HAL
open science

Bivariate integer-autoregressive process with an application to mutual fund flows

Serge Darolles, Gaëlle Le Fol, Yang Lu, Ran Sun

To cite this version:

Serge Darolles, Gaëlle Le Fol, Yang Lu, Ran Sun. Bivariate integer-autoregressive process with an application to mutual fund flows. Journal of Multivariate Analysis, 2019, 173, pp.181-203. 10.1016/j.jmva.2019.02.015 . halshs-02418967

HAL Id: halshs-02418967

https://shs.hal.science/halshs-02418967v1

Submitted on 20 Dec 2019

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

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

Bivariate INAR Processes with an Application to Mutual Fund Flows

Serge Darolles[*], Gaëlle Le Fol[†], Yang Lu[‡], and Ran Sun[§]

February 4, 2019

Abstract: We propose a new family of bivariate nonnegative integer-autoregressive (BINAR) models for count process data. We first generalize the existing BINAR(1) model by allowing for dependent thinning operators and arbitrary innovation distribution. The extended family allows for intuitive interpretation, as well as tractable aggregation and stationarity properties. We then introduce higher order BINAR(p) and BINAR(∞) dynamics to accommodate more flexible serial dependence patterns. So far, the literature has regarded such models to be computationally intractable. Our contribution is to show that the extended BINAR family allows for closed-form predictive distributions at any horizons and for any values of p, which significantly facilitates non-linear forecasting and likelihood based estimation. Finally, a BINAR(∞) model with memory persistence is applied to open-ended mutual fund purchase and redemption order counts.

Keywords: Multivariate low event count process, non-linear forecasting, memory persistence, liquidity risk, mutual funds. MSC code: 62-15, JEL code: C32, C53

Acknowledgement We thank an associate editor as well as two referees for insightful comments.

[*]University of Paris Dauphine, Department of Finance.

[†]University of Paris Dauphine, Department of Finance.

[‡]Corresponding author. University of Paris 13, Department of Economics (CEPN). Email: luyang000278@gmail.com

[§]University of Paris Dauphine, Department of Finance.

1 Introduction

Nonnegative low count processes has been widely used in domains such as marketing [Böckenholt (1998)], economics [Blundell et al. (1999)], finance [Heinen and Rengifo (2007)], insurance [Gouriéroux and Jasiak (2004)] and beyond, ever since the seminal work of McKenzie (1985). Our interest in this paper lies in the monitoring of the liquidity risk of an open-ended mutual fund (MF). A MF channels investors' cash investment into less liquid assets, and is thus structurally vulnerable to liquidity risk. This risk has recently received much attention of the regulators [see, e.g., Securities and Exchange Commission (2015); Darolles (2018)], but its quantification and management remains difficult. Indeed, from the modelling point of view, the liquidity risk is quite different from traditional market risks in that they involve the daily counts of redemption and purchase orders, which are *i*) most of the time low integers or zero, but also have a non-null probability of taking mildly large values; *ii*) both cross-sectionally and serially dependent, with significant heteroscedasticity. Recently, the MF industry has started to record purchase and redemption order count data separately. This allows to distinguish auto-correlation effects and cross-effects between the two count processes, which have different economic interpretations. For instance, the clustering of the redemption counts corresponds to fund run, whereas a fund manager usually reacts to past redemptions by seeking new investors in order to stabilize the fund size, leading to a positive feedback effect between past redemption and current purchase counts. Therefore, a bivariate count analysis can be of great interest to understand clients' behaviour and the manager's reaction to exogenous liquidity shock.

Yet the literature on bivariate count processes is still in its infancy. The benchmark approach is the Bivariate INteger-valued AutoRegressive (BINAR) model [Latour (1997); Pedeli and Karlis (2013b)], which assumes that for each t:

$$\begin{aligned} X_{1,t} &= \alpha_{11} \circ X_{1,t-1} + \alpha_{12} \circ X_{2,t-1} + \epsilon_{1,t}, \\ X_{2,t} &= \alpha_{21} \circ X_{1,t-1} + \alpha_{22} \circ X_{2,t-1} + \epsilon_{2,t}, \end{aligned} \tag{1.1}$$

where given $\boldsymbol{X}_{t-1} = (X_{1,t-1}, X_{2,t-1})^\top$, the binomial thinning operators are defined as follows: for each $i, j = 1, 2$, variable $\alpha_{i,j} \circ X_{j,t-1}$ follows the binomial distribution with size $X_{j,t-1}$ and success probability $\alpha_{i,j} \in [0, 1]$. Moreover these variables are conditionally independent, and are

also independent of the i.i.d. innovation sequence $\boldsymbol{\epsilon}_t = (\epsilon_{1,t}, \epsilon_{2,t})^\top$. This approach has several drawbacks. First, the conditional independence assumption between the thinning operators restricts significantly the dependence feature. Secondly, so far only Latour (1997) has considered higher-order models, but the author suggests to base the estimation and forecasting solely on conditional expectation, that is as if the observations are continuous, and nothing is said about estimation. This is due to the fact that the term structure of predictive distributions of higher-order BINAR process has yet to be derived and is so far (wrongly) considered intractable.These downsides seriously limit their usefulness for risk management and forecasting purpose.

Besides BINAR processes, other non-thinning-based models have also been introduced. Quoreshi (2017); Livsey et al. (2018) propose parameter-driven models with flexible (auto-)correlation, but the likelihood estimation and forecasting in these models are way too cumbersome to be feasible. Another popular approach is the bivariate INGARCH model [see, e.g., Liu (2012)], which assumes that given the past, $X_{1,t}$ and $X_{2,t}$ follow some simple (say, Poisson) distributions, with parameters that are exponentially weighted average of past observations. Then the contemporaneous conditional dependence between $X_{1,t}$ and $X_{2,t}$ is captured by a copula [Heinen and Rengifo (2007); Bien et al. (2011); Doukhan et al. (2017)]. The downsides of this latter approach are that *i*) it is sometimes unclear whether such processes are strictly stationary and how memory persistence can be introduced; *ii*) forecasting formulas necessitate numerical integration, except for the conditional expectation *iii*) it remains an open question as to whether the copula function is identified in the count data setting [see Genest and Nešlehová (2007); Trivedi and Zimmer (2017)]; Recently, some alternative, non-INGARCH models are proposed by Scotto et al. (2014); Cui and Zhu (2018); Gouriéroux and Lu (2018), which allow for rather flexible conditional correlation. However it is computationally difficult to extend them to models of arbitrary orders (The model of Cui and Zhu (2018), as well as INGARCH models, are necessarily infinite order Markov, whereas that of Gouriéroux and Lu (2018) is first order Markov). Moreover, in these models, forecasting formulas are complicated either when the forecast horizon h is large. Further, Cui and Zhu (2018); Gouriéroux and Lu (2018) assume *ex ante* the conditional distributions of X_{1t} and X_{2t} given the past to be equi- (resp. over-) dispersed. The paper that is closest to ours is Scotto et al. (2014). While they are mainly interested in bounded counts, they mention in their conclusion some possible extensions of model (1.1), which they conjecture to be appropriate for

unbounded count data. In this paper we show that one of these extensions has indeed tractable properties, even after extension to higher-order cases.

More precisely, we contribute to the BINAR literature in two ways. First, we extend model (1.1) (called independent BINAR(1) henceforth) by introducing (positively or negatively) dependent thinning operators and arbitrary innovation distribution. We also show that the process belongs to the compound autoregressive (CaR) family, and possesses intuitive aggregation and stationarity properties. More importantly, we clarify that in this large family of models, the predictive distributions at various horizons are easily computable via a matrix-based algorithm. This largely facilitates likelihood-based inference and non-linear forecasting, especially when it comes to the prediction of extreme events. Second, we extend our model to higher-order dependent BINAR(p) and BINAR(∞) processes, which can better capture slowly decaying serial correlation patterns.

The paper is organized as follows. The dependent BINAR(1) model is introduced in Section 2 and extended in Section 3 to higher-order BINAR(p) and BINAR(∞) models. Section 4 computes the predictive distributions. Section 5 applies the model to forecast the counts of share purchase and redemption of a MF. Section 6 concludes. Proofs and technical details are gathered in Appendices.

2 Dependent BINAR(1) process

2.1 The dynamic specification

Definition 1. We say that the bivariate count process $(\boldsymbol{X}_t) = (X_{1t}, X_{2t})^\top$ with domain $\mathbb{N}^2_{\geq 0}$ is dependent BINAR(1) if it has the stochastic representation:

$$\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \sum_{i=1}^{X_{1,t-1}} \begin{pmatrix} Z_{1,i,t} \\ Z_{2,i,t} \end{pmatrix} + \sum_{j=1}^{X_{2,t-1}} \begin{pmatrix} Z_{3,j,t} \\ Z_{4,j,t} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}, \qquad \forall t, \tag{2.1}$$

where given $\boldsymbol{X}_{t-1}$:

- for i, j, t varying, random vectors $(Z_{1,i,t}, Z_{2,i,t})^\top$, $(Z_{3,j,t}, Z_{4,j,t})^\top$, $(\epsilon_{1,t}, \epsilon_{2,t})^\top$ are mutually independent copies of $(Z_1, Z_2)^\top$, $(Z_3, Z_4)^\top$, and $(\epsilon_1, \epsilon_2)^\top$, respectively.

- Couples $(Z_1, Z_2)^\top$ and $(Z_3, Z_4)^\top$ are independent and bivariate Bernoulli distributed. That is, they have Bernoulli marginal distribution with success probability parameters given by:

$$\begin{pmatrix} \mathbb{P}[Z_1 = 1], & \mathbb{P}[Z_3 = 1] \\ \mathbb{P}[Z_2 = 1], & \mathbb{P}[Z_4 = 1] \end{pmatrix} = \begin{pmatrix} \alpha_{11}, & \alpha_{12} \\ \alpha_{21}, & \alpha_{22} \end{pmatrix} := A,$$

  where all the entries of matrix A are nonnegative, whereas the two joint probabilities are respectively:

$$\mathbb{P}[Z_1 = Z_2 = 1] = q_1, \qquad \mathbb{P}[Z_3 = Z_4 = 1] = q_2.$$

- Innovation $\boldsymbol{\epsilon}_t$ is independent of $\boldsymbol{X}_{t-1}, \boldsymbol{X}_{t-2}, ...,$ and is i.i.d. across t. Moreover, it is nonnegative and has finite variance.

The bivariate Bernoulli distribution is first introduced by Wicksell (1916) [see also Marshall and Olkin (1985)], and is recently used by Scotto et al. (2014) to model bounded count processes.In order for it to be well defined, its parameters have to satisfy the following constraint:

Lemma 1 (Joe (1997), page 210)**.** *Parameters* $(\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}, q_1, q_2)$ *of the bivariate Bernoulli distribution can take any values satisfying the following two inequalities:*

$$\max(\alpha_{11} + \alpha_{21} - 1, 0) \leq q_1 \leq \min(\alpha_{11}, \alpha_{21}), \tag{2.2}$$

$$\max(\alpha_{12} + \alpha_{22} - 1, 0) \leq q_2 \leq \min(\alpha_{12}, \alpha_{22}). \tag{2.3}$$

The covariance between Z_1 and Z_2 (resp. Z_3 and Z_4) is $q_1 - \alpha_{11}\alpha_{21}$. In particular, if $q_1 = \max(\alpha_{11} + \alpha_{21} - 1, 0)$, then the covariance is nonpositive; if $q_1 = \min(\alpha_{11}, \alpha_{21})$, the covariance is nonnegative; if $q_1 = \alpha_{11}\alpha_{21}$ and $q_2 = \alpha_{12}\alpha_{22}$, we get the independent BINAR(1) model (1.1).

Birth-death-immigration interpretation. Let us temporarily assume that $X_{1,t}$ and $X_{2,t}$ count individuals of type 1 and 2 at time t, respectively. A type j_1 ($j_1 = 1, 2$) individual produces a type j_2 ($j_2 = 1, 2$) off-spring with marginal probability α_{j_1,j_2}, and joint probability q_{j_1}. Then the population of type j ($j = 1, 2$) at the next period $t + 1$ is composed of such off-springs, plus

$\epsilon_{j,t+1}$ immigrants of type j. In particular, if $q_1 = \alpha_{11}\alpha_{21}$ and $q_2 = \alpha_{12}\alpha_{22}$, the productions of the two types of off-springs are independent events. If instead $q_1 = q_2 = 0$, the productions are mutually exclusive, i.e., each individual can only produce up to one off-spring.

2.2 Conditional distribution

2.2.1 First two conditional moments

Since we have:

$$\mathbb{E}[\boldsymbol{X}_t|\boldsymbol{X}_{t-1}] = A\boldsymbol{X}_{t-1} + \mathbb{E}[\boldsymbol{\epsilon}_t], \tag{2.4}$$

we abbreviate model (2.1) into:

$$\boldsymbol{X}_t = A(q_1, q_2) \circ \boldsymbol{X}_{t-1} + \boldsymbol{\epsilon}_t, \tag{2.5}$$

where the operator $A(q_1, q_2)\circ$ is linear and will be called the dependent matrix thinning operator.

The conditional covariance matrix is:

$$\begin{pmatrix} \mathbb{V}[X_{1,t}|\boldsymbol{X}_{t-1}], & * \\ \mathbb{C}ov[X_{1,t}, X_{2,t}|\boldsymbol{X}_{t-1}], & \mathbb{V}[X_{2,t}|\boldsymbol{X}_{t-1}] \end{pmatrix} = \gamma_\epsilon + X_{1,t-1}\gamma_{12} + X_{2,t-1}\gamma_{34}, \tag{2.6}$$

where γ_ϵ, γ_{12}, γ_{34} are the covariance matrices of $(\epsilon_{1t}, \epsilon_{2t})$, $(Z_{1,t}, Z_{2,t})$ and $(Z_{3,t}, Z_{4,t})$, respectively:

$$\gamma_{12} = \begin{pmatrix} \alpha_{11}(1-\alpha_{11}), & * \\ q_1 - \alpha_{11}\alpha_{21}, & \alpha_{21}(1-\alpha_{21}) \end{pmatrix}, \qquad \gamma_{34} = \begin{pmatrix} \alpha_{12}(1-\alpha_{12}), & * \\ q_2 - \alpha_{12}\alpha_{22}, & \alpha_{22}(1-\alpha_{22}) \end{pmatrix}, \tag{2.7}$$

and the conditional correlation between $X_{1,t}$ and $X_{2,t}$ is:

$$\mathrm{corr}[X_{1,t}, X_{2,t}|\boldsymbol{X}_{t-1}] = \frac{\mathbb{C}ov[\epsilon_{1,t}, \epsilon_{2,t}] + X_{1,t-1}\mathbb{C}ov[Z_1, Z_2] + X_{2,t-1}\mathbb{C}ov[Z_3, Z_4]}{\sqrt{\big(\mathbb{V}[\epsilon_{1t}] + X_{1,t-1}\mathbb{V}[Z_1] + X_{2,t-1}\mathbb{V}[Z_3]\big)\big(\mathbb{V}[\epsilon_{2t}] + X_{1,t-1}\mathbb{V}[Z_2] + X_{2,t-1}\mathbb{V}[Z_4]\big)}}.$$

In the above equation, the denominator of the right hand side does not depend on q_1 or q_2, whereas the numerator is increasing in q_1 and q_2. In the independent BINAR(1) model with $q_1 = \alpha_{11}\alpha_{21}$, $q_2 = \alpha_{22}\alpha_{12}$, the above correlation coefficient is (in absolute value) non larger

than $\mathrm{corr}(\epsilon_{1t}, \epsilon_{2t})$, and becomes small whenever $X_{1,t-1}$ and/or $X_{2,t-1}$ are large, thus the conditional heteroscedasticity cannot be well captured. This downside exists also in several other competing bivariate count process models. For instance, in the bivariate Poisson autoregression of Liu (2012), the conditional correlation coefficient goes to zero when X_{1t} and X_{2t} are large. In copula-based bivariate count processes [Heinen and Rengifo (2007)], this conditional correlation coefficient can only be computed numerically and it is not clear how it behaves when components of $\boldsymbol{X}_{t-1}$ are large; in our model, when both $X_{1,t-1}$ and $X_{2,t-1}$ are large, the conditional correlation is approximately:

$$\mathrm{corr}[X_{1,t}, X_{2,t}|\boldsymbol{X}_{t-1}] \approx \frac{X_{1,t-1}\mathbb{C}ov[Z_1, Z_2] + X_{2,t-1}\mathbb{C}ov[Z_3, Z_4]}{\sqrt{\big(X_{1,t-1}\mathbb{V}[Z_1] + X_{2,t-1}\mathbb{V}[Z_3]\big)\big(X_{1,t-1}\mathbb{V}[Z_2] + X_{2,t-1}\mathbb{V}[Z_4]\big)}},$$

which can be closer to 1 (resp. -1) if q_1 and q_2 are close to their upper (resp. lower) bounds.

