UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 13-18
Statistics and Econometrics Series 17
June 2013

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

The change-point problem and segmentation of processes with conditional heteroskedasticity

Ana Laura Badagian, Regina Kaiser and Daniel Peña*

Abstract

In this paper we explore, analyse and apply the change-points detection and location procedures to conditional heteroskedastic processes. We focus on processes that have constant conditional mean, but present a dynamic behavior in the conditional variance and which can also be affected by structural changes. Thus, the goal is to explore, analyse and apply the change-point detection and estimation methods to the situation when the conditional variance of a univariate process is heteroskedastic and exhibits change-points. Based on the fact that a GARCH process can be expressed as an ARMA model in the squares of the variable, we propose to detect and locate change-points by using the Bayesian Information Criterion as an extension of its application in linear models. The proposed procedure is characterized by its computational simplicity, reducing difficulties of the change-point detection in the complex non-linear processes.

We compare this procedure with others available in the literature, which are based on cusum methods (Inclán and Tiao (1994), Kokoszka and Leipus (1999), Lee et al. (2004)), informational approach (Fukuda, 2010), minimum description length principle (Davis and Rodriguez-Yam (2008)), and the time varying spectrum (Ombao et al (2002)). We compute the empirical size and power properties by Monte Carlo simulation experiments considering several scenarios. We obtained a good size and power properties in detecting even small magnitudes of change and for low levels of persistence.

The procedures were applied to the S\&P500 log returns time series, in order to compare with the results in Andreou and Ghysels (2002) and Davis and Rodriguez-Yam (2008). Change-points detected by the proposed procedure were similar to the breaks found by the other procedures, and their location can be related with the Southeast Asia financial crisis and with other known financial events.

Keywords: heteroskedastic time series, segmentation, change-points

*Departamento de Estadística, Universidad Carlos III de Madrid, C/ Madrid 126, 28903 Getafe (Madrid), e-mail: abadagia@est-econ.uc3m.es, kaiser@est-econ.uc3m.es and dpena@est-econ.uc3m.es

1 Introduction

This paper deals with the detection, location and estimation of change-points in the unconditional variance of heteroskedastic time series. This kind of processes have a great importance in finance, but also in other fields as neurology, cardiology, seismology, meteorology and atmospheric physics. Testing for changes in the unconditional variance of a time series has received considerable attention, but most of the testing procedures assumed constant conditional variance (Inclán and Tiao (1994), Chen and Gupta (1997)). However, procedures for the change-point problem in the variance allowing a heteroskedastic behavior of the time series have been less investigated.

Suppose that $\{x_t\},\ t = 1, ..., T$ is a time series of independent random variables with zero mean[1],and conditional variance equal to σ_t^2. We assume that σ_t^2 is a function that evolves through time and can exhibit a piecewise behavior. Thus, the purpose of this paper is to explore, analyse and apply the change-point detection and estimation procedures to the situation when the conditional variance of a univariate process exhibited change-points.
The hypotheses of interest are:

$$\begin{aligned} H_0 &: \ \sigma_t^2 \ \text{ is a function with constant parameters over } \ x_1, ..., x_T \\ H_1 &: \ \sigma_t^2 \ \text{ is a function with changing parameters over } \ x_1, ..., x_T. \end{aligned}$$

Under H_0, the parameters defining the variance function are constant over time, meanwhile under H_1, there is, at least one point $t = k^* < T$, at which a change in the parameters of the variance function occurs.

The research issue is to present, evaluate and apply several statistics and procedures in order to find and locate a change point in the conditional variance of a time series process. One of these procedures, the one that we propose, is a model-based method using the Bayesian Information Criterion (BIC). The merit of this approach, comparing with other procedures based on BIC, is that it is not necessary to use non-linear estimation methods and the algorithm involved becomes more efficient.

The paper is organized as follows. Section 2 introduces the conditional variance models and their dependence properties. Section 3 explains the importance of detecting and estimating change-points

[1] The assumption of zero or constant conditional mean is made in order to focus the analysis in the variance of the process, but it can be changed for a stationary behavior in the mean, allowing serial correlation in the data.

in the conditional heteroskedastic processes. Section 4 presents a number of statistics and procedures to detect a change-point in processes with conditional heteroskedasticity, and, in Section 5, we discuss their strengths and limitations. In Section 6, we propose a procedure to detect and locate change-points by using the BIC as an extension of its application in linear models. In Section 7, we perform Monte Carlo simulation experiments to asses the behavior of seven different procedures to test and detect change-points, analysing their size and power properties, both for heteroskedastic processes with a single or multiple change-points. Section 8 presents the application of the procedures to S&P500 return index and Section 9 concludes.

2 Review of conditional heteroskedastic volatility models

Over the eighties and nineties, several models of conditional variance or volatility (as it is known among econometricians), for time series have been proposed. The common element to all these approaches is the notion that volatility can be decomposed into predictable and unpredictable components. Empirical applications of this idea have been made in financial time series, where the interest has centered on the determinants of the predictable part because the risk premium is a function of it.

To formalize this idea, we denote the conditional mean of a time series $\{x_1, ..., x_T\}$ by $\mu_t = E\left(x_t/x_{t-1}, x_{t-2}, ... = E_{t-1}\left(x_t\right)\right.$ and the conditional variance by

$$\sigma_t^2 = E\left[\left(x_t - \mu_t\right)^2 / x_{t-1}, x_{t-2}, ...\right] = E_{t-1}\left(x_t - \mu_t\right)^2.$$

Engle (1982) proposed to model σ_t^2 as

$$\begin{aligned} x_t &= \epsilon_t \sigma_t, \\ \sigma_t^2 &= \omega + \sum_{k=1}^{p} \alpha_k x_{t-k}^2, \end{aligned}$$

where ϵ_t is an iid process with zero mean and variance 1. This process is called Autoregressive Conditional Heteroscedastic of order p (ARCH(p)) model.

To simplify the exposition, consider the ARCH(1) model, where the conditional variance is $\sigma_t^2 = \omega + \alpha x_{t-1}^2$, with $\omega > 0$ and $\alpha \geq 0$ to be positive at every t. Although the conditional variance evolves through time, the unconditional or marginal variance of such a process is constant and equal to $\omega / \left(1 - \alpha\right)$. Thus, the constant ω is related to the scale or the marginal variance of the process while the parameter α models the evolutive part of the variance. When $\alpha = 0$, the variance is constant over time and the process is homoscedastic, while when $\alpha \neq 0$, σ_t^2 evolves depending on the past values of x_t: if x_t was large in a given t, the next period variance is going to be large, while if x_t was small, the next period variance is also small, resulting in a clustering of variance behavior. In this sense, α represents the persistence in the variance evolution and the weakly stationarity condition is $\alpha < 1$.

The ARCH(1) model can be written as an AR(1) in the squares of x_t:

$$x_t^2 = \sigma_t^2 + \sigma_t^2 \left(\epsilon_t^2 - 1\right) = \omega + \alpha x_{t-1}^2 + u_t, \tag{1}$$

where $u_t = \sigma_t^2 \left(\epsilon_t^2 - 1\right)$, which has zero mean and is uncorrelated but conditionally heteroskedastic.

The problem of ARCH models is that many lags are needed to adequately represent the dynamic evolution of the conditional variance. Following the idea in the Wold theorem, Bollerslev (1986) generalized ARCH models to

$$\sigma_t^2 = \omega + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 + \sum_{k=1}^{p} \alpha_k x_{t-k}^2,$$

the Generalized ARCH (GARCH(p, q)) models. The most recurrent model in financial applications is the GARCH(1,1) given by:

$$\sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha x_{t-1}^2. \tag{2}$$

The parameters have to be restricted to guarantee the positiveness of the conditional variance. In particular, $\omega > 0$, $\alpha \geq 0$ and $\beta \geq 0$. The marginal variance for the GARCH(1,1) is constant and equal to $\omega / \left(1 - \alpha - \beta\right)$. Alternatively, the GARCH(1,1) model can be written as an ARMA(1,1) model for squared residuals as follows:

$$\begin{aligned} x_t^2 &= \omega + \alpha x_{t-1}^2 + \beta \sigma_{t-1}^2 + u_t \\ &\quad \omega + \left(\alpha + \beta\right) x_{t-1}^2 - \beta \left(x_{t-1}^2 - \sigma_{t-1}^2\right) + u_t \\ &\quad \omega + \left(\alpha + \beta\right) x_{t-1}^2 - \beta \sigma_{t-1}^2 \left(\epsilon_{t-1}^2 - 1\right) + u_t \\ &\quad \omega + \left(\alpha + \beta\right) x_{t-1}^2 - \beta u_{t-1} + u_t. \end{aligned} \tag{3}$$

The sum of the parameters, $\alpha + \beta$, is related with the persistence of shocks to the volatility. The weak stationarity condition of the GARCH(1,1) model is $\alpha + \beta < 1$.

GARCH model has a very important limitation: it is very rigid to represent simultaneously series with high kurtosis and small autocorrelations of squares. Only when the persistence is very close to one, the GARCH model is able to represent both characteristics. Moreover, when the GARCH(1,1) is applied to financial returns, it is often observed that $\hat{\alpha} + \hat{\beta}$ is almost 1. For this reason, Engle and Bollerslev (1986) proposed the IGARCH(1,1) model which is given by

$$\begin{aligned} \sigma_t^2 &= \omega + \alpha x_{t-1}^2 + (1-\alpha)\, \sigma_{t-1}^2 \\ &= \omega + \sigma_{t-1}^2 + \alpha \left(x_{t-1}^2 - \sigma_{t-1}^2 \right) \end{aligned}$$

In the IGARCH model, the conditional volatility is modeled with a random walk plus drift.

Other approaches for modeling conditional variance are based on the idea that it has a predictable component that depends on past information and an unexpected noise. This type of models are called Stochastic Volatility Models (SVM), where the variance is an unobserved variable. In the simplest SVM, the log-volatility follows an AR(1) process (Andersen (1994)), where

$$\begin{aligned} x_t &= \epsilon_t \sigma_t^* \\ \log(\sigma_t^*) &= \mu + \phi \log(\sigma_{t-1}^*) + \eta_t \end{aligned}$$

with ϵ_t a strict white noise with variance 1, η has a normal distribution with zero mean and variance σ_η^2 and the parameter μ is related with the marginal variance of the process. The noise of the volatility equation, η_t, is assumed to be a Gaussian white noise with variance σ_η^2, independent of the noise of the level, ϵ_t. The Gaussianity of η_t, means that the log-volatility process has a normal distribution. In this model, the parameter ϕ measures the persistency in the conditional variance.

We focus the change-point problem in the parameters of the GARCH family models, letting this study of change-points in the SVM model for future research.

3 Motivation

GARCH(p, q) models are composed of a constant term, ω, related to the scale of the process, and a dynamic term, generated through the past values, which is driven by the parameters α and β. Thus, there are two sources of a change-point in conditional heteroskedastic processes: a) changes in the parameter related to the scale, ω and, b) changes in the parameters α and/or β.

Changes in α and β are related with changes in the persistence of the condicional variance and had been analysed in several papers, specially those related with finance, because the degree to which the conditional variance is affected by its past values is a very important economic or financial issue of daily stock returns.

