



Munich Personal RePEc Archive

Negative binomial quasi-likelihood inference for general integer-valued time series models

Aknouche, Abdelhakim and Bendjedou, Sara and Touche,
Nassim

Faculty of Mathematics, USTHB, Algeria, Mathematics
Department, Qassim University, Saudi Arabia, Department of
Operational Research, Faculty of Exact Sciences, University of
Bejaia, Bejaia, Algeria

6 December 2016

Online at <https://mpra.ub.uni-muenchen.de/83082/>
MPRA Paper No. 83082, posted 04 Dec 2017 07:19 UTC

Negative binomial quasi-likelihood inference for general integer-valued time series models

Abdelhakim Aknouche^{*◦}, Sara Bendjeddou^{*} and Nassim Touche^{*†}

December 2, 2017

Abstract

Two *negative binomial quasi-maximum likelihood estimates* (NB-QMLE's) for a general class of count time series models are proposed. The first one is the *profile* NB-QMLE calculated while arbitrarily fixing the dispersion parameter of the negative binomial likelihood. The second one, termed *two-stage* NB-QMLE, consists of four stages estimating both conditional mean and dispersion parameters. It is shown that the two estimates are consistent and asymptotically Gaussian under mild conditions. Moreover, the two-stage NB-QMLE enjoys a certain asymptotic efficiency property provided that a negative binomial *link function* relating the conditional mean and conditional variance is specified. The proposed NB-QMLE's are compared with the Poisson QMLE asymptotically and in finite samples for various well-known particular classes of count time series models such as the (Poisson and negative binomial) Integer-valued GARCH model and the INAR(1) model. Applications to two real datasets are given.

^{*}Faculty of Mathematics, University of Science and Technology Houari Boumediene. [◦]Mathematics department, Qassim University.

[†]This material is a long version of a forthcoming paper (Aknouche et al, 2017) to be published in *Journal of Time Series Analysis*.

Keywords and phrases: Integer-valued time series models, Integer-valued GARCH, Integer-valued AR, Generalized Linear Models, Quasi-likelihood, Geometric QMLE, Negative Binomial QMLE, Poisson QMLE, consistency and asymptotic normality.

1. Introduction

Integer-valued time series such as count and binary data are well observed in a broad range of applications (e.g. economics, finance, epidemiology, medicine, telecommunications...). They are characterized by some stylized facts such as small values, overfrequency of zeros, locally constant behavior, overdispersion, positive autocorrelation structure, and asymmetric marginal distributions (see e.g. Kedem and Fokianos, 2002; McKenzie, 2003; Fokianos, 2012; Cameron and Trivedi, 2013; Silva, 2015; Davis et al, 2016). It is well documented that continuous-valued time series models such as ARMA-like processes are inappropriate for modeling such integer-valued series. This is why considerable interest has been paid in recent decades to alternative integer-valued time series models. Numerous models have been introduced so it appears difficult to classify them. However, two major classes of integer-valued models have played a central role. The first one is the class of models based on integer-valued regressions such as *generalized* ARMA (GARMA) models, Poisson autoregression and especially Poisson *Integer Generalized Conditional Heteroskedastic* (INGARCH) models (e.g. Grunwald et al, 2000; Rydberg and Shephard, 2000; Benjamin et al, 2003; Heimen, 2003; Ferland et al, 2006; Fokianos et al, 2009; Zhu, 2011-2012a-2012b; Doukhan et al, 2012; Christou and Fokianos, 2014; Davis and Liu, 2016; Chen et al, 2016). The second class, however, concerns stochastic difference equations involving the *thinning* operator where the best known example is the *INteger* AR (INAR) model (e.g. McKenzie, 1985-2003; Al-Osh and Alzaid, 1987; Silva, 2015; Bourguignon, 2016).

Ahmad and Francq (2016) recently introduced a more general integer-valued time series model that encompasses many models of the two aforementioned classes. This model is defined through specifying its conditional mean as a measurable parametric function of

the infinite past of the observed process. Important subclasses of this model are the general *Poisson autoregression* (Doukhan et al, 2012; Doukhan and Kengne, 2015; Kengne, 2015), the INGARCH model and the INAR model. Ahmad and Francq (2016) established consistency and asymptotic normality of the *Poisson quasi-maximum likelihood estimate* (P-QMLE), which is calculated as if the conditional distribution of the model were Poissonian. The P-QMLE has in fact many advantages: i) first, it is robust to misspecification of the true conditional distribution whenever the conditional expectation is well specified. This is due to the fact that the Poisson likelihood belongs to the linear exponential family (White, 1982; Gouriéroux et al, 1984a). ii) Second, it is asymptotically efficient when the true conditional distribution of the model is Poissonian. iii) Third, when the conditional variance and conditional mean of the model are proportional, the P-QMLE is asymptotically efficient in the class of all QMLE's whose likelihood belongs to the linear exponential family (see Gouriéroux et al, 1984a). The latter proportionality between the conditional mean and conditional variance is usually called the Poisson *Generalized Linear Model* (henceforth GLM) variance *assumption* (or *link function*). Despite these advantages, it is known that the Poisson model is less flexible in modeling overdispersed series than a model based on an overdispersed conditional distribution such as the negative binomial one (e.g. Christou and Fokianos, 2014; Zhu, 2011). Therefore, it is likely that the P-QMLE does not reach its full asymptotic efficiency in the presence of overdispersed data which are frequently observed in practice. Thus a QMLE calculated using an overdispersed likelihood while belonging to the linear exponential family would be an interesting complementary to the P-QMLE.

For the model considered by Ahmad and Francq (2016), we propose two variants of the *negative binomial* QMLE (NB-QMLE). These estimates are calculated on the basis of the negative binomial likelihood, belonging to the linear exponential family. The first one we call "profile NB-QMLE" (pNB-QMLE) consists in maximizing the negative binomial likelihood over the conditional mean parameter letting the corresponding dispersion parameter arbitrarily fixed. In particular, when the latter parameter equals one, the resulting estimate reduces to the *geometric* QMLE (Aknouche and Bendjeddou, 2017). The second one,

however, consists of four stages: a two-stage NB-QMLE to estimate the conditional mean parameter of the model and a two-stage weighted least squares estimate for the dispersion parameter. For this, the underlying model should satisfy a negative binomial GLM *link function* involving the unknown dispersion parameter to be estimated. In the context of static integer-valued regression, a similar three-stage estimate was termed "*quasi-generalized pseudo-maximum likelihood estimate*" by Gouriéroux et al (1984b) and "*two-stage negative binomial quasi-maximum likelihood estimate*" (2SNB-QMLE) by Wooldridge (1997). Adopting the latter notation, the four-stage estimate we propose will be denoted by 2SNB-QMLE. It will be shown under some mild assumptions that the two proposed estimates are consistent and asymptotically Gaussian without fully specifying the conditional distribution of the model. Moreover, under the negative binomial GLM link function, the 2SNB-QMLE is asymptotically efficient in the class of all QMLE's belonging to the linear exponential family, including the P-QMLE.

The rest of this paper is outlined as follows. Section 2 presents the general integer-valued time series model we deal with and some of its specific cases. Section 3 defines some negative binomial QML criteria and establishes consistency and asymptotic normality of the pNB-QMLE and the 2SNB-QMLE. As a result, Section 4 compares the asymptotic variance of the proposed NB-QMLE's with that of the P-QMLE under some specific GLM assumptions as well as on particular classes of the general model. In particular, the *Poisson INGARCH* model, the *negative binomial INGARCH* model, the INAR(1) model, the Double Poisson INGARCH model and the Generalized Poisson INGARCH model are examined. Moreover, these estimates are compared in finite samples via some simulation experiments. Application to the number of poliomyelitis cases in the United States (Polio data, Zeger, 1988) and the number of transactions of the Ericsson *B* stock (Transaction data, Fokianos et al, 2009; Christou and Fokianos, 2014) under the negative binomial INGARCH framework are considered. Section 6 concludes while proofs of the main results are left to Section 7.

In what follows, we heavily use the following notations and conventions: All random variables and sequences we consider are defined on a probability space (Ω, \mathcal{F}, P) . The sym-

bols $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$, $\mathbb{N} = \{0, 1, \dots\}$ and $\mathbb{N}^* = \mathbb{N}/\{0\}$ denote respectively the set of integers, the set of nonnegative integers and the set of positive integers. The notation $Y \sim \mathcal{P}(\mu)$ means that the random variable Y has a Poisson distribution with parameter $\mu > 0$. Similarly, $X \sim \mathcal{NB}(r, p)$ means that X has the negative binomial distribution (also called mixture Poisson-Gamma distribution). This distribution is given for any $x \in \mathbb{N}$ by $f_X(x) := P(X = x) = \frac{\Gamma(x+r)}{x!\Gamma(r)} p^r (1-p)^x$, where $r > 0$ is a positive real number called the dispersion parameter, $p \in (0, 1)$ is a probability parameter, Γ is the gamma function and $x!$ is the factorial of x . When $r \in \mathbb{N}^*$ has to be a positive integer, the factor $\frac{\Gamma(x+r)}{x!\Gamma(r)}$ may be replaced by the binomial coefficient $\binom{x+r-1}{x}$. In particular, when $r = 1$ we find the geometric distribution and we simply write $X \sim \mathcal{G}(p)$. Following Cameron and Trivedi (1986, 2013), the negative binomial $_K$ (NB_K) conditional distribution given a σ -algebra $\mathcal{B} \subset \mathcal{F}$ is defined by $X|\mathcal{B} \sim \mathcal{NB}\left(r\lambda^{2-K}, \frac{r\lambda^{2-K}}{r\lambda^{2-K}+\lambda}\right)$ where $\lambda = E(X|\mathcal{B})$ and $r > 0$. Two important cases of the latter model are the negative binomial $_1$ (NB_1) conditional distribution corresponding to $K = 1$ and the negative binomial $_2$ (NB_2) model for which $K = 2$. Finally, the symbols $\xrightarrow[n \rightarrow \infty]{a.s.}$, $\xrightarrow[n \rightarrow \infty]{p}$ and $\xrightarrow[n \rightarrow \infty]{\mathcal{L}}$ denote respectively almost sure convergence, convergence in probability and convergence in distribution as $n \rightarrow \infty$ while $o_p(1)$ and $o_{a.s.}(1)$ are respectively: a term converging in probability to zero and a term converging almost surely (*a.s.*) to zero as $n \rightarrow \infty$.

2. A general class of count time series models

Let $\theta_0 \in \Theta \subset \mathbb{R}^m$ ($m \in \mathbb{N}^*$) be an unknown "true" parameter and consider a measurable positive real-valued function $\lambda : \mathbb{N}^\infty \times \Theta \rightarrow (0, \infty)$. A general class of count time series models, as proposed by Ahmad and Francq (2016), is given through an observable integer-valued stochastic process $\{X_t, t \in \mathbb{Z}\}$, which is defined on (Ω, \mathcal{F}, P) with conditional expectation specified as follows

$$E(X_t | \mathcal{F}_{t-1}) = \lambda(X_{t-1}, X_{t-2}, \dots; \theta_0) := \lambda_t(\theta_0) := \lambda_t, \quad t \in \mathbb{Z}, \quad (2.1)$$

where $\mathcal{F}_t \subset \mathcal{F}$ is the σ -algebra generated by $\{X_t, X_{t-1}, \dots\}$. Without any constraints on (2.1), any discrete-time stochastic process satisfies (2.1). However, two restrictions are made here. The first one is that $\{X_t, t \in \mathbb{Z}\}$ is restricted to be integer-valued, i.e. with sample space $\{0, 1, 2, \dots\}$. The second one is that the function λ depends on the parameter θ_0 so (2.1) is a parametric model. Letting

$$e_t := e_t(\theta_0) = X_t - E(X_t | \mathcal{F}_{t-1}),$$

model (2.1), which is defined through the conditional mean representation (2.1), may also be written in the following stochastic difference equation (or in innovation form, cf. Grunwald et al, 2000)

$$X_t = \lambda(X_{t-1}, X_{t-2}, \dots; \theta_0) + e_t, \quad t \in \mathbb{Z}. \quad (2.2)$$

Equation (2.2), which is driven by the $\{\mathcal{F}_t, t \in \mathbb{Z}\}$ -martingale difference $\{e_t, t \in \mathbb{Z}\}$, appears to be an infinite nonlinear autoregression with an integer-valued solution $\{X_t, t \in \mathbb{Z}\}$. In fact, model (2.1)-(2.2) is very general and encompasses many important classes of integer-valued time series models such as the (stable) *Poisson* INGARCH model (Grunwald et al, 2000; Rydberg and Shephard, 2000; Heinen, 2003; Ferland et al, 2006), the general Poisson autoregression (Doukhan et al, 2012; Doukhan and Kengne, 2015; Kengne, 2015), the stable negative binomial₂ INGARCH model (Zhu, 2011; Christou and Fokianos, 2014; Davis and Liu, 2016; Diop and Kengne, 2017) and the INAR model (McKenzie, 1985; Al-Osh and Alzaid, 1987).

Note that the generality of model (2.1) stems not only from the general form of the function $\lambda(\cdot)$, but also from the fact that apart from the conditional mean, no other specification concerning the conditional distribution of the process $\{X_t, t \in \mathbb{N}\}$ is required. However, it is sometimes important to specify a *link function* relating the conditional variance and the conditional mean of model (2.1), i.e.

$$\text{Var}(X_t | \mathcal{F}_{t-1}) = l(E(X_t | \mathcal{F}_{t-1})), \quad (2.3)$$

where $l : (0, \infty) \rightarrow (0, \infty)$ is a positive real function. In the literature on *generalized linear models* (e.g. Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989), such a link func-

tion is also called the GLM *nominal variance assumption* and is induced either by the conditional distribution of the model when it is fully specified or by the structure of the model. For example, when the conditional distribution corresponding to (2.1) is Poissonian with parameter λ_t , which reduces to a special case of the general Poisson autoregression proposed by Doukhan et al (2012), the Poisson GLM link function for model (2.1) is given by the linear form $l(x) = x$. A more general linear link function $l(x) = \left(1 + \frac{1}{r_0}\right)x$, for some $r_0 > 0$, is induced by the negative binomial₁ conditional distribution, i.e. $X_t | \mathcal{F}_{t-1} \sim \mathcal{NB}\left(r_0 \lambda_t, \frac{r_0 \lambda_t}{r_0 \lambda_t + \lambda_t}\right)$, $r_0 > 0$ (see Cameron and Trivedi, 1986 and Section 4.1 below). Furthermore, the link function implied by the negative binomial₂ conditional distribution, that is $\mathcal{NB}\left(r_0, \frac{r_0}{r_0 + \lambda_t}\right)$, is given by

$$l(x) = x \left(1 + x \frac{1}{r_0}\right), \quad r_0 > 0. \quad (2.4)$$

When $r_0 = 1$, we find the link function corresponding to the Geometric distribution. On the other hand, a link function may be obtained even when the conditional distribution of the model is misspecified. In Section 4.1.4 we will see that the GLM link function for the INAR(1) model is always an affine function regardless of the conditional distribution of this model.

In this paper we are interested in estimating the unknown conditional mean parameter θ_0 using a series X_1, X_2, \dots, X_n ($n \in \mathbb{N}^*$) generated from (2.1). When a negative binomial₂ link function such as (2.3)-(2.4) is specified we are also interested in estimating the dispersion parameter r_0 . In fact, two instances of (2.1) are considered:

Case 1: Only the conditional mean (2.1) is specified so that we only have to estimate the conditional mean parameter θ_0 .

Case 2: Equation (2.1) and the NB₂ variance GLM assumption (2.3)-(2.4) are both specified so we have to estimate both θ_0 and r_0 .

A particularly important instance of **Case 2** appears when the full conditional distribution of the model is specified as a negative binomial₂ one, i.e. $X_t | \mathcal{F}_{t-1} \sim \mathcal{NB}\left(r_0, \frac{r_0}{r_0 + \lambda_t}\right)$, where a special case is the NB₂-INGARCH model (see Davis and Liu, 2016; Zhu, 2011; Christou and Fokianos, 2014-2015 and Section 4.1.3 below).

For our estimation purposes we make the following regularity assumption on (2.1).

A0 *The process $\{X_t, t \in \mathbb{Z}\}$ given by (2.1) is strictly stationary and ergodic.*

For some particular classes of (2.1) such as the INGARCH and INAR models, assumption **A0** may be expressed more explicitly as a stability condition on θ_0 (see Ahmad and Francq, 2016 and Section 4.1 below). Furthermore, when the conditional distribution of (2.1) is Poissonian, Doukhan et al (2012) provided general conditions on the function λ in (2.1) for strict stationarity, ergodicity and weak dependence of the model. Their results were based on Doukhan and Wintenberger (2008).

Now, given a generic parameter $\theta \in \Theta$, the conditional mean function given by

$$\lambda(X_{t-1}, X_{t-2}, \dots; \theta) := \lambda_t(\theta), \quad t \in \mathbb{N},$$

clearly coincides with the conditional mean in (2.1) when $\theta = \theta_0$. It is unobservable because of the unobservable values $X_0, X_{-1}, X_{-2}, \dots$. For any arbitrary fixed initial values $\tilde{X}_0, \tilde{X}_{-1}, \tilde{X}_{-2}, \dots$, let

$$\tilde{\lambda}_t(\theta) = \lambda\left(X_{t-1}, X_{t-2}, \dots, X_1, \tilde{X}_0, \tilde{X}_{-1}, \dots; \theta\right), \quad t \in \mathbb{N}^*,$$

be an observable proxy for $\lambda_t(\theta)$. The latter approximation serves in calculating various QMLE-type of θ_0 we intend to study below.

3. Negative binomial QMLE's

This Section considers two negative binomial QMLE's of model (2.1) given a realization X_1, \dots, X_n thereof. To describe these estimates consider **Case 2** of model (2.1)-(2.4) with unknown parameters θ_0 and r_0 . For any generic $\theta \in \Theta$ and $r > 0$, the negative binomial (log) likelihood, $\tilde{L}_{NB}(\theta, r)$, based on the NB₂ conditional distribution, $\mathcal{NB}\left(r, \frac{r}{r+\lambda_t(\theta)}\right)$, is given by

$$\begin{aligned} \tilde{L}_{NB}(\theta, r) &= \frac{1}{n} \sum_{t=1}^n \tilde{l}_t(\theta, r), \\ \text{with } \tilde{l}_t(\theta, r) &= r \log\left(\frac{r}{r+\tilde{\lambda}_t(\theta)}\right) + X_t \log\left(\frac{\tilde{\lambda}_t(\theta)}{r+\tilde{\lambda}_t(\theta)}\right) + \frac{\Gamma(X_t+r)}{X_t! \Gamma(r)}. \end{aligned} \tag{3.1}$$

A negative binomial quasi-maximum likelihood estimate (NB-QMLE) of (θ_0, r_0) is a maximizer of $\tilde{L}_{NB}(\theta, r)$ over $\theta \in \Theta$ and $r > 0$.

