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

Aknouche, Abdelhakim and Francq, Christian

USTHB and Qassim University, 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. We give conditions for consistency and asymptotic normality of the Exponential Quasi-Maximum Likelihood Estimator (QMLE) of the conditional mean parameters. 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.

*I am grateful to the Agence Nationale de la Recherche (ANR), which supported this work via the Project MultiRisk (ANR-16-CE26-0015-02). I also thanks the labex ECODEC.

Keywords: Absolute regularity, Autoregressive Conditional Duration, Count time series models, Distance correlation, Ergodicity, Exponential QMLE, Integer-valued GARCH, Mixing.

1 Introduction

Models for nonnegative time series include the Autoregressive Conditional Duration (ACD) model introduced by Engle and Russell (1998) to analyze 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, 2019). 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 Markov chains with continuous state space (see Meyn and Tweedie, 2009). 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 (2019) 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 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), and Zhu (2012)).

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 explanatory 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 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 $\boldsymbol{\pi} = (\pi_1, \dots, \pi_r)^\top \geq 0$ componentwise. We also consider more general specifications of the form

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

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

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 studies the existence and properties of a process (Y_t) with conditional mean λ_t and cdf satisfying (1.5). Subsection 3.1 assumes a linear conditional mean of the form (1.3) and Subsection 3.2 considers the more general specification (1.4). 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. Section 4 considers the estimation of the parameters involved in the conditional mean λ_t. 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.

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 (2016) 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, $Y \sim NB(r_0, p_0)$, of parameters $r_0 > 0$ and $p_0 \in (0, 1)$ if

$$P(Y = k) = \frac{\Gamma(k + r_0)}{k!\Gamma(r_0)} p_0^{r_0} (1 - p_0)^k, \quad k \in \mathbb{N}.$$

We have $\lambda = r_0(1 - p_0)/p_0$. This distribution also belongs to the exponential family when $p_0 = r_0/(\lambda + r_0))$ and r_0 is fixed (with $\eta = \log(1 - p_0)$).

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 Additive duration models

An alternative to the multiplicative ACD model (1.2) is the additive model

$$Y_t = \lambda_t - E\epsilon_1 + \epsilon_t, \tag{2.3}$$

where (ϵ_t) is a stationary sequence of positive random variables, ϵ_t and λ_t are independent, λ_t satisfies (1.3) or (1.4) with $\lambda_t \geq \omega$, and $\omega \geq E\epsilon_t$ to ensure positivity of λ_t. Any model of this form satisfies (1.5) because $F_\lambda(y) := P(Y_t \leq y \mid \lambda_t = \lambda) = P(\epsilon_1 \leq y + E\epsilon_1 - \lambda)$ is a decreasing function of λ.

2.4 Negative binomial NB(r_0, p_0) with fixed p_0

For any fixed p_0, the negative binomial distribution F_λ with parameter $r_0 = p_0\lambda/(1-p_0)$ apparently does not belong to the one-parameter exponential family. The next Lemma shows that this family of distributions 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 $X \sim NB(r_1, p_0)$ and $Y \sim NB(r_2, p_0)$ 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.

2.5 Gamma distributions

A random variable Y is said to be Gamma distributed $\Gamma(a, b)$ with shape parameter $a > 0$ 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 $\mathrm{Var}(Y_t \mid \mathcal{F}_{t-1}) = \lambda_t/b$. 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.

2.6 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) and Zhu (2012)). One solution consists in assuming that a random element Y of the data set has a zero-inflated Poisson (ZIP) distribution, given by

$$P(Y = k) = \begin{cases} \tau + (1-\tau)e^{-\mu} & \text{if} \quad k = 0 \\ (1-\tau)e^{-\mu}\frac{\mu^k}{k!} & \text{if} \quad k > 0. \end{cases} \tag{2.4}$$

If $\tau \in [0,1]$, the ZIP(τ, μ) distribution (2.4) 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 μ. When $\tau \in [-e^{-\mu}/(1-e^{-\mu}), 0)$ and $\mu > 0$, 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

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

More generally, assume that the baseline cdf is not necessarily Poisson $\mathcal{P}(\mu)$ but the cdf F_λ, and define two zero-inflated distributions by

$$P(Y \le y) = \tau + (1-\tau)F_\lambda(y), \qquad P(Y^* \le y) = \tau + (1-\tau)F_{\lambda^*}(y), \tag{2.5}$$

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* (2.5) *hold true, then $EY \le EY^*$ entails $Y \le_{st} Y^*$.*

3 Probabilistic properties

We first consider the strict stationarity and ergodicity of the linear 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. We also discuss the existence of moments in the case $p = q = 1$. We then extend the stationarity results for general conditional means of the form (1.4), and show the geometric decay of the mixing coefficients in the case where Y_t is valued in $\mathbb{N}$.

3.1 The linear conditional mean case

Theorem 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*

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

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.

Remark 3.1 (The exogenous variables do not matter for stationarity) *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 (2019) made a similar comment on GARCH models with exogenous variables.*

Remark 3.2 (Markovian representation) *The proof of Theorem 3.1 shows the existence of a solution of the form*

$$Y_t = F^{-}_{\lambda_t}(U_t),$$

where λ_t satisfies (1.1) or (1.3) with (3.2), F_λ satisfies (1.5), the sequences (U_t) and $(\boldsymbol{X}_t)$ are independent and (U_t) is iid uniformly distributed in $[0,1]$. It follows that, given $(\boldsymbol{X}_t)$, the process $\boldsymbol{Z}_t := (Y_{t-1}, \dots, Y_{t-q}, \lambda_{t-1}, \dots, \lambda_{t-p})^\top$ is a Markovian process. First note that this excludes conditional means of AR(∞)-type $\lambda_t = \lambda(X_{t-1}, X_{t-2}, \dots)$. This also suggests using Markov chain techniques, as in Meyn and Tweedie (2009). However, when Y_t is integer-valued, those techniques seem difficult to apply. Note also that, in the case (1.1) with $p = q = 1$, the conditional mean satisfies a Stochastic Recurrence Equation (SRE) of the form $\lambda_t = \varphi(\lambda_{t-1}, U_{t-1})$ where $\varphi(\lambda, u) = \omega + \alpha F_\lambda^-(u) + \beta\lambda$. It is also difficult to apply the SRE theory, as developed in Bougerol (1993) and Straumann and Mikosch (2006), because the application $\lambda \mapsto F_\lambda^-(u)$ is not continuous when Y_t is integer-valued, and thus it seems impossible to impose the Cauchy root test constraint

$$E \log \sup_{\lambda \neq \lambda^*} \frac{|\varphi(\lambda, U_1) - \varphi(\lambda^*, U_1)|}{|\lambda - \lambda^*|} < 0$$

required by the SRE theory (see Bougerol, 1993).

Remark 3.3 (Joint stationarity with the exogenous variables) *The stationary solution defined in the proof has a causal Bernoulli shift representation of the form*

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

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

Remark 3.4 (Link with the stationarity of ACD and GARCH) *The square of a GARCH is an ACD model. It has been shown in Subsection 2.2 that any conditional distribution of an ACD model satisfies (1.5). Therefore, when Y_t in Theorem 3.1 corresponds to the square of a GARCH whose squared volatility λ_t follows (1.1), we retrieve the very well known result that an ACD is stationary with finite first-order moments (or a GARCH is stationary with finite second-order moments) if and only if (3.2) holds true.*

From Theorem 3.1, we retrieve that (3.2) ensures the stationarity and ergodicity of the Poisson-INGARCH(p, q) model (see Ferland, Latour and Oraichi, 2006) and of the NB(r_0, p_t)-INGARCH(1,1) model with $p_t = r_0/(\lambda_t + r_0)$ (see Zhu (2011), Christou and Fokianos (2014) and Davis and Liu (2016)). The theorem also provides new stationarity results, examples of which are given in the following corollaries.

Corollary 3.1 (NB(r_t, p_0)-INGARCH) *There exists a stationary and ergodic sequence (Y_t) such that the distribution of Y_t conditional to $\mathcal{F}_{t-1}$ is $NB(p_0\lambda_t/(1-p_0), p_0)$ where λ_t satisfies either (1.1) or (1.3) with $(\boldsymbol{X}_t)$ stationary and ergodic if (3.2) holds.*

Conversely, if there exists (Y_t) such that $Y_t \mid \mathcal{F}_{t-1} \sim NB(p_0\lambda_t/(1-p_0), p_0)$ with $EY_t = m < \infty$ and λ_t satisfies (1.3), then $E\boldsymbol{\pi}^\top \boldsymbol{X}_t < \infty$ and (3.2) holds.

Corollary 3.1 is a direct consequence of Theorem 3.1 and Subsection 2.4. This result has been conjectured by Aknouche, Bendjeddou and Touche (2018) but, to our knowledge, it had not yet been formally proven.

We now consider a ZIP(τ, μ) distribution of the form (2.4). Zhu (2012) investigated such conditional distributions with an INGARCH dynamics on the parameter μ. Denoting by λ the mean of the ZIP(τ, μ) distribution, we have $\mu = \lambda/(1-\tau)$. To make the link between Zhu (2012) and our framework, note that if τ is fixed and $\mu_t = \omega + \alpha Y_{t-1} + \beta\mu_{t-1}$ then $\lambda_t = (1-\tau)\omega + (1-\tau)\alpha Y_{t-1} + \beta\mu_{t-1}$. Therefore, a linear dynamics on μ_t (as in Zhu 2012) is equivalent to a linear dynamics on λ_t, under an appropriate change of notation. Since τ is fixed, denote by F_λ^{ZIP} the cdf of the ZIP$(\tau, \lambda/(1-\tau))$ distribution.

Corollary 3.2 (ZIP) *There exists a stationary and ergodic sequence (Y_t) such that $Y_t \mid \mathcal{F}_{t-1} \sim F_{\lambda_t}^{ZIP}$ with $\tau \in [0, 1]$ and λ_t satisfies either (1.1) or (1.3), $(\boldsymbol{X}_t)$ being stationary and ergodic, if (3.2) holds.*

Conversely, if there exists (Y_t) such that $Y_t \mid \mathcal{F}_{t-1} \sim F_{\lambda_t}^{ZIP}$ with $EY_t = m < \infty$ and λ_t satisfies (1.3), then $E\boldsymbol{\pi}^\top \boldsymbol{X}_t < \infty$ and (3.2) holds.

Corollary 3.2, which is a direct consequence of Theorem 3.1 and Subsection 2.6, shows the strict stationarity and ergodicity under (3.2), Zhu (2012) having showed the mean stationarity under the same condition. The same results could be trivially obtained for zero-inflated negative binomial conditional distributions.

We now give conditions for the existence of moments for the POLI(1,1) model. For simplicity of notation, we write α and β instead of α_1 and β_1. Theorem 3.1 showed that, for strict stationarity (and ergodicity), the precise form of the conditional distribution is not important (provided it satisfies the stochastic-equal-mean order property (1.5)). For the second-order stationarity, and more generally for the existence of moments, the next proposition shows that the shape of the conditional distribution matters.

