Memory and infrequent breaks

Christian Gourieroux\textsuperscript{a}, Joann Jasiak\textsuperscript{b,\*}

\textsuperscript{a}CREST and CEPREMAP, Paris, France
\textsuperscript{b}Department of Economics, York University, 4700 Keele Street, Toronto, Ontario, Canada M3J 1P3

Received 9 December 1999; accepted 15 June 2000

Abstract

We study how processes with infrequent regime switching may generate a long memory effect in the autocorrelation function. In such a case, the use of a strong fractional I(d) model for economic or financial analysis may lead to spurious results. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Long memory; Switching regime; Heavy tail

JEL classification: C22

1. Introduction

Inference on the dynamics of economic or financial time series is usually based on the autocorrelation function whose decay pattern is used to assess the persistence range of processes. Those, displaying a geometric decay rate are modelled as Autoregressive Moving Averages whereas strong fractional I(d) models are used to fit hyperbolic decay rates of so-called long memory processes. However the analysis adequate for linear dynamics may often become misleading if the true underlying dynamics is nonlinear. This point is of special importance for the ‘long memory’ property, which is often observed in macroeconomic series (see, e.g. the study by Diebold and Rudebusch (1989)), and financial series of either volatility of returns (Ding et al., 1993; Andersen and Bollerslev, 1997), or intertrade durations (Gourieroux and Jasiak, 1998; Jasiak, 1999).

Even though the hyperbolic decay of the autocorrelation function [acf henceforth] of long memory processes is a stylized fact, it is widely known that the estimated decay rate may not be accurate, and that linear formulas of the type:

\[(1 - L)^d Y_t = \epsilon_t\]

\textsuperscript{*}Corresponding author. Tel.: +1-416-736-5083; fax: +1-416-736-5987.
E-mail address: jasiakj@yorku.ca (J. Jasiak).
where $\epsilon$ is a strong white noise, show poor performance in terms of nonlinear prediction criteria.

Following a recurrent idea (see Klemes, 1974; Balke and Fomby, 1989; Lobato and Savin, 1997; Granger, 1999; Granger and Hyung, 1999; Granger and Terasvirta, 1999; Diebold and Inoue, 1999), this paper aims to explain how infrequent stochastic breaks can create strong persistence in the estimated serial correlation. We develop two approaches. In Section 2 we consider a stationary switching regime model with rather large spells of each regime, and directly relate the decay rate of autocorrelations to the tails of the duration distributions. In Section 3, we introduce a dynamic model in which the switching probabilities tend to zero when the number of observations tends to infinity. The associated limiting model features a finite, but random number of regime switches. This limiting model is used to derive the asymptotic properties of the estimated autocorrelogram. We conclude in Section 4.

2. Switching regimes with large spells

In this section we describe a regime switching model with two regimes 0 and 1. We assume that the durations of the successive spells of regimes 0 and 1 are independent and identically distributed when they correspond to the same regime. Next, we study the dynamic properties of the binary process indicating regime 1, and especially the pattern of its autocorrelation function.

2.1. Assumption and notation

We introduce a time origin 0 and assume that the system is in regime 0 at this date. The sequence of durations spent in the two possible states is denoted by: $D_0^0, D_1^1, D_0^0, D_1^1, \ldots; D_j^0$ is the duration of the $j$th spell of regime 0, whereas $D_j^1$ is the duration of the $j$th spell of regime 1. The successive regime switching dates are: $\tau_1 = D_0^0, \tau_2 = D_0^0 + D_1^1, \tau_3 = D_1^1 + D_0^0 + D_1^1, \ldots; \tau_{2p} = \Sigma_{j=1}^p (D_j^0 + D_j^1)$.

We assume that the various durations $D_0^0, D_1^1, j$ varying are independent, and that the duration $D_j^0, j$ varying [resp. $D_j^1, j$ varying] follow the same distribution $F_0$ [resp. $F_1$]. These distributions are discrete with values in $N^*$.

The binary process $(Z_t, t \in N^*)$ is defined by:

$$Z_t = \begin{cases} 1, & \text{if there exist } p \text{ such that } \tau_{2p-1} < t \leq \tau_{2p} \\ 0, & \text{otherwise} \end{cases}$$

During a given time interval $\{1, \ldots, T\}$ there is a number $N_T$ of switching dates, $N_T^0$ of spells of regime 0, $N_T^1$ of spells in regime 1, with $N_T^0 + N_T^1 = N_T + 1$. The total time spent in regime 0 [resp. regime 1] is denoted by $A_T^0$ [resp. $A_T^1$] with $A_T^0 + A_T^1 = T$. Note that $A_T^0$ and $A_T^1$ are not necessarily sums of the basic durations due to the censoring effect of the window of observations.

