Testing for Multiple Structural Breaks in Multivariate Long Memory Time Series

Philipp Sibbertsen[a*] Kai Wenger[a] Simon Wingert[a]

[a]Leibniz University Hannover

September 28, 2020

Abstract

This paper considers estimation and testing of multiple breaks that occur at unknown dates in multivariate long-memory time series. We propose a likelihood ratio based approach for estimating breaks in the mean and the covariance of a system of long-memory time series. The limiting distribution of these estimates as well as consistency of the estimators is derived. A testing procedure to determine the unknown number of break points is given based on iterative testing on the regression residuals. A Monte Carlo exercise shows the finite sample performance of our method. An empirical application to inflation series illustrates the usefulness of our procedures.

Key words: Multivariate Long Memory · Multiple Structural Breaks · Hypothesis Testing.

JEL classification: C12, C22, C58, G15

1 Introduction

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

*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

instance in Perron (2006).

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

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

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

The validity of this approach in finite samples is shown in a Monte Carlo study, while the applicability in practice is demonstrated in an empirical example where we examine a system of inflation series.

The rest of the paper is organized as follows. In Section 2 we provide the model and our assumptions. Section 3 contains the estimators for the break points and Section 4 provides the testing procedure. Section 5 contains of the Monte Carlo study and Section 6 illustrates the empirical example before Section 7 concludes. All proofs are gathered in the appendix.

2 The Model and Assumptions

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

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

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

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

$$y_t = (I \otimes z_t')S\beta_j + u_t, \tag{2}$$

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

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

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

We rewrite Equation (2) in order to lighten the notation by introducing the $p \times n$ matrix x_t that is defined by $x_t' = (I \otimes z_t')S$. This gives

$$y_t = x_t'\beta_j + u_t \tag{3}$$

for $T_{j-1} + 1 \leq t \leq T_j$ ($j = 1, \ldots, m+1$). Furthermore, we rewrite Equation (3) using matrix notation. To this end let $Y = (y_1', \ldots, y_T')'$ be the nT vector of dependent variables, $U = (u_1', \ldots, u_T')'$ the error vector and let the $nT \times p$ matrix of regressors be $X = (x_1, \ldots, x_T)'$. Now form the block partition $\bar{X}$ of the matrix X: For a given partition of the sample using the breaks $(T_1, \ldots, T_m)$ we define $\bar{X}$ as the $nT \times p(m+1)$ matrix $\bar{X} = \text{diag}(X_1, \ldots, X_{m+1})$, where $X_j (j = 1, \ldots, m+1)$ is the $n(T_j - T_{j-1}) \times p$ subset of X that corresponds to observations in regime j. Similarly we define the subvector U_j of U. Using these symbols we can express the regression system (3) as $Y = \bar{X}\beta + U$. In the following the true values of the parameters are denoted with a 0 superscript, e.g. the data generating process is given by $Y = \bar{X}^0\beta^0 + U$. Here, the term $\bar{X}^0$ is the diagonal partition of X using the partition of true break dates $(T_1^0, \ldots, T_m^0)$. We impose the following set of assumptions. Note that the assumptions are similar to those of Qu and Perron (2007) with the difference that we allow the errors u_t to be long-range dependent. However, we assume that the memory is only introduced through the errors u_t and thus that the regressors x_t are mostly short-range dependent such that the process $x_t u_t$ is of the same order of integration as u_t. This is a simplyfying assumption which is not necessary. Technically it would be possible to allow also for long-memory regressors with an order of integration smaller or equal than one as long as they are independent of the errors. However, this would complicate the proof of our property five below and furthermore lead to identification problems in practice as the order of integration of the observation would be determined by the maximum of the integration order of the regressors and errors. As the proof of our property five in the case of integrated regressors rely on a correct differencing of the regressors our procedure may become infeasible in practice.

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

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

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

Assumption 4. It holds that

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

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

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

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

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

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

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

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

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

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

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

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

$\Omega = \text{Var}(\text{vec}(\xi_D(1)))$. Also assume that $E[\eta_{tk}\eta_{tl}\eta_{th}] = 0$ for all k, l, h and for every t.

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

3 Estimation of the Break Dates and Model Parameters

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

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

where

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

The quasi-likelihood ratio statistic is given by

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

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

$$RLR_T(\mathcal{T}, \beta, \Sigma) = LR_T(\mathcal{T}, \beta, \Sigma) + \lambda^{'} g(\beta, \text{vec}(\Sigma)). \tag{4}$$

We need one further assumption about the minimal regime length.

Assumption 11. The maximization of the objective function (4) is taken over all partitions $\mathcal{T} = (T_1, \ldots, T_m) = (T\lambda_1, \ldots, T\lambda_m)$ for some $\varepsilon > 0$ in the set

$$\Lambda_\varepsilon = \{(\lambda_1, \ldots, \lambda_m) \colon |\lambda_{j+1} - \lambda_j| \geq \varepsilon, \lambda_1 \geq \varepsilon, \lambda_m \leq 1 - \varepsilon\}. \tag{5}$$

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

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

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

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

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

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

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

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

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

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

$$\frac{\Delta_{1,j}^2}{\Gamma_{1,j}^2} T^{1-2d} \nu_T^2 (\hat{T}_j - T_j^0) \Rightarrow \begin{cases} -\frac{|u|}{2} + W_{j,d}(u), & \text{for } u \le 0 \\ -\frac{|u|}{2}\frac{\Delta_{2,j}}{\Delta_{1,j}} + \frac{\Gamma_{2,j}}{\Gamma_{1,j}} W_{j,d}(u), & \text{for } u > 0, \end{cases} \tag{6}$$

where

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

and

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

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

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

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

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

4 Testing for Multiple Breaks in Multivariate Time Series

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

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

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

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

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

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

First, we introduce a likelihood ratio test of no break versus the alternative hypothesis of precisely

m breaks under long memory, i.e.

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

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

$$\begin{aligned} \frac{1}{T^{2d}} \sup LR_T(m, p_b, n_{bd}, n_{bo}, \varepsilon) &= \frac{1}{T^{2d}} \sup_{(\lambda_1,\ldots,\lambda_m)\in\Lambda_\varepsilon} 2\left[\log \hat{L}_T(T_1, \ldots, T_m) - \log \tilde{L}_T\right] \\ &= \frac{2}{T^{2d}}[\log \hat{L}_T(\hat{T}_1, \ldots, \hat{T}_m) - \log \tilde{L}_T], \end{aligned} \tag{7}$$

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

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

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

with

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

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

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

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

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

The second test statistic tests the null hypothesis of no break against the alternative of m breaks,

$1 \leq m \leq M$ for some upper bound M, i.e.

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

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

$$UDmax\, LR_T(m, p_b, n_{bd}, n_{bo}, \varepsilon) = \max_{1 \leq m \leq M} \sup LR_T(m, p_b, n_{bd}, n_{bo}, \varepsilon). \tag{8}$$

The asymptotic distribution for this test statistic can be obtained in the setting of Theorem 2. We have

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

Critical values for different values of d and m are given in Tables 4 to 6 in the appendix.

For our iterative procedure below it is also essential to mention that the UDmax test enjoys pitman efficiency. This follows directly by noting that the tests are likelihood ratio type tests and applying the usual Taylor expansion argument to derive consistency of likelihood ratio tests also delivers the result in our set-up.

Now we are in the position to introduce an iterative method that can be used to determine the unknown number of breaks in a multivariate system of long-memory time series. It is inspired by applying the $UDmax\, LR_T$ test in (8) repeatedly. Therefore, the method requires fixing an upper bound on the number of breaks M in advance. It is a residual based iterative procedure, so that we shorten it as REBIT. It proceeds as follows:

(1) Set $m = 0$.

(2) Estimate m breaks in the original system of time series y_t and save the residuals.

(3) Conduct the $UDmax\, LR_T$ test with $H_0 \colon l = 0$ vs. $H_1 \colon 1 \leq l \leq M - m$ on the residuals.

(4a) If the test rejects and $m < (M-1)$: set $m = m + 1$ and reiterate from (2).

(4b) If the test cannot reject: the detected number of breaks is m. Furthermore, if $m = M - 1$: the number of breaks is greater or equal than the previously chosen upper bound M.

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

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

Note that the estimation of the break dates in step (2) is always performed on the original time series. That is the residuals are always estimated from a global optimization. Hence, the estimated break dates from different iterations do not depend on each other. Therefore, our procedure avoids the usually problematic situation of using residuals of residuals. From Lemma 1 we thus obtain consistency of our break point estimates in each step. The break point estimates in underspecified models are consistent as has been shown by Bai (1997a) and Bai and Perron (1998) for breaks in the mean. By similar methods one could therefore show that our procedure estimates some true break points if the number of true break points is underspecified.

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

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

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

Theorem 3. *Let α be the significance level of the break point test in Theorem 2. Under Assumptions 1-11 the REBIT procedure has a hit rate of $(1-\alpha)\%$.*

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

5 Simulation results

We conduct two Monte Carlo simulation studies to examine the finite sample properties of the break point estimator and our proposed REBIT procedure. We consider a bivariate model of fractionally integrated white noise processes

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

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

with long-memory parameters chosen as $D = (d_1, d_2) = \{(0,0), (0.2,0), (0.2,0.2), (0.4,0), (0.2,0.4), (0.39,0.4), (0.4,0.4)\}$. To estimate D, we apply the multivariate local Whittle estimator by Shimotsu (2007) using a bandwidth of $\lfloor T^{2/3} \rfloor$. To test the procedure the nominal significance level is $\alpha = 5\%$, we choose $\epsilon = 0.05$, and $M = 1,000$ replications.

We choose $m = 0, 1, 2, 3$ breaks which are uniformly allocated to the two series such that the distance between the breaks across both series is the same. Whether the breaks are positive or negative is randomly chosen. We make the break size dependent on the memory parameter as follows

$$\beta = \kappa T^{d-1/2}, \tag{9}$$

where β is the break size, $d = \max(d_1, d_2)$, and κ is a finite constant.