Theorem 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 $\ell \geq 2$, there exist nonnegative coefficients $a_j(0), a_j(1), \ldots, a_j(j)$ for all $j \leq \ell$ such that*

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

Under (3.2), let (Y_t) be a stationary sequence such that $P(Y_t \leq y \mid \mathcal{F}_{t-1}) = F_{\lambda_t}(y)$, 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_0, p_t)) *The first moments $m_i = EY^i$ of Y following the $BN(r_0, r_0/(\lambda + r_0))$ distribution are*

$$\begin{aligned} m_1 &= \lambda, \quad m_2 = \lambda + \frac{1+r_0}{r_0}\lambda^2, \quad m_3 = \lambda + 3\frac{1+r_0}{r_0}\lambda^2 + \frac{2+3r_0+r_0^2}{r_0^2}\lambda^3, \\ m_4 &= \lambda + 7\frac{1+r_0}{r_0}\lambda^2 + 6\frac{2+3r_0+r_0^2}{r_0^2}\lambda^3 + \frac{6+11r_0+6r_0^2+r_0^3}{r_0^3}\lambda^4. \end{aligned}$$

It follows that (3.3) holds with

$$a(2) = \frac{1+r_0}{r_0}, \quad a(3) = \frac{2+3r_0+r_0^2}{r_0^2}, \quad a(4) = \frac{6+11r_0+6r_0^2+r_0^3}{r_0^3}.$$

Theorem 3.2 shows that the POLI(1,1) model with BN$(r_0, r_0/(\lambda_t + r_0))$ *conditional distribution admits a moment of*

$$\text{order } 2 \quad \text{iff} \quad (\alpha+\beta)^2 + \frac{\alpha^2}{r_0} < 1, \tag{3.5}$$

$$\text{order } 3 \quad \text{iff} \quad (\alpha+\beta)^3 + \frac{3\alpha^2(\alpha+\beta)}{r_0} + \frac{2\alpha^3}{r_0^2} < 1, \tag{3.6}$$

$$\text{order } 4 \quad \text{iff} \quad (\alpha+\beta)^4 + \frac{6\alpha^2(\alpha+\beta)^2}{r_0} + \frac{\alpha^3(11\alpha+8\beta)}{r_0^2} + \frac{6\alpha^4}{r_0^3} < 1. \tag{3.7}$$

Figure 1 displays these moment conditions when $r_0 = 1$.

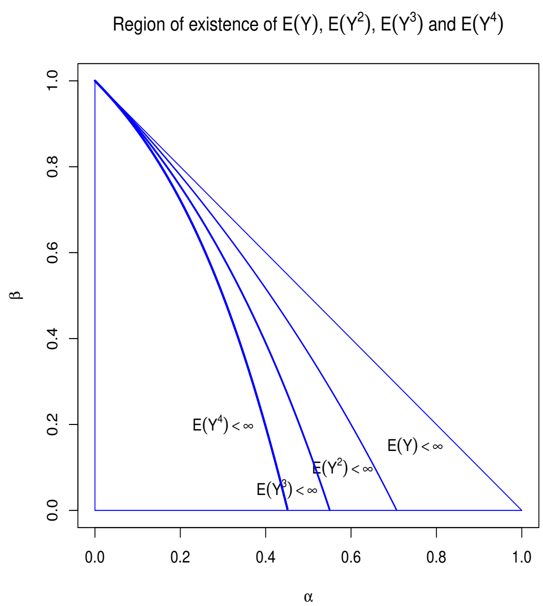

Figure 1: Moment conditions for the INGARCH(1,1) process with NB(r_0, p_t) conditional distribution.

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_0)) *Now consider the INGARCH(1,1) model with BN*$(p_0\lambda_t/(1-p_0), p_0)$ *conditional distribution. By Jain and Consul (1971), the moments* $m_\ell = EY^\ell$ *of*

$Y \sim NB(r, p_0)$ *satisfy*

$$m_\ell = p_0 \lambda \sum_{j=0}^{\ell-1} \binom{\ell-1}{j} \left(m_j + \tfrac{1-p_0}{\lambda p_0} m_{j+1} \right), \qquad \ell \geq 1.$$

It follows that

$$m_1 = \lambda, \qquad m_2 = \lambda^2 + \frac{1}{p_0}\lambda, \qquad m_3 = \lambda^3 + \frac{3}{p_0}\lambda^2 + \frac{2-p_0}{p_0^2}\lambda,$$

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

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

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

3.2 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)

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

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 ergodicity, which entails the central limit theorem under some moment conditions.

Theorem 3.3 *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* $(y_i, y_i') \in [0, +\infty)^2$, $i = 1, \dots, q$ *and for all* $(\lambda_j, \lambda_j') \in (0,\infty)^2$, $j = 1, \dots, p$,

$$\begin{aligned} &\left| g(y_1, \dots, y_q, \lambda_1, \dots, \lambda_p) - g(y_1', \dots, y_q', \lambda_1', \dots, \lambda_p') \right| \\ &\leq \sum_{i=1}^{q} \alpha_i |y_i - y_i'| + \sum_{j=1}^{p} \beta_j |\lambda_j - \lambda_j'|. \end{aligned} \tag{3.8}$$

If

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1, \tag{3.9}$$

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*

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

Remark 3.5 (On the integer value assumption) *Showing the ergodicity is much more difficult for count time series models than for standard time series models such as ARMA, GARCH or ACD. Surprisingly enough, when the stationarity is established, showing geometric mixing seems simpler for integer valued observations than for continuous state space observations. We used a simple coupling technique that works when observations are integer valued. Establishing a mixing property without that assumption remains an open problem.*

Note also that (3.8) *is satisfied when* (1.3) *holds. Therefore* (3.9) *and* Y_t *valued in* $\mathbb{N}$ *also entail geometric mixing in the linear case* (1.3)*.*

4 Exponential QMLE of the conditional mean

The previous section showed that simple stationarity and ergodicity conditions can be obtained when the conditional distribution is not fully specified, but satisfies the stochastic-equal-mean order property (1.5). This section shows that the conditional mean parameter can be consistently estimated by using a QMLE based on a member of the exponential family. We concentrate on the Exponential QMLE because this estimator coincides with the Maximum Likelihood Estimator (MLE) in the benchmark ACD model (1.2) when z_t follows the Exponential $\Gamma(1, 1)$ distribution.

Let $Y_1, .., Y_n$ be observations with conditional mean of the form (1.4):

$$\lambda_t = \lambda_t(\boldsymbol{\theta}_0) = g(Y_{t-1}, \ldots, Y_{t-q}, \lambda_{t-1}, \ldots, \lambda_{t-p}; \boldsymbol{\theta}_0) + \pi(\boldsymbol{X}_{t-1}; \boldsymbol{\theta}_0), \tag{4.1}$$

where $(\boldsymbol{X}_t)$ is a stationary and ergodic process and $\boldsymbol{\theta}_0$, the true parameter, belongs to some parametric space $\Theta \subset \mathbb{R}^d$. The conditional distribution of the model may be unknown, but assume:

A1 $Y_t \mid \mathcal{F}_{t-1} \sim F_{\lambda_t}$ where F_λ satisfies (1.5).

Let us approximate $\lambda_t(\boldsymbol{\theta})$ by the observable proxy $\widetilde{\lambda}_t(\boldsymbol{\theta})$, given by

$$\widetilde{\lambda}_t(\boldsymbol{\theta}) = g(Y_{t-1}, \ldots, Y_{t-q}, \widetilde{\lambda}_{t-1}, \ldots, \widetilde{\lambda}_{t-p}; \boldsymbol{\theta}) + \pi(\boldsymbol{X}_{t-1}; \boldsymbol{\theta}), \quad t \geq q+1,$$

where $\widetilde{\lambda}_q(\boldsymbol{\theta}), \ldots, \widetilde{\lambda}_{q+1-p}(\boldsymbol{\theta})$ are fixed initial values for any $\boldsymbol{\theta} \in \Theta$. When (1.4) reduces to (1.3), we have $\boldsymbol{\theta} = (\omega, \alpha_1, \ldots, \beta_p, \boldsymbol{\pi}^\top)^\top$ and

$$\widetilde{\lambda}_t(\boldsymbol{\theta}) = \omega + \sum_{i=1}^{q} \alpha_i Y_{t-i} + \sum_{j=1}^{p} \beta_j \widetilde{\lambda}_{t-j}(\boldsymbol{\theta}) + \boldsymbol{\pi}^\top \boldsymbol{X}_{t-1}, \quad t \geq q+1. \tag{4.2}$$

Wedderburn (1974) and Gouriéroux, Monfort and Trognon (1984) demonstrated that, under some high-level assumptions, a MLE is a QMLE – that is the estimator remains consistent even when the conditional distribution is misspecified – for estimating a conditional mean parameter if and only if it is based on a member of the exponential family (like Poisson or Exponential). Ahmad and Francq (2016) give regularity conditions for consistency and asymptotic normality (CAN) of the Poisson QMLE (PQMLE) defined by

$$\widehat{\boldsymbol{\theta}}_P = \arg\max_{\boldsymbol{\theta} \in \Theta} \sum_{t=q+1}^{n} \left\{ Y_t \log\left(\widetilde{\lambda}_t(\boldsymbol{\theta})\right) - \widetilde{\lambda}_t(\boldsymbol{\theta}) \right\}.$$

Aknouche, Bendjeddou and Touche (2018) considered the (profile) negative binomial QMLE

$$\widehat{\boldsymbol{\theta}}_{NB} = \arg\max_{\boldsymbol{\theta} \in \Theta} \sum_{t=q+1}^{n} Y_t \log\left(\frac{\widetilde{\lambda}_t(\boldsymbol{\theta})}{r + \widetilde{\lambda}_t(\boldsymbol{\theta})}\right) - r \log\left\{ r + \widetilde{\lambda}_t(\boldsymbol{\theta}) \right\}.$$

For integer-valued observations, these two estimators may seem natural because they give the maximum likelihood estimate (MLE) in the benchmark Poisson or negative binomial INGARCH models, respectively. For positive observations, these estimators remain generally consistent. However, in case of duration data, the Exponential QMLE (EQMLE) given by

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \Theta} \sum_{t=q+1}^{n} \widetilde{l}_t(\boldsymbol{\theta}), \quad \widetilde{l}_t(\boldsymbol{\theta}) = Y_t / \widetilde{\lambda}_t(\boldsymbol{\theta}) + \log \widetilde{\lambda}_t(\boldsymbol{\theta}), \tag{4.3}$$

might be preferred because it corresponds to the MLE when the DGP is the standard Exponential ACD model. In this section we give regularity conditions for CAN of this EQMLE. The main condition is the stochastic-equal-mean order property (1.5). In addition we need to consider the following assumptions, similar to those made by Ahmad and Francq (2016) for the strong consistency of their PQMLE.

