Contributions to change-point analysis under long-range dependencies

Von der Wirtschaftswissenschaftlichen Fakultät der

Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des akademischen Grades

Doktor der Wirtschaftswissenschaften

– Doctor rerum politicarum –

genehmigte Dissertation

von

Diplom - Ökonom Juliane Willert

geboren am 17. Februar 1984 in Bergen auf Rügen

2012

Referent: Prof. Dr. Philipp Sibbertsen, Leibniz Universität Hannover

Koreferent: Prof. Dr. Matei Demetrescu, Universität Bonn

Tag der Promotion: 02.04.2012

Acknowledgements

This thesis was built on the support of so many. First and foremost I am deeply grateful to my mentor and co-author Prof. Dr. Philipp Sibbertsen. Despite his more than full workload he contributed so much encouragement and guidance and invested his knowledge to keep track of everything. I am also thankful to Prof. Dr. Matei Demetrescu who agreed straightforwardly to be the second examiner and to Prof. Dr. Olaf Hübler for chairing the examination board.

During the last 5 years I was fortunate enough to work with wonderful companions who spend time and load. My appreciation goes especially to Dr. Florian Heinen who pioneered for all of us in dissertation technicalities and co-authored with me, to Corinna Luedtke for constant motivation during wellness hours, to Philip Bertram for his clear view on the entirety, to Johannes Rohde for making endless hours over exams so much more delightful, to Claudia Grote for bringing spunk into the gang and to Hendrik Kaufmann for sharing not only the office but so many discussions with me. It always gave me great pleasure to work with such colleagues and superior. I will treasure so many memories, hopefully for the rest of my life.

For investing into my education and constantly looking after my well-being I am very grateful to my family, especially my parents. The thesis is the gratification for all the unexplainable problems and questions and the vague forecasts concerning the progress of my work.

This thesis would not exist without the devotion of my husband. He knows what he had to listen to.

Hannover, April 2012

Juliane Willert

Kurzzusammenfassung

Die Analyse von Strukturbrüchen umfasst die Ermittlung von Anzahl und Position von strukturellen Brüchen in Zeitreihen. Durch das wachsende Interesse im Finanzmarktbereich in den letzten 50 Jahren wurde die Forschung für das korrekte Auffinden von Brüchen immer wichtiger, um entsprechende Prozesse präzise modellieren, testen und prognostizieren zu können. Diese 4 Beiträge untersuchen verschiedene Ansätze bei dem Vorherrschen von langfristigen Abhängigkeiten in den Zeitreihen, auch bezeichnet als langes Gedächtnis.

Kapitel 2 und 3 basieren auf dem Ansatz der atheoretischen Regressionsbäume (ART). In einem ersten Schritt wird ein stark angepasster Baum aufgespannt, der die potentiellen Bruchpunkte enthält. Er entsteht durch das Anpassen von stückweise linearen Funktionen an die Zeitreihe. In dem zweiten Schritt wird die Überanpassung korrigiert mit Hilfe einer Zurückschneideprozedur, die die Äste mit dem geringsten Erklärungsbeitrag entfernt. In Kapitel 2 wird gezeigt, dass das häufig verwendete BIC (Bayesianische Informationskriterium) als Zurückschneideprozedur unter langem Gedächtnis nicht gut arbeitet aufgrund seines zu schwachen Strafterms. Eine einfache, aber effektive Methode für das Zurückschneiden wird vorgestellt, die den zu geringen Einfluss des Strafterms entsprechend ausgleicht. In Kapitel 3 (gemeinsam verfasst mit Philipp Sibbertsen) wird eine Modifikation des BIC, das LWZ (Liu, Wu und Zidek (1997)), vorgestellt, welches die gut erforschten Eigenschaften des BIC und die Besonderheiten bei langem Gedächtnis miteinander vereint. Dies wird nun mit alternativen Zurückschneideverfahren wie dem BIC und dem LIC (Lavielle und Moulines (2002)) verglichen und Konsistenz der Schätzung auf Basis der atheoretischen Regressionsbäume kann gezeigt werden. ART stellt sich als überaus schneller Ansatz zur Schätzung der Anzahl und Position von Sturkurbrüchen heraus.

Die folgenden Beiträge in Kapitel 4 und 5 befassen sich mit Problemen der Strukturbruchanalyse in Bezug auf die Testverfahren CUSUM und MOSUM. Zusätzlich zum Mittelwert einer Zeitreihe kann ebenfalls der Lange-Gedächtnis-Parameter zeitabhängig sein. In Kapitel 4 (gemeinsam verfasst mit Philipp Sibbertsen) wird ein CUSUM-Quadrat-Test basierend auf Leybourne et al. (2007) verwendet, um das Verhalten des langen Gedächtnis gegen einen Bruch in diesem zu testen. Die Testalternative umfasst den Bruch in der Persistenz sowohl vom stationären in den instationären Bereich als auch umgekehrt. Bedauerlicherweise ist diese Testprozedur nicht robust gegenüber zusätzlichen Brüchen im Mittelwert und erleidet starke Verzerrungen in der Size. Deshalb sind adjustierte kritische Werte unerlässlich, wenn bekannt ist, dass ein Mittelwertbruch im datengenerierenden Prozess vorliegt.

Ein anderer Ansatz bezüglich des Zustand abhängigen Verhaltens von Parametern wird im abschließenden Kapitel 5 (gemeinsam verfasst mit Florian Heinen) beleuchtet. Die Testidee des CUSUM Testes wird modifiziert zu einem Monitoring-Ansatz. Dieser erlaubt die schnelle Entdeckung einer Änderung in der langfristigen Abhängigkeitsstruktur, also dem Parameter des langen Gedächtnis. Der MOSUM Test kann unproblematisch erneut ausgeführt werden, so bald neue Daten vorliegen, ohne in Probleme des multiplen Testens zu geraten.

Schlagwörter: Langes Gedächtnis, Strukturbrüche, ART, Informationskriterien, CUSUM, MOSUM

Short summary

In time series analysis the change-point analysis describes the detection and localization of structural breaks. During the last 50 years the growing interest in financial markets nourished the research in finding breaks in different parameters to model, test and forecast the underlying process correctly. These four contributions investigate different approaches when it comes to long-range dependencies, named long-memory behavior.
Chapter 2 and 3 focus on approaches based on atheoretical regression trees (ART). In the first step a tree is constructed well overfitted with potential breakpoints due to the fitting of piecewise constant functions to the time series. In the second step the overestimation is adjusted through a pruning procedure that cuts back branches with the lowest contribution. In chapter 2 it is shown that the bayesian information criterion (BIC), which is commonly used as a pruning method, does not operate well in the long memory framework because of an inferior penalty term. A simple but effective procedure is presented to deal with this underweight impact of the penalty term. In chapter 3, co-authored with Philipp Sibbertsen, a modification of the BIC, the LWZ (Liu, Wu and Zidek (1997)), is presented to overcome long-range dependence issues and use the well-researched properties of the BIC at the same time. It is compared to alternative pruning criteria like the BIC or LIC (Lavielle and Moulines (2002)). Also consistency of the estimation using tree-based methods is shown. ART are highlighted as a fast approach for change-point detection that can estimate the number and location of structural breaks both in a single algorithm with minor impacts through long memory behavior.
The following essays in chapter 4 and 5 overcome problems regarding change-point analysis in the context of CUSUM and MOSUM testing. Based on the idea that not only the mean is at risk of changing over time the long memory parameter could additionally be time-dependent. In chapter 4, co-authored with Philipp Sibbertsen, the CUSUM-squared based test for a change in persistence by Leybourne et al. (2007) tests long memory behavior versus a break in persistence from stationary to non-stationary long memory and vice versa in the alternative. Unfortunately this test procedure is not robust against shifts in the mean and suffers from serious size distortions when mean shifts occur. Therefore, adjusted critical values are needed when it is known that the data generating process has a mean shift.
A different perspective on the regime changing behavior is taken in the concluding chapter 5, co-authored with Florian Heinen. The CUSUM idea is modified to a monitoring technique that allows the detection of a single change in the long-run correlation structure of a time series at some unknown future point in time. The MOSUM test can be executed once new data arrives without running into multiple testing problems. Different forms of boundary functions for the test are derived and the finite sample performance is investigated.

Keywords: long memory, structural breaks, ART, information criteria, CUSUM, MOSUM

Contents

Chapter 1

Introduction

Introduction

In the last 50 years the study of detection and location of structural breaks in time series developed effectively both in the statistical and econometric literature. The growing interest in financial markets and at the same time strong shocks like world wars and global economic and oil crises led to the awareness that conventional time series models were not sufficient any longer. Change-point analysis became an area of research that gained attention and spread out not only in finance but in medicine, chemistry, meteorology, physics, computer science and engineering.

Early contributions to the change-point analysis made by Chow (1960), who suggested a test for structural break detection at a known date, and Brown et al. (1975), who developed the theory for tests of significance for cumulative summation (CUSUM), laid the foundation for multiple break detection. By examining macroeconomic time series Nelson and Plosser (1982) began to model the mean through a stochastic model rather than a deterministic trend and later Perron (1989) modeled breaks for explaining shocks. He found out that a misspecification of these shocks would bias unit root tests. Zivot and Andrews (1992) reconsidered Perrons findings and saw disadvantages in his choice to set the breakpoints. Therefore Zivot and Andrews (1992) modeled their own breakpoint estimator. On that basis Andrews (1993) proposed a test for detection of a break at unknown break dates however with average sample properties. For further overviews see Hansen (2001) and Banerjee and Urga (2005).

More recently Bai and Perron (1998, 2003) extended their work to a multiple breakpoint estimator when the date is unknown. Their estimator, later referred as the Bai-Perron-estimator, provides a basis for change-point analyses not only for time series. It is a consistent estimator with good small sample properties and can serve as a benchmark when it comes to breakpoint detection. However, the Bai-Perron-estimator is computationally intensive and therefore not feasible for long time series (Cappelli et al. (2008) see more than 600 time points too long as a rule of thumb). Additionally the Bai-Perron-estimator depends on the pre-specification of the maximum number of breaks which increases the computational time disproportionately with larger maxima. Moreover recent studies have shown that the Bai-Perron-estimator is unsuitable for long memory time series (Rea (2008)).

Long-range dependency models have been most successful for economic time series. Large evidence for the effective modeling based on long memory processes can be found at e.g. Christensen and Nielsen (2007), Shimotsu (2006), Bhardwaj and Swanson (2006), Deo et al. (2006), Hurvich et al. (2005), Granger and Hyung (2004), Breidt et al. (1998) and Andersen and Bollerslev (1997). The persistence indicates local trends and long cycles and can handle e.g. long-term dependencies on financial markets. On the other hand this behavior makes it very challenging for breakpoint detection procedures to find the correct breaks (see Sibbertsen (2004)). The biggest challenge is to distinguish between true long memory behavior and regular breaks in the mean and inevitably the literature started discovering spurious long memory behavior. For instance Perron and Qu (2010), Granger and Hyung (2004), Gourieroux and Jasiak (2001) and Diebold and Inoue (2001) find examples where long memory can easily be confused with breaks in the mean and conclude that the distinguishing is very hard because both processes are almost observationally equivalent (Shimotsu (2006)). Nevertheless they do not consider that certain behavior can be explained through different modeling approaches and it does not indicate the true data generating process. Choi and Zivot (2007) showed that even after adjusting for breaks in the mean there is still substantial evidence for long memory. That's why it is so crucial to detect all mean shifts regardless of the persistence in order to avoid misleading conclusions. The estimation of the long memory parameter e.g. is heavily biased when there are structural changes in the mean or in the long memory parameter itself (see Granger and Hyung (1999) and Diebold and Inoue (2001)). What makes it even more appealing for current research is the fact that the well-established Bai-Perron-estimator tends to fail finding the correct number and location of breakpoints when it comes to high persistence and hence reasonable alternatives are required (Rea (2008)).

This thesis focusses on detecting structural breaks in the mean occurring at unknown dates when there is long-term persistence, named long-memory behavior. To this purpose the use of a fast non-parametric procedure based on regression trees is suggested. In the first step the tree is spanned and constructs a well overfitted tree with potential breakpoints due to the fitting of piecewise constant functions to the time series. In the second step the overestimation, especially for short series, is adjusted through a pruning procedure that cuts back branches with the lowest contribution to the deviance reduction to gain the optimal partitioning. This binary splitting in time series analysis was first justified by Hartigan (1975) and later Breiman et al. (1993) derived the large sample theory that is seen by Wu et al. (2008) as a preferred method when it comes to partitioning. When applying atheoretical regression trees (ART) to time series some open questions following Rea et al. (2010) are:

- What is the best tree selection and pruning procedure?
- Do ART find or add breaks through the fitting of piecewise constant functions?
- What are the effects of serial correlation on the performance?
- Are ART robust to any kind of noise structure or a lack of breaks?
- How do ART handle long-range dependencies?

The first two questions are fundamental for the breakpoint estimation. How to construct optimal break detection procedures and whether it provides consistent estimates for the number and location of the breaks is the key element of change-point analysis. The robustness can be checked via monte carlo studies. Obviously an increase in the length of the series leads to more robust results. The natural focus here marks the impact on the performance of ART when it comes to long memory behavior.

In chapter 2 and 3 two different approaches for the pruning procedure will be introduced. Both show good robustness properties when it comes to serial correlation and long-range dependencies and reveal superior performance to alternatives. In chapter 2 it is shown that the bayesian information criterion (BIC), which is commonly used as a pruning method, does not operate well in the long memory framework because of an inferior penalty term. A simple but effective procedure is presented to deal with this underweight impact of the penalty term. In chapter 3, co-authored with Philipp Sibbertsen, a modification of the BIC, the LWZ (Liu, Wu and Zidek (1997)), is presented to overcome long-range dependence issues and use the well-researched properties of the BIC at the same time. It is compared to alternative pruning criteria like the BIC or LIC (Lavielle and Moulines (2002)). Also consistency of the estimation using tree-based methods is shown. ART are highlighted as a fast approach for change-point detection that can estimate the number and location of structural breaks both in a single algorithm with minor impacts through long memory behavior.

The following essays in chapter 4 and 5 overcome problems regarding change-point analysis in the context of CUSUM and MOSUM testing. Based on the idea that not only the mean is at risk of changing over time the long memory parameter could additionally be time-dependent. In chapter 4, co-authored with Philipp Sibbertsen, the CUSUM-squared based test for a change in persistence by Leybourne et al. (2007) tests long memory behavior versus a break in persistence from stationary to non-stationary long memory and vice versa in the alternative. Unfortunately this test procedure is not robust against shifts in the mean and suffers from serious size distortions when mean shifts occur. Therefore, adjusted critical values are needed when it is known that the data generating process has a mean shift. These are given for the case of one mean break. Response curves for the critical values are derived and a monte carlo study showing the size and power properties under general de-trending is given.

A different perspective on the regime changing behavior is taken in the concluding chapter 5, co-authored with Florian Heinen. The CUSUM idea is modified to a monitoring technique that allows the detection of a single change in the long-run correlation structure of a time series at some unknown future point in time. The MOSUM test can be executed once new data arrives without running into multiple testing problems. We focus on the detection of an increasing persistence with a process that is becoming non-stationary under the alternative. Different forms of boundary functions for the test are derived and the finite sample performance is investigated. The concluding application shows that loss of controllability indicated through an increasing persistence is indeed a highly probable outcome for economic time series.

Chapter 2

Mean Shift detection under long-range dependencies with ART

Mean Shift detection under long-range dependencies with ART

Published as Leibniz University of Hannover Discussion Paper No. 437.

Revised and resubmitted in Statistical Methods and Applications.

2.1 Introduction

It is an ongoing problem to detect changes in the mean. In the long-memory framework it gets even more difficult to specify number and location correctly because of the high persistence in the time series. The long cycles and local trends challenge every breakpoint estimator and make it hard to distinguish between long memory and mean shifts (see e.g. Sibbertsen (2004)). In addition undetected shifts in the mean strongly bias estimators e.g. for the memory parameter and create therefore misleading results.

Granger and Hyung (1999) as well as Diebold and Inoue (2001) showed that long memory behavior can easily be confused with mean shifts and that their properties are very similar. That's why standard break detection procedures can struggle and are at risk to fail.

There are several methods to specify the presence of structural breaks. Chow (1960) was the first creating a test on structural changes based on the F statistic when the breakpoint was known. There are Brown, Durbin and Evans (1975) who suggested the CUSUM approach and Ploberger and Krämer (1992) who based a structural change test on the cumulative sums of recursive residuals. Bai and Perron (1998) modeled their own break date estimator and allowed to have multiple breaks in the mean. Their method was a breakpoint estimator based on OLS regression which works reasonable for short memory time series. Hence it became the standard procedure for breakpoint estimation.

The methodology of classification and regression trees of Breiman et al. (1993) was applied to time series analysis by Cappelli et al. (2008) and Rea et al. (2010). They showed that atheoretical regression trees (ART) have reasonable performance in detecting and locating structural breaks in short-memory time series and perform impressively in comparison with Bai and Perron (1998). However in the long-memory framework the Bai Perron procedure does not work properly (see Rea (2008)), so least squares regression trees could be a reasonable alternative.

Regression trees operate in two steps. First the growing step spans a tree which is often overfitted (see Rea et al. (2010)) and therefore the second step, the pruning of the tree, is the much more important part. Given that the common pruning techniques fail in the long memory framework a new pruning method called elbow criterion will be modeled to overcome this problem. It still maintains the good properties of the regression trees to specify the number of mean shifts and detect their location. Additionally it overcomes the problem of overestimation due to long memory behavior by penalizing accordingly.

The rest of the paper is organized as follows. In section 2.2 the method of atheoretical regression trees is introduced and different pruning techniques are discussed. The common pruning methods will be replaced by the elbow criterion. Section 2.3 contains an extensive Monte Carlo study to analyze the performance of the elbow criterion and its advantage in comparison to other pruning techniques. In section 2.4 an application to CPI inflation rates is given. Section 2.5 concludes.

2.2 Atheoretical regression trees

ART is a nonparametric procedure that is used to detect and locate structural breaks. It does not require distributional assumptions about the data or the residuals and hence it is well suited for a variety of time series. A simple breakpoint model reads

$$\begin{aligned} y_t &= \mu_p + \epsilon_t \\ \mu_p &= \sum_{i=1}^{p} I_{(t_{i-1}<t<t_i)}\mu_i \end{aligned}$$

where y_t is the value of the time series at time t, ϵ_t is the error term which is assumed to be stationary and μ_p is the mean of the time series up to the breakpoint p. $I_{t\in R}$ is an indicator function which is 1 if t is in the regime i and 0 otherwise. t_i with $i = 1,...,p$ are the breakpoints with the mean of the regime μ_i.

A regression tree fits piecewise constant functions to the data and determines thereby potential breakpoints. The construction of the tree uses a greedy algorithm. That means that at each step the best split is determined and there is no reconsideration of the already set splits. The only exogenous predictor variable for the OLS regression is the time t. Though it operates more like a counter rather than a true predictor.

