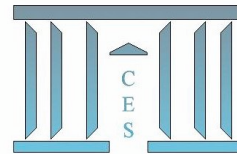




# Documents de Travail du Centre d'Économie de la Sorbonne

C  
E  
S  
  
W  
o  
r  
k  
i  
n  
g  
  
P  
a  
p  
e  
r  
s



## Global and Local stationary modelling in finance : Theory and empirical evidence

Dominique GUEGAN

**2007.53**



CENTRE NATIONAL  
DE LA RECHERCHE  
SCIENTIFIQUE

# Global and Local stationary modelling in finance: Theory and empirical evidence

Guégan D. \*

April 8, 2007

## Abstract

In this paper we deal with the problem of non - stationarity encountered in a lot of data sets coming from existence of multiple seasonalities, jumps, volatility, distorsion, aggregation, etc. We study the problem caused by these non stationarities on the estimation of the sample autocorrelation function and give several examples of models for which spurious behaviors is created by this fact. It concerns Markov switching processes, Stopbreak models and SETAR processes. Then, new strategies are suggested to study locally these data sets. We propose first a test based on the  $k$ -th cumulants and mainly the construction of a meta-distribution based on copulas for the data set which will permit to take into account all the non-stationarities. This approach suggests that we can be able to do risk management for portfolio containing non stationary assets and also to obtain the distribution function of some specific models.

*Keywords:* Non-Stationarity - Distribution function - Copula - Long-memory - Switching - SETAR - StopBreak Models - Cumulants - Estimation. theory.

*JEL classification:* C32, C51, G12

---

\*E.N.S. Cachan, Département d'Economie et Gestion, Centre d'Economie de la Sorbonne (CES), Antenne Cachan, UMR 7184, 61 Avenue du President Wilson, 94235, Cachan Cedex, France, e-mail : guegan@ecogest.ens-cachan.fr

# 1 Introduction

Adequate framework to study the behavior of stock returns, and more generally asset prices, is important for several reasons. First, the nature of asset prices behavior is necessary to get robust forecasts. Second the knowledge of the probabilistic properties of these asset prices is fundamental to the formulation of the concept of risks: indeed the measurement of risks depends heavily on properties of the empirical distribution such as stationarity, long tailedness, finiteness of the second and higher order moments. Third, various tests for the empirical validity of financial models and the application of these models rely on the robustness of statistical tools which can be deficient in specific context. Fourth, several important pricing models for stock options and other similar financial instruments usually require explicit estimates of stock return variances. The usefulness of these models depends largely on the adequacy and the stationarity of almost the second order moments. Indeed, to model real data sets using classical stochastic processes imposes that the data sets verify almost the second order stationarity condition. This stationarity condition concerns the unconditional moments of the process. It is in that context that most of the models developed from the sixties' have been studied; We refer to the ARMA processes (Brockwell and Davis, 1988), the ARCH, GARCH and EGARCH models (Engle, 1982, Bollerslev, 1986, Nelson, 1990), the SETAR process (Lim and Tong, 1980 and Tong, 1990), the bilinear model (Granger and Andersen, 1978, Guégan, 1994), the EXPAR model (Haggan and Ozaki, 1980), the long memory process (Granger and Joyeux, 1980, Hosking, 1981, Gray, Zang and Woodward, 1989, Beran, 1994, Giraitis and Leipus, 1995, Guégan, 2000), the Markov switching process (Hamilton, 1988). For all these models, we get an invertible causal solution under specific conditions on the parameters, then the forecast points and the forecast intervals are available.

Thus, the stationarity assumption is the basis for a general asymptotic theory for identification, estimation and forecasting. It guarantees that the increase of the sample size leads to more and more information of the same kind which is basic for an asymptotic theory to make sense.

Moreover, non-stationarity modelling has also a long tradition in econometrics. This one is based on the conditional moments of the data generating process. It appears mainly in the heteroscedastic and volatility models, like the GARCH and related models, and stochastic volatility processes (Ghysels, Harvey and Renault (1997)). This non stationarity appears also in a different way with structural changes models like the Markov switching model (Hamil-

ton, 1988), the Stopbreak model (Diebold and Inoue, 2001, Breidt and Hsu, 2002, Granger and Hyung, 2004) and the SETAR model (Tong, 1990), for instance. It can also be observed for linear models with time varying coefficients (Nicholls and Quinn, 1982, Tsay, 1987).

Thus, using stationary unconditional moments suggest a global stationarity for the model, but using non-stationary unconditional or conditional moments suggest that this global stationarity fails and that we only observe a local stationary behavior.

The growing evidence of instability in the stochastic behavior of stocks, of exchange rates, of some economic data sets like growth rates for instance, characterized by volatility, explosions or jumps inside the variance or on the levels of the prices imposes to discuss the assumption of global stationarity and its consequence in modelling, particularly in forecasting and risk management strategy. There exist a voluminous literature concerning the modelling of stock returns, and most of the published research work relies on the stationary assumption that is embedded in almost conventional statistical methods: it is assumed that the distributional shapes are invariant and the variance finite. Thus we can address several questions with respect to these remarks.

1. What kinds of non-stationarity affect the major financial and economic data sets? How can we detect them?
2. What is the impact of evidence of non stationarity on the statistics computed from the global non stationary data sets?
3. How can we analyze data sets in the non stationary global framework? Does the asymptotic theory work in non-stationary framework?
4. Can we use a local modelling to avoid a global non-stationarity? How to detect it?

Based on the fact that we cannot expect market return variability to be a constant value over time, the conventional concept of risk cannot be defined on a stationary basis. Thus the purpose of studying the non constancy of market return variability will permit to gather better insights into the complex mechanism underlying the evolution of these returns in order to better forecast and to define a robust risk management strategy.

These questions begin to be discussed in some papers in the economic literature. For some of these questions, the answers are known, for others,

very few works exist, we refer for instance to Lobato and Savin, 1998, Engle and Smith, 1999, Granger and Terasvirta, 1999, Diebold and Inoue, 2001, Granger and Hyung, 2004, Mikosch and Starica, 2004. In this paper we discuss all these problems and we propose new strategies to detect non stationarity based on the cumulants spectral density function and on copulas, and modellings to solve them in order to propose a new approach for risk measures. Several interesting topics in empirical finance awaiting future research are also discussed.

In the next Section, we recall the notion of global stationarity and short and long memory behaviors. In Section three, we introduce the different types of non stationarities observed in real data sets and we examine what are the specific behaviors observed in presence of non stationarity on real data sets. In Section four, we study the asymptotic behavior of the sample ACF in presence of non stationarity. In Section 5, we introduce the notion of local stationarity. Section 6 is devoted to new results concerning the detection of local stationarity and the building of a kind of meta-distribution. This last approach permits to compute the empirical distribution function for a whole non-stationary sample using a local approach and also to give new possibilities to build the distribution function for specific models characterizing by certain non stationarities. Finally Section 7 concludes. In the different annexes, we found proof of the proposition 4.1 and some simulations that illustrate problems discussed all along the paper.

## 2 Some notations and recalls

In order to find a model which characterizes a real data set, we use a theoretical framework assuming that the observations of any data set are dependent random variables obtained at different moments of time under stationarity and corresponding to the realization of a process  $(Y_t)_t$ . We first recall the notion of stationary stochastic process.

A stochastic process is a sequence of random variables  $(Y_t)_t$  defined on a probability space  $(\Omega, \mathcal{A}, P)$ . Then  $\forall t$  fixed,  $Y_t$  is a function  $Y_t(\cdot)$  on the space  $\Omega$ , and  $\forall \omega \in \Omega$  fix,  $Y_t(\omega)$  is a function on  $\mathbb{Z}$ . The functions  $(Y_t(\omega))_{\omega \in \Omega}$  defined on  $\mathbb{Z}$  are realizations of the process  $(Y_t)_t$ . A second order stochastic process  $(Y_t)_t$  is such that,  $\forall t$ ,  $EY_t^2 < \infty$ . For a second order stochastic process, the mean  $\mu_t = EY_t$  exists,  $\forall t$ , as the variance and the covariance. The covariance  $\gamma(\cdot, \cdot)$  of a second order stochastic process  $(Y_t)_t$  exists and is defined by

$$\forall h, \forall t \in \mathbb{Z}, cov(Y_t, Y_{t+h}) = \gamma(h, t) < \infty. \quad (1)$$

A stochastic process is completely known as soon as we know its probability distribution function. When several realizations of a process are available, the theory of stochastic processes can be applied to study this distribution, however, in most empirical problems, only a single realization is available. Each observation in a time series is a realization of each random variable of the process.

$$\begin{array}{c} Y_1, Y_2, \dots, Y_T \\ \downarrow, \downarrow, \dots, \downarrow \\ y_1, y_2, \dots, y_T. \end{array}$$

Consequently, we have one realization of each random variable and inference is not possible. We have to restrict the properties of the process to carry out inference. To allow estimation, we need to restrict the process to be stationary. There exist several concepts of stationarity. We recall the notions of strict and second order stationarities.

**Definition 2.1** *A stochastic process  $(Y_t)_t$  is strictly stationary if the joint distribution of  $Y_{t_1}, Y_{t_2}, \dots, Y_{t_k}$  is identical to that of  $Y_{t_1+h}, Y_{t_2+h}, \dots, Y_{t_k+h}$ , for all  $h$ , where  $k$  is an arbitrary positive integer and  $t_1, t_2, \dots, t_k$  is a collection of  $k$  positive integers.*

Strict stationarity means intuitively that the graphs over two equal-length time intervals of a realization of a time series should exhibit similar statistical characteristics. It means also that  $(Y_{t_1}, \dots, Y_{t_k})$  and  $(Y_{t_1+h}, \dots, Y_{t_k+h})$  have the same joint distribution for all positive integers  $k$  and  $h$ . Therefore, strict stationarity requires that the distribution of  $Y_{t_1}, Y_{t_2}, \dots, Y_{t_k}$  is invariant under time shifts. In that case, we speak of global stationarity.

**Definition 2.2** *A stochastic process  $(Y_t)_t$  is weakly stationary if  $\forall t, E(Y_t) = \mu$ ,  $Var(Y_t) = \sigma^2$  and  $cov(Y_t, Y_{t+h}) = \gamma(h)$ .*

In this definition, the first property implies that the mean of each random variable in the process is the same regardless of the particular random variable chosen. The two other conditions have similar interpretations. Weak stationarity requires that only the two first moments of the process are time invariant. It is obvious to show that weak stationarity does not imply strict stationarity, as the latter requires that the whole distribution function is invariant. However, it is important to note that strict stationarity implies weak stationarity if the variance of the process is finite. There are examples of strictly stationary processes without finite variances that are not weakly stationary.

Under the two previous conditions (Definitions 2.1 and 2.2), and under the strong condition  $\sum_{j \in \mathbb{Z}} a_j^2 < \infty$ , generally we get an invertible causal solution for the process  $(Y_t)_t$ :

$$Y_t = \sum_{j \in \mathbb{Z}} a_j \varepsilon_{t-j},$$

where  $(\varepsilon_t)_t$  is a strong white noise. Now, under the restricted condition  $\sum_{j \in \mathbb{Z}} |a_j| < \infty$ , the process  $(\varepsilon_t)_t$  would be the innovations process associated to  $(Y_t)_t$ .

These stationary conditions are needed to guarantee the asymptotic properties of the sample mean, variance and covariances and are different for each of these estimators. In a nonlinear setting, we use ergodicity conditions which require that observations sufficiently far apart should be almost uncorrelated. Under all these conditions, the process is always globally stationary: this means that this property remains true on the whole sample. In all cases this means that inference can be done for such processes and that the asymptotic theory works. In particular forecasts are available and confidence intervals can be provided.

