Count and duration time series with equal conditional stochastic and mean orders

Aknouche, Abdelhakim and Francq, Christian

University of Science and Technology Houari Boumediene, CREST
and University of Lille

11 November 2018

Count and duration time series with equal conditional stochastic and mean orders

Abdelhakim Aknouche

University of Science and Technology Houari Boumediene

and

Christian Francq *

CREST and University of Lille

Abstract

We consider a positive-valued time series whose conditional distribution has a time-varying mean, which may depend on exogenous variables. The main applications concern count or duration data. Under a contraction condition on the mean function, it is shown that stationarity and ergodicity hold when the mean and stochastic orders of the conditional distribution are the same. The latter condition holds for the exponential family parametrized by the mean, but also for many other distributions. We also provide conditions for the existence of marginal moments and for the geometric decay of the beta-mixing coefficients. Simulation experiments and illustrations on series of stock market volumes and of greenhouse gas concentrations show that the multiplicative-error form of usual duration models deserves to be relaxed, as allowed in the present paper.

Keywords: Absolute regularity, Autoregressive Conditional Duration, Count time series models, Ergodicity, Integer $GARCH$, Mixing.

1 Introduction

Models for nonnegative time series include the Autoregressive Conditional Duration (ACD) model introduced by Engle and Russell (1998) to analyse durations between events (such as trades, quotes, price changes), the Conditional AutoRegressive Range (CARR) model introduced by Chou (2005) to study the range of an asset during a trading day, the more general Multiplicative Error Model (MEM) introduced by Engle (2002) and count time series models such as the INteger-valued AutoRegressive (INAR) studied by Al-Osh and Alzaid (1987) or the Poisson INteger GARCH (INGARCH) studied by Ferland, Latour and Oraichi (2006). Count time series models have been used in various domains, in particular economics, finance, insurance, environmental science, social science and epidemiology (see Davis, Holan, Lund and Ravishanker (2016) and the references therein). For MEM-like models, the stationary solutions are obtained explicitly, like for GARCH models, as function of the parameters and the rescaled iid innovations of the model (see *e.g.* Francq and Zakoïan, 2010). INGARCH-type count time series models are not defined by means of an iid white noise, but by assuming a discrete conditional distribution with a time-varying parameter depending on the past values. Since the primary goal of these time series models is to forecast the future level of the observed series, that parameter is generally the conditional mean. The absence of an iid sequence in the definition of these models prevents exhibiting an explicit solution. The fact that the support of the conditional distribution is countable also prevents using the theory of the Markov chains with continuous state space (see Meyn and Tweedie, 2012). As a consequence, studying the probabilistic structure of most count time series models is not obvious (see Fokianos, Rahbek and Tjøstheim, 2009, Tjøstheim, 2012, Davis, Holan, Lund and Ravishanker, 2016). Ferland, Latour and Oraichi (2006) obtained stationarity results for INGARCH models with Poisson conditional distribution of linear intensity parameter. Neumann (2011) proved the absolute regularity and relaxed the linearity assumption on the Poisson intensity parameter. Doukhan and Neumann (2017) showed the absolute regularity for a much broader class of processes. Franke (2010) and Doukhan, Fokianos and Tjøstheim (2012, 2013) studied the weak dependence of nonlinear Poisson autoregressions.

Douc, Doukhan and Moulines (2013) gave conditions on the associated Markov kernel for stationarity and ergodicity of a first-order observation-driven time series valued in $\mathbb{N}$. These results have been extended to more general observation-driven models by Douc, Roueff and Sim (2015, 2016) and Sim, Douc and Roueff (2016). Gonçalves, Mendes-Lopes and Silva (2015) showed the stationarity and ergodicity of the INGARCH model with compound Poisson conditional distributions. Davis and Liu (2016) showed stationarity and mixing properties when the conditional distribution belongs to the one-parameter exponential family of distributions. The latter reference builds explicit solutions of count time series as limits of functions of quantiles of an iid sequence. The present paper adopts the same strategy. The assumption that the conditional distribution belongs to the exponential family is however restrictive. In particular, that assumption precludes the zero-inflated distributions and hurdle models, which proved to be useful to deal with count data sets that have an excess of zero counts (see *e.g.* Gurmu and Trivedi, 1996).

The main aim of the present paper is to give stationarity and ergodicity conditions for conditional distributions that are not restricted to belong to the one-parameter exponential family. In addition we will allow the conditional mean to depend on covariates, which seems relevant for some applications.

We thus consider a stochastic process of interest $\{Y_t, t \in \mathbb{Z}\}$ valued in the set $[0, \infty)$, and a stochastic process of exogenous explonatory variables $\{\boldsymbol{X}_t, t \in \mathbb{Z}\}$ valued in $\mathbb{R}^r$. Let $\mathcal{F}_t$ be the information set available at time t, *i.e.* the sigma-field generated by $\{Y_u, \boldsymbol{X}_u, u \leq t\}$. When there is no exogenous variable, *i.e.* when $\mathcal{F}_t = \sigma(Y_u, u \leq t)$, the most frequent specifications of $\lambda_t := E(Y_t \mid \mathcal{F}_{t-1})$ is the linear equation

$$\lambda_t = \omega + \sum_{i=1}^{q} \alpha_i Y_{t-i} + \sum_{j=1}^{p} \beta_j \lambda_{t-j}, \tag{1.1}$$

where $\omega > 0, \alpha_i \geq 0$ and $\beta_j \geq 0$. The standard ACD duration models and MEMs are of the form

$$Y_t = \lambda_t z_t, \tag{1.2}$$

where (λ_t) satisfies (1.1) and (z_t) is an iid sequence of positive variables of mean 1, for

3

instance of exponential distribution of rate parameter 1. Note that for time series of counts, *i.e.* when Y_t is valued in $\mathbb{N}$, the sequence $z_t = Y_t/\lambda_t$ cannot be independent, in general. Even for duration models for which the support of Y_t is $[0, \infty)$, assuming that z_t and λ_t are independent is very restrictive. In particular, this implies that the conditional variance $\mathrm{Var}(Y_t \mid \mathcal{F}_{t-1})$ is proportional to λ_t^2, whatever the distribution of z_t. In the numerical part of this paper, the independence between z_t and λ_t will be assessed by bootstrapping the distance covariance test of Székely, Rizzo and Bakirov (2007). For more versatile duration time series models, it is thus of interest to relax the MEM specification (1.2), by only specifying a conditional distribution with mean λ_t.

We refer to a distribution of Y_t given $\mathcal{F}_{t-1}$ with mean (1.1) as a positive linear POLI(p, q) model. If, as for INGARCH (p, q) models, the distribution of Y_t given $\mathcal{F}_{t-1}$ is integer-valued, the model is intended to represent time series of counts. If, as for the above-mentioned extension of the ACD models, the distribution of Y_t given $\mathcal{F}_{t-1}$ is valued in $(0, \infty)$, the POLI model could suit for some time series of duration or volume, for instance.

Even if many references mention the possibility of adding exogenous variables in count or duration time series models (see *e.g.* Cameron and Trivedi, 2001), we are only aware of few references focusing on exogenous variables, the paper on Poisson autoregresssion with exogenous covariates (PARX) by Agosto, Cavaliere, Kristensen and Rahbek (2016) and that of Liboschik, Fokianos and Fried (2017) which also considers negative binomial conditional distributions and has the R companion package `tscount` (see also the R package `acp` of Siakoulis, 2015). In the PARX model, we have

$$\lambda_t = \omega + \sum_{i=1}^{q} \alpha_i Y_{t-i} + \sum_{j=1}^{p} \beta_j \lambda_{t-j} + \boldsymbol{\pi}^\top \boldsymbol{X}_{t-1}, \tag{1.3}$$

where the components of $\boldsymbol{X}_t = (x_{1,t}, \ldots, x_{r,t})^\top$ are (transformed to) nonnegative numbers and $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_r)^\top \geq 0$ componentwise. We also consider more general specifications of the form

$$\lambda_t = g(Y_{t-1}, \ldots, Y_{t-q}, \lambda_{t-1}, \ldots, \lambda_{t-p}) + \pi(\boldsymbol{X}_{t-1}), \tag{1.4}$$

where the functions g and π are values in $[0, \infty)$.

