Geometric ergodicity, regularity of the invariant distribution and inference for a threshold bilinear Markov process

Marco Ferrante
Dip. di Matematica Pura ed Appl.
Università degli Studi di Padova
via Belzoni 7
35131 Padova, Italy
E-Mail: ferrante@math.unipd.it

Giovanni Fonseca
Dipartimento di Economia
Università degli Studi dell'Insubria
Via Ravasi, 2
21100 Varese, Italy
E-Mail: gfonseca@eco.uninsubria.it

Paolo Vidoni
Dipartimento di Scienze Statistiche
Università degli Studi di Udine
via Treppo, 18
33100 Udine , Italy
E-Mail: vidoni@dss.uniud.it

Abstract

In this paper we consider a first order threshold bilinear Markov process, which can be viewed as an AR model with ARCH-type errors and may be useful for modelling economic or financial time series. We study the main features of this process within a wider family of nonlinear models, where the threshold term is replaced by a smooth approximating function. Under suitable general assumptions, we provide sufficient conditions for the geometric ergodicity of the processes of this class and for the existence of their finite moments of a given order. Furthermore, we state regularity conditions for the invariant measures and we prove that the invariant measures of the smooth models weakly converge to that of the threshold one. The problem of estimating the parameters, including the threshold parameter, is studied and a simple semiparametric procedure based on the theory of optimal estimating functions is proposed.

Keywords: Autoregressive conditional heteroskedasticity; estimating function; invariant probability; stationary process; threshold model.

AMS 2000 SCI: 60J05, 62M05

1 Introduction

In this paper we study the first order threshold bilinear Markov process

$$X_t = aX_{t-1} + \left(b_1 1_{\{X_{t-1}<c\}} + b_2 1_{\{X_{t-1}\geq c\}}\right) X_{t-1}e_t + de_t, \tag{1.1}$$

$t \in I\!N^+ = I\!N \setminus \{0\}$, where $\{e_t, t \in I\!N^+\}$ is a sequence of independent identically distributed (i.i.d.) absolutely continuous random variables, X_0 is a given random variable independent of $\{e_t, t \in I\!N^+\}$, a, b_1, b_2, c, and d are real numbers. This process can be viewed as an autoregressive (AR) process with autoregressive conditionally heteroscedastic (ARCH) errors, similar to those introduced in a more general framework by Weiss (1984), and it is the simplest non-trivial way to mix together the bilinear and the threshold models. For this reason, the mathematical interest in such a new model is evident and we believe that there is an empirical interest as well, since (1.1) is a new element in the class of AR processes with ARCH-type errors. More precisely, this process combines the usual AR part, which targets on the conditional mean of X_t given the past, with a description for the conditional variance of X_t given the past, similar to that one given in the TARCH (threshold ARCH) models proposed by Rabemananjara and Zakoian (1993) (see also Liu, Li and Li (1996)). Here, alternative regimes for the conditional variance are allowed (in order to account for asymmetries) and the non-negativity constraint on the parameters is relaxed (in order to introduce new nonlinear time paths). The present model is analytically tractable and it may be considered as a prototype for a new larger class of processes useful for modelling economic or financial time series.

Process (1.1) is clearly discontinuous, since it involves two different bilinear sub-models, namely two alternative regimes, with change point specified by the threshold parameter c. It is convenient, in order to provide a useful inferential procedure, to introduce, as in Chan and Tong (1986), a class of processes encompassing (1.1), where the switching between the two regimes takes place in a smoother way. The main properties of (1.1) are studied within this wider family of nonlinear models.

The paper is organized as follows. In Section 2, we obtain sufficient conditions for the geometric ergodicity and for the existence of finite moments of process (1.1) and of processes belonging to the associated family of smooth models. In order to apply the drift-criteria of Meyn and Tweedie (1993), we prove a general result on irreducibility and T-continuity for first-order nonlinear discontinuous processes. For nonlinear models that are not irreducible and T-continuous, or for which it is too difficult to prove these two properties, results in Fonseca and Tweedie (2002) may be applied in order to obtain sufficient conditions for the existence of a stationary measure. In Section 3, we provide sufficient conditions so that the invariant probability measures of the processes turn out to be absolutely continuous with respect to Lebesgue measure. Indeed, we prove the weak convergence of the invariant measures of the smooth models to that of the threshold one. Hence, the first order threshold bilinear Markov process (1.1) may be approximated within this family of smooth nonlinear models. This result will be considered in the final section in order to define a useful estimation procedure for the parameters, including the threshold coefficient, based on the theory of optimal estimating functions.

2 Sufficient conditions for geometric ergodicity

The aim of the present section is to derive sufficient conditions, on the coefficients of the equation (1.1) and of similar systems, for the existence of a stationary solution process and of its moments of order p. Throughout this section, we use the notation adopted by Meyn and Tweedie (1993).

It is convenient to rewrite the model (1.1) in the equivalent form

$$X_t = aX_{t-1} + \left(b_1 + b_3 1_{\{X_{t-1} \geq c\}}\right) X_{t-1}e_t + de_t, \tag{2.1}$$

$t \in I\!N^+$, where $\{e_t, t \in I\!N^+\}$, X_0, a, b_1, c and d are defined as previously and the real parameter $b_3 = b_2 - b_1 \neq 0$. Furthermore, as in Chan and Tong (1986), we introduce a class of models where the switching between the regimes takes place in a smoother way. More precisely, letting $\Phi(\cdot)$ denote the standard normal distribution function, we define the AR model with STARCH (Smooth TARCH) errors by

$$X_t = aX_{t-1} + \Big\{b_1 + b_3\Phi\Big(\frac{X_{t-1} - c}{z}\Big)\Big\}X_{t-1}e_t + de_t, \tag{2.2}$$

$t \in I\!N^+$, with $\{e_t, t \in I\!N^+\}$, X_0, a, b_1, b_3, c and d defined previously and $z > 0$. The parameter z is called smoothing parameter, since the larger is z the smoother is the switching between the two bilinear sub-models. Instead of $\Phi(\cdot)$, it is possible to consider different smoothing functions, such as the logistic distribution function.

Since models (2.1) and (2.2) are quite similar, we present a general result and we obtain the conditions for (2.1) and (2.2) as a simple corollary. Let us consider the model

$$X_t = f(X_{t-1}) + g(X_{t-1})e_t, \tag{2.3}$$

where $t \in I\!N^+$, $\{e_t, t \in I\!N^+\}$ is a sequence of i.i.d. random variables and X_0 is a given random variable. Hereafter, given h a real function, we define $N_h = \{x \in I\!R : \liminf_{x_n \to x} |h(x_n)| = 0 \text{ or } h(x) = 0\}$, we denote by D_h the set of its discontinuities and by $h^{(k)}(\,\cdot\,)$ the k-times composition of h with itself. It is immediate to see that if $x \in N_h \setminus D_h$ then $h(x) = 0$.

We consider the following assumptions:

(H1) f, g are locally bounded real functions, the sets N_g and $D_g \cup D_f$ are countable without finite limit points and disjoint, and $\forall\, x \in N_g,\ \exists\, k \in I\!N$ such that $f^{(k)}(x) \notin N_g \cup D_g \cup D_f$ and $\forall\, j < k,\ f^{(j)}(x) \in N_g$;

(H2) e_1 is absolutely continuous with respect to Lebesgue measure λ, with density $p(\cdot)$ strictly positive almost everywhere and lower semicontinuous (l.s.c.).