A2 $g(\cdot) = g(\cdot; \boldsymbol{\theta}_0)$ is a contraction in the sense of (3.8) and (3.9), substituting $\boldsymbol{\theta}_0$ for $\boldsymbol{\theta}$. In addition, for all $\boldsymbol{\theta} \in \Theta$, $\sum_{j=1}^{p} \beta_j < 1$.

A3 $\boldsymbol{\theta} \mapsto \lambda_t(\boldsymbol{\theta})$ is *a.s.* continuous and valued in $(\underline{\omega}, \infty)$ and $\forall t \geq 1$, $\widetilde{\lambda}_t(\boldsymbol{\theta}) > \underline{\omega}$, *a.s.* for some $\underline{\omega} > 0$.

A4 $EY_1 < \infty$.

A5 $\lambda_t(\boldsymbol{\theta}) = \lambda_t(\boldsymbol{\theta}_0)$ *a.s. iff* $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

A6 $\boldsymbol{\theta}_0 \in \Theta$ and Θ is compact.

By Theorem 3.3, Assumptions **A1** and **A2** ensure the stationarity and ergodicity of $\{Y_t, t \in \mathbb{Z}\}$. Assumption **A3** holds true if, for instance, the function $\pi(\boldsymbol{x}; \cdot)$ is continuous for all $\boldsymbol{x} \in \mathbb{R}^r$, and the function $g(\boldsymbol{x}; \cdot)$ is continuous and valued in $(\underline{\omega}, \infty)$ for all $\boldsymbol{x} \in \mathbb{R}^{p+q}$. In the proof of Theorem 3.3, λ_t is defined as the limit in L^1 of a Cauchy sequence $(\lambda_t^{(k)})_k$. Under the assumption that $E\pi(\boldsymbol{X}_1) < \infty$, $\lambda_t^{(k)}$ belongs to L^1 for all k. By the L^1 completeness theorem, the limit λ_t also belongs to L^1. It follows that $EY_t = E\lambda_t < \infty$, and thus **A4** is satisfied by the solution given in the proof of Theorem 3.3 when $E\pi(\boldsymbol{X}_1) < \infty$. Assumption **A5** is an identifiability condition, and the compactness assumption **A6** is standard.

Now, let us further comment the assumptions in the linear case (1.3). First note that **A3** is satisfied when $\inf_\Theta w > 0$. Under **A1**, let the polynomials $\mathcal{A}_{\boldsymbol{\theta}}(z) = \sum_{i=1}^{q} \alpha_i z^i$ and $\mathcal{B}_{\boldsymbol{\theta}}(z) = 1 - \sum_{i=1}^{p} \beta_i z^i$. Consider the case where $r = 0$ (no exogenous variables). When $p = 0$, it is easy to see that **A5** is satisfied when, for all $\lambda > 0$, the conditional distribution F_λ is not degenerated. When $p > 0$, it suffices to assume further that $\mathcal{A}_{\boldsymbol{\theta}_0}(z)$ and $\mathcal{B}_{\boldsymbol{\theta}_0}(z)$ have no common root, $\mathcal{A}_{\boldsymbol{\theta}_0}(1) \neq 0$ and $\alpha_{0q} + \beta_{0p} \neq 0$ (see A4 page 174 in Francq and Zakoian (2019) for an analog condition in the GARCH(p, q) framework). The case $r > 0$ is trickier. Obviously, it is necessary that the components of the vector $\boldsymbol{X}_t$ are not linearly dependent.

Using the arguments of Theorem 1 in Francq and Thieu (2019), the identifiability condition **A5** can be shown by assuming, in addition, that Y_t is not a measurable function of $(\boldsymbol{X}_u)$. Note that this condition is satisfied for the solution given in the proof of Theorem 3.3 because $(\boldsymbol{X}_t)$ and (U_t) are supposed to be independent and $F_{\boldsymbol{\lambda}}^{-}(U_t)$ is not degenerated.

Theorem 4.1 *Let $\{Y_t, t \in \mathbb{Z}\}$ be a strictly stationary and ergodic process and $\widehat{\boldsymbol{\theta}}$ a sequence of estimators satisfying (4.3). Under* ***A1*****–*****A6****, we have*

$$\widehat{\boldsymbol{\theta}} \to \boldsymbol{\theta}_0 \quad a.s. \quad as \quad n \to \infty.$$

Remark 4.1 (Consistency of the PQMLE) *Ahmad and Francq (2016) studied $\widehat{\boldsymbol{\theta}}_P$ in the case of integer-valued observations, without exogenous variables, but it is easy to see that the PQMLE remains consistent in the present framework, under the assumptions of Theorem 4.1, except that* **A4** *is replaced by the marginally stronger assumption*

A4' $EY_1^{1+\varepsilon} < \infty$ *for some* $\varepsilon > 0$.

This assumption is required to show that $EY_t|\log \lambda_t(\boldsymbol{\theta})| < \infty$ (instead of showing that $EY_t/\lambda_t(\boldsymbol{\theta}) < \infty$ for the EQMLE).

For $\boldsymbol{y} \in \mathbb{R}^q$ and $\boldsymbol{\lambda} \in \mathbb{R}^p$, consider the partial derivatives

$$\boldsymbol{D}_{\boldsymbol{\theta}}\left(\boldsymbol{y}^\top, \boldsymbol{\lambda}^\top, \boldsymbol{\theta}\right) = \frac{\partial}{\partial \boldsymbol{\theta}} g\left(\boldsymbol{y}^\top, \boldsymbol{\lambda}^\top; \boldsymbol{\theta}\right), \qquad \boldsymbol{D}_{\boldsymbol{\lambda}}\left(\boldsymbol{y}^\top, \boldsymbol{\lambda}^\top, \boldsymbol{\theta}\right) = \frac{\partial}{\partial \boldsymbol{\lambda}} g\left(\boldsymbol{y}^\top, \boldsymbol{\lambda}^\top; \boldsymbol{\theta}\right).$$

By the chain rule, with the R notation for indices, we have

$$\frac{\partial}{\partial \boldsymbol{\theta}} g\left(Y_{t-1:q}, \lambda_{t-1:p}; \boldsymbol{\theta}\right) = \boldsymbol{D}_{\boldsymbol{\theta}} + \left(\frac{\partial \lambda_{t-1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \cdots \frac{\partial \lambda_{t-p}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) \boldsymbol{D}_{\boldsymbol{\lambda}}, \tag{4.4}$$

where

$$\boldsymbol{D}_{\boldsymbol{\theta}} = \boldsymbol{D}_{\boldsymbol{\theta}}\left(Y_{t-1:q}, \lambda_{t-1:p}; \boldsymbol{\theta}\right), \quad \boldsymbol{D}_{\boldsymbol{\lambda}} = \boldsymbol{D}_{\boldsymbol{\lambda}}\left(Y_{t-1:q}, \lambda_{t-1:p}; \boldsymbol{\theta}\right).$$

Denote by $\rho(\boldsymbol{A})$ the spectral radius of a square matrix $\boldsymbol{A}$ and let I_p be the identity matrix of order p. The following assumption is used to show that the initial values are unimportant for the asymptotic distribution.

A7 For $\boldsymbol{y} \in \mathbb{R}^q$ and $\boldsymbol{\lambda} \in \mathbb{R}^p$, the function $\boldsymbol{\theta} \mapsto g\left(\boldsymbol{y}^\top, \boldsymbol{\lambda}^\top; \boldsymbol{\theta}\right)$ and $\boldsymbol{\lambda} \mapsto g\left(\boldsymbol{y}^\top, \boldsymbol{\lambda}^\top; \boldsymbol{\theta}\right)$ are continuously differentiable. The random variable

$$u_t = \sup_{\boldsymbol{\theta}\in\Theta}\left\{\|\boldsymbol{D_\theta}\| + \left\|\frac{\partial\pi(\boldsymbol{X}_{t-1};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\| + \sup_{\boldsymbol{\lambda}\geq\boldsymbol{0}}\left(\left\|\frac{\partial\boldsymbol{D_\theta}\left(Y_{t-1:q},\boldsymbol{\lambda}^\top;\boldsymbol{\theta}\right)}{\partial\boldsymbol{\lambda}^\top}\right\| + \left\|\frac{\partial\boldsymbol{D_\lambda}\left(Y_{t-1:q},\boldsymbol{\lambda}^\top;\boldsymbol{\theta}\right)}{\partial\boldsymbol{\lambda}}\right\|\right)\right\}$$

In the linear case (1.3), we have

$$\boldsymbol{D_\theta} = \left(1, Y_{t-1}, \dots, Y_{t-q}, \lambda_{t-1}, \dots, \lambda_{t-p}, \boldsymbol{0}^\top\right)^\top, \qquad \boldsymbol{D_\lambda} = \left(\beta_1, \dots, \beta_p\right)\top.$$

It is thus easy to verify that, under **A2**, Assumption **A7** is always satisfied in the linear case. Let $l_t(\boldsymbol{\theta})$ be defined in the same way as $\widetilde{l}_t(\boldsymbol{\theta})$ in (4.3) with $\lambda_t(\boldsymbol{\theta})$ in place of $\widetilde{\lambda}_t(\boldsymbol{\theta})$. The following extra assumptions are standard.

A8 $\boldsymbol{\theta}_0$ belongs to the interior of Θ.

A9 The conditional variance $\upsilon_t(\boldsymbol{\theta}_0) := \mathrm{Var}(Y_t | \mathcal{F}_{t-1})$ is *a.s.* finite.

A10 $\frac{\partial^2\lambda_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}$ and $\frac{\partial^2\widetilde{\lambda}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}$ exist and are continuous, the matrices

$$\boldsymbol{I} = E\left(\frac{\upsilon_t(\boldsymbol{\theta}_0)}{\lambda_t^4(\boldsymbol{\theta}_0)}\frac{\partial\lambda_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial\lambda_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\right) \quad \text{and} \quad \boldsymbol{J} = E\left(\frac{1}{\lambda_t^2(\boldsymbol{\theta}_0)}\frac{\partial\lambda_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial\lambda_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\right)$$

are finite, and $\boldsymbol{J}$ is nonsingular.

A11 There is a neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ such that $E\left(\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}\left\|\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\|\right) < \infty$.

Let us go back to the linear case (1.3). By adapting Remark 2.3 of Ahmad and Francq (2016) to the presence of exogenous variables, it is easy to see that $\boldsymbol{J}$ exists under **A2**, **A4**, **A8** and $E\|\boldsymbol{X}_1\| < \infty$. If, in addition, $E\upsilon_t^{1+\varepsilon}(\boldsymbol{\theta}_0) < \infty$ for some $\varepsilon > 0$ then $\boldsymbol{I}$ also exists. The invertibility of $\boldsymbol{J}$ is a consequence of the identifiability conditions discussed before the statement of Theorem 4.1. Similarly, it can be shown that **A11** is entailed by the previous assumptions and **A4′**.

The symbol $\xrightarrow{\mathcal{L}} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ denotes the convergence in distribution to a Gaussian vector with zero mean and variance $\boldsymbol{\Sigma}$ as $n \to \infty$.

