2018-05

Testing for common breaks in a multiple equations system

Testing for Common Breaks in a Multiple Equations System*

Tatsushi Oka† **Pierre Perron**‡

January 11, 2018

Abstract

The issue addressed in this paper is that of testing for common breaks across or within equations of a multivariate system. Our framework is very general and allows integrated regressors and trends as well as stationary regressors. The null hypothesis is that breaks in different parameters occur at common locations and are separated by some positive fraction of the sample size unless they occur across different equations. Under the alternative hypothesis, the break dates across parameters are not the same and also need not be separated by a positive fraction of the sample size whether within or across equations. The test considered is the quasi-likelihood ratio test assuming normal errors, though as usual the limit distribution of the test remains valid with non-normal errors. Of independent interest, we provide results about the rate of convergence of the estimates when searching over all possible partitions subject only to the requirement that each regime contains at least as many observations as some positive fraction of the sample size, allowing break dates not separated by a positive fraction of the sample size across equations. Simulations show that the test has good finite sample properties. We also provide an application to issues related to level shifts and persistence for various measures of inflation to illustrate its usefulness.

Keywords: change-point, segmented regressions, break dates, hypothesis testing, multiple equations systems.

JEL codes: C32

*We thank the Editor, Oliver Linton, an Associate Editor and three anonymous referees for their constructive comments, which improved the paper. We would like to thank Jushan Bai, Alastair Hall, Eiji Kurozumi, James Morley, Zhongjun Qu, Mototsugu Shintani, Denis Tkachenko, seminar participants at Boston University and participants at the 2009 Far East and South Asia Meeting of the Econometric Society for useful comments. We are also grateful to Douglas Sondak for advices on the computations. Oka gratefully acknowledges the financial support from Singapore Ministry of Education Academic Research Fund Tier 1 (FY2015-FRC3-003) and also gratefully acknowledges the financial support from Monash Business School.

†Department of Econometrics and Business Statistics, Monash University (`tatsushi.oka@monash.edu`).

‡Department of Economics, Boston University, USA (`perron@bu.edu`).

1 Introduction

Issues related to structural change have been extensively studied in the statistics and econometrics literature (see Csörgö and Horváth, 1997; Perron, 2006, for comprehensive reviews). In the last twenty years or so, substantial advances have been made in the econometrics literature to cover models at a level of generality that makes them relevant across time-series applications in the context of unknown change points. For example, Bai (1994, 1997) studies the least squares estimation of a single change point in regressions involving stationary and/or trending regressors. Bai and Perron (1998, 2003) extend the testing and estimation analysis to the case of multiple structural changes and present an efficient algorithm. Hansen (1992) and Kejriwal and Perron (2008) consider regressions with integrated variables. Andrews (1993) and Hall and Sen (1999) consider nonlinear models estimated by generalized method of moments. Bai (1995, 1998) studies structural changes in least absolute deviation regressions, while Qu (2008), Su and Xiao (2008) and Oka and Qu (2011) analyze structural changes in regression quantiles. Hall, Han, and Boldea (2012) and Perron and Yamamoto (2014, 2015) consider structural changes in linear models with endogenous regressors. Studies about structural changes in panel data models include Bai (2010), Kim (2011), Baltagi, Feng, and Kao (2016) and Qian and Su (2016) for linear panel data models and Breitung and Eickmeier (2011), Cheng, Liao, and Schorfheide (2016), Corradi and Swanson (2014), Han and Inoue (2015) and Yamamoto and Tanaka (2015) for factor models.

The literature on structural breaks in a multiple equations system includes Bai et al. (1998), Bai (2000) and Qu and Perron (2007), among others. Their analysis relies on a common breaks assumption, under which breaks in different basic parameters (regression coefficients and elements of the covariance matrix of the errors) occur at a common location or are separated by some positive fraction of the sample size (i.e., asymptotically distinct).[1] Bai et al. (1998) assume a single common break across equations for a multivariate system with stationary regressors and trends as well as for cointegrated systems. For the case of multiple common breaks, Bai (2000) analyzes vector autoregressive models for stationary variables and Qu and Perron (2007) cover multiple system equations, allowing for more general stationary regressors and arbitrary restrictions across parameters. Under the framework of Qu and Perron (2007), Kurozumi and Tuvaandorj (2011) propose model selection procedures for a system of equations with multiple common breaks and Eo and Morley (2015) consider a confidence set for the common break date based on inverting the likelihood ratio test. In this literature, it has been documented that common breaks allow more precise

[1] The concept of common breaks here is quite distinct from the notion of co-breaking or co-trending (e.g., Hatanaka and Yamada, 2003; Hendry and Mizon, 1998). In this literature, the focus is on whether some linear combination of series with breaks do not have a break, a concept akin to that of cointegration.

estimates of the break dates in multivariate systems. Given unknown break dates, however, an issue of interest for most applications concerns the validity of the assumption of common breaks.[2] To our knowledge, no test has been proposed to address this issue.

Our paper addresses three outstanding issues about testing for common breaks. First, we propose a quasi-likelihood ratio test under a very general framework.[3] We consider a multiple equations system under a likelihood framework with normal errors, though the limit distribution of the proposed test remains valid with non-normal, serially dependent and heteroskedastic errors. Our framework allows integrated regressors and trends as well as stationary regressors as in Bai et al. (1998) and also accommodates multiple breaks and arbitrary restrictions across parameters as in Qu and Perron (2007). Thus, our results apply for general systems of multiple equations considered in existing studies. A case not covered in our framework is when the regressors depend on the break date. This occurs when considering joint segmented trends and this issue was analyzed in Kim et al. (2017).

Second, we propose a test for common breaks not only across equations within a multivariate system, but also within an equation. As in Bai et al. (1998), the issue of common breaks is often associated with breaks occurring across equations, whereas one may want to test for common breaks in the parameters within a regression equation, whether a single equation or a system of multiple equations are considered. More precisely, the null hypothesis of interest is that some subsets of the basic parameters share one or more common break dates, so that each regime is separated by some positive fraction of the sample size. Under the alternative hypothesis, the break dates are not the same and also need not be separated by a positive fraction of the sample size, or be asymptotically distinct.

Third, we derive the asymptotic properties of the quasi-likelihood and the parameter estimates, allowing for the possibility that the break dates associated with different basic parameters may not be asymptotically distinct. This poses an additional layer of difficulty, since existing studies establish the consistency and rate of convergence of estimators only when the break dates are assumed to either have a common location or be asymptotically distinct, at least under the level of generality adopted here. Moreover, we establish the results in the presence of integrated regressors and trends as well as stationary regressors. This is by itself a noteworthy contribution. These asymptotic results will allow us to derive the limit

[2]The common breaks assumption is also used in the literature on panel data (e.g. Bai, 2010; Kim, 2011; Baltagi et al., 2016). In this paper, we consider a multiple equations system in which the number of equations are relatively small, and thus panel data models are outside our scope. However, testing for common breaks in a system with a large number of equations is an interesting avenue for future research.

[3]One may also consider other type of tests, such as LM-type tests. The literature on structural breaks, however, documents that even though LM-type tests have simple asymptotic representations, they tend to exhibit poor finite sample properties with respect to power. Thus, this paper focuses on the LR test (see Deng and Perron, 2008; Kim and Perron, 2009; Perron and Yamamoto, 2016, for instance).

distribution of our test statistic under the null hypothesis and also facilitate asymptotic power analyses under fixed and local alternatives. We can show that our test is consistent under fixed alternatives and also has non-trivial local power.

There is one additional layer of difficulty compared to Bai and Perron (1998) or Qu and Perron (2007). In their analysis, it is possible to transform the limit distribution so that it can be evaluated using a closed form solution and thus critical values can be tabulated. Here, no such solution is available and we need to obtain critical values for each case through simulations. This involves simulating the Wiener processes with consistent parameter estimates and evaluating each realization of the limit distribution with and without the restriction of common breaks. While it is conceptually straightforward and quick enough to be feasible for common applications, the procedure needs to be repeated many times to obtain the relevant quantities and can be quite computationally intensive. This is because we need to search over many possible combinations of all the permutations of the break locations for each replication of the simulations. To reduce the computational burden, we propose an alternative procedure based on the particle swarm optimization method developed by Eberhart and Kennedy (1995) with the Karhunen-Loève representation of stochastic processes. Our simulation results suggest that the test proposed has reasonably good size and power performance even in small samples under both computation procedures. Also, we apply our test to inflation series, following the work of Clark (2006) to illustrate its usefulness.

The remainder of the paper is as follows. Section 2 introduces the models with and without the common breaks assumption and describes the estimation methods under the quasi-likelihood framework. Section 3 presents the assumptions and asymptotic results including the asymptotic null distribution and asymptotic power analyses. Section 4 examines the finite sample properties of our procedure via Monte Carlo simulations. Section 5 presents an empirical application and Section 6 concludes. An appendix contains all the proofs.

2 Models and quasi-likelihood method

In this section, we first introduce models for a multiple equations system with and without common breaks. Subsequently, we describe the quasi-likelihood estimation method assuming normal errors and then propose the quasi-likelihood ratio test for common breaks. For illustration purpose, we also discuss some examples.

As a matter of notation, "$\xrightarrow{p}$" denotes convergence in probability, "$\xrightarrow{d}$" convergence in distribution and "$\Rightarrow$" weak convergence in the space $D[0,\infty)$ under the Skorohod topology. We use $\mathbb{R}$, $\mathbb{Z}$ and $\mathbb{N}$ to denote the set of all real numbers, all integers and all positive integers, respectively. For a vector x, we use $\|\cdot\|$ to denote the Euclidean norm (i.e., $\|x\| = \sqrt{x'x}$), while for a matrix A, we use the vector-induced norm (i.e., $\|A\| = \sup_{x \neq 0} \|Ax\| / \|x\|$). Define the L_r-norm of a random matrix X as $\|X\|_r = (\sum_i \sum_j E|X_{ij}|^r)^{1/r}$ for $r \geq 1$. Also, $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$ for any $a, b \in \mathbb{R}$. Let $\circ$ denote the Hadamard product (entry-wise product) and let $\otimes$ denote the Kronecker product. Define $\mathbb{1}_{\{\cdot\}}$ as the indicator function taking value one when its argument is true, and zero otherwise and e_i as a unit vector having 1 at the i^{th} entry and 0 for the others. We use the operator $\text{vec}(\cdot)$ to convert a matrix into a column vector by stacking the columns of the matrix and the operator $\text{tr}(\cdot)$ to denote the trace of a matrix. The largest integer not greater than $a \in \mathbb{R}$ is denoted by $[a]$ and the sign function is defined as $\text{sgn}(a) = -1, 0, 1$ if $a > 0$, $a = 0$ or $a < 0$, respectively.

2.1 The models with and without common breaks

Let the data consist of observations $\{(y_t, x_{tT})\}_{t=1}^T$, where y_t is an $n \times 1$ vector of dependent variables and x_{tT} is a $q \times 1$ vector of explanatory variables for $n, q \in \mathbb{N}$ with a subscript t indexing a temporal observation and T denoting the sample size. We allow the regressors x_{tT} to include stationary variables, time trends and integrated processes, while scaling by the sample size T so that the order of all components is the same. In what follows, we consider

$$x_{tT} = \left(z_t', \varphi(t/T)', T^{-1/2} w_t'\right)'.$$

Here, z_t, $\varphi(t/T)$ and w_t respectively denote vectors of stationary, trending and integrated variables with sizes being $q_z \times 1$, $q_\varphi \times 1$ and $q_w \times 1$, so that $q \equiv q_z + q_\varphi + q_w$.[4] Also,

$$\varphi(t/T) := [(t/T), (t/T)^2, \ldots, (t/T)^{q_\varphi}]' \quad \text{and} \quad w_t = w_{t-1} + u_{wt},$$

where w_0 is assumed, for simplicity, to be either $O_p(1)$ random variables or fixed finite constants, and u_{wt} is a vector of unobserved random variables with zero means. We label the variables z_t as $I(0)$ if the partial sums of the associated noise components satisfy a functional central limit theorem, while we label a variable as $I(1)$ if it is the accumulation of an $I(0)$ process. We discuss in more details the specific conditions in Section 3.

We first explain the case of common breaks through a model in which all of the parameters including those of the covariance matrix of the errors change, i.e., a pure structural change model. The model of interest is a multiple equations system with n equations and T time periods, excluding the initial conditions if lagged dependent variables are used as regressors. We denote the break dates in the system by $T_1, \ldots, T_m$ with m denoting the total number of structural changes and we use the convention that $T_0 = 0$ and $T_{m+1} = T$.

[4]The normalization is simply a theoretical device to reduce notational burden. Without it, we would need to handle different convergence rates of the estimates by introducing additional notations.

With a subscript j indexing a regime for $j = 1, ..., m+1$, the model is given by

$$y_t = (x'_{tT} \otimes I_n) S\beta_j + u_t, \quad \text{for } T_{j-1} + 1 \leq t \leq T_j, \tag{1}$$

where I_n is an $n \times n$ identity matrix, S is an $nq \times p$ selection matrix with full column rank, β_j is a $p \times 1$ vector of unknown coefficients, and u_t is an $n \times 1$ vector of errors having zero means and covariance matrix Σ_j.[5] The selection matrix S usually consists of elements that are 0 or 1 and, hence, specifies which regressors appear in each equation, although in principle it is allowed to have entries that are arbitrary constants. To ease notation, define the $n \times p$ matrix $X_{tT} := S'(x_{tT} \otimes I_n)$ so that (1) becomes, for $j = 1, ..., m+1$,

$$y_t = X'_{tT}\beta_j + u_t, \quad \text{for } T_{j-1} + 1 \leq t \leq T_j. \tag{2}$$

The set of basic parameters in the j^{th} regime consists of the coefficients β_j and the elements of the covariance matrix Σ_j, and we denote it by $\theta_j := (\beta_j, \Sigma_j)$ for each regime $j = 1, \ldots, m+1$. We use $\Theta_j \subset \mathbb{R}^p \times \mathbb{R}^{n \times n}$ to denote a parameter space for θ_j and we also define a product space $\Theta := \Theta_1 \times \cdots \times \Theta_{m+1}$ for $\theta := (\theta_1, \ldots, \theta_{m+1})$. In model (2), we allow for the imposition of a set of r restrictions through a function $R : \Theta \to \mathbb{R}^r$, given by

$$R(\theta) = 0. \tag{3}$$

Note that the equation in (3) can impose restrictions both within and across equations and regimes. Thus the model in (2) with some restrictions of the form (3) can accommodate structural break models other than a pure structural change model, such as partial structural change models in which a part of the basic parameters are constant across regimes. For a discussion of how general the framework is, see Qu and Perron (2007).

Next, we consider a pure structural change model allowing for the possibility that the break dates are not necessarily common across basic parameters. In the equations system with the $p \times 1$ vector of coefficients, we can assign each coefficient an index from 1 to p and we then group the p indices into disjoint subsets $\mathcal{G}_1, \ldots, \mathcal{G}_G \subset \{1, \ldots, p\}$ with G standing for the total number of groups, such that coefficients indexed by elements of $\mathcal{G}_g$ share the same break dates for each group $g = 1, \ldots, G$ and $\cup_{g=1}^{G} \mathcal{G}_g = \{1, ..., p\}$. Given a collection $\{\mathcal{G}_g\}_{g=1}^{G}$, we define, for $(g, j) \in \{1, \ldots, G\} \times \{1, ..., m+1\}$,

$$\beta_{gj} := \sum_{l \in \mathcal{G}_g} e_l \circ \beta_j. \tag{4}$$

Without loss of generality, we assume that the elements of the covariance matrix Σ_j have break dates that are common to those in the last group G. If none of the regression coefficients

[5] An example of models involving stationary and integrated variables is the dynamic ordinary least squares method to estimate cointegrating vectors (e.g. Saikkonen, 1991; Stock and Watson, 1993).

change at the same time as the elements of the covariance matrix Σ_j, then $\mathcal{G}_G$ is simply an empty set.[6] Here, we introduce groups of basic parameters to accommodate a wide range of empirical applications under our framework. Sometimes, researchers have economic models of interest or empirical knowledge that suggest specific parameter groups having common breaks. Even when one has no knowledge to form parameter groups, our analysis can be applied by considering all basic parameters as separate groups.

To denote the break date for regime j and group g, we use k_{gj} for $(g, j) \in \{1, \ldots, G\} \times \{1, \ldots, m\}$ with the convention that $k_{g0} = 0$ and $k_{g,m+1} = T$ for any $g = 1, \ldots, G$. Also, define a collection of break dates as,

$$\mathcal{K} := \{\mathcal{K}_1, \ldots, \mathcal{K}_G\} \text{ with } \mathcal{K}_g := (k_{g1}, \ldots, k_{gm}) \text{ for } g = 1, \ldots, G.$$

The regression model can be expressed as one depending on time-varying basic parameters according to the collection $\mathcal{K}$:

$$y_t = X'_{tT}\beta_{t,\mathcal{K}} + u_t, \tag{5}$$

where $\beta_{t,\mathcal{K}} := \sum_{g=1}^{G} \beta_{g,t,\mathcal{K}}$ and $E[u_t u_t'] = \Sigma_{t,\mathcal{K}}$ with

$$\beta_{g,t,\mathcal{K}} := \beta_{gj} \text{ for } k_{g,j-1} + 1 \leq t \leq k_{gj} \quad \text{and} \quad \Sigma_{t,\mathcal{K}} := \Sigma_j \text{ for } k_{G,j-1} + 1 \leq t \leq k_{Gj}, \tag{6}$$

for $(g, j) \in \{1, \ldots, G\} \times \{1, ..., m+1\}$. We also use $\theta_{t,\mathcal{K}} := (\beta_{t,\mathcal{K}}, \Sigma_{t,\mathcal{K}})$ to denote time-varying basic parameters depending on the collection of break dates $\mathcal{K}$. Thus the restrictions (3) can be imposed on the system (5) to accommodate more general models with structural breaks as in the one with common breaks.

In model (5), the basic parameters, break dates and the number of breaks are unknown and have to be estimated. To select the total number of structural changes, we can apply existing sequential testing procedures or information criteria. For example, if the breaks are common within each equation under both null and alternative hypotheses, but may differ across equations (see Example 1 below), sequential testing procedures proposed by Bai and Perron (1998) can be used to select the number of structural changes in each equation of a system (see Bai and Perron, 1998, p. 65, for a discussion of the statistical properties of such sequential procedures). In a similar way, the sequential testing procedure in Qu and Perron (2007) can be applied for sets of equations of a system separately. In order to handle more complex cases, we can alternatively use the Bayesian information criterion or the minimum description length principle as in Kurozumi and Tuvaandorj (2011), Lee (2000) and Aue and

[6] We assume that the different elements of the covariance matrix of the errors change at the same time. The results can be extended to the case where different parameters have distinct break dates, although additional notations would be needed. For the sake of notational simplicity, we only consider the case where the break dates are common within all elements of the covariance matrix.

Lee (2011). Because we use the likelihood framework, a likelihood function with a relevant penalty can be computed with the use of genetic algorithms (see Davis, 1991, for example), which consistently selects the number of structural breaks, as in Lee (2000) and Aue and Lee (2011). Thus, our analysis in what follows focuses on unknown basic parameters and breaks dates, given a total number of structural changes.

We use a 0 superscript to denote the true values of the parameters in both (2) and (5). Thus, the true basic parameters and break dates in (2) are denoted by $\{(\beta_j^0, \Sigma_j^0)\}_{j=1}^{m+1}$ and $\{T_j^0\}_{j=1}^m$, respectively, with the convention that $T_0^0 = 0$ and $T_{m+1}^0 = T$, whereas the ones in (5) are denoted by $\{\beta_{1j}^0, \ldots, \beta_{Gj}^0, \Sigma_j^0\}_{j=1}^{m+1}$ and $\mathcal{K}_g^0 := (k_{g1}^0, \ldots, k_{gm}^0)$ with $k_{g0}^0 = 0$ and $k_{g,m+1}^0 = T$ for $g = 1, \ldots, G$. Also let $\mathcal{K}^0 := \{\mathcal{K}_1^0, \ldots, \mathcal{K}_G^0\}$. Given a collection of break dates $\mathcal{K}$, let $\theta_{t,\mathcal{K}}^0 := (\beta_{t,\mathcal{K}}^0, \Sigma_{t,\mathcal{K}}^0)$ with a 0 superscript to denote time-varying true basic parameters θ^0, where $\theta^0 := (\theta_1^0, \ldots, \theta_{m+1}^0)$ with $\theta_j^0 := (\beta_j^0, \Sigma_j^0)$ for $j = 1, \ldots, m+1$.

2.2 The estimation and test under the quasi-likelihood framework

We consider the quasi-maximum likelihood estimation method with serially uncorrelated Gaussian errors for model (5) with restrictions given by (3).[7] Given the collection of break dates $\mathcal{K}$ and the basic parameters θ, the Gaussian quasi-likelihood function is defined as

$$L_T(\mathcal{K}, \theta) := \prod_{t=1}^{T} f(y_t | X_{tT}, \theta_{t,\mathcal{K}}),$$

where

$$f(y_t | X_{tT}, \theta_{t,\mathcal{K}}) := \frac{1}{(2\pi)^{n/2} |\Sigma_{t,\mathcal{K}}|^{1/2}} \exp\Big(-\frac{1}{2} \big\| \Sigma_{t,\mathcal{K}}^{-1/2} (y_t - X_{tT}' \beta_{t,\mathcal{K}}) \big\|^2 \Big).$$

To obtain maximum likelihood estimators, we impose a restriction on the set of permissible partitions with a trimming parameter $\nu > 0$ as follows[8]:

$$\Xi_\nu := \Big\{ \mathcal{K} : \min_{1 \le g \le G} \min_{1 \le j \le m+1} (k_{gj} - k_{g,j-1}) \ge T\nu \Big\}.$$

This set of permissible partitions ensures that there are enough observations between any break dates within the same group $\mathcal{K}_g$, while it accommodates the possibility that the break dates across different groups are not separated by a positive fraction of the sample size.

We propose a test for common breaks under the quasi-likelihood framework. The null hypothesis of common breaks in model (2) can be stated as

$$H_0 : \mathcal{K}_{g_1}^0 = \mathcal{K}_{g_2}^0 \quad \text{for all } g_1, g_2 \in \{1, \ldots, G\}, \tag{7}$$

[7] Our framework includes OLS-based estimation by setting the covariance matrix to be an identity matrix.

[8] For the asymptotic analysis, the trimming value ν can be an arbitrary small constant such that a positive fraction of the sample size $T\nu$ diverges at rate T.

and the alternative hypothesis is

$$H_1 : \mathcal{K}^0_{g_1} \neq \mathcal{K}^0_{g_2} \quad \text{for some } g_1, g_2 \in \{1, \ldots, G\}. \tag{8}$$

The set of permissible partitions under the null hypothesis can be expressed as

$$\Xi_{\nu,H_0} := \{\mathcal{K} \in \Xi_\nu : \mathcal{K}_1 = \cdots = \mathcal{K}_G\}.$$

The test considered is simply the quasi-likelihood ratio test that compares the values of the likelihood function with and without the common breaks restrictions. The quasi-maximum likelihood estimates under the null hypothesis, denoted by $(\widetilde{\mathcal{K}}, \widetilde{\theta})$, can be obtained from the following maximization problem with a restricted set of candidate break dates:

$$(\widetilde{\mathcal{K}}, \widetilde{\theta}) := \underset{(\mathcal{K},\theta)\in\Xi_{\nu,H_0}\times\Theta}{\arg\max} \log L_T(\mathcal{K}, \theta) \quad \text{s.t.} \quad R(\theta) = 0,$$

where $\widetilde{\mathcal{K}} := (\widetilde{\mathcal{K}}_1, \ldots, \widetilde{\mathcal{K}}_G)$ with $\widetilde{\mathcal{K}}_g := (\widetilde{k}_1, \ldots, \widetilde{k}_m)$ for all $g = 1, \ldots, G$, $\widetilde{\theta} := (\widetilde{\beta}, \widetilde{\Sigma})$ with $\widetilde{\beta} := (\widetilde{\beta}_1, \ldots, \widetilde{\beta}_{m+1})$ and $\widetilde{\Sigma} := (\widetilde{\Sigma}_1, \ldots, \widetilde{\Sigma}_{m+1})$. Also, the quasi-maximum likelihood estimates under the alternative, denoted by $(\hat{\mathcal{K}}, \hat{\theta})$, are obtained from the following problem:

$$(\hat{\mathcal{K}}, \hat{\theta}) := \underset{(\mathcal{K},\theta)\in\Xi_\nu\times\Theta}{\arg\max} \log L_T(\mathcal{K}, \theta) \quad \text{s.t.} \quad R(\theta) = 0, \tag{9}$$

where $\hat{\mathcal{K}} := (\hat{\mathcal{K}}_1, \ldots, \hat{\mathcal{K}}_G)$ with $\hat{\mathcal{K}}_g := (\hat{k}_{g1}, \ldots, \hat{k}_{gm})$ for $g = 1, \ldots, G$, $\hat{\theta} := (\hat{\beta}, \hat{\Sigma})$ with $\hat{\beta} := (\hat{\beta}_1, \ldots, \hat{\beta}_{m+1})$ and $\hat{\Sigma} := (\hat{\Sigma}_1, \ldots, \hat{\Sigma}_{m+1})$. Using the estimates $\hat{\theta}$, we can define $\hat{\beta}_{gj}$ as in (4) and $\hat{\theta}_{t,\mathcal{K}} := (\hat{\beta}_{t,\mathcal{K}}, \hat{\Sigma}_{t,\mathcal{K}})$ as in (6) given a collection of break dates $\mathcal{K}$.

We define the quasi-likelihood ratio test for common breaks as

$$CB_T := 2\{\log L_T(\hat{\mathcal{K}}, \hat{\theta}) - \log L_T(\widetilde{\mathcal{K}}, \widetilde{\theta})\}.$$

For the asymptotic analysis, it is useful to employ a normalization by using the log-likelihood function evaluated at the true parameters $(\mathcal{K}^0, \theta^0)$ and we consider

$$CB_T = 2\{\ell_T(\hat{\mathcal{K}}, \hat{\theta}) - \ell_T(\widetilde{\mathcal{K}}, \widetilde{\theta})\},$$

where $\ell_T(\mathcal{K}, \theta) := \log L_T(\mathcal{K}, \theta) - \log L_T(\mathcal{K}^0, \theta^0)$ for any $(\mathcal{K}, \theta) \in \Xi_\nu \times \Theta$. The common break test CB_T depends on two log-likelihoods with and without the common breaks assumption. The break date estimates $\widetilde{\mathcal{K}}$ under the null hypothesis are required to either have common locations or be separated by a positive fraction of the sample size. Without common breaks restrictions, however, the break date estimates $\hat{\mathcal{K}}$ are simply allowed to be distinct but not necessarily separated by a positive fraction of the sample size across groups. This will be important since the setup of Bai (2000) and Qu and Perron (2007) requires the maximization to be taken over asymptotically distinct elements and their proof for the convergence rate of the estimates relies on this premise. Hence, we will need to provide a detailed proof of the convergence rate under this less restrictive maximization problem (see Section 3).

2.3 Examples

Given that the notation is rather complex, it is useful to illustrate the framework explained in the preceding subsection via examples.

