

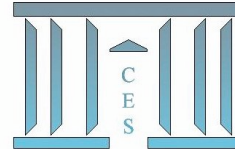


Documents de Travail du Centre d'Economie de la Sorbonne

C
E
S

W
o
r
k
i
n
g

P
a
p
e
r
s



Non-stationarity and meta-distribution

Dominique GUEGAN

2008.26



CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE

Non-stationarity and meta-distribution

Guégan D. *

March 28, 2008

Abstract

In this paper we deal with the problem of non-stationarity encountered in a lot of data sets, mainly in financial and economics domains, coming from the presence of multiple seasonalities, jumps, volatility, distortion, aggregation, etc. Existence of non-stationarity involves spurious behaviors in estimated statistics as soon as we work with finite samples. We illustrate this fact using Markov switching processes, Stopbreak models and SETAR processes. Thus, working with a theoretical framework based on the existence of an invariant measure for a whole sample is not satisfactory. Empirically alternative strategies have been developed introducing dynamics inside modelling mainly through the parameter with the use of rolling windows. A specific framework has not yet been proposed to study such non-invariant data sets. The question is difficult. Here, we address a discussion on this topic proposing the concept of meta-distribution which can be used to improve risk management strategies or forecasts.

Keywords: Non-Stationarity - Switching processes - SETAR processes - Jumps - Forecast - Risk management - Copula - Probability Distribution Function.

JEL classification: C32, C51, G12

1 Introduction

The strict stationarity assumption, or ergodicity, is the basis for a general asymptotic theory for stochastic non-linear processes and their identification,

*Paris School of Economics, CES-MSE, University Paris 1 Panthéon-Sorbonne, 106 boulevard de l'hôpital, 75013, Paris, France, e-mail : dominique.guegan@univ-paris1.fr

estimation and forecasting procedures. It guarantees that the increase of the sample size leads to more and more information which permits to an asymptotic theory to make sense. Generally speaking, stationarity is thought to be verified by the process that we study, for all $t \in R$ or $t \in Z$, or by the whole sample data set when we work empirically. Thus stationarity means "global" stationarity.

Nevertheless, it has been pointed out, from a long time, some evidence of non-stationarity inside data sets, when we observe the process over long periods of time. This problem has been discussed for instance through the problem of forecasting, Brooks (1997). Structural models have then been built to include adjustment mechanisms to improve the forecastings. All these works try to stay inside a stationary framework developing non-linear models including ARCH models, Engle (1982), GARCH models, (Bollerslev (1986), RCA models, Nicholls and Quinn (1982) and so on.

To circumvent this problem, Dalhaus (1997) used a rescaled technique to define the notion of local stationarity in order to avoid the problem of global non-stationarity. He develops, in that context, an asymptotic theory, mainly applied for linear models or close form models. Another strategy has been developed by various authors creating local approximations of non-stationary models by the way of local stationary models, using time-varying state-space representations, Subba Rao (2005) or wavelets representations, Fryzlewiczi, Van Bellegem, Von Sachz (2003). The local covariance representation for these local approximations have also been investigated, Stephan and Skander (2002).

Here our purpose is different. We develop a framework permitting to work in a stationary setting including the structural non-stationarities. We assume that a process $(X_t)_t$ is characterized by changes inside the k -order moments all along the information set (corresponding in practice to the observed trajectory). This corresponds to structural changes in financial time series causing the time series over long intervals to deviate significantly from stationarity. This means that we assume that non-global stationarity is verified for the sample. Then, it is plausible that by relaxing the assumptions of stationarity in an adequate way, we may obtain better fit and then robust forecasts and management theory for instance. Doing that, we will see that we can get new insight to approximate the unknown distribution function for complex non-stationary process.

Avoiding to use the whole sample, source of non-stationarity, we define a

new way to analyse and model this information set dividing it in subsamples on which stationarity is achieved. In this paper, our objectives are twofold. First, we show that the non-stationarities observed on the empirical moments pollute the theoretical properties of the statistics defined inside a "global" stationary framework, and thus a new framework needs to be developed. Second we propose a new way to study finite sample data sets in presence of k -order non-stationarity.

