



Munich Personal RePEc Archive

Poisson qmle of count time series models

Ahmad, Ali and Francq, Christian

University of Lille, Crest and University of Lille

November 2014

Online at <https://mpra.ub.uni-muenchen.de/59804/>

MPRA Paper No. 59804, posted 04 Jan 2015 21:12 UTC

Poisson qmle of count time series models*

ALI AHMAD

Université Lille 3 (EQUIPPE)

BP 60 149, 59653 Villeneuve d'Ascq cedex, France.

E-mail: ali.ahmad@etu.univ-lille3.fr

and

CHRISTIAN FRANCO

CREST and Université Lille 3 (EQUIPPE)

E-mail: christian.francq@univ-lille3.fr

Abstract: Regularity conditions are given for the consistency of the Poisson quasi-maximum likelihood estimator of the conditional mean parameter of a count time series. The asymptotic distribution of the estimator is studied when the parameter belongs to the interior of the parameter space and when it lies at the boundary. Tests for the significance of the parameters and for constant conditional mean are deduced. Applications to specific INAR and INGARCH models are considered. Numerical illustrations, on Monte Carlo simulations and real data series, are provided.

Keywords: Boundary of the parameter space, Consistency and asymptotic normality, Integer-valued AR and GARCH models, Non-normal asymptotic distribution, Poisson quasi-maximum likelihood estimator, Time series of counts.

*The authors gratefully acknowledge financial support of the Agence Nationale de la Recherche (ANR) through the ORA Program (Project PRAM ANR-10-ORAR-008-01).

1 Introduction

The literature on time series of counts is becoming increasingly abundant, with applications in numerous domains (see *e.g.* the monographs by [Christou \(2013\)](#) and [Liu \(2012\)](#), and the references therein). It is common to assume a conditional Poisson distribution with the intensity parameter depending on the past values. This leads to models that are quite tractable¹, but extremely constrained, since their conditional variance and conditional mean coincide. Many extensions and alternative conditional distributions have been proposed, but either the conditional distribution remains relatively constrained or it contains extra parameters that are difficult to estimate and interpret.

In the present paper we adopt a semi-parametric approach, in which only the conditional mean is specified. Since the works of [Wedderburn \(1974\)](#), [White \(1982\)](#), [McCullagh \(1983\)](#) and [Gourieroux et al. \(1984\)](#), it is known that certain maximum likelihood estimators (MLEs) can be consistent and asymptotically normal (CAN) for the parameters of the conditional mean and variance, even if the actual conditional distribution is not that assumed by the MLE. In particular, the Gaussian quasi-maximum likelihood estimator (QMLE), in which the conditional mean and variance parameters are estimated by maximizing a pseudo-likelihood written as if the condition mean were Gaussian, is the method of choice for estimating ARMA-GARCH type models. For time series of counts, the Poisson QMLE (PQMLE) can be employed to identify the conditional mean.

In this paper, we give general regularity conditions under which the PQMLE is CAN. We also consider the case where the above-mentioned regularity conditions are violated because the parameter stands at the boundary of the parameter space. In that case the asymptotic distribution is not Gaussian. This situation must be considered for testing the nullity of some conditional mean parameters. For important classes of time series of counts, such as the INGARCH models, the significance test statistics are not asymptotically distributed as a standard chi-square, but as chi-bar-square. The general results are applied to specific models, namely the integer-valued autoregressive (INAR) and the integer-valued GARCH (INGARCH) and the log-linear models, with different specifications of the conditional distribution.

Thus, the main contribution of the present paper is threefold. Firstly, the asymptotic theory of the PQMLE is developed. To our knowledge, this is the first time a QMLE is studied for

¹even if the probabilistic structure, in particular the ergodicity, of these models is not easy to derive (see [Tjøstheim \(2012, 2014\)](#) and the references therein)

general count time series models. This allows for obtaining optimal predictions without having to specify entirely the conditional distribution. Second, the asymptotic distribution of the estimator is obtained without positivity constraint on the coefficients, which is also new for count time series. Third, Wald-type significance tests are proposed. Due to boundary effects, the asymptotic distribution of these tests is not standard, but they can however be easily implemented and are obviously useful to model identification. These theoretical results are illustrated by Monte Carlo simulations and applications on financial data.

The paper is organized as follows. Section 2 contains the main results concerning the asymptotic behavior of the Poisson QMLE and of the related significance tests. Section 3 applies the general results to particular observation-driven and parameter-driven models (according to the nomenclature introduced by Cox et al. (1981)). Section 4 studies the finite sample properties of the PQMLE and of the significance tests, via a set of Monte Carlo experiments. In Section 5, we use the PQMLE to fit INGARCH(p, q) models on daily series of the number of transactions of stocks. Section 6 concludes, and the proofs are collected in Section 7.

2 Asymptotic distribution of the Poisson QMLE

Assume that we have observations X_1, \dots, X_n of a times series valued in \mathbb{N} , such that

$$E(X_t | X_u, u < t) = \lambda(X_{t-1}, X_{t-2}, \dots; \theta_0), \quad (2.1)$$

where

$$\lambda \text{ is a measurable function valued in } (\underline{\omega}, +\infty) \text{ for some } \underline{\omega} > 0 \quad (2.2)$$

and θ_0 is an unknown parameter belonging to some parameter space $\Theta \subset \mathbb{R}^d$. The marginal distribution is assumed to have a moment slightly greater than 1

$$EX_t^{1+\varepsilon} < \infty, \text{ for some } \varepsilon > 0, \quad (2.3)$$

which entails the existence of the conditional mean (2.1). For all $\theta \in \Theta$, $x_0 \in \mathbb{N}$ and $t \geq 1$, let

$$\lambda_t(\theta) = \lambda(X_{t-1}, X_{t-2}, \dots; \theta) \text{ and } \tilde{\lambda}_t(\theta) = \lambda(X_{t-1}, X_{t-2}, \dots, X_1, x_0, x_0, \dots; \theta).$$

Note that $\tilde{\lambda}_t(\theta)$ will serve as a proxy for $\lambda_t(\theta)$. It is obtained by setting to some integer x_0 the unknown initial values X_0, X_{-1}, \dots involved in $\lambda_t(\theta)$. This value x_0 can either be a fixed integer, for instance $x_0 = 0$, or a value depending on θ , or a value depending on the observations. For example, when $\lambda_t(\theta) = \omega + \alpha X_{t-1} + \beta \lambda_{t-1}(\theta)$ with $\theta = (\omega, \alpha, \beta)$, one can take $\tilde{\lambda}_t(\theta) = \omega + \alpha X_{t-1} +$

$\beta\tilde{\lambda}_{t-1}(\theta)$ with $\tilde{\lambda}_1(\theta) = \omega/(1 - \beta)$ (which corresponds to $x_0 = 0$), or with $\tilde{\lambda}_1(\theta) = \omega/(1 - \alpha - \beta)$ (which corresponds to $x_0 = \omega/(1 - \alpha - \beta)$), or with $\tilde{\lambda}_1(\theta) = \bar{X}_5$ (the average of the working days of the first week, for daily data). It will be shown that the choice of x_0 is asymptotically unimportant, provided we have a.s.

$$\lim_{t \rightarrow \infty} a_t = 0 \quad \text{and} \quad \lim_{t \rightarrow \infty} X_t a_t = 0, \quad \text{where } a_t = \sup_{\theta \in \Theta} \left| \tilde{\lambda}_t(\theta) - \lambda_t(\theta) \right|, \quad (2.4)$$

and

$$\tilde{\lambda}_t(\theta) \geq \underline{\omega}, \quad \forall t \geq 1, \quad \forall \theta \in \Theta. \quad (2.5)$$

Assuming that

$$\theta \mapsto \lambda_t(\theta) \text{ is almost surely continuous and } \Theta \text{ is a compact set}, \quad (2.6)$$

a Poisson Quasi Maximum Likelihood Estimator of θ_0 is defined as any measurable solution of

$$\hat{\theta}_n = \arg \max_{\theta \in \Theta} \tilde{L}_n(\theta), \quad \tilde{L}_n(\theta) = \frac{1}{n} \sum_{t=s+1}^n \tilde{\ell}_t(\theta), \quad (2.7)$$

where $\tilde{\ell}_t(\theta) = -\tilde{\lambda}_t(\theta) + X_t \log \tilde{\lambda}_t(\theta)$ and s is an integer. The value of s is asymptotically unimportant, but it can affect the finite sample behavior of the PQMLE by reducing the impact of the initial value x_0 . Note that $\hat{\theta}_n$ is equal to the maximum likelihood estimator (MLE) of θ_0 if the conditional distribution of X_t is Poisson with parameter $\lambda_t(\theta_0)$. Since we do not make any specific assumption on the conditional distribution of X_t , the estimator is called "quasi" MLE (QMLE). The reader is referred to [Gourieroux et al. \(1984\)](#) for a general reference on QMLE.

2.1 Consistency of the PQMLE

As shown by the following theorem, the essential assumption required for the consistency of the PQMLE is that the conditional mean be well specified. Obviously, the following identifiability assumption is also required:

$$\lambda_1(\theta) = \lambda_1(\theta_0) \text{ almost surely if and only if } \theta = \theta_0. \quad (2.8)$$

Theorem 2.1 *Assume that (X_t) is an ergodic strictly stationary sequence valued in \mathbb{N} , satisfying (2.1)-(2.6) and (2.8). Then the PQMLE defined by (2.7) satisfies*

$$\hat{\theta}_n \rightarrow \theta_0 \quad \text{a.s.} \quad \text{as } n \rightarrow \infty.$$

In the sequel, K and ρ denote generic constants, or random variables depending on $\{X_u, u \leq 0\}$, such that $K > 0$ and $\rho \in (0, 1)$. It is often assumed that $\lambda_t(\theta)$ is a linear function of the past values. In that case, the regularity conditions become much simpler.

Remark 2.1 (linear conditional mean) *Assume that Θ is a compact subset of $(0, \infty) \times [0, \infty)^{p+q}$, that $\theta_0 = (\omega_0, \alpha_{01}, \dots, \beta_{0p})$, and that*

$$\lambda_t(\theta_0) = \omega_0 + \sum_{i=1}^q \alpha_{0i} X_{t-i} + \sum_{j=1}^p \beta_{0j} \lambda_{t-j}(\theta_0). \quad (2.9)$$

Assume also that for all $\theta = (\omega, \alpha_1, \dots, \beta_p) \in \Theta$, we have $\sum_{i=1}^p \beta_i < 1$. Noting that the equation (2.9) is similar to that satisfied by the volatility in a GARCH(p, q) model, it is easy to show (by for instance the arguments used to show (7.30) in [Francq and Zakoian \(2010\)](#), denoted hereafter FZ) that $a_t \leq K\rho^t$. The first condition in (2.4) directly follows. To show the second convergence of (2.4), it suffices to use the Borel-Cantelli lemma and $P(\rho^t X_t \geq \varepsilon) \leq \rho^t EX_t / \varepsilon$, for $\varepsilon > 0$. The conditions (2.2) and (2.4)-(2.6) are thus satisfied without any additional constraint. Let the polynomials $\mathcal{A}_\theta(z) = \sum_{i=1}^q \alpha_i z^i$ and $\mathcal{B}_\theta(z) = 1 - \sum_{i=1}^p \beta_i z^i$. As in the proof of (b) Page 157 in FZ, the identifiability condition (2.8) is satisfied by assuming that

$$\begin{aligned} & \text{if } p > 0, \mathcal{A}_{\theta_0}(z) \text{ and } \mathcal{B}_{\theta_0}(z) \text{ have no common root,} \\ & \text{at least one } \alpha_i \neq 0 \text{ for } i = 1, \dots, q, \text{ and } \beta_p \neq 0 \text{ if } \alpha_q = 0. \end{aligned} \quad (2.10)$$

In the case $q = 1$, the conditions (2.10) simply amount to assuming $\alpha_{01} > 0$.

The stationarity and ergodicity issues will be discussed for particular classes of count models in [Section 3](#).

