Two-stage weighted least squares estimator of the conditional mean of observation-driven time series models

Aknouche, Abdelhakim and Francq, Christian

USTHB and Qassim University, University of Lille and CREST

1 December 2019

Two-stage weighted least squares estimator of the conditional mean of observation-driven time series models

Abdelhakim Aknouche* and Christian Francq † ‡

Abstract

General parametric forms are assumed for the conditional mean $\lambda_t(\theta_0)$ and variance $\upsilon_t(\xi_0)$ of a time series. These conditional moments can for instance be derived from count time series, Autoregressive Conditional Duration (ACD) or Generalized Autoregressive Score (GAS) models. In this paper, our aim is to estimate the conditional mean parameter θ_0, trying to be as agnostic as possible about the conditional distribution of the observations. Quasi-Maximum Likelihood Estimators (QMLEs) based on the linear exponential family fulfill this goal, but they may be inefficient and have complicated asymptotic distributions when θ_0 contains zero coefficients. We thus study alternative weighted least square estimators (WLSEs), which enjoy the same consistency property as the QMLEs when the conditional distribution is misspecified, but have simpler asymptotic distributions when components of θ_0 are null and gain in efficiency when υ_t is well specified. We compare the asymptotic properties of the QMLEs and WLSEs, and determine a data driven strategy for finding an asymptotically optimal WLSE. Simulation experiments and illustrations on realized volatility forecasting are presented.

*University of Science and Technology Houari Boumediene, and Qassim University.

†Corresponding author, CREST and University of Lille, e-mail: francq@univ-lille.fr.

‡Acknowledgments: The second author is grateful to the Agence Nationale de la Recherche (ANR), which supported this work via the Project MultiRisk (ANR-16-CE26-0015-02) and also thanks the labex ECODEC.

Keywords: Autoregressive Conditional Duration model; Exponential, Poisson, Negative Binomial QMLE; INteger-valued AR; INteger-valued GARCH; Weighted LSE.
JEL Classification: C13, C32, C53, C58.

1 Estimating the conditional mean

Consider a real-valued stochastic process $\{X_t, t \in \mathbb{Z}\}$. Let $\mathcal{F}_t$ be the sigma-field generated by $\{X_u, u \leq t\}$. Assume a parametric form for the conditional mean :

$$E\left(X_t \mid \mathcal{F}_{t-1}\right) = \lambda\left(X_{t-1}, X_{t-2}, ...; \theta_0\right) = \lambda_t\left(\theta_0\right) = \lambda_t, \qquad t \in \mathbb{Z}. \tag{1.1}$$

Important classes of count time series models, in particular the Poisson INteger GARCH (INGARCH), the Negative Binomial INGARCH and the INteger AR (INAR), that will be considered in Section 3 below, have a conditional mean of the form (1.1). The most frequent, and maybe most natural, specification for λ_t is the INGARCH(p, q)-type equation

$$\lambda_t = \omega_0 + \sum_{i=1}^{q} \alpha_{0i} X_{t-i} + \sum_{j=1}^{p} \beta_{0j} \lambda_{t-j}. \tag{1.2}$$

For the INAR models, the conditional mean has also the parametric form (1.2), with $p = 0$. In (1.2) the unknown parameter is $\theta_0 = (\omega_0, \alpha_{01}, \ldots, \beta_{0p})$. For modeling positive time series, such as durations or volumes, Engle and Russell (1998) proposed the ACD model of the form

$$X_t = \lambda_t z_t, \tag{1.3}$$

where (λ_t) satisfies (1.2) and (z_t) is an iid sequence of positive variables of mean 1, for instance of exponential distribution of rate parameter 1. Standard ARMA models are also of the form $X_t = \lambda_t + \epsilon_t$ with (ϵ_t) an iid noise and λ_t satisfying (1.2).

Time series models with linear conditional mean (1.2) are thus very frequent. A drawback of this linear specification is that it is very sensitive to large "outliers" in X_{t-i}. Following Creal, Koopman and Lucas (2011, 2013), Harvey (2013) and Blasques, Koopman, Lucas (2015), Generalized Autoregressive Score (GAS) alternative updating equations can be considered. For example, by assuming that z_t in (1.3) follows the square of a Student distribution of degree of freedom $\nu_0 > 2$, standardized in such a way that $Ez_t^2 = 1$, the GAS approach developed in Harvey and Chakravarty (2008) leads to the Beta-t-ACD model[1] in which

$$\lambda_t = \omega_0 + \beta_0 \lambda_{t-1} + \alpha_0 \frac{\nu_0 + 1}{\nu_0 - 2 + \frac{X_{t-1}}{\lambda_{t-1}}} X_{t-1}. \tag{1.4}$$

When ν_0 is large, this equation is close to an INGARCH(1,1), but when ν_0 is small or moderate, λ_t is less sensitive to an extreme value of X_{t-1} in Model (1.4) than in Model (1.2), which can be a highly desirable robustness property. As far as possible, we thus prefer to consider the general model (1.1) than the linear specification (1.2).

Estimating θ_0 is obviously of primary importance, in particular for predicting X_{t+h} given $\mathcal{F}_t$ for $h \geq 1$. The maximum-likelihood estimator (MLE) is often readily computable – except for parameter-driven models like the INAR model (see Cox, 1981) – but it requires to specify a conditional distribution. Each parametric specification of the conditional distribution function (cdf) leads to a parameterization of the conditional variance (when existing)

$$\mathrm{Var}\left(X_t \mid \mathcal{F}_{t-1}\right) = \upsilon\left(X_{t-1}, X_{t-2}, \ldots; \xi_0\right) = \upsilon_t\left(\xi_0\right) = \upsilon_t. \tag{1.5}$$

In practice, the choice of the cdf is an issue. There exists actually no natural choice for the cdf, or even for the conditional variance (1.5). For example, for count time series, the choice of the Poisson distribution with intensity λ_t entails $\upsilon_t = \lambda_t$, and is thus questionable since it has been empirically observed that numerous count time series exhibit conditional overdispersion (see *e.g* Christou and Fokianos, 2014). For positive observations, the ACD model (1.3) entails a conditional variance proportional to the square of the conditional mean, $\upsilon_t = \lambda_t^2(Ez_t^2 - 1)$. An additive ARMA-type model of the form $X_t = \lambda_t + \epsilon_t$ entails a constant conditional variance $\upsilon_t = E\epsilon_t^2$. In practice, one can easily conceive that the conditional variance may have other forms. Obviously, the choice of a wrong cdf may affect the efficiency, or even the consistency, of the misspecified MLE.

In the present work, we focus on the estimation of the parameter θ_0 of the conditional mean (1.1), without assuming a specific form for the cdf F_θ of the observations. In particular,

[1]The original version of this model was proposed for GARCH, but the ACD version is direct because an ACD is nothing else than the square of a GARCH.

we are interested in estimators that could be consistent even if the conditional variance (1.5) is misspecified. Since the works of Wedderburn (1974) and Gouriéroux, Monfort and Trognon (1984), it is known that, under general regularity conditions, a MLE is a QMLE – that is a MLE based on a cdf F_θ which remains consistent when the true cdf is not F_θ – if and only if F_θ is a particular member of the linear exponential family (defined by (2.19) below). For positive observations $X_1, \dots, X_n$, an example of such misspecification-consistent estimator is the Exponential QMLE (EQMLE), defined by

$$\widehat{\theta}_E = \arg\min_{\theta\in\Theta} \sum_{t=1}^{n} \left\{ X_t/\widetilde{\lambda}_t(\theta) + \log \widetilde{\lambda}_t(\theta) \right\}, \tag{1.6}$$

where Θ denotes the parameter space and $\widetilde{\lambda}_t(\theta) = \lambda(X_{t-1}, \dots, X_1, \widetilde{X}_0, \widetilde{X}_{-1}, \dots; \theta)$ for given initial values $\widetilde{X}_0, \widetilde{X}_{-1}, \dots$ This estimator coincides with the MLE when the cdf of the observations is the exponential distribution of parameter rate 1, but the EQMLE is consistent and asymptotically normal (CAN) for a much broader class of cdf's (see Aknouche and Francq, 2019). Another example of QMLE is the Poisson Quasi-MLE (PQMLE), defined by

$$\widehat{\theta}_P = \arg\max_{\theta\in\Theta} \sum_{t=1}^{n} \left\{ X_t \log\left(\widetilde{\lambda}_t(\theta)\right) - \widetilde{\lambda}_t(\theta) \right\}. \tag{1.7}$$

This estimator, which coincides with the MLE when the cdf of the observations is Poisson $\mathcal{P}_{\lambda_t}$, is CAN for the mean parameter of count time series (see Ahmad and Francq, 2016) or duration-type (see Aknouche and Francq, 2019) models. However, this estimator is in general inefficient when $\upsilon_t \neq \lambda_t$. Motivated by the existence of overdispersed series for which $\upsilon_t > \lambda_t$, Aknouche, Bendjeddou and Touche (2018) studied the profile Negative Binomial QMLE (NBQMLE), defined by

$$\widehat{\theta}_{NB} = \arg\max_{\theta\in\Theta} \sum_{t=1}^{n} X_t \log\left(\frac{\widetilde{\lambda}_t(\theta)}{r + \widetilde{\lambda}_t(\theta)}\right) - r\log\left\{r + \widetilde{\lambda}_t(\theta)\right\}, \tag{1.8}$$

where the parameter r is fixed. An intuition for the CAN of the QMLEs is obtained by looking at the first order conditions. Any QMLE $\widehat{\theta}$ satisfies

$$s_n(\widehat{\theta}) = 0, \qquad s_n(\theta) = \sum_{t=1}^{n} \frac{X_t - \widetilde{\lambda}_t(\theta)}{\widetilde{\upsilon}_t(\theta)} \frac{\partial \widetilde{\lambda}_t(\theta)}{\partial\theta}, \tag{1.9}$$

where $\widetilde{\upsilon}_t(\theta)$ is an approximation of the conditional variance υ_t of a given member of the exponential family. For the Exponential, Poisson and Negative Binomial QMLE, we have respectively $\widetilde{\upsilon}_t(\theta) = \widetilde{\lambda}_t^2(\theta)$, $\widetilde{\upsilon}_t(\theta) = \widetilde{\lambda}_t(\theta)$ and $\widetilde{\upsilon}_t(\theta) = \widetilde{\lambda}_t(\theta)(1 + \widetilde{\lambda}_t(\theta)/r)$. Each of these estimators is optimal within the class of the QMLEs when the conditional variance υ_t is well specified. The possible value of υ_t is however restricted by the fact that it must match the conditional variance of an exponential family distribution. For example, it is not possible to have $\upsilon_t = \lambda_t$ or $\upsilon_t = \lambda_t^2$ when the support of the observations is $\mathbb{R}$ (see Table 1 in Morris, 1982).

The aim of this paper is to propose and study alternative estimators which enjoy the same consistency property as the QMLEs when the cdf is misspecified, but gain in efficiency when υ_t is well specified.

Given a theoretical weight function $w_t = w(X_{t-1}, X_{t-2}, \dots)$, where w is a measurable function from $\mathbb{R}^\infty$ to $(0, \infty)$, and its observation-proxy

$$\widetilde{w}_t = w(X_{t-1}, \dots, X_1, \widetilde{X}_0, \widetilde{X}_{-1}, \dots) \geq \underline{w} > 0, \tag{1.10}$$

a first weighted least square estimator (WLSE) is defined by

$$\widehat{\theta}_{1WLS} \quad = \quad \arg\min_{\theta \in \Theta} \widetilde{L}_n\left(\theta, \widetilde{w}\right), \tag{1.11}$$

where

$$\widetilde{L}_n\left(\theta, \widetilde{w}\right) \quad = \quad \frac{1}{n}\sum_{t=1}^{n} \widetilde{l}_t\left(\theta, \widetilde{w}_t\right) \qquad \text{with} \qquad \widetilde{l}_t\left(\theta, w_t\right) = \frac{\left(X_t - \widetilde{\lambda}_t(\theta)\right)^2}{w_t}. \tag{1.12}$$

The role of the weighting sequence $\widetilde{w} = \left(\widetilde{w}_t\right)_{t \geq 1}$ is twofold: it allows the WLSE to be CAN without too strong moment conditions, and it may reduce the asymptotic variance of the estimator.

As will be seen in Section 2, the optimal choice of $\widetilde{w}$ is (proportional to) $\upsilon = \left(\upsilon_t\right)_{t \geq 1}$. In practice, the actual value of υ_t is generally unknown. Assuming for the conditional variance a parametric specification of the form

$$\upsilon^*\left(X_{t-1}, X_{t-2}, ...; \xi_0^*\right) = \upsilon_t^*\left(\xi_0^*\right), \tag{1.13}$$

the optimal sequence of weights may be estimated by

$$\{\widehat{w}_{t,n}\}_t, \quad \widehat{w}_{t,n} = \upsilon^* \left(X_{t-1}, X_{t-2}, ..., X_1, \widetilde{X}_0, \widetilde{X}_{-1}, \dots; \widehat{\xi}_n \right), \tag{1.14}$$

where $\widehat{\xi}_n$ is a first-step estimator of ξ_0^* (which is often function of the estimator $\widehat{\theta}_{1WLS}$ of θ_0, and possibly of estimates of some extra parameter ς_0). This leads to a two-stage WLSE, defined by

$$\widehat{\theta}_{2WLS} = \arg\min_{\theta\in\Theta} \widetilde{L}_n \left(\theta, \{\widehat{w}_{t,n}\}_t\right). \tag{1.15}$$

We will see that, even when the conditional variance is misspecified (*i.e.* $\upsilon_t^*(\xi_0^*) \neq \upsilon_t$), the two-stage estimator $\widehat{\theta}_{2WLS}$ is a consistent estimator of θ_0 under mild regularity conditions.

For an informal comparison with the QMLEs, note that the first order conditions entail

$$s_n(\widehat{\theta}_{2WLS}) = 0, \qquad s_n(\theta) = \sum_{t=1}^{n} \frac{X_t - \widetilde{\lambda}_t(\theta)}{\widehat{\upsilon}_t} \frac{\partial \widetilde{\lambda}_t(\theta)}{\partial\theta}, \tag{1.16}$$

where $\widehat{\upsilon}_t = \widehat{w}_{t,n}$ is a first-step estimator of υ_t. The main difference with (1.9) is that there is particular constraint on the conditional variance. We will see that this can lead to efficiency gains of the WLSE compared to QMLEs.

The rest of the paper is organized as follow. Section 2 provides general regularity conditions for CAN of the WLS estimators and compares these estimators with the MLE and QMLEs. In Section 3, more explicit CAN conditions are given for particular time series models. Section 4 proposes a method to select one estimator within a set of possible WLSEs. Monte Carlo experiments and illustrations on real data sets are presented in Section 5. Proofs are collected in Section 6.

2 Asymptotic behavior of the WLS estimators

Using a WLSE of the form (1.11), we assume that $\lambda : \mathbb{R}^\infty \times \Theta \to (-\infty, \infty)$ is a known measurable function satisfying (1.1), with θ_0 an unknown parameter belonging to some compact parameter space $\Theta \subset \mathbb{R}^m$. The WLSEs are semi-parametric estimators in the sense that, except for the mean, they are totally agnostic about the cdf of the observations.

2.1 CAN of the estimators

The CAN of the WLSE can be shown under the following assumptions.

A1 There exists a strictly stationary and ergodic process $\{X_t, t \in \mathbb{N}\}$ satisfying (1.1).

A2 Letting $a_t = \sup_{\theta\in\Theta} \left|\widetilde{\lambda}_t(\theta) - \lambda_t(\theta)\right|$, a.s. $\lim_{t\to\infty} \{\sup_{\theta\in\Theta} |\lambda_t(\theta)| + |X_t| + 1\} a_t = 0$.

A3 $\lambda_t(\theta) = \lambda_t(\theta_0)$ a.s. if and only if $\theta = \theta_0$.

