PLISKA

STUDIA MATHEMATICA BULGARICA

ПЛИСКА

БЪЛГАРСКИ МАТЕМАТИЧЕСКИ СТУДИИ

Pliska Stud. Math. Bulgar. **19** (2009), 135–156

PLISKA
STUDIA MATHEMATICA
BULGARICA

STRONG CONSISTENCY OF THE CONDITIONAL LEAST SQUARES ESTIMATOR FOR A NONSTATIONARY PROCESS. EXAMPLE OF THE *GARCH* MODEL

Christine Jacob

We consider the Conditional Least Squares Estimator (CLSE) of a unknown parameter $\theta_0 \in \mathbb{R}^p$ of the conditional expectation of a real stochastic process $\{Y_n\}$ having finite first two conditional moments $E(Y_n|\mathcal{F}_{n-1}) \overset{a.s.}{<} \infty$, $E(Y_n^2|\mathcal{F}_{n-1}) \overset{a.s.}{<} \infty$ at each time n, where $E(Y_n|\mathcal{F}_{n-1})$ is Lipschitz and may be nonlinear in θ_0 and $\{\mathcal{F}_n\}$ is an increasing sequence of σ-algebra. We generalize to this class of processes the necessary and sufficient condition got for the strong consistency of the CLSE of θ_0 in the particular linear deterministic (or linear stochastic if $p = 1$) model $E(Y_n|\mathcal{F}_{n-1}) = \theta_0^T W_n$. We illustrate this theoretical result with examples, mainly a nonstationary $GARCH(1,1)$ model.

1. Introduction

Let $\{Y_n\}_{n\in\mathbb{N}}$ be an observed one-dimensional real stochastic process defined on a probability space $(\Omega, \mathcal{F}, P)$ and let $\{\mathcal{F}_n\}$ be an increasing sequence of σ-algebra depending on observed processes. We assume that $\{Y_n\}$ satisfies:

2000 *Mathematics Subject Classification:* Primary: 62M10, 62J02, 62F12, 62M05, 62P05, 62P10; secondary: 60G46, 60F15.

Key words: Stochastic nonlinear regression, Heteroscedasticity, Nonstationary, Conditional Least Squares Estimator, Consistency, *GARCH* model.

$$
\begin{aligned}
M_Y: \quad & E(Y_n|\mathcal{F}_{n-1}) =: f_n(\theta_0, \nu_0) = f_n^{(1)}(\theta_0) + f_n^{(2)}(\theta_0, \nu_0) \\
& \theta_0 \in \Theta \subset \mathbb{R}^p,\ 0 < p < \infty, \nu_0 \in \mathcal{N} \subset \mathbb{R}^q, 0 \leq q \leq \infty, \\
(1) \qquad & \overline{\lim_n} E(Y_n^2|\mathcal{F}_{n-1}) \overset{a.s.}{<} \infty,
\end{aligned}
$$

where θ_0 is a unknown parameter that we want to estimate, $f_n^{(1)}(\theta_0)$ is the $\mathcal{F}_{n-1}$-measurable parametric part of the model that may be nonlinear in θ_0, and $f_n^{(2)}(\theta_0, \nu_0)$ is a $\mathcal{F}_{n-1}$-measurable nuisance part that may be nonparametric. In this case $q = \infty$ and $\nu_0 = \{f_n^{(2)}(\theta_0, \nu_0)\}$. The case $q = 0$ is defined by $f_n^{(2)}(\theta_0, \nu_0) = 0$, for all n, and corresponds to the classical parametric setting. We assume that, as $n \to \infty$, $\{f_n^{(2)}(\theta_0, \nu_0)\}$ is asymptotically relatively negligible compared to $\{f_n^{(1)}(\theta_0)\}$ (see section 3.). So finally

$$Y_n = f_n^{(1)}(\theta_0) + f_n^{(2)}(\theta_0, \nu_0) + \eta_n,$$

where by construction η_n is a martingale difference, that is $E(\eta_n|\mathcal{F}_{n-1}) = 0$.

Some examples of processes satisfying M_Y are nonlinear regression models with random covariates and heteroscedastic variances, noisy dynamical models, nonlinear time series model (TARMA, SETAR, bilinear processes, ...), financial models ($ARCH$ initiated by [3], $GARCH$ and others), and branching processes.

We deal here with the CLSE (Conditional Least Squares Estimators) of θ_0 in the approximate model $\{f_k(\theta_0, \widehat{\nu})\}_{k \leq n}$, where $\widehat{\nu} = \{\widehat{\nu}_n\}$ is any sequence of estimations of ν_{0n} when $q = \infty$, or ν_0 when $q < \infty$, with either $f_k(\theta_0, \widehat{\nu}) = f_k(\theta_0, \widehat{\nu}_n)$, for all $k \leq n$, if $q < \infty$, and $f_k(\theta_0, \widehat{\nu}) = f_k(\theta_0, \widehat{\nu}_k)$, for all k, if $q = \infty$. The CLSE of θ_0 is defined by:

$$(2) \qquad \widehat{\theta}_n = \arg\min_{\theta \in \Theta} S_{n|\widehat{\nu}}(\theta),\ S_{n|\widehat{\nu}}(\theta) = \sum_{k=1}^{n} (Y_k - f_k(\theta, \widehat{\nu}))^2.$$

It is well-known in classical regression theory that the rate of convergence of $\{\widehat{\theta}_n\}$ is optimal when $E(\eta_n^2|\mathcal{F}_{n-1})$ is constant in n. Generalizing this property, $\{\widehat{\theta}_n\}$ will have an optimal convergence rate under the following assumption

$$(3) \qquad 0 < \underline{\lim_n} E_(\eta_n^2|\mathcal{F}_{n-1}) \overset{a.s.}{\leq} \overline{\lim_n} E(\eta_n^2|\mathcal{F}_{n-1}) \overset{a.s.}{<} \infty.$$

So $\{Y_n\}$ will be generally derived from the studied original process $\{Z_n\}$ by some suitable normalization: $Y_n = Z_n \lambda_n^{-1/2}$, where λ_n is $\mathcal{F}_{n-1}$ measurable and independent of the unkown parameters of the model. The particular cases $q = \infty$ with $\widehat{\nu} = 0$ and $q = 0$ correspond to the estimations of θ_0 in the parametric model $\{f_k^{(1)}(\theta_0)\}_{k \leq n}$ considered respectively as an approximate model or an exact one.

We will focus here on the strong consistency which is particularly useful when the goal of the estimation is either the knowledge of the true parameter or the prediction of the future behavior of the process from the estimated model or the estimation of the residual distribution.

The consistency property of $\{\widehat{\theta}_n\}$ is a classical topic in the parametric setting $q = 0$. The proofs and the required conditions depend on the linearity or nonlinearity of $f_n(.)$ in θ_0, since in the linear case $\{\widehat{\theta}_n\}$ has an explicit expression allowing direct proofs contrary to the general nonlinear case. They also depend on the properties of $\{\eta_n\}$, if they are independent, stationary or only check the more general martingale difference property. We assume here only this last property, particular useful when η_n may be nonstationary. Published conditions for getting the strong consistency of $\{\widehat{\theta}_n\}$ in model M_Y involve the strong identifiability of the parameter and this last condition is a necessary and sufficient one in the particular linear model $f_n(\theta) = \theta^T W_n$, where W_n is either a deterministic vector and $\{\eta_n\}$ are i.i.d. ([9]) or W_n is stochastic with $p = 1$ ([10]): $\lim_n \widehat{\theta}_n \stackrel{a.s.}{=} \theta_0$ if and only if

$$\lim_n \lambda_{\min} \left\{ \sum_{k=1}^n W_k W_k^T \right\} \stackrel{a.s.}{=} \infty, \tag{4}$$

where $\lambda_{\min} \left\{ \sum_{k=1}^n W_k W_k^T \right\}$ is the smallest eigenvalue of $\sum_{k=1}^n W_k W_k^T$. Defining

$$D_n(\theta) = \sum_{k=1}^n (f_k(\theta) - f_k(\theta_0))^2, \tag{5}$$

(4) is equivalent to $\lim_n D_n(\theta) \stackrel{a.s.}{=} \infty$, for all $\theta \neq \theta_0$. This quantity is the identifiability criterion of θ_0 in the model.

But in the general nonlinear parametric setting M_Y with $q = 0$, under some Lipschitz property of the model, the published theorems of consistency require additional sufficient conditions. One condition is $\overline{\lim_n} E(\eta_n^2 | \mathcal{F}_{n-1}) \stackrel{a.s.}{<} \infty$. The other ones concern some rate of convergence to ∞ of $\{D_n(\theta)\}$, and differ from one author to another one ([1], [5], [7], [10], [11], [12], [17], [18], [21], ...). We removed here these last unnecessary conditions, thus generalizing the necessary and sufficient condition (4) to our general setting M_Y with $0 \le q \le \infty$. The generalization of (4) that we called strong identifiability of θ_0, and reduced to (4) in the linear case, is the following when $q = 0$:

$$(6) \qquad SI(\{D_n(\theta)\}) : \lim_n \inf_{\theta : \|\theta - \theta_0\| \ge \delta} D_n(\theta) \overset{a.s.}{=} \infty, \forall\, \delta > 0,$$

where $\|\cdot\|$ is any norm in $\mathbb{R}^p$. This result is got thanks to a SLLNSM (strong law of large numbers for submartingales).

