Title: Analysis of break-points in non-stationary time series.

By: Jean Remy Habimana

Abstract

Previous studies have shown that the structure of time series may change at any time due to the change in mean and/or variance of the series, which violates the assumption of stationarity and creates uncertainty in parameter estimates and forecasting. This paper, discusses a method of analyzing non-stationary time series, by breaking the series into smaller locally stationary series. The specific focus in this paper will be on financial time series. To achieve the main goal, which is locating beak-points: a point in the series where occurs a sudden change in mean and/or variance, this method uses Maximum Likelihood Estimation (MLE) by means of Simulated Annealing (SA) optimization. Then the originally non-stationary time series is divided into locally stationary time series according to the number of break points found. The analysis by this method considers each interval on which the series is stationary as an independent time series with its specific parameters, which enhances the accuracy in parameter estimates and forecasting. Applications to simulated and real financial data are presented.

Introduction

Time series are generally series of random values observed at equally spaced intervals of time; this is because we are unsure what is going to happen in the future. Even if we can predict the future based on the historical information, our predictions are not exact; there is always a given uncertainty. This randomness becomes even more interesting when at some point in time there is a significant change in the structure of the series; which should get the attention of analysts because these structural changes make the series non-stationary. Some suggestions have already been given about how to solve this problem of nonstationarity; Starica and Granger(2005), proposed a method of detecting break-points in a time series by using a goodness of fit test to see if the observations in a given interval come from the same distribution or if they should be allowed to have different distributions.

Other methods of modeling a nonstationary time series, by allowing certain intervals to have their own distributions were proposed in previous research on time series analysis. For example, Ozaki and Tong (1975), suggested that we can assume there is a change point in the series and use the likelihood function to estimate the location of the change-point, and apply final prediction error (FPE) method to assess the performance of model fitted to the entire dataset and the one fitted locally, considering two independent distributions in the series, so as to see which one is better.

Fitting a model to a nonstationary time series by breaking it into locally stationary series is reasonable because in practice it is likely to have a time series whose structure may change over time. This is why we need more methods to detect the break-points in a time series.

In real life, we can observe this unpredictable significant change in mean and variance at any time, but in this paper we are focusing on financial time series, in which this structure is more observed due to the stochastic changes of the financial market. (Aue et al.,2011). According to Starica and Granger (2005), the sample auto correlation function (ACF) of financial time series is likely to have a structure characterized by a slow decay for the

first lags that ends up being stable at higher lags with positive correlation; they add that this structure is due to the change in unconditional mean and variance. Remember that this is the same structure we observe in non-stationary time series. Two questions arise from this observation: either the ACF in financial time series indicates that there is a long-range dependence; in other words, there is a significant correlation at different lags in the series, or simply the series is not stationary. Mikosch and Satrica (2004) emphasize that it is strongly reasonable to say that the slow decay observed in the ACF is a result of change in mean and variance of the series rather than saying it is due to the strong correlation in the series.

Aue and Horvath (2011) say that is not advisable to assume long-range dependence in a series based only on the slow decay observed in the ACF because the same structure can be observed in short memory processes with structural breaks. Consequently, the idea of considering financial time series stationary and assuming that the slow decay that stabilizes at higher lags is due to a long-range dependence in the series is dubious. Hence, there are strong reasons to consider financial time series as nonstationary due the structural change in time.

From these previous works, considerable studies should be done to find different ways of analysis of financial time series that are not based on the ideas of long-range dependence in stochastic data like financial market data.

In this paper, based on real life financial market random feature, we are developing a new method to estimate the random changes in the structure of a time series, which means that we are dealing with nonstationary time series characterized by unpredictable changes in unconditional mean and variance. In this methodology, our first goal is to locate the break-points and divide the series into small intervals having the same mean, variance and with the same covariance; hence the assumption of stationarity will work on each interval which is going to make the analysis easy and gives more accurate results. After breaking the series into locally stationary intervals, we are calculating Bayesian Information Criterion (BIC) for the series without any break and the BIC for the series that is divided into small stationary series to see if dividing the series into locally stationary series gives smaller BIC. Finally, we are comparing the BIC for different models with different break-points to see which one is better.

We tested our methodology on a simulation study by drawing observations from a normal distribution with the same mean and variance for each small interval, and compare the result with simulated data by considering same mean and variance for the entire series. We apply the method to the real data S&P 500 series of return to see if breaking the series into small locally stationary series gives a better model than assuming the whole series to be stationary.

1. Understanding financial time series and stationarity

1.1. Financial time series

A time series in general, is a set of data collected at equally spaced intervals of time. It may be for instance, daily, weekly, monthly…data. Depending on the nature of the data, time series may have different characteristics, but structural change may occur in any time series. In this paper we are considering the case of financial time series because it is one of the series where the structure is likely to change over time.

As all time series, financial market data are generally random data due to the fact that it is not easy to predict how the market will be in the future, and we don't know when the structure of the series may change. That is why the series of returns is random and its mean and/or variance may change at any time (Starica & Granger, 2005). So, when modeling time series of returns we should care not only about long range dependent and autocorrelation, but also about the structural change. Consequently, financial time series are most likely not to be stationary. In addition to the change in unconditional mean of the series, the plot of a financial time series does not generally have a consistent trend. Moreover, there is a high volatility, which is a measure of dispersion of return for a given security or market index, in financial market data; this implies that the unconditional variance is also unpredictable and changing over time (Mikosch and Starica, 2004).

Financial data are generally presented in terms of returns such as stock returns, market index returns, and any other form of returns. For a better analysis of return time series, some transformations may be done, especially to make the series satisfy the assumption of normality in the data because the returns are not usually normally distributed. First, it is better to use absolute return to avoid having negative values of return, and then take the log of the absolute return because financial time series are heavy tailed; so this will make it approximately normal (Starica & Granger, 2005).

This can be done mathematically as follows,

Let r_t be a series of return and X_t the series after the two above transformations.

$X_t = \log |r_t|$, the series of log of the absolute return.

Considering two consecutive observations of a financial series, $P_t \; and \; P_{t-1}$, we can define the return as

$$r_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1$$

Note that in the analysis of the series of return we will use the series X_t , the series of log of absolute returns that satisfies the assumption of normality.

1.2. Stationarity in financial time series

As explained by Aue and Horvath (2011), a time series is said to be second-order stationary if its statistical parameters, mean, variance and covariance, do not change in time. Based on this definition, it's not advisable to assume that financial time series are stationary because they are characterized by random changes in unconditional mean and variance.

But in fact, even if the entire time series is not likely to be stationary, there are some specific intervals on which the mean and variance are constant; the series is stationary on these small intervals. This means that despite the whole series being non-stationary, we can find intervals on which the series is stationary by breaking it into smaller series on which the mean and variance are constant; we call these intervals, intervals of homogeneity (Starica & Granger, 2005).

Regardless of the complex nature of financial time series, considerable research has proven that there may be many ways to delimitate these small intervals of almost homogenous data, with constant mean and variance. The following methodology used in this paper

describes well how to break the entire non-stationary time series into smaller stationary time series, and explains in details how to use maximum likelihood estimation to locate the break points and estimate the parameters for each interval.

I. Methodology

1. Homogeneous intervals

Homogeneous intervals are small intervals on which the mean and variance remain approximately constant. The number of these homogeneous intervals depends on the structure of the data. Some series may have many or few intervals depending on how the unconditional mean of the series changes in time.

