Essays on Long Memory Estimation and Testing for Structural Breaks under Long-Range Dependent Errors

Von der Wirtschaftswissenschaftlichen Fakultät der
Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des akademischen Grades

Doktor der Wirtschaftswissenschaften
– Doctor rerum politicarum –

genehmigte Dissertation

von

M.Sc. Simon Wingert

geboren am 24. Februar 1992 in Gehrden

2020

Referent: Prof. Dr. Philipp Sibbertsen, Leibniz Universität Hannover
Korreferent: Prof. Dr. Philipp Otto, Leibniz Universität Hannover
Tag der Promotion: 15.06.2020

Abstract

This thesis contains five essays on estimating the long-memory parameter and testing for structural change under long-range dependent errors. After an introduction in the first Chapter, Chapter 2 suggests a modification of the popular local Whittle long memory estimator in the presence of deterministic low frequency seasonality. The basic idea is motivated by the observation that deterministic seasonality introduces bounded peaks in the spectral density of the time series only at the seasonal frequencies and their harmonics.

Chapter 3 introduces an estimator for time-varying long memory in locally stationary processes that is based on a Whittle-like constrast function. The estimator uses the wavelet transform, which seems natural as wavelets are localized in both time and frequency. Other estimators known in the literature are regression-based so that our new estimator is expected to have lower variance at the price of conceptual complexity.

In Chapter 4 we consider the problem of estimating and testing for multiple strucutral breaks under long memory in multivariate systems. The model allows for a considerable diversity of situations. Structural breaks here refer to breaks in the regression coefficients or the contemporaneous correlation matrix of the errors. We introduce a test for the number of breaks and integrate this test in a procedure that allows us to test for the unknown number of breaks in a given multivariate time series.

Chapter 5 suggests a testing procedure for structural breaks that can distinguish three situations: no break, a break in the mean and a break in persistence. First, we argue that a simple CUSUM based test statistic is able to detect both of the considered forms of structural breaks and rejects the null hypothesis of no break. Second, we suggest to split the sample at an estimated break point and proceed to test which of the forms of a break is present in the time series at hand.

Finally, in Chapter 6 we test electricity load series for a break in mean. To this end we introduce a new test statistic that is robust to long memory and deterministic seasonality. We study the theoretical properties of the test statistic under the null and a specific alternative. Comparing our estimator to known estimators we moreover find higher power against a break at the beginning or at the end of the sample.

Keywords: Long Memory, Seasonality, Whittle estimation, Wavelet analysis, Change-point, break dates, break in persistence

Contents

List of Figures **IV**

List of Tables **IV**

1 Introduction **1**

2 Seasonality Robust Local Whittle Estimation **4**

3 Local Whittle Wavelet Estimation for Locally Stationary Processes **5**
3.1 Introduction 5
3.2 The Model and Assumptions 6
3.2.1 Discrete Wavelet Transform (DWT) 7
3.2.2 The Tangent Process and its Relation to the Long Memory Parameter 9
3.3 Local Whittle Wavelet Estimator 11
3.4 Monte Carlo Simulation 13
3.5 Time-varying Long Memory of Realized Volatility 15
3.6 Conclusion 17
3.7 Appendix 17
3.7.1 Consistency 19
3.7.2 Central Limit Theorem 20

4 Testing for Multiple Structural Breaks in Multivariate Long Memory Time Series **22**
4.1 Introduction 22
4.2 Model 23
4.3 Estimation of the Break Dates and Model Parameters 26
4.4 Testing for Multiple Breaks in Multivariate Time Series 29
4.5 Simulation results 32
4.6 Empirical Application 34
4.7 Conclusions 38
4.8 Appendix 38
4.8.1 Generalised Hájek-Rényi Inequality, SLLN, FCLT 38
4.8.2 10 Properties of the Quasi-Likelihood Ratio 41
4.8.3 Proof of Lemma 4.1 42
4.8.4 Proof of Theorem 4.1 44
4.8.5 Proof of Theorem 4.2 46
4.8.6 Proof of Theorem 4.3 52
4.8.7 Proof of 10 Properties 52

4.8.8 Critical values of the $UDmax\, LR_T$ test 58

5 Distinguishing between Breaks in the Mean and Breaks in Persistence under Long Memory 61

5.1 Introduction 61
5.2 Detecting breaks in the mean or in the persistence 62
5.3 Break in mean or break in persistence? 64
5.4 A Simulation Study 67
5.5 An Application to Inflation Rates 68
5.6 Conclusion 69
5.7 Appendix 70

6 Testing for a break in mean in electricity load data 71

6.1 Introduction 71
6.2 Testing for a break in mean under long-range dependence 72
6.3 The effect of seasonality on the test statistic 74
6.4 A Monte Carlo Study 75
6.5 Breaks in mean in electricity load data 76
6.6 Conclusion 77
6.7 Appendix 78
6.7.1 Proof of Theorems 78
6.7.2 Lemmas 80

Bibliography 81

List of Figures

3.1 Time-varying long-memory functions. 14
3.2 Estimated time-varying long memory parameter of realized volatility. 16

4.1 Hit ratio of REBIT procedure for different long-memory values. . . . 34
4.2 Monthly inflation rates of France and Germany from 1970 to 2019. . . 35

5.1 Inflation rates for Netherlands and France. 69

6.1 Observations of exemplary electricity demand time series. 77

List of Tables

3.1 Bias, Variance and MSE of LWW and LRW estimators with $\phi = 0$. . 14
3.2 Bias, Variance and MSE of LWW and LRW estimators with $\phi = 0.5$. 15

4.1 Persistence of the system and estimated number of breaks. 37
4.2 Asymptotic critical values for the $UDmax\, LR_T$ test, $\alpha = 90\,\%$. 58
4.3 Asymptotic critical values for the $UDmax\, LR_T$ test, $\alpha = 95\,\%$. 59
4.4 Asymptotic critical values for the $UDmax\, LR_T$ test, $\alpha = 99\,\%$. 60

5.1 Size and power results of the test of Section 5.2 under scenario II. . . 67
5.2 Power results of the test of Section 5.2 under scenario III. 67
5.3 Size results of the test of Section 5.3. 68
5.4 Power results of the test of Section 5.3. 68
5.5 P-values for the test statistic R_T and for the test statistic R^*. 69

6.1 Critical values $\lambda_{c,d}(\alpha)$ with $\alpha = 0.95$ and $c = 0.15$. 73
6.2 Size results of the change-point test. 75
6.3 Power results of the test suggested. 76
6.4 P-values for the test statistic $\Lambda_d(n, r)$ for different countries. 77

Chapter 1

Introduction

This thesis deals with long-memory time series. Long memory is a property of second-order stationary time series that concerns the dependence structure of a time series. Loosely speaking, a time series is long-range dependent if the dependence between two points in time decays rather slowly as the distance between the points in time increases.

Long memory has been studied in-depth and many applications of long memory have been found. To just name a few applications in economics we mention inflation rates (Kumar and Okimoto, 2007), interest rates (W.-J. Tsay, 2000), or trading volumes (Fleming and Kirby, 2011). Interestingly, not the stock market returns themselves show characteristics of long memory (Ding et al., 1993), but the second moment of this random quantity. The long-memory nature of volatility (Bollerslev and Mikkelsen, 1996) or realized volatility (Deo et al., 2006) are well established. Moreover, we use electricity load time series in this thesis as an application for our methods, for which long-memory properties have been found by Lacir Jorge Soares and Leonardo Rocha Souza (2006) or Sadaei et al. (2017), among others.

In this thesis we consider broadly two problem areas: the estimation of the long memory parameter and the testing for structural breaks in the presence of long memory. Mathematically, several definitions of long memory exist that are not equivalent. In this thesis we will make use of different definitions of long memory. Thus, it seems appropriate to point out the relationship of these definitions here. Details can be found in Chapter 2 of Pipiras and Taqqu (2017).

The most specific definition of long memory arises if we specify a time series x_t to have a linear representation, i.e. $x_t = \mu + \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}$ for some white noise process ε_t with weights satisfying $\psi_k \sim c_1 k^{2d-1}$, $0 < d < 1/2$ as $k \to \infty$ for some constant $c_1 > 0$. The ARFIMA(p, d, q) model is an example that complies with this definition and we will see an application of this model in Chapter 5.

The second definition used in this thesis is based on the behavior of the spectral density f_x of x_t. If the spectral density obeys $f_x(\lambda) \sim c_2 \lambda^{-2d}$, $0 < d < 1/2$ as $\lambda \to 0+$ for some constant $c_2 > 0$ we call x_t long-range dependent. This definition of long-range dependence is implied by our first definition, but it is more general. It will be used in Chapters 2, 3, and 4 (in a multivariate version). This definition is prevalent in the literature on estimation because many estimating procedures work in the spectral domain due to numerical issues of likelihood-based estimation in the time domain.[1]

[1] This definition is, however, not equivalent to defining long memory by a power-like decaying

A third definition of long memory refers to the behavior of the variance of partial sums $S_n = \sum_{k=1}^{n} x_k$. A time series is long-range dependent if $\text{Var}(S_n) \sim c_3 n^{2d+1}$, $0 < d < 1/2$ as $n \to \infty$ for some constant $c_3 > 0$. It is the most general definition of long memory and implied by the other definitions. The practical importance of this definition can be seen by the fact that under additional assumptions like linearity or Gaussianity the partial sum process S_n converges to fractional Brownian motion for all finite dimensional distributions (see Propositions 2.8.7 and 2.8.8 of Pipiras and Taqqu (2017)). This will lead us to the assumption we make about the long-memory time series in Chapter 6.

Chapters 2 and 3 deal with the estimation of the long-memory parameter d. Parametric procedures are typically at a disadvantage when compared to semiparametric approaches because they require a full specification of the model albeit their potentially greater efficiency if correctly specified. Therefore, we focus on semiparametric estimators. Probably the most well-known semi-parametric estimators are the GPH estimator of Geweke and Porter-Hudak (1983) and the local Whittle estimator of Robinson (1995a). While the former may be conceptually simpler, the latter benefits from its lower variance.

In Chapter 2 we study the effects of deterministic seasonality on the estimation of the long-memory parameter d. Seasonality perturbs the estimation if the seasonality has long periods relative to the observation frequency. Deterministic seasonality affects the periodogram in a similar way as structural breaks or non-periodic trends. Thus, they may be regarded as a type of spurious long memory. Although deterministic seasonality influences the periodogram only at the seasonal frequencies and their harmonics, it causes a sizeable bias of the local Whittle estimator. We document this bias and suggest a solution by omitting affected periodogram ordinates. A similar approach has been suggested by Ooms and Hassler (1997) for the GPH estimator. In our contribution we compare both estimators in a Monte Carlo study. The potential of our proposed estimator is shown in an application to electricity load series. Electricity load series are revisited in Chapter 6.

In Chapter 3 we study the estimation of time-varying long memory for locally stationary processes. Locally stationary processes can roughly be described as processes "which locally at each time point are close to a stationary process but whose characteristics (covariances, parameters, etc.) are gradually changing in an unspecific way as time evolves" (Dahlhaus, 2012). Using estimators based on the wavelet transform that is localized both in the spectral and in the time domain simultaneously is a natural choice when constructing estimators for the long-memory parameter function $d(t)$, $t = 1, \ldots, T$. Our contribution is motivated by the paper of Roueff and Von Sachs (2011), who introduce a wavelet based estimator similar to the GPH estimator in the sense that their estimator is regression based. On the contrary, the estimator introduced here is based on the optimization of a Whittle likelihood. We compare both estimators in a Monte Carlo study and show an application to realized volatility of national stock indices.

Chapters 4, 5, and 6 deal with testing for structural change under long-range dependent errors. Structural change can be seen as a complement to time-varying models. Where the latter assumes that the data generating system changes smoothly, a structural break model takes the standpoint that the system changes rapidly at one instant in time. In this thesis we consider breaks in the long-memory parameter,

autocovariance function: $\gamma(n) \sim \tilde{c}_2 n^{2d-1}$ $(n \to \infty)$ for some constant $\tilde{c}_2$, see Gubner (2005).

i.e. breaks in persistence as well as breaks in the mean and in the variance. As we consider the null hypothesis of no break in all articles, the main theoretical device will be the functional central limit theorem for long-memory time series to derive the limiting distributions. The limiting distribution thus depends on the long-memory parameter d as a nuisance parameter. Hence, we pay special attention to simulation studies that examine how well the limiting distribution is approximated in finite samples.

In Chapter 4 we study a multivariate regression model allowing for multiple breaks in the mean and in the variance. Earlier contributions that allowed for long memory only considered univariate models with breaks in the mean (see inter alia L. Wang (2008); Shao (2011); Iacone et al. (2014); Betken (2016)). Our approach is motivated by Qu and Perron (2007), who analyze a similar model but did not take long-memory errors into account. We suggest to estimate regression coefficients, the contemporaneous covariance matrix of the errors, and the break points by quasi-maximum likelihood. In a first step the theoretical properties of the estimators are studied. We show that the estimators are consistent and give their limiting distributions. In a second step we propose a test statistic and a procedure that allows us to identify the a priori unknown number of break points in a given time series.

Chapter 5 deals with a setup in which we wish to distinguish the null hypothesis of no structural break against two alternatives: a break in the mean and a break in persistence. We devise to use a simple CUSUM test statistic. If a structural break is indeed present, we then show how to distinguish which of the aforementioned forms of a structural break is in place for a time series at hand. The approach is inspired by Aue et al. (2009). We derive the limiting distribution and study the finite sample properties of the test in a Monte Carlo study. An application to inflation rates illustrates the practical utility of this idea.

In Chapter 6 we reconsider the case of electricity load data, for which we want to test the null hypothesis of no change in mean. As indicated above, several testing procedures have been proposed in the literature before. Yet, we suggest a new hypothesis test for detecting a break in mean under long-range dependent errors. Two reasons provide support for this decision. First, we find that existing tests have rather low power against breaks that happen in the beginning or end of the sample. Second, we are able to show that our test statistic is theoretically not affected by deterministic seasonality, which is crucial when using the test with electricity load data. Our idea is motivated by a test of Wu (2004). The test statistic is based on isotonic and antitonic regression. We derive the limiting distribution under the null hypothesis allowing for deterministic seasonality and show consistency of the test against a certain alternative. The test is then applied to electricity load time series.

Chapter 2

Seasonality Robust Local Whittle Estimation

Co-authored with Christian Leschinski and Philipp Sibbertsen.
Applied Economics Letters

Online available at: `https://doi.org/10.1080/13504851.2019.1691710`

Chapter 3

Local Whittle Wavelet Estimation for Locally Stationary Processes

Co-authored with Philipp Sibbertsen.

3.1 Introduction

Long memory has been recognized as a feature of many practically relevant time series, see e.g. the recent monograph of Pipiras and Taqqu (2017). However, in many applications the long memory parameter d has been found to be changing over time. A few examples of data that seems to have time-varying long memory includes classical examples like the Nile river minima (Beran (2009)), individual stock and index returns (Ray and R. S. Tsay (2002), Z. Lu and Guegan (2011)), video data traffic (Beran et al. (1995)), and electricity prices (Haldrup and Nielsen (2006)).

In order to analyze the long memory parameter of such time series one may stick with the well-known semiparametric estimators in the frequency domain. In particular, there is the Geweke-Porter-Hudak (GPH) estimator that has been introduced by Geweke and Porter-Hudak (1983) and analyzed by Robinson (1995b). The GPH estimator is regression-based and uses the log-linear relationship of the long memory parameter and the periodogram ordinates. On the other hand one can use the likelihood-based approach in the frequency domain called local Whittle (Fourier) estimation (LWF). This estimator was proposed by Künsch (1987) and analyzed by Robinson (1995a). However, as the Fourier approach is not well localized in time and always "utilizes" information of the whole sample these methods seem questionable for time-varying long memory.

In contrast, wavelet-based methods seem to be better suited for the case of time-varying long memory as has been pointed out by Whitcher and Jensen (2000) or Roueff and Von Sachs (2011). The localization of wavelets not only in frequency, but also in the time domain makes them a natural candidate for an estimator in the context of time-varying long memory. Similar to the GPH estimator in the frequency domain there exists the local regression wavelets estimator that has been analyzed by Moulines et al. (2007b) and Moulines et al. (2007a). Additionally, the counterpart of the LWF estimator is the local Whittle wavelet estimator which was introduced by Moulines et al. (2008). An overview and a concise comparison of the different methods in the frequency and wavelet domain can be found in Faÿ et al. (2009).

In this paper we focus on the semi-parametric estimation of the long memory parameter of locally stationary processes. Roueff and Von Sachs (2011) introduced a locally stationary long memory model and provided an semi-parametric estimator that is regression-based. In this model the spectral density obeys for all observations t, $1 \leq t \leq T$, the relation

$$f(\lambda) \sim \lambda^{-2d(t)}\, C(t), \quad C(t) > 0, \qquad \text{as } \lambda \to 0+,$$

where "$\sim$" means that the fraction of both sides tends to 1 asymptotically. Roueff and Von Sachs (2011) give several examples of how usual time series models may be written with time-varying long memory. Here, we will use the same suggested model and develop a local Whittle wavelet estimator for this process.

Locally stationary processes have been introduced by Dahlhaus (1997) and Dahlhaus (2000). A special property when estimating parameters in the model is the kind of asymptotics that is referred to as infill asymptotics (Dahlhaus (2012)). This term describes the fact that one imagines the increasing sample to become more densely sampled. That is, the time domain is rescaled to the unit interval and for any time point $u \in [0,1]$ one wishes to estimate $d(u)$ by $d(t/T)$ if $u \approx t/T$ where T describes the number of observations. Increasing the number of observations can be seen to improve this approximation $u \approx t/T$ for any time point u. We are able to prove consistency and asymptotic normality of our estimator in this framework.

The paper is structured as follows. The purpose of Section 3.2 is three-part. At first we introduce the model and define locally stationary long memory. Then we present the assumptions on the wavelets in 3.2.1 and define the so-called tangent process in 3.2.2, which provides us with the relation between our estimator and the local long memory parameter. In Section 3.3 we introduce the estimator and state its asymptotic properties, i.e. consistency and asymptotic normality. With a Monte Carlo study we explore its small sample properties in the subsequent Section 3.4. We give a practical example and examine the time-varying long memory property of several financial assets in Section 3.5. Section 3.6 concludes. Proofs are postponed to the Appendix.

3.2 The Model and Assumptions

We define locally stationary long memory processes via its spectral representation. For $p \in \mathbb{N}_0$ and an array of $L^2([-\pi,\pi])$ functions $A^0_{t,T}(\lambda)$ with real-valued Fourier coefficients we define

$$\Delta^p X_{t,T} = \int_{-\pi}^{\pi} A^0_{t,T}(\lambda)\exp(i\lambda t) dZ(\lambda), \quad t = 1, \ldots, T, \tag{3.1}$$

where $dZ(\lambda)$ is the spectral representation of a centered weak white noise with unit variance, i.e.

$$\varepsilon_t = \int_{-\pi}^{\pi} \exp(i\lambda t) dZ(\lambda), \quad t \in \mathbb{Z}. \tag{3.2}$$

Thus, the mueasure $Z(\lambda)$ is Hermitian and complex valued with stationary orthogonal increments on $[-\pi,\pi]$. We assume that there exists a function $A(u,\lambda)$ in $L^2([0,1] \times [-\pi,\pi])$ and two constants $c > 0$ and $D < 1/2$ such that

$$|A^0_{t,T}(\lambda) - A(t/T,\lambda)| \leq c\,T^{-1}\,|\lambda|^{-D}, \qquad 1 \leq t \leq T, \quad -\pi \leq \lambda \leq \pi, \tag{3.3}$$

and

$$|A(u,\lambda) - A(v,\lambda)| \leq c\,|v-u|\,|\lambda|^{-D}, \qquad 0 \leq u,v \leq 1, \quad -\pi \leq \lambda \leq \pi. \tag{3.4}$$

This definition has been suggested by Roueff and Von Sachs (2011) and it allows a singularity at the zero frequency bounded by $|\lambda|^{-D}$. In addition we assume that the spectral representation of the process $X_{t,T}$ has a density f that is called time-varying generalized spectral density by Roueff and Von Sachs (2011):

$$f(u,\lambda) = |1-\exp(-i\lambda)|^{-2p}|A(u,\lambda)|^2. \tag{3.5}$$

Definition 1. *The process* $\{X_{t,T} : t = 1,\ldots,T,\ T \geq 1\}$ *has a* local memory parameter $d_0(u) \in (-\infty, p+1/2)$ *at time* $u \in [0,1]$ *if it satisfies Equations* (3.1), (3.3) *and* (3.4). *In addition its time-varying generalized spectral density satisfies the following conditions*

$$f(u,\lambda) = |1-\exp|^{-2d(u)} f^*(u,\lambda), \qquad \lambda \in [-\pi,\pi],$$

where $f^*(u,0) > 0$ *and*

$$|f^*(u,\lambda) - f^*(u,0)| \leq C\, f^*(u,0)\, |\lambda|^{\beta}, \qquad \lambda \in [-\pi,\pi], \tag{3.6}$$

where $C > 0$ *and* $\beta \in (0,2]$.

Using terminology introduced by Moulines et al. (2008) we may say that $Z \in \mathscr{H}(\beta, C, \pi)$ where $\mathscr{H}(\beta, C, \pi)$ is the class of finite nonnegative symmetric measures on $[-\pi,\pi]$ that have a density f^* that complies with Equation (3.6).

The following assumption is from Roueff and Von Sachs (2011).

Assumption 1. *The array* $\{X_{t,T}\}$ *of real-valued random variables has local memory parameter* $d_0(u) \in (-\infty, p+1/2)$ *at time* $u \in [0,1]$. *Moreover,* $\{\varepsilon_t\}$ *in Equation* (3.2) *is a weak white noise such that* $E[\varepsilon_0] = 0$, $\mathrm{Var}(\varepsilon_0) = 1$, $E[\varepsilon_t^4] < \infty$ *for all* $t \in \mathbb{Z}$ *and the fourth-order cumulants of its spectral representation* $dZ(\lambda)$ *satisfy*

$$\mathrm{Cum}(dZ(\lambda_k), 1 \leq k \leq 4) = \hat{\kappa}_4(\lambda) d\mu(\lambda), \quad \lambda = (\lambda_k)_{1\leq k\leq 4} \in [-\pi,\pi]^4,$$

where $\hat{\kappa}_4(\lambda) = \hat{\kappa}_4(\lambda_1,\lambda_2,\lambda_3)$ *is a bounded function on* $[-\pi,\pi]^3$ *and* μ *is the measure on* $[-\pi,\pi]^4$ *such that for any* 2π*-periodic functions* $A_k, 1 \leq l \leq 4$,

$$\int_{[-\pi,\pi]^4} \prod_{k=1}^{4} A_k(\lambda_k) d\mu(\lambda) = \int_{[-\pi,\pi]^3} A_4(-\lambda_1-\lambda_2-\lambda_3) \prod_{k=1}^{3} A_k(\lambda_k) d\lambda.$$

3.2.1 Discrete Wavelet Transform (DWT)

We largely follow the descriptions of Moulines et al. (2007b) about the discrete wavelet transform (DWT) of $\{X_{t,T}, 1 \leq t \leq T\}$ for a given scale function ϕ and wavelet ψ. The requirements for resp. assumptions about these functions are listed below. Also, one may check Percival and Walden (2006) for more information on the topic.

The wavelet coefficients $W_{j,k;T}$ for $j \geq 0$ and $1 \leq t \leq T$ are defined by

$$W_{j,k;T} = \sum_{t=1}^{T} h_{j,2^j k - t} X_{t,T}, \quad k = 0,\ldots,T_j - 1, \tag{3.7}$$

where $h_{j,t}, t \in \mathbb{Z}$ denotes the wavelet filter at scale j and T_j is the number of available wavelet coefficients at scale j. The filter coefficients are related to ϕ and ψ by

$$h_{j,t} = 2^{-j/2} \int_{-\infty}^{\infty} \phi(u+t)\psi(2^{-j}u)du.$$

Note that we adopt from the engineering literature the convention that large j correspond to coarser scales. The number T_j of available wavelet coefficients at scale j obeys for some constant c independent of the scale j

$$T2^{-j} - c \leq T_j \leq T2^{-j}. \tag{3.8}$$

The filter coefficients $h_{j,\cdot}$ and the number T_j have support $\{t\colon h_{j,2^jk-t} \neq 0\}$ included in $\{1, \ldots, T\}$ for $k = 0, \ldots, T_j - 1$. In order to state the assumptions on ϕ and ψ we introduce the filter transfer function of the wavelet filter $h_{j,t}$

$$H_j(\lambda) = \sum_{t \in \mathbb{Z}} h_{j,t} \exp(-it\lambda).$$

The following set of assumption is common in the literature (Moulines et al. (2007b), Moulines et al. (2008) or Roueff and Von Sachs (2011)).

Assumption 2. *We assume that the scale function ϕ and wavelet function ψ meet the following requirements for $M \in \mathbb{N}$ and $\alpha \in \mathbb{R}$.*

(a) *ϕ and ψ are compactly supported, integrable, $\int_{-\infty}^{\infty} \phi(t)dt = 1$ and $\int_{-\infty}^{\infty} \psi^2(t)dt = 1$.*

(b) *There exists $\alpha > 1$ such that $\sup_{\xi \in \mathbb{R}} |\hat{\psi}(\xi)|(1 + |\xi|)^{\alpha} < \infty$, where $\hat{\psi}(\xi) = \int_{-\infty}^{\infty} \psi(t) \exp(-it\xi)dt$ denotes the Fourier transform of ψ.*

(c) *The function ψ has M vanishing moments, $\int_{-\infty}^{\infty} t^m \psi(t)dt = 0$ for all $m = 0, \ldots, M-1$.*

(d) *The function $\sum_{k \in \mathbb{Z}} k^m \phi(\cdot - k)$ is a polynomial of degree m for all $m = 0, \ldots, M-1$.*

(e) *For all $u \in [0,1]$ the values M, α, β are such that $M \geq p \vee d_0(u)$ und $d_0(u) > (1+\beta)/2 - \alpha$.*

For a more detailed interpretation of Assumption 2 we refer to the literature. However, they are mainly needed to guarantee that the emerging wavelet coefficients are stationary. Under **(c)**-**(d)** the filter $h_{j,t}$ can be described as the convolution of the difference filter Δ^M and a finite impulse response filter $\tilde{h}_{j,t}$. When $M \geq p$ we can rewrite Equation (3.7) as

$$W_{j,k;T} = \sum_{t=1}^{T} \tilde{h}_{j,2^jk-t} (\Delta^p X)_{t,T}, \quad k = 0, \ldots, T_j - 1,$$

where $h_{j,\cdot} = \tilde{h}_{j,\cdot} * \Delta^p$.

3.2.2 The Tangent Process and its Relation to the Long Memory Parameter

We now review the concept of the local wavelet spectrum $\sigma^2(u) = \{\sigma_j^2(u), j \geq 0\}$ where for $j \geq 0$ the quantity $\sigma_j^2(u)$ is the variance of the wavelet coefficients at scale j of the so-called tangent stationary process $\Delta^p X_t(u)$. Under the assumption $M \geq p$ this quantity $\sigma^2(u)$ is well-defined because the wavelet coefficients are weakly stationary. In particular, from (3.5) we have

$$\sigma_j^2(u) = \int_{-\pi}^{\pi} |H_j(\lambda)|^2 f(u,\lambda) d\lambda.$$

For $u \in [0,1]$ we can define a tangent stationary process for the p-th increment which is weakly stationary

$$\Delta^p X_t(u) = \int_{-\pi}^{\pi} A(u,\lambda) \exp(it\lambda) dZ(\lambda).$$

Thus, we can define the wavelet coefficients of the tangent process

$$\begin{aligned} W_{j,k}(u) &= \sum_{t=1}^{T} \tilde{h}_{j,2^j k - t} (\Delta^p X)_t(u) \\ &= \int_{-\pi}^{\pi} \tilde{H}_j(\lambda) A(u,\lambda) \exp(i\lambda 2^j k) dZ(\lambda). \end{aligned}$$

The preceding quantities are connected by the relation $\sigma_j^2(u) = E[W_{j,k}^2(u)]$. Note that these wavelet coefficients are those of a process with generalized spectral density $f(u,\cdot)$. Fixing $u \in [0,1]$ we thus can use the results of Moulines et al. (2007b) on the spectral properties of wavelet coefficients $W_{j,k}(u)$.

We now introduce the notion of the so-called tangent scalogram $\tilde{\sigma}_{j,T}^2$ whose asymptotic properties are the main ingredient for our estimator to be defined below. In order to do so, we average the wavelet coefficients of the tangent process by using a kernel $\gamma_{j,T}(k)$ that is concentrated around the index $k \approx uT_j$. As usual for kernel weights the weights are assumed to be non-negative and sum to one, i.e. $\sum_{k=0}^{T_j-1} \gamma_{j,T}(k) = 1$. The tangent scalogram is

$$\tilde{\sigma}_{j,T}^2(u) = \sum_{k=0}^{T_j-1} \gamma_{j,T}(k) W_{j,k}^2(u).$$

However, this tool is not an estimator as it cannot be computed from the observations $X_{1,T}, \ldots, X_{T,T}$. Instead we calculate the local scalogram defined as

$$\hat{\sigma}_{j,T}^2(u) = \sum_{k=0}^{T_j-1} \gamma_{j,T}(k) W_{j,k;T}^2.$$

Being able to approximate the local scalogram $\hat{\sigma}_{j,T}^2(u)$ asymptotically by the tangent scalogram $\tilde{\sigma}_{j,T}^2(u)$ requires additional assumptions on the weights $\gamma_{j,T}$ as introduced by Roueff and Von Sachs (2011). For this reason the authors use a quantity $\Gamma_q(u;j,T)$ for describing the localization property of the weights defined as

$$\Gamma_q(u;j,T) = \sum_{k=0}^{T_j-1} |\gamma_{j,T}(k)| \, |k - Tu2^{-j}|^q.$$

Furthermore they denote the Fourier transform of the weights as

$$\Phi_{j,T}(\lambda; i, v) = \sum_{l \in \mathcal{T}_j(i,v)} \gamma_{j-i,T}(2^i l + v) \exp(il\lambda),$$

where $0 \leq i \leq j, v \in \{0, \ldots, 2^i - 1\}$ and $\mathcal{T}_j(i, v)$ describes the number of available wavelet coefficients at scale $j - i$, i.e.

$$\mathcal{T}_j(i, v) = \{l \colon 0 \leq l \leq 2^{-i}(T_{j-i} - v)\}.$$

Lastly, the largest weight is called

$$\delta_{j,T} = \sup_{k=0,\ldots,T_j-1} \gamma_{j,T}(k).$$

With these definitions we can state the set of assumptions on the weights.

Assumption 3. *The index j must depend on T in such a way that the weights $(\gamma_{j,T}(k))_k$ satisfy the following asymptotic properties as $T \to \infty$.*

(a) $\delta_{j,T} \to 0$ *and for any fixed integer i, $\delta_{j+i,T} \sim 2^i \delta_{j,T}$.*

(b) *For all $i, i' \geq 0, v \in \{0, \ldots, 2^i - 1\}$ and $v' \in \{0, \ldots, 2^{i'} - 1\}$ there exists a constant $V(i, v; i', v')$ such that*

$$\delta_{j,T}^{-1} \int_{-\pi}^{\pi} \Phi_{j,T}(\lambda; i, v) \overline{\Phi_{j,T}(\lambda; i', v')} d\lambda \to V(i, v; i', v').$$

(c) *For all $\eta > 0, i \geq 0$ and $v \in \{0, \ldots, 2^{i'} - 1\}$ we have*

$$\delta_{j,T}^{-1/2} \sup_{\eta \leq |\lambda| \leq \pi} |\Phi_{j,T}(\lambda; i, v)| \to 0.$$

(d) *For $q = 0, 1, 2$ we have*

$$\Gamma_q(u; j, T) = O(\delta_{j,T}^{-q}).$$

Roueff and Von Sachs (2011) give examples of weights that meet these requirements and in our Monte Carlo simulation in Section 3.4 we will use these suggested weights.

The authors prove the following relation between the local scalogram $\hat{\sigma}_{j,T}^2$ and the true local memory parameter. It is the key in order to understand the asymptotic properties of our estimator.

For the statement we define the function $\mathbf{K}$ as

$$\mathbf{K}(d) = \int_{-\pi}^{\pi} |\xi|^{-2d} \, |\hat{\psi}(\xi)|^2 \, d\xi. \tag{3.9}$$

Theorem 3.1 (Roueff and Von Sachs (2011) (Theorem 1))**.** *Let $u \in [0, 1]$ and consider a model satisfying Assumptions 1 and 2. Then we have, as $j \to \infty$,*

$$\sigma_j^2(u) = f^*(u, 0)\mathbf{K}(d_0(u))2^{2jd_0(u)}\{1 + O(2^{-\beta j})\}. \tag{3.10}$$

Suppose moreover that Assumption 3 holds and that

$$2^{\{3+2(p-d_0(u))\}L} T^{-2} \delta_{L,T}^{-2} \;\to\; 0.$$

Then we have for $j = L + i$ with $i = 0, \dots, l$,

$$E\left[\left(2^{-2Ld_0(u)}\hat{\sigma}^2_{j,T}(u) - f^*(u,0)\mathbf{K}(d_0(u))2^{id_0(u)}\right)^2\right] = O\left(\delta_{L,T} + 2^{(3+2(p-d_0(u)))L}T^{-2}\delta^{-2}_{L,T} + 2^{-2\beta L}\right). \tag{3.11}$$

The first part connects the true local memory parameter $d_0(u)$ and the local wavelet spectrum $\sigma^2_j(u)$. The second part gives a bound on the mean squared error between the local scalogram $\hat{\sigma}^2_{j,T}(u)$ and the asymptotic form of the local wavelet spectrum. These Equations equip us with the right tool as they render the variance of the local scalogram normalized by the local wavelet spectrum tractable. This quantity denoted as

$$\mathrm{Var}\left(\hat{\sigma}^2_{j,T}(u) \,/\, \sigma^2_j(u)\right), \qquad j = L + i, \quad i = 0, \dots, l,$$

is considered in Moulines et al. (2008) in their Condition 1. This condition must be fulfilled by any process for their estimator to be applicable.

3.3 Local Whittle Wavelet Estimator

Following the reasoning of Moulines et al. (2008) we define our estimator as the minimum of a pseudo negative log-likelihood:

$$\hat{d}_\ell = \hat{d}_\ell(u) = \arg\min_{d\in\mathbb{R}} \hat{L}_\ell(d(u)),$$

where for a set ℓ of wavelet coefficients to be specified later

$$\hat{L}_\ell = \hat{L}_\ell(d(u)) = \log \sum_{(j,k)\in\ell} 2^{2d(u)(\langle\ell\rangle - j)}\,\gamma_{j,T}(k)\,W^2_{j,k;T}.$$

Here, $|\ell|$ denotes the number of elements of the set ℓ and $\langle\ell\rangle$ is the average scale

$$\langle\ell\rangle = \frac{1}{|\ell|}\sum_{(j,k)\in\ell} j.$$

In this paper we only consider the case that the set of wavelet coefficients ℓ involves a fixed number of scales. The case of an asymptotically increasing number of scales that has been studied by Moulines et al. (2008) cannot be dealt with by the framework of Roueff and Von Sachs (2011). We introduce some notation to state this idea formally. Define for an integer T and two bounds of scales $j_0 \le j_1$ the set

$$\ell_T(j_0, j_1) = \{(j,k) : \ j_0 \le j \le j_1, \ 0 \le k \le T_j\},$$

where T_j is the number of available wavelet coefficients at scale j described by (3.8). We call the sequence of the lower scale index $\{L_T\}$, the number of involved scale indices l and the sequence of the maximal scale indices $\{J_T\}$ for all T. So, we must have

$$0 \le L_T < L_T + l \le J_T, \qquad J_T = \max\{j : T_j \ge 1\}.$$

Therefore, the scale indices that are included in the pseudo-likelihood function is indicated by the set $\ell_T(L_T, L_T + l)$ and the estimator is $\hat{d}_{\ell_T(L_T,L_T+l)}(u)$.