As the non-stationarity affects nearly all the moments of a time series when it is present, we first studied the impact of the non-stationarity on a non-linear transformation of the observed data set $(Y_t)_t$ considering $(Y_t^\delta)_t$, for any $\delta \in R^+$, looking at its sample autocovariance function (ACF). We exhibit the strange behavior of this ACF in presence of non stationarity and illustrate it through several modellings. Then, in order to avoid mis-specification using the sample ACF when we work with a practical point of view, we propose to work with the distribution function. In case of non-stationarity, this one appears also non-adequate, and justifies the necessity to adopt a new strategy. Thus, we focus on the building of sequence of invariant distribution functions using the notion of homogeneity intervals introduced by Starica and Granger (2005). This methodology will conduce us to propose the notion of meta-distribution associated to a sample in presence of non-stationarity. This last notion lies on both the use of copula and sequence of homogeneity intervals characterized by invariant distribution functions.

Thanks to this new approach, we will see that we can propose new insight for robust forecastings, risk management theory and solutions of complex probabilistic problems.

The plan of the paper is the following. In Section two we recall the notion of strict stationarity and we exhibit the specific behavior of the sample ACF for a $(Y_t^\delta)_t$ process in presence of non-stationarity leading to the creation of spurious behaviors that we describe through three different modellings. In Section three we specify an homogeneity test based on higher order cumulants and we show how the copula concept is usefull in presence of non-stationarity. The notion of meta-distribution is introduced. Some applications are proposed for econometricians and risk managers. Section four concludes proposing new extensions at this work.

2 Empirical evidence

A stochastic process is a sequence of random variables $(Y_t)_t$ defined on a probability space (Ω, \mathcal{A}, P) . Then, $\forall t$ fixed Y_t is a function $Y_t(\cdot)$ on the space Ω , 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 $\forall t$ and also the variance and the covariance. The covariance $\gamma(\cdot, \cdot)$ of a second order stochastic process $(Y_t)_t$ exists and is defined by

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

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 used to study this distribution function. 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. 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 strictly stationary, because we work mainly with non-linear models.

Definition 2.1 *A stochastic process $(Y_t)_t$ is strictly stationary if the joint distribution of $Y_{t_1}, Y_{t_2}, \dots, Y_{t_p}$ is identical to that of $Y_{t_1+h}, Y_{t_2+h}, \dots, Y_{t_p+h}$, for all h , where p is an arbitrary positive integer and t_1, t_2, \dots, t_p 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 $(Y_{t_1}, \dots, Y_{t_p})$ and $(Y_{t_1+h}, \dots, Y_{t_p+h})$ have the same joint distribution for all positive integers h and p , and thus all the same k -order moments. Therefore, strict stationarity requires that the distribution of $Y_{t_1}, Y_{t_2}, \dots, Y_{t_p}$ is invariant under time shifts. In that case, we speak of global stationarity.

2.1 Description of data sets

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. This latter one is generally modelled using time varying function. In presence of seasonality the covariance depends on time producing evidence of non stationarity. 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. Distorsion effects correspond to explosions that one cannot remove from any transformation. This behavior can also be viewed as a structural effect. Existence of explosions means that some higher order moments of the distribution function do not exist. Concatenated data sets used to produce specific behavior cannot have the same probability distribution function on the whole period as soon as it is a juxtaposition of several data sets. Aggregation of independent or weakly dependent random variables is a source of specific features. All these behaviors provoke the non existence of higher order moments or erratic behaviors of the sample ACF.

Until now, a lot of authors tried to take into account these non-stationarities through models. The simple one consists to take the square (or any transformation) of the data to model the conditional variance, Engle (1982), Nelson (1990), Ding and Granger (1993). Now, whatever the chosen methodology, the main tool to study the data sets remains the use of the sample autocorrelation function. Or the sample ACF computed from $(Y_t^\delta)_t$, $\delta \in R^+$, presents inappropriate behaviors under non stationarity. We exhibit below the asymptotic behavior of the sample autocorrelation function in that context and illustrate it through some modellings.