To determine the best split a measure of node impurity is needed. The sum of squared residuals (RSS) is used to determine where the node will be set. The mean squared error is given as a risk function by

$$R(t) = \frac{1}{n(t)} \sum_{x_i \in t} (y_i - \bar{y}(t))^2$$

where

$$\bar{y}(t) = \frac{1}{n(t)} \sum_{x_i \in t} y_i.$$

x_i are the predictor variables (time points) which belong to one regime and $n(t)$ is the number of elements in node t. The tree construction splits a node t into a left t_L and a right t_R child node for which the sum of the RSS of the left and right node is minimized.

$$\min_t (R(t_L) + R(t_R)) = \min_t \left(\frac{1}{n(t_L)} \sum_{x_i \in t_L} (y_i - \bar{y}(t_L))^2 + \frac{1}{n(t_R)} \sum_{x_i \in t_R} (y_i - \bar{y}(t_R))^2 \right)$$

This can also be written as a maximization of the improvement through the splitting into t_L and t_R.

$$\max_t (R(t) - R(t_L) - R(t_R))$$

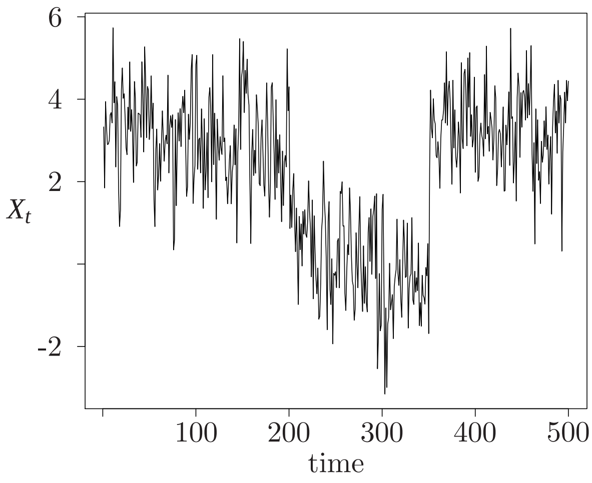
ART requires at any node $O(n(t))$ steps to identify the best split (see Rea (2008)). The recursive partitioning produces a hierarchical structure of nodes and leaves (terminal nodes). Every terminal node represents a regime with a shifted mean. The tree growth until no improvement can be made by splitting the time series. Thus the location and number of breaks in the data are determined (see also Zheng et al. (2008)).

An example will be introduced. Considering an ARFIMA(0,d,0) process

$$(1 - L)^d X_t = \epsilon_t,$$

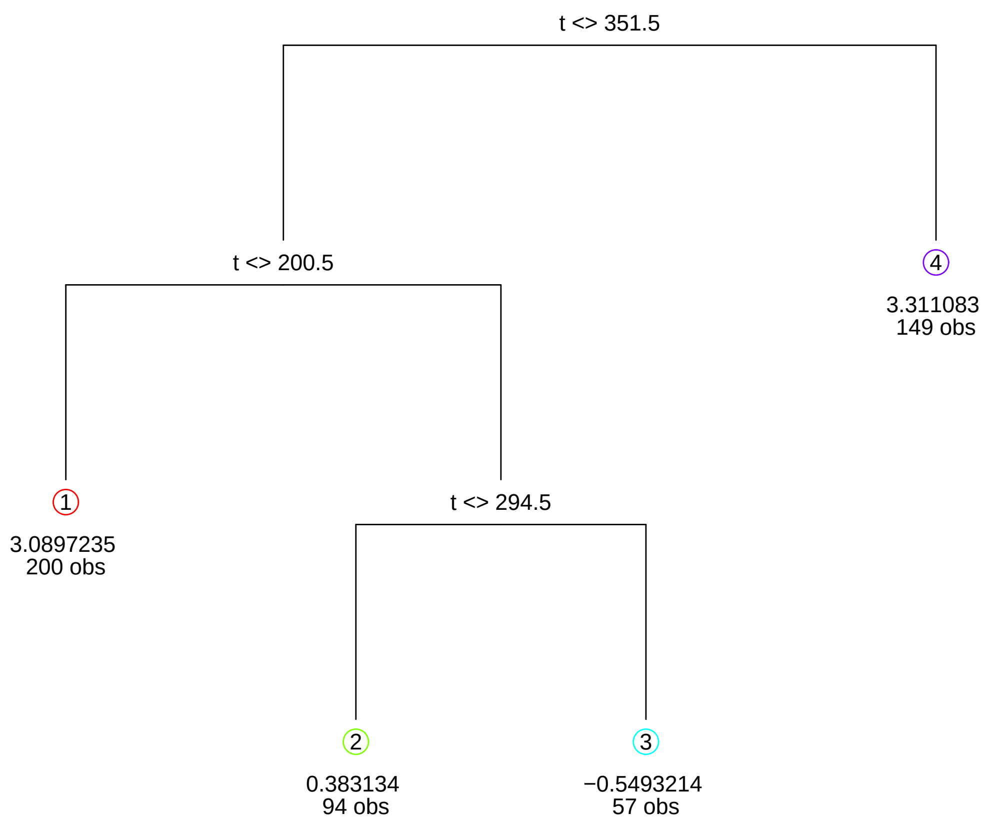
where L is the lag operator, ϵ_t are iid random variables with zero mean and the variance σ^2 and the degree of integration is determinded by the long memory parameter d. A stationary long memory process is characterized by the value of d in the interval between $[0, 0.5]$.
For $d = 0.2$, a sample size of $T = 500$ and two breaks from $\mu_1 = 3$ to $\mu_2 = 0$ and $\mu_3 = 3$ at $t_1 = 200$ and $t_2 = 350$ an exemplary time series is shown in figure 2.1.

Figure 2.1: Exemplary time series with two breaks in the mean



In figure 2.2 the spanned regression tree is presented. There are four leaves and each is representing a regime with a different mean. The nodes represent the breakpoints which are detected at $t_1 = 200$, $t_2 = 294$ and $t_3 = 351$. The different estimated mean levels are noted below the encircled numbers.

Figure 2.2: Regression tree after growing



The growing of the tree is literally driven by the data. After the growing process a very well fitted tree is build, because the only stopping rule is a lack of improvement in the sum of RSS. In fact the tree often gets quite large and is overfitted (see Rea et al. (2010)). That's why pruning techniques are needed to determine which of the nodes are redundant. There is the possibility of manual pruning which is a quite reasonable way if a priori knowledge can be used.

A nested hierarchy of regimes has been built and can be pruned back by a pruning method. Pruning works from bottom to top. That means that the first node to cut would be the one which was grown last, so which gained the weakest node impurity improvement. In our example this would be the node at $t = 294$. In figure 2.2 it is easy to see that this branch was built last.

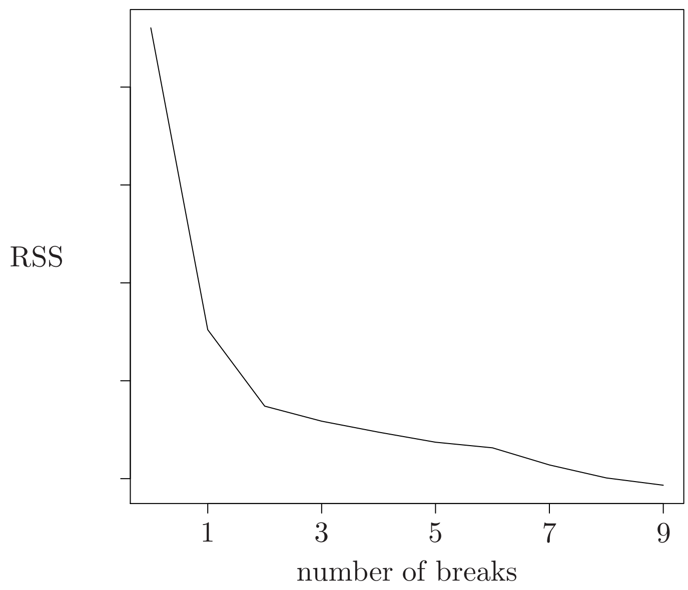
Widely used pruning methods are e.g. the cost-complexity pruning (see Breiman et al. (1993)), predictive cross validation, the one-standard-error-rule, Mallows Cp or an information criteria such as the BIC.

Rea (2008) showed that the cost-complexity pruning is difficult to handle because a complexity parameter (penalty parameter) has to be chosen. The same dilemma appears with Mallows Cp. A complexity parameter needs to be determined which directly controls the penalty parameter and hence the number of chosen breaks. The one-standard-error-rule and predictive cross validation are dealing also with handling problems and are too vulnerable to the high persistence behavior.

In the short-memory framework the BIC is the best information criterion (see Bai and Perron (1998) and Bai Perron (2003)). The penalty term of the BIC depends on the size of the time series T and the number of terminal nodes p. Kokoszka and Leipus (2002) show that the Bai Perron procedure which is similar to the BIC information criterion excludes linear sequences with long-range dependence. Regarding to that it is not astonishing that the BIC does not handle long memory reliably, which can also be seen in section 2.3.

A new pruning method will be suggested to overcome this problem. The idea of the *elbow criterion* is that the optimal break number is reached when the improvement of the sum of RSS is highest. A typical shape of the sum of the squared residuals shows that there is always a better fit by including more breaks due to its convex characteristics (see figure 2.3). The key point is that some splits downsize the risk function even more than others.

Figure 2.3: Typical shape of the sum of squared residuals depending on the break number



The largest improvement in the RSS is made where the trend has the biggest bend. To determine this bend the slopes of the piecewise constant functions are considered. The RSS function is extended to the right with a slope of zero, because the tree stopped splitting at that point. Therefore it be assumed that no improvement of the RSS could be achieved anymore. Calculating the difference between two adjacent slopes provides a measure for the improvement benefit through this splitting. The highest benefit is defined as the optimal number of breaks.

This procedure is independent of the length of the time series and the number of terminal nodes. It determines the optimal number of breaks where the highest improvement can be made through splitting at that point. The advantage is that the overfitted tree which was grown can be counterbalanced because all the small RSS improvements become irrelevant. In comparison the BIC does depend on the size of its penalty term and though it depends on the amount of suggested breakpoints (see Bai and Perron (2003)).

The elbow criterion considers an absolute deviation between the levels of the RSS function and by this it can easily respond to different levels of the RSS function through different time series and persistences respectively. Returning to the example given before the optimal number of breaks would be 2. In figure 2.3 you can see that at two breaks the improvement through splitting the sample is highest which expresses in the biggest bend (and smallest angle) of the RSS function.

2.3 Monte Carlo study

An extensive Monte Carlo study will demonstrate the performance of the new pruning method for the long-memory framework in comparison to the BIC. All simulations are computed with the open-source programming language R (2008) with package support (Zeileis et al. (2002) and Ripley (2005)). The number of replications is set to $M = 1000$ and we consider a sample size of $T = 500$ in order to illustrate the good performance in small samples. All results improve when using larger samples.

The data generating process is an ARFIMA (0,d,0) with $d = 0.2$ and $d = 0.4$ respectively. The levels of the mean are chosen relatively small on purpose. The considered changes correspond to the standard deviation of the noise distribution ($s_{\epsilon_t} = 1$) and half of s_{ϵ_t} respectively (leaned on Bai and Perron (2006)). Small changes e.g. from $\mu_1 = 1$ to $\mu_2 = 2$ are harder to determine than large level shifts. Also returning breaks (e.g. $\mu_1 = 1$ to $\mu_2 = 2$ and back to $\mu_3 = 1$) are challenging, because the small peak in between can be easily overlooked. In case of one mean shift the break location is set to the 300th observation. Besides it will be shown that the position does not have a big influence on the results. Considering more mean shifts the break locations will be spaced equally.

Be aware that the case of no mean shifts is not encountered by this procedure. The regression trees are build to split data. Finding no break at all is only given by the rather unlikely case that no improvement in the RSS over the whole time series during the first step can be found. No splitting is not an option during the growing procedure or in other words: the tree has always at least two branches.

Comparing the widespread BIC and the elbow criterion underpin the findings of Kokoszka and Leipus (2002) and Banerjee and Urga (2005). The BIC is not able to handle the long-range dependencies because of the high persistence and dependencies. The tree misspecifies local trends and cycles as additional breakpoints and the penalty term of the BIC is not strong enough to penalize the high persistence. The BIC leads to choose the maximum number of breakpoints which is spanned by the regression tree, so in most cases no real pruning takes place.

Table 2.1: Performance of BIC and elbow criterion
when there is one mean shift

	elbow criterion			BIC		
$d = 0.2$	mean	s.d.	% correct	mean	s.d.	% correct
$\mu_1 = 1; \mu_2 = 2$	1.03	0.35	98.60	3.82	1.67	7.40
$\mu_1 = 2; \mu_2 = 1$	1.04	0.34	98.30	3.78	1.68	7.80
$\mu_1 = 1; \mu_2 = 1.5$	1.66	1.37	72.40	4.01	1.89	8.20
$\mu_1 = 1.5; \mu_2 = 1$	1.64	1.30	70.60	3.86	1.78	8.40
$d = 0.4$						
$\mu_1 = 1; \mu_2 = 2$	1.53	1.28	78.10	6.44	1.86	0.40
$\mu_1 = 1; \mu_2 = 1.5$	1.91	1.53	59.00	6.63	1.89	0.10

Table 2.1 displays the huge problems of the BIC to find only one mean shift. It overestimates the quantity by multiple times. The higher the persistence the more mean shifts will be detected and the lower is the quantity of a correct determination. For the elbow criterion it is easier to determine this one mean shift in a stationary long memory process. The higher the level of the mean shift and the lower the persistence the more accurate is the criterion. Hence the mean is very close to the correct number of breaks, a very small standard deviation is obtained and the percentage of a correct chosen number of breaks is high.

The direction of the shift (from a high level to a lower one or vice versa) influences neither the pruning criterion nor the tree growing process. The following table 2.2 shows that the position of the mean shift barely influences the performance of the pruning method.

Table 2.2: Performance of BIC and elbow criterion
when the position of the break varies and there is one mean shift

$d = 0.2$; $\mu_1 = 1; \mu_2 = 2$	elbow criterion			BIC		
break at observation	mean	s.d.	% correct	mean	s.d.	% correct
50	1.39	1.02	81.10	4.00	1.88	9.90
250	1.03	0.31	98.90	3.77	1.63	8.00
450	1.39	0.98	80.70	4.07	1.85	8.70
$d = 0.4$; $\mu_1 = 1; \mu_2 = 2$						
50	2.02	1.67	57.00	6.63	1.79	0.30
250	1.44	1.10	80.00	6.31	1.83	0.30
450	1.87	1.46	60.80	6.70	1.73	0.00

The results for multiple mean shifts are reported in table 2.3 and 2.4. The elbow criterion handles more breaks solid and gives good results in detecting the mean shifts. The positions of the breakpoints are spaced equally.

Table 2.3: Performance of BIC and elbow criterion when there are two mean shifts

	elbow criterion			BIC		
$d = 0.2$	mean	s.d.	% correct	mean	s.d.	% correct
$\mu_1 = 1; \mu_2 = 3; \mu_3 = 1$	2.15	0.39	87.20	3.36	1.13	23.90
$\mu_1 = 1; \mu_2 = 2; \mu_3 = 1$	2.04	0.64	67.00	4.52	1.48	7.80
$\mu_1 = 1.5; \mu_2 = 2; \mu_3 = 1$	1.51	0.85	33.40	4.31	1.61	9.50
$d = 0.4$						
$\mu_1 = 1; \mu_2 = 3; \mu_3 = 1$	1.92	0.75	55.10	5.85	1.58	1.20
$\mu_1 = 1; \mu_2 = 2; \mu_3 = 1$	1.87	1.21	31.20	6.71	1.65	0.00
$\mu_1 = 1.5; \mu_2 = 2; \mu_3 = 1$	1.88	1.39	22.90	6.59	1.79	0.70

Table 2.4: Performance of BIC and elbow criterion for multiple mean shifts

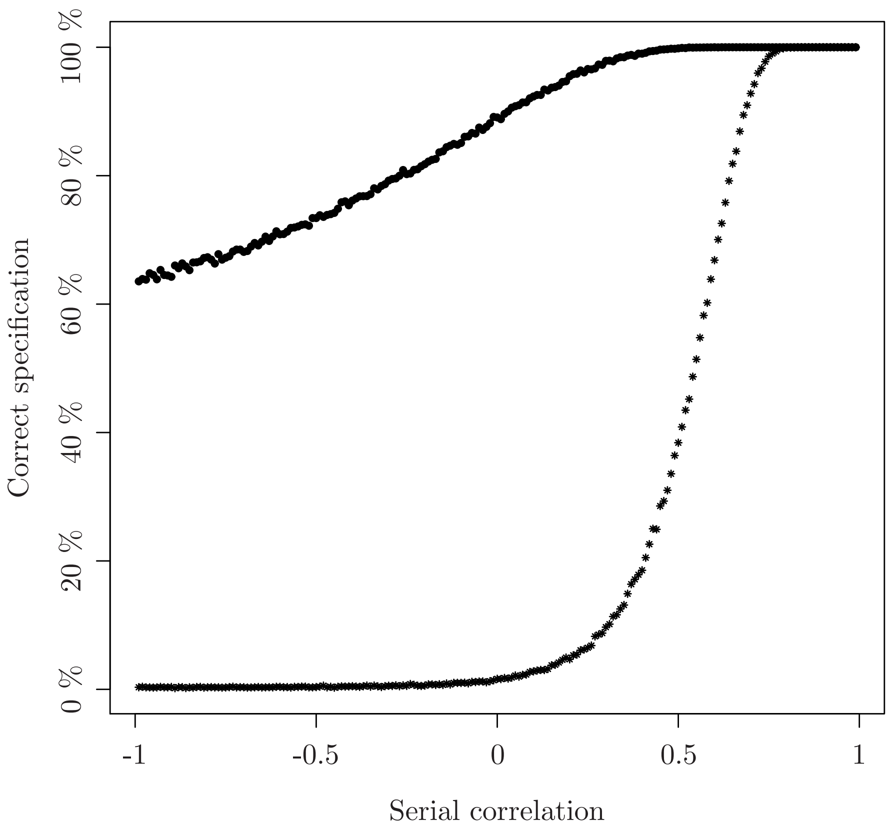
	elbow criterion			BIC		
$d = 0.2$	mean	s.d.	% correct	mean	s.d.	% correct
$\mu_1 = 1; \mu_2 = 2; \mu_3 = 1; \mu_4 = 2$	2.59	1.14	39.90	5.12	1.27	8.30
$\mu_1 = 1; \mu_2 = 1.5; \mu_3 = 2.5; \mu_4 = 1$	2.18	0.73	31.40	4.69	1.35	16.00
$\mu_1 = 1; \mu_2 = 2; \mu_3 = 1; \mu_4 = 2; \mu_5 = 1$	3.17	1.66	19.40	5.94	1.26	9.40
$\mu_1 = 1; \mu_2 = 3; \mu_3 = 1; \mu_4 = 3; \mu_5 = 1$	4.05	1.13	53.00	5.20	1.00	26.50
$d = 0.4$						
$\mu_1 = 1; \mu_2 = 2; \mu_3 = 1; \mu_4 = 2$	2.08	1.35	19.30	6.81	1.63	1.50
$\mu_1 = 1; \mu_2 = 1.5; \mu_3 = 2.5; \mu_4 = 1$	1.83	1.05	16.90	6.43	1.70	0.30
$\mu_1 = 1; \mu_2 = 2; \mu_3 = 1; \mu_4 = 2; \mu_5 = 1$	2.34	1.62	8.30	7.03	1.61	3.30
$\mu_1 = 1; \mu_2 = 3; \mu_3 = 1; \mu_4 = 3; \mu_5 = 1$	2.77	1.61	16.20	6.82	1.36	2.90

In the case of multiple mean shifts the elbow criterion tends to underestimate the number of mean shifts which implies that some of the small shifts can not be identified. Nevertheless the chosen transitions are quite regular which is much more difficult to detect for a breakpoint estimator than extreme breaks. This almost cyclic behavior (from $\mu_1 = 1$ to $\mu_2 = 2$ and back to $\mu_3 = 1$ and $\mu_4 = 2$) simulates the most challenging break pattern with local cycles and persistences best. Hence the good behavior in this cases are very founded results for more obvious (easier to be detected) breaks.

Studying various noise distributions including normal, t- and double exponential distribution shows that the noise distribution has no effect at the elbow criterion and the BIC at all. Bai and Perron (2006) also note in a large monte carlo study that the BIC is not affected by the distribution of the noise parameter. The impact of serial correlation is more serious and also extensively illustrated on the BIC in Bai and Perron (2006). They suggest using the LWZ criterion (by Liu et al. (1997)) to impose a higher penalty term within the information criterion.

The elbow criterion reacts to serial correlation surprisingly well. Figure 2.4 illustrates that with positive correlation the criteria choosing more often the correct quantity of breaks and with negative correlation less often. The elbow criterion (solid dots) performs with at least 75% of correct specifications, where as the BIC (stars) goes down to zero very quickly.