Remark 2.1 Assumption (H1) states that functions f and g are piecewise continuous and that $|g|$ is bounded away from zero in a neighborhood of any discontinuity. Moreover, if (2.3) starts from $x \in N_g$, the solution X_t cannot be a deterministic process (i.e. $X_t = f^{(t)}(x)\ \forall t$), since after a finite number of steps $f^{(k)}(x) \notin N_g$. (H1) is fulfilled for (2.1) and (2.2) under the mild conditions that $|a| \neq 1$, $d \neq 0$, $(b_1 + b_3)c + d \neq 0$ and $b_1 c + d \neq 0$.

Proposition 2.1 *Under (H1) and (H2), the process solution to (2.3) is a λ-irreducible, aperiodic T-chain.*

Proof: By (H1) it is straightforward to prove the aperiodicity of the Markov chain $\{X_t, t \in I\!N\}$, solution to (2.3). As regards irreducibility, we consider an arbitrary set $A \in \mathcal{B}(I\!R)$, where $\mathcal{B}(I\!R)$ is the Borel σ-field, such that $\lambda(A) > 0$. We denote by $P(x, A)$ and $P^k(x, A)$, $k \in I\!N^+$, the transition and the k-th step transition probability of $\{X_t, t \in I\!N\}$. If $x \notin N_g$, then $P(x, A) > 0$. Otherwise, by (H1), there exists $k < +\infty$ such that $P^{k+1}(x, A) = P(f^{(k)}(x), A) > 0$. Therefore $\sum_{i=1}^{\infty} P^i(x, A) 2^{-(i+1)} > 0$ and the Markov chain is λ-irreducible.

We prove now that the process $\{X_t, t \in I\!N\}$, solution to (2.3), is a T-chain. By Proposition 6.2.4 in Meyn and Tweedie (1993), it will be sufficient to show that for each $x \in I\!R$, there exists a $k \in I\!N^+$ and a non-trivial substochastic transition kernel $T_x(\cdot, \cdot)$, l.s.c. in the first variable, such that $P^k(y, A) \geq T_x(y, A)$ for each $y \in I\!R$ and $A \in \mathcal{B}(I\!R)$. For future convenience, we state a result which will be widely used in the following:

(LSC) the product of l.s.c. functions is a l.s.c. function and the composition of a l.s.c. with a continuous function is l.s.c.

For $x \in I\!R$, assume first that $x \notin N_g \cup D_g \cup D_f$. By (H1), we may define $I_\varepsilon(x) = (x - \varepsilon, x + \varepsilon)$ with $\varepsilon > 0$, such that for every $y \in I_\varepsilon(x)$, $y \notin N_g \cup D_g \cup D_f$. Then, by (LSC) and Fatou's Lemma we have that, for every $y \in I_\varepsilon(x)$,

$$P(y, A) = \int_A p(y, z) dz = \int_A \frac{1}{|g(y)|}\, p \left\{ \frac{z - f(y)}{g(y)} \right\}\, dz$$

is a positive l.s.c. function whenever $\lambda(A) > 0$. Therefore, taking $\varphi(\,\cdot\,)$ as a smooth positive function equal to 1 on $I_{\frac{\varepsilon}{2}}(x)$ and 0 outside $I_\varepsilon(x)$, we define, for every $y \in I\!R$, the function

$T_x(y,A) = \varphi(y)P(y,A)$, which is a l.s.c. non-trivial substochastic transition kernel. If we take $x \in D_f \cup D_g$, by (H1), $x \notin N_g$ and there exists $I_\varepsilon(x)$ such that for every $y \in I_\varepsilon(x)$ with $y \neq x$, we have that $y \notin N_g \cup D_g \cup D_f$. By the assumptions, $\liminf_{y\to x} P(y,A) \geq \int_A \liminf_{y\to x} p(y,z)\ dz = \int_A h(x,z)\ dz > 0$ whenever $\lambda(A) > 0$. Taking $\varphi(\,\cdot\,)$ as above we define, for each $y \in I\!R$, $T_x(y,A) = \varphi(y)\int_A \min\{p(y,z), h(x,z)\}dz$, again a l.s.c. non-trivial substochastic transition kernel. Finally, we consider the case with $x \in N_g$ and, for simplicity, assume that $f^{(k)}(x) \notin N_g \cup D_g \cup D_f$ with $k = 1$ (the case with $k > 1$ follows by similar arguments). By (H.1), $x \notin D_g \cup D_f$ and there exists $I_{\varepsilon_0}(x)$ such that for every $y \in I_{\varepsilon_0}(x)$, with $y \neq x$, we have that $y \notin N_g \cup D_g \cup D_f$. Moreover, there exist $I_{\varepsilon_1}(x) \subset I_{\varepsilon_0}(x)$ and $\delta_1 > 0$ such that $f(y) + g(y)e_1 \notin N_g \cup D_g \cup D_f$, for every $y \in I_{\varepsilon_1}(x)$ and $|e_1| < \delta_1$. Hence, there exists $C > 0$ such that, for every $y \in I_{\varepsilon_1}(x)$ and $A \in \mathcal{B}(I\!R)$,

$$\begin{aligned}
P^2(y,A) &= P[f\{f(y)+g(y)e_1\} + g\{f(y)+g(y)e_1\}e_2 \in A] \\
&= \int_{I\!R}\left[\int_A \frac{1}{|g(f(y)+g(y)s_1)|} p\left\{\frac{s_2 - f(f(y)+g(y)s_1)}{g(f(y)+g(y)s_1)}\right\} ds_2\right] p(s_1)\, ds_1 \\
&\geq \int_A \left[\int_{(-\delta_1,\delta_1)} \frac{1}{|g(f(y)+g(y)s_1)|} p\left\{\frac{s_2 - f(f(y)+g(y)s_1)}{g(f(y)+g(y)s_1)}\right\} p(s_1)\, ds_1\right] ds_2 \\
&\geq C\int_A \left[\int_{(-\delta_1,\delta_1)} p\left\{\frac{s_2 - f(f(y)+g(y)s_1)}{g(f(y)+g(y)s_1)}\right\} p(s_1)\, ds_1\right] ds_2 = \widetilde{T}(y,A) \quad .
\end{aligned}$$

By (LSC), $\widetilde{T}(y,A)$ is a positive l.s.c. function and, as above, we can define the real function $T_x(y,A) = \varphi(y)\widetilde{T}(y,A)$, a l.s.c. nontrivial substochastic transition kernel. The proof is now complete. □

Remark 2.2 It is interesting to note that the previous result is not a consequence of the general conditions proved by Cline and Pu (1998), since the hypothesis of local boundedness of $1/g(\cdot)$ is not fulfilled. By its generality, the previous proposition could be interesting by itself, although it requires much more regularity on the functions involved.

Now assume that $f(x) = \alpha x$ and $g(x) = \beta(x)x + \delta$, with $\alpha, \delta \in I\!R$ and $\beta(\cdot)$ a suitable bounded real function. We provide sufficient conditions for the geometric ergodicity of the solution process $\{X_t, t \in I\!N\}$ to (2.3), and for the existence of its moments.

