



Asymptotic Properties of Conditional Least-squares Estimators for Array Time Series

Rajae Azral
Mohammed V University of Rabat

Guy Mélard
ECARES, Université libre de Bruxelles

April 2020

ECARES working paper 2020-12

ASYMPTOTIC PROPERTIES OF CONDITIONAL LEAST-SQUARES ESTIMATORS FOR ARRAY TIME SERIES

BY RAJAE AZRAK^{*}, and GUY MÉLARD[†]

Abstract The paper provides a kind of Klimko-Nelson theorems alternative in the case of conditional least-squares and M-estimators for array time series, when the assumptions of almost sure convergence cannot be established. We do not assume stationarity nor even local stationarity. In addition, we provide sufficient conditions for two of the assumptions and two theorems for the evaluation of the information matrix in array time series. In addition to time-dependent models, illustrations to a threshold model and to a count data model are given.

On behalf of all authors, the corresponding author states that there is no conflict of interest.

1. Introduction. Klimko and Nelson (1978) theorems, simply denoted KN theorems in the sequel, offer a powerful tool for proving almost sure convergence and asymptotic normality in a large number of situations in the context of dependent observations. Their usefulness is attested by their inclusion in well known books like Hall and Heyde (1980), Tong (1990), Taniguchi and Kakizawa (2000), and a large number of citations in papers, starting with Nelson (1980) and Tong and Lim (1980). Papers like those of Tjøstheim (1986, 1994) have had a huge influence to show the usefulness of KN theorems for deriving the asymptotic properties of estimators in many non linear time series models. But they were also used for stochastic regression models (Lai, 1994), functional time series models (Chen and Tsay, 1993), smooth transition autoregressive models (Teräsvirta, 1994), estimation of linear representation of nonlinear processes (Francq and Zakoïan, 1998), etc.

Of course KN theorems lay on assumptions that need to be checked but this is often much easier than proving consistency and asymptotic normality

^{*}Mohammed V University of Rabat, Faculty of Law, Economics and Social Sciences, Rabat, Morocco, (e-mail: rajae.azrak@gmail.com).

[†]Université libre de Bruxelles, SBS-EM, ECARES, avenue Franklin Roosevelt 50 CP 114/04, B-1050 Brussels, Belgium (e-mail: gmelard@ulb.ac.be).

MSC 2010 subject classifications: Primary 62M10, 60K35; secondary 60G12

Keywords and phrases: Klimko-Nelson theorems, non-stationary process, multivariate time series, time-varying models, information matrix

from scratch. There are however cases where it is not possible to prove the almost sure convergence conditions in some of their assumptions. In that case it is tempting to consider a weak version of KN theorems where almost sure convergence is replaced, both in the hypotheses and in the thesis, by convergence in probability. That was done in Theorem 1' of Azrak and M elard (2006) who have faced a case where almost sure convergence could not be obtained for estimators in ARMA models with time-dependent coefficients that depend also on the length of the series, n . In that context, the series cannot be considered as generated by a stochastic process but well by a triangular array process, see e.g. Alj et al. (2014).

Since weak KN theorems have not been established until now, we have preferred to propose a light alternative, Theorem 2.1 below, where convergence in probability of a general M-estimator (not necessarily a conditional least-squares estimator) is obtained, provided two of the hypotheses are valid in probability instead of almost surely. Then asymptotic normality is deduced in Theorem 2.2. In the context of conditional least-squares estimators, we have added Theorem 2.3, a sufficient condition for one of our assumptions, and Theorem 2.4 to check convergence in probability for averages in the context of array processes. Also, for any estimation problem, we propose Theorem 2.5 in order to prove the existence of a limit in order to compute the Fisher information matrix. Together with the previous results, this provides better foundations for the results sketched in Azrak and M elard (2006). This is not enough, however, to evaluate the Fisher information matrix in other examples. For that reason, we provide Theorem 2.6 where the limit of an average over time is replaced, in some cases, by the computation of an integral. The notations and the theorems are given in Section 2. Several applications are discussed in Section 3, not only for time-dependent ARMA models but also a threshold model and a model for count data. In the last two cases, our approach avoids the need to obtain a stationary and ergodic solution to the model equation. The proofs are collected in Section 4.

2. Theorems. All our random variables are supposed to belong to a probability space (Ω, F, P_θ) whose distribution depends on parameters stored in a vector $\theta \in \Theta$, the interior of a compact subset of \mathbb{R}^m . Let $E_\theta(\cdot)$ and $E_\theta(\cdot|\cdot)$ denote expectation and conditional expectation under P_θ . The true value of θ is denoted by θ^0 . As usual we suppose that θ^0 is an interior point of Θ . All probabilities, almost sure (a.s.) assertions, unqualified expectations and conditional expectations will be taken with respect to P_{θ^0} , although we will sometimes indicate θ^0 . Limit in probability is denoted by plim .

For each $n \in \mathbb{N}$, let $x^{(n)} = \{x_t^{(n)}; t = 1, \dots, n\}$ be a r -dimensional vector

stochastic process defined on (Ω, F, P_θ) , $r \geq 1$. This is called an array or a triangular array process. In some lemmas we will deal with scalar array processes i.e. the case where $r = 1$.

Let $Q^{(n)}(\theta) = Q^{(n)}(\theta, x^{(n)})$ be a general penalty function which depends on the observations, to be minimized with respect to the parameters θ and supposed to be at least two times differentiable with respect to θ in a neighbourhood of θ^0 . Let us consider the limited Taylor expansion of $Q^{(n)}(\theta)$ around θ^0

$$\begin{aligned} Q^{(n)}(\theta) &= Q^{(n)}(\theta^0) + \sum_{j=1}^m (\theta_j - \theta_j^0) \frac{\partial Q^{(n)}(\theta)}{\partial \theta_j} \Big|_{\theta=\theta^0} \\ &+ \frac{1}{2} \sum_{j=1}^m \sum_{k=1}^m (\theta_j - \theta_j^0)(\theta_k - \theta_k^0) \frac{\partial^2 Q^{(n)}(\theta)}{\partial \theta_j \partial \theta_k} \Big|_{\theta=\theta'_{jk}}, \end{aligned}$$

where θ'_{jk} are intermediate points on the straight line between θ and θ^0 . The second derivatives in the last term compose a matrix that will be denoted $\partial^2 Q^{(n)}(\theta) / \partial \theta \partial \theta^T$, where T denotes transposition. Let $\hat{\theta}^{(n)} = \arg \min_{\theta \in \Theta} Q^{(n)}(\theta)$ be a sequence of estimators.

Let us first recall KN theorems, see also Hall and Heyde (1980, pp. 174-176), in a single theorem, like Taniguchi and Kakizawa (2000, pp. 97-98). We follow the latter reference and Tjøstheim (1986) by providing a general penalty version of the theorems.

Theorem 1. Assume that

- i $\lim_{n \rightarrow \infty} \frac{1}{n} \frac{\partial Q^{(n)}(\theta)}{\partial \theta^T} \Big|_{\theta=\theta^0} = 0$ a.s. ;
- ii $\lim_{n \rightarrow \infty} \frac{1}{n} \frac{\partial^2 Q^{(n)}(\theta)}{\partial \theta \partial \theta^T} \Big|_{\theta=\theta^0} =_{def} V$ a.s., where V is a strictly positive definite matrix ;
- iii $\limsup_{n \rightarrow \infty} \frac{1}{n\Delta} \left(\left| \frac{\partial^2 Q^{(n)}(\theta)}{\partial \theta_j \partial \theta_k} \Big|_{\theta=\theta'_{jk}} - \frac{\partial^2 Q^{(n)}(\theta)}{\partial \theta_j \partial \theta_k} \Big|_{\theta=\theta^0} \right| \right) < \infty$ a.s., where θ'_{jk} is such that $|\theta'_{jk} - \theta^0| < \Delta$, $\Delta > 0$, $j, k = 1, \dots, m$;
- iv $\frac{1}{\sqrt{n}} \frac{\partial Q^{(n)}(\theta)}{\partial \theta^T} \Big|_{\theta=\theta^0} \xrightarrow{L} N(0, W)$ when $n \rightarrow \infty$, where \xrightarrow{L} indicates convergence in law, and W is a positive definite matrix.

Then, when $n \rightarrow \infty$, $\hat{\theta}^{(n)} \rightarrow \theta^0$ a.s., and $n^{1/2}(\hat{\theta}^{(n)} - \theta^0) \xrightarrow{L} N(0, V^{-1}WV^{-1})$.

The usual and original KN approach is to consider a conditional least squares criterion, that can be expressed in an array setting as follows, with an additional weighting. Of course, we do not retain stationarity and ergodicity

assumed by Taniguchi & Kakizawa (2000). We suppose there exists a filtration, more precisely a triangular array of increasing σ -fields $\{F_t^{(n)}; t \in \mathbb{N}_0\}$ with $F_0^{(n)} = \{\emptyset, \Omega\}$ and $F_0^{(n)} \subseteq F_1^{(n)} \subseteq \dots \subseteq F_t^{(n)}$, where $t = 1, \dots, n$, $n \in \mathbb{N}$, and, for each $n \in \mathbb{N}$, $F_t^{(n)}$ is generated by an arbitrary subset of $\{x_1^{(n)}, x_2^{(n)}, \dots, x_t^{(n)}\}$. The interpretation is that, for each n , $x_t^{(n)}$ is measurable with respect to $F_t^{(n)}$ for all t . Given a set of observations $\{x_t^{(n)}; t = 1, \dots, n\}$, we estimate θ by minimizing the conditional M-estimation criterion (see Martin, 1984, Bai and Wu, 1997, generalized here in an array context) $Q^{(n)}(\theta) = \sum_{t=1}^n \alpha_t^{(n)}(\theta)$ with respect to θ , satisfying the following:

Assumption on $\alpha_t^{(n)}(\theta)$

$\alpha_t^{(n)}(\theta)$ is two times continuously differentiable with respect to θ , in some neighborhood of θ^0 and such that $\{\partial \alpha_t^{(n)}(\theta)/\partial \theta, F_t^{(n)}; t = 1, \dots, n, n \in \mathbb{N}\}$, where $\partial \alpha_t^{(n)}(\theta)/\partial \theta$ is a column vector, and

$$(2.1) \quad \left\{ \left(\frac{\partial^2 \alpha_t^{(n)}(\theta)}{\partial \theta \partial \theta^T} - E \left(\frac{\partial^2 \alpha_t^{(n)}(\theta)}{\partial \theta \partial \theta^T} \middle| F_{t-1}^{(n)} \right) \right), F_t^{(n)}; t = 1, \dots, n, n \in \mathbb{N} \right\},$$

where $\partial^2 \alpha_t^{(n)}(\theta)/\partial \theta \partial \theta^T$ is a square matrix, are both martingale difference arrays.