In Tables 1 and 2 we report the bias resp. the MSE of the estimator if the number of breaks m is known. We observe that increasing κ, i.e. the break size, or σ, i.e. the correlation in the bivariate system, decreases bias and MSE.

		Breaks	1		2		3	
d_1	d_2	σ / κ	3	6	3	6	3	6
0	0	0	0.000 696 10	0.000 741 50	0.106 608 00	0.063 163 25	0.085 305 86	0.054 599 75
		.5	0.000 175 70	0.002 811 20	0.099 840 33	0.050 717 58	0.081 558 18	0.048 461 56
.2	0	0	0.001 078 60	0.000 040 50	0.001 394 58	0.000 123 47	0.011 312 72	0.000 257 08
		.5	0.000 791 60	0.000 038 20	0.001 114 13	0.000 229 42	0.003 724 34	0.000 082 71
.4	0	0	0.000 019 30	0.000 011 60	0.000 049 40	0.000 005 47	0.000 042 55	0.000 010 97
		.5	0.000 007 30	0.000 003 80	0.000 024 57	0.000 009 88	0.000 007 14	0.000 008 01
.2	.2	0	0.000 551 00	0.000 004 80	0.002 377 49	0.000 344 37	0.011 612 07	0.000 376 60
		.5	0.000 586 10	0.000 026 60	0.001 655 75	0.000 115 98	0.003 805 34	0.000 133 23
.4	.2	0	0.000 013 40	0.000 000 90	0.000 034 23	0.000 006 97	0.000 019 87	0.000 009 67
		.5	0.000 017 40	0.000 003 10	0.000 011 36	0.000 003 32	0.000 029 16	0.000 008 34

Table 1: Bias of the suggested estimator.

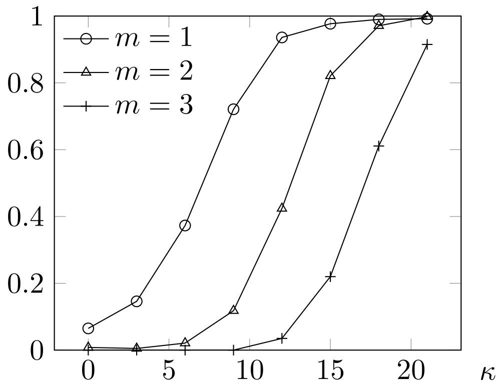
		Breaks	1		2		3	
d_1	d_2	σ / κ	3	6	3	6	3	6
0	0	0	0.086 847 57	0.044 528 43	0.123 645 93	0.082 614 74	0.130 690 19	0.087 243 98
		.5	0.080 877 34	0.031 975 87	0.115 343 19	0.069 150 92	0.122 001 24	0.079 260 33
.2	0	0	0.002 929 25	0.000 091 31	0.013 965 03	0.000 204 09	0.032 434 71	0.000 346 61
		.5	0.001 287 93	0.000 049 59	0.005 817 66	0.000 108 77	0.017 118 46	0.000 174 99
.4	0	0	0.000 005 02	0.000 000 26	0.000 010 77	0.000 000 51	0.000 016 25	0.000 000 77
		.5	0.000 002 91	0.000 000 12	0.000 005 51	0.000 000 25	0.000 008 58	0.000 000 37
.2	.2	0	0.003 393 87	0.000 094 32	0.013 845 58	0.000 206 51	0.031 960 04	0.000 373 27
		.5	0.001 331 91	0.000 049 85	0.005 360 97	0.000 108 77	0.016 500 71	0.000 181 54
.4	.2	0	0.000 004 87	0.000 000 24	0.000 011 32	0.000 000 51	0.000 016 52	0.000 000 79
		.5	0.000 002 91	0.000 000 12	0.000 005 77	0.000 000 26	0.000 009 38	0.000 000 37

Table 2: MSE of the suggested estimator.

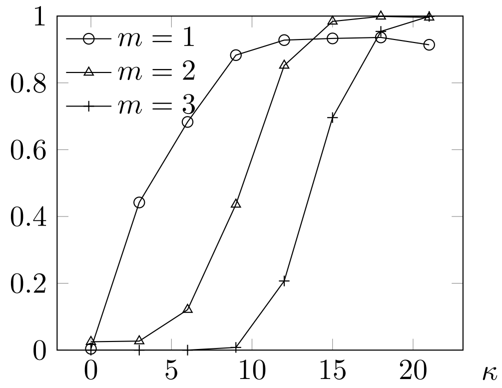
The asymptotic critical values we use are simulated for different combinations of $D = (d_1,d_2)$ by approximating the stochastic integrals by partial sums and can be found in Table 4 to 6 in our online appendix. They are based on 10,000 Monte Carlo replications with $1,000$ increments per path of the fractional Brownian motion. We obtain the corresponding critical value for an estimated value that is between two simulated d-values by linear interpolation between these two values.

In Figure 1 we report the hit ratio, i.e. how often our procedure detects the true number of breaks, dependent on κ with $T = 1,000$. The case of $m = 0$ is implicitly given for $\kappa = 0$.

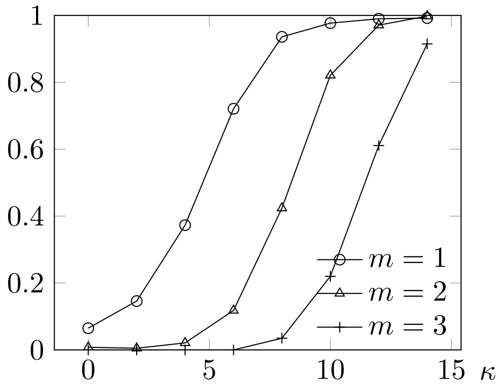
First, we observe that in all cases we obtain a hit ratio smaller than 5% (our nominal significance level) when $\kappa = 0$, i.e. we have no breaks in the series. For all combinations of $D = (d_1, d_2)$ we observe that the hit ration increases as the break size increases. Additionally, we see in all graphs that the more breaks we have the larger their size needs to be to obtain a higher hit ratio.



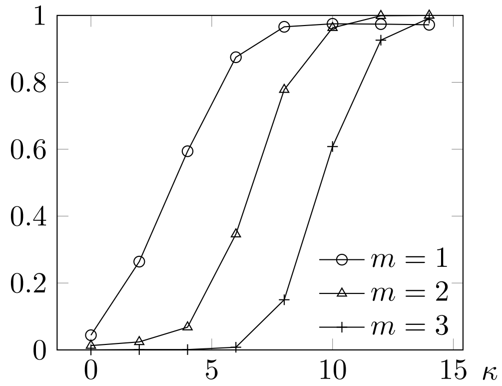
(a) $d_1 = 0$, $d_2 = 0$, Correlation $\rho = 0$



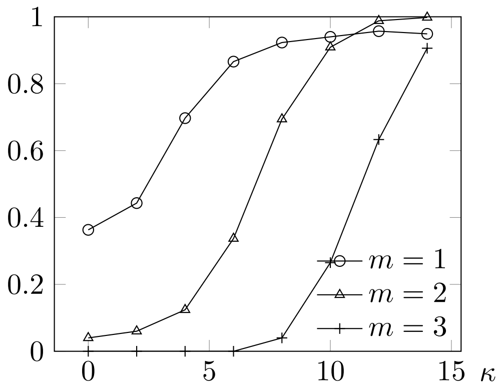
(b) $d_1 = 0$, $d_2 = 0$, Correlation $\rho = 0.5$



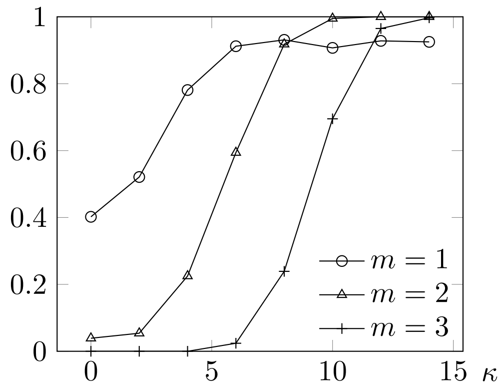
(c) $d_1 = 0.2$, $d_2 = 0$, Correlation $\rho = 0$



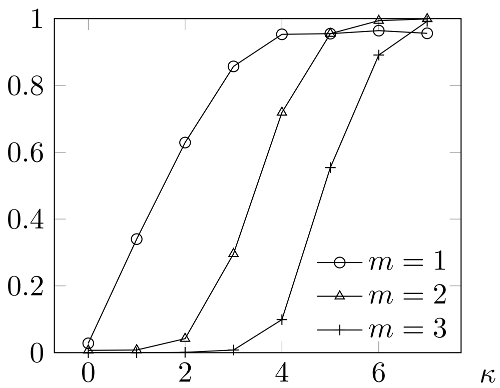
(d) $d_1 = 0.2$, $d_2 = 0$, Correlation $\rho = 0.5$



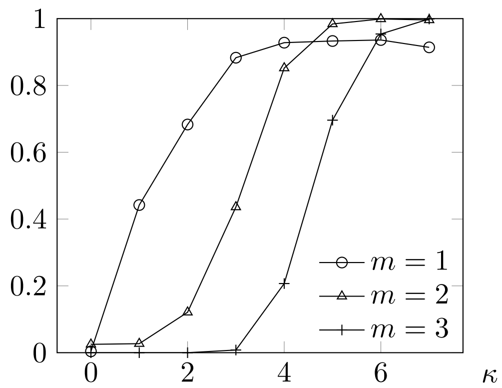
(e) $d_1 = 0.2$, $d_2 = 0.2$, Correlation $\rho = 0$