Note, however, that $\tilde{l}_t(\theta, r)$ given by (3.1) is not a member of the linear exponential family in the sense of Gourieroux et al (1984a). So any maximizer of (3.1) might be inconsistent under misspecification of the true conditional distribution of model (2.1), which constitutes a serious limitation. In lieu of maximizing directly (3.1) and picking up the estimate component corresponding to θ_0 , we may consider a four-stage approach which is rather robust to misspecification of the true conditional distribution and which consists in:

i) Fixing r in (3.1) arbitrarily to any known positive number, say $r^* > 0$, and estimating θ_0 while maximizing (3.1) with respect to θ , giving a first-step QMLE $\hat{\theta}_{r^*}$.

ii) Estimating r_0 under the GLM link function (2.3)-(2.4) using a weighted least squares estimate \hat{r}_1 while replacing θ_0 in the weight by its QMLE, $\hat{\theta}_{r^*}$, obtained in *i*).

iii) Re-estimating θ_0 by maximizing a variation of (3.1) obtained while replacing r by the estimate \hat{r}_1 obtained in *ii*), giving $\hat{\theta}_{\hat{r}_1}$.

iv) Re-estimating r_0 using the same weighted least squares method in *ii*) but while replacing θ_0 by $\hat{\theta}_{\hat{r}_1}$ obtained in *iii*).

For a similar approach in the context of static count regression, see Gourieroux et al (1984a, 1984b) and Wooldridge (1997, 2002). In the above first and third steps, maximization of (3.1) is carried out with respect to θ letting r fixed. So the last term in (3.1) may be left out and (3.1) is simply replaced by the following "*profile negative binomial likelihood*"

$$\tilde{L}_{n,r}(\theta) = \frac{1}{n} \sum_{t=1}^n \tilde{l}_{t,r}(\theta) \quad \text{with} \quad \tilde{l}_{t,r}(\theta) = r \log \left(\frac{r}{r + \tilde{\lambda}_t(\theta)} \right) + X_t \log \left(\frac{\tilde{\lambda}_t(\theta)}{r + \tilde{\lambda}_t(\theta)} \right). \quad (3.2)$$

It should be noted that $\tilde{l}_{t,r}(\theta)$ in (3.2) rather belongs to the linear exponential family. Therefore any maximizer of (3.2) with respect to θ would be robust to misspecification of the conditional distribution, whenever correctly specifying the conditional mean such as (2.1). It turns out that for any fixed $r > 0$, $\tilde{L}_{n,r}(\theta)$ is the Wedderburn quasi-likelihood function (Wedderburn, 1974) based on the NB₂ variance GLM assumption (2.3)-(2.4) (with r in place of r_0).

On the other hand, if we consider **Case 1** of model (2.1) where only the conditional mean is specified, then only θ_0 has to be estimated and r in (3.1) can be set to any positive real value. So maximization of (3.1) will only be done with respect to θ , which again amounts to maximizing (3.2). In summary, for both **Case 1** and **Case 2**, we have to maximize the profile (or Quasi-) likelihood (3.2) with respect to θ .

In the rest of this Section we shall study asymptotics of two QML-type estimates that maximize (3.2) over $\theta \in \Theta$. Section 3.1 examines consistency and asymptotic normality of a maximizer of (3.2) for arbitrarily fixed $r > 0$. The resulting estimate will be called *profile* (or *marginal*) NB-QMLE (pNB-QMLE). In Section 3.2, consistency and asymptotic normality of the four-stage estimate (see *i-iv*) above) are established assuming the NB₂ variance GLM assumption (2.3)-(2.4) for an unknown $r_0 > 0$.

3.1. Profile negative binomial QMLE

Consider **Case 1** of model (2.1). A *profile negative binomial quasi-maximum likelihood estimate* (pNB-QMLE) of θ_0 is any measurable solution of the following problem

$$\hat{\theta}_r = \arg \max_{\theta \in \Theta} \left(\tilde{L}_{n,r}(\theta) \right), \quad (3.3)$$

for some Θ and some fixed known $r > 0$, where $\tilde{L}_{n,r}(\theta)$ is given by (3.2). When $r = 1$, $\hat{\theta}_1$ reduces to the geometric QMLE (G-QMLE) studied by Aknouche and Bendjeddou (2017). The choice of $(\tilde{X}_0, \tilde{X}_{-1}, \dots)$ is of no asymptotic importance, but may influence the accuracy of estimate in finite samples. In general, one assumes that $\tilde{X}_0 = x, \tilde{X}_{-1} = x, \dots$ with x depending on the function λ or on the observations (see Ahmad and Francq, 2016). To study consistency of the pNB-QMLE, $\hat{\theta}_r$, we need the following assumptions:

A1 $\theta \mapsto \lambda_t(\theta)$ is a.s. continuous; $\lambda_t(\theta) > c$ and $\tilde{\lambda}_t(\theta) > c$, a.s. for some $c > 0$.

A2 $a_t \xrightarrow{a.s.} 0$ and $a_t X_t \xrightarrow{a.s.} 0$ where $a_t = \sup_{\theta \in \Theta} \left| \tilde{\lambda}_t(\theta) - \lambda_t(\theta) \right|$.

A3 $E(X_t^\delta) < \infty$ for some $\delta > 1$.

A4 $\lambda_t(\theta) = \lambda_t(\theta_0)$ a.s. if and only if $\theta = \theta_0$.

A5 Θ is compact.

Assumptions **A1-A5** are standard and may be made more explicit for some particular models of (2.1) (cf. Section 4.1). Similar assumptions were considered by Ahmad and Franco (2016) for the strong consistency of their P-QMLE.

Theorem 3.1 *Under (2.1) and **A0-A5**,*

$$\widehat{\theta}_r \xrightarrow[n \rightarrow \infty]{a.s.} \theta_0, \quad \text{for all } r > 0. \quad (3.4)$$

The latter result shows that, like the P-QMLE, the pNB-QMLE is robust to misspecification of the true conditional distribution where only (2.1) has to be specified. This is not surprising as the profile negative binomial log-likelihood (3.2) belongs to the linear exponential family (see *Gourieroux et al, 1984a*).

We now examine the asymptotic normality of the pNB-QMLE. Let $l_{t,r}(\theta)$ be defined in the same way as $\widetilde{l}_{t,r}(\theta)$ in (3.2) with $\lambda_t(\theta)$ in place of $\widetilde{\lambda}_t(\theta)$ and set $L_{n,r}(\theta) = \frac{1}{n} \sum_{t=1}^n l_{t,r}(\theta)$. Consider the following supplementary assumptions.

A6 *The variables $c_t, c_t X_t, a_t d_t, a_t d_t X_t$ and $b_t d_t X_t$ are of order $O(t^{-\tau})$ a.s. for some $\tau > 1/2$, where $b_t = \sup_{\theta \in \Theta} \left| \widetilde{\lambda}_t^2(\theta) - \lambda_t^2(\theta) \right|$, $c_t = \sup_{\theta \in \Theta} \left\| \frac{\partial(\widetilde{\lambda}_t(\theta) - \lambda_t(\theta))}{\partial \theta} \right\|$ and*

$$d_t = \sup_{\theta \in \Theta} \max \left(\left\| \frac{1}{\widetilde{\lambda}_t(\theta)(r + \widetilde{\lambda}_t(\theta))} \frac{\partial \widetilde{\lambda}_t(\theta)}{\partial \theta} \right\|, \left\| \frac{1}{\lambda_t(\theta)(r + \lambda_t(\theta))} \frac{\partial \lambda_t(\theta)}{\partial \theta} \right\| \right).$$

A7 *The true θ_0 belongs to the interior of Θ .*

A8 *The conditional variance $v_t(\theta_0) := \text{Var}(X_t | \mathcal{F}_{t-1}) = E(X_t^2 | \mathcal{F}_{t-1}) - \lambda_t^2(\theta_0)$ is a.s. finite.*

A9 *The derivatives $\frac{\partial^2 \lambda_t(\theta)}{\partial \theta \partial \theta'}$ and $\frac{\partial^2 \widetilde{\lambda}_t(\theta)}{\partial \theta \partial \theta'}$ exist and are continuous, the matrices*

$$I_r = E \left(\frac{v_t(\theta_0)}{\lambda_t^2(\theta_0)(r + \lambda_t(\theta_0))^2} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \lambda_t(\theta_0)}{\partial \theta'} \right) \quad \text{and} \quad J_r = E \left(\frac{1}{\lambda_t(\theta_0)(r + \lambda_t(\theta_0))} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \lambda_t(\theta_0)}{\partial \theta'} \right),$$

are finite, and J_r is nonsingular for all $r > 0$.

A10 *There is a neighborhood $V(\theta_0)$ of θ_0 such that $E \left(\sup_{\theta \in V(\theta_0)} \left\| \frac{\partial^2 l_{t,r}(\theta)}{\partial \theta \partial \theta'} \right\| \right) < \infty$ for all $r > 0$.*

Like consistency conditions, assumptions **A6-A10** may be made more explicit for specific cases of (2.1). Now we have the following asymptotic normality result.

Theorem 3.2 Under (2.1) and **A0-A10**,

$$\sqrt{n} \left(\widehat{\theta}_r - \theta_0 \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N \left(0, J_r^{-1} I_r J_r^{-1} \right) \quad \text{for all } r > 0. \quad (3.5)$$

Some remarks are in order:

- When the conditional distribution of the data generating process (2.1) is NB₂ with parameters r_0 and $\frac{r_0}{r_0 + \lambda_t}$, i.e. $X_t | \mathcal{F}_{t-1} \sim \mathcal{NB} \left(r_0, \frac{r_0}{r_0 + \lambda_t} \right)$, then (3.5) holds with $I_r = \frac{1}{r_0} E \left(\frac{r_0 + \lambda_t(\theta_0)}{\lambda_t(\theta_0)(r_0 + \lambda_t(\theta_0))^2} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \lambda_t(\theta_0)}{\partial \theta'} \right)$. In particular, when r in (3.2)-(3.3) coincides with the "true" r_0 in (2.3)-(2.4) then $I_{r_0} = \frac{1}{r_0} J_{r_0}$, so (3.5) becomes

$$\sqrt{n} \left(\widehat{\theta}_{r_0} - \theta_0 \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N \left(0, \frac{1}{r_0} J_{r_0}^{-1} \right). \quad (3.6)$$

- A weak result, which does not require specifying the full conditional distribution is that under the following more general *negative binomial₂ GLM link function*

$$\text{Var} (X_t | \mathcal{F}_{t-1}) = \delta^2 E (X_t | \mathcal{F}_{t-1}) \left(1 + \frac{1}{r_0} E (X_t | \mathcal{F}_{t-1}) \right) \quad \text{for some } \delta^2 > 0, r_0 > 0, \quad (3.7)$$

which generalizes (2.3)-(2.4), $\widehat{\theta}_{r_0}$ is asymptotically efficient in the class of all QMLE's in the linear exponential family (see e.g. Gourieroux et al (1984a, 1984b) and Wooldridge (1997) in the context of QML inference for static integer-valued regression models). In that case we have

$$\sqrt{n} \left(\widehat{\theta}_{r_0} - \theta_0 \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N \left(0, \delta^2 J_{r_0}^{-1} \right). \quad (3.8)$$

Note, however, that r_0 is generally unknown and (3.6) and (3.8) does not hold unless r_0 is consistently estimated under (3.7) as we will see in the following subsection.

Now an important issue is to estimate the asymptotic variance of the pNB-QMLE. Similarly to Ahmad and Francq (2016), a consistent estimate of the asymptotic variance $J_r^{-1} I_r J_r^{-1}$ of the pNB-QMLE, $\widehat{\theta}_r$, is $\widehat{J}_r^{-1} \widehat{I}_r \widehat{J}_r^{-1}$ with

$$\widehat{I}_r = \frac{1}{n} \sum_{t=1}^n \left(\frac{X_t - \widetilde{\lambda}_t(\widehat{\theta}_r)}{\widetilde{\lambda}_t(\widehat{\theta}_r)(r + \widetilde{\lambda}_t(\widehat{\theta}_r))} \right)^2 \frac{\partial \widetilde{\lambda}_t(\widehat{\theta}_r)}{\partial \theta} \frac{\partial \widetilde{\lambda}_t(\widehat{\theta}_r)}{\partial \theta'}. \quad (3.9)$$

$$\widehat{J}_r = \frac{1}{n} \sum_{t=1}^n \frac{1}{\widetilde{\lambda}_t(\widehat{\theta}_r)(r + \widetilde{\lambda}_t(\widehat{\theta}_r))} \frac{\partial \widetilde{\lambda}_t(\widehat{\theta}_r)}{\partial \theta} \frac{\partial \widetilde{\lambda}_t(\widehat{\theta}_r)}{\partial \theta'}. \quad (3.10)$$

3.2. Two-stage negative binomial QMLE

Consider **Case 2** of model (2.1)-(2.4) for which we study the aforementioned four-stage procedure i)-iv). Here, the second and fourth steps are described in more details. Under the GLM assumption (2.3)-(2.4), if we set

$$u_t = (X_t - \lambda_t)^2 - E((X_t - \lambda_t)^2 | \mathcal{F}_{t-1}) = (X_t - \lambda_t)^2 - \left(1 + \frac{1}{r_0} \lambda_t\right) \lambda_t, \quad (3.11a)$$

then $E(u_t | \mathcal{F}_{t-1}) = 0$ and

$$\frac{(X_t - \lambda_t(\theta_0))^2 - \lambda_t(\theta_0)}{\lambda_t^2(\theta_0)} = \gamma_0 + \frac{u_t}{\lambda_t^2(\theta_0)}, \quad (3.11b)$$

where $\gamma_0 = \frac{1}{r_0}$. Regression (3.11) is not ready to be used to estimate γ_0 since its regressand, $\frac{(X_t - \lambda_t(\theta_0))^2 - \lambda_t(\theta_0)}{\lambda_t^2(\theta_0)}$, depends on the unknown θ_0 and is then unobservable. If a consistent estimate of θ_0 , say $\hat{\theta}$, is available then we may form the following modified (observable-regressand) regression

$$\frac{(X_t - \hat{\lambda}_t)^2 - \hat{\lambda}_t}{\hat{\lambda}_t^2} = \gamma_0 + \frac{\hat{u}_t}{\hat{\lambda}_t^2}, \quad (3.12a)$$

where

$$\hat{u}_t = \left(X_t - \hat{\lambda}_t\right)^2 - \left(1 + \frac{1}{r_0} \hat{\lambda}_t\right) \hat{\lambda}_t \text{ and } \hat{\lambda}_t = \lambda_t(\hat{\theta}). \quad (3.12b)$$

From (3.12a) a consistent estimate of r_0 is \hat{r} , the inverse of the weighted least squares estimate $\hat{\gamma}$ of γ_0 given by

$$\hat{r} = \left(\frac{1}{n} \sum_{t=1}^n \frac{((X_t - \hat{\lambda}_t)^2 - \hat{\lambda}_t)}{\hat{\lambda}_t^2} \right)^{-1}, \quad \hat{\gamma} = \hat{r}^{-1}, \quad (3.13)$$

where $\hat{\lambda}_t = \tilde{\lambda}_t(\hat{\theta})$. Note that the estimate \hat{r} we use here is a dynamic adaptation of the estimate proposed by *Gourieroux et al (1984b)* in the context of static negative binomial regression. Now, with (3.13) the following algorithm summarizes the four-stage approach i)-iv) described above.

Algorithm 3.1 (Two-stage NB-QMLE)

Given a fixed known $r^ > 0$, the two-stage NB-QMLE of (θ_0, r_0) in (2.1)-(2.4) consists of a quadruple $(\hat{\theta}_{r^*}, \hat{r}_1, \hat{\theta}_{\hat{r}_1}, \hat{r}_2)$, which is described by the following steps:*

Step 1 *Set $\hat{\theta}_{r^*} = \arg \max_{\theta \in \Theta} \tilde{L}_{n,r^*}(\theta)$, a solution to the problem (3.3) while replacing r by r^* . Let $\hat{\lambda}_{1t} = \tilde{\lambda}_t(\hat{\theta}_{r^*})$, $(1 \leq t \leq n)$.*

Step 2 Set $\hat{\gamma}_1 = \frac{1}{n} \sum_{t=1}^n \frac{(X_t - \hat{\lambda}_{1t})^2 - \hat{\lambda}_{1t}}{\hat{\lambda}_{1t}^2}$ and $\hat{r}_1 = \hat{\gamma}_1^{-1}$.

Step 3 Let $\hat{\theta}_{\hat{r}_1} = \arg \max_{\theta \in \Theta} \tilde{L}_{n, \hat{r}_1}(\theta)$ be a solution of the problem (3.3) while replacing the generic r by \hat{r}_1 . Get $\hat{\lambda}_{2t} = \tilde{\lambda}_t(\hat{\theta}_{\hat{r}_1})$, $(1 \leq t \leq n)$.

Step 4 Set $\hat{\gamma}_2 = \frac{1}{n} \sum_{t=1}^n \frac{(X_t - \hat{\lambda}_{2t})^2 - \hat{\lambda}_{2t}}{\hat{\lambda}_{2t}^2}$ and $\hat{r}_2 = \hat{\gamma}_2^{-1}$.

To get asymptotic properties of the quadruple $(\hat{\theta}_{r^*}, \hat{r}_1, \hat{\theta}_{\hat{r}_1}, \hat{r}_2)$, note first that $\hat{\theta}_{r^*}$ is no other than the profile NB-QMLE proposed in Section 3.1 whose asymptotic properties were given by Theorem 3.1 and Theorem 3.2. So it remains to study the triple $(\hat{r}_1, \hat{\theta}_{\hat{r}_1}, \hat{r}_2)$, asymptotic properties of which are given by the following result.

Theorem 3.3 Under (2.1), (2.3)-(2.4) and **A0-A10**,

$$\hat{r}_1 \xrightarrow[n \rightarrow \infty]{a.s.} r_0, \quad (3.14a)$$

$$\sqrt{n}(\hat{\gamma}_1 - \gamma_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N\left(0, E\left(\frac{((X_t - \lambda_t(\theta_0))^2 - (\lambda_t(\theta_0) + \frac{1}{r_0} \lambda_t^2(\theta_0)))^2}{\lambda_t^4(\theta_0)}\right)\right), \hat{\gamma}_2 \stackrel{A.D.}{=} \hat{\gamma}_1, \quad (3.14b)$$

$$\hat{\theta}_{\hat{r}_1} \xrightarrow[n \rightarrow \infty]{a.s.} \theta_0, \quad (3.14c)$$

$$\sqrt{n}(\hat{\theta}_{\hat{r}_1} - \theta_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N\left(0, \frac{1}{r_0} J_{r_0}^{-1}\right), \quad (3.14d)$$

where $\stackrel{A.D.}{=}$ stands for equality in asymptotic distribution.

A few broad conclusions can be drawn.

- Strong consistency of $\hat{\theta}_{\hat{r}_1}$ directly follows from strong consistency of $\hat{\theta}_r$ (for all $r > 0$) and \hat{r}_1 .

