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

Abdelhakim Aknouche and Sara Bendjeddou

Faculty of Mathematics, USTHB, Mathematics Department, Qassim University

6 December 2016

Online at https://mpra.ub.uni-muenchen.de/76574/
MPRA Paper No. 76574, posted 4 February 2017 08:48 UTC
Negative binomial quasi-likelihood inference for general integer-valued time series models

Abdelhakim Aknouche* and Sara Bendjeddou*

February 3, 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 GARCH model and the INAR(1) model. Applications to two real datasets are given.

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

*Faculty of Mathematics, University of Science and Technology Houari Boumediene. °Mathematics department, Qassim University.
1. Introduction

Integer-valued time series like 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 like generalized ARMA (GARMA) models, Poisson autoregression and especially Integer Generalized Conditional Heteroskedastic (INGARCH) models (e.g. Benjamin et al, 2003; Heinen, 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 we call INteger Generalized AutoRegression (henceforth INGAR) 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. For the INGAR 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) firstly, 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; Gourieroux et al, 1984a). ii) Secondly, it is asymptotically efficient when the true conditional distribution of the model is Poissonian. iii) Thirdly, 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 Gourieroux et al, 1984a). The latter proportionality between the conditional mean and conditional variance is usually called the Poisson Generalized Linear Model (henceforth GLM) assumption (or link function). However, despite these advantages, the Poisson distribution, which is known to be equidispersed, fit badly to overdispersed series that are frequently observed in practice. Therefore, it is likely that the P-QMLE does not reach its full asymptotic efficiency in the presence of overdispersed data. Thus a quasi-maximum likelihood (QML) estimate, which is calculated using an overdispersed likelihood while belonging to the linear exponential family would be an interesting complementary to the P-QMLE.

For the INGAR 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, which 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 INGAR 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 Gourieroux 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 INGAR model and the corresponding negative binomial QML criteria. Section 3 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 INGAR model. In particular, the Poisson INGARCH model, the negative binomial INGARCH model and the INAR(1) 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 \((\Omega, \mathcal{F}, P)\). The symbols \(\mathbb{Z} = \{ ..., -1, 0, 1, ... \}\), \(\mathbb{N} = \{ 0, 1, ... \}\) 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, \(\Gamma\) is the gamma function and
$x!$ is the factorial of $x$. When $r \in \mathbb{N}^*$ has to be a positive integer, the factor $\frac{\Gamma(x+r)}{2!\Gamma(r)}$ 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 (2013), the negative binomial-$K$ conditional distribution given a $\sigma$-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-$I$ conditional distribution corresponding to $K = 1$ and the negative binomial-$II$ model for which $K = 2$. Finally, the symbols $\xrightarrow{a.s.}, \xrightarrow{p}$ and $\xrightarrow{L}$ denote respectively almost sure convergence, convergence in probability and convergence in distribution as $n \to \infty$ while $o_p(1), o_{a.s.}(1)$ and $O_p$ are respectively: a term converging in probability to zero, a term converging almost surely ($a.s.$) to zero and a term bounded in probability as $n \to \infty$.

2. The INGAR model: 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 $\lambda : \mathbb{N}^\times \Theta \to (0, \infty)$. 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 $(\Omega, \mathcal{F}, P)$ with conditional expectation specified as follows

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

where $\mathcal{F}_t \subset \mathcal{F}$ is the $\sigma$-algebra generated by $\{X_t, X_{t-1}, \ldots\}$. 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)

$$X_t = \lambda(X_{t-1}, X_{t-2}, \ldots; \theta_0) + e_t, \quad t \in \mathbb{Z}.$$
Equation (2.2), which is driven by the \( \{F_t, t \in \mathbb{Z}\} \)-martingale difference \( \{e_t, t \in \mathbb{Z}\} \), appears to be an infinite generalized autoregression with integer-valued solution \( \{X_t, t \in \mathbb{Z}\} \). The term "generalized" refers to the general form of the function \( \lambda \), which may be linear or nonlinear. This is why the model is termed \textit{IN}teger \textit{G}eneralized \textit{A}uto\textit{R}egression (\textit{INGAR}). In fact, the \textit{INGAR} model (2.1)-(2.2) is quite general and encompasses many important classes of integer-valued time series models such as the (stable) \textit{P}oisson \textit{INGARCH} model (Heinen, 2003; Ferland et al, 2006), the general Poisson autoregression (Doukhan et al, 2012; Doukhan and Kengne, 2015; Kengne, 2015), the stable negative binomial \textit{INGARCH} model (Zhu, 2011; Christou and Fokianos, 2014; Davis and Liu, 2016; Diop and Kengne, 2016) and the \textit{INAR} model (Al-Osh and Alzaid, 1987).

Note that the generality of the \textit{INGAR} model (2.1) stems not only from the general form of the function \( \lambda(.) \) (see also Doukhan and Wintenberger, 2008), 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 \textit{link function} relating the conditional variance and the conditional mean of model (2.1), i.e.

\[
Var(X_t/F_{t-1}) = l(E(X_t/F_{t-1})), \tag{2.3}
\]

where \( l: (0, \infty) \to (0, \infty) \) is a positive real function. In the literature on \textit{generalized linear models} (e.g. Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989), such a link function is also called the \textit{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 \( \lambda_t \), which reduces to a special case of the general Poisson autoregression proposed by Doukhan et al (2012), the Poisson \textit{GLM} link function for model (2.1) is given by the linear form \( l(x) = x \). A more general linear link function

\[
l(x) = \left(1 + \frac{1}{r_0}\right) x, \text{ for some } r_0 > 0,
\]

is induced by the conditional negative binomial-\textit{I} conditional distribution, i.e. \( X_t/F_{t-1} \sim \)}
\[ NB \left( r_0 \lambda_t, \frac{r_0 \lambda_t}{r_0 + \lambda_t} \right), \quad r_0 > 0 \] (see Cameron and Trivedi, 1986 and Section 4.1 below). Furthermore, the link function implied by the negative binomial-\( II \) conditional distribution, that is \[ NB \left( r_0, \frac{r_0}{r_0 + \lambda_t} \right), \] is given by

\[ l(x) = x \left( 1 + x \frac{1}{r_0} \right), \quad r_0 > 0. \] (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 exhibited 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 \( \theta_0 \) using a series \( X_1, X_2, ..., X_n \ (n \in \mathbb{N}^*) \) generated from (2.1). When a negative binomial-\( II \) link function like (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 \( \theta_0 \).

**Case 2**: Equation (2.1) and the negative binomial-\( II \) GLM link function (2.3)-(2.4) are both specified so we have to estimate both \( \theta_0 \) and \( r_0 \).

A particularly important instance of **Case 2** appears when the full conditional distribution of the model is specified as a negative binomial-\( II \) one, i.e. \( X_t / \mathcal{F}_{t-1} \sim NB \left( r_0, \frac{r_0}{r_0 + \lambda_t} \right), \) where a special case is the negative binomial-\( II \) \( INGARCH \) model (see Davis and Liu, 2016; Zhu, 2011; Christou and Fokianos, 2014 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) like the \( INGARCH \) and \( INAR \) models, assumption **A0** may be expressed more explicitly as a stability condition on \( \theta_0 \) (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 \( \lambda \) in (2.1) for strict stationarity and ergodicity of the model.
Now, given a generic parameter $\theta \in \Theta$, the conditional mean function given by

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

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

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

be an observable proxy for $\lambda_t(\theta)$. The latter approximation serves in calculating various QMLE-type of $\theta_0$ we intend to study below.

### 3. Negative binomial QMLE’s of the INGAR model

This Section considers two negative binomial QMLE’s of the INGAR model (2.1) given a realization $X_1, \ldots, X_n$ of (2.1). To describe these estimates consider Case 2 of model (2.1)-(2.4) with unknown parameters $\theta_0$ 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 negative binomial-II conditional distribution, $\text{NB}(r; \frac{r}{r+\lambda_t(\theta)})$, is given by

$$\tilde{L}_{NB}(\theta, r) = \frac{1}{n} \sum_{t=1}^{n} \tilde{l}_t(\theta, r), \quad (3.1)$$

with $\tilde{l}_t(\theta, r) = r \log \left( \frac{r}{r+\tilde{\lambda}_t(\theta)} \right) + X_t \log \left( \frac{\tilde{\lambda}_t(\theta)}{r+\tilde{\lambda}_t(\theta)} \right) + \frac{r(X_t+r)}{X_t r(\theta)}$.

