INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Structural Breaks: Testing and Applications in Finance

Adriana Lourenço Calado Gomes Santo

Master in Finance

Supervisor:
PhD José Joaquim Dias Curto, Associate Professor,
ISCTE-IUL Business School, Department of Quantitative Methods for Management and Economics (IBS)

October, 2023

BUSINESS
SCHOOL

Department of Finance

Structural Breaks: Testing and Applications in Finance

Adriana Lourenço Calado Gomes Santo

Master in Finance

Supervisor:
PhD José Joaquim Dias Curto, Associate Professor,
ISCTE-IUL Business School, Department of Quantitative Methods for Management and Economics (IBS)

October, 2023

Para os meus pais, Anabela e Jaime, por tudo o que a eles devo.
Para o Rui, por ser o meu grande apoio.

Resumo

Uma quebra de estrutura é uma mudança inesperada ao longo do tempo nos parâmetros estimados de modelos de regressão em séries temporais. Em Econometria, quebras estruturais num modelo de previsão podem invalidar testes de significância convencionais, levando a uma má performance nas previsões e à falta de confiança no modelo em geral. Em primeiro lugar, o principal objetivo desta dissertação é revisitar os desenvolvimentos feitos nos últimos anos no que diz respeito aos testes desenvolvidos para a existência de quebras estruturais, e como os resultados dos testes de estacionariedade (raiz unitária) podem ser comprometidos na presença de uma quebra estrutural. Os testes serão listados e descritos de acordo com as necessidades do modelo em termos de uso e aplicação.

Em segundo lugar, foi desenvolvido um modelo ARMA com uma quebra estrutural forçada para investigar como os testes de raiz unitária se comportam na presença de uma quebra estrutural, de forma a antecipar como os testes deveriam se comportar com dados reais.

Por fim, cada teste será realizado para detectar quebras estruturais e avaliar a estacionaridade usando dados reais de séries temporais.

Classificação JEL:

C58, G15

Palavras-Chave: Quebra de estrutura, Séries temporais, Estacionariedade, Modelo ARMA, Raíz unitária, Econometria

Abstract

A structural break is an unexpected change over time in the estimated parameters of regression models in time series. In Econometrics, structural breaks in an estimated forecasting model can invalidate conventional significance testing, leading to lousy forecasting performance and unreliability of the model in general. Firstly, the primary purpose of this dissertation is to review developments made in the last years as they relate to testing models on the existence of structural breaks, and how the results for unit root's testing can be compromised in the presence of a structural break. The tests will be listed and described according to model needs in terms of use and application.

Secondly, an ARMA model was designed with a forced structural break, to address how unit root's tests perform in the presence of a structural break, as an anticipation on how the tests should perform.

Lastly, each test will be performed to detect structural breaks and access stationarity using real time-series data.

JEL Classification:

C58, G15

Key-words: Structural break, Time series, Stationarity, Regression model, ARMA model, Unit root, Econometrics

Contents

Resumo iii

Abstract v

Chapter 1. Introduction 1

Chapter 2. Literature review 3
2.1. Unit root's tests 4

Chapter 3. Methodology 7

Chapter 4. Tests for Structural Breaks 11
4.1. With known breaking points 11
4.2. Without known breaking points 13
4.3. Structural breaks in unconditional variance 20

Chapter 5. Results and discussion - ARMA model simulation 25
5.1. ARMA model simulation results 25
5.2. Discussion 26

Chapter 6. Results and discussion - Tests on real financial data 29
6.1. Allianz 29
6.1.1. Unit root's tests 30
6.1.2. Structural breaks tests 31
6.2. CAC Index 40
6.2.1. Unit root's tests 40
6.2.2. Structural breaks' tests 41
6.3. FRED Employment level 46
6.3.1. Unit root's tests 46
6.3.2. Structural breaks' tests 47

Chapter 7. Conclusion 53

References 55

Annex A 59

Annex B 69

CHAPTER 1

Introduction

The study of time-series data is crucial in many fields, particularly Finance, where understanding trends and patterns in financial data can inform investment decisions and Risk Management strategies. A critical aspect of time-series analysis is the concept of stationarity, which refers to the property of a time series where its statistical properties, as primarily mean and variance and autocorrelation, remain constant over time. However, much real-world time series, particularly in Finance, exhibit changes in their underlying statistical properties, known as structural breaks.

Structural breaks can occur for various reasons, such as changes in economic policy, shifts in consumer behaviour, or the introduction of new technologies. These breaks can significantly impact the stationarity of a time series and lead to misleading conclusions if not adequately accounted for in the analysis.

Despite the importance of structural breaks in time-series analysis, much is still not understood about their behaviour and impact on stationarity. This thesis aims to contribute to the existing literature by studying the effect of structural breaks on time-series stationarity and exploring their applications in Finance.

This thesis aims to answer the following questions: How do structural breaks affect time-series stationarity, and what are the most appropriate methods to detect them? Can the existence of structural breaks lead to misconceptions regarding stationarity of time series?

The objectives are as follows:

- List the main tests available in literature regarding structural breaks and the most commonly used to address stationarity of time series (through unit root testing).
- Simulate an ARMA model with a fixed structural break, and through multiple iterations assess whether the unit root tests can correctly identify the series starionarity or if they produce conflicting conclusions.
- Collect and test real financial time-series data on series' stationarity and existence of structural breaks.
- Compare the performance of different methods for detecting structural breaks and determine the most appropriate methods.

Specifically, this research methodology investigates the most commonly known methods for detecting structural breaks in financial time series as well as some more recently developed tests and the impact of these breaks on stationarity detection through unit root detection tests.

The study of structural breaks is a broad area with a wealth of literature. To identify changes in the mean and variance of time-series data, a variety of techniques and tests have been created throughout the years. The main goal in this thesis is to give a thorough overview of the most popular and current methods for spotting structural breaks in both mean and variance. The wide variety of approaches that are accessible might be intimidating, thus the goal is to condense this knowledge for clarity and real-world use. To help scholars and practitioners better grasp these crucial time-series analysis techniques, the most important tests currently in use will be discussed and identified in the chapters that follow.

The remainder of this paper is organized as follows: Chapter 2 begins the literature review, providing a brief introduction to the topic and mentioning the unit root tests that will be applied. In Chapter 3, the methodology applied to both the ARMA simulation model and the analysis conducted on real financial data is explained. Chapter 4 continues the literature review by describing each of the tests used for detecting structural breaks. Chapter 5 presents the results and initiates the discussion on the ARMA simulation model, while Chapter 6 concludes the discussion, focusing this time on the testing of real financial data.

Additionally, in Annex A, the R code is extensively commented upon regarding the implementation of tests and the ARMA simulation model. In Annex B, results related to two other financial time series that could not be accommodated within the appropriate length of the thesis can be found. These results are included because the tests were conducted, and the writer deems them to be relevant.

CHAPTER 2

Literature review

Time-series modeling is an evolving field that has captured the interest of researchers in the recent decades. Its main goal is to analyse past observations of a series and develop a suitable model to understand its underlying structure. This model is then utilized to forecast future values, allowing users to predict the future based on understanding of the past. Its importance has been noted in fields such as Finance, Science, Engineering and even Health (G. E. Box, Jenkins, Reinsel, & Ljung, 2015).

A time series is influenced by four main components that can be distinguished from the observed data: trend, cyclical, seasonal, and random components.

The trend component refers to the general tendency of a time series to increase, decrease, or remain stable over a long period of time, representing the long-term movement.

The seasonal component refers to fluctuations that occur within a year, typically associated with specific seasons. Factors such as climate, weather conditions, customs, and traditions contribute to the seasonal variations, a component with high value for businesses to make informed future plans.

The cyclical variation refers to medium-term changes in a time series that occur in repetitive cycles. These cycles usually span two or more years and are influenced by recurring circumstances. Economic and financial time series often exhibit cyclical patterns. A typical example is the business cycle, which consists of four phases.

Finally, the random component in a time series results from unpredictable influences that do not follow a specific pattern. These variations can be caused by events like wars, strikes, earthquakes, floods, or revolutions. There is still no defined statistical technique for measuring random fluctuations in a time series (Adhikari & Agrawal, 2013).

Stationarity is the foundation of time-series analysis, and it basically states that the probability principles governing a process' behavior do not vary over time. This is done to make sure that the time-series process is statistically in equilibrium, which would improve the statistical environment for describing and drawing conclusions about the structure of data that vary in some unpredictably ways (Shumway & Stoffer, 2011; Johnson, 2009). A process is deemed strictly stationary if the entire probability structure is forced to depend solely on time differences (G. Box, Jenkins, & Reinsel, 2008).

A less stringent criteria known as weak stationarity of order k states that moments up to a certain order k, or even to different lags of k, rely exclusively on time delays and that rigorous stationarity may be produced using second order stationarity alone and the normality assumption (Tsay, 2010). A time series is considered stable for ease of use if its mean, variance, as auto-covariance function remains constant across time (Pankratz,

2008). The white noise process, which is defined as a series of independent (uncorrelated) and identically distributed random variables with a zero mean and constant variance (Wei, 2006), is one of the most significant and fundamental examples of a stationary process, often being used as a baseline or null model in time-series analysis.

Essentially, a time series is said to be stationary when its characteristics are not dependant on time in which data was observed (Cerqueira, Torgo, & Mozetič, 2020), hence having a constant mean, variance, and covariance.

As stated, linear correlations between dependent and explanatory variables hold significant importance. To explore these concepts, researchers gather observations across time for one or more cross-sectional units, such as company performance, indices, or macroeconomic information regarding nations. By using this data, regression model coefficients can be estimated. However, there is typically one critical assumption in this process - the coefficients remain constant over time.

This assumption becomes problematic, especially over longer periods, due to significant disruptive events like financial crises, as the Global Financial Crisis in 2008, or as more recently the 2020 COVID-19 pandemic. Such events can cause parameter instability, meaning the coefficients no longer hold steady. As a result, estimation and inference can be negatively affected, leading to costly mistakes in decision-making.

Typically, statisticians refer to the periods when these parameters change as "change points", while economists term them as "structural breaks". These changes in the coefficients can significantly impact the validity of the regression model, and thus require careful consideration under all analysis (Ditzen, Karavias, & Westerlund, 2021).

Hence, structural breaks refer to significant changes in the underlying statistical properties of a financial time series. These changes can be caused by various internal and external factors, such as economic policies, market regulations, technological innovations, and natural disasters. The identification and analysis of structural breaks are important for various applications in Finance, including Risk Management, Asset Allocation, and Forecasting.

2.1. Unit root's tests

Furthermore, other concepts that are relevant to introduce at this point are unit root and unit root's tests. The concepts of structural breaks and unit root are related in the field of time-series analysis and Econometrics, particularly when dealing with non-stationary time-series data.

A unit root is a statistical property of a time series that indicates non-stationarity. As already stated, a non-stationary time series is one where the mean or variance is not constant over time. It implies that the time series has a stochastic trend and is sensitive to shocks, which can lead to persistent deviations from the mean. This concept is the basis for several tests for time-series stationarity.

In most cases, limiting distributions defined as functions of Brownian movements are used to simulate critical values for unit root testing. The simplest example is a random

walk, such as $x_t = x_{t-1} + \varepsilon_t$, where ε_t are random disturbances. In Finance, a typical representation of random walks is when the logarithm of stock prices is modeled as a random walk: $\log(S_t) = \log(S_{t-1}) + \varepsilon_t$, which is equivalent to modeling log returns as a stationary process: $\log(\frac{St}{S_{t-1}}) = \varepsilon_t$.

Assuming that the disturbances ε_t are a white noise, independent and identically distributed with $E(\varepsilon_t) = 0$ and $\text{var}(\varepsilon_t) = \sigma^2$, it can be deduced that the process x_t has a constant mean ($E(x_t) = x_0$) and a time-dependent variance ($\text{var}(x_t) = t\sigma^2$). The time-dependent variance implies that the spread of the values in the process increases with time. While the mean remains constant, the changing variance indicates a lack of stationarity as the variability in the data grows over time (Herranz, 2017).

The interaction between structural changes and unit root has also received a lot of attention in the literature, especially in light of the fact that both kinds of processes share certain qualitative characteristics. For instance, when the genuine process is vulnerable to structural changes but is otherwise (trend) stationary within regimes defined by the break dates, the majority of tests that seek to discriminate between a unit root and a (trend) stationary process will prefer the unit root model. Additionally, when the process has a unit root component but constant model parameters, the majority of tests used to determine if structural change is present will reject the null hypothesis of no structural change (Perron, 2005).

The literature review shall continue in the Chapter 4, as it served as foundation for the writing of all the tests used in this paper, to serve the purpose of listing the main tests available in the literature in a more reader-friendly way.

CHAPTER 3

Methodology

The methodology used in this thesis is divided into three parts: (i) the gathering of information spread in the literature regarding tests to detect structural breaks in mean and variance, as well as the most common tests to assess whether the time-series parameters can be said to be stationary, (ii) a model simulation that consists of an ARMA with a break and the application of said tests to the simulated model, to anticipate the results, (iii) and the analysis of real time-series data.

The traditional structural break detection methods, such as the Chow test, CUSUM, MOSUM, Quandt-Andrews and the Bai-Perron test were applied to the simulated series using the package "*strucchange*" in R to study changes in the mean of the parameters. For the change in variance, Inclan and Tiao test was performed using the "$ICSS$" package. In addition to these traditional methods, more recent structural break detection methods, such as the CPT test, with the Binary Segmentation and PELT search algorithms, were also applied to the data series using the package "*changepoint*" in R. These packages may be obtained from the Comprehensive R Archive Network (CRAN) at http://cran.r-project.org/. More details regarding these tests can be found in Chapter 4.

The second part of the methodology involved the simulation study that was conducted using an ARMA model with 5000 datapoints, which was simulated to include a structural break at point 2000. This simulated series was used as a benchmark to test various methods for detecting the stationarity of the series in the presence of a known structural break.

To evaluate the stationarity of the time-series, unit root tests such as the Augmented Dickey-Fuller (ADF), Phillips-Perron (PP), Kwiatkowski-Phillips-Schmidt-Shin (KPSS), Elliott-Rothenberg-Stock (ERS) and Zivot-Andrews (ZA) tests were performed. They where performed by recurring to the "*urca*" package, available in R. The results of the simulation study were compared and discussed. Conclusions were drawn regarding the efficiency of different methods for detecting stationarity in the presence of structural breaks.

For the purpose of conduction the ARMA model simulation, an R code was written to define a function that generates a time series of synthetic financial returns with specified characteristics. It can be found described in Annex A. The function takes the following parameters:

- `num_observations`: Number of observations in the generated time series.
- `mean_min` and `mean_max`: Minimum and maximum values for the mean of the returns.

- `sigma_min` and `sigma_max`: Minimum and maximum values for the standard deviation of the returns.
- `break_point`: The observation number at which a structural break occurs, changing the mean and standard deviation of the returns.

In a step-by-step approach, the basis behind the function is as follows:

(1) ARMA model generation: It generates an ARMA(1,1) time-series model with specific coefficients (`ar_coefs` and `ma_coefs`) and a fixed standard deviation of 0.1796 (arbitrary value).

(2) Randomly setting means and standard deviations: It randomly selects two means (`mean1` and `mean2`) and two standard deviations (`sigma1` and `sigma2`) within specified ranges.

(3) Generating returns: Using the ARMA model, it generates synthetic returns. Before the `break_point`, the returns have mean `mean1` and standard deviation `sigma1`, and after the `break_point`, the returns have mean `mean2` and standard deviation `sigma2`.

The function is called with specific parameter values to generate a time series with 5000 observations, where the means and standard deviations change at observation 2000.

To perform the unit root's tests, the following logic was applied:

- Package installation and loading: The code starts by installing and loading the `urca` package, which provides functions for unit root tests.
- The code runs a loop `n` times (specified as 1.000 times in this case) to generate synthetic financial return data and conduct unit root tests on each generated dataset.
- For each iteration of the loop, synthetic financial return data is generated using the `Test_generate` function. This function creates time-series data with different specified means, standard deviations, and a structural break at observation 2000.
- Five different unit root tests are performed on the generated data.

For the ADF test: `ur.df` function is used with drift term and lag selection based on Bayesian Information Criterion (BIC). If the test statistic is less than the critical value (tau3), the data is considered non-stationary (`resultdf = 0`); otherwise, it is stationary (`resultdf = 1`).

For the PP test: `ur.pp` function is used with constant term and short lag specification. If the test statistic is less than the critical value (`c.val`), the data is non-stationary (`resultpp = 0`); otherwise, it is stationary (`resultpp = 1`).

For the KPSS test: `ur.kpss` function is used with the tau statistic and short lag specification. If the test statistic is greater than the critical value (`kpss@cval[1,2]`), the data is non-stationary (`resultkpss = 0`); otherwise, it is stationary (`resultkpss = 1`).

For the ERS test: `ur.ers` function is used. If the test statistic is less than the critical value (`ers@cval[1,2]`), the data is non-stationary (`resulters = 0`); otherwise, it is stationary (`resulters = 1`).

For the ZA test: `ur.za` function is used. If the test statistic falls bellow the critical value (`za@cval[2]`), the series is considered to be stationary (`resultza = 1`); otherwise is non stationary (`resultza = 0`).

The results of these tests (0 for non-stationary, 1 for stationary) are stored in the `unit_roots` data frame.

After conducting the unit root tests, the code calculates the proportions of stationary outcomes for each test and prints the results. These proportions are calculated by summing the corresponding columns in the `unit_roots` data frame and dividing the sums by the total number of iterations. The `cat` function concatenates these results into a single output string, which is then printed to the console or output file.

For each iteration, the code determines whether the generated time series is stationary or non-stationary according to each test. By calculating the proportions of stationary outcomes and displaying them, it can be assessed whether the effectiveness of these tests in correctly identifying stationary processes, a valuable process to understand if the performance of different unit root tests under specific conditions is correct, helping to evaluate the reliability of these tests and their ability to distinguish between stationary and non-stationary time-series data.

In summary, the code generates synthetic financial return data, applies different unit root tests, and records whether each dataset is stationary or non-stationary.

Finally, the third part consists of analysing real financial time-series data, including:

- Daily stock prices of companies Allianz and RollsRoyce (the data being retrieved from Yahoo Finance under tickers ALV.DE and RR.L, respectively), from 2003 until 2022.
- Daily index values for CAC 40 (the data being retrieved from Yahoo Finance under ticker FHCI), from 2003 until 2022.
- Monthly values for the Employment Level, thousands of persons, seasonally adjusted (the data being retrieved from FRED, Federal Reserve Bank of St. Louis, under ticker CE16OV), from 2003 until 2022.
- Monthly values for the Federal Funds Effective Rate, percent, not seasonally adjusted (the data being retrieved from FRED, Federal Reserve Bank of St. Louis, under ticker FEDFUNDS), from 2001 until 2022.

The results of the real data analysis were discussed, and conclusions were drawn regarding the efficiency of different methods for detecting structural breaks and their impact on stationarity assessment.

Overall, the methodology used in this thesis is designed to comprehensively evaluate different methods for detecting structural breaks and their impact on time-series stationarity, as well as understand the unit root's test accuracy when in the presence of a structural break.

CHAPTER 4

Tests for Structural Breaks

In this chapter, a detailed exploration of various statistical tests designed to identify structural breaks within time-series data will be described. The chapter is structured around tests that focus on detecting breaks in both the mean and unconditional variance of the data. For mean shifts, essential tests such as the Chow Breakpoint Test, Chow Forecast Test, CUSUM, MOSUM, Quandt Andrews, and Bai Perron tests will be covered. Additionally, tests for unconditional variance shifts, including ICSS and tests provided by the *changepoint* package utilizing Binary Segmentation and PELT algorithms will also be addressed.

4.1. With known breaking points

Chow Breakpoint test

Chow's tests are a widely used method for detecting structural breaks in time-series data, shown as the first tests for structural breaks in economic literature. It involves comparing the fit of a model with and without a structural break at a particular point in time and evaluating the significance of the difference, hence assuming known structural breaks (Chow, 1960). The tests are based on the assumption that the time series can be divided into two distinct subsamples, each with its own set of estimated parameters.