Hendry (1986), Diebold (1986), Lamoureux and Lastrapes (1990) and Mikosch and Starica (2004) suggested that the persistence in volatility must be combined with the presence of change-points, since as it happens for linear processes, when modeling time varying volatility, we require that the parameters which describe the data generating process of volatility be stable over time. Parameter instability is an evidence of model miss-specification and standard econometric theory no longer applies. Furthermore, West et al. (1999) and Starica et al. (2005) showed that the presence of structural breaks could affect forecasting. Forecasting improve considerably, taking in account the change-points in the variance of the return series.

Lamoureux and Lastrapes (1990) demonstrated that breaks in the unconditional level of variance drove the estimated persistence of variance towards IGARCH. However, an IGARCH model implies an infinite unconditional variance for the time series, and in particular, assets returns does not exhibit this property. Thus, ignoring the presence of those change-points produces higher values of the estimated persistence. We illustrate this fact with the example presented in equations (4) and the Figure (1).

Let assume a stochastic GARCH(1,1) as follows:

$$x_t = \epsilon_t \sigma_t, \tag{4}$$
$$\sigma_t^2 = \begin{cases} 0.001 + 0.7\sigma_{t-1}^2 + 0.03x_{t-1}^2 & \text{if } \ t \leq 2048 \\ 0.001 + 0.8\sigma_{t-1}^2 + 0.1x_{t-1}^2 & \text{if } \ t > 2048, \end{cases}$$

where ϵ_t is N(0,1). In this example, the marginal variance of x_t increases in $t = 2048$ from $0.001/(1 - 0.03 - 0.7) = 0.0037$ to $0.001/(1 - 0.1 - 0.8) = 0.01$. If we ignore this change-point and fit a GARCH(1,1) with constant parameters, the estimated model is:

$$0.000009 + 0.981\sigma_{t-1}^2 + 0.0177x_{t-1}^2,$$

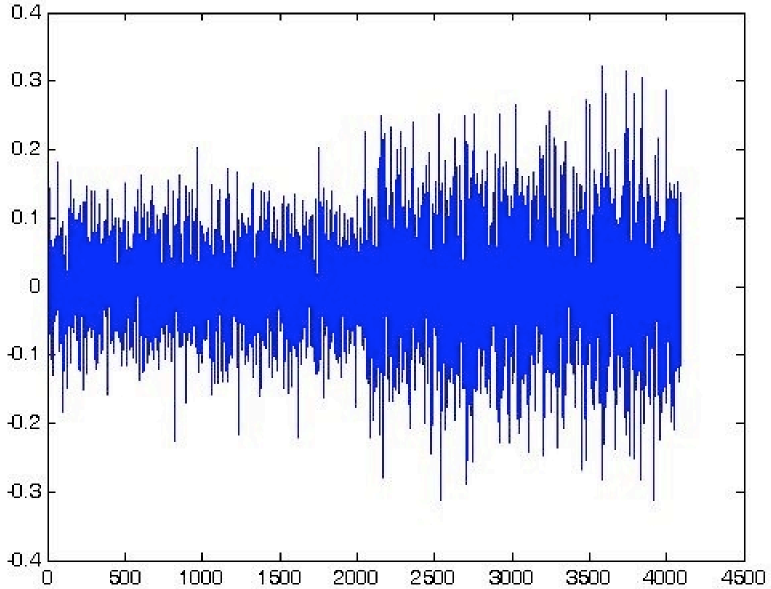

Figure 1: Simulated GARCH(1,1) defined in equations (4)

resulting in a persistence equal to $\hat{\alpha} + \hat{\beta} = 0.9987$, which is greater than the true persistence: 0.73 in $1 \leq t \leq 2048$ and 0.9 in $2048 \leq t \leq 4096$, and the marginal variance is $0.000009/(1-0.981-0.0177) = 0.00692$, which lies between the marginal variances computed for the true model.

As the example shows, it is important to detect change-points in the conditional variance. The observation of equations (1) and (3), where the presence of evolutive behavior in the variance is reflected by a linear behavior in the squares of the time series, suggest to apply the change-point tests for linear processes presented in Chapter 2 to squared transformation of the time series. Several articles demonstrated that it can be done under the fulfilment of some asymptotic properties[2] (Carrasco and Chen (2002), Fryzlewicz and Subba Rao (2011)). The models and change-points procedures presented in this Chapter are assumed to satisfy those properties.

There are many approaches based on cusums, informational criteria, minimum description length (MDL) and the spectrum to detect and locate breaks in the parameters of the conditional het-

[2] These conditions are called mixing properties. Intuitivelly, they imply that the distant future is essentially independent of the past or present of the process, and they are very important in order to apply tests for the change-point problem, because they allow to show asymptotic normality of the sums of $\{x_t\}$ and $\{x_t^2\}$ consistency of change-point detection schemes for non-linear time series. Several of those tests requiere some conditions on the dependence between the elements of a random sequence to be consistent.

eroskedastic variance. We present them in the following section.

4 Procedures for the change-point problem in conditional heteroskedastic processes

A procedure for detecting a change-point is composed of two elements:

- a statistic or loss function, useful for detect and locate a change-point, and,
- a multiple change-point searching method.

We concentrate our literature review of the procedures, in the following statistics for detecting, locating and estimating change-points in the conditional variance: cusum methods presented by Inclán and Tiao (1994), Kokoszka and Leipus (1999) and Lee et al. (2004); BIC proposed by Fukuda (2010); minimum description length (Davis et al. (2008)) and the spectrum (Ombao et al. (2002)).

In what follows, we denote as $\hat{k}$ the estimation of the true change point location k^* by applying a test statistic.

4.1 Cusum type procedures

A cusum statistic is a cumulative sum of terms (usually original data or residuals, in levels or squared) and when this sum is statistically high, it is assumed that a change-point had ocurred. When the parameter exhibiting the potential change-point is the variance, the cusum statistic is usually computed adding the squares of the data. The pioneer paper using cumulative sums of squares for the detection of changes in variance is Inclán & Tiao (1994). Their statistic was proposed for independent observations with constant conditional variance.

Let $x_1, x_2, ..., x_T$ be a sequence of independent normal random variables with parameters $\left(0, \sigma_1^2\right)$, $\left(0, \sigma_2^2\right)$, $\left(0, \sigma_T^2\right)$ respectively. The hypothesis to test is:

$$H_0 : \sigma_1^2 = ... = \sigma_T^2 = \sigma^2 \tag{5}$$

versus the alternative:

$$\sigma_1^2 = ... = \sigma_{k_1}^2 = \eta_0^2 \leq \sigma_{k_1+1}^2 = ... = \sigma_{k_2}^2 = \eta_1^2 \leq ... \leq \sigma_{k_m+1}^2 = ...\sigma_T^2 = \eta_{m+1}^2. \tag{6}$$

where m is the number of change-points and $1 < k_1 < k_2 < ... < k_m < T$ are the unknown positions of the change-points, respectively. The test statistic is defined as:

$$IT = \sqrt{T/2}\ \max_k |D_k| \tag{7}$$

where

$$D_k = \sum_{t=1}^{k} X_t / \sum_{t=1}^{T} X_t - k/T, \tag{8}$$

X_t is usually the process x_t (for testing shifts in the mean) or x_t^2 (for changes in the variance), and $0 < k < T$. The null hypothesis of no break is rejected when the maximum value of the function IT is greater than the critical value and we conclude that there is a change-point at period $k = \hat{k}$, where the maximum is achieved:

$$\hat{k} = \min\{k : IT > \text{c.v.}\}.$$

where c.v. is the corresponding critical value. The asymptotic distribution of the statistic IT is the supremum of a Brownian bridge ($B(k)$):

$$\sup\{IT(k)\} \rightarrow_{D[0,1]} \sup\{B(k) : k \in [0,1]\}$$

This establishes a Kolmogorov-Smirnov type asymptotic distribution.

Kokoszka and Leipus (1999, 2000) proposed a cusum statistic (hereinafter, KL) for which the main difference with the IT statistic, is that it was designed to analyse the existence and location of structural breaks in the conditional variance of a time series. This gives to the KL statistic the advantage of being a valid test under a wide class of strongly dependent processes, including long memory, autoregressive conditional heteroskedasticity (ARCH) and stochastic volatility (SV) type processes which have important empirical application examining financial time series.

The statistic in order to test for breaks in an ARCH(∞) process is:

$$U_T(k) = \frac{1}{\sqrt{T}} \left\{ \sum_{j=1}^{k} X_j - \frac{k}{T} \sum_{j=1}^{T} X_j \right\}$$

where $0 < k < T$. Kokoszka and Leipus (2000) considered $X_t = x_t^2$ for ARCH(∞) and $X_t = |x_t|$ for long memory processes, where x_t are the returns. The CUSUM type estimator $\hat{k}$ of a change point k^* is defined as:

$$\hat{k} = \min\{k : |U_T(k)| = \max |U_T(j)|\}$$

The asymptotic distribution of the statistic $U_T(k)$ is the same as the one of IT, this means a Kolmogorov-Smirnov type asymptotic distribution.

$$\sup\{|U_T(k)| / \hat{\sigma}\} \rightarrow_{D[0,1]} \sup\{B(k) : k \in [0,1]\}$$

where $B(k)$ is a Brownian bridge. The computation of this statistic depends on $\hat{\sigma}$, which is the estimator of square root of $\sigma^2 = \sum_{j=-\infty}^{\infty} \mathrm{Cov}(X_j, X_0)$. There are several of such estimators depending of the kernel function used. Kokoszka and Leipus (1999) suggested:

$$\hat{\sigma}^2_{T,q} = \sum_{|j|\leq q} \omega_j(q)\, \hat{\gamma}_j,$$

where $\hat{\gamma}_j$ are the sample covariances:

$$\hat{\gamma}_j = \frac{1}{T} \sum_{i=1}^{T-|j|} \left(X_i - \bar{X}\right)\left(X_{i+|j|} - \bar{X}\right) \quad |j| < T,$$

$\bar{X}$ is the sample mean $T^{-1} \sum_{j=1}^{T} X_j$ and

$$\omega_j(q) = 1 - \frac{|j|}{q+1} \quad |j| \leq q,$$

are the Bartlett weights, assuming that $q \rightarrow \infty$ and $q/T \rightarrow 0$.

Lee et al. (2004) performed a simulation study and concluded that the test for GARCH(1,1) models, using the cusum statistic based on the squares is unstable and produces low power. They considered a cusum test based on the squares of $\hat{\epsilon}_t = x_t/\hat{\sigma}$ instead of x_t, where $\hat{\sigma}$ is obtained via estimating the unknown parameters of a GARCH process. The test statistic is

$$T_T = \frac{1}{\sqrt{T}\tau} \max_{1\leq k\leq T} \left| \sum_{t=1}^{k} \hat{\epsilon}_t^2 - \left(\frac{k}{T}\right) \sum_{t=1}^{T} \hat{\epsilon}_t^2 \right|, \tag{9}$$

where $\tau^2 = Var(\hat{\epsilon}_1^2)$. Since the iid property of the true errors remains when there are no changes, this statistic is capable of detect changes in the parameters, with more stability and better powers. The parameter τ^2 is estimated as

$$\hat{\tau}^2 = \frac{1}{T-p-q} \sum_{j=p+q+1}^{T} \epsilon_j^4 - \left(\frac{1}{T-p-q} \sum_{j=p+q+1}^{T} \epsilon_j^2 \right)^2.$$

Substituting τ by $\hat{\tau}$ in the expression (9) the result is $\hat{T}_T$, which under H_0

$$\hat{T}_T \to^d \sup_{0 \leq u \leq 1} \left| B^0 \left(u \right) \right|, \quad T \to \infty,$$

where B^0 is a Brownian bridge.