For example, assume existence of $g_t^{(n)}(\theta, F_{t-1}^{(n)}) = E_\theta(x^{(n)} | F_{t-1}^{(n)})$, $t = 1, \dots, n$, and that it is a.s. two times continuously differentiable with respect to θ , in some neighborhood of θ^0 , and a function ρ satisfying appropriate conditions, then $\alpha_t^{(n)}(\theta) = \rho(x_t^{(n)} - g_t^{(n)}(\theta, F_{t-1}^{(n)}))$ which corresponds to M-estimation. In particular, the assumptions on $\alpha_t^{(n)}(\theta)$ are satisfied in the case of conditional weighted least-squares estimation, where

$$(2.2) \quad \alpha_t^{(n)}(\theta) = \{x_t^{(n)} - g_t^{(n)}(\theta, F_{t-1}^{(n)})\}^T \Sigma_t^{(n)-1}(\theta) \{x_t^{(n)} - g_t^{(n)}(\theta, F_{t-1}^{(n)})\},$$

where $\Sigma_t^{(n)}(\theta)$ is a $r \times r$ symmetric strictly positive definite scale matrix supposed to be two times continuously differentiable with respect to θ in a suitable neighborhood of θ^0 . The assumption on $g_t^{(n)}(\theta, F_{t-1}^{(n)})$ implies that $\alpha_t^{(n)}(\theta)$ is two times continuously differentiable with respect to θ , in some neighborhood of θ^0 , and hence also $Q^{(n)}(\theta)$. By definition of $\alpha_t^{(n)}(\theta)$, we have $E_{\theta^0}(\partial \alpha_t^{(n)}(\theta)/\partial \theta | F_{t-1}^{(n)}) = 0$, for all $t = 1, \dots, n$ and for all $n \in \mathbb{N}$, which means that $\{\partial \alpha_t^{(n)}(\theta)/\partial \theta |_{\beta=\beta^0}, F_t^{(n)}; t = 1, 2, \dots, n, n \in \mathbb{N}\}$ is a martingale difference array. The same is true of (2.1).

Here are our main results. The proofs will be given in Section 4. In the univariate case, Azrak and Méléard (2006) show that the KN theorems lead

to an almost sure sequence-oriented version of the following two theorems, proved here in a multivariate, limit in probability and array-oriented context. Moreover they use exponent 4 instead of $2+\delta$ in Assumption $\mathbf{H}_{2.1}$ of Theorem 2.1, which is a significant improvement.

Theorem 2.1. We suppose that there exist two positive constants C_1 and C_2 such that for all $t = 1, \dots, n$, and uniformly in n :

$\mathbf{H}_{2.1}$ for $i = 1, \dots, m$, and with $\delta > 0$

$$E_{\theta^0} \left(\left| \frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta_i} \right|^{2+\delta} \right) \leq C_1;$$

$\mathbf{H}_{2.2}$ for $i, j = 1, \dots, m$

$$E_{\theta^0} \left(\left| \frac{\partial^2 \alpha_t^{(n)}(\theta)}{\partial \theta_i \partial \theta_j} - E_{\theta} \left(\frac{\partial^2 \alpha_t^{(n)}(\theta)}{\partial \theta_i \partial \theta_j} \middle| F_{t-1}^{(n)} \right) \right|^2 \right) \leq C_2.$$

Suppose further that

$\mathbf{H}_{2.3}$

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n E_{\theta^0} \left(\frac{\partial^2 \alpha_t^{(n)}(\theta)}{\partial \theta_i \partial \theta_j} \middle| F_{t-1}^{(n)} \right) = V_{ij} \text{ for } i, j = 1, \dots, m,$$

where $V = (V_{ij})_{1 \leq i, j \leq m}$ is a strictly positive definite matrix of constants, called the Fisher information matrix;

$\mathbf{H}_{2.4}$

$$\text{plimsup}_{n \rightarrow \infty} (n\Delta)^{-1} \left| \sum_{t=1}^n \left(\left\{ \frac{\partial^2 \alpha_t^{(n)}(\theta)}{\partial \theta_i \partial \theta_j} \right\}_{\theta=\theta_{ij}^*} - \left\{ \frac{\partial^2 \alpha_t^{(n)}(\theta)}{\partial \theta_i \partial \theta_j} \right\}_{\theta=\theta^0} \right) \right| < \infty$$

for $i, j = 1, \dots, m$, where θ_{ij}^* is a point of the straight line joining θ^0 to every θ , such that $\|\theta - \theta^0\| < \Delta$, $0 < \Delta$, where $\|\cdot\|$ is the Euclidean norm.

Then there exists a sequence of estimators $\widehat{\theta}^{(n)}$ such that $\text{plim} \widehat{\theta}^{(n)} = \theta^0$ when $n \rightarrow \infty$.

Theorem 2.2. If Assumptions $\mathbf{H}_{2.1} - \mathbf{H}_{2.4}$ of Theorem 2.1 are satisfied, as well as $\mathbf{H}_{2.5}$ and $\mathbf{H}_{2.6}$:

H_{2.5} for $i, j = 1, \dots, m$

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \left\{ E_{\theta^0} \left(\frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta_i} \frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta_j} \middle| F_{t-1} \right) - E_{\theta^0} \left(\frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta_i} \frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta_j} \right) \right\} = 0;$$

H_{2.6} there exists a positive definite matrix $W = (W_{ij})_{1 \leq i, j \leq m}$ defined by

$$W_{ij} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n E_{\theta^0} \left(\frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta_i} \frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta_j} \right).$$

Then

$$n^{1/2}(\hat{\theta}^{(n)} - \theta^0) \xrightarrow{L} \mathcal{N}(0, V^{-1}WV^{-1}) \text{ when } n \rightarrow \infty.$$

REMARK 2.1. *In the case of a simple stochastic process $\{x_t; t = 1, 2, \dots\}$, see Alj et al. (2017, Theorem 1), the superscripts $^{(n)}$ can be omitted everywhere except on $Q^{(n)}(\theta)$ and $\hat{\theta}^{(n)}$.*

For a direct application of Theorems 2.1 and 2.2, see e.g. Alj et al. (2020).

The following theorem provides a sufficient condition for Assumption **H_{2.1}**. It can be used to replace the stringent assumption of existence of moments of order 8 to a weaker assumption of existence of moments of order $4 + 2\delta$, $\delta > 0$. Theorem 2.3 can then be used to reduce Assumption $H_{2.4}(a)$ in Alj et al. (2017) to existence of moments of order $4 + 2\delta$ instead of 8, $\delta > 0$.

Theorem 2.3. Assume that $\alpha_t^{(n)}(\theta)$ has the form (2.2). Denote $e_t^{(n)} = x_t^{(n)} - g_t^{(n)}(\theta, F_{t-1}^{(n)})$ and $\|\cdot\|_F$, the Frobenius norm of a matrix. Suppose that for some $\delta > 0$, there exist some constants K_4 , m_2 , P_1 , and P_2 such that we have for all $t = 1, \dots, n$ and n , and $i = 1, \dots, m$,

$$\begin{aligned} \left\| \frac{\partial \Sigma_t^{(n)-1}(\theta)}{\partial \theta_i} \middle|_{\theta=\theta^0} \right\|_F^2 &\leq K_4, & \left\| \Sigma_t^{(n)-1}(\theta^0) \right\|_F^2 &\leq m_2, \\ E_{\theta^0} \left(\left| e_t^{(n)T}(\theta) e_t^{(n)}(\theta) \right|^{2+\delta} \right) &\leq P_1, & E_{\theta^0} \left(\left| \frac{\partial e_t^{(n)T}(\theta)}{\partial \theta_i} \frac{\partial e_t^{(n)}(\theta)}{\partial \theta_i} \right|^{1+\delta/2} \right) &\leq P_2, \end{aligned}$$

and that $e_t^{(n)}(\theta^0)$ and $\partial e_t^{(n)}(\theta)/\partial \theta_i|_{\theta=\theta^0}$ are independent. Then, Assumption **H_{2.1}** is satisfied for that δ , which means that there exists a positive constant

C_1 such that for all t and all n , and $i = 1, \dots, m$,

$$(2.3) \quad E_{\theta^0} \left(\left| \frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta_i} \right|^{2+\delta} \right) \leq C_1.$$

REMARK 2.2. *Theorem 2.3 can be seen as an alternative version of Lemma 4.11 in the Technical Appendix (denoted here TA) of Alj et al. (2017).*

REMARK 2.3. *Note that $\alpha_t^{(n)}(\theta)$ in Alj et al. (2017) has an additional term of the form $\log\{\det(\Sigma_t^{(n)}(\theta))\}$ but it is easy to see that an upper bound also exists by using Hölder inequality. Then TA Lemma 3.2 of Alj et al. (2017) stated there for integral p can be adapted, taking $p = 2 + \delta$, since its proof then results from Hölder inequality.*

COROLLARY 2.1. *If $\Sigma_t^{(n)}(\theta)$ does not depend on θ , then the third inequality of Theorem 2.3, can be replaced by $E_{\theta^0} \left(\left| e_t^{(n)T}(\theta) e_t^{(n)}(\theta) \right|^2 \right) \leq P_1$ while the thesis remains true.*

The following theorem provides an alternative to check convergence in probability for averages in the context of array processes.

Theorem 2.4. Let $\{x_t^{(n)}, t = 1, \dots, n\}$ be, for each $n \in N$, a scalar array process with second-order moments, and such that

$$E \left(\frac{1}{n} \sum_{t=1}^n x_t^{(n)} \right)^2 = O \left(\frac{1}{n} \right).$$

Then, $n^{-1} \sum_{t=1}^n x_t^{(n)}$ converges in probability to zero when n tends to infinity.

In previous papers, the following theorem is also used implicitly to prove existence of the information matrix V in $\mathbf{H}_{2.3}$ of our Theorem 2.1, for instance in their Examples 2, 3 and 4 of several autoregressive-moving average (ARMA) models with time-dependent coefficients. Contrarily to Theorem 2.3, Theorem 2.5 does not help in the examples in Alj et al. (2017). It can also be used to prove existence of the limit W in Assumption $\mathbf{H}_{2.6}$.