A negative binomial quasi-maximum likelihood estimate (NB-QMLE) of $(\theta_0, r_0)$ 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 $\theta_0$, 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 $\theta_0$ while maximizing (3.1) with respect to $\theta$, 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 $\theta_0$ in the weight by its QMLE, $\hat{\theta}_{r^*}$, obtained in i).

iii) Re-estimating $\theta_0$ 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 $\theta_0$ 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 $\theta$ 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"

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

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 $\theta$ would be robust to misspecification of the conditional distribution, whenever correctly specifying the conditional mean like (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 negative binomial GLM link function (2.3)-(2.4) (with $r$ in place of $r_0$).

On the other hand, considering **Case 1** of model (2.1) where only the conditional mean is specified, then only $\theta_0$ 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 $\theta$, 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 $\theta$.

In the rest of this Section we shall study asymptotics of two QML-type estimates that maximize (3.2) over $\theta \in \Theta$. Subsection 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) negative binomial quasi-maximum likelihood estimate (pNB-QMLE).

In Subsection 3.2, consistency and asymptotic normality of the four-stage estimate (see i)-iv) above) are established assuming the nominal GLM link function (2.3)-(2.4) for an unknown \( r_0 > 0 \).

**3.1. Profile negative binomial QMLE**

Consider **Case 1** of the INGAR model where only (2.1) is required. A profile negative binomial quasi-maximum likelihood estimate (pNB-QMLE) of \( \theta_0 \) is any measurable solution of the following problem

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

for some \( \Theta \) 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 \( (\tilde{X}_0, \tilde{X}_{-1}, \ldots) \) is of no asymptotic importance, but may influence the accuracy of estimate in finite samples. In general, one assumes that \( \tilde{X}_0 = x, \tilde{X}_{-1} = x, \ldots \) with \( x \) depending on the function \( \lambda \) or on the observations (see Ahmad and Francq, 2016). To study consistency of the pNB-QMLE, \( \hat{\theta}_r \), we need the following assumptions:

\begin{enumerate}
  \item[\textbf{A1}] \( \theta \mapsto \lambda_t (\theta) \) is a.s. continuous; \( \lambda_t (\theta) > c \) and \( \tilde{\lambda}_t (\theta) > c \), a.s. for some \( c > 0 \).
  \item[\textbf{A2}] \( a_t \xrightarrow{a.s.} 0 \) and \( \frac{a_t X_t}{a_t} \xrightarrow{a.s.} 0 \) where \( a_t = \sup_{\theta \in \Theta} \left| \tilde{\lambda}_t (\theta) - \lambda_t (\theta) \right| \).
  \item[\textbf{A3}] \( E \left( X_t^\delta \right) < \infty \) for some \( \delta > 1 \).
  \item[\textbf{A4}] \( \lambda_t (\theta) = \lambda_t (\theta_0) \) a.s. if and only if \( \theta = \theta_0 \).
  \item[\textbf{A5}] \( \Theta \) is compact.
\end{enumerate}

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

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

\[
\hat{\theta}_r \xrightarrow{a.s.} \theta_0, \quad \text{for all } r > 0.
\]
The latter result shows that, like the \( P-QMLE \), the \( pNB-QMLE \) is robust to mis-specification 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 \( \tilde{l}_{t,r}(\theta) \) in (3.2) with \( \lambda_t(\theta) \) in place of \( \tilde{\lambda}_t(\theta) \) 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 \( b_t = \sup_{\theta \in \Theta} \left| \lambda_t^2(\theta) - \lambda_t^2(\theta) \right| \), \( c_t = \sup_{\theta \in \Theta} \left\| \frac{\partial (\tilde{\lambda}_t(\theta) - \lambda_t(\theta))}{\partial \theta} \right\| \) and
\[
d_t = \sup_{\theta \in \Theta} \max \left( \left\| \frac{1}{\lambda_t(\theta)(t+\lambda_t(\theta))} \frac{\partial^2 \tilde{\lambda}_t(\theta)}{\partial \theta^2} \right\|, \left\| \frac{1}{\lambda_t(\theta)(t+\lambda_t(\theta))} \frac{\partial \lambda_t(\theta)}{\partial \theta} \right\| \right).
\]

**A7** The true \( \theta_0 \) belongs to the interior of \( \Theta \).

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

**A9** The derivatives \( \frac{\partial^2 \lambda_t(\theta)}{\partial \theta \partial \theta} \) and \( \frac{\partial^2 \tilde{\lambda}_t(\theta)}{\partial \theta \partial \theta} \) exist and are continuous, the matrices
\[
I_r = E\left( \frac{v_t(\theta_0)}{\lambda_t^2(\theta_0)(t+\lambda_t(\theta_0))^2} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \right) \quad \text{and} \quad J_r = E\left( \frac{1}{\lambda_t(\theta_0)(t+\lambda_t(\theta_0))} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \right),
\]
are finite, and \( J_r \) is nonsingular for all \( r > 0 \).

**A10** There is a neighborhood \( V(\theta_0) \) of \( \theta_0 \) such that \( E\left( \sup_{\theta \in V(\theta_0)} \left\| \frac{\partial^2 \tilde{l}_{t,r}(\theta)}{\partial \theta \partial \theta} \right\| \right) < \infty \) 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,
\[
\sqrt{n} \left( \tilde{\theta}_r - \theta_0 \right) \overset{L}{\underset{n \to \infty}{\rightarrow}} N(0, J^{-1}_r I_r J^{-1}_r) \quad \text{for all } r > 0. \tag{3.5}
\]
Some remarks are in order:
When the conditional distribution of the data generating process (2.1) is negative binomial-II with parameters $r_0$ and $\frac{r_0}{r_0+\lambda}$, i.e. $X_t/\mathcal{F}_{t-1} \sim \mathcal{NB}(r_0, \frac{r_0}{r_0+\lambda})$, then (3.5) holds with $I_r = \frac{1}{r_0} E \left( \frac{r_0+\lambda}{\lambda(\theta_0)(r+\lambda(\theta_0))} \frac{\partial \lambda(\theta_0) \partial \lambda(\theta_0)}{\partial \theta^2} \right)$. In particular, when $r$ in (3.2)-(3.3) coincides with the "true" $r_0$ in (2.3)-(2.4), then $\hat{\theta}_{r_0}$ becomes the maximum likelihood estimate (MLE), which is then asymptotically efficient with

$$I_{r_0} = \frac{1}{r_0^2} J_{r_0}.$$  

Therefore, (3.5) becomes

$$\sqrt{n} \left( \hat{\theta}_{r_0} - \theta_0 \right) \xrightarrow{D} N \left( 0, \frac{1}{r_0^2} J_{r_0}^{-1} \right). \quad (3.6)$$

- A weaker result, which does not require specifying the full conditional distribution is that under the following more general negative binomial-II GLM link function

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

which generalizes (2.3)-(2.4), $\hat{\theta}_{r_0}$ 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

$$\sqrt{n} \left( \hat{\theta}_{r_0} - \theta_0 \right) \xrightarrow{D} N \left( 0, \delta^2 J_{r_0}^{-1} \right). \quad (3.8)$$

