n° 2018-03
Negative Binomial Autoregressive Process
C.GOURIEROUX[1]
Y.LU[2]

[1] University of Toronto; Toulouse School of Economics; CREST. E-mail: gouriero@ensae.fr
[2] University of Paris 13, Department of Economics (CEPN). E-mail: luyang000278@gmail.com

Negative Binomial Autoregressive Process

Christian Gouriéroux,[*] and Yang Lu[†]

March 5, 2018

Abstract: We introduce Negative Binomial Autoregressive (NBAR) processes for (univariate and bivariate) count time series. The univariate NBAR process is defined jointly with an underlying intensity process, which is autoregressive gamma. The resulting count process is Markov, with negative binomial conditional and marginal distributions. The process is then extended to the bivariate case with a Wishart autoregressive matrix intensity process. The NBAR processes are Compound Autoregressive, which allows for simple stationarity condition and quasi-closed form nonlinear forecasting formulas at any horizon, as well as a computationally tractable generalized method of moment estimator. The model is applied to a pairwise analysis of weekly occurrence counts of a contagious disease between the greater Paris region and other French regions.

Keywords: Negative Binomial Process, Autoregressive Gamma, Poisson-Gamma Conjugacy, Intensity, Compound Autoregressive Process, Common Factor, Pairwise Analysis, Health Insurance.

MSC code: 62-15, JEL code: C32

---

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

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

1 Introduction

Since the seminal paper of McKenzie (1985), time series models for nonnegative integers have become more and more popular in domains such as marketing [Bockenholt (1998)], economics [Blundell et al. (1999)], finance [Heinen and Rengifo (2007)] and insurance [Gouriéroux and Jasiak (2004)]. There are two approaches in the literature. The first one is based on the (Poisson, first-order) Integer Autoregressive [INAR(1)] process [McKenzie (1985), (1988), Al-Osh and Azaid (1987)]. This model has the following advantages: i) the stationary distribution is tractable and ii) the multi-step-ahead predictive density has a closed form expression. However the INAR(1) process features conditional under-dispersion and is thus not suitable for data with conditional over-dispersion.

Another popular model is the Poisson autoregression [see Rydberg and Shephard (2000), Fokianos et al. (2009)] as well as its higher-order extensions called INGARCH [Ferland et al. (2006)]. Such models assume that, given the past history $\underline{X_t}$, variable X_{t+1} is Poisson distributed. Thus the Poisson autoregression features neither conditional over-dispersion, nor under-dispersion. Moreover it does not allow for tractable multi-step-ahead nonlinear prediction formulas.

Although many extensions of the univariate INAR(1) and Poisson autoregressions have been proposed to allow for more flexible conditional or marginal distributions [see McKenzie (1986), Zhu (2011)], the literature on multivariate count processes is rather sparse. The existing models either are not suitable for conditional over-dispersed data [see Pedeli and Karlis (2013) for bivariate INAR and Doukhan et al. (2017) for bivariate Poisson autoregression], or do not allow for tractable stationarity condition and forecasting formulas [see Heinen and Rengifo (2007); Doukhan et al. (2017)].

The aim of this paper is to introduce a new family of (univariate and bivariate) processes, called the Negative Binomial AutoRegressive (NBAR) processes, which solves the two difficulties above. We explore the Poisson-gamma conjugacy to define a univariate count process with a gamma type intermediate stochastic intensity process, and then use the matrix extension of the gamma distribution, that is the Wishart distribution, to define a bivariate count process. We show that the NBAR process belongs to the class of Compound AutoRegressive (CaR) processes,

which facilitates: i) the analysis of stationarity; ii) the Laplace transform based GMM estimation; iii) the nonlinear probabilistic forecasting at any horizon.

The paper is organized as follows. In Section 2, we introduce the univariate NBAR process. The univariate model is extended to the bivariate case in Section 3. Section 4 discusses the parameter identification and statistical inference of the NBAR model. Section 5 proposes a real data application to analyze the diffusion of chickenpox in neighbouring French regions. Section 6 concludes. Proofs are gathered in Appendices.

2 Univariate NBAR process

In this section we introduce a first-order Markov process for univariate count data called Negative Binomial Autoregressive process (NBAR). This terminology is motivated by the property that both the transition distribution at horizon 1 and the stationary distribution (*i.e.* the conditional distribution at infinite horizon) are negative binomial.

2.1 Dynamic specification

The model is based on the Poisson-gamma conjugacy, reviewed in Appendix 1. We denote by $\mathcal{P}(\lambda)$ the Poisson distribution with parameter λ, by $\gamma(\delta, \beta, c)$ the gamma distribution with degree of freedom δ, noncentrality parameter β and scale parameter c, and by $NB(\delta, \beta)$ the negative binomial (NB) distribution with positive parameters δ, β. This parametrization relies on the interpretation of the NB distribution as a Poisson distribution with gamma stochastic intensity, with δ the degree of freedom of the gamma distribution, and β the scale of this intensity. The process (X_t) is defined jointly with a real positive intensity process (Y_t) in the following way:

Definition 1. The process (X_t) is NBAR if:

i) the conditional distribution of X_{t+1} given $\underline{Y_{t+1}}, \underline{X_t}$ is Poisson $\mathcal{P}(\beta Y_{t+1})$,

ii) the conditional distribution of Y_{t+1} given $\underline{X_t}, \underline{Y_t}$ is centered gamma with shape parameter $\delta + X_t$ and scale parameter c: $\gamma(\delta + X_t, 0, c)$,

where $\underline{X_t} = (X_t, X_{t-1}, ..., X_1, ...)$, $\underline{Y_t} = (Y_t, Y_{t-1}, ..., Y_1, ...)$, and β, δ, c are positive scalars.

3

In other words (Y_{t+1}) is the stochastic intensity of the count X_{t+1} at time $t+1$, which depends on past count X_t. Thus we have a causal chain :

$$\ldots X_t \to Y_{t+1} \to X_{t+1} \to Y_{t+2} \ldots$$

In this chain each variable depends on all previous variables via its nearest left neighbor only. As a consequence, both the count process (X_t) and the intensity process (Y_t) are individually Markov with respect to their own histories and their respective dynamics are the following ones:

Proposition 1. *1. The count process (X_t) is Markov and the conditional distribution X_{t+1} given X_t is negative binomial $NB(\delta + X_t, \beta c)$.*

2. The intensity process (Y_t) is Markov and the conditional distribution of Y_{t+1} given Y_t is noncentered gamma $\gamma(\delta, \beta Y_t, c)$.

Proof. The proposition is an immediate consequence of Property A.1 in Appendix 1. □

Such an intensity process (Y_t) is called (first order) AutoRegressive Gamma, or ARG(1) [see Gouriéroux and Jasiak (2006)]. It is the exact time discretization of the continuous time Cox-Ingersoll-Ross process. Using the iterated expectation formula. We have:

$$\mathbb{E}[Y_{t+1}|Y_t] = \mathbb{E}\Big[\mathbb{E}[Y_{t+1}|X_t]|Y_t\Big] = \mathbb{E}[c(\delta + X_t)|Y_t] = c\delta + \beta c Y_t, \tag{2.1}$$

$$\mathbb{E}[X_{t+1}|X_t] = \mathbb{E}\Big[\mathbb{E}[X_{t+1}|Y_t]|X_t\Big] = \mathbb{E}[\beta Y_t|X_t] = \beta c(\delta + X_t). \tag{2.2}$$

Both processes (X_t) and (Y_t) have a (weak) linear AR(1) representation, with the same positive autocorrelation coefficient $\rho = \beta c$, if the latter is strictly smaller than 1.

The transition distributions are characterized by their (real) Laplace transforms[1], that are:

$$\mathbb{E}[\exp(-uX_{t+1})|X_t] = \frac{1}{[1 + \beta c(1 - \exp(-u))]^{\delta + X_t}}, \qquad \forall u \geq 0, \tag{2.3}$$

$$\mathbb{E}[\exp(-uY_{t+1})|Y_t] = \frac{1}{(1 + cu)^{\delta}} \exp(-\frac{\beta cu}{1 + cu} Y_t), \qquad \forall u \geq 0. \tag{2.4}$$

---

[1]Let us recall that the Laplace transform with nonnegative argument characterizes the distribution of a positive variable [Feller (1968)].

Both log-Laplace transforms are affine functions of the conditioning variable. Thus both processes are Compound AutoRegressive (CaR) [Darolles, Gouriéroux, Jasiak (2006)], also called affine processes in the continuous time framework [see Duffie, Filipovic, Schachermayer (2003)] and thinning based processes in the count process literature [see Steutel and van Harn (1979), Latour (1998), Zhu and Joe (2010)]. The following corollary explores the link between the NBAR process and its two alternatives within the CaR framework, that are the INAR(1) process and the (linear) Poisson autoregression.

Corollary 1. *The NBAR process satisfies the representation:*

$$X_{t+1} = \sum_{i=1}^{X_t} Z_{i,t+1} + \epsilon_{t+1}, \tag{2.5}$$

where given X_t, variables $(Z_{i,t+1})$ are i.i.d. with $NB(1, \beta c)$ distribution (that is a geometric distribution with probability $\frac{\beta c}{\beta c+1}$), whereas (ϵ_{t+1}) is i.i.d., independent of $Z_{i,t+1}$, with $NB(\delta, \beta c)$ distribution.

This result follows from Proposition 1 and the infinite divisibility of the NB distribution.

Conversely, any process (X_t) satisfying (2.5), with $Z_{i,t+1}$ and ϵ_{t+1} mutually independent and independent of X_t, not necessarily negative binomial distributed, is CaR since:

$$\mathbb{E}[e^{-uX_{t+1}}|X_t] = \left(\mathbb{E}[e^{-uZ_{1,t+1}}]\right)^{X_t} \mathbb{E}[e^{-u\epsilon_{t+1}}],$$

is exponential affine in X_t. Thus both INAR(1) and Poisson autoregression are CaR. Indeed:

- The process (X_t) is Poisson INAR(1) if $Z_{i,t+1}$ follows Bernoulli distribution with parameter p and ϵ_{t+1} follows $\mathcal{P}(\lambda)$;

- The process (X_t) is Poisson autoregression if $Z_{i,t+1}$ follows $\mathcal{P}(\lambda_0)$ and ϵ_{t+1} follows $\mathcal{P}(\lambda_1)$.

2.2 Conditional overdispersion

Because of its stochastic intensity, the NBAR process (X_t) features conditional over-dispersion. Indeed the conditional variance is

$$\mathbb{V}[X_{t+1}|X_t] = \mathbb{E}\Big[\mathbb{V}[X_{t+1}|X_t, Y_{t+1}] \mid X_t\Big] + \mathbb{V}\Big[\mathbb{E}[X_{t+1} \mid X_t, Y_{t+1}] \mid X_t\Big] = \beta c(\delta + X_t) + \beta^2 c^2(\delta + X_t)$$

Thus the coefficient of (conditional) over-dispersion is:

$$\frac{\mathbb{V}[X_{t+1}|X_t]}{\mathbb{E}[X_{t+1}|X_t]} = 1 + \beta c > 1. \tag{2.6}$$

It is constant independent of the conditioning variable X_t. The larger the serial correlation $\rho = \beta c$, the more important the conditional over-dispersion.

As a comparison, using the representation (2.5), it is easily checked that a Poisson INAR(1) process is conditionally under-dispersed, since for such a process:

$$\mathbb{V}[X_{t+1}|X_t] = p(1 - p)X_t + \lambda \le \mathbb{E}[X_{t+1}|X_t] = pX_t + \lambda,$$

whereas a Poisson autoregression features neither conditional under-dispersion, nor conditional over-dispersion, since:

$$\mathbb{E}[X_{t+1}|X_t] = \lambda_0 X_t + \lambda_1 = \mathbb{V}[X_{t+1}|X_t].$$

2.3 Term structure of nonlinear predictions and the stationarity

The advantage of CaR process is to facilitate the computation of nonlinear predictions at any horizon. Such non-linear forecasts are essential for count data, as linear expectations are inconsistent with the integer state space [see McCabe et al. (2011)]. The following proposition is proved in Appendix 2.

Proposition 2. *At horizon h, the transition of the count process is such that for any $u \geq 0$:*

$$\mathbb{E}[e^{-uX_{t+h}}|X_t] = \frac{\left(1 + \rho\frac{1-\rho^{h-1}}{1-\rho}[1 - \exp(-u)]\right)^{X_t}}{\left(1 + \rho\frac{1-\rho^h}{1-\rho}[1 - \exp(-u)]\right)^{\delta+X_t}} = \frac{\left(1 + \beta c_{h-1}[1 - \exp(-u)]\right)^{X_t}}{\left(1 + \beta c_h[1 - \exp(-u)]\right)^{\delta+X_t}}, \quad (2.7)$$

where $\rho = \beta c$, and the sequence (c_h) is defined by $c_h = c\frac{1-\rho^h}{1-\rho}$.