- The third-step estimate $\hat{\theta}_{\hat{r}_1}$ is clearly more asymptotically efficient than the first-step estimate $\hat{\theta}_{r^*}$.

- No supplementary moment assumptions apart those required by **A0-A10** are needed for consistency and asymptotic normality of $\hat{\gamma}_1$. Other methods for estimating γ are available (e.g. Christou and Fokianos, 2014), but they may involve higher order moment conditions.

- Asymptotic distribution of \hat{r}_1 is a *reciprocal normal distribution*, which is bimodal and has no first moment.

- Since $\hat{\gamma}_1$ and $\hat{\gamma}_2$ have the same asymptotic distribution, **Step 4** is optional and may be left out. However, for finite-samples considerations, we keep it here because it allows to

re-estimate r_0 using $\widehat{\lambda}_{2t}$ and hence $\widehat{\theta}_{\widehat{r}_1}$, which is more asymptotically efficient than $\widehat{\theta}_{r^*}$ we used in **Step 2**.

- A consistent estimate of the asymptotic variance $\frac{1}{r_0} J_{r_0}^{-1}$ of the third-step estimate, $\widehat{\theta}_{\widehat{r}_1}$, is

$$\frac{1}{\widehat{r}_2} \widehat{J}_{\widehat{r}_2}^{-1}, \quad (3.15)$$

where $\widehat{J}_{\widehat{r}_2}$ is given by (3.10). Note that since here $I_r = J_r$, then (3.9) may also be used.

- A consistent estimate of the asymptotic variance of $\widehat{\gamma}_2$ in (3.14b) is

$$\frac{1}{n} \sum_{t=1}^n \frac{\left((X_t - \lambda_t(\widehat{\theta}_{\widehat{r}_1}))^2 - (\lambda_t(\widehat{\theta}_{\widehat{r}_1}) + \frac{1}{r_0} \lambda_t^2(\widehat{\theta}_{\widehat{r}_1})) \right)^2}{\lambda_t^4(\widehat{\theta}_{\widehat{r}_1})}. \quad (3.16)$$

- The outputs of the 2SNB-QMLE method are $\widehat{r}_2 = (\widehat{\gamma}_2)^{-1}$ and $\widehat{\theta}_{\widehat{r}_1}$.

4. Comparison between the NB-QMLE's and the Poisson QMLE

For the conditional mean parameter θ_0 of model (2.1), Ahmad and Francq (2016) proposed a Poisson QMLE (P-QMLE), which is defined as a measurable solution to the following problem

$$\widehat{\theta}_P = \arg \max_{\theta \in \Theta} \left(\widetilde{L}_{P,n}(\theta) \right), \quad (4.1a)$$

where

$$\widetilde{L}_{P,n}(\theta) = \frac{1}{n} \sum_{t=1}^n \left(-\widetilde{\lambda}_t(\theta) + X_t \log \left(\widetilde{\lambda}_t(\theta) \right) \right). \quad (4.1b)$$

Under similar assumptions to **A0-A10**, Ahmad and Francq (2016) showed consistency and asymptotic normality of the P-QMLE with

$$\sqrt{n} \left(\widehat{\theta}_P - \theta_0 \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N \left(0, J_P^{-1} I_P J_P^{-1} \right), \quad (4.2)$$

where $I_P = E \left(\frac{v_t(\theta_0)}{\lambda_t^2(\theta_0)} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \lambda_t(\theta_0)}{\partial \theta'} \right)$ and $J_P = E \left(\frac{1}{\lambda_t(\theta_0)} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \lambda_t(\theta_0)}{\partial \theta'} \right)$. One important property of the P-QMLE is its robustness to misspecification of the true conditional distribution of model (2.1). In this Section we will compare the NB-QMLE's and P-QMLE with regard

to asymptotic relative efficiency for some well-known specific cases of (2.1) and also on some particular GLM link functions of (2.3). We also compare these estimates in finite samples through some simulation experiments.

4.1. Comparison on asymptotic relative efficiency for specific models

4.1.1. The Poisson INGARCH model (Poisson autoregression)

The Poisson integer-valued GARCH (INGARCH(p, q)) process $\{X_t, t \in \mathbb{Z}\}$, as proposed by Rydberg and Shephard (2000) and Grunwald et al (2000), is defined to have a Poisson conditional distribution

$$X_t | \mathcal{F}_{t-1} \sim \mathcal{P}(\lambda_t), \quad t \in \mathbb{Z}, \quad (4.3a)$$

with conditional mean $\lambda_t = \lambda_t(\theta_0)$ specified as follows

$$\lambda_t(\theta_0) = \omega_0 + \sum_{i=1}^q \alpha_{0i} X_{t-i} + \sum_{j=1}^p \beta_{0j} \lambda_{t-j}(\theta_0), \quad (4.3b)$$

where $\theta_0 = (\omega_0, \alpha_{01}, \dots, \alpha_{0q}, \beta_{01}, \dots, \beta_{0p})'$ is such that $\omega_0 > 0$, $\alpha_{0i} \geq 0$, $\beta_{0j} \geq 0$. It is also assumed that X_t is nondegenerate and that if $p > 0$ $A_{\theta_0}(z) = \sum_{i=1}^q \alpha_{0i} z^i$ and $B_{\theta_0}(z) = 1 - \sum_{i=1}^p \beta_{0i} z^i$ have no common roots (cf. Ahmad and Francq, 2016). Ferland et al (2006) showed that under the following stability condition

$$\sum_{i=1}^q \alpha_{0i} + \sum_{j=1}^p \beta_{0j} < 1, \quad (4.4)$$

the process $\{X_t, t \in \mathbb{Z}\}$ given by (4.3) is strictly stationary (see also Franke, 2010). The ergodicity of the Poisson INGARCH(p, q) model (4.3) has been established first by Grunwald et al (2000) for $p = 0$, by Fokianos et al (2009), Neumann (2011), Davis and Liu (2016) and Douc et al, (2013) for the case $p = q = 1$, and by Doukhan et al (2012) and Gonçalves et al (2015) for general p and q . Under $\sum_{j=1}^p \beta_{0j} < 1$, the conditional mean λ_t of the process may be written in the form (2.1). Hence model (4.3) is a special case of (2.1). In particular,

it is characterized by the following "identity" GLM link function

$$\text{Var}(X_t | \mathcal{F}_{t-1}) = E(X_t | \mathcal{F}_{t-1}). \quad (4.5)$$

On the other hand, the P-QMLE of (4.3) reduces to the maximum likelihood estimate, which is asymptotically efficient and is then more asymptotically efficient than the pNB-QMLE. In particular $I_P = J_P$ follows from (4.2) and (4.5). Furthermore, assumptions **A0-A10** simplify in the case of the Poisson INGARCH model (4.3) as in Ahmad and Francq (2016). For instance, **A0** is implied by (4.4) while **A1** and **A4** follow from the linear form of λ_t in (4.3b). Since (4.4) entails the existence of moments of any order (Ferland et al, 2006) then **A3** and **A8** and **A9** are obviously satisfied. A similar argument shows that **A2** and **A6** are satisfied. As a result, I_r defined in **A9** reduces to $I_r = E\left(\frac{1}{\lambda_t(\theta_0)(r+\lambda_t(\theta_0))^2} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \lambda_t(\theta_0)}{\partial \theta'}\right)$. Note finally that the 2SNB-QMLE given by Section 3.2 is ill-defined in the present Poisson INGARCH case since the **Step 2** of Algorithm 3.1 is derived under the GLM assumption (2.3)-(2.4), which is different from the link function (4.5) characterizing the Poisson INGARCH model (4.3).

4.1.2. The negative binomial₁-INGARCH model

Here we follow Cameron and Trivedi (1986, 2013) who proposed the negative binomial_K conditional distribution in the context of static integer-valued regression. We say that $\{X_t, t \in \mathbb{Z}\}$ is a negative binomial_K-INGARCH (NB_K-INGARCH(p, q)) process if its conditional distribution is a negative binomial one,

$$X_t | \mathcal{F}_{t-1} \sim \mathcal{NB}(r_t, \pi_t), \quad t \in \mathbb{Z}, \quad (4.6a)$$

with parameters

$$r_t = r_0 \lambda_t^{2-K} \quad \text{and} \quad \pi_t = \frac{r_0 \lambda_t^{2-K}}{r_0 \lambda_t^{2-K} + \lambda_t}, \quad (4.6b)$$

where $K \in \mathbb{Z}$, $r_0 > 0$ and $\lambda_t = \lambda_t(\theta_0)$ satisfies the INGARCH(p, q) representation (4.3b). Model (4.6) in which $E(X_t | \mathcal{F}_{t-1}) = \lambda_t$ satisfies the following GLM link function

$$\text{Var}(X_t | \mathcal{F}_{t-1}) = E(X_t | \mathcal{F}_{t-1}) \left(1 + \frac{1}{r_0} (E(X_t | \mathcal{F}_{t-1}))^{K-1}\right), \quad (4.7)$$

which implies the process is conditionally overdispersed since $\text{Var}(X_t | \mathcal{F}_{t-1}) > E(X_t | \mathcal{F}_{t-1})$.

Now consider the $\text{NB}_1\text{-INGARCH}(p, q)$ model corresponding to $K = 1$, i.e.

$$X_t | \mathcal{F}_{t-1} \sim \mathcal{NB} \left(r_0 \lambda_t, \frac{r_0 \lambda_t}{r_0 \lambda_t + \lambda_t} \right) \equiv \mathcal{NB} \left(r_0 \lambda_t, \frac{r_0}{r_0 + 1} \right). \quad (4.8a)$$

for which (4.7) reduces to the following linear form

$$\text{Var}(X_t | \mathcal{F}_{t-1}) = \left(1 + \frac{1}{r_0} \right) E(X_t | \mathcal{F}_{t-1}). \quad (4.8b)$$

This is a strict generalization of the Poisson GLM condition (4.5) implied by the Poisson INGARCH model. In view of (4.5) and (4.8b), the $\text{NB}_1\text{-INGARCH}$ model (4.8a) presents some similarities with the Poisson INGARCH model (4.3). Indeed, we conjecture that the $\text{NB}_1\text{-INGARCH}$ is strictly stationary with finite second moment and ergodic under the same stationarity condition (4.4) for the Poisson INGARCH model. Moreover, from (4.2) and (4.8b), it follows under similar assumptions to **A0-A10** (see Ahmad and Francq, 2016) that

$$\sqrt{n} \left(\widehat{\theta}_P - \theta_0 \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N \left(0, \left(1 + \frac{1}{r_0} \right) \left(E \left(\frac{1}{\lambda_t(\theta_0)} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \lambda_t(\theta_0)}{\partial \theta'} \right) \right)^{-1} \right).$$

A more important result is that under the Poisson GLM condition (4.8b), it is easily seen that the P-QMLE is asymptotically efficient in the class of all QMLE's belonging to the linear exponential family. So the P-QMLE is more asymptotically efficient than the pNB-QMLE (see Gouriéroux et al (1984a, 1984b) in the case of static integer-valued regression models where adaptation to the present dynamic case is trivial). In fact, under **A0-A10** and in view of (3.5) and (4.8b), the asymptotic variance of the pNB-QMLE, $\widehat{\theta}_r$, is in "sandwich" form with $I_r = \left(1 + \frac{1}{r_0} \right) E \left(\frac{1}{\lambda_t(\theta_0)(r + \lambda_t(\theta_0))^2} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \lambda_t(\theta_0)}{\partial \theta'} \right)$. Note finally that like the Poisson INGARCH case, the 2SNB-QMLE given by Section 3.2 is ill-defined.

4.1.3. The negative binomial₂-INGARCH model

Consider the $\text{NB}_2\text{-INGARCH}(p, q)$ model corresponding to (4.6) with $K = 2$, i.e.

$$X_t | \mathcal{F}_{t-1} \sim \mathcal{NB} \left(r_0, \frac{r_0}{r_0 + \lambda_t} \right), \quad (4.9)$$

where $r_0 > 0$ and λ_t is given by (4.3b). The same parameter restriction for the Poisson INGARCH is assumed here. Model (4.9) has been considered by Zhu (2011), Davis and Liu (2016) and Christou and Fokianos (2014) who gave for $p = q = 1$ the following strict stationarity condition

$$\alpha_0^2 \left(1 + \frac{1}{r_0}\right) + 2\alpha_0\beta_0 + \beta_0^2 < 1,$$

with finite second moment. The formulation of Zhu (2011) is in fact,

$$X_t | \mathcal{F}_{t-1} \sim \mathcal{NB} \left(r_0, \frac{1}{1+\mu_t} \right), \quad (4.10)$$

where $r_0 \in \mathbb{N}^*$ is restricted to be a positive integer and μ_t satisfies (4.3b). However, the latter may be written in the form (4.9) while taking $\lambda_t = \frac{\mu_t}{r_0}$. For model (4.9), the link function (4.7) clearly reduces to the NB₂ GLM condition (3.7) (with $\delta^2 = 1$), i.e.

$$\text{Var}(X_t | \mathcal{F}_{t-1}) = E(X_t | \mathcal{F}_{t-1}) \left(1 + \frac{1}{r_0} E(X_t | \mathcal{F}_{t-1})\right), \quad r_0 > 0, \quad (4.11)$$

under which the 2SNB-QMLE is derived. Christou and Fokianos (2014) used the Poisson QMLE for estimating model (4.9) and proved its consistency and asymptotic normality with asymptotic variance in sandwich form like (4.2) where, in view of (4.4), $I_P = \frac{1}{r_0} E \left(\frac{(r_0 + \lambda_t(\theta_0))}{\lambda_t(\theta_0)} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \lambda_t(\theta_0)}{\partial \theta'} \right)$. Ahmad and Francq (2016) showed how their assumptions of consistency and asymptotic normality for the general model (2.1) simplify for model (4.9).

Concerning the pNB-QMLE, assumptions **A0-A10** simplify for $p = q = 1$ as in the Poisson INGARCH case. A notable difference is that one should assume the existence of $E(X_t^4) < \infty$, a condition of which is given by Ahmad and Francq (2016). For general p and q , a general moment condition seems difficult to obtain. Note that as $I_r = \frac{1}{r_0} E \left(\frac{(r_0 + \lambda_t(\theta_0))}{(r_0 + \lambda_t(\theta_0))} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \lambda_t(\theta_0)}{\partial \theta'} \right)$, none of the pNB-QMLE and P-QMLE is asymptotically superior than the other, unless r_0 would be known. In that case, one takes $r = r_0$ and the resulting pNB-QMLE, $\hat{\theta}_{r_0}$, would be asymptotically efficient. For instance, consider the *Geometric* INGARCH model which is a special case of the NB₂-INGARCH model (4.9) in which $r_0 = 1$, i.e. $X_t | \mathcal{F}_{t-1} \sim \mathcal{G} \left(\frac{1}{1+\lambda_t} \right)$. For this model, the Geometric QMLE (G-QMLE), which is a particular case of pNB-QMLE corresponding to $r = 1$, reduces to the maximum likelihood estimate and is then asymptotically efficient.

However, whether or not r_0 is known, the 2SNB-QMLE has the nice property of being asymptotically efficient in the class of all QMLE's belonging to the linear exponential family (cf. Theorem 3.3). Hence, it is more asymptotically efficient than the P-QMLE.

Finally, it is worth noting that when $K \notin \{1, 2\}$, the link function (4.7) corresponding to the NB_K -INGARCH model is different from both the Poisson GLM condition (4.8b) and the NB_2 variance assumption (4.11). Therefore, the 2SNB-QMLE is ill-defined and none of P-QMLE and pNB-QMLE is asymptotically preferred than the other.

4.1.4. The INAR(1) model

A well-known particular case of (2.1) is the first-order integer-valued autoregressive model (INAR(1)) proposed by McKenzie (1985) and Al-Osh and Alzaid (1987). This model has the following form

$$X_t = \alpha_0 \circ X_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}, \quad (4.12)$$

where $\{\varepsilon_t, t \in \mathbb{Z}\}$ is an independent and identically distributed (*iid*) sequence of non-negative integer-valued random variables with mean $E(\varepsilon_t) = \omega_0 > 0$ and variance $\text{Var}(\varepsilon_t) = \sigma_0^2 > 0$. The symbol \circ denotes the binomial thinning operator (cf. Steutel and Van Harn, 1979) defined for any non-negative integer-valued random variable X by $\alpha_0 \circ X = \sum_{i=1}^X Y_i$, where $\{Y_i, i \in \mathbb{N}\}$ is an *iid* Bernoulli random sequence such that $P(Y_i = 1) = \alpha_0 \in (0, 1)$. It is well known that

$$E(X_t | \mathcal{F}_{t-1}) = \lambda_t(\theta_0) = \alpha_0 X_{t-1} + \omega_0, \quad \text{with } \theta_0 = (\alpha_0, \omega_0)',$$

and that assumption **A0** reduces in term of α_0 to $\alpha_0 < 1$ (cf. Al-Osh and Alzaid, 1987). Furthermore, the INAR(1) model (4.12) obeys to the following *affine* GLM link function

$$\text{Var}(X_t | \mathcal{F}_{t-1}) = \alpha_0(1 - \alpha_0)X_{t-1} + \sigma_0^2 = (1 - \alpha_0)E(X_t | \mathcal{F}_{t-1}) + \sigma_0^2 - (1 - \alpha_0)\omega_0. \quad (4.13)$$

Note that if $\frac{\sigma_0^2}{\omega_0} = 1 - \alpha_0 < 1$, so that the innovation term ε_t would be *underdispersed*, then the affine link function (4.13) reduces to the linear Poisson GLM condition (4.8b) with proportionality constant $1 - \alpha_0$. Therefore, the P-QMLE would be asymptotically efficient

in the class of all QMLE's belonging to the linear exponential family and hence it would be more asymptotically efficient than the pNB-QMLE. Specifically,

$$\sqrt{n} \left(\widehat{\theta}_P - \theta_0 \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N \left(0, (1 - \alpha_0) \left(E \left(\frac{1}{\lambda_t(\theta_0)} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \lambda_t(\theta_0)}{\partial \theta'} \right) \right)^{-1} \right).$$

If, however, $\frac{\sigma_0^2}{\omega_0} \neq (1 - \alpha_0)$, then none of the two estimates P-QMLE and pNB-QMLE is more asymptotically efficient than the other. Moreover, in all cases the 2SNB-QMLE is ill-defined.