Proposition 2.2 *Let $\eta(p) = |\alpha| + \|\beta\|_\infty\, \|e_1\|_p$, where $\|\cdot\|_\infty$ and $\|\cdot\|_p$ denotes the sup- and the L^p-norms, respectively, and let $\{X_t, t \in I\!N\}$ be the solution process of (2.3), with $f(x) = \alpha x$ and $g(x) = \beta(x)x + \delta$. Under (H1) and (H2):*

i) if $\eta(1) < 1$ then $\{X_t, t \in I\!N\}$ is geometrically ergodic;

ii) if $\eta(p) < 1$, *then* $\|X_t^\pi\|_p < +\infty$, *where* $\{X_t^\pi, t \in I\!N\}$ *denotes the stationary solution process to (2.3).*

Remark 2.3 It is immediate to prove that $\eta(p) \leq \eta(q)$, whenever $1 \leq p \leq q$ and therefore *ii)* implies *i)*.

Corollary 2.1 *The solution processes to (2.1) and (2.2) are irreducible, aperiodic, T-chains if* $d \neq 0$ *and* $|a| \neq 1$. *Moreover, if*

$$|a| \ + \ \sup\{|b_1|, |b_2|\} \ \|e_1\|_1 < 1, \tag{2.4}$$

then the solution processes are geometrically ergodic, while if

$$|a| \ + \ \sup\{|b_1|, |b_2|\} \ \|e_1\|_p < 1, \tag{2.5}$$

the stationary solution processes have finite moments up to order p.

Proof (of Proposition 2.2): Adopting a standard procedure (see Cappuccio, Ferrante and Fonseca (1998)), fix $p \geq 1$ and define the Lyapounov function $V_p(x) = (1 + |x|)^p$ and the compact set $K_p = \left\{x \in I\!R : (1 + |x|) \leq \frac{C_p + 1 - \eta(p)}{\lambda - \eta(p)}\right\}$, where $\eta(p) < \lambda < 1$ and $C_p = |\delta| \|e_1\|_p$. Since the solution process of (2.3) is an irreducible, aperiodic T-chain, then K_p is a petite set and it holds that $[\mathrm{E}\{V_p(X_t)|X_{t-1} = x\}]^{\frac{1}{p}} \leq 1 + |x|\eta(p) + C_p$. From this inequality we have that function V_p complies with condition (iii) of Theorem 15.0.1 in Meyn and Tweedie (1993). The geometric ergodicity ($p = 1$) and the existence of finite moments up to order p for the stationary solution process follow immediately. □

3 Properties of the invariant probability measures

In this section we assume that (2.1) and (2.2) admit a stationary solution process. Here, we consider only the case where $c > 0, d > 0, b_1 > 0, b_1 + b_3 < 0$ (which implies $b_3 < 0$),$(b_1+b_3)c+d > 0$ and $0 \leq a < 1$. The other cases can be considered following the same lines. The stochastic difference equation (2.2) defines a net of Markov processes $\{X_t^z, t \in I\!N\}$, indexed by $z \in I\!R^+$, encompassing models of the form (2.1). We prove that, under suitable regularity assumptions on the density of the variables of the noise sequence, the invariant probability measures associated to (2.1) and (2.2), denoted respectively by π_0 and $\{\pi_z, z \in I\!R^+\}$, admit density functions with respect to the Lebesgue measure. These densities, defined by a suitable integral equation, are functions continuous everywhere except at two points, where they diverge to $+\infty$. Moreover, we prove that π_z converge weakly to π_0, as z goes to zero. The approach is similar to that used by Chan and Tong (1986) to prove that the invariant probability measures associated to a suitable net of STAR (smooth threshold AR) models approximate that of a SETAR model. The results proved in this section allow us to consider a suitable AR model with STARCH errors as an approximation for the AR model with TARCH errors given by (2.1). This can be useful in order to estimate the threshold parameter c, as shown at the end of the next section.

First, we need to prove the following lemma.

Lemma 3.1 *Functions* $g_z(x) = \Big\{b_1 + b_3\Phi\Big(\frac{x-c}{z}\Big)\Big\}x + d$, *for a sufficiently small* $z \in I\!R^+$, *and* $g_0(x) = \Big\{b_1 + b_3 1_{[c,+\infty)}(x)\Big\}x + d$ *have two zeroes and, for all* $x \neq c$, $g_z(x) \to g_0(x)$ *as* $z \to 0$.

Proof : Since the pointwise convergence of $g_z(x)$ to $g_0(x)$ for $x \neq c$ is straightforward, we only prove that $g_z(x)$ and $g_0(x)$ have two zeroes. With the above assumptions on the parameters, it is easy to see that $g_0(\cdot)$ has two zeroes, namely $x_1^0 = -d/b_1 < 0$ and $x_2^0 = -d/(b_1 + b_3) > c$. Moreover, we have that $g_0(c^-) = b_1 c + d > 0$ and $g_0(c) = (b_1 + b_3)c + d > 0$. With regard to $g_z(x)$, since $\Phi(0) = \frac{1}{2}$, we have $g_z(c) = \Big(b_1 + \frac{b_3}{2}\Big)\, c + d$ and $g_0(c^-) > g_z(c) > g_0(c) > 0$. For $x < c$, we have $g_z(x) = g_0(x) + b_3\Phi\left(\frac{x-c}{z}\right)x$, while for $x \geq c$, $g_z(x) = g_0(x) + b_3\left\{\Phi\left(\frac{x-c}{z}\right) - 1\right\}x$. Simple computations give that $g_z(x)$ is positive on (x_1^0, x_2^0). For a sufficiently small z, it is immediate to prove that $g_z(x)$ has two zeros, x_1^z, x_2^z, with $x_1^z \uparrow x_1^0 < 0$ and $x_2^z \downarrow x_2^0 > 0$, as z goes to zero. □

Remark 3.1 For different choices of the coefficients than those fixed at the beginning of the present section, $g_z(x)$ may possess one more zero than $g_0(x)$. For example, this happens if $b_1 > 0$, $b_1 + b_3 > 0$, $(b_1 + b_3)c + d < 0$ and $b_1 c + d > 0$. The extra zero converges to c as z goes to zero.

From now on we take $z \in [0, \eta]$, $\eta > 0$, such that the functions $g_z(x)$ have two zeros. The next proposition states that, under suitable regularity assumptions on the noise sequence, the invariant probability measures admit densities. For this purpose, we follow the procedure considered by Tong ((1990), Theorem 4.5).

Proposition 3.1 *Suppose that* e_1 *satisfies (H2) with a bounded and uniformly continuous density* $p_e(\cdot)$. *Then, for each* $z \geq 0$, *the invariant probability measure* π_z *is absolutely continuous with density function* $p_z(\cdot)$, *continuous on* $I\!R \setminus \{ax_1^z, ax_2^z\}$ *and such that* $p_z(x) \uparrow +\infty$ *as* $x \to ax_i^z$, $i = 1, 2$.

