Testing for Multiple Structural Breaks in Multivariate Long Memory Regression Models

Vivien Less[a], Paulo M. M. Rodrigues[b] and Philipp Sibbertsen[a*]

[a] Leibniz University Hannover

[b] Banco de Portugal and Nova School of Business and Economics

February 11, 2025

Abstract

This paper focuses on the estimation and testing of multiple breaks that occur at unknown dates in multivariate long memory time series regression models, allowing for fractional cointegration. A likelihood-ratio based approach for estimating the breaks in the parameters and in the covariance of a system of long memory time series regressions is proposed. The limiting distributions as well as the consistency of the estimators are derived. Furthermore, a testing procedure to determine the unknown number of breaks is introduced which is based on iterative testing on the regression residuals. A Monte Carlo exercise shows the good finite sample properties of our novel approach, and empirical applications on inflation series of France and Germany and on benchmark government bonds of eight EMU countries illustrate the usefulness of the proposed procedures.

Key words: Multivariate Long Memory · Fractional Cointegration · Multiple Structural Breaks · Hypothesis Testing · Inflation · Government Bonds.

JEL classification: C12, C22, C58, G15

*Corresponding Author. Leibniz University Hannover, School of Economics and Management, Institute of Statistics, Königsworther Platz 1, D-30167 Hannover, Germany. E-Mail: sibbertsen@statistik.uni-hannover.de. Phone: +49-511-762-3783

1 Introduction

Testing for structural changes and estimating break points occurring at unknown dates has been a topic of significant interest in the economics and econometrics literature for a long time. A vast amount of research has focused on estimation and testing for single structural breaks in a univariate time series regression framework with weak correlations. Bai and Perron (1998) extended this literature by introducing estimators and tests for multiple break points at unknown dates in a univariate time series regression (see, for instance, Perron (2006) and Casini and Perron (2019), for interesting reviews).

Developments on estimation and testing for breaks in multivariate regression models are more limited. Bai (1997b) considers estimation of a single break in a multivariate regression set-up and, Bai et al. (1998) and Preuss et al. (2015) provide tests and estimators for common breaks in a multivariate system of short-memory time series. Qu and Perron (2007) propose a versatile framework for estimating and testing multiple, not necessarily common, breaks that occur at unknown dates in a multivariate short-memory time series regression framework, allowing for breaks in the mean as well as in the covariance of the system.

In contrast, in the long memory context, literature on structural breaks testing is more limited. The challenge in testing for structural breaks in time series that display long memory behaviour resides in the fact that both phenomena (structural breaks and long memory) are observationally equivalent in finite samples, i.e., long memory can cause false rejections of tests for structural changes and vice versa (see, for instance, Sibbertsen (2004) and Hassler et al. (2014), for an overview of literature addressing this problem).

Several approaches have been proposed to test for a single structural break in univariate long memory time series models; e.g., Wang (2008), Shao (2011), Dehling et al. (2013), Iacone et al. (2014), Betken (2016), Wenger and Leschinski (2019), and Wenger and Less (2020). A recent overview is provided in Wenger et al. (2019). Lavielle and Moulines (2000) and Yao (1988) estimated multiple breaks in a univariate framework using information criteria, with Lavielle and Moulines (2000) allowing for long-range dependence and Yao (1988) assuming an independent and identically distributed framework. However, empirical evidence suggests that information criteria tend to over-estimate the number of break points (see, for instance, Hall et al. (2013)).

In this paper we contribute to the literature on structural breaks in a multivariate regression context by introducing estimators and tests for multiple structural breaks that occur at unknown dates in a multivariate long memory time series regression framework with deterministic or stochastic regressors, allowing for fractional cointegration. To the best of our knowledge this is the first paper providing tests for multiple breaks under long memory and to consider breaks in a multivariate, possibly fractionally cointegrated, system.

Specifically, we generalise the framework of Qu and Perron (2007) in three directions. First, we use a likelihood-ratio based approach to estimate breaks in the mean and covariance of the system of long memory time series. We show consistency and provide the limiting distributions of these estimators under long memory. Second, we provide tests for multiple structural changes, generalizing the testing ideas of Bai and Perron (1998), which are based on the segmentation of the time series and repeated testing for breaks within these segments. Importantly, however, the limiting distribution of the Bai and Perron (1998) approach strongly depends on the assumption of at most weakly correlated segments, since it is derived as the product of the limiting distributions for each segment. However, this assumption is infeasible under long-range dependence, as the segments are strongly correlated and, consequently, the limiting distribution of the Bai and Perron (1998) test does not hold in this situation. We circumvent this problem by suggesting to repeatedly test for breaks in the residuals, after applying our consistent break point estimator and eliminating the largest break in each step. Third, we extend our procedures to stochastic, possibly non-stationary, fractionally integrated regressors and allow testing for breaks in fractional cointegration.

The procedures that we introduce only depend on the maximal memory parameter of the multivariate system of long memory time series. The limiting distribution of our test is different when all memory parameters are equal compared to when at least two of them are different. Additionally, in order to prove our results we need to derive a multivariate generalized Hájek-Rényi-type inequality under long-range dependence. The finite sample performance of the test procedures is analysed in a Monte Carlo study, and the empirical usefulness illustrated with two applications to real data, one where we examine a system of inflation series and another that focuses on benchmark government bonds.

This paper focuses on testing for structural breaks in the coefficients and the covariance matrix of a multivariate regression model with long-range dependence, assuming that the memory parameters remain constant over time, hence ruling out breaks in persistence. The latter are considered in, among others, Sibbertsen and Kruse (2009), Yamaguchi (2011), Martins and Rodrigues (2014) and Hassler and Meller (2014), and can easily be mistaken for mean shifts as they can lead to divergence of the test statistic due to wrong standardization; see Sibbertsen and Willert (2012). A procedure for disentangling these two phenomena in a univariate time series context can be found in Wingert et al. (2022). However, a generalization of this approach to the situation considered here is beyond the scope of this paper.

The remainder of the paper is organized as follows. Section 2 provides the model and the underlying assumptions. Section 3 introduces the break point estimator and Section 4 the testing procedure considering that regressors are deterministic. The case of stochastic, possibly fractionally integrated, regressors is given in Section 5. Section 6 contains a Monte Carlo study of the finite sample properties of the break point estimators and of the test procedures, and Section 7 provides two empirical applications. Section 8 concludes. Finally, an appendix collects the proofs for all results presented in the main text.

2 The Model and Assumptions

This paper focuses on the detection of structural changes in multivariate regression models allowing for long memory innovations. Specifically, an n dimensional system of time series $\boldsymbol{u}_t$ is said to exhibit multivariate long-range dependence or long memory with integration order $\boldsymbol{D} = (d_1, \dots, d_n)'$ and $-1/2 < d_k < 1/2$, for $k = 1, \dots, n$, if its spectral density behaves local to the origin, i.e.,

$$\boldsymbol{f}(\boldsymbol{\lambda}) \sim \boldsymbol{\Lambda}(\boldsymbol{D})\boldsymbol{G}\boldsymbol{\Lambda}(\boldsymbol{D})^*, \tag{1}$$

where $\boldsymbol{\Lambda}(\boldsymbol{D}) = \text{diag}(\Lambda_1(d_1), \dots, \Lambda_n(d_n))$, $\Lambda_k(d_k) = \lambda^{-d_k} e^{\mathrm{i}(\pi-\lambda)d_k/2}$, $k = 1, \dots, n$, i is the imaginary number and d_k the memory parameter of series k. $\boldsymbol{G}$ is a real, positive definite, finite and symmetric matrix and $\boldsymbol{\Lambda}(\boldsymbol{D})^*$ denotes the complex conjugate of $\boldsymbol{\Lambda}(\boldsymbol{D})$. Furthermore, in what follows, $d = \max\{d_1, \dots, d_n\}$.

Remark 1. *The assumption on $\boldsymbol{G}$ is standard in defining multivariate long memory and excludes fractional cointegration. However, for the estimators and tests we propose next it is of no relevance whether the series are fractionally cointegrated. We therefore stick to the standard assumption keeping in mind that relaxing the assumptions on $\boldsymbol{G}$ will not affect our procedures for now. We will, however, relax this assumption below in order to be able to handle breaks in a (fractionally) cointegrated framework.* ♢

Consider a system of n time series each of length T, with a total number m of structural changes. The break dates in the system are denoted by the $m \times 1$ vector $\mathcal{T} = (T_1, \ldots, T_m)$ and, for ease of calculation, we set $T_0 = 1$ and $T_{m+1} = T$. We use the convention that a subscript j indexes a regime ($j = 1, \ldots, m+1$), a subscript t indexes a temporal observation ($t = 1, \ldots, T$) and a subscript i indexes an equation ($i = 1, \ldots, n$). The number of regressors in each equation i is defined as q and $\boldsymbol{z}_t$ is the set which includes the regressors at time t from all equations, i.e., $\boldsymbol{z}_t = (\boldsymbol{z}'_{1t}, \ldots, \boldsymbol{z}'_{nt})'$ is an $nq \times 1$ vector. In what follows, we first consider the case of deterministic regressors but relax this assumption to allow for stochastic regressors in Section 5.

Consider the model,

$$\boldsymbol{y}_t = \boldsymbol{x}'_t \boldsymbol{\beta}_j + \boldsymbol{u}_t, \quad j = 1, ..., m+1, \tag{2}$$

where $\boldsymbol{y}_t$ is an $n \times 1$ vector, $\boldsymbol{x}_t = [(I \otimes \boldsymbol{z}'_t)\boldsymbol{S}]'$ is an $n \times p$ with $p \leq q$ matrix, I is an $n \times n$ identity matrix, $\boldsymbol{S}$ is an $nq \times p$ full column rank selection matrix with entries 0 and 1, and $\boldsymbol{u}_t$ is an $n \times 1$ vector error process (whose properties will be detailed below) with zero vector mean and covariance matrix $\boldsymbol{\Sigma}_j$, for $T_{j-1} + 1 \leq t \leq T_j$, $j = 1, \ldots, m+1$. The parameters to be estimated in regime j are given by the $p \times 1$ vector of parameters $\boldsymbol{\beta}_j$ and the matrix $\boldsymbol{\Sigma}_j$. To allow for restrictions in our model we consider,

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

where $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \ldots, \boldsymbol{\beta}'_{m+1})'$ is a $(m+1)p \times 1$ vector of parameters, $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_{m+1})$ is an $n \times (m+1)n$ matrix, and $\boldsymbol{g}(\cdot)$ is an r-dimensional vector of restrictions. If required, it is also possible to allow for cross-equation restrictions across regimes in this setting.

To simplify notation, we rewrite (2) using matrix notation. Let $\boldsymbol{Y} = (\boldsymbol{y}'_1, \ldots, \boldsymbol{y}'_T)'$ be the $nT \times 1$ vector of dependent variables, $\boldsymbol{U} = (\boldsymbol{u}'_1, \ldots, \boldsymbol{u}'_T)'$ the $nT \times 1$ error vector and $\boldsymbol{X} = (\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_T)'$ the $nT \times p$ matrix of regressors. Form the block partition of matrix $\boldsymbol{X}$, which we denote as $\widetilde{\boldsymbol{X}}$, such that for a given partition of the sample using the breaks at $T_1, \ldots, T_m$, $\widetilde{\boldsymbol{X}}$ is defined as the $nT \times p(m+1)$ matrix $\widetilde{\boldsymbol{X}} = \mathrm{diag}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{m+1})$, where $\boldsymbol{X}_j$, $j = 1, \ldots, m+1$, is the $n(T_j - T_{j-1}) \times p$ subset of $\boldsymbol{X}$ that corresponds to observations in regime j. Similarly, define the subvector $\boldsymbol{U}_j$ of $\boldsymbol{U}$. Hence, using this notation, the regression system in (2) can be expressed as $\boldsymbol{Y} = \widetilde{\boldsymbol{X}}\boldsymbol{\beta} + \boldsymbol{U}$. Denoting the true values of the parameters with a 0 superscript, the data generating process is $\boldsymbol{Y} = \widetilde{\boldsymbol{X}}^0 \boldsymbol{\beta}^0 + \boldsymbol{U}$. Here, $\widetilde{\boldsymbol{X}}^0$ is the diagonal partition of $\boldsymbol{X}$ based on the true break dates $T^0_1, \ldots, T^0_m$.

To derive our approach we impose the following set of assumptions, which are similar to those of Qu and Perron (2007) but, with the important difference that the errors, $\boldsymbol{u}_t$, are allowed to be long-range dependent. We assume that the memory is introduced through $\boldsymbol{u}_t$ and thus that $\boldsymbol{x}_t$ is mostly short-range dependent so that the process $\boldsymbol{x}_t'\boldsymbol{u}_t$ is of the same order of integration as $\boldsymbol{u}_t$. This is a simplifying assumption which is not necessary and will be relaxed when introducing the stochastic regressors framework.

Assumption 1. For each $j = 1, \ldots, m+1$ and $l_j \leq T_j^0 - T_{j-1}^0$, $l_j^{-1}\sum_{t=T_{j-1}^0+1}^{T_{j-1}^0+l_j} \boldsymbol{x}_t\boldsymbol{x}_t' \xrightarrow{\text{a.s.}} \boldsymbol{Q}_j^0$ as $l_j \to \infty$, with $\boldsymbol{Q}_j^0$ 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 $l^{-1}\sum_{t=T_j^0+1}^{T_j^0+l} \boldsymbol{x}_t\boldsymbol{x}_t'$ and of $l^{-1}\sum_{t=T_j^0-l}^{T_j^0} \boldsymbol{x}_t\boldsymbol{x}_t'$, $j = 1, ..., m$, are bounded away from zero.

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

Assumption 4. $\boldsymbol{u}_t$ has multivariate long memory dynamics, as defined in (1), with

$$\boldsymbol{u}_t - E[\boldsymbol{u}_t] = \boldsymbol{A}(L)\boldsymbol{\varepsilon}_t = \sum_{j=0}^{\infty} \boldsymbol{A}_j\boldsymbol{\varepsilon}_{t-j},$$

where $\sum_{j=0}^{\infty} \|\boldsymbol{A}_j\|^2 < \infty$ and $\|\cdot\|$ denotes the supremum norm. It is assumed that $E(\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1}) = 0$, and $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^{'}|\mathcal{F}_{t-1}) = \boldsymbol{I}_q$ a.s. for $t = 0, \pm 1, \pm 2, \ldots$, where $\mathcal{F}_t$ denotes the σ-field generated by $\boldsymbol{\varepsilon}_s$, $s \leq t$ and $\boldsymbol{I}_q$ is a $q \times q$ identity matrix. Furthermore, there exists a scalar random variable ϵ such that $E(\epsilon^2) < \infty$, and for all $\tau > 0$ and some $K > 0$ it holds that $P(\|\epsilon\|^2 > \tau) \leq KP(\epsilon^2 > \tau)$. In addition, for $a, b, c, d = 1, 2$, and $t = 0, \pm 1, \pm 2, \ldots$, $E(\varepsilon_{at}\varepsilon_{bt}\varepsilon_{ct}|\mathcal{F}_{t-1}) = \mu_{abc}$ a.s. and $E(\varepsilon_{at}\varepsilon_{bt}\varepsilon_{ct}\varepsilon_{dt}|\mathcal{F}_{t-1}) = \mu_{abcd}$ a.s., where $|\mu_{abc}| < \infty$ and $|\mu_{abcd}| < \infty$.

Assumption 5. Assumption 4 holds with $\boldsymbol{u}_t$ replaced by $\boldsymbol{x}_t'\boldsymbol{u}_t$ or $\boldsymbol{u}_t\boldsymbol{u}_t' - \boldsymbol{\Sigma}_j^0$, for $T_{j-1}^0 < t \leq T_j^0$, $j = 1, \ldots, m+1$.

Assumption 6. The magnitudes of the shifts satisfy $\boldsymbol{\beta}_{T,j+1}^0 - \boldsymbol{\beta}_{T,j}^0 = \nu_T\boldsymbol{c}_{1j}$ and $\boldsymbol{\Sigma}_{j+1,T}^0 - \boldsymbol{\Sigma}_{j,T}^0 = \nu_T\boldsymbol{c}_{2j}$, where $(\boldsymbol{c}_{1j}, \boldsymbol{c}_{2j}) \neq 0$ are vectors of constants 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. $(\boldsymbol{\beta}^0, \boldsymbol{\Sigma}^0) \in \bar{\boldsymbol{\Theta}}$ with $\bar{\boldsymbol{\Theta}} = \{(\boldsymbol{\beta}, \boldsymbol{\Sigma}) : \|\boldsymbol{\beta}\| \leq \boldsymbol{c}_1, \lambda_{\min}(\boldsymbol{\Sigma}) \geq \boldsymbol{c}_2,$ and $\lambda_{\max}(\boldsymbol{\Sigma}) \leq \boldsymbol{c}_3\}$ for some $\boldsymbol{c}_1 < \infty, 0 < \boldsymbol{c}_2 \leq \boldsymbol{c}_3 < \infty$, where $\lambda_{\min}$ and $\lambda_{\max}$ denote the smallest and largest eigenvalues, respectively.

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

Our assumptions include the standard fractionally integrated vector autoregressive moving average (FIVARMA) model as well as long memory panel data models and regression models with exogenous regressors and long memory errors. However, unit root regressors are ruled out by Assumption 1. Moreover, 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 to ensure break identification. Assumption 3 is a standard invertibility assumption. Assumptions 4 and 5 state that we consider a long memory regression framework and that the order of integration is solely determined from $\boldsymbol{u}_t$. Our conditions allow for conditional heteroscedasticity in $\boldsymbol{u}_t$. However, perturbation in the error term is excluded as this would bias estimation of the memory parameter and thus lead to inconsistent testing. 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 date 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 stronger the persistence of the errors the larger the break needs 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. No other assumptions on the breaks are needed. This includes that the breaks do not need to occur contemporaneously in each series. So we allow each series to have breaks at different times or not to break at all.

Additionally, in order to derive the limiting distribution of the test under the null hypothesis of no structural change, we also require the following additional assumptions.

Assumption 9. $T^{-1}\sum_{t=1}^{\lfloor Ts \rfloor}\boldsymbol{x}_t\boldsymbol{x}_t' \xrightarrow{p} s\boldsymbol{Q}$, uniformly in $s \in [0,1]$, with $\boldsymbol{Q}$ being some positive definite matrix.

Assumption 10. The errors $\{\boldsymbol{u}_t\}$ form an array of long-range dependent processes as defined in Assumption 4 and, additionally, $E(\boldsymbol{u}_t\boldsymbol{u}_t') = \boldsymbol{\Sigma}^0$ for all t, and $T^{-1/2-D}\sum_{t=1}^{[Ts]}\boldsymbol{x}_t'\boldsymbol{u}_t \Rightarrow \Phi^{1/2}\boldsymbol{W}_D(s)$, where $\Phi = \operatorname*{plim}_{T\to\infty} T^{-1}\boldsymbol{X}'(\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}^0)\boldsymbol{X}$ and $\boldsymbol{W}_D(s)$ is a vector of independent fractional Brownian motions of type I[1]. Also, with $\boldsymbol{\eta}_t \equiv (\eta_{t1},\ldots,\eta_{tn})' = (\boldsymbol{\Sigma}^0)^{-1/2}\boldsymbol{u}_t$, we have $T^{-1/2-D}\sum_{t=1}^{[Ts]}(\boldsymbol{\eta}_t\boldsymbol{\eta}_t' - \boldsymbol{I}_n) \Rightarrow \boldsymbol{\xi}_D(s)$, where $\boldsymbol{\xi}_D(s)$ is an $n \times n$ matrix of fractional Brownian motion processes with $\boldsymbol{\Omega} = \mathrm{Var}(\mathrm{vec}(\boldsymbol{\xi}_D(1)))$. Moreover, $E[\eta_{tk}\eta_{tl}\eta_{th}] = 0$, for all k, l, h and for every t.

[1]For a detailed discussion of fractional Brownian motions of type I and type II cf. Marinucci and Robinson (1999).

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 that 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.

3 Estimation of the Break Dates and Model Parameters

The dates and number of breaks are estimated by restricted Quasi-Maximum Likelihood (QML) conditional on a given partition of the sample $\mathcal{T} = (T_1, \ldots, T_m)$, and the tests for the number of breaks are based on likelihood-ratio statistics.

Assuming Gaussian serially uncorrelated errors the quasi-likelihood function is,

$$L_T(\mathcal{T}, \boldsymbol{\beta}, \boldsymbol{\Sigma}) = \prod_{j=1}^{m+1} \prod_{t=T_{j-1}+1}^{T_j} f(\boldsymbol{y}_t | \boldsymbol{x}_t; \boldsymbol{\beta}_j, \boldsymbol{\Sigma}_j),$$

where

$$f(\boldsymbol{y}_t | \boldsymbol{x}_t; \boldsymbol{\beta}_j, \boldsymbol{\Sigma}_j) = \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}_j|^{1/2}} \exp\left(-\frac{1}{2} [\boldsymbol{y}_t - \boldsymbol{x}_t' \boldsymbol{\beta}_j]' \boldsymbol{\Sigma}_j^{-1} [\boldsymbol{y}_t - \boldsymbol{x}_t' \boldsymbol{\beta}_j] \right),$$

and the quasi-likelihood-ratio statistic is,

$$LR_T(\mathcal{T}, \boldsymbol{\beta}, \boldsymbol{\Sigma}) = \frac{\prod_{j=1}^{m+1} \prod_{t=T_{j-1}+1}^{T_j} f(\boldsymbol{y}_t | \boldsymbol{x}_t; \boldsymbol{\beta}_j, \boldsymbol{\Sigma}_j)}{\prod_{j=1}^{m+1} \prod_{t=T_{j-1}^0+1}^{T_j^0} f(\boldsymbol{y}_t | \boldsymbol{x}_t; \boldsymbol{\beta}_j^0, \boldsymbol{\Sigma}_j^0)}.$$

We aim to estimate the values of $(T_1, \ldots, T_m, \boldsymbol{\beta}, \boldsymbol{\Sigma})$ under the restriction $\boldsymbol{g}(\boldsymbol{\beta}, \mathrm{vec}(\boldsymbol{\Sigma})) = \mathbf{0}$, by maximizing the restricted quasi-likelihood ratio statistic,

$$RLR_T(\mathcal{T}, \boldsymbol{\beta}, \boldsymbol{\Sigma}) = LR_T(\mathcal{T}, \boldsymbol{\beta}, \boldsymbol{\Sigma}) + \boldsymbol{l}' \boldsymbol{g}(\boldsymbol{\beta}, \mathrm{vec}(\boldsymbol{\Sigma})), \tag{3}$$

where $\boldsymbol{l}$ is an r-dimensional vector of weights.

Before proceeding, the following assumption on the minimal regime length is also required.

Assumption 11. The maximization of (3) is taken over all partitions $\mathcal{T} = (T_1, \ldots, T_m) = (T\lambda_1, \ldots, T\lambda_m)$, for some $\epsilon > 0$, in the set

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

This assumption is standard in the structural breaks literature and states that some percentage of the data needs to be skipped at the beginning and end of the sample before maximization of the likelihood and thus, that potential breaks cannot occur in a possible small window at the beginning and end of the sample period. 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 break point 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 no longer holds under long memory and needs to be replaced by a law of iterated logarithm for fractional Brownian motions, the arguments used by Qu and Perron (2007) to prove Property 2 are no longer valid and the property does not apply. However, these arguments are needed for the endpoints and therefore Assumption 11 circumvents this problem.

We can now establish the rate of convergence of this estimator under long-range dependence. Note that all results that follow depend on the vector of memory parameters $\boldsymbol{D}$. For the time being, we assume $\boldsymbol{D}$ to be known and comment on the estimation of the memory parameters, to make the procedures feasible, after presenting all of our results.