Now, if a process  $(Y_t)_t$  is strongly mixing with geometric rate, then the theoretical auto-correlation function (ACF) of the process  $(Y_t)_t$  decays to zero at exponential rate. This theoretical behavior characterizes a short memory behavior for the process  $(Y_t)_t$ . Thus, this short memory property should be reflected in the behavior of the sample ACF computed using a data set issued from this process. If the series has not this last behavior, then generally we speak about long memory behavior.

**Definition 2.3** *Let  $(Y_t)_t$  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  $\rho_Y(h)$  satisfies*

$$\lim_{h \rightarrow \infty} \rho_Y(h) = Ch^{2d-1}, \quad (2)$$

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

For a review on the concepts of long memory behavior, we refer to Guégan (2005).

### 3 Non-stationary behaviors

Even if we work always in a global stationary framework, a lot of non stationarities are observed on real data sets. In this Section, we specify some of the

non-stationarities which affect the major financial and economic data sets. These questions are the base of a lot of problems concerning the modelling of real data sets. Indeed, structural behaviors like volatility, jumps, explosions and seasonality provoke non-stationarity. Now specific transformations on the data sets like concatenation, aggregation or distortion are also at the origin of non-stationarity.

All these features imply that the property of global stationarity fails. Indeed, existence of volatility imposes that the variance depends on time. In presence of seasonality the covariance depends on time. The existence of states induces changes in mean or in variance all along the trajectory. Concatenated or distorted models cannot have the same probability distribution function on the whole period. Aggregation is a source of specific features. For some of the previous situations the higher order moments do not exist implying that the covariance function is not defined. Thus, we need - in terms of modelling - to specify the framework in which we are going to work. First of all, we specify how these non stationarities are actually taking into account.

The existence of volatility inside a real data set is generally modelled using time varying conditional variance, via the classical heteroscedastic models, for a review, Bollerslev, Chou and Kroner, (1992). But it can also be modelled using time-varying second unconditional moments, Nicholls and Quinn, (1982), Léorat, (2000) and Starica and Granger, (2005). Under stationary conditions, all these models have short memory behavior, in the covariance sense. The detection of presence of volatility is obtained through different tests like the Wald test. This test attests changes in volatility along the trajectory, thus the conditional variance evolves with time. This non global stationarity can affect the forecasts.

A strong cyclical component inside financial data (monthly for instance or hourly for high frequency data) produces evidence of non stationarity. Practically, if we observe seasonality inside the data, some transformations can suppress this feature. If after some transformations (like stochastic differentiating filter), the seasonal effect remains, then we will consider it as a structural property and its analysis will be included in a more general model. Specific models take into account this feature. The class of GARMA models and related models describes these features, see Gray, Zhang and Woodward (1989), Giraitis and Leipus (1995), Guégan (2001) and Ferrara and Guégan (2006). The estimation of the fractional parameter associated to a Gegenbauer filter can be used as a test to detect presence of fractional seasonality, see for instance Arteche (2003).

Existence of jumps produces several regimes inside data sets. These different regimes can characterize the level of the data or its volatility. Changes in mean or in variance affect the properties of the distribution function characterizing the underlying process. Indeed, this distribution function cannot be invariant under time-shifts and thus a global stationarity cannot be assumed. These behaviors are often modelled using Markov switching models, SETAR or STAR processes, Stop-Break processes, sign processes, etc. If for instance we use a Markov switching model with two states on the mean:  $\mu_1$  and  $\mu_2$ , this model is locally stationary, as soon as we stay in one regime but it is not globally stationary. The distribution function of the underlying process has not always the same mean. This means that we have to test  $\mu_1 - \mu_2 = 0$ , under the constraint that the model has changes in regimes. This is not easy because we have constrained parameters like the transition probabilities  $p_{11}$  and  $p_{22}$ , for a switching process or the condition  $Y_{t-d} < c$  for a SETAR process, if  $(Y_t)_t$  is the underlying process. Some tests exist in that context, see Hansen, (1992), Charfeddine and Guégan, (2006) and references therein. The question of forecasting is still opened. How to predict correctly when we move from one regime to another one? What is the effect of this shift on the behavior of the whole sample, knowing that the Markov switching and the SETAR processes under global stationary assumption have a short memory behavior?

We say that we are in presence of a distortion effect inside a real data set when we observe explosions that one cannot remove from any transformation. This behavior can also be viewed as a structural effect. Explosions imply existence of non stationarity inside the data set. This effect can be characterized by the fact that some higher order moments of the distribution function do not exist. Models with coefficients close to the non stationary domain provoke also this kind of effect. For instance, this behavior can be produced by AR(1) model with Cauchy distribution, ARCH(1) process with Pareto distribution or some IGARCH processes. The empirical distribution function has to be studied in detail.

When we juxtapose (or concatenate) different stationary linear or non linear processes, we get specific second order behavior for the whole model. The probabilistic properties of the whole process cannot be the same than those of each process. In practice, this situation seems a little bit artificial because we do not know a priori when we are in presence of concatenated processes. Nevertheless, it has been shown on different examples that some data sets can correspond to this situation (Mikosch and Starica, 2004). Thus, it is

important to have a deeper understanding - in practice - of this phenomena.

Aggregation of independent or weakly dependent random variables can create specific dependence. Some aggregated models use time - varying coefficients with specific probability distribution functions, (Robinson, 1978, Granger, 1980), and they appear to be non-stationary. In 1980, Granger was probably one of the first, in economy, who points out the specific properties obtained by aggregating dynamic equations. With his construction, he creates long memory behavior aggregating short memory processes. Moreover he shows that these models are useful in improving long run forecasts in economics and also in finding stronger distributed lag relationships between economic variables.

Thus, a lot of non-stationarities affect real data sets. It is necessary to understand how these non-stationarities do affect the well known statistics used to model and understand the evolution of these data sets. It seems that it is mainly the behavior of the sample ACF which is affected and also the properties of the empirical distribution function. In the next Section, we discuss the properties of the sample ACF.

## 4 Behavior of sample ACF under a global non stationary framework

We focus on the behavior of the sample ACF which is one of the main statistical tool used to identify second order stochastic processes. The class of short memory, second order stationary processes includes ARMA, ARCH, GARCH, Markov Switching, SETAR, Stop Break processes. For all these models, the theoretical ACF under second order stationarity has an exponential decay towards zero, (for a review of most of these models, see Guégan, 2003a, chapter 8). What do happen when we compute the sample ACF in presence of some kind of non-stationarity? In order to illustrate this problem, we are going to compute the sample ACF of a process on different subsamples and then, we consider its behavior on the whole sample. The general case is the folowing.

We assume that we observe a sample size  $T$ ,  $Y = (Y_1, Y_2, \dots, Y_T)$  and we divide it in  $r$  subsamples consisting each of distinct stationary ergodic processes with finite second order moments. We denote  $p_j \in R^+$ ,  $j = 1, \dots, r$  such that  $p_1 + p_2 + \dots + p_r = 1$ . Here  $p_j$  is the proportion of ob-

servations from the  $j$ th subsample in the full sample. If we define now  $q_j = p_1 + p_2 + \dots + p_j$ ,  $j = 1, \dots, r$ , thus the whole sample is written as:  $Y = (Y_1^{(1)}, \dots, Y_{Tq_1}^{(1)}, Y_{Tq_1+1}^{(2)}, \dots, Y_{Tq_{r-1}+1}^{(r)}, \dots, Y_T^{(r)})$ . Now, the sample autocovariance function for the series  $(Y_t)_t$  is equal to

$$\tilde{\gamma}_Y(h) = \frac{1}{T} \sum_{t=1}^{T-h} (Y_t - \underline{Y}_T)(Y_{t+h} - \underline{Y}_T), \quad (3)$$

where  $\underline{Y}_T$  is the sample mean. To get  $\tilde{\gamma}_Y(h)$ , we can compute the sample autocovariance on each subsample and then sum it. We get the following result, whose a detailed proof is given in the Annex:

**Proposition 4.1** *Let be  $r$  subsamples  $Y_1^{(i)}, \dots, Y_{Tq_i}^{(i)}$ ,  $i = 1, \dots, r$ , coming from the sample  $Y$ , each subsample corresponding to a distinct stationary ergodic process with finite second order moments, whose sample covariance is equal to  $\tilde{\gamma}_{Y^{(i)}}(h)$ , then*

$$\tilde{\gamma}_Y(h) \rightarrow \sum_{i=1}^r p_i \tilde{\gamma}_{Y^{(i)}}(h) + \sum_{1 \leq i \leq j \leq r} p_i p_j (EY^{(j)} - EY^{(i)})^2, \quad h \rightarrow \infty. \quad (4)$$

Now, if the expectations of the subsequences  $(Y^{(i)})$  differ, and because the autocovariances  $\tilde{\gamma}_{Y^{(i)}}(h)$  decay to zero exponentially as  $h \rightarrow \infty$  (due to the ergodic property of the subsequences), the sample ACF  $\tilde{\gamma}_Y(h)$  for sufficiently large  $h$  is close to a strictly positive constant given by the second term of the equation (4). This property can be interpreted as presence of long memory in such series when we analyse them through the behavior of the sample ACF using the whole sample. Thus, here we show how shifts in the means could explain long memory effect in the sample ACF. This effect can be observed in a lot of data sets generated a priori by short memory processes. We give now some examples.

Let be a two states Markov switching process

$$Y_t = \mu_{s_t} + \varepsilon_t. \quad (5)$$

The process  $(s_t)_t$  is a Markov chain which permits to switch from one state to another one with respect to the transition matrix  $P$ , whose elements are the fixed probabilities  $p_{ij}$  defined by  $p_{ij} = P[s_t = i | s_{t-1} = j]$ ,  $i, j = 1, 2$ ,  $0 \leq p_{ij} \leq 1$  and  $\sum_{i,j=1}^2 p_{ij} = 1$ . The process  $(\varepsilon_t)_t$  is a strong white noise, independent to  $(s_t)_t$ . The process  $(Y_t)_t$  switches from level  $\mu_1$  to level  $\mu_2$  with respect to the Markov chain. This model has been studied by Andel

(1993) and then by Poskitt and Chung (1996). The theoretical behavior of the autocorrelation function of such a model, under stationarity conditions, is similar to the one of an ARMA(1,1) process, thus it has theoretically a short memory behavior. Indeed, its autocorrelation function decreases with an exponential rate towards zero for large  $h$ . Nevertheless respecting the stationary conditions, it is possible to exhibit sample ACFs which have a very slow decay, (see Guégan and Rioublanc, 2005, for a lot of examples). This effect can be explained in that case by the behavior of the second term of the relationship (4). Note that this behavior is also observed for more general Markov switching  $(Y_t)_t$  defined  $\forall t$ , by  $Y_t = \mu_{s_t} + \phi_{s_t} + \varepsilon_t$ . Their theoretical ACF decrease with an exponential rate towards zero for large  $h$ , see Timmerman (2000), but it is also possible to exhibit a lot of patterns of the sample ACFs decreasing in an hyperbolic way towards zero, see Guégan and Rioublanc (2003).

Coming back to the model (5), it is possible to be more precise concerning the slow behavior of the whole sample ACF. We can show that it does not depend only on the means  $\mu_i, i = 1, 2$ , but also on the transition probabilities  $p_{ii}, i = 1, 2$ , (see, Guégan and Rioublanc, 2005, for more details). In case of two high transition probabilities, the autocorrelation function of model (5) decreases slowly. When  $p_{11} + p_{22}$  is close to 1, the decay of the autocorrelation function is quick. 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 number influences also the behavior of the autocorrelation function. A close relationship between  $p_{ii}$  and  $n_{s_t}$  is not available but empirical studies can illustrate this situation. In the Section 9, we provide some illustrative examples. This remark shows that the effect of non-stationarity is more complex than the only behavior of the sample ACF. Further works need to be developed.