A4 Almost surely, as $t \to \infty$

$$|w_t - \widetilde{w}_t| \left\{1 + X_t^2 + \sup_{\theta\in\Theta} \lambda_t^2(\theta)\right\} \to 0.$$

A5 $E\left(\frac{\upsilon_1}{w_1}\right) < \infty$ with $\upsilon_t = \mathrm{Var}(X_t \mid \mathcal{F}_{t-1})$.

A6 The matrices $I(\theta_0, w) = E\left(\frac{\upsilon_t}{w_t^2}\frac{\partial\lambda_t(\theta_0)\partial\lambda_t(\theta_0)}{\partial\theta\partial\theta'}\right)$ and $J(\theta_0, w) = E\left(\frac{1}{w_t}\frac{\partial\lambda_t(\theta_0)\partial\lambda_t(\theta_0)}{\partial\theta\partial\theta'}\right)$ exist and $J(\theta_0, w)$ is invertible.

A7 Almost surely, the function $\lambda_t(\cdot)$ admits continuous second-order derivatives in a neighbourhood $V(\theta_0)$ of θ_0, and we have $Ew_t^{-1} \sup_{\theta\in V(\theta_0)} \{X_t - \lambda_t(\theta)\}^2 < \infty$,

$$Ew_t^{-1} \sup_{\theta\in V(\theta_0)} \left\|\frac{\partial^2\lambda_t(\theta)}{\partial\theta\partial\theta'}\right\|^2 < \infty \quad \text{and} \quad Ew_t^{-1} \sup_{\theta\in V(\theta_0)} \left\|\frac{\partial\lambda_t(\theta)}{\partial\theta}\frac{\partial\lambda_t(\theta)}{\partial\theta'}\right\| < \infty. \tag{2.1}$$

A8 Letting $b_t = \sup_{\theta\in\Theta} \left\|\partial\widetilde{\lambda}_t(\theta)/\partial\theta - \partial\lambda_t(\theta)/\partial\theta\right\|$, the sequences

$$b_t\left\{|X_t| + \sup_{\theta\in\Theta}|\lambda_t(\theta)|\right\}, \quad a_t \sup_{\theta\in\Theta}\left\|\frac{\partial\lambda_t(\theta)}{\partial\theta}\right\|, \quad |w_t - \widetilde{w}_t| \sup_{\theta\in\Theta}\left\|\frac{\partial\lambda_t(\theta)}{\partial\theta}\right\|\left\{|X_t| + \sup_{\theta\in\Theta}|\lambda_t(\theta)|\right\}$$

are a.s. of order $O(t^{-\kappa})$ for some $\kappa > 1/2$.

A9 The true parameter θ_0 belongs to the interior $\overset{\circ}{\Theta}$ of Θ.

Assumptions **A1**–**A3** are used by Ahmad and Francq (2016) for showing the consistency of the PQMLE in the case of count time series. Assumptions **A2** and **A4** are used to show that the initial values $\widetilde{X}_0, \widetilde{X}_{-1}, \ldots$ are asymptotically unimportant. The choice of the weight function w_t is guided by **A5**. If υ_t is assumed to be (bounded by) a linear function of $|X_{t-1}|, \ldots, |X_{t-r}|$, then **A5** is automatically satisfied if, for instance, $w_t = 1 + \sum_{i=1}^r |X_{t-i}|$. If w_t is chosen to be constant then the moment condition $EX_t^2 < \infty$ is required. These assumptions will be made more explicit in specific examples discussed in Section 3 below. Right now, it has to be emphasized that **A9** is less restrictive for WLSE than for the QMLEs.

Remark 2.1 (The WLS estimators avoid boundary problems) *Consider the case of positive observations (for instance* (X_t) *represents a time series of counts or volumes). For the estimators in* (1.6)–(1.8) *be well defined, it is necessary to be able to compute* $\log\left(\widetilde{\lambda}_t(\theta)\right)$ *for all* $\theta \in \Theta$*. For this reason, the condition*

$$\lambda : [0,\infty)^\infty \times \Theta \to [\underline{\lambda},\infty) \quad \textit{for some } \underline{\lambda} > 0 \tag{2.2}$$

is imposed for these QMLEs. In the INGARCH case (1.2)*, the latter condition is satisfied by imposing* $\omega \geq \underline{\lambda}$*,* $\alpha_i \geq 0$ *and* $\beta_j \geq 0$*. Indeed, if for instance* $\alpha < 0$ *is allowed, then* $\lambda_t(\theta) := \omega + \alpha X_{t-1} + \beta\lambda_{t-1}(\theta)$ *can take negative values with non zero probability, and the QMLEs may fail. When one or several coefficients in* (1.2) *are equal to zero,* θ_0 *thus lies at the boundary of* Θ*, and* **A9** *is not satisfied. In this situation, appearing in particular when testing the significance of the INGARCH coefficients, Ahmad and Francq (2016) showed that the PQMLE has a non Gaussian asymptotic distribution, which entails serious practical difficulties. For the WSLE, it is possible to have* $\widetilde{\lambda}_t(\theta) < 0$ *for some values of* θ*—although we must have* $\lambda_t(\theta_0) \geq 0$ *for positive observations—and thus* **A9** *may hold even if* θ_0 *has zero components (see Section 3.1).*

Theorem 2.1 *Under the assumptions* **A1**-**A5***, and* (1.10)

$$\widehat{\theta}_{1WLS} \to \theta_0 \qquad a.s. \qquad as \qquad n \to \infty. \tag{2.3}$$

If in addition **A6**-**A9** *hold, as* $n \to \infty$

$$\sqrt{n}\left(\widehat{\theta}_{1WLS} - \theta_0\right) \xrightarrow{d} \mathcal{N}(0,\Sigma) \qquad \Sigma = \Sigma(\theta_0,w) = J^{-1}(\theta_0,w)\, I(\theta_0,w)\, J^{-1}(\theta_0,w)\,. \tag{2.4}$$

Note that the consistency of the two-stage WLSE cannot be directly deduced from that of the one-step WLSE because, contrary to w_t, $\widehat{w}_{t,n}$ is not $\mathcal{F}_t$-measurable. Let $\widetilde{v}_t^*(\xi) = v^*\left(X_{t-1}, X_{t-2}, ..., X_1, \widetilde{X}_0, \widetilde{X}_{-1}, \ldots; \xi\right)$, so that $\widehat{w}_{t,n} = \widetilde{v}_t^*(\widehat{\xi}_n)$, and let $w_t = v_t^*(\xi_0^*)$. From now on, K denotes a generic positive constant, or a positive random variable $\mathcal{F}_0$-measurable, and ρ a generic constant belonging to $[0,1)$. For consistency of the two-stage WLSE, we replace **A4** by the following assumption.

A4* There exists $\underline{\sigma} > 0$ such that, almost surely, $w_t > \underline{\sigma}$ and $\widehat{w}_{t,n} > \underline{\sigma}$ for n large enough. Assume $\widehat{\xi}_n$ is a strongly consistent estimator of ξ_0^*, the function $\upsilon_t^*(\cdot)$ is almost surely continuously differentiable,

$$\sup_{\xi \in V(\xi_0^*)} |\widetilde{\upsilon}_t^*(\xi) - \upsilon_t^*(\xi)| \leq K\rho^t \quad \text{and} \quad E\frac{1}{w_t} \sup_{\xi \in V(\xi_0^*)} \left\| \frac{\partial \upsilon_t^*(\xi)}{\partial \xi} \right\| \sup_{\theta \in \Theta} \{X_t - \lambda_t(\theta)\}^2 < \infty, \quad (2.5)$$

where $V(\xi_0^*)$ is a neighborhood of ξ_0^*. Moreover, assume

$$E \sup_{\theta \in \Theta} |X_t - \lambda_t(\theta)|^s < \infty \quad \text{for some } s > 0. \quad (2.6)$$

To show the asymptotic normality, we need to slightly modify other assumptions. First of all, when υ_t is well specified, **A6** simplifies as follows.

A6* The matrix $I = E\left(\frac{1}{\upsilon_t} \frac{\partial \lambda_t(\theta_0) \partial \lambda_t(\theta_0)}{\partial \theta \partial \theta'}\right)$ exists and is invertible.

Let **A7*** be obtained by adding in **A7** the assumption that $\sqrt{n}\left(\widehat{\xi}_n - \xi_0^*\right) = O_P(1)$ and

$$E\frac{1}{w_t} \sup_{\xi \in V(\xi_0^*)} \left\| \frac{\partial \upsilon_t^*(\xi)}{\partial \xi} \right\|^2 \left[1 + \sup_{\theta \in V(\theta_0)} \{X_t - \lambda_t(\theta)\}^2\right] < \infty. \quad (2.7)$$

Let **A8*** be the assumption obtained by replacing $|\widetilde{w}_t - w_t|$ by $\sup_{\xi \in V(\xi_0^*)} |\widetilde{\upsilon}_t(\xi) - \upsilon_t(\xi)|$ in **A8**, for some neighborhood $V(\xi_0^*)$ of ξ_0^*.

The following theorem establishes the asymptotic distribution of the two-stage WLSE when the conditional variance is well specified ($i.e.$ $\upsilon_t^*(\xi_0^*) = \upsilon_t$) or when it is misspecified, and shows its relative efficiency with respect to the one-step WLSE under correct specification of υ_t.

Theorem 2.2 *Under* **A1**-**A3**, (1.10), **A4*** *and* **A5** *(which is satisfied when υ_t is well specified)*

$$\widehat{\theta}_{2WLS} \to \theta_0 \qquad a.s. \qquad as \quad n \to \infty. \quad (2.8)$$

Under the previous assumptions and **A6**, **A7***, **A8*** *and* **A9**, *as $n \to \infty$,*

$$\sqrt{n}\left(\widehat{\theta}_{2WLS} - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, \Sigma) \qquad \Sigma = \Sigma(\theta_0, w) = J^{-1}(\theta_0, w)\, I(\theta_0, w)\, J^{-1}(\theta_0, w). \quad (2.9)$$

If in addition the conditional variance is well specified up to a positive constant, that is (1.5) *and* (1.13) *hold with* $\xi_0^* = \xi_0$ *and* $\upsilon^*(\cdot) = k\upsilon(\cdot)$ *for some* $k > 0$, *then* **A6** *can be replaced by* **A6*** *and*

$$\sqrt{n}\left(\widehat{\theta}_{2WLS} - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, I^{-1}\right) \qquad as \qquad n \to \infty. \tag{2.10}$$

Moreover the matrix $\Sigma - I^{-1}$ *is positive semi-definite.*

2.2 The linear conditional mean case

Assume that $X_t \geq 0$ almost surely and that the conditional distribution of X_t given $\mathcal{F}_{t-1}$, denoted by F_{λ_t}, depends on its conditional mean λ_t (and maybe of other fixed parameters). Consider the case where λ_t follows the linear model (1.2). We assume that the stochastic order of the cdf increases with its mean. More precisely, let F_λ be a family of cumulative distribution functions indexed by the mean $\lambda = \int y dF_\lambda(y) \in [0, \infty)$. Assume that, within this family, the stochastic order is equal to the mean order, *i.e.*

$$\lambda \leq \lambda^* \quad \Rightarrow \quad F_\lambda(x) \geq F_{\lambda^*}(x), \quad \forall x \in \mathbb{R}. \tag{2.11}$$

Aknouche and Francq (2019) showed that if $P\left(X_t \leq x \mid \mathcal{F}_{t-1}\right) = F_{\lambda_t}(x)$ and λ_t satisfies (1.2), then **A1** holds true when $\{F_\lambda, \lambda \in (0, \infty)\}$ satisfies (2.11) and

$$\sum_{i=1}^{q} \alpha_{0i} + \sum_{j=1}^{p} \beta_{0j} < 1. \tag{2.12}$$

Moreover, the solution is such that $EX_t < \infty$. By Remark 2.1 in Ahmad and Francq (2016), Assumption **A2** is satisfied when

$$\sum_{j=1}^{p} \beta_j < 1 \text{ for all } \theta \in \Theta. \tag{2.13}$$

In the latter reference, it is also shown that **A3** is satisfied if $q > 0$ and

$$\begin{gathered}\mathcal{A}_{\theta_0}(z) := \sum_{i=1}^{q} \alpha_{0i} z^i \text{ and } \mathcal{B}_{\theta_0}(z) := 1 - \sum_{i=1}^{p} \beta_{0i} z^i \text{ have no common root,} \\ \text{at least one } \alpha_{0i} \neq 0 \text{ for } i = 1, \dots, q, \text{ and } \beta_{0p} \neq 0 \text{ if } \alpha_{0q} = 0.\end{gathered} \tag{2.14}$$

Now suppose that the weighting sequence $\widetilde{w}$ is defined by

$$\widetilde{w}_t = c + aX_{t-1} + b\widetilde{w}_{t-1}$$

with $c > 0$, $a > 0$ and $b \in (0, 1)$. We thus have $w_t = \sum_{i=0}^{\infty} b^i (c + aX_{t-i-1})$ and

$$w_t - \widetilde{w}_t = b^{t-1}(w_1 - \widetilde{w}_1) = b^{t-1} \sum_{i=0}^{\infty} b^i a \left(X_{-i} - \widetilde{X}_{-i} \right)$$

with, for instance, $\widetilde{X}_t = 0$ for $t \leq 0$, and thus $\widetilde{w}_1 = c$. By the Borel-Cantelli lemma, it is then easy to show that **A4** holds true. It is also clear that **A4** holds true for many other forms of the weighting sequence $\widetilde{w}$. Assumptions such as **A5**, as well as the choice of the weighting sequence for the two-stage estimator, depend on the particular form of F_λ and are thus discussed in Section 3 below.

Let us discuss the other assumptions in the case $p = q = 1$, the results extending to general orders p and q with the same arguments but heavier notations. We have

$$\lambda_t(\theta) - \widetilde{\lambda}_t(\theta) = \beta \left\{ \lambda_{t-1}(\theta) - \widetilde{\lambda}_{t-1}(\theta) \right\} = \beta^{t-1} \sum_{i=0}^{\infty} \beta^i \alpha \left(X_{-i} - \widetilde{X}_{-i} \right)$$

and

$$\frac{\partial \lambda_t(\theta)}{\partial \theta} = \begin{pmatrix} 1 \\ X_{t-1} \\ \lambda_{t-1}(\theta) \end{pmatrix} + \beta \frac{\partial \lambda_{t-1}(\theta)}{\partial \theta}.$$

This entails that

$$a_t \leq K\rho^t, \quad b_t \leq Kt\rho^t, \quad \sup_{\theta \in \Theta} |\lambda_t(\theta)| \leq K \sum_{i=0}^{\infty} \rho^i \{1 + |X_{t-i}|\}$$

and

$$\sup_{\theta \in \Theta} \left\| \frac{\partial \lambda_t(\theta)}{\partial \theta} \right\| + \sup_{\theta \in \Theta} \left\| \frac{\partial^2 \lambda_t(\theta)}{\partial \theta \partial \theta'} \right\| \leq K \sum_{i=0}^{\infty} \rho^i \left(1 + |X_{t-i}| + \sup_{\theta \in \Theta} |\lambda_{t-i}(\theta)| \right). \tag{2.15}$$

It follows that, for all weighting sequence satisfying (1.10) and **A4**, Assumptions **A7** is satisfied whenever $EX_t^2 < \infty$. By the Borel-Cantelli lemma and Markov inequality, we also deduce that, for weighting sequences satisfying

$$|w_t - \widetilde{w}_t| \leq K\rho^t, \tag{2.16}$$

A8 is satisfied under the same moment condition. The existence of $I(\theta_0, w)$ for any sequence $w_t \geq \underline{w} > 0$ is ensured by the moment condition $EX_t^4 < \infty$. By the arguments given in Remark 2.3 of Ahmad and Francq (2016), $J(\theta_0, w)$ is invertible under the identifiability condition (2.14). Assumptions **A6** is thus satisfied when $EX_t^4 < \infty$. When the weighting sequence is optimally chosen, the moment conditions are weaker. In particular Assumptions **A6*** is satisfied when $EX_t^2 < \infty$. Now let us further discuss Assumption **A9**, for simplicity in the case $p = q = 1$. For the reasons given in Remark 2.1, for computing the PQMLE the components of θ must be positively constrained, so that (2.2) holds true. The parameter space of the PQMLE is thus typically chosen of the form

$$\Theta = [\underline{\omega}, \overline{\omega}] \times [0, \overline{\alpha}] \times [0, \overline{\beta}], \tag{2.17}$$

with $0 < \underline{\omega} < \overline{\omega}$, $0 < \overline{\alpha}$ and $0 < \overline{\beta} < 1$ (the last inequality ensuring (2.13)). The WLS estimators can be computed without imposing any positivity constraints, so that the parameter space can be chosen, for instance, of the form

$$\Theta = [-\overline{\omega}, \overline{\omega}] \times [-\overline{\alpha}, \overline{\alpha}] \times [-\overline{\beta}, \overline{\beta}]. \tag{2.18}$$

When Θ is like (2.17), Assumption **A9** is quite restrictive because it precludes, in particular, a parameter of the form $\theta_0 = (\omega_0, \alpha_0, 0)$, *i.e.* the interesting situation where the DGP is an Integer ARCH (see Section 3.4 below). On the contrary, for Θ of the form (2.18), Assumption **A9** is always satisfied, provided $\overline{\omega}$, $\overline{\alpha}$ and $\overline{\beta}$ are chosen large enough.

