Department of Economics

Nonlinear Autoregressive Models and Long Memory

George Kapetanios

Working Paper No. 516 July 2004 ISSN 1473-0278

Queen Mary
University of London

Nonlinear Autoregressive Models and Long Memory

George Kapetanios*
Queen Mary, University of London

May 2004

Abstract

This note shows that regime switching nonlinear autoregressive models widely used in the time series literature can exhibit arbitrary degrees of long memory via appropriate definition of the model regimes.

Key Words: Long Memory, Nonlinearity.
JEL Classification: C15

1 Introduction

Nonlinear time series models have been used extensively in recent years to investigate economic phenomena. A number of classes of models have been popularised in the literature. Two of the main classes considered are threshold models and Markov-Switching models. The main characteristic of both classes is the non-constancy of the response of the dependent variables to the explanatory variables. This response which is, in linear regression models, simply the coefficient of the explanatory variable, is allowed to vary depending on the occurrence of given trigger events. The main difference between

*Department of Economics, Queen Mary, University of London, Mile End Road, London E1 4NS. email: G.Kapetanios@qmul.ac.uk

threshold models and Markov Switching models is that the first class assumes that coefficients change with respect to the occurrence of some event in an observed variable whereas for Markov Switching models the responses change with respect to a change in state of an unobserved discrete Markov Chain.

A related strand of the literature on the investigation of highly persistent processes looks at the presence of long memory in the data. Long memory and nonlinearity have rarely been jointly analysed. Exceptions include Davidson and Sibbertsen (2002), Diebold and Inoue (2001), van Dijk, Frances, and Paap (2002) and Kapetanios and Shin (2002). Within this small set of papers two strands are apparent. One strand considers long-memory and nonlinearity as alternative representations which maybe confused and tries to investigate their similarities and differences. Diebold and Inoue (2001) juxtapose the variance structures of long memory and Markov switching models. Davidson and Sibbertsen (2002) discuss one class of nonlinear models which have a similar variance structure to long memory models. Kapetanios and Shin (2002) suggest a formal test for distinguishing between nonstationary long memory and nonlinear geometrically ergodic models in small samples. On the other hand van Dijk, Frances, and Paap (2002) investigate the possibility that the nature of the process driving the long memory process is nonlinear. They apply such a model to US unemployment data with interesting results.

This short paper is in the spirit of the first strand and in particular extends the theoretical analysis of Diebold and Inoue (2001). That paper considered simple Markov Switching models where the only explanatory variable that was affected by the unobserved Markov chain was the constant implying the process switched means at particular time periods. By allowing the time intervals in which the processes did not switch mean to grow to infinity on average, Diebold and Inoue (2001) showed that the variance of the partial sums of the process would grow at a similar rate to long memory processes.

The current paper takes up this idea and explores both Markov Switching and threshold models whose autoregressive parameter is allowed to switch regimes. By appropriately specifying the model we can see that long memory may emerge. Unlike Diebold and Inoue (2001) who focused on variances of partial sums of the processes, we focus more on the autocovariance of the process itself. This appears to be of greater interest as it relates more closely to the economic concept of slow decay of shock effects. In statistical terms we obtain autocovariances which decay hyperborically rather than exponentially. Further, our model specification bridges the gap between standard stationary autoregressive models whose autocovariances decay exponentially and random walk models. A further advantage of our specification is the ability to provide a somewhat structural interpretation of the emergence of long memory rather than simply a statistical construct.

The note is structured as follows: Section 2 provides the setup of our analysis. Section 3 provides some theoretical results. Finally, Section 4 concludes.