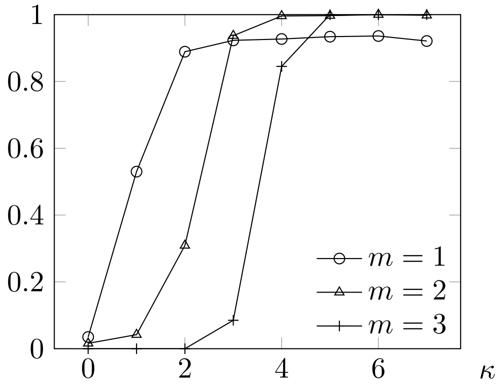
(f) $d_1 = 0.2$, $d_2 = 0.2$, Correlation $\rho = 0.5$



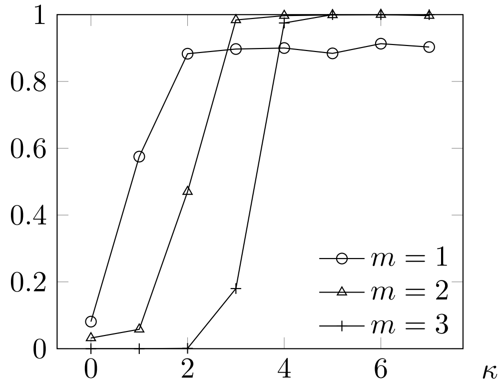
(g) $d_1 = 0.4$, $d_2 = 0$, Correlation $\rho = 0$



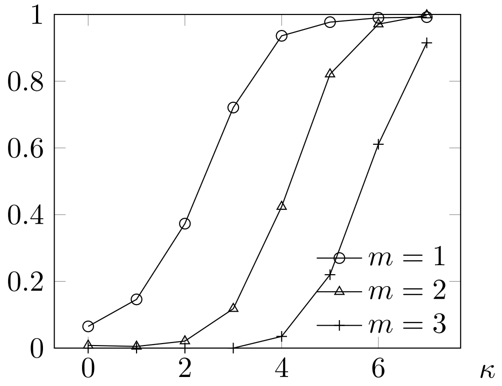
(h) $d_1 = 0.4$, $d_2 = 0$, Correlation $\rho = 0.5$



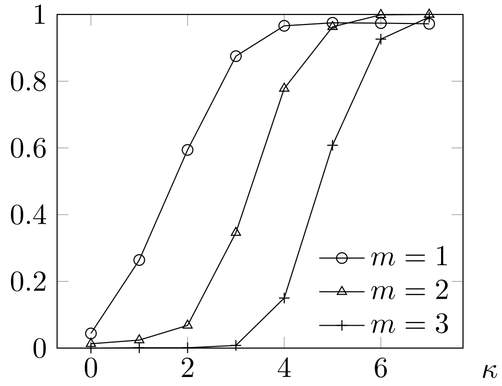
(i) $d_1 = 0.2$, $d_2 = 0.4$, Correlation $\rho = 0$



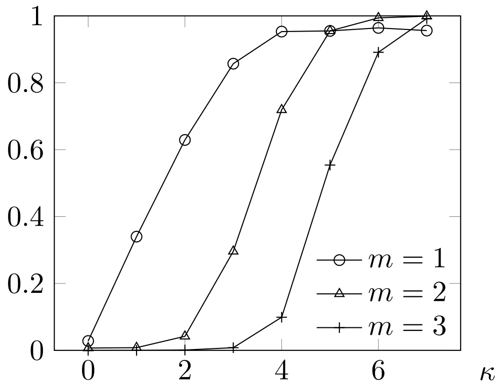
(j) $d_1 = 0.2$, $d_2 = 0.4$, Correlation $\rho = 0.5$



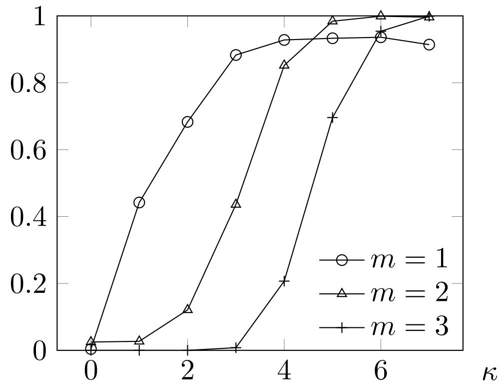
(k) $d_1 = 0.4$, $d_2 = 0.4$, Correlation $\rho = 0$



(l) $d_1 = 0.4$, $d_2 = 0.4$, Correlation $\rho = 0.5$



(m) $d_1 = 0.39$, $d_2 = 0.4$, Correlation $\rho = 0$



(n) $d_1 = 0.39$, $d_2 = 0.4$, Correlation $\rho = 0.5$

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

6 Empirical Application

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

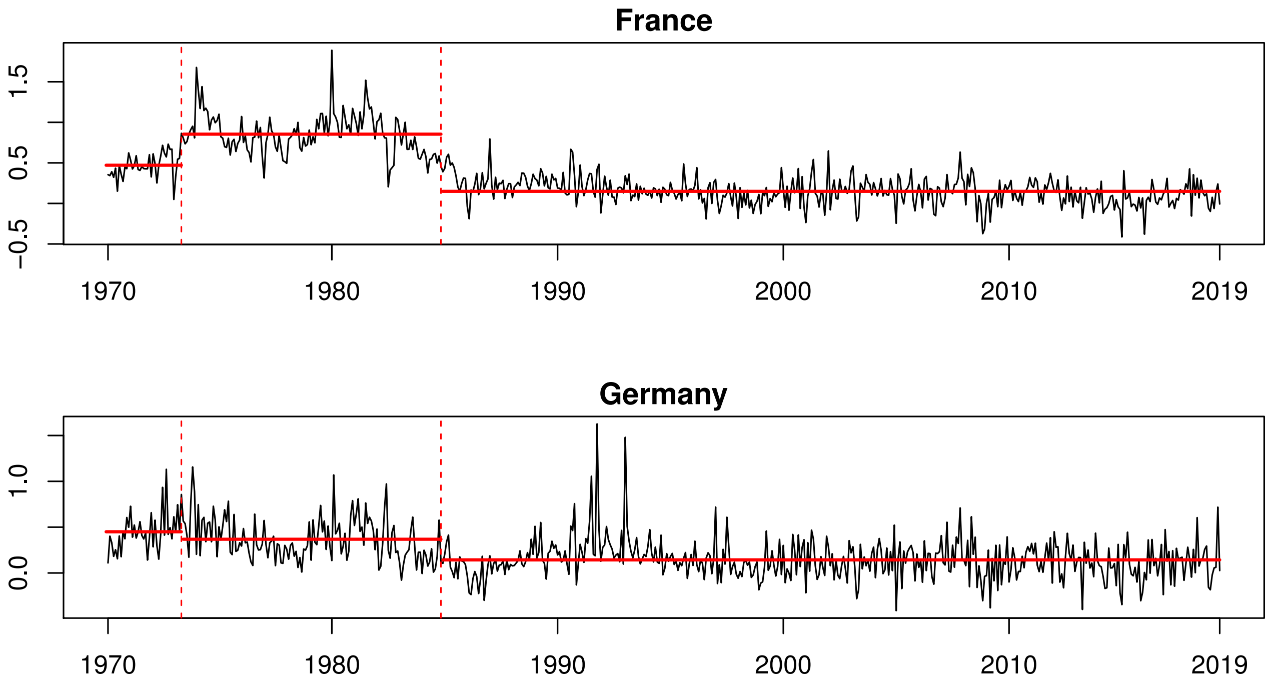


Figure 2: Monthly inflation rates of France and Germany from 1970 to 2019. The dotted red vertical lines refer to the mean shifts our procedure detected. The bold red horizontal lines refer to the estimated means in each partition.

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

Standard estimation procedures of the long-memory parameter are biased upwards in the presence of breaks. The other way around standard testing procedures for shifts detect too many breaks in a long-memory time series. The literature is therefore unclear about the nature of the underlying process of inflation time series. On the one hand Hassler and Wolters (1995) and Baum et al. (1999) argue, for example, that an ARFIMA model can describe inflation rates well. Bos et al. (1999) and Morana (2002) on the other hand find evidence of structural breaks

in international inflation rates and Gadea et al. (2004) shows that the memory of the series is reduced when structural changes are allowed. Many recent contributions favor a mixture of long-memory models and structural breaks (Kumar and Okimoto (2007), among others).
However, whether the series follow a pure long-memory process, a short-memory process with breaks or a mixture of long memory and breaks is of major importance for policy makers: if inflation rates are persistent, monetary policy actions need more time to unfold their effect, which is more expensive.
Our testing procedure allows to detect the true number of breaks for multivariate time series that are allowed to possess long memory. Therefore, it can be used to examine the properties of the underlying process of inflation rates. We use monthly CPI data (P_t) from January 1970 until May 2019 of Germany and France available from the OECD[1] to calculate inflation rates (π_t) as

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

As a result we have 592 observations that are further seasonally adjusted. Figure 2 illustrates the bivariate time series along with the detected break points and partitions. Table 3 provides further results of our testing procedure and regarding the persistence of the raw and the demeaned inflation series.
The left hand side of Panel A of Table 3 shows results regarding the persistence of the inflation series. In line with earlier empirical results (see, for example, Hassler and Wolters (1995) or Bos et al. (1999)) the multivariate local Whittle estimator (GSE) by Shimotsu (2007) estimates high values of d for the raw data such that both series seem to be highly persistent. However, there is evidence that the long-memory time series are contaminated by breaks, which lead to an upward bias of the memory parameter estimates of the GSE (Mikosch and Stărică (2004)). First, the multivariate test against spurious long memory (MLWS) by Sibbertsen et al. (2018) rejects the null hypothesis of pure long-memory processes at a 1% significance level. Second, applying the (univariate) trimmed log-periodogram estimator (tGPH) by McCloskey and Perron (2013) on both inflation series, we observe that the memory decreases. Therefore, we apply our procedure that can consistently detect and estimate multiple shifts in the bivariate system of inflation series. The results can be seen in Panel B of Table 3. We observe that the first structural break our procedure detects was in May 1973, which is a few month before the first oil crisis. We further see that the mean in the second partition of the inflation series of France increases heavily while the mean in the series of Germany decreases. The second structural break is detected in November

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

Panel A: persistence

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

Panel B: breaks

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

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

1984 which could be connected to the 1980s oil glut. We observe that the mean of both inflation series strongly decrease in the third partition of the series.