2.2 Asymptotic distribution

As expected, under mild regularity conditions, the Poisson QMLE turns out to be asymptotically normal when the parameter belongs to the interior of the parameter space. In the more general situation where the parameter may lie at the boundary of the parameter space, its asymptotic distribution is the projection of a Gaussian random vector on a convex cone. Estimators with similar asymptotic distributions have been studied by *e.g.* [Andrews \(1999\)](#), [Francq and Zakoian \(2009\)](#) and the references therein.

2.2.1 When θ_0 belongs to the interior of Θ

To give conditions for the asymptotic normality of the PQMLE, we need to assume the existence of

$$E(X_t^2 | X_u, u < t) := v_t(\theta_0) + \lambda_t^2(\theta_0), \quad (2.11)$$

the existence of continuous second-order derivatives for $\lambda_t(\cdot)$ and $\tilde{\lambda}_t(\cdot)$, as well as the existence of the information matrices

$$J = E \frac{1}{\lambda_t(\theta_0)} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \lambda_t(\theta_0)}{\partial \theta'}, \quad I = E \frac{v_t(\theta_0)}{\lambda_t^2(\theta_0)} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \lambda_t(\theta_0)}{\partial \theta'}. \quad (2.12)$$

It is easy to see that the matrix J is invertible when

$$c' \frac{\partial \lambda_t(\theta_0)}{\partial \theta} = 0 \quad a.s. \Rightarrow c = 0. \quad (2.13)$$

We also assume that there exists a neighborhood $V(\theta_0)$ of θ_0 such that, for all $(i, j) \in \{1, \dots, d\}$,

$$E \sup_{\theta \in V(\theta_0)} \left| \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ell_t(\theta) \right| < \infty. \quad (2.14)$$

Assume also that, a.s.,

$$b_t, b_t X_t \text{ and } a_t d_t X_t \text{ are of order } O(t^{-\kappa}) \text{ for some } \kappa > 1/2, \quad (2.15)$$

where

$$b_t = \sup_{\theta \in \Theta} \left\| \frac{\partial \tilde{\lambda}_t(\theta)}{\partial \theta} - \frac{\partial \lambda_t(\theta)}{\partial \theta} \right\|, \quad d_t = \sup_{\theta \in \Theta} \max \left\{ \left\| \frac{1}{\lambda_t(\theta)} \frac{\partial \lambda_t(\theta)}{\partial \theta} \right\|, \left\| \frac{1}{\tilde{\lambda}_t(\theta)} \frac{\partial \tilde{\lambda}_t(\theta)}{\partial \theta} \right\| \right\}.$$

Theorem 2.2 *Assume that (X_t) satisfies the conditions of Theorem 2.1. Assume also (2.11)-(2.14) and (2.15). If $\theta_0 \in \overset{\circ}{\Theta}$, where $\overset{\circ}{\Theta}$ denotes the interior of Θ , then*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma := J^{-1} I J^{-1}) \quad \text{as } n \rightarrow \infty.$$

Note that when the distribution of X_t conditional to its past is Poisson, we have $I = J$, and thus $\Sigma = J^{-1}$, as established in [Ferland et al. \(2006\)](#).

It can be shown that, under the assumptions of Theorem 2.2, the asymptotic variance of the PQMLE can be consistently estimated by $\hat{\Sigma} = \hat{J}^{-1} \hat{I} \hat{J}^{-1}$ with

$$\hat{J} = \frac{1}{n} \sum_{t=s+1}^n \frac{1}{\tilde{\lambda}_t(\hat{\theta}_n)} \frac{\partial \tilde{\lambda}_t(\hat{\theta}_n)}{\partial \theta} \frac{\partial \tilde{\lambda}_t(\hat{\theta}_n)}{\partial \theta'}, \quad (2.16)$$

$$\hat{I} = \frac{1}{n} \sum_{t=s+1}^n \left(\frac{X_t}{\tilde{\lambda}_t(\hat{\theta}_n)} - 1 \right)^2 \frac{\partial \tilde{\lambda}_t(\hat{\theta}_n)}{\partial \theta} \frac{\partial \tilde{\lambda}_t(\hat{\theta}_n)}{\partial \theta'}. \quad (2.17)$$

Remark 2.2 (alternative conditions to (2.12) and (2.14)) *Note that*

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\theta) = \left(\frac{X_t}{\lambda_t(\theta)} - 1\right) \frac{\partial^2}{\partial\theta_i\partial\theta_j}\lambda_t(\theta) - \frac{X_t}{\lambda_t^2(\theta)} \frac{\partial}{\partial\theta_i}\lambda_t(\theta) \frac{\partial}{\partial\theta_j}\lambda_t(\theta). \quad (2.18)$$

Using Hölder's inequality and (2.3), one can thus obtain (2.14) by showing that

$$E \sup_{\theta \in V(\theta_0)} \left| \frac{\partial^2}{\partial\theta_i\partial\theta_j}\lambda_t(\theta) \right| < \infty, \quad (2.19)$$

and

$$E \sup_{\theta \in V(\theta_0)} \left| \frac{1}{\lambda_t(\theta)} \frac{\partial}{\partial\theta_i}\lambda_t(\theta) \right|^r < \infty, \quad E \sup_{\theta \in V(\theta_0)} \left| \frac{1}{\lambda_t(\theta)} \frac{\partial^2}{\partial\theta_i\partial\theta_j}\lambda_t(\theta) \right|^r < \infty \quad (2.20)$$

for all $r > 0$. Note also that (2.20) entails the existence of J . For the existence of I an extra assumption is needed. In particular, (2.12) is obtained under the conditions (2.20) and

$$E \left(\frac{v_t(\theta_0)}{\lambda_t(\theta_0)} \right)^{1+\varepsilon} < \infty, \text{ for some } \varepsilon > 0, \quad (2.21)$$

Remark 2.3 (linear conditional mean) *Let us come back to the framework of Remark 2.1.*

Because the roots of the polynomial $\mathcal{B}_\theta(z)$ are outside the unit disk, we have $\lambda_t(\theta) = \pi_0(\theta) + \sum_{k=1}^{\infty} \pi_k(\theta) X_{t-k}$ where

$$\sum_{k=1}^{\infty} \pi_k(\theta) z^k = \mathcal{B}_\theta^{-1}(z) \mathcal{A}_\theta(z) \quad \text{and} \quad \sup_{\theta \in \Theta} |\pi_k(\theta)| \leq K \rho^k.$$

We also have

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j}\lambda_t(\theta) = \pi_0^{(i,j)}(\theta) + \sum_{k=1}^{\infty} \pi_k^{(i,j)}(\theta) X_{t-k} \quad \text{with} \quad \sup_{\theta \in \Theta} |\pi_k^{(i,j)}(\theta)| \leq K \rho^k.$$

Under the moment assumption (2.3), the condition (2.19) is thus satisfied, whatever the neighborhood $V(\theta_0)$ included in Θ . One can show (2.20) by the arguments used to prove (7.54) in FZ. We thus obtain (2.14), and (2.12) under (2.21).

Now, note that $b_t \leq K \rho^t$ and that d_t admits moments at any order, by arguments already given. We thus have $E|t^\kappa b_t X_t| \leq K t^\kappa \rho^t$ and $E|t^\kappa a_t d_t X_t| \leq K t^\kappa \rho^t$, which entails (2.15) by the Borel-Cantelli lemma and the Markov inequality.

Finally, note that (2.13) is a consequence of (2.10), by the arguments used to show (b) Page 162 in FZ.

2.2.2 When θ_0 stands at the boundary of Θ

For the computation of the PQMLE it is obviously necessary to have $\tilde{\lambda}_t(\theta) > 0$ almost surely, for any $\theta \in \Theta$. For that reason, the parameter space Θ must be constrained. Very often,

one or several components of θ are constrained to be positive or equal to zero. For example, when we have an INGARCH(1,1) model of the form $\lambda_t(\theta) = \omega + \alpha X_{t-1} + \beta \lambda_{t-1}(\theta)$ then $\theta = (\omega, \alpha, \beta) \in \Theta \subset [\underline{\omega}, \infty) \times [0, \infty)^2$. Following the celebrated Box-Jenkins time series methodology, it is often interesting to test if the model can be simplified. For the INGARCH(1,1) example, it is of interest to test if the true parameter is of the form $\theta_0 = (\omega_0, \alpha_0, 0)$, *i.e* if the DGP is an INARCH(1). Theorem 2.2 does not apply because, in that situation, $\theta_0 \notin \overset{\circ}{\Theta}$. Moreover the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is not Gaussian because the positivity constraints entail that $\sqrt{n}(\hat{\beta}_n - \beta_0) \geq 0$ with probability one when $\beta_0 = 0$.

We now come back to the general model. The component i of the parameter θ is said to be positively constrained if the i -th section of Θ is of the form $[0, \bar{\theta}_i]$ with $\bar{\theta}_i > 0$. For example, for the linear model of Remark 2.1, the first component is not positively constrained, but the other components are. Let $d_2 = d - d_1 \in \{0, \dots, d\}$ be the number of positively constrained components of θ . Without loss of generality, assume that these d_2 constrained components are the last ones. The parameter θ_0 can belong to $\overset{\circ}{\Theta}$ even if one or several of its first d_1 components are equal to zero. However if one of the last d_2 components of θ_0 is equal to zero, then Theorem 2.2 does not apply because θ_0 stands at the boundary of Θ . We assume that $\Theta - \theta_0$ is large enough to contain an hypercube of the form $\prod_{i=1}^d [\underline{\theta}_i, \bar{\theta}_i]$ where, for all $i \in \{1, \dots, d\}$, $\underline{\theta}_i = 0$ if $\theta_{0i} = 0$ with $i > d_1$, $\underline{\theta}_i < 0$ otherwise, and $\underline{\theta}_i > 0$. Under this assumption we have

$$\lim_{n \rightarrow \infty} \sqrt{n}(\Theta - \theta_0) = \mathcal{C}, \quad (2.22)$$

where $\mathcal{C} = \prod_{i=1}^d \mathcal{C}_i$, in which $\mathcal{C}_i = \mathbb{R}$ when $i \leq d_1$ or $\theta_{0i} > 0$ and $\mathcal{C}_i = [0, +\infty)$ when $i > d_1$ and $\theta_{0i} = 0$.

Similarly to (2.15), assume that, a.s.,

$$\lim_{t \rightarrow \infty} c_t + X_t (a_t e_t + c_t + a_t d_t^2 + b_t d_t) = 0, \quad (2.23)$$

where

$$c_t = \sup_{\theta \in \Theta} \left\| \frac{\partial^2 \tilde{\lambda}_t(\theta)}{\partial \theta \partial \theta} - \frac{\partial^2 \lambda_t(\theta)}{\partial \theta \partial \theta'} \right\|,$$

$$e_t = \sup_{\theta \in \Theta} \max \left\{ \left\| \frac{1}{\lambda_t(\theta)} \frac{\partial^2 \lambda_t(\theta)}{\partial \theta \partial \theta'} \right\|, \left\| \frac{1}{\tilde{\lambda}_t(\theta)} \frac{\partial^2 \tilde{\lambda}_t(\theta)}{\partial \theta \partial \theta'} \right\| \right\}.$$

Note that, in the framework of Remarks 2.1 and 2.3, the condition (2.23) is always satisfied. Since J is positive definite, one can consider the norm $\|x\|_J^2 = x' J x$ and the scalar product $\langle x, y \rangle_J = x' J y$ for $x, y \in \mathbb{R}^d$. With this metric, the projection of a vector $Z \in \mathbb{R}^d$ on the convex

cone \mathcal{C} is defined by

$$Z^{\mathcal{C}} = \arg \inf_{C \in \mathcal{C}} \|C - Z\|_J$$

or equivalently by

$$Z^{\mathcal{C}} \in \mathcal{C} \quad \text{and} \quad \langle Z - Z^{\mathcal{C}}, C - Z^{\mathcal{C}} \rangle_J \leq 0, \quad \forall C \in \mathcal{C}. \quad (2.24)$$

Theorem 2.3 *Assume the conditions of Theorem 2.2 (except that $\theta_0 \in \overset{\circ}{\Theta}$) and (2.22), (2.23). Then, as $n \rightarrow \infty$,*

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} Z^{\mathcal{C}} = \arg \inf_{C \in \mathcal{C}} \|C - Z\|_J, \quad \text{where } Z \sim \mathcal{N}(0, \Sigma).$$

Note that, when $\theta_0 \in \overset{\circ}{\Theta}$ we have $\mathcal{C} = \mathbb{R}^d$ and $Z^{\mathcal{C}} = Z$. In that case, we retrieve the CAN of the PQMLE, as stated in Theorem 2.2. When $\theta_0 \notin \overset{\circ}{\Theta}$, the conditions required for the existence of the information matrices I and J can however be more demanding in terms of moments of X_t .