Consider the traditional linear regression model:

$$Y_t = \beta_1 + \beta_2 X_{2t} + ... + \beta_k X_{kt} - \varepsilon_t, \text{ equivalent to } Y = \beta X - \varepsilon \tag{4.1}$$

with k parameters and divided into two distinct subsamples, T_1 and T_2, with $T_1 + T_2 = T$ the total number of observations.

You may determine whether the regression coefficients are different for divided data sets using the Chow test. In essence, it examines which of two distinct regression models or one unique regression model best fits a split set of data. By doing this method, if the estimated parameters of both models are equal, then the subsample T_1 and T_2 can be expressed as single regression model, i.e., $\beta_1 = \beta_2 = \beta$.

Chow presented two tests: Chow Breakpoint test and Chow Forecast test. Under Chow Breakpoint test, the null hypothesis is that there is no structural break, meaning that the data set can be described by a single regression line. This test's methodology begins with the proposed initial partition of the data into two subsamples, if there is only one known breakpoint *a priori*. The number of partitions of the data, m, is dependent on the known breakpoints. If there is the suspicion of more than one, $m > 1$, there will be an equivalent number of subsamples of data. The estimation of the model, done for

each subsample of data, is followed by the assessment of whether the differences in each estimated coefficients β are statistically significant. By default, the test determines if there has been a structural change in each of the equation's parameters.

Chow Breakpoint test is based on the F-statistic that is derived from comparing the limited residuals sum of squares obtained by fitting one equation to the whole sample with the unconstrained residuals sum of squares ("RSS") obtained by fitting the equation to each subsample. The vector of residuals sum of squares is obtained by $RSS_m = \varepsilon'_m \varepsilon_m$, where $\varepsilon_m = Y_m - X_m \beta_m$, for each subsample, m, obtaining the F test computed as

$$F = \frac{[RSS - (RSS_1 + RSS_2)]/k}{(RSS_1 + RSS_2)/(T - 2K)} \sim F[k; T - 2k] \tag{4.2}$$

for $m = 2$ as is the case of the single breakpoint test proposed by Chow.

If the errors are independent, identically distributed normal random variables, then the F-statistic has an accurate finite sample F-distribution. According to the test's logic, if the coefficients are identical, then the total of RSS_1 and RSS_2 should equal the sum of the squared residuals from the full sample estimate RSS, and the F test should result in a value of zero. The fact that the assumption of equal coefficients will be rejected increases with increasing F value.

Chow Forecast test

The Chow Breakpoint test has a significant limitation in that each subsample needs at least the same number of observations as calculated parameters. If, for instance, there are fewer observations of a particular event in time and you wish to test for structural change between the longer, "normal period" and the particular event, this might be an issue. In these circumstances, the Chow Forecast test should be applied.

Two models are estimated using the Chow Forecast test, one utilizing the entire set of data T and the other using a lengthy subperiod T_1 and the shorter period T_2. The calculated relation's stability during the sample period is questioned by differences in the results for the two estimated models. Both least squares and two-stage least squares regressions can be used.

The Chow Forecast test is used to determine whether or not the model can forecast the dependent variable's most recent values. In that case, it is claimed that the parameters are stable concerning the break date.

The stability of the coefficients across the sample period is questioned by differences between the findings for the two data sets. When doing the Chow Forecast test, one must compute the following:

$$F = \frac{(RSS - RSS_1)/T_2)}{RSS_1/(T_1 - k)} \sim F[T_2; T_1 - k] \tag{4.3}$$

As previously, in the case where the errors are independent and identically distributed, this statistic follows an exact finite sample F-distribution.

The relationship between the returns of the data sample and the subsample appears stable if there is statistical evidence pointing towards no structural change in the coefficients before and after the breakpoint, as there is no statistical evidence from the subsample to conclude in favour of a break in the coefficients of the sample.

The Chow tests do have the substantial drawback of requiring a prior determination of the break date. A researcher only has two choices: either select a potential break point at random date or select a break date based on a well-known observable data feature. In the first instance, it is possible that the true break date was ignored, rendering the Chow tests possibly ineffective. The Chow tests may be misleading since they may falsely indicate the existence of a break when none actually exists, as the proposed break date in the second scenario is endogenous. Additionally, it is relatively simple for different researchers to get very different findings as the results may be considerably impacted by these arbitrary choices, which is hardly an example of sound scientific practice (Hansen, 2001).

4.2. Without known breaking points

CUSUM test

The CUSUM (Cumulative Sum) test is a non-parametric method for detecting changes in the mean of a time series. It involves computing the cumulative sum of the residuals from a model fit to the data and comparing it to a predetermined threshold. If the cumulative sum exceeds the threshold, it indicates that a structural break has occurred.

Furthermore, scenarios in which neither the amount nor the date of potential changes in the regression parameters are known are taken into account. The CUSUM test is based on the cumulative sum of recursive residuals, was proposed back in 1975 ((Brown, Durbin, & Evans, 1975)). Standardized one-step-ahead prediction mistakes are referred to as recursive residuals. Assume that there are T total observations in the data sample. When the regression is calculated using only the first $t-1$ data, the t^{th} recursive residual is the one-step-ahead prediction error for y_t. The parameter estimates are obtained based on $t-1$ observations and utilize them to forecast the dependent variable's subsequent observation.

Being x_t the regressor of the t observation and β_{t-1} the least squares coefficient computed with the first $t-1$ observations of the data sample, the residual is given by:

$$\varepsilon_{rt} = y_t - \beta_{(t-1)} x_t \tag{4.4}$$

where x_t stands for the t observation of the regressors' vector and $\beta_{(t-1)}$ stands for the least squares coefficients computed using the first $t-1$ observations.

The next step is to obtain the scaled residual, ω_t, given by:

$$\omega_t = \frac{\varepsilon_{rt}}{\sqrt{1 + x_t (X'_{t-1} X_{t-1})^{-1} x'_t}} \sim N(o, \sigma^2) \tag{4.5}$$

based on the assumption that the coefficients remain constant and that w_t is independent of w_k for all $t \neq s$

In this test, being based on a visual illustration of the model to identify structural breaks, two barriers shall be determined. By representing $W_t = \sum_{j=k+1}^{t} \frac{w_t}{\hat{\sigma}}$, $t = k+1...T$ as the cumulative sum of recursive residuals, the stability of the model can be represented as it staying within the the -2 and $+2$ barriers, meaning that a structural break is identified when such barrier is surpassed These barriers are given by $\pm 2\hat{\sigma}\sqrt{1 + x_t(X'_{t-1}X_{t-1})^{-1}x'_t}$.

In the case where the data available is perfectly fitted by the model, the sum of the forecasting errors should be very close to zero. The CUSUM test enables us to determine whether the total cumulated sum of forecast errors is statistically different from zero.

The CUSUM test is given by:

$$CUSUM = \max_{k+1<r<T} \left| \frac{\sum_{t=k+1}^{r} \omega_t}{\hat{\sigma}\sqrt{T-k}} \right| / (1 + 2\frac{r-k}{T-k}) \tag{4.6}$$

where $\hat{\sigma}$ is an estimate for the standard deviation of the errors ε.

The null hypothesis of the CUSUM test is of no structural change, translated into W_t having mean of zero and a variance equal to the sum of the number of residuals, as each has a variance equal to one and they are considered to be independent.

The CUSUM test shows that if there is just one structural change point at $t = r$, the recursive residuals will only have a zero mean up to that point. Departing from its mean after $t = r$, the process's path should thus be near to 0 up until then, with the coefficients being constant up until time $t = r$ and change after that. The scaled recursive residuals w_t will have a zero mean up until time $t = r$, but will typically have non zero means beyond that. Consequently, it is reasonable to assume that a plot of the CUSUM will reveal some information regarding prospective structural alterations.

W_t is plotted against t and represents the cumulative sum of recursive residuals. $E(W_t) = 0$ for constant parameters, however W_t will tend to deviate from the mean value line when the values are not constant. By making use of two lines that pass through the points and are symmetrically above and below the line $W_t = 0$, it is possible to determine the significance of the deviation from the zero line. Theses two lines can be defined by $\left[k, \pm a\sqrt{T-k}\right]$ and $\left[T, \pm 3a\sqrt{T-k}\right]$, a being the parameter that depends on the test's selected significance level, α. 90, 95, and 99 percent are represented by the values of α of 0,850, 0,948, and 1,143, respectively. If W_t deviates from the bounds, the null hypothesis should be rejected (at significance level α). If the empirical process route surpasses these boundaries, W_t is unreasonably large.

A nonzero mean of the recursive residuals caused by changes in the model parameters is what the CUSUM test is intended to find. If there are several parameter adjustments that might offset their effects on the means of the recursive residuals, the test may not be very powerful (Luetkepohl & Krätzig, 2004).

As stated, this test has some limitations, as it is also sensitive to the choice of reference value and threshold, and it may not be as effective at detecting breaks that occur gradually over time (Kleiber, 2016).

Despite these considerations, the CUSUM test is relatively easy to implement and interpret and is also sensitive to small changes in the data, which makes it useful for detecting subtle structural breaks, characteristics that allowed this test to have been adapted for use in statistical analysis for detecting structural breaks in data since its early development.

MOSUM test

Analyzing shifting sums of residuals is another way to spot a structural change. The resultant process does not include the sum of all residuals up to a certain time t, but rather the sum of a specified number of residuals in a data window, the size of which is set by the bandwidth parameter $h \in (0, 1)$ and spans the whole sample period (Zeileis, Leisch, Hornik, & Kleiber, 2002).

The definition of the moving sums of recursive residuals is given by:

$$M(j,h) = [\sigma[Th]]^{-1/2} \sum_{t=j+11}^{j+[Th]} w_t, \text{ t=j+1,..., j+[Th]} \tag{4.7}$$

With the statistic being defined as:

$$T_{rec} = \max_{1 \le j < T-[Th]} |M(j,h)| \tag{4.8}$$

For $j = 0, 1, \ldots, T - [Th]$, the jth moving sum of least-squares residual up until which it is calculated is given by

$$M_{\text{ls}}(j,h) = \frac{1}{[\hat{\sigma}[Th]]^{1/2}} \left(\sum_{n=j+1}^{[Th]} e_n \right),$$

Each moving sum in the MOSUM test has a set number of residuals, unlike cumulated sums, which have a greater number of residuals. The selection of bandwidth h for the MOSUM test is crucial. If h is big, there are not enough moving sums to identify potential changes since each one contains too many residuals. As a result, moving sums with a big h are less susceptible to parameter change. This being said, if h is small, the moving sums' sample variation is likely to be high and the limit distribution may not be a fair approximation (Kim, 2011).

The normal distribution applies to $M_{j,h}$ of MOSUM test's statistic under constancy of the parameters. It's distribution, however, is normal only in the asymptotic situation, and these moments still hold if σ is replaced by its estimate s, which is found to be the square root of the average of the squared recursive residuals.

The discrepancies of the MOSUM and its expectations will be non-systematic as long as the assumption of constancy is not broken, but after a structural break, systematic deviations will emerge. Therefore, the test for the presence of deviations may be used to identify parameter inconsistency (Hackl, 2016).

The MOSUM test was developed after the CUSUM test as a way to address some of the limitations of the CUSUM test and improve its performance. It has become a popular

tool for detecting structural breaks in various types of data. However, the MOSUM test also has some limitations.

For determining model stability, MOSUM tests allow for more reliable estimations than CUSUM tests and provide more information about the model (Zeileis et al., 2002).

Quandt-Andrews Breakpoint test

The Quandt-Andrews test works by comparing the variance of the residuals before and after the suspected break point in the data. To perform the test, the data is first divided into two segments at the suspected break point. A linear regression model is then fitted to each segment separately, and the residuals are calculated for each segment. Again, this is a test for a single structural break. The Quandt-Andrews test is a parametric test, meaning it assumes that the data follows a specific distribution (usually normal), as opposed to the Chow test, a non-parametric test (Quandt, 1960).

The null hypothesis for the test is that there is no structural break, as the variance of the residuals is the same in both segments.

The Chow Breakpoint test (Chow, 1960) was refined into Quandt-Andrews, who also proposed the Quandt likelihood ratio (QLR) test. By leaving out the initial and last 15% of the data, the QLR statistic performs better while analyzing the uncertain break date and evaluating the Chow Breakpoint test at each observation.

To begin with, a single Chow Breakpoint test is performed at every observation between two selected dates, t_1 and t_2. The $k = t_2 - t_1 + 1$ results from the Chow Breakpoint test are compiled into three different statistics: the Maximum Statistic (the max function of the individual Chow's F-statistic), the Exp Statistic, and the Ave Statistic (Andrews & Ploberger, 1994; Andrews, 1993). The three are defined as follows:

$$\text{Maximum Statistic, } MaxF = \max_{t_1 \leq t \leq t_2} \Big[F(t)\Big] \tag{4.9}$$

$$\text{Exp Statistic, } ExpF = \ln\left[\frac{1}{k}\sum_{t=t_1}^{t_2} \exp\left(\frac{1}{2}F(t)\right)\right] \tag{4.10}$$

$$\text{Ave Statistic, } AveF = \frac{1}{k}\sum_{t=t_1}^{t_2} F(t) \tag{4.11}$$

The distribution of these test statistics is non-standard and have asymptotic null distributions (Andrews, 1993). Approximate asymptotic $p - values$ for the validation of three can also be found in the literature (Hansen, 1997).

The test statistic for the Quandt-Andrews test is based on the F-test, which is used to test whether the variances of two populations are equal. Specifically, the test statistic for the Quandt-Andrews test is calculated as the ratio of the variance of the data points before the break point to the variance of the data points after the break point. If this ratio is significantly different from 1, in can be concluded that there is a structural break at the specified location in the data.

The distribution of the stated statistics becomes degenerated at the beginning and at the end of the sample. Therefore, the suggestion is to exclude the sample first and last observations from the testing procedure, by removing the first and last 15% of the sample observations.

If the $p - value$ is less than the chosen significance level (usually $\alpha = 0.05$), or if the test statistics are greater than the critical value for the chosen level of significance, or too large, the null hypothesis can be rejected in favour of the alternative hypothesis, and it is concluded that there is a structural break in the data.

The Quandt-Andrews Test can show some limitations in case of multiple structural changes, in terms of practical implementation, as it is sensitive to the choice of the break point, and the results can be affected by the location of the suspected break point.

Bai Perron tests

The Bai-Perron family of tests includes a number of different approaches for estimating multiple breaks. These methods include hybrid versions that include both sequential analysis and global maximization. Discussing the Bai and Perron models involves exploring aspects such as generalized serial correlation, different error and regressor distributions across segments, lagged dependent variables, and trending regressors, as well as various distributions for the errors and regressors across segments. Thereafter, a partial structural change model was also developed, in which not all parameters are subject to changes.

The Quandt-Andrews framework serves as base to these tests (Bai & Perron, 1998, 2003) expanded by Bai and Perron by allowing for numerous unknown breakpoints. The work presented in this literature is considered the foundation for the structural breaks method, where this approach considers that not all the parameters in the model are constant and can allow for breaks to occur at a limited number of m points (Bai & Perron, 1998). The purpose of the study was to figure out what causes different types of breaks and to estimate how many breaks happen. They looked at a model where all the variables were fixed and instead of trying to find where the breaks were, they wanted to find the best way to predict them. They also looked at the idea of no structural changes happening versus changes happening.

The purpose is to estimate the unknown coefficients and the dates of the break points $(T1, T2, \ldots, Tm)$ in a linear regression model, using T observations of (y_t, x_t, z_t). The break points are considered unknown and are estimated along with the coefficients. Bai and Perron provided three tests in their study to handle estimating various structural changes in this type of model: test of no structural change versus a fixed number of breaks, test of no structural change versus an unknown number of breaks and sequential tests.

Consider the typical multiple linear regression model with T observations and m possible breaks (creating $K = m + 1$ regimes) into consideration. The regression model to

estimate for the observations in regime j (where $k = 0, 1, 2, ..., , m$) is as follows:

$$y_t = x_t'\beta + z_t'\delta_k + \varepsilon_t,\ t = T_{k-1} + 1, ..., T_k \tag{4.12}$$

being y_t is the observed dependent variable at time t, x_t and z_t vectors of covariates, β and δ_k the vectors of coefficients, δ_k vary across regimes opposed to the coefficients associated to x_t. β is constant.

This model can be rewritten in matrix format, originating:

$$Y = X\beta + \overline{Z}\delta + E \tag{4.13}$$

being $Y = (y_1, ..., y_T)', X = (x_1, ..., x_T)', E = (\varepsilon_1, ..., \varepsilon_T)', \delta = (\delta_1, ..., \delta_{m+1})'$, and $\overline{Z}$ a matrix where the diagonal is composed of Z elements at $(T_1, ..., T_m)$.

Additionally, Bai and Perron have also considered the variance break model, where the breaks may occur at variance level of the error:

$$y_t = x_t'\beta + u_t,\ var(u_t) = \sigma_1^2, t \leq T_1,\ var(u_t) = \sigma_2^2, t > T_1, \tag{4.14}$$

To perform the test, the data is divided into two sub-samples, one before the suspected break point and one after the suspected break point. The test statistic is then calculated as the difference between the means of the two sub-samples, divided by an estimate of the standard error of the difference. One advantage of the Bai-Perron test is that it is robust to the presence of outliers in the data. It is also relatively easy to implement and can be used with a wide range of data types.

The null hypothesis of the test is that there is no structural break in the data, which means that the mean of the entire data set is constant over time. The procedure begins by determining the point at which the structural break is believed to have occurred. This is often done by visually inspecting the data and identifying any obvious breaks.

Test of no structural change versus a fixed number of breaks

For each partition, given each observation in time T, the estimates for β and δ_k, through the least squares method, are the ones minimizing

$$(Y - X\beta - \overline{Z}\delta)'(Y - X\beta - \overline{Z}\delta) = \sum_{i=1}^{m+1} \sum_{t=T_{1-1}+1}^{T_i} [y_t - x_t'\beta - z_t'\delta_i]^2 \tag{4.15}$$

The estimates for the given m number partitions are represented by $\hat{\beta}(T_k)$ and $\hat{\delta}(T_k)$. These coefficients and partitions are selected as the optimal solution for minimizing the sum of the squared residuals across all of the partitions

$$(\hat{T}_1, ..., \hat{T}_m) = argmin_{T_1,...,T_m} S_T(T_1, ..., T_m) \tag{4.16}$$

where ${}_T(T_1, ..., T_m)$ represents the sum of squared residuals given by the coefficients $\hat{\beta}(T_k)$ and $\hat{\delta}(T_k)$.

In order to select which combination of m segments yields the minimum value, a dynamic programming algorithm was proposed by Bai and Perron (Bai & Perron, 2003).

The Wald test is used for testing the null hypothesis of no change in the parameters $H_0 : \delta_1 = ... = \delta_{m+1}$ against pre-defined m breaks is set as

$$W_T(\lambda_1, ..., \lambda_m, q) = \left(\frac{T - (m+1)q - p}{m} \right) \frac{\hat{\delta}' R' [(\overline{Z}' M_x \overline{Z})^{-1} R'] R\hat{\delta}}{RSS_m} \tag{4.17}$$

with $\hat{\delta}$ the optimal m-break estimates of δ and RSS_m the sum of squared residuals of the alternative hypothesis.

Under the hypothesis of serial correlation and/or heteroskedasticity of the parameters is the residuals, the statistic of Wald test is

$$W_T(\lambda_1, ..., \lambda_m, q) = \frac{1}{T} \left(\frac{T - (m+1)q - p}{m} \right) \frac{\hat{\delta}' R' R\hat{\delta}}{RV(\hat{\delta})R'} \tag{4.18}$$

$V(\hat{\delta})$ the estimation for the covariance matrix of the estimated value $\hat{delta}$, designed to handle issues of serial correlation and heteroskedasticity. The specific form of $V(\hat{\delta})$ is dependent on the assumptions made about the distribution of the data and the errors within different segments.

