Bernoulli **18**(2), 2012, 735–745
DOI: 10.3150/11-BEJ352

On moving-average models with feedback

DONG LI[1,*], SHIQING LING[1,**] and HOWELL TONG[2]

[1] *Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong. E-mail:* [*] *malidong@ust.hk;* [**] *maling@ust.hk*
[2] *Department of Statistics, London School of Economics and Political Science, Houghton Street, London, WC2A 2AE, UK. E-mail: howell.tong@gmail.com*

Moving average models, linear or nonlinear, are characterized by their short memory. This paper shows that, in the presence of feedback in the dynamics, the above characteristic can disappear.

Keywords: ACF; ergodicity; existence; feedback; leptokurticity; memory; stationarity; thresholds

1. Introduction

Since the introduction by Slutsky [13], moving average models have played a significant role in time series analysis, especially in finance and economics. The models have been extended to include measurable (nonlinear) functions of independent and identically distributed random variables, representing unobservable and purely random impulses, for example, Robinson [12]. The characterizing feature of these models is the cut-off of the auto-covariance functions when they exist, implying that they are models of short memory. Another interesting feature of these models is the homogeneity of the random impulses, free from any feedback in the generating mechanism. Now, Slutsky developed these models in an economic context; the random impulses may correspond to unobservable political factors. In such a context, as well as in other contexts for which these models are relevant (e.g., business studies), it can be argued that feedback is often present: political decisions are often predicated on economic conditions. One simple way to incorporate feedback in these models is through the notion of thresholds, that is, on-off feedback controllers.

Since Tong [14] initiated the threshold notion in time series modelling, the notion has been extensively used in the literature, especially for the threshold autoregressive (TAR) or TAR-type models. For these models, some basic and probabilistic properties were given in Chan *et al.* [3] and Chan and Tong [4]. More related results can be found in An and Huang [1], Brockwell *et al.* [2], Chen and Tsay [5], Cline and Pu [6, 7], Ling [8], Ling *et al.* [9], Liu and Susko [10] and Lu [11], among others. A fairly comprehensive review of threshold models is available in Tong [15] and a selective survey of the history of threshold models is given by Tong [16].

However, most work to-date on the threshold model has primarily concentrated on the TAR or the TAR-type model. The threshold moving average (TMA) model, that is a moving average model with a simple on-off feedback control mechanism, has not attracted as much attention. As far as we know, only a few results are available for the TMA model. Brockwell *et al.* [2] investigated a threshold autoregressive and moving-average (TARMA) model and obtained a strictly stationary and ergodic solution to the model when the MA part does not contain any threshold component. Unfortunately, their TARMA model does not cover the TMA model as a special case. Using the Markov chain theory, Liu and Susko [10] provided the existence of the strictly stationary solution to the TMA model without any restriction on the coefficients. However, they neither gave an explicit form of the solution nor proved the ergodicity. A similar result can be found in Ling [8]. Ling *et al.* [9] gave a sufficient condition for the ergodicity of the solution for a first order TMA model under some restrictive conditions. These results have been extended to the first-order TMA model with more than two regimes. However, the uniqueness and the ergodicity of the solution are still open problems for higher-order TMA models.

In this paper, we use a different approach to study the TMA model without resorting to the Markov chain theory. Note that the TMA model involves a feedback control mechanism. An intuitive and simple idea is to seek a closed form of the solution in terms of the above mechanism, which is expressible as an indicator function. We can show that for the TMA model there always exists a unique strictly stationary and ergodic solution without any restriction on the coefficients of the TMA model. More importantly, for the first time in the literature, an explicit/closed form of the solution is derived. In addition, for the correlation structure, we show that the ACF (when it exists) of the TMA model typically does not cut off. In fact, it has a much richer structure. For example, it can exhibit almost long memory, although it generally decays at an exponential rate. Furthermore, the difference between the joint two-dimensional distribution and the corresponding product of its marginal distributions also decays to zero at an exponential rate as the lag tends to infinity.

The rest of the paper is organized as follows. Section 2 discusses the strict stationarity and ergodicity of the TMA model. Section 3 studies the asymptotic behaviour of the ACF of the TMA model and other correlation structure. We conclude in Section 4. All proofs of the theorems are relegated to the Appendix.