4.2 Informational approach

Information criteria were used to detect changes in the marginal mean and variance by Yao (1988), but they have been less used for changes in the parameters of conditional variance. The paper of Lavielle and Moulines (2000) which was cited in many papers, proposed a very general least square test combined with a penalty function based on the BIC to avoiding oversegmentation. As a particular application, this test can be used with the squared data for detecting, locating and estimating change-point in the variance Andreou and Ghysels (2002).

We focus in the recent approach presented by Fukuda (2010), where the segmentation is based on the minimization of the BIC: the parameters of a piecewise GARCH(1,1) are estimated, jointly with the number and location of change-points.

Fukuda (2010) consider the case in which the time series x_t is divided into $m+1$ pieces generated from different GARCH(1,1) models. Thus, the i-th segment is modeled by:

$$x_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim N(0,1),$$
$$\sigma_t^2 = \omega_i + \alpha_i x_{t-1}^2 + \beta_i \sigma_{t-1}^2,$$

with $\omega_i > 0$ and $\alpha_i + \beta_i < 1$. The log likelihood (L_i) of the piece i (or the interval $[k_{i-1}, k_i]$, with $k_0 = 0$ and $k_{m+1} = T$) is given by:

$$L_i = \frac{k_i - k_{i-1}}{2 \log \left(2\pi \right)} - \left(\frac{1}{2} \right) \sum_{t=k_{i-1}+1}^{k_i} \log \left(\sigma_t^2 \right) - \left(\frac{1}{2} \right) \sum_{t=k_{i-1}+1}^{k_i} \frac{x_t^2}{\sigma_t^2}.$$

The BIC is obtained as follows:

$$BIC = -2 \sum_{i=1}^{m+1} L_i + \left\{ 3 \left(m+1 \right) + m \right\} \log T. \tag{10}$$

Moreover, it is imposed a minimum length constraint on the segments, then:

$$k_i - k_{i-1} \geq L, \quad i = 1, ..., m+1.$$

L and the maximum number of change-points is predetermined using a visual inspection of the data. The vector of parameters $(k_1, ..., k_m, \omega_1, ..., \omega_{m+1}, \alpha_1, ..., \alpha_{m+1}, \beta_1, ...\beta_{m+1})$ was obtained by fixing the maximum value of m and minimizing the BIC from the situation of $m = 0$, different locations of $m = 1$, until different locations of the $m_{\max}$ change-points.

For change-points in SVM, information criteria approach was less investigated. Berg et al. (2004) proposed a Bayesian approach based on the deviance information criterion for comparing SVM, but it has not been used for the change-point problem.

4.3 Minimum Description Length and Auto-SEG

In Chapter 2 we presented the Minimum Description Length (MDL) as a criterion to select the model that achieves the best compression of the data, in particular for piecewise autoregressive processes as in the method Auto-PARM (Davis et al. (2006)). Following the same idea, in Davis et al. (2008) MDL is used in a more general way by the Auto-SEG (for automatic segmentation) procedure, for GARCH and SVM among others.

Let m be the unknown number of change-points of the process x_t of length T. Moreover, let k_j, $j = 1, ..., m$ be the change-point between the j-th and $(j+1)$-th segments, and set $k_0 = 1$ and $k_{m+1} = T + 1$. The j-th piece of the time series $\{x_t\}$ is modeled by a stationary time series $\{x_{t,j}\}$ such that:

$$x_t = x_{t+1-k_{j-1},j} \quad k_{j-1} \leq t < k_j,$$

where the pieces are independent with stationary distribution $p_{\theta_j}(.)$, and θ_j is a member of the parameter space Θ_j with $\theta_j \neq \theta_{j+1}$, $j = 1, ..., m$. The dimension of θ_j and its parameter space Θ_j may vary with j and can be unknown.

The j-th piece of the process $\{x_t\}$ can be modelled, for example, as

- a GARCH(p_j,q_j) model; i.e

$$
\begin{aligned}
x_{t,j} &= \sigma_{t,j}\epsilon_{t,j}, \quad t = ..., -1, 0, 1, ..., \qquad (11)\\
\sigma^2_{t,j} &= \alpha_{0,j} + \alpha_{1,j}x^2_{t-1,j} + ... + \alpha_{p_j,j}x^2_{t-p_j,j}\\
&\quad +\beta_{1,j}\sigma^2_{t-1,j} + ... + \beta_{q_j,j}\sigma^2_{t-q_j,j} \quad t = ..., -1, 0, 1, ...,
\end{aligned}
$$

subject to constraints $\alpha_{0,j} > 0$, $\alpha_{i,j} \geq 0$, $\beta_{i,j} \geq 0$, $i = 1, ..., m+1$ and $\alpha_{1,j} + ... + \alpha_{p_j,j} + \beta_{1,j} + ... + \beta_{q_j,j} < 1$. With p_j and q_j unknown, then $\theta_j = (p_j, q_j, \alpha_{0,j}, \alpha_j, \beta_j)$, where α_j and β_j are the vectors of α_js and β_js in equation (12) respectively.

- a ARSV(p_j) model; i.e.

$$
\begin{aligned}
x_{t,j} &= \sigma_{t,j}\epsilon_{t,j}, \quad t = ..., -1, 0, 1, ..., \qquad (12)\\
\log(\sigma_{t^*,j}) &= \mu_{0,j} + \phi_{1,j}\log(\sigma_{t^*-1}) + ... + \phi_{p_j,j}\log(\sigma_{t^*-p_j,j}) + \eta_{t,j}, \quad t = ..., -1, 0, 1, ...
\end{aligned}
$$

The problem of finding the best segmentation is solved using the Minimum Description Length (MDL) principle of Rissanen (1989). As mentioned in the Chapter 2, using the MDL for the model selection problem, consists of select the model $\mathcal{M} \in \mathcal{F}$ that achieves the best compression of the data, where $\mathcal{F}$ is a family of candidate models. Thus, the MDL principle defines as the best model of $\mathcal{F}$, as the one that produces the shortest code length that completely describes the observed data $\mathbf{x} = (x_1, x_2, ..., x_T)$.

The code length of an object is defined as the memory space required to store that object. In the applications of MDL principle, a classical way to store $\mathbf{x}$ is to split it in two components: the adjusted model $\hat{\mathcal{M}}$ and the portion of $\mathbf{x}$ not explained by the model, the residuals, denoted by $\hat{\mathbf{e}} = \mathbf{x} - \hat{\mathbf{x}}$, where $\hat{\mathbf{x}}$ is the fitted vector for $\mathbf{x}$. If $CL_{\mathcal{M}}(z)$ denotes the code length of the object z using model $\mathcal{M}$, then is obtained the following decomposition:

$$
CL_{\mathcal{M}}(\mathbf{x}) = CL_{\mathcal{M}}\left(\hat{\mathcal{M}}\right) + CL_{\mathcal{M}}\left(\hat{\mathbf{e}}/\hat{\mathcal{M}}\right),
$$

where $CL_{\mathcal{M}}\left(\hat{\mathcal{M}}\right)$ represent the code length of the fitted model and $CL_{\mathcal{M}}\left(\hat{\mathbf{e}}/\hat{\mathcal{M}}\right)$ and is the code length of the corresponding residuals conditional on the fitted model $\hat{\mathcal{M}}$. The MDL principle suggests that the best piecewise model $\hat{\mathcal{M}}$ is the minimizer of $CL_{\mathcal{M}}(\mathbf{x})$. The authors decompose $CL_{\mathcal{M}}\left(\hat{\mathcal{M}}\right)$ in:

$$
\begin{aligned}
&CL_{\mathcal{M}}(m) + CL_{\mathcal{M}}(k_1, ..., k_m) + CL_{\mathcal{M}}(\xi_1, ..., \xi_{m+1}) + CL_{\mathcal{M}}\left(\hat{\psi}_1, ..., \hat{\psi}_{m+1}\right)\\
&= CL_{\mathcal{M}}(m) + CL_{\mathcal{M}}(T_1, ..., T_{m+1}) + CL_{\mathcal{M}}(\xi_1, ..., \xi_{m+1}) + CL_{\mathcal{M}}\left(\hat{\psi}_1, ..., \hat{\psi}_{m+1}\right).
\end{aligned}
$$

where k_j are the change-points periods, m the number of change-points, ξ_j collects the c_j integer-valued parameters (i.e. unknown orders of the model) and ψ_j contains the d_j real-valued parameters. Behind the last equation, the idea is that complete knowledge of $(k_1, ..., k_m)$ implies the complete knowledge of $(T_1, ..., T_{m+1})$ and *viceversa*. The $CL_{\mathcal{M}}(m) = \log_2 m$, $CL_{\mathcal{M}}(T_j) = \log_2 T$ for all j, and, $CL_{\mathcal{M}}\left(\hat{\psi}_j\right) = \frac{d_j}{2}\log_2 T_j$. Moreover, $CL_{\mathcal{M}}(\xi_j) = \sum_{k=1}^{c} {}_j\log_2 \xi_{kj}$, where ξ_{kj} is the kth entry of ξ_j.

Combining these results is obtained:

$$CL_{\mathcal{M}}\left(\hat{\mathcal{M}}\right) = \log_2 m + (m+1)\log_2 T + \sum_{j=1}^{m+1}\sum_{k=1}^{c_j}\log_{2kj} + \sum_{j=1}^{m+1}\frac{d_j}{2}\log_2 T_j.$$

The code length for the residuals, $CL_{\mathcal{M}}\left(\hat{\mathbf{e}}/\hat{\mathcal{M}}\right)$ as demostrated by Rissanen, is equal to the negative of the log-likelihood of the fitted model $\hat{\mathcal{M}}$, denoted as $L(\psi_j, \mathbf{x_j})$.

Thus, the formula of the MDL is given by

$$\log_2 m + (m+1)\log_2 T + \sum_{j=1}^{m+1}\sum_{k=1}^{c_j}\log_2 \xi_{kj} + \sum_{j=1}^{m+1}\frac{d_j}{2}\log_2 T_j - \sum_{j=1}^{m+1} L\left(\hat{\psi}_j, \mathbf{x_j}\right), \tag{13}$$

where the last addend is obtained from the assumption that the pieces are independent.