The paper is organized in the following way. We show in Section 2. that condition $SI(\{D_n(\theta)\})$ is a necessary and sufficient condition of consistency in the general parametric setting M_Y with $q = 0$ provided that the model satisfies the following Lipschitz condition:

$LIP(\{f_k(\theta)\})$: for all k, there exists a nonnegative $\mathcal{F}_{k-1}$-measurable function g_k and a function $h(\cdot) : \mathbb{R}^+ \to \mathbb{R}^+$ with $\lim_{x \searrow 0} h(x) = 0$, such that $\forall \theta_1 \in \Theta, \theta_2 \in \Theta$, $|f_k(\theta_1) - f_k(\theta_2)| \overset{a.s.}{\le} h(\|\theta_1 - \theta_2\|) g_k$.

This result is got thanks to a SLLNSM given in Section 5..

In Section 3., the consistency conditions are generalized to M_Y with $0 \le q \le \infty$. This result shows the robustness of the consistency property since if $\widehat{\theta}_n$ is strongly consistent in a given model, then it is strongly consistent in every model "close" to this given model.

In Section 4., we illustrate this result by some examples. The main one is a $GARCH(1,1)$ model $\xi_n = s_n(\theta) U_n$, $\{U_n\}$ i.i.d. (0,1), where $s_n(\theta_0)$ is the volatility of the process. The strong consistency of the CLSE is proved without assuming the stationarity of the process contrary to existing references based on quasi-likelihood or CLSE ([2], [16], [15]). We show that the strong consistency of the nonweighted CLSE is ensured as soon as $\{\xi_n^2\}$ does not die out as $n \to \infty$, and, when dealing with weighted CLSE, if in addition, the sequence of successive upper records $\{\xi_{L_m}^2\}$ satisfies $\xi_{L_m}^2 \overset{a.s.}{=} O([(L_{m+1} - L_m) m]^{1/2})$, where L_m is the index of the mth record. The nonstationarity allows burst phenomena with a very large amplitude compared to the stationarity setting.

Other examples and detailed proofs may be found in [6].

2. Consistency in M_Y under $q = 0$

In all existing results based on the martingale difference property of $\{\eta_n\}$, $D_n(\theta)$ defined by (5) plays a crucial role. This quantity is an identifiability criteria of the parameter: θ_0 is identifiable in $\{f_n(\cdot)\}_n$ if for all $\theta \ne \theta_0$, $\{f_n(\theta)\}_n \overset{a.s.}{\ne} \{f_n(\theta_0)\}_n$

which is equivalent to $\lim_n D_n(\theta) \overset{a.s.}{\neq} 0$. In the linear model $f_n(\theta) = \theta_0^T W_n$, this condition is equivalent to $\lim_n \lambda_{\min} \left\{ \sum_{k=1}^{n} W_k W_k^T \right\} \overset{a.s.}{\neq} 0$. Moreover the consistency of $\{\widehat{\theta}_n\}$ in this setting, when W_n is deterministic for all n ([9]) or W_n is stochastic with $p = 1$ ([10]), is got if and only if (4) is checked. For example, let $p = 1$ and $Y_n = \theta_0 W_n + \eta_n$, $\{\eta_n\}$ i.i.d. $(0, \sigma_0^2)$, $\{W_n\}$ deterministic. Then $\widehat{\theta}_n - \theta_0 = \left[\sum_{k=1}^{n} (Y_k - \theta_0 W_k) W_k\right] \left[\sum_{k=1}^{n} W_k^2\right]^{-1}$. So $Var(\widehat{\theta}_n - \theta_0) = \sigma^2 \left[\sum_{k=1}^{n} W_k^2\right]^{-1}$ which does not tend to 0 if $\sum_{k=1}^{\infty} W_k^2 < \infty$.

In this section, we generalize (4) to the setting of model M_Y under $q = 0$. Assume $\theta_0 \in \Theta$, where Θ is an open set and $\overline{\Theta}$ is compact. let B_δ^c be the complementary in Θ of the open ball of center θ_0 and radius δ. Denote $d_k(\theta) = f_k(\theta_0) - f_k(\theta)$, and $L_n(\theta) = \sum_{k=1}^{n} \eta_k d_k(\theta)$.

Proposition 2.1. *Let $p < \infty$ and assume that there exists $\Omega_\infty \subset \Omega$ with $P(\Omega_\infty) > 0$ and such that on Ω_∞, $SI(\{D_n(\theta)\})$ and $LIP(\{f_k(\theta)\})$ are checked. Then $\lim_n \widehat{\theta}_n \overset{a.s.}{=} \theta_0$ on Ω_∞.*

Proof. We use Wu's Lemma ([20], see Lemma 5.1) and Wu's decomposition based on $Y_k - f_k(\theta) = \eta_k + d_k(\theta)$ ([20]) implying $S_n(\theta) - S_n(\theta_0) = D_n(\theta) + 2L_n(\theta)$ and consequently,

$$\inf_{\theta \in B_\delta^c} S_n(\theta) - S_n(\theta_0) \geq \inf_{\theta \in B_\delta^c} D_n(\theta)[1 - 2 \sup_{\theta \in B_\delta^c} |L_n(\theta)| [D_n(\theta)]^{-1}] \tag{7}$$

Then the proof directly follows from the SLLNSM (Proposition 5.1) applied on $d_k(\theta) = f_k(\theta_0) - f_k(\theta)$ and $\widetilde{\Theta} = B_\delta^c$. □

Assume now that $\widehat{\theta}_{h,n} = \arg\min_{\theta \in \Theta} S_{h,n}(\theta)$, $S_{h,n}(\theta) = \sum_{k=h+1}^{n} (Y_k - f_k(\theta))^2$, where h may depend on n (for example $n - h$ is constant, for all n), and denote $L_n(\theta) - L_h(\theta) =: L_{h,n}(\theta)$ and $D_n(\theta) - D_h(\theta) =: D_{h,n}(\theta)$.

Proposition 2.2. *Assume that there exists $\Omega_\infty \subset \Omega$ with $P(\Omega_\infty) > 0$ and such that, on Ω_∞, $SI(\{D_n(\theta)\})$, $LIP(\{f_k(\theta)\})$ are checked, and moreover $\overline{\lim_n} \sup_{\theta \in B_\delta^c} D_n(\theta)[D_{h,n}(\theta)]^{-1} \overset{a.s.}{<} \infty$. Then $\lim_n \widehat{\theta}_{h,n} \overset{a.s.}{=} \theta_0$ on Ω_∞.*

When $n - h$ is constant, this setting is particularly suitable for branching processes studied on their nonextinction set since the proposition requires that $D_{h,n}(\theta)$ tends to ∞ at the same rate as $D_n(\theta)$.

Proof. According to (7) written with $S_{h,n}(\theta)$ instead of $S_n(\theta)$, it is sufficient to prove that $\lim_n \sup_\theta L_{h,n}(\theta)[D_{h,n}(\theta)]^{-1} = 0$. For that, we use

$$\frac{L_{h,n}(\theta)}{D_{h,n}(\theta)} = \frac{L_n(\theta)}{D_n(\theta)}\frac{D_n(\theta)}{D_{h,n}(\theta)} - \frac{L_h(\theta)}{D_h(\theta)}\frac{D_h(\theta)}{D_{h,n}(\theta)}. \qquad \square$$

3. Consistency in M_Y containing a nuisance part

Assume now that $\{f_n^{(2)}(\theta_0, \nu_0)\}$ is not identically null. We consider the CLSE of θ_0 in the approximate model $\{f_k(\theta, \widehat{\nu})\}_{k \le n}$, defined by (2). Since we deal with a.s. consistency, we assume in the proofs that $\widehat{\nu}$ is a nonrandom sequence of estimations.

Let

$$D_n(\theta, \widehat{\nu}) := \sum_{k=1}^n [d_k(\theta, \widehat{\nu})]^2, \quad D_n^{(1)}(\theta) := \sum_{k=1}^n [d_k^{(1)}(\theta)]^2,$$

$$d_k^{(1)}(\theta) := f_k^{(1)}(\theta_0) - f_k^{(1)}(\theta), \quad D_n^{(2)}(\theta, \widehat{\nu}) := \sum_{k=1}^n [d_k^{(2)}(\theta, \widehat{\nu})]^2,$$

$$d_k^{(2)}(\theta, \widehat{\nu}) := f_k^{(2)}(\theta_0, \nu_0) - f_k^{(2)}(\theta, \widehat{\nu}), \quad L_n^{(2)}(\theta, \widehat{\nu}) := \sum_{k=1}^n \eta_k d_k^{(2)}(\theta, \widehat{\nu})$$

and so on.