A StopBreak model permits also to switch from one state to another one. let be the process defined by

$$X_t = \mu_t + \varepsilon_t, \quad (6)$$

where

$$\mu_t = (1 - \alpha\delta_t)\mu_{t-1} + \delta_t\eta_t, \quad (7)$$

with  $\alpha \in [0, 2]$ ,  $(\delta_t)_t$  a sequence of independant identically distributed Bernouilli  $(\lambda)$  random variables and  $(\varepsilon_t)_t$  and  $(\eta_t)_t$  two independent strong white noises, (Breidt and Hsu, 2002). It is know that for fixed  $\lambda$ , this process - which models switches with breaks - has short memory behavior. This one is observed

for long samples, but as soon as the jumps are rare relatively to sample size, the short memory behavior does not appear so evident. Even if the asymptotic theory describes a short memory behavior, a sample experiment for a short sample size looks much like the corresponding characteristics for long memory processes. This effect can be explained by the relationship (4). Indeed, for different values of  $\alpha$  and  $\lambda$ , the means  $\mu_1$  and  $\mu_2$  are different, thus the second term of the relationship (4) is bounded and the sample ACF of model (6)-(7) does not decrease towards zero. We provide an example of this behavior in Section 12. This effect is also due to the number  $n_{s_t}$  of changes between the states. Note that this model (6)-(7) includes a lot of models with switches. For example, for  $\lambda = 0$ , we have a constant; for  $\alpha = 0$  and  $0 < \lambda < 1$ , we get the mean plus noise process; for  $\alpha = 0$  and  $\lambda = 1$ , we get a random walk; for  $0 < \alpha < 2$  and  $\lambda = 1$ , we get an AR(1) process; for  $\alpha = 1$  and  $\lambda = 1$ , we get a white noise; for  $\alpha = 1$  and  $0 < \lambda < 1$ , we come back on some regime-switching models. Thus, a more understanding of the class of all the models has to be done to control the non-stationary behavior observed on different sample experiments.

Consider now a SETAR process whose a simple representation is

$$Y_t = \mu_1 I(Y_{t-1} > 0) + \mu_2 I(Y_{t-1} \leq 0) + \varepsilon_t, \quad (8)$$

where  $I(\cdot)$  is the indicator function. This model permits to shift from the mean  $\mu_1$  to the mean  $\mu_2$  with respect to the value taken by  $Y_{t-1}$ . SETAR processes are known to be short memory, see Tong (1990). But it is also possible to exhibit sample ACFs which present slow decay. This slow decay can also be explained by the second term of the relationship (4), and also by the time spent in each state. We exhibit some simulations experiment in Section 13.

Mikosch and Starica (2004) using the proposition 4.1 have explained why returns which exhibit volatility - that they model using GARCH processes (which are short memory) - have a sample ACF which decreases slowly. They consider processes  $(Y_t^2)_t$  and  $(|Y_t|)_t$  in order to model the variance of the log returns. Their results can be extended for any process  $(Y_t^\delta)_t$ ,  $\delta \in R$ . Indeed, Mikosch and Starica (2004) use a linear function of past square residuals  $\varepsilon_t^2$  which are correlated to each other to model this volatility. But we can remark that  $|\varepsilon_t|^\delta$  or  $|Y_t|^\delta$ , where  $\delta$  is a positive number, are also correlated with each other, thus, we can compute the correlation between the  $|\varepsilon_t|^\delta$  or  $|Y_t|^\delta$ , and extend the result of the proposition 4.1. Recall that in 1996, Dinh and Granger propose to model the volatility of the log returns using the process

$(Y_t^\delta)_t$ . We define now the sample size  $T$ ,  $Y^\delta = (Y_1^\delta, Y_2^\delta, \dots, Y_T^\delta)$ ,  $\delta \in R^+$ , that we can consider, using the previous notation as the reunion of  $r$  subsamples  $Y^\delta = ((Y_1^{(1)})^\delta, \dots, (Y_{T_{q_1}}^{(1)})^\delta, (Y_{T_{q_1+1}}^{(2)})^\delta, \dots, (Y_{T_{q_{r-1}+1}}^{(r-1)})^\delta, \dots, (Y_T^{(r)})^\delta)$ .

**Lemma 4.2** *Let be  $r$  subsamples  $(Y_1^{(i)})^\delta, \dots, (Y_{T_{q_i}}^{(i)})^\delta$ ,  $i = 1, \dots, r$  and  $\delta \in R^+$ , coming from the sample  $Y^\delta$ , each subsample corresponding to a stationary distinct ergodic process with finite second order moments, whose sample covariance is equal to  $\tilde{\gamma}_{(Y^{(i)})^\delta}(h)$ , then the sample autocorrelation function  $\tilde{\gamma}_{Y^\delta}$  of the sample  $Y^\delta$  is such that:*

$$\tilde{\gamma}_{Y^\delta}(h) \rightarrow \sum_{i=1}^r p_i \tilde{\gamma}_{(Y^{(i)})^\delta}(h) + \sum_{1 \leq i \leq j \leq r} p_i p_j (E(Y^{(j)})^\delta - E(Y^{(i)})^\delta)^2, \quad h \rightarrow \infty. \quad (9)$$

**Proof:** To get this result, we replace  $Y_t$  by  $Y_t^\delta$  and  $(Y_t^{(i)})$  by  $(Y_t^{(i)})^\delta$ , for  $i = 1, \dots, r$ , in the proof given in Annex 8.

Under the property of stationary ergodicity, the ACF of each process has an exponential decay. Thus, the sample  $Y^\delta$  has its sample ACF  $\tilde{\gamma}_{Y^\delta}(h)$  that decays quickly for the first lags and then approach positive constants given by  $\sum_{1 \leq i \leq j \leq r} p_i p_j (E(Y^{(j)})^\delta - E(Y^{(i)})^\delta)^2$ . This explains some long memory effect observed on the ACF of the series when we analyse them with the whole sample  $Y^\delta$ . Here we show how shifts in the variances (modelled using  $Y_t^\delta$ ) could explain long memory effect inside the sample ACF.

Thus, the lemma 4.2 is very useful to understand what happens when we observe shifts inside the volatilities. Mikosch and Starica (2004) develop a lot of situations using ARCH and GARCH processes. Their approach can be easily extended with A-P-ARCH models introduced by Ding and Granger (1996). Moreover, Markov switching models are also useful to model existence of shifts inside volatilities. Let be the simplest model  $(Y_t)_t$  whose the volatility parameter depends on a Markov chain  $(s_t)_t$  (defined as previously), such that:

$$Y_t = \mu + \sigma_{s_t} \varepsilon_t. \quad (10)$$

The autocorrelation function of this process  $(Y_t)_t$  is similar to the one of a white noise:

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

Thus, this model will always exhibit a short memory behavior looking at its ACF. the lemma 4.2 (for  $\delta = 1$ ), explains why the sample autocorrelation function of process (10) decreases exponentially quick towards zero thanks

to the first term (the second term being zero). We exhibit such an example of process (10) in Section 11. This example illustrates the important difference between the models (5) and (10) in terms of volatility modelling. An empirical approach cannot create long memory using the process (10).

Another non stationary setting arises when we are in presence of different seasonals inside real data sets. Presence of seasonals indicate that there exist significant correlations between random variables at present  $t$ , and in the past and in the future. This correlation creates dependence in each subsequence, which implies non-stationarity on the whole sample, and then a slow decay of the autocorrelation function. Thus, for this kind of data sets, the non decreasing of the sample ACF comes from the first term of the relationship (4).

In summary, existence of volatility, structural changes like jumps or several seasonals are important sources of problems to analyse correctly the behavior of the sample ACF. These data sets, traditionally are modelled by short memory processes (related GARCH, StopBreak, etc), but the behavior of their sample ACF suggests to model them using long memory processes like FIGARCH or GIGARCH processes, (Bollerslev and Mikkelsen, 1996, Guégan, 2003b and Ashley and patterson, 2007). This situation can provoke confusion.

Moreover, the slow decay of the sample ACF can be observed in other situations. It can be dû to the presence of distorsion inside the data. Distorsion corresponds to data sets for which the variance is infinite. This effect, dû to an important shock, provokes a strong persistence. Note that, if the variance is infinite, the ACF is not well defined. Aggregated models can also exhibit long memory behavior. For instance, AR(1) processes for which the parameters follow a Beta distribution on the range  $(0, 1)$ , Granger (1980). High frequency data often present multiple volatility components. This means in particular that there exists a mixture of distributions beside these behaviors, which correspond to multiple seasonalities inside these high frequencies data. If we make temporal aggregation of these intraday data, then we can purge them with periodicities of less than one day. In another way, this aggregation represents heterogeneous information arrival processes, some with short - run volatility dependencies, others possessing more highly persistent volatility patterns. As time passes, the more highly persistent processes remain influential and this aggregation shows that the dependence inside the volatility process dissipates at a slow hyperbolic rate.

These different kinds of situations increases the degree of confusion between different kinds of data sets:

1. data generated by true long memory processes corresponding to existence of cycles,
2. data generated by sequences of short memory processes whose the whole sample ACF has a behavior assimilated to the one of processes with long cycles,
3. data coming from specific transformations (aggregation),
4. data submitted to a strong shock (distorsion) or corresponding to specific sampling (high frequency).

If the study of the behavior of the ACF can appear limited in that context, it seems more interesting to analyse the existence of an invariant empirical distribution function for this kind of data sets. In case of distorsion, the research of extreme values distributions will be privilege. The aggregated data can be characterized by very specific distribution functions depending on the dynamics of the coefficients. The high frequency data need generally to use mixtures of distributions. For all these cases, we are very far from a classical framework, often based on elliptical distributions. In some previous examples, we can also noted that some specific features imply that the variance of the distribution function cannot exist. This situation questions the use of the sample ACF and we need to develop other strategies to identify the underlying process(es). In particular models whose distribution functions belong to the generalized Pareto distribution functions class.

We can also remark that, when we use  $\delta = 3$  or  $\delta = 4$  inside the expression (9), we obtain an information which concerns the behavior of the sample autocorrelation function using the random variables  $Y_t^3$  or  $Y_t^4$ . Thus, if the second term in the right hand side of the expression (9) is not zero, this means that the third moment (which characterizes the skewness) and the fourth moment (which characterizes the kurtosis) are not equal. This gives us a deeper understanding of the real data set under study.

This means in particular that when real data sets present some non stationary features, if we want to adjust the same model using the whole sample, we can use different strategies using, models with dynamic parameters, or models with a distribution function evolving in a dynamical way all along the whole sample. In order to decide between a unique distribution with dynamic

parameters or a set of different distributions, we have to introduce a local analysis. We begin to specify the notion of local stationarity.

## 5 The local stationarity

In the previous Section we have observed that, when the global stationarity is not realized, then the empirical statistics fail to have the true theoretical behavior. In particular, some kind of specific behavior of the sample ACF is observed (slow decay) even though we expect another kind of behavior (exponential decay).