Figure 2.4: Correct break quantity specification for $d = 0.3$ and the break pattern $\mu_1 = 1 \rightarrow \mu_2 = 2$ when there is serial correlation in the DGP



The diversing performance can be explained through the mixture of long-memory behavior and serial correlation. High positive correlation overlays the long memory correlation behavior due to "small" sample problems and therefore both criteria improve because of assumed short memory. High negative correlation on the other hand intensifies the correlation of the long-range dependencies and challenges the criteria even more.

Finally you can say that the BIC overestimates the number of breaks with high standard deviations (see also Bai and Perron (2006)). The percentage of correctly chosen breaks is often so small that even educated guessing would be more successful. The ability of the elbow criterion on the other hand stays reasonable even if there is more than one mean shift. When the persistence increases the criterion tends to underestimate the number of mean shifts. The elbow criterion as a pruning technique of the atheoretical regression trees shows very good properties even when multiple mean shifts with small level changes occur in a long memory time series. There is still a correct detection and specification with high probability. Its good properties still hold when applying different noise distributions and serial correlation in the error term.

2.4 Application on inflation rates

To illustrate the good performance of the atheoretical regression trees an application to CPI inflation rates is given. The time series data starts in January of 1960 (except Australia starts in 1971) and ends in June 2009. The following table 2.5 shows the results of some OECD countries when ART with the elbow criterion is applied.

Table 2.5: Breakpoints in inflation rates of selected OECD countries

Country	1st break	2nd break
Australia	Jan 91	-
Canada	Aug 72	Dec 91
Germany	Sep 70	May 83
Japan	Dec 81	-
New Zealand	Sep 70	Jun 90
Switzerland	Oct 93	-
UK	Sep 73	Nov 82
US	Jul 73	Nov 82

The atheoretical regression trees find one or two breaks in the inflation rates. Corvoisier and Mojon (2005) determined three waves where breaks in inflation rates occur. In their opinion since 1960 most OECD countries had breaks around 1970, 1982 and 1991. This can be very well encountered by the estimated breakpoints via ART. Hsu (2005) identifies the breakpoints under the assumption of two known breaks and finds for Germany the breaks at October 1969 and July 1982 and for the US at January 1973 and September 1981. Under the assumption of one appearing break he determines for the japanese inflation rate the breakpoint at May 1981. Hence most of his results are very close to the specified breaks by the elbow criterion, however Hsu has to know a priori how many breaks will occur.

After demeaning the inflation rates using the specified breakpoints the long memory parameter can be computed by the GPH estimator. In the following table 2.6 the mean of each regime and the d parameter after demeaning is displayed.

Table 2.6: Mean of each break regime and demeaned d estimation of selected OECD countries

Country	mean			d estimation
	start to 1st break	1st to 2nd break	2nd (1st) break to end	
Australia	9.2991	-	2.6299	0.68
Canada	2.7330	7.2467	1.8732	0.75
Germany	2.6175	5.1386	2.0153	0.50
Japan	7.0455	-	0.8459	0.58
New Zealand	3.3628	11.8101	2.2907	0.40
Switzerland	3.9000	-	0.9489	0.71
UK	4.7109	14.7415	3.7510	0.26
US	2.9175	9.0408	3.0724	0.54

The level differences of the detected breaks are quite high. When there are two breaks in the inflation rate the mean before the first break and after the second break is often almost the same and a large peak in between the breaks can be detected. In this situation (when the transitions are quite regular) ART shows good properties (see section 2.3) and hence underpin that these breakpoint findings are reliable. After demeaning the data accordingly to the estimated breakpoints long-range dependencies are still present in the data. This implies that an approach which accounts for both, long memory and mean shifts, is very rational.

2.5 Conclusion

In this paper a new pruning technique for atheoretical regression trees is introduced. When the data generating process is long memory and has shifts in the mean function it performs very reasonable and much better than common pruning methods like the BIC. In a stationary long memory framework the elbow criterion accomplishes the detection of the breaks no matter how many shifts appear and where they are situated, even in small samples. With increasing persistence and decreasing shift level the determination gets slightly underestimated. As the procedure is well grounded it can also be extended for smooth transition trees (da Rosa et al. (2008)) and to trend or volatility shifts.

Chapter 3

Estimating the number of mean shifts under long memory

Estimating the number of mean shifts under long memory

Co-authored with Philipp Sibbertsen.

3.1 Introduction

The detection of changes in the mean is a fundamental issue for many areas of time series analysis. To specify the number and location of a mean shift can be even more challenging when the underlying framework consists of long memory behavior (see Sibbertsen (2004)). The high persistence in the time series with local trends and long cycles makes it hard for every breakpoint estimator. Therefore, the biggest challenge is distinguishing between the true long memory behavior and regular mean shifts. Undetected mean shifts can lead to misleading conclusions e.g. by biased estimation of the long memory parameter (see Granger and Hyung (1999) and Diebold and Inoue (2001) for further details).

Bai and Perron (1998) developed a method to specify number and location of mean shifts which is performing well in a short memory framework. Rea (2008) investigated that the Bai and Perron procedure does not work properly when it comes to long memory data. It tends to fail when high persistent behavior becomes too severe. To overcome this problem we adopt the fast approach of Breiman et al. (1984) via atheoretical regression trees (ART).

Regression trees split a time series into a left and right partition and continue by splitting the subpartitions recursively. The split choice is based on the location where the highest reduction in the residual sum of squares can be made. In this first phase the tree is spanned and builds a well overfitted tree of potential partitions and breakpoints (see Rea et al. (2010)). In the second phase the pruning technique tries to cut back branches with low contribution to the deviance reduction to locate the optimal partition of the time series.

The application of ART to time series analysis by Cappelli and Reale (2005) shows the enormous utility regarding breakpoint analysis and opens a new perspective when it comes to structural break estimators. They showed that regression trees have reasonable performance in detecting and locating structural breaks. In comparison with Bai and Perron (1998) the least squares regression trees perform convincingly even in short-memory time series.

To locate the redundant mean shifts during the pruning phase of ART information criteria are used. Common pruning techniques such as the BIC fail when it comes to long memory behavior. Lavielle and Moulines (2002) suggested the LIC for the long memory case, which takes the long memory parameter into account. However, this requires a pre-specification that the underlying process is indeed long memory and an estimation of the long memory behavior when there are potential mean shifts coexistent. Thus a new information criteria, also Schwarz information criteria based, will be used to overcome this problem and still maintain the good properties of the regression trees to specify the number of mean shifts. The LWZ information criterion, first suggested by Liu, Wu and Zidek (1997), retains consistency but is constructed in a more flexible way with two parameters that are determined throughout the data generating process. It will be shown that it performs also in the long memory framework with superior results in comparison to the alternative pruning criteria.

The remainder of this paper is organized as follows. Section 3.2 outlines the tree-based procedure and their characteristics. Section 3.3 describes the new LWZ based pruning procedure and section 3.4 presents the results of the simulation study. It compares the LWZ with the procedure of Bai and Perron (1998, 2003) and the LIC (Laville and Moulines (2002)). Section 3.5 provides the conclusion.

3.2 Atheoretical regression trees

Atheoretical regression trees are used to detect and locate structural breaks. Using a nonparametric approach no distributional assumptions are required and a good fit to any kind of time series can be expected. Our breakpoint model is defined by

$$\begin{aligned} y_t &= \mu + \epsilon_t \\ \mu &= (\mu_1, \dots, \mu_m) \\ \mu_k &= I_{(T_k+1<\dots<T_{k+1})}\delta_k \text{ with } \delta_k \in \mathbb{R} \end{aligned}$$

where y_t is the value of the time series at time t, ϵ_t is the error term which is assumed to be stationary and μ_k is the mean of the time series in regime k up to the breakpoint m. The indicator function is 1 if you are in the regime k and 0 otherwise. $k = 1, ..., m$ are the breakpoints with the mean of the regime μ_k.

The regression tree determines breakpoints through fitting piecewise constant functions in an OLS regression framework. The exogenous predictor variable is the time t which works more like a counter than a predictor. At each regression step the best split of the time series is determined and an estimated breakpoint is not reconsidered but set fix in the further analysis.

The determination of the best split is identified with a node impurity measurement. Usually the sum of squared residuals (RSS) is used as the risk function. The mean squared error is given by

$$R(t) = \frac{1}{n(t)} \sum_{x_i \in t} (y_i - \bar{y}(t))^2$$

with

$$\bar{y}(t) = \frac{1}{n(t)} \sum_{x_i \in t} y_i.$$

The predictor variable x_i represents the time points which belong to one regime and $n(t)$ is the number of elements in node t. A node symbolizes a part of the time series with length $n(t)$ i.e. the root node reflects the whole time series. To construct the tree a node t is split into a left child node t_L and a right child node t_R where the sum of the RSS of the left side and the right side of the node is minimized. That means we start by cutting the time series into two parts where the minimization of the RSS is highest. The minimization problem describes as follows.

$$\min_t (R(t_L) + R(t_R)) = \min_t \left(\frac{1}{n(t_L)} \sum_{x_i \in t_L} (y_i - \bar{y}(t_L))^2 + \frac{1}{n(t_R)} \sum_{x_i \in t_R} (y_i - \bar{y}(t_R))^2 \right)$$

The total sum of squares can be rewritten as a minimization of the within child nodes sum of squares. This can also be written as a maximization problem regarding the improvement through the splitting into t_L and t_R which maximally distinguishes the time series in the left and right nodes by generating the highest drop in deviance (see Rea et al. (2010)).

$$\max_t (R(t) - R(t_L) - R(t_R)) =$$
$$\max_t \left(\frac{1}{n(t)} \sum_{x_i \in t} (y_i - \bar{y}(t))^2 - \frac{1}{n(t_L)} \sum_{x_i \in t_L} (y_i - \bar{y}(t_L))^2 - \frac{1}{n(t_R)} \sum_{x_i \in t_R} (y_i - \bar{y}(t_R))^2 \right)$$

Each splitting process is a binary decision whether a node is found or not. This is applied separately to each subgroup recursively until no improvement of the criterion can be achieved. Thereby a hierarchical structure is build through the recursive partitionment of the time series into nodes and terminal nodes (leaves), where every terminal node represents a final regime with a shifted mean.

The growing process of the tree continues until no further improvement by splitting the time series can be made. In practice this would lead to as many terminal nodes as observations and therefore a minimum number of observations in each child node or a minimum within-node deviance is set. Denote in what follows the estimated breakpoints by $\hat{\kappa} = (\hat{\kappa}_1, \ldots, \hat{\kappa}_m) = (\hat{T}_1/T, \ldots, \hat{T}_m/T)$ with true values $\kappa^0 = (\kappa_1^0, \ldots, \kappa_m^0)$. Under assumption 3.1 that b_T with $T \geq 1$ are nonnegative constants with probability one, we show adopting arguments similar to those in Bai and Perron (1998).

Assumption 3.1:

$$P_T(t) \geq b_T \frac{\log T}{T} \text{ for } T \geq 1 \text{ and } t \in \hat{T}_T \tag{3.1}$$

$P_T(t)$ denotes the empirical distribution of a random sample.

Lemma 3.1: Let ϵ_t be $I(d)$ with $d \in [0, 1/2)$. Then under assumption 3.1, $\hat{\kappa} \to \kappa^0$.

Proof: Denote by $\hat{\epsilon}_t$ the estimated residuals

$$\hat{\epsilon}_t = y_t - \hat{\mu}_k \qquad \text{for} \qquad t \in [\hat{T}_{k-1}+1, \hat{T}_k]. \tag{3.2}$$

Here, $\hat{\mu}_k = \overline{y}(t) = \frac{1}{n(t)} \sum_{t \in [\hat{T}_{k-1}+1, \hat{T}_k]} y_i$ and $n(t)$ gives the number of time points t in $[\hat{T}_{k-1}+1, \hat{T}_k]$. Thus in our model the mean is piecewise estimated with the arithmetic mean of the respective observations. It holds

$$\frac{1}{T}\sum_{t=1}^{T} \hat{\epsilon}_t \leq \frac{1}{T}\sum_{t=1}^{T} \epsilon_t. \tag{3.3}$$

Furthermore, we have with $d_t = \hat{\mu} - \mu^0$ for $t \in [\hat{T}_{k-1}+1, \hat{T}_k]$ and $\hat{\epsilon}_t = \epsilon_t - d_t$

$$\frac{1}{T}\sum_{t=1}^{T} \hat{\epsilon}_t^2 = \frac{1}{T}\sum_{t=1}^{T} \epsilon_t^2 + \frac{1}{T}\sum_{t=1}^{T} d_t^2 - 2\frac{1}{T}\sum_{t=1}^{T} \epsilon_t d_t. \tag{3.4}$$

Using Lemma 1 in Bai and Perron (1998) which holds also in the long-memory context for $d < 1/2$ (Beran et al. (1998)) and states that $\frac{1}{T}\sum_{t=1}^{T} \epsilon_t d_t = o_P(1)$ and the equations (3.3) and (3.4) it can be seen that $\frac{1}{T}\sum_{t=1}^{T} d_t^2 \xrightarrow{P} 0$. This states that $\hat{\kappa}$ contains the correct breakpoints among possible other incorrectly estimated mean shifts. Therefore, the regression tree is overfitted. However, pruning the tree by any under $I(d)$ consistent information criteria gives the desired consistency for the number and location of the mean shifts. $\diamond$

3.3 Pruning by means of the LWZ information criterion

The process of pruning is the ex post discarding of branches whose proportion to the error reduction is negligible. In order to find out the optimal sequence of partitions and breakpoints of all candidates a model selection criteria can be employed. The well-established BIC fails in the presence of long-range dependencies. It retains its consistency but is outperformed in finite sample studies (Bai and Perron (2004)).

Lavielle and Moulines (2002) suggested an information criterion based on the bayesian information criterion that penalizes the estimation with a term including the long memory parameter d. The LIC is defined by

$$LIC = \min_{1 \leq k \leq m} \min_{\kappa_1, \ldots, \kappa_m} \sum_{k=1}^{m+1} \sum_{t=[\kappa_{k-1}T]+1}^{[\kappa_k T]} (y_t - \hat{\mu}_k)^2 + \frac{4k \log T}{T^{1-2d}}.$$

The penalization is chosen in order to obtain a consistent estimator for the change-point and balances the number of over- and underestimation (see Lavielle and Moulines (2002)). The information criterion is built exclusively for the long memory case and leads to a necessary pre-specification of the underlying framework. Also the long memory parameter has to be estimated without being biased through potential mean shifts.

Liu, Wu and Zidek (1997) suggested a modified Schwarz criterion to estimate the number of sections of their multivariate regression model which is denoted as LWZ. This criterion takes the form

$$LWZ(m) = \ln(S_T(\hat{T}_1, \ldots, \hat{T}_m)/(T - p^*)) + (p^*/T)c_o(\ln(T))^{2+\delta_0},$$

where $c_0 > 0$ and $\delta_0 > 0$ are some constants and p^* describes the total number of fitted parameters. T denotes the total number of observations and $\hat{T}_i$ the number of observations of regime i. The idea is to change the well-established Schwarz criterion as little as possible to retain consistency but also to embrace the desire to construct a more flexible information criterion accordingly.

By minimizing the sum of squares of the residuals a model dependent best criterion is given. A reasonable choice of c_0 and δ_0 is suggested by Liu, Wu and Zidek (1997) for short memory processes. They set a small δ_0 (=0.1) to reduce the potential risk of underestimation with a normal noise distribution and estimate $c_0 = 0.299$ by equalizing the LWZ to the Schwarz information criterion, but call for further research to develop a globally optimal pair of c_0 and δ_0 under a variety of specifications which will be done in section 3.4.

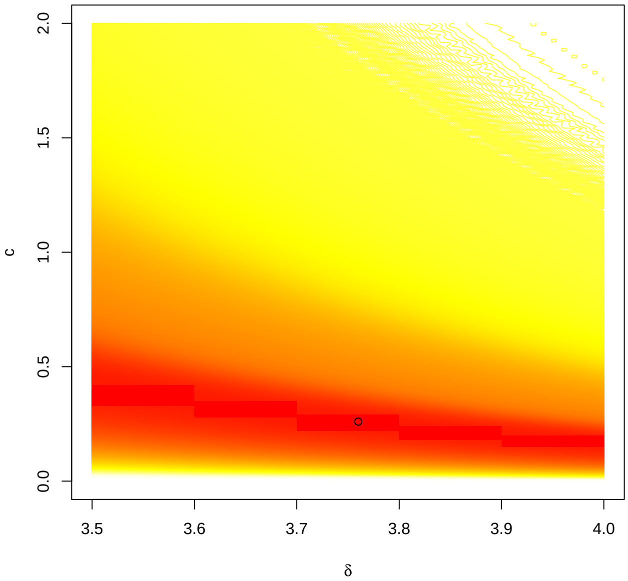
Bai and Perron (2004) show that the LWZ outperforms the BIC in all short memory cases including serial correlation. Under long memory the BIC is generally outperformed (see Rea (2008) and Rea et al. (2010) for demonstrative comparison). The classic BIC is therefore no competitor when it comes to performance questions.

3.4 Monte Carlo study

In the long memory context a simulation to specify a globally optimal pair for (c_0, δ_0) of the LWZ is done. Based on an ARFIMA(p,d,q) process with negligible short memory components for differentiation reasons no, one and two shifts in the mean of the time series is considered. For ten different values of the long memory parameter d (stationary and non-stationary) and a level shift height equally to the variance of the noise distribution (constantly 1) an overall distribution regarding the percentage of correctly specified breakpoints is computed. Under normal, t- and double exponential noise all combinations are examined.

Through a two-dimensional grid search procedure for all considered cases the optimal parameter pair ($c_0 = 0.26$, $\delta_0 = 3.76$, marked with a dot in figure 3.1) leads to 83% correct specifications. The performance deficit of 17% is based on high (nonstationary) d values and challenging break patterns when there are two mean shifts in the data.

Figure 3.1: Contour lines for correct specifications over all parameter combinations



See figure 3.1 for the contour plot of all considered parameter combinations. Yellow lines represent a low percentage of correct specifications and the more red the contour level line the higher the percentage of correct specification over all considered cases. The parameter combination with the highest percentage (83%) is marked with a dot in figure 3.1 and lies at $c_0 = 0.26$ and $\delta_0 = 3.76$. The LWZ would be accordingly

$$LWZ(m) = ln(S_T(\hat{T}_1, \ldots, \hat{T}_m)/(T - p^*)) + (p^*/T)0.26(ln(T))^{5.76}.$$

Not surprisingly, the penalization is typically higher than in the BIC (see Yao (1988)). As the BIC was constructed based on the iid case, the penalty term has to be somewhat stronger to balance the long-range dependance structure.

Besides an optimal parameter pair the graph also tells us that there is a rather wide central corridor for results of roughly equally good quality. That implies that the exact parameter combination is subordinate because of the stability of the results. The combination suggested by Liu, Wu and Zidek (1997) ($\delta_0 = 0.1$, $c_0 = 0.299$) is situated at the edge of the red corridor. Due to the fact that this combination leads to good results in the short memory case and outperforms the BIC, in general the LWZ is supposed to lead to good specification results as long as the penalty term is higher than the BIC.

In the short memory case the optimal parameter pair for long memory ($c_0 = 0.26$, $\delta_0 = 3.76$) leads to 89% correct specifications which makes the criterion safe to use for both frameworks without previous specification analysis. In the short memory case the *optimal* parameter pair would be a smaller value for c_0 with the same constant δ_0 or vice versa.

The Monte Carlo study serves as a comparison between the new adjusted LWZ criterion, the ordinary BIC as a benchmark information criterion and the LIC which is specialized in long memory cases. For ART we used tree growing procedures as implemented in 'tree' (Ripley (2005)) as a contributed package in the 'R' software. A time series with a length of 500 observations will be used and 100,000 replications are made since the computation time is not an issue.

The question that needs to be addressed after applying regression trees to time series according to Rea et al. (2010) is whether the pruning method under- or overestimates mean shifts and is robust against e.g. serial correlation. When there is no mean shift present in the data the results for the estimated number of mean shifts is given in table 3.1.