Remark 2.4 (alternative conditions to (2.12) and (2.14)) *When θ does not belong to the interior of Θ , the conditions (2.20) generally impose restrictive moment conditions on the observed process. For example, in the linear case considered in Remark 2.1 they may impose the existence of EX_t^r . By (2.18) and Hölder's inequality, (2.14) can be obtained by showing (2.19), (2.20) for $r = 3$ and $EX_t^3 < \infty$. In the linear case of Remark 2.1 this is equivalent to $EX_t^3 < \infty$. For (2.12), in particular the existence of I , the additional moment condition $Ev_t^3(\theta_0) < \infty$ on the conditional distribution of X_t can be imposed. Alternatively, one can impose $Ev_t^2(\theta_0) < \infty$ and (2.20) for $r = 4$.*

Note that the matrices I and J are still estimated by (2.16) and (2.17). As an application of Theorem 2.3, let us assume $d_2 > 0$ and consider the testing problem

$$H_0 : \theta_{0d} = 0 \quad \text{against} \quad H_1 : \theta_{0d} > 0.$$

Denote by $\widehat{\theta}_{nd}$ the last component of $\widehat{\theta}_n$ and denote by $\chi_k^2(\underline{\alpha})$ the $\underline{\alpha}$ -quantile of a chi-squared distribution with k degrees of freedom χ_k^2 . If one also assume that, under the null, only the last component of θ_0 is at the boundary (see Example 8.2 and Section 8.3.3 in FZ), then the test of rejection region

$$\left\{ \frac{n\widehat{\theta}_{nd}^2}{\widehat{\Sigma}(d, d)} \geq \chi_1^2(1 - 2\underline{\alpha}) \right\} \quad (2.25)$$

has the asymptotic level $\underline{\alpha}$. Note that when the last component is not positively constrained (*i.e.* when $d_2 = 0$) the PQMLE has a normal asymptotic distribution and the critical value of

the test is $\chi_1^2(1 - \underline{\alpha})$. Another application of Theorem 2.3 is given in the following corollary. Denote by δ_0 the Dirac mass at 0 and denote by $p_0\delta_0 + \sum_{i=1}^q p_i\chi_{k_i}^2$ the mixture of δ_0 and of $\chi_{k_i}^2$ -distributions, with the mixture weights p_0, \dots, p_q .

Corollary 2.1 (testing for constant conditional mean) *Consider an ergodic strictly stationary process (X_t) with a linear conditional mean of the form $\lambda_t(\theta) = \omega + \sum_{i=1}^q \alpha_i X_{t-i}$. Assume that the conditional distribution of X_t depends on the past only through $\lambda_t(\theta)$. If $\theta_0 = (\omega_0, 0, \dots, 0)' \in \Theta$, where Θ is a compact subset of $(0, \infty) \times [0, \infty)^q$, and if $EX_t^4 < \infty$, then the statistic*

$$S_n = n \sum_{i=1}^q \hat{\alpha}_{ni}^2 \xrightarrow{d} \frac{1}{2^q} \delta_0 + \sum_{i=1}^q \binom{q}{i} \frac{1}{2^q} \chi_i^2 \quad \text{as } n \rightarrow \infty, \quad (2.26)$$

where $\hat{\mu}_2$ and $\hat{\omega}_0$ denote consistent estimators of $\mu_2 = EX_1^2$ and $\omega_0 = EX_1$, and $\hat{\alpha}_{ni} = \hat{\theta}_{n,i+1}$.

The asymptotic distribution is known as a chi-bar-square distribution, and has been tabulated. By simply choosing $\hat{\mu}_2 = n^{-1} \sum_{i=1}^n X_t^2$ and $\hat{\omega}_0 = n^{-1} \sum_{i=1}^n X_t$, one can reject the constant conditional mean assumption at the asymptotic level $\underline{\alpha}$ if $\{S_n > c_{q,\underline{\alpha}}\}$, where the critical value $c_{q,\underline{\alpha}}$ can be found in Table 8.2 of FZ.

3 Application to particular models

We now show that the regularity conditions required for the asymptotic results of the previous section can be made explicit for the most popular classes of observation-driven and parameter-driven models for time series of counts (see Cox et al. (1981) for the distinction between observation-driven and parameter-driven models).

3.1 Poisson INGARCH model

One of the most natural count time series model is the Poisson INGARCH model, which has been studied by Ferland et al. (2006). This model is also called Autoregressive Conditional Poisson in Heinen (2003). The INGARCH(p, q) model is obtained by assuming that the conditional distribution of X_t given its past values is Poisson with intensity parameter of the linear form (2.9). Ferland et al. (2006) showed that there exists a stationary process (X_t) satisfying the INGARCH model, with second-order moments, under the assumption

$$\sum_{i=1}^r \alpha_{0i} + \beta_{0j} < 1 \quad (3.1)$$

with $r = \max\{p, q\}$ and the convention $\alpha_{0i} = 0$ when $i > q$ and $\beta_{0i} = 0$ when $i > p$. As shown in Tjøstheim (2012, 2014), the ergodicity of the stationary solution is a difficult issue. Fokianos et al. (2009) showed that this model can be approximated by an ergodic process, and applied this result to the likelihood inference. By using different techniques and different frameworks encompassing the Poisson INGARCH model, Neumann et al. (2011), Liu (2012), Davis and Liu (2012) and Christou and Fokianos (2013) showed the ergodicity. Under (3.1) and the assumptions of Remark 2.1, Theorem 2.1 thus establishes the strong consistency of the PQMLE. Since (3.1) also entails the existence of moments of any order (see Christou and Fokianos, 2013), the condition (2.21) is obviously verified, and Remark 2.3 entails that the conclusion of Theorem 2.2 holds true when θ_0 belongs to the interior of the parameter space. This was quite expected because the PQMLE is actually the MLE in the framework of this section. Similarly, the regularity conditions required for Theorem 2.3 and Corollary 2.1 are satisfied. To our knowledge, the asymptotic behaviour of the MLE had never been studied for count time series models with parameter at the boundary of the parameter space.

3.2 Negative binomial INGARCH model

As an alternative to the conditional Poisson distribution, Zhu (2011) and Christou and Fokianos (2013) considered the Negative Binomial distribution $\mathcal{NB}(r, p_t)$ with parameters $r > 0$ and $p_t = r/(\lambda_t + r)$ where λ_t is, for instance, of the form (2.9). We still have $E(X_t | X_u, u < t) = \lambda_t$, but the conditional variance $\lambda_t + \lambda_t^2/r$ is larger than the conditional variance of the Poisson case, which reflects the conditional overdispersion that is suspected to be present on real series (see Christou and Fokianos, 2013). From Proposition 3.4.1 in Liu (2012), Condition (3.1) entails the existence of an ergodic and strictly stationary solution (X_t) . In the case $(p, q) = (1, 1)$, it can be shown (see Christou and Fokianos, 2013), that the stationary solution is such that $EX_t^2 < \infty$ if and only if

$$(\alpha_0 + \beta_0)^2 + \frac{\alpha_0^2}{r} < 1, \quad (3.2)$$

writing α_0 and β_0 instead of α_{01} and β_{01} . Always in the case $(p, q) = (1, 1)$, it can be shown (see the appendix), that $EX_t^4 < \infty$ if and only if

$$(\alpha_0 + \beta_0)^4 + \frac{6\alpha_0^2(\alpha_0 + \beta_0)^2}{r} + \frac{\alpha_0^3(11\alpha_0 + 8\beta_0)}{r^2} + \frac{6\alpha_0^4}{r^3} < 1. \quad (3.3)$$

The conditions ensuring the existence of EX_t^2 are much more complicated for general orders p and q (see Theorem 2 in Zhu, 2011). The regularity conditions required for Theorems 2.1

and 2.2 are thus explicit, at least in the case $(p, q) = (1, 1)$. Christou and Fokianos (2013) had already noted that the Poisson QMLE is consistent in the case of a Negative binomial conditional distribution. To our knowledge the result stated in Theorem 2.3 and Corollary 2.1 are however new.

3.3 Double-Poisson INGARCH model (DACP model)

Count time series often exhibit over-dispersion, *i.e.* the variance larger than the mean, but the opposite phenomenon may be encountered. The Poisson and negative binomial INGARCH models can not take into account the under-dispersion. To tackle the problem, Heinen (2003) proposes a model based on the Double-Poisson distribution of Efron (1986). This distribution, which has two parameters $\lambda > 0$ and $\gamma > 0$, is defined by

$$P(X = x | \lambda, \gamma) = c(\lambda, \gamma) \left(\frac{e^{-x} x^x}{x!} \right) \left(\frac{e\lambda}{x} \right)^{\gamma x}, \quad x = 0, 1, \dots$$

where $c(\lambda, \gamma)$ is a normalization constant. We then use the notation $X \sim \mathcal{DP}(\lambda, \gamma)$. Efron (1986) shows that the mean of the $\mathcal{DP}(\lambda, \gamma)$ distribution is λ , and that its variance is approximately equal to λ/γ . The Double-Poisson INGARCH model is defined by assuming that the conditional distribution of X_t given its past values is $\mathcal{DP}(\lambda_t, \gamma)$ with parameters λ_t of the form (2.9). For $(p, q) = (1, 1)$, according to Propositions 3.1 and 3.2 in Heinen (2003), the condition (3.1) entails the existence of a stationary solution (X_t) such that

$$E(X_t) = \frac{\omega_0}{1 - \alpha_0 - \beta_0}, \quad \text{Var}(X_t) = \frac{1 - (\alpha_0 + \beta_0)^2 + \alpha_0^2 E[X_t]}{1 - (\alpha_0 + \beta_0)^2} \frac{E[X_t]}{\gamma}.$$

In view of Remark 2.1, the consistency result of Theorems 2.1 thus holds true in the case $(p, q) = (1, 1)$ when $\alpha_{01} + \beta_{01} < 1$ and $\alpha_{01} > 0$. Similarly, the conditions required for Theorems 2.2 and 2.3 are explicit in the INGARCH(1,1) case.

3.4 Log-linear model

One drawback of the previous models is that their coefficients are positively constrained, which entails statistical difficulties when a coefficient is equal to zero (see Theorem 2.3) and makes difficult to add exogenous explanatory variables to λ_t . Another drawback is that the autocovariances $\text{cov}(X_t, X_{t-h})$ are nonnegative at any lag h (see Christou and Fokianos (2013) for the explicit expression of these autocovariances for first-order models). To tackle these problems, Fokianos and Tjøstheim (2011) proposed a model in which the conditional distribution of X_t

given its past values is Poisson with intensity parameter $\lambda_t = e^{v_t}$, where

$$v_t = \omega_0 + \alpha_0 \log(X_{t-1} + 1) + \beta_0 v_{t-1}. \quad (3.4)$$

Under the conditions $|\beta_0| < 1$ and

$$|\alpha_0 + \beta_0| < 1 \text{ when } \alpha_0 > 0 \quad \text{and} \quad |\beta_0| |\alpha_0 + \beta_0| < 1 \text{ when } \alpha_0 < 0, \quad (3.5)$$

Fokianos and Tjøstheim (2011) showed that a slightly perturbed version of the log-linear model defined by (3.4) has a stationary and ergodic solution admitting moments of any order. Similar perturbed versions have been introduced by Fokianos et al. (2009) for INGARCH models. In view of the recent results on the stationarity and ergodicity of the INGARCH models, (see Neumann et al., 2011, Liu, 2012, Davis and Liu, 2012 and Christou and Fokianos, 2013), one can conjecture that the log-linear model itself admits a stationary and ergodic solution with moments of any order under (3.5). If this is the case, it is easy to verify that all the assumptions required for Theorems 2.1, 2.2 and 2.3, as well as for Corollary 2.1, are satisfied under the conditions $\alpha_0 \neq 0$ and (3.5).