Let us now provide the corresponding $h-$step-ahead probability mass function (p.m.f.). We first write the conditional Laplace transform as an infinite sum:

$$\mathbb{E}[\exp(-uX_{t+h})|X_t] = \sum_{n=0}^{\infty} e^{-un}\mathbb{P}[X_{t+h} = n|X_t]. \quad (2.8)$$

The RHS of equation (2.8) is analytic in e^{-u}, thus $\mathbb{P}[X_{t+h} = n|X_t]$ is equal to the coefficient of the term e^{-nu} in the Taylor's expansion in e^{-u} of the LHS. By equation (2.7), this expansion is equal to:

$$\left(\frac{1+\beta c_{h-1}}{1+\beta c_h}\right)^{X_t} \frac{\left[1 - \frac{\beta c_{h-1}}{1+\beta c_{h-1}}e^{-u}\right]^{X_t}}{\left[1 - \frac{\beta c_h}{1+\beta c_h}e^{-u}\right]^{\delta+X_t}}$$

$$= \left(\frac{1+\beta c_{h-1}}{1+\beta c_h}\right)^{X_t} \sum_{i=0}^{X_t} \binom{X_t}{i}\left(-\frac{\beta c_{h-1}}{1+\beta c_{h-1}}e^{-u}\right)^i \sum_{j=0}^{\infty}\left(\frac{\beta c_h}{1+\beta c_h}e^{-u}\right)^j \frac{\Gamma(\delta+X_t+j)}{\Gamma(\delta+X_t)j!}.$$

Thus we have the next corollary:

Corollary 2. *The $h-$step-ahead conditional p.m.f. is given by:*

$$\mathbb{P}[X_{t+h} = n|X_t] = \left(\frac{1+\beta c_{h-1}}{1+\beta c_h}\right)^{X_t} \sum_{i=0}^{\min(n,X_t)}\left[\binom{X_t}{i}\left(-\frac{\beta c_{h-1}}{1+\beta c_{h-1}}\right)^i\left(\frac{\beta c_h}{1+\beta c_h}\right)^{n-i}\frac{\Gamma(\delta+X_t+n-i)}{\Gamma(\delta+X_t)(n-i)!}\right].$$

To get the stationarity condition, let us get back to Proposition 2. If $\rho < 1$, then $c_h \xrightarrow{h\infty} \frac{c}{1-\rho}$. Thus for large h, the Laplace transform $\mathbb{E}[e^{-uX_{t+h}}|X_t]$ converges to $\frac{1}{[1+\beta c(1-e^{-u})]^\delta}$ and we get the next corollary:

Corollary 3. *Both the counting process (X_t) and the intensity process (Y_t) are strongly stationary if $\rho = \beta c < 1$. The stationary distributions are the centered gamma distribution $\gamma(\delta, 0, \frac{c}{1-\beta c})$*

for the intensity process, and the negative binomial distribution $NB(\delta, \beta c)$ for the count process, respectively.

The count process is such that both the short term transition $(h = 1)$ and the long term transition $(h = \infty)$ are negative binomial, which motivates the terminology of NBAR. However, the transitions at intermediate horizons are not NB.

As an illustration, we plot, in Figure 1, a simulated sample path of the joint process (X_t, Y_t) with parameters set to $c = 1$, $\beta = 0.69$, and $\delta = 1.3$.
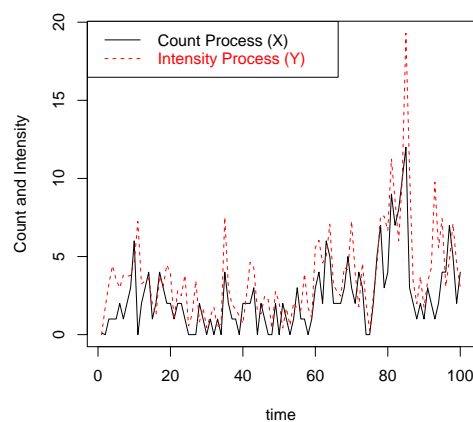

Figure 1: Paths of count process (X_t) (full line) and intensity process (Y_t) (dashed line). The two paths are plotted at the same scale.

In the simulation, the magnitude of the autocorrelation $\rho = \beta c = 0.69$ [see equations (2.1), (2.2)] has been chosen smaller than, but rather close to 1 to ensure stationarity and high large intensity and count clusterings.

2.4 Comparison with other negative binomial processes

Our model is different from other existing Markov processes with NB stationary distribution. Most of these models are also based on representation (2.5), but involve rather complicated p.m.f. for variables ϵ_{t+1} and/or $Z_{j,t+1}$. For instance,

- McKenzie (1986) assume that $Z_{j,t+1}$ is binomial, but $\epsilon_{j,t+1}$ follows a mixture distribution

8

of the point mass at zero and a geometric distribution;

- Al-Osh and Aly (1992) assume that the distribution of $Z_{j,t+1}$ is a mixture of point mass at zero and a geometric distribution, whereas the distribution of ϵ_t is deduced under the assumption of NB marginal distribution.

- Ristic and Nastić (2009) assume that $Z_{j,t+1}$ is geometric, but the distribution of ϵ_{t+1} is a mixture of two geometric distributions.

- Zhu and Joe (2010) propose a model in which $Z_{j,t+1}$ and ϵ_{t+1} have intractable p.m.f. and are defined via Laplace transforms.

- Finally, Joe (1996) consider a model in which $Z_{j,t+1}, j = 1, ...$ are independent and Bernoulli distributed conditionally on a stochastic, beta distributed parameter α, whereas ϵ_{t+1} is NB. This latter model does not have the representation (2.5) due to the dependence induced by the common stochastic parameter.

Finally, these models do not reflect the Poisson-gamma mixture interpretation of the NB distribution. As a consequence, none of them allow for tractable higher-horizon nonlinear forecasting formulas and it is not clear whether they can be extended to the bivariate case.

3 Bivariate NBAR processes

Let us now extend the NBAR process to the bivariate case. In Section 3.1 we first consider a model in which the two component processes share a single gamma common intensity factor. Then in Section 3.2 we generalize the approach to a model with a matrix valued common factor, by using the matrix-variate extension of the gamma distribution, that is the Wishart distribution.

3.1 Bivariate count process with single common intensity factor

3.1.1 The model

Definition 2. The bivariate NBAR process $X_t = (X_{1,t}, X_{2,t})'$ with single common intensity factor Y_t is such that:

- Given $\underline{X_t}$, variables $Z_{1,t+1}$, $Z_{2,t+1}$, Y_{t+1} are independent, with gamma conditional distributions $\gamma(\delta_1 + X_{1t}, c_1)$, $\gamma(\delta_2 + X_{2t}, c_2)$ and $\gamma(\delta_3 + \alpha_1 X_{1,t} + \alpha_2 X_{2,t}, c_3)$, respectively.

- $X_{1,t+1}$ and $X_{2,t+1}$ are independent conditionally on $\underline{X_t}, \underline{Y_{t+1}}, \underline{Z_{1,t+1}}$, and $\underline{Z_{2,t+1}}$. Their conditional distributions are Poisson, with parameters $\beta_1 Y_{t+1} + \kappa_1 Z_{1,t+1}$ and $\beta_2 Y_{t+1} + \kappa_2 Z_{2,t+1}$, respectively,

where $\delta_1, \delta_2, \delta_3, c_1, c_2, c_3, \beta_1, \beta_2, \alpha_1, \alpha_2, \kappa_1, \kappa_2$ are nonnegative parameters.

Thus both conditional Poisson intensities are sums of two components: the first components $\beta_j Y_{t+1}, j = 1, 2$, depend on the common factor, whereas the second components $\kappa_j Z_{j,t+1}, j = 1, 2$ are conditionally independent and depend only on the past history via the individual count processes X_{1t} (resp. $X_{2,t}$). These are the specific intensities.

As in the univariate case, the joint process $(X_t) = (X_{1,t}, X_{2,t})$ is Markov and we have the following causal scheme:

$$
\begin{array}{ccccccccc}
X_{1,t-1} & \longrightarrow & Z_{1,t} & \longrightarrow & X_{1,t} & \longrightarrow & Z_{1,t+1} & \longrightarrow & X_{1,t+1} \\
 & \searrow & & \nearrow & & \searrow & & \nearrow & \\
 & & Y_t & & & & Y_{t+1} & & \\
 & \nearrow & & \searrow & & \nearrow & & \searrow & \\
X_{2,t-1} & \longrightarrow & Z_{2,t} & \longrightarrow & X_{2,t} & \longrightarrow & Z_{2,t+1} & \longrightarrow & X_{2,t+1}
\end{array}
\tag{2.9}
$$

Example 1: In the special case $\kappa_1 = \kappa_2 = 0$, $\alpha_1 = \alpha_2 = 1$, we can check that given Y_{t+1}, $X_{1,t} + X_{2,t}$ is $\mathcal{P}\big((\beta_1 + \beta_2) Y_{t+1}\big)$, whereas given $X_{1,t} + X_{2,t}$, intensity Y_{t+1} is $\gamma(\delta_3 + X_{1,t} + X_{2,t}, 0, c_3)$. Thus the sum process $(X_{1,t} + X_{2,t})$ is univariate NBAR, the intensity process (Y_t) is ARG(1), and the conditional distribution of $X_{1,t}$ given $(X_{1,t} + X_{2,t}, Y_t)$ is binomial $Bin(X_{1,t} + X_{2,t}, \frac{\beta_1}{\beta_1 + \beta_2})$, which is independent of Y_t.

3.1.2 Linear prediction

By iterated expectation we derive the linear prediction:

$$
\begin{aligned}
\mathbb{E}[X_{t+1}|X_t] &= \mathbb{E}[Y_t|X_t] \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \kappa_1 \delta_1 c_1 + \kappa_1 c_1 X_{1,t} \\ \kappa_2 \delta_2 c_2 + \kappa_2 c_2 X_{2,t} \end{pmatrix} \\
&= \underbrace{\begin{pmatrix} \kappa_1 c_1 + \alpha_1 \beta_1 c_3 & \alpha_1 \beta_2 c_3 \\ \alpha_2 \beta_1 c_3 & \kappa_2 c_2 + \alpha_2 \beta_2 c_3 \end{pmatrix}}_{:=A} X_t + \begin{pmatrix} \kappa_1 \delta_1 c_1 + \beta_1 \delta_3 c_3 \\ \kappa_2 \delta_2 c_2 + \beta_2 \delta_3 c_3 \end{pmatrix},
\end{aligned} \tag{2.10}
$$

The matrix A plays a key role when studying the stationarity properties of the process (see subsection 3.1.4).

3.1.3 Nonlinear predictions

The conditional Laplace transform of X_{t+1} given X_t is:

$$
\mathbb{E}[\exp(-u_1 X_{1,t+1} - u_2 X_{2,t+1})|X_t] = \exp\Big(-a_1(u_1,u_2)X_{1,t} - a_2(u_1,u_2)X_{2,t} - b(u_1,u_2)\Big),
$$

where:
$$
a_1(u_1,u_2) = \log[1 + (1 - e^{-u_1})\kappa_1 c_1] + \alpha_1 \log\Big(1 + \beta_1 c_3(1 - e^{-u_1}) + \beta_2 c_3(1 - e^{-u_2})\Big),
$$

$$
a_2(u_1,u_2) = \log[1 + (1 - e^{-u_2})\kappa_2 c_2] + \alpha_2 \log\Big(1 + \beta_1 c_3(1 - e^{-u_1}) + \beta_2 c_3(1 - e^{-u_2})\Big),
$$

$$
b(u_1,u_2) = \delta_1 \log[1 + (1 - e^{-u_1})\kappa_1 c_1] + \delta_2 \log[1 + (1 - e^{-u_2})\kappa_2 c_2]
$$
$$
+ \delta_3 \log\Big(1 + \beta_1 c_3(1 - e^{-u_1}) + \beta_2 c_3(1 - e^{-u_2})\Big).
$$

The bivariate NBAR with single common factor is also compound autoregressive. By applying this formula recursively, we get the higher horizon Laplace transform, which is also exponential affine in X_t:

Corollary 4. *The conditional Laplace transform of (X_t) at horizon h is:*

$$
\mathbb{E}[e^{-u_1 X_{1,t+h} - u_2 X_{2,t+h}}|X_t] = \exp\Big(-a_1^{(h)}(u_1,u_2)X_{1,t} - a_2^{(h)}(u_1,u_2)X_{2,t} - b^{(h)}(u_1,u_2)\Big), \tag{2.11}
$$

$$\text{where: } a_1^{(h)}(u_1, u_2) = a_1(a_1^{(h-1)}(u_1, u_2), a_2^{(h-1)}(u_1, u_2)),$$

$$a_2^{(h)}(u_1, u_2) = a_2(a_1^{(h-1)}(u_1, u_2), a_2^{(h-1)}(u_1, u_2)),$$

$$b^{(h)}(u_1, u_2) = b(a_1^{(h-1)}(u_1, u_2), a_2^{(h-1)}(u_1, u_2)), \qquad h \geq 2.$$

This formula can be used for non-linear prediction at any horizon.

3.1.4   Stationarity condition

Let us now discuss the stationarity of the bivariate count process. First, a necessary condition for mean-variance stationarity is that the means $\mu_1 = \mathbb{E}[X_{1,t}]$ and $\mu_2 = \mathbb{E}[X_{2,t}]$ are positive and finite. In this case they satisfy the following system, obtained by taking expectations in equation (2.10):

$$(1 - \beta_1 c_3 \alpha_1 - \kappa_1 c_1)\mu_1 = \beta_1 c_3 \delta_3 + \kappa_1 c_1 \delta_1 + \beta_1 \alpha_2 c_3 \mu_2,$$

$$(1 - \beta_2 c_3 \alpha_2 - \kappa_2 c_2)\mu_2 = \beta_2 c_3 \delta_3 + \kappa_2 c_2 \delta_2 + \alpha_1 \beta_2 c_3 \mu_1.$$

Thus a necessary condition for mean-variance stationarity is:

$$1 - \kappa_1 c_1 - \alpha_1 \beta_1 c_3 > 0, \qquad \text{and} \qquad 1 - \kappa_2 c_2 - \alpha_2 \beta_2 c_3 > 0. \tag{2.12}$$

A necessary and sufficient condition for the strict stationarity of a CaR process has been derived in Darolles, Gouriéroux and Jasiak (2006), Prop. 7.

Proposition 3 (DGJ (2006)). *The process (X_t) is strictly stationary if and only if all the eigenvalues of matrix:*

$$M := \begin{pmatrix} \frac{\partial a_1(0,0)}{\partial u_1} & \frac{\partial a_1(0,0)}{\partial u_2} \\ \frac{\partial a_2(0,0)}{\partial u_1} & \frac{\partial a_2(0,0)}{\partial u_2} \end{pmatrix},$$

are in modulus smaller than 1.