2. Stationarity and ergodicity of TMA(q) models

We first consider a TMA(q) model which satisfies the following equation:

$$y_n = \begin{cases} \mu_1 + e_n + \sum_{i=1}^{q} \phi_i e_{n-i}, & \text{if } y_{n-d} \le r, \\ \mu_2 + e_n + \sum_{i=1}^{q} \psi_i e_{n-i}, & \text{if } y_{n-d} > r, \end{cases} \tag{2.1}$$

where $\{e_n\}$ is a sequence of i.i.d. random variables. Here, q and d are positive integers, $r \in \mathbb{R}$, the real line, is the threshold parameter, and μ_1, μ_2, ϕ_i and ψ_i, $i = 1, \ldots, q$, are real coefficients.

For the sake of simplicity, we adopt the following notation:

$$U_n = 1(a_n \le r) \quad \text{and} \quad W_n = 1(b_n \le r) - 1(a_n \le r),$$

where $1(\cdot)$ is an indicator function,

$$a_n = \mu_2 + e_n + \sum_{i=1}^{q} \psi_i e_{n-i} \quad \text{and} \quad b_n = \mu_1 + e_n + \sum_{i=1}^{q} \phi_i e_{n-i}. \tag{2.2}$$

The following theorem gives the strict stationarity and ergodicity of model (2.1).

Theorem 2.1. *Suppose that $\{e_n\}$ is a sequence of i.i.d. random variables with $\mathbb{P}(a_n \le r, b_n \le r) + \mathbb{P}(a_n > r, b_n > r) \ne 0$. Then y_n has a unique strictly stationary and ergodic solution expressed by*

$$y_n = \mu_2 + e_n + \sum_{i=1}^{q} \psi_i e_{n-i} + \left[(\mu_1 - \mu_2) + \sum_{i=1}^{q} (\phi_i - \psi_i) e_{n-i}\right] \alpha_{n-d}, \qquad a.s.,$$

where

$$\alpha_{n-d} = \sum_{j=1}^{\infty} \left[\left(\prod_{s=1}^{j-1} W_{n-sd} \right) U_{n-jd} \right], \qquad \text{in } L^1 \text{ and a.s.}$$

If e_1 has a strictly and continuously positive density on $\mathbb{R}$ (e.g., normal, Student's t_v or double exponential distribution), then $\mathbb{P}(a_n \le r, b_n \le r) + \mathbb{P}(a_n > r, b_n > r) \ne 0$. The basic idea for Theorem 2.1 is a direct and concrete expression in terms of $1(y_{n-d} \le r)$, without resorting to the Markov chain theory. Theorem 2.1 shows that the TMA(q) model is always stationary and ergodic as is the MA(q) model.

3. The ACF of TMA(q) models

The ACF plays a crucial role in studying the correlation structure of weakly stationary time series. It is well known that for a causal ARMA(p, q) model, its ACF ρ_k goes to zero at an exponential rate as k diverges to infinity. The exact formula for ACF can be obtained although its closed form is not compact. However, for a general nonlinear time series model, it is rather difficult to obtain an exact formula for the ACF and to study the asymptotic behaviour. Additionally, the notion of memory, short or long, is closely associated with the ACF. One significant fact is that a causal ARMA(p, q) model is short-memory. For a general nonlinear time series model, due to its complicated structure, there is no universally accepted criterion for determining whether or not it is short-memory. As for some specific time series model, an ad hoc approach is usually adopted.

One important characteristic of the MA(q) model is that its ACF cuts off after lag q. Interestingly, this property is not generally inherited by the TMA model; this is not surprising theoretically because the TMA model involves some nonlinear feedback. Another interesting fact is that although a TMA model is generally short-memory, in some cases it can exhibit some almost long-memory phenomena; see Example 3.3. The following theorem characterizes the ACF of model (2.1).

Theorem 3.1. *Suppose that the condition in Theorem 2.1 is satisfied and* $\mathbb{E}|e_1|^2 < \infty$*. Then there exists a constant* $\rho \in (0,1)$ *such that* $\rho_k = \mathrm{O}(\rho^k)$*.*

Theorem 3.1 indicates that the TMA model (2.1) is short-memory. The next theorem describes the relationship between the two-dimensional joint distribution and the corresponding marginal distributions.