4

We do not make a specific parametric assumption on the conditional distribution of Y_t given $\mathcal{F}_{t-1}$, but we assume that its stochastic order increases with its mean. More precisely, let F_λ be a family of cumulative distribution functions (cdf) indexed by the mean $\lambda = \int y\, dF_\lambda(y) \in \mathbb{R}$. Assume that, within this family, the stochastic order is equal to the mean order, *i.e.*

$$\lambda \leq \lambda^* \quad \Rightarrow \quad F_\lambda(y) \geq F_{\lambda^*}(y), \quad \forall y \in \mathbb{R}. \tag{1.5}$$

We shall refer to (1.5) as the stochastic-equal-mean order property. Section 2 gives examples of cdf satisfying this property. Section 3 assumes a linear conditional mean of the form (1.3) and Section 4 considers the more general specification (1.4). Section 5 proposes a test of independence between z_t and λ_t in the duration model (1.2). Monte Carlo experiments and illustrations on series of trading volume and greenhouse gas concentrations are presented. Concluding remarks are given in Section 6. It is shown that a positive-valued time series whose conditional cdf satisfies (1.5) and the mean verifies mild regularity conditions is stationary and ergodic. When Y_t is valued in $\mathbb{N}$, we show that the β-mixing coefficients have exponential decay rate. For some particular POLI models, necessary and sufficient conditions for the existence of moments are also provided.

2 Examples of distributions with stochastic-equal-mean order

We first recall that the exponential family is included in the class of the distributions for which the conditional stochastic order is equal to the conditional mean order, and we notice that the conditional distribution of any ACD-MEM model also satisfies the stochastic-equal-mean order property. We then give other examples of such conditional distributions which, to our knowledge, are not yet fully considered in existing count or duration time series models.

2.1 One-parameter exponential family

Using Yu (2009), Davis and Liu (2012) demonstrated (see Proposition 6 and the discussion after (2.1) in their paper) that (1.5) holds true when F_λ is the cdf of a one-parameter exponential family on $[0, \infty)$. A distribution F_λ is said to belong to such an exponential family if, with respect to a σ-finite measure, it admits a density of the form

$$g_\lambda(y) = h(y) \exp\{\eta y - A(\eta)\} 1_{\{y \geq 0\}}, \tag{2.1}$$

for some scalar natural parameter $\eta = \eta(\lambda)$ and some twice differentiable cumulant generating function $A(\eta)$. It is known that $\lambda = A'(\eta)$. For example F_λ can be the cdf of the Poisson distribution with intensity parameter $\lambda = e^\eta$. Recall that a random variable Y follows a negative binomial, $Y \sim NB(r, p)$, of parameters $r > 0$ and $p \in (0, 1)$ if

$$P(Y = k) = \frac{\Gamma(k + r)}{k! \Gamma(r)} p^r (1 - p)^k, \quad k \in \mathbb{N}.$$

We have $\lambda = r(1-p)/p$. This distribution also belongs to the exponential family when $p = r/(\lambda + r))$ and r is fixed (with $\eta = \log(1 - p)$).

2.2 Standard multiplicative ACD-type models

Let F_λ^- be the quantile function associated to the cdf F_λ. Note that (1.5) is equivalent to

$$\lambda \leq \lambda^* \quad \Rightarrow \quad F_\lambda^-(u) \leq F_{\lambda^*}^-(u), \ \forall u \in (0, 1). \tag{2.2}$$

By positive homogeneity of the quantile function, conditional on $\mathcal{F}_{t-1}$, the quantile function of Y_t satisfying (1.2) is

$$F_{\lambda_t}^-(\alpha) = \lambda_t F^-(\alpha),$$

where F^- is the quantile function of z_t. Therefore the conditional distribution of any standard ACD model satisfies the stochastic-equal-mean order property (2.2).

2.3 Negative binomial $NB(r, p)$ with fixed p

For any fixed p, the negative binomial distribution F_λ with parameter $r = p\lambda/(1-p)$ apparently does not belong to the one-parameter exponential family. The next Lemma shows that this family of distribution however satisfies (1.5). Write $X \leq_{st} Y$ when the random variable Y stochastically dominates the random variable X, *i.e.* if $P(X \leq y) \geq P(Y \leq y)$ for all y.

Lemma 2.1 *If $X \sim NB(r_1, p)$ and $Y \sim NB(r_2, p)$ with $r_1 \leq r_2$, then $X \leq_{st} Y$.*

The previous lemma is quite obvious and can probably be found somewhere in the literature, but we did not find a precise reference of such a result. For completeness, we thus give a proof in Appendix.

2.4 Gamma distributions

A random variable Y is said to be Gamma distributed $\Gamma(a, b)$ with shape parameters $a > 0$ and rate parameter $b > 0$ if it admits the density $g(y) = \Gamma^{-1}(a)b^a y^{a-1}e^{-by}1_{\{y>0\}}$. We have $\lambda := EY = a/b$. For fixed a, the distribution $\Gamma(a, a/\lambda)$ readily belongs to the exponential family (2.1). For fixed b, the distribution $\Gamma(\lambda b, b)$ is not of the form (2.1). However, denoting by $g_\lambda(y)$ the density of that $\Gamma(\lambda b, b)$ distribution, it can be seen that when $\lambda < \lambda^*$ the likelihood ratio $g_\lambda(y)/g_{\lambda^*}(y)$ is a decreasing function, which entails (1.5). Note that if $Y_t \mid \mathcal{F}_{t-1} \sim \Gamma(\lambda_t b, b)$, then $\text{Var}(Y_t \mid \mathcal{F}_{t-1}) = \lambda_t/b$. This entails that (Y_t) does not follow an ACD model of the form (1.2), for which the variance is proportional to λ_t^2.

2.5 Zero-inflated distributions

There exists numerous instances of count data sets with excess zeros with respect to a baseline model, for example the Poisson distribution (see *e.g.* Ridout, Demétrio and Hinde, 1998). One solution consists in assuming that a random element Y of the data set has a

zero-inflated Poisson (ZIP) distribution, given by

$$P(Y = k) = \begin{cases} \tau + (1 - \tau)e^{-\lambda} & \text{if} \quad k = 0 \\ (1 - \tau)e^{-\lambda}\frac{\lambda^k}{k!} & \text{if} \quad k > 0. \end{cases} \tag{2.3}$$

If $\tau \in [0, 1]$ the ZIP distribution is that of a mixture of a proportion τ of variables that structurally always take the zero value and a proportion $1 - \tau$ of variables that follow the Poisson distribution with intensity λ. When $\tau \in [-e^{-\lambda}/(1 - e^{-\lambda}), 0)$ and $\lambda > 0$, the ZIP distribution is actually zero-deflated. The same law can be obtained with the hurdle model which assumes that a proportion τ of variables always take the zero value and a proportion $1 - \tau$ of variables follow the zero-truncated Poisson distribution

$$P(Y = k) = \begin{cases} \tau & \text{if} \quad k = 0 \\ \frac{(1-\tau)e^{-\lambda}\lambda^k}{(1-e^{-\lambda})k!} & \text{if} \quad k > 0. \end{cases}$$

More generally, assume that the baseline cdf is not necessarily Poisson but the cdf F_λ, and define two zero-inflated distributions by

$$P(Y \le y) = \tau + (1 - \tau)F_\lambda(y), \qquad P(Y^* \le y) = \tau + (1 - \tau)F_{\lambda^*}(y), \tag{2.4}$$

for all $y \ge 0$ and $P(Y \le y) = P(Y^* \le y) = 0$ for all $y < 0$, where $\tau \in [0, 1]$ is some extra zero probability. The following lemma shows that if the family of distributions F_λ satisfies (1.5) then this is also the case for the zero-inflated distributions.

Lemma 2.2 *If* (1.5) *and* (2.4) *hold true, then* $EY \le EY^*$ *entails* $Y \le_{st} Y^*$.

3 Stationarity in the linear conditional mean case

First consider the strict stationarity and ergodicity of the general POLI-X model (1.3). Ergodicity entails the strong law of large numbers, and is thus a fundamental tool for studying the asymptotic properties of estimators and test statistics.