Let $X_{1:T}$ be the entire series, T the total number of observations in the series. The series $X_{1:T}$will be $X_1, X_2, X_3,\ldots X_{T-2}, X_{T-1,} X_T$. We need to locate the times at which the mean of the series changes, and we call these points, break-points. For i = 1,2 …k, where k is the number of break-points, let us denote t_i any break-point; the series $X_1, X_2, X_3,\ldots X_{T-2}, X_{T-1,} X_T$ can be written in interval form as

$$X_{1:T} = \begin{cases} x_1, x_2 \ldots x_{t_1}, & with\ mean\ \mu_1\ and\ variance\ \sigma_1^2 \\ x_{t_1+1}, \ldots x_{t2} & with\ mean\ \mu_2\ and\ variance\ \sigma_2^2 \\ x_{t_2+1} \ldots x_{t3} & with\ mean\ \mu_3\ and\ variance\ \sigma_3^2 \\ & \vdots \\ x_{t_{k-1}+1} \ldots x_{tk}, & with\ mean\ \mu_k\ and\ variance\ \sigma_k^2 \end{cases}$$

With k number of break-points in a series, the series will have k+1 homogeneous intervals on which the mean and variance are constant. In other words, every time there is a significant change in the mean of the series, we have a new interval.

Assuming there is no linear dependence in the series on each homogeneous interval, the observations in every single interval of homogeneity are independent random variables with the same mean and variance. This also means that in every interval of homogeneity the series is stationary due to the fact that the mean and variance remain constant.

To conclude, on each homogeneous interval we have a series of independent random variables with the same mean and variance.

2. Delimitation of homogeneous intervals and parameter estimation

The main goal in this section is to explain in detail how to locate the break-points and find estimates of the parameters in each local stationary interval.

Let $X_1, X_2, X_3,\ldots X_{T-2}, X_{T-1,} X_T$ be the entire non-stationary series with T total number of observations.

Let k be the number of break-points, which gives us k+1 homogeneous intervals.

Let t_i with i =1,2,…k denote the break-points. The number of observations in each homogeneous interval is calculated as follows:

The first interval contains n_1 observations
The last interval contains T -t_k observations
The middle interval contains $t_i - t_{i-1}$ observations.

As we discussed before the series of log absolute return is approximately normal, in each local stationary interval, the observations are independent approximately normally distributed random variables with the same mean and variance, which enables us to assume that the probability distribution function (pdf) on each homogeneous interval is that of the normal distribution with specific mean and variance.

Let f_i with i =1,2,…k+1, be the pdf for the observations in each homogeneous interval; since all of the observations x_t of the series are independent and normally distributed, for every interval i, with i= 1,2,3,…k+1, f_i is defined as follows:

$$f(X_{(t_{i-1}+1):t_i}) = \left(\frac{1}{2\pi\sigma_i^2}\right)^{\frac{n_i}{2}} exp\left\{-\frac{1}{2\sigma_i^2}\sum_{t=t_{i-1}+1}^{t_i}(X_t - \mu_i)^2\right\},$$

with $n_i = t_i - t_{i-1}$, the number of observations in each interval i and the corresponding mean and variance μ_i and σ_i^2 respectively.

Note that $t_0 = 0$ and $t_{k+1} = T$.

In order to locate the break points and estimate all of the means and variances from each interval, the maximum likelihood estimation is used.

3. Bayesian Information Criterion (BIC) and parameter estimation

3.1. Bayesian Information Criterion (BIC)

BIC is a model selection tool that is used to select the best model for a given dataset among a finite number of models. It is defined in terms of the likelihood function, but with a penalty term for the number of parameters. Since the penalty term increases with the number of parameters, it avoids overfitting.

BIC is given by the following formula:

BIC = -2*$\hat{l}$ +Ω*log (n), where Ω is the total number of parameters, $\hat{l} =$ is the maximized loglikelihood, and n the sample size. In our case, n = T, Ω = 2*(k+1) + k where k is the number of break-points and (k+1) the number of homogeneous intervals. So, the BIC is given by

BIC = -2*$\hat{l}$ + (3*k+2)*log (T). When using BIC as a model selection tool, among all the candidates, we choose the model that gives the lowest BIC.

3.2. Parameter estimation

The best estimates for our parameters, break-points, is a set of break-points that give the minimum BIC. Let θ be the set of all parameters to be estimated, local means and variances and all break-points. Since the data from each interval is an independent set of random variables, the overall likelihood function is a product of local likelihood functions. This overall likelihood is denoted by $L(\theta; X_{1:T})$

Of all possible values of the parameters the estimates of the parameters obtain from maximum likelihood estimation are those for which the likelihood function is the largest.

Mathematically, the overall likelihood function is obtained by multiplying all the local probability distributions $f(X_{(t_{i-1}+1):t_i} \mid \theta)$.

$$L(\theta; X_{1:T}) = \prod_{i=1}^{k+1} \left(\frac{1}{2\pi\hat{\sigma}_i^2} \right)^{\frac{n_i}{2}} exp\left\{ -\frac{1}{2\hat{\sigma}_i^2} \sum_{t=t_{i-1}+1}^{t_i} (X_t - \hat{\mu}_i)^2 \right\},$$

$$= \prod_{i=1}^{k+1} \left(\frac{1}{\hat{\sigma}_i^2} \right)^{\frac{n_i}{2}} exp\left\{ -\frac{1}{2\hat{\sigma}_i^2} \sum_{t=t_{i-1}+1}^{t_i} (X_t - \hat{\mu}_i)^2 \right\}, 2\pi \text{ is}$$

dropped because it is a constant.

For a given set of break-points, the log likelihood function can be maximized in closed form with respect to μ_i and σ_i^2 , I = 1 ,2 …k+1.

From standard calculation for normal samples,

$\hat{\mu}_i = \frac{1}{n_i} \sum_{t=t_{i-1}+1}^{t_i} X_t$ and $\hat{\sigma}_i^2 = \frac{1}{n_i} \sum_{t=t_{i-1}+1}^{t_i} (X_t - \hat{\mu}_i)^2$; then it follows that

the maximized log likelihood is $l(\hat{\theta}; X_{1:T}) = \log L(\hat{\theta}; X_{1:T})$

$$= \log\left(\prod_{i=1}^{k+1} \left(\frac{1}{\hat{\sigma}_i^2} \right)^{\frac{n_i}{2}} exp\left\{ -\frac{1}{2\hat{\sigma}_i^2} \sum_{t=t_{i-1}+1}^{t_i} (X_t - \hat{\mu}_i)^2 \right\} \right)$$

$$= \sum_{i=1}^{k+1} \left(\log\left(\frac{1}{\hat{\sigma}_i^2} \right)^{\frac{n_i}{2}} exp\left\{ -\frac{1}{2\hat{\sigma}_i^2} \sum_{t=t_{i-1}+1}^{t_i} (X_t - \hat{\mu}_i)^2 \right\} \right)$$

$$= \sum_{i=1}^{k+1} \left(-\frac{n_i}{2} \log(\hat{\sigma}_i^2) - \frac{1}{2\hat{\sigma}_i^2} \sum_{t=t_{i-1}+1}^{t_i} (X_t - \hat{\mu}_i)^2 \right)$$

$$= \sum_{i=1}^{k+1} \left(-\frac{n_i}{2} \log(\hat{\sigma}_i^2) - \frac{n_i}{2} \right)$$

$$= \sum_{i=1}^{k+1} \left(-\frac{n_i}{2} (\log \hat{\sigma}_i^2 + 1 \right)$$

$$= -\sum_{i=1}^{k+1} \left(\frac{n_i}{2} (1 + \log \hat{\sigma}_i^2 \right).$$

Hence, $l(\hat{\theta}; X_{1:T}) = -\sum_{i=1}^{k+1} \left(\frac{n_i}{2} (1 + \log \hat{\sigma}_i^2 \right)$,

with n_i the number of observations and $\hat{\sigma}_i^2$ the variance in i homogeneous interval.