Finally, if ϵ_{1t}, ϵ_{2t} have Poisson distributions, then both components are conditionally under-dispersed:

$$\mathbb{V}[X_{jt}|\boldsymbol{X}_{t-1}] \leq \mathbb{E}[X_{jt}|\boldsymbol{X}_{t-1}], \qquad j = 1, 2.$$

This differs from the models of Cui and Zhu (2018); Gouriéroux and Lu (2018), which are conditionally equi-dispersed and over-dispersed, respectively. In general, by leaving the distribution of $\boldsymbol{\epsilon}_t$ unconstrained, the BINAR(1) family allows for flexible conditional dispersion.

2.2.2 Conditional p.g.f.

The dynamics of process $(\boldsymbol{X}_t)$ is characterized by its conditional probability generating function (p.g.f.), which is equal to:

$$\begin{aligned}\mathbb{E}[u^{X_{1,t}}v^{X_{2,t}}|\boldsymbol{X}_{t-1}] &= \big(\mathbb{E}[u^{Z_1}v^{Z_2}]\big)^{X_{1,t-1}}\big(\mathbb{E}[u^{Z_3}v^{Z_4}]\big)^{X_{2,t-1}}\mathbb{E}[u^{\epsilon_{1,t}}v^{\epsilon_{2,t}}] \\ &= a_1(u,v)^{X_{1,t-1}}a_2(u,v)^{X_{2,t-1}}b(u,v), \qquad \forall u, v \geq 0, \end{aligned} \tag{2.8}$$

where $b(u,v)$ is the p.g.f. of $(\epsilon_{1,t}, \epsilon_{2,t})$ and

$$a_1(u,v) = q_1 uv + (\alpha_{11} - q_1)u + (\alpha_{21} - q_1)v + (1 + q_1 - \alpha_{11} - \alpha_{21}), \qquad \forall u, v \geq 0, \tag{2.9}$$

$$a_2(u,v) = q_2 uv + (\alpha_{12} - q_2)u + (\alpha_{22} - q_2)v + (1 + q_2 - \alpha_{12} - \alpha_{22}), \qquad \forall u, v \geq 0 \tag{2.10}$$

are the p.g.f. of (Z_1, Z_2) [resp. (Z_3, Z_4)] respectively. This conditional p.g.f. is exponential affine in $\boldsymbol{X}_{t-1}$. Such a process is called compound autoregressive (CaR) [see Darolles et al. (2006)]. A remarkable property of such processes is that the multiple-step-ahead conditional p.g.f. remains exponential affine in $\boldsymbol{X}_{t-1}$:

Proposition 1. *In model (2.1), we have, for any horizon $h \geq 0$:*

$$\mathbb{E}[u^{X_{1,t+h-1}} v^{X_{2,t+h-1}} | \boldsymbol{X}_{t-1}] = a_1^{(h)}(u,v)^{X_{1,t-1}} a_2^{(h)}(u,v)^{X_{2,t-1}} b^{(h)}(u,v), \qquad \forall u, v \geq 0, \tag{2.11}$$

where functions $a_1^{(h)}(u,v), a_2^{(h)}(u,v)$ and $b^{(h)}(u,v)$ are defined by the recursion:

$$\begin{aligned}
a_1^{(h+1)}(u,v) &= 1 + \alpha_{11}[a_1^{(h)}(u,v) - 1] + \alpha_{21}[a_2^{(h)}(u,v) - 1] + q_1[a_1^{(h)}(u,v) - 1][a_2^{(h)}(u,v) - 1], \\
a_2^{(h+1)}(u,v) &= 1 + \alpha_{12}[a_1^{(h)}(u,v) - 1] + \alpha_{22}[a_2^{(h)}(u,v) - 1] + q_2[a_1^{(h)}(u,v) - 1][a_2^{(h)}(u,v) - 1], \\
b^{(h+1)}(u,v) &= b^{(h)}(u,v) b(a_1^{(h)}(u,v), a_2^{(h)}(u,v)), \qquad \forall u, v \geq 0, h \in \mathbb{N},
\end{aligned}$$

with initial conditions $a_1^{(0)}(u,v) = u, a_2^{(0)}(u,v) = v$, and $b^{(0)}(u,v) = 1$.

The proof (by induction) is straightforward and thus omitted. This proposition implies that for each $h \geq 1$, $a_1^{(h)}(u,v)$ and $a_2^{(h)}(u,v)$ are polynomials of degree 2^{h-1} in both arguments, except when $q_1 = q_2 = 0$. More precisely:

Corollary 1. *If $q_1 = q_2 = 0$, then:*

$$\begin{pmatrix} a_1^{(h)}(u,v) - 1 \\ a_2^{(h)}(u,v) - 1 \end{pmatrix} = A^\top \begin{pmatrix} a_1^{(h-1)}(u,v) - 1 \\ a_2^{(h-1)}(u,v) - 1 \end{pmatrix} = (A^\top)^h \begin{pmatrix} u - 1 \\ v - 1 \end{pmatrix}. \tag{2.12}$$

Thus in this case $a_1^{(h)}(u,v)$ and $a_2^{(h)}(u,v)$ are affine instead of polynomial. In other words, the conditional p.g.f. of $[A(0,0)\circ]^{(h)} \boldsymbol{X}_{t-1}$ given $\boldsymbol{X}_{t-1}$ at any horizon h has the same functional

form as that of $A^h(0,0) \circ \boldsymbol{X}_{t-1}$, or equivalently, the operator $[A(0,0)\circ]^{(h)} = A^h(0,0)\circ$ is still a dependent matrix thinning operator. As a consequence, in terms of temporal aggregation, when observed at a lower frequency of h, the process $(X_{th})_t$ is still BINAR(1).

Remark 1. Corollary 1 is easily explained by the birth-immigration interpretation: if $q_1 = q_2 = 0$, each individual produces at most one off-spring in the next period, and thus at most one off-spring at horizon $h \geq 2$. Hence the identity $[A(0,0)\circ]^{(h)} = A^h(0,0)\circ$. Moreover, this constrained model has also a similar queuing interpretation as the univariate INAR(1) model [see e.g. Schweer and Wichelhaus (2015)]. More precisely, we can think of $X_{1,t}, X_{2,t}$ as the number of individuals in queues 1 and 2 at date t, respectively. Both queues have infinite capacity. At date $t+1$, $\epsilon_{1,t+1}$ (resp. $\epsilon_{2,t+1}$) new customers join queue 1 (resp. queue 2), whereas $X_{1,t}$ (resp. $X_{2,t}$) customers that were in queue 1 (resp. queue 2) at date t can either stay in the same queue, or go to the other queue, or leave the queues after being served, with probabilities α_{11}, α_{21} and $1-\alpha_{11}-\alpha_{21}$ (resp. α_{22}, α_{12} and $1-\alpha_{22}-\alpha_{12}$).

Remark 2. Note that the literature [see, e.g., Boudreault and Charpentier (2011), Thm. 2.8, as well as Pedeli and Karlis (2013b), eq. 15] usually claims that in the independent BINAR model where $q_1 = \alpha_{11}\alpha_{21}, q_2 = \alpha_{21}\alpha_{22}$, the composite operator $[A(q_1,q_2)\circ]^{(h)}$ is equal to the bivariate thinning operator $A_h(q_{1,h}, q_{2,h})\circ$, where matrix $A_h = A^h$, and $q_{1,h}$ (resp. $q_{2,h}$) is the products of the two entries of the first (resp. second) column of A^h. Then these authors deduce that the $h-$step-ahead conditional p.g.f. still has the CaR form of equation (1), but with functions $a_1^{(h)}$ and $a_2^{(h)}$ being affine rather than higher-order polynomial. From the above analysis we can see that this assertion is incorrect.

Notice also that equation (2.11) corresponds to the decomposition:

$$\boldsymbol{X}_{t+h-1} = [A(q_1,q_2)\circ]^{(h)}\boldsymbol{X}_{t-1} + [A(q_1,q_2)\circ]^{(h-1)}\boldsymbol{\epsilon}_t + \cdots + A(q_1,q_2)\circ\boldsymbol{\epsilon}_{t+h-2} + \boldsymbol{\epsilon}_{t+h-1},$$

where the additive terms on the right hand side are conditionally independent given $\boldsymbol{X}_{t-1}$, with conditional p.g.f.'s $a_1^{(h)}(u,v)^{X_{1,t-1}} a_2^{(h)}(u,v)^{X_{2,t-1}}$ for $[A(q_1,q_2)\circ]^{(h)}\boldsymbol{X}_{t-1}$, $b(a_1^{(h-1)}(u,v), a_2^{(h-1)}(u,v))$ for $[A(q_1,q_2)\circ]^{(h-1)}\boldsymbol{\epsilon}_t$,..., and $b(u,v)$ for $\boldsymbol{\epsilon}_{t+h-1}$, respectively.

Example 1. Let us consider the case where ϵ_{1t}, ϵ_{2t} are mutually independent, Poisson $\mathcal{P}(\lambda_1), \mathcal{P}(\lambda_2)$

distributed, respectively. We have $b(u,v) = \exp\big(\lambda_1(u-1)+\lambda_2(v-1)\big)$, and the p.g.f. of $A(q_1,q_2)\circ\boldsymbol{\epsilon}_{t+h-2}$ is:

$$\begin{aligned} b(a_1(u,v),a_2(u,v)) &= \exp\Big(\lambda_1(a_1(u,v)-1)+\lambda_2(a_2(u,v)-1)\Big) \\ &= \exp\Big(\lambda_1\big[(\alpha_{11}-q_1)(u-1)+(\alpha_{21}-q_1)(v-1)+q_1(uv-1)\big] \quad (2.13) \\ &\quad +\lambda_2\big[(\alpha_{12}-q_2)(u-1)+(\alpha_{22}-q_2)(v-1)+q_2(uv-1)\big]\Big) \\ &= \exp\Big(m_1(u-1)+m_2(v-1)+m_3(uv-1)\Big), \quad (2.14) \end{aligned}$$

where $m_1=\lambda_1(\alpha_{11}-q_1)+\lambda_2(\alpha_{12}-q_2)$, $\quad m_2=\lambda_2(\alpha_{21}-q_1)+\lambda_2(\alpha_{22}-q_2)$, $\quad m_3=(\lambda_1q_1+\lambda_2q_2)$.

This is the p.g.f. of a bivariate (dependent) Poisson distribution $BP(m_1,m_2,m_3)$ with trivariate reduction [see, e.g., Marshall and Olkin (1985)].Its correlation coefficient is:

$$\rho_1=\frac{m_3}{\sqrt{(m_1+m_3)(m_2+m_3)}}=\frac{\lambda_1q_1+\lambda_2q_2}{\sqrt{(\lambda_1\alpha_{11}+\lambda_2\alpha_{21})(\lambda_1\alpha_{12}+\lambda_2\alpha_{22})}},$$

which is nonnegative, and increasing in q_1 and q_2. As a consequence, the conditional distribution $\boldsymbol{X}_{t+1}|\boldsymbol{X}_{t-1}$ is $BP(m_1+\lambda_1,m_2+\lambda_2,m_3)$.

Under the same assumption, the p.g.f. of $[A(q_1,q_2)\circ]^{(2)}\boldsymbol{\epsilon}_{t+h-3}$ is:

$$b(a_1^{(2)}(u,v),a_2^{(2)}(u,v))=\exp\Big(\lambda_1[a_1^{(2)}(u,v)-1]+\lambda_2[a_2^{(2)}(u,v)-1]\Big), \quad (2.15)$$

which is exponential quadratic in u and v. The family of distributions with exponential quadratic p.g.f. is called bivariate Hermite (BH). It nests the bivariate Poisson as special case and is closed under convolution [see Kemp and Papageorgiou (1982)]. In particular, the conditional distribution $\boldsymbol{X}_{t+1}|\boldsymbol{X}_{t-1}$ still belongs to the BH family.

2.3 Cross-sectional aggregation

In the previous subsection we have analyzed the frequency aggregation property of the BINAR(1) process. Brännäs et al. (2002) also consider the cross-sectional aggregation of univariate INAR(1) models. Let us now extend this analysis to the BINAR(1) models. Consider the dynamics of the

sum process $(X_{1,t} + X_{2,t})$:

$$X_{1,t} + X_{2,t} = \sum_{j=1}^{X_{1,t-1}} (Z_{1,j,t} + Z_{2,j,t}) + \sum_{j=1}^{X_{2,t-1}} (Z_{3,j,t} + Z_{4,j,t}) + \epsilon_{1,t} + \epsilon_{2,t}. \tag{2.16}$$

We can check that variables $Z_{1,j,t}+Z_{2,j,t}$ and $Z_{3,j,t}+Z_{4,j,t}$ are Bernoulli with parameters $\alpha_{11}+\alpha_{21}$ and $\alpha_{12}+\alpha_{22}$, respectively, if and only if they only take values 0 and 1, that is when $q_1 = q_2 = 0$. Nevertheless, given $X_{1,t-1} + X_{2,t-1}$, the sum $\sum_{j=1}^{X_{1,t-1}}(Z_{1,j,t} + Z_{2,j,t}) + \sum_{j=1}^{X_{2,t-1}}(Z_{3,j,t} + Z_{4,j,t})$ is generically not Binomial, except when $\alpha_{11} + \alpha_{21} = \alpha_{12} + \alpha_{22}$, in which case (2.16) becomes:

$$X_{1,t} + X_{2,t} = \sum_{j=1}^{X_{1,t-1}+X_{2,t-1}} (Z_{1,j,t} + Z_{2,j,t}) + \epsilon_{1,t} + \epsilon_{2,t}, \tag{2.17}$$

where $Z_{1,j,t} + Z_{2,j,t}$, j varying are independent of $\boldsymbol{X}_{t-1}, \boldsymbol{X}_{t-2}, ...$, and the innovation $\epsilon_{1,t} + \epsilon_{2,t}$ is independent of the variables $Z_{1,j,t} + Z_{2,j,t}$. To summarize, we have the following property:

Proposition 2. *In the BINAR(1) model with $q_1 = q_2 = 0$ and $\alpha_{11} + \alpha_{21} = \alpha_{12} + \alpha_{22}$, the sum process $(X_{1,t} + X_{2,t})$ is also univariate INAR(1) with autocorrelation coefficient $\alpha_{11} + \alpha_{21}$.*

Thus the sum process is Markov with respect to its own history. Moreover, eq. (2.17) says that: $\ell(X_{1,t} + X_{2,t}|X_{1,t-1}, X_{2,t-1})$ depends on $X_{1,t-1}, X_{2,t-1}$ only via the sum $X_{1,t-1} + X_{2,t-1}$. In other words, the sum process can be viewed as an exogenous, common Markov factor.

It is also possible to interpret condition $\alpha_{11} + \alpha_{21} = \alpha_{12} + \alpha_{22}$ in terms of the second order dynamics of the process. To this end we exclude two degenerate cases where matrix A is diagonal or anti-diagonal. Then condition $\alpha_{11} + \alpha_{21} = \alpha_{12} + \alpha_{22}$ implies that

$$(1,1)^\top A = (\alpha_{11} + \alpha_{21})(1,1)^\top,$$

that is, $\alpha_{11} + \alpha_{21}$ is an eigenvalue of matrix A associated with eigenvector $(1,1)^\top$. Since matrix A and vector $(1,1)^\top$ are positive, by the Perron-Frobenius theorem, $\alpha_{11} + \alpha_{21}$ is, in modulus, the simple largest eigenvalue of A. Thus, among all the linear combinations of components of $\boldsymbol{X}_t$, process $(X_{1,t} + X_{2,t})$ has the largest autocorrelation coefficient $\alpha_{11} + \alpha_{21}$. This justifies its interpretation as a common factor.

2.4 Stationarity

2.4.1 The stationarity condition

The strict stationarity condition of the BINAR(1) process is given in the next proposition:

Proposition 3. *Process* $(\boldsymbol{X}_t)$ *is both strictly and mean-variance stationary if and only if:*

$$(1-\alpha_{11})(1-\alpha_{22}) > \alpha_{12}\alpha_{21}, \tag{2.18}$$

or equivalently, if and only if the eigenvalues of A *are smaller than 1 in modulus.*

Note also that under condition (2.18), inequalities $\alpha_{21}+\alpha_{11} > 1$ and $\alpha_{22}+\alpha_{12} > 1$ cannot hold simultaneously. Thus in inequalities (2.2) and (2.3), at least one of the lower bounds are effectively zero.