For a bivariate NBAR with single intensity, we check that $M = A$, where matrix A is defined in (2.10). Thus Proposition 3 can be equivalently written as follows:

Corollary 5. *The process (X_t) is strictly and mean-variance stationary if and only if the parameters satisfy inequalities (2.12) and*

$$(1 - \kappa_1 c_1 - \alpha_1 \beta_1 c_3)(1 - \kappa_2 c_2 - \alpha_2 \beta_2 c_3) > \alpha_1 \alpha_2 \beta_1 \beta_2 c_3^2. \qquad (2.13)$$

Proof. See Appendix 3. □

As an illustration of a stationary NBAR process, we plot, in Figure 2, a simulated sample path of the joint process (X_t, Y_t). The parameters are set to $c_1 = c_2 = c_3 = 1$, $\beta_1 = \beta_2 = 0.5$, $\delta = 1.4$, $\alpha_1 = \alpha_2 = 1$, and $\kappa_1 = \kappa_2 = 0.2$.
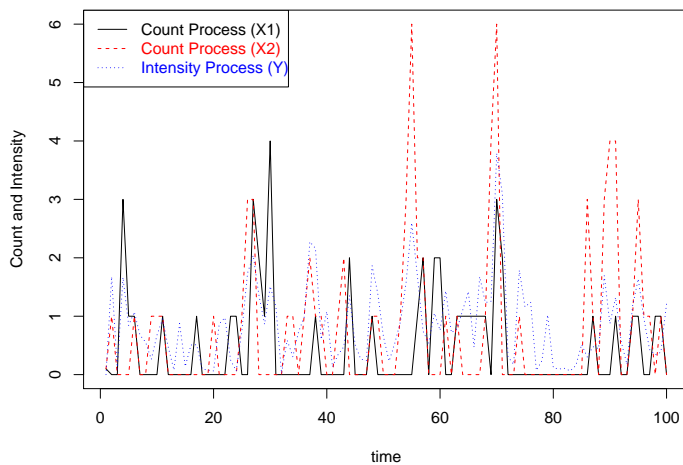

Figure 2: Paths of count processes $(X_{1,t}), (X_{2,t})$, in full and dashed line, respectively, as well as path of intensity process (Y_t), in dotted line.

We observe in Figure 2 the co-movement of the two count processes driven by the single common intensity (Y_t). The magnitude of the serial (cross-)correlation is measured by the maximal modulus of the eigenvalues of matrix A. For this set of parameters the two eigenvalues are equal to 0.8, 0.4, respectively, which explains the serial clustering of intensities as well as of the mortality events of the two types.

13

3.1.5 Conditional overdispersion

Let us now define the notion of conditional over-dispersion in the bivariate framework.

Definition 3: The bivariate process (X_t) features conditional over-dispersion if for any $\theta_1, \theta_2 \geq 0$ such that $\theta_1 + \theta_2 \geq 1$, we have:

$$\mathbb{V}[\theta_1 X_{1,t+1} + \theta_2 X_{2,t+1}|X_t] \geq \mathbb{E}[\theta_1 X_{1,t+1} + \theta_2 X_{2,t+1}|X_t]. \tag{2.14}$$

The definition of over-dispersion in the bivariate framework has to be interpreted in terms of costs: θ_1, θ_2 are the unitary costs of the two types of events and $\theta_1 X_1 + \theta_2 X_2$ their total cost. The condition $\theta_1 + \theta_2 \geq 1$ is introduced, since a measure of over-dispersion is not homogeneous, that is not invariant by a change of numeraire. Thus condition (2.14) is the condition of over-dispersion on the total cost. Typically for car insurance the event of type 1 can correspond to claims for which the policyholder is totally at fault, with normalized cost $\theta_1 = 1$, whereas claims of type 2 are those for which he/she is partially at fault with normalized cost $\theta_2 < 1$.

Condition (2.14) is also equivalent to:

$$\mathbb{V}[\theta X_{1,t+1} + (1-\theta) X_{2,t+1}|X_t] \geq \mathbb{E}[\theta X_{1,t+1} + (1-\theta) X_{2,t+1}|X_t], \tag{2.15}$$

for any $\theta \in [0,1]$, or:

$$\min_{\theta \in [0,1]} \left(\mathbb{V}[\theta X_{1,t+1} + (1-\theta) X_{2,t+1}|X_t] - \mathbb{E}[\theta X_{1,t+1} + (1-\theta) X_{2,t+1}|X_t] \right) \geq 0. \tag{2.16}$$

The LHS is a constrained quadratic optimisation, which admits either an interior, or a boundary solution, depending on the values of the parameters and on those of X_t. This condition is given in Appendix 4.

3.2 Bivariate count process with common matrix intensity factor

3.2.1 The model

Let us now replace the univariate common factor by a symmetric positive definite matrix intensity $Y_t := \begin{pmatrix} Y_{11,t} & Y_{12,t} \\ Y_{12,t} & Y_{22,t} \end{pmatrix}$. The NBAR model is based on the matrix extension of the gamma distribution, *i.e.* the Wishart distribution.

Definition 3. The bivariate NBAR process with symmetric matrix intensity is such that:

- The counts $X_{1,t+1}$ and $X_{2,t+1}$ are conditionally independent given $\underline{Y_{t+1}}$, $\underline{X_t}$ and $(Z_{1,t+1}, Z_{2,t+1})$, with Poisson distributions:

$$X_{1,t+1} =_{(d)} \mathcal{P}\Big(\mathrm{Tr}(BY_{t+1}) + \kappa_1 Z_{1,t+1}\Big), \qquad X_{2,t+1} =_{(d)} \mathcal{P}\Big(\mathrm{Tr}(DY_{t+1}) + \kappa_2 Z_{2,t+1}\Big),$$

  where the symbol Tr denotes the trace of a matrix and matrices B, D are symmetric nonnegative definite [in this case the traces $\mathrm{Tr}(BY_{t+1})$ and $\mathrm{Tr}(DY_{t+1})$ are nonnegative[2]], and parameters κ_1, κ_2 are nonnegative.

- Given $\underline{X_t}$, variables $Z_{1,t+1}, Z_{2,t+1}$ driving the specific intensities are independent, with conditionally gamma distribution $\gamma(\delta_1 + X_{1,t}, c_1)$ and $\gamma(\delta_2 + X_{2,t}, c_2)$, respectively.

- The conditional distribution of the common matrix intensity factor Y_{t+1} is Wishart[3], characterized by the conditional Laplace transform:

$$\mathbb{E}[e^{-\mathrm{Tr}(\Gamma Y_{t+1})}|X_t] = \frac{1}{\det(\mathrm{Id} + \Gamma\Sigma)^{\delta_3 + \alpha' X_t}} \tag{2.17}$$

  where $\alpha' X_t = (\alpha_1, \alpha_2)X_t = \alpha_1 X_{1t} + \alpha_2 X_{2t}$, the matrix scale parameter $\Sigma = (\sigma_{i,j})_{1 \le i,j \le 2}$ is symmetric positive definite, the scalar shape parameter δ_3 is larger or equal to $1/2$, and the argument $\Gamma = (\gamma_{i,j})_{1 \le i,j \le 2}$ is any symmetric positive definite matrix[4].

---

[2]Indeed, by commuting within the trace operator, we have $\mathrm{Tr}(BY_{t+1}) = \mathrm{Tr}(B^{1/2}Y_{t+1}B^{1/2})$, which is the trace of a symmetric positive definite matrix.

[3]This is the Wishart distribution with scale matrix parameter $\Sigma/2$ and degree of freedom parameter $2(\delta_3 + \alpha' X_t)$. In order for it to be properly defined for any values of X_t, we require that $2\alpha_3 > 1$.

[4]For any symmetric positive definite Γ, we have $\det(\mathrm{Id} + \Gamma\Sigma) = \det[\Gamma^{1/2}(\mathrm{Id} + \Gamma^{1/2}\Sigma\Gamma^{1/2})\Gamma^{-1/2}] = \det(\mathrm{Id} + \Gamma^{1/2}\Sigma\Gamma^{1/2}) > 0$ since the last matrix is clearly symmetric positive definite.

The matrix intensity Y_{t+1} has two effects. First, it introduces serial dependency between X_t and X_{t+1} through $\delta_3 + \alpha' X_t$, since $\mathbb{E}[Y_{t+1}|X_t] = (\delta_3 + \alpha' X_t)\Sigma$. On the other hand, given X_t, the different components of Y_{t+1} are correlated. Thus Y_{t+1} introduces also contemporaneous dependence between $X_{1,t+1}$ and $X_{2,t+1}$.

Example 2. Let us consider the case where $B = \text{diag}(1,0)$, $D = \text{diag}(0,1)$, that is, $Tr(BY_{t+1}) = Y_{11,t+1}$, $Tr(DY_{t+1}) = Y_{22,t+1}$. By equation (2.17), the conditional Laplace transform of $(Y_{11,t+1}, Y_{22,t+1})$ given X_t is:

$$
\begin{aligned}
\mathbb{E}[e^{-u_1 Y_{11,t+1} - u_2 Y_{22,t+1}}|X_t] &= \frac{1}{\left[\det(Id + \text{diag}(u_1,u_2)\Sigma)\right]^{\delta_3+\alpha' X_t}} \\
&= \frac{1}{\left[1 + u_1\sigma_{11} + u_2\sigma_{22} + u_1 u_2(\sigma_{11}\sigma_{22} - \sigma_{12}^2)\right]^{\delta_3+\alpha' X_t}}, \qquad \forall u_1, u_2 \geq 0.
\end{aligned}
$$

Both $Y_{11,t+1}$ and $Y_{22,t+1}$ are marginally gamma distributed given X_t, with the same shape parameter. In particular, when $\kappa_1 = \kappa_2 = 0$, the conditional marginal distributions of $X_{1,t+1}$ and $X_{2,t+1}$ given $\underline{X_t}$ are negative binomial. Hence the terminology of **bivariate NBAR** process. The joint distribution of $Y_{11,t+1}$ and $Y_{22,t+1}$ given X_t is called bivariate gamma [see Vere-Jones (1967)]. A simple calculation gives:

$$
\text{Corr}[Y_{11,t+1}, Y_{22,t+1}|X_t] = \frac{\sigma_{12}^2}{\sigma_{11}\sigma_{22}} \geq 0.
$$

which lies between 0 and 1, since Σ is positive definite. Let us now analyze the following two limiting cases:

- If $\sigma_{11}\sigma_{22} - \sigma_{12}^2 = 0$, the Laplace transform is the Laplace transform of a single gamma variable, i.e. $Y_{11,t+1} = Y_{22,t+1}$. We get the bivariate NBAR process with a single stochastic factor (see Section 4.1).

- If $\sigma_{12} = 0$, $Y_{11,t+1}$ and $Y_{22,t+1}$ are conditionally independent and gamma distributed. Then the intensity (Y_t) introduces only specific risks (conditional on X_t).

16

3.2.2 Nonlinear predictions

They can be derived by using the conditional Laplace transform of X_{t+1} given its own past:

$$\mathbb{E}[e^{-u_1 X_{1,t+1} - u_2 X_{2,t+1}} \mid X_t] = \mathbb{E}\Big(\mathbb{E}[e^{-u_1 X_{1,t+1} - u_2 X_{2,t+1}} \mid Y_{t+1}] \mid X_t\Big)$$

$$= \mathbb{E}\Big[\exp\Big(-(1 - e^{-u_1})(\mathrm{Tr}BY_{t+1} + \kappa_1 Z_{1,t+1}) - (1 - e^{-u_2})(\mathrm{Tr}DY_{t+1} + \kappa_2 Z_{2,t+1})\Big) \mid X_t\Big]$$

$$= \exp\Big(-a_1(u_1, u_2)X_{1,t} - a_2(u_1, u_2)X_{2,t} - b(u_1, u_2)\Big), \tag{2.18}$$

where

$$a_1(u_1, u_2) = \log[1 + (1 - e^{-u_1})\kappa_1 c_1] + \alpha_1 \log \det\Big[\mathrm{Id} + (1 - e^{-u_1})B\Sigma + (1 - e^{-u_2})D\Sigma\Big],$$

$$a_2(u_1, u_2) = \log[1 + (1 - e^{-u_2})\kappa_2 c_2] + \alpha_2 \log \det\Big[\mathrm{Id} + (1 - e^{-u_1})B\Sigma + (1 - e^{-u_2})D\Sigma\Big],$$

$$b(u_1, u_2) = \delta_1 \log[1 + (1 - e^{-u_1})\kappa_1 c_1] + \delta_2 \log[1 + (1 - e^{-u_2})\kappa_2 c_2] + \delta_3 \log \det\Big[\mathrm{Id} + (1 - e^{-u_1})B\Sigma + (1 - e^{-u_2})D\Sigma\Big].$$

The bivariate process (X_t) is also CaR, and thus the $h-$step-ahead conditional Laplace transform $\mathbb{E}[e^{-u_1 X_{1,t+h} - u_2 X_{2,t+h}} | X_t]$ can be computed recursively as in equation (2.11). Then, by mimicking Corollary 1, the corresponding conditional p.m.f. $\mathbb{P}[X_{1,t+1} = n_1, X_{2,t+1} = n_2 | X_t]$ can be obtained by considering the double Taylor's expansion of $\mathbb{E}[e^{-u_1 X_{1,t+h} - u_2 X_{2,t+h}} \mid X_t]$ with respect to e^{-u_1} and e^{-u_2}. This calculation is omitted.

