Multiple Breaks Detection in Financial Interval-Valued Time Series

Carmela Cappelli[a,*], Roy Cerqueti[b,c,†], Pierpaolo D'Urso[b‡] and Francesca Di Iorio[a§]

[a] Department of Political Sciences, University of Naples Federico II, Italy

[b] Department of Social and Economic Sciences, Sapienza University of Rome, Italy

[c] School of Business, London South Bank University, London SE1 0AA, UK

Abstract

Multiple structural breaks detection for Interval-Valued Time Series (IVTS) is undoubtedly relevant under practical perspectives and challenging under the point of view of the analysis of expert systems. In this respect, financial time series usually show high variability and outliers; moreover, they often exhibit the property of being of high frequency nature; thus, it is naturally advisable to consider them as IVTS type for a given time unit. Despite this relevance, scarce effort has been spent by scholars to apply the methodological advancements in breaks detection for IVTS to the crucial environment of financial time series. This paper contributes to fill this gap. It employs the Atheoretical Regression Trees framework – a very recent tool that is able to automatically locale multiple breaks occurring to unknown dates – to stock prices. Such a procedure is able to estimate in an efficient way the structural breaks of the considered

* *email*: carcappe@unina.it

†Corresponding author: Roy Cerqueti, Department of Social and Economic Sciences, Sapienza University of Rome. Piazzale Aldo Moro, 5 - I-00185, Rome, Italy, *email:*roy.cerqueti@uniroma1.it.

‡ *email*: pierpaolo.durso@uniroma1.it

§ *email*: fdiiorio@unina.it

series; at the same time, it keeps into account the main characteristics of the intervals describing the IVTS. For our purposes, we adopt a theoretical proposal of reading daily stock prices as intervals whose bounds are defined through the closing prices. Empirical experiments on the American International Group – whose daily prices have experienced structural breaks in the past – validate the theoretical model and show the usefulness of the proposed procedure.

Keyword: Interval-valued time series, multiple structural breaks, Atheoretical Regression Trees, stock prices.

1 Introduction

Since the seminal work of Billard and Diday (2003) on symbolic data, interval-valued data have attracted the attention of the statistical community. Indeed, several real life phenomena are described by data which are naturally of interval type; hence, summarizing the original data into single values may imply a relevant loss of information. This explains the strong and still growing effort spent by academicians and practitioners to account for the interval nature of the data; in particular, a variety of methods that either extend classical approaches or develop specific new ones are still at the core of the scientific debate. Examples can be found in different empirical fields, such as chemometrics (D'Urso and De Giovanni, 2011), ecotoxicology (de Almeida et al., 2013), meteorology (Hajjar and Hamdan, 2013), physics (Groenen et al., 2006), pattern recognition (Douzal-Chouakria et al., 2011), telecommunications (D'Urso and De Giovanni, 2011), economics and finance (Gonzalez-Rivera and Lin, 2013).

Interval-based observations arise often in intra-period time series analysis where, usually, several values are collected at each time period (week, day, hour, even minute); noticeable examples are the time series of the temperatures or the air pollutant concentrations in meteorology; the systolic or diastolic pressure of a patience in medicine; the volatility of an asset in finance. Such cases are associated to phenomena which are measured more than once in a unitary time period. Thus, interval-valued data meet the requirement of being suitable in the context of big data, mainly in presence of high frequency observations.

The perspective of the interval-valued data relies on the context of the fuzzy logic, where numbers are suitably conceptualized to include several aspects of the modeled phenomena. In particular, financial data can be effectively viewed as realizations of expert systems, and they can be described by employing the instruments of the fuzzy logic. In this respect, some relevant literature contributions are worth mentioning.

Agliardi and Agliardi (2009) propose a fuzzy model for defaultable bonds, by fuzzifying the assets value of the considered firms. In so doing, the Authors face the relevant issue of the management of uncertainty when dealing with some types of financial data. Chen et al. (2009) deal with the issue of the identification of possible (fuzzy) trends in financial series, which can be viewed as specific regularities appearing in the considered data. They provide extensive arguments supporting the need of employing a fuzzy logic-based approach for transforming the difference between two consecutive punctual observations into angles. Lai et al. (2009) face the relevant theme of the dependence of stock prices on many factors by transforming prices through a fuzzy decision tree. The Authors implement forecasting exercises in this specific framework. On the same line, Lee et al. (2006) provide a fuzzy model for financial prediction which is able to include the psychology of the financial markets and how it affects investment strategies. Mansour et al. (2019) develop an optimal portfolio model by considering fuzzy parameters for the considered assets. In so doing, the quoted paper includes uncertainty and different sources of information on the main quantities involved in portfolio selection. Under a similar perspective, Rubio et al. (2017) elaborate on the usefulness of fuzzy logic for providing high-quality information to the investors. The Authors support the theoretical methodology by showing the effectiveness of a related stock index forecasting application.

The quoted papers move from the same premise of the present one. Indeed, as we will see in details below, we also provide a fuzzy view of financial time series – assets prices and returns, in this study – for including their most relevant characteristics.

Under a more specific point of view, interval-valued time series (IVTS henceforth) represent the natural extensions of low and high prices of an asset over time, and the intervals – along with their ranges – are strictly connected to volatility because they account for the intra-period price variability. Moreover, high frequency data can be efficiently explored when several observations are synthesized together, as in the data science framework (see e.g. Li et al., 2015 and references therein contained). Actually, financial IVTS's – with specific reference to the stock prices – represent a special case of IVTS's. Indeed, also the last daily price – the so-called "closing price" – is available, besides the extreme values (lowest and highest prices) of the corresponding daily interval. Importantly, the closing price is the most widely used price value in standard financial analysis. Moreover, financial time series are characterized by high variability and presence of outliers; thus, the width of the intervals cannot be ignored, as it conveys information on both such aspects.

Most of the financial literature focuses, as expected, on the crucial task of forecasting or on properly clustering IVTS's. Arroyo et al. (2011) discuss different approaches for predicting the IVTS associated to daily prices. Maia et al. (2008) use a hybrid Box-Jenkins and ANN model for improving the forecasting outcomes of financial prices. Rodriguez and Nazarii (2015) employ threshold models for forecasting financial IVTS. As in the Maia et al. (2008)'s approach, also Rodriguez and Nazarii (2015) provide evidence that a combination of models is more effective than the individual ones for having good forecasting performances. Following the same research line, Xiong et al. (2017) deal with a combination of models for making prediction of financial IVTS. We particularly appreciate the empirical instances related to energy prices of Xiong et al. (2017); indeed, it is out of doubts that the price of energy is a challenging empirical test, in that such series presents hard features like spikes, seasonality and volatility clustering. The four papers mentioned above share the same strategy for defining a IVTS; in details, they present a very simple procedure to transform punctual data into intervals by considering the difference between the highest and the lowest intra-period values. In a very different way, Wang et al. (2019) propose a clustering procedure of IVTS by taking into account also the possibility of an IVTS not defined as such a difference. The Authors present a stepwise algorithm for the definition of IVTS which leads to the inclusion of the time and space characteristics of the phenomenon under investigation.