Test of no structural change versus an unknown number of breaks

In their work, tests are proposed for determining the absence of structural changes, while also taking into account the possibility of an unknown number of breaks, given an upper bound M for the number of breaks (Bai & Perron, 1998). These tests are referred to as "double maximum tests", and include one using equal weights (UDMax) and the other utilizing weights that result in equal marginal $p - values$ across all m (WDMax). A more thorough explanation can be found in the original literature.

UDMax is defined as $WT(M, q) = \max_{1 \leq m \leq M} W_T \left(\hat{\lambda_1}, ..., \hat{\lambda_m}, q \right)$, the latter being the estimates of breakpoints that are determined by utilizing a global optimization technique to minimize the sum of squared residuals.

WDMax follows the $maxF_T(M, q)$, with the choice of M for the upper bound to be usually set to 5, with critical values varying little as M increases past this value.

In the case of non-standard distribution of these test statistics, the work of Bai and Perron in 2003 (Bai & Perron, 2003) has established methods for determining critical values and analyzing the impact of different trimming parameters, such as minimum sample size and number of regressors, on the number of breaks estimated.

Sequential tests

Bai and Perron also present a method for testing the presence of a specific number of breaks in a time series, which can also be used as the basis for a sequential testing procedure (Bai & Perron, 1998). For a fixed number m of breaks, $(\hat{T}_1, ..., \hat{T}_m)$ the estimated break points are obtained by minimizing the residual sum of squares globally. The test is applied to each segment containing the observations $\hat{T}_{k-1} + 1$ to $\hat{T}_k$ for $k = 1, 2, ..., m + 1$.

If the overall minimal value of the residuals sum of squares (over all segments where an additional break is included) is sufficiently smaller than the residuals sum of squares from the m breaks model, the null hypothesis of m breaks is rejected in favor of a model with $(m + 1)$ breaks. If rejected, the sample is divided, and tests are performed in each subsample, adding breakpoints whenever necessary. This process continues until no structural break is detected in all subsamples or until the maximum allowed breakpoints or subsample intervals are reached.

This test can be used in a sequential procedure by starting with $m = 0$, and repeatedly applying the test until the null hypothesis of no structural break is not rejected in all subsamples or until a maximum number of breakpoints or maximum subsample intervals to test is reached.

Empirical applications using historical data demonstrate that statistically identified changes in the mean or coefficients of linear regression align with significant historical, political, or economic events, indicating the practical relevance of the method (Zeileis, Kleiber, Kramer, & Hornik, 2003).

4.3. Structural breaks in unconditional variance

A quick glance at historical data of asset returns over a prolonged period may lead one to question: whether volatility remains constant over time. It is commonly observed in financial markets that volatility tends to fluctuate, with distinct periods of relative calm followed by periods of heightened volatility, a phenomenon known as clustering.

The pioneer econometric literature originally assumed that the asset price process $S_{tt=0,\dots,T}$ followed a normal distribution, but later research has shown that this assumption is not always accurate. Many studies have shown that the *logreturns*, $R_t = log\left(\frac{S_t}{S_{t-1}}\right)$, exhibit leptokurticity, which is a characteristic of a non-normal distribution. As a result, the assumption of normality has been relaxed and more sophisticated models such as the GARCH model have been proposed. However, these models still assume that the unconditional variance is constant, which may not be accurate and can lead to incorrect statistical inference and poor forecasting performance.

A variety of approaches have been proposed for detecting volatility breaks, resulting in a range of concurrent detection algorithms. One early method, proposed by Inclán and Tiao, was a CUSUM-type test for detecting changes in variance in *i.i.d.* Gaussian data, which was later iteratively extended to handle multiple breaks and allowed the definition of the ICSS algorithm (Inclan & Tiao, 1994). However, this method has faced criticism due to its assumptions not being met in real-world data, with issues such as non-Gaussian distributions and serial dependence causing poor performance. THE NPCP algorithm (Ross, 2012) was also presented as a non-parametric change-point algorithm, an approach based on Mood's rank test (that dates back to 1954 (Mood, 1954)).

ICSS of Inclán and Tiao test

Back in 1994, Inclán and Tiao developed a theoretical result that sustains the algorithm behind the test:

$$C_k = \sum_{i=1}^{k} R_i^2 \text{ and } D_k = \frac{C_k}{C_T} - \frac{k}{T}, \text{ with } R_{tt=1,\dots,T} \sim \text{ iid } N(0, \sigma^2) \tag{4.19}$$

with $R_t = \log\left(\frac{S_t}{S_{t-1}}\right)$ the log returns.

When $T \to \infty$ the weak convergence $\sqrt{\frac{T}{2}} \max_{1 \le k \le T} |D_k| \implies \sup_{t \in [0,1]} |B_t^0|$ with B_t^0 a Brownian bridge on $[0, 1]$. The outcome above facilitates identifying a singular volatility structural break by examining the null hypothesis of homogeneity in variance, $H_0 : \sigma^2 = const$ versus H_1: a shift in variance occurs at some point $1 < \tau < T$. The formal assessment method disproves the null hypothesis at a pre-determined significance level α if $\sqrt{\frac{T}{2}} \max_{1 \le k \le T} |D_k| > D_\alpha^*$, D_α^* which originates from the presence of a Brownian bridge in the analysis. A significance level of $\alpha = 0.05$ is typically utilized and a numerical simulation by Inclan and Tiao has determined that the value of $D_{0.05}^*$ at this level is 1.358 (Stawiarski, 2015).

If the inequality in the latter equation holds true, the point of deviation in variance (and therefore volatility) is identified as the moment at which the maximum value occurs. In the event that multiple breaks are present, the ICSS algorithm is applied iteratively by dividing the data set into smaller subsets, testing each until all change points have been detected.

Changepoint package Tests

To hold changepoint analysis objects, the "*changepoint*" R package adds a new object type named 'cpt'. The class has been designed in such a way that the 'cpt' object has the essential features needed for a structural changes analysis and subsequent summaries. Each of them is kept in a 'cpt' class slot entry. The class slots are as follows:

- data.set: a time-series object holding the data's numeric values.
- cpttype: characters specifying the sort of changepoint requested, such as mean and variance.
- method: characters indicating whether a single or multiple changepoint search technique was used.
- test.stat: denotes the test statistic.
- pen.type: characters indicating the penalty type, such as AIC, BIC, or manual.
- pen.value: the value of the penalty utilized in the analysis.
- cpts: a numeric vector containing the estimated changepoint positions.
- ncpts.max: the maximum number of changepoints that are being searched for.
- param.est: a set of parameters, each of which is a vector of the estimated numeric parameter values for each segment.

This function includes a number of typical penalty functions used in changepoint analysis. There are four of them: the SIC (Schwarz Information Criterion), the BIC (Bayesian Information Criterion), the AIC (Akaike Information Criterion), and the Hannan-Quinn. The proper penalty is still a matter of concern, and it is often determined by a number of criteria, including the amount of the modifications and the length of segments, both of which are unknown before to analysis (Killick & Eckley, 2014).

For single structural change detection, a hypothesis test can be given to identify a single changepoint. The null hypothesis, H0, corresponds to no changepoint ($m = 0$), whereas the alternative hypothesis, H1, corresponds to one changepoint ($m = 1$). The simplest version of cpt.mean or cpt.var functions can be used, without detailing on the search methods or penalties associated.

For multiple structural changepoints search within the data, there will be used two algorithms: Binary Segmentation (Sen & Srivastava, 1975) and PELT (Killick, Fearnhead, & Eckley, 2012).

The Binary Segmentation search algorithm recursively examines for a single change on distinct subsets of the data. If a change point is discovered, the signal is divided into two segments, one before and one after the change. The technique is repeated for both segments, and so forth. Thus, Binary Segmentation greedily searches by making local judgments on which the algorithm's next iteration is based. Under Binary Segmentation search algorithm, an indicative number of maximum breaks must be provided, contrary to the PELT search algorithm, built within the "*changepoint*" package.

The Pruned Exact Linear Time (PELT) algorithm is another method for detecting changepoints in time-series data. It aims to find the optimal segmentation of the time series into segments of different statistical properties, such as mean or variance. The key advantage of PELT is that it can be considered highly appropriate for large datasets.

PELT utilizes a cost function, denoted as $C(t, \tau)$, which represents the cost of dividing the time series up to time t into τ segments. The cost typically involves a measure of goodness-of-fit within segments and a penalty for introducing new segments. Common cost functions include sum of squares within segments and information criteria penalties.

PELT employs dynamic programming to efficiently compute the optimal segmentation. It calculates the optimal cost for all possible changepoint locations up to time t, written as $C(t, \tau)$, by considering the optimal cost up to a previous changepoints s and adding the cost of the segments from $s + 1$ to t.

$$C(t, \tau) = \min_{s<t} \left(C(s, \tau - 1) + \text{Cost}(s + 1, t) \right) \tag{4.20}$$

PELT incorporates pruning steps to eliminate suboptimal solutions. During the dynamic programming step, if a potential changepoint does not improve the cost function significantly, it is pruned, hence the name, reducing the computational complexity. After computing the cost for all possible changepoint locations up to the end of the time series,

the algorithm identifies the locations where introducing a changepoint results in a significant reduction in the overall cost (Gachomo, Gichuhi, & Wanjoya, 2015).

cpt.mean and cpt.var

The *cpt.mean* function is used to retrieve all change in mean methods inside the changepoint package.

The *cpt.var* function is used to access all change in variance methods inside the changepoint package, needing the data to have a fixed value mean across time, and so this periodic mean must be eliminated prior to analysis.

The considered arguments are as follows:

- `data`: The input time-series data.
- `penalty`: The penalty value used in the cost function. Higher penalty values lead to fewer changepoints. Common choices include `"Manual"` for user-defined penalties or specific numeric values like `"SIC"` for Schwarz Information Criterion or `"BIC"` for Bayesian Information Criterion. For the purpose of this paper, MBIC penalty will be used.
- `method`: The method used for segmenting the data. Options include `"BinSeg"` for Binary Segmentation, `"PELT"` for Pruned Exact Linear Time.
- `Q`: The user-defined penalty value if `penalty` is set to `"Manual"`.
- `minseglen`: The minimum segment length. Detected segments shorter than this length are ignored.

CHAPTER 5

Results and discussion - ARMA model simulation

In this chapter, an in-depth analysis of time-series data is conducted, employing a variety of statistical methods and tests. The chapter contains the studies using the simulated ARMA model with unit root tests such as ADF, PP, KPSS, ERS, and Zivot-Andrews.

Here, the focus lies on the simulation of the ARMA model, allowing for the exploration of the behavior of different unit root tests under controlled conditions. Through rigorous analysis using ADF, PP, KPSS, ERS, and Zivot-Andrews tests, the aim is to gain insights into their effectiveness in detecting unit roots in data that contains structural breaks and is stationary, trying to understand comparative strengths and limitations.

5.1. ARMA model simulation results

As explained previously in the Methodology chapter, this process involves generating synthetic financial data with means and standard deviations that vary between each iteration using an ARMA(1,1) model. The generated data is then subjected to five unit root tests to assess the stationarity of the time series. The results (0 for not stationary, 1 for stationary) from these tests are aggregated over multiple iterations.

After that, the proportions of stationary outcomes for each test are calculated and summarized, offering insights into the effectiveness of these tests in identifying stationary processes under specific conditions.

As can be visually observed in the next graph, the series generated points towards being stationary around two mean levels, with the variance also being potentially constant at each level:

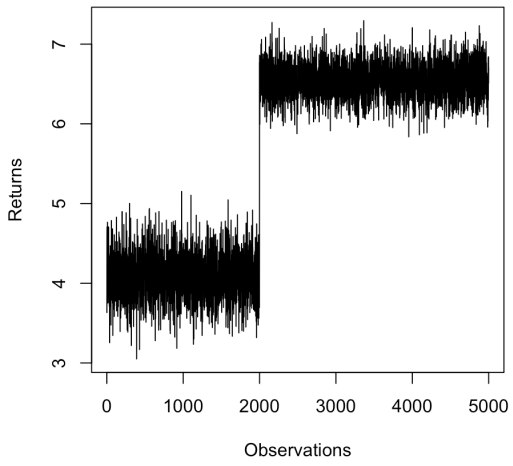


FIGURE 1. ARMA (1,1) model generation

The results for the unit root's tests calculated are stored as follows:

Iterations	**ADF**	**PP**	**KPSS**	**ERS**	**ZA**
1	0	0	0	0	1
2	0	1	0	1	1
3	0	0	0	0	1
4	0	0	0	1	1
5	0	0	0	1	1
6	...	...	...	...	...

TABLE 1. Results for the unit root's tests, where 0 stands for not stationary and 1 for stationary results

The proportion of stationary outcomes for each test are:

Test	ADF	PP	KPSS	ERS	ZA
Result	0	0.231	0.024	0.541	1

TABLE 2. Results of the unit root tests

5.2. Discussion

The interpretation of such results is as follows:

(1) ADF Test (Result: 0% - non-stationary):
- The ADF test consistently indicates that the data is non-stationary. This test is widely used and generally reliable, and its unanimous verdict is a strong indication against the stationarity hypothesis.

(2) PP Test (Result: 23,1% - Partial evidence for stationarity):
- The PP test shows a mixed outcome, with approximately 23% of the tests suggesting stationarity. While this might imply some possibility of stationarity, the majority of tests still point towards non-stationarity.

(3) KPSS Test (Result: 2,4% - Almost unanimously non-stationary):
- The KPSS test provides robust evidence against stationarity, with only about 2,4% of the tests suggesting stationarity. This near-unanimous agreement among KPSS tests reinforces the notion of non-stationarity in the dataset.

(4) ERS Test (Result: 54,1% - Mixed evidence for stationarity):
- The ERS test results are divided, with approximately 54,1% of the tests indicating stationarity. This indicates a significant possibility of the data being stationary, but it's important to note that almost half of the tests still suggest non-stationarity.

(5) ZA (Result: 100% - Complete evidence towards stationarity):
- The Zivot-Andrews test results are clear in terms of assessing that the data is stationary. This result is very important as the series is, indeed, stationary around the stuctural break at observation 2000.

The overall conclusion is conflictual. While the traditional ADF and KPSS tests strongly indicate non-stationarity, the PP test provides conflicting signals, the ERS tests are divided and the Zivot-Andrews points clearly to stationarity of the series. This mixed evidence underscores the complexity of the data's behavior. It suggests that there might be specific segments or aspects within the data that exhibit stationary properties while other parts do not. These results do not support the starting point conclusions, either by observation of the graph or the choice of parameters for the ARMA model, as the test was built to have two segments of stationary data at mean level with a structural break at observation 2000, hence the conclusion to the overall stationarity of the series was expected.

Collectively, and considering that different tests provide varying results, analyzing different segments of the data separately is proven to be beneficial, which corroborates the main point of this paper of testing time series for structural breaks as well as for stationarity of its variables.

This approach is essential for a comprehensive understanding of the data's behavior and corroborates the main argument of this paper: the importance of testing time series for structural breaks and account for ther to assess stationarity in their variables. Utilizing tests specifically designed to account for structural breaks is essential in capturing the nuances of complex data patterns. These results underscore the significance of employing a variety of tests that consider different aspects of the data, ultimately enhancing the accuracy and reliability of the analysis.

CHAPTER 6

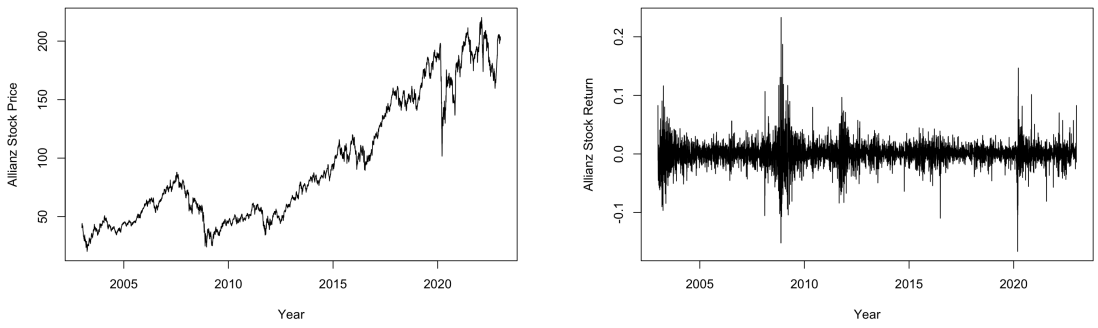
Results and discussion - Tests on real financial data

This chapter is organized as follows: (i) First, the results are presented regarding unit root's tests conducted on the daily stock prices and returns of the selected company, Allianz, randomly chosen. These tests serve to determine the stationarity of the company's stock prices and returns over time. (ii) After, the focus shifts to examining structural breaks within the data. Structural breaks tests aim to identify significant shifts or disruptions in the return patterns that may have occurred during the observation period. By diving into these tests, the intention is to gain deeper understandings of the stability and underlying dynamics of the daily stock prices and returns of this company. (iii) Later, the same tests, in a more condensed way, will be applied to daily CAC Index values from 2003 to 2022, and to monthly values for Employment Level, thousands of persons, seasonally adjusted from FRED, from 2003 to 2022.

The results regarding daily stock prices of RollsRoyce and monthly rates of Federal Funds Effective Rate can be found in Annex B.

6.1. Allianz

Firstly, the historical stock prices of Allianz are extracted, and a time-series object is created, spanning from January 1st, 2003, to December 31st, 2022, with a frequency of around 255 observations per year. Subsequently, the daily returns of Allianz stocks are computed and transformed into another time series, the returns, with the log returns function. Visualizations of both the stock prices and the corresponding returns are presented through plots, offering a graphical insight into the patterns and trends within the data:



(a) Allianz daily stock prices, 2003-2022 (b) Allianz daily stock returns, 2003-2022

FIGURE 2. Allianz stock prices and returns

6.1.1. Unit root's tests

ADF test

Value of test-statistic: -0.4697 ; 0.9546	
Critical values for test statistics:	
τ^2	-3.43 -2.86 -2.57
ϕ^1	6.43 4.59 3.78

TABLE 3. ADF test statistics - Allianz stock prices

The test-statistic (-0.4697) is higher than the critical values at all common significance levels. Therefore, the null hypothesis (presence of a unit root) cannot be rejected. This implies that the time series data is non-stationary and has a unit root, indicating a lack of a stable, long-term trend.

PP test

Value of test-statistic, type: Z-tau -0.5201	
Critical values for Z statistics:	1% Level: -3.434737
	5% Level: -2.862664
	10% Level: -2.567396

TABLE 4. PP test statistics - Allianz stock prices

The test-statistic (-0.5201) is higher than the critical values at all common significance levels. Therefore, the null hypothesis (presence of a unit root) cannot be rejected. This implies that the time series data is non-stationary and has a unit root.

KPSS test

Value of test-statistic: 8.5045	
Critical values for significance level:	1% Level: 0.216
	5% Level: 0.146
	10% Level: 0.119

TABLE 5. KPSS test statistics - Allianz stock prices

In this case, the test-statistic (8.5045) exceeds the critical values at all common significance levels (10%, 5%, 2.5%, and 1%). Therefore, the null hypothesis of the KPSS test (that the series is stationary around a deterministic trend) is rejected. This suggests that the data is not stationary and exhibits a unit root, indicating a need for differencing to achieve stationarity.

ERS test

Value of test-statistic: 0.532	
Critical values for significance level:	1% Level: -2.57
	5% Level: -1.94
	10% Level: -1.62

TABLE 6. ERS test statistics - Allianz stock prices

Tn this case, the test-statistic (0.532) does not exceed the critical values at any common significance level. Therefore, the null hypothesis of the ERS test (that the series has a unit root) is not rejected, indicating that series may be non-stationary.