2.4.2 The stationary distribution

Let us denote by b_∞ the p.g.f. of the stationary distribution. By taking expectation in (2.8), we get:

$$b_\infty(u,v) = b_\infty(a_1(u,v), a_2(u,v))b(u,v) \Longleftrightarrow b_\infty(u,v) = \prod_{i=1}^{\infty} b(a_1^{(h)}(u,v), a_2^{(h)}(u,v)).$$

This latter expression is generically complicated, but can be simplified in the special case considered in Corollary 1:

Proposition 4. *If:*

- *Processes* $(\epsilon_{1,t})$ *and* $(\epsilon_{2,t})$ *are mutually independent and Poisson* $\mathcal{P}(\lambda_1), \mathcal{P}(\lambda_2)$ *distributed;*
- *and* $q_1 = q_2 = 0$ *(which is possible only if* $\alpha_{11}+\alpha_{21} < 1$ *and* $\alpha_{12}+\alpha_{22} < 1$*),*

then:

- *the stationary distribution of process* $(\boldsymbol{X}_t)$ *is such that* X_{1t} *and* X_{2t} *are independent, Poisson* $\mathcal{P}(\lambda_{1,\infty})$*,* $\mathcal{P}(\lambda_{2,\infty})$ *distributed, with parameters given by:*

$$(\lambda_{1,\infty}, \lambda_{2,\infty})^\top = (Id - A)^{-1}(\lambda_1, \lambda_2)^\top,$$

- *the sum process* $(Y_t) = (X_{1,t} + X_{2,t})$ *is Poisson INAR(1), with representation:*

$$Y_t = \sum_{j=1}^{Y_{t-1}} B_{jt} + \eta_t, \tag{2.19}$$

where given Y_{t-1}*, variables* $(B_{jt})_j$ *are i.i.d. Bernoulli with probability* $\alpha = 1 - \frac{\lambda_1+\lambda_2}{\lambda_{1,\infty}+\lambda_{2,\infty}}$*, and* $(\eta_t) = (\epsilon_{1,t} + \epsilon_{2,t})$ *is i.i.d., independent of* B_{jt} *and* Y_{t-1}*, and Poisson* $\mathcal{P}(\lambda_1 + \lambda_2)$ *distributed.*

Proof. See Appendix A.1. □

Proposition 3 is the bivariate analogue of the well known result that the stationary distribution of a univariate Poisson-INAR(1) process is Poisson [see McKenzie (1985)]. It is interesting to compare it with Proposition 2, since in both cases we have assumed $q_1 = q_2 = 0$ and the sum process is INAR(1). On the one hand, Proposition 2 requires condition $\alpha_{11} + \alpha_{21} = \alpha_{12} + \alpha_{22}$, but leaves the distribution of the innovation $(\boldsymbol{\epsilon}_t)$ unconstrained. On the other hand, Proposition 4 does not restrict matrix A, but is based on the independent Poisson assumption of $(\boldsymbol{\epsilon}_t)$.

2.4.3 The marginal moments

The simple conditional expectation allows us to derive the first two marginal moments of the process. We have:

Proposition 5. *The marginal expectation of the process is given by:*

$$\mathbb{E}[\boldsymbol{X}_t] = (Id - A)^{-1}\mathbb{E}[\boldsymbol{\epsilon}_t], \tag{2.20}$$

the covariance matrix is:

$$\Gamma(0) := \begin{pmatrix} \mathbb{V}[X_{1,t}], & * \\ \mathbb{C}ov[X_{1,t}, X_{2,t}], & \mathbb{V}[X_{2,t}] \end{pmatrix} = \sum_{h=0}^{\infty} A^h \Big(\gamma_\epsilon + \mathbb{E}[X_{1t}]\gamma_{12} + \mathbb{E}[X_{2t}]\gamma_{34}\Big)(A^\top)^h, \tag{2.21}$$

whereas the autocovariance matrices are:

$$\Gamma(h) := \begin{pmatrix} \mathbb{C}ov[X_{1,t}, X_{1,t-h}], & \mathbb{C}ov[X_{1,t}, X_{2,t-h}] \\ \mathbb{C}ov[X_{2,t}, X_{1,t-h}], & \mathbb{C}ov[X_{2,t}, X_{2,t-h}] \end{pmatrix} = A^h \Gamma(0), \qquad \forall h \in \mathbb{N}. \tag{2.22}$$

Proof. See Appendix A.2. □

Thus, since matrix A is nonnegative, the variances $\mathbb{V}[X_{1t}], \mathbb{V}[X_{2t}]$ are increasing in q_1, q_2, *ceteris paribus*, whereas the marginal expectations do not depend on these two probabilities. Thus the marginal over-dispersion coefficients $\frac{\mathbb{V}[X_{1,t}]}{\mathbb{E}[X_{1,t}]}$, $\frac{\mathbb{V}[X_{2,t}]}{\mathbb{E}[X_{2,t}]}$ has a wider range than the independent BINAR model.

To illustrate this larger flexibility, we compare three models which differ only by q_1, q_2. We set $A = \begin{pmatrix} 0.5, & 0.3 \\ 0.4, & 0.5 \end{pmatrix}$ and assume $\epsilon_{1,t}$, ϵ_{2t} to be independent, $\mathcal{P}(1)$ distributed. Finally, q_1, q_2 are specified as follows:

- In Model 1, $q_1 = \min(\alpha_{1,1}, \alpha_{1,2}) = 0.4$, $q_2 = \min(\alpha_{2,1}, \alpha_{2,2}) = 0.3$, i.e. both bivariate Bernoulli variables $(Z_1, Z_2)^\top$ and $(Z_3, Z_4)^\top$ have maximal correlation.
- In the [independent BINAR(1)] Model 2, $q_1 = \alpha_{1,1}\alpha_{1,2} = 0.2$, $q_2 = \alpha_{2,1}\alpha_{2,2} = 0.15$, i.e., both bivariate Bernoulli variables have zero correlation.
- In Model 3, $q_1 = q_2 = 0$, i.e. both bivariate Bernoulli variables have minimal correlation.

Table 1 reports the over-dispersion coefficients and correlation coefficients computed using Proposition 4, under the three above models.

	$\mathbb{V}[X_{1,t}]/\mathbb{E}[X_{1,t}]$	$\mathbb{V}[X_{1,t}]/\mathbb{E}[X_{1,t}]$	$\text{corr}[X_{1,t}, X_{2,t}]$
Model 1	1.95	1.80	0.84
Model 2	1.47	1.40	0.55
Model 3	1	1	0

Table 1: Comparison of over-dispersion coefficients and correlation coefficients of the three models.

Note that in Model 3, The two over-dispersion coefficients as well as the correlation coefficient are equal to 1, 1, 0, respectively. Indeed in this case the marginal stationary distribution is bivariate independent Poisson. As expected, these coefficients are largest (resp. smallest) in

Model 1 (resp. Model 2). In other words, by letting q_1, q_2 vary, the dependent BINAR model allows for a much larger range of over-dispersion and coerrelation coefficients than the benchmark Model 2.

3 Higher-order BINAR processes

3.1 BINAR(p) process

Similar as the INAR(p) process introduced by Du and Li (1991), we define the dependent BINAR (p) process as follows:

Definition 2. We say that process $(\boldsymbol{X}_t)$ is dependent BINAR(p) if it has the representation:

$$\boldsymbol{X}_t = \sum_{i=1}^{p} A_i(q_{1,i}, q_{2,i}) \circ \boldsymbol{X}_{t-i} + \boldsymbol{\epsilon}_t, \tag{3.1}$$

where given $\underline{\boldsymbol{X}_{t-1}} = \{\boldsymbol{X}_{t-1}, \boldsymbol{X}_{t-2}, ...\}$, bivariate count variables $A_i(q_{1,i}, q_{2,i}) \circ \boldsymbol{X}_{t-i}$, $i = 1, ..., p$ are mutually independent, and are independent of $\boldsymbol{\epsilon}_t$. Moreover, $A_i(q_{1,i}, q_{2,i}) \circ \boldsymbol{X}_{t-i}$ is the sum of $X_{1,t-i}$ independent copies of bivariate Bernoulli variable with marginal (resp. joint) probabilities $\alpha_{11,i}$, $\alpha_{21,i}$ (resp. $q_{1,i}$), as well as $X_{2,t-i}$ independent copies of bivariate Bernoulli variable with marginal (resp. joint) probabilities $\alpha_{12,i}$, $\alpha_{22,i}$ (resp. $q_{2,i}$).

Thus compared with the BINAR(1) process, the extended model (3.1) has a slightly different interpretation since each individual can produce off-springs of both types at the next p periods, and these production outcomes are independent across these periods.

The stationarity condition of the BINAR(p) process is given below.

Proposition 6. *Process* (3.1) *is both strictly and mean-variance stationary if and only if:*

$$\sum_{i=1}^{p} \alpha_{11,i} < 1, \qquad \sum_{i=1}^{p} \alpha_{22,i} < 1, \tag{3.2}$$

$$\textit{and } \Big(1 - \sum_{i=1}^{p} \alpha_{11,i}\Big)\Big(1 - \sum_{i=1}^{p} \alpha_{22,i}\Big) > \Big(\sum_{i=1}^{p} \alpha_{21,i}\Big)\Big(\sum_{i=1}^{p} \alpha_{12,i}\Big). \tag{3.3}$$

or equivalently, if and only if the eigenvalues of $\sum_{i=1}^{p} A_i$ *are smaller than 1 in modulus.*

The BINAR(p) process has a weak VAR(p) representation since:

$$\mathbb{E}[\boldsymbol{X}_{t+1}|\underline{\boldsymbol{X}_t}] = \sum_{i=1}^{p} A_i \boldsymbol{X}_{t+1-p} + \mathbb{E}[\epsilon_{t+1}], \tag{3.4}$$

and its conditional p.g.f. is:

$$\mathbb{E}[u^{X_{1,t}} v^{X_{2,t}}|\underline{\boldsymbol{X}_{t-1}}] = \exp\Big[\log b(u,v) + \sum_{i=1}^{p} X_{1,t-i} \log a_{i,1}(u,v) + \sum_{i=1}^{p} X_{2,t-i} \log a_{i,2}(u,v)\Big], \tag{3.5}$$

$$\text{where } a_{i,1}(u,v) = q_{1,i}uv + (\alpha_{11,i} - q_{1,i})u + (\alpha_{21,i} - q_{1,i})v + (1 + q_{1,i} - \alpha_{11,i} - \alpha_{21,i}),$$
$$a_{i,2}(u,v) = q_{2,i}uv + (\alpha_{12,i} - q_{2,i})u + (\alpha_{22,i} - q_{2,i})v + (1 + q_{2,i} - \alpha_{12,i} - \alpha_{22,i}).$$

Thus the conditional p.g.f. is exponential affine in $\boldsymbol{X}_{t-1}, ..., \boldsymbol{X}_{t-p}$, i.e., process $(\boldsymbol{X}_t)$ is CaR of order p [CaR(p)]. Similar as the BINAR(1) process, its $h-$step$-$ahead conditional p.g.f. is still exponential affine.

Corollary 2. *We have:*

$$\mathbb{E}[u^{X_{1,t+h}} v^{X_{2,t+h}}|\underline{\boldsymbol{X}_t}] = \exp\Big(B_{h,0}(u,v) + \sum_{i=1}^{p} B_{h,i}^{\top}(u,v)\boldsymbol{X}_{t+1-i}\Big), \qquad \forall h \geq 1, \tag{3.6}$$

where $B_{h,0}$ and $B_{h,i}$ are univariate and bivariate functions, respectively. For $h = 1$, their values are given by equation (3.5)*:*

$$B_{1,0}(u,v) = \log\big(b(u,v)\big),$$
$$B_{1,i}(u,v) = \begin{pmatrix} \log\Big[q_{1,i}uv + (\alpha_{11,i} - q_{1,i})u + (\alpha_{21,i} - q_{1,i})v + (1 + q_{1,i} - \alpha_{11,i} - \alpha_{21,i})\Big] \\ \log\Big[q_{2,i}uv + (\alpha_{12,i} - q_{2,i})u + (\alpha_{22,i} - q_{2,i})v + (1 + q_{2,i} - \alpha_{12,i} - \alpha_{22,i})\Big] \end{pmatrix},$$

whereas for $h > 1$, we have the following recursions:

$$B_{h+1,0}(u,v) = B_{h,0}(u,v) + \log b\big(B_{h,1}(u,v)\big),$$
$$B_{h+1,i}(u,v) = \mathbb{1}_{i\neq p} B_{h,i+1}(u,v) + B_{1,i}(B_{h,i}(u,v)), \qquad \forall i = 1, ..., p.$$

The proof is a direct consequence of equation (3.5) and is omitted.

Finally, the marginal mean and variance-covariance matrix of the BINAR(p) also have closed form. Their formulas are derived in Appendix A.3.

3.2 BINAR(∞) process

3.2.1 Definition, stationarity, and memory persistence

A natural extension of the BINAR(p) model is to let the order p go to infinity. More precisely:

Definition 3. We say that process $(\boldsymbol{X}_t)$ is dependent BINAR(∞) if it has the representation:

$$\boldsymbol{X}_t = \sum_{i=1}^{\infty} A_i(q_{1,i}, q_{2,i}) \circ \boldsymbol{X}_{t-i} + \boldsymbol{\epsilon}_t, \qquad \forall t, \tag{3.7}$$

where variables $A_i(q_{1,i}, q_{2,i}) \circ \boldsymbol{X}_{t-i}$ are defined in the same way as in Definition 2.

In the above definition, the partial sum $\sum_{i=1}^{p} A_i(q_{1,i}, q_{2,i}) \circ \boldsymbol{X}_{t-i}$ given $\underline{\boldsymbol{X}_{t-1}}$ converges almost surely to a non degenerate limit when p goes to infinity. Moreover, since all the terms on the right hand side are integer-valued, variables $A_i(q_{1,i}, q_{2,i}) \circ \boldsymbol{X}_{t-i}$ are null for i larger than a certain stochastic threshold τ_t.

To our knowledge, this is the first infinite order INAR-type model in the literature [see Kirchner (2016) for infinite order, univariate INARCH process]. Let us first provide its stationarity condition.

Proposition 7. *Process* (3.1) *is both strictly and mean-variance stationary if and only if:*

$$\sum_{i=1}^{\infty} \alpha_{11,i} < 1, \qquad \sum_{i=1}^{\infty} \alpha_{22,i} < 1, \tag{3.8}$$

$$\textit{and} \sum_{i=1}^{\infty} \alpha_{12,i} < \infty, \qquad \sum_{i=1}^{\infty} \alpha_{21,i} < \infty, \tag{3.9}$$

$$\textit{and} \left(1 - \sum_{i=1}^{\infty} \alpha_{11,i}\right)\left(1 - \sum_{i=1}^{\infty} \alpha_{22,i}\right) > \left(\sum_{i=1}^{\infty} \alpha_{21,i}\right)\left(\sum_{i=1}^{\infty} \alpha_{12,i}\right). \tag{3.10}$$

or equivalently, if and only if matrix $\sum_{i=1}^{\infty} A_i$ is finite and its eigenvalues are all smaller than 1 in modulus.

This proposition nests Proposition 3 and 6. Its proof is provided in Appendix A.4.

This stationarity condition has also implications for the ranges of $q_{1,i}, q_{2,i}$, which have to satisfy the analogue of inequalities (2.2) and (2.3), for each i. Since sequences $(\alpha_{11,i})_i$ and $(\alpha_{21,i})_i$ converge to zero, for large i we have: $\alpha_{11,i} + \alpha_{21,i} < 1$. Then the ranges become:

$$0 \leq q_{1,i} \leq \min(\alpha_{11,i}, \alpha_{21,i}), \qquad 0 \leq q_{2,i} \leq \min(\alpha_{12,i}, \alpha_{22,i}), \qquad \text{for large } i. \tag{3.11}$$

Finally, the conditional expectation and conditional p.g.f. of a BINAR(∞) process can be deduced from (3.4) and (3.6) by replacing p by ∞. Under the stationary condition, these infinite summation are all finite.