For instance, in the GARCH(1,1) model presented in (2), $p_j = 1$, $q_j = 1$ are the integer-valued parameters $_j$ representing the model orders and ω, β and α are the real-valued parameters ψ_j. Thus, $\theta_j = (1, 1, \omega, \beta, \alpha)$, $c_j = 2$ and $d_j = 3$. The corresponding MDL is then,

$$\log_2 m + (m+1)\log_2 n + \sum_{j=1}^{m+1}\frac{3}{2}\log n_j - \sum_{j=1}^{m+1} L\left(\hat{\psi}_j, \mathbf{x_j}\right).$$

Davis et al. (2006) showed that the best-fitted model obtained by the minimization of the MDL principle is a non trivial issue because the search space has a enormous dimension. They use a genetic algorithm to solve this problem, providing an automatic method for multiple change-point detection, location and estimation.

4.4 The spectrum of locally stationary processes and Auto-SLEX

This is a non-parametric procedure introduced by Ombao et al. (2002) for detecting change-point in the variance of a time series. Given its non-parametric nature, it does not depend on the model assumed. Thus, in principle, Auto-SLEX could be applied to data with conditional heteroskedasticity.

The basis is the Cramer representation of locally stationary processes, that generalizes the Fourier vectors which are perfectly localized in frequency, but they cannot adequately represent non stationary time series, i.e., the time series with spectra that change over time. Since Fourier vectors are perfectly localized in frequency, they are ideal at representing stationary time series. However, they cannot adequately represent non stationary time series, i.e., the time series with spectra that change over time. Ombao et al. (2002) create SLEX vectors which are simultaneously orthogonal and localized in time and frequency. They are calculated by applying a projection operator on the Fourier vectors, consisting on two specially constructed smooth windows. Then, a SLEX basis vector $\phi_{S,\omega}(t)$ for the time block $[\alpha_0, \alpha_1]$ and oscillating at frequency ω, has support on the discrete time block $S = \{\alpha_0 - \epsilon + 1, ..., \alpha_1 - \epsilon\}$ and has the form

$$\phi_{S,\omega}(t) = \Psi_{S,+}(t) \exp\left(i2\pi\omega \frac{t}{|S|}\right) + \Psi_{S,-}(t) \exp\left(-i2\pi\omega \frac{t}{|S|}\right) \tag{14}$$

where $\omega \in [-1/2, 1/2]$, $|S| = \alpha_1 - \alpha_0$, ϵ is a small overlap between two consecutive time blocks which ensures smoothness in the transition between them. In Huang et al. (2004), the windows $\Psi_{S,+}(t)$ and $\Psi_{S,-}(t)$ take the form

$$\begin{aligned} \Psi_{S,+}(t) &= r^2\left(\frac{t-\alpha_0}{\epsilon}\right) r^2\left(\frac{\alpha_1 - t}{\epsilon}\right) \\ \Psi_{S,-}(t) &= r\left(\frac{t-\alpha_0}{\epsilon}\right) r\left(\frac{\alpha_0 - t}{\epsilon}\right) - r\left(\frac{t-\alpha_1}{\epsilon}\right) r\left(\frac{\alpha_1 - t}{\epsilon}\right) \end{aligned}$$

where $r(.)$ is called a "rising cut-off function". Huang et al. (2004) use the sine rising cut-off function

$$r(u) = \sin\left(\frac{\pi}{4}(1+u)\right), \qquad \text{where } u \in [-1, 1]. \tag{15}$$

Other types of rising cut-off functions may be used (see Wickerhauser and Chui (1994) for details).

The SLEX library is a collection of bases, each having orthogonal vectors with time support, which are obtained by segmenting the time series, of length T, in a dyadic way. We explain the dyadic algorithm in section (**??**). Let $S(j, b)$ be the block b on level j and $M_j = T/2^j$ the length of the block j, with $j = 0, ..., J$ and J the finest resolution level. The SLEX transform consists of the set of coefficients corresponding to all the SLEX vectors defined in the library. The SLEX coefficients on block $S = S(j, b)$ are defined by

$$\hat{\theta}_{S,k} = \frac{1}{\sqrt{M_j}} \sum_t x_{t,T} \overline{\phi_{S,\omega_k}(t)}, \tag{16}$$

where the fundamental frequency is $\omega_k = k/M_j$ and $k = -M_j/2 + 1, ..., M_j/2$. The SLEX periodogram, an analogue of the Fourier periodogram for a stationary process, is defined to be

$$\hat{\alpha}_{S,k} = \left|\hat{\theta}_{S,k}\right|^2. \tag{17}$$

After computing the SLEX transform a well-defined cost is computed at each of the blocks. For example, the cost function of the block $S(j,b)$ could be

$$\text{Cost}(j,b) = \sum_{k=-M_j/2+1}^{M_j/2} \log \hat{\alpha}_{S,k} + \beta\sqrt{M_j}, \tag{18}$$

where β is a complexity penalty parameter. The penalty term $\beta\sqrt{M_j}$ safeguards the procedure from obtaining a segmentation that has too many, or too few, blocks. A small value of β leads to a procedure that tends to select a segmentation with too many small blocks, and this favors the existence of less bias due to the non stationarity. However, having less observations within each block leads to inflated variances of the estimates. A large value of β, on the other hand, leads to a procedure that tends to select a segmentation with very few blocks. Although variance of the estimates is reduced, having too few blocks may lead to bias due to non stationarity (i.e. error due to not splitting a non stationary block). The penalty parameter β can be either approximated or computed via a data-driven procedure. Ombao et al. (2002) set $\beta = 1$ motivated by Donoho et al. (1998).

The cost for a particular segmentation of the time series is the sum of the costs at all the blocks defining that segmentation. The Best Basis Algorithm is applied to the SLEX transform to obtain the unique orthonormal transform in the SLEX library that has the smallest cost. So, the Best Basis in the SLEX library is the segmentation having the smallest cost.

Let B_T the best basis selected from the SLEX library and $\cup S_i$ be the blocks in B_T (a particular dyadic segmentation of the time series). Define M_i to be the numbers of points on the block S_i. Let J_T to be the highest time resolution level in B_T, i.e., the smallest time block in B_T has length $T/2^{J_T}$. The frequencies defined on S_i are the grid frequencies $\omega_{k_i} = k_i/M_i$ for $k_i = -M_i/2+1, ..., M_i/2$. The spectral representation of $x_{t,T}$ is

$$x_{t,T} = \sum_{\cup S_i \sim B_T} \frac{1}{\sqrt{M_i}} \sum_{k=-M_i/2+1}^{M_i/2} \theta_{i,k,T} \phi_{i,k}(t)\, z_{i,k} \tag{19}$$

where $\theta_{i,k,T}$ is the transfer function on time block S_i and frequency k; $\phi_{i,k}$ is the SLEX basis vector oscillating at frequency k and having support at block S_i; and $z_{i,k}$ is a orthonormal random process with finite fourth moment.

5 Strenghs and limitations of the previous procedures

The procedures presented above were examined by several authors by studying the theoretical properties of the statistics and performing Monte Carlo simulation experiment for assessing their size and power properties. In general, cusum methods have the advantage of being non-parametrical o semi-parametrical methods that can be easily implemented, and do not require parameter estimation. The same issue is valid for Auto-SLEX, which is a non-parametric procedure.

The main strengh of the model-based procedures is that they consider the theoretical properties of the data generating process, taking into account the dynamic structure of the time series. The advantage of this aspect is that using parametric procedures, it is possible to determine which is the parameter shifting. Galeano and Tsay (2010) stated that depending on what is the parameter changing, the effects on the time series could be very different. They examined the consequences of a shift in the individual parameters of the GARCH(1,1) on the unconditional variance and the kurtosis. They showed that a change in ω remains constant the kurtosis, but produces a permanent change in the volatility level. A change in α or in β produces a permanent change in both the volatility level and the excess kurtosis, such that the variance level increases if α and/or β increases, and it decreases otherwise. However, if the innovations ϵ_t are normally distributed, a change in ω has a larger influence in the excess kurtosis than a change in β, if both have the same size.

In what follows we present some findings presented in the literature about the procedures mentioned above.

With respect to IT statistic, it was designed by Inclán and Tiao (1994) for iid data with zero mean. When the analysed data is serial correlated, its power porperties are severely affected and the statistic does not perform well when the process is not iid.

IT is affected by the presence of outliers. Thus, Inclán and Tiao (1994) suggested to complement the test for variance changes with a procedure for outlier detection. For instance, looking at the plots of D_k, because a big outlier would create a significant peak that might not be due to a variance change. In most cases it is easy to detect outliers affecting the D_k plot, because they will appear as sudden jumps; the slope of the D_k would not be changed.

As showed by Inclán and Tiao (1994), IT puts more weight near the middle of the series. They demonstrated that the estimator $\hat{k}^*$, which is the point where the maximum is achieved, is skewed distributed and biased to the middle of the time series. The skewness depends both of k^* and the variance ratio of H_0 with respect H_1. What makes the statistics work well is that the mode of $\hat{k}^*$ is exactly at the point where the change in variance occurs. The values of $\hat{k}^*$ become increasingly concentrated around the true change-point as the sample size increases or as the variance ratio increases. Another implication of the skewness in the distribution of $\hat{k}^*$ is that if the smaller variance correspond to the shorter segment of the series, then it will be harder to find the change-point using the statistic proposed.

Aggarwal et al. (1999), Malik (2003), Malik and Hassan (2004), Morana and Beltratti (2004), Nouira et al. (2004), Hyung et al. (2009) among others used the IT statistic to detect change-points in time series of financial returns. Kim et al. (2000) considered the application of IT statistic to GARCH(1,1) processes taking in account of the fact that the unconditional variance is a functional of GARCH parameters, and their change can be detected by examining the existence of a change in the unconditional variance. They modify the IT test, allowing GARCH errors and concluded that their test performs appropriately in GARCH models under some limited conditions. Andreou and Ghysels (2002) showed via Monte Carlo simulation that IT test has power and size distorsions when applied to dependent data, particularly GARCH models, though it is not as powerful as other tests like KL.

The most important virtue of KL cusum test is that it was designed to detect and locate changes in the unconditional variance, when the time series is heteroskedastic. Kokoszka and Leipus (2000) prove its consistency. Sansó et al. (2004) studied the KL statistic performance with conditional heteroskedastic processes and suggested that it is a robust method, valid for detecting structural breaks under fairly general conditions. The properties of KL statistic were also analysed by Andreou and Ghysels (2002) and Pooter and Dijk (2004), finding that the test has good power but it can suffer of severe size distorsions. In the paper of Pooter and Dijk (2004), KL is applied to examine changes in the unconditional variance of a set of emerging stock market returns.

Unlike the other cusum procedures, the LEE cusum statistic needs the estimation of the parameters of the GARCH model. Lee et al. (2003) argumented that in linear processes a change in the variance of the observations, imply a change in the errors. Thus, a test for a variance change can be performed based on the errors rather than the observations themselves. Furthermore, the test based on the errors performs better than the one based on the observations since the latter is subject to serious power losses when the data is highly correlated. Other authors were interested in cusums of squares of the residuals of a GARCH model. For instance, Kulperger and Yu (2005) constructed high moment partial sum processes of residuals (from squares until fourth moment) in a GARCH model and provide interesting diagnostic tools.