Finally, by equation (2.18), process (X_t) has the CaR representation:

$$\begin{pmatrix} X_{1,t+1} \\ X_{2,t+1} \end{pmatrix} = \sum_{i=1}^{X_{1t}} \begin{pmatrix} Z_{1,i,t+1} + Z_{2,i,t+1} \\ Z_{3,i,t+1} \end{pmatrix} + \sum_{j=1}^{X_{2t}} \begin{pmatrix} Z_{4,j,t+1} \\ Z_{5,j,t+1} + Z_{6,j,t+1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t+1} + \epsilon_{3,t+1} \\ \epsilon_{2,t+1} + \epsilon_{4,t+1} \end{pmatrix}, \tag{2.19}$$

where all the terms on the RHS are mutually independent count variables when i and t vary. The distributions of $Z_{1,i,t+1}, Z_{6,j,t+1}, \epsilon_{1,t+1}, \epsilon_{2,t+1}$ are $NB(1, \kappa_1 c_1)$, $NB(1, \kappa_2 c_2)$, $NB(\delta_1, \kappa_1 c_1)$, $NB(\delta_2, \kappa_2 c_2)$, respectively, whereas the distributions of the pairs $(Z_{2,i,t+1}, Z_{3,i,t+1})$, $(Z_{4,i,t+1}, Z_{5,i,t+1})$, and $(\epsilon_{1,t+1}, \epsilon_{2,t+1})]$ have joint Laplace transforms $\exp(-\alpha_1 \mathcal{L}(u_1, u_2))$, $\exp(-\alpha_2 \mathcal{L}(u_1, u_2))$ and $\exp(-\delta_3 \mathcal{L}(u_1, u_2))$, respectively, with function $\mathcal{L}(u_1, u_2)$ given by: $\mathcal{L}(u_1, u_2) = \log \det\Big[\mathrm{Id} + (1 - e^{-u_1})B\Sigma + (1 - e^{-u_2})D\Sigma\Big]$. Moreover all these (univariate or bivariate) variables are mutually independent.

3.2.3 First two conditional moments

The conditional mean of the process is, by the law of iterated expectation:

$$\mathbb{E}[X_{1,t+1}|X_t] = \text{Tr}(B\mathbb{E}[Y_{t+1} \mid X_t]) + \kappa_1(X_{1,t} + \delta_1) = (\delta_3 + \alpha'X_t)\text{Tr}(B\Sigma) + \kappa_1 c_1(X_{1,t} + \delta_1),$$

$$\mathbb{E}[X_{2,t+1}|X_t] = (\delta_3 + \alpha'X_t)\text{Tr}(D\Sigma) + \kappa_2 c_2(X_{2,t} + \delta_2),$$

or in matrix form:

$$\mathbb{E}[X_{t+1}|X_t] = \underbrace{\begin{pmatrix} \alpha_1\text{Tr}(B\Sigma) + \kappa_1 c_1 & \alpha_2\text{Tr}(B\Sigma) \\ \alpha_1\text{Tr}(D\Sigma) & \alpha_2\text{Tr}(D\Sigma) + \kappa_2 c_2 \end{pmatrix}}_{:=A} X_t + \begin{pmatrix} \delta_3\text{Tr}(B\Sigma) + \kappa_1 c_1\delta_1 \\ \delta_3\text{Tr}(D\Sigma) + \kappa_2 c_2\delta_1 \end{pmatrix}. \quad (2.20)$$

The corresponding conditional (co-)variances are given in the next proposition:

Proposition 4.

$$\mathbb{V}[X_{1,t+1}|X_t] = (\delta_3 + \alpha'X_t)\Big[\mathit{Tr}(B\Sigma) + \mathit{Tr}(B\Sigma B\Sigma)\Big] + (\delta_1 + X_{1,t})(\kappa_1 c_1 + \kappa_1^2 c_1^2),$$

$$\mathbb{V}[X_{2,t+1}|X_t] = (\delta_3 + \alpha'X_t)\Big[\mathit{Tr}(D\Sigma) + \mathit{Tr}(D\Sigma D\Sigma)\Big] + (\delta_2 + X_{2,t})(\kappa_2 c_2 + \kappa_2^2 c_2^2),$$

$$Cov[X_{1,t+1}, X_{2,t+1}|X_t] = (\delta_3 + \alpha'X_t)\,\mathit{Tr}(B\Sigma D\Sigma).$$

Proof. See Appendix 5. ☐

Each component is conditionally over-dispersed. Indeed $\mathbb{V}[X_{j,t+1}|X_t] > \mathbb{E}[X_{j,t+1}|X_t]$, $j = 1, 2$, since $\text{Tr}(B\Sigma B\Sigma)$ and $\text{Tr}(D\Sigma D\Sigma)$ are positive[5]. Moreover, since $\text{Tr}(B\Sigma D\Sigma)$ is also positive, the model implies a positive conditional correlation between X_{1t} and X_{2t}.

3.2.4 Stationarity condition

The following proposition is a direct consequence of Proposition 3:

Proposition 5. *The process (X_t) is strictly stationary if and only if the eigenvalues of matrix A defined in (2.20) are smaller than 1 in modulus, or equivalently, if and only if, the following*

---

[5]Since B and $\Sigma B\Sigma$ are both symmetric positive definite.

18

conditions (2.21) *and* (2.22) *are satisfied:*

$$1 - \alpha_1 \, Tr(B\Sigma) - \kappa_1 c_1 > 0, \qquad 1 - \alpha_2 \, Tr(D\Sigma) - \kappa_2 c_2 > 0, \tag{2.21}$$

$$\Big[1 - \alpha_1 \, Tr(B\Sigma) - \kappa_1 c_1\Big]\Big[1 - \alpha_2 \, Tr(D\Sigma) - \kappa_2 c_2\Big] > \alpha_1 \alpha_2 \, Tr(B\Sigma) \, Tr(D\Sigma). \tag{2.22}$$

3.2.5 Level of conditional overdispersion

Let us now analyze the coefficient of conditional overdispersion. In the univariate NBAR model, this coefficient, given by (2.6), is necessarily smaller than 2 under the stationarity condition. This is no longer the case in the Wishart intensity based bivariate NBAR model. For instance, if $\kappa_1 = \kappa_2 = 0$, $\Sigma = Id_2$, then coefficient of conditional over-dispersion over X_{1t} is:

$$1 + \frac{\mathrm{Tr}(B^2)}{\mathrm{Tr}(B)} = 1 + \frac{b_{11}^2 + b_{22}^2 + 2b_{12}^2}{\mathrm{Tr}(B)} \leq 1 + \frac{[\mathrm{Tr}(B)]^2}{\mathrm{Tr}(B)} = 1 + \mathrm{Tr}(B). \tag{2.23}$$

since B is nonnegative definite, and there is equality if and only if $\det B = b_{11}b_{22} - b_{12}^2$ is zero. Thus a sharp upper bound of (2.23) is $1 + \mathrm{Tr}(B)$. On the other hand, for this model, the stationarity condition of Proposition 5 becomes

$$\alpha_1 \mathrm{Tr}(B) + \alpha_2 \mathrm{Tr}(D) < 1.$$

Thus $\mathrm{Tr}(B)$ and $\mathrm{Tr}(D)$ can take arbitrarily large values so long as α_1, α_2 take correspondingly small values. Thus the coefficient of conditional over-dispersion can be as large as possible.

3.2.6 Stationary distribution

The bivariate stationary distribution has no tractable expression (unless in the special case of Example 2). Nevertheless, the two first marginal moments of the latter can be easily derived from their conditional counterparts derived in section 3.2.3. They are summarized in the following corollary, whose proof can be found in a standard textbook on VAR [see Lütkepohl (2005)]:

Corollary 6. *The expectation of the process is given by:*

$$\begin{pmatrix} \mathbb{E}[X_{1,t}] \\ \mathbb{E}[X_{2,t}] \end{pmatrix} = (Id_2 - A)^{-1} \begin{pmatrix} \delta_3 \, Tr(B\Sigma) + \kappa_1 c_1 \delta_1 \\ \delta_3 \, Tr(D\Sigma) + \kappa_2 c_2 \delta_2 \end{pmatrix},$$

and the sequence of cross-covariance functions

$$\Gamma(h) := \begin{pmatrix} \mathbb{C}ov(X_{1,t+h}, X_{1t}) & \mathbb{C}ov(X_{1,t+h}, X_{2t}) \\ \mathbb{C}ov(X_{2,t+h}, X_{1t}) & \mathbb{C}ov(X_{2,t+h}, X_{2t}) \end{pmatrix}, \qquad \forall h \geq 0$$

are equal to:

$$\Gamma(0) = \sum_{h=0}^{\infty} A^h \Big[V_1 + \mathbb{E}[X_{1t}]V_2 + \mathbb{E}[X_{2t}]V_3 \Big] (A')^h, \tag{2.24}$$

and

$$\Gamma(h) = A^h \Gamma(0), \qquad \forall h \geq 1,$$

where the symmetric positive definite matrices V_1, V_2, V_3 are defined by:

$$V_1 = \delta_3 \left(\begin{bmatrix} Tr(B\Sigma) + Tr(B\Sigma B\Sigma) \end{bmatrix} & Tr(B\Sigma D\Sigma) \\ Tr(B\Sigma D\Sigma) & \begin{bmatrix} Tr(D\Sigma) + Tr(D\Sigma D\Sigma) \end{bmatrix} \end{pmatrix} \right) + \begin{pmatrix} \delta_1(\kappa_1 c_1 + \kappa_1^2 c_1^2) & 0 \\ 0 & \delta_2(\kappa_2 c_2 + \kappa_2^2 c_2^2) \end{pmatrix},$$

$$V_2 = \alpha_1 \left(\begin{bmatrix} Tr(B\Sigma) + Tr(B\Sigma B\Sigma) \end{bmatrix} & Tr(B\Sigma D\Sigma) \\ Tr(B\Sigma D\Sigma) & \begin{bmatrix} Tr(D\Sigma) + Tr(D\Sigma D\Sigma) \end{bmatrix} \end{pmatrix} \right) + \begin{pmatrix} \kappa_1 c_1 + \kappa_1^2 c_1^2 & 0 \\ 0 & 0 \end{pmatrix},$$

$$V_3 = \alpha_2 \left(\begin{bmatrix} Tr(B\Sigma) + Tr(B\Sigma B\Sigma) \end{bmatrix} & Tr(B\Sigma D\Sigma) \\ Tr(B\Sigma D\Sigma) & \begin{bmatrix} Tr(D\Sigma) + Tr(D\Sigma D\Sigma) \end{bmatrix} \end{pmatrix} \right) + \begin{pmatrix} 0 & 0 \\ 0 & \kappa_2 c_2 + \kappa_2^2 c_2^2 \end{pmatrix}.$$

3.2.7 Smoothing the common factor

The common factor (Y_t) is the source of contemporaneous conditional dependence between the two count variables. Thus it is useful to infer the unobserved path of the common factor, that is $\ell(Y_t|\underline{X_t})$, when analyzing the common risk, or systemic risk. Due to the Markov chain structure [see (2.9)], the conditional density of Y_t given $\underline{X_t}$ depends only on X_t and X_{t-1}. By the Bayes

theorem, this density is equal to:

$$\ell(y_t|\underline{x_t}) \propto \ell(y_t|x_{t-1})\ell(x_t|x_{t-1}, y_t), \tag{2.25}$$

where $\ell(y_t|x_{t-1})$ is the density of a Wishart distribution $W_2\big(\Sigma/2, 2(\delta_3 + \alpha'x_{t-1})\big)$. Then we remark that X_{1t} and X_{2t} are independent given X_{t-1} and Y_t, and both conditional distributions are the convolution of a Poisson distribution and a negative binomial distribution. Thus we have:

$$\ell(x_t|x_{t-1}, y_t) = \left[\sum_{i=0}^{x_{1t}} \frac{Tr^i(By_t)e^{-Tr(By_t)}}{i!} \frac{\kappa_1^{x_{1t}-i}\Gamma(\delta_1 + x_{1t} - i + x_{1,t-1})}{(x_{1t}-i)!\Gamma(\delta_1 + x_{1,t-1})(\kappa_1 + 1)^{\delta_1 + x_{1t} - i + x_{1,t-1}}} \right]$$
$$\times \left[\sum_{j=0}^{x_{2t}} \frac{Tr^j(Dy_t)e^{-Tr(Dy_t)}}{j!} \frac{\kappa_2^{x_{2t}-j}\Gamma(\delta_2 + x_{2t} - j + x_{2,t-1})}{(x_{2t}-j)!\Gamma(\delta_2 + x_{2,t-1})(\kappa_2 + 1)^{\delta_2 + x_{2t} - j + x_{1,t-1}}} \right].$$

Since Wishart random matrices can be simulated using standard statistical packages, $\mathbb{E}[Y_t|\underline{X_t}]$ can be approximated using importance sampling. More precisely, let $y_{j,t}, j = 1, ..., N$ be i.i.d. draws from the Wishart distribution $W_2\big(\Sigma/2, 2(\delta_3 + \alpha'x_{t-1})\big)$, then we have:

$$\frac{\sum_{j=1}^{N} y_{j,t}\ell(x_t|x_{t-1}, y_{j,t})}{\sum_{j=1}^{N} \ell(x_t|x_{t-1}, y_{j,t})} \xrightarrow[N\to\infty]{} \mathbb{E}[Y_t|\underline{X_t}]. \tag{2.26}$$

4 Statistical Inference

In this section we first discuss the identification of parameters in both the univariate and the bivariate models. Then we discuss the maximum likelihood estimation approach and a GMM approach based on conditional Laplace transforms.

4.1 Identification of the univariate NBAR model

As NBAR models involve unobservable intensity processes, we need to discuss which parameters can be identified from the observation of the count process(es) only. Since the conditional Laplace transform characterizes the dynamics of the process, it is equivalent to study how this Laplace transform depends on the model parameters.

Let us first consider the univariate NBAR model, when Y_t is not observed, but the count

process X_t is observed. Since in equation (2.3), the conditional Laplace transform depends on parameters β, c, δ through $\rho := \beta c$ and δ only, β and c are not separately identifiable. Thus, without loss of generality, we can introduce the following identification restriction:

Identification Restriction 1. Scale parameter $c = 1$.