Let us now consider  $Y_1, \dots, Y_T$  observations from an arbitrary non-stationary process  $(Y_t)_{t=1, \dots, T}$ : we work here for a finite sample size. For simplicity, to model non stationary process we assume that this process has the following representation:

$$Y_t = \sum_{j=0}^{\infty} a_j(t) \varepsilon_{t-j},$$

assuming that  $(\varepsilon_t)_t$  is a strong white noise. Then, we are going to assume that we "observe"  $(a_j(t))$  on a finer grid (but on the same interval), that is, we have :

$$Y_{t,T} = \sum_{j=0}^{\infty} a_j\left(\frac{t}{T}\right) \varepsilon_{t-j}, \quad \text{for } t = 1, \dots, T. \quad (11)$$

( $a$  is now rescaled to the interval  $[0, 1]$ ). For such a process, we can define its first and second order moments. The mean:  $\mu\left(\frac{t}{T}\right)$ ; the variance:  $\sigma\left(\frac{t}{T}\right)$  and the covariance :  $\gamma\left(\frac{t}{T}, h\right) = \text{cov}(Y_{\left(\frac{t}{T}\right)}, Y_{\left(\frac{t}{T}+h\right)})$ .

**Definition 5.1** *A sequence of stochastic process  $(Y_{t,T})_{t=1, \dots, T}$ , is called locally stationary if there exists a representation:*

$$Y_{t,T} = \mu\left(\frac{t}{T}\right) + \int_{-\infty}^{\infty} \exp(i\lambda t) A\left(\frac{t}{T}, \lambda\right) dF(\lambda), \quad (12)$$

where  $F(\lambda)$  is a spectral distribution on  $[-\pi, \pi]$  with  $\overline{F}(\lambda) = F(-\lambda)$ , and where

$$A\left(\frac{t}{T}, \lambda\right) = \sum_{j=0}^{\infty} a_j\left(\frac{t}{T}\right) \exp(-i\lambda j).$$

The representation (12) is the time varying spectral representation (similar to the analogous representation for stationary process) of the process  $(Y_t)_t$ .

The equation (11) has not exactly but only approximately a solution of the form (12). If (12) holds then the process has locally stationary behavior.

The above definition does not mean that a fixed continuous time process is discretized on a finer grid as  $T$  tends to infinity. Instead it is an abstract setting for asymptotic statistical inference which means that with increasing  $T$ , more and more data of each local structure are available.

With this framework and the above definition of a locally stationary process we deal an adequate set up for asymptotic theory. Then, the classical asymptotic theory for stationary process is a special case of the theory developed for the process (12).

Letting  $T$  tends to infinity no longer means looking in the future. Nevertheless a prediction theory using the representation (12) is possible. For instance, we assume that  $Y_{t,T}$  is observed for  $t \leq T/2$  (i.e. on the interval  $(0, 1/2)$ ) and one predicts the next observation.

The time varying spectral density associated to the process (12) is given by:

$$f(u, \lambda) = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \text{cov}(Y_{[uT-s/2],T}, Y_{[uT+s/2],T}) \exp(-i\lambda s). \quad (13)$$

The spectral density can be used to estimate the parameters of the model using the Whittle likelihood, see Whittle (1951) and Dalhaus (1996). On regular assumptions the estimates are asymptotically Gaussian.

In a practical application, the problem of model selection arises. Now the asymptotic theory developed for non stationary process like (12) still hold if a stationary process is fitted, see Dalhaus, 1997.

Most of the previous discussion is based on the fact that the real data sets are composed of juxtaposed models like AR, ARCH, SETAR, Switching and so on: this is not realistic. But it is possible to think that in practice the data switches from one regime to another one with respect to a persistent shock. Thus, we can imagine a model permitting sharp switches with parameters which change continuously. This behavior is related to the non stability of the invariant distribution of the underlying process or its heavy tailness behavior. Thus, we are close to the situation created by concatenated processes. The question is how to model this kind of behavior, but also how to detect it. We cannot use the global property of the model that we want to

adjust on the data sets and we know that this one is locally stationary. The asymptotic theory on local stationary processes tells us that we can work with the same tools developed for global stationary processes. But, we need to verify that the statistics that we estimate in the former case has a sense.

Detecting (or building) local stationary processes from the data sets need to develop new approaches based on local stationarity. One possibility is to build subsets which are locally stationary, from the original data sets, with respect to the different kinds of non stationarity.

In the previous Section, we show that the main problem is the presence of a kind of long memory behavior suggested by the sample ACF. This means that in most cases, if we use the whole sample we are going to estimate a long memory parameter  $d$  using parametric models like the FARMA models or the Gegenbauer models. This method creates some confusion in the modelling of the data because this behavior is closer to a spurious behavior than to a true behavior. Moreover a lot of statistics used in the detection and estimation of long memory are characterized by their lack of power to discriminate between possible scenarios, in particular in presence or not of true long memory behavior, see for several examples Bhattacharya, Gupta and Waymire (1983), Teverovsky and Taqqu (1997), Bisaglia and Guégan (1997) and Andersen and Bollerslev (1997). In another hand, in non-stationary specific cases, we have the robustness of the estimators of the long memory parameters, see for instance Hurwitch and Ray (1995) for the FARMA models and Lobato and Velasco (2000) for the Gegenbauer models.

In the Section 10 we illustrate, for Markov switching models like (5) the spurious long memory effect creates with respect to the choice of the probability  $p_{ii}$ . This example permits to convince of the necessity to work with a local approach to avoid confusion in modelling.

## 6 Detection of local stationarity

This Section concerns the discussion of several points of views in order to detect and model local stationarity. It is mainly a methodology discussion. In a first insight, we can consider an extension of the approach proposed by Starica and Granger (2005) using moments beyond the second moments. Another possibility is to transform an univariate approach in a multivariate one and to detect the copula which links the different invariant probability distribution functions adjusted on each subsample.

## 6.1 Approach based on empirical moments

In a recent and interesting paper Starica and Granger (2005) propose to test successively on different subsamples of a time series  $(Y_t)_t$  the invariance of the spectral density. They propose a specific test and their strategy is the following. They consider a subset  $(Y_{m_1}, \dots, Y_{m_2})$ ,  $\forall m_1, m_2 \in N$ , on which they apply their test and build confidence intervals. Then, they consider another subset, for some  $p \in N$ ,  $(Y_{m_2+1}, \dots, Y_{m_2+p})$ . They apply again the test and verify if the value of the statistic belongs to the confidence interval previously built or not. If it belongs to the confidence interval, they continue with a new subset. If not, they consider  $(Y_{m_1}, \dots, Y_{m_2+p})$  as an interval of homogeneity and analyse the next subset  $(X_{m_2+p+1}, \dots, X_{m_2+2p})$  and define new confidence intervals from their statistics. When they finish, they estimate a model on each subset of homogeneity (which means that the parameters change). They use these intervals to forecast.

The approach proposed by Starica and Granger (2005) is based on the spectral density which is built using the second order moments of a process, we propose to consider an approach based on empirical moments above moments of order 2. Here, we suggest to estimate the different moments on the whole sample and on subsamples and to compare their evolution, in a first insight by kinds of Student tests or Wald tests. This procedure permits in a first insight to have an idea of existence of non-linearity and non-stationarity.

Indeed, a cumulant test can be build that will permit to obtain intervals of homogeneity, see Fofana and Guégan (2007). This test is built using the spectral representation of the cumulants of order  $k$ , denoted  $c_k$ . For a stationary process  $(Y_t)_t$ , we consider the spectral density of cumulant of order  $k$  denoted  $f_{c_k, Y}$ . For a sample size  $T$ , we can build its estimate,  $I_{c_k, Y, T}$ . Then we consider the following statistic :

$$\tilde{T}(T, Y) = \sup_{\lambda \in [-\pi, \pi]} \left| \int_{[-\pi, \pi]^{k-1}} \left( \frac{I_{c_k, Y, T}(z)}{f_{c_k, Y}} - \frac{\tilde{c}_k}{c_k} \right) dz \right|, \quad (14)$$

where  $\tilde{c}_k$  is an estimate of  $c_k$ . We show that - under the null that the cumulants of order  $k$  are invariant on the subsamples we consider - this statistic  $\tilde{T}(T, Y)$  converges in distribution to  $\frac{(2\pi)^{k-1}}{c_k} B(\sum_{j=1}^{k-1} \lambda_j)$  where  $B(\cdot)$  is the Brownian bridge. Thanks to the knowledge of the critical values of this statistic, we can build homogeneity intervals, using moving windows. With this statistic, we use a complete information from the data set in order to build homogeneity intervals on which we can get an invariant distribution function.

Another approach could consist in testing the changes of models on the different subsets and to detect if we are in presence of spurious long memory. For this strategy we can use LR or LM tests, see for applications Charffedine and Guégan (2006) and references therein.

## 6.2 Approach based on the empirical distribution function

Now it seems also important to investigate the distribution function of the whole sample and on the subsamples. Indeed, the distribution function of the underlying process is useful in order to take into account the non linearity of the process which can create the non stationarity and to propose a robust risk management strategy for data sets that compose a portfolio. The previous work on the empirical moments permit to know when there is a change in the distribution function which characterizes the data under study. We are going to use these intervals of homogeneity to define a sequence of invariant distribution functions.

let be a process  $(Y_t)_t$  whose distribution function is  $F_Y$  and assume that we observe  $Y_1, \dots, Y_T$ , a sample size  $T$ . We are interested to know the joint distribution function  $F_Y = P[Y_1 \leq y_1, \dots, Y_T \leq y_T]$  for the process  $(Y_t)_t$ . The knowledge of this distribution function will permit to do forecasting or to make a risk management strategy for the data sets under interest. Now we assume that the process  $(Y_t)_t$  is non stationary, thus its distribution function is not invariant on the whole sample. We give an example on figure 1. On this figure, we have identified a sequence of homogeneity intervals characterized by changes in mean or in variance for instance. Thus, we observe that we can adjust different distribution functions on each subsample noted (1), (2), (3), (4).

---

Insert Figure 1

Example of a sequence of invariant distribution functions

---

Thus, from the observations  $Y_1, \dots, Y_T$ , we want to build a sequence of distribution functions which are invariant on each subsamples. For the moment we assume that the subasamples are clearly identified. We choice here, for simplicity, to adjust parametric distribution functions on each subsample. These distribution functions can belong to the same class but with different

parameters or can belong to different classes of distribution functions. We will use tests, as the Kolmogorov-Smirnov test or the  $\chi^2$  test to detect the parametric distributions.

Now we assume that we have identified a sequence of  $r$  stationary ergodic subsamples  $Y_1^{(i)}, \dots, Y_{Tq_i}^{(i)}$ ,  $i = 1, \dots, r$ , each characterized by an invariant distribution function  $F_{Y^{(i)}}$ ,  $i = 1, \dots, r$ . As soon as, we have built the stationary subsamples with their invariant distribution function, we are going to build the empirical distribution characterizing the whole sample. In a certain sense we want to get some meta-distribution for the data set. To determine this distribution function  $F_Y$ , we can use a copula distribution function. We develop now this strategy. We briefly recall the notion of copulas in a two dimensional setting and we will extend it to an  $r$  dimensional setting.

Consider a general random vector  $Z = (X, Y)^T$  and assume that it has a joint distribution function  $F(x, y) = \mathbb{P}[X \leq x, Y \leq y]$  and that each random variable  $X$  and  $Y$  has a *continuous* marginal distribution function respectively denoted  $F_X$  and  $F_Y$ . It has been shown by Sklar (1959) that every 2-dimensional distribution function  $F$  with margins  $F_X$  and  $F_Y$  can be written as  $F(x, y) = C(F_X(x), F_Y(y))$  for an *unique* (because the marginals are continuous) function  $C$  that is known as the copula of  $F$  (this result is also true in the  $r$ -dimensional setting). Generally a copula will depend almost on one parameter, then we denote it  $C_\alpha$  and we have the following relationship:

$$F(x, y) = C_\alpha(F_X(x), F_Y(y)). \quad (15)$$

Here, a copula  $C_\alpha$  is a bivariate distribution with uniform marginals and it has the important property that it does not change under strictly increasing transformations of the random variables  $X$  and  $Y$ . Moreover, it makes sense to interpret  $C_\alpha$  as the dependence structure of the vector  $Z$ . In the literature, this function has been called "dependence function" by Deheuvels (1978), "uniform representation" by Kimeldorf and Sampson (1975) and "copula" by Sklar (1959). It is the last denomination which is now the more popular, and we use it here.

