

Estimating and testing structural changes in multivariate regressions *

Zhongjun Qu[†]

Boston University

Pierre Perron[‡]

Boston University

This version: March 2, 2005

Abstract

This paper considers issues related to estimation, inference and computation with multiple structural changes occurring at unknown dates in a system of equations. Changes can occur in the regression coefficients and/or the covariance matrix of the errors. We also allow arbitrary restrictions on these parameters, which permits the analysis of partial structural change models, common breaks occurring in all equations, breaks occurring in a subset of equations, etc. The method of estimation is quasi maximum likelihood based on Normal errors. The limiting distributions are obtained under more general assumptions than previous studies. Of special interest is the fact that substantial efficiency gains can be obtained by casting a regression affected by changes in a system even if the other equations are not affected by breaks, provided there is non-zero correlation between the errors. For testing, we propose likelihood ratio type statistics to test the null hypothesis of no structural change and to select the number of changes. Structural change tests with restrictions on the parameters can be constructed to achieve higher power when prior information is present. We propose an algorithm for an efficient procedure to construct the estimates and test statistics. We also introduce a novel locally ordered breaks model, which allows the breaks in different equations to be related yet not occurring at the same dates.

Keywords: change-point, segmented regressions, break dates, hypothesis testing, model selection, system of regressions.

*Pierre Perron acknowledges financial support from the National Science Foundation under Grant SES-0078492.

[†]Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (qu@bu.edu).

[‡]Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (perron@bu.edu).

1 Introduction

Both the statistics and econometrics literature contain a vast amount of work on issues related to structural changes with unknown break dates, most of it specifically designed for the case of a single change. The problem of multiple structural changes has received more attention recently mostly in the context of a single regression. Bai and Perron (1998, 2003a) provide a comprehensive treatment of various issues in the context of multiple structural change models: consistency of estimates of the break dates, tests for structural changes, confidence intervals for the break dates, methods to select the number of breaks and efficient algorithms to compute the estimates. Perron and Qu (2004) extend this analysis to the case where arbitrary linear restrictions are imposed on the coefficients of the model. Related contributions include Hawkins (1976) who presents a comprehensive treatment of estimation based on a dynamic programming algorithm. Also, Liu, Wu and Zidek (1997) consider multiple structural changes in the context of a more general threshold model and propose an information criterion for the selection of the number of changes.

Work on structural change issues arising in the context of a system of multivariate equations is relatively scarce. Bai, Lumsdaine and Stock (1998) consider asymptotically valid inference for the estimate of a single break date in multivariate time series allowing stationary or integrated regressors as well as trends. They show that the width of the confidence interval decreases in an important way when series having a common break are treated as a group and estimation is carried using a quasi maximum likelihood (QML) procedure. Also, Bai (2000) considers the consistency, rate of convergence and limiting distribution of estimated break dates in a segmented stationary VAR model estimated again by QML when the break can occur in the parameters of the conditional mean, the variance of the error term or both. Hansen (2003) considers multiple structural changes in a cointegrated system, though his analysis is restricted to the case of known break dates.

The aim of this paper is to provide a comprehensive treatment of issues related to estimation, inference and computation with multiple structural changes occurring at unknown dates in linear multivariate regression models that include VAR, certain linear panel data models and SUR. Changes can occur in the parameters of the conditional mean, the covariance matrix of the errors or both and the distribution of the regressors is allowed to change across regimes. The assumptions regarding the distribution of the error processes are quite mild such that both conditional heteroskedasticity and autocorrelation are allowed. Also, our general framework allows for the incorporation of arbitrary valid restrictions on

the parameters. This permits a host of practically interesting models to be analyzed. For example, our framework applies to partial structural change models where a subset of the coefficients do not change across regimes, to models where only a subset of the regressions are affected by breaks, to models with common breaks across regressions, and others. Also, these restrictions can be non-linear in the parameters.