3.5 INAR

One of the most popular count time series model is the integer-valued autoregressive (INAR) process. Contrary to the previous models, the INAR is parameter-driven. Since it is not obvious to compute the MLE of a parameter-driven model, the PQMLE model seems particularly attractive in this framework. The INAR(1) defines X_t as the convolution of a binomial distribution $\mathcal{B}(X_{t-1}, \alpha_0)$ (with the convention $\mathcal{B}(X_{t-1}, \alpha_0) = 0$ when $X_{t-1} = 0$) with a distribution ϵ_t on the integers. One can interpret $\mathcal{B}(X_{t-1}, \alpha_0)$ as the number of survivors from the population X_{t-1} and ϵ_t as the number of new arrivals, which is assumed to be independent of X_{t-1} . With this model we have (2.1) with $\lambda_t = \omega_0 + \alpha_0 X_{t-1}$ and $\omega_0 = E\epsilon_1$, obviously under the assumption that the expectation exists. In this case, and when the sequence (ϵ_t) is iid and $\alpha_0 < 1$, (X_t) is always stationary and $EX_1 = E\epsilon_1/(1 - \alpha_0)$. If $E\epsilon_1 \neq 0$, one can choose Θ such that (2.5) holds true. It is easy to see that the identifiability condition (2.8) is satisfied when the conditional distribution of X_t is not degenerated, which is the case when $\alpha_0 \neq 0$ or $\text{Var}(\epsilon_1) > 0$. Now, note that $v_t(\theta_0) = \alpha_0(1 - \alpha_0)X_{t-1} + \text{Var}(\epsilon_1)$, with $\theta_0 = (\omega_0, \alpha_0)$. Therefore (2.11) is satisfied. The information matrices I and J in (2.12) exist because we have $|v_t(\theta_0)/\lambda_t(\theta_0)| \leq c_0$, $\left\| \frac{1}{\lambda_t(\theta_0)} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \lambda_t(\theta_0)}{\partial \theta'} \right\| \leq c_0 + c_1 X_1$ for some constants c_0 and c_1 . We show (2.13) by the argument used to show (2.8). The second-order derivatives of $\lambda_t(\theta)$ being equal to zero, (2.14) is easily

verified. Since $a_t = b_t = c_t = d_t = e_t = 0$ for $t \geq 2$, the conditions (2.4), (2.15) and (2.23) are trivially satisfied.

4 Numerical illustrations

The first part of this section examines the finite sample behaviour of the PQMLE. The second part presents a simulation study concerning the test of nullity of one coefficient and the test of constant conditional mean. All the results of this section are based on $N = 1000$ independent replications of Monte Carlo simulations of different sample sizes n . For each simulation, the first 100 observations are omitted, so that the process approaches its stationary regime.

4.1 Finite sample behaviour of the PQMLE

The PQML function \tilde{L}_n , defined in (2.7), is optimized numerically, using the PORT routine (implemented by the function `nlminb()` of R).

The first Monte Carlo experiments concern the INAR(1) model. When the innovation ϵ_t follows a Poisson $\mathcal{P}(\lambda)$ distribution, then the conditional mean is $\lambda_t = \omega_0 + \alpha_0 X_{t-1}$ with $\omega_0 = \lambda$. When ϵ_t follows the geometric distribution $\mathcal{G}(p)$ with parameter $p \in (0, 1)$, then $\omega_0 = (1 - p)/p$. When $\epsilon_t \sim \mathcal{NB}(r, p)$ then $\omega_0 = rp/(1-p)$. We also simulated INGARCH(1,1) and Log-linear(1,1) models, with Poisson, Double-Poisson and binomial negative conditional distributions.

The means of the estimated values of θ_0 are given in the rows "PQMLE" of Table 1. This table also gives four different estimators of the root-mean-square deviation $\sqrt{E(\hat{\theta}_n - \theta_0)^2}$: the empirical standard error (ESE), the estimated standard error based on the asymptotic theory (ASE), the theoretical standard error based on the asymptotic theory (TSE), and the Poisson standard error based on the asymptotic theory assuming a Poisson conditional distribution (PSE). The ESE is equal to the root mean square error of estimation over the N replications. The ASE of the estimator of the i -th parameter is equal to the empirical mean of the estimated standard errors $\sqrt{\hat{\Sigma}(i, i)/n}$, where $\hat{\Sigma}$ is obtained from (2.16) and (2.17). The TSE is defined like the ASE, except that $\hat{\Sigma}$ is replaced by Σ computed from a very large simulation ($n = 5000$). The PSE is equal to the empirical mean of $\sqrt{\hat{J}^{-1}(i, i)/n}$ (noting that $\Sigma = J^{-1}$ when the conditional distribution is Poisson). The ESE offers the best view of the finite sample standard error of the PQMLE but, on real data series, only ASE and PSE are computable.

Table 1 shows that, for all the models, the means of the estimated parameters are satisfac-

Table 1: Finite sample behaviour of the PQMLE

INAR(1)										
n		$\epsilon_t \sim \mathcal{P}(2)$			$\epsilon_t \sim \mathcal{G}(0.5)$			$\epsilon_t \sim \mathcal{NB}(2, 0.5)$		
		$\omega_0=2$	$\alpha_0=0.9$		$\omega_0=1$	$\alpha_0=0.9$		$\omega_0=2$	$\alpha_0=0.9$	
500	PQMLE	2.159	0.892		1.075	0.892		2.173	0.892	
	ESE	0.412	0.021		0.232	0.023		0.442	0.022	
	ASE	0.404	0.020		0.225	0.022		0.425	0.020	
	TSE	0.409	0.020		0.229	0.022		0.424	0.021	
	PSE	0.895	0.046		0.340	0.037		0.713	0.037	
1000	PQMLE	2.066	0.897		1.048	0.895		2.070	0.896	
	ESE	0.286	0.014		0.169	0.016		0.287	0.014	
	ASE	0.283	0.014		0.159	0.016		0.294	0.015	
	TSE	0.285	0.014		0.161	0.016		0.297	0.015	
	PSE	0.620	0.032		0.236	0.025		0.497	0.026	
INGARCH(1,1)										
n		$\mathcal{P}(\lambda_t)$			$\mathcal{DP}(\lambda_t, 0.5)$			$\mathcal{NB}(3, p_t)$		
		$\omega_0=2$	$\alpha_0=0.3$	$\beta_0=0.6$	$\omega_0=2$	$\alpha_0=0.3$	$\beta_0=0.6$	$\omega_0=2$	$\alpha_0=0.3$	$\beta_0=0.6$
500	PQMLE	2.229	0.296	0.592	2.194	0.297	0.592	2.239	0.292	0.595
	ESE	0.703	0.039	0.059	0.658	0.038	0.058	0.758	0.045	0.064
	ASE	0.653	0.038	0.057	0.630	0.038	0.057	0.643	0.047	0.064
	TSE	0.619	0.038	0.056	0.616	0.038	0.056	0.667	0.047	0.063
	PSE	0.658	0.038	0.058	0.236	0.015	0.021	0.445	0.027	0.041
1000	PQMLE	2.134	0.298	0.595	2.087	0.297	0.593	2.168	0.298	0.596
	ESE	0.476	0.026	0.040	0.448	0.026	0.040	0.496	0.033	0.046
	ASE	0.444	0.027	0.040	0.435	0.027	0.040	0.481	0.032	0.045
	TSE	0.438	0.027	0.039	0.427	0.027	0.039	0.468	0.033	0.044
	PSE	0.446	0.027	0.040	0.167	0.010	0.015	0.313	0.019	0.028
Log-Linear(1,1)										
n		$\mathcal{P}(\lambda_t)$			$\mathcal{DP}(\lambda_t, 0.5)$			$\mathcal{NB}(3, p_t)$		
		$\omega_0=2$	$\alpha_0=0.3$	$\beta_0=-0.6$	$\omega_0=2$	$\alpha_0=0.3$	$\beta_0=-0.6$	$\omega_0=2$	$\alpha_0=0.3$	$\beta_0=-0.6$
500	PQMLE	1.946	0.304	-0.570	1.958	0.302	-0.586	1.976	0.301	-0.596
	ESE	0.214	0.048	0.116	0.203	0.047	0.107	0.217	0.052	0.116
	ASE	0.206	0.049	0.108	0.196	0.047	0.099	0.208	0.050	0.106
	TSE	0.205	0.050	0.105	0.146	0.047	0.062	0.177	0.051	0.084
	PSE	0.208	0.049	0.105	0.131	0.032	0.066	0.140	0.034	0.070
1000	PQMLE	1.975	0.303	-0.587	1.976	0.301	-0.595	1.990	0.300	-0.598
	ESE	0.150	0.035	0.076	0.142	0.034	0.072	0.155	0.037	0.079
	ASE	0.146	0.035	0.073	0.138	0.033	0.070	0.148	0.036	0.075
	TSE	0.146	0.035	0.073	0.111	0.033	0.050	0.136	0.036	0.067
	PSE	0.146	0.035	0.074	0.093	0.023	0.046	0.099	0.024	0.050

torily close to their theoretical values, especially for large sample sizes. Moreover the first three estimations of the standard deviations, the ESE, ASE and TSE, are very similar. The ASE and TSE are close because Σ is well estimated by (2.16)–(2.17). The closeness between ESE and ASE means that the asymptotic theory provides a reliable view on the actual standard error of the PQMLE. As expected, the standard errors decrease as the sample sizes increase. It is important to note that the PSE is different from the other estimators when the conditional distribution is not Poisson (*i.e.* for all the models but the two ones of the first columns of INGARCH(1,1) and Log-Linear(1,1)). The fact that PSE may be more than twice smaller or larger than the ESE demonstrates that, for a valid inference based on the PQMLE, it is crucial to rely on the asymptotic variance $\Sigma = J^{-1}IJ^{-1}$ instead of $\Sigma = J^{-1}$. From Table 1, we can thus draw the conclusion that ASE is a much more robust estimator of the PQMLE standard deviation than PSE.

Figures 4.1 displays the boxplot and histogram of the $N = 1000$ values of the PQMLE (centred and reduced) for simulations of length $n = 3000$ of an INAR(1) with $\epsilon_1 \sim \mathcal{NB}(3, 0.6)$ and $\alpha_0 = 0.9$. In agreement with Theorem 2.2, the empirical distribution of the estimator resembles the standard Gaussian law. Other simulation experiments, not presented here for sake of conciseness, reveal similar behaviors for other models and other values of parameters, provided they are sufficiently far from zero. Indeed, in accordance with Theorem 2.3, the empirical distribution of the PQMLE moves away the Gaussian when the parameter gets closer to the boundary of the parameter space. Table 2 gives the p -values of the Kolmogorov-Smirnov test of normality for the N values of the PQMLE, computed on simulations of size $n = 3000$ of each of the models considered in Table 1. The normality assumption is never rejected.