4.1.5. The double Poisson INGARCH (DP-INGARCH) model

The double-Poisson INGARCH model as proposed by Heinen (2003) is defined by

$$X_t | \mathcal{F}_{t-1} \sim \mathcal{DP}(\lambda_t, \gamma) \tag{4.14}$$

where λ_t is given by (4.3b), $\gamma > 0$ and $X \sim \mathcal{DP}(\lambda, \gamma)$ means that X has a double Poisson distribution (cf. Efron, 1986) given by

$$P(X = x) = c(\lambda, \gamma) \frac{e^{-x} x^x}{x} \left(\frac{e\lambda}{x} \right)^{\gamma x}, \quad x = \mathbb{N},$$

$c(\lambda, \gamma)$ being a normalizing constant. It is well known (e.g. Ahmad and Francq, 2016) that $E(X_t | \mathcal{F}_{t-1}) = \lambda_t$ so the DP-INGARCH is a particular case of (2.1). It is especially recommended for representing underdispersion. Ahmad and Francq conjectured that the DP-INGARCH(1, 1) is ergodic under the same condition (4.4) for the Poisson INGARCH(1, 1). Under that condition, assumptions **A0-A10** may be made explicit in the same way as in Ahmad and Francq (2016). Note that $\text{Var}(X_t | \mathcal{F}_{t-1})$ is approximately equal to λ_t/γ , which is an approximation of the Poisson GLM variance assumption (4.8b). Therefore, the P-QMLE is "approximately" more asymptotically efficient than the pNB-QMLE.

4.1.6. The generalized Poisson INGARCH model

A random variable X is said to have a Generalized Poisson (\mathcal{GP}) distribution with parameters $\lambda > 0$ and $0 \leq \tau < 1$, that is $X \sim \mathcal{GP}(\lambda, \tau)$, if

$$P(X = x) = \frac{\lambda(\lambda + \tau x)^{x-1} e^{-(\lambda + \tau x)}}{x!}, \quad x \in \mathbb{N}.$$

The latter definition can be extended for $\tau < 0$. In order to model both conditional overdispersion and conditional underdispersion, Zhu (2012a) introduced the generalized Poisson INGARCH (GP-INGARCH) model as follows

$$X_t | \mathcal{F}_{t-1} \sim \mathcal{GP}(\lambda_t(1 - \tau), \tau), \quad (4.15)$$

where λ_t is given by (4.3b). Since $E(X) = \lambda/(1 - \tau)$, it is clear that the GP-INGARCH is a particular case of (2.1). Gonçalves et al (2015) showed that the latter model is ergodic under the same condition as that of the Poisson INGARCH model. Ahmad and Francq (2016) showed the consistency of the P-QMLE under similar conditions to **A0-A10**. These assumptions hold under the ergodicity condition (4.4). Note that the conditional variance has a rather complicated expression of the conditional mean, which is different from the Poisson and the NB₂ GLM variance assumptions. Therefore, none of the two estimates P-QMLE and pNB-QMLE is asymptotically preferred than the other. For the same reason, the 2SNB-QMLE is ill-defined.

4.2. Comparison in finite samples

We now examine the finite-sample performance of the proposed NB-QMLE's on simulated series with sample size $n = 1000$. These series are generated from six instances of (2.1), namely:

i) The Poisson INGARCH(1, 1) model (4.3) with parameter $\theta_0 = (2, 0.6, 0.3)'$ (cf. Table 4.1).

ii) The geometric INGARCH(1, 1) model corresponding to (4.9) with $r_0 = 1$ and $\theta_0 = (2, 0.3, 0.6)'$ (cf. Table 4.2).

iii) The NB₂-INGARCH(1, 1) model (4.9) with parameters $r_0 = 3$ and $\theta_0 = (2, 0.6, 0.3)'$ (cf. Table 4.3).

iv) The DP-INGARCH(1, 1) model (4.14) with $\gamma = 2$ and $\theta_0 = (2, 0.6, 0.3)'$ (cf. Table 4.4).

v) The GP-INGARCH(1, 1) model (4.15) with $\tau = 0.4$ and $\theta_0 = (2, 0.6, 0.3)'$ (cf. Table

4.5).

vi) The (Poisson) INAR(1) model (4.12) with $\alpha = 0.9$ and $\varepsilon_t \sim \mathcal{P}(2)$, that is $\theta_0 = (2, 0.9)'$ (cf. Table 4.6).

Three QMLE's are compared on these models: i) The Poisson QMLE ($\widehat{\theta}_P$, Ahmad and Francq, 2016) given by (4.1), ii) the Geometric QMLE, $\widehat{\theta}_1$, corresponding to (3.3) with $r = 1$, and iii) the profile negative binomial QMLE, $\widehat{\theta}_4$, given by (3.3) with $r = 4$. For the NB₂-INGARCH(1, 1) model (4.9) we also run the two-stage NB-QMLE, $(\widehat{\theta}_{r^*}, \widehat{r}_1, \widehat{\theta}_{\widehat{r}_1}, \widehat{r}_2)$, given by Algorithm 3.1. These estimates are calculated using 500 Monte Carlo replications for the three mentioned models. In implementing the NB-QMLE's, we used the same devices. The starting parameter value, $\theta^{(0)} = (\omega^{(0)}, \alpha^{(0)}, \beta^{(0)})'$, of the nonlinear optimization routine (3.3) is set to the value obtained while preliminarily running a pNB-QMLE starting from an initial parameter $\theta^{(-1)} = (2, 0.3, 0.6)'$ and $r^{(-1)} = 3$. The unobservable starting values X_0 and $\lambda_0(\theta)$ of the INGARCH(1, 1) equation are estimated respectively by

$$\widetilde{X}_0 = \bar{X} \text{ and } \widetilde{\lambda}_0(\theta) = \frac{\omega + \alpha \bar{X}}{1 - \beta} \simeq E(\lambda_t(\theta)) \text{ for } \theta = (\omega, \alpha, \beta)' \in \Theta, \quad (4.16)$$

where \bar{X} is the empirical mean of the series X_1, \dots, X_n . Concerning Algorithm 3.1, which is only applied in the case of the NB₂-INGARCH model (4.9), we need to estimate the initial dispersion parameter r^* . For this we mimic the negative binomial₂ GLM assumption (4.11), taking r^* to be a solution to the equation, $S^2 = \bar{X} \left(1 + \frac{1}{r^*} \bar{X}\right)$, i.e.

$$r^* = \frac{(\bar{X})^2}{S^2 - \bar{X}}, \quad (4.17)$$

where S^2 is the sample variance of X_1, \dots, X_n . Of course, there is no theoretical justification for this choice. We have just replaced in (4.11) the conditional variance and conditional mean by their unconditional sample counterparts. For that choice, the series X_1, \dots, X_n should be overdispersed (i.e. $S^2 > \bar{X}$), otherwise r^* would be negative, which is not valid.

Mean of estimates, their standard deviation (StD), their *Asymptotic Standard Errors* (ASE) and their empirical Root Minimum Square Error (RMSE) over the 500 replications are reported in Tables 4.1-4.6. The RMSE of an estimate $\widehat{\theta}$ of θ_0 is calculated from the formula $\text{RMSE} = \sqrt{\text{bias}^2 + \text{StD}^2}$, where bias is the sample mean of $\widehat{\theta} - \theta_0$ over the 500

replications. The ASE's are obtained from the asymptotic variances of the NB-QMLE's given by Theorem 3.2 and Theorem 3.3, and that of the P-QMLE (cf. Ahmad and Francq, 2016 and Section 4.1 above).

θ_0		$\widehat{\theta}_P$	$\widehat{\theta}_1$	$\widehat{\theta}_4$
$\omega = 2$	Mean	1.9983	1.8804	1.9931
	StD	0.4067	0.3991	0.4094
	ASE	0.3979	0.4017	0.4041
	RMSE	0.4067	0.4166	0.4094
$\alpha_0 = 0.6$	Mean	0.6006	0.6147	0.6016
	StD	0.0363	0.0373	0.0377
	ASE	0.0360	0.0401	0.0398
	RMSE	0.0363	0.0400	0.0377
$\beta_0 = 0.3$	Mean	0.2982	0.2901	0.2977
	StD	0.0260	0.0266	0.0268
	ASE	0.0278	0.0280	0.0275
	RMSE	0.0260	0.0266	0.0268

Table 4.1: Mean, standard deviation, asymptotic standard error and empirical RMSE of $\widehat{\theta}_r$ ($r = 1, 4$) and $\widehat{\theta}_P$ for Poisson INGARCH(1, 1) series with $\theta_0 = (2, 0.6, 0.3)'$ and $n = 1000$.

θ_0		$\widehat{\theta}_P$	$\widehat{\theta}_1$	$\widehat{\theta}_4$
$\omega = 2$	Mean	1.9375	2.0008	2.0516
	StD	0.6018	0.4465	0.4519
	ASE	0.5501	0.4565	0.4619
	RMSE	0.6050	0.4465	0.4548
$\alpha_0 = 0.6$	Mean	0.5721	0.5967	0.5844
	StD	0.0777	0.0459	0.0462
	ASE	0.0687	0.0470	0.0502
	RMSE	0.0825	0.0460	0.0488
$\beta_0 = 0.3$	Mean	0.3218	0.3031	0.3141
	StD	0.0744	0.0370	0.0376
	ASE	0.0802	0.0382	0.0388
	RMSE	0.0775	0.0371	0.0401

Table 4.2: Mean, standard deviation, asymptotic standard error and empirical RMSE of $\widehat{\theta}_r$ ($r = 1, 4$) and $\widehat{\theta}_P$ for geometric INGARCH(1, 1) series with $\theta_0 = (2, 0.6, 0.3)'$ and $n = 1000$.

θ_0		$\hat{\theta}_P$	$\hat{\theta}_1$	$\hat{\theta}_3$	$\hat{\theta}_{\hat{r}}$
$\omega = 2$	Mean	2.1088	2.0711	2.0702	2.0316
	StD	0.4793	0.4658	0.4558	0.4508
	ASE	0.4981	0.4702	0.4664	0.4601
	RMSE	0.4914	0.4711	0.4611	0.4519
$\alpha_0 = 0.6$	Mean	0.6174	0.6109	0.6119	0.6166
	StD	0.0443	0.0491	0.0430	0.0427
	ASE	0.0533	0.0483	0.0446	0.0435
	RMSE	0.0475	0.0502	0.0446	0.0458
$\beta_0 = 0.3$	Mean	0.2712	0.2807	0.2796	0.2870
	StD	0.0305	0.0481	0.0296	0.0232
	ASE	0.0370	0.0450	0.0302	0.0262
	RMSE	0.0419	0.0518	0.0359	0.0265
$r_0 = 3$	Mean				2.9995
	StD	-	-	-	0.0330
	RMSE				0.0331

Table 4.3: Mean, standard deviation, asymptotic standard error and empirical RMSE of $\hat{\theta}_r$ ($r = 1, 4$), $\hat{\theta}_P$, $\hat{\theta}_{\hat{r}}$ and \hat{r}_2 for NB₂-INGARCH(1, 1) series with $r_0 = 3$, $\theta_0 = (2, 0.6, 0.3)'$ and $n = 1000$.

θ_0		$\widehat{\theta}_P$	$\widehat{\theta}_1$	$\widehat{\theta}_4$
$\omega = 2$	Mean	1.9857	1.9083	1.9090
	StD	0.5127	0.5303	0.5219
	ASE	0.5010	0.5200	0.5239
	RMSE	0.5128	0.5381	0.5297
$\alpha_0 = 0.6$	Mean	0.5943	0.5833	0.5829
	StD	0.0465	0.0493	0.0482
	ASE	0.0442	0.0510	0.0495
	RMSE	0.0408	0.0520	0.0511
$\beta_0 = 0.3$	Mean	0.3160	0.3208	0.3212
	StD	0.0351	0.0362	0.0357
	ASE	0.0348	0.0394	0.0376
	RMSE	0.0385	0.0417	0.0415

Table 4.4: Mean, standard deviation, asymptotic standard error and empirical RMSE of $\widehat{\theta}_r$ ($r = 1, 4$) and $\widehat{\theta}_P$ for DP-INGARCH(1, 1) series with $\gamma = 2$, $\theta_0 = (2, 0.6, 0.3)'$ and $n = 1000$.

θ_0		$\widehat{\theta}_P$	$\widehat{\theta}_1$	$\widehat{\theta}_4$
$\omega = 2$	Mean	1.7540	1.7681	1.7826
	StD	0.2248	0.2274	0.2271
	ASE	0.2448	0.2344	0.2260
	RMSE	0.3332	0.3247	0.3143
$\alpha_0 = 0.6$	Mean	0.5590	0.5471	0.5395
	StD	0.0615	0.0621	0.0620
	ASE	0.0665	0.0642	0.0654
	RMSE	0.0739	0.0815	0.0866
$\beta_0 = 0.3$	Mean	0.2570	0.2585	0.2589
	StD	0.0141	0.0143	0.0143
	ASE	0.0243	0.0221	0.0205
	RMSE	0.0452	0.0438	0.0435

Table 4.5: Mean, standard deviation and, asymptotic standard error and empirical RMSE of $\widehat{\theta}_r$ ($r = 1, 4$) and $\widehat{\theta}_P$ for GP-INGARCH(1, 1) series with $\tau = 0.4$, $\theta_0 = (2, 0.6, 0.3)'$ and $n = 1000$.

θ_0		$\hat{\theta}_P$	$\hat{\theta}_1$	$\hat{\theta}_4$
$\omega = 2$	Mean	2.1342	2.1018	2.1326
	StD	0.2862	0.3122	0.3495
	ASE	0.3098	0.5709	0.4894
	RMSE	0.3161	0.3284	0.3738
$\alpha_0 = 0.9$	Mean	0.8932	0.8948	0.8932
	StD	0.014	0.0152	0.0173
	ASE	0.0155	0.0294	0.0251
	RMSE	0.0155	0.0161	0.0186

Table 4.6: Mean, standard deviation, asymptotic standard error and empirical RMSE of $\hat{\theta}_r$ ($r = 1, 4$) and $\hat{\theta}_P$ for Poisson INAR(1) series with $\theta_0 = (2, 0.9)'$ and $n = 1000$.

From Tables 4.1-4.6 our Monte Carlo analysis broadly reveals that the parameters are well estimated by all accessed methods and the results are consistent with asymptotic theory. More precisely, when the INGARCH(1,1) model has a given conditional distribution, the QMLE calculated on that distribution is the best one compared to the other estimates in view of its smallest RMSE and ASE. Specifically, in the Poisson INGARCH(1,1) case (cf. Table 4.1) the P-QMLE outperforms the G-QMLE and the pNB-QMLE. Similarly, for the Geometric INGARCH(1,1) model (cf. Table 4.2), the G-QMLE has smaller RMSE than the P-QMLE and pNB-QMLE, $\hat{\theta}_4$. For the NB₂-INGARCH(1,1) model with dispersion parameters $r_0 = 3$ (cf. Table 4.3), the four-stage estimate $\hat{\theta}_{\hat{r}}$ outperforms the P-QMLE, the G-QMLE and the pNB-QMLE, $\hat{\theta}_4$. As expected, it can be seen from Table 4.4 that the P-QMLE outperforms the pNB-QMLE's for the DP-INGARCH(1,1) because the conditional variance is approximately proportional to the conditional mean (cf. Section 4.1.4). For the GP-INGARCH case (cf. Table 4.5), the pNB-QMLE's give better estimates than the P-QMLE in terms of the empirical RMSE and the ASE criteria. Finally, the P-QMLE outperforms the pNB-QMLE's in the case of the Poisson INAR(1) (cf. Table 4.6).

5. Real applications

For illustration purposes, we propose to apply the two-stage NB-QMLE given by Algorithm 3.1 to two famous integer-valued time series under the $\text{NB}_2\text{-INGARCH}(1, 1)$ framework. The first one is the *Polio data* (Zeger, 1988) while the second one is the *Transaction data* (Fokianos et al, 2009). The choice of the $\text{NB}_2\text{-INGARCH}(1, 1)$ model is motivated by the overdispersion of the mentioned series. Moreover, these two real series were considered by Zhu (2011) and Christou and Fokianos (2014) respectively using the $\text{NB}_2\text{-INGARCH}(1, 1)$ model, but via different estimation methods. This allows us to compare their methods with our proposed 2SNB-QMLE. All procedures have been applied on a personal computer using *R*. The optimization (3.3) is carried out using the function `constrOptim()` of *R*.

5.1. The polio data

The first dataset is the monthly number of poliomyelitis cases in the United States over the sample period from 1970 to 1983 with a total of $n = 168$ observations (cf. Figure 5.1). This series was originally modelled by Zeger (1988) and used later by many authors (see Zeger and Qaqish, 1988; Davis et al, 1999; Benjamin et al, 2003; Heinen, 2003; Davis and Wu, 2009; Zhu, 2011 among others). The Polio series with a sample mean of 1.3333 and a sample variance of 3.5050 is clearly overdispersed. It has a large frequency of zeros, has an asymmetric marginal distribution and is characterized by a locally constant behavior (cf. Figure 5.1, see also Zeger, 1987; Benjamin et al, 2003; Zhu, 2011).

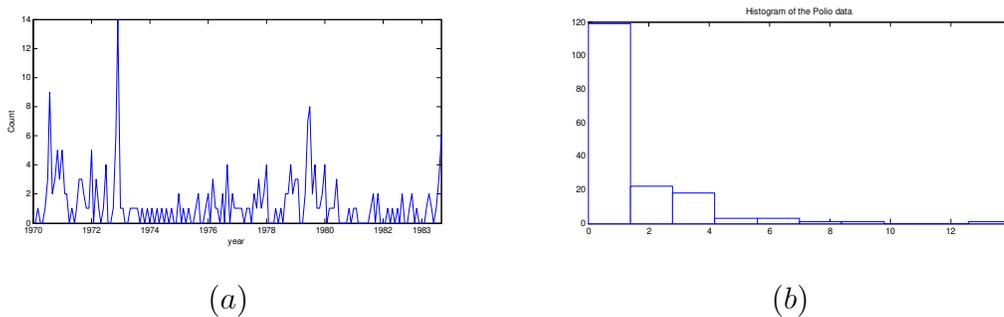


Figure 5.1: Monthly number of poliomyelitis cases in the United States from 1970 to 1983.

(a) Series. (b) Histogram.

Zhu (2011) fitted a NB₂-INGARCH(1, 1) model of the form (4.10) to the polio series. As emphasized above, this model is slightly different from the model (4.9). First, the dispersion parameter in (4.10) is taken to be a positive integer, which is somewhat restrictive. Second, the probability parameter is $\frac{1}{1+\mu_t}$ rather than $\frac{r_0}{r_0+\lambda_t}$ in (4.9). So the conditional mean of model (4.10) is not in the form (2.1). However, by taking $\lambda_t = \frac{\mu_t}{r_0}$ we find model (4.9) with a different parametrization. Zhu (2011) estimated model (4.10) using an approximate maximum likelihood estimate. This estimate consists in maximizing the negative binomial likelihood over θ for fixed r and then choosing θ with largest likelihood over all selected values of $r \in \{1, \dots, \bar{r}\}$, for some fixed positive integer \bar{r} . The estimated model of Zhu (2011) is given by

$$X_t | \mathcal{F}_{t-1} \sim \mathcal{NB} \left(\hat{r}, \frac{1}{1+\hat{\mu}_t} \right), \quad (5.1)$$

$$\hat{r} = 2, \begin{cases} \hat{\mu}_t = 0.31190 + 0.1843X_{t-1} + 0.1815\hat{\mu}_{t-1}, & 2 \leq t \leq 168 \\ \hat{\mu}_1 = \bar{X}, \end{cases}$$

from which the estimate of $E(X_t)$ is $2 \times \frac{0.3119}{1-(0.1843+0.1815)} = 0.9836$ and the persistence (or stability) parameter is $0.1843 + 0.1815 = 0.3658$.