3.2.2 Exact simulation

While simulating a BINAR(p) process is straightforward for finite p, this is no longer the case if $p = \infty$. Let us now derive an exact simulation method for this latter. The basic idea is that the infinite summation in (3.7) is almost surely finite, and thus it suffices to sample the stochastic threshold $\tau_t \in \mathbb{N}$ such that the infinite summation actually stops at order τ_t, i.e., $A_{\tau_t} \circ \boldsymbol{X}_{t-\tau_t}$ is non null but $A_i \circ \boldsymbol{X}_{t-i} = 0$ is zero for any $i \geq \tau_t + 1$. First we define:

$$\delta_i(\underline{\boldsymbol{X}_{t-1}}) := (1 + q_{1,i} - \alpha_{11,i} - \alpha_{21,i})^{X_{1,t-i}}(1 + q_{2,i} - \alpha_{12,i} - \alpha_{22,i})^{X_{2,t-i}}, \qquad \forall i \geq 1,$$

which is equal to $\mathbb{P}[A_i \circ \boldsymbol{X}_{t-i} = 0 | \underline{\boldsymbol{X}_{t-1}}]$. We also assume without loss of generality that:

Assumption 1.

$$\alpha_{11,i} + \alpha_{21,i} < 1, \qquad \alpha_{12,i} + \alpha_{22,i} < 1, \qquad \forall i \geq 1.$$

Indeed if the first few terms $q_{i,1}, q_{i,2}$ do not satisfy this condition, we can leave the corresponding variables $A_i \circ \boldsymbol{X}_{t-i}$ out of the infinite sum and simulate them separately. This assumption implies that $\delta_i(\underline{\boldsymbol{X}_{t-1}}) > 0$, for any i. Then the conditional CDF of the count variable τ is:

$$\mathbb{P}[\tau_t \leq i | \underline{\boldsymbol{X}_{t-1}}] = \prod_{j=i+1}^{\infty} \delta_j(\underline{\boldsymbol{X}_{t-1}}) := F(i+1 | \underline{\boldsymbol{X}_{t-1}}), \qquad \forall i \in \mathbb{N}.$$

This CDF has the following property:

Lemma 2. *Function $i \mapsto F(i|\underline{\boldsymbol{X}_{t-1}})$ is nondecreasing. Its upper limit is $\lim_{i\to\infty} F(i|\underline{\boldsymbol{X}_{t-1}}) = 1$, whereas its lower limit $F(0|\underline{\boldsymbol{X}_{t-1}})$ is strictly positive.*

Proof. See Appendix A.5. □

Since $F(i+1|\underline{\boldsymbol{X}_{t-1}})$ can be easily computed, we can simulate $\boldsymbol{X}_t$ given $\underline{\boldsymbol{X}_{t-1}}$ as follows:

- Draw U from the uniform distribution on $[0,1]$.
- Find the unique integer $\tau_t \geq 0$ such that $F(\tau_t - 1|\underline{\boldsymbol{X}_{t-1}}) < U \leq F(\tau_t|\underline{\boldsymbol{X}_{t-1}})$, where by convention we set $F(-1|\underline{\boldsymbol{X}_{t-1}}) = 0$. In particular, by definition $\boldsymbol{X}_{t-\tau_t}$ cannot be zero since $\delta_{\tau_t}(\underline{\boldsymbol{X}_{t-1}}) = \frac{F(\tau_t|\underline{\boldsymbol{X}_{t-1}})}{F(\tau_t-1|\underline{\boldsymbol{X}_{t-1}})} > 1$.
- Sample a certain number of independent copies of $A_{\tau_t} \circ \boldsymbol{X}_{t-\tau_t}$ until we obtain the first non null sample. This is possible due to Assumption 1.
- Sample random vectors $A_i \circ \boldsymbol{X}_{t-i}$ for $i = 1, ..., \tau_t - 1$, as well as $\boldsymbol{\epsilon}_t$. Then a sample of $\boldsymbol{X}_t$ is given by $\sum_{i=1}^{\tau_t} A_i \circ \boldsymbol{X}_{t-i} + \boldsymbol{\epsilon}_t$.

3.2.3 A constrained specification with persistent memory

To avoid the curse of dimensionality, in the application of this paper, we will focus on the following constrained BINAR(∞) specification:

$$\alpha_{11,i} = \frac{\alpha_{11}}{i^d}, \quad \alpha_{21,i} = \frac{\alpha_{21}}{i^d}, \quad \alpha_{12,i} = \frac{\alpha_{12}}{i^d}, \quad \alpha_{22,i} = \frac{\alpha_{22}}{i^d}, \quad q_{1,i} = \frac{q_1}{i^d}, \quad q_{2,i} = \frac{q_2}{i^d}, \quad \forall i = 1, ... \tag{3.12}$$

where the power index $d > 1$ to ensure that $\sum_i A_i$ is finite, and for each i, probabilities $q_{1,i}$ and $q_{2,i}$ satisfy the constraint:

$$\max(\alpha_{11,i} + \alpha_{21,i} - 1, 0) \leq q_{1,i} \leq \min(\alpha_{11,i}, \alpha_{21,i}), \quad \max(\alpha_{12,i} + \alpha_{22,i} - 1, 0) \leq q_{2,i} \leq \min(\alpha_{12,i}, \alpha_{22,i}).$$

It is easily checked that this is true if and only if these two inequalities hold for $i = 1$.

While in a BINAR(p) model, the autocovariance function decays geometrically in the lag h, a distinct feature of the BINAR(∞) model is that it allows the autocovariance to have a hyperbolic decay rate. More precisely we have the following result:

Lemma 3. *In model* (3.12)*, the autocovariance matrix* $\Gamma(h) = \mathbb{E}[(\boldsymbol{X}_t - \mathbb{E}[\boldsymbol{X}_t])(\boldsymbol{X}_{t-h} - \mathbb{E}[\boldsymbol{X}_t])^\top]$ *of process* $(\boldsymbol{X}_t)$ *decays also at the hyperbolic rate* d*.*

Proof. See Appendix A.6. □

This model has a similar spirit as the (univariate) ARCH(∞) model for asset returns [see Giraitis et al. (2000); Zaffaroni (2004)]. Since $d > 1$, the autocovariance matrix $\Gamma(h)$ is summable, ruling out the possibility that $\sum_{h=0}^{\infty} \Gamma(h) = \infty$ [see Giraitis et al. (2000) for a similar difficulty in the ARCH(∞) literature].

We plot in the following figure the simulated path of a BINAR(∞) process with:

$$A = \begin{pmatrix} 0.12, & 0.03 \\ 0.06, & 0.15 \end{pmatrix}, \qquad d = 1.3, \qquad q_1 = 0.015, \qquad q_2 = 0.03, \qquad (\epsilon_{1t}, \epsilon_{2t}) \sim BP(1, 1, 0.5).$$

In this model, the eigenvalues of $(\sum_{i=1}^{\infty} 1/i^d)A_1$ are 0.7 and 0.35, respectively, thus the persistence of the process is quite strong.

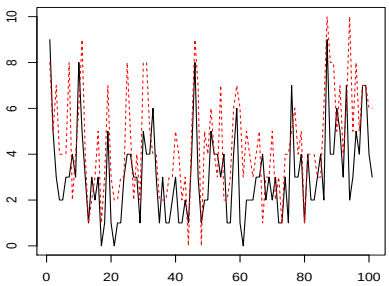

Figure 1: Simulated trajectory of the BINAR(∞) process under constraint (3.12).

The simulated means and variances of the two processes are: $\hat{\mathbb{E}}[X_{1t}] = 3.4$, $\hat{\mathbb{E}}[X_{2t}] = 4.6$, $\hat{\mathbb{V}}[X_{1t}] = 3.7$, $\hat{\mathbb{V}}[X_{2t}] = 5.1$, corresponding to a process with mild over-dispersion. Below we also report the autocorrelation functions (ACF) as well as the cross-correlation function (CCF) of the two component processes. These functions are computed using a simulated sample of 10000 observations (although in Figure 1 only a subset of 100 neighbouring observations is plotted).

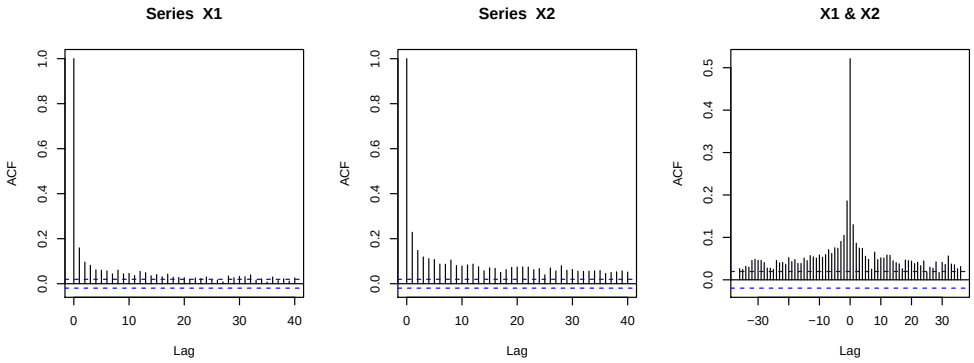

Figure 2: ACF and CCF function of the same simulated BINAR(∞) process as in Figure 1. As is expected, both functions decay rather slowly.

4 Predictive distributions

McCabe and Martin (2005) argue that one of the essential properties of a count process model is the tractability of the predictive p.m.f. of $\boldsymbol{X}_{t+h}|\underline{\boldsymbol{X}_t}$, for both estimation and forecasting purposes.

Indeed, in terms of estimation, there are two natural approaches for the BINAR(p) model. The first one is the Generalized Method of Moments (GMM), based on moment restrictions derived from the conditional p.g.f. [see e.g. Gouriéroux and Lu (2018)]. While this approach is computationally simple, it usually induces an efficiency loss. The second approach is the maximum likelihood estimation. Although more efficient, the difficulty of the latter lies in the computation of the conditional p.m.f. $\ell(\boldsymbol{X}_t|\underline{\boldsymbol{X}_{t-1}})$. Indeed, since this distribution is the convolution of $(p+1)$ bivariate discrete distributions, its expression is highly cumbersome if we apply brute-force convolution [see e.g. Marshall and Olkin (1985) for the expression of $\sum_{i=1}^{n}(Z_{1,j}, Z_{2,j})^\top$]. This has been identified by Pedeli and Karlis (2013b) as the major downside of higher-order (B)INAR models.

As for forecasting, the conditional expectations of count processes are non integer-valued and thus incompatible with count data. While some approximation methods [see, e.g., Jung and Tremayne (2006); McCabe et al. (2011)] have been proposed in the univariate INAR framework, they are generically time consuming and induce significant errors [see Lu (2018) for such an

analysis].

In this section we clarify that in the dependent (and, *a fortiori*, independent) BINAR(p) model, the conditional p.m.f. $\ell(\boldsymbol{X}_{t+h-1}|\underline{\boldsymbol{X}_{t-1}})$ at any horizon $h = 1, ...$ can be computed exactly. This makes the likelihood-based estimation feasible, and eliminates the approximation error induced for the forecasting. For expository purpose we focus on BINAR(p) models with a finite, potentially large p. We first consider the horizon $h = 1$, before explaining how to adapt the algorithm to higher horizons.

4.1 One-step-ahead predictive distribution

Our aim in this subsection is to compute the probabilities $\mathbb{P}[\boldsymbol{X}_t = (x_1, x_2)^\top|\underline{\boldsymbol{X}_{t-1}}]$ simultaneously for all couples $(x_1, x_2) \in [|0, m|] \times [|0, n|]$, where the bounds m, n are chosen such that the residual probability $\mathbb{P}[X_{1t} > m \text{ or } X_{2t} > n|\underline{\boldsymbol{X}_{t-1}}]$ is negligible.

Let us first remark that for any count process, the conditional p.g.f. and p.m.f. are linked via:

$$\mathbb{E}[u^{X_{1,t}} v^{X_{2,t}}|\underline{\boldsymbol{X}_{t-1}}] = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \mathbb{P}[\boldsymbol{X}_t = (i, j)^\top|\underline{\boldsymbol{X}_{t-1}}]u^i v^j, \qquad \forall u, v \geq 0. \tag{4.1}$$

Thus $\mathbb{P}[\boldsymbol{X}_t = (x_1, x_2)^\top|\underline{\boldsymbol{X}_{t-1}}]$ is equal to the $(x_1, x_2)-$th order coefficient in the above Taylor expansion with respect to (u, v), at $(0, 0)$. Let us now make use of the simple conditional p.g.f. to compute the Taylor expansion up to order (m, n). First, we rewrite (3.5) into:

$$\begin{aligned}\mathbb{E}[u^{X_{1,t}} v^{X_{2,t}}|\underline{\boldsymbol{X}_{t-1}}] &= \exp\Big(\log b(0,0) + \sum_{i=1}^{p} X_{1,t-i}\log a_1(0,0) + \sum_{i=1}^{p} X_{2,t-i}\log a_2(0,0)\Big)\\ &\times \exp\Big(\log\frac{b(u,v)}{b(0,0)} + \sum_{i=1}^{p} X_{1,t-i}\log\frac{a_{i,1}(u,v)}{a_{i,1}(0,0)} + \sum_{i=1}^{p} X_{2,t-i}\log\frac{a_{i,2}(u,v)}{a_{i,2}(0,0)}\Big).\end{aligned} \tag{4.2}$$

Then we perform the $(m, n)-$th order Taylor expansion of $\log[b(u, v)/b(0, 0)]$ and $\log[a_{i,j}(u, v)/a_{i,j}(0, 0)]$ at $(0, 0)$, where $i = 1, ..., p$, and $j = 1, 2$, respectively. For most standard bivariate count distributions, $\log b(u, v)$ is Taylor-expandable. Examples include the bivariate (independent or dependent) Poisson distribution [see equation (2.13)], the bivariate Hermite distribution [see equation (2.15)], as well as the bivariate negative binomial distribution [see Edwards and Gurland (1961)]

with p.g.f.:

$$b(u,v) = (1-b_1-b_2)^{\theta}/(1-b_1u-b_2v)^{\theta}, \quad \forall u,v \geq 0 \text{ such that } b_1u+b_2v<1,$$

where $b_1, b_2, \theta > 0$, $1-b_1-b_2>0$.

As for $\log \frac{a_{i,j}(u,v)}{a_{i,j}(0,0)}$, we have:

$$\log \frac{a_{i,j}(u,v)}{a_{i,j}(0,0)} = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} P_{i,j}^k(u,v) = \sum_{k=1}^{m+n} \frac{(-1)^{k-1}}{k} P_{i,j}^k(u,v) + o_{m,n}(u,v), \tag{4.3}$$

$$\text{where } P_{i,1}(u,v) = \frac{a_{i,1}(u,v)}{a_{i,1}(0,0)} - 1 = \frac{q_{1,i}uv + (\alpha_{11,i}-q_{1,i})u + (\alpha_{21,i}-q_{1,i})v}{1+q_{1,i}-\alpha_{11,i}-\alpha_{21,i}}, \qquad i=1,...,p,$$

$$P_{i,2}(u,v) = \frac{a_{i,2}(u,v)}{a_{i,2}(0,0)} - 1 = \frac{q_{2,i}uv + (\alpha_{12,i}-q_{2,i})u + (\alpha_{22,i}-q_{2,i})v}{1+q_{2,i}-\alpha_{12,i}-\alpha_{22,i}}, \qquad i=1,...,p,$$

are polynomials in u and v without constant term, whereas $o_{m,n}(u,v)$ represents the omitted higher-order terms in the expansion. Let us explain why the truncation in (4.3) stops at order $m+n$. The polynomial $P_{i,j}^k(u,v)$ is a linear combination of terms $u^{k_1}v^{k_2}$, where $k_1+k_2 \geq k$. Thus if $k > m+n$, then either $k_1 > m$ or $k_2 > n$, and $u^{k_1}v^{k_2}$ is omitted in the $(m,n)-$th order Taylor expansion. Therefore, we only need to expand recursively each $P_{i,j}^k(u,v)$ for $k=1,...,m+n$, $i=1,...,p$ and $j=1,2$ and truncate these polynomials at order (m,n). This latter can be achieved by the following algorithm:

Proposition 8. *If we represent the $(m,n)-$th order truncation of a polynomial P:*

$$P(u,v) = \sum_{k_1=0}^{m} \sum_{k_2=0}^{n} c_{k_1,k_2} u^{k_1} v^{k_2} + o_{m,n}(u,v),$$

by the column vector:

$$(\underbrace{c_{0,0}, c_{0,1}, ..., c_{0,n}}_{n+1 \ terms}, \underbrace{c_{1,0}, c_{1,1}, ..., c_{1,n}}_{n+1 \ terms}, \cdots, \underbrace{c_{m,0}, c_{m,1}, ..., c_{m,n}}_{n+1 \ terms})^{\top} \in \mathbb{R}^{(m+1)(n+1)},$$

then the $(m,n)-$th order truncation of polynomial $P^k(u,v)$ is represented by the column vector:

$$\underbrace{\begin{pmatrix} M_0 & 0 & 0 & 0 & \cdots \\ M_1 & M_0 & 0 & \cdots & \cdots \\ M_2 & M_1 & M_0 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ M_m & M_{m-1} & \cdots & \cdots & M_0 \end{pmatrix}}_{:=M}^{k} \begin{pmatrix} e_1 \\ 0 \\ \cdots \\ \cdots \\ 0 \end{pmatrix},$$

where column vector $e_1 = (1,0,0,\cdots,0)^\top \in \mathbb{R}^{n+1}$; square matrix $M \in \mathcal{M}_{(m+1)(n+1)}(\mathbb{R})$ is $(m+1)\times(m+1)$ block lower triangular Toeplitz, i.e.,

$$M = \Big[\mathbb{1}(i \geq j) M_{i-j}\Big]_{0\leq i,j\leq m},$$

and the block matrices $M_i \in \mathcal{M}_{n+1}(\mathbb{R})$ are themselves lower triangular Toeplitz:

$$M_i = \Big[\mathbb{1}(k_1 \geq k_2) c_{i,k_1-k_2}\Big]_{0\leq k_1,k_2\leq n}, \qquad \forall i = 0,...,m.$$

As an illustration, if $m = n = 2$ we have:

$$M = \left(\begin{array}{ccc|ccc|ccc} c_{00} & 0 & 0 & & & & & & \\ c_{01} & c_{00} & 0 & & \mathbf{0} & & & \mathbf{0} & \\ c_{02} & c_{01} & c_{00} & & & & & & \\ \hline c_{10} & 0 & 0 & c_{00} & 0 & 0 & & & \\ c_{11} & c_{10} & 0 & c_{01} & c_{00} & 0 & & \mathbf{0} & \\ c_{12} & c_{11} & c_{10} & c_{02} & c_{01} & c_{00} & & & \\ \hline c_{20} & 0 & 0 & c_{10} & 0 & 0 & c_{00} & 0 & 0 \\ c_{21} & c_{20} & 0 & c_{11} & c_{10} & 0 & c_{01} & c_{00} & 0 \\ c_{22} & c_{21} & c_{20} & c_{12} & c_{11} & c_{10} & c_{02} & c_{01} & c_{00} \end{array}\right).$$

This matrix has $(m+1)^2 = 9$ blocks, each block is a $(n+1)\times(n+1) = 3\times 3$ matrix.

The proof of this proposition is obvious and omitted.

After Taylor-expanding $\log b(u,v)$ and $\log[1+P_{i,j}(u,v)]$, $i=1,...,p$, $j=1,2$, we sum up these expansions and get:

$$\begin{aligned}\mathbb{E}[u^{X_{1,t}}v^{X_{2,t}}|\underline{\boldsymbol{X}_{t-1}}] &= c\exp\Big(\sum_{k_1=0}^{m}\sum_{k_2=0}^{n} f_{k_1,k_2}(\underline{\boldsymbol{X}_{t-1}})u^{k_1}v^{k_2}+R_m(u,v)\Big)\\ &= c\sum_{k=0}^{m+n}\frac{1}{k!}\Big(\sum_{k_1=0}^{m}\sum_{k_2=0}^{n} f_{k_1,k_2}(\underline{\boldsymbol{X}_{t-1}})u^{k_1}v^{k_2}\Big)^k+o_{m,n}(u,v), \qquad (4.4)\end{aligned}$$

where coefficients $f_{k_1,k_2}(\underline{\boldsymbol{X}_{t-1}})$ are affine in $\underline{\boldsymbol{X}_{t-1}}$; whereas constant c is equal to:

$$c=\exp\Big(\sum_{i=1}^{p}X_{1,t-i}\log(1+q_{1,i}-\alpha_{11,i}-\alpha_{21,i})+\sum_{i=1}^{p}X_{2,t-i}\log(1+q_{2,i}-\alpha_{12,i}-\alpha_{22,i})+\log b(0,0)\Big).$$

In equation (4.4), the expansion stops at order $m+n$ for the same reason as in (4.3). Then we apply Proposition 8 and obtain the $(m,n)-$th order truncation of polynomials

$$\Big(\sum_{k_1=0}^{m}\sum_{k_2=0}^{n} f_{k_1,k_2}(\underline{\boldsymbol{X}_{t-1}})u^{k_1}v^{k_2}\Big)^k, k=1,...,m+n.$$

Finally by coefficient matching we get the p.m.f. of $\boldsymbol{X}_t|\underline{\boldsymbol{X}_{t-1}}$.

In terms of the computational cost, the Taylor expansions of $\log b(u,v)$ and $\log[1+P_{i,j}(u,v)]$, $i=1,...,p, j=1,2$ are conducted only once when t varies. Thus for each t, the computation of $\ell(\boldsymbol{X}_t|\underline{\boldsymbol{X}_{t-1}})$ involves essentially the computation of the right hand side of equation (4.4), whose cost is *independent* of p. Thus this method is applicable even for large p.

The tractability of conditional distribution has several important implications. Firstly, it allows for efficient maximum likelihood estimation. In Appendix 7, we propose a small comparison between the MLE and a GMM estimator via Monte-Carlo experiments. Secondly, the likelihood function can also be used for model selection via the information criteria [see e.g. Weiss (2018)], as an alternative to Box-Jenkins type approach [see e.g. Bu and McCabe (2008) for an application on univariate INAR(p) models]. Thirdly, it allows to conduct likelihood ratio type tests for statistical significance of the parameters.

4.2 Comparison with the simulation-based forecasting approach

Let us now illustrate how the exact forecasting approach fares against the state-of-the-art simulation-based method [McCabe and Martin (2005); Jung and Tremayne (2006)] when it comes to the computation of the one-step-ahead predictive distribution. While this latter method is general and applies to other non-BINAR count process model, it will be shown that one of the advantages of BINAR models is that the exact approach outperforms significantly the simulation approach, both in terms of computational time and forecasting accuracy (note that McCabe et al. (2011) propose another method to approximate the predictive p.m.f. They approximate the a univariate INAR(p) process by a Markov chain with S states, where S is a large integer. This method involves the computation of matrices of dimension $S^{2p} \times S^{2p}$, which is extremely cumbersome when $p \geq 2$).

To this end we consider the BINAR(p) model with:

$$A_i = \frac{1}{i^d} \begin{pmatrix} 0.12, & 0.06 \\ 0.03, & 0.15 \end{pmatrix}, \quad i = 1, ..., p, \qquad q_1 = 0.015, \qquad q_2 = 0.03,$$

whereas the innovation ($\boldsymbol{\epsilon}_t$) follows the bivariate Poisson distribution $BP(2, 2, 2)$. Given the past observation $\underline{X_T}$, the simulation-based method consists in drawing a large number of possible future values $\boldsymbol{X}_{T+1}^{(n)}, n = 1, ..., N$. Then the conditional p.m.f. is approximated by:

$$\mathbb{P}[\boldsymbol{X}_{T+1} = (i, j)^\top | \underline{\boldsymbol{X}_T}] \approx \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}\big(\boldsymbol{X}_{T+1}^{(n)} = (i, j)^\top\big), \qquad i, j.$$

We first report below the run time of the two methods (for the simulation-based method we consider two values for the number of draws). Both methods are implemented in R using the same laptop (intel i5, 3.0 GHz, 8GB RAM) and the program is available from the authors upon request.

$p = 1$, $\boldsymbol{X}_T = (1,4)^\top$			
Method	Exact method	Simulation-based method	Simulation-based method
Number of draws	0	$N = 10000$	$N = 100000$
CPU time	0.01 s	0.05 s	0.5 s
$p = 4$, $\boldsymbol{X}_T = \boldsymbol{X}_{t-1} = \boldsymbol{X}_{T-2} = \boldsymbol{X}_{T-3} = (1,4)^\top$			
Method	Exact method	Simulation-based method	Simulation-based method
Number of draws	0	$N = 10000$	$N = 100000$
CPU time	0.01 s	0.2 s	2 s

Table 2: CPU time of the two methods applied to BINAR(1) and BINAR(4) models. For the exact approach, we compute the value of the conditional p.m.f. $\ell((i,j)^\top|\underline{\boldsymbol{X}_T})$ for any i, j ranging from 0 to 15, although in the table we only report their values for i, j non larger than 8.

We see that the exact method is 10 to 100 times faster than the simulation method. Moreover this ratio becomes larger when the order p increases. This is expected since the run time of the simulation-based method is roughly proportional to p, whereas we have argued in the previous section that it remains roughly the same in our approach.

Let us now illustrate the accuracy of the simulation-based approach. We focus on the above BINAR(1) model, and express the approximated conditional p.m.f.'s as a percentage of the corresponding exact values.

p.m.f. computed using the exact method					
	$X_{1,T+1}=0$	$X_{1,T+1}=2$	$X_{1,T+1}=4$	$X_{1,T+1}=6$	$X_{1,T+1}=8$
$X_{2,T+1}=0$	0.00096	0.00248	0.00104	0.00017	0.00001
$X_{2,T+1}=2$	0.00324	0.02270	0.01944	0.00552	0.00074
$X_{2,T+1}=4$	0.00172	0.02491	0.04171	0.02079	0.00447
$X_{2,T+1}=6$	0.00035	0.00885	0.02625	0.02250	0.00782
$X_{2,T+1}=8$	0.00003	0.00145	0.00698	0.00975	0.00542
Relative accuracy of the simulation-based method, $N=10000$					
	$X_{1,T+1}=0$	$X_{1,T+1}=2$	$X_{1,T+1}=4$	$X_{1,T+1}=6$	$X_{1,T+1}=8$
$X_{2,T+1}=0$	72 %	132 %	105 %	115 %	0 %
$X_{2,T+1}=2$	111 %	101 %	104 %	104 %	93 %
$X_{2,T+1}=4$	87 %	96 %	98 %	99 %	109 %
$X_{2,T+1}=6$	85 %	103 %	106 %	108 %	91 %
$X_{2,T+1}=8$	0 %	89 %	91 %	108 %	101 %
Relative accuracy of the simulation-based method, $N=100000$					
	$X_{1,T+1}=0$	$X_{1,T+1}=2$	$X_{1,T+1}=4$	$X_{1,T+1}=6$	$X_{1,T+1}=8$
$X_{2,T+1}=0$	118 %	99 %	105 %	92 %	65 %
$X_{2,T+1}=2$	101 %	97 %	101 %	96 %	115 %
$X_{2,T+1}=4$	96 %	98 %	100 %	100 %	100 %
$X_{2,T+1}=6$	99 %	96 %	101 %	97 %	98 %
$X_{2,T+1}=8$	80 %	95 %	97 %	105 %	100 %

Table 3: Relative accuracy of the simulation approach compared to the exact approach. Due to the space constraint we have only reported the results for odd values of $X_{1,T+1}$ and $X_{2,T+1}$.

We can see that the approximation error of the simulation approach is substantial, even with a huge number of draws ($N = 100000$). This is particularly the case for the probabilities of "extreme events", that is when either $X_{1,t+1}$ or/and $X_{2,t+1}$ is large. This is a serious downside of the standard simulation approach, since in finance, predicting extreme events is key to the risk management.

4.3 Multiple-step-ahead predictive distributions

Let us now adapt the above algorithm for the computation of $\ell(\boldsymbol{X}_{t+h}|\underline{\boldsymbol{X}_t})$. For expository purpose, we focus on the BINAR(1) process. To Taylor-expand the conditional p.g.f. given in Proposition 1, we use the following procedure:

- First, we use Proposition 8 to compute the $(m,m)-$th order Taylor expansion of $\log a_1^{(h)}(u,v)$, $\log a_2^{(h)}(u,v)$ and $b^{(h)}(u,v)$ at $(u,v)=(0,0)$, where m is chosen such that the probability of either component processes taking values larger than m is negligible.Note that although

by Proposition 1, $a_1^{(h)}$ and $a_2^{(h)}$ are 2^h-th order polynomial in (u,v), which can be large for large h, only the terms of degree lower than $\min(m, 2^h)$ contribute to the $(m,m)-$th Taylor expansion. Thus this step involves roughly the same computational effort for different values of h.

- Then we compute the $(m,m)-$th order Taylor expansion of

$$\begin{aligned}
&\exp\Big(X_{1,t}\log a_1^{(h)}(u,v) + X_{2,t}\log a_2^{(h)}(u,v) + \log b^{(h)}(u,v)\Big)\\
&= c\exp\Big(X_{1,t}\log\frac{a_1^{(h)}(u,v)}{a_1^{(h)}(0,0)} + X_{2,t}\log\frac{a_2^{(h)}(u,v)}{a_2^{(h)}(0,0)} + \log b^{(h)}(u,v)\Big)\\
&= c\sum_{k=0}^{2m}\frac{1}{k!}\Big(X_{1,t}\log\frac{a_1^{(h)}(u,v)}{a_1^{(h)}(0,0)} + X_{2,t}\log\frac{a_2^{(h)}(u,v)}{a_2^{(h)}(0,0)} + \log\frac{b^{(h)}(u,v)}{b(0,0)}\Big)^k + o_{m,m}(u,v),
\end{aligned}$$

where $c = \exp\Big(X_{1,t}\log a_1^{(h)}(0,0) + X_{2,t}\log a_2^{(h)}(0,0) + \log b(0,0)\Big)$. This is conducted using the same method as for $h=1$. Finally we deduce $\mathbb{P}[\boldsymbol{X}_{t+h} = (i,j)^\top|\underline{\boldsymbol{X}_t}]$ for any $(i,j) \in [|0,m|]^2$ by coefficient matching.

5 Application

5.1 The mutual fund industry

Mutual funds (MF) are investment vehicles who invest in a wide range of assets ranging from liquid ones such as stocks, bonds, to highly illiquid ones such as hedge funds. Their clients include, for instance, insurance companies, private banks, large corporations as well as retail investors (In some funds, including the one we study, retail investors' orders are first centralized by a broker before being transferred to the fund. Thus from the fund's point of view, its client is the broker). They are traditionally much less regulated than commercial banks, but this potential loophole has recently received much attention of the regulator.

Most MF are open-ended, *i.e.*, they allow investors to purchase new shares, or redeem their shares on a daily basis. Thus the size of the MF can feature important short-term fluctuations, making them vulnerable to liquidity risk. In particular, during a market turmoil, investors' redemption decisions tend to cluster. If the fund manager's cash holding is insufficient to meet

the redemption requests, he might be foced to sell its illiquid assets, whose market liquidity would have also plunged due to the crisis. Such fire selling usually leads to significant investment loss, which in turn creates panics and triggers further redemptions. This phenomenon is called fund run. On the other hand, while inflow, that is the purchase orders of the MF, or cash holding can offset the outflow due to redemption, a sudden large inflow also dilutes the investment performance of the fund due to the lack of immediate investment opportunities. Thus they can also trigger subsequent (large) outflows. As a consequence, it is essential for the fund manager to monitor parallely the outflow and the inflow pattern of its clients, on a daily basis. The current MF literature usually focuses on the outflow only, or the net outflow, that is the difference between the outflow volume and inflow volume [see, e.g., Schmidt et al. (2016); Desmettre et al. (2018)]. Moreover, many of these studies are based on weekly data only.

Finally, while prior studies focus on the volume of the outflow/inflow, our attention is on the number of purchase $X_{1,t}$ and redemption orders $X_{2,t}$. These variables are closely related since on each trading day, the volume $Y_{1,t}$ and $Y_{2,t}$ of the outflow/inflow have the compound representation:

$$Y_{1,t} = \sum_{j=1}^{X_{2,t}} S_{j,t}, \qquad Y_{2,t} = \sum_{j=1}^{X_{2,t}} B_{j,t}, \tag{5.1}$$

where $S_{j,t}$ (resp. $B_{j,t}$) denotes the volume of the $j-$th redemption (resp. purchase). Moreover, when j and t vary, $(S_{j,t})$ and $(B_{j,t})$ can be reasonably assumed to i.i.d.