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

To further investigate whether the REBIT procedure detects the relevant breaks, we examine the demeaned inflation series (demeaning is executed with the two breaks our procedure detected). The results of the GSE estimator and MLWS test can be seen on the right hand side of Panel A of Table 3. The estimated memory by the GSE strongly decreases to a value around 0.24 which is similar to the tGPH estimate of the raw series. Furthermore, evidence of spurious long memory also lessens since the MLWS test is just significant at the 10% significance level. Therefore, we conclude that our procedure detects all relevant breaks of the bivariate inflation system.

7 Conclusions

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

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

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

8 Acknowledgment

Financial support of the Deutsche Forschungsgemeinschaft (DFG) is gratefully acknowledged. We are also grateful to Michael Massmann, Christian Leschinski, the participants of the 13th International Conference on Computational and Financial Econometrics 2019 in London, the Joint Statistical Meeting of the American Statistical Association 2020 in Philadelphia and the SNDE meeting 2020 for inspiring discussion and comments on this paper.

References

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

Bai, Jushan (1997b). “Estimation of a change point in multiple regression models”. In: *Review of Economics and Statistics* 79(4), pp. 551–563.

Bai, Jushan (2000). “Vector Autoregressive Models with Structural Changes in Regression Coefficients and in Variance-Covariance Matrices”. In: *Annals of Economics and Finance* 1(2), pp. 303–339.

Bai, Jushan, Robin L Lumsdaine, and James H Stock (1998). “Testing for and dating common breaks in multivariate time series”. In: *The Review of Economic Studies* 65(3), pp. 395–432.

Bai, Jushan and Pierre Perron (1998). “Estimating and testing linear models with multiple structural changes”. In: *Econometrica*, pp. 47–78.

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

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

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

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

Davidson, James and Robert M. de Jong (2000). “The functional central limit theorem and weak convergence to stochastic integrals II: Fractionally Integrated Processes”. In: *Econometric Theory* 16(5), pp. 643–666.

Dehling, Herold, Aeneas Rooch, and Murad S Taqqu (2013). “Non-Parametric Change-Point Tests for Long-Range Dependent Data”. In: *Scandinavian Journal of Statistics* 40(1), pp. 153–173.

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

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

Hassler, Uwe and Jürgen Wolters (1995). "Long memory in inflation rates: International evidence". In: *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". In: *Journal of Time Series Analysis* 35(1), pp. 40–54.

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

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

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

Marinucci, D. and P.M. Robinson (1999). "Alternative forms of fractional Brownian motion". In: *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". In: *Stochastic Processes and their Applications* 86(1), pp. 103–120.

McCloskey, Adam and Pierre Perron (2013). "Memory parameter estimation in the presence of level shifts and deterministic trends". In: *Econometric Theory* 29(6), pp. 1196–1237.

Mikosch, Thomas and Cătălin Stărică (2004). "Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects". In: *Review of Economics and Statistics* 86(1), pp. 378–390.

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

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

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

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

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

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

Sibbertsen, Philipp, Christian Leschinski, and Marie Busch (2018). “A multivariate test against spurious long memory”. In: *Journal of Econometrics* 203(1), pp. 33–49.

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

Wenger, Kai and Christian Leschinski (2019). “Fixed-bandwidth CUSUM tests under long memory”. In: *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”. In: *AStA Advances in Statistical Analysis* 103(2), pp. 237–256.

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

Appendix to

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

Not Intended for Publication!

Will be Provided as Online Appendix

Appendix

A Proofs

This section contains the proofs of Lemma 1 and Theorem 1 and 2. In order to prove these results we need a generalized Hájek-Rényi inequality, a strong law of large numbers (SLLN) and a functional central limit theorem (FCLT) that hold under our stated assumptions and in particular under long memory. We collect them in separate Lemmas in Section A.1. Afterwards we show in Section A.2 that under our set of Assumptions 10 properties of the quasi-likelihood are satisfied that have been considered in Bai et al. (1998), Bai (2000) and Qu and Perron (2007). The proofs can be found in an online appendix. We prove consistency of the break point estimators, i.e. Lemma 1. In Section A.4 we proof the limiting distribution of our test statistic, i.e. Theorem 2.

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

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

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

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

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

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

Suppose (A.2) holds. Then we can write

$$\begin{aligned} P(\max_{2^p m\leq k\leq 2^{p+1}m}\frac{1}{k}\|M_{1:k}\|>\delta) \leq & P(\frac{1}{2^p m}\|M_{1:m}\|>\frac{\delta}{2}) \\ & + P(\max_{2^p m+1\leq k\leq 2^{p+1}m}\frac{1}{k}\|M_{2^p m+1:2^{p+1}m}\|>\frac{\delta}{2}) \\ \leq & 4\frac{A(d)C(\varepsilon)}{\delta^2}(2^p m)^{2d-2}+4\frac{A(d)C(\varepsilon)}{\delta^2}(2^p m)^{2d}\sum_{t=2^p m+1}^{2^{p+1}}\left(\frac{1}{t}\right)^2 \\ \leq & 8\frac{A(d)C(\varepsilon)}{\delta^2}(2^p m)^{2d-1}. \end{aligned} \tag{A.3}$$

Using equation (A.1) we have

$$P\left(\max_{k\geq m}\frac{1}{k}\|M_{1:k}\|>\delta\right) \leq 8\frac{A(d)C(\varepsilon)}{\delta^2}\sum_{p=0}^{\infty}(2^p m)^{2d-1} \leq \frac{L}{\delta^2}m^{2d-1}, \tag{A.4}$$

where $L<\infty$ is a constant.

We prove equation (A.2) by the Markov inequality. To simplify notation we write $S_{i:j} = \max_{k=i,\ldots,j} 1/k\,\|M_{1:k}\|$. Specifically, we set out to prove

$$E\left(S_{1:n}^2\right) \leq C(\varepsilon)A(d)n^{2d}\sum_{t=1}^{n}\frac{1}{t^2}. \tag{A.5}$$

If equation (A.5) holds, our auxiliary result (A.2) is proven by the Markov inequality. The claim in (A.5) is proved by induction on n. For $n=1$ the inequality is obvious for $A(d)=1$ because of the following inequality: Kechagias and Pipiras (2015) proved that for the partial sums $M_{i:j}$ there exists $C(\varepsilon)<\infty$ such that, for all i,j,

$$E\left(\|M_{i:j}\|^2\right) \leq C(\varepsilon)\,|j-i+1|^{2d+1}. \tag{A.6}$$

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

$$\max_{k=1,\ldots,n}\frac{1}{k}\,\|M_{1:k}\| \leq \frac{1}{m}M_{1:m}+\left(\left(\max_{k=1,\ldots,m-1}\frac{1}{k}\,\|M_{1:k}\|\right)^2+\left(\max_{k=m+1,\ldots,n}\frac{1}{k}\,\|M_{1:k}\|\right)^2\right)^{1/2}. \tag{A.7}$$

Applying the Minkowski inequality to the above inequality yields

$$\begin{aligned} E\left(S_{1:n}^2\right)^{1/2} \leq & \frac{1}{m}(E(\|M_{1:m}\|^2)^{1/2}+\left(E(S_{1:m-1}^2)+E(S_{m+1:n}^2)\right)^{1/2} \\ \leq & \frac{1}{m}\left(C(\varepsilon)m^{2d+1}\right)^{1/2}+\left(A(d)C(\varepsilon)\left((m-1)^{2d}\sum_{t=1}^{m-1}\frac{1}{t^2}+(n-m)^{2d}\sum_{t=m+1}^{n}\frac{1}{t^2}\right)\right)^{1/2} \\ \leq & \left(C(\varepsilon)m^{2d}\sum_{t=1}^{n}\frac{1}{t^2}\right)^{1/2}+\left(A(d)C(\varepsilon)\left(\frac{n}{2}\right)^{2d}\left(\sum_{t=1}^{m-1}\frac{1}{t^2}+\sum_{t=m+1}^{n}\frac{1}{t^2}\right)\right)^{1/2} \\ \leq & \left(C(\varepsilon)n^{2d}\sum_{t=1}^{n}\frac{1}{t^2}\right)^{1/2}\left(1+\left(\frac{A(d)}{2^{2d}}\right)^{1/2}\right), \end{aligned} \tag{A.8}$$

where we used equation (A.6) and the induction hypothesis in the second line and the fact that $1\leq\sum_{t=1}^{m}1/t^2$ in the third line. Now we choose $A(d)$ such that

$$1+\frac{A(d)^{1/2}}{2^d} \leq A(d)^{1/2} \quad\Leftrightarrow\quad A(d) \geq \left(1-\frac{1}{2^d}\right)^{-2} \geq 1. \tag{A.9}$$

The induction step is proven and thus this concludes the proof of inequality (A.5). □

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

$$W_{d_i}(t) = \frac{1}{\Gamma(d_i+1)} \left(\int_0^t (t-s)^{d_i} dW_0^{(i)}(s) + \int_{-\infty}^0 \left((t-s)^{d_i} - (-s)^{d_i} \right) dW_0^{(i)}(s) \right), \qquad \text{(A.10)}$$

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

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

(a) (FCLT)

$$\operatorname{diag}(T^{-1/2-d_1}, \ldots, T^{-1/2-d_n}) \sum_{t=1}^{[Tr]} \xi_t \Rightarrow \Omega W_D(r), \qquad \text{(A.11)}$$

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

(b) (SLLN)

$$k^{-1} \sum_{i=1}^{k} \xi_i \xrightarrow{a.s.} 0 \qquad \text{as} \qquad k \to \infty; \qquad \text{(A.12)}$$