2.2 Asymptotic behavior of the ACF of the $(Y_t^\delta)_t$ process

We focus on the behavior of the sample autocovariance function of a specific data set. We assume that we observe a sample size T , $Y = (Y_1, Y_2, \dots, Y_T)$ from which we build $Y^\delta = (Y_1^\delta, Y_2^\delta, \dots, Y_T^\delta)$, $\delta \in R^+$, and we divide it in r subsamples consisting each of distinct stationary processes with finite k -order moments, $k \in N$. 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 observations 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 the reunion of r subsamples $Y^\delta = ((Y_1^\delta)^{(1)}, \dots, (Y_{Tq_1}^\delta)^{(1)}, (Y_{Tq_1+1}^\delta)^{(2)}, \dots, (Y_{Tq_{r-1}+1}^\delta)^{(r)}, \dots, (Y_T^\delta)^{(r)})$.

The sample auto-covariance function for the series $(Y_t^\delta)_t$ is equal to

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

where \underline{Y}_T^δ is the sample mean of the process $(Y_t^\delta)_t$.

Proposition 2.2 *Let be r subsamples $((Y_1^\delta)^{(i)}, \dots, (Y_{Tq_i}^\delta)^{(i)})$, $i = 1, \dots, r$ and $\delta \in R^+$, coming from the sample $(Y_t^\delta)_t$, each subsample corresponding to a strict stationary distinct process with finite k -order moments, whose sample covariance is equal to $\tilde{\gamma}_{(Y^\delta)^{(i)}}(h)$. Under these previous conditions the sample autocorrelation function $\tilde{\gamma}_{Y^\delta}$ of the sample $(Y_t^\delta)_t$ is such that:*

$$\tilde{\gamma}_{Y^\delta}(h) \rightarrow \sum_{i=1}^r p_i \tilde{\gamma}_{(Y^\delta)^{(i)}}(h) + \sum_{1 \leq i \leq j \leq r} p_i p_j [E(Y^\delta)^{(i)} - E(Y^\delta)^{(j)}]^2, \quad h \rightarrow \infty. \quad (3)$$

The proof is postponed at the end of the article in an Annex.

Under the property of strict stationary, the ACF of each process has an exponential decay. Thus, the sample $(Y_t^\delta)_t$ has its sample ACF $\tilde{\gamma}_{Y^\delta}(h)$ that decays quickly for the first lags and then approach positive constants given by $\sum_{1 \leq i \leq j \leq r} p_i p_j [E[(Y_t^\delta)^{(i)}] - E[(Y_t^\delta)^{(j)}]]^2$. Thus, in presence of non-stationarity, this last term explains the existence of persistence observed on the sample ACF when we compute it using the whole sample $(Y_t^\delta)_t$. When $\delta = 1$, this proposition permits to explain how shifts in the means could provoke a slow decay of the autocorrelation function - which can be associated to a long memory behavior - and the same behavior is observed for the variance as soon as we modelled it using Y_t^δ , for any δ . We refer to Guégan (2005) for a review on the long memory concepts. We illustrate now these previous facts using different modellings.

Let be a two states Markov switching process $(Y_t)_t$:

$$Y_t = \mu_{s_t} + \varepsilon_t. \quad (4)$$

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 μ_1 to level μ_2 with respect to the Markov chain. This model has been studied by Andel (1993). 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: 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. This effect is explained in that case by the behavior of the second term of the relationship (3) which stays always bounded. We exhibit in Figure 1 this kind of behaviors for some models (4).