Fukuda (2010) approach using the BIC to detect, locate and estimate change-points for GARCH models was examined and compared with KL statistic and other information criterion by performing some simulation experiments. The test size resulted very small (0.000) and the frequency count of correctly selecting one-change model was not high, particularly when the magnitude of the change in the parameters is small. Another important aspect that can be noted from the Table 1 in **??** is that the success of the procedure seems to be sensitive to the selection of the parameter L, or to the relationship beetween L and the sample size T. For instance, for a GARCH(1,1) model with $\alpha = 0.1$ and $\beta = 0.8$, where in the first half of the time series, $\omega = 0.1$ changing in the second half to $\omega = 0.2$, the percentage of detected breaks changed from 14.8% to 63%, when T and L changed from 1000 to 2000 and from 300 to 800, respectively. The method is applicated to financial Japanese data. Compared with KL statistic, the BIC computed by estimating a GARCH model obtained a lower power. The merit of the procedure, compared with cusum methods, is that the number and location of the change-points are determined based on a piecewise GARCH model and the estimates in each segment are jointly estimated.

The same applies for Auto-SEG, since it is a model-based procedure, the detection and location of the change-points and the estimation of the parameters of each piece are jointly obtained. In Davis et al. (2008) Auto-SEG was applied to analyse change-points in the SP 500.

Finally, Auto-SLEX was designed as a segmentation method to detect and locate change-points in the unconditional variance. The main problem with this non-parametric procedure is that the segmentation is performed in a dyadic way, making it difficult to properly locate changes away from the cutoffs.

6 ARMA models and BIC for detecting and locating change-points in the conditional heteroskedastic processes

Given the good performance of the procedure based on the BIC found in the previous Chapter, we propose an alternative to Fukuda (2010), for detecting changes in the parameters of the conditional heteroskedastic processes.

From equations (1) and (3) a GARCH process can be represented as an ARMA(p,q) in the squares of x_t. Recall that an ARCH(1) model can be expressed as an AR(1) in squares, such that

$$x_t^2 = \omega + \alpha x_{t-1}^2 + u_t,$$

where $u_t = x_t^2 - \sigma^2$. Analogously, a GARCH(1,1) model, can be expressed as an ARMA(1,1), where

$$x_t^2 = \omega + (\alpha + \beta)\, x_{t-1}^2 - \beta u_{t-1} + u_t.$$

Then, we propose to detect a single change-point as follows:

- Estimating an ARMA(p,q) process for the squares, x_t^2, $t = 1, ..., T$ and compute the BIC using the formula (**??**), denoted as BIC_0. In that formula, $\hat{\sigma}^2$ is the conditional maximum likelihood estimator of the variance assuming an ARMA(p,q) model for x_t^2.

  For instance, an ARCH(1) process can be represented by an AR(1) in the squares of x_t. Then, the BIC under the hypothesis of no change in the conditional variance can be assessed with

  $$\mathrm{BIC}_0 = (T-1)\log\hat{\sigma}^2 + 3\log(T-1)$$

  where $\hat{\sigma}^2 = \frac{1}{T-1}\sum_{t=2}^{T}\left(x_t^2 - \hat{\omega} - \hat{\alpha}x_{t-1}^2\right)^2$, $\hat{\omega}$, $\hat{\alpha}$ are the conditional maximum likelihood estimators of σ^2, ω, α.

  For the GARCH(1,1) model, one more parameter is estimated. Then,

  $$\mathrm{BIC}_0 = (T-1)\log\hat{\sigma}^2 + 4\log(T-1)$$

where $\hat{\sigma}^2 = \frac{1}{T-1}\sum_{t=2}^{T}\left(x_t^2 - \hat{\omega} - \left(\hat{\alpha} + \hat{\beta}\right)x_{t-1}^2 - \hat{\beta}\hat{u}_{t-1}\right)^2$, $\hat{\omega}$, $\hat{\alpha}$, $\hat{\beta}$ are the conditional maximum likelihood estimators of σ^2, ω, α, β, respectively, $\hat{u}_t = x_t^2 - \hat{\sigma}_t^2$ and $\hat{u}_0 = 0$.

- Estimating a piecewise ARMA(p,q) for the pieces $x^2\left(1:k\right)$ and $x^2\left(k+1:T\right)$, where $k = 1,...,T$. We denote with $x^2\left(i:j\right)$ the vector of squares of x_t, from the observation i to the observation j. Compute the BIC corresponding to each segmentation by using the equation (6) and take the minimum, denoted as BIC_1. In that formula, $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are the conditional maximum likelihood estimators of the variance before and after the change-point, assuming a piecewise ARMA(p,q) model for x_t^2. Thus, for the ARCH(1) process with a single change-point in $k = 1,...,T$,

$$\text{BIC}_1 = (k-1)\text{log}\hat{\sigma}_1^2 + (T-k)\text{log}\hat{\sigma}_2^2 + 6\text{log}\left(T-1\right) \tag{20}$$

where $\hat{\sigma}_1^2 = \frac{1}{k-1}\sum_{i=2}^{k}(x_t^2 - \hat{\omega}_1 - \hat{\alpha}_1 x_{t-1}^2)^2$ and $\hat{\sigma}_2^2 = \frac{1}{T-k}\sum_{i=k+1}^{T}(x_t^2 - \hat{\omega}_2 - \hat{\alpha}_2 x_{t-1}^2)^2$, $\hat{\omega_1}$, $\hat{\alpha_1}$, $\hat{\omega_2}$ and $\hat{\alpha_2}$ are the conditional maximum likelihood estimators of σ_1^2, σ_2^2, ω_1, α_1, ω_2 and α_2, with α_i, ω_i, $i = 1,2$, the parameters before and after the change-point, respectively.

For the GARCH(1,1) model,

$$\text{BIC}_1 = (k-1)\text{log}\hat{\sigma}_1^2 + (T-k)\text{log}\hat{\sigma}_2^2 + 8\text{log}\left(T-1\right)$$

where $\hat{\sigma}_1^2 = \frac{1}{k-1}\sum_{i=2}^{k}(x_t^2 - \hat{\omega}_1 - \left(\hat{\alpha}_1 + \hat{\beta}_1\right)x_{t-1}^2 - \hat{\beta}_1\hat{u}_{t-1})^2$ and $\hat{\sigma}_2^2 = \frac{1}{T-k}\sum_{i=k+1}^{T}(x_t^2 - \hat{\omega}_2 - \left(\hat{\alpha}_2 + \hat{\beta}_2\right)x_{t-1}^2 - \hat{\beta}_1\hat{u}_{t-1})^2$, $\hat{\omega_1}$, $\hat{\alpha_1}$, $\hat{\beta_1}$, $\hat{\omega_2}$, $\hat{\alpha_2}$ and $\hat{\beta_2}$ are the conditional maximum likelihood estimators of σ_1^2, σ_2^2, ω_1, α_1, β_1, ω_2, α_2 and β_2, with ω_i, α_i, β_i, $i = 1,2$, the parameters before and after the change-point, respectively, $\hat{u}_t = x_t^2 - \hat{\sigma}_t^2$ and $\hat{u}_0 = 0$.

- If $\text{BIC}_0 \leq \text{BIC}_1$, there is not a change-point, else there is a change-point in $\hat{k} = \arg\min \text{BIC}_1$.

If there are multiple change-points, by sequentially repeating this procedure using binary segmentation, multiple change-points can be detected.

The merit of this approach, comparing with that in Fukuda (2010) is that, it arises as an extension of the change-point problem in piecewise linear autocorrelated processes to the case of conditional heteroskedastic processes. Thus, it is not necessary to use non-linear estimation methods and the algorithm involved becomes more efficient. Since in previous studies, Fukuda (2010) approach resulted with very small size and not enough power compared with other procedures, we proposed this informational approach to compare with the other methods presented above.

7 Monte Carlo simulation experiments

In this section we report and discuss results from a set of Monte Carlo simulations experiments, designed to examine and compare the procedures presented above. We will analyse the performance of IT, KL, LEE, BIC from an ARMA(p, q) applied to x_t^2 (in what follows, BICx2), BIC obtained from the GARCH model (hereinafter, BICgarch, with fixing $L = 300$), Auto-SEG and Auto-SLEX by computing the size and the power for simulated data. For multiple change-point detection by KL, LEE and BIC statistics are combined with binary segmentation.

Since, in the following section we are going to apply these procedures to a financial dataset, and the GARCH(1,1) was the most recurrent model in financial applications, we consider this model in order to perform the simulation experiments. For comparing with Auto-SEG, the same GARCH(1,1) models included in Davis et al. (2008) have been used. Simulations are performed with 500 replicates with $T = 1000$ (1024 for Auto-SLEX). We denote the GARCH parameters as ω_i, α_i and β_i, $i = 1, 2$, where the subscript i refers to the corresponding piece of the process. Simulated processes are presented in Table (1).

Table 1: Piecewise GARCH(1,1) simulated processes

	$(\omega_1, \alpha_1, \beta_1)$	$(\omega_2, \alpha_2, \beta_2)$	Marginal variance 1^{st} piece	Marginal variance 2^{nd} piece
1	$(0.4, 0.1, 0.5)$	$(0.4, 0.1, 0.5)$	1.000	1.000
2	$(0.1, 0.1, 0.8)$	$(0.1, 0.1, 0.8)$	1.000	1.000
3	$(0.4, 0.1, 0.5)$	$(0.4, 0.1, 0.6)$	1.000	1.333
4	$(0.4, 0.1, 0.5)$	$(0.4, 0.1, 0.8)$	1.000	4.000
5	$(0.1, 0.1, 0.8)$	$(0.1, 0.1, 0.7)$	1.000	0.500
6	$(0.1, 0.1, 0.8)$	$(0.1, 0.1, 0.4)$	1.000	0.200
7	$(0.4, 0.1, 0.5)$	$(0.5, 0.1, 0.5)$	1.000	1.250
8	$(0.4, 0.1, 0.5)$	$(0.8, 0.1, 0.5)$	1.000	2.000
9	$(0.1, 0.1, 0.8)$	$(0.3, 0.1, 0.8)$	1.000	3.000
10	$(0.1, 0.1, 0.8)$	$(0.5, 0.1, 0.8)$	1.000	5.000

Note that in the cases 1 and 2 the GARCH parameters do not change. In the cases 3, 4, 5 and 6 the persistence of the variance is changing, in the first two cases the persistence is low/moderate and the marginal variance exhibits an increasing, and in the second two cases, the persistence in the first piece is high and the unconditional variance decreases. Finally, in the cases 7, 8, 9 and 10 the constant term in the GARCH is increasing; in both the first two cases with a low/moderate persistence and in the second two with a high persistence. We also analyse the sensitiveness to the magnitude of the break in the cases 3 to 10.

Table 2 presents the proportion of simulation runs for which the correct number of change-points (zero for models 1 and 2; one for the rest) has been detected, for the seven procedures. The Auto-SEG values were taken from Table I of Davis et al. (2008) and are also based on 500 replicates.

As a general feature, the detection rate is influenced by the size of the change in the unconditional variance. The larger is the magnitude of change, the higher is the detection rate. Besides the processes in 1 and 2, this conclusion can be noted by comparing the "even" processes which the change in the marginal variance is higher than the change in the "odd" ones.