In this framework, our preference for studying the counts rather than the volume is motivated by the following reasons. First, very often a large outflow volume is due to the redemption order of a large investor. These "VIP" clients usually have privileged relationship with the fund manager and in the case of a large redemption, they also tend to (privately) inform the fund manager sufficiently in advance so that the latter can avoid a massive fire selling. On the other hand, a large $X_{2,t}$ spells a collective withdrawal, that is the fund run, which is the most dangerous scenario. More importantly, the non-linear forecasting of the net outflow, that is $Y_{2,t} - Y_{1,t}$, can be easily deduced from representation (5.1). Indeed, the conditional Laplace transform of $Y_{2,t} - Y_{1,t}$ given $\underline{\boldsymbol{X}_{t-1}}$ is:

$$\mathbb{E}[e^{-u(Y_{2,t}-Y_{1,t})}|\underline{\boldsymbol{X}_{t-1}}] = \mathbb{E}\Big[(\mathbb{E}[e^{uB_{1,t}}])^{X_{2,t}}(\mathbb{E}[e^{-uS_{1,t}}])^{X_{1,t}}|\underline{\boldsymbol{X}_{t-1}}\Big], \qquad \forall u \in \mathbb{R}. \tag{5.2}$$

The right hand side is the conditional p.g.f. of the count process evaluated at $(\mathbb{E}[e^{uS_{1,t}}], \mathbb{E}[e^{-uB_{1,t}}])$, which has closed form under suitable distributional assumptions of $S_{1,t}$ and $B_{1,t}$ (such as gamma). Thus the left hand side of (5.2) is readily available. Then the conditional Value-at-Risk of the net outflow, say, can be accurately approximated without simulation [see Gordy (2002)]. As a consequence, in the paper we will only focus on counts.

5.2 Data description

Our dataset comes from a French equity-focused MF. We observe the daily number of purchase orders X_{1t} and redemption orders X_{2t}, during about 1000 trading days. The following three figures plot the trajectory of the two count processes during a sub-sample of 100 trading days, as well as the histograms of X_{1t} and X_{2t}.

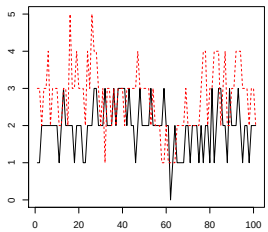

Figure 3: Joint trajectory of the purchase count (red dashed line) and redemption count (black full line) during 100 trading days.

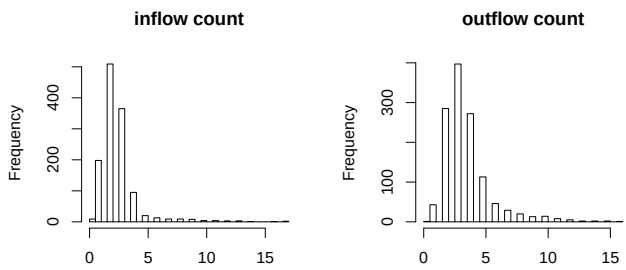

Figure 4: Histogram of the daily purchase/redemption order counts. We can see that the probability of $X_{1,t}$ and $X_{2,t}$ taking large values is extremely small.

We report below the empirical marginal moments of the two processes.

$\hat{\mathbb{E}}[X_{1t}]$	$\hat{\mathbb{E}}[X_{2t}]$	$\hat{\mathbb{V}}[X_{1t}]$	$\hat{\mathbb{V}}[X_{2t}]$	$\widehat{\text{corr}}[X_{1t}, X_{2t}]$
2.63	3.06	3.67	3.97	0.27

Table 4: Summary empirical moments of the two processes.

Both processes feature mild unconditional over-dispersion, a typical feature of BINAR processes. Figure 5 plots their autocorrelation patterns.

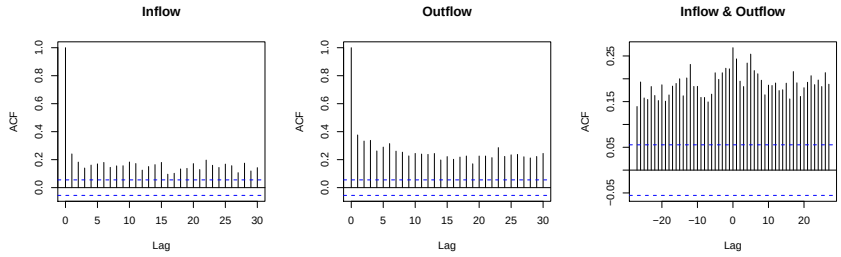

Figure 5: ACF's and CCF of inflow count process ($X_{1,t}$) and outflow process count (X_{2t}).

These ACF's and CCF decay rather slowly, resembling the simulated patterns of section 3.2.3 for a model with hyperbolic decaying coefficients $A_i, q_{1,i}, q_{2,i}$. Thus in this section, we will focus on the estimation of this latter model. Note that we can interpret the thinning part of model (2.1) as trades by clients who also made purchase/sell orders in the previous periods. Then the binomial distributional assumption can be explained by the fact that usually clients only trade on a daily basis (even if they place several orders within a day, usually the fund only executes the orders once each day at the market closure, and thus only one aggregate order is counted for each client). Thus this interpretation suggests model (2.1) with the further constraint that $q_1 = q_2 = 0$, that is, clients do not make simultaneous buy and sell orders as these two will partially cancel out. In practice, in the application, although this constrained version of model (2.1) is easier to interpret, we have considered a BINAR(∞) as this latter reflects better the persistence of the ACF function.

Since the CCF indicates a positive (marginal) correlation between the two processes, it is reasonable to assume that the innovation process ($\boldsymbol{\epsilon}_t$) features also positive correlation between its two components. Moreover, the degree of unconditional over-dispersion of the two count variables are rather weak (see Table 4), suggesting the Poisson distribution for the marginal distribution of $\boldsymbol{\epsilon}_t$. Thus we assume the distribution of the latter to be bivariate Poisson BP($\lambda_1, \lambda_2, \lambda_3$). Then

the set of parameters is:

$$\theta = (\alpha_{11}, \alpha_{22}, \alpha_{12}, \alpha_{21}, q_1, q_2, \lambda_1, \lambda_2, \lambda_3, d).$$

Let us now interpret the regression parameters $\alpha_{i,j}$. Parameter α_{11} measures how purchase decisions of investors are (positively) correlated, due to the so-called reputation effect. Moreover, since here the investors' benchmark is past yearly performance and our observations concern daily inflow/outflow movements, it is not surprising that this reputation effect decays rather slowly when daily data is used, as is shown by the ACF of X_{1t}. Similarly, parameter α_{22} measures the panic effect among investors. Parameter α_{12} captures the propensity of redemption following large recent inflows, as such inflows might dilute the fund's performance due to lack of sufficient investment opportunities. This is consistent with the CCF given in Figure 5, which indicates a positive cross correlation between X_{2t} and the lagged values of $X_{1,t}$. Finally, parameter α_{21} captures the propensity of purchase following large outflows. This can be interpreted as the fund manager's capability of attracting new investment in order to stabilize the fund size.

5.3 Model estimation

We estimate the parameters by maximum likelihood. A practical difficulty of applying an infinite order model is that our number of observations (which is around 1000 trading days) is finite. Thus the "∞" in $\sum_{i=1}^{\infty} A_i \circ \boldsymbol{X}_{t-i}$ should be replaced by a finite, but large p. In the estimation we take $p = 300$, and regard the first p values of X_t as initial values rather than observations. We have checked that for reasonable values of parameters, further increasing p only marginally impacts the value of the likelihood function. The following table reports the maximum likelihood estimates:

	Parameter estimate
α_{11}	0.174 (0.055)
α_{21}	0.0355 (0.016)
α_{12}	0.120 (0.052)
α_{22}	0.285 (0.12)
q_1	0.0219 (0.012)
q_2	0.0173 (0.0073)
λ_1	0.712 (0.21)
λ_2	0.695 (0.16)
λ_3	0.354 (0.18)
d	1.83 (0.1)

Table 5: Parameter estimate along with the standard errors of the estimators.

The parameter estimate $\hat{\alpha}_{22}(= 0.285)$ is larger than $\hat{\alpha}_{11}(= 0.174)$. In other words, the panic effect is more important than the reputation effect. Second, $\hat{\alpha}_{12}(= 0.12)$ is much larger than $\hat{\alpha}_{21}(= 0.0355)$, which means that existing investors are quite sensitive to large inflows and tend to redeem their shares for fear of performance drop after a large inflow. This highlights the importance of monitoring the purchase counts separately from the redemption counts. On the other hand, it seems to be difficult for the fund to attract new investors after large outflows. Finally, the two eigenvalues of the matrix $(\sum_{i=1}^{\infty} 1/i^d)A_1$ are 0.58 and 0.26, respectively, which are both smaller than 1. Thus the joint process seems to be stationary.

Let us now compute the conditional p.m.f. We focus on horizon 1, since the fund manager usually adjusts the positions on a daily basis. For each trading day t in our observation period, we compute $\mathbb{P}[X_{1,t+1} = m, X_{2,t+1} = n|\underline{\boldsymbol{X}_t}]$ for all $m, n \leq M$, where we set M to be the largest past observation $M = \max_{t,j} X_{j,t} = 17$.Then we follow McCabe and Martin (2005) and compute the conditional mode $(\hat{X}_{1,t+1}, \hat{X}_{2,t+1})$ defined by:

$$(\hat{X}_{1,t+1}, \hat{X}_{2,t+1}) = \arg \max_{0 \leq m,n \leq M} \mathbb{P}[X_{1,t+1} = m, X_{2,t+1} = n|\underline{\boldsymbol{X}_t}].$$

Figure 6 displays the evolution of the conditional mode $\hat{X}_{2,t}$ against the corresponding realized value $X_{2,t}$ for all the past dates.

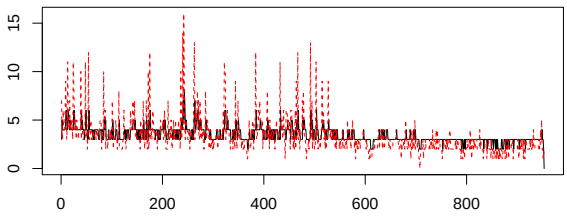

Figure 6: Joint evolution of $\hat{X}_{2,t+1}$ (in black full line) and the realized redemption count $X_{2,t+1}$ (in red dashed line).

Globally, we can see that the mode forecast can quite well capture the local tendency of the count process, although the realized paths tends to be more erratic.

Diagnostic check Let us finally conduct some adequacy checks of the estimated model. We first compute the ACF/CCF of the estimated model, as well as some summary statistics of the empirical Pearson residuals [see e.g. Weiss (2018) for their definitions]. For the first one, since in the paper we have only derived the ACF/CCF for finite p, we resort to Monte-Carlo simulation to obtain approximations of these functions for this BINAR(∞) model. Due to space constraint this figure is provided in Appendix 8. We can remark that globally, the ACF/CCF of the estimated process are quite similar to their empirical counterparts reported in Figure 5. Moreover, the Pearson residuals seem to be well uncorrelated across different lags.

6 Conclusion

We have extended the BINAR(1) model to allow for dependent thinning, arbitrary errors, and higher-order dynamics. This family has intuitive interpretations, tractable stationarity properties, and are rather flexible compared to existing models. More importantly, we have derived tractable expressions for the predictive distributions, allowing for likelihood based estimation and non-linear forecasting. The model has been applied to a new application area, i.e., fund liquidity risk.

In the paper we have followed the literature by focusing on bivariate models. Is it possible

to extend our model into higher dimensions? From the fund manager's point of view, this can be of interest since different investors-large corporates and bank/insurance companies, say, can have different behaviour. If count series, two for each client category, becomes available, then a multivariate analysis allows to study cross sectional dynamic effects between different clients and may improve the quality of forecasts. The answer is (partially) affirmative, since equation (2.1) can be extended to the multivariate case, using the multivariate Bernoulli distribution [Chaganty and Joe (2006)]. However such extensions are not without downsides. Firstly, this extended model is not closed under margins; for example the bivariate margins of the trivariate extension no longer have representation (2.1). Secondly and most importantly, as many multivariate models, when the dimension increases, both the number of parameters and the computational burden increases. These issues might be mitigated by introducing constrained specifications (see, e.g., Proposition 2), or by conducting pairwise analysis [see Pedeli and Karlis (2013a); Gouriéroux and Lu (2018)]These await future research.

Appendix

A.1. Proof of Proposition 4

Under the assumptions of the proposition, we have:

$$\log b^{(1)}(u,v) = \lambda_1(u-1) + \lambda_2(v-1) = (\lambda_1, \lambda_2)(u-1, v-1)^\top,$$

where $\lambda_1 = \mathbb{E}[\epsilon_{1t}]$ and $\lambda_2 = \mathbb{E}[\epsilon_{2t}]$. By equation (2.12) we get:

$$\begin{aligned}\log b^{(\infty)}(u,v) &= \sum_{h=1}^{\infty} \log b^{(h)}(u,v) \\ &= \sum_{h=1}^{\infty} (\lambda_1, \lambda_2)(A^\top)^h (u-1, v-1)^\top = (\lambda_1, \lambda_2)(\mathrm{Id} - A^\top)^{-1}(u-1, v-1)^\top,\end{aligned}$$

which is the log p.g.f. of a bivariate independent Poisson distribution, with expectations:

$$(\lambda_{1,\infty}, \lambda_{2,\infty})^\top = (\mathrm{Id} - A)^{-1}(\lambda_1, \lambda_2)^\top.$$

Let us now check that $(Y_t) = (X_{1,t} + X_{2,t})$ follows a Poisson INAR(1) process. The joint p.g.f. of $(X_{1,t-1} + X_{2,t-1}, X_{1t} + X_{2t})$ is:

$$
\begin{aligned}
&\mathbb{E}[u^{X_{1,t-1}+X_{2,t-1}} v^{X_{1,t}+X_{2,t}}] \\
&= \mathbb{E}\Big[u^{X_{1,t-1}+X_{2,t-1}}\big(1+(\alpha_{11}+\alpha_{21})(v-1)\big)^{X_{1,t-1}}\big(1+(\alpha_{12}+\alpha_{22})(v-1)\big)^{X_{2,t-1}} e^{(\lambda_1+\lambda_2)(v-1)}\Big] \\
&= \exp\Big\{\lambda_{1,\infty}\big[u\big(1+(\alpha_{11}+\alpha_{21})(v-1)\big)-1\big] + \lambda_{2,\infty}\big[u\big(1+(\alpha_{12}+\alpha_{22})(v-1)\big)-1\big] + (\lambda_1+\lambda_2)(v-1)\Big\} \\
&= \exp\Big\{[\lambda_{1,\infty}(\alpha_{11}+\alpha_{21}) + \lambda_{2,\infty}(\alpha_{12}+\alpha_{22})]uv + u\underbrace{[\lambda_{1,\infty}(1-\alpha_{11}-\alpha_{21}) + \lambda_{2,\infty}(1-\alpha_{12}-\alpha_{22})]}_{=\lambda_1+\lambda_2 \text{ by equation (2.20)}} \\
&\quad + (\lambda_1+\lambda_2)(u_2-1) - \lambda_{1,\infty} - \lambda_{2,\infty}\Big\} \qquad \text{(eq.a.1)}
\end{aligned}
$$

A quick calculation shows that for a univariate Poisson INAR(1) process (Z_t) satisfying $Z_t = \alpha \circ Z_{t-1} + \eta_t$ where $\alpha\circ$ is the univariate binomial thinning operator, and η_t is i.i.d. $\mathcal{P}(\lambda)$ distributed, the joint p.g.f. is:

$$
\mathbb{E}[u^{Z_t} v^{Z_{t-1}}] = \exp\Big[\frac{\alpha\lambda}{1-\alpha}uv + \lambda u + \lambda v - \lambda(1+\frac{1}{1-\alpha})\Big]. \qquad \text{(eq.a.2)}
$$

By matching equations (eq.a.1) and (eq.a.2), we conclude that $(X_{1t} + X_{2t})$ follows Poisson INAR(1) with $\lambda = \lambda_1 + \lambda_2$, and $\alpha = 1 - (\lambda_1 + \lambda_2)/(\lambda_{1,\infty} + \lambda_{2,\infty})$.

