# Regime switching model: real or spurious long memory?

Dominique Guegan, Stéphanie Rioublanc

# Regime switching models : real or spurious

# long memory ?

Dominique GUEGAN, IDHE

Stéphanie RIOUBLANC, IDHE

**2005.100**

# Regime Switching Models: Real or Spurious Long Memory?

Guégan Dominique [*]        Rioublanc Stéphanie [†]

**Abstract**

In this paper, we analyze the possible confusion in terms of long memory behavior of the autocorrelation function of a Markov switching model. Such a model is known to have a short memory behavior. Analyzing the value of sum of the transition probabilities and the number of switches inside such a model, we show their impact to create long memory. The ability of the true Markov switching model to predict is compared with the forecasts obtained from a long memory process adjusted on data derived from the former model. It is shown that, in certain cases, this spurious long memory behavior can be benefit to get better forecasts.

*JEL classification:* C22.

*Keywords:* Markov switching models, Structural breaks, Long memory behavior, ARFIMA models, Forecasting.

[*]ENS-Cachan, IDHE-MORA, UMR, CNRS 8534, 61 avenue du Président Wilson, 94235 Cachan Cedex, France, guegan@ecogest.ens-cachan.fr.

[†]Banque de France, DEER, Centre de recherche 41-1391, 31 rue Croix des Petits Champs, 75001 Paris, Tél. : + 33 (1) 42 92 49 95, stephanie.rioublanc@banque-france.fr, ENS-Cachan, IDHE-MORA, UMR, CNRS 8534, 61 avenue du Président Wilson, 94235 Cachan Cedex, France.

# 1 Introduction

From a long time, structural breaks have been observed in many economic and financial time series. Recently, a huge literature proposing models capable to capture existence of changes' structure has been developed. One of the most popular models able to take into account sudden changes in times series being the Markov Switching model introduced by Hamilton (1988). Nevertheless, other competitive models have also been developed, see for instance Granger and Terasvirta (1999), Diebold and Inoue (2001), Breidt and Hsu (2002) or Granger and Hyung (2004). All these models point out the fact that they can exhibit also long memory in covariance sense or in Allan's sense (Allan, 1966) with respect to specific assumptions on their parameters or only through simulations, see Guégan (2004) for some review on this kind of situation. Or one of the characteristic of these different models is the quick decay of their autocorrelation function, which implies a short memory behavior and thus a good capability to predict for short term prediction and a bad performance for the long term forecast prediction. Thus, if it is possible to observe a slow decay of the autocorrelation for these models although they are short memory, some confusion can arise in terms of modelling. The possible confusion between long memory and existence of jumps inside models with short memory behavior stands in the fact that a slow decay of the empirical autocorrelation function would not necessarily indicate the presence of long memory, but might be due to breaks. This long memory behavior property in the data due to structural breaks or regime switches is called "spurious long memory" in some papers. Diebold and Inoue (2001), in their paper mention that it is the case for the switching model if we impose some specific assumptions on the transition probabilities.

We consider, in this paper a particular Markov switching model and we show, without specific assumptions on the transition probabilities related to the size of the sample, that through the simulations this model exhibits a long memory behavior in the covariance sense and can be confused with a FI(d) model (see Granger and Joyeux, 1980 and Hosking, 1981) in terms of modelling. But we will see also that this confusion is not as prejudicial as we can think because, in certain cases that we specific latter, the forecasts are better using the spurious model in place of the true model.

To illustrate our approach, we investigate the sum of the transition probabilities and the occurrence in states changes. Our intuition is that according to the means values and the sum of the transition probabilities, we can observe different behaviors of the autocorrelation function. We also think that the more the process switches from one state to the other, the more the decay of the autocorrelation function becomes quick. In our simulations, we increase the number of times for which the series changes from one state to the other through a panel of transition probabilities in order to check those ideas.

The paper is organized as follows. Section 2 introduces the models explored in the paper and recalls some properties of these models. In Section 3, through a simulation experiment, we analyze graphically the autocorrelation functions' behavior in order to detect when this Markov switching model exhibits short or long memory. In Section 4, we adjust a FI(d) process to the previous simulated Markov switching processes and analyze the estimations obtained for the long memory parameter $d$. In Section 5, we compare the forecast performance of the estimated FI($d$) model and the true Markov switching model. Section 6 concludes.

3

## 2 Real or spurious long memory?

As we specify in the introduction, it has been observed that stationary models with jumps or specific states which are theoretically short memory in covariance sense can exhibit long memory. Here, to illustrate this situation, we use a simple Markov switching model and we analyze the role of its different parameters to create long memory behavior. In this Section, we specify the notion of long memory we consider and the model with which we work all along this paper.

There exist different criteria to define the existence of long memory behavior inside real data, see Guégan (2004) for a survey. The definition we use here characterizes the long memory behavior in terms of asymptotic decay of the autocorrelation function.

**Definition 1** *Let $(X_t)_t$ be a stationary process for which the following holds: there exists a real number $d \in ]0, 1/2[$ and a constant $C > 0$ such that the autocorrelation function $\Gamma(h)$ satisfies*

$$\lim_{h \to \infty} \Gamma(h) = Ch^{2d-1}.$$

*Then $(X_t)_t$ is a stationary process with a long memory behavior in covariance.*

The autocorrelation function of such a process $(X_t)_t$ has an asymptotic hyperbolic decay. Empirically it is very difficult to observe such asymptotic decay, thus in this article, for convenience and because we work with finite samples, we assume that if a process has an empirical autocorrelation function $\widehat{\Gamma}_X(h)$, for which $\widehat{\Gamma}_X(h) \neq 0 \ \forall h > 10$, then the series has a long memory behavior, otherwise, we assume that it has a short memory behavior.

Now we specify the Markov switching model under study. Let $(X_t)_t$ the process defined by the following equation, $\forall t$:

$$X_t = \begin{cases} \mu_1 + \varepsilon_t & \text{if } s_t = 1, \\ \mu_2 + \varepsilon_t & \text{if } s_t = 2. \end{cases} \tag{1}$$

In the model (1), there exist two states in level which represent expansion and recession for instance in a time series. The process $(s_t)_t$ is an hidden ergodic Markov chain, characterized by its transition matrix $P$, whose elements are the fixed transition probabilities $p_{ij}$, defined $\forall i, j = 1, 2$, by:

$$p_{ij} = P[s_t = j | s_{t-1} = i], \quad 0 \leq p_{ij} \leq 1, \quad \sum_{i,j=1}^{2} p_{ij} = 1. \tag{2}$$

The process $(X_t)_t$ switches from level $\mu_1$ to level $\mu_2$ with respect to this Markov chain. The process $(\varepsilon_t)_t$ in (1) is a centered Gaussian strong white noise with variance one, independent of the Markov chain $(s_t)_t$. The autocorrelation function of the Markov switching process (1) decreases exponentially towards zero. Indeed:

**Lemma 1** *The autocorrelation function $\Gamma(h)$ of the process $(X_t)_t$ defined by (1) is equal to, $\forall h$:*

$$\Gamma(h) = \frac{(\mu_1 - \mu_2)^2 (1 - p_{11})(1 - p_{22}) \rho^h}{(2 - p_{11} - p_{22})^2 [\pi_1 \mu_1^2 + \pi_2 \mu_2^2 + 1 - (\pi_1 \mu_1 + \pi_2 \mu_2)^2]}, \tag{3}$$