In defining the IVTS of the daily prices, we follow the idea of Wang et al. (2019) and adopt the definition of Han et al. (2008). The IVTS is built by considering not only highest and lowest intra-day prices, but also the closing daily price. For a discussion on the conceptualization of the IVTS, we refer to the comments to the formal definition provided in Section 2.

For what concerns the financial target of the paper, we add to the financial modeling of the IVTS's while, at the same time, departing from the existing literature. Specifically, we here face the problem of change point detection. Indeed, as acknowledged also by authoritative studies, breaks detection can be important in the context of predictions, mainly on a short-term horizon (see e.g. Bardwell et al., 2019).

Change point analysis comprises various tools developed to determine if and when one or more structural breaks have occurred in a data set; thus, in the last decades, it has emerged as a relevant research topic whose usefulness can be easily appreciated under several perspectives. First, change point analysis can reveal specific behaviors of the time series, that could otherwise be misunderstood and modeled inadequately; in this respect, it is well known that a process with structural breaks can be easily confused with a long memory process (see Granger and Hyung, 2004). Second, as already mentioned above when referring to the context of forecasting, change point detection might serve as a sort of preliminary data treatment; indeed, in the peculiar case of long time series, a model estimated on a recent segment of the series might provide more accurate forecasts on the future evolution of the phenomenon, especially when the considered segment is free from breaks (see e.g. Choi et al., 2008 and the recent contribution of Moews et al., 2019 for the specific case of the analysis of the stock markets). Eventually, the identification of breaks might reveal the presence of outliers and thus the need for adjusting the data see e.g. Cappelli et al., 2008.

The most challenging task in change point analysis is the detection of multiple changes occurring at unknown dates; in this context, the most relevant contributions are due to Bai and Perron (2003 and 2006), but see also Billard and Diday (2003). In this respect, a change in persistence might be associated to a change point, as discussed in Busetti and Taylor (2004), Kim (2000), Kim et al. (2002), Harvey et al. (2006) for the case of time series and, more recently, by Cerqueti et al. (2016) in the general panel data case.

In the same spirit of dealing with multiple changes, Cappelli et al. (2008) proposed a method called Atheoretical Regression Trees (ART). In the quoted paper, extensive simulation studies, comparison with other widely employed methods and applications to various real time series provide evidence of the usefulness of the approach (see Rea et al., 2010). Moreover, such a methodological device has been successfully extended to analyze imprecise, vaguely observed time series by using a fuzzy approach (see Cappelli et al., 2013). Thus, ART is particularly suitable for dealing with IVTS. This statement represents the methodological scientific ground of the present paper, where

ART has been employed for detecting multiple structural breaks of financial IVTS.

To pursue this scope, we firstly present the daily series of the prices of a financial asset as a financial IVTS. In particular, as already preannounced above, we follow the approach proposed by Han et al. (2008) and define the lower and upper bound of the time unit observations (intra-period lowest and highest prices of a financial asset) as a function of the closing value. Then, we adopt the framework of ART and consider the distance based on the deviance measure proposed by Cappelli et al. (2015); in so doing, we account for the width of the intervals.

Under a purely financial perspective, our approach shows important advantages and novelties. Firstly, it is tailored to conduct change point analysis of financial IVTS by considering all the relevant values (daily extremes and closing value); this is a real improvement with respect to the standard methods, which analyze only single values (e.g. mean or extreme values). Moreover, we are far from the relevant contribution by Cappelli et al. (2015) – where the change point assessment procedure of IVTS adopted here is implemented to the air pollution context – either for what concerns the target of the study as well as for the reference literature. Indeed, we here present a very novel approach to the financial breaks detection theory, since the standard financial breaks detection framework is associated to the cases of time series or panel data. Furthermore, the nonparametric nature of the tree methods inherited by ART is considered as a main advantage in the framework of expert systems and machine learning, since the procedure is entirely data driven; moreover, the recursive partitioning approach tends to separate small groups of observations from large ones, hence revealing the possible presence of outliers. In so doing, the procedure is able to certify the need for adjusting the data (see e.g. Cappelli et al., 2008). Finally, the model-free property of the considered approach lets it be suitable in contexts of nonstandard distribution assumptions for the prices and in the very challenging framework of the commodity markets. In fact, commodities prices are characterized by spikes, seasonality and mean reversion properties; hence, they cannot be modeled through standard distributions of Gaussian type.

To demonstrate the effectiveness of the proposed device, we present an empirical experiment on the stock prices of a very troublesome asset: the American International Group (AIG), over the period Jan 2005 – Dec 2018. In so doing, we are in line with a long list of authoritative literature contributions which test methodological proposals over one reasonable empirical dataset (see e.g. Park et al., 2014 and Zhang et al. 2009).

In this respect, it is worth mentioning the specific econometric context of structural breaks detection, where several crucial papers have developed new test statistics and validated them by means of a unique paradigmatic empirical application. As an example, Leybourne et al. (2003) use the US quarterly inflation rate over a period ranging from 1959 to 2000 for testing their change in persistence assessment procedure. The quoted paper is in line with other ones when using US inflation rate for testing structural breaks detectors (see e.g. Kim, 2000 and Busetti and Taylor, 2004). The selection of the US inflation rate as ground of empirical experiments on structural breaks detection is driven by the statistical characteristics of such a macroeconomic variable. In particular, there is evidence of a real change in persistence of the US inflation rate in the seventies, so that inflation rate switched from stationarity to the presence of unit root.

The present paper is highly indebted to the econometric perspective of empirically validating a breaks detection method. Indeed, under the guide of all the papers quoted above, we have properly selected the AIG dataset, in that it represents in our context one of the most relevant empirical settings – hence, being particularly effective for supporting the worthiness of our methodological advancement – for some reasons. Firstly, AIG has been for a long period the most important insurance company and – in general – one of the biggest companies in the world. At the peak of its success, its credit ranking was AAA and its business activity involved more than 130 countries with more than 70 million customers. This explains why such a giant company has been at the center of the high-level scientific debates on business studies (see e.g. McDonald and Paulson, 2015 and Jin et al., 2015). Secondly, AIG has experienced some clearly visible breaks during the analyzed time period, including the financial crash of 2008. For this reason, such a Group has been often at the center of the debate on systemic risk and crisis (see e.g. MCDonald and Paulson, 2014 and the long list of their followers). Therefore, AIG can be viewed as a noticeable benchmark for empirically illustrating the effectiveness of the presented methodology by offering relevant material for dealing with structural breaks detection.

The remainder of the paper is organized as follows. In Section 2, we theoretically build the financial the IVTS we are dealing with. Section 3 briefly provides the necessary background on multiple change points detection; furthermore, it recalls the ART method by illustrating how it can be employed to detect change points in financial IVTS's. In Section 4, we present an empirical experiment by implementing the change point analysis of the stock prices of the American International Group (AIG) – which is a very troublesome asset that was a central player during the 2008 financial crisis. Such a section contains also a subsection which provides some arguments pointing to possible explanations of the structural breaks. Last Section offers some conclusive remarks.