Example 1 (changes in intercepts): We consider a two-equations system of autoregressions with structural changes in intercepts, for $j = 1, 2$,

$$y_{1t} = \mu_{1j} + \alpha_1 y_{1,t-1} + u_{1t} \quad \text{and} \quad y_{2t} = \mu_{2j} + \alpha_2 y_{2,t-1} + u_{2t}, \quad \text{for } T_{j-1} + 1 \leq t \leq T_j,$$

where $(u_{1t}, u_{2t})'$ have a covariance matrix Σ. In this model, the basic parameters except the intercepts are assumed to be constant and the intercepts change at a common break date T_1. In equation (1), we have $x_{tT} = (1, y_{1,t-1}, y_{2,t-1})'$, $\beta_j = (\mu_{1j}, \alpha_{1j}, \mu_{2j}, \alpha_{2j})'$ and $E[u_t u_t'] = \Sigma_j$. The selection matrix $S =< s_{ij} >$ is a 6×4 matrix taking value 1 at the entries s_{11},s_{22}, s_{33} and s_{64} and 0 elsewhere. Also, by setting $R(\theta) = \big(\alpha_{11} - \alpha_{12}, \alpha_{21} - \alpha_{22}, \text{vec}(\Sigma_1) - \text{vec}(\Sigma_2)\big)' = 0$ in (3), we impose restrictions on the basic parameters so that a partial structural change model is considered with no changes in the autoregressive parameters and the covariance matrix of the errors. On the other hand, when we allow the possibility that break dates can differ across the two equations as in the model (5), we consider the following system, for $j = 1, 2$,

$$\begin{aligned} y_{1t} &= \mu_{1j} + \alpha_1 y_{1,t-1} + u_{1t}, \quad \text{for } k_{1,j-1} + 1 \leq t \leq k_{1j}, \\ y_{2t} &= \mu_{2j} + \alpha_2 y_{2,t-1} + u_{2t}, \quad \text{for } k_{2,j-1} + 1 \leq t \leq k_{2j}. \end{aligned}$$

Here, we separate β_j into $\beta_{1j} = (\mu_{1j}, \alpha_{1j}, 0, 0)'$ and $\beta_{2j} = (0, 0, \mu_{2j}, \alpha_{2j})'$, so that we can set $\mathcal{G}_1 = \{1, 2\}$ and $\mathcal{G}_2 = \{3, 4\}$. We have two possibly distinct break dates k_{11} and k_{21} for the parameter groups $\{\beta_{1j}\}_{j=1}^2$ and $\{(\beta_{2j}, \Sigma_j)\}_{j=1}^2$, respectively. We address the issue of testing the null hypothesis $H_0 : k_{11} = k_{21}$ against the alternative hypothesis $H_1 : k_{11} \neq k_{21}$.

Example 2 (a single equation model): Consider a single equation model:

$$y_{1t} = \mu + \alpha_j z_{1,t} + \gamma_j (t/T) + \rho_j T^{-1/2} w_{1t} + u_{1t},$$

for $T_{j-1} + 1 \leq t \leq T_j$ with $j = 1, 2, 3$, where u_{1t} denotes the error term with $E[u_{1t}] = 0$ and $E[u_{1t}^2] = \sigma_j^2$. In this example, the basic parameters other than the intercepts have two structural changes. Under model (2) with break dates T_1 and T_2, we have $x_{tT} = (1, z_{1t}, t/T, T^{-1/2} w_{1t})'$, $S = I_4$, $\beta_j = (\mu_j, \alpha_j, \gamma_j, \rho_j)'$. Restrictions of the form (3) are imposed by the function $R(\theta) = (\mu_1 - \mu_2, \mu_2 - \mu_3)' = 0$. We consider a test for common breaks against the alternative that all coefficients change at distinct break dates, while the coefficient ρ_j and the variance σ_j^2 change at the same break dates. In this case, we separate β_j into three vectors $\beta_{1j} = (\mu_j, \alpha_j, 0, 0)$, $\beta_{2j} = (0, 0, \gamma_j, 0)$ and $\beta_{3j} = (0, 0, 0, \rho_j)$. For these parameters groups, we assign a set of break dates $\mathcal{K}_g = (k_{g1}, k_{g2})$ for $g = 1, \ldots, 3$ and we set $\mathcal{G}_1 = \{1, 2\}$, $\mathcal{G}_2 = \{3\}$ and $\mathcal{G}_3 = \{4\}$. The break dates for the last group, $\mathcal{K}_3$, are also the ones for the variance. This example shows that our framework can accommodate common breaks not only across equations in a system but also within an equation.

3 Asymptotic results

This section presents the relevant asymptotic results. We first provide the convergence rates of the estimates of the break dates and the basic parameters, allowing for the possibility that the break dates of different basic parameters may not be asymptotically distinct. This condition is substantially less restrictive than the ones usually assumed in the existing literature and particularly includes the assumption of common breaks as a special case. Next, we provide the limiting distribution of the quasi-likelihood ratio test for common breaks under the null hypothesis. Finally, we provide asymptotic power analyses of the test under a fixed alternative as well as a local one. Our result shows non-trivial asymptotic power.

3.1 The rate of convergence of the estimates.

We consider the case where we obtain the quasi-likelihood estimates $(\hat{\mathcal{K}}, \hat{\theta})$ as in (9), using the observations $\{(y_t, x_{tT})\}_{t=1}^{T}$ generated by model (5) with collections of true parameter values $(\mathcal{K}^0, \theta^0)$. The results presented in this subsection can apply for the estimates obtained from the model under the null hypothesis since it is a special case of the setup adopted. To obtain the asymptotic results, the following assumptions are imposed.

Assumptions:

A1. There exists a constant $k_0 > 0$ such that for all $k > k_0$, the minimum eigenvalues of the matrices $k^{-1} \sum_{t=s}^{s+k} x_{tT} x_{tT}'$ are bounded away from zero for every $s = 1, \ldots, T-k$.

A2. Define the sigma-algebra $\mathcal{F}_t := \sigma(\{z_s, u_{ws}, \eta_s\}_{s \leq t})$ for $t \in \mathbb{Z}$, where $\eta_s := (\Sigma^0_{s,\mathcal{K}^0})^{-1/2} u_s$. (a) Define $\zeta_t := (z_t', u_{wt}')'$ and let z_t include a constant term. The sequence $\{\zeta_t \otimes \eta_t, \mathcal{F}_t\}_{t \in \mathbb{Z}}$ forms a strongly mixing (α-mixing) sequence with size $-(4+\delta)/\delta$ for some $\delta \in (0, 1/2)$ and satisfies $E[z_t \otimes \eta_t] = 0$ and $\sup_{t \in \mathbb{Z}} \|\zeta_t \otimes \eta_t\|_{4+\delta} < \infty$. (b) It is also assumed that $\{\eta_t \eta_t' - I_n\}_{t \in \mathbb{Z}}$ satisfies the same mixing and moment conditions as in (a). (c) The sequence $\{w_0 \otimes \eta_t\}_{t \in \mathbb{Z}}$ forms a strong mixing sequence as in (a) with $\sup_{t \in \mathbb{Z}} \|w_0 \otimes \eta_t\|_{4+\delta} < \infty$ and the initial condition w_0 is $\mathcal{F}_0$-measurable.

A3. The collection of the true break dates $\mathcal{K}^0$ is included in Ξ_ν and satisfies $k^0_{gj} = [T\lambda^0_{gj}]$ for every $(g, j) \in \{1, \ldots, G\} \times \{1, ..., m\}$, where $0 < \lambda^0_{g1} < \cdots < \lambda^0_{gm} < 1$.

A4. For every parameter group g and regime j, there exists a $p \times 1$ vector δ_{gj} and an $n \times n$ matrix Φ_j such that $\beta^0_{g,j+1} - \beta^0_{gj} = v_T \delta_{gj}$ and $\Sigma^0_{j+1} - \Sigma^0_j = v_T \Phi_j$, where both δ_{gj} and Φ_j are independent of T, and $v_T > 0$ is a scalar satisfying $v_T \to 0$ and $\sqrt{T} v_T / \log T \to \infty$ as $T \to \infty$. Let $\delta_j := \sum_{g=1}^{G} \delta_{gj}$ for $j = 1, \ldots, m+1$.

A5. The true basic parameters (β^0, Σ^0) belong to the compact parameter space

$$\Theta := \left\{ \theta : \max_{1 \leq j \leq m+1} \|\beta_j\| \leq c_1,\ c_2 \leq \min_{1 \leq j \leq m+1} \lambda_{\min}(\Sigma_j),\ \max_{1 \leq j \leq m+1} \lambda_{\max}(\Sigma_j) \leq c_3 \right\},$$

for some constants $c_1 < \infty$, $0 < c_2 \leq c_3 < \infty$, where $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ denote the smallest and largest eigenvalues of the matrix in its argument, respectively.

Assumption A1 ensures that there is no local collinearity problem so that a standard invertibility requirement holds if the number of observations in some sub-sample is greater than k_0, not depending on T. Assumption A2 determines the dependence structure of $\{\zeta_t \otimes \eta_t\}$, $\{\eta_t \eta_t' - I_n\}$ and $\{w_0 \otimes \eta_t\}$ to guarantee that they are short memory processes and have bounded fourth moments. The assumptions are imposed to obtain a functional central limit theorem and a generalized Hájek and Rényi (1955) type inequality that allow us to derive the relevant convergence rates. Assumption A2 also specifies that the stationary regressors are contemporaneously uncorrelated with the errors and that a constant term is included in z_t. The former is a standard requirement to obtain consistent estimates and the latter is for notational simplicity since the results reported below are the same without a constant term.[9,10] It is important to note that no assumption is imposed on the correlation between the innovations to the $I(1)$ regressors and the errors. Hence, we allow endogenous $I(1)$ regressors. Assumption A3 ensures that $\lambda^0_{gj} - \lambda^0_{g,j-1} > \nu$ holds for every pair of group and regime (g, j) and thus implies asymptotically distinct breaks within each parameter group, but not necessarily across groups. Assumption A4 implies a shrinking shifts asymptotic framework whereby the magnitudes of the shifts converge to zero as the sample size increases. This condition is necessary to develop a limit distribution theory for the estimates of the break dates that does not depend on the exact distributions of the regressors and the errors, as commonly used in the literature (e.g., Bai, 1997; Bai and Perron, 1998; Bai et al., 1998). Assumption A5 implies that the data are generated by a model with a finite conditional mean and innovations having a non-degenerate covariance matrix.

[9]One can use the usual ordinary least squares framework to simply estimate the break dates and test for structural change even in the presence of the correlation between the stationary regressors and the errors (see Perron and Yamamoto, 2015). One may also use a two-stage least squares method if relevant instrumental variables are available (see Hall et al., 2012; Perron and Yamamoto, 2014).

[10]When a constant term is not included in z_t, in contrast to Assumption A2, one additionally needs to assume that the sequence $\{\eta_t\}_{t \in \mathbb{Z}}$ satisfies the same mixing and moment conditions as in Assumption A2(a).

As stated above, the break dates are estimated from a set Ξ_ν, which requires candidate break dates to be separated by some fraction of the sample size only within parameter groups. Thus, we cannot appeal to the results in Bai (2000) and Qu and Perron (2007) about the rate of convergence of the estimates, and more general results are needed. The following theorem presents results about the convergence rates of the estimates.

Theorem 1. *Suppose that Assumptions A1-A5 hold. Then,*
(a) uniformly in $(g, j) \in \{1, ..., G\} \times \{1, ..., m\}$,

$$v_T^2(\hat{k}_{gj} - k_{gj}^0) = O_p(1),$$

(b) uniformly in $(g, j) \in \{1, ..., G\}\times\{1, ..., m+1\}$,

$$\sqrt{T}(\hat{\beta}_{gj} - \beta_{gj}^0) = O_p(1) \;\; \text{and} \;\; \sqrt{T}(\hat{\Sigma}_j - \Sigma_j^0) = O_p(1).$$

This theorem establishes the convergence rates obtained in Bai and Perron (1998), Bai et al. (1998), Bai (2000) and Qu and Perron (2007), while assuming less restrictive conditions regarding the optimization problem and the time-series properties of the regressors.

The importance of these results is that they will allow us to analyze the properties of our test under compact sets for the parameters, namely, for some $M > 0$,

$$\begin{aligned}
\bar{\Xi}_M &:= \left\{\mathcal{K} \in \Xi_\nu : \max_{1\leq g\leq G} \max_{1\leq j\leq m} |k_{gj} - k_{gj}^0| \leq M v_T^{-2}\right\} \\
\bar{\Theta}_M &:= \left\{\theta \in \Theta : \max_{1\leq g\leq G} \max_{1\leq j\leq m+1} \|\beta_{gj} - \beta_{gj}^0\| \leq M T^{-1/2}, \max_{1\leq j\leq m+1} \|\Sigma_j - \Sigma_j^0\| \leq M T^{-1/2}\right\}.
\end{aligned}$$

We also have a result that expresses the restricted likelihood in two parts: one that involves only the break dates and the true values of the coefficients; the other involving the true values of the break dates, the basic parameters and the restrictions. Thus, asymptotically the estimates of the break dates are not affected by the restrictions imposed on the coefficients, while the limiting distributions of these estimates are influenced by the restrictions.

Theorem 2. *Suppose that Assumptions A1-A5 hold. Then,*

$$\sup_{(\mathcal{K},\theta)\in\bar{\Xi}_M\times\bar{\Theta}_M} \ell_{T,R}(\mathcal{K}, \theta) = \sup_{\mathcal{K}\in\bar{\Xi}_M} \ell_T(\mathcal{K}, \theta^0) + \sup_{\theta\in\bar{\Theta}_M} \ell_{T,R}(\mathcal{K}^0, \theta) + o_p(1), \tag{10}$$

where $\ell_{T,R}(\mathcal{K}, \theta) := \ell_T(\mathcal{K}, \theta) + \gamma' R(\theta)$ *with a Lagrange multiplier* γ.

The result in Theorem 2 implies that when analyzing the asymptotic properties of the break date estimates, one can ignore the restrictions in (3). This will prove especially convenient to obtain the limit distribution of our test. Since the quasi-likelihood ratio test can be expressed as a difference of two normalized log likelihoods evaluated at different break dates, the second term on the right-hand side of (10) is canceled out in the test statistic. The result in Theorem 2 has been obtained in Bai (2000) for vector autoregressive models and Qu and Perron (2007) for more general stationary regressors, when break dates are assumed to either have a common location or be asymptotically distinct. We establish the results, allowing for the possibility that the break dates associated with different basic parameters may not be asymptotically distinct, and thus expand the scope of prior work such as Bai et al. (1998), Bai (2000) and Qu and Perron (2007).

3.2 The limit distribution of the likelihood ratio test

We now establish the limit distribution of the quasi-likelihood ratio test under the null hypothesis of common breaks in (7). To this end, let the data consist of the observations $\{(y_t, x_{tT})\}_{t=1}^T$ from model (2) with true basic parameters $\theta^0 = (\beta^0, \Sigma^0)$ and true break dates $\mathcal{T}^0$ consisting of $T_1^0, \ldots, T_m^0$. Theorem 1(a) shows that, uniformly in $(g, j) \in \{1, \ldots, G\} \times \{1, \ldots, m\}$, there exists a sufficiently large M such that $|\hat{k}_{gj} - T_j^0| \leq M v_T^{-2}$ and $|\widetilde{k}_j - T_j^0| \leq M v_T^{-2}$ with probability approaching 1. This implies that we can restrict our analysis to an interval centered at the true break T_j^0 with length $2Mv_T^{-2}$ for each regime $j \in \{1, \ldots, m\}$. More precisely, given a sufficiently large M, we have that $\theta^0_{t,\hat{\mathcal{K}}} = \theta^0_{t,\mathcal{T}^0}$ and $\theta^0_{t,\widetilde{\mathcal{K}}} = \theta^0_{t,\mathcal{T}^0}$ for all $t \notin \cup_{j=1}^m [T_j^0 - Mv_T^{-2}, T_j^0 + Mv_T^{-2}]$, with probability approaching 1. This follows since the break dates estimates are asymptotically in neighborhoods of the true break dates; hence that there are some miss-classification of regimes around the neighborhoods, while the regimes are correctly classified outside of the neighborhoods. This together with Theorem 2 yields that, under the null hypothesis specified by (7),

$$\begin{aligned} CB_T &= 2 \max_{\mathcal{K} \in \bar{\Xi}_M} \sum_{j=1}^m \sum_{\underline{k}_j+1}^{\bar{k}_j} \left\{ \log f(y_t | X_{tT}, \theta^0_{t,\mathcal{K}}) - \log f(y_t | X_{tT}, \theta^0_{t,\mathcal{T}^0}) \right\} \\ &\quad -2 \max_{\mathcal{K} \in \bar{\Xi}_{M,H_0}} \sum_{j=1}^m \sum_{\underline{k}_j+1}^{\bar{k}_j} \left\{ \log f(y_t | X_{tT}, \theta^0_{t,\mathcal{K}}) - \log f(y_t | X_{tT}, \theta^0_{t,\mathcal{T}^0}) \right\} + o_p(1), \end{aligned}$$

where $\overline{k}_j := \max\{k_{1j}, \ldots, k_{Gj}, T_j^0\}$, $\underline{k}_j := \min\{k_{1j}, \ldots, k_{Gj}, T_j^0\}$, and $\Xi_{M,H_0} = \Xi_M \cap \Xi_{\eta,H_0}$. Under the null hypothesis, the true break dates $T_1^0, \ldots, T_m^0$ are separated by some positive fraction of the sample size and we can obtain the limit distribution of the common break test by separately analysing terms of the test for each neighborhood of the true break date. We consider a shrinking framework under which the break date estimates $\hat{k}_{gj}$ and $\widetilde{k}_j$ diverge to ∞ as v_T decreases and thus an application of a Functional Central Limit Theorem for each neighborhood yields a limit distribution of the test which does not depend on the exact distributions. To derive the limit distribution, we make the following additional assumptions.

Assumptions:

A6. The matrix $(\Delta T_j^0)^{-1}\sum_{t=T_{j-1}^0+1}^{T_j^0} x_{tT}x_{tT}$ converges to a (possibly) random matrix not necessarily the same for all $j = 1,\dots,m+1$, as $\Delta T_j^0 := (T_j^0 - T_{j-1}^0) \to \infty$. Also, $(\Delta T_j^0)^{-1}\sum_{t=T_{j-1}^0+1}^{T_{j-1}^0+[s\Delta T_j^0]} z_t \xrightarrow{p} s\mu_{z,j}$ and $(\Delta T_j^0)^{-1}\sum_{t=T_{j-1}^0+1}^{T_{j-1}^0+[s\Delta T_j^0]} z_t z_t' \xrightarrow{p} sQ_{zz,j}$ uniformly in $s \in [0,1]$ as $\Delta T_j^0 \to \infty$, where $Q_{zz,j}$ is a non-random positive definite matrix.

A7. Define $S_{k,j}(l) := \sum_{T_{j-1}^0+l+1}^{T_{j-1}^0+l+k}(\zeta_t \otimes \eta_t)$ for $k,l \in \mathbb{N}$ and for $j = 1,...,m+1$. (i) If $\{\zeta_t \otimes \eta_t\}_{t\in\mathbb{Z}}$ is weakly stationary within each segment, then, for any vector $e \in \mathbb{R}^{(q_z+q_w)n}$ with $\|e\| = 1$, $\mathrm{var}\big(e'S_{k,j}(0)\big) \geq v(k)$ for some function $v(k) \to \infty$ as $k \to \infty$. (ii) If $\{\zeta_t \otimes \eta_t\}_{t\in\mathbb{Z}}$ is not weakly stationary within each segment, we additionally assume that there is a positive definite matrix $\Omega = [w_{i,s}]$ such that for any $i, s \in \{1,...,p\}$, we have, uniformly in ℓ, $\left|k^{-1}E\big[\big(S_{k,j}(\ell)\big)_i\big(S_{k,j}(\ell)\big)_s\big] - w_{i,s}\right| \leq k^{-\psi}$, for some $C > 0$ and for some $\psi > 0$. We also assume the same conditions for $\{\eta_t\eta_t' - I_n\}_{t\in\mathbb{Z}}$.

A8. Let $V_{T,w}(r) := T^{-1/2}\sum_{t=1}^{[Tr]} u_{wt}$ for $r \in [0,1]$. $V_{T,w}(\cdot) \Rightarrow \mathbb{V}_w(\cdot)$, where $\mathbb{V}_w(\cdot)$ is a Wiener processes having a covariance function $\mathrm{cov}(\mathbb{V}_w(r), \mathbb{V}_w(s)) = (r \wedge s)\Omega_w$ for $r, s \in [0,1]$ with a positive definite matrix $\Omega_w := \lim_{T\to\infty}\mathrm{var}\big(T^{-1/2}\sum_{t=1}^{T} u_{wt}\big)$.

A9. For all $1 \leq s,t \leq T$, (a) $E[(z_t \otimes \eta_t)w_s'] = 0$, (b) $E[(z_t \otimes \eta_t)\mathrm{vec}(\eta_s\eta_s')'] = 0$, and (c) $E[(u_{zt} \otimes \eta_t)\mathrm{vec}(\eta_s\eta_s')'] = 0$.

Assumption A6 rules out trending variables in the stationary regressors z_t. Assumption A7 is mild in the sense that the conditions allow for substantial conditional heteroskedasticity and autocorrelation. It can be shown to apply to a large class of linear processes including those generated by all stationary and invertible ARMA models. This assumption is useful to describe the asymptotic behavior of the test and in particular to characterize the limit distribution. Here, we introduce some processes used later. For each $j = 1,\dots,m$, let $\mathbb{V}_{z\eta,j}^{(1)}(\cdot)$ and $\mathbb{V}_{z\eta,j}^{(2)}(\cdot)$ be Brownian motions defined on the space $D[0,\infty)^{nq}$ with zero means and covariance functions given by, for $l = 1,2$ and for $s_1, s_2 > 0$,

$$E\big[\mathbb{V}_{z\eta,j}^{(l)}(s_1)\mathbb{V}_{z\eta,j}^{(l)}(s_2)'\big] = (s_1 \wedge s_2)\lim_{T\to\infty}\mathrm{var}\big(\bar{V}_{T,z\eta,j}^{(l)}\big),$$

where $\bar{V}_{T,z\eta,j}^{(1)} := (\Delta T_j^0)^{-1/2}\sum_{t=T_{j-1}^0+1}^{T_j^0}(z_t \otimes \eta_t)$ and $\bar{V}_{T,z\eta,j}^{(2)} := (\Delta T_{j+1}^0)^{-1/2}\sum_{t=T_j^0+1}^{T_{j+1}^0}(z_t \otimes \eta_t)$. Similarly, define $\mathbb{V}_{\eta\eta,j}^{(1)}(\cdot)$ and $\mathbb{V}_{\eta\eta,j}^{(2)}(\cdot)$ as Brownian motions defined on the space $D[0,\infty)^{n^2}$ with zero means and covariance functions given by, for $l = 1,2$ and for $s_1, s_2 > 0$,

$$E\big[\mathrm{vec}\big(\mathbb{V}_{\eta\eta,j}^{(l)}(s_1)\big)\mathrm{vec}\big(\mathbb{V}_{\eta\eta,j}^{(l)}(s_2)\big)'\big] = (s_1 \wedge s_2)\lim_{T\to\infty}\mathrm{var}\big\{\mathrm{vec}\big(\bar{V}_{T,\eta\eta,j}^{(l)}\big)\big\},$$

where $\bar{V}^{(1)}_{T,\eta\eta,j} := (\Delta T^0_j)^{-1/2} \sum_{t=T^0_{j-1}+1}^{T^0_j} (\eta_t \eta_t' - I_n)$ and $\bar{V}^{(2)}_{T,\eta\eta,j} := (\Delta T^0_{j+1})^{-1/2} \sum_{t=T^0_j+1}^{T^0_{j+1}} (\eta_t \eta_t' - I_n)$. We define the following two-sided Brownian motions

$$\mathbb{V}_{z\eta,j}(s) := \begin{cases} \mathbb{V}^{(1)}_{z\eta,j}(-s), & s \leq 0 \\ \mathbb{V}^{(2)}_{z\eta,j}(s), & s > 0 \end{cases} \quad \text{and} \quad \mathbb{V}_{\eta\eta,j}(s) := \begin{cases} \mathbb{V}^{(1)}_{\eta\eta,j}(-s), & s \leq 0 \\ \mathbb{V}^{(2)}_{\eta\eta,j}(s), & s > 0. \end{cases}$$

Under Assumption A2, z_t is assumed to include a constant term and the process $\mathbb{V}^{(l)}_{z\eta,j}(\cdot)$ includes some process depending purely on $\{\eta_t\}$. We denote it by $\mathbb{V}^{(l)}_{\eta,j}(\cdot)$ for each $l = 1, 2$ and also define a two-sided Brownian motion, denoted by $\mathbb{V}_{\eta,j}(\cdot)$, as before.

Assumption A8 requires the integrated regressors to follow a homogeneous distribution throughout the sample. Allowing for heterogeneity in the distribution of the errors underlying the $I(1)$ regressors would be considerably more difficult, since we would, instead of having the limit distribution in terms of standard Wiener processes, have time-deformed Wiener processes according to the variance profile of the errors through time; see, e.g., Cavaliere and Taylor (2007). This would lead to important complications given that, as shown below, the limit distribution of the estimates of the break dates depends on the whole time profile of the limit Wiener processes. It is possible to allow for trends in the $I(1)$ regressors. The limiting distributions of the test to be derived will remain valid under different Wiener processes (see Hansen, 1992). The positive definiteness of the matrix Ω_w rules out cointegration among the $I(1)$ regressors and is needed to ensure a set of regressors that has a positive definite limit.

Assumption A9 is quiet mild and is sufficient but not necessary to obtain a manageable limit distribution of the test. It requires the independence of most Wiener processes described above. Condition (a) ensures that the autocovariance structure of the $I(0)$ regressors and the errors are uncorrelated with the $I(1)$ variables. This guarantees that $\mathbb{V}_{z\eta,j}(\cdot)$ and $\mathbb{V}_{w,j}(\cdot)$ are uncorrelated and thus independent because of Gaussianity. Without these conditions, the analysis would be much more complex. Similarly, the conditions (b) and (c) imply the independence between $\mathbb{V}_{z\eta,j}(\cdot)$ and $\mathbb{V}_{\eta\eta}(\cdot)$. See Kejriwal and Perron (2008) for more details.

In order to characterize the limit distribution of CB_T it is useful to first state some preliminary results about the limit distribution of some quantities. For $s \in \mathbb{R}$ and for $j = 1, \ldots, m$, let $\overline{T}_j(s) := \max\{T_j(s), T^0_j\}$ and $\underline{T}_j(s) := \min\{T_j(s), T^0_j\}$ where $T_j(s) := T^0_j + [s v_T^{-2}]$. For $s, r \in \mathbb{R}$, we define $B_{T,j}(s,r) := v_T^2 \sum_{t=\underline{T}^0_j(s)+1}^{\overline{T}_j(s)} X_{tT} (\Sigma^0_{j+\mathbb{1}_{\{T_j(r)<t\}}})^{-1} X_{tT}'$ and $W_{T,j}(s,r) := v_T \sum_{t=\underline{T}^0_j(s)+1}^{\overline{T}_j(s)} X_{tT} (\Sigma^0_{j+\mathbb{1}_{\{T_j(r)<t\}}})^{-1} u_t$ for $j \in \{1, \ldots, m\}$.

Lemma 1. *Suppose that Assumptions A1-A9 hold. Then,*

$$\left\{B_{T,j}(\cdot,\cdot), W_{T,j}(\cdot,\cdot)\right\}_{j=1}^m \Rightarrow \left\{\mathbb{B}_j(\cdot,\cdot), \mathbb{W}_j(\cdot,\cdot)\right\}_{j=1}^m,$$

where

$$\mathbb{B}_j(s,r) := |s| S' \mathbb{D}_j(s) \otimes (\Sigma^0_{j+\mathbb{1}_{\{r \leq s\}}})^{-1} S - \mathbb{1}_{\{|r| \leq |s|\}} |r| S' \mathbb{D}_j(s) \otimes \{(\Sigma^0_{j+1})^{-1} - (\Sigma^0_j)^{-1}\} S,$$

and

$$\mathbb{W}_j(s,r) := S'\big(I_q \otimes (\Sigma^0_{j+\mathbb{1}_{\{r\le s\}}})^{-1}\big)\mathbb{V}_j(s) - \mathrm{sgn}(r)\mathbb{1}_{\{|r|\le|s|\}}S'\big[I_q \otimes \{(\Sigma^0_{j+1})^{-1} - (\Sigma^0_j)^{-1}\}\big]\mathbb{V}_j(r),$$

with $\mathbb{V}_j(s) := \big(I_q \otimes (\Sigma^0_{j+\mathbb{1}_{\{0\le s\}}})^{1/2}\big)\big[\mathbb{V}_{z\eta,j}(s)', \varphi(\lambda^0_j)' \otimes \mathbb{V}_{\eta,j}(s)', \mathbb{V}_w(\lambda^0_j)' \otimes \mathbb{V}_{\eta,j}(s)'\big]'$ *and*

$$\mathbb{D}_j(s) := \begin{pmatrix} Q_{zz,j+\mathbb{1}_{\{0<s\}}} & \mu_{z,j+\mathbb{1}_{\{0<s\}}}\varphi(\lambda^0_j)' & \mu_{z,j+\mathbb{1}_{\{0<s\}}}\mathbb{V}_w(\lambda^0_j)' \\ \varphi(\lambda^0_j)\mu'_{z,j+\mathbb{1}_{\{0<s\}}} & \varphi(\lambda^0_j)\varphi(\lambda^0_j)' & \varphi(\lambda^0_j)\mathbb{V}_w(\lambda^0_j)' \\ \mathbb{V}_w(\lambda^0_j)\mu'_{z,j+\mathbb{1}_{\{0<s\}}} & \mathbb{V}_w(\lambda^0_j)\varphi(\lambda^0_j)' & \mathbb{V}_w(\lambda^0_j)\mathbb{V}_w(\lambda^0_j)' \end{pmatrix}.$$

The theorem below presents the main result of the paper concerning the limit distribution of the test statistic, which can be expressed as the difference of the maxima of a limit process with and without restrictions implied by the assumption of common breaks.