2.3 Optimality of the 2WLSE

Under **A1**-**A3**, assumptions similar to **A6**-**A8**, and **A9** with (2.2) (see Remark 2.1), Ahmad and Francq (2016) established CAN of the PQMLE in the case of integer-valued observations. They showed that

$$\sqrt{n}\left(\widehat{\theta}_P - \theta_0\right) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}\left(0, \Sigma_P\right), \quad \Sigma_P = J_P^{-1} I_P J_P^{-1}$$

with

$$I_P = E\left(\frac{\upsilon_t(\theta_0)}{\lambda_t^2(\theta_0)} \frac{\partial\lambda_t(\theta_0)\partial\lambda_t(\theta_0)}{\partial\theta\partial\theta'}\right) \quad \text{and} \quad J_P = E\left(\frac{1}{\lambda_t(\theta_0)} \frac{\partial\lambda_t(\theta_0)\partial\lambda_t(\theta_0)}{\partial\theta\partial\theta'}\right).$$

Note that $I_P = I(\theta_0, \omega)$ and $J_P = J(\theta_0, \omega)$ with $\omega = \{\lambda_t\}$. In the same framework, Aknouche *et al.* (2018) showed that under certain regularity conditions we have

$$\sqrt{n}\left(\widehat{\theta}_{NB} - \theta_0\right) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}\left(0, \Sigma_{NB}\right), \quad \Sigma_{NB} = \Sigma(\theta_0, \omega), \quad \omega = \left\{\lambda_t(1+\lambda_t/r)\right\}.$$

For positive observations Aknouche and Francq (2019) gave conditions for

$$\sqrt{n}\left(\widehat{\theta}_{E} - \theta_0\right) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}\left(0, \Sigma_{E}\right), \quad \Sigma_{E} = \Sigma(\theta_0, \omega), \quad \omega = \left\{\lambda_t^2\right\}.$$

Note that, as for the last one, the CAN of the first 2 QMLEs is valid not only for count series but also for positive data in general (see Remark 4.1 in Aknouche and Francq, 2019). The optimal WLSE is never asymptotically less efficient than a QMLE.

Corollary 2.1 *Assume $X_t \geq 0$ almost surely and the CAN of the WLSEs and QMLEs. If the conditional variance is well specified, the two-stage WLSE is asymptotically more efficient than the QMLEs, in the sense that the matrices $\Sigma_P - I^{-1}$, $\Sigma_{NB} - I^{-1}$ and $\Sigma_E - I^{-1}$ are all positive semi-definite.*

We now show that $\widehat{\theta}_{2WLS}$ is asymptotically efficient when the true cdf of X_t belongs to the versatile class of the linear exponential distributions. With respect to some σ-finite measure μ (in general the Lebesgue measure or the counting measure), let f_λ be the density of a real random variable of mean $\lambda = \int f_\lambda(x)d\mu(x)$. Let Λ be a nonempty open subspace of $\mathbb{R}$. It is said that the set $\{f_\lambda,\ \lambda \in \Lambda\}$ constitutes a one-parameter linear exponential family if for all $\lambda \in \Lambda$

$$f_\lambda(x) = h(x)e^{\eta(\lambda)x - a(\lambda)}, \tag{2.19}$$

for some two times differentiable functions $\eta(\cdot)$ and $a(\cdot)$. For example f_λ can be the Exponential density of rate parameter $1/\lambda = -\eta$, or the Poisson distribution with intensity parameter $\lambda = e^\eta$, or the negative binomial distribution with parameters r and $p = r/(\lambda + r)$, assuming that r is fixed.

Corollary 2.2 *Assume* **A1** *where $\lambda_t(\cdot)$ admits continuous second-order derivatives. Suppose that* **A2**, **A3**, **A8** *and* **A9** *are satisfied. Assume also that the conditional distribution of X_t*

given $\lambda_t = \lambda$ has the linear exponential form (2.19)*, and that $\lambda_t(\theta_0)$ belongs almost surely to the interior of Λ. The optimal two-stage WLSE is then asymptotically as efficient as the MLE of θ_0.*

To apply Theorem 2.2, it is necessary to estimate the matrix Σ involved in (2.9). This can be done by using the empirical estimator $\widehat{\Sigma} = \widehat{J}^{-1}\widehat{I}\widehat{J}^{-1}$, where

$$\widehat{J} = \frac{1}{n}\sum_{t=1}^{n} \frac{1}{\widehat{w}_{t,n}} \frac{\partial \widetilde{\lambda}_t(\widehat{\theta}_{2WLS})}{\partial \theta} \frac{\partial \widetilde{\lambda}_t(\widehat{\theta}_{2WLS})}{\partial \theta'}, \tag{2.20}$$

$$\widehat{I} = \frac{1}{n}\sum_{t=1}^{n} \frac{\left\{X_t - \widetilde{\lambda}_t(\widehat{\theta}_{2WLS})\right\}^2}{\widehat{w}_{t,n}^2} \frac{\partial \widetilde{\lambda}_t(\widehat{\theta}_{2WLS})}{\partial \theta} \frac{\partial \widetilde{\lambda}_t(\widehat{\theta}_{2WLS})}{\partial \theta'}. \tag{2.21}$$

To estimate the matrix Σ involved in (2.4), it suffices to replace $\widehat{w}_{t,n}$ and $\widetilde{\theta}_{2WLS}$ by w_t and $\widetilde{\theta}_{1WLS}$ in the previous matrices.

3 Application to particular models

We now give primitive conditions ensuring CAN of the WLS estimators for some specific count time series models, an ACD model and a GAS model. We compare the relative asymptotic efficiency of the WLSE with respect to the MLE and QMLEs.

3.1 The Poisson INGARCH model

A leading example of count time series satisfying (1.1) is the Poisson Integer GARCH model proposed by Heinen (2003), in which the distribution of X_t conditional on $\mathcal{F}_{t-1}$ is Poisson $\mathcal{P}(\lambda_t)$ with intensity parameter $\lambda_t = \lambda_t(\theta_0)$ of the form (1.2), where $\omega_0 > 0$, $\alpha_{0i} \geq 0$, $\beta_{0j} \geq 0$. Ferland et *al* (2006) showed that under the condition (2.12) there exists a strictly stationarity solution to the Poisson INGARCH model. The ergodicity of the solution has been shown by Davis and Liu (2016). As discussed in Section 2.2, the result is not only true for the Poisson cdf, but for any class of conditional distributions satisfying (2.11). Note also that under the condition (2.12) we have $EX_t^r < \infty$ for any $r > 0$ (see Christou and Fokianos, 2014). Since

$E\upsilon_t = E\lambda_t < \infty$ under (2.12), **A5** is satisfied for any sequence of weight $w_t > 0$. Using Section 2.2 and Theorem 2.1, we thus have the following result.

Corollary 3.1 *Assume that* $X_t \mid \mathcal{F}_{t-1} \sim \mathcal{P}(\lambda_t)$ *where* $\lambda_t = \lambda_t(\theta_0)$ *follows* (1.2) *with* (2.12) *and* (2.14). *Assume* $\theta_0 \in \Theta$ *with* (2.13). *For any sequence of weights* (w_t) *satisfying* (1.10) *and* (2.16), *the WLSE is strongly consistent in the sense* (2.3). *When* $\theta_0 \in \overset{\circ}{\Theta}$ *the estimator is asymptotically normal, in the sense* (2.4).

For the two-stage estimator, let us take the weighting sequence $\widehat{w}_{t,n} = \widetilde{\lambda}_t\left(\widehat{\theta}_{1WLS}\right)$ (which satisfies (1.10) and (2.16)). We then set $\widehat{\theta}_{2WLS} = \widehat{\theta}^{(P)}_{2WLS}$ where

$$\widehat{\theta}^{(P)}_{2WLS} = \arg\min_{\theta\in\Theta} \sum_{t=1}^{n} \frac{\left(X_t - \widetilde{\lambda}_t(\theta)\right)^2}{\widehat{w}_{t,n}}, \quad \widehat{w}_{t,n} = \widetilde{\lambda}_t\left(\widehat{\theta}_{1WLS}\right). \tag{3.1}$$

Using Section 2.2 and Theorem 2.2, it is easy to verify that we have the following result.

Corollary 3.2 *Under the assumptions of Corollary 3.1, and if* Θ *is chosen sufficiently large so that* $\theta_0 \in \overset{\circ}{\Theta}$, *the 2-stage WLSE* $\widehat{\theta}^{(P)}_{2WLS}$ *is CAN with asymptotic variance*

$$\Sigma = E\left(\frac{1}{\lambda_t(\theta_0)}\frac{\partial\lambda_t(\theta_0)\partial\lambda_t(\theta_0)}{\partial\theta\partial\theta'}\right)^{-1}.$$

Note that, in accordance with Corollary 2.2, $\widehat{\theta}^{(P)}_{2WLS}$ has the same asymptotic distribution as the (PQ)MLE under **A9**. When one or several coefficients α_{0i} or β_{0j} are equal to zero, the CAN of the 2WLSE may still hold (if Θ is chosen large enough), whereas the asymptotic distribution of the (PQ)MLE is more complicated (see the previous discussion and Ahmad and Francq, 2016).

3.2 The Exponential ACD model

Denote by Exp(λ) the exponential distribution of mean λ, which has the density $f(x) = \lambda^{-1}\exp(-x/\lambda)1_{x>0}$. Assume the standard ACD model (1.3) where λ_t follows (1.2) and $z_t \sim$ Exp(1). In this case, the optimal 2-stage WLSE is

$$\widehat{\theta}^{(E)}_{2WLS} = \arg\min_{\theta\in\Theta} \sum_{t=1}^{n} \frac{\left(X_t - \widetilde{\lambda}_t(\theta)\right)^2}{\widehat{w}_{t,n}}, \quad \widehat{w}_{t,n} = \widetilde{\lambda}^2_t\left(\widehat{\theta}_{1WLS}\right). \tag{3.2}$$

For simplicity the following result concerns the first-order model $p = q = 1$, but it could be easily extended to higher-orders.

Corollary 3.3 *Let the ACD model* $X_t \mid \mathcal{F}_{t-1} \sim Exp(\lambda_t)$ *where* $\lambda_t = \lambda_t(\theta_0)$ *follows* (1.2) *with* $p = q = 1$ *and* $\theta_0 = (\omega_0, \alpha_0, \beta_0)$. *Assume that* $E \log(\alpha_0 z_1 + \beta_0) < 0$ *and* $\theta_0 \in \Theta$ *where* Θ *is a compact subset of* $(0, \infty)^2 \times [0, 1)$. *For any sequence of weights* (w_t) *satisfying* (1.10), (2.16) *and* $E(\lambda_t^2/w_t) < \infty$, *the WLSE is strongly consistent in the sense* (2.3). *If* $(\alpha_0 + \beta_0)^2 + \alpha_0^2 < 1$, *then the WLSE is strongly consistent for any sequence of weights* (w_t) *satisfying* (1.10) *and* (2.16). *When, moreover,* $\theta_0 \in \overset{\circ}{\Theta}$ *and*

$$24\alpha_0^4 + 24\alpha_0^3\beta_0 + 12\alpha_0^2\beta_0^2 + 4\alpha_0\beta_0^3 + \beta_0^4 < 1 \tag{3.3}$$

the estimator is asymptotically normal, in the sense (2.4). *The optimal 2-stage WLSE is* $\widehat{\theta}_{2WLS}^{(E)}$. *Under the previous assumptions, this estimator is CAN with asymptotic variance*

$$\Sigma = E\left(\tfrac{1}{\lambda_t^2(\theta_0)}\tfrac{\partial\lambda_t(\theta_0)\partial\lambda_t(\theta_0)}{\partial\theta\partial\theta'}\right)^{-1}. \tag{3.4}$$

Comments similar to those in the last section can be made. The 2WLSE $\widehat{\theta}_{2WLS}^{(E)}$ has the same asymptotic distribution as the MLE $\widehat{\theta}_E$, but does not suffer from boundary problems.

3.3 The Negative Binomial-S-INGARCH model

A random variable X follows a negative binomial, $X \sim \mathrm{NB}\,(r, p)$, of parameters $r > 0$ and $p \in (0, 1)$ if

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

The parameters are related to the first and second order moments by

$$EX = \frac{(1 - p)r}{p} \qquad \text{and} \qquad \mathrm{Var}(X) = \frac{(1 - p)r}{p^2}. \tag{3.5}$$

Inspired by Cameron and Trivedi (1998, p. 73), we now introduced a dynamic version of the negative binomial distribution with a particular parameterization for $r = r_t$ and $p = p_t$. The process $\{X_t, t \in \mathbb{Z}\}$ is said to follow a Negative Binomial-S-INGARCH (NB-S-INGARCH) model if

$$X_t \mid \mathcal{F}_{t-1} \sim \mathrm{NB}\,(r_t, p_t)\,, \qquad p_t = \frac{r_t}{r_t + \lambda_t} \qquad r_t = \varsigma_0 \lambda_t^{2-S}, \tag{3.6}$$

where $S \in \mathbb{R}$, $\varsigma_0 > 0$ and, as in the Poisson INGARCH, λ_t follows (1.2). With this parameterization, in view of (3.5), we have (1.1) and (1.5) with

$$\upsilon_t = \frac{(1-p_t)\, r_t}{p_t^2} = \lambda_t \left(1 + \frac{\lambda_t^{S-1}}{\varsigma_0}\right). \tag{3.7}$$

Since $\upsilon_t > \lambda_t$, the NB-S-INGARCH model can take into account the conditional overdispersion that is often observed in count time series (see Christou and Fokianos, 2014). The cdf (3.6) was proposed by Cameron and Trivedi (1998) in the context of regression count data (*i.e* when λ_t depends on exogenous variables, but not on lagged values of X_t). It is clear from (3.7) that the parameter S plays a key role in the NB-S-INGARCH model. The case $S = 1$, corresponding to the Negative Binomial-I-distribution proposed by Cameron and Trivedi (1986), is close to the Poisson distribution when ς_0 is large. Christou and Fokianos (2014) and Ahmad and Francq (2016) considered the $\mathrm{NB}\,(r, p_t)$ distribution with $p_t = r/(r + \lambda_t)$, which corresponds to (3.6) with $S = 2$. Note that the NB-II distribution $\{\mathrm{NB}(r, r/(r+\lambda)), \lambda > 0\}$ belongs to the linear exponential family (2.19), whereas this is not the case for the NB-I distribution $\mathrm{NB}(p(1-p)^{-1}\lambda, p)$. We now detail these two particular models, corresponding to $S = 1$ and $S = 2$.