2 Financial IVTS

Given a nonnegative integer T, a sample of length T of an IVTS can be seen as a sequence $(\widetilde{Y}_t : t \in 1, \dots, T)$ such that $\widetilde{Y}_t = [\tilde{l}_t, \tilde{u}_t], \quad t = 1, \dots, T$, i.e. each time unit is characterized by a lower and an upper bound $\tilde{l}_t$ and $\tilde{u}_t$ respectively. A time unit t corresponds then to a time period where some pointwise observations are available. We here intend the $\widetilde{Y}$'s as stock prices.

We follow the arguments developed by Han et al. (2008) and Wang et al. (2019), which reasonably state that the lowest and the highest intra-day prices might be not reliable devices for the description of the daily prices. Indeed, the intra-day fluctuations of the daily prices might be generated by market shocks of different nature, which are exceptional events to be not included in a reasonable model of the daily prices. Therefore, while evidently offering the maximum level of information in an interval-valued series context, the minimum and maximum daily prices may be also viewed as a source of bias for modeling purposes.

Furthermore, the closing price brings a crucial information on the intra-day series of the daily prices. Indeed, even if a standard trading day involves several investors, empirical evidence suggests that the most part of them do not trade usually more than once on the same stock in a single day; thus, the closing prices are those related to the widest part of positions in the market. This explains also the popularity of the closing prices for describing financial time series.

Therefore, we here follow the approach of Han et al. (2008) and employ the daily closing prices for building the IVTS of the prices. Specifically, the lower and upper bounds $\tilde{l}_t, \tilde{u}_t$ are defined as functions of the lowest (l_t), the highest (u_t) and the closing price (denoted by c_t) at the time period labeled by t. In particular, if $u_t \neq l_t$ we set

$$\left[\tilde{l}_t = c_t - \frac{c_t - l_t}{u_t - l_t}(c_t - l_t), \;\; \tilde{u}_t = c_t + \frac{u_t - c_t}{u_t - l_t}(u_t - c_t)\right], \tag{1}$$

and we adopt the financially reasonable conventional agreement that $u_t = l_t$ implies $\tilde{u}_t = \tilde{l}_t = u_t = l_t = c_t$.

Formula (1) suggests that the definition of the upper and lower bounds of the intervals is given as a weighted mean of the lowest and closing price (lower bound) and highest and closing price (upper bound). The weights are conceptualized to give credit jointly to the presence of large deviations in intra-day prices and to the value of the closing prices. Specifically, they are given by the relative distance between the closing price and the highest/lowest ones with respect to the overall variation range between the highest and the lowest prices. Indeed, we can rewrite (1) as follows:

$$\tilde{l}_t = \omega_1(t) l_t + \omega_2(t) c_t; \qquad \tilde{u}_t = \omega_1(t) c_t + \omega_2(t) u_t, \tag{2}$$

where

$$\omega_1(t) = \frac{c_t - l_t}{u_t - l_t}; \;\; \omega_2(t) = \frac{u_t - c_t}{u_t - l_t}. \tag{3}$$

Formula (2) – with weights given in (3) – gives that the upper and the lower bounds of the intervals are driven by the relationship between the highest price u_t and the lowest price l_t with the closing price c_t. In particular, we have that $c_t = l_t$ implies $\tilde{u}_t = u_t$, while $c_t = u_t$ implies $\tilde{l}_t = l_t$.

Thus, the standard case of an IVTS defined through the difference between the highest and the lowest price is a subcase of our conceptualization of IVTS, which then brings a more relevant information content on the considered financial quantity.

Then, as a counterpart of standard return series, the IVTS of the returns can be constructed by setting $\tilde{\tilde{Y}}_t = \left[\tilde{\tilde{l}}_t, \tilde{\tilde{u}}_t\right]$, for each $t = 1, \ldots, T$, where

$$\tilde{\tilde{l}}_t = \ln\left(\frac{\tilde{l}_t}{\tilde{u}_{t-1}}\right), \;\; \tilde{\tilde{u}}_t = \ln\left(\frac{\tilde{u}_t}{\tilde{l}_{t-1}}\right). \tag{4}$$

Note that each interval return can be represented by a vertical segment defined by the *center* (midpoint) $m_t = \frac{\tilde{\tilde{l}}_t + \tilde{\tilde{u}}_t}{2}$ and *radius* (spread) $r_t = \frac{\tilde{\tilde{u}}_t - \tilde{\tilde{l}}_t}{2}$ of the interval. Thus, the IVTS of the returns can be rewritten in terms of centers and radii $\tilde{\tilde{Y}}_t = [m_t, r_t]$, for each $t = 1, \ldots, T$.

3 ART for financial IVTS

We now recall the main features of the ART procedure introduced in Cappelli and Reale (2005), to illustrate how it can be applied to the financial IVTS described in the previous Section.

ART exploits the idea underlying the Least Square Regression Trees (LSRT, for short). LSRT is a method to model the relationship between a response variable and a set of covariates in the form of a binary tree. In details, every node – which represents a group of observations – is divided into two subgroups, i.e. the left and the right child nodes. Such subgroups are subsets of observations which are more homogeneous – in terms of the response variable – than the father node. In LSRT, the covariate space is partitioned into regions (the tree nodes) and a constant is fitted within each region/node; this constant is the mean of the response values in the given node; therefore, LSRT can be seen as piecewise constant regression models.

Specifically, given a numerical response variable Y and a set of covariates $(X_1, \ldots, X_p)$ observed on a sample of N units, the binary tree arises by recursively splitting the training set $(y_i, x_{i1}, \ldots, x_{ip})_{i=1}^n$, into two subsets. The splitting procedure is implemented by choosing at any internal node h the best split – i.e., the binary division – of the current node, where the optimization criterion is given by the maximization of the reduction in deviance

$$\Delta SS(h, t) = SS(h) - [SS(h_l) + SS(h_r)], \tag{5}$$

where $SS(h) = \sum_{y_i \in h}(y_i - \hat{\mu}(h))$ is the sum of squares for node h, and $SS(h_l)$ and $SS(h_r)$ are the sums of squares for the left and right child nodes h_l and h_r, respectively. The procedure is then recursively applied to the descendants until a specific stopping rule is met, which in our case means that we are in presence of a minimum level of node deviance or in case of minimal node size (see Breiman et al., 1984 for further details).

In the case of ART, the response variable is a time series Y_t and a single artificial covariate – which is a sequence of strictly ascending numbers $K = 1, \ldots, T$ – is employed in the partitioning process. By tree regressing Y_t on K, the procedure returns a partition of Y_t into G contiguous segments $\mu_1, \ldots, \mu_G$ such that $\sup \mu_g = \inf \mu_{g+1}$ with $g = 1, \ldots, G$. The points separating the segments are labeled as split points, and the segments can be viewed as regimes. Thus, by construction, the procedure can be successfully employed in time series analysis to detect multiple changes in mean occurring at unknown dates; specifically, the split points in the tree diagram identify the set of estimated break points (see e.g. Cappelli et al., 2008 and Rea et al., 2010).