Lemma 1. *Under Assumptions 1 to 8 and 11 it follows, for $j = 1, \dots, m$, that $\nu_T^2(\hat{T}_j - T_j^0) = O_P(1)$, and for $j = 1, \dots, m+1$, that $T^{1/2-d}(\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j^0) = O_P(1)$ and $T^{1/2-d}(\hat{\boldsymbol{\Sigma}}_j - \boldsymbol{\Sigma}_j^0) = O_P(1)$.*

The results in Lemma 1 are similar to those in Bai (1997b), Bai and Perron (1998), Bai (2000), and Qu and Perron (2007), but with a major difference, in that they hold under long-range dependence in the error terms. Also, the rate for the break date estimator is fast enough not to affect the estimation of the model parameters asymptotically. Therefore, we have the following result that we state without proof.

Lemma 2. *Under the Assumptions of Lemma 1, the limiting distribution of $T^{1/2-d}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)$ when the break dates are consistently estimated is the same as that under known break dates.*

These results are necessary for our tests on the number of potential break points later. 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 containing the break dates and the true parameter values, so that restrictions to these values do not affect the estimation of the break dates; and the other part involving the true values of the break dates, model parameters and restrictions, so 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, this enables us to show that Theorem 2 of Qu and Perron (2007) also holds under long-range dependence. This result concerns the limiting distribution of the break dates. However, the drawback of this result is that the limiting distribution of the break dates depends on the true error distribution, which is a standard problem in the structural breaks literature and is usually accounted for by assuming shrinking breaks with an increasing sample size. 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 $s \in [0,1]$, $(\Delta T_j^0)^{-1} \sum_{t=T_{j-1}^0+1}^{T_{j-1}^0+[s\Delta T_j^0]} \boldsymbol{x}_t \boldsymbol{x}_t' \xrightarrow{P} s\boldsymbol{Q}_j^0$, with $\boldsymbol{Q}_j^0$ being a nonrandom positive definite matrix not necessarily the same for all j.

With this additional assumption the following Theorem characterizing the limiting distribution for the break date estimators can be stated.

Theorem 1. *Let $\boldsymbol{\eta}_t = (\eta_{t1}, \ldots, \eta_{tn}) = (\boldsymbol{\Sigma}_j^0)^{-1/2}\boldsymbol{u}_t$, 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, 11 and 12 with $(\boldsymbol{c}_{1,j}, \boldsymbol{c}_{2,j})$ as in Assumption 6 and $\nu_t \to 0$, such that $T^{1/2-d}\nu_T/(\log T)^2 \to \infty$, it follows, for $j = 1, \ldots, m$, as $T \to \infty$ that,*

$$\frac{\boldsymbol{\Delta}_{1,j}^2}{\boldsymbol{\Gamma}_{1,j}^2}\nu_T^2(\hat{T}_j - T_j^0) \Rightarrow \begin{cases} -\frac{|u|}{2} + W_{j,d}(u), & \text{for } u \leq 0 \\ -\frac{|u|}{2}\frac{\boldsymbol{\Delta}_{2,j}}{\boldsymbol{\Delta}_{1,j}} + \frac{\boldsymbol{\Gamma}_{2,j}}{\boldsymbol{\Gamma}_{1,j}}W_{j,d}(u), & \text{for } u > 0, \end{cases} \tag{4}$$

where $\Rightarrow$ denotes weak convergence under the Skorokhod topology, $W_{j,d}(u)$ is a fractional Wiener *process defined on the real line and*

$$\begin{aligned}
\boldsymbol{\Delta}_{1,j} =& \frac{1}{2}\operatorname{tr}(\boldsymbol{A}_{1,j}^2 + \boldsymbol{c}_{1,j}^{'}\boldsymbol{Q}_{1,j}\boldsymbol{c}_{1,j}),\\
\boldsymbol{\Delta}_{2,j} =& \frac{1}{2}\operatorname{tr}(\boldsymbol{A}_{2,j}^2 + \boldsymbol{c}_{1,j}^{'}\boldsymbol{Q}_{2,j}\boldsymbol{c}_{1,j}),\\
\boldsymbol{A}_{1,j} =& (\boldsymbol{\Sigma}_j^0)^{1/2}(\boldsymbol{\Sigma}_{j+1}^0)^{-1}\boldsymbol{c}_{2,j}(\boldsymbol{\Sigma}_j^0)^{-1/2},\\
\boldsymbol{A}_{2,j} =& (\boldsymbol{\Sigma}_{j+1}^0)^{1/2}(\boldsymbol{\Sigma}_j^0)^{-1}\boldsymbol{c}_{2,j}(\boldsymbol{\Sigma}_{j+1}^0)^{-1/2},\\
\boldsymbol{\Gamma}_{1,j} =& \left(\frac{1}{4}\operatorname{vec}(\boldsymbol{A}_{1,j})^{'}\boldsymbol{\Omega}_{1,j}^0\operatorname{vec}(\boldsymbol{A}_{1,j}) + \boldsymbol{c}_{1,j}^{'}\boldsymbol{\Pi}_{1,j}\boldsymbol{c}_{1,j}\right)^{1/2},\\
\boldsymbol{\Gamma}_{2,j} =& \left(\frac{1}{4}\operatorname{vec}(\boldsymbol{A}_{2,j})^{'}\boldsymbol{\Omega}_{2,j}^0\operatorname{vec}(\boldsymbol{A}_{2,j}) + \boldsymbol{c}_{1,j}^{'}\boldsymbol{\Pi}_{2,j}\boldsymbol{c}_{1,j}\right)^{1/2},\\
\boldsymbol{Q}_{1,j} =& \operatorname*{plim}_{T\to\infty}(T_j^0 - T_{j-1}^0)^{-1}\sum_{t=T_{j-1}^0+1}^{T_j^0}\boldsymbol{x}_t(\boldsymbol{\Sigma}_{j+1}^0)^{-1}\boldsymbol{x}_t^{'},\\
\boldsymbol{Q}_{2,j} =& \operatorname*{plim}_{T\to\infty}(T_{j+1}^0 - T_j^0)^{-1}\sum_{t=T_j^0+1}^{T_{j+1}^0}\boldsymbol{x}_t(\boldsymbol{\Sigma}_j^0)^{-1}\boldsymbol{x}_t^{'},
\end{aligned}$$

$$\boldsymbol{\Pi}_{1,j} = \lim_{T\to\infty}\operatorname{Var}\left\{(T_j^0 - T_{j-1}^0)^{-1/2}\left[\sum_{t=T_{j-1}^0+1}^{T_j^0}\boldsymbol{x}_t(\boldsymbol{\Sigma}_{j+1}^0)^{-1}(\boldsymbol{\Sigma}_j^0)^{1/2}\boldsymbol{\eta}_t\right]\right\},$$

$$\boldsymbol{\Pi}_{2,j} = \lim_{T\to\infty}\operatorname{Var}\left\{(T_{j+1}^0 - T_j^0)^{-1/2}\left[\sum_{t=T_j^0+1}^{T_{j+1}^0}\boldsymbol{x}_t(\boldsymbol{\Sigma}_j^0)^{-1}(\boldsymbol{\Sigma}_{j+1}^0)^{1/2}\boldsymbol{\eta}_t\right]\right\},$$

and

$$\boldsymbol{\Omega}_{1,j}^0 = \lim_{T\to\infty}\operatorname{Var}\left(\operatorname{vec}\left[(T_j^0 - T_{j-1}^0)^{-1/2}\sum_{t=T_{j-1}^0+1}^{T_j^0}(\boldsymbol{\eta}_t\boldsymbol{\eta}_t^{'} - \boldsymbol{I}_n)\right]\right),$$

$$\boldsymbol{\Omega}_{2,j}^0 = \lim_{T\to\infty}\operatorname{Var}\left(\operatorname{vec}\left[(T_{j+1}^0 - T_j^0)^{-1/2}\sum_{t=T_j^0+1}^{T_{j+1}^0}(\boldsymbol{\eta}_t\boldsymbol{\eta}_t^{'} - \boldsymbol{I}_n)\right]\right).$$

4 Testing for Multiple Breaks in Multivariate Time Series

In this section we 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 breaks against the alternative of a prespecified number of breaks, whereas the second approach 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 testing all parameters or only a subset of the coefficients of the regressors, $\boldsymbol{\beta}$, or of the covariance matrix of the errors, $\boldsymbol{\Sigma}_j$, for change per regime j, where $1 \leq j \leq m$. We acknowledge this dependence in the specification of our test statistic by considering the total number, p_b, of coefficients allowed to change across regimes, and allowing n_{bd} diagonal entries of $\boldsymbol{\Sigma}_j$ and n_{bo} entries in the upper triangle of $\boldsymbol{\Sigma}_j$ to change across regimes.

Specifically, consider the system specification,

$$\boldsymbol{y}_t = \boldsymbol{x}'_{at}\boldsymbol{\beta}_a + \boldsymbol{x}'_{bt}\boldsymbol{\beta}_{bj} + \boldsymbol{u}_t, \quad \text{for } T_{j-1}+1 \leq t < T_j \text{ and } j = 1, \ldots, m+1,$$

where $\boldsymbol{\beta}_{bj}$ is a p_b dimensional vector. Moreover, the covariance matrix of the errors is,

$$\boldsymbol{\Sigma}_j = E(\boldsymbol{u}_t\boldsymbol{u}'_t), \quad \text{for } T_{j-1}+1 \leq t < T_j \text{ and } j = 1, \ldots, m+1.$$

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

4.1 The Likelihood-Ratio Test

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

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

Denoting the log-likelihood value by $\log \hat{L}_T(T_1, \ldots, T_m)$, the test statistic is the supremum of the likelihood-ratio over all admissible partitions in the set $\boldsymbol{\Lambda}_\epsilon$ defined in Assumption 11, i.e.,

$$\begin{aligned}\frac{1}{T^{2d}} \sup LR_T(m, p_b, n_{bd}, n_{bo}, \epsilon) &= \frac{2}{T^{2d}} \sup_{(\lambda_1,\ldots,\lambda_m)\in\Lambda_\epsilon} \left[\log \hat{L}_T(T_1, \ldots, T_m) - \log \tilde{L}_T\right] \\ &= \frac{2}{T^{2d}} [\log \hat{L}_T(\hat{T}_1, \ldots, \hat{T}_m) - \log \tilde{L}_T], \end{aligned} \tag{5}$$

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

Theorem 2. *Under Assumptions 1-11 with the* $\sup LR_T(m,\ p_b,\ n_{bd},\ n_{bo},\ \epsilon)$ *test in (5) constructed for an alternative hypothesis* $\mathcal{H}_1$ *in the class of models described in this Section, as* $T \to \infty$ *it follows that,*

$$\frac{1}{T^{2d}} \sup LR_T(m, p_b, n_{bd}, n_{bo}, \epsilon) \Rightarrow \sup_{(\lambda_1,\ldots,\lambda_m)\in\boldsymbol{\Lambda}_\epsilon} \sum_{j=1}^{m} LR_j(\boldsymbol{\lambda}, d, p_b, n_b^*), \tag{6}$$

with

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

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

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

where $W_{d_j}(.)$ *is a univariate fractional Brownian motion of type I with memory parameter* d_j.

Remark 2. *Note that the limiting distribution in (6) depends on the number of series with the highest (or maximum) memory parameter d, i.e., only the series in the test statistic with the highest memory parameter in the vector* $\boldsymbol{D} = (d_1, \dots, d_n)$ *contribute asymptotically to the limiting distribution. This, however, poses a problem in practice as it is unlikely for estimated memory parameters* d_i, $i = 1, ..., n$, *to be equal so that the dimension of the limiting distribution may always be one in practice. To circumvent this problem we suggest applying the Robinson and Yajima (2002) test for equality of the memory parameters to get the correct dimension of the limiting distribution.* ◇

4.2 The Double Maximum Test

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

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

Following Bai and Perron (1998), we consider the double maximum test statistic,

$$
UDmax\, LR_T(m, p_b, n_{bd}, n_{bo}, \epsilon) = \max_{1\le m\le M} \sup LR_T(m, p_b, n_{bd}, n_{bo}, \epsilon). \tag{7}
$$

The asymptotic distribution for this test can be obtained in the setting of Theorem 2, and the following Theorem can be provided.

Theorem 3. *Under the assumptions of Theorem 2 it follows that,*

$$UDmax\, LR_T(m, p_b, n_{bd}, n_{bo}, \epsilon) \;\Rightarrow\; \max_{1\leq m\leq M} \sup_{(\lambda_1,\ldots,\lambda_m)\in\boldsymbol{\Lambda}_\epsilon} \sum_{j=1}^{m} LR_j(\boldsymbol{\lambda}, d, p_b, n_b^*).$$

An important property of the $UDmax\, LR_T$ test is that it 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.

4.3 The Breaks Determination Approach

In this section we introduce an iterative method to determine the unknown number of breaks in a multivariate system of long memory time series. It is based on repeated application of the $UDmax\, LR_T$ test in (7). The method requires fixing in advance an upper bound, M, on the number of breaks. For convenience of notation we define the novel residual based iterative procedure as REBIT.

Algorithm 1 (The REBIT algorithm)**.**

(i) *Set* $m = 0$.

(ii) *Estimate* m *breaks in the original system of time series* $\boldsymbol{y}_t$ *and save the residuals.*

(iii) *Estimate the memory of the process and conduct the* $UDmax\, LR_T$ *test with* $H_0\colon l = 0$ *vs.* $H_1\colon 1 \leq l \leq M - m$ *on the residuals.*

(iv) *:*

- *(iv.a)* *If the test rejects the null hypothesis and* $m < (M-1)$ *then set* $m = m+1$ *and reiterate from (ii).*
- *(iv.b)* *If the test cannot reject the null hypothesis then the detected number of breaks is* m. *Furthermore, if* $m = M-1$ *then the number of breaks is greater or equal to the previously chosen upper bound* M.

(v) *The approach ends when the test can either not reject the null hypothesis or the user chosen upper bound is reached - number of detected breaks is* m.

Remark 3. *The* $\sup LR_T$ *is not applicable in the suggested algorithm since a true break number* k, *such that* $k \neq 0$ *or* $k \neq m$, *is neither covered by the null nor by the alternative hypotheses.* $\diamondsuit$

Estimation of the break dates in step (ii) 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 1 we thus obtain consistency of our break point estimates in each step. The break point estimates in underspecified models are consistent, as shown by Bai (1997a) and Bai and Perron (1998) for breaks in the mean. By similar methods one can show that our procedure estimates some true break points even if the number of true break points is underspecified.

This iterative procedure avoids splitting the sample, as suggested, for example, in Bai (1997b), which is not possible under long memory and allows us to use the limiting results in Theorems 1, 2 and 3 which are derived under long-range dependence. The following Theorem states that our procedure has a hit rate of $(1-\alpha)$, where α is the nominal significance level of the break point test in Theorem 2.[2]

Theorem 4. *Let α be the nominal significance level of the break point test in Theorem 3. Under Assumptions 1-11, as $T \to \infty$, the asymptotic hit rate of the REBIT procedure is $(1-\alpha)$.*

Remark 4. *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 considered here as the sample size is fixed in practice.* ♢

Remark 5. *All the limiting distributions in this section and also those in the next sections depend on the underlying memory parameter d which is usually unknown in practice. As, to the best of our knowledge, there is no $\sqrt{T}$-consistent estimator for d available if series have structural changes, we use the following approach to make our procedure feasible in practice. The memory parameter is always estimated in the residuals of the test regression. As the break dates can be estimated consistently without knowledge of the memory parameter we start by estimating the break dates first. In a second step we estimate the regression parameters segment wise and compute the residuals which under the null hypothesis do not contain any structural breaks. Here, the memory parameter can be $\sqrt{T}$-consistently estimated by Maximum Likelihood.* ♢

[2] The hit rate gives the proportion of correctly specified number of breaks in the system.

5 Stochastic Regressors

In what follows we relax the assumption of deterministic regressors to allow for stochastic regressors including the case of fractional cointegration. Consider, as in (2), the linear regression model,

$$\boldsymbol{y}_t = \boldsymbol{x}_t^{'}\boldsymbol{\beta}_j + \boldsymbol{u}_t, \quad j = 1, ..., m+1, \tag{8}$$

where $\boldsymbol{x}_t = \left[(\boldsymbol{I} \otimes \boldsymbol{z}_t^{'})\boldsymbol{S}\right]^{'}$, is as in Section 4, but where now $\boldsymbol{x}_t$ can also be fractionally integrated of order d_x, with $1/2 < d_x \leq 1$. For the error term $\boldsymbol{u}_t$ we assume that $d_u < d_x$, which implies fractional cointegration. The set-up includes as a special case the $I(1) - I(0)$ cointegration case, if $d_x = 1$ and $d_u = 0$. We do not restrict the memory parameters to be equal across the individual time series.

In addition to Assumptions 1 - 12 we need to impose the following assumptions:

Assumption 13. $\boldsymbol{x}_t$ is $I(d_x)$ with $1/2 < d_x \leq 1$.

Assumption 14. The sequences $\boldsymbol{\varepsilon}_t$ from Assumption 4 and $\boldsymbol{x}_t$ are independent, that is,

$$E\left([\boldsymbol{x}_t'\boldsymbol{\varepsilon}_t][\boldsymbol{x}_t'\boldsymbol{\varepsilon}_t]'\right) = \boldsymbol{\Omega}_{\boldsymbol{\varepsilon}}\boldsymbol{\Omega}_{\boldsymbol{x}},$$

where $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \boldsymbol{\Omega}_{\boldsymbol{\varepsilon}}$ and $E(\boldsymbol{x}_t\boldsymbol{x}_t') = \boldsymbol{\Omega}_{\boldsymbol{x}}$

In addition we need to strengthen Assumption 6 as, in the case of fractional cointegration, the breaks in the cointegrating regressors would dominate breaks in the variance of the error term.

Assumption 15. The magnitudes of the shifts satisfy $\boldsymbol{\beta}_{T,j+1}^0 - \boldsymbol{\beta}_{T,j}^0 = \nu_T T^{-d_x/2}\boldsymbol{c}_{1j}$ and $\boldsymbol{\Sigma}_{j+1,T}^0 - \boldsymbol{\Sigma}_{j,T}^0 = \nu_T \boldsymbol{c}_{2j}$, where $(\boldsymbol{c}_{1j}, \boldsymbol{c}_{2j}) \neq 0$ are vectors of constants which are 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 $T^{1/2-d_u}\nu_T/(\log T)^2 \to \infty$.

Our assumptions allow for a very general setting of fractional cointegration. We rule out endogeneity by Assumption 14. The assumptions on the breaks and the break dates remain the same as for the deterministic regressors case discussed in the previous section. Moreover, although a combination of stochastic and deterministic regressors, such as linear trends, could also be considered, for the sake of clarity of the representation, this will be omitted here.

With these assumptions we can state that the consistency of the break point estimator remains unchanged with the same rate of convergence as in the case of deterministic regressors. Specifically, the following Lemma can be stated.

Lemma 3. *Under Assumptions 13 to 15, Lemmas 1 and 2 still hold for model (8) with a new convergence rate for the regression parameters,* $T^{d_x-d_u}(\hat{\boldsymbol{\beta}}_j-\boldsymbol{\beta}_j^0)$ *for* $d_x+d_u>1$ *and* $T^{2d_x-1}(\hat{\boldsymbol{\beta}}_j-\boldsymbol{\beta}_j^0)$ *for* $d_x+d_u\leq 1$*. In addition, Theorem 1 still holds.*

Remark 6. *Note that the limiting distribution of the break dates estimator in the case of stochastic regressors is no longer Gaussian as it contains the limit of sums of the regressors and the error terms.* ♢

As now stochastic regressors are considered, the limiting distribution of the test changes and becomes non-Gaussian. Specifically, define the process,

$$\Xi_{Qu}=\frac{1}{\Gamma(d_u)}\int_0^1(1-s)^{d_u-1}\int_0^r Q(s)dW(s)ds,$$

where $W(s)$ denotes a standard Brownian motion and $Q(s)$ is a fractional Brownian motion with parameter $1-d_x$ as defined in Assumption 13. For the case when $d_x+d_u>1$ we use the limiting result, $T^{d_x-d_u}\sum_{t=1}^T x_t u_t \xrightarrow{d} \Xi_{Qu}$ and when $d_x+d_u\leq 1$, $T^{2d_x-1}\sum_{t=1}^T x_t u_t \xrightarrow{d} \Xi_{Qu}$.

Theorem 5. *Under Assumptions 4, 7, 8 and 13 to 15 with the* $\sup LR_T(m, p_b, n_{bd}, n_{bo}, \epsilon)$ *test (adopting the notation from the previous section), constructed for an alternative hypothesis* $\mathcal{H}_1$ *in the class of models described in the previous Section, it follows that:*

i)

$$\frac{1}{T^{2d_u}}\sup LR_T(m,p_b,n_{bd},n_{bo},\epsilon)\Rightarrow \sup_{(\lambda_1,\dots,\lambda_m)\in\boldsymbol{\Lambda}_\epsilon}\sum_{j=1}^m LR_j(\boldsymbol{\lambda},d,p_b,n_b^*),$$

with

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

where $\boldsymbol{\lambda}=(\lambda_1,\dots,\lambda_m)$ *and* $\lambda_{m+1}=1$*. The vectors* $\boldsymbol{\Xi}^*_{d,p_b}(\cdot)$ *and* $\boldsymbol{W}^*_{d,n_b^*}(\cdot)$ *are of dimensions* p_b *and* $n_b^*=(n_{bd}+2n_{bo})$*, respectively.*

ii) The limiting distribution of the $UDmaxLR_T$ test in Theorem 3 changes accordingly.

iii) The REBIT procedure and Theorem 4 hold unchanged.

6 Simulation results

In what follows we conduct a Monte Carlo simulation exercise to analyse the finite sample properties of the break point estimator and of our proposed REBIT procedure. The data generation process is the bivariate model,

$$y_{1t} = x_{1t}\beta_{1j} + u_{1t}, \tag{9}$$

$$y_{2t} = x_{2t}\beta_{2j} + u_{2t}, \quad j = 1, ..., m+1, \tag{10}$$

with $\boldsymbol{u}_t = (u_{1t}, u_{2t})'$ a fractionally integrated white noise process, $\boldsymbol{x}_t = (x_{1t}, x_{2t})'$ a fractionally integrated process, and β_{ij}, $i = 1, 2$ the cointegration parameter in regime j. The vector of long memory parameters of the error term is $\boldsymbol{D}_u = (d_{1u}, d_{2u}) = \{(0.2, 0.2), (0.4, 0.2)\}$ and of the stochastic regressors $\boldsymbol{D}_x = (d_{1x}, d_{2x}) = \{(0.6, 0.4), (0.6, 0.6), (0.6, 0.8)\}$. The simulation study focuses on breaks in the fractional cointegration relation in a multivariate setting. To estimate the elements of $\boldsymbol{D}_u$, we apply the MLE approach of Beran (1995) to each series individually and apply the testing procedure of Robinson and Yajima (2002) to check for the equality of the memory parameters. The error term is modeled as $\boldsymbol{u}_t \sim N(0, \boldsymbol{\Sigma}_j)$. All results are based on a 5% nominal significance level, $\epsilon = 0.05$, meaning that the break fraction is in the interval $[0.05, 0.95]$, $T = 500$ and $1,000$ Monte Carlo replications.

We choose $m \in \{0, 1, 2, 3\}$ breaks which are uniformly allocated to the two series such that the distance between the breaks across both series is the same. Furthermore, breaks are not constrained to occur simultaneously in the two time series. The break size is set as,

$$\Delta\beta_{ij} = \kappa, \tag{11}$$

where κ is a finite constant.