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, $\hat{\theta}_r$, is $\hat{J}_{r}^{-1} \hat{I}_{r} \hat{J}_{r}^{-1}$ with

$$\hat{I}_{r} = \frac{1}{n} \sum_{t=1}^{n} \left( \frac{x_t - \lambda_t(\hat{\theta}_r)}{\lambda_t(\hat{\theta}_r)(r+\lambda_t(\hat{\theta}_r))} \right)^2 \frac{\partial \lambda_t(\hat{\theta}_r) \partial \lambda_t(\hat{\theta}_r)}{\partial \theta^2}. \quad (3.9)$$

$$\hat{J}_{r} = \frac{1}{n} \sum_{t=1}^{n} \lambda_t(\hat{\theta}_r) \left( \frac{1}{r+\lambda_t(\hat{\theta}_r)} \right) \frac{\partial \lambda_t(\hat{\theta}_r) \partial \lambda_t(\hat{\theta}_r)}{\partial \theta^2}. \quad (3.10)$$
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

\[ u_t = (X_t - \lambda_t)^2 - E \left( \frac{(X_t - \lambda_t)^2}{F_{t-1}} \right) = (X_t - \lambda_t)^2 - \left( 1 + \frac{1}{\beta_0} \lambda_t \right) \lambda_t, \]

then \( E(u_t/F_{t-1}) = 0 \) and

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

where \( \gamma_0 = \frac{1}{\beta_0} \). Regression (3.11) is not ready to be used to estimate \( \gamma_0 \) since its regressand, \( \frac{(X_t - \lambda_t)^2 - \lambda_t}{\lambda_t^2(\theta_0)} \), depends on the unknown \( \theta_0 \) and is then unobservable. If a consistent estimate of \( \theta_0 \), say \( \hat{\theta}_0 \), is available then we may form the following modified (observable-regressand) regression

\[ \frac{(X_t - \hat{\lambda}_t)^2 - \hat{\lambda}_t}{\hat{\lambda}_t^2} = \gamma_0 + \frac{u_t}{\hat{\lambda}_t^2}, \quad (3.12) \]

from which a consistent estimate of \( \beta_0 \) is \( \hat{\gamma}_0 \), the inverse of the weighted least squares estimate \( \hat{\gamma} \) of \( \gamma_0 \) given by

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

where \( \hat{\lambda}_t = \tilde{\lambda}_t \left( \hat{\theta}_0 \right) \). Note that the estimate \( \hat{\gamma} \) we use here is a dynamic INGAR 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 \( (\theta_0, r_0) \) in (2.1)-(2.4) consists of a quadruple \( \left( \hat{\theta}_{r^*}, \tilde{r}_1, \tilde{r}_2, \tilde{r}_2 \right) \), 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 \( r \) par \( r^* \). Let \( \tilde{\lambda}_{1t} = \tilde{\lambda}_t \left( \hat{\theta}_{r^*} \right) \), \( 1 \leq t \leq n \).

**Step 2** Set \( \tilde{\gamma}_1 = \frac{1}{n} \sum_{t=1}^{n} \frac{(X_t - \tilde{\lambda}_{1t})^2 - \tilde{\lambda}_{1t}}{\tilde{\lambda}_{1t}^2} \) and \( \tilde{r}_1 = \tilde{\gamma}_1^{-1} \).
Step 3 Let $\widehat{\theta}_r = \arg\max_{\theta \in \Theta} \tilde{L}_{n,r_1}(\theta)$ be a solution of the problem (3.3) while replacing the generic $r$ by $\widehat{r}_1$. Get $\widehat{\lambda}_{2t} = \tilde{\lambda}_t\left(\widehat{\theta}_r\right)$, $(1 \leq t \leq n)$.

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

To get asymptotic properties of the quadruple $\left(\widehat{\theta}_r, \widehat{\gamma}_1, \widehat{\theta}_{r_1}, \widehat{r}_2\right)$, note first that $\widehat{\theta}_r$ is no other than the profile $\text{NB-QMLE}$ 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 $\left(\widehat{r}_1, \widehat{\theta}_{r_1}, \widehat{r}_2\right)$, asymptotic properties of which are given by the following result.

**Theorem 3.3** Under $(2.1)$, $(2.3)$-$(2.4)$ and $\textbf{A0-A10}$,

\begin{align}
\widehat{r}_1 & \xrightarrow{a.s.} r_0, \\
\sqrt{n}(\widehat{\gamma}_1 - \gamma_0) & \xrightarrow{L} N\left(0, E\left(\frac{\left((X_t - \lambda_t(\theta_0))^2 - \lambda_t(\theta_0) + \frac{1}{r_0} \lambda_t^2(\theta_0)\right)^2}{\lambda_t^2(\theta_0)}\right)\right),
\widehat{\gamma}_2 & \xrightarrow{A.D} \widehat{\gamma}_1, \\
\sqrt{n}(\widehat{\theta}_{r_1} - \theta_0) & \xrightarrow{L} N\left(0, \frac{1}{r_0} J^{-1}_r\right),
\end{align}

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

A few broad conclusions can be drawn.

- Strong consistency of $\widehat{\theta}_{r_1}$ directly follows from strong consistency of $\widehat{\theta}_r$ (for all $r > 0$) and $\widehat{\gamma}_1$.

- The third-step estimate $\widehat{\theta}_{r_1}$ is clearly more asymptotically efficient than the first-step estimate $\widehat{\theta}_r$.

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

- Asymptotic distribution of $\widehat{\gamma}_1$ is a reciprocal normal distribution, which is bimodal and has no first moment.

- Since $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ 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 $\tilde{\lambda}_{2t}$ and hence $\widehat{\theta}_{r_1}$, 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, \( \hat{\gamma}_{r_1} \), is

\[
\frac{1}{r} \hat{J}_{r}^{-1},
\]

(3.15)

where \( \hat{J}_{r} \) 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 \( \hat{\gamma}_{2} \) in (3.14b) is

\[
\frac{1}{n} \sum_{t=1}^{n} \frac{\left((X_t - \lambda_t(\theta_{r_1}))^2 - \left(\lambda_t(\theta_{r_1}) + \frac{1}{r_0} \lambda^2_t(\theta_{r_1})\right)\right)^2}{\lambda^4_t(\theta_{r_1})}.
\]

(3.16)

- The outputs of the 2SNB-QMLE method are \( \hat{r}_2 = (\hat{\gamma}_{2})^{-1} \) and \( \hat{\theta}_{r_1} \).

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

For the conditional mean parameter \( \theta_0 \) of the INGAR model (2.1), Ahmad and Francq (2016) proposed a Poisson QMLE (P-QMLE), which is defined as a measurable solution to the following problem

\[
\hat{\theta}_P = \arg \max_{\theta \in \Theta} \left( \tilde{L}_{P,n}(\theta) \right),
\]

(4.1a)

where

\[
\tilde{L}_{P,n}(\theta) = \frac{1}{n} \sum_{t=1}^{n} \left( -\tilde{\lambda}_t(\theta) + X_t \log \left( \tilde{\lambda}_t(\theta) \right) \right).
\]

(4.1b)

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

\[
\sqrt{n} \left( \hat{\theta}_P - \theta_0 \right) \xrightarrow{d} N \left( 0, J_P^{-1} I_P J_P^{-1} \right),
\]

(4.2)

where \( I_P = E \left( \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \right) \) 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 GARCH (INGARCH \((p, q)\)) process \(\{X_t, t \in \mathbb{Z}\}\), as proposed by Heinen (2003) and Ferland et al (2006), is defined to have a Poisson conditional distribution

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

(4.3a)

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

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

(4.3b)

where \(\theta_0 = (\omega_0, \alpha_{01}, ..., \alpha_{0q}, \beta_{01}, ..., \beta_{0p})'\) is such that \(\omega_0 > 0, \alpha_{0i} \geq 0, \beta_{0j} \geq 0\). Ferland et al (2006) showed that under the following stability condition

\[
\sum_{i=1}^{q} \alpha_{0i} + \sum_{j=1}^{p} \beta_{0j} < 1,
\]

(4.4)

the process \(\{X_t, t \in \mathbb{Z}\}\) given by (4.3) is strictly stationary and ergodic (see also Douc et al, 2013; Gonçalves et al, 2015; Davis and Liu, 2016). Under \(\sum_{j=1}^{p} \beta_{0j} < 1\), the conditional mean \(\lambda_t\) 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

\[
\text{Var} (X_t | \mathcal{F}_{t-1}) = E (X_t | \mathcal{F}_{t-1}).
\]