ZA test

Value of test-statistic: -5.0403	
Critical values for significance level:	1% Level: -5.57
	5% Level: -5.08
	10% Level: -4.82
Potential Break Point Position: 1462	

TABLE 7. Zivot-Andrews test statistics - Allianz stock prices

The Zivot-Andrews test was conducted to assess the stationarity of the time series data with a potential structural break, suggesting the position at observation 1462. The test statistic (-5.0403) falls below the critical values at 10% significance level(-4.82). Therefore, the null hypothesis of a unit root is rejected at this level, but not at the typical of $\alpha = 0.05$. Hence, there is only statistical significance to reject the null at 10% but not at 5%.

Additionally, the test results suggest a potential structural break at position 1462, correspondent to date 2008-09-02, which can also be observed in Figure 2. A structural break implies that there might be a significant change in the underlying data-generating process at that specific point in time, that could be justified by the Global Financial Crisis that took place in 2008.

As expected, the overall results point to non stationarity of the data, with Zivot-Andrews test considering the series stationary at 10% significance level, with a break at 2008-09-02.

6.1.2. Structural breaks tests

Chow's Breakpoint test

The Chow's test is employed to assess the presence of structural breaks in the stock price data. By using the stock prices, the aim is to find structural breaks at mean level.

By observation, the test was conducted for the 5th of March, 2020, corresponding to observation 4380 in the dataset. The choice of this data point was due to the coronavirus

pandemic, as the declaration of the pandemic by the World Health Organization led to a sharp increase in market volatility. Investors were concerned about the economic impact of the virus and the potential disruptions to businesses and supply chains. The test could also be performed at datapoint 1462 as Zivot-Andrews test suggested.

The purpose of the test is to examine whether there is a significant structural change in the stock prices of Allianz during this period. The sctest() function is utilized with the type parameter set to "Chow" and the point parameter set to 4380 to focus on this specific observation as there could be a potential structural break.

The results are:

Test Statistic	3988.4	**P-Value**	< 2.2e-16

TABLE 8. Chow's Breakpoint test results - Allianz stock prices

Based on the results, the null hypothesis of no structural change is rejected. Therefore, there is substantial statistical evidence to suggest the presence of a structural break in the selected data point.

If the same test is applied to returns, on the same data point, the results are:

Test Statistic	0.017146	**P-Value**	0.8958

TABLE 9. Chow's Breakpoint test results - Allianz returns

As expected, since the mean is removed from the returns when applying the log function, the null hypothesis of no structural change is not rejected. Therefore, there is no substantial statistical evidence to suggest the presence of a structural break in the selected data point (a conclusion further reinforced by visual inspection). No further tests that are dedicated to changes in mean will be applied to stock returns.

Chow's Forecast test

Initially, two linear regression models are constructed: reg0 that represents the entire dataset, and reg1, that represents the subset of data up to a specific point (in this case, the same observation was used - observation 4380). The test statistic (Ftest) is computed by comparing the residual sums of squares of these models, adjusted for the sample sizes and the number of coefficients in the models. The resulting Ftest value and its associated p-value are:

Test Statistic	16.92678	**P-Value**	0

TABLE 10. Chow's Forecast test results - Allianz stock prices

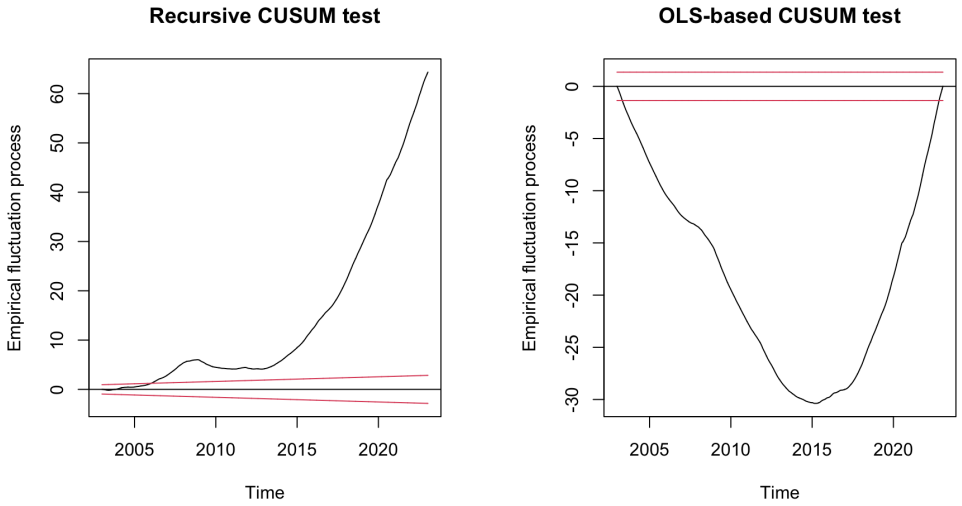
The calculated test statistic and the p-value suggest strong evidence against the null hypothesis, at any significance level. This implies that there is a significant difference in the model fit before and after the specified point, indicating a structural change in the

Allianz stock data at observation 4380.

CUSUM test

In this analysis, two tests will be considered: the Recursive Cumulative Sum (CUSUM Rec) and the OLS Cumulative Sum (CUSUM OLS). The first is applied to the Allianz stock data to identify subtle structural changes within the time series. The test is specifically designed to pinpoint sequential shifts in the data. Regarding OLS Cusum test, it checks for deviations in the OLS residuals from the expected pattern, aiming to detect any significant structural changes in the time series.

The resulting plots, that aim to illustrate any deviations from the expected pattern in the stock data, are the following:



(a) Recursive CUSUM Test (b) OLS CUSUM Test

FIGURE 3. CUSUM tests results - Allianz stock prices

The test statistics corroborate the visual inspection:

	Rec CUSUM	OLS CUSUM
Test Statistic	22.004	30.849
P-Value	2.2e-16	2.2e-16

TABLE 11. CUSUM test results - Allianz stock prices

The test results indicate a significant deviation from the expected pattern in the Allianz stock data. Both p-values are extremely small (2.2e-16), providing strong evidence against the null hypothesis, suggesting that there is a substantial structural change in the data.

MOSUM test

Two MOSUM tests are performed on the Allianz stock data, used to identify structural changes in the time series. The results are visually presented through two plots, that aim to illustrate any deviations from the expected pattern in the Allianz stock data. The visuals are:

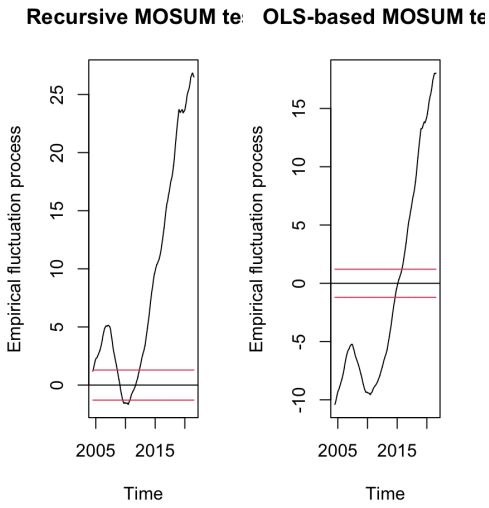


FIGURE 4. MOSUM tests results - Allianz stock prices

The test statistics corroborate the visual inspection:

	Rec MOSUM	OLS MOSUM
Test Statistic	26.802	17.567
P-Value	0.01	0.01

TABLE 12. MOSUM test results - Allianz stock prices

Both conducted tests reveal a significant deviation from the expected pattern, with the associated p-values being 0.01, indicating evidence against the null hypothesis, under the typical significance level at 5%. This suggests a structural change in the data.

The comparison between CUSUM and MOSUM tests centers on their sensitivity to structural changes in data. CUSUM excels at detecting abrupt shifts, while MOSUM, a modified version, is more versatile, capturing gradual changes and multiple shifts effectively. The tests differ in their underlying statistics, with CUSUM using cumulative sums and MOSUM employing adaptable statistics like weighted differences. MOSUM's flexibility allows it to identify diverse deviations, making it sensitive to minor changes. While CUSUM is simpler, MOSUM's complexity suits intricate patterns. The choice between them depends on the data's nature. Typically, MOSUM is preferred for nuanced analyses, ensuring a precise understanding of structural shifts.

Quandt-Andrews Breakpoint test

The Quandt Andrews test involves dividing the dataset into segments and conducting F-statistic tests to identify potential structural changes. Moreover, the analysis identifies the specific point within the dataset where the highest probability of a structural change occurs. The obtained results are:

Test Type	**Test Statistic**	**P-Value**
supF test	22036	$< 2.2 \times 10^{-16}$
aveF test	8869.4	$< 2.2 \times 10^{-16}$
expF test	∞	NA

TABLE 13. Quandt Andrews test results - Allianz stock prices

Based on the results, it can be concluded that it points towards non-stationarity of the underlying data, with the highest probability of structural change detected at datapoint 3558, at 2016-12-05.

Bai Perron test

The Bai-Perron test operates by detecting points where significant shifts occur in the data, indicating changes in the statistical properties like mean or variance. The test refers to the breakpoints (and accompanying) breakdates for all segmentations up to the maximum number of breaks, as well as the corresponding RSS and BIC.

The results are as follows:

Segments $(m+1)$	**0**	**1**	**2**	**3**	**4**
1	-	-	-	3558	-
2	-	-	2749	3687	-
3	-	-	2647	3558	4323
4	765	-	2760	3559	4324
5	765	1530	2647	3558	4323

Segments (m)	**0**	**1**	**2**	**3**	**4**
RSS	14,588,183	2,740,861	1,374,191	1,179,651	1,066,062
BIC	55,080	46,570	43,066	42,305	41,805

TABLE 14. Optimal $(m+1)$-segment partition and corresponding fit statistics - Allianz stock prices

(1) Breakpoints $(m+1)$ test indicates an optimal segment partition of 6, the results being:
   - For $m = 5$, breakpoints are at observations 751, 1530, 2647, 3558, and 4323, corresponding to the dates of 2005-12-06, 2008-12-05, 2013-05-02, 2016-12-05, and 2019-12-10, respectively.

(2) Fit Statistics:

- The Residual Sum of Squares (RSS) decreases as the number of segments (m) increases, indicating improved model fitting with more segments. Lower RSS values suggest better representation of the data.
- The Bayesian Information Criterion (BIC) is used for model selection. Smaller BIC values indicate better model fit. BIC values decrease as the number of segments increases, reflecting the trade-off between model complexity and fit to the data.

The analysis suggests that the stock data can be effectively segmented into an optimal number of 6 partitions, each characterized by different statistical properties.

The confidence intervals correspondent to each segment are given by:

Breakpoints at Observation Number			
	2.5% breakpoints	Breakpoints	97.5% breakpoints
1	762	765	766
2	1526	1530	1537
3	2646	2647	2648
4	3556	3558	3559
5	4317	4323	4327

TABLE 15. Breakpoints and confidence intervals

Bai-Perron test can also calculate F-statistics for various possible breakpoints, identifying the optimal breakpoints based on these statistics, and then ploting them, along with vertical lines indicating the detected breakpoints, as the next graph shows:

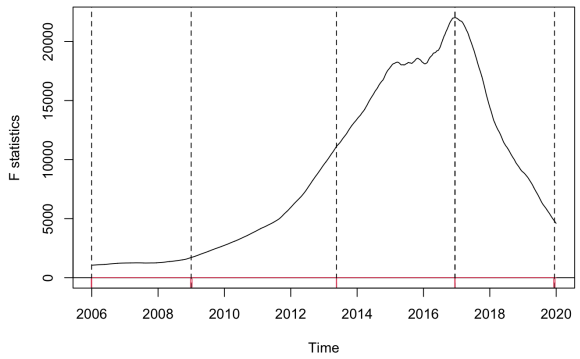


FIGURE 5. Fstat for Allianz stock prices and optimal breakpoints

ICSS of Inclán and Tiao test

This test is a robust statistical method utilized for detecting structural breakpoints within time-series data in the variance of a time series, rather than the mean. These breakpoints indicate instances where there are significant shifts or changes in the underlying statistical properties of the data set.

The detected changepoints results are:

Index	665	979	1414	1843
Date	2005-07-19	2006-10-05	2008-06-26	2010-03-05
2486	2765	3116	3725	4558
2012-09-06	2013-10-16	2015-03-11	2017-08-01	2020-11-16

TABLE 16. Breakpoints and corresponding dates for the ICSS test - Allianz stock prices

In graphical terms, these breaks can be observed at returns level:

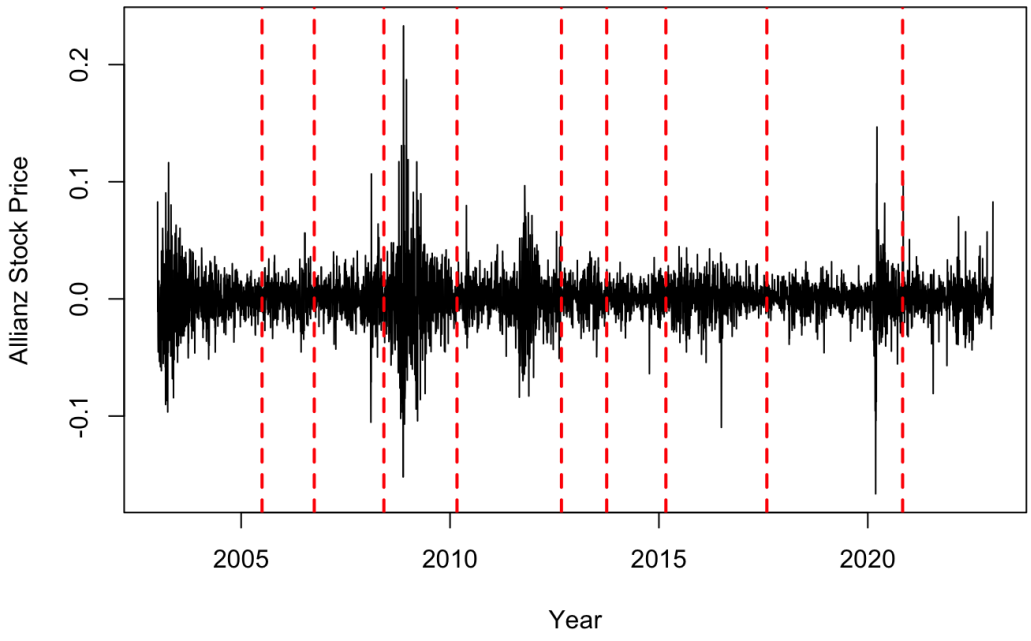


FIGURE 6. ICSS variance test results - Allianz log returns

The results point towards 9 structural breaks at variance level, what appears to be supported by the visual inspection. Yet, the test provided by "*ICSS*" package has no adjustable parameter that allows the user to better suit the test to the available data, hence producing the amount of segmentations that should be considered in order to fit the data to different models.

Changepoint package test

For single structural changepoint detection, without specifying the exact nature of the data or the statistical methods used, the results are as follows:

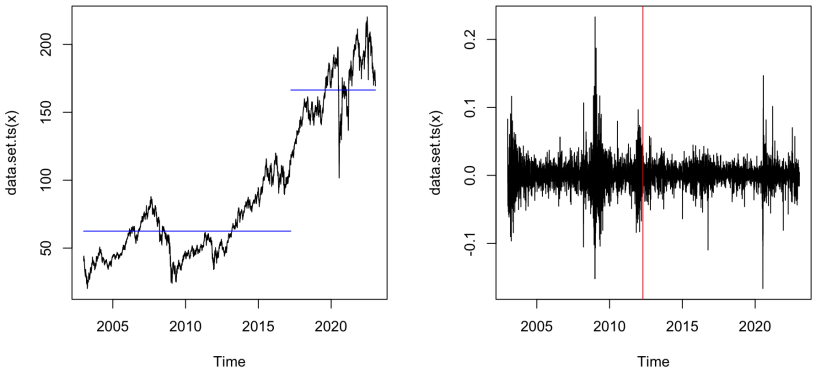


FIGURE 7. CPT mean and variance test results - Allianz stock prices and log returns

Suggesting a break at variance level at point 2323 that corresponds to date 2012-01-18 and 3558, that corresponds to 2016-12-05 for mean level.

Now, let us address the results under the Binary Segmentation algorithm search, allowing for up to 10 structural breaks in data:

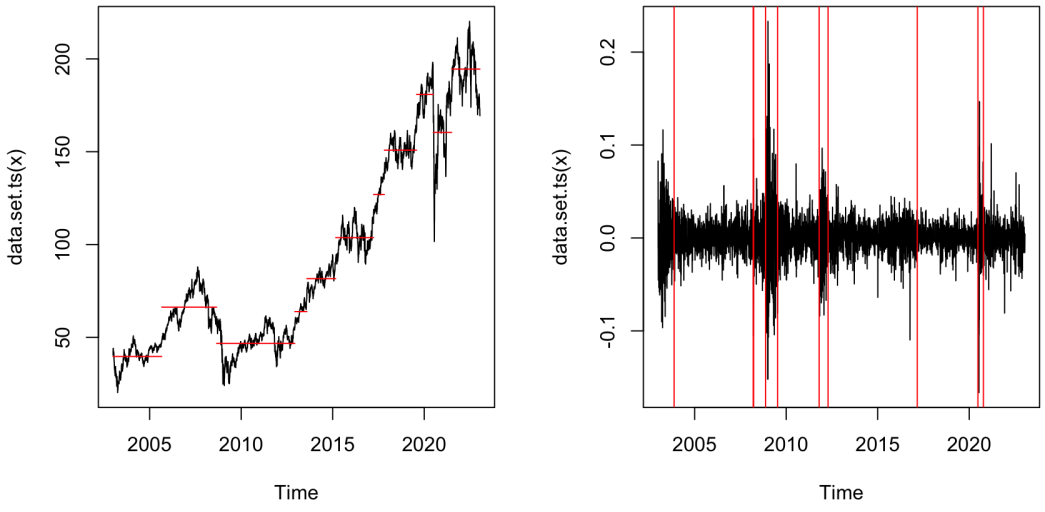
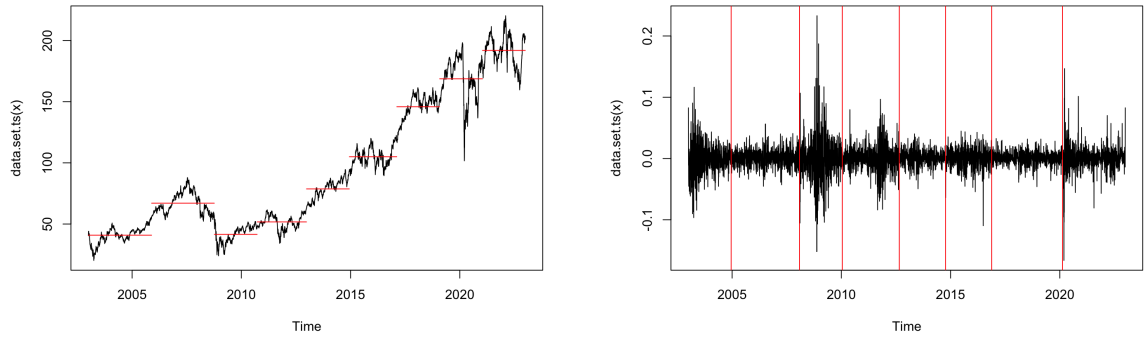


FIGURE 8. CPT mean and variance under Binary Segmentation search algorithm test results - Allianz stock prices and log returns

The results differ from the first ones as they are getting more and more adjusted to the data. If the parameter Q=10 was set to NULL, the resulting partition would tend to infinity (or an exagerated number of structural breaks), which removes the point of

finding models to fit the data. Yet, the knowledge of this Q parameter can be very biased, being set to 10 as a standard based only on visual inspection.