2.2. The autocovariance function

We now discuss the pattern of the autocovariance function of the binary process $Z$. For this purpose we first consider the empirical autocovariance, and next its limit obtained when the number of observations tends to infinity.
The empirical autocovariance function is given by:

$$\hat{\gamma}_T(h) = \frac{1}{T} \sum_{t=h+1}^{T} Z_t Z_{t-h} - \frac{T-h}{T} \bar{Z}_T^2$$

The cross terms $Z_t Z_{t-h}$ take value one if and only if $t$ and $t-h$ both belong to spells of regime 1. We can distinguish these terms depending if they are in the same spell or in different spells of this type. We denote $(t, t-h) \in I_{j,T}$ if $t$ and $t-h$ belong to two spells of regime 0 separated by $j$ spells of regime 0. Then we get:

$$\hat{\gamma}_T(h) = \sum_{j=0}^{\infty} \left[ \frac{1}{T} \sum_{t,t-h \in I_{j,T}} Z_t Z_{t-h} \right] - \frac{T-h}{T} \bar{Z}_T^2$$

When the durations spent in regime 0 and 1 are rather large, i.e. the switching dates are not very frequent, we can expect that the second term in the decomposition is small with respect to the first one. Therefore we focus the attention on the first and the third terms, and denote by $R_T(h)$ the residual second term. We get:

$$\hat{\gamma}_T(h) = \frac{1}{T} \sum_{i=1}^{N_T} (D_i^1 - h)^+ + R_T(h) - \frac{T-h}{T} \left( \frac{A_T^1}{T} \right)^2$$

(1)

where $(D - h)^+ = \max(D - h, 0)$, or equivalently:

$$\hat{\gamma}_T(h) = \frac{N_T}{T} \sum_{i=1}^{N_T} (D_i^1 - h)^+ + R_T(h) - \frac{T-h}{T} \left( \frac{A_T^1}{T} \right)^2$$

(2)

Under standard ergodicity conditions, the various terms converge to constant limits:

$$\lim_{T \to \infty} \frac{N_T}{T} = \pi, \text{ i.e. the limiting switching probability}$$

$$\lim_{T \to \infty} \frac{N_T^{1/2}}{N_T^{1/2}} = \alpha^1, \text{ i.e. the limiting proportion of spells in regime 1}$$

$$\lim_{T \to \infty} \frac{A_T^1}{T} = \lim_{T \to \infty} \frac{1}{T} \sum_{i=1}^{N_T^{1/2}} D_i^1 = \lim_{T \to \infty} \frac{N_T}{T} \frac{1}{N_T^{1/2}} \sum_{i=1}^{N_T^{1/2}} D_i^1 = \pi \alpha E D^1$$

$$\lim_{T \to \infty} R_T(h) = R_\infty(h), \text{ (say)}$$

Therefore we get:

\footnote{We neglect the possible censoring of the last duration $D_{N_T^{1/2}}$, without loss of generality.}
\[ \gamma(h) = \lim_{t \to h} \hat{g}_t(h) \]
\[ = \pi \alpha_1 E(D^1 - h)^+ + R_\alpha(h) - \pi^2 \alpha_1^2 (ED^1)^2 \]  
(3)

Whenever \( R_\alpha(h) \) is small with respect to the first term \( \pi \alpha_1 E(D^1 - h)^+ \), the pattern of the autocovariance function is the same as the pattern of \( E(D^1 - h)^+ \) considered as a function of \( h \).

It is well known that: \( E(D^1 - h)^+ = \int_u^\infty \tilde{F}_1(u) \, du \), where \( \tilde{F}_1 \) is the survivor function associated with \( F_1 \), and that this quantity measures the magnitude of the tails of \( D^1 \). Therefore the pattern of the autocovariance function is directly related to the type of tails of the duration distribution \( F_1 \). For instance the autocovariance function has an hyperbolic decay if the survivor function \( \tilde{F}_1(u) \sim u^{-\delta} \) for large \( u \), i.e. if the duration distribution is of Pareto type.