Theorem 2.5. Let $\{u_t^{(n)}, t = 1, \dots, n\}$ and $\{v_t^{(n)}, t = 1, \dots, n\}$ be two triangular arrays of real numbers such that $(1/n) \sum_{t=1}^n v_t^{(n)}$ absolutely converges and $\lim_{n \rightarrow \infty} (1/n) \sum_{t=1}^n v_t^{(n)} = L$, and that $\{u_t^{(n)}\}$ converges to $C > 0$ when $t \rightarrow \infty$, hence $n \rightarrow \infty$. Then $(1/n) \sum_{t=1}^n u_t^{(n)} v_t^{(n)}$ converges when $n \rightarrow \infty$ and its limit is CL .

Theorem 2.5 is not enough to evaluate limits of $(1/n) \sum_{t=1}^n v_t^{(n)}$ (or $(1/n) \sum_{t=1}^n u_t^{(n)} v_t^{(n)}$) in all cases. The following theorem can help sometimes. If needed, the integral can of course be evaluated numerically.

Theorem 2.6. Consider $(1/n) \sum_{t=1}^n v_t^{(n)}$. Assume that there exists a Riemann-integrable function $V(x)$ defined on $[0, 1]$ such that $V(t/n) = v_t^{(n)}$ for $t = 1, \dots, n$ and all $n > 0$. Then

$$(2.4) \quad \lim_{n \rightarrow \infty} (1/n) \sum_{t=1}^n v_t^{(n)} = \int_0^1 V(x) dx$$

3. Several applications. As already mentioned, the theorems of Section 2 can be used to provide better foundations for time dependent ARMA models in the array case. These models will be denoted $\text{tdARMA}^{(n)}$. This will be the subject of Section 3.1. All the results of Section 2 are also used in Alj et al. (2020) for vector $\text{tdARMA}^{(n)}$ models or $\text{tdVARMA}^{(n)}$ models, the array version of Alj et al. (2017). Nevertheless other applications are possible and will be sketched. In Section 3.2, we will consider a self-excited autoregressive threshold (SETAR in brief) models. In Section 3.3, we investigate an application to count data time series. That does not mean that any process can be the subject of the asymptotics Section 2. We will comment on this in Section 3.4.

3.1. Time dependent ARMA models. As intended, the theorems of Section 2 provide better foundations for time dependent ARMA (tdARMA) models in the array case, i.e. when at least one of the autoregressive or moving average coefficient, or the function involved in marginal heteroscedasticity, depends on the length of the series n . We denote $\text{tdARMA}^{(n)}(p, q)$ such a tdARMA model of orders p and q .

Theorem 2.1 and 2.2 replace Theorem 1' in Azrak and M elard (2006) (denoted AM2006 in this Section) advantageously with its slightly modified assumption $\mathbf{H}_{2.1}$. Their proofs is now independent of a conjectured weak version of Theorem 1 in Section 2. They are, moreover, valid in a wider framework, with a multivariate M-estimation setting, not necessarily

for univariate least-squares estimation. Except in Section 3.4, we focus on variants of autoregressive or ARMA models to simplify the exposition and consider a Gaussian quasi-likelihood estimation method with $\alpha_t(\theta)$ defined by (2.2).

Theorem 2.3 allows to reduce the moment condition on the ϵ_t 's from 8 to $4 + 2\delta$, $\delta > 0$, in AM2006. Indeed if the exponent $2 + \delta$ in (2.3) is replaced by 4, as originally, given that $\alpha_t^{(n)}$ is a quadratic function of the ϵ_t , it is the only place where existence of moments of order 8 is required.

Let us look at a special case in order to illustrate Section 2 for these models. We take the example of a univariate tdAR(1)⁽ⁿ⁾ model, i.e. when $p = 1$ and $q = 0$, defined by

$$(3.1) \quad x_t^{(n)} = \phi_t^{(n)}(\theta)x_{t-1}^{(n)} + g_t^{(n)}(\theta)\epsilon_t,$$

with independent ϵ_t 's with zero mean and finite variance σ^2 . For the moment, $g_t^{(n)}(\theta) > 0$ is assumed not to depend on the parameters $\theta = (\theta_1, \dots, \theta_m)^T$, for simplicity, and is denoted simply $g_t^{(n)}$. We take $\Sigma_t^{(n)} = g_t^{(n)2}$ in (2.2). We also assume that there exist $M_1 > 0$ and $M_2 < \infty$ such that $M_1 < g_t^{(n)} < M_2$. Then, this is a special case of Example 3 of AM2006, except that, by application of Theorem 2.3, the moment condition on the ϵ_t is reduced.

Although it is expressed slightly differently, the main conditions in Theorem 2' of AM2006 are about the moving average development of the first three derivatives of the residuals with respect to the parameters. The residuals are defined by $e_t^{(n)}(\theta) = x_t^{(n)} - \phi_t^{(n)}(\theta)x_{t-1}^{(n)}$, $t = 1, \dots, n$, like in Theorem 2.3. For example, for the first-order derivative, we have the following moving average representation $\partial e_t^{(n)}(\theta)/\partial \theta_i = \sum_{k=1}^{t-1} \psi_{tik}^{(n)}(\theta)g_{t-k}\epsilon_{t-k}$, $i = 1, 2$. On the basis of Example 1 of AM2006

$$(3.2) \quad \psi_{tik}^{(n)}(\theta) = \frac{\partial \phi_t^{(n)}(\theta)}{\partial \theta_i} \prod_{\ell=1}^{k-1} \phi_{t-\ell}^{(n)}(\theta).$$

Except the existence of V , it suffices that the coefficients $\psi_{tik}(\theta^0)$ be bounded by Φ^k , $k = 1, \dots, m$, for some $\Phi < 1$, and this for $i = 1, \dots, m$, $t = 1, \dots, n$. Consequently a sufficient condition is that $|\phi_t^{(n)}(\theta^0)| < \Phi < 1$, for all $t = 1, \dots, n$, and uniformly with respect to n .

There remains to show the existence of the matrices V and W , V being strictly positive definite and W being positive definite. We have to show existence of the Hessian matrix $V = \lim_{n \rightarrow \infty} (1/n) \sum_{t=1}^n V_t^{(n)}$, where the

elements of $V_t^{(n)}$ are obtained like in Alj et al. (2017, Equation A.21)

$$(3.3) \quad V_{t,ij}^{(n)} = \left\{ \frac{\partial \phi_t^{(n)}(\theta)}{\partial \theta_i} \frac{\partial \phi_t^{(n)}(\theta)}{\partial \theta_j} \right\}_{\theta=\theta^0} E(x_{t-1}^{(n)})^2,$$

$i, j = 1, \dots, m$. Provided $|\Phi| < 1$, it is shown by AM2006 (p. 298) that $u_t^{(n)} = E(x_{t-1}^{(n)})^2$ is a convergent series with a positive limit, say C . As shown, this is true in particular if $g_t^{(n)}$ is an exponential function like $\exp\{(t - (n + 1)/2)/(n - 1)\}$.

We will now illustrate Theorem 2.5. For each $i, j = 1, \dots, m$,

$$\frac{1}{n} \sum_{t=1}^n v_t^{(n)} = \frac{1}{n} \sum_{t=1}^n \left\{ \frac{\partial \phi_t^{(n)}(\theta)}{\partial \theta_i} \frac{\partial \phi_t^{(n)}(\theta)}{\partial \theta_j} \right\}_{\theta=\theta^0}$$

is absolutely convergent, and converges to a limit L_{ij} , then V does exist by application of Theorem 2.5, and its element V_{ij} is equal to CL_{ij} . As shown by AM2006 (p. 298), this is the case, in particular, if $\phi_t^{(n)}(\theta) = \theta_1 + \theta_2(t - (n + 1)/2)/(n - 1)$, at least when $\theta_1^0 = \phi^0$ and $\theta_2^0 = 0$.

REMARK 3.1. *Note that a basic result on Cesàro means (that if a sequence v_n converges to V , then the Cesàro means $V_n = (1/n) \sum_{i=1}^n v_i$ converges to V) can also be used in some cases, like for $i = j = 1$ or the term corresponding to $-(n + 1)/(2(n - 1))$ for $i = 1$ and $j = 2$.*

We will not discuss the existence of the outer product of gradient matrix W since this is done in Alj et al. (2020) in a more general set-up. Anyway, in this case, because no parameter is involved in $g_t^{(n)}$, we have that $W = V$, irrespective of the distribution of ϵ_t .

We can also illustrate Theorem 2.6 in a tdAR(1)⁽ⁿ⁾ model (3.1), except that we suppose now that $g_t^{(n)}(\theta)$ depends on parameters in θ . Remind that, to simplify the proofs, we have supposed in AM2006 (p. 288) that the parameters involved in $g_t^{(n)}(\theta)$ are functionally independent of those involved in the AR and MA coefficients.

Let $g_{t,\theta}^{(n)} = g_t^{(n)}(\theta) = \exp\{\theta(t - (n + 1)/2)/(n - 1)\}$ for $\theta \geq 0$, $t = 1, \dots, n$, see Examples 2, 3 and 4 of AM2006. Suppose that, for $\xi \geq 0$ and $\eta > 0$

$$(3.4) \quad v_t^{(n)} = \frac{g_{t,\xi}^{(n)}}{(1 + g_{t,\eta}^{(n)})^2}.$$

Note that $g_{1,\theta}^{(n)} = \exp(-\theta/2) \leq g_{t,\theta}^{(n)} \leq g_{n,\theta}^{(n)} = \exp(\theta/2)$. Hence

$$\frac{g_{1,\xi}^{(n)}}{(1 + g_{n,\eta}^{(n)})^2} \leq v_t^{(n)} \leq \frac{g_{n,\xi}^{(n)}}{(1 + g_{1,\eta}^{(n)})^2}.$$

We can even say that $g_{t,\theta}^{(n)} \leq 1$ for $t \leq (n+1)/2$ and $g_{t,\theta}^{(n)} \geq 1$ for $t \geq (n+1)/2$. Hence we have upper and lower bounds for $(1/n) \sum_{t=1}^n v_t^{(n)}$. Using a variation of Theorem 2.6 (where, for convenience $(t - (n+1)/2)/(n-1)$ is replaced by x), we have a simple evaluation of $\lim_{n \rightarrow \infty} (1/n) \sum_{t=1}^n v_t^{(n)}$ as

$$(3.5) \quad \int_{-0.5}^{0.5} \frac{e^{\xi x}}{(1 + e^{\eta x})^2} dx.$$

For $\xi > 0$, a primitive of the integrand in (3.5) is

$$(3.6) \quad \frac{e^{\xi x}(\xi - (\xi - \eta)(1 + e^{\eta x}))F(1, \frac{\xi}{\eta}, \frac{\xi + \eta}{\eta}; e^{-\eta x})}{\xi \eta (1 + e^{\eta x})} + C,$$

where $F(a, b, c; z)$ is the hypergeometric function, see Abramowitz and Stegun (1965, Chapter 15) or Erdélyi (1953), and C is a constant. If $\xi = 0$, a simpler primitive is found:

$$x + \frac{1}{\eta} \left\{ \frac{1}{1 + e^{\eta x}} - \log(1 + e^{\eta x}) \right\} + C.$$

Hence $\lim_{n \rightarrow \infty} (1/n) \sum_{t=1}^n v_t^{(n)}$ is equal to

$$(3.7) \quad \left[x + \frac{1}{\eta} \left\{ \frac{1}{1 + e^{\eta x}} - \log(1 + e^{\eta x}) \right\} \right]_{-1/2}^{1/2} \\ = 1 + \frac{1}{\eta} \left\{ \frac{1}{1 + e^{\eta/2}} - \frac{1}{1 + e^{-\eta/2}} - \log \left(\frac{1 + e^{\eta/2}}{1 + e^{-\eta/2}} \right) \right\}.$$

For example, if $n = 200$, $\xi = 0$ and $\eta = 1.3$, the average in (2.4) with terms (3.4) is 0.258527 whereas the integral in (3.7) yields 0.258445.