Table 3.1: Simulation results for pruning criteria when there are no mean shifts present

	LWZ			BIC			LIC		
d	% correct	mean	s.d.	% correct	mean	s.d.	% correct	mean	s.d.
0,05	100,00%	0,00	0,00	59,41%	0,65	0,95	50,23%	0,83	1,03
0,15	99,99%	0,00	0,00	18,86%	1,99	1,48	15,92%	2,15	1,49
0,25	99,94%	0,00	0,03	0,04%	3,36	1,52	0,03%	3,45	1,50
0,35	96,13%	0,04	0,19	0,00%	4,33	1,35	0,04%	4,36	1,33
0,45	77,62%	0,23	0,44	0,00%	4,87	1,18	0,10%	4,48	1,29
0,55	50,38%	0,57	0,65	0,00%	5,11	1,11	1,37%	3,26	1,45
0,65	27,21%	1,02	0,86	0,00%	5,17	1,10	6,14%	2,05	1,22
0,75	13,18%	1,56	1,09	0,00%	5,12	1,12	14,90%	1,34	0,93
0,85	5,91%	2,17	1,26	0,00%	4,99	1,15	24,92%	0,96	0,74
0,95	2,66%	2,70	1,32	0,00%	4,82	1,16	34,03%	0,76	0,65

The LWZ performs well when it comes to low and moderate long memory. For high values of d the increasing process variance of the underlying long memory tends to cover the true behavior of the mean. The BIC fails and tends to find at least one mean shift. The LIC develops a valley distribution. The shape of the estimation with the LIC is conditioned on the penalty term. With T^{2d-1} it degenerates for d values close to 0.5 and increases very strong for higher d values. For very small d values it performs well again because of the negligible long-range dependency. That's why for the LIC rather good results can be observed for low and high d values but not for moderate ones.

When it comes to a single mean shift at midpoint of the series the characteristics of the pruning criteria hold. For different break sizes that correspond to the standard deviation of the noise distribution ($s_{\epsilon_t} = 1$) see table 3.2. The position of the mean shift does not affect the estimations strongly though mean shifts in the boundary area weaken every criterion.

Table 3.2: Simulation results for pruning criteria when there is one mean shift present

$\mu_1 = 1,\ \mu_2 = 3$	LWZ			BIC			LIC		
d	% correct	mean	s.d	% correct	mean	s.d.	% correct	mean	s.d.
0,05	100,00%	1,00	0,00	86,76%	1,14	0,38	86,76%	1,14	0,38
0,15	100,00%	1,00	0,00	49,20%	1,69	0,81	49,20%	1,69	0,81
0,25	99,90%	1,00	0,03	16,40%	2,63	1,13	16,40%	2,63	1,13
0,35	96,52%	0,97	0,18	3,92%	3,60	1,24	3,92%	3,60	1,24
0,45	82,65%	0,89	0,41	0,85%	4,35	1,23	1,04%	4,27	1,23
0,55	65,85%	0,91	0,64	0,27%	4,82	1,17	7,33%	3,54	1,37
0,65	54,00%	1,16	0,89	0,12%	5,03	1,14	29,05%	2,26	1,25
0,75	42,95%	1,63	1,11	0,09%	5,06	1,14	50,13%	1,42	0,96
0,85	29,96%	2,18	1,26	0,07%	4,96	1,15	58,60%	0,99	0,76
0,95	18,83%	2,71	1,32	0,07%	4,81	1,16	57,99%	0,77	0,65
$\mu_1 = 1,\ \mu_2 = 2$									
0,05	38,39%	0,38	0,49	60,40%	1,51	0,72	59,78%	1,52	0,72
0,15	40,94%	0,41	0,49	23,57%	2,40	1,11	22,99%	2,41	1,10
0,25	41,48%	0,41	0,49	6,36%	3,42	1,27	6,14%	3,42	1,27
0,35	42,73%	0,43	0,50	1,59%	4,22	1,26	1,54%	4,23	1,26
0,45	45,51%	0,48	0,53	0,45%	4,77	1,19	1,13%	4,49	1,25
0,55	50,50%	0,67	0,66	0,15%	5,04	1,13	10,22%	3,36	1,44
0,65	51,64%	1,06	0,87	0,09%	5,13	1,11	32,14%	2,10	1,23
0,75	43,09%	1,58	1,10	0,06%	5,10	1,13	50,64%	1,36	0,94
0,85	30,11%	2,17	1,27	0,07%	4,98	1,15	58,33%	0,98	0,75
0,95	19,08%	2,71	1,33	0,05%	4,82	1,16	57,50%	0,76	0,65

The BIC again performs inferior with an average break estimation higher than 1. The LIC holds its shape and outperforms the LWZ for several combinations. The problem of the LIC still holds that d has to be estimated first and therefore can lead due to the deviance of the criterion easily to false results in a practical setting. The LWZ stays comparably constant when the long memory parameter changes and tends to underestimate the number of mean shifts for stationary long memory. Tree-based procedures in general overfit for small breaks and short observation length (see Rea et al. (2010)), hence a criterion which does not exceed this behavior could be a more than welcome technique.
For more than one mean shift the criteria weaken and are highly dependent on the break size but fortunately not on the break pattern.

3.5 Conclusion

Estimating the number of mean shifts in a long-memory time series can be challenging. Tree-based procedures are presented as a powerful yet simple technique (see De'ath, G., Fabricius, K. (2000)) and are therefore useful for the practitioner (Rea et al. (2010)). To prune the overfitting of atheoretical regression trees the BIC is widely used in a short memory framework and surprisingly outperformed by the LWZ under multiple specifications (Bai and Perron (2004)). The LIC which was derived for long memory shows good properties as well and partially outperforms the LWZ for some combinations of d. Though the disadvantages of the LIC to depend on the true value of d last. The LWZ keeps reasonable results even when the framework contains long memory and thus needs no beforehand knowledge of the data generating process. It is therefore preferable to the BIC and LIC.

Chapter 4

Testing for a break in persistence under long-range dependencies and mean shifts

Testing for a break in persistence under long-range dependencies and mean shifts

Co-authored with Philipp Sibbertsen.

Published as Leibniz University of Hannover Discussion Paper No. 422.

Forthcoming in Statistical Papers.

4.1 Introduction

It is well known that structural breaks in the mean of a time series can easily be confused with long-range dependence. Shifts in the mean can heavily bias estimators for the memory parameter and therefore create misleading results. For an overview about the problem of spurious long memory due to mean shifts see Sibbertsen (2004). In the recent years a change of the persistence of a time series, this is a change of the order of integration, has come more and more into the focus of empirical and theoretical researchers. Beginning with Banerjee et al. (1992) several authors proposed tests for a change in persistence in the classical $I(0)/I(1)$ framework. A popular stationarity test against a break in persistence was introduced by Kim (2000). Kim's test has the disadvantage to reject the null if the data generating process is constantly $I(1)$ during the whole sample what is theoretically correct but not desirable. Leybourne et al. (2007) suggest a CUSUM-squares based test to solve this problem. Sibbertsen and Kruse (2009) generalized this test to the long memory framework by allowing for fractional degrees of integration.

Belaire-Franch (2005) proved that Kim's test is not robust against mean shifts in the sense that it has an asymptotic size of one when the data generating process is $I(0)$ with a break in the mean. Unfortunately, we show that the Leybourne et al. test does not overcome this problem as it is not robust against mean shifts either. We therefore derive adjusted critical values for the test under a generalized de-trending allowing for one mean shift.

The rest of the paper is organized as follows. In section 4.2 the test for changes in persistence is briefly described. Section 4.3 derives its properties under mean shifts and section 4.4 contains some Monte Carlo studies. Section 4.5 gives critical values of the test under a generalized de-trending procedure. Size and power results are given as well. Section 4.6 contains an empirical example showing the usefulness of our de-trending procedure in practice and Section 4.7 concludes.

4.2 Testing for a break in persistence under long memory

We assume that the data generating process follows an ARFIMA$(0,d,0)$ process. Sibbertsen and Kruse (2009) generalized a CUSUM of squares-based type test proposed by Leybourne et al. (2007) to test in this model framework the hypothesis of constant long-range dependencies versus a change in persistence. The alternative can be either a change in persistence from stationary to non-stationary long memory or vice versa. The null hypothesis tested is

$$H_0 : d = d_0 \quad \text{for} \quad t = 1,\dots,T,$$

where we assume $1/2 < d_0 < 3/2$. The alternative hypothesis is either

$$H_{01} \quad : \quad \begin{cases} d = d_1 \in (0,1/2) & \text{for } t = 1,\dots,[\tau T] \\ d = d_2 \in (1/2,3/2) & \text{for } t = [\tau T]+1,\dots,T \end{cases}$$

or

$$H_{10} \quad : \quad \begin{cases} d = d_2 \in (1/2,3/2) & \text{for } t = 1,\dots,[\tau T] \\ d = d_1 \in (0,1/2) & \text{for } t = [\tau T]+1,\dots,T. \end{cases}$$

The CUSUM of squares-based test statistic R used in Sibbertsen and Kruse (2009) is given by

$$R = \frac{\inf_{\tau\in\Lambda} K^f(\tau)}{\inf_{\tau\in\Lambda} K^r(\tau)}$$

with the forward statistic

$$K^f(\tau) = [\tau T]^{-2d_0} \sum_{t=1}^{[\tau T]} \hat{v}_{t,\tau}^2$$

and the reversed statistic of the data generating process

$$K^r(\tau) = (T - [\tau T])^{-2d_0} \sum_{t=1}^{T-[\tau T]} \tilde{v}_{t,\tau}^2.$$

Here τ is the relative breakpoint where we assume that $\tau \subset \Lambda$ and $\Lambda \subset (0,1)$ and is symmetric around 0.5. For now we assume τ to be fixed though unknown. $[x]$ is the ceiling function of x and $\hat{v}_{t,\tau}$ is the residual from the OLS regression of X_t on a constant $z_t = 1 \; \forall t$ based on the observations up to $[\tau T]$. This is

$$\hat{v}_{t,\tau} = X_t - \bar{X}(\tau)$$

with $\bar{X}(\tau) = [\tau T]^{-1} \sum_{t=1}^{[\tau T]} X_t$. Similarly $\tilde{v}_{t,\tau}$ is defined for the reversed series $y_t = X_{T-t+1}$. Thus, it is given by

$$\tilde{v}_{t,\tau} = y_t - \bar{y}(1-\tau)$$

with $\overline{y}(1-\tau) = (T - [\tau T])^{-1} \sum_{t=1}^{T-[\tau T]} y_t$.

Sibbertsen and Kruse (2009) derive the limiting distribution of this test statistic and provide response curves in order to compute critical values for different hypothetical memory parameters d_0.

4.3 Behavior of Test under mean shifts

In order to analyze how the CUSUM of squares-based test behaves under mean shifts let us introduce some notation first. In what follows τ denotes the relative breakpoint in the memory parameter d and λ denotes the relative position of the mean shift. For the sake of notational simplicity we only consider the easiest break in mean model allowing only for abrupt changes. Our model is given by

$$y_t = \alpha + \delta D_t + \varepsilon_t \tag{4.1}$$

with $D_t = 1(t \geq [\lambda T] + 1)$ with $1(\cdot)$ being the indicator function. In this model a level shift from α to δ occurs at some unknown breakpoint $[\lambda T]$. We further assume that $\varepsilon_t \sim I(d)$ with $0 \leq d \leq 1.5$. Thus, a possible choice for ε_t is an $ARFIMA(0,d,0)$ model. Let furthermore $\overset{P}{\rightarrow}$ denote convergence in probability.

Theorem 4.1.
Given model (4.1) with the assumptions given above. Then:

1. *for* $1/2 < d < 3/2$ *the value of the test statistic is*

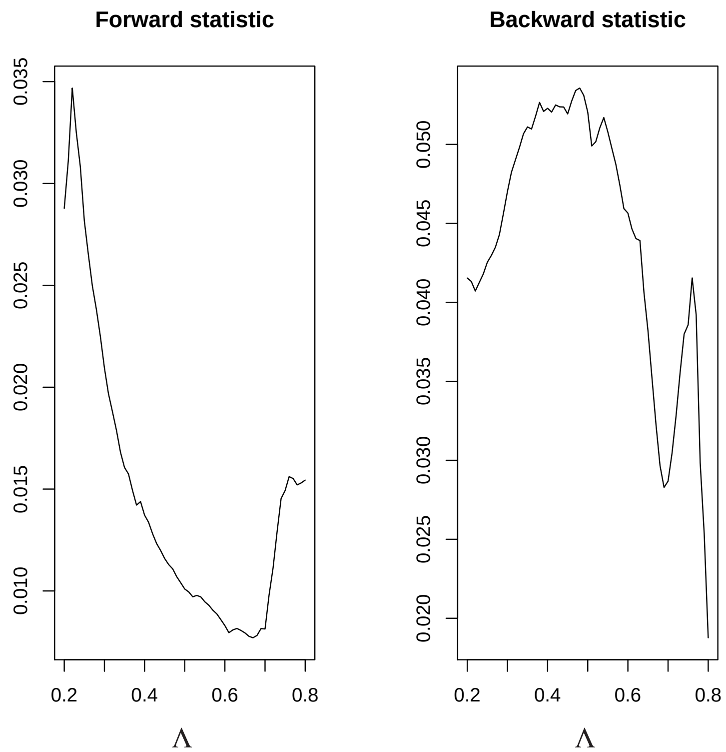
$$R = \frac{\inf_{\tau \in \Lambda} K^f(\tau)}{\inf_{\tau \in \Lambda} K^r(\tau)} = \frac{\inf_{\tau \leq \lambda} K^f(\tau)}{\inf_{\tau \geq \lambda} K^r(\tau)};$$

2. *for* $0 \leq d < 1/2$ *we have* $R \overset{P}{\rightarrow} 1$.

The results can also be derived for a general de-trending. The ideas are the same. It only introduces more notational difficulties and is therefore left out here.

The result means that the minimization takes place over a restricted interval up to the point where the mean shift occurs or beginning from this point. The further the mean shift is on the limits of Λ the smaller is this interval either for the forward or reversed statistic. Therefore, the occurrence of the minimum in this interval becomes less likely. This can be seen when considering a typical shape of the forward and reversed statistic as given in Figure 4.1. At $\lambda = 0.7$ the forward statistic increases immediately and so the minimum can only be found before the mean shift distorts the forward statistic. This distortion is big enough for the test statistic to reject the null in most cases. It should be mentioned that we cannot prove inconsistency of the test in the sense that the test statistic diverges when a mean shift occurs. This is not the case and thus allows us to readjust the critical values in the case of mean shifts as it is done in section 4.5.

Figure 4.1: Forward and backward statistic with $\lambda = 0.7$, $\alpha = 0$, $\delta = 5$ and $d = 0.8$



The size distortions are smallest for a mean shift at $\lambda = 0.5$ considering that the interval for the forward and backward statistic have the same length. Therefore, it is less likely that both minima findings are distorted. Interestingly, these results do not hold for a stationary data generating process. In this case the test statistic is still conservative. Some Monte Carlo underpinning these findings is given in the next section.

4.4 Monte Carlo study

Our theoretical findings in section 4.3 can be backed up with Monte Carlo studies. All simulations are computed with the open-source programming language R (2008). The number of replications is set to $M = 2000$ and we consider a sample size of $T = 1000$, set so high in order to illustrate the asymptotic results. When there is a mean shift from $\alpha = 0$ to $\delta = 5$ in model (1), the size varies with the relative position of the mean shift λ as follows.

Table 4.1: Empirical size when there is a mean shift using estimated response curves

	$d = 0.6$					$d = 0.8$				
λ	0.10	0.25	0.50	0.75	0.90	0.10	0.25	0.50	0.75	0.90
1L	0.15	0.00	3.75	43.95	34.30	0.35	0.60	1.00	2.60	2.90
5L	0.70	0.65	11.75	67.45	57.20	2.40	3.05	5.15	10.00	11.60
10L	1.20	1.05	19.45	77.15	71.35	5.35	7.00	10.55	17.95	19.90
10U	70.40	74.80	15.80	0.50	0.75	19.20	18.45	10.05	6.75	6.60
5U	55.25	62.45	8.40	0.10	0.4	12.00	11.70	5.60	3.30	3.05
1U	30.45	38.30	2.65	0.00	0.10	3.40	2.70	0.80	0.60	0.50
	$d = 1.0$					$d = 1.4$				
λ	0.10	0.25	0.50	0.75	0.90	0.10	0.25	0.50	0.75	0.90
1L	0.75	1.55	1.20	1.20	0.85	0.15	0.20	0.15	0.1	0.15
5L	4.10	6.00	5.10	4.90	4.90	2.25	2.70	2.15	1.70	2.00
10L	9.30	11.95	9.60	10.05	10.45	6.55	5.65	5.50	4.3	5.25
10U	9.45	8.65	9.65	10.35	10.40	5.55	7.00	6.55	6.45	6.1
5U	4.60	4.85	5.40	5.00	6.30	2.00	2.55	2.45	2.55	2.4
1U	1.00	1.30	1.20	1.00	1.00	0.00	0.30	0.10	0.15	0.05

As shown in section 4.3 it leads to distorted size results for $1/2 < d < 3/2$ no matter what shift size is used. For $d < 1$ it remains most likely above the significance level. The size distortion increases by getting closer to the limits of the Λ interval. For $d = 1$ as well as for $\lambda = 0.5$ the smallest size distortion can be observed. For $d > 1$ the test statistic tends to conservative size results. The test statistic does not diverge because of a mean shift and tends to reject not properly. Because of the missing mean reverting characteristic for long memory with $d > 1$ and the thereby explosive performance of the time series, the mean shift no matter what size has no such strong impact on the test statistic and hence on the size results.

For this onesided test depending on whether λ is smaller or greater than the interval Λ, elevated size values appear at the upper and lower bound respectively as shown in Figures 4.2 and 4.3. Due to the fact, that the true position of the break is unknown, distorted size results can always appear.