Upon the choice of a proper deviance measure, ART can be employed to locate multiple changes in non standard time series. In particular, as in Cappelli et al. (2015), for an IVTS $(\widetilde{Y}_t : t = 1, \dots, T)$ the deviation is derived by the Euclidean distance proposed in the context of time series fuzzy clustering by D'Urso et al. (2015) for comparing interval-valued data:

$$d(t, t') = \sqrt{(m_t - m_{t'})^2 + (r_t - r_{t'})^2}. \tag{6}$$

The conceptualization of distance in (6) provides a combined information on the considered interval-valued time units, by taking into account their centers (midpoints m's) and widths (radii r's).

Then, the deviation of the entire time series of the interval returns over the given sample period is defined as

$$SS(\widetilde{\widetilde{Y}}_t) = \sum_{t=1}^{T} [(m_t - \bar{m})^2 + (r_t - \bar{r})^2],$$

with $\bar{m}$ and $\bar{r}$ representing the mean values of the midpoints and radii of the interval returns, respectively.

For a generic node h in the ART's recursive partitioning approach, the best binary division minimizes

$$SSR(h_l) + SSR(h_r) = \sum_{i \in \{l,r\}} \sum_{t=1}^{T(h_i)} [(m_t - \bar{m}(h_i))^2 + (r_t - \bar{r}(h_i))^2], \tag{7}$$

where $T(h_i)$ denotes the length of the subseries associated to node h_i, $\bar{m}(h_i)$ is the mean value of the return centers and $\bar{r}(h_i)$ is the mean value of the return radii for the same node.

Since the procedure tends to grow a large tree, a pruning CART-like method is adopted for the generation of a sequence of nested subtrees, each of them providing an alternative set of estimates of the change points models. A standard approach for selecting an optimal model – in the present case, the partition and the set of change points – is usually based on information criteria. In this respect, we here consider the modified Bayesian Information Criterion (BIC) described in Cappelli et al. (2015); for a partition into m segments, such a criterion is defined as follows:

$$BIC(m) = \ln \hat{\sigma}^2(m) + p \ln(T)/T,$$

where $\hat{\sigma}^2(m) = T^{-1} SSR(\tilde{\tilde{Y}}_t)$ is the sum of the squared residuals of the m-partition of the IVTS and $p = (m+1) \times (q+1)$. In our specific framework, we set $q = 2$, in that two parameters (mean center and radius) are estimated in each regime. In so doing, the BIC values associated with the identified subtrees – which represent here the change point models – can be visualized by plotting them against the number of terminal nodes. Then, one can select the optimal subtree according either to the minimum BIC or by the visual inspection of the graph by seeking – for instance – for an elbow point.

4 Empirical experiments

The approach described in Section 3 has been employed for conducting change point analysis on the time series of daily prices of the American International Group (AIG) collected from January 3rd, 2005 to December 18th , 2018. The series is freely available at https://finance.yahoo.com/quote/AIG?p=AIG.

AIG represents a meaningful instance for our scopes and multiple break points are expected. Indeed, it is an American multinational insurance corporation, which operates through three businesses: insurance products for commercial, institutional and individual customers, life insurance and retirement services in USA, and mortgage guaranty insurance and mortgage insurance. Thus, the AIG was a central player in the financial crisis of 2008 as it was bailed out by the federal government for 180 billion dollars, and the government took control of it.

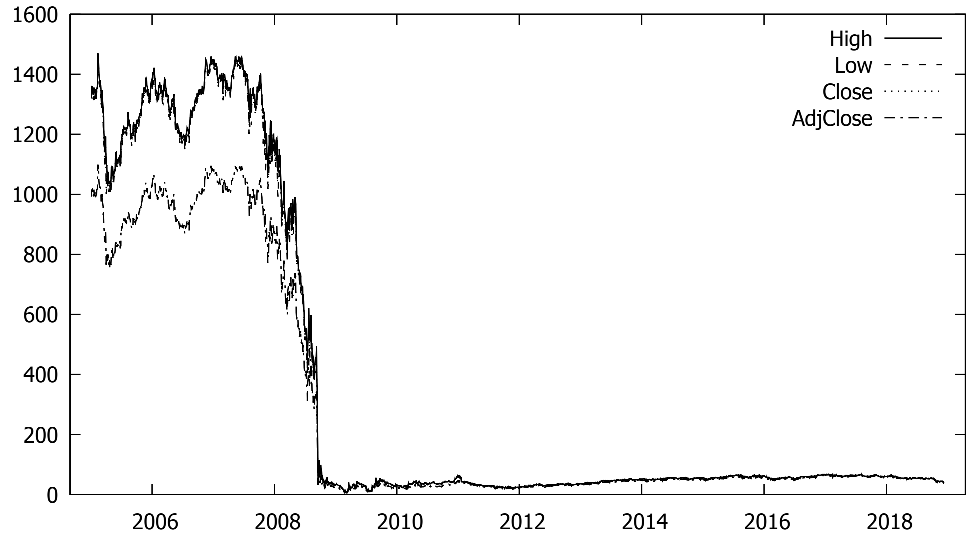


Figure 1: AIG: High, Low, Close and Adjusted Closing daily prices. The reference period is January 3rd, 2005 – December 7th, 2018.

Share prices referring to High, Low, Close and Adjusted Closing are plotted in Figure 1. Adjusted Closing price is presented only for the sake of completeness. Indeed, following Han et al. (2008), we do not consider the Adjusted Closing in our context of financial IVTS since it accounts for all corporate actions such as stock splits, dividends/distributions and rights offerings; such terms might lead to price values in period t outside the range $[\tilde{l}_t, \tilde{u}_t]$.

The upper panel in Figure 2 displays the usual returns of the AIG daily prices obtained form the log-difference of the adjusted close and the absolute range (high-low), while the lower panel shows the interval return defined according to formula (4). As we can see, the behavior of the interval return series mimics that of the standard returns.

In the ART framework, we set a minimum number of observations per segment of seven days; then, on the sequence of pruned subtrees, the modified BIC – whose values against the number of terminal nodes are plotted in Figure 3 – has been computed. As we can see, the curve becomes almost flat between two and five change points after a strong reduction corresponding to the split of the root nodes; then, the values begin to rise. We selected the partition with five changes whose BIC is slightly below that corresponding to the elbow point (only one break) as the final one.

The detected changes are highlighted by the vertical dashed lines in Figure 4, which displays the midpoint and the radius, e.g. the interval-valued series $\widetilde{\widetilde{Y}}_t$; the corresponding change point model is plotted in the form of a tree diagram, whose split points report the break dates in Figure 5. Stylized facts related to the entire series and the sub-periods identified by the change points are reported in Table 1.

The first and strongest change point detected by the procedure occurs at the end of 2011 (December 19th, 2011) and it roughly separates the period before and after the crisis. Further changes are detected during the years of the crisis; such breaks isolate periods characterized by different degrees of instability (high variability) and outliers, specifically at June 13th, 2008 and September 24th, 2009, respectively. A further break is located at the beginning of the crisis in October 23rd, 2007, which then acts as isolating the financial pre-crisis period. Similarly, another break is detected at June 18th, 2013, that separates a short post-crisis period, still characterized by a relatively high variability.