Theorem 4.2 *Under the assumptions of Theorem 4.1 and* **A7**–**A11**

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}), \qquad \textit{where} \qquad \boldsymbol{\Sigma} = \boldsymbol{J}^{-1}\boldsymbol{I}\boldsymbol{J}^{-1}.$$

Remark 4.2 (Optimality of the EQMLE) *When the conditional distribution of Y_t is exponential with mean $\lambda_t(\boldsymbol{\theta}_0)$, the conditional variance of Y_t is $\upsilon_t(\boldsymbol{\theta}_0) = \lambda_t^2(\boldsymbol{\theta}_0)$, thus $\boldsymbol{I} = \boldsymbol{J}$ and $\boldsymbol{\Sigma} = \boldsymbol{J}^{-1}$. In such a case, $\widehat{\boldsymbol{\theta}}$ is asymptotically efficient. More generally, $\widehat{\boldsymbol{\theta}}$ is asymptotically efficient within the class of the QMLE's of the linear exponential family (see e.g. Gouriéroux, Monfort and Trognon (1984), Wooldridge (1999)) under the so-called exponential nominal (quadratic) variance assumption*

$$\upsilon_t(\boldsymbol{\theta}_0) = \kappa\lambda_t^2(\boldsymbol{\theta}_0) \text{ for some } \kappa > 0, \tag{4.5}$$

and we then have

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \kappa\boldsymbol{J}^{-1}\right).$$

For example, if $Y_t/\mathcal{F}_{t-1} \sim \Gamma(a, a/\lambda_t)$ then (4.5) holds with $\kappa = \frac{1}{a}$, and the EQMLE is thus an asymptotically optimal QMLE.

Remark 4.3 (Comparison with the PQMLE) *Ahmad and Francq (2016) established CAN of the PQMLE:*

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_P - \boldsymbol{\theta}_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_P), \tag{4.6}$$

where $\boldsymbol{\Sigma}_P = \boldsymbol{J}_P^{-1}\boldsymbol{I}_P\boldsymbol{J}_P^{-1}$, $\boldsymbol{I}_P = E\left(\frac{\upsilon_t(\boldsymbol{\theta}_0)}{\lambda_t^2(\boldsymbol{\theta}_0)}\frac{\partial\lambda_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial\lambda_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\right)$ and $\boldsymbol{J}_P = E\left(\frac{1}{\lambda_t(\boldsymbol{\theta}_0)}\frac{\partial\lambda_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial\lambda_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\right)$. Let us compare the asymptotic variances of the EQMLE and PQMLE for some particular POLI models.

i) *For the conditional distribution $\Gamma(a, a/\lambda_t)$ we have seen in Remark 4.2 that EQMLE is optimal. It can be seen that EQMLE is indeed strictly more efficient than PQMLE.*

ii) *When $Y_t/\mathcal{F}_{t-1} \sim \Gamma(b\lambda_t, b)$, the model satisfies the Poisson nominal (linear) variance assumption (cf. Wooldridge, 1999)*

$$\upsilon_t(\boldsymbol{\theta}_0) = \tfrac{1}{b}\lambda_t(\boldsymbol{\theta}_0),$$

under which PQMLE is the most efficient estimate within all the QMLEs belonging to the exponential family. Thus, somewhat surprisingly, PQMLE (which is built from a *discrete distribution) is asymptotically more efficient than EQMLE in this continuous distribution framework, with*

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_P-\boldsymbol{\theta}_0\right)\xrightarrow{\mathcal{L}}\mathcal{N}\left(\mathbf{0},\tfrac{1}{b}\boldsymbol{J}_P^{-1}\right),\qquad \tfrac{1}{b}\boldsymbol{J}_P^{-1}\prec\boldsymbol{\Sigma}=\boldsymbol{J}^{-1}\boldsymbol{I}\boldsymbol{J}^{-1},\tag{4.7}$$

where $\boldsymbol{A}\prec\boldsymbol{B}$ means that $\boldsymbol{B}-\boldsymbol{A}$ is definite positive. Indeed, omitting "$(\boldsymbol{\theta}_0)$" we have

$$Var\left(\boldsymbol{J}^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{F^{-}_{\lambda_t}(U_t)-\lambda_t}{\lambda_t^2}\frac{\partial\lambda_t}{\partial\boldsymbol{\theta}}-\boldsymbol{J}_P^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{F^{-}_{\lambda_t}(U_t)-\lambda_t}{\lambda_t}\frac{\partial\lambda_t}{\partial\boldsymbol{\theta}}\right)=\boldsymbol{\Sigma}-\tfrac{1}{b}\boldsymbol{J}_P^{-1}.$$

Similarly to Ahmad and Francq (2016), a consistent estimate of the asymptotic variance $\boldsymbol{\Sigma}$ is $\widehat{\boldsymbol{\Sigma}}=\widehat{\boldsymbol{J}}^{-1}\widehat{\boldsymbol{I}}\widehat{\boldsymbol{J}}^{-1}$ with

$$\widehat{\boldsymbol{I}}=\frac{1}{n}\sum_{t=1}^{n}\left(\frac{Y_t-\widetilde{\lambda}_t(\widehat{\boldsymbol{\theta}})}{\widetilde{\lambda}_t^2(\widehat{\boldsymbol{\theta}})}\right)^2\frac{\partial\widetilde{\lambda}_t(\widehat{\boldsymbol{\theta}})\partial\widetilde{\lambda}_t(\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\text{ and }\widehat{\boldsymbol{J}}=\frac{1}{n}\sum_{t=1}^{n}\frac{1}{\widetilde{\lambda}_t^2(\widehat{\boldsymbol{\theta}})}\frac{\partial\widetilde{\lambda}_t(\widehat{\boldsymbol{\theta}})\partial\widetilde{\lambda}_t(\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}.$$

Monte Carlo experiments, not presented here for the sake of brevity, confirm the asymptotic results of this section in finite samples.

5 Testing the multiplicative form of duration models

Instead of a standard ACD duration model (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 $Y_1,\dots,Y_n$, the hypothesis H_0 can be tested by using the empirical distance covariance (see Székely et al. (2007), Rizzo and Székely (2016), and the references therein)

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

where $\widehat{\varphi}_{z,\lambda}$, $\widehat{\varphi}_z$ and $\widehat{\varphi}_\lambda$ 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 $\widetilde{\lambda}_t(\widehat{\boldsymbol{\theta}})$ defined by (4.2). 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 the QMLE $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}_n(Y_1, \dots, Y_n)$ defined by (4.3), the test statistics $\mathcal{V}_n^2 = \mathcal{V}_n^2(Y_1, \dots, Y_n)$, and the residuals $\widehat{z}_t = Y_t/\widetilde{\lambda}_t(\widehat{\boldsymbol{\theta}})$ for $t = q+1, \dots, n$. Denote by F_n the empirical distribution of $\{\widehat{z}_t/s_n, t = 1+q, \dots, n\}$ where $s_n = \sum_{t=q+1}^n \widehat{z}_t/(n-q)$ (with this scaling factor, the expectation of the distribution F_n is equal to 1).

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

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

(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 $\mathcal{V}_{n,(1)}^{*2} \leq \dots \leq \mathcal{V}_{n,(B)}^{*2}$ 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 (in the numerical illustration of the next subsection, we also used $B = 999$ and noticed that the results were similar for $B = 99$ and $B = 999$). 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

	$n = 500$			$n = 1000$		
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.

kind is well controlled when $\alpha = 1\%$, but the test is 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 model (1.1) with $(p, q) = (2, 1)$, the parameter estimates of the QMLE (4.3) are $\widehat{\omega} = 0.680$, $\widehat{\alpha}_1 = 0.498$, $\widehat{\beta}_1 = 0.271$, $\widehat{\beta}_2 = 0.040$. As shown in the bottom-left panel of Figure 2, the autocorrelation function (ACF) of the residuals $\widehat{z}_t = Y_t/\widetilde{\lambda}_t(\widehat{\boldsymbol{\theta}})$ 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 independent. Indeed, a kernel density estimator of the bootstrapped 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 $B = 999$ bootstrap replications used to approximate the distribution of $\mathcal{V}_n^2$ under the null. The estimated p-value is thus $1/1000 = 0.001$. On a personal computer with a 2.80 GHz processor, the bootstrap-based test run time was around 600 seconds.

The distance covariance test concludes that a non-multiplicative POLI model is better than an ACD model for this particular series, but the test is not informative about the distribution F_λ. We therefore tried several specifications for the conditional distribution F_λ: the Exponential (ACD), the $\Gamma(a, a/\lambda_t)$ (G-ACD), the $\Gamma(b\lambda_t, b)$ (G-POLI), and two additive models of the form (2.3) in which ϵ_t is assumed to follow a $\Gamma(a, b)$ distribution (G-Add) or a Fisk distribution (F-Add) with density $f(y) = ab(ay)^{b-1}/(1 + (ay)^b)^2 1_{y>0}$, where $a > 0$

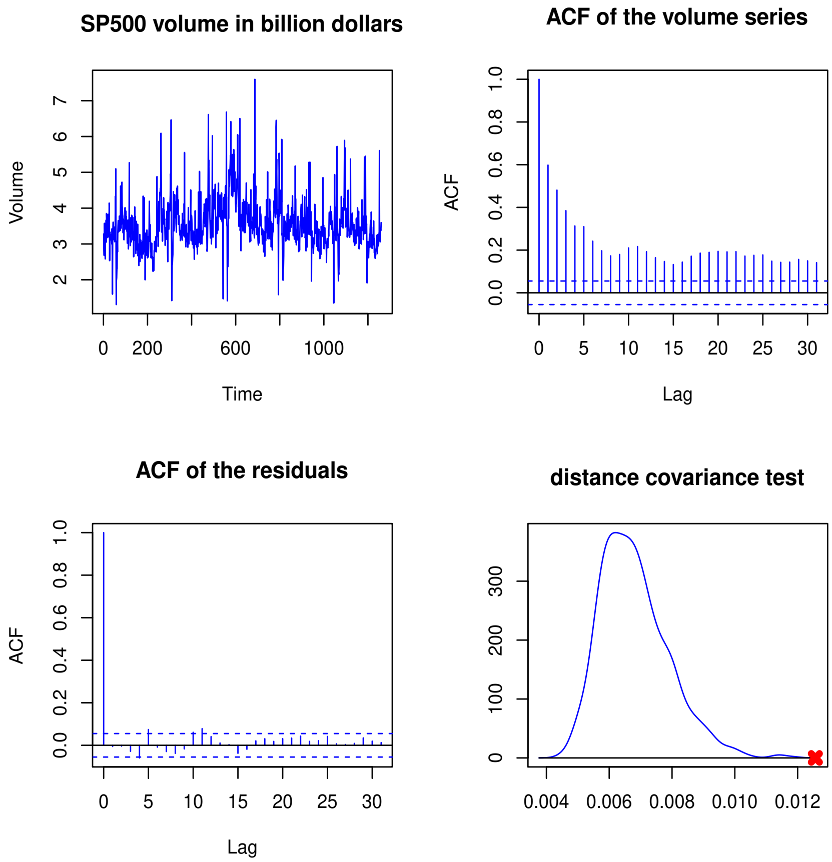

Figure 2: S&P 500 transaction volume from 3/10/2013 to 3/10/2018, ACF on the observed series, ACF on the residuals of the POLI(2,2) model, distribution of the distance covariance under the null hypothesis of multiplicative form, and observed distance covariance (cross symbol).