Proposition 3.1 *Let $\{F_\lambda, \ \lambda \in (0, \infty)\}$ be a family of cdf on $[0, \infty)$ (i.e. $F_\lambda(y) = 0$ for all $y < 0$) satisfying* (1.5). *There exists a stationary (and ergodic) sequence* (Y_t) *such that*

$$P\left(Y_t \le y \mid \mathcal{F}_{t-1}\right) = F_{\lambda_t}(y), \tag{3.1}$$

where λ_t satisfies either (1.1) *or* (1.3) *with* $(\boldsymbol{X}_t)$ *stationary and ergodic, if*

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1. \tag{3.2}$$

Conversely, if there exists a solution of (3.1) *such that $EY_t = m < \infty$, then $E\boldsymbol{\pi}^\top \boldsymbol{X}_t < \infty$ and* (3.2) *holds.*

The strict stationarity condition (3.2) does not depend on the exogenous variables. This is not surprising since adding covariates remains to substitute a stationary intercept $\omega_t = \omega + \sum_{i=1}^{r} \pi_i x_{i,t-1}$ for the constant ω in λ_t, and it is known (at least for conditional cdf belonging to the exponential family) that the stationarity condition does not depend on the intercept. Francq and Thieu (2018) made a similar comment on GARCH models with exogenous variables. Note also that the stationary solution defined in the proof has a causal Bernoulli shift representation of the form

$$Y_t = \varphi(U_t, U_{t-1}, \ldots; \boldsymbol{X}_{t-1}, \boldsymbol{X}_{t-2}, \ldots),$$

where the sequences (U_t) and $(\boldsymbol{X}_t)$ are independent and (U_t) is iid. It follows that, under the conditions of Proposition 3.1, the condition (3.2) also entails that the multivariate process $(Y_t \boldsymbol{X}_t^\top)^\top$ is stationary and ergodic.

We now give conditions for the existence of moments for the POLI(1,1) model.

Proposition 3.2 *Let $\{F_\lambda, \ \lambda \in (0, \infty)\}$ be a family of cdf on $[0, \infty)$ satisfying* (1.5). *Assume that, for $Y \sim F_\lambda(y)$ and some integer $\ell \ge 2$, there exist nonnegative coefficients $a_j(0), a_j(1), \ldots, a_j(j)$ for all $j \le \ell$ such that*

$$EY^j = \sum_{i=0}^{j} a_j(i) \lambda^i \ for \ j = 1, \ldots, \ell. \tag{3.3}$$

9

Under (3.2), let (Y_t) be a stationary sequence such that $P\left(Y_t \leq y \mid \mathcal{F}_{t-1}\right) = F_{\lambda_t}(y)$, where λ_t satisfies (1.1) with $p = q = 1$. We have $EY_t^\ell < \infty$ if and only if

$$\sum_{j=0}^{\ell} a(j)\binom{\ell}{j}\alpha^j\beta^{\ell-j} < 1, \tag{3.4}$$

where $a(0) = a(1) = 1$ and $a(j) = a_j(j)$ for $j \geq 2$.

Example 3.1 (NB(r, p_t)) *The first moments $m_i = EY^i$ of Y following the $NB\left(r, r/(\lambda + r)\right)$ distribution are*

$$
\begin{aligned}
m_1 &= \lambda, \quad m_2 = \lambda + \frac{1+r}{r}\lambda^2, \quad m_3 = \lambda + 3\frac{1+r}{r}\lambda^2 + \frac{2+3r+r^2}{r^2}\lambda^3, \\
m_4 &= \lambda + 7\frac{1+r}{r}\lambda^2 + 6\frac{2+3r+r^2}{r^2}\lambda^3 + \frac{6+11r+6r^2+r^3}{r^3}\lambda^4.
\end{aligned}
$$

It follows that (3.3) holds with

$$a(2) = \frac{1+r}{r}, \quad a(3) = \frac{2+3r+r^2}{r^2}, \quad a(4) = \frac{6+11r+6r^2+r^3}{r^3}.$$

Proposition 3.2 shows that the POLI(1,1) model with $NB(r, r/(\lambda_t + r))$ conditional distribution admits a moment of

$$
\begin{aligned}
&order\ 2 \quad iff \quad (\alpha+\beta)^2 + \frac{\alpha^2}{r} < 1, \tag{3.5} \\
&order\ 3 \quad iff \quad (\alpha+\beta)^3 + \frac{3\alpha^2(\alpha+\beta)}{r} + \frac{2\alpha^3}{r^2} < 1, \tag{3.6} \\
&order\ 4 \quad iff \quad (\alpha+\beta)^4 + \frac{6\alpha^2(\alpha+\beta)^2}{r} + \frac{\alpha^3(11\alpha+8\beta)}{r^2} + \frac{6\alpha^4}{r^3} < 1. \tag{3.7}
\end{aligned}
$$

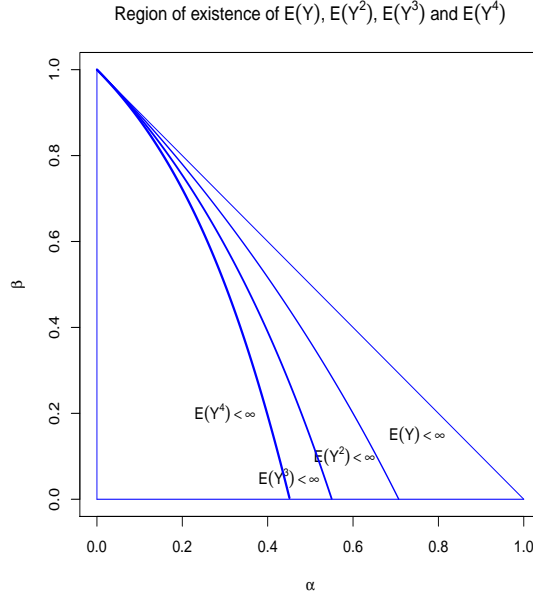*Figure 1 displays these moment conditions when $r = 1$.*

Figure 1: Moment conditions for the INGARCH(1,1) process with $NB(r, p_t)$ conditional distribution.

The condition (3.5) has been given by Christou and Fokianos (2014) and (3.7) by Ahmad and Francq (2016), but without formal proof.

Example 3.2 ($NB(r_t, p)$) *Now consider the INGARCH(1,1) model with $NB(p\lambda_t/(1-p), p)$ conditional distribution. By Jain and Consul (1971), the moments $m_\ell = EY^\ell$ of $Y \sim NB(r, p)$ satisfy*

$$m_\ell = p\lambda \sum_{j=0}^{\ell-1} \binom{\ell-1}{j} \left(m_j + \tfrac{1-p}{\lambda p} m_{j+1}\right), \qquad \ell \geq 1.$$

It follows that

$$m_1 = \lambda, \qquad m_2 = \lambda^2 + \frac{1}{p}\lambda, \qquad m_3 = \lambda^3 + \frac{3}{p}\lambda^2 + \frac{2-p}{p^2}\lambda,$$

and, more generally, (3.3) holds with $a(j) = a_j(j) = 1$ for all j. We then have

$$\sum_{j=0}^{\ell} a(j) \binom{\ell}{j} \alpha^j \beta^{\ell-j} = (\alpha + \beta)^j,$$

11

and Proposition 3.2 shows that this INGARCH(1,1) model admits moments of any orders if and only if $\alpha + \beta < 1$.

4 Extension to nonlinear conditional means

Let $\mathcal{B}$ be the Borel sigma-algebra of $\mathbb{R}^\infty$. For $h \geq 0$, let the β-mixing coefficient (also called absolute regularity coefficient)

$$\beta(h) = E \sup_{A \in \mathcal{B}} |P\{(Y_h, Y_{h+1}, \dots) \in A \mid Y_0, Y_{-1}, \dots\} - P\{(Y_h, Y_{h+1}, \dots) \in A\}|.$$

We now give conditions for stationarity and ergodicity when the conditional mean has the general form (1.4). For integer-valued observations, we also show the geometric decrease of the β-mixing coefficients. The geometric decrease of the β-mixing coefficients is a stronger property than the ergodicity, which entails the central limit theorem under some moment conditions.