In order to highlight the higher ability of the interval-based approach with respect to point-based analysis for providing a rich description of the data, we have also applied two standard time series procedures – the baseline ART and Bai and Perron's methods – to the transformed series of the centers. In both cases a single break occurring at the beginning of March, 2009 (March 3rd, 2009) – that roughly corresponds to the change clearly appearing in Figure 1 – is detected. To further validate the supremacy of the interval-valued time series with respect to the point-based ones in this special context, the application of either baseline ART or Bai and Perron's procedures to the highest, lowest and closing prices in the original time series has been implemented. Also in these cases, both methods identify a single break located at June 20th, 2008, June 17th, 2008 and June 16th, 2008, respectively. It is worth noting that a very close break-point $13 - Giu - 2008$ is also identified by our procedure as a split of node 2 in Figure 5 (see also Figures 1 and 4).

Differently, the modified ART procedure provides the existence of multiple breaks. This means that the ART procedure for interval-valued data leads to a richer description of the phenomenon under investigation than those provided when the standard point-based methods are used.

Table 1: Stylized facts

	$\bar{c}$	$\bar{r}$	sd_c	sd_r
Entire series				
03-Jan-2005:07-Dec-2018	-0.00102	0.022	0.038	0.033
Regimes				
03-Jan-2005:23-Oct-2007	-5.e-5	0.010	0.010	0.006
24-Oct-2007: 13-Jun-2008	-0.004	0.027	0.028	0.011
13-Jun-2008: 24-Sep-2009	-0.008	0.114	0.089	0.073
25-Sep-2009: 19-Dec-2011	-0.001	0.028	0.030	0.046
20-Dec-2011: 18-Jun-2013	0.002	0.016	0.017	0.007
19-Jun-2013: 11-Dec-2018	-1.e-4	0.012	0.011	0.005

As we can see from Table 1, the mean center, the radius and their standard deviations vary across the sub-periods also with respect to the entire period covered by the series, hence confirming that the detected change points identify quite different regimes.

Further insights on the procedure can be obtained by looking at the behavior of the objective function over the segments induced by the detected change points (see Figure 6). Indeed, if we compare the upper left side panel in Figure 6 with the single break detected by the baseline ART and Bai and Perron's procedure in Figure 7, we notice that the behavior of the objective function in the first case is quite irregular, with some peaks and not so well defined minima. This circumstance is due to the presence of the the radii in the computation of the deviance within the sub periods, which accounts for variability and possible presence of outliers.

Eventually, it is worth noting the different scales of the vertical axis in the three central sub-periods, which corresponds to changes at June 13th, 2008, December 19th, 2011 and September 24th, 2009, where the sum of the squares is an order of magnitude greater than the one of the other two sub-periods before and after the crisis.

4.1 Some financial comments on the breaks

It is interesting to observe also some peculiar occurrences, which seem to link the break dates detected by the ART procedure and the real life of the AIG company. We here advance a related discussion, on the basis of historical facts regarding the considered empirical instance. The analysis here presented can be seen as the starting point of a full exploration of the breaks. Indeed, the complete investigation of them is out of the scopes of the paper and deserves more detailed studies.

For the breaks in 2007-2008, we can refer to McDonald and Paulson (2015). The AIG company has experienced a severe financial distress in the second semester of 2008; such a negative outcome can be reasonably seen as announced by the structural break in June 2008. Importantly, two days after the break in 2008 – i.e., on June 15th, 2008 – there has been the resignment of the CEO of the AIG, and the consequent appointment of Robert B. Willumstad as new CEO. The reasons behind such a change can be found in the sudden disclosure of financial losses and the loss of trustfulness of the CEO of the company. Before this, the company had a continuous expansion which culminated with the peak in the third quarter of 2007. Such an expansion is likely the driver of the subsequent break in October 2007. On August 3rd, 2009 – quite close to the break in September 2009 – there has been a change of the CEO of the AIG. The new CEO was Robert Benmosche, who is reputed as the man driving the AIG to a period of financial prosperity. Moreover, exactly on September 2009, the Wall Street Journal came up with very positive news about the financial strategies implemented by the Pacific Century Group and involving the AIG. In November 2011 – close to the break of December – the value of the AIG shares collapsed as a consequence of the losses experienced in the latest part of that year. In this context regarding the AIG company, one has also to bear in mind the contagion effects of the global financial crisis of 2007 – with the Dow Jones hitting its negative peak on October 2007, close to the 2007 structural break – or the sovereign debt crisis, whose effects in 2013 might suggest a reason explaining the break in June of that year.

5 Concluding remarks

This paper has addressed the problem of detecting multiple change points occurring at unknown times in financial interval-valued time series. In particular, we have shown the positive features of the breaks detection in the case of IVTS when dealing with financial time series.

Our proposal employs a distance metric that accounts for the width of the intervals – hence, reflecting the variability and the presence of outliers – whose bounds are expressed as a function of the closing price. The reference framework is the one of the Atheoretical Regression Trees.

The procedure has been empirically tested over a challenging financial time series: the daily prices of the AIG stock. Results have shown that the proposed methodology is effective in locating changes and isolating the periods characterized by high variability and outliers.

In general, we here contribute to the debate on the change point analysis as a useful tool for monitoring and control which – when referring to the IVTS's – needs appropriate methodological advancements to account for the interval structure of the time ordered units and to avoid the loss of relevant changes. In the special case of the financial IVTS of the prices, a suitable procedure requires to consider the closing price besides the width of the intervals. In so doing, such a procedure reflects the high variability and the presence of outliers usually shown by these types of time series.

Moreover, the proposed procedure is advantageous from several points of view: it is model free – i.e., no parametric model is assumed within the regimes – automatic, fast, and it can be easily implemented in any software that grows classification and regression trees. Thus, the considered procedure is particularly useful for applied time series analysis when several time series are handled.

Ongoing research involves the extension of the procedure to the analysis of co-breaking in financial IVTS's.

References

Agliardi, E., Agliardi, R. (2009). Fuzzy defaultable bonds. *Fuzzy Sets and Systems*, **160**(18), 2597–2607.

de Almeida C.W.D., de Souza R.M.C.R., Candeias A.L.B. (2013). Fuzzy Kohonen clustering networks for interval data. *Neurocomputing*, **99(1)**, 65–75.

Arroyo J., Espínola R., Maté C. (2011). Different approaches to forecast interval time series: a comparison in finance. *Computational Economics*, **37**, 169–191.

Bai, J., Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, **18**, 1-22.

Bai, J., Perron, P. (2006). Multiple structural change models: a simulation analysis. *In*: M. Broy, E. Dener, (Eds), Econometric theory and practice: frontiers of analysis and applied research, 212–237. Cambridge University Press.

Bardwell L., Fearnhead P., Eckley I. A., Smith S., Spott M. (2019). Most recent changepoint detection in panel data. *Technometrics*, **61**(1), 88–98.

Billard, L., Diday, E. (2003). From the statistics of data to the statistics of knowledge: symbolic data analysis. *Journal of the American Statistical Association*, **98**, 470-487.

Breiman L., Friedman J.H., Olshen R.A., Stone C.J. (1984). *Classification and Regression Trees*. Wadsworth & Brooks: Monterey (CA).

Busetti F., Taylor R.A.M. (2004). Tests of stationarity against a change in persistence. *Journal of Econometrics*, **123**, 33–66.