3.2. SETAR models. We consider a SETAR(2; 1) model here. More general results will be presented by Mélard and Niglio (2020). The model is a two regime model defined by the equation

$$(3.8) \quad x_t = \begin{cases} \theta_1 x_{t-1} + \epsilon_t, & \text{if } x_{t-1} \leq 0, \\ \theta_2 x_{t-1} + \epsilon_t, & \text{if } x_{t-1} > 0, \end{cases}$$

where θ_1 and θ_2 are parameters and the ϵ_t 's are independent zero mean random variables with a constant variance σ^2 . Introducing the indicator I_{t-1} such that $I_{t-1} = 1$, if $x_{t-1} \leq 0$ and $I_{t-1} = 0$, elsewhere, we can write (3.8) under the form

$$(3.9) \quad x_t = \{\theta_1 I_{t-1} + \theta_2(1 - I_{t-1})\}x_{t-1} + \epsilon_t,$$

which resembles the equation of a tdAR(1) process as in (3.1), but without (n) . Nevertheless, the specification was not considered before so that the assumptions of their Theorem 2 are to be checked.

Except the existence of V , the main conditions are about the coefficients $\psi_{tik}(\theta)$ of the moving average development of the derivatives of the residuals with respect to the two parameters. A sufficient condition is that the coefficients $\psi_{tik}(\theta^0)$ in absolute value be bounded by Φ^k for some $\Phi < 1$. Denote $\phi_t(\theta) = \theta_1 I_{t-1} + \theta_2(1 - I_{t-1})$, the residual $e_t(\theta) = x_t - \phi_t(\theta)x_{t-1}$, and the moving average representation of their derivatives $\partial e_t(\theta)/\partial \theta_i = \sum_{k=1}^{t-1} \psi_{tik}(\theta)\epsilon_{t-k}$, $i = 1, 2$.

Suppressing the superscript (n) in (3.2), and since $\partial \phi_t(\theta)/\partial \theta_i = I_{t-1}$, if $i = 1$, or $(1 - I_{t-1})$, if $i = 2$, and there remains the product $\prod_{\ell=1}^{k-1} \phi_{t-\ell}(\theta)$ in (3.2) is equal to $\theta_1^{N_{t,k-1}} \theta_2^{k-1-N_{t,k-1}}$, where $N_{t,k-1} = \sum_{\ell=1}^{k-1} I_{t-1-\ell}$ is the number of occurrences of the first regime between $t-1$ and $t-k+1$. Denote $\Phi(\theta) = \max(|\theta_1|, |\theta_2|)$, and $\Phi = \Phi(\theta^0)$. Then $|\psi_{tik}(\theta^0)| < \Phi^k$. Hence a sufficient but not necessary condition is that $\max(|\theta_1^0|, |\theta_2^0|) < 1$. Note that these conditions are more restrictive than those of Petruccelli and Woolford (1984) which are $\theta_1^0 < 1$, $\theta_2^0 < 1$, and $\theta_1^0 \theta_2^0 < 1$. According to Giordano et al. (2013), this is specific to the first-order autoregressive threshold model, and does even not extend to second-order models for which they provide conditions.

Note that, as presented, the coefficients of the model do not depend on n but there will be no problem to introduce time-dependency and even dependency on n to the autoregressive coefficients θ_1 and θ_2 in (3.8). This would be impossible in the classical approach of proving existence of a stationary and ergodic solution. That would also be true if the two regimes had different innovation standard deviation.

There remains that the existence of a strictly definite positive matrix V is dependent on the stochastic behavior according to (3.3) of the I_t 's. Here, by direct computation of the Hessian of the Gaussian log-likelihood, proportional to the sum of squared residuals, we have that $V_{12} = V_{21} = 0$. For the diagonal elements, at this time, their existence is an assumption. We can consider the alternative model of Amendola et al. (2007) where, defining $x_t^{(i)}$, $i = 1, 2$ as the two second hand sides of (3.8), we suppose $x_t = x_t^{(1)}$

if $x_{t-1} \leq 0$, and $x_t = x_t^{(2)}$, if $x_{t-1} > 0$. Using that the variance for the two regimes are, respectively, $\sigma^2/(1 - \theta_1^2)$ and $\sigma^2/(1 - \theta_2^2)$, we can then use Proposition 1 in Amendola et al. (2007) to obtain

$$(3.10) \quad V_{11} = V_{22} = \frac{1}{2} \left(\frac{1}{1 - \theta_1^2} + \frac{1}{1 - \theta_2^2} \right).$$

3.3. Count data models. Liboschik et al. (2017) have presented a generalized linear model approach for handling time series for count data. Remember that asymptotic results are only available for very specific count data models. Indeed, for the same reason as for SETAR models, the asymptotic theory relies on the existence of a stationary ergodic solution of the model equation at the true value of the model parameters. Here, we will just handle a particular model which is close to a model for which a theory is available, which is not true for more complex models.

Let $x_t \in \mathbb{N}$ be the observation. We suppose that the x_t 's are independent random variables, each with a Poisson distribution of parameter λ_t and that these λ_t are themselves random variables satisfying the equation

$$(3.11) \quad \log(\lambda_t + 1) = \theta_1 \log(\lambda_{t-1} + 1) + \epsilon_t + \theta_2 \epsilon_{t-1},$$

where θ_1 and θ_2 are parameters and the ϵ_t 's are independent zero mean random variables with a constant variance σ^2 . There is some similitude with Davis et al. (2003) GLARMA model for which a simple specification is

$$(3.12) \quad \log(\lambda_t) = \theta_1 \log(\lambda_{t-1}) + \theta_2 \epsilon_{t-1},$$

but where ϵ_t is related to the observation at time t by $\epsilon_t = (x_t - \lambda_t)/\lambda_t^\delta$ where $0 < \delta \leq 1$. The general log-linear model for count time series of Liboschik et al. (2017) is illustrated by a similar but different equation

$$(3.13) \quad \log(\lambda_t) = \theta_0 + \theta_1 \log(x_{t-1} + 1) + \theta_2 \log(\lambda_{t-1}).$$

A first difference of both (3.12) and (3.13) with (3.11) is that the logarithm of λ_t is used, not of $\lambda_t + 1$. The problem in the three formulations is that λ_t is not observable. In our case we use the observations x_{t-k} instead of λ_{t-k} , $k = 0, 1$ in (3.11), so that the parameters are estimated by using the following model:

$$(3.14) \quad \log(x_t + 1) = \theta_1 \log(x_{t-1} + 1) + \epsilon_t + \theta_2 \epsilon_{t-1}.$$

A second difference with respect to Davis et al. (2003) and Liboschik et al. (2017) is that a Gaussian quasi-likelihood is used here, instead of a

Poisson quasi-likelihood. Using that Gaussian quasi-likelihood, the objective likelihood appears as a univariate version of the sum of the α_t 's in (2.2). Using Mélard (2020a) who has derived the asymptotic properties of standard ARMA and VARMA models from Alj et al. (2017), we can deduce the asymptotic results for $\hat{\theta}_n$ provided that the true values of θ fulfil $|\theta_1^0| < 1$ and $|\theta_2^0| < 1$. Although the distribution of $\log(x_t + 1)$ is not normal, we do not need to use W because there is no marginal heteroscedasticity depending on a parameter. Contrarily to the case of the Poisson quasi-likelihood, this is done again, like in Section 3.2, without having to find a stationary solution and proving its ergodicity.

Again, like in Section 3.2, we are not in an array case but the coefficients θ_1 and θ_2 can be replaced by function of t , and possibly of n , depending on a new vector of parameters θ . If suitable conditions are satisfied, the asymptotic properties are obtained. Mélard (2020b) consider ARMA(p, q) models, seasonal models, and even regression. Monte Carlo results are given and examples taken from the literature are shown.

3.4. *Critical examination of other models.* The previous subsections can leave the impression that the assumptions in Section 2 are weak. This is not true. The models considered in Sections 3.2 and 3.3 are largely linear models. For true nonlinear models, the assumptions do not seem to be valid.

A simple example is the integer GARCH(1,1) (INGARCH) model which is a Poisson process like in Section 3.3 except that (3.14) is replaced by $\lambda_t = \lambda_t(\theta)$ given by

$$(3.15) \quad \log(\lambda_t) = \theta_0 + \theta_1 \log(\lambda_{t-1}) + \theta_2 x_{t-1},$$

where $\theta_0, \theta_1, \theta_2 > 0$. Then, using the Poisson quasi-likelihood instead of the Gaussian likelihood used in Section 3.3, $\alpha_t(\theta) = -\lambda_t(\theta) + x_t \log \lambda_t(\theta)$, hence $\partial \alpha_t(\theta) / \partial \theta = (x_t / \lambda_t(\theta) - 1) \partial \lambda_t(\theta) / \partial \theta$. In a more general context of weak dependent processes, Doukhan et al. (2012, Corollary 2.1) show the existence of a unique strictly stationary and ergodic process $\{(x_t, \lambda_t), t \in \mathbf{Z}\}$ such that the moments of order s exist for all integers $s \geq 1$, if $0 < \theta_1 + \theta_2 < 1$. An alternative proof is provided by Fokianos et al. (2020, Proposition 3.2) in a multivariate count autoregression framework. In the same paper, in the proof of Lemma 4.1, it is shown that the expectation of $|\partial \lambda_t(\theta) / \partial \theta_i|^4$ is bounded, $i = 1, 2, 3$, here. Because of stationarity, there is a constant C_1 such that $\mathbf{H}_{2,1}$ is satisfied. There remains, of course, to check the other assumptions. Note, however, that the checking of $\mathbf{H}_{2,1}$ is a consequence of the existence of a unique strictly stationary solution and does not seem to have been established without it. Therefore, Theorem 2.1 cannot be used in the case of the INGARCH model.