Proof : Fix $z \geq 0$. The stationary probability distribution of X_t^z satisfies

$$\begin{aligned} F_z(x) &= Pr[X_t^z \leq x] = Pr[aX_{t-1}^z + g_z(X_{t-1}^z)e_t \leq x] \\ &= \int_{I\!R} \left\{ \int_{av + g_z(v)u \leq x} p_e(u)\, du \right\} dF_z(v). \end{aligned}$$

We first prove that F_z is continuous everywhere. Given $x \in I\!R$ and $\varepsilon > 0$, we have

$$\begin{aligned} Pr[X_t^z \in (x-\varepsilon, x+\varepsilon)] &= Pr[x-\varepsilon < aX_{t-1}^z + g_z(X_{t-1}^z)e_t < x+\varepsilon] = \\ &= \int_{I\!R} \left\{ \int_{x-\varepsilon<av+g_z(v)u<x+\varepsilon} p_e(u)\,du \right\} dF_z(v) \\ &= \int_{I\!R\setminus\{x_1^z,x_2^z\}} \left\{ \int_{x-\varepsilon<av+g_z(v)u<x+\varepsilon} p_e(u)\,du \right\} dF_z(v) \\ &\quad + \sum_{i=1,2} \int_{\{x_i^z\}} \left\{ \int_{x-\varepsilon<av+g_z(v)u<x+\varepsilon} p_e(u)\,du \right\} dF_z(v) \quad . \end{aligned}$$

Define the sequence of measurable functions $f_n(v) = \int_{x-\frac{1}{n}<av+g_z(v)u<x+\frac{1}{n}} p_e(u)\,du$. If $x \neq ax_i^z$, for $i = 1,2$, it is immediate to prove that $\lim_{n\to+\infty} f_n(v) = 0$ for every $v \in I\!R$ (since $x \neq ax_i^z$, for $i = 1,2$, when $v = x_i^z$, $f_n(v) \equiv 0$ for n sufficiently big) and, by the Dominated Convergence Theorem, we obtain that $\lim_{n\to+\infty} \int_{I\!R} f_n(v)\,dF_z(v) = 0$. Therefore F_z is continuous. Now consider $x = ax_1^z$ (and the same can be done when $x = ax_2^z$); we have $\lim\limits_{n\to+\infty} f_n(v) = 0$ for every $v \in I\!R \setminus \{x_1^z\}$, while $f_n(x_1^z) \equiv 1$ for any n. In this case we obtain that

$$\begin{aligned} Pr[X_t^z = ax_1^z] &= \lim_{n\to+\infty} \int_{I\!R} f_n(v)\,dF_z(v) \\ &= \lim_{n\to+\infty} \int_{I\!R\setminus\{x_1^z\}} f_n(v)\,dF_z(v) + \lim_{n\to+\infty} \int_{\{x_1^z\}} f_n(v)\,dF_z(v) \ , \\ &= Pr[X_t^z = x_1^z] = 0 \end{aligned}$$

and the result is proved.

By Lemma 3.1 we have that $g_z(\,\cdot\,)$ is positive on (x_1^z, x_2^z) and negative otherwise. Thus

$$F_z(x) = \sum_{i=1}^{3} \int_{B_i} \left\{ \int_{av+g_z(v)u\le x} p_e(u)\,du \right\} dF_z(v),$$

where $B_1 = (-\infty, x_1^z)$, $B_2 = (x_1^z, x_2^z)$ and $B_3 = (x_2^z, +\infty)$. By the assumptions on $p_e(\,\cdot\,)$, for every x such that $|x - ax_i^z| \ge \delta > 0$, $i = 1,2$, there exists $k_\delta < \infty$ such that $\sup_v \frac{1}{|g_z(v)|} p_e \left\{ \frac{x-av}{g_z(v)} \right\} \le k_\delta$. Therefore, for x bounded away from ax_i^z, $i = 1,2$, we may differentiate under the integral sign, obtaining

$$\begin{aligned} \frac{d}{dx} F_z(x) &= \sum_{i=1}^{3} \int_{B_i} \frac{1}{|g_z(v)|} p_e \left\{ \frac{x-av}{g_z(v)} \right\} dF_z(v) \\ &= \int_{I\!R} \frac{1}{|g_z(v)|} p_e \left\{ \frac{x-av}{g_z(v)} \right\} dF_z(v) \ > \ 0. \end{aligned}$$

We conclude that π_z is absolutely continuous, with density function defined by

$$p_z(x) = \int_{I\!R} \frac{1}{|g_z(v)|} p_e \left\{ \frac{x-av}{g_z(v)} \right\} p_z(v)dv, \tag{3.1}$$

for $x \neq ax_i^z$, $i = 1, 2$.
Now fix $x \neq ax_i^z$, $i = 1, 2$ and $0 < \delta < \min_{i=1,2}\{|x - ax_i^z|\}$. Given $|x - y| \leq \delta$, we have

$$|p_z(x) - p_z(y)| \leq \int_{I\!R} \left| \frac{1}{|g_z(v)|} p_e \left\{ \frac{x - av}{g_z(v)} \right\} - \frac{1}{|g_z(v)|} p_e \left\{ \frac{y - av}{g_z(v)} \right\} \right| p_z(v) dv = F(x, y, I\!R).$$

Denoting by $I_\varepsilon(x)$ the interval $(x - \varepsilon, x + \varepsilon)$, with $\varepsilon > 0$, we have that

$$F(x, y, I\!R) = F\{x, y, I_\varepsilon(x_1^z) \cup I_\varepsilon(x_2^z)\} + F\{x, y, (I_\varepsilon(x_1^z) \cup I_\varepsilon(x_2^z))^c\}.$$

By the uniform continuity of $p_e(\cdot)$ and the continuity of $g_z(\cdot)$, we easily get that, for any positive ε and ε_1, there exists a $\delta \in (0, \min_{i=1,2}\{|x - ax_i^z|\})$ such that $F\{x, y, (I_\varepsilon(x_1^z) \cup I_\varepsilon(x_2^z))^c\} \leq \varepsilon_1$. Since p_e is uniformly continuous and bounded, we have that $\lim_{v \to x_i^z} \frac{1}{|g_z(v)|} p_e \left\{ \frac{x-av}{g_z(v)} \right\} = 0$ for any $x \neq ax_i^z$; therefore, for every fixed δ, we can choose ε sufficiently small so that $F\{x, y, I_\varepsilon(x_1^z) \cup I_\varepsilon(x_2^z)\} \leq \varepsilon_1$, and the continuity of $p_z(\cdot)$ on $I\!R \setminus \{ax_1^z, ax_2^z\}$ is proved.
Let us now prove that $p_z(x) \uparrow +\infty$, as $x \to ax_i^z$, $i = 1, 2$. Indeed, let $x \uparrow ax_1^z$ (the other cases follow analogously). By (3.1), we have that

$$p_z(x) \geq \int_{[x_1^z - 1, x_1^z - \varepsilon]} \frac{1}{|g_z(v)|} p_e \left\{ \frac{x - av}{g_z(v)} \right\} p_z(v) dv, \tag{3.2}$$

for any $x \leq ax_1^z$ and any $0 < \varepsilon \leq 1$. Since $x_1^z < ax_1^z$, we have that $p_z(\cdot)$ is continuous on $[x_1^z - 1, x_1^z]$ and, therefore, there exists $C_1 > 0$ such that