Cappelli C., D'Urso P., Di Iorio F. (2013). Change point analysis of imprecise time series. *Fuzzy Sets & Systems*, **225**, 23–38.

Cappelli, C., D'Urso, P., Di Iorio, F. (2015). Regime change analysis of interval-valued time series with an application to PM10. *Chemometrics and Intelligent Laboratory Systems*, **146**, 337-346.

Cappelli C., Penny R., Rea W. , Reale M. (2008). Detecting multiple mean breaks at unknown points in official statistic. *Mathematics and Computers in Simulation*, **78**, 351–356.

Cappelli, C. and Reale, M. (2005). Detecting changes in mean levels with atheoretical regression tree. *In*: Provasi C. (ed), S.Co.2005: Modelli Complessi e Metodi Computazionali Intensivi per la Stima e la Previsione, 479-484. Cleup, Padova.

Cerqueti R., Costantini M., Gutierrez L., Westerlund J. (2016). Panel stationary tests against changes in persistence. *Statistical Papers*, https://doi.org/10.1007/s00362-016-0864-6.

Chen, C.H., Hong, T.P., Tseng, V.S. (2009). Mining fuzzy frequent trends from time series. *Expert Systems with Applications*, **36**(2), 4147–4153.

Choi H., Ombao H., Ray B. (2008). Sequential change-point detection methods for nonstationary time series. *Technometrics*, **50(1)**, 40–52.

D'Urso P., De Giovanni L. (2011). Midpoint radius self-organizing maps for interval-valued data with telecommunications application. *Applied Soft Computing*, **11**, 3877–3886.

D'Urso, P., De Giovanni, L., Massari, R., (2015) Time series clustering by a robust autoregressive metric with application to air pollution. *Chemometrics and Intelligent Laboratory Systems*, **141**, 107–124.

Douzal-Chouakria A., Billard L., Diday E. (2011). Principal Component Analysis for Interval-Valued Observations. *Statistical Analysis and Data Mining*, **4**, 229–246.

Gonzalez-Rivera G., Lin W. (2013). Constrained Regression for Interval-Valued Data. *Journal of Business & Economic Statistics*, **31(4)**, 473–490.

Granger, C.W.J., Hyung, J. (2004). Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance*, **11**, 399–421.

Groenen P.J.F., Winsberg S., Rodrguez O., Diday E. (2006). I-Scal: Multidimensional scaling of interval dissimilarities. *Computational Statistics & Data Analysis*, **51**, 360–378.

Hajjar C., Hamdan H. (2013). Interval data clustering using self-organizing maps based on adaptive Mahalanobis distances. *Neural Networks*, **46**, 124–132.

Han, A., Hong Y., Lai, K.K., Wang, S. (2008). Interval time series analysis with an application to the sterling-dollar exchange rate. *Journal of Systems Science and Complexity*, **21**(4), 558–573.

Harvey D.I., Leybourne S.J., Taylor R.A.M. (2006). Modified tests for a change in persistence. *Journal of Econometrics*, **134**, 441–469.

Jin, Z., Yang, H., Yin, G. (2015). Optimal debt ratio and dividend payment strategies with reinsurance. *Insurance: Mathematics and Economics*, **64**, 351-363.

Kim J.Y. (2000). Detection of change in persistence of a linear time series. *Journal of Econo-*

metrics, **95**, 97–116.

Kim J.Y., Franch J.B., Amador R.B. (2002). Corrigendum to “Detection of change in persistence of a linear time series”. *Journal of Econometrics*, **109**, 389–392.

Lai, R.K., Fan, C.Y., Huang, W.H., Chang, P.C. (2009). Evolving and clustering fuzzy decision tree for financial time series data forecasting. *Expert Systems with Applications*, **36**(2), 3761–3773.

Lee, C.H.L., Liu, A., Chen, W.S. (2006). Pattern discovery of fuzzy time series for financial prediction. *IEEE Transactions on Knowledge and Data Engineering*, **18**(5), 613–625.

Leybourne, S., Kim, T. H., Smith, V., Newbold, P. (2003). Tests for a change in persistence against the null of difference-stationarity. *The Econometrics Journal*, **6**(2), 291-311.

Li, W., Wong, M.C., Cenev, J. (2015). High frequency analysis of macro news releases on the foreign exchange market: A survey of literature. *Big Data Research*, **2**(1), 33-48.

Maia A.L.S., de A.T. de Carvalho F. , Ludermir T.B. (2008), Forecasting models for interval-valued time series. *Neurocomputing* **71**, 3344–3352.

Mansour, N., Cherif, M.S., Abdelfattah, W. (2019). Multi-objective imprecise programming for financial portfolio selection with fuzzy returns. *Expert Systems with Applications*, **138**, 112810.

McDonald, R., Paulson, A. (2015). AIG in Hindsight. *Journal of Economic Perspectives*, **29(2)**, 81–106.

Moews, B., Herrmann, J.M., Ibikunle, G. (2019). Lagged correlation-based deep learning for directional trend change prediction in financial time series. *Expert Systems with Applications*, **120**, 197–206.

Park, H., Kim, N., Lee, J. (2014). Parametric models and non-parametric machine learning models for predicting option prices: Empirical comparison study over KOSPI 200 Index options. *Expert Systems with Applications*, **41**(11), 5227-5237.

Rea W., Reale M., Cappelli C., Brown J.A. (2010). Identification of changes in mean with regression trees: an application to market research. *Econometric Reviews*, **29**, 754–777.

Rodriguez P. M. M., Nazarii S. (2015). Modeling and forecasting interval time series with threshold models *Advances in Data Analysis and Classification*, **15**, 41–57.

Rubio, A., Bermudez, J.D., Vercher, E. (2017). Improving stock index forecasts by using a new weighted fuzzy-trend time series method. *Expert Systems with Applications*, **76**, 12–20.

Wang, X., Yu, F., Pedrycz, W., Yu, L. (2019). Clustering of interval-valued time series of unequal length based on improved dynamic time warping. *Expert Systems with Applications*, **125**, 293–304.

Xiong T., Li C., Bao Y. (2017). Interval-valued time series forecasting using a novel hybrid HoltI and MSVR model. *Economic Modeling*, **60**, 11–23.

Zhang, W. G., Xiao, W. L., He, C. X. (2009). Equity warrants pricing model under Fractional Brownian motion and an empirical study. *Expert Systems with Applications*, **36**(2), 3056-3065.