Proposition 4.1 *Let $\{F_\lambda,\ \lambda \in (0, \infty)\}$ be a family of cdf on $[0, \infty)$ satisfying (1.5), and let $(\boldsymbol{X}_t)$ be a stationary and ergodic process. Assume that the function $g(y_1, \dots, y_q, \lambda_1, \dots, \lambda_p)$ is such that, for all $(y_i, y_i') \in [0, +\infty)^2$, $i = 1, \dots, q$ for all $(\lambda_j, \lambda_j') \in (0, \infty)^2$, $j = 1, \dots, p$,*

$$\left| g(y_1, \dots, y_q, \lambda_1, \dots, \lambda_p) - g(y_1', \dots, y_q', \lambda_1', \dots, \lambda_p') \right|$$
$$\leq \sum_{i=1}^{q} \alpha_i |y_i - y_i'| + \sum_{j=1}^{p} \beta_j |\lambda_j - \lambda_j'|.$$

If

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1, \tag{4.1}$$

then there exists a stationary and ergodic sequence (Y_t) such that the distribution of Y_t conditional on $\mathcal{F}_{t-1}$ is F_{λ_t}, where λ_t satisfies (1.4). Moreover, if Y_t is valued in $\mathbb{N}$, there exist constants $K > 0$ and $\rho \in (0, 1)$ such that

$$\beta(h) \leq K\rho^h, \qquad h \geq 0.$$

5   Testing the multiplicative form of duration models

Instead of a standard ACD duration model with the multiplicative (MEM) form (1.2), the present paper suggests a more general POLI model with a conditional distribution that is not constrained by the MEM structure. The variable $z_t = Y_t/\lambda_t$ is independent of $\lambda_t :=$ $E(Y_t \mid \mathcal{F}_{t-1})$ in model (1.2), whereas the two variables are uncorrelated but not necessarily independent in the POLI model. In particular the conditional variance of a POLI model is not constrained to be proportional to λ_t^2. It is thus of interest to test

$$H_0: \quad z_t \text{ and } \lambda_t \text{ are independent,} \tag{5.1}$$

without specifying a particular alternative model. Based on observations $Y_1, \ldots, Y_n$, the hypothesis H_0 can be tested by using the empirical distance covariance (see Székely et al. (2007), Rizzo and Szḱely (2016), and the references therein)

$$\mathcal{V}_n^2 = \int \left| \hat{\varphi}_{z,\lambda}(t,s) - \hat{\varphi}_z(t)\hat{\varphi}_\lambda(s) \right|^2 w(t,s)\,dt\,ds,$$

where $\hat{\varphi}_{z,\lambda}$, $\hat{\varphi}_z$ and $\hat{\varphi}_\lambda$ are respectively empirical estimators of the characteristic functions of (z_t, λ_t), z_t and λ_t. As shown in Székely, Rizzo and Bakirov (2007), a relevant choice of weighting function is $w(t,s)$ proportional to $t^{-2}s^{-2}$. Under the null and the existence of marginal moments, $n\mathcal{V}_n^2$ converges in distribution. The limiting distribution depends on the marginal laws of the two variables z_t and λ_t in the iid case. Davis, Matsui, Mikosch and Wan (2018) recently showed that the nice properties of the distance covariance and correlation can also be extended to time series. In our framework, the sequence $(z_t, \lambda_t)_{t \geq 1}$ is not iid under the null, and λ_t is not directly observable, but can be approximated by $\tilde{\lambda}_t(\hat{\boldsymbol{\theta}})$ where $\hat{\boldsymbol{\theta}}$ is a consistent estimator of the parameter $\boldsymbol{\theta} = (\omega, \alpha_1, \ldots, \beta_p, \boldsymbol{\pi}^\top)^\top$ and

$$\tilde{\lambda}_t(\boldsymbol{\theta}) = \omega + \sum_{i=1}^q \alpha_i Y_{t-i} + \sum_{j=1}^p \beta_j \tilde{\lambda}_{t-j}(\boldsymbol{\theta}) + \boldsymbol{\pi}^\top \boldsymbol{X}_{t-1}, \quad t \geq q+1,$$

where $\tilde{\lambda}_q(\boldsymbol{\theta}), \ldots, \tilde{\lambda}_{q+1-p}(\boldsymbol{\theta})$ are fixed initial values. A natural choice for the estimator of the unknown true parameter value $\boldsymbol{\theta}_0 \in \Theta$ is the exponential quasi-maximum likelihood

13

estimator (QMLE) defined by

$$\hat{\boldsymbol{\theta}} = \arg\min_{\theta \in \Theta} \sum_{t=q+1}^{n} Y_t / \tilde{\lambda}_t(\boldsymbol{\theta}) + \log(\tilde{\lambda}_t(\boldsymbol{\theta})). \qquad (5.2)$$

We propose to approximate the distribution of $\mathcal{V}_n^2$ by the bootstrap distribution of the variable $\mathcal{V}_n^{*2}$ defined in the following resampling scheme:

(i) Calculate $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_n(Y_1, \ldots, Y_n)$, the test statistics $\mathcal{V}_n^2 = \mathcal{V}_n^2(Y_1, \ldots, Y_n)$, and the residuals $\hat{z}_t = Y_t / \tilde{\lambda}_t(\hat{\boldsymbol{\theta}})$ for $t = q + 1, \ldots, n$. Denote by F_n the empirical distribution of $\{\hat{z}_t / s_n, t = 1 + q, \ldots, n\}$ where $s_n = \sum_{t=q+1}^{n} \hat{z}_t / (n - q)$ (with this scaling factor, the expectation of the distribution F_n is equal to 1).

(ii) Generate $Y_1^*, \ldots, Y_n^*$ where $Y_t^* = z_t^* \tilde{\lambda}_t^*(\hat{\boldsymbol{\theta}})$, the z_t^*'s are independent and F_n-distributed, and $\tilde{\lambda}_t^*(\boldsymbol{\theta})$ is defined as $\tilde{\lambda}_t(\boldsymbol{\theta})$ with Y_{t-i} replaced by Y_{t-i}^*. Calculate $\hat{\boldsymbol{\theta}}^* = \boldsymbol{\theta}_n(Y_1^*, \ldots, Y_n^*)$ and the test statistics $\mathcal{V}_n^{*2} = \mathcal{V}_n^2(Y_1^*, \ldots, Y_n^*)$

(iv) Repeat step (ii) B times and calculate the corresponding test statistics $\mathcal{V}_{n,1}^{*2}, \ldots, \mathcal{V}_{n,B}^{*2}$.

(v) At the nominal significance level $\alpha \in (0, 1)$, reject H_0 if $\mathcal{V}_n^2 > \mathcal{V}_{n,(B-[\alpha B])}^{*2}$, where $\mathcal{V}_{n,(1)}^{*2} \le \ldots \le \mathcal{V}_{n,(B)}^{*2}$ denote the corresponding order statistics.

The validity, *i.e.* the consistency under the null and the alternative, of an apparently similar resampling scheme has been proven in Francq, Jiménez-Gamero and Meintanis (2017). However, our framework is not the same, since the above-mentioned paper concerns sphericity tests based on the empirical characteristic function. Proving the validity of the present algorithm does not seem trivial and will be the topic of future research.

Of course, when one wants to test a given ACD model against a particular POLI model, a standard–and often more efficient–alternative to the previous omnibus test consists in comparing the likelihood of the two models. This will be illustrated in an empirical application below.