As in section 2. assume $\theta_0 \in \Theta$, where Θ is an open set and $\overline{\Theta}$ is compact, and assume $\nu \in \mathcal{N}$, where $\overline{\mathcal{N}}$ is compact when $q < \infty$.

Proposition 3.1. *Let $p < \infty$ and assume that there exists $\Omega_\infty \subset \Omega$ with $P(\Omega_\infty) > 0$ and such that on Ω_∞,*

1. *$SI(\{D_n^{(1)}(\theta)\})$ is checked and for all $\delta > 0$,*

$$\overline{\lim_n} \sup_{\theta \in B_\delta^c} D_n^{(2)}(\theta, \widehat{\nu}) \left[\inf_{\theta \in B_\delta^c} D_n^{(1)}(\theta)\right]^{-1} \overset{a.s.}{=} 0.$$

2. *$LIP(\{f_k(\theta, \nu)\})$ is checked in θ, ν when $q < \infty$ (resp. $LIP(\{f_k(\theta, \widehat{\nu})\})$ is checked in θ when $q = \infty$).*

Then $\lim_n \widehat{\theta}_{n|\widehat{\nu}} \stackrel{a.s.}{=} \theta_0$ *on* Ω_∞.

P r o o f. We have

$$S_{n|\widehat{\nu}}(\theta) - S_{n|\widehat{\nu}}(\theta_0) = D_n(\theta, \widehat{\nu}_n) - D_n^{(2)}(\theta_0, \widehat{\nu}_n) + 2L_n(\theta, \widehat{\nu}_n) - 2L_n^{(2)}(\theta_0, \widehat{\nu}_n).$$

Since $D_n(\theta, \widehat{\nu}) = D_n^{(1)}(\theta) + D_n^{(2)}(\theta, \widehat{\nu}) + 2\sum_{k=1}^n d_k^{(1)}(\theta) d_k^{(2)}(\theta, \widehat{\nu})$, using Hölder's inequality, we get

$$\inf_{\theta \in B_\delta^c} S_{n|\widehat{\nu}}(\theta) - S_{n|\widehat{\nu}}(\theta_0) \geq \inf_{\theta \in B_\delta^c} D_n^{(1)}(\theta) \left[1 - 2 \sup_{\theta \in B_\delta^c} \left[\frac{D_n^{(2)}(\theta, \widehat{\nu})}{D_n^{(1)}(\theta)} \right]^{1/2} \right.$$

$$\left. - \sup_{\theta \in B_\delta^c} \frac{D_n^{(2)}(\theta_0, \widehat{\nu})}{D_n^{(1)}(\theta)} - 2 \sup_{\theta \in B_\delta^c} \frac{|L_n(\theta, \widehat{\nu})|}{D_n^{(1)}(\theta)} - 2 \sup_{\theta \in B_\delta^c} \frac{|L_n^{(2)}(\theta_0, \widehat{\nu})|}{D_n^{(1)}(\theta)} \right].$$

Then the result follows from Wu's Lemma 5.1 and Proposition 5.1, applied on $\sup_{\theta \in B_\delta^c, \nu} L_n(\theta, \nu)[D_n^{(1)}(\theta)]^{-1}$ and $\sup_{\theta \in B_\delta^c, \nu} L_n^{(2)}(\theta, \nu)[D_n^{(1)}(\theta)]^{-1}$ when $q < \infty$. □

4. Examples

1. *Example 1.* $f_n(\theta) = m + \mu n^{-\alpha}$, $\alpha > 0$, $\theta = (m, \mu, \alpha)$. We have

$$\begin{aligned} |f_n(\theta_0) - f_n(\theta)| &\leq |m_0 - m| + |\mu_0 - \mu| n^{-\alpha_*} + |\alpha_0 - \alpha| \mu_* \ln(n) n^{-\alpha_*} \\ &\leq \|\theta_0 - \theta\|_{L^1}, n \text{ large enough} \end{aligned}$$

where (μ_*, α_*) lies between (μ_0, α_0) and (μ, α), and moreover $\inf_{\theta \in B_\delta^c} D_n(\theta) = O\left(\inf_{\|\alpha - \alpha_0\| \geq \delta} \sum_{k=1}^n k^{-2\alpha} \right)$ which is equal to $O\left(\inf_{\|\alpha - \alpha_0\| \geq \delta} \{ n^{1-2\alpha} 1_{\{2\alpha \neq 1\}} + \ln(n) 1_{\{2\alpha = 1\}} \} \right)$, implying $\lim_n \inf_{\theta \in B_\delta^c} D_n(\theta) = \infty$ if $0 \leq 2\alpha \leq 1$. Since Θ is open, then we assume $2\alpha < 1$ (the limit value $\alpha = 1/2$ is on the boundary $\overline{\Theta} \setminus \Theta$). Therefore for $2\alpha < 1$, $\lim_n (\widehat{m}_n, \widehat{\mu}_n, \widehat{\alpha}_n) \stackrel{a.s.}{=} (m_0, \mu_0, \alpha_0)$.

Now if α_0 is known then $\lim_n (\widehat{m}_n, \widehat{\mu}_n) \stackrel{a.s.}{=} (m_0, \mu_0)$ if $2\alpha_0 \leq 1$ which is reduced to the necessary and sufficient condition given in the linear model ([9]).

2. *Example 2.* Let the Bienaymé-Galton-Watson process $N_n = \sum_{i=1}^{N_{n-1}} X_{n,i}$, where the $\{X_{n,i}\}$ are i.i.d. (m_0, σ_0^2), implying $Y_n = N_n N_{n-1}^{-1/2} = m_0 N_{n-1}^{1/2} + \eta_n$. Let $\theta = m$. Then $|f_n(m_1) - f_n(m_2)| \leq \|m_1 - m_2\| N_{n-1}^{1/2}$ (Lipschitz) and $\inf_{m \in B_\delta^c} D_n(m) = (m_0 - m)^2 \sum_{k=1}^{n} N_{k-1} \overset{a.s.}{\longrightarrow} \infty$ on the nonextinction set Ω_∞, where $P(\Omega_\infty) > 0$ for $m_0 > 1$. Therefore $\lim_n \widehat{m}_n \overset{a.s.}{=} m_0$ on Ω_∞, assuming $m_0 > 1$.

Remark 4.1. (direct proof) *Recall that the CLSE of m_0 is the Harris estimator*

$$\widehat{m}_n = \frac{\sum_{k=1}^{n} N_k}{\sum_{k=1}^{n} N_{k-1}} = \frac{\sum_{k=1}^{n} (N_k m_0^{-k}) m_0^k}{\sum_{k=1}^{n} m_0^k} \frac{m_0 \sum_{k=1}^{n} m_0^{k-1}}{\sum_{k=1}^{n} (N_{k-1} m_0^{-(k-1)}) m_0^{k-1}}.$$

Then use $\lim_n N_n m_0^{-n} \overset{a.s.}{=} W$, where W is a nonnegative variable, and Toeplitz lemma. This implies $\lim_n \widehat{m}_n \overset{a.s.}{=} m_0$ on the nonextinction set Ω_∞.

Notice that the indirect proof based on Proposition 2.1 does not require the knowledge of the asymptotic behavior of the process nor an explicit expression for the estimator as it is the case in the direct proof.