6.1 Bias and MSE of the break date estimator

Tables 1 and 2 present the bias and mean squared error (MSE) of the break date estimator for the stochastic regressor case, assuming that the total number of breaks is known. We observe that as the break size (κ) increases, both bias and MSE generally decrease, meaning that larger breaks are easier to detect. However, this does not always hold, particularly in cases with high persistence in the regressors and errors (d values closer to 1) or when the errors are correlated ($\rho = 0.5$). This introduces estimation challenges, leading to higher bias and MSE. Additionally, larger sample sizes improve estimation accuracy, as seen in the consistently lower bias and MSE values when comparing results for $\kappa = 4$ and $\kappa = 8$.

Moreover, we can also see that estimating multiple breaks is more difficult than detecting a single break, especially when persistence is high. When errors are correlated, the estimator's accuracy decreases, leading to less precise break date identification. In some cases, bias is negative (indicating breaks are estimated too early), while in others, it is positive (suggesting breaks are identified too late). This variation shows how persistence and correlation affect estimation performance.

			Number of Breaks					
			1		2		3	
d_{1x} d_{1u}	d_{2x} d_{2u}	ρ / κ	4	8	4	8	4	8
0.6	0.4	0.0	0.073	-0.052	-0.097	0.010	0.132	0.005
0.4	0.2	0.5	0.034	0.011	-0.102	-0.014	0.146	0.005
0.6	0.6	0.0	0.215	0.067	0.857	-0.189	0.170	-0.078
0.2	0.2	0.5	-0.741	0.095	0.391	-0.041	0.550	0.099
0.6	0.6	0.0	0.050	-0.017	-0.101	-0.029	0.022	0.002
0.4	0.2	0.5	-0.112	0.039	0.083	0.027	-0.060	-0.024
0.6	0.8	0.0	0.026	-0.019	0.026	0.009	0.082	-0.005
0.4	0.2	0.5	-0.002	0.028	-0.085	-0.005	0.028	0.013

Table 1: Bias of the break date estimator.

			Number of Breaks					
			1		2		3	
d_{1x} d_{1u}	d_{2x} d_{2u}	ρ / κ	4	8	4	8	4	8
0.6	0.4	0.0	0.0053	0.0027	0.0106	0.0001	0.0207	0.0004
0.4	0.2	0.5	0.0340	0.0001	0.0106	0.0051	0.0254	0.0002
0.6	0.6	0.0	0.0462	0.0045	0.7343	0.0403	0.0506	0.0193
0.2	0.2	0.5	0.5491	0.0090	0.1836	0.0021	0.3628	0.0299
0.6	0.6	0.0	0.0025	0.0003	0.0102	0.0023	0.0005	0.0003
0.4	0.2	0.5	0.0125	0.0015	0.0167	0.0015	0.0114	0.0007
0.6	0.8	0.0	0.0007	0.0004	0.0015	0.0002	0.0071	0.0001
0.4	0.2	0.5	0.0000	0.0009	0.0105	0.0001	0.0009	0.0009

Table 2: MSE of the break date estimator.

6.2 The hit rate of the REBIT approach

In order to evaluate the hit rate of the novel REBIT approach introduced in this paper, the necessary asymptotic critical values are simulated for different combinations of $\boldsymbol{D} = (d_1, d_2)$ by approximating the stochastic integrals by partial sums. They are based on 10,000 Monte Carlo replications with $1,000$ increments per path of the fractional Brownian motion.

Figure 1 reports the hit ratio, i.e., how often our procedure detects the true number of breaks when $\boldsymbol{\beta}$ is breaking, depending on κ in a sample of size $T = 500$. This Figure consists of six panels, illustrating the hit rate of our test for determining break dates under different parameter configurations. κ corresponds to the magnitude of the break, and m represents the number of breaks considered in the model.

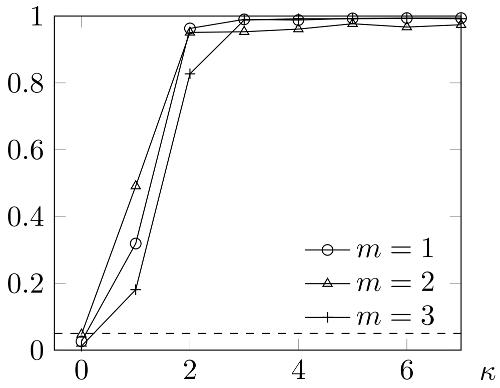
In general, across all panels, the hit rate increases with κ, reflecting that larger break magnitudes are easier to detect. We observe that for the majority of cases when $\kappa = 0$, i.e. in the case of no breaks ($m = 0$) in the series, we obtain a hit rate smaller or equal to 5% (our nominal significance level). Although, the rejection frequencies under the null hypothesis suggest that our test is conservative, we also observe that for large values of κ, the hit rate approaches $1 - \alpha = 0.95$, suggesting a highly effective test.

We can also observe that the total number of breaks, m, plays a crucial role in the test's performance. When fewer breaks ($m = 1$) are present in the series the test generally achieves higher hit rates across all scenarios, as they are simpler and involve fewer parameters to estimate. As m increases, the complexity of the model grows, making it harder to detect breaks precisely, particularly for small values of κ.

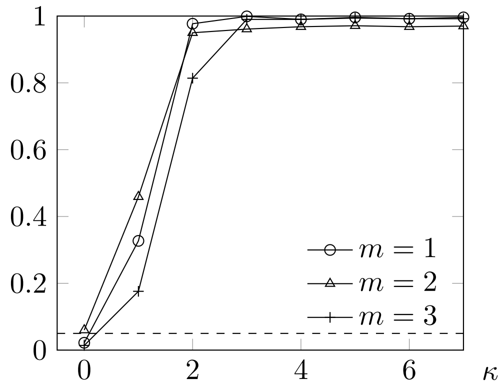
Panels (a) and (b): These panels correspond to the configuration $(d_{1x}, d_{2x}) = (0.6, 0.6)$, $(d_{1u}, d_{2u}) = (0.2, 0.2)$, and $\rho = 0$ and $\rho = 5$, respectively, where ρ represents the correlation between errors. In Panel (a) the test achieves high hit rates as κ increases, with hit rates close to 0.95 for $\kappa \geq 3$. The performance gap between $m = 1$, $m = 2$, and $m = 3$ is more noticeable for small κ, indicating higher sensitivity to simple models when break magnitudes are small. In Panel (b) the hit rate is slightly reduced compared to (a), especially for small κ, due to the difficulty of isolating the break effect in the presence of correlated errors. However, as κ grows, the hit rate converges to 0.95, and the difference between the $m = 1$, $m = 2$, and $m = 3$ cases diminishes.

Panels (c) and (d): Here, the persistence parameters are $(d_{1x}, d_{2x}) = (0.6, 0.4)$ and $(d_{1u}, d_{2u}) = (0.4, 0.2)$, with $\rho = 0$ and $\rho = 0.5$, respectively. In these panels heterogeneity in persistence for both the regressors and the errors is introduced. In Panel (c) the hit rate is slightly lower for small κ compared to Panel (a), reflecting the increased complexity of the test due to the heterogeneous persistence. Nevertheless, for $\kappa \geq 3$, the test performs well, with hit rate close to 0.95. Panel (d) represents the most challenging scenario, with the slowest increase in the hit rate as κ grows. The hit rate is noticeably lower for $m = 2$ and $m = 3$ when κ is small, and the convergence to 0.95 is more gradual compared to earlier panels.

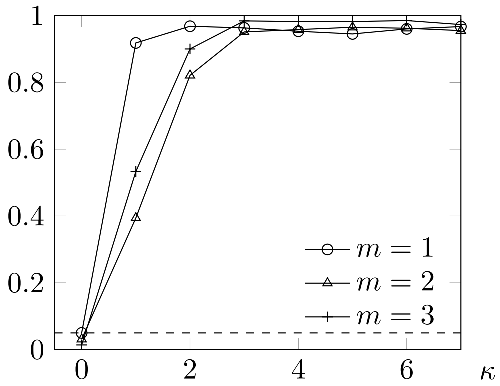
Panels (e) and (f): Finally, in Panels (e) and (f) we consider the configuration, $(d_{1x}, d_{2x}) = (0.6, 0.8)$, $(d_{1u}, d_{2u}) = (0.4, 0.2)$, and $\rho = 0$ and $\rho = 0.5$, respectively. In the case of uncorrelated errors ($\rho = 0$) the higher persistence of the second regressor ($d_{2x} = 0.8$) makes the break easier to detect for large κ, and the hit rate increases rapidly. For small κ, the hit rate is slightly lower for $m = 2$ and $m = 3$ compared to $m = 1$, but this gap diminishes as κ increases. However, the combination of high persistence and correlated errors leads to reduced hit rates for small κ. Nevertheless, the hit rate improves significantly as κ grows, with convergence to 0.95 occurring for all m values.



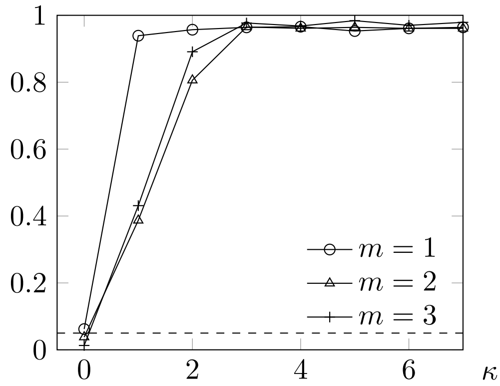
(a) $(d_{1x}, d_{2x}) = (0.6, 0.6)$, $(d_{1u}, d_{2u}) = (0.2, 0.2)$, and $\rho = 0$



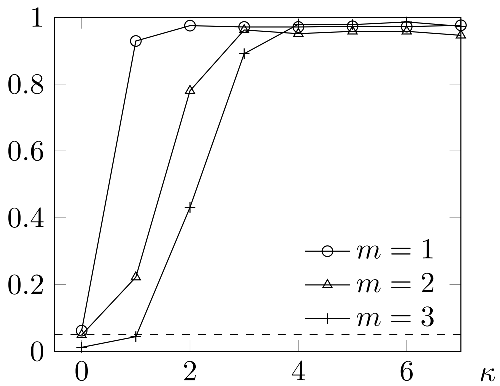
(b) $(d_{1x}, d_{2x}) = (0.6, 0.6)$, $(d_{1u}, d_{2u}) = (0.2, 0.2)$, and $\rho = 0.5$



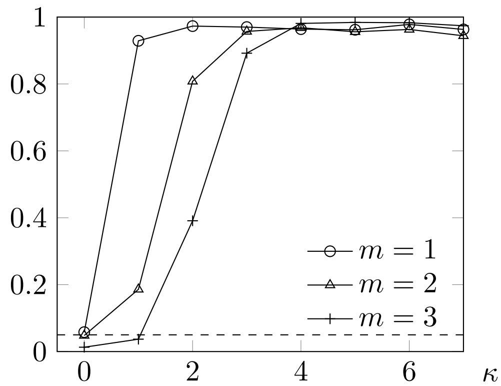
(c) $(d_{1x}, d_{2x}) = (0.6, 0.4)$, $(d_{1u}, d_{2u}) = (0.4, 0.2)$, and $\rho = 0$



(d) $(d_{1x}, d_{2x}) = (0.6, 0.4)$, $(d_{1u}, d_{2u}) = (0.2, 0.4)$, and $\rho = 0.5$



(e) $(d_{1x}, d_{2x}) = (0.6, 0.8)$, $(d_{1u}, d_{2u}) = (0.4, 0.2)$, and $\rho = 0$



(f) $(d_{1x}, d_{2x}) = (0.6, 0.8)$, $(d_{1u}, d_{2u}) = (0.4, 0.2)$, and $\rho = 0.5$

Figure 1: Hit ratio of our REBIT procedure for different orders of integration, d_{1x} and d_{2x}, for the stochastic regressors and d_{1u} and d_{2u} for the error terms, where the true number of breaks is m and the breaks are occurring in the cointegration relation. 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. the fraction the true number of breaks is detected. The memory parameter is estimated based on the approach of Beran (1995), ρ is the correlation of the series, and the sample size is $T = 500$. The dashed line indicates the 5% nominal significance level.

7 Empirical Applications

7.1 Inflation rates

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 autocorrelation at large lags and a pole in the periodogram at Fourier frequencies local to zero (e.g. Hassler and Wolters (1995) or Kumar and Okimoto (2007)). This can be seen as an indication that inflation rates follow a long memory process, but similar time series features can also be generated by short memory processes that are contaminated with breaks, which in the literature is referred to as spurious long memory (see, e.g., Diebold and Inoue (2001), Granger and Hyung (2004), and Mikosch and Stărică (2004)).

Standard estimation procedures of the long memory parameter are biased upwards in the presence of breaks, and 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. For instance, Hassler and Wolters (1995) and Baum et al. (1999) argue that ARFIMA models can describe inflation rates well; Bos et al. (1999) and Morana (2002) find evidence of structural breaks in international inflation rates; and Gadea et al. (2004) show that the memory of the series is reduced when structural changes are allowed for.

Many recent contributions favor a mixture of long memory models and structural breaks (Kumar and Okimoto (2007)). However, being able to determine 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, since if inflation rates are persistent, monetary policy actions need more time to unfold their effect, which can be more costly. The testing procedure we propose in this paper is therefore of importance as it allows to detect the true number of breaks in multivariate time series that are allowed to possess long memory dynamics. The approach can therefore be used to examine the properties of the underlying process of inflation rates.

Considering monthly CPI data (P_t), from January 1970 to May 2019, from Germany and

France, available from the OECD[3] we compute monthly 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 by applying X-13ARIMA-SEATS of Sax and Eddelbuettel (2018) to remove a yearly seasonality present in the german inflation series and a half-yearly seasonality in the french series. Figure 2 illustrates the bivariate time series along with the detected break points and partitions. We find that the nature of the break points resulting into spurious long memory are of the level shift kind.

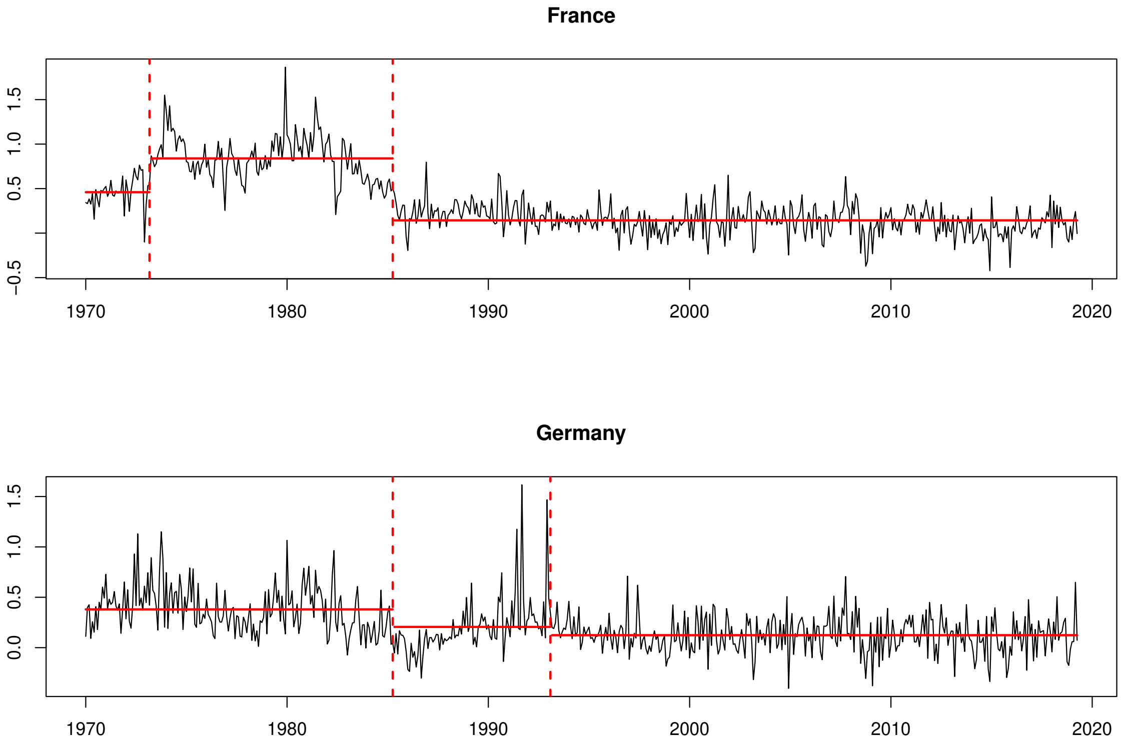


Figure 2: Monthly inflation rates for France and Germany from 1970 to 2019. The dashed red vertical lines refer to the mean shifts detected by our procedure. The solid bold red horizontal lines refer to the estimated means in each partition.

Table 3 provides further results of our testing procedure, as well as, the estimates of the memory parameter of the raw and the breaks corrected inflation series.

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

Panel A: Persistence of Inflation Series

	raw series			breaks corrected series	
	d_{GSE}	MLWS	d_{tGPH}	d_{GSE}	MLWS
France	0.567	2.526***	0.242	0.275	0.7813
Germany	0.381		0.214	0.283	

Panel B: Number and Dates of Breaks

	# breaks	break dates
REBIT	3	03/73, 04/85, 02/93

Table 3: 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 with the multivariate local Whittle estimator (GSE) of Shimotsu (2007) with a bandwidth of $b = \lfloor T^{2/3} \rfloor$, and the trimmed log-periodogram estimator (tGPH) of McCloskey and Perron (2013), which is robust against shifts, with $b = \lfloor T^{0.8} \rfloor$ and the constant that determines the trimming set at $\epsilon = 0.05$. Furthermore, results of the multivariate test against spurious long memory (MLWS) by Sibbertsen et al. (2018) are given with $b = \lfloor T^{2/3} \rfloor$ and trimming parameter $\epsilon = 0.02$. Here, ***, **, * denote significance at 1%, 5% and 10%, respectively. The GSE estimates of the memory on the right hand side of the table as well as the result of the MLWS test are given for the break corrected time series. The break correction 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.

The left hand side of Panel A of Table 3 shows results regarding the persistence of the inflation series. In line with earlier empirical results (e.g., Hassler and Wolters (1995) and Bos et al. (1999)) the multivariate local Whittle estimator (GSE) of 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 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 et al. (2018) rejects the null hypothesis of pure long memory processes at the 1% significance level. Second, applying the (univariate) trimmed log-periodogram estimator (tGPH) of McCloskey and Perron (2013) on both inflation series shows that the memory of these series decreases. Therefore, we apply our procedure that can consistently detect and estimate multiple shifts in the bivariate system of inflation series (see results in Panel B of Table 3).

We observe that the first structural break detected is in the french time series in March 1973, which is a few month before the first oil crisis. We further note that the mean in the second partition of the inflation series of France shows a significant increase. The second structural break is detected in both time series simultaneously in April 1985 which may be connected to the 1980s oil glut. We observe that the mean of both inflation series strongly decreases after the break in this partition of the series. It should be noted that the break point estimator is not constrained to estimate only simultaneous occurring breaks. The last break detected occurs in the german inflation rate series in February 1993. The break results in another small reduction of the mean of the process. This break is likely caused by a severe recession that occurred in Germany during this time.

The other two procedures we also 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, and 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 break corrected inflation series. The results of the GSE estimator and MLWS test can be seen on the right hand side of Panel A of Table 3. The estimated memory by the GSE strongly decreases to a value around 0.28 which is similar to the tGPH estimate of the raw series. Furthermore, the MLWS test is no longer significant suggesting that there is no evidence of spurious long memory. Therefore, we conclude that our procedure detects all relevant breaks of the bivariate inflation system.

7.2 EMU Government Bond Markets

The second empirical analysis is based on daily observations of 10-year-to-maturity government bonds of eight EMU countries from 01.01.2006 to 06.06.2024, including a total of 4809 observations per country. The analysed countries are Portugal, Ireland, Belgium, Austria, Finland, Netherlands, France, and Germany. The data is obtained from Datastream.

In general, since the introduction of the euro the market is seen as an integrated market (see, among others, Hartmann et al. (2003) and Abad et al. (2010)). This is reasoned by a reduction of the trading barriers, for example, by eliminating exchange-rate risk and increasing homogeneity in the market itself. All in all, the reduction of trading barriers and transaction costs, and an increase in liquidity leads to a market which is said to co-move in an equilibrium state, so that a fully integrated market is not affected by country specific risk anymore.

One way of modelling co-movement or market integration can be done in terms of (fractional) cointegration, as introduced by Engle and Granger (1987) and Johansen (1988), in order to express the long-term equilibrium relationship given in this market. However, the standard (fractional) cointegration paradigm assumes a constant relationship between the non-stationary time series. This might be a rather restrictive assumption especially in the case of the EMU government bonds. As seen in Figure 3 the time series are moving relatively close to each other which may be seen as evidence of markets convergence. However, this behaviour changes as some of the yields start to drift apart from the beginning of the financial crisis in 2008 and remain diverged as a consequence of the European sovereign debt crisis starting shortly after the financial crisis in 2009. This raises doubts about the integration of the European bond market. The discussion about integration or disintegration of the European bond market led to a new strand of literature addressing time variation in the integration of the bond market, as it is broadly accepted in the literature that the introduction of the euro led to an almost fully integrated bond market. However, the crises introduced some kind of structural change or temporary shock in the market (see e.g. Christiansen (2014), Babecký et al. (2017), Sehgal et al. (2017), Qin et al. (2023), and Rodrigues et al. (2024)).

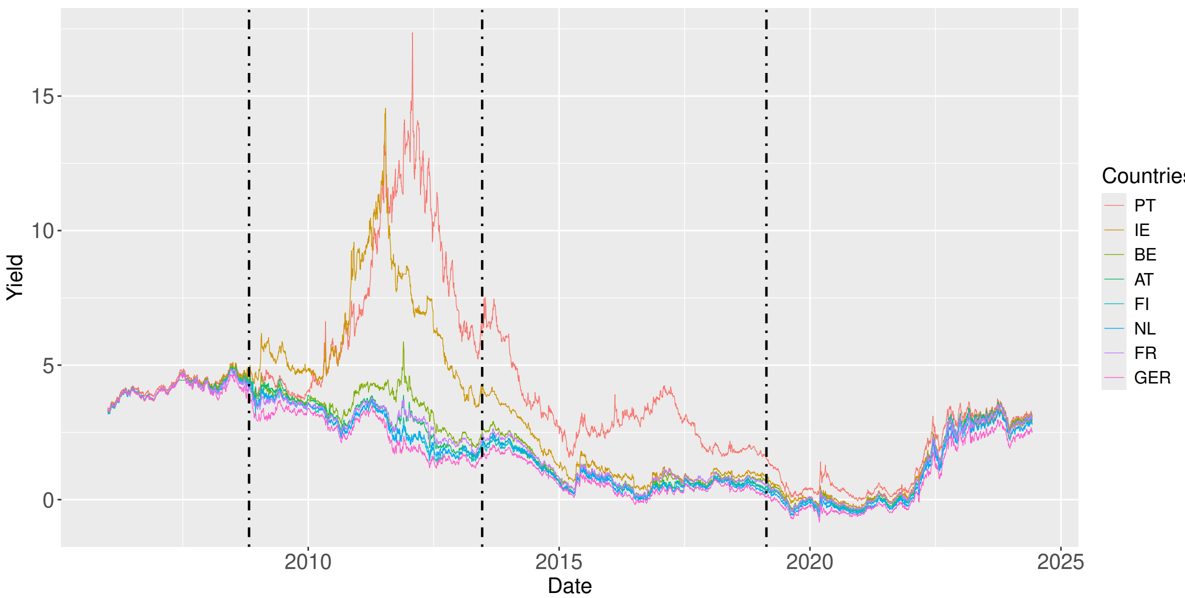


Figure 3: Yields of the EMU government bonds from 2006 to 2024. The black vertical lines refer to the estimated break points our procedure detected.