Now we can state the consistency of the estimator.

Theorem 3.2. *Let Assumptions* $1-3$ *hold and assume that the sequences* $\{L_T\}$ *describes the lower index of scales and suppose that as* $T \to \infty$

$$2^{\{3+2(p-d_0(u))\}L_T}T^{-2}\delta_{L,T}^{-2} + L_T^2(T2^{-L_T})^{-1/4} + L_T^{-1} \;\to\; 0.$$

Then for any $u \in [0,1]$ *the estimator* $\hat{d}_{\ell_T(L_T,L_T+l)}$ *is consistent with*

$$\hat{d}_{\ell_T(L_T,L_T+l)}(u) = d_0(u) + O_P\left(\left(T2^{-L_T}\right)^{-1/2} + 2^{-\beta L_T}\right).$$

We derive a central limit theorem for the estimator $\hat{d}_{\ell_T(L_T,L_T+l)}$ of $d_0(u)$ for $u \in [0,1]$ when further assuming that $X_{t,T}$ is Gaussian. In order to state the asymptotic variance in the following theorem we need some further notation. For $0 \le p \le i \le l$ we define

$$\Sigma_{i,p}(u) = 2^{i+1}\sum_{v=0}^{2^p-1} V(0,0;p,v)\int_{-\pi}^{\pi}|\mathbf{D}_{\infty,p,v}(\lambda;d_0(u))|^2 d\lambda,$$

where $\mathbf{D}_{\infty,p,v}(\lambda;d_0(u))$ is defined in Moulines et al. (2008) (equation (11)). We denote for integers $l \ge 1$,

$$\eta_l = \sum_{j=0}^{l} j\frac{2^{-j}}{2-2^{-l}} \quad \text{and} \quad \kappa_l = \sum_{j=0}^{l}(j-\eta_l)^2\frac{2^{-j}}{2-2^{-l}}. \tag{3.12}$$

Finally for $u \in [0,1]$,

$$V(d_0(u),l) = \frac{1}{\left((2-2^{-l})\,\kappa_l\,\log(2)\,\mathbf{K}(d_0(u))\right)^2}\times \left\{\sum_{i=0}^{l}(i-\kappa_L)^2\,\Sigma_{i,0}(u) + 2\sum_{i=1}^{l}\sum_{p=1}^{i}\Sigma_{i,p}(u)\,2^{2d_0(u)p}\,(i-\eta_l)(i+u-\eta_l)\right\} \tag{3.13}$$

where $\mathbf{K}(d)$ is defined in (3.9).

Theorem 3.3. *Let Assumptions* $1-3$ *hold, where the innovations in Equation* (3.2) *follow a normal distribution* $\varepsilon_t \sim N(0,s^2)$*, and assume that the sequence* $\{L_T\}$ *describes the lower index of scales. Additionally, let* $\{L_T\}$ *be a sequence such that*

$$2^{\{3+2(p-d_0(u))\}L_T}T^{-2}\delta_{L_T,T}^{-3} + 2^{-2\beta L_T}\delta_{L,T}^{-1} + L_T^2(T2^{-L_T})^{-1/4} + L_T^{-1} \;\to\; 0. \tag{3.14}$$

Then as $T \to \infty$ *for any* $u \in [0,1]$,

$$T2^{-L_T}\,\delta_{L_T,T}^{-1/2}\,(\hat{d}_{\ell_T(L_T,L_T+l)} - d_0(u)) \;\xrightarrow{\mathcal{L}}\; \mathcal{N}\left(0, V(d_0(u),l)\right).$$

Notice the similarity of the the first and second term of the condition (3.14) in the central limit theorem with the condition (50) of Theorem 3 of Roueff and Von Sachs (2011). Additionally, one may note that the remaining terms are unchanged compared to the condition of Theorem 3.2. Moulines et al. (2008) need to change their condition (31) in order to control for the bias, which we do effectively by our second term in (3.14).

3.4 Monte Carlo Simulation

We investigate the small sample properties of our local Whittle wavelet estimator (which we denote by LWW in the following) and compare its performance in this Section with the local wavelet regression estimator of Roueff and Von Sachs (2011) (denoted by LRW in the following). For the local Whittle wavelet estimator and the estimator of Roueff and Von Sachs (2011) we use weights $\gamma_{j,T}$ that have been suggested by these authors. Specifically, for a given bandwidth b_T we define the weights by

$$\gamma_{j,T}(k) = \rho_{j,T}^{-1} \mathbb{1}_{[-1/2,1/2]} \left(\frac{uT_j - k}{b_T T_j} \right), \qquad k = 1, \ldots, T,$$

where $\rho_{j,T}$ is normalizing the weights in order for them to sum to one. We set $b_T = 0.25$ in accordance with the aforementioned source. Additionally, we assume that as $T \to \infty$ we have $b_T \to 0$ and $T_j b_T \to \infty$. Then the weights satisfy Assumption 3.

In order to get Wavelet coefficients that fulfill Assumption 2 we use the Daubechies and Coiflet wavelets (cf. Faÿ et al. (2009)). These wavelets can produce M vanishing moments. In the following we refer to Daubechies wavelets with, say, $M = 6$ vanishing moments as $D6$ and to Coiflet wavelets with $M = 6$ as $C6$. In our experiments we found small values of M to perform best so we do not report higher values of vanishing moments M.

Our Monte Carlo simulation is performed by repeating estimation $10,000$ times. We set the length of the simulated time series to $T = 2^{10} = 1024$. As we estimate $d(u)$ for $u \in [0.2; 0.8]$ we report here the integrated squared bias and the integrated variance of the estimators. Therefore, we have

$$\mathrm{Bias}(\hat{d}) = \sum_{t=[0.2\,T]}^{[0.8\,T]} \left(\mathrm{E}(\hat{d}(t/T)) - d_0(t/T) \right)^2,$$

$$\mathrm{Var}(\hat{d}) = \sum_{t=[0.2\,T]}^{[0.8\,T]} \mathrm{E}\left(\left(\hat{d}(t/T) - \mathrm{E}(\hat{d}(t/T)) \right)^2 \right).$$

Obviously, the expectation in the preceding equations is estimated by the mean of the Monte Carlo sample.

We use four different functions for $d_0(u)$ in order to analyze the applicability of the estimator in various situations. The different functions can be seen in Figure 3.1. The functions are defined as follows

(i) cosine: $d_0^1(u) = 0.6 + (1 - \cos(u\,\pi/2))/3,$

(ii) linear: $d_0^2(u) = 0.6 + 0.3\,u,$

(iii) cubic: $d_0^3(u) = 0.6 + 0.3\,u^3,$

(iv) logistic: $d_0^4(u) = 0.6 + 0.3\,\dfrac{1}{1 + \exp(-8\,(u - 0.5))}.$

The values of the long memory parameter are in the non-stationary region $d > 0.5$ as these values are often found in applications, e.g. the long memory of volatility (Wenger, Leschinski, and Sibbertsen (2018b)). This choice is further motivated by the fact that it has often been pointed out that wavelet methods seem to be perfect suited for nonstationary long memory (Whitcher and Jensen (2000)).

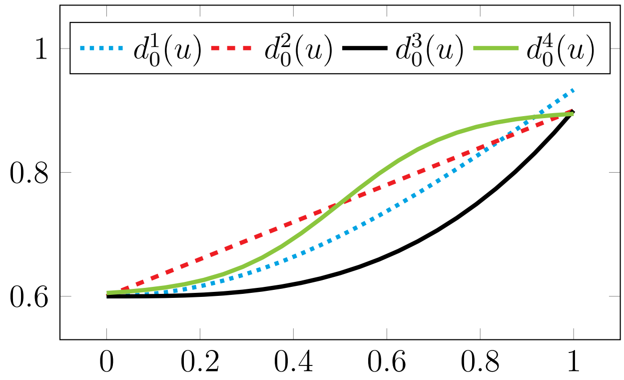


Figure 3.1: Functions $d_0^i(u)$, $i \in \{1, 2, 3, 4\}$ that are used in the Monte Carlo simulations.

In the experiments we use an tv-ARFIMA$(p, d(u), 0)$ model which has been introduced by Roueff and Von Sachs (2011). In order to study the effects of short-run components we choose $p = 0$ and $p = 1$ with an AR-coefficient $\phi_1 = 0.5$.

The results of the Monte Carlo simulation can be found in the Tables 3.1 and 3.2. Overall, we see that the performance of our estimator measured by the MSE is superior to the estimator of Roueff and Von Sachs (2011). Unsurprisingly, using the higher scales $4 - 6$ provide best estimates, thus we just report their performance. Also the estimates are worse if the simulated process has an AR-coefficient.

				LWW			LRW		
d^i	Wavelet	L	U	Bias($\hat{d}$)	Var($\hat{d}$)	MSE	Bias($\hat{d}$)	Var($\hat{d}$)	MSE
d^1	D6	4	5	7.2	43.8	51.0	9.1	56.0	65.1
		4	6	7.3	43.8	51.0	9.1	56.1	65.2
	C6	4	5	7.6	42.6	50.3	9.7	52.7	62.4
		4	6	7.6	42.5	50.2	9.7	52.6	62.3
d^2	D6	4	5	9.1	43.8	52.9	12.2	57.8	69.9
		4	6	9.1	43.8	52.9	12.2	57.7	69.9
	C6	4	5	7.2	40.6	47.8	9.0	51.9	60.9
		4	6	7.2	40.6	47.8	9.0	51.8	60.8
d^3	D6	4	5	6.9	43.2	50.1	9.1	53.6	62.7
		4	6	6.9	43.2	50.1	9.1	53.5	62.6
	C6	4	5	7.8	47.8	55.7	10.5	60.1	70.5
		4	6	7.8	47.9	55.7	10.5	60.1	70.5
d^4	D6	4	5	4.6	44.1	48.7	4.2	56.2	60.4
		4	6	4.6	44.1	48.7	4.2	56.3	60.5
	C6	4	5	5.1	44.5	49.5	4.8	60.7	65.5
		4	6	5.1	44.5	49.6	4.8	60.8	65.6

Table 3.1: Bias, Variance and MSE of LWW and LRW of Roueff and Von Sachs (2011) estimators for the tv-ARFIMA using Daubechies wavelets and without short-run dynamics, i.e. $\phi = 0$.

In Table 3.1 we consider no AR-coefficient for the data generating process and find that the LWW estimator performs better in terms of squared Bias and Variance for d^1, d^2 and d^3. However, its Bias is slightly worse for the long memory parameter d^4 although its Variance is much lower. All in all it is not clear which wavelet - Daubechies or Coiflet - perform better as it seems to depend on the form of the long memory parameter.

				LWW			LRW		
d^i	Wavelet	L	U	Bias($\hat{d}$)	Var($\hat{d}$)	MSE	Bias($\hat{d}$)	Var($\hat{d}$)	MSE
d^1	D6	4	5	56.8	43.8	100.6	53.1	54.4	107.6
		4	6	56.8	43.8	100.6	53.2	54.4	107.6
	C6	4	5	52.7	44.4	97.1	50.1	57.0	107.1
		4	6	52.9	44.4	97.3	50.3	57.0	107.3
d^2	D6	4	5	54.0	46.0	100.0	50.7	59.5	110.2
		4	6	54.0	46.1	100.1	50.8	59.6	110.3
	C6	4	5	46.1	43.3	89.3	43.0	54.0	97.0
		4	6	46.2	43.2	89.3	43.1	53.9	97.0
d^3	D6	4	5	53.6	43.2	96.8	50.4	58.7	109.0
		4	6	53.7	43.1	96.9	50.5	58.6	109.1
	C6	4	5	48.5	42.9	91.4	46.4	55.6	102.1
		4	6	48.5	42.9	91.5	46.5	55.7	102.1
d^4	D6	4	5	82.3	45.0	127.2	75.9	57.3	133.2
		4	6	82.4	44.9	127.3	76.0	57.2	133.2
	C6	4	5	77.1	43.2	120.3	71.4	57.0	128.4
		4	6	77.2	43.2	120.4	71.6	57.0	128.5

Table 3.2: Bias, Variance and MSE of LWW and LRW of Roueff and Von Sachs (2011) estimators for the tv-ARFIMA with short-run dynamics, i.e. $\phi = 0.5$.

In Table 3.2 we introduce short-run dynamics in the form of an AR-coefficient $\phi = 0.5$. Still, the LWW estimator is better in terms of the MSE. Nevertheless, closer inspection show that its squared Bias is larger and its Variance is lower than the LRW estimator. We find that the Coiflet wavelets perform much better in this case for both estimators.

3.5 Time-varying Long Memory of Realized Volatility

In this section we estimate the time-varying long memory parameter of several major national stock indices like the S&P 500 index for the United States of America. This analysis is related to and inspired by the findings of Jensen and Whitcher (2014) who study the long memory of exchange rates. They take their starting point in the stylized fact about financial markets that their volatility has long memory (Ding et al. (1993); Baillie (1996); Andersen et al. (2001)). However, long memory may arise from structural breaks, day-of-the-week seasonality or macroeconmic effects like market crashes and policy changes (Banerjee and Urga (2005); Martens et al. (2009)). Therefore one may argue that the environment in which long memory

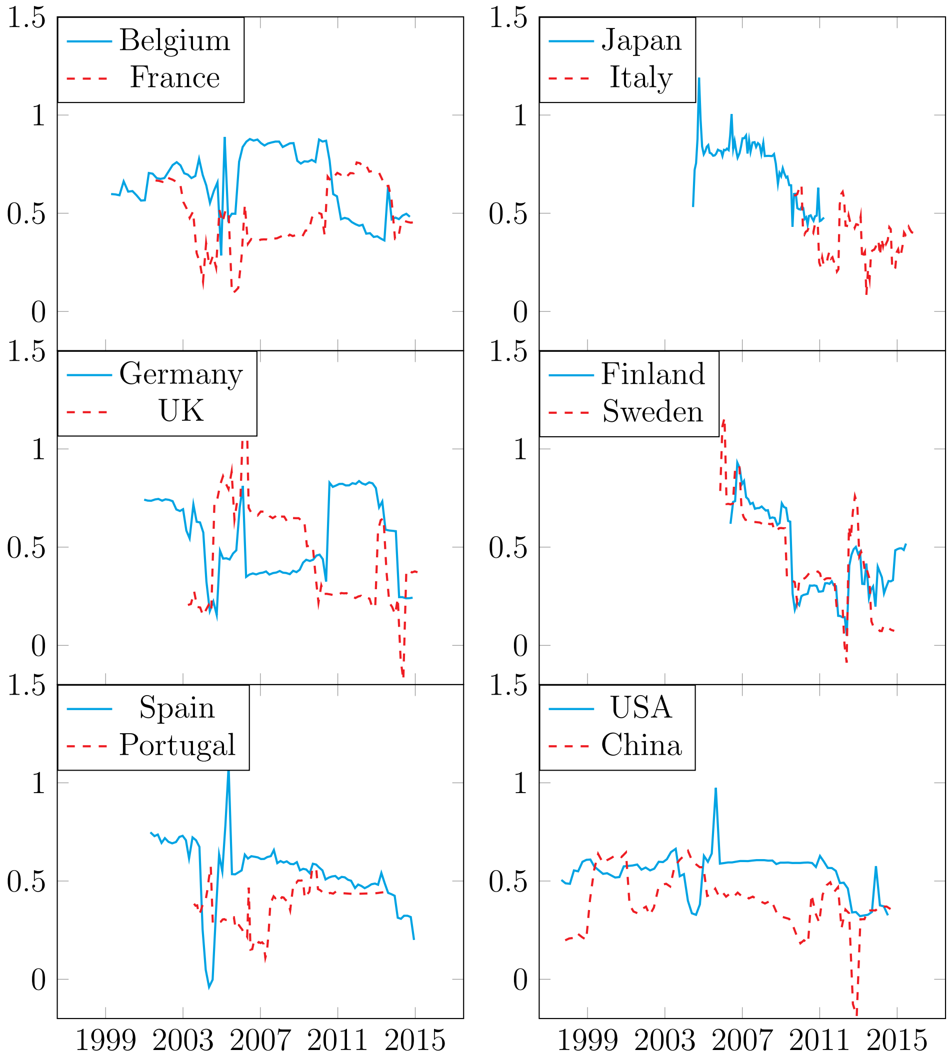


Figure 3.2: Estimated time-varying long memory parameter of realized volatility of several national stock indices over the years 1999 − 2016.

is generated is not stationary, but changing. This in turn may influence the long memory parameter.

Jensen and Whitcher (2014) reason along these lines and thus analyze the Deutsche Mark/Dollar foreign exchange rates. They find time-varying long memory and can even relate the form of the estimated function $\hat{d}(u)$ to historical dates. Specifically, they explain the estimated curve with events like the Russian crisis in 1993 or the scheduled announcement of macroeconomic data of the US.

We use daily data on major stock indices and estimate the long memory parameter $\hat{d}(u)$ of the realized volatility by the wavelet local Whittle estimator for the years 1999 to 2016. The results can be found in Figure 3.2. We use Coiflet wavelets with $M = 6$ vanishing moments and use the scales 10 to 14 which corresponds to the 5 largest scales. Other choices of the scales yield similar results.

Overall, we find values of the long memory parameter that are slightly in the non-stationary region $d > 0.5$ most of the time. At some points in time the volatility appears even to be antipersistent for a short period. One may have expected to see effects of the financial crisis 2008 on the long memory parameter. Unfortunately, we are not able to undoubtedly identify these effects on the long memory parameter for every country. Indeed, some countries seem to experience relatively stable long memory parameters for the realized volatility of their stock indices. Interestingly, the long memory parameter of the Scandinavian countries Finland and Sweden show very similar patterns. This may point to some form of cointegration of the long memory parameter.

3.6 Conclusion

In this paper we introduce a new Whittle-type estimator for time-varying long memory of locally stationary processes. The estimator utilizes wavelets that are well-suited due to their time-frequency resolution and can deal with long memory in the non-stationary region of $d > 1/2$. We prove its consistency and derive its limiting distribution. A small Monte Carlo study shows that the wavelet Whittle estimator has a better MSE than the wavelet regression estimator of Roueff and Von Sachs (2011) in many cases. Lastly, we have shown how to apply the estimator to analyze time-varying features of realized volatility time series of common stock indices.

3.7 Appendix

As we prove pointwise convergence to $d_0(u)$ for any $u \in [0,1]$ let u be fixed in this whole Section. Moreover, in the following we require Assumptions 1-3 to hold.

Write the contrast process as

$$\tilde{L}_\ell(d) = L_\ell(d) + E_\ell(d) + \log(|\ell|\sigma^2 2^{2d_0(u)\langle \ell \rangle}), \tag{3.15}$$

where $L_\ell(d)$ is defined by Equation (47) of Moulines et al. (2008) and

$$E_\ell(d) = \log\left[1 + \sum_{(j,k)\in\ell} \frac{2^{2(d_0(u)-d)}}{\sum_\ell 2^{2(d_0(u)-d)}}\left(\frac{\gamma_{j,T}(k)W^2_{j,k;T}}{\sigma^2 2^{2d_0(u)j}} - 1\right)\right] \tag{3.16}$$

with $\sigma^2 = f^*(u,0)\mathbf{K}(d_0(u))$ and $\mathbf{K}$ specified in (3.9). We know about the behavior of L_ℓ from Proposition 6 of Moulines et al. (2008). For the study of E_ℓ we will modify their Proposition 7. Introduce some notation for this task. For $\rho > 0$ and $q \geq 0$ the set of real-valued sequences $\{\mu_j\}_{j\geq 0}$ is defined as

$$\mathcal{B}(\rho, q) = \left\{\{\mu_j\}_{j\geq 0} \colon |\mu_j| \leq \rho(1 + j^q) \text{ for all } j \geq 0\right\}.$$

Additionally we define for any $u \in [0,1]$, any $T \geq 1$, any sequence $\{\mu_j\}_{j\geq 0}$ and for scales $0 \leq j_0 \leq j_0 + l \leq J_T$,

$$\tilde{S}_{u,T,j_0,l}(\mu) = \sum_{j=j_0}^{j_0+l} \mu_{j-j_0} \sum_{k=0}^{T_j - 1}\left[\frac{\gamma_{j,T}(k)W^2_{j,k;T}}{\sigma^2(u)2^{2d_0(u)j}} - 1\right].$$

Lemma 3.1. *For any $u \in [0,1]$ and any $q \geq 0$, there exists $C > 0$ such that for all $\rho \geq 0, T \geq 1$ and $j_0 = 1, \dots, J_T - l$,*

$$\left(E \sup_{\mu \in \mathcal{B}(\rho,q)} \sup_{j_0=1,\dots,J_T-l} |\tilde{S}_{u,T,j_0,l}(\mu)|^2 \right)^{1/2} \leq C\rho T 2^{-j_0} \left[(T2^{-j_0})^{-1/2} + 2^{-\beta j_0} \right],$$

if we assume that

$$2^{\{3+2(p-d_0(u))\}j_0} T^{-2} \delta_{j_0,T}^{-2} \to 0.$$

Proof. This result mainly follows using the same arguments as Moulines et al. (2008) in the proof of their Proposition 7 combined with an application of their so-called Condition 1. Technically we must approximate $\hat{\sigma}_{j,T}(u)$ by $\tilde{\sigma}^2_{j,T}(u)$ and bound the approximation error. We set $\rho = 1$ without loss of generality and write

$$\begin{aligned} \tilde{S}_{u,T,j_0,l}(\mu) =& \sum_{j=j_0}^{j_0+l} \frac{\sigma_j^2(u)}{\sigma^2 2^{2d_0(u)j}} \mu_{j-j_0} \left[\frac{\hat{\sigma}^2_{j,T}(u)}{\sigma_j^2(u)} - T_j \right] + \sum_{j=j_0}^{j_0+l} T_j \mu_{j-j_0} \left[\frac{\sigma_j^2(u)}{\sigma^2 2^{2d_0(u)j}} - 1 \right] \\ =& \tilde{S}^{(1)}_{u,T,j_0,l}(\mu) + \tilde{S}^{(2)}_{u,T,j_0,l}(\mu). \end{aligned}$$

The last term can be treated as in Proposition 7 of Moulines et al. (2008) to obtain a bound on its absolute value. We have

$$\sup_{j \geq 1} 2^{\beta j} \left| \frac{\sigma_j^2(u)}{\sigma^2 2^{2d_0(u)j}} - 1 \right| = \sup_{j \geq 1} 2^{\beta j} \left| O(2^{-\beta j}) \right| < \infty$$

by using (3.10). Using $T_j \leq T2^{-j}$ this gives

$$\sup_{\mu \in \mathcal{B}(\rho,q)} \sup_{j_0=1,\dots,J_T-l} |\tilde{S}^{(2)}_{u,T,j_0,l}(\mu)| \leq CT \sum_{j=j_0}^{j_0+l} (1 + (j-j_0)^q) 2^{-j(1+\beta)} = O(T2^{-(1+\beta)j_0}). \quad (3.17)$$

When focusing on the expectation of the first term we use the Minkowski inequality so it suffices to analyze the inner term. One finds

$$E\left[\left| \frac{\hat{\sigma}^2_{j,T}(u)}{\sigma_j^2(u)} - T_j \right|^2 \right] = \mathrm{Var}\left(\frac{\hat{\sigma}^2_{j,T}(u)}{\sigma_j^2(u)} \right) + E\left[\frac{\hat{\sigma}^2_{j,T}(u)}{\sigma_j^2(u)} - T_j \right].$$

To bound the variance we use the result from Theorem 1 of Roueff and Von Sachs (2011). But, instead of directly using this result we show that another result from Moulines et al. (2008) holds and prove that their so-called Condition 1 holds. This allows us to pursue the path of the proof without too many changes. Using Equation (3.11)

$$\sup_{T \geq 1} \sup_{j=1,\dots,J_T} (1 + T_j 2^{-2j\beta})^{-1} T_j^{-1} \mathrm{Var}\left(\frac{\hat{\sigma}^2_{j,T}(u)}{\sigma_j^2(u)} \right) \leq \sup_{T \geq 1} \sup_{j=1,\dots,J_T} (1 + T_j 2^{-2j\beta})^{-1} T_j^{-1} E\left[\left(\frac{\hat{\sigma}^2_{j,T}(u)}{\sigma_j^2(u)} - \sigma^2 \right)^2 \right] < \infty.$$

Overall, this gives

$$\left(E \sup_{\mu\in\mathcal{B}(\rho,q)} \sup_{j_0=1,\ldots,J_T-l} |\tilde{S}^{(1)}_{u,T,j_0,l}(\mu)|^2\right)^{1/2} \leq$$
$$CT\sum_{j=j_0}^{j_0+l}(1+(j-j_0)^q)\left[(T2^{-j})^{-1/2}+T2^{-(1+\beta)j}\right] =$$
$$O((T2^{-j_0})^{-1/2}+T2^{-(1+\beta)j_0}).$$

Combining (3.17) and (3.7) gives the result. □

The following Corollary summarizes the behavior of E_ℓ and is obtained by applying Lemma 3.1. The proof is virtually the same as the proof of Corollary 8 of Moulines et al. (2008) and is thus omitted.

Corollary 1. *Let $\{L_T\}$ be a sequence such that $2^{\{3+2(p-d_0(u))\}j_0}T^{-2}\delta^{-2}_{j_0,T} + L_T^{-1} + (T2^{-L_T})^{-1} \to 0$ as $T\to\infty$ and let $E_\ell(d)$ be defined as in (3.16). Under Condition 1 we have as $T\to\infty$ for any $l\geq 0$:*

$$\sup_{d\in\mathbb{R}}|E_{\ell_T(L_T,L_T+l)}(d)| = O_p\left((T2^{-L_T})^{-1/2}+2^{-\beta L_T}\right).$$

3.7.1 Consistency

In a first step the consistency of the estimator is proved with a suboptimal rate. This result is later used to prove the optimal rate. Consider the decomposition of the contrast in Equation (3.15) again. One has

$$0 \geq \tilde{L}_\ell(\hat{d}_\ell) - \tilde{L}_\ell(\hat{d}_0) = L_\ell(\hat{d}_\ell) + E_\ell(\hat{d}_\ell) - E_\ell(\hat{d}_0).$$

The proof of the following Proposition follows from Proposition 6 of Moulines et al. (2008) which shows that the function $\tilde{L}(d)$ behaves as $(d-d_0(u))^2$ up to a multiplicative positive constant and from our Lemma 3.1 resp. Corollary 1.

Lemma 3.2. *Let $\{L_T\}$ be a sequence such that $L_T^{-1}+(T2^{-L_T})^{-1}\to 0$ as $T\to\infty$. Under Condition 1 we have as $T\to\infty$ and for any $l\geq 0$,*

$$|\hat{d}_{\ell_T(L_T,L_T+l)} - d_0(u)| = O_p\left((T2^{-L_T})^{-1/4}+2^{-\beta L_T/2}\right).$$

Proof. The result follows directly using the same arguments as in Step 1 of the proof of Proposition 9 of Moulines et al. (2008). □

Proof of Theorem 3.2. The proof is mainly identical to the proof of Theorem 3 of Moulines et al. (2008) after noticing Lemma 3.1.

However, as we will make use of certain equations in the proof of the central limit theorem below we state them here. As the derivative of the objective function L_ℓ at the estimate $\hat{d}_\ell$ is zero we find that the following function $\hat{S}_\ell(d_0(u))$ is zero, too:

$$\hat{S}_\ell(d_0(u)) = \sum_{(j,k)\in\ell}(j-\langle\ell\rangle)\,2^{-2jd_0(u)}\gamma_{j,T}(k)W^2_{j,k;T}. \tag{3.18}$$

A Taylor expansion of $\hat{S}_\ell$ around $d = \hat{d}_\ell$ yields for some $\tilde{d}_T$ between $d_0(u)$ and $\hat{d}_\ell$ that

$$\hat{S}_\ell(d_0(u)) = 2\log(2)\left(\hat{d}_\ell - d_0(u)\right)\sum_{(j,k)\in\ell}(j - \langle\ell\rangle)\, j\, 2^{-2j\tilde{d}_T}\gamma_{j,T}(k)W^2_{j,k;T}.$$

Following the lines of the proof of Moulines et al. (2008) one has to bound $\hat{S}_\ell(d_0(u))$ from above and verify that the sum has a strictly positive limit after normalization. In particular, one can find the bound

$$\sum_{(j,k)\in\ell}(j - \langle\ell\rangle)\, j\,\frac{\gamma_{j,T}(k)W^2_{j,k;T}}{2^{2\tilde{d}_T j}} = (T2^{-L_T})\left\{f^*(u,0)\mathbf{K}(d_0(u))\left(2 - 2^{-l}\right)\kappa_l + o_p(1)\right\}, \quad (3.19)$$

where κ_l is defined in (3.12). Since $\kappa_l > 0$ for any integer $l \geq 1$ this gives the required positive lower bound. □

3.7.2 Central Limit Theorem

Proof of Theorem 3.3. Using (3.18) and (3.19) we can write

$$(T2^{-L_T})^1(\hat{d}_\ell - d_0(u)) = \frac{\hat{S}_T(d_0(u))}{2\log(2)f^*(u,0)\mathbf{K}(d_0(u))(2-2^{-l})\kappa_l}(1 + o_p(1)).$$

In order to simplify notation we define $d_0 = d_0(u)$ and write

$$\hat{S}_\ell(d_0) = \hat{S}_\ell(d_0) - \tilde{S}_\ell(d_0) + E[\tilde{S}_\ell(d_0)] + \tilde{S}_\ell(d_0) - E[\tilde{S}_\ell(d_0)],$$

where $\tilde{S}_\ell(d_0)$ is defined as $\hat{S}_\ell(d_0)$, but with $W^2_{j,k;T}$ replaced by $W^2_{j,k}(u)$.

Analyzing the equation termwise and since $\sum_{(j,k)\in\ell_n}(j - \langle\ell_n\rangle) = 0$ and $E[W^2_{j,k}(u)] = \sigma^2_j(u)$ we can see that

$$\begin{aligned}E[\tilde{S}_\ell(d_0)] &= \sum_{(j,k)\in\ell_T}(j - \langle\ell_T\rangle)2^{-2d_0 j}\sigma^2_j(u)\\ &= f^*(u,0)\mathbf{K}(d_0)\sum_{(j,k)\in\ell_T}(j - \langle\ell_T\rangle)O(2^{-\beta j})\\ &= O(2^{-\beta L_T}) = o(1),\end{aligned}$$

where we have used (3.10) in the second line.

Turning to the term $\hat{S}_\ell(d_0) - \tilde{S}_\ell(d_0)$ we write

$$\begin{aligned}\hat{S}_\ell(d_0) - \tilde{S}_\ell(d_0) &= \sum_{(j,k)\in\ell_T}(j - \langle\ell_T\rangle)2^{-2jd_0}\left[W^2_{j,k;T} - W^2_{j,k}(u)\right]\\ &= \sum_{j=L_T}^{L_T+l}(j - \langle\ell_T\rangle)\frac{\hat{\sigma}^2_{j,T}(u) - \tilde{\sigma}^2_{j,T}(u)}{2^{2jd_0}}.\end{aligned}$$

Noticing Proposition 1 of Roueff and Von Sachs (2011) we can use Markov's, Lyapunov's and Minkowski's inequaltiy to find that

$$\hat{S}_\ell(d_0) - \tilde{S}_\ell(d_0) = O_p(2^{(3/2+p-d_0)L_T}T^{-1}\delta^{-1}_{L,T}) = o_p(1).$$

Hence, we may apply Lemma 12 of Moulines et al. (2008) in order to obtain the limiting distribution of the numerator if we verify its prerequisites. To do so we notice that our problem can be written as

$$\tilde{S}_\ell(d_0) = \xi_T^T A_T \xi_T, \quad \text{where} \quad \xi_T = \left[|j - \langle \ell_T \rangle|^{1/2}\, 2^{-d_0 j}\, \gamma_{j,T}(k)\, W_{j,k}(u) \right]_{(j,k)\in \ell_T}$$

and A_T is the diagonal matrix with entries $\delta_{L,T}^{-1/2}\,\text{sign}(j - \langle \ell_n \rangle)$ for all $(j,k)\times(j,k) \in \ell_T$. Furthermore, we denote by Γ_T the covariance matrix of ξ_T. Thence, to apply the Lemma 12 it suffices to show that firstly

$$\rho(A_T)\rho(\Gamma_T) \to 0, \qquad \text{as} \qquad T \to \infty,$$

where $\rho(A)$ is the spectral radius of the square matrix A, and that secondly $\delta_{L,T}^{-1}$ $\text{Var}(\tilde{S}_\ell(d_0))$ has a finite limit.

Obviously, $\rho(A_T) = \delta_{L,T}^{-1/2}$ so we turn to the covariance matrix and find

$$\rho(\Gamma_T) \le \sum_{j=L_T}^{L_T+l} |j - \langle \ell_T \rangle|\; 2^{-2d_0 j} \max_{0\le i\le l} \delta_{L+i,T}^2\; 2\pi \sup_{\lambda\in(-\pi,\pi)} |D_{j,0}(\lambda;\nu)|.$$

Now, from Theorem 1 of Moulines et al. (2007b) the supremum is of order $O(2^{2d_0 L_T})$. Using Lemma 3.3 we can deduce that

$$\rho(A_T)\rho(\Gamma_T) = O(\delta_{L,T}^{3/2}) = o(1),$$

from Assumption 3.

It remains to calculate the limiting variance. Using Theorem 2 of Roueff and Von Sachs (2011) we can introduce the following quantity that converges under the stated assumptions

$$c_T(j,p) = 2^{-4d_0 j} \delta_{L,T}^{-1/2}\, \text{Cov}(\tilde{\sigma}_j^2, \tilde{\sigma}_{j-p}^2) \;\to\; (f^*(u,0))^2\, \Sigma_{j,p}(u).$$

We obtain

$$\delta_{L,T}^{-1}\,\text{Var}(\hat{S}_T(d_0)) = \sum_{i=0}^{l} (i + L_T - \langle \ell_L \rangle)^2\; c_T(L+i,0) +$$
$$2\sum_{i=0}^{l} \sum_{p=1}^{i} (i + L_T - \langle \ell_L \rangle)\,(i - p + L_T - \langle \ell_L \rangle)\; 2^{2d_0 p}\; c_T(L+i,u)$$

By Lemma 13 of Moulines et al. (2008) $\langle \ell_L \rangle - L_T \to \eta_l$ so that the limiting variance (3.13) follows. The claim then follows from applying Lemma 12 of Moulines et al. (2008). □

Lemma 3.3. *For large enough T we have*

$$\sup_{j=1,\dots,J_T} \langle \ell_T(j, j+l) \rangle < L_T + 1$$

Proof. This follows directly from Lemma 13 of Moulines et al. (2008) because $T 2^{-L_T} \to \infty$. □

Chapter 4

Testing for Multiple Structural Breaks in Multivariate Long Memory Time Series

Co-authored with Kai Wenger and Philipp Sibbertsen.

4.1 Introduction

The problem of testing for structural changes and estimating break points that occur at unknown dates has long been discussed in the econometric literature. Most of the literature has focused on issues about estimating and testing of single structural breaks in a univariate time series regression framework with weak correlations. A review of this literature can be found for instance in Perron (2006).

Bai and Perron (1998) extended this literature by suggesting estimators and tests for multiple break points that occur at unknown dates in a univariate time series regression. Bai (1997b) considered estimation of a single break in a multivariate regression set-up and Bai, Lumsdaine, et al. (1998) provide tests and estimators for common breaks in a multivariate system of short-memory time series. Qu and Perron (2007) provide a versatile framework for estimating and testing multiple and not necessarily common breaks that occur at unknown dates in a multivariate short-memory time series regression framework. They allow for breaks in the mean as well as in the covariance of the system. The estimators and tests of Qu and Perron (2007) are based on a likelihood ratio approach.