Theorem 3.2. *Suppose that* $\{e_n\}$ *is i.i.d. random variables having a continuously, boundedly and strictly positive density. Then, for any* $u, v \in \mathbb{R}$ *and* $k \geq 1$*, there exists a constant* $\rho \in (0,1)$ *such that*

$$|\mathbb{P}(y_0 \leq u, y_k \leq v) - \mathbb{P}(y_0 \leq u)\mathbb{P}(y_k \leq v)| = \mathrm{O}(\rho^k).$$

Actually, Theorem 3.2 still holds for $\mathrm{Cov}(1(u_1 < y_0 \leq u_2), 1(v_1 < y_k \leq v_2))$ where $-\infty \leq u_1 < u_2 \leq \infty$ and $-\infty \leq v_1 < v_2 \leq \infty$. Next, we consider some special TMA models and study their ACFs as well as some other properties.

Example 3.1. Suppose that y_n is defined as

$$y_n = \begin{cases} \mu_1 + e_n, & \text{if } y_{n-1} \leq r, \\ \mu_2 + e_n, & \text{if } y_{n-1} > r, \end{cases}$$

where $\{e_n\}$ satisfies the condition in Theorem 2.1 with mean 0 and finite variance σ^2.

This example can also be regarded as a special case of the TAR model, which was studied in Tong [15], Question 29, page 212. By calculation, we have the ACF of $\{y_n\}$

$$\rho_k = \frac{(\mu_1 - \mu_2)\lambda_k + (\mu_1 - \mu_2)^2 \delta_0 (1 - \delta_0)\beta^k}{\sigma^2 + (\mu_1 - \mu_2)^2 \delta_0 (1 - \delta_0)} \qquad \text{for } k \geq 1,$$

where $\lambda_k = \mathbb{E}[e_{n-k} 1(y_{n-1} \leq r)]$, $\beta = [G(r - \mu_1) - G(r - \mu_2)] \in (-1, 1)$ and $\delta_0 = G(r - \mu_2)/[1 - G(r - \mu_1) + G(r - \mu_2)]$. Here, $G(x)$ is the distribution function of e_1.

Clearly, the ACF does not possess the cut-off property except for $\mu_1 = \mu_2$. Generally, ρ_k decays exponentially since $\lambda_k = \mathrm{O}(\rho^k)$ for some $\rho \in (0,1)$ by the proof of Theorem 3.2. In the nonlinear time series literature, the search for a nonlinear AR model with long memory has been largely in vain. Against this background, it is interesting to note that as $\mu_1 \to \infty$ and $\mu_2 \to -\infty$, ρ_k can exhibit almost long memory in that ρ_k can be made to decay arbitrarily slowly. Note that Example 3.1 can be driven by a white noise process with a thin tailed distribution. The skewness and the kurtosis of y_n are also available explicitly and interesting. Specifically,

$$\text{skewness} = \frac{\mathbb{E}e_1^3 + (\mu_1 - \mu_2)^3(\delta_0 - 3\delta_0^2 + 2\delta_0^3)}{[\sigma^2 + (\mu_1 - \mu_2)^2 \delta_0 (1 - \delta_0)]^{3/2}}$$