3.3.1 The Negative Binomial-I-INGARCH

The NB-I-INGARCH model is obtained when $S = 1$ in (3.6), so that $r_t = \varsigma_0 \lambda_t$ and $p_t = \varsigma_0/(\varsigma_0 + 1)$ is constant. Note that $\upsilon_t = \lambda_t \left(1 + \varsigma_0^{-1}\right)$ is proportional to λ_t. Therefore an asymptotically optimal two-stage WLSE is $\widehat{\theta}^{(P)}_{2WLS}$ defined by (3.1).

Corollary 3.4 *Let the NB-I-INGARCH(1,1) model $X_t \mid \mathcal{F}_{t-1} \sim NB(\varsigma_0\lambda_t, \varsigma_0/(\varsigma_0+1))$ where $\varsigma_0 > 0$, $\lambda_t = \lambda_t(\theta_0)$ follows (1.2) with $p = q = 1$ and $\theta_0 = (\omega_0, \alpha_0, \beta_0)$. Assume $\alpha_0 + \beta_0 < 1$ and $\theta_0 \in \Theta$ where Θ is a compact subset of $(0,\infty)^2 \times [0,1)$. For any sequence of weights (w_t) satisfying (1.10) and (2.16), the WLSE is strongly consistent in the sense (2.3). When $\theta_0 \in \overset{\circ}{\Theta}$ the estimator is asymptotically normal, in the sense (2.4). An optimal 2-stage WLSE is $\widehat{\theta}^{(P)}_{2WLS}$. Under the previous assumptions, this estimator is CAN with asymptotic variance*

$$\Sigma = \left(1 + \frac{1}{\varsigma_0}\right) E\left(\tfrac{1}{\lambda_t(\theta_0)} \tfrac{\partial \lambda_t(\theta_0) \partial \lambda_t(\theta_0)}{\partial \theta \partial \theta'}\right)^{-1}.$$

3.3.2 The Negative Binomial-II-INGARCH

In view of (3.7), when $S = 2$ in (3.6), an asymptotically optimal two-stage WLSE is

$$\widehat{\theta}_{2WLS}^{(NB)} = \arg\min_{\theta \in \Theta} \sum_{t=1}^{n} \frac{\left(X_t - \widetilde{\lambda}_t(\theta)\right)^2}{\widehat{w}_{t,n}}, \quad \widehat{w}_{t,n} = \widetilde{\lambda}_t\left(\widehat{\theta}_{1WLS}\right)\left(1 + \frac{\widetilde{\lambda}_t(\widehat{\theta}_{1WLS})}{\widehat{r}}\right), \quad (3.8)$$

where $\widehat{r}$ is a consistent estimator of $r = \varsigma_0$. Noting that

$$E\frac{(X_t - \lambda_t)^2 - \lambda_t}{\lambda_t^2} = \frac{1}{\varsigma_0},$$

one can take the estimator proposed by Gouriéroux *et al.* (1984) in a static negative binomial regression context:

$$\widehat{r} = \left(\frac{1}{n}\sum_{t=1}^{n} \frac{(X_t - \widehat{\lambda}_t)^2 - \widehat{\lambda}_t}{\widehat{\lambda}_t^2}\right)^{-1}, \quad \widehat{\lambda}_t = \widetilde{\lambda}_t(\widehat{\theta}_{1WLS}). \quad (3.9)$$

Corollary 3.5 *Let the NB-II-INGARCH(1,1) model $X_t \mid \mathcal{F}_{t-1} \sim NB(\varsigma_0, \varsigma_0/(\varsigma_0 + \lambda_t))$ where $\varsigma_0 > 0$, $\lambda_t = \lambda_t(\theta_0)$ follows (1.2) with $p = q = 1$ and $\theta_0 = (\omega_0, \alpha_0, \beta_0)$. Assume $\alpha_0 + \beta_0 < 1$ and $\theta_0 \in \Theta$ where Θ is a compact subset of $(0, \infty)^2 \times [0, 1)$. For any sequence of weights (w_t) satisfying (1.10), (2.16) and $E(\lambda_t^2/w_t) < \infty$, the WLSE is strongly consistent in the sense (2.3). If*

$$(\alpha_0 + \beta_0)^2 + \frac{\alpha_0^2}{\varsigma_0} < 1, \quad (3.10)$$

then the WLSE is strongly consistent for any sequence of weights (w_t) satisfying (1.10) and (2.16). If in addition $\theta_0 \in \overset{\circ}{\Theta}$ and

$$(\alpha_0 + \beta_0)^4 + \frac{6\alpha_0^2(\alpha_0 + \beta_0)^2}{\varsigma_0} + \frac{\alpha_0^3(11\alpha_0 + 8\beta_0)}{\varsigma_0^2} + \frac{6\alpha_0^4}{\varsigma_0^3} < 1, \quad (3.11)$$

the estimator is asymptotically normal, in the sense (2.4). An optimal 2-stage WLSE is $\widehat{\theta}_{2WLS}^{(NB)}$. Under the previous assumptions, this estimator is CAN with asymptotic variance

$$\Sigma = \frac{1}{\varsigma_0} E\left(\frac{1}{\lambda_t(\theta_0)(\varsigma_0 + \lambda_t(\theta_0))} \frac{\partial \lambda_t(\theta_0) \partial \lambda_t(\theta_0)}{\partial \theta \partial \theta'}\right)^{-1}.$$

Note that, as Corollary 2.2 implies, $\widehat{\theta}_{2WLS}^{(NB)}$ has the same asymptotic distribution as the (Q)MLE $\widehat{\theta}_{NB}$ (when r is estimated by (3.9), see Aknouche et al., 2018, Theorem 3.3). Therefore, the two-stage WLSE is asymptotically efficient.

3.4 INARCH models

An INARCH model is a particular INGARCH, obtained when λ_t satisfies (1.2) with $p = 0$. In this case, the conditional mean function is linear in θ. Indeed, we have $\lambda_t(\theta) = \theta' \chi_t$ with $\chi_t = (1, X_{t-1}, ..., X_{t-q})'$. A numerically attractive feature of the WLS estimators is that they have explicit forms for estimating INARCH parameters. More precisely, we have

$$\widehat{\theta}_{1WLS} = \left(\sum_{t=1}^{n} \frac{\chi_t \chi_t'}{w_t} \right)^{-1} \sum_{t=1}^{n} \frac{X_t \chi_t}{w_t}. \tag{3.12}$$

If the weight function is chosen of the form $\widehat{w}_{t,n} = \chi_t' \widehat{\theta}_{1WLS}$, we obtain the two-stage WLSE $\widehat{\theta}_{2WLS} = \widehat{\theta}_{2WLS}^{(P)}$, with

$$\widehat{\theta}_{2WLS}^{(P)} = \left(\sum_{t=1}^{n} \frac{\chi_t \chi_t'}{\chi_t' \widehat{\theta}_{1WLS}} \right)^{-1} \sum_{t=1}^{n} \frac{X_t \chi_t}{\chi_t' \widehat{\theta}_{1WLS}}. \tag{3.13}$$

When the cdf of X_t is $\mathcal{P}(\lambda_t)$, the estimator $\widehat{\theta}_{2WLS}^{(P)}$ is efficient, in the sense that it has exactly the same asymptotic distribution as the MLE. More generally, *i.e.* when the cdf of X_t is not necessarily Poisson, the estimator $\widehat{\theta}_{2WLS}^{(P)}$ has the same asymptotic distribution as the Poisson QMLE. The two-stage WLSE is however numerically simpler than the Poisson (Q)MLE because it does not require any numerical optimization.

Assuming a conditional variance equal (or proportional) to that of a NB-II-INGARCH, we obtain the two-stage WLSE $\widehat{\theta}_{2WLS} = \widehat{\theta}_{2WLS}^{(NB)}$, where

$$\widehat{\theta}_{2WLS}^{(NB)} = \left(\sum_{t=1}^{n} \frac{\chi_t \chi_t'}{\chi_t' \widehat{\theta}_{1WLS} \left(1 + \frac{\chi_t' \widehat{\theta}_{1WLS}}{\widehat{r}} \right)} \right)^{-1} \sum_{t=1}^{n} \frac{X_t \chi_t}{\chi_t' \widehat{\theta}_{1WLS} \left(1 + \frac{\chi_t' \widehat{\theta}_{1WLS}}{\widehat{r}} \right)} \tag{3.14}$$

where $\widehat{r}$ is defined by (3.9). Numerical experiments showed that the two estimators $\widehat{\theta}_{2WLS}^{(P)}$ and $\widehat{\theta}_{2WLS}^{(NB)}$ have similar behaviours when the data generating process (DGP) is INGARCH with Poisson or NB-II cdf. For other cdf's (such as the Double-Poisson considered in Section 5 below) the optimal weights can be proportional to the inverse of the conditional mean, which leads to set $\widehat{\theta}_{2WLS} = \widehat{\theta}_{2WLS}^{(Inv)}$ with

$$\widehat{\theta}_{2WLS}^{(Inv)} = \left(\sum_{t=1}^{n} \chi_t' \widehat{\theta}_{1WLS} \chi_t \chi_t' \right)^{-1} \sum_{t=1}^{n} \chi_t' \widehat{\theta}_{1WLS} X_t \chi_t. \tag{3.15}$$

3.5 The INAR(p) model

The p-th order integer-valued autoregressive (INAR(p)) model proposed by Du and Li (1991) is given by the following equation

$$X_t = \alpha_{01} \circ X_{t-1} + ... + \alpha_{0p} \circ X_{t-p} + \varepsilon_t, \quad t \in \mathbb{Z}, \tag{3.16}$$

where $\{\varepsilon_t, t \in \mathbb{Z}\}$ is an *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 $\alpha \circ X = \sum_{i=1}^{X} Y_i$, where $\{Y_i, i \in \mathbb{N}\}$ is an *iid* Bernoulli random sequence which is independent of X with $P(Y_i = 1) = \alpha \in [0, 1]$. It is assumed that conditionally on $\mathcal{F}_{t-1}$, the sequence $\{\alpha_{0i} \circ X_{t-i}, 1 \le i \le p\}$ is independent. Clearly, the INAR(p) model (3.16) is a particular case of (1.2) since

$$E(X_t \mid \mathcal{F}_{t-1}) = \omega_0 + \alpha_{01} X_{t-1} + ... + \alpha_{0p} X_{t-p} = \lambda_t = \chi_t' \theta_0, \tag{3.17}$$

where $\theta_0 = (\omega_0, \alpha_0')'$, $\alpha_0 = (\alpha_{01}, ..., \alpha_{0p})'$ and $\chi_t = (1, X_{t-1}, ..., X_{t-p})'$. The conditional mean $\chi_t'\theta_0$ is linear in the parameter θ_0 and the conditional variance $\upsilon_t = \text{Var}(X \mid \mathcal{F}_{t-1})$ is given by (cf. Zheng et *al*, 2006, p. 413)

$$\upsilon_t = \text{Var}(X_t \mid \mathcal{F}_{t-1}) = \sum_{i=1}^{p} \alpha_{0i}(1 - \alpha_{0i}) X_{t-i} + \sigma_0^2 := \upsilon_t\left(\alpha_0, \sigma_0^2\right). \tag{3.18}$$

That conditional variance depends on the mean parameter α_0 and on the nuisance parameter σ_0^2. Note that a similar INAR(p) specification has been earlier proposed by Alzaid and Al-Osh (1990), but in which $\{\alpha_{0i} \circ X_{t-i}, 1 \le i \le p\}$ is not a sequence of independent variables.

From Du and Li (1991), Model (3.16) admits a strictly stationary and ergodic solution if

$$\alpha_{01} + \alpha_{02} + ... + \alpha_{0p} < 1. \tag{3.19}$$

Thus under this condition **A1** holds. Moreover, the unconditional mean of the model is given by $E(X_t) = \omega_0 / \left(1 - \sum_{i=1}^{p} \alpha_{0i}\right)$. Since $\sigma_0^2 > 0$ then **A3** is satisfied. Assumption **A5** is obviously satisfied by taking a weighting function of the form

$$w_t = c_0 + \sum_{i=1}^{p} c_j X_{t-i}, \tag{3.20}$$

for some positive constants $c_0, ..., c_p$ and $\widetilde{w}_t = w_t$ for $t \geq p+1$. Assumptions **A2** and **A4** are then satisfied. This completes the proof of the consistency of $\widehat{\theta}_{1WLS}$ defined by (3.12). Let

$$\widehat{w}_{t,n} = \upsilon_t\left(\widehat{\theta}_{1WLS}, \widehat{\sigma}^2\right) = \sum_{i=1}^{p} \widehat{\alpha}_i \left(1-\widehat{\alpha}_i\right) X_{t-i} + \widehat{\sigma}^2,$$

where $\widehat{\theta}_{1WLS} = (\widehat{\omega}_1, \widehat{\alpha}_1, ..., \widehat{\alpha}_p)'$ and $\widehat{\sigma}^2$ is a consistent estimate of σ_0^2, for example

$$\widehat{\sigma}^2 = \frac{1}{n-p}\sum_{t=p}^{n}\left[\left(X_t - \widehat{\omega} - \sum_{i=1}^{p}\widehat{\alpha}_i X_{t-i}\right)^2 - \sum_{i=1}^{p}\widehat{\alpha}_i\left(1-\widehat{\alpha}_i\right)X_{t-i}\right]. \tag{3.21}$$

An optimal WLSE of the INAR model is then

$$\widehat{\theta}_{2WLS}^{(INAR)} = \left(\sum_{t=1}^{n}\frac{\chi_t\chi_t'}{\sum_{j=1}^{p}\widehat{\alpha}_j(1-\widehat{\alpha}_j)X_{t-j}+\widehat{\sigma}^2}\right)^{-1}\sum_{t=1}^{n}\frac{X_t\chi_t}{\sum_{j=1}^{p}\widehat{\alpha}_j(1-\widehat{\alpha}_j)X_{t-j}+\widehat{\sigma}^2}.$$

We then obtain the following result.

Corollary 3.6 *Let the INAR model* (3.16). *Assume* (3.19) *and* (3.20). *If* $\theta_0 \in \Theta$, *the WLSE is consistent. If* $\theta_0 \in \overset{\circ}{\Theta}$ *and* $E\epsilon_t^4 < \infty$, *this estimator is asymptotically normal and satisfies* (2.4). *An optimal 2-stage WLSE is* $\widehat{\theta}_{2WLS}^{(INAR)}$, *which is CAN.*

3.6 The GAS Beta-t-ACD model

The equation (1.4) is a Stochastic Recursive Equation (SRE) of the form

$$\lambda_t = \omega_0 + a(z_{t-1})\lambda_{t-1}, \quad a(z) = \alpha_0\frac{\nu_0+1}{\nu_0-2+z}z + \beta_0.$$

Bougerol (1993) and Straumann and Mikosch (2006) developed a general theory of SRE. From these works, or simply by using the Cauchy root test for convergence of positive series, it is known that when $E\log a(z_1) < 0$ there exists a stationary solution, explicitly given by

$$\lambda_t = \omega_0\left\{1 + \sum_{i=1}^{\infty} a(z_{t-1})\cdots a(z_{t-i})\right\}.$$

For practical use, λ_t needs to be written as function of past observations, as in (1.1). When $\lambda_t(\theta) = \lambda\left(X_{t-1}, X_{t-2}, ...; \theta\right)$ is well defined for all $\theta \in \Theta$ the model is said to be uniformly invertible. The condition (2.13) ensures the uniform invertibility of the linear INGARCH model. For a non linear model of the form (1.4), finding invertibility conditions is much more difficult. The problem has been investigated by Blasques, Gorgi, Koopman and Wintenberger (2018). Given a starting value $\widetilde{\lambda}_1(\theta)$, we approximate $\lambda_t(\theta)$ of model (1.4) by

$$\widetilde{\lambda}_t(\theta) = \omega + \beta\widetilde{\lambda}_{t-1}(\theta) + \alpha\frac{\nu+1}{\nu-2+\frac{X_{t-1}}{\widetilde{\lambda}_{t-1}(\theta)}}X_{t-1}, \quad t \geq 2.$$

Under non explicit conditions on Θ, θ_0 and the distribution of z_1, it is known that there exists a stationary solution $\{\lambda_t(\theta)\}$ to the filter

$$\lambda_t(\theta) = \omega + \beta\lambda_{t-1}(\theta) + \alpha\frac{\nu+1}{\nu-2+\frac{X_{t-1}}{\lambda_{t-1}(\theta)}}X_{t-1}, \quad t \in \mathbb{Z},$$

and that there exits $\rho \in (0,1)$ such that

$$\frac{1}{\rho^t}\sup_{\theta\in\Theta}\left|\widetilde{\lambda}_t(\theta) - \lambda_t(\theta)\right| \to 0 \text{ a.s. as } t \to \infty, \tag{3.22}$$

for all $\widetilde{\lambda}_1(\theta)$ belonging to some fixed set of initial values.