Practically, to get the joint distribution  $F$  of the random vector  $Z = (X, Y)^T$  given the marginal distribution functions  $F_X$  and  $F_Y$  of  $X$  and  $Y$  respectively, we have to choose a copula that we apply to these margins. There exists different families of copulas: the elliptical copulas, the archimedean copulas, the meta-copulas and so on. We refer to Ling (1965), Cambanis *et al.* (1981), Fang *et al.* (2002) and Joe (1997) for a precise presentation of all these cop-

ulas.

In order to choose the best copula adjusted for a pair of random variables, we need to estimate the parameters of the copula and to estimate the copulas. Different strategies can be done to solve these problems.

To estimate the parameters, we can use:

1. The Kendall's tau. There exists a fairly relationship between this coefficient and the parameters of the Gaussian and Student copulas for elliptical family. For applications, see Guégan and Ladoucette (2002, 2004).
2. The relationship between the Kendall's tau and the parameter  $\alpha$  of the copula  $C_\alpha$ , in case of the Archimedean copulas. These relationships can be found for instance in Nelsen, 1999.

To estimate the copula, there exists several possibilities:

1. A non-parametric approach. We compare the empirical non parametric copula which has been introduced by Deheuvels, (1979), with the empirical joint distribution function estimated from the data set.
2. A parametric approach. We use the maximum likelihood function. For a set of copulas, we estimate the parameters of the copula, following the works of Oakes (1982) maximising the pseudo log-likelihood function defined, for  $\alpha$ , by:

$$\tilde{\alpha} = \arg \max_{\alpha} \sum_{k=1}^T \log L(\alpha; \hat{F}_X(x_k), \hat{F}_Y(y_k)), \quad (16)$$

where  $\hat{F}_X$  is defined for instance by:

$$\hat{F}_X(x_p) = \frac{1}{T+1} \sum_{k=1}^T 1_{\{x_k < x_p\}}, \quad (17)$$

with  $L(\alpha; u, v) = \frac{\partial^2}{\partial u \partial v} C_\alpha(u, v)$ .

3. A semi-parametric method: the IFM (inference for the margins) method. We decompose the previous approach in two steps. As a first step, we estimate the margin's parameters, by performing the estimation of the univariate marginal distribution (this has been done in the previous

Subsection). In a second step , we perform the estimation of the parameter copula using a expression close to (16), but it is not the same because we already have estimated the margins. For details see Joe and Xu (1996).

Now, to discriminate between a lot of the copulas, we can use several methods:

1. the Akaike Information criteria provides by each estimated parameter: the minimum value of the AIC gives the best copula, in that sense.
2. a diagnostic based on the  $L^2$  distance computed between the distribution functions, as:

$$D_2 = \sum_{m=0}^T \sum_{n=0}^T \left| C_{\hat{\alpha}}(\hat{F}_X(x_m), \hat{F}_Y(y_n)) - \hat{F}(m/T, n/T) \right|^2,$$

where  $\hat{F}_X$  is defined in (17) and  $\hat{F}$  is the empirical distribution function. Thus, for each pair of process  $((X_t)_t, (Y_t)_t)$ , we get  $(\hat{F}_X, \hat{F}_Y)$ . The copula  $C_{\hat{\alpha}}$  for which we will get the minimum distance  $D_2$  will be chosen as the best approximation, in that sense.

3. a graphical method. Let be a copula  $C$ , if  $U$  and  $V$  are two uniform random variables then:

$$C(V|U) = \frac{\partial C}{\partial U}(U, V)$$

and

$$C(U|V) = \frac{\partial C}{\partial V}(U, V)$$

are also uniformly distributed. We can use this property to compare the joint empirical distribution function and the different copulas that we want to adjust on our data set and for which we have estimated the parameters. This can be done using the QQ-plot method. To apply this method, we need to know the partial derivatives of the copulas. In order to use this method, we provide the derivatives of some of the most famous Archimedean copulas used in applications.

- Gumbel Copula:

$$G_{\alpha}(v|u) = G_{\alpha}(u, v) \frac{|\log u|^{\alpha}}{u} \left( |\log u|^{\alpha+1} + |\log v|^{\alpha+1} \right).$$

- Cook et Johnson copula:

$$J_\alpha(v|u) = \left(1 + u^\alpha(v^{-\alpha} - 1)\right)^{-(1+\alpha)/\alpha}.$$

- Ali-Mikhail-Haq copula:

$$A_\alpha(v|u) = \frac{v(1 - \alpha(1 - u)(1 - v)) - uv\alpha(1 - v)}{(1 - \alpha(1 - u)(1 - v))^2}.$$

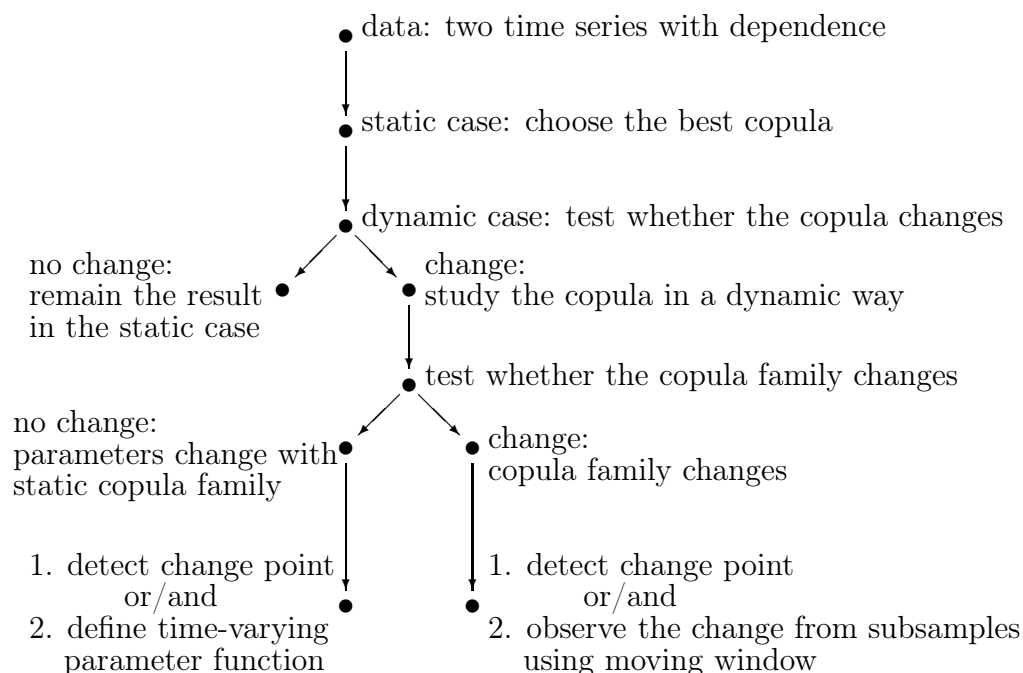
- Frank copula:

$$F_\alpha(v|u) = \left(1 + \frac{(\alpha^u - 1)(\alpha^v - 1)}{\alpha + 1}\right)^{-1} \frac{\alpha^u}{\alpha + 1}(\alpha^v - 1).$$

As the distributions of  $\hat{F}_X$  and  $\hat{F}_Y$  are uniform, for each copula  $C_\alpha$ , we represent the empirical distribution function  $C_{\hat{\alpha}}(\hat{F}_Y|\hat{F}_X)$  and the uniform distribution function. More close to a line is the representation, better is the adjustment between  $\hat{F}$  and  $C_{\hat{\alpha}}$ . For some applications, we refer to Caillault and Guégan (2005)

We have presented the method to adjust a copula in case of two processes. We can extend dynamically the adjustment for a  $r$  sample of distribution functions  $F_{Y^{(i)}}$ ,  $i = 1, \dots, r$ . Thus, we will work step by step working with two subsamples at each step. This will permit to detect if the copula that we look after is the same all along the samples, or if it is the same but with different parameters or if it changes with respect of the subsamples. We will use the method developped in Guégan and Zhang (2006) to retain the best copula. The process is the following. First of all, we consider the static case for which the copula does not change and the parameters can change, and secondly the dynamical case for which the copula family changes. We can use a series of nested tests based on the conditional pseudo copula (see, e.g., Fermanian and Wegkamp, 2004) and goodness-of-fit (GOF) test (see, e.g., Fermanian, 2005) to decide in what case we are. We test at first whether the copula does change during the considered time period. If the copula seems changeless, we remain the result in the static case. If we detect some changes, we have to decide between the parameters and the copulas. If the result of the test shows that the copula family may not change, we can only deal with the changes of copula parameters. Otherwise, if the result of the test tells us that the copula family may change, then we handle the changes of copula family.

Through the nested tests mentioned above, we analyze the changes. If we admit that only the copula parameters change, we can apply the change-point analysis as in Csörgő and Horváth (1997), Gombay and Horváth (1999), Dias and Embrechts (2003) to decide the change time. Moreover, considering that change-point tests have less power in case of “small” changes, we can assume that the parameters change according to a time-varying function of predetermined variables. In the other case, if we cannot deny the change of the copula’s family, we can apply  $U$ -statistics to change-point analysis as in Gombay (2001) and Gombay and Horváth (2002) to detect the change time. More tractably, we can decide the best copula on subsamples using the moving window, and then observe the changes. Then we can summarize the different steps as follows.



We can use this method on the different subsamples of our present study and determine an unique dynamic copula  $C_{\alpha_t}$  linking the sequences of invariant distribution functions  $F_{Y^{(i)}}$ ,  $i = 1, \dots, r$  which would permit to characterize the joint distribution function  $F_Y$ , using a parametric model for the dynamical parameter  $\alpha_t$ . But we can also by this method built a sequence of copulas if we detect different copulas linking several subsamples.

The method developed in this Section seems interesting in the sense that it permits to give new developments concerning several problems pointed in the introduction.

1. Concerning forecasting in presence of non stationarity: we can use the linking copula  $C_{\alpha_t}$  to get a suitable forecast for the process  $(Y_t)_t$  assuming the knowledge of the whole information set  $I_T = \sigma(Y_t, t < T)$ : we compute  $E_{C_{\alpha_t}}[Y_{t+h}|I_T]$ . We can also decide to do forecast using, as an information set, one or several subsamples defined as homogeneity intervals. For instance if we consider the last homogeneity interval, we will compute  $E_{F_Y^{(r)}}[Y_{t+h}|I_r]$ , where  $I_r$  is the information set generated by the random variables  $(Y_{T_{q_{r-1}+1}}^{(r)}, \dots, Y_T^{(r)})$  and  $F_Y^{(r)}$ , the margin characterizing this subset. If we take two homogeneity intervals, we will compute the expectation under the copula linking the two margins corresponding to each subsample and the information set that is the reunion of the two subsamples, etc.
2. Concerning the risk management strategy and the computation of the measures of risks associated to a portfolio which contains financial assets which are non stationary, we will use - for each asset - the corresponding distribution function  $C_{\alpha_t}$ . This one will appear as a margin of the global distribution function of the portfolio, computed also by copulas method, see Caillault and Guégan, 2005 for instance.
3. This new approach will permit also to solve the problem of the knowledge of the distribution function for specific models. Indeed, we can use this approach to compute the distribution function of any Markov switching model, for instance.

## 7 conclusion

In this paper, we have discussed deeply the influence of non-stationarity on the stylized facts observed on the data sets and on specific statistics. For these statistics, a lack of robustness is observed in presence of non stationarity.