The method of estimation is quasi maximum likelihood based on Normal errors. We derive the consistency, rate of convergence and limiting distribution of all parameters, which appears to be the first asymptotic treatment in a multivariate setting with a general error process. An important, and at first sight surprising, result is that the limiting distribution of the estimates of the break dates is unaffected by the imposition of valid restrictions on the other parameters of the model. In large samples, no efficiency gains emerge from such restrictions. Hence, the limiting distribution of the estimates of the break dates is only affected by the underlying structure of the system. Our results include that of Bai, Lumsdaine and Stock (1998) as a special case, namely that with common breaks across equations the precision of the estimates increases with the number of equations in the system. Hence, their result for the single break case extends to multiple structural changes. More importantly, our general framework allows to uncover other important features that affect the precision of the estimates. One that is particularly interesting is that the precision of the estimate of a particular break date in one equation can increase when the system includes other equations even if the parameters of the latter are invariant across regimes. All that is needed is that the correlation between the errors be non-zero. While surprising, this result is ex-post fairly intuitive since a poorly estimated break in one regression affects the likelihood function through the residual variance of that equation and also via the correlation with the rest of the regressions. Hence, by including ancillary equations without breaks, additional forces are in play to better pinpoint the break dates.

For the parameters of the conditional mean and the covariance matrix of the errors, the limiting distributions are the same with estimated break dates as they are when the break dates are known. Hence, standard distributions apply and these parameters are indeed more efficiently estimated with valid restrictions imposed. Hence, for this reason, and others related to the power of tests for structural changes, it is important to have a feasible and efficient algorithm that permits the computation of all estimates. We discuss a procedure that extends the work of Hawkins (1976) and Bai and Perron (2003a) based on a dynamic programming algorithm that searches for an optimal partition by efficiently looking at various combinations of the globally maximized likelihood function implied by a given partition.

Since we use a QMLE method, we apply an iterative Feasible Generalized Least Squares (FGLS) procedure to obtain the estimates. We use results by Tobing and McGilchrist (1992) to construct the recursive residuals from each FGLS iteration needed to apply the dynamic programming algorithm to search for the optimal partition. The presence of restrictions on the parameters imposes an additional layer of iterations. Nevertheless, the method is very efficient as we only need least squares computations of order $O(T)$ and matrix inversions of only order $O(n)$ where n is the number of equations in the system.

To determine the number of breaks in the system, Bai (2000) proposes the use of an information criterion, following the work of Yao (1988). We instead consider, as in Bai and Perron (1998), the use of testing procedures. We derive the likelihood ratio test of no change versus some specific number of changes, say k , for the following cases: a) changes in the coefficients of the conditional mean; b) changes in the coefficient of the covariance matrix of the residuals, c) changes in both. Our procedure appears to be first that can test jointly changes in the coefficients and in the covariance matrix. We also consider a sequential procedure based on a test of, say, l versus $l + 1$ changes and discuss how to incorporate valid restrictions into the structural change tests to achieve higher power. We show that, for important classes of restrictions, the limiting distributions of the tests depend only on a parameter related to the number of coefficients allowed to change and that we can rely on critical values already available. This includes the partial structural change model and models where breaks occur only in a subset of the equations. For other forms of restricted models, which include globally ordered breaks model and model with switching regimes, the distributions of the tests differ from the existing ones but can easily be simulated.

We also introduce a novel structure that we label “locally ordered breaks”. Consider a two equations system where one equation is a policy reaction function (e.g., a monetary policy function) and the other is some market clearing equation (e.g., an expectation augmented Phillips curve equation; see, e.g. Alogoskoufis and Smith, 1991, and Bai and Perron, 2003a). Suppose there is a change in the monetary policy regime which implies a break in the policy function at some date. The policy change is expected to have an impact on the Phillips curve equation but the change may not be simultaneous and may occur with a lag, say because of some adjustments due to frictions or incomplete information. However, it is expected to take place soon after the break in the policy function. Here, the breaks across the two equations are “ordered” in the sense that we have the prior knowledge that the break in the Phillips curve equation occurs after the break in the policy function. The breaks are also “local” in the sense that the time span between their occurrence is expected to be short. We label

such a structure as “locally ordered breaks”, in which case the usual asymptotic framework that requires breaks to be separated by a positive fraction of the sample does not apply. Accordingly, we provide appropriate methods for estimation, inference and testing.