We also apply the P-QMLE in Christou and Fokianos (2014) to the Polio data, giving the results in Table 5.1a.

INGARCH parameters	Estimates:	ASE
ω_0	0.6401	0.1670
α_0	0.3501	0.1289
β_0	0.1821	0.0673

Table 5.1a: P-QML estimates and their asymptotic standard errors for the NB₂-INGARCH(1, 1) model from the Polio data.

The fitted model is then given by

$$X_t | \mathcal{F}_{t-1} \sim \mathcal{P} \left(\hat{\lambda}_t^p \right), \begin{cases} \hat{\lambda}_t^p = 0.6401 + 0.3501X_{t-1} + 0.1821\hat{\lambda}_{t-1}^p, & 2 \leq t \leq 168 \\ \hat{\lambda}_1^p = \bar{X} = 1.3333, \end{cases} \quad (5.2)$$

with persistence parameter 0.5322 and estimated mean $\frac{0.6401}{1-(0.3501+0.1821)} = 1.3683$.

To compare with Zhu's (2011) fit and the Poisson QMLE of Christou and Fokianos (2014) we estimated a NB₂-INGARCH(1, 1) model (4.9) using the 2SNB-QMLE (Algorithm 3.1). In implementing Algorithm 3.1 we used the same devices as in Section 4.2. More precisely, the initial dispersion parameter r^* is calculated using (4.17) giving $r^* = \frac{(1.3333)^2}{3.5050-1.3333} = 0.8186$ while the starting values of the INGARCH(1, 1) equation (4.3b) are taken as in (4.16). The initial conditional mean parameter $\theta^{(0)}$ of the optimization problem (3.3) is obtained while preliminarily running the Geometric QMLE on the polio series with initial parameter (2, 0.3, 0.6)'. The estimated parameters of the model and their ASE are summarized in Table 5.1b. The ASE's are calculated from the asymptotic distribution of the 2SNB-QMLE given by Theorem 3.3. In particular, the ASE of $\hat{\gamma}_2 = (\hat{r}_2)^{-1}$ is computed from (3.14b) and (3.16) while the ASE of $\hat{\theta}_{\hat{r}_2}$ is obtained from (3.14d) and (3.15). Note that the ASE of \hat{r}_2 is not available since the asymptotic distribution of \hat{r}_2 has not a usual form, but may be estimated using parametric bootstrap.

NB ₂ -INGARCH parameters	Estimates: $(\hat{\theta}_{\hat{r}_1}, \hat{\gamma}_2, \hat{r}_2)$	ASE of $\hat{\theta}_{\hat{r}_1}, \hat{\gamma}_2$
ω_0	0.6564	0.2050
α_0	0.3743	0.1580
β_0	0.1511	0.0935
$\gamma_0 = \frac{1}{r_0}$	0.3843	0.1945
r_0	2.6023	—

Table 5.1b: 2SNB-QML estimates and their asymptotic standard errors for the NB₂-INGARCH(1, 1) model from the Polio data.

The fitted model (4.9) using the 2SNB-QMLE is given by

$$X_t | \mathcal{F}_{t-1} \sim \mathcal{NB} \left(\hat{r}_2, \frac{\hat{r}_2}{\hat{r}_2 + \hat{\lambda}_t} \right), \quad (5.3)$$

$$\hat{r}_2 = 2.6023, \quad \begin{cases} \hat{\lambda}_t = 0.6564 + 0.3743X_{t-1} + 0.1511\hat{\lambda}_{t-1}, & 2 \leq t \leq 168 \\ \hat{\lambda}_1 = \bar{X} = 1.3333, \end{cases}$$

with persistence parameter $0.3743 + 0.1511 = 0.5254$. Note that our estimate of the mean $E(X_t)$ is $\frac{0.6564}{1-(0.3743+0.1511)} = 1.3834$, which is closer to the sample mean $\bar{X} = 1.3333$ than the estimated mean, 0.9836, given by Zhu's (2011) model. However, the estimated mean given by the P-QMLE is slightly better than the estimated mean given by the $2SNB-QMLE$. On the other hand, some properties of the residuals are shown in Figure 5.2a. From the sample autocorrelation and partial autocorrelation functions in Figure 5.2a (panels (a) and (b)), the residuals look like a white noise. However, a visual inspection (cf. Figure 5.2a, panels (c) and (d)) reveals that the normality assumption of the residuals is untenable.

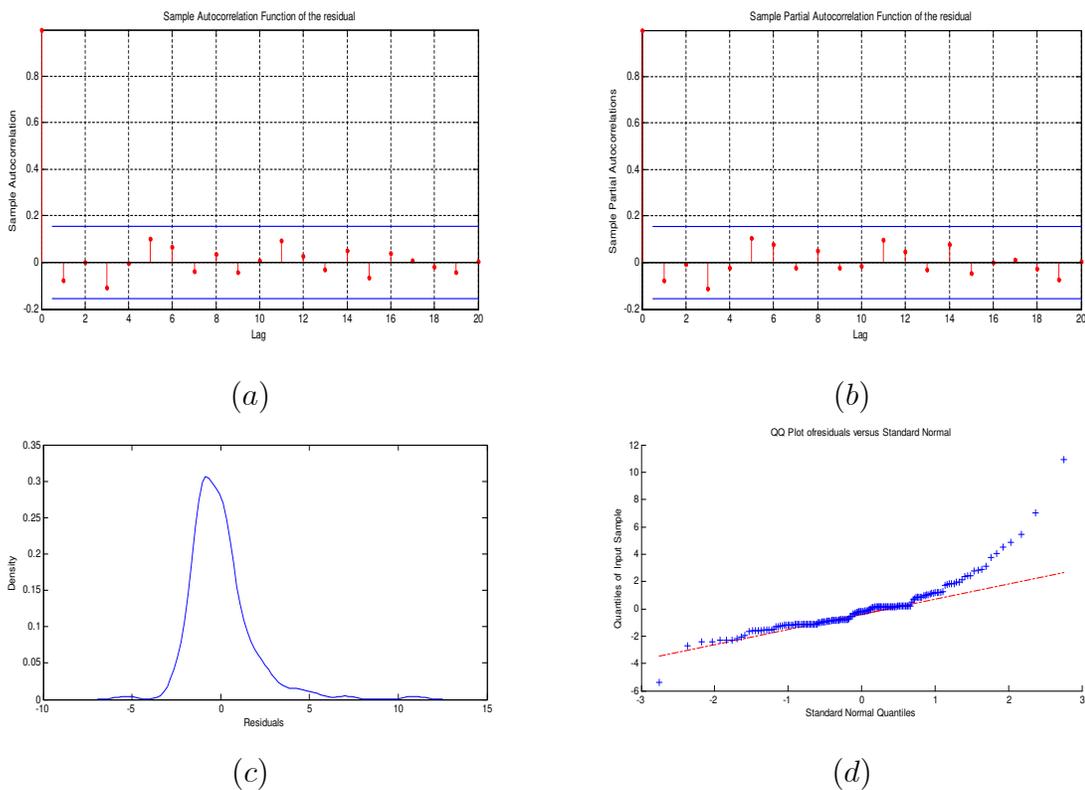


Figure 5.2a: Residual analysis for the Polio series.

(a) Sample autocorrelations of residuals. (b) Sample partial autocorrelations of residuals.

(c) Kernel density of residuals.

(d) QQ-plot of the residuals versus the standard normal distribution.

To get a more meaningful conclusion about the $2SNB-QMLE$ fitting we rather use the ran-

domized quantile residuals used in Zhu (2011). The residuals are given by $\hat{e}_t = \Phi^{-1}(p_t)$, where Φ^{-1} is the inverse of the standard normal cumulative distribution and p_t is a random number uniformly chosen in the interval

$$\left[F \left(X_t - 1; \hat{\theta}_{\hat{r}}, \hat{r}_2 \right), F \left(X_t; \hat{\theta}_{\hat{r}}, \hat{r}_2 \right) \right],$$

$F \left(x; \hat{\theta}_{\hat{r}}, \hat{r}_2 \right)$ being the cumulative function of the NB₂ distribution evaluated at x with parameters $\hat{\theta}_{\hat{r}}$ and \hat{r} (cf. Zhu, 2011; Benjamin et al, 2003). In summary, regarding the stability of the estimated model, the significance of its coefficients and the randomized residual analysis in Figure 5.2b, it can be concluded that the estimated model is acceptable.

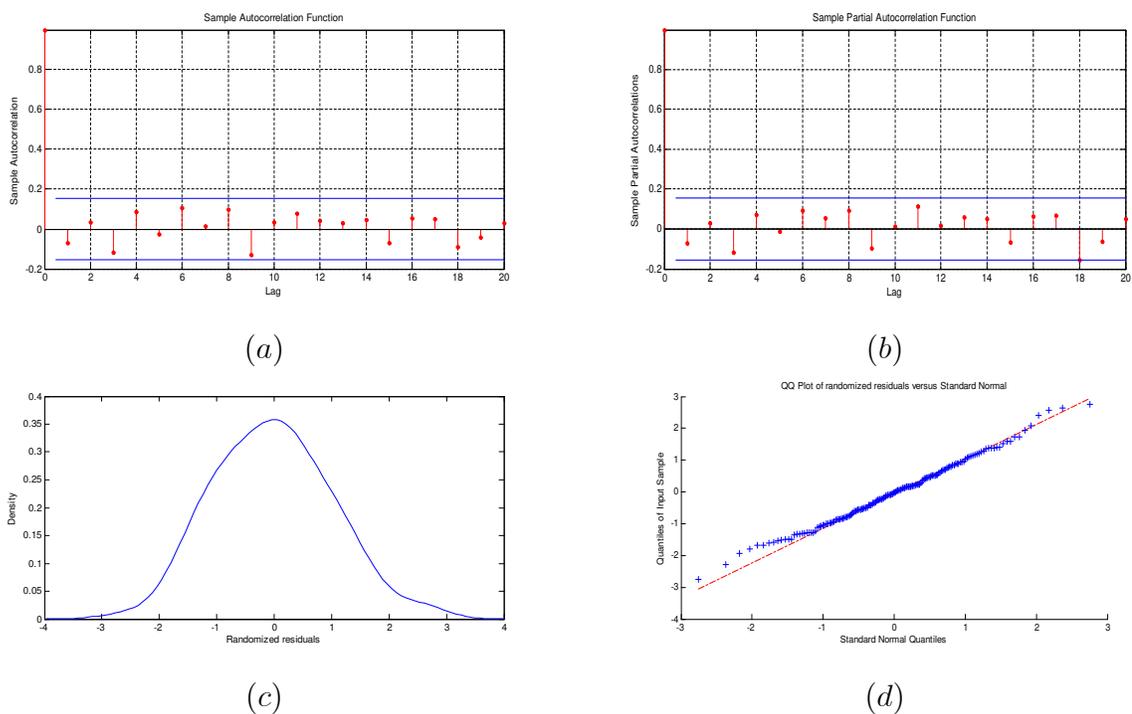


Figure 5.2b: Randomized residual analysis for the Polio series. (a) Sample autocorrelations of randomized residuals. (b) Sample partial autocorrelations of randomized residuals. (c) Kernel density of randomized residuals. (d) QQ-plot of the randomized residuals versus the standard normal distribution.

Now we compare in-sample performance of our fit (5.1) with that of Zhu (2011) and the P-QMLE in Christou and Fokianos (2014). Table 5.2 provides the *Residual Sum of*

Squares (RSS) induced by models (5.1), (5.2) and (5.3). These RSS's are given respectively by $\text{RSS}(\hat{\lambda}_t) = \sum_{t=2}^{168} (X_t - \hat{\lambda}_t)^2$, $\text{RSS}(2\hat{\mu}_t) = \sum_{t=2}^{168} (X_t - 2\hat{\mu}_t)^2$ and $\text{RSS}(\hat{\lambda}_t^p) = \sum_{t=2}^{168} (X_t - \hat{\lambda}_t^p)^2$ starting from initial values $\hat{\lambda}_1 = \hat{\lambda}_1^p = \hat{\mu}_1 = \bar{X}$. The latter initial value was considered by Zhu (2011).

Predictors	$\hat{\lambda}_t$	$2\hat{\mu}_t$	$\hat{\lambda}_t^p$
RSS	535.1793	540.6634	533.5275

Table 5.2: Residual sum of squares (RSS) of the predictors

$\hat{\lambda}_t$ (5.2), $2\hat{\mu}_t$ (5.1) and $\hat{\lambda}_t^p$ (5.3) for the Polio series.

From Table 5.2 it can be seen that our model estimated by the 2SNB-QMLE (Algorithm 3.1) outperforms the model of Zhu (2011) with smaller Residual Sum of Squares (RSS). Since the conditional mean may be influenced by the choice of the initial values, we have calculated several RSS corresponding to models (5.1), (5.2) and (5.3), starting from several initial values $\hat{\lambda}_1$, $\hat{\lambda}_1^p$ and $\hat{\mu}_1$. The unreported results were virtually the same. However, the model obtained from the P-QMLE is slightly better than our model with RSS equaling 533.5275.

Finally, Figure 5.3 displays the polio data together with the 2SNB-QML estimated conditional mean $\hat{\lambda}_t$ and the estimated conditional variance given by $\hat{v}_t = \hat{\lambda}_t \left(1 + \frac{1}{r_2} \hat{\lambda}_t\right)$, where the conditional overdispersion phenomenon seems reproduced.

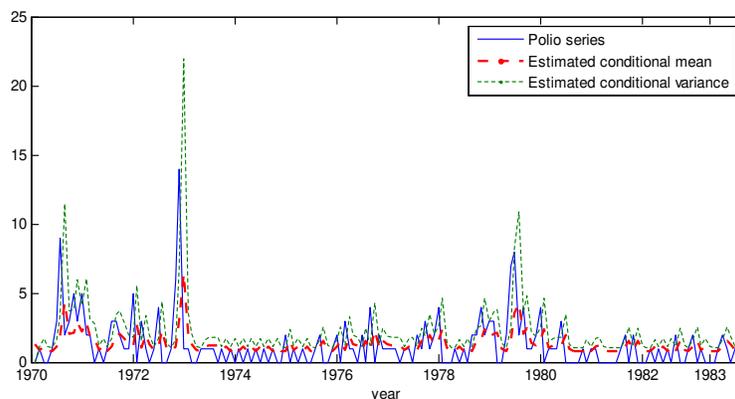


Figure 5.3: The Polio series and its (2SNB-QML) estimated conditional mean and conditional variance.

5.2. Transaction data

The second dataset is the number of transactions per minute for the stock Ericsson B during July 05, 2002. This series has a total of $n = 460$ observations representing the transaction of approximately 8 hours (from 09:35 through 17:14, cf. Figure 5.4). It was used by Fokianos *et al* (2009), Davis and Liu (2016) and Christou and Fokianos (2014) among others. Like the Polio data, the Transaction series is overdispersed in view of its sample mean and sample variance, which are equal to 9.8239 and 23.7532 respectively. It is characterized by small values, an asymmetric marginal distribution and a locally constant behavior (cf. Figure 5.4).

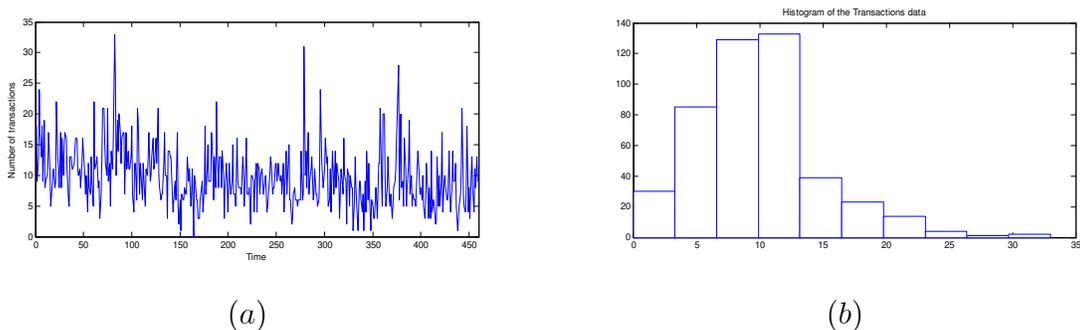


Figure 5.4: Number of transactions per minute for the stock Ericsson B during July 05, 2002.

(a) Series. (b) Histogram.

Using the Poisson QMLE, Christou and Fokianos (2014) fitted a NB_2 -INGARCH(1, 1) model (4.9) to the Transaction data. They found the following specification

$$X_t | \mathcal{F}_{t-1} \sim \mathcal{NB} \left(\hat{r}, \frac{\hat{r}}{\hat{r} + \hat{\mu}_t} \right), \quad (5.4)$$

$$\hat{r} = 7.0220, \quad \begin{cases} \hat{\mu}_t = 0.5808 + 0.1986X_{t-1} + 0.7445\hat{\mu}_{t-1}, & 2 \leq t \leq 460 \\ \hat{\mu}_1 = 0, \end{cases}$$

with a strong persistence parameter 0.9431 and an estimated mean $\frac{0.5808}{1-0.9431} = 10.2070$.

Motivated by the fact that the 2SNB-QMLE (Algorithm 3.1) is more asymptotically efficient than the P-QMLE in the context of the NB_2 -INGARCH model (cf. Section 4.1.3), we applied the former estimate to the Transaction series using the same devices as for the Polio data. Indeed, from (4.17), the initial dispersion parameter is taken to be $r^* = \frac{(9.8239)^2}{23.7532 - 9.8239} = 6.9285$, while the starting values of the INGARCH(1, 1) equation (4.3b) are

set according to (4.16). The parameter estimates and their ASE are summarized in Table 5.3.

NB ₂ -INGARCH parameters	Estimates: $(\hat{\theta}_{\hat{r}_1}, \hat{\gamma}_2, \hat{r}_2)$	ASE of $\hat{\theta}_{\hat{r}_1}, \hat{\gamma}_2$
ω_0	0.7996	0.4034
α_0	0.7928	0.0650
β_0	0.1249	0.0340
$\gamma_0 = \frac{1}{r_0}$	0.1279	0.0241
r_0	7.8199	—

Table 5.3: 2SNB-QML estimates and their asymptotic standard errors for the NB₂-INGARCH(1, 1) model from the Transaction data.