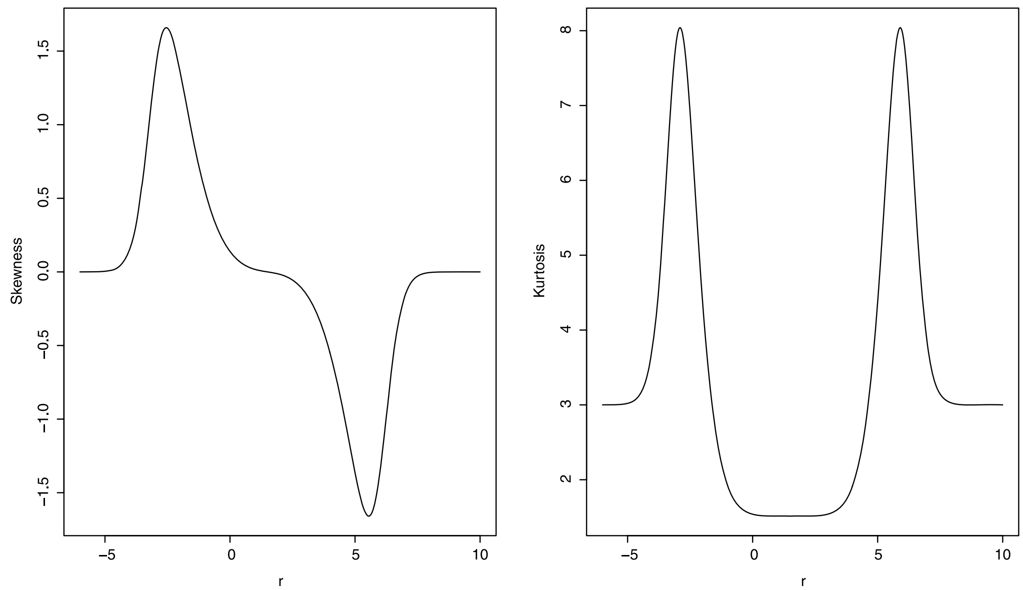

Figure 1. The skewness (left) and the kurtosis (right) of y_n as functions of r when e_1 is standard normal.

and

$$\text{kurtosis} = \frac{\mathbb{E}e_1^4 + 6\sigma^2(\mu_1-\mu_2)^2\delta_0(1-\delta_0) + (\mu_1-\mu_2)^4(\delta_0 - 4\delta_0^2 + 6\delta_0^3 - 3\delta_0^4)}{[\sigma^2 + (\mu_1-\mu_2)^2\delta_0(1-\delta_0)]^2},$$

respectively. The impact of the threshold parameter r is related to the bi-modality of the marginal density, which can be established by simple calculation. When e_1 is standard normal and $(\mu_1, \mu_2) = (4, -1)$, Figure 1 shows the skewness and the kurtosis of y_n as functions of r.

Example 3.2. Suppose that $\{y_n\}$ follows a TMA(1) model without drift:

$$y_n = \begin{cases} e_n + \phi e_{n-1}, & \text{if } y_{n-2} \le r, \\ e_n + \psi e_{n-1}, & \text{if } y_{n-2} > r, \end{cases}$$

where $\{e_n\}$ satisfies the condition in Theorem 2.1, having zero mean and finite variance.

After simple calculation, we have the ACF

$$\rho_k = \begin{cases} \dfrac{\psi + (\phi-\psi)\varrho}{1+\psi^2+(\phi^2-\psi^2)\varrho}, & \text{if } k=1, \\ 0, & \text{if } k \ge 1, \end{cases}$$

where $\varrho = \mathbb{P}(e_2 + \psi e_1 \le r)/[\mathbb{P}(e_2 + \phi e_1 > r) + \mathbb{P}(e_2 + \psi e_1 \le r)] \in [0, 1)$.

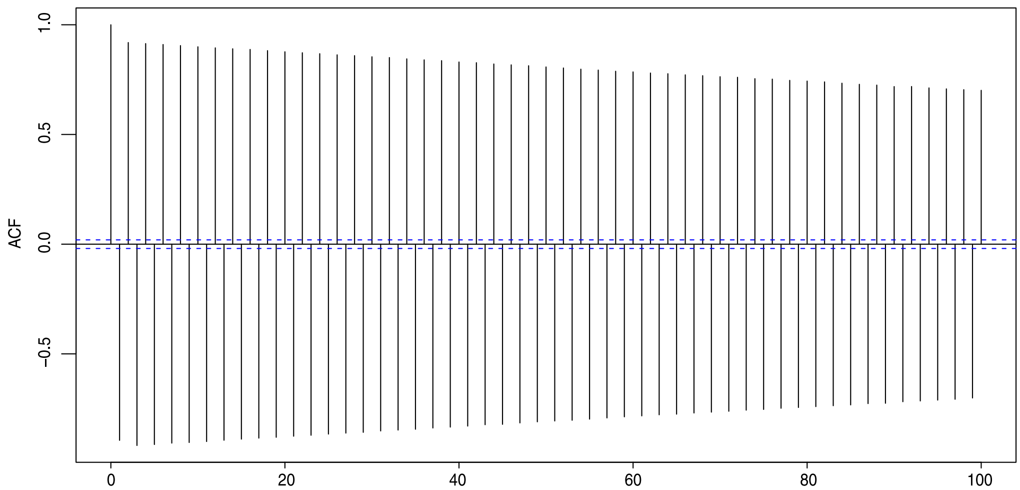

Figure 2. The sample ACFs of model (3.1).