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 10 Properties of the Quasi-Likelihood Ratio

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

$$\mathcal{L}(1,k;\beta,\Sigma) = \frac{\prod_{t=1}^{k} f(y_t|x_t,...,\beta,\Sigma)}{\prod_{t=1}^{k} f(y_t|x_t,...,\beta_0,\Sigma_0)}, \tag{A.13}$$

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

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

$$\sup_{T\delta\le k\le T} \mathcal{L}(1,k;\hat{\beta}_{(k)},\hat{\Sigma}_{(k)}) = O_p(1), \tag{A.14}$$

$$\sup_{T\delta\le k\le T} (\|\hat{\beta}_{(k)}-\beta_0\| + \|\hat{\Sigma}_{(k)}-\Sigma_0\|) = O_p(T^{d-1/2}). \tag{A.15}$$

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

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

$$\Pr\left(\sup_{\delta T\le k\le T} T^{-B}\mathcal{L}(1,k;\hat{\beta}_{(k)},\hat{\Sigma}_{(k)}) > 1\right) < \varepsilon \tag{A.16}$$

for all large T*.*

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

$$\Pr\left(\sup_{k\ge\delta T}\sup_{(\beta,\Sigma)\in S_T} T^{D}\mathcal{L}(1,k;\beta,\Sigma) > 1\right) < \varepsilon. \tag{A.17}$$

Property 4. *Not needed.*

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

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

$$\Pr\left(\sup_{k\geq Ah_T}\sup_{(\beta,\Sigma)\in S_T\cap\bar{\Theta}_3}\mathcal{L}(1,k;\beta,\Sigma)>\varepsilon\right)<\varepsilon \tag{A.18}$$

when T is large.

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

$$\sup_{1\leq k\leq T^{1/2-d}v_T^{-1}}\sup_{\lambda,\Xi}\frac{\mathcal{L}(1,k;\beta+T^{-1/2+d}\lambda,\Sigma+T^{-1/2+d}\Xi)}{\mathcal{L}(1,k;\beta,\Sigma)}=o_p(1). \tag{A.19}$$

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

$$\sup_{1\leq k\leq Mv_T^{-2}}\sup_{\lambda,\Xi}\log\frac{\mathcal{L}(1,k;\beta+T^{-1/2+d}\lambda,\Sigma+T^{-1/2+d}\Xi)}{\mathcal{L}(1,k;\beta,\Sigma)}=o_p(1). \tag{A.20}$$

Property 8. *We have*

$$\sup_{T\delta\leq k\leq T}\sup_{\beta^*,\Sigma^*,\lambda,\Xi}\log\frac{\mathcal{L}(1,k;\beta_0+T^{-1/2+d}\beta^*+T^{-1+2d}\lambda,\Sigma_0+T^{-1/2+d}\Sigma^*+T^{-1+2d}\Xi)}{\mathcal{L}(1,k;\beta_0+T^{-1/2+d}\beta^*,\Sigma_0+T^{-1/2+d}\Sigma^*)}=o_p(1), \tag{A.21}$$

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

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

$$y_t = x_t'\beta_1^0+\Sigma_1^0\eta_t, \qquad (t=1,\ldots,T_1), \tag{A.22}$$

$$y_t = x_t'\beta_2^0+\Sigma_2^0\eta_t, \qquad (t=T_1+1,\ldots,T_1+T_2), \tag{A.23}$$

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

Property 10. *Not needed.*

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

$$P(\sup_{(k_1,...,k_m)\in A_j} LR_T(k_1, \dots, k_m) > \varepsilon) < \varepsilon. \tag{A.24}$$

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

$$LR_T(k_1, \dots, k_m) \leq LR_T(k_1, \dots, k_m, k_1^0, \dots, k_{j-1}^0, k_j^0 - N, k_j^0 + N, k_{j+1}^0, \dots, k_m^0). \tag{A.25}$$

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

$$D(k, l, \beta, \Sigma) = \frac{\prod_{t=k+1}^{l} f(y_t|x_t; \beta, \Sigma)}{\prod_{t=k+1}^{l} f(y_t|x_t; \beta^0, \Sigma^0)} \tag{A.26}$$

and its optimal value

$$D(k, l) = \sup_{\beta,\Sigma} D(k, l, \beta, \Sigma). \tag{A.27}$$

The likelihood ratio of the entire sample can be written as

$$LR_T(k_1, \dots, k_m) = D(0, k_1) \cdot D(k_1, k_2) \cdot \dots \cdot D(k_m, T). \tag{A.28}$$

The right hand side of (A.25) can be written as the product of at most $(2m+2)$ terms expressible as $D(l, k)$ as in (A.28). There are at most $(2m+2)$ terms because k_i may coincide with k_l^0 for some i and l. One of these $(2m+2)$ terms is $D(k_j^0 - N, k_j^0 + N)$ and all the rest can be written as $D(l, k)$ with $[l, k] \subset [k_1^0 + 1; k_{i+1}^0]$ for some i. By Property 1 and 2, $\log D(l, k) = O_p(\log T)$ uniformly in l, k such that $k_i^0 + 1 \leq l < k \leq k_{i+1}^0$ with $|l - k| > T\nu$. That is, $D(k, l) = O_p(T^B)$ for some $B > 0$. Thus,

$$LR_T(k_1, \dots, k_m) \leq O_p(T^{(2m+1)B})\, D(k_j^0 - N, k_j^0 + N). \tag{A.29}$$

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

$$\begin{aligned} D(k_j^0 - N, k_j^0 + N) &= \sup_{\beta,\Sigma}[D(k_j^0 - N, k_j^0; \beta, \Sigma) \cdot D(k_j^0, k_j^0 + N; \beta, \Sigma)] \\ &= \sup_{\beta,\Sigma}[LR_T^*(k_j^0 - N, k_j^0; N^{1/2}(\beta - \beta_j^0), N^{1/2}(\Sigma - \Sigma_j^0)) \qquad \text{(A.30)} \\ &\quad \times (LR_T^*(k_j^0, k_j^0 + N; N^{1/2}(\beta - \beta_{j+1}^0), N^{1/2}(\Sigma - \Sigma_{j+1}^0))]. \end{aligned}$$

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

$$\max\{N^{1/2}\|\beta - \beta_j^0\|, N^{1/2}\|\beta - \beta_{j+1}^0\|\} \geq N^{1/2}\|\beta_j^0 - \beta_{j+1}^0\|/2 \qquad \text{(A.31)}$$

$$\max\{N^{1/2}\|\Sigma - \Sigma_j^0\|, N^{1/2}\|\Sigma - \Sigma_{j+1}^0\|\} \geq N^{1/2}\|\Sigma_j^0 - \Sigma_{j+1}^0\|/2. \qquad \text{(A.32)}$$

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

$$N^{1/2}\|\beta_j^0 - \beta_{j+1}^0\|/2 = (T^{1/2}\nu_T^{-1})^{1/2}\nu_T C = C(T^{1/2}\nu_T)^{\frac{1}{2}} \geq \log T \geq \log N. \qquad \text{(A.33)}$$

Now suppose that $N^{1/2}\|\beta_j^0 - \beta_{j+1}^0\|/2 \geq \log N$. Then we have either (i) $N^{1/2}\|\beta - \beta_j^0\| \geq \log N$ or (ii) $N^{1/2}\|\beta - \beta_{j+1}^0\| \geq \log N$. For case (i) we can apply Property 3 to the first term inside the brackets of (A.30) to obtain

$$LR_T^*(k_j^0 - N, k_j^0; N^{1/2}(\beta - \beta_j^0), N^{1/2}(\Sigma - \Sigma_j^0)) = O_p(N^{-A}) \qquad \text{(A.34)}$$

for every $A > 0$. Moreover, by Property 2 the second term inside the bracket of (A.30) is bounded by $O_p(\log T)$. Similarly, for case (ii), we can apply Property 3 to show that the second term of (A.30) is $O_p(N^{-A})$ and the first term is bounded by $O_p(\log T)$. So for each case, we have

$$D(k_j^0 - N, k_j^0 + N) = \log T \, O_p(N^{-A}) \qquad \text{(A.35)}$$

for an arbitrary $A < 0$. It is further $N^{-A} \leq T^{-A/2}$ since $N \geq T^{1/2}$ for all large T. Thus from (A.29), $LR_T(k_1, \ldots, k_m) \leq O_p(T^{(2m+1)B - \frac{1}{2}A}) \log T \xrightarrow{p} 0$ for a large A. This proves (A.24).