Figure 4.2: Behavior of the size at the lower 5% tail

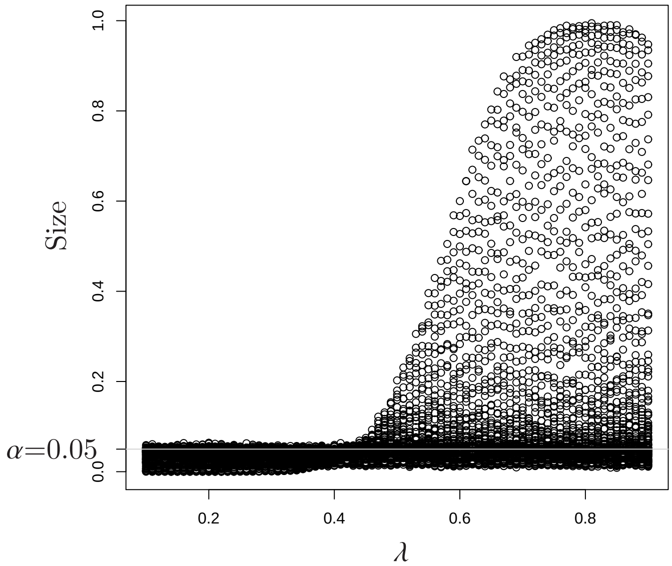
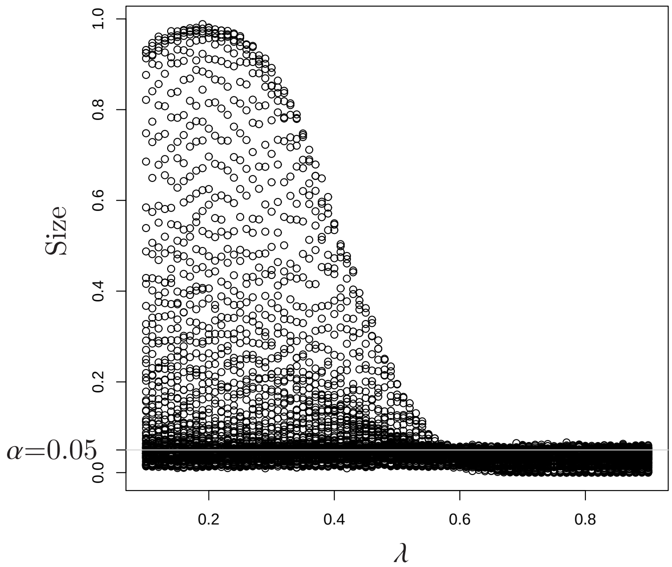
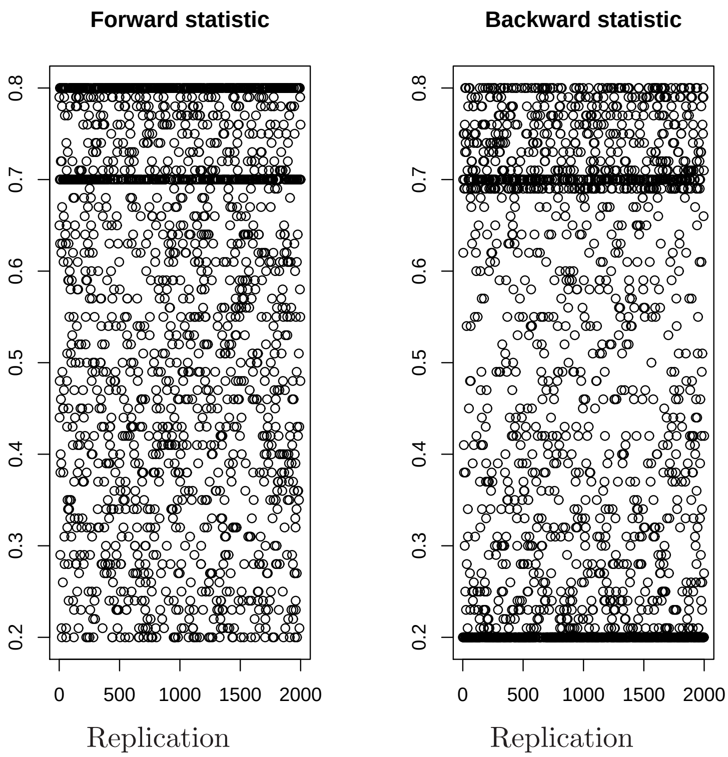


Figure 4.3: Behavior of the size at the upper 5% tail



The smaller the long memory parameter, the more distinctive is this size behavior. Hence at the boundaries of the time series the test decision is strongly biased by the mean shift and leads to a false rejection of the null. The following graphic show the distribution of the minima of the forward and backward statistics for $\lambda = 0.7$ and $d = 0.8$. It shows that the minima of the forward statistic cumulate at the boundary of 0.8 and around $\lambda = 0.7$. The reversed statistic shows similar findings with a cumulation at 0.2.

Figure 4.4: Empirical minima of the forward and backward statistics



4.5 Adjustment of critical values

Due to the size distortion at the boundaries it is reasonable to adjust the critical values and take the mean shifts into account. The adjustment of the critical values takes place under the allowance for one break in the mean. In addition to this situation which is detailed in the theoretical part we consider also the adjustment of the critical values if there is a break in the mean and in the slope of the linear trend. This adjustment procedure goes conform with the situation of breaks in the conditional mean in the de-trending case. Furthermore, we allow for a smooth transition between the regimes allowing a higher flexibility in the trend function. The smooth transition is driven by a logistic transition function. The abrupt mean shift model is a special case of this more general mean shift model. It should be mentioned that for our adjustment procedure the existence of the mean shift has to be known. Estimating mean shifts within a long memory model with breaking persistence is a difficult task and beyond the scope of this paper. It should be mentioned that the response curves given in this chapter and thus the critical values of the test depend on λ. However, as in most applications there are at least rough if not exact ideas about mean shifts in the data, we consider our procedure still as useful for the practitioner.

We simulate the asymptotic distribution of the test statistic depending on d for the cases $d = 0.51$ to $d = 1.49$ with $\lambda = 0.5$. Due to the wide range of possible values of d we fit polynomial functions to the sequence of critical values depending on d. The adjusted critical values can be displayed in response curves given by

$$q_\alpha(d) = \sum_{i=0}^{s} \beta_i d^i.$$

q_α denotes the α-quantile of the asymptotic distribution and s the maximal polynomial order which is set to nine. The parameters β_i are estimated with OLS. For different values of λ the response curves are parallel so the functional form remains unchanged for different values of λ though the parameters change.

Table 4.2: Estimated response curve when a mean shift occurs

	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
1L	0.965	0	0	0	-21.507	66.386	-89.993	65.334	-24.904	3.927
5L	-162.657	1700.856	-7729.333	20149.046	-33212.416	35896.289	-25444.859	11411.826	-2940.170	331.773
10L	0.931	0	0	0	-2.550	2.475	0	0	-0.675	0.283
10U	1.132	0	0	0	0	0	10.268	-18.031	11.346	-2.557
5U	1.161	0	0	0	0	0	16.821	-30.738	20.506	-4.932
1U	0.975	0	0	0	18.784	-41.418	39.564	-13.136	0	0

OLS estimates for β_i ($i = 0,1,\dots 9$) are reported in columns; $\beta_i = 0$ means that the parameter is set equal to zero.

Table 4.3: Estimated response curve when a mean and a slope shift occurs

	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
1L	0.9277	0	0	-1.8787	1.352	0	0	-0.0764	0	0
5L	0.9022	0	0	0	0	-5.2669	9.6113	-6.0405	1.3009	0
10L	1.035	0	0	0	-21.792	86.363	-144.9	123.989	-53.224	9.103
10U	1.068	0	0	0	1.395	-0.732	0	0	0	0
5U	1.065	0	0	0	2.129	-1.083	0	0	0	0
1U	1.1842	0	0	0	2.1661	0	0	-0.2962	0	0

OLS estimates for β_i ($i = 0,1,\dots 9$) are reported in columns; $\beta_i = 0$ means that the parameter is set equal to zero.

The size and power properties of the test using the estimated response curves for one break in the mean are reported in Tables 4.4 and 4.6, and for a break in the mean and the slope in Table 4.5 and 4.7 respectively.

Table 4.4: Empirical size for mean shift

d	0.55	0.70	0.85	1.25
1L	0.4	1.0	0.7	2.2
5L	3.3	5.3	3.9	3.4
10L	8.5	9.7	9.8	8.1
10U	10.4	9.3	11.1	10.2
5U	5.6	5.0	5.8	5.8
1U	1.0	1.2	1.3	1.2

Table 4.5: Empirical size for mean and slope shift

d	0.55	0.70	0.85	1.25
1L	0.6	0.8	1.3	1.0
5L	3.5	5.4	6.2	5.4
10L	8.9	9.9	11.7	9.8
10U	9.6	8.5	8.9	10.1
5U	5.9	5.0	4.5	5.2
1U	1.0	1.0	1.1	1.0

Table 4.6: Power Experiment for one break at the 5% level

d	$0.8 \rightarrow 0.4$	$0.4 \rightarrow 0.8$	$0.6 \rightarrow 0.0$	$0.0 \rightarrow 0.6$	$0.6 \rightarrow 0.4$	$0.4 \rightarrow 0.6$
	83.7	96.2	96.0	77.0	58.5	54.9

Table 4.7: Power Experiment for one break in the mean and slope at the 5% level

d	$0.8 \rightarrow 0.4$	$0.4 \rightarrow 0.8$	$0.6 \rightarrow 0.0$	$0.0 \rightarrow 0.6$	$0.6 \rightarrow 0.4$	$0.4 \rightarrow 0.6$
	96.5	95.6	92.1	92.0	67.3	64.9

The size experiments with these adjusted critical values show that it is useful to correct for the effect of the mean shift. When it is known or likely that the time series contains a mean shift the test gains good size properties and appropriate power results. This is very helpful to know when you consider the additional size distortion if the mean shift is neglected. It should be mentioned that the model can also be extended to more than one break.

4.6 Empirical Example

In order to show the usefulness of our adjustment procedure we consider harmonized monthly CPI inflation rates for the UK from 01.1989 to 03.2008 and USA from 01.1950 to 03.2008. The data is obtained from Datastream. The series are depicted in Figure 4.5 and 4.6 below where also the fitted trend functions, the residuals and the break in the persistence parameter is shown. The data for the USA is the same as in Sibbertsen and Kruse (2009) to obtain comparability. Both series exhibit long-range dependence before and after de-trending. For the UK we obtain for the whole series a value of $d = 1.151$ before the de-trending and $d = 0.835$ after de-trending. For the USA we have $d = 1.215$ before and $d = 1.227$ after de-trending. Both series do have long-range dependencies even after a general de-trending allowing for shifts in the trend.

Applying the test of Sibbertsen and Kruse (2009) for the constancy of the persistence and neglect possible breaks in the trend the null of no break in the persistence parameter is rejected for both series at the 10% level indicating a change in the memory. For the USA this finding goes conform with the findings in Sibbertsen and Kruse (2009).

However, these results change when allowing for a break in the mean and the slope of the trend function and applying our general de-trending procedure before applying the test for a break in persistence. The series with the fitted trend function and the residuals are shown in Figures 4.5 and 4.6.

Figure 4.5: UK inflation with trend and residuals

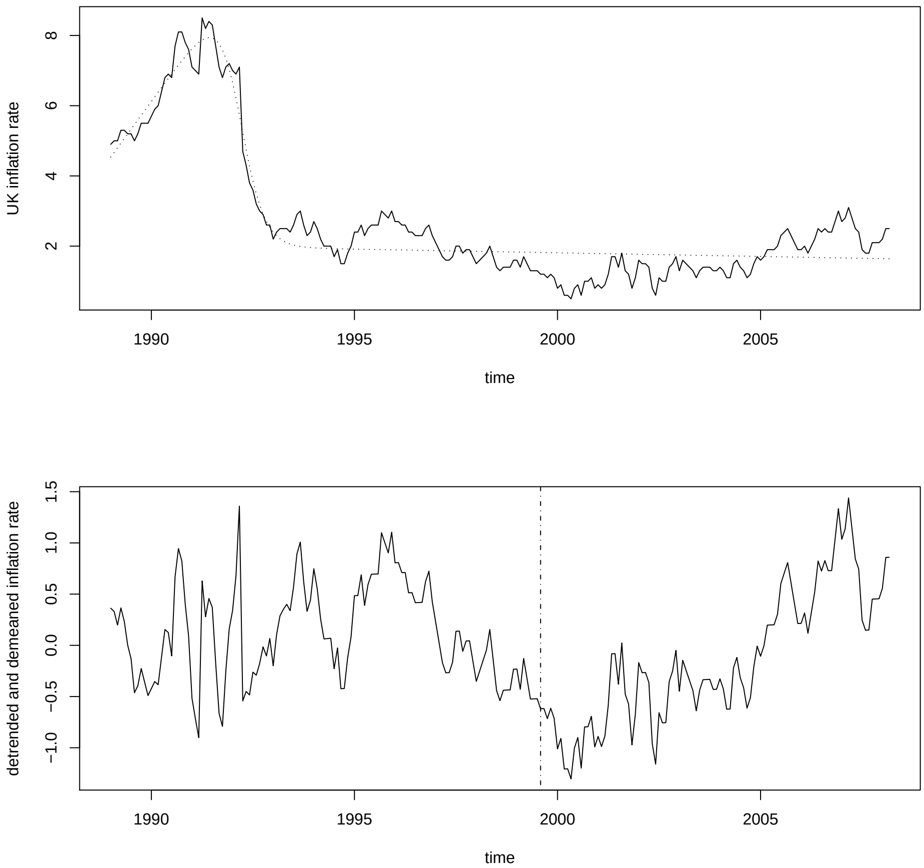
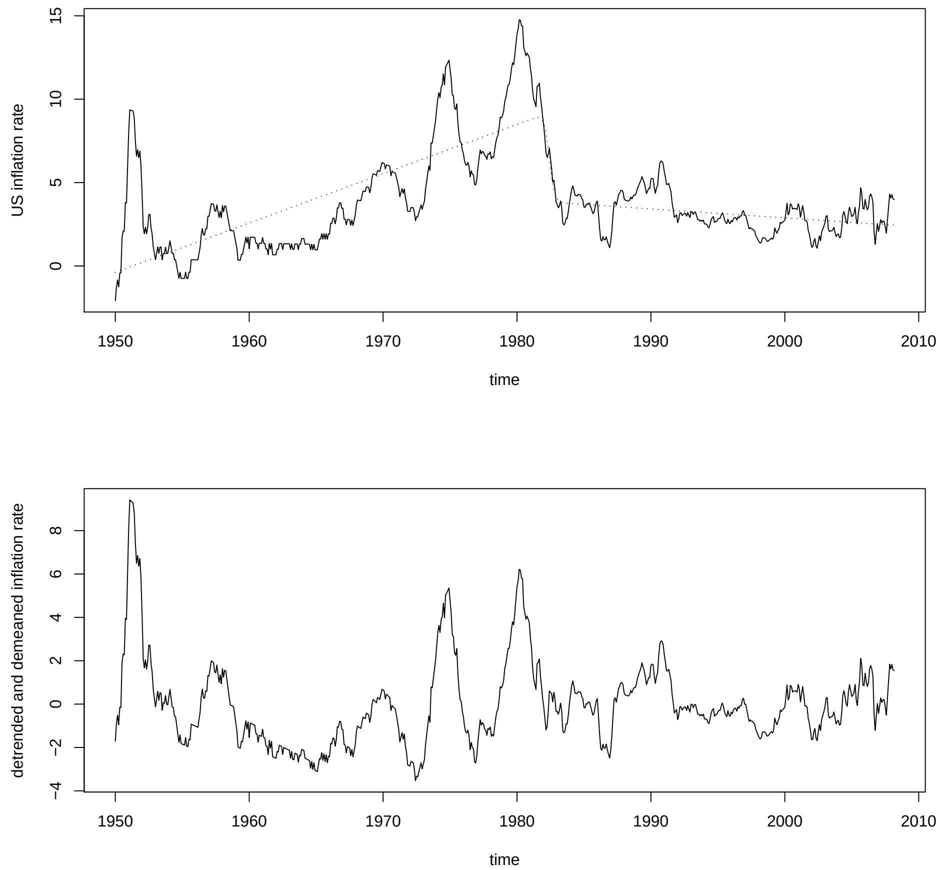


Figure 4.6: US inflation with trend and residuals



Both series have clear mean shifts. However, after eliminating these mean shifts there still seems to be a persistence change within the residuals of the UK inflation rates whereas the residuals of the US inflation do not look like having a breaking persistence. Application of the test for changing persistence confirms this. Whereas the null of a constant persistence cannot be rejected for the US inflation at any level of significance the null for the UK inflation can be rejected at the 1% level of significance indicating both a breaking trend plus a breaking persistence.

Estimating the breakpoint for the persistence break in the UK inflation shows a break at $\tau = 0.54$ which is 07.1999. The breakpoint is indicated by the dashed line in Figure 4.5. Estimation of the memory parameter gives $d = 0.799$ before the break and $d = 1.034$ after the break suggesting that the UK inflation follows basically a random walk since mid of 1999.

4.7 Conclusion

In this paper we show that the Leybourne et al. (2007) test on a break in persistence becomes biased when the data generating process has a shift in the mean function. The test is therefore not robust against mean shifts. The size of the test is most likely even higher than the chosen significance level. Therefore, the null of no change in persistence is falsely rejected by the test due to mean shifts. Mean shifts do effect the test decision even more when they occur at the extreme ends of the sampling period.

As the test is distorted when a mean shift occurs, it is useful to correct for this effect when it is known or likely to have mean shifts in the data. We give adjusted critical values for the case of one mean shift (and trend shift) and provide response curves for them. It is shown that the test has good size and reasonable power properties.

Without applying our adjustment procedure the null of a constant persistence has to be rejected for both series. After the application of our adjustment procedure to monthly inflation rates of the UK and the US it can be seen that the null of constant persistence cannot be rejected any more for the US after a general de-trending whereas it still gets rejected for the UK. Hence, we find a breaking trend and a breaking persistence for the UK inflation.

4.8 Appendix

4.8.1 Proof

Proof of Theorem 4.1:

1. Let us first assume that $0.5 < d < 1.5$. Let us furthermore assume that $\tau \leq \lambda$. The case $\tau \geq \lambda$ is analogous with an interchange of the forward and reverse statistics.

   The main advantage of our simple breakpoint model is that we only have to consider the case of a de-meaning of the time series. Due to the fact that a level shift occurs we consider the case of de-meaning instead of de-trending which would be appropriate in the case of a broken trend. For the residuals of (4.1) we have before the persistence break

$$\hat{e}_j = \varepsilon_j - [\tau T]^{-1} \sum_{t=1}^{[\tau T]} \varepsilon_t$$

   respectively afterwards

$$\hat{e}_j = \varepsilon_j - [(1-\tau)T]^{-1} \sum_{t=[\tau T]+1}^{T} \varepsilon_t - [(1-\tau)T]^{-1} \delta \sum_{t=[\tau T]+1}^{T} D_t + \delta D_j.$$

   Assume $\tau \leq b \leq \lambda$ and $t = [bT]$. For a fixed τ the mean shift is behind the assumed persistence shift and thus the forward statistics remains unchanged:

$$K^f(\tau) = [\tau T]^{-2d_0} \sum_{t=1}^{[\tau T]} \hat{v}_{t,\tau}^2 \to L_d^f(\tau).$$

   Have in mind that the test always works under the alternative and therefore the existence of a persistence shift is assumed.

For the reversed statistic $\tilde{v}_{t,\tau}$ we obtain:

$$\begin{aligned}
\tilde{v}_{[bT],\tau} &= \tilde{v}_{[\lambda T],\tau} + \tilde{v}_{[\lambda T]+1,\tau} \\
&= \sum_{j=1}^{T-[\lambda T]} \varepsilon_j - \sum_{j=1}^{T-[\lambda T]} \bar{\varepsilon} - \delta \sum_{j=1}^{T-[\lambda T]} \bar{D} \\
&\quad + \sum_{j=T-([\lambda T]+1)}^{T-[bT]} \varepsilon_j - \sum_{j=T-([\lambda T]+1)}^{T-[bT]} \bar{\varepsilon} - \delta \sum_{j=T-([\lambda T]+1)}^{T-[bT]} \bar{D} + \sum_{j=T-([\lambda T]+1)}^{T-[bT]} \delta D_j
\end{aligned}$$

with $\bar{\varepsilon}$ and $\bar{D}$ being the mean of ε and D over the respective time interval.
If $\lambda \le \tau$ the reversed statistic remains unchanged and we have for the forward statistic:

$$\begin{aligned}
\hat{v}_{[bT],\tau} &= \hat{v}_{[\lambda T],\tau} + \hat{v}_{[\lambda T]+1,\tau} \\
&= \sum_{j=1}^{[\lambda T]} \varepsilon_j - \sum_{j=1}^{[\lambda T]} \bar{\varepsilon} - \delta \sum_{j=1}^{[\lambda T]} \bar{D} \\
&\quad + \sum_{j=[\lambda T]+1}^{[bT]} \varepsilon_j - \sum_{j=[\lambda T]+1}^{[bT]} \bar{\varepsilon} - \delta \sum_{j=[\lambda T]+1}^{[bT]} \bar{D} + \sum_{j=[\lambda T]+1}^{[bT]} \delta D_j.
\end{aligned}$$

The statistic is minimized over all $\tau \in \Lambda$ up to λ in the first situation and afterwards in the second. This means that up to $\tau = \lambda$ the forward statistic remains unchanged and afterwards the mean shift will effect the residuals by reason that the de-meaning has to consider the mean shift. Thus, for $\tau > \lambda$ the square of the forward statistic increases and therefore the minimum is in the interval $\tau \le \lambda$ and it is greater or equal the minimum which is obtained without a mean shift.

We have a similar argument for the reversed statistic. For $\tau > \lambda$ it remains unchanged. The changing mean does not affect the recursive de-meaning and thus the residuals remain unchanged. For $\tau < \lambda$ the reversed statistic increases and the minimum is thus in the interval $\tau \ge \lambda$. This proves the first part of the theorem.