This example shows that for some special TMA(q) model, the ACF may be cut off after lag q. In particular, if $\phi = \psi$, then the ACF coincides with that of the classical linear MA(1) model. Unfortunately, for general TMA models with $d \leq q$, there are no explicit expressions available for the ACFs due to the extremely complicated dependence of y_{t-d} on $\{e_{t-j}, d \leq j \leq q\}$. However, we can obtain the sample ACFs of TMA models by simulation.

Example 3.3. Suppose that $\{y_n\}$ follows the TMA(1) model:

$$y_n = \begin{cases} 5 + e_n + 0.2e_{n-1}, & \text{if } y_{n-1} \leq 0.5, \\ -3 + e_n + 0.8e_{n-1}, & \text{if } y_{n-1} > 0.5, \end{cases} \tag{3.1}$$

where $\{e_n\}$ is i.i.d. standard normal.

This model produces a time series that mimics a unit root and long memory. In Figure 2, the sample ACF of model (3.1) decays slowly, although model (3.1) is stationary.

4. Concluding remarks

Conventional moving average models, whether linear or nonlinear, assume absence of any feedback control mechanism. This paper shows that the introduction of simple feedback can enrich the structure of moving average models. For example, their ACF need not cut off but can now exhibit (near) long memory. Their distributions can be leptokurtic even when driven by Gaussian white noise. In nonlinear time series modeling, moving average models have been overshadowed by autoregressive models. Our study suggests that, by introducing a simple feedback mechanism, the notion of moving average possesses some unexpected properties beyond the shadow.

Appendix A: Proofs of theorems

A.1. Proof of Theorem 2.1

From model (2.1), $1(y_n \le r) = U_n + W_n 1(y_{n-d} \le r)$. Iterating $k \ge 1$ steps, we have

$$1(y_n \le r) = \sum_{j=0}^{k-1}\left[\left(\prod_{s=0}^{j-1} W_{n-sd}\right)U_{n-jd}\right] + \left(\prod_{i=0}^{k-1} W_{n-id}\right)1(y_{n-kd} \le r)$$

with the convention $\prod_0^{-1} = 1$. Let

$$\alpha_{n,k} = \sum_{j=0}^{k-1}\left[\left(\prod_{s=0}^{j-1} W_{n-sd}\right)U_{n-jd}\right].$$

For given d and q, there exists a unique nonnegative integer m such that $md < \max(d, q+1) \le (m+1)d$. Let $\delta = \mathbb{E}|W_1|$. Under the condition in Theorem 2.1, it is not difficult to prove that $0 \le \delta < 1$. Observing that both $\{U_n\}$ and $\{W_n\}$ are q-dependent sequences, we can extract an independent subsequence $\{W_{n-j(m+1)d}, j = 0, 1, \dots, [\frac{k-1}{m+1}]\}$ from the sequence $\{W_{n-id}, i = 0, 1, 2, \dots, k-1\}$, where $[a]$ denotes the integral part of a. Since $|U_n| \le 1$ and $|W_n| \le 1$, it yields that

$$\mathbb{E}\left|\left(\prod_{i=0}^{k-1} W_{n-id}\right)U_{n-kd}\right| \le (\mathbb{E}|W_1|)^{[(k-1)/(m+1)]},$$

implying

$$\sum_{j=1}^{\infty}\mathbb{E}\left|\left(\prod_{i=0}^{j-1} W_{n-id}\right)U_{n-jd}\right| \le \sum_{j=1}^{\infty}\delta^{[(j-1)/(m+1)]} = (m+1)\sum_{k=0}^{\infty}\delta^k < \infty.$$