*where*

$$\rho = -1 + p_{11} + p_{22}, \tag{4}$$

*and, $\pi_1 = \frac{1 - p_{22}}{2 - p_{11} - p_{22}}$ and $\pi_2 = \frac{1 - p_{11}}{2 - p_{11} - p_{22}}$ are the non conditional probabilities.*

The autocorrelation function $\Gamma(h)$, can be rewritten as:

$$\Gamma(h) = C_{\mu_i, p_{ii}} \, \rho^h, \quad i = 1, 2,$$

with

$$C_{\mu_i, p_{ii}} = \frac{(\mu_1 - \mu_2)^2 (1 - p_{11})(1 - p_{22})}{(2 - p_{11} - p_{22})^2 [\pi_1 \mu_1^2 + \pi_2 \mu_2^2 + 1 - (\pi_1 \mu_1 + \pi_2 \mu_2)^2]}, \quad i = 1, 2.$$

Thus, the autocorrelation function's decay of the model (1) depends on the means levels $\mu_i$ and the transition probabilities $p_{ii}$, $i = 1, 2$. Its convergence's rate is in $\rho^h = (-1 + p_{11} + p_{22})^h$ and $\rho$ varies between -1 and 1 according to the values of the transition probabilities $p_{ii}$, $i = 1, 2$. Thus, in case of two high transition probabilities, $\rho$ is close to 1, and the autocorrelation function decreases slowly. When $p_{11} + p_{22}$ is close to 1, the decay of the autocorrelation function is quicker because $\rho$ is close to 0: this situation can arise even if the two transition probabilities are weak. Now, it is important to remark that the values chosen for the transition probabilities $p_{ii}$ influence the number $n_{s_t}$ of changes inside the two states: this is this number which influences the behavior of the autocorrelation function. It does not exist a fairly relationship between this number and the transition probabilities, so in order to understand this link we are going to make some simulations making varying the $p_{ii}$ and thus the $n_{s_t}$, inside the model (1). This will permit us to determine some range of values for which the long memory behavior is observed. Then, we will measure this behavior adjusting a long memory process on the simulated data and we will compare the forecasting performances both from the true Markov switching and the estimated long memory process.

## 3 How to create a long memory behavior from the model (1) ?

In order to investigate the autocorrelation function's behavior of simulated Markov switching processes such (1), we make varying the transition prob-

abilities $p_{ii}$, $i = 1,\ 2$, and the levels $\mu_i$. All along this Section, we impose

$$p = p_{11} = p_{22}. \tag{5}$$

In the expression (4), it is the sum of the transition probabilities $p_{ii}$, $i = 1,\ 2$ which appears and thus this sum determines the value of $\rho$. This sum will be an important task for the following because it influences the behavior of the autocorrelation function of model (1). Indeed, according to the value of $\rho \in\ ]$-$1,\ 1[$, when $h \to \infty$, $\rho^h$ decreases slowly, quickly or switches constantly between positive and negative values. We are also interested to examine the influence of the values of $n_{s_t}$ on the decay of the autocorrelation function. Even if the relationship between this number $n_{s_t}$ and the transition probabilities $p_{ij}$, $i, j = 1,\ 2$ has not a clear analytical form, the intuition is: the more the transition probabilities $p_{ij}$ are high, the less the series switches from one state to the other, and then the weak $n_{s_t}$ is. In the literature, it has been mentioned that the larger $n_{s_t}$ is, the quicker the autocorrelation function's decay is, see Granger and Terasvirta (1999), Gourieroux and Jasiak (2001) and Granger and Hyung (2004). Here we specify this kind of intuition.

The simulations' experiment is the following: for two pairs of levels: $(\mu_1,\ \mu_2) = (5, -5)$ and $(\mu_1,\ \mu_2) = (0.5, -0.5)$, and 3 sample sizes: $T = 1000$, $T = 5000$ and $T = 25000$, making varying the transition probability from $p = 0.99$ to $0.01$, decreasing $0.01$ at each step, we simulate a lot of models (1). For each $p$ defined in (5), we get a state vector $s_t$ which provides a number $n_{s_t}$. At each step, we investigate the autocorrelation function of the simulated series in order to study how the quantity $p_{11} + p_{22}$ and the number $n_{s_t}$ influence the asymptotic behavior of the autocorrelation function of the model (1). We analyze below the results obtained for $T = 1000$. The results obtained for $T = 5000$ and $T = 25000$ are nearly similar and given in Appendix 7.1, see
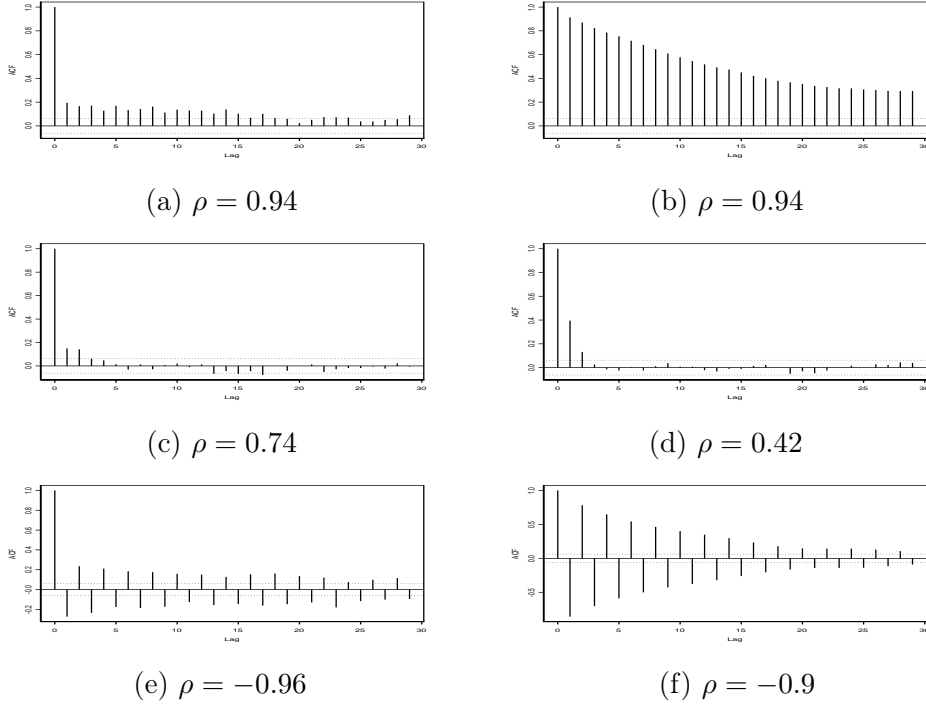
(a) $\rho = 0.94$      (b) $\rho = 0.94$

(c) $\rho = 0.74$      (d) $\rho = 0.42$

(e) $\rho = -0.96$      (f) $\rho = -0.9$

Figure 1: Behaviors of the autocorrelation functions of some simulated series issued from model (1), with respect to $\rho$. Left column: $(\mu_1, \ \mu_2) = (0.5, -0.5)$ and right column: $(\mu_1, \ \mu_2) = (5, -5)$.