(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 A1-A10 simplify in the case of the Poisson INGARCH model (4.3) as in Ahmad and Francq
(2016). For instance, $I_r$ defined in A9 reduces to $I_r = E \left( \frac{1}{\lambda_t(\theta_0) (r + \lambda_t(\theta_0))} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \right)$. 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 binomial-I 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,

$$X_t/\mathcal{F}_{t-1} \sim NB\left(r_t, \pi_t\right), \quad t \in \mathbb{Z},$$

with parameters

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

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

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

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

Now consider the NB-I-INGARCH $(p, q)$ model corresponding to $K = 1$, i.e.

$$X_t/\mathcal{F}_{t-1} \sim NB\left( r_0 \lambda_t; \frac{r_0 \lambda_t}{r_0 \lambda_t + \lambda_t} \right) \equiv NB \left( r_0, \frac{r_0}{r_0 + 1} \right),$$

for which (4.7) reduces to the following linear form

$$Var \left( X_t/\mathcal{F}_{t-1} \right) = \left( 1 + \frac{1}{r_0} \right) E \left( X_t/\mathcal{F}_{t-1} \right),$$

which 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 NB-I-INGARCH model (4.8a) presents
some similarities with the Poisson $INGARCH$ model (4.3). Indeed, it is straightforward to show that the $NB-I-INGARCH$ 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

$$\sqrt{n} \left( \hat{\theta}_p - \theta_0 \right) \xrightarrow{d} N \left( 0, \left( 1 + \frac{1}{r_0} \right) \left( E \left( \frac{1}{\lambda_t(\theta_0)} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \lambda_t(\theta_0)}{\partial \theta'} \right) \right)^{-1} \right).$$

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 Gourieroux 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 $pNB-QMLE$, $\hat{\theta}_r$, 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 as in the Poisson $INGARCH$ case, the $2SNB-QMLE$ given by Section 3.2 is ill-defined.

### 4.1.3. The negative binomial-II $INGARCH$ model

Consider the $NB-II-INGARCH (p, q)$ model corresponding to (4.6) with $K = 2$, i.e.

$$X_t/F_{t-1} \sim NB \left( r_0, \frac{r_0}{r_0 + \lambda_t} \right),$$

where $r_0 > 0$ and $\lambda_t$ is given by (4.3b). 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,

$$X_t/F_{t-1} \sim NB \left( r_0, \frac{1}{1+\mu_t} \right),$$

18
where \( r_0 \in \mathbb{N}^* \) is restricted to be a positive integer and \( \mu_t \) satisfying (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 negative binomial-II GLM condition (3.7) (with \( \delta^2 = 1 \)), i.e.

\[
Var (X_t/\mathcal{F}_{t-1}) = E (X_t/\mathcal{F}_{t-1}) \left( 1 + \frac{1}{r_0} E (X_t/\mathcal{F}_{t-1}) \right), \quad r_0 > 0, \tag{4.11}
\]

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),

\[
I_p = \frac{1}{r_0} E \left( \frac{(r_0 + \lambda_t(\theta_0)) \partial \lambda_t(\theta_0) \partial \lambda_t(\theta_0)}{\lambda_t(\theta_0)} \right).
\]

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 it is clear that

\[
I_r = \frac{1}{r_0} E \left( \frac{(r_0 + \lambda_t(\theta_0)) \partial \lambda_t(\theta_0) \partial \lambda_t(\theta_0)}{r + \lambda_t(\theta_0)} \right).
\]

Thus none of the pNB-QMLE and P-QMLE is asymptotically superior than the other, unless \( r_0 \) would be known. In that case, one can take \( r = r_0 \) and the resulting pNB-QMLE, \( \widehat{\theta}_{r_0} \), would be asymptotically efficient. For instance, consider the Geometric INGARCH model which is a special case of the NB-II-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 Negative binomial-\textit{II} assumption (4.11). Therefore, the 2SNB-QMLE is ill-defined and none of \textit{P-QMLE} and \textit{pNB-QMLE} is asymptotically preferred than the other.

4.1.4. The \textit{INAR(1)} model

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

\[ X_t = \alpha_0 \circ X_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}, \]  

(4.12)

where \{\varepsilon_t, t \in \mathbb{Z}\} is an independent and identically distributed (\textit{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 \textit{iid} Bernoulli random sequence such that \( P(Y_i = 1) = \alpha_0 \in (0, 1) \). It is well known that

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

and that assumption \textbf{A0} reduces in term of \( \alpha_0 \) to

\[ \alpha_0 < 1, \]

(cf. Al-Osh and Alzaid, 1987). Furthermore, the \textit{INAR(1)} model (4.12) obeys to the following \textit{affine GLM} link function

\[ \text{Var}(X_t/\mathcal{F}_{t-1}) = \alpha_0(1 - \alpha_0) X_{t-1} + \sigma_0^2 \]

\[ = (1 - \alpha_0) E(X_t/\mathcal{F}_{t-1}) + \sigma_0^2 - (1 - \alpha_0) \omega_0. \]  

(4.13)

Note that if \( \frac{\sigma_0^2}{\omega_0} = 1 - \alpha_0 < 1 \), so that the innovation term \( \varepsilon_t \) should be \textit{underdispersed}, then the affine link function (4.13) reduces to the linear Poisson \textit{GLM} condition (4.8b) with proportionality constant \( 1 - \alpha_0 \). Therefore, the \textit{P-QMLE} would be asymptotically efficient in the class of all \textit{QMLE}'s belonging to the linear exponential family and hence it would be
more asymptotically efficient than the $pNB$-QMLE. Specifically,

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

If, however, \( \frac{\sigma^2}{\omega_0} \neq (1 - \alpha_0) \), 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.

### 4.2. 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 three instances of (2.1), namely:

i) The Poisson $INGARCH(1, 1)$ model (4.3) with parameter $\theta_0 = (2, 0.3, 0.6)'$ (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 negative binomial-II $INGARCH(1, 1)$ model (4.9) with parameters $r_0 = 3$ and $\theta_0 = (2, 0.3, 0.6)'$ (cf. Table 4.3).

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 $NB$-II $INGARCH(1, 1)$ model (4.9) we also run the two-stage $NB$-QMLE, \( \left( \widehat{\theta}_r, \tilde{\gamma}_1, \tilde{\gamma}_2 \right) \), 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, $\theta^{(0)} = \left( \omega^{(0)}, \alpha^{(0)}, \beta^{(0)} \right)'$, 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

\[ \tilde{X}_0 = \bar{X} \quad \text{and} \quad \tilde{\lambda}_0(\theta) = \frac{\omega + \alpha \bar{X}}{1 - \beta} \simeq E \left( \lambda_t(\theta) \right), \quad \text{for} \quad \theta = (\omega, \alpha, \beta)' \in \Theta, \]
where $\bar{X}$ is the empirical mean of the series $X_1, \ldots, X_n$. Concerning Algorithm 3.1, which is only applied in the case of the $NB-II$-INGARCH model (4.9), we need to estimate the initial dispersion parameter $r^*$. For this we mime the negative binomial-$II$ 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.

$$r^* = \frac{(\bar{X})^2}{S^2 - \bar{X}}, \quad (4.15)$$

where $S^2$ is the sample variance of $X_1, \ldots, X_n$. 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 $X_1, \ldots, X_n$ should be overdispersed (i.e. $S^2 > \bar{X}$), otherwise $r^*$ would be negative, which is not valid.

Mean of estimates, their standard deviation ($StD$) and their empirical Root Minimum Square Error ($RMSE$) over the 500 replications are reported in Tables 4.1-4.3. The $RMSE$ of an estimate $\hat{\theta}$ of $\theta_0$ is calculated from the formula $RMSE = \sqrt{bias^2 + StD^2}$, where $bias$ is the sample mean of $\hat{\theta} - \theta_0$ over the 500 replications.