4. Proofs. Here are the proofs of the theorems in Section 2. For the proofs of Theorems 2.1 and 2.2, we need a certain number of propositions. We first need a weak law of large numbers for martingale arrays.

Proposition 4.1. Let $\{X_t^{(n)}, F_t^{(n)}; t = 1, \dots, k_n, n \in \mathbb{N}\}$ be a scalar martingale difference array, and $\{k_n\}$ an increasing integer sequence with $\{k_n\} \uparrow \infty$.

Let $\{c_t^{(n)}; t = 1, \dots, k_n, n \in \mathbb{N}\}$ a positive constant array. If, for $1 \leq \pi \leq 2$,

(i) $\{|X_t^{(n)}/c_t^{(n)}|^\pi\}$ is uniformly integrable with respect to t and n ,

(ii) $\limsup_{n \rightarrow \infty} \sum_{t=1}^{k_n} c_t^{(n)} < \infty$,

(iii) $\lim_{n \rightarrow \infty} \sum_{t=1}^{k_n} c_t^{(n)2} = 0$,

then $\sum_{t=1}^{k_n} X_t^{(n)} \xrightarrow{L_\pi} 0$, as $n \rightarrow \infty$, where L_π indicates convergence in L_π norm, hence $\text{plim} \sum_{t=1}^{k_n} X_t^{(n)} = 0$ as $n \rightarrow \infty$.

Proof. See Davidson (1994, Theorem 19.7 p. 299), and Davidson (1994, Corollary 19.9 p. 301). \square

Note that uniform integrability for a scalar array process $\{x_t^{(n)}; t = 1, \dots, k_n, n \in \mathbb{N}\}$ is a precise assumption, i.e. that $\lim_{M \rightarrow \infty} \sup_{t,n} E(|x_t^{(n)}| 1_{|x_t^{(n)}| > M}) = 0$. see e.g. Davidson (1994, p. 188). A sufficient condition is that $E\{|x_t^{(n)}|^{1+\delta}\} < \infty$, for all n and t , for some $\delta > 0$, see Davidson (1994, p. 190).

From now on, we use the notations of Section 2.

Proposition 4.2. Under the assumptions on $\alpha_t^{(n)}(\theta)$, plus Assumption $\mathbf{H}_{2.1}$ of Theorem 2.1, for $i = 1, \dots, m$,

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} \frac{\partial Q^{(n)}(\theta)}{\partial \theta_i} \Big|_{\theta=\theta^0} = 0.$$

Proof. Let $X_{t,i}^{(n)} = (1/n)(\partial \alpha_t^{(n)}(\theta)/\partial \theta_i)|_{\theta=\theta^0}$, $i = 1, \dots, m$. By definition of $\alpha_t^{(n)}(\theta)$, $\{X_{t,i}^{(n)}, F_t^{(n)}; t = 1, \dots, n, n \in \mathbb{N}\}$ is a martingale difference array. We use Proposition 4.1 with $k_n = n$, $c_t^{(n)} = 1/n$, $\pi = 2$, and $X_{t,i}^{(n)}/c_t^{(n)} = \partial \alpha_t^{(n)}(\theta)/\partial \theta_i|_{\theta=\theta^0}$. Conditions (ii) and (iii) are satisfied, and also (i) since $|X_{t,i}^{(n)}/c_t^{(n)}|^2$ is uniformly integrable as a consequence of Assumption $\mathbf{H}_{2.1}$, because $E_{\theta^0}|\partial \alpha_t^{(n)}(\theta)/\partial \theta_i|^{2+\delta}$ is uniformly bounded by C_1 . Hence

$$\sum_{t=1}^n X_{t,i}^{(n)} = \frac{1}{n} \sum_{t=1}^n \frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta_i} \Big|_{\theta=\theta^0} \xrightarrow{L_2} 0$$

when $n \rightarrow \infty$, $i = 1, \dots, m$ which implies convergence in probability. \square

Proposition 4.3. Under the assumptions on $\alpha_t^{(n)}(\theta)$, for $i, j = 1, \dots, m$ and for all t and n , let

$$Z_{t,ij}^{(n)} = \left\{ \frac{\partial^2 \alpha_t^{(n)}(\theta)}{\partial \theta_i \partial \theta_j} - E \left(\frac{\partial^2 \alpha_t^{(n)}(\theta)}{\partial \theta_i \partial \theta_j} \middle| F_{t-1}^{(n)} \right) \right\}_{\theta=\theta^0}.$$

Then, under Assumption **H**_{2.2} of Theorem 2.1, $\text{plim}_{n \rightarrow \infty} \sum_{t=1}^n Z_{t,ij}^{(n)}/n = 0$.

Proof. Let $X_{t,ij}^{(n)} = (1/n)Z_{t,ij}^{(n)}$, $i, j = 1, \dots, m$. Then $\{X_{t,ij}^{(n)}, F_t^{(n)}; t = 1, \dots, n, n \in \mathbb{N}\}$ is obviously a martingale difference array. We use Proposition 4.1 with $k_n = n$, $c_t^{(n)} = 1/n$, $\pi = 1$, and $X_{t,ij}^{(n)}/c_t^{(n)} = Z_{t,ij}^{(n)}$. Conditions (ii) and (iii) are satisfied like in Proposition 4.2 and also (i) since $|X_{t,ij}^{(n)}/c_t^{(n)}|$ is uniformly integrable as a consequence of Assumption **H**_{2.2}. Hence

$$\sum_{t=1}^n X_{t,ij}^{(n)} = \frac{1}{n} \sum_{t=1}^n \left\{ \frac{\partial^2 \alpha_t^{(n)}(\theta)}{\partial \theta_i \partial \theta_j} - E \left(\frac{\partial^2 \alpha_t^{(n)}(\theta)}{\partial \theta_i \partial \theta_j} \middle| F_{t-1}^{(n)} \right) \right\}_{\theta=\theta^0} \xrightarrow{L_1} 0$$

when $n \rightarrow \infty$, $i, j = 1, \dots, m$, which implies convergence in probability. \square

In the sequence case (i.e. when the model does not depend on n), Azrak and Melard (2006) used a theorem taken from Basawa and Prakasa Rao (1980, p. 388). In the array case, Azrak and Melard (2006) invoked Corollary 3.1 of Hall and Heyde (1980, p. 58) after having replaced the original conditional Lindeberg condition by an unconditional Lyapunov condition. As shown by Gaenssler et al. (1978), the unconditional Lindeberg condition implies the conditional Lindeberg condition. Since Lyapunov conditions are stronger than Lindeberg conditions, Alj et al. (2014) have fully justified the replacement mentioned above.

Proposition 4.4. Let $\{Y_t^{(n)}, F_t^{(n)}; t = 1, \dots, k_n, n \in \mathbb{N}\}$ be a scalar zero-mean square-integrable martingale difference array, and $\{k_n\}$ an increasing integer sequence with $\{k_n\} \uparrow \infty$. Suppose that for some $\delta > 0$, there exists a constant $B > 0$ such that

$$(4.1) \quad E(|Y_t^{(n)}|^{2+\delta}) < B < \infty$$

for all n and $t = 1, \dots, k_n$, and suppose that

$$\text{plim}_{n \rightarrow \infty} \frac{1}{k_n} \sum_{t=1}^n E(Y_t^{(n)2} | F_{t-1}^{(n)}) = \sigma^2,$$

then

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{k_n} Y_t^{(n)} \xrightarrow{L} Z$$

when $n \rightarrow \infty$, where Z has a $N(0, \sigma^2)$ distribution.

Proof. It is similar to Alj et al. (2014) (except that there $k_n = n$ and $\sigma = 1$ were assumed) where it is shown that it is a consequence of Brown (1971), more precisely Brown and Eagleson (1971), and of Gaenssler et al. (1978).

Proposition 4.5. Under the assumptions on $\alpha_t^{(n)}(\theta)$, plus Assumptions $\mathbf{H}_{2.1}$, $\mathbf{H}_{2.5}$, and $\mathbf{H}_{2.6}$ of Theorems 2.1 and 2.2, $Q^{(n)}(\theta) = \sum_{t=1}^n \alpha_t^{(n)}(\theta)$ satisfies

$$(4.2) \quad \frac{1}{\sqrt{n}} \frac{\partial Q^{(n)}(\theta)}{\partial \theta} \Big|_{\theta=\theta^0} = \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta} \Big|_{\theta=\theta^0} \xrightarrow{L} N(0, W),$$

when $n \rightarrow \infty$.

Proof. We consider an arbitrary $m \times 1$ vector λ . Let

$$Y_t^{(n)} = \lambda^T \frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta} \Big|_{\theta=\theta^0}.$$

By definition of $\alpha_t^{(n)}(\theta)$, $\{Y_t^{(n)}, F_t^{(n)}; t = 1, \dots, n, n \in \mathbb{N}\}$ is a martingale difference array. Using the Cramér-Wold device, we want to prove that $(1/\sqrt{n}) \sum_{t=1}^n Y_t^{(n)}$ converges in law to a normal distribution, for all λ . This will be done by using Proposition 4.4. For that purpose, let us check that there exists $\delta > 0$ and $M(\lambda)$, such that $E_{\theta^0}(|Y_t^{(n)}|^{2+\delta}) \leq M(\lambda) < \infty$, for all $t = 1, \dots, n, n \in \mathbb{N}$. We have

$$\left| \sum_{i=1}^m \lambda_i \frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta_i} \right| \leq \|\lambda\|_{\infty} \sum_{i=1}^m \left| \frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta_i} \right|,$$

using the maximum norm $\|\cdot\|_{\infty}$ such that $\|\lambda\|_{\infty} = \max_{i=1, \dots, m} |\lambda_i|$. But

$$\sum_{i=1}^m \left| \frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta_i} \right| \leq \left(\sum_{i=1}^m \left| \frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta_i} \right|^{2+\delta} \right)^{\frac{1}{2+\delta}} m^{\frac{1+\delta}{2+\delta}},$$

taking δ as defined in Assumption $\mathbf{H}_{2.1}$, and using Hölder inequality for sums with Hölder conjugates $p = 2 + \delta$ and $q = (2 + \delta)/(1 + \delta)$. Consequently, we have the upper bound