Based on recent literature we suggest to test for possible breaks in the (fractional) cointegration relationship in the EMU bond market. Therefore, we first examine whether there is a fractional cointgeration relationship in the system. To do so we investigate, by using the ADF test (with constant and trend) and the KPSS test the order of integration of the single series. The ADF test does not reject the null of a unit root and the KPSS test rejects stationarity for all cases. In order to proceed with the analysis of possible (fractional) cointegration we apply different tests. First we test for equality of the estimated memory parameters in the system using the test by Robinson and Yajima (2002) and Nielsen and Shimotsu (2007). All results are presented in Panel A of Table 4. We are not able to reject the null of equal memory parameters.

Next, we test for fractional cointegration for the time series all together as well as for the single time series in a bivariate fashion to ensure overall fractional cointgeration. We apply the tests of Chen and Hurvich (2006), Robinson (2008), and Nielsen (2010) where the null of no fractional cointegration can be rejected for all the testing procedures considered.
This result seems to be in line with a general long-run equilibrium and in favor of the integration of the EMU bond markets. However, based on recent findings and assumptions regarding the time-varying behaviour of market integration in Europe, this framework may not be flexibel enough. Therefore, we apply our REBIT procedure to check whether recent crises might have caused some changes in the long-run equilibrium relationship.

The number of breaks and the respective break dates can be found in Panel B of Table 4. In general, our procedure is able to detect three breaks in total in the system. The system is set up in such a way that we investigate fractional cointegration between Germany and the other EMU countries as our model set-up requires an explanatory variable. The first two break dates are close to breaks reported in the literature. Babecký et al. (2017) find periods of interruption of the financial integration in these segments and Rodrigues et al. (2024) confirms this by finding segments where the bond market is not fractionally cointegrated before and after the sovereign debt crisis.

We contribute to these findings in that we can estimate the long-run coefficients for the single segments. Panel C of Table 4 includes the estimates of the long-run relationship between Germany and the single time series. In the first segment we receive for all the countries a $\boldsymbol{\beta}$ estimate which is close to one. That indicates and supports the assumption of a very close co-moving or fully integrated market, where the long-run coefficient is indicating an almost one to one relationship. Applying the aforementioned fractional cointegration tests we can reject the null of no fractional cointegration for all cases. The second segment and the first estimated break point coincides with the financial crisis and covers the period over the sovereign debt crisis as well. The $\boldsymbol{\beta}$ estimates over this segment show a clear change in the long-run coefficients compared to the first segment and show a huge change in the long-run coefficient for Portugal and Ireland during the eurozone crisis. Although, not all countries where severely influenced by the debt crisis the change of the estimates of the long-run coefficients show how the countries were all together influenced by the crisis. This supports the idea of a fully integrated market as the countries are affected as a whole. Interestingly, when applying the testing procedures to detect whether the crises just changed the long-run coefficient or lead to an interruption of the franctional cointegration relationship we are again able to reject the null of no fractional cointegration for all the applied testing procedures. The second break coincides with the end of the European sovereign debt crisis where Portugal and Ireland improved their financial stability and managed to exit their bailout programs in mid 2014, e.g. in the third segment our procedure detected.

The analyzed time series in Babecký et al. (2017) and Rodrigues et al. (2024) do not exceed 2017 while we are able to find another break in the long-run equilibrium after this. The break is at the beginning of 2019 and could coincide with the ongoing debate about the Brexit and the Irish backstop. In February the president of the European commission claimed to support Ireland in case of a no-deal Brexit. The respective $\boldsymbol{\beta}$ estimates for the last segment show and support the overall picture we see in Figure 3 that the market starts to co-move again although not as close as in the first segment where the long-run coefficients where close to one for all investigated time series. Here, the majority of estimates are around 1.2. All fractional cointegration tests reject the null of no fractional cointegration in this period.

In general, the EMU bond market seems to be integrated in the long-run, but this long-run relationship is dynamic and changing in time.

Panel A: Persistence of the Bond Series

	raw series			
	ADF	KPSS	$\hat{d}$	$\mathcal{T}_0$
PT	0.76	0.01	0.96	
IE	0.79	0.01	0.99	
BE	0.98	0.01	0.96	
AT	0.99	0.01	0.99	0.40
FI	0.99	0.01	1.01	
NL	0.99	0.01	0.99	
FR	0.99	0.01	0.98	
GER	0.99	0.01	0.98	

Panel B: Number and Dates of Breaks

	# breaks	break dates
REBIT	3	27.10.08, 20.06.13, 18.02.19

Panel C: Long-run Coefficient Estimates

	2006 - 2008	2008 - 2013	2013 - 2019	2019
$\hat{\beta}_{PT}$	1.0612	2.5053	3.5959	1.2800
$\hat{\beta}_{IE}$	1.0380	2.3719	2.0229	1.1536
$\hat{\beta}_{BE}$	1.0383	1.3549	1.4560	1.2350
$\hat{\beta}_{AT}$	1.0282	1.2230	1.2505	1.2305
$\hat{\beta}_{FI}$	1.0230	1.1247	1.1827	1.2094
$\hat{\beta}_{NL}$	1.0210	1.1276	1.2057	1.1237
$\hat{\beta}_{FR}$	1.0223	1.1888	1.3939	1.1965

Table 4: Panel A includes results regarding the persistence of the time series. The order of integration, d, of the time series is estimated by the univariate local Whittle estimator of Robinson (1995) using a bandwith of $b = \lfloor T^{2/3} \rfloor$ applied to the each time series individually. This is necessary in order to prevent inconsistency of the multivariate local Whittle estimator proposed by Shimotsu (2007) in the case of (fractional) cointegration. The remaining columns include the p-values of the ADF-, KPSS-test and a testing procedure proposed by Robinson and Yajima (2002) and Nielsen and Shimotsu (2007) to test the equality of the memory parameters in the single time series, $\mathcal{T}_0$. Panel B includes the number of breaks and the corresponding break dates detected by our procedure. Panel C includes the estimates of the long-run coefficient for the estimated segments.

8 Conclusions

This paper introduces, to the best of our knowledge, the first procedure for testing for multiple breaks in a multivariate long memory framework covering the case of fractional cointegration. We embed our procedure into a multivariate system of long memory time series allowing for breaks in the regression parameters 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.

The procedure consists of iterative testing for m structural breaks with m increasing in each step. It therefore avoids splitting the sample into segments as in Bai (1997b) and others, which is not possible under long memory. Our test and break point estimator is Likelihood-ratio based. The consistency and limiting distributions of both procedures are derived. Interestingly, the limiting distribution of the test depends only on the maximum of all memory parameters and on the number of series having this maximum memory.

A Monte Carlo study demonstrates the good finite sample properties of our procedure and an application to inflation rates of France and Germany and EMU government bonds its usefulness in practice.

9 Acknowledgment

Less and Sibbertsen gratefully acknowledge the financial support of the Deutsche Forschungsgemeinschaft (DFG), and Rodrigues thanks the Portuguese Science Foundation (FCT) for financial support through projects 2023.15199.PEX (https://doi.org/10.54499/2023.15199.PEX) and UIDB/00124/2020, UIDP/00124/2020 and Social Sciences DataLab - PINFRA/22209/2016, POR Lisboa and POR Norte (Social Sciences DataLab,PINFRA/22209/2016). We are also grateful to Jan Beran, Liudas Giraitis, Harry Haupt, Christian Leschinski, Michael Massmann, Morten Nielsen, Kai Wenger, Simon Wingert as well as the participants of the Aarhus Long Memory Workshop 2024 for useful comments and suggestions.

References

Abad, Pilar, Helena Chuliá, and Marta Gómez-Puig (2010). “EMU and European government bond market integration”. *Journal of Banking & Finance* 34(12), pp. 2851–2860.

Babecký, Jan, Luboš Komárek, and Zlatuše Komárková (2017). "Financial integration at times of crisis and recovery". *Economic Imbalances and Institutional Changes to the Euro and the European Union*. Vol. 18. Emerald Publishing Limited, pp. 173–191.

Bai, Jushan (1997a). "Estimating multiple breaks one at a time". *Econometric Theory* 13(3), pp. 315–352.

Bai, Jushan (1997b). "Estimation of a change point in multiple regression models". *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". *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". *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". *Econometrica*, pp. 47–78.

Baum, Christopher F, John T Barkoulas, and Mustafa Caglayan (1999). "Persistence in international inflation rates". *Southern Economic Journal*, pp. 900–913.

Beran, Jan (1995). "Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models". *Journal of the Royal Statistical Society: Series B (Methodological)* 57(4), pp. 659–672.

Betken, Annika (2016). "Testing for Change-Points in Long-Range Dependent Time Series by Means of a Self-Normalized Wilcoxon Test". *Journal of Time Series Analysis* 37(6), pp. 785–809.

Bos, Charles S, Philip Hans Franses, and Marius Ooms (1999). "Long memory and level shifts: Re-analyzing inflation rates". *Empirical Economics* 24(3), pp. 427–449.

Casini, Alessandro and Pierre Perron (2019). "Structural Breaks in Time Series". *Oxford Research Encyclopedia of Economics and Finance*. DOI: 10.1093/acrefore/9780190625979.013.179.

Chen, Willa W and Clifford M Hurvich (2006). "Semiparametric estimation of fractional cointegrating subspaces". *The Annals of Statistics*, pp. 2939–2979.

Christiansen, Charlotte (2014). "Integration of European bond markets". *Journal of Banking & Finance* 42, pp. 191–198.

Chung, Ching-Fan (2002). "Sample means, sample autocovariances, and linear regression of stationary multivariate long memory processes". *Econometric Theory* 18(1), pp. 51–78.

Davidson, James and Robert M. de Jong (2000). "The functional central limit theorem and weak convergence to stochastic integrals II: Fractionally Integrated Processes". *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". *Scandinavian Journal of Statistics* 40(1), pp. 153–173.

Diebold, Francis X and Atsushi Inoue (2001). "Long memory and regime switching". *Journal of Econometrics* 105(1), pp. 131–159.

Engle, Robert F and Clive WJ Granger (1987). "Co-integration and error correction: representation, estimation, and testing". *Econometrica*, pp. 251–276.

Gadea, Marıa Dolores, Marcela Sabaté, and José Marıa Serrano (2004). "Structural breaks and their trace in the memory: Inflation rate series in the long-run". *Journal of International Financial Markets, Institutions and Money* 14(2), pp. 117–134.

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". *Journal of Empirical Finance* 11(3), pp. 399–421.

Hall, Alastair R., Denise R. Osborn, and Nikolaos Sakkas (2013). "Inference on Structural Breaks using Information Criteria". *The Manchester School* 81(S3), pp. 54–81.

Hartmann, Philipp, Angela Maddaloni, and Simone Manganelli (2003). "The Euro-area financial system: Structure, integration, and policy initiatives". *Oxford Review of Economic Policy* 19(1), pp. 180–213.

Hassler, Uwe and Barbara Meller (2014). "Detecting multiple breaks in long memory the case of U.S. inflation". *Empirical Economics* 46(2), pp. 653–680.

Hassler, Uwe, Paulo M.M. Rodrigues, and Antonio Rubia (2014). "Persistence in the banking industry: Fractional integration and breaks in memory". *Journal of Empirical Finance* 29, pp. 95–112.

Hassler, Uwe and Jürgen Wolters (1995). "Long memory in inflation rates: International evidence". *Journal of Business & Economic Statistics* 13(1), pp. 37–45.

Iacone, Fabrizio, Stephen J Leybourne, and Robert AM Taylor (2014). “A Fixed-B Test for a Break in Level at an Unknown Time Under Fractional Integration”. *Journal of Time Series Analysis* 35(1), pp. 40–54.

Johansen, Søren (1988). “Statistical analysis of cointegration vectors”. *Journal of Economic Dynamics and Control* 12(2-3), pp. 231–254.

Kechagias, Stefanos and Vladas Pipiras (2015). “Definitions and representations of multivariate long-range dependent time series”. *Journal of Time Series Analysis* 36(1), pp. 1–25.

Kumar, Manmohan S and Tatsuyoshi Okimoto (2007). “Dynamics of persistence in international inflation rates”. *Journal of Money, Credit and Banking* 39(6), pp. 1457–1479.

Lavielle, Marc and Eric Moulines (2000). “Least-squares Estimation of an Unknown Number of Shifts in a Time Series”. *Journal of Time Series Analysis* 21(1), pp. 33–59.

Marinucci, D. and P.M. Robinson (1999). “Alternative forms of fractional Brownian motion”. *Journal of Statistical Planning and Inference* 80(1-2), pp. 111–122.

Marinucci, D. and P.M. Robinson (2000). “Weak convergence of multivariate fractional processes”. *Stochastic Processes and Their Applications* 86(1), pp. 103–120.

Martins, Luis F. and Paulo M.M. Rodrigues (2014). “Testing for persistence change in fractionally integrated models: An application to world inflation rates”. *Computational Statistics & Data Analysis* 76. CFEnetwork: The Annals of Computational and Financial Econometrics, pp. 502–522.

McCloskey, Adam and Pierre Perron (2013). “Memory parameter estimation in the presence of level shifts and deterministic trends”. *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”. *Review of Economics and Statistics* 86(1), pp. 378–390.

Morana, Claudio (2002). “Common persistent factors in inflation and excess nominal money growth and a new measure of core inflation”. *Studies in Nonlinear Dynamics & Econometrics* 6(3).

Nielsen, Morten Ørregaard (2010). “Nonparametric cointegration analysis of fractional systems with unknown integration orders”. *Journal of Econometrics* 155(2), pp. 170–187.

Nielsen, Morten Ørregaard and Katsumi Shimotsu (2007). “Determining the cointegrating rank in nonstationary fractional systems by the exact local Whittle approach”. *Journal of Econometrics* 141(2), pp. 574–596.

Perron, Pierre (2006). "Dealing with structural breaks". *Palgrave Handbook of Econometrics* 1(2), pp. 278–352.

Preuss, Philip, Ruprecht Puchstein, and Holger Dette (2015). "Detection of Multiple Structural Breaks in Multivariate Time Series". *Journal of the American Statistical Association* 110, pp. 654–668.

Qin, Weiping, Sungjun Cho, and Stuart Hyde (2023). "Time-varying bond market integration and the impact of financial crises". *International Review of Financial Analysis* 90, p. 102909.

Qu, Zhongjun and Pierre Perron (2007). "Estimating and testing structural changes in multivariate regressions". *Econometrica* 75(2), pp. 459–502.

Robinson, Peter (1995). "Gaussian semiparametric estimation of long range dependence". *The Annals of statistics*, pp. 1630–1661.

Robinson, Peter M. (2008). "Diagnostic testing for cointegration". *Journal of econometrics* 143(1), pp. 206–225.

Robinson, Peter Michael and Yoshihiro Yajima (2002). "Determination of cointegrating rank in fractional systems". *Journal of Econometrics* 106(2), pp. 217–241.

Rodrigues, Paulo MM, Philipp Sibbertsen, and Michelle Voges (2024). "The stability of government bond markets' equilibrium and the interdependence of lending rates". *Empirical Economics*, pp. 1–36.

Sax, Christoph and Dirk Eddelbuettel (2018). "Seasonal adjustment by x-13arima-seats in r". *Journal of Statistical Software* 87, pp. 1–17.

Sehgal, Sanjay, Priyanshi Gupta, and Florent Deisting (2017). "Assessing time-varying stock market integration in Economic and Monetary Union for normal and crisis periods". *The European Journal of Finance* 23(11), pp. 1025–1058.

Shao, Xiaofeng (2011). "A simple test of changes in mean in the possible presence of long-range dependence". *Journal of Time Series Analysis* 32(6), pp. 598–606.

Shimotsu, Katsumi (2007). "Gaussian semiparametric estimation of multivariate fractionally integrated processes". *Journal of Econometrics* 137(2), pp. 277–310.

Sibbertsen, Philipp (2004). "Long memory versus structural breaks: An overview". *Statistical Papers* 45(4), pp. 465–515.

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

Sibbertsen, Philipp, Christian Leschinski, and Marie Busch (2018). “A multivariate test against spurious long memory”. *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”. *Statistical Papers* 53, pp. 357–370.

Wang, Lihong (2008). “Change-in-mean problem for long memory time series models with applications”. *Journal of Statistical Computation and Simulation* 78(7), pp. 653–668.

Wenger, Kai and Christian Leschinski (2019). “Fixed-bandwidth CUSUM tests under long memory”. *Econometrics and Statistics.*

Wenger, Kai, Christian Leschinski, and Philipp Sibbertsen (2019). “Change-in-mean tests in long-memory time series: a review of recent developments”. *AStA Advances in Statistical Analysis* 103(2), pp. 237–256.

Wenger, Kai and Vivien Less (2020). “A modified Wilcoxon test for change points in long-range dependent time series”. *Economics Letters* 192, p. 109237.

Wingert, Simon, Mwasi Mboya, and Philipp Sibbertsen (2022). “Distinguishing between Breaks in the Mean and Breaks in Persistence under Long Memory”. *Economics Letters* 193, p. 109338.

Wu, Wei Biao (2007). “Strong invariance principles for dependent random variables”. *The Annals of Probability* 35(6), pp. 2294–2320.

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

Yao, Yi-Ching (1988). “Estimating the number of change-points via Schwarz’ criterion”. *Statistics and Probability Letters* 6, pp. 181–189.

On-Line Supplementary Appendix

to

"Testing for Multiple Structural Breaks in Multivariate Long Memory Time Series"

by

Vivien Less, Paulo M. M. Rodrigues and Philipp Sibbertsen

A Technical Proofs

This section contains the proofs of Lemma 1 and Theorems 1 and 2. To prove these results we require 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 A.1. Afterwards, in Section A.2, we show that under our set of Assumptions, the ten properties of the quasi-likelihood considered in Bai et al. (1998), Bai (2000) and Qu and Perron (2007) are satisfied. We prove consistency of the break point estimators next in Lemma A.1 and in Section A.4 we provide the proof of the limiting distribution of our test statistic, i.e. Theorem 2.

In what follows, if not stated otherwise, $d = d_u$ is considered.

A.1 Proof Generalised Hájek-Rényi Inequality, SLLN, FCLT

Lemma A.1 (Generalised Hájek-Rényi Inequality)**.** *Let $\boldsymbol{\xi} = (\boldsymbol{\xi}_i)_{i\geq 1}$ be a sequence of mean zero $\mathbb{R}^d$-valued random vectors. Define $\mathcal{F}_k$ as an increasing σ-field generated by $(\boldsymbol{\xi}_i)_{i\geq k}$. Consider that $(\boldsymbol{\xi}_i)_{i\geq 1}$ satisfies Assumption 4 with $\boldsymbol{x}_i'\boldsymbol{u}_i$ replaced by $\boldsymbol{\xi}_i$. Then, there exists an $L < \infty$ such that, for every $\delta > 0$ and $m > 0$, $P(\sup_{k\geq m} k^{-1}\|\sum_{t=1}^{k}\boldsymbol{\xi}_t\| > \delta) \leq (L/\delta^2 m^{2d-1})$, where $d = \mathrm{d}_{\max}$ is the largest memory parameter of the elements of the vector $\boldsymbol{\xi}_i$.*

Proof. In the following we define $\boldsymbol{M}_{i:j} = \sum_{t=i}^{j}\boldsymbol{\xi}_t$. We start by noting that,

$$\Pr\left(\max_{k\geq m}\frac{1}{k}\|\boldsymbol{M}_{1:k}\| > \delta\right) \leq \sum_{p=0}^{\infty}\Pr\left(\max_{2^p\leq k\leq 2^{p+1}m}\frac{1}{k}\|\boldsymbol{M}_{1:k}\| > \delta\right), \tag{A.1}$$

and that,

$$\Pr\left(\max_{1\leq k\leq n}\frac{1}{k}\|\boldsymbol{M}_{1:k}\| > \delta\right) \leq 4\frac{\boldsymbol{A}(d)\boldsymbol{C}(\boldsymbol{\xi})}{\delta^2}n^{2d}\sum_{t=1}^{n}\left(\frac{1}{t}\right)^2. \tag{A.2}$$

Here $\boldsymbol{C}(\boldsymbol{\xi}) < \infty$ such that $\forall i,j$, $E(\|M_{i:j}\|^2) \leq \boldsymbol{C}(\boldsymbol{\xi})|j-i+1|^{2d+1}$. Suppose (A.2) holds.

Then,

$$
\begin{aligned}
\Pr\left(\max_{2^p m \le k \le 2^{p+1} m} \frac{1}{k}\|\boldsymbol{M}_{1:k}\| > \delta\right) &\le \Pr\left(\frac{1}{2^p m}\|\boldsymbol{M}_{1:m}\| > \frac{\delta}{2}\right) \\
&\quad + \Pr\left(\max_{2^p m+1 \le k \le 2^{p+1} m} \frac{1}{k}\|\boldsymbol{M}_{2^p m+1:2^{p+1} m}\| > \frac{\delta}{2}\right) \\
&\le 4\frac{\boldsymbol{A}(d)\boldsymbol{C}(\boldsymbol{\xi})}{\delta^2}(2^p m)^{2d-2} + 4\frac{\boldsymbol{A}(d)\boldsymbol{C}(\boldsymbol{\xi})}{\delta^2}(2^p m)^{2d} \sum_{t=2^p m+1}^{2^{p+1}} \left(\frac{1}{t}\right)^2 \\
&\le 8\frac{\boldsymbol{A}(d)\boldsymbol{C}(\boldsymbol{\xi})}{\delta^2}(2^p m)^{2d-1}.
\end{aligned}
$$

Moreover, using (A.1) we have that,

$$
\Pr\left(\max_{k \ge m} \frac{1}{k}\|\boldsymbol{M}_{1:k}\| > \delta\right) \;\le\; 8\frac{\boldsymbol{A}(d)\boldsymbol{C}(\boldsymbol{\xi})}{\delta^2}\sum_{p=0}^{\infty}(2^p m)^{2d-1} \;\le\; \frac{L}{\delta^2} m^{2d-1},
$$

where $L < \infty$ is a constant.