Corollary 3.7 *Let the ACD model* (1.3) *where λ_t satisfies the Beta-t updating equation* (1.4)*. Assume $E\log a(z_1) < 0$, the support of the distribution of z_1 contains at least 3 points,* (3.22) *and $\theta_0 = (\omega_0, \alpha_0, \beta_0, \nu_0)' \in \Theta \subset (0,\infty)^2 \times [0,1) \times (2,\infty)$. For any sequence of weights (w_t) satisfying* (1.10), (2.16) *and $E(\lambda_t^2/w_t) < \infty$, the WLSE is strongly consistent in the sense of* (2.3)*. If $Ea^2(z_1) < 1$ then the WLSE is strongly consistent for any sequence of weights (w_t) satisfying* (1.10) *and* (2.16)*. If in addition $\theta_0 \in \mathring{\Theta}$, $Ea^4(z_1) < 1$ and* (3.22) *holds when $\widetilde{\lambda}_t(\theta)$ and $\lambda_t(\theta)$ are replaced by their partial derivatives, the estimator is asymptotically normal, in the sense* (2.4)*. An optimal 2-stage WLSE is $\widehat{\theta}^{(E)}_{2WLS}$, which is CAN with asymptotic variance* (3.4)*.*

4 Data driven choice of the optimal WLSE

We have seen that an asymptotically optimal two-stage WLSE is obtained by taking a sequence of weights $(\widehat{w}_{t,n})$ such that, as $n \to \infty$, $\widehat{w}_{t,n}$ converges to a weight of the form $w_t = c\upsilon_t$ with $c > 0$ and $\upsilon_t = E\{(X_t - \lambda_t)^2 \mid \mathcal{F}_{t-1}\}$.

In other words, up to a positive multiplicative constant c, the optimal weighting sequence is the conditional variance, that is the best predictor of $(X_t - \lambda_t)^2$. It is then natural to select the weighting sequence $\widehat{w}_{t,n}$ by minimizing in $(\widehat{w}_{t,n})$ the MSE-like loss

$$\mathrm{MSE}_n(\widehat{w}_{t,n}) = \min_c \frac{1}{n}\sum_{t=1}^n \left\{ \left(X_t - \widehat{\lambda}_t \right)^2 - c\widehat{w}_{t,n} \right\}^2 = \frac{1}{n}\sum_{t=1}^n \left\{ \left(X_t - \widehat{\lambda}_t \right)^2 - \widehat{c}_n\widehat{w}_{t,n} \right\}^2,$$

with

$$\widehat{c}_n = \frac{\sum_{t=1}^n \left(X_t - \widehat{\lambda}_t \right)^2 \widehat{w}_{t,n}}{\sum_{t=1}^n \widehat{w}_{t,n}^2}.$$

Inspired by Patton (2011), we also investigate the method that selects the two-stage WLSE by minimizing the QLIKE loss

$$\mathrm{QLIK}_n(\widehat{w}_{t,n}) = \frac{1}{n}\sum_{t=1}^n \left\{ \frac{\left(X_t - \widehat{\lambda}_t \right)^2}{\widehat{c}_n\widehat{w}_{t,n}} + \log\left(\widehat{c}_n\widehat{w}_{t,n}\right) \right\}, \qquad \widehat{c}_n = \frac{1}{n}\sum_{t=1}^n \frac{\left(X_t - \widehat{\lambda}_t \right)^2}{\widehat{w}_{t,n}}.$$

The general theoretical justification for using these two loss functions is that

$$EZ = \arg\min_{m\in\mathbb{R}} E(Z-m)^2 = \arg\min_{m>0} E\frac{Z}{m} + \log m, \tag{4.1}$$

where the fist equality requires a random variable Z such that $EZ^2 < \infty$ and the second one requires $Z \geq 0$ and $0 < EZ < \infty$. In agreement with Patton (2011), we found that the method based on the QLIKE loss works much better in practice than that based on the MSE. In accordance with (4.1) and the following asymptotic result, the fact that the MSE selection method does not work very well in practice, is certainly related to the requirement of higher order moments.

Proposition 4.1 *Assume* **A1** *where $\lambda_t(\cdot)$ has the linear form* (1.2) *with* (2.12) *and* (2.14). *Assume $\theta_0 \in \overset{\circ}{\Theta}$ with* (2.13) *and set $\widehat{\lambda}_t = \widetilde{\lambda}_t(\widehat{\theta})$, where $\widehat{\theta}$ is a consistent estimator of θ_0. Let $w_t = v_t^*(\xi_0^*) \in \mathcal{F}_{t-1}$ be an assumed parametric specification of the conditional variance of X_t, and let its estimation $\widehat{w}_{t,n} = \widetilde{v}_t^*(\widehat{\xi}_n)$. Assume there exists $\underline{\sigma} > 0$ such that, almost surely, $w_t > \underline{\sigma}$, the estimator $\widehat{\xi}_n$ converges almost surely to ξ_0^*, the function $v_t^*(\cdot)$ is almost surely continuously differentiable,*

$$\sup_{\xi\in V(\xi_0^*)} |\widetilde{v}_t^*(\xi) - v_t^*(\xi)| \leq K\rho^t, \quad E\sup_{\xi\in V(\xi_0^*)} v_t^{*2}(\xi) < \infty, \quad E\sup_{\xi\in V(\xi_0^*)} \left\| \frac{\partial v_t^*(\xi)}{\partial \xi} \right\|^2 < \infty, \tag{4.2}$$

for some neighbourhood $V(\xi_0^*)$ of ξ_0^*. Let another sequence of weights (w_t^*) and its approximation $(\widehat{w}_{t,n}^*)$ satisfying the same assumptions (for another potential parametric specification of the conditional variance).

If $EX_t^4 < \infty$ and

$$0 < MSE(w_t) < MSE(w_t^*), \quad MSE(w_t) := \min_c E\left\{(X_t - \lambda_t)^2 - cw_t\right\}^2,$$

then, almost surely

$$(\widehat{w}_{t,n}) = \arg\min\{MSE_n(\widehat{w}_{t,n}^*), MSE_n(\widehat{w}_{t,n})\} \quad \textit{for } n \textit{ large enough.}$$

If $E(\upsilon_t/w_t) < \infty$ and $Ew_t^s < \infty$ for some $s > 0$, the last two conditions in (4.2) are replaced by

$$\left\|\frac{X_t^2}{w_t}\right\|_{p_1} + \left\|\frac{1}{w_t}\sup_{\xi\in V(\xi_0^*)}\left\|\frac{\partial\upsilon_t^*(\xi)}{\partial\xi}\right\|\right\|_{p_2} + \left\|\sup_{\xi\in V(\xi_0^*)}\frac{w_t}{\upsilon_t^*(\xi)}\right\|_{p_3} < \infty, \tag{4.3}$$

$$\left\|\frac{X_t}{\sqrt{w_t}}\right\|_{p_4} + \left\|\frac{1}{\sqrt{w_t}}\sup_{\theta\in V(\theta_0)}\left\|\frac{\partial\lambda_t(\theta)}{\partial\theta}\right\|\right\|_{p_5} < \infty \tag{4.4}$$

$$\left\|\frac{X_t}{\sqrt{w_t}}\right\|_{p_6} + \left\|\frac{1}{\sqrt{w_t}}\sup_{\theta\in V(\theta_0)}\|\lambda_t(\theta)\|\right\|_{p_7} + \left\|\frac{1}{w_t}\sup_{\xi\in V(\xi_0^*)}\left\|\frac{\partial\upsilon_t^*(\xi)}{\partial\xi}\right\|\right\|_{p_8} + \left\|\sup_{\xi\in V(\xi_0^*)}\frac{w_t}{\upsilon_t^*(\xi)}\right\|_{p_9} < \infty, \tag{4.5}$$

$$\left\|\frac{1}{w_t}\sup_{\theta\in V(\theta_0)}\|\lambda_t(\theta)\|^2\right\|_{p_{10}} + \left\|\frac{1}{w_t}\sup_{\xi\in V(\xi_0^*)}\left\|\frac{\partial\upsilon_t^*(\xi)}{\partial\xi}\right\|\right\|_{p_{11}} + \left\|\sup_{\xi\in V(\xi_0^*)}\frac{w_t}{\upsilon_t^*(\xi)}\right\|_{p_{12}} < \infty, \tag{4.6}$$

$$\left\|\frac{1}{\sqrt{w_t}}\sup_{\theta\in V(\theta_0)}\|\lambda_t(\theta)\|\right\|_{p_{13}} + \left\|\frac{1}{\sqrt{w_t}}\sup_{\theta\in V(\theta_0)}\left\|\frac{\partial\lambda_t(\theta)}{\partial\theta}\right\|\right\|_{p_{14}} < \infty \tag{4.7}$$

for some neighbourhood $V(\theta_0)$ of θ_0 and some neighbourhood $V(\xi_0^*)$ of ξ_0^*, where the p_i's are positive numbers such that $\sum_{i=1}^3 p_i = \sum_{i=4}^5 p_i = \sum_{i=6}^9 p_i = \sum_{i=10}^{12} p_i = \sum_{i=13}^{14} p_i = 1$, and

$$0 < QLIK(w_t) < QLIK(w_t^*), \quad QLIK(w_t) := \min_{c>0} E\left\{\frac{(X_t - \lambda_t)^2}{cw_t} + \log(cw_t)\right\},$$

then, almost surely

$$(\widehat{w}_{t,n}) = \arg\min\{QLIK_n(\widehat{w}_{t,n}^*), QLIK_n(\widehat{w}_{t,n})\} \quad \textit{for } n \textit{ large enough.}$$

Remark 4.1 (Moments for the QLIK-based weight selection method) *Note that the moment conditions* (4.3)–(4.7) *are quite mild when* w_t *is well chosen. As for a Poisson INGARCH model, assume that* $X_t \geq 0$ *and*

$$\upsilon_t^*(\xi) = c_0(\xi) + \sum_{i=1}^{\infty} c_i(\xi) X_{t-i}, \quad \xi \mapsto c_i(\xi) \in (0, \infty) \text{ continuous uniformly in } i,$$

with $\inf_\xi c_0(\xi) > \underline{\sigma} > 0$ *and* $c_i(\xi) \sim K\rho^i$ *as* $i \to \infty$*, where* $K > 0$ *and* $\rho \in (0,1)$*. Using the inequality* $x/(1+x) \leq x^s$ *for all* $x \geq 0$ *and* $s \in (0,1)$*, we have*

$$\sup_{\xi \in V(\xi_0^*)} \frac{w_t}{\upsilon_t^*(\xi)} \leq \frac{c_0(\xi_0^*)}{\underline{\sigma}} + \sum_{i=1}^{\infty} \frac{c_i(\xi_0^*)}{c_i(\xi)} \frac{\frac{c_i(\xi)X_{t-i}}{\underline{\sigma}}}{1 + \frac{c_i(\xi)X_{t-i}}{\underline{\sigma}}} \leq \frac{c_0(\xi_0^*)}{\underline{\sigma}} + \sum_{i=1}^{\infty} \frac{c_i(\xi_0^*)}{c_i(\xi)} \frac{c_i^s(\xi) X_{t-i}^s}{\underline{\sigma}^s}.$$

Assuming only $EX_t^{s_0} < \infty$ *for some* $s_0 > 0$*, we have* $\|X_t^s\|_p < \infty$ *whenever* s *is chosen small enough (*i.e. $s < s_0/p$*). If* ρ *is sufficiently close to* ρ^**, where* $\rho^* \in (0,1)$ *such that* $c_i(\xi_0^*) \sim K\rho^{*i}$*, we have*

$$\left\| \frac{c_i(\xi_0^*)}{c_i(\xi)} c_i^s(\xi) X_{t-i}^s \right\|_p \sim K \left(\frac{\rho^* \rho^s}{\rho} \right)^i, \qquad \frac{\rho^* \rho^s}{\rho} < 1.$$

It follows that, for any $p \geq 1$*,* $\left\| \sup_{\xi \in V(\xi_0^*)} \frac{w_t}{\upsilon_t^*(\xi)} \right\|_p < \infty$ *when* $V(\xi_0^*)$ *is sufficiently small. Assuming, as it is the case when the power series* $\sum_{i=1}^{\infty} c_i(\xi) z^i$ *is the ratio of two polynomials, that*

$$\left\| \frac{\partial c_i(\xi)}{\partial \xi} \right\| \leq K\rho^i,$$

the previous arguments show that for any $p \geq 1$*,*

$$\left\| \frac{1}{w_t} \sup_{\xi \in V(\xi_0^*)} \left\| \frac{\partial \upsilon_t^*(\xi)}{\partial \xi} \right\| \right\|_p < \infty$$

when $V(\xi_0^*)$ *is sufficiently small. It follows that, in this situation, conditions* (4.3)–(4.7) *can be considerably weakened.*

Note that, applying (4.1) with $Z \sim X_t \mid \mathcal{F}_{t-1}$, under the assumptions of Proposition 4.1, we have $\mathrm{MSE}(\upsilon_t) \leq \mathrm{MSE}(w_t)$ and $\mathrm{QLIK}(\upsilon_t) \leq \mathrm{QLIK}(w_t)$ for any weighting sequence w_t. Therefore, provided the moments and the other regularity conditions hold, the optimal 2-stage WLSE will be asymptotically found by minimizing either MSE_n or QLIK_n over a finite of possible weighting sequence for which one of them converges to cv_t, $c > 0$. When the set of the potential weighting sequences does not contain such an optimal sequence, Proposition 4.1 guarantees that some kind of sub-optimality is however asymptotically found. The following example illustrates that point, as well as the moment conditions required with the MSE and QLIK losses.

Example 4.1 *Consider an ACD model of the form $X_t = \lambda_t z_t$ where (z_t) is iid with distribution Exp(1). Assume that $EX_t^4 < \infty$. Noting that*

$$\arg\min_c E\left\{(X_t - \lambda_t)^2 - cw_t\right\}^2 = \frac{E(X_t - \lambda_t)^2 w_t}{Ew_t^2} = \frac{E\lambda_t^2 w_t}{Ew_t^2},$$

we have

$$MSE(w_t) = 9E\lambda_t^4 - \frac{\left(E\lambda_t^2 w_t\right)^2}{Ew_t^2}.$$

It follows that

$$MSE(\lambda_t^2) = 8E\lambda_t^4 \leq MSE(\lambda_t) = 9E\lambda_t^4 - \frac{(E\lambda_t^3)^2}{E\lambda_t^2} \leq MSE(1) = 9E\lambda_t^4 - (E\lambda_t^2)^2 \qquad (4.8)$$

where the inequalities are strict in general (in particular, when λ_t is not degenerated). If the set of the potential weighting sequences contains $\widehat{w}_{t,n} = \widehat{\lambda_t^2}$ and the assumptions of Proposition 4.1 are satisfied, the optimal WLSE is found when n is large enough by minimising either the MSE_n or the $QLIK_n$ criterion. If we have only the two potential weighting sequences $\widehat{w}_{t,n} = \widehat{\lambda}_t$ and $\widehat{w}_{t,n} = 1$, the MSE_n criterion will select asymptotically the first sequence. We do not know what would be the choice of the $QLIK_n$ criterion in the same situation because

$$QLIK(w_t) = 1 + \log E\frac{\lambda_t^2}{w_t} + E\log w_t$$

does not seem to be explicitly computable.