Tables 5 and 6. Empirically, we observe that when $T = 1000$, and $p$ varies from 0.99 to 0.01, then the number of switches $n_{s_t}$ varies from 7 to 990. The below analysis is only graphical.

1. $(\mu_1, \ \mu_2) = (0.5, -0.5)$.

   - When $p$ varies from 0.99 to 0.91, then $0.82 \leq \rho \leq 0.98$ and $n_{s_t}$ varies from 7 to 105. We observe graphically on Figure 1 (a), a slow decay of the autocorrelation function, and thus some long memory behavior.

   - For $p$ varying from 0.9 to 0.1, then $-0.8 \leq \rho \leq 0.8$ and $n_{s_t}$ is

8

such that: $112 \leq n_{s_t} \leq 889$. The behavior of the autocorrelation functions of the simulated series on Figure 1 (c), is close to a short memory behavior, although some lags remain slightly significant.

- When $0 < p < 0.1$, then $-1 < \rho < -0.8$ and $n_{s_t}$ varies from 920 to 990. In that case, some kind of seasonality appears on the autocorrelation functions on Figure 1 (e), which is probably created by the fact that $\rho$ is close to -1.

- Thus, it appears that the model (1) exhibits empirically a long memory behavior when $0.8 < \rho < 1$ and a short memory behavior when $-0.8 < \rho < 0.8$.

2. $(\mu_1, \, \mu_2) = (5, -5)$.

- When $p$ varies from 0.99 to 0.81, this corresponds to $0.62 \leq \rho \leq 0.98$ and $7 \leq n_{s_t} \leq 192$. The autocorrelation functions decrease slowly.

- When $p$ varies from 0.8 to 0.7, then $0.4 \leq \rho \leq 0.6$ and $208 \leq n_{s_t} \leq 285$. Various situations arise. Although most of the autocorrelation functions decrease quickly towards 0, some others still decline slowly.

- When $0.7 < p < 0.76$, which corresponds to $0.4 < \rho < 0.52$ and $247 \leq n_{s_t} \leq 297$, the autocorrelation functions given on Figure 1 (d) decrease very quickly and this decreasing rate increases when $0.25 < p < 0.7$, i.e. when $309 \leq n_{s_t} \leq 750$ and $-0.5 < \rho < 0.4$.

- When $0 < p < 0.24$, then $-1 < \rho < -0.52$ and $760 \leq n_{s_t} \leq 990$, and we observe some kind of seasonality inside the autocorrelation functions.

- In summary, for $\rho > 0.6$, we are in presence of a slow decay of the autocorrelation function, for $-0.5 < \rho < 0.5$, the autocorrelation function exhibits a quick decay, and for $\rho < -0.5$, we create seasonality in the autocorrelation function as Figure 1 (f) shows.

3. When we compare the results obtained for these two classes of levels: $(\mu_1, \ \mu_2) = (0.5, -0.5)$ and $(\mu_1, \ \mu_2) = (5, -5)$, we see that for fixed transition probabilities $p$, there exists a slower decay of the autocorrelation functions of the model (1) in the latter case. This difference of behavior can be observed on Figures 1 (a) and (b). Thus, the levels $(\mu_1, \ \mu_2)$ have also an impact on the convergence's speed of the autocorrelation function of the model (1).

In summary, by varying $p$, and thus $\rho$, we show evidence of the influence of the sum of the transition probabilities and of the means' values $\mu_i, i = 1, \ 2$ on the asymptotic behavior of the autocorrelation function of the model (1). The parameter $\rho$ is not the only cause of the speed convergence of the autocorrelation function of the model (1). This is obvious from the expression (3). Indeed, for a fixed value of $p$ defined in (5), if $(\mu_1, \ \mu_2) = (0.5, -0.5)$, then $C = 0.2$, and for $(\mu_1, \ \mu_2) = (5, -5)$, then $C \simeq 0.96$. Or $\Gamma(h) = C\rho^h$, thus, the autocorrelation function decreases quicker in the former case than in the latter case. This explains also, when $\rho$ is close to - 1, why the autocorrelation function can exhibit some seasonnality.

Now, if we analyze the results with respect to the $n_{s_t}$ values, when the shifts are rare, we observe that the empirical autocorrelation functions of the model (1) decrease slowly. For a sample size $T = 1000$, this behavior remains until $n_{s_t} = 105$ for $(\mu_1, \ \mu_2) = (0.5, \ -0.5)$ and until $n_{s_t} = 192$ for

10

$(\mu_1, \ \mu_2) = (5, \ -5)$. In both cases, compared to the sample size $T = 1000$, these values of $n_{s_t}$ correspond to a great number of shifts. This appears in contradiction with some remark made in the literature saying that only a small number of shifts provoke long memory, see Diebold and Inoue (2001) for instance. Our intuition is that weak transition probabilities can create the phenomenon of anti-persistence inside the data and high transition probabilities create a long memory behavior. To characterize the notion of high transition probabilities is difficult because other parameters provoke also this long memory behavior.

All the results discussed previously are summarized in Table 1. For $T = 5000$ and for $T = 25000$, the results are reported in Tables 5 and 6 in the Appendix 7.1.

| $\rho$ | -1 | -0.82 | -0.52 | 0.6 | 0.8 | 1 |
|:---:|:---:|:---:|:---:|:---:|:---:|:---:|
| (0.5, -0.5) | AP | | SM | | | LM |
| (5, -5) | AP | | SM | | LM | |

Table 1: Autocorrelation functions behavior for the model (1) according to the values of $\rho$ and $(\mu_1, \ \mu_2)$, for $T = 1000$. LM stands for long memory, SM for short memory and AP for anti-persistence.

# 4  Estimation of the long memory parameter $\widehat{d}$

In order to measure the existence of long memory behavior inside the simulated previous models (1), in this Section we adjusted on these simulated

11

data sets, a FI(d) process defined by:

$$(1 - B)^d X_t = \varepsilon_t, \tag{6}$$

where $B$ represents the lag operator, $(\varepsilon_t)_t$ a strong white noise. The fractional difference operator $(1 - B)^d$, for $d \in \mathbb{R}$, being defined by:

$$(1 - B)^d = \sum_{i=0}^{\infty} \left( \begin{smallmatrix} d \\ i \end{smallmatrix} \right) (-B)^i. \tag{7}$$

We proceed by Monte-Carlo experiment. For each sample size $T = 1000$, $5000$ and $25000$ and each transition probability $p$ defined in (5), decreasing from 1 to 0.01, by step 0.01, and two sets of levels $(\mu_1,\ \mu_2) = (0.5, -0.5)$ and $(\mu_1,\ \mu_2) = (5, -5)$, we replicate 100 simulations of processes issued from the model (1). For each experiment, we fit a FI(d) process defined by equations (6)-(7), and in fine, we retain the estimated parameter $\widehat{d}$ obtained by averaging all the estimated parameters. The long memory parameter $d$ is estimated using the Whittle approach, see Yajima (1985).

On Figure 2, we exhibit the estimated parameters $\widehat{d}$ with their confidence interval for the sample size $T = 1000$. We can remark:

- For $p$ varying from 0.99 to nearly 0.5, which corresponds to $7 \leq n_{s_t} \leq 510$, the estimated parameter $\widehat{d}$ is positive. Nevertheless, we observe differences in the estimation according to the means values $\mu_i$. Indeed, for $(\mu_1,\ \mu_2) = (0.5, -0.5)$, we have $0 < \widehat{d} < 0.5$; for $(\mu_1,\ \mu_2) = (5, -5)$, $\widehat{d}$ is positive, but can be greater than 1. The confidence interval are very large.