Under this restriction, parameters β and δ are identifiable.

4.2 Identification of the bivariate NBAR model

Let us now turn to the bivariate NBAR model with common matrix intensity factor, when (X_t) is observed but $(Y_t, Z_{1,t}, Z_{2,t})$ are not observed. We prove in Appendix 6 that there is no loss of generality to assume that:

Identification Restriction 2. • Scalar parameters $c_1 = c_2 = 1$.

- Matrix $\Sigma = \mathrm{Id}_2$.

- Matrix B is diagonal, with $b_{11} \geq b_{22}$.

Then we have the following identification theorem:

Proposition 6. *Under Identification Restriction 2, and the extra condition $b_{11} \neq b_{22}$,[6] we can identify the model parameters as follows:*

- *Parameters $\kappa_1, \kappa_2, \alpha_1, \alpha_2, \delta_1, \delta_2, \delta_3$ are identified;*

- *b_{11} and b_{22} are identified;*

- *d_{11} and d_{22} are identified;*

- *d_{12} is identified up to a sign, i.e., we can assume without loss of generality that d_{12} is nonnegative.*

The proof, as well as the discussion of the case where $b_{11} = b_{22}$, are provided in Appendix 6.

4.3 Maximum likelihood estimation

We provide below the expressions of the log-likelihood functions. In particular, for bivariate NBAR models, these functions involve series expansions and can be computationally demanding.

---

[6]That is to say, $\mathrm{Tr}^2(B) - 4 \det(B) = (b_{11} + b_{22})^2 - 4 b_{11} b_{22} = (b_{11} - b_{22})^2 > 0$.

4.3.1 Univariate NBAR

For observed counts $X_1, X_2, ..., X_T$, the log-likelihood function is:

$$\log \ell(\theta) = \sum_{t=2}^{T} \log p(x_t | x_{t-1}, \theta), \qquad (2.27)$$

where $p(x_t | x_{t-1}, \theta)$ is the p.m.f. of a $NB(x_{t-1} + \delta, \beta)$:

$$p(x_t | x_{t-1}, \theta) = \frac{\beta^{x_t} \Gamma(\delta + x_t + x_{t-1})}{\Gamma(\delta + x_{t-1})(\beta + 1)^{\delta + x_t + x_{t-1}} x_t!},$$

and $\theta = (\beta, \delta)$ is the identifiable model parameter.

4.3.2 Bivariate NBAR with single common factor

After integrating out the value of Y_t, the conditional p.m.f. of the bivariate NBAR with single common factor is:

$$
p(x_t | x_{t-1}, \theta) = \sum_{i=0}^{x_{1t}} \sum_{j=0}^{x_{2t}} \left[\frac{\kappa_1^{x_{1t}-i} \Gamma(\delta_1 + x_{1t} - i + x_{1,t-1})}{(x_{1t}-i)! \Gamma(\delta_1 + x_{1,t-1})(\kappa_1 + 1)^{\delta_1 + x_{1t} - i + x_{1,t-1}}} \frac{\kappa_2^{x_{2t}-j} \Gamma(\delta_2 + x_{2t} - j + x_{2,t-1})}{(x_{2t}-j)! \Gamma(\delta_2 + x_{2,t-1})(\kappa_2 + 1)^{\delta_2 + x_{2t} - j + x_{1,t-1}}} \right.
$$
$$
\times \left. \int_0^\infty \frac{e^{-\beta_1 y}(\beta_1 y)^i}{i!} \frac{e^{-\beta_2 y}(\beta_2 y)^j}{j!} \frac{y^{\delta + \alpha_1 x_{1,t-1} + \alpha_2 x_{2,t-1}} e^{-y}}{\Gamma(\delta + \alpha_1 x_{1,t-1} + \alpha_2 x_{2,t-1})} dy \right]
$$
$$
= \sum_{i=0}^{x_{1t}} \sum_{j=0}^{x_{2t}} \left[\frac{\kappa_1^{x_{1t}-i} \kappa_2^{x_{2t}-j} \Gamma(\kappa_1 + x_{1t} - i + x_{1,t-1}) \Gamma(\kappa_2 + x_{2t} - j + x_{2,t-1})}{\Gamma(\kappa_1 + x_{1,t-1})(\kappa_1 + 1)^{\kappa_1 + x_{1t} - i + x_{1,t-1}} \Gamma(\kappa_2 + x_{2,t-1})(\kappa_2 + 1)^{\kappa_2 + x_{2t} - j + x_{1,t-1}} i! j! (x_{1t} - i)! (x_{2t} - j)!} \right.
$$
$$
\times \left. \frac{\Gamma(\delta + \alpha_1 x_{1,t-1} + \alpha_2 x_{2,t-1} + x_{1t} + x_{2t} - i - j)}{\Gamma(\delta + \alpha_1 x_{1,t-1} + \alpha_2 x_{2,t-1})(1 + \beta_1 + \beta_2)^{x_{1t} + x_{2t} - i - j}} \right].
$$

4.3.3 Bivariate NBAR with common matrix intensity factor

The same approach is still possible but is more computationally intensive. We get:

$$
p(x_t | x_{t-1}, \theta) = \sum_{i=0}^{x_{1,t}} \sum_{j=0}^{x_{2,t}} \frac{\kappa_1^{x_{1,t}-i} \Gamma(\delta_1 + x_{1,t} - i + x_{1,t-1}) \kappa_2^{x_{2,t}-j} \Gamma(\delta_2 + x_{2,t} - j + x_{2,t-1})}{\Gamma(\delta_1 + x_{1,t-1})(\kappa_1 + 1)^{\delta_1 + x_{1,t} - i + x_{t-1}} \Gamma(\delta_2 + x_{2,t-1})(\kappa_2 + 1)^{\delta_2 + x_{2,t} - j + x_{t-1}} i! j! (x_{1t} - i)! (x_{2t} - j)!}
$$
$$
\times \mathbb{E}\left[(\mathrm{Tr} BY_t)^i (\mathrm{Tr} DY_t)^j e^{-\mathrm{Tr} DY_t - \mathrm{Tr} BY_t} | x_{t-1}, \theta \right], \qquad (2.28)
$$

where

$$\mathbb{E}\left[(\mathrm{Tr}BY_t)^i(\mathrm{Tr}DY_t)^j e^{-\mathrm{Tr}DY_t-\mathrm{Tr}BY_t}|x_{t-1},\theta\right] = (-1)^{i+j}\frac{\partial^{i+j}}{\partial s_1^i \partial s_2^j}\mathbb{E}\left[e^{-s_1\mathrm{Tr}DY_t-s_2\mathrm{Tr}BY_t}|x_{t-1},\theta\right]|_{s_1=s_2=0}$$

$$= (-1)^{i+j}\frac{\partial^{i+j}}{\partial s_1^i \partial s_2^j}\frac{1}{\det(\mathrm{Id}+s_1B+s_2D)^{\delta_3+\alpha'x_{t-1}}}|_{s_1=s_2=0}\;.$$

For each pair (i,j), the RHS of the equation above is a function of $x_{1,t-1}$ and $x_{2,t-1}$ only. Its functional form can be obtained by a symbolic computation package such as Mathematica. In total, to compute the likelihood function, we have to compute $(N_1+1)(N_2+1)$ partial derivatives symbolically, where N_1 (resp. N_2) is the maximal observed value of process (X_{1t}) (resp. of (X_{2t})). Thus the downside of this ML approach is that when N_1 and N_2 are large, higher order partial derivatives tend to be complicated, and the number of partial derivatives to compute is large. The Generalized Method of Moments (GMM) proposed below is still consistent without this computational burden.

4.4   Laplace transform based GMM

In the literature of (continuously valued) affine processes, GMM has been suggested by Singleton (2001), who used moment conditions derived from the conditional characteristic function. Here, since the process (X_t) is nonnegative, we use instead the (real) conditional Laplace transform of X_t given X_{t-1}. Let us denote by $\mathcal{U}$ a grid of positive real numbers; then, by the conditional Laplace transform formula we get a set of moment conditions:

$$\mathbb{E}\left[Z_{u,v}(X_{1,t-1},X_{2,t-1})\left(e^{-uX_{1,t}-vX_{2,t}} - \underbrace{e^{-a_1(u,v)X_{1,t-1}-a_2(u,v)X_{2,t-1}-b(u,v)}}_{=\mathbb{E}[e^{-uX_{1,t}-vX_{2,t}}|X_{t-1}]}\right)\right] = 0, \qquad \forall u,v \in \mathcal{U},$$

(2.29)

where $Z_{u,v}(X_{1,t-1},X_{2,t-1})$ can be any instrumental function. For convenience we choose Z to be:

$$Z_{u,v}(X_{1,t-1},X_{2,t-1}) = \frac{1}{\sqrt{\mathbb{V}[e^{-uX_{1,t}-vX_{2,t}}|X_{t-1}]}},$$

(2.30)

24

in order that the integrand in equation (2.29) has unitary unconditional variance. Using the compound autoregressive property, the expression of Z can be obtained explicitly since:

$$\mathbb{V}[e^{-uX_{1,t}-vX_{2,t}}|X_{t-1}] = \mathbb{E}[e^{-2uX_{1,t}-2vX_{2,t}}|X_{t-1}] - \left[\mathbb{E}[e^{-uX_{1,t}-vX_{2,t}}|X_{t-1}]\right]^2$$

$$= \exp\left(-a_1(2u,2v)X_{1,t-1} - a_2(2u,2v)X_{2,t-1} - b(2u,2v)\right) - \exp\left(-2a_1(u,v)X_{1,t-1} - 2a_2(u,v)X_{2,t-1} - 2b(u,v)\right).$$

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

$$\hat{\theta}_T = \arg\min_{\theta} \left[\frac{1}{T-1}\sum_{t=2}^{T} g(X_{t-1}, X_t, \theta)\right]' W \left[\frac{1}{T-1}\sum_{t=2}^{T} g(X_{t-1}, X_t, \theta)\right],$$

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

$$g_{Z,u,v}(X_{t-1}, X_t) = Z_{u,v}(X_{1,t-1}, X_{2,t-1})(e^{-uX_{1,t}-vX_{2,t}} - e^{-a_1(u,v)X_{1,t-1}-a_2(u,v)X_{2,t-1}-b(u,v)}), \qquad \forall u, v \in \mathcal{U}, Z \in \mathcal{Z},$$

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:

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \longrightarrow_{(d)} \mathcal{N}\left(0, (G'WG)^{-1}GW\Omega W'G(G'W'G)^{-1}\right),$$

where θ_0 is the true parameter value, $G = \mathbb{E}[\nabla g(X_{t-1}, X_t, \theta_0)]$, with ∇ representing the differential with respect to argument θ, and $\Omega = \mathbb{E}[g(X_{t-1}, X_t, \theta_0)g(X_{t-1}, X_t, \theta_0)']$. Moreover, the optimal choice of the weighting matrix W is given by $W = \Omega^{-1}$, and for this choice we have:

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \longrightarrow_{(d)} \mathcal{N}\left(0, (G'\Omega^{-1}G)^{-1}\right).$$

Finally, the model parameters need to satisfy the nonlinear stationarity constraints given by Proposition 5. In Appendix 7, we propose an alternative re-parametrization of the model that ensures that these conditions are automatically satisfied.

5 Application to the diffusion of a disease

5.1 The data

We consider chickenpox cases in French regions. Chickenpox, also called varicella, is a highly contagious disease which spreads easily through the coughs and sneezes of an infected person. The data are downloaded from the Sentinelles Network[7], which is a group of physicians collecting real-time epidemiological data to be used for analysis and forecasting. The website provides, for each region, the weekly number of chickenpox cases, from the first week of 1990 until the first week of 2016, that is a total of 1358 observations. The left panel of Figure 3 plots the histogram of the weekly count in the Ile-de-France region (IdF, the greater Paris Area).
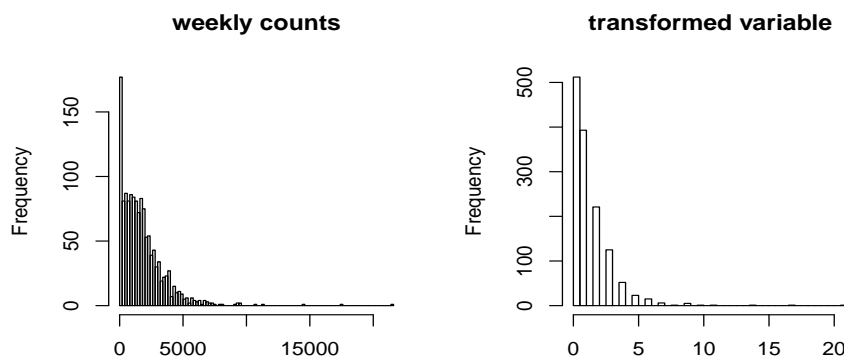

Figure 3: Histogram of the weekly count of chickenpox cases in IdF, as well as the corresponding transformed variable.

This distribution is different from a negative binomial-type distribution due to i) zero-inflation, i.e. a large proportion of weeks with no chickenpox cases; and ii) a high proportion of extremely large observations. Thus the initial weekly counts are not suitable for a NBAR type model. Hence, we consider, for each region with index i ($i = 0$ for Ile-de-France, and $i > 0$ for other regions), the integer part of the weekly count divided by 1000. These new variables can take any nonnegative integer values. The right panel of Figure 3 plots the histogram of the transformed count variable $X_{0,t}$ for IdF.