5.1 Monte Carlo experiments

We simulated two data generating processes (DGP), one which satisfies H_0 and the other which does not. The first DGP is an ACD(1,1) model $Y_t = \lambda_t z_t$ where $\lambda_t = \omega + \alpha Y_{t-1} + \beta \lambda_{t-1}$ with $(\omega, \alpha, \beta) = (0.5, 0.1, 0.89)$, and the z_t's are independent with exponential distribution of mean 1. The other DGP (denoted H_1 in Table 1) is a POLI model of conditional distribution $\Gamma(b\lambda_t, b)$ with $b = 0.01$ and λ_t which follows the same equation as in the first DGP. We used the resampling algorithm with $B = 99$ replications. Table 1 displays the empirical relative frequency of rejection over $N = 1000$ independent replications of the two DGP's, for the sample sizes $n = 500$ and $n = 1000$. The exercise is computationally demanding since $N \times (B+1) \times 2 \times 2 = 400000$ models have to be estimated and as many distance covariances have to be computed (leading to around 3 days of computations on a personal laptop). Table 1 shows that the error of first kind is well controlled when $\alpha = 1\%$, but the test is

	$n = 500$			$n = 1000$		
DGP	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$
H_0	1.2	3.0	5.8	0.7	3.8	6.7
H_1	54.0	86.0	95.2	73.8	96.5	99.2

Table 1: Percentages of rejections of the bootstrapped distance covariance test.

slightly conservative at levels $\alpha = 5\%$ and $\alpha = 10\%$. Indeed, over $N = 1000$ replications of a test with nominal level $\alpha = 1\%$ (respectively 5% and 10%), the empirical relative frequency of rejection should vary between 0.2% and 1.9% (respectively 3.2% and 6.9%, and 7.5% and 12.5%) with probability 0.99. Despite the fact it is conservative, the distance covariance test is surprisingly powerful in our Monte Carlo setting. Of course, for other alternative models, that omnibus test of independence may be less powerful. For instance, when the conditional distribution of the DGP is $\Gamma(b\lambda_t, b)$ with larger b, the power is smaller. This is not surprising because the variance λ_t/b of $z_t \sim \Gamma(b\lambda_t, b)$ is a decreasing function of b and, since the variable z_t tends to become constant when b increases, it is harder and harder to detect a relationship

between z_t and any other variable.

5.2   S&P 500 transaction volume

Consider the series (Y_t) of the S&P 500 transaction volume from 3/10/2013 to 3/10/2018, which corresponds to 1260 values (downloaded on Yahoo! Finance). Fitting a POLI(2,2) model (1.1), the parameter estimates of the QMLE (5.2) are $\hat{\omega} = 0.870$, $\hat{\alpha}_1 = 0.499$, $\hat{\alpha}_2 = 0.130$, $\hat{\beta}_1 = 0.004$, $\hat{\beta}_2 = 0.122$. As shown in the bottom-left panel of Figure 2, the autocorrelation function (ACF) of the residuals $\hat{z}_t = Y_t / \tilde{\lambda}_t(\hat{\boldsymbol{\theta}})$ no longer shows any sign of dynamics. The distance covariance test however rejects the standard MEM-ACD model in which z_t and λ_t are independendent. Indeed, a kernel density estimator of the boostrapped distribution of $\mathcal{V}_n^2$ under the null is displayed at the bottom-right panel of Figure 2. The value of $\mathcal{V}_n^2$ computed on the observations, indicated by a cross on the figure, is located at the extreme right of the distribution, which gives strong evidence for rejecting the null. Actually, the observed value of the distance covariance is larger than all the $B = 199$ bootstrapp replications used to approximate the distribution of $\mathcal{V}_n^2$ under the null. The estimated p-value is thus $1/200 = 0.005$.

5.3   Greenhouse gas concentrations

Lucas *et al.* (2015) studied a large network data set of greenhouse gas (GHG) concentrations collected by tracers located at different areas in California. The left panel of Figure 3 displays the time series obtained by one of these tracers. The partial autocorrelogram suggests that a simple POLI$(1,0)$ model could summarize the dynamics of the conditional mean. The distance covariance test is not conclusive, since the p-values of the test generally vary between 2% and 14% among the different series of GHG concentrations. On the time series plot, one can see a concentration of observations around zero, which precludes a continuous conditional distribution such as the Gamma law. We thus investigated the use of zero-inflated conditional distributions. In particular, we found that for a large majority of series, a zero-inflated $\Gamma(\lambda b, b)$ distribution has a higher likelihood than a zero-inflated $\Gamma(a, a/\lambda)$
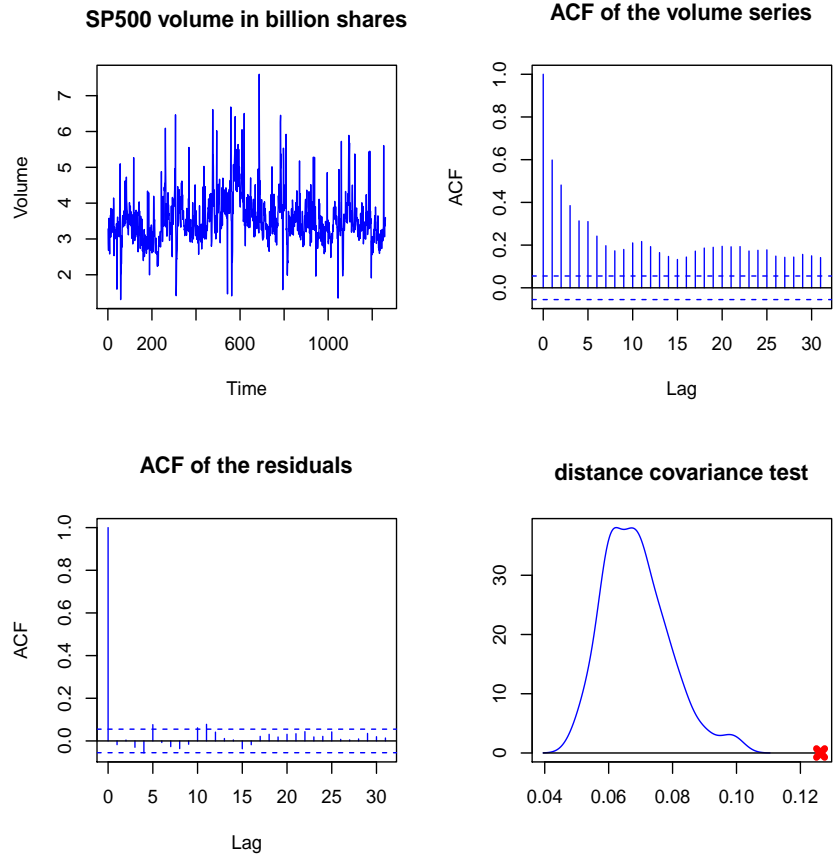
16

Figure 2: S&P 500 transaction volume from 3/10/2013 to 3/10/2018, ACF on the observed series, ACF on the residuals of the POLI(2,2) model, distribution of the distance covariance under the null hypothesis of multiplicative form, and observed distance covariance (cross symbol).

distribution. Since the variance of the former law is not proportional to the square of its mean, this suggests that a multiplicative model of the form (1.2) is inadequate. For the greenhouse gas time series, we thus propose the model

$$\lambda_t = \omega + \alpha Y_{t-1}, \quad Y_t \mid \mathcal{F}_{t-1} \sim \tau \delta_0 + (1-\tau)\Gamma(\lambda_t b, b),$$

with obvious notation for the mixing distribution. On the series displayed in Figure 3, the maximum-likelihood estimates of the parameters are $\hat{\omega} = 0.0024$, $\hat{\alpha} = 0.834$, $\hat{\tau} = 0.186$ and $\hat{b} = 245.2$.
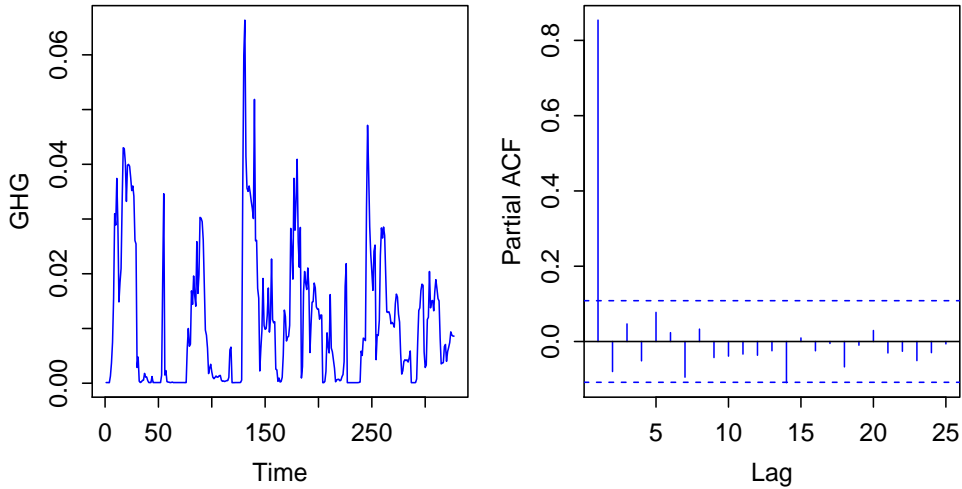

Figure 3: Greenhouse gas time series concentration every 6 hours from May 10 to July 31, 2010, and empirical partial autocorrelations of the time series.