To detect existence of local or global stationarity on data sets, we propose different strategies. One is based on the study of the empirical moments (beyond to 2). Another one concerns the detection of invariant distribution functions on different subsamples, and then their link using a dynamical copula on the whole sample. We propose nested tests to construct this copula. The approach that we propose here is interesting in the sense that it uses a vector approach.

We have also specified in the previous section, several new developments that this approach will permit. Most of these points would be developed in companion papers. Now we want to specify also some extended researches that are link to the new ideas developed in this paper. They concern:

- The use of the change point theory to verify the date at which we get the homogeneity intervals. This could be a nice task. Indeed, most of the works concerning the change point theory concern detection of breaks in mean or in volatility. These works have to be reexamined taking into account the fact that breaks can provoke spurious long memory. Indeed, in that latter case, the using of the covariance matrix appears as the problem in the sense that we cannot observe change point in the covariance matrix.
- The time spend in each state when breaks are observed. This random variable appears very important in order to characterize the existence of states. In a lot of papers, empirical evidence has been discussed. It will be interesting to know exactly (or to know how to estimate) the distribution function of this time spend in each state, denoted  $n_{s_t}$  in this paper.
- The discussion of models taking into account sharp switches and time varying parameters. A theory has to be developed to answer to a lot of questions coming from practitioners. If the model proposed by Hyung and Franses (2005) appears interesting in that context, because it nests several related models by imposing certain parameter restrictions (AR, ARFI, STOPBREAK, models for instance, etc.), more identification theory concerning this model need to be done to understand how it can permit to give some answer to the problematic developed in this paper. Recent developments proposed by Baillie and Kapetanios (2007) can be also considered.

## References

- [1] Andél J. (1993). A Time Series Model with Suddenly Changing Parameters, *Journal of Time Series Analysis*, 14, 111-123.
- [2] Andersen T.G., Bollerslev T. (1997). Heterogeneous information arrivals and return volatility dynamics: un covering the long run in high frequency returns, *The journal of Finance*, 52, 975 - 1005.

- [3] Arteche J. (2003). Semiparametric robust tests as seasonal or cyclical long memory time series, *Journal of Time Series Analysis*, 23, 251 - 285.
- [4] Ashley R., Patterson D. (2007). Apparent long memory in a times series and the role of trends: a moving mean filtering alternative to the fractionally integrated model, Presentation to the SNDE in Paris, France, March 2007.
- [5] Baillie R., Kapetanios G. (2007). Non-linear models with strongly dependent processes and applications to Forward Premia and real Exchanges Rates, Presentation to the SNDE in Paris, France, March 2007.
- [6] Bhattacharya R.N., V.K. Gupka, E. Waymire (1983). The Hurst effect under trend, *Journal of applied probability*, 20, 649 - 662.
- [7] Beran, J. (1994). *Statistics for Long Memory Processes*, Chapman and Hall, New-York.
- [8] Bisaglia L., Guégan D. (1997). A review of techniques of estimation in long memory processes: application to intra-day data , *Computational Statistics and Data Analysis*, 61 - 81, 1997.
- [9] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 31, 307-327.
- [10] Bollerslev, T., Mikkelsen H.O. (1996). Modelling and pricing long memory in stock market volatility, *Journal of Econometrics*, 73, 151 - 174.
- [11] Bollerslev, T., Chou R.Y., Kroner K.F. (1992). Arch modelling in finance: A review of the theory and empirical evidence, *Journal of Econometrics*, 52, 5 - 59.
- [12] Breidt F.J., Hsu N.J. (2002). A class of nearly long-memory time series models, *International Journal of forecasting*, 18, 265 - 281.
- [13] Brockwell P.J., Davis R.A. (1996). *Introduction to time series analysis*, Springer Texts in Statistics, Springer Verlag, New York.
- [14] Caillault C., Guégan D. (2005). Empirical Estimation of Tail Dependence Using Copulas. Application to Asian Markets, *Quantitative Finance*, 5, 489 - 501.
- [15] Cambanis S., Huang S., Simons G. (1981). On the theory of elliptical contoured distributions, *JMA*, 11,, 368-385.

- [16] Charffedine L., Guégan D. (2006). Is it possible to discriminate between structural break models and long memory processes? An empirical approach, WP 2006 - 07, CES - Cachan, France.
- [17] Csörgő, M., Horváth, L., (1997). *Limit Theorems in Change-point Analysis*. Wiley, Chichester.
- [18] Dalhaus R. (1997). Fitting time series models to nonstationary processes, *The annals of Statistics*, 25, 1-37.
- [19] Deheuvels P. (1978). Caractérisation complète des lois extrêmes multivariées et de la convergence des types extrêmes", *Publications de l'Institut de Statistique de l'Université de Paris*, 23, 1 - 36.
- [20] Diebold F.X., Inoue A. (2001). Long memory and regime switching, *Journal of Econometrics*, 105, 131 - 159.
- [21] Dinh Z., Granger C.W.J. (1996). Modelling volatility persistence of speculative returns: a new approach, *Journal of Econometrics*, 73, 185 - 215.
- [22] Dias, A., Embrechts, P., (2004). Change-point analysis for dependence structures in finance and insurance, *Risk Measures for the 21st Century*, 321-335.
- [23] Engle R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of the United Kingdom, *Econometrica*, 50, 987 - 1007.
- [24] Engle R. F., Smith A.D. (1999). Stochastic permanent breaks, *Review of Economics and Statistics*, 81, 553 - 574.
- [25] Fang H.-B., Fang K.-T., Kotz S. (2002). The Meta-elliptical distributions with given marginals, *JMA*, 82, 1-16.
- [26] Fermanian, J-D., (2005). Goodness of fit tests for copulas. *Journal of multivariate analysis*, 95, 119-152.
- [27] Fermanian, J-D., Wegkamp, M., (2004). Time dependent copulas. Preprint INSEE, Paris, France.
- [28] Ferrara L. , Guégan D (2006). Fractional seasonality: Models and Applications to Economic Activity of the Euro Area, *EuroIndicators Papers Collection*, Eurostat Publications, Luxembourg.

- [29] Fofana S., Guégan D (2007). Spectral cumulant test to detect homogeneity intervals in non-stationary framework, work in progress, CES Paris 1 and ENS Cachan, France.
- [30] Ghysels E., Harvey A, Renault E. (1997). Stochastic volatility, in *Statistical methods: handbook of Statistics*, G.S. Madala, C.R. Rao (eds.), 14, 119 - 191.
- [31] Giraitis, L. and Leipus, R. (1995). A generalized fractionally differencing approach in long memory modeling, *Lithuanian Mathematical Journal*, 35, 65 - 81.
- [32] Gombay, E., (2001). *U*-statistics for change under alternatives, *Journal of Multivariate Analysis*, 78, 139-158.
- [33] Gombay, E., Horváth, L., (2002). Rates of convergence for *U*-statistic processes and their bootstrapped versions, *Journal of Statistical Planning and Inference*, 26, 247-272.
- [34] Granger, C.W.J. (1980). long memory relationships and the aggregation of dynamic models, *Journal of econometrics*, 14, 227 - 238.
- [35] Granger, C.W.J., A.P. Andersen (1978). *An introduction to bilinear time series analysis*, Vandenhoeck and Ruprecht, Gottingen.
- [36] Granger, C.W.J., R Joyeux (1980). An introduction to long memory time series models and fractional differencing, *Journal of Time Series Analysis* , 1, 15-29.
- [37] Granger, C.W.J., Terasvirta T. (1999). A simple non linear time series model with misleading linear properties, *Economic Letters*, 62, 161 - 165.
- [38] Granger, C.W.J., Hyung N. (2004). Occasional structural breaks and long memory with an application to the SandP 500 absolute stock returns, *Journal of empirical finance*, 11, 399 - 421.
- [39] Gray H., N. Zhang, W. Woodward (1989). On generalized fractional processes, *Journal of Time Series Analysis*, 10, 233-257.
- [40] Guégan D. (1994). *Séries chronologiques non linéaires à temps discret*, Economica, Série Mathématique et Probabilité, 301 pages., Paris, France.

- [41] Guégan D. (2000). A new model: the  $k$ -factor GIGARCH process, *Journal of Signal Processing*, 4, 265 - 271.
- [42] Guégan D. (2001). Long memory behavior for simulated chaotic time series, *IEICE Trans. Fundamentals*, E84-A, 2145 - 2155.
- [43] Guégan D. (2003a). *Les chaos en finance : Approche statistique*, Economica, Série Mathématique et Probabilité, 420 pages, Paris, France.
- [44] Guégan D. (2003b). A prospective study of the  $k$ -factor Gegenbauer processes with heteroscedastic errors and an application to inflation rates, *Finance India*, XVII, 1.
- [45] Guégan D. (2005). How can we define the concept of long memory ? An econometric survey, *Econometric Reviews*, 24, 2.
- [46] Guégan D., Ladoucette S. (2002). What is the best approach to measure the interdependence between different markets ? NER 95, 64p, Banque de France, Paris, France.
- [47] Guégan D. , S. Rioublanc (2003). Study of regime switching models. Do they provide long memory behavior? An empirical approach, Note MORA, 13-2003, ENS Cachan, France.
- [48] Guégan D. , S. Ladoucette (2004). Dependence modelling of the joint extremes in a portfolio using Archimedean copulas: application to MSCI indices " - Note de Recherche IDHE-MORA , 05-2004, Cachan, France.
- [49] Guégan D., S. Rioublanc (2005). Regime switching models: real or spurious long memory, Note de recherche MORA - IDHE - n° 02-2005, Février 2005, Cachan, France.
- [50] Guégan D. , C Zhang (2006). Change analysis of dynamic copula for measuring dependence in multivariate financial data, WP 2006 - 06, CES - Paris 1 and Cachan, France.
- [51] Haggan V., Ozaki T. (1981). A study of application of modelling non linear vibrations using an amplitude-dependent autoregressive time series analysis, *Biometrika*, 68, 189-196.
- [52] Hamilton, J.D. (1988). Rational expectations econometric analysis of change in regime: an investigation of the term structure of interest rates, *Journal of Economic Dynamics and Control*, 12, 385-423.

- [53] Hansen B.E. (1992). The likelihood ratio test under non-stationary conditions: testing the markov Switching model of GNP, *Journal of applied econometrics*, 7, S61-S82.
- [54] Hoshing J.R.M. (1981). Fractional differencing, *Biometrika*, 68, 165 - 176.
- [55] Hurwitch C.M., Ray B.K. (1995). Estimation of the long parameter for non-stationary and non-invertible fractionally integrated processes, *Journal of time series analysis*, 16, 17 - 41.
- [56] Hyung N., Franses P.H. (2005). Forecasting time series with long memory and level shifts, *Journal of forecasting*, 24, 1 - 16.
- [57] Joe H. (1997). *Multivariate models and dependence concepts*, Monograph on Statistics and Applied Probability, Chapman and Hall.
- [58] Joe H., Xu J.J. (1996). The estimation method of inference functions for margins for multivariate models, Depart. of Statistics University of British Columbia, Tech. Report 166.
- [59] Kimeldorf G., Sampson A.R. (1975). Uniform representations of bivariate distributions, *Comm. Stat.*, 4, 617 - 627.
- [60] Léorat G. (2000). Processus stochastiques et déterministes à paramètres dynamiques, PhD thesis, Ecole Nationale Supérieure de Télécoms, Paris.
- [61] Ling C.H. (1965). Representation of associative functions, *Publ. Math. Debrecen*, 12, 189 - 212.
- [62] Lobato I.N., Savin N.E. (1998). Real and spurious long memory properties of stock market data, *Journal of business and economic statistics*, 16, 261 - 268.
- [63] Lobato I.N., Velasco C. (2000). Long memory in stock market trading volume, *Journal of time series analysis*, 18, 410 - 427.
- [64] Mikosch T. , Starica C. (2004). Nonstationarities in financial time series, the long range dependence and the IGARCH effects, *The review of Economics and Statistics*, 86, 378 - 390.
- [65] Nelsen, R. (1999). *An Introduction to Copulas*, Springer, New York.
- [66] Nelson D.B. (1990). Stationary and persistence in the GARCH(1,1) model, *Econometric theory*, 6, 318-334.

