration $m_i = \frac{B(\alpha,\beta+i)}{B(\alpha,\beta)}$. In other words the sequence of integer moments of $1-p_t$ coincides with those of a beta distribution $\mathcal{B}eta(\beta,\alpha)$. Thus p_t follows $\mathcal{B}eta(\alpha,\beta)$ distribution. Let us finally get back to equation (a.49) and show that, in this case, the distribution of (ϵ_t) is negative binomial. Summing up this equation for $i = 1, 2, \dots$ yields:

$$\beta\mu = (\beta + \mu)\frac{P_\epsilon(1)}{P_\epsilon(0)}, \qquad (a.52)$$

where μ is the expectation of ϵ_t. Thus equation (a.49) becomes:

$$P_\epsilon(i) = P_\epsilon(i-1)\frac{\beta+i-1}{\beta i}\frac{P_\epsilon(1)}{P_\epsilon(0)} = P_\epsilon(i-1)\frac{\beta+i-1}{i}\frac{\mu}{\mu+\beta}, \qquad \forall i \geq 1. \tag{a.53}$$

By iteration we deduce that (ϵ_t) follows a NB distribution with stopping parameter β and mean μ.

Acknowledgments

References

[1] 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. MR0903755

[2] Al-Osh, M. and Alzaid, A. A. (1988). Integer-Valued Moving Average (INMA) Process. *Statistical Papers*, 29(1):281–300. MR1000749

[3] Aly, E.-E. A. and Bouzar, N. (1994). Explicit Stationary Distributions for some Galton-Watson Processes with Immigration. *Stochastic Models*, 10(2):499–517. MR1268562

[4] Barreto-Souza, W. (2019). Mixed Poisson INAR (1) Processes. *Statistical Papers*, 60(6):2119–2139. MR4027638

[5] 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.

[6] 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. MR1404684

[7] Christoph, G. and Schreiber, K. (1998). Discrete Stable Random Variables. *Statistics & Probability Letters*, 37(3):243–247. MR1614930

[8] Darolles, S., Florens, J.-P., and Gourieroux, C. (2004). Kernel-Based Nonlinear Canonical Analysis and Time Reversibility. *Journal of Econometrics*, 119(2):323–353. MR2057103

[9] Darolles, S., Gourieroux, C., and Jasiak, J. (2006). Structural Laplace Transform and Compound Autoregressive Models. *Journal of Time Series Analysis*, 27(4):477–503. MR2245710

[10] Demeny, P., McNicoll, G., and Hodgson, D. (2003). *Encyclopedia of Population*. Hodgson, Dennis (2003). Contemporary Population Thought.

[11] 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. MR2488633

[12] 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.

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

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

[15] Fries, S. (2019). Conditional Moments of Anticipative α-Stable Markov Processes and the Prediction of Bubble Crash Odds. *arXiv preprint arXiv:1805.05397, revised Dec. 2019*.

[16] Fries, S. and Zakoian, J.-M. (2019). Mixed Causal-Noncausal AR Processes and the Modelling of Explosive Bubbles. *Econometric Theory*, 35(6):1–37. MR4028969

[17] Gouriéroux, C., Hencic, A., and Jasiak, J. (2021). Forecast Performance and Bubble Analysis in Noncausal MAR (1, 1) Processes. *Journal of Forecasting*, 40(2):301–326. MR4212339

[18] Gouriéroux, C. and Lu, Y. (2019a). Negative Binomial Autoregressive Process with Stochastic Intensity. *Journal of Time Series Analysis*, 40(2):225–247. MR3915528

[19] Gouriéroux, C. and Lu, Y. (2019b). Noncausal Affine Processes with Application to Derivative Pricing. *CREST DP, URL=crest. science/RePEc/ wpstorage/ 2019-02. pdf*.

[20] 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. MR3641405

[21] Hecq, A. and Sun, L. (2021). Selecting Between Causal and Noncausal Models with Quantile Autoregressions. *forthcoming Studies in Nonlinear Dynamics & Econometrics*.

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

[23] Joe, H. (1996). Time Series Models with Univariate Margins in the Convolution-Closed Infinitely Divisible Class. *Journal of Applied Probability*, 33(3):664–677. MR1401464

[24] 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. MR1622447

[25] 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. MR1938173

[26] Lanne, M. and Saikkonen, P. (2013). Noncausal Vector Autoregression. *Econometric Theory*, 29(3):447–481. MR3064046

[27] 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. MR3773399

[28] Lu, Y. (2021). The Predictive Distributions of Thinning-Based Count Pro-

cesses. *Scandinavian Journal of Statistics*, 48(1):42–67. MR4233164

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

[30] McKenzie, E. (1985). Some Simple Models for Discrete Variate Time Series. *Journal of the American Water Resources Association*, 21(4):645–650. MR1973555

[31] McKenzie, E. (1986). Autoregressive Moving-Average Processes with Negative-Binomial and Geometric Marginal Distributions. *Advances in Applied Probability*, 18(3):679–705. MR0857325

[32] Moran, P. (1952). A Characteristic Property of the Poisson Distribution. *Mathematical Proceedings of the Cambridge Philosophical Society*, 48(1):206–207. MR0045968

[33] Pickands, I. J. and Stine, R. A. (1997). Estimation for an $M/G/\infty$ Queue with Incomplete Information. *Biometrika*, 84(2):295–308. MR1467048

[34] Resnick, S. and Greenwood, P. (1979). A Bivariate Stable Characterization and Domains of Attraction. *Journal of Multivariate Analysis*, 9(2):206–221. MR0538402

[35] Ristić, M. M., Bakouch, H. S., and Nastić, A. S. (2009). A New Geometric First-order Integer-valued Autoregressive (NGINAR (1)) Process. *Journal of Statistical Planning and Inference*, 139(7):2218–2226. MR2507983

[36] 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. MR3336436

[37] 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. MR3210062

[38] Schweer, S. and Wichelhaus, C. (2015). Queueing Systems of INAR (1) Processes with Compound Poisson Arrivals. *Stochastic Models*, 31(4):618–635. MR3411347

[39] Steutel, F. and Van Harn, K. (1979). Discrete Analogues of Self-Decomposability and Stability. *Annals of Probability*, 7(5):893–899. MR0542141

[40] Weiss, C. H. (2008). Thinning Operations for Modeling Time Series of Counts-A Survey. *AStA Advances in Statistical Analysis*, 92(3):319. MR2426093

[41] Weiss, C. H. (2010). INARCH (1) Processes: Higher-Order Moments and Jumps. *Statistics & Probability Letters*, 80(23-24):1771–1780. MR2734241

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

[43] Willmot, G. E. (1987). The Poisson-Inverse Gaussian Distribution as an Alternative to the Negative Binomial. *Scandinavian Actuarial Journal*, 1987(3-4):113–127. MR0943576

[44] 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. MR2292852

[45] 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. MR2547940