$$\min_{v \in [x_1^z - 1, x_1^z]} p_z(v) \geq C_1. \tag{3.3}$$

Now consider the function $(x, v) \mapsto u(x, v) := \dfrac{x - av}{g_z(v)}$ on $(-\infty, ax_1^z] \times (-\infty, x_1^z)$. It is sufficient to consider the behaviour of the function $u(x, v)$ on the rectangle $[a(x_1^z - \varepsilon), ax_1^z] \times [x_1^z - 1, x_1^z - \varepsilon]$. The upper left corner is on the line $x = av$, where $u(x, v(x)) \equiv 0$, and for all the other points (x, v) in the rectangle $u(x, v) < 0$. Since the function u is decreasing in x and increasing in v, the maximum of $|u(x, v)|$ on the rectangle is at the lower right corner $(x, v) = (ax_1^z, x_1^z - 1)$ which is fixed for every value of ε. Hence if $M = |u(ax_1^z, x_1^z - 1)| = \dfrac{-a}{g_z(x_1^z - 1)}$, for any $0 < \varepsilon < 1$ we have $\max_{(x,v) \in [ax_1^z - a\varepsilon, ax_1^z] \times [x_1^z - 1, x_1^z - \varepsilon]} |u(x, v)| = M$. The continuity of $p_e(\cdot)$ ensures that there exists a constant $C_2 = C_2(M)$ such that

$$\min_{(x,v) \in [ax_1^z - a\varepsilon, ax_1^z] \times [x_1^z - 1, x_1^z - \varepsilon]} p_e(u(x, v)) \geq C_2 \quad . \tag{3.4}$$

By (3.2), (3.3) and (3.4) we have

$$\inf_{x \in [ax_1^z - a\varepsilon, ax_1^z)} p_z(x) \geq C_1 C_2 \int_{[x_1^z - 1, x_1^z - \varepsilon]} \frac{1}{|g_z(v)|} dv \quad .$$

Since $\int_{[x_1^z-1,x_1^z)} \frac{1}{|g_z(v)|}\,dv = +\infty$, we have that $p_z(x) \uparrow +\infty$ as $x \uparrow ax_1^z$. $\square$

The aim now is to prove that the invariant probability measures π_z, $z \in I\!R^+$, converge in some sense to π_0, as z goes to zero. To do this, we prove that the family $\{\pi_z, z \in [0,\eta]\}$ is tight and that the weak limit (abbreviated as $\stackrel{w}{\to}$) of a suitable sequence $\{\pi_{z_n}, n \in I\!N\}$ coincides with π_0, as z_n goes to zero.

Proposition 3.2 *Suppose that e_1 satisfies (H2) with a bounded and uniformly continuous density $p_e(\,\cdot\,)$, and that (2.4) holds. Then the family of the invariant probability measures $\{\pi_z, z \in [0,\eta]\}$ is tight and $\pi_{z_n} \stackrel{w}{\to} \pi_0$, as $z_n \to 0$.*

Proof: By Proposition 3.1, we know that, for $z \in [0,\eta]$, π_z is an absolutely continuous probability measure. To prove the tightness of the family $\{\pi_z, z \in [0,\eta]\}$, we define $\gamma = \sup_{\theta\in[0,1]} |b_1 + \theta b_3| = \sup\{|b_1|, |b_2|\}$. Recalling (2.1) and (2.2), we find $|X_t^z| \le (|a| + \gamma|e_t|)\,|X_{t-1}^z| + |d||e_t|$. Therefore, if we define recursively the process $\{Y_t, t \in I\!N\}$ as $Y_t = (|a| + \gamma|e_t|)\,Y_{t-1} + |d||e_t|$, with the same sequence of error terms $\{e_t, t \in I\!N^+\}$, we have that

$$|X_t^z| \le Y_t, \quad t \in I\!N. \tag{3.5}$$

Furthermore, by condition (2.4), the process $\{Y_t, t \in I\!N\}$ is geometrically ergodic and, hence, it possess a unique invariant stationary distribution π^y. If the processes $\{Y_t, t \in I\!N\}$ and $\{X_t^z, t \in I\!N\}$ have the same positive starting value $Y_0 = X_0^z$, by (3.5), we conclude that

$$\pi_z([-k,k]) \ge \pi^y([0,k]) \to 1 \tag{3.6}$$

as $k \to \infty$. This implies that, $\forall\varepsilon > 0$, $\exists k > 0$ such that $1 - \varepsilon \le \pi_z([-k,k]) \le 1$ and, hence, the family of distributions $\{\pi_z\, z \in [0,\delta]\}$ is tight.
By Prokhorov's theorem, for every sequence $\{\pi_{z_n}, n \in I\!N\}$ with $z_n \to 0$, there exists a subsequence $\{\pi_{z_{n_k}}, k \in I\!N\}$ and a probability measure $\hat\pi$ such that $\pi_{z_{n_k}} \stackrel{w}{\to} \hat\pi$. For convenience, we refer to such a subsequence as $\{\pi_{z_n}, n \in I\!N\}$ and we prove that $\hat\pi$ is an invariant measure satisfying

$$\hat\pi(x) = \int_{I\!R} \frac{1}{|g_0(v)|} p_e\left\{\frac{x - av}{g_0(v)}\right\} \hat\pi(dv),$$

for every $x \ne ax_1^0, ax_2^0$. Consider again the set $I_\varepsilon(x) = (x - \varepsilon, x + \varepsilon)$, with $\varepsilon > 0$. For every function $f(\cdot) \in C_b(I\!R)$ (where $C_b(I\!R)$ is the set of the continuous and bounded function over $I\!R$), with $\operatorname{supp} f \subset I\!R \setminus (I_\varepsilon(ax_1^0) \cup I_\varepsilon(ax_2^0))$, we have that

$$\int_{I\!R} f(x)\pi_{z_n}(dx) \to \int_{I\!R} f(x)\hat\pi(dx), \tag{3.7}$$

as $n \to \infty$. Define $h(x,z,v) = \frac{1}{|g_z(v)|} p_e \left\{ \frac{x - av}{g_z(v)} \right\}$ and $\bar{h}(x,z,v) = h(x,z,v) - h(x,0,v)$. By Fubini's theorem, the left-hand side of (3.7) is

$$\int_{\mathbb{R}} f(x) \left\{ \int_{\mathbb{R}} p_{z_n}(v) h(x, z_n, v)\, dv \right\} dx = \int_{\mathbb{R}} p_{z_n}(v) \left\{ \int_{\mathbb{R}} f(x) h(x, z_n, v)\, dx \right\} dv$$
$$= \int_{\mathbb{R}} p_{z_n}(v) \left\{ \int_{\mathbb{R}} f(x) h(x, 0, v)\, dx \right\} dv + \int_{\mathbb{R}} p_{z_n}(v) \left\{ \int_{\mathbb{R}} f(x) \bar{h}(x, z_n, v)\, dx \right\} dv.$$