is a scale parameter and $b > 0$ is a shape parameter. For instance, the Fisk distribution is used for hydrological stream flow modeling, or for the distribution of wealth in economics. These models being fully parametric, they have been estimated by maximum-likelihood. Table 2 shows that, according to the usual Akaike and bayesian information criteria (AIC and BIC), the F-Add model outperforms the other models. This is certainly due to the fact that the Fisk distribution can better take into account the fat tails of the conditional distribution of the series (see the top-left panel of Figure 2) than the Gamma distribution. Note that the Fisk distribution admits finite moments of order less than b only, while the Gamma distribution admits moments of any order. Figure 3 compares the histograms of the Probability Integral Transform (PIT) of the ACD and F-Add models, *i.e.* the empirical distributions of $\widehat{F}_{\widehat{\lambda}_t}(Y_t)$, where λ_t and F_λ are estimated by the MLE of the two models. Note that if the actual conditional distribution of Y_t is the continuous cdf F_{λ_t}, then $F_{\lambda_t}(Y_t)$ is uniformly distributed on $[0, 1]$. Given this graph, the ACD is clearly rejected, while there is no visible evidence against the F-Add model. Indeed, similar PIT histograms are obtained on simulations of the F-Add model.

	ACD	G-ACD	G-POLI	G-Add	F-Add
AIC	5657.409	1871.527	1888.031	1927.031	1636.941
BIC	5677.933	1897.181	1913.685	1957.816	1667.726

Table 2: AIC and BIC of the different models for the S&P 500 transaction volume series.

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 4 displays the time series obtained by one of these tracers. The partial autocorrelogram suggests that the simple model (1.1) with $q = 1$ and $p = 0$ could be sufficient to summarize the dynamics of the conditional mean. The distance covariance test is not conclusive, since the

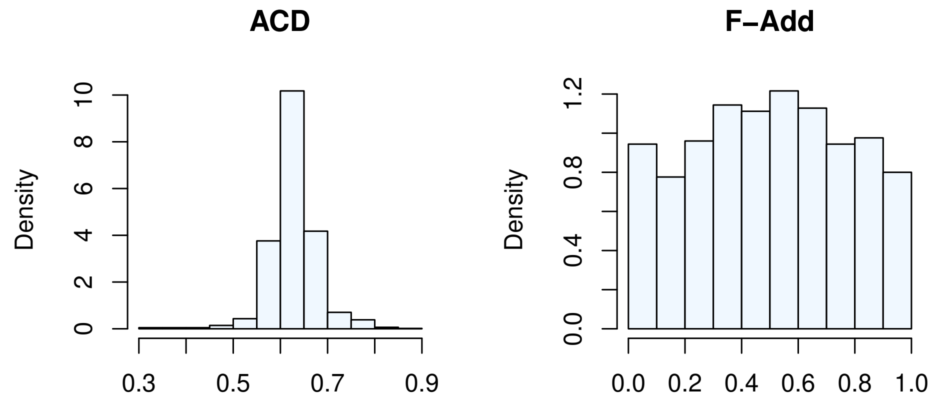

Figure 3: Probability integral transform (PIT) histograms for the ACD and F-Add models.

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. We denote by ZIE-ACD the model of the ACD form (1.2) where z_t follows a zero-inflated exponential distribution, $i.e.$ the model

$$Y_t = \lambda_t z_t, \quad \lambda_t = \omega + \alpha Y_{t-1}, \quad z_t \sim \tau\delta_0(x) + (1-\tau)\mu e^{-\mu x} 1_{x>0},$$

with standard notation for the mixture distribution, and $\mu = 1-\tau$ in order to have $Ez_t = 1$. We denoted by ZIG-ACD the same model where, in the Radon-Nikodym density of z_t, the exponential distribution is replaced by the $\Gamma(a, (1-\tau)a)$ law. Note that the conditional distribution of Y_t is then $Y_t \mid \mathcal{F}_{t-1} \sim \tau\delta_0 + (1-\tau)\Gamma(a, (1-\tau)a/\lambda_t)$. We also considered the model

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

Since this model can not be written in ACD multiplicative form (1.2) (its conditional variance is not proportional to the square of its mean), we called it ZIG-POLI. The three models have been estimated by maximum-likelihood on 15 series of GHG concentrations. Table 3 shows that, according to the AIC and BIC criteria, the POLI model is almost always preferable to the ACD models. On the series displayed in Figure 4 (corresponding to Series 1 of Table 3), the maximum-likelihood estimates of the ZIG-POLI parameters are $\widehat{\omega} = 0.0020$, $\widehat{\alpha} = 0.6888$, $\widehat{\tau} = 0.1743$ and $\widehat{b} = 297.0$.

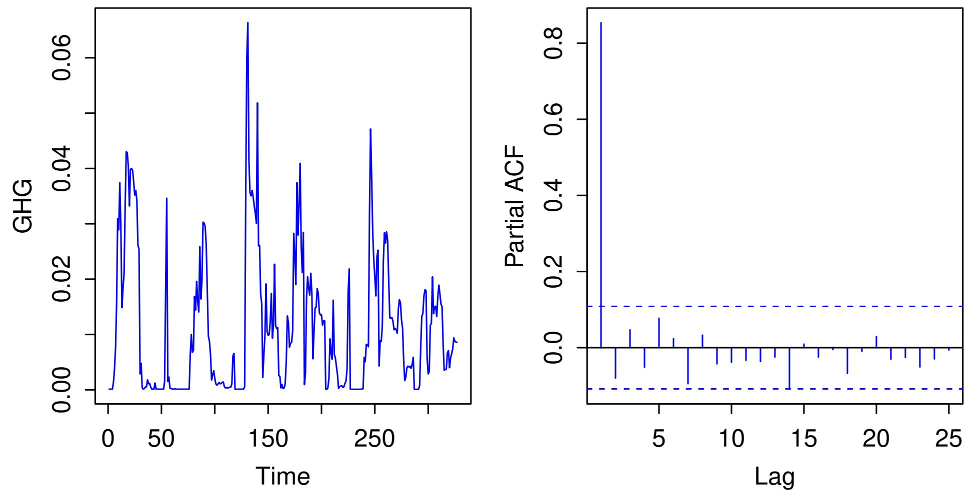

Figure 4: Greenhouse gas time series concentration every 6 hours from May 10 to July 31, 2010, and empirical partial autocorrelations of the time series.

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. In Sections 2-3, we present a unified approach to investigate stationarity and other probabilistic properties of many, seemingly distinct, models of count and durations time series. Section 4 shows that the approach also allows for a unified treatment in terms of estimation of the conditional mean. 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).

	AIC			BIC		
	ZIE-ACD	ZIG-ACD	ZIG-POLI	ZIE-ACD	ZIG-ACD	ZIG-POLI
Series 1	-1573.66	-1626.76	**-1703.70**	-1562.39	-1611.72	**-1688.67**
Series 2	-293.66	-312.56	**-417.66**	-282.38	-297.52	**-402.62**
Series 3	-114.97	-123.42	**-233.31**	-103.69	-108.39	**-218.27**
Series 4	-1154.97	-1172.95	**-1210.19**	-1143.70	-1157.91	**-1195.15**
Series 5	-1552.91	-1571.89	**-1627.29**	-1541.64	-1556.86	**-1612.26**
Series 6	-1089.47	-1090.13	**-1251.73**	-1078.20	-1075.10	**-1236.70**
Series 7	1021.05	1019.35	**949.97**	1032.33	1034.39	**965.01**
Series 8	322.52	308.68	**304.59**	333.80	323.72	**319.62**
Series 9	327.65	324.13	**213.92**	338.93	339.17	**228.96**
Series 10	-911.84	-959.47	**-965.92**	-900.57	-944.43	**-950.89**
Series 11	1103.19	1063.01	**1005.96**	1114.46	1078.04	**1020.99**
Series 12	1611.99	1404.65	**1403.94**	1623.26	1419.69	**1418.98**
Series 13	-862.05	-879.64	**-915.15**	-850.77	-864.60	**-900.11**
Series 14	2586.31	**1061.56**	1068.98	2597.59	**1076.60**	1084.02
Series 15	779.00	775.85	**734.78**	790.27	790.89	**749.82**

Table 3: Information criteria of ACD and POLI models on 15 series of GHG concentrations (the minimal information criteria are displayed in boldface).

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 research.

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

$$\frac{P\left\{NB(r_2,p_0)=k\right\}}{P\left\{NB(r_1,p_0)=k\right\}} = p_0^{r_2-r_1}\prod_{i=1}^{k}\frac{r_2+k-i}{r_1+k-i}$$

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

Proof of Lemma 2.2 Assume (1.5), (2.5) 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. $\square$

Proof of Theorem 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 $m = EY_t = E\lambda_t$ for all t, then

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

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

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

For $k \geq 2$, we have

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

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

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

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

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

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,

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

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

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

Therefore the inequalities in (A.3) are shown by induction. Now note that $EY_t^{(k)} = E\lambda_t^{(k)}$ 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. $\square$

Proof of Theorem 3.2 Let the notation $m_s = EY_t^s$ 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 $EY_t^2 < \infty$ entails (A.5). To show that (A.5) is also sufficient, recall that it has been shown in the proof of Theorem 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

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

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 Theorem 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

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

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 $\ell = 2$. $\square$

Proof of Theorem 3.3 As in the proof of Theorem 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 $k > 0$,

$$\begin{aligned} Y_t^{(k)} &= F^-_{\lambda_t^{(k)}}(U_t), \\ \lambda_t^{(k)} &= g(Y_{t-1}^{(k-1)}, \ldots, Y_{t-q}^{(k-q)}, \lambda_{t-1}^{(k-1)}, \ldots, \lambda_{t-p}^{(k-p)}) + \pi(\boldsymbol{X}_{t-1}). \end{aligned} \tag{A.6}$$

By the argument of the proof of Theorem 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

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

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 (3.9),

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

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

Now assume (3.9) 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

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

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. $\square$

Lemma A.1 *Let $\{Y_t, t \in \mathbb{Z}\}$ be a strictly stationary and ergodic sequence satisfying* ***A1*** *and* ***A2****. Assume that Θ satisfies the compactness assumption* ***A6****. There exist a $\mathcal{F}_0$-measurable random variable $K > 0$ and a constant $\rho \in (0, 1)$ such that*

$$\sup_{\boldsymbol{\theta} \in \Theta} \left|\lambda_t(\boldsymbol{\theta}) - \widetilde{\lambda}_t(\boldsymbol{\theta})\right| < K\rho^t.$$