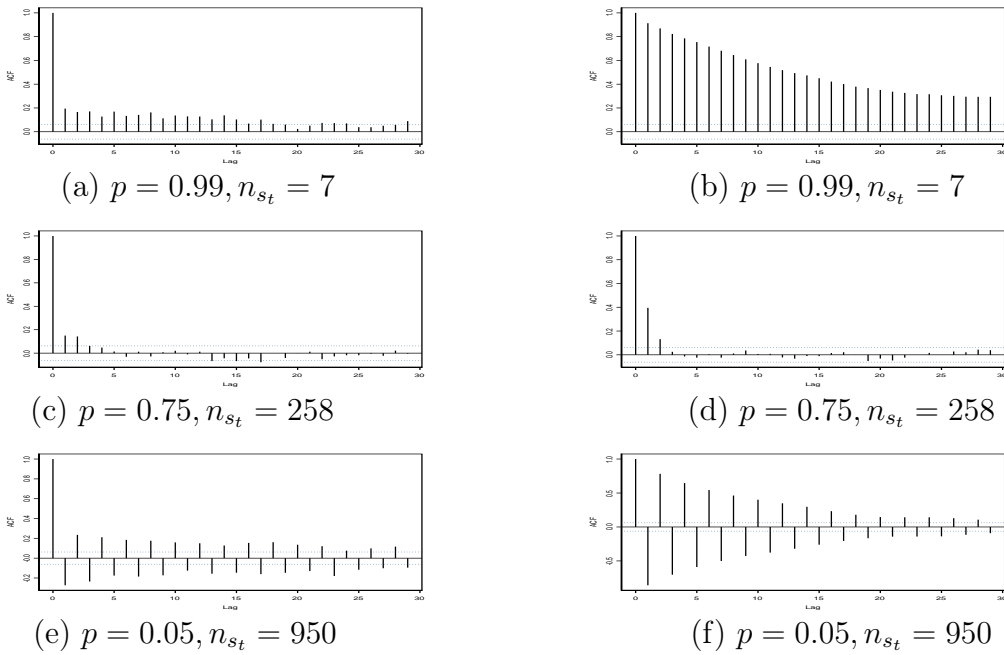


Figure 1: Behaviors of the autocorrelation functions of some simulated series issued from model (4) with $p_{11} = p_{22} = p$. Left column: $(\mu_1, \mu_2) = (0.5, -0.5)$ and right column: $(\mu_1, \mu_2) = (5, -5)$. Sample size $T = 1000$.

A StopBreak model permits also to switch from one state to another one. let be the process $(Y_t)_t$ defined by

$$Y_t = \mu_t + \varepsilon_t, \quad (5)$$

where

$$\mu_t = (1 - \alpha\delta_t)\mu_{t-1} + \delta_t\eta_t, \quad (6)$$

with $\alpha \in [0, 2]$. $(\delta_t)_t$ is a sequence of independent identically distributed Bernoulli (λ) random variables and $(\varepsilon_t)_t$ and $(\eta_t)_t$ are two independent strong white noises, Breidt and Hsu (2002). It is known that for fixed λ , 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 (3). Indeed, for different values of α and λ , the means μ_1 and μ_2 are different, thus the second term of the relationship (3) is bounded and the sample ACF of model (5)-(6) does not decrease towards zero. We provide in Figure 2 an example of this behavior.

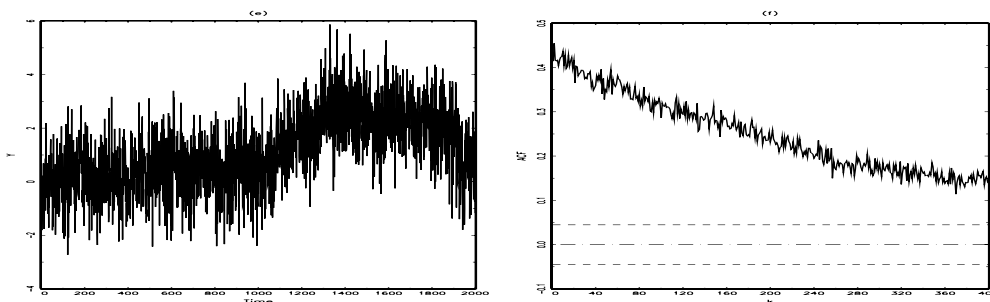


Figure 2: The trajectory and ACF of the model (5)-(6) with $\lambda = 0.01$ and $\alpha = 0.9$, $T = 2000$, $\sigma_\varepsilon^2 = \sigma_\eta^2 = 1$.