Finally, the results under the PELT algorithm search, without a maximum number of structural breaks indicated, are as follows:



(a) PELT cpt.mean results - stock prices (b) PELT cpt.var results - log returns

FIGURE 9. PELT tests results - Allianz

One important note is at the PELT search algorithm, as in $cpt.mean(\text{"}data\text{"}, penalty = \text{"}MBIC\text{"}, method = \text{"}PELT\text{"}, minseglen = 500)$, the same applies to $cpt.var$, the last variable allows to turn the test less sensitive to changes in mean, increasing the length between data where structural breaks can be found, thus providing a smaller number of breaks. This adjustment was conducted given that the first results where not adjusted to the expectations.

The main differences between these results mainly derive from:

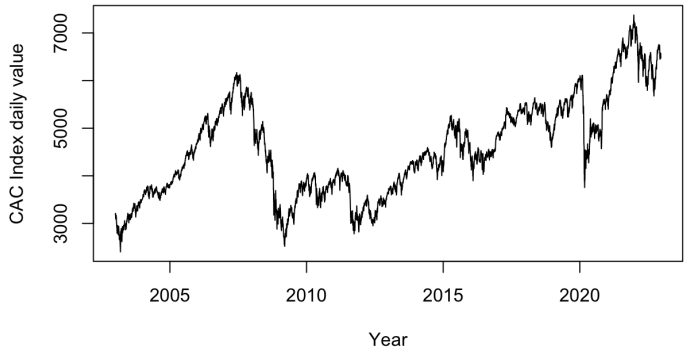
- PELT: more precise but intensive in terms of hardware computational needs. It often provides exact solutions to the changepoint problem, ensuring that it finds the optimal set of changepoints with respect to the chosen penalty function and its performance is generally less affected by the sample size.
- Binary Segmentation: On the other hand, it is more computationally efficient but might not find the exact optimal solution. It is based on a recursive binary segmentation approach, which divides the data into segments recursively. While this one is faster, it may not always find the optimal changepoints, as the user must provide that maximum number of segmentations to the algorithm. It may also become less accurate with very large datasets due to its recursive nature and potential limitations in processing power

As a conclusion, it can be recomended to use PELT when precision is critical, the dataset is extense and you have the computational resources to handle the processing needs and use Binary Segmentation when there is need for a faster analysis, or want a quicker overview of possible changepoints, being more suitable for initial exploratory analysis.

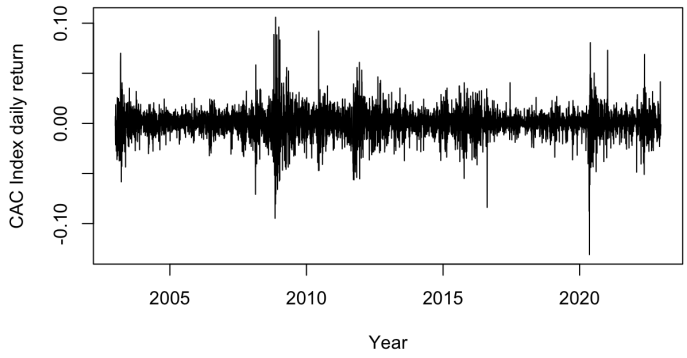
6.2. CAC Index

As previously stated, this section will contain the results' discussion as the tests are provided, but not under the same extent of explanation as the one provided in the previous section, as the introductory notes are similar.

Firstly, the historical values of CAC Index are extracted, and a time-series object is created, spanning from January 1st, 2003, to December 30th, 2022, with a frequency of around 255 observations per year. Subsequently, the daily returns are computed and transformed into another time series, with the log returns function. Visualizations of both series are presented through plots, offering a graphical insight into the patterns and trends within the data:



(a) CAC Index daily values, 2003-2022



(b) CAC Index log returns, 2003-2022

FIGURE 10. CAC Index daily values and log returns

6.2.1. Unit root's tests

The results for unit root's test for CAC index daily values are as follows:

Test	Result
ADF Test	Value of test-statistic: -1.8163 2.0002 Critical values for test statistics: τ^2: -3.43 -2.86 -2.57 ϕ^1: 6.43 4.59 3.78
PP Test	Value of test-statistic, type: Z-tau: -1.7438 Critical values for Z statistics: 1% Level: -3.43467 5% Level: -2.862634 10% Level: -2.56738
KPSS Test	Value of test-statistic: 4.6831 Critical values for significance level: 1% Level: 0.216 5% Level: 0.146 10% Level: 0.119
ERS Test	Value of test-statistic: -0.394 Critical values for significance level: 1% Level: -2.57 5% Level: -1.94 10% Level: -1.62
Zivot-Andrews Test	Value of test-statistic: -5.6266 Critical values for significance level: 1% Level: -5.57 5% Level: -5.08 10% Level: -4.82 Potential Break Point Position: 2239

TABLE 17. Unit root test statistics - CAC Index daily values

All the required tests point for non-stationarity of the time-series data, except for the Zivot-Andrews that considers the time-series to be stationary even at 1% significance level. Additionally, this last test suggests a structural break at observation 1281, that corresponds to the date of 2007-12-12. This date can be related to the Global Crisis that took place around 2008. Sadly, the Zivot-Andrews is not suggesting the breakpoint to be around 2020, pointing towards the pandemic. Yet, it is a very acceptable result, supported by visual inspection.

6.2.2. Structural breaks' tests

Chow's tests

Chow's Breakpoint Test Statistic	2442.7	**P-Value**	< 2.2e-16
Chow's Forecast Test Statistic	4.0791	**P-Value**	< 4.1893e-188

TABLE 18. Chow's Breakpoint test results - CAC Index daily values

Chow's tests' results both point towards a breakpoint occuring at datapoint 4380 (chosen as explained earlier at Allianz's analysis, as well as by observation), implying there is

a significant difference in the model fit before and after the specified point.

CUSUM and MOSUM tests

As for the CUSUM test, the produced graphs are as follows:

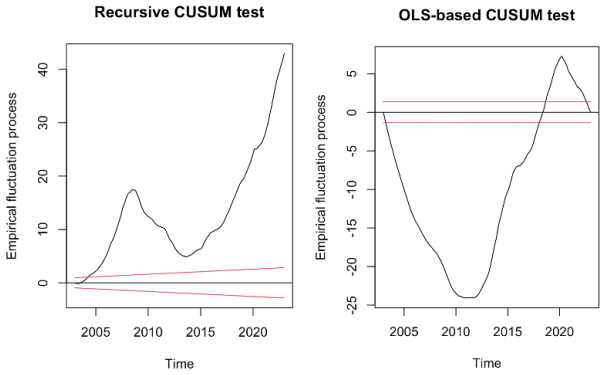


FIGURE 11. CUSUM Recursive and OLS-bases test results - CAC Index daily values

The test statistics corroborate the visual inspection:

	Rec CUSUM	OLS CUSUM
Test Statistic	14.311	22.769
P-Value	2.2e-16	2.2e-16

TABLE 19. CUSUM Test Results - CAC Index daily prices

The test results indicate a significant deviation from the expected pattern in the CAC Index values. Both p-values are well below significance level, suggesting that there is a substantial structural change in the data.

In terms of the MOSUM results, the produced graphs are as follows:

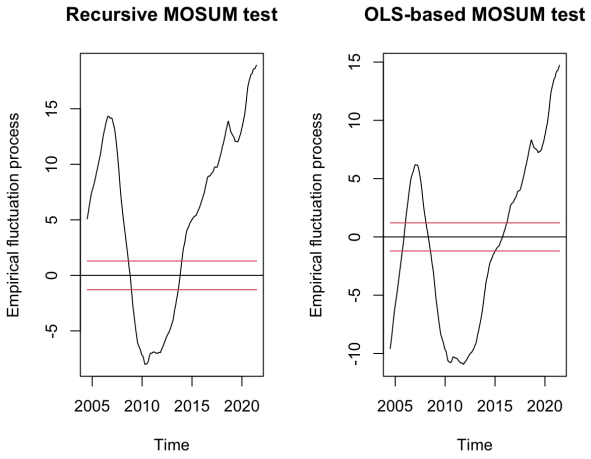


FIGURE 12. MOSUM Recursive and OLS-bases test results - CAC Index daily prices

The test statistics corroborate the visual inspection:

	Rec MOSUM	OLS MOSUM
Test Statistic	18.912	14.72
P-Value	0.01	0.01

TABLE 20. MOSUM test results - CAC Index daily prices

Both conducted tests reveal a significant deviation from the expected pattern, with the associated p-values indicating evidence against the null hypothesis, under the typical significance level at 5%, suggesting a structural change in the data.

Quandt-Andrews Breakpoint test

The obtained result are:

Test Type	**Test Statistic**	**P-Value**
supF test	4509	$< 2.2 \times 10^{-16}$
aveF test	2036.7	$< 2.2 \times 10^{-16}$
expF test	∞	NA

TABLE 21. Quandt Andrews test results - CAC Index daily prices

With the highest probability of structural change detected at datapoint 3636, at 2017-02-28, and unexpected result that is not as easily corroborated as the others by visual inspection.

Bai Perron test

The results are as follows:

Segments $(m+1)$	**0**	**1**	**2**	**3**	**4**
1	-	-	-	3636	-
2	-	-	3096	-	4330
3	769	1538	-	3583	-
4	769	1538	2719	3672	-
5	769	1538	2719	3577	4346

Segments (m)	**0**	**1**	**2**	**3**	**4**	**5**
RSS	5.328e+09	2.836e+09	2.614e+09	1.917e+09	1.446e+09	1.296e+09
BIC	8.567e+04	8.245e+04	8.205e+04	8.048e+04	7.905e+04	7.850e+04

TABLE 22. Optimal $(m+1)$-segment partition and corresponding fit statistics - CAC Index daily prices

(1) Breakpoints $(m+1)$ test indicates an optimal segment partition of 6, the results being:

- For $m = 5$, breakpoints are at observations 769, 1538, 2719, 3577, and 4346, corresponding to the dates of 2005-12-08, 2008-12-15, 2013-07-31, 2016-12-06, and 2019-12-10, respectively.

The analysis suggests that the stock data can be effectively segmented into an optimal number of 6 partitions, each characterized by different statistical properties.

The confidence intervals correspondent to each segment are given by:

Breakpoints at Observation Number			
	2.5% breakpoints	Breakpoints	97.5% breakpoints
1	766	769	771
2	1537	1538	1541
3	2717	2719	2721
4	3575	3577	3579
5	4330	4346	4347

TABLE 23. Breakpoints and confidence intervals

ICSS of Inclán and Tiao test

The detected changepoints results are:

Index	485	816	1473	2719	3637	4648
Date	2004-11-10	2006-02-16	2008-09-15	2013-07-01	2017-03-01	2021-02-15

TABLE 24. Breakpoints and corresponding dates for the ICSS Test - CAC Index daily values

In graphical terms, these breaks can be observed at returns level:

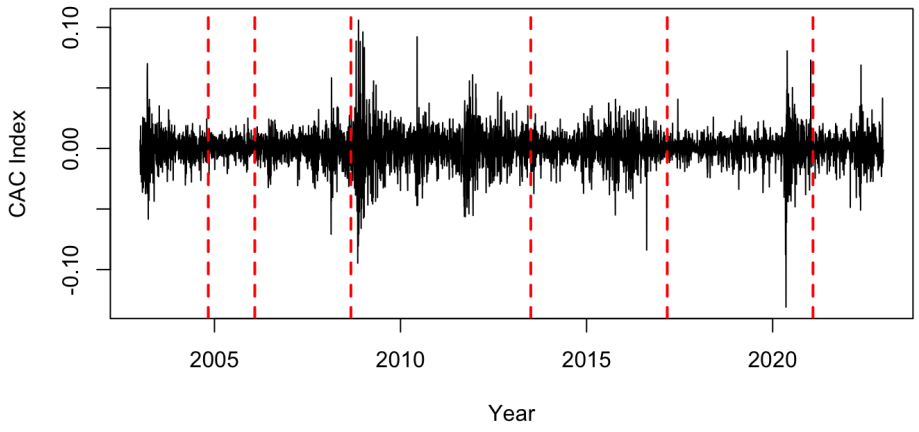
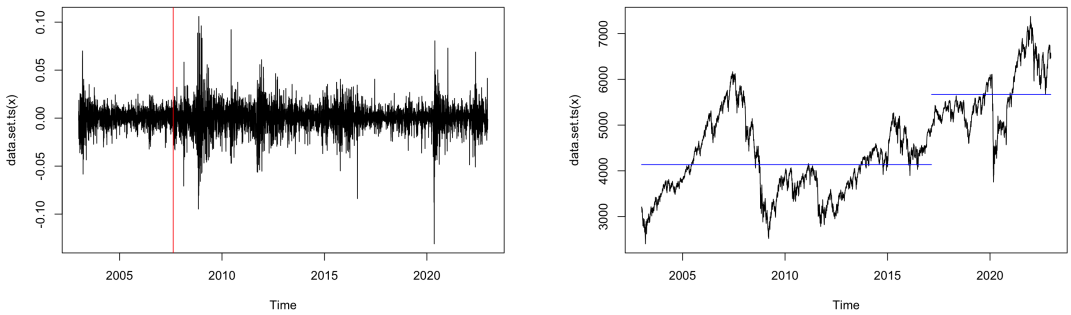


FIGURE 13. ICSS variance test results - CAC Index daily log returns

Changepoint package tests

For single structural changepoint detection, without specifying the exact nature of the data or the statistical methods used, the results are as follows:

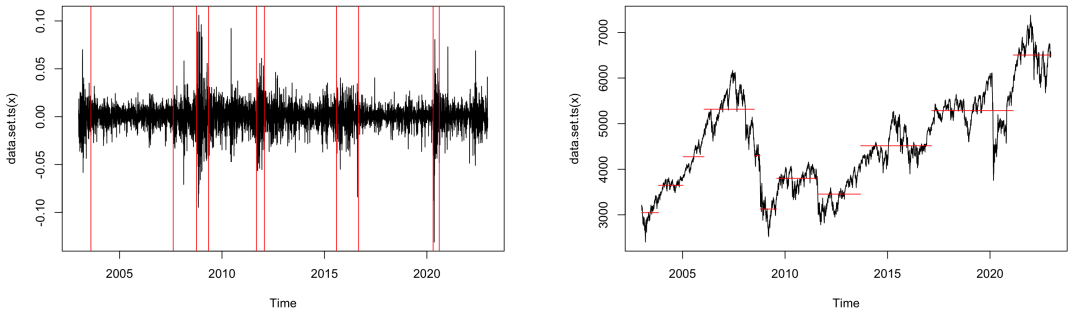


(a) Cpt.var tests' results - daily log returns (b) Cpt.mean tests' results - daily values

FIGURE 14. cpt.var and cpt.mean results - CAC Index

Suggesting a break at variance level at point 1174 that corresponds to date 2007-07-16 (closer to the one pointed by Zivot-Andrews test) and 3636, that corresponds to 2017-02-28 for mean level.

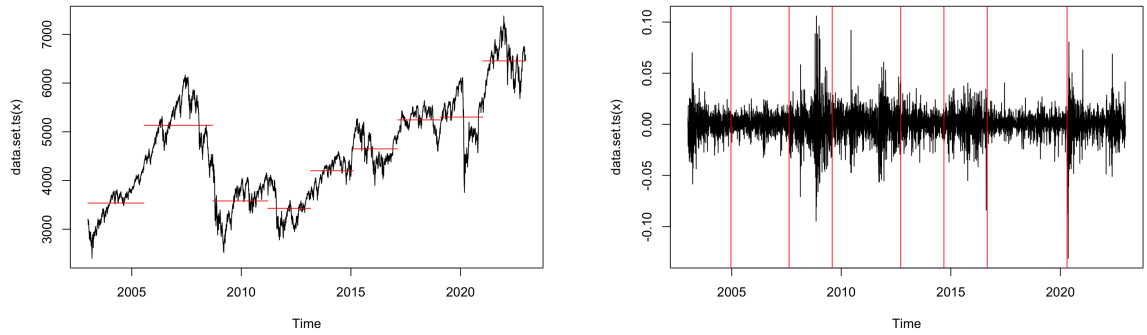
Now, let us address the results under the Binary Segmentation algorithm search, again, allowing for up to 10 structural breaks in data:



(a) Cpt.var tests' with Binary Segmentation results - daily log returns (b) Cpt.mean tests' with Binary Segmentation results - daily values

FIGURE 15. cpt.var and cpt.mean with Binary Segmentation results - CAC Index

Finally, the results under the PELT algorithm search, without a maximum number of structural breaks indicated, are as follows:



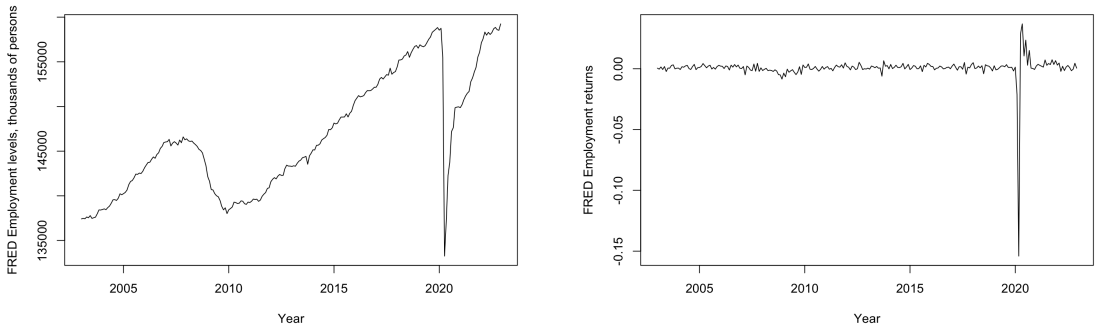
(a) PELT cpt.mean results - daily values (b) PELT cpt.var results - daily log returns

FIGURE 16. PELT Tests' results - CAC Index

The results are rather close, with PELT providing fewer optimal breaks of the data. Again, both are pointing towards the same areas, the choice of each test being more related to prior knowledge of data, sample size and execution capacity.

6.3. FRED Employment level

Firstly, the monthly historical values of FRED Employment monthly levels, thousands of persons, seasonally adjusted, are extracted, and a time-series object is created, spanning from January 1st, 2003, to December 1st, 2022, with a frequency of around 12 observations per year. Subsequently, the returns are computed and transformed into another time series, with the log returns function. Visualizations of both series are presented through plots, offering a graphical insight into the patterns and trends within the data:



(a) FRED Employment monthly levels, thousands of persons, seasonally adjusted, 2003-2022 (b) FRED Employment monthly log returns, 2003-2022

FIGURE 17. FRED Employment monthly levels and log returns

6.3.1. Unit root's tests

The results for unit root's test are as follows:

Test	Result
ADF Test	Value of test-statistic: -1.7839 ; 1.9507 Critical values for test statistics: τ^2: -3.46 -2.88 -2.57 ϕ^1: 6.52 4.63 3.81
PP Test	Value of test-statistic, type: Z-tau: -1.5235 Critical values for Z statistics: 1% Level: -3.459112 5% Level: -2.873702 10% Level: -2.573195
KPSS Test	Value of test-statistic: 0.3602 Critical values for significance level: 1% Level: 0.216 5% Level: 0.146 10% Level: 0.119
ERS Test	Value of test-statistic: -0.0078 Critical values for significance level: 1% Level: -2.57 5% Level: -1.94 10% Level: -1.62
Zivot-Andrews Test	Value of test-statistic: -4.7377 Critical values for significance level: 1% Level: -5.57 5% Level: -5.08 10% Level: -4.82 Potential Break Point Position: 206

TABLE 25. Unit Root test statistics - FRED Employment monthly levels

All the required tests point for non-stationarity of the time-series data. Additionally, the Zivot-Andrews test suggests a structural break at 206 datapoint, that corresponds to the date of February 2020. This date is very much likely related to the pandemic that begun in late 2019/early 2020, proven to have had a strong impact at employment level. Plus, it is also supported by visual inspection.