3. *Example 3.* Let the size-dependent branching process $N_n = \sum_{i=1}^{N_{n-1}} X_{n,i}$, $\{X_{n,i}\}$ i.i.d. with $E(X_{n,1} | N_{n-1} = N) = m(N) = m_0 + \mu_0 N^{-\alpha_0}$, $Var(X_{n,1} | N_{n-1} = N) = \sigma^2(N) = O(N^{\beta_0})$, $\alpha_0 > 0$, $\beta_0 < 1$ known, $m_0 > 1$, $\theta = (m, \mu, \alpha)$. The asymptotic behavior of this model has been studied by Klebaner ([8]). Model M is $Y_n = N_n N_{n-1}^{-(1+\beta_0)/2}$ implying

$$f_n(\theta) = (m + \mu N_{n-1}^{-\alpha}) N_{n-1}^{1-(1+\beta_0)/2} = m N_{n-1}^{(1-\beta_0)/2} + \mu N_{n-1}^{(1-(2\alpha+\beta_0))/2}.$$

Therefore

$$\begin{aligned} \inf_{\theta \in B_\delta^c} D_n(\theta) &= O\left(\inf_{\|\alpha - \alpha_0\| \geq \delta} \sum_{k=1}^{n} N_{k-1}^{(1-(2\alpha+\beta_0))} \right) \\ &= O\left(\inf_{\|\alpha - \alpha_0\| \geq \delta} \{ m^{n((1-(2\alpha+\beta_0)))} 1_{\{2\alpha+\beta_0 \neq 1\}} + n 1_{\{2\alpha+\beta_0 = 1\}} \} \right) \end{aligned}$$

Then $\lim_n \inf_{\theta\in B_\delta^c} D_n(\theta)=\infty$ on the nonextinction set Ω_∞, for $2\alpha+\beta_0\le 1$. Moreover, for all θ_1, all θ_2, there exists (μ_*,α_*) lying between (μ_1,α_1) and (μ_2,α_2) such that on Ω_∞,

$$f_n(\theta_1)-f_n(\theta_2) = (m_1-m_2)N_{n-1}^{(1-\beta_0)/2}+(\mu_1-\mu_2)N_{n-1}^{(1-(2\alpha_*+\beta_0))/2}- (\alpha_1-\alpha_2)\mu_*\ln(N_{n-1})N_{n-1}^{(1-(2\alpha_*+\beta_0))/2}.$$

Therefore $|f_n(\theta_1)-f_n(\theta_2)|\le\|\theta_1-\theta_2\|_{L^1}N_{n-1}^{(1-\beta_0)/2}$, for n large enough since $P(\lim_n N_n=0\cup\lim_n N_n=\infty)=1$.

So since Θ is open, for $2\alpha+\beta_0<1$, we get $\lim_n(\widehat{m}_{h,n},\widehat{\mu}_{h,n},\widehat{\alpha}_{h,n})=(m_0,\mu_0,\alpha_0)$ on Ω_∞. In the same way if $\theta=(m,\mu)$, then $\lim_n(\widehat{m}_{h,n},\widehat{\mu}_{h,n})=(m_0,\mu_0)$ on Ω_∞ for $2\alpha_0+\beta_0\le 1$, while Lai and Wei's condition in this linear stochastic setting ([10]) is $2\alpha_0+\beta_0<1$.

4. Other examples in branching processes may be found in ([13], [14]) and ([5]).

5. *Example 4.* the $GARCH(p,q)$ model $\xi_n=s_n(\theta_0)U_n$, where the $\{U_n\}$ are i.i.d.$(0,1)$, $s_n(\theta_0)\ge 0$, $\mathcal{F}_{n-1}$ is generated by $\{\xi_k^2\}_{k\le n-1}$, $s_n^2(\theta_0)$ is $\mathcal{F}_{n-1}$-measurable with $B_\theta(L)(s_n^2(\theta))=A_\theta(L)(\xi_n^2)$, $B_\theta(L)=1-\sum_{j=1}^{p}\beta_jL^j$, $A_\theta(L)=\alpha_0+\sum_{j=1}^{q}\alpha_jL^j$, and L is the time lag operator, that is $s_n^2(\theta)=\alpha_0+\sum_{j=1}^{q}\alpha_j\xi_{n-j}^2+\sum_{j=1}^{p}\beta_js_{n-j}^2(\theta)$. Process $\{s_n(\theta_0)\}$ is called volatility. Then $E(\xi_n^2|\mathcal{F}_{n-1})=s_n^2(\theta_0)$ and $\{\xi_n\}$ follows a $GARCH(p,q)$ model. Assume that the observations are $\{X_n\}_{n\ge 0}$, where $X_n=\gamma_0+\xi_n$, γ_0 being an unknown parameter. The particular case $\gamma_0=0$ corresponds to the observation of $\{\xi_n\}_{n\ge 0}$. We assume here for simplification $p=1$, $q=1$, and $\alpha_{00}>0$, $\alpha_{10}>0$, $\beta_{10}>0$. By definition,

$$s_n^2(\theta)=\alpha_0+\alpha_1\xi_{n-1}^2+\beta_1s_{n-1}^2(\theta)=\alpha_0+(\alpha_1U_{n-1}^2+\beta_1)s_{n-1}^2(\theta).$$

This implies, when assuming $E(U_n^2|\mathcal{F}_{n-1})=E(U_n^2)$, that the following condition $E(\alpha_{10}U_1^2+\beta_{10})<1 \iff \alpha_{10}+\beta_{10}<1$ is a necessary condition for the stationarity of $\{s_n^2(\theta_0)\}$ and equivalently for $\{\xi_n\}$. This is the usual classical setting for deriving the asymptotic properties of any estimator of $(\alpha_{00},\alpha_{10},\beta_{10})$: under the stronger assumption of stationarity and ergodicity of the observed process, the CLSE is strongly consistent ([15]).

But here we do not assume any stationarity condition. The nonstationarity of $\{\xi_n\}$ (and therefore of $\{s_n^2(\theta_0)\}$) allows burst phenomena that are illustrated by simulations of the $GARCH(1,1)$ model with $\{U_n\}$ i.i.d., $U_n^2 \sim \exp(1)$, U_n independent of $s_n(\theta_0)$. In fig. 1 we see that as the nonstationarity increases, the magnitude of the highest burst increases which shows that appearant high nonlinearity may be modelled by nonstationary processes.

The parameter space is $\Theta =]0, c_M[\times]0, \alpha_{1M}[\times]0,1[$. Let the notations

$$\begin{aligned} Z_n(\gamma_0) &:= \xi_n^2 = s_n^2(\theta_0) + s_n^2(\theta_0)(U_n^2 - 1) =: g_n(\theta_0, \nu_0) + \sigma_n(\theta_0, \mu_0)\epsilon_n \\ \xi_n &:= X_n - \gamma_0. \end{aligned}$$

Assuming $0 < \beta_{10} < 1$, we can write

$$s_n^2(\theta) = B_\theta(L)^{-1} A_\theta(L)(\xi_n^2) = \sum_{l=0}^{\infty} \beta_1^l(\alpha_0 + \alpha_1 \xi_{n-(l+1)}^2).$$

We define 0 as the time origin for the observation of the process, that is ξ_l^2 is not observed for $l < 0$. We write $s_n^2(\theta) = g_n(\theta, \nu) = g_n^{(1)}(\theta) + g_n^{(2)}(\theta, \nu)$, where, for $n \geq 1$,

$$\begin{aligned} g_n^{(1)}(\theta) &= c + \alpha_1 \sum_{l=0}^{n-1} \beta_1^{n-(l+1)} \xi_l^2, \; c = \alpha_0(1-\beta_1)^{-1}, \theta = (c, \alpha_1, \beta_1) \\ g_n^{(2)}(\theta, \nu) &= \alpha_1 \beta_1^{n-1} \sum_{l=-1}^{-\infty} \beta_1^{-l} \xi_l^2. \end{aligned}$$

Assume that for any $0 < \beta_1 < 1$, $\sum_{l=-1}^{-\infty} \beta_1^{-l} \xi_l^2 \overset{a.s.}{<} \infty$, which means that the process starts at some finite negative time $-n_0$. Then either $\lim_n g_n^{(2)}(\theta_0, \nu_0) \overset{a.s.}{=} 0$, or $g_n^{(2)}(\theta_0, \nu_0) \overset{a.s.}{=} 0$, for all $n \geq 1$, when $\xi_k = 0$, for $k < 0$.

Let $T_n(\beta) = \sum_{l=0}^{n-1} \beta^{n-(l+1)} \xi_l^2$. Then $g_n^{(1)}(\theta_0) = c_0 + \alpha_{10} T_n(\beta_{10})$. Let $\lambda_n^{1/2} = 1 + dT_n(\beta_*)$, for some $\beta_* \in [0,1[$, $d \geq 0$ (in λ_k, $k \leq n$, d may be replaced by any bounded nonrandom $d_n \geq 0$).