Table 2: p -values of the Kolmogorov-Smirnov test of normality of the PQMLE

	INAR(1)			INGARCH(1,1)			Log-Linear(1,1)		
	$\mathcal{P}(2)$	$\mathcal{G}(0.5)$	$\mathcal{NB}(2, 0.5)$	$\mathcal{P}(\lambda_t)$	$\mathcal{DP}(\lambda_t, 0.5)$	$\mathcal{NB}(3, p_t)$	$\mathcal{P}(\lambda_t)$	$\mathcal{DP}(\lambda_t, 0.5)$	$\mathcal{NB}(3, p_t)$
$\hat{\omega}$	0.244	0.395	0.302	0.396	0.080	0.257	0.936	0.961	0.385
$\hat{\alpha}$	0.318	0.449	0.768	0.542	0.707	0.841	0.756	0.487	0.584
$\hat{\beta}$				0.848	0.384	0.851	0.658	0.890	0.969

INAR(1)

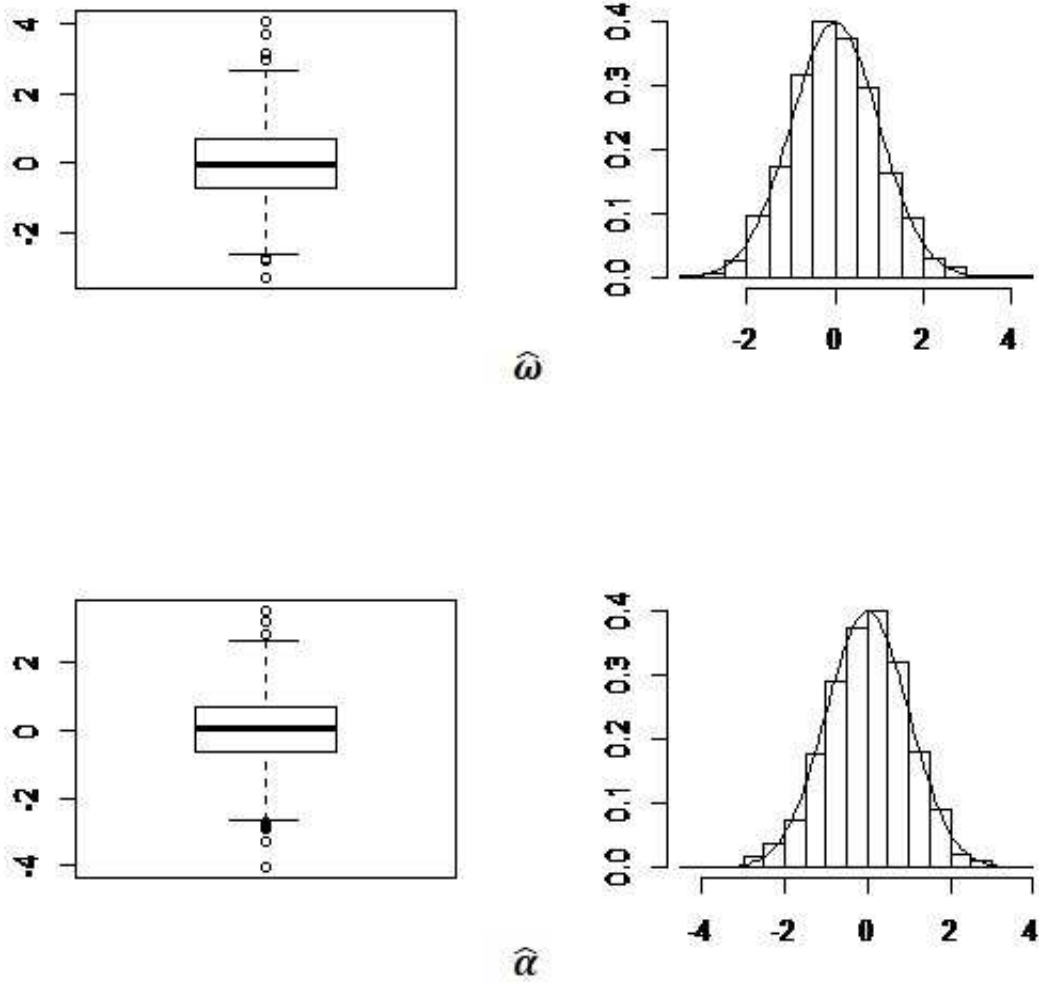


Figure 1: Boxplot and histogram of the standardized distribution of $\hat{\theta} = (\hat{\omega}, \hat{\alpha})$ for an INAR(1) model with negative binomial innovations. Superimposed is the standard normal density function.

4.2 Significance tests based on the PQMLE

We now report a Monte Carlo experiment for examining the performance of two adequacy tests for the conditional mean: the test that one coefficient is equal to zero, and the test of constant conditional mean. The simulation is implemented to obtain the sizes and the powers of the tests for different sample sizes. The tests are carried out at asymptotic level $\underline{\alpha} = 5\%$.

4.2.1 Empirical behavior of the tests under the null

For the test of nullity of one coefficient, we consider two different classes of DGPs. The first DGP's are INGARCH(1,1) models with $(\omega_0, \alpha_0, \beta_0) = (2, 0.5, 0)$ and three different conditional distributions, a Poisson, a negative binomial and a double-Poisson. On each of the $N = 1000$ simulations, we fit an INGARCH(1,1) model by PQMLE, and carry out the test of $H_0 : \beta_0 = 0$ against $H_1 : \beta_0 > 0$. The null is rejected for large values of the test statistic $n\hat{\beta}_n^2/\hat{\Sigma}(3, 3)$. As the parameters of the INGARCH(1,1) are positively constrained, according to (2.25), we use the critical value $\chi_1^2(1 - 2\underline{\alpha})$. The second class of DGP's is that of the Log-Linear(1,1) models with $(\omega_0, \alpha_0, \beta_0) = (2, -0.5, 0)$. We test the same hypotheses and use the same test statistic. However, as the regression parameters of the Log-Linear model are not positively constrained, we use the usual critical value $\chi_1^2(1 - \underline{\alpha})$. The relative rejection frequencies are shown in Table 3. Recall that, over $N = 1000$ independent replications of a test having the exact level 5%, the relative rejection frequency should vary between 3.6% and 6.4% with probability 95%. For the sample size $n = 1000$, the empirical sizes of the tests are thus in perfect agreement with the nominal level $\underline{\alpha} = 5\%$.

Table 3: Size of the test of nullity of β_0

INGARCH(1,1)			
n	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(3, p_t)$	$\mathcal{DP}(\lambda_t, 0.5)$
100	6.6	6.3	7.1
1000	4.1	3.8	5.2
Log-Linear(1,1)			
n	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(3, p_t)$	$\mathcal{DP}(\lambda_t, 0.5)$
100	16	18.4	14.4
1000	6.1	5.26	5.4

For the test of constant conditional mean, we simulate INARCH(3) models with $(\omega_0, \alpha_{01}, \alpha_{02}, \alpha_{03}) =$

(2, 0, 0, 0). We then carry out the test of

$$H_0 : \alpha_{01} = \alpha_{02} = \alpha_{03} = 0 \quad \text{against} \quad H_1 : \text{at least one } \alpha_{0i} > 0 \text{ for } i = 1, 2, 3.$$

In view of Corollary 2.1, the null is rejected when the statistic $S_n = n (\hat{\alpha}_{n1}^2 + \hat{\alpha}_{n2}^2 + \hat{\alpha}_{n3}^2)$ exceeds the $\underline{\alpha}$ -quantile $c_{3,\underline{\alpha}} = 5.43$ of the chi-bar-square distribution. The relative frequencies of rejection are given in Table 4. We can note that, at least when $n = 1000$, the observed relative rejection frequencies of all the tests are not significantly different from the theoretical level 5%.

Table 4: Size of the test of constant conditional mean

INGARCH(1,1)			
n	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(3, p_t)$	$\mathcal{DP}(\lambda_t, 0.5)$
100	3.7	3.2	4.9
1000	3.9	5.7	4.5

4.2.2 Empirical behavior of the tests under the alternative

To study the power of the tests, we now simulate INGARCH(1,1) processes with $(\omega_0, \alpha_0) = (2, 0.3)$ and $\beta_0 \in \{0.05, 0.2, 0.6\}$, and Log-Linear(1,1) processes with $(\omega_0, \alpha_0) = (2, 0.3)$ and $\beta_0 \in \{-0.1, -0.3, -0.6\}$. We carry out the test of nullity of the coefficient β_0 for both kind of models. Table 5 shows that the test works as expected: the power increases as the sample size increases and as the value of β_0 increases.

For the test of constant conditional mean, we simulate INARCH(3) models with $(\omega_0, \alpha_{01}, \alpha_{02}) = (2, 0, 0)$ and $\alpha_{03} \in \{0.05, 0.1, 0.4\}$. Table 6 shows that this test also works reasonably well.

A way to visualize the power of a test is to plot the function of the relative rejection frequencies (RRF)

$$RRF(z) = \frac{1}{N} \sum_{j=1}^N I(p_j < z), \quad z \in [0, 1],$$

where p_j denotes the observed p -value for the j -th replication of the test, and $I(p_j < z)$ is an indicator function that takes the value 1 if its argument is true and 0 otherwise. Figure 2 displays the RRF functions of the test of nullity of one coefficient (2.25) and of the test of constant conditional mean (2.26), for different sample sizes n . The first test is applied with the null $H_0 : \beta_0 = 0$ on simulations of the INGARCH(1,1) process with $(\omega_0, \alpha_0, \beta_0) = (2, 0.6, 0.1)$ and the conditional distribution $\mathcal{NB}(3, p_t)$. The second test is applied to the INARCH(3) process with the conditional distribution $\mathcal{NB}(3, p_t)$ and $(\omega_0, \alpha_{01}, \alpha_{02}, \alpha_{03}) = (2, 0, 0, 0.1)$. In

Table 5: Power of the test of nullity of β_0

INGARCH(1,1)				
n	β_0	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(3, p_t)$	$\mathcal{DP}(\lambda_t, 0.5)$
100	$\beta_0 = 0.05$	11.9	12.6	12.3
	$\beta_0 = 0.2$	22.3	20.0	21.6
	$\beta_0 = 0.6$	86.4	84.5	85.2
1000	$\beta_0 = 0.05$	14.2	14.2	11.5
	$\beta_0 = 0.2$	68.2	62.9	65.8
	$\beta_0 = 0.6$	100	100	100
Log-Linear(1,1)				
n	β_0	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(3, p_t)$	$\mathcal{DP}(\lambda_t, 0.5)$
100	$\beta_0 = -0.1$	20.1	18.2	17.3
	$\beta_0 = -0.3$	23.6	26.9	27.7
	$\beta_0 = -0.6$	60.4	65.5	63.7
1000	$\beta_0 = -0.1$	32.4	16.3	18.8
	$\beta_0 = -0.3$	73.1	78.7	82.8
	$\beta_0 = -0.6$	100	100	100

Table 6: Power of the test of constant conditional mean

INGARCH(1,1)				
n	θ_0	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(3, p_t)$	$\mathcal{DP}(\lambda_t, 0.5)$
100	$\theta_0 = (2, 0, 0, 0.05)$	5.5	7.7	11
	$\theta_0 = (2, 0, 0, 0.1)$	13.3	13.1	13.8
	$\theta_0 = (2, 0, 0, 0.4)$	92.6	90.1	91.1
1000	$\theta_0 = (2, 0, 0, 0.05)$	29.9	29	24.5
	$\theta_0 = (2, 0, 0, 0.1)$	83.3	81.1	76
	$\theta_0 = (2, 0, 0, 0.4)$	100	100	100

Figure 2, the more concave the shape of a curve is, the better the corresponding test is in terms of power. Note that, for the first test, $RRF(z)$ does not reach 1 when $z = 1$. This is due to the fact that when the test statistic takes the value zero (which appears with non zero probability, even under the alternative) the p -value is equal to 1 (*i.e.* the probability that a chi-bar-square distribution be positive or equal to zero).

5 Real data application

In this section, we report an application of the PQMLE to financial time series data. The data set is obtained from the QUANDL search engine and it contains the daily number of

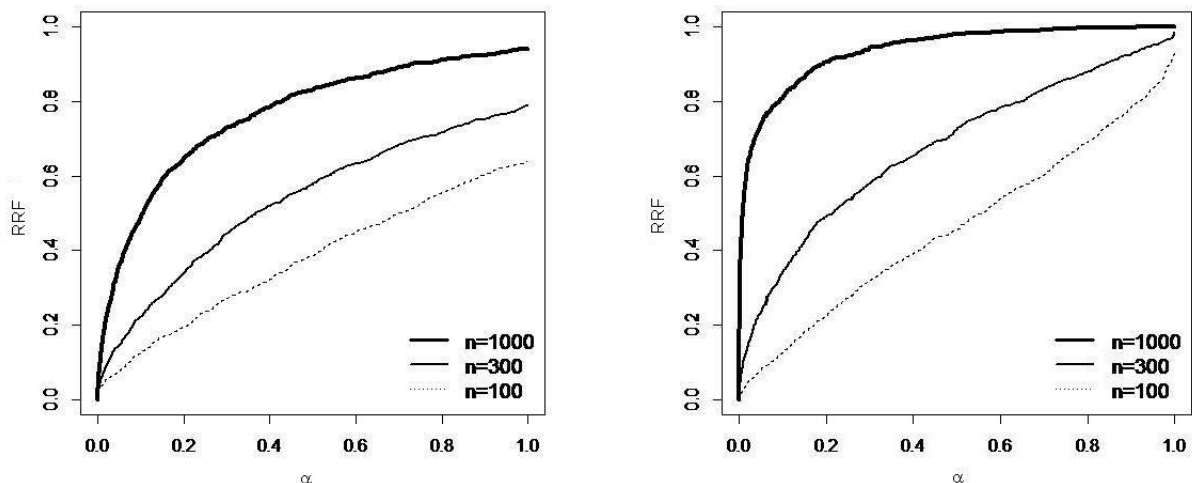


Figure 2: Empirical power of the test of nullity of one coefficient (left plot) and of the test of constant conditional mean (right plot), measured by the function of the relative rejection frequencies (RRF)

trades of 6 stocks listed in the NYSE Euronext group, namely CR.FONC.MONACO, SIRAGA, TECHNOFIRST, SIPAREX CROISSANCE, PROXIMIDIA and ACHMEA (see Figure 3). The size of the series varies from 1006 to 3633.