6.3.2. Structural breaks' tests

Chow's tests

Chow's Breakpoint Test Statistic	38.077	**P-Value**	< 2.903e-09
Chow's Forecast Test Statistic	2.193924	**P-Value**	< 0.0004256

TABLE 26. Chow's Breakpoint test results - FRED Employment monthly levels

Particularly for this time series, the chosen point to apply Chow's tests was at observation 206, that corresponds to February 2020, as Zivot-Andrews' test suggested. The results both point towards a breakpoint occuring at this datapoint, implying there is a significant difference in the model fit before and after the specified point.

CUSUM and MOSUM tests

As for the CUSUM test, the produced graphs are as follows:

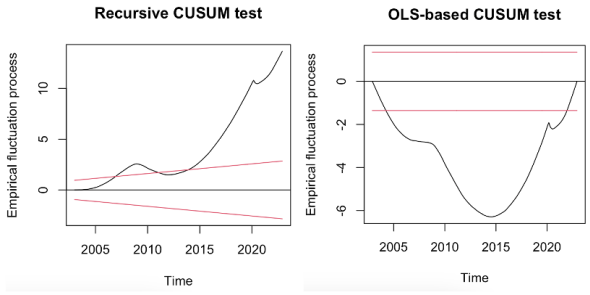


FIGURE 18. CUSUM test results - FRED Emplyment monthly levels

The test results indicate a significant deviation from the expected pattern in the FRED Employment monthly levels. Both p-values are well below significance level, suggesting that there is a substantial structural change in the data.

In terms of the MOSUM results, the produced graphs are as follows:

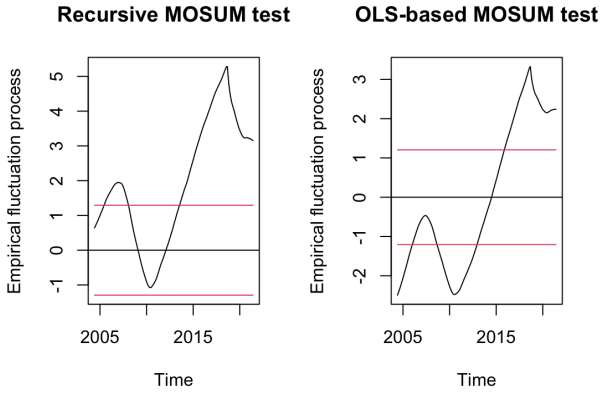


FIGURE 19. MOSUM test results - FRED Employment monthly levels

Both conducted tests reveal a significant deviation from the expected pattern, with the associated p-values indicating evidence against the null hypothesis, under the typical significance level at 5%, suggesting a structural change in the data.

Quandt-Andrews Breakpoint test

The obtained result are:

Test Type	Test Statistic	P-Value
supF test	512.72	$< 2.2 \times 10^{-16}$
aveF test	220.85	$< 2.2 \times 10^{-16}$
expF test	252.45	$< 2.2 \times 10^{-16}$

TABLE 27. Quandt Andrews test results - FRED Employment monthly levels

With the highest probability of structural change detected at datapoint 144, at 2014-12-01. Again, not an anticipated result, when facing all the other options and the visuals of the data.

Bai Perron test

The results are as follows:

Segments $(m+1)$	**0**	**1**	**2**	**3**	**4**
1	-	-	-	144	-
2	-	-	-	131	167
3	36	-	-	134	-
4	36	72	117	155	-
5	36	72	108	144	180

Segments (m)	**0**	**1**	**2**	**3**	**4**	**5**
RSS	1.025e+10	3.250e+09	2.842e+09	2.589e+09	2.127e+09	2.038e+09
BIC	4.909e+03	4.644e+03	4.623e+03	4.612e+03	4.575e+03	4.576e+03

TABLE 28. Optimal $(m+1)$-segment partition and corresponding fit statistics - FRED Employment monthly levels

(1) Breakpoints $(m+1)$ test indicates an optimal segment partition of 5, the results being:
   - For $m = 4$, breakpoints are at observations 36, 72, 117, and 155, corresponding to the dates of December 2005, December 2008, September 2012 and November 2015. respectively.

The analysis suggests that the stock data can be effectively segmented into an optimal number of 5 partitions, each characterized by different statistical properties.

The confidence intervals correspondent to each segment are given by:

Breakpoints at Observation Number			
	2.5% breakpoints	Breakpoints	97.5% breakpoints
1	35	36	38
2	71	72	73
3	115	117	118
4	150	155	156

TABLE 29. Breakpoints and confidence intervals

ICSS of Inclán and Tiao test

The detected changepoints results are:

Index	206	208	214
Date	February 2020	April 2020	October 2020

TABLE 30. Breakpoints and Corresponding Dates for the ICSS Test - FRED Employment monthly levels

In graphical terms, these breaks can be observed at returns level:

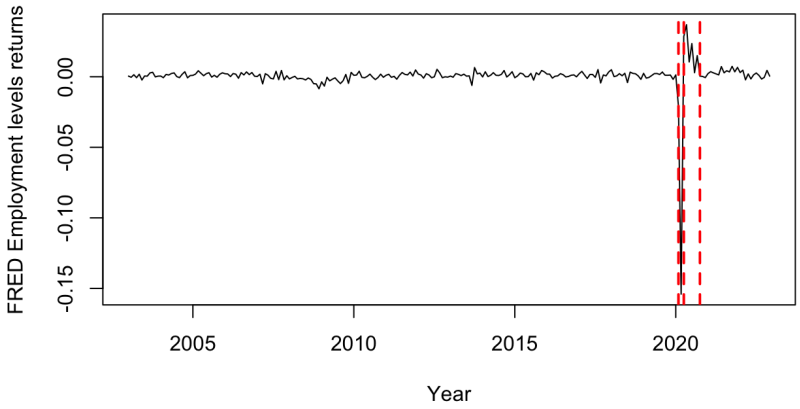
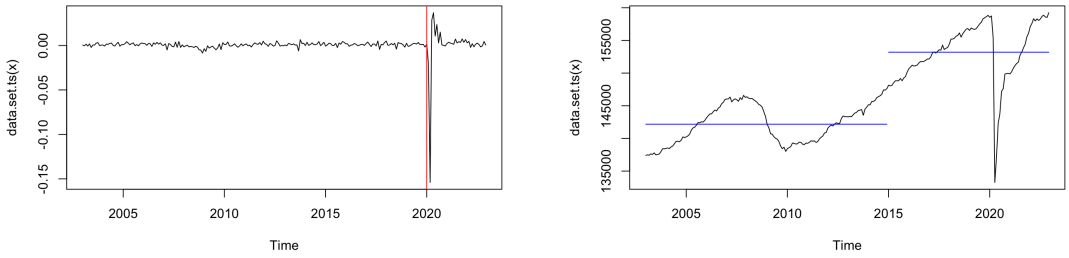


FIGURE 20. ICSS variance test results - FRED Employment monthly log returns

These results are very aligned with the expectations, as at variance level there is a clear disruption at the data around 2020.

Changepoint package tests

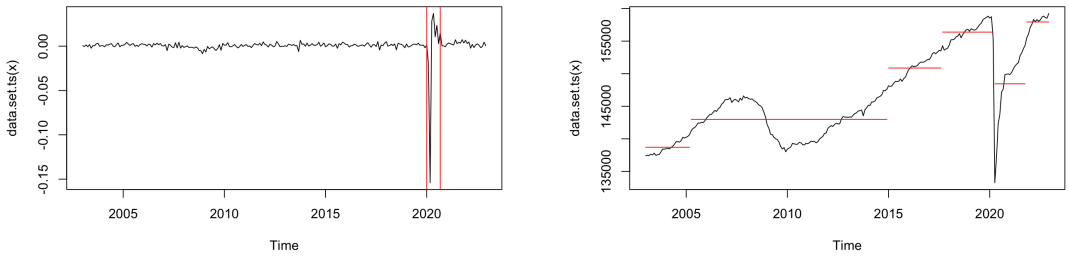
For single structural changepoint detection, the results are as follows:



(a) Cpt.var tests' results - monthly log returns (b) Cpt.mean tests' results - monthly levels

FIGURE 21. FRED Employment levels cpt.var and cpt.mean results

Suggesting a break at variance level at point 205 that corresponds to January 2020 (as anticipated) and 144, that corresponds to December 2014 for mean level. Now, as for the Binary Segmentation algorithm search, allowing for up to 5 structural breaks in data:



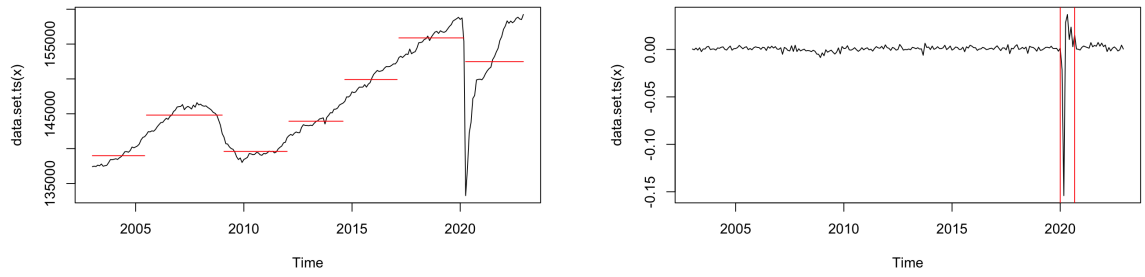
(a) Cpt.var tests' with Binary Segmentation results - monthly log returns

(b) Cpt.mean tests' with Binary Segmentation results - monthly levels

FIGURE 22. FRED Employment levels cpt.var and cpt.mean with Binary Segmentation results

Here, at mean level, the Binary Segmentation produced 5 breakpoints (as it typically does) but at variance level it recognized only two.

Finally, the results under the PELT algorithm search, without a maximum number of structural breaks indicated, are really close to the ones obtained with Binary Segmentation serch algorithm, and one observation that can lead to these results rather that the data's type is also the number of datapoints, as this time series accounts only for around 240 entrances. Again, both are pointing towards the same areas, the choice of each test being more related to prior knowledge of data, sample size and execution capacity. The results for PELT are as follows:



(a) PELT cpt.mean results - monthly levels

(b) PELT cpt.var results - monthly log returns

FIGURE 23. PELT tests' results - FRED Employment levels

CHAPTER 7

Conclusion

In conclusion, the analysis of the time series data, employing various unit root tests, reveals a complex and nuanced picture of its stationarity. The divergent results applied to the ARMA model obtained from traditional tests such as ADF, KPSS, PP, and ERS highlight the intricacies involved in assessing the underlying patterns. However, the introduction of the Zivot-Andrews (ZA) test, especially in the context of identifying structural breaks, sheds significant light on the dataset's behavior.

The findings underscore the critical importance of considering specific structural shifts when examining time series data. The correct assessment of stationarity identification in the presence of a structural break at point=2000 by the ZA test emphasizes that traditional stationarity tests may oversimplify the analysis. Accounting for these breaks is essential, as they significantly influence the behavior of the series. This study demonstrates that structural breaks can obscure or even reverse conclusions drawn from conventional unit root tests. Therefore, incorporating tests designed to detect these breaks is imperative for a comprehensive understanding of the data.

Moreover, the mixed evidence from different tests applied to real financial data highlights the complexity inherent in time series data. The variability in outcomes among tests indicates that a one-size-fits-all approach is inadequate when assessing stationarity. Instead, a nuanced methodology that incorporates multiple tests, each addressing specific aspects of the data, is necessary for accurate analyses.

Furthermore, this study reaffirms the necessity of exploring new models and techniques continuously. The ever-evolving landscape of time series analysis demands adaptive approaches. Relying solely on established methods might lead to overlooking crucial structural breaks or misinterpreting the data's behavior. Embracing emerging models and methodologies, tailored to capture the complexities of real-world data, is essential for robust and insightful analyses.

In summary, the study's results emphasize the significance of testing time series for structural breaks and accounting for them when assessing stationarity in their variables. The conflicting signals from various tests highlight the need for a holistic approach, incorporating diverse methodologies and remaining open to innovative techniques. By doing so, researchers can enhance the accuracy and reliability of their analyses, leading to a more profound understanding of the underlying patterns in time series data.

References

Adhikari, R., & Agrawal, R. (2013). *An introductory study on time series modeling and forecasting*. doi: 10.13140/2.1.2771.8084

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, *61*(4), 821–856. Retrieved 2023-03-09, from http://www.jstor.org/stable/2951764

Andrews, D. W. K., & Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*, *62*(6), 1383–1414. Retrieved from http://www.jstor.org/stable/2951753

Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, *66*(1), 47–78. Retrieved 2023-07-07, from http://www.jstor.org/stable/2998540

Bai, J., & Perron, P. (2003, 01). Computation and analysis of multiple structural-change. *Journal of Applied Econometrics*, *18*.

Box, G., Jenkins, G., & Reinsel, G. (2008). *Time series analysis: Forecasting and control*. Wiley.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Brown, R. L., Durbin, J., & Evans, J. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society Series B*, *37*, 149-192.

Cerqueira, V., Torgo, L., & Mozetič, I. (2020, 11). Evaluating time series forecasting models: an empirical study on performance estimation methods. *Machine Learning*, *109*, 1-32. doi: 10.1007/s10994-020-05910-7

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, *28*(3), 591–605.

Ditzen, J., Karavias, Y., & Westerlund, J. (2021). *Testing and estimating structural breaks in time series and panel data in stata.*

Gachomo, D. W., Gichuhi, A. W., & Wanjoya, A. (2015, November). The power of the pruned exact linear time (pelt) test in multiple changepoint detection. *American Journal of Theoretical and Applied Statistics*, *4*(6), 581-586. doi: 10.11648/j.ajtas.20150406.30

Hackl, P. (2016, 08). Moving sums (mosum). In (p. 1-6). doi: 10.1002/9781118445112.stat03107.pub2

Hansen, B. E. (1997). Approximate asymptotic p values for structural-change tests.

Journal of Business Economic Statistics, *15*(1), 60-67.

Hansen, B. E. (2001). The new econometrics of structural change: Dating breaks in u.s. labor productivity. *The Journal of Economic Perspectives*, *15*(4), 117–128.

Herranz, E. (2017, 03). Unit root tests. *Wiley Interdisciplinary Reviews: Computational Statistics*, *9*, e1396. doi: 10.1002/wics.1396

Inclan, C., & Tiao, G. C. (1994). Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association*, *89*(427), 913–923. Retrieved from `http://www.jstor.org/stable/2290916`

Johnson, T. (2009, 03). Time series analysis with applications in r, 2nd edition by cryer, j. d. and chan, k.-s. *Biometrics*, *65*, 337-337. doi: 10.1111/j.1541-0420.2009.01208$_1$2.x

Killick, R., & Eckley, I. A. (2014). changepoint: An r package for changepoint analysis. *Journal of Statistical Software*, *58*(3), 1,19. Retrieved from `https://www.jstatsoft.org/index.php/jss/article/view/v058i03` doi: 10.18637/jss.v058.i03

Killick, R., Fearnhead, P., & Eckley, I. (2012, 12). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, *107*, 1590-1598. doi: 10.1080/01621459.2012.737745

Kim, J.-H. (2011, December 31). Comparison of structural change tests in linear regression models. *Korean Journal of Applied Statistics*, *24*(6), 1197. doi: 10.5351/kjas.2011.24.6.1197

Kleiber, C. (2016). Structural change in (economic) time series [Working papers]. Retrieved from `https://EconPapers.repec.org/RePEc:bsl:wpaper:2016/06`

Luetkepohl, H., & Krätzig, M. (2004, 08). In applied time series econometrics. *Applied Time Series Econometrics*, 53. doi: 10.1017/CBO9780511606885

Mood, A. M. (1954). On the Asymptotic Efficiency of Certain Nonparametric Two-Sample Tests. *The Annals of Mathematical Statistics*, *25*(3), 514 – 522. Retrieved from `https://doi.org/10.1214/aoms/1177728719` doi: 10.1214/aoms/1177728719

Pankratz, A. (2008, 05). Forecasting with univariate box-jenkins models: Concepts and cases. , 553-555. doi: 10.1002/9780470316566.refs

Perron, P. (2005, 05). Dealing with structural breaks. *Palgrave handbook of econometrics, volume 1: econometric theory*.

Quandt, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association*, *55*, 324-330. doi: 10.1080/01621459.1960.10482067

Ross, G. (2012, 12). Modeling financial volatility in the presence of abrupt changes. *Physica A: Statistical Mechanics and its Applications*, *392*. doi: 10.1016/j.physa.2012.08.015

Sen, A., & Srivastava, M. S. (1975). On Tests for Detecting Change in Mean. *The Annals of Statistics*, *3*(1), 98 – 108. Retrieved from `https://doi.org/10.1214/aos/1176343001` doi: 10.1214/aos/1176343001

Shumway, R., & Stoffer, D. (2011). *Time series analysis and its applications with r examples* (Vol. 9). doi: 10.1007/978-1-4419-7865-3

Stawiarski, B. (2015, 03). Selected techniques of detecting structural breaks in financial volatility. *e-Finanse*, *11*. doi: 10.1515/fiqf-2016-0104

Tsay, R. (2010, 08). Analysis of financial time series: Third edition. *Analysis of Financial Time Series: Third Edition*. doi: 10.1002/9780470644560

Wei, W. (2006). *Time series analysis: Univariate and multivariate methods, 2nd edition, 2006*.