6   Conclusion

Proving the ergodicity of count time series models is a notorious tricky problem, for which the present paper gives a simple solution. This also applies to more general positive-valued series. The illustrations presented in Section 5 suggest that some real series are better represented by a POLI model than by a model of the form (1.2). This gives a motivation for relaxing the

usual multiplicative form of the ACD-like models, even if the probabilistic structure of the model is then complicated by the absence of an explicit iid innovation sequence. Note that the positivity of the observations is not fundamental for some of the results. In particular, one could easily obtain sufficient stationarity conditions without this assumption. Moreover, our results can be applied to positive-valued transformations of a non-positive series ϵ_t. For example, the square of a GARCH has the ACD form $\epsilon_t^2 = \sigma_t^2 \eta_t^2$ where the volatility σ_t is independent of the iid sequence η_t. Since the multiplicative form of the GARCH model entails strong restrictions, such as a constant conditional kurtosis, it could be of interest to consider a POLI model on ϵ_t^2. This is a topic that we leave for future researches.

A   Proofs

Proof of Lemma 2.1

Note that the result is trivial when the number of failures r_1 and r_2 are integers. More generally, note that the likelihood ratio

$$\frac{P\{NB(r_2, p) = k\}}{P\{NB(r_1, p) = k\}} = p^{r_2 - r_1} \prod_{i=1}^{k} \frac{r_2 + k - i}{r_1 + k - i}$$

increases with k, which is known to entail the required stochastic dominance (see *e.g.* Theorem 1 in Lehmann (1955)). ∎

Proof of Lemma 2.2

Assume (1.5), (2.4) and $EY = (1 - \tau)\lambda \leq EY^* = (1 - \tau)\lambda^*$. Then for $y \geq 0$ we have $P(Y \leq y) = \tau + (1 - \tau)F_\lambda(y) \geq \tau + (1 - \tau)F_{\lambda^*}(y) = P(Y^* \leq y)$ and the result follows. ∎

Proof of Proposition 3.1

Assume (1.3) with $(\boldsymbol{X}_t)$ stationary and ergodic, for which (1.1) can be considered as a particular case.

If there exists $m \in (0, \infty)$ such that such that $m = EY_t = E\lambda_t$ for all t, then

$$\left(1 - \sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j\right) m = \omega + E\boldsymbol{\pi}^\top \boldsymbol{X}_t.$$

19

Under the positivity constraints on the parameters and exogenous variables, this equality entails (3.2) and $E\boldsymbol{\pi}^\top \boldsymbol{X}_t < \infty$.

It thus remains to show that (3.2) is sufficient for the existence of a strictly stationary and ergodic solution to (3.1). Let (U_t) be an iid sequence of random variables uniformly distributed in $[0, 1]$, independent of the sequence $(\boldsymbol{X}_t)$. For $t \in \mathbb{Z}$, let $Y_t^{(k)} = \lambda_t^{(k)} = 0$ when $k \leq 0$ and, for $k > 0$, let

$$Y_t^{(k)} = F_{\lambda_t^{(k)}}^-(U_t), \qquad \lambda_t^{(k)} = \omega + \sum_{i=1}^q \alpha_i Y_{t-i}^{(k-i)} + \sum_{j=1}^p \beta_j \lambda_{t-j}^{(k-j)} + \boldsymbol{\pi}^\top \boldsymbol{X}_{t-1}. \qquad (A.1)$$

For $k \geq 2$, we have

$$\lambda_t^{(k)} = \psi_k(U_{t-1}, \ldots, U_{t-k+1}; \boldsymbol{X}_s, \, s < t),$$

where $\psi_k : [0, 1]^k \times [0, \infty)^\infty \to [0, \infty)$ is a measurable function. Therefore, for any k, the sequences $\left(\lambda_t^{(k)}\right)_t$ and $\left(Y_t^{(k)}\right)_t$ are stationary and ergodic. Let $\mathcal{F}_{t-1}^{(k)}$ and $\mathcal{F}_{t-1}^*$ be the sigma-fields generated by $\left\{Y_{t-i}^{(k-i)}, i > 0; \boldsymbol{X}_s, s < t\right\}$ and $\{U_s, \boldsymbol{X}_s, s < t\}$, respectively. We have

$$E\left(Y_t^{(k)} \mid \mathcal{F}_{t-1}^{(k)}\right) = E\left(Y_t^{(k)} \mid \mathcal{F}_{t-1}^*\right) = \lambda_t^{(k)},$$

$$P\left(Y_t^{(k)} \leq y \mid \mathcal{F}_{t-1}^{(k)}\right) = P\left(F_{\lambda_t^{(k)}}^-(U_t) \leq y \mid \mathcal{F}_{t-1}^*\right) = F_{\lambda_t^{(k)}}(y).$$

We have used the well known result that $F_\lambda^-(U)$ has the cdf F_λ when U is uniformly distributed in $[0, 1]$. To show the existence of a solution to (3.1), with $\mathcal{F}_{t-1}$ replaced by $\mathcal{F}_{t-1}^*$, it is now sufficient to show that

$$\lambda_t = \lim_{k \to \infty} \lambda_t^{(k)} \text{ exists almost surely (a.s.) in } [0, +\infty). \qquad (A.2)$$

Taking the limit as $k \to \infty$ in both sides of the equalities in (A.1), the solution will be then given by $Y_t = \lim_{k \to \infty} Y_t^{(k)} = F_{\lambda_t}^-(U_t)$ a.s. We then note that the distribution of Y_t given $\mathcal{F}_{t-1}^*$ is the same as that of Y_t given $\mathcal{F}_{t-1}$ since λ_t is $\mathcal{F}_{t-1}$-measurable.

We now show (A.2) under (3.2). We first prove that, for all k,

$$0 \leq \lambda_t^{(k-1)} \leq \lambda_t^{(k)} \text{ a.s.} \qquad (A.3)$$

and

$$E\left(Y_t^{(k)} - Y_t^{(k-1)}\right) = E\left(\lambda_t^{(k)} - \lambda_t^{(k-1)}\right) \in [0, \infty). \tag{A.4}$$

Clearly, (A.3) and (A.4) hold true for $k \leq 0$. Assume (A.3) is satisfied for $k \leq k_0$, then using (2.2) we have

$$\lambda_t^{(k_0+1)} = \omega + \sum_{i=1}^q \alpha_i F^-_{\lambda_{t-i}^{(k_0+1-i)}}(U_{t-i}) + \sum_{j=1}^p \beta_j \lambda_{t-j}^{(k_0+1-j)} + \sum_{i=1}^r \pi_i x_{i,t-1}$$

$$\geq \omega + \sum_{i=1}^q \alpha_i F^-_{\lambda_{t-i}^{(k_0-i)}}(U_{t-i}) + \sum_{j=1}^p \beta_j \lambda_{t-j}^{(k_0-j)} + \sum_{i=1}^r \pi_i x_{i,t-1} = \lambda_t^{(k_0)}.$$

Therefore the inequalities in (A.3) are shown by induction. Now note that $EX_t^{(k)} = E\lambda_t^{(k)}$ exists for any fixed k, and for all positive parameters. It follows that (A.4) holds true. In the case $p = q = 1$, we then have

$$E\left|\lambda_t^{(k)} - \lambda_t^{(k-1)}\right| = (\alpha + \beta) E\left(\lambda_{t-1}^{(k-1)} - \lambda_{t-1}^{(k-2)}\right) = (\alpha + \beta)^{k-1} \omega.$$

More generally, with obvious convention, under (3.2) we have

$$E\left|\lambda_t^{(k)} - \lambda_t^{(k-1)}\right| = \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) E\left(\lambda_{t-i}^{(k-i)} - \lambda_{t-i}^{(k-i-1)}\right) \leq K\rho^k, \quad \forall k \geq 1,$$

with $K > 0$ and $\rho \in (0, 1)$. This entails that the sequence $\left\{\lambda_t^{(k)}\right\}_k$ converges in L^1 and a.s. under (3.2). Moreover, since

$$\lambda_t = \psi(U_{t-1}, U_{t-2}, \ldots; \boldsymbol{X}_{t-1}, \boldsymbol{X}_{t-2}, \ldots),$$

where $\psi : [0,1]^\infty \times [0,\infty)^\infty \to [0,\infty)$ is a measurable function, the sequence (λ_t) is ergodic.