<table>
<thead>
<tr>
<th>$\theta_0$</th>
<th>$\hat{\theta}_P$</th>
<th>$\hat{\theta}_1$</th>
<th>$\hat{\theta}_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega = 2$</td>
<td>$Mean$ 1.9891</td>
<td>2.0111</td>
<td>2.0587</td>
</tr>
<tr>
<td>$\alpha_0 = 0.3$</td>
<td>$StD$ 0.2205</td>
<td>0.2977</td>
<td>0.3298</td>
</tr>
<tr>
<td>$\beta_0 = 0.6$</td>
<td>$RMSE$ 0.2208</td>
<td>0.2979</td>
<td>0.3350</td>
</tr>
<tr>
<td>$Mean$ 0.3144</td>
<td>0.3322</td>
<td>0.3248</td>
<td></td>
</tr>
<tr>
<td>$StD$ 0.0215</td>
<td>0.0290</td>
<td>0.0328</td>
<td></td>
</tr>
<tr>
<td>$RMSE$ 0.0259</td>
<td>0.0433</td>
<td>0.0411</td>
<td></td>
</tr>
<tr>
<td>$Mean$ 0.5850</td>
<td>0.5669</td>
<td>0.5713</td>
<td></td>
</tr>
<tr>
<td>$StD$ 0.0253</td>
<td>0.0357</td>
<td>0.0372</td>
<td></td>
</tr>
<tr>
<td>$RMSE$ 0.0294</td>
<td>0.0487</td>
<td>0.0470</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1. Mean, Standard Deviation and empirical $RMSE$ of $\hat{\theta}_r$ ($r = 1, 4$) and $\hat{\theta}_P$ for Poisson INGARCH($1, 1$) series with $\theta_0 = (2, 0.3, 0.6)'$ and $n = 1000$. 

22
<table>
<thead>
<tr>
<th>(\theta_0)</th>
<th>(\hat{\theta}_r)</th>
<th>(\hat{\theta}_1)</th>
<th>(\hat{\theta}_4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\omega = 2)</td>
<td>Mean</td>
<td>2.2428</td>
<td>2.0316</td>
</tr>
<tr>
<td></td>
<td>StD</td>
<td>0.4957</td>
<td>0.3227</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.5520</td>
<td>0.3242</td>
</tr>
<tr>
<td>(\alpha_0 = 0.3)</td>
<td>Mean</td>
<td>0.2965</td>
<td>0.2973</td>
</tr>
<tr>
<td></td>
<td>StD</td>
<td>0.0422</td>
<td>0.0325</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.0423</td>
<td>0.0326</td>
</tr>
<tr>
<td>(\beta_0 = 0.6)</td>
<td>Mean</td>
<td>0.5896</td>
<td>0.6006</td>
</tr>
<tr>
<td></td>
<td>StD</td>
<td>0.0528</td>
<td>0.0296</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.0538</td>
<td>0.0296</td>
</tr>
</tbody>
</table>

Table 4.2. Mean, Standard Deviation and empirical \(RMSE\) of \(\hat{\theta}_r\) \((r = 1, 4)\) and \(\hat{\theta}_P\) for geometric \(INGARCH(1, 1)\) series with \(\theta_0 = (2, 0.3, 0.6)'\) and \(n = 1000\).
Table 4.3. Mean, Standard Deviation and empirical RMSE of $\hat{\theta}_r$ ($r = 1, 4$), $\hat{\theta}_P$, $\hat{\theta}_3$ and $\hat{\theta}_\tilde{r}$ for $NB-II$-INGARCH$(1, 1)$ series with $r_0 = 3$, $\theta_0 = (2, 0.3, 0.6)'$ and $n = 1000$.

From Tables 4.1-4.3 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 conditional distribution of the INGARCH$(1, 1)$ model follows a given distribution, the QMLE calculated on that distribution is the best one compared to the others regarding to its smallest RMSE. 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 the profile $NB$-QMLE, $\hat{\theta}_4$. Finally, for the $NB-II$-INGARCH$(1, 1)$ model with dispersion parameters $r_0 = 3$ (cf. Table 4.3), the four-stage estimate $\hat{\theta}_\tilde{r}$ outperforms the Poisson QMLE, the geometric QMLE and the profile $NB$-QMLE, $\hat{\theta}_4$. 

24
5. Real applications

For illustration purposes, we propose to apply the two-stage \textit{NB-QMLE} given by Algorithm 3.1 to two famous integer-valued time series under the negative binomial-\textit{II INGARCH} (1, 1) framework. The first one is the 	extit{Polio data} (Zeger, 1988) while the second one is the \textit{Transaction data} (Fokianos et al, 2009). The choice of the \textit{NB-II-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 \textit{NB-II-INGARCH} (1, 1) model, but via different estimation methods. This allows us to compare their methods with our proposed \textit{2SNB-QMLE}. All procedures have been applied on a personal computer using \textit{R}. The optimization (3.3) is carried out using the function constrOptim() of \textit{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: Monthly number of poliomyelitis cases in the United States from 1970 to 1983.

(a) Series, (b) Histogram.

Zhu (2011) fitted a $NB-II-INGARCH (1,1)$ 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 $\theta$ for fixed $r$ and then choosing $\theta$ with largest likelihood over all selected values of $r \in \{1, \ldots, \tau\}$, for some fixed positive integer $\tau$. The estimated model of Zhu (2011) is given by

$$X_t/\mathcal{F}_{t-1} \sim NB\left(\hat{\gamma}, \frac{1}{1+\hat{\mu}_t}\right),$$

$$\hat{\gamma} = 2,$$

$$\left\{ \begin{array}{l}
\hat{\mu}_t = 0.31190 + 0.1843X_{t-1} + 0.1815\hat{\mu}_{t-1}, \quad 2 \leq t \leq 168 \\
\hat{\mu}_1 = \bar{X},
\end{array} \right.$$ 

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 $0.1843 + 0.1815 = 0.3658$. 

26
To compare with Zhu’s (2011) fit, we estimated a \textit{NB-II-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 (4.15) giving
\[
r^* = \frac{(1.3333)^2}{3.5050-1.3333} = 0.8186,
\]
while the starting values of the \textit{INGARCH} (1, 1) equation (4.3b) are taken as in (4.14). The initial conditional mean parameter \(\theta^{(0)}\) of the optimization problem (3.3) is obtained while preliminarily running the Geometric \textit{QMLE} on the polio series with initial parameter \((2, 0.3, 0.6)^\prime\). The estimated parameters of the model and their \textit{Asymptotic Standard Errors (ASE)} are summarized in Table 5.1. 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 \(\hat{\theta}_{r_2}\) is obtained from (3.14d) and (3.15). Note that the ASE of \(\hat{r}_2\) is not available since the distribution of \(\hat{r}_2\) has not a usual form.

<table>
<thead>
<tr>
<th>\textit{NB-II-INGARCH} \textit{parameters}</th>
<th>Estimates : ((\hat{\theta}_{\hat{r}_1}, \hat{\gamma}_2, \hat{r}_2))</th>
<th>ASE of ((\hat{\theta}_{\hat{r}_1}, \hat{\gamma}_2))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\omega_0)</td>
<td>0.6564</td>
<td>0.2050</td>
</tr>
<tr>
<td>(\alpha_0)</td>
<td>0.3743</td>
<td>0.1580</td>
</tr>
<tr>
<td>(\beta_0)</td>
<td>0.1511</td>
<td>0.0935</td>
</tr>
<tr>
<td>(\gamma_0 = \frac{1}{r_0})</td>
<td>0.3843</td>
<td>0.1945</td>
</tr>
<tr>
<td>(r_0)</td>
<td>2.6023</td>
<td>–</td>
</tr>
</tbody>
</table>

Table 5.1: 2S-NBQML estimates and their asymptotic standard errors for the \textit{NB-II-INGARCH}(1, 1) model from the Polio data.