2. Let us finally consider the case where $0 \le d < 0.5$. Because of the arguments used before, the minimum of the forward statistic is located earlier than λ and that of the backward statistic later than λ. Therefore, we are in a similar situation as in Sibbertsen and Kruse (2009), Theorem 4, and can therefore adopt the same arguments as in their proof. $\diamond$

Chapter 5

Monitoring a change in persistence of a long range dependent time series

Monitoring a change in persistence of a long range dependent time series

Co-authored with Florian Heinen.

Published as Leibniz University of Hannover Discussion Paper No. 479.

5.1 Introduction

The assumption of structural stability of an econometric model is a major issue in time series econometrics. If the parameter estimates stem from an unstable relationship they are not meaningful and additionally inference can be biased and forecasts yield inaccurate results (see e.g. Hansen (2001), Andrews and Fair (1988), Ghysel et al. (1997), Garcia and Perron (1996) or Clements and Hendry (1998)). In reaction to these findings a large amount of literature emerged that incorporated structural change in the inference techniques or analyzes forecasting subject to structural change more closely (see e.g. Perron (1989), Zivot and Andrews (1992) or Pesaran and Timmermann (2005)). Recently the possibility of a change in persistence, i.e. a change in the memory structure of the time series as a special case of structural instability, has become object of study (see e.g. Kim (2000), Kim et al. (2002), Busetti and Taylor (2004), Banerjee et al. (1992), Leybourne et al. (2003) or Leybourne et al. (2007)). This work has been placed within the $I(0)$ vs. $I(1)$, or vice versa, framework where the focus lies on short memory time series with an exponentially decaying autocorrelation structure.

However, since the seminal papers of Granger and Joyeux (1980) and Hosking (1981), long memory time series have become widely used in economics to model highly persistent time series as diverse as inflation rates or realized volatility (see e.g. Hassler and Wolters (1995) and Corsi et al. (2008)). Baillie (1996) provides an overview about various applications of long memory time series in economics.

Despite these facts little work has been done to test for a change in persistence in long range dependent time series. Notable exceptions are Beran and Terrin (1996), Ray and Tsay (2002), Sibbertsen and Kruse (2009) or Yamaguchi (2011). These test belong to the class of so-called ”one-shot” tests (see Chu et al. (1996, p. 1045)), i.e. tests that are applied a posteriori to detect a structural break within a historical data set.

Because breaks can occur at any given time and also new data arrives steadily it is desirable for the applied econometrician to detect a change in persistence as soon as possible. This leads to a sequential testing problem (see Siegmund (1985) for an overview). As the usual ”one-shot” tests work with constant critical values they cannot be applied sequentially given that the true null of no change would eventually be rejected with probability one (see Robbins (1970)). Starting with Bauer and Hackl (1978) a strand of literature has emerged that studies monitoring procedures that allow to detect structural change whenever new data arrives. Important contributions on this field are Chu et al. (1995), Kuan and Hornik (1995), Chu et al. (1996), Leisch et al. (2000), Altissimo and Corradi (2003), Zeileis et al. (2005), Andreou and Ghysels (2006) and Hsu (2007). These papers contribute to the literature on monitoring structural stability on different levels ranging from theoretical contributions to detecting structural change in the conditional mean or the conditional variance or comparing different types of rejection regions for the null.

In this paper we use a monitoring approach based on moving sums of residuals and place it into a long memory framework. We develop a procedure to detect an increase in persistence for the case that the process becomes non-stationary. This is important because an increase in persistence implies a loss of controllability for important macroeconomic time series such as inflation rate or the European overnight rate (EONIA) (see Sibbertsen and Kruse (2009) and Hassler and Nautz (2008)). Further, a change in persistence also affects forecast accuracy in long memory time series (see Heinen at al. (2009)).

The rest of the paper is organized as follows: In section 5.2 we describe the test procedure we use and develop the asymptotic behavior. We further discuss and motivate different forms of boundary functions for the test. In section 5.3 we undertake a simulation study to asses the finite sample performance of the monitoring test. Section 5.4 contains an empirical application before section 5.5 concludes. All proofs are collected in the appendix 5.6.

5.2 Monitoring a change in persistence

We assume that the data generating process follows an ARFIMA(p,d,q) process as proposed by Granger and Joyeux (1980)

$$\Phi(L)(1-L)^d y_t = \Theta(L)\varepsilon_t, \text{ with } \varepsilon_t \overset{iid}{\sim} (0,\sigma^2) \text{ and } t = 1,2,\ldots,T\ . \tag{5.1}$$

The differencing parameter d can take fractional values but is assumed to be $|d| < \frac{1}{2}$. Thus the process y_t is in the stationary region (see e.g. Beran (1995)).

Bauer and Hackl (1978) propose the use of moving sums of cumulated residuals (MOSUM) to detect parameter changes in regression models. These tests are further investigated by Chu et al. (1995).

We are interested in detecting a change in persistence, i.e. a change in the fractional differencing parameter d, in the monitoring period $T+1$ up to $[T\tau]$, $\tau > 1$. Where $[\cdot]$ denotes the integer part of its argument.

In particular, we test the null of no change in persistence, i.e. $d = d_0$ within the monitoring period where $|d_0| < \frac{1}{2}$, against the alternative of an increase in persistence. More formally we test the null that

$$H_0 : d_\ell = d_0, \quad \ell = T+1, \ldots, [T\tau], \tag{5.2}$$

against the alternative that at some point in the monitoring period the persistence increases and $\frac{1}{2} < d_\ell < \frac{3}{2}$. Thus we test whether the process stays in the stationary region throughout the whole monitoring period or changes into the non-stationary region with an infinite variance at some point in the monitoring period. For the period from $t = 1, \ldots, T$ we follow Chu et al. (1996) and make the "noncontamination" assumption that

$$d_t = d_0, \quad t = 1, \ldots, T,$$

with $|d_0| < \frac{1}{2}$. Consider for simplicity the case of an ARFIMA(0,d,0) process.

Let $\hat{e}_t$ be an ARFIMA(0,d,0) process as in (5.1) and $\hat{\sigma}^2 = T^{-1}\sum_{i=1}^{T}\hat{e}_i^2$ a consistent estimator of σ^2. Based on a moving sum of residuals obtained from a fixed window size $[Th]$, $0 < h \leq 1$, the prototypical MOSUM test reads

$$MS_{T,h,d} = \max_{T+1 \leq k \leq [T\tau]} \sigma^{-1} T^{-\frac{1}{2}-d} \left| \sum_{i=k-[Th]+1}^{k} \hat{e}_i - \frac{[Th]}{T} \sum_{i=1}^{T} \hat{e}_i \right|, \tag{5.3}$$

for each value k in the monitoring period $T+1$ through $[T\tau]$.

The next theorem gives the asymptotic behavior of the test statistic in (5.3) if y_t follows a long range dependent process as in (5.1) and (5.2).

Theorem 5.1.
Assume the process y_t follows an ARFIMA$(0,d,0)$ process as in (5.1) with $|d| < \frac{1}{2}$. Then, as $T \to \infty$, we have for $MS_{T,h,d}$ in (5.3) that

$$MS_{T,h,d} \Rightarrow \frac{1}{\sigma} \max_{t \in [1,\tau]} \left| BB^0(t,d) - BB^0(t-h,d) \right|,$$

where $BB^0(t,d)$ denotes a fractional Brownian Bridge depending on fractional Brownian motion with parameter d. $\Rightarrow$ denotes weak convergence on a function space.

Under the alternative of a break in persistence the test is consistent.

The limiting distribution thus depends on the increments of a fractional Brownian bridge which in turn depends on the differencing parameter d of the data generating process. Therefore the asymptotic critical values of $MS_{T,h,d}$ are determined by the boundary crossing probabilities of the increments of a fractional Brownian bridge:

$$\mathbb{P}\{MS_{T,h,d} \leq b\} = \mathbb{P}\left\{\left|BB^0(t,d) - BB^0(t-h,d)\right| \leq b\right\}. \tag{5.4}$$

The use of the test statistic in (5.3) is beneficial because the sequential application of usual CUSUM tests as in Sibbertsen and Kruse (2009) with constant critical values will eventually reject a correct null of no change in persistence with probability one (see Robbins (1970)).

Generally, every strictly increasing function $b(t) = zq(t)$ could serve as a boundary function where z is some suitable scaling factor and $q(t)$ is some monotonically increasing function in time. However if the boundary function grows too slowly the monitoring test will commit the type one error almost surely as it will detect a break in persistence with probability one. On the contrary if the boundary grows too quickly the test will loose power because a break in persistence cannot be detected anymore. For the short memory case a variety of different boundary functions have been proposed (see Andreou and Ghysels (2006, p. 92) for an overview). In particular Altissimo and Corradi (2003) derive a boundary function based on the almost sure asymptotically uniform equicontinuity of the Brownian bridge obtaining an almost sure boundary function. This is convenient because it gives the rate of convergence with which the sequence of functions converges to a relatively compact set in the sense of an Arzelà-Ascoli theorem (see e.g. Davidson (1994, p. 335)). This provides useful information as we are interested in the behavior of the limiting distribution independently of the test statistic. We also derive almost sure results similar to the ones obtained by Altissimo and Corradi (2003) which are collected in the next theorem.

Theorem 5.2.
Let $BB^0(t,d) = B(t,d) - tB(1,d)$ be a fractional Brownian bridge. Then, $d_T^{-1}|BB^0(t,d)|$ is almost surely asymptotically uniform equicontinuous in $t \in [0,1]$. With $d_T := \sqrt{2T^{2d+1}\log\log(T)}$.

The use of this theorem is that it provides the rate with which the increment of the fractional brownian bridge becomes asymptotically uniform equicontinuous. In the proof this derived to be $\sqrt{2\log\log(T)}$. Hence, if we use this growth rate for the boundary function we will obtain a slowly growing function and therefore detect a change in persistence but at the same time the growth rate of this function is independent of the long memory parameter under the null d_0.

Different forms of the boundary function are possible. For example one could use the boundary function

$$b_1(t) = z\sqrt{2t\log_2(t)}\,, \tag{5.5}$$

where $\log_2(t) := \log(\log(t))$. This boundary function is based on the law of iterated logarithm and is motivated by the fastest detection of change because it grows as slowly as possible. From theorem 5.2 we deduce the boundary function

$$b_2(t) = z\sqrt{2\log_2(t)}\,. \tag{5.6}$$

Because both boundary functions rely on the square root of a logarithm one needs to find a way to deal with values $\leq \log(1)$ to ensure real valued boundaries. One way of doing so is to define

$$\log_2'(t) := \begin{cases} 1, & \text{if } t \leq \exp(1) \\ \log\log(t), & \text{if } t > \exp(1)\,, \end{cases}$$

similar to Leisch et al. (2000). Another way which avoids the constant behavior of the boundary function at the beginning of the monitoring period is to define

$$\log_2''(t) := \begin{cases} t, & \text{if } t \leq \exp(1) \\ \log\log(t), & \text{if } t > \exp(1)\,. \end{cases}$$

Formally this leads to four possible boundary functions

$$b_3(t) = z\sqrt{2t\log_2'(t)} \tag{5.7}$$

$$b_4(t) = z\sqrt{2t\log_2''(t)} \tag{5.8}$$

$$b_5(t) = z\sqrt{2\log_2'(t)} \tag{5.9}$$

$$b_6(t) = z\sqrt{2\log_2''(t)}\,. \tag{5.10}$$

One could think of different boundary functions such as functions that are dependent on the long memory parameter under the null to account for the gradually increasing variance of the process. However, unreported simulations showed that such a boundary function does not perform satisfactorily and we therefore restrict ourselves to the above boundary functions.

5.3 Monte Carlo evidence

We start by providing some Monte Carlo evidence on the small sample behavior of the usual MOSUM test as considered in Leisch et al. (2000) under long range dependence. Table 5.1 shows some of the simulation results.

	$\tau = 4$			$\tau = 6$			$\tau = 8$		
d	$h = 0.25$	$h = 0.5$	$h = 1$	$h = 0.25$	$h = 0.5$	$h = 1$	$h = 0.25$	$h = 0.5$	$h = 1$
0.1	66.22	57.26	51.08	70.04	62.80	53.72	72.52	67.12	57.02
0.2	98.06	96.38	91.90	99.28	97.94	94.90	99.76	98.98	96.86
0.3	99.96	99.90	99.64	100.00	100.00	99.96	100.00	100.00	99.96
0.4	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table 5.1: Empirical size of the fluctuation test by Leisch et al. (2000) [in %] for $T = 250$ and $\alpha = 5\%$.

As expected the generalized fluctuation test does not keep its size. Even if the long memory is only moderately present the test does not allow a secure conclusion whether a change in persistence is present or not because the boundary functions are too narrow.

In order to assess the finite sample performance of the monitoring procedure described in section 5.2 we consider different values for the long memory parameter $d = 0.1, 0.2, 0.3, 0.4$, the monitoring window $h = 0.25, 0.5, 0.75, 1$ and the out-of-sample monitoring period $\tau = 2, 4, 6, 8, 10$. We also consider different sample sizes of $T = 200, 250, 300$ and the different boundary functions $b_i(t)$, for $i = 3, \ldots, 6$, from (5.7) to (5.10) for the simulations. The number of Monte Carlo repetitions is set to $M = 10000$ and the levels of significance are set to $\alpha = 1\%, 5\%, 10\%$.[1]

Boundary function $b_3(t)$									
	$\tau = 4$			$\tau = 6$			$\tau = 8$		
d	$h = 0.5$	$h = 0.75$	$h = 1$	$h = 0.5$	$h = 0.75$	$h = 1$	$h = 0.5$	$h = 0.75$	$h = 1$
0.1	6.86	7.10	8.59	6.26	6.95	9.03	7.07	6.97	8.60
0.2	6.77	6.80	8.73	6.47	6.56	8.52	5.75	6.82	8.97
0.3	5.68	7.12	9.76	6.16	7.08	9.54	6.21	7.08	10.24
0.4	10.77	12.29	16.43	9.96	12.82	16.18	10.07	12.24	16.09

Table 5.2: Empirical size of the monitoring procedure [in %] for $T = 250$ and $\alpha = 5\%$.

Table 5.2 shows the size results for the boundary function motivated by the law of iterated logarithm. Using this boundary we obtain a procedure that is generally oversized. This overrejection of the correct null becomes more severe as the degree of persistence increases and/or the monitoring window h increases.

Table 5.3 displays the respective results based on the almost sure results from theorem 5.2. These results are more promising compared to the ones of boundary $b_3(t)$ as the nominal size level is better adhered to. Looking at the dependencies between the size, the long memory parameter d, the monitoring window h and the monitoring period τ we see that a moderate window size of $h = 0.5$ or $h = 0.75$ is generally preferable regardless of the monitoring period τ. If the persistence increases a reduced window size of $h = 0.5$ yields the most accurate size results. Reducing the window size even further to $h = 0.25$, however, leads to overrejection again as unreported results show.

As the boundary function $b_6(t)$ is only a slight modification of boundary function $b_5(t)$ the same argument as above applies to the results in table 5.4. The only difference is that the test overrejects somewhat when using boundary function $b_6(t)$.

[1]Some of the results here and in the sequel are unreported to save space but can be obtained from the authors on request.

Boundary function $b_5(t)$									
	$\tau = 4$			$\tau = 6$			$\tau = 8$		
d	$h = 0.5$	$h = 0.75$	$h = 1$	$h = 0.5$	$h = 0.75$	$h = 1$	$h = 0.5$	$h = 0.75$	$h = 1$
0.1	7.55	7.02	7.09	7.12	6.16	6.69	7.30	6.56	6.29
0.2	6.76	5.85	6.56	6.64	5.85	5.60	5.90	5.66	5.23
0.3	5.15	4.87	4.87	5.08	4.28	4.29	5.16	3.89	3.92
0.4	5.61	5.10	5.78	4.86	4.12	4.28	4.32	3.63	3.89

Table 5.3: Empirical size of the monitoring procedure [in %] for $T = 250$ and $\alpha = 5\%$.

Boundary function $b_6(t)$									
	$\tau = 4$			$\tau = 6$			$\tau = 8$		
d	$h = 0.5$	$h = 0.75$	$h = 1$	$h = 0.5$	$h = 0.75$	$h = 1$	$h = 0.5$	$h = 0.75$	$h = 1$
0.1	7.00	7.52	8.65	6.75	6.71	8.42	7.34	7.05	7.59
0.2	7.13	6.92	8.60	6.69	6.31	7.33	6.02	6.20	6.70
0.3	6.12	6.96	8.94	5.96	5.86	7.26	5.84	5.58	6.49
0.4	10.26	11.22	13.96	8.39	9.61	10.83	7.67	8.06	9.53

Table 5.4: Empirical size of the monitoring procedure [in %] for $T = 250$ and $\alpha = 5\%$.

The size results for the $\alpha = 10\%$ level are unreported but show the same general behavior of the previously discussed results. However, in this setting it becomes even more obvious that the boundary function $b_5(t)$ yields the best performance over all considered settings. Generally the size distortions are minor and acceptable and also comparable to the short memory case as reported in Leisch et al. (2000).

In an empirical setting the long memory parameter d_0 is unknown and has to be estimated. We therefore conduct the size experiment again but this time using an estimated d_0. Generally every consistent estimation method is applicable but estimators that converge faster than the asymptotic distribution to the true value of d_0 are preferable. One such estimator is the approximate maximum likelihood estimator proposed by Beran (1995) which is $\sqrt{T}$ consistent. Another popular method to estimate d_0 is the log-periodogram regression (see Geweke and Porter-Hudak (1983)). The rate of convergence of this estimator is $\sqrt{m}$ where m is the number of frequencies used. The estimator is consistent as long as $(m\log(m))/n \to 0$ as $m,n \to \infty$, with n being the sample size (see Hurvich et al. (1998)). In our simulations we use this estimator with $T^{4/5}$ frequencies. The results are reported for the $\alpha = 5\%$ level in table 5.5.

Boundary function $b_5(t)$									
	$\tau = 4$			$\tau = 6$			$\tau = 8$		
d	$h = 0.5$	$h = 0.75$	$h = 1$	$h = 0.5$	$h = 0.75$	$h = 1$	$h = 0.5$	$h = 0.75$	$h = 1$
0.1	8.68	7.62	8.50	8.62	7.14	7.48	7.72	7.20	6.42
0.2	7.12	6.62	6.02	7.44	6.08	5.82	6.36	5.86	5.42
0.3	5.74	4.98	5.26	5.14	4.28	4.40	4.26	4.24	4.04
0.4	5.36	4.86	5.38	4.32	4.36	4.34	3.74	3.42	4.04

Table 5.5: Empirical size of the monitoring procedure with estimated d_0.

We observe small size distortions for smaller values of d_0 and larger monitoring periods but generally the size is well kept even if we estimate the long memory parameter.
When the persistence changes from stationary to non-stationary the MOSUM test will eventually detect this with probability one due to consistency (see theorem 5.1).[2] Therefore it is more interesting how fast a change in persistence can be detected.
To study the detection delay we consider breaks from the stationary region, namely $d_0 = 0.1, 0.2, 0.3, 0.4$, to the non-stationary region, $d_1 = 0.6, 0.7, 0.8, 0.9, 1$. The break occurs within the monitoring period at $t^* = [\rho\tau T]$, where $\rho = 0.3, 0.5, 0.7$ and $\tau = 2, 4, 6, 8, 10$ as above and $[\cdot]$ denotes the integer part of its argument. We use a sample size of $T = 250$ and the boundary functions $b_i(t)$, for $i = 3, \ldots, 6$, from (5.7) to (5.10). As an example the average detection delay for the $\alpha = 5\%$ level for the boundary function $b_5(t)$ for different breaks is displayed in tables 5.6, 5.7 and 5.8.

[2]This has also been confirmed in unreported simulations.