We prove (A.2) by the Markov inequality. To simplify notation we define $\boldsymbol{S}_{i:j} = \max_{k=i,\ldots,j} \frac{1}{k}\|\boldsymbol{M}_{1:k}\|$. We need to show that,

$$
E\left(\boldsymbol{S}_{1:n}^2\right) \;\le\; \boldsymbol{C}(\boldsymbol{\xi})\boldsymbol{A}(d) n^{2d} \sum_{t=1}^{n} \frac{1}{t^2}. \tag{A.3}
$$

If (A.3) holds, our auxiliary result in (A.2) is proven by the Markov inequality. The claim in (A.3) is proven by induction on n. For $n = 1$ the inequality is obvious for $\boldsymbol{A}(d) = 1$, from the result in Kechagias and Pipiras (2015) who showed that for the partial sums $\boldsymbol{M}_{i:j}$ there exists a $\boldsymbol{C}(\boldsymbol{\xi}) < \infty$ such that, for all i, j,

$$
E\left(\|\boldsymbol{M}_{i:j}\|^2\right) \;\le\; \boldsymbol{C}(\boldsymbol{\xi})\,|j - i + 1|^{2d+1}. \tag{A.4}
$$

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

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

Applying the Minkowski inequality to the above inequality yields,

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

where we used (A.4) 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 $\boldsymbol{A}(d)$ such that,

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

The induction step is proven and thus this concludes the proof of inequality (A.3). □

Denote $\boldsymbol{W}_D(t) = (W_{d_1}(t), \ldots, W_{d_n}(t))'$ an n-dimensional fractional Brownian motion with n different memory parameters $\boldsymbol{D} = (d_1, \ldots, d_n)'$ (cf. Marinucci and Robinson (2000), Davidson and Jong (2000), Chung (2002)). Each $W_{d_i}(t)$, $i = 1, ..., n$, is a one-dimensional fractional Brownian motion defined as,

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

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

Lemma A.2 (FCLT, SLLN)**.** *Let $(\boldsymbol{\xi}_i)_{i\geq 1}$ be a sequence of mean zero $\mathbb{R}^n$-valued random vectors that satisfy Assumption 4. Then,*

(a) (FCLT)

$$
\mathrm{diag}(T^{-1/2-d_1}, \ldots, T^{-1/2-d_n}) \sum_{t=1}^{[Tr]} \boldsymbol{\xi}_t \Rightarrow \boldsymbol{\Omega}\boldsymbol{W}_D(r), \tag{A.5}
$$

where $\boldsymbol{W}_D(r)$ is an $n \times 1$ vector of independent fractional Wiener processes and $\Rightarrow$ denotes weak convergence under the Skorokhod topology;

(b) (SLLN)

$$k^{-1}\sum_{i=1}^{k}\boldsymbol{\xi}_i \xrightarrow{a.s.} \mathbf{0}, \qquad as \qquad k\to\infty. \tag{A.6}$$

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. □

A.2 The Ten Properties of the Quasi-Likelihood Ratio

This section provides the properties of the quasi-likelihood ratio and parameter estimates, which will be relevant in the subsequent proofs of our results. Consider,

$$\mathcal{L}(1,k;\boldsymbol{\beta},\boldsymbol{\Sigma}) = \frac{\prod_{t=1}^{k} f(\boldsymbol{y}_t|\boldsymbol{x}_t,...,\boldsymbol{\beta},\boldsymbol{\Sigma})}{\prod_{t=1}^{k} f(\boldsymbol{y}_t|\boldsymbol{x}_t,...,\boldsymbol{\beta}_0,\boldsymbol{\Sigma}_0)},$$

where $\boldsymbol{\beta}_0$ and $\boldsymbol{\Sigma}_0$ describe the true coefficients and variance matrix, respectively. In the following, we denote the estimates obtained from maximizing $\mathcal{L}(1,k;\boldsymbol{\beta},\boldsymbol{\Sigma})$ as $\hat{\boldsymbol{\beta}}_{(k)}$ and $\hat{\boldsymbol{\Sigma}}_{(k)}$. Then the following properties hold:

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

$$\sup_{\lfloor T\delta\rfloor\le k\le T}\mathcal{L}(1,k;\hat{\boldsymbol{\beta}}_{(k)},\hat{\boldsymbol{\Sigma}}_{(k)}) = O_p(1),$$

$$\sup_{\lfloor T\delta\rfloor\le k\le T}(\|\hat{\boldsymbol{\beta}}_{(k)}-\boldsymbol{\beta}_{\mathbf{0}}\| + \|\hat{\boldsymbol{\Sigma}}_{(k)}-\boldsymbol{\Sigma}_0\|) = O_p(T^{d-1/2}).$$

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

$$\hat{\boldsymbol{\beta}}_{(k)}-\boldsymbol{\beta}_0 = \left(\sum_{t=1}^{k}\boldsymbol{x}_t\hat{\boldsymbol{\Sigma}}_{(k)}^{-1}\boldsymbol{x}_t'\right)^{-1}\sum_{t=1}^{k}\boldsymbol{x}_t\hat{\boldsymbol{\Sigma}}_{(k)}^{-1}\boldsymbol{u}_t,$$

and apply the generalized Hájek-Rényi inequality introduced in Lemma A.1 on $\sum_{t=1}^{k}\boldsymbol{x}_t(\boldsymbol{\Sigma}_0)^{-1}\boldsymbol{u}_t$. Together with the strong consistency of $\hat{\boldsymbol{\Sigma}}_{(k)}$ it follows that $\sup_{\lfloor T\delta\rfloor\le k\le T}\|\hat{\boldsymbol{\beta}}_{(k)}-\boldsymbol{\beta}_0\| = O_p(T^{-1/2+d})$.

Furthermore,

$$\hat{\boldsymbol{\Sigma}}_{(k)}-\boldsymbol{\Sigma}_0 = \frac{1}{k}\sum_{t=1}^{k}(\boldsymbol{u}_t-\boldsymbol{x}_t'(\hat{\boldsymbol{\beta}}_{(k)}-\boldsymbol{\beta}_{\mathbf{0}}))(\boldsymbol{u}_t-\boldsymbol{x}_t'(\hat{\boldsymbol{\beta}}_{(k)}-\boldsymbol{\beta}_0))'-\boldsymbol{\Sigma}_0.$$

Applying again the generalized Hájek-Rényi inequality provided in Lemma A.1 it follows that $\sup_{\lfloor T\delta\rfloor\leq k\leq T}\|\hat{\boldsymbol{\Sigma}}_{(k)}-\boldsymbol{\Sigma}_0\| = O_p(T^{-1/2+d})$, and as a direct consequence $\sup_{\lfloor T\delta\rfloor\leq k\leq T}\mathcal{L}(1,k;\hat{\boldsymbol{\beta}}_{(k)},\hat{\boldsymbol{\Sigma}}_{(k)}) = O_p(1)$. □

Property 2 next is modified compared to property 2 of Qu and Perron (2007). Instead of considering the supremum of the likelihood over $1\leq k\leq T$ we consider the supremum over $\lfloor \delta T\rfloor \leq k\leq T$, for some $\delta\in(0,1)$.

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

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

for all large T.

Proof. This result is a direct consequence of Property 1. □

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

$$\Pr\left(\sup_{k\geq\lfloor\delta T\rfloor}\ \sup_{(\boldsymbol{\beta},\boldsymbol{\Sigma})\in S_T} T^{D}\mathcal{L}(1,k;\boldsymbol{\beta},\boldsymbol{\Sigma})>1\right)<\epsilon. \tag{A.7}$$

Proof. To prove this result 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,

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

where $\lambda_{\min}$ and $\lambda_{\max}$ denote the smallest and largest eigenvalues of $\boldsymbol{\Sigma}$ and the finite constants d_1, d_2 and d_3 are chosen in such a way that $(\boldsymbol{\beta}_0,\boldsymbol{\Sigma}_0)$ is an inner point of $\overline{\boldsymbol{\Theta}}_2$. As noted above, we first show (A.7) with the second supremum taken over $\boldsymbol{S}_T\cap\overline{\boldsymbol{\Theta}}_2$ which is compact. We decompose the segmential log-likelihood as $\log\boldsymbol{\mathcal{L}}(1,k;\boldsymbol{\beta},\boldsymbol{\Sigma})=\boldsymbol{\mathcal{L}}_{1,T}+\boldsymbol{\mathcal{L}}_{2,T}$, such that,

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

and

$$\boldsymbol{\mathcal{L}}_{2,T}=\boldsymbol{\beta}^{*'}\sum_{t=1}^{k}\boldsymbol{x}_t\boldsymbol{\Sigma}^{-1}\boldsymbol{u}_t-\frac{k}{2}\boldsymbol{\beta}^{*'}\left(\frac{1}{k}\sum_{t=1}^{k}\boldsymbol{x}_t\boldsymbol{\Sigma}^{-1}\boldsymbol{x}_t'\right)\boldsymbol{\beta}^{*},$$

where $\boldsymbol{\beta}^* = \boldsymbol{\beta} - \boldsymbol{\beta}_0, \boldsymbol{\Sigma}^* = \boldsymbol{\Sigma} - \boldsymbol{\Sigma}_0, \boldsymbol{\eta}_t = (\boldsymbol{\Sigma}_0)^{-1}\boldsymbol{u}_t$ and $\boldsymbol{\Psi}_T = (\boldsymbol{\Sigma}_0)^{-1/2}\boldsymbol{\Sigma}^*(\boldsymbol{\Sigma}_0)^{-1/2}$. We note that only $\mathcal{L}_{2,T}$ depends on $\boldsymbol{\beta}^*$. We split the parameter space $\boldsymbol{S}_T = \boldsymbol{S}_{1,T} \cup \boldsymbol{S}_{2,T}$ with,

$$\boldsymbol{S}_{1,T} = \{(\boldsymbol{\beta}, \boldsymbol{\Sigma})\colon \|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_0\| \geq T^{-1/2+d} \log T,\ \boldsymbol{\beta} \text{ arbitrary}\}$$

and

$$\boldsymbol{S}_{2,T} = \{(\boldsymbol{\beta}, \boldsymbol{\Sigma})\colon \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \geq T^{-1/2+d} \log T \text{ and } \|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_0\| \leq T^{-1/2+d} \log T\}.$$

It has to be shown that,

$$\Pr\left(\sup_{k \geq \lfloor T\delta \rfloor} \sup_{(\boldsymbol{\beta}, \boldsymbol{\Sigma}) \in \boldsymbol{S}_{1,T} \cap \overline{\boldsymbol{\Theta}}_2} T^D \mathcal{L}(1, k; \boldsymbol{\beta}, \boldsymbol{\Sigma}) > 1\right) < \epsilon \tag{A.8}$$

and

$$\Pr\left(\sup_{k \geq \lfloor T\delta \rfloor} \sup_{(\boldsymbol{\beta}, \boldsymbol{\Sigma}) \in \boldsymbol{S}_{2,T} \cap \overline{\boldsymbol{\Theta}}_2} T^D \mathcal{L}(1, k; \boldsymbol{\beta}, \boldsymbol{\Sigma}) > 1\right) < \epsilon. \tag{A.9}$$

We start to show (A.8). On $\boldsymbol{S}_{1,T}$, $\boldsymbol{\mathcal{L}}_{2,T}$ is a quadratic function of $\boldsymbol{\beta}^*$ and has maximum value,

$$\sup_{\boldsymbol{S}_{1,T}} \mathcal{L}_{2,T} = \frac{k}{2}\left(\frac{1}{k}\sum_{t=1}^{k} \boldsymbol{x}_t \boldsymbol{\Sigma}^{-1}\boldsymbol{u}_t\right)' \left(\frac{1}{k}\sum_{t=1}^{k} \boldsymbol{x}_t \boldsymbol{\Sigma}^{-1}\boldsymbol{x}_t'\right)^{-1} \left(\frac{1}{k}\sum_{t=1}^{k} \boldsymbol{x}_t \boldsymbol{\Sigma}^{-1}\boldsymbol{u}_t\right).$$

Applying Property 1 gives,

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

Additionally,

$$\begin{aligned}
\sup_{k \geq \lfloor T\delta \rfloor} \left\|\frac{1}{k}\sum_{t=1}^{k} \boldsymbol{x}_t \boldsymbol{\Sigma}^{-1}\boldsymbol{u}_t\right\| &= \sup_{k \geq \lfloor T\delta \rfloor} \left\|\frac{1}{k}\sum_{t=1}^{k} S'(\boldsymbol{I}_n \otimes \boldsymbol{z}_t)\boldsymbol{\Sigma}^{-1}\boldsymbol{u}_t\right\| \\
&= \sup_{k \geq \lfloor T\delta \rfloor} \left\|S'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\frac{1}{k}\sum_{t=1}^{k}(\boldsymbol{I}_n \otimes \boldsymbol{z}_t)\boldsymbol{u}_t\right\| \\
&\leq \sup_{k \geq \lfloor T\delta \rfloor} \left\|\frac{1}{k}\sum_{t=1}^{k}(\boldsymbol{I}_n \otimes \boldsymbol{z}_t)\boldsymbol{u}_t\right\| \left\|S'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\right\|.
\end{aligned}$$

From the FCLT of Lemma A.2 we have for fixed $r > 0$ that,

$$\lim_{T \to \infty} \Pr\left(\sup_{k \geq \lfloor T\delta \rfloor} \left\|\frac{1}{k}\sum_{t=1}^{k}(\boldsymbol{I}_n \otimes \boldsymbol{z}_t)\boldsymbol{u}_t\right\| > rT^{d-1/2}\log^{1/2} T\right) = 0,$$

while $\|S'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\| = \sum_{i=1}^{n}(1+\lambda_i)^{-1} O_p(1)$, where λ_i, $i = 1, ..., n$ are the eigenvalues of $(\boldsymbol{\Sigma}_0)^{-1/2}\boldsymbol{\Sigma}^*(\boldsymbol{\Sigma}_0)^{-1/2}$. Hence,

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

which implies

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

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 $\boldsymbol{U}$ that diagonalizes $(\boldsymbol{I} + \boldsymbol{\Psi}_T)^{-1}$. Then we have that,

$$\frac{1}{k} \sum_{t=1}^{k} \boldsymbol{\eta}_t'((\boldsymbol{I} + \boldsymbol{\Psi}_T)^{-1} - \boldsymbol{I})\boldsymbol{\eta}_t = \mathrm{tr}\left(\mathrm{diag}\left\{ \frac{1}{1+\lambda_i} - 1 \right\} \left(\frac{1}{k} \boldsymbol{U} \sum_{t=1}^{k} \boldsymbol{\eta}_t \boldsymbol{\eta}_t' \boldsymbol{U}' \right) \right).$$

Because $\|\boldsymbol{U}\| = 1$ it suffices to investigate whether,

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

Then, for any $a > 0$ by the FCLT of Lemma A.2, it follows that,

$$\lim_{T \to \infty} \Pr\left(\sup_{k \geq \lfloor T\delta \rfloor} \frac{1}{k} \sum_{t=1}^{k} \|(\boldsymbol{\eta}_t \boldsymbol{\eta}_t' - \boldsymbol{I})\| > a b_T \right) = 0.$$

Hence, arguing as in Bai et al. (1998), we may show that,

$$\sup_{k \geq \lfloor T\delta \rfloor} \sup_{\boldsymbol{S}_{1,T} \cap \bar{\Theta}_2} \mathcal{L}_{1,T} \leq -\frac{k}{2} \left[\sum_{i=1}^{n} \left(\log(1+\lambda_i) + \left(\frac{1}{1+\lambda_i} - 1 \right) (1 + \mathrm{sign}(\lambda_i) a b_T) \right) \right],$$

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,

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

It is now straightforward to see that using similar arguments as in Bai et al. (1998) one can show that (A.9) holds. Therefore, the claim is shown on the compact parameter space $\overline{\boldsymbol{\Theta}}_2$. But as in Qu and Perron (2007) we conclude that the result is valid also for an unrestricted parameter space. Therefore the proof is complete. □

Property 4. *This property is not needed in our framework.*

Property 5 next is different from Property 5 in 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 $\overline{\boldsymbol{\Theta}}_3 = \{(\boldsymbol{\beta}, \boldsymbol{\Sigma})\colon \|\boldsymbol{\beta}\| \leq p_1, \lambda_{\min}(\boldsymbol{\Sigma}) \geq p_2, \lambda_{\max}(\boldsymbol{\Sigma}) \leq p_3\}$, where p_1, p_2 and p_3 are arbitrary constants that satisfy $p_1 < \infty$, $0 < p_2 \leq p_3 < \infty$. Define $S_T = \{(\boldsymbol{\beta}, \boldsymbol{\Sigma})\colon \|\boldsymbol{\beta} - \boldsymbol{\beta}^0\| \geq T^{-1/2+d}\log T$ or $\|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}^0\| \geq T^{-1/2+d}\log T\}$. Then, for any $\epsilon > 0$, there exists an $A > 0$, such that, when T is large,*

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

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

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

or

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

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

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

where

$$\left(\sum_{t=1}^{k}\boldsymbol{x}_t\boldsymbol{\Sigma}^{-1}\boldsymbol{x}_t'\right)^{-1} = \left(\sum_{t=1}^{k}S'(\boldsymbol{I}\otimes\boldsymbol{z}_t)\boldsymbol{\Sigma}^{-1}(I\otimes\boldsymbol{z}_t')S\right)^{-1} = \left(S'(\boldsymbol{\Sigma}^{-1}\otimes\sum_{t=1}^{k}\boldsymbol{z}_t\boldsymbol{z}_t')S\right)^{-1}.$$

From $l^{-1}\sum_{t=1}^{l}\boldsymbol{z}_t\boldsymbol{z}_t' \xrightarrow{a.s.} \boldsymbol{Q}_z$, for a given $\epsilon > 0$, we can always find a $k_1 > 0$ such that,

$$P(\sup_{k\geq k_1}\|\frac{1}{k}\sum_{t=1}^{k}\boldsymbol{z}_t\boldsymbol{z}_t' - \boldsymbol{Q}_z\| > \epsilon) < \epsilon.$$

Define $\boldsymbol{Q}_{\triangle} := k^{-1}\sum_{t=1}^{k}\boldsymbol{z}_t\boldsymbol{z}_t' - \boldsymbol{Q}_z$. Then,

$$
\begin{aligned}
&(S'(\boldsymbol{\Sigma}^{-1}\otimes\frac{1}{k}\sum_{t=1}^{k}\boldsymbol{z}_t\ \boldsymbol{z}_t')S)^{-1} - (S'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{Q}_z)S)^{-1}\\
=&(S'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{Q}_z)S + S'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{Q}_{\triangle})S)^{-1} - (S'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{Q}_z)S)^{-1}\\
=&-\boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{A}+\boldsymbol{B})^{-1},
\end{aligned}
$$

where $\boldsymbol{A} = S'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{Q}_z)S$ and $\boldsymbol{B} = S'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{Q}_{\triangle})S$. Because $\boldsymbol{\Sigma}^{-1}$ has uniformly bounded eigenvalues and $k^{-1}\sum_{t=1}^{k}\boldsymbol{z}_t\boldsymbol{z}_t'$ is positive definite for large k, $\boldsymbol{A}^{-1}$ and $\boldsymbol{B}^{-1}$ have bounded eigenvalues. Because $\boldsymbol{B}$ is uniformly small, $-\boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{A}+\boldsymbol{B})^{-1}$ is uniformly small for large k. That is,

$$
(S'(\boldsymbol{\Sigma}^{-1}\otimes k^{-1}\sum_{t=1}^{k}\boldsymbol{z}_t\boldsymbol{z}_t')S)^{-1} - (S'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{Q}_z)S)^{-1} \overset{\text{a.s.}}{=} o(1) \qquad \text{as} \quad k\to\infty.
$$

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

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

Now,

$$
\begin{aligned}
\sup_{k\geq Ah_T}\|\frac{1}{k}\sum_{t=1}^{k}\boldsymbol{x}_t\boldsymbol{\Sigma}^{-1}\boldsymbol{u}_t\| &= \sup_{k\geq Ah_T}\|\frac{1}{k}\sum_{t=1}^{k}S'(\boldsymbol{I}_n\otimes\boldsymbol{z}_t)\boldsymbol{\Sigma}^{-1}\boldsymbol{u}_t\|\\
&\leq \sup_{k\geq Ah_T}\|\frac{1}{k}\sum_{t=1}^{k}(\boldsymbol{I}_n\otimes\boldsymbol{z}_t)\boldsymbol{u}_t\|\|S'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{I}_n)\|. \qquad \text{(A.10)}
\end{aligned}
$$

From Lemma A.1 we have,

$$
P(\sup_{k\geq Ah_T}\|\frac{1}{k}\sum_{t=1}^{k}(\boldsymbol{I}_n\otimes\boldsymbol{z}_t)\boldsymbol{u}_t\| > ab_T) \leq \frac{\boldsymbol{C}_1}{\boldsymbol{A}h_Ta^2b_T} < \frac{2\boldsymbol{C}_1}{\boldsymbol{A}a^2h}, \qquad \text{(A.11)}
$$

for some $\boldsymbol{C}_1>0$, where the bound can be made arbitrarily small by choosing a large $\boldsymbol{A}$. For the second component,

$$
\|S'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{I}_n)\| \leq n\boldsymbol{C}_2\sum_{i=1}^{n}\frac{1}{1+\lambda_i}, \qquad \text{(A.12)}
$$

for some $0<\boldsymbol{C}_2<\infty$, which depends on the matrix S. Now, combining (A.10)-(A.12), we have, for any $\epsilon > 0$ there exists an $\bar{\boldsymbol{A}} > 0$, such that with probability no less than $1 - \epsilon$,

$$\sup_{k \geq \bar{\boldsymbol{A}} h_T} \sup_{(\boldsymbol{\beta},\boldsymbol{\Sigma}) \in S_{1,T} \cap \bar{\boldsymbol{\Theta}}_3} |\boldsymbol{\mathcal{L}}_{2T}| < k a^2 b_T^2 n^2 \boldsymbol{C}_2^2 \boldsymbol{M} (\sum_{i=1}^{n} \frac{1}{1+\lambda_i})^2 \leq \quad \frac{k}{2} \sum_{i=1}^{n} \frac{G a^2 b_T^2}{1+\lambda_i} = \frac{k}{2} \sum_{i=1}^{n} \frac{\gamma^2 b_T^2}{1+\lambda_i}$$

with $\boldsymbol{G} = 2n^3 \boldsymbol{C}_2^2 \boldsymbol{M}/p_2$. Because a^2 can be made arbitrarily small by choosing a large $\boldsymbol{A}$, so can γ^2. Hence, Property 5 follows. □

The next properties (Properties 6 - 10) are the same as Lemmas 6 – 10 of Bai (2000), and because the proofs are similar, they are omitted for the sake of space.

Property 6. *With ν_T satisfying Assumption 6, for each $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ such that $\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \leq \boldsymbol{M} v_T$ and $\|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_0\| \leq \boldsymbol{M} v_T$, with $\boldsymbol{M} < \infty$, we have*

$$\sup_{1 \leq k \leq T^{1/2-d} v_T^{-1}} \sup_{\boldsymbol{\lambda},\boldsymbol{\Xi}} \frac{\mathcal{L}(1,k;\boldsymbol{\beta} + T^{-1/2+d}\boldsymbol{\lambda}, \boldsymbol{\Sigma} + T^{-1/2+d}\boldsymbol{\Xi})}{\mathcal{L}(1,k;\boldsymbol{\beta},\boldsymbol{\Sigma})} = o_p(1). \tag{A.13}$$

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

$$\sup_{1 \leq k \leq M v_T^{-2}} \sup_{\boldsymbol{\lambda},\boldsymbol{\Xi}} \log \frac{\mathcal{L}(1,k;\boldsymbol{\beta} + T^{-1/2+d}\boldsymbol{\lambda}, \boldsymbol{\Sigma} + T^{-1/2+d}\boldsymbol{\Xi})}{\mathcal{L}(1,k;\boldsymbol{\beta},\boldsymbol{\Sigma})} = o_p(1). \tag{A.14}$$

Property 8. *We have,*

$$\sup_{\lfloor T\delta \rfloor \leq k \leq T} \sup_{\boldsymbol{\beta}^*,\boldsymbol{\Sigma}^*,\boldsymbol{\lambda},\boldsymbol{\Xi}} \log \frac{\mathcal{L}(1,k;\boldsymbol{\beta}_0 + T^{-1/2+d}\boldsymbol{\beta}^* + T^{-1+2d}\boldsymbol{\lambda}, \boldsymbol{\Sigma}_0 + T^{-1/2+d}\boldsymbol{\Sigma}^* + T^{-1+2d}\boldsymbol{\Xi})}{\mathcal{L}(1,k;\boldsymbol{\beta}_0 + T^{-1/2+d}\boldsymbol{\beta}^*, \boldsymbol{\Sigma}_0 + T^{-1/2+d}\boldsymbol{\Sigma}^*)} = o_p(1),$$

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

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

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{x}_t' \boldsymbol{\beta}_1^0 + \boldsymbol{\Sigma}_1^0 \boldsymbol{\eta}_t, & t &= 1, \ldots, T_1, \\ \boldsymbol{y}_t &= \boldsymbol{x}_t' \boldsymbol{\beta}_2^0 + \boldsymbol{\Sigma}_2^0 \boldsymbol{\eta}_t, & t &= T_1 + 1, \ldots, T_1 + T_2, \end{aligned}$$

where $\|\boldsymbol{\beta}_1^0 - \boldsymbol{\beta}_2^0\| \leq \boldsymbol{M} v_T$ and $\|\boldsymbol{\Sigma}_1^0 - \boldsymbol{\Sigma}_2^0\| \leq \boldsymbol{M} v_T$ for some $\boldsymbol{M} < \infty$. Let $k = T_1 + T_2$ be the size of the pooled sample and let $(\hat{\boldsymbol{\beta}}_n, \hat{\boldsymbol{\Sigma}}_n)$ be the associated estimates. Then $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_1^0 = O_p(T^{d-1/2})$ and $\hat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_1^0 = O_p(T^{d-1/2})$.