Zeileis, A., Kleiber, C. K., Kramer, W., & Hornik, K. (2003). Testing and dating of structural changes in practice. *Computational Statistics Data Analysis*, *44*(1), 109-123. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0167947303000306` doi: https://doi.org/10.1016/S0167-9473(03)00030-6

Zeileis, A., Leisch, F., Hornik, K., & Kleiber, C. (2002, 07). Strucchange: An r package for testing for structural change in linear regression models. *Journal of Statistical Software*, *7*, 1-38. doi: 10.18637/jss.v007.i02

Annex A

Computation methods

The following lines of code reflect what was used to perform the listed tests to real data retrieved from the markets, in R language, with detailed explanations of each step. The same code was used for each data type, with appropriate changes given the different data sets.

```
# Load required libraries
library(readxl) # Load readxl library for reading Excel files
library(ggplot2) # Load ggplot2 library for creating plots
library(tseries) # Load tseries library for time-series analysis
library(strucchange) # Load strucchange library for structural change
    testing

# Read data from an Excel file
Data <- read_excel("File path")

# Extract relevant columns
allianz <- Data$Allianz # Extract Allianz stock prices
date <- Data$Date # Extract corresponding dates

# Convert stock prices to a time series with specified start and end
    dates and frequency adjusted to data
stock.allianz <- ts(allianz, start = c(2003, 1, 1), end = c(2023, 0, 1),
    frequency = 255)

# Compute log returns of the Allianz stock prices and convert to a time
    series
ret.allianz <- ts(diff(log(allianz)), start = c(2003, 1, 1), end = c
    (2023, 0, 1), frequency = 255)

# Plot time-series data
plot(stock.allianz, xlab = "Year", ylab = "Allianz Stock Price") # Plot
    Allianz stock prices
```

```
plot(ret.allianz, xlab = "Year", ylab = "Allianz␣Stock␣Return") # Plot
   Allianz stock returns

#Unit root's tests

# Perform Augmented Dickey-Fuller unit root test
summary(ur.df(stock.allianz, lags=4, type="drift", selectlags="BIC"))

# Perform Phillips-Perron unit root test
summary(ur.pp(stock.allianz, type="Z-tau", model="constant", lags="short"
   ))

# Perform KPSS unit root test
summary(ur.kpss(stock.allianz, type="tau", lags="short"))

# Perform Elliott, Rothenberg, and Stock (ERS) unit root test
summary(ur.ers(stock.allianz, type = c("DF-GLS"), model = c("constant"),
   lag.max = 4))

# Perform Zivot Andrews unit root test
summary(ur.za(stock.allianz, model = "both", lag=NULL))


# Perform structural break tests

# Chow's test for stock prices and returns
sctest(stock.allianz ~ 1, type = "Chow", point = 4380, data = Data) #
   Chow's test for stock prices
sctest(ret.allianz ~ 1, type = "Chow", point = 4380, data = Data) # Chow'
   s test for returns

# Chow's predictive test for stock prices
reg0 <- lm(stock.allianz ~ 1) # Fit initial regression model
k <- length(coef(reg0)) # Count coefficients in the model
breakp <- 4380 # Specified breakpoint
Tobs <- length(stock.allianz) # Total observations
T1 <- length(stock.allianz[1:breakp]) # Observations before the
   breakpoint
T2 <- Tobs - T1 # Observations after the breakpoint
```

```
reg1 <- lm(stock.allianz[1:breakp] ~ 1) # Fit regression model before
    the breakpoint
rss0 <- sum(residuals(reg0)^2) # Residual sum of squares for the initial
     model
rss1 <- sum(residuals(reg1)^2) # Residual sum of squares for the model
    before the breakpoint
(Ftest <- ((rss0 - rss1) / T2) / (rss1 / (T1 - k))) # Compute F-statistic
     for Chow's predictive test
(pvalue <- pf(Ftest, T2, T1 - k, lower.tail = FALSE, log.p = FALSE)) #
    Compute p-value for the F-test

# CUSUM tests for stock prices
cusumrec <- efp(stock.allianz ~ 1, type = "Rec-CUSUM") # CUSUM test with
    recursive residuals
plot(cusumrec, alpha = 0.05) # Plot CUSUM test results
sctest(cusumrec, type = "Rec-CUSUM", data = Data) # Perform CUSUM test
    with recursive residuals

# CUSUM tests for OLS residuals
cusumols <- efp(stock.allianz ~ 1, type = "OLS-CUSUM") # CUSUM test with
    OLS residuals
plot(cusumols, alpha = 0.05) # Plot CUSUM test results
sctest(cusumols, type = "OLS-CUSUM", data = Data) # Perform CUSUM test
    with OLS residuals

# MOSUM tests for Recursive and OLS residuals
par(mfrow = c(1, 2)) # Set up a 1x2 grid for plotting
mosumrec <- efp(stock.allianz ~ 1, type = "Rec-MOSUM") # MOSUM test with
    recursive residuals
mosumols <- efp(stock.allianz ~ 1, type = "OLS-MOSUM") # MOSUM test with
    OLS residuals
plot(mosumrec, alpha = 0.05) # Plot MOSUM test results with recursive
    residuals
plot(mosumols, alpha = 0.05) # Plot MOSUM test results with OLS
    residuals
sctest(mosumrec, type = "Rec-MOSUM", data = Data) # Perform MOSUM test
    with recursive residuals
sctest(mosumols, type = "OLS-MOSUM", data = Data) # Perform MOSUM test
    with OLS residuals
```

```
# Quandt Andrews test
stocksqa <- cbind(Data[2:5101,], stock.allianz) # Combine relevant
   columns
regfstat <- lm(stock.allianz ~ 1, data = Data) # Fit initial regression
   model
qa <- Fstats(regfstat, from = 0.15, to = 0.85, data = stocksqa) # Compute
    F-statistics for Quandt Andrews test
plot(qa, pval = TRUE) # Plot Quandt Andrews test results
plot(qa) # Plot Quandt Andrews test results without p-values
sctest(qa, type = "supF", from = 0.15, to = 0.85, data = stocksqa) #
   Perform Quandt Andrews test (supF)
sctest(qa, type = "aveF", from = 0.15, to = 0.85, data = stocksqa) #
   Perform Quandt Andrews test (aveF)
sctest(qa, type = "expF", from = 0.15, to = 0.85, data = stocksqa) #
   Perform Quandt Andrews test (expF)
qa$Fstats[which.max(qa$Fstats)] # Point with highest probability of
   structural change
qa$breakpoint # Detected breakpoint index
date[qa$breakpoint] # Date corresponding to the detected breakpoint

# Bai-Perron test
bp.test <- breakpoints(stock.allianz ~ 1) # Bai-Perron structural break
   test
summary(bp.test) # Summary of the Bai-Perron test
breakpoints(bp.test) # Breakpoint indices
result<-bp.test$breakpoints # To address the first line of results
View(result)
date[result] # To provide more complete dates for the breakpoints
   identified
ci.allianz <- confint(bp.test) # Confidence interval for breakpoints
ci.allianz # Display confidence intervals
lines(ci.allianz) # Add confidence intervals to the plot
bp.test2 <- Fstats(stock.allianz ~ 1) # Compute F-statistics for all
   possible breakpoints
plot(bp.test2) # Plot F-statistics for all possible breakpoints
breakpoints(bp.test2) # Detected breakpoints
lines(breakpoints(bp.test2)) # Add detected breakpoints to the plot

## INCLAN TIAO test
# Install and load ICSS package
```

```
install.packages("ICSS")
library(ICSS)

# Perform INCLAN-TIAO test
inc.tiao <- ICSS(ret.allianz, demean=FALSE) # INCLAN-TIAO test
print(inc.tiao) # Vector list of breakpoints
plot(ret.allianz) # Plot the time-series data
summary(inc.tiao) # Summary of the INCLAN-TIAO test results
date[inc.tiao] # Summaty of corresponding dates for the test results


## Package changepoint
install.packages("changepoint")
library(changepoint)

## CPT - Change in variance, single-point identification
ansvar <- cpt.var(ret.allianz)
plot(ansvar) # Plot the change in variance
print(ansvar) # Identify changepoint

## Change in mean
ansmean <- cpt.mean(stock.allianz)
plot(ansmean, cpt.col='blue') # Plot the change in mean with blue color
print(ansmean) # Identify changepoints in mean


## Binary Segmentation
bin.segmean <- cpt.mean(stock.allianz, penalty="MBIC", pen.value=0,
   method="BinSeg", Q=10, test.stat="Normal", class=TRUE, param.estimates
   =TRUE, minseglen=2)
summary(bin.segmean) # Summary of mean changepoints using Binary
   Segmentation

bin.segvar <- cpt.var(ret.allianz, penalty="MBIC", pen.value=0, know.mean
   =FALSE, mu=NA, method="BinSeg", Q=20, test.stat="Normal", class=TRUE,
   param.estimates=TRUE, minseglen=2)
summary(bin.segvar) # Summary of variance changepoints using Binary
   Segmentation
```

```
bin.segmv <- cpt.meanvar(stock.allianz, penalty="MBIC", pen.value=0,
   method="BinSeg", Q=2',␣test.stat="Normal",␣class=TRUE,␣param.estimates
   =TRUE,␣shape=1,␣minseglen=2)
summary(bin.segmv)␣#␣Summary␣of␣mean␣and␣variance␣changepoints␣using␣
   Binary␣Segmentation

##␣PELT␣(Pruned␣Exact␣Linear␣Time)
peltmean␣<-␣cpt.mean(stock.allianz,␣penalty="MBIC",␣method="PELT",␣
   minseglen=500)␣#this␣last␣variable␣allows␣to␣turn␣the␣test␣less␣
   sensitive␣to␣changes␣in␣mean␣at␣data␣level,␣and␣provide␣a␣smaller␣
   number␣of␣breaks
summary(peltmean)␣#␣Summary␣of␣mean␣changepoints␣using␣PELT

peltvar␣<-␣cpt.var(ret.allianz,␣penalty="MBIC",␣method="PELT",␣Q=5,␣
   minseglen=500)
summary(peltvar)␣#␣Summary␣of␣variance␣changepoints␣using␣PELT

peltmv␣<-␣cpt.meanvar(stock.allianz,␣penalty="MBIC",␣method="PELT",␣
   minseglen=500)
summary(peltmv)␣#␣Summary␣of␣mean␣and␣variance␣changepoints␣using␣PELT
```

The following lines of code reflect the simulation applied to an ARMA model that was created, in R language, with detailed explanations in each step, that was later used to perform the listed test above regarding unit root.

```
# Define a function called Test_generate that generates synthetic
   financial return data
Test_generate <- function(num_observations, mean_min, mean_max, sigma_min
   , sigma_max, break_point) {
  # Define ARMA coefficients
  ar_coefs <- c(0.2, -0.2)
  ma_coefs <- c(0.3)

  # Generate an ARMA(1,1) model with specified coefficients and standard
      deviation
  arma_model <- list(arima.sim(n = num_observations, list(ar = ar_coefs,
     ma = ma_coefs), sd = sqrt(0.1796)))

  # Randomly set means and standard deviations within specified ranges
  mean1 <- runif(1, mean_min, mean_max)
  mean2 <- runif(1, mean1, mean_max)
```

```
  sigma1 <- runif(1, sigma_min, sigma_max)
  sigma2 <- runif(1, sigma_min, sigma_max)

  # Generate synthetic returns with a structural break at break_point
  returns <- arima.sim(num_observations, model = arma_model, rand.gen =
     function(n) {
    rnorm(n, mean = ifelse(1:n < break_point, mean1, mean2), sd = ifelse
       (1:n < break_point, sigma1, sigma2))
  })

  # Return the generated synthetic financial returns
  return(returns)
}

# Generate synthetic financial returns with specified parameters
returns_arma <- Test_generate(5000, 0, 10, 0.01, 0.5, 2000)

# Plot the generated synthetic financial returns
plot(returns_arma, type = "l", xlab = "Observations", ylab = "Returns")

# Install and load the urca package for unit root tests
install.packages("urca")
library(urca)

# Set the number of iterations
n <- 1000

# Initialize a data frame to store unit root test results
unit_roots <- data.frame(matrix(ncol = 5, nrow = n))
colnames(unit_roots) <- c("ADF", "PP", "kpss", "ERS", "ZA")

# Loop for n iterations to perform unit root tests on generated data
for (i in 1:n) {
  # Generate synthetic financial returns for each iteration
  returns_arma <- Test_generate(5000, 0, 10, 0.01, 0.5, 2000)

  df <- ur.df(returns_arma, lags = 4, type = "drift", selectlags = "BIC")
  pp <-ur.pp(returns_arma, type="Z-tau", model="constant", lags="short")
  kpss<-ur.kpss(returns_arma, type="tau", lags="short")
```

```
ers<-ur.ers(returns_arma, type = c("DF-GLS"), model = c("constant"),
   lag.max = 4)
za<-ur.za(returns_arma, model = "both", lag=NULL)

# Perform ADF test
df <- ur.df(returns_arma, lags = 4, type = "drift", selectlags = "BIC")
tau3 <- attr(df, 'cval')[4]
teststat <- attr(df, 'teststat')[1]
resultdf <- ifelse(teststat < tau3, 0, 1)

# Perform PP test
pp <- ur.pp(returns_arma, type = "Z-tau", model = "constant", lags = "
   short")
c.val <- attr(pp, 'cval')[2]
Zstat <- attr(pp, 'teststat')
resultpp <- ifelse(Zstat < c.val, 0, 1)

# Perform KPSS test
kpss <- ur.kpss(returns_arma, type = "tau", lags = "short")
kpss_crit_val <- kpss@cval[1, 2]
resultkpss <- ifelse(kpss@teststat > kpss_crit_val, 0, 1)

# Perform ERS test
ers <- ur.ers(returns_arma, type = c("DF-GLS"), model = c("constant"),
   lag.max = 4)
ers_crit_val <- ers@cval[1, 2]
resulters <- ifelse(ers@teststat < ers_crit_val, 0, 1)

# Perform ZA test
za.stat <- za@teststat
za.cv <- za@cval[2]
if (za.stat<za.cv) {resultza<-1 # reject the null -> the value
   attributed is "1" of stationary
} else{ resultza<-0}
resultza

# Store the results in the unit_roots data frame
unit_roots[i, 1] <- resultdf
unit_roots[i, 2] <- resultpp
unit_roots[i, 3] <- resultkpss
```

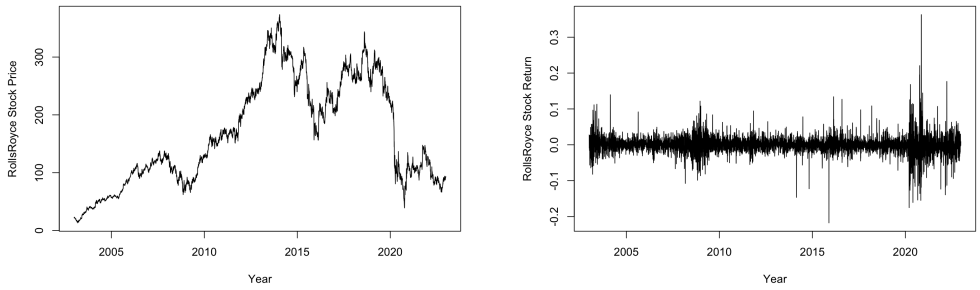
```
  unit_roots[i, 4] <- resulters
  unit_roots[i, 5] <- resultza
}

# Print the proportions of stationary outcomes for each test
cat("Result␣ADF:", sum(unit_roots[1])/n,"Result␣PP:", sum(unit_roots[2])/
   n,"Result␣KPSS:", sum(unit_roots[3])/n,"Result␣ERS:", sum(unit_roots
   [4])/n,"Result␣ZA:", sum(unit_roots[5])/n)
```

Annex B

Results - other conducted studies - RollsRoyce

Visualizations of both series are presented through plots, offering a graphical insight into the patterns and trends within the data:



(a) RollsRoyce daily stock prices, 2003-2022 (b) RollsRoyce daily stock returns, 2003-2022

FIGURE 24. RollsRoyce stock prices and returns

Unit root's tests

Test	Result
ADF Test	Value of test-statistic: -1.7689 Critical values for test statistics: τ^2: -3.43 -2.86 -2.57
PP Test	Value of test-statistic, type: Z-tau: -1.6962 Critical values for Z statistics:10% Level: -2.567396
KPSS Test	Value of test-statistic: 7.4876 Critical values for significance level: 10% Level: 0.119
ERS Test	Value of test-statistic: -0.6771 Critical values for significance level: 10% Level: -1.62
Zivot-Andrews Test	Value of test-statistic: -3.7425 Critical values for significance level: 1% Level: 10% Level: -4.82 Potential Break Point Position: 2239

TABLE 31. Unit Root test statistics - RollsRoyce stock prices

All the required tests point for non-stationarity of the time-series data. The Zivot-Andrews test suggests a structural break at 2239 datapoint, that corresponds to the date of 2011-10-04.

Structural breaks' tests

Chow's tests

Chow's Breakpoint Test Statistic	477.97	**P-Value**	< 2.2e-16
Chow's Forecast Test Statistic	0.6383	**P-Value**	< 1

TABLE 32. Chow's Breakpoint test results - RollsRoyce stock prices

Chow's tests produce contradictory conclusions towards a breakpoint occurring at point 4380 (chosen as explained earlier at Allianz's analysis, as well as by observation), with Chow's Forecast test pointing towards the non rejection of the null hypothesis, implying there is not a significant difference in the model fit before and after the specified point.

CUSUM and MOSUM Tests

As for the CUSUM test, the produces graphs are as follows:

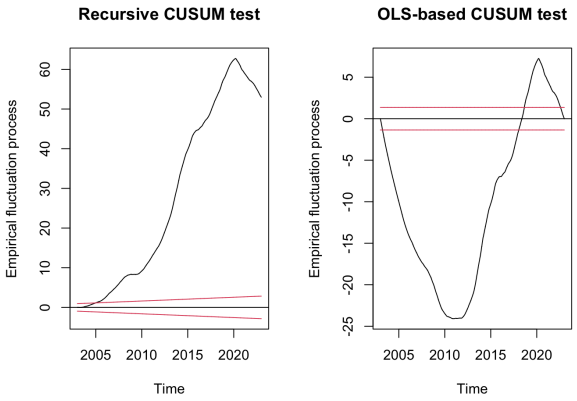


FIGURE 25. CUSUM Recursive and OLS-bases test results - RollsRoyce stock prices

The test statistics corroborate the visual inspection:

	Rec CUSUM	OLS CUSUM
Test Statistic	23.106	24.088
P-Value	2.2e-16	2.2e-16

TABLE 33. CUSUM test results - RollsRoyce stock prices

The test results indicate a significant deviation from the expected pattern in the RollsRoyce stock data. Both p-values are well below significance level, suggesting that there is a substantial structural change in the data.

In terms of the MOSUM results, the produced graphs are as follows:

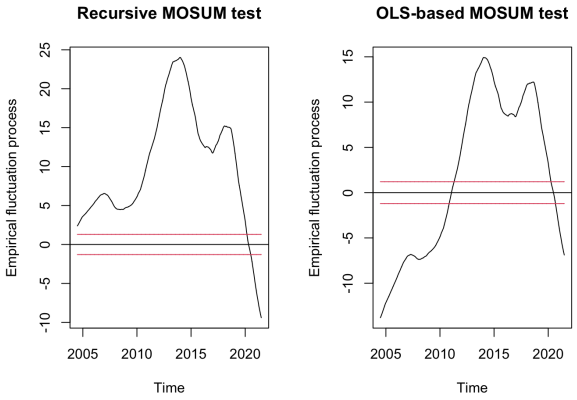


FIGURE 26. MOSUM Recursive and OLS-bases test results - RollsRoyce stock prices

The test statistics corroborate the visual inspection:

	Rec MOSUM	OLS MOSUM
Test Statistic	24.017	14.929
P-Value	0.01	0.01

TABLE 34. MOSUM test results - RollsRoyce stock prices

Both conducted tests reveal a significant deviation from the expected pattern, with the associated p-values indicating evidence against the null hypothesis, under the typical significance level at 5%, suggesting a structural change in the data.

Quandt-Andrews Breakpoint test

The obtained result are:

Test Type	**Test Statistic**	**P-Value**
supF test	4739.3	$< 2.2 \times 10^{-16}$
aveF test	2011.1	$< 2.2 \times 10^{-16}$
expF test	∞	NA

TABLE 35. Quandt Andrews test results - RollsRoyce stock prices

With the highest probability of structural change detected at datapoint 1839, at 2010-03-04, and unexpected result that is not a corroborated as the others by visual inspection.

Bai Perron test

The results are as follows:

Segments $(m+1)$	**0**	**1**	**2**	**3**	**4**
1	-	1839	-	-	-
2	-	-	2245	-	4317
3	761	-	2279	-	4317
4	761	1714	2475	-	4317
5	761	1662	2423	3184	4317

Segments (m)	**0**	**1**	**2**	**3**	**4**	**5**
RSS	41,302,978	21,359,921	9,064,415	6,422,831	5,730,713	4,926,964
BIC	60,149	56,818	52,482	50,750	50,188	49,437

TABLE 36. Optimal $(m+1)$-segment partition and corresponding fit statistics

(1) Breakpoints $(m+1)$ test indicates an optimal segment partition of 6, the results being:
   - For $m = 5$, breakpoints are at observations 761, 1662, 2423, 3184, and 4317, corresponding to the dates of 2005-11-30, 2009-06-24, 2012-06-29, 2015-07-03, and 2019-12-23, respectively.

The analysis suggests that the stock data can be effectively segmented into an optimal number of 6 partitions, each characterized by different statistical properties.