The above reasoning is valid if the second term \( R_\alpha(h) \) is sufficiently small, i.e. when, intuitively, \( h \) is small with respect to the values of \( D^1 \). For example, if the expected duration of regime 0 is close to 200, we get an insight into the decay rate for \( h \leq 500 \), which corresponds to degree \( \delta \). This result completes the property by Heath et al. (1997), who investigated the limiting behaviour of the autocorrelation function when \( h \) tends to infinity. They prove that if \( F_1 \) is of Pareto type with tail parameter \( \delta \), \( 1 < \delta < 2 \), \( F_0 \) of Pareto type with tail parameter \( \delta^* > \delta \), then the autocorrelation function admits a hyperbolic decay of degree \( \delta - 1 \). We infer that the estimated autocorrelogram likely features a hyperbolic decay at a rate varying between \( \delta \) and \( \delta - 1 \) depending on the window selected for the lag.

2.3. Simulation

As an illustration we consider a simulated path of a binary series \( Z \), of length \( T = 29040 \). The duration distributions \( F_0 \) and \( F_1 \) are identical Pareto distributions with mean 200, and tail parameter 1.5. We provide in Fig. 1 the estimated autocorrelogram for \( h \leq 200 \), which clearly features a slow decay typical for long memory processes (see Granger and Terasvirta (1999) for a similar pattern with a model with endogenous switching regimes). Moreover we observe smooth waves for \( h \geq 200 \), indicating different hyperbolic rates of decay over varying ranges of lags.

Similar patterns are obtained when the regimes are perturbed by a noise. Let us consider the series defined by:

\[ Y_t = X_{0,t}(1 - Z_t) + X_{1,t}Z_t \]  
(4)

where \( X_{0,t} \), \( X_{1,t} \) are independent variables, independent of \( Z_t \), with gaussian distributions: \( X_{0,t} \sim N[m_0, \sigma^2_0] \), \( X_{1,t} \sim N[m_1, \sigma^2_1] \). We get: \( Y_t = m_0 + (m_1 - m_0) Z_t + \sigma_0 U_{0,t}(1 - Z_t) + \sigma_1 U_{1,t}Z_t \), where

---

2This aspect is discussed in Appendix 1 where we compute explicitly the term corresponding to \( I_{1,T} \).

3Note that for financial tick by tick data, autocorrelations are often computed up to lag 1000 or 2000. In our case, where average intertrade durations are very short, this would span two or three trading days in calendar time.

4Note that the above results are different from the properties derived by Cioczek-Georges and Mandelbrot (1995), Taqqu and Levy (1986), Taqu et al. (1997) and Willinger et al. (1997), who derive the fractional Brownian motion by averaging renewal processes. In our case, the results are derived from one single path.
$U_0, U_1, \ldots$ are the associated standard gaussian variables. The quantitative series $Y$ will likely share some properties of temporal dependence of $Z$ if $m_0 \neq m_1$. To illustrate this point we provide in Fig. 2 the autocorrelation function [acf, henceforth] of $Y$ for $m_0 = 0, m_1 = 3, \sigma_0 = \sigma_1 = 1$, and in Fig. 3 the acf of $Y$ and $Y^2$ for $m_0 = m_1 = 0, \sigma_0 = 1, \sigma_1 = 3$.

We see that the long memory effect is still observed in the presence of breaks in the mean, and it disappears when breaks occur in the variance. Indeed in the latter case the $Y$ process is a martingale difference sequence and therefore a weak white noise. However the persistence effect is observed from the acf of its squares. As noted by Granger and Marmol (1998), 'a frequent property of data, particularly in the financial area, is that the correlogram is low but remains positive for many lags'. To take into account this stylized fact, Granger and Marmol (1998) proposed to add a simple noise to a long memory process, while Ghysels et al. (1997) introduced nonlinear unobservable factors. Figs. 2 and 3 show that noisy infrequent breaks can create this specific effect. This phenomenon can easily be explained. Indeed, we obtain:
Fig. 2. Autocorrelogram for switching mean.

\[
\text{Cov}(Y_t, Y_{t-h}) = \text{Cov}[E(Y_t|Z), E(Y_{t-h}|Z)] + E\text{Cov}(Y_t, Y_{t-h}|Z)
\]
\[
= \text{Cov}[(m_0 - m_1)Z_t, (m_0 - m_1)Z_{t-h}]
\]
\[
= (m_0 - m_1)^2 \text{Cov}(Z_t, Z_{t-h}), \text{ for } h \neq 0
\]

\[
VY_t = VE(Y_t|Z) + E\text{V}(Y_t|Z)
\]
\[
= \left[(m_0 - m_1)^2 + (\sigma_0^2 - \sigma_1^2)^2\right]VZ_t
\]

Therefore the autocorrelation function of \( Y \) is proportional to the autocorrelation function of \( Z \), for \( h \geq 1 \), with a factor related to the ratio \( (m_0 - m_1)^2/(\sigma_0^2 - \sigma_1^2)^2 \).