Using the above inequalities, we can prove that $\mathbb{E}|\alpha_{n,s} - \alpha_{n,t}| \to 0$ as $s, t \to \infty$ for each fixed n. By the Cauchy criterion, $\alpha_{n,k}$ converges in L^1 as $k \to \infty$. Write the limit as

$$\alpha_n = \sum_{j=0}^{\infty}\left[\left(\prod_{s=0}^{j-1} W_{n-sd}\right)U_{n-jd}\right].$$

Applying the inequalities above again, it is easy to get

$$\sum_{k=1}^{\infty}\mathbb{E}|\alpha_{n,k} - \alpha_n| \le \sum_{k=1}^{\infty}\sum_{j=k}^{\infty}\delta^{[(j-1)/(m+1)]} < \infty,$$

yielding that

$$\lim_{k\to\infty} \alpha_{n,k} = \alpha_n, \qquad \text{in } L^1 \text{ and a.s.}$$

Furthermore, recall that $U_n = 1(a_n \le r)$ and $W_n = 1(b_n \le r) - 1(a_n \le r)$, where a_n and b_n are defined in (2.2), we have the iterative sequence: $\alpha_{n,1} = U_n$ and

$$\alpha_{n,k} = U_n + W_n \alpha_{n-d,k-1} = (1-\alpha_{n-d,k-1})1(a_n \le r) + \alpha_{n-d,k-1}1(b_n \le r)$$

for each n and $k \ge 1$. Note that $\alpha_{n,k}$ and $\alpha_{n-d,k}$ have the same distribution for fixed k since the error $\{e_i\}$ is i.i.d. By induction over k, we have that $\alpha_{n,k}$ only takes two values 0 and 1 a.s. since $\alpha_{n,1}$ only takes 0 and 1. Thus, α_n at most takes two values 0 and 1 a.s., namely, $\alpha_n = 1(\alpha_n = 1)$ a.s. Define a new sequence $\{S_n\}$

$$S_n = \mu_2 + e_n + \sum_{i=1}^{q} \psi_i e_{n-i} + \left[(\mu_1 - \mu_2) + \sum_{i=1}^{q} (\phi_i - \psi_i) e_{n-i}\right] \alpha_{n-d}.$$

By simple calculation, we have

$$\begin{aligned} 1(S_n \le r) &= 1(a_n \le r)1(\alpha_{n-d} = 0) + 1(b_n \le r)1(\alpha_{n-d} = 1) \\ &= U_n + W_n 1(\alpha_{n-d} = 1) \\ &= U_n + W_n \alpha_{n-d} = \alpha_n, \qquad \text{a.s.} \end{aligned}$$

Hence,

$$S_n = \mu_2 + e_n + \sum_{i=1}^{q} \psi_i e_{n-i} + \left[(\mu_1 - \mu_2) + \sum_{i=1}^{q} (\phi_i - \psi_i) e_{n-i}\right] 1(S_{n-d} \le r), \qquad \text{a.s.}$$

Thus, $\{S_n\}$ is the solution of model (2.1) which is strictly stationary and ergodic.

To uniqueness, suppose that $\tilde{S}_n$ is a solution to model (2.1), then

$$1(\tilde{S}_n \le r) = U_n + W_n 1(\tilde{S}_{n-d} \le r).$$

Iterating the above equation, one can get for $k \ge 1$

$$1(\tilde{S}_n \le r) = \alpha_{n,k} + \left(\prod_{i=0}^{k-1} W_{n-id}\right) 1(\tilde{S}_{n-kd} \le r).$$

We can show that the second term of the previous equation converges to zero a.s. Thus, we have $1(\tilde{S}_n \le r) = \alpha_n$ a.s. Therefore,

$$\tilde{S}_n = \mu_2 + e_n + \sum_{i=1}^{q} \psi_i e_{n-i} + \left[(\mu_1 - \mu_2) + \sum_{i=1}^{q} (\phi_i - \psi_i) e_{n-i}\right] \alpha_{n-d}, \qquad \text{a.s.,}$$

that is, $\tilde{S}_n = S_n$ a.s. The proof is complete.