Theorem 3. *Let* $\boldsymbol{s}_j = (s_{1j}, \ldots, s_{Gj})'$ *for* $j = 1, \ldots, m$ *and let* $\mathbf{1}$ *be a* $G \times 1$ *vector having 1 at all entries. Suppose Assumptions A1-A9 hold. Then, under the null hypothesis (7),*

$$CB_T \Rightarrow CB_\infty := \sup_{\boldsymbol{s}_1,\ldots,\boldsymbol{s}_m} \sum_{j=1}^m CB^{(j)}_\infty(\boldsymbol{s}_j) - \sup_{s_1,\ldots,s_m} \sum_{j=1}^m CB^{(j)}_\infty(s_j \cdot \mathbf{1}),$$

where

$$\begin{aligned} CB^{(j)}_\infty(\boldsymbol{s}_j) &:= \mathrm{tr}\Big(\Pi_j(s_{Gj})\mathbb{V}_{\eta\eta,j}(s_G)\Big) + \frac{|s_{Gj}|}{2}\mathrm{tr}\big(\{\Pi_j(s_{Gj})\}^2\big) - 2\sum_{g=1}^G \mathrm{sgn}(s_{gj})\Delta'_{gj}\mathbb{W}_j(s_{gj}, s_{Gj}) \\ &\quad - \sum_{g=1}^G\sum_{h=1}^G \Delta'_{gj}\Big\{\mathbb{1}_{\{s_{gj}\vee s_{hj}\le 0\}}\mathbb{B}_j\big(s_{gj}\vee s_{hj}, s_{Gj}\big) + \mathbb{1}_{\{0<s_{gj}\wedge s_{hg}\}}\mathbb{B}_j\big(s_{gj}\wedge s_{hj}, s_{Gj}\big)\Big\}\Delta_{hj}, \\ \Pi_j(s_{Gj}) &:= \begin{cases} \big(\Sigma^0_j\big)^{-1/2}\Upsilon_j\big(\Sigma^0_{j+1}\big)^{-1}\big(\Sigma^0_j\big)^{1/2}, & \text{if } s_{Gj} \le 0 \\ -\big(\Sigma^0_{j+1}\big)^{-1/2}\Upsilon_j\big(\Sigma^0_j\big)^{-1}\big(\Sigma^0_{j+1}\big)^{1/2}, & \text{if } s_{Gj} > 0 \end{cases}, \end{aligned} \tag{11}$$

with $\Delta_{gj} := \big(\|\delta_j\|^2 + \mathrm{tr}(\Phi_j^2)\big)^{-1/2}\delta_{gj}$ *and* $\Upsilon_j := \big(\|\delta_j\|^2 + \mathrm{tr}(\Phi_j^2)\big)^{-1/2}\Phi_j$.

The limit distribution in Theorem 3 is quite complex and depends on nuisance parameters. However, they can be consistently estimated and it is easy to show that the coverage rates will be asymptotically valid provided $\sqrt{T}$-consistent estimates are used instead of the true values. The various quantities can be estimated as follows: for $\Delta\widetilde{k}_j := \widetilde{k}_j - \widetilde{k}_{j-1}$, we can use $\widetilde{Q}_{zz,j} = (\Delta\widetilde{k}_j)^{-1}\sum_{t=\widetilde{k}_{j-1}+1}^{\widetilde{k}_j} z_t z_t'$, $\widetilde{\mu}_{z,j} = (\Delta\widetilde{k}_j)^{-1}\sum_{t=\widetilde{k}_{j-1}+1}^{\widetilde{k}_j} z_t$, $\Delta\widetilde{\beta}_j := \widetilde{\beta}_j - \widetilde{\beta}_{j-1}$ and $\widetilde{\Sigma}_j = (\Delta\widetilde{k}_j)^{-1}\sum_{t=\widetilde{k}_{j-1}+1}^{\widetilde{k}_j} \widetilde{u}_t\widetilde{u}_t'$, $\widetilde{\Delta}_{gj} := \big\{\|\Delta\widetilde{\beta}_j\|^2 + \mathrm{tr}\big((\Delta\widetilde{\Sigma}_j)^2\big)\big\}^{-1/2}\sum_{l\in\mathcal{G}_g} e_l \circ \Delta\widetilde{\beta}_{j+1}$ and $\widetilde{\Upsilon}_j := \big\{\|\Delta\widetilde{\beta}_j\|^2 + \mathrm{tr}\big((\Delta\widetilde{\Sigma}_j)^2\big)\big\}^{-1/2}\Delta\widetilde{\Sigma}_j$, where $\Delta\widetilde{\beta}_j := \widetilde{\beta}_j - \widetilde{\beta}_{j-1}$ and $\Delta\widetilde{\Sigma}_j := \widetilde{\Sigma}_j - \widetilde{\Sigma}_{j-1}$. Also, the estimates of the long run variances of $\{z_t \otimes \eta_t\}$ and $\{\eta_t\eta_t - I_n\}$ can be constructed using a method based on a weighted sum of sample autocovariances of the relevant quantities, as discussed in Andrews (1991), for instance. Though only $\sqrt{T}$-consistent estimates of (β, Σ) are needed, it is likely that more precise estimates of these parameters will lead to better finite sample coverage rates. Hence, it is recommended to use the estimates obtained imposing the restrictions in (3) even though imposing restrictions does not have a first-order effect on the limiting distribution of the estimates of the break dates.

In some cases, the limit distribution of the common breaks test can be derived and expressed in a simpler manner. For illustration purpose, our supplemental material states the limit distribution of the test under the setup of Examples 1 and 2. When the covariance matrix is constant over time (i.e., $\Sigma_j^0 = \Sigma^0$ for $j = 1, \ldots, m+1$), the limit distribution above can be further simplified as stated in the following corollary.

Corollary 1. *Let $\boldsymbol{s}_j = (s_{1j}, \ldots, s_{Gj})'$ for $j = 1, \ldots, m$ and let $\mathbf{1}$ be a $G \times 1$ vector having 1 at all entries. Suppose that Assumptions A1-A9 hold and also that the covariance matrix Σ_j^0 is constant over time. Then, under the null hypothesis (7),*

$$CB_T \Rightarrow \widetilde{CB}_\infty := \sup_{\boldsymbol{s}_1,\ldots,\boldsymbol{s}_m} \sum_{j=1}^{m} \widetilde{CB}_\infty^{(j)}(\boldsymbol{s}_j) - \sup_{s_1,\ldots,s_m} \sum_{j=1}^{m} \widetilde{CB}_\infty^{(j)}(s_j \cdot \mathbf{1}),$$

where

$$\begin{aligned}\widetilde{CB}_\infty^{(j)}(\boldsymbol{s}_j) \;\; := \;\; & -2\sum_{g=1}^{G} \operatorname{sgn}(s_{gj})\Delta_{gj}' \widetilde{\mathbb{W}}_j(s_{gj}) \\ & - \sum_{g=1}^{G}\sum_{h=1}^{G} \Delta_{gj}' \Big\{ \mathbb{1}_{\{s_{gj}\vee s_{hj}\le 0\}} \widetilde{\mathbb{B}}_j\big(s_{gj}\vee s_{hj}\big) + \mathbb{1}_{\{0<s_{gj}\wedge s_{hg}\}} \widetilde{\mathbb{B}}_j\big(s_{gj}\wedge s_{hj}\big) \Big\} \Delta_{hj},\end{aligned}$$

with $\widetilde{\mathbb{W}}_j(s) := S'\big(I_q \otimes (\Sigma^0)^{-1/2}\big)\big[\mathbb{V}_{z\eta,j}(s)', \varphi(\lambda_j^0)' \otimes \mathbb{V}_{\eta,j}(s)', \mathbb{V}_w(\lambda_j^0)' \otimes \mathbb{V}_{\eta,j}(s)'\big]'$ and $\widetilde{\mathbb{B}}_j(s) := |s| S'\mathbb{D}_j(s) \otimes (\Sigma^0)^{-1} S$ for $s \in \mathbb{R}$.

As another immediate corollary to Theorem 3, when no integrated variables are present, the limit distribution of the test for a common break date only involves the pre and post break date regimes, as is the case for the limit distribution of the estimates when multiple breaks are present (e.g. Bai and Perron, 1998). Also, the above result can be easily extended to test the hypothesis of common break dates for a part of the parameter groups, while the break dates of the other groups are not necessarily common. We illustrate the application of the test for common breaks in (7) and its variant through an application in Section 5.

As discussed in Section 1, there is one additional layer of difficulty compared to Bai and Perron (1998) or Qu and Perron (2007). In their analysis, the limit distribution can be evaluated using a closed form solution after some transformation, while no such solution is available here and thus we need to resort simulations to obtain the critical values. This involves first simulating the Wiener processes appearing in the various Brownian motion processes by partial sums of *i.i.d.* normal random vectors (independent of each others given Assumption A9). One can then evaluate one realization of the limit distribution by replacing unknown values by their estimates as stated above. The procedure is then repeated many times to obtain the relevant quantiles. While conceptually straightforward, this procedure is nevertheless computationally intensive. The reason is that for each replication we need to search over many possible combinations of all the permutations of the locations of the break dates. The procedure suggested is nevertheless quick enough to be feasible for common applications involving testing for few common break dates but the computational burden increases exponentially with the number of common breaks being tested. In Section 4, we propose an alternative approach to alleviate this issue and examine its performance.

3.3 Asymptotic power analysis

In this subsection, we provide an asymptotic power analysis of the test statistic CB_T when using a critical value c_α^* at the significance level α from the asymptotic null distribution CB_∞. As a fixed alternative hypothesis, we consider, for some $\delta > 0$

$$H_1 : \max_{1 \le g_1, g_2 \le G} |k^0_{g_1,j} - k^0_{g_2,j}| \ge \delta T \quad \text{for some } j = 1, \ldots, m. \tag{12}$$

Given that $k^0_{gj} = [T\lambda^0_{gj}]$ for $(g, j) \in \{1, \ldots, G\} \times \{1, ..., m\}$ under Assumption A3, the above condition is asymptotically equivalent to $\max_{1 \le g_1, g_2 \le G} |\lambda^0_{g_1,j} - \lambda^0_{g_2,j}| \ge \delta$ for some $j = 1, \ldots, m$, and thus can be considered as a fixed alternative hypothesis in term of break fractions. As a local alternative hypothesis, we consider

$$H_{1T} : \max_{1 \le g_1, g_2 \le G} |k^0_{g_1,j} - k^0_{g_2,j}| \ge M v_T^{-2} \quad \text{for some } j = 1, \ldots, m, \tag{13}$$

for some constant $M > 0$, where v_T satisfies the condition in Assumption A4. We can also express (13) as $\max_{1 \le g_1, g_2 \le G} |\lambda^0_{g_1,j} - \lambda^0_{g_2,j}| \ge M(\sqrt{T} v_T)^{-2}$ for some $j = 1, \ldots, m$. The following theorem shows that the proposed test statistic is consistent against fixed alternatives and also has non-trivial local power against local alternatives.

Theorem 4. *Let* $c_\alpha^* := \inf\left\{c \in \mathbb{R} : \Pr\{CB_\infty \le c\} \ge 1 - \alpha\right\}$. *Suppose that Assumptions A1-A9 hold. Then, (a) under the fixed alternative (12) with any* $\delta \in (0, 1]$,

$$\lim_{T \to \infty} \Pr\left\{CB_T > c_\alpha^*\right\} = 1,$$

(b) under the local alternative (13), for any $\epsilon > 0$, *there exits an* M *defined in (13) such that*

$$\lim_{T \to \infty} \Pr\left\{CB_T > c_\alpha^*\right\} > 1 - \epsilon.$$

4 Monte Carlo simulations

This section provides simulation results about the finite sample performance of the test in terms of size and power. We first consider a direct simulation-based approach to obtain the critical values and then a more computationally efficient algorithm. As a data generating process (DGP), we adopt a similar setup to the one used in Bai et al. (1998), namely a bivariate autoregressive system with a single break in intercepts as in Example 1. Hence, only the intercepts are allowed to change at some dates k_{i1} for equation $i \in \{1, 2\}$. We test the null hypothesis $H_0 : k_{11} = k_{21}$ against the alternative hypothesis $H_1 : k_{11} \neq k_{21}$. The number of observations is set to $T = 100$, and we use 500 replications. Results are reported for autoregressive parameters $\alpha \in \{0.0, 0.4, 0.8\}$. We set $\mu_{i1} = 1$ and let $\delta_i := \mu_{i2} - \mu_{i1}$, the magnitude of the mean shift, take values $\{0.50, 0.75, 1.00, 1.25, 1.50\}$.

A direct simulation-based approach: We first present results when we resort direct simulations to obtain the critical values, which involves simulating the Wiener processes by partial sums of i.i.d. normal random vectors and searching over all possible combinations of the break dates. Given the computational cost, we choose a simple setup and focus on limited cases. To examine the empirical sizes and power, we here consider the errors $(u_{1t}, u_{2t})'$ following *i.i.d.* $N(0, I_2)$ and we use 3,000 repetitions to generate the critical values.

We first examine the empirical rejection frequencies under the null hypothesis that $k_{11} = k_{21} = 50$ with a trimming parameter $\nu = 0.15$. The results are reported in Table 1 for nominal sizes of 10%, 5% and 1%. First, when the autoregressive process has no or moderate dependency ($\alpha = 0.0$ or $\alpha = 0.4$), the empirical size of the test is either slightly conservative or close to the nominal size. Given the small sample size, this size property is satisfactory. When the autoregressive parameter is close to the boundary of the non-stationary region, e.g. $\alpha = 0.8$, as expected there are some liberal size distortions. When the magnitudes of the breaks are small, the test tends to over-reject the null hypothesis. This is due to the fact that for very small breaks the break date estimates are quite imprecise and are more likely to be affected by the highly dependent series than the break sizes themselves, so that the test depends on the log likelihoods evaluated outside neighborhoods of the true break dates. When the magnitude of the break sizes increases, the size of the test quickly approaches the nominal level. These results are encouraging given the small sample size.

To analyze power, we also set $\mu_{i1} = 1$, while we consider values $\{0.50, 1.00, 1.50\}$ for the magnitude of the mean shift. The break date in the first equation is kept fixed at $k_1 = 35$, while the break date in the second equation takes values $k_2 = 35, 40, 45, 50, 55$. The power is a function of the difference between the break dates, $k_2 - k_1$. The results are presented in Figure 1, where the horizontal axis in each box represents the difference $k_2 - k_1$ and the vertical axis shows the empirical rejection frequency. As before, when the magnitudes of the breaks are small, the data are not informative enough to reject the common breaks null hypothesis and the test has little power. However, when the magnitudes of the changes reach 1, the power increases rapidly as the distance between the break dates increases. The results are qualitatively similar for all values of α considered.

An alternative approach: The direct simulation-based procedure involves a combinatorial optimization problem and the computational burden increases exponentially with the number of common breaks being tested. Such a procedure may be feasible for a small number of breaks in a parsimonious system. However, in more general cases, it may be prohibitive. Hence, we also propose an alternative approach that solves this problem, using heuristic algorithms that find approximate, if not optimal, solutions. Because heuristic algorithms have mainly been developed to optimize functions having explicit forms, we use the Karhunen-Loève (KL) representation of stochastic processes, which expresses a Brownian motion as an infinite sum of sine functions with independent Gaussian random multipliers (see Bosq, 2012, p. 26, for instance). A truncated series of the KL representation was used to obtain critical values by Durbin (1970) and Krivyakov et al. (1978), among others. Similarly, we use a truncated series with 500 terms and apply a change of variables to approximately obtain an explicit form of the objects being maximized in the limit distribution of the common breaks test. Also, we use the particle swarm optimization method, which is an evolutionary computation algorithm developed by Eberhart and Kennedy (1995).[11]

We examine the performance of the common breaks test using the alternative algorithm under various setups in order to show that similar good finite sample properties are obtained compared to the direct optimization method. In addition to the setup used above, we consider a trimming value $\nu = 0.10$, a pair of break dates (35, 35) and normal errors with correlation coefficient being 0.5 across equations. Columns (1)-(4) of Table 2 present empirical rejection frequencies under the null hypothesis for a nominal size of 5%. Whether the errors are correlated or not, the empirical size of the test is either conservative or close to the nominal size in cases of moderate dependency ($\alpha = 0.0$ or $\alpha = 0.4$). Also the trimming parameter has little impact. With uncorrelated errors, there are size distortions in cases of high dependency ($\alpha = 0.8$) and small break sizes. When the errors are correlated, however, the empirical sizes get closer to the nominal level in all cases. This is likely due to efficiency gains from using a SUR estimation method. Columns (5)-(6) of Table 2 report the empirical power for the case $(k_1, k_2) = (35, 50)$ and the results show satisfactory power, comparable to the direct method.

[11]For our simulations, we use the particle swarm algorithm "*particleswarm*" of the Matlab Global Optimization Toolbox. We also tried the genetic algorithm "*ga*" from Matlab and found that the two algorithms yield very similar, frequently the same, critical values, while the particle swarm algorithm is faster.

5 Application

In this section, we apply the common breaks test to inflation series, following Clark (2006). He analyzes the persistence of a number of disaggregated inflation series based on the sum of the autoregressive (AR) coefficients in an AR model, and documents that the persistence is very high and close to one without allowing for a mean shift, whereas the persistence declines substantially when allowing for one. Although such features have been documented theoretically in the literature (e.g. Perron, 1990), he finds that the decline in persistence is more pronounced amongst disaggregated measures compared to various aggregate measures. The issue of importance is that Clark (2006) assumes a common mean shift for all series, following Bai et al. (1998), but the validity of this assumption is not established.

We consider a subset of the series analyzed in Clark (2006), namely the inflation measures for durables, nondurables and services. These are taken from the NIPA accounts and cover the period 1984-2002 at the quarterly frequency; see Clark (2006) for more details. Let $\{(y_{1t}, y_{2t}, y_{3t})\}_{t=1}^{T}$ denote the inflation series of durables, nondurables and services and consider an AR model allowing for a mean shift for each series $i = 1, 2, 3$:

$$y_{it} = \mu_i + \delta_i \mathbb{1}_{\{k_i+1 \leq t\}} + \alpha_i^{(1)} y_{i,t-1} + \cdots + \alpha_i^{(p_i)} y_{i,t-p_i} + u_{it}, \qquad t = 1, \ldots, T,$$

where μ_i is an intercept parameter, δ_i is the magnitude of the mean shift with k_i being a break date. The parameters, $\alpha_i^{(1)}, \ldots, \alpha_i^{(p_i)}$, are AR coefficients with p_i denoting the lag length and u_{it} is an error term. The persistence of each series is measured by the sum $\alpha_i^{(1)} + \cdots + \alpha_i^{(p_i)}$ for $i = 1, 2, 3$. Clark (2006) uses the Akaike information criterion (AIC) to select the AR lag length such that $(p_1, p_2, p_3) = (4, 5, 3)$ and also presents some evidence to support a mean shift in the AR models by applying break tests for each series and for groups.

We present our empirical results in Table 3. We first replicate a part of the results in Clark (2006). We find that when not allowing for a mean shift, the persistence measure is indeed quite high ranging from 0.855 to 0.921. Also, the persistence measure decreases to a large extent for non-durables and services but not so much for durables when a common break is imposed for the intercept at the break date 1993:Q1, which is not estimated but treated as known in Clark (2006). When we use the Seemingly Unrelated Regressions (SUR) method with an unknown common break date, following Bai et al. (1998), the point estimates are similar expect that the break date is estimated at 1992:Q1.

We now use our test to assess the validity of the common breaks specification. In Table 3, we report values of the test statistic for several null hypotheses as well as critical values corresponding to a 5% significance level, obtained through the computationally efficient algorithm described in Section 4 with 3,000 repetitions. First, we consider the null hypothesis of common breaks in the three inflation series, i.e., $H_0 : k_1 = k_2 = k_3$. The value of the test statistic is 9.015 and the critical value is 5.242, so that the test rejects the null hypothesis of common breaks at the 5% significance level. Next, we test for common breaks in two inflation series within the full system of the three inflation series, separately. That is, we separately calculate the test statistic for $H_0 : k_1 = k_2$, $H_0 : k_1 = k_3$, and $H_0 : k_2 = k_3$. The values of the test statistic are 9.735 and 7.684 with corresponding critical values 3.473 and 3.259 for $H_0 : k_1 = k_2$ and $H_0 : k_1 = k_3$, respectively, and thus both hypotheses are rejected at the 5% significance level. On the other hand, the value of the statistic for $H_0 : k_2 = k_3$ is 0.749 with a critical value of 2.501. Thus, we cannot reject the null hypothesis of common breaks in the nondurables and service series.

We then estimate a system with the three inflation series imposing a common break only in the nondurables and service series (i.e., $k_2 = k_3$), estimated at 1992:Q1, which is the same as when allowing for an unknown common break date in all series (the parameter estimates are also broadly similar). Things are quite different for the durables series. In this case, the estimate of the break date is 1995:Q1. What is interesting is that with this break date the decrease in persistence is very important with an estimate of 0.324 compared to 0.805 obtained assuming a common break date across the three series. Hence, allowing for different break dates for durables and the other series, we document a substantial decline in the persistence measure across all three series. Moreover, we report the 95% confidence intervals for the estimated break dates: [1994:Q2, 1995:Q4] for durables and [1991:Q3, 1992:Q3] for the others. These non-overlapping intervals are consistent with our results.

6 Conclusion

This paper provides a procedure to test for common breaks across or within equations. Our framework is very general and allows integrated regressors and trends as well as stationary regressors. The test considered is the quasi-likelihood ratio test assuming normal errors, though as usual the limit distribution of the test remains valid with non-normal errors. Of independent interest, we provide results about the rate of convergence when searching over all possible partitions subject only to the requirement that each regime contains at least as many observations as some positive fraction of the sample size, allowing break dates not separated by a positive fraction of the sample size across equations. We propose two approaches to obtain critical values. Simulations show that the test has good finite sample properties. We also provide an application to issues related to level shifts and persistence for various measures of inflation to illustrate its usefulness.

References

Andrews, D. W. K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59 (3), 817–58.

Andrews, D. W. K., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61 (4), 821–56.

Aue, A., Lee, T. C. M., 2011. On image segmentation using information theoretic criteria. Ann. Statist. 39 (6), 2912–2935.

Bai, J., 1994. Least squares estimation of a shift in linear processes. Journal of Time Series Analysis 15 (5), 453–472.

Bai, J., 1995. Least absolute deviation estimation of a shift. Econometric Theory 11 (03), 403–436.

Bai, J., 1997. Estimation of a change point in multiple regression models. The Review of Economics and Statistics 79 (4), 551–563.

Bai, J., 1998. Estimation of multiple-regime regressions with least absolutes deviation. Journal of Statistical Planning and Inference 74 (1), 103–134.

Bai, J., 2000. Vector autoregressive models with structural changes in regression coefficients and in variance-covariance matrices. Annals of Economics and Finance 1 (2), 303–339.

Bai, J., 2010. Common breaks in means and variances for panel data. Journal of Econometrics 157 (1), 78–92.

Bai, J., Lumsdaine, R. L., Stock, J. H., 1998. Testing for and dating common breaks in multivariate time series. Review of Economic Studies 65 (3), 395–432.

Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66 (1), 47–78.

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

Baltagi, B. H., Feng, Q., Kao, C., 2016. Estimation of heterogeneous panels with structural breaks. Journal of Econometrics 191 (1), 176–195.

Bosq, D., 2012. Linear Processes in Function Spaces: Theory and Applications. Vol. 149. Springer Science & Business Media.

Breitung, J., Eickmeier, S., 2011. Testing for structural breaks in dynamic factor models. Journal of Econometrics 163 (1), 71–84.

Cavaliere, G., Taylor, A. R., 2007. Testing for unit roots in time series models with non-stationary volatility. Journal of Econometrics 140 (2), 919–947.

Cheng, X., Liao, Z., Schorfheide, F., 2016. Shrinkage estimation of high-dimensional factor models with structural instabilities. The Review of Economic Studies 83 (4), 1511–1543.

Clark, T. E., 2006. Disaggregate evidence on the persistence of consumer price inflation. Journal of Applied Econometrics 21 (5), 563–587.

Corradi, V., 1999. Deciding between I(0) and I(1) via flil-based bounds. Econometric Theory 15 (05), 643–663.

Corradi, V., Swanson, N. R., 2014. Testing for structural stability of factor augmented forecasting models. Journal of Econometrics 182 (1), 100–118.

Csörgö, M., Horváth, L., 1997. Limit Theorems in Change-Point Analysis. John Wiley & Sons Inc.

Davidson, J., 1994. Stochastic Limit Theory: An Introduction for Econometricians. Oxford University Press, UK.

Davis, L. (Ed.), 1991. Handbook of Genetic Algorithms. Van Nostrand Reinhold.

Deng, A., Perron, P., 2008. A non-local perspective on the power properties of the cusum and cusum of squares tests for structural change. Journal of Econometrics 142 (1), 212–240.

Dragomir, S. S., 2016. Logarithmic inequalities for two positive numbers via Taylor's expansion with integral remainder. RGMIA Research Report Collection 19, 139.

Durbin, J., 1970. Asymptotic distributions of some statistics based on the bivariate sample distribution functions. In: Puri, M. L. (Ed.), Nonparametric Techniques in Statistical Inference. London: Cambridge Univ. Press, pp. 435–449.

Eberhart, R., Kennedy, J., 1995. A new optimizer using particle swarm theory. In: Micro Machine and Human Science, 1995. MHS'95., Proceedings of the Sixth International Symposium on. IEEE, pp. 39–43.

Eberlein, E., 1986. On strong invariance principles under dependence assumptions. The Annals of Probability 14 (1), 260–270.

Eo, Y., Morley, J., 2015. Likelihood-ratio-based confidence sets for the timing of structural breaks. Quantitative Economics 6 (2), 463–497.

Hájek, J., Rényi, A., 1955. Generalization of an inequality of Kolmogorov. Acta Mathematica Hungarica 6 (3-4), 281–283.

Hall, A. R., Han, S., Boldea, O., 2012. Inference regarding multiple structural changes in linear models with endogenous regressors. Journal of Econometrics 170 (2), 281–302.

Hall, A. R., Sen, A., 1999. Structural stability testing in models estimated by generalized method of moments. Journal of Business & Economic Statistics 17 (3), 335–48.

Han, X., Inoue, A., 2015. Tests for parameter instability in dynamic factor models. Econometric Theory 31 (05), 1117–1152.

Hansen, B. E., 1992. Tests for parameter instability in regressions with I(1) processes. Journal of Business & Economic Statistics 10 (3), 321–35.

Hatanaka, M., Yamada, H., 2003. Co-trending. Springer.

Hendry, D. F., Mizon, G. E., 1998. Exogeneity, causality, and co-breaking in economic policy analysis of a small econometric model of money in the UK. Empirical Economics 23 (3), 267–294.