Now by Proposition 2 of Bai (2000) we can deduce that $\hat{k}_j - k_j^0 = O_p(\nu_T^2)$ for $j = 1, \ldots, m$ using the preliminary convergence order given by Equation (A.24). The convergence rate for the estimated regression coefficients β_j and covariances Σ_j follows as in Bai (1997b) and Bai and Perron (1998) due to the fast convergence of the estimated break points. □

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)

$$\begin{aligned}
&\frac{1}{2}\sum_{t=T_j^0+[s/\nu_T^2]}^{T_j^0} u_t^T\Big((\Sigma_j^0)^{-1} - (\Sigma_{j+1}^0)^{-1}\Big)u_t \qquad \text{(A.36)}\\
&\qquad = \frac{1}{2}tr\Big((\Sigma_j^0)^{\frac{1}{2}}(\Sigma_{j+1}^0)^{-1}\Phi_j(\Sigma_j^0)^{-\frac{1}{2}}\nu_T\sum_{t=T_j^0+[s/\nu_T^2]}^{T_j^0}(\eta_t\ \eta_t^T - I)\\
&\qquad - \frac{r}{2}\nu_T\ tr\Big((\Sigma_{j+1}^0)^{-1}\Phi_j\Big).
\end{aligned}$$

For the second term we have

$$-\frac{r}{2}(\log|\Sigma_j^0| - \log|\Sigma_{j+1}^0|) = \frac{r}{2}\nu_T\ tr\Big(\Phi_j(\Sigma_{j+1}^0)^{-1}\Big) + \frac{r}{4}\nu_T^2\ tr\Big([\Phi_j(\Sigma_{j+1}^0)^{-1}]^2\Big). \qquad \text{(A.37)}$$

The sum of the first two terms is

$$\begin{aligned}
&\frac{1}{2}\sum_{T_j^0+[s/\nu_T^2]}^{T_j^0} u_t^T\Big((\Sigma_j^0)^{-1} - (\Sigma_{j+1}^0)^{-1}\Big)u_t - \frac{r}{2}(log|\Sigma_j^0| - log|\Sigma_{j+1}^0|) \qquad \text{(A.38)}\\
&= \frac{1}{2}tr\Big((\Sigma_j^0)^{\frac{1}{2}}(\Sigma_{j+1}^0)^{-1}\Phi_j(\Sigma_j^0)^{-\frac{1}{2}}\nu_T\sum_{T_j^0+[s/\nu_T^2]}^{T_j^0}(\eta_t\ \eta_t^T - I)\Big)\\
&+ \frac{r}{4}\nu_T^2\ tr\Big([\Phi_j(\Sigma_{j+1}^0)^{-1}]^2\Big) = \mathrm{I} + \mathrm{II}.
\end{aligned}$$

Now

$$\begin{aligned}
T^{1-2d}(\mathrm{I}+\mathrm{II}) &\xrightarrow{d} \frac{1}{2}tr\Big((\Sigma_j^0)^{\frac{1}{2}}(\Sigma_{j+1}^0)^{-1}\Phi_j(\Sigma_j^0)^{-\frac{1}{2}}\xi_{1,d,j}(s)\Big) \qquad \text{(A.39)}\\
&+ \frac{s}{4}tr\Big([(\Sigma_{j+1}^0)^{-1}\Phi_j]^2\Big)\\
&= \frac{1}{2}tr(A_{1,j}\xi_{1,d,j}(s)) + \frac{s}{4}tr(A_{1,j}^2),
\end{aligned}$$

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

$$-\frac{1}{2}\sum_{t=T_j^0+[s/\nu_T^2]}^{T_j^0}(\beta_j^0-\beta_{j+1}^0)^T x_t(\Sigma_{j+1}^0)^{-1}x_t^T(\beta_j^0-\beta_{j+1}^0)\xrightarrow{P}\frac{1}{2}s\ \delta_j^T\ Q_{1,j}\ \delta_j. \tag{A.40}$$

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

$$-T^{1-2d}\sum_{t=T_j^0+[s/\nu_T^2]}^{T_j^0}(\beta_j^0-\beta_{j+1}^0)^T x_t(\Sigma_{j+1}^0)^{-1}u_t\xrightarrow{d}\delta_j^T(\Pi_{1,j})^{\frac{1}{2}}\zeta_{1,d,j}(s) \tag{A.41}$$

with

$$\Pi_{1,j}=\lim_{T\to\infty}Var\Big\{(T_j^0-T_{j-1}^0)^{-\frac{1}{2}}\Big[\sum_{t=T_j^0+[s/\nu_T^2]}^{T_j^0}x_t(\Sigma_{j+1}^0)^{-1}(\Sigma_j^0)^{\frac{1}{2}}\eta_t\Big]\Big\}. \tag{A.42}$$

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

$$\begin{aligned}T^{1-2d}lr_j^1\Big([\frac{s}{\nu_T^2}]\Big)\xrightarrow{d}&-\frac{|s|}{2}\Big[\frac{1}{2}tr(A_{1,j}^2)+\delta_j^T\ Q_{1,j}\ \delta_j\Big] \\ &+\frac{1}{2}\,\mathrm{vec}(A_{1,j})^T\,\mathrm{vec}(\xi_{1,d,j}(s))+\delta_j^T(\pi_{1,j})^{\frac{1}{2}}\zeta_{1,d,j}(s).\end{aligned} \tag{A.43}$$

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

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

$$\Big(vec(A_{1,j})^T\Omega^0_{1,j}vec(A_{1,j})/4\Big)^{\frac{1}{2}}V_{1,d,j}(s)+\Big(\delta_j^T(\pi_{1,j})\delta_j\Big)^{\frac{1}{2}}U_{1,d,j}(s) \tag{A.44}$$

$$\stackrel{d}{=}\Big(vec(A_{1,j})^T\Omega^0_{1,j}vec(A_{1,j})/4+\delta_j^T(\pi_{1,j})\delta_j\Big)^{\frac{1}{2}}W_{1,j,d}(s) \tag{A.45}$$

$$\equiv T_{1,j}\ W_{1,j,d}(s), \tag{A.46}$$

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

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

$$T^{1-2d}lr_j^1\Big([\frac{s}{\nu_T^2}]\Big)\xrightarrow{d}-\frac{|s|}{2}\Delta_{1,j}+T_{1,j}\ W_{1,j,d}(s) \tag{A.47}$$

The proof for $s>0$ is similar:

$$T^{1-2d}lr_j^1\Big([\frac{s}{\nu_T^2}]\Big)\xrightarrow{d}-\frac{|s|}{2}\Delta_{2,j}+T_{2,j}\ W_{2,j,d}(s) \tag{A.48}$$

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

$$T_{2,j}=\Big[vec(A_{2,j})^T\Omega^0_{2,j}vec(A_{2,j})/4+\delta_j^T(\pi_{2,j})\delta_j\Big]^{\frac{1}{2}}. \tag{A.49}$$

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

Hence,

$$T^{1-2d}\nu_T^2(\hat{T}_j - T_j^0) \xrightarrow{d} \underset{s}{argmax} \begin{cases} -\frac{|s|}{2}\Delta_{1,j} + T_{1,j}\ W_{j,d}(s), s \leq 0, \\ -\frac{|s|}{2}\Delta_{2,j} + T_{2,j}\ W_{j,d}(s), s > 0, \end{cases} \tag{A.50}$$

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

A.5 Proof of Theorem 2

Proof of Theorem 2. We introduce some notation first. Let

$$\tilde{\Sigma}_{1,j} = \frac{1}{T_j}\sum_{t=1}^{T_j}(y_t - x'_{at}\tilde{\beta}_a - x'_{bt}\tilde{\beta}_{b1,j})(y_t - x'_{at}\tilde{\beta}_a - x'_{bt}\tilde{\beta}_{b1,j}) \tag{A.51}$$

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

$$\tilde{\beta}_{b1,j} = (\sum_{t=1}^{T_j} x_{bt}\tilde{\Sigma}_{1,j}^{-1}x'_{bt})^{-1}\sum_{t=1}^{T_j} x_t\tilde{\Sigma}_{1,j}^{-1}(y_t - x'_{at}\tilde{\beta}_a). \tag{A.52}$$

Additionally,

$$\hat{\Sigma}_j = \frac{1}{T_j - T_{j-1}}\sum_{t=T_{j-1}+1}^{T_j}(y_t - x'_{at}\hat{\beta}_a - x'_{bt}\hat{\beta}_{bj})(y_t - x'_{at}\hat{\beta}_a - x'_{bt}\hat{\beta}_{bj})' \tag{A.53}$$

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

$$\hat{\beta}_{b,j} = (\sum_{t=T_{j-1}+1}^{T_j} x_{bt}\hat{\Sigma}_j^{-1}x'_{bt})^{-1}\sum_{t=T_{j-1}+1}^{T_j} x_t\hat{\Sigma}_j^{-1}(y_t - x'_{at}\hat{\beta}_a). \tag{A.54}$$

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

$$\begin{aligned} LR_T(T_1,\dots,T_m) &= \frac{2}{T^{2d}}\log\hat{L}_T(T_1,\dots,T_m) - \frac{2}{T^{2d}}\log\tilde{L}_T = \frac{T}{T^{2d}}\log|\tilde{\Sigma}| - \frac{T}{T^{2d}}\log|\hat{\Sigma}| \qquad \text{(A.55)}\\ &= \frac{1}{T^{2d}}\sum_{j=1}^{m}(T_{j+1}\log|\tilde{\Sigma}_{1,j+1}| - T_j\log|\tilde{\Sigma}_{1,j}| - (T_{j+1}-T_j)\log|\hat{\Sigma}_{j+1}|)\\ &=: \frac{1}{T^{2d}}\sum_{j=1}^{m}F_T^j. \end{aligned}$$

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

$$\begin{aligned} \log|\tilde{\Sigma}_{1,j+1}| = &\log|\Sigma^0| + \operatorname{tr}((\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j+1} - \Sigma^0)) \qquad \text{(A.56)}\\ &- \frac{1}{2}\operatorname{tr}((\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j+1} - \Sigma^0)(\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j+1} - \Sigma^0))\\ &+ o_p(T^{-1}),\\ \log|\tilde{\Sigma}_{1,j}| = &\log|\Sigma^0| + \operatorname{tr}((\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j} - \Sigma^0)) \qquad \text{(A.57)}\\ &- \frac{1}{2}\operatorname{tr}((\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j} - \Sigma^0)(\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j} - \Sigma^0)) + o_p(T^{-1}),\\ \log|\hat{\Sigma}_{j+1}| = &\log|\Sigma^0| + \operatorname{tr}((\Sigma^0)^{-1}(\hat{\Sigma}_{j+1} - \Sigma^0)) \qquad \text{(A.58)}\\ &- \frac{1}{2}\operatorname{tr}((\Sigma^0)^{-1}(\hat{\Sigma}_{j+1} - \Sigma^0)(\Sigma^0)^{-1}(\hat{\Sigma}_{j+1} - \Sigma^0)) + o_p(T^{-1}). \end{aligned}$$