The fitted model (4.9) using the 2SNB-QMLE is given by
\[
X_t/F_{t-1} \sim \text{NB}\left(\frac{\hat{r}_2}{\hat{r}_2^2 + \lambda_t}, \frac{\hat{r}_2}{\hat{r}_2^2 + \lambda_t}\right),
\]
\[
\hat{r}_2 = 2.6023,
\]
\[
\begin{cases}
\hat{\lambda}_t = 0.6564 + 0.3743X_{t-1} + 0.1511\hat{\lambda}_{t-1}, & 2 \leq t \leq 168 \\
\hat{\lambda}_1 = X = 1.3333,
\end{cases}
\]
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 $\overline{X} = 1.3333$ than the estimated mean, 0.9836, given by Zhu’s (2011) model. On the other hand, some properties of the residuals are shown in Figure 5.2. Indeed, from the sample autocorrelation and partial autocorrelation functions in Figure 5.2 (panels (a) and (b)), the residuals look like a white noise. However, a visual inspection (cf. Figure 5.2, panels (c) and (d)) reveals that the normality assumption of the residuals is untenable. In sum, regarding the stability of the estimated model, the significance of its coefficients and the residual analysis in Figure 5.2, it can be concluded that the estimated model is acceptable.

Figure 5.2: 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.

Now we compare in-sample performance of our fit (5.1) with that of Zhu (2011). Table

28
5.2 provides the residual sum of squares (RSS) induced by models (5.1) and (5.2). These RSS’s are given respectively by

\[
\text{RSS} \left( \hat{\lambda}_t \right) = \sum_{t=2}^{168} (X_t - \hat{\lambda}_t)^2,
\]

\[
\text{RSS} \left( 2\hat{\mu}_t \right) = \sum_{t=2}^{168} (X_t - 2\hat{\mu}_t)^2,
\]

starting from initial values \( \hat{\lambda}_1 = \hat{\mu}_1 = \bar{X} \). The latter initial value was considered by Zhu (2011).

<table>
<thead>
<tr>
<th>Predictors</th>
<th>( \hat{\lambda}_t )</th>
<th>( 2\hat{\mu}_t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>RSS</td>
<td>535.1793</td>
<td>540.6634</td>
</tr>
</tbody>
</table>

Table 5.2: Residual sum of squares (RSS) of the predictors \( \hat{\lambda}_t \) (5.2) and \( 2\hat{\mu}_t \) (5.1) for the Polio series.

From Table 5.2 it can be seen that our model estimated by the 2SNB-QMLE (Algorithm 3.1) slightly 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) and (5.2) starting from several initial values \( \hat{\lambda}_1 \) and \( \hat{\mu}_1 \); the unreported results were virtually the same. Finally, Figure 5.3 displays the polio data together with the estimated conditional mean \( \hat{\lambda}_t \) and the estimated conditional variance given by \( \hat{\nu}_t = \hat{\lambda}_t \left( 1 + \frac{1}{\hat{\nu}_2} \hat{\lambda}_t \right) \), where the overdispersion phenomenon seems reproduced.

![Figure 5.3: Polio series and its estimated conditional mean and conditional variance.](image)
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 (2009) and Christou and Fokianos (2014) among others. Like the Polio data, the Transaction series is overdispersed viewing 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).

![Graph](image_url)

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 \( NB-II-INGARCH(1,1) \) model (4.9) to the Transaction data. They found the following specification

\[
X_t/\mathcal{F}_{t-1} \sim NB\left(\hat{r}, \frac{\hat{r}}{\hat{r}+\hat{\mu}}\right),
\]

\[
\hat{r} = 7.0220,
\]

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

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 \( NB-II-INGARCH \) 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 (4.15), the initial dispersion parameter $r^*$ 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 (4.14). The parameter estimates and their Asymptotic Standard Errors (ASE) are summarized in Table 5.3.

<table>
<thead>
<tr>
<th>$NB-II-INGARCH$ parameters</th>
<th>Estimates : $\hat{\theta}_{r_1, \gamma_2, \beta_2}$</th>
<th>ASE of $\hat{\theta}_{r_1, \gamma_2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_0$</td>
<td>0.7996</td>
<td>0.4034</td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>0.7928</td>
<td>0.0650</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>0.1249</td>
<td>0.0340</td>
</tr>
<tr>
<td>$\gamma_0 = \frac{1}{r_0}$</td>
<td>0.1279</td>
<td>0.0241</td>
</tr>
<tr>
<td>$r_0$</td>
<td>7.8199</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 5.3: 2S-NBQML estimates and their asymptotic standard errors for the $NB-II-INGARCH (1,1)$ model from the Transaction data.

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

$$X_t/\mathcal{F}_{t-1} \sim NB \left( \hat{r}_2, \frac{\hat{\gamma}_2}{\hat{\gamma}_2 + \lambda_t} \right),$$

$$\hat{r}_2 = 7.8199,$$

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

with a strong persistence parameter of 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 5.5 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 $(5.4)$. It turns out that the hypothesis that the residuals form a non-Gaussian white noise is strongly tenable.

![Sample Autocorrelation Function](image1)

![Sample Partial Autocorrelation Function](image2)

![Kernel density](image3)

![QQ-plot](image4)

Figure 5.5: 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 the residuals versus the standard normal distribution.

Next we compare the $RSS$ of our fit $(5.4)$ with that of Christou and Fokianos (2014) given by $(5.3)$. 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 $(5.3)$ and $(5.4)$ from several initial values (cf. Table 5.4)
although Christou and Fokianos (2014) have taken $\hat{\mu}_1 = 0$.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>$\hat{\lambda}_t$</th>
<th>$\hat{\mu}_t$</th>
<th>$\hat{\lambda}_t$</th>
<th>$\hat{\mu}_t$</th>
<th>$\hat{\lambda}_t$</th>
<th>$\hat{\mu}_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial values</td>
<td>$\hat{\lambda}_1, \hat{\mu}_1$</td>
<td>0</td>
<td>0</td>
<td>9.8239</td>
<td>9.8239</td>
<td>10.2070</td>
</tr>
</tbody>
</table>

$RSS$ | 10400.6733 | 10422.8003 | 9809.6645 | 9943.0150 | 9796.8644 | 9933.0780 |

Table 5.4: Residual sum of squares ($RSS$) of the predictors $\hat{\lambda}_t$ (5.4) and $\hat{\mu}_t$ (5.3) for the Transaction data.

It can be seen from Table 5.4 that model (5.4) 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 $\hat{\nu}_t = \hat{\lambda}_t \left(1 + \frac{1}{r^2} \hat{\lambda}_t\right)$, where the overdispersion phenomenon is highlighted.

![Figure 5.6: Transaction series and its estimated conditional mean and conditional variance.](image)

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

On the other hand, we have seen how the proposed \textit{NB-QMLE}'s can be applied to some specific integer-valued models like the Poisson and negative binomial \textit{INGARCH} models and also to the \textit{INAR} equation. Other famous particular cases of the \textit{INGAR} model like the log-\textit{INGARCH} model (Fokianos and Tjøstheim, 2011), the double Poisson \textit{INGARCH} model (Heinen, 2003; Ahmad and Francq, 2016), the generalized Poisson \textit{INGARCH} model (Zhu, 2012a) and Integer-valued \textit{ARMA} (\textit{INARMA}) models also apply in the framework of our methods. Finally, generalizations of the proposed methods to multivariate versions of the \textit{INGAR} model 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.