Since, by Lemma 3.1, $\forall\, \bar{\varepsilon} > 0$ there exists $N \in \mathbb{N}$ such that $\sup_{(v,x) \in \mathbb{R} \times \operatorname{supp} f} \; |\bar{h}(x, z_n, v)| < \bar{\varepsilon}$, for every $n > N$, we obtain that

$$\begin{aligned} \int_{\mathbb{R}} p_{z_n}(v) \left\{ \int_{\mathbb{R}} f(x) h(x, z_n, v)\, dx \right\} dv \quad &\to \quad \int_{\mathbb{R}} \left\{ \int_{\mathbb{R}} f(x) h(x, 0, v)\, dx \right\} \hat{\pi}(dv) \\ &= \quad \int_{\mathbb{R}} f(x) \left\{ \int_{\mathbb{R}} h(x, 0, v) \hat{\pi}(dv) \right\} dx, \end{aligned}$$

as $n \to \infty$. By (3.7), we get $\int_{\mathbb{R}} f(x) \hat{\pi}(dx) = \int_{\mathbb{R}} f(x) \left\{ \int_{\mathbb{R}} h(x,0,v) \hat{\pi}(dv) \right\} dx$ and $\hat{\pi}$ has density function $\hat{p}(\,\cdot\,)$, on $\mathbb{R} \setminus \{ax_1^0, ax_2^0\}$, such that $\hat{p}(x) = \int_{\mathbb{R}} h(x,0,v) \hat{\pi}(dv)$. By the uniqueness of the invariant probability measure, we can conclude that $\hat{\pi} = \pi_0$ and then $\pi_{z_n} \xrightarrow{w} \pi_0$, as $n \to \infty$. The conclusion clearly holds for the full sequence. □

Corollary 3.1 *Assume that the hypotheses of Proposition 3.2 are satisfied and that (2.5) is fulfilled for some $p > 1$. Then, for every $1 \le q < p$, $E[X^q_{\pi_{z_n}}] \to E[X^q_{\pi_0}]$ as $n \to \infty$, where X_{π_z} is a random variable with probability distribution π_z.*

Proof : By (3.6) and (2.5) we have that $\sup_n \mathrm{E}[|X_{\pi_{z_n}}|^p] < +\infty$. Then, by the Corollary to Theorem 25.12 in Billingsley (1995), we are done. □

Remark 3.2 For bilinear models (Tong (1990), p. 159) and for STAR models (Chan and Tong (1986)), there are stronger convergence results than those stated here, involving pointwise or uniform convergence of the density functions. The mean square convergence of the invariant distributions, in the former case, and the uniformly boundedness of the invariant densities, in the latter case, play a fundamental role. None of these conditions are satisfied here.

4 Simulations and inference using estimating functions

This final section provides a simple simulation study describing the sample path behaviour of processes satisfying (1.1). Moreover an inferential procedure, based on the theory of optimal estimating functions, is proposed for the class of AR models with (Smooth)TARCH errors.

Consider (1.1), with $\{e_t, t \in I\!N\}$ a sequence of i.i.d. random variables, such that $E(e_t) = 0$, $E(e_t^2) = 1$, $E(e_t^3) = 0$ and $E(e_t^4) = \rho$, $t \in I\!N$, $\rho \in I\!R^+$. Indeed, X_0 is assumed to be distributed as e_0. We study the cases A) $\{e_t, t \in I\!N\}$ is a sequence of standard normally distributed random variables, and B) e_t, $t \in I\!N$ are distributed as $Z\{(\nu - 2)/\nu\}^{1/2}$, where Z follows a Student's t-distribution with $\nu > 2$ degrees of freedom. Assumption B) assures that $E(e_t^2) = 1$ and $E(e_t^4) = 3\frac{\nu-2}{\nu-4}$, whenever $\nu > 4$. Figure 1 gives simulated sample paths of the model (1.1), for different parameter values. For each set of values, simulations are presented for both A) and B), with $\nu = 5$. The extreme behaviour of $\{X_t, t \in I\!N\}$ under B) shows larger fluctuations than under A). This is a consequence of the well-known fact that the density of the Student's t-distribution has heavier tails than those of the standard Gaussian density.

INSERT FIGURE 1

With regard to the estimation problem, a useful and relatively simple solution is based on the theory of optimal estimating functions, first considered for nonlinear time series models by Thavaneswaran and Abraham (1988). This approach, which approximates in some sense that based on likelihood, is usually applied when the likelihood function is not explicitly known or is not computationally tractable. We adopt this method since (1.1) is defined without any particular distributional assumption (such as the Gaussian one). Furthermore maximum likelihood inference, with the standard normal distribution for $\{e_t, t \in I\!N^+\}$, is obtained as a special case. This general semiparametric specification may be relevant, for example, in the field of finance where there is enough evidence for rejecting the usual assumption of normality. It is known that the unconditional distribution of returns to financial assets presents fatter tails than a Gaussian distribution and that ARCH-type models, under the assumption of conditional normality, describe such a behaviour. However, leptokurtosis may be found in the conditional distribution as well (see, for example, Bollerslev (1987)) and this fact justifies a non-Gaussian distributional assumption or, in a more general framework, the above semiparametric specification.

Let us briefly recall some basic definitions and fundamental results on estimating functions for stochastic processes. For an extensive introduction, see Godambe (1985), Godambe and Heyde (1987), Heyde (1988) and Heyde (1997). Consider $X_0, \ldots, X_n$ as a discrete time sample from a stochastic process whose probability measure is parameterized by an unknown parameter $\theta \in \Theta \subseteq I\!R^q$. Let $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ denote the corresponding class of finite-parameterized distributions on $I\!R^{n+1}$ and $\mathcal{F}_m$ the σ-field generated by $X_0, \ldots, X_m$, $m = 0, \ldots, n$. In order to simplify the exposition, take $q = 1$; the extension to the multi-dimensional parameter case is straightforward.

An estimating function is a suitable real function of $X_0, \ldots, X_n$ and θ given by $G_n(\theta) = G_n(\theta; X_0, \ldots, X_n)$. We get an estimator for θ by solving $G_n(\theta) = 0$. Since $G_n(\theta)$ is to substitute for the score function when it is unknown or computationally untreatable, one wishes to choose a function which is, in some sense, similar to that one. Thus, it may be natural to consider martingale estimating functions, namely to assume that $\{G_n(\theta), n \in I\!N\}$ is a martingale with respect to $\{\mathcal{F}_n, n \in I\!N\}$, when θ is the true parameter value. Indeed, an estimating function is called unbiased if $E_\theta\{G_n(\theta)\} = 0$, where the expectation is with respect to P_θ, for every $\theta \in \Theta$ and $n \in I\!N$. Godambe and Heyde (1987) discuss the problem of choosing a suitable estimating function and present criteria for determining, within a class of martingale estimating functions, the one closest to the true score function (fixed sample criterion) or the one with the smallest asymptotic variance as n tends to infinity (asymptotic criterion).

We assume that $X_0, \ldots, X_n$ is a sample from a Markov process and we consider the class $\mathcal{G}$ of unbiased martingale estimating functions of the general form

$$G_n(\theta) = \sum_{i=1}^{n} \sum_{j=1}^{r} \alpha_j(X_{i-1}; \theta) h_j(X_{i-1}, X_i; \theta),$$