2 The setup

We consider the following simple model

$$y_t = \rho_t y_{t-1} + \epsilon_t, \quad t = 1, \ldots, T \tag{1}$$

where $\rho_t = 1$ if $\mathcal{I}_t = 1$ and $\rho_t = \rho$, $|\rho| < 1$ if $\mathcal{I}_t = 0$ and ϵ_t is an i.i.d. process with finite variance σ^2. Depending on the specification of $\mathcal{I}_t$ this setup encompasses both threshold models and Markov-Switching models. More specifically, if $\mathcal{I}_t = I_{\{s_t=1\}}$ where $I_{\{.\}}$ denotes the indicator function and s_t is an ergodic Markov Chain with transition matrix P, whose elements do not depend on T, taking the values 1,2 then this is a Markov Switching model. If, on the other hand, $\mathcal{I}_t = I_{\{|x_{t-d}|<r\}}$, for finite r and some process x_t (possibly $x_t = y_t$) then this is a particular type of threshold model. As they stand these models describe strictly stationary and I(0) process, where an I(0) process is defined to be a process whose normalised partial sums converge to a Brownian motion. This is easy to see for both models. For the threshold models note that by the drift condition of Tweedie (1975) the model is easily seen to be geometrically ergodic, hence asymptotically stationary and β-mixing. Note that the geometric ergodicity result requires that x_t is absolutely continuous with uniformly continuous and positive pdf. Then, the result follows. For the Markov Switching model it is easily seen that the model is a near epoque dependent (NED) process, (see Davidson (1994)) of any arbitrary size. Hence, it satisfies a functional central limit theorem and it therefore is I(0). We do not provide more details on these properties at this stage as the analysis in the next section will elucidate our claims.

To obtain long memory behaviour and in particular slowly declining autocovariances, we will regulate the occurrence of the event $\mathcal{I}_t = 0$. In particular, the standard models assume that $\mathcal{I}_t = 0$ occurs for a proportion of periods in the sample which is bounded away from zero. But if we allow $\mathcal{I}_t = 0$ to occur increasingly rarely then the random walk behaviour will increasingly affect the persistence of the process. So the idea is relatively simple. Whereas short memory models allow random walk behaviour for some proportion of time bounded away from one, if we allow increasingly long periods of random walk behaviour then long memory emerges.

A structural interpretation of this phenomenon is viewing the events $\mathcal{I}_t = 0$ as increasingly rare events that constrain the evolution of the process. For example, in economics one can come up with a number of economic processes which may exhibit such behaviour. Processes such as exchange rates which may be usually left to evolve freely in the financial markets but may be constrained at particular points in time due to exceptional circumstances such as financial crises come to mind.

3 Theoretical Results

3.1 Markov-Switching Models

In this section we provide our theoretical results. We first examine the Markov Switching model. For simplicity we specify the transition matrix for s_t, for every T, by

$$P = \begin{pmatrix} 1-\varpi_T & 1-\varpi_T \\ \varpi_T & \varpi_T \end{pmatrix} \tag{2}$$

We introduce dependence of the transition matrix on T but do not specify any more details. These will be provided below. It is easy to establish that the stationary distribution of the chain, for given T, is $(\pi_1, \pi_2) = (1-\varpi_T, \varpi_T)$. Then we prove the following theorem

Theorem 1 *The process given by (1) with $\mathcal{I}_t = I_{\{s_t=1\}}$, where s_t is a two state discreet Markov chain with transition matrix P given by (2) is covariance stationary as long as (4) holds for ϖ_T.*

Proof

We have the following representation for y_t

$$y_t = \sum_{i=0}^{\infty} \rho^{\sum_{j=1}^{i} I_{\{s_{t-j}=2\}}} \epsilon_{t-i} = \sum_{i=0}^{\infty} c_{i,t} \epsilon_{t-i} \quad (3)$$

The condition for covariance stationarity of the process will be borne out of the ensuing analysis. We start by examining the decay properties of the coefficients of the above MA representation of the process. Note that (3) is not the same as the Wold representation of the process, since firstly the Wold representation will not necessarily exist unless y_t is stationary and secondly the coefficients $c_{i,t}$ depend on t. Note however, that $c_{i,t}$ is a stationary process for each i.

We examine $c_{i,t}$ for large i. Asymptotically, we have

$$\rho^{\sum_{j=1}^{i} I_{\{s_{t-j}=2\}}} = \left(\rho^{\frac{1}{i}\sum_{j=1}^{i} I_{\{s_{t-j}=2\}}}\right)^{i}$$

But, by the ergodicity of the chain and the standard law of large numbers we have that

$$plim_{i\to\infty} \frac{1}{i} \sum_{j=1}^{i} I_{\{s_{i-j}=2\}} = lim_{i\to\infty} \varpi_i$$