Thus our fitted NB₂-INGARCH(1, 1) model from the Transaction series using the 2SNB-QMLE is given by

$$X_t | \mathcal{F}_{t-1} \sim \mathcal{NB} \left(\hat{r}_2, \frac{\hat{r}_2}{\hat{r}_2 + \hat{\lambda}_t} \right), \quad (5.5)$$

$$\hat{r}_2 = 7.8199, \quad \begin{cases} \hat{\lambda}_t = 0.7996 + 0.7928X_{t-1} + 0.1249\hat{\lambda}_{t-1}, & 2 \leq t \leq 460 \\ \hat{\lambda}_1 = \bar{X} = 9.8134, \end{cases}$$

with a strong persistence parameter, 0.9177, and an estimated mean, $\frac{0.7996}{1-0.9177} = 9.7157$, which is closer to the sample mean $\bar{X} = 9.8239$ than the estimated mean obtained from the specification of Christou and Fokianos (2014).

Figure 5.5a shows the sample autocorrelation function (panel (a)), the sample partial autocorrelation function (panel (b)), the Kernel density (panel (c)) and the QQ-plot (panel (d)) of the residuals of model (5.5). It turns out that the hypothesis that the residuals form

a non-Gaussian white noise is strongly tenable.

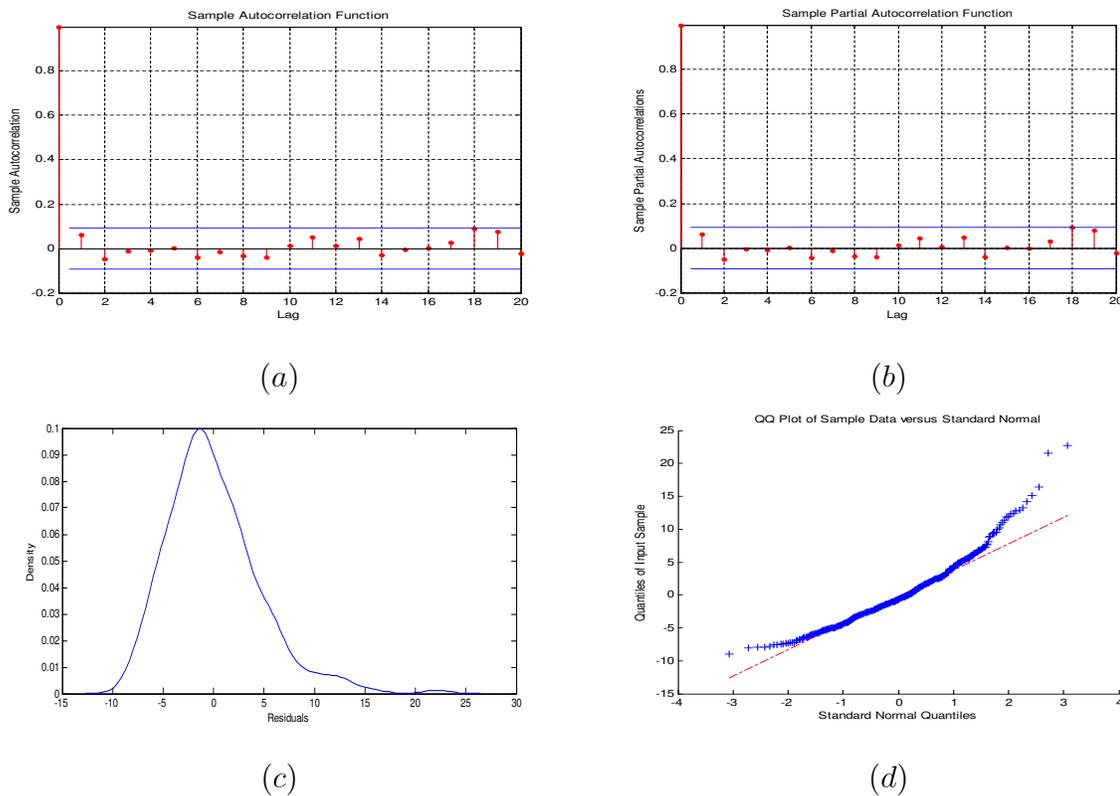


Figure 5.5a: Residual analysis for the Transaction series.

(a) Sample autocorrelations of residuals. (b) Sample partial autocorrelations of residuals. (c) Kernel density of residuals. (d) QQ-plot of residuals versus the standard normal distribution.

However, from Figure 5.5b it turns out that the randomized residuals calculated as above look like a Gaussian white noise regarding the sample autocorrelations and partial correlations (panels (a) and (b)), the kernel density estimate (panel (c)) and the QQ-plot versus

the normal distribution (panel (d)).

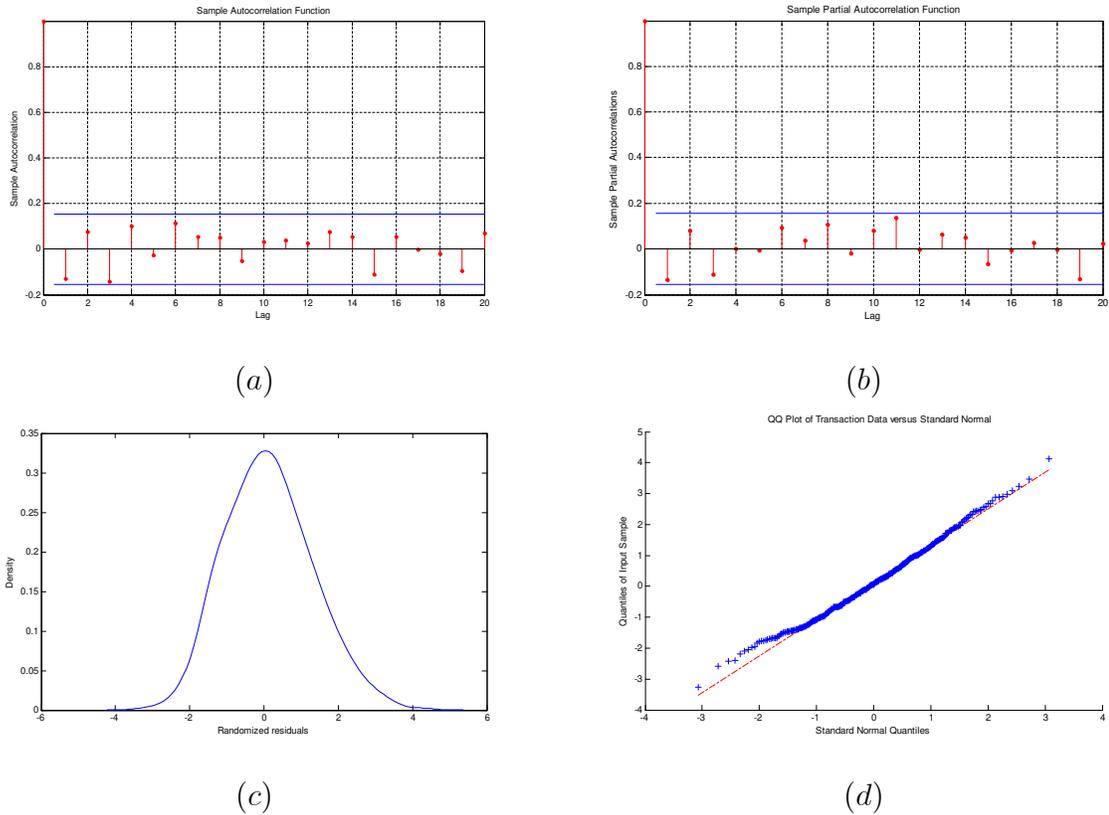


Figure 5.5b: Randomized residual analysis for the Transaction series.

(a) Sample autocorrelations of randomized residuals. (b) Sample partial autocorrelations of randomized residuals. (c) Kernel density of randomized residuals.

(d) QQ-plot of the randomized residuals versus the standard normal distribution.

Next we compare the RSS of our fit (5.5) with that of Christou and Fokianos (2014) given by (5.4). Because of the high persistence parameters in both models, the RSS's may be influenced by the starting values for the moderate sample size of the Transaction series. We therefore started the equations (5.4) and (5.5) from several initial values (cf. Table 5.4)

although Christou and Fokianos (2014) have taken $\hat{\mu}_1 = 0$.

Predictors	$\hat{\lambda}_t$	$\hat{\mu}_t$	$\hat{\lambda}_t$	$\hat{\mu}_t$	$\hat{\lambda}_t$	$\hat{\mu}_t$
Initial values $\hat{\lambda}_1, \hat{\mu}_1$	0	0	9.8239	9.8239	10.2070	10.2070
RSS	10400.6733	10422.8003	9809.6645	9943.0150	9796.8644	9933.0780

Table 5.4: Residual sum of squares (RSS) of the predictors $\hat{\lambda}_t$ (5.5) and $\hat{\mu}_t$ (5.4) for the Transaction data.

It can be seen from Table 5.4 that model (5.5) estimated by the 2SNB-QMLE has the smallest RSS for all chosen initial values. Figure 5.6 shows the Transaction series together with the estimated conditional mean $\hat{\lambda}_t$ and the estimated conditional variance given by $\hat{v}_t = \hat{\lambda}_t \left(1 + \frac{1}{r_2} \hat{\lambda}_t\right)$, where the conditional overdispersion phenomenon is highlighted.

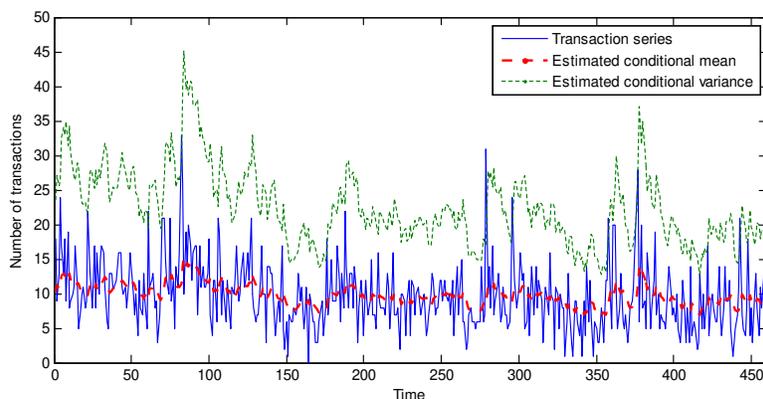


Figure 5.6: The Transaction series and its (2SNB-QML) estimated conditional mean and conditional variance.

6. Conclusion

In this paper we proposed two negative binomial QMLE's, namely the profile NB-QMLE and the two-stage NB-QMLE, for a general class of integer-valued time series models. These estimates are consistent and asymptotically Gaussian under general weak assumptions. In particular, they are robust to misspecification of the true conditional distribution of the

model provided that the conditional mean is well specified. Moreover, under the negative binomial₂-GLM link function, the two-stage NB-QMLE is more asymptotically efficient than the Poisson QMLE and is especially well adapted to overdispersed series. Furthermore, it is asymptotically efficient in the class of all QMLE's belonging to the linear exponential family. In fact, the two-stage NB-QMLE may be seen as a good alternative to the maximum likelihood estimate (for models with negative binomial₂ conditional distributions), which suffers from the non-robustness to misspecification of the true conditional distribution and whose calculation is quite tedious. From asymptotics of the NB-QMLE's (Theorems 3.1-3.3), portmanteau tests for goodness-of-fit in the framework of the general model (2.1) may be easily derived.

On the other hand, we have seen how the proposed NB-QMLE's can be applied to some specific integer-valued models such as the Poisson and negative binomial INGARCH models and also to the Double Poisson INGARCH, the Generalized Poisson INGARCH and the INAR equation. Other famous particular cases of model (2.1) like the log-INGARCH model (Fokianos and Tjøstheim, 2011), the COM-Poisson INGARCH (Zhu, 2012c) and the Integer-valued ARMA (INARMA) models also apply in the framework of our methods. Finally, generalizations of the proposed methods to multivariate versions of model (2.1) are appealing.

7. Proofs

7.1. Proof of Theorem 3.1

Following Wald's approach, the proof of Theorem 3.1 is based on the following three lemmas.

Lemma 7.1 *Under **A1-A2** $\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} \left| L_{n,r}(\theta) - \tilde{L}_{n,r}(\theta) \right| = 0$, a.s.*

Proof Using the inequality $\log(x) \leq x - 1$, the fact that $\tilde{\lambda}_t(\theta) > 0$, the assumptions

A1-A2 and the Césaro lemma it follows that

$$\begin{aligned}
\sup_{\theta \in \Theta} \left| L_{n,r}(\theta) - \tilde{L}_{n,r}(\theta) \right| &= \frac{1}{n} \sup_{\theta \in \Theta} \left| \sum_{t=1}^n \left(\log \left(\frac{r+\lambda_t(\theta)}{r+\tilde{\lambda}_t(\theta)} \right) + X_t \log \left(\frac{\tilde{\lambda}_t(\theta)(r+\lambda_t(\theta))}{\lambda_t(\theta)(r+\tilde{\lambda}_t(\theta))} \right) \right) \right| \\
&\leq \frac{1}{n} \sum_{t=1}^n \left(\frac{1}{r} \sup_{\theta \in \Theta} \left| \lambda_t(\theta) - \tilde{\lambda}_t(\theta) \right| + X_t \sup_{\theta \in \Theta} \left| \tilde{\lambda}_t(\theta) - \lambda_t(\theta) \right| \frac{r}{cr} \right) \\
&= \frac{1}{n} \sum_{t=1}^n \left(\frac{1}{r} a_t + \frac{1}{c} X_t a_t \right) \xrightarrow[n \rightarrow \infty]{a.s.} 0. \blacksquare
\end{aligned}$$

Lemma 7.2 Under **A0-A4**,

i) $E(l_{1,r}(\theta_0)) < \infty$.

ii) $E(l_{1,r}(\theta_0)) \geq E(l_{1,r}(\theta))$ for all $\theta \in \Theta$.

iii) $E(l_{1,r}(\theta)) = E(l_{1,r}(\theta_0)) \Rightarrow \theta = \theta_0$.

Proof Under **A1** the random variables $\log \left(\frac{r}{r+\lambda_t(\theta)} \right)$ and $\log \left(\frac{\lambda_t(\theta)}{r+\lambda_t(\theta)} \right)$ are bounded. Hence, they admit finite moments of all order. By the Jensen and Hölder inequalities together with **A3** it follows that

$$\begin{aligned}
|E(l_{1,r}(\theta_0))| &\leq E(|l_{1,r}(\theta_0)|) \leq E \left(\left| \log \left(\frac{r}{r+\lambda_t(\theta_0)} \right) \right| \right) + E \left(\left| X_t \log \left(\frac{\lambda_t(\theta_0)}{r+\lambda_t(\theta_0)} \right) \right| \right) \\
&\leq E \left(\left| \log \left(\frac{r}{r+\lambda_t(\theta_0)} \right) \right| \right) + (E(X_t^\delta))^{1/\delta} \left(E \left(\left| \log \frac{\lambda_t(\theta_0)}{r+\lambda_t(\theta_0)} \right|^{\frac{\delta}{\delta-1}} \right) \right)^{\frac{\delta-1}{\delta}} < \infty. \quad (7.1)
\end{aligned}$$

On the other hand, using again the inequality $\log(x) \leq x - 1$, we have

$$\begin{aligned}
E(l_{1,r}(\theta) - l_{1,r}(\theta_0)) &= E \left(r \log \left(\frac{r+\lambda_t(\theta_0)}{r+\lambda_t(\theta)} \right) + X_t \log \left(\frac{\lambda_t(\theta)(r+\lambda_t(\theta_0))}{\lambda_t(\theta_0)(r+\lambda_t(\theta))} \right) \right) \\
&\leq E \left(r \left(\frac{r+\lambda_t(\theta_0)}{r+\lambda_t(\theta)} - 1 \right) + X_t \left(\frac{\lambda_t(\theta)(r+\lambda_t(\theta_0))}{\lambda_t(\theta_0)(r+\lambda_t(\theta))} - 1 \right) \right) \\
&= rE \left(\frac{\lambda_t(\theta_0) - \lambda_t(\theta)}{r+\lambda_t(\theta)} + \frac{\lambda_t(\theta) - \lambda_t(\theta_0)}{r+\lambda_t(\theta)} \right) = 0. \quad (7.2)
\end{aligned}$$

By (7.1) and (7.2) it follows that $E(l_{1,r}(\theta) - l_{1,r}(\theta_0)) \in [-\infty, 0]$. Finally, inequality (7.2) reduces to equality if and only if

$$E \left(r \log \left(\frac{r+\lambda_t(\theta_0)}{r+\lambda_t(\theta)} \right) + X_t \log \left(\frac{\lambda_t(\theta)(r+\lambda_t(\theta_0))}{\lambda_t(\theta_0)(r+\lambda_t(\theta))} \right) \right) = 0,$$

which holds if and only if $\lambda_t(\theta) = \lambda_t(\theta_0)$ and hence, by the identifiability assumption **A4**, if and only if $\theta = \theta_0$. \blacksquare

Lemma 7.3 Under **A0-A5**, there exists for all $\theta \neq \theta_0$ a neighborhood $V(\theta)$ such that

$$\limsup_{n \rightarrow \infty} \sup_{\theta^* \in V(\theta)} \tilde{L}_{n,r}(\theta^*) < \limsup_{n \rightarrow \infty} \tilde{L}_{n,r}(\theta_0) \quad a.s. \quad (7.3)$$

Proof For all $\bar{\theta} \in \Theta$ and $k \in \mathbb{N}^*$ let $V_k(\bar{\theta})$ be the open ball with center $\bar{\theta}$ and radius $1/k$. Since $\sup_{\theta \in V_k(\bar{\theta}) \cap \Theta} l_{t,r}(\theta)$ is a measurable function of the terms of $\{X_t, t \in \mathbb{Z}\}$, which is strictly stationary and ergodic under **A0**, then $\left\{ \sup_{\theta \in V_k(\bar{\theta}) \cap \Theta} l_{t,r}(\theta), t \in \mathbb{Z} \right\}$ is also strictly stationary and ergodic where by Lemma 7.2 $E \left(\sup_{\theta \in V_k(\bar{\theta}) \cap \Theta} l_{t,r}(\theta) \right) \in [-\infty, +\infty[$. Therefore, in view of Lemma 7.1 and the ergodic theorem (Billingsley, 2008) it follows that

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in V_k(\bar{\theta}) \cap \Theta} \tilde{L}_{n,r}(\theta) = \limsup_{n \rightarrow \infty} \sup_{\theta \in V_k(\bar{\theta}) \cap \Theta} L_{n,r}(\theta) \leq E \left(\sup_{\theta \in V_k(\bar{\theta}) \cap \Theta} l_{1,r}(\theta) \right).$$

By the Beppo-Levi theorem $E \left(\sup_{\theta \in V_k(\bar{\theta}) \cap \Theta} l_{1,r}(\theta) \right)$ converges while decreasing to $E(l_{1,r}(\bar{\theta}))$ as $k \rightarrow \infty$. Hence, (7.3) follows from Lemma 7.2, ii). ■

In view of Lemmas 7.1-7.3, we have shown that for all $\bar{\theta} \neq \theta_0$ there exists a neighborhood $V(\bar{\theta})$ such that

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in V_k(\bar{\theta}) \cap \Theta} \tilde{L}_{n,r}(\theta) < \limsup_{n \rightarrow \infty} \tilde{L}_{n,r}(\theta_0) = \limsup_{n \rightarrow \infty} L_{n,r}(\theta_0) = E(l_{1,r}(\theta_0)).$$

Thus, from standard arguments the proof of Theorem 3.1 is completed while using assumption **A5** of compactness of Θ .