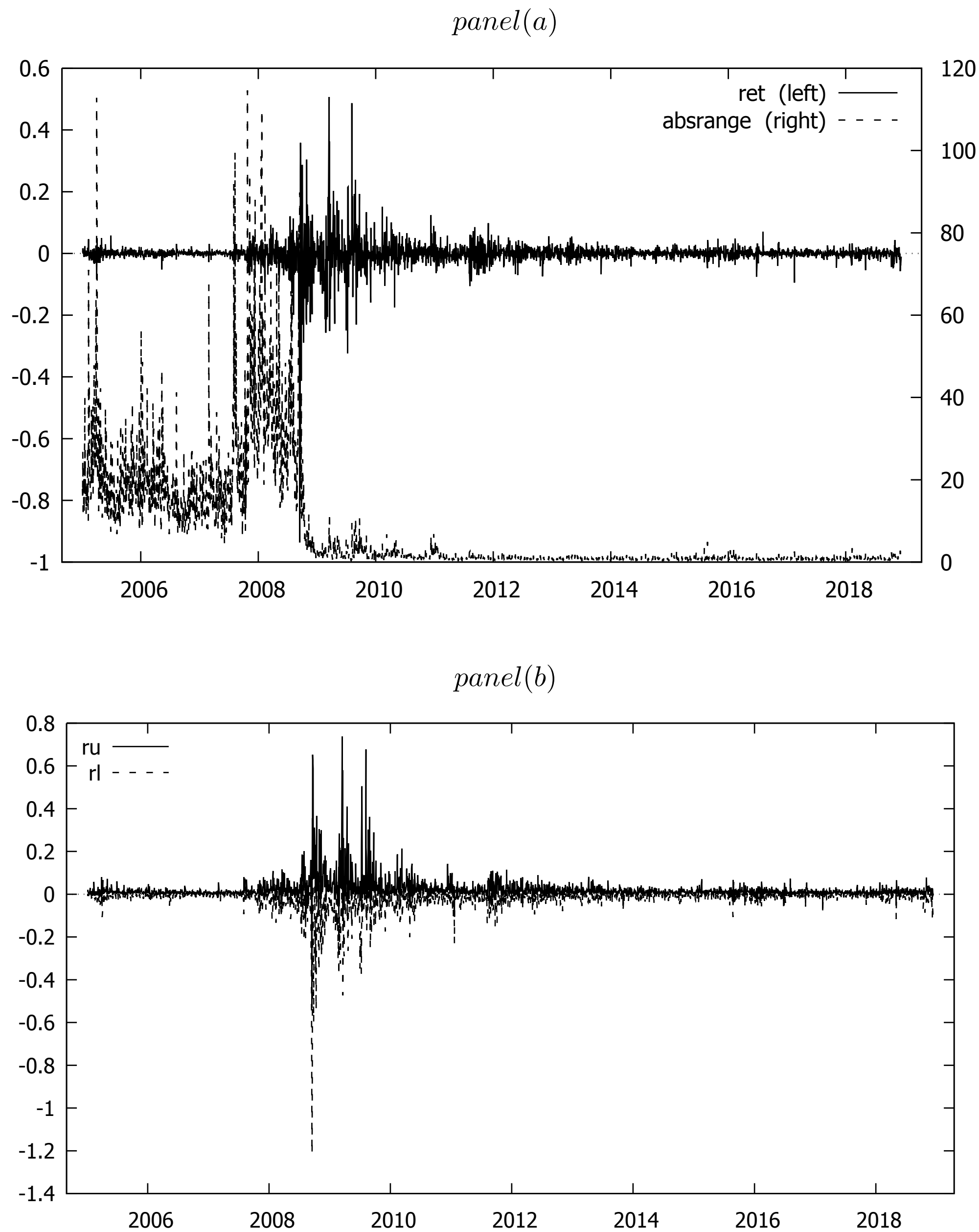


Figure 2: Panel (a): standard daily return series (log-difference of the adjusted closing price and absolute range) of the AIG. Panel (b): interval-valued daily return series, lower and upper bounds.

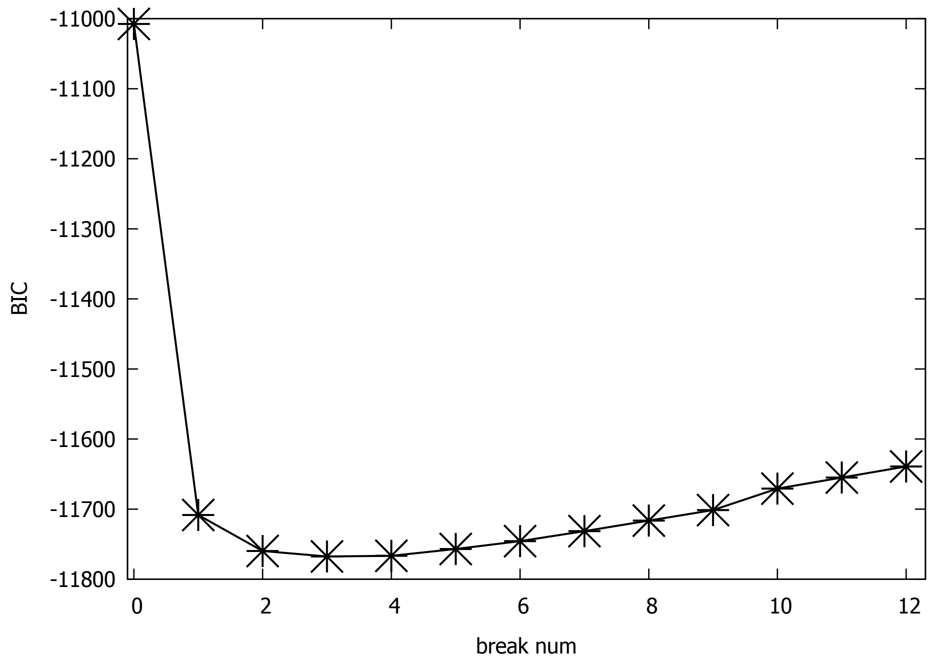


Figure 3: Values of the modified BIC against the number of terminal nodes in the pruning sequence.

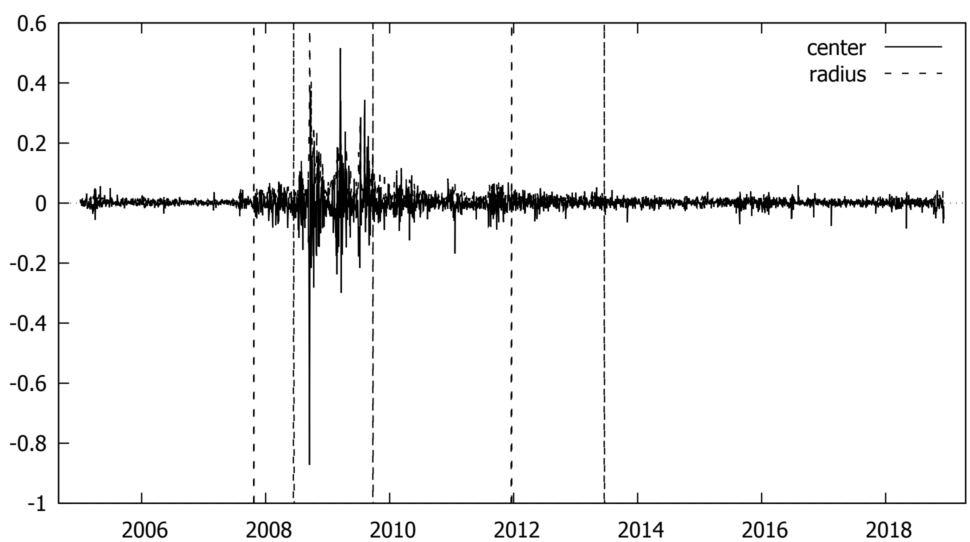


Figure 4: Centers radii of the series $\widetilde{\widetilde{Y}}_t$; the vertical dashed lines highlight the location of the detected change points. The breaks related to highest, lowest and closing prices are not reported, for the sake of maintaining a visually appealing graphical representation.