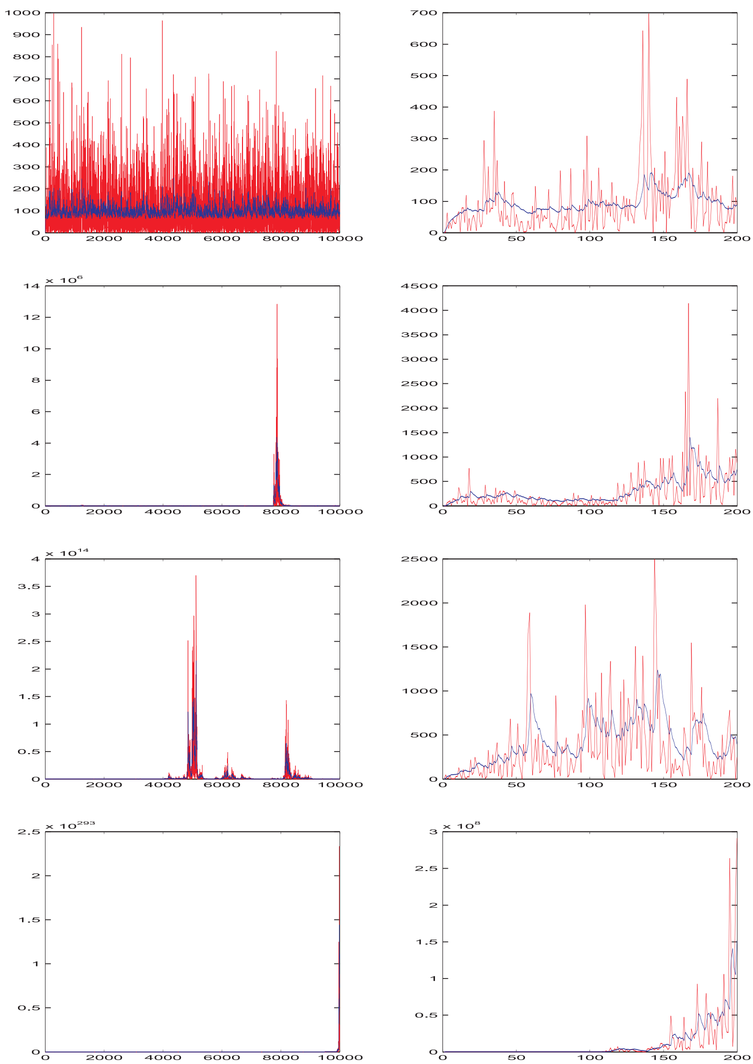

Figure 1: The most erratic line represents $\{\xi_n^2\}$ while the smoothest one represents the squared volatility $\{s_n^2(\theta)\}$. The right hand figures are zoom of the left hand figures until time 200. On the first line, the stationarity setting $\theta_0 = (10, 0.1, 0.8)$, where $\theta = (\alpha_0, \alpha_1, \beta_1)$. Then on the following lines, $\{\xi_n\}$ is nonstationary with from the second line to the last line: $\theta_0 = (10, 0.2, 0.8)$, $\theta_0 = (10, 0.22, 0.8)$, $\theta_0 = (10, 0.3, 0.8)$

Then model M_Y is defined by

$$\begin{array}{rcl} Y_n(\gamma_0) := \xi_n^2 \lambda_n^{-1/2} & =: & f_n(\theta_0, \nu_0) + \eta_n, \\ f_n(\theta_0, \nu_0) & := & f_n^{(1)}(\theta_0) + f_n^{(2)}(\theta_0, \nu_0), \\ f_n^{(1)}(\theta_0) & := & g_n^{(1)}(\theta_0)\lambda_n^{-1/2}, \\ \{f_n^{(2)}(\theta_0, \nu_0)\} & := & \{g_n^{(2)}(\theta_0, \nu_0)\lambda_n^{-1/2}\} =: \nu_0, \end{array}$$

where $E(\eta_n^2 | \mathcal{F}_{n-1}) \propto s_n^4(\theta_0)\lambda_n^{-1}$.

The CLSE of (γ_0, θ_0) is defined by:

$$(8) \quad \widehat{\gamma}_n = \arg\min_{\gamma} S_n(\gamma), \; S_n(\gamma) = \sum_{k=1}^{n} (X_k - \gamma)^2 \lambda_k^{-1/2}$$

$$(9) \quad \widehat{\theta}_n = \arg\min_{\gamma} S_{n|\widehat{\gamma},\widehat{\nu}}(\theta), \; S_{n|\widehat{\gamma},\widehat{\nu}}(\theta) = \sum_{k=1}^{n} (Y_k(\widehat{\gamma}_n) - f_k(\theta, \widehat{\nu}_k))^2,$$

where, for all k, $\widehat{\nu}_k = \widehat{g_k^{(2)}(\theta)} = 0$, that is $f_k(\theta, \widehat{\nu}_k) = f_k^{(1)}(\theta)$.

Proposition 4.1. *Let R_m be the upper mth record of $\{\xi_n\}$, that is $R_m := \xi_{L_m}^2$, where $\{L_m\}$ is the sequence of record indices, and let $\{n_k^\varepsilon\}_k := \{n : \xi_n^2 \geq \varepsilon\}$, for some $\varepsilon > 0$. Then the CLSE of (γ_0, θ_0) is strongly consistent on the set $\{\{R_m \overset{a.s.}{=} O([(L_{m+1} - L_m)m]^{1/2})\} \cap \{\overline{\lim_k}(n_{k+1}^\varepsilon - n_k^\varepsilon) \overset{a.s.}{<} \infty\}$, when $d > 0$, and on the set $\{\overline{\lim_k}(n_{k+1}^\varepsilon - n_k^\varepsilon) \overset{a.s.}{<} \infty\}$, when $d = 0$.*

P r o o f. Assume here the general nonstationary setting. Then the optimal rate of convergence is ensured under (3), that is under

$$(10) \qquad 0 \overset{a.s.}{<} \underline{\lim_n} \frac{c_0 + \alpha_{10} T_n(\beta_{10})}{1 + d T_n(\beta_*)} \leq \overline{\lim_n} \frac{c_0 + \alpha_{10} T_n(\beta_{10})}{1 + d T_n(\beta_*)} \overset{a.s.}{<} \infty.$$

Assume first $d = 0$. Then since $T_n(\beta_{10}) \geq \xi_{n-1}^2$, $\overline{\lim_n} T_n(\beta_{10}) \geq \lim_m \xi_{L_m}^2$ which is infinite on the set $\tilde{\Omega}_\infty := \left\{\lim_m \xi_{L_m}^2 = \infty\right\}$, we cannot prove the optimality in this case. Assume now $d > 0$.

On $\tilde{\Omega}_\infty^c$, (10) is checked. It remains to check (10) on $\tilde{\Omega}_\infty$. We denote, for all k, $M_{k-1} = \sup_{l \leq k-1} \xi_l^2$ and $n_k = \inf\{0 \leq l \leq k-1 : \xi_l^2 = M_{k-1}\}$, that is, there exists m such that $L_m = n_k$, for all $k : L_m < k \le L_{m+1}$ and $\xi^2_{n_k} = \xi^2_{L_m} = R_m$ (R_m is the mth record and L_m is its random index). Assume $\beta_{10} \le \beta_* < 1$. Then $\overline{\lim_n}\, T_n(\beta_{10})[T_n(\beta_*)]^{-1} \overset{a.s.}{<} 1$. Since $T_{n_k+1}(\beta) = \sum_{l=0}^{n_k} \beta^{n_k-l}\xi_l^2 = \beta \sum_{l=0}^{n_k-1} \beta^{n_k-1-l}\xi_l^2 + \xi^2_{n_k}$, then

$$\begin{aligned} \frac{T_{n_k+1}(\beta_{10})}{T_{n_k+1}(\beta_*)} &> \frac{\inf\limits_{l\le n_k-1}\{\xi_l^2\xi_{n_k}^{-2}\}\beta_{10}(1-\beta_{10}^{n_k})(1-\beta_{10})^{-1}+1}{\sup\limits_{l\le n_k-1}\{\xi_l^2\xi_{n_k}^{-2}\}\beta_*(1-\beta_*^{n_k})(1-\beta_*)^{-1}+1} \\ &> [\beta_*(1-\beta_*)^{-1}+1]^{-1} \end{aligned} \tag{11}$$

yielding to $\underline{\lim}_m\, T_{L_m+1}(\beta_{10})[T_{L_m+1}(\beta_*)]^{-1} \overset{a.s.}{>} 0$. Next for $L_m = n_k < n < L_{m+1}$, then $\xi^2_n \le \xi^2_{n_k}$, yielding

$$\frac{1+d_0T_{n+1}(\beta_{10})}{1+dT_{n+1}(\beta_*)} \ge \frac{1+d_0\beta_{10}^{n-n_k}T_{n_k+1}(\beta_{10})}{1+d\beta_*^{n-n_k}T_{n_k+1}(\beta_*)+(1-\beta_*)^{-1}\xi^2_{n_k}}, \tag{12}$$

where $d_0 := \alpha_{10}/c_0$. But (12) does not imply $\underline{\lim}_n\, T_n(\beta_{10})[T_n(\beta_*)]^{-1} \overset{a.s.}{>} 0$. But according to (10), on the set of trajectories ensuring the strong consistency of the CLSE, we can improve the estimator by first calculating $\widehat{\theta}_n^{(1)}$ with $\lambda_k^{1/2} = 1 + T_k(\beta_*)$, $k \le n$, and then calculate $\widehat{\theta}_n^{(2)}$ using $\lambda_k^{1/2} = 1 + \widehat{\alpha}_{1n}^{(1)}/\widehat{c}_n^{(1)}T_k(\widehat{\beta}_n^{(1)})$, $k \le n$, and so on. Another method is to calculate the CLSE for estimating $\{U_n\}$, then deriving the distribution of U_1 from $\{\widehat{U}_n\}$, and then calculating the MLE based on this distribution if this one belongs to a parametric family.