Consider now a SETAR process $(Y_t)_t$ whose a simple representation is

$$Y_t = \mu_1 I(Y_{t-1} > 0) + \mu_2 I(Y_{t-1} \leq 0) + \varepsilon_t, \quad (7)$$

where $I(\cdot)$ is the indicator function, $(\varepsilon_t)_t$ being a strong white noise. This model permits to shift from the mean μ_1 to the mean μ_2 with respect to the value taken by Y_{t-1} . SETAR processes are known to be short memory, 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 (3), and also by the time spent in each state. We exhibit in Figure 3 an experiment corresponding to this fact.

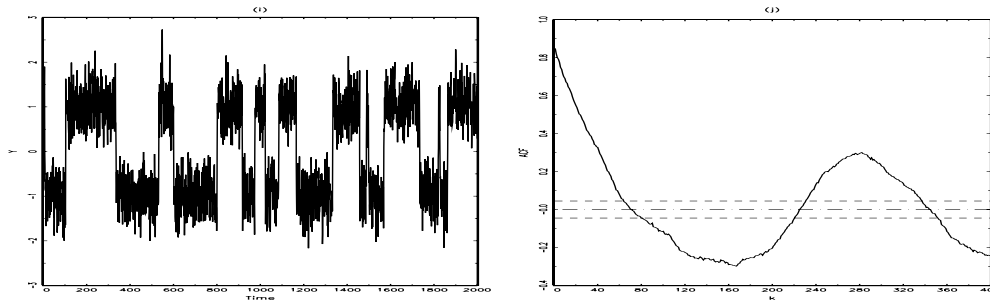


Figure 3: Trajectory and ACF of the Threshold Auto-Regressive model defined by equation (7) with $T = 2000$, $\sigma_\epsilon^2 = 0.2$ and $\mu_0 = -\mu_1 = -1$.

Thus the use of the sample ACF creates confusion in the modelling of some data sets in presence of non-stationarity. This last statistical tool appears non sufficient to characterize the behavior of any data set in that context. Thus, a new strategy needs to be developed as soon as the second order properties fail to give correct information for modelling. Moreover, it appears important to use the characteristics of the higher order moments or of the distribution function to solve this problem.

One way will be to test the invariance of the sample higher order moments and of the empirical distribution function all along the whole sample. With respect to the result of these tests, various strategies can be used. In presence of non-stationarity, we can model dynamical parameters models, consider models with a distribution function evolving in a dynamical way all along the whole sample or we can consider a sequence of stationary models. This means that we can define two strategies to study such a data set: the use of an unique distribution function with dynamic parameters or a set of several distribution functions invariant on each subsample.

3 A meta-distribution function

This Section concerns the discussion of several points of views in order to take into account non-stationarity and, to detect and model local stationarity. It is mainly a methodology discussion.

The first problem is to detect non-stationarities or to test them. In a first insight, we can consider an extension of the approach proposed by Starica and Granger (2005) using moments up to 2. Then, we will transform an univariate study in a multivariate one using the copula tool to determine the

invariant joint distribution function of the sample. In that latter case, we are interested to detect the copula which links the different invariant probability distribution functions adjusted on each subsample.

First step: detection of homogeneity intervals. In a recent 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 $(Y_{m_2+p+1}, \dots, Y_{m_2+2p})$ and define new confidence intervals from their statistic. At the end, they estimate a model on each homogeneity interval. 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. It is possible to extend this method using empirical higher order moments. using higher order moments is mainly justified by the fact that the moments up to 2 are non-stationary inside financial data sets. These higher order moments are estimated on the whole sample and on subsamples as before. Then a test based on these higher order moments or their cumulants can be built permitting to obtain intervals of homogeneity, Fofana and Guégan (2007). The spectral representation of the cumulants of order k , denoted c_k are used to build the test statistic. If we denote $f_{c_k, Y}$ the spectral density of cumulants of order k for the process $(Y_t)_t$ and $I_{c_k, Y, T}$, its estimate using a sample (Y_1, \dots, Y_T) , then we define the following statistic :

$$\tilde{T}(T, Y) = \sup_{\lambda \in [-\pi, \pi]} \left| \int_{[-\pi, \pi]^{k-1}} \left(\frac{I_{c_k, Y, T}(z)}{f_{c_k, Y}} - \frac{\tilde{c}_k}{c_k} \right) dz \right|, \quad (8)$$

where \tilde{c}_k is an estimate of c_k . The authors show that - under the null that the cumulants of order k are invariant on the subsamples - the statistic (8) 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, one can build homogeneity intervals, using moving windows. This statistic permits to use a more complete information from the data set in order to build homogeneity intervals.