Property 10. *Property 10 is not needed in our framework.*

A.3 Proof of Lemma 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 $N := \lfloor T^{\frac{1}{2}-d}\nu_T^{-1}\rfloor$ and $\boldsymbol{A}_j = \{(k_1, ..., k_m) \in \boldsymbol{\Lambda}_\epsilon \colon |k_i - k_j^0| > N, i = 1, \ldots, m\}$, where $\boldsymbol{\Lambda}_\varepsilon$ is given in Assumption 11. Because $LR_T(\hat{k}_1, ..., \hat{k}_m) \geq LR_T(k_1^0, ..., k_m^0) \geq LR_T(k_1^0, ..., k_m^0, \boldsymbol{\beta}^0, \boldsymbol{\Sigma}^0) = 1$, to show $(\hat{k}_1, ..., \hat{k}_m) \notin \boldsymbol{A}_j$, it suffices in a first step to show that,

$$\Pr(\sup_{(k_1,...,k_m)\in\boldsymbol{A}_j} LR_T(k_1, \ldots, k_m) > \epsilon) < \epsilon. \tag{A.15}$$

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 $(k_1, \ldots, k_m) \in \boldsymbol{A}_j$,

$$LR_T(k_1, \ldots, k_m) \leq LR_T(k_1, \ldots, k_m, k_1^0, \ldots, k_{j-1}^0, k_j^0 - N, k_j^0 + N, k_{j+1}^0, \ldots, k_m^0). \tag{A.16}$$

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

$$D(k, l, \boldsymbol{\beta}, \boldsymbol{\Sigma}) = \frac{\prod_{t=k+1}^{l} f(\boldsymbol{y}_t|\boldsymbol{x}_t; \boldsymbol{\beta}, \boldsymbol{\Sigma})}{\prod_{t=k+1}^{l} f(\boldsymbol{y}_t|\boldsymbol{x}_t; \boldsymbol{\beta}^0, \boldsymbol{\Sigma}^0)},$$

and its optimal value,

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

The likelihood-ratio of the entire sample can be written as,

$$LR_T(k_1, \ldots, k_m) = D(0, k_1) \cdot D(k_1, k_2) \cdot \cdots \cdot D(k_m, T). \tag{A.17}$$

The right hand side of (A.16) can be written as the product of at most $(2m+2)$ terms expressible as $D(l, k)$ as in (A.17). 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 Properties 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,

$$LR_T(k_1, \ldots, k_m) \leq O_p(T^{(2m+1)B})\, D(k_j^0 - N, k_j^0 + N). \tag{A.18}$$

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

$$\begin{aligned} D(k_j^0 - N, k_j^0 + N) &= \sup_{\boldsymbol{\beta}, \boldsymbol{\Sigma}} [D(k_j^0 - N, k_j^0; \boldsymbol{\beta}, \boldsymbol{\Sigma}) \cdot D(k_j^0, k_j^0 + N; \boldsymbol{\beta}, \boldsymbol{\Sigma})] \\ &= \sup_{\boldsymbol{\beta}, \boldsymbol{\Sigma}} [LR_T^*(k_j^0 - N, k_j^0; N^{1/2}(\boldsymbol{\beta} - \boldsymbol{\beta}_j^0), N^{1/2}(\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_j^0)) \\ &\quad \times (LR_T^*(k_j^0, k_j^0 + N; N^{1/2}(\boldsymbol{\beta} - \boldsymbol{\beta}_{j+1}^0), N^{1/2}(\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_{j+1}^0))]. \end{aligned} \tag{A.19}$$

This follows from the definition of LR_T^* and the fact that $(\boldsymbol{\beta}_j^0, \boldsymbol{\Sigma}_j^0)$ is the true parameter for the segment $[k_j^0 - N, k_j^0]$ and $(\boldsymbol{\beta}_{j+1}^0, \boldsymbol{\Sigma}_{j+1}^0)$ 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 $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ that,

$$\begin{aligned} \max\{N^{1/2}\|\boldsymbol{\beta} - \boldsymbol{\beta}_j^0\|, N^{1/2}\|\boldsymbol{\beta} - \boldsymbol{\beta}_{j+1}^0\|\} &\geq N^{1/2}\|\boldsymbol{\beta}_j^0 - \boldsymbol{\beta}_{j+1}^0\|/2 \\ \max\{N^{1/2}\|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_j^0\|, N^{1/2}\|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_{j+1}^0\|\} &\geq N^{1/2}\|\boldsymbol{\Sigma}_j^0 - \boldsymbol{\Sigma}_{j+1}^0\|/2. \end{aligned}$$

By Assumption 6, we either have $N^{1/2}\|\boldsymbol{\beta}_j^0 - \boldsymbol{\beta}_{j+1}^0\|/2 \geq \log N$ or $N^{1/2}\|\boldsymbol{\Sigma}_j^0 - \boldsymbol{\Sigma}_{j+1}^0\|/2 \geq \log N$. This follows from $\|\boldsymbol{\beta}_j^0 - \boldsymbol{\beta}_{j+1}^0\| \geq \nu_T \boldsymbol{C}$ for some $\boldsymbol{C} > \mathbf{0}$, then,

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

Now suppose that $N^{1/2}\|\boldsymbol{\beta}_j^0 - \boldsymbol{\beta}_{j+1}^0\|/2 \geq \log N$. Then we have either (i) $N^{1/2}\|\boldsymbol{\beta} - \boldsymbol{\beta}_j^0\| \geq \log N$ or (ii) $N^{1/2}\|\boldsymbol{\beta} - \boldsymbol{\beta}_{j+1}^0\| \geq \log N$. For case (i) we can apply Property 3 to the first term inside the brackets of (A.19) to obtain,

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

for every $\boldsymbol{A} > 0$. Moreover, by Property 2 the second term inside the bracket of (A.19) is bounded by $O_p(\log T)$. Similarly, for case (ii), we can apply Property 3 to show that the second term of (A.19) 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 $A > 0$. Moreover, $N^{-A} \leq T^{-A/2}$ since $N \geq T^{1/2}$ for all large T. Thus, from (A.18), $LR_T(k_1, \dots, k_m) \leq O_p(T^{(2m+1)B-\frac{1}{2}A}) \log T \xrightarrow{p} 0$ for a large A. This proves (A.15).

Now by Proposition 2 of Bai (2000) we can deduce that $\hat{k}_j - k_j^0 = O_p(\nu_T^2)$, for $j = 1, \dots, m$, using the preliminary convergence order given by (A.15). The convergence rate for the estimated regression coefficients $\boldsymbol{\beta}_j$ and covariances $\boldsymbol{\Sigma}_j$ follows as in Bai (1997b) and Bai and Perron (1998) due to the fast convergence of the estimated break points. □

A.4 Proof of Theorem 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), that,

$$\frac{1}{2}\sum_{t=T_j^0+\lfloor s/\nu_T^2 \rfloor}^{T_j^0} \boldsymbol{u}_t'\Big((\boldsymbol{\Sigma}_j^0)^{-1} - (\boldsymbol{\Sigma}_{j+1}^0)^{-1}\Big)\boldsymbol{u}_t = \tfrac{1}{2}tr\Big((\boldsymbol{\Sigma}_j^0)^{\frac{1}{2}}(\boldsymbol{\Sigma}_{j+1}^0)^{-1}\boldsymbol{c}_{2,j}(\boldsymbol{\Sigma}_j^0)^{-\frac{1}{2}}\nu_T \sum_{t=T_j^0+\lfloor s/\nu_T^2 \rfloor}^{T_j^0}(\boldsymbol{\eta}_t\, \boldsymbol{\eta}_t' - \boldsymbol{I}\Big)$$
$$-\tfrac{r}{2}\nu_T\; tr\Big((\boldsymbol{\Sigma}_{j+1}^0)^{-1}\boldsymbol{c}_{2,j}\Big),$$

and for the second term,

$$-\frac{r}{2}(\log|\boldsymbol{\Sigma}_j^0| - \log|\boldsymbol{\Sigma}_{j+1}^0|) = \frac{r}{2}\nu_T\; tr\Big(\boldsymbol{c}_{2,j}(\boldsymbol{\Sigma}_{j+1}^0)^{-1}\Big) + \frac{r}{4}\nu_T^2\; tr\Big([\boldsymbol{c}_{2,j}(\boldsymbol{\Sigma}_{j+1}^0)^{-1}]^2\Big).$$

The sum of the first two terms is,

$$\frac{1}{2}\sum_{T_j^0+\lfloor s/\nu_T^2 \rfloor}^{T_j^0} \boldsymbol{u}_t'\Big((\boldsymbol{\Sigma}_j^0)^{-1} - (\boldsymbol{\Sigma}_{j+1}^0)^{-1}\Big)\boldsymbol{u}_t - \frac{r}{2}(log|\boldsymbol{\Sigma}_j^0| - log|\boldsymbol{\Sigma}_{j+1}^0|)$$
$$= \frac{1}{2}tr\Big((\boldsymbol{\Sigma}_j^0)^{\frac{1}{2}}(\boldsymbol{\Sigma}_{j+1}^0)^{-1}\boldsymbol{c}_{2,j}(\boldsymbol{\Sigma}_j^0)^{-\frac{1}{2}}\nu_T \sum_{T_j^0+\lfloor s/\nu_T^2 \rfloor}^{T_j^0}(\boldsymbol{\eta}_t\, \boldsymbol{\eta}_t' - \boldsymbol{I})\Big)$$
$$+ \frac{r}{4}\nu_T^2\; tr\Big([\boldsymbol{c}_{2,j}(\boldsymbol{\Sigma}_{j+1}^0)^{-1}]^2\Big) = \mathrm{I} + \mathrm{II}.$$

Now,

$$T^{1-2d}(\mathrm{I}+\mathrm{II}) \xrightarrow{d} \frac{1}{2}tr\Big((\boldsymbol{\Sigma}_j^0)^{\frac{1}{2}}(\boldsymbol{\Sigma}_{j+1}^0)^{-1}\boldsymbol{c}_{2,j}(\boldsymbol{\Sigma}_j^0)^{-\frac{1}{2}}\boldsymbol{\xi}_{1,d,j}(s)\Big) + \frac{s}{4}tr\Big([(\boldsymbol{\Sigma}_{j+1}^0)^{-1}\boldsymbol{c}_{2,j}]^2\Big)$$
$$= \frac{1}{2}tr(\boldsymbol{A}_{1,j}\boldsymbol{\xi}_{1,d,j}(s)) + \frac{s}{4}tr(\boldsymbol{A}_{1,j}^2),$$

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

$$-\frac{1}{2}\sum_{t=T_j^0+[s/\nu_T^2]}^{T_j^0}(\boldsymbol{\beta}_j^0-\boldsymbol{\beta}_{j+1}^0)^T\boldsymbol{x}_t(\boldsymbol{\Sigma}_{j+1}^0)^{-1}\boldsymbol{x}_t'(\boldsymbol{\beta}_j^0-\boldsymbol{\beta}_{j+1}^0) \xrightarrow{P} \frac{1}{2}s\ \boldsymbol{c}_{1,j}'\ \boldsymbol{Q}_{1,j}\ \boldsymbol{c}_{1,j}.$$

Note that $\boldsymbol{x}_t$ 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,

$$-T^{1-2d}\sum_{t=T_j^0+\lfloor s/\nu_T^2\rfloor}^{T_j^0}(\boldsymbol{\beta}_j^0-\boldsymbol{\beta}_{j+1}^0)'\boldsymbol{x}_t(\boldsymbol{\Sigma}_{j+1}^0)^{-1}\boldsymbol{u}_t \xrightarrow{d} \boldsymbol{c}_{1,j}'(\boldsymbol{\Pi}_{1,j})^{\frac{1}{2}}\boldsymbol{\xi}_{1,d,j}(s)$$

with

$$\boldsymbol{\Pi}_{1,j} = \lim_{T\to\infty} Var\Big\{(T_j^0-T_{j-1}^0)^{-\frac{1}{2}}\Big[\sum_{t=T_j^0+\lfloor s/\nu_T^2\rfloor}^{T_j^0}\boldsymbol{x}_t(\boldsymbol{\Sigma}_{j+1}^0)^{-1}(\boldsymbol{\Sigma}_j^0)^{\frac{1}{2}}\boldsymbol{\eta}_t\Big]\Big\}.$$

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

$$T^{1-2d}lr_j^1\Big([\frac{s}{\nu_T^2}]\Big) \xrightarrow{d} -\frac{|s|}{2}\Big[\frac{1}{2}tr(\boldsymbol{A}_{1,j}^2) + \boldsymbol{c}_{1,j}'\ \boldsymbol{Q}_{1,j}\ \boldsymbol{c}_{1,j}\Big]$$
$$+\frac{1}{2}\,\mathrm{vec}(\boldsymbol{A}_{1,j}')\,\mathrm{vec}(\boldsymbol{\xi}_{1,d,j}(s)) + \boldsymbol{c}_{1,j}'(\pi_{1,j})^{\frac{1}{2}}\boldsymbol{\xi}_{1,d,j}(s).$$

Now, $\mathrm{vec}(\boldsymbol{A}_{1,j})'\,\mathrm{vec}(\boldsymbol{\xi}_{1,d,j}(s)) \overset{d}{=} \Big(vec(\boldsymbol{A}_{1,j})^T\boldsymbol{\Omega}_{1,j}^0 vec(A_{1,j})\Big)^{\frac{1}{2}}V_{1,d,j}(s)$, where $V_{1,d,j}(s)$ is a standard fractional Brownian motion.

Similarly, $\boldsymbol{c}_{1,j}'(\boldsymbol{\Pi}_{1,j})^{\frac{1}{2}}\boldsymbol{\xi}_{1,d,j}(s) \overset{d}{=} (\boldsymbol{c}_{1,j}'\ \Pi_{1,j}\ \boldsymbol{c}_{1,j})^{\frac{1}{2}}U_{1,d,j}(s)$ 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,

$$\Big(vec(\boldsymbol{A}_{1,j}')\boldsymbol{\Omega}_{1,j}^0 vec(\boldsymbol{A}_{1,j})/4\Big)^{\frac{1}{2}}V_{1,d,j}(s) + \Big(\boldsymbol{c}_{1,j}'(\pi_{1,j})\boldsymbol{c}_{1,j}\Big)^{\frac{1}{2}}U_{1,d,j}(s)$$
$$\overset{d}{=}\Big(vec(\boldsymbol{A}_{1,j})\boldsymbol{\Omega}_{1,j}^0 vec(\boldsymbol{A}_{1,j})/4 + \boldsymbol{c}_{1,j}'(\pi_{1,j})\boldsymbol{c}_{1,j}\Big)^{\frac{1}{2}}W_{1,j,d}(s)$$
$$\equiv T_{1,j}\ W_{1,j,d}(s),$$

where $W_{1,j,d}(s)$ is a unit fractional Brownian motion.

Hence, with $\boldsymbol{\Delta}_{1,j} = tr(\boldsymbol{A}_{1,j}^2)/2 + \boldsymbol{c}_{1,j}' \ \boldsymbol{Q}_{1,j} \ \boldsymbol{c}_{1,j}$, we have,

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

The proof for $s > 0$ is similar:

$$T^{1-2d} lr_j^1\Big([\frac{s}{\nu_T^2}]\Big) \xrightarrow{d} -\frac{|s|}{2}\boldsymbol{\Delta}_{2,j} + T_{2,j} \ W_{2,j,d}(s), \tag{A.20}$$

with $\boldsymbol{\Delta}_{2,j} = tr(\boldsymbol{A}_{2,j}^2)/2 + \boldsymbol{c}_{1,j}' \ \boldsymbol{Q}_{2,j} \ \boldsymbol{c}_{1,j}$ and

$$T_{2,j} = \Big[vec(\boldsymbol{A}_{2,j}')\boldsymbol{\Omega}_{2,j}^0 vec(\boldsymbol{A}_{2,j})/4 + \boldsymbol{c}_{1,j}'(\pi_{2,j})\boldsymbol{c}_{1,j}\Big]^{\frac{1}{2}}.$$

By definition $lr_j^1(0) = 0$. 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,

$$T^{1-2d}\nu_T^2(\hat{T}_j - T_j^0) \xrightarrow{d} \underset{s}{argmax} \begin{cases} -\frac{|s|}{2}\Delta_{1,j} + T_{1,j} \ W_{j,d}(s), s \leq 0, \\ -\frac{|s|}{2}\Delta_{2,j} + T_{2,j} \ W_{j,d}(s), s > 0, \end{cases} \tag{A.21}$$

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 $\boldsymbol{u} = (\boldsymbol{\Delta}_{1,j}^2/T_{1,j}^2)s$, we obtain Theorem 1. □

A.5 Proof of Theorem 2

Proof of Theorem 2. Before proceeding with the proof we first introduce some notation. Let

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

be the estimated covariance matrix using the full sample estimate of $\boldsymbol{\beta}_a$ obtained under the null hypothesis of no change and using the estimate of $\boldsymbol{\beta}_b$ based on data up to the last date of regime j, defined as,

$$\tilde{\boldsymbol{\beta}}_{b1,j} = \left(\sum_{t=1}^{T_j}\boldsymbol{x}_{bt}\tilde{\boldsymbol{\Sigma}}_{1,j}^{-1}\boldsymbol{x}_{bt}'\right)^{-1}\sum_{t=1}^{T_j}\boldsymbol{x}_t\tilde{\boldsymbol{\Sigma}}_{1,j}^{-1}(\boldsymbol{y}_t - \boldsymbol{x}_{at}'\tilde{\boldsymbol{\beta}}_a).$$

Additionally,

$$\hat{\boldsymbol{\Sigma}}_j = \frac{1}{T_j - T_{j-1}} \sum_{t=T_{j-1}+1}^{T_j} (\boldsymbol{y}_t - \boldsymbol{x}'_{at}\hat{\boldsymbol{\beta}}_a - \boldsymbol{x}'_{bt}\hat{\boldsymbol{\beta}}_{bj})(\boldsymbol{y}_t - \boldsymbol{x}'_{at}\hat{\boldsymbol{\beta}}_a - \boldsymbol{x}'_{bt}\hat{\boldsymbol{\beta}}_{bj})',$$

is the estimate of the covariance matrix of the errors under the alternative hypothesis using the full sample estimate of $\boldsymbol{\beta}_a$ and using the estimate of $\boldsymbol{\beta}_b$ based on data from regime j only, that is,

$$\hat{\boldsymbol{\beta}}_{b,j} = \left(\sum_{t=T_{j-1}+1}^{T_j} \boldsymbol{x}_{bt}\hat{\boldsymbol{\Sigma}}_j^{-1}\boldsymbol{x}'_{bt} \right)^{-1} \sum_{t=T_{j-1}+1}^{T_j} \boldsymbol{x}_t\hat{\boldsymbol{\Sigma}}_j^{-1}(\boldsymbol{y}_t - \boldsymbol{x}'_{at}\hat{\boldsymbol{\beta}}_a).$$

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

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

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

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

Applying this to the terms F_T^j,

$$\begin{aligned}
F_T^j :=& F_{1,T}^j + F_{2,T}^j \\
=& \operatorname{tr}(T_{j+1}(\boldsymbol{\Sigma}^0)^{-1}(\tilde{\boldsymbol{\Sigma}}_{1,j+1} - \boldsymbol{\Sigma}^0) - T_j(\boldsymbol{\Sigma}^0)^{-1}(\tilde{\boldsymbol{\Sigma}}_{1,j} - \boldsymbol{\Sigma}^0)) \qquad \text{(A.22)}\\
&- (T_{j+1} - T_j)(\boldsymbol{\Sigma}^0)^{-1}(\hat{\boldsymbol{\Sigma}}_{j+1} - \boldsymbol{\Sigma}^0)) \\
&- \frac{1}{2}\operatorname{tr}(T_{j+1}[(\boldsymbol{\Sigma}^0)^{-1}(\tilde{\boldsymbol{\Sigma}}_{1,j+1} - \boldsymbol{\Sigma}^0)]^2 \qquad \text{(A.23)}\\
&- T_j[(\boldsymbol{\Sigma}^0)^{-1}(\tilde{\boldsymbol{\Sigma}}_{1,j} - \boldsymbol{\Sigma}^0)]^2 - (T_{j+1} - T_j)[(\boldsymbol{\Sigma}^0)^{-1}(\hat{\boldsymbol{\Sigma}}_{j+1} - \boldsymbol{\Sigma}^0)]^2).
\end{aligned}$$

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

$$\boldsymbol{Y} = \boldsymbol{X}_a\boldsymbol{\beta}_a + \boldsymbol{X}_b\boldsymbol{\beta}_b + \boldsymbol{U},$$

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

$$\boldsymbol{Y}_{1,j} = \boldsymbol{X}_{a1,j}\boldsymbol{\beta}_a + \boldsymbol{X}_{b1,j}\boldsymbol{\beta}_{b1,j} + \boldsymbol{U}_{1,j}.$$

We now define $\boldsymbol{Y}_{1,j}^d = (\boldsymbol{I}_T \otimes \tilde{\boldsymbol{\Sigma}}_{1,j}^{-1/2})\boldsymbol{Y}_{1,j}$, $\boldsymbol{W}_{1,j} = (\boldsymbol{I}_T \otimes \tilde{\boldsymbol{\Sigma}}_{1,j}^{-1/2})\boldsymbol{X}_{a1,j}$, $\boldsymbol{Z}_{1,j} = (\boldsymbol{I}_T \otimes \tilde{\boldsymbol{\Sigma}}_{1,j}^{-1/2})\boldsymbol{X}_{b1,j}$ and $\boldsymbol{U}_{1,j}^d = (\boldsymbol{I}_T \otimes \tilde{\boldsymbol{\Sigma}}_{1,j}^{-1/2})\boldsymbol{U}_{1,j}$. Then, omitting the subscript when the full sample is used, we have,

$$\tilde{\boldsymbol{\beta}}_a = [\boldsymbol{W}'\boldsymbol{M}_{\boldsymbol{Z}}\boldsymbol{W}]^{-1}\boldsymbol{W}'\boldsymbol{M}_{\boldsymbol{Z}}\boldsymbol{Y}^d, \qquad \text{(A.24)}$$

$$\tilde{\boldsymbol{\beta}}_{b1,j} = (\boldsymbol{Z}_{1,j}'\boldsymbol{Z}_{1,j})^{-1}\boldsymbol{Z}_{1,j}'(\boldsymbol{Y}_{1,j}^d - \boldsymbol{W}_{1,j}\tilde{\boldsymbol{\beta}}_a), \qquad \text{(A.25)}$$

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

$$\boldsymbol{Y}_{j+1} = \boldsymbol{X}_{a,j+1}\boldsymbol{\beta}_a + \boldsymbol{X}_{b,j+1}\boldsymbol{\beta}_{b,j+1} + \boldsymbol{U}_{j+1}. \qquad \text{(A.26)}$$