Ibragimov, I. A., 1962. Some limit theorems for stationary processes. Theory of Probability & Its Applications 7 (4), 349–382.

Kejriwal, M., Perron, P., 2008. The limit distribution of the estimates in cointegrated regression models with multiple structural changes. Journal of Econometrics 146 (1), 59–73.

Kim, D., 2011. Estimating a common deterministic time trend break in large panels with cross sectional dependence. Journal of Econometrics 164 (2), 310–330.

Kim, D., Oka, T., Estrada, F., Perron, P., 2017. Inference related to common breaks in a multivariate system with joined segmented trends with applications to global and hemispheric temperatures, manuscript, Department of Economics, Boston University.

Kim, D., Perron, P., 2009. Assessing the relative power of structural break tests using a framework based on the approximate Bahadur slope. Journal of Econometrics 149 (1), 26–51.

Krivyakov, E., Martynov, G., Tyurin, Y. N., 1978. On the distribution of the ω^2 statistics in the multi-dimensional case. Theory of Probability & Its Applications 22 (2), 406–410.

Kurozumi, E., Tuvaandorj, P., 2011. Model selection criteria in multivariate models with multiple structural changes. Journal of Econometrics 164 (2), 218–238.

Lee, T. C. M., 2000. A minimum description length-based image segmentation procedure, and its comparison with a cross-validation-based segmentation procedure. Journal of the American Statistical Association 95 (449), 259–270.

Oka, T., Qu, Z., 2011. Estimating structural changes in regression quantiles. Journal of Econometrics 162 (2), 248–267.

Perron, P., 1990. Testing for a unit root in a time series with a changing mean. Journal of Business & Economic Statistics 8 (2), 153–62.

Perron, P., 2006. Dealing with structural breaks. In: Patterson, K., Mills, T. (Eds.), Palgrave Handbook of Econometrics, Vol. 1: Econometric Theory. Palgrave Macmillan, pp. 278–352.

Perron, P., Yamamoto, Y., 2014. A note on estimating and testing for multiple structural changes in models with endogenous regressors via 2SLS. Econometric Theory 30 (02), 491–507.

Perron, P., Yamamoto, Y., 2015. Using OLS to estimate and test for structural changes in models with endogenous regressors. Journal of Applied Econometrics 30 (1), 119–144.

Perron, P., Yamamoto, Y., 2016. On the usefulness or lack thereof of optimality criteria for structural change tests. Econometric Reviews 35 (5), 782–844.

Qian, J., Su, L., 2016. Shrinkage estimation of common breaks in panel data models via adaptive group fused lasso. Journal of Econometrics 191 (1), 86–109.

Qu, Z., 2008. Testing for structural change in regression quantiles. Journal of Econometrics 146 (1), 170–184.

Qu, Z., Perron, P., 2007. Estimating and testing structural changes in multivariate regressions. Econometrica 75 (2), 459–502.

Saikkonen, P., 1991. Asymptotically efficient estimation of cointegration regressions. Econometric Theory 7, 1–21.

Stock, J. H., Watson, M. W., 1993. A simple estimator of cointegrating vectors in higher order integrated systems. Econometrica 61, 783–820.

Su, L., Xiao, Z., 2008. Testing for parameter stability in quantile regression models. Statistics & Probability Letters 78 (16), 2768–2775.

Yamamoto, Y., Tanaka, S., 2015. Testing for factor loading structural change under common breaks. Journal of Econometrics 189 (1), 187–206.

Appendix

Throughout the appendix, we use C, C_1,C_2, ... to denote generic positive constants without further clarification. Also, we use diag$(\cdot)$ to denote the operator that generates a square diagonal matrix with its diagonal entries being equal to its inputs. The key ingredients in the proofs are a Strong Approximation Theorem (SAT), a Functional Central Limit Theorem (FCLT) and a generalized Hajek-Renyi inequality. We first state two technical lemmas.

Lemma A.1. *Let $\{\varsigma_t\}_{t\in\mathbb{Z}}$ be a sequence of mean-zero, $\mathbb{R}^d$-valued random vectors satisfying Assumptions A2 and A7. Define $S_k(\ell)=\sum_{t=\ell+1}^{\ell+k}\varsigma_t$, then, (a) (SAT) the covariance matrix of $k^{-1/2}S_k(\ell)$, Ω_k, converge, with the limit denoted by Ω, and there exists a Brownian Motion $(W(t))_{t\geq 0}$ with covariance matrix Ω such that $\sum_{i=1}^{t}\varsigma_i - W(t) = O_{a.s}(t^{1/2-\kappa})$ for some $\kappa>0$; (b) (FCLT) $T^{-1/2}\sum_{t=1}^{[Tr]}\varsigma_t \Rightarrow \Omega^{1/2}W^*(r)$, where $W^*(r)$ is a $\mathbb{R}^d$-valued vector of independent Wiener processes and "$\Rightarrow$" denotes weak convergence under the Skorohod topology.*

The above lemma is proved in Lemma A.1 of Qu and Perron (2007), who use Theorem 2 in Eberlein (1986) together with the arguments of Corradi (1999). The following lemma is an extension of the Hajek-Renyi inequality.

Lemma A.2. *Suppose that Assumptions A1, A2 and A5 hold. Let $\{b_k\}_{k\in\mathbb{N}}$ be a sequence of positive, non-increasing constants and let $\{\xi_{tT}\}$ denote either $\{X_{tT}\Sigma_{t,\mathcal{K}}^{-1}u_t\}$ or $\{\eta_t\eta_t'-I_n\}$. Then, for any $B>0$ and for any $k_1,k_2\in\mathbb{N}$ with $k_1<k_2$,*

$$\Pr\left\{\sup_{k_1\leq k\leq k_2}\frac{1}{kb_k}\left\|\sum_{t=1}^{k}\xi_{tT}\right\|>B\right\}\leq\frac{C}{B^2}\left(\frac{1}{k_1b_{k_1}^2}+\sum_{k=k_1+1}^{k_2}\frac{1}{(kb_k)^2}\right).$$

***Proof*.** The assertion is proved if we show that $\{X_{tT}\Sigma_{t,\mathcal{K}}^{-1}u_t\}$ and $\{\eta_t\eta_t'-I_n\}$ satisfy the L^2-mixingale condition in Lemma A6 of Bai and Perron (1998), which shows the HajeK-Renyi inequality for a L^2-mixingale sequence.[12] We consider only $\{X_{tT}\Sigma_{t,\mathcal{K}}^{-1}u_t\}$ because the proof for $\{\eta_t\eta_t'-I_n\}$ is similar and actually simpler. We use the notation $E_t(\cdot):=E(\cdot|\mathcal{F}_t)$ for $t\in\mathbb{Z}$.

We can write $X_{tT}\Sigma_{t,\mathcal{K}}^{-1}u_t = S'(I_q\otimes\Sigma_{t,\mathcal{K}}^{-1}(\Sigma_{t,\mathcal{K}}^0)^{1/2})(x_{tT}\otimes\eta_t)$, where $\|S'(I_q\otimes\Sigma_{t,\mathcal{K}}^{-1}(\Sigma_{t,\mathcal{K}}^0)^{1/2})\|\leq C_1$ from Assumption A5 and the term $(x_{tT}\otimes\eta_t)$ is $\mathcal{F}_t$-measurable. Thus, it suffices to show that there exist non-negative constants $\{\psi_j\}_{j\geq 0}$ such that, for all $t\geq 1$ and $j\geq 0$,

$$\left\|E_{t-j}\big(x_{tT}\otimes\eta_t\big)-E\big(x_{tT}\otimes\eta_t\big)\right\|_2\leq C_2\psi_j, \tag{A.1}$$

as well as $\psi_j\to 0$ as $j\to\infty$ and $\sum_{j=1}^{\infty}j^{1+\vartheta}\psi_j<\infty$ for some $\vartheta>0$.

In order to show (A.1), we write $x_{tT}\otimes\eta_t=\left[z_t'\otimes\eta_t',\varphi(t/T)'\otimes\eta_t',T^{-1/2}w_t'\otimes\eta_t'\right]'$ and observe that $E[z_t\otimes\eta_t]=0$ and $E[\eta_t]=0$. It follows from Minkowski's inequality that

$$\begin{aligned}\left\|E_{t-j}(x_{tT}\otimes\eta_t)-E(x_{tT}\otimes\eta_t)\right\|_2 &\leq \left\|E_{t-j}(z_t\otimes\eta_t)\right\|_2+\left\|\varphi(t/T)\otimes E_{t-j}(\eta_t)\right\|_2\\ &\quad +T^{-1/2}\left\|E_{t-j}(w_t\otimes\eta_t)-E(w_t\otimes\eta_t)\right\|_2\\ &=: A_1+A_2+A_3.\end{aligned}$$

[12] Lemma A6 of Bai and Perron (1998) obtains a Hajek-Renyi inequality with the the supremum taken over $[k_1,\infty]$ rather than the original one with the the supremum taken over a finite range $[k_1,k_2]$ as in the assertion of this lemma. Their argument, however, can easily be extended to cover the case considered here.

For A_1 and A_2, an application of the mixing inequality of Ibragimov (1962) yields that[13]

$$A_1 \leq 2(\sqrt{2}+1)\alpha_j^{1/2-1/\phi}\|z_t \otimes \eta_t\|_\phi \quad \text{and} \quad A_2 \leq 2(\sqrt{2}+1)\alpha_j^{1/2-1/\phi}\|\eta_t\|_\phi, \tag{A.2}$$

where $\phi := 4+\delta$ with δ defined in Assumption A2. For the term A_3, we separately consider two cases: (i) $t < j$ and (ii) $t \geq j$, given $t \geq 1$. First, we consider case (i), i.e., $t-j < 0$. We have $w_t = w_0 + \sum_{l=0}^{t-1} u_{w,t-l}$, which with Minkowski's inequality implies that

$$\sqrt{T}A_3 \leq \big\|E_{t-j}(w_0 \otimes \eta_t) - E(w_0 \otimes \eta_t)\big\|_2 + \sum_{l=0}^{t-1} \big\|E_{t-j}\big(u_{w,t-l} \otimes \eta_t\big) - E\big(u_{w,t-l} \otimes \eta_t\big)\big\|_2.$$

Since $\|E_{t-j}(V) - E(V)\|_2 \leq \|E_{t-j}(V)\|_2$ for a random vector V, an application of Jensen's inequality and Corollary 14.3 of Davidson (1994) (a covariance inequality for a α-mixing sequence) yields that

$$\big\|E_{t-j}(w_0 \otimes \eta_t) - E(w_0 \otimes \eta_t)\big\|_2 \leq \big\|w_0 \otimes \eta_t\big\|_2 \leq C_3\alpha_t^{1/2-1/\phi}, \tag{A.3}$$

and that, for $0 \leq l \leq t-1$,

$$\big\|E_{t-j}\big(u_{w,t-l} \otimes \eta_t\big) - E\big(u_{w,t-l} \otimes \eta_t\big)\big\|_2 \leq \big\|u_{w,t-l} \otimes \eta_t\big\|_2 \leq C_4\alpha_l^{1/2-1/\phi}. \tag{A.4}$$

Also, using the mixing inequality of Ibragimov (1962), we can show that

$$\big\|E_{t-j}(w_0 \otimes \eta_t) - E(w_0 \otimes \eta_t)\big\|_2 \leq 2(\sqrt{2}+1)\alpha_{j-t}^{1/2-1/\phi}\big\|w_0 \otimes \eta_t\big\|_\phi, \tag{A.5}$$

and that, for $0 \leq l \leq t-1$,

$$\big\|E_{t-j}\big(u_{w,t-l} \otimes \eta_t\big) - E\big(u_{w,t-l} \otimes \eta_t\big)\big\|_2 \leq 2(\sqrt{2}+1)\alpha_{j-l}^{1/2-1/\phi}\big\|u_{w,t-l} \otimes \eta_t\big\|_\phi, \tag{A.6}$$

where both moments on the right-hand side of (A.5) and (A.6) are bounded from Assumption A2. It follows from (A.3)-(A.6) that, when $t < j$, we have

$$A_3 \leq C_5 T^{-1/2}\sum_{l=0}^{t}\min\{\alpha_l^{1/2-1/\phi}, \alpha_{j-l}^{1/2-1/\phi}\} \leq C_5 j^{1/2}\alpha_{[j/2]}^{1/2-1/\phi}, \tag{A.7}$$

where the last inequality is due to the fact that $\min\{\alpha_l^{1/2-1/\phi}, \alpha_{j-l}^{1/2-1/\phi}\} \leq \alpha_{[j/2]}^{1/2-1/\phi}$ for every $0 \leq l \leq t$ and that $T^{-1/2}t \leq t^{1/2} \leq j^{1/2}$ for $t < j$.

Next, we consider case (ii), i.e., $0 \leq t-j$. Since $w_t = w_{t-j} + \sum_{l=0}^{j-1} u_{w,t-l}$, Minkowski's inequality leads to

$$\sqrt{T}A_3 \leq \big\|w_{t-j} \otimes E_{t-j}(\eta_t)\big\|_2 + \sum_{l=0}^{j-1}\big\|E_{t-j}\big(u_{w,t-l} \otimes \eta_t\big) - E\big(u_{w,t-l} \otimes \eta_t\big)\big\|_2. \tag{A.8}$$

Using the Cauchy-Schwarz and Ibragimov's mixing inequalities, we can show that

$$\big\|w_{t-j} \otimes E_{t-j}(\eta_t)\big\|_2 \leq \big\|w_{t-j}\big\|_2\big\|E_{t-j}(\eta_t)\big\|_2 \leq \big\|w_{t-j}\big\|_2 C_6\alpha_j^{1/2-1/\phi}. \tag{A.9}$$

Furthermore, we can write $\|w_{t-j}\|_2^2 = \sum_{s=1}^{t-j} E[u_{ws}'u_{ws}] + 2\sum_{k=1}^{t-j-1}\sum_{s=1}^{t-j-k} E[u_{ws}'u_{w,s+k}]$, which with Corollary 14.3 of Davidson (1994) implies

$$T^{-1}\|w_{t-j}\|_2^2 \leq C_7\Big(\frac{t-j}{T} + \sum_{k=1}^{t-j-1}\frac{t-j-k}{T}\alpha_k^{1/2-1/\phi}\Big) \leq C_8.$$

[13]For A_2, we use the fact $\|\varphi(t/T) \otimes \eta_t\|_2^2 = E[(\varphi(t/T) \otimes \eta_t)'(\varphi(t/T) \otimes \eta_t)] = \varphi(t/T)'\varphi(t/T)E[\eta_t'\eta_t]$, which implies that $\|\varphi(t/T) \otimes \eta_t\|_2 \leq C\|\eta_t\|_2$.

Also, applying the same arguments used in case (i), we can show that

$$\sum_{l=0}^{j-1} \left\| E_{t-j}(u_{w,t-l} \otimes \eta_t) - E(u_{w,t-l} \otimes \eta_t) \right\|_2 \leq C_9 \sum_{l=0}^{j-1} \min\{\alpha_l^{1/2-1/\phi}, \alpha_{j-l}^{1/2-1/\phi}\}. \qquad (A.10)$$

Combining the results in (A.9)-(A.10), we obtain

$$A_3 \leq C_{10}\big(\alpha_j^{1/2-1/\phi} + T^{-1/2} j \alpha_{[j/2]}^{1/2-1/\phi}\big) \leq C_{11} j^{1/2} \alpha_{[j/2]}^{1/2-1/\phi}.$$

Thus, from the above equation and (A.7), we obtain that $A_3 \leq C_{12} j^{1/2} \alpha_{[j/2]}^{1/2-1/\phi}$ for every $t \geq 1$. This result together with (A.2) and (A.8) yields

$$\left\| E_{t-j}(x_{tT} \otimes \eta_t) - E(x_{tT} \otimes \eta_t) \right\|_2 \leq C_{13} j^{1/2} \alpha_{[j/2]}^{1/2-1/\phi}.$$

We set $\psi_j = j^{1/2} \alpha_{[j/2]}^{1/2-1/\phi}$ and it remains to show that $\sum_{j=1}^{\infty} j^{1+\vartheta} \psi_j < \infty$ for some $\vartheta > 0$. Observe that $\alpha_{[j/2]}^{1/2-1/\phi} = O(j^{\frac{5}{2} - \frac{1-2\delta}{\delta}})$ under Assumption A2. Thus, for $\vartheta < (1-2\delta)/\delta$, we can show that $\sum_{j=1}^{\infty} j^{1+\vartheta} \psi_j \leq C_{14} \sum_{j=1}^{\infty} j^{-1-\frac{1-2\delta}{\delta}+\vartheta} < \infty$. This completes the proof. ■

In what follows, we shall use a collection of sub-intervals $\{[\tau_{l-1}+1, \tau_l]\}_{l=1}^{N}$ with $\tau_0 = 0$ and $\tau_N = T$ as a partition of the interval $[1, T]$ according to sets of break dates $\mathcal{K}$ and $\mathcal{K}^0$, such that both the true basic parameters and their estimates are constant within each sub-interval and N is set to be the smallest number of such sub-intervals; that is, $(\beta_{t,\mathcal{K}}, \beta^0_{t,\mathcal{K}^0}, \Sigma_{t,\mathcal{K}}, \Sigma^0_{t,\mathcal{K}^0}) = (\beta_{\tau_l,\mathcal{K}}, \beta^0_{\tau_l,\mathcal{K}^0}, \Sigma_{\tau_l,\mathcal{K}}, \Sigma^0_{\tau_l,\mathcal{K}^0})$ for $\tau_{l-1}+1 \leq t \leq \tau_l$. For each parameter group $g \in \{1, \ldots, G\}$, we similarly consider a collection $\{[\tau_{g,l-1}+1, \tau_{gl}]\}_{l=1}^{N_g}$ with $\tau_0 = 0$ and $\tau_{N_g} = T$ as a partition of the interval $[1, T]$ given $\mathcal{K}_g$ and $\mathcal{K}_g^0$, where both the true basic parameters and their estimates for the g^{th} group are constant within each sub-interval and N_g is the smallest number of such intervals. Thus we have $(\beta_{g,t,\mathcal{K}}, \beta^0_{g,t,\mathcal{K}}) = (\beta_{g,\tau_{gl},\mathcal{K}}, \beta^0_{g,\tau_{gl},\mathcal{K}^0})$ for $\tau_{g,l-1}+1 \leq t \leq \tau_{gl}$ and $(\Sigma_{t,\mathcal{K}}, \Sigma^0_{t,\mathcal{K}^0}) = (\Sigma_{\tau_{G,l},\mathcal{K}}, \Sigma^0_{\tau_{G,l},\mathcal{K}^0})$ for $\tau_{G,l-1}+1 \leq t \leq \tau_{G,l}$, whereas the basic parameters of the other groups may change. For $\tau_{G,l-1}+1 \leq t \leq \tau_{G,l}$ with $l \in \{1, \ldots, N_g\}$, we define

$$\Psi_l := (\Sigma^0_{t,\mathcal{K}^0})^{-1/2} (\Sigma_{t,\mathcal{K}} - \Sigma^0_{t,\mathcal{K}^0}) (\Sigma^0_{t,\mathcal{K}^0})^{-1/2}, \qquad (A.11)$$

where we have $I_n + \Psi_l = (\Sigma^0_{\tau_{G,l},\mathcal{K}^0})^{-1/2} \Sigma_{\tau_{G,l},\mathcal{K}} (\Sigma^0_{\tau_{G,l},\mathcal{K}^0})^{-1/2}$. Since Ψ_l is an $n \times n$ symmetric matrix, there exits an orthogonal matrix U such that

$$U\Psi U' = \text{diag}\{\lambda^{\Psi}_{l1}, ..., \lambda^{\Psi}_{ln}\} \quad \text{and} \quad U(I_n + \Psi)U' = \text{diag}\{1 + \lambda^{\Psi}_{l1}, ..., 1 + \lambda^{\Psi}_{ln}\},$$

where $\lambda^{\Psi}_{l1}, \ldots, \lambda^{\Psi}_{ln}$ are the eigenvalues of Ψ_l.

In the lemma below, we shall obtain an upper bound for the normalized log likelihood based on sub-intervals. As a short-hand notation, we define, for $1 \leq t \leq T$ and $1 \leq g \leq G$,

$$\Delta\beta_{t,\mathcal{K}} := \beta_{t,\mathcal{K}} - \beta^0_{t,\mathcal{K}^0} \quad \text{and} \quad \Delta\beta_{g,t,\mathcal{K}} := \beta_{g,t,\mathcal{K}} - \beta^0_{g,t,\mathcal{K}^0}.$$

Lemma A.3. *Suppose that Assumptions A1-A5 hold. Then,*

$$\ell_T(\mathcal{K}, \theta) \leq C\bigg\{ \sum_{g=1}^{G} \sum_{l=1}^{N_g} \bar{\ell}_{g,l}(\mathcal{K}, \theta) + \sum_{l=1}^{N_G} \bar{\ell}_{G+1,l}(\mathcal{K}, \theta) + \Delta_T(\mathcal{K}, \theta) \bigg\},$$

where, for $g = 1, \ldots, G$ *and* $l = 1, \ldots, N_g$,

$$\begin{aligned}
\bar{\ell}_{g,l}(\mathcal{K}, \theta) &:= \left(\Big\| \sum_{t=\tau_{g,l-1}+1}^{\tau_{gl}} X_{tT} \Sigma_{t,\mathcal{K}}^{-1} u_t \Big\| - (\tau_{gl} - \tau_{g,l-1}) \big\| \Delta\beta_{g,\tau_{gl},\mathcal{K}} \big\| \right) \big\| \Delta\beta_{g,\tau_{gl},\mathcal{K}} \big\|, \\
\bar{\ell}_{G+1,l}(\mathcal{K}, \theta) &:= \sum_{i=1}^{n} \left(\Big\| \sum_{t=\tau_{G,l-1}+1}^{\tau_{Gl}} (\eta_t \eta_t' - I_n) \Big\| - (\tau_{Gl} - \tau_{G,l-1}) |\lambda_{il}^{\Psi}| \right) |\lambda_{il}^{\Psi}|, \\
\Delta_T(\mathcal{K}, \theta) &:= \max_{1 \le t \le T} \|\Delta\beta_{t,\mathcal{K}}\|.
\end{aligned}$$

Proof. We can write $\log f(y_t|X_{tT}, \theta_{t,\mathcal{K}}) = -(1/2)\big(\log(2\pi)^n + \log|\Sigma_{t,\mathcal{K}}| + \|\Sigma_{t,\mathcal{K}}^{-1/2}(u_t - X_{tT}'\Delta\beta_{t,\mathcal{K}})\|^2\big)$, which implies that

$$\begin{aligned}
\ell_T(\mathcal{K}, \theta) &= -\frac{1}{2} \sum_{t=1}^{T} \Big(\log\big|\Sigma_{t,\mathcal{K}}\big| - \log\big|\Sigma_{t,\mathcal{K}^0}^0\big| + \big\|\Sigma_{t,\mathcal{K}}^{-1/2} u_t\big\|^2 - \big\|(\Sigma_{t,\mathcal{K}^0}^0)^{-1/2} u_t\big\|^2 \Big) \\
&\quad + \sum_{t=1}^{T} \Delta\beta_{t,\mathcal{K}}' X_{tT} \Sigma_{t,\mathcal{K}}^{-1} u_t - \frac{1}{2} \sum_{t=1}^{T} \big\|\Sigma_{t,\mathcal{K}}^{-1/2} X_{tT}' \Delta\beta_{t,\mathcal{K}}\big\|^2 \\
&=: A_1 + A_2 + A_3.
\end{aligned}$$

For the term A_1, we write $\log\big|\Sigma_{t,\mathcal{K}}\big| - \log\big|\Sigma_{t,\mathcal{K}^0}^0\big| = \log\big|(\Sigma_{t,\mathcal{K}^0}^0)^{-1/2} \Sigma_{t,\mathcal{K}} (\Sigma_{t,\mathcal{K}^0}^0)^{-1/2}\big|$ and also $u_t = (\Sigma_{t,\mathcal{K}^0}^0)^{1/2} \eta_t$. Since A_1 depends only on $\mathcal{K}_G$ and $\mathcal{K}_G^0$, we have

$$A_1 = \sum_{l=1}^{N_G} \left\{ -\frac{1}{2} \sum_{t=\tau_{G,l-1}+1}^{\tau_{Gl}} \Big(\log\big|I_n + \Psi_l\big| + \mathrm{tr}\big((I_n + \Psi_l)^{-1} \eta_t \eta_t'\big) - \mathrm{tr}\big(\eta_t \eta_t'\big) \Big) \right\} =: \sum_{l=1}^{N_G} A_{1,l}.$$

For every $l = 1, \ldots, N_G$, we have that $\log\big|I_n + \Psi_l\big| = \sum_{i=1}^{n} \log(1 + \lambda_{li}^{\Psi})$ and that

$$\mathrm{tr}\big((I_n + \Psi_l)^{-1} \eta_t \eta_t'\big) = \mathrm{tr}\left(\mathrm{diag}\left(\left\{ \frac{1}{1 + \lambda_{li}^{\Psi}} \right\}_{i=1}^{n} \right) U' \big(\eta_t \eta_t'\big) U \right),$$

which leads to

$$A_{1,l} = -\frac{\tau_{Gl} - \tau_{G,l-1}}{2} \sum_{i=1}^{n} \log(1 + \lambda_{li}^{\Psi}) + \frac{1}{2} \mathrm{tr}\left(\mathrm{diag}\left(\left\{ \frac{\lambda_{li}^{\Psi}}{1 + \lambda_{li}^{\Psi}} \right\}_{i=1}^{n} \right) U' \Big(\sum_{t=\tau_{G,l-1}+1}^{\tau_{Gl}} \eta_t \eta_t' \Big) U \right).$$

We can show that $-\log(1 + a) + a/(1 + a) \le -a^2/(1 + a)$ for $0 < a < \infty$ (see Dragomir, 2016, for instance). Thus,

$$A_{1,l} \le -\frac{\tau_{Gl} - \tau_{G,l-1}}{2} \sum_{i=1}^{n} \frac{|\lambda_i^{\Psi}|^2}{1 + \lambda_{li}^{\Psi}} + \frac{1}{2} \mathrm{tr}\left(\mathrm{diag}\left(\left\{ \frac{\lambda_{li}^{\Psi}}{1 + \lambda_{li}^{\Psi}} \right\}_{i=1}^{n} \right) U' \Big(\sum_{t=\tau_{G,l-1}+1}^{\tau_{Gl}} (\eta_t \eta_t' - I_n) \Big) U \right).$$

Since the maximum of the diagonal elements of $U'\big(\sum_{t=\tau_{G,l-1}+1}^{\tau_{Gl}} (\eta_t \eta_t' - I_n)\big)U$ is bounded from above by $\|U'\big(\sum_{t=\tau_{G,l-1}+1}^{\tau_{Gl}} (\eta_t \eta_t' - I_n)\big)U\|$ with $\|U\| = 1$, we have

$$A_{1,l} \le \frac{1}{2} \sum_{i=1}^{n} \left\{ -(\tau_{Gl} - \tau_{G,l-1}) \frac{|\lambda_{li}^{\Psi}|^2}{1 + \lambda_{li}^{\Psi}} + \frac{|\lambda_{li}^{\Psi}|}{1 + \lambda_{li}^{\Psi}} \Big\| \sum_{t=\tau_{G,l-1}+1}^{\tau_{Gl}} (\eta_t \eta_t' - I_n) \Big\| \right\}. \qquad \text{(A.12)}$$

From the compactness of Θ and (A.11), we have $\max_{1 \le i \le n}(1 + \lambda_{li}^{\Psi}) = \|I_n + \Psi_l\| \le C_1$ and

$$1 + \min_{1 \le i \le n} \lambda_{li}^{\Psi} = \min_{a \in \mathbb{R}^n} \frac{a'(I_n + \Psi_l)a}{a'a} \ge \left(\min_{b \in \mathbb{R}^n} \frac{b' \Sigma_{\tau_{Gl},\mathcal{K}} b}{b'b} \right) \times \left(\min_{a \in \mathbb{R}^n} \frac{a' (\Sigma_{\tau_{Gl},\mathcal{K}^0}^0)^{-1} a}{a'a} \right) \ge C_2.$$

Thus we have that $C_2 \leq 1 + \lambda_{li}^{\Psi} \leq C_1$ for all $i = 1, \ldots, n$. This together with (A.12) yields

$$A_{1,l} \leq C_3 \sum_{i=1}^{n} \left\{ -(\tau_{Gl} - \tau_{G,l-1})|\lambda_{li}^{\Psi}|^2 + |\lambda_{li}^{\Psi}| \left\| \sum_{t=\tau_{G,l-1}+1}^{\tau_{Gl}} (\eta_t \eta_t' - I_n) \right\| \right\}.$$

It follows that $A_1 \leq C_4 \sum_{l=1}^{N_G} \bar{\ell}_{G+1,l}(\mathcal{K}, \theta)$.