To find the break-points we need to maximize $l(\hat{\theta}; X_{1:T})$ over t_1 , t_2 , …t_k by using an optimization method that deals with discrete data because the break points are not continuous.

4. Simulated annealing

We have seen that the number of all possible break-points depends on the structure and the size of the data; the bigger the size or the more the variability in the data, the larger the number of all possible break-points. With size larger than 100 data points generally, it's not practically easy to explore all possible combinations of break points to see which one gives the best model for the data, small BIC. For example, with n=100, the number of possible combinations of two, three, and four break-points are 4851, 152096, and 3464840 respectively. In this paper we are using simulated annealing to solve this problem.

As it is defined by Bertsimas and Tsitsiklis (1993), simulated annealing is a probabilistic strategy of locating a global minimum of a function that has many different local minima. In this process you start with an initial random solution and let it change to another solution, better or worse, to help explore a big range of solutions from which we can get one close to the optimal solution, since we cannot explore all possible solution to make sure that we have obtained the optimal solution.

In our case, the function to be minimized is the BIC which is a function of the likelihood function; we also need a function that will propose the different number and positions of break-points and observe how the BIC changes depending on the number of break-points and/or the position of break-points.

How does simulated annealing work?

Simulated annealing has its origin in metal work in which the goal is to obtain low energy states of a solid metal.

How does annealing of metals work?

1. A metal is heated up in a heat bath until it melts; at this stage the energy level is very high.
2. Then it is cooled gradually until it reaches the lowest energy state. This will alter the internal structure of the metal, which will change its physical properties. But, this happens when the temperature is decreased slowly and carefully.
3. To get the lowest energy state, we are decreasing the temperature, which causes the particles in the metal to move slowly and hence the energy reduces. So, the energy state of the metal depends on the temperature.

For the algorithm to move from a high energy state to a lower energy state, which is our goal, there should be some conditions:

Acceptance conditions:

Here we are describing how we move from a current energy state to a neighbor energy state in order to reach the ground state. In other words, how do we decide which solution to accept!

- First, the algorithm checks if the proposed solution is better than, or is the same as the current solution. If it is, accept it with no conditions.
- If the proposed solution is worse than the current solution, we accept it with some probability depending on:

- How worse the proposed solution is compared to the current solution and also depending on
- The temperature of the system.

To make this clear, let x_c be the current configuration and x_p the proposed configuration, and T the temperature of the system. Recall that a better solution here means lower energy state.Then if $g(x_p) \leq g(x_c)$, we accept the proposed solution and it becomes the new solution. If however, $g(x_p) > g(x_c)$, we will accept the proposed solution with probability

$\alpha = \exp\left\{-\frac{(g(xp)-g(xc))}{T}\right\}$. Otherwise, keep the current solution.

Note that in simulation annealing accepting some worse solutions enables the algorithm not to get stuck in the local minima.

In the above formula if $g(x_p) > g(x_c)$, we will always have a negative exponent and the value of α will depend on the difference between the two solutions and the temperature of the system.

More explicit,

- ✓ For high temperatures, the exponent in the above formula approaches zero, and the probability of accepting a worse solution approaches 1
- ✓ As the temperature decreases to zero, the exponent in the above formula goes to -∞, and the probability of accepting a worse solution goes to zero. So, we are less likely to accept a worse solution as the number of iterations increases.

We can illustrate by example how the probability of accepting a neighbor solution depends on the temperature and how much worse the neighbor solution is in the following table:

Table1: Probability of accepting a neighbor solution in simulated annealing.

	Probability of accepting a neighbor solution		
Temperature	$g(x_p)$-$g(x_c)$ = 1	$g(x_p)$-$g(x_c)$ = 2	$g(x_p)$-$g(x_c)$ = 3
10	0.9	0.82	0.74
1	0.37	0.14	0.05
0.25	0.018	0.0003	0.000006
0.1	0.00005	2×10^{-9}	9×10^{-14}

From the above table we observe that:

- For a fixed temperature, the probability of accepting a worse solution decreases as the proposed solution gets farther from the current solution,
- And for a fixed difference in solution, the probability of accepting a worse solution decreases as temperature also decreases.

Remark: the algorithm stops after a fixed number of iterations.

In optimization problem we simulate this process, where we have a random variable that changes at each stage to propose a new solution and follow the same acceptance conditions in order to reach the global minimum of our function.

In this paper, our state variable is a set of break-points we propose and calculate the BIC. Each time we randomly select a new set of beak-points and calculate the value of BIC.

If the proposed set of break-points gives a smaller than, or equal to the current value of BIC, we accept it with no conditions. But if the proposed set of break-points gives a higher value for BIC, we accept it with the probability we described above.

4.1. How to change the break-points and their positions

The idea here is trying to explore many possible positions of break-points. To minimize the BIC we need an initial set of break-points, which will be randomly changed to get different values for BIC until we reach the minimum value. These changes are either positions or the number of the break-points. Here we are using three ways to obtain a new set of break-points.

- First, adding one more random break-point if you haven't reached the max possible number of break-points,
- Second, delete one break-point if there are at least two of them
- Third, keep the same number of break-points and propose to move a random break-point to a new position.

Generally, give the same chance to delete or insert a new break-point; and slightly more chance to change the position since we want to explore many different positions to locate where may be a break-point in the series.

The initial set of break-points may be any break points less than the maximum possible number of break-points, but we have observed that starting with one initial random break-point gives more accurate solution and is easier to choose.

Remark: The maximum number of break-points is (T-2)/2 if n is even, and (T-3)/2 if n is odd. Therefore, the initial or any other set of break-points cannot be bigger than these values. So, to get a reasonable set of break-points there are some restrictions when proposing a new set of break points by changing the number or positions.

4.2. Reasonable break-points

For any dataset, it is not reasonable to propose a break-point at the first and last position in the series.

When inserting a break-point between two other break-points, make sure that the difference between their positions is at least five, and check if after inserting a break-point, the difference between all consecutive break-points is at least two.

Note: We should avoid constructing an interval containing only one observation because we are not considering the distribution of a single observation.

With all of the above conditions satisfied, every change made in the break-points, change of position or number, the main function to be minimized, BIC, returns a different value. Therefore, we can find which of these values is the smallest, and then note the corresponding number of break-points and their positions, which give us the best model.

Remember that the goal of this paper is to develop a methodology to find the break-points in a given times series that exhibits structural change behavior.

We are using Monte Carlo simulation study to analyze the distribution of the outcome, which enables to assess the properties of the estimation process.

II. Monte Carlo method and Simulation study

1. Simulation study

Since the log absolute return has approximately a normal distribution, in this simulated study we are considering a series from a normal distribution with known means and variances.

Consider a series with specified break-points, with known means, variance and size for each homogeneous interval.