We have

$$S_{n|\widehat{\gamma},\widehat{\nu}}(\theta) - S_{n|\widehat{\gamma},\widehat{\nu}}(\theta_0) =$$
$$\sum_{k=1}^n [(Y_k(\widehat{\gamma}_n) - Y_k(\gamma_0) + \eta_k + (f_k(\theta_0,\nu_0) - f_k^{(1)}(\theta))]^2 -$$

$$\sum_{k=1}^{n}[Y_k(\widehat{\gamma}_n) - Y_k(\gamma_0) + \eta_k + (f_k(\theta_0,\nu_0) - f_k^{(1)}(\theta_0))]^2 =$$
$$\sum_{k=1}^{n}(f_k(\theta_0,\nu_0) - f_k^{(1)}(\theta))^2 - \sum_{k=1}^{n}(f_k(\theta_0,\nu_0) - f_k^{(1)}(\theta_0))^2 +$$
$$2\sum_{k=1}^{n}[Y_k(\widehat{\gamma}_n) - Y_k(\gamma_0) + \eta_k][(f_k^{(1)}(\theta_0)) - f_k^{(1)}(\theta))].$$

Using now $Y_k(\widehat{\gamma}_n) - Y_k(\gamma_0) = (\gamma_0 - \widehat{\gamma}_n)^2\lambda_k^{-1/2} + 2\xi_k(\gamma_0 - \widehat{\gamma}_n)\lambda_k^{-1/2}$, the previous quantity becomes

$$S_{n|\widehat{\gamma},\widehat{\nu}}(\theta) - S_{n|\widehat{\gamma},\widehat{\nu}}(\theta_0) =$$
$$D_n^{(1)}(\theta) + 2\sum_{k=1}^{n} f_k^{(2)}(\theta_0,\nu_0)d_k^{(1)}(\theta) + 2(\gamma_0 - \widehat{\gamma}_n)^2\sum_{k=1}^{n}\lambda_k^{-1/2}d_k^{(1)}(\theta) +$$
$$4(\gamma_0 - \widehat{\gamma}_n)\sum_{k=1}^{n}\xi_k\lambda_k^{-1/2}d_k^{(1)}(\theta) + 2\sum_{k=1}^{n}\eta_k d_k^{(1)}(\theta). \tag{13}$$

Then, according to Bienaymé-Tchebyshev's inequality, we get

$$\inf_{\theta\in B_\delta^c} S_{n|\widehat{\gamma},\widehat{\nu}}(\theta) - S_{n|\widehat{\gamma},\widehat{\nu}}(\theta_0) \geq$$
$$\inf_{\theta\in B_\delta^c} D_n^{(1)}(\theta)\left[1 - 2\left[\frac{\sum_{k=1}^{n}(f_k^{(2)}(\theta_0,\nu_0))^2}{\inf_{\theta\in B_\delta^c} D_n^{(1)}(\theta)}\right]^{1/2} - \right.$$
$$2(\gamma_0 - \widehat{\gamma}_n)^2\left[\frac{\sum_{k=1}^{n}\lambda_k^{-1}}{\inf_{\theta\in B_\delta^c} D_n^{(1)}(\theta)}\right]^{1/2} - 2\sup_{\theta\in B_\delta^c}\left|\frac{\sum_{k=1}^{n}\eta_k d_k^{(1)}(\theta)}{D_n^{(1)}(\theta)}\right| -$$
$$\left. 4|\gamma_0 - \widehat{\gamma}_n|\sup_{\theta\in B_\delta^c}\left|\frac{\sum_{k=1}^{n}\xi_k\lambda_k^{-1/2}d_k^{(1)}(\theta)}{D_n^{(1)}(\theta)}\right|\right]. \tag{14}$$

Assume first $\gamma_0 = \widehat{\gamma}_n = 0$, equivalent to $\{\xi_k\}_{k\geq 0}$ observed. Then $\lim_n \widehat{\theta}_n \overset{a.s.}{=} \theta_0$ under (15) and (16) thanks to proposition 3.1: for all $\delta > 0$,

$$\text{(15)} \qquad \lim_n \inf_{\theta \in B_\delta^c} \sum_{k=1}^n (g_k^{(1)}(\theta_0) - g_k^{(1)}(\theta))^2 \lambda_k^{-1} \overset{a.s.}{=} \infty,$$

$$\text{(16)} \qquad \overline{\lim}_n \frac{\sum_{k=1}^n (g_k^{(2)}(\theta_0, \nu_0))^2 \lambda_k^{-1}}{\inf_{\theta \in B_\delta^c} \sum_{k=1}^n (g_k^{(1)}(\theta_0) - g_k^{(1)}(\theta))^2 \lambda_k^{-1}} \overset{a.s.}{=} 0.$$

The first condition, which is $SI(\{D_n^{(1)}(\theta)\})$, is satisfied if $\lim_n \inf_{|c-c_0|\geq\delta} D_n^{(1)}(c, \alpha_{10}, \beta_{10}) \overset{a.s.}{=} \infty$, equivalent to $\lim_n \sum_{k=1}^n \lambda_k^{-1} \overset{a.s.}{=} \infty$, if $\lim_n \inf_{|\beta_1-\beta_{10}|\geq\delta} D_n^{(1)}(c_0, \alpha_{10}, \beta_1) \overset{a.s.}{=} \infty$, and if $\lim_n \inf_{|\alpha_1-\alpha_{10}|\geq\delta} D_n^{(1)}(c_0, \alpha_1, \beta_{10}) \overset{a.s.}{=} \infty$. First of all

$$\sum_{k=1}^n \lambda_k^{-1} = \sum_{k=1}^n \left[1 + d \sum_{l=0}^{k-1} \beta_*^{k-(l+1)} \xi_l^2\right]^{-2} \geq \sum_{k=1}^n \left[1 + d(1-\beta_*)^{-1} \xi_{n_k}^2\right]^{-2}.$$

Therefore $\sum_{k=1}^n \lambda_k^{-1} \overset{a.s.}{=} \infty$ for $d = 0$, and for $d > 0$ either on $\tilde{\Omega}_\infty^c$, or on $\tilde{\Omega}_\infty$ using $\sum_k \lambda_k^{-1} \geq \sum_m (L_{m+1} - L_m)[1 + d(1-\beta_*)^{-1} \xi_{L_m}^2]^{-2}$, if the sequence of records satisfies

$$\text{(17)} \quad R_m := \xi_{L_m}^2 \overset{a.s.}{=} O([(L_{m+1} - L_m)m]^{1/2}), \text{ on } \tilde{\Omega}_\infty, \text{ for } d > 0.$$

In the standard extreme value theory for i.i.d. variables, L_m behaves asymptotically as $\exp m$ and is independent of the distribution of these variables. In a general $GARCH(1,1)$ model, the distribution of the records indices and values $\{(L_m, R_m)\}$ may be calculated from the distribution of $\{U_n\}$, when given, and $\{s_n(\theta)\}$.

Next, consider $\lim_n \inf_{|\beta_1-\beta_{10}|\geq\delta} D_n^{(1)}(c_0, \alpha_{10}, \beta_1) \overset{a.s.}{=} \infty$. Define, for all $\varepsilon > 0$

small enough, the subsequence of indices $\{n_k^\varepsilon\}_k = \{n : \xi_n^2 \geq \varepsilon\}$. Then

$$\begin{aligned} D_n^{(1)}(c_0, \alpha_{10}, \beta_1) &\geq \sum_{n_k^\varepsilon \leq n-2} \frac{\left[\sum_{l=0}^{n_k^\varepsilon+1} (\beta_{10}^{n_k^\varepsilon+2-(l+1)} - \beta_1^{n_k^\varepsilon+2-(l+1)})\xi_l^2 \xi_{n_k^\varepsilon}^{-2}\right]^2}{\left[(1 + d \sum_{l=0}^{n_k^\varepsilon+1} \beta_*^{n_k^\varepsilon+2-(l+1)} \xi_l^2)\xi_{n_k^\varepsilon}^{-2}\right]^2} \\ &\geq \sum_{n_k^\varepsilon \leq n-2} \frac{(\beta_{10} - \beta_1)^2}{[\varepsilon^{-2} + d(1 - \beta_*)^{-1} \sup_{l \leq n_k^\varepsilon+1} \xi_l^2 \varepsilon^{-2}]^2} \end{aligned}$$

implying $\lim_n \inf_{|\beta_1 - \beta_{10}| \geq \delta} D_n^{(1)}(c_0, \alpha_{10}, \beta_1) \overset{a.s.}{=} \infty$ on $\tilde{\Omega}_\infty^c$ for $d = 0$, and for $d > 0$ on $\tilde{\Omega}_\infty^c$ under (18):