Second step: the use of copula concept. When the homogeneity intervals are determined, we can associate to each subsample an invariant distribution function and compare it to the distribution function derived using the whole sample. Thus, we define a sequence of invariant distribution functions all along the sample. The principle is the following.

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 this set of information. The knowledge of this distribution function will permit to do forecasting or to propose 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 F_Y is not invariant on the whole sample.

We provide an example of such a situation on figure 4. On this figure, we have identified a sequence of homogeneity intervals characterized by changes in mean or in variance. Looking at this example (figure 4), we observe that it appears more appropriate to build, on each subsample, an adequate distribution function than to define directly a joint distribution function.

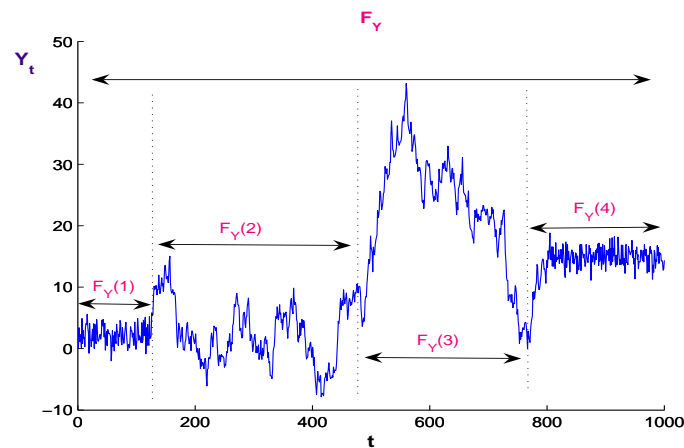


Figure 4: Example of a sequence of invariant distribution functions

In order to find F_Y , we are going to determine on each homogeneity interval an invariant distribution function. These distribution functions can belong to the same class of distribution functions but with different parameters or

can belong to different classes of distribution functions. To find them, we can use tests, as the Kolmogorov-Smirnov test, Q-Q plots or the χ^2 test.

Thus, we get a sequence of r stationary subsamples $Y_1^{(i)}, \dots, Y_{Tq_i}^{(i)}$, $i = 1, \dots, r$, each characterized by an invariant distribution function $F_{Y^{(i)}}$, $i = 1, \dots, r$. Using this sequence of invariant distribution functions, we can build the distribution function characterizing the whole sample, using copulas. We will call this probability distribution function, the meta-distribution function associated to the whole sample. In order to build it, we briefly recall the notion of copulas in a two dimensional setting and we will extend it to a r dimensional setting.

Consider a general random vector $Z = (X, Y)'$, where $'$ denotes the transpose, 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_α and we have the following relationship:

$$F(x, y) = C_\alpha(F_X(x), F_Y(y)). \quad (9)$$

Here, the copula C_α is a bivariate distribution function 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_α as the dependence structure of the vector Z .

Practically, to get the joint distribution function F of the random vector $Z = (X, Y)'$ 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 one, the archimedean one, the meta-copulas, Joe (1997) and Cherubini and Luciano (2004).

We have presented the method to adjust a copula in case of two processes. We can extend dynamically the adjustment for a sequence of r processes with invariant 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. In order to achieve this process, we can use nested tests, Guégan and Zhang (2006) for applications.