3. Asymptotics for small switching probabilities

In this section we study asymptotic properties of the empirical autocovariance function when the switching probabilities are small. We consider a Markov transition model, for a finite number of
observations, which allows us to focus on the effect of infrequent switches on memory, and provides a tractable limit model.

3.1. The finite sample model and the limiting model

For a finite sample of size $T$, we consider a standard transition model, where $p_T^j$, $j = 0, 1$ denotes the transition probability from state $j$ to the complementary state between consecutive dates. These probabilities depend on $T$ as well as on the corresponding distributions of durations of the two regimes. To emphasize this dependence, we index by $T$ the various variables, i.e. $\tau_{\alpha,T}, D^0_{j,T}, D^1_{j,T} \ldots$. The duration variables are independent with geometric distributions of parameters $p_0^T$ and $p_1^T$, respectively. Under the assumption:

\[ \lim_{T \to \infty} Tp_0^T = \lambda_0, \quad \lim_{T \to \infty} Tp_1^T = \lambda_1 \]  

(A.1.)

it is known that the finite sample process tends in distribution to a limiting process after an appropriate change of time. More precisely the normalized durations $D^0_{j,T}/T$ and $D^1_{j,T}/T$ tend in distribution to limiting duration variables $D^0_{j,\infty}, D^1_{j,\infty}$, which are independent with exponential distributions with respective parameters $\lambda_0, \lambda_1$. The normalized observation period becomes $[0,1]$, whereas the limiting

\[ ^5 \text{See Granger and Hyung (1999) for a similar assumption.} \]
variables \( \lim_{T \to \infty} \tau_{n,T}/T = \tau_{n,\infty} \), \( \lim_{T \to \infty} N_T = N_{\infty} \), \( \lim_{T \to \infty} A_T^T/T = A_{\infty} \) are deduced from the limiting durations as they were in the finite sample model.

Note that in the limiting model the number of switching dates \( N_{\infty} \) is finite, but stochastic. For instance in the special case \( \lambda_1 = \lambda_0 \), the distribution of \( N_{\infty} \) is a Poisson distribution with a parameter \( \lambda_1 = \lambda_0 \).

3.2. Limiting behaviour of the autocorrelogram

We study directly the asymptotic properties of the autocorrelogram or equivalently of the empirical autocovariance function. In fact \( \hat{\gamma}_r(h) \) tends in distribution to a limiting variable \( \gamma_\infty(h) \) which is not degenerate. This variable admits a mixture distribution, which takes into account the limiting number of switching dates. More precisely let us consider the first possible number of regimes and distinguish the cases of fixed and large lags \( h \).

(i) If there is only one limiting regime \( N_{\infty} = 0 \), with probability \( \pi_0 \), say, we get:

\[
\hat{\gamma}_r(h) = 0, \forall h, \text{ and also } \gamma_\infty(h) = 0, \forall h
\]

(ii) If there are two limiting regimes, \( N_{\infty} = 1 \), with probability \( \pi_1 \), we get:

\[
\hat{\gamma}_r(h) = \frac{1}{T}(T - D_{1,T}^0 - h)^+ - T - h \left( \frac{T - D_{1,T}^0}{T} \right)^2
\]

\[
= \left( 1 - \frac{D_{1,T}^0}{T} - \frac{h}{T} \right)^+ - \left( 1 - \frac{h}{T} \right) \left( 1 - \frac{D_{1,T}^0}{T} \right)^2
\]

- If \( h \) is fixed, we deduce that:

\[
\lim_{T \to \infty} \hat{\gamma}_r(h) = (1 - D_{1,\infty}^0) D_{1,\infty}^0
\]

The limiting pattern corresponds to a constant function with a stochastic level.

- If \( h \) is large proportional to \( T \), \( h_T = \alpha T \), we get:

\[
\lim_{T \to \infty} \hat{\gamma}_r(\alpha T) = (1 - D_{1,\infty}^0 - \alpha)^+ - (1 - \alpha)(1 - D_{1,\infty}^0)^2,
\]

or:

\[
\hat{\gamma}_r(h) \sim \left( 1 - D_{1,\infty}^0 - \frac{h}{T} \right)^+ - \left( 1 - \frac{h}{T} \right) \left( 1 - D_{1,\infty}^0 \right)^2
\]

The limiting pattern corresponds to a piecewise linear function in the lag, featuring a rather slow decay which may be confused with a hyperbolic decay in practice.