- For $(\mu_1,\ \mu_2) = (0.5, -0.5)$, $\widehat{d} = 0$ for $p$ varying from 0.55 to 0.53, which corresponds to $n_{s_t}$ varying from 460 to 473 and $0.06 \leq \rho \leq 0.1$. For $(\mu_1,\ \mu_2) = (5, -5)$, $\widehat{d} = 0$ only for $p = 0.5$, which corresponds to $n_{s_t} = 510$ and $\rho = 0$.

12
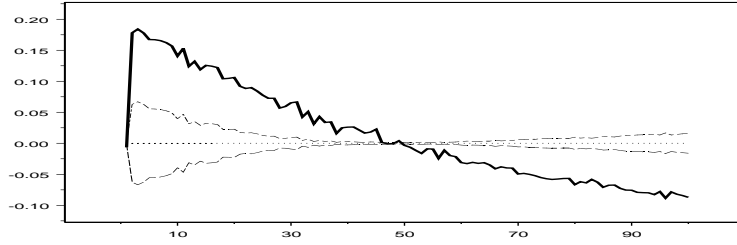
(a) $(\mu_1, \ \mu_2) = (0.5, -0.5)$.
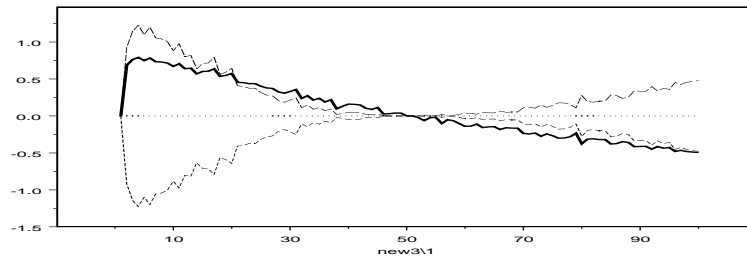


(b) $(\mu_1, \ \mu_2) = (5, -5)$.

Figure 2: Estimated memory parameters $\widehat{d}$ (solid line) and their confidence interval (dotted line) for the switching model (1). The abscissa represent the transition probability $p$ from 1 to 0.01 with a step of 0.01 and $T = 1000$.

- For $p < 0.5$, that is when $\rho < 0$ and $n_{s_t} \geq 518$, then $\widehat{d}$ is negative. This range of values characterizes existence of anti-persistence.

- Notice that the switch from long memory to anti persistence, looking at the values of $\widehat{d}$, occurs for the same transition probabilities in both cases.

- We obtain similar results for $T = 5000$ and $T = 25000$, see Figure 5 in Appendix 7.2.

To apply a FI(d) model on the simulated switching process (1) permits to confirm that, for some range of transition probabilities, some long memory behavior can be detected and measured. Thus, even if the Markov switching

13

model can be classified as a short memory process, existence of spurious long memory is detected. The value $\widehat{d} = 0$ is obtained in few cases. This method reveals also that it is possible to get some $\widehat{d}$ which are greater than 0.5. This should imply that the data are non stationary, but we know that they are issued from a stationary process.

Now, we can think that this spurious long memory, detected by the previous method, can have dramatic incidence when we make such mistake on real data, particularly if we need to make forecasts from the observed data sets. We detail such incidence in the next Section.

## 5 Forecasting

In this Section, we compare the forecasting performances of the true Markov switching model (1) and the estimated FI(d) model, for simulated data sets issued from model (1), for two pairs of levels $(\mu_1, \ \mu_2) = (5, -5)$ and $(\mu_1, \ \mu_2) = (0.5, -0.5)$, and for different transition probabilities $p$, using $T = 1000$.

To forecast data which exhibit long memory in the covariance sense is now well documented in the literature for FI($d$) processes or Gegenbauer processes, see for instance Guégan and Ferrara (2001). To compute the predictor $\widehat{X}_{t+h|h}$ from the FI($d$) model, we use an expression derived from (6) using the estimated parameter $\widehat{d}$.

Forecasting with Markov switching model is a more complicated work. Following the works of Krolzig (1997), we define the optimal predictor $\widehat{X}_{t+h|h}$

14

for the model (1) as:

$$\widehat{X}_{t+h|h} = \mu_1\pi_1 + \mu_2\pi_2 + (\mu_1 - \mu_2)\rho^h\widehat{\zeta}_{t|t} \tag{8}$$

where $\widehat{\zeta}_{t|t}$ is the vector of filtered regime probabilities $P[s_t = i|X_t, \ldots, X_0]$, $i = 1, 2$, based on information up to time $t$. In practice, the vector $\widehat{\zeta}_{t|t}$ can be calculated recursively following the algorithm given in Hamilton (1988). We obtain them by iterating the following equation:

$$\widehat{\zeta}_{t|t} = \frac{(\widehat{\zeta}_{t|t-1} \odot \eta_t)}{1_2'(\widehat{\zeta}_{t|t-1} \odot \eta_t)},$$

where $\odot$ represents the element-by-element multiplication, $1_2 = (1\ 1)'$ and $(\eta_t)_t$ is a process whose components are the conditional densities of the process $(X_t)_t$:

$$\eta_t = \begin{pmatrix} \frac{1}{\sqrt{2\pi}} \exp \frac{-(X_t - \mu_1)^2}{2} \\ \frac{1}{\sqrt{2\pi}} \exp \frac{-(X_t - \mu_2)^2}{2} \end{pmatrix}.$$

To assess the prediction performance of the model FI($d$) and the switching model (1), we perform the following procedure. First, the two models are estimated on samples size $T = 800$, and the forecasts are computed for the horizons $h = 1, \ldots, 10$. We roll the forecast origin forward 10 observations and repeat the procedure 20 times in order to obtain forecasts for the horizons $T = 801, \ldots, 1000$. We compute the Root Mean Squared Errors (RMSE) for each forecast horizon. The results are given in the Tables 2 and 3, and we detail now these results.

1. $(\mu_1,\ \mu_2) = (5, -5)$.

   • When $p = 0.99$ (i.e. $n_{s_t} = 7$), the model (1) provides better forecasts at each horizon than the FI($d$) model.