Table 7: The dispersion of the data

	C.F.M	SIRAGA	TECHNOFIRST	SIPAREX	PROXIMIDIA	ACHMEA
Mean	2.226	3.589	4.132	10.019	1.736	23.788
Variance	2.963	17.578	19.562	129.730	1.586	234.854

Table 7 shows that the series are overdispersed (their empirical variances are larger than their means), with the exception of the PROXIMIDIA stock which is underdispersed. For each series, we fitted INGARCH(3,1), INGARCH(2,1), INGARCH(1,1) and INARCH(1) models. The estimated parameters are shown in Table 8. This table also gives, into parentheses, the p -values of the test (2.25) of nullity of the corresponding coefficient. The p -values that are less than 0.05 are underlined. To illustrate the table, take the example of the daily number of transactions of the CR.FONC.MONACO stock. For the full INGARCH(3,1) model, the parameter $\hat{\beta}_3$ is not statistically significant. Constrained INGARCH(3,1) models (assuming $\beta_{01} = 0$, or $\beta_{02} = 0$, or

even $\beta_{01} = \beta_{02} = 0$) have also been tried, but these constrained models do not seem adapted to this series. The first beta coefficient of the INGARCH(2,1) model does not appear to be significant, thus a conditional mean of the form $\lambda_t = 0.340 + 0.112X_{t-1} + 0.735\lambda_{t-2}$ is retained for this series. The residuals of the simpler INGARCH(1,1) and INARCH(1) models (not presented here) present signs of correlatedness. It is interesting to note that, for all the series, the sum of the estimated values of the α and β coefficients is close to 0.9, which indicates a strong persistence in the dynamics. This is in accordance with the clusters of high values that are observed on the series plotted in Figure 3.

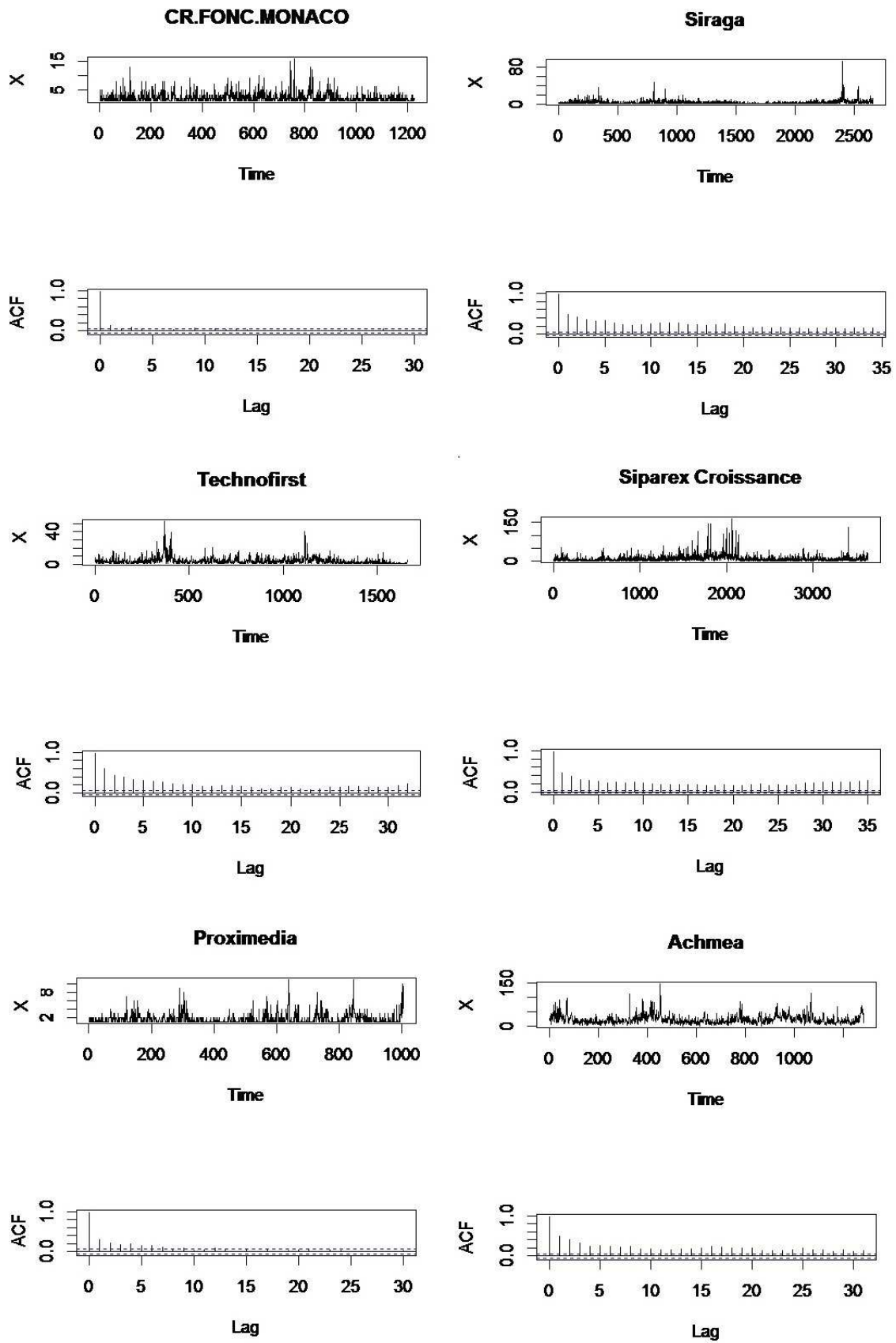


Figure 3: The time series and their sample autocorrelation functions

Table 8: PQMLE of INGARCH models for the expected number of transactions (and p -value of the test of nullity of the coefficient)

INGARCH(3,1)						
$\hat{\theta}$	C.F.M	SIRAGA	TECHNOFIRST	SIPAREX	PROXIMIDIA	ACHMEA
$\hat{\omega}$	0.280	0.187	0.492	0.728	0.165	2.413
$\hat{\alpha}$	0.111(0.030)	0.029(0.000)	0.384(0.000)	0.314(0.000)	0.224(0.000)	0.335(0.000)
$\hat{\beta}_1$	0.075(0.761)	0.469(0.000)	0.123(0.109)	0.435(0.000)	0.453(0.032)	0.433(0.000)
$\hat{\beta}_2$	0.689(0.000)	0.000(1.000)	0.230(0.004)	0.000(1.000)	0.086(0.756)	0.000(1.000)
$\hat{\beta}_3$	0.000(1.000)	0.189(0.023)	0.144(0.021)	0.178(0.003)	0.145(0.394)	0.130(0.169)
$\hat{\omega}$	0.283	0.200	0.522	0.805	0.202	2.646
$\hat{\alpha}$	0.111(0.000)	0.307(0.000)	0.397(0.000)	0.341(0.000)	0.261(0.000)	0.357(0.000)
$\hat{\beta}_2$	0.710(0.000)	0.425(0.000)	0.288(0.000)	0.305(0.000)	0.228(0.141)	0.410(0.000)
$\hat{\beta}_3$	0.051(0.462)	0.213(0.000)	0.188(0.000)	0.274(0.000)	0.398(0.024)	0.122(0.044)
$\hat{\omega}$	0.267	0.187	0.482	0.728	0.164	2.413
$\hat{\alpha}$	0.096(0.003)	0.290(0.000)	0.368(0.000)	0.314(0.000)	0.221(0.000)	0.335(0.000)
$\hat{\beta}_1$	0.676(0.001)	0.469(0.000)	0.281(0.000)	0.435(0.000)	0.500(0.000)	0.433(0.000)
$\hat{\beta}_3$	0.108(0.532)	0.189(0.006)	0.234(0.000)	0.178(0.000)	0.187(0.061)	0.130(0.025)
$\hat{\omega}$	0.854	0.309	0.657	1.299	0.404	4.126
$\hat{\alpha}$	0.122(0.000)	0.333(0.000)	0.408(0.000)	0.376(0.000)	0.285(0.000)	0.401(0.000)
$\hat{\beta}_3$	0.495(0.001)	0.582(0.000)	0.433(0.000)	0.495(0.000)	0.485(0.000)	0.426(0.000)
INGARCH(2,1)						
$\hat{\omega}$	0.279	0.198	0.522	0.775	0.157	2.686
$\hat{\alpha}$	0.111(0.000)	0.278(0.000)	0.377(0.000)	0.308(0.000)	0.202(0.000)	0.338(0.000)
$\hat{\beta}_1$	0.075(0.410)	0.485(0.000)	0.169(0.013)	0.417(0.000)	0.564(0.012)	0.406(0.000)
$\hat{\beta}_2$	0.689(0.000)	0.182(0.122)	0.328(0.000)	0.197(0.005)	0.145(0.448)	0.143(0.140)
$\hat{\omega}$	0.340	0.257	0.657	1.037	0.265	3.416
$\hat{\alpha}$	0.112(0.000)	0.298(0.000)	0.404(0.000)	0.335(0.000)	0.221(0.000)	0.372(0.000)
$\hat{\beta}_2$	0.735(0.000)	0.631(0.000)	0.437(0.000)	0.562(0.000)	0.628(0.000)	0.485(0.000)
INGARCH(1,1)						
$\hat{\omega}$	0.290	0.201	0.567	0.802	0.148	2.842
$\hat{\alpha}$	0.090(0.000)	0.278(0.000)	0.343(0.000)	0.288(0.000)	0.184(0.000)	0.326(0.000)
$\hat{\beta}_1$	0.780(0.000)	0.687(0.000)	0.520(0.000)	0.632(0.000)	0.733(0.000)	0.554(0.000)
INARCH(1)						
$\hat{\omega}$	1.905	1.686	1.948	4.569	1.190	11.458
$\hat{\alpha}$	0.145(0.000)	0.531(0.000)	0.529(0.000)	0.545(0.000)	0.315(0.000)	0.518(0.000)

6 Conclusion

The PQMLE provides a general approach for estimating the conditional mean parameters of time series of counts. If the conditional mean is correctly specified, under some regularity conditions, the PQMLE is CAN, even if the conditional distribution is not Poisson. For the asymptotic variance, it is however important to employ the robust expression $\Sigma = J^{-1}IJ^{-1}$ instead of the expression $\Sigma = J^{-1}$ which may be invalid when the conditional distribution is not Poisson. When the parameter stands at the boundary of the parameter space, the asymptotic distribution of the PQMLE is no more Gaussian. This is a usual framework which appears, for instance, when we have an over-identified INGARCH(p, q) model (*i.e.* p or q is larger than necessary). When assessing the significance of the estimated parameters, we thus have to take into account the fact that the PQMLE has a special non Gaussian asymptotic distribution under the null. This leads to adequacy tests with chi-bar-square distributions instead of usual chi-square distributions. Note that these particular distributions of the estimator and its related tests also hold for the MLE (*i.e.* when the conditional distribution is Poisson).