- (iii) If there are three limiting regimes, \( N_{\infty} = 2 \), with probability \( \pi_2 \), we get:

\[
\lim_{T \to \infty} \hat{\gamma}_r(h) = D_{1,\infty}^1(1 - D_{1,\infty}^1)
\]

\[
\lim_{T \to \infty} \hat{\gamma}_r(\alpha T) = (D_{1,\infty}^1 - \alpha)^+ - (1 - \alpha)(D_{1,\infty}^1)^2
\]
And so on. It is important to note that the limiting distributions depend on the initial regime fixed at 0 in our computations.

3.3. Simulation

To illustrate the limiting behaviour of the empirical autocovariance we have simulated a series of length $T = 3000$, with switching probabilities $p^0 = p^1 = 0.005$. Despite the large number of observations, the limiting distributions of the first and second order moments are not reduced to a point mass (Figs, 4 and 5).

The effect of state transitions on the pattern of the autocorrelogram and also on the estimated fractional degree can be observed by considering the joint distribution of estimated autocorrelations at two different lags (Fig. 6).

These limiting distributions may explain the large variation of the estimator of fractional coefficient for high frequency data, where the $d$ parameter is estimated by rolling, since the estimated value is very sensitive to the subsample it is computed from.

Fig. 4. Distribution of the mean.
4. Concluding remarks

To discuss the relation between infrequent breaks and long memory, we focused our study on the estimated autocorrelogram, rather than on the estimation of the fractional degree from the autocorrelogram. Indeed the main difficulties encountered in practice can be due to the estimated autocorrelogram itself, and not to the approach used to derive the parameter $d$. We first observe that long memory may be found in nonlinear time series with infrequent breaks, and that, in the limiting case of very small switching probabilities, the estimated autocorrelogram may converge to a nondegenerate distribution.

Therefore the hyperbolic decay rate of the estimated autocorrelograms may result from nonlinear dynamics with infrequent switching regimes and not from linear fractional dynamics with i.i.d. innovations.

The choice between these specifications is important, for instance, when we consider nonlinear predictions. If a fractional model is retained, the predictions are based on a long history of the observed series; if a nonlinear model with infrequent switches is selected, the predictions will be computed from moving average techniques based on short memory to retain observations from the same regime with a large probability. This approach is common in applied finance (see e.g. the computation of the Value at Risk proposed by the Basle Committee) and robust to infrequent breaks.
Finally note that it is a common practice to round up the series to a given number of decimals either for computational ease, or due to market rules, for example, in asset prices which are multiples of a fixed tick. These rounding procedures introduce artificial switching regimes, which may create spurious long memory in the observed series.

Acknowledgements

The second author gratefully acknowledges financial support of the Natural Sciences and Engineering Research Council of Canada.

Appendix 1. Second term of the autocovariance expansion

Let us consider the term \( \frac{1}{T} \sum_{t, t-h \in I_{t, t}} Z_{t} Z_{t-h} = A_{1, T} \) (say).

We have:
\[ Z_t Z_{t-h} = 1 \]

\( \Leftrightarrow \exists i: \tau_{2i-1} \leq t - h \leq \tau_{2i-1} + D_i^1 \text{ and} \)

\[ \tau_{2i-1} + D_i^1 + D_{i+1}^0 \leq t \leq \tau_{2i-1} + D_i^1 + D_{i+1}^0 + D_{i+1}^1 \]

\( \Leftrightarrow \exists i: h \leq t - \tau_{2i-1} \leq h + D_i^1 \text{ and} \)

\[ D_i^1 + D_{i+1}^0 \leq t - \tau_{2i-1} \leq D_i^1 + D_{i+1}^0 + D_{i+1}^1 \]

\( \Leftrightarrow \exists i: t - \tau_{2i-1} \in [\max(h, D_i^0 + D_i^1), \min(h + D_i^1, D_i^1 + D_{i+1}^0 + D_{i+1}^1)] \)

We deduce (up to the final truncation effect which is negligible):

\[ A_{1,T} = \frac{1}{T} \sum_{i=1}^{N_1} [\min(h + D_i^1, D_i^1 + D_{i+1}^0 + D_{i+1}^1) - \max(h, D_i^0 + D_i^1)]^+ \]

This quantity tends to:

\[ A_{1,T} = \pi \alpha_i E[\min(h + D_i^1, D_i^1 + D_{i+1}^0 + D_{i+1}^1) - \max(h, D_i^0 + D_i^1)]^+ \]

If the lag \( h \) is small with respect to the values taken by the durations spent in regime 0, we get:

\[ A_{1,T} \sim \pi \alpha_i E[h + D_i^1 - (D_i^0 + D_i^1)]^+ = \pi \alpha_i E[h - D_i^0]^+ \sim 0 \]

References