Define $\bar{\boldsymbol{Y}}_{j+1}^d = (\boldsymbol{I}_T \otimes \hat{\boldsymbol{\Sigma}}_{j+1}^{-1/2})\boldsymbol{Y}_{j+1}$, $\bar{\boldsymbol{W}}_{j+1} = (\boldsymbol{I}_T \otimes \hat{\boldsymbol{\Sigma}}_{j+1}^{-1/2})\boldsymbol{X}_{a,j+1}$, $\bar{\boldsymbol{Z}}_{j+1} = (\boldsymbol{I}_T \otimes \hat{\boldsymbol{\Sigma}}_{j+1}^{-1/2})\boldsymbol{X}_{b,j+1}$, $\bar{\boldsymbol{U}}_{j+1}^d = (\boldsymbol{I}_T \otimes \hat{\boldsymbol{\Sigma}}_{j+1}^{-1/2})\boldsymbol{U}_{j+1}$, $\bar{\boldsymbol{Z}} = \operatorname{diag}(\bar{Z}_1, \ldots, \bar{Z}_{m+1})$. Then, omitting the subscript when the full sample is used, we have

$$\hat{\boldsymbol{\beta}}_a = [\bar{\boldsymbol{W}}'\boldsymbol{M}_{\bar{Z}}\bar{\boldsymbol{W}}]^{-1}\bar{\boldsymbol{W}}'\boldsymbol{M}_{\bar{Z}}\bar{\boldsymbol{Y}}^d, \qquad \text{(A.27)}$$

$$\hat{\boldsymbol{\beta}}_{b,j+1} = (\bar{\boldsymbol{Z}}_{j+1}'\bar{\boldsymbol{Z}}_{j+1})^{-1}\bar{\boldsymbol{Z}}_{j+1}'(\bar{\boldsymbol{Y}}_{j+1}^d - \bar{W}_{j+1}\hat{\beta}_a). \qquad \text{(A.28)}$$

Note that the choice of the estimate of the covariance matrix in (A.24) to (A.28) will have no effect provided a consistent one is used. As in Qu and Perron (2007) (supplement, pp. 25-26) we can show for the first component of $F^{j}_{1,T}$ (or with obvious changes for the second component) that,

$$T_{j+1}\,\mathrm{tr}((\mathbf{\Sigma}^0)^{-1}\tilde{\mathbf{\Sigma}}_{j+1}) = \boldsymbol{A}'_T\boldsymbol{W}'_{1,j+1}\boldsymbol{M}_{Z_{1,j+1}}\boldsymbol{W}_{1,j+1}\boldsymbol{A}_T - \boldsymbol{U}^{d'}_{1,j+1}\boldsymbol{P}_{Z_{1,j+1}}\boldsymbol{U}^{d}_{1,j+1} - 2(\boldsymbol{M}_{Z_{1,j+1}}\boldsymbol{W}_{1,j+1}\boldsymbol{A}_T)'\boldsymbol{U}^{d}_{1,j+1} + \boldsymbol{U}'_{1,j+1}(I_T\otimes(\mathbf{\Sigma}^0)^{-1})\boldsymbol{U}_{1,j+1} + o_p(1),$$

where $\boldsymbol{A}_T = [\boldsymbol{W'M_ZW}]^{-1}\boldsymbol{W'M_ZU}^d$. For the third component of $F^{j}_{1,T}$ it can be shown that,

$$\begin{aligned}&(T_{j+1}-T_j)\,\mathrm{tr}((\mathbf{\Sigma}^0)^{-1}\hat{\mathbf{\Sigma}}_{j+1})\\ &=\bar{\boldsymbol{A}}'_T\bar{\boldsymbol{W}}'_{j+1}\bar{\boldsymbol{M}}_{\bar{\boldsymbol{Z}}_{j+1}}\bar{\boldsymbol{W}}_{j+1}\bar{\boldsymbol{A}}_T - \bar{\boldsymbol{U}}^{d'}_{j+1}\bar{\boldsymbol{P}}_{\bar{Z}_{j+1}}\bar{\boldsymbol{U}}^{d}_{j+1}\\ &\quad - 2(\bar{\boldsymbol{M}}_{\bar{Z}_{j+1}}\bar{\boldsymbol{W}}_{j+1}\bar{\boldsymbol{A}}_T)'\bar{\boldsymbol{U}}^{d}_{j+1} + \bar{\boldsymbol{U}}'_{j+1}(I_T\otimes(\mathbf{\Sigma}^0)^{-1})\bar{\boldsymbol{U}}_{j+1} + o_p(1),\end{aligned}$$

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

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

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

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

Using similar arguments,

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

and

$$T^{-2d}\boldsymbol{U}_{j+1}^{d'}\boldsymbol{P}_{\bar{Z}_{j+1}}\boldsymbol{U}_{j+1}^{d}$$
$$\Rightarrow (\boldsymbol{W}_{D,p_b}^{*}(\lambda_{j+1})-\boldsymbol{W}_{D,p_b}^{*}(\lambda_j))'(\boldsymbol{W}_{D,p_b}^{*}(\lambda_{j+1})-\boldsymbol{W}_{D,p_b}^{*}(\lambda_j))/(\lambda_{j+1}-\lambda_j).$$

These results imply that the first component in (A.22) has the limit,

$$F_{1,T}^{j}\Rightarrow\frac{(\lambda_j\boldsymbol{W}_{D,p_b}^{*}(\lambda_{j+1})-\lambda_{j+1}\boldsymbol{W}_{D,p_b}^{*}(\lambda_j))'(\lambda_j\boldsymbol{W}_{D,p_b}^{*}(\lambda_{j+1})-\lambda_{j+1}\boldsymbol{W}_{D,p_b}^{*}(\lambda_j))}{(\lambda_{j+1}-\lambda_j)\lambda_j\lambda_{j+1}}. \quad (A.29)$$

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

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

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

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

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

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

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

$((\boldsymbol{\Sigma}^0)^{-1}\tilde{\boldsymbol{\Sigma}}_{1,j} - \boldsymbol{I})^F$, $((\boldsymbol{\Sigma}^0)^{-1}\tilde{\boldsymbol{\Sigma}}_{1,j} - \boldsymbol{I})^S$, $((\boldsymbol{\Sigma}^0)^{-1}\hat{\boldsymbol{\Sigma}}_{j+1} - \boldsymbol{I})^F$ and $((\boldsymbol{\Sigma}^0)^{-1}\hat{\boldsymbol{\Sigma}}_{j+1} - \boldsymbol{I})^S$. Then,

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

and, in view of (A.23),

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

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

$$\begin{aligned}&\frac{T_{j+1}}{T^{2d}}((\boldsymbol{\Sigma}^0)^{-1}\tilde{\boldsymbol{\Sigma}}_{1,j+1} - \boldsymbol{I})^S((\boldsymbol{\Sigma}^0)^{-1}\tilde{\boldsymbol{\Sigma}}_{1,j+1} - \boldsymbol{I})^S \\ &\quad = \frac{T}{T_{j+1}}\left(T^{-1/2-d}\sum_{t=1}^{T_{j+1}}[(\boldsymbol{\Sigma}^0)^{-1}\boldsymbol{u}_t\boldsymbol{u}_t' - \boldsymbol{I}]\right)^S\left(T^{-1/2-d}\sum_{t=1}^{T_{j+1}}[(\boldsymbol{\Sigma}^0)^{-1}\boldsymbol{u}_t\boldsymbol{u}_t' - \boldsymbol{I}]\right)^S + o_p(1) \\ &\quad \Rightarrow \frac{\xi_n^d(\lambda_{j+1})^S\xi_n^d(\lambda_{j+1})^S}{\lambda_{j+1}}; \\ &\frac{T_j}{T^{2d}}((\boldsymbol{\Sigma}^0)^{-1}\tilde{\boldsymbol{\Sigma}}_{1,j} - \boldsymbol{I})^S((\boldsymbol{\Sigma}^0)^{-1}\tilde{\boldsymbol{\Sigma}}_{1,j} - \boldsymbol{I})^S \\ &\quad = \frac{T}{T_j}\left(T^{-1/2-d}\sum_{t=1}^{T_j}[(\boldsymbol{\Sigma}^0)^{-1}\boldsymbol{u}_t\boldsymbol{u}_t' - I]\right)^S\left(T^{-1/2-d}\sum_{t=1}^{T_j}[(\boldsymbol{\Sigma}^0)^{-1}\boldsymbol{u}_t\boldsymbol{u}_t' - I]\right)^S + o_p(1) \\ &\quad \Rightarrow \frac{\xi_n^d(\lambda_j)^S\xi_n^d(\lambda_j)^S}{\lambda_j};\end{aligned}$$

and

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

where $\boldsymbol{\xi}_D^*(\cdot)$ is an $n\times n$ matrix whose elements are,

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

and where $\boldsymbol{\xi}_D$ is a (nonstandard) fractional Brownian motion defined on $[0,1]$ such that $\mathrm{Var}(\mathrm{vec}(\boldsymbol{\xi}_D(1)))=\boldsymbol{\Omega}$ (which follows from Theorem 4.8.2 of Giraitis et al. (2012) p.109). Hence,

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

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

$$\begin{aligned}
\mathrm{vec}(\boldsymbol{\xi}_{D,n}^*(\lambda_{j+1})^S)'\,\mathrm{vec}(\boldsymbol{\xi}_{D,n}^*(\lambda_{j+1})^S)&=\mathrm{vec}(\boldsymbol{\xi}_{D,n}^*(\lambda_{j+1}))'\boldsymbol{H}'\boldsymbol{H}\,\mathrm{vec}(\boldsymbol{\xi}_{D,n}^*(\lambda_{j+1}))\\
&\overset{d}{=}\boldsymbol{W}_{D,n_b^*}^*(\lambda_{j+1})'\boldsymbol{H}\boldsymbol{\Omega}\boldsymbol{H}'\boldsymbol{W}_{D,n_b^*}^*(\lambda_{j+1}),
\end{aligned}$$

where $\boldsymbol{W}^*_{D,n_b^*}$ is an n_b^* vector of processes as defined in Theorem 2. Hence, we have

$$\sum_{j=1}^{m} F_{2,T}^{j} \Rightarrow -\frac{1}{2}\Bigg[\frac{\boldsymbol{W}^*_{D,n_b^*}(\lambda_{j+1})'\boldsymbol{H}'\boldsymbol{\Omega}\boldsymbol{H}\boldsymbol{W}^*_{D,n_b^*}(\lambda_{j+1})}{\lambda_{j+1}} - \frac{\boldsymbol{W}^*_{D,n_b^*}(\lambda_j)'\boldsymbol{H}'\boldsymbol{\Omega}\boldsymbol{H}\boldsymbol{W}^*_{D,n_b^*}(\lambda_j)}{\lambda_j}$$
$$- \frac{(\boldsymbol{W}^*_{D,n_b^*}(\lambda_{j+1}) - \boldsymbol{W}^*_{D,n_b^*}(\lambda_j))'\boldsymbol{H}'\boldsymbol{\Omega}\boldsymbol{H}(\boldsymbol{W}^*_{D,n_b^*}(\lambda_{j+1}) - \boldsymbol{W}^*_{D,n_b^*}(\lambda_j))}{\lambda_{j+1} - \lambda_j}\Bigg]$$
$$=(\lambda_j \boldsymbol{W}^*_{D,n_b^*}(\lambda_{j+1}) - \lambda_{j+1}\boldsymbol{W}^*_{D,n_b^*}(\lambda_j))'\boldsymbol{H}'\boldsymbol{\Omega}\boldsymbol{H}$$
$$\times (\lambda_j \boldsymbol{W}^*_{D,n_b^*}(\lambda_{j+1}) - \lambda_{j+1}\boldsymbol{W}^*_{D,n_b^*}(\lambda_j))/(\lambda_j\lambda_{j+1}(\lambda_{j+1} - \lambda_j)). \quad \text{(A.30)}$$

By combining (A.29) and (A.30) we have shown the limiting distribution of our test. □

A.6 Proof of Theorem 4

Proof. From Theorem 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 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 $(1-\alpha)\%$. □

A.7 Proof of Lemma 3

Proof. The proof of Lemma 3 is based on the reparameterization of $\boldsymbol{\beta}$ which comes from the rate of integration of the regressors $\boldsymbol{x}$. We use the fact that the regression residuals are,

$$\begin{aligned} \boldsymbol{u}_t(\boldsymbol{\beta}) &= \boldsymbol{y}_t - \tilde{\boldsymbol{x}}_t'(\boldsymbol{\beta}_0 + T^{-d_x}\boldsymbol{\beta}) \\ &= \boldsymbol{u}_t - T^{-d_x}\tilde{\boldsymbol{x}}_t'\boldsymbol{\beta}. \end{aligned}$$

Therefore, using from now on $\boldsymbol{x}_t = T^{-d_x/2}\tilde{\boldsymbol{x}}_t$, we have that,

$$\boldsymbol{u}_t(\boldsymbol{\beta}) = \boldsymbol{u}_t - T^{-d_x/2}\boldsymbol{x}_t'\boldsymbol{\beta}.$$

In this way the likelihood-ratio for stochastic regressors has the same form as the likelihood-ratio for deterministic regressors, except that $\boldsymbol{u}_t$ is replaced by $\boldsymbol{u}_t(\boldsymbol{\beta})$. All notations in the proof of Lemma 1 remain unchanged as long as the $I(d_x)$ regressors are considered as divided by $T^{d_x/2}$. To prove the lemma it is enough to show that our ten properties still hold for the stochastic, possibly non-stationary, regressors, which is the main problem. Properties 1, 2 and 6 to 10 still hold in this case. Adaptions are needed for Properties 3 and 5 so we will give the proofs here. □

Property 11. *Let* $S_T = \{(\boldsymbol{\beta}, \boldsymbol{\Sigma})\colon \|\boldsymbol{\beta} - \boldsymbol{\beta}^0\| \geq T^{d_u+d_x} \log T$ *or* $\|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}^0\| \geq T^{-1/2+d_u} \log T\}$. *For any* $\delta \in (0,1), D > 0$ *and* $\epsilon > 0$ *the following statement holds when* T *is large:*

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

Proof. For the 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,

$$\overline{\Theta}_2 = \{(\boldsymbol{\beta}, \boldsymbol{\Sigma})\colon \|\boldsymbol{\beta}\| \leq a_1, \lambda_{\min}(\boldsymbol{\Sigma}) \geq a_2, \lambda_{\max}(\boldsymbol{\Sigma}) \leq a_3\},$$

where $\lambda_{\min}$ and $\lambda_{\max}$ denote the smallest and largest eigenvalues of $\boldsymbol{\Sigma}$ and the finite constants $a_1,\ a_2$ and a_3 are chosen in such a way that $(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0)$ is an inner point of $\overline{\boldsymbol{\Theta}}_2$. As explained we first show (A.7) with the second supremum taken over $S_T \cap \overline{\Theta}_2$ which is compact. We decompose the segmential log-likelihood as $\log \mathcal{L}(1,k;\boldsymbol{\beta},\boldsymbol{\Sigma}) = \mathcal{L}_{1,T} + \mathcal{L}_{2,T}$, where

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

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

$\boldsymbol{\beta}^* = \boldsymbol{\beta} - \boldsymbol{\beta}_0, \boldsymbol{\Sigma}^* = \boldsymbol{\Sigma} - \boldsymbol{\Sigma}_0, \boldsymbol{\eta}_t = (\boldsymbol{\Sigma}_0)^{-1}\boldsymbol{u}_t$ and $\Psi_T = (\boldsymbol{\Sigma}_0)^{-1/2}\boldsymbol{\Sigma}^*(\boldsymbol{\Sigma}_0)^{-1/2}$. We note that only $\mathcal{L}_{2,T}$ depends on $\boldsymbol{\beta}^*$. We split the parameter space $S_T = S_{1,T} \cup S_{2,T}$ where,

$$S_{1,T} = \{(\boldsymbol{\beta},\boldsymbol{\Sigma})\colon \|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_0\| \geq T^{-1/2+d_u}\log T,\ \boldsymbol{\beta} \text{ arbitrary}\}$$

and

$$S_{2,T} = \{(\boldsymbol{\beta},\boldsymbol{\Sigma})\colon \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \geq T^{d_u-d_x}\log T \text{ and } \|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_0\| \leq T^{-1/2+d_u}\log T\}.$$

It has to be shown that,

$$\Pr\left(\sup_{k\geq\lfloor T\delta\rfloor}\ \sup_{(\boldsymbol{\beta},\boldsymbol{\Sigma})\in S_{1,T}\cap\overline{\Theta}_2} T^D \mathcal{L}(1,k;\boldsymbol{\beta},\boldsymbol{\Sigma}) > 1\right) < \epsilon \tag{A.31}$$

and

$$\Pr\left(\sup_{k\geq\lfloor T\delta\rfloor}\sup_{(\boldsymbol{\beta},\boldsymbol{\Sigma})\in S_{2,T}\cap\overline{\Theta}_2} T^D\mathcal{L}(1,k;\boldsymbol{\beta},\boldsymbol{\Sigma})>1\right)<\epsilon. \tag{A.32}$$

We start to show (A.31). On $S_{1,T}$, $\mathcal{L}_{2,T}$ is a quadratic function of $\boldsymbol{\beta}^*$ and has maximum value,

$$\sup_{\boldsymbol{S}_{1,T}}\mathcal{L}_{2,T}=\frac{k}{2}\left(\frac{1}{k}\sum_{t=1}^{k}\boldsymbol{x}_t\boldsymbol{\Sigma}^{-1}\boldsymbol{u}_t\right)'\left(\frac{1}{k}\sum_{t=1}^{k}\boldsymbol{x}_t\boldsymbol{\Sigma}^{-1}\boldsymbol{x}_t'\right)^{-1}\left(\frac{1}{k}\sum_{t=1}^{k}\boldsymbol{x}_t\boldsymbol{\Sigma}^{-1}\boldsymbol{u}_t\right).$$

Applying Property 1 gives,

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

Additionally we see that,

$$\begin{aligned}\sup_{k\geq\lfloor T\delta\rfloor}\left\|\frac{1}{k}\sum_{t=1}^{k}\boldsymbol{x}_t\boldsymbol{\Sigma}^{-1}\boldsymbol{u}_t\right\| &= \sup_{k\geq\lfloor T\delta\rfloor}\left\|\frac{1}{k}\sum_{t=1}^{k}S'(\boldsymbol{I}_n\otimes\boldsymbol{z}_t)\boldsymbol{\Sigma}^{-1}\boldsymbol{u}_t\right\|\\ &= \sup_{k\geq\lfloor T\delta\rfloor}\left\|S'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{I}_n)\frac{1}{k}\sum_{t=1}^{k}(\boldsymbol{I}_n\otimes\boldsymbol{z}_t)\boldsymbol{u}_t\right\|\\ &\leq \sup_{k\geq\lfloor T\delta\rfloor}\left\|\frac{1}{k}\sum_{t=1}^{k}(\boldsymbol{I}_n\otimes\boldsymbol{z}_t)\boldsymbol{u}_t\right\|\left\|S'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{I}_n)\right\|.\end{aligned}$$

From the FCLT of Lemma A.2 we have for fixed $r>0$,

$$\lim_{T\to\infty}\Pr\left(\sup_{k\geq\lfloor T\delta\rfloor}\left\|\frac{1}{k}\sum_{t=1}^{k}(\boldsymbol{I}_n\otimes\boldsymbol{z}_t)\boldsymbol{u}_t\right\|>rT^{d_u-1/2}\log^{1/2}T\right)=0,$$

while $\|S'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{I}_n)\|=\sum_{i=1}^{n}(1+\lambda_i)^{-1}O_p(1)$, where λ_i $i=1,...,n$, are the eigenvalues of $(\boldsymbol{\Sigma}_0)^{-1/2}\boldsymbol{\Sigma}^*(\boldsymbol{\Sigma}_0)^{-1/2}$. Hence,

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

which implies

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

where $b_T=T^{d_u-1/2}\log T$ 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

$\boldsymbol{U}$ that diagonalizes $(\boldsymbol{I}+\boldsymbol{\Psi}_T)^{-1}$. Then we have,

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

Because $\|\boldsymbol{U}\|=1$ it suffices to investigate whether,

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

Then, for any $a>0$ by the FCLT of Lemma A.2 it follows that,

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

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

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

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,

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

It is now straightforward to see, using similar arguments as in Bai et al. (1998), that one can show that equation (A.32) holds. Therefore, the claim is shown on the compact parameter space $\overline{\Theta}_2$. Additionally as in Qu and Perron (2007) we can further conclude that the result is valid also on an unrestricted parameter space. Therefore, the proof is complete. □

Property 12 that follows is different from Qu and Perron (2007) in that we do not assume that the limit of $(h_Td_T^2)/T$ exists. Instead, as pointed out by Bai (2000), we assume the sufficient condition that $\liminf_{T\to\infty}(h_Td_T^2)/T\geq h>0$.