∎

Proof of Proposition 3.2

Let the notation $m_s = EX_t^s$ when the moment exists, and $b(\ell) = \sum_{i=0}^{\ell-1} a_\ell(i) E\lambda_t^i$. Then (3.3) entails $m_\ell = a(\ell) E\lambda_t^\ell + b(\ell)$.

We first show $EY_t^2 < \infty$ iff (3.4) holds with $\ell = 2$. The latter condition writes

$$\rho := (\alpha + \beta)^2 + \{a(2) - 1\}\alpha^2 < 1. \tag{A.5}$$

Since $m_2 = a(2)E\lambda_t^2 + b(2)$, we have

$$
\begin{aligned}
m_2 &= a(2)\left\{\omega^2 + \alpha^2 m_2 + 2\omega(\alpha+\beta)m_1\right\} + (\beta^2 + 2\alpha\beta)\left\{m_2 - b(2)\right\} + b(2) \\
&= \left\{a(2)\alpha^2 + \beta^2 + 2\alpha\beta\right\} m_2 + K,
\end{aligned}
$$

where

$$
K = a(2)\left\{\omega^2 + 2\omega(\alpha+\beta)m_1\right\} + b(2)\left(1 - \beta^2 - 2\alpha\beta\right) > 0.
$$

Therefore $EX_t^2 < \infty$ entails (A.5). To show that (A.5) is also sufficient, recall that it has been shown in the proof of Proposition 3.1 that

$$
Y_t = \lim_{k\to\infty} \uparrow Y_t^{(k)}.
$$

By the monotone convergence theorem, to prove that m_2 exists it thus suffices to prove that $\lim_{k\to\infty} m_2^{(k)}$ is finite, where $m_s^{(k)}$ denotes $EY_t^{(k)s}$ (which is finite for all $s \geq 0$ and all k). Letting $\mu_s^{(k)} = E\lambda_t^{(k)s}$ and $b^{(k)}(\ell) = \sum_{i=0}^{\ell-1} a_\ell(i)E\lambda_t^{(k)i}$ we have

$$
\begin{aligned}
m_2^{(k)} &= a(2)\mu_2^{(k)} + b^{(k)}(2) \\
&= a(2)\left\{\omega^2 + \alpha^2 m_2^{(k-1)} + 2\omega(\alpha+\beta)m_1^{(k-1)}\right\} \\
&\quad + (\beta^2 + 2\alpha\beta)\left\{m_2^{(k-1)} - b^{(k-1)}(2)\right\} + b^{(k)}(2) \\
&= \left\{a(2)\alpha^2 + \beta^2 + 2\alpha\beta\right\} m_2^{(k-1)} + K^{(k)},
\end{aligned}
$$

where

$$
K^{(k)} = a(2)\left\{\omega^2 + 2\omega(\alpha+\beta)m_1^{(k-1)}\right\} + b^{(k)}(2) - b^{(k-1)}(2)\left(\beta^2 + 2\alpha\beta\right) \to K
$$

a.s. as $k \to \infty$, since we have seen in the proof of Proposition 3.1 that (3.2) entails $\lim_{k\to\infty} m_1^{(k)} = \lim_{k\to\infty} \mu_1^{(k)} = m_1$. We thus have

$$
m_2^{(k)} \leq \rho m_2^{(k-1)} + 2K \leq 2K \sum_{i=0}^{\infty} \rho^i < \infty
$$

under (A.5). It follows that $m_2 = \lim_{k\to\infty} \uparrow m_2^{(k)} < \infty$ under (A.5).

The proof of (3.4) is complete in the case $\ell = 2$. Now consider the general case, arguing by induction on $\ell \geq 3$. We have

$$
\begin{aligned}
m_\ell &= a(\ell) \left\{ \sum_{j=0}^{\ell} \binom{\ell}{j} \alpha^j \beta^{\ell-j} E X_{t-1}^j \lambda_{t-1}^{\ell-j} + R_\ell \right\} + b(\ell) \\
&= a(\ell) \alpha^\ell m_\ell + \sum_{j=0}^{\ell-1} a(j) \binom{\ell}{j} \alpha^j \beta^{\ell-j} \{ m_\ell - b(\ell) \} + a(\ell) R(\ell) + b(\ell),
\end{aligned}
$$

where the term $R(\ell)$ is a linear combination of $1, E\lambda_t, \ldots, E\lambda_t^{\ell-1}$ with positive coefficients. By induction, one can assume that $R(\ell)$ and $b(\ell)$ are finite under (3.4). It follows that (3.4) is necessary to have m_ℓ finite. The converse is shown as in the case $\ell = 2$. ∎

Proof of Proposition 4.1

As in the proof of Proposition 3.1, consider an iid sequence (U_t) of random variables uniformly distributed in $[0, 1]$, independent of the sequence $(\boldsymbol{X}_t)$, and define $Y_t^{(k)} = \lambda_t^{(k)} = 0$ when $k \leq 0$ and, when $k > 0$,

$$
Y_t^{(k)} = F_{\lambda_t^{(k)}}^{-}(U_t), \tag{A.6}
$$

$$
\lambda_t^{(k)} = g(Y_{t-1}^{(k-1)}, \ldots, Y_{t-q}^{(k-q)}, \lambda_{t-1}^{(k-1)}, \ldots, \lambda_{t-p}^{(k-p)}) + \pi(\boldsymbol{X}_{t-1}).
$$

By the argument of the proof of Proposition 3.1, to show the existence of a stationary solution it suffices to show the almost sure convergence (A.2) of $\lambda_t^{(k)}$ as $k \to \infty$. In view of (2.2), we have

$$
E \left\{ |Y_t^{(k)} - Y_t^{(k-1)} \mid \lambda_t^{(k)}, \lambda_t^{(k-1)} \right\} = E \left| \lambda_t^{(k)} - \lambda_t^{(k-1)} \right|.
$$

Therefore

$$
E \left| Y_t^{(k)} - Y_t^{(k-1)} \right| = E \left| \lambda_t^{(k)} - \lambda_t^{(k-1)} \right|.
$$

It follows that, under (4.1),

$$
E \left| \lambda_t^{(k)} - \lambda_t^{(k-1)} \right| \leq \sum_{i=1}^{p \vee q} (\alpha_i + \beta_i) E \left| \lambda_{t-i}^{(k-i)} - \lambda_{t-i}^{(k-i-1)} \right| \leq K\rho^k, \quad \forall k \geq 1,
$$

for some constans $K > 0$ and $\rho \in (0, 1)$. The proof of the existence of a stationary solution follows.

23

Now assume (4.1) and Y_t is valued in $\mathbb{N}$. For $i = 1, 2$, define stationary processes by

$$Y_t^{[i]} = F_{\lambda_t^{[i]}}^{-}(U_t), \qquad \lambda_t^{[i]} = g(Y_{t-1}^{[i]}, \ldots, Y_{t-q}^{[i]}, \lambda_{t-1}^{[i]}, \ldots, \lambda_{t-p}^{[i]}) + \pi(\boldsymbol{X}_{t-1}),$$

for $t \geq 1$, where

$$\boldsymbol{Z}_0^{[1]} = (Y_0^{[1]}, \ldots, Y_{1-q}^{[1]}, \lambda_0^{[1]}, \ldots, \lambda_{1-p}^{[1]})$$

and

$$\boldsymbol{Z}_0^{[2]} = (Y_0^{[2]}, \ldots, Y_{1-q}^{[2]}, \lambda_0^{[2]}, \ldots, \lambda_{1-p}^{[2]})$$

are independent and follow the stationary law of

$$\boldsymbol{Z}_t := (Y_{t-1}, \ldots, Y_{t-q}, \lambda_{t-1}, \ldots, \lambda_{t-p}).$$