In view of [Fokianos \(2012\)](#), the INGARCH model admits a weak ARMA representation. The result still holds true when the conditional distribution is not Poisson. In principle, we could thus use general diagnostic checking tools of weak ARMA models (as in [Francq et al. 2005](#)) for identifying the orders p and q of a conditional mean of the INGARCH(p, q) form (2.9). The problem deserves however more thought, and is left for future work. Other possible extensions of the present work include models for time series valued in \mathbb{Z} (see *e.g.* [Kachour and Truquet, 2011](#) and [Andersson and Karlis, 2014](#)) which appear naturally, in particular when a count time series is differenced, and for which a QMLE could be searched.

7 Proofs

Proof of Theorem 2.1. Let $\ell_t(\theta)$ and $L_n(\theta)$ be the random variables obtained by replacing $\tilde{\lambda}_t(\theta)$ by $\lambda_t(\theta)$ in $\tilde{\ell}_t(\theta)$ and $\tilde{L}_n(\theta)$, respectively. Using (2.2), (2.5) and the inequality $\log(1+x) \leq x$, we have

$$\left| \log \tilde{\lambda}_t(\theta) - \log \lambda_t(\theta) \right| = \left| \log \left(1 + \frac{\tilde{\lambda}_t(\theta) - \lambda_t(\theta)}{\lambda_t(\theta)} \right) \right| \leq \frac{a_t}{\underline{\omega}}.$$

By (2.4), as $n \rightarrow \infty$

$$\sup_{\theta \in \Theta} \left| \tilde{L}_n(\theta) - L_n(\theta) \right| \leq \frac{1}{n} \sum_{t=s+1}^n a_t + X_t \frac{a_t}{\underline{\omega}} \rightarrow 0 \quad \text{a.s.} \quad (7.1)$$

Now note that, using (2.1), and again $\log(x) \leq x - 1$ for $x > 0$,

$$\begin{aligned} E \{ \ell_1(\theta) - \ell_1(\theta_0) \} &= E \left\{ \lambda_1(\theta_0) \log \frac{\lambda_1(\theta)}{\lambda_1(\theta_0)} - \lambda_1(\theta) + \lambda_1(\theta_0) \right\} \\ &\leq E \left\{ \lambda_1(\theta_0) \left(\frac{\lambda_1(\theta)}{\lambda_1(\theta_0)} - 1 \right) - \lambda_1(\theta) + \lambda_1(\theta_0) \right\} = 0 \end{aligned}$$

with equality iff $\theta = \theta_0$ by (2.8).

From (2.2) and (2.3), it can be seen that $|\log \lambda_1(\theta_0)|$ admits moments of any order. Hölder's inequality and (2.3) then entail that

$$E|X_1 \log \lambda_0(\theta_0)| \leq \|X_1\|_{1+\varepsilon} \|\log \lambda_1(\theta_0)\|_{1+1/\varepsilon} < \infty.$$

We thus have $E|\ell_1(\theta_0)| < \infty$. Therefore $E \{ \ell_1(\theta) - \ell_1(\theta_0) \}$ belongs *a priori* to $[-\infty, 0]$, and one can deduce

$$E\ell_1(\theta) < E\ell_1(\theta_0), \quad \forall \theta \neq \theta_0. \quad (7.2)$$

For $k \in \mathbb{N}^*$ and $\theta_1 \in \Theta$, let $V_k(\theta_1)$ be the open ball of center θ_1 and radius $1/k$. Note that $\left\{ \sup_{\theta \in V_k(\theta_1) \cap \Theta} \ell_t(\theta) \right\}_t$ is an ergodic stationary sequence, as a measurable function of the ergodic stationary process (X_t) . Note also that $E \sup_{\theta \in V_k(\theta_1) \cap \Theta} \ell_t(\theta)$ belongs to $\mathbb{R} \cup \{-\infty\}$. In view of (7.1) and the ergodic theorem (see Billingsley, 2008, pp. 284 and 495) we thus obtain

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in V_k(\theta_1) \cap \Theta} \tilde{L}_n(\theta) = \limsup_{n \rightarrow \infty} \sup_{\theta \in V_k(\theta_1) \cap \Theta} L_n(\theta) \leq E \sup_{\theta \in V_k(\theta_1) \cap \Theta} \ell_1(\theta).$$

By Beppo Levi's theorem, $E \sup_{\theta \in V_k(\theta_1) \cap \Theta} \ell_1(\theta)$ decreases to $E\ell_1(\theta_1)$ as $k \rightarrow \infty$. In view of (7.2), we have shown that for all $\theta_1 \neq \theta_0$ there exists a neighborhood $V(\theta_1)$ of θ_1 such that

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in V(\theta_1) \cap \Theta} \tilde{L}_n(\theta) < \limsup_{n \rightarrow \infty} \tilde{L}_n(\theta_0) = E\ell_1(\theta_0). \quad (7.3)$$

The conclusion follows from a standard argument, using the compactness of Θ . \square

Proof of Theorem 2.2. First consider the impact of the initial values. We have

$$\begin{aligned} \sqrt{n} \sup_{\theta \in \Theta} \left\| \frac{\partial}{\partial \theta} \tilde{L}_n(\theta) - \frac{\partial}{\partial \theta} L_n(\theta) \right\| &\leq \frac{1}{\sqrt{n}} \sum_{t=s+1}^n \left\{ b_t + X_t \left(\frac{a_t}{\underline{\omega}} d_t + \frac{b_t}{\underline{\omega}} \right) \right\} \\ &= o(1) \end{aligned} \quad (7.4)$$

almost surely, by (2.15). For n large enough, $\hat{\theta}_n$ does not lie at the boundary of Θ and thus we have

$$0 = \sqrt{n} \frac{\partial}{\partial \theta} \tilde{L}_n(\hat{\theta}_n) \stackrel{o(1)}{=} \sqrt{n} \frac{\partial}{\partial \theta} L_n(\hat{\theta}_n) = \sqrt{n} \frac{\partial}{\partial \theta} L_n(\theta_0) - J_n^* \sqrt{n}(\hat{\theta}_n - \theta_0), \quad (7.5)$$

where $a \stackrel{c}{=} b$ stands for $a = b + c$, and J_n^* is a matrix whose generic term is of the form $-\partial^2 L_n(\theta_{ij}^*)/\partial\theta_i\partial\theta_j$, for some θ_{ij}^* between $\hat{\theta}_n$ and θ_0 . Note that

$$\sqrt{n}\frac{\partial}{\partial\theta}L_n(\theta_0) = \frac{1}{\sqrt{n}}\sum_{t=s+1}^n U_t, \quad U_t = \left(\frac{X_t}{\lambda_t(\theta_0)} - 1\right)\frac{\partial\lambda_t(\theta_0)}{\partial\theta}, \quad (7.6)$$

where $\{U_t, \mathcal{F}_t\}$ is a stationary martingale difference, \mathcal{F}_t denoting the σ -field generated by $\{X_u, u \leq t\}$. In view of (2.11) and (2.12) we have $EU_tU_t' = I$. The central limit theorem of Billingsley (1961) for square-integrable stationary martingale difference then entails that

$$\sqrt{n}\frac{\partial}{\partial\theta}L_n(\theta_0) \xrightarrow{d} \mathcal{N}(0, I) \quad \text{as } n \rightarrow \infty. \quad (7.7)$$

Let $V_m(\theta_0)$ be the ball of center θ_0 and radius $1/m$. Assume that m is large enough so that $V_m(\theta_0)$ is included in the neighborhood $V(\theta_0)$ defined in (2.14). Suppose that n is sufficiently large, so that $\theta_{ij}^* \in V_m(\theta_0)$. With probability one,

$$\begin{aligned} |J_n^*(i, j) - J(i, j)| &\leq \frac{1}{n} \sum_{t=s+1}^n \sup_{\theta \in V_m(\theta_0)} \left| \frac{\partial^2}{\partial\theta_i\partial\theta_j} \ell_t(\theta) - E \frac{\partial^2}{\partial\theta_i\partial\theta_j} \ell_t(\theta_0) \right| \\ &\rightarrow E \sup_{\theta \in V_m(\theta_0)} \left| \frac{\partial^2}{\partial\theta_i\partial\theta_j} \ell_t(\theta) - E \frac{\partial^2}{\partial\theta_i\partial\theta_j} \ell_t(\theta_0) \right| \end{aligned} \quad (7.8)$$

as $n \rightarrow \infty$. Under (2.14), the Lebesgue dominated convergence theorem entails that

$$\lim_{m \rightarrow \infty} E \sup_{\theta \in V_m(\theta_0)} \left| \frac{\partial^2}{\partial\theta_i\partial\theta_j} \ell_t(\theta) - E \frac{\partial^2}{\partial\theta_i\partial\theta_j} \ell_t(\theta_0) \right| = 0. \quad (7.9)$$

It follows that

$$J_n^* \rightarrow J \quad \text{a.s.}$$

The conclusion follows from (7.5), (7.7) and (2.13). \square

Proof of Theorem 2.3. For all $\theta \in \Theta$, a second order Taylor expansion of $\tilde{L}_n(\theta)$ at θ_0 yields

$$\tilde{L}_n(\theta) - \tilde{L}_n(\theta_0) = \frac{\partial L_n(\theta_0)}{\partial\theta'}(\theta - \theta_0) - \frac{1}{2}(\theta - \theta_0)'J(\theta - \theta_0) + R_n(\theta),$$

where

$$R_n(\theta) = \left\{ \frac{\partial \tilde{L}_n(\theta_0)}{\partial\theta'} - \frac{\partial L_n(\theta_0)}{\partial\theta'} \right\} (\theta - \theta_0) + \frac{1}{2}(\theta - \theta_0)' \left(\frac{\partial^2 \tilde{L}_n(\theta^*)}{\partial\theta\partial\theta'} + J \right) (\theta - \theta_0),$$

and θ^* is between θ and θ_0 . Note that $\partial L_n(\theta_0)/\partial\theta'$ has to be understood as a vector of right derivatives. Even if θ_0 contains null components, this vector of right-derivatives is well defined, and is equal to $n^{-1} \sum_{t=s+1}^n U_t$, as in (7.6). Introducing the vector

$$Z_n = J^{-1} \sqrt{n} \frac{\partial L_n(\theta_0)}{\partial\theta},$$

we can write

$$\tilde{L}_n(\theta) - \tilde{L}_n(\theta_0) = \frac{1}{2n} \|Z_n\|_J^2 - \frac{1}{2n} \|Z_n - \sqrt{n}(\theta - \theta_0)\|_J^2 + R_n(\theta).$$

Let the projection of Z_n on \mathcal{C}

$$Z_n^{\mathcal{C}} = \arg \inf_{C \in \mathcal{C}} \|C - Z_n\|_J.$$

Define also

$$\theta_{Z_n} = \arg \inf_{\theta \in \Theta} \|\sqrt{n}(\theta - \theta_0) - Z_n\|_J.$$

In view of (2.22), we have

$$\sqrt{n}(\theta_{Z_n} - \theta_0) = Z_n^{\mathcal{C}} \text{ for } n \text{ large enough.} \quad (7.10)$$

By definition of θ_{Z_n} and $\hat{\theta}_n$, and Lemma 7.1 below, we have

$$\begin{aligned} 0 &\leq \|\sqrt{n}(\hat{\theta}_n - \theta_0) - Z_n\|_J^2 - \|\sqrt{n}(\theta_{Z_n} - \theta_0) - Z_n\|_J^2 \\ &= 2n \left\{ \tilde{L}_n(\theta_{Z_n}) - \tilde{L}_n(\hat{\theta}_n) \right\} + 2n \left\{ R_n(\hat{\theta}_n) - R_n(\theta_{Z_n}) \right\} \\ &\leq 2n \left\{ R_n(\hat{\theta}_n) - R_n(\theta_{Z_n}) \right\} = o_P(1). \end{aligned}$$

By (7.10) it follows that

$$\|\sqrt{n}(\hat{\theta}_n - \theta_0) - Z_n\|_J^2 - \|Z_n^{\mathcal{C}} - Z_n\|_J^2 = o_P(1).$$

In view of (2.24) we have

$$\begin{aligned} \|\sqrt{n}(\hat{\theta}_n - \theta_0) - Z_n\|_J^2 &= \|\sqrt{n}(\hat{\theta}_n - \theta_0) - Z_n^{\mathcal{C}}\|_J^2 + \|Z_n^{\mathcal{C}} - Z_n\|_J^2 \\ &\quad + 2 \left\langle \sqrt{n}(\hat{\theta}_n - \theta_0) - Z_n^{\mathcal{C}}, Z_n^{\mathcal{C}} - Z_n \right\rangle_J \\ &\geq \|\sqrt{n}(\hat{\theta}_n - \theta_0) - Z_n^{\mathcal{C}}\|_J^2 + \|Z_n^{\mathcal{C}} - Z_n\|_J^2. \end{aligned}$$

We thus obtain

$$\|\sqrt{n}(\hat{\theta}_n - \theta_0) - Z_n^{\mathcal{C}}\|_J^2 \leq \|\sqrt{n}(\hat{\theta}_n - \theta_0) - Z_n\|_J^2 - \|Z_n^{\mathcal{C}} - Z_n\|_J^2 = o_P(1).$$

Noting that, by central limit theorem of Billingsley (1961),

$$Z_n \xrightarrow{d} Z \sim \mathcal{N}(0, J^{-1} I J^{-1}) \quad \text{as } n \rightarrow \infty, \quad (7.11)$$

we have $Z_n^{\mathcal{C}} \xrightarrow{d} Z^{\mathcal{C}}$ and the conclusion follows. \square

Lemma 7.1 Under the assumptions of Theorem 2.3, $R_n(\theta_{Z_n}) = o_P(n^{-1})$ and $R_n(\hat{\theta}_n) = o_P(n^{-1})$ as $n \rightarrow \infty$.