We now consider A_2 and A_3. Note that $\Delta\beta_{t,\mathcal{K}} = \sum_{g=1}^{G} \Delta\beta_{g,t,\mathcal{K}}$, and

$$A_2 = \sum_{g=1}^{G} \sum_{t=1}^{T} \Delta\beta_{g,t,\mathcal{K}}' X_{tT} \Sigma_{t,\mathcal{K}}^{-1} u_t. \tag{A.13}$$

Also, given $X_{tT} \Sigma_{t,\mathcal{K}}^{-1} X_{tT}' = S'(x_{tT} x_{tT}' \otimes \Sigma_{\tau_l,\mathcal{K}}^{-1})S$ for $\tau_{l-1} + 1 \leq t \leq \tau_l$, we can show that

$$A_3 = \sum_{l=1}^{N} \left\{ -\frac{1}{2} \left\| \left(\sum_{t=\tau_{l-1}+1}^{\tau_l} x_{tT} x_{tT}' \otimes \Sigma_{\tau_l,\mathcal{K}}^{-1} \right)^{1/2} S\Delta\beta_{\tau_l,\mathcal{K}} \right\|^2 \right\} =: \sum_{l=1}^{N} A_{3,l}.$$

Under Assumption A1, there exists a finite integer k_0 such that the minimum eigenvalue of $(\tau_l - \tau_{l-1})^{-1} \sum_{t=\tau_{l-1}+1}^{\tau_l} x_{tT} x_{tT}'$ is strictly positive for every $(\tau_l - \tau_{l-1}) \geq k_0$ and also the eigenvalues of $\Sigma_{\tau_l,\mathcal{K}}$ take finite positive values in Θ from Assumption A5. Thus, an application of the result that $\min_{1 \leq i \leq n} \lambda_i(A)\|b\|^2 \leq b'Ab \leq \max_{1 \leq i \leq n} \lambda_i(A)\|b\|^2$ for an $n \times 1$ vector b and an $n \times n$ symmetric matrix A with eigenvalues $\{\lambda_i(A)\}_{i=1}^{n}$ yields that, when $\tau_l - \tau_{l-1} \geq k_0$,

$$A_{3,l} \leq -C_5(\tau_l - \tau_{l-1})\|S\Delta\beta_{\tau_l,\mathcal{K}}\|^2 \leq -C_6(\tau_l - \tau_{l-1})\|\Delta\beta_{\tau_l,\mathcal{K}}\|^2, \tag{A.14}$$

where the last inequality is due to the fact that $S'S$ is positive definite.[14] When $\tau_l - \tau_{l-1} < k_0$, we have that $(\tau_l - \tau_{l-1})\|\Delta\beta_{\tau_l,\mathcal{K}}\|^2 \leq C_7\|\Delta\beta_{\tau_l,\mathcal{K}}\|^2$, which yields

$$A_{3,l} \leq 0 \leq -C_8(\tau_l - \tau_{l-1})\|\Delta\beta_{\tau_l,\mathcal{K}}\|^2 + C_9\|\Delta\beta_{\tau_l,\mathcal{K}}\|^2. \tag{A.15}$$

It follows from (A.14) and (A.15) that $A_3 \leq -C_{10} \sum_{l=1}^{N} (\tau_l - \tau_{l-1}) \left\| \beta_{\tau_l,\mathcal{K}} - \beta_{\tau_l,\mathcal{K}^0}^0 \right\|^2 + C_{11}\Delta_T(\mathcal{K}, \theta)$. Also, we can show that $\sum_{l=1}^{N} (\tau_l - \tau_{l-1}) \left\| \Delta\beta_{\tau_l,\mathcal{K}} \right\|^2 = \sum_{t=1}^{T} \left\| \Delta\beta_{t,\mathcal{K}} \right\|^2$ and that $\left\| \Delta\beta_{t,\mathcal{K}} \right\|^2 = \sum_{g=1}^{G} \left\| \Delta\beta_{g,t,\mathcal{K}} \right\|^2$ because $(\Delta\beta_{g_1,t,\mathcal{K}})' \Delta\beta_{g_2,t,\mathcal{K}} = 0$ for all $g_1, g_2 \in \{1, \ldots, G\}$ with $g_1 \neq g_2$. Thus,

$$A_3 \leq -C_{12} \sum_{g=1}^{G} \sum_{t=1}^{T} \left\| \Delta\beta_{g,t,\mathcal{K}} \right\|^2 + C_{13}\Delta_T(\mathcal{K}, \theta). \tag{A.16}$$

For each $g = 1, \ldots, G$, we have partitions $\{[\tau_{g,l-1} + 1, \tau_{gl}]\}$ of an interval $[1, T]$. From, (A.13) and (A.16), $A_2 + A_3 \leq C_{14}\{\sum_{g=1}^{G} \sum_{l=1}^{N_g} \bar{\ell}_{g,l}(\mathcal{K}, \theta) + \Delta_T(\mathcal{K}, \theta)\}$. Hence, the result follows. ■

We shall establish several properties of the terms $\{\bar{\ell}_{g,l}(\mathcal{K}, \theta)\}_{g=1}^{G+1}$ based on subsamples free from structural changes. To this end, we consider a sequence $\{\xi_t\}_{t=1}^{T}$ of some random vectors or matrices satisfying the condition under which the Hajek-Renyi inequality in Lemma A.2 holds. Let γ be a parameter vector or matrix as an element of the bounded parameter space $\Gamma := \{\gamma : \|\gamma\| \leq C\}$. We define an object depending on a subsample of k observations free from structural changes in γ, namely for $k = 1, \ldots, T$,

$$\ell_k^{(0)}(\gamma) := \left(\left\| \sum_{t=1}^{k} \xi_t \right\| - k\|\gamma\| \right) \|\gamma\|.$$

[14]The selection matrix S is of dimension $nq \times p$ with full column rank and thus $Sv \neq 0$ for all $v \in \mathbb{R}^p$ with $v \neq 0$. It follows that $v'S'Sv \neq 0$ for all $v \in \mathbb{R}^p$ with $v \neq 0$ and $S'S$ positive definite. This implies that there exists a constant $c > 0$ such that $\|Sb\| \geq c\|b\|$ for any $b \in \mathbb{R}^p$.

We now establish a series of properties related to the likelihood function that will enable us to prove the rate of convergence of the estimates. Under the level of generality adopted here, one can apply the arguments used in Bai et al. (1998) to prove the properties of the likelihood function with some modifications. However, since these properties are key ingredients to prove theorems, we provide the whole proof.

Property 1. $\sup_{1\le k\le T}\sup_{\gamma\in\Gamma}\ell_k^{(0)}(\gamma)\le\big|O_p\big(\log T\big)\big|$.

Proof. Let $D>0$ and define $\Gamma_{1,k}(D):=\{\gamma\in\mathcal{G}:\sqrt{k}\|\gamma\|\le D(\log T)^{1/2}\}$ for $1\le k\le T$. We can write $\ell_k^{(0)}(\gamma)=\big(k^{-1/2}\|\sum_{t=1}^k\xi_t\|-\sqrt{k}\|\gamma\|\big)\sqrt{k}\|\gamma\|$ for every $1\le k\le T$. It follows that, for any $1\le k\le T$,

$$\sup_{\gamma\in\Gamma\backslash\Gamma_{1,k}(D)}\ell_k^{(0)}(\gamma)\le\sup_{\gamma\in\Gamma\backslash\Gamma_{1,k}(D)}\left(\frac{1}{\sqrt{k}}\left\|\sum_{t=1}^k\xi_t\right\|-D(\log T)^{1/2}\right)\sqrt{k}\|\gamma\|,$$

and

$$\sup_{\gamma\in\Gamma_{1,k}(D)}\ell_k^{(0)}(\gamma)\le\frac{1}{\sqrt{k}}\left\|\sum_{t=1}^k\xi_t\right\|D(\log T)^{1/2}.$$

Lemma A.2 implies that, for any $B_1>0$,

$$\Pr\left\{\sup_{1\le k\le T}\frac{1}{\sqrt{k\log T}}\left\|\sum_{t=1}^k\xi_t\right\|\ge B_1\right\}\le\frac{C_1}{B_1^2\log T}\sum_{k=1}^T\frac{1}{k}.$$

The right-hand side of the above inequality becomes arbitrarily small for a sufficiently large B_1 because $\sum_{k=1}^T k^{-1}=O(\log T)$. Thus, $\sup_{1\le k\le T}k^{-1/2}\|\sum_{t=1}^k\xi_t\|-D(\log T)^{1/2}<0$ with probability approaching 1 for a sufficiently large D, so that

$$\sup_{1\le k\le T}\sup_{\gamma\in\Gamma\backslash\Gamma_{1,k}(D)}\ell_k^{(0)}(\gamma)\le-C_2D^2\log T\quad\text{and}\quad\sup_{1\le k\le T}\sup_{\gamma\in\Gamma_{1,k}(D)}\ell_k^{(0)}(\gamma)\le C_3D\log T,$$

with probability approaching 1. Hence, the desired conclusion follows. ■

Property 2. *For any $D>0$, there exists a constant $A>0$ such that, for any deterministic sequence $m_T\ge Av_T^{-2}$,*

$$\sup_{m_T\le k\le T}\sup_{\gamma:\|\gamma\|\ge Dv_T}\ell_k^{(0)}\big(\gamma\big)\le-\big|O_p\big((Dv_T)^2m_T\big)\big|.$$

Proof. Let $D>0$ be fixed. We have, for every $1\le k\le T$,

$$\sup_{\gamma:\|\gamma\|\ge Dv_T}\frac{1}{k}\ell_k^{(0)}(\gamma)\le\sup_{\gamma:\|\gamma\|\ge Dv_T}\left(\frac{1}{k}\left\|\sum_{t=1}^k\xi_t\right\|-Dv_T\right)\|\gamma\|.$$

Lemma A.2 yields that, for any $A>0$ and for any $\epsilon>0$,

$$\Pr\left\{\sup_{Av_T^{-2}\le k\le T}\frac{1}{kv_T}\left\|\sum_{t=1}^k\xi_{tT}\right\|>\epsilon\right\}\le\frac{C_1}{\epsilon^2}\left(\frac{1}{A}+\frac{1}{v_T^2}\sum_{k=Av_T^{-2}}^T\frac{1}{k^2}\right).\tag{A.17}$$

Because $\sum_{k=Av_T^{-2}}^T k^{-2}=O\big((Av_T^{-2})^{-1}\big)$, we can show that the right-hand side of (A.17) becomes arbitrarily small for a sufficiently large $A>0$. Since ϵ can be arbitrarily small, there exists an A such that

$$\sup_{Av_T^{-2}\le k\le T}\sup_{\gamma:\|\gamma\|\ge Dv_T}\frac{1}{k}\ell_k^{(0)}\big(\gamma\big)\le-C_2(Dv_T)^2.$$

with probability approaching 1. The result follows because $-m_T^{-1} \leq -k^{-1}$ when $k \geq m_T$. ■

Property 3. *Let* $\Gamma_3(D) := \{\gamma \in \mathcal{G} : \sqrt{T}\|\gamma\| \leq D\}$ *for any* $D > 0$. *Then, for any* $\delta \in (0,1)$, *(a) there exists a* $D > 0$ *such that*

$$\sup_{\delta T \leq k \leq T} \sup_{\gamma \in \Gamma \backslash \Gamma_3(D)} \ell_k^{(0)}(\gamma) \leq -|O_p(D^2)|,$$

(b) for any $D > 0$,

$$\sup_{\delta T \leq k \leq T} \sup_{\gamma \in \Gamma_3(D)} \ell_k^{(0)}(\gamma) = O_p(D).$$

Proof. Let $\delta \in (0,1)$ be fixed. Then, we have, for every $\delta T \leq k \leq T$ and for any $D > 0$,

$$\sup_{\gamma \in \Gamma \backslash \Gamma_3(D)} \ell_k^{(0)}(\gamma) \leq \sup_{\gamma \in \Gamma \backslash \Gamma_3(D)} \left(\frac{1}{\sqrt{T}} \left\| \sum_{t=1}^{k} \xi_t \right\| - \delta D \right) \sqrt{T}\|\gamma\|, \tag{A.18}$$

and

$$\sup_{\gamma \in \Gamma_3(D)} |\ell_k^{(0)}(\gamma)| \leq \frac{1}{\sqrt{T}} \left\| \sum_{t=1}^{k} \xi_t \right\| D. \tag{A.19}$$

Lemma A.2 implies that $\sup_{\delta T \leq k \leq T} \left\| \sum_{t=1}^{k} \xi_t \right\| = O_p(\sqrt{T})$. It follows from (A.18) that, for some $D > 0$, $\sup_{\delta T \leq k \leq T} \sup_{\gamma \in \Gamma \backslash \Gamma_3(D)} \ell_k^{(0)}(\gamma) \leq -C_1 D^2$ with probability approaching 1, while it follows from (A.19) that $\sup_{\delta T \leq k \leq T} \sup_{\gamma \in \Gamma_3(D)} |\ell_k^{(0)}(\gamma)| \leq C_2 D$ with probability approaching 1, for any $D > 0$. Hence, the desired result follows. ■

Property 4. *For any constant* $M > 0$ *and a deterministic sequence* $b_T > 0$, *we have*

$$\sup_{1 \leq k \leq M v_T^{-2}} \sup_{\gamma : \|\gamma\| \leq b_T} \ell_k^{(0)}(\gamma) = O_p(M^{1/2} v_T^{-1} b_T).$$

Proof. We have that $\sup_{1 \leq k \leq M v_T^{-2}} \sup_{\gamma : \|\gamma\| \leq b_T} |\ell_k^{(0)}(\gamma)| \leq \sup_{1 \leq k \leq M v_T^{-2}} \| \sum_{t=1}^{k} \xi_t \| b_T$ for any $M > 0$. Lemma A.2 yields $\sup_{1 \leq k \leq M v_T^{-2}} \| \sum_{t=1}^{k} \xi_t \| \leq O_p\big((M v_T^{-2})^{1/2}\big)$. ■

For $\tau_{G,l-1} + 1 \leq t \leq \tau_{Gl}$, we can show that

$$\|\Psi_l\| \leq \|(\Sigma_{t,\mathcal{K}^0}^0)^{-1/2}\|^2 \|\Sigma_{t,\mathcal{K}} - \Sigma_{t,\mathcal{K}^0}^0\| \quad \text{and} \quad \|\Sigma_{t,\mathcal{K}} - \Sigma_{t,\mathcal{K}^0}^0\| \leq \|(\Sigma_{t,\mathcal{K}^0}^0)^{1/2}\|^2 \|\Psi_l\|,$$

Since $\|(\Sigma_{t,\mathcal{K}^0}^0)^{1/2}\|$ and $\|(\Sigma_{t,\mathcal{K}^0}^0)^{-1/2}\|$ are bounded and $\|\Psi_l\| = \max_{1 \leq i \leq n} |\lambda_{il}^{\Psi}|$, we have

$$d_1 \|\Sigma_{t,\mathcal{K}} - \Sigma_{t,\mathcal{K}^0}^0\| \leq \max_{1 \leq i \leq n} |\lambda_{il}^{\Psi}| \leq d_2 \|\Sigma_{t,\mathcal{K}} - \Sigma_{t,\mathcal{K}^0}^0\|,$$

for some constants $d_1, d_2 > 0$. This relation will be used when we restrict the space for the covariance matrix of the error. The next proposition presents a result about the break date estimates.

Proposition A.1. *Under Assumptions A1-A5, there exists a* $B > 0$ *such that*

$$\lim_{T \to \infty} \Pr\left\{ \left| \hat{k}_{gj} - k_{gj}^0 \right| > B v_T^{-2} \log T \right\} = 0,$$

for every $(g,j) \in \{1, ..., G\} \times \{1, ..., m\}$.

Proof. For a constant $B > 0$, define

$$\ddot{\Xi}(B) := \Big\{\mathcal{K} \in \Xi_\nu : \max_{1\le g\le G} \max_{1\le j\le m} |k_{gj} - k_{gj}^0| \le B v_T^{-2} \log T\Big\}.$$

To prove the assertion, we shall show that, for a sufficiently large $B > 0$,

$$\lim_{T\to\infty} \Pr\Big\{ \sup_{(\mathcal{K},\theta)\in\Xi_\nu\setminus\ddot{\Xi}(B)\times\Theta} \ell_T\big(K,\theta\big) \ge 0\Big\} = 0. \tag{A.20}$$

Since the normalized log likelihood evaluated at the maximum likelihood estimates should be non-negative, the desired conclusion follows from (A.20).

To show (A.20), we examine the upper bound in Lemma A.3 given sets of break dates $\mathcal{K} \notin \ddot{\Xi}(B)$ and $\mathcal{K}^0$. First, observe that Property 1 provides a not necessarily sharp but general upper bound in probability and that the parameter space is bounded. Thus,

$$\sup_{(\mathcal{K},\theta)\in\Xi_\nu\setminus\ddot{\Xi}(B)\times\Theta} \bar{\ell}_{g,l}(\mathcal{K},\theta) \le |O_P(\log T)| \quad \text{and} \quad \sup_{(\mathcal{K},\theta)\in\Xi_\nu\setminus\ddot{\Xi}(B)\times\Theta} \Delta(\mathcal{K},\theta) \le C_1, \tag{A.21}$$

for every $1 \le g \le G+1$ and $1 \le l \le 2(m{+}1)$.

Next, for $\mathcal{K} \notin \ddot{\Xi}(B)$, there exits a pair $(g,j) \in \{1,...,G\} \times \{1,...,m\}$ such that some neighborhood $\mathcal{N}_{gj} := \{t \in [1,T] : |t - k_{gj}^0| \le B v_T^{-2} \log T\}$ of a true break date, k_{gj}^0, contains none of the break dates $\mathcal{K}_g$ of the g^{th} group, i.e., $\mathcal{K}_g \not\subset \mathcal{N}_{gj}$. This implies that there is a $\tau_{gl} = k_{gj}^0$ with a union of sub-intervals

$$[\tau_{g,l-1}{+}1, \tau_{gl}] \cup [\tau_{gl}{+}1, \tau_{g,l+1}] \quad \text{with} \quad \min_{l\le j\le l+1} (\tau_{gj} - \tau_{g,j-1}) \ge B v_T^{-2} \log T.$$

Since $\mathcal{K}_g \not\subset (\tau_{g,l-1}, \tau_{g,l+1})$, the g^{th} group estimates are constant for $\tau_{g,l-1} + 1 \le t \le \tau_{g,l+1}$ and both $\bar{\ell}_{g,l}(\mathcal{K},\theta)$ and $\bar{\ell}_{g,l+1}(\mathcal{K},\theta)$ depend on the same g^{th} group estimates. Note that the triangle inequality yields that

$$C_2 v_T \le 2\max\big\{\big\|\beta_{g,\tau_{g,l+1},\mathcal{K}} - \beta^0_{g,\tau_{gl},\mathcal{K}^0}\big\|, \big\|\beta_{g,\tau_{g,l+1},\mathcal{K}} - \beta^0_{g,\tau_{g,l+1},\mathcal{K}^0}\big\|\big\},$$

and additionally when $g = G$,

$$C_3 v_T \le 2\max\big\{\big\|\Sigma_{\tau_{G,l+1},\mathcal{K}} - \Sigma^0_{\tau_{Gl},\mathcal{K}^0}\big\|, \big\|\Sigma_{\tau_{G,l+1},\mathcal{K}} - \Sigma^0_{\tau_{G,l+1},\mathcal{K}^0}\big\|\big\}.$$

This implies that either $\bar{\ell}_{g,l}(\mathcal{K},\theta)$ or $\bar{\ell}_{g,l+1}(\mathcal{K},\theta)$ satisfies the condition in Property 2 with $m_T = B v_T^{-2} \log T$, which together with (A.21) implies that, for a sufficiently large B,

$$\sup_{(\mathcal{K},\theta)\in\Xi_\nu\setminus\ddot{\Xi}(B)\times\Theta} \ell_T(\mathcal{K},\theta) \le -|O_p(B\log T)| + O_p(\log T).$$

This yields (A.20) and thus completes the proof. ■

Proposition A.2. *Suppose that Assumptions A1-A5 hold. Then,*

$$\hat{\beta}_{gj} - \beta_{gj}^0 = o_p(v_T) \quad \text{and} \quad \hat{\Sigma}_j - \Sigma_j^0 = o_p(v_T),$$

for every $(g,j) \in \{1,...,G\} \times \{1,...,m+1\}$.

Proof. Let $\epsilon > 0$ be fixed and define a subset of the parameter space Θ:

$$\ddot{\Theta}(\epsilon) := \Big\{\theta \in \Theta : \max_{1\le g\le G} \max_{1\le j\le m+1} \|\beta_{gj} - \beta_{gj}^0\| \le \epsilon v_T \text{ and } \max_{1\le j\le m+1} \|\Sigma_j - \Sigma_j^0\| \le \epsilon v_T\Big\}.$$

Proposition A.1 shows that the break date estimates $\hat{\mathcal{K}}$ are included in $\ddot{\Xi}(B)$ with probability approaching 1 for a sufficiently large B and thus we consider the case where $\mathcal{K} \in \ddot{\Xi}(B)$. For $\theta \in \Theta \setminus \ddot{\Theta}(\epsilon)$, there exists a pair $(g,j) \in \{1,...,G\} \times \{1,...,m\}$ such that either

$$\|\beta_{gj} - \beta_{gj}^0\| \ge \epsilon v_T \quad \text{or} \quad \|\Sigma_j - \Sigma_j^0\| \ge \epsilon v_T. \tag{A.22}$$

Observe that $k_{gj} - k_{g,j-1} \geq \nu T$ and $k^0_{gj} - k^0_{g,j-1} \geq \nu T$, while $|k_{gj} - k^0_{gj}| \leq Bv_T^{-2} \log T$. For some $l \in \{1, \dots, N_g\}$, we have $\tau_{g,l-1} = \max\{k_{g,j-1}, k^0_{g,j-1}\}$ and $\tau_{gl} = \max\{k_{g,j}, k^0_{gj}\}$ satisfying $\tau_{gl} - \tau_{g,l-1} \geq \delta T$ for some $\delta \in (0, 1)$ and that (A.22) holds over a sub-interval $[\tau_{g,l-1} + 1, \tau_{gl}]$. Thus, Property 2 with $m_T = \delta T$ implies that

$$\sup_{(\mathcal{K},\theta)\in\ddot{\Xi}(B)\times\Theta\backslash\ddot{\Theta}(\epsilon)} \bar{\ell}_{g,l}(\mathcal{K}, \theta) \leq -|O_p(\epsilon^2 T v_T^2)|.$$

For the other sub-intervals, Property 1 provides an upper bound of order $|O_p(\log T)|$. Since $\sqrt{T} v_T / \log T \to \infty$ as $T \to \infty$, we can show that

$$\sup_{(\mathcal{K},\theta)\in\ddot{\Xi}(B)\times\Theta\backslash\ddot{\Theta}(\epsilon)} \ell_T\big(\mathcal{K}, \theta\big) \leq -|O_p(\epsilon^2 T v_T^2)|.$$

This leads to the desired result. ■

Propositions A.1 and A.2 are important intermediate steps to establish the convergence rates of the estimates as stated in the theorem below. A similar approach was used by Bai et al. (1998), Bai (2000) and Qu and Perron (2007) when break dates are assumed to either have a common location or be asymptotically distinct. A key difference between their approach and ours is that we allow for the possibility that the break dates associated with different basic parameters may not be asymptotically distinct.

Proof of Theorem 1. **(a)** Proposition A.1 shows that $\hat{\mathcal{K}} \in \ddot{\Xi}(B)$ with probability approaching 1 for some $B > 0$, while both $\hat{\mathcal{K}}$ and $\mathcal{K}^0$ are included in Ξ_ν. Thus, it suffices to consider the case where either $\tau_{gl} - \tau_{g,l-1} \geq \delta T$ for some $\delta > 0$ or $\tau_{gl} - \tau_{g,l-1} \leq Bv_T^{-2} \log T$ for every $(g, l) \in \{1, \dots, G\} \times \{1, \dots, N\}$. If $\tau_{gl} - \tau_{g,l-1} \geq \delta T$, then Property 3 implies that

$$\sup_{(\mathcal{K},\theta)\in\ddot{\Xi}(B)\times\Theta} \bar{\ell}_{g,l}(\mathcal{K}, \theta) \leq |O_p(1)|. \tag{A.23}$$

When $\tau_{gl} - \tau_{g,l-1} \leq Bv_T^{-2} \log T$, there are two cases: $Mv_T^{-2} \leq \tau_{gl} - \tau_{g,l-1} \leq Bv_T^{-2} \log T$ and $\tau_{gl} - \tau_{g,l-1} \leq Mv_T^{-2}$ for some $M > 0$. For sake of concreteness, let $\tau_{g,l-1} = k^0_{gj}$ and $\tau_{gl} = \hat{k}_{gj}$ in both cases. When $k^0_{gj} + 1 \leq t \leq \hat{k}_{gj}$, we have $(\hat{\beta}_{g,t,\hat{\mathcal{K}}}, \beta^0_{g,t,\mathcal{K}^0}) = (\hat{\beta}_{gj}, \beta^0_{g,j+1})$ for $1 \leq g \leq G$ and $(\hat{\Sigma}_{t,\hat{\mathcal{K}}}, \Sigma^0_{t,\mathcal{K}^0}) = (\hat{\Sigma}_j, \Sigma^0_{j+1})$ for $g = G$. Since $\|\beta^0_{g,j+1} - \beta^0_{gj}\| = v_T \|\delta_{gj}\|$ and $\|\Sigma^0_{j+1} - \Sigma^0_j\| = v_T \|\Phi_j\|$, we can show[15]

$$\Big| \|\hat{\beta}_{gj} - \beta^0_{g,j+1}\| - v_T \|\delta_{gj}\| \Big| \leq \|\hat{\beta}_{gj} - \beta^0_{gj}\| \quad \text{and} \quad \Big| \|\hat{\Sigma}_j - \Sigma^0_{j+1}\| - v_T \|\Phi_j\| \Big| \leq \|\hat{\Sigma}_j - \Sigma^0_j\|.$$

Moreover, Proposition A.2 shows that $\|\hat{\beta}_{gj} - \beta^0_{gj}\| = o_p(v_T)$ and $\|\hat{\Sigma}_j - \Sigma^0_j\| = o_p(v_T)$. Thus,

$$\|\hat{\beta}_{gj} - \beta^0_{g,j+1}\| = v_T \|\delta_{gj}\| + o_p(v_T) \quad \text{and} \quad \|\hat{\Sigma}_j - \Sigma^0_{j+1}\| = v_T \|\Phi_j\| + o_p(v_T). \tag{A.24}$$

When $Mv_T^{-2} \leq \tau_{gl} - \tau_{g,l-1} \leq Bv_T^{-2} \log T$, Property 2 together with (A.24) implies that

$$\bar{\ell}_{g,l}(\hat{\mathcal{K}}, \hat{\theta}) \leq -|O_p(M)|, \tag{A.25}$$

for a sufficiently large M, while, for $\tau_{gl} - \tau_{g,l-1} \leq Mv_T^{-2}$, Property 4 with (A.24) implies

$$\bar{\ell}_{g,l}(\hat{\mathcal{K}}, \hat{\theta}) \ = O_p(M^{1/2}). \tag{A.26}$$

Since $\sup_{(\mathcal{K},\theta)\in\ddot{\Xi}(B)\times\ddot{\Theta}(\epsilon)} \Delta(\mathcal{K}, \theta) = o(1)$, Lemma A.3 with (A.23), (A.25) and (A.26) implies

$$\sup_{(\mathcal{K},\theta)\in\ddot{\Xi}(B)\backslash\ddot{\Xi}_M\times\ddot{\Theta}(\epsilon)} \ell_T(\mathcal{K}, \theta) < -|O_p(M)|,$$

[15] To prove this, we use the inequality, $\|a - b\| - \|b - c\| \leq \|a - c\| \leq \|a - b\| + \|b - c\|$ for any elements a, b and c of some space with the norm $\| \cdot \|$, which is due to the triangle inequality.

for a sufficiently large M. This completes the proof of part (a).