Through the nested tests, we analyze the changes. If we admit that only the copula parameters change, we can apply the change-point analysis as in Dias and Embrechts (2004) 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 and test it. More tractably, we can decide the best copula on subsamples using the moving window, and then observe the changes. We apply this method on the r homogeneity subsamples and determine an unique dynamic copula C_{α_t} linking the sequences of invariant distribution functions $F_{Y^{(i)}}$, $i = 1, \dots, r$. This copula will characterize the joint distribution function F_Y of the sample. It will provide an analytical expression of the joint distribution function. This distribution function will characterize the non-stationarity of the whole sample by the fact that it is built via a sequence of invariant distribution functions which can be all different or belonging to the same class with various parameters. This copula can be also characterized by sequence of parameters α_t evolving in time. It is in that sense that we call it a meta-distribution function. Thus, we have the following representation :

$$F(Y_t^{(1)}, \dots, Y_t^{(r)}) = C_{\alpha_t}(F(Y_t^{(1)}), \dots, F(Y_t^{(r)})), \quad (10)$$

where $Y_t^{(i)}$ represent the observations in each i subsample, $i = 1, \dots, r$, following the notation of the Subsection 2.2.

This new method permits now to propose new developments concerning several problems pointed in the introduction.

1. Concerning forecasting in presence of non-stationarity: we can use the linking copula C_{α_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 $(Y_{Tq_{r-1}+1}^{(r)}, \dots, Y_T^{(r)})$ and $F_{Y^{(r)}}$, the margin associated to this subset. If we use two homogeneity intervals, we will compute the expectation under the copula linking the two margins

corresponding to each subsample and the information set will be the reunion of the two subsamples.

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 adjust - on each asset - the corresponding margins and then we compute C_{α_t} linking these margins and associated to the portfolio. We use it to compute risk measures for the portfolio, like the VaR or the expected shortfall, Caillault and Guégan, (2005).
3. This new methodology gives also a way to find the distribution function of specific models. For instance, this approach can be used to compute the distribution function of any Markov switching model. In that latter case the joint distribution function is non known. On each state, we can estimate a margin and use the meta-distribution based on these margins as an approximation of the true distribution function. For instance, if we consider the model introduced in (4), the probability distribution function of $(Y_t)_t$ is unknown. Now, assuming that we observe (Y_1, \dots, Y_T) a sample set size T , when we are in the state 1, we can estimate the distribution function associated to this state, denoted F_1 and the same for the state 2, F_2 . Those distribution functions can also be computed theoretically if we assume known the distribution function $D(., .)$ of the noise $(\varepsilon_t)_t$. Under regular conditions, we know that a way to approach the joint distribution function is to consider a mixing of these two previous distribution functions computing it as $F(Y_1, Y_2) = \pi_{11}F_1(Y_1) + \pi_{22}F_2(Y_2)$, where π_{11} is the probability for Y_t to be in the state 1 and π_{22} , its probability to be in the state 2. Y_1 will represent the observations of the sample on the state 1 and Y_2 the observations in the state 2. We can also define the joint distribution function in the following way $F(Y_1, Y_2) = C_\alpha(F_1(Y_1), F_2(Y_2))$, where C_α will be the copula associated to the two margins F_1 and F_2 . This approach avoids independence conditions between $(\varepsilon_t)_t$ and $(s_t)_t$, the knowledge of $D(., .)$, and the estimation of the probabilities $\pi_l, l = 1, 2$, which is not an easy task.

4 Concluding remarks

In this paper, we discuss deeply the influence of presence of non-stationarity inside data sets on specific statistics, for which a lack of robustness is observed.

To detect existence of local or global stationarity on data sets, we have proposed a new test based on the empirical moments up to 2. Then we introduce the concept of meta-distribution to characterize the joint distribution function of a non-stationary sample. We observe that with this approach, some current open problems find an interesting solution. It concerns forecasting, risk management strategy and the obtention of the probability distribution function of non-stationary non-linear models. Now some extensions can also be proposed from this work.

- 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, using the covariance matrix is a problem in the sense that we cannot observe change point in it.
- 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.
- 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.

Aknowledgements We would like to thank colleagues whose suggestions during several presentations in Singapore inside the department of Mathematics, in Brisbane inside the Business School in QUT or during the ISI Platinum Jubilee Conference ICSRAR-08 in Kolkota have been determinant to improve this work. We particularly thank Professors L. Chen, R. Wolff and Dr. Piort Fryzlewiczi.