The confidence intervals correspondent to each segment are given by:

	Breakpoints at Observation Number		
	2.5% breakpoints	Breakpoints	97.5% breakpoints
1	759	761	763
2	1658	1662	1663
3	2422	2423	2424
4	3175	3184	3192
5	4316	4317	4318

TABLE 37. Breakpoints and confidence intervals

ICSS of Inclán and Tiao test

The detected changepoints results are:

Index	444	781	1315	1824
Date	2004-09-13	2005-12-28	2008-02-07	2010-02-11
2280	2586	3180	3634	4359
2011-11-30	2013-02-19	2015-06-29	2017-04-11	2020-02-24

TABLE 38. Breakpoints and corresponding dates for the ICSS Test - RollsRoyce stock prices

In graphical terms, these breaks can be observed at returns level:

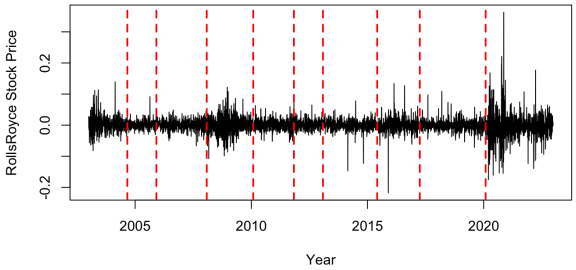
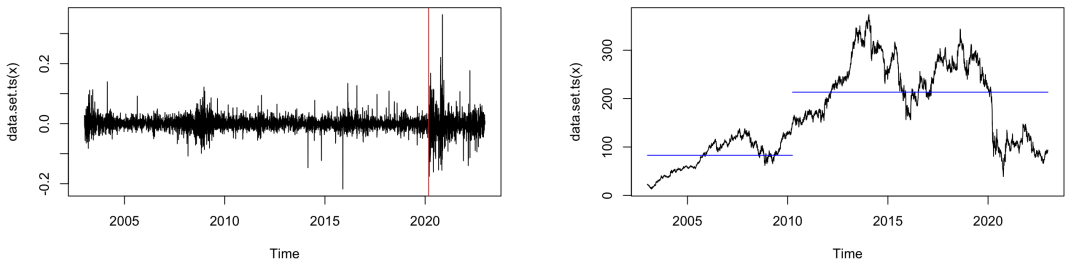


FIGURE 27. ICSS variance test results - RollsRoyce log returns

Changepoint package tests

For single structural changepoint detection, without specifying the exact nature of the data or the statistical methods used, the results are as follows:

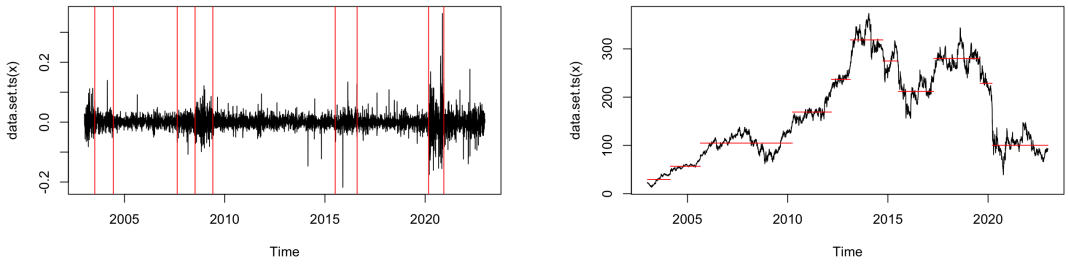


(a) Cpt.var tests' results - log returns (b) Cpt.mean tests' results - stock prices

FIGURE 28. RollsRoyce cpt.var and cpt.mean results

Suggesting a break at variance level at point 4326 that corresponds to date 2020-02-28 and 1839, that corresponds to 2010-03-04 for mean level.

Now, let us address the results under the Binary Segmentation algorithm search, again, allowing for up to 10 structural breaks in data:



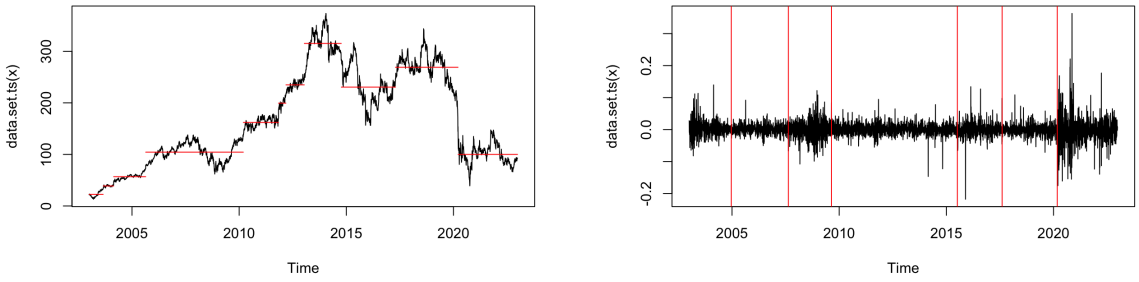
(a) Cpt.var tests' with Binary Segmentation results - log returns

(b) Cpt.mean tests' with Binary Segmentation results - stock prices

FIGURE 29. RollsRoyce cpt.var and cpt.mean with Binary Segmentation results

Again, Binary Segmentation allows for a more suited analysis to the given data other then just cpt.mean and cpt.var functions without further specifications. Yet, the user is free to consider up to as many breaks as he wishes, and with this power comes bigger responsibility.

Finally, the results under the PELT algorithm search, without a maximum number of structural breaks indicated, are as follows:



(a) PELT cpt.mean results - stock prices

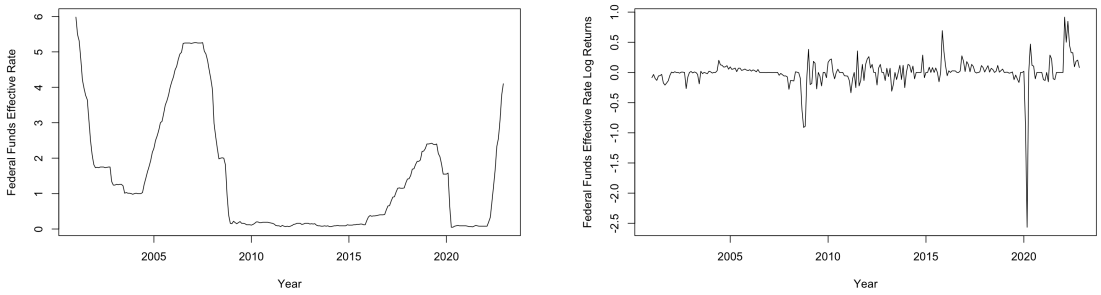
(b) PELT cpt.var results - log returns

FIGURE 30. PELT Tests' results - RollsRoyce

The results are rather close, with PELT providing fewer optimal breaks of the data. By visual inspection, Binary Segmentation seems to have provided more suited results. Yet, there was need to indicate the maximum number of segments to be considered.

Results - other conducted studies - Federal Funds Effective Rate

Visualizations of both series are presented through plots, offering a graphical insight into the patterns and trends within the data:



(a) Federal Funds Effective Rate, monthly rate, 2001-2022

(b) Federal Funds Effective Rate, monthly log returns, 2001-2022

FIGURE 31. Federal Funds Effective monthly rate and log returns

Unit root's tests

The results for unit root's test are as follows:

Test	**Result**
ADF Test	Value of test-statistic: -1.6763 1.443 Critical values for test statistics: τ^2: -3.44 -2.87 -2.57 ϕ^1: 6.47 4.61 3.79
PP Test	Value of test-statistic, type: Z-tau: -2.5269 Critical values for Z statistics: 1% Level: -3.456733 5% Level: -2.872632 10% Level: -2.572632
KPSS Test	Value of test-statistic: 0.3093 Critical values for significance level: 1% Level: 0.216 5% Level: 0.146 10% Level: 0.119
ERS Test	Value of test-statistic: -1.1484 Critical values for significance level: 1% Level: -2.57 5% Level: -1.94 10% Level: -1.62
Zivot-Andrews Test	Value of test-statistic: -5.8907 Critical values for significance level: 1% Level: -5.57 5% Level: -5.08 10% Level: -4.82 Potential Break Point Position: 82

TABLE 39. Unit Root Test Statistics - FFER monthly rate

All the required tests point for non-stationarity of the time-series data, except the Zivot-Andrews that considers the time-series to be stationary even at 1% significance level.

Additionally, the test suggests a structural break at obervation 82, that corresponds to the month of October, 2007, supported by visual inspection. In reaction to deteriorating economic conditions, the Federal Open Market Committee reduced its federal funds rate objective from 4.5 percent at the end of 2007 to 2 percent at the start of September 2008. As the financial crisis and economic downturn worsened in the fall of 2008, the FOMC expedited interest rate reduction, bringing the rate down to its effective floor - a target range of 0 to 25 basis points - by the end of the year. This information is supported by the result Zivot-Andrews' test provided.

Structural breaks' tests

Chow's tests

Similarly to the previous tests conducted in other time series, the chosen point to apply Chow's tests was at observation 229, that corresponds to January 2020, in an attempt to address if the pandemic can also be considered to have an effect similar to a structural break in the data. The choice will not rely on Zivot-Andrews' test suggestion in an attempt to search for other structural breaks (pointed by visual inspection both at rate and log returns level).

Chow's Breakpoint Test Statistic	8.7879	**P-Value**	0.003312
Chow's Forecast Test Statistic	0.691818	**P-Value**	0.9040226

TABLE 40. Chow's Breakpoint test results - FFER monthly rate

The results for Chow's Breakpoint test points towards a breakpoint occuring at this datapoint, yet the Chow's Forecast Test does not imply there is a significant difference in the model fit before and after the specified point, as the p-value is bigger than any significance level to reject the null hypothesis.

If the second test is conducted at observation 82, the result is also for the non-rejection of the null hypothesis, with a p-value of 0.691362.

CUSUM and MOSUM tests

As for the CUSUM test, the produced graphs are as follows:

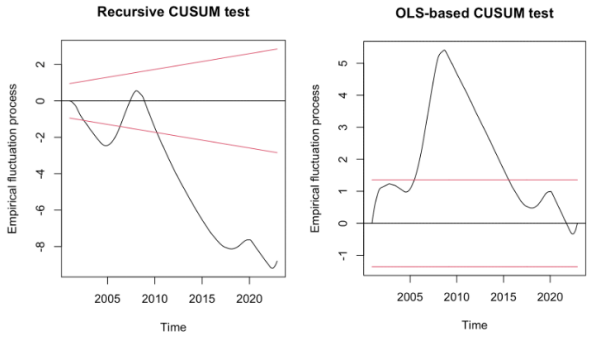


FIGURE 32. CUSUM Recursive and OLS-bases test results - FFER monthly rate

The test statistics corroborate the visual inspection:

	Rec CUSUM	OLS CUSUM
Test Statistic	3.2192	5.4145
P-Value	2.2e-16	2.2e-16

TABLE 41. CUSUM Test Results - FFER monthly rate

The test results indicate a significant deviation from the expected pattern in the Federal Funds Effective Rate, monthly rates. Both p-values are well below significance level, suggesting that there is a substantial structural change in the data.

In terms of the MOSUM results, the produced graphs are as follows:

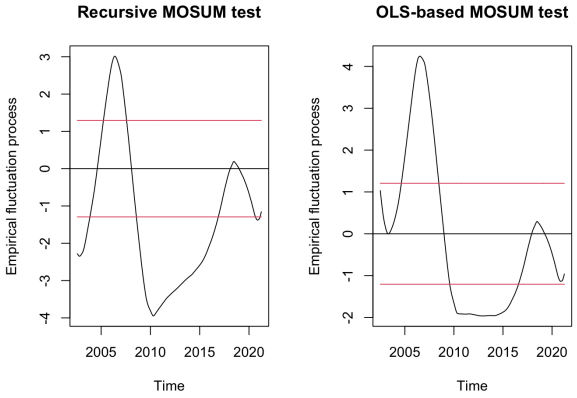


FIGURE 33. MOSUM test results - FFER monthly rate

The test statistics corroborate the visual inspection:

	Rec MOSUM	OLS MOSUM
Test Statistic	3.9471	4.2423
P-Value	0.01	0.01

TABLE 42. MOSUM test Results - FFER monthly ratee

Both conducted tests reveal a significant deviation from the expected pattern, with the associated p-values indicating evidence against the null hypothesis, under the typical significance level at 5%, suggesting a structural change in the data.

Quandt-Andrews Breakpoint test

The obtained result are:

Test Type	Test Statistic	P-Value
supF test	250.26	$< 2.2 \times 10^{-16}$
aveF test	60.669	$< 2.2 \times 10^{-16}$
expF test	120.98	$< 2.2 \times 10^{-16}$

TABLE 43. Quandt Andrews test results - FFER monthly rate

With the highest probability of structural change detected at observation 93, at September 2008. Regarding the explanation provided around the execution of the Zivot-Andrews test, Quandt Andrews provides a suggestion to a structural break aligned with the expectations.

Bai Perron test

The results are as follows:

Segments $(m+1)$	**0**	**1**	**2**	**3**	**4**
1	-	93	-	-	-
2	49	88	-	-	-
3	51	90	-	194	-
4	51	90	-	186	225
5	50	89	128	186	225

Segments (m)	**0**	**1**	**2**	**3**	**4**	**5**
RSS	693.1	354.5	244.0	206.1	200.1	199.2
BIC	1015.2	849.3	761.9	728.5	731.7	741.7

TABLE 44. Optimal $(m+1)$-segment partition and corresponding fit statistics - FFER monthly rate

(1) Breakpoints $(m+1)$ test indicates an optimal segment partition of 4, the results being:
   - For $m = 3$, breakpoints are at observations 51, 90 and 194, corresponding to the dates of March 2005, June 2008 and February 2017, respectively.

The analysis suggests that the data can be effectively segmented into an optimal number of 4 partitions, each characterized by different statistical properties, with both RSS and BIC descending values up to that partition.

The confidence intervals correspondent to each segment are given by:

Breakpoints at Observation Number			
	2.5% breakpoints	Breakpoints	97.5% breakpoints
1	48	51	55
2	89	90	91
3	181	194	195

TABLE 45. Breakpoints and confidence intervals

ICSS of Inclán and Tiao test

The detected changepoints results are:

Index	11	55	88	196	259
Date	November 2001	July 2005	April 2008	April 2017	July 2022

TABLE 46. Breakpoints and corresponding dates for the ICSS test FFER monthly rate

In graphical terms, these breaks can be observed at returns level:

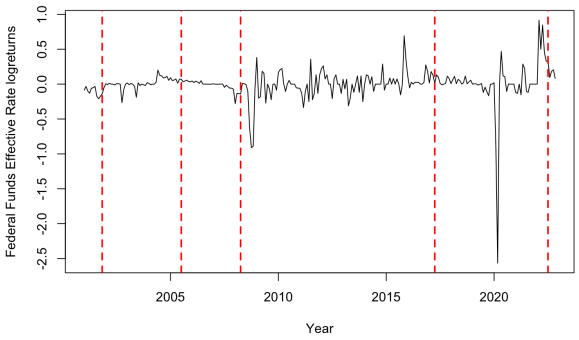
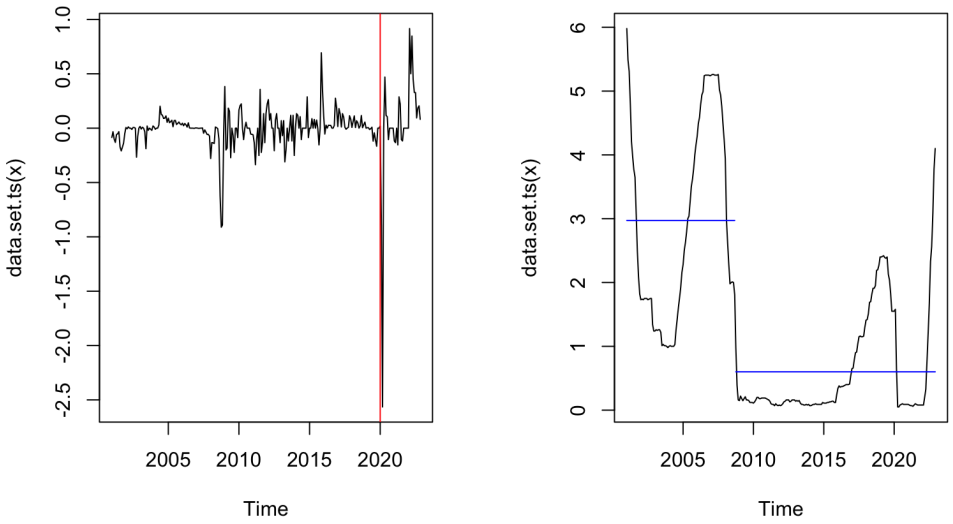


FIGURE 34. ICSS variance test results - FFER log returns

These results are very aligned with the expectations, as they point the disruptions in terms of practitioned rates around the time period being considered.

Changepoint package tests

For single structural changepoint detection, without specifying the exact nature of the data or the statistical methods used, the results are as follows:

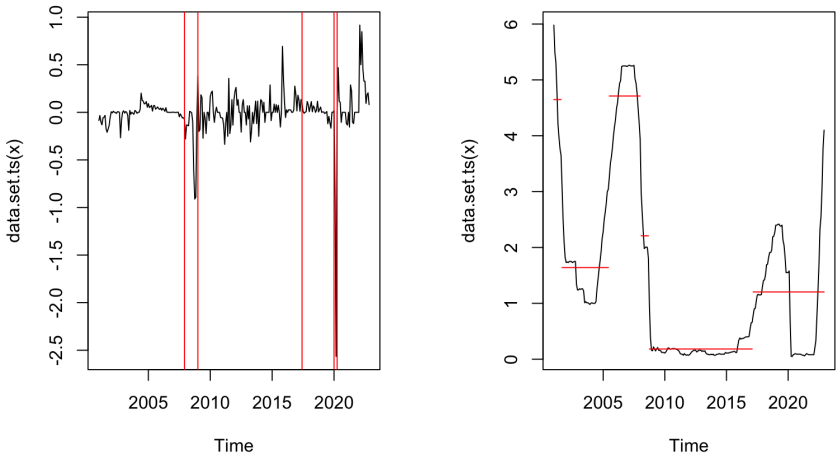


(a) Cpt.var tests' results - log returns (b) Cpt.mean tests' results - monthly rate

FIGURE 35. Federal Funds Effective Rate cpt.var and cpt.mean results

Suggesting a break at variance level at point 229 that corresponds to January 2020 (as anticipated) and 93, that corresponds to September 2008 for mean level, also aligned with the previously explained reasons.

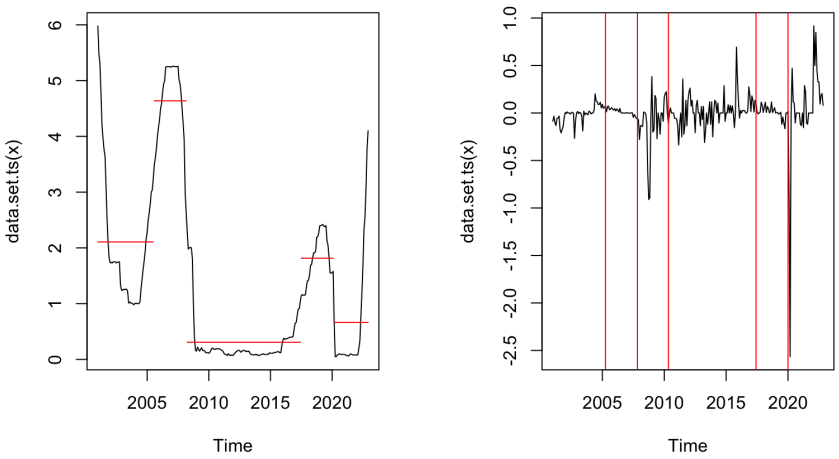
Now, let us address the results under the Binary Segmentation algorithm search, again, allowing for up to 5 structural breaks in data:



(a) Cpt.var tests' with Binary Segmentation results - log returns (b) Cpt.mean tests' with Binary Segmentation results - monthly rate

FIGURE 36. Federal Funds Effective Rate cpt.var and cpt.mean with Binary Segmentation results

Here, at mean level, the Binary Segmentation produced 5 breakpoints (as it typically does) and 5 breakpoints at variance level. Finally, the results under the PELT algorithm search, without a maximum number of structural breaks indicated, are as follows:



(a) PELT cpt.mean results - monthly (b) PELT cpt.var results - log returns rate

FIGURE 37. PELT Tests' results - Federal Funds Effective Rate

The results are really close, and one observation that can lead to these results rather that the type of data is also the number of observations, as this time series accounts only for around 260 entrances. Again, both are pointing towards the same areas, the choice of each test being more related to prior knowledge of data, sample size and execution capacity.