\textbf{Lemma 7.1 Under \textit{A1-A2}}

\[
\lim_{n \to \infty} \sup_{\theta \in \Theta} \left| L_{n,r}(\theta) - \tilde{L}_{n,r}(\theta) \right| = 0, \quad a.s.
\]

\textbf{Proof} Using the inequality \( \log(x) \leq x - 1 \), the fact that \( \tilde{\lambda}_i(\theta) > 0 \), the assumptions
A1-A2 and the Césaro lemma it follows that

\[
\sup_{\theta \in \Theta} \left| L_{n,r} (\theta) - \tilde{L}_{n,r} (\theta) \right| = \frac{1}{n} \sup_{\theta \in \Theta} \left| \sum_{t=1}^{n} \left( \log \left( \frac{r+\lambda_t(\theta)}{r+\lambda_t(\theta_0)} \right) + X_t \log \left( \frac{\lambda_t(\theta)(r+\lambda_t(\theta))}{\lambda_t(\theta_0)(r+\lambda_t(\theta_0))} \right) \right) \right|
\]

\[
= \frac{1}{n} \sup_{\theta \in \Theta} \left| \sum_{t=1}^{n} \left( \log \left( \frac{\lambda_t(\theta)}{r+\lambda_t(\theta)} + 1 \right) + X_t \log \left( \frac{\lambda_t(\theta) - \lambda_t(\theta) + 1}{\lambda_t(\theta)(r+\lambda_t(\theta)) + 1} \right) \right) \right|
\]

\[
\leq \frac{1}{n} \sum_{t=1}^{n} \left( \frac{1}{r} \sup_{\theta \in \Theta} \left| \lambda_t(\theta) - \tilde{\lambda}_t(\theta) \right| + X_t \sup_{\theta \in \Theta} \left| \tilde{\lambda}_t(\theta) - \lambda_t(\theta) \right| \right) \frac{r}{r+\lambda_t(\theta)}
\]

\[
= \frac{1}{n} \sum_{t=1}^{n} \left( \frac{1}{r} a_t + \frac{1}{c} X_t a_t \right) \xrightarrow{a.s.} 0.
\]

\[\blacksquare\]

**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

\[
|E (l_{1,r} (\theta_0))| \leq E (|l_{1,r} (\theta_0)|) \leq E \left( \left| \log \left( \frac{r}{r+\lambda_t(\theta_0)} \right) \right| \right) + E \left( \left| X_t \log \left( \frac{\lambda_t(\theta_0)}{r+\lambda_t(\theta_0)} \right) \right| \right)
\]

\[
\leq E \left( \left| \log \left( \frac{r}{r+\lambda_t(\theta_0)} \right) \right| \right) + E \left( X_t^\delta \right)^{1/\delta} \left( E \left( \log \left( \frac{\lambda_t(\theta_0)}{r+\lambda_t(\theta_0)} \right) \right) \right)^{\delta - 1/\delta} < \infty.
\]

(7.1)

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

\[
E (l_{1,r} (\theta) - l_{1,r} (\theta_0)) = E \left( r \log \left( \frac{r+\lambda_t(\theta_0)}{r+\lambda_t(\theta)} \right) + X_t \log \left( \frac{\lambda_t(\theta)(r+\lambda_t(\theta))}{\lambda_t(\theta_0)(r+\lambda_t(\theta_0))} \right) \right)
\]

\[
\leq r E \left( \frac{r+\lambda_t(\theta_0)}{r+\lambda_t(\theta)} - 1 \right) + X_t \left( \frac{\lambda_t(\theta)(r+\lambda_t(\theta))}{\lambda_t(\theta_0)(r+\lambda_t(\theta_0))} - 1 \right)
\]

\[
= r E \left( \frac{\lambda_t(\theta)-\lambda_t(\theta_0)}{r+\lambda_t(\theta)} \right) + X_t \left( \frac{\lambda_t(\theta)-\lambda_t(\theta_0)}{r+\lambda_t(\theta)} \right)
\]

\[
= r E \left( \frac{\lambda_t(\theta_0)-\lambda_t(\theta)}{r+\lambda_t(\theta)} \right) = 0,
\]

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

\[
 r E \left( \log \frac{r+\lambda_1(\theta_0)}{r+\lambda_1(\theta)} + X_t \log \frac{\lambda_1(\theta)(r+\lambda_1(\theta))}{\lambda_1(\theta_0)(r+\lambda_1(\theta_0))} \right) = 0,
\]

which holds if and only if \( \lambda_t(\theta) = \lambda_t(\theta_0) \) and then, by the identifiability assumption \( A_4 \), if and only if \( \theta = \theta_0 \). ■

**Lemma 7.3** Under \( A_0-A_5 \), there exists for all \( \theta \neq \theta_0 \) a neighborhood \( V(\theta) \) such that

\[
 \limsup_{n \to \infty} \sup_{\theta^* \in V(\theta)} \bar{L}_{n,r}(\theta^*) < \limsup_{n \to \infty} \bar{L}_{n,r}(\theta_0) \quad a.s. \tag{7.3}
\]

**Proof** For all \( \bar{\theta} \in \Theta \) and \( k \in \mathbb{N}^* \) let \( V_k(\bar{\theta}) \) be the open ball of 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 \( A_0 \), 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

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

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

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

\[
 \limsup_{n \to \infty} \sup_{\theta \in V_k(\bar{\theta}) \cap \Theta} \bar{L}_{n,r}(\theta) \leq \limsup_{n \to \infty} \bar{L}_{n,r}(\theta_0) = \limsup_{n \to \infty} L_{n,r}(\theta_0) = E \left( l_{1,r}(\theta_0) \right).
\]

Thus from standard arguments the proof of Theorem 3.1 is completed while using assumption \( A_5 \) of compactness of \( \Theta \).
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 \( \Theta \) for \( n \) sufficiently large. Hence, a Taylor expansion of \( \frac{\partial L_n, r(\hat{\theta}_r)}{\partial \theta} \) at \( \theta_0 \) yields

\[
0 = \sqrt{n} \frac{\partial L_n, r(\hat{\theta}_r)}{\partial \theta} = \sqrt{n} \frac{\partial L_n, r(\hat{\theta}_r)}{\partial \theta} + \sqrt{n} \left( \frac{\partial L_n, r(\theta)}{\partial \theta} - \frac{\partial L_n, r(\hat{\theta}_r)}{\partial \theta} \right) = \sqrt{n} \frac{\partial^2 L_n, r(\theta^*)}{\partial \theta \partial \theta}(\hat{\theta}_r - \theta_0) + \sqrt{n} \left( \frac{\partial L_n, r(\theta)}{\partial \theta} - \frac{\partial L_n, r(\hat{\theta}_r)}{\partial \theta} \right),
\]

(7.4)

for a certain \( \theta^* \) between \( \hat{\theta}_r \) and \( \theta_0 \). 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 \( n \to \infty \). 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 convergence of the matrix in the second term of (7.4).

**Lemma 7.4 Under A0-A10**

\[
\sqrt{n} \sup_{\theta \in \Theta} \left| \frac{\partial L_n, r(\theta)}{\partial \theta} - \frac{\partial L_n, r(\theta_0)}{\partial \theta} \right| \xrightarrow{a.s.} 0, \quad n \to \infty.
\]

**Proof** Using \( A2 \) and \( A6 \) it follows that

\[
\sqrt{n} \sup_{\theta \in \Theta} \left| \frac{\partial L_n, r(\theta)}{\partial \theta} - \frac{\partial L_n, r(\theta_0)}{\partial \theta} \right| = \frac{1}{\sqrt{n}} \sup_{\theta \in \Theta} \left| \sum_{t=1}^{n} \left[ \frac{\partial}{\partial \theta} \left( \log \left( \frac{r + \lambda_t(\theta)}{r + \lambda_t(\theta_0)} \right) + X_t \log \left( \frac{\lambda_t(\theta)}{r + \lambda_t(\theta)} \right) \right) \right] \right| \leq \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( c_t + a_t d_t + X_t \left( \frac{c_t}{\lambda_t} + \frac{(a_t + b_t) d_t}{e^t r^2} \right) \right) \xrightarrow{a.s.} 0, \quad n \to \infty.
\]

**Lemma 7.5 Under A8-A9,**

\[
\sqrt{n} \frac{\partial L_n, r(\theta_0)}{\partial \theta} \xrightarrow{L^2} \mathcal{N}(0, I_r).
\]