Testing and estimating structural breaks in long-memory time series is problematic as both phenomena are observationally equivalent in finite samples and long memory can cause false rejections of tests for structural changes. An overview about the literature regarding this problem gives for instance Sibbertsen (2004). Nevertheless, recently some approaches are published to test for a single structural break in a univariate long-memory time series model. Among those are L. Wang (2008), Shao (2011), Dehling et al. (2013), Iacone et al. (2014), Betken (2016), and Wenger and Leschinski (in press). A recent overview is provided in Wenger, Leschinski, and Sibbertsen (2019). Estimation of multiple breaks in a univariate set-up allowing also for long-range dependence has been considered in Lavielle and Moulines (2000) by applying information criteria.

This paper contributes to the literature by considering estimators and tests for multiple structural breaks that occur at unknown dates in a multivariate long-memory time series regression framework. To the best of our knowledge this is the first paper providing tests for multiple breaks under long memory and the first paper considering breaks in a multivariate system of long-memory time series. We extend the general framework of Qu and Perron (2007) in two directions. First, we use a likelihood ratio based approach for estimating breaks in the mean and the covariance of the system. We obtain consistency and the limiting distribution of these estimates under long memory. Second, we provide tests on multiple structural changes generalizing the testing ideas of Bai and Perron (1998). The tests of Bai and Perron (1998) are based on segmentation of the time series and repeated testing for breaks within these segments. The limiting distribution strongly depends on the assumption of at most weak correlations as it is derived as the product of the limiting distributions for each segment. This does not hold true under long-range dependence as the segments are strongly correlated and the limiting distribution of the test proves wrong in this situation. We circumvent this problem by suggesting to repeatedly test for breaks on the residuals after applying our consistent break point estimator and eliminate the largest break in each step. It turns out that all of our procedures only depend on the maximal memory parameter of the multivariate system of long-memory time series. Interestingly, the limiting distribution of our test is different for the case where all memory parameter are equal compared to the case where at least two of them are not equal.

In order to prove our results we derive a multivariate generalized Hájek-Rényi-type inequality under long-range dependence. The validity of this approach in finite samples is shown in a Monte Carlo study, while the applicability in practice is demonstrated in an empirical example where we examine a system of inflation series.

The rest of the paper is organized as follows. In Section 4.2 we provide the model and our assumptions. Section 4.3 contains the estimators for the break points and Section 4.4 provides the testing procedure. Section 4.5 contains of the Monte Carlo study and Section 4.6 illustrates the empirical example before Section 4.7 concludes. All proofs are gathered in the appendix.

4.2 Model

In this paper we consider issues regarding the detection of structural changes in a multivariate regression model allowing for long-memory errors. An n dimensional system of time series u_t is said to exhibit multivariate long-range dependence or long memory with $D = (d_1, \ldots, d_n)'$ and $-1/2 < d_i < 1/2$ for $i = 1, \ldots, n$ if its spectral density behaves local to the origin as

$$f(\lambda) \sim \Lambda(D) G \Lambda(D)^*,$$

where $\Lambda(D) = \text{diag}(\Lambda_1(d_1), \ldots, \Lambda_n(d_n))$ and $\Lambda_k(d_k) = \lambda^{-d_k} e^{i(\pi-\lambda)d_k/2}$ for $k = 1, \ldots, n$. G is a real, positive definite, finite and symmetric matrix and the asterix A^* denotes the complex conjugate of the matrix A. Further, the imaginary number is denoted by i and d_k is the memory parameter of series k. Furthermore, define $d = \max\{d_1, \ldots, d_n\}$. The assumption on G is standard in defining multivariate long memory and excludes fractional cointegration as it stands. However, for our estimators and tests proposed later it is of no relevance whether or not the series are fractionally cointegrated. We therefore stick to the standard assumption keeping in mind that relaxing the assumptions on G would not effect our procedures.

For the regression model consider a system of n time series each of length T. We denote by m the total number of structural changes in the system. The break dates in the system are denoted by the m vector $\mathcal{T} = (T_1, \dots, T_m)$ and for the ease of calculation we use $T_0 = 1$ and $T_{m+1} = T$. We use the convention that a subscript j indexes a regime ($j = 1, \dots, m+1$), a subscript t indexes a temporal observation ($t = 1, \dots, T$) and a subscript i indexes the equation ($i = 1, \dots, n$). The number of regressors is named q and z_t is the set which includes the regressors at a point in time t from all equations $z_t = (z_{1t}, \dots, z_{qt})'$. Consider the model

$$y_t = (I \otimes z_t')S\beta_j + u_t, \tag{4.1}$$

where u_t is the error process to be specified more precisely below with mean 0 and covariance matrix Σ_j for $T_{j-1} + 1 \le t \le T_j$ ($j = 1, \dots, m+1$). The matrix S is a selection matrix with entries 0 and 1. It is of dimension $nq \times p$ with full column rank. In regime j the parameters to be estimated are given by the p vector β_j and the matrix Σ_j. Restrictions on the parameters should be allowed by our model so we introduce r restrictions given by

$$g(\beta, \mathrm{vec}(\Sigma)) \;=\; 0,$$

where $\beta = (\beta_1', \dots, \beta_{m+1}')$, $\Sigma = (\Sigma_1, \dots, \Sigma_{m+1})$ and $g(\cdot)$ is an r-dimensional vector. This setting is even capable of expressing cross-equations restrictions across regimes.

We rewrite Equation (4.1) in order to lighten the notation by introducing the $p \times n$ matrix x_t that is defined by $x_t' = (I \otimes z_t')S$. This gives

$$y_t = x_t'\beta_j + u_t \tag{4.2}$$

for $T_{j-1} + 1 \le t \le T_j$ ($j = 1, \dots, m+1$). Furthermore, we rewrite Equation (4.2) using matrix notation. To this end let $Y = (y_1', \dots, y_T')'$ be the nT vector of dependent variables, $U = (u_1', \dots, u_T')'$ the error vector and let the $nT \times p$ matrix of regressors be $X = (x_1, \dots, x_T)'$. Now form the block partition $\bar{X}$ of the matrix X: For a given partition of the sample using the breaks $(T_1, \dots, T_m)$ we define $\bar{X}$ as the $nT \times p(m+1)$ matrix $\bar{X} = \mathrm{diag}(X_1, \dots, X_{m+1})$, where $X_j (j = 1, \dots, m+1)$ is the $n(T_j - T_{j-1}) \times p$ subset of X that corresponds to observations in regime j. Similarly we define the subvector U_j of U. Using these symbols we can express the regression system (4.2) as $Y = \bar{X}\beta + U$. In the following the true values of the parameters are denoted with a 0 superscript, e.g. the data generating process is given by $Y = \bar{X}^0\beta^0 + U$. Here, the term $\bar{X}^0$ is the diagonal partition of X using the partition of true break dates $(T_1^0, \dots, T_m^0)$.

We impose the following set of assumptions. Note that the assumptions are similar to those of Qu and Perron (2007) with the difference that we allow the errors u_t to be long-range dependent. However, we assume that the memory is only introduced through the errors u_t and thus that the regressors x_t are mostly short-range dependent such that the process $x_t u_t$ is of the same order of integration as u_t. This is a simplyfying assumption which is not necessary. Technically it would be possible to allow also for long-memory regressors with an order of integration smaller or equal than one as long as they are independent of the errors. However, this would complicate the proof of our property five below and furthermore lead to identification problems in practice as the order of integration of the observation would be determined by the maximum of the integration order of the regressors and errors. As the proof of our property five in the case of integrated regressors rely on a correct differencing of the regressors our procedure may become infeasible in practice.

Assumption 1. For each $j = 1, \dots, m+1$ and $l_j \leq T_j^0 - T_{j-1}^0$, $l_j^{-1} \sum_{t=T_{j-1}^0+1}^{T_{j-1}^0+l_j} x_t x_t' \xrightarrow{\text{a.s.}} Q_j^0$ as $l_j \to \infty$, with Q_j^0 being a nonrandom positive definite matrix not necessarily the same for all j.

Assumption 2. There exists an $l_0 > 0$ such that for all $l > l_0$, the minimum eigenvalues of $l^{-1} \sum_{j=T_j^0+1}^{T_j^0+l} x_t x_t'$ and of $l^{-1} \sum_{t=T_j^0-l}^{T_j^0} x_t x_t'$ are bounded away from zero $(j = 1, \dots, m)$.

Assumption 3. The matrix $\sum_{t=k}^{l} x_t x_t'$ is invertible for $l - k \geq k_0$ for some $0 < k_0 < \infty$.

Assumption 4. It holds that

$$u_t = A(L)\varepsilon_t = \sum_{j=0}^{\infty} A_j \varepsilon_{t-j},$$

where the innovations $\varepsilon_t = (\varepsilon_{t,1}, \dots, \varepsilon_{t,n})$ are n-dimensional martingale differences with respect to the σ-field $\mathfrak{F}_t$ generated by ε_s, $s \leq t$. Hence $E(\varepsilon_t | \mathfrak{F}_{t-1}) = 0$ and it is assumed that $E(\varepsilon_t \varepsilon_t^{'} | \mathfrak{F}_{t-1}) = I_n$ a.s. Additionally for some $\delta > 0$ we assume the moment condition $\sup_t E|\varepsilon_{t,k}|^{2+\delta} < \infty$ for $k = 1, \dots, n$. For the coefficients A_j we assume that asymptotically

$$A_j \sim \operatorname{diag}\left(\frac{j^{d_1-1}}{\Gamma(d_1)}, \dots, \frac{j^{d_n-1}}{\Gamma(d_n)}\right) \Pi, \qquad \text{as} \quad j \to \infty,$$

where Π is an $n \times n$ matrix independent of $D = (d_1, \dots, d_n)$.

Assumption 5. Assumption 4 holds with u_t replaced by $x_t u_t$ or $u_t u_t' - \Sigma_j^0$ for $T_{j-1}^0 < t \leq T_j^0$ $(j = 1, \dots, m+1)$.

Assumption 6. The magnitudes of the shifts satisfy $\beta_{T,j+1}^0 - \beta_{T,j}^0 = \nu_T \delta_j$ and $\Sigma_{j+1,T}^0 - \Sigma_{j,T}^0 = \nu_T \Phi_j$, where $(\delta_j, \Phi_j) \neq 0$ and independent of T. Moreover, ν_T is either a positive number independent of T or a sequence of positive numbers that satisfy $\nu_T \to 0$ and $T^{1/2-d} \nu_T / (\log T)^2 \to \infty$.

Assumption 7. We have $(\beta^0, \Sigma^0) \in \bar{\Theta}$ with $\bar{\Theta} = \{(\beta, \Sigma) : \|\beta\| \leq c_1, \lambda_{\min}(\Sigma) \geq c_2, \lambda_{\max}(\Sigma) \leq c_3\}$ for some $c_1 < \infty, 0 < c_2 \leq c_3 < \infty$ and $\lambda_{\min}$ and $\lambda_{\max}$ denote the smallest resp. largest eigenvalue.

Assumption 8. We have $0 < \lambda_1^0 < \dots < \lambda_m^0 < 1$ with $T_i^0 = [T\lambda_i^0]$.

Our assumptions include the standard FIVARMA model as well as long-memory panel models and regression models with exogenous regressors and long-memory errors. However, unit root regressors are ruled out by Assumption 1 although in general regressors may be trending. Moreover, the regressors can have different distributions in different regimes. This is necessary because a change in a dynamic model leads to changes in the moments of the regressors. Assumption 2 rules out the case of local collinearity which makes the breaks identifiable. Assumption 3 is a standard invertibility assumption. Assumption 4 to 5 state that we consider a long-memory regression framework and that the order of integration is solely determined by the errors u_t. We additionally assume bounded moments of order $2+\delta$ for some $\delta > 0$ for u_t, $x_t u_t$ and $u_t u_t^{'}$ to obtain strongly consistent estimates of the parameters and a well-behaved likelihood. Assumption 6 ensures that the breaks are asymptotically non-negligible. Using a fixed ν_T captures large breaks whereas a shrinking ν_T gives small and intermediate breaks in finite samples. The latter ensures an asymptotic theory for the break dates estimators which does not depend on the actual distribution of the regressors and errors. It should be noted that we assume the break size to depend on the memory of the errors. The higher the persistence of the errors is the larger the break need to be in order to be detected. Assumption 7 makes sure that the errors have a non-degenerate covariance matrix and a finite conditional mean and Assumption 8 ensures distinct breaks. It should be mentioned that no other assumptions on the breaks are needed. This includes that the breaks do not need to be contemporaneously in each series. So we allow each series to have breaks at different times or not to break at all.

Later when we introduce our testing procedure in order to derive the limiting distribution of the test under the null hypothesis of no structural change, we impose the following additional assumptions.

Assumption 9. We have $T^{-1}\sum_{t=1}^{[Ts]} x_t x_t' \xrightarrow{p} sQ$, uniformly in $s \in [0,1]$, for Q being some positive definite matrix.

Assumption 10. The errors $\{u_t\}$ form an array of long-range dependent processes as defined in Assumption 4 and, additionally, $E(u_t u_t') = \Sigma^0$ for all t and $T^{-1/2-D}\sum_{t=1}^{[Ts]} x_t u_t \Rightarrow \Phi^{1/2} W_D(s)$, where $\Phi = \text{plim}_{T\to\infty} T^{-1} X'(I_n \otimes \Sigma^0)X$ and $W_D(s)$ is a vector of independent fractional Brownian motions of type I. Also, with $\eta_t \equiv (\eta_{t1}, \ldots, \eta_{tn})' = (\Sigma^0)^{-1/2} u_t$, we have $T^{-1/2-D}\sum_{t=1}^{[Ts]}(\eta_t \eta_t' - I_n) \Rightarrow \xi_D(s)$, where $\xi_D(s)$ is an $n \times n$ matrix of fractional Brownian motion processes with $\Omega = \text{Var}(\text{vec}(\xi_D(1)))$. Also assume that $E[\eta_{tk}\eta_{tl}\eta_{th}] = 0$ for all k, l, h and for every t.

Assumption 9 rules out trending regressors and requires that the second moment matrix of the regressors converges in probability to the same limiting matrix throughout the sample. This entails we do not allow for a change in the distribution of the regressors without a change in the coefficients of the regressors. In addition, Assumption 10 requires the error process to be stable throughout the sample so that a functional central limit theorem applies to the product of regressors and errors. For a detailed discussion of fractional Brownian motions of type I and type II cf. Marinucci and Robinson (1999)

4.3 Estimation of the Break Dates and Model Parameters

We estimate the break dates and the number of breaks by restricted Quasi-Maximum Likelihood conditional on a given partition of the sample $\mathcal{T} = (T_1, \ldots, T_m)$. Our tests for the number of breaks is then based on the likelihood ratio statistic. Assuming Gaussian serially uncorrelated errors the quasi-likelihood function is given by

$$L_T(\mathcal{T},\beta,\Sigma) = \prod_{j=1}^{m+1} \prod_{t=T_{j-1}+1}^{T_j} f(y_t|x_t;\beta_j,\Sigma_j),$$

where

$$f(y_t|x_t;\beta_j,\Sigma_j) = \frac{1}{(2\pi)^{n/2}|\Sigma_j|^{1/2}} \exp\left(-\frac{1}{2}[y_t - x_t'\beta_j]'\Sigma_j^{-1}[y_t - x_t'\beta_j]\right).$$

The quasi-likelihood ratio statistic is given by

$$LR_T(\mathcal{T},\beta,\Sigma) = \frac{\prod_{j=1}^{m+1}\prod_{t=T_{j-1}+1}^{T_j} f(y_t|x_t;\beta_j,\Sigma_j)}{\prod_{j=1}^{m+1}\prod_{t=T_{j-1}^0+1}^{T_j^0} f(y_t|x_t;\beta_j^0,\Sigma_j^0)}.$$

We aim now to estimate the values of $(T_1,\ldots,T_m,\beta,\Sigma)$ under the restriction $g(\beta,\text{vec}(\Sigma)) = 0$. This is done by maximizing the objective function

$$RLR_T(\mathcal{T},\beta,\Sigma) = LR_T(\mathcal{T},\beta,\Sigma) + \lambda' g(\beta,\text{vec}(\Sigma)). \tag{4.3}$$

We need one further assumption about the minimal regime length.

Assumption 11. The maximization of the objective function (4.3) is taken over all partitions $\mathcal{T} = (T_1,\ldots,T_m) = (T\lambda_1,\ldots,T\lambda_m)$ for some $\varepsilon > 0$ in the set

$$\Lambda_\varepsilon = \{(\lambda_1,\ldots,\lambda_m)\colon |\lambda_{j+1}-\lambda_j| \geq \varepsilon, \lambda_1 \geq \varepsilon, \lambda_m \leq 1-\varepsilon\}.$$

This assumption is standard in the structural breaks literature and says that some percentage of the data needs to be skipped at the beginning and the end of the observation period before the maximization of the likelihood and thus that potential breaks cannot happen in a possible small environment of the first and the last observation. Other than in Qu and Perron (2007) this assumption is essential for our procedure to work as property 2 in the appendix and therefore the consistency of the breakpoint estimators proves wrong otherwise. Qu and Perron (2007) prove this property and consistency of the estimator when maximizing over the whole sample by means of the standard law of iterated logarithm. As this does no longer hold under long memory and needs to be replaced by a law of iterated logarithm for fractional Brownian motions the arguments used to prove property 2 do no longer hold and the property does not apply. However, these arguments are needed for the endpoints only and therefore assuming assumption 11 circumvents this problem.

We can now establish the rate of convergence of these estimators under long-range dependencies.

Lemma 4.1. Under Assumptions 1 to 8 and 11 we have for $j = 1,\ldots,m$, $T^{1-2d}\nu_T^2(\hat{T}_j - T_j^0) = O_P(1)$ and for $j = 1,\ldots,m+1$, $T^{1/2-d}(\hat{\beta}_j - \beta_j^0) = O_P(1)$ and $T^{1/2-d}(\hat{\Sigma}_j - \Sigma_j^0) = O_P(1)$.

The proof of this and all following results can be found in the appendix. These results are similar as those in Bai (1997b), Bai and Perron (1998), Bai (2000), and Qu and Perron (2007), but account for the long-range dependencies in the error terms. Also in our case the rate for the break dates is fast enough not to effect the estimation of the model parameter asymptotically. Therefore, we have the following result that we state without proof.

Lemma 4.2. Under the Assumptions of Lemma 4.1, the limiting distribution of $T^{1/2-d}(\hat{\beta} - \beta^0)$ is the same as that for known break dates.

These results are necessary to our tests on the number of potential break points later. However, it allows us also to derive results regarding the limiting distribution of the restricted likelihood under long memory. We can now split the restricted likelihood in one part only containing the break dates and the true parameter values so that restrictions to these values do not affect the estimation of the break dates. The other part involves the true values of the break dates and model parameters and the restrictions such that the limiting distribution of the model parameters is affected by these restrictions but not by the estimation of the break dates. With these comments in mind it is obvious that Theorem 1 of Qu and Perron (2007) still holds under our set of assumptions, where the aforementioned split of the maximization problem in a term concerning the estimate of the break dates and a term that does not involve the break date estimates is made mathematically precise.

Moreover we would be able to show that Theorem 2 of Qu and Perron (2007) still holds under long-range dependence. This result concerns the limiting distribution of the break dates. The drawback of this result is that the the limiting distribution of the break dates depends on the true error distribution. This is a standard problem in the structural breaks literature and is usually accounted for by assuming shrinking breaks with an increasing sample size. However, to do so trending regressors need to be ruled out.

Assumption 12. Let $\Delta T_j^0 = T_j^0 - T_{j-1}^0$. For $j = 1, \ldots, m$, as $\Delta T_j^0 \to \infty$, uniformly in $s \in [0, 1]$, $(\Delta T_j^0)^{-1} \sum_{t=T_{j-1}^0+1}^{T_{j-1}^0+[s\Delta T_j^0]} x_t x_t' \xrightarrow{P} sQ_j^0$ with Q_j^0 being a nonrandom positive definite matrix not necessarily the same for all j.

With this assumption we obtain the following limiting distribution for the break dates.

Theorem 4.1. Let $\eta_t = (\eta_{t1}, \ldots, \eta_{tn}) = (\Sigma_j^0)^{-1/2} u_t$ for $t \in [T_{j-1}^0 + 1, T_j^0]$ and assume that $E[\eta_{tk}\eta_{tl}\eta_{th}] = 0$ for all k, l, h and for every t. Under Assumptions 1 to 8 and 11 and 12 with $\nu_t \to 0$ such that $T^{1/2-d}\nu_T/(\log T)^2 \to \infty$ as $T \to \infty$ and with $\Rightarrow$ denoting weak convergence under the Skorohod topology, we have, for $j = 1, \ldots, m$

$$\frac{\Delta_{1,j}^2}{\Gamma_{1,j}^2} T^{1-2d} \nu_T^2 (\hat{T}_j - T_j^0) \Rightarrow \begin{cases} -\frac{|u|}{2} + W_{j,d}(u), & \text{for } u \leq 0 \\ -\frac{|u|}{2} \frac{\Delta_{2,j}}{\Delta_{1,j}} + \frac{\Gamma_{2,j}}{\Gamma_{1,j}} W_{j,d}(u), & \text{for } u > 0, \end{cases}$$

where

$$\begin{aligned}
\Delta_{1,j} =& \frac{1}{2}\,\mathrm{tr}(A_{1,j}^2 + \delta' Q_{1,j}\delta_j),\\
\Delta_{2,j} =& \frac{1}{2}\,\mathrm{tr}(A_{2,j}^2 + \delta' Q_{2,j}\delta_j),\\
A_{1,j} =& (\Sigma_j^0)^{1/2}(\Sigma_{j+1}^0)^{-1}\Phi_j(\Sigma_j^0)^{-1/2}\\
A_{2,j} =& (\Sigma_{j+1}^0)^{1/2}(\Sigma_j^0)^{-1}\Phi_j(\Sigma_{j+1}^0)^{-1/2}\\
\Gamma_{1,j} =& \left(\frac{1}{4}\,\mathrm{vec}(A_{1,j})'\Omega_{1,j}^0\,\mathrm{vec}(A_{1,j} + \delta_j'\Pi_{1,j}\delta_j\right)^{1/2}\\
\Gamma_{2,j} =& \left(\frac{1}{4}\,\mathrm{vec}(A_{2,j})'\Omega_{2,j}^0\,\mathrm{vec}(A_{2,j} + \delta_j'\Pi_{2,j}\delta_j\right)^{1/2}\\
Q_{1,j} =& \mathrm{plim}_{T\to\infty}(T_j^0 - T_{j-1}^0)^{-1}\sum_{t=T_{j-1}^0+1}^{T_j^0} x_t(\Sigma_{j+1}^0)^{-1}x_t'\\
Q_{2,j} =& \mathrm{plim}_{T\to\infty}(T_{j+1}^0 - T_j^0)^{-1}\sum_{t=T_j^0+1}^{T_{j+1}^0} x_t(\Sigma_j^0)^{-1}x_t',
\end{aligned}$$

and

$$\begin{aligned}
\Pi_{1,j} &= \lim_{T\to\infty}\mathrm{Var}\left\{(T_j^0 - T_{j-1}^0)^{-1/2}\left[\sum_{t=T_{j-1}^0+1}^{T_j^0} x_t(\Sigma_{j+1}^0)^{-1}(\Sigma_j^0)^{1/2}\eta_t\right]\right\},\\
\Pi_{2,j} &= \lim_{T\to\infty}\mathrm{Var}\left\{(T_{j+1}^0 - T_j^0)^{-1/2}\left[\sum_{t=T_j^0+1}^{T_{j+1}^0} x_t(\Sigma_j^0)^{-1}(\Sigma_{j+1}^0)^{1/2}\eta_t\right]\right\},
\end{aligned}$$

with $W_{j,d}(s)$ a fractional Wiener process defined on the real line and

$$\begin{aligned}
\Omega_{1,j}^0 &= \lim_{T\to\infty}\mathrm{Var}\left(\mathrm{vec}\left[(T_j^0 - T_{j-1}^0)^{-1/2}\sum_{t=T_{j-1}^0+1}^{T_j^0}(\eta_t\eta_t' - I_n)\right]\right),\\
\Omega_{2,j}^0 &= \lim_{T\to\infty}\mathrm{Var}\left(\mathrm{vec}\left[(T_{j+1}^0 - T_j^0)^{-1/2}\sum_{t=T_j^0+1}^{T_{j+1}^0}(\eta_t\eta_t' - I_n)\right]\right).
\end{aligned}$$

4.4 Testing for Multiple Breaks in Multivariate Time Series

In this section we first introduce two likelihood ratio based tests for multiple breaks in a multivariate system of long-memory time series. The first procedure tests the null of no break against the alternative of a prespecified number of breaks whereas the second tests against the alternative of an unknown number of breaks given an upper bound. Iterative application of the second procedure is one of the main ingredients of our proposed procedure to identify multiple breaks in a long-memory framework.

Our tests allow only a subset of the coefficients of the regressors β or of the covariance matrix of the errors Σ_j to change per regime j, where $1 \leq j \leq m$. We acknowledge this dependence on the specification in our test statistic by introducing the numbers p_b, n_{bd} and n_{bo}. Considering the system specification

$$y_t = x'_{at}\beta_a + x'_{bt}\beta_{bj} + u_t \qquad \text{for } T_{j-1}+1 \leq t < T_j \qquad (j = 1, \ldots, m+1),$$

p_b describes the total number of coefficients allowed to change across regimes, i.e. β_{bj} is a p_b dimensional vector. Moreover for the covariance matrix of the errors

$$\Sigma_j = E(u_t u_t') \qquad T_{j-1}+1 \leq t < T_j \qquad (j = 1, \ldots, m+1),$$

we allow n_{bd} diagonal entries of Σ_j and n_{bo} entries in the upper triangle of Σ_j to change across regimes. To simplify notation we also need the full row rank matrix H of dimension $(n_{bd}+2n_{bo}) \times n^2$. This is chosen such that $H\,\mathrm{vec}(\Sigma)$ is the $n_{bd}+2n_{bo}$ dimensional vector of the entries allowed to change. Thus, it contains both upper and lower triangle covariance entries.

First, we introduce a likelihood ratio test of no break versus the alternative hypothesis of precisely m breaks under long memory, i.e.

$$\mathcal{H}_0\colon K = 0 \qquad \text{vs} \qquad \mathcal{H}_1\colon K = m.$$

We denote the log-likelihood value by $\log \hat{L}_T(T_1, \ldots, T_m)$. Then the test is the maximal value of the likelihood ratio over all admissible partitions in the set Λ_ε defined by Assumption 11, that is,

$$\begin{aligned}\frac{1}{T^{2d}} \sup LR_T(m, p_b, n_{bd}, n_{bo}, \varepsilon) &= \frac{1}{T^{2d}} \sup_{(\lambda_1,\ldots,\lambda_m)\in\Lambda_\varepsilon} 2\left[\log \hat{L}_T(T_1, \ldots, T_m) - \log \tilde{L}_T\right] \\ &= \frac{2}{T^{2d}}[\log \hat{L}_T(\hat{T}_1, \ldots, \hat{T}_m) - \log \tilde{L}_T],\end{aligned}$$

where the log-likelihood $\log \tilde{L}_T$ is obtained by estimating β and Σ under the null hypothesis of no break. The list of estimated break points $(\hat{T}_1, \ldots, \hat{T}_m)$ contains the QMLE obtained by considering only those partitions in Λ_ε. As we assume a minimal length ε for each segment this parameter will affect the limiting distribution of the test.

Theorem 4.2. Under Assumptions 1-11 with the $\sup LR_T(m,\, p_b,\, n_{bd},\, n_{bo},\, \varepsilon)$ test constructed for an alternative hypothesis $\mathcal{H}_1$ in the class of models described in this Section,

$$\frac{1}{T^{2d}} \sup LR_T(m, p_b, n_{bd}, n_{bo}, \varepsilon) \Rightarrow \sup_{(\lambda_1,\ldots,\lambda_m)\in\Lambda_\varepsilon} \sum_{j=1}^{m} LR_j(\lambda, d, p_b, n_b^*)$$

with

$$\begin{aligned}LR_j(\lambda, d, p_b, n_b^*) =& \frac{\|\lambda_j W^*_{d,p_b}(\lambda_{j+1}) - \lambda_{j+1} W^*_{d,p_b}(\lambda_j)\|^2}{(\lambda_{j+1}-\lambda_j)\lambda_j\lambda_{j+1}} \\ &+ \frac{1}{2}\big(\lambda_j W^*_{d,n_b^*}(\lambda_{j+1}) - \lambda_{j+1} W^*_{d,n_b^*}(\lambda_j)\big)' H\Omega H' \\ &\times \big(\lambda_j W^*_{d,n_b^*}(\lambda_{j+1}) - \lambda_{j+1} W^*_{d,n_b^*}(\lambda_j)\big) / ((\lambda_{j+1}-\lambda_j)\lambda_j\lambda_{j+1}),\end{aligned}$$

where $\lambda = (\lambda_1, ..., \lambda_m)$ and $\lambda_{m+1} = 1$. The vectors $W^*_{d,p_b}(\cdot)$ and $W^*_{d,n_b^*}(\cdot)$ are of dimension p_b resp. $n_b^* = (n_{bd} + 2n_{bo})$ writing $d = (d_1, ..., d_q)$ with $q \in \{p_b, n_b^*\}$ and $n_b^* = \text{rank}(H)$. They are defined as

$$W^*_{D,n}(\cdot) = \left(W^*_{d_j}(\cdot)\right)_{j=1,\dots,n}, \qquad W^*_{d_j}(\cdot) = \begin{cases} W_{d_j}(\cdot) & \text{if } d_j = \max_{1\le i\le n} d_i, \\ 0 & \text{else,} \end{cases}$$

where W_d is univariate fractional Brownian motion of type I with memory parameter d.

Note that the limiting distribution depends on the number of series having the maximal memory parameter. Only the series in the test statistic with the maximal memory parameter in the vector of memory parameters $D = (d_1, \dots, d_n)$ contribute asymptotically to the limiting distribution.

The second test statistic tests the null hypothesis of no break against the alternative of m breaks, $1 \le m \le M$ for some upper bound M, i.e.

$$\mathcal{H}_0' \colon K = 0 \qquad \text{vs} \qquad \mathcal{H}_1' \colon 1 \le K \le M.$$

Bai and Perron (1998) suggest to use a so-called double maximum test. The test statistic is given by

$$UDmax\, LR_T(m, p_b, n_{bd}, n_{bo}, \varepsilon) = \max_{1\le m\le M} \sup LR_T(m, p_b, n_{bd}, n_{bo}, \varepsilon). \tag{4.4}$$

The asymptotic distribution for this test statistic can be obtained in the setting of Theorem 4.2. We have

$$UDmax\, LR_T(m, p_b, n_{bd}, n_{bo}, \varepsilon) \Rightarrow \max_{1\le m\le M} \sup_{(\lambda_1,\dots,\lambda_m)\in\Lambda_\varepsilon} \sum_{j=1}^{m} LR_j(\lambda, d, p_b, n_b^*).$$

Critical values for different values of d and m are given in Tables 4.2 to 4.4 in the appendix.
For our iterative procedure below it is also essential to mention that the UDmax test enjoys pitman efficiency. This follows directly by noting that the tests are likelihood ratio type tests and applying the usual Taylor expansion argument to derive consistency of likelihood ratio tests also delivers the result in our set-up.
Now we are in the position to introduce an iterative method that can be used to determine the unknown number of breaks in a multivariate system of long-memory time series. It is inspired by applying the $UDmax\, LR_T$ test in (4.4) repeatedly. Therefore, the method requires fixing an upper bound on the number of breaks M in advance. It is a residual based iterative procedure, so that we shorten it as REBIT. It proceeds as follows:

(1) Set $m = 0$.
(2) Estimate m breaks in the original system of time series y_t and save the residuals.
(3) Conduct the $UDmax\, LR_T$ test with $H_0 \colon l = 0$ vs. $H_1 \colon 1 \le l \le M - m$ on the residuals.
(4a) If the test rejects and $m < (M - 1)$: set $m = m + 1$ and reiterate from (2).
(4b) If the test cannot reject: the detected number of breaks is m. Furthermore, if $m = M - 1$: the number of breaks is greater or equal than the previously chosen upper bound M.

The method ends if the applied test cannot reject (or the user chosen upper bound is reached). Therefore, in a situation with an unknown number of breaks the iterative method suggests m breaks where m is the number returned by the method.

It should be mentioned that the $\sup LR_T$ is not applicable in the suggested procedure since a true break number k such that $k \neq 0$ and $k \neq m$ is neither covered by the null nor by the alternative hypothesis.

Note that the estimation of the break dates in step (2) is always performed on the original time series. That is the residuals are always estimated from a global optimization. Hence, the estimated break dates from different iterations do not depend on each other. Therefore, our procedure avoids the usually problematic situation of using residuals of residuals. From Lemma 4.1 we thus obtain consistency of our break point estimates in each step. The break point estimates in underspecified models are consistent as has been shown by Bai (1997a) and Bai and Perron (1998) for breaks in the mean. By similar methods one could therefore show that our procedure estimates some true break points if the number of true break points is underspecified.

Note that we do not estimate any break in step (2) if $m = 0$ (first iteration), so "saving the residuals" refers to using the original time series in the following steps.

Whereas the estimation in step (2) is done on the original system of time series, the testing in step (3) is conducted on the residuals.

This iterative procedure avoids splitting up the sample as suggested in for example Bai (1997b) which is not possible under long memory and allows us to use the limiting results in Theorem 4.1 and 4.2 which are derived under long-range dependencies. The following Theorem states that our procedure has a hit rate of $(1-\alpha)\%$ where α is the level of the break point test in Theorem 4.2.

Theorem 4.3. Let α be the significance level of the break point test in Theorem 4.2. Under Assumptions 1-11 the REBIT procedure has a hit rate of $(1 - \alpha)\%$.

The hit rate can be made converging to one by choosing the critical value of the break point test to be sample size dependent α/T. However, this is not further considered here as the sample size is given fixed in practice.

4.5 Simulation results

We conduct a Monte Carlo simulation study to examine the finite sample properties of our proposed REBIT procedure. We consider a bivariate model of fractionally integrated white noise processes

$$X_{1t} = \Delta^{d_1} u_{1t}$$
$$X_{2t} = \Delta^{d_2} u_{2t}$$

with long-memory parameters chosen as $D = (d_1, d_2) = [(0.4, 0), (0.2, 0), (0.2, 0.4)]$. To estimate D, we apply the multivariate local Whittle estimator by Shimotsu (2007) using a bandwidth of $\lfloor T^{2/3} \rfloor$. The nominal significance level is $\alpha = 5\%$, we choose $\epsilon = 0.05$, and $M = 1,000$ replications.
The asymptotic critical values we use are simulated for different combinations of $D = (d_1,d_2)$ by approximating the stochastic integrals by partial sums and can be found in Table 4.2 to 4.4 in the appendix. They are based on 10,000 Monte Carlo

replications with $1,000$ increments per path of the fractional Brownian motion. We obtain the corresponding critical value for an estimated value that is between two simulated d-values by linear interpolation between these two values.
We choose $m = 0, 1, 2, 3$ breaks which are uniformly allocated to the two series such that the distance between the breaks across both series is the same. Whether the breaks are positive or negative is randomly chosen. We make the break size dependent on the memory parameter as follows

$$\beta = \kappa T^{d-1/2},$$

where β is the break size, $d = \max(d_1, d_2)$, and κ is a finite constant. In Figure 4.1 we report the hit ratio, i.e. how often our procedure detects the true number of breaks, dependent on κ with $T = 1,000$. The case of $m = 0$ is implicitly given for $\kappa = 0$.
First, we observe that in all cases we obtain a hit ratio smaller than 5% (our nominal significance level) when $\kappa = 0$, i.e. we have no breaks in the series. For all combinations of $D = (d_1, d_2)$ we observe that the hit ration increases as the break size increases. Additionally, we see in all graphs that the more breaks we have the larger their size needs to be to obtain a higher hit ratio.