The same results holds in mean square (rather than in probability) since $I_{\{s_i=2\}}$ is a uniformly integrable sequence. Hence, asymptotically

$$\rho^{\sum_{j=1}^{T} I_{\{s_{t-j}=2\}}} \sim \rho^{T\varpi_T}$$

The process will be covariance stationary as long as $\rho^{2T\varpi} = o(\frac{1}{T})$ since then $\sum_{i=0}^{\infty} E(c_{i,t}^2) < \infty$. In fact if $\rho^{2T\varpi} = O(\frac{1}{T^{2-2d}})$ for $0 < d < 0.5$ the process exhibits long memory similar to that of ARFIMA processes with long memory parameter d. This follows from the asymptotic form of the coefficients of the MA representation of an ARFIMA process as given in (2.48) of Beran (1997). So, we need to investigate conditions for that. We set $\varpi_T = a \ln T/T$. We have

$$\rho^{2T\varpi_T} = \rho^{2a \ln T} = T^{2a \ln \rho}$$

So, we need

$$2a \ln \rho < 2d - 2$$

which gives for $d = 0.5$

$$a > \frac{-1}{2 \ln \rho} \tag{4}$$

This is a condition for stationarity. A condition for stationary long memory is

$$\frac{-1}{\ln \rho} > a > \frac{-1}{2 \ln \rho} \tag{5}$$

Obviously, for smaller a we have nonstationary long memory.

Q.E.D.

3.2 Threshold models

Now, let us examine threshold models. A simple threshold model, we focus on, takes the form

$$y_t = y_{t-1} I_{\{|v_t| < r\}} + \rho y_{t-1} I_{\{|v_t| \geq r\}} + \epsilon_t \tag{6}$$

where v_t is an i.i.d. sequence which is independent of ϵ_t and ϵ_t is again i.i.d. with variance σ^2. Then, the following theorem holds

Theorem 2 *The process given by (6) is covariance stationary as long as (7) holds.*

Proof

Similarly to the analysis of Markov-Switching models, we can write:

$$y_t = \sum_{i=0}^{\infty} \rho^{\sum_{j=1}^{i} I_{\{|v_{t-j}|\geq r\}}} \epsilon_{t-i}$$

We therefore focus on the behaviour of $\sum_{j=1}^{i} I_{\{|v_{t-j}|\geq r\}}$ as $i \to \infty$. From the proof of theorem 1, we know that a sufficient condition for stationarity is that $\sum_{j=1}^{i} I_{\{|v_{t-j}|\geq r\}} = o_{m.s.}(\ln i)$. In other words, we are interested in the behaviour of extreme realisations of v_t. To pose the problem more concretely, we are looking for r as a function of T such that only $o_{m.s.}(\ln T)$ realisations of $|v_t|$ out of a sample of size T exceed r. The theory of intermediate order statistics provides answers to this problem. Define the i-th order statistic of the sequence $\{v_t\}_1^T$ as $v_{i:T}$. Then, the focus of interest is the behaviour of $v_{T-k+1:T}$ when $T \to \infty$, $k \to \infty$ and $k/T \to 0$. Denote the distribution and density functions of v_t as F_v and f_v respectively. Then, by Theorem 2.1 of Falk (1989) we know that there exist sequences of constants $c_{T,k}$ and $d_{T,k}$ depending on F_v such that

$$sup_{B\in\mathcal{B}} \left| P\left(c_{T,k}^{-1}(v_{T-k+1:T} - d_{T,k}) \in B\right) - N_{(0,1)}(B)\right| \to 0$$

where $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}$, $P(x \in B)$ denotes the probability measure of x and $N_{(0,1)}(B)$ is $P(x \in B)$ when $x \sim N(0,1)$. Hence,

$$r = r_T = o(c_{T,\ln T} a + d_{T,\ln T}) \tag{7}$$

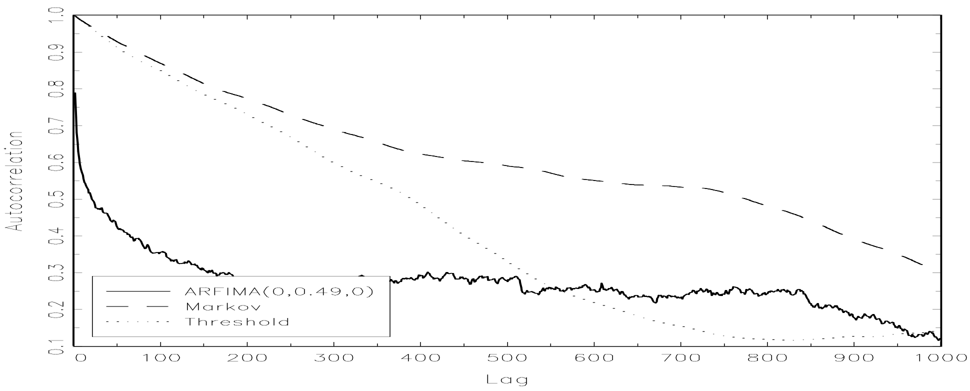
for some constant a is a sufficient condition for stationarity. For example, if $v_t \sim N(0,1)$ then

$$c_{T,\ln T} = T(\ln T)^{-1/2} \phi(\Phi^{-1}(1 - \ln T/T))$$

and

$$d_{T,\ln T} = \Phi^{-1}(1 - \ln T/T)$$

Figure 1:



In general, the sequences $c_{T,\ln T}$ and $d_{T,\ln T}$ are such that

$$\lim_{T\to\infty} c_{T,\ln T}/a_{T,\ln T} = 1$$

and

$$\lim_{T\to\infty} (d_{T,\ln T} - b_{T,\ln T})/a_{T,\ln T} = 0$$

where

$$b_{T,\ln T} = F_v^{-1}(1 - \ln T/T)$$

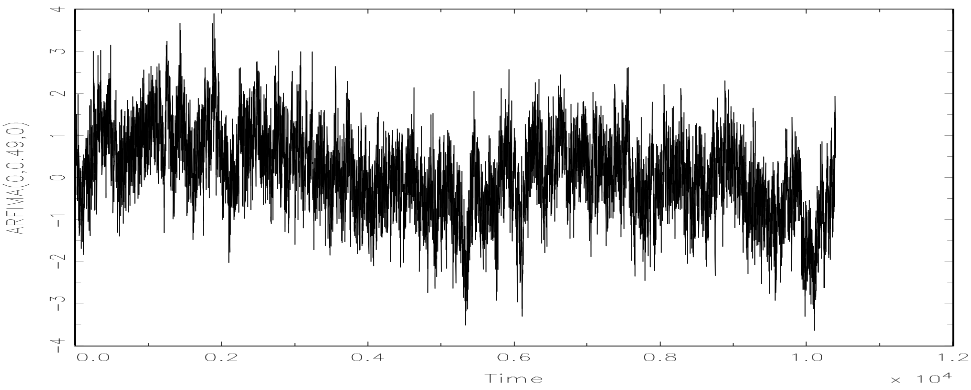
and

$$a_{T,\ln T} = (lnT)^{1/2}/(Tf_v(b_{T,\ln T}))$$

Q.E.D.

Of course these results can, in principle, be generalised for dependent processes v_t. In particular, mixing processes in the sense of definition 3.7.1 of Galambos (1978) (this is similar to standard strong mixing but applied

Figure 2:

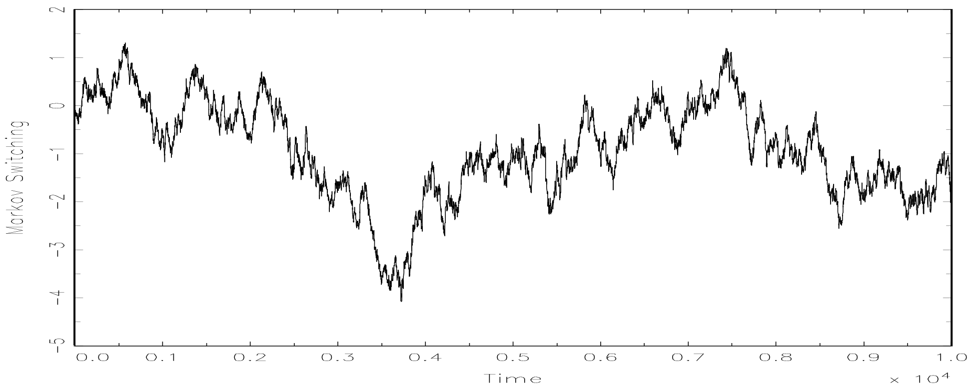


only to the tails of the relevant distribution) which also satisfy a bound on the probability that any two observations will take large values have been shown to obey similar laws concerning the behaviour of order statistics as i.i.d. processes. (see (Galambos, 1978, Ch. 3 Sec. 7))

Our analysis has been focused on nonlinear models with one lag. Multi-lag extensions of these results are obviously possible. In particular, what is needed is simply the existence of two regimes, one of which has an autoregressive polynomial with a unit root and occurs most of the time and another regime whose autoregressive polynomial has roots lying outside the unit circle and occurs rarely.