Except for BICgarch, which resulted undersized and IT, KL and Auto-SLEX, with a high size, the sizes of the different procedures were appropriate. The undersize exhibited for BICgarch is coherent with the findings in Fukuda (2010) where the frequency count of incorrectly selecting one change model was also 0.000. In other hand, the size distorsions of IT and KL was also obtained in the simulations performed by Andreou and Ghysels (2002) for both statistics, and by Fukuda (2010), for the second one, where the critical values for 95% percentile were on average higher than the asymptotic critical value of 1.36, obtained by the supremum of the Brownian Bridge.

Both in the cases 3 and 5, the persistence is changing in a small magnitude, but in case 3 the procedures exhibited a lower power detecting one change-point than in case 5. This result can have both two explanation: a) on one side, the unconditional variance is changing more in the case 5, which increases the power with respect to the processes in the case 3, and, B) on the other side, in the case 5, the persistence given by $\alpha + \beta$ is closer to 1, which call an interesting issue referred byAndreou and Ghysels (2002), who showed that closer to the boundaries (i.e. $\alpha + \beta \approx 1$) the power of the procedures seems to improve. The explanation that we found is that, when the persistence is close to 1 and exhibits a change-point, the dynamic of a GARCH(1,1) process varies more and is easier to detect the change with respect a GARCH(1,1) process with less persistence and a chang-point of the same magnitude. Finally, for the case 3, where detecting the break seems to be difficult for all the procedures, the proposed BICx2 procedure exhibited the highest power (0.728), and only, LEE statistic obtained a nearby proportion of detection (0.694).

When the magnitude of the change in the persistence of the process is higher (cases 4 and 6), all the procedures improve the power, and the same as before, near to the boundaries, the break was more frequently detected (case 6).

For the breaks in the constant of the GARCH(1,1) (cases 7, 8, 9 and 10) results were similar. For the case 7, with a small persistence and a small magnitude of change, procedures obtained a small rate of detection. The proposed procedure, BICx2, obtained the highest power (0.670), followed by LEE and KL with a similar proportion of detection (0.594 and 0.584, respectively).

In cases 8, 9 and 10, IT procedure pretty improved the power, showing that is a more appropriate test for detecting change-points in the constant of the conditional heteroskedastic processes. Auto-SEG was the procedure exhibiting the highest power for this cases, but the proportion of one break detected by BICx2 exceeded 0.800.

Before applying the statistics and procedures to real datasets is convenient to evaluate their performance for detecting and locating multiple change-point. For this task, except for IT and Auto-SLEX which have their own multiple points searching algorithm, we combine the other statistics with binary segmentation[3]. To illustrate how the procedure work with multiple changes we simulate 1000 replications of the following process:

1. GARCH(1,1) process with $\omega_1 = 0.1$, $\omega_2 = \omega_3 = 0.5$, $\alpha_1 = \alpha_2 = \alpha_3 = 0.03$, $\beta_1 = \beta_2 = 0.8$ and $\beta_3 = 0.9$,

where ω_i, α_i and β_i, $i = 1, 2, 3$, denote the parameters in each piece of the process. The changes are located in $k_1^* = 340$ and $k_2^* = 680$ and the length of the time series is 1024. The first change-point

[3]Binary segmentation (Scott and Knott (1974), Sen and Srivastava (1975), Vostrikova (1981)) addresses the issue of multiple change-points detection as an extension of the single change-point problem. The segmentation procedure sequentially or iteratively applies the single change-point detection procedure, i.e. it applies the test to the total sample of observations, and if a break is detected, the sample is then segmented into two sub-samples and the test is reapplied. This procedure continues until no further change-points are found. This simple method can consistently estimate the number of breaks (e.g. Bai (1997), Inclán and Tiao (1994)) and is computationally efficient, resulting in an $O(n \log n)$ calculation (Killick et al. (2011). In practice, binary segmentation become less accurate with either small changes or changes that are very close on time.

is due to the parameter related to the scale, ω, and the second is induced by a change in the persistence.

As for the single change-point examples, we clasify the results of this experiment by the proportion of breaks detected, which are presented in the Table (3).

The main conclusions of Table (3) are:

- All the procedures have the ability of detecting, at least, one change-point. Only in few cases IT and less Auto-SEG did not detected a break.
- Except for LEE and Auto-SLEX, the procedures exhibit a big rate of detecting only one change.
- Auto-SEG, BICx2 and LEE obtained a similar and the higher proportion of detecting two breaks, around 70%.
- The procedures detecting less times two breaks are Auto-SLEX and IT.
- Auto-SLEX, LEE and, in lesser extent KL, detected an important proportion of processes with more than two changes.

To complement this information, in Figure (2) the histograms of the locations detected are showed, in order to examine the shape of the sampling distribution of the change-points estimators, and a bar graph of the total number of breaks detected by each procedure.

Except for Auto-SLEX, the sampling distributions of the estimators are bimodal around the true change-points. A general feature of the plots is that the spread of the estimators seems to be higher for the second break than for the first one. By observing the histograms, we realized that the high rate of detecting only one change-point of IT, KL, BICx2, BICgarch and Auto-SEG presented in Table (3), has to do with the detection of the first true break. According to the simulations for a single change-point, this can be explained mainly by two factors: first, it seems that a break in the parameter ω is, in general, detected with more success than a change in the other parameters; second, the magnitude of change introduced in β is smaller than the shift in ω, which makes harder to find the second change.

While LEE, BICx2 and Auto-SEG procedures detected a similar proportion of two breaks, their performance is very different. The histogram reflects the bimodal sampling distribution of the estimators, but the latter with a smaller dispersion than the other.

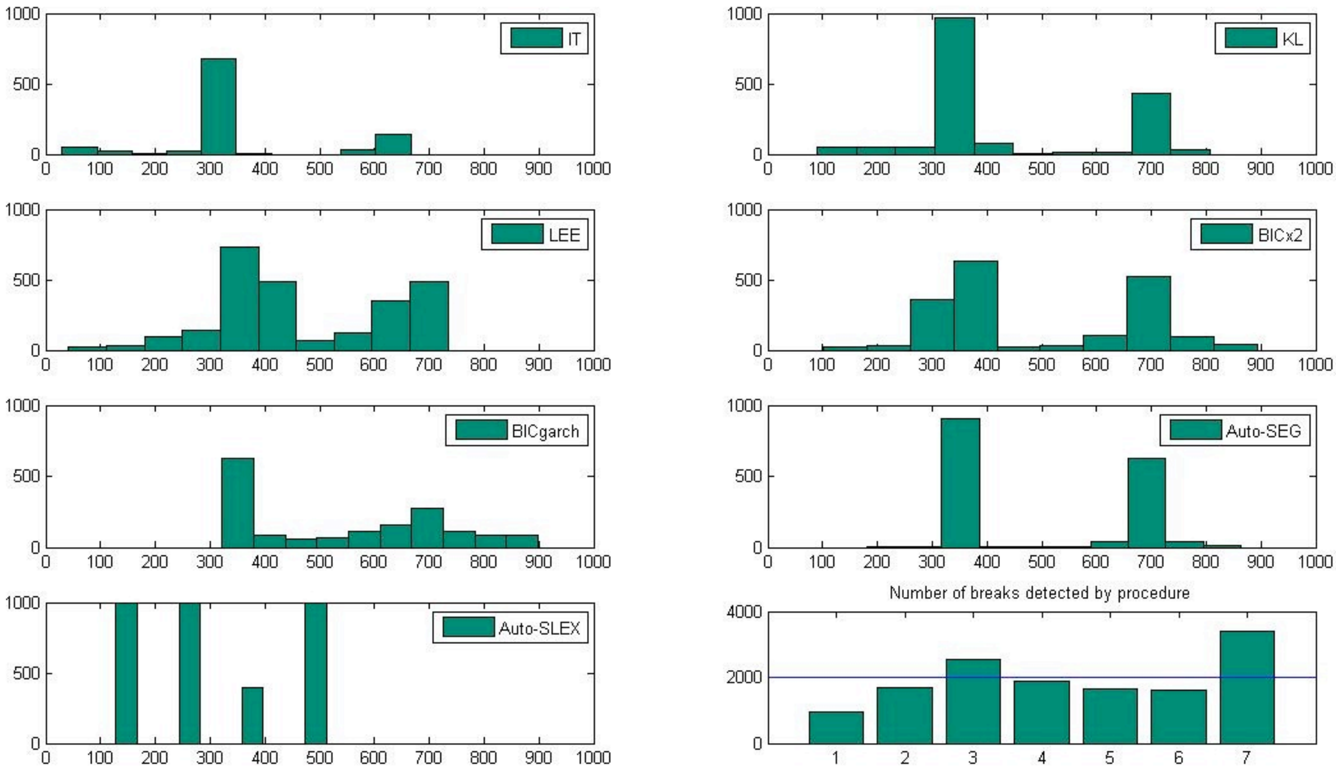

Figure 2: Sampling distributions of $\hat{k}_1^*, \hat{k}_2^*$, based on 1000 replications of the GARCH(1,1) with $\omega_1 = 1$, $\omega_2 = \omega_3 = 1.5$, $\alpha_1 = \alpha_2 = \alpha_3 = 0.03$, $\beta_1 = \beta_2 = 0.8$ and $\beta_3 = 0.9$ and the total number of breaks detected by each procedure (1: IT, 2: KL, 3: LEE, 4: BICx2, 5:BICgarch, 6: Auto-SEG, 7: Auto-SLEX).

By watching the histogram of Auto-SLEX, we can conclude that the dyadic segmentation is not able of detecting, with a good performance, multiple changes, when the breaks do not coincide with the dyadic boundaries.

Finally, in the last bar plot, the total number of breaks detected by each procedure is presented. The horizontal blue line marks the total true number of breaks, which is 2000. Bars exceeding this line indicate oversegmentation. LEE and Auto-SLEX appear as the procedures detecting extra breaks, a feature that we noticed in Table (3). While for Auto-SLEX the segmentation had a bad performance, for LEE, both the two modes of the sampling distribution remained close to the true breaks. The other procedures resulted in less than 2000 breaks, given that many of the replications were segmented in only two pieces.

8 Application to real dataset: changes in the conditional volatility of the S&P 500 index

In this section we study the existence of change-points in the conditional volatility of the S&P 500 daily log returns from 5th January 1989 to 19th October 2001 ($T = 3230$) by applying the procedures compared in the previous section. Data is presented in the Figure 3. This stock market time series was also analysed by Andreou and Ghysels (2002) and Davis et al. (2008), where the goal was to study the impact of Asian and Russian Financial crises begining in July 1997.