$$\begin{aligned} E_{\theta^0}(|Y_t^{(n)}|^{2+\delta}) &\leq m^{1+\delta} \|\lambda\|_\infty^{2+\delta} E_{\theta^0} \left(\sum_{i=1}^m \left| \frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta_i} \right|^{2+\delta} \right) \\ &\leq m^{2+\delta} \|\lambda\|_\infty^{2+\delta} C_1 < \infty, \end{aligned}$$

using Assumption $\mathbf{H}_{2.1}$. Hence the unconditional Lyapunov condition (4.1) is satisfied. There remains to prove that

$$(4.3) \quad \text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n E_{\theta^0}(Y_t^{(n)2} | F_{t-1}^{(n)}) = \sigma^2(\lambda),$$

where $\sigma^2(\lambda) > 0$. The left hand side of (4.3) is equal to

$$\begin{aligned} &\text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n E_{\theta^0} \left(\lambda^T \frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta} \frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta^T} \lambda \middle| F_{t-1}^{(n)} \right) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n E_{\theta^0} \left(\lambda^T \frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta} \frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta^T} \lambda \right) = \lambda^T W \lambda > 0, \end{aligned}$$

using Assumption $\mathbf{H}_{2.5}$ and the definition of W in Assumption $\mathbf{H}_{2.6}$. We take $\lambda^T W \lambda$ as $\sigma^2(\lambda)$. All the assumptions of Proposition 4.4 are then verified. Consequently, we have

$$n^{-1/2} \sum_{t=1}^n Y_t^{(n)} \xrightarrow{L} N(0, \sigma^2(\lambda)) \quad \text{or} \quad n^{-1/2} \lambda^T \sum_{t=1}^n \frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta} \bigg|_{\theta=\theta^0} \xrightarrow{L} N(0, \lambda^T W \lambda)$$

when $n \rightarrow \infty$. Since it is true for all λ , we deduce (4.2). \square

REMARK 4.1.

- (i) The thesis of Proposition 4.2 can also be seen as a consequence of Proposition 4.5 when assuming $\mathbf{H}_{2.5}$, and $\mathbf{H}_{2.6}$ in addition to $\mathbf{H}_{2.1}$.
- (ii) With respect Azrak and Mélard (2006), the proposition here uses a weaker moment assumption of order $2 + \delta$ instead of 4.

Proposition 4.6. Let $(T_1^{(n)}, \dots, T_m^{(n)})$, $m \in \mathbb{N}$, $n \in \mathbb{N}$ be a sequence of scalar random vectors tending weakly to (T_1, \dots, T_m) as $n \rightarrow \infty$, such that the

distribution of (T_1, \dots, T_m) has a density with respect to Lebesgue measure over \mathbb{R}^m , and suppose that for each fixed i and j , $i, j = 1, \dots, m$, $A_{ij}^{(n)}$ is a sequence of random variables tending in probability to constants A_{ij} for which the matrix $A = (A_{ij})$ is non singular. Let $B = (B_{ij}) = A^{-1}$. Then the solutions $(Y_1^{(n)}, \dots, Y_m^{(n)})$ of

$$\sum_{j=1}^m A_{ij}^{(n)} Y_j^{(n)} = T_i^{(n)}, \quad i = 1, \dots, m,$$

tend in probability to the solutions (Y_1, \dots, Y_m) of

$$\sum_{j=1}^m A_{ij} Y_j = T_i, \quad i = 1, \dots, m,$$

given by $Y_i = \sum_{j=1}^m B_{ij} T_j$.

Proof. See Lehmann and Casella (1998, Lemma 5.2, pp. 465-466). \square

Proof of Theorem 2.1. We follow Lehmann and Casella (1998, Section 6.5), given there for maximum likelihood estimators for i.i.d. observations (except that we use a second-order instead of a third-order Taylor expansion). Let

$$\begin{aligned} \alpha_t^{(n)}(\theta) - \alpha_t^{(n)}(\theta^0) &= \sum_{i=1}^m (\theta_i - \theta_i^0) \left. \frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta_i} \right|_{\theta=\theta^0} + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (\theta_i - \theta_i^0)(\theta_j - \theta_j^0) \left. \frac{\partial^2 \alpha_t^{(n)}(\theta)}{\partial \theta_i \partial \theta_j} \right|_{\theta=\theta^0} \\ &+ \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (\theta_i - \theta_i^0)(\theta_j - \theta_j^0) \left\{ \left. \frac{\partial^2 \alpha_t^{(n)}(\theta)}{\partial \theta_i \partial \theta_j} \right|_{\theta=\theta'_{ij}} - \left. \frac{\partial^2 \alpha_t^{(n)}(\theta)}{\partial \theta_i \partial \theta_j} \right|_{\theta=\theta^0} \right\}, \end{aligned}$$

where the θ'_{ij} are points in Θ located on the straight line joining θ and θ^0 . Summing over t and dividing by n gives

$$\frac{Q^{(n)}(\theta) - Q^{(n)}(\theta^0)}{n} = S_1^{(n)}(\theta, \theta^0) + S_2^{(n)}(\theta, \theta^0) + S_3^{(n)}(\theta, \theta^0),$$

where we denote

$$\begin{aligned} S_1^{(n)}(\theta, \theta^0) &= \sum_{i=1}^m (\theta_i - \theta_i^0) d_{1,i}^{(n)}, \\ S_2^{(n)}(\theta, \theta^0) &= \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (\theta_i - \theta_i^0)(\theta_j - \theta_j^0) d_{2,ij}^{(n)}, \\ S_3^{(n)}(\theta, \theta^0) &= \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (\theta_i - \theta_i^0)(\theta_j - \theta_j^0) d_{3,ij}^{(n)}(\theta'_{ij}), \end{aligned}$$

with

$$\begin{aligned} d_{1,i}^{(n)} &= \frac{1}{n} \sum_{t=1}^n \frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta_i} \Big|_{\theta=\theta^0}, \\ d_{2,ij}^{(n)} &= \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 \alpha_t^{(n)}(\theta)}{\partial \theta_i \partial \theta_j} \Big|_{\theta=\theta^0}, \\ d_{3,ij}^{(n)}(\theta'_{ij}) &= \frac{1}{n} \sum_{t=1}^n \left\{ \frac{\partial^2 \alpha_t^{(n)}(\theta)}{\partial \theta_i \partial \theta_j} \Big|_{\theta=\theta'_{ij}} - \frac{\partial^2 \alpha_t^{(n)}(\theta)}{\partial \theta_i \partial \theta_j} \Big|_{\theta=\theta^0} \right\}. \end{aligned}$$

In order to study $d_{1,i}^{(n)}$, $i = 1, \dots, m$, using Assumption $\mathbf{H}_{2.1}$, we refer to Proposition 4.2 which implies

$$(4.4) \quad \text{plim}_{n \rightarrow \infty} d_{1,i}^{(n)} = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta_i} \Big|_{\theta=\theta^0} = 0.$$

Let us now consider $d_{2,ij}^{(n)}$, for $i, j = 1, \dots, m$. Similarly, using Proposition 4.3 and Assumption $\mathbf{H}_{2.2}$,

$$(4.5) \quad \text{plim}_{n \rightarrow \infty} d_{2,ij}^{(n)} = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 \alpha_t^{(n)}(\theta)}{\partial \theta_i \partial \theta_j} \Big|_{\theta=\theta^0} = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n E_{\theta^0} \left(\frac{\partial^2 \alpha_t^{(n)}(\theta)}{\partial \theta_i \partial \theta_j} \Big|_{F_{t-1}^{(n)}} \right),$$

and, because of Assumption $\mathbf{H}_{2.3}$, $\text{plim}_{n \rightarrow \infty} d_{2,ij}^{(n)} = V_{ij}$.

We consider a sphere $B_\rho \subset \Theta$ with center at the true point θ^0 and radius ρ . First on B_ρ we have

$$|S_1^{(n)}(\theta, \theta^0)| \leq \rho \sum_{i=1}^m |d_{1,i}^{(n)}|.$$

For any given $\rho > 0$, it follows from (4.4), with probability tending to 1, that $|d_{1,i}^{(n)}| < \rho^2$ and hence that $|S_1^{(n)}(\theta, \theta^0)| < m\rho^3$. Next consider

$$\begin{aligned} 2S_2^{(n)}(\theta, \theta^0) &= \sum_{i=1}^m \sum_{j=1}^m V_{ij}(\theta_i - \theta_i^0)(\theta_j - \theta_j^0) \\ &+ \sum_{i=1}^m \sum_{j=1}^m (d_{2,ij}^{(n)} - V_{ij})(\theta_i - \theta_i^0)(\theta_j - \theta_j^0). \end{aligned}$$

By using (4.5) the absolute value of the second term is less than $m^2\rho^3$ with probability tending to 1, using an argument analogous to that for $S_1^{(n)}(\theta, \theta^0)$. The first term is a positive non-random quadratic form in the variables $(\theta_i - \theta_i^0)$. By an orthogonal transformation this can be reduced to diagonal form $\sum_{i=1}^m \lambda_i \xi_i^2$ with B_ρ becoming $\sum_{i=1}^m \xi_i^2 = \rho^2$. Suppose that the λ 's that are positive are numbered so that $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_m$. Then $\sum_{i=1}^m \lambda_i \xi_i^2 \geq \lambda_1 \sum_{i=1}^m \xi_i^2 = \lambda_1 \rho^2$. Combining the first and second terms, we see that there exist $c > 0$, $\rho_0 > 0$ such that for $\rho < \rho_0$, $S_2^{(n)}(\theta, \theta^0) > c\rho^2$ with probability tending to 1. By Assumption **H**_{2.4}, $d_{3,ij}^{(n)}$ is bounded in probability, uniformly in $n \in \mathbb{N}$, and $\theta \in \Theta$ and hence there exists $b > 0$ such that $|S_3^{(n)}(\theta, \theta^0)| < b\rho^3$ on B_ρ . Combining the three inequalities, we see that

$$\min(S_1^{(n)} + S_2^{(n)} + S_3^{(n)}) > c\rho^2 - (b + m)\rho^3$$

which is greater than zero if $\rho < c/(b + m)$.