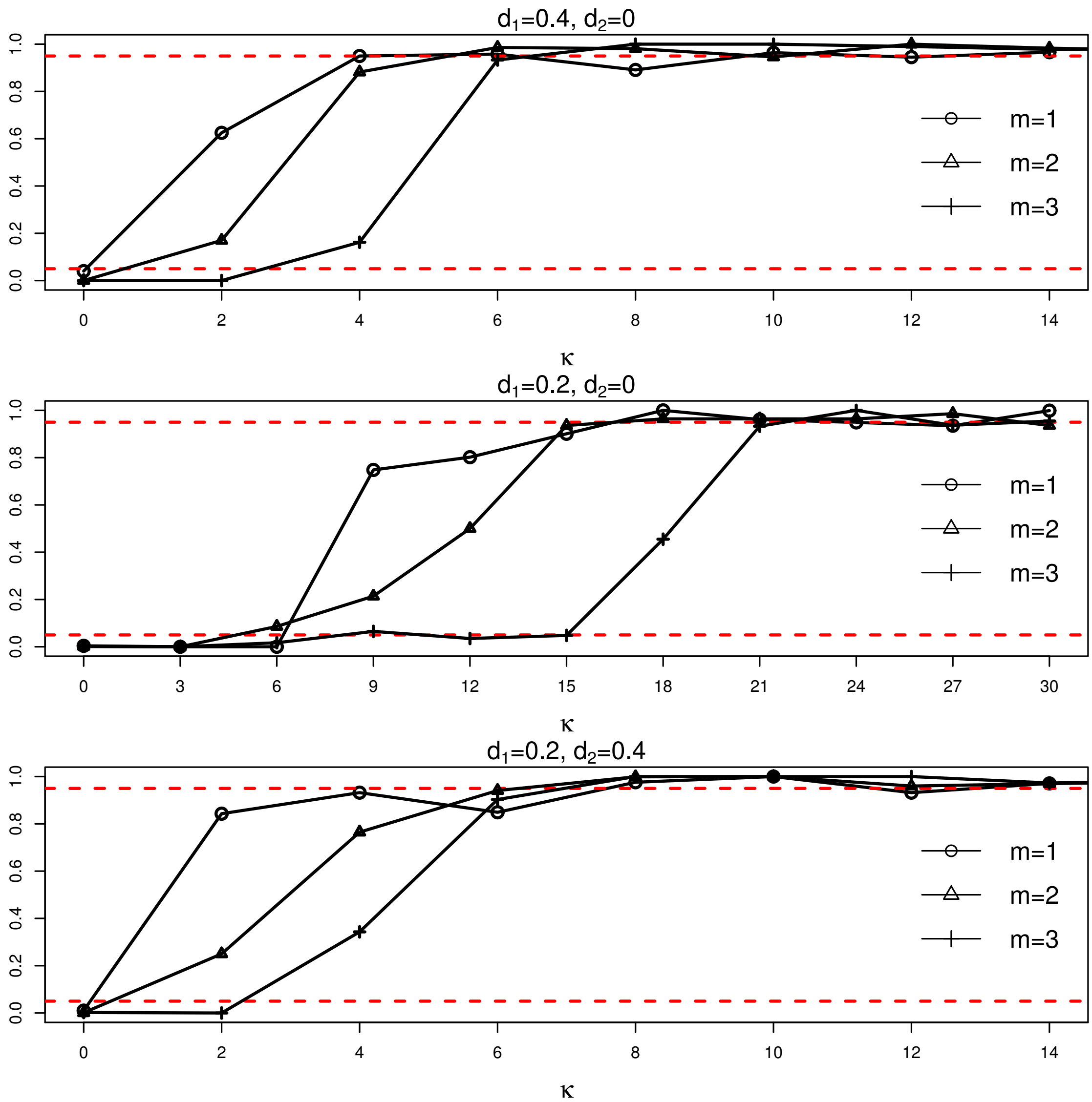


Figure 4.1: Hit ratio of our REBIT procedure for different values of d_1 and d_2 where the true number of breaks is m. The parameter κ on the x-axis is related to the break size, which increases as κ increases. The value on the y-axis provides the hit ratio of our test, i.e. whether the true number of breaks is detected. $T = 1,000$, the memory parameter is estimated by multivariate local Whittle estimation, the series are uncorrelated.

4.6 Empirical Application

Inflation is one of the key variables in macroeconomics since it is assumed to determine unemployment and national output. Over the past years numerous empirical studies found that inflation rates possess significant autocorrelations at large lags and a pole at the periodogram at Fourier frequencies local to zero (Hassler and Wolters (1995) or Kumar and Okimoto (2007) among others).

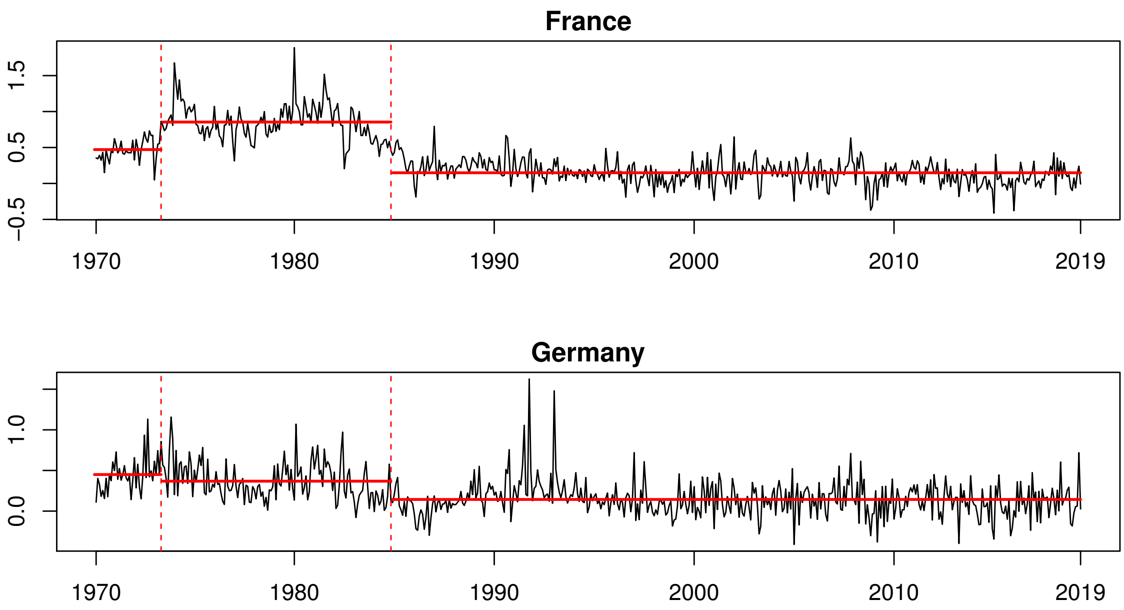


Figure 4.2: Monthly inflation rates of France and Germany from 1970 to 2019. The dotted red vertical lines refer to the mean shifts our procedure detected. The bold red horizontal lines refer to the estimated means in each partition.

This can be seen as an indication that inflation rates follow a pure long-memory process, but similar time series features can also be generated by short-memory processes that are contaminated with breaks, which is referred to as spurious long memory (see for example Diebold and Inoue (2001), Granger and Hyung (2004), Mikosch and Stărică (2004)).

Standard estimation procedures of the long-memory parameter are biased upwards in the presence of breaks. The other way around standard testing procedures for shifts detect too many breaks in a long-memory time series. The literature is therefore unclear about the nature of the underlying process of inflation time series. On the one hand Hassler and Wolters (1995) and Baum et al. (1999) argue, for example, that an ARFIMA model can describe inflation rates well. Bos et al. (1999) and Morana (2002) on the other hand find evidence of structural breaks in international inflation rates and Gadea et al. (2004) shows that the memory of the series is reduced when structural changes are allowed. Many recent contributions favor a mixture of long-memory models and structural breaks (Kumar and Okimoto (2007), among others).

However, whether the series follow a pure long-memory process, a short-memory process with breaks or a mixture of long memory and breaks is of major importance for policy makers: if inflation rates are persistent, monetary policy actions need more time to unfold their effect, which is more expensive.

Our testing procedure allows to detect the true number of breaks for multivariate time series that are allowed to possess long memory. Therefore, it can be used to examine the properties of the underlying process of inflation rates. We use monthly CPI data (P_t) from January 1970 until May 2019 of Germany and France available from the OECD[1] to calculate inflation rates (π_t) as

$$\pi_t = 100(\log P_t - \log P_{t-1}).$$

As a result we have 592 observations that are further seasonally adjusted. Figure 4.2 illustrates the bivariate time series along with the detected break points and partitions. Table 4.1 provides further results of our testing procedure and regarding the persistence of the raw and the demeaned inflation series.

The left hand side of Panel A of Table 4.1 shows results regarding the persistence of the inflation series. In line with earlier empirical results (see, for example, Hassler and Wolters (1995) or Bos et al. (1999)) the multivariate local Whittle estimator (GSE) by Shimotsu (2007) estimates high values of d for the raw data such that both series seem to be highly persistent. However, there is evidence that the long-memory time series are contaminated by breaks, which lead to an upward bias of the memory parameter estimates of the GSE (Mikosch and Stărică (2004)). First, the multivariate test against spurious long memory (MLWS) by Sibbertsen, Leschinski, et al. (2018) rejects the null hypothesis of pure long-memory processes at a 1% significance level. Second, applying the (univariate) trimmed log-periodogram estimator (tGPH) by McCloskey and Perron (2013) on both inflation series, we observe that the memory decreases. Therefore, we apply our procedure that can consistently detect and estimate multiple shifts in the bivariate system of inflation series.

The results can be seen in Panel B of Table 4.1.

[1] http://data.oecd.org/price/inflation-cpi.htm

Panel A: persistence

	raw series			demeaned series	
	d_{GSE}	MLWS	d_{tGPH}	d_{GSE}	MLWS
France	0.567	2.518***	0.231	0.234	1.227*
Germany	0.375		0.230	0.250	

Panel B: breaks

	# breaks	breakdates
REBIT	2	05/73, 11/84

Table 4.1: Panel A presents results regarding the persistence of the system. On the left hand side of the table the memory of the raw inflation series is estimated applying on the one hand the multivariate local Whittle estimator (GSE) by Shimotsu (2007) with a bandwidth of $m = \lfloor T^{2/3} \rfloor$ and on the other hand the trimmed log-periodogram estimator (tGPH) by McCloskey and Perron (2013), which is robust against shifts, with $m = \lfloor T^{0.8} \rfloor$ and the constant that determines the trimming of $\epsilon = 0.05$. Furthermore, the test statistic of the multivariate test against spurious long memory (MLWS) by Sibbertsen, Leschinski, et al. (2018) is given with $m = \lfloor T^{2/3} \rfloor$ and trimming parameter $\epsilon = 0.02$. Here, *** denotes significance at 1%, ** significance at 5% and * significance at 10%.
On the right hand side of the table the GSE estimates of the memory as well as the result of the MLWS test are given for the demeaned time series. The demeaning was executed with regard to the break dates detected by our REBIT procedure.
Panel B presents the number of breaks and corresponding break dates detected by our REBIT procedure.

We observe that the first structural break our procedure detects was in May 1973, which is a few month before the first oil crisis. We further see that the mean in the second partition of the inflation series of France increases heavily while the mean in the series of Germany decreases. The second structural break is detected in November 1984 which could be connected to the 1980s oil glut. We observe that the mean of both inflation series strongly decrease in the third partition of the series.

The other two procedures we consider are the SEQ($l+1|l$) test of Qu and Perron (2007) and the F($l+1|l$) test of Bai and Perron (1998). The F($l+1|l$) test detects 12 breaks in the inflation series, the SEQ($l+1|l$) test more than 19. Some of the detected break dates are similar to the ones found by the REBIT test, but the other two tests find more breaks especially at the end of the sample. This can be reasoned by the fact that both procedures are not robust under long memory. The robust tGPH estimator by McCloskey and Perron (2013) indicates that there is still memory left apart from the upward bias in standard long-memory estimation methods induced by breaks.

To further investigate whether the REBIT procedure detects the relevant breaks, we examine the demeaned inflation series (demeaning is executed with the two breaks our procedure detected). The results of the GSE estimator and MLWS test can be seen on the right hand side of Panel A of Table 4.1. The estimated memory by the GSE strongly decreases to a value around 0.24 which is similar to the tGPH estimate of the raw series. Furthermore, evidence of spurious long memory also lessens since the MLWS test is just significant at the 10% significance level. Therefore, we conclude that our procedure detects all relevant breaks of the bivariate inflation system.

4.7 Conclusions

This paper contains to the best of our knowledge the first procedure for testing for multiple breaks in a long-memory time series framework. We embed our procedure into a multivariate system of long-memory time series allowing for breaks in the mean as well as in the covariance matrix. The breaks are allowed to appear contemporaneously or at different times. Our assumptions on the breaks are fairly general basically just assuming that the size of the breaks depends of the memory of the underlying time series.

The procedure consists of iteratively testing for m structural breaks with m increasing in each step. It therefore avoids splitting the sample in segments as in Bai (1998) and others which is not possible under long memory. Our test and break point estimator in each step is likelihood-ratio based. The consistency and limiting distribution of both procedures are derived. Interestingly, the limiting distribution of the test depends on the one hand only on the maximum of all memory parameters but on the other hand on the number of series having this maximum memory.

A Monte Carlo study demonstrates the finite sample properties of our procedure and an application to inflation rates its usefulness in practice.

4.8 Appendix

This section contains the proofs of Lemma 4.1 and Theorem 4.1 and 4.2. In order to prove these results we need a generalized Hájek-Rényi inequality, a strong law of large numbers (SLLN) and a functional central limit theorem (FCLT) that hold under our stated assumptions and in particular under long memory. We collect them in separate Lemmas in Section 4.8.1. Afterwards we state in Section 4.8.2 10 properties of the quasi-likelihood that have been considered in Bai, Lumsdaine, et al. (1998), Bai (2000) and Qu and Perron (2007). The proofs can be found in Section 4.8.7. We prove consistency of the break point estimators in Section 4.8.3, i.e. Lemma 4.1. In Section 4.8.5 we proof the limiting distribution of our test statistic, i.e. Theorem 4.2.

4.8.1 Generalised Hájek-Rényi Inequality, SLLN, FCLT

Lemma 4.3 (Generalised Hájek-Rényi Inequality)**.** Let $(\xi_i)_{i\geq 1}$ be a sequence of mean zero $\mathbb{R}^d$-valued random vectors. Define $\mathcal{F}_k$ as an increasing σ-field generated by $(\xi_i)_{i\geq k}$. Suppose $(\xi_i)_{i\geq 1}$ satisfies Assumption 4 with $x_i u_i$ replaced by ξ_i. Then there exists an $L < \infty$ such that, for every $\delta > 0$ and $m > 0$, $P(\sup_{k\geq m} k^{-1}\|\sum_{t=1}^{k} \xi_t\| > \delta) \leq (L/\delta^2 m^{2d-1})$, where $d = \mathrm{d}_{\max}$ is the largest memory parameter of the elements of the vector ξ.

Proof. In the following we write for the partial sums $M_{i:j} = \sum_{t=i}^{j} \xi_t$. We start by noting

$$P\left(\max_{k\geq m}\frac{1}{k}\|M_{1:k}\|>\delta\right)\;\leq\;\sum_{p=0}^{\infty}P\left(\max_{2^p\leq k\leq 2^{p+1}m}\frac{1}{k}\|M_{1:k}\|>\delta\right). \tag{4.5}$$

To simplify notation we write $S_{i:j}=\max_{k=i,\ldots,j}1/k\,\|M_{1:k}\|$. We need the following auxilliary result:

$$P\left(\max_{1\leq k\leq n}1/k\|M_{1:k}\|>\delta\right)\leq 4\frac{A(d)C(\varepsilon)}{\delta^2}n^{2d}\sum_{t=1}^{n}\left(\frac{1}{t}\right)^2. \tag{4.6}$$

Suppose (4.6) holds. Then we can write

$$\begin{aligned}P(\max_{2^pm\leq k\leq 2^{p+1}m}\frac{1}{k}\|M_{1:k}\|>\delta)\leq &P(\frac{1}{2^pm}\|M_{1:m}\|>\frac{\delta}{2})\\&+P(\max_{2^pm+1\leq k\leq 2^{p+1}m}\frac{1}{k}\|M_{2^pm+1:2^{p+1}m}\|>\frac{\delta}{2})\\&\leq 4\frac{A(d)C(\varepsilon)}{\delta^2}(2^pm)^{2d-2}+4\frac{A(d)C(\varepsilon)}{\delta^2}(2^pm)^{2d}\sum_{t=2^pm+1}^{2^{p+1}}\left(\frac{1}{t}\right)^2\\&\leq 8\frac{A(d)C(\varepsilon)}{\delta^2}(2^pm)^{2d-1}.\end{aligned}$$

Using equation (4.5) we have

$$P\left(\max_{k\geq m}\frac{1}{k}\|M_{1:k}\|>\delta\right)\;\leq\;8\frac{A(d)C(\varepsilon)}{\delta^2}\sum_{p=0}^{\infty}(2^pm)^{2d-1}\;\leq\;\frac{L}{\delta^2}m^{2d-1},$$

where $L<\infty$ is a constant.

We prove equation (4.6) by the Markov inequality. Specifically, we set out to prove

$$E\left(S_{1:n}^2\right)\;\leq\;C(\varepsilon)A(d)n^{2d}\sum_{t=1}^{n}\frac{1}{t^2}. \tag{4.7}$$

If equation (4.7) holds, our auxiliary result (4.6) is proven by the Markov inequality. The claim in (4.7) is proved by induction on n. For $n=1$ the inequality is obvious for $A(d)=1$ because of the following inequality: Kechagias and Pipiras (2015) proved that for the partial sums $M_{i:j}$ there exists $C(\varepsilon)<\infty$ such that, for all i,j,

$$E\left(\|M_{i:j}\|^2\right)\;\leq\;C(\varepsilon)\,|j-i+1|^{2d+1}. \tag{4.8}$$

For the induction step we set $m=\lceil\frac{n}{2}\rceil+1$. Then, we note that

$$\max_{k=1,\ldots,n}\frac{1}{k}\|M_{1:k}\|\;\leq\;\frac{1}{m}M_{1:m}+\left(\left(\max_{k=1,\ldots,m-1}\frac{1}{k}\|M_{1:k}\|\right)^2+\left(\max_{k=m+1,\ldots,n}\frac{1}{k}\|M_{1:k}\|\right)^2\right)^{1/2}.$$

Applying the Minkowski inequality to the above inequality yields

$$
\begin{aligned}
E\left(S_{1:n}^2\right)^{1/2} \leq & \frac{1}{m}(E(\|M_{1:m}\|^2)^{1/2} + \left(E(S_{1:m-1}^2) + E(S_{m+1:n}^2)\right)^{1/2} \\
\leq & \frac{1}{m}\left(C(\varepsilon)m^{2d+1}\right)^{1/2} + \left(A(d)C(\varepsilon)\left((m-1)^{2d}\sum_{t=1}^{m-1}\frac{1}{t^2} + (n-m)^{2d}\sum_{t=m+1}^{n}\frac{1}{t^2}\right)\right)^{1/2} \\
\leq & \left(C(\varepsilon)m^{2d}\sum_{t=1}^{n}\frac{1}{t^2}\right)^{1/2} + \left(A(d)C(\varepsilon)\left(\frac{n}{2}\right)^{2d}\left(\sum_{t=1}^{m-1}\frac{1}{t^2} + \sum_{t=m+1}^{n}\frac{1}{t^2}\right)\right)^{1/2} \\
\leq & \left(C(\varepsilon)n^{2d}\sum_{t=1}^{n}\frac{1}{t^2}\right)^{1/2}\left(1 + \left(\frac{A(d)}{2^{2d}}\right)^{1/2}\right),
\end{aligned}
$$

where we used equation (4.8) and the induction hypothesis in the second line and the fact that $1 \leq \sum_{t=1}^{m} 1/t^2$ in the third line. Now we choose $A(d)$ such that

$$
1 + \frac{A(d)^{1/2}}{2^d} \leq A(d)^{1/2} \quad \Leftrightarrow \quad A(d) \geq \left(1 - \frac{1}{2^d}\right)^{-2} \geq 1.
$$

The induction step is proven and thus this concludes the proof of inequality (4.7). □

To state the following Lemma we repeat the notion of multivariate fractional Brownian motion (cf. Marinucci and Robinson (2000), Davidson and de Jong (2000), Chung (2002)). We denote by $W_D(t) = (W_{d_1}(t), \ldots, W_{d_n}(t))'$ an n-dimensional fractional Brownian motion with n different memory parameters $D = (d_1, \ldots, d_n)'$. Each $W_{d_i}(t)$ is a one-dimensional fractional Brownian motion defined by

$$
W_{d_i}(t) = \frac{1}{\Gamma(d_i+1)}\left(\int_0^t (t-s)^{d_i} dW_0^{(i)}(s) + \int_{-\infty}^{0}\left((t-s)^{d_i} - (-s)^{d_i}\right) dW_0^{(i)}(s)\right),
$$

where $W_0^{(i)}(t)$ is the ith element of an n-dimensional Brownian motion with the covariance matrix Ω.

Lemma 4.4 (FCLT, SLLN)**.** Let $(\xi_i)_{i\geq 1}$ be a sequence of mean zero $\mathbb{R}^d$-valued random vectors that satisfy Assumption 4. Then

(a) (FCLT)

$$
\operatorname{diag}(T^{-1/2-d_1}, \ldots, T^{-1/2-d_n}) \sum_{t=1}^{[Tr]} \xi_t \Rightarrow \Omega W_D(r),
$$

where $W_D(r)$ is an n vector of independent fractional Wiener processes and $\Rightarrow$ denotes weak convergence under the Skorohod topology;

(b) (SLLN)

$$
k^{-1}\sum_{i=1}^{k}\xi_i \xrightarrow{a.s.} 0 \qquad \text{as} \qquad k \to \infty;
$$

Proof. a) Under our Assumptions Theorem 1 of Chung (2002) holds and gives this result.

b) Under our Assumptions Corollary 3 of Wu (2007) applies and gives this result. □

4.8.2 10 Properties of the Quasi-Likelihood Ratio

This section contains 10 properties of the quasi-likelihood ratio and parameter estimates. We need them in subsequent proofs. In this subsection we write

$$\mathcal{L}(1, k; \beta, \Sigma) = \frac{\prod_{t=1}^{k} f(y_t|x_t, ..., \beta, \Sigma)}{\prod_{t=1}^{k} f(y_t|x_t, ..., \beta_0, \Sigma_0)},$$

where β^0 and Σ^0 describe the true values of the coefficients. In the following we denote by $\hat{\beta}_{(k)}$ and $\hat{\Sigma}_{(k)}$ estimates obtained from maximizing $\mathcal{L}(1, k; \beta, \Sigma)$. Then the following properties hold:

Property 1. For each $\delta \in (0, 1]$

$$\sup_{T\delta \le k \le T} \mathcal{L}(1, k; \hat{\beta}_{(k)}, \hat{\Sigma}_{(k)}) = O_p(1),$$
$$\sup_{T\delta \le k \le T} (\|\hat{\beta}_{(k)} - \beta_0\| + \|\hat{\Sigma}_{(k)} - \Sigma_0\|) = O_p(T^{d-1/2}).$$

The following property is modified compared to property 2 of Qu and Perron (2007). Instead of considering the supremum of the likelihoood over $1 \le k \le T$ we consider here the supremum over $\delta T \le k \le T$ for some $\delta \in (0, 1)$.

Property 2. For some $\delta \in (0, 1)$, each $\varepsilon > 0$, there exists a $B > 0$ such that

$$\Pr\left(\sup_{\delta T \le k \le T} T^{-B} \mathcal{L}(1, k; \hat{\beta}_{(k)}, \hat{\Sigma}_{(k)}) > 1\right) < \varepsilon$$

for all large T.

Property 3. Let $S_T = \{(\beta, \Sigma) \colon \|\beta - \beta^0\| \ge T^{-1/2+d} \log T \text{ or } \|\Sigma - \Sigma^0\| \ge T^{-1/2+d} \log T\}$. For any $\delta \in (0, 1), D > 0$ and $\varepsilon > 0$ the following statement holds when T is large:

$$\Pr\left(\sup_{k \ge \delta T} \sup_{(\beta, \Sigma) \in S_T} T^D \mathcal{L}(1, k; \beta, \Sigma) > 1\right) < \varepsilon. \tag{4.9}$$

Property 4. Not needed.

The following property is different from Qu and Perron (2007) in that we do not assume that the limit of $(h_T d_T^2)/T$ exists. Instead as pointed out by Bai (2000) we assume the sufficient condition that $\liminf_{T\to\infty} (h_T d_T^2)/T \ge h > 0$.

Property 5. Let h_T and d_T be positive sequences such that h_T is nondecreasing, $d_T \to \infty$ and $\liminf_{T\to\infty} (h_T d_T^2)/T \ge h > 0$. Define $\overline{\Theta}_3 = \{(\beta, \Sigma) \colon \|\beta\| \le p_1, \lambda_{\min}(\Sigma) \ge p_2, \lambda_{\max}(\Sigma) \le p_3\}$, where p_1, p_2 and p_3 are arbitrary constants that satisfy $p_1 < \infty, 0 < p_2 \le p_3 < \infty$. Define $S_T = \{(\beta, \Sigma) \colon \|\beta - \beta^0\| \ge T^{-1/2+d} \log T$ or $\|\Sigma - \Sigma^0\| \ge T^{-1/2+d} \log T\}$. Then, for any $\varepsilon > 0$, there exists an $A > 0$, such that

$$\Pr\left(\sup_{k \ge A h_T} \sup_{(\beta, \Sigma) \in S_T \cap \bar{\Theta}_3} \mathcal{L}(1, k; \beta, \Sigma) > \varepsilon\right) < \varepsilon$$

when T is large.

Property 6. With ν_T satisfying Assumption 6, for each β and Σ such that $\|\beta - \beta_0\| \leq Mv_T$ and $\|\Sigma - \Sigma_0\| \leq Mv_T$, with $M < \infty$, we have

$$\sup_{1 \leq k \leq T^{1/2-d} v_T^{-1}} \sup_{\lambda, \Xi} \frac{\mathcal{L}(1, k; \beta + T^{-1/2+d}\lambda, \Sigma + T^{-1/2+d}\Xi)}{\mathcal{L}(1, k; \beta, \Sigma)} = o_p(1).$$

Property 7. Under the conditions of Property 6, we have

$$\sup_{1 \leq k \leq M v_T^{-2}} \sup_{\lambda, \Xi} \log \frac{\mathcal{L}(1, k; \beta + T^{-1/2+d}\lambda, \Sigma + T^{-1/2+d}\Xi)}{\mathcal{L}(1, k; \beta, \Sigma)} = o_p(1).$$

Property 8. We have

$$\sup_{T\delta \leq k \leq T} \sup_{\beta^*, \Sigma^*, \lambda, \Xi} \log \frac{\mathcal{L}(1, k; \beta_0 + T^{-1/2+d}\beta^* + T^{-1+2d}\lambda, \Sigma_0 + T^{-1/2+d}\Sigma^* + T^{-1+2d}\Xi)}{\mathcal{L}(1, k; \beta_0 + T^{-1/2+d}\beta^*, \Sigma_0 + T^{-1/2+d}\Sigma^*)} = o_p(1),$$

where the supremum with respect to $\beta^*, \Sigma^*, \lambda, \Xi$ is taken over an arbitrary compact set.

Property 9. Let $T_1 = [aT]$ for some $a \in (0, 1]$ and let $T_2 = [T^{1/2-d} v_T^{-1}]$, where v_T satisfies Assumption 6. Consider

$$\begin{aligned} y_t =& x_t' \beta_1^0 + \Sigma_1^0 \eta_t, && (t = 1, \ldots, T_1), \\ y_t =& x_t' \beta_2^0 + \Sigma_2^0 \eta_t, && (t = T_1 + 1, \ldots, T_1 + T_2), \end{aligned}$$

where $\|\beta_1^0 - \beta_2^0\| \leq Mv_T$ and $\|\Sigma_1^0 - \Sigma_2^0\| \leq Mv_T$ for some $M < \infty$. Let $k = T_1 + T_2$ be the size of the pooled sample and let $(\hat{\beta}_n, \hat{\Sigma}_n)$ be the associated estimates. Then $\hat{\beta}_n - \beta_1^0 = O_p(T^{d-1/2})$ and $\hat{\Sigma}_n - \Sigma_1^0 = O_p(T^{d-1/2})$.

Property 10. Not needed.

4.8.3 Proof of Lemma 4.1

Proof. We show the consistency in two steps: First we prove an auxiliary result on the convergence rate of the break point estimates. Second we use results from Bai (2000) to justify the statement.

Let $N := [T^{\frac{1}{2}-d} \nu_T^{-1}]$. Let $A_j = \{(k_1, ..., k_m) \in \Lambda_\varepsilon \colon |k_i - k_j^0| > N, i = 1, \ldots, m\}$, where Λ_ε is given in Assumption 11. Because $LR_T(\hat{k}_1, ..., \hat{k}_m) \geq LR_T(k_1^0, ..., k_m^0) \geq LR_T(k_1^0, ..., k_m^0, \beta^0, \Sigma^0) = 1$, to show $(\hat{k}_1, ..., \hat{k}_m) \notin A_j$, it suffices in a first step to show

$$P(\sup_{(k_1, \ldots, k_m) \in A_j} LR_T(k_1, \ldots, k_m) > \varepsilon) < \varepsilon. \tag{4.10}$$

We extend the definition of LR_T to every subset $\{l_1, \ldots, l_r\}$ of $\{1, 2, ..., T-1\}$ such that $LR_T(l_1, \ldots, l_r) = LR_T(l_{(1)}, \ldots, l_{(r)})$ where $0 < l_{(1)} < \cdots < l_{(r)}$ are the ordered versions of $l_{(1)}, \ldots, l_{(r)}$. For every $(k_1, \ldots, k_m) \in A_j$,

$$LR_T(k_1, \ldots, k_m) \leq LR_T(k_1, \ldots, k_m, k_1^0, \ldots, k_{j-1}^0, k_j^0 - N, k_j^0 + N, k_{j+1}^0, \ldots, k_m^0). \tag{4.11}$$

Denote the likelihood ratio of the segment $[k, l]$ by

$$D(k, l, \beta, \Sigma) = \frac{\prod_{t=k+1}^{l} f(y_t|x_t; \beta, \Sigma)}{\prod_{t=k+1}^{l} f(y_t|x_t; \beta^0, \Sigma^0)}$$

and its optimal value

$$D(k, l) = \sup_{\beta, \Sigma} D(k, l, \beta, \Sigma).$$

The likelihood ratio of the entire sample can be written as

$$LR_T(k_1, \ldots, k_m) = D(0, k_1) \cdot D(k_1, k_2) \cdot \cdots \cdot D(k_m, T). \tag{4.12}$$

The right hand side of (4.11) can be written as the product of at most $(2m+2)$ terms expressible as $D(l, k)$ as in (4.12). There are at most $(2m+2)$ terms because k_i may coincide with k_l^0 for some i and l. One of these $(2m+2)$ terms is $D(k_j^0 - N, k_j^0 + N)$ and all the rest can be written as $D(l, k)$ with $[l, k] \subset [k_1^0+1; k_{i+1}^0]$ for some i. By Property 1 and 2, $\log D(l, k) = O_p(\log T)$ uniformly in l, k such that $k_i^0 + 1 \leq l < k \leq k_{i+1}^0$ with $|l - k| > T\nu$. That is, $D(k, l) = O_p(T^B)$ for some $B > 0$. Thus,

$$LR_T(k_1, \ldots, k_m) \leq O_p(T^{(2m+1)B})\, D(k_j^0 - N, k_j^0 + N). \tag{4.13}$$

We now show that $D(k_j^0 - N, k_j^0 + N)$ is small. Introduce the reparameterization. $LR_T^*(k, l, \beta, \Sigma) = D(k, l, \beta^0 + (l-k)^{-1/2}\beta, \Sigma^0 + (l-k)^{-1/2}\Sigma)$ assuming that (β^0, Σ^0) is the true parameter of the segment $[k, l]$. We note that

$$\begin{aligned} D(k_j^0 - N, k_j^0 + N) &= \sup_{\beta, \Sigma} [D(k_j^0 - N, k_j^0; \beta, \Sigma) \cdot D(k_j^0, k_j^0 + N; \beta, \Sigma)] \\ &= \sup_{\beta, \Sigma} [LR_T^*(k_j^0 - N, k_j^0; N^{1/2}(\beta - \beta_j^0), N^{1/2}(\Sigma - \Sigma_j^0)) \\ &\quad \times (LR_T^*(k_j^0, k_j^0 + N; N^{1/2}(\beta - \beta_{j+1}^0), N^{1/2}(\Sigma - \Sigma_{j+1}^0))]. \end{aligned} \tag{4.14}$$

This follows from the definition of LR_T^* and the fact that (β_j^0, Σ_j^0) is the true parameter for the segment $[k_j^0 - N, k_j^0]$ and $(\beta_{j+1}^0, \Sigma_{j+1}^0)$ is the true parameter for the segment $[k_j^0 + 1, k_j^0 + N]$. From $\max\{\|x - z\|, \|y - z\|\} \geq \|x - y\|/2$ for all (x, y, z), we have for all β and Σ

$$\begin{aligned} \max\{N^{1/2}\|\beta - \beta_j^0\|, N^{1/2}\|\beta - \beta_{j+1}^0\|\} &\geq N^{1/2}\|\beta_j^0 - \beta_{j+1}^0\|/2 \\ \max\{N^{1/2}\|\Sigma - \Sigma_j^0\|, N^{1/2}\|\Sigma - \Sigma_{j+1}^0\|\} &\geq N^{1/2}\|\Sigma_j^0 - \Sigma_{j+1}^0\|/2. \end{aligned}$$

By Assumption 6, we either have $N^{1/2}\|\beta_j^0 - \beta_{j+1}^0\|/2 \geq \log N$ or $N^{1/2}\|\Sigma_j^0 - \Sigma_{j+1}^0\|/2 \geq \log N$. This follows from if $\|\beta_j^0 - \beta_{j+1}^0\| \geq \nu_T C$ for some $C > 0$, then

$$N^{1/2}\|\beta_j^0 - \beta_{j+1}^0\|/2 = (T^{1/2}\nu_T^{-1})^{1/2}\nu_T C = C(T^{1/2}\nu_T)^{\frac{1}{2}} \geq \log T \geq \log N.$$

Now suppose that $N^{1/2}\|\beta_j^0 - \beta_{j+1}^0\|/2 \geq \log N$. Then we have either (i) $N^{1/2}\|\beta - \beta_j^0\| \geq \log N$ or (ii) $N^{1/2}\|\beta - \beta_{j+1}^0\| \geq \log N$. For case (i) we can apply Property 3 to the first term inside the brackets of (4.14) to obtain

$$LR_T^*(k_j^0 - N, k_j^0; N^{1/2}(\beta - \beta_j^0), N^{1/2}(\Sigma - \Sigma_j^0)) = O_p(N^{-A})$$

for every $A > 0$. Moreover, by Property 2 the second term inside the bracket of (4.14) is bounded by $O_p(\log T)$. Similarly, for case (ii), we can apply Property 3 to show that the second term of (4.14) is $O_p(N^{-A})$ and the first term is bounded by $O_p(\log T)$. So for each case, we have

$$D(k_j^0 - N, k_j^0 + N) = \log T\, O_p(N^{-A})$$

for an arbitrary $A < 0$. It is further $N^{-A} \leq T^{-A/2}$ since $N \geq T^{1/2}$ for all large T. Thus from (4.13), $LR_T(k_1, \ldots, k_m) \leq O_p(T^{(2m+1)B-\frac{1}{2}A}) \log T \xrightarrow{p} 0$ for a large A. This proves (4.10).