(b) From part (a), there exists an $M > 0$ such that $\max_{1\le g\le G}\max_{1\le j\le m}|\hat{k}_{gj} - k^0_{gj}| \le Mv_T^{-2}$ with probability approaching 1. Thus it suffices to consider the case where either $\tau_{gl} - \tau_{g,l-1} \le Mv_T^{-2}$ or $\tau_{gl} - \tau_{g,l-1} > \delta T$ for some $\delta > 0$. As in (A.23) and (A.26), we can show that $\bar{\ell}_{g,l}(\hat{\mathcal{K}}, \hat{\theta})$ is bounded by a term of order $|O_p(1)|$ for every $(g,l) \in \{1,\dots,G{+}1\}\times\{1,\dots,2(m{+}1)\}$. If $\sqrt{T}\|\hat{\beta}_{gj} - \beta^0_{gj}\| \ge M$ for some group and regime (g,j) and for some $M > 0$, then there is a corresponding sub-interval $[\tau_{g,l-1}+1, \tau_{gl}]$ with $\tau_{gl} - \tau_{g,l-1} > \delta T$ and thus Property 3(a) implies that $\bar{\ell}_{g,l}(\hat{\mathcal{K}}, \hat{\theta}) \le -|O_p(M^2)|$ for a sufficiently large M. Thus, on the event that $\max_{1\le g\le G}\max_{1\le j\le m+1}\|\hat{\beta}_{gj} - \beta^0_{gj}\| \ge MT^{-1/2}$ for a sufficiently large M, Lemma A.3 implies that the normalized log likelihood takes negative value with probability approaching 1. The same result holds when $\max_{1\le j\le m+1}\|\hat{\Sigma}_j - \Sigma^0_j\| \ge MT^{-1/2}$ for a sufficiently large M. ■

Having established the convergence rates of the estimates, we are now in a position to prove results about the asymptotic independence of the break date estimates and the estimates of the basic parameters. In order to proceed, we let the likelihood based on the observations in the interval $[t_1, t_2] \subset [1, T]$ be denoted as $L(t_1, t_2; \mathcal{K}, \theta) = \prod_{t=t_1}^{t_2} f(y_t|X_{tT}, \theta_{t,\mathcal{K}})$. Then, using the partition $\{[\tau_{l-1}+1, \tau_l]\}_{l=1}^N$ of an interval $[1,T]$ given $\mathcal{K}$ and $\mathcal{K}^0$, we can express the normalized log likelihood as

$$\ell_T(\mathcal{K}, \theta) = \sum_{l=1}^{N}\big\{\log L(\tau_{l-1}+1, \tau_l; \mathcal{K}, \theta) - \log L(\tau_{l-1}+1, \tau_l; \mathcal{K}^0, \theta^0)\big\}.$$

Proof of Theorem 2. Consider the case where $(\mathcal{K}, \theta) \in \bar{\Xi}_M \times \bar{\Theta}_M$ for a sufficiently large M with the restriction $R(\theta) = 0$. By definition, we can write

$$\begin{aligned}&\ell_T(\mathcal{K}, \theta) - \ell_T(\mathcal{K}^0, \theta) - \ell_T(\mathcal{K}, \theta^0)\\ &\quad= \sum_{l=1}^{N}\big\{\log L(\tau_{l-1}+1, \tau_l; \mathcal{K}, \theta) - \log L(\tau_{l-1}+1, \tau_l; \mathcal{K}^0, \theta)\big\} \qquad \text{(A.27)}\\ &\qquad - \sum_{l=1}^{N}\big\{\log L(\tau_{l-1}+1, \tau_l; \mathcal{K}, \theta^0) - \log L(\tau_{l-1}+1, \tau_l; \mathcal{K}^0, \theta^0)\big\}. \qquad \text{(A.28)}\end{aligned}$$

If $\tau_l - \tau_{l-1} > Mv_T^{-2}$, then we have $\theta_{t,\mathcal{K}} = \theta_{t,\mathcal{K}^0}$ and $\theta^0_{t,\mathcal{K}} = \theta^0_{t,\mathcal{K}^0}$ for all $\tau_{l-1}+1 \le t \le \tau_l$. Thus, it suffices to consider the quantities in (A.27) and (A.28) with the index l satisfying $\tau_l - \tau_{l-1} \le Mv_T^{-2}$. Property 4 with $b_T = MT^{-1/2}$ implies that, uniformly in $(\mathcal{K}, \theta) \in \bar{\Xi}_M \times \bar{\Theta}_M$,

$$\ell_T(\mathcal{K}, \theta) = \ell_T(\mathcal{K}, \theta^0) + \ell_T(\mathcal{K}^0, \theta) + O_p\big((\sqrt{T}v_T)^{-1}\big).$$

Hence, we obtain the desired result. ■

To derive the limit distribution of the test, we first present a technical lemma, which is a direct consequence of Lemma A.1(b). To this end, we introduce some notation. For $j = 1, \dots, m$, we define, for $s < 0$,

$$V^{(1)}_{T,z\eta,j}(-s) := v_T \sum_{t=T^0_j+[sv_T^{-2}]}^{T^0_j} (z_t \otimes \eta_t) \quad \text{and} \quad V^{(1)}_{T,\eta\eta,j}(-s) := v_T \sum_{t=T^0_j+[sv_T^{-2}]}^{T^0_j} (\eta_t\eta_t{}' - I_n),$$

and, for $s > 0$,

$$V^{(2)}_{T,z\eta,j}(s) := v_T \sum_{t=T^0_j}^{T^0_j+[sv_T^{-2}]} (z_t \otimes \eta_t) \quad \text{and} \quad V^{(2)}_{T,\eta\eta,j}(s) := v_T \sum_{t=T^0_j}^{T^0_j+[sv_T^{-2}]} (\eta_t\eta_t' - I_n).$$

Lemma A.4. *Under Assumptions A6-A9 with a sequence v_T defined in Assumption A4, we have, for $j = 1, \dots, m$,*

$$V^{(1)}_{T,z\eta,j}(\cdot) \Rightarrow \mathbb{V}^{(1)}_{z\eta,j}(\cdot) \quad \text{and} \quad V^{(2)}_{T,z\eta,j}(\cdot) \Rightarrow \mathbb{V}^{(2)}_{z\eta,j}(\cdot),$$

where the weak convergence is in the space $D[0,\infty)^{nq}$ and the Brownian motions $\mathbb{V}^{(1)}_{z\eta,j}(\cdot)$ and $\mathbb{V}^{(2)}_{z\eta,j}(\cdot)$ are defined in the main text. Furthermore, for $j = 1, \dots, m$,

$$V^{(1)}_{T,\eta\eta,j}(\cdot) \Rightarrow \mathbb{V}^{(1)}_{\eta\eta,j}(\cdot) \quad \text{and} \quad V^{(2)}_{T,\eta\eta,j}(\cdot) \Rightarrow \mathbb{V}^{(2)}_{\eta\eta,j}(\cdot),$$

where the weak convergence is in the space $D[0,\infty)^{n^2}$ and the $n \times n$ matrices $\mathbb{V}^{(1)}_{\eta\eta,j}(\cdot)$ and $\mathbb{V}^{(2)}_{\eta\eta,j}(\cdot)$ are Brownian motion defined in the main text.

Proof of Lemma 1. Consider a regime $j \in \{1, \dots, m\}$. For $s \in \mathbb{R}$ and for $\underline{T}^0_j(s) \le t \le \overline{T}^0_j(s)$, observe that

$$(\Sigma^0_{t,j+\mathbb{1}_{\{T_j(r)\le t\}}})^{-1} = \begin{cases} (\Sigma^0_{j+1})^{-1}, & \text{if } T_j(r) \le \underline{T}^0_j(s) \\ (\Sigma^0_{j+1})^{-1} - \mathbb{1}_{\{T^0_j < t \le T_j(r)\}}\{(\Sigma^0_{j+1})^{-1} - (\Sigma^0_j))^{-1}\}, & \text{if } T^0_j < T_j(r) \le T^0_j(s) \\ (\Sigma^0_j)^{-1} \quad + \mathbb{1}_{\{T_j(r) < t \le T^0_j\}}\{(\Sigma^0_{j+1})^{-1} - (\Sigma^0_j)^{-1}\}, & \text{if } T^0_j(s) < T_j(r) \le T^0_j \\ (\Sigma^0_j)^{-1}, & \text{if } \overline{T}^0_j(s) \le T_j(r), \end{cases}$$

which yields

$$(\Sigma^0_{t,j+\mathbb{1}_{\{T_j(r)\le t\}}})^{-1} = (\Sigma^0_{j+\mathbb{1}_{\{r\le s\}}})^{-1} - \text{sgn}(r)\mathbb{1}_{\{|r|\le|s|\}}\{(\Sigma^0_{j+1})^{-1} - (\Sigma^0_j)^{-1}\}.$$

Let $D_{T,j}(s) := v_T^2 \sum_{t=\underline{T}^0_j(s)+1}^{\overline{T}^0_j(s)} x_{tT}x_{tT}'$. We have, for every $\underline{T}^0_j(s) \le t \le \overline{T}^0_j(s)$ and for $r \in \mathbb{R}$,

$$\begin{aligned} B_{T,j}(s,r) &= S'D_{T,j}(s) \otimes (\Sigma^0_{j+\mathbb{1}_{\{r\le s\}}})^{-1}S \\ &\quad -\text{sgn}(r)\mathbb{1}_{\{|r|\le|s|\}}S'D_{T,j}(r) \otimes \{(\Sigma^0_{j+1})^{-1} - (\Sigma^0_j)^{-1}\}S, \end{aligned}$$

since $X_{tT}(\Sigma^0_{t,j+\mathbb{1}_{\{T_j(r)\le t\}}})^{-1}X_{tT}' = S'x_{tT}x_{tT}' \otimes (\Sigma^0_{t,j+\mathbb{1}_{\{T_j(r)\le t\}}})^{-1}S$, and also

$$\varphi(t/T) = \varphi(\lambda^0_j) + O\big((\sqrt{T}v_T)^{-2}\big) \quad \text{and} \quad w_t = w_{T^0_j} + O\big((\sqrt{T}v_T)^{-2}\big), \tag{A.29}$$

uniformly in $s \in \mathbb{R}$.[16] Under Assumption A6, we can show that, uniformly in $s \in \mathbb{R}$,

$$v_T^2 \sum_{t=\underline{T}^0_j(s)+1}^{\overline{T}^0_j(s)} z_t = |s|\mu_{z,j+\mathbb{1}_{\{0<s\}}} + o_p(1) \quad \text{and} \quad v_T^2 \sum_{t=\underline{T}^0_j(s)+1}^{\overline{T}^0_j(s)} z_tz_t' = |s|Q_{zz,j+\mathbb{1}_{\{0<s\}}} + o_p(1).$$

[16] We have that $a^r - b^r = (a-b)\sum_{l=0}^{r-1} a^{r-1-l}b^l$ for $a, b \in \mathbb{R}$ and for an integer $r \ge 2$. It follows that $|(t/T)^r - (T^0_j/T)^r| \le C|(t - T^0_j)/T|$.

It follows that, uniformly in $s \in \mathbb{R}$,

$$D_{T,j}(s) = |s| \begin{pmatrix} Q_{zz,j+\mathbb{1}_{\{0<s\}}} & \mu_{z,j+\mathbb{1}_{\{0<s\}}}\varphi(\lambda_j^0)' & \mu_{z,j+\mathbb{1}_{\{0<s\}}}T^{-1/2}w'_{T_j^0} \\ \varphi(\lambda_j^0)\mu'_{z,j+\mathbb{1}_{\{0<s\}}} & \varphi(\lambda_j^0)\varphi(\lambda_j^0)' & \varphi(\lambda_j^0)T^{-1/2}w'_{T_j^0} \\ T^{-1/2}w_{T_j^0}\mu'_{z,j+\mathbb{1}_{\{0<s\}}} & T^{-1/2}w_{T_j^0}\varphi(\lambda_j^0)' & (T^{-1/2}w_{T_j^0})(T^{-1/2}w_{T_j^0})' \end{pmatrix} + o_p(1).$$

Also, we have $X_{tT}(\Sigma^0_{t,j+\mathbb{1}_{\{T_j(r)\le t\}}})^{-1}u_t = S'\big(I \otimes (\Sigma^0_{t,j+\mathbb{1}_{\{T_j(r)\le t\}}})^{-1}\big)(x_{tT} \otimes u_t)$ and $u_t = (\Sigma^0_{j+\mathbb{1}\{0<s\}})^{1/2}\eta_t$. Thus, for $\underline{T}_j^0(s) \le t \le \overline{T}_j^0(s)$,

$$\begin{aligned} W_{T,j}(s,r) &= S'\big(I_q \otimes (\Sigma^0_{j+\mathbb{1}_{\{r\le s\}}})^{-1}\big)V_{T,j}(s) \\ &\quad -\mathrm{sgn}(r)\mathbb{1}_{\{|r|\le|s|\}}S'\big(I_q \otimes \{(\Sigma^0_{j+1})^{-1} - (\Sigma^0_j)^{-1}\}\big)V_{T,j}(r), \end{aligned}$$

where $V_{T,j}(s) := \big(I_q \otimes (\Sigma^0_{j+\mathbb{1}\{0<s\}})^{1/2}\big)v_T \sum_{t=\underline{T}_j^0(s)+1}^{\overline{T}_j^0(s)} (x_{tT} \otimes \eta_t)$. It follows from (A.29) that

$$v_T \sum_{t=\underline{T}_j^0(s)+1}^{\overline{T}_j^0(s)} (x_{tT} \otimes \eta_t) = \left(v_T \sum_{t=\underline{T}_j^0(s)+1}^{\overline{T}_j^0(s)} (z_t \otimes \eta_t)', \Big(\varphi(\lambda_j^0)', T^{-1/2}w'_{T_j^0}\Big) \otimes v_T \sum_{t=\underline{T}_j^0(s)+1}^{\overline{T}_j^0(s)} \eta'_t \right)' + o_p(1),$$

uniformly in $s \in \mathbb{R}$. Hence, Lemma A.4 with the continuous mapping theorem yields $\{B_{T,j}(\cdot), W_{T,j}(\cdot)\}_{j=1}^m \Rightarrow \{\mathbb{B}_j(\cdot), \mathbb{W}_j(\cdot)\}_{j=1}^m$. ■

Proof of Theorem 3. Theorems 1 and 2 imply that, for a sufficiently large $M > 0$,

$$CB_T = 2\Big\{ \sup_{\mathcal{K}\in\bar{\Xi}_M} \ell_T(\mathcal{K},\theta^0) - \sup_{\mathcal{K}\in\bar{\Xi}_{M,H_0}} \ell_T(\mathcal{K},\theta^0)\Big\} + o_p(1). \tag{A.30}$$

Let M be an arbitrary large constant. For $(g,j) \in \{1,\ldots,G\} \times \{1,\ldots,m\}$, define $\boldsymbol{r}_j := (r_{1j},\ldots,r_{Gj})'$ with $r_{gj} \in [-M,M]$ and consider $\mathcal{K} \in \bar{\Xi}_M$ such that $k_{gj} = T_j^0 + [r_{gj}v_T^{-2}]$. Then, we can write $\underline{k}_j = T_j^0 + \min\{[r_{1j}v_T^{-2}],\ldots,[r_{Gj}v_T^{-2}],0\}$ and $\overline{k}_j = T_j^0 + \max\{[r_{1j}v_T^{-2}],\ldots,[r_{Gj}v_T^{-2}],0\}$. Also, $\ell_T(\mathcal{K},\theta^0) = \sum_{j=1}^m \ell_T^{(j)}(\mathbf{r}_j)$, where $\ell_T^{(j)}(\mathbf{r}_j) := \sum_{\underline{k}_j+1}^{\overline{k}_j} \big\{ \log f(y_t|X_{tT},\theta^0_{t,\mathcal{K}}) - \log f(y_t|X_{tT},\theta^0_{t,\mathcal{T}^0})\big\}$. Observe that, for $1 \le t \le T$,

$$\begin{aligned} \log f(y_t|X_{tT},\theta^0_{t,\mathcal{K}}) &= -\frac{1}{2}\Big\{ \log(2\pi)^n + \log|\Sigma^0_{t,\mathcal{K}}| + \|(\Sigma^0_{t,\mathcal{K}})^{-1/2}u_t\|^2 \\ &\qquad -2(\Delta\beta^0_{t,\mathcal{K}})'X_{tT}(\Sigma^0_{t,\mathcal{K}})u_t + \|(\Sigma^0_{t,\mathcal{K}})^{-1/2}X'_{tT}\Delta\beta^0_{t,\mathcal{K}}\|^2\Big\}. \end{aligned}$$

Let $\underline{k}_{Gj} := T_j^0 + \min\{[r_{Gj}v_T^{-2}],0\}$ and $\overline{k}_{Gj} := T_j^0 + \max\{[r_{Gj}v_T^{-2}],0\}$ for $j \in \{1,\ldots,m\}$. Then,

$$\ell_T^{(j)}(\boldsymbol{r}_j) = \ell_{T,1}^{(j)}(\boldsymbol{r}_j) + \ell_{T,2}^{(j)}(\boldsymbol{r}_j),$$

where

$$\begin{aligned} \ell_{T,1}^{(j)}(\boldsymbol{r}_j) &:= \frac{1}{2}\sum_{t=\underline{k}_{Gj}+1}^{\overline{k}_{Gj}} \Big\{ \log\big|\Sigma^0_{t,\mathcal{T}^0}(\Sigma^0_{t,\mathcal{K}})^{-1}\big| + \mathrm{tr}\Big(\big\{(\Sigma^0_{t,\mathcal{T}^0})^{-1} - (\Sigma^0_{t,\mathcal{K}})^{-1}\big\}u_tu'_t\Big)\Big\}, \\ \ell_{T,2}^{(j)}(\boldsymbol{r}_j) &:= \frac{1}{2}\sum_{t=\underline{k}_j+1}^{\overline{k}_j} \big\{2(\Delta\beta^0_{t,\mathcal{K}})'X_{tT}(\Sigma^0_{t,\mathcal{T}^0})^{-1}u_t - \|(\Sigma^0_{t,\mathcal{T}^0})^{-1/2}X'_{tT}\Delta\beta^0_{t,\mathcal{K}}\|^2\big\}. \end{aligned}$$

First, we consider the term $\ell_{T,1}^{(j)}(\boldsymbol{r}_j)$. We can write $\Sigma^0_{t,\mathcal{T}^0}(\Sigma^0_{t,\mathcal{K}})^{-1} = I_n - (\Sigma^0_{t,\mathcal{K}} - \Sigma^0_{t,\mathcal{T}^0})(\Sigma^0_{t,\mathcal{K}})^{-1}$ and $\Sigma^0_{t,\mathcal{K}} - \Sigma^0_{t,\mathcal{T}^0} = v_T\Phi_{t,\mathcal{K}}$, where $\Phi_{t,\mathcal{K}} = \Phi_j$ if $k_{Gj} < t \le T_j^0$ and $\Phi_{t,\mathcal{K}} = -\Phi_j$ if $T_j^0 < t \le k_{Gj}$.

Thus, an application of the Taylor series expansion yields that, for $\underline{k}_{Gj} \le t \le \overline{k}_{Gj}$,

$$\log|\Sigma^0_{t,\mathcal{T}^0}(\Sigma^0_{t,\mathcal{K}})^{-1}| = \mathrm{tr}\big(-v_T\Phi_{t,\mathcal{K}}(\Sigma^0_{t,\mathcal{K}})^{-1}\big) + \frac{1}{2}\mathrm{tr}\big(\big\{v_T\Phi_{t,\mathcal{K}}(\Sigma^0_{t,\mathcal{K}})^{-1}\big\}^2\big) + O_p(v_T^3). \quad \text{(A.31)}$$

Also we can write $(\Sigma^0_{t,\mathcal{T}^0})^{-1}-(\Sigma^0_{t,\mathcal{K}})^{-1} = (\Sigma^0_{t,\mathcal{T}^0})^{-1}(\Sigma^0_{t,\mathcal{K}}-\Sigma^0_{t,\mathcal{T}^0})(\Sigma^0_{t,\mathcal{K}})^{-1}$ and $u_t = (\Sigma^0_{t,\mathcal{T}^0})^{1/2}\eta_t$, which implies, for $\underline{k}_{Gj} \le t \le \overline{k}_{Gj}$,

$$\mathrm{tr}\big(\big\{(\Sigma^0_{t,\mathcal{T}^0})^{-1} - (\Sigma^0_{t,\mathcal{K}})^{-1}\big\}u_tu_t'\big) = \mathrm{tr}\Big((\Sigma^0_{t,\mathcal{T}^0})^{-1/2}v_T\Phi_{t,\mathcal{K}}(\Sigma^0_{t,\mathcal{K}})^{-1}(\Sigma^0_{t,\mathcal{T}^0})^{1/2}\eta_t\eta_t'\Big). \quad \text{(A.32)}$$

For $\underline{k}_{Gj} \le t \le \overline{k}_{Gj}$, we have

$$(\Phi_{t,\mathcal{K}}, \Sigma^0_{t,\mathcal{T}^0}, \Sigma^0_{t,\mathcal{K}}) = \begin{cases} (\Phi_j, \Sigma^0_j, \Sigma^0_{j+1}), & \text{if } r_{Gj} \le 0 \\ (-\Phi_j, \Sigma^0_{j+1}, \Sigma^0_j), & \text{if } r_{Gj} > 0. \end{cases}$$

Using (A.31) and (A.32) with $\pi_j(r_{Gj}) := (\Sigma^0_{t,\mathcal{T}^0})^{-1/2}\Phi_{t,\mathcal{K}}(\Sigma^0_{t,\mathcal{K}})^{-1}(\Sigma^0_{t,\mathcal{T}^0})^{1/2}$, we obtain

$$\ell^{(j)}_{T,1}(\boldsymbol{r}_j) = \frac{1}{2}\mathrm{tr}\Big(\pi_j(r_{Gj})V_{T,\eta\eta,j}(r_{Gj})\Big) + \frac{|r_{Gj}|}{4}\mathrm{tr}\big(\{\pi_j(r_{Gj})\}^2\big) + o_p(1), \quad \text{(A.33)}$$

where $V_{T,\eta\eta,j}(r_{Gj}) := v_T\sum_{t=\underline{k}_{Gj}+1}^{\overline{k}_{Gj}}(\eta_t\eta_t' - I_n)$.

Next, we consider the term $\ell^{(j)}_{T,2}(\boldsymbol{r}_j)$. Define $\Delta\beta^0_{g,t,\mathcal{K}} := \sum_{i\in\mathcal{G}_g} e_i \circ (\beta^0_{t,\mathcal{K}} - \beta^0_{t,\mathcal{T}^0})$. Then $\Delta\beta^0_{t,\mathcal{K}} = \sum_{g=1}^G \Delta\beta^0_{g,t,\mathcal{K}}$ and we have

$$\ell^{(j)}_{T,2}(\boldsymbol{r}_j) = \sum_{t=\underline{k}_j+1}^{\overline{k}_j}\Bigg(\sum_{g=1}^G(\Delta\beta^0_{g,t,\mathcal{K}})'X_{tT}(\Sigma^0_{t,\mathcal{K}})^{-1}u_t - \frac{1}{2}\sum_{g=1}^G\sum_{l=1}^G(\Delta\beta^0_{g,t,\mathcal{K}})'X_{tT}(\Sigma^0_{t,\mathcal{K}})^{-1}X_{tT}'\Delta\beta^0_{l,t,\mathcal{K}}\Bigg).$$

For a group $g \in \{1,\dots,G\}$, we have that $\Delta\beta^0_{g,t,\mathcal{K}} = \beta^0_{g,j+1} - \beta^0_{gj}$ for $k_{gj} < t \le T^0_j$ and that $\Delta\beta^0_{g,t,\mathcal{K}} = -(\beta^0_{g,j+1} - \beta^0_{gj})$ for $T^0_j < t \le k_{gj}$. It follows that

$$\sum_{t=\underline{k}_j+1}^{\overline{k}_j}(\Delta\beta^0_{g,t,\mathcal{K}})'X_{tT}(\Sigma^0_{t,\mathcal{K}})^{-1}u_t = -\mathrm{sgn}(r_{gj})\delta_{gj}'W_{T,j}(r_{gj}, r_{Gj}).$$

Similarly, for groups $g, h \in \{1,\dots,G\}$, we have that

$$\begin{aligned}&\sum_{t=\underline{k}_j+1}^{\overline{k}_j}(\Delta\beta^0_{g,t,\mathcal{K}})'X_{tT}(\Sigma^0_{t,\mathcal{K}})^{-1}X_{tT}'\Delta\beta^0_{h,t,\mathcal{K}} \\ &\quad = \mathbb{1}_{\{k_{gj}\vee k_{hj}\le T^0_j\}}\delta_{gj}'B_{T,j}(r_{gj}\vee r_{hj}, r_{Gj})\delta_{hj} + \mathbb{1}_{\{T^0_j<k_{gj}\wedge k_{hj}\}}\delta_{gj}'B_{T,j}(r_{gj}\wedge r_{hj}, r_{Gj})\delta_{hj}.\end{aligned}$$

Thus, we have

$$\begin{aligned}\ell^{(j)}_{T,2}(\boldsymbol{r}_j) \;&=\; -\sum_{g=1}^G\mathrm{sgn}(r_{gj})\delta_{gj}'W_{T,j}(r_{gj}, r_{Gj}) \\ &\quad -\frac{1}{2}\sum_{g=1}^G\sum_{l=1}^G\delta_{gj}'\Big\{\mathbb{1}_{\{r_{gj}\vee r_{lg}\le 0\}}B_{T,j}\big(r_{gj}\vee r_{lj}, r_{Gj}\big) + \mathbb{1}_{\{0<r_{gj}\wedge r_{lg}\}}B_{T,j}\big(r_{gj}\wedge r_{lj}, r_{Gj}\big)\Big\}\delta_{lj}.\end{aligned}$$

Applying Lemma 1 with (A.33) and the above equation, we can obtain

$$\big(\ell^{(1)}_T(\boldsymbol{r}_1),\dots,\ell^{(m)}_T(\boldsymbol{r}_m)\big) \Rightarrow \big(\ell^{(1)}_\infty(\boldsymbol{r}_1),\dots,\ell^{(m)}_\infty(\boldsymbol{r}_m)\big),$$

where, for $j = 1, \ldots, m$,

$$\ell_\infty^{(j)}(\boldsymbol{r}_j) \;:=\; \frac{1}{2}\mathrm{tr}\Big(\pi_j(r_{Gj})\mathbb{V}_{\eta\eta,j}(r_G)\Big) + \frac{|r_{Gj}|}{4}\mathrm{tr}\Big(\big\{\pi_j(r_{Gj})\big\}^2\Big) - \sum_{g=1}^{G}\mathrm{sgn}(r_{gj})\delta_{gj}'\mathbb{W}_j(r_{gj}, r_{Gj})$$
$$-\frac{1}{2}\sum_{g=1}^{G}\sum_{h=1}^{G}\delta_{gj}'\Big\{\mathbb{1}_{\{r_{gj}\vee r_{hg}\le 0\}}\mathbb{B}_j\big(r_{gj}\vee r_{hj}, r_{Gj}\big) + \mathbb{1}_{\{0<r_{gj}\wedge r_{hg}\}}\mathbb{B}_j\big(r_{gj}\wedge r_{hj}, r_{Gj}\big)\Big\}\delta_{hj}.$$

Applying a change of variables with $\boldsymbol{s}_j := \big(\|\delta_j\|^2 + \mathrm{tr}(\Phi_j^2)\big)\boldsymbol{r}_j$ with $\boldsymbol{s}_j = (s_1, \ldots, s_m)'$ for $j = 1, \ldots, m$, we can show that $2\ell_\infty^{(j)}(\boldsymbol{r}_j) = CB_\infty^{(j)}(\boldsymbol{s}_j)$ for all $j = 1, \ldots, m$. Thus, the continuous mapping theorem leads to the desired result. ■

Proof of Theorem 4. Under both alternatives H_1 and H_{1T}, the convergence rates of Theorem 1 apply to the estimates $\hat{\theta}$ and $\hat{\mathcal{K}}$. Thus, given collections of break dates $\hat{\mathcal{K}}$ and $\mathcal{K}^0$, the sub-intervals $\{[\tau_{g,l-1}+1, \tau_{gl}]\}_{l=1}^{N_g}$ for each group g satisfy either $\tau_{gl} - \tau_{g,l-1} \ge \nu T$ or $\tau_{gl} - \tau_{g,l-1} \le Mv_T^{-2}$ for some $M > 0$. If $\tau_{gl} - \tau_{g,l-1} \ge \nu T$, then the arguments used to prove Property 3(b) with $\sqrt{T}$-consistent estimate $\hat{\theta}$ show that $\bar{\ell}_{g,l}(\hat{\mathcal{K}}, \hat{\theta}) = O_p(1)$, while the arguments to obtain (A.26) show that $\bar{\ell}_{g,l}(\hat{\mathcal{K}}, \hat{\theta}) = O_p(1)$ if $\tau_{gl} - \tau_{gl} \le Mv_T^{-2}$. Also, Theorem 1(b) implies that $\Delta(\hat{\mathcal{K}}, \hat{\theta}) = o_p(1)$. It follows from Lemma A.3 that

$$\ell_T(\hat{\mathcal{K}}, \hat{\theta}) = O_p(1). \tag{A.34}$$

It remains to consider the normalized likelihood $\ell_T(\widetilde{\mathcal{K}}, \widetilde{\theta})$ under the null hypothesis H_0.