A.2. Proof of Proposition 5

The marginal expectation is obtained by taking expectation in equation (2.4). By the co-variance decomposition formula we get:

$$\begin{pmatrix} \mathbb{V}[X_{1,t}], & * \\ \mathbb{C}ov[X_{1,t}, X_{2,t}], & \mathbb{V}[X_{2,t}] \end{pmatrix} = \begin{pmatrix} \mathbb{V}[\alpha_{11}X_{1,t-1} + \alpha_{12}X_{2,t-1}], & * \\ \mathbb{C}ov[\alpha_{11}X_{1,t-1} + \alpha_{12}X_{2,t-1}, \alpha_{21}X_{1,t-1} + \alpha_{22}X_{2,t-1}], & \mathbb{V}[\alpha_{21}X_{1,t-1} + \alpha_{22}X_{2,t-1}] \end{pmatrix}$$

$$+ \mathbb{E}\Big[\begin{pmatrix} \mathbb{V}[X_{1,t}|\boldsymbol{X}_{t-1}], & * \\ \mathbb{C}ov[X_{1,t}, X_{2,t}|\boldsymbol{X}_{t-1}], & \mathbb{V}[X_{2,t}|\boldsymbol{X}_{t-1}] \end{pmatrix} \Big]$$

$$= A \begin{pmatrix} \mathbb{V}[X_{1,t}], & * \\ \mathbb{C}ov[X_{1,t}, X_{2,t}], & \mathbb{V}[X_{2,t}] \end{pmatrix} A^\top + \gamma_\epsilon + \mathbb{E}[X_{1t}]\gamma_{12} + \mathbb{E}[X_{2t}]\gamma_{34}.$$

Solving this linear matrix equation yields solution (2.21). Finally, equation (2.22) is a direct consequence of equation (2.4).

A.3. The first two marginal moments of BINAR(p) model

Under the stationarity condition, the marginal expectation of a BINAR(p) model satisfies:

$$\mathbb{E}[\boldsymbol{X}_t] = \sum_{i=1}^{p} A_i \mathbb{E}[\boldsymbol{X}_t] + \mathbb{E}[\boldsymbol{\epsilon}_t] \Longleftrightarrow \mathbb{E}[\boldsymbol{X}_t] = (Id - \sum_{i=1}^{p} A_i)^{-1}\mathbb{E}[\boldsymbol{\epsilon}_t].$$

where matrix $Id - \sum_{i=1}^{p} A_i$ is invertible by conditions (3.2) and (3.3). The covariance matrix can be obtained using the companion CaR(1) form. More precisely, let us denote by V the $(2p \times 2p)$ covariance matrix of $(\boldsymbol{Y}_t) = (\boldsymbol{X}_t, \boldsymbol{X}_{t-1}, ..., \boldsymbol{X}_{t-p+1})$, i.e.:

$$V := \mathbb{E}\Big[(\boldsymbol{Y}_t - \mathbb{E}[\boldsymbol{Y}_t])(\boldsymbol{Y}_t - \mathbb{E}[\boldsymbol{Y}_t])^\top\Big] = \begin{pmatrix} \Gamma(0) & \Gamma(1) & \cdots & \Gamma(p-1) \\ \Gamma^\top(1) & \Gamma(0) & \cdots & \Gamma(p-2) \\ \cdots & \cdots & \cdots & \cdots \\ \Gamma(p-1)^\top & \Gamma(p-2)^\top & \cdots & \Gamma(0) \end{pmatrix},$$

where $\Gamma(h)$ is the auto-covariance function of $(\boldsymbol{X}_t)$ at lag h. Since $\boldsymbol{Y}_t$ has the VAR(1) representation:

$$\mathbb{E}[\boldsymbol{Y}_t|\boldsymbol{Y}_{t-1}] = \underbrace{\begin{pmatrix} A_1 & A_2 & \cdots & A_p \\ I_2 & 0 & \cdots & \cdots \\ 0 & I_2 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}}_{:=A} \boldsymbol{Y}_{t-1} + \begin{pmatrix} \mathbb{E}[\boldsymbol{\epsilon}_t] \\ 0 \\ \cdots \\ \cdots \end{pmatrix},$$

we have:

$$V = AVA^\top + \begin{pmatrix} \gamma_\epsilon + \sum_{i=1}^p \left(\mathbb{E}[X_{1,t}]\gamma_{12,i} + X_{2,t}\gamma_{34,i}\right) & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{eq.a.3}$$

where $\gamma_{12,i}, \gamma_{34,i}$ are defined in a similar way as in (2.7). Thus $V = \sum_{h=0}^\infty (A^\top)^h V_0 A^h$, where V_0 is the second matrix term on the right hand side of (eq.a.3).

A.4. Proof of Proposition 7

We first reformulate the condition given in Proposition 7 in terms of the eigenvalues of A:

Lemma 4. *For any matrix $A = (\alpha_{i,j})_{1\le i,j\le 2}$ with nonnegative entries only, the two following sets of conditions are equivalent:*

1. *$\alpha_{11} < 1, \alpha_{22} < 1$, and $(1-\alpha_{11})(1-\alpha_{22}) > \alpha_{12}\alpha_{21}$.*
2. *The eigenvalues of A are smaller than 1 in modulus.*

The proof is straightforward and omitted [see e.g. Gouriéroux and Lu (2018)].

Let $(\boldsymbol{\epsilon}_t)$ be an i.i.d. sequence of innovations and $(Z_{1,t,i,j}, Z_{2,t,i,j})$, $(Z_{3,t,i,j}, Z_{4,t,i,j})$ i.i.d. bivariate Bernoulli variables for any $t \in \mathbb{Z}, i, j \in \mathbb{N}$. Then we define the doubly indexed sequences

$(\boldsymbol{X}_t^{(n)}) = (X_{1,t}^{(n)}, X_{2,t}^{(n)})_t^\top$ recursively by:

$$\boldsymbol{X}_t^{(n)} = \begin{cases} \boldsymbol{\epsilon}_t, & n = 0, \quad \forall t; \\ \boldsymbol{\epsilon}_t + \sum_{i=1}^n \sum_{j=1}^{X_{1,t-i}^{(n-1)}} \begin{pmatrix} Z_{1,t,i,j} \\ Z_{2,t,i,j} \end{pmatrix} + \sum_{i=1}^n \sum_{j=1}^{X_{2,t-i}^{(n-1)}} \begin{pmatrix} Z_{3,t,i,j} \\ Z_{4,t,i,j} \end{pmatrix}, & n > 0, \quad \forall t; \end{cases}$$

We will show that the $n-$indexed sequence $(\boldsymbol{X}_t^{(n)})_n$ converges almost surely to a limit $\boldsymbol{X}_t$, and that the limiting process $(\boldsymbol{X}_t)$ satisfies the definition (3.7).

A.5.1. Almost sure and $\mathbb{L}^1(\mathbb{P})$ convergence of $\boldsymbol{X}_t^{(n)}$

By induction (with respect to n) it is easily checked that:

$$X_{1,t}^{(n)} \geq X_{1,t}^{(n-1)}, \qquad \forall t \geq 0, \qquad \text{and} \qquad X_{2,t}^{(n)} \geq X_{2,t}^{(n-1)} \geq 0, \qquad \forall t \geq 0.$$

Thus for each t, both $n-$indexed sequences $(X_{1,t}^{(n)})_n$ and $(X_{2,t}^{(n)})_n$ converge almost surely, to limits, say, $X_{1,t}$ and $X_{2,t}$, respectively. Let us now consider the $\mathbb{L}^1(\mathbb{P})$ convergence. We have:

$$\mathbb{E}[\boldsymbol{X}_t^{(n)}] = \mathbb{E}[\boldsymbol{\epsilon}_t] + \sum_{i=1}^n A_i \mathbb{E}[\boldsymbol{X}_t^{(n-1)}] \leq \mathbb{E}[\boldsymbol{\epsilon}_t] + \sum_{i=1}^n A_i \mathbb{E}[\boldsymbol{X}_t^{(n)}] \leq \mathbb{E}[\boldsymbol{\epsilon}_t] + \underbrace{\Big(\sum_{i=1}^\infty A_i\Big)}_{:=A_\infty} \mathbb{E}[\boldsymbol{X}_t^{(n)}],$$

where the inequalities hold componentwise. Thus we have:

$$(Id - A_\infty)\mathbb{E}[\boldsymbol{X}_t^{(n)}] \leq \mathbb{E}[\boldsymbol{\epsilon}_t].$$

Then we remark that the (2×2) matrix $(Id - A_\infty)$ is invertible and all the entries of its inverse are nonnegative. Thus we can multiply both sides by $(Id - A_\infty)^{-1}$ and deduce that:

$$\mathbb{E}[\boldsymbol{X}_t^{(n)}] \leq (Id - A_\infty)^{-1}\mathbb{E}[\boldsymbol{\epsilon}_t]$$

is upper bounded when n increases. Hence $(\boldsymbol{X}_t^{(n)})$ converges also in $\mathbb{L}^1(\mathbb{P})$ to $\boldsymbol{X}_t$ for each t by monotonous convergence theorem.

A.5.2. Strict and second-order stationarity of process $(\boldsymbol{X}_t)_t$

By definition, for each n, process $(\boldsymbol{X}_t^{(n)})_t$ is strictly stationary. Thus the limiting process $(\boldsymbol{X}_t)$ is also strictly stationary. By the $\mathbb{L}^1(\mathbb{P})$ convergence, $(\boldsymbol{X}_t)$ is mean-stationary. To show the covariance stationarity, let us check that the sequence $(X_{1t}^{(n)})_n$ is bounded in $\mathbb{L}^2(\mathbb{P})$. We have:

$$\begin{aligned}
\mathbb{V}[X_{1t}^{(n)}] &= \mathbb{V}[\epsilon_{1t}] + \mathbb{V}\Big[\sum_{i=1}^{n}\sum_{j=1}^{X_{1,t-i}^{(n-1)}} Z_{1,t,i,j} + \sum_{i=1}^{n}\sum_{j=1}^{X_{2,t-i}^{(n-1)}} Z_{3,t,i,j}\Big] \\
&\leq \mathbb{V}[\epsilon_{1t}] + \mathbb{E}\Big[\Big(\sum_{i=1}^{n}\sum_{j=1}^{X_{1,t-i}^{(n-i)}} Z_{1,t,i,j} + \sum_{i=1}^{n}\sum_{j=1}^{X_{2,t-i}^{(n-i)}} Z_{3,t,i,j}\Big)^2\Big] \\
&\leq \mathbb{V}[\epsilon_{1t}] + \mathbb{E}\Big[\Big(\sum_{i=1}^{n}\sum_{j=1}^{X_{1,t}^{(n)}} Z_{1,t,i,j} + \sum_{i=1}^{n}\sum_{j=1}^{X_{2,t}^{(n)}} Z_{3,t,i,j}\Big)^2\Big] \\
&\leq C_1 + \mathbb{V}\Big[(\sum_{i=1}^{\infty}\alpha_{11,i})X_{1t}^{(n)} + (\sum_{i=1}^{\infty}\alpha_{12,i})X_{2t}^{(n)}\Big],
\end{aligned}$$

where constant C_1 is independent of t and n. Similar upper bounds can also be obtained for $\mathbb{V}[X_{1t}^{(n)}]$ and $\mathbb{V}[X_{2t}^{(n)}]$ and in matrix form we have:

$$V_n \leq A_\infty V_n A_\infty^\top + C,$$

where V_n is the covariance matrix of $\boldsymbol{X}_t^{(n)}$, and C is a constant matrix. Thus we get:

$$V_n \leq \sum_{i=0}^{\infty} A_\infty^i C (A_\infty^\top)^i,$$

which is uniformly bounded. Thus by the dominated convergence theorem, $(\boldsymbol{X}_t^{(n)})_n$ also converges to $(\boldsymbol{X}_t)$ in $\mathbb{L}^2(\mathbb{P})$.

A.5.3. Conditional distribution of X_t given its past

It remains to check that the above limiting process $(\boldsymbol{X}_t)$ satisfies the representation (3.7). It suffices to show that, for a fixed t, the sequence

$$\boldsymbol{r}_t^{(n)} = \boldsymbol{\epsilon}_t + \sum_{i=1}^{n} \sum_{j=1}^{X_{1,t-i}} (Z_{1,t,i,j}, Z_{2,t,i,j})^\top + \sum_{i=1}^{n} \sum_{j=1}^{X_{2,t-i}} (Z_{3,t,i,j}, Z_{4,t,i,j})^\top$$

converges to $\boldsymbol{X}_t$ almost surely (*a.s.*), or equivalently, since

$$\boldsymbol{X}_t - \boldsymbol{r}_t^{(n)} = \boldsymbol{X}_t - \boldsymbol{X}_t^{(n)} + \boldsymbol{X}_t^{(n)} - \boldsymbol{r}_t^{(n)},$$

it suffices to show that $\boldsymbol{r}_t^{(n)} - \boldsymbol{X}_t^{(n)}$ converges to zero *a.s.* But this sequence is non-decreasing, thus it suffices to find an *a.s.* convergent subsequence. Its $\mathbb{L}^1(\mathbb{P})$ norm is:

$$\mathbb{E}[\boldsymbol{r}_t^{(n)} - \boldsymbol{X}_t^{(n)}] = \sum_{i=1}^{n} A_i \mathbb{E}[\boldsymbol{X}_t] - \sum_{i=1}^{n} A_i \mathbb{E}[\boldsymbol{X}_t^{(n)}] \longrightarrow 0$$

when n goes to infinity. Thus $(\boldsymbol{r}_t^{(n)} - \boldsymbol{X}_t^{(n)})_n$ converges in $\mathbb{L}^1(\mathbb{P})$ and admits a subsequence that is *a.s.* convergent. Thus process $(\boldsymbol{X}_t)$ satisfies equation (3.7).

A.5.4. The necessary condition of stationarity

Thus under conditions (3.8), (3.10), (3.9), the BINAR(∞) process exists and is stationary. Let us now show that these conditions are also necessary. Taking expectation in (3.7), we get:

$$\mathbb{E}[\boldsymbol{X}_t] = \mathbb{E}[\boldsymbol{\epsilon}_t] + A_\infty \mathbb{E}[\boldsymbol{X}_t], \tag{eq.a.4}$$

thus all the entries of A_∞ are finite, hence inequality (3.8). Finally, by iteration we have:

$$\mathbb{E}[\boldsymbol{X}_t] = (Id + A_\infty + A_\infty^2 + \cdots + A_\infty^n)\mathbb{E}[\boldsymbol{\epsilon}_t] + A_\infty^{n+1}\mathbb{E}[\boldsymbol{X}_t], \qquad \forall n \geq 1.$$

Thus the largest eigenvalue of A_∞ is smaller than one in modulus. Then by Lemma 4 we get conditions (3.9) and (3.10).

A.5. Proof of Lemma 2

Under Assumption 1, each $\delta_i(\underline{\boldsymbol{X}_{t-1}})$ lies between 0 and 1, thus $F(i|\underline{\boldsymbol{X}_{t-1}})$ is nondecreasing, and upper bounded by 1. Its lower limit is strictly positive under the stationarity condition since:

$$F(0|\underline{\boldsymbol{X}_{t-1}}) = \exp\Big(\sum_{i=1}^{\infty} X_{1,t-i}\log(1+q_{1,i}-\alpha_{11,i}-\alpha_{21,i}) + \sum_{i=1}^{\infty} X_{2,t-i}\log(1+q_{2,i}-\alpha_{12,i}-\alpha_{22,i})\Big)$$
$$\geq \exp\Big(-\sum_{i=1}^{\infty} X_{1,t-i}\log(1-\alpha_{11,i}-\alpha_{21,i}) - \sum_{i=1}^{\infty} X_{2,t-i}\log(1-\alpha_{12,i}-\alpha_{22,i})\Big). \qquad \text{(eq.a.5)}$$

As $\alpha_{11,i}+\alpha_{21,i}\to 0$ and $\alpha_{12,i}+\alpha_{22,i}\to 0$ when i goes to infinity, for large i, we have:

$$-\log(1-\alpha_{12,i}-\alpha_{22,i}) > -2(\alpha_{12,i}+\alpha_{22,i}), \qquad -\log(1-\alpha_{12,i}-\alpha_{22,i}) > -2(\alpha_{12,i}+\alpha_{22,i}).$$

Then since $\sum_{i=1}^{\infty}(\alpha_{12,i}+\alpha_{22,i})X_{1,t-i} + \sum_{i=1}^{\infty}(\alpha_{12,i}+\alpha_{22,i})X_{2,t-i}$ is finite, the right hand side of (eq.a.5) is positive.

A.6. Proof of Lemma 3

The proof is based on the fact that process $(\boldsymbol{X}_t)$ has the weak VAR(∞) representation:

$$\boldsymbol{X}_t = \sum_{i=1}^{\infty} A_i \boldsymbol{X}_{t-i} + \boldsymbol{\eta}_t,$$

where $\boldsymbol{\eta}_t$ is a weak white noise. Then we revert this VAR(∞) representation into the Vector MA(∞) representation $\boldsymbol{X}_t = (Id + \sum_{i=1}^{\infty} B_i L^i)\boldsymbol{\eta}_t$, where B_j are matrices and L is the lag operator. By mimicking the proof of Theorem 2 in Zaffaroni (2004), we can show that$B_i = O(i^d)D$ for some constant matrix D and the autocovariance function also decays at the same hyperbolic rate i^d.