Now by Proposition 2 of Bai (2000) we can deduce that $\hat{k}_j - k_j^0 = O_p(\nu_T^2)$ for $j = 1, \ldots, m$ using the preliminary convergence order given by Equation (4.10). The convergence rate for the estimated regression coefficients β_j and covariances Σ_j follows as in Bai (1997b) and Bai and Perron (1998) due to the fast convergence of the estimated break points. □

4.8.4 Proof of Theorem 4.1

Proof. Without loss of generality, consider the j-th break date and start with the case where the candidate estimate is before the true break date. We obtain an expansion for $lr_j^1([s/\nu_r^2])$ as defined in Theorem 1. Note that s is implicitly defined by $s = \nu_T^2(T_i - T_i^0) = r\nu_T^2$. We deal with each term separately.

For the first term, we have as in Qu and Perron (2007)

$$\frac{1}{2}\sum_{t=T_j^0+[s/\nu_T^2]}^{T_j^0} u_t^T\Big((\Sigma_j^0)^{-1} - (\Sigma_{j+1}^0)^{-1}\Big)u_t$$

$$= \frac{1}{2} tr\Big((\Sigma_j^0)^{\frac{1}{2}}(\Sigma_{j+1}^0)^{-1}\Phi_j(\Sigma_j^0)^{-\frac{1}{2}}\nu_T \sum_{t=T_j^0+[s/\nu_T^2]}^{T_j^0} (\eta_t\, \eta_t^T - I)$$

$$- \frac{r}{2}\nu_T\; tr\Big((\Sigma_{j+1}^0)^{-1}\Phi_j\Big).$$

For the second term we have

$$-\frac{r}{2}(\log|\Sigma_j^0| - \log|\Sigma_{j+1}^0|) = \frac{r}{2}\nu_T\; tr\Big(\Phi_j(\Sigma_{j+1}^0)^{-1}\Big) + \frac{r}{4}\nu_T^2\; tr\Big([\Phi_j(\Sigma_{j+1}^0)^{-1}]^2\Big).$$

The sum of the first two terms is

$$\frac{1}{2}\sum_{T_j^0+[s/\nu_T^2]}^{T_j^0} u_t^T\Big((\Sigma_j^0)^{-1} - (\Sigma_{j+1}^0)^{-1}\Big)u_t - \frac{r}{2}(log|\Sigma_j^0| - log|\Sigma_{j+1}^0|)$$

$$= \frac{1}{2} tr\Big((\Sigma_j^0)^{\frac{1}{2}}(\Sigma_{j+1}^0)^{-1}\Phi_j(\Sigma_j^0)^{-\frac{1}{2}}\nu_T \sum_{T_j^0+[s/\nu_T^2]}^{T_j^0} (\eta_t\, \eta_t^T - I)\Big)$$

$$+ \frac{r}{4}\nu_T^2\; tr\Big([\Phi_j(\Sigma_{j+1}^0)^{-1}]^2\Big) = \text{I} + \text{II}.$$

Now

$$\begin{aligned} T^{1-2d}(\mathrm{I}+\mathrm{II}) &\xrightarrow{d} \frac{1}{2}tr\Big((\Sigma_j^0)^{\frac{1}{2}}(\Sigma_{j+1}^0)^{-1}\Phi_j(\Sigma_j^0)^{-\frac{1}{2}}\xi_{1,d,j}(s)\Big) \\ &+\frac{s}{4}tr\Big([(\Sigma_{j+1}^0)^{-1}\Phi_j]^2\Big) \\ &=\frac{1}{2}tr(A_{1,j}\xi_{1,d,j}(s))+\frac{s}{4}tr(A_{1,j}^2), \end{aligned}$$

where $\xi_{1,d,j}$ is a nonstandard Brownian motion process with $var\Big[vec(\xi_{1,d,j}(s))\Big]=\Omega_{1,j}^0$. For the third term we have

$$-\frac{1}{2}\sum_{t=T_j^0+[s/\nu_T^2]}^{T_j^0}(\beta_j^0-\beta_{j+1}^0)^T x_t(\Sigma_{j+1}^0)^{-1}x_t^T(\beta_j^0-\beta_{j+1}^0)\xrightarrow{P}\frac{1}{2}s\ \delta_j^T\ Q_{1,j}\ \delta_j.$$

Note that x_t belongs to regime j, but it is scaled by the covariance matrix of regime $j+1$ because the estimate of the break occurs before the true break date. For the fourth term,

$$-T^{1-2d}\sum_{t=T_j^0+[s/\nu_T^2]}^{T_j^0}(\beta_j^0-\beta_{j+1}^0)^T x_t(\Sigma_{j+1}^0)^{-1}u_t\xrightarrow{d}\delta_j^T(\Pi_{1,j})^{\frac{1}{2}}\zeta_{1,d,j}(s)$$

with

$$\Pi_{1,j}=\lim_{T\to\infty}Var\Big\{(T_j^0-T_{j-1}^0)^{-\frac{1}{2}}\Big[\sum_{t=T_j^0+[s/\nu_T^2]}^{T_j^0}x_t(\Sigma_{j+1}^0)^{-1}(\Sigma_j^0)^{\frac{1}{2}}\eta_t\Big]\Big\}.$$

Combining these results, we have, for $s<0$

$$\begin{aligned} T^{1-2d}lr_j^1\Big([\frac{s}{\nu_T^2}]\Big) &\xrightarrow{d} -\frac{|s|}{2}\Big[\frac{1}{2}tr(A_{1,j}^2)+\delta_j^T\ Q_{1,j}\ \delta_j\Big] \\ &+\frac{1}{2}\operatorname{vec}(A_{1,j})^T\operatorname{vec}(\xi_{1,d,j}(s))+\delta_j^T(\pi_{1,j})^{\frac{1}{2}}\zeta_{1,d,j}(s). \end{aligned}$$

Now, $\operatorname{vec}(A_{1,j})^T\operatorname{vec}(\xi_{1,d,j}(s))\overset{d}{=}\Big(vec(A_{1,j})^T\Omega_{1,j}^0 vec(A_{1,j})\Big)^{\frac{1}{2}}V_{1,d,j}(s)$, where $V_{1,d,j}(s)$ is a standard fractional Brownian motion.

Similarly, $\delta_j^T(\Pi_{1,j})^{\frac{1}{2}}\zeta_{1,d,j}(s)\overset{d}{=}(\delta_j^T\ \Pi_{1,j}\ \delta_j)^{\frac{1}{2}}U_{1,d,j}(s)$ and $U_{1,d,j}(s)$ is a standard fractional Brownian motion. Under the stated conditions, $V_{1,d,j}(s)$ and $U_{1,d,j}(s)$ are independent. Then,

$$\begin{aligned} &\Big(vec(A_{1,j})^T\Omega_{1,j}^0 vec(A_{1,j})/4\Big)^{\frac{1}{2}}V_{1,d,j}(s)+\Big(\delta_j^T(\pi_{1,j})\delta_j\Big)^{\frac{1}{2}}U_{1,d,j}(s) \\ &\overset{d}{=}\Big(vec(A_{1,j})^T\Omega_{1,j}^0 vec(A_{1,j})/4+\delta_j^T(\pi_{1,j})\delta_j\Big)^{\frac{1}{2}}W_{1,j,d}(s) \\ &\equiv T_{1,j}\ W_{1,j,d}(s), \end{aligned}$$

where $W_d(s)$ is a unit fractional Brownian motion.

Hence with $\Delta_{1,j} = tr(A_{1,j}^2)/2 + \delta_j^T\ Q_{1,j}\ \delta_j$, we have

$$T^{1-2d} lr_j^1\Big([\frac{s}{\nu_T^2}]\Big) \xrightarrow{d} -\frac{|s|}{2}\Delta_{1,j} + T_{1,j}\ W_{1,j,d}(s)$$

The proof for $s > 0$ is similar:

$$T^{1-2d} lr_j^1\Big([\frac{s}{\nu_T^2}]\Big) \xrightarrow{d} -\frac{|s|}{2}\Delta_{2,j} + T_{2,j}\ W_{2,j,d}(s)$$

with $\Delta_{2,j} = tr(A_{2,j}^2)/2 + \delta_j^T\ Q_{2,j}\ \delta_j$ and

$$T_{2,j} = \Big[vec(A_{2,j})^T \Omega_{2,j}^0 vec(A_{2,j})/4 + \delta_j^T(\pi_{2,j})\delta_j\Big]^{\frac{1}{2}}.$$

By definition it is $lr_j^1(0) = 0$. Given that $s = \nu_T^2(T_j - T_j^0)$, the argmax yields the scaled estimate $\nu_T^2(\hat{T}_j - T_j^0)$. The result follows because we can take the argmax over the compact set C_M and with Lemma 1, this is equivalent to taking the argmax in an unrestricted set because with probability arbitrarily close to 1, the estimates will be contained in C_M

Hence,

$$T^{1-2d}\nu_T^2(\hat{T}_j - T_j^0) \xrightarrow{d} \underset{s}{\text{argmax}} \begin{cases} -\frac{|s|}{2}\Delta_{1,j} + T_{1,j}\ W_{j,d}(s), s \leq 0, \\ -\frac{|s|}{2}\Delta_{2,j} + T_{2,j}\ W_{j,d}(s), s > 0, \end{cases}$$

where $W_{j,d}(s) = W_{1,j,d}(s)$ for $s \leq 0$ and $W_{j,d}(s) = W_{2,j,d}(s)$ for $s > 0$. Multiplying by $\Delta_{1,j}/T_{1,j}^2$ and applying a change of variable with $u = (\Delta_{1,j}^2/T_{1,j}^2)s$, we obtain Theorem 4.1. □

4.8.5 Proof of Theorem 4.2

Proof of Theorem 4.2. We introduce some notation first. Let

$$\tilde{\Sigma}_{1,j} = \frac{1}{T_j}\sum_{t=1}^{T_j}(y_t - x_{at}'\tilde{\beta}_a - x_{bt}'\tilde{\beta}_{b1,j})(y_t - x_{at}'\tilde{\beta}_a - x_{bt}'\tilde{\beta}_{b1,j})$$

be the estimated covariance matrix using the full sample estimate of β_a obtained under the null hypothesis of no change and using the estimate of β_b based on data up to the last date of regime j, defined as

$$\tilde{\beta}_{b1,j} = (\sum_{t=1}^{T_j} x_{bt}\tilde{\Sigma}_{1,j}^{-1}x_{bt}')^{-1}\sum_{t=1}^{T_j} x_t\tilde{\Sigma}_{1,j}^{-1}(y_t - x_{at}'\tilde{\beta}_a).$$

Additionally,

$$\hat{\Sigma}_j = \frac{1}{T_j - T_{j-1}}\sum_{t=T_{j-1}+1}^{T_j}(y_t - x_{at}'\hat{\beta}_a - x_{bt}'\hat{\beta}_{bj})(y_t - x_{at}'\hat{\beta}_a - x_{bt}'\hat{\beta}_{bj})'$$

is the estimate of the covariance matrix of the errors under the alternative hypothesis using the full sample estimate of β_a and using the estimate of β_b based on data from regime j only, that is,

$$\hat{\beta}_{b,j} = (\sum_{t=T_{j-1}+1}^{T_j} x_{bt}\hat{\Sigma}_j^{-1}x'_{bt})^{-1} \sum_{t=T_{j-1}+1}^{T_j} x_t\hat{\Sigma}_j^{-1}(y_t - x'_{at}\hat{\beta}_a).$$

Consider the log-likelihood of a given partition of the sample

$$\begin{aligned}
LR_T(T_1,\dots,T_m) &= \frac{2}{T^{2d}}\log \hat{L}_T(T_1,\dots,T_m) - \frac{2}{T^{2d}}\log \tilde{L}_T = \frac{T}{T^{2d}}\log|\tilde{\Sigma}| - \frac{T}{T^{2d}}\log|\hat{\Sigma}| \\
&= \frac{1}{T^{2d}}\sum_{j=1}^{m}(T_{j+1}\log|\tilde{\Sigma}_{1,j+1}| - T_j\log|\tilde{\Sigma}_{1,j}| - (T_{j+1}-T_j)\log|\hat{\Sigma}_{j+1}|) \\
&=: \frac{1}{T^{2d}}\sum_{j=1}^{m} F_T^j.
\end{aligned}$$

Using a second-order Taylor series expansion of each term gives

$$\begin{aligned}
\log|\tilde{\Sigma}_{1,j+1}| =& \log|\Sigma^0| + \mathrm{tr}((\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j+1} - \Sigma^0)) \\
&- \frac{1}{2}\mathrm{tr}((\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j+1} - \Sigma^0)(\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j+1} - \Sigma^0)) \\
&+ o_p(T^{-1}), \\
\log|\tilde{\Sigma}_{1,j}| =& \log|\Sigma^0| + \mathrm{tr}((\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j} - \Sigma^0)) \\
&- \frac{1}{2}\mathrm{tr}((\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j} - \Sigma^0)(\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j} - \Sigma^0)) + o_p(T^{-1}), \\
\log|\hat{\Sigma}_{j+1}| =& \log|\Sigma^0| + \mathrm{tr}((\Sigma^0)^{-1}(\hat{\Sigma}_{j+1} - \Sigma^0)) \\
&- \frac{1}{2}\mathrm{tr}((\Sigma^0)^{-1}(\hat{\Sigma}_{j+1} - \Sigma^0)(\Sigma^0)^{-1}(\hat{\Sigma}_{j+1} - \Sigma^0)) + o_p(T^{-1}).
\end{aligned}$$

Applying this to the terms F_T^j,

$$\begin{aligned}
F_T^j :=& F_{1,T}^j + F_{2,T}^j \\
=& \mathrm{tr}(T_{j+1}(\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j+1} - \Sigma^0) - T_j{(\Sigma^0)}^{-1}(\tilde{\Sigma}_{1,j} - \Sigma^0)) \qquad (4.15) \\
&- (T_{j+1} - T_j)(\Sigma^0)^{-1}(\hat{\Sigma}_{j+1} - \Sigma^0)) \\
&- \frac{1}{2}\mathrm{tr}(T_{j+1}[(\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j+1} - \Sigma^0)]^2 \qquad (4.16) \\
&- T_j[(\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j} - \Sigma^0)]^2 - (T_{j+1} - T_j)[{(\Sigma^0)}^{-1}(\hat{\Sigma}_{j+1} - \Sigma^0)]^2).
\end{aligned}$$

First we consider $F_{1,T}^j$ and write the regression in matrix form. Under the null hypothesis, we have

$$Y = X_a\beta_a + X_b\beta_b + U$$

with $E(UU') = I_T \otimes \Sigma^0$. If only data up to the last date of regime j are included, we have

$$Y_{1,j} = X_{a1,j}\beta_a + X_{b1,j}\beta_{b1,j} + U_{1,j}.$$

We now define $Y^d_{1,j} = (I_T \otimes \tilde{\Sigma}^{-1/2}_{1,j})Y_{1,j}$, $W_{1,j} = (I_T \otimes \tilde{\Sigma}^{-1/2}_{1,j})X_{a1,j}$, $Z_{1,j} = (I_T \otimes \tilde{\Sigma}^{-1/2}_{1,j})X_{b1,j}$ and $U^d_{1,j} = (I_T \otimes \tilde{\Sigma}^{-1/2}_{1,j})U_{1,j}$. Then, omitting the subscript when the full sample is used, we have

$$\begin{aligned}\tilde{\beta}_a &= [W'M_Z W]^{-1}W'M_Z Y^d, \qquad (4.17)\\ \tilde{\beta}_{b1,j} &= (Z'_{1,j}Z_{1,j})^{-1}Z'_{1,j}(Y^d_{1,j} - W_{1,j}\tilde{\beta}_a),\end{aligned}$$

where $M_Z = I - Z(Z'Z)^{-1}Z'$. The regression equation using only regime $(j+1)$ is

$$Y_{j+1} = X_{a,j+1}\beta_a + X_{b,j+1}\beta_{b,j+1} + U_{j+1}.$$

Define $\bar{Y}^d_{j+1} = (I_T \otimes \hat{\Sigma}^{-1/2}_{j+1})Y_{j+1}$, $\bar{W}_{j+1} = (I_T \otimes \hat{\Sigma}^{-1/2}_{j+1})X_{a,j+1}$, $\bar{Z}_{j+1} = (I_T \otimes \hat{\Sigma}^{-1/2}_{j+1})X_{b,j+1}$, $\bar{U}^d_{j+1} = (I_T \otimes \hat{\Sigma}^{-1/2}_{j+1})U_{j+1}$, $\bar{Z} = \text{diag}(\bar{Z}_1, \ldots, \bar{Z}_{m+1})$. Then, omitting the subscript when the full sample is used, we have

$$\begin{aligned}\hat{\beta}_a &= [\bar{W}'M_{\bar{Z}}\bar{W}]^{-1}\bar{W}'M_{\bar{Z}}\bar{Y}^d,\\ \hat{\beta}_{b,j+1} &= (\bar{Z}'_{j+1}\bar{Z}_{j+1})^{-1}\bar{Z}'_{j+1}(\bar{Y}^d_{j+1} - \bar{W}_{j+1}\hat{\beta}_a). \qquad (4.18)\end{aligned}$$

Note that the choice of the estimate of the covariance matrix in Equations (4.17) to (4.18) will have no effect provided a consistent one is used. As Qu and Perron (2007) (supplement, p. 25−26) we can show for the first component of $F^j_{1,T}$ (or with obvious changes for the second component) that

$$\begin{aligned}&T_{j+1}\,\text{tr}((\Sigma^0)^{-1}\tilde{\Sigma}_{j+1})\\ &\quad= A'_T W'_{1,j+1}M_{Z_{1,j+1}}W_{1,j+1}A_T - U^{d'}_{1,j+1}P_{Z_{1,j+1}}U^d_{1,j+1}\\ &\quad\quad - 2(M_{Z_{1,j+1}}W_{1,j+1}A_T)'U^d_{1,j+1} + U'_{1,j+1}(I_T \otimes (\Sigma^0)^{-1})U_{1,j+1} + o_p(1),\end{aligned}$$

where $A_T = [W'M_Z W]^{-1}W'M_Z U^d$. For the third component of $F^j_{1,T}$ it can be shown that

$$\begin{aligned}&(T_{j+1} - T_j)\,\text{tr}((\Sigma^0)^{-1}\hat{\Sigma}_{j+1})\\ &\quad=\bar{A}'_T\bar{W}'_{j+1}M_{\bar{Z}_{j+1}}\bar{W}_{j+1}\bar{A}_T - \bar{U}^{d'}_{j+1}P_{\bar{Z}_{j+1}}\bar{U}^d_{j+1}\\ &\quad\quad - 2(M_{\bar{Z}_{j+1}}\bar{W}_{j+1}\bar{A}_T)'\bar{U}^d_{j+1} + U'_{j+1}(I_T \otimes (\Sigma^0)^{-1})U_{j+1} + o_p(1),\end{aligned}$$

where $\bar{A}_T = [\bar{W}'M_{\bar{Z}}\bar{W}]^{-1}\bar{W}'M_{\bar{Z}}\bar{U}^d$. Following the same arguments as in Bai and Perron (1998, p.75), we have $\text{plim}_{T\to\infty} T^{1/2}\bar{A}_T = \text{plim}_{T\to\infty} T^{1/2}A_T$ Hence, all terms that involve $\bar{A}_T$ and A_T eventually cancel and

$$F^j_{1,T} = U^{d'}_{1,j}P_{Z_{1,j}}U^d_{1,j} + U^{d'}_{j+1}P_{\bar{Z}_{j+1}}U^d_{j+1} - U^{d'}_{1,j+1}P_{Z_{1,j+1}}U^d_{1,j+1} + o_p(1).$$

Now, $T^{-d}Z'_{1,j}U^d_{1,j} \Rightarrow Q_b^{1/2}W^*_{D,p_b}(\lambda_i)$ and $T^{-1}\sum_{t=1}^{T_j} x_{bt}(\Sigma^0)^{-1}x'_{bt} \to^p \lambda_i Q_b$ where $W^*_{D,p_b}(\lambda_i)$ is a p_b vector of zeros and independent fractional Wiener processes defined on $[0,1]$ as given in Theorem 4.2 and where Q_b is the appropriate submatrix of Q that corresponds to the elements of x_{bt}. Hence,

$$T^{-2d}\,U^{d'}_{1,j+1}P_{Z_{1,j+1}}U^d_{1,j+1} \Rightarrow [W^*_{D,p_b}(\lambda_{j+1})'W^*_{D,p_b}(\lambda_{j+1})]/\lambda_{j+1}.$$

Using similar arguments

$$T^{-2d}\,U^{d'}_{1,j}P_{Z_{1,j}}U^d_{1,j} \Rightarrow [W^*_{D,p_b}(\lambda_j)'W^*_{D,p_b}(\lambda_j)]/\lambda_j$$

and

$$T^{-2d} U_{j+1}^{d'} P_{\bar{Z}_{j+1}} U_{j+1}^{d} \Rightarrow (W_{D,p_b}^*(\lambda_{j+1}) - W_{D,p_b}^*(\lambda_j))'(W_{D,p_b}^*(\lambda_{j+1}) - W_{D,p_b}^*(\lambda_j))/(\lambda_{j+1} - \lambda_j).$$

These results imply that the first component in (4.15) has the limit

$$F_{1,T}^{j} \Rightarrow \frac{(\lambda_j W_{D,p_b}^*(\lambda_{j+1}) - \lambda_{j+1} W_{D,p_b}^*(\lambda_j))'(\lambda_j W_{D,p_b}^*(\lambda_{j+1}) - \lambda_{j+1} W_{D,p_b}^*(\lambda_j))}{(\lambda_{j+1} - \lambda_j)\lambda_j \lambda_{j+1}}. \quad (4.19)$$

Consider now the limit of $\sum_{j=1}^{m} F_{2,T}^{j}$ when changes in Σ^0 are allowed. We have

$$
\begin{aligned}
F_{2,T}^{j} = & -\frac{1}{2}\sum_{j=1}^{m} \operatorname{tr}(T_{j+1}((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I)^2) \\
& - T_j((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j} - I)^2 - (T_{j+1} - T_j)((\Sigma^0)^{-1}\hat{\Sigma}_{j+1} - I)^2.
\end{aligned}
$$

Let $((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I)^F$ (”F” for full sample) be the matrix whose entries are those of $((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I)$ for the corresponding entries of Σ^0 that are not allowed to vary across regimes; the remaining entries are filled with zeros. Then

$$
\left[((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I)^F\right]_{i,k} = \frac{\sigma^{ik}}{T}\sum_{t=1}^{T}(y_{it} - x_{it}'\tilde{\beta})'(y_{kt} - x_{kt}'\tilde{\beta}) - I_{i,k},
$$

where σ^{ik} is the (i,k) element of $(\Sigma^0)^{-1}$ and $I_{i,k}$ is the (i,k) element of I. Also let $((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I)^S$ (”S” for relevant segments) be the matrix whose entries are those of $((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I)$ for the corresponding entries of Σ^0 that are allowed to vary across regimes, the remaining entries being filled with zeros. Then

$$
\left[((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I)^S\right]_{i,k} = \frac{\sigma^{ik}}{T_{j+1}}\sum_{t=1}^{T_{j+1}}(y_{it} - x_{it}'\tilde{\beta})'(y_{kt} - x_{kt}'\tilde{\beta}) - I_{i,k}.
$$

Note that the entries for $((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I)^F$ are the same for all segments. Define similarly $((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j} - I)^F$, $((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j} - I)^S$, $((\Sigma^0)^{-1}\hat{\Sigma}_{j+1} - I)^F$ and $((\Sigma^0)^{-1}\hat{\Sigma}_{j+1} - I)^S$. Then

$$
\begin{aligned}
((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I) &= ((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I)^F + ((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I)^S, \\
((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j} - I) &= ((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j} - I)^F + ((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j} - I)^S, \\
((\Sigma^0)^{-1}\hat{\Sigma}_{j+1} - I) &= ((\Sigma^0)^{-1}\hat{\Sigma}_{j+1} - I)^F + ((\Sigma^0)^{-1}\hat{\Sigma}_{j+1} - I)^S,
\end{aligned}
$$

and, in view of (4.16),

$$
\begin{aligned}
\sum_{j=1}^{m} F_{2,T}^{j} = & -\frac{1}{2}\operatorname{tr}(\sum_{j=1}^{m}[T_{j+1}((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I)^S((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I)^S \\
& - T_j((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j} - I)^S((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j} - I)^S \\
& - (T_{j+1} - T_j)((\Sigma^0)^{-1}\hat{\Sigma}_{j+1}^{S} - I)^S((\Sigma^0)^{-1}\hat{\Sigma}_{j+1}^{S} - I)^S]) \\
& + o_p(1)
\end{aligned}
$$

Now, because $\tilde{\beta} - \beta^0 = O_p(T^{-1/2+d})$, we have

$$\frac{T_{j+1}}{T^{2d}}((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I)^S((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I)^S$$

$$= \frac{T}{T_{j+1}} \left(T^{-1/2-d} \sum_{t=1}^{T_{j+1}} [(\Sigma^0)^{-1} u_t u_t' - I] \right)^S \left(T^{-1/2-d} \sum_{t=1}^{T_{j+1}} [(\Sigma^0)^{-1} u_t u_t' - I] \right)^S + o_p(1)$$

$$\Rightarrow \frac{\xi_n^d(\lambda_{j+1})^S \xi_n^d(\lambda_{j+1})^S}{\lambda_{j+1}}$$

$$\frac{T_j}{T^{2d}}((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j} - I)^S((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j} - I)^S$$

$$= \frac{T}{T_j} \left(T^{-1/2-d} \sum_{t=1}^{T_j} [(\Sigma^0)^{-1} u_t u_t' - I] \right)^S \left(T^{-1/2-d} \sum_{t=1}^{T_j} [(\Sigma^0)^{-1} u_t u_t' - I] \right)^S + o_p(1)$$

$$\Rightarrow \frac{\xi_n^d(\lambda_j)^S \xi_n^d(\lambda_j)^S}{\lambda_j}$$

and

$$\begin{aligned}
&\frac{(T_{j+1} - T_j)}{T^{2d}}(\Sigma^0)^{-1}\hat{\Sigma}_{j+1}^S - I)^S((\Sigma^0)^{-1}\hat{\Sigma}_{j+1}^S - I)^S \\
&= \frac{T}{T_{j+1} - T_j} \left(T^{-1/2-d} \sum_{t=T_j+1}^{T_{j+1}} [(\Sigma^0)^{-1} u_t u_t' - I] \right)^S \\
&\quad \times \left(T^{-1/2-d} \sum_{t=T_j+1}^{T_{j+1}} [(\Sigma^0)^{-1} u_t u_t' - I] \right)^S + o_p(1) \\
&\Rightarrow \frac{(\xi_{D,n}^*(\lambda_{j+1}) - \xi_{D,n}^*(\lambda_j))^S (\xi_{D,n}^*(\lambda_{j+1}) - \xi_{D,n}^*(\lambda_j))^S}{\lambda_{j+1} - \lambda_j}
\end{aligned}$$

where $\xi_D^*(\cdot)$ is an $n \times n$ matrix whose elements are

$$[\xi_D^*(\cdot)]_{i,j} = \begin{cases} [\xi_D(\cdot)]_{i,j}, & \text{if } d_i = d_j = \max_{1 \le k \le n} d_k, \\ 0, & \text{else,} \end{cases}$$

and where ξ_D is (nonstandard) fractional Brownian motions defined on $[0, 1]$ such that $\text{Var}(\text{vec}(\xi_D(1))) = \Omega$ (which follows from Theorem 4.8.2 of Giraitis et al. (2012)

p.109). Hence,

$$\begin{aligned}\sum_{j=1}^{m} F_{2,T}^{j} \Rightarrow & -\frac{1}{2}\operatorname{tr}\left(\frac{\xi_{D,n}^{*}(\lambda_{j+1})^{S}\xi_{D,n}^{*}(\lambda_{j+1})^{S}}{\lambda_{j+1}} - \frac{\xi_{D,n}^{*}(\lambda_{j})^{S}\xi_{D,n}^{*}(\lambda_{j})^{S}}{\lambda_{j}}\right. \\ & \left. + \frac{(\xi_{D,n}^{*}(\lambda_{j+1}) - \xi_{D,n}^{*}(\lambda_{j}))^{S}(\xi_{D,n}^{*}(\lambda_{j+1}) - \xi_{D,n}^{*}(\lambda_{j}))^{S}}{\lambda_{j+1}-\lambda_{j}}\right) \\ = & -\frac{1}{2}\left[\frac{\operatorname{vec}(\xi_{D,n}^{*}(\lambda_{j+1})^{S})' \operatorname{vec}(\xi_{D,n}^{*}(\lambda_{j+1})^{S})}{\lambda_{j+1}}\right. \\ & - \frac{\operatorname{vec}(\xi_{D,n}^{*}(\lambda_{j})^{S})' \operatorname{vec}(\xi_{D,n}^{*}(\lambda_{j})^{S})}{\lambda_{j}} \\ & \times \left(\operatorname{vec}(\xi_{D,n}^{*}(\lambda_{j+1})^{S}) - \operatorname{vec}(\xi_{D,n}^{*}(\lambda_{j})^{S})\right)' \\ & \left. \times \frac{(\operatorname{vec}(\xi_{D,n}^{*}(\lambda_{j+1})^{S}) - \operatorname{vec}(\xi_{D,n}^{*}(\lambda_{j})^{S})}{(\lambda_{i+1}-\lambda_{i})}\right]\end{aligned}$$

using the fact that $\operatorname{tr}(AA) = \operatorname{vec}(A)' \operatorname{vec}(A)$ for a symmetric matrix A. Now let H be the matrix that selects the elements of $\operatorname{vec}(\Sigma^0)$ that are allowed to change. Then

$$\begin{aligned}\operatorname{vec}(\xi_{D,n}^{*}(\lambda_{j+1})^{S})' \operatorname{vec}(\xi_{D,n}^{*}(\lambda_{j+1})^{S}) &= \operatorname{vec}(\xi_{D,n}^{*}(\lambda_{j+1}))' H' H \operatorname{vec}(\xi_{D,n}^{*}(\lambda_{j+1})) \\ &\overset{d}{=} W_{D,n_b^*}^{*}(\lambda_{j+1})' H\Omega H' W_{D,n_b^*}^{*}(\lambda_{j+1}),\end{aligned}$$

where $W_{D,n_b^*}^{*}$ is an n_b^* vector of processes as defined in Theorem 4.2. Hence, we have

$$\begin{aligned}\sum_{j=1}^{m} F_{2,T}^{j} \Rightarrow & -\frac{1}{2}\left[\frac{W_{D,n_b^*}^{*}(\lambda_{j+1})' H'\Omega H W_{D,n_b^*}^{*}(\lambda_{j+1})}{\lambda_{j+1}} - \frac{W_{D,n_b^*}^{*}(\lambda_{j})' H'\Omega H W_{D,n_b^*}^{*}(\lambda_{j})}{\lambda_{j}}\right. \\ & \left. - \frac{(W_{D,n_b^*}^{*}(\lambda_{j+1}) - W_{D,n_b^*}^{*}(\lambda_{j}))' H'\Omega H (W_{D,n_b^*}^{*}(\lambda_{j+1}) - W_{D,n_b^*}^{*}(\lambda_{j})}{\lambda_{j+1}-\lambda_{j}}\right] \\ = & (\lambda_j W_{D,n_b^*}^{*}(\lambda_{j+1}) - \lambda_{j+1} W_{D,n_b^*}^{*}(\lambda_{j}))' H'\Omega H \\ & \times (\lambda_j W_{D,n_b^*}^{*}(\lambda_{j+1}) - \lambda_{j+1} W_{D,n_b^*}^{*}(\lambda_{j}))/(\lambda_j\lambda_{j+1}(\lambda_{j+1}-\lambda_j)).\end{aligned} \tag{4.20}$$

By combining equations (4.19) and (4.20) we have shown the limiting distribution of our test. □

4.8.6 Proof of Theorem 4.3

Proof. From Theorem 4.1 we have the consistency of our break point estimates at each iteration. If we have m^0 break points in the data the break point test of Theorem 4.2 rejects in each iteration $m < m^0$ with a probability tending to one for $T \to \infty$ due to the Pitman efficiency of the test. In iteration m^0 the test has a type-I error of α and thus the hit rate of our procedure is $(1-\alpha)\%$. □

4.8.7 Proof of 10 Properties

This section contains proofs for properties of the quasi-likelihood ratio and parameter estimates.

Property 1. For each $\delta \in (0,1]$

$$\sup_{T\delta\leq k\leq T} \mathcal{L}(1,k;\hat{\beta}_{(k)},\hat{\Sigma}_{(k)}) = O_p(1),$$
$$\sup_{T\delta\leq k\leq T} (\|\hat{\beta}_{(k)}-\beta_0\| + \|\hat{\Sigma}_{(k)}-\Sigma_0\|) = O_p(T^{d-1/2}).$$

Proof. The strong consistency of $(\hat{\beta}_{(k)},\hat{\Sigma}_{(k)})$ follows using the arguments of Qu and Perron (2007). Then we can write

$$\hat{\beta}_{(k)}-\beta_0 = \left(\sum_{t=)1}^{k} x_t\hat{\Sigma}_{(k)}^{-1}x_t'\right)^{-1}\sum_{t=1}^{k} x_t\hat{\Sigma}_{(k)}^{-1}u_t$$

and apply the generalized Hájek-Rényi inequality on $\sum_{t=1}^{k} x_t(\Sigma_0)^{-1}u_t$. Together with the strong consistency of $\hat{\Sigma}_{(k)}$ this gives $\sup_{T\delta\leq k\leq T}\|\hat{\beta}_{(k)}-\beta_0\| = O_p(T^{-1/2+d})$.

Furthermore, we have

$$\hat{\Sigma}_{(k)}-\Sigma_0 = \frac{1}{k}\sum_{t=1}^{k}(u_t - x_t'(\hat{\beta}_{(k)}-\beta_0))(u_t - x_t'(\hat{\beta}_{(k)}-\beta_0))^1 - \Sigma_0.$$

Applying again the generalized Hájek-Rényi inequality gives $\sup_{T\delta\leq k\leq T}\|\hat{\Sigma}_{(k)}-\Sigma_0\| = O_p(T^{-1/2+d})$. As a direct consequence this yields $\sup_{T\delta\leq k\leq T}\mathcal{L}(1,k;\hat{\beta}_{(k)},\hat{\Sigma}_{(k)}) = O_p(1)$. □

The following property is modified compared to property 2 of Qu and Perron (2007). Instead of considering the supremum of the likelihoood over $1\leq k\leq T$ we consider here the supremum over $\delta T\leq k\leq T$ for some $\delta\in(0,1)$.

Property 2. For some $\delta\in(0,1)$, each $\varepsilon>0$, there exists a $B>0$ such that

$$Pr\left(\sup_{\delta T\leq k\leq T} T^{-B}\mathcal{L}(1,k;\hat{\beta}_{(k)},\hat{\Sigma}_{(k)}) > 1\right) < \varepsilon$$

for all large T.