- [67] Nicholls D.F., Quinn G.G. (1982). *Random coefficient autoregressive models, An introduction*, Lecture notes in statistics, Springer Verlag.
- [68] Oakes D. (1982). A model for association in bivariate survival data, *J.R.S.S. B*, 60, 516 - 522.
- [69] Poskitt D.S., Chung S.H. (1996). Markov Chain Models, Time Series Analysis and Extreme Value Theory, *Advances in Applied Probability*, 28, 405-425.
- [70] Robinson P.M. (1978). Statistical inference for a random coefficient autoregressive model, *Scandinavian Journal of Statistics*, 5, 163 - 168.
- [71] Sklar A. (1959). Fonctions de répartition à  $n$  dimensions et leurs marges, *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229 - 231.
- [72] Starica C., Granger C. (2005). Nonstationarities in stock returns, *The review of Economics and Statistics*, 87, 3.
- [73] Teverovsky V., Taqqu M. (1997). Testing for long range dependence in the presence of shifting means or a slowing declining trend using a variance type estimator, *Journal of time series analysis*, 18, 279 - 304.
- [74] Timmermann A. (2000). Moments of Markov switching models, *Journal of Econometrics*, 96, 75 - 111
- [75] Tong H., Lim K.S. (1980). Threshold Autoregression, Limit Cycles and Cyclical Data, *Journal of the Royal Statistical Society*, B, 1980, 42, 245 - 292.
- [76] Tong, H. (1990). *Non-linear time series: a dynamical approach*, Oxford Scientific Publications, Oxford.
- [77] Tsay R.S. (1987). Conditional heteroscedastic time series models, *Journal of the American Statistical association*, 82, 590 - 604.
- [78] Whittle P. (1951). *Hypothesis testing in time series analysis*, Hafner, New York.
- [79] Yagima Y. (1985). On estimation of long memory time series models, *Austr. J. Stat.*, 27, 303 - 320.

## 8 Annex 1: Proof of the proposition 4.1

Let  $Y = (Y_1, Y_2, \dots, Y_T)$  be a sample size  $T$  of a process  $(Y_t)_t$ . We consider  $r$  subsamples consisting of distinct ergodic stationary processes with finite second moment. Let  $p_j \in R^+$ ,  $j = 1, \dots, r$  such that  $p_1 + p_2 + \dots + p_r = 1$ . Hence  $p_j$  is the proportion of observations from the  $j$ th subsample in the whole sample. We define  $q_j = p_1 + p_2 + \dots + p_j$ ,  $j = 1, \dots, r$ . Thus the sample is written as  $Y = (Y_1^{(1)}, \dots, Y_{Tq_1}^{(1)}, Y_{Tq_1+1}^{(2)}, \dots, Y_{Tq_{r-1}+1}^{(r)}, \dots, Y_T^{(r)})$ . We define the sample autocovariances of the sequence  $(Y_t)_t$  as follows:

$$\tilde{\gamma}_Y(h) = \frac{1}{T} \sum_{t=1}^{T-h} (Y_t - \underline{Y}_T)(Y_{t+h} - \underline{Y}_T).$$

We develop the right hand side of the previous relationship

$$\tilde{\gamma}_Y(h) = \frac{1}{T} \sum_{t=1}^{T-h} Y_t Y_{t+h} - \frac{\underline{Y}_T}{T} \sum_{t=1}^{T-h} (Y_t + Y_{t+h}) + \frac{1}{T} \sum_{t=1}^{T-h} \underline{Y}_T^2.$$

Let

$$A = \frac{1}{T} \sum_{t=1}^{T-h} Y_t Y_{t+h}$$

and

$$B = -\frac{\underline{Y}_T}{T} \sum_{t=1}^{T-h} (Y_t + Y_{t+h}) + \frac{1}{T} \sum_{t=1}^{T-h} \underline{Y}_T^2.$$

Thus  $\tilde{\gamma}_Y(h) = A + B$ . First, we compute A.

$$\begin{aligned} A &= \frac{1}{T} \sum_{i=1}^r \sum_{t=Tq_{i-1}+1}^{Tq_i-h} Y_t^{(i)} Y_{t+h}^{(i)} \\ &+ \frac{1}{T} \sum_{i=1}^r \left[ \sum_{t=Tq_i-h+1}^{Tq_{i+1}-h} Y_t^{(i)} Y_{t+h}^{(i+1)} + \dots + \sum_{t=Tq_{r-1}-h+1}^{Tq_r-h} Y_t^{(i)} Y_{t+h}^{(r)} \right]. \end{aligned}$$

Now, we know that  $cov(Y_t^{(i)}, Y_t^{(j)}) = 0$  for all  $i \neq j$  by building, thus

$$A = \frac{1}{T} \sum_{i=1}^r \sum_{t=Tq_{i-1}+1}^{Tq_i-h} Y_t^{(i)} Y_{t+h}^{(i)} + O(1).$$

We develop the term of the right hand of the previous relationship. Thus we get

$$\frac{1}{T} \sum_{i=1}^r \sum_{t=Tq_{i-1}+1}^{Tq_i-h} Y_t^{(i)} Y_{t+h}^{(i)} = \sum_{i=1}^r p_i \frac{1}{T p_i} \sum_{t=Tq_{i-1}+1}^{Tq_i-h} Y_t^{(i)} Y_{t+h}^{(i)}$$

$$+ \sum_{i=1}^r p_i E[Y_t^{(i)}]^2 - \sum_{i=1}^r p_i E[Y_t^{(i)}]^2.$$

Thus

$$\begin{aligned} \frac{1}{T} \sum_{i=1}^r \sum_{t=Tq_{i-1}+1}^{Tq_i-h} Y_t^{(i)} Y_{t+h}^{(i)} &= \sum_{i=1}^r p_i E[Y_0^{(i)} Y_h^{(i)}] - \sum_{i=1}^r p_i E[Y_t^{(i)}]^2 + \sum_{i=1}^r p_i E[Y_t^{(i)}]^2 \\ &= \sum_{i=1}^r p_i \gamma_{Y^{(i)}}(h) + E[Y_t^{(i)}]^2. \end{aligned}$$

And, in probability,  $A \rightarrow \sum_{i=1}^r p_i \gamma_{Y^{(i)}}(h) + \sum_{i=1}^r p_i E[Y_t^{(i)}]^2$ .

Now we compute  $B$ . Using the same remark as before,  $B$  can be simplified and we get:

$$B = -\underline{Y}_T^2 + O(1).$$

Or

$$\begin{aligned} -\underline{Y}_T^2 &= -\left(\sum_{i=1}^r p_i E[Y_t^{(i)}]\right)^2 = -\sum_{i=1}^r \sum_{j=1}^r p_i p_j E[Y_t^{(i)}] E[Y_t^{(j)}] \\ &= -\sum_{i=1}^r (p_i E[Y_t^{(i)}])^2 - 2 \sum_{1 \leq i < j \leq r} p_i p_j E[Y_t^{(i)}] E[Y_t^{(j)}]. \end{aligned}$$

Moreover  $p_i = p_i^2 + p_i \sum_{j \neq i, j=1}^r p_j$ . Thus

$$-\underline{Y}_T^2 = -\sum_{i=1}^r p_i (E[Y_t^{(i)}])^2 + \sum_{1 \leq i < j \leq r} p_i p_j (E[Y_t^{(i)}] - E[Y_t^{(j)}])^2.$$

Then

$$B \rightarrow -\sum_{i=1}^r p_i (E[Y_t^{(i)}])^2 + \sum_{1 \leq i < j \leq r} p_i p_j (E[Y_t^{(i)}] - E[Y_t^{(j)}])^2.$$

Now, using expressions found for  $A$  and  $B$  we get:

$$\begin{aligned} A+B &= \sum_{i=1}^r p_i \gamma_{Y^{(i)}}(h) + \sum_{i=1}^r p_i E[Y_t^{(i)}]^2 - \sum_{i=1}^r p_i (E[Y_t^{(i)}])^2 + \sum_{1 \leq i < j \leq r} p_i p_j (E[Y_t^{(i)}] - E[Y_t^{(j)}])^2 \\ &= \sum_{i=1}^r p_i \gamma_{Y^{(i)}}(h) + \sum_{1 \leq i < j \leq r} p_i p_j (E[Y_t^{(i)}] - E[Y_t^{(j)}])^2. \end{aligned}$$

Hence the proposition (4.1).

## 9 Annex 2: Examples of sample ACF for model (5)

We present the behaviors of the sample ACF for some Markov switching models such that (5), using a strong Gaussian white noise  $(0,1)$  for  $(\varepsilon_t)_t$ . We observe very specific patterns for these sample ACFs, which are totally different from the classical behavior of AR or ARMA processes' sample ACFs. We give the figures for a sample size  $T = 1000$ . We provide the numbers  $n_{st}$  of switches inside the trajectory sample for the different values of the transition probabilities  $p_{ii}$ ,  $i = 1, 2$ .

---

Insert Figure 2

Behaviors of the autocorrelation functions of some simulated series issued from model (5)

---

## 10 Annex 3: Estimation of the long memory parameter for model (5)

In order to measure the existence of long memory behavior inside the simulated models (5), we adjusted on these simulated data sets, a FI(d) process defined by:

$$(1 - B)^d X_t = \varepsilon_t, \quad (18)$$

where  $B$  represents the lag operator,  $(\varepsilon_t)_t$  a strong white noise.

We proceed by Monte-Carlo experiment. For a sample size  $T = 1000$  and each transition probability  $p = p_{11} = p_{22}$ , 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 (5). For each experiment, we fit a FI(d) process defined by equations (18), and in fine, we retain the estimated parameter  $\hat{d}$  obtained by averaging all the estimated parameters. The long memory parameter  $d$  is estimated using the Whittle approach, see Yajima (1985).

On Figure 3, we exhibit the estimated parameters  $\hat{d}$  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  $7 \leq n_{st} \leq 510$ , the estimated parameter  $\hat{d}$  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  $0 < \hat{d} < 0.5$ ; for  $(\mu_1, \mu_2) = (5, -5)$ ,  $\hat{d}$  is positive, but can be greater than 1. The confidence interval are very large.

---

Insert Figure 3

Estimated memory parameters  $\hat{d}$  (solid line) and their confidence interval (dotted line) for the switching model (5).

---

- For  $(\mu_1, \mu_2) = (0.5, -0.5)$ ,  $\hat{d} = 0$  for  $p$  varying from 0.55 to 0.53, which corresponds to  $n_{st}$  varying from 460 to 473. For  $(\mu_1, \mu_2) = (5, -5)$ ,  $\hat{d} = 0$  only for  $p = 0.5$ , which corresponds to  $n_{st} = 510$ .
- For  $p < 0.5$ , that is when  $n_{st} \geq 518$ , then  $\hat{d}$  is negative. This range of values characterizes existence of anti-persistence.
- Note that the switch from long memory to anti persistence, looking at the values of  $\hat{d}$ , occurs for the same transition probabilities in both cases.