Applying this to the terms F_T^j,

$$
\begin{aligned}
F_T^j :=& F_{1,T}^j + F_{2,T}^j \\
=& \operatorname{tr}(T_{j+1}(\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j+1} - \Sigma^0) - T_j(\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j} - \Sigma^0)) \qquad (A.59)\\
& - (T_{j+1} - T_j)(\Sigma^0)^{-1}(\hat{\Sigma}_{j+1} - \Sigma^0)) \\
& - \frac{1}{2}\operatorname{tr}(T_{j+1}[(\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j+1} - \Sigma^0)]^2 \qquad (A.60)\\
& - T_j[(\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j} - \Sigma^0)]^2 - (T_{j+1} - T_j)[(\Sigma^0)^{-1}(\hat{\Sigma}_{j+1} - \Sigma^0)]^2).
\end{aligned}
$$

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

$$Y = X_a\beta_a + X_b\beta_b + U \qquad (A.61)$$

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

$$Y_{1,j} = X_{a1,j}\beta_a + X_{b1,j}\beta_{b1,j} + U_{1,j}. \qquad (A.62)$$

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

$$\tilde{\beta}_a = [W'M_ZW]^{-1}W'M_ZY^d, \qquad (A.63)$$

$$\tilde{\beta}_{b1,j} = (Z_{1,j}'Z_{1,j})^{-1}Z_{1,j}'(Y_{1,j}^d - W_{1,j}\tilde{\beta}_a), \qquad (A.64)$$

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

$$Y_{j+1} = X_{a,j+1}\beta_a + X_{b,j+1}\beta_{b,j+1} + U_{j+1}. \qquad (A.65)$$

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

$$\hat{\beta}_a = [\bar{W}'M_{\bar{Z}}\bar{W}]^{-1}\bar{W}'M_{\bar{Z}}\bar{Y}^d, \qquad (A.66)$$

$$\hat{\beta}_{b,j+1} = (\bar{Z}_{j+1}'\bar{Z}_{j+1})^{-1}\bar{Z}_{j+1}'(\bar{Y}_{j+1}^d - \bar{W}_{j+1}\hat{\beta}_a). \qquad (A.67)$$

Note that the choice of the estimate of the covariance matrix in Equations (A.63) to (A.67) 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 $F_{1,T}^j$ (or with obvious changes for the second component) that

$$T_{j+1}\,\mathrm{tr}((\Sigma^0)^{-1}\tilde{\Sigma}_{j+1})$$
$$= A_T'W_{1,j+1}'M_{Z_{1,j+1}}W_{1,j+1}A_T - U_{1,j+1}^{d'}P_{Z_{1,j+1}}U_{1,j+1}^d$$
$$- 2(M_{Z_{1,j+1}}W_{1,j+1}A_T)'U_{1,j+1}^d + U_{1,j+1}'(I_T\otimes(\Sigma^0)^{-1})U_{1,j+1} + o_p(1),$$

where $A_T = [W'M_ZW]^{-1}W'M_ZU^d$. For the third component of $F_{1,T}^j$ it can be shown that

$$(T_{j+1}-T_j)\,\mathrm{tr}((\Sigma^0)^{-1}\hat{\Sigma}_{j+1}) \tag{A.68}$$
$$=\bar{A}_T'\bar{W}_{j+1}'M_{\bar{Z}_{j+1}}\bar{W}_{j+1}\bar{A}_T - \bar{U}_{j+1}^{d'}P_{\bar{Z}_{j+1}}\bar{U}_{j+1}^d$$
$$- 2(M_{\bar{Z}_{j+1}}\bar{W}_{j+1}\bar{A}_T)'\bar{U}_{j+1}^d + U_{j+1}'(I_T\otimes(\Sigma^0)^{-1})U_{j+1} + o_p(1),$$

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

$$F_{1,T}^j = U_{1,j}^{d'}P_{Z_{1,j}}U_{1,j}^d + U_{j+1}^{d'}P_{\bar{Z}_{j+1}}U_{j+1}^d - U_{1,j+1}^{d'}P_{Z_{1,j+1}}U_{1,j+1}^d + o_p(1). \tag{A.69}$$

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

$$T^{-2d}\,U_{1,j+1}^{d'}P_{Z_{1,j+1}}U_{1,j+1}^d \Rightarrow [W_{D,p_b}^*(\lambda_{j+1})'W_{D,p_b}^*(\lambda_{j+1})]/\lambda_{j+1}. \tag{A.70}$$

Using similar arguments

$$T^{-2d}\,U_{1,j}^{d'}P_{Z_{1,j}}U_{1,j}^d \Rightarrow [W_{D,p_b}^*(\lambda_j)'W_{D,p_b}^*(\lambda_j)]/\lambda_j \tag{A.71}$$

and

$$T^{-2d}\,U_{j+1}^{d'}P_{\bar{Z}_{j+1}}U_{j+1}^d \tag{A.72}$$
$$\Rightarrow (W_{D,p_b}^*(\lambda_{j+1})-W_{D,p_b}^*(\lambda_j))'(W_{D,p_b}^*(\lambda_{j+1})-W_{D,p_b}^*(\lambda_j))/(\lambda_{j+1}-\lambda_j).$$

These results imply that the first component in (A.59) has the limit

$$F_{1,T}^j \Rightarrow \frac{(\lambda_jW_{D,p_b}^*(\lambda_{j+1})-\lambda_{j+1}W_{D,p_b}^*(\lambda_j))'(\lambda_jW_{D,p_b}^*(\lambda_{j+1})-\lambda_{j+1}W_{D,p_b}^*(\lambda_j))}{(\lambda_{j+1}-\lambda_j)\lambda_j\lambda_{j+1}}. \tag{A.73}$$

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

$$F_{2,T}^{j} = -\frac{1}{2}\sum_{j=1}^{m} \mathrm{tr}(T_{j+1}((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I)^2) \qquad \text{(A.74)}$$

$$- T_j((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j} - I)^2 - (T_{j+1} - T_j)((\Sigma^0)^{-1}\hat{\Sigma}_{j+1} - I)^2.$$

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

$$\left[((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I)^F\right]_{i,k} = \frac{\sigma^{ik}}{T}\sum_{t=1}^{T}(y_{it} - x_{it}'\tilde{\beta})'(y_{kt} - x_{kt}'\tilde{\beta}) - I_{i,k}, \qquad \text{(A.75)}$$

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

$$\left[((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I)^S\right]_{i,k} = \frac{\sigma^{ik}}{T_{j+1}}\sum_{t=1}^{T_{j+1}}(y_{it} - x_{it}'\tilde{\beta})'(y_{kt} - x_{kt}'\tilde{\beta}) - I_{i,k}. \qquad \text{(A.76)}$$

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

$$((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I) = ((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I)^F + ((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I)^S, \qquad \text{(A.77)}$$

$$((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j} - I) = ((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j} - I)^F + ((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j} - I)^S, \qquad \text{(A.78)}$$

$$((\Sigma^0)^{-1}\hat{\Sigma}_{j+1} - I) = ((\Sigma^0)^{-1}\hat{\Sigma}_{j+1} - I)^F + ((\Sigma^0)^{-1}\hat{\Sigma}_{j+1} - I)^S, \qquad \text{(A.79)}$$

and, in view of (A.60),

$$\sum_{j=1}^{m} F_{2,T}^{j} = -\frac{1}{2}\mathrm{tr}(\sum_{j=1}^{m}[T_{j+1}((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I)^S((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I)^S \qquad \text{(A.80)}$$

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

$$- (T_{j+1} - T_j)((\Sigma^0)^{-1}\hat{\Sigma}_{j+1}^S - I)^S((\Sigma^0)^{-1}\hat{\Sigma}_{j+1}^S - I)^S])$$

$$+ o_p(1)$$

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

$$\frac{T_{j+1}}{T^{2d}}((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I)^S((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I)^S \tag{A.81}$$
$$= \frac{T}{T_{j+1}}\left(T^{-1/2-d}\sum_{t=1}^{T_{j+1}}[(\Sigma^0)^{-1}u_t u_t' - I]\right)^S\left(T^{-1/2-d}\sum_{t=1}^{T_{j+1}}[(\Sigma^0)^{-1}u_t u_t' - I]\right)^S + o_p(1)$$
$$\Rightarrow \frac{\xi_n^d(\lambda_{j+1})^S\xi_n^d(\lambda_{j+1})^S}{\lambda_{j+1}}$$

$$\frac{T_j}{T^{2d}}((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j} - I)^S((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j} - I)^S \tag{A.82}$$
$$= \frac{T}{T_j}\left(T^{-1/2-d}\sum_{t=1}^{T_j}[(\Sigma^0)^{-1}u_t u_t' - I]\right)^S\left(T^{-1/2-d}\sum_{t=1}^{T_j}[(\Sigma^0)^{-1}u_t u_t' - I]\right)^S + o_p(1)$$
$$\Rightarrow \frac{\xi_n^d(\lambda_j)^S\xi_n^d(\lambda_j)^S}{\lambda_j}$$

and

$$\frac{(T_{j+1} - T_j)}{T^{2d}}(\Sigma^0)^{-1}\hat{\Sigma}_{j+1}^S - I)^S((\Sigma^0)^{-1}\hat{\Sigma}_{j+1}^S - I)^S \tag{A.83}$$
$$= \frac{T}{T_{j+1} - T_j}\left(T^{-1/2-d}\sum_{t=T_j+1}^{T_{j+1}}[(\Sigma^0)^{-1}u_t u_t' - I]\right)^S$$
$$\times\left(T^{-1/2-d}\sum_{t=T_j+1}^{T_{j+1}}[(\Sigma^0)^{-1}u_t u_t' - I]\right)^S + o_p(1)$$
$$\Rightarrow \frac{(\xi_{D,n}^*(\lambda_{j+1}) - \xi_{D,n}^*(\lambda_j))^S(\xi_{D,n}^*(\lambda_{j+1}) - \xi_{D,n}^*(\lambda_j))^S}{\lambda_{j+1} - \lambda_j}$$