Property 12. *Let h_T and d_T be positive sequences such that h_T is nondecreasing, $d_T\to\infty$ and $\liminf_{T\to\infty}(h_Td_T^2)/T\geq h>0$. Define $\overline{\Theta}_3=\{(\boldsymbol{\beta},\boldsymbol{\Sigma})\colon\|\boldsymbol{\beta}\|\leq p_1,\lambda_{\min}(\boldsymbol{\Sigma})\geq p_2,\lambda_{\max}(\boldsymbol{\Sigma})\leq p_3\}$, where p_1,p_2 and p_3 are arbitrary constants that satisfy $p_1<\infty$, $0<p_2\leq p_3<\infty$. Define $S_T=\{(\boldsymbol{\beta},\boldsymbol{\Sigma})\colon\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\|\geq T^{d_u-d_x}\log T$ or $\|\boldsymbol{\Sigma}-\boldsymbol{\Sigma}^0\|\geq T^{-1/2+d_u}\log T\}$. Then, for any $\epsilon>0$,* *there exists an* $A > 0$*, such that when* T *is large,*

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

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

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

or

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

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

$$\sup_{S_{1,T}}\mathcal{L}_{2T}=\frac{k}{2}(\frac{1}{k}\sum_{t=1}^{k}\boldsymbol{x}_t\boldsymbol{\Sigma}^{-1}\boldsymbol{u}_t)'(\frac{1}{k}\sum_{t=1}^{k}\boldsymbol{x}_t\boldsymbol{\Sigma}^{-1}\boldsymbol{x}_t')^{-1}(\frac{1}{k}\sum_{t=1}^{k}\boldsymbol{x}_t\boldsymbol{\Sigma}^{-1}\boldsymbol{u}_t),$$

where

$$(\sum_{t=1}^{k}\boldsymbol{x}_t\boldsymbol{\Sigma}^{-1}\boldsymbol{x}_t')^{-1}=(\sum_{t=1}^{k}S'(\boldsymbol{I}\otimes\boldsymbol{z}_t)\boldsymbol{\Sigma}^{-1}(\boldsymbol{I}\otimes\boldsymbol{z}_t')S)^{-1}=(S'(\boldsymbol{\Sigma}^{-1}\otimes\sum_{t=1}^{k}\boldsymbol{z}_t\boldsymbol{z}_t')S)^{-1}.$$

From $l^{-1}\sum_{t=1}^{l}\boldsymbol{z}_t\boldsymbol{z}_t'\xrightarrow{a.s.}\boldsymbol{Q}_z$, for a given $\epsilon>0$ we can always find a $k_1>0$ such that,

$$P(\sup_{k\geq k_1}\|\frac{1}{k}\sum_{t=1}^{k}\boldsymbol{z}_t\boldsymbol{z}_t'-\boldsymbol{Q}_z\|>\epsilon)<\epsilon.$$

Define $\boldsymbol{Q}_\triangle := k^{-1}\sum_{t=1}^{k}\boldsymbol{z}_t\boldsymbol{z}_t'-\boldsymbol{Q}_z$. Then

$$\begin{aligned}&(S'(\boldsymbol{\Sigma}^{-1}\otimes\frac{1}{k}\sum_{t=1}^{k}\boldsymbol{z}_t\ \boldsymbol{z}_t')S)^{-1}-(S'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{Q}_z)S)^{-1}\\&=(S'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{Q}_z)S+S'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{Q}_\triangle)S)^{-1}-(S'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{Q}_z)S)^{-1}\\&=-\boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{A}+\boldsymbol{B})^{-1},\end{aligned}$$

where $\boldsymbol{A}=S'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{Q}_z)S$ and $\boldsymbol{B}=S'(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{Q}_\triangle)S$. Because $\boldsymbol{\Sigma}^{-1}$ has uniformly bounded eigenvalues and $k^{-1}\sum_{t=1}^{k}\boldsymbol{z}_t\boldsymbol{z}_t'$ is positive definite for large k, $\boldsymbol{A}^{-1}$ and $\boldsymbol{B}^{-1}$ have bounded eigenvalues. Because $\boldsymbol{B}$ is uniformly small, $-\boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{A}+\boldsymbol{B})^{-1}$ is uniformly small for large k. This is,

$$(S'(\boldsymbol{\Sigma}^{-1} \otimes k^{-1} \sum_{t=1}^{k} \boldsymbol{z}_t \boldsymbol{z}_t') S)^{-1} - (S'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{Q}_z) S)^{-1} \overset{\text{a.s.}}{=} o(1) \qquad \text{as} \quad k \to \infty.$$

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

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

Now,

$$\begin{aligned}
\sup_{k \geq Ah_T} \|\frac{1}{k} \sum_{t=1}^{k} \boldsymbol{x}_t \boldsymbol{\Sigma}^{-1} \boldsymbol{u}_t\| &= \sup_{k \geq Ah_T} \|\frac{1}{k} \sum_{t=1}^{k} S'(\boldsymbol{I}_n \otimes \boldsymbol{z}_t) \boldsymbol{\Sigma}^{-1} \boldsymbol{u}_t\| \\
&\leq \sup_{k \geq Ah_T} \|\frac{1}{k} \sum_{t=1}^{k} (\boldsymbol{I}_n \otimes \boldsymbol{z}_t) \boldsymbol{u}_t\| \|S'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\|. \qquad \text{(A.33)}
\end{aligned}$$

From Lemma A.1 we have,

$$P(\sup_{k \geq Ah_T} \|\frac{1}{k} \sum_{t=1}^{k} (I_n \otimes \boldsymbol{z}_t) \boldsymbol{u}_t\| > ab_T) \leq \frac{C_1}{Ah_T a^2 b_T} < \frac{2C_1}{Aa^2 h} \qquad \text{(A.34)}$$

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

$$\|S'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\| \leq n\boldsymbol{C}_2 \sum_{i=1}^{n} \frac{1}{1+\lambda_i} \qquad \text{(A.35)}$$

for some $0 < \boldsymbol{C}_2 < \infty$, which depends on the matrix S. Now, combining (A.33)-(A.35), we have, for any $\epsilon > 0$ that there exists an $\bar{A} > 0$, such that with probability no less than $1 - \epsilon$,

$$\sup_{k \geq \bar{A}h_T} \sup_{(\boldsymbol{\beta},\boldsymbol{\Sigma}) \in S_{1,T} \cap \bar{\Theta}_3} |L_{2T}| < ka^2 b_T^2 n^2 \boldsymbol{C}_2^2 M (\sum_{i=1}^{n} \frac{1}{1+\lambda_i})^2 \leq \frac{k}{2} \sum_{i=1}^{n} \frac{\boldsymbol{G} a^2 b_T^2}{1+\lambda_i} = \frac{k}{2} \sum_{i=1}^{n} \frac{\gamma^2 b_T^2}{1+\lambda_i}$$

with $\boldsymbol{G} = 2n^3 \boldsymbol{C}_2^2 M / p_2$. Because a^2 can be made arbitrarily small by choosing a large A, so can γ^2. Hence Property 12 follows. □

A.8 Proof of Theorem 5

Proof. The proof of Theorem 5 is very similar to the proof of Theorem 2 only that the limiting distribution changes. However, for the ease of reading we repeat the whole proof here.

We introduce some notation first. Let,

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

be the estimated covariance matrix using the full sample estimate of $\boldsymbol{\beta}_a$ obtained under the null hypothesis of no change and using the estimate of $\boldsymbol{\beta}_b$ based on data up to the last date of regime j, defined as

$$\tilde{\beta}_{b1,j} = (\sum_{t=1}^{T_j}\boldsymbol{x}_{bt}\tilde{\boldsymbol{\Sigma}}_{1,j}^{-1}\boldsymbol{x}'_{bt})^{-1}\sum_{t=1}^{T_j}\boldsymbol{x}_t\tilde{\boldsymbol{\Sigma}}_{1,j}^{-1}(\boldsymbol{y}_t - \boldsymbol{x}'_{at}\tilde{\boldsymbol{\beta}}_a).$$

Additionally,

$$\hat{\boldsymbol{\Sigma}}_j = \frac{1}{T_j - T_{j-1}}\sum_{t=T_{j-1}+1}^{T_j}(\boldsymbol{y}_t - \boldsymbol{x}'_{at}\hat{\boldsymbol{\beta}}_a - \boldsymbol{x}'_{bt}\hat{\boldsymbol{\beta}}_{bj})(\boldsymbol{y}_t - \boldsymbol{x}'_{at}\hat{\boldsymbol{\beta}}_a - \boldsymbol{x}'_{bt}\hat{\boldsymbol{\beta}}_{bj})',$$

is the estimate of the covariance matrix of the errors under the alternative hypothesis using the full sample estimate of $\boldsymbol{\beta}_a$ and using the estimate of $\boldsymbol{\beta}_b$ based on data from regime j only, that is,

$$\hat{\boldsymbol{\beta}}_{b,j} = (\sum_{t=T_{j-1}+1}^{T_j}\boldsymbol{x}_{bt}\hat{\boldsymbol{\Sigma}}_j^{-1}\boldsymbol{x}'_{bt})^{-1}\sum_{t=T_{j-1}+1}^{T_j}\boldsymbol{x}_t\hat{\boldsymbol{\Sigma}}_j^{-1}(\boldsymbol{y}_t - \boldsymbol{x}'_{at}\hat{\boldsymbol{\beta}}_a). \tag{A.36}$$

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

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

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

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

Applying this to the terms F_T^j,

$$\begin{aligned}
F_T^j :=& F_{1,T}^j + F_{2,T}^j \\
=& \operatorname{tr}(T_{j+1}(\boldsymbol{\Sigma}^0)^{-1}(\tilde{\boldsymbol{\Sigma}}_{1,j+1} - \boldsymbol{\Sigma}^0) - T_j(\boldsymbol{\Sigma}^0)^{-1}(\tilde{\boldsymbol{\Sigma}}_{1,j} - \boldsymbol{\Sigma}^0)) \qquad \text{(A.37)} \\
& - (T_{j+1} - T_j)(\boldsymbol{\Sigma}^0)^{-1}(\hat{\boldsymbol{\Sigma}}_{j+1} - \boldsymbol{\Sigma}^0)) \\
& - \frac{1}{2}\operatorname{tr}(T_{j+1}[(\boldsymbol{\Sigma}^0)^{-1}(\tilde{\boldsymbol{\Sigma}}_{1,j+1} - \boldsymbol{\Sigma}^0)]^2 \qquad \text{(A.38)} \\
& - T_j[(\boldsymbol{\Sigma}^0)^{-1}(\tilde{\boldsymbol{\Sigma}}_{1,j} - \boldsymbol{\Sigma}^0)]^2 - (T_{j+1} - T_j)[(\boldsymbol{\Sigma}^0)^{-1}(\hat{\boldsymbol{\Sigma}}_{j+1} - \boldsymbol{\Sigma}^0)]^2).
\end{aligned}$$

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

$$\boldsymbol{Y} = \boldsymbol{X}_a\boldsymbol{\beta}_a + \boldsymbol{X}_b\boldsymbol{\beta}_b + \boldsymbol{U}$$

with $E(\boldsymbol{U}\boldsymbol{U}') = \boldsymbol{I}_T \otimes \boldsymbol{\Sigma}^0$. If only data up to the last date of regime j are included, we have,

$$\boldsymbol{Y}_{1,j} = \boldsymbol{X}_{a1,j}\boldsymbol{\beta}_a + \boldsymbol{X}_{b1,j}\boldsymbol{\beta}_{b1,j} + \boldsymbol{U}_{1,j}.$$

We now define $\boldsymbol{Y}_{1,j}^d = (\boldsymbol{I}_T \otimes \tilde{\boldsymbol{\Sigma}}_{1,j}^{-1/2})\boldsymbol{Y}_{1,j}$, $\boldsymbol{W}_{1,j} = (\boldsymbol{I}_T \otimes \tilde{\boldsymbol{\Sigma}}_{1,j}^{-1/2})\boldsymbol{X}_{a1,j}$, $\boldsymbol{Z}_{1,j} = (\boldsymbol{I}_T \otimes \tilde{\boldsymbol{\Sigma}}_{1,j}^{-1/2})\boldsymbol{X}_{b1,j}$ and $\boldsymbol{U}_{1,j}^d = (\boldsymbol{I}_T \otimes \tilde{\boldsymbol{\Sigma}}_{1,j}^{-1/2})\boldsymbol{U}_{1,j}$. Then, omitting the subscript when the full sample is used, we have

$$\tilde{\boldsymbol{\beta}}_a = [\boldsymbol{W}'\boldsymbol{M}_{\boldsymbol{Z}}\boldsymbol{W}]^{-1}\boldsymbol{W}'\boldsymbol{M}_{\boldsymbol{Z}}\boldsymbol{Y}^d, \qquad \text{(A.39)}$$

$$\tilde{\boldsymbol{\beta}}_{b1,j} = (\boldsymbol{Z}_{1,j}'\boldsymbol{Z}_{1,j})^{-1}\boldsymbol{Z}_{1,j}'(\boldsymbol{Y}_{1,j}^d - \boldsymbol{W}_{1,j}\tilde{\boldsymbol{\beta}}_a), \qquad \text{(A.40)}$$

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

$$\boldsymbol{Y}_{j+1} = \boldsymbol{X}_{a,j+1}\boldsymbol{\beta}_a + \boldsymbol{X}_{b,j+1}\boldsymbol{\beta}_{b,j+1} + \boldsymbol{U}_{j+1}.$$

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

$$\hat{\boldsymbol{\beta}}_a = [\bar{\boldsymbol{W}}'\boldsymbol{M}_{\bar{Z}}\bar{\boldsymbol{W}}]^{-1}\bar{\boldsymbol{W}}'\boldsymbol{M}_{\bar{Z}}\bar{\boldsymbol{Y}}^d, \tag{A.41}$$

$$\hat{\boldsymbol{\beta}}_{b,j+1} = (\bar{\boldsymbol{Z}}'_{j+1}\bar{\boldsymbol{Z}}_{j+1})^{-1}\bar{\boldsymbol{Z}}'_{j+1}(\bar{\boldsymbol{Y}}^d_{j+1} - \bar{\boldsymbol{W}}_{j+1}\hat{\boldsymbol{\beta}}_a). \tag{A.42}$$

Note that the choice of the estimate of the covariance matrix in (A.39) to (A.42) 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,

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

where $\boldsymbol{A}_T = [\boldsymbol{W}'\boldsymbol{M}_Z\boldsymbol{W}]^{-1}\boldsymbol{W}'\boldsymbol{M}_Z U^d$. For the third component of $F^j_{1,T}$ it can be shown that

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

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

$$F^j_{1,T} = U^{d'}_{1,j}P_{Z_{1,j}}U^d_{1,j} + U^{d'}_{j+1}P_{\bar{Z}_{j+1}}U^d_{j+1} - U^{d'}_{1,j+1}P_{Z_{1,j+1}}U^d_{1,j+1} + o_p(1). \tag{A.43}$$

Now, $T^{-d_u}\boldsymbol{Z}'_{1,j}\boldsymbol{U}^d_{1,j} \Rightarrow \boldsymbol{Q}^{1/2}_b\boldsymbol{\Xi}^*_{x,u,p_b}(\lambda_i)$ and $T^{-1}\sum_{t=1}^{T_j}\boldsymbol{x}_{bt}(\boldsymbol{\Sigma}^0)^{-1}\boldsymbol{x}'_{bt} \to^p \lambda_i\boldsymbol{Q}_b$ where $\boldsymbol{\Xi}^*_{x,u,p_b}(\lambda_i)$

is a p_b vector of zeros and independent integrals of fractional Wiener processes defined on $[0,1]$ as given in Theorem 2 and where $\boldsymbol{Q}_b$ is the appropriate submatrix of $\boldsymbol{Q}$ that corresponds to $\int_0^1 \boldsymbol{x}_{bt}\boldsymbol{x}_{bt}' dt$. Define the projection matrix $\boldsymbol{P}_z = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$. Hence,

$$T^{-2d_u}\, \boldsymbol{U}_{1,j+1}^{d'} \boldsymbol{P}_{Z_{1,j+1}} \boldsymbol{U}_{1,j+1}^{d} \Rightarrow [\boldsymbol{\Xi}_{x,u,p_b}^{*}(\lambda_{j+1})' \boldsymbol{\Xi}_{x,u,p_b}^{*}(\lambda_{j+1})]/\lambda_{j+1}. \tag{A.44}$$

Using similar arguments

$$T^{-2d_u}\, \boldsymbol{U}_{1,j}^{d'} \boldsymbol{P}_{Z_{1,j}} \boldsymbol{U}_{1,j}^{d} \Rightarrow [\boldsymbol{\Xi}_{x,u,p_b}^{*}(\lambda_j)' \boldsymbol{\Xi}_{x,u,p_b}^{*}(\lambda_j)]/\lambda_j \tag{A.45}$$

and

$$\begin{aligned} &T^{-2d_u}\, \boldsymbol{U}_{j+1}^{d'} \boldsymbol{P}_{\bar{\boldsymbol{Z}}_{j+1}} \boldsymbol{U}_{j+1}^{d} \\ &\qquad \Rightarrow (\boldsymbol{\Xi}_{x,u,p_b}^{*}(\lambda_{j+1}) - \boldsymbol{\Xi}_{x,u,p_b}^{*}(\lambda_j))'(\boldsymbol{\Xi}_{x,u,p_b}^{*}(\lambda_{j+1}) - \boldsymbol{\Xi}_{x,u,p_b}^{*}(\lambda_j))/(\lambda_{j+1} - \lambda_j). \end{aligned} \tag{A.46}$$

These results imply that the first component in (A.37) has the limit

$$F_{1,T}^{j} \Rightarrow \frac{(\lambda_j \boldsymbol{\Xi}_{x,u,p_b}^{*}(\lambda_{j+1}) - \lambda_{j+1}\boldsymbol{\Xi}_{x,u,p_b}^{*}(\lambda_j))'(\lambda_j \boldsymbol{\Xi}_{x,u,p_b}^{*}(\lambda_{j+1}) - \lambda_{j+1}\boldsymbol{\Xi}_{x,u,p_b}^{*}(\lambda_j))}{(\lambda_{j+1} - \lambda_j)\lambda_j\lambda_{j+1}}. \tag{A.47}$$

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

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

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

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

where $\boldsymbol{\Sigma}^{ik}$ is the (i,k) element of $(\boldsymbol{\Sigma}^0)^{-1}$ and $I_{i,k}$ is the (i,k) element of I.

Also let $((\boldsymbol{\Sigma}^0)^{-1}\tilde{\boldsymbol{\Sigma}}_{1,j+1} - I)^S$ ("S" for relevant segments) be the matrix whose entries are those of $((\boldsymbol{\Sigma}^0)^{-1}\tilde{\boldsymbol{\Sigma}}_{1,j+1} - I)$ for the corresponding entries of $\boldsymbol{\Sigma}^0$ that are allowed to vary across regimes, the remaining entries being filled with zeros. Then,

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

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

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

and, in view of (A.38),

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

Now, because $\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0 = O_p(T^{d_u+d_x})$, we have,

$$\begin{aligned}
&\frac{T_{j+1}}{T^{2d_u}}((\mathbf{\Sigma}^0)^{-1}\tilde{\mathbf{\Sigma}}_{1,j+1} - I)^S((\mathbf{\Sigma}^0)^{-1}\tilde{\mathbf{\Sigma}}_{1,j+1} - I)^S \\
&\quad = \frac{T}{T_{j+1}}\left(T^{-1/2-d_u}\sum_{t=1}^{T_{j+1}}[(\mathbf{\Sigma}^0)^{-1}\boldsymbol{u}_t\boldsymbol{u}_t' - I]\right)^S \left(T^{-1/2-d}\sum_{t=1}^{T_{j+1}}[(\mathbf{\Sigma}^0)^{-1}\boldsymbol{u}_t\boldsymbol{u}_t' - I]\right)^S + o_p(1) \\
&\quad \Rightarrow \frac{\boldsymbol{\xi}_n^d(\lambda_{j+1})^S\boldsymbol{\xi}_n^d(\lambda_{j+1})^S}{\lambda_{j+1}} \\
&\frac{T_j}{T^{2d_u}}((\mathbf{\Sigma}^0)^{-1}\tilde{\mathbf{\Sigma}}_{1,j} - I)^S((\mathbf{\Sigma}^0)^{-1}\tilde{\mathbf{\Sigma}}_{1,j} - I)^S \\
&\quad = \frac{T}{T_j}\left(T^{-1/2-d_u}\sum_{t=1}^{T_j}[(\mathbf{\Sigma}^0)^{-1}\boldsymbol{u}_t\boldsymbol{u}_t' - I]\right)^S \left(T^{-1/2-d_u}\sum_{t=1}^{T_j}[(\mathbf{\Sigma}^0)^{-1}\boldsymbol{u}_t\boldsymbol{u}_t' - I]\right)^S + o_p(1) \\
&\quad \Rightarrow \frac{\xi_n^d(\lambda_j)^S\boldsymbol{\xi}_n^d(\lambda_j)^S}{\lambda_j}
\end{aligned}$$

and

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

where $\boldsymbol{\xi}_D^*(\cdot)$ is an $n \times n$ matrix whose elements are

$$
[\boldsymbol{\xi}_D^*(\cdot)]_{i,j} = \begin{cases} [\boldsymbol{\xi}_D(\cdot)]_{i,j}, & \text{if } d_i = d_j = \max_{1\le k\le n} d_k, \\ 0, & \text{else,} \end{cases} \tag{A.48}
$$

and where $\boldsymbol{\xi}_D$ is (nonstandard) fractional Brownian motion defined on $[0,1]$ such that $\mathrm{Var}(\mathrm{vec}(\boldsymbol{\xi}_D(1))) = \boldsymbol{\Omega}$ (which follows from Theorem 4.8.2 of Giraitis et al. (2012) p.109). Hence,

$$
\begin{aligned}
\sum_{j=1}^{m}\boldsymbol{F}_{2,T}^j \Rightarrow & -\frac{1}{2}\operatorname{tr}\left(\frac{\boldsymbol{\xi}_{D,n}^*(\lambda_{j+1})^S\boldsymbol{\xi}_{D,n}^*(\lambda_{j+1})^S}{\lambda_{j+1}} - \frac{\boldsymbol{\xi}_{D,n}^*(\lambda_j)^S\boldsymbol{\xi}_{D,n}^*(\lambda_j)^S}{\lambda_j}\right. \\
&\left. + \frac{(\boldsymbol{\xi}_{D,n}^*(\lambda_{j+1}) - \boldsymbol{\xi}_{D,n}^*(\lambda_j))^S(\boldsymbol{\xi}_{D,n}^*(\lambda_{j+1}) - \boldsymbol{\xi}_{D,n}^*(\lambda_j))^S}{\lambda_{j+1}-\lambda_j}\right) \\
= & -\frac{1}{2}\left[\frac{\mathrm{vec}(\boldsymbol{\xi}_{D,n}^*(\lambda_{j+1})^S)'\,\mathrm{vec}(\boldsymbol{\xi}_{D,n}^*(\lambda_{j+1})^S)}{\lambda_{j+1}}\right. \\
& - \frac{\mathrm{vec}(\boldsymbol{\xi}_{D,n}^*(\lambda_j)^S)'\,\mathrm{vec}(\boldsymbol{\xi}_{D,n}^*(\lambda_j)^S)}{\lambda_j} \\
& \times \left(\mathrm{vec}(\boldsymbol{\xi}_{D,n}^*(\lambda_{j+1})^S) - \mathrm{vec}(\boldsymbol{\xi}_{D,n}^*(\lambda_j)^S)\right)' \\
& \left. \times \frac{(\mathrm{vec}(\boldsymbol{\xi}_{D,n}^*(\lambda_{j+1})^S) - \mathrm{vec}(\boldsymbol{\xi}_{D,n}^*(\lambda_j)^S)}{(\lambda_{i+1}-\lambda_i)}\right]
\end{aligned} \tag{A.49}
$$

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

$$
\begin{aligned}
\mathrm{vec}(\boldsymbol{\xi}_{D,n}^*(\lambda_{j+1})^S)'\,\mathrm{vec}(\boldsymbol{\xi}_{D,n}^*(\lambda_{j+1})^S) &= \mathrm{vec}(\boldsymbol{\xi}_{D,n}^*(\lambda_{j+1}))'\boldsymbol{H}'\boldsymbol{H}\,\mathrm{vec}(\boldsymbol{\xi}_{D,n}^*(\lambda_{j+1})) \\
&\overset{d}{=} \boldsymbol{W}_{D,n_b^*}^*(\lambda_{j+1})'\boldsymbol{H}\boldsymbol{\Omega}\boldsymbol{H}'\boldsymbol{W}_{D,n_b^*}^*(\lambda_{j+1}),
\end{aligned}
$$

where $\boldsymbol{W}^*_{D,n_b^*}$ is an n_b^* vector of processes as defined in Theorem 2. Hence, we have

$$\begin{aligned}\sum_{j=1}^{m}\boldsymbol{F}_{2,T}^{j} \Rightarrow & -\frac{1}{2}\Bigg[\frac{\boldsymbol{W}^*_{D,n_b^*}(\lambda_{j+1})'\boldsymbol{H}'\boldsymbol{\Omega}\boldsymbol{H}\boldsymbol{W}^*_{D,n_b^*}(\lambda_{j+1})}{\lambda_{j+1}} - \frac{\boldsymbol{W}^*_{D,n_b^*}(\lambda_j)'\boldsymbol{H}'\boldsymbol{\Omega}H\boldsymbol{W}^*_{D,n_b^*}(\lambda_j)}{\lambda_j} \\ & - \frac{(\boldsymbol{W}^*_{D,n_b^*}(\lambda_{j+1}) - \boldsymbol{W}^*_{D,n_b^*}(\lambda_j))'\boldsymbol{H}'\boldsymbol{\Omega}\boldsymbol{H}(\boldsymbol{W}^*_{D,n_b^*}(\lambda_{j+1}) - \boldsymbol{W}^*_{D,n_b^*}(\lambda_j)}{\lambda_{j+1}-\lambda_j}\Bigg] \\ =&(\lambda_j\boldsymbol{W}^*_{D,n_b^*}(\lambda_{j+1}) - \lambda_{j+1}\boldsymbol{W}^*_{D,n_b^*}(\lambda_j))'\boldsymbol{H}'\boldsymbol{\Omega}\boldsymbol{H} \\ & \times(\lambda_j\boldsymbol{W}^*_{D,n_b^*}(\lambda_{j+1}) - \lambda_{j+1}\boldsymbol{W}^*_{D,n_b^*}(\lambda_j))/(\lambda_j\lambda_{j+1}(\lambda_{j+1}-\lambda_j)). \end{aligned} \tag{A.50}$$

By combining equations (A.47) and (A.50) we have shown the limiting distribution of our test.

□