Recall that the existence of the MSEs in (4.8) require $EX_t^4 < \infty$. Assume that $\lambda_t = \omega_0 + \alpha_0 X_{t-1} + \beta_0 \lambda_{t-1}$. Standard computations show that, in the ACD(1,1) case, $EX_t^4 < \infty$ if and only if

$$\mu_4\alpha_0^4 + 4\mu_3\alpha_0^3\beta_0 + 6\mu_2\alpha_0^2\beta_0^2 + 4\mu_1\alpha_0\beta_0^3 + \beta_0^4 < 1$$

where $\mu_n = n!$. This condition entails strong restrictions on α_0 and β_0. The condition for the existence of a strictly stationary solution to the ACD model is $\gamma := E\log(\alpha_0 z_1 + \beta_0) < 0$. Figure 1 shows that the region of strict stationarity is much wider than that of the existence of fourth-order moments. Under $\gamma < 0$, it is know that $EX_t^s < \infty$ for some $s > 0$, and thus $E\log\lambda_t^k < \infty$ for any k. It follows that $QLIK(\lambda_t^2)$ is finite whenever $\gamma < 0$, and that $QLIK(\lambda_t) < \infty$ (respectively $QLIK(1) < \infty$) iff $E\lambda_t < \infty$ (respectively $E\lambda_t^2 < \infty$).

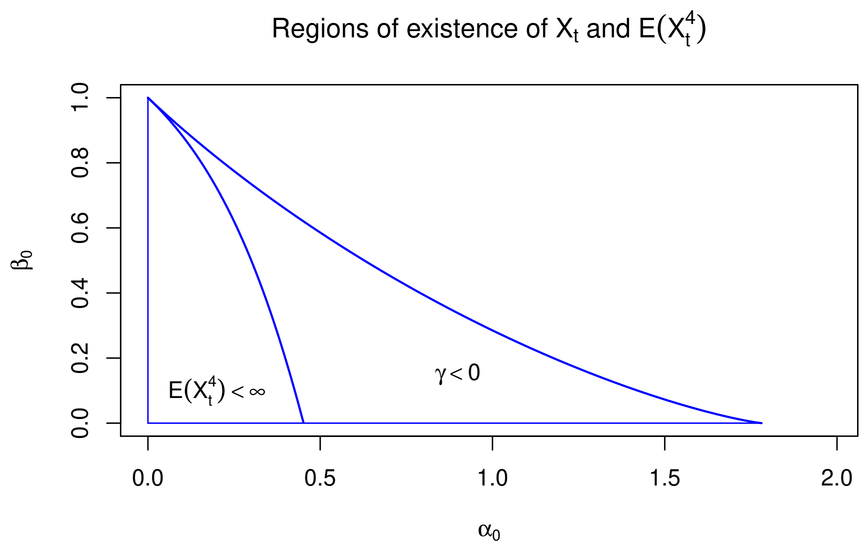

Figure 1: Region of strict stationarity $\gamma < 0$ and region of existence of the fourth-order moment for the ACD process $X_t = \lambda_t z_t$ where $z_t \sim \text{Exp}(1)$ and $\lambda_t = \omega_0 + \alpha_0 X_{t-1} + \beta_0\lambda_{t-1}$.

5 Numerical illustrations

In this section we first present a small Monte Carlo experiment that compares the finite sample performance of different estimators of θ_0. We then applied our methodology for predicting a realized volatility series. Other numerical illustrations are available from the authors.

5.1 A simulation study

We simulated $N = 1000$ independent replications of length $n = 500$ and $n = 2000$ of INARCH(q) models, and compared the finite-sample performance of the following estimators: the PQMLE (1.7), the NBQMLE (1.8) with r=1, the WLSE (1.11) with $\widetilde{w} \equiv 1$, and the two-stage WLS estimators (3.13), (3.14) and (3.15). For choosing between the different versions of the two-stage WLSE, we used the data-driven methods presented in Section 4. Since the criterion based of the QLIK loss works much better than that based of the MSE, we only present the estimator selected by the former (denoted by $\widehat{\theta}^*_{2WLS}$).

When the cdf is Poisson or Negative Binomial, there is no much difference between the estimators (thus we do not present these results). Table 1 displays the results for an INARCH(3) with parameter $\theta_0 = (\omega_0, \alpha_{01}, \alpha_{02}, \alpha_{03}) = (1, 0.3, 0.1, 0.5)$, when the cdf is the Double-Poisson of Efron (1986) of parameters such that the conditional variance is s/λ_t with $s = 50$.[2] As expected, the version $\widehat{\theta}^{(Inv)}_{2WLS}$ of the two-stage WLSE clearly outperforms the other estimators, both in terms of bias and Root Mean Square Error (RMSE) of estimation. Interestingly, the data-chosen WLSE $\widehat{\theta}^*_{2WLS}$ always coincides with the optimal two-stage WLSE.

Of course, we made other numerical experiments, that we do not present here to save space. In particular, we compared the computation time of the different estimators on INARCH(q) models for increasing values of q. We found that, when the number $q + 1$ of parameters becomes large the computation time of the QMLEs tends to be prohibitive because these estimators require numerical optimizations, which is not the case for the WLS estimators. We also performed Monte Carlo experiments showing that, for all the estimators, the estimated standard errors based on the asymptotic theory, using the estimators (2.20) and (2.21), are close to the observed RMSEs on simulations of INGARCH models.

[2] For small values of s the variance is small and, as a consequence, the weighting sequence w_t has little effect on the estimator.

Table 1: Bias and RMSE of estimators of the mean parameters when the DGP is a Double-Poisson INARCH(3).

	ω		α_1		α_2		α_3	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
				$n = 500$				
$\widehat{\theta}_E$	1.178	1.354	-0.046	0.086	-0.016	0.072	-0.054	0.094
$\widehat{\theta}_P$	0.836	0.960	-0.029	0.057	-0.011	0.053	-0.039	0.065
$\widehat{\theta}_{NB}$	1.130	1.294	-0.044	0.080	-0.016	0.069	-0.053	0.089
$\widehat{\theta}_{1WLS}$	0.462	0.596	-0.014	0.042	-0.006	0.043	-0.022	0.047
$\widehat{\theta}_{2WLS}^{(E)}$	1.438	1.890	-0.052	0.101	-0.031	0.100	-0.064	0.114
$\widehat{\theta}_{2WLS}^{(P)}$	0.851	0.984	-0.029	0.057	-0.013	0.055	-0.039	0.066
$\widehat{\theta}_{2WLS}^{(NB)}$	1.006	1.200	-0.034	0.071	-0.019	0.070	-0.044	0.080
$\widehat{\theta}_{2WLS}^{(Inv)}$	0.248	0.479	-0.006	0.038	-0.004	0.041	-0.012	0.041
$\widehat{\theta}_{2WLS}^{*}$	0.248	0.479	-0.006	0.038	-0.004	0.041	-0.012	0.041
				$n = 2000$				
$\widehat{\theta}_E$	1.246	1.306	-0.051	0.065	-0.020	0.050	-0.054	0.068
$\widehat{\theta}_P$	0.773	0.813	-0.029	0.039	-0.009	0.031	-0.034	0.043
$\widehat{\theta}_{NB}$	1.164	1.221	-0.047	0.060	-0.018	0.047	-0.051	0.064
$\widehat{\theta}_{1WLS}$	0.333	0.371	-0.011	0.023	-0.003	0.022	-0.015	0.025
$\widehat{\theta}_{2WLS}^{(E)}$	0.333	0.371	-0.011	0.023	-0.003	0.022	-0.015	0.025
$\widehat{\theta}_{2WLS}^{(P)}$	0.778	0.820	-0.029	0.039	-0.009	0.031	-0.034	0.044
$\widehat{\theta}_{2WLS}^{(NB)}$	0.906	0.960	-0.035	0.047	-0.012	0.037	-0.039	0.051
$\widehat{\theta}_{2WLS}^{(Inv)}$	0.115	0.217	-0.003	0.019	-0.001	0.021	-0.006	0.020
$\widehat{\theta}_{2WLS}^{*}$	0.115	0.217	-0.003	0.019	-0.001	0.021	-0.006	0.020

Table 2: WLS estimation results for the CAT series.

	$\widehat{\omega}$	$\widehat{\alpha}$	$\widehat{\beta}$	QLIK
$\widehat{\theta}_{1WLS}$	0.147 (0.0912)	0.380 (0.0878)	0.580 (0.0895)	3.357
$\widehat{\theta}^{(E)}_{2WLS}$	0.093 (0.0197)	0.342 (0.0233)	0.632 (0.0238)	**1.870**
$\widehat{\theta}^{(P)}_{2WLS}$	0.099 (0.0331)	0.357 (0.0364)	0.616 (0.0383)	2.189
$\widehat{\theta}^{(NB)}_{2WLS}$	0.090 (0.0200)	0.345 (0.0235)	0.631 (0.0243)	1.884
$\widehat{\theta}^{(Inv)}_{2WLS}$	0.298 (0.185)	0.349 (0.135)	0.595 (0.148)	5.231

5.2 Predicting a realized volatility series

Considerable interest has been paid in recent years to modeling and forecasting daily realized volatility, which is defined as an integrate variability of high frequency intra-day asset returns (see e.g. Barndorff-Nielsen and Shephard, 2002). We consider in this subsection the daily series of Caterpillar Inc. (CAT) realized volatility, from 01/04/1999 to 31/12/2008, which corresponds to the sample size $n = 2489$. On this series, we fitted an ACD model with linear conditional mean (1.2). We found that the first orders $p = q = 1$ are sufficient (for larger orders, the usual information criteria AIC and BIC are not smaller, and the estimated additional parameters are not significantly different from zero). To estimate the mean parameter of the ACD model, we used the previously described five WLSEs. Table 2 shows that the estimated values of the parameters are close, while their estimated standard deviations (in parentheses) vary more. The QLIK criterion of Section 4 selects $\widehat{\theta}^{(E)}_{2WLS}$ as the best WLSE. Note also that $\widehat{\theta}^{(E)}_{2WLS}$ and $\widehat{\theta}^{(NB)}_{2WLS}$ (now calculated while replacing the estimate in (3.9) by the value 1) provide almost the same results.

We also compared the performance of the different WLSEs by means of out-of-sample forecasts. Consider the first n_c observations on which we calculate the five WLSEs. The realized volatility forecast at time $t > n_c$ (and horizon 1)

$$\widehat{X}_t = \widehat{\lambda}_t = \widehat{\omega} + \sum_{i=1}^{q} \widehat{\alpha}_i X_{t-i} + \sum_{j=1}^{p} \widehat{\beta}_j \widehat{\lambda}_{t-j}$$

is compared to the actual value X_t, for $t = n_c + 1, ..., n$. We used seven loss functions considered in Paton (2011): the mean square error prediction, the mean absolute error prediction, the mean QLIKE, the mean square log-error prediction, the mean absolute log-error prediction, the mean square root error prediction, and the mean absolute root error prediction, respectively defined by

$$\text{MSEP} = \frac{1}{n} \sum_{t=n_c+1}^{n} \left(X_t - \widehat{\lambda}_t \right)^2, \ \text{MAEP} = \frac{1}{n} \sum_{t=n_c+1}^{n} \left| X_t - \widehat{\lambda}_t \right|, \ \text{MQLIK} = \frac{1}{n} \sum_{t=n_c+1}^{n} \log \widehat{\lambda}_t + \frac{X_t}{\widehat{\lambda}_t},$$

$$\text{MSLEP} = \frac{1}{n} \sum_{t=n_c+1}^{n} \left(\log X_t - \log \widehat{\lambda}_t \right)^2, \quad \text{MALEP} = \frac{1}{n} \sum_{t=n_c+1}^{n} \left| \log X_t - \log \widehat{\lambda}_t \right|,$$

$$\text{MSREP} = \frac{1}{n} \sum_{t=n_c+1}^{n} \left(\sqrt{X_t} - \sqrt{\widehat{\lambda}_t} \right)^2, \quad \text{MAREP} = \frac{1}{n} \sum_{t=n_c+1}^{n} \left(\sqrt{X_t} - \sqrt{\widehat{\lambda}_t} \right)^2.$$

Table 3 displays the loss functions when the learning sample size is $n_c = 500$. Very similar results have been obtained for $n_c = 1000$ and $n_c = 1500$. Using the `R` package `MCS` developed by Bernardi and Catania (2014), which implements the Model Confidence Set procedure of Hansen, Lunde, and Nason (2011), we found that the models estimated by $\widehat{\theta}^{(E)}_{2WLS}$ and $\widehat{\theta}^{(NB)}_{2WLS}$ generally constitute the so-called Superior Set Models. These results comfort those obtained in-sample: the best estimator is $\widehat{\theta}^{(E)}_{2WLS}$, closely followed by $\widehat{\theta}^{(NB)}_{2WLS}$.

6 Proofs of the main results

Proof of Theorem 2.1 Let $L_n(\theta, w)$ and $l_t(\theta, w_t)$ be the random variables obtained by replacing $\widetilde{\lambda}_t(\theta)$ by $\lambda_t(\theta)$ in $\widetilde{L}_n(\theta, w)$ and $\widetilde{l}_t(\theta, w_t)$. In view of **A4** and (1.10), one can assume

Table 3: Loss functions for out-of-sample predictions of the CAT realized volatility.

	$\widehat{\theta}^{(P)}_{2WLS}$	$\widehat{\theta}^{(NB)}_{2WLS}$	$\widehat{\theta}^{(Inv)}_{2WLS}$	$\widehat{\theta}_{1WLS}$	$\widehat{\theta}^{(E)}_{2WLS}$
MSEP	12.7204	12.6798	12.8184	12.7693	12.6741
MAEP	1.5669	1.5552	1.5785	1.5784	1.5518
MQLIK	1.9122	1.9100	1.9143	1.9144	1.9094
MSLEP	0.3736	0.3660	0.3792	0.3807	0.3636
MALEP	0.5024	0.4966	0.5067	0.5077	0.4948
MSREP	0.2752	0.2719	0.2787	0.2784	0.2710
MAREP	0.3992	0.3950	0.4027	0.4032	0.3937

without loss of generality that $w_t \geq \underline{w} > 0$. We have

$$
\begin{aligned}
&\sup_{\theta\in\Theta}\left|l_t(\theta,w_t)-\widetilde{l}_t(\theta,\widetilde{w}_t)\right| \\
&=\sup_{\theta\in\Theta}\left|\frac{\left\{\widetilde{\lambda}_t(\theta)-\lambda_t(\theta)\right\}\left\{\lambda_t(\theta)+\widetilde{\lambda}_t(\theta)-2X_t\right\}}{\widetilde{w}_t}+\frac{(w_t-\widetilde{w}_t)\left\{X_t-\lambda_t(\theta)\right\}^2}{w_t\widetilde{w}_t}\right| \\
&\leq\sup_{\theta\in\Theta}\left|\lambda_t(\theta)-\widetilde{\lambda}_t(\theta)\right|\frac{2}{\underline{w}}\left\{1+|X_t|+\sup_{\theta\in\Theta}|\lambda_t(\theta)|\right\}+|w_t-\widetilde{w}_t|\frac{2}{\underline{w}^2}\left\{X_t^2+\sup_{\theta\in\Theta}\lambda_t^2(\theta)\right\} \qquad (6.1)
\end{aligned}
$$

for t large enough. Therefore, under **A2** and **A4**, by Cesàro's lemma we have

$$\sup_{\theta\in\Theta}\left|\widetilde{L}_n(\theta,\widetilde{w})-L_n(\theta,w)\right|\to 0 \qquad \text{a.s. as } n\to\infty. \qquad (6.2)$$

Now, noting that $\{w_t,\lambda_t(\theta),X_t\}$ is a stationary and ergodic process, $\lim_{n\to\infty} L_n(\theta,w) = El_t(\theta,w_t)\in[0,\infty]$ a.s. Moreover

$$El_t(\theta_0,w_t)=E\frac{(X_t-\lambda_t)^2}{w_t}=E\left\{E\left(\frac{(X_t-\lambda_t)^2}{w_t}\mid\mathcal{F}_{t-1}\right)\right\}=E\frac{\upsilon_t}{w_t}<\infty$$

under **A5**. Obviously, **A3** then implies $El_t(\theta_0,w_t)\leq El_t(\theta,w_t)$ with equality if and only if $\theta=\theta_0$.