---

[7]See their website: `https://websenti.u707.jussieu.fr/sentiweb/?lang=en`

The Metropolitan France has a total of 21 regions, among which IdF is the most populous. While the bivariate NBAR model is extendable to higher dimensions, such an extension may suffer from curse of dimensionality. Instead we will focus on the pairwise analysis between IdF and another region, to understand how the chickenpox is transmitted. For expository purpose, we limit ourselves to 6 non-IdF regions: Provence-Alpes-Côte-d'Azur (PACA), Haute Normandie, Rhone-Alpes, Centre and Bourgogne (Burgundy). These regions are either neighbors of the IdF (such as Haute Normandie and Burgundy), or are well connected with Paris, both economically, and through high speed train (such as PACA and Rhone-Alpes, which house the Marseille and Lyon Metropolitan Areas, respectively).

Table 1 reports summary statistics of the count processes $(X_{i,t})$ of these regions. For comparison purpose, we also provide the size of the population of each region (in million).

	Mean	Variance	Maximum	$\frac{\text{Variance}}{\text{Mean}}$	Population	$\frac{\text{Mean}}{\text{Population}}$
Ile-de-France	1.32	2.87	21	2.17	12.1	0.110
Picardie	0.189	0.453	9	2.38	1.8	0.105
Bourgogne	0.0603	0.0744	4	1.23	1.6	0.037
Centre	0.136	0.197	4	1.44	2.5	0.054
Haute-Normandie	0.133	0.218	5	1.64	1.8	0.073
Provence-Alpes-Côte-d'Azur	0.634	1.35	14	2.13	4.8	0.132
Rhône-Alpes	0.715	0.915	7	1.28	6.0	0.119

Table 1: Summary Statistics

We see that i) all count variables $X_{i,t}$ feature (marginal) over-dispersion; ii) the marginal mean of variable $X_{0,t}$ is significantly larger than that of other regions. This can be explained by either the superior size of the IdF, or the larger concentration of the population.

The next figure plots the auto-correlation (ACF) and cross-correlation functions (CCF) of the two count processes corresponding to IdF and PACA, respectively.
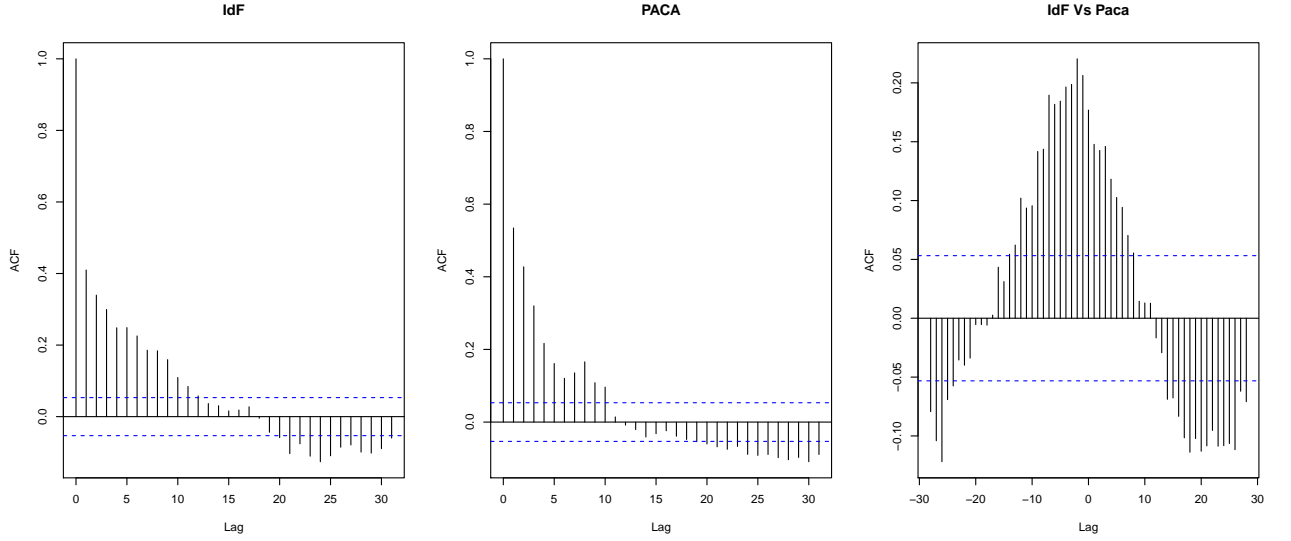
Figure 4: ACF and CCF of the two count processes.

The two ACF's show that both processes have rather persistent autocorrelation. They are positive up to two months and a half, and become negative around 6 months. The CCF confirms that there are Granger directional causality from one process to the other, as well as instantaneous causality.

5.2 Estimation of bivariate NBAR

For each region $i = 1, ..., 6$, we estimate a bivariate NBAR model with matrix common factor for the joint process (X_{0t}, X_{it}). We use the GMM approach, and choose the grid set $\mathcal{U}$ defining the instruments to be 0.1, 0.2,...,1. In order to account for the stationary condition, we first reparameterize the parameters as discussed in Appendix 7. Once the model is estimated, we use the inverse transformation to recover the corresponding estimates of the initial parameters. Table 2 reports parameter estimates for each pair $(X_{0,t}, X_{i,t})$.

	Picardie	Bourgogne	Centre	Haute Normandie	PACA	Rhone Alpes
κ_1	0.119	0.018	0.091	0.057	0.514	0.098
	(0.053)	(0.005)	(0.039)	(0.029)	(0.047)	(0.015)
κ_2	0.047	0.162	0.040	0.038	0.023	0.204
	(0.010)	(0.015)	(0.026)	(0.039)	(0.018)	(0.027)
α_1	0.035	0.019	0.064	0.226	0.809	0.018
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.019)
α_2	0.484	3.041	0.956	0.094	0.543	0.531
	(0.012)	(0.013)	(0.006)	(0.011)	(0.004)	(0.032)
b_{11}	0.830	1.041	0.986	0.840	0.150	0.737
	(0.197)	(0.181)	(0.133)	(0.23)	(0.214)	(0.154)
b_{22}	0.203	0.019	0.382	0.193	0.011	0.012
	(0.007)	(0.114)	(0.099)	(0.046)	(0.164)	(0.090)
d_{11}	0.136	0.038	0.140	0.065	0.204	0.227
	(0.190)	(0.028)	(0.064)	(0.113)	(0.157)	(0.075)
d_{22}	0.009	0.007	0.074	0.027	0.017	0.08
	(0.061)	(0.038)	(0.002)	(0.017)	(0.058)	(0.066)
d_{12}	0.011	0.010	0.081	0.037	0.039	0.095
	(0.050)	(0.030)	(0.006)	(0.028)	(0.034)	(0.008)
δ_1	3.434	1.113	1.660	3.223	0.624	1.422
	(0.397)	(0.204)	(0.315)	(0.426)	(0.174)	(0.492)
δ_2	1.744	0.088	0.377	1.488	4.588	0.716
	(0.182)	(0.004)	(0.030)	(0.025)	(0.129)	(0.159)
δ_3	0.535	1.046	0.540	0.653	0.711	0.929
	(0.003)	(0.009)	(0.035)	(0.013)	(0.040)	(0.007)

Table 2: Parameter estimates along with the standard deviations provided below each model estimate in parenthesis.

The results of the pairwise analysis have to be interpreted with caution. Each pairwise model i, $i = 1, ..., 6$ is a dynamic model for the series (X_{0t}, X_{1t}) with the same benchmark X_{0t}, corresponding to greater Paris. When we focus on the transition joint p.m.f., the information sets $(\underline{X_{0t}}, \underline{X_{it}})$ depends on i, as the parameters and the underlying factor. These ones would have to be indexed by i, $i = 1, ..., n$, even if this index has been omitted for expository purpose. These pairwise models are well specified, if they can be derived from a joint model for $(X_{0t}, X_{1t}, ..., X_{6t})$, which is not explicitly written here. The existence of such an underlying joint model implies minimal coherence restrictions between the pairwise models. Roughly speaking, these pairwise models have to be such that the resulting individual dynamics of the benchmark process (X_{0t}) does not depend on i [see Gouriéroux and Monfort (2017) for a discussion of such coherency conditions in a similar framework with stochastic volatilities]. We have checked that different

(cross-)moments of X_{0t} are approximated in a same way in the pairwise analyses. For instance we provide in Table 3 first row the estimated expectation of X_{0t}. They are not significantly different and are close to the observed historical mean. The same remark applies to the estimates of $\mathbb{V}[X_{0t}]$ and of $\mathrm{corr}(X_{0t}, X_{0,t-1})$, whose theoretical values are given by Corollary 6.

	Picardie $i=1$	Bourgogne $i=2$	Centre $i=3$	Haute Normandie $i=4$	PACA $i=5$	Rhone Alpes $i=6$
Model implied $\mathbb{E}[X_{0t}]$	1.25	1.24	1.21	1.21	1.33	1.23
Empirical mean of X_{0t}	1.31	1.31	1.31	1.31	1.31	1.31
Model implied $\mathbb{E}[X_{it}]$	0.19	0.04	0.17	0.14	0.58	0.75
Empirical mean of X_{it}	0.19	0.06	0.14	0.13	0.63	0.72
Model implied $\mathbb{V}[X_{0t}]$	1.92	2.39	2.88	2.67	2.47	2.56
Empirical variance of X_{0t}	2.87	2.87	2.87	2.87	2.87	2.87
Model implied $\mathbb{V}[X_{it}]$	0.21	0.06	0.24	0.23	1.01	1.22
Empirical variance of X_{it}	0.45	0.07	0.19	0.21	1.35	0.91
Model implied $\mathrm{corr}(X_{0t}, X_{0,t-1})$	0.35	0.27	0.35	0.27	0.23	0.12
Empirical autocorrelation of X_{0t}	0.41	0.41	0.41	0.41	0.41	0.41

Table 3: Model-implied theoretical moments vs their empirical counterparts

The other parameters that are not characterizing the dynamics of (X_{0t}) depend in general on i. In other words, each of these parameters include a fixed effect, linked to the size of the region i or the density of its population. But the size and density are not sufficient to explain the heterogeneity of the κ_2 coefficient, which is the partial derivative of $\mathbb{E}[X_{1,t+1}|Y_{t+1}, Z_{1,t+1}]$ with respect to the standardized specific factor $Z_{1,t+1}$.

By Proposition 5, the joint process is stationary if the largest eigenvalue of matrix A defined in (2.20) is smaller than one in modulus. The estimated eigenvalues are reported in the following table:

	Picardie	Bourgogne	Centre	Haute Normandie	PACA	Rhone-Alpes
The larger eigenvalue	0.39	0.25	0.36	0.28	0.41	0.15
The smaller eigenvalue	0.05	0.10	0.08	0.02	0.09	0.06

Table 4: Estimated eigenvalues of matrix A in the pairwise models

The largest estimated eigenvalues are quite similar and all smaller than 1.

5.3 The effects of common and specific factors

The average importance of the common and specific factors can be deduced from the expressions of the conditional first- and second-order moments. We get the following decompositions:

	Common factor	Specific factor
Average effect on $\mathbb{E}[X_{1,t+1}|X_t]$	$(\delta_3 + \alpha'\mathbb{E}[X_t])\mathrm{Tr}(B)$	$\kappa_1(\delta_1 + \mathbb{E}[X_{1t}])$
Average effect on $\mathbb{E}[X_{2,t+1}|X_t]$	$(\delta_3 + \alpha'\mathbb{E}[X_t])\mathrm{Tr}(D)$	$\kappa_2(\delta_2 + \mathbb{E}[X_{2t}])$
Average effect on $\mathbb{V}[X_{1,t+1}|X_t]$	$(\delta_3 + \alpha'\mathbb{E}[X_t])[\mathrm{Tr}(B) + \mathrm{Tr}(B^2)]$	$(\kappa_1 + \kappa_1^2)(\delta_1 + \mathbb{E}[X_{1t}])$
Average effect on $\mathbb{V}[X_{2,t+1}|X_t]$	$(\delta_3 + \alpha'\mathbb{E}[X_t])[\mathrm{Tr}(D) + \mathrm{Tr}(D^2)]$	$(\kappa_2 + \kappa_2^2)(\delta_2 + \mathbb{E}[X_{2t}])$

Table 5: Decompositions of the average contribution of common vs specific factors to conditional first and second moments

These decompositions can be estimated pairwise. As an illustration we report the values for the pair IdF and Picardie in Table 6.

	Common factor	Specific factor
Average effect on $\mathbb{E}[X_{1,t+1}|X_t]$	1.00	0.23
Average effect on $\mathbb{E}[X_{2,t+1}|X_t]$	0.16	0.02
Average effect on $\mathbb{V}[X_{1,t+1}|X_t]$	1.42	0.26
Average effect on $\mathbb{V}[X_{2,t+1}|X_t]$	0.18	0.02

Table 6: Decomposition (relative magnitude) common vs specific factors

From Table 5, we can also decompose the relative overdispersion effect due to the common and specific factors on (X_{1t}), which are $\frac{\mathrm{Tr}(B^2)}{\mathrm{Tr}(B)}$ and κ_1, respectively. Table 7 reports the specific and common effects for different pairwise models. The over-dispersion effect is systematically much larger for the common matrix factor.