$$\text{(18)} \qquad \exists \varepsilon > 0 : \lim_n \sum_{k=1}^n 1_{\{\xi_k^2 \geq \varepsilon\}} \overset{a.s.}{=} \infty \text{ on } \tilde{\Omega}_\infty^c, \text{ for } d > 0,$$

which means that $\{\xi_k^2\}$ does not die out as $k \to \infty$; (18) is satisfied under the assumption that the $\{U_n\}$ are i.i.d. with a continuous distribution. Consider next $\tilde{\Omega}_\infty$. We have, in the same way,

$$\begin{aligned} D_n^{(1)}(c_0, \alpha_{10}, \beta_1) &\geq \sum_{n_k \leq n-2} \frac{\left[\sum_{l=0}^{n_k+1} (\beta_{10}^{n_k+2-(l+1)} - \beta_1^{n_k+2-(l+1)})\xi_l^2 \xi_{n_k}^{-2}\right]^2}{\left[(1 + d \sum_{l=0}^{n_k+1} \beta_*^{n_k+2-(l+1)} \xi_l^2)\xi_{n_k}^{-2}\right]^2} \\ &\geq \sum_{n_k : n_k \leq n-2, \xi_{n_k+1}^2 \leq \xi_{n_k}^2} \frac{(\beta_{10} - \beta_1)^2}{[\xi_{n_k}^{-2} + d(1 - \beta_*)^{-1}]^2} \end{aligned}$$

which converges a.s. to ∞ on $\tilde{\Omega}_\infty$ for $d \geq 0$. The proof is similar for $\lim_n \inf_{|\alpha_1 - \alpha_{10}| \geq \delta} D_n^{(1)}(c_0, \alpha_1, \beta_{10}) \overset{a.s.}{=} \infty$.

Study now (16), assuming (15). This condition is satisfied if $\{\xi_k^2\}_{k<0}$ is null.

Otherwise, we have

$$\frac{\sum\limits_{k=1}^{n}(g_k^{(2)}(\theta_0))^2\lambda_k^{-1}}{\sum\limits_{k=1}^{n}(g_k^{(1)}(\theta_0)-g_k^{(1)}(\theta))^2\lambda_k^{-1}} \leq \frac{\sup\limits_{k\leq N}(g_k^{(2)}(\theta_0))^2\sum\limits_{k=1}^{N}\lambda_k^{-1}}{\sum\limits_{k=1}^{n}(g_k^{(1)}(\theta_0)-g_k^{(1)}(\theta))^2\lambda_k^{-1}} + \sup_{N<k\leq n}(g_k^{(2)}(\theta_0))^2\frac{\sum\limits_{k=1}^{n}\lambda_k^{-1}}{\sum\limits_{k=1}^{n}(g_k^{(1)}(\theta_0)-g_k^{(1)}(\theta))^2\lambda_k^{-1}}.$$

Since $\lim\limits_n g_n^{(2)}(\theta_0) \overset{a.s.}{=} 0$, then (16) is checked if

$$\underline{\lim}_n Q_n \overset{a.s.}{>} 0,\ Q_n := \frac{\inf\limits_{\theta\in B_\delta^c}\sum\limits_{k=1}^{n}(g_k^{(1)}(\theta_0)-g_k^{(1)}(\theta))^2\lambda_k^{-1}}{\sum\limits_{k=1}^{n}\lambda_k^{-1}}. \tag{19}$$

Then (19) is checked if $\underline{\lim}_k \inf_\theta (g_k^{(1)}(\theta_0)-g_k^{(1)}(\theta))^2 \overset{a.s.}{>} 0$. Defining $k_k^\varepsilon = \sup\{l \leq k-1 : \xi_l^2 \geq \varepsilon\} = \sup\{n_l^\varepsilon \leq k-1\}$, this is checked if

$$\underline{\lim}_k \inf_{|\beta_1-\beta_{10}|\geq\delta}\left(\beta_1^{k-(k_k^\varepsilon+1)}-\beta_{10}^{k-(k_k^\varepsilon+1)}\right)^2 \overset{a.s.}{>} 0$$

which is itself satisfied when

$$\overline{\lim}_k (n_{k+1}^\varepsilon - n_k^\varepsilon) \overset{a.s.}{<} \infty. \tag{20}$$

This means that the durations of the time periods during which $\{\xi_k^2\}$ is null or converges to 0 are bounded. Since, for $d > 0$, this assumption is stronger than (18), $\lim\limits_n \widehat{\theta}_n \overset{a.s.}{=} \theta_0$ under (15) and (16), or under the stronger condition (17) with (20) (when $\{\xi_k^2\}_{k<0}$ may be nonnull) or with (18) (when $\{\xi_k^2\}_{k<0}$ is null).

Assume next that $\{X_k\}$ is observed. Then according to proposition 2.1, $\lim_n \widehat{\gamma}_n \overset{a.s.}{=} \gamma_0$ if

$$\lim_n \sum_{k=1}^{n}\lambda_k^{-1/2} \overset{a.s.}{=} \infty \tag{21}$$

which is checked under (17). So the consistency of $\{\widehat{\theta}_n\}$ requires (15), (16), (21) (all satisfied under (17) and (20)) and the following assumptions (22) and (23) coming from (14): for all $\delta > 0$,

$$\text{(22)} \qquad \underline{\lim}_n \left[\inf_{\theta \in B_\delta^c} \sum_{k=1}^n [g_k^{(1)}(\theta_0) - g_k^{(1)}(\theta)]^2 \lambda_k^{-1} \right] \left[\sum_{k=1}^n \lambda_k^{-1} \right]^{-1} \overset{a.s.}{>} 0,$$

$$\text{(23)} \qquad \overline{\lim}_n \sup_{\theta \in B_\delta^c} \left| \left[\sum_{k=1}^n \xi_k \lambda_k^{-1/2} d_k^{(1)}(\theta) \right] \left[\sum_{k=1}^n [d_k^{(1)}(\theta)]^2 \right]^{-1} \right| \overset{a.s.}{<} \infty.$$

Since $\lambda_k \geq 1$, then $\lambda_k^{-1} \leq 1$ leading to (23) according to the SLLNSM (proposition 5.1), and (22) is (19) which is checked under (20).

So finally under (17) and (20), $\lim_n(\widehat{\gamma}_n, \widehat{\theta}_n) \overset{a.s.}{=} (\gamma_0, \theta_0)$. The proof is similar if more generally $\{\xi_n\}$ is the innovation of a linear autoregressive process. □

5. Strong Law of Large Numbers for SubMartingales

Proposition 5.1. *Let $\widetilde{\Theta} \subset \mathbb{R}^p$, $\overline{\widetilde{\Theta}}$ compact, $p < \infty$. Let $\{\mathcal{F}_k\}$ be an increasing sequence of σ-algebra, and $L_n(\theta) = \sum_{k=1}^n \eta_k d_k(\theta)$, $\theta \in \widetilde{\Theta}$, where, for all k, η_k is $\mathcal{F}_k$-measurable with $E(\eta_k|\mathcal{F}_{k-1}) = 0$, $E(\eta_k^2|\mathcal{F}_{k-1}) = \sigma_k^2$, $\overline{\lim}_k \sigma_k^2 \overset{a.s.}{<} \infty$, and $d_k(\theta)$ is $\mathcal{F}_{k-1}$-measurable. For all k, n, let $d_{*k}(\theta)$ be $\mathcal{F}_{k-1}$-measurable, $D_{*n}(\theta) = \sum_{k=1}^n d_{*k}^2(\theta)$, $D_n(\theta) = \sum_{k=1}^n d_k^2(\theta)$. Assume that there exists $\Omega_\infty \subset \Omega$ with $P(\Omega_\infty) > 0$ and such that on Ω_∞, $\overline{\lim}_n \sup_\theta D_n(\theta)[D_{*n}(\theta)]^{-1} \overset{a.s.}{<} \infty$, $SI(\{D_{*n}(\theta)\})$, $LIP(\{d_k(\theta)\})$, and $LIP(\{d_{*k}(\theta)\})$ are checked. Then*

$$\text{(24)} \qquad \lim_n \sup_{\theta \in \widetilde{\Theta}} |L_n(\theta)| [D_{*n}(\theta)]^{-1} \overset{a.s.}{=} 0 \text{ on } \Omega_\infty.$$

P r o o f. If $\widetilde{\Theta}$ is a finite set (i.e. $Card\, \widetilde{\Theta} < \infty$), thanks to the SLLNM (Strong Law of Large Numbers for Martingales, th. 2.18, [4]),

$$\lim_n \sup_{\theta \in B_\delta^c} \left| \left[\sum_{k=1}^n \eta_k d_k(\theta) \right] [D_n(\theta)]^{-1} \right| \overset{a.s.}{=} 0 \text{ on } \left\{ \lim_n \inf_{\theta \in B_\delta^c} D_n(\theta) \overset{a.s.}{=} 0 \right\}.$$