Proof. First consider the impact of the initial values on the second-order derivatives of the objective function. We have

$$\begin{aligned} & \sup_{\theta \in \Theta} \left\| \frac{\partial^2}{\partial \theta \partial \theta'} \tilde{L}_n(\theta) - \frac{\partial^2}{\partial \theta \partial \theta'} L_n(\theta) \right\| \\ & \leq \frac{1}{n} \sum_{t=s+1}^n \left\{ c_t + X_t \left(\frac{a_t}{\underline{\omega}} e_t + \frac{c_t}{\underline{\omega}} + \frac{a_t}{\underline{\omega}} d_t^2 + \frac{b_t}{\underline{\omega}} d_t \right) \right\} = o(1) \end{aligned} \quad (7.12)$$

almost surely, by (2.23). In view of (7.4), (7.12) and (7.8)-(7.9), as $n \rightarrow \infty$ we have

$$nR_n(\theta_n) = o_P \{ \sqrt{n}(\theta_n - \theta_0) \} + o_P \{ n \|\theta_n - \theta_0\|^2 \} \quad (7.13)$$

when $\theta_n - \theta_0 = o_P(1)$. Therefore

$$nR_n(\theta_n) = o_P(1) \quad \text{when} \quad \sqrt{n}(\theta_n - \theta_0) = O_P(1). \quad (7.14)$$

By definition of θ_{Z_n} , and since $\theta_0 \in \Theta$, we also have

$$\| \sqrt{n}(\theta_{Z_n} - \theta_0) - Z_n \|_J \leq \| Z_n \|_J.$$

The Minskowski inequality then entails that

$$\| \sqrt{n}(\theta_{Z_n} - \theta_0) \|_J \leq \| \sqrt{n}(\theta_{Z_n} - \theta_0) - Z_n \|_J + \| Z_n \|_J \leq 2 \| Z_n \|_J.$$

By (7.11), we have $\| Z_n \|_J = O_P(1)$, and thus $\sqrt{n}(\theta_{Z_n} - \theta_0) = O_P(1)$. In view of (7.14), this entails $nR_n(\theta_{Z_n}) = o_P(1)$.

It remains to show the second convergence. By definition of $\hat{\theta}_n$, we have

$$0 \leq 2n\tilde{L}_n(\hat{\theta}_n) - 2n\tilde{L}_n(\theta_0) = \| Z_n \|_J^2 - \left\| Z_n - \sqrt{n}(\hat{\theta}_n - \theta_0) \right\|_J^2 + 2nR_n(\hat{\theta}_n).$$

It follows that

$$\begin{aligned} \left\| \sqrt{n}(\hat{\theta}_n - \theta_0) \right\|_J^2 & \leq 2 \left(\left\| \sqrt{n}(\hat{\theta}_n - \theta_0) - Z_n \right\|_J^2 + \| Z_n \|_J^2 \right) \\ & \leq 4 \| Z_n \|_J^2 + 4nR_n(\hat{\theta}_n). \end{aligned}$$

The consistency of $\hat{\theta}_n$ and (7.13) entail that $nR_n(\hat{\theta}_n) = o_P \left(\left\| \sqrt{n}(\hat{\theta}_n - \theta_0) \right\|_J^2 \right)$. It follows that $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_P(1)$, and the conclusion comes from (7.14). \square

Proof of Corollary 2.1. Note that, because $\lambda_t(\theta_0)$ is fixed, we have $\mu_2 = E(X_t^2 | X_u, u < t)$.

For notational convenience, write the information matrices in the case $q = 3$. We have

$$J = \frac{1}{\omega_0} \begin{pmatrix} 1 & \omega_0 & \omega_0 & \omega_0 \\ \omega_0 & \mu_2 & \omega_0^2 & \omega_0^2 \\ \omega_0 & \omega_0^2 & \mu_2 & \omega_0^2 \\ \omega_0 & \omega_0^2 & \omega_0^2 & \mu_2 \end{pmatrix} \quad \text{and} \quad I = \frac{\mu_2 - \omega_0^2}{\omega_0} J.$$

We thus obtain

$$\Sigma = \begin{pmatrix} (q-1)\omega_0^2 + \mu_2 & -\omega_0 & \cdots & -\omega_0 \\ & -\omega_0 & & \\ & \vdots & I_q & \\ & -\omega_0 & & \end{pmatrix}.$$

By the arguments given in Section 8.3.4 of FZ, the conclusion then follows from Theorem 2.3 and Remarks 2.3 and 2.4. \square

Sketch of proof of (3.3). The detailed proof, which is long and tedious, is available from the authors. We only give here the main computations showing that the condition (3.3) is necessary for the existence of EX_t^4 . Let $\epsilon_t = X_t - \lambda_t$. Note that for all $r \geq 2$, $EX_t^r < \infty$ iff $E|\epsilon_t|^r < \infty$. We have $E(\epsilon_t^2 | \mathcal{F}_t) = \lambda_t + \lambda_t^2/r$. Therefore $EX_t^2 < \infty$ iff $E\lambda_t^2 < \infty$. Writing $\lambda_t = \omega_0 + \alpha_0\epsilon_{t-1} + (\alpha_0 + \beta_0)\lambda_{t-1}$, and assuming that λ_t is stationary with second-order moments, we obtain

$$\begin{aligned} E\lambda_t^2 &= \omega_0^2 + \alpha_0^2 E\epsilon_{t-1}^2 + (\alpha_0 + \beta_0)^2 E\lambda_{t-1}^2 + 2\omega_0(\alpha_0 + \beta_0)E\lambda_{t-1} \\ &= \left\{ \frac{\alpha_0^2}{r} + (\alpha_0 + \beta_0)^2 \right\} E\lambda_t^2 + K, \end{aligned}$$

where, here and in the sequel, K denotes a generic positive constant whose value is unimportant. Therefore $EX_t^2 < \infty$ entails (3.2), which was already known from Christou and Fokianos (2013). Now, from the expression of the centred third and fourth order moments of the negative binomial distribution, we have

$$E(\epsilon_t^3 | \mathcal{F}_t) = 2\frac{\lambda_t^3}{r^2} + R_t^{(2)}, \quad E(\epsilon_t^4 | \mathcal{F}_t) = (6 + 3r)\frac{\lambda_t^4}{r^3} + R_t^{(3)},$$

where, for $i = 2, 3$, $R_t^{(i)}$ is a polynomial in λ_t of degree i with positive coefficients. Therefore $EX_t^4 < \infty$ iff $E\lambda_t^4 < \infty$ and, after some tedious computations,

$$E\lambda_t^4 = \alpha_0^4 E\epsilon_{t-1}^4 + 8\alpha_0^3(\alpha_0 + \beta_0)\frac{E\lambda_{t-1}^4}{r^2} + 6\alpha_0^2(\alpha_0 + \beta_0)^2\frac{E\lambda_{t-1}^4}{r} + (\alpha_0 + \beta_0)^4 E\lambda_{t-1}^4 + K,$$

and the conclusion follows. \square

References

- Andersson, J. and Karlis, D. (2014). A parametric time series model with covariates for integers in *Z. Statistical Modelling*, 14(2):135–156.
- Andrews, D. W. (1999). Estimation when a parameter is on a boundary. *Econometrica*, 67(6):1341–1383.
- Billingsley, P. (1961). The lindeberg-levy theorem for martingales. In *Proc. Amer. Math. Soc.*, volume 12, pages 88–792.
- Billingsley, P. (2008). *Probability and measure*. John Wiley & Sons.
- Christou, V. (2013). *Statistical theory for mixed poisson time series models*. PhD thesis, University of Cyprus.
- Christou, V. and Fokianos, K. (2013). Quasi-likelihood inference for negative binomial time series models. *Journal of Time Series Analysis*, (*wileyonlinelibrary.com*) DOI: 10.1111/jtsa.12050.
- Cox, D. R., Gudmundsson, G., Lindgren, G., Bondesson, L., Harsaae, E., Laake, P., Juselius, K., and Lauritzen, S. L. (1981). Statistical analysis of time series: Some recent developments [with discussion and reply]. *Scandinavian Journal of Statistics*, pages 93–115.
- Davis, R. A. and Liu, H. (2012). Theory and inference for a class of observation-driven models with application to time series of counts. *arXiv preprint arXiv:1204.3915*. <http://adsabs.harvard.edu/abs/2012arXiv1204.3915D>.
- Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association*, 81(395):709–721.
- Ferland, R., Latour, A., and Oraichi, D. (2006). Integer-valued GARCH process. *Journal of Time Series Analysis*, 27(6):923–942.
- Fokianos, K. (2012). Count time series models. *Handbook in statistics. Time series analysis: Methods and Applications*, 30:315–348.
- Fokianos, K., Rahbek, A., and Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association*, 104(488):1430–1439.

- Fokianos, K. and Tjøstheim, D. (2011). Log-linear poisson autoregression. *Journal of Multivariate Analysis*, 102(3):563–578.
- Francq, C., Roy, R., and Zakoïan, J.-M. (2005). Diagnostic checking in ARMA models with uncorrelated errors. *Journal of the American Statistical Association*, 100(470):532–544.
- Francq, C. and Zakoïan, J.-M. (2009). *A tour in the asymptotic theory of GARCH estimation*. Edts: T. G. Andersen, R.A. Davis, J-P. Kreiss, T. Mikosch. Springer Statistics.
- Francq, C. and Zakoïan, J.-M. (2010). *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons.
- Gourieroux, C., Monfort, A., and Trognon, A. (1984). Pseudo maximum likelihood methods: Theory. *Econometrica*, 52(3):681–700.
- Heinen, A. (2003). Modelling time series count data: an autoregressive conditional poisson model. *Available at SSRN 1117187*.
- Kachour, M. and Truquet, L. (2011). A p-order signed integer-valued autoregressive (SINAR(p)) model. *Journal of Time Series Analysis*, 32(3):223–236.
- Liu, H. (2012). *Some models for time series of counts*. PhD thesis, Columbia university.
- McCullagh, P. (1983). Quasi-likelihood functions. *The Annals of Statistics*, pages 59–67.
- Neumann, M. H. et al. (2011). Absolute regularity and ergodicity of poisson count processes. *Bernoulli*, 17(4):1268–1284.
- Tjøstheim, D. (2012). Some recent theory for autoregressive count time series. *Test*, 21(3):413–438.
- Tjøstheim, D. (2014). *Statistical analysis of count time series models*. Springer.
- Wedderburn, R. W. (1974). Quasi-likelihood functions, generalized linear models, and the gauss-newton method. *Biometrika*, 61(3):439–447.
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25.
- Zhu, F. (2011). A negative binomial integer-valued GARCH model. *Journal of Time Series Analysis*, 32(1):54–67.