Proof of Lemma A.1 By (3.8), for $t \geq q+1$ we have

$$\delta_t := \left|\lambda_t(\boldsymbol{\theta}) - \widetilde{\lambda}_t(\boldsymbol{\theta})\right| \leq \sum_{j=1}^{p} \beta_j \left|\lambda_{t-j}(\boldsymbol{\theta}) - \widetilde{\lambda}_{t-j}(\boldsymbol{\theta})\right| \leq \beta \max_{j=1,\dots,p} \delta_{t-j},$$

where $\beta := \sup_{\boldsymbol{\theta}\in\Theta} \sum_{j=1}^{p} \beta_j < 1$ by **A2** and **A6**. Iterating the previous inequality, and setting $K_0 = \sup_{\boldsymbol{\theta}\in\Theta} \max_{j=1,\dots,p} \delta_{q+1-j}$, we obtain

$$\delta_{q+1} \leq K_0\beta, \qquad \delta_{q+2} \leq \beta \max\{\delta_{q+1}, K_0\} \leq K_0\beta, \qquad \delta_{q+j} \leq K_0\beta,\ j=1,\dots,p$$
$$\delta_{q+p+j} \leq K_0\beta^2,\ j=1,\dots,p, \qquad \delta_{q+kp+j} \leq K_0\beta^{k+1},\ j=1,\dots,p.$$

When $\beta = 0$, the result is obvious. When $\beta > 0$, the result holds with $K = K_0\beta^{-q/p}$ and $\rho = \beta^{1/p}$. $\square$

Lemma A.2 *Let $\{Y_t, t \in \mathbb{Z}\}$ be a strictly stationary and ergodic sequence satisfying **A1**, **A2** and **A4**, and assume **A6**. We have*

$$E \sup_{\boldsymbol{\theta}\in\Theta} \lambda_t(\boldsymbol{\theta}) < \infty.$$

Proof of Lemma A.2 Note that, by (3.8),

$$\lambda_t(\boldsymbol{\theta}) \leq c_t(\boldsymbol{\theta}) + \sum_{i=1}^{p} \beta_j \lambda_{t-j}(\boldsymbol{\theta}), \qquad c_t(\boldsymbol{\theta}) = g(\mathbf{0}^\top; \boldsymbol{\theta}) + \pi(\boldsymbol{X}_{t-1}) + \sum_{i=1}^{q} \alpha_i Y_{t-i}.$$

Let $\boldsymbol{\lambda}_t(\boldsymbol{\theta}) = (\lambda_t(\boldsymbol{\theta}), \dots, \lambda_{t-p+1}(\boldsymbol{\theta}))^\top$, $\boldsymbol{c}_t(\boldsymbol{\theta}) = \left(c_t(\boldsymbol{\theta}), \mathbf{0}^\top\right)^\top$ and $\boldsymbol{B}$ a companion-like matrix such that the previous inequality yields $\boldsymbol{\lambda}_t(\boldsymbol{\theta}) \leq \boldsymbol{c}_t(\boldsymbol{\theta}) + \boldsymbol{B}\boldsymbol{\lambda}_{t-1}(\boldsymbol{\theta})$. Letting $\boldsymbol{\lambda}_t = \sup_{\boldsymbol{\theta}\in\Theta} \boldsymbol{\lambda}_t(\boldsymbol{\theta})$ and $\boldsymbol{c}_t = \sup_{\boldsymbol{\theta}\in\Theta} \boldsymbol{c}_t(\boldsymbol{\theta})$ componentwise, we obtain

$$\|\boldsymbol{\lambda}_t\| \leq \|\boldsymbol{c}_t\| \sum_{i=0}^{\infty} \sup_{\boldsymbol{\theta}\in\Theta} \|\boldsymbol{B}\|^i < \infty$$

because **A2** and **A6** entail $\sup_{\boldsymbol{\theta}\in\Theta} \rho(\boldsymbol{B}) < 1$ (see *e.g.* (7.27) in Francq and Zakoian, 2019). The conclusion follows. $\square$

Proof of Theorem 4.1 Set $\widetilde{L}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=q+1}^{n} \widetilde{l}_t(\boldsymbol{\theta})$ and $L_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=q+1}^{n} l_t(\boldsymbol{\theta})$. Using the inequality $\log(x) \le x-1$, **A3** and Lemma A.1, it follows that

$$\begin{aligned}
\sup_{\boldsymbol{\theta}\in\Theta}\left|L_n(\boldsymbol{\theta})-\widetilde{L}_n(\boldsymbol{\theta})\right| &= \frac{1}{n}\sup_{\boldsymbol{\theta}\in\Theta}\left|\sum_{t=1}^{n}\left(Y_t\left(\frac{1}{\lambda_t(\boldsymbol{\theta})}-\frac{1}{\widetilde{\lambda}_t(\boldsymbol{\theta})}\right)+\log\left(\frac{\lambda_t(\boldsymbol{\theta})}{\widetilde{\lambda}_t(\boldsymbol{\theta})}\right)\right)\right| \\
&\le \frac{1}{n}\sum_{t=1}^{n}\left(\frac{Y_t\sup_{\boldsymbol{\theta}\in\Theta}\left|\lambda_t(\boldsymbol{\theta})-\widetilde{\lambda}_t(\boldsymbol{\theta})\right|}{\lambda_t(\boldsymbol{\theta})\widetilde{\lambda}_t(\boldsymbol{\theta})}+\frac{\sup_{\boldsymbol{\theta}\in\Theta}\left|\widetilde{\lambda}_t(\boldsymbol{\theta})-\lambda_t(\boldsymbol{\theta})\right|}{\widetilde{\lambda}_t(\boldsymbol{\theta})}\right) \\
&\le \frac{K}{n}\sum_{t=1}^{n}\left(\frac{Y_t\rho^t}{\underline{\omega}^2}+\frac{\rho^t}{\underline{\omega}}\right)\to 0,\ a.s.\text{ as } n\to\infty. \qquad (A.7)
\end{aligned}$$

By **A3**, **A4** and Lemma A.2, $|\log\lambda_t(\boldsymbol{\theta})|$ admits moments of any order, and we have

$$E\left|l_1(\boldsymbol{\theta})\right| \le \frac{E\left|Y_1\right|}{\underline{\omega}}+E\left|\log(\lambda_1(\boldsymbol{\theta}))\right|<\infty.$$

Moreover, using again the inequality $\log(x)\le x-1$, we have

$$\begin{aligned}
E\left(l_1(\boldsymbol{\theta}_0)-l_1(\boldsymbol{\theta})\right) &= E\left(Y_1\frac{\lambda_1(\boldsymbol{\theta})-\lambda_1(\boldsymbol{\theta}_0)}{\lambda_1(\boldsymbol{\theta})\lambda_1(\boldsymbol{\theta}_0)}+\log\frac{\lambda_1(\boldsymbol{\theta}_0)}{\lambda_1(\boldsymbol{\theta})}\right) \\
&\le EE\left(Y_1\frac{\lambda_1(\boldsymbol{\theta})-\lambda_1(\boldsymbol{\theta}_0)}{\lambda_1(\boldsymbol{\theta})\lambda_1(\boldsymbol{\theta}_0)}\middle|\mathcal{F}_{t-1}\right)+E\left(\frac{\lambda_1(\boldsymbol{\theta}_0)-\lambda_1(\boldsymbol{\theta})}{\lambda_1(\boldsymbol{\theta})}\right) \\
&= E\left(\frac{\lambda_1(\boldsymbol{\theta})-\lambda_1(\boldsymbol{\theta}_0)}{\lambda_1(\boldsymbol{\theta})}\right)+E\left(\frac{\lambda_1(\boldsymbol{\theta}_0)-\lambda_1(\boldsymbol{\theta})}{\lambda_1(\boldsymbol{\theta})}\right)=0,
\end{aligned}$$

with equality iff $\lambda_t(\boldsymbol{\theta})=\lambda_t(\boldsymbol{\theta}_0)$, that is, by **A5**, iff $\boldsymbol{\theta}=\boldsymbol{\theta}_0$. It follows that

$$E\left(l_1(\boldsymbol{\theta}_0)\right)<E\left(l_1(\boldsymbol{\theta})\right),\quad \forall\boldsymbol{\theta}\ne\boldsymbol{\theta}_0. \qquad (A.8)$$

Let $V_k(\boldsymbol{\theta}_1)$ ($\boldsymbol{\theta}_1\in\Theta$ and $k\in\mathbb{N}^*$) be the open ball with center $\boldsymbol{\theta}_1$ and radius $1/k$. Since $\sup_{\boldsymbol{\theta}\in V_k(\boldsymbol{\theta}_1)\cap\Theta} l_t(\boldsymbol{\theta})$ is a measurable function of the terms of $\{Y_t, t\in\mathbb{Z}\}$, the process $\left\{\sup_{\boldsymbol{\theta}\in V_k(\boldsymbol{\theta}_1)\cap\Theta} l_t(\boldsymbol{\theta}),\ t\in\mathbb{Z}\right\}$ is strictly stationary and ergodic with $E\left|\sup_{\boldsymbol{\theta}\in V_k(\boldsymbol{\theta}_1)\cap\Theta} l_t(\boldsymbol{\theta})\right|<\infty$ by Lemma A.2. The ergodic theorem and (A.7) thus entail

$$\limsup_{n\to\infty}\sup_{\boldsymbol{\theta}\in V_k(\boldsymbol{\theta}_1)\cap\Theta}\widetilde{L}_n(\boldsymbol{\theta})=\limsup_{n\to\infty}\sup_{\boldsymbol{\theta}\in V_k(\boldsymbol{\theta}_1)\cap\Theta}L_n(\boldsymbol{\theta})\ge E\left(\sup_{\boldsymbol{\theta}\in V_k(\boldsymbol{\theta}_1)\cap\Theta}l_1(\boldsymbol{\theta})\right).$$

By the Beppo-Levi theorem, $E\left(\sup_{\boldsymbol{\theta}\in V_k(\boldsymbol{\theta}_1)\cap\Theta} l_1(\boldsymbol{\theta})\right)$ decreases to $E\left(l_1(\boldsymbol{\theta}_1)\right)$ as $k\to\infty$. Thus, in view of (A.8), we have shown that for all $\boldsymbol{\theta}_1\ne\boldsymbol{\theta}_0$, there exists a neighborhood $V(\boldsymbol{\theta}_1)$ such that

$$\limsup_{n\to\infty}\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_1)\cap\Theta}\widetilde{L}_n(\boldsymbol{\theta})>\limsup_{n\to\infty}\widetilde{L}_n(\boldsymbol{\theta}_0)=\limsup_{n\to\infty}L_n(\boldsymbol{\theta}_0)=E\left(l_1(\boldsymbol{\theta}_0)\right).$$

By standard arguments the proof of Theorem 4.1 is completed, using compactness of Θ. $\square$