15

| | $h,p$ | 0.99 | 0.95 | 0.85 | 0.75 | 0.65 | 0.55 | 0.45 | 0.35 | 0.25 | 0.15 | 0.05 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FI($d$) | 1 | 20.039 | 7.667 | 11.097 | **4.928** | **5.084** | **5.009** | **4.821** | **4.782** | **5.611** | **5.571** | 4.191 |
| | 2 | 13.187 | 6.868 | 7.984 | **5.771** | **4.894** | **5.204** | **5.490** | **4.627** | **4.997** | **4.913** | 5.266 |
| | 3 | 11.834 | 6.217 | 6.748 | **4.662** | **4.936** | **5.309** | **5.280** | **4.909** | **5.383** | 5.101 | 4.904 |
| | 4 | 10.111 | 6.044 | 6.204 | **4.944** | **5.042** | **5.300** | **4.829** | **5.443** | **5.153** | **4.890** | 5.336 |
| | 5 | 9.327 | **5.396** | **5.673** | **5.005** | **5.351** | **4.881** | **5.197** | **5.017** | **5.410** | **4.816** | 5.061 |
| | 6 | 8.794 | **4.938** | **5.807** | **4.736** | **4.922** | **4.721** | **5.500** | **5.396** | **5.324** | **5.162** | 5.147 |
| | 7 | 8.894 | **5.126** | **5.471** | **5.715** | **4.864** | **5.196** | **5.189** | **5.338** | **5.072** | **4.827** | 4.938 |
| | 8 | 8.195 | **5.395** | **4.906** | **5.175** | **4.870** | **5.212** | **5.157** | **5.222** | **5.060** | **5.203** | 4.956 |
| | 9 | 7.817 | **4.846** | **5.116** | **5.285** | **5.495** | **5.131** | **5.157** | **4.886** | **5.107** | **5.293** | 4.969 |
| | 10 | 7.855 | **5.163** | **4.928** | **5.802** | **5.511** | 5.39 | **4.985** | **4.879** | **5.471** | **5.054** | 5.303 |
| SW | 1 | 0.717 | 3.887 | 4.456 | **4.710** | **4.864** | **5.071** | **4.906** | **4.293** | **5.580** | **5.193** | 2.701 |
| | 2 | 0.831 | 3.996 | 5.434 | **5.482** | **4.954** | **5.192** | **5.444** | **4.537** | **5.062** | **4.665** | 2.118 |
| | 3 | 1.205 | 4.498 | 5.051 | **4.519** | **4.830** | **5.290** | **5.300** | **4.784** | **4.995** | 4.475 | 3.131 |
| | 4 | 1.135 | 4.468 | 5.232 | **5.126** | **4.955** | **5.286** | **4.755** | **5.287** | **5.374** | **4.787** | 3.670 |
| | 5 | 0.862 | **5.703** | **5.229** | **5.173** | **5.275** | **4.910** | **5.176** | **4.947** | **5.309** | **4.584** | 3.787 |
| | 6 | 0.737 | **5.138** | **5.292** | **5.019** | **4.994** | **4.724** | **5.500** | **5.464** | **5.397** | **5.165** | 4.150 |
| | 7 | 1.374 | **5.226** | **5.034** | **5.419** | **4.894** | **5.256** | **5.271** | **5.287** | **5.038** | **4.750** | 4.058 |
| | 8 | 1.274 | **5.273** | **5.118** | **5.017** | **4.957** | **5.255** | **5.163** | **5.219** | **4.997** | **5.269** | 4.015 |
| | 9 | 0.928 | **4.838** | **5.433** | **5.102** | **5.496** | **5.177** | **5.179** | **4.783** | **5.085** | **5.066** | 4.574 |
| | 10 | 2.377 | **4.710** | **5.127** | **5.521** | **5.308** | **5.343** | **4.981** | **4.959** | **5.421** | **5.194** | 4.677 |

Table 2: RMSE of the forecasts obtained for the model FI($d$) and for the model (1) for the couple $(\mu_1, \mu_2) = (5, -5)$. Bold values represent values for which the difference between the two models is inferior to 0.5.

- This remains true for high transition probability $p$ and small forecasting horizons $h$.

- The same result appears for $p = 0.05$, which corresponds to $n_{s_t} = 950$ and $\rho = -0.9$.

- Nevertheless, when $h$ becomes larger, even for high transition probability, the FI($d$) model becomes competitive with the switching model and can give better forecasts than the model (1).

- When the shifts are more frequent, for $p$ varying from 0.75 to 0.15 for instance, which corresponds to $258 \leq n_{s_t} \leq 834$ and

$-0.7 \leq \rho \leq 0.5$, the predictions' performance of the two models are quite similar.

2. $(\mu_1, \ \mu_2) = (0.5, -0.5)$. The forecasts obtained for the FI($d$) model are almost as good as those obtained for the switching model (1), except in a few cases.

Tables 2 and 3 provide the forecasts for the values of the process $(X_t)_t$, but working with switching models, practitioners are also interested in knowing the ability of the model to predict the "good" state in which the data have to be. Thus, at each step of the procedure, we have computed the probability to predict the true state in which the data have to be. The results are provided in Table 4. For instance, for the lag $h = 1$ and the transition probability $p = 0.99$, the model (1) correctly predicts to be in the state 1 in 95% of the cases. This allows us to measure the quality of forecasting of Markov switching model at each horizon $h$. For $(\mu_1, \ \mu_2) = (5, -5)$, the quality of the prediction diminishes with the transition probability $p$ or with the horizon $h$. Nevertheless, for $(\mu_1, \ \mu_2) = (0.5, -0.5)$, this quality of prediction is weak whatever the horizon $h$ and the transition probability $p \neq 0.99$. This confirm the weak ability of the Markov switching model to predict with a high probability the next state for the data.

In summary, it appears that the switching models provide better forecasts for high transition probabilities and small horizon $h$ if the estimated means levels $\widehat{\mu}_i, i = 1, 2$ are far from each other. Nevertheless, for long horizons, the long memory model provides similar forecasts. In case of close means values, $(\mu_1, \ \mu_2) = (0.5, -0.5)$, the switching model has not a bigger capability to forecast than the FI(d) model.

17

| | h, p | 0.99 | 0.95 | 0.85 | 0.75 | 0.65 | 0.55 | 0.45 | 0.35 | 0.25 | 0.15 | 0.05 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FI(d) | 1 | 1.261 | 1.303 | **1.193** | **0.762** | **1.042** | **1.092** | **1.125** | **1.216** | **0.938** | **1.18** | 1.263 |
| | 2 | 1.13 | **1.301** | 1.141 | **0.924** | 1.196 | **1.16** | **1.033** | **1.222** | **0.892** | **1.097** | **1.056** |
| | 3 | **0.779** | **1.165** | 1.28 | **0.814** | **1.088** | **1.134** | **1.087** | 0.919 | **0.975** | **0.878** | 1.277 |
| | 4 | 1.302 | 1.183 | **0.914** | **1.029** | **1.065** | **1.249** | 1.401 | **1.181** | 1.193 | **1.086** | **1.155** |
| | 5 | **1.409** | **1.116** | **1.022** | **0.707** | **1** | **0.959** | **0.927** | **1.07** | **1.154** | **1.092** | **0.958** |
| | 6 | **1.169** | **1.071** | **1.05** | 1.136 | 1.206 | **1.297** | **1.311** | **1.01** | **1.162** | **1.053** | 0.766 |
| | 7 | **1.142** | **0.952** | 1.119 | **1.238** | **1.075** | **1.057** | 1.218 | 1.18 | **1.005** | 1.682 | **1.292** |
| | 8 | **1.212** | **0.997** | **0.863** | 1.102 | **1.123** | **1.041** | 1.118 | **1.195** | **1.237** | **1.138** | **0.838** |
| | 9 | **0.881** | 1.351 | **1.019** | 1.037 | 1.175 | **0.865** | **1.345** | **1.234** | **0.904** | **0.69** | 1.167 |
| | 10 | **0.789** | **1.211** | **1.076** | **0.96** | **0.971** | **1.344** | **1.217** | **1.006** | 1.52 | 1.16 | **1.125** |
| SW | 1 | 1.247 | 1.321 | **1.184** | **0.762** | **1.061** | **1.094** | **1.125** | **1.249** | **0.923** | **1.19** | 1.037 |
| | 2 | 1 | **1.327** | 1.120 | **0.932** | 1.182 | **1.161** | **1.039** | **1.216** | **0.94** | **1.163** | **1.388** |
| | 3 | **0.825** | **1.129** | 1.259 | **0.817** | **1.093** | **1.141** | **1.088** | 0.909 | **1.005** | **0.846** | 0.721 |
| | 4 | 1.249 | 1.169 | **0.922** | **1.026** | **1.07** | **1.24** | 1.396 | **1.177** | 1.185 | **1.107** | **1.321** |
| | 5 | **1.506** | **1.111** | **1.018** | **0.713** | **1.011** | **0.955** | **0.928** | **1.078** | **1.153** | **1.097** | **0.951** |
| | 6 | **1.176** | **1.083** | **1.067** | 1.137 | 1.227 | **1.299** | **1.312** | **1.011** | **1.167** | **1.069** | 1.375 |
| | 7 | **1.137** | **0.955** | 1.120 | **1.231** | **1.088** | **1.055** | 1.22 | 1.193 | **0.987** | 1.692 | **1.299** |
| | 8 | **1.163** | **1** | **0.865** | 1.105 | **1.094** | **1.047** | 1.116 | **1.176** | **1.243** | **1.128** | **0.879** |
| | 9 | **0.922** | 1.359 | **1.017** | 1.044 | 1.22 | **0.865** | **1.343** | **1.204** | **0.9** | **0.684** | 1.053 |
| | 10 | **0.883** | **1.21** | **1.084** | **0.955** | **0.978** | **1.344** | **1.218** | **1** | 1.522 | 1.169 | **1.24** |