A.7. Comparison with the GMM and ML estimators

For illustrative purpose, in this section we focus on a BINAR(1) model with independent, Poisson innovations, and compare the small sample behavior of the GMM and ML estimators. Let us

first briefly recall the GMM method, which has recently been suggested by Gouriéroux and Lu (2018) for CaR count processes.

A.7.1. The GMM method

We denote by $\mathcal{U}$ a grid of positive real numbers, then the conditional pgf provides a set of moment constraints:

$$\mathbb{E}\Big[\psi_{u,v}(\boldsymbol{X}_{t-1})\Big(u^{X_{1,t}}v^{X_{2,t}} - \underbrace{a_1(u,v)^{X_{1,t-1}}a_2(u,v)^{X_{2,t-1}}b(u,v)}_{=\mathbb{E}[u^{X_{1,t}}v^{X_{2,t}}|\boldsymbol{X}_{t-1}]}\Big)\Big] = 0, \qquad \forall u,v \in \mathcal{U}, \quad \text{(eq.a.6)}$$

where $\psi_{u,v}(\boldsymbol{X}_{t-1})$ can be any instrumental function depending only on $\boldsymbol{X}_{t-1}$. We choose ψ to be:

$$\psi_{u,v}(\boldsymbol{X}_{t-1}) = \frac{1}{\sqrt{\mathbb{V}[u^{X_{1,t}}v^{X_{2,t}}|\boldsymbol{X}_{t-1}]}},$$

where
$$\begin{aligned}\mathbb{V}[u^{X_{1,t}}v^{X_{2,t}}|\boldsymbol{X}_{t-1}] &= \mathbb{E}[u^{2X_{1,t}}v^{2X_{2,t}}|\boldsymbol{X}_{t-1}] - \Big[\mathbb{E}[u^{X_{1,t}}v^{X_{2,t}}|\boldsymbol{X}_{t-1}]\Big]^2 \\ &= a_1(u^2,v^2)^{X_{1,t-1}}a_2(u^2,v^2)^{X_{2,t-1}}b(u^2,v^2) - a_1(u,v)^{2X_{1,t-1}}a_2(u,v)^{2X_{2,t-1}}b^2(u,v),\end{aligned}$$

such that the integrand in equation (eq.a.6) has unitary unconditional variance.

Then a simple GMM estimator is obtained by minimizing the error of the empirical counterparts of these orthogonality conditions:

$$\hat{\theta}_T = \arg\min_{\theta}\Big[\frac{1}{T-1}\sum_{t=2}^{T} g(\boldsymbol{X}_{t-1},\boldsymbol{X}_t,\theta)\Big]^\top W\Big[\frac{1}{T-1}\sum_{t=2}^{T} g(\boldsymbol{X}_{t-1},\boldsymbol{X}_t,\theta)\Big],$$

where g is a vector function of dimension $\Big[\text{Card}(\mathcal{U})\Big]^2$ given by:

$$g_{\psi,u,v}(\boldsymbol{X}_{t-1},\boldsymbol{X}_t) = \psi_{u,v}(\boldsymbol{X}_{t-1})\Big(u^{X_{1,t}}v^{X_{2,t}} - a_1(u,v)^{X_{1,t-1}}a_2(u,v)^{X_{2,t-1}}b(u,v)\Big), \qquad \forall u,v \in \mathcal{U},$$

and W is a symmetric positive definite weighting matrix. To ensure the asymptotic consistency, the number of moment conditions $[\text{Card}(\mathcal{U})]^2$ has to be larger than the number of parameters. Then, under standard regularity conditions, the GMM estimator is consistent and asymptotically

normally distributed:

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \longrightarrow_{(d)} \mathcal{N}\Big(0, (G^\top W G)^{-1} G W \Omega W^\top G (G^\top W^\top G)^{-1}\Big),$$

where θ_0 is the true parameter value, $G = \mathbb{E}[\nabla g(\boldsymbol{X}_{t-1}, \boldsymbol{X}_t, \theta_0)]$, with ∇ representing the differential with respect to argument θ, and $\Omega = \mathbb{E}[g(\boldsymbol{X}_{t-1}, \boldsymbol{X}_t, \theta_0) g(\boldsymbol{X}_{t-1}, \boldsymbol{X}_t, \theta_0)^\top]$. Moreover, the optimal choice of the weighting matrix W is given by $W = \Omega^{-1}$, and for this choice we have:

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \longrightarrow_{(d)} \mathcal{N}\Big(0, (G^\top \Omega^{-1} G)^{-1}\Big).$$

Thus the asymptotic behaviour of the GMM estimator will usually depend on the choice of the set $\mathcal{U}$. As far as we know, this issue has yet to be considered for count data and thus is well beyond the scope of the present paper. In the simulation exercise below, we take:

$$\mathcal{U} = \{0.11, 0.22, ..., 0.99, 1.1\},$$

which corresponds to a total of $[\text{Card}(\mathcal{U})]^2 = 121$ moment conditions, which is well beyond the typical number of parameters of a BINAR(p) model.

A.7.2. A Monte-Carlo comparison of GMM and ML estimators

In this subsection we consider the following dependent BINAR(1) model with independent, Poisson $\mathcal{P}(\lambda_1)$ (resp. $\mathcal{P}(\lambda_2)$) distributed error terms $\epsilon_{1,t}$ and $\epsilon_{2,t}$. The parameters of the model are set as:

$$A = \begin{pmatrix} 0.4 & 0.2 \\ 0.3 & 0.5 \end{pmatrix}, \qquad q_1 = q_2 = 0.1, \qquad \lambda_1 = \lambda_2 = 1. \tag{A.7.1}$$

Then we simulate $m = 100$ independent trajectories of the process, each with a sample size of $T = 500$. For each simulated dataset $j = 1, ..., m$, we estimate the model using both the GMM and ML methods to obtain two estimators $\hat{\theta}_{1,j}$ and $\hat{\theta}_{2,j}$. Boxplots of the empirical distribution of the parameters $(\hat{\theta}_{1,j})_j$ and $(\hat{\theta}_{2,j})_j$ are given in the following figures.

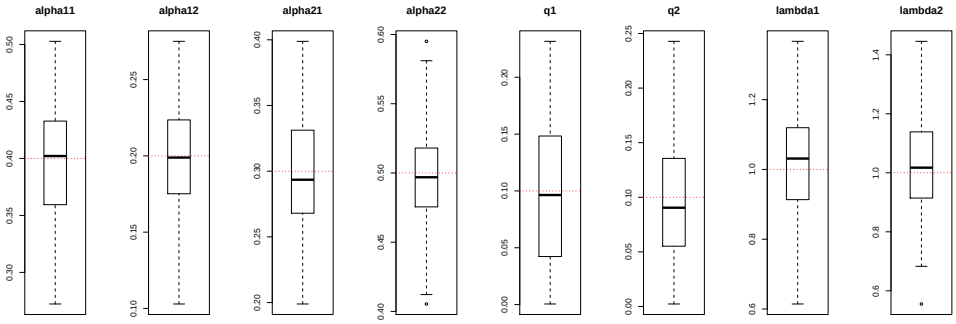

Figure A.7.1: Boxplots of the GMM estimators of the BINAR(1) model (2.1) with parameter values given by (A.7.1). The real value of each component of θ is plotted in red dotted line.

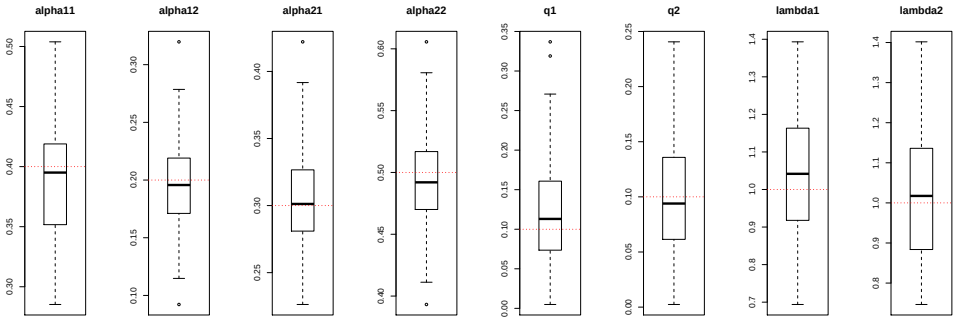

Figure A.7.2: Boxplots of the MLE estimators of the same model. The real value of each component of θ is plotted in red dotted line.

We can remark that, first, since the sample size T is only moderately large, the distribution of the estimators are quite different from the limiting Gaussian distribution. In particular, due to the constraint that all components of the estimators have to be bounded from below (by zero), and most of them (including $\alpha_{11}, .., \alpha_{22}, q_1, q_2$ are also from above by 1, their distribution can be either left skewed, or right skewed. Secondly, by comparing the above two figures, we can remark a slight better finite sample behavior for the MLE estimator, in the sense that for most components, the range of the MLE estimator is (slightly) more concentrated than for the GMM estimator. This is expected, since in most applications the MLE is asymptotically more efficient than a GMM estimator.

A.8. Additional figures

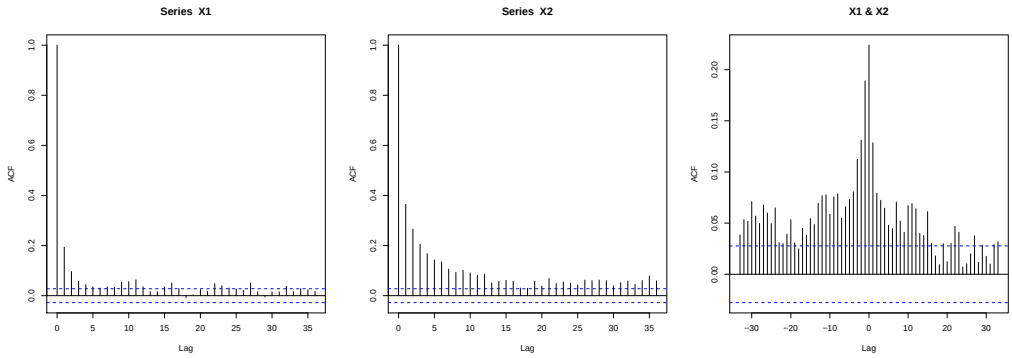

Figure A.8.1: ACF/CCF of the estimated process.

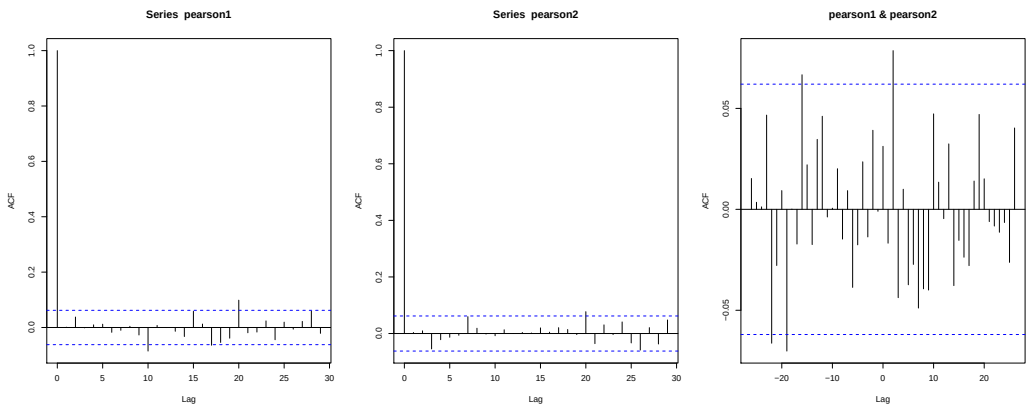

Figure A.8.2: ACF/CCF of the Pearson's residuals.

References

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

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

Böckenholt, U. (1998). Mixed INAR (1) Poisson Regression Models: Analyzing Heterogeneity

and Serial Dependencies in Longitudinal Count Data. *Journal of Econometrics*, 89(1-2):317–338.

Boudreault, M. and Charpentier, A. (2011). Multivariate Integer-Valued Autoregressive Models Applied to Earthquake Counts. *UQAM DP, arXiv preprint arXiv:1112.0929.*

Brännäs, K., Hellström, J., and Nordström, J. (2002). A New Approach to Modelling and Forecasting Monthly Guest Nights in Hotels. *International Journal of Forecasting*, 18(1):19–30.

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

Chaganty, N. R. and Joe, H. (2006). Range of Correlation Matrices for Dependent Bernoulli Random Variables. *Biometrika*, 93(1):197–206.

Cui, Y. and Zhu, F. (2018). A new bivariate integer-valued garch model allowing for negative cross-correlation. *TEST*, 27(2):428–452.

Darolles, S. (2018). Liquidity Risk and Investor Behavior: Issues, Data and Models. *Autorité des Marchés Financiers Scientific Advisory Board Review.*

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

Desmettre, S., de Kock, J., Ruckdeschel, P., and Seifried, F. T. (2018). Generalized Pareto Processes and Liquidity. *Quantitative Finance*, 18(8):1327–1343.

Doukhan, P., Fokianos, K., Støve, B., and Tjøstheim, D. (2017). Multivariate Count Autoregression. *University of Cergy-Pontoise DP, arXiv:1704.02097.*

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

Edwards, C. B. and Gurland, J. (1961). A Class of Distributions Applicable to Accidents. *Journal of the American Statistical Association*, 56(295):503–517.

Genest, C. and Nešlehová, J. (2007). A Primer on Copulas for Count Data. *Astin Bulletin*, 37(2):475–515.

Giraitis, L., Kokoszka, P., and Leipus, R. (2000). Stationary ARCH models: Dependence Structure and Central Limit Theorem. *Econometric theory*, 16(1):3–22.

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

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

Gouriéroux, C. and Lu, Y. (2018). Negative Binomial Autoregressive Process with Stochastic Intensity. *Journal of Time Series Analysis*.

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

Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*. CRC Press.

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

Kemp, C. and Papageorgiou, H. (1982). Bivariate Hermite Distributions. *Sankhyā: The Indian Journal of Statistics, Series A*, 44(2):269–280.

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

Latour, A. (1997). The Multivariate GINAR (p) Process. *Advances in Applied Probability*, 29(1):228–248.

Liu, H. (2012). *Some Models for Time Series of Counts*. PhD thesis, Columbia University.

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

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

Marshall, A. W. and Olkin, I. (1985). A Family of Bivariate Distributions Generated by the Bivariate Bernoulli Distribution. *Journal of the American Statistical Association*, 80(390):332–338.

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

McCabe, B. P., Martin, G. M., and Harris, D. (2011). Efficient Probabilistic Forecasts for Counts. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):253–272.

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

Pedeli, X. and Karlis, D. (2013a). On Composite Likelihood Estimation of a Multivariate INAR (1) Model. *Journal of Time Series Analysis*, 34(2):206–220.

Pedeli, X. and Karlis, D. (2013b). Some Properties of Multivariate INAR (1) Processes. *Computational Statistics & Data Analysis*, 67:213–225.

Quoreshi, A. S. (2017). A Bivariate Integer-Valued Long-Memory Model for High-Frequency Financial Count Data. *Communications in Statistics-Theory and Methods*, 46(3):1080–1089.

Schmidt, L., Timmermann, A., and Wermers, R. (2016). Runs on Money Market Mutual Funds. *American Economic Review*, 106(9):2625–57.

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

Scotto, M. G., Weiß, C. H., Silva, M. E., and Pereira, I. (2014). Bivariate Binomial Autoregressive Models. *Journal of Multivariate Analysis*, 125:233–251.

Securities and Exchange Commission (2015). Open-End Fund Liquidity Risk Management Programs; Swing Pricing; Re-Opening of Comment Period for Investment Company Reporting Modernization Release.

Trivedi, P. and Zimmer, D. (2017). A Note on Identification of Bivariate Copulas for Discrete Count Data. *Econometrics*, 5(1):1–11.

Weiss, C. H. (2018). *An Introduction to Discrete-valued Time Series*. John Wiley & Sons.

Wicksell, S. (1916). Some theorems in the theory of probability, with special reference to their importance in the theory of homograde correlations. *Svenska Aktuarieforeningens Tidskrift*, pages 165–213.

Zaffaroni, P. (2004). Stationarity and Memory of ARCH (∞) Models. *Econometric Theory*, 20(1):147–160.