A.2. Proof of Theorem 3.1

The notations a_n and b_n are defined by (2.2), m and δ are the same as those in the proof of Theorem 2.1. From Theorem 2.1, we have $y_n = a_n + (b_n - a_n)\alpha_{n-d}$. For $n \geq (m+2)d + q$, we decompose α_{n-d} into two parts

$$\begin{aligned}\alpha_{n-d} &= \sum_{j=1}^{[(n-q)/d]-1}\left[\left(\prod_{s=1}^{j-1} W_{n-sd}\right)U_{n-jd}\right] + \sum_{j=[(n-q)/d]}^{\infty}\left[\left(\prod_{s=1}^{j-1} W_{n-sd}\right)U_{n-jd}\right]\\ &\equiv I_1 + I_2.\end{aligned}$$

Clearly, $I_1 \in \mathcal{F}_d^{n-d}$ and $I_2 \in \mathcal{F}_{-\infty}^{n-d}$, where $\mathcal{F}_m^n = \sigma(e_m, \ldots, e_n)$. By calculation, we have for $n \geq (m+2)d + q$

$$\begin{aligned}|\operatorname{Cov}(y_0, y_n)| &\leq \sum_{j=[(n-q)/d]}^{\infty}\left[\mathbb{E}\left(\prod_{s=1}^{j-1}|W_{n-sd}|\right)\right]^{1/2}[\mathbb{E}(b_n - a_n)^2(y_0 - \mathbb{E}y_0)^2]^{1/2}\\ &\leq \sum_{j=[(n-q)/d]}^{\infty}\sqrt{\delta}^{[(j-1)/(m+1)]}[\mathbb{E}(b_n - a_n)^2]^{1/2}[\mathbb{E}(y_0 - \mathbb{E}y_0)^2]^{1/2}\\ &\leq \frac{H(m+1)}{1-\sqrt{\delta}}\sqrt{\delta}^{[(n-q-d)/(d(m+1))]-1}\end{aligned}$$

by Hölder's inequality, the boundedness of W_n and U_n, and the independence of $\{b_n, a_n\}$ and y_0, where

$$H = \left[|\mu_1 - \mu_2| + (\mathbb{E}e_1^2)^{1/2}\sum_{i=1}^{q}|\phi_i - \psi_i|\right]\left[|\mu_1| + |\mu_2| + (\mathbb{E}e_1^2)^{1/2}\sum_{i=1}^{q}(|\phi_i| + |\psi_i|)\right].$$

Thus, the conclusion holds.

A.3. Proof of Theorem 3.2

Let $x_n = a_n + (b_n - a_n)I_1$. Then $y_n - x_n = (b_n - a_n)I_2$. Clearly, $x_n \in \mathcal{F}_d^n$ and $\mathbb{E}|y_n - x_n| = \mathrm{O}(\rho^n)$ for large enough n, where $\rho \in (0,1)$. So, using the independence of x_n and y_0, for large enough n, we have

$$\begin{aligned}&|\mathbb{P}(y_0 \leq u, y_n \leq v) - \mathbb{P}(y_0 \leq u)\mathbb{P}(y_n \leq v)|\\ &\quad = |\mathbb{E}\{[1(y_n \leq v) - 1(x_n \leq v)][1(y_0 \leq u) - \mathbb{E}1(y_0 \leq u)]\}|\\ &\quad \leq \mathbb{E}|1(y_n \leq v) - 1(x_n \leq v)|.\end{aligned}$$

On noting the independence between e_n and $\bar{e}_{n-1}$, where $\bar{e}_{n-1} = \mu_2 + \sum_{i=1}^{q}\psi_i e_{n-i} + (b_n - a_n)\alpha_{n-d}$, the density of y_n is $f_y(x) = \int_{\mathbb{R}} h(x-y)\,\mathrm{d}G_{\bar{e}}(y)$, where $h(x)$ is the density function of e_1 and $G_{\bar{e}}(y)$ is the distribution function of $\bar{e}_{n-1}$. On using the property of convolution, $f_y(x)$ is continuous and bounded. Write $\|f_y\|_\infty = \max\{|f_y(x)|\colon x \in \mathbb{R}\}$. On the one hand, using the following inequality

$$|1(x \le t) - 1(y \le t)|1(|x-y| \le \varepsilon) \le 1(t-\varepsilon \le x \le t+\varepsilon),$$

we can get

$$\mathbb{E}[|1(y_n \le v) - 1(x_n \le v)|1(|y_n - x_n| \le \varepsilon)] \le \mathbb{P}(v-\varepsilon \le y_n \le v+\varepsilon) \le 2\|f_y\|_\infty \varepsilon.$$