Table 3: RMSE of the forecasts obtained for the model FI($d$) and for the model (1) for the couple $(\mu_1, \mu_2) = (0.5, -0.5)$. Bold values represent values for which the difference between the two models is inferior to 0.25.

## 6   Conclusion

In the literature, many studies have shown that stationary processes with structural breaks can produce slowly decaying autocorrelations and can be assimilated to long memory processes. This phenomenon is observed in many economic and financial time series (Baillie, 1996, Ooms and Doornik, 1999 and Starica, 2004). Considering a Markov switching model, we show theoretically that the means values as well as the sum of the transition probabilities have a huge influence on the autocorrelation function's behavior. In this

| $h, p$ | 0.99 | 0.95 | 0.85 | 0.75 | 0.65 | 0.55 | 0.45 | 0.35 | 0.25 | 0.15 | 0.05 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| (5, -5) | | | | | | | | | | | |
| 1 | 0.95 | 0.8 | 0.7 | 0.55 | 0.6 | 0.6 | 0.65 | 0.25 | 0.55 | 0.5 | 0.15 |
| 2 | 0.95 | 0.75 | 0.55 | 0.5 | 0.75 | 0.7 | 0.3 | 0.3 | 0.6 | 0.55 | 0.15 |
| 3 | 0.95 | 0.7 | 0.55 | 0.45 | 0.45 | 0.25 | 0.35 | 0.4 | 0.55 | 0.45 | 0.2 |
| 4 | 0.95 | 0.7 | 0.55 | 0.45 | 0.4 | 0.35 | 0.5 | 0.4 | 0.3 | 0.45 | 0.25 |
| 5 | 0.95 | 0.55 | 0.5 | 0.55 | 0.4 | 0.45 | 0.45 | 0.55 | 0.25 | 0.45 | 0.25 |
| 6 | 0.95 | 0.65 | 0.6 | 0.4 | 0.45 | 0.55 | 0.5 | 0.4 | 0.45 | 0.5 | 0.2 |
| 7 | 0.95 | 0.65 | 0.65 | 0.45 | 0.6 | 0.45 | 0.3 | 0.5 | 0.4 | 0.55 | 0.2 |
| 8 | 0.95 | 0.65 | 0.3 | 0.6 | 0.5 | 0.5 | 0.6 | 0.65 | 0.35 | 0.5 | 0.2 |
| 9 | 0.95 | 0.7 | 0.4 | 0.55 | 0.25 | 0.5 | 0.4 | 0.65 | 0.3 | 0.45 | 0.25 |
| 10 | 0.9 | 0.75 | 0.45 | 0.45 | 0.5 | 0.4 | 0.65 | 0.4 | 0.65 | 0.55 | 0.35 |
| (0.5, -0.5) | | | | | | | | | | | |
| 1 | 0.85 | 0.7 | 0.65 | 0.55 | 0.65 | 0.65 | 0.55 | 0.4 | 0.55 | 0.55 | 0.25 |
| 2 | 0.85 | 0.65 | 0.5 | 0.55 | 0.6 | 0.45 | 0.4 | 0.55 | 0.55 | 0.5 | 0.25 |
| 3 | 0.85 | 0.6 | 0.35 | 0.45 | 0.45 | 0.45 | 0.55 | 0.45 | 0.65 | 0.6 | 0.3 |
| 4 | 0.85 | 0.65 | 0.5 | 0.45 | 0.4 | 0.45 | 0.5 | 0.55 | 0.4 | 0.35 | 0.35 |
| 5 | 0.85 | 0.5 | 0.45 | 0.6 | 0.35 | 0.55 | 0.55 | 0.5 | 0.75 | 0.6 | 0.35 |
| 6 | 0.85 | 0.5 | 0.45 | 0.45 | 0.5 | 0.55 | 0.6 | 0.6 | 0.35 | 0.4 | 0.3 |
| 7 | 0.85 | 0.5 | 0.55 | 0.5 | 0.45 | 0.65 | 0.3 | 0.3 | 0.6 | 0.55 | 0.3 |
| 8 | 0.85 | 0.5 | 0.6 | 0.7 | 0.55 | 0.6 | 0.6 | 0.55 | 0.4 | 0.5 | 0.3 |
| 9 | 0.85 | 0.55 | 0.6 | 0.55 | 0.6 | 0.7 | 0.4 | 0.55 | 0.6 | 0.75 | 0.35 |
| 10 | 0.8 | 0.6 | 0.55 | 0.45 | 0.45 | 0.3 | 0.65 | 0.5 | 0.45 | 0.35 | 0.35 |

Table 4: Probability computed to be in the actual "good" state, for an horizon $h$ and a probability $p$, using the model (1).

paper, from simulations and according to the values of the different parameters of the switching model under study, we exhibit several behaviors for this Markov switching model's autocorrelation function: long memory behavior, short memory behaviors and anti-persistence. With a first graphical approach, we are unable to decide precisely for which transition probability $p$, we are in presence of long or short memory behavior. Thus, we estimate an FI(d) model on several series simulated from the Markov switching model (1) in order to get a measure of the long memory parameter $\widehat{d}$ and see if it permits to decide when we are in presence of long or short memory behavior. An important result is that $\widehat{d} = 0$ arises only in few cases and thus

the presence of short memory behavior for the model (1) is rarely accepted. Then, we have compared the predictions obtained from the Markov switching model and the long memory process. We show that in certain cases the long memory process provides better forecasts, in particular for long term horizon: in that case we cannot speak of spurious long memory process.