To make calculations not too complicated, let k, the number of break-points, be 5. So, we will have k+1 homogeneous intervals of size n_i, which gives us 6 intervals with 6 different means and variances. It is better to consider homogeneous intervals with different sizes.

Let us choose the sizes for the first to the last interval to be:

$$n_1 = 90, n_2 = 85,\ n_3 = 70, n_4 = \ 75, = n_5 = \ 80,\ n_6 = 100.$$

Note that in this paper we are using the size greater or equal to 500 to increase the penalty term in BIC.

Let $\mu_1, \mu_2, \mu_3, \ \mu_4\ , \mu_5\ , \mu_6$, the mean for each interval be 25,35,45,40,30,25 respectively, and variances $\boldsymbol{\sigma}_1{}^2, \boldsymbol{\sigma}_2{}^2, \ \boldsymbol{\sigma}_3{}^2\ , \boldsymbol{\sigma}_4{}^2, \boldsymbol{\sigma}_5{}^2, \ \boldsymbol{\sigma}_6{}^2$, be 1.5, 2, 2.5, 0.8, 1.5, 1 respectively. Means and variances do not have any specific restrictions as long as the chosen choose values are not very close to each other. This simulation gives a series with five break-points and six homogeneous intervals with specified different means and variances

Let us summarize our simulation in the following table:

Table2: Simulation study summarized

Homogeneous intervals	Number of observations in each interval	Mean	Variance
First interval	$n_1 = 90$	μ_1= 25	$\boldsymbol{\sigma}_1{}^2$=1.5
Second interval	$n_2 = 85$	μ_2=35	$\boldsymbol{\sigma}_2{}^2$=2
Third interval	$n_3 = 70$	μ_3=45	$\boldsymbol{\sigma}_3{}^2$=2.5
Forth interval	$n_4 = 75$	μ_4=40	$\boldsymbol{\sigma}_4{}^2$=0.8
fifth interval	$n_5 = 80$	μ_5=30	$\boldsymbol{\sigma}_5{}^2$=1.5
Sixth interval	$n_6 = 100$	μ_6=25	$\boldsymbol{\sigma}_6{}^2$=1

The main goal in this simulation study is to see if we can use the method described above, using simulated annealing, to recover the proposed break-points. We know in advance that the data contains five break-points located at the following positions where there is a change in mean and variance of the series:

$$t_1 = n_1 = 90 ,$$

$$t_2 = t_1 + n_2 = 175,$$

$$t_3 = t_2 + n_3 = 245,$$

$$t_4 = t_3 + n_4 = 320,$$

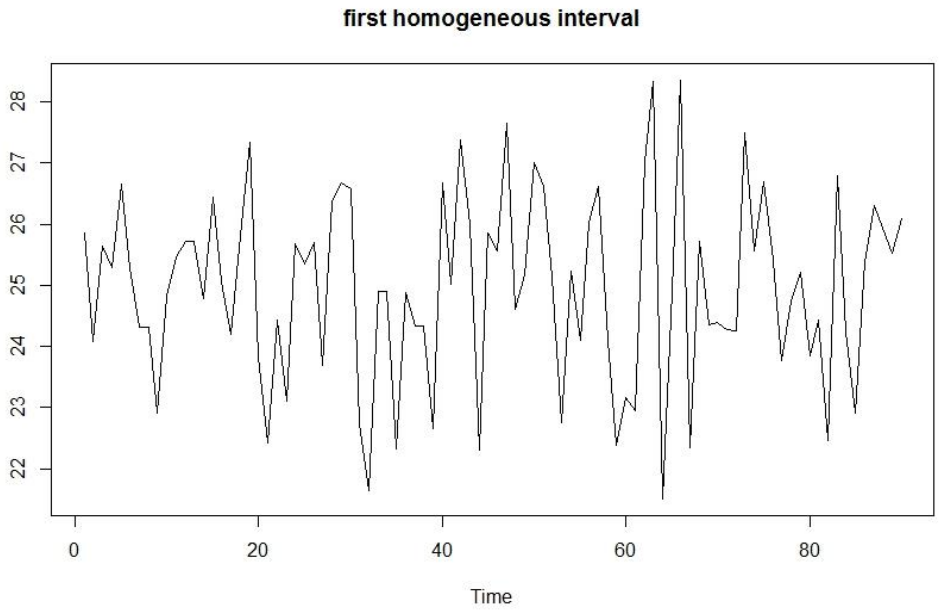
$$t_5 = t_4 + n_5 = 400.$$

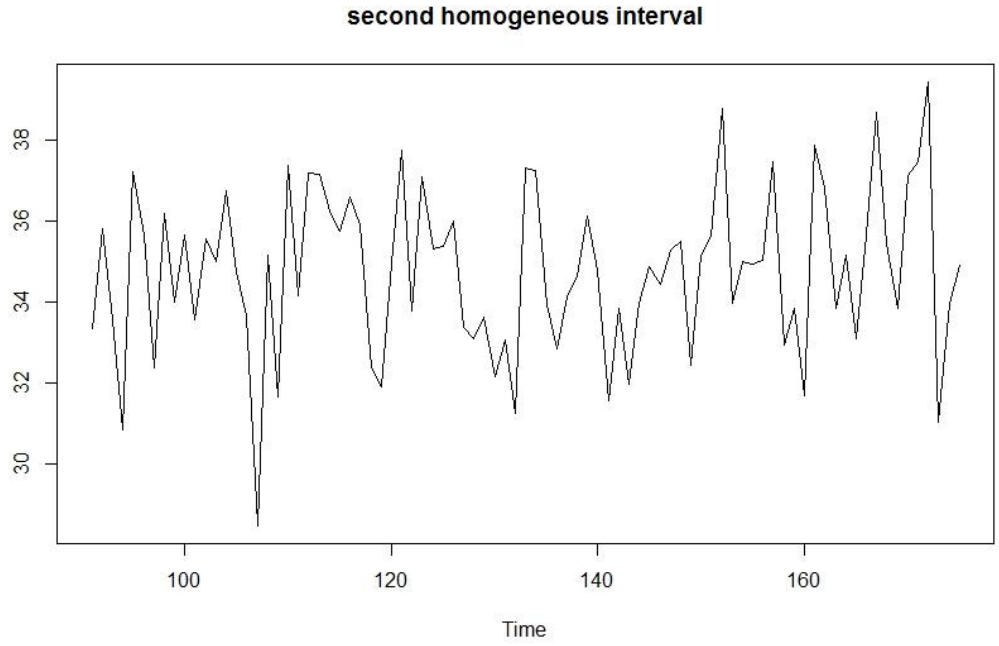
t_i is an i break-points with i =1,2…5.