where $h_j(X_{i-1}, X_i; \theta)$, $j = 1, \ldots, r$, are real valued functions such that $E_\theta\{h_j(X_{i-1}, X_i; \theta)|\mathcal{F}_{i-1}\} = 0$, for all θ and $i = 1, \ldots, n$. Note that the above inferential procedure is conditioned on $X_0 = x_0$. Whenever θ is a q-dimensional parameter with $q > 1$, functions $\alpha_j(\cdot)$, $j = 1, \ldots, r$, and therefore $G_n(\theta)$, are q-dimensional. A theorem by Kessler (1995) gives a procedure for finding, within the class $\mathcal{G}$, the optimal estimating function with respect to both the fixed sample criterion and the asymptotic criterion of Godambe and Heyde (1987). The result applies to general Markov processes and to discretely observed continuous-time Markov processes; some regularity assumptions on the transition probabilities of the Markov model and the conditional square integrability of $h_j(X_{i-1}, X_i; \theta)$, $j = 1, \ldots, r$, $i = 1, \ldots, n$, are usually required. For a Markov process model, we may consider estimating functions of the class $\mathcal{G}$ with $h^{(1)}(X_{i-1}, X_i; \theta) = X_i - \Gamma(X_{i-1}; \theta)$ and $h^{(2)}(X_{i-1}, X_i; \theta) = \{X_i - \Gamma(X_{i-1}; \theta)\}^2 - \Delta(X_{i-1}; \theta)$, where $\Gamma(X_{i-1}; \theta) = E_\theta(X_i|\mathcal{F}_{i-1})$ and $\Delta(X_{i-1}; \theta) = Var_\theta(X_i|\mathcal{F}_{i-1})$. Thus, we can define quadratic martingale estimating functions

$$\begin{aligned} G_n(\theta) &= \sum_{i=1}^{n} [\alpha_1(X_{i-1}; \theta)\{X_i - \Gamma(X_{i-1}; \theta)\} \\ &+ \alpha_2(X_{i-1}; \theta)\{(X_i - \Gamma(X_{i-1}; \theta))^2 - \Delta(X_{i-1}; \theta)\}], \end{aligned}$$

which are useful for inference within AR models with (Smooth)TARCH errors. From the theory of optimal martingale estimating functions (see, for example, Kessler (1995)), the optimal choice for α_1 and α_2 is

$$\alpha_1(X_{i-1};\theta) = \frac{\{\partial_\theta \Delta(X_{i-1};\theta)\}\Lambda(X_{i-1};\theta) - \{\partial_\theta \Gamma(X_{i-1};\theta)\}\Psi(X_{i-1};\theta)}{\Delta(X_{i-1};\theta)\Psi(X_{i-1};\theta) - \Lambda(X_{i-1};\theta)^2},$$

$$\alpha_2(X_{i-1};\theta) = \frac{\{\partial_\theta \Gamma(X_{i-1};\theta)\}\Lambda(X_{i-1};\theta) - \{\partial_\theta \Delta(X_{i-1};\theta)\}\Delta(X_{i-1};\theta)}{\Delta(X_{i-1};\theta)\Psi(X_{i-1};\theta) - \Lambda(X_{i-1};\theta)^2},$$

where $\Lambda(X_{i-1};\theta) = E_\theta[\{X_i - \Gamma(X_{i-1};\theta)\}^3|\mathcal{F}_{i-1}]$, $\Psi(X_{i-1};\theta) = E_\theta[\{X_i - \Gamma(X_{i-1};\theta)\}^4|\mathcal{F}_{i-1}] - \{\Delta(X_{i-1};\theta)\}^2$ and $\partial_\theta g(\theta)$ denotes the (column vector of partial) derivatives of a function $g(\theta)$ with respect to θ. Usually, the above conditional moments are unknown and they have to be found by simulation or by numerical methods.

We now consider an application of these general results to the threshold bilinear model expressed in the equivalent form (2.1). We assume that the random variables $\{e_t, t \in I\!N^+\}$ satisfy the above assumptions on the first four moments, without considering any specific distribution. The extension to the more general case where $E(e_t^3) = \tau$, with $\tau \in I\!R$, is straightforward. We first assume that ρ is known and denote $\theta = (a, b_1, b_3, c, d)$. In general, estimating thresholds in non-linear time series models is not a simple problem and, within the class of TAR (Threshold AR) processes, is usually solved by means of a two stage conditional least squares estimation procedure (see, for example, Chan (1993)). Alternatively, estimation is performed within an enlarged class of smooth models, such as those introduced in Section 2, which encompass the initial one (Chan and Tong (1986)). However, we consider the smooth models not as a new enlarged class of models but as an approximating model useful in order to define an estimating function suitable as well for the threshold c. Thus, the smoothing parameter z is fixed at a convenient value close to zero.

Recall the following well-known result on the tail probabilities of the standard normal distribution: for $u \to +\infty$, $1 - \Phi(u) = (u\sqrt{2\pi})^{-1}\exp(-u^2/2)\{1 + O(u^{-2})\}$. Therefore, if the random variables $\{e_t, t \in I\!N^+\}$ are absolutely continuous with respect to the Lebesgue measure, it is easy to see that, for $z \to 0$,

$$\Phi\Big(\frac{X_{t-1} - c}{z}\Big) = 1_{\{X_{t-1} \geq c\}} + O(z\exp(-(X_{t-1} - c)^2/(2z^2))), \quad t \in I\!N^+,$$

almost surely, since $Pr(X_{t-1} = c) = 0$. A possible choice for the smoothing parameter could be $z = n^{-1}$, where n is the sample size, so that the error term is of order $o(n^{-1})$.

With regard to the smooth model (2.2), which approximates (in the sense of Proposition 3.2) model (2.1), the (approximated) optimal quadratic martingale estimating function for θ, is

$$G_n(\theta) = (G_{n,a}(\theta), G_{n,b_1}(\theta), G_{n,b_3}(\theta), G_{n,c}(\theta), G_{n,d}(\theta))^{\mathrm{T}}. \tag{4.1}$$

Here, since

$$E_\theta(X_i|\mathcal{F}_{i-1}) = \Gamma(X_{i-1};\theta) = aX_{i-1},$$
$$Var_\theta(X_i|\mathcal{F}_{i-1}) = \Delta(X_{i-1};\theta) = [\{b_1 + b_3\Phi(\frac{X_{i-1}-c}{\mathrm{z}})\}X_{i-1} + d]^2,$$
$$E_\theta[\{X_i - E_\theta(X_i|\mathcal{F}_{i-1})\}^3|\mathcal{F}_{i-1}] = \Lambda(X_{i-1};\theta) = 0,$$
$$E_\theta[\{X_i - E_\theta(X_i|\mathcal{F}_{i-1})\}^4|\mathcal{F}_{i-1}] = \rho\{\Delta(X_{i-1};\theta)\}^2,$$

deleting a constant term depending on ρ, we obtain

$$G_{n,a}(\theta) = \sum_{i=1}^{n} \frac{X_{i-1}(X_i - aX_{i-1})}{\Delta(X_{i-1};\theta)},$$
$$G_{n,b_1}(\theta) = \sum_{i=1}^{n} \frac{X_{i-1}\{(X_i - aX_{i-1})^2 - \Delta(X_{i-1};\theta)\}}{\Delta(X_{i-1};\theta)^{3/2}},$$
$$G_{n,b_3}(\theta) = \sum_{i=1}^{n} \frac{X_{i-1}\{(X_i - aX_{i-1})^2 - \Delta(X_{i-1};\theta)\}\Phi(\frac{X_{i-1}-c}{z})}{\Delta(X_{i-1};\theta)^{3/2}},$$
$$G_{n,c}(\theta) = \sum_{i=1}^{n} \frac{X_{i-1}\{(X_i - aX_{i-1})^2 - \Delta(X_{i-1};\theta)\}\phi(\frac{X_{i-1}-c}{z})(-b_3z^{-1})}{\Delta(X_{i-1};\theta)^{3/2}},$$
$$G_{n,d}(\theta) = \sum_{i=1}^{n} \frac{\{(X_i - aX_{i-1})^2 - \Delta(X_{i-1};\theta)\}}{\Delta(X_{i-1};\theta)^{3/2}},$$

with $\phi(\cdot)$ the standard normal density function. The estimator $\hat{\theta}$, obtained as a solution to $G_n(\theta) = 0$, does not depend on the kurtosis parameter ρ, which can be eventually estimated by $\hat{\rho} = \frac{1}{n}\sum_{i=1}^{n} \frac{(X_i - \hat{a}X_{i-1})^4}{\Delta(X_{i-1};\hat{\theta})^2}$. When the threshold parameter c is assumed to be known, the optimal quadratic estimating function may be easily obtained, by considering the model in the form (2.1). Note that, in this context, conditional least squares methods are not appropriate, since they provide estimates only for the regression parameter a.