Proof. This is a direct consequence of property 1. □

Property 3. Let $S_T = \{(\beta,\Sigma)\colon \|\beta-\beta^0\| \geq T^{-1/2+d}\log T \text{ or } \|\Sigma-\Sigma^0\| \geq T^{-1/2+d}\log T\}$. For any $\delta\in(0,1), D>0$ and $\varepsilon>0$ the following statement holds when T is large:

$$\Pr\left(\sup_{k\geq\delta T}\sup_{(\beta,\Sigma)\in S_T} T^D\mathcal{L}(1,k;\beta,\Sigma) > 1\right) < \varepsilon.$$

Proof. We proceed in two steps: First we consider the behaviour of the likelihood function over a compact set and show that the claim is true. Second we argue why this is still true once we remove the requirement of a compact parameter subset. Define

$$\overline{\Theta}_2 = \{(\beta,\Sigma)\colon \|\beta\|\leq d_1, \lambda_{\min}(\Sigma)\geq d_2, \lambda_{\max}(\Sigma)\leq d_3\},$$

where $\lambda_{\min}$ and $\lambda_{\max}$ denote the smallest and largest eigenvalues of Σ and the finite constants d_1, d_2 and d_3 are chosen in such a way that (β_0,Σ_0) is an inner point of

$\overline{\Theta}_2$. As explained we first show (4.9) with the second supremum taken over $S_T \cap \overline{\Theta}_2$ which is compact. We decompose the segmential log likelihood as $\log \mathcal{L}(1,k;\beta,\Sigma) = \mathcal{L}_{1,T} + \mathcal{L}_{2,T}$, where

$$\mathcal{L}_{1,T} = -\frac{k}{2}\log|I+\Psi_T| - \frac{k}{2}\left[\frac{1}{k}\sum_{t=1}^{k}\eta_t'(I+\Psi_T)^{-1}\eta_t - \frac{1}{k}\sum_{t=1}^{k}\eta_t'\eta_t\right]$$

and

$$\mathcal{L}_{2,T} = \beta^{*\prime}\sum_{t=1}^{k}x_t\Sigma^{-1}u_t - \frac{k}{2}\beta^{*\prime}\left(\frac{1}{k}\sum_{t=1}^{k}x_t\Sigma^{-1}x_t'\right)\beta^*,$$

where $\beta^* = \beta - \beta_0, \Sigma^* = \Sigma - \Sigma_0, \eta_t = (\Sigma_0)^{-1}u_t$ and $\Psi_T = (\Sigma_0)^{-1/2}\Sigma^*(\Sigma_0)^{-1/2}$. We note that only $\mathcal{L}_{2,T}$ depends on β^*. We split the parameter space $S_T = S_{1,T} \cup S_{2,T}$ with

$$S_{1,T} = \{(\beta,\Sigma)\colon \|\Sigma - \Sigma_0\| \geq T^{-1/2+d}\log T,\ \beta \text{ arbitrary}\}$$

and

$$S_{2,T} = \{(\beta,\Sigma)\colon \|\beta - \beta_0\| \geq T^{-1/2+d}\log T \text{ and } \|\Sigma - \Sigma_0\| \leq T^{-1/2+d}\log T\}.$$

It has to be shown that

$$\Pr\left(\sup_{k\geq T\delta}\ \sup_{(\beta,\Sigma)\in S_{1,T}\cap\overline{\Theta}_2} T^D\mathcal{L}(1,k;\beta,\Sigma) > 1\right) < \varepsilon \tag{S.3}$$

and

$$\Pr\left(\sup_{k\geq T\delta}\ \sup_{(\beta,\Sigma)\in S_{2,T}\cap\overline{\Theta}_2} T^D\mathcal{L}(1,k;\beta,\Sigma) > 1\right) < \varepsilon. \tag{S.4}$$

We start to show (S.3). On $S_{1,T}$, $\mathcal{L}_{2,T}$ is a quadratic function of β^* and has maximum value

$$\sup_{S_{1,T}}\mathcal{L}_{2,T} = \frac{k}{2}\left(\frac{1}{k}\sum_{t=1}^{k}x_t\Sigma^{-1}u_t\right)'\left(\frac{1}{k}\sum_{t=1}^{k}x_t\Sigma^{-1}x_t'\right)^{-1}\left(\frac{1}{k}\sum_{t=1}^{k}x_t\Sigma^{-1}u_t\right).$$

Applying Property 1 gives

$$\sup_{k\geq T\delta}\sup_{\overline{\Theta}_2}\left\|\left(\frac{1}{k}\sum_{t=1}^{k}x_t\Sigma^{-1}x_t'\right)^{-1}\right\| = O_p(1).$$

Additionally we see

$$\begin{aligned}\sup_{k\geq T\delta}\left\|\frac{1}{k}\sum_{t=1}^{k}x_t\Sigma^{-1}u_t\right\| &= \sup_{k\geq T\delta}\left\|\frac{1}{k}\sum_{t=1}^{k}S'(I_n\otimes z_t)\Sigma^{-1}u_t\right\| \\ &= \sup_{k\geq T\delta}\left\|S'(\Sigma^{-1}\otimes I_n)\frac{1}{k}\sum_{t=1}^{k}(I_n\otimes z_t)u_t\right\| \\ &\leq \sup_{k\geq T\delta}\left\|\frac{1}{k}\sum_{t=1}^{k}(I_n\otimes z_t)u_t\right\|\left\|S'(\Sigma^{-1}\otimes I_n)\right\|.\end{aligned}$$

From the FCLT of Lemma 4.4 we have for fixed $r > 0$

$$\lim_{T\to\infty} \Pr\left(\sup_{k\geq T\delta}\left\|\frac{1}{k}\sum_{t=1}^{k}(I_n\otimes z_t)u_t\right\| > rT^{d-1/2}\log^{1/2}T\right) = 0,$$

while $\|S'(\Sigma^{-1}\otimes I_n)\| = \sum_{i=1}^{n}(1+\lambda_i)^{-1}O_p(1)$, where λ_i $(i = 1, ..., n)$ are the eigenvalues of $(\Sigma_0)^{-1/2}\Sigma^*(\Sigma_0)^{-1/2}$. Hence,

$$\sup_{k\geq T\delta}\sup_{S_{1,T}\cap\overline{\Theta}_2}\mathcal{L}_{2,T} \leq \frac{k}{2}\left(\sum_{i=1}^{n}\frac{1}{1+\lambda_i}\right)^2 (r^2T^{2d-1}\log T),$$

which implies

$$\sup_{k\geq T\delta}\sup_{S_{1,T}\cap\bar{\Theta}_2}\mathcal{L}_{2,T} \leq \frac{k}{2}\sum_{i=1}^{n}\frac{1}{1+\lambda_i}r^2b_T^2,$$

where $b_T = T^{d-1/2}\log T$ with the inequality holding with probability arbitrarily close to 1 for large T. For $\mathcal{L}_{1,T}$ we start by considering the term in brackets. Introduce an orthogonal matrix U that diagonalizes $(I+\Psi_T)^{-1}$. Then we have

$$\frac{1}{k}\sum_{t=1}^{k}\eta_t'((I+\Psi_T)^{-1}-I)\eta_t = \operatorname{tr}\left(\operatorname{diag}\left\{\frac{1}{1+\lambda_i}-1\right\}\left(\frac{1}{k}U\sum_{t=1}^{k}\eta_t\eta_t'U'\right)\right).$$

Because $\|U\| = 1$ it suffices to investigate

$$\left\|\frac{1}{k}U\sum_{t=1}^{k}\eta_t\eta_t'U' - I\right\| \leq \frac{1}{k}\left\|\sum_{t=1}^{k}(\eta_t\eta_t' - I)\right\|.$$

Then for any $a > 0$ by the FCLT of Lemma 4.4

$$\lim_{T\to\infty}\Pr\left(\sup_{k\geq T\delta}\frac{1}{k}\sum_{t=1}^{k}\|(\eta_t\eta_t' - I)\| > ab_T\right) = 0.$$

Then arguing as Bai, Lumsdaine, et al. (1998) we may show that

$$\sup_{k\geq T\delta}\sup_{S_{1,T}\cap\overline{\Theta}_2}\mathcal{L}_{1,T} \leq -\frac{k}{2}\left[\sum_{i=1}^{n}\left(\log(1+\lambda_i) + \left(\frac{1}{1+\lambda_i}-1\right)(1+\operatorname{sign}(\lambda_i)ab_T)\right)\right]$$

with probability arbitrarily close to 1 for large T, where a is a fixed positive number which can be made arbitrarily small. Combining the preceding two inequalities we can show that

$$\Pr\left(\sup_{k\geq T\delta}\sup_{(\beta,\Sigma)\in S_{1,T}\cap\overline{\Theta}_2}\mathcal{L}_{1,T}+\mathcal{L}_{2,T} > -D\log T\right) < \varepsilon.$$

It is now straightforward to see that using the similar arguments as Bai, Lumsdaine, et al. (1998) one can show that equation (S.4) holds. Therefore the claim is shown on the compact parameter space $\overline{\Theta}_2$. But as in Qu and Perron (2007) we can conclude that the result is valid also on an unrestricted parameter space. Therefore the proof is complete. □

Property 4. Not needed.

The following property is different from Qu and Perron (2007) in that we do not assume that the limit of $(h_T d_T^2)/T$ exists. Instead as pointed out by Bai (2000) we assume the sufficient condition that $\liminf_{T\to\infty}(h_T d_T^2)/T \geq h > 0$.

Property 5. Let h_T and d_T be positive sequences such that h_T is nondecreasing, $d_T \to \infty$ and $\liminf_{T\to\infty}(h_T d_T^2)/T \geq h > 0$. Define $\overline{\Theta}_3 = \{(\beta, \Sigma)\colon \|\beta\| \leq p_1, \lambda_{\min}(\Sigma) \geq p_2, \lambda_{\max}(\Sigma) \leq p_3\}$, where p_1, p_2 and p_3 are arbitrary constants that satisfy $p_1 < \infty$, $0 < p_2 \leq p_3 < \infty$. Define $S_T = \{(\beta, \Sigma)\colon \|\beta - \beta^0\| \geq T^{-1/2+d}\log T$ or $\|\Sigma - \Sigma^0\| \geq T^{-1/2+d}\log T\}$. Then, for any $\varepsilon > 0$, there exists an $A > 0$, such that

$$\Pr\left(\sup_{k\geq Ah_T}\ \sup_{(\beta,\Sigma)\in S_T\cap\bar{\Theta}_3} \mathcal{L}(1,k;\beta,\Sigma) > \varepsilon\right) < \varepsilon$$

when T is large.

Proof. As in Property 3 we only need to look at the behaviour of L_{2T} over $S_{1,T}\cap\bar{\Theta}_3$. The rest of the proof is as in Bai, Lumsdaine, et al. (1998). We need to show

$$P(\sup_{k\geq Ah_T}\ \sup_{(\beta,\Sigma)\in S_{1,T}\cap\bar{\Theta}_3} \mathcal{L}(1,k;\beta,\Sigma) > \varepsilon) < \varepsilon$$

or

$$P(\sup_{k\geq Ah_T}\ \sup_{(\beta,\Sigma)\in S_{1,T}\cap\bar{\Theta}_3} \mathcal{L}_{1T} + \mathcal{L}_{2T} > \varepsilon) < \varepsilon.$$

Define $b_T := T^{-1/2}d_T$. Now all the arguments in the proof of Property 3 still hold. Thus, we have

$$\sup_{S_{1,T}} \mathcal{L}_{2T} = \frac{k}{2}\left(\frac{1}{k}\sum_{t=1}^{k} x_t\Sigma^{-1}u_t\right)'\left(\frac{1}{k}\sum_{t=1}^{k} x_t\Sigma^{-1}x_t'\right)^{-1}\left(\frac{1}{k}\sum_{t=1}^{k} x_t\Sigma^{-1}u_t\right),$$

where

$$\left(\sum_{t=1}^{k} x_t\Sigma^{-1}x_t'\right)^{-1} = \left(\sum_{t=1}^{k} S'(I\otimes z_t)\Sigma^{-1}(I\otimes z_t')S\right)^{-1} = \left(S'(\Sigma^{-1}\otimes\sum_{t=1}^{k} z_t z_t')S\right)^{-1}.$$

From $l^{-1}\sum_{t=1}^{l} z_t z_t' \xrightarrow{a.s.} Q_z$, for a given $\varepsilon > 0$ we can always find a $k_1 > 0$ such that

$$P(\sup_{k\geq k_1}\|\frac{1}{k}\sum_{t=1}^{k} z_t z_t' - Q_z\| > \varepsilon) < \varepsilon.$$

Define $Q_\triangle := k^{-1}\sum_{t=1}^{k} z_t z_t' - Q_z$. Then

$$\begin{aligned}
&(S'(\Sigma^{-1}\otimes\frac{1}{k}\sum_{t=1}^{k} z_t\ z_t')S)^{-1} - (S'(\Sigma^{-1}\otimes Q_z)S)^{-1}\\
=&(S'(\Sigma^{-1}\otimes Q_z)S + S'(\Sigma^{-1}\otimes Q_\triangle)S)^{-1} - (S'(\Sigma^{-1}\otimes Q_z)S)^{-1}\\
=&-A^{-1}B(A+B)^{-1},
\end{aligned}$$

where $A = S'(\Sigma^{-1} \otimes Q_z)S$ and $B = S'(\Sigma^{-1} \otimes Q_{\triangle})S$. Because Σ^{-1} has uniformly bounded eigenvalues and $k^{-1}\sum_{t=1}^{k} z_t z_t'$ is positive definite for large k, A^{-1} and B^{-1} have bounded eigenvalues. Because B is uniformly small, $-A^{-1}B(A+B)^{-1}$ is uniformly small for large k. This is

$$(S'(\Sigma^{-1} \otimes k^{-1}\sum_{t=1}^{k} z_t z_t')S)^{-1} - (S'(\Sigma^{-1} \otimes Q_z)S)^{-1} \stackrel{\text{a.s.}}{=} o(1) \qquad \text{as} \quad k \to \infty.$$

Now there exists an $M > 0$ such that $\sup_{(\beta,\Sigma)\in S_{1,T}\cap\bar{\Theta}_3} |(S'(\Sigma^{-1} \otimes Q_z)S)^{-1}| < M$, and we have, for any $\varepsilon > 0$, that there exists an $A > 0$ such that

$$P(\sup_{k\geq Ah_T} \sup_{(\beta,\Sigma)\in S_{1,T}\cap\bar{\Theta}_3} \|(\frac{1}{k}\sum_{t=1}^{k} x_t \Sigma^{-1} x_t')^{-1}\| > 2M) < \varepsilon.$$

Now,

$$\begin{aligned}\sup_{k\geq Ah_T} \|\frac{1}{k}\sum_{t=1}^{k} x_t \Sigma^{-1} u_t\| &= \sup_{k\geq Ah_T} \|\frac{1}{k}\sum_{t=1}^{k} S'(I_n \otimes z_t)\Sigma^{-1} u_t\| \\ &\leq \sup_{k\geq Ah_T} \|\frac{1}{k}\sum_{t=1}^{k} (I_n \otimes z_t) u_t\| \|S'(\Sigma^{-1} \otimes I_n)\|. \qquad (4.21)\end{aligned}$$

From Lemma 4.3 we have

$$P(\sup_{k\geq Ah_T} \|\frac{1}{k}\sum_{t=1}^{k} (I_n \otimes z_t) u_t\| > ab_T) \leq \frac{C_1}{Ah_T a^2 b_T} < \frac{2C_1}{Aa^2 h}$$

for some $C_1 > 0$, where the bound can be made arbitrarily small by choosing a large A. For the second component,

$$\|S'(\Sigma^{-1} \otimes I_n)\| \leq nC_2 \sum_{i=1}^{n} \frac{1}{1+\lambda_i} \qquad (4.22)$$

for some $0 < C_2 < \infty$, which depends on the matrix S. Now, combining (4.21)-(4.22), we have, for any $\varepsilon > 0$ that there exists an $\bar{A} > 0$, such that with probability no less than $1 - \varepsilon$,

$$\sup_{k\geq \bar{A}h_T} \sup_{(\beta,\Sigma)\in S_{1,T}\cap\bar{\Theta}_3} |L_{2T}| < ka^2 b_T^2 n^2 C_2^2 M (\sum_{i=1}^{n} \frac{1}{1+\lambda_i})^2 \leq \frac{k}{2}\sum_{i=1}^{n} \frac{Ga^2 b_T^2}{1+\lambda_i} = \frac{k}{2}\sum_{i=1}^{n} \frac{\gamma^2 b_T^2}{1+\lambda_i}$$

with $G = 2n^3 C_2^2 M/p_2$. Because a^2 can be made arbitrarily small by choosing a large A, so can y^2. Hence Property 5 follows. □

The next properties are the same as Lemmas 6 − 10 of Bai (2000). Because the proofs are similar, they are omitted.

Property 6. With ν_T satisfying Assumption 6, for each β and Σ such that $\|\beta - \beta_0\| \leq Mv_T$ and $\|\Sigma - \Sigma_0\| \leq Mv_T$, with $M < \infty$, we have

$$\sup_{1\leq k\leq T^{1/2-d}v_T^{-1}} \sup_{\lambda,\Xi} \frac{\mathcal{L}(1,k;\beta + T^{-1/2+d}\lambda, \Sigma + T^{-1/2+d}\Xi)}{\mathcal{L}(1,k;\beta,\Sigma)} = o_p(1).$$

Property 7. Under the conditions of Property 6, we have

$$\sup_{1\le k\le Mv_T^{-2}} \sup_{\lambda,\Xi} \log \frac{\mathcal{L}(1,k;\beta+T^{-1/2+d}\lambda,\Sigma+T^{-1/2+d}\Xi)}{\mathcal{L}(1,k;\beta,\Sigma)} = o_p(1).$$

Property 8. We have

$$\sup_{T\delta\le k\le T} \sup_{\beta^*,\Sigma^*,\lambda,\Xi} \log \frac{\mathcal{L}(1,k;\beta_0+T^{-1/2+d}\beta^*+T^{-1+2d}\lambda,\Sigma_0+T^{-1/2+d}\Sigma^*+T^{-1+2d}\Xi)}{\mathcal{L}(1,k;\beta_0+T^{-1/2+d}\beta^*,\Sigma_0+T^{-1/2+d}\Sigma^*)} = o_p(1),$$

where the supremum with respect to $\beta^*, \Sigma^*, \lambda, \Xi$ is taken over an arbitrary compact set.

Property 9. Let $T_1 = [aT]$ for some $a \in (0,1]$ and let $T_2 = [T^{1/2-d}v_T^{-1}]$, where v_T satisfies Assumption 6. Consider

$$\begin{aligned} y_t &= x_t'\beta_1^0 + \Sigma_1^0\eta_t, && (t = 1,\ldots,T_1), \\ y_t &= x_t'\beta_2^0 + \Sigma_2^0\eta_t, && (t = T_1+1,\ldots,T_1+T_2), \end{aligned}$$

where $\|\beta_1^0 - \beta_2^0\| \le Mv_T$ and $\|\Sigma_1^0 - \Sigma_2^0\| \le Mv_T$ for some $M < \infty$. Let $k = T_1 + T_2$ be the size of the pooled sample and let $(\hat{\beta}_n, \hat{\Sigma}_n)$ be the associated estimates. Then $\hat{\beta}_n - \beta_1^0 = O_p(T^{d-1/2})$ and $\hat{\Sigma}_n - \Sigma_1^0 = O_p(T^{d-1/2})$.

Property 10. Not needed.

4.8.8 Critical values of the $UDmax\, LR_T$ test

	90%												
m/d	-0.49	-0.48	-0.46	-0.44	-0.42	-0.4	-0.38	-0.36	-0.34	-0.32	-0.3	-0.28	-0.26
1	8.410	17.846	26.491	34.309	39.193	35.994	39.441	37.912	37.097	36.898	30.827	30.891	28.523
2	24.276	44.007	72.370	93.699	101.748	105.425	118.393	95.369	96.470	89.007	76.508	81.029	71.999
3	34.579	61.969	104.355	130.856	145.328	146.919	145.735	131.450	135.958	124.616	104.839	108.244	96.204
4	40.347	75.998	120.862	150.517	170.464	171.259	169.404	153.776	157.105	143.193	126.115	124.425	110.262
5	46.367	79.971	133.176	164.134	191.383	189.105	182.596	171.821	175.159	157.489	143.972	137.865	120.730
	-0.24	-0.22	-0.2	-0.18	-0.16	-0.14	-0.12	-0.1	-0.08	-0.06	-0.04	-0.02	0
1	27.775	23.931	51.253	21.261	16.834	18.119	15.610	13.599	12.198	10.345	10.935	10.199	7.848
2	64.060	57.967	67.182	48.206	40.386	35.945	34.010	29.195	26.547	23.113	21.761	19.792	15.505
3	85.440	78.278	76.534	61.710	53.235	52.144	44.940	38.279	33.728	30.242	27.125	25.244	21.152
4	98.523	88.057	92.015	71.680	62.012	58.889	50.653	42.137	38.695	34.269	31.291	28.241	23.645
5	110.394	97.934	92.015	78.337	67.372	62.605	56.043	47.731	41.459	37.547	33.179	31.488	25.394
	0.02	0.04	0.06	0.08	0.1	0.12	0.14	0.16	0.18	0.2	0.22	0.24	0.26
1	7.846	6.954	6.557	5.856	5.233	4.911	4.094	4.095	3.576	3.205	2.787	2.518	2.362
2	14.977	12.956	11.182	9.573	9.155	8.411	6.695	6.517	5.412	5.108	4.225	3.781	3.359
3	19.608	16.915	14.468	12.606	11.838	10.411	8.073	7.720	6.824	6.213	5.161	4.598	4.022
4	21.508	18.819	16.197	14.229	13.254	11.335	8.819	8.696	7.567	6.789	5.572	4.961	4.334
5	22.755	20.164	17.649	15.519	13.629	12.200	9.652	9.381	7.968	6.996	5.940	5.264	4.543
	0.28	0.3	0.32	0.34	0.36	0.38	0.4	0.42	0.44	0.46	0.48	0.49	
1	2.050	1.698	1.556	1.430	1.071	0.894	0.769	0.612	0.433	0.245	0.118	0.059	
2	3.045	2.447	2.240	2.008	1.469	1.238	1.026	0.780	0.557	0.326	0.201	0.074	
3	3.465	2.851	2.496	2.218	1.657	1.383	1.170	0.911	0.622	0.360	0.218	0.086	
4	3.853	3.100	2.727	2.429	1.866	1.510	1.277	0.982	0.659	0.390	0.225	0.092	
5	4.089	3.269	2.809	2.576	1.896	1.598	1.350	1.019	0.679	0.393	0.234	0.096	

Table 4.2: Asymptotic critical values for the $UDmax\, LR_T$ test, using $\epsilon = 0.05$. These are obtained from 10,000 Monte Carlo replications with $1,000$ increments.

	95%												
m/d	-0.49	-0.48	-0.46	-0.44	-0.42	-0.4	-0.38	-0.36	-0.34	-0.32	-0.3	-0.28	-0.26
1	10.126	20.489	31.669	39.739	46.815	43.505	44.244	41.594	43.839	44.949	38.438	35.138	33.807
2	27.227	48.129	79.026	99.144	111.514	113.695	125.206	107.440	109.675	102.952	87.820	85.306	78.340
3	36.716	66.400	110.651	138.349	160.067	156.946	160.882	147.343	151.418	139.355	117.222	119.344	104.340
4	43.673	78.056	131.583	162.124	182.726	183.895	186.022	172.254	171.460	157.489	137.487	139.132	120.228
5	47.814	84.137	145.187	174.638	212.650	202.412	202.834	188.747	186.820	172.324	163.417	151.156	130.191
	-0.24	-0.22	-0.2	-0.18	-0.16	-0.14	-0.12	-0.1	-0.08	-0.06	-0.04	-0.02	0
1	32.781	27.874	51.253	23.557	22.012	19.800	17.962	16.039	14.821	12.274	13.550	12.221	9.251
2	71.130	64.897	67.182	51.598	46.696	41.533	39.718	31.871	29.452	26.338	24.790	21.774	17.878
3	92.669	84.269	76.534	67.133	60.688	54.381	49.486	41.422	36.861	33.659	30.639	27.085	22.343
4	108.611	97.592	92.015	77.021	67.439	62.084	54.301	47.102	42.486	38.406	33.979	30.485	24.647
5	118.056	104.760	92.015	84.535	71.714	66.745	60.788	54.403	45.939	40.872	36.741	34.249	26.749
	0.02	0.04	0.06	0.08	0.1	0.12	0.14	0.16	0.18	0.2	0.22	0.24	0.26
1	8.711	8.717	7.536	6.765	6.162	6.046	4.873	4.658	4.065	3.843	3.466	3.451	2.858
2	17.725	15.480	13.012	11.595	10.331	9.496	7.880	7.220	6.562	5.863	4.996	4.666	3.884
3	21.508	19.152	16.213	13.972	12.852	11.413	9.127	8.485	7.695	6.671	5.880	5.783	4.421
4	23.171	20.711	17.974	15.002	14.723	12.469	10.143	9.385	8.320	7.302	6.427	5.989	4.944
5	25.497	22.891	19.263	16.671	14.806	13.379	10.782	10.043	8.841	8.001	7.129	6.097	5.196
	0.28	0.3	0.32	0.34	0.36	0.38	0.4	0.42	0.44	0.46	0.48	0.49	
1	2.432	2.088	1.946	1.645	1.548	1.144	0.938	0.868	0.571	0.343	0.198	0.071	
2	3.380	2.927	2.520	2.311	1.917	1.453	1.161	0.995	0.685	0.390	0.259	0.088	
3	3.990	3.336	2.860	2.731	2.083	1.698	1.344	1.136	0.782	0.434	0.323	0.096	
4	4.457	3.701	3.159	2.885	2.179	1.762	1.447	1.175	0.812	0.454	0.338	0.101	
5	4.605	3.847	3.330	2.914	2.239	1.844	1.497	1.208	0.829	0.469	0.344	0.104	

Table 4.3: Asymptotic critical values for the $UDmax\,LR_T$ test, using $\epsilon = 0.05$. These are obtained from 10,000 Monte Carlo replications with $1,000$ increments.

	99%												
m/d	-0.49	-0.48	-0.46	-0.44	-0.42	-0.4	-0.38	-0.36	-0.34	-0.32	-0.3	-0.28	-0.26
1	13.242	27.302	41.304	51.161	60.097	58.974	56.234	54.063	61.337	62.121	66.550	48.803	45.649
2	31.624	55.570	101.378	114.483	137.152	149.019	133.183	132.250	127.398	150.242	103.128	98.281	100.489
3	44.328	73.390	135.962	161.676	193.512	218.208	187.724	182.320	173.746	215.210	127.916	156.701	128.524
4	48.597	82.612	149.202	184.739	214.974	254.592	209.437	215.628	194.856	233.235	163.417	159.445	140.256
5	56.680	94.172	161.373	204.324	240.566	254.592	227.372	233.447	216.968	261.370	185.785	189.944	151.578
	-0.24	-0.22	-0.2	-0.18	-0.16	-0.14	-0.12	-0.1	-0.08	-0.06	-0.04	-0.02	0
1	41.889	38.796	51.253	28.561	28.390	26.587	23.533	20.137	19.944	18.043	16.801	15.665	13.096
2	85.590	76.726	67.182	58.089	55.900	53.785	44.964	37.432	34.804	31.151	28.722	26.991	21.998
3	108.814	117.656	86.158	81.578	74.409	61.891	55.258	46.322	42.761	40.785	34.559	31.962	27.031
4	130.189	117.656	93.050	92.660	84.612	73.290	66.677	55.068	51.097	44.669	40.851	38.593	31.099
5	144.329	133.404	103.493	97.502	92.949	77.167	75.029	63.593	55.843	52.113	44.947	44.315	33.184
	0.02	0.04	0.06	0.08	0.1	0.12	0.14	0.16	0.18	0.2	0.22	0.24	0.26
1	12.715	12.620	10.390	8.238	8.038	8.043	6.287	6.585	5.762	5.210	4.668	4.846	3.750
2	20.337	18.698	15.991	13.043	11.535	10.982	9.198	8.892	7.711	7.418	6.078	5.937	4.783
3	26.003	23.042	20.218	15.911	14.540	13.318	10.874	10.415	9.294	8.218	7.079	6.719	5.654
4	29.820	26.954	23.581	18.106	19.267	15.089	12.023	11.551	9.829	8.951	7.758	7.592	6.059
5	31.235	30.150	23.869	19.116	20.125	15.521	12.795	12.379	10.410	9.495	8.092	7.900	6.341
	0.28	0.3	0.32	0.34	0.36	0.38	0.4	0.42	0.44	0.46	0.48	0.49	
1	3.395	3.052	2.556	1.984	2.068	1.895	1.313	1.260	0.763	0.470	0.305	0.104	
2	4.300	3.740	3.639	2.851	2.337	2.387	1.551	1.479	0.841	0.535	0.359	0.125	
3	4.895	4.482	4.059	3.089	2.671	2.606	1.636	1.609	0.938	0.578	0.391	0.141	
4	5.402	4.720	4.269	3.357	2.737	2.634	1.754	1.625	0.962	0.645	0.394	0.151	
5	5.619	4.875	4.359	3.471	2.860	2.680	1.786	1.632	0.982	0.662	0.397	0.156	

Table 4.4: Asymptotic critical values for the $UDmax\, LR_T$ test, using $\epsilon = 0.05$. These are obtained from 10,000 Monte Carlo replications with $1,000$ increments.

Chapter 5

Distinguishing between Breaks in the Mean and Breaks in Persistence under Long Memory

Co-authored with Mwasi Mboya and Philipp Sibbertsen.

5.1 Introduction

Change point detection and estimation is a classical topic in econometrics and statistics (see the monograph of Csorgo and Horváth, 1997). Here we focus on changes in the mean and changes in the persistence in the long-memory context. Long memory as a feature of time series has been recognized in many fields of study. Examples of time series that exhibit long memory include volatilies (Y. K. Lu and Perron, 2010), inflation rates (Hsu, 2005) or trading volumes (Fleming and Kirby, 2011).

Changes in the mean under long-range dependence have been studied extensively. Iacone et al. (2014) use a sub-Wald type test, while Betken (2016) employs a Wilcoxon two-sample test statistic. However, many contributions focus on CUSUM-based test statistics (L. Wang, 2008; Shao, 2011; Wenger, Leschinski, and Sibbertsen, 2018a). Our paper builds on these findings but further expands on the idea that one may confuse breaks in the mean and breaks in persistence.

Leybourne et al. (2007) advise a CUSUM of squares test to detect a break in persistence of a time series in the $I(0)/I(1)$ framework. This test is extended by Sibbertsen and Kruse (2009) to fractional integration $I(d)$, i.e. they are able to detect if the order of integration breaks from stationarity ($0 \leq d < 1/2$) to non-stationarity ($1/2 < d < 3/2$) or vice versa. However, as pointed out by Sibbertsen and Willert (2012) this test is not robust in the presence of breaks in the mean. This finding motivates the present paper.

In this paper we introduce a novel procedure to discriminate between no break, a break in the mean or a break in persistence. The procedure consists of two steps: In a first step we determine whether or not a structural break affects the process at all by a statistical test. Afterwards in a second step we test if the break is a break in the mean or a break in persistence.

Our methodology follows Berkes et al. (2006) who study a procedure to distinguish a process with a break in the mean (but without long memory) under the null hypothesis from a process with true long memory. The idea is to split up the sample at the estimated break point under the null in two segments. The authors then investigate the behaviour of CUSUM-based test statistics formed from both segments. Aue et al. (2009) extend the procedure of Berkes et al. (2006) to distinguish breaks in the mean from random walks. Our contribution introduces long memory in this setting.

The paper is structured as follows: In Section 5.2 we introduce a test of the null of stationarity $I(d)$ $(0 \leq d < 1/2)$ against a break in the mean or a break in persistence. Afterwards in Section 5.3 we show how to distinguish between the two types of structural break. The finite sample properties are studied by a Monte Carlo simulation in Section 5.4. An application to inflation rates is given in Section 5.5. Section 5.6 concludes.

5.2 Detecting breaks in the mean or in the persistence

We want to study the process X_t given by

$$X_t \;=\; \mu_t + Y_t, \qquad t = 1, \dots, T,$$

where μ_t describes the mean of the observed time series and Y_t is the innovation process possibly having long memory. In this paper we only consider long memory generated according to the ARFIMA(p, d, q) model as proposed by Granger and Joyeux (1980):

$$\Phi(B)\,(1-B)^d\,Y_t \;=\; \Psi(B)\;\varepsilon_t, \qquad \text{as} \quad t = 1, \dots, T,$$

where ε_t is i.i.d. white noise with mean 0, variance σ_ε^2 and $E|\varepsilon_t|^{2+\delta} < \infty$ for some $\delta > 0$. The AR- and MA-polynomials $\Phi(B)$ and $\Psi(B)$ are assumed to have all roots outside the unit circle. For notational convenience we write $Y_t \sim$ ARFIMA(p, d, q). We consider different scenarios for both the mean process and the innovation process.

Remark 1. We note that if $Y_t \sim$ ARFIMA(p, d, q) the process satisfies the functional central limit theorem (FCLT), i.e. for $d \in [0, 1/2)$ and with some $\sigma > 0$ it holds

$$\frac{1}{T^{1/2+d}} \sum_{k=1}^{[Tt]} Y_k \overset{D}{\Rightarrow} \sigma W_d(t) \qquad (t \in [0,1], T \to \infty),$$

where $[\cdot]$ denotes the integer part and $W_d(t), t \in [0, 1]$ denotes standard fractional Brownian motion and $\overset{D}{\Rightarrow}$ stands for weak convergence in the Skorohod topology.

The first scenario I characterizes the situation where we observe no break in the mean and no break in persistence.

Scenario I. Under the null hypothesis of stability of the parameters the mean process μ_t is constant and the innovation process Y_t is stationary, i.e.

$$\begin{aligned} \mathcal{H}_0\colon\; \mu_t = \mu \qquad &\text{for all} \quad t = 1, \dots, T, \\ Y_t \sim \text{ARFIMA}&(p, d, q) \end{aligned}$$

for some $d \in [0, 1/2)$.

The alternative scenarios II and III have in common that they involve a change-point at an unknown point in time k^*. As is typical in the literature on change-point analysis we assume that, loosely speaking, the observations before and after the change point tend both to infinity with increasing sample size, i.e.

$$k^* = [\theta T] \text{ for some } \theta \in (0,1).$$

We now introduce the scenarios formally.

Scenario II. Under the first alternative we have a mean shift from μ to $\mu + \Delta$ at k^*, while the innovation process Y_t is unchanged, i.e.

$$\mathcal{H}_A^{(1)}: \ \mu_t = \begin{cases} \mu, & t = 1, \dots, k^* \\ \mu + \Delta, & t = k^* + 1, \dots, T, \end{cases}$$
$$Y_t \sim \text{ARFIMA}(p, d, q)$$

for some $d \in [0, 1/2)$ and $\Delta \neq 0$.

Scenario III. Under the second alternative we observe a constant mean μ_t while the long memory parameter d of Y_t changes from stationarity to non-stationarity ($\mathcal{H}_A^{(2.1)}$) or vice versa ($\mathcal{H}_A^{(2.2)}$), i.e. there exist $d_1 \in [0, 1/2)$ and $d_2 \in (1/2, 3/2)$ such that

$$\begin{aligned} \mathcal{H}_A^{(2.1)}: \ & \mu_t = \mu \qquad \text{for all} \quad t = 1, \dots, T, \\ & Y_t \sim \text{ARFIMA}(p, d_1, q), t = 1, \dots, k^*, \\ & Y_t \sim \text{ARFIMA}(p, d_2, q), t = k^* + 1, \dots, T, \\ \mathcal{H}_A^{(2.2)}: \ & \mu_t = \mu \qquad \text{for all} \quad t = 1, \dots, T, \\ & Y_t \sim \text{ARFIMA}(p, d_2, q), t = 1, \dots, k^* \\ & Y_t \sim \text{ARFIMA}(p, d_1, q), t = k^* + 1, \dots, T. \end{aligned}$$

Our test statistic is the CUSUM statistic given by

$$R_T \ = \ \frac{1}{T^{1/2+\hat{d}}\hat{\sigma}(\hat{d})} \max_{1 \le k \le T} \left| \sum_{i=1}^{k} (X_i - \bar{X}_T) \right| ,$$

where $\bar{X}_T$ denotes the mean of the whole series X_t and $\hat{\sigma}(\hat{d})$ is an estimate of the long-run variance of X_t as in remark 1. We will employ the local Whittle estimator of Robinson (1995a) that satisfies $\hat{d} - d_0 = o_p(m^{-1})$ where $m = T^b$ the bandwidth parameter with $b \in (0, 1)$. Furthermore for $\hat{\sigma}(\hat{d})$ we use the MAC estimator of Robinson (2005). First, we consider the asymptotics of our test statistic under the null hypothesis.