In this paper, we consider the Markov switching model defined by the equation (1). It will be interesting to transpose this study on Markov switching model with autoregressive parameters. Moreover, we have also restricted our study to the case $p_{11} = p_{22}$, but it will be interesting to consider many pairs of transition probabilities, because it is possible that two processes, whose transition probabilities are different but having the same sum, can exhibit similar autocorrelation functions' behaviors. For instance, choose for $(p_{11}, \ p_{22}) = (0.9, 0.6)$ and $(p_{11}, \ p_{22}) = (0.75, 0.75)$, then $\rho = 0.5$ in both cases. For these two couples of transition probabilities, the simulated series switch from one state to the other respectively 143 and 271 times. Thus, the series behave very differently. For each couple of means $(\mu_1, \ \mu_2) = (0.5, -0.5)$ and $(\mu_1, \ \mu_2) = (5, -5)$, we observe, on Figure 3, that although the values of the transition probabilities are different with the same sum, the autocorrelation functions of the simulated processes issued from model (1) behave similarly. Thus, in terms of modelization, the model cannot be identified even if this situation agrees with the theoretical expression of the autocorrelation function (3).

Markov switching models are also a useful tool to detect periods of different volatilities inside the data. By considering the previous approach, we study how periods of different volatilities have an impact on the short or long
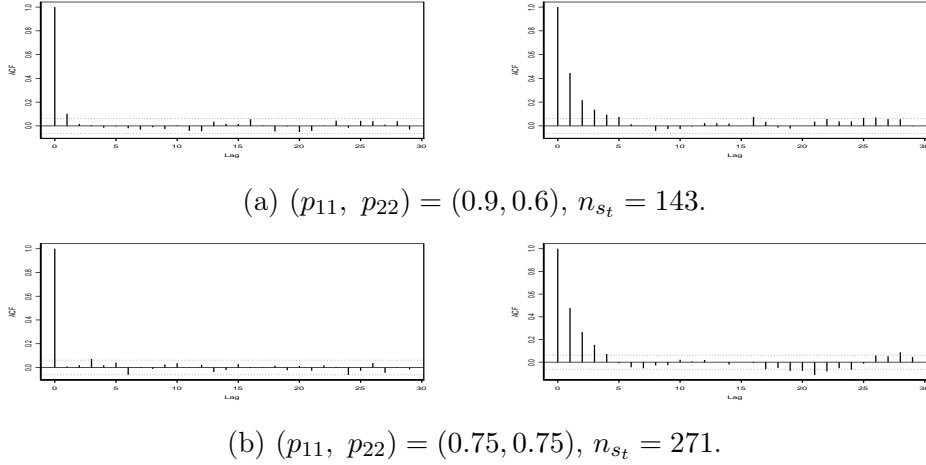
20

(a) $(p_{11}, \ p_{22}) = (0.9, 0.6)$, $n_{s_t} = 143$.



(b) $(p_{11}, \ p_{22}) = (0.75, 0.75)$, $n_{s_t} = 271$.

Figure 3: Autocorrelation functions issued from the model (1) with $p_{11} + p_{22} = 1.5$. Left column: $(\mu_1, \ \mu_2) = (0.5, -0.5)$ and right column: $(\mu_1, \ \mu_2) = (5, -5)$.

memory behavior of Markov switching models. Consider the simplest model which has a switch on the volatility parameter defined by:

$$X_t = \mu + \sigma_{s_t} \varepsilon_t. \tag{9}$$

Its autocorrelation function is similar to the white noise's one:

$$\Gamma(h) = 0, \quad \forall h > 0.$$

Thus, this model will always have a short memory behavior and cannot exhibit a long memory behavior. This is really the big difference between the model (1) and the model (9). The simulations are in accord with this result. To illustrate this property, we provide on Figure 4, the trajectories and the autocorrelation functions of two simulated series issued from model (9) with two different pairs of volatilities: $(\sigma_1, \ \sigma_2) = (1, \ 4)$ and $(\sigma_1, \ \sigma_2) = (1, \ 20)$ for $p = 0.99$ and $T = 1000$. The difference between the trajectories of Figures 4 (a) and (b) stands in the range of the values. Notice that the switches inside

21

the data are rare, indeed $n_{s_t} = 10$ here. Whereas the processes have periods of high and low volatilities, their autocorrelation functions are similar to the white noise's one, and thus exhibit a short memory behavior. This means that, in presence of high and low volatility inside real data, spurious long memory behavior would be rare.
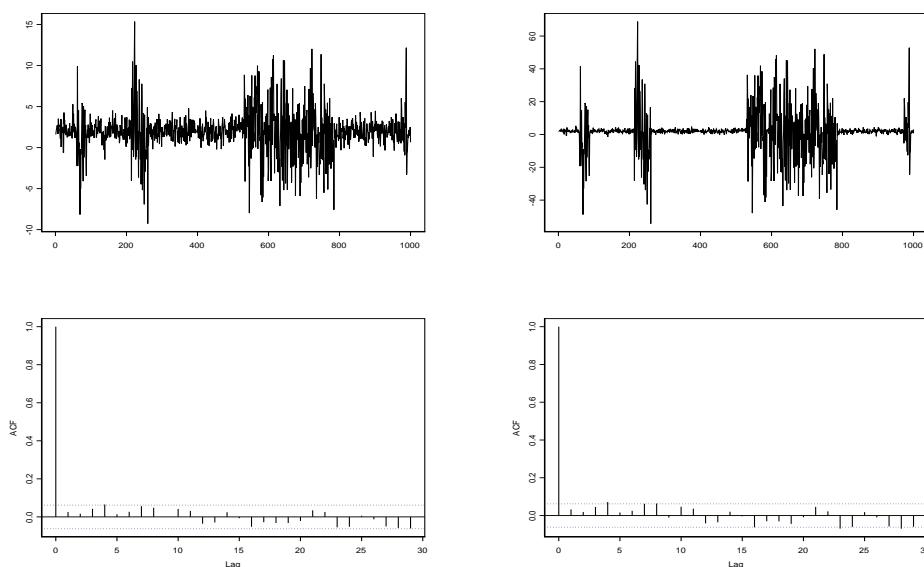


(a) $(\sigma_1, \ \sigma_2) = (1, \ 4)$        (b) $(\sigma_1, \ \sigma_2) = (1, \ 20)$

Figure 4: Trajectories and autocorrelation functions of two simulated series issued from model (9), with $\mu = 2$ and $p = 0.99$, for $T = 1000$. Here, $n_{s_t} = 10$.