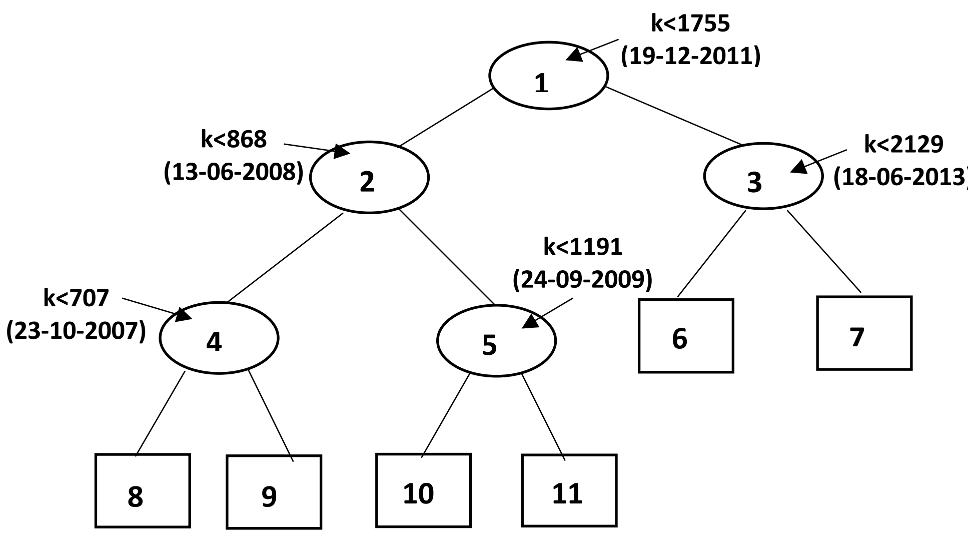
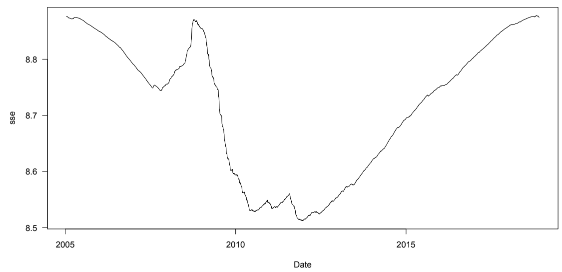


Figure 5: Change points represented as a tree diagram; the split points correspond to the break dates.

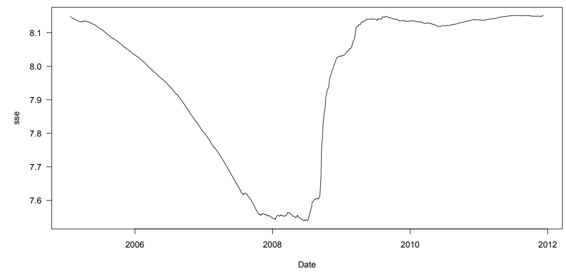
Segment: 03-Jan-2005:07-Dec-2018

change point at $19 - Dec - 2011$



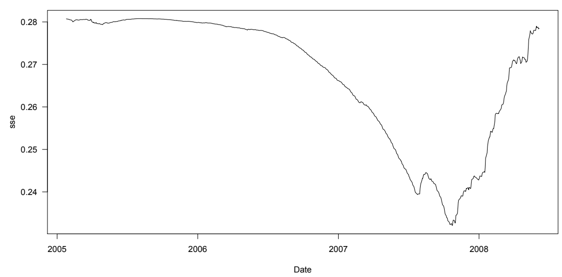
Segment: 03-Jan-2005:19-Dec-2011

change point at 13-Jun-2008



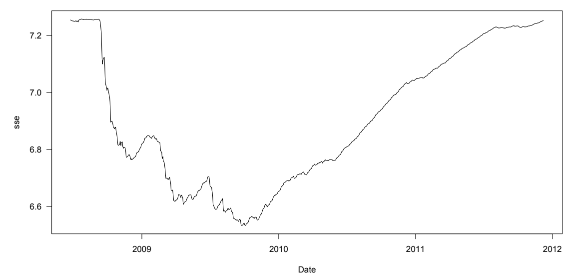
Segment: 03-Jan-2005:13-Jun-2008

change point at 23-Oct-2007



Segment: 16-Jun-2008: 19-Dec-2011

change point at 24-Sept-2009



Segment: 20-Dec-2011:07-Dec-2018

change point at 18-06-2013

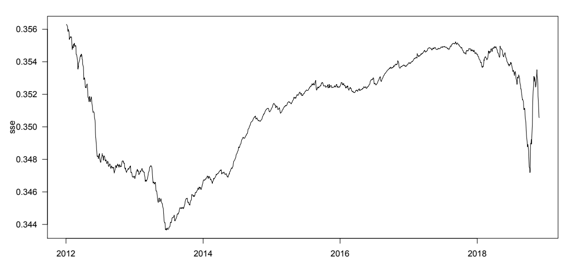


Figure 6: Behavior of the minimized objective function for the detected change points in AIG transformed interval-valued time series (centers and radii).

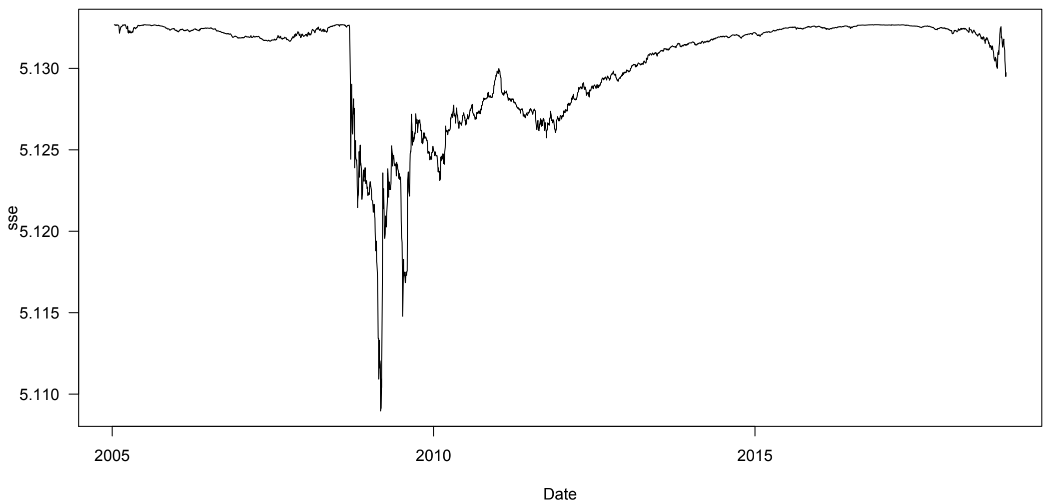


Figure 7: Behavior of the minimized objective function for the change point detected by baseline ART and Bai and Perron's procedure on the time series of the transformed centers.