Lemma A.3 *Under the assumptions of Theorem 4.1 and* **A7** *we have*

$$E \sup_{\boldsymbol{\theta}\in\Theta} \left\| \frac{\partial \boldsymbol{\lambda}_t^\top(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|^\varepsilon < \infty \tag{A.9}$$

for some $\varepsilon > 0$*, and*

$$\sup_{\boldsymbol{\theta}\in\Theta} \left\| \frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial \widetilde{\lambda}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| < K v_t \rho^t,$$

where K *and* ρ *are as in Lemma A.1 and* $\sup_t E v_t^\varepsilon < \infty$ *for some* $\varepsilon > 0$*.*

Proof of Lemma A.3 Let $k \in \{1, \dots, d\}$ and $\boldsymbol{e}_k$ the k-th column of I_d. With the notation of the proof of Lemma A.2, we have

$$\frac{\partial}{\partial \theta_k} \boldsymbol{\lambda}_t(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{e}_k^\top \boldsymbol{D}_{\boldsymbol{\theta}} + \frac{\partial}{\theta_k} \pi(\boldsymbol{X}_{t-1}; \boldsymbol{\theta}) \\ \boldsymbol{0}_{p-1} \end{pmatrix} + \boldsymbol{A} \frac{\partial}{\partial \theta_k} \boldsymbol{\lambda}_{t-1}(\boldsymbol{\theta}).$$

Thus (A.9) follows by **A7**. Now, by (4.4), note that

$$\begin{aligned}
\frac{\partial \lambda_t}{\partial \theta_k} - \frac{\partial \widetilde{\lambda}_t}{\partial \theta_k} =& \boldsymbol{e}_k^\top \boldsymbol{D}_{\boldsymbol{\theta}} \left(Y_{t-1:q}, \lambda_{t-1:p}; \boldsymbol{\theta}\right) - \boldsymbol{e}_k^\top \boldsymbol{D}_{\boldsymbol{\theta}} \left(Y_{t-1:q}, \widetilde{\lambda}_{t-1:p}; \boldsymbol{\theta}\right) \\
&+ \frac{\partial \boldsymbol{\lambda}_{t-1}^\top(\boldsymbol{\theta})}{\partial \theta_k} \left\{ \boldsymbol{D}_{\boldsymbol{\lambda}} \left(Y_{t-1:q}, \lambda_{t-1:p}; \boldsymbol{\theta}\right) - \boldsymbol{D}_{\boldsymbol{\lambda}} \left(Y_{t-1:q}, \widetilde{\lambda}_{t-1:p}; \boldsymbol{\theta}\right) \right\} \\
&+ \left(\frac{\partial \boldsymbol{\lambda}_{t-1}^\top(\boldsymbol{\theta})}{\partial \theta_k} - \frac{\partial \widetilde{\boldsymbol{\lambda}}_{t-1}^\top(\boldsymbol{\theta})}{\partial \theta_k} \right) \boldsymbol{D}_{\boldsymbol{\lambda}} \left(Y_{t-1:q}, \widetilde{\lambda}_{t-1:p}; \boldsymbol{\theta}\right).
\end{aligned} \tag{A.10}$$

In matrix form

$$\frac{\partial \boldsymbol{\lambda}_t}{\partial \theta_k} - \frac{\partial \widetilde{\boldsymbol{\lambda}}_t}{\partial \theta_k} = \begin{pmatrix} d_t \\ \boldsymbol{0} \end{pmatrix} + \boldsymbol{A} \left\{ \frac{\partial \boldsymbol{\lambda}_{t-1}}{\partial \theta_k} - \frac{\partial \widetilde{\boldsymbol{\lambda}}_{t-1}}{\partial \theta_k} \right\}$$

where d_t is the sum of the first two terms of the right-hand side of (A.10). By the mean value theorem, **A7**, (A.9) and Lemma A.1, we have $|d_t| \le w_t \rho_1^t$ where $E|w_t|^\varepsilon < \infty$ for some $\varepsilon > 0$ and $\rho_1 < 1$. We thus have

$$\left\| \frac{\partial \boldsymbol{\lambda}_t}{\partial \theta_k} - \frac{\partial \widetilde{\boldsymbol{\lambda}}_t}{\partial \theta_k} \right\| \le K \rho_2^t (w_t + w_{t-1} + \cdots + w_1) + K \rho_2^t d_0$$

for some $K > 0$ and $\max\{\rho_1, \rho(\boldsymbol{A})\} < \rho_2 < 1$. The conclusion follows by taking, for instance, $\rho = \rho_2^{1/2}$ and $v_t = \rho_2^{t/2}(w_t + \cdots + w_1 + d_0)$. $\square$

Proof of Theorem 4.2 Since by **A8** and Theorem 4.1, $\widehat{\boldsymbol{\theta}}$ cannot be at the boundary of Θ for n sufficiently large, a Taylor expansion of $\frac{\partial L_n(\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}$ at $\boldsymbol{\theta}_0$ yields

$$\sqrt{n}\frac{\partial L_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + \sqrt{n}\frac{\partial^2 L_n(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) + \sqrt{n}\left(\frac{\partial \widetilde{L}_n(\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} - \frac{\partial L_n(\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\right) = 0 \tag{A.11}$$

for some $\boldsymbol{\theta}^*$ between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$.

We first show that the third term in the left hand side of (A.11) is $a.s.$ negligeable. By **A3**, Lemma A.1 and Lemma A.3 it follows that $a.s.$

$$\sqrt{n}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial \widetilde{L}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial L_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\| \leq \frac{K}{\sqrt{n}}\sum_{t=1}^{n}(1+Y_t)\left\{\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial \lambda_t}{\partial \boldsymbol{\theta}}\right\|\rho^t + \sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial \lambda_t}{\partial \boldsymbol{\theta}} - \frac{\partial \widetilde{\lambda}_t}{\partial \boldsymbol{\theta}}\right\|\right\} = o(1)\,. \tag{A.12}$$

For the last equality, we used the fact that

$$E\left(\sum_{t=1}^{\infty}(1+Y_t)\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial \lambda_t}{\partial \boldsymbol{\theta}} - \frac{\partial \widetilde{\lambda}_t}{\partial \boldsymbol{\theta}}\right\|\right)^{\varepsilon/2} \leq \sum_{t=1}^{\infty}\sqrt{E\left(1+Y_t\right)^{\varepsilon}}\sqrt{E\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial \lambda_t}{\partial \boldsymbol{\theta}} - \frac{\partial \widetilde{\lambda}_t}{\partial \boldsymbol{\theta}}\right\|^{\varepsilon}} < \infty$$

for $\epsilon \in (0,1]$ satisfying Lemma A.3.

Now, it is easy to check that $\left\{\sqrt{n}\frac{\partial L_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}, t \in \mathbb{Z}\right\}$ is a martingale with respect to $\{\mathcal{F}_t, t \in \mathbb{Z}\}$ where

$$\sqrt{n}\frac{\partial L_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \quad \text{and} \quad \frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = \frac{\lambda_t(\boldsymbol{\theta}_0) - Y_t}{\lambda_t^2(\boldsymbol{\theta}_0)}\frac{\partial \lambda_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}.$$

By **A9** and **A10** we get

$$E\left(\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right) = E\left(\frac{v_t(\boldsymbol{\theta}_0)}{\lambda_t^4(\boldsymbol{\theta}_0)}\frac{\partial \lambda_t(\boldsymbol{\theta}_0)\partial \lambda_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right) = \boldsymbol{I}.$$

From the martingale central limit theorem (e.g. Billingsley, (2008), Hall and Heyde, (1980)), it follows that

$$\sqrt{n}\frac{\partial L_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}\left(\mathbf{0}, \boldsymbol{I}\right). \tag{A.13}$$

We finally show the convergence of the second term in the left-hand side of (A.11). Let $V_k(\boldsymbol{\theta}_0)$ $(k \in \mathbb{N}^*)$ be the open ball with center $\boldsymbol{\theta}_0$ and radius $1/k$, where k is supposed large enough so that $V_k(\boldsymbol{\theta}_0)$ is contained in $V(\boldsymbol{\theta}_0)$ defined by **A11**. Assume that n is large enough so that $\boldsymbol{\theta}^*$ belongs to $V_k(\boldsymbol{\theta}_0)$. By stationarity and ergodicity of $\left\{\frac{\partial^2 l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j}\right\}_t$ and $\left\{\sup_{\boldsymbol{\theta}\in V_k(\boldsymbol{\theta}_0)}\left|\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j}-\frac{\partial^2 l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j}\right|\right\}_t$, it follows that

$$\begin{aligned}\left|\frac{\partial^2 L_n(\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j}-\boldsymbol{J}(i,j)\right| &\leq \frac{1}{n}\sum_{t=1}^{n}\sup_{\boldsymbol{\theta}\in V_k(\boldsymbol{\theta}_0)}\left|\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j}-\frac{\partial^2 l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j}\right|+\left|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j}-E\left(\frac{\partial^2 l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j}\right)\right| \\ &\to E\left(\sup_{\boldsymbol{\theta}\in V_k(\boldsymbol{\theta}_0)}\left|\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j}-E\left(\frac{\partial^2 l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j}\right)\right|\right)\end{aligned}$$

a.s. as $n\to\infty$. The Lebesgue dominated convergence theorem and **A10** then yield

$$\lim_{k\to\infty}E\left(\sup_{\boldsymbol{\theta}\in V_k(\boldsymbol{\theta}_0)}\left|\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j}-\frac{\partial^2 l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j}\right|\right)=E\left(\lim_{k\to\infty}\sup_{\boldsymbol{\theta}\in V_k(\boldsymbol{\theta}_0)}\left|\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j}-\frac{\partial^2 l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j}\right|\right)=0. \qquad (A.14)$$

The conclusion then follows from (A.11), (A.12), **A10**, (A.13) and (A.14). $\square$

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] Aknouche, A., Bendjeddou, S. and Touche, N. (2018) Negative Binomial Quasi-Likelihood Inference for General Integer-Valued Time Series Models. *Journal of Time Series Analysis* 39, 192–211.

[4] 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.

[5] Billingsley P. (2008) *Probability and measure.* John Wiley & Sons, Third Edition.

[6] Bougerol, P. (1993) Kalman filtering with random coefficients and contractions. *SIAM Journal on Control and Optimization* 31, 942–959.

[7] 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.

[8] 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.

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

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

[11] 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.

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

[13] 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.

[14] 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.

[15] 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.

[16] Doukhan, P. and Neumann, M.H. (2019) Absolute regularity of semi-contractive GARCH-type processes. *Journal of Applied Probability* 56, 91-115.

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

[18] 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.

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

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

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

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

[23] Francq, C., and Thieu, L. (2019) Qml inference for volatility models with covariates. *Econometric Theory* 35, 37–72.

[24] 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.

[25] Francq, C. and Zakoian, J.-M. (2019) *GARCH models: structure, statistical inference and financial applications.* John Wiley & Sons, Second Edition.