**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

\[
\sqrt{n} \frac{\partial L_n, r(\theta_0)}{\partial \theta} = \sum_{t=1}^{n} \frac{1}{\sqrt{n}} \frac{\partial L_n, r(\theta_0)}{\partial \theta},
\]

\[
\frac{\partial L_n, r(\theta_0)}{\partial \theta} = \frac{\partial \lambda_t(\theta_0)}{\partial \theta} X_t - \lambda_t(\theta_0) \lambda_t(\theta_0)(1 + \lambda_t(\theta_0)).
\]
By A8-A9 we have

\[
E \left( \frac{\partial^2 L_{n,r}(\theta_0)}{\partial \theta_i \partial \theta_j} \right) = E \left( \frac{v_r(\theta_0)}{\lambda^2_r(\theta_0)(1+\lambda_r(\theta_0))^2} \frac{\partial \lambda_r(\theta_0) \partial \lambda_r(\theta_0)}{\partial \theta_i \partial \theta_j} \right) = I_r.
\]

Thus Lemma 7.4 follows from the martingale central limit theorem (e.g. Billingsley, 2008).

\[ \blacksquare \]

**Lemma 7.6** Under A8-A10,

\[
\frac{\partial^2 L_{n,r}(\theta^*)}{\partial \theta_i \partial \theta_j} \xrightarrow{a.s.} n \to \infty J_r.
\]

**Proof** Let \( V_k(\theta_0) \ (k \in \mathbb{N}^*) \) be the open ball with center \( \theta_0 \) 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 \( \theta^* \) belongs to \( V_k(\theta_0) \). By stationarity and ergodicity of

\[
\left\{ \sup_{\theta \in V_k(\theta_0)} \left| \frac{\partial^2 L_{n,r}(\theta)}{\partial \theta_i \partial \theta_j} - E \left( \frac{\partial^2 L_{n,r}(\theta_0)}{\partial \theta_i \partial \theta_j} \right) \right| \right\},
\]

we have

\[
\left| \frac{\partial^2 L_{n,r}(\theta^*)}{\partial \theta_i \partial \theta_j} - J_r(i, j) \right| = \left| \frac{\partial^2 L_{n,r}(\theta^*)}{\partial \theta_i \partial \theta_j} - E \left( \frac{\partial^2 L_{n,r}(\theta_0)}{\partial \theta_i \partial \theta_j} \right) \right|
\]

\[
= \frac{1}{n} \left| \sum_{t=1}^{n} \frac{\partial^2 L_{t,r}(\theta^*)}{\partial \theta_i \partial \theta_j} - E \left( \frac{\partial^2 L_{t,r}(\theta_0)}{\partial \theta_i \partial \theta_j} \right) \right|
\]

\[
\leq \frac{1}{n} \sup_{\theta \in V_k(\theta_0)} \left| \sum_{t=1}^{n} \frac{\partial^2 L_{t,r}(\theta)}{\partial \theta_i \partial \theta_j} - E \left( \frac{\partial^2 L_{t,r}(\theta_0)}{\partial \theta_i \partial \theta_j} \right) \right|
\]

\[
\leq \frac{1}{n} \sup_{\theta \in V_k(\theta_0)} \left| \sum_{t=1}^{n} \frac{\partial^2 L_{t,r}(\theta)}{\partial \theta_i \partial \theta_j} - E \left( \frac{\partial^2 L_{t,r}(\theta_0)}{\partial \theta_i \partial \theta_j} \right) \right|
\]

\[
\xrightarrow{a.s.} n \to \infty E \left( \sup_{\theta \in V_k(\theta_0)} \left| \frac{\partial^2 L_{t,r}(\theta)}{\partial \theta_i \partial \theta_j} - E \left( \frac{\partial^2 L_{t,r}(\theta_0)}{\partial \theta_i \partial \theta_j} \right) \right| \right).
\]

In view of A10, the Lebesgue dominated convergence theorem yields

\[
\lim_{k \to \infty} E \left( \sup_{\theta \in V_k(\theta_0)} \left| \frac{\partial^2 L_{t,r}(\theta)}{\partial \theta_i \partial \theta_j} - E \left( \frac{\partial^2 L_{t,r}(\theta_0)}{\partial \theta_i \partial \theta_j} \right) \right| \right) = E \left( \lim_{k \to \infty} \sup_{\theta \in V_k(\theta_0)} \left| \frac{\partial^2 L_{t,r}(\theta)}{\partial \theta_i \partial \theta_j} - E \left( \frac{\partial^2 L_{t,r}(\theta_0)}{\partial \theta_i \partial \theta_j} \right) \right| \right)
\]

\[
= 0,
\]

which completes the proof of the lemma. \( \blacksquare \)
7.3. Proof of Theorem 3.3

i) Proof of (3.14a) It suffices to prove strong consistency of \( \hat{\gamma} \). From (3.12) and (3.13) we have
\[
\hat{\gamma} - \gamma_0 = \frac{1}{n} \sum_{t=1}^{n} \frac{u_t}{\lambda_t^2} = \frac{1}{n} \sum_{t=1}^{n} \frac{u_t}{\lambda_t^2} + \frac{1}{n} \sum_{t=1}^{n} \frac{u_t}{\lambda_t^2} \left( \frac{1}{\lambda_t} - \frac{1}{\lambda_t^2} \right) .
\] (7.5)

By the ergodic theorem the first term in the right hand side of (7.5) satisfies the following limiting result
\[
\frac{1}{n} \sum_{t=1}^{n} \frac{u_t}{\lambda_t^2} \xrightarrow{a.s.} \frac{1}{\lambda^2} E \left( \frac{u_t}{\lambda_t^2} \right) = E \left( \frac{1}{\lambda_t^2} E (u_t|\mathcal{F}_{t-1}) \right) = 0.
\]
So it remains to show that
\[
\frac{1}{n} \sum_{t=1}^{n} \frac{u_t}{\lambda_t^2} \left( \frac{1}{\lambda_t} - \frac{1}{\lambda_t^2} \right) = o_{a.s.} (1) .
\] (7.6)

Using a Taylor expansion of \( \frac{1}{\lambda_t^2(\theta_r)} \) around \( \theta_0 \), we have
\[
\frac{1}{\lambda_t} - \frac{1}{\lambda_t^2} = \frac{1}{\lambda_t^2(\theta_r)} - \frac{1}{\lambda_t^2(\theta_0)} = -2 \frac{\lambda_t(\theta_0)}{\lambda_t^2(\theta_0)} \left( \theta_r - \theta_0 \right) ,
\]
where \( \theta^* \) is between \( \hat{\theta}_r \) and \( \theta_0 \). Thus (7.6) follows from A1, A10, the strong consistency of \( \hat{\theta}_r \) and the ergodic theorem.

ii) Proof of (3.14b) Rewrite (7.5) as follows
\[
\sqrt{n} (\hat{\gamma} - \gamma_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{u_t}{\lambda_t^2} + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{u_t}{\lambda_t^2} \left( \frac{1}{\lambda_t} - \frac{1}{\lambda_t^2} \right) .
\]

If we show that
\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{u_t}{\lambda_t^2} \left( \frac{1}{\lambda_t} - \frac{1}{\lambda_t^2} \right) = o_p(1) ,
\] (7.7)
then (3.14b) would follow from the martingale central limit theorem applied for the \( \{\mathcal{F}_t, t \in \mathbb{Z}\} \)-martingale difference \( \left\{ \frac{u_t}{\lambda_t^2}, t \in \mathbb{Z} \right\} \). Now by a Taylor expansion of \( \frac{1}{\lambda_t^2(\theta_r)} \) around \( \theta_0 \), the left-hand side of (7.7) becomes
\[
-2(\theta_r - \theta_0)' \sqrt{n} \sum_{t=1}^{n} \frac{u_t}{\lambda_t^2} \frac{\partial \lambda_t(\theta^*)}{\partial \theta^*} .
\]
and (7.7) follows from the assumptions A1 and A10, the asymptotic normality of \(\sqrt{n}(\hat{\theta}_r - \theta_0)\), which implies that

\[
\hat{\theta}_r - \theta_0 = n^{-1/2}O_p(1),
\]

and the ergodic theorem.

**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 would like to thank Prof. M.L. Diop for providing us the "Transaction data".

**References**