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

$$[\xi_D^*(\cdot)]_{i,j} = \begin{cases} [\xi_D(\cdot)]_{i,j}, & \text{if } d_i = d_j = \max_{1\le k\le n} d_k, \\ 0, & \text{else,} \end{cases} \tag{A.84}$$

and where ξ_D is (nonstandard) fractional Brownian motions defined on $[0, 1]$ such that $\text{Var}(\text{vec}(\xi_D(1))) =$

Ω (which follows from Theorem 4.8.2 of Giraitis et al. (2012) p.109). Hence,

$$\sum_{j=1}^{m} F_{2,T}^{j} \Rightarrow -\frac{1}{2}\operatorname{tr}\Bigg(\frac{\xi_{D,n}^{*}(\lambda_{j+1})^{S}\xi_{D,n}^{*}(\lambda_{j+1})^{S}}{\lambda_{j+1}} - \frac{\xi_{D,n}^{*}(\lambda_{j})^{S}\xi_{D,n}^{*}(\lambda_{j})^{S}}{\lambda_{j}} \tag{A.85}$$

$$+ \frac{(\xi_{D,n}^{*}(\lambda_{j+1}) - \xi_{D,n}^{*}(\lambda_{j}))^{S}(\xi_{D,n}^{*}(\lambda_{j+1}) - \xi_{D,n}^{*}(\lambda_{j}))^{S}}{\lambda_{j+1} - \lambda_{j}}\Bigg)$$

$$= -\frac{1}{2}\Bigg[\frac{\operatorname{vec}(\xi_{D,n}^{*}(\lambda_{j+1})^{S})' \operatorname{vec}(\xi_{D,n}^{*}(\lambda_{j+1})^{S})}{\lambda_{j+1}}$$

$$- \frac{\operatorname{vec}(\xi_{D,n}^{*}(\lambda_{j})^{S})' \operatorname{vec}(\xi_{D,n}^{*}(\lambda_{j})^{S})}{\lambda_{j}}$$

$$\times \left(\operatorname{vec}(\xi_{D,n}^{*}(\lambda_{j+1})^{S}) - \operatorname{vec}(\xi_{D,n}^{*}(\lambda_{j})^{S})\right)'$$

$$\times \frac{(\operatorname{vec}(\xi_{D,n}^{*}(\lambda_{j+1})^{S}) - \operatorname{vec}(\xi_{D,n}^{*}(\lambda_{j})^{S})}{(\lambda_{i+1} - \lambda_{i})}\Bigg]$$

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

$$\operatorname{vec}(\xi_{D,n}^{*}(\lambda_{j+1})^{S})' \operatorname{vec}(\xi_{D,n}^{*}(\lambda_{j+1})^{S}) = \operatorname{vec}(\xi_{D,n}^{*}(\lambda_{j+1}))' H' H \operatorname{vec}(\xi_{D,n}^{*}(\lambda_{j+1})) \tag{A.86}$$

$$\stackrel{d}{=} W_{D,n_b^*}^{*}(\lambda_{j+1})' H\Omega H' W_{D,n_b^*}^{*}(\lambda_{j+1}),$$

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

$$\sum_{j=1}^{m} F_{2,T}^{j} \Rightarrow -\frac{1}{2}\Bigg[\frac{W_{D,n_b^*}^{*}(\lambda_{j+1})' H'\Omega H W_{D,n_b^*}^{*}(\lambda_{j+1})}{\lambda_{j+1}} - \frac{W_{D,n_b^*}^{*}(\lambda_{j})' H'\Omega H W_{D,n_b^*}^{*}(\lambda_{j})}{\lambda_{j}} \tag{A.87}$$

$$- \frac{(W_{D,n_b^*}^{*}(\lambda_{j+1}) - W_{D,n_b^*}^{*}(\lambda_{j}))' H'\Omega H (W_{D,n_b^*}^{*}(\lambda_{j+1}) - W_{D,n_b^*}^{*}(\lambda_{j})}{\lambda_{j+1} - \lambda_{j}}\Bigg]$$

$$= (\lambda_j W_{D,n_b^*}^{*}(\lambda_{j+1}) - \lambda_{j+1} W_{D,n_b^*}^{*}(\lambda_{j}))' H'\Omega H$$

$$\times (\lambda_j W_{D,n_b^*}^{*}(\lambda_{j+1}) - \lambda_{j+1} W_{D,n_b^*}^{*}(\lambda_{j}))/(\lambda_j \lambda_{j+1}(\lambda_{j+1} - \lambda_j)).$$

By combining equations (A.73) and (A.87) we have shown the limiting distribution of our test.

□

A.6 Proof of Theorem 3

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 $(1-\alpha)\%$. $\square$

B Proof of 10 Properties

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

Property 11. *For each* $\delta \ \in (0,1]$

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

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

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

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

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

Furthermore, we have

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

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

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

Property 12. *For some* $\delta \in (0,1)$*, each* $\varepsilon > 0$*, there exists a* $B > 0$ *such that*

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

for all large T.

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

Property 13. *Let* $S_T = \{(\beta,\Sigma)\colon \|\beta-\beta^0\| \geq T^{-1/2+d}\log T$ *or* $\|\Sigma-\Sigma^0\| \geq T^{-1/2+d}\log T\}$. *For any* $\delta \in (0,1), D>0$ *and* $\varepsilon > 0$ *the following statement holds when* T *is large:*

$$\Pr\left(\sup_{k\geq\delta T}\sup_{(\beta,\Sigma)\in S_T} T^D\mathcal{L}(1,k;\beta,\Sigma) > 1\right) < \varepsilon. \tag{A.88}$$

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

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

where $\lambda_{\min}$ and $\lambda_{\max}$ denote the smallest and largest eigenvalues of Σ and the finite constants d_1, d_2 and d_3 are chosen in such a way that (β_0,Σ_0) is an inner point of $\overline{\Theta}_2$. As explained we first show (A.17) with the second supremum taken over $S_T \cap \overline{\Theta}_2$ which is compact. We decompose the segmential log likelihood as $\log\mathcal{L}(1,k;\beta,\Sigma) = \mathcal{L}_{1,T} + \mathcal{L}_{2,T}$, where

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

and

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

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

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

and

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

It has to be shown that

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

and

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

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

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

Applying Property 1 gives

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

Additionally we see

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

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

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

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

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

which implies

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

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

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

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

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

Then for any $a>0$ by the FCLT of Lemma A.2

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

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

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

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

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

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

Property 14. *Not needed.*

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

Property 15. *Let h_T and d_T be positive sequences such that h_T is nondecreasing, $d_T \to \infty$ and $\liminf_{T\to\infty}(h_T d_T^2)/T \geq h > 0$. Define $\overline{\Theta}_3 = \{(\beta, \Sigma)\colon \|\beta\| \leq p_1, \lambda_{\min}(\Sigma) \geq p_2, \lambda_{\max}(\Sigma) \leq p_3\}$, where p_1, p_2 and p_3 are arbitrary constants that satisfy $p_1 < \infty$, $0 < p_2 \leq p_3 < \infty$. Define $S_T = \{(\beta, \Sigma)\colon \|\beta - \beta^0\| \geq T^{-1/2+d}\log T$ or $\|\Sigma - \Sigma^0\| \geq T^{-1/2+d}\log T\}$. Then, for any $\varepsilon > 0$, there exists an $A > 0$, such that*

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

when T is large.

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

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

or

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

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

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

where

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

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

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

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

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

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

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

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

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

Now,

$$\begin{aligned}\sup_{k\geq Ah_T}\|\frac{1}{k}\sum_{t=1}^k x_t\Sigma^{-1}u_t\| &= \sup_{k\geq Ah_T}\|\frac{1}{k}\sum_{t=1}^k S'(I_n\otimes z_t)\Sigma^{-1}u_t\|\\ &\leq \sup_{k\geq Ah_T}\|\frac{1}{k}\sum_{t=1}^k (I_n\otimes z_t)u_t\|\|S'(\Sigma^{-1}\otimes I_n)\|. \qquad (A.89)\end{aligned}$$

From Lemma A.1 we have

$$P(\sup_{k\geq Ah_T}\|\frac{1}{k}\sum_{t=1}^k (I_n\otimes z_t)u_t\| > ab_T) \leq \frac{C_1}{Ah_T a^2 b_T} < \frac{2C_1}{Aa^2h} \qquad (A.90)$$

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

$$\|S'(\Sigma^{-1}\otimes I_n)\| \leq nC_2\sum_{i=1}^n \frac{1}{1+\lambda_i} \qquad (A.91)$$

for some $0<C_2<\infty$, which depends on the matrix S. Now, combining (A.89)-(A.91), we have, for any $\varepsilon>0$ that there exists an $\bar{A}>0$, such that with probability no less than $1-\varepsilon$,

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

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

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

Property 16. *With ν_T satisfying Assumption 6, for each β and Σ such that $\|\beta-\beta_0\|\leq Mv_T$*

and $\|\Sigma - \Sigma_0\| \leq M v_T$*, with* $M < \infty$*, we have*

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

Property 17. *Under the conditions of Property 6, we have*

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

Property 18. *We have*

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

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

Property 19. *Let* $T_1 = [aT]$ *for some* $a \in (0,1]$ *and let* $T_2 = [T^{1/2-d} v_T^{-1}]$*, where* v_T *satisfies Assumption 6. Consider*

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

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

Property 20. *Not needed.*

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

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

Table 4: Asymptotic critical values for the $UDmax\ LR_T$ test, using $\epsilon = 0.05$. These are obtained from 10,000 Monte Carlo replications with $1,000$ increments.

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

Table 5: Asymptotic critical values for the $UDmax\ LR_T$ test, using $\epsilon = 0.05$. These are obtained from 10,000 Monte Carlo replications with $1,000$ increments.

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

Table 6: Asymptotic critical values for the $UDmax\, LR_T$ test, using $\epsilon = 0.05$. These are obtained from 10,000 Monte Carlo replications with $1,000$ increments.