(a) Let $\delta \in (0,1)$ be fixed. If $\max_{1\le j\le m}\max_{1\le g_1,g_2\le G}|k_{g_1j}^0 - k_{g_2j}^0| \ge \delta T$, then we have $\max_{1\le j\le m}\max_{1\le g\le G}|\widetilde{k}_j - k_{gj}^0| \ge \delta T/2$. Applying a similar argument used in Proposition A.1, we can show that Properties 1 and 2 with $m_T = \delta T/2$ imply that

$$\ell_T(\widetilde{\mathcal{K}}, \widetilde{\theta}) \le -|O_p(Tv_T^2)|. \tag{A.35}$$

It follows from (A.34) and (A.35) that $CB_T = 2\{\ell_T(\hat{\mathcal{K}}, \hat{\theta}) - \ell_T(\widetilde{\mathcal{K}}, \widetilde{\theta})\} \ge |O_p(Tv_T^2)|$. Since the critical value c_α^* is a finite value, we obtain the desired result.

(b) If $\max_{1\le j\le m}\max_{1\le g_1,g_2\le G}|k_{g_1j}^0 - k_{g_2j}^0| \ge Mv_T^{-2}$ for some constant $M > 0$, then we have $\max_{1\le j\le m}\max_{1\le g\le G}|\widetilde{k}_j - k_{gj}^0| \ge Mv_T^{-2}/2$. When $\max_{1\le j\le m}\max_{1\le g\le G}|\widetilde{k}_j - k_{gj}^0| \ge Dv_T^{-2}\log T$ for a sufficiently large D, it was shown that $\ell_T(\widetilde{\mathcal{K}}, \widetilde{\theta}) \le -|O_p(M)|$ in the proof of Proposition A.1. When $Mv_T^{-2} \le \max_{1\le j\le m}\max_{1\le g\le G}|\widetilde{k}_j - k_{gj}^0| \le Dv_T^{-2}\log T$, it follows from the proof of Theorem 1(a) that $\ell_T(\widetilde{\mathcal{K}}, \widetilde{\theta}) \le -|O_p(M)|$ for a sufficiently large $M > 0$. Thus, there is some $M > 0$ such that $CB_T \ge |O_p(M)|$ and the proof is completed. ■

Table 1. Empirical Rejection Frequencies under the Null Hypotheses

		AR Coefficient								
		$\alpha = 0.0$			$\alpha = 0.4$			$\alpha = 0.8$		
Break Size		Nominal Size			Nominal Size			Nominal Size		
δ_1	δ_2	10%	5%	1%	10%	5%	1%	10%	5%	1%
0.50	0.50	0.064	0.036	0.004	0.086	0.050	0.004	0.162	0.104	0.032
	0.75	0.070	0.036	0.004	0.094	0.054	0.006	0.158	0.088	0.032
	1.00	0.084	0.036	0.004	0.106	0.060	0.010	0.170	0.098	0.038
	1.25	0.086	0.044	0.004	0.108	0.058	0.014	0.182	0.104	0.040
	1.50	0.096	0.050	0.006	0.120	0.056	0.010	0.186	0.108	0.036
0.75	0.75	0.084	0.032	0.004	0.112	0.046	0.004	0.158	0.086	0.030
	1.00	0.088	0.040	0.004	0.108	0.050	0.010	0.154	0.082	0.030
	1.25	0.086	0.050	0.006	0.104	0.060	0.006	0.156	0.088	0.028
	1.50	0.090	0.052	0.006	0.118	0.058	0.010	0.166	0.090	0.028
1.00	1.00	0.090	0.044	0.008	0.104	0.060	0.012	0.150	0.078	0.022
	1.25	0.086	0.050	0.010	0.090	0.060	0.010	0.140	0.072	0.026
	1.50	0.092	0.050	0.012	0.096	0.056	0.012	0.152	0.070	0.026
1.25	1.25	0.080	0.044	0.008	0.084	0.052	0.012	0.118	0.058	0.018
	1.50	0.074	0.042	0.010	0.080	0.044	0.010	0.112	0.056	0.018
1.50	1.50	0.074	0.038	0.010	0.088	0.040	0.010	0.106	0.048	0.018

Notes: The data generating process is the bivariate system:

$$y_{1t} = 1 + \delta_1 \mathbb{1}_{\{k_1<t\}} + \alpha y_{1,t-1} + u_{1t} \quad \text{(EQ1)}$$
$$y_{2t} = 1 + \delta_2 \mathbb{1}_{\{k_2<t\}} + \alpha y_{2,t-1} + u_{2t}, \quad \text{(EQ2)}$$

for $t = 1, \dots, T$, where $(u_{1t}, u_{2t})' \sim i.i.d. N(0, I_2)$ and δ_i is the break size for the i^{th} equation for $i = 1, 2$. We set the sample size $T = 100$, the break date $k_1 = k_2 = 50$ and the trimming value $\nu = 0.15$.

Table 2. Empirical Rejection Frequencies under the Null and Alternative Hypotheses
(the significance level: 5%)

				(1)	(2)	(3)	(4)	(5)	(6)
				Break dates (k_1, k_2)					
				(50,50)		(35, 35)		(35, 50)	
	AR	Break Size		Trimming value		Trimming value		Trimming value	
Correlation	α	δ_1	δ_2	0.15	0.10	0.15	0.10	0.15	0.10
0.0	0.0	0.5	0.5	0.024	0.030	0.018	0.030	0.05	0.06
			1.0	0.030	0.034	0.026	0.038	0.154	0.166
			1.5	0.036	0.038	0.034	0.048	0.226	0.228
		1.0	1.0	0.032	0.034	0.048	0.028	0.550	0.554
			1.5	0.036	0.038	0.022	0.022	0.728	0.730
		1.5	1.5	0.034	0.034	0.012	0.012	0.932	0.932
	0.4	0.5	0.5	0.036	0.044	0.026	0.040	0.064	0.080
			1.0	0.038	0.050	0.040	0.056	0.182	0.188
			1.5	0.048	0.056	0.036	0.050	0.250	0.300
		1.0	1.0	0.044	0.044	0.054	0.036	0.586	0.569
			1.5	0.048	0.048	0.062	0.032	0.732	0.734
		1.5	1.5	0.036	0.036	0.018	0.018	0.934	0.945
	0.8	0.5	0.5	0.082	0.092	0.096	0.102	0.172	0.215
			1.0	0.078	0.084	0.100	0.104	0.300	0.390
			1.5	0.090	0.104	0.178	0.096	0.370	0.445
		1.0	1.0	0.068	0.068	0.080	0.082	0.668	0.710
			1.5	0.056	0.056	0.056	0.056	0.774	0.805
		1.5	1.5	0.044	0.044	0.032	0.032	0.942	0.955
0.5	0.0	0.5	0.5	0.018	0.022	0.020	0.026	0.106	0.106
			1.0	0.028	0.034	0.038	0.038	0.256	0.248
			1.5	0.038	0.038	0.040	0.046	0.300	0.298
		1.0	1.0	0.028	0.028	0.026	0.028	0.730	0.730
			1.5	0.036	0.036	0.030	0.030	0.826	0.828
		1.5	1.5	0.020	0.020	0.020	0.020	0.978	0.978
	0.4	0.5	0.5	0.022	0.034	0.030	0.038	0.130	0.138
			1.0	0.044	0.044	0.032	0.038	0.262	0.268
			1.5	0.044	0.046	0.048	0.052	0.318	0.324
		1.0	1.0	0.038	0.038	0.036	0.042	0.752	0.752
			1.5	0.036	0.036	0.034	0.034	0.832	0.834
		1.5	1.5	0.022	0.022	0.022	0.022	0.978	0.978
	0.8	0.5	0.5	0.060	0.070	0.074	0.082	0.214	0.214
			1.0	0.068	0.070	0.076	0.084	0.362	0.364
			1.5	0.062	0.064	0.068	0.074	0.396	0.400
		1.0	1.0	0.046	0.046	0.052	0.056	0.778	0.776
			1.5	0.044	0.044	0.042	0.044	0.838	0.838
		1.5	1.5	0.026	0.026	0.026	0.026	0.978	0.978

Notes: The data generating process is the bivariate system as in (EQ1) and (EQ2) of Table 1 and standard normal errors $(u_{1t}, u_{2t})'$ are either uncorrelated or correlated with $cov(u_{1t}, u_{2t}) = 0.5$. The number of observations T is set to 100. Columns (1)-(4) report empirical size at a 5% nominal level and Columns (5)-(6) show empirical power given break dates $(k_1, k_2) = (35, 50)$ and critical values at a 5% significance level. The AR coefficient α is set to 0.0, 0.4 and 0.8. We use 0.5, 1.0 and 1.5 as magnitude of the break sizes.

Table 3. Structural breaks in the U.S. disaggregated inflation series

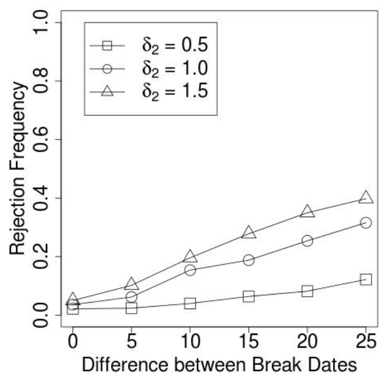
Replication of the results in Clark (2006)			
	OLS without breaks		
	Durables	Nondurables	Service
Persistency	0.921	0.878	0.855
	OLS with common break		
	Durables	Nondurables	Service
Persistency	0.800	0.367	0.137
Break Date (Known)	93:Q1		
Evidence from SUR system			
	SUR with common breaks ($k_1 = k_2 = k_3$)		
	Durables	Nondurables	Service
Persistency	0.805	0.356	0.166
Break Date	92:Q1		
Test for common break			
Null Hypothesis	LR test	Critical value (5%)	
$H_0 : k_1 = k_2 = k_3$	9.015	5.242	
$H_0 : k_1 = k_2$	9.735	3.473	
$H_0 : k_1 = k_3$	7.684	3.259	
$H_0 : k_2 = k_3$	0.749	2.501	
	SUR with common break ($k_2 = k_3$)		
	Durables	Nondurables	Service
Persistency	0.324	0.406	0.153
Break Date	95:Q1	92:Q1	
95% C.I.	[94:Q2, 95:Q4]	[91:Q3, 92:Q3]	

Notes: The sample period is 1984 to 2002. The estimated model is the AR model with the intercept and the AR lag length selected by the AIC is 4, 5 or 3 for durables, nondurables or service, respectively. *Persistency* is measured by the sum of AR coefficients. The critical values at the 5% significance level are obtained through a computationally efficient algorithm with 3,000 repetitions. C.I. denotes the 95% confidence interval of the break date.

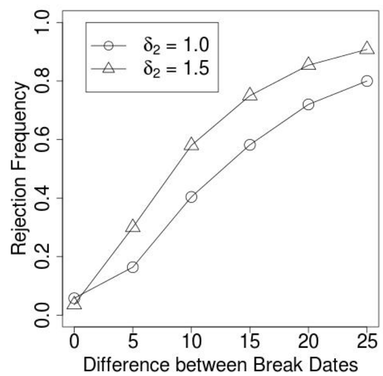
Figure 1: Finite-sample power of the test

Panel A: AR Coefficient = 0.00

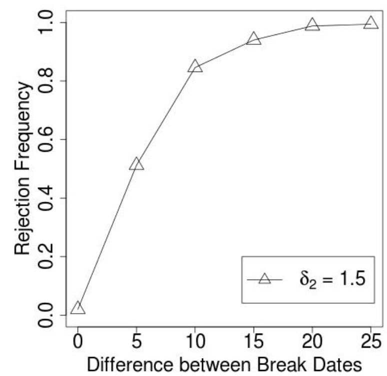
(a) Break Size in EQ1: 0.5



(b) Break Size in EQ1: 1.0

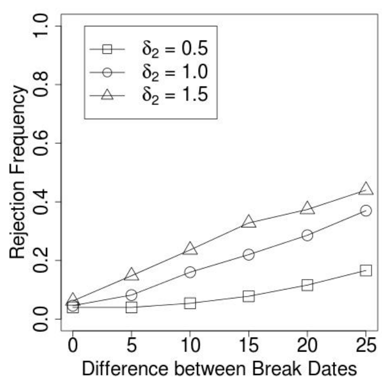


(c) Break Size in EQ1: 1.5

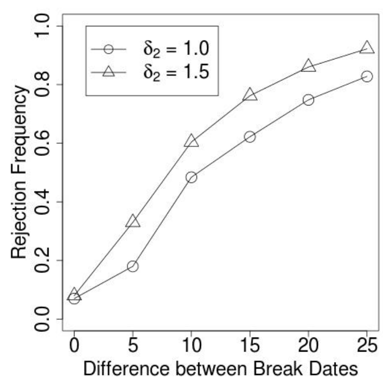


Panel B: AR Coefficient = 0.40

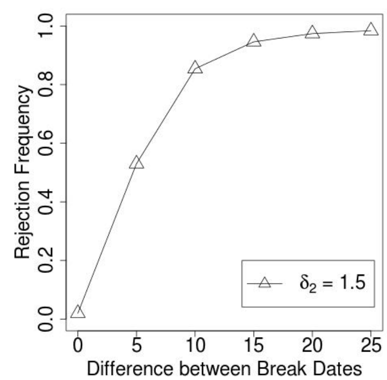
(d) Break Size in EQ1: 0.5



(e) Break Size in EQ1: 1.0

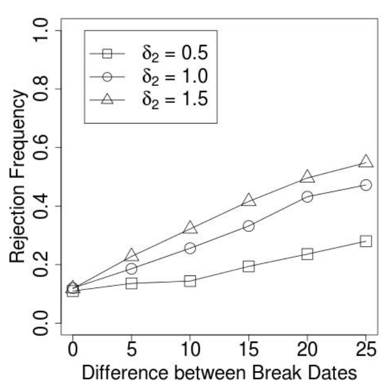


(f) Break Size in EQ1: 1.5

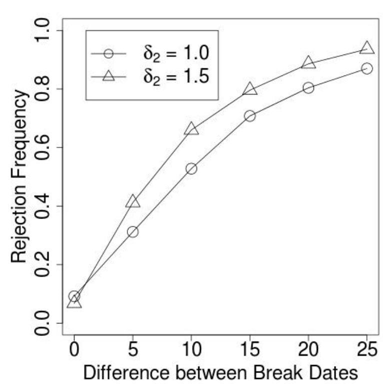


Panel C: AR Coefficient = 0.80

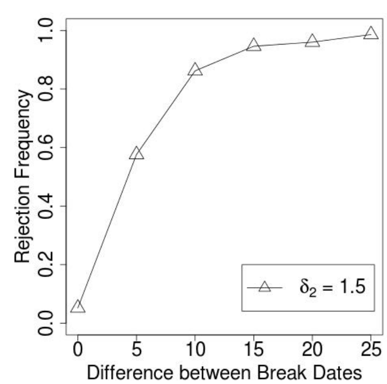
(g) Break Size in EQ1: 0.5



(h) Break Size in EQ1: 1.0



(i) Break Size in EQ1: 1.5



Notes: The data generating process is the bivariate system as in (EQ1) and (EQ2) of Table 1. The number of observations T is set to 100. The break date k_1 in (EQ1) is kept fixed at $k_1 = 35$, while the break date k_2 in (EQ2) changes from 30 to 55. The horizontal axis shows the difference between break dates: $k_2 - k_1$. The AR coefficient α is set to 0.0, 0.4 and 0.8 for Panel A, B and C, respectively. The break size δ_1 in (EQ1) changes across panel (a)-(c), (d)-(f) and (g)-(i), while the break size δ_2 in (EQ2) changes within each panel. We use 0.5, 1.0 and 1.5 as magnitude of the break size.

Supplemental Material for
"Testing for Common Breaks in a Multiple Equations System"

Tatsushi Oka **Pierre Perron**

May 28, 2017

1 Introduction

In this supplemental material, we derive a limit distribution of a test for common breaks in two examples of our main text. We use the same notations as in the main text. Our quasi-likelihood ratio test for common breaks is written as a difference of two normalized log likelihoods with and without a common breaks restriction. Define a normalized log likelihood as

$$\ell_T(\mathcal{K},\theta) := \log L_T(\mathcal{K},\theta) - \log L_T(\mathcal{T}^0,\theta^0).$$

Then, a quasi-likelihood ratio test for common breaks is given by

$$CB_T = 2\{\ell_T(\hat{\mathcal{K}},\hat{\theta}) - \ell_T(\widetilde{\mathcal{K}},\widetilde{\theta})\},$$

where $\hat{\mathcal{K}}$ and $\hat{\theta}$ are estimates without a common breaks restriction and $\widetilde{\mathcal{K}}$ and $\widetilde{\theta}$ are estimates with the restriction. Under the null hypothesis of common breaks, Theorems 1 and 2 yield

$$CB_T = 2\Big\{\sup_{\mathcal{K}\in\bar{\Xi}_M} \ell_T(\mathcal{K},\theta^0) - \sup_{\mathcal{K}\in\bar{\Xi}_{M,H_0}} \ell_T(\mathcal{K},\theta^0)\Big\} + o_p(1).$$

To obtain the limit distribution of the common breaks test, it suffices to derive the limit distribution of the normalized likelihood $\ell_T(\mathcal{K},\theta^0)$ evaluated at the true basic parameters over a restricted break date space $\bar{\Xi}_M$, because the desired result will follow from a simple application of the continuous mapping theorem.

2 Example 1 (changes in intercepts)

Under the null hypothesis, the true model is a two-equations system of autoregressive (AR) model with structural changes in intercepts:

$$\begin{aligned} y_{1t} &= \mu_{1j}^0 + \alpha_1^0 y_{1,t-1} + u_{1t}, \\ y_{2t} &= \mu_{2j}^0 + \alpha_2^0 y_{2,t-1} + u_{2t}, \end{aligned}$$

for $T_{j-1}^0 + 1 \le t \le T_j^0$ with $j = 1, 2$, where $(u_{1t}, u_{2t})'$ have mean zeros and a covariance matrix Σ^0. In the above model, basic parameters other than intercepts are assumed to be constant and intercepts change at a common break date T_1^0. Suppose that we observe a data set $\{(y_{1t}, y_{2t})\}_{t=0}^T$, generated by the model above.

Allowing the possibility that break dates can differ across two equations, we consider the following estimation model, for $j = 1, 2$,

$$\begin{aligned} y_{1t} &= \mu_{1j} + \alpha_1 y_{1,t-1} + u_{1t}, \quad k_{1,j-1} + 1 \le t \le k_{1j}, \\ y_{2t} &= \mu_{2j} + \alpha_2 y_{2,t-1} + u_{2t}, \quad k_{2,j-1} + 1 \le t \le k_{2j}. \end{aligned}$$

We consider the issue of testing for the null and alternative hypotheses:

$$H_0 : k_{11} = k_{21} \quad \text{and} \quad H_1 : k_{11} \neq k_{21}.$$

Alternatively, the estimation model above can be written as

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 1 & y_{1,t-1} & 0 & 0 \\ 0 & 0 & 1 & y_{2,t-1} \end{bmatrix} \begin{bmatrix} \mu_{1,1+\mathbb{1}_{\{k_{11}<t\}}} \\ \alpha_1 \\ \mu_{2,1+\mathbb{1}_{\{k_{21}<t\}}} \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix},$$

or

$$y_t = X_t' \beta_{t,\mathcal{K}} + u_t,$$

where $\beta_{t,\mathcal{K}} = \sum_{g=1}^2 \beta_{g,t,\mathcal{K}}$ with

$$\begin{aligned} \beta_{1,t,\mathcal{K}} &= (\mu_{1j}, \alpha_1, 0, 0)', \quad k_{1,j-1} + 1 \le t \le k_{1j}, \\ \beta_{2,t,\mathcal{K}} &= (0, 0, \mu_{2j}, \alpha_2)', \quad k_{2,j-1} + 1 \le t \le k_{12j}, \end{aligned}$$

for $j = 1, 2$. For break dates $\mathcal{K} \in \bar{\Xi}_M$, the log likelihood function evaluated at the true basic parameter is given by

$$\log L_T(\mathcal{K}, \theta^0) = -\frac{T}{2} \log(2\pi)^2 - \frac{T}{2} \log|\Sigma^0| - \frac{1}{2} \sum_{t=1}^T \|(\Sigma^0)^{-1/2}(y_t - X_t' \beta_{t,\mathcal{K}}^0)\|^2,$$

which leads to the following normalized log likelihood:

$$\ell_T(\mathcal{K}, \theta^0) = -\frac{1}{2} \sum_{t=1}^T \left\{ \|(\Sigma^0)^{-1/2}(y_t - X_t' \beta_{t,\mathcal{K}}^0)\|^2 - \|(\Sigma^0)^{-1/2} u_t\|^2 \right\}.$$

Letting $\Delta\beta_{t,\mathcal{K}}^0 = \beta_{t,\mathcal{K}}^0 - \beta_{t,\mathcal{T}^0}^0$, we have $y_t - X_t'\beta_{t,\mathcal{K}}^0 = u_t - X_t'\Delta\beta_{t,\mathcal{K}}^0$ and also $\Delta\beta_{t,\mathcal{K}}^0 = 0$ for $t \notin [\underline{k}_1 + 1, \bar{k}_1]$ with $\underline{k}_1 = \min\{k_{11}, k_{21}, T_1^0\}$ and $\bar{k}_1 = \max\{k_{11}, k_{21}, T_1^0\}$. Thus, we have

$$\ell_T(\mathcal{K}, \theta^0) = \sum_{t=\underline{k}_1+1}^{\bar{k}_1} (\Delta\beta_{t,\mathcal{K}}^0)' X_t (\Sigma^0)^{-1} u_t - \frac{1}{2} \sum_{t=\underline{k}_1+1}^{\bar{k}_1} \|(\Sigma^0)^{-1/2} X_t' \Delta\beta_{t,\mathcal{K}}^0\|^2.$$

For $\mathcal{K} \in \bar{\Xi}_M$, we can write $k_{g1} = T_1^0 + [r_g v_T^{-2}]$ for some $r_g \in [-M, M]$ and for $g = 1, 2$. Let $\boldsymbol{r} = (r_1, r_2)'$ and we define

$$\ell_T^{(1)}(\boldsymbol{r}) := \ell_T(\mathcal{K}, \theta^0).$$

We shall derive a limit distribution of $\ell_T^{(1)}(\boldsymbol{r})$ for two cases: (1) $r_1 \leq 0$ and $r_2 \leq 0$ and (2) $r_1 \leq 0$ and $0 < r_2$. A similar analysis can apply for the remaining cases. Under a shrinking framework, let $\delta^{(g)}$ denote a constant satisfying that $\mu_{g2} - \mu_{g1} = v_T \delta^{(g)}$ for $g = 1, 2$.