On the other hand, using Markov's inequality, we have

$$\mathbb{E}\{|1(y_n \le v) - 1(x_n \le v)|1(|y_n - x_n| > \varepsilon)\} \le \mathbb{E}1(|y_n - x_n| > \varepsilon)\} \le \varepsilon^{-1}\mathbb{E}|y_n - x_n|.$$

Choosing $\varepsilon = \mathrm{O}(\rho^{n/2})$, we can obtain

$$|\mathbb{P}(y_0 \le u, y_n \le v) - \mathbb{P}(y_0 \le u)\mathbb{P}(y_n \le v)| = \mathrm{O}(\rho^{n/2}).$$

Hence, the result holds.

Acknowledgements

We thank the referee, the Associate Editor and the Editor for their very helpful comments and suggestions. The research was partially supported by Hong Kong Research Grants Commission Grants HKUST601607 and HKUST602609, the Saw Swee Hock professorship at the National University of Singapore and the Distinguished Visiting Professorship at the University of Hong Kong (both to HT).

References

[1] An, H.Z. and Huang, F.C. (1996). The geometrical ergodicity of nonlinear autoregressive models. *Statist. Sinica* **6** 943–956. MR1422412

[2] Brockwell, P.J., Liu, J. and Tweedie, R.L. (1992). On the existence of stationary threshold autoregressive moving-average processes. *J. Time Ser. Anal.* **13** 95–107. MR1165659

[3] Chan, K.S., Petruccelli, J.D., Tong, H. and Woolford, S.W. (1985). A multiple-threshold AR(1) model. *J. Appl. Probab.* **22** 267–279. MR0789351

[4] Chan, K.S. and Tong, H. (1985). On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Adv. in Appl. Probab.* **17** 666–678. MR0798881

[5] Chen, R. and Tsay, R.S. (1991). On the ergodicity of TAR(1) processes. *Ann. Appl. Probab.* **1** 613–634. MR1129777

[6] Cline, D.B.H. and Pu, H.m.H. (1999). Geometric ergodicity of nonlinear time series. *Statist. Sinica* **9** 1103–1118. MR1744827

[7] Cline, D.B.H. and Pu, H.M.H. (2004). Stability and the Lyapounov exponent of threshold AR-ARCH models. *Ann. Appl. Probab.* **14** 1920–1949. MR2099657

[8] Ling, S. (1999). On the probabilistic properties of a double threshold ARMA conditional heteroskedastic model. *J. Appl. Probab.* **36** 688–705. MR1737046

[9] Ling, S., Tong, H. and Li, D. (2007). Ergodicity and invertibility of threshold moving-average models. *Bernoulli* **13** 161–168. MR2307400

[10] Liu, J. and Susko, E. (1992). On strict stationarity and ergodicity of a nonlinear ARMA model. *J. Appl. Probab.* **29** 363–373. MR1165221

[11] Lu, Z. (1998). On the geometric ergodicity of a non-linear autoregressive model with an autoregressive conditional heteroscedastic term. *Statist. Sinica* **8** 1205–1217. MR1666249

[12] Robinson, P.M. (1977). The estimation of a nonlinear moving average model. *Stochastic Process. Appl.* **5** 81–90. MR0428654

[13] Slutsky, E. (1927). The summation of random causes as the source of cyclic processes. *Voprosy Koniunktury* **3** 34–64. {English translation in *Econometrika* **5** 105–146.}.

[14] Tong, H. (1978). On a thresold model. In *Pattern Recognition and Signal Processing* (C.H. Chen, ed.) 575–586. Amsterdam: Sijthoff and Noordhoff.

[15] Tong, H. (1990). *Nonlinear Time Series: A Dynamical System Approach*. *Oxford Statistical Science Series* **6**. New York: Oxford Univ. Press. MR1079320

[16] Tong, H. (2011). Threshold models in time series analysis – 30 years on. *Stat. Interface* **4** 107–118.

Received July 2010 and revised December 2010