Boundary function $b_5(t)$									
$\tau = 2$	$h = 0.25$			$h = 0.5$			$h = 0.75$		
d_0	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$
0.1	53.33	43.40	37.23	77.61	63.15	49.58	91.34	72.67	56.20
0.2	79.23	63.41	54.45	112.58	95.82	81.91	126.86	112.28	95.50
0.3	112.44	92.12	76.31	144.79	137.08	122.53	158.95	156.00	148.96
0.4	133.96	118.80	99.73	150.59	158.25	151.10	157.33	172.92	172.75
$\tau = 4$	$h = 0.25$			$h = 0.5$			$h = 0.75$		
d_0	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$
0.1	57.89	46.73	38.50	91.25	71.51	57.13	116.31	90.09	69.96
0.2	94.31	72.12	59.49	149.92	118.26	98.78	186.03	152.03	127.12
0.3	160.65	109.40	86.93	222.47	183.72	152.57	265.43	240.25	210.58
0.4	225.64	167.07	121.51	272.32	258.26	214.56	293.90	309.68	282.91
$\tau = 6$	$h = 0.25$			$h = 0.5$			$h = 0.75$		
d_0	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$
0.1	53.52	41.29	32.36	96.13	71.78	56.47	124.22	94.77	72.26
0.2	98.55	69.55	55.80	167.16	124.39	104.44	217.05	169.95	138.28
0.3	190.16	111.91	85.96	282.65	204.74	165.30	339.06	276.86	231.03
0.4	315.36	192.13	126.20	380.04	315.26	239.72	419.28	396.05	320.28

Table 5.6: Average detection delay of the monitoring procedure for $T = 250$, $\alpha = 5\%$ and $\rho = 0.3$.

Table 5.6 shows the results for the case of an early break within the monitoring period. As one expects the detection is easier and therefore faster if the difference between d_0 and d_1 is large. Consequently the detection delay is rather small if the persistence changes from stationary, say $d_0 = 0.2$, long memory to non-stationary, say $d_1 = 0.8$, and even faster if the process becomes a unit root process after the break. In fact, the detection delay for larger breaks is comparable with the short memory case (see table 3 in Leisch et al. (2000)). This is encouraging given the well known slow rate of convergence in long memory time series. Another result is that it is easier and faster to detect a change in persistence if the width of the monitoring window $[Th]$ is rather small. Detection delays for values of $h = 0.25$ and $h = 0.5$ are generally smaller compared to larger values of h. This is also in line with the findings of Leisch et al. (2000) for the short memory case. It is well known also in related areas of the structural change literature (see e.g Pesaran and Zimmermann (2005) for results regarding forecasts under structural breaks) that smaller windows of data are usually better to detect and deal with structural change. The results for later breaks within the monitoring period are shown in tables 5.7 and 5.8. The general conclusions from above remain valid but the detection delay becomes even smaller if the breaks occurs later. This is also a similar behavior to the short memory case reported in Leisch et al. (2000).

Boundary function $b_5(t)$									
$\tau = 2$	$h = 0.25$			$h = 0.5$			$h = 0.75$		
d_0	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$
0.1	44.55	35.39	28.67	65.85	54.45	43.62	74.23	60.25	48.20
0.2	65.49	55.62	47.31	89.16	82.66	72.46	96.55	91.89	81.20
0.3	85.56	78.49	68.26	110.59	112.02	106.93	118.37	123.46	118.71
0.4	89.62	92.35	84.78	99.62	115.70	117.50	95.06	110.72	125.06
$\tau = 4$	$h = 0.25$			$h = 0.5$			$h = 0.75$		
d_0	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$
0.1	51.01	35.76	29.86	88.67	68.46	52.90	113.94	87.37	67.87
0.2	88.44	64.41	51.54	141.47	116.89	98.79	172.76	150.65	129.69
0.3	143.60	108.29	82.25	190.48	174.08	152.71	223.04	217.80	204.48
0.4	169.54	153.31	114.90	204.31	218.35	201.99	225.85	250.99	241.94
$\tau = 6$	$h = 0.25$			$h = 0.5$			$h = 0.75$		
d_0	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$
0.1	36.59	24.66	16.15	87.80	65.26	47.83	122.35	90.52	67.80
0.2	86.31	58.32	41.59	158.73	120.80	99.18	204.27	169.67	140.36
0.3	174.06	106.13	74.87	242.96	201.27	163.20	289.19	264.62	229.60
0.4	235.73	181.23	121.38	287.28	279.41	231.81	303.02	332.77	302.43

Table 5.7: Average detection delay of the monitoring procedure for $T = 250$, $\alpha = 5\%$ and $\rho = 0.5$.

Boundary function $b_5(t)$									
$\tau = 2$	$h = 0.25$			$h = 0.5$			$h = 0.75$		
d_0	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$
0.1	28.02	24.45	18.84	39.55	35.59	30.44	41.95	38.81	31.69
0.2	39.24	38.55	33.06	53.02	52.25	50.44	54.65	53.90	52.27
0.3	45.85	53.02	50.46	61.29	69.01	69.73	50.52	60.77	69.68
0.4	27.52	44.96	50.19	21.71	38.06	52.93	-77.72	-86.60	-61.95
$\tau = 4$	$h = 0.25$			$h = 0.5$			$h = 0.75$		
d_0	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$
0.1	38.81	24.41	16.72	69.65	58.49	44.95	86.56	79.32	63.26
0.2	66.51	51.73	40.02	101.49	101.71	92.10	121.38	117.65	112.30
0.3	90.24	82.78	69.57	124.50	133.20	131.38	146.83	156.16	159.54
0.4	77.71	100.90	92.53	96.81	138.45	144.58	93.48	140.91	166.83
$\tau = 6$	$h = 0.25$			$h = 0.5$			$h = 0.75$		
d_0	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$
0.1	12.08	4.92	-5.71	68.34	50.98	33.94	99.05	76.97	55.14
0.2	63.08	37.87	20.33	120.52	103.78	87.11	155.37	142.54	128.07
0.3	104.40	85.53	57.51	163.12	161.03	145.76	188.54	197.29	196.56
0.4	110.33	132.11	96.32	139.56	185.50	185.40	148.29	215.15	222.53

Table 5.8: Average detection delay of the monitoring procedure for $T = 250$, $\alpha = 5\%$ and $\rho = 0.7$.

5.4 Empirical Application

To illustrate the use of the monitoring approach we analyze monthly US price inflation series from Stock and Watson (2005).[3] In particular we consider the first difference of the logarithmic implied price deflator for durable goods. This series has also been under investigation from Cavaliere and Taylor (2008) who report a change in persistence from $I(0)$ to $I(1)$. However, they did not consider the possibility of fractional integration in the series although inflation related time series are likely to show long memory behavior (see e.g. Hassler and Wolters (1995)). The sample spans from 01/1959 to 12/2003. The series is depicted in figure 5.1.

[3]The data is available at Mark Watson's website at: http://www.princeton.edu/~mwatson/wp.html.

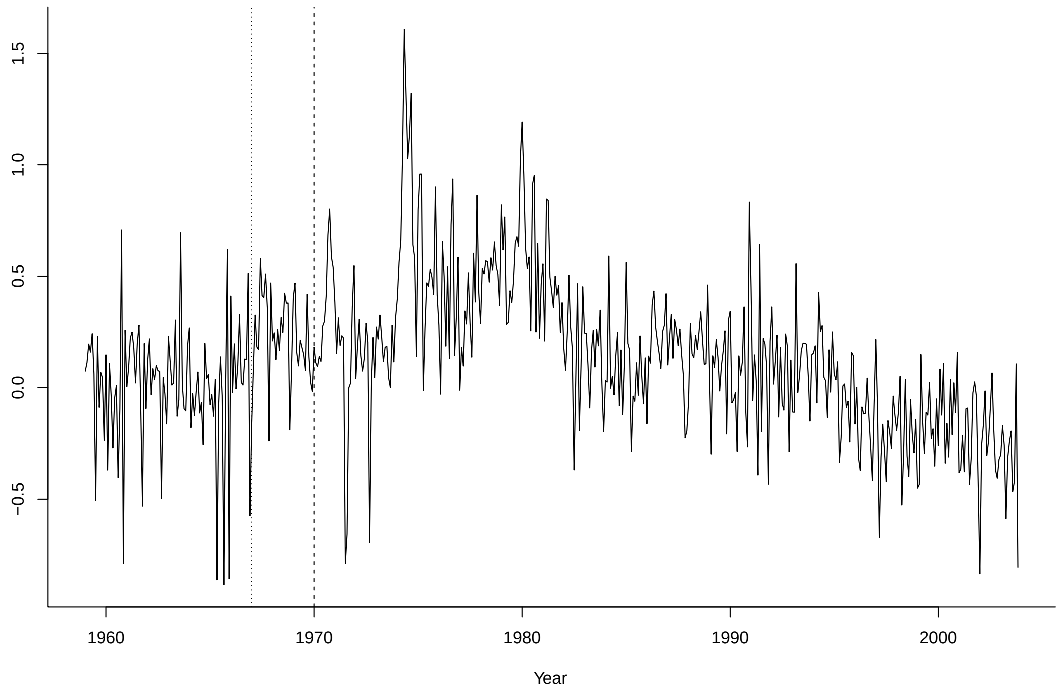


Figure 5.1: First difference of logarithmic price deflator for durable goods.

To determine the value of the long memory parameter we use log-periodogram regression as proposed by Geweke and Porter-Hudak (1983). The decision of how many frequencies should be used in the regression is a trade-off between reducing the bias and reducing the asymptotic variance. We use $T^{1/2}$ frequencies to deal with potential short memory components in the data (see e.g. Agiakloglou et al. (1993)). For the whole sample this yields an estimate of $\hat{d} = 0.61$. This value is highly significant as judged by its p-value which is $< 1e-03$.

To test whether a change in persistence can be detected in the data we apply the CUSUM of squares test for a change in persistence proposed by Sibbertsen and Kruse (2009) to the whole sample. This leads to a test statistic of $R = 0.0373$ which is significant at the $\alpha = 5\%$ level in favor of an increasing persistence. The estimated breakpoint is at $t^* = 107$ which is 11/1967 (the dotted line in figure 5.1).

To use the monitoring approach we split the sample in an in-sample part ranging from 01/1959 to 12/1965 and leave the rest as monitoring period. This yields a $\tau \approx 5$. The estimated d_0 within the in-sample period is $\hat{d}_0 = 0.23$.

For the application of the MOSUM test we use the boundary function $b_5(t)$ and set $h = 0.5$. The first time the sequence of test statistics exceeds the $\alpha = 1\%$ boundary function is at $t = 55$ in the monitoring period. This is equivalent to an estimated breakpoint at $t^* = 139$ which is 06/1970 (the dashed line in figure 5.1). The first time the sequence of test statistics exceeds the $\alpha = 5\%$ and $\alpha = 10\%$ boundary functions is only one period earlier.

The estimation of d_1 in the monitoring period yields $\hat{d}_1 = 0.68$. Thus we can confirm a change in persistence with high probability from stationary long memory to non-stationary long memory.

Notably the detection delay is rather short and we obtain a fast indication of the change in persistence from using the monitoring procedure.

5.5 Conclusion

Detecting a change in persistence as soon as possible is of paramount interest because structural change affects the subsequent analysis of the data heavily. The usual approach is to use one-shot tests to detect a change in persistence a posteriori. However, these tests cannot be applied sequentially because a correct null of no change would eventually be rejected with probability one. We propose a monitoring procedure based on moving sums that allows to detect a change in the long memory parameter of a long range dependent time series whenever new data arrives. By means of a Monte Carlo experiment we show good size properties and also study the detection delay when a change in persistence occurs. Depending on the width of the monitoring window and the difference between the pre- and post-break long memory parameter the detection is rather fast. Smaller monitoring windows generally prove more useful to detect a change in persistence early and also larger differences between the long memory parameters are detected faster.

In an empirical illustration of the method we are able to confirm a change in persistence from stationary to non-stationary long memory in an inflation time series.

5.6 Appendix

5.6.1 Proof of Theorem 5.1

First, let $k = [Tt]$ for each value in the monitoring period then write the test statistic as

$$\begin{aligned} MS_{T,h,d} &= \max_{T+1\le k\le [T\tau]} \sigma^{-1} T^{-\frac{1}{2}-d} \left| \sum_{i=k-[Th]+1}^{k} \hat{e}_i - \frac{[Th]}{T} \sum_{i=1}^{T} \hat{e}_i \right| \\ &= \max_{T+1\le [Tt]\le [T\tau]} \sigma^{-1} T^{-\frac{1}{2}-d} \left| \sum_{i=1}^{[Tt]} \hat{e}_i - \frac{[Tt]}{T} \sum_{i=1}^{T} \hat{e}_i - \sum_{i=1}^{[Tt]-[Th]} \hat{e}_i + \frac{[Tt]-[Th]}{T} \sum_{i=1}^{T} \hat{e}_i \right| . \end{aligned}$$

Then using the FCLT for fractionally integrated processes (see Sowell (1990) and Davidson and de Jong (2000)) and the continuous mapping theorem (CMT) we have

$$\begin{aligned} MS_{T,h,d} &\Rightarrow \max_{T+1\le [Tt]\le [T\tau]} \sigma^{-1} \left| B(t,d) - tB(1,d) - B(t-h,d) + (t-h)B(1,d) \right| \\ &= \max_{T+1\le [Tt]\le [T\tau]} \sigma^{-1} \left| BB^0(t,d) - [B(t-h,d) - (t-h)B(1,d)] \right| \\ &= \max_{T+1\le [Tt]\le [T\tau]} \sigma^{-1} \left| BB^0(t,d) - BB^0(t-h,d) \right| , \end{aligned}$$

where $BB^0(t,d)$ denotes a fractional Brownian bridge.

To prove consistency we consider that at some point in the monitoring period, say k^*, the persistence changes from stationary long memory with $0 < d_0 < \frac{1}{2}$ to non-stationary long memory with $\frac{1}{2} < d_1 < \frac{3}{2}$ and then split the test statistic into its stationary and non-stationary parts. We write the test statistic as

$$\begin{aligned} MS_{T,h,d_0} &= \max_{T+1\le k\le[T\tau]} \sigma^{-1}T^{-\frac{1}{2}-d_0}\left|\sum_{i=k-[Th]+1}^{k}\hat{e}_i - \frac{[Th]}{T}\sum_{i=1}^{T}\hat{e}_i\right| \\ &= \max_{T+1\le[rT]\le[T\tau]} \sigma^{-1}T^{-\frac{1}{2}-d_0}\left|\underbrace{\sum_{i=1}^{[rT]}\hat{e}_i}_{I} - \underbrace{\sum_{i=1}^{[rT]-[hT]}\hat{e}_i}_{II} - \frac{[Th]}{T}\underbrace{\sum_{i=1}^{T}\hat{e}_i}_{III}\right|, \end{aligned}$$

where $k = [rT]$ for some $r > 1$. Part III only contains $I(d_0)$ variables due to the noncontamination assumption.

We have to distinguish two cases:

(i) $k^* \le [rT]-[Th] \Rightarrow$ in this case both I and II contain $I(d_1)$ variables

(ii) $[rT]-[Th] \le k^* \le [rT] \Rightarrow$ in this case only I contains $I(d_1)$ variables.

Ad (i):

The case (i) is depicted in figure 5.2 where $[rT]$ is denoted by k_1 and $[rT]-[Th]$ is denoted by k_0. The gray shaded area is the monitoring window.

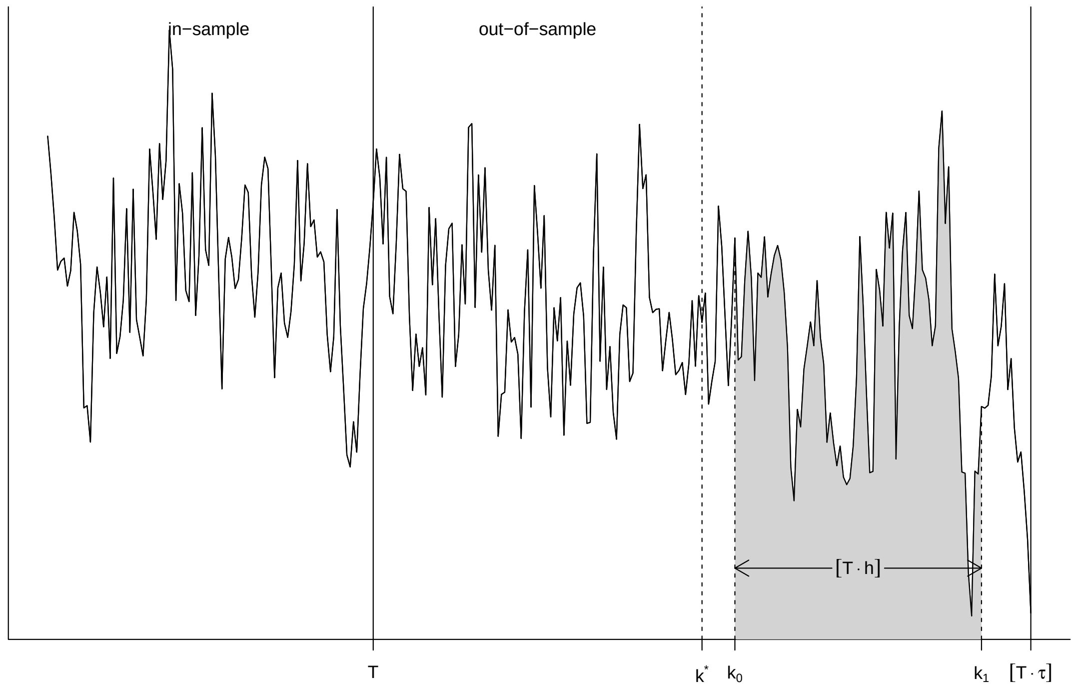


Figure 5.2: MOSUM case (i).

Write the test statistic as

$$\begin{aligned} MS_{T,h,d_0} &= \max_{T+1\le[rT]\le[T\tau]} \sigma^{-1}T^{-\frac{1}{2}-d_0}\left|\sum_{i=1}^{k^*}\hat{e}_i + \sum_{i=k^*+1}^{[rT]}\hat{e}_i - \sum_{i=1}^{k^*}\hat{e}_i - \sum_{i=k*+1}^{[rT]-[hT]}\hat{e}_i - \frac{[hT]}{T}\sum_{i=1}^{T}\hat{e}_i\right| \\ &= \max_{T+1\le[rT]\le[T\tau]} \sigma^{-1}T^{-\frac{1}{2}-d_0}\left|-\frac{[hT]}{T}\sum_{i=1}^{T}\hat{e}_i\right| + \max_{T+1\le[rT]\le[T\tau]} \sigma^{-1}T^{-\frac{1}{2}-d_0}\left|\sum_{i=[rT]-[hT]}^{[rT]}\hat{e}_i\right| . \end{aligned}$$

Now, the first part is $I(d_0)$ and is correctly standardized. Therefore, using the arguments from above it converges to $-hB(1, d_0)$ which is the standard deviation of the fractional Brownian motion. For the second part the standardization is obtained from d_0 but the variables are $I(d_1)$ and so the expression diverges and we obtain

$$MS_{T,h,d_0} = o_p(1) + O_p\left(T^{d_1-d_0}\right) . \tag{5.11}$$

Ad (ii):
The situation (ii) is depicted in figure 5.3.