To apply a FI(d) model on the simulated switching process (5) 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 model can be classified as a short memory process, existence of spurious long memory is detected. The value  $\hat{d} = 0$  is obtained in few cases. This method reveals also that it is possible to get some  $\hat{d}$  which are greater than 0.5. This should imply that the data sets are non stationary.

## 11 Annex 4: Examples of sample ACF for model (10)

To illustrate the specific behavior for the sample ACF of the model (10), we provide on Figure 4, the trajectories and the autocorrelation functions of two

simulated series issued from this model with two different pairs of volatilities:  $(\sigma_1, \sigma_2) = (1, 4)$  and  $(\sigma_1, \sigma_2) = (1, 20)$  for  $p_{11} = p_{22} = p = 0.99$  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 the data are rare, indeed  $n_{st} = 10$  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.

---

Insert Figure 4

Trajectories and autocorrelation functions of two simulated series issued from model (10), with  $\mu = 2$  and  $p = 0.99$ , for  $T = 1000$ .

---

## 12 Annex 5: Examples of sample ACF for model (6)-(7)

Here we consider a particular case of the model (6)-(7). The Figure 5 shows many breaks on the trajectory of the model, and the corresponding sample autocorrelation function decreases very slowly as the lag  $h$  increases. Notice that under the binomial distribution the model (6)-(7) implies only a sudden changes.

---

Insert Figure 5

The trajectory and ACF of the model (6)-(7) with  $\lambda = 0.01$  and  $\alpha = 0.9$ ,  $T = 2000$ ,  $\sigma_\epsilon^2 = \sigma_\eta^2 = 1$ .

---

## 13 Annex 6: Examples of sample ACF for model (8)

We consider the model defined by the relationship (8). It is a simple form of the Threshold Auto-Regressive model introduced by Lim and Tong (1980).

We can also consider it as a special case of Markov switching model. In this latter case the Markov chain  $I\{Y_{t-1} \leq r\}$  is not exogenous for the model. We provide the trajectory and the autocorrelation function of this model on figure 6. We observe switches on the trajectory and slow decay of the sample ACF, eventhough this model is classified as short memory process for the values of parameters used here.

---

Insert Figure 6

Trajectory and ACF of the Threshold Auto-Regressive model defined by equation (8) with  $T = 2000$ ,  $\sigma_\epsilon^2 = 0.2$  and  $\mu_0 = -\mu_1 = -1$ .

---

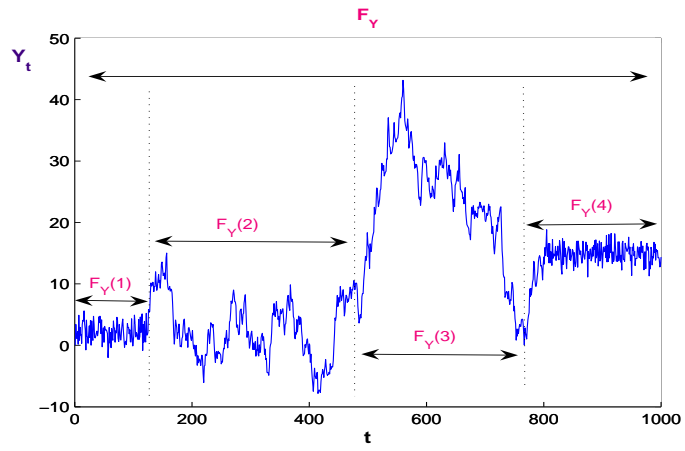
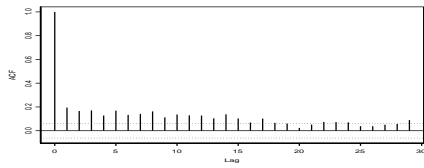
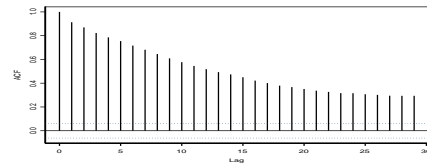


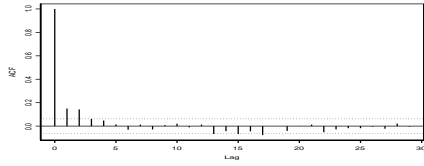
Figure 1: Example of a sequence of invariant distribution functions



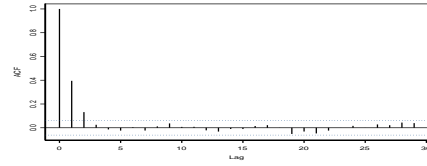
(a)  $p = 0.99, n_{st} = 7$



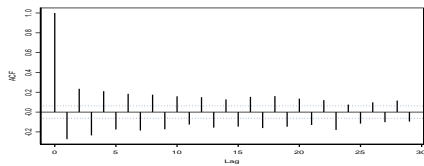
(b)  $p = 0.99, n_{st} = 7$



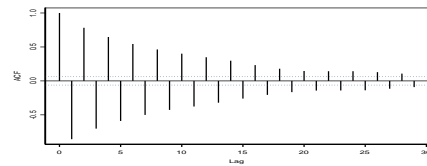
(c)  $p = 0.75, n_{st} = 258$



(d)  $p = 0.75, n_{st} = 258$

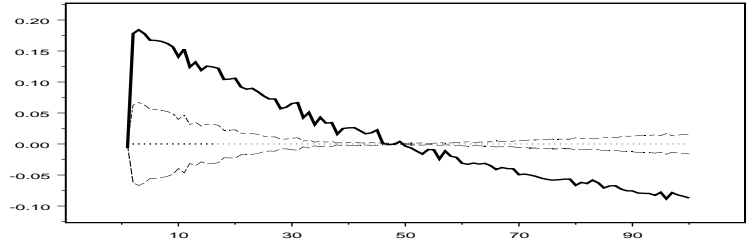


(e)  $p = 0.05, n_{st} = 950$

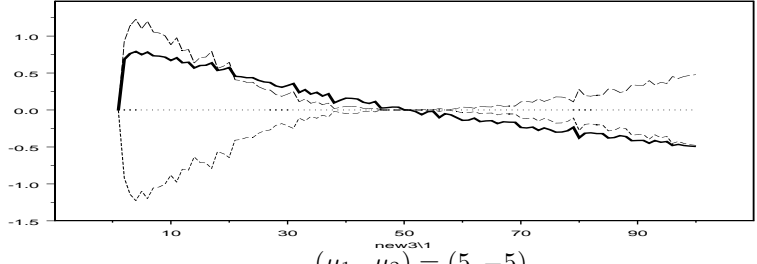


(f)  $p = 0.05, n_{st} = 950$

Figure 2: Behaviors of the autocorrelation functions of some simulated series issued from model (5) with  $p_{11} = p_{22} = p$ . Left column:  $(\mu_1, \mu_2) = (0.5, -0.5)$  and right column:  $(\mu_1, \mu_2) = (5, -5)$ . Sample size  $T = 1000$ .



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



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

Figure 3: Estimated memory parameters  $\hat{d}$  (solid line) and their confidence interval (dotted line) for the switching model (5). The abscissa represent the transition probability  $p$  from 1 to 0.01 with a step of 0.01 and  $T = 1000$ .

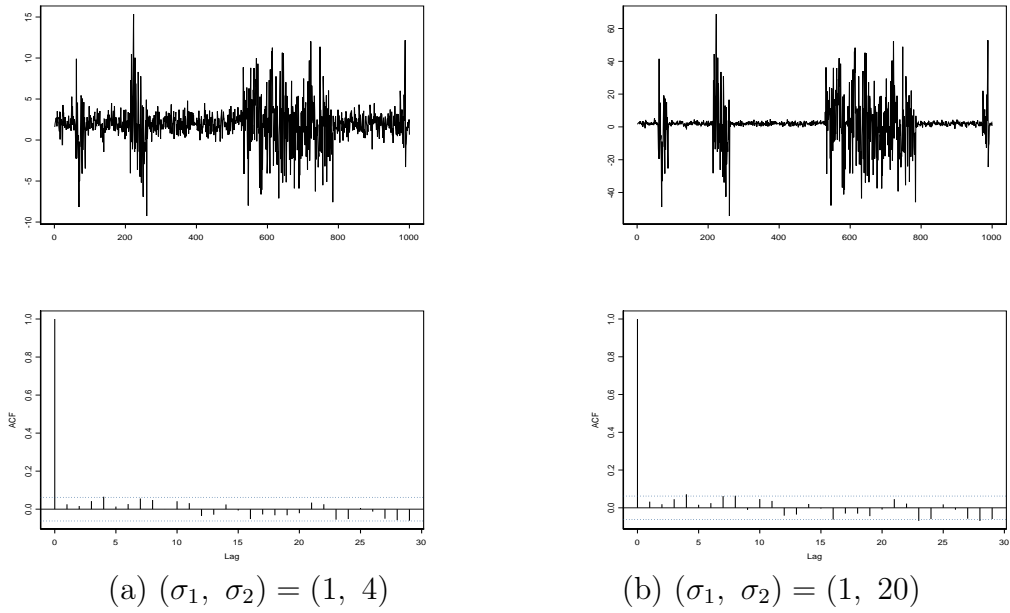


Figure 4: Trajectories and autocorrelation functions of two simulated series issued from model (10), with  $\mu = 2$  and  $p = 0.99$ , for  $T = 1000$ . Here,  $n_{st} = 10$ .

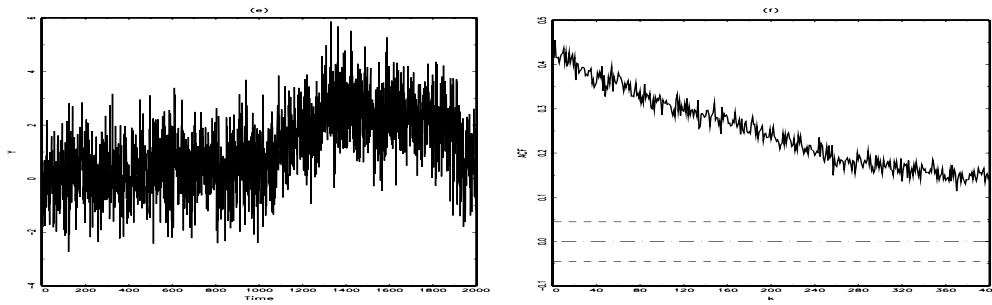


Figure 5: The trajectory and ACF of the model (6)-(7) with  $\lambda = 0.01$  and  $\alpha = 0.9$ ,  $T = 2000$ ,  $\sigma_\epsilon^2 = \sigma_\eta^2 = 1$ .

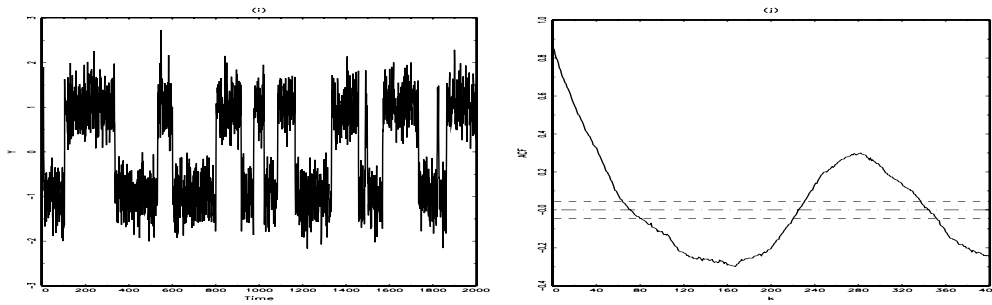


Figure 6: Trajectory and ACF of the Threshold Auto-Regressive model defined by equation (8) with  $T = 2000$ ,  $\sigma_\epsilon^2 = 0.2$  and  $\mu_0 = -\mu_1 = -1$ .