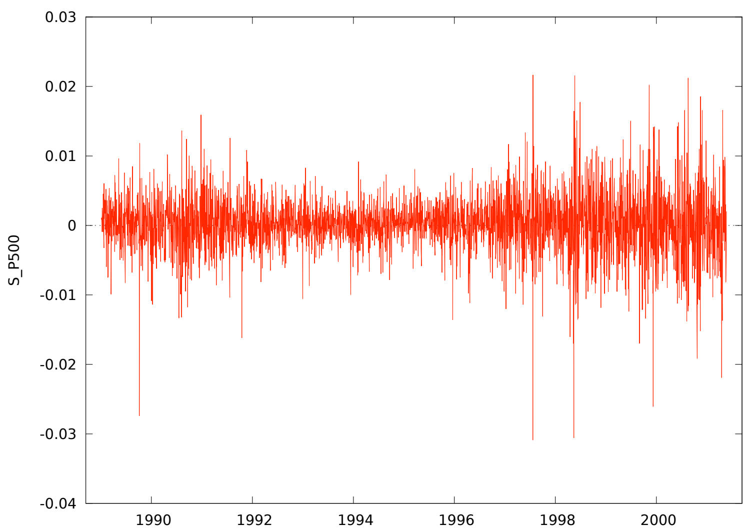

Figure 3: Daily log returns of S&P 500 from 5th January 1989 to 19th October 2001 ($T = 3230$)

In Table 3.4 we present the location of the breaks detected by each procedure. Results for Auto-SEG are taken from Davis et al. (2008). It can be noted the similarities between the break-points detected by the different procedures. IT and KL detected almost the same changes. The three break-points detected by Auto-SEG were also found almost in the same dates by BICx2. Two of the change-points detected by BICgarch are similar to those of IT and KL; the third one was found also by BICx2 and Auto-SEG, and the fourth one was not found by other procedure. Since the Auto-SLEX method gives such results that always have the form of two to the power of a positive integer number, the detected change points are different than the ones detected by other procedures. However, some of them might be very similar to the breaks detected by other methods.

Many of the detected change-points can be related with some shocks affecting the evolution of the S&P 500 Index. The change-point in October 1989 corresponds to the Black Friday mini-crash caused by a reaction to the news of the breakdown of an agreement leveraged buyout of 6750 million for UAL Corporation, the parent company of United Airlines. When the UAL deal failed, helped trigger the collapse of the junk bond market.

Begining 1990 and following in 1991 United States economy exhibited a large stock market recession mainly attributable to the workings of the business cycle and restrictive monetary policy. In December 1991, (change-point detected by all the procedures), the stock market recovered from the recession and resumed a largely stable upward trajectory until the onset of the great stock market bubble began in April 1997 (break detected by KL and BICgarch in March). In 1997 the wold economy was affected by the Asian crisis, which the main impacts were in the second half of 1997 (found by BICx2 and Auto-SEG) and the spread to Russia in August 1998 (detected by LEE), when it was increased the perceived riskiness of the largest corporate cash flows. Finally, the last change-point detected by BICx2 in September 2001 can be related with the Twin Towers attack when the S&P 500 sank 11.6 percent in four days and the volatility increased.

For the detected breaks by each methodology a piecewise GARCH(1,1) is estimated. The summary of the fitted model is presented in Tables 5 and 6 and the conditional volatilities resulting from each model jointly with the estimated breaks are plotted in Figure 4.

The estimations for the three pieces detected by IT and KL are very similar. In the first change-point, $\hat{\omega}$ decreases and $\hat{\beta}$ increases. This change resulted in a smaller marginal variance and a bigger persistence, as can be observed in Figure 4, and can be explained by the stock market recovery from

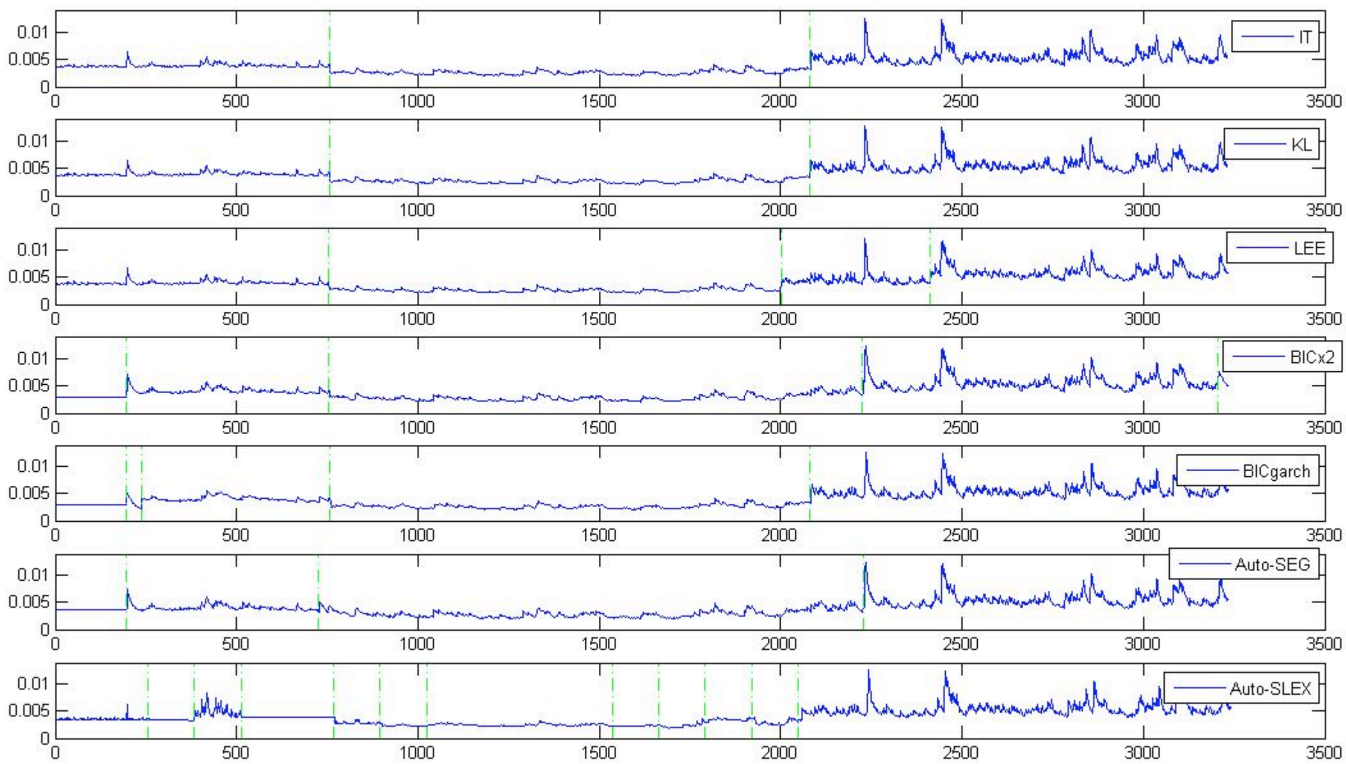

Figure 4: Estimated volatility of the S&P 500 log returns by fitting a piecewise GARCH(1,1) models using the break-points of each procedure

the recession. In the second change-point, the three parameters varied their estimations, increasing $\hat{\omega}$ and $\hat{\alpha}$, and reducing $\hat{\beta}$, and getting an increased marginal variance and less persistence, that can be observed in the Figure 4, which is a expected result, given the instability begining and during the Asian crisis.

The segmentation performed by using LEE is very similar to the one performed by IT and KL, but the last piece is segmented in two intervals. Although both of them have a similar level of persistence, in the second interval $\hat{\alpha}$ is smaller and $\hat{\beta}$ greater. Also the estimation of the constant ω decreased in the second interval with respect to the first one, probably because the diminishing of the Asian crisis effects.

The estimations of the piecewise GARCH(1,1) processes are very similar for BICx2 and Auto-SEG. They found a constant conditional variance until the first change-point, the Black Friday in 1989, as occurred also by using BICgarch. After that, the dynamic of the conditional variance appeared as heteroskedastic, with a persistence slightly higher than 0.9, and a marginal variance which is greater than the previous interval. The stock market recovery increased the persistence of the S&P index conditional variance, which is reflected in the value of $\hat{\beta}$ and a reduction of the marginal variance given by the decreasing of $\hat{\omega}$, both of them in the piece 3. With the Asian crisis, $\hat{\omega}$ and $\hat{\alpha}$ resulted higher and $\hat{\beta}$ smaller than the respective estimations in the previous piece. The persistence was remained relatively constant, but the marginal variance increased. Finally, with BICx2, one more change-point is detected, corresponding to the 11S, where $\hat{\omega}$ and $\hat{\alpha}$ decreased and $\hat{\beta}$ increased.

The change-points detected by Auto-SLEX resulted in periods of heteroskedastic and homoskedastic behavior of the conditional variance. The most notorious result is that after the Asian crises the marginal variance increased, as the other procedures showed.

Finally, as a measure of the goodness of the segmentation performed we estimated piecewise GARCH(1,1) models according to the change-points detected by each procedure, and computed the BIC for that models to obtain a measure of the segmentation goodness. The smallest BIC is obtained for the segmentation performed by Auto-SEG.

9 Conclusions

In this paper we explored, analysed and applied the change-points detection and estimation procedures to conditional heteroskedastic processes. Based on the fact that a GARCH process can be expressed as an ARMA model in the squares of the variable, we proposed to detect and locate change-points by using the BIC as an extension of its application in linear models.

As cusum methods, BICx2 s characterized by computational simplicity, reducing difficulties of the change-point detection in the complex non-linear processes. By the simulation performed, we obtained a good size and power properties in detecting even small magnitudes of change and for low levels of persistence. Since we focused on GARCH(1,1) processes with Gaussian perturbations, we suggest to analyse the performance of the proposed procedure both to GARCH(1,1) processes with t-student perturbations and to Stochastic Volatility models.

Finally, the procedures were applied to the S&P500 log returns time series, in order to compare with the results in Andreou and Ghysels (2002) and Davis et al. (2008). Change-points detected by BICx2 were similar to the breaks found by the other procedures, and their location can be related with the Southeast Asia financial crisis and with other known financial events.

References

Aggarwal, R., C. Inclan, and R. Leal (1999). Volatility in emerging stock markets. *Journal of Financial and Quantitative Analysis 34*(1), 33–55.

Andersen, T. (1994). Stochastic autoregressive volatility: a framework for volatility modeling. *Mathematical finance 4*(2), 75–102.

Andreou, E. and E. Ghysels (2002). Detecting multiple breaks in financial market volatility dynamics. *Journal of Applied Econometrics 17*(5), 579–600.

Bai, J. (1997). Estimating multiple breaks one at a time. *Econometric Theory 13*, 315–352.

Berg, A., R. Meyer, and J. Yu (2004). Deviance information criterion for comparing stochastic volatility models. *Journal of Business and Economic Statistics 22*(1), 107–120.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics 31*(3), 307–327.

Carrasco, M. and X. Chen (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory 18*(1), 17–39.

Chen, J. and A. Gupta (1997). Testing and locating variance changepoints with application to stock prices. *Journal of the American Statistical Association 92*(438), 739–747.

Davis, R., T. Lee, and G. Rodriguez-Yam (2006). Structural Break Estimation for Nonstationary Time Series Models. *Journal of the American Statistical Association 101*(473), 229–239.

Davis, R., T. Lee, and G. Rodriguez-Yam (2008). Break detection for a class of nonlinear time series models. *Journal of Time Series Analysis 29*(5), 834–867.

Donoho, D., S. Mallat, and R. von Sachs (1998). Estimating Covariances of Locally Stationary Processes: Rates of Convergence of Best Basis Methods. *Technical Report. Dept. Statist., Stanford Univ., Stanford, CA 517*, 1–64.

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society 50*(4), 987–1007.