(1) Consider the case where $r_1 \leq 0$ and $r_2 \leq 0$. Then, we can write

$$\begin{aligned} \beta_{1,t,\mathcal{K}}^0 - \beta_{1,t,\mathcal{T}^0}^0 &= [\mu_{12}^0, \alpha_1^0, 0, 0]' - [\mu_{11}^0, \alpha_1^0, 0, 0]' = v_T \delta_{11}, \quad k_{11} + 1 \leq t \leq T_1^0 \\ \beta_{2,t,\mathcal{K}}^0 - \beta_{2,t,\mathcal{T}^0}^0 &= [0, 0, \mu_{22}^0, \alpha_2^0]' - [0, 0, \mu_{21}^0, \alpha_2^0]' = v_T \delta_{21}, \quad k_{21} + 1 \leq t \leq T_1^0, \end{aligned}$$

where $\delta_{11} := [\delta^{(1)}, 0, 0, 0]'$ and $\delta_{21} := [0, 0, v_T \delta^{(2)}, 0]'$. Thus, we have

$$\sum_{t=\underline{k}_1+1}^{\bar{k}_1} (\Delta\beta_{t,\mathcal{K}}^0)' X_t (\Sigma^0)^{-1} u_t = \sum_{g=1}^{2} v_T \delta_{g1}' (\Sigma^0)^{-1} \sum_{t=k_{g1}+1}^{T_1^0} u_t = \sum_{g=1}^{2} \delta_{g1}' (\Sigma^0)^{-1/2} V_{T,\eta}^{(1)}(-r_g),$$

where $V_{T,\eta}^{(1)}(-r) := v_T \sum_{t=T_1^0+[rv_T^{-2}]+1}^{T_1^0} \eta_t$ for $r < 0$. Also we have

$$\begin{aligned} \sum_{t=\underline{k}_1+1}^{\bar{k}_1} \|(\Sigma^0)^{-1/2} X_t' \Delta\beta_{t,\mathcal{K}}^0\|^2 &= \sum_{g=1}^{2}\sum_{h=1}^{2} \left| v_T^2 \big([r_g v_T^{-2}] \wedge [r_h v_T^{-2}] \big) \right| \delta_{g1}' (\Sigma^0)^{-1} \delta_{h1} \\ &= \sum_{g=1}^{2}\sum_{h=1}^{2} \big(|r_g| \wedge |r_h| \big) d_{gh} + o(1), \end{aligned}$$

where $d_{gh} := \delta_{g1}' (\Sigma^0)^{-1} \delta_{h1}$ for $g, h \in \{1, 2\}$. We have

$$\ell_T^{(1)}(\boldsymbol{r}) = \sum_{g=1}^{2} \delta_{g1}' (\Sigma^0)^{-1/2} V_{T,\eta}^{(1)}(-r_g) - \frac{1}{2} \sum_{g=1}^{2}\sum_{h=1}^{2} \big(|r_g| \wedge |r_h| \big) d_{gh} + o(1).$$

Using the FCLT, we have, uniformly in $\boldsymbol{r} \in [-M, 0] \times [-M, 0]$,

$$\ell_T^{(1)}(\boldsymbol{r}) \Rightarrow \sum_{g=1}^{2} \delta_{g1}' (\Sigma^0)^{-1/2} \mathbb{V}_\eta(r_g) - \frac{1}{2} \sum_{g=1}^{2}\sum_{h=1}^{2} \big(|r_g| \wedge |r_h| \big) d_{gh}.$$

(2) Consider the case where $r_1 \leq 0$ and $0 < r_2$. Then, we can write, for $k_{11} + 1 \leq t \leq T_1^0$,

$$\beta_{1,t,\mathcal{K}} - \beta_{1,t,\mathcal{T}^0} = v_T \delta_{11}$$

and, for $T_1^0 + 1 \leq t \leq k_{21}$,

$$\beta_{2,t,\mathcal{K}} - \beta_{2,t,\mathcal{T}^0} = [0, 0, \mu_{21}^0, \alpha_2^0]' - [0, 0, \mu_{22}^0, \alpha_2^0]' = -v_T \delta_{21}.$$

We have

$$X_t'(\beta_{1,t,\mathcal{K}} - \beta_{1,t,\mathcal{T}^0}) = v_T \delta_{11} \quad \text{and} \quad X_t'(\beta_{2,t,\mathcal{K}} - \beta_{2,t,\mathcal{T}^0}) = -v_T \delta_{21},$$

for $k_{11} + 1 \leq t \leq T_1^0$ and $T_1^0 + 1 \leq t \leq k_{21}$, respectively. Thus, we have

$$\sum_{t=\underline{k}_1+1}^{\bar{k}_1} (\Delta\beta_{t,\mathcal{K}}^0)' X_t (\Sigma^0)^{-1} u_t = \delta_{11}' (\Sigma^0)^{-1/2} V_{T,\eta}^{(1)}(-r_1) - \delta_{21}' (\Sigma^0)^{-1/2} V_{T,\eta}^{(2)}(r_2),$$

where $V^{(2)}_{T,\eta}(r) := v_T \sum_{t=T_1^0+1}^{T_1^0+[rv_T^{-2}]} \eta_t$ for $r > 0$. It follows from the FCLT that

$$\begin{aligned}\sum_{t=\underline{k}_1+1}^{\bar{k}_1} (\Delta\beta^0_{t,\mathcal{K}})' X_t (\Sigma^0)^{-1} u_t &\Rightarrow \delta'_{11}(\Sigma^0)^{-1/2}\mathbb{V}_\eta(r_1) - \delta'_{21}(\Sigma^0)^{-1/2}\mathbb{V}_\eta(r_2) \\ &= -\sum_{g=1}^{2} \operatorname{sgn}(r_g)\delta'_{g1}(\Sigma^0)^{-1/2}\mathbb{V}_\eta(r_g).\end{aligned}$$

Also, we have

$$X_t'\Delta\beta^0_{t,\mathcal{K}} = \begin{cases} v_T\delta_{11}, & \text{if } k_{11} < t \le T_1^0 \\ -v_T\delta_{21}, & \text{if } T_1^0 < t \le k_{21}. \end{cases}$$

Thus, we can show that

$$\begin{aligned}\sum_{t=\underline{k}_1+1}^{\bar{k}_1} \|(\Sigma^0)^{-1/2} X_t'\Delta\beta^0_{t,\mathcal{K}}\|^2 &= \left|v_T^2[r_1 v_T^{-2}]\right|\delta'_{11}(\Sigma^0)^{-1}\delta_{11} + \left|v_T^2[r_2 v_T^{-2}]\right|\delta_{21}(\Sigma^0)^{-1}\delta_{21} \\ &= |r_1|d_{11} + |r_2|d_{22}.\end{aligned}$$

Thus we have

$$\ell^{(1)}_T(\boldsymbol{r}) \Rightarrow -\sum_{g=1}^{2}\operatorname{sgn}(r_g)\delta'_{g1}(\Sigma^0)^{-1/2}\mathbb{V}_\eta(r_g) - \frac{1}{2}\sum_{g=1}^{2}|r_g|d_{gg}.$$

A similar argument used above can apply to the remaining cases. The results obtained above shows that when r_1 and r_2 take the same sign, the drift term has an interaction term of $\delta^{(1)}$ and $\delta^{(2)}$, while there is no such interaction term when signs of r_1 and r_2 are different. Thus we can obtain

$$\ell^{(1)}_T(\boldsymbol{r}) \Rightarrow \ell^{(1)}_\infty(\boldsymbol{r}),$$

where

$$\begin{aligned}\ell^{(1)}_\infty(\boldsymbol{r}) &= -\sum_{g=1}^{2}\operatorname{sgn}(r_g)\delta'_{g1}(\Sigma^0)^{-1/2}\mathbb{V}_\eta(r_g) \\ &\quad -\frac{1}{2}\sum_{g=1}^{2}\sum_{h=1}^{2}\mathbb{1}_{\{\operatorname{sgn}(s_g)=\operatorname{sgn}(s_h)\}}\big(|r_g| \wedge |r_h|\big)d_{gh}.\end{aligned}$$

This together with the continuous mapping theorem yields

$$CB_T \Rightarrow 2\Big\{\sup_{\boldsymbol{r}}\ell^{(1)}_\infty(\boldsymbol{r}) - \sup_{\boldsymbol{r}:r_1=r_2}\ell^{(1)}_\infty(\boldsymbol{r})\Big\}.$$

The limiting distribution depends on nuisance parameters. We replace the nuisance parameters by their estimates. For this end, we first rewrite the limiting distribution using a change of variable. Let $\boldsymbol{s} := \|\delta\|^2 \boldsymbol{r}$ with $\delta = (\delta^{(1)}, \delta^{(2)})'$ and $\boldsymbol{s} = (s_1, s_2)'$. Then, an application of a change of variables yields $2\ell_\infty^{(1)}(\boldsymbol{r}) = CB_\infty^{(1)}(\boldsymbol{s})$, where

$$\begin{aligned} CB_\infty^{(1)}(\boldsymbol{s}) &= -\frac{1}{\|\delta\|} 2 \sum_{g=1}^{2} \operatorname{sgn}(s_g) \delta_g (\Sigma^0)^{-1/2} \mathbb{V}_\eta(s_g) \\ &\quad - \frac{1}{\|\delta\|} \sum_{g=1}^{2} \sum_{h=1}^{2} \mathbb{1}_{\{\operatorname{sgn}(s_g) = \operatorname{sgn}(s_h)\}} \big(|s_g| \wedge |s_h|\big) d_{gh}. \end{aligned}$$

Therefore, we have

$$CB_T \Rightarrow \sup_{\boldsymbol{s}} CB_\infty^{(1)}(\boldsymbol{s}) - \sup_{\boldsymbol{s}: s_1 = s_2} CB_\infty^{(1)}(\boldsymbol{s}).$$

Here, we can have consistent estimators for nuisance parameters. To see this, let $\hat{\mu}_{gj}$ be a consistent estimator for μ_{gj} for $g, j \in \{1, 2\}$ and define $\Delta\hat{\mu} = [\Delta\hat{\mu}_1, \Delta\hat{\mu}_2]'$ with $\Delta\hat{\mu}_g = \hat{\mu}_{g2} - \hat{\mu}_{g1}$ for $g = 1, 2$. Also, let $\hat{\Sigma}$ be a consistent estimator Σ^0. Then we have, for $g, h \in \{1, 2\}$,

$$\frac{1}{\|\Delta\hat{\mu}\|} \Delta\hat{\mu}_g = \frac{1}{\|\delta\|} \delta^{(g)} + o_p(1) \quad \text{and} \quad \frac{1}{\|\Delta\hat{\mu}\|} \Delta\hat{\mu}_g' (\hat{\Sigma})^{-1} \Delta\hat{\mu}_g = \frac{1}{\|\delta\|} d_{gh} + o_p(1).$$

3 Example 2 (a single equation model)

We consider a single equation model:

$$y_{1t} = \mu^0 + \alpha_j^0 z_{1t} + \gamma_j^0 (t/T) + \rho_j^0 T^{-1/2} w_{1t} + u_{1t},$$

for $T_{j-1}^0 + 1 \le t \le T_j^0$ with $j = 1, 2, 3$, where u_{1t} denotes the error term with $E[u_{1t}] = 0$ and $E[u_{1t}^2] = (\sigma_j^0)^2$. In this example, basic parameters other than an intercept have two structural changes at the true break dates T_1^0 and T_2^0. Suppose that we observe data $\{(y_{1t}, z_{1t}, w_{1t})\}_{t=1}^T$ generated by the model above.

We consider a test of common breaks against the alternative that all coefficients change at distinct break dates, while the coefficient δ_j and the variance σ_j^2 change at the same break dates. More specifically, an estimation model with distinct break dates among basic parameters is given by

$$y_{1t} = \mu + \alpha_{t,\mathcal{K}} z_{1t} + \gamma_{t,\mathcal{K}} (t/T) + \rho_{t,\mathcal{K}} T^{-1/2} w_{1t} + u_{1t}, \quad E[u_{1t}^2] = \sigma_{t,\mathcal{K}}^2$$

where $\alpha_{t,\mathcal{K}} := \alpha_j$ for $k_{1,j-1} + 1 \le t \le k_{1j}$, $\gamma_{t,\mathcal{K}} := \gamma_j$ for $k_{2,j-1} + 1 \le t \le k_{2j}$ and $(\rho_{t,\mathcal{K}}, \sigma_{t,\mathcal{K}}^2) := (\rho_j, \sigma_j^2)$ for $k_{3,j-1} + 1 \le t \le k_{3j}$ with $j = 1, 2, 3$. This model can be written as

$$y_{1t} = x_{1t}' \beta_{t,\mathcal{K}} + u_t,$$

where

$$x_{1t} := (1, z_{1t}, t/T, T^{-1/2} w_{1t})' \quad \text{and} \quad \beta_{t,\mathcal{K}} := (\mu, \alpha_{t,\mathcal{K}}, \gamma_{t,\mathcal{K}}, \rho_{t,\mathcal{K}})'.$$

The log likelihood function evaluated at the true basic parameters is given by

$$\log L_T(\mathcal{K},\theta^0) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left\{\log(\sigma^0_{t,\mathcal{K}})^2 - \frac{(y_{1t}-x_{1t}'\beta^0_{t,\mathcal{K}})^2}{(\sigma^0_{t,\mathcal{K}})^2}\right\}.$$

Thus, for $\mathcal{K}\in\bar{\Xi}_M$, a normalized log likelihood is given by

$$\ell_T(\mathcal{K},\theta^0) = \frac{1}{2}\sum_{j=1}^{2}\sum_{t=\underline{k}_j+1}^{\overline{k}_j}\left\{\log\left(\frac{\sigma^0_{t,\mathcal{T}^0}}{\sigma^0_{t,\mathcal{K}}}\right)^2 - \frac{(y_{1t}-x_{1t}'\beta^0_{t,\mathcal{K}})^2}{(\sigma^0_{t,\mathcal{K}})^2} + \frac{u_{1t}^2}{(\sigma^0_{t,\mathcal{T}^0})^2}\right\},$$

where $\underline{k}_j = \min\{k_{1j},k_{2j},k_{3j}T_j^0\}$ and $\overline{k}_j = \max\{k_{1j},k_{2j},k_{3j},T_j^0\}$. Let $\underline{k}_{gj} := \min\{k_{gj},T_j^0\}$ and $\overline{k}_{3j} := \max\{k_{gj},T_j^0\}$. Furthermore, we have $k_{gj} = T_j^0 + [v_T^{-2}r_{gj}]$ with $r_{gj}\in[-M,M]$ for all $g=1,2,3$ and $j=1,2$. Letting $\boldsymbol{r}_j := (r_{1j},r_{2j},r_{3j})'$ with $j=1,2$, we can write

$$\ell_T(\mathcal{K},\theta^0) = \sum_{j=1}^{2}\ell_T^{(j)}(\boldsymbol{r}_j),$$

where

$$\ell_T^{(j)}(\boldsymbol{r}_j) = \frac{1}{2}\sum_{t=\underline{k}_j+1}^{\overline{k}_j}\left\{\log\left(\frac{\sigma^0_{t,\mathcal{T}^0}}{\sigma^0_{t,\mathcal{K}}}\right)^2 - \frac{(y_{1t}-x_{1t}'\beta^0_{t,\mathcal{K}})^2}{(\sigma^0_{t,\mathcal{K}})^2} + \frac{u_{1t}^2}{(\sigma^0_{t,\mathcal{T}^0})^2}\right\}.$$

We have that $y_{1t}-x_{1t}'\beta^0_{t,\mathcal{K}} = u_{1t}-x_{1t}'\Delta\beta^0_{t,\mathcal{K}}$, where $\Delta\beta^0_{t,\mathcal{K}} = \beta^0_{t,\mathcal{K}}-\beta^0_{t,\mathcal{T}^0}$. Thus, we have

$$\ell_T^{(j)}(\boldsymbol{r}_j) = \ell_{T,1}^{(j)}(\boldsymbol{r}_j) + \ell_{T,2}^{(j)}(\boldsymbol{r}_j),$$

where

$$\ell_{T,1}^{(j)}(\boldsymbol{r}_j) = \frac{1}{2}\sum_{t=\underline{k}_{3j}+1}^{\overline{k}_{3j}}\left\{\log\left(\frac{\sigma^0_{t,\mathcal{T}^0}}{\sigma^0_{t,\mathcal{K}}}\right)^2 + \left(1-\frac{(\sigma^0_{t,\mathcal{T}^0})^2}{(\sigma^0_{t,\mathcal{K}})^2}\right)\frac{u_{1t}^2}{(\sigma^0_{t,\mathcal{T}^0})^2}\right\} \tag{S.1}$$

$$\ell_{T,2}^{(j)}(\boldsymbol{r}_j) = \frac{1}{2}\sum_{t=\underline{k}_j+1}^{\overline{k}_j}\frac{2\Delta\beta^{0\prime}_{t,\mathcal{K}}x_{1t}u_{1t}-(x_{1t}'\Delta\beta^0_{t,\mathcal{K}})^2}{(\sigma^0_{t,\mathcal{K}})^2}. \tag{S.2}$$

Under a shrinking framework, we can write, for $j=1,2$

$$\left(\alpha^0_{j+1},\gamma^0_{j+1},\rho^0_{j+1}\right) - \left(\alpha^0_j,\gamma^0_j,\rho^0_j\right) = v_T\left(\delta_j^{(1)},\delta_j^{(2)},\delta_j^{(3)}\right) \quad\text{and}\quad (\sigma^0_{j+1})^2-(\sigma^0_j)^2 = v_T\phi_j,$$

for some $\delta_j^{(1)},\delta_j^{(2)},\delta_j^{(3)},\phi_j\in\mathbb{R}$. Then, we have, for $\underline{k}_j+1\le t\le\overline{k}_j$,

$$\left(\frac{\sigma^0_{t,\mathcal{T}^0}}{\sigma^0_{t,\mathcal{K}}}\right)^2 = 1-\frac{(\sigma^0_{t,\mathcal{K}})^2-(\sigma^0_{t,\mathcal{T}^0})^2}{(\sigma^0_{t,\mathcal{K}})^2} = 1-\frac{v_T\phi_{t,\mathcal{K}}}{(\sigma^0_{t,\mathcal{K}})^2},$$

where

$$\phi_{t,\mathcal{K}} := \begin{cases} \phi_j, & \text{if } k_{3j} < t \le T_j^0 \\ -\phi_j, & \text{if } T_j^0 < t \le k_{3j}. \end{cases}$$

Thus, an application of Taylor's expansion yields

$$\log\Big(\frac{\sigma^0_{t,\mathcal{T}^0}}{\sigma^0_{t,\mathcal{K}}}\Big)^2 = -\frac{\upsilon_T\phi_{t,\mathcal{K}}}{(\sigma^0_{t,\mathcal{K}})^2} + \frac{1}{2}\Big(\frac{\upsilon_T\phi_{t,\mathcal{K}}}{(\sigma^0_{t,\mathcal{K}})^2}\Big)^2 + O_p(\upsilon_T^3), \tag{S.3}$$

and also we have

$$\Big(1 - \frac{(\sigma^0_{t,\mathcal{T}^0})^2}{(\sigma^0_{t,\mathcal{K}})^2}\Big)\frac{u_{1t}^2}{(\sigma^0_{t,\mathcal{T}^0})^2} = \frac{\upsilon_T\phi_{t,\mathcal{K}}}{(\sigma^0_{t,\mathcal{K}})^2}\eta_{1t}^2, \tag{S.4}$$

where $\eta_{1t} = u_{1t}/\sigma^0_{t,\mathcal{T}^0}$. It follows from (S.1)-(S.4) that

$$\ell_{T,1}^{(j)}(\boldsymbol{r}_j) = \frac{1}{2}\upsilon_T\sum_{t=\underline{k}_{3j}+1}^{\overline{k}_{3j}}\frac{\phi_{t,\mathcal{K}}}{(\sigma^0_{t,\mathcal{K}})^2}\big(\eta_{1t}^2 - 1\big) + \frac{1}{4}\upsilon_T^2\sum_{t=\underline{k}_{3j}+1}^{\overline{k}_{3j}}\frac{\phi^2_{t,\mathcal{K}}}{(\sigma^0_{t,\mathcal{K}})^4} + o_p(1). \tag{S.5}$$

From (S.2) and (S.5), we can show that, when $r_{3j} < 0$,

$$\ell_{T,1}^{(j)}(\boldsymbol{r}_j) = \frac{\phi_j}{2(\sigma^0_{j+1})^2}\upsilon_T\sum_{t=k_{3j}+1}^{T_j^0}\big(\eta_{1t}^2 - 1\big) - \frac{\phi_j^2}{4(\sigma^0_{j+1})^4}r_{3j} + o_p(1) \tag{S.6}$$

$$\ell_{T,2}^{(j)}(\boldsymbol{r}_j) = \frac{1}{2(\sigma^0_{j+1})^2}\sum_{t=\underline{k}_j+1}^{\overline{k}_j}\big(2\Delta\beta^{0\prime}_{t,\mathcal{K}}x_{1t}u_{1t} - (x_{1t}'\Delta\beta^0_{t,\mathcal{K}})^2\big). \tag{S.7}$$

On the other hand, when $0 \le r_{3j}$, we have

$$\ell_{T,1}^{(j)}(\boldsymbol{r}_j) = -\frac{1}{2}\frac{\phi_j}{(\sigma^0_j)^2}\upsilon_T\sum_{t=T_j^0+1}^{k_{3j}}\big(\eta_{1t}^2 - 1\big) + \frac{\phi_j^2}{4(\sigma^0_j)^4}r_{3j} + o_p(1) \tag{S.8}$$

$$\ell_{T,2}^{(j)}(\boldsymbol{r}_j) = \frac{1}{2(\sigma^0_j)^2}\sum_{t=\underline{k}_j+1}^{\overline{k}_j}\big(2\Delta\beta^{0\prime}_{t,\mathcal{K}}x_{1t}u_{1t} - (x_{1t}'\Delta\beta^0_{t,\mathcal{K}})^2\big). \tag{S.9}$$

Let $V_{T,\eta\eta,j}(r_{3j}) := \upsilon_T\sum_{t=\underline{k}_{3j}+1}^{\overline{k}_{3j}}(\eta_{1t}^2 - 1)$. From (S.6) and (S.8), we have

$$\ell_{T,1}^{(j)}(\boldsymbol{r}_j) = -\text{sgn}(r_{3j})\frac{1}{2}\frac{\phi_j}{(\sigma^0_{j+\mathbb{1}_{\{r_{3j}<0\}}})^2}V_{T,\eta\eta}(r_{3j}) + \frac{\phi_j^2}{4(\sigma^0_{j+\mathbb{1}_{\{r_{3j}<0\}}})^4}|r_{3j}| + o_p(1). \tag{S.10}$$

Furthermore, for $j = 1, 2$, we can write $\Delta\beta^0_{t,\mathcal{K}} = \sum_{g=1}^3 \Delta\beta_{g,t,\mathcal{K}}$, where

$$\begin{aligned} \Delta\beta_{1,t,\mathcal{K}} &= -\text{sgn}(r_{1j})\upsilon_T\delta_{1j}, \text{ for } \underline{k}_{1j} < t \le \overline{k}_{1j}, \\ \Delta\beta_{2,t,\mathcal{K}} &= -\text{sgn}(r_{2j})\upsilon_T\delta_{2j}, \text{ for } \underline{k}_{2j} < t \le \overline{k}_{2j}, \\ \Delta\beta_{3,t,\mathcal{K}} &= -\text{sgn}(r_{3j})\upsilon_T\delta_{3j}, \text{ for } \underline{k}_{3j} < t \le \overline{k}_{3j}, \end{aligned}$$

with $\delta_{1j} = (0, \delta_j^{(1)}, 0, 0)'$, $\delta_{2j} = (0, 0, \delta_j^{(2)}, 0)'$ and $\delta_{3j} = (0, 0, 0, \delta_j^{(3)})'$, and $\Delta\beta_{g,t,\mathcal{K}} = (0, 0, 0, 0)$ if $t \notin (\underline{k}_{gj}, \overline{k}_{gj}]$ for each $g = 1, 2, 3$. We have

$$\sum_{t=\underline{k}_j+1}^{\overline{k}_j} \Delta\beta_{t,\mathcal{K}}^{0\prime} x_{1t}u_{1t} = \sum_{g=1}^{3} \sum_{t=\underline{k}_{gj}+1}^{\overline{k}_{gj}} \Delta\beta_{g,t,\mathcal{K}}^{0\prime} x_{1t}u_{1t} = -\sum_{g=1}^{3} \mathrm{sgn}(r_{gj}) v_T \delta_{gj}' \Big(v_T \sum_{t=\underline{k}_{gj}+1}^{\overline{k}_{gj}} x_{1t}u_{1t} \Big).$$

Also, we have

$$\sum_{t=\underline{k}_j+1}^{\overline{k}_j} (x_{1t}'\Delta\beta_{t,\mathcal{K}}^0)^2 = \sum_{t=\underline{k}_j+1}^{\overline{k}_j} \sum_{g=1}^{3} \sum_{h=1}^{3} \Delta\beta_{g,t,\mathcal{K}}^{0\prime} x_{1t}x_{1t}' \Delta\beta_{h,t,\mathcal{K}}^0,$$

and moreover

$$\sum_{t=\underline{k}_j+1}^{\overline{k}_j} \Delta\beta_{g,t,\mathcal{K}}^{0\prime} x_{1t}x_{1t}' \Delta\beta_{h,t,\mathcal{K}}^0 = \begin{cases} \delta_{gj}' \big(v_T^2 \sum_{t=k_{gj}\wedge k_{hj}+1}^{T_j^0} x_{1t}x_{1t}' \big) \delta_{hj}, & \text{if } r_{gj}, r_{hj} < 0 \\ \delta_{gj}' \big(v_T^2 \sum_{t=T_j^0+1}^{k_{gj}\vee k_{hj}} x_{1t}x_{1t}' \big) \delta_{hj}, & \text{if } 0 < r_{gj}, r_{hj} \\ 0, & \text{otherwise.} \end{cases}$$

Define

$$W_{T,j}(r_{gj}, r_{3j}) = \frac{1}{(\sigma_{j+\mathbb{1}_{\{r_{3j}<0\}}}^0)^2} v_T \sum_{t=\underline{k}_{gj}+1}^{\overline{k}_{gj}} x_{1t}u_{1t} \quad \text{and} \quad B_{T,j}(r_{gj}, r_{3j}) = \frac{1}{(\sigma_{j+\mathbb{1}_{\{r_{3j}<0\}}}^0)^2} v_T \sum_{t=\underline{k}_{gj}+1}^{\overline{k}_{gj}} x_{1t}x_{1t}'.$$

Thus,

$$\begin{aligned} &\ell_{T,2}^{(j)}(\boldsymbol{r}_j) \\ &= -\sum_{g=1}^{3} \mathrm{sgn}(r_{gj}) v_T \delta_{gj}' W_{T,j}(r_{gj}, r_{3j}) \\ &\quad -\frac{1}{2} \sum_{g=1}^{3} \sum_{h=1}^{3} \delta_{gj}' \Big\{ \mathbb{1}_{\{r_{gj}\wedge r_{hj}\le 0\}} B_{T,j}(r_{gj}\wedge r_{hj}, r_{3j}) + \mathbb{1}_{\{r_{gj}\vee r_{hj}\le 0\}} B_{T,j}(r_{gj}\vee r_{hj}, r_{3j}) \Big\} \delta_{hj} + o_p(1). \end{aligned}$$

Applying the FCLT to (S.10) and the equation above, we can show that

$$\big(\ell_T^{(1)}(\boldsymbol{r}_1), \ell_T^{(2)}(\boldsymbol{r}_2)\big) \Rightarrow \big(\ell_\infty^{(1)}(\boldsymbol{r}_1), \ell_\infty^{(2)}(\boldsymbol{r}_2)\big),$$

where

$$\begin{aligned} 2\ell_\infty^{(j)}(\boldsymbol{r}_j) \;=\; & -\mathrm{sgn}(r_{3j}) \frac{\phi_j}{(\sigma_{j+\mathbb{1}_{\{r_{3j}<0\}}}^0)^2} \mathbb{V}_{\eta\eta}(r_{3j}) + \frac{\phi_j^2}{2(\sigma_{j+\mathbb{1}_{\{r_{3j}<0\}}}^0)^4} r_{3j} \\ & -2 \sum_{g=1}^{3} \mathrm{sgn}(r_{gj}) v_T \delta_{gj}' \mathbb{W}_j(r_{gj}, r_{3j}) \\ & -\sum_{g=1}^{3} \sum_{h=1}^{3} \delta_{gj}' \Big\{ \mathbb{1}_{\{r_{gj}\wedge r_{hj}\le 0\}} \mathbb{B}_j(r_{gj}\wedge r_{hj}, r_{3j}) + \mathbb{1}_{\{r_{gj}\vee r_{hj}\le 0\}} \mathbb{B}_{T,j}(r_{gj}\vee r_{hj}, r_{3j}) \Big\} \delta_{hj}. \end{aligned}$$

We apply the same argument used in the previous section. Define $\delta_j = (\delta_j^{(1)}, \delta_j^{(2)}, \delta_j^{(3)})'$ and $\boldsymbol{s_j} = (s_{1j}, s_{2j}, s_{3j})' \in \mathbb{R}^3$. Let $\boldsymbol{s}_j := \omega \boldsymbol{r}_j$ with $\omega := \|\delta_j\|^2 + |\phi_j|^2$ Then, $2\ell_\infty^{(j)}(\boldsymbol{r}_j) = CB_\infty^{(j)}(\boldsymbol{s}_j)$ for $j = 1, 2$, where

$$\begin{aligned} CB_\infty^{(j)}(\boldsymbol{s}_j) \;=\; & -\mathrm{sgn}(r_{3j})\omega^{-1/2}\frac{\phi_j}{(\sigma^0_{j+\mathbb{1}_{\{r_{3j}<0\}}})^2}\mathbb{V}_{\eta\eta}(r_{3j}) + \omega^{-1}\frac{\phi_j^2}{2(\sigma^0_{j+\mathbb{1}_{\{r_{3j}<0\}}})^4} r_{3j} \\ & -2\omega^{-1/2}\sum_{g=1}^{3}\mathrm{sgn}(r_{gj})\upsilon_T\delta'_{gj}\mathbb{W}_j(r_{gj}, r_{3j}) \\ & -\omega^{-1}\sum_{g=1}^{3}\sum_{h=1}^{3}\delta'_{gj}\Big\{\mathbb{1}_{\{r_{gj}\wedge r_{hj}\le 0\}}\mathbb{B}_j(r_{gj}\wedge r_{hj}, r_{3j}) + \mathbb{1}_{\{r_{gj}\vee r_{hj}\le 0\}}\mathbb{B}_{T,j}(r_{gj}\vee r_{hj}, r_{3j})\Big\}\delta_{hj}. \end{aligned}$$

Therefore, we have

$$CB_T \Rightarrow \sup_{\boldsymbol{s}_1,\boldsymbol{s}_2}\sum_{j=1}^{2} CB_\infty^{(j)}(\boldsymbol{s}_j) - \sup_{s_1,s_2}\sum_{j=1}^{2} CB_\infty^{(j)}(s_j\cdot\mathbf{1}),$$

where $\mathbf{1}$ is a 3×1 vector having 1 at all entries. In practice, unknown parameters in this limit distribution can be replaced by consist estimates as explained in the previous section.