Hence for any sufficiently small ρ the probability tends to 1 that

$$Q^{(n)}(\theta) - Q^{(n)}(\theta^0) = \sum_{t=1}^n \alpha_t^{(n)}(\theta) - \sum_{t=1}^n \alpha_t^{(n)}(\theta^0) > 0$$

at all points on the surface of B_ρ , and hence $Q^{(n)}(\theta)$ has a local minimum in the interior of B_ρ . Since the derivative of the criterion at the local minimum will be equal to zero, it will follow that for any $\rho > 0$, with probability tending to 1, there will be a solution, denoted by $\theta^{(n)}(\rho)$, within B_ρ , of the equation

$$Q^{(n)}(\theta) = 0.$$

By using the arguments of Lehmann and Casella (1998, pp. 447) for the maximum likelihood estimator of a single parameter in the case of i.i.d. observations, we can determine a sequence of estimators which doesn't depend on ρ .

This achieves the proof for the weak consistency. \square

Proof of Theorem 2.2. Let us now prove the convergence of $n^{1/2}(\hat{\theta}^{(n)} - \theta^0)$ to a normal distribution, again by following Lehmann and Casella (1998, pp. 466). For $i = 1, \dots, m$, let

$$\begin{aligned} \frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta_i} &= \left. \frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta_i} \right|_{\theta=\theta^0} + \sum_{j=1}^m (\theta_j - \theta_j^0) \left. \frac{\partial^2 \alpha_t^{(n)}(\theta)}{\partial \theta_i \partial \theta_j} \right|_{\theta=\theta^0} \\ &+ \sum_{j=1}^m (\theta_j - \theta_j^0) \left\{ \left. \frac{\partial^2 \alpha_t^{(n)}(\theta)}{\partial \theta_i \partial \theta_j} \right|_{\theta=\theta''_{ij}} - \left. \frac{\partial^2 \alpha_t^{(n)}(\theta)}{\partial \theta_i \partial \theta_j} \right|_{\theta=\theta^0} \right\} \end{aligned}$$

where the θ''_{ij} are located on the straight line joining θ and θ^0 . Summing over t and dividing by n gives

$$\begin{aligned} \frac{1}{n} \frac{\partial Q^{(n)}(\theta)}{\partial \theta_i} &= \frac{1}{n} \sum_{t=1}^n \left. \frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta_i} \right|_{\theta=\theta^0} \\ &+ \frac{1}{n} \sum_{t=1}^n \sum_{j=1}^m (\theta_j - \theta_j^0) \left. \frac{\partial^2 \alpha_t^{(n)}(\theta)}{\partial \theta_i \partial \theta_j} \right|_{\theta=\theta^0} \\ &+ \frac{1}{n} \sum_{t=1}^n \sum_{j=1}^m (\theta_j - \theta_j^0) \left\{ \left. \frac{\partial^2 \alpha_t^{(n)}(\theta)}{\partial \theta_i \partial \theta_j} \right|_{\theta=\theta''_{ij}} - \left. \frac{\partial^2 \alpha_t^{(n)}(\theta)}{\partial \theta_i \partial \theta_j} \right|_{\theta=\theta^0} \right\} \\ &= d_{1,i}^{(n)} + \sum_{j=1}^m (\theta_j - \theta_j^0) d_{2,ij}^{(n)} + \sum_{j=1}^m (\theta_j - \theta_j^0) d_{3,ij}^{(n)}(\theta''_{ij}). \end{aligned}$$

The estimator $\hat{\theta}^{(n)}$ is defined by the system of equations for $i = 1, \dots, m$

$$\frac{1}{n} \sum_{t=1}^n \left. \frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta_i} \right|_{\theta=\hat{\theta}^{(n)}} = 0$$

hence

$$0 = d_{1,i}^{(n)} + \sum_{j=1}^m (\hat{\theta}_j^{(n)} - \theta_j^0) d_{2,ij}^{(n)} + \sum_{j=1}^m (\hat{\theta}_j^{(n)} - \theta_j^0) d_{3,ij}^{(n)}(\theta''_{ij})$$

or

$$\sum_{j=1}^m \sqrt{n}(\hat{\theta}_j^{(n)} - \theta_j^0) \left\{ d_{2,ij}^{(n)} + d_{3,ij}^{(n)}(\theta''_{ij}) \right\} = -\sqrt{n}d_{1,i}^{(n)}.$$

These have the form in Proposition 4.6, $i, j = 1, \dots, m$, with

$$\begin{aligned} Y_j^{(n)} &= \sqrt{n}(\hat{\theta}_j^{(n)} - \theta_j^0) \\ A_{ij}^{(n)} &= d_{2,ij}^{(n)} + d_{3,ij}^{(n)}(\theta_{ij}''^{(n)}) \\ T_i^{(n)} &= -\sqrt{n}d_{1,i}^{(n)} = -\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial \alpha_t^{(n)}(\theta)}{\partial \theta_i} \Big|_{\theta=\theta^0}. \end{aligned}$$

By Proposition 4.5, given Assumptions $\mathbf{H}_{2.5}$ - $\mathbf{H}_{2.6}$ in Theorem 2.2, the $T_i^{(n)}$, $i = 1, \dots, m$, have a joint multivariate normal distribution with covariance matrix W . Using Theorem 2.1 and Assumption $\mathbf{H}_{2.4}$, the second term of the $A_{ij}^{(n)}$, $d_{3,ij}^{(n)}(\theta_{ij}''^{(n)})$, converges in probability to 0 and the first term $d_{2,ij}^{(n)}$, by (4.5), converges in probability to V_{ij} when $n \rightarrow \infty$. By Proposition 4.6, the limit distribution of the Y 's is therefore that of the solution (Y_1, \dots, Y_m) of the equations

$$\sum_{j=1}^m V_{ij} Y_j = T_i,$$

$i = 1, \dots, m$, where $T = (T_1, \dots, T_m)$ is multivariate normal with mean zero and covariance matrix W . It follows that the distribution of Y is that of $V^{-1}T$, which is multivariate normal with zero mean and covariance matrix $V^{-1}WV^{-1}$.

□

Proof of Theorem 2.3. For $i = 1, \dots, m$, denoting ∂_i the partial derivative with respect to θ_i , since $\partial_i \alpha_t(\theta)$ can be written as a sum of two terms $2e_t^{(n)T}(\theta)\Sigma_t^{(n)-1}(\theta)\partial_i e_t^{(n)}(\theta)$ and $e_t^{(n)T}(\theta)\partial_i \Sigma_t^{(n)-1}(\theta)e_t^{(n)}(\theta)$, by using Hölder inequality it is enough to prove that the $2 + \delta$ power of the absolute value of the expectation at $\theta = \theta^0$ of each term is uniformly bounded.

We have first for $i = 1, \dots, m$

$$\begin{aligned} \left| 2e_t^{(n)T}(\theta)\Sigma_t^{(n)-1}(\theta)\partial_i e_t^{(n)}(\theta) \right|^2 &= \text{tr}^2 \left(2e_t^{(n)T}(\theta)\Sigma_t^{(n)-1}(\theta)\partial_i e_t^{(n)}(\theta) \right) \\ &= 4 \text{tr}^2 \left(\Sigma_t^{(n)-1}(\theta)\partial_i e_t^{(n)}(\theta)e_t^{(n)T}(\theta) \right) \\ &\leq 4 \left\| \Sigma_t^{(n)-1}(\theta) \right\|_F^2 \\ &\quad \times \text{tr} \left(e_t^{(n)}(\theta)\partial_i e_t^{(n)T}(\theta)\partial_i e_t^{(n)}(\theta)e_t^{(n)T}(\theta) \right) \end{aligned}$$

by using TA Lemmas 3.4 and 3.3 of Alj et al. (2017). The trace in the last factor can be written $\text{tr}(\partial_i e_t^{(n)T}(\theta)\partial_i e_t^{(n)}(\theta)e_t^{(n)T}(\theta)e_t^{(n)}(\theta))$. This is equal to

the product of the scalars $|\partial_i e_t^{(n)T}(\theta) \partial_i e_t^{(n)}(\theta)|$ and $|e_t^{(n)T}(\theta) e_t^{(n)}(\theta)|$. Hence

$$\begin{aligned} & \left| 2e_t^{(n)T}(\theta) \Sigma_t^{(n)-1}(\theta) \partial_i e_t^{(n)}(\theta) \right|^{2+\delta} \\ & \leq 2^{2+\delta} \left\| \Sigma_t^{(n)-1}(\theta) \right\|_F^{2+\delta} \left| \partial_i e_t^{(n)T}(\theta) \partial_i e_t^{(n)}(\theta) \right|^{1+\delta/2} \left| e_t^{(n)T}(\theta) e_t^{(n)}(\theta) \right|^{1+\delta/2}. \end{aligned}$$

Taking the expectation of that product at $\theta = \theta^0$, using the assumptions, and the fact that for any random variable X with finite moment of order $q \geq p > 0$, $(E|X|^p)^{1/p} \leq (E|X|^q)^{1/q}$, with $p = 2 + \delta$ and $q = 4$, we conclude that the $2 + \delta$ power of the absolute value of the expectation at $\theta = \theta^0$ of the first term $2e_t^{(n)T}(\theta) \Sigma_t^{(n)-1}(\theta) \partial_i e_t^{(n)}(\theta)$ is uniformly bounded.

We proceed similarly for the second term $e_t^{(n)T}(\theta) \partial_i \Sigma_t^{(n)-1}(\theta) e_t^{(n)}(\theta)$. Its square equals

$$\begin{aligned} & \text{tr}^2 \left(e_t^{(n)T}(\theta) \partial_i \Sigma_t^{(n)-1}(\theta) e_t^{(n)}(\theta) \right) \\ & = \text{tr}^2 \left(\partial_i \Sigma_t^{(n)-1}(\theta) e_t^{(n)}(\theta) e_t^{(n)T}(\theta) \right) \\ & \leq \left\| \partial_i \Sigma_t^{(n)-1}(\theta) \right\|_F^2 \text{tr} \left(e_t^{(n)}(\theta) e_t^{(n)T}(\theta) e_t^{(n)}(\theta) e_t^{(n)T}(\theta) \right) \end{aligned}$$

The trace in the last factor can be written $\text{tr}(e_t^{(n)T}(\theta) e_t^{(n)}(\theta) e_t^{(n)T}(\theta) e_t^{(n)}(\theta))$. This is equal to the square of the scalar $|e_t^{(n)T}(\theta) e_t^{(n)}(\theta)|$. Hence

$$\left| e_t^{(n)T}(\theta) \partial_i \Sigma_t^{(n)-1}(\theta) e_t^{(n)}(\theta) \right|^{2+\delta} \leq \left\| \partial_i \Sigma_t^{(n)-1}(\theta) \right\|_F^{2+\delta} \left| e_t^{(n)T}(\theta) e_t^{(n)}(\theta) \right|^{2+\delta}.$$