Theorem 5.1 (Theorem 2.1, L. Wang (2008))**.** *Under scenario I we have*

$$R_T \overset{D}{\Rightarrow} \sup_{0 \le \tau \le 1} |\tilde{W}_d(\tau)|, \qquad \textit{as} \quad T \to \infty,$$

where $\tilde{W}_d(\tau) = W_d(\tau) - \tau W_d(1)$ *is a fractional Brownian bridge.*

Now, we consider the asymptotics of our test statistic under the first alternative $\mathcal{H}_A^{(1)}$. The following is known

Theorem 5.2 (Theorem 2.2, L. Wang (2008))**.** *Under scenario II we have*

$$R_T \xrightarrow{p} \infty \qquad as \quad T \to \infty.$$

For the case of alternatives $\mathcal{H}_A^{(2.1)}$ resp. $\mathcal{H}_A^{(2.2)}$ we provide the following theorem about the test statistic

Theorem 5.3. *Under scenario III if* $2d_2 - (2+b)\hat{d} > 0$ *we have*

$$R_T \xrightarrow{p} \infty \qquad as \quad T \to \infty.$$

The proof of the theorem can be found in the appendix. The Theorems 5.1 to 5.3 suggest a testing procedure of the null hypothesis $\mathcal{H}_0$ against the alternative hypotheses $\mathcal{H}_A$.

5.3 Break in mean or break in persistence?

After having detected a break in mean or in persistence in the previous section we now want to distinguish these two forms of a break. Therefore we use an estimator for the break point k^* in case of scenario II. We use

$$\hat{k} = \arg\min_{1 \le k \le T} \sum_{i=1}^{k} (X_i - \bar{X}_{1,k})^2 + \sum_{i=k+1}^{T} (X_i - \bar{X}_{k+1,T})^2,$$

where $\bar{X}_{i,j}$ denotes the mean value of the observations $X_i, \dots, X_j$. By Theorem 3 of Lavielle and Moulines (2000) it holds in the case of scenario II that $\hat{k} \xrightarrow{p} k^*$ as $T \to \infty$. We therefore suggest to split the sample based on this estimator $\hat{k}$ into $X_1, \dots, X_{\hat{k}}$ and $X_{\hat{k}+1}, \dots, X_T$. Then we take the CUSUM statistics before and after the estimated break as our basic test statistics:

$$R_{T,1} = \frac{1}{\hat{k}^{1/2+\hat{d}_1} \hat{\sigma_1}(\hat{d}_1)} \max_{1 \le k \le \hat{k}} \left| \sum_{i=1}^{k} X_i - \frac{k}{\hat{k}} \sum_{i=1}^{\hat{k}} X_i \right|,$$

$$R_{T,2} = \frac{1}{(T-\hat{k})^{1/2+\hat{d}_2} \hat{\sigma_2}(\hat{d}_2)} \max_{\hat{k}+1 \le k \le T} \left| \sum_{i=\hat{k}+1}^{k} X_i - \frac{k-\hat{k}}{T-\hat{k}} \sum_{i=\hat{k}+1}^{T} X_i \right|,$$

where the long memory parameters estimates $\hat{d}_1, \hat{d}_2$ and the long-run variance estimates $\hat{\sigma_1}(\hat{d}_1), \hat{\sigma_2}(\hat{d}_2)$ are obtained on the respective subsamples $X_1, \dots, X_{\hat{k}}$ and $X_{\hat{k}+1}, \dots, X_T$. The new test statistic is obtained by determining the maximum of $R_{T,1}$ and $R_{T,2}$, i.e.

$$R^* = \max\{R_{T,1}, R_{T,2}\}.$$

Then we have the following

Theorem 5.4. *Suppose that under scenario II*

$$\hat{d}_1 \xrightarrow{p} d, \quad \hat{d}_2 \xrightarrow{p} d, \quad \hat{\sigma}_1(\hat{d}_1) \xrightarrow{p} \sigma, \quad \textit{and} \quad \hat{\sigma}_2(\hat{d}_2) \xrightarrow{p} \sigma. \tag{5.1}$$

Then the test statistic R^* has the following limit distribution

$$R^* \overset{D}{\Rightarrow} \max\{ \sup_{0\le t\le \theta} \frac{\sigma}{\theta^{1/2+d}} \left| W_d(t) - \frac{t}{\theta} W_d(\theta) \right|,$$
$$\sup_{\theta\le t\le 1} \frac{\sigma}{(1-\theta)^{1/2+d}} \left| W_d(t) - W_d(\theta) - \frac{t-\theta}{1-\theta}(W_d(1) - W_d(\theta)) \right| \},$$

as $T \to \infty$.

Proof. By the strong approximation principle (Corollary 1.1 of Q. Wang et al. (2003)) we know

$$\sup_{0\le t\le 1} \left| \sum_{i=1}^{T} Y_i - \sigma W_d(Tt) \right| = o_p(T^{1/2+d}), \qquad \text{as} \quad T \to \infty. \tag{5.2}$$

First, we want to prove that

$$\max_{1\le k\le T} \Bigg| \sum_{i=1}^{[Tt]} (X_i - \bar{X}_{\hat{k}}) 1\{[Tt] \le \hat{k}\} -$$
$$\sigma \left(W_d(Tt) - \frac{t}{\theta} W_d(T\theta) \right) 1\{t \le \theta\} \Bigg| = o_p(T^{1/2+d}). \tag{5.3}$$

Notice that by Theorem 3 of Lavielle and Moulines (2000) we have

$$\max_{1\le k\le \hat{k}} \left| \sum_{i=1}^{k} (X_i - \bar{X}_{\hat{k}}) - \sum_{i=1}^{k} (Y_i - \bar{Y}_{\hat{k}}) \right| \le 2|\hat{k} - k^*|\,|\Delta| = o_p(T^{1/2+d}),$$

as $T \to \infty$ where $\bar{Y}_{\hat{k}} = 1/\hat{k} \sum_{i=1}^{\hat{k}} Y_i$. Using (5.2) we further have as $T \to \infty$

$$\max_{1\le k\le \hat{k}} \left| \sum_{i=1}^{k} (Y_i - \bar{Y}_{\hat{k}}) - \sigma \left(W_d(k) - \frac{\hat{k}}{T} W_d(\hat{k}) \right) \right| = o_p(T^{1/2+d}).$$

By the almost sure continuity of fractional Brownian motion (Nualart, 2006) we deduce

$$\left| \frac{T}{\hat{k}} W_d(\hat{k}) - \frac{1}{\theta} W_d(\theta T) \right| = o_p(T^{1/2+d}), \qquad (T \to \infty).$$

Now, we write

$$\max_{1\le k\le T} \left| \left(W_d(k) - \frac{k}{\hat{k}} W_d(\hat{k}) \right) 1\{k \le \hat{k}\} - \left(W_d(k) - \frac{k}{T\theta} W_d(T\theta) \right) 1\{k \le T\theta\} \right|$$
$$\le \max_{1\le k\le T} \left| \left[\frac{k}{\hat{k}} - \frac{k}{T\theta} \right] W_d(\hat{k}) \right| + \max_{1\le k\le T} \left| \frac{k}{T\theta} (W_d(\hat{k}) - W_d(T\theta)) \right|$$
$$+ \max_{\min(\hat{k},T\theta)\le k\le \max(\hat{k},T\theta)} \left| W_d(k) - \frac{k}{T\theta} W_d(T\theta) \right|$$
$$=: A_1 + A_2 + A_3.$$

For the first term because $|W_d(\hat{k})| \leq |\sup_{0\leq t\leq 1}|W_d(t)| = O_p(T^{1/2+d})$ we have as $\hat{k}/T \xrightarrow{p} \theta$ that $A_1 = o_p(T^{1/2+d})$. For the second term we have by self-similarity

$$T^{-1/2-d} \sup_{|h|\leq\varepsilon}|W_d(T\theta) - W_d(T(\theta+h))| \overset{D}{=} \sup_{|h|\leq\varepsilon}|W_d(\theta) - W_d(\theta+h)| \to 0, \qquad \text{a.s.},$$

as $\varepsilon \to 0$. Thus $A_2 = o_p(T^{1/2+d})$. For the third term we use the rescaling property to arrive at

$$\begin{aligned} &T^{-1/2-d} \sup_{T(\theta-\varepsilon)\leq t\leq T(\theta+\varepsilon)} |W_d(t) - \frac{t}{T\theta}W_d(T\theta)| \\ \overset{D}{=}\ &\theta^{1/2+d} \sup_{1-\varepsilon/\theta\leq s\leq 1+\varepsilon/\theta} |W_d(s) - sW_d(1)| \to 0 \qquad \text{a.s.}, \end{aligned}$$

as $\varepsilon \to 0$ because of the almost sure continuity of W_d at 1. This gives $A_3 = o_p(T^{1/2+d})$. Hence we have proved Equation (5.3). Arguing similarly one can show that

$$\begin{aligned} \max_{1\leq k\leq T} \Bigg| &\sum_{i=1}^{[Tt]}(X_i - \bar{X}_{\hat{k}})1\{[Tt] > \hat{k}\} - \\ &\sigma\left(W_d(Tt) - W_d(T\theta) - \frac{t-\theta}{1-\theta}(W_d(T) - W_d(T\theta))\right)1\{t > \theta\}\Bigg| \\ &= o_p(T^{1/2+d}). \end{aligned} \tag{5.4}$$

Using again the self-similarity of fractional Brownian motion Equations (5.3) and (5.4) imply the joint weak convergence of the test statistics

$$\begin{aligned} (R_{T,1}, R_{T,2}) \overset{D}{\Rightarrow} \sigma\Bigg(&\sup_{0\leq t\leq\theta} \frac{\sigma}{\theta^{1/2+d}}\left|W_d(t) - \frac{t}{\theta}W_d(\theta)\right|, \\ &\sup_{\theta\leq t\leq 1} \frac{\sigma}{(1-\theta)^{1/2+d}}\left|W_d(t) - W_d(\theta) - \frac{t-\theta}{1-\theta}(W_d(1) - W_d(\theta))\right|\Bigg). \end{aligned}$$

By the continuous mapping theorem we finally prove the theorem. □

Remark 2. Establishing Equation (5.1) for the local Whittle and the MAC estimator is out of the scope for this paper as elaborate arguments concerning randomly stopped sums of random variables are needed. However, in Section 5.4 we find promising results using these estimators in a Monte Carlo experiment.

The test statistic is asymptotically consistent against the alternative of a break in persistence as the next result shows.

Theorem 5.5. *Under scenario III if* $2d_2 - (2+b)\max\{\hat{d}_1, \hat{d}_2\} > 0$ *the test statistic diverges:*

$$R^* \xrightarrow{p} \infty \qquad \text{as} \quad T \to \infty.$$

The proof of the theorem can be found in the appendix. We see that Theorems 5.4 and 5.5 suggest a test of a break in mean against a break in persistence.

5.4 A Simulation Study

We study the finite sample size and power properties of the test by a Monte Carlo simulation study. Therefore we simulate fractionally integrated standard normal random variables and choose $d \in \{0, 0.2, 0.4\}$ and sample size $T = 1000$. The nominal significance level is $\alpha = 5\,\%$ and we report rejection frequencies obtained from $10,000$ replications.

At first we analyze how well the test of Section 5.2 is able to detect if a break in mean or in persistence exists. In Table 5.1 we show the size and the power against scenario II of the test. We consider a break in the mean of one standard deviation in the middle of the sample. Similar results have been obtained for example by Wenger, Leschinski, and Sibbertsen (2018a), therefore we do not investigate other sample sizes.

	Size			Power (scenario II)		
b	0.6	0.7	0.8	0.6	0.7	0.8
$d = 0.0$	0.0151	0.0221	0.0277	0.9956	1	1
$d = 0.2$	0.0209	0.0248	0.0347	0.7913	0.9644	0.9938
$d = 0.4$	0.0105	0.0156	0.0270	0.6368	0.6176	0.6982

Table 5.1: Size and power results of the test of Section 5.2 where power results are obtained if scenario II is true. We set $T = 1,000$ and use bandwidth $m = [T^b]$.

In Table 5.2 we show the power of the test against scenario III. We consider a break in persistence from d_1 to d_2 in the middle of the sample. As the power is quite high we do not investigate other sample sizes. Power results for breaks in persistence from non-stationarity to stationarity are not shown here, but yield similar results.

b	0.6			0.7			0.8		
d_1/d_2	0.6	0.8	1	0.6	0.8	1	0.6	0.8	1
0	0.4473	0.6844	0.8549	0.4639	0.7522	0.9233	0.4503	0.7903	0.9549
0.2	0.394	0.6674	0.8484	0.4057	0.7338	0.9161	0.3965	0.7798	0.9578
0.4	0.3164	0.63	0.8451	0.3414	0.7213	0.9118	0.3457	0.7788	0.9545

Table 5.2: Power results of the test of Section 5.2 if the DGP is obtained under scenario III. We set $T = 1,000$ and use bandwidth $m = [T^b]$.

Subsequently, we analyze how the test of Section 5.3 performs to distinguish breaks in mean from breaks in persistence. As before we consider one break in the middle of the sample specified as above. It should be noted that the limiting distribution in Theorem 5.4 depends on two parameters the break point θ and the memory parameter d both which must be estimated in practice. For the break point θ we suggest to use the break point estimate $\hat{\theta} = \hat{k}/T$. For the memory parameter we simply take the mean of the two estimated parameters $\hat{d}_1, \hat{d}_2$, i.e. $\hat{d} = (\hat{d}_1 + \hat{d}_2)/2$ as this is a consistent estimate of d under the null hypothesis. In Table 5.3 we see that the test is conservative as it does not reach its nominal level.

b	0.6		0.7		0.8	
T	1,000	10,000	1,000	10,000	1,000	10,000
$d = 0.0$	0.0138	0.0197	0.0118	0.0288	0.0161	0.0370
$d = 0.2$	0.0332	0.0569	0.0361	0.0332	0.0316	0.0398
$d = 0.4$	0.1700	0.0607	0.1117	0.0321	0.0638	0.0294

Table 5.3: Size results of the test of Section 5.3. We set $T = 1,000$ and use bandwidth $m = [T^b]$.

Lastly, we report in Table 5.4 the power results of the test of Section 5.4. Clearly, the test only has reasonable power if the difference of d_1 to d_2 is large enough. Hence for a break in persistence from $d_1 = 0.4$ to $d_2 = 0.6$ for a smaller sample size $T = 1,000$ the test has almost no power.

	b	0.6			0.7			0.8		
T	d_1/d_2	0.6	0.8	1	0.6	0.8	1	0.6	0.8	1
$1,000$	0.0	0.1665	0.2873	0.5784	0.2234	0.3527	0.7108	0.3458	0.4245	0.7592
$10,000$		0.1788	0.6216	0.9203	0.2289	0.7285	0.9732	0.3941	0.7674	0.9794
$1,000$	0.2	0.1065	0.2503	0.5682	0.1114	0.316	0.7022	0.1686	0.358	0.768
$10,000$		0.1278	0.5945	0.908	0.1604	0.7115	0.9711	0.2387	0.7472	0.9824
$1,000$	0.4	0.0453	0.2194	0.563	0.0512	0.3245	0.7282	0.0539	0.4008	0.82
$10,000$		0.0878	0.587	0.922	0.1101	0.7228	0.9761	0.1253	0.8141	0.9919

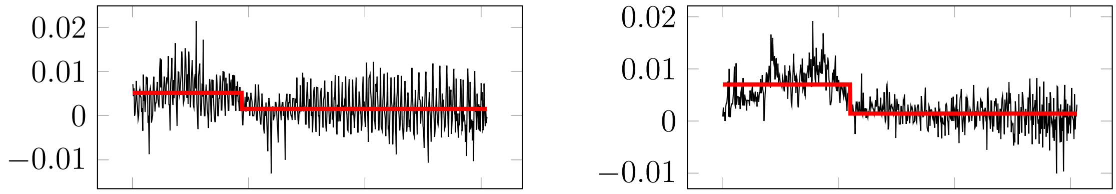
Table 5.4: Power results of the test of Section 5.3 if the DGP is obtained under scenario III. We use bandwidth $m = [T^b]$.

5.5 An Application to Inflation Rates

In this section we reconsider the case of inflation rates. This study is inspired by Kumar and Okimoto (2007) who investigated inflation rates in the US and found that they can be successfully explained by fractional integration. Moreover, they find decreasing persistence when estimating the long memory parameter in different time spans. In a similar vein, Caporin and Gupta (2017) use a model with a long memory coefficient that may vary for different time periods. Contrarily, Gadea et al. (2004) or Hsu (2005) argue that structural breaks should be taken into account when modelling inflation rates. We contribute to this strand of literature by analyzing the behavior of inflation rates for European countries.

We obtain our data from the OECD[1] and use the monthly CPI for 10 countries for a time span from 1967 to 2017. We deseasonalize the data first and transform the data to inflation rates π_t by taking differences of the log of the data, i.e. $\pi_t = \log(\mathrm{CPI}_t) - \log(\mathrm{CPI}_{t-1})$, which is common in the literature.

[1]Dataset from https://data.oecd.org/price/inflation-cpi.htm.



(a) Netherlands (b) France

Figure 5.1: Inflation Rates for Netherlands and France (not deseasonalized) where our procedure suggest a break in mean and a break in persistence respectively. In red we show the means of the subsamples $X_1, \ldots, X_{\hat{k}}$ and $X_{\hat{k}+1}, \ldots, X_T$.

Country	AUT	BEL	FIN	FRA	DEU	ESP	ITA	LUX	NLD	PRT
P-value I	0.0001	0.0121	0.0048	0.0022	0.0305	0	0.0255	0	0.0001	0.0001
P-value II	0.331	0.4074	0.0204	0	0.1431	0.0001	0.0092	0.4324	0.7443	0.0162

Table 5.5: P-values for the test statistic R_T (P-value I) and for the test statistic R^* (P-value II) for different countries. We use bandwidth $b = 0.7$, i.e. $m = [T^{0.7}]$.

In Table 5.5 we show the p-values for both steps of the procedure for the studied countries. All countries have either a break in mean or break in persistence if we use a significance level $\alpha = 5\%$. Therefore we calculate the p-values for the second step of the procedure. For the significance level $\alpha = 5\%$ there are 5 countries whose inflation rate cannot be described by a break in mean. In Figure 5.1 we try to make the results of our procedure plausible by showing examples, where we suggest a break in mean (5.1a) and a break in persistence (5.1b).

5.6 Conclusion

In this paper we introduce a new method to decide whether a time series is affected by a break in the mean or by a break in persistence. The procedure consists of two tests: We derive their theoretical limiting distributions under the null and show that they are consistent against the alternative. Furthermore we show in a simulation study how the test performs in small samples and apply the procedure to inflation rates.

5.7 Appendix

Proof of Theorem 5.3. We begin by proving the result under hypothesis $\mathcal{H}_A^{(2.1)}$ and write

$$\begin{aligned}
&\frac{1}{T^{1/2+\hat{d}}} \max_{1\leq k\leq T} \left| \sum_{i=1}^{k} X_i - \frac{k}{T}\sum_{i=1}^{T} X_i \right| \\
&\quad \geq \frac{1}{T^{1/2+\hat{d}}} \left| \sum_{i=1}^{k^*} X_i - \frac{k^*}{T}\sum_{i=1}^{T} X_i \right| \\
&\quad \geq \left| \left| \frac{(1-\theta)}{T^{1/2+\hat{d}}} \sum_{i=1}^{k^*} Y_i \right| - \left| \frac{\theta}{T^{1/2+\hat{d}}} \sum_{i=k^*+1}^{T} Y_i \right| \right|. \qquad (5.5)
\end{aligned}$$

The Y_i are stationary for $i = 1, \ldots, k^*$ under $\mathcal{H}_A^{(2.1)}$, hence the first term is $O_p(T^{d_1-\hat{d}})$. On the other hand the second term is of order $O_p(T^{d_2-\hat{d}})$. Furthermore we note that $\hat{d}$ is estimated under $\mathcal{H}_0$ so that $\hat{d} < 1/2$. So, we conclude that the second term in Equation (5.5) dominates the first term. The MAC estimator $\hat{\sigma}(\hat{d})$ is given by (cf. Abadir, Distaso, and Giraitis (2009))

$$\hat{\sigma}^2(\hat{d}) = \frac{p(d)}{m} \sum_{j=1}^{m} \lambda_j^{2d} I_T(\lambda_j), \qquad I_T(\lambda_j) = (2\pi T)^{-1} \left| \sum_{t=1}^{T} e^{it\lambda_j} X_t \right|^2,$$

where $\lambda_j = 2\pi j/T$ and $p(d)$ is uniformly bounded for $d \in [0, 1/2]$. Thus $\hat{\sigma}(\hat{d})$ is seen to be of order $O_p(m^{\hat{d}}\, T^{d_2-\hat{d}}) = O_p(T^{(1+b)\hat{d}-d_2})$ as $m = T^b$. Combining this with (5.5) we find

$$\frac{1}{T^{1/2+\hat{d}}\hat{\sigma}(\hat{d})} \max_{1\leq k\leq T} \left| \sum_{i=1}^{k} X_i - \frac{k}{T}\sum_{i=1}^{T} X_i \right| = O_p(T^{2d_2-(2+b)\hat{d}}),$$

which diverges under the premises of the Theorem. It is seen that under $\mathcal{H}_A^{(2.2)}$ the proof is similar. □

Proof of Theorem 5.5. The proof is similar to the proof of Theorem 5.3 as we just need to show that the maximum of the CUSUM statistics $R_{T,1}$ and $R_{T,2}$ diverges, which is shown by carefully considering all cases. □

Chapter 6

Testing for a break in mean in electricity load data

Co-authored with Theoplasti Kolaiti.

6.1 Introduction

Historically, electricity markets have developed from monopolistic providers to deregulated and competitive markets (cf. Weron (2007)). Therefore it is of great importance for system operators to schedule power generators in order to harmonize production and demand. Typically, scheduling algorithms use forecasts of electricity loads based on time series models.

Many papers study forecasting of short-term (Taylor (2003); Taylor et al. (2006)), mid-term (Mirasgedis et al. (2006)), and long-term (Hyndman and Fan (2009)) electricity demand. However, most of the papers on forecasting do not take into account the possibility of breaks in the mean in the time series data. The paper of Lacir J. Soares and Medeiros (2008) is a notable exception where the authors seem to identify the structural break by the “eyeball technique”. Another example is the paper of Leonardo R. Souza and Lacir J. Soares (2007) who identify a break in the Brazilian electricity demand after a nationwide rationing scheme was implemented.

Yet, for electricity prices structural breaks are well-documented for many countries: the EU (Kirat and Ahamada (2011)), Australia (Apergis and Lau (2015)), or China (Liu et al. (2013)). In this paper we argue that breaks in electricity demand may also be present. Detecting them by a statistical test is useful to avoid model misspecification and misleading forecasts. We introduce a novel hypothesis test to detect breaks in mean as well as an application to electricity load data.

Electricity loads display intriguing features that have been extensively researched in time series analysis so far. First of all, electricity loads exhibit strong seasonal variation. The climate conditions change throughout the year so that temperature or daylight hours cause varying demand by consumers. On the other hand, the supply side may shift following a yearly pattern, e.g. when considering hydro units or photovoltaic cells.

Moreover, electricity load is characterized by long memory. That is the autocorrelations of the process decay hyperbolically, instead of exponentially as in ARMA models. Lacir Jorge Soares and Leonardo Rocha Souza (2006) introduce a model that is based on generalized long memory processes using GARMA models. Similarly, Sadaei et al. (2017) employ a SARFIMA model to capture the long-range dependence of the process using fractional integration.

The preceding discussion motivates to include seasonality and long memory in our model. We assume that, electricity load data, follow a model of the form

$$Y_t = \psi_t + \theta_t + Z_t, \qquad t = 1, \ldots, n, \tag{6.1}$$

where ψ_t is the long-time trend, θ_t is the seasonal component and Z_k is a long memory error process to be specified below in greater detail. We also assume that the long-time trend is given by $\psi_t = f(t/n)$; for f being a piecewise continuous function with finitely many segments. Testing for a break in mean then corresponds to the hypotheses

$$\mathcal{H}_0\colon f = \text{constant} \qquad \text{against} \qquad \mathcal{H}_A\colon f \text{ is not constant.}$$

This article is motivated by the setup of Wu (2004) whose test is inspired by isotonic regression. The simulation study indicates that our test's perfomance is more powerful than the alternative ones for a break in mean and particularly when the break date is at the beginning or end of the sample.

The paper is structured as follows: In Section 6.2 we introduce the test for a break in the mean under long memory. In Section 6.3 we show that it is not affected by seasonality theoretically. The finite sample properties of the test are investigated in a Monte Carlo study in Section 6.4. Then in Section 6.5 we apply the test to electricity load data and discuss the results. Finally, in Section 6.6 we conclude.

6.2 Testing for a break in mean under long-range dependence

To motivate our procedure we first consider the special model $X_t = \psi_t + Z_t$, that has no seasonal component, θ_t. For the error process we assume that for some $\sigma > 0$ and $d \in [0, 1/2)$ a functional central limit theorem (FCLT) holds:

$$\frac{1}{\sigma n^{1/2+d}} \sum_{i=1}^{[nt]} Z_i \Rightarrow W_d(t), \qquad 0 \le t \le 1, \tag{6.2}$$

where $[\cdot]$ denotes the integer part and W_d is fractional Brownian motion. Note that this assumption is satisfied by the ARFIMA model if the innovation sequence of the ARFIMA has bounded moments of order 2 (Abadir, Distaso, Giraitis, and Koul, 2014). Following Wu (2004) we suggest testing for $\psi_t = f(k/n) =$ constant by considering

$$\begin{aligned} \Lambda_d(n, r) &= \frac{1}{n^{2d}\hat{\sigma}^2} \sum_{k=1}^{n} (\underline{\psi}_{k,r} - \bar{X}_n)^2 + \frac{1}{n^{2d}\hat{\sigma}^2} \sum_{k=1}^{n} (\overline{\psi}_{k,r} - \bar{X}_n)^2 \\ &=: \underline{\Lambda}_d(n, r) + \overline{\Lambda}_d(n, r), \end{aligned} \tag{6.3}$$

where $\bar{X}_n$ is the sample mean of X_t and $\hat{\sigma}$ is an estimate of the square root of the long-run variance σ of the process and

$$\underline{\psi}_{k,r} = \max_{i \le k} \min_{j \ge k} \frac{X_{i,r} + \cdots + X_{j,r}}{j - i + 1}, \quad \overline{\psi}_{k,r} = \min_{i \le k} \max_{j \ge k} \frac{X_{i,-r} + \cdots + X_{j,-r}}{j - i + 1},$$

where $X_{1,r} = X_1 + r\,n^{1/2+d}$, $X_{i,r} = X_i$ for $2 \le i \le n-1$ and $X_{n,r} = X_n - r\,n^{1/2+d}$. The quantities denoted as $\underline{\psi}_{k,r}$ and $\overline{\psi}_{k,r}$ are the isotonic resp. antitonic regression functions of the sample $X_{1,r}, \ldots, X_{n,r}$ resp. $X_{1,-r}, \ldots, X_{n,-r}$ (cf. (Robertson et al., 1988:p. 24)). In the following we demonstrate the reasoning for the isotonic regression function $\underline{\psi}_{k,r}$. However, a similar interpretation can be given for the antitonic regression function $\overline{\psi}_{k,r}$.

If the X_i were iid $N(0, \sigma^2)$ random variables, then the isotonic regression function $\underline{\psi}_{k,r}$ can be seen to be a maximum likelihood estimate. Consider the set of isotonic functions $\psi = (\psi_1, \ldots, \psi_n)$, i.e. $\psi_1 \le \psi_2 \le \cdots \le \psi_n$. Then the isotonic regression function is the value that maximizes the penalized log-likelihood function $\mathcal{L}(\psi) = -(2\sigma^2)^{-1}\sum_{i=1}^{n}(X_i - \psi_i)^2 - r\,n^{1/2+d}(\psi_n - \psi_1)/\sigma^2$ with respect to the restriction that ψ is isotonic, where C is a constant. The penalization by a value $r > 0$ is introduced in order to avoid estimating the first ψ_1 and ψ_n with bias, the so-called "spiking problem".

On a sidenote we remark that $\underline{\Lambda}_0(n, r)$ has been considered before by Wu et al. (2001) as a test statistic to detect a break in the mean. In their setup the null hypothesis "$\psi_t =$ constant" is contrasted with the specific alternative $\psi_t < \psi_{t+1}$ for some t. Their test is based on isotonic regression.

In preparation for our results below we denote for a function H on $[0, 1]$ by $\underline{H}$ its greatest convex minorant and by $\overline{H} = -\underline{H}$ its least concave majorant. Moreover we use $\underline{h}$ and $\overline{h}$ for the left-hand derivatives of $\underline{H}$ and $\overline{H}$ respectively.

The asymptotic distribution of the test statistic in equation (6.3) will be given below. To prepare this result we consider the local alternative

$$f(t) \;=\; \mu + \sigma\phi(t)/n^{1/2-d}, \tag{6.4}$$

where ϕ is a right-continuous function on $[0, 1]$ for which $\int_0^1 \phi(t)dt = 0$. Further we denote by W_d standard fractional Brownian motion with memory parameter $d \in [0, 1/2)$. For $c > 0$ let $B^{\phi}_{d,c}(t) = W_d(t) - tW_d(1) + \int_0^t \phi(s)ds + c\mathbb{1}_{(0,1)}(t)$, $\underline{b}^{\phi}_{d,c}(t)$ and $\overline{b}^{\phi}_{d,-c}(t)$ be the left-hand derivatives of $\underline{B}^{\phi}_{d,c}(t)$ and $\overline{B}^{\phi}_{d,-c}(t)$ respectively.

Theorem 6.1. *Assume that* (6.4) *holds and let* $\hat{\sigma}^2_n$ *be a consistent estimator* σ^2. *Then by the FCLT* (6.2) *for* $c = r/\sigma$

$$\Lambda_d(n, r) \;\Rightarrow\; \int_0^1 [\underline{b}^{\phi}_{d,c}(t)]^2 + [\overline{b}^{\phi}_{d,-c}(t)]^2 dt.$$

The proof of this and the other theoretical results are given in the appendix. Theorem 6.1 provides us with the asymptotic distribution under the null hypothesis setting $\phi \equiv 0$. Critical values $\lambda_{c,d}(\alpha)$ depend on the long memory parameter d which has to be estimated in practice and a user-chosen parameter c. We present critical values for $c = 0.15$ in Table 6.1. Note that in practice the value r must be set to $c\hat{\sigma}_n$.

d	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45
$\lambda_{c,d}(\alpha)$	9.7587	7.0551	5.0859	3.7797	2.7147	1.9368	1.2896	0.82	0.3902	0.0995

Table 6.1: Critical values $\lambda_{c,d}(\alpha)$ of the test statistic $\Lambda(n, r)$ for $\alpha = 0.95$ and $c = 0.15$.

To study the power of our test we consider the specific alternative for some $\delta_n > 0$

$$f(t) = \frac{\delta_n}{n^{1/2+d}}\phi\left(\frac{t}{n}\right), \qquad t = 1, \ldots, n, \tag{6.5}$$

where ϕ is a nonzero function, nondecreasing and with $\int_0^1 \phi(t)dt = 0$. Then by Theorem 6.1 the asymptotic power of our test based on test statistic (6.3) can be computed. Below we show that the power converges to 1 if $\delta_n \to \infty$.

Proposition 6.1. *Assume the setting of Theorem 6.1 with* (6.5) *instead of* (6.4). *Then*

$$P\left(\int_0^1 [\underline{b}^{\phi}_{d,c}(t)]^2 + [\overline{b}^{\phi}_{d,-c}(t)]^2 dt > \lambda_{c,d}(\alpha)\right) = 1, \qquad (\delta_n \to \infty).$$

6.3 The effect of seasonality on the test statistic

We now expand the time series model from the previous section by a seasonal component θ_t, i.e. $Y_t = \psi_t + \theta_t + Z_t$, and study the implications for the test statistic $\Lambda(n, r)$. For the seasonal component we assume that $\theta_t = \sum_{i=1}^{I} A_i \cos(k\omega_i + \alpha_i)$ where $0 < \omega_i < 2\pi$ are frequencies and $A_i > 0$ are amplitudes. In this section we establish, casually speaking, that on the one hand the long-term trend ψ_t has no effect on identifying the spectral frequencies ω_i of the seasonal component and on the other hand that the seasonal component θ_t does not influence the test statistic.

As before let $X_t = \psi_t + Z_t$ be the time series without seasonal component and $S_t = \theta_t + Z_t$ be the time series without long-time trend. Furthermore, let the discrete Fourier transform of X_t denoted by $w_{n,X}(\omega) = \sum_{k=1}^{n} X_k e^{ik\omega}$ and the periodogram $I_{n,X}(\omega) = |w_{n,X}(\omega)|^2$ for $\omega \in (0, 2\pi)$. Then the difference in periodogram ordinates of X_t and S_t is asymptotically negligible:

Proposition 6.2. *If f is of bounded variation then for $\omega \in (0, 2\pi)$ it holds that $I_{n,X} - I_{n,S} = o_p(1)$.*

The proposition allows us to identify the seasonal components frequencies ω_i by looking at the periodogram of X_t since the periodogram $I_{n,S}(\omega)$ of S_t has a magnitude of order n if ω is one of the seasonal frequencies ω_i. We demonstrate this in Section 6.5 when analyzing electricity load data.

Now let, similar to X_t, $Y_{1,r} = Y_1 + r\, n^{1/2+d}$, $Y_{i,r} = Y_i$ for $2 \le i \le n-1$ and $Y_{n,r} = Y_n - r\, n^{1/2+d}$ and define

$$\underline{\nu}_{k,r} = \max_{i \le k} \min_{j \ge k} \frac{Y_{i,r} + \cdots + Y_{j,r}}{j - i + 1}, \quad \overline{\nu}_{k,r} = \min_{i \le k} \max_{j \ge k} \frac{Y_{i,-r} + \cdots + Y_{j,-r}}{j - i + 1}.$$

Then the test statistic $\Lambda_d(n, r)$ based on X_t and Y_t are asymptotically equivalent.

Proposition 6.3. *It holds that*

$$\sum_{k=1}^{n} (\underline{\psi}_{k,r} - \underline{\nu}_{k,r})^2 + \sum_{k=1}^{n} (\overline{\psi}_{k,r} - \overline{\nu}_{k,r})^2 = o_p(1).$$

We note that the estimation of d and σ are not affected by the introduction of the seasonality θ_t asymptotically. Therefore Proposition 6.3 immediately implies that the results from the previous section hold under model (6.1) with a seasonal component θ_t. Thus the asymptotic distribution for our model is given by Theorem 6.1.

6.4 A Monte Carlo Study

In order to compare the small sample properties of the test we conduct a Monte Carlo experiment. We study the properties of our test by considering in accordance with Equation (6.1) the process

$$y_t - \psi_t - \theta_t = (1-L)^{-d}\,\varepsilon_t, \qquad t = 1, \ldots, T, \tag{6.6}$$

where ψ_t describes the mean, θ_t the seasonal component and ε_t is standard normal white noise. For the long memory parameter d we choose $d \in \{0;\, 0.2;\, 0.4\}$ and for the sample size $T \in \{500;\, 1,000;\, 2,000\}$. Below we use a nominal significance level $\alpha = 5\%$ and we report the rejection frequencies for $M = 10,000$ replications.

Moreover, for estimating the long-memory parameter we use the local Whittle estimator of Robinson (1995a) and we use his MAC estimator (Robinson (2005)) for the long-run variance σ^2 of the process y_t. This estimator depends on the choice of a bandwidth parameter $b \in (0,1)$. Both estimators use the first $m = [n^b]$ periodogram ordinates.