Engle, R. and T. Bollerslev (1986). Modelling the persistence of conditional variances. *Econometric reviews 5*(1), 1–50.

Fryzlewicz, P. and S. Subba Rao (2011). Mixing properties of ARCH and time-varying ARCH processes. *Bernoulli 17*(1), 320–346.

Fukuda, K. (2010). Parameter changes in GARCH model. *Journal of Applied Statistics 37*(7), 1123–1135.

Galeano, P. and R. Tsay (2010). Shifts in individual parameters of a GARCH model. *Journal of Financial Econometrics 8*(1), 122–153.

Huang, H., H. Ombao, and D. Stoffer (2004). Discrimination and Classification of Nonstationary Time Series Using the SLEX Model. *Journal of the American Statistical Association 99*(467), 763–774.

Hyung, N., S. Poon, and C. Granger (2009). A source of long memory in volatility. *Frontiers of Economics and Globalization 3*, 329–380.

Inclán, C. and G. Tiao (1994). Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association 89*(427), 913–923.

Kim, S., S. Cho, and S. Lee (2000). On the cusum test for parameter changes in GARCH (1,1) models. *Communications in Statistics-Theory and Methods 29*(2), 445–462.

Kokoszka, P. and R. Leipus (1999). Testing for parameter changes in ARCH models. *Lithuanian Mathematical Journal 39*(2), 182–195.

Kokoszka, P. and R. Leipus (2000). Change-point estimation in ARCH models. *Bernoulli 6*(3), 513–539.

Kulperger, R. and H. Yu (2005). High moment partial sum processes of residuals in GARCH models and their applications. *The Annals of Statistics 33*(5), 2395–2422.

Lamoureux, C. and W. Lastrapes (1990). Heteroskedasticity in stock return data: volume versus GARCH effects. *Journal of Finance 45*(1), 221–229.

Lavielle, M. and E. Moulines (2000). Least-squares Estimation of an Unknown Number of Shifts in a Time Series. *Journal of time series analysis 21*(1), 33–59.

Lee, S., J. Ha, O. Na, and S. Na (2003). The cusum test for parameter change in time series models. *Scandinavian Journal of Statistics 30*(4), 781–796.

Lee, S., O. Na, and S. Na (2003). On the cusum of squares test for variance change in nonstationary and nonparametric time series models. *Annals of the Institute of Statistical Mathematics 55*(3), 467–485.

Lee, S., Y. Tokutsu, and K. Maekawa (2004). The CUSUM test for parameter change in regression models with ARCH errors. *Journal of the Japanese Statistical Society 34*, 173–188.

Malik, F. (2003). Sudden changes in variance and volatility persistence in foreign exchange markets. *Journal of Multinational Financial Management 13*(3), 217–230.

Malik, F. and S. Hassan (2004). Modeling volatility in sector index returns with GARCH models using an iterated algorithm. *Journal of Economics and Finance 28*(2), 211–225.

Mikosch, T. and C. Starica (2004). Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *Review of Economics and Statistics 86*(1), 378–390.

Morana, C. and A. Beltratti (2004). Structural change and long-range dependence in volatility of exchange rates: either, neither or both? *Journal of Empirical Finance 11*(5), 629–658.

Nouira, L., I. Ahamada, J. Jouini, and A. Nurbel (2004). Long-memory and shifts in the unconditional variance in the exchange rate euro/us dollar returns. *Applied Economics Letters 11*(9), 591–594.

Ombao, H., J. Raz, R. von Sachs, and W. Guo (2002). The SLEX Model of a Non-Stationary Random Process. *Annals of the Institute of Statistical Mathematics 54*(1), 171–200.

Pooter, M. and D. Dijk (2004). Testing for changes in volatility in heteroskedastic time series: a further examination. Technical report, Erasmus School of Economics (ESE).

Sansó, A., V. Aragó, and J. Carrion (2004). Testing for changes in the unconditional variance of financial time series. *Revista de Economía Financiera 4*, 32–53.

Starica, C., S. Herzel, and T. Nord (2005). Why does the GARCH (1,1) model fail to provide sensible longer-horizon volatility forecasts.

West, M., R. Prado, and A. Krystal (1999). Evaluation and Comparison of EEG Traces: Latent Structure in Nonstationary Time Series. *Journal of the American Statistical Association 94*(446), 375–376.

Wickerhauser, M. V. and C. K. Chui (1994). *Adapted wavelet analysis from theory to software.* AK Peters Wellesley.

Yao, Y. (1988). Estimating the number of change-points via Schwarz criterion. *Statistics & Probability Letters 6*(3), 181–189.

Table 2: Proportion of estimated change-points based on 500 replications when there is a break at $t = 501$ in the GARCH process

Process	Procedure	IT	KL	LEE	BICx2	BICgarch	Auto-SEG	Auto-SLEX
	No break	0.870	0.850	0.958	0.922	**1.000**	0.958	0.902
1	1 break	0.130	0.116	0.038	0.078	0.000	0.042	0.092
	≥ 2 break	0.000	0.034	0.004	0.000	0.000	0.000	0.006
	No break	0.772	0.904	0.976	0.941	**1.000**	0.956	0.820
2	1 break	0.218	0.050	0.024	0.059	0.000	0.044	0.142
	≥ 2 break	0.000	0.048	0.000	0.000	0.000	0.000	0.038
	No break	0.974	0.500	0.226	0.268	0.984	0.804	0.648
3	1 break	0.026	0.488	0.694	**0.728**	0.016	0.192	0.336
	≥ 2 break	0.080	0.000	0.012	0.004	0.000	0.004	0.016
	No break	0.835	0.500	0.200	0.000	0.004	0.000	0.756
4	1 break	0.165	0.280	0.796	0.900	**0.976**	0.964	0.266
	≥ 2 break	0.016	0.220	0.004	0.100	0.020	0.036	0.018
	No break	0.418	0.006	0.136	0.014	0.394	0.370	0.094
5	1 break	0.578	0.524	**0.862**	0.806	0.602	0.626	0.652
	≥ 2 break	0.004	0.470	0.012	0.180	0.004	0.004	0.254
	No break	0.000	0.000	0.000	0.000	0.008	0.004	0.000
6	1 break	0.576	0.742	0.976	0.956	**0.978**	0.978	0.670
	≥ 2 break	0.424	0.258	0.024	0.044	0.014	0.018	0.330
	No break	0.996	0.370	0.386	0.268	0.876	0.878	0.786
7	1 break	0.004	0.584	0.594	**0.670**	0.124	0.122	0.198
	≥ 2 break	0.000	0.046	0.020	0.062	0.000	0.000	0.016
	No break	0.066	0.000	0.004	0.000	0.200	0.072	0.050
8	1 break	0.744	0.888	0.892	0.818	0.710	**0.912**	0.778
	≥ 2 break	0.190	0.112	0.104	0.182	0.090	0.016	0.172
	No break	0.002	0.000	0.360	0.000	0.294	0.068	0.002
9	1 break	0.778	0.530	0.638	0.822	0.704	**0.910**	0.668
	≥ 2 break	0.020	0.470	0.002	0.122	0.002	0.022	0.330
	No break	0.000	0.024	0.100	0.000	0.050	0.008	0.000
10	1 break	0.601	0.688	0.820	0.898	0.950	**0.952**	0.594
	≥ 2 break	0.399	0.288	0.080	0.102	0.000	0.040	0.406

Table 3: Proportion of estimated change-points based on 1000 replications when there are two breaks at $k_1^* = 340$ and $k_2^* = 680$ in the GARCH process with parameters $\omega_1 = 1$, $\omega_2 = \omega_3 = 1.5$, $\alpha_1 = \alpha_2 = \alpha_3 = 0.03$, $\beta_1 = \beta_2 = 0.8$ and $\beta_3 = 0.9$

	IT	KL	LEE	BICx2	BICgarch	Auto-SEG	Auto-SLEX
No break	0.048	0.000	0.000	0.000	0.000	0.001	0.000
1 break	0.780	0.444	0.004	0.307	0.371	0.310	0.000
2 breaks	0.172	0.446	0.679	0.678	0.620	**0.720**	0.000
$\geq$2 breaks	0.000	0.110	0.317	0.015	0.009	0.002	1.000

Table 4: Change-point locations detected by all the procedures

Procedure	Number of breaks	Location
IT	2	31/12/91, 27/03/97
KL	2	2/01/92, 26/03/97
LEE	3	30/12/91, 2/12/96, 20/07/98
BICx2	4	13/10/89, 30/12/91, 21/10/97, 10/09/01
BICgarch	4	13/10/89, 11/12/89, 31/12/91, 27/03/97
Auto-SEG	3	13/10/89, 15/11/91, 27/10/97
Auto-SLEX	11	9/01/90, 12/07/90, 14/01/91, 17/01/92, 21/07/92, 21/01/93, 16/12/94, 3/08/95, 5/02/96, 7/08/96, 7/02/97

Table 5: Estimated coefficients of the piecewise GARCH(1,1) processes (Part 1)

Parameter	IT	KL	LEE	BICx2	BICgarch	Auto-SEG
PIECE 1						
ω	2.2e-06	2.1e-06	2.4e-06	8.96e-06	8.96e-06	1.35e-05
β	0.820	0.830	0.809	0.000	0.000	0.000
α	0.039	0.036	0.041	0.000	0.000	0.000
PIECE 2						
ω	2e-07	2e-07	2e-07	1.49e-06	2e-07	1.46e-06
β	0.931	0.932	0.935	0.868	0.937	0.862
α	0.042	0.041	0.036	0.042	0.000	0.049
PIECE 3						
ω	2.32e-06	2.4e-06	2.74e-06	2e-07	5.3e-07	2e-07
β	0.819	0.817	0.761	0.928	0.942	0.917
α	0.110	0.111	0.105	0.049	0.027	0.064
PIECE 4						
ω	-	-	2.35e-06	1.55e-06	2e-07	1.84e-06
β	-	-	0.846	0.858	0.931	0.843
α	-	-	0.009	0.095	0.042	0.101
PIECE 5						
ω	-	-	-	2e-07	2.32e-06	-
α	-	-	-	0.960	0.819	-
β	-	-	-	0.010	0.110	-
BIC	-2.6833	-2.6840	-2.6821	-2.6834	-2.6844	-3.5382

Table 6: Estimated coefficients of the piecewise GARCH(1,1) processes (Part 2)

Parameter	Auto-SLEX	
	PIECE 1	PIECE 7
ω	1.24e-05	2e-07
β	0.000	0.943
α	0.000	0.024
	PIECE 2	PIECE 8
ω	2e-07	2e-07
β	0.981	0.952
α	0.000	0.011
	PIECE 3	PIECE 9
ω	4.92e-05	2e-07
β	0.619	0.906
α	0.190	0.067
	PIECE 4	PIECE 10
ω	1.31e-05	1.13e-06
β	0.094	0.889
α	0.000	0.019
	PIECE 5	PIECE 11
ω	1.75e-06	2.42e-07
α	0.711	0.915
β	0.059	0.061
	PIECE 6	PIECE 12
ω	3.52e-07	2.31e-06
α	0.911	0.820
β	0.023	0.109
BIC	-2.6685	