The rest of the proof of the consistency (2.3) follows from standard arguments (see *e.g.* the proof of Theorem 2.1 in Ahmad and Francq, 2016).

We now show that the choice of the initial values does not modify the asymptotic distribution of the estimator. Indeed, we have

$$\sqrt{n}\sup_{\theta\in\Theta}\left\|\frac{\partial \widetilde{L}_n(\theta,\widetilde{w})}{\partial\theta}-\frac{\partial L_n(\theta,w)}{\partial\theta}\right\|$$

$$\leq \frac{2}{\sqrt{n}}\sum_{t=1}^{n}\left[\frac{a_t}{\underline{w}}+\frac{|w_t-\widetilde{w}_t|\left\{|X_t|+a_t+\sup_{\theta\in\Theta}|\lambda_t(\theta)|\right\}}{\underline{w}^2}\right]\sup_{\theta\in\Theta}\left\|\frac{\partial\lambda_t}{\partial\theta}\right\|+\frac{b_t\left\{|X_t|+a_t+\sup_{\theta\in\Theta}|\lambda_t(\theta)|\right\}}{\underline{w}},$$

which tends to zero almost surely, by **A8**. Now noting that $\{e_t,\mathcal{F}_t\}_t$, where $e_t=X_t-\lambda_t(\theta_0)$, is a stationary martingale difference sequence, under **A6** we have

$$\sqrt{n}\frac{\partial L_n(\theta_0,w)}{\partial\theta}=\frac{-2}{\sqrt{n}}\sum_{t=1}^{n}\frac{e_t}{w_t}\frac{\partial\lambda_t(\theta_0)}{\partial\theta}\xrightarrow{d}\mathcal{N}\left\{0,4I\left(\theta_0,w\right)\right\}\qquad\text{as } n\to\infty. \tag{6.3}$$

Using Taylor expansions and standard arguments (see *e.g.* the proof of Theorem 2.2 in Ahmad and Francq, 2016), the convergence in law (2.4) is then proven by showing

$$\frac{\partial^2 L_n(\theta_n,w)}{\partial\theta\partial\theta'}\to 2J\left(\theta_0,w\right)\qquad\text{as } n\to\infty \tag{6.4}$$

for any sequence θ_n tending to θ_0 as $n\to\infty$. The convergence result (6.4) can be shown by using the ergodic theorem, the dominated convergence theorem, the continuity of the second order derivatives of $l_t(\cdot,w_t)$ and **A7** (see the proof of Theorem 2.2 where, in a more complex framework, this part of the demonstration is detailed). □

Proof of Theorem 2.2 First note that (2.5) entails that for n large enough

$$|\widehat{w}_{t,n}-w_t|\leq K\rho^t+\left|\upsilon_t^*(\widehat{\xi}_n)-\upsilon_t^*(\xi_0^*)\right|\leq K\rho^t+\|\widehat{\xi}_n-\xi_0^*\|Z_t$$

where $Z_t=\sup_{\xi\in V(\xi_0^*)}\|\partial\upsilon_t^*(\xi)/\partial\xi\|$. Therefore, in view of (6.1) and **A4***, we have

$$\sup_{\theta\in\Theta}\left|l_t(\theta,w_t)-\widetilde{l}_t(\theta,\widehat{w}_{t,n})\right|$$

$$\leq\frac{2a_t}{\underline{\sigma}}\left\{1+|X_t|+\sup_{\theta\in\Theta}|\lambda_t(\theta)|\right\}+\frac{2(K\rho^t+\|\xi_0^*-\widehat{\xi}_n\|Z_t)}{\underline{\sigma}w_t}\sup_{\theta\in\Theta}\left\{X_t-\lambda_t(\theta)\right\}^2.$$

Under (2.6) with $s<2$, we have

$$E\left\{\sum_{t=1}^{\infty}\rho^t\sup_{\theta\in V(\theta_0)}\left\{X_t-\lambda_t(\theta)\right\}^2\right\}^{s/2}\leq\sum_{t=1}^{\infty}\rho^{ts/2}E\sup_{\theta\in V(\theta_0)}|X_t-\lambda_t(\theta)|^s<\infty.$$

Therefore, in the left-hand side of the previous non-strict inequality, the sum into brackets is almost surely finite. It follows that the analogue of (6.2) holds true under **A2** and **A4***. Therefore (2.8) follows as in the proof of Theorem 2.1.

The asymptotic irrelevance of the initial values is shown as in Theorem 2.1, using **A8***. Using the lightened notation $\widehat{\theta} = \widehat{\theta}_{2WLS}$, under **A9** we thus have

$$0 = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{X_t - \widetilde{\lambda}_t(\widehat{\theta})}{\widetilde{\upsilon}_t^{\,*}(\widehat{\xi}_n)} \frac{\partial \widetilde{\lambda}_t(\widehat{\theta})}{\partial \theta} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{X_t - \lambda_t(\widehat{\theta})}{\upsilon_t^*(\widehat{\xi}_n)} \frac{\partial \lambda_t(\widehat{\theta})}{\partial \theta} + o_p(1)$$

as $n \to \infty$. Taylor expansions then yield

$$o_P(1) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{e_t}{\upsilon_t^*(\widehat{\xi}_n)} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} + J_n^* \sqrt{n}(\widehat{\theta}_{2WLS} - \theta_0), \tag{6.5}$$

where the element of the i-th row and j-th column of J_n^* is

$$\frac{1}{n} \sum_{t=1}^{n} \frac{X_t - \lambda_t(\theta^*)}{\upsilon_t^*(\widehat{\xi}_n)} \frac{\partial^2 \lambda_t(\theta^*)}{\partial \theta_i \partial \theta_j} - \frac{1}{n} \sum_{t=1}^{n} \frac{1}{\upsilon_t^*(\widehat{\xi}_n)} \frac{\partial \lambda_t(\theta^*)}{\partial \theta_i} \frac{\partial \lambda_t(\theta^*)}{\partial \partial \theta_j} =: a_n(\theta^*) + b_n(\theta^*),$$

for some θ^* between $\widehat{\theta}_{2WLS}$ and θ_0. Let

$$a_n^*(\theta) = \frac{1}{n} \sum_{t=1}^{n} \frac{X_t - \lambda_t(\theta)}{w_t} \frac{\partial^2 \lambda_t(\theta)}{\partial \theta_i \partial \theta_j}, \qquad b_n^*(\theta) = -\frac{1}{n} \sum_{t=1}^{n} \frac{1}{w_t} \frac{\partial \lambda_t(\theta)}{\partial \theta_i} \frac{\partial \lambda_t(\theta)}{\partial \partial \theta_j}.$$

A Taylor expansion and the convergence of $\widehat{\xi}_n$ to ξ_0^* show that, for n large enough,

$$\sup_{\theta \in V(\theta_0)} |a_n(\theta) - a_n^*(\theta)| \leq \frac{\|\xi_0^* - \widehat{\xi}_n\|}{n} \sum_{t=1}^{n} \frac{\sup_{\xi \in V(\xi_0^*)} \left\| \frac{\partial v_t(\xi)}{\partial \xi} \right\|}{k \underline{\sigma} w_t} \sup_{\theta \in V(\theta_0)} \left| \{X_t - \lambda_t(\theta)\} \frac{\partial^2 \lambda_t(\theta)}{\partial \theta_i \partial \theta_j} \right|.$$

By (2.1) and (2.7), the Cauchy-Schwarz inequality entails

$$E \frac{1}{w_t} \sup_{\xi \in V(\xi_0^*)} \left\| \frac{\partial v_t(\xi)}{\partial \xi} \right\| \sup_{\theta \in V(\theta_0)} \left| \{X_t - \lambda_t(\theta)\} \frac{\partial^2 \lambda_t(\theta)}{\partial \theta_i \partial \theta_j} \right| < \infty.$$

Using also the ergodic theorem and the strong consistency of $\widehat{\xi}_n$, it follows that

$$\sup_{\theta \in V(\theta_0)} |a_n(\theta) - a_n^*(\theta)| \to 0.$$

Now, note that the ergodic theorem, the Lebesgue-dominated convergence theorem and **A7*** entail that for any $\epsilon > 0$

$$\lim_{n \to \infty} \sup_{\theta \in V(\theta_0)} |a_n^*(\theta) - a_n^*(\theta_0)| \leq E \sup_{\theta \in V(\theta_0)} \left| \frac{X_t - \lambda_t(\theta)}{w_t} \frac{\partial^2 \lambda_t(\theta)}{\partial \theta_i \partial \theta_j} - \frac{e_t}{w_t} \frac{\partial^2 \lambda_t(\theta_0)}{\partial \theta_i \partial \theta_j} \right| < \epsilon$$

if $V(\theta_0)$ is small enough. The ergodic theorem also entails $a_n^*(\theta_0) \to 0$ a.s. By similar arguments, it can be shown that $\lim_{n\to\infty} b_n(\theta^*) = \lim_{n\to\infty} b_n^*(\theta_0) = -J(\theta_0, w)(i,j)$. We thus have shown that $J_n^* \to -J(\theta_0, w)$. In view of (6.5), it remains to show that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{e_t}{\upsilon_t^*(\widehat{\xi}_n)} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \xrightarrow{d} \mathcal{N}\left(0, I(\theta_0, w)\right). \tag{6.6}$$

For $i \in \{1, \ldots, d\}$, we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{e_t}{\upsilon_t^*(\widehat{\xi}_n)} \frac{\partial \lambda_t(\theta_0)}{\partial \theta_i} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{e_t}{w_t} \frac{\partial \lambda_t(\theta_0)}{\partial \theta_i} + \lambda_n'(\xi_n)\sqrt{n}\left(\widehat{\xi}_n - \xi_0\right)$$

where ξ_n is between $\widehat{\xi}_n$ and ξ_0^*, and

$$\lambda_n(\xi) = \frac{-1}{n} \sum_{t=1}^{n} \frac{e_t}{\upsilon_t^{*2}(\xi)} \frac{\partial \lambda_t(\theta_0)}{\partial \theta_i} \frac{\partial \upsilon_t^*(\xi)}{\partial \xi}.$$

Noting that $E\lambda_n(\xi) = 0$ and using the consistency of $\widehat{\xi}_n$, already used arguments show that $\lambda_n(\xi_n) \to 0$ a.s. Since $\sqrt{n}\left(\widehat{\xi}_n - \xi_0\right) = O_P(1)$, we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{e_t}{\upsilon_t^*(\widehat{\xi}_n)} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{e_t}{w_t} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} + o_P(1),$$

and (6.6) follows from the CLT for stationary square integrable martingale differences.

To complete the proof, notice that

$$\mathrm{Var}\left(J^{-1}(\theta_0, w) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{e_t}{w_t} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} - I^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{e_t}{\upsilon_t} \frac{\partial \lambda_t(\theta_0)}{\partial \theta}\right) = \Sigma - I^{-1}. \qquad \square$$

Proof of Corollary 2.1. This is a consequence of the last result of Theorem 2.2, noting that Σ_E, Σ_P and Σ_P are each equal to the matrix $\Sigma(\theta_0, w)$ for some particular w. $\square$

Proof of Corollary 2.2. Recall that a random variable X with mean λ and density f_λ belonging to a regular exponential family satisfies

$$a'(\lambda) = \eta'(\lambda)\lambda, \qquad a''(\lambda) = \eta''(\lambda)\lambda + \eta'(\lambda), \qquad a''(\lambda) = \{\eta'(\lambda)\}^2 \mathrm{Var}(X) + \eta''(\lambda)\lambda.$$

These well-known equalities are respectively obtained from

$$0 = \frac{\partial}{\partial \lambda} \int f_\lambda(x) d\mu(x) = \int h(x) e^{\eta(\lambda)x - a(\lambda)} \{x\eta'(\lambda) - a'(\lambda)\} d\mu(x) = \eta'(\lambda)\lambda - a'(\lambda),$$

the derivative of the previous equality, and

$$0=\frac{\partial^2}{\partial\lambda^2}\int f_\lambda(x)d\mu(x)=\int h(x)e^{\eta(\lambda)x-a(\lambda)}\left\{x\eta'(\lambda)-a'(\lambda)\right\}^2d\mu(x)+\eta''(\lambda)\lambda-a''(\lambda).$$

It follows that

$$\eta'(\lambda)=\left\{\eta'(\lambda)\right\}^2\mathrm{Var}(X)=\frac{1}{\mathrm{Var}(X)}.$$

Note that the conditional log-likelihood of X_t given $\mathcal{F}_{t-1}$ is

$$\ell_t(\theta)=\log f_{\lambda_t(\theta)}(X_t)=\eta\left\{\lambda_t(\theta)\right\}X_t-a\left\{\lambda_t(\theta)\right\}.$$

Under the assumed regularity conditions, as $n\to\infty$, the MLE of θ_0 satisfies

$$\sqrt{n}\left(\widehat{\theta}_n-\theta_0\right)\xrightarrow{d}\mathcal{N}\left(0,I^{-1}\right),$$

with

$$I=E\frac{\partial\ell_t(\theta_0)}{\partial\theta}\frac{\partial\ell_t(\theta_0)}{\partial\theta'}=E\left\{\eta'(\lambda_t)\right\}^2(X_t-\lambda_t)^2\frac{\partial\lambda_t}{\partial\theta}\frac{\partial\lambda_t}{\partial\theta'}=E\frac{1}{\mathrm{Var}(X_t)}\frac{\partial\lambda_t}{\partial\theta}\frac{\partial\lambda_t}{\partial\theta'}.$$

We conclude by noting that I is the inverse of the asymptotic variance of the two-stage WLSE, as defined in **A6***. □

Proof of Corollary 3.3. By the Cauchy root test, when $E\log(\alpha_0z_1+\beta_0)<0$, the ACD equation admits the stationary and ergodic solution

$$X_t=\lambda_tz_t,\quad \lambda_t=\omega_0\left\{1+\sum_{i=1}^{\infty}\prod_{j=1}^{i}(\alpha_0z_{t-j}+\beta_0)\right\}.$$

It follows that **A1** is satisfied. Moreover the condition $E\log(\alpha_0z_1+\beta_0)<0$ entails $E|X_t|^s<\infty$ for some $s>0$ (see *e.g.* Corollary 2.3 in Francq and Zakoian, 2019). We also have $a_t\le K\rho^t$ because $\sup_{\theta=(\omega,\alpha,\beta)\in\Theta}\beta<1$. Assumptions **A2** and **A4** follow. Because $\alpha_0>0$ and the law of z_t is not degenerated, the identfiability assumption **A3** holds true. Since $\upsilon_t=\lambda_t^2$, and $EX_t^2<\infty$ when $(\alpha_0+\beta_0)^2+\alpha_0^2<1$ (see *e.g.* Example 2.3 in Francq and Zakoian (2019), arguing that an ACD is the square of a GARCH), **A5** is satisfied under the conditions on (w_t), and the consistency results hold. Under the assumption (3.3) entailing $EX_t^4<\infty$, Section 2.2 shows that **A6-A8** are also satisfied. The conclusion follows. □