By the coupling arguments used to show (5.6) in Davis and Liu (2016) or (5.9) in Neumann (2011), we have

$$\begin{aligned}
\beta(h) &= E \sup_{A \in \mathcal{B}} |P\{(Y_h, Y_{h+1}, \ldots) \in A \mid \boldsymbol{Z}_0\} - P\{(Y_h, Y_{h+1}, \ldots) \in A\}| \\
&= E \sup_{A \in \mathcal{B}} \left| P\left\{(Y_h^{[1]}, Y_{h+1}^{[1]}, \ldots) \in A \mid \boldsymbol{Z}_0^{[1]}\right\} - P\left\{(Y_h^{[2]}, Y_{h+1}^{[2]}, \ldots) \in A \mid \boldsymbol{Z}_0^{[1]}\right\} \right| \\
&\leq \sum_{k=0}^{\infty} P\left(Y_{h+k}^{[1]} \neq Y_{h+k}^{[2]}\right) \leq \sum_{k=0}^{\infty} E\left|Y_{h+k}^{[1]} - Y_{h+k}^{[2]}\right|,
\end{aligned}$$

with obvious notation. The last inequality holds because $\left|Y_{h+k}^{[1]} - Y_{h+k}^{[2]}\right|$ is valued in $\mathbb{N}$. Now, note that (2.2) implies that

$$E\left(|Y_t^{[1]} - Y_t^{[2]}| \mid \lambda_t^{[1]}, \lambda_t^{[2]}\right) = |\lambda_t^{[1]} - \lambda_t^{[2]}|.$$

Therefore

$$E|Y_t^{[1]} - Y_t^{[2]}| = E|\lambda_t^{[1]} - \lambda_t^{[2]}| \leq \sum_{i=1}^{q} \alpha_i E|Y_{t-i}^{[1]} - Y_{t-i}^{[2]}| + \sum_{j=1}^{p} \beta_j E|\lambda_{t-j}^{[1]} - \lambda_{t-j}^{[2]}| \leq K\rho^t,$$

and the conclusion follows. ∎

References

[1] Agosto, A., Cavaliere, G., Kristensen, D. and Rahbek, A. (2016) Modeling corporate defaults: Poisson autoregressions with exogenous covariates (PARX). *Journal of Empirical Finance* 38, 640-663.

[2] Ahmad, A. and Francq, C. (2016) Poisson qmle of count time series models. *Journal of Time Series analysis* 37, 291-314.

[3] Al-Osh, M.A. and Alzaid, A.A. (1987) First-order integer-valued autoregressive ($INAR$(1)) process. *Journal of Time Series Analysis* 8, 261-275.

[4] Cameron, A.C. and Trivedi, P.K. (2001) Essentials of count data regression, in Baltagi, B. Hani, *A companion to theoretical econometrics*, 331-348, Blackwell.

[5] Chou, R.Y. (2005) Forecasting financial volatilities with extreme values: The conditional autoregressive range (CARR) Model. *Journal of Money, Credit, and Banking* 37, 561–582.

[6] Christou, V. and Fokianos, K. (2014) Quasi-likelihood inference for negative binomial time series models. *Journal of Time Series Analysis* 35, 55-78.

[7] Davis, R.A., Holan, S.H., Lund, R. and Ravishanker, N. (2016) *Handbook of discrete-valued time series*. Chapman and Hall.

[8] Davis, R.A. and Liu, H. (2016) Theory and inference for a class of nonlinear models with application to time series of counts. *Statistica Sinica* 26, 1673-1707.

[9] Davis, R.A., Matsui, M., Mikosch, T. and Wan, P. (2018) Applications of distance correlation to time series. *Bernoulli* 24, 3087–3116.

[10] Douc, R., Doukhan, P. and Moulines, E. (2013) Ergodicity of observation-driven time series models and consistency of the maximum likelihood estimator. *Stochastic Processes and their Applications* 123, 2620–2647.

[11] Douc, R., Roueff, F. and Sim, T. (2015) Handy sufficient conditions for the convergence of the maximum likelihood estimator in observation-driven models. *Lithuanian Mathematical Journal* 55, 367–392.

[12] Douc, R., Roueff, F. and Sim, T. (2016) The maximizing set of the asymptotic normalized log-likelihood for partially observed Markov chains. *The Annals of Applied Probability* 26, 2357–2383.

[13] Doukhan, P. and Neumann, M.H. (2017) Absolute regularity of semi-contractive GARCH-type processes. *arXiv preprint arXiv:1711.04282*.

[14] Doukhan, P., Fokianos, K. and Tjøstheim, D. (2012) On weak dependence conditions for Poisson autoregressions. *Statistics and Probability Letters* 82, 942–948.

[15] Doukhan, P., Fokianos, K. and Tjøstheim, D. (2013) Correction to "On weak dependence conditions for Poisson autoregressions" [Statist. Probab. Lett. 82 (2012) 942–948]. *Statistics and Probability Letters* 83, 1926–1927.

[16] Engle, R. (2002) New frontiers for Arch models. *Journal of Applied Econometrics* 17, 425–446.

[17] Engle, R. and Russell, J. (1998) Autoregressive conditional duration: A new model for irregular spaced transaction data. *Econometrica* 66, 1127–1162.

[18] Ferland, R., Latour, A., and Oraichi, D. (2006) Integer-valued *GARCH* process. *Journal of Time Series Analysis* 27, 923-942.

[19] Fokianos K, Rahbek A, Tjøstheim D. (2009) Poisson autoregression. *Journal of the American Statistical Association* 140, 1430–1439.

[20] Francq, C., and Thieu, L. (2018) Qml inference for volatility models with covariates. *Econometric Theory* 1-36.

[21] Francq, C., Jiménez-Gamero, M.D., and Meintanis, S.G. (2017) Tests for conditional ellipticity in multivariate GARCH models. *Journal of Econometrics* 196, 305–319.

[22] Francq, C. and Zakoian, J.-M. (2010) *GARCH models: structure, statistical inference and financial applications.* John Wiley & Sons.

[23] Franke J. (2010) Weak dependence of functional INGARCH processes. *Technical report, University of Kaiserslautern.*

[24] Gonçalves, E., Mendes-Lopes N. and Silva F. (2015) Infinitely divisible distributions in integer-valued GARCH models. *Journal of Time Series Analysis* 36, 503-527.

[25] Gurmu, S. and Trivedi, P.K. (1996) Excess Zeros in Count Models for Recreational Trips. *Journal of Business & Economic Statistics* 14, 469-477.

[26] Jain, G.C. and Consul, P.C. (1971) A generalized negative binomial distribution. *SIAM Journal on Applied Mathematics* 21, 501–513.

[27] Lehmann, E. L. (1955). Ordered Families of Distributions. *Ann. Math. Statist.* 26, 399-419.

[28] Liboschik T., Fokianos K. and Fried, R. (2017) tscount: An R Package for Analysis of Count Time Series Following Generalized Linear Models. *Journal of Statistical Software* 82, 1–51.

[29] Lucas, D.D., Yver Kwok, C., Cameron-Smith, P., Graven, H., Bergmann, D., Guilderson, T.P, Weiss, R. and Keeling, R. (2015) Designing optimal greenhouse gas observing networks that consider performance and cost. *Geoscientific Instrumentation Methods and Data Systems* 4, 121—137.

[30] Meyn, S.P. and Tweedie, R.L. (2012) *Markov chains and stochastic stability.* 3nd edition, Springer Science & Business Media.

[31] Neumann, M.H. (2011) Absolute regularity and ergodicity of Poisson count processes. *Bernoulli* 17, 1268-1284.

[32] Ridout, M., Demétrio, C.G. and Hinde, J. (1998) Models for count data with many zeros. *In Proceedings of the XIXth international biometric conference* 19, 179-192.

[33] Rizzo, M.L. and Székely, G.J. (2016) Energy distance. *Wiley Interdisciplinary Reviews: Computational Statistics* 8, 27–38.

[34] Siakoulis, V. (2015) acp: Autoregressive Conditional Poisson. *R package version 2.1.*

[35] Sim, T., Douc, R. and Roueff, F. (2016) General-order observation-driven models. Hal preprint Nb hal-01383554.

[36] Székely, G.J., Rizzo, M.L. and Bakirov, N.K. (2007) Measuring and testing dependence by correlation of distances. *Annals of Statististics* 35, 2769–2794.

[37] Tjøstheim D. (2012) Some recent theory for autoregressive count time series. *Test* 21, 413–438.

[38] Yu, Y. (2009) Stochastic ordering of exponential family distributions and their mixtures. *Journal of Applied Probability* 46, 244–254.