7.2. Proof of Theorem 3.2

By **A7** and Theorem 3.1 we know that $\hat{\theta}_r$ cannot be at the boundary of Θ for n sufficiently large. Hence, a Taylor expansion of $\frac{\partial L_{n,r}(\hat{\theta}_r)}{\partial \theta}$ at θ_0 yields

$$\begin{aligned} 0 &= \sqrt{n} \frac{\partial \tilde{L}_{n,r}(\hat{\theta}_r)}{\partial \theta} = \sqrt{n} \frac{\partial L_{n,r}(\hat{\theta}_r)}{\partial \theta} + \sqrt{n} \left(\frac{\partial \tilde{L}_{n,r}(\theta)}{\partial \theta} - \frac{\partial L_{n,r}(\theta)}{\partial \theta} \right) \\ &= \sqrt{n} \frac{\partial L_{n,r}(\theta_0)}{\partial \theta} + \sqrt{n} \frac{\partial^2 L_{n,r}(\theta^*)}{\partial \theta \partial \theta'} \left(\hat{\theta}_r - \theta_0 \right) + \sqrt{n} \left(\frac{\partial \tilde{L}_{n,r}(\theta)}{\partial \theta} - \frac{\partial L_{n,r}(\theta)}{\partial \theta} \right), \end{aligned} \quad (7.4)$$

for a certain θ^* between $\hat{\theta}_r$ and θ_0 . In view of (7.4), the proof of Theorem 3.2 is based on the following three lemmas. Lemma 7.4 shows that the last term in (7.4) is *a.s.* negligible

as $n \rightarrow \infty$. Lemma 7.5 establishes the convergence in law of the first term of (7.4) using a martingale central limit theorem while Lemma 7.6 shows the *a.s.* convergence of the matrix in the second term of (7.4).

Lemma 7.4 Under **A0-A10** $\sqrt{n} \sup_{\theta \in \Theta} \left\| \frac{\partial \tilde{L}_{n,r}(\theta)}{\partial \theta} - \frac{\partial L_{n,r}(\theta)}{\partial \theta} \right\| \xrightarrow[n \rightarrow \infty]{a.s.} 0$.

Proof Using **A2** and **A6** it follows that

$$\sqrt{n} \sup_{\theta \in \Theta} \left\| \frac{\partial \tilde{L}_{n,r}(\theta)}{\partial \theta} - \frac{\partial L_{n,r}(\theta)}{\partial \theta} \right\| = \frac{1}{\sqrt{n}} \sup_{\theta \in \Theta} \left\| \sum_{t=1}^n \left[\frac{\partial}{\partial \theta} \left(\log \left(\frac{r}{r + \tilde{\lambda}_t(\theta)} \right) + X_t \log \left(\frac{\tilde{\lambda}_t(\theta)}{r + \tilde{\lambda}_t(\theta)} \right) \right) - \frac{\partial}{\partial \theta} \left(\log \left(\frac{r}{r + \lambda_t(\theta)} \right) + X_t \log \left(\frac{\lambda_t(\theta)}{r + \lambda_t(\theta)} \right) \right) \right] \right\| \leq \frac{1}{\sqrt{n}} \sum_{t=1}^n \left(c_t + a_t d_t + X_t \left(\frac{c_t}{c r} + \frac{(a_t + b_t) d_t}{c^2 r^2} \right) \right) \xrightarrow[n \rightarrow \infty]{a.s.} 0.$$

Lemma 7.5 Under **A8-A9**, $\sqrt{n} \frac{\partial L_{n,r}(\theta_0)}{\partial \theta} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, I_r)$.

Proof It is clear that $\left\{ \sqrt{n} \frac{\partial L_{n,r}(\theta_0)}{\partial \theta}, t \in \mathbb{Z} \right\}$ is a martingale with respect to $\{\mathcal{F}_t, t \in \mathbb{Z}\}$

where

$$\sqrt{n} \frac{\partial L_{n,r}(\theta_0)}{\partial \theta} = \sum_{t=1}^n \frac{1}{\sqrt{n}} \frac{\partial l_{t,r}(\theta_0)}{\partial \theta} \quad \text{and} \quad \frac{\partial l_{t,r}(\theta_0)}{\partial \theta} = \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{X_t - \lambda_t(\theta_0)}{\lambda_t(\theta_0)(1 + \lambda_t(\theta_0))}.$$

By **A8-A9** we have

$$E \left(\frac{\partial l_{t,r}(\theta_0)}{\partial \theta} \frac{\partial l_{t,r}(\theta_0)}{\partial \theta'} \right) = E \left(\frac{v_t(\theta_0)}{\lambda_t^2(\theta_0)(1 + \lambda_t(\theta_0))^2} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \lambda_t(\theta_0)}{\partial \theta'} \right) = I_r.$$

Thus, Lemma 7.4 follows from the martingale central limit theorem (e.g. Billingsley, 2008; Hall and Heyde, 1980). ■

Lemma 7.6 Under **A8-A10**, $\frac{\partial^2 L_{n,r}(\theta^*)}{\partial \theta \partial \theta'} \xrightarrow[n \rightarrow \infty]{a.s.} J_r$.

Proof Let $V_k(\theta_0)$ ($k \in \mathbb{N}^*$) be the open ball with center θ_0 and radius $1/k$ where k is supposed large enough so that $V_k(\theta_0)$ is contained in $V(\theta_0)$ defined by **A10**. Assume that n is large enough so that θ^* belongs to $V_k(\theta_0)$. By stationarity and ergodicity of

$$\left\{ \sup_{\theta \in V_k(\theta_0)} \left| \frac{\partial^2 l_{t,r}(\theta)}{\partial \theta_i \partial \theta_j} - E \left(\frac{\partial^2 l_{t,r}(\theta_0)}{\partial \theta_i \partial \theta_j} \right) \right| \right\},$$

we have

$$\begin{aligned} \left| \frac{\partial^2 L_{n,r}(\theta^*)}{\partial \theta_i \partial \theta_j} - J_r(i, j) \right| &= \left| \frac{\partial^2 L_{n,r}(\theta^*)}{\partial \theta_i \partial \theta_j} - E \left(\frac{\partial^2 L_{n,r}(\theta_0)}{\partial \theta_i \partial \theta_j} \right) \right| = \frac{1}{n} \left| \sum_{t=1}^n \frac{\partial^2 l_{t,r}(\theta^*)}{\partial \theta_i \partial \theta_j} - E \left(\frac{\partial^2 l_{t,r}(\theta_0)}{\partial \theta_i \partial \theta_j} \right) \right| \\ &\leq \frac{1}{n} \sum_{t=1}^n \sup_{\theta \in V_k(\theta_0)} \left| \frac{\partial^2 l_{t,r}(\theta)}{\partial \theta_i \partial \theta_j} - E \left(\frac{\partial^2 l_{t,r}(\theta_0)}{\partial \theta_i \partial \theta_j} \right) \right| \\ &\xrightarrow[n \rightarrow \infty]{a.s.} E \left(\sup_{\theta \in V_k(\theta_0)} \left| \frac{\partial^2 l_{t,r}(\theta)}{\partial \theta_i \partial \theta_j} - E \left(\frac{\partial^2 l_{t,r}(\theta_0)}{\partial \theta_i \partial \theta_j} \right) \right| \right). \end{aligned}$$

In view of **A10**, the Lebesgue dominated convergence theorem yields

$$\lim_{k \rightarrow \infty} E \left(\sup_{\theta \in V_k(\theta_0)} \left| \frac{\partial^2 l_{t,r}(\theta)}{\partial \theta_i \partial \theta_j} - E \left(\frac{\partial^2 l_{t,r}(\theta_0)}{\partial \theta_i \partial \theta_j} \right) \right| \right) = E \left(\lim_{k \rightarrow \infty} \sup_{\theta \in V_k(\theta_0)} \left| \frac{\partial^2 l_{t,r}(\theta)}{\partial \theta_i \partial \theta_j} - E \left(\frac{\partial^2 l_{t,r}(\theta_0)}{\partial \theta_i \partial \theta_j} \right) \right| \right) = 0,$$

which completes the proof of the lemma. \blacksquare

7.3. Proof of Theorem 3.3

i) Proof of (3.14a)

It suffices to prove the strong consistency of $\widehat{\gamma}_1$. From (3.12a) and **Step 2** (cf. Algorithm 3.1) we have

$$\begin{aligned} \widehat{\gamma}_1 - \gamma_0 &= \frac{1}{n} \sum_{t=1}^n \frac{\widehat{u}_{1t}}{\widehat{\lambda}_t^2} \\ &= \frac{1}{n} \sum_{t=1}^n \frac{u_t}{\lambda_t^2} + \frac{1}{n} \sum_{t=1}^n \frac{\widehat{u}_{1t} - u_t}{\lambda_t^2} + \frac{1}{n} \sum_{t=1}^n \left(\frac{1}{\widehat{\lambda}_{1t}^2} - \frac{1}{\lambda_t^2} \right) u_t + \frac{1}{n} \sum_{t=1}^n \left(\frac{1}{\widehat{\lambda}_{1t}^2} - \frac{1}{\lambda_t^2} \right) (\widehat{u}_{1t} - u_t), \end{aligned} \quad (7.5)$$

where u_t is given by (3.11a) and \widehat{u}_{1t} is defined as in (3.12b) while replacing $\widehat{\lambda}_t$ by $\widehat{\lambda}_{1t} = \widetilde{\lambda}_t(\widehat{\theta}_{r^*})$. By the ergodic theorem, the first term in the right hand side of (7.5) satisfies the following limiting result

$$\frac{1}{n} \sum_{t=1}^n \frac{u_t}{\lambda_t^2} \xrightarrow[n \rightarrow \infty]{a.s.} E \left(\frac{u_t}{\lambda_t^2} \right) = E \left(\frac{1}{\lambda_t^2} E(u_t | \mathcal{F}_{t-1}) \right) = 0.$$

So it remains to prove the following three lemmas which show that the last three terms in the right hand side of (7.5) vanish asymptotically *a.s.*

Lemma 7.7 Under **A0-A10** $\frac{1}{n} \sum_{t=1}^n \frac{\widehat{u}_{1t} - u_t}{\lambda_t^2} = o_{a.s.}(1)$.

Proof Note that

$$\begin{aligned} \widehat{u}_{1t} - u_t &= \left(X_t - \widehat{\lambda}_{1t} \right)^2 - \left(X_t - \lambda_t \right)^2 + \left(\lambda_t + \frac{1}{r_0} \lambda_t^2 \right) - \left(\widehat{\lambda}_{1t} + \frac{1}{r_0} \widehat{\lambda}_{1t}^2 \right) \\ &= \left(1 - \frac{1}{r_0} \right) \left(\left(\widetilde{\lambda}_t^2(\widehat{\theta}_{r^*}) - \lambda_t^2(\widehat{\theta}_{r^*}) + \left(\lambda_t^2(\widehat{\theta}_{r^*}) - \lambda_t^2(\theta_0) \right) \right) \right) \\ &\quad - (2X_t + 1) \left(\widetilde{\lambda}_t(\widehat{\theta}_{r^*}) - \lambda_t(\widehat{\theta}_{r^*}) + \left(\lambda_t(\widehat{\theta}_{r^*}) - \lambda_t(\theta_0) \right) \right). \end{aligned} \quad (7.6)$$

Hence,

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n \frac{\widehat{u}_{1t} - u_t}{\lambda_t^2} &= \left(1 - \frac{1}{r_0}\right) \left[\frac{1}{n} \sum_{t=1}^n \frac{\widetilde{\lambda}_t^2(\widehat{\theta}_{r^*}) - \lambda_t^2(\widehat{\theta}_{r^*})}{\lambda_t^2(\theta_0)} + \frac{1}{n} \sum_{t=1}^n \frac{\lambda_t^2(\widehat{\theta}_{r^*}) - \lambda_t^2(\theta_0)}{\lambda_t^2(\theta_0)} \right] - \\ &\left[\frac{1}{n} \sum_{t=1}^n \frac{(2X_t+1)(\widetilde{\lambda}_t(\widehat{\theta}_{r^*}) - \lambda_t(\widehat{\theta}_{r^*}))}{\lambda_t^2(\theta_0)} + \frac{1}{n} \sum_{t=1}^n \frac{(2X_t+1)(\lambda_t(\widehat{\theta}_{r^*}) - \lambda_t(\theta_0))}{\lambda_t^2(\theta_0)} \right]. \end{aligned} \quad (7.7)$$

We will show that each term in the right hand side of (7.7) is $o_{a.s.}(1)$ as $n \rightarrow \infty$. In view of **A2** we have $\left| \widetilde{\lambda}_t(\widehat{\theta}_{r^*}) - \lambda_t(\widehat{\theta}_{r^*}) \right| \leq a_t \xrightarrow[t \rightarrow \infty]{a.s.} 0$. By **A1** we know that $\lambda_t^{-2} < c^{-2}$ and under **A3** it follows that

$$E \left(\frac{2X_t+1}{\lambda_t^2(\theta_0)} \right) = E \left(\lambda_t^{-2} E(2X_t + 1 | \mathcal{F}_{t-1}) \right) \leq c^{-2} E(2X_t + 1) < \infty.$$

Therefore, $\frac{(2X_t+1)(\widetilde{\lambda}_t(\widehat{\theta}_{r^*}) - \lambda_t(\widehat{\theta}_{r^*}))}{\lambda_t^2} \xrightarrow[t \rightarrow \infty]{a.s.} 0^1$ and by the Césaro lemma

$$\frac{1}{n} \sum_{t=1}^n \frac{(2X_t+1)(\lambda_t(\widehat{\theta}_{r^*}) - \lambda_t(\theta_0))}{\lambda_t^2(\theta_0)} \xrightarrow[n \rightarrow \infty]{a.s.} 0.$$

Now under **A1** and **A6** we have $\frac{|\widetilde{\lambda}_t^2(\widehat{\theta}_{r^*}) - \lambda_t^2(\widehat{\theta}_{r^*})|}{\lambda_t^2(\theta_0)} \leq c^{-2} b_t \xrightarrow[t \rightarrow \infty]{a.s.} 0$, so by the Césaro lemma it follows that $\frac{1}{n} \sum_{t=1}^n \frac{\widetilde{\lambda}_t^2(\widehat{\theta}_{r^*}) - \lambda_t^2(\widehat{\theta}_{r^*})}{\lambda_t^2} \xrightarrow[n \rightarrow \infty]{a.s.} 0$. Next we show that

$$\frac{1}{n} \sum_{t=1}^n \frac{(2X_t+1)(\lambda_t(\widehat{\theta}_{r^*}) - \lambda_t(\theta_0))}{\lambda_t^2(\theta_0)} \xrightarrow[n \rightarrow \infty]{a.s.} 0. \quad (7.8)$$

A Taylor expansion of $\lambda_t(\widehat{\theta}_{r^*})$ around θ_0 gives

$$\lambda_t(\widehat{\theta}_{r^*}) - \lambda_t(\theta_0) = \frac{\partial \lambda_t(\bar{\theta})}{\partial \theta'}, \quad (7.9)$$

where $\bar{\theta}$ is between $\widehat{\theta}_{r^*}$ and θ_0 . Under **A9** we know that $E \left(\frac{\partial \lambda_t(\bar{\theta})}{\partial \theta'} \frac{(2X_t+1)}{\lambda_t^2(\theta_0)} \right) < \infty$. By the strong consistency of $\widehat{\theta}_{r^*}$ it follows that $(\widehat{\theta}_{r^*} - \theta_0) \frac{1}{n} \sum_{t=1}^n \frac{(2X_t+1) \frac{\partial \lambda_t(\bar{\theta})}{\partial \theta'}}{\lambda_t^2(\theta_0)} \xrightarrow[n \rightarrow \infty]{a.s.} 0$. Thus it remains to show that

$$\frac{1}{n} \sum_{t=1}^n \frac{\lambda_t^2(\widehat{\theta}_{r^*}) - \lambda_t^2(\theta_0)}{\lambda_t^2(\theta_0)} \xrightarrow[n \rightarrow \infty]{a.s.} 0.$$

¹We used the fact that if $E(X) < \infty$ and $Z_t \xrightarrow[t \rightarrow \infty]{a.s.} 0$ then $Z_t X \xrightarrow[t \rightarrow \infty]{a.s.} 0$, see Franq and Zakoain (2010) for a proof.

By a Taylor expansion of $\lambda_t^2(\widehat{\theta}_{r^*})$ around θ_0 we have

$$\lambda_t^2(\widehat{\theta}_{r^*}) - \lambda_t^2(\theta_0) = 2\lambda_t(\bar{\theta}) \frac{\partial \lambda_t(\bar{\theta})}{\partial \theta'} (\widehat{\theta}_{r^*} - \theta_0), \quad (7.10)$$

where $\bar{\theta}$ is between $\widehat{\theta}_{r^*}$ and θ_0 . Since by **A9** $E\left(\frac{\lambda_t(\bar{\theta})}{\lambda_t^2(\theta_0)} \frac{\partial \lambda_t(\bar{\theta})}{\partial \theta'}\right) < \infty$, the strong consistency of $\widehat{\theta}_{r^*}$ yields $(\widehat{\theta}_{r^*} - \theta_0) \frac{1}{n} \sum_{t=1}^n \frac{2\lambda_t(\bar{\theta})}{\lambda_t^2(\theta_0)} \frac{\partial \lambda_t(\bar{\theta})}{\partial \theta'} \xrightarrow[n \rightarrow \infty]{a.s.} 0$, so the conclusion follows.

Lemma 7.8 Under **A0-A10** $\frac{1}{n} \sum_{t=1}^n \left(\frac{1}{\widehat{\lambda}_{1t}^2} - \frac{1}{\lambda_t^2}\right) u_t = o_{a.s.}(1)$.