	Picardie	Bourgogne	Centre	Haute Normandie	PACA	Rhone-Alpes
Common	1.9	1.2	0.8	1.6	1.0	0.7
Specific	0.061	0.121	0.148	0.271	0.125	0.175

Table 7: Relative over-dispersion effect on X_{0t}, of common vs specific factors

5.4 Recovering the underlying common factor Y_t

Let us now illustrate the smoothing of the intermediate matrix intensity factor in the pairwise model of IdF and Picardie. Figure 5 below plots the paths of $\mathbb{E}[\mathrm{Tr}(BY_t)|\underline{X_t}]$ and $\mathbb{E}[\mathrm{Tr}(DY_t)|\underline{X_t}]$

along with the paths of the two count processes (X_{1t}) and (X_{2t}). These smoothed values are computed using equation (2.26) in which the number of draws at each time t is set to be $N = 500$. The two smoothed intensities capture quite well the dynamics of the two count processes. This is expected since the effect of the common intensity factor is much more important than the region-specific factors.
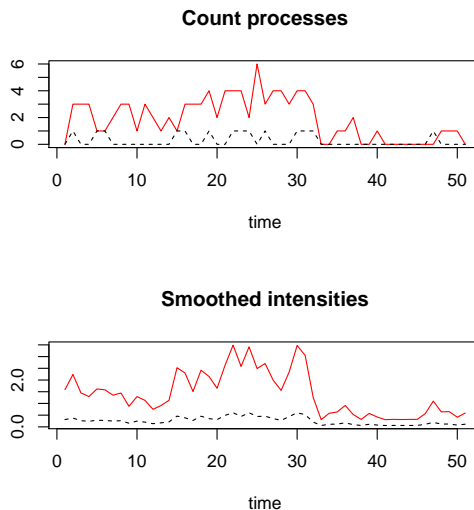
Count processes

Smoothed intensities

Figure 5: The upper panel plots the evolution of (X_{1t}) (in red dotted line) and (X_{2t}) (in black full line) during 40 periods. The lower panel plots the smoothed paths of $\mathrm{Tr}(BY_t)$ (in red dotted line) and $\mathrm{Tr}(DY_t)$ (in black full line).

6   Conclusion

We have introduced the Negative Binomial Autoregressive(NBAR(1)) process as an alternative to the standard INAR(1) process and Poisson autoregressions, which cannot capture conditional over-dispersion. The NBAR has two nice technical features. First, it is introduced jointly with an intensity process by exploring the Poisson-gamma conjugacy. Secondly, it belongs to the family of Compound Autoregressive processes. These two properties lead to tractable formulas for nonlinear prediction as well as stationarity distribution, and allows for simple GMM estimation method based on Laplace transform. The univariate NBAR process can be naturally extended into the bivariate case with a matrix intensity process. The bivariate NBAR(1) model has been

32

illustrated by a pairwise analysis of the diffusion of a disease in France with the Greater Paris as the benchmark region.

Appendix 1 Review on Poisson, gamma and negative binomial distributions

Appendix 1.1 The elementary distributions

a) Poisson distribution $\mathcal{P}(\lambda)$ with parameter $\lambda > 0$:

Its p.m.f. is : $p(x) = \dfrac{\exp(-\lambda)\lambda^x}{x!}$.

Its Laplace transform is : $\mathbb{E}[\exp(-uX)] = e^{-\lambda(1-e^{-u})}$, $\forall u \in \mathbb{N}$.

b) Centered gamma distribution $\gamma(\delta, 0, c)$ with parameter $\delta, c > 0$:

Its density function is : $f_Y(y) = \dfrac{1}{\Gamma(\delta)} \exp(-y/c) \dfrac{y^{\delta-1}}{c^\delta} \mathbf{1}_{y>0}$.

Its Laplace transform is : $\mathbb{E}[\exp(-uY)] = \dfrac{1}{(1+cu)^\delta}$, $\forall u \geq 0$.

c) Noncentral gamma distribution : $\gamma(\delta, \beta, c)$ with parameters $\delta, \beta, c > 0$:

We have $Y \sim \gamma(\delta, \beta, c)$, if and only if we can write:

$$Y|X \sim \gamma(\delta + X, 0, c), \text{ with } X \sim \mathcal{P}(\beta).$$

Its density is : $f_Y(y) = \exp(-y/c - \beta) \dfrac{y^{\delta-1}}{c^\delta} \left\{ \displaystyle\sum_{x=0}^{\infty} \dfrac{1}{\Gamma(\delta+x)} \dfrac{1}{x!} \left(\dfrac{y\beta}{c}\right)^x \right\}$.

Its Laplace transform is : $\mathbb{E}[\exp(-uY)] = \dfrac{1}{(1+cu)^\delta} \exp\left(\dfrac{-\beta cu}{1+cu}\right)$, $\forall u \geq 0$.

d) Negative Binomial distribution : $NB(\delta, \beta)$ with parameters $\delta, \beta > 0$:

We have $X \sim NB(\delta, \beta)$ if and only if we can write:

$$X|Y \sim \mathcal{P}(\beta Y), \text{ with } Y \sim \gamma(\delta, 0, 1),$$

Its p.m.f. is : $p(x) = \dfrac{1}{(1+\beta)^\delta} \left(\dfrac{\beta}{1+\beta}\right)^x \dfrac{\Gamma(x+\delta)}{x!\Gamma(\delta)}$.

Its Laplace transform is : $\mathbb{E}[e^{-uX}] = \dfrac{1}{[1 + \beta(1 - e^{-u})]^\delta}$, $\forall u \geq 0$.

When $\beta \to 0$, $\delta \to \infty$ such that $\beta\delta \to \lambda$, the negative binomial distribution $NB(\delta, \beta)$ reduces to the Poisson distribution $\mathcal{P}(\lambda)$.

Note that we have parametrized the negative binomial distribution in order to get directly its interpretation as a Poisson mixture with a gamma mixing variable. In the literature it is often parametrized by δ and probability parameter $p = \frac{\beta}{\beta+1}$.

Appendix 1.2   A count-continuous distribution

Let us consider a couple (X, Y), where X is a count variable and Y a real positive variable. We assume that its joint density (with respect to $\nu \otimes \lambda^+$, i.e. the product measure between the counting measure ν on $\mathbb{N}$ and the Lebesgue measure λ^+ on $I\!R^+$) is:

$$f(x,y) = \frac{\exp[-y(\beta + \frac{1-\beta c}{c})]}{x!\Gamma(\delta)} y^{x+\delta-1}\beta^x(\frac{1-\beta c}{c})^\delta, \tag{a.31}$$

with $\beta c \neq 1$. This joint density can be written as:

$$f(x,y) = \left[\frac{\exp(-\beta y)(\beta y)^x}{x!}\right] \left[\frac{1}{\Gamma(\delta)}\left(\frac{1-\beta c}{c}\right)^\delta \exp\left(-\frac{1-\beta c}{c}y\right) y^{\delta-1}\right],$$

where the first term on the RHS is the p.m.f. of $\mathcal{P}(\beta Y)$, and the second term is the density of $\gamma(\delta, 0, \frac{c}{1-\beta c})$. Symmetrically, we can write:

$$f(x,y) = \left[\frac{\Gamma(x+\delta)}{x!\Gamma(\delta)}(\beta c)^x(1-\beta c)^\delta\right] \left[\frac{1}{\Gamma(x+\delta)}\frac{y^{x+\delta-1}}{c^{x+\delta}}\exp(-y/c)\right],$$

where the first term on the RHS is the p.m.f. of $NB(\delta, \frac{\beta c}{1-\beta c})$, whereas the second-term is the p.d.f. of $\gamma(\delta + x, 0, c)$. We deduce the following property :

Property A.1 : The pair (X, Y) with joint distribution given in equation (a.1) is such that :

- the conditional distribution of X given Y is : $\mathcal{P}(\beta Y)$;

- the conditional distribution of Y given X is : $\gamma(\delta + X, 0, c)$;

- the unconditional distribution of X is : $NB(\delta, \frac{\beta c}{1-\beta c})$;

- the unconditional distribution of Y is : $\gamma(\delta, 0, \frac{c}{1 - \beta c})$.

Appendix 2   Proof of Proposition 2

Let us proceed by induction. Assume that equation (2.7) holds for a certain h, then:

$$\mathbb{E}[\exp(-uX_{t+h+1})|X_t] = \mathbb{E}\Big[\mathbb{E}[\exp(-uX_{t+h+1})|X_{t+1}]|X_t\Big]$$

$$= \mathbb{E}\Big[\frac{\Big(1 + \beta c_{h-1}(1 - e^{-u})\Big)^{X_{t+1}}}{\Big(1 + \beta c_h(1 - e^{-u})\Big)^{\delta + X_{t+1}}}|X_t\Big] \tag{a.32}$$

$$= \frac{1}{\Big(1 + \beta c_h(1 - e^{-u})\Big)^{\delta}\Big(1 + \beta c \frac{\beta(c_h - c_{h-1})(1 - e^{-u})}{1 + \beta c_h(1 - e^{-u})}\Big)^{\delta + X_t}}, \tag{a.33}$$

$$= \frac{[1 + \beta c_h(1 - e^{-u})]^{X_t}}{\Big[1 + \beta c \frac{1 - (\beta c)^{h+1}}{1 - \beta c}(1 - e^{-u})\Big]^{\delta + X_t}} = \frac{[1 + \beta c_h(1 - e^{-u})]^{X_t}}{\Big[1 + \beta c_{h+1}(1 - e^{-u})\Big]^{\delta + X_t}},$$

where we have used the one-step-ahead conditional Laplace transform formula to pass from equation (a.32) to equation (a.33). Thus equation (2.7) holds also for horizon $h + 1$. Thus it holds for any horizon.

Appendix 3   Proof of Corollary 5

Let us first prove a lemma:

Lemma 1. *Let $A = (a_{i,j})_{1 \leq i,j \leq 2}$ be a matrix with nonnegative elements. Then the eigenvalues of A are of modulus smaller than 1 if and only if:*

$$\alpha_{11} < 1, \alpha_{22} < 1$$

, and $(1 - \alpha_{11})(1 - \alpha_{22}) > \alpha_{12}\alpha_{21}$.

Proof. The eigenvalues are solutions of the determinantal equation:

$$x^2 - x\mathrm{Tr}A + \det A = 0,$$

35

with discriminant: $\Delta = (\text{Tr} A)^2 - 4\det A = (\alpha_{11} - \alpha_{22})^2 + 4\alpha_{12}\alpha_{21} \geq 0$. Thus this polynomial has two real roots (or a double root). They are smaller than 1 in modulus if and only if:

$$-1 < \frac{1}{2}(Tr(A) - \sqrt{\Delta}), \text{ and } \frac{1}{2}(Tr(A) + \sqrt{\Delta}) < 1$$

$$\Longleftrightarrow \sqrt{\Delta} < 2 - Tr(A)$$

$$\Longleftrightarrow \alpha_{11} + \alpha_{22} < 2, \text{ and } \Delta < (2 - Tr(A))^2$$

$$\Longleftrightarrow \alpha_{11} < 1, \text{ and } \alpha_{22} < 1, \text{ and } (1 - \alpha_{11})(1 - \alpha_{22}) > \alpha_{12}\alpha_{21}.$$

$\square$

Finally, we check that Corollary 5 is a direct consequence of Proposition 3 and the above lemma.

Appendix 4   The overdispersion condition

Let us now derive the necessary and sufficient condition for a bivariate count process to be over-dispersed (see Definition 3). Let us denote by v the vector $(c, 1 - c)$, e the unitary vector $(1, 1)$, Σ_t the conditional variance-covariance matrix, and m_t the conditional expectation vector at time t. Then the condition for global over-dispersion is:

$$\min_{v, v'e=1} v'\Sigma_t v - v'm_t \geq 0. \tag{a.34}$$

The solution of the optimization $\min_{v, v'e=1} v'\Sigma_t v - v'm_t$ is the same as that of the optimization of the Lagrangean:

$$\mathcal{L}(\lambda, v) = v'\Sigma_t v - v'm_t - \lambda v'e,$$

where λ denotes the Lagrange multiplier associated with the constraint: $v'e = 1$. Differentiating $\mathcal{L}(\lambda, v)$ with respect to v leads to the optimal value of v: $v = \frac{1}{2}\Sigma_t^{-1}(m_t + \lambda e)$, and the objective

function in (a.34) becomes:

$$\frac{1}{4}(m_t + \lambda e)'\Sigma_t^{-1}(m_t + \lambda e) - \frac{1}{2}m'\Sigma_t^{-1}(m + \lambda e)$$

$$= \frac{1}{4}(m_t + \lambda e - 2m_t)'\Sigma_t^{-1}(m_t + \lambda e)$$

$$= \frac{1}{4}(\lambda e - m_t)'\Sigma_t^{-1}(m_t + \lambda e).$$

Using the constraint $v'e = 1$, we get the optimal value of λ:

$$\lambda = \frac{2 - e'\Sigma_t^{-1}m_t}{e'\Sigma_t^{-1}e}.$$

Thus condition (a.34) is equivalent to:

$$m_t'\Sigma_t^{-1}m_t + \lambda^2 e'\Sigma_t^{-1}e - 2\lambda m_t'\Sigma_t^{-1}e \geq 0$$

$$\Longleftrightarrow (m_t'\Sigma_t^{-1}m_t)(e'\Sigma_t^{-1}e) + (3e'\Sigma_t^{-1}m_t - 2)(e'\Sigma_t^{-1}m_t - 2) \geq 0 \qquad (a.35)$$

The optimisation problem above has been solved without taking into account the constraint that both entries of v are nonnegative. Therefore, the global over-dispersion condition is equivalent to equation (a.35) if both entries of vector:

$$\frac{1}{2}\Sigma_t^{-1}(m_t + \lambda e) = \frac{1}{2}\Sigma_t^{-1}(m_t + \frac{2 - e'\Sigma_t^{-1}m_t}{e'\Sigma_t^{-1}e}e)$$

are nonnegative. It is equivalent to:

$$\mathbb{V}[X_{1,t+1}|X_t] \geq \mathbb{E}[X_{1,t+1}|X_t], \qquad \mathbb{V}[X_{2,t+1}|X_t] \geq \mathbb{E}[X_{2,t+1}|X_t],$$

otherwise.