Taking the expectation of that product at $\theta = \theta^0$ and using the assumptions implies that the $2 + \delta$ power of the absolute value of the expectation at $\theta = \theta^0$ of the second term $e_t^{(n)T}(\theta) \partial_i \Sigma_t^{(n)-1}(\theta) e_t^{(n)}(\theta)$ is uniformly bounded. Consequently, Assumption $\mathbf{H}_{2.1}$ of Theorem 2.1 is satisfied. \square

Proof of Theorem 2.4. We know that

$$E \left(\left| \frac{1}{n} \sum_{t=1}^n x_t^{(n)} \right|^2 \right) < \frac{k}{n}$$

for some $k > 0$. We start from Markov inequality with exponent 2, for any $\epsilon > 0$,

$$P \left(\left| \frac{1}{n} \sum_{t=1}^n x_t^{(n)} \right| > \epsilon \right) \leq P \left(\left| \frac{1}{n} \sum_{t=1}^n x_t^{(n)} \right| \geq \epsilon \right) \leq \frac{E \left| \frac{1}{n} \sum_{t=1}^n x_t^{(n)} \right|^2}{\epsilon^2} \leq \frac{k}{n\epsilon^2}$$

We need to show that for every $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n} \sum_{t=1}^n x_t^{(n)} \right| \leq \epsilon \right) = 1$$

that is to say that for every $\epsilon > 0$, and for every $\eta > 0$, there exists $N(\epsilon, \eta)$ such that for all $n > N(\epsilon, \eta)$

$$P \left(\left| \frac{1}{n} \sum_{t=1}^n x_t^{(n)} \right| \leq \epsilon \right) > 1 - \eta.$$

We choose $N(\epsilon, \eta) = k/(\eta\epsilon^2)$ such that

$$P \left(\left| \frac{1}{n} \sum_{t=1}^n x_t^{(n)} \right| \leq \epsilon \right) > 1 - \frac{k}{n\epsilon^2} > 1 - \frac{k}{N(\epsilon, \eta)\epsilon^2} = 1 - \eta.$$

□

Proof of Theorem 2.5. Let $M = \sup_n \{(1/n) \sum_{t=1}^n |v_t^{(n)}|\}$. We know that, for any $\epsilon > 0$:

$$\exists N_1 \in \mathbb{N} \text{ such that } \forall n > N_1 : \left| \frac{1}{n} \sum_{t=1}^n v_t^{(n)} - L \right| < \epsilon/3C$$

$$\exists N_2 \in \mathbb{N} \text{ such that } \forall n > N_2, \forall t \in \{N_2 + 1, \dots, n\} : |u_t^{(n)} - C| < \epsilon/3M.$$

Write $u_t^{(n)}v_t^{(n)} - CL = C(v_t^{(n)} - L) + (u_t^{(n)} - C)v_t^{(n)}$. Hence, because $C > 0$,

$$\begin{aligned} \left| \frac{1}{n} \sum_{t=1}^n u_t^{(n)}v_t^{(n)} - CL \right| &= \left| \frac{1}{n} \sum_{t=1}^n C(v_t^{(n)} - L) + \frac{1}{n} \sum_{t=N_2+1}^n (u_t^{(n)} - C)v_t^{(n)} \right. \\ (4.6) \quad &+ \left. \frac{1}{n} \sum_{t=1}^{N_2} (u_t^{(n)} - C)v_t^{(n)} \right| \\ &\leq \frac{C}{n} \left| \sum_{t=1}^n (v_t^{(n)} - L) \right| + \frac{1}{n} \sum_{t=N_2+1}^n |u_t^{(n)} - C| |v_t^{(n)}| \\ &+ \frac{1}{n} \sum_{t=1}^{N_2} |u_t^{(n)} - C| |v_t^{(n)}|. \end{aligned}$$

Each of the first two terms is bounded by $\epsilon/3$. Let us show that this is also the case for the third term. Let $U_{N_2} = \max_{n=1}^{N_2} \max_{t=1}^n \{|u_t^{(n)} - C| |v_t^{(n)}|\}$. Consequently, the third term is bounded by $N_2 U_{N_2}/n$. Taking $N_3 = 3N_2 U_{N_2}/\epsilon$, for $n > \max\{N_1, N_2, N_3\}$, the left hand side of (4.6) is bounded by ϵ and this achieves the proof. □

Proof of Theorem 2.6. This is a direct consequence of the definition of Riemann integration using a partition of the interval $[0, 1]$ into n sub-intervals of length $1/n$. \square

Acknowledgments. Guy M elard has benefited from a Belgian research grant F.R.S.-FNRS 1.5.261.09. We thank Prof. Nikolaos Kourogenis, University of Piraeus (Greece), for providing the reference of Brown and Eagleson (1971). We thank also Abdelkamel Alj for his comments on the proof of Theorem 2.1. The second author thanks Marcella Niglio (University of Salerno, Italy) and Luisa Bisaglia (University of Padua, Italy) for their invitations in their universities, Marcella Niglio for her remarks, and Konstantinos Fokianos for references and discussions on count data models. This paper was presented at several places by the second author who thanks Aboubacar Amiri, Marianne Clausel, Christian Francq, Marc Hallin, Yacouba Boubacar Mainassara, and Davy Paindaveine, among others.

References.

- Abramowitz, M. & Stegun I. A. (1965). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover Publications, New York.
- Alj, A., Azrak, R., Ley, C., & M elard, G. (2017). Asymptotic properties of QML estimators for VARMA models with time-dependent coefficients. *Scandinavian Journal of Statistics* **44**, 617-635. DOI: 10.1111/sjos.12268.
- Alj, A., Azrak, R. & M elard, G. (2014). On conditions in central limit theorems for martingale difference arrays. *Economics Letters* **123**, 305-307.
- Alj, A., Azrak, R. & M elard, G. (2020). General estimation results for td-VARMA models. In preparation.
- Amendola, A., Niglio, M. & Vitale, C. (2007). The autocorrelation functions in SETARMA models, in Kontoghiorghes, E. J., and Gatu, C. (Eds.) *Optimisation, Econometric and Financial Analysis*, Springer-Verlag, Berlin, pp. 127-142.
- Azrak, R. & M elard, G. (2006). Asymptotic properties of quasi-likelihood estimators for ARMA models with time-dependent coefficients. *Statistical Inference for Stochastic Processes* **9**, 279-330.
- Bai, Z. D. & Wu, Y. (1997). General M-estimation. *Journal of Multivariate Analysis* **63**, 117-135.
- Basawa, I. V. & Prakasa Rao, B. L. S. (1980). *Statistical Inference for Stochastic Processes*. Academic Press, New York.
- Brown, B. M. (1971). Martingale central limit theorems. *Annals of Mathematical Statistics* **42**, 59-66
- Brown, B. M. & Eagleson, G.K. (1971). Martingale convergence to infinitely divisible laws with finite variances. *Transactions of the American Mathematical Society* **162**, 449-453.
- Chen, R. & Tsay, R.S. (1993). Functional-coefficient autoregressive models. *Journal of the American Statistical Association* **88**, 298-308.
- Davidson, J. E. H. (1994). *Stochastic Limit Theorems. An Introduction for Econometricians*. Oxford University Press, Oxford.

- Davis, R. A., Dunsmuir, W. T. M., & Streett, S. B. (2003). Observation driven models for Poisson counts, *Biometrika* **90**, 777-790.
- Doukhan, P, Fokianos, K., & Tjøstheim, D. (2012). On weak dependence conditions for Poisson autoregressions, *Statistics and Probability Letters* **82**, 942-948.
- Erdélyi, A. (1953). *Higher Transcendental Functions, Vol. 1*, McGraw-Hill, New York.
- Franco, C. & Zakoian, J.-M. (1998). Estimating linear representations of nonlinear processes. *Journal of Statistical Planning and Inference* **68** 145-165.
- Fokianos, K., Støve, B., Tjøstheim, D., & Doukhan, P, (2020). Multivariate count autoregression, *Bernoulli* **26**(1), 471-499.
- Gaenssler, P., Strobel, J., & Stute, W. (1978). On central limit theorems for martingale arrays. *Acta Mathematica Academiae Scientiarum Hungaricae*, **31**, 205-216.
- Giordano, F., Niglio, M., & Vitale, C.D. (2013). Ergodicity of threshold autoregressive models: the two regimes case. https://www.sis-statistica.it/old_upload/contenuti/2013/09/RS12-On-the-stationarity-of-the-Threshold.pdf
- Hall, P. & Heyde, C. C. (1980). *Martingale Limit Theory and its Application*. Academic Press, New York.
- Klimko, L. A. & Nelson, P. I. (1978). On conditional least squares estimation for stochastic processes. *Annals of Statistics* **6**, 629-642.
- Lai, T.L. (1994). Asymptotic properties of nonlinear least squares estimates in stochastic regression models. *Annals of Statistics* **22**, 1917-1930.
- Lehmann, E. L. & Casella, G. (1998). *Theory of Point Estimation*. Springer Verlag, New York.
- Liboschik, T., Fokianos, K., & Fried, R. (2017). tscount: An R package for analysis of count time series following generalized linear models, *Journal of Statistical Software* **82** (5). doi: 10.18637/jss.v082.i05
- Liu Yan Asymptotics for M-estimators in Time Series ASTE : advances in science, technology and environmentology : special issue
- Martin, R.D. (1981). Robust methods for time series, in Findley D. F. (Ed) *Applied Time Series II*, Academic Press, New York, pp. 683-759.
- Mélard, G. (2020a). An indirect proof for the asymptotic properties of VARMA model estimators, submitted.
- Mélard, G. (2020b). On an alternative count time series model, in preparation.
- Mélard, G., & Niglio, M. (2020). Another approach for asymptotic properties of threshold models, in preparation.
- Petrucelli, J.D., & Woolford, S.W. (1984). A threshold AR(1) model. *Journal of Applied Probability* **21**, 270-286.
- Nelson, P.I. (1980). A note on strong consistency of least squares estimators in regression models with martingale difference errors. *Annals of Statistics* **8**, 1057-1064.
- Taniguchi, M., & Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*. Springer Verlag, New York.
- Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association* **89**, 208-218.
- Tjøstheim, D. (1986). Estimation in nonlinear time series models. *Stochastic Processes and their Applications* **21**, 251-273.
- Tjøstheim, D. (1994). Non-linear time series: a selective review. *Scandinavian Journal of Statistics* **21**, 97-130.
- Tong, H. (1990). *Nonlinear Time Series: A Dynamical Systems Approach*. Oxford University Press, Oxford.
- Tong, H. & Lim, K.S. (1980). Threshold autoregression, limit cycles and cyclical data, *Journal of the Royal Statistical Society Series B* **42**, 245-292.