Now assume the general case $\widetilde{\Theta} \subset \mathbb{R}^p$. We define for each k a discretization of $\mathbb{R}^p$ by a random grid G_k with fixed directions, a fixed origin, and a random mesh size ϵ_k $\mathcal{F}_{k-1}$-measurable and converging a.s. to 0 sufficiently rapidly as $k \to \infty$. Let $\{\theta_{k,i}\}_i$ be the vertices of $G_k \cap \widetilde{\Theta}$, let $\theta_k(\theta) \in \{\theta_{k,i}\}_i$ such that $\|\theta_k(\theta) - \theta\| \le c\epsilon_k$, where c is a constant (take $c = \sqrt{p}$ if the norm in condition $LIP(\{d_k(\theta)\})$ is the euclidean norm). Then, denoting $\mathcal{G}_n(\theta) := \{\theta_k(\theta)\}_{k \le n}$, we get

$$\begin{aligned}
\overline{\lim_n} \sup_\theta \frac{|L_n(\theta)|}{D_{*n}(\theta)} &\le \overline{\lim_n} \sup_\theta \frac{|L_n(\theta, \mathcal{G}_n(\theta))|}{D_{*n}(\theta)} + \overline{\lim_n} \sup_\theta \frac{|L_n(\mathcal{G}_n(\theta))|}{D_{*n}(\theta)} \\
L_n(\theta, \mathcal{G}_n(\theta)) &:= \sum_{k=1}^n \eta_k (d_k(\theta) - d_k(\theta_k(\theta))), \\
L_n(\mathcal{G}_n(\theta)) &:= \sum_{k=1}^n \eta_k d_k(\theta_k(\theta)).
\end{aligned}$$

Defining $D_n(\mathcal{G}_n(\theta)) := \sum_{k=1}^n [d_k(\theta_k(\theta))]^2$, $D_{*n}(\mathcal{G}_n(\theta)) := \sum_{k=1}^n [d_{*k}(\theta_k(\theta))]^2$,

$$\sup_\theta \frac{|L_n(\mathcal{G}_n(\theta))|}{D_{*n}(\theta)} \le \sup_\theta \frac{|L_n(\mathcal{G}_n(\theta))|}{D_{*n}(\mathcal{G}_n(\theta))} \left[\sup_\theta \frac{|D_{*n}(\mathcal{G}_n(\theta)) - D_{*n}(\theta)|}{D_{*n}(\theta)} + 1 \right] \tag{25}$$

which converges a.s. to 0 thanks to the SLLNM ([4]) and $LIP(\{d_k(\theta)\})$.

Next we must show that $\lim_n \sup_\theta |L_n(\theta, \mathcal{G}_n(\theta))| [D_{*n}(\theta)]^{-1} \overset{a.s.}{=} 0$. For that the successive steps are the following ones. We define $U_{m,n}(\theta, \mathcal{G}_n(\theta)) = \sum_{k=m}^n \eta_k (d_k(\theta) - d_k(\theta_k(\theta)))[D_k(\theta)]^{-1}$. Then we use the following property for submartingales (th. 2.1, [4], see also th. 5.1 further):

$$\lambda P \left(\max_{n: m \le n \le m'} \sup_\theta |U_{m,n}(\theta, \mathcal{G}_n(\theta))| > \lambda \right) \le E \left(\sup_\theta |U_{m,m'}(\theta, \mathcal{G}_{m'}(\theta))| \right). \tag{26}$$

Using $LIP(\{d_k(\theta)\})$ and the convergence of $\{\epsilon_k\}$ to 0 fast enough, we show

$$\lim_m \lim_{m'} E \left(\sup_\theta |U_{m,m'}(\theta, \mathcal{G}_{m'}(\theta))| \right) = 0$$

from which we deduce $\lim_m \sup_{n \ge m} \sup_\theta |U_{m,n}(\theta, \mathcal{G}_n(\theta))| \overset{P}{=} 0$ thanks to (26), and then $\lim_m \overline{\lim_{n \ge m}} \sup_\theta |U_{m,n}(\theta, \mathcal{G}_n(\theta))| \overset{a.s.}{=} 0$. Finally the result will follow from

$$|L_n(\theta, \mathcal{G}_n(\theta))| \, [D_{*n}(\theta)]^{-1} = \left| \sum_{k=1}^n U_{k,n}(\theta, \mathcal{G}_n(\theta)) d_k^2(\theta) \right| [D_{*n}(\theta)]^{-1}$$

together with a generalized Toeplitz's lemma applied on the $\sup_\theta$ of this quantity. □

Lemma 5.1. (Wu's lemma (1981), [20]) *If for all* $\delta > 0$, $\underline{\lim}_n \inf_{\theta \in B_\delta^c} (S_n(\theta) - S_n(\theta_0)) \overset{a.s.(P.)}{>} 0$, *then* $\lim_n \widehat{\theta}_n \overset{a.s.(P)}{=} \theta_0$.

Theorem 5.1. (Theorem 2.1, [4]) *If* $\{S_i, \mathcal{F}_i, 1 \le i \le n\}$ *is a submartingale, then for each real* λ, $\lambda P\left(\max_{i\le n} S_i > \lambda\right) \le E\left[S_n 1_{\{\max_{i\le n} S_i > \lambda\}}\right]$.

Acknowledgments

The author is grateful to V. Vatutin and A. Zubkov from Steklov Institute (Moscow) for their fruitful comments.

REFERENCES

[1] T. W. Anderson, J. B. Taylor. Strong consistency of least squares estimates in dynamic models. *Ann. Statist.* **7** (1979), 484–489.

[2] F. Comte, O. Lieberman. Asymptotic theory for multivariate *GARCH* processes. *J. Multiv. Anal.* **84**, No 1 (2003), 61–84.

[3] R. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, No 4 (1982), 987–1007.

[4] P. Hall, C. C. Heyde. Martingale limit theory and its application. Probability and Mathematical Statistics, New York, 1980.

[5] C. Jacob, N. Lalam, N. Yanev. Statistical inference for processes depending on exogenous inputs and application in regenerative processes. *Pliska Stud. Math. Bulg.* **17** (2005), 109–136

[6] C. Jacob. Conditional Least Squares Estimation in nonlinear and nonstationary regression models: asymptotic properties and examples. *Technical Report.*, UR341, INRA, Jouy-en-Josas, 2008.

[7] R. I. Jennrich. Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.* **40** (1969), 633–643.

[8] F. C. Klebaner. On population-size dependent branching processes. *Adv. in Appl. Probab.* **16** (1984), 30–55.

[9] T. L. Lai, H. Robbins, C. Z. Wei. Strong consistency of estimates in multiple regression II, *J. Multivariate Anal.* **9** (1979), 343–361.

[10] T. L. Lai, C. Z. Wei. Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems, *Ann. Statist.* **10** (1982), 154–166.

[11] T. L. Lai, C. Z. Wei. Asymptotic properties of general autoregressive models and strong consistency of least-squares estimates of their parameters. *J. Multivariate Anal.* **13** (1983), 1–23.

[12] T. L. Lai. Asymptotic properties of nonlinear least squares estimates in stochastic regression models. *Ann. Statist.* **22** (1994), 1917–1930.

[13] N. Lalam, C. Jacob. Estimation of the offspring mean in a supercritical or near-critical size-dependent branching process. *Adv. in Appl. Probab.* **36**, No 2 (2004), 582–601.

[14] N. Lalam, C. Jacob, P. Jagers. Modelling of the PCR amplification process by a size-dependent branching process and estimation of the efficiency. *Adv. in Appl. Probab.* **36**, No 2 (2004), 602–615.

[15] J. Ngatchou-Wandji. Estimation in a class of nonlinear heteroscedastic time-series models. *Electronical Journal of Statistics* **2** (2008), 40–62.

[16] P. Posedel. Properties and Estimation of $GARCH(1,1)$ Model. *Metodoloski zvezki* **2**, No 2 (2005), 243–257.

[17] K. Skouras. Strong consistency in nonlinear stochastic regression models. *Ann. Statist.* **28** (2000), 871–879.

[18] C. Z. Wei. Asymptotic properties of least-squares estimates in stochastic regression models. *Ann. Statist.* **13** (1985), 1498–1508.

[19] C. Z. Wei. Convergence rates for the critical branching process with immigration. *Statist. Sinica* **1** (1991), 175–184.

[20] C. Z. Wei. Asymptotic theory of nonlinear least squares estimation. *Ann. Statist.* **9** (1981), 501–513.

[21] J. F. Yao, On least squares estimation for stable nonlinear AR processes. *Ann. Inst. Statist. Math.* **52** (2000), 316–331.

Christine Jacob
Mathematics and Informatics unit
UR341, INRA
F-78352 Jouy-en-Josas, France
e-mail: `christine.jacob@jouy.inra.fr`