To give an empirical flavour of our results we have simulated processes for the Markov Switching and threshold models we have discussed. More specifically, for the Markov switching model we have used $\rho = 0.9$ and

Figure 3:

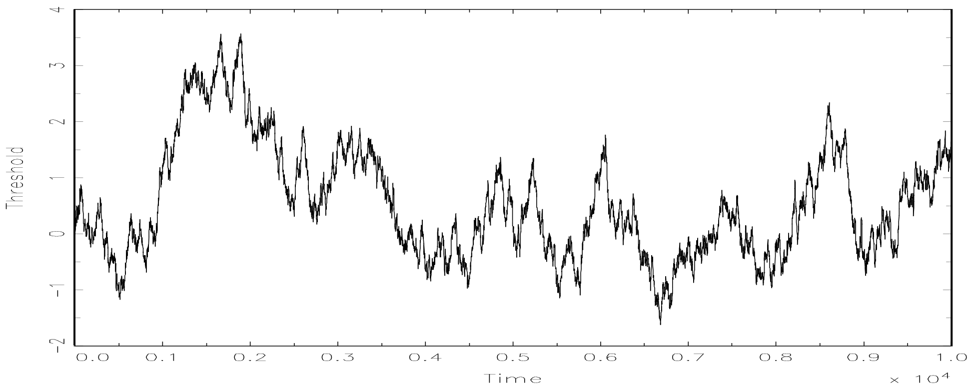


$a = \frac{-1.9}{2\ln\rho}$. For the threshold model we have used $\rho = 0.9$ and $r_T = 0.25 * T(\ln T)^{-1/2}\phi(\Phi^{-1}(1 - \ln T/T))$. Throughout $\sigma^2 = 1$. For comparison we have also simulated an ARFIMA(0,0.49,0) process. Again the variance of the i.i.d. process driving the ARFIMA process is $\sigma^2 = 1$. For all processes $T = 10000$. All noise terms are standard normal. In Figure 1, we present the autocorrelation functions as estimated from the data. In Figures 2-4, we present the process realisations. Clearly the processes are very persistent. Looking at the realisations, it is clear that the processes although covariance stationary look much more like a random walk than the ARFIMA process.

4 Conclusion

This short paper has illustrated the potential of persistent nonlinear autoregressive processes to produce long memory behaviour. We have extended in a number of directions the results of Diebold and Inoue (2001) who looked at

Figure 4:



simple (not autoregressive) Markov Switching models and found configurations that resulted in long memory. In particular, we have shown that simple nonlinear regime switching autoregressive models can exhibit long memory behaviour as long as the less persistent regime occurs increasingly rarely.

References

BERAN, J. (1997): *Statistics for Long Memory Processes.* Chapman & Hall.

DAVIDSON, J. (1994): *Stochastic Limit Theory*, Advanced Tests in Econometrics. Oxford University Press.

DAVIDSON, J., AND SIBBERTSEN (2002): "Generating Schemes for Long Memory Processes: Regimes, Aggregation and Linearity," *Mimeo, Cardiff University.*

DIEBOLD, F. X., AND A. INOUE (2001): "Long Memory and Regime Switching," *Journal of Econometrics*, 105, 131–159.

FALK, M. (1989): "A Note on Uniform Asymptotic Normality of Intermediate Order Statistics," *Annals of the Institute of Statistical Mathematics*, 41(1), 19–29.

GALAMBOS, J. (1978): *The Asymptotic Theory of Extreme Order Statistics*. Wiley.

KAPETANIOS, G., AND Y. SHIN (2002): "Testing for nonstationary long memory against nonlinear ergodic models," Unpublished Manuscript, University of Edinburgh.

TWEEDIE, R. L. (1975): "Sufficient Conditions for Ergodicity and Recurrence of Markov Chains on a General State Space," *Stochastic Processes Appl.*, 3, 385–403.

VAN DIJK, D., P. H. FRANCES, AND R. PAAP (2002): "A Nonlinear Long Memory Model with Application to US Unemployment," *Journal of Econometrics*, 110(2), 135–165.

Queen Mary
University of London

This working paper has been produced by
the Department of Economics at
Queen Mary, University of London

Department of Economics
Queen Mary, University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 5096
Fax: +44 (0)20 8983 3580
Website: www.econ.qmul.ac.uk/papers/wp.htm