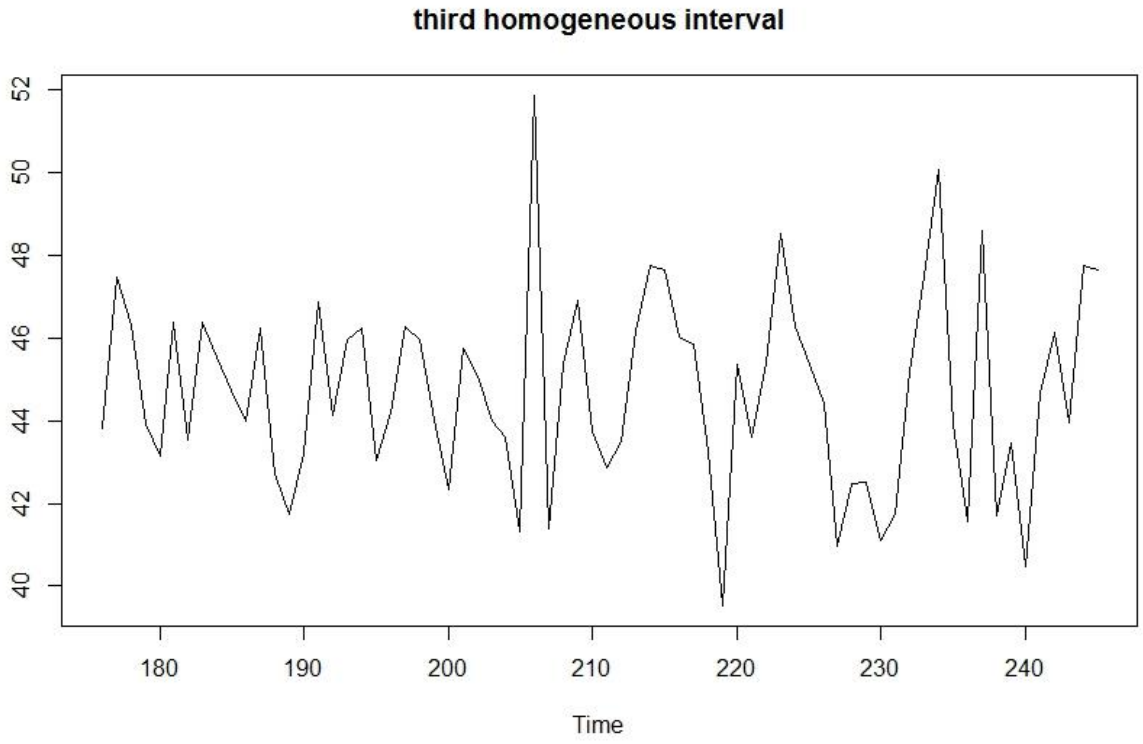
n_i is the number of observations in i interval, with $n_i = t_{i+1} - t_i$, for middle intervals,

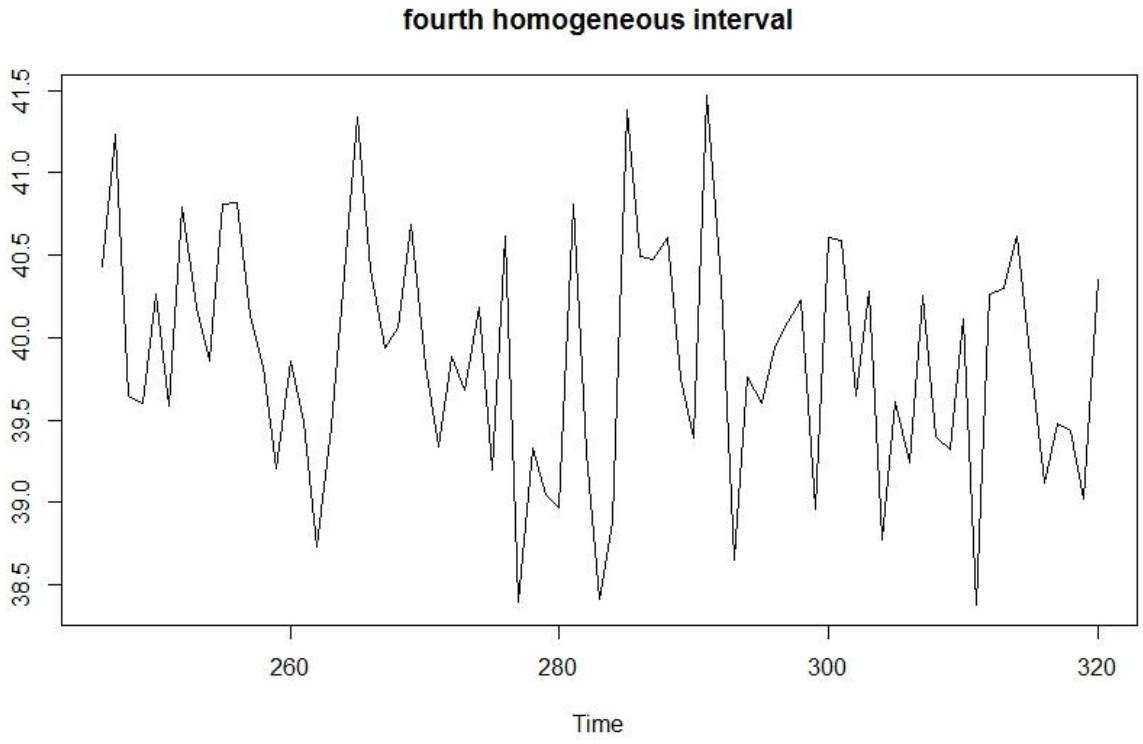
n_6 = T - t_5 for the last interval; T the total number of observations; T = 500.

Graphically, our simulated data looks like this for each interval:


first homogeneous interval
22
23
24
25
26
27
28
0
20
40
60
80
Time


second homogeneous interval
30
32
34
36
38
100
120
140
160
Time


third homogeneous interval
52
50
48
46
44
42
40
180
190
200
210
220
230
240
Time


fourth homogeneous interval
41.5
41.0
40.5
40.0
39.5
39.0
38.5
260
280
300
320
Time