**Aknowledgments**

# References

[1] Allan D.W., 1966, Statistics of Atomic Frequency Standards, *Proceedings of the IEEE, 54, 221Ũ230.*

[2] Baillie R.T., 1996, Long Memory Processes and Fractional Integration in Econometrics, *Journal of Econometrics, 73, 5-59.*

[3] Breidt F.J., N.J. Hsu, 2002, A Class of Nearly Long Memory Time Series Models, *International Journal of Forecasting, 18, 265-281.*

[4] Chan K.S., 1990, Testing for Threshold Autoregression, *The Annals of Statistics, 18, 1886-1894.*

[5] Diebold F.X., Inoue A., 2001, Long Memory and Regime Switching, *Journal of Econometrics 105, 131-159.*

[6] Dufrénot G., Guégan D., Péguin-Feissole A., 2005, Modelling Squared Returns using a SETAR Model with Long Memory Dynamics , *Economics Letters, 86, 237-243.*

[7] Gourieroux C., Jasiak J., 2001, Memory and Infrequent Breaks, *Economic Letters, 70, 29-47 .*

[8] Granger C.W.J., Joyeux R., 1980, An Introduction to Long Memory Time Series and Fractional Fifferencing, *Journal of Time Series Analysis, 1, 15-30.*

[9] Granger C.W.J., Hyung N., 2004, Occasional Structural Breaks and Long Memory with an Application to the S&P 500 Absolute Stock Returns, *Journal of Empirical Finance, 11, 399-421.*

[10] Granger C.W.J., Terasvirta T., 1999, A Simple Nonlinear Time Series Model with Misleading Linear Properties, *Economic Letters, 62, 161-165.*

[11] Guégan D., 2004, How Can we Define the Concept of Long Memory? An Econometric Survey? Discussion Paper no. 178, School of Economics and Finance, Brisbane.

[12] Guégan D., Ferrara L., 2001, Forecasting with k-Factor Gegenbauer Processes: Theory and Applications, *Journal of Forecasting, 20, 581-601.*

[13] Guégan D., Rioublanc S., 2003, Study of regime switching models? Do they provide long memory behavior? An empirical approach, Note de Recherche IDHE-MORA, 2003-13, Ecole Normale Supérieure de Cachan, France.

[14] Guégan, D., S. Rioublanc, 2004, Long Memory Behavior of Markov Switching Models, Document de travail, Ecole Normale Supérieure de Cachan, France.

[15] Hamilton J.D., 1988, Rational-Expectations Econometric Analysis of Changes in Regime: An Investigation of the Term Structure of Interest Rates, *Journal of Economic Dynamics and Control 12, 385-423.*

[16] Hosking J.R.M., 1981, Fractional Differencing, *Biometrika, 68, 165-176.*

[17] Krolzig H.-M., 1997, *Markov-Switching Vector Autoregressions. Modelling, Statistical Inference and Application to Business Cycle Analysis,* Lecture Notes on Economics and Mathematical Systems, Volume 454, Berlin: Springer.

[18] Mikosch T., Starica C., 2004, Non-stationarities in Financial Time Series, the Long Range Dependence and the IGARCH Effects, *Review of Economics and Statistics, 86, 378-390.*

[19] Ooms M., Doornik J., 1999, Inference and Forecasting for Fractional Autoregressive Integrated Moving Average Models with an Application to US and UK Inflation, *Econometrics Institute Report 9947-A, Erasmus University Rotterdam.*

[20] Yajima Y., 1985, On Estimation of Long-Memory Times Series Models, *Australian Journal of Statistics, 27, 303-320.*

# 7 Appendix

## 7.1 Tables

Here, we provide a summary on the autocorrelation function's behavior for the model (1), detected graphically with respect to $\rho$, for the large sample size: $T = 5000$ and $T = 25000$.

| $\rho$ | -1 | -0.76 | -0.6 | 0.48 | 0.68 | 1 |
|---|---|---|---|---|---|---|
| (0.5, -0.5) | AP | | SM | | | LM |
| (5, -5) | AP | | SM | | LM | |

Table 5: Autocorrelation functions' behavior for the model (1) according to $\rho$ and $(\mu_1, \mu_2)$. LM stands for long memory, SM for short memory and AP for anti-persistence, $T = 5000$.

| $\rho$ | -1 | -0.76 | -0.54 | 0.48 | 0.6 | 1 |
|---|---|---|---|---|---|---|
| (0.5, -0.5) | AP | | SM | | | LM |
| (5, -5) | AP | | SM | | LM | |

Table 6: Autocorrelation functions' behavior for the model (1) according to $\rho$ and $(\mu_1, \mu_2)$. LM stands for long memory, SM for short memory and AP for anti-persistence, $T = 25000$.
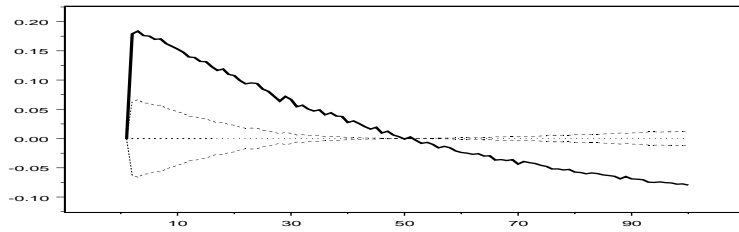
## 7.2 Figures

We provide here the behavior of the estimated long memory parameter $d$ when the sample size is large: $T = 5000$ and $T = 25000$.
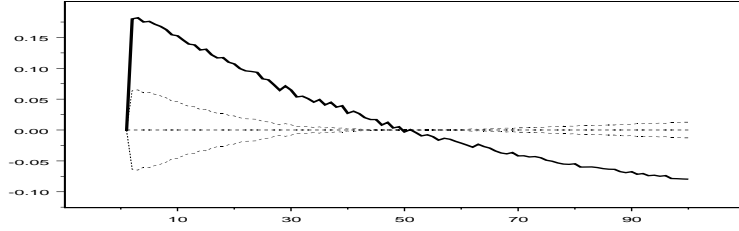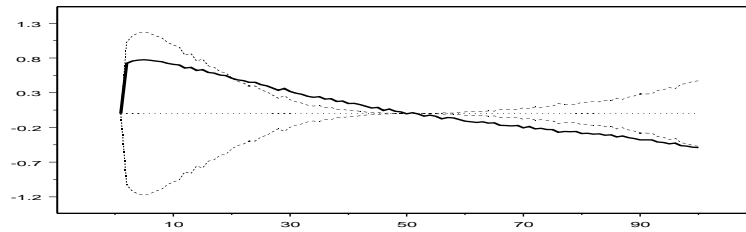
(a) $T = 5000$, $(\mu_1,\ \mu_2) = (0.5, -0.5)$.
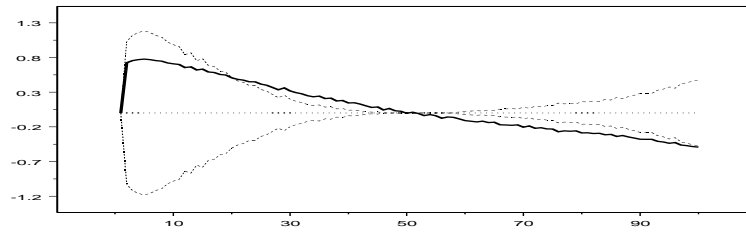


(b) $T = 5000$, $(\mu_1,\ \mu_2) = (5, -5)$.



(c) $T = 25000$, $(\mu_1,\ \mu_2) = (0.5, -0.5)$.



(d) $T = 25000$, $(\mu_1,\ \mu_2) = (5, -5)$.

Figure 5: Estimated memory parameters $\widehat{d}$ (solid line) and their confidence interval (dotted line) for the switching model (1). The abscissa represent the transition probabilities $p$ from 1 to 0.01 with a step of 0.01, $T = 5000$ and $T = 25000$.