Proof of Corollary 3.4. It is shown in Aknouche and Francq (2019, Lemma 2.1) that the family $\{\mathrm{NB}(p(1-p)^{-1}\lambda,p),\lambda>0\}$ satisfies (2.11). Therefore, when λ_t follows the

INGARCH equation (1.2), Assumption **A1** is satisfied under (2.12). By the arguments given in Section 3.1, Assumptions **A2** and **A3** are satisfied under (2.13)–(2.14). In Aknouche and Francq (2019, Example 3.2) it is also shown that, in the first order case $p = q = 1$, X_t admits moments of any order under (2.12). Therefore, in view of Section 2.2, **A4**–**A7** hold. The CAN follows from Theorem 2.1. For the two stage estimator, note that $\upsilon_t = \lambda_t \left(1 + \varsigma_0^{-1}\right)$ is proportional to λ_t. Therefore the weighting sequence $\widehat{w}_{t,n} = \widetilde{\lambda}_t \left(\widehat{\theta}_{1WLS}\right)$ is asymptotically optimal. The CAN of $\widehat{\theta}_{2WLS}^{(P)}$ is obtained without additional constraint. □

Proof of Corollary 3.5. The proof of the CAN of the WLSE is similar to those of the previous corollaries, using the fact that $EX_t^2 < \infty$ if and only if (3.10) and $EX_t^4 < \infty$ if and only if (3.11) (see Ahmad and Francq (2016) and the references therein). □

Proof of Corollary 3.6. The proof uses already given arguments, after showing that (3.19) and $E\epsilon_t^r < \infty$ imply $EX_t^r < \infty$, for all $r > 0$ (we did not find a reference for this technical result, which should already be known, but a proof is available from the authors). For the consistency of $\widehat{\sigma}^2$, note that the strong convergence of $\widehat{\theta}_{1WLS}$ to θ_0 shows that

$$\widehat{\sigma}^2 = \sigma_0^2 + \frac{1}{n}\sum_{t=1}^{n} u_t + o(1) = \sigma_0^2 + o(1)$$

a.s. since

$$\begin{aligned} u_t &= \left(X_t - \omega_0 - \sum_{i=1}^{p} \alpha_{0i} X_{t-i}\right)^2 - \sigma_0^2 - \sum_{i=1}^{p} \alpha_{0i}\left(1 - \alpha_{0i}\right) X_{t-i} \\ &= \left(X_t - E\left(X_t \mid \mathcal{F}_{t-1}\right)\right)^2 - \operatorname{Var}\left(X_t \mid \mathcal{F}_{t-1}\right) \end{aligned}$$

is a martingale difference with finite second moment (see Csörgö, 1968). □

Proof of Corollary 3.7. Given $\mathcal{F}_{t-2}$, let the constants $a = \alpha(\nu+1)\lambda_{t-1}$, $a_0 = \alpha_0(\nu_0+1)\lambda_{t-1}$, $b = \nu - 2$, $b_0 = \nu_0 - 2$ and $c = \omega - \omega_0 + \beta\lambda_{t-1}(\theta) - \beta_0\lambda_{t-1}$. Given $\mathcal{F}_{t-2}$, the equation $\lambda_t = \lambda_t(\theta)$ a.s. is equivalent to

$$\frac{a_0 z}{b_0 + z} - \frac{az}{b + z} = c$$

for almost all z belonging to the support S of the distribution of z_{t-1}. This is equivalent to

$$z^2(a - a_0 + c) + z(ab_0 - a_0 b + c(b_0 + b)) + cbb_0 = 0.$$

Since S contains at least 3 points, the 3 coefficients of that second order polynomial are equal to 0. We thus have $c = 0$, $a_0 = a$ and $b = b_0$ (since $a_0 \neq 0$), that is

$$\beta\lambda_{t-1}(\theta) - \beta_0\lambda_{t-1} = \omega_0 - \omega, \qquad \alpha_0(\nu_0 + 1)\lambda_{t-1} = \alpha(\nu + 1)\lambda_{t-1}(\theta), \qquad \nu = \nu_0,$$

which entails $\theta = \theta_0$ when $\lambda_t = \lambda_t(\theta)$ a.s. If follows that the identifiability condition **A3** holds true. It can be shown that, under the strict stationarity condition $E \log a(z_1) < 0$, there exists $s > 0$ such that $EX_t^s < \infty$ and $E\lambda_t^s < \infty$ (see the proof of Corollary 3.3). Noting that

$$\lambda_t(\theta) \leq \omega + \alpha\frac{\nu + 1}{\nu - 2}X_{t-1} + \beta\lambda_{t-1}(\theta) \leq \sum_{i=0}^{\infty} \beta^i \left(\omega + +\alpha\frac{\nu + 1}{\nu - 2}X_{t-i-1}\right),$$

and Θ is compact, we also have $E \sup_{\theta\in\Theta} \lambda_t^s(\theta) < \infty$. This entails that **A2** and **A4** hold true. The consistency of the WLSE follows. The rest of the proof is shown by already given arguments, noting that, for $r \geq 1$, $EX_t^r < \infty$ when $\|a(z_1)\|_r < 1$. □

Proof of Proposition 4.1 Note that it suffices to show that

$$\text{MSE}_n(\widehat{w}_{t,n}) - \text{MSE}(w_t) \to 0 \qquad \text{and} \qquad \text{QLIK}_n(\widehat{w}_{t,n}) - \text{QLIK}(w_t) \to 0 \text{ a.s.} \tag{6.7}$$

Indeed, since the assumptions on the two sequences of weights are the same, the convergences (6.7) hold when $\widehat{w}_{t,n}$ and w_t are replaced by $\widehat{w}^*_{t,n}$ and w^*_t, and the conclusion follows. Since

$$\begin{aligned}
\text{MSE}_n(\widehat{w}_{t,n}) =& \frac{1}{n}\sum_{t=1}^{n}(X_t - \widehat{\lambda}_t)^4 - \frac{(\frac{1}{n}\sum_{t=1}^{n}(X_t - \widehat{\lambda}_t)^2\widehat{w}_{t,n})^2}{\frac{1}{n}\sum_{t=1}^{n}\widehat{w}^2_{t,n}}, \\
\text{MSE}(w_t) =& E(X_t - \lambda_t)^4 - \frac{(E(X_t - \lambda_t)^2 w_t)^2}{Ew_t^2},
\end{aligned}$$

the first convergence in (6.7) is obtained by showing

$$\frac{1}{n}\sum_{t=1}^{n} X_t^{4-i}\widehat{\lambda}_t^i \to EX_t^{4-i}\lambda_t^i, \quad \frac{1}{n}\sum_{t=1}^{n} X_t^{2-j}\widehat{\lambda}_t^j\widehat{w}_{t,n} \to EX_t^{2-j}\lambda_t^j w_t, \quad \frac{1}{n}\sum_{t=1}^{n}\widehat{w}^2_{t,n} \to Ew_t^2 \tag{6.8}$$

for $i = 0, 1, \ldots, 4$ and $j = 0, 1, 2$. Let us show the second convergence for $j = 1$. First consider the initial values. By the first inequalities of (2.5) and (2.15), and the consistency of $\widehat{\xi}_n$, for n large enough we have

$$\left|X_t\widehat{\lambda}_t\widehat{w}_{t,n} - X_t\lambda_t(\widehat{\theta})\upsilon_t^*(\widehat{\xi}_n)\right| \leq K\rho^t u_{1,t}, \quad u_{1,t} = |X_t|\left(\sup_{\xi\in V(\xi_0^*)}\upsilon_t^*(\xi) + \sup_{\theta\in\Theta}\lambda_t(\theta) + 1\right).$$

Now Taylor expansions yield

$$\left|X_t\lambda_t(\widehat{\theta})\upsilon_t^*(\widehat{\xi}_n) - X_t\lambda_t w_t\right| \leq u_{2,t}\left\|\widehat{\theta}-\theta_0\right\| + u_{3,t}\left\|\widehat{\xi}_n - \xi_0^*\right\|.$$

with

$$u_{2,t} = |X_t| \sup_{\xi\in V(\xi_0^*)} \upsilon_t^*(\xi) \sup_{\theta\in\Theta}\left\|\frac{\partial\lambda_t(\theta)}{\partial\theta}\right\|, \qquad u_{3,t} = |X_t||\lambda_t| \sup_{\xi\in V(\xi_0^*)}\left\|\frac{\partial\upsilon_t^*(\xi)}{\partial\xi}\right\|.$$

Since, for $i=1,2,3$, the processes $(u_{i,t})_t$ are stationary and ergodic processes with finite first order moments, we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^n X_t\widehat{\lambda}_t\widehat{w}_{t,n} = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^n X_t\lambda_t w_t = EX_t\lambda_t w_t \quad \text{a.s.}$$

The other convergences in (6.8) are shown by the same arguments, and the first result in (6.7) follows.

For the second result, noting that

$$\begin{aligned}\mathrm{QLIK}_n(\widehat{w}_{t,n}) &= 1 + \log\frac{1}{n}\sum_{t=1}^n\frac{(X_t-\widehat{\lambda}_t)^2}{\widehat{w}_{t,n}} + \frac{1}{n}\sum_{t=1}^n\log\widehat{w}_{t,n},\\ \mathrm{QLIK}(w_t) &= 1 + \log E\frac{(X_t-\lambda_t)^2}{w_t} + E\log w_t = 1 + \log E\frac{\upsilon_t}{w_t} + E\log w_t,\end{aligned}$$

we have to show that

$$\frac{1}{n}\sum_{t=1}^n\frac{X_t^{2-j}\widehat{\lambda}_t^j}{\widehat{w}_{t,n}} \to E\frac{X_t^{2-j}\lambda_t^j}{w_t} \quad \text{and} \quad \frac{1}{n}\sum_{t=1}^n\log\widehat{w}_{t,n}\to E\log w_t \tag{6.9}$$

for $j=0,1,2$. Let us detail the proof of the first convergence for $j=1$, that is

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^n\frac{X_t\widehat{\lambda}_t}{\widehat{w}_{t,n}} = E\frac{X_t\lambda_t}{w_t} = E\frac{\lambda_t^2}{w_t}. \tag{6.10}$$

The initial values are treated as previously. We thus show (6.10) by the ergodic theorem, noting that Taylor expansions entail that for n large enough

$$\frac{1}{n}\sum_{t=1}^n\left|\frac{X_t\lambda_t(\widehat{\theta})}{\upsilon_t^*(\widehat{\xi}_n)} - \frac{X_t\lambda_t}{w_t}\right| \leq \left\|\widehat{\theta}-\theta_0\right\|\frac{1}{n}\sum_{t=1}^n u_{4,t} + \left\|\widehat{\xi}_n-\xi_0^*\right\|\frac{1}{n}\sum_{t=1}^n u_{5,t},$$

where

$$u_{4,t} = \frac{|X_t|\sup_{\theta\in V(\theta_0)}\left\|\frac{\partial\lambda_t(\theta)}{\partial\theta}\right\|}{w_t}, \qquad u_{5,t} = \frac{|X_t|\sup_{\theta\in V(\theta_0)}|\lambda_t(\theta)|\sup_{\xi\in V(\xi_0^*)}\left\|\frac{\partial\upsilon_t^*(\xi)}{\partial\xi}\right\|}{w_t^2}\sup_{\xi\in V(\xi_0^*)}\frac{w_t}{\upsilon_t^*(\xi)},$$

admit finite expectations in view of (4.4) and (4.5). The first convergence in (6.9) for $j = 0$ is shown similarly, using (4.3). The convergence for $j = 2$ is shown by using (4.6) and(4.7). The last convergence in (6.9) also comes by doing a Taylor expansion, noting that

$$E \sup_{\xi \in V(\xi_0^*)} \left\| \frac{1}{v_t^*(\xi)} \frac{\partial v_t^*(\xi)}{\partial \xi} \right\| < \infty$$

is entailed by (4.6). The proof is complete. □

7 Conclusion

We proposed a class of WLS estimators for the conditional mean of a time series, which do not require the whole knowledge of the cdf of the observations. The asymptotic and finite sample properties of these estimators have been studied. Compared to the QMLEs, the WLSE presents the advantages of: 1) being of higher efficiency in some situations; 2) be asymptotically efficient when the cdf belongs to the linear exponential family; 3) have a standard asymptotic normal distribution even when one or several coefficients of the conditional mean are equal to zero; 4) be explicit and do not require any optimisation routine in INARCH models. We applied our general results to standard count and duration models. We studied selection methods of the optimal WLSE based on the MSE and QLIK loss functions, and demonstrated the theoretical and empirical superiority of the QLIK-based approach.

References

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

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

[3] Aknouche, A. and Francq, C. (2019). Count and duration time series with equal conditional stochastic and mean orders. *MPRA Working paper No. 90838*.

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

[5] Alzaid, A.A. and Al-Osh, M.A. (1990). An integer-valued pth-order autoregressive structure (INAR(p)) process. *Journal of Applied Probability*, **27**, 314-324.

[6] Barndorff-Nielsen, O.E., Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society*, **64**, 253-280.

[7] Bernardi, M., and Catania, L. (2014). The model confidence set package for R. *arXiv preprint arXiv:1410.8504v1*.

[8] Blasques, F., Gorgi, P., Koopman, S.J., and Wintenberger, O. (2018). Feasible invertibility conditions and maximum likelihood estimation for observation-driven models. *Electronic Journal of Statistics*, **12**, 1019-1052.

[9] Blasques, F., Koopman, S.J. and Lucas, A. (2015). Information-theoretic optimality of observation-driven time series models for continuous responses. *Biometrika*, **102**, 325-343.

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

[11] Cameron, A.C. and Trivedi P.K. (1986). Econometric models based on count data: comparisons and applications of some estimators and tests. *Journal of Applied Econometrics*, **1**, 29-53.

[12] Cameron, C. and Trivedi, P. (1998). *Regression analysis of count data*. Cambridge University Press, New York.

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

[14] Cox, D.R. (1981). Statistical analysis of time series: Some recent developments. *Scandinavian Journal of Statistics*, **8**, 93-115.

[15] Creal, D., Koopman, S.J. and Lucas, A. (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business and Economic Statistics*, **29**, 552-563.

[16] Creal, D., Koopman, S.J. and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, **28**, 777-795.

[17] Csörgö, M. (1968). On the strong law of large numbers and the central limit theorem for martingales. *Transactions of the American Mathematical Society*, **131**, 259-275.

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

[19] Du, J.-G. and Li, Y. (1991). The integer-valued autoregressive (INAR(p)) model. *Journal of Time Series Analysis*, **12**, 130-142.

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

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

[22] Francq, C. and J-M. Zakoian (2019). *GARCH models: structure, statistical inference and financial applications.* Chichester: John Wiley, second edition.

[23] Hansen, P. R., Lunde, A., and Nason, J. M. (2011). The model confidence set. *Econometrica*, 79, 453-497.

[24] Harvey, A.C. (2013). *Dynamic models for Volatility and Heavy Tails*. Econometric Society Monograph. Cambridge University Press.

[25] Harvey, A.C. and Chakravarty, T. (2008). Beta-t-(E)GARCH. University of Cambridge, Faculty of Economics.

[26] Heinen, A. (2003). Modelling time series count data: an autoregressive conditional poisson model. *CORE Discussion Paper 2003/62, Université Catholique de Louvain.*

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

[28] Morris, C. (1982). Natural Exponential Families with Quadratic Variance Functions. *The Annals of Statistics*, **10**, 65-80.

[29] Patton, A.J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, **160**, 246-256.

[30] Steutel, F.W. and Van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability*, **5**, 893-899.

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

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

[33] Zheng, H., Basawa, I.V., and Datta, S. (2006). Inference for pth-order random coefficient integer valued autoregressive processes. *Journal of Time Series Analysis*, **27**, 411-440.