Appendix 5 Proof of Proposition 4

The conditional variances and covariances of Y_{t+1} given X_t are obtained by differentiating the conditional Laplace transform. We have:

Lemma 2. *For any symmetric positive definite matrix S, we have:*

$$\mathbb{V}[Tr(SY_{t+1})|X_t] = (\delta_3 + \alpha'X_t)\,Tr(S\Sigma S\Sigma) \geq 0. \tag{a.36}$$

The RHS of the equation is necessarily nonnegative since both S and $\Sigma S\Sigma$ are symmetric positive definite. Thus we get:

$$\mathbb{V}[Y_{11,t+1}|X_t] = (\delta_3 + \alpha'X_t)\sigma_{11}^2, \qquad \mathbb{V}[Y_{22,t+1}|X_t] = (\delta_3 + \alpha'X_t)\sigma_{22}^2,$$

$$\mathbb{V}[Y_{12,t+1}|X_t] = (\delta_3 + \alpha'X_t)\frac{\sigma_{11}\sigma_{22} + \sigma_{12}^2}{2}, \qquad \text{Cov}[Y_{11,t+1}, Y_{22,t+1}|X_t] = (\delta_3 + \alpha'X_t)\sigma_{12}^2,$$

$$\text{Cov}[Y_{11,t+1}, Y_{12,t+1}|X_t] = (\delta_3 + \alpha'X_t)\sigma_{11}\sigma_{12}, \qquad \text{Cov}[Y_{22,t+1}, Y_{12,t+1}|X_t] = (\delta_3 + \alpha'X_t)\sigma_{12}\sigma_{22}.$$

Then, by the variance decomposition formula, we get the conditional (co-)variance of X_{t+1} given X_t announced in Proposition 4.

Appendix 6   Model identification

Appendix 6.1   Proof of Proposition 6

Let us first show that we can, without loss of generality, make Identification Restriction 2. Equation (2.18) depends on (κ_1, c_1) (resp. (κ_2, c_2)) only through $\kappa_1 c_1$ (resp. $\kappa_2 c_2$); thus without loss of generality, we can assume $c_1 = c_2 = 1$.

Second, for any symmetric positive definite matrix T, we have:

$$\det\left[\text{Id} + (1 - e^{-u_1})B\Sigma + (1 - e^{-u_2})D\Sigma\right] = \det\left[T[\text{Id} + (1 - e^{-u_1})B\Sigma + (1 - e^{-u_2})D\Sigma]T^{-1}\right]$$

$$= \det\left[\text{Id} + (1 - e^{-u_1})TBTT^{-1}\Sigma T^{-1} + (1 - e^{-u_2})TDTT^{-1}\Sigma T^{-1}\right]. \tag{a.37}$$

Thus we can always replace the triplet (B, D, Σ) by the triplet $(TBT, TDT, T^{-1}\Sigma T^{-1})$, without changing the conditional Laplace transform. As a consequence, since Σ is symmetric positive definite, we can take $T = \Sigma^{1/2}$ and $T^{-1}\Sigma T^{-1}$ becomes Id_2. Thus we can assume, without loss of generality, that $\Sigma = \text{Id}_2$.

Under this identification restriction, equation (a.37) becomes, for any orthogonal matrix O:

$$\det\left[\mathrm{Id} + (1 - e^{-u_1})B + (1 - e^{-u_2})D\right] = \det\left[\mathrm{Id} + (1 - e^{-u_1})OBO' + (1 - e^{-u_2})ODO'\right].$$

Thus B, C are still only identified up to a common orthonormal change of basis. Since B is symmetric, it is diagonalizable within an orthonormal basis. Therefore, we assume without loss of generality that B is diagonal, with $b_{11} \geq b_{22}$. Under these identification restrictions, equation (a.37) becomes:

$$\det\left[\mathrm{Id} + (1 - e^{-u_1})B\Sigma + (1 - e^{-u_2})D\Sigma\right]$$
$$= 1 + (1 - e^{-u_1})(b_{11} + b_{22}) + (1 - e^{-u_2})(d_{11} + d_{22}) + (1 - e^{-u_1})^2 b_{11} b_{22}$$
$$+ (1 - e^{-u_2})^2(d_{11}d_{22} - d_{12}^2) + (1 - e^{-u_1})(1 - e^{-u_2})(d_{11}b_{22} + d_{22}b_{11}) \qquad \text{(a.38)}$$

From equation (a.38), we identify $b_{11} + b_{22}$ and $b_{11}b_{22}$, which allows the identification of b_{11}, b_{22}.

Then the knowledge of $d_{11} + d_{22}$ and $d_{11}b_{22} + d_{22}b_{11}$ identifies d_{11}, d_{22}, so long as $b_{11} \neq b_{22}$.

Finally $d_{11}d_{22} - d_{12}^2 = \det(D)$ identifies d_{12} up to a sign, and the result of Proposition 6 follows.

Appendix 6.2   The limiting case $b_{11} = b_{22}$

Let us now discuss the case where we know *ex-ante* that $b_{11} = b_{22}$.

- If $b_{11} = b_{22} = b$ is unknown, the conditional Laplace transform depends on B and D through

$$\det\left[\mathrm{Id} + (1 - e^{-u_1})B\Sigma + (1 - e^{-u_2})D\Sigma\right]$$
$$= 1 + (1 - e^{-u_1})2b + (1 - e^{-u_2})(d_{11} + d_{22}) + (1 - e^{-u_2})^2(d_{11}d_{22} - d_{12}^2) + (1 - e^{-u_1})(1 - e^{-u_2})b(d_{11} + d_{22}).$$

    Thus we can only identify b, $d_{11} + d_{22}$ and $d_{11}d_{22} - d_{12}^2$, but not separately the entries of D.

39

- If $B = 0$, we have:

$$\det\left[\text{Id} + (1 - e^{-u_1})B\Sigma + (1 - e^{-u_2})D\Sigma\right] = 1 + (1 - e^{-u_2})(d_{11} + d_{22}) + (1 - e^{-u_2})^2(d_{11}d_{22} - d_{12}^2).$$

Thus we can only identify $d_{11} + d_{22}$ and $d_{11}d_{22} - d_{12}^2$, but not separately the entries of D.

Appendix 7 An alternative parameterization

In the bivariate model with matrix common factor, the set of identifiable parameters is:

$$\theta = (\delta_1, \delta_2, \delta_3, \alpha_1, \alpha_2, \kappa_1, \kappa_2, b_{11}, b_{22}, d_{11}, d_{22}, d_{12}).$$

All these parameters are positive[8], and satisfy the nonlinear constraints:

$$b_{11} \geq b_{22}, \qquad d_{11} \geq d_{22}, \qquad d_{11}d_{22} \geq d_{12}^2,$$

as well as stationarity conditions (2.21) and (2.22). Since the numerical optimization with non-linear inequality constraints is usually difficult, let us propose a re-parametrization of the model. First, we set:

$$b_{11} = Tr(B)\rho_b, \qquad b_{22} = Tr(B)(1 - \rho_b), \tag{a.39}$$

$$d_{11} = Tr(D)\rho_d, \qquad d_{22} = Tr(D)(1 - \rho_d), \tag{a.40}$$

$$d_{12} = \sqrt{d_{11}d_{22}}\rho_1, \tag{a.41}$$

where $Tr(B), Tr(D) \geq 0$, $\rho_b, \rho_d \in]0.5, 1[$, $\rho_1 \in]0, 1[$.

Since $1 - \alpha_1 \text{Tr}(B) - \kappa_1$, $1 - \alpha_2 \text{Tr}(D) - \kappa_2$ are positive, we set:

$$\alpha_1 = \frac{z_1}{\text{Tr}(B)}, \qquad \alpha_2 = \frac{z_2}{\text{Tr}(D)}, \tag{a.42}$$

$$\kappa_1 = (1 - z_1)(1 - z_3), \qquad \kappa_2 = (1 - z_2)(1 - z_4), \tag{a.43}$$

---

[8]Parameter d_{12} can be set positive due to the identification analysis.

with $z_1, z_2, z_3, z_4 \in]0, 1[$. However, these four new variables cannot be chosen arbitrarily, since they have to satisfy condition (2.22), or equivalently:

$$z_3 z_4 (1 - z_1)(1 - z_2) > z_1 z_2. \tag{a.44}$$

We can remark that:

$$(1 - z_1)(1 - z_2) > (1 - z_1)(1 - z_2) z_3 z_4 > z_1 z_2,$$

that is $z_1 + z_2 < 1$. Thus we set:

$$z_1 = \rho_2 \rho_3, \qquad z_2 = \rho_2 (1 - \rho_3), \tag{a.45}$$

with $\rho_2 = z_1 + z_2$ and $\rho_3 = \frac{z_1}{z_1 + z_2}$ lying between 0 and 1. Thus (a.44) becomes:

$$z_3 z_4 > \frac{z_1 z_2}{(1 - z_1)(1 - z_2)} \tag{a.46}$$

Thus given z_1, z_2, the possible range of values of z_3 and z_4 is $]\frac{z_1 z_2}{(1-z_1)(1-z_2)}, 1[$. Thus we can set

$$z_3 = \frac{z_1 z_2}{(1 - z_1)(1 - z_2)} \rho_4 + (1 - \rho_4) = \frac{\rho_2^2 \rho_3 (1 - \rho_3)}{(1 - \rho_2 \rho_3)(1 - \rho_2 + \rho_2 \rho_3)} \rho_4 + 1 - \rho_4, \tag{a.47}$$

where $\rho_4 \in]0, 1[$. Hence (a.46) becomes:

$$z_4 > \frac{\frac{z_1 z_2}{(1 - z_1)(1 - z_2)}}{\frac{z_1 z_2}{(1 - z_1)(1 - z_2)} \rho_4 + (1 - \rho_4)}. \tag{a.48}$$

Finally z_4 can be represented as

$$z_4 = \frac{\frac{z_1 z_2}{(1 - z_1)(1 - z_2)}}{\frac{z_1 z_2}{(1 - z_1)(1 - z_2)} \rho_4 + (1 - \rho_4)} \rho_5 + 1 - \rho_5 \tag{a.49}$$

where $\rho_5 \in]0, 1[$. Thus we get a reparametrization:

$$\theta' = (Tr(B), Tr(D), \rho_b, \rho_d, \rho_1, \rho_2, \rho_3, \rho_4, \rho_5, \delta_1, \delta_2, \delta_2)$$

41

We check that, the alternative parameterization (a.40), (a.41), (a.42), (a.43) and (a.45), (a.47), (a.49) leads to a set of parameters θ that satisfy all the nonlinear constraints.

REFERENCES

Al-Osh, M., and A., Azaid (1987) : "First-Order Integer Valued Autoregressive (INAR (1)) Processes", Journal of Time Series Analysis, 8, 261-275.

Al-Osh, M., and E., Aly (1992) : "First-Order Autoregressive Time Series with Negative Binomial and Geometric Marginals", Communications in Statistics-Theory and Methods 21 (9), 2483-2492.

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.

Bockenholt, 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.

Darolles, S., Gouriéroux, C., and J., Jasiak (2006) : "Structural Laplace Transform and Compound Autoregressive Models", Journal of Time Series Analysis, 27, 477-503.

Doukhan, P., Fokianos, K., Stove.B, and D., Tjostheim (2017) : "Multivariate Count Autoregression", University of Cergy-Pontoise DP.

Duffie, D., Filipovic, D., and W., Schachermayer (2003) : "Affine Processes and Applications in Finance", Annals of Applied Probability, 13, 984-1053.

Feller, W. (1968) : "An Introduction to Probability Theory and its Applications", Wiley.

Ferland, R., Latour, A., and D., Oraichi(2006), "Integer-Valued GARCH Process." Journal of Time Series Analysis 27(6), 923-942.

Fokianos, K., Rahbek, A., and D. Tjoestheim (2009) : "Poisson Autoregression", Journal of

the American Statistical Association, 104, 1430-1439.

Gouriéroux, C., and J., Jasiak (2006) : "Autoregressive Gamma Processes", Journal of Forecasting, 25, 129-152.

Gouriéroux, C., and A., Monfort (2017) : "Pairwise Analysis of High Dimensional Stochastic Volatility with Market Benchmark", CREST DP.

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

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

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

Latour, A. (1998) : "Existence and Stochastic Structure of a Non-negative Integer-Valued Autoregressive Process." Journal of Time Series Analysis 19(4), 439-455.

Lütkepohl, H. (2005) : "New Introduction to Multiple Time Series Analysis." Springer Science & Business Media.

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

McKenzie, E. (1985) : "Some Simple Models for Discrete Variate Time Series", Water Resources Bulletin, 21, 645-650.

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

McKenzie, E. (1988) : "Some ARMA Models for Dependent Sequences of Poisson Counts", Advances in Applied Probability, 20, 822-835.

Pedeli, X., and D., Karlis (2013) : "Some Properties of Multivariate INAR(1) Processes",

Computational Statistics and Data Analysis, 67, 213-225.

Ristić M.M., and A.S., Nastić (2009) : "A New Geometric First-Order Integer-Valued Autoregressive (NGINAR(1)) Process", Journal of Statistical Planning and Inference, 139, 2218-2226.

Rydberg, T. H., and N., Shephard (2000) : "A Modeling Framework for the Prices and Times of Trades on the New York Stock Exchange," in Nonlinear and Nonstationary Signal Processing, eds. W. J. Fitzgerlad, R. L. Smith, A. T. Waiden, and P. C. Young, Cambridge: Isaac Newton Institute and Cambridge University Press, 217-246

Singleton, K. (2001) :"Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function", Journal of Econometrics, 102, 111-141.

Steutel, F.W., and K., van Harn (1979) : "Discrete Analogues of Self-Decomposability and Stability", Annals of Probability, 7, 893-899

Vere-Jones, D. (1967) : "The Infinite Divisibility of a Bivariate Gamma Distribution", Sankhya, Series A, 29, 421-442.

Zhu, F. (2011) : "A Negative Binomial Integer-Valued GARCH Model." Journal of Time Series Analysis, 32(1), 54-67.

Zhu, R., and H., Joe (2010) : "Negative Binomial Time Series Models Based on Expectation Thinning Operators," Journal of Statistical Planning and Inference, 140, 1874-1888.