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

[27] Gouriéroux, C., Monfort, A. and Trognon, A. (1984) Pseudo maximum likelihood methods: Theory. *Econometrica* 52, 681–700.

[28] 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.

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

[30] Hall P, Heyde CC. (1980) *Martingale Limit Theory and its Applications.* Academic Press, New York.

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

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

[33] 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.

[34] 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.

[35] Meyn, S.P. and Tweedie, R.L. (2009) *Markov chains and stochastic stability.* Springer Science & Business Media, Second Edition.

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

[37] 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.

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

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

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

[41] Straumann, D. and Mikosch, T. (2006) Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. *The Annals of Statistics* 34, 2449–2495.

[42] 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.

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

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

[45] Wedderburn, R.W. (1974) Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* 61, 439–447.

[46] Wooldridge, J.M. (1999) *Quasi-Likelihood Methods for Count Data.* In M.H. Pesaran and P. Schmidt (ed.), *Handbook of Applied Econometrics, Volume 2: Microeconomics, (pp. 352—406).* Oxford: Blackwell.

[47] Zhu, F. (2011) A negative binomial integer-valued GARCH model. *Journal of Time Series Analysis* 32, 54–67.

[48] Zhu, F. (2012) Zero-inflated Poisson and negative binomial integer-valued GARCH models. *Journal of Statistical Planning and Inference* 142, 826–839.

Supplement to "Count and duration time series with equal conditional stochastic and mean orders"

Finite-sample properties of the EQMLE

Finite-sample behavior of EQMLE are examined through a simulation study. We consider three models satisfying the stochastic-equal-mean order property (cf. (1.5) in Aknouche and Francq, 2019, henceforth AF), namely the exponential conditional distribution with mean λ_t ($Y_t/\mathcal{F}_{t-1} \sim \Gamma(1, 1/\lambda_t)$), the quadratic Gamma distribution, $\Gamma(0.5, 0.5/\lambda_t)$, and the linear Gamma distribution $\Gamma(\lambda_t/2, 1/2)$. For each model, we generate $N = 1000$ replications with sample-sizes $n = 500$, $n = 1000$ and $n = 3000$. The conditional mean is generated from a linear POLI model (cf. AF, (1.1)) with $p = q = 1$ and true parameter $\theta_0 = (1, 0.6, 0.2)^\top$. EQMLE and PQMLE are computed for each model. Mean of EQML and PQML estimates over the 1000 replications are reported in bold, in Table 1 for model $\Gamma(1, 1/\lambda_t)$, in Table 2 for model $\Gamma(0.5, 0.5/\lambda_t)$, and in Table 3 for model $\Gamma(0.5\lambda_t, 0.5)$. These tables also show four estimates of the mean square error $\mathrm{E}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)^2$ (see also Ahmad and Francq, 2016). These estimates are i) the estimated standard error (ESE) given by $\mathrm{ESE}(\boldsymbol{\theta}_{0j}) = \frac{1}{N}\sum_{i=1}^{N}\left(\widehat{\boldsymbol{\theta}}_j^{(i)} - \boldsymbol{\theta}_{0j}\right)^2$ ($\widehat{\boldsymbol{\theta}}_j^{(i)}$ being the estimate of $\boldsymbol{\theta}_{0j}$ at the ith replication, $j = 1, 2, 3$), ii) the asymptotic standard error (ASE) defined by $\mathrm{ASE}(\boldsymbol{\theta}_{0j}) = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\frac{1}{n}\left(\widehat{\boldsymbol{\Sigma}}^{(i)}\right)^{-1}(j,j)}$, iii) the theoretical standard error (TSE) given by $\mathrm{TSE}(\boldsymbol{\theta}_{0j}) = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\frac{1}{n}\left(\boldsymbol{\Sigma}^{(i)}\right)^{-1}(j,j)}$ where $\boldsymbol{\Sigma}$ is computed from a very large series ($n = 20000$), and finally iv) the eXponential standard error (XSE) computed similarly to ASE while replacing $\widehat{\boldsymbol{\Sigma}}^{(i)}$ by $\widehat{\boldsymbol{J}}^{(i)}$. The same measures are considered for PQMLE but are rather based on the asymptotic results given by (4.6) in AF. In particular, XSE is replaced by the Poisson standard error (PSE) computed from (4.7) in AF with $b = 1$ (see, Ahmad and Francq 2016).

		$\Gamma(1, 1/\lambda_t)$						
		ω_0	α_0	β_0		ω_0	α_0	β_0
n	θ_0	1	0.6	0.2		1	0.6	0.2
500	**EQMLE**	**1.1286**	**0.5918**	**0.1743**	PQMLE	**1.1450**	**0.5698**	**0.1867**
	ESE	0.3059	0.0728	0.0772	ESE	0.3034	0.0923	0.0913
	ASE	0.1945	0.0699	0.0675	ASE	0.2500	0.0932	0.0828
	TSE	0.1819	0.0706	0.0672	TSE	0.2752	0.1076	0.0996
	XSE	0.1955	0.0703	0.0663	PSE	0.1072	0.0251	0.0310
1000	EQMLE	**1.0641**	**0.5971**	**0.1860**	PQMLE	**1.0841**	**0.5827**	**0.1899**
	ESE	0.1694	0.0495	0.0494	ESE	0.1961	0.0707	0.0658
	ASE	0.1318	0.0496	0.0464	ASE	0.1837	0.0727	0.0618
	TSE	0.1286	0.0499	0.0475	TSE	0.1946	0.0761	0.0704
	XSE	0.1332	0.0498	0.0463	PSE	0.0723	0.0177	0.0210
3000	EQMLE	**1.0247**	**0.6001**	**0.1945**	PQMLE	**1.0368**	**0.5916**	**0.1960**
	ESE	0.0857	0.0305	0.0285	ESE	0.1215	0.0430	0.0425
	ASE	0.0750	0.0287	0.0266	ASE	0.1153	0.0462	0.0389
	TSE	0.0743	0.0288	0.0274	TSE	0.1124	0.0439	0.0407
	XSE	0.0749	0.0288	0.0265	PSE	0.0400	0.0101	0.0116

Table 1. Estimation results for EQMLE and PQMLE for model $\Gamma(1, 1/\lambda_t)$.

		$\Gamma(0.5, 0.5/\lambda_t)$						
		ω_0	α_0	β_0		ω_0	α_0	β_0
n	θ_0	1	0.6	0.2		1	0.6	0.2
500	EQMLE	**1.0710**	**0.5968**	**0.1836**	PQMLE	**1.2129**	**0.5960**	**0.1597**
	ESE	0.2457	0.0943	0.0789	ESE	0.3268	0.0535	0.0826
	ASE	0.1960	0.0950	0.0723	ASE	0.2236	0.0539	0.0691
	TSE	0.1954	0.0946	0.0738	TSE	0.2047	0.0553	0.0670
	XSE	0.1412	0.0681	0.0518	PSE	0.2157	0.0467	0.0648
1000	EQMLE	**1.0409**	**0.5970**	**0.1890**	PQMLE	**1.0979**	**0.5973**	**0.1827**
	ESE	0.1596	0.0704	0.0552	ESE	0.1882	0.0389	0.0517
	ASE	0.1365	0.0678	0.0514	ASE	0.1524	0.0388	0.0481
	TSE	0.1382	0.0669	0.0522	TSE	0.1447	0.0391	0.0474
	XSE	0.0980	0.0482	0.0367	PSE	0.1437	0.0328	0.0444
3000	EQMLE	**1.0182**	**0.6006**	**0.1950**	PQMLE	**1.0326**	**0.6004**	**0.1930**
	ESE	0.0852	0.0383	0.0306	ESE	0.0940	0.0231	0.0285
	ASE	0.0783	0.0394	0.0296	ASE	0.0865	0.0227	0.0276
	TSE	0.0798	0.0386	0.0301	TSE	0.0836	0.0226	0.0274
	XSE	0.0557	0.0279	0.0211	PSE	0.0799	0.0189	0.0251

Table 2. Estimation results for EQMLE and PQMLE for model $\Gamma(0.5, 0.5/\lambda_t)$.

		$\Gamma\left(0.5\lambda_t, 0.5\right)$						
		ω_0	α_0	β_0		ω_0	α_0	β_0
n	θ_0	1	0.6	0.2		1	0.6	0.2
500	EQMLE	**1.2100**	**0.6158**	**0.1395**	PQMLE	**1.1290**	**0.6018**	**0.1719**
	ESE	0.3543	0.0603	0.1042	ESE	0.2601	0.0529	0.0783
	ASE	0.2324	0.0582	0.0775	ASE	0.2049	0.0504	0.0684
	TSE	0.2093	0.0597	0.0759	TSE	0.1909	0.0516	0.0677
	XSE	0.2796	0.0840	0.1038	PSE	0.1474	0.0362	0.0490
1000	EQMLE	**0.1286**	**0.6112**	**0.1620**	PQMLE	**1.0565**	**0.6019**	**0.1868**
	ESE	0.2265	0.0426	0.0680	ESE	0.1547	0.0367	0.0500
	ASE	0.1615	0.0414	0.0548	ASE	0.1401	0.0358	0.0477
	TSE	0.1480	0.0422	0.0537	TSE	0.1350	0.0365	0.0479
	XSE	0.1891	0.0594	0.0728	PSE	0.0999	0.0255	0.0340
3000	EQMLE	**1.0433**	**0.6040**	**0.1856**	PQMLE	**1.0231**	**0.5992**	**0.1964**
	ESE	0.1044	0.0241	0.0351	ESE	0.0852	0.0208	0.0279
	ASE	0.0905	0.0240	0.0312	ASE	0.0800	0.0207	0.0274
	TSE	0.0855	0.0244	0.0310	TSE	0.0779	0.0211	0.0277
	XSE	0.1033	0.0341	0.0414	PSE	0.0567	0.0147	0.0195

Table 3. Estimation results for EQMLE and PQMLE for model $\Gamma\left(0.5\lambda_t, 0.5\right)$.

From the latter simulations some broad conclusions may be drawn. Firstly, the parameters are well estimated by the two methods regarding their small bias and their various estimated standard errors. The latter are quite close to each other implying a well reliability of the estimates. Secondly, the estimation results are consistent with asymptotic theory as their accuracies increase with the sample size. Thirdly, as expected, the EQMLE gives better results under the conditional exponential distribution but is less accurate than the PQMLE if we depart from the exponential distribution. Note finally that EQMLE largely outperforms PQMLE under the conditional exponential model but its superiority is less pronounced in the Gamma $\Gamma(0.5, 0.5/\lambda_t)$ case. However, the PQMLE dominates EQMLE for the Gamma $\Gamma(0.5\lambda_t, 0.5)$ model with linear conditional variance, which is in accordance with Remark 4.3 in AF. The estimation methods were implemented in Matlab on a desktop with Intel Core i7. The optimization routines were developed using the fminunc function of Matlab.

References

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

[2] Aknouche, A. and Francq, C. (2019). Count and duration time series with equal conditional stochastic and mean orders. Submitted preprint.