Proof We have

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n \left(\frac{1}{\widehat{\lambda}_{1t}^2} - \frac{1}{\lambda_t^2}\right) u_t &= \frac{1}{n} \sum_{t=1}^n \left(\frac{1}{\widetilde{\lambda}_t^2(\widehat{\theta}_{r^*})} - \frac{1}{\lambda_t^2(\theta_0)}\right) u_t \\ &= \frac{1}{n} \sum_{t=1}^n \left(\frac{1}{\widetilde{\lambda}_t^2(\widehat{\theta}_{r^*})} - \frac{1}{\lambda_t^2(\widehat{\theta}_{r^*})}\right) u_t + \frac{1}{n} \sum_{t=1}^n \left(\frac{1}{\lambda_t^2(\widehat{\theta}_{r^*})} - \frac{1}{\lambda_t^2(\theta_0)}\right) u_t \\ &= \frac{1}{n} \sum_{t=1}^n \left(\frac{\lambda_t^2(\widehat{\theta}_{r^*}) - \widetilde{\lambda}_t^2(\widehat{\theta}_{r^*})}{\widetilde{\lambda}_t^2(\widehat{\theta}_{r^*}) \lambda_t^2(\widehat{\theta}_{r^*})}\right) u_t + \frac{1}{n} \sum_{t=1}^n \left(\frac{1}{\lambda_t^2(\widehat{\theta}_{r^*})} - \frac{1}{\lambda_t^2(\theta_0)}\right) u_t. \end{aligned} \quad (7.11)$$

The first term in the right hand side of (7.11) is $o_{a.s.}(1)$ since by (3.11a) and **A8**, $E(u_t) < \infty$ and from **A1** and **A6**

$$\frac{|\lambda_t^2(\widehat{\theta}_{r^*}) - \widetilde{\lambda}_t^2(\widehat{\theta}_{r^*})|}{\widetilde{\lambda}_t^2(\widehat{\theta}_{r^*}) \lambda_t^2(\widehat{\theta}_{r^*})} \leq c^{-4} b_t \xrightarrow[t \rightarrow \infty]{a.s.} 0.$$

It remains to show that

$$\frac{1}{n} \sum_{t=1}^n \left(\frac{1}{\lambda_t^2(\widehat{\theta}_{r^*})} - \frac{1}{\lambda_t^2(\theta_0)}\right) u_t \xrightarrow[n \rightarrow \infty]{a.s.} 0. \quad (7.12)$$

A Taylor expansion of $\frac{1}{\lambda_t^2(\widehat{\theta}_{r^*})}$ around θ_0 gives

$$\frac{1}{\lambda_t^2(\widehat{\theta}_{r^*})} - \frac{1}{\lambda_t^2(\theta_0)} = -\frac{2}{\lambda_t^3(\bar{\theta})} \frac{\partial \lambda_t(\bar{\theta})}{\partial \theta'} (\widehat{\theta}_{r^*} - \theta_0), \quad (7.13)$$

where $\bar{\theta}$ is between $\widehat{\theta}_{r^*}$ and θ_0 . Thus (7.12) follows from **A1**, **A10**, the strong consistency of $\widehat{\theta}_{r^*}$ and the Césaro lemma.

Lemma 7.9 Under **A0-A10** $\frac{1}{n} \sum_{t=1}^n \left(\frac{1}{\widehat{\lambda}_{1t}^2} - \frac{1}{\lambda_t^2}\right) (\widehat{u}_{1t} - u_t) = o_{a.s.}(1)$.

Proof From Lemma 7.7 we know that $\widehat{u}_{1t} - u_t \xrightarrow[t \rightarrow \infty]{a.s.} 0$ and from Lemma 7.8 we already shown that $\frac{1}{\widehat{\lambda}_{1t}^2} - \frac{1}{\lambda_t^2} \xrightarrow[t \rightarrow \infty]{a.s.} 0$ so the conclusion follows from the Césaro lemma.

ii) Proof of (3.14b)

We rewrite (7.5) as follows

$$\begin{aligned}\sqrt{n}(\widehat{\gamma}_1 - \gamma_0) &= \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\widehat{u}_{1t}}{\lambda_t^2} = \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{u_t}{\lambda_t^2} + \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\widehat{u}_{1t} - u_t}{\lambda_t^2} + \\ &\frac{1}{\sqrt{n}} \sum_{t=1}^n \left(\frac{1}{\widehat{\lambda}_{1t}^2} - \frac{1}{\lambda_t^2} \right) u_t + \frac{1}{\sqrt{n}} \sum_{t=1}^n \left(\frac{1}{\widehat{\lambda}_{1t}^2} - \frac{1}{\lambda_t^2} \right) (\widehat{u}_{1t} - u_t).\end{aligned}\quad (7.14)$$

The first term in the right hand side of (7.14) is clearly a martingale with respect to $\{\mathcal{F}_t, t \in \mathbb{Z}\}$. By the martingale central limit theorem (e.g. Hall and Heyde, 1980, Corollary 3.1) it follows that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{u_t}{\lambda_t^2} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N\left(0, E\left(\frac{u_t^2}{\lambda_t^4}\right)\right).$$

Therefore, it remains to establish the following three lemmas which show that the last three terms in the right hand side of (7.14) are asymptotically negligible in probability.

Lemma 7.10 *Under **A0-A10** $\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\widehat{u}_{1t} - u_t}{\lambda_t^2} = o_p(1)$.*

Proof Let us rewrite (7.7) as follows

$$\begin{aligned}\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\widehat{u}_{1t} - u_t}{\lambda_t^2} &= -\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{(2X_t+1)(\widetilde{\lambda}_t(\widehat{\theta}_{r^*}) - \lambda_t(\widehat{\theta}_{r^*}))}{\lambda_t^2(\theta_0)} + \left(1 - \frac{1}{r_0}\right) \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\widetilde{\lambda}_t^2(\widehat{\theta}_{r^*}) - \lambda_t^2(\widehat{\theta}_{r^*})}{\lambda_t^2(\theta_0)} - \\ &\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{(2X_t+1)(\lambda_t(\widehat{\theta}_{r^*}) - \lambda_t(\theta_0))}{\lambda_t^2(\theta_0)} + \left(1 - \frac{1}{r_0}\right) \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\lambda_t^2(\widehat{\theta}_{r^*}) - \lambda_t^2(\theta_0)}{\lambda_t^2(\theta_0)}.\end{aligned}\quad (7.15)$$

Under **A6** we know that $\left| \widetilde{\lambda}_t(\widehat{\theta}_{r^*}) - \lambda_t(\widehat{\theta}_{r^*}) \right| \leq a_t$, so

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{(2X_t+1) \left| \widetilde{\lambda}_t(\widehat{\theta}_{r^*}) - \lambda_t(\widehat{\theta}_{r^*}) \right|}{\lambda_t^2(\theta_0)} \leq \frac{1}{\sqrt{n}} \sum_{t=1}^n a_t \frac{(2X_t+1)}{\lambda_t^2(\theta_0)} \xrightarrow[n \rightarrow \infty]{a.s.} 0.$$

By the same argument, we have under **A6**

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\left| \widetilde{\lambda}_t^2(\widehat{\theta}_{r^*}) - \lambda_t^2(\widehat{\theta}_{r^*}) \right|}{\lambda_t^2(\theta_0)} \leq \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{b_t}{\lambda_t^2(\theta_0)} \xrightarrow[n \rightarrow \infty]{a.s.} 0.$$

The third term in the right hand side of (7.15) is $o_p(1)$ using the Taylor expansion (7.9) and the \sqrt{n} -consistency of $\widehat{\theta}_{r^*}$. The same argument, while using the Taylor expansion (7.10), shows that $\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\lambda_t^2(\widehat{\theta}_{r^*}) - \lambda_t^2(\theta_0)}{\lambda_t^2(\theta_0)} = o_p(1)$.

Lemma 7.11 *Under **A0-A10** $\frac{1}{\sqrt{n}} \sum_{t=1}^n \left(\frac{1}{\widehat{\lambda}_{1t}^2} - \frac{1}{\lambda_t^2} \right) u_t = o_p(1)$.*

Proof From (7.11) we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \left(\frac{1}{\tilde{\lambda}_{1t}^2} - \frac{1}{\lambda_t^2} \right) u_t = \frac{1}{\sqrt{n}} \sum_{t=1}^n \left(\frac{\lambda_t^2(\hat{\theta}_{r^*}) - \tilde{\lambda}_t^2(\hat{\theta}_{r^*})}{\tilde{\lambda}_t^2(\hat{\theta}_{r^*}) \lambda_t^2(\hat{\theta}_{r^*})} \right) u_t + \frac{1}{\sqrt{n}} \sum_{t=1}^n \left(\frac{1}{\lambda_t^2(\hat{\theta}_{r^*})} - \frac{1}{\lambda_t^2(\theta_0)} \right) u_t.$$

Under **A1** we know that $\frac{|\lambda_t^2(\hat{\theta}_{r^*}) - \tilde{\lambda}_t^2(\hat{\theta}_{r^*})|}{\tilde{\lambda}_t^2(\hat{\theta}_{r^*}) \lambda_t^2(\hat{\theta}_{r^*})} \leq c^{-4} b_t$, so in view of **A6** we find

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{|(\lambda_t^2(\hat{\theta}_{r^*}) - \tilde{\lambda}_t^2(\hat{\theta}_{r^*})) u_t|}{\tilde{\lambda}_t^2(\hat{\theta}_{r^*}) \lambda_t^2(\hat{\theta}_{r^*})} \leq \frac{1}{\sqrt{n}} \sum_{t=1}^n c^{-4} b_t |u_t| \xrightarrow[n \rightarrow \infty]{a.s.} 0.$$

Using the Taylor expansion (7.13) we therefore get

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \left(\frac{1}{\lambda_t^2(\hat{\theta}_{r^*})} - \frac{1}{\lambda_t^2(\theta_0)} \right) u_t = \frac{-2(\hat{\theta}_{r^*} - \theta_0)'}{\sqrt{n}} \sum_{t=1}^n \frac{u_t}{\lambda_t^3(\bar{\theta})} \frac{\partial \lambda_t(\bar{\theta})}{\partial \theta},$$

and the conclusion follows from the \sqrt{n} -consistency of $\hat{\theta}_{r^*}$, the finiteness of $E \left(\frac{1}{\lambda_t^3(\bar{\theta})} \frac{\partial \lambda_t(\bar{\theta})}{\partial \theta} \right)$, the ergodic theorem and the fact that $E(u_t) = 0$.

Lemma 7.12 Under **A0-A10** $\frac{1}{\sqrt{n}} \sum_{t=1}^n \left(\frac{1}{\tilde{\lambda}_{1t}^2} - \frac{1}{\lambda_t^2} \right) (\hat{u}_{1t} - u_t) = o_p(1)$.

Proof The result easily follows while combining Lemma 7.10 and Lemma 7.11.

iii) Proof of (3.14c) Result (3.14c) is an obvious consequence of the strong consistency of $\hat{\theta}_{r^*}$ (cf. (3.4)) for all $r^* > 0$.

iv) Proof of (3.14d) From the consistency of \hat{r}_1 and the \sqrt{n} -consistency of $\hat{\theta}_r$ for all $r > 0$ we have

$$\sqrt{n} (\hat{\theta}_{\hat{r}_1} - \theta_0) = \sqrt{n} (\hat{\theta}_{r_0} - \theta_0) + \sqrt{n} (\hat{\theta}_{\hat{r}_1} - \hat{\theta}_{r_0}) = \sqrt{n} (\hat{\theta}_{r_0} - \theta_0) + o_p(1),$$

so the result follows from Theorem 3.2 while replacing r by r_0 (cf. (3.6)) and using the fact that, under (4.11), $I_{r_0} = J_{r_0}$. ■

Acknowledgements We are deeply grateful to the Editor-in-Chief, a Co-Editor and two reviewers for their relevant comments and useful suggestions that have allowed us to improve upon the earlier versions of the manuscript. We also thank Prof. F. Zhu and Prof. M.L. Diop for providing us with the "*Polio data*" and the "*Transaction data*".

References

- [1] Ahmad A, Francq C. 2016. Poisson qmle of count time series models. *Journal of Time Series Analysis* **37**: 291-314.
- [2] Aknouche A, Bendjeddou S. 2017. Estimateur du quasi-maximum de vraisemblance géométrique d'une classe générale de modèles de séries chronologiques à valeurs entières. *Comptes Rendus Mathématique* **355**: 99-104.
- [3] Aknouche A, Bendjeddou S, Touche N. 2017. Negative binomial quasi-likelihood inference for general integer-valued time series models. *Journal of Time Series Analysis, forthcoming*, DOI: 10.1111/jtsa.12277.
- [4] Al-Osh MA, Alzaid AA. 1987. First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis* **8**: 261-275.
- [5] Benjamin MA, Rigby RA, Stasinopoulos DM. 2003. Generalized autoregressive moving average models. *Journal of the American Statistical Association* **98**: 214-223.
- [6] Billingsley P. 2008. *Probability and measure* (3rd edn). John Wiley.
- [7] Bourguignon M. 2016. Poisson-geometric INAR(1) process for modeling count time series with overdispersion. *Statistica Neerlandica* **70**: 176-192.
- [8] Cameron AC, Trivedi PK. 1986. Econometric models based on count data: Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics* **1**: 29-53.
- [9] Cameron C, Trivedi P. 2013. *Regression analysis of count data* (2nd edn). New York: Cambridge University Press.
- [10] Chen CWS, So M, Li JC, Sriboonchitta S. 2016. Autoregressive conditional negative binomial model applied to over-dispersed time series of counts. *Statistical Methodology* **31**: 73-90.

- [11] Christou V, Fokianos K. 2014. Quasi-likelihood inference for negative binomial time series models. *Journal of Time Series Analysis* **35**: 55-78.
- [12] Christou V, Fokianos K. 2015. Estimation and testing linearity for non-linear mixed Poisson autoregressions. *Electronic Journal of Statistics* **9**: 1357-1377.
- [13] Davis RA, Liu H. 2016. Theory and inference for a class of observation-driven models with application to time series of counts. *Statistica Sinica* **26**: 1673-1707.
- [14] Davis R, Wu R. 2009. A negative binomial model for time series of counts. *Biometrika* **96**: 735-749.
- [15] Davis RA, Dunsmuir WTM, Wang Y. 1999. *Modelling time series of count data*. In Asymptotics, Nonparametrics and Time Series (edn, Subir Ghosh). New York: Taylor and Francis.
- [16] Davis RA, Holan SH, Lund R, Ravishanker N. 2016. *Handbook of discrete-valued time series*. Chapman and Hall.
- [17] Diop ML, Kengne W. 2017. Testing for parameter change in general integer-valued time series. *Journal of Time Series Analysis* **38**: 880-894.
- [18] Douc R, Doukhan P, 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.
- [19] Doukhan P, Fokianos K, Tjøstheim D. 2012. On weak dependence conditions for Poisson autoregressions. *Statistics and Probability Letters* **82**: 942-948.
- [20] Doukhan P, Kengne W. 2015. Inference and testing for structural change in general Poisson autoregressive models. *Electronic Journal of Statistics* **9**: 1267-1314.
- [21] Doukhan P, Wintenberger O. 2008. Weakly dependent chains with infinite memory. *Stochastic Processes and their Applications* **118**: 1997-2013.

- [22] Efron B. 1986. Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association* **81**: 709-721.
- [23] Ferland R, Latour A, Oraichi D. 2006. Integer-valued GARCH process. *Journal of Time Series Analysis* **27**: 923-942.
- [24] Fokianos K. 2012. *Count time series models. Handbook of Statistics. Time Series Analysis: Methods and Applications* **30**: 315-348.
- [25] Fokianos K, Rahbek A, Tjøstheim D. 2009. Poisson autoregression. *Journal of the American Statistical Association* **140**: 1430-1439.
- [26] Fokianos K. Tjøstheim D. 2011. Log-linear Poisson autoregression. *Journal of Multivariate Analysis* **102**: 563-578.
- [27] Francq C, Zakoïan JM. 2010. *GARCH models: Structure, statistical inference and financial applications*. New York: Wiley.
- [28] Franke J. 2010. Weak dependence of functional INGARCH processes. *Technical report, University of Kaiserslautern*.
- [29] Heinen A. 2003. Modelling time series count data: an autoregressive conditional Poisson model. *Available at SSRN 1117187*.
- [30] Gonçalves E, Mendes-Lopes N, Silva F. 2015. Infinitely divisible distributions in integer-valued GARCH models. *Journal of Time Series Analysis* **36**: 503-527.
- [31] Gouriéroux C, Monfort A, Trognon A. 1984a. Pseudo maximum likelihood methods: Theory. *Econometrica* **52**: 681-700.
- [32] Gouriéroux C, Monfort A, Trognon A. 1984b. Pseudo maximum likelihood methods: Applications to Poisson models. *Econometrica* **52**: 701-720.
- [33] Grunwald G, Hyndman R, Tedesco L, Tweedie R. 2000. Non-Gaussian conditional linear AR(1) models. *Australian and New Zealand Journal of Statistics* **42**: 479-495.

- [34] Hall P, Heyde CC. 1980. *Martingale Limit Theory and its Applications*. New York: Academic Press.
- [35] Kedem B, Fokianos K. 2002. *Regression models for time series analysis*. New York: Wiley.
- [36] Kengne W. 2015. Sequential change-point detection in Poisson autoregressive models. *Journal de la Société Française de Statistique* **156**: 98-112.
- [37] McCullagh P, Nelder JA. 1989. *Generalized Linear Models* (2nd edn). London: Chapman and Hall.
- [38] McKenzie E. 1985. Some simple models for discrete variate time series. *Water Resources Bulletin* **21**: 645-650.
- [39] McKenzie E. 2003. *Discrete variate time series*, in Handbook of statistics. Amsterdam: Elsevier Science.
- [40] Nelder JA, Wedderburn RW. 1972. Generalized Linear Models. *Journal of the Royal Statistical Society, Series A* **135**: 370-384.
- [41] Neumann MH. 2011. Absolute regularity and ergodicity of Poisson count processes. *Bernoulli* **17**: 1268-1284.
- [42] Rydberg TH, Shephard N. 2000. BIN models for trade-by-trade data. Modelling the number of trades in a fixed interval of time. In World Conference Econometric Society, 2000, Seattle. *Contributed Paper 0740*.
- [43] Silva ME. 2015. Modelling time series of counts: an INAR approach. *Textos de Matemática* **47**: 107-121.
- [44] Steutel FW, Van Harn K. 1979. Discrete analogues of self-decomposability and stability. *Annals of Probability* **5**: 893-899.

- [45] Wedderburn RW. 1974. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* **61**: 439-447.
- [46] White H. 1982. Maximum likelihood of misspecified models. *Econometrica* **50**: 1-25.
- [47] Wooldridge JM. 1997. *Quasi-likelihood methods for count data*. In *Handbook of Applied Econometrics*. Oxford: Blackwell.
- [48] Wooldridge JM. 2002. *Econometric analysis of cross section and panel data*. Cambridge, MA, MIT Press.
- [49] Zeger SL. 1988. A regression model for time series of counts. *Biometrika* **75**: 621-629.
- [50] Zeger SL, Qaqish B. 1988. Markov regression models for time series: a quasi-likelihood approach. *Biometrics* **44**: 1019-1031.
- [51] Zhu F. 2011. A negative binomial integer-valued GARCH model. *Journal of Time Series Analysis* **32**: 54-67.
- [52] Zhu F. 2012a. Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models. *Journal of Mathematical Analysis and its Applications* **389**: 58-71.
- [53] Zhu F. 2012b. Zero-inflated Poisson and negative binomial integer-valued GARCH models. *Journal of Statistical Planning and Inference* **142**: 826-839.
- [54] Zhu F. 2012c. Modeling time series of counts with COM-Poisson INGARCH models. *Mathematical and Computer Modelling* **56**: 191-203.