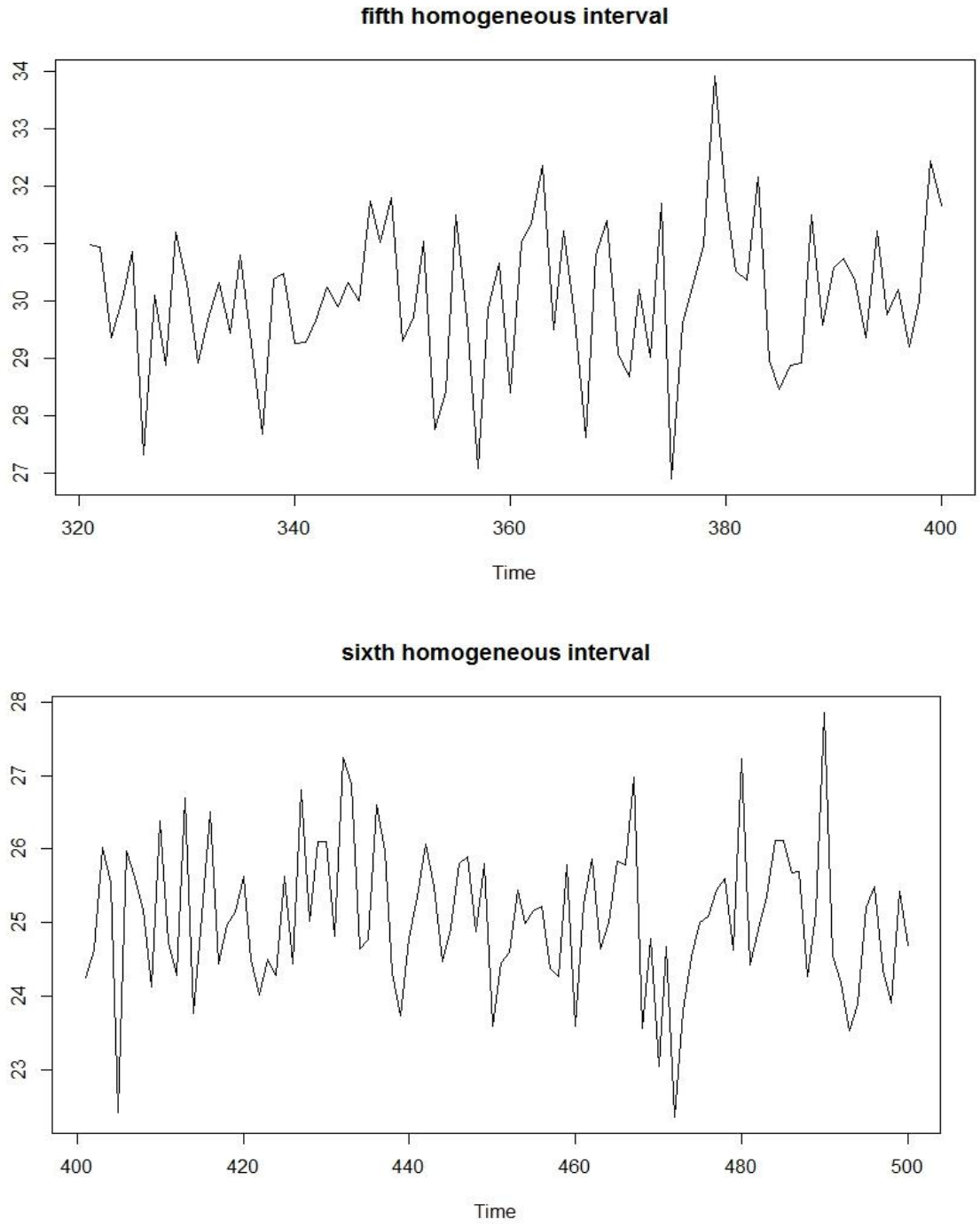


Figure1: plots of every local stationary time series

The plots show that the series is stationary for each homogeneous interval, and the autocorrelation function (ACF) shows that the observations are independent on each interval. Moreover, there is no slow decay in the ACF, which confirms that there is no sign of non-stationarity over these homogeneous intervals. Next, let us look at the entire series consisting of the six stationary small series and their ACF.

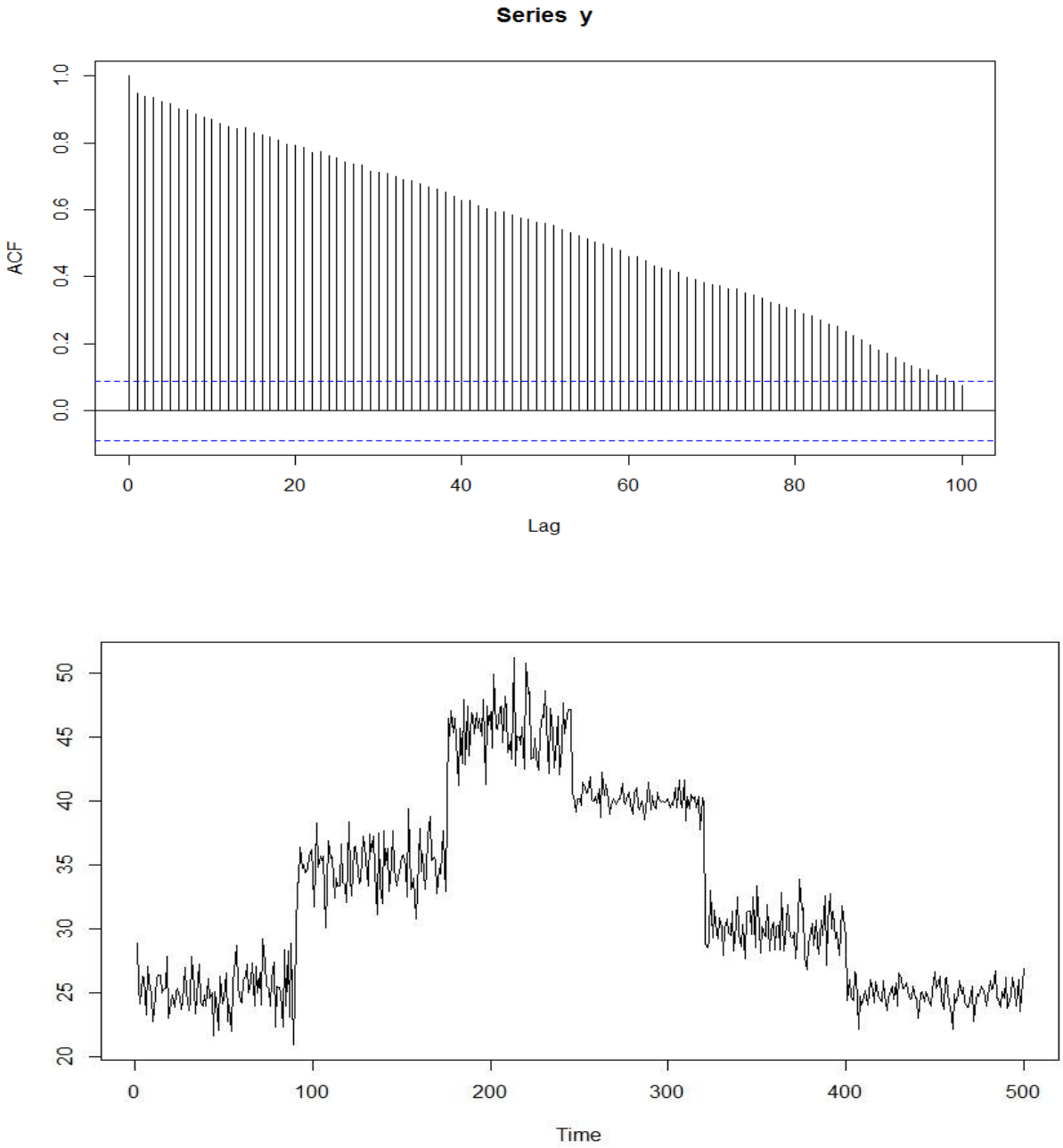


Figure2: A non-stationary time series composed by local stationary time series.

Above is the plot of the series, below is its autocorrelation function. The first graph in figure2, shows different structural breraks; it illustrates how the series is composed by small chunks that have their own specific distributions, and it clearly shows where the break-point on the graph is. From the graph we observe that there is a change in mean and variance of the series at some points, but it is not always easy to tell exactly from the graph what time that happened, especially when the change is not very significant; that is why we need some methods that locate the exact point where the structure of the series changed.

The second graph in figure2 illustrates the structure of an ACF that is a typical example of an acf of a non-stationary time series. This means that when local stationary series are combined together they may result in a nonstationary time series, as it is the case on the graph above.

Note: We know that the data was simulated randomly from a normal distribution; So, there is no assumption of linear dependence. As a result, the structure observed in the autocorrelation function is a result of structural break in the series.

In conclusion, it is generally known that the structure of financial time series may change at any time causing the series to have different structures during some periods of time. This means that the acf that decays slowly in financial time series should not be considered to be a result of linear dependence but it is caused by the change in the mean and variance of the series.

2. Monte Carlo simulation

Monte Carlo method or Monte Carlo simulation is a technique that is based on repeating random processes or other processes with uncertain outcomes to create a set of all possible outcomes of the experiment; generally, it is used to study the distribution of all outcomes (Harrison, 2010).