References

- [1] Andél J. (1993). A Time Series Model with Suddenly Changing Parameters, *Journal of Time Series Analysis*, 14, 111-123.
- [2] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 31, 307-327.
- [3] Breidt F.J., Hsu N.J. (2002). A class of nearly long-memory time series models, *International Journal of forecasting*, 18, 265 - 281.
- [4] Brooks C. (1997) Linear and non-linear (non)-forecastability of high frequency exchange rates, *Journal of forecasting*, 16, 125 - 145.
- [5] Cherubini, U. Luciano, E. (2004) *Copula methods in finance*, Wiley.
- [6] Caillault C., Guégan D. (2005). Empirical Estimation of Tail Dependence Using Copulas. Application to Asian Markets, *Quantitative Finance*, 5, 489 - 501.
- [7] Dalhaus R. (1997). Fitting time series models to nonstationary processes, *The annals of Statistics*, 25, 1-37.
- [8] Dias, A., Embrechts, P., (2004). Change-point analysis for dependence structures in finance and insurance, *Risk Measures for the 21st Century*, 321-335.
- [9] Engle R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of the United Kingdom, *Econometrica*, 50, 987 - 1007.
- [10] Fofana S., Guégan D (2007). Spectral cumulant test to detect homogeneity intervals in non-stationary framework, mimeo, University Paris 1, France.
- [11] Fryzlewiczi P., Van Bellegem S., Von Sachz R. (2003) Forecasting non-stationary time series by wavelet process modelling, *Ann. Inst. Statist. Math.* 55, 737 - 764.
- [12] Guégan D. (2005). How can we define the concept of long memory ? An econometric survey, *Econometric Reviews*, 24, 2.
- [13] 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.

- [14] Hyung N., Franses P.H. (2005). Forecasting time series with long memory and level shifts, *Journal of forecasting*, 24, 1 - 16.
- [15] Joe H. (1997) *Multivariate models and dependence concepts*, Monograph on Statistics and Applied Probability, Chapman and Hall.
- [16] Mikosch T. , Starica C. (2005) Nonstationarities in financial time series, the long range dependence and the IGARCH effects, *The review of Economics and Statistics*, 86, 378 - 390.
- [17] Nicholls D., Quinn B. (1982) *Random Coefficient Autoregressive Models, An introduction*, New-York, Springer Verlag.
- [18] 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.
- [19] Stephan C., Skander S. (2002) *Statistical analysis of financial time series under the assumption of local stationarity*, WP, University ParisX, Nanterre, France.
- [20] Subba Rao S. (2005) *On some nonstationary, nonlinear random processes and their stationary approximations*, technical Report, Heidelberg university, Germany.
- [21] Tong, H. (1990) *Non-linear time series: a dynamical approach*, Oxford Scientific Publications, Oxford.

5 Annex: Proof of the proposition 2.2

We develop the right hand side of the relationship (2):

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

Let

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

and

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

Thus $\tilde{\gamma}_{Y^\delta}(h) = A + B$. First, we compute A .

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

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

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

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

$$\frac{1}{T} \sum_{i=1}^r \sum_{t=Tq_{i-1}+1}^{Tq_i-h} (Y_t^\delta)^{(i)} (Y_{t+h}^\delta)^{(i)} = \sum_{i=1}^r p_i \frac{1}{T p_i} \sum_{t=Tq_{i-1}+1}^{Tq_i-h} (Y_t^\delta)^{(i)} (Y_{t+h}^\delta)^{(i)} \\ + \sum_{i=1}^r p_i E[(Y_t^\delta)^{(i)}]^2 - \sum_{i=1}^r p_i E[(Y_t^\delta)^{(i)}]^2.$$

Thus

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

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

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

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

Or

$$-\underline{Y}^{2\delta}_T = -\left(\sum_{i=1}^r p_i E[(Y^{\delta_t})^{(i)}]^2 \right) = -\sum_{i=1}^r \sum_{j=1}^r p_i p_j E[(Y_t^\delta)^{(i)}] E[(Y_t^\delta)^{(j)}]$$

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

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

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

Then

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

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

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

Hence the proposition (2.2).