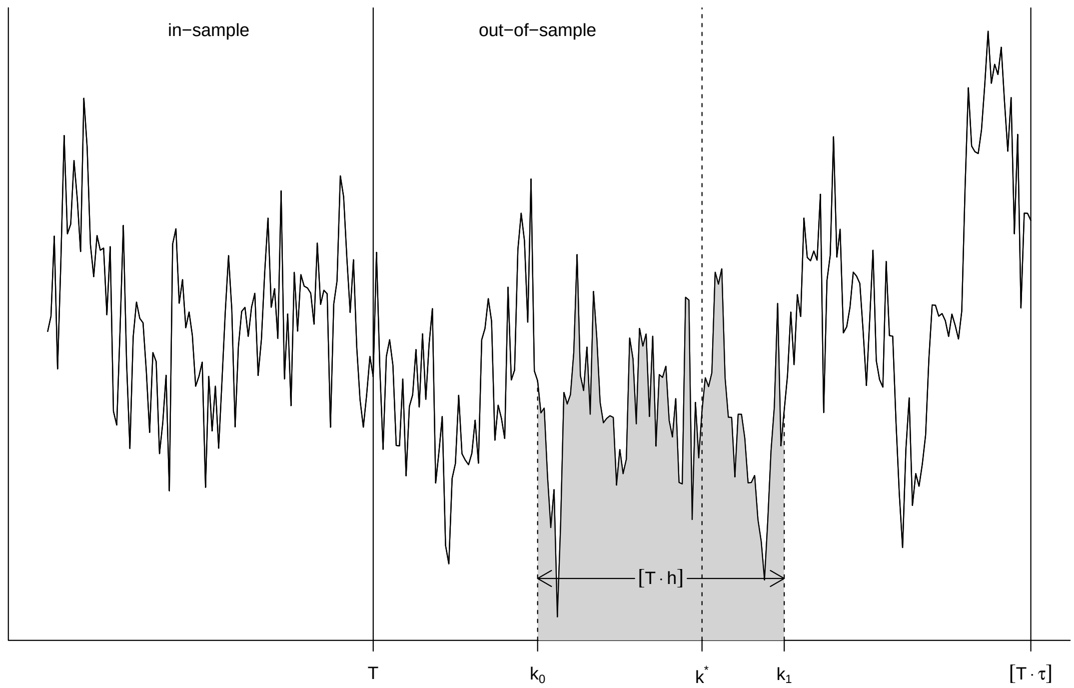


Figure 5.3: MOSUM case (ii).

Now only I contains $I(d_1)$ variables. Write the test statistic as

$$\begin{aligned} MS_{T,h,d_0} &= \max_{T+1\leq[rT]\leq[T\tau]} \sigma^{-1}T^{-\frac{1}{2}-d_0}\left|\sum_{i=1}^{[rT]}\hat{e}_i - \sum_{i=1}^{[rT]-[hT]}\hat{e}_i - \frac{[hT]}{T}\sum_{i=1}^{T}\hat{e}_i\right| \\ &= \max_{T+1\leq[rT]\leq[T\tau]} \sigma^{-1}T^{-\frac{1}{2}-d_0}\left|\sum_{i=1}^{k^*}\hat{e}_i + \sum_{i=k^*+1}^{[rT]}\hat{e}_i - \sum_{i=1}^{[rT]-[hT]}\hat{e}_i - \frac{[hT]}{T}\sum_{i=1}^{T}\hat{e}_i\right| \\ &= \max_{T+1\leq[rT]\leq[T\tau]} \sigma^{-1}T^{-\frac{1}{2}-d_0}\left|\sum_{i=k^*+1}^{[rT]}\hat{e}_i\right| + \max_{T+1\leq[rT]\leq[T\tau]} \sigma^{-1}T^{-\frac{1}{2}-d_0}\left|\sum_{i=[rT]-[hT]}^{k^*}\hat{e}_i\right| \\ &\quad - \max_{T+1\leq[rT]\leq[T\tau]} \sigma^{-1}T^{-\frac{1}{2}-d_0}\left|\frac{[hT]}{T}\sum_{i=1}^{T}\hat{e}_i\right| . \end{aligned}$$

With the arguments from case (i) we obtain

$$MS_{T,h,d_0} = O_p\left(T^{d_1-d_0}\right) + o_p(1) + o_p(1) , \tag{5.12}$$

where the second part of the above expression does not expand with T anymore and therefore vanishes as $T \to \infty$. □

5.6.2 Proof of Theorem 5.2

Denote by $d_T := \sqrt{2T^{2d+1}\log\log(T)}$. By the reverse triangle inequality we have for some $r \in [0,1]$

$$d_T^{-1}\left|B(Tr,d) - rB(T,d) - (B(Tr',d) - r'B(T,d))\right| \leq d_T^{-1}\left|B(Tr,d) - B(Tr',d)\right| + d_T^{-1}\left|(r-r')B(T,d)\right| ,$$

for distinct values r and r'. Using the notation from Altissimo and Corradi (2003, p. 232) we write $S(r,\delta) = (r' : |r-r'| \leq \delta)$. Now, by the fact that (see Davidson (1994, p. 335ff.))

$$\sup_{\theta\in\Theta}\sup_{\theta'\in S(\theta,\delta)}\left|f_n(\theta') - f_n(\theta)\right| \leq 2\sup_{\theta\in\Theta}|f_n(\theta)|$$

and the LIL for the fractional Brownian motion (see e.g. Taqqu (1977)) we have for the second part of the right side

$$\limsup_{T\to\infty}\sup_{r\in[0,1]}\sup_{r'\in S(r,\delta)} d_T^{-1}\left|(r-r')B(T,d)\right| \leq 2\delta\sigma ,$$

with σ the variance of the fractional Brownian Motion. As $\delta \to 0$ the whole part approaches zero which ensures the asymptotic uniform equicontinuity almost surely.
For the first part of the right hand side we have by self-similarity

$$\begin{aligned} \limsup_{T\to\infty}\sup_{r\in[0,1]}\sup_{r'\in S(r,\delta)} d_T^{-1}\left|B(Tr) - B(Tr')\right| &= \limsup_{T\to\infty}\sup_{r\in[0,1]}\sup_{r'\in S(r,\delta)} d_T^{-1}\left|T^{d+1/2}B(r) - T^{d+1/2}B(r')\right| \\ &= \limsup_{T\to\infty}\sup_{r\in[0,1]}\sup_{r'\in S(r,\delta)} T^{d+1/2}d_T^{-1}\left|B(r) - B(r')\right| . \end{aligned}$$

Now note that

$$d_T = \sqrt{2T^{2d+1}\log\log(T)} = \sqrt{T^{2d+1}}\sqrt{2\log\log(T)} = T^{d+1/2}\sqrt{2\log\log(T)}\,.$$

Therefore we obtain

$$\limsup_{T\to\infty}(2\log\log(T))^{-\frac{1}{2}}\sup_{r\in[0,1]}\sup_{r'\in S(r,\delta)}\left|B(r)-B(r')\right|\,.$$

Because $|B(r)-B(r')|$ is almost surely Hölder continuous of order strictly less than H (see Biagini et al. (2008, p. 11)) and $\limsup_{T\to\infty}(2\log\log(T))^{-\frac{1}{2}}$ tends to zero as $T\to\infty$ it follows that the above expression is almost surely asymptotically uniform equicontinuous. □

References

Agiakloglou, C., Newbold, P., Wohar, M. (1993): Bias of an estimator of the fractional difference parameter. *Journal of Time Series Analysis* 14, 235 – 246.

Altissimo, F., Corradi, V. (2003): Strong rules for detecting the number of breaks in a time series. *Journal of Econometrics* 117, 207 – 244.

Andersen, T.G., Bollerslev, T. (1997): Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance* 4, 115 – 158.

Andreou, E., Ghysels, E. (2006): Monitoring disruptions in financial markets. *Journal of Econometrics* 135, 77 – 124.

Andrews, D.W.K. (1993): Tests for parameter instability and structural change with unknown change point. *Econometrica* 61(4), 821 – 856.

Andrews, D.W.K., Fair, R.C. (1988): Inference in nonlinear econometric models with structural change. *The Review of Economic Studies* 55, 615 – 639.

Bai, J., Perron, P. (1998): Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica* 66, 47 – 78.

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

Bai, J., Perron, P. (2004): Multiple Structural Change Models: A Simulation Analysis. In: *Econometric theory and practice: frontiers of analysis and applied research by Corbae, D., Durlauf, S.N. and Hansen, B.E.*, Cambridge University Press, 212 – 237.

Bai, J., Perron, P. (2006): Econometric Theory and Practice: Frontiers of Analysis and Applied Research. In: *Theory and Applications of Long-Range Dependence.* Cambridge University Press.

Baillie, R.T. (1996): Long memory processes and fractional integration in econometrics. *Journal of Econometrics* 73, 5 – 59.

Banerjee, A., Lumsdaine, R., Stock, J. (1992): Recursive and sequential tests of the unit root and trend break hypothesis: Theory and international evidence. *Journal of Business & Economic Statistics* 10, 271 – 288.

Banerjee, A., Urga, G. (2005): Modelling structural breaks, long memory and stock market volatility: an overview. *Journal of Econometrics* 129, 1 – 34.

Bauer, P., Hackl, P. (1978): The use of MOSUMs for quality control. *Technometrics* 20, 431 – 436.

Belaire-Franch, J. (2005): A proof of the power of Kim's test against stationary processes with structural breaks. *Econometric Theory* 21, 1172 – 1176.

Beran, J. (1995): Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *Journal of the Royal Statistical Society, Series B* 57, 659 – 672.

Beran, J. Bhansali, R.J., Ocker, D. (1998): On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes. *Biometrika* 85(4), 921 – 934.

Beran, J., Terrin, N. (1996): Testing for a change of the long-memory parameter. *Biometrika* 83, 627 – 638.

Bhardwaj, G., Swanson, N.R. (2006): An empirical investigation of the usefulness of ARFIMA models for predicting macroeconomic and financial time series. *Journal of Econometrics* 131, 539 – 578.

Biagini, F., Hu, Y., Øksendahl, B., Zhang, T. (2008): Stochastic calculus for fractional brownian motion and applications. *Springer, New York.*

Breidt, F.J., Crato, N., Lima, P. de (1998): The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics* 83, 325 – 348.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. (1993): Classification and Regression Trees. *Chapman & Hall, New York.*

Brown, R.L., Durbin, J., Evans, J.M. (1975): Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society B* 37, 149 – 163.

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

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

Cappelli, C., Reale, M. (2005): Detecting Changes in Mean Levels with Atheoretical Regression Trees. *Research Report UCMSD* 2005/2, Department of Mathematics and Statistics, University of Canterbury.

Cavaliere, G., Taylor, A.M.R. (2008): Testing for a change in persistence in the presence of non-stationary volatility. *Journal of Econometrics* 147, 84 – 98.

Choi, K., Zivot, E. (2007): Long memory and structural changes in the forward discount: an empirical investigation. *Journal of International Money and Finance* 26(3), 342 – 363.

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

Christensen, B.J., Nielsen, M. (2007): The effects of long memory in volatility on stock market fluctuations. *The Review of Economics and Statistics* 89(4), 684 – 700.

Chu, C., Hornik, K., Kuan, C. (1995): MOSUM tests for parameter constancy. *Biometrika* 82, 603 – 617.

Chu, C., Stinchcombe, M., White, H. (1996): Monitoring structural change. *Econometrica* 64, 1045 – 1065.

Clements, M.P., Hendry, D.F. (1998): Forecasting economic time series. *Cambridge University Press.*

Cooper, S.J. (1998): Multiple Regimes in US Output Fluctuations. *Journal of Business and Economic Statistics* 28(3), 92 – 100.

Corsi, F., Mittnik, S., Pigorsch, C., Pigorsch, U. (2008): The volatility of realized volatility. *Econometric Reviews* 27, 46 – 78.

Corvoisier, S., Mojon, B. (2005): Breaks in the Mean of Inflation: How they happen and what to do with them. *ECB Working Paper* No. 451.

da Rosa, J.C., Veiga, A., Medeiros, M.C. (2008): Tree-Structured Smooth Transition Regression Models Based on CART Algorithm. *Journal of Computational Statistics and Data Analysis* 52, 2469 – 2488.

Davidson, J. (1994): Stochastic Limit Theorem. *Oxford University Press.*

Davidson, J., de Jong, R. (2000): The functional central limit theorem and weak convergence to stochastic integrals II: Fractional integrated processes. *Econometric Theory* 16, 643 – 666.

De'ath, G., Fabricius, K. (2000): Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* 81(11), 3178 – 3192.

Deo, R., Hurvich, C., Lu, Y. (2006): Forecasting realized volatility using a long-memory stochastic volatility model: estimation, prediction and seasonal adjustment. *Journal of Econometrics* 131, 29 – 58.

Diebold, F.X., Inoue, A. (2001): Long memory and regime switching. *Journal of Econometrics* 105, 131 – 159.

Fisher, W.D. (1958): On grouping for maximum homogeneity. *Journal of the American Statistical Association* 53, 789 – 798.

Garcia, R., Perron, P. (1996): An analysis of the real interest rate under regime shifts. *The Review of Economics and Statistics* 78, 111 – 125.

Geweke, J., Porter-Hudak, S. (1983): The estimation and application of long-memory time series models. *Journal of Time Series Analysis* 4, 221 – 238.

Ghysel, E., Guay, A., Hall, A. (1997): Predictive tests for structural change with unknown breakpoint. *Journal of Econometrics* 82, 209 – 233.

Gourieroux, C., Jasiak, J. (2001): Financial Econometrics: Problems, Models, and Methods. *Princeton University Press.*

Granger, C., Hyung, N. (1999): Occasional Structural Breaks and Long Memory. *University of California, San Diego, Discussion Paper* 99.14.

Granger, C., Hyung, N. (2004): Occasional Structural Breaks and Long Memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance* 11, 213 – 228.

Granger, C., Joyeux, R. (1980): An introduction to long-range time series models and fractional integration. *Journal of Time Series Analysis* 1, 15 – 30.

Hansen, B.E. (1997): Approximate asymptotic p values for structural-change tests. *Journal of Business & Economic Statistics* 15, 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, 117 – 128.

Hartigan, J.A. (1975): *Clustering Algorithms.* John Wiley & Sons, New York.

Hassler, U., Nautz, D. (2008): On the persistence of the Eonia spread. *Economics Letters* 10, 184 – 187.

Hassler, U., Wolters, J. (1995): Long memory in inflation rates: International evidence. *Journal of Business & Economic Statistics* 13, 37 – 45.

Heinen, F., Sibbertsen, P., Kruse, R. (2009): Forecasting long memory time series under a break in persistence. *CREATES Research Paper* 53.

Hosking, J.R.M. (1981): Fractional Differencing. *Biometrika* 68, 165 – 176.

Hsu, C.-C. (2005): Long memory or structural changes: An empirical examination on inflation rates. *Economics Letters* 88, 289 – 294.

Hsu, C.-C. (2007): The MOSUM of squares test for monitoring variance changes. *Finance Research Letters* 4, 254 – 260.

Hurvich, C.M., Deo, R., Brodsky, J. (1998): The mean-squared error of Geweke and Porter-Hudaks estimator of the memory parameter of a long-memory time series. *Journal of Time Series Analysis* 19, 19 – 46.

Hurvich, C.M., Moulines, E., Soulier, P. (2005): Estimating Long Memory in Volatility. *Econometrica* 73(4), 1283 – 1328.

Kim, J. (2000): Detection of a change in persistence of a linear time series. *Journal of Econometrics* 95, 97 – 116.

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

Kokoszka, P., Leipus, R. (2002): Detection and estimation of changes in regime. In: *Long-range Dependence: Theory and Applications by Doukhan, P., Oppenheim, G. and Taqqu, M. S., eds.* Birkhauser, Boston, 325 – 337.

Krämer, W., Sibbertsen, P. (2002): Testing for structural change in the presence of long memory. *International Journal of Business and Economics* 1(3), 235 – 242.

Kuan, C.-M., Hornik, K. (1995): The generalized fluctuation test: A unifying view. *Econometric Reviews* 14, 135 – 161.

Kuan, C.-M., Hsu, C.-C. (1998): Change-Point Estimation of fractionally integrated processes. *Journal of Time Series Analyses* 19(6), 693 – 708.

Lavielle, M., Moulines, E. (2002): Least-squares estimation of an unknown number of shifts in a time series. *Journal of Time Series Analysis* 21(1), 33 – 59.

Leisch, F., Hornik, K., Kuan, C.-M. (2000): Monitoring structural changes with the generalized fluctuation test. *Econometric Theory* 16, 835 – 854.

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

Leybourne, S., Taylor, R., Kim, T. (2007): CUSUM of squares-based tests for a change in persistence. *Journal of Time Series Analysis* 28, 408 – 433.

Liu, J., Wu, S., Zidek, J. (1997): On segmented multivariate regression. *Statistica Sinica* 7, 497 – 525.

Nelson, C.R., Plosser, C.R. (1982): Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics* 10 (2), 139 – 162.

Perron, P. (1989): The Great Crash, the Oil Price Shock and the Unit Root Hypothesis. *Econometrica* 57, 1361 – 1401.

Perron, P., Qu, Z. (2010): Long-Memory and Level Shifts in the Volatility of Stock Market Return Indices. *Journal of Business and Economic Statistics* 28(2), 275 – 290.

Pesaran, M.H., Zimmermann, A. (2005): Small sample properties of forecasts from autoregressive models under structural breaks. *Journal of Econometrics* 129, 183 – 217.

Ploberger, W., Krämer, W. (1992): The CUSUM test with OLS residuals. *Econometrica* 60(2), 271 – 285.

R Development Core Team (2011): R: A Language and Environment for Statistical Computing. *Vienna, Austria, http://www.R-project.org.*

Ray, B.K., Tsay, R.S. (2002): Bayesian methods for change-point detection in long-range dependent processes. *Journal of Time Series Analysis* 23, 687 – 705.

Rea, W.S. (2008): The Application of Atheoretical Regression Trees to Problems in Time Series Analysis. *PhD Thesis.* Department of Mathematics and Statistics, University of Canterbury.

Rea, W.S., Reale, M., Cappelli, C., Brown, J.A. (2010): Identification of Changes in Mean with Regression Trees: An Application to Market Research. *Econometric Reviews.* 29(5), 754 – 777.

Ripley, B. (2005): tree: Classification and regression trees. *R package version 1.0-19.*

Robbins, H. (1970): Statistical methods related to the law of iterated logarithms. *The Annals of Mathematical Statistics* 41, 1397 – 1409.

Schwarz, G. (1978): Estimating the dimension of a model. *The Annals of Statistics* 6, 461 – 464.

Shimotsu, K. (2006): Simple (but effective) tests of long memory versus structural breaks. *Queen's Economics Department Working Paper* 1101.

Sibbertsen, P. (2004): Long-memory versus structural change: An overview. *Statistical Papers* 45, 465 – 515.

Sibbertsen, P. and Kruse, R. (2009): Testing for a break in persistence under long-range dependencies. *Journal of Time Series Analysis* 30, 263 – 285.

Siegmund, D. (1985): Sequential Analysis. *Springer, New York.*

Sowell, F. (1990): The fractional unit root distribution. *Econometrica* 58, 495 – 505.

Stock, J.H., Watson, M.W. (2005): Implications of dynamic factor analysis for VAR models. *NBER working paper* 11467.

Taqqu, M. (1977): Law of the iterated logarithm for sums of non-linear functions of gaussian variables that exhibit a long range dependence. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 40, 203 – 238.

Toth, D., Eltinge, J.L. (2010): Building Consistent Regression Trees from Complex Sample Data. *Bureau of Labor Statistics*, Office of Survey Methods Research (OSMR).

Wright, J.H. (1998): Testing for a structural break at unknown date with long-memory disturbances using intervention analysis. *Journal of Time Series Analysis* 19(3), 369 – 376.

Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S., Zhou, Z.-H., Steinbach, M., Hand, D.J., Steinberg, D. (2008): Top 10 algorithms in data mining. *Knowledge and Information Systems* 14, 1 – 37.

Yamaguchi, K. (2011): Estimating a change point in the long memory parameter. *Journal of Time Series Analysis* 32, 304 – 314.

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

Zeileis, A., Leisch, F., Hansen, B., Hornik, K., Kleiber, C . (2002): strucchange: An R Package for Testing for Structural Change in Linear Regression Models. *Journal of Statistical Software* 7(2), 1 – 38.

Zeileis, A., Leisch, F., Kleiber, C., Hornik, K. (2005): Monitoring structural change in dynamic econometric models. *Journal of Applied Econometrics* 20, 99 – 121.

Zheng, G.Y., Penny, R., Reale, M., Chacko, E. (2008): An Empirical Study of Break Point Detection for Seasonal Change in an External Migration Series. *Paper presented at Econometric Society of Australasia Meeting* Wellington, New Zealand.

Zivot, E., Andrews, D.W.K. (1992): Further evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis. *Journal of Business & Economic Statistics* 10 (3), 251 – 270.