Monte Carlo simulation can be used in an optimization problem, forecasting, and other problems dealing with risk analysis. The most important feature of Monte Carlo simulation is that it provides not only possible outcomes of the experiment, but it also provides the likelihood of each outcome. This is very important in stochastic optimization problems where the output of the algorithm depends in an non-deterministic way on the starting value.

In Risk analysis, Monte Carlo method provides possible results from your decision, and their respective likelihoods so that decision makers may consider the risks they have before making certain decisions.

This is the same idea in forecasting; since we are estimating an uncertain value by using the historical values or based on the experience, to predict the future, there is some uncertainty but the model does not tell us how likely our estimate is. So we use Monte Carlo to provide a range of estimates with their likelihoods so that we may choose the one that is more likely, or just to know how likely our estimate is. There are many other ways Monte Carlo simulation may be applied; these were just some examples.

So, this will help us build a frequency distribution of break-points to see how many times the minimum BIC corresponds to the true model, or to calculate the expected number of breaks and compare it to the true model. Recall that in simulation study we know the true model, the true number and position of the breaks-points.

In addition, from Monte Carlo outcomes we can figure out how frequent the true model is obtained, which tells us about the performance of our methodology. For a better assessment of our methodology, we performed Monte Carlo simulation study using different sample size. Furthermore, we can explore the distribution of BIC's to see how frequent the model with breaks gives us smaller BIC compared to the model without breaks.

3. Results of Monte Carlo method on simulated data

In these results, we want to see how likely we are to locate the true break-points from the true model we defined above.

Remark: In our simulation study, we have realized that in the optimization process using an initial vector with one break-point generally gives the smallest BIC compared to the BIC obtained by using an initial vector with a different number of break-points. This initial break-point can be at any reasonable location of a break-point as defined earlier.

Table3: Monte Carlo results from simulation study with T = 500

Number of breaks	Set of breaks	frequencies	percent
5	90 175 245 320 400	963	96.3%
6	90 175 245 320 331 400	16	1.6%
4	90 175 245 320	12	1.2%
1	6,7,8	4 2 3	.9%

Since the true model was t_1 =90, t_2 = 175, t_3 = 245, 4_1 = 320, t_5 = 400, the above results shows that 96.3% of the time we can estimate correctly the true model. Also, with the true number of break-points being five, the expected number of break-points obtained by our method is 4.968, which is not really different from five we have in the true model.

For T =300 and n= 400 we have the following results respectively:

True model: t_1 = 50 , t_2 = 115, t_3 = 165, t_4 = 200, t_5 = 240

Table4: Monte Carlo results from simulation study with T = 300

Number of breaks	Set of break-point	frequencies	%
1	6,7,8	15	1.5%
6	50 115 165 168 200 240	3	0.3%
7	50 115 165 168 200 213 240	535	53.5%
8	50 115 165 168 200 213 238 240 50 115 165 168 200 214 238 240 50 115 165 168 200 220 238 240	227 188 27	44.2%
9	50 115 165 168 200 214 236 238 240	3	0.3%

10	50 66 68 115 165 168 200 213 238 240	2	0.2%

For n = 300, the expected number of break-points is 7.361

Table5: Monte Carlo results from simulation study with T= 400

Number of breaks	Set of breaks	frequencies	percent
1	6,8,10	7	.7%
4	80 165 235 300	3	.3%
5	80 165 235 300 340 (true model)	214	21.4%
6	80 103 107 165 235 300	18	1.8%
7	80 103 107 165 235 300 340	570	57%
8	42 45 80 103 107 165 235 300	17	1.7%
9	42 45 80 103 107 165 235 300 340	165	16.5%
11	42 45 75 77 80 103 107 165 235 300 340	6	6%

For T = 400, the expected number of break-points is 6.829.

Note: in the above three tables the frequencies represent how many times each of those set of break-points was found to be the one that minimize the BIC. Our purpose is to see how likely each set of break-points is to minimize the BIC.

Remarks: When keeping fixed means and variances but changing the number of data points in our simulation study, we realize that for samples of a size less than 500, The BIC is likely to choose the model with more break-points. For example with T = 300, T =400 we have the expected number of break-points 7.361 and 6.829 respectively. But with $T \geq 500$, the expected number of break-points is very close to five.

Note that in all of the above cases, when the series is broken into homogeneous intervals the BIC is smaller than the one for the model that ignores the structural changes, and considers the entire series to be stationary.

From the above results, we can deduce that:

1. For a small sample size, the BIC is likely to select a model with more breaks; it prefers the model with more parameters. In this case, we can correctly estimate the time when there is a change in mean, but with additional change points that we did not have in the true model.

2. With the sample of five hundreds and above, we can correctly estimate the true locations of the break-points without other additional break-points at more than 90%.
3. With $T \geq 500$, the expected number of parameters, break-points, is most of the time the same as the true model or almost the same; which means that with $T \geq 500$, this method is efficient in estimating the structural break in a time series. So, we can use it to estimate the time when there has been a change in mean and variance of a time series.
4. According to the above percentages, with $T \geq 500$ more that 85% we have obtained exactly the true model, and notice that with five breaks, the expected number of breaks is 4.968 when $T = 500$ and 5.075 when $T = 1000$. From the above results, we have evidences that this method does a great job in estimating the time where there is a structural change in a time series.

III. Applying the methodology to the real data.

The real data used in this paper is the S&P 500 returns given as percentages, collected from 500 leading companies in leading industries in the economy of the United States of America. In this paper we are using a time series data from 2006-11-03 to 2016-10-18 containing 520 observations that are weekly data from Monday to Friday. The following is the plot of the data:

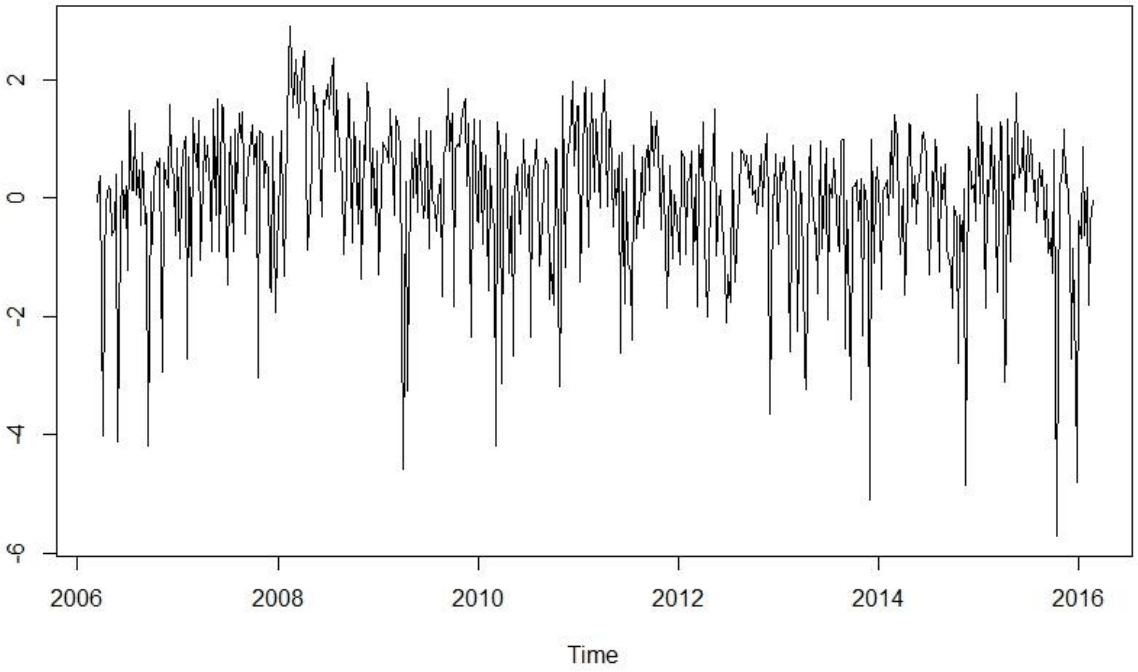


Figure3: plots of the log of absolute return

This figure shows that at some points, the structure of the series might have changed; either in the mean or in variance. So, we want to know how many times these changes have occurred and when it occurred by using our method, which consist of simulated

annealing optimization to find the minimum BIC and estimate the parameters. The following table illustrate the results:

Table6: Summary of the real data results

Number of breaks	Set of breaks	BIC	frequencies
1	More than 50 positions	784.25≤ BIC ≤ 787.8128	10.4%
2	303 491	785.5484	8.6%
3	100 109 159	759.5484	66.9%
4	100 109 159 500	755.3354	9.9%
5	100 109 159 477 494 100 109 159 334 350	757.8098 755.1888	4.2%

Note: for the real data we repeated the process a thousand times by changing the initial starting point to see how simulated annealing performs starting at different locations. Recall that in our simulation study we have realized that the initial vector with one break-point gives a better solution than starting with many initial break-points. So, to find the results we have above for the real data, every time we started with one random initial break –point and repeated the process a thousand times.

Remark: As discussed above, using an initial vector with one break-point generally gives better results compared to the results obtained by using initial vector with a different number of break-points. To obtain the above results in table5, we tried different initial positions to see which one gives the lowest BIC.

Looking at the above table, we realize that the smallest BIC is 755.1888, which tell us that based on the values of BIC, the best model among all of the above candidates is the one with the following five break-points: $t_1 = 100$, $t_2 = 109$, $t_3 = 159$, $t_4 = 334$, $t_5 = 350$, since it is the one that gives the lowest BIC. Notice that if we ignore the change points in the model, and calculate the BIC, the BIC is always bigger than what we get with the model divided into locally stationary models. For this data, the BIC for the model with no break is 787.8128, which is significantly bigger than the model with two to five breaks.

We can summarize our model in the following table:

Table7: Explicit real data results

Homogeneous interval	Homogeneous interval With the corresponding time	Number of observations	mean	variance
1 to 100	2006-11-03 to 2008-09-26	100	0.03339238	1.518884
101 to 109	2008-10-03 to 2008-11-28	9	2.08091	0.2323811

110 to 159	2008-12-05 to 2009-11-13	50	0.7478358	0.8301838
160 to 334	2009-11-20 to 2013-03-22	175	-0.01794629	1.539549
335 to 350	2013-03-29 to 2013-07-12	16	0.352495	0.182704
351 to 520	2013-07-19 to 2016-10-14	170	-0.3011363	1.785894

The purpose of the bove table is to show how the mean and/or variance changes from one interval to an other; this can be seen by looking at the last two columns of this table. The results show that there have been a significant change in mean everytime we move from interval to another. Again, we can see that the variance has also changed even if the cahnge is not as big as the one occurred in mean.

V.I. Conclusion and direction on future researches

1. Conclusion

The main goal of this paper was to propose a new method to locate the break-points in a time series by using simulated annealing optimization process to find a set of break-points corresponding to the minimum BIC.

We observed from our simulation study that with big samples, $T \geq 500$, about 90% of time we can correctly locate the exact break-points, the true model. When we get a different set of break-points than the true model, it's most likely the true model with some additional breaks, or almost all of the break-points from the true model without one or two of them. With this big sample size, almost all of the time, the BIC is likely to choose the true model as the best model over the other models with less or more breaks.

With a sample size less than 500, the BIC is biased; it is likely to choose a model with more break-points. So in this case, the BIC is smaller for models with more break-points. So, in general, when the size of the data is less than 500, BIC is not doing a good job as a model selection criterion. Hence, for small data sets it is not advisable to use BIC as a model selection criterion.

On the other hand, if you are told in advance that the series has a certain number of break-points, you can estimate these change-points by only using the maximum likelihood estimation but this works only with small samples with few breaks because it is when we can explore all possible positions of these break-points. So, since in reality we do not know how many breaks we have, in this paper we used simulated annealing to give opportunity to any number of break-points to be a candidate to our solution, but it's not the only way to find the minimum BIC.

The results we got from the real data, S&P500 weekly data on price index return, has shown that we can actually locate where the change in mean and /or in variance occurred. The results we have from table 5 show that the structure of the series has significantly changed over time especially in variance. Notice that even if by looking at the plot you may see some other times where the structure may seem to have changed, but if the change is not significant it's not likely to be identified as a break-point.

2. Directions on future research.

Recall that the main purpose of this paper was to find the time when the structure of the series changes, which allows us to divide the series that is not stationary into smaller stationary series. This was done by minimizing the BIC. The issue with this is that the function we are minimizing, BIC, tends to produce models that include also spurious break-points, at least for small sample sizes.

Notice that in this method AIC is worse than BIC, since it has a smaller penalty term in terms of parameters, so it will choose the model with more parameters than BIC does. According to all of our results, more work is needed to find a model selection criterion that works for any size of the data. For future work, our suggestions to solve this problem of model selection is assessing the performance of these different models by comparing their predictive accuracy or use other model selection criterion such as Bayesian model selection. Bayesian model selection will give us the evidence from the data about a given number of break-points; in other words, we want to calculate posterior model probabilities: the probability of each model given the data, so we can choose the model that is more likely based on the data and not the one corresponding to the minimum BIC. In addition, we want to compare the model with no break-points, the non-stationary model, with the locally stationary model considering the structural change in the series, but this comparison cannot be accurate by using BIC because it is in favor of the locally stationary model, the one with more parameters.

References

1. Aue, A., & Hovath, L. (2011).Structural break in time series. Retrieved from http://www.wileyonlinelibrary.com

2. Bertsimas, D. & Tsitsiklis, J. (1993). Simulated Annealing. *Statistical Science,* vol. 8, No. 1, 10-15.

3. Harrison, R. L. (2010). Introduction to Monte Carlo Simulation. *US National library of Medicine.* 1204: 17–21. doi: 10.1063/1.3295638

4. Kitagawa, G., & Akaike, H. (1978). A procedure for the modeling of non-stationary time series. *Ann. Ins. Statist. Math, 20(1978),* Part B, 351-363.

5. Mikosch, T., & Starica, C. (2004). Nonstationarity in Financial Time Series, the Long-Range Dependence, and the IGARCH Effects. *The review of economics and Statistics,* 86(1): 378-390.

6. Ozaki, T.,&Tong, W.(1975). On the Fitting of Non-Stationary Autoregressive Models in Time series Analysis. Retrieved from http://www.researchgate.net/publication/30526390

7. Poole, D.,& Mackwort, A. (2010). *Artificial intelligence: Foundations of computational* agents. Cambridge University Press, New York

8. Starica, C., & Granger,C. (2005). Nonstationarity in Stock return. *The review of economics and Statistics, 87(3):503-522.*