In Table 6.2 we study the size of the test. First we set $\theta_t \equiv 0$ in the left half of the table. This results in the situation of Section 6.2. The test seems to hold its size better for larger samples sizes. Overall the test seems conservative as its rejection frequency is below its nominal significance level $\alpha = 5\%$ in most cases. Furthermore the rejection frequencies seem to be closer to 5% for the larger bandwidth $b = 0.8$.

		$\theta_t \equiv 0$			$\theta_t = \cos(2\pi 100 t)$		
d	n/b	0.7	0.75	0.8	0.7	0.75	0.8
0	500	0.0542	0.0577	0.0425	0.0501	0.0474	0.0452
	1000	0.0552	0.056	0.044	0.0484	0.046	0.049
	2000	0.0549	0.0537	0.0501	0.0486	0.0509	0.0494
0.2	500	0.0175	0.0161	0.0179	0.0192	0.0233	0.0224
	1000	0.0212	0.0182	0.0243	0.0266	0.0282	0.0298
	2000	0.0246	0.0262	0.0305	0.0274	0.0331	0.0332
0.4	500	0.0397	0.0384	0.033	0.0333	0.0322	0.0298
	1000	0.0483	0.0499	0.0392	0.0378	0.0343	0.0347
	2000	0.062	0.0532	0.0496	0.0453	0.0391	0.0393

Table 6.2: Size results of the change-point test suggested in Section 6.2 using the DGP (6.6). On the left we use no seasonal component, on the right we specify a low-frequency seasonality.

In the right half of table 6.2 we use a low-frequency seasonality in Equation (6.6). We set $\theta_t = \cos(2\pi\, n/100\; t/n) = \cos(2\pi 100\, t)$. Note that the spectral density of θ_t has a peak at frequency $2\pi\, n/100$. Therefore θ_t has an effect on the periodogram of X_t at a frequency that grows with rate n. However, as the estimators for d and σ use only the first $m = [n^b]$ periodogram ordinates the seasonality θ_t has no influence asymptotically on the estimation. By Proposition 6.3 the test should not be influenced by the seasonality. In the table we see that this is indeed the case. The rejection frequency is reduced in most parameter configurations.

	b	0.7			0.75			0.8		
d	n/τ	0.3	0.5	0.7	0.3	0.5	0.7	0.3	0.5	0.7
0	500	1	1	1	1	1	1	1	1	1
	1000	1	1	1	1	1	1	1	1	1
	2000	1	1	1	1	1	1	1	1	1
0.2	500	0.8291	0.9084	0.8197	0.848	0.9141	0.8524	0.8845	0.9376	0.8753
	1000	0.9718	0.9905	0.9718	0.982	0.9924	0.9795	0.9859	0.9952	0.9836
	2000	0.9994	1	0.9993	0.9996	0.9999	0.9992	0.9999	1	0.9997
0.4	500	0.3051	0.3062	0.2986	0.4384	0.4539	0.4285	0.5866	0.6197	0.5719
	1000	0.4729	0.4735	0.4563	0.6231	0.6544	0.6201	0.755	0.8098	0.7435
	2000	0.6643	0.6782	0.6495	0.805	0.8479	0.7975	0.8505	0.9097	0.8577
W: 0.4	2000	0.4385	0.7045	0.4378	0.5122	0.7637	0.5109	0.6403	0.84	0.6262
B: 0.4	2000	0.2989	0.627	0.3045	0.3114	0.6392	0.3062	0.3106	0.6443	0.3153

Table 6.3: Power results of the test suggested in Section 6.2 using the DGP (6.6). In the last two rows we show the power of the test of L. Wang (2008) (W) and Betken (2016) (B). Throughout we use $\theta_t \equiv 0$.

In Table 6.3 we study the power of our test under the alternative

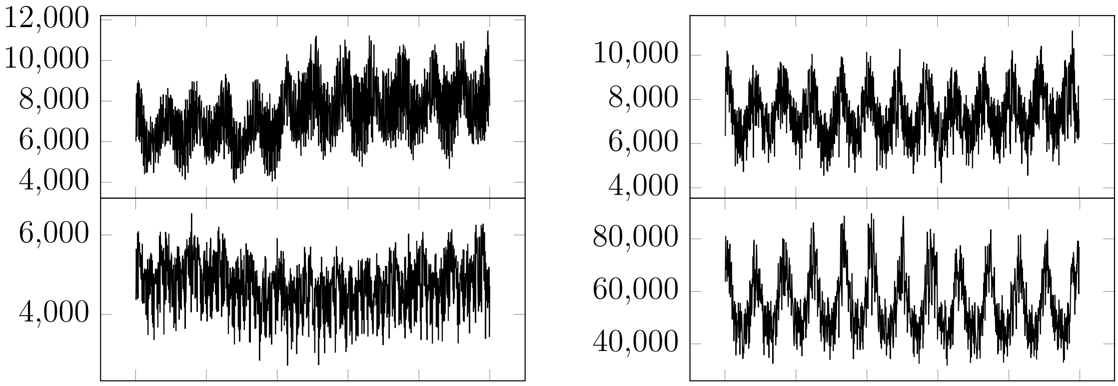
$$\psi_t = \mathbb{1}_{[\tau,1]}(t/n)\,\sigma$$

for $\tau \in \{0.3, 0.5, 0.7\}$ the break point. We find that the test has reasonable power even for large d although the power is diminishing with increasing d. Moreover we compare our test with the power of other testing procedures that have been suggested in the literature before: a CUSUM-type test of L. Wang (2008) and a Wilcoxon-type test of Betken (2016). It is interesting to see that our test performs similar if the break point is in the middle of the sample, but better if the break is at the beginning or end of the sample.

6.5 Breaks in mean in electricity load data

Our dataset of electricity load data[1] is publicly available and contains hourly electricity load data from 24 European countries. The observation length ranges from 9 to 12 years and spans the years 2006 to 2017. Looking at the periodogram of the series we can clearly identify a seasonality at the yearly, weekly and daily frequencies and their harmonics. The identification is justified by Proposition 6.2. This again confirms our model for electricity load given in Equation (6.1).

[1]Dataset from https://data.open-power-system-data.org/time_series/.



(a) Austria (top), Hungary (bottom)

(b) Czech Republic (top), France (bottom)

Figure 6.1: Observations of electricity demand of exemplary countries over the last 12 years.

Country	p-value	Country	p-value	Country	p-value
Austria	0	Bosnia	0	Belgium	0.1371
Bulgaria	0.0185	Czech Rep.	0.8089	Switzerland	0.0185
Germany	0.5949	Estonia	0.6443	Spain	0.0709
France	0.7234	Greece	0	Croatia	0
Hungary	0	Italy	0.354	Luxembourg	0.2724
Montenegro	0	Macedonia	0	Netherlands	0.056
Poland	0.0122	Portugal	0.2602	Romania	0.1485
Serbia	0.1068	Slovenia	0	Slovakia	0.0409

Table 6.4: P-values for the test statistic $\Lambda(n,r)$ for different countries when the test is applied to electricity demand. We use $c = 0.15$ and bandwidth $b = 0.75$, i.e. $m = [n^{0.75}]$.

We apply the test from section 6.2 using $c = 0.15$ and the estimators for d and σ as in the previous section. In Table 6.4 we show the p-values for the hourly electricity demand series for the 24 European countries. For 8 countries we can reject the null hypothesis of no break in mean for a significance level of $\alpha = 0.01$.

In order to make these results plausible we plot observations of hourly electricy demand in Figure 6.1. In the subplot 6.1a we have plots of Austria and Hungary for which the null hypothesis of no break is rejected by our test. On the other hand in the subplot 6.1b we have plots of Czech Republic and France which have p-values of 0.8 and 0.72 respectively. From visual inspection these examples show that the results seem reasonable.

6.6 Conclusion

In this paper we have introduced a change in mean test under long-range dependent errors and seasonality. The test is inspired by isotonic regression. We derive its asymptotic distribution and show that the test is consistent against a local alternative. Finite sample properties of the test are studied in a Monte Carlo study. Our test is applied to electricity load data from European countries. We can reject the null hypothesis of no break in the mean for the electricity load data in a various countries.

6.7 Appendix

6.7.1 Proof of Theorems

Proof of Theorem 6.1. Following Wu et al. (2001) lines, we prove the convergences of $\underline{\Lambda}_d(n,r)$ and $\overline{\Lambda}_d(n,r)$ separately: First

$$\underline{\Lambda}_d(n,r) \Rightarrow \int_0^1 [\underline{b}^{\phi}_{d,c}(t)]^2 dt. \tag{6.7}$$

Let $S_{k,r} = \sum_{i=1}^{k} X_{i,r}$ and $G_{n,r}(t)$ for $t \in [0,1]$ be a continuous, piecewise linear function, where $G_{n,r}(k/n) = S_{k,r}/n$ for $k = 1,\dots,n$. For $r = 0$ we denote S_k and G_n for simplicity. Analogously, let T_n be a continuous, piecewise linear function for which $T_n(t) = \sum_{i=1}^{k} Z_i$ for $t = k/n$, $k = 1,\dots,n$. Then we define

$$H_{n,r}(t) = \frac{n^{1/2-d}}{\sigma}(G_{n,r}(t) - \bar{X}_n t).$$

Furthermore, let $L_{n,r}$ and Φ_n be continuous, piecewise linear functions for which

$$\begin{aligned} L_{n,r}(0) =& L_{n,r}(1) = 0, \quad L_{n,r}(\frac{1}{n}) = L_{n,r}(1-\frac{1}{n}) = \frac{r}{n^{1/2}}, \\ \Phi_n(\frac{k}{n}) =& n^{-1}\sum_{p=1}^{k}\phi(\frac{p}{n}). \end{aligned}$$

So we may express $H_{n,r}$ in terms of these functions

$$H_{n,r}(t) = \frac{T_n(t) - tT_n(1)}{\sigma n^{1/2+d}} + \Phi_n(t) - t\Phi_n(1) + \frac{n^{1/2}}{\sigma}L_{n,r}(t).$$

Finally, for a bounded function H on $[0,1]$ we denote by $\underline{H}$ the greatest convex minorant of H and by $\underline{H}$ the left-hand derivative of $\underline{H}$. This justifies our choice of $\underline{b}^{\phi}_{d,c}(t)$ for the left-hand derivative of the GCM of $B^{\phi}_{d,c}$.

From Robertson et al. (1988), p. 7, it is known that $\underline{\psi}_{k,r} = \underline{g}_{n,r}(k/n)$ for $k = 1,\dots,n$. Hence $\underline{\psi}_{k,r} - \bar{X}_n = \sigma \underline{h}_{n,r}(k/n)/n^{1/2-d}$ and thus

$$\frac{1}{\sigma^2 n^{2d}}\sum_{k=1}^{n}(\underline{\psi}_{n,r} - \bar{X}_n)^2 = \int_0^1 \left(\underline{h}_{n,r}(t)\right)^2 dt.$$

So it is sufficient to show that

$$\int_0^1 \left(\underline{h}_{n,r}(t) - \underline{b}^{\phi}_{d,c}(t)\right)^2 dt \xrightarrow{p} 0.$$

This follows along the lines of the proof of Theorem 1 of Wu et al. (2001), p. 802, where we use Lemmas 6.1-6.4. The Lemmas can be found below. After having proved (6.7) the proof of the convergence of $\overline{\Lambda}_d(n,r)$ can be shown similarly. □

Proof of Proposition 6.1. Consider the function $J_n(t) = \Phi(t) + \delta_n^{-1} B_{d,c}^0(t)$ for $t \in [0,1]$. By Marshal's Lemma we have $\|\underline{J}_n - \Phi\| = O_p(\delta_n^{-1})$. Furthermore we see that

$$\underline{j}_n(0+) = \inf_{0<t<1} \frac{J_n(t)}{t} = O_p(1)$$

and similarly that $\underline{j}_n(1-) = O_p(1)$. Hence we may apply Lemma 6.3, which gives us

$$\|\underline{j}_n\|_2^2 \; = \; \|\phi\|_2^2 + O_p(\delta_n^{-1/2})$$

by the Cauchy-Schwarz inequality. Similarly we may deduce that $\|\overline{j}_n\|_2^2 = \|\phi\|_2^2 + O_p(\delta_n^{-1/2})$. Therefore the power of the test converges to 1 if $\delta_n \to \infty$. □

Proof of Proposition 6.2. We write $\kappa_n(\omega) = \sum_{k=1}^n \exp(\omega k i)$. For a fixed $\omega \in (0, 2\pi)$ it holds that $\sup_{n\geq 0}|\kappa_n(\omega)| \leq 2/|1 - \exp(\omega i)| = O(1)$. Then we may write using summation by parts

$$\begin{aligned}
\frac{|w_{n,X}(\omega) - w_{n,S}(\omega)|}{n^{1/2}} =& \frac{1}{n^{1/2}} \left| \sum_{k=1}^n f(k/n) \exp(\omega k i) \right| \\
=& \frac{1}{n^{1/2}} \left| \sum_{k=2}^n [f(k/n) - f((k-1)/n)] \kappa_k(\omega) \right| + O\left(\frac{1}{n^{1/2}}\right) \\
=& O\left(\frac{1}{n^{1/2}}\right),
\end{aligned}$$

where we use the bounded variation of f. Furthermore we have

$$\frac{|w_{n,X}(\omega) + w_{n,S}(\omega)|}{n^{1/2}} = O_p(n^d).$$

Combining both estimates yields

$$|I_{n,X} - I_{n,S}| = |w_{n,X} - w_{n,S}|\,|w_{n,X} + w_{n,S}| = O_p(n^{d-1/2}) = o_p(1).$$

□

Proof of Proposition 6.3. The proof essentially follows from the proof of Theorem 3 of Wu (2004). From the proof of Theorem 6.1 we recall that $G_{n,r}(k/n) = \sum_{i=1}^n X_{i,r}/n$ and $H_{n,r}(t) = n^{1/2-d}[G_{n,r}(t) - \bar{X}_n t]/\sigma$. Similarly, we define $P_{n,r}(t) = n^{1/2-d}[R_{n,r}(t) - \bar{Y}_n t]/\sigma$, where $R_{n,r}(k/n) = \sum_{i=1}^n Y_{i,r}/n$. By Lemma 6.3 we have

$$\begin{aligned}
&\int_0^1 [\underline{\psi}_{k,r}(t) - \underline{\nu}_{k,r}(t)]^2 dt \\
&\qquad \leq \|\underline{P}_{n,r} - \underline{H}_{n,r}\| \; [\underline{\psi}_{k,r}(1-) - \underline{\psi}_{k,r}(0+) + \underline{\nu}_{k,r}(1-) - \underline{\nu}_{k,r}(0+)].
\end{aligned}$$

The right term of the right-hand side stochastically bounded (see the proof of Lemma 6.4). However, for the left term we have by Marshall's lemma,

$$\|\underline{P}_{n,r} - \underline{H}_{n,r}\| \; \leq \; \|P_{n,r} - H_{n,r}\| = O\left(\sup_{1\leq k \leq n} \left| \sum_{i=1}^k \theta_i \right| / n^{1/2+d} \right) = o(1).$$

Therefore $\|\underline{\psi}_{k,r} - \underline{\nu}_{k,r}\|_2^2 = o_p(1)$ and we can show the same for $\|\overline{\psi}_{k,r} - \overline{\nu}_{k,r}\|_2^2$ using similar techniques. □

6.7.2 Lemmas

The following lemmas 6.1-6.3 are from Wu et al. (2001). Here, G and H are bounded functions on $[0,1]$ and $\|G\| = \sup_{0\le t\le 1}|G(t)|$.

Lemma 6.1. *If G is lower semi-continuous at 0 (resp. 1), then $\underline{G}(0) = G(0)$ (resp. $\underline{G}(1) = G(1)$).*

Lemma 6.2. *If $B \subset [0,1]$, $|G(t) - H(t)| \le \varepsilon$ for all $t \in B$ and $|\underline{G}(t) - \underline{H}(t)| \le \varepsilon$ for all $t \in [0,1] \setminus B$, then $\|\underline{G} - \underline{H}\| \le \varepsilon$.*

Lemma 6.3. *If*

$$\underline{G}(0) = \underline{H}(0), \underline{G}(1) = \underline{H}(1)$$
$$(-\infty < \underline{g}(0+) \le \underline{g}(1-) < \infty, -\infty < \underline{h}(0+) \le \underline{h}(1-) < \infty),$$

then

$$\int_0^1 \left(\underline{h} - \underline{g}\right)^2 dt \le \|\underline{H} - \underline{G}\| \left(\underline{g}(1-) - \underline{g}(0+) + \underline{h}(1-) - \underline{h}(0+)\right).$$

Lemma 6.4. *For any $c > 0$, $-\infty < \underline{b}^{\phi}_{d,c}(0+) \le \underline{b}^{\phi}_{d,c}(1-) < \infty$ with probability 1. Furthermore, $\underline{h}_{n,r}(0+)$ and $\underline{h}_{n,r}(1-)$ are stochastically bounded.*

Proof. First, it is clear that for a fractional Brownian bridge $\mathbb{B}_d$

$$0 \ge \underline{b}_{d,r}(0+) = \inf_{0<t<1} \frac{\mathbb{B}_d(t) + \Phi(t) + c}{t} \wedge 0 > -\infty.$$

Analogously, one shows the result for $\underline{b}_{d,r}(1-) < \infty$. Next, for the stochastic boundedness we write

$$\begin{aligned}\Pr(\underline{h}_{n,r}(0+) < -\lambda) \le& \Pr\left(n^{1/2-d} \max_{k\le\delta n} \frac{S_{n,r}}{\sigma k} > \lambda\right)\\ \le& \Pr\left(\max_{k\le\delta n}\left|\frac{S_k}{\sigma(\delta n)^{1/2+d}}\right| > \frac{c}{\delta^{1/2+d}}\right)\\ &+ \Pr\left(\max_{\delta n\le k\le n}\left|\frac{S_k}{\sigma n^{1/2+d}}\right| > \lambda\delta^{1/2+d}\right),\end{aligned}$$

where $\delta = 1/\sqrt{\lambda}$. But $\max_{k\le n} S_k/n^{1/2+d}$ is stochastically bounded, so $\Pr(\underline{h}_{n,r}(0+) < -\lambda) \to 0$ as first $n \to \infty$ and then $\lambda \to \infty$. The same result holds for the right endpoint $\underline{h}_{n,r}(1-)$ by a similar calculation. □

Bibliography

Abadir, Karim M., Walter Distaso, and Liudas Giraitis (2009). "Two estimators of the long-run variance: beyond short memory". In: *Journal of Econometrics* 150(1), pp. 56–70.

Abadir, Karim M., Walter Distaso, Liudas Giraitis, and Hira L. Koul (2014). "Asymptotic normality for weighted sums of linear processes". In: *Econometric Theory* 30(1), pp. 252–284.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Heiko Ebens (2001). "The distribution of realized stock return volatility". In: *Journal of Financial Economics* 61(1), pp. 43–76.

Apergis, Nicholas and Marco Chi Keung Lau (2015). "Structural breaks and electricity prices: Further evidence on the role of climate policy uncertainties in the Australian electricity market". In: *Energy Economics* 52, pp. 176–182.

Aue, Alexander, Lajos Horváth, Marie Hušková, and Shiqing Ling (2009). "On distinguishing between random walk and change in the mean alternatives". In: *Econometric Theory* 25(2), pp. 411–441.

Bai, Jushan (1997a). "Estimating multiple breaks one at a time". In: *Econometric Theory* 13(3), pp. 315–352.

Bai, Jushan (1997b). "Estimation of a change point in multiple regression models". In: *Review of Economics and Statistics* 79(4), pp. 551–563.

Bai, Jushan (2000). "Vector autoregressive models with structural changes in regression coefficients and in variance-covariance matrices". In: *Annals of Economics and Finance* 1(2), pp. 303–339.

Bai, Jushan, Robin L Lumsdaine, and James H Stock (1998). "Testing for and dating common breaks in multivariate time series". In: *The Review of Economic Studies* 65(3), pp. 395–432.

Bai, Jushan and Pierre Perron (1998). "Estimating and testing linear models with multiple structural changes". In: *Econometrica*, pp. 47–78.

Baillie, Richard T (1996). "Long memory processes and fractional integration in econometrics". In: *Journal of Econometrics* 73(1), pp. 5–59.

Banerjee, Anindya and Giovanni Urga (2005). "Modelling structural breaks, long memory and stock market volatility: an overview". In: *Journal of Econometrics* 129(1), pp. 1–34.

Baum, Christopher F, John T Barkoulas, and Mustafa Caglayan (1999). "Persistence in international inflation rates". In: *Southern Economic Journal*, pp. 900–913.

Beran, Jan (2009). "On parameter estimation for locally stationary long-memory processes". In: *Journal of Statistical Planning and Inference* 139(3), pp. 900–915.

Beran, Jan, R. Sherman, Murad S. Taqqu, and Walter Willinger (1995). "Long-range dependence in variable-bit-rate video traffic". In: *IEEE Transactions on Communications* 43(2/3/4), pp. 1566–1579.

Berkes, István, Lajos Horváth, Piotr Kokoszka, and Qi-Man Shao (2006). "On discriminating between long-range dependence and changes in mean". In: *The Annals of Statistics* 34(3), pp. 1140–1165.

Betken, Annika (2016). "Testing for change-points in long-range dependent time series by means of a self-normalized wilcoxon test". In: *Journal of Time Series Analysis* 37(6), pp. 785–809.

Bollerslev, Tim and Hans Ole Mikkelsen (1996). "Modeling and pricing long memory in stock market volatility". In: *Journal of Econometrics* 73(1), pp. 151–184.

Bos, Charles S, Philip Hans Franses, and Marius Ooms (1999). "Long memory and level shifts: Re-analyzing inflation rates". In: *Empirical Economics* 24(3), pp. 427–449.

Caporin, Massimiliano and Rangan Gupta (2017). "Time-varying persistence in US inflation". In: *Empirical Economics* 53(2), pp. 423–439.

Chung, Ching-Fan (2002). "Sample means, sample autocovariances, and linear regression of stationary multivariate long memory processes". In: *Econometric Theory* 18(1), pp. 51–78.

Csorgo, Miklos and Lajos Horváth (1997). *Limit theorems in change-point analysis.* John Wiley & Sons.

Dahlhaus, Rainer (1997). "Fitting time series models to nonstationary processes". In: *The Annals of Statistics* 25(1), pp. 1–37.

Dahlhaus, Rainer (2000). "A likelihood approximation for locally stationary processes". In: *The Annals of Statistics* 28(6), pp. 1762–1794.

Dahlhaus, Rainer (2012). "Locally stationary processes". In: *Handbook of statistics.* Vol. 30. Elsevier, pp. 351–413.

Davidson, James and Robert M. de Jong (2000). "The functional central limit theorem and weak convergence to stochastic integrals II: Fractionally Integrated Processes". In: *Econometric Theory* 16(5), pp. 643–666.

Dehling, Herold, Aeneas Rooch, and Murad S. Taqqu (2013). "Non-parametric change-point tests for long-range dependent data". In: *Scandinavian Journal of Statistics* 40(1), pp. 153–173.

Deo, Rohit, Clifford Hurvich, and Yi Lu (2006). "Forecasting realized volatility using a long-memory stochastic volatility model: estimation, prediction and seasonal adjustment". In: *Journal of Econometrics* 131(1-2), pp. 29–58.

Diebold, Francis X and Atsushi Inoue (2001). "Long memory and regime switching". In: *Journal of Econometrics* 105(1), pp. 131–159.

Ding, Zhuanxin, Clive WJ Granger, and Robert F Engle (1993). "A long memory property of stock market returns and a new model". In: *Journal of Empirical Finance* 1(1), pp. 83–106.

Faÿ, Gilles, Eric Moulines, François Roueff, and Murad S. Taqqu (2009). "Estimators of long-memory: Fourier versus wavelets". In: *Journal of Econometrics* 151(2), pp. 159–177.

Fleming, Jeff and Chris Kirby (2011). "Long memory in volatility and trading volume". In: *Journal of Banking & Finance* 35(7), pp. 1714–1726.

Gadea, Mariá Dolores, Marcela Sabaté, and José Marıa Serrano (2004). "Structural breaks and their trace in the memory: Inflation rate series in the long-run".

In: *Journal of International Financial Markets, Institutions and Money* 14(2), pp. 117–134.

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

Giraitis, Liudas, Hira L Koul, and Donatas Surgailis (2012). *Large sample inference for long memory processes*. World Scientific Publishing Company.

Granger, Clive WJ and Namwon Hyung (2004). "Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns". In: *Journal of Empirical Finance* 11(3), pp. 399–421.

Granger, Clive WJ and Roselyne Joyeux (1980). "An introduction to long-memory time series models and fractional differencing". In: *Journal of Time Series Analysis* 1(1), pp. 15–29.

Gubner, J. A. (2005). "Theorems and fallacies in the theory of long-range-dependent processes". In: *IEEE Transactions on Information Theory* 51(3), pp. 1234–1239.

Haldrup, Niels and Morten Ørregaard Nielsen (2006). "A regime switching long memory model for electricity prices". In: *Journal of Econometrics* 135(1-2), pp. 349–376.

Hassler, Uwe and Jürgen Wolters (1995). "Long memory in inflation rates: International evidence". In: *Journal of Business & Economic Statistics* 13(1), pp. 37–45.

Hsu, Chih-Chiang (2005). "Long memory or structural changes: An empirical examination on inflation rates". In: *Economics Letters* 88(2), pp. 289–294.

Hyndman, Rob J and Shu Fan (2009). "Density forecasting for long-term peak electricity demand". In: *IEEE Transactions on Power Systems* 25(2), pp. 1142–1153.

Iacone, Fabrizio, Stephen J Leybourne, and AM Robert Taylor (2014). "A fixed-b test for a break in level at an unknown time under fractional integration". In: *Journal of Time Series Analysis* 35(1), pp. 40–54.

Jensen, Mark J and Brandon Whitcher (2014). "Measuring the impact intradaily events have on the persistent nature of volatility". In: *Wavelet Applications in Economics and Finance*. Ed. by Marco Gallegati and Willi Semmler. Springer International Publishing. Chap. 5, pp. 103–129.

Kechagias, Stefanos and Vladas Pipiras (2015). "Definitions and representations of multivariate long-range dependent time series". In: *Journal of Time Series Analysis* 36(1), pp. 1–25.

Kirat, Djamel and Ibrahim Ahamada (2011). "The impact of the European Union emission trading scheme on the electricity-generation sector". In: *Energy Economics* 33(5), pp. 995–1003.

Kumar, Manmohan S and Tatsuyoshi Okimoto (2007). "Dynamics of persistence in international inflation rates". In: *Journal of Money, Credit and Banking* 39(6), pp. 1457–1479.

Künsch, Hans Rudolph (1987). "Statistical aspects of self-similar processes". In: *Proceedings of the First World Congress of the Bernoulli Society, 1987*. Vol. 1. VNU Science Press, pp. 67–74.

Lavielle, Marc and Eric Moulines (2000). "Least-squares estimation of an unknown number of shifts in a time series". In: *Journal of time series analysis* 21(1), pp. 33–59.

Leybourne, Stephen, Robert Taylor, and Tae-Hwan Kim (2007). "CUSUM of squares-based tests for a change in persistence". In: *Journal of Time Series Analysis* 28(3), pp. 408–433.

Liu, Ming-Hua, Dimitris Margaritis, and Yang Zhang (2013). "Market-driven coal prices and state-administered electricity prices in China". In: *Energy Economics* 40, pp. 167–175.

Lu, Yang K and Pierre Perron (2010). "Modeling and forecasting stock return volatility using a random level shift model". In: *Journal of Empirical Finance* 17(1), pp. 138–156.

Lu, Zhiping and Dominique Guegan (2011). "Estimation of time-varying long memory parameter using wavelet method". In: *Communications in Statistics—Simulation and Computation* 40(4), pp. 596–613.

Marinucci, Domenico and Peter Robinson (1999). "Alternative forms of fractional Brownian motion". In: *Journal of statistical planning and inference* 80(1-2), pp. 111–122.

Marinucci, Domenico and Peter Robinson (2000). "Weak convergence of multivariate fractional processes". In: *Stochastic Processes and their Applications* 86(1), pp. 103–120.

Martens, Martin, Dick van Dijk, and Michiel de Pooter (2009). "Forecasting S&P 500 volatility: Long memory, level shifts, leverage effects, day-of-the-week seasonality, and macroeconomic announcements". In: *International Journal of Forecasting* 25(2), pp. 282–303.

McCloskey, Adam and Pierre Perron (2013). "Memory parameter estimation in the presence of level shifts and deterministic trends". In: *Econometric Theory* 29(6), pp. 1196–1237.

Mikosch, Thomas and Cătălin Stărică (2004). "Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects". In: *Review of Economics and Statistics* 86(1), pp. 378–390.

Mirasgedis, S, Y Sarafidis, E Georgopoulou, DP Lalas, M Moschovits, F Karagiannis, and D Papakonstantinou (2006). "Models for mid-term electricity demand forecasting incorporating weather influences". In: *Energy* 31(2-3), pp. 208–227.

Morana, Claudio (2002). "Common persistent factors in inflation and excess nominal money growth and a new measure of core inflation". In: *Studies in Nonlinear Dynamics & Econometrics* 6(3).

Moulines, Eric, François Roueff, and Murad S. Taqqu (2007a). "Central Limit Theorem for the log-regression wavelet estimation of the memory parameter in the Gaussian semi-parametric context". In: *Fractals* 15(04), pp. 301–313.

Moulines, Eric, François Roueff, and Murad S. Taqqu (2007b). "On the spectral density of the wavelet coefficients of long-memory time series with application to the log-regression estimation of the memory parameter". In: *Journal of Time Series Analysis* 28(2), pp. 155–187.

Moulines, Eric, François Roueff, and Murad S. Taqqu (2008). "A wavelet Whittle estimator of the memory parameter of a nonstationary Gaussian time series". In: *The Annals of Statistics* 36(4), pp. 1925–1956.

Nualart, David (2006). "Fractional Brownian motion". In: *The Malliavin Calculus and Related Topics*. Springer: Berlin, Heidelberg, pp. 273–320.

Ooms, Marius and Uwe Hassler (1997). "On the effect of seasonal adjustment on the log-periodogram regression". In: *Economics Letters* 56(2), pp. 135–141.

Percival, Donald B and Andrew T Walden (2006). *Wavelet methods for time series analysis*. Cambridge University Press.

Perron, Pierre (2006). "Dealing with structural breaks". In: *Palgrave Handbook of Econometrics* 1(2), pp. 278–352.

Pipiras, Vladas and Murad S. Taqqu (2017). *Long-Range Dependence and Self-Similarity*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Qu, Zhongjun and Pierre Perron (2007). "Estimating and testing structural changes in multivariate regressions". In: *Econometrica* 75(2), pp. 459–502.

Ray, Bonnie K and Ruey S Tsay (2002). "Bayesian methods for change-point detection in long-range dependent processes". In: *Journal of Time Series Analysis* 23(6), pp. 687–705.

Robertson, Tim, F.T. Wright, and Dykstra R.L. (1988). *Order restricted statistical inference*. John Wiley & Sons: Chichester.

Robinson, Peter (1995a). "Gaussian semiparametric estimation of long range dependence". In: *The Annals of Statistics* 23(5), pp. 1630–1661.

Robinson, Peter (1995b). "Log-periodogram regression of time series with long range dependence". In: *The Annals of Statistics*, pp. 1048–1072.

Robinson, Peter (2005). "Robust covariance matrix estimation: HAC estimates with long memory/antipersistence correction". In: *Econometric Theory* 21(1), pp. 171–180.

Roueff, François and Rainer Von Sachs (2011). "Locally stationary long memory estimation". In: *Stochastic Processes and their Applications* 121(4), pp. 813–844.

Sadaei, Hossein Javedani, Frederico Gadelha Guimarães, Cidiney José da Silva, Muhammad Hisyam Lee, and Tayyebeh Eslami (2017). "Short-term load forecasting method based on fuzzy time series, seasonality and long memory process". In: *International Journal of Approximate Reasoning* 83, pp. 196–217.

Shao, Xiaofeng (2011). "A simple test of changes in mean in the possible presence of long-range dependence". In: *Journal of Time Series Analysis* 32(6), pp. 598–606.

Shimotsu, Katsumi (2007). "Gaussian semiparametric estimation of multivariate fractionally integrated processes". In: *Journal of Econometrics* 137(2), pp. 277–310.

Sibbertsen, Philipp (2004). "Long memory versus structural breaks: An overview". In: *Statistical Papers* 45(4), pp. 465–515.

Sibbertsen, Philipp and Robinson Kruse (2009). "Testing for a break in persistence under long-range dependencies". In: *Journal of Time Series Analysis* 30(3), pp. 263–285.

Sibbertsen, Philipp, Christian Leschinski, and Marie Busch (2018). "A multivariate test against spurious long memory". In: *Journal of Econometrics* 203(1), pp. 33–49.

Sibbertsen, Philipp and Juliane Willert (2012). "Testing for a break in persistence under long-range dependencies and mean shifts". In: *Statistical Papers* 53(2), pp. 357–370.

Soares, Lacir J. and Marcelo C. Medeiros (2008). "Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data". In: *International Journal of Forecasting* 24(4), pp. 630–644.

Soares, Lacir Jorge and Leonardo Rocha Souza (2006). "Forecasting electricity demand using generalized long memory". In: *International Journal of Forecasting* 22(1), pp. 17–28.

Souza, Leonardo R. and Lacir J. Soares (2007). "Electricity rationing and public response". In: *Energy Economics* 29(2), pp. 296–311.

Taylor, James W (2003). "Short-term electricity demand forecasting using double seasonal exponential smoothing". In: *Journal of the Operational Research Society* 54(8), pp. 799–805.

Taylor, James W, Lilian M De Menezes, and Patrick E McSharry (2006). "A comparison of univariate methods for forecasting electricity demand up to a day ahead". In: *International Journal of Forecasting* 22(1), pp. 1–16.

Tsay, Wen-Jen (2000). "Long memory story of the real interest rate". In: *Economics Letters* 67(3), pp. 325–330.

Wang, Lihong (2008). "Change-in-mean problem for long memory time series models with applications". In: *Journal of Statistical Computation and Simulation* 78(7), pp. 653–668.

Wang, Qiying, Yan-Xia Lin, and Chandra M Gulati (2003). "Strong approximation for long memory processes with applications". In: *Journal of Theoretical Probability* 16(2), pp. 377–389.

Wenger, Kai and Christian Leschinski (in press). "Fixed-bandwidth CUSUM tests under long memory". In: *Econometrics and Statistics*. in press.

Wenger, Kai, Christian Leschinski, and Philipp Sibbertsen (2018a). "A simple test on structural change in long-memory time series". In: *Economics Letters* 163, pp. 90–94.

Wenger, Kai, Christian Leschinski, and Philipp Sibbertsen (2018b). "The memory of volatility". In: *Quantitative Finance and Economics* 2(1), pp. 137–159.

Wenger, Kai, Christian Leschinski, and Philipp Sibbertsen (2019). "Change-in-mean tests in long-memory time series: a review of recent developments". In: *AStA Advances in Statistical Analysis* 103(2), pp. 237–256.

Weron, Rafal (2007). *Modeling and forecasting electricity loads and prices: A statistical approach*. John Wiley & Sons.

Whitcher, Brandon and Mark J Jensen (2000). "Wavelet estimation of a local long memory parameter". In: *Exploration Geophysics* 31(1/2), pp. 94–103.

Wu, Wei Biao (2004). "A test for detecting changes in mean". In: *Time Series Analysis and Applications to Geophysical Systems*. Springer, pp. 105–121.

Wu, Wei Biao (2007). "Strong invariance principles for dependent random variables". In: *The Annals of Probability* 35(6), pp. 2294–2320.

Wu, Wei Biao, Michael Woodroofe, and Graciela Mentz (2001). "Isotonic regression: Another look at the changepoint problem". In: *Biometrika* 88(3), pp. 793–804.