A simple numerical example is presented to illustrate the usefulness of this inferential procedure. We analyze observations from the model (2.1), with b_1, b_3 and c unknown and known $a = 0.2$, $d = 1$, $\rho = 6$. Estimation is based on the (approximated) quadratic estimating function for (b_1, b_3, c), obtained by choosing the second, the third and the fourth components of (4.1). We consider 1,000 simulated sample paths of dimension $n = 50, 100, 200, 300$ from a threshold bilinear model with $b_1 = -0.5$, $b_3 = 0.7$, $c = 0.5$ and e_t, $t \in I\!N$, distributed as $Z\{(\nu - 2)/\nu\}^{1/2}$, where Z follows a Student's t-distribution with $\nu = 6$ degrees of freedom.

The estimates for b_1, b_3 and c, with respect to a given sample path, cannot be obtained by a simple application of a Newton-Raphson iterative procedure since, in this case, the algorithm does not converge. This is related to the fact that, when the smoothing parameter z is close to zero (in this application we assume $z = n^{-1}$), the slope of the smoothing function $\Phi[(x-c)z^{-1}]$ is steep and the component of the estimating function referred to the threshold c is extremely peaked. In order to overcome this problem, our suggestion is to consider a set of possible values for c; a convenient grid of values may be $C = \{(X_{i-1} + X_i)/2, i = 1, \ldots, n\}$. With respect to a given sample path, the estimates for b_1, b_3 and c may be obtained by the following two steps procedure. First, for each fixed value $c \in C$, we determine the conditional estimates $\hat{b}_1(c)$, $\hat{b}_3(c)$ by means of an iterative procedure of Newton-Raphson type with initial values $(0, 0)$. Second, we consider as estimates for b_1, b_3 and c the quantities $\hat{b}_1(c^*)$, $\hat{b}_3(c^*)$, c^*, respectively, where c^* is the element $c \in C$ such that the component of the estimating function referred to the threshold, namely $G_{n,c}(\hat{b}_1(c), \hat{b}_3(c))$, is closest to zero.

Table 1 presents the sample mean values, and the associated standard deviations, for the corresponding estimators. These simulations show that, as expected, the larger the sample size the better the conformance of the estimators to the theoretical parameter values. Although the results are encouraging, this numerical example is very simple. A further numerical investigation is needed to verify the usefulness of the method for estimating, in particular, the threshold parameter c.

INSERT TABLE 1

Acknowledgements

The authors would like to thank an associate editor and an anonymous referee for their suggestions and careful reading. This research was partially supported by a grant from Ministero dell'Istruzione, dell'Università e della Ricerca, Italy.

References

[1] P. Billingsley (1995). Probability and Measure (3rd edn). John Wiley & Sons, New York.

[2] T.P.Bollerslev (1987). A conditional heteroskedastic time-series model for security prices and rates of return data. Rev. Econom. Statist. 69, 542-547.

[3] N.Cappuccio, M.Ferrante and G.Fonseca (1998). A note on the stationarity of a threshold first-order bilinear process. Statist. Probab. Lett. 40, 379-384.

[4] K.S.Chan (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. Ann. Statist. 21, 520-533.

[5] K.S.Chan and H.Tong (1986). On estimating thresholds in autoregressive models. J. Time Ser. Anal. 7, 179-190.

[6] D.B.Cline and H.H.Pu (1998). Verifying irreducibility and continuity of a nonlinear time series. Statist. Probab. Lett. 40, 139-148.

[7] G.Fonseca and R.L.Tweedie (2002). Stationary measures for non-irreducible non-continuous Markov chains with times series applications. Statistica Sinica 12, 651-660.

[8] V.P.Godambe (1985). The foundations of finite sample estimation in stochastic processes. Biometrika 72, 419-428.

[9] V.P.Godambe and C.C.Heyde (1987). Quasi likelihood and optimal estimation. Int. Stat. Rev. 55, 231-244.

[10] C.C.Heyde (1988). Fixed sample and asymptotic optimality for classes of estimating functions. Contemporary Mathematics 80, 241-247.

[11] C.C.Heyde (1997). Quasi-Likelihood and Its Application. Springer-Verlag, New York.

[12] M.Kessler (1995). Martingale estimating functions for a Markov chain. Preprint, Laboratoire de Probabilités, Université Paris VI.

[13] J.Liu, W.K.Li and C.W.Li (1997). On a threshold autoregression with conditional heteroscedastic variances. J. Statist. Plann. Inference 62, 279-300.

[14] S.P.Meyn and R.L.Tweedie (1993). Markov Chains and Stochastic Stability. Springer-Verlag, London.

[15] R.Rabemananjara and J.M.Zakoian (1993). Threshold ARCH models and asymmetries in volatility. J. Appl. Econometrics 8, 31-49.

[16] A.Thavaneswaran and B.Abraham (1988). Estimation for non-linear time series models using estimating equations. J. Time Ser. Anal. 9, 99-108.

[17] H.Tong (1990). Non-linear Time Series. A Dynamical System Approach. Oxford University Press, New York.

[18] A.A.Weiss (1984). ARMA models with ARCH errors. J. Time Ser. Anal. 3, 750-783.

Table 1: Sample means and standard deviations (s.d.) for the estimators of b_1, b_3 and c. Computation based on 1,000 replications of samples of size n from model (2.1), with $b_1 = -0.5$, $b_3 = 0.7$, $a = 0.2$, $c = 0.5$, $d = 1$, $\rho = 6$, and underlying Student's t-distribution.

True value	Sample	$n = 50$	$n = 100$	$n = 200$	$n = 300$
$b_1 = -0.5$	mean	-0.401	-0.463	-0.483	-0.488
	s.d.	0.379	0.210	0.148	0.121
$b_3 = 0.7$	mean	0.492	0.603	0.666	0.672
	s.d.	0.622	0.362	0.196	0.156
$c = 0.5$	mean	0.545	0.509	0.501	0.499
	s.d.	0.242	0.111	0.053	0.035

Figure 1: Simulated sample paths of the model (1.1), with parameters: $a = 0.2$, $b_1 = -0.5$, $b_2 = 0.2$, $c = 0$, $d = 1$ (figures (a) and (b)) and $a = 0.2$, $b_1 = 1.1$, $b_2 = 0.3$, $c = 0$, $d = 1$ (figures (c) and (d)); e_t, $t \in I\!N$ follow a standard normal distribution (figures (a) and (c)) or are distributed as $Z\sqrt{3/5}$, where Z is a Student's t-distributed random variable with $\nu = 5$ (figures (b) and (d)).