The structure of the paper is as follows. Section 2 presents the general model with restrictions and the assumptions imposed on the regressors, errors and parameters. In Section 3, we derive the consistency, rate of convergence and limiting distribution of the estimates. An important ingredient is a result that expresses the restricted likelihood ratio in two parts, one that involves only the break dates and the true values of the coefficients (so that the estimates of the break dates are not affected by the restrictions imposed on the coefficients), the other involving the parameters of the model, the true value of the break date and the restrictions (showing that the limiting distributions of these estimates are influenced by the restrictions but not by the estimation of the break dates). Section 4 considers the algorithm to construct the estimates of the model. Section 5 introduces locally ordered breaks, methods to estimate such break dates and discusses their limiting distributions. Section 6 considers hypothesis testing related to the presence or absence of structural changes and the determination of the number of breaks. Section 7 offers brief conclusions and an appendix contains the proofs of the results stated in the text.

2 The model and the assumptions

We first define the notation used throughout. We have n equations and T observations excluding the initial conditions if lagged dependent variables are used as regressors. The total number of structural changes in the system is m . The break dates are denoted by the m vector $\mathcal{T} = (T_1, \dots, T_m)$ and we use the convention that $T_0 = 1$ and $T_{m+1} = T$. A subscript j indexes a regime ($j = 1, \dots, m + 1$). A subscript t indexes a temporal observation ($t = 1, \dots, T$) and a subscript i indexes the equation ($i = 1, \dots, n$) to which a scalar dependent variable y_{it} is associated. The parameter q is the number of regressors and z_t is the set that includes the regressors from all equations $z_t = (z_{1t}, \dots, z_{qt})'$. The model considered is of the form

$$y_t = (I \otimes z_t')S\beta_j + u_t \tag{1}$$

with u_t having mean 0 and covariance matrix Σ_j for $T_{j-1} + 1 \leq t \leq T_j$ ($j = 1, \dots, m + 1$). The matrix S is of dimension nq by p with full column rank. Though, in principle it is allowed to have entries that are arbitrary constants, it is usually a selection matrix involving elements that are 0 or 1 and, hence, specifies which regressors appear in each equation. The set of

basic parameters in regime j consists of the p vector β_j and Σ_j . We allow for the imposition of a set of r restrictions of the form:

$$g(\beta, \text{vec}(\Sigma)) = 0 \quad (2)$$

where $\beta = (\beta'_1, \dots, \beta'_{m+1})'$, $\Sigma = (\Sigma_1, \dots, \Sigma_{m+1})$ and $g(\cdot)$ is an r dimensional vector. Note that we allow within and cross equation restrictions and in each case within or across regimes.

Remark 1 *The model considered is very general. It includes the standard SUR model when $p = q$ and S is a selection matrix. The standard VAR model applies when we further have $z_t = (y_{t-1}, \dots, y_{t-q})'$. A panel data model can be obtained using an appropriate selection matrix for S . A partial structural change model obtains when the restrictions (2) imposes that a particular subset of β_j is the same for all j (e.g., a dynamic panel data model with a sufficient time span where the parameter on the lagged dependent variable is the same for all units and regimes but the fixed effects are allowed to change across units and regimes). Similarly, breaks occurring in only a subset of the equations applies when the restrictions specify that the relevant subset of β_j is the same for all j . Of course, more complex models are applicable using other forms of restrictions.*

Remark 2 *The case where breaks occur at known dates in a subset of equations can easily be handled. Here, the regressors z_{it} for this subset are constructed according to the known partition, e.g. use $z_{it}^* = \{z_{it}, z_{it}1(t > T_1)\}$ in the case of a single break. Then, restrictions are imposed to constrain the coefficients of this subset not to change across regimes. This can be useful, for instance, when one equation represents a policy function with a change occurring at a known date and the other subset contains behavioral equations that should exhibit a change in response to the change in policy but with an unknown delay.*

Remark 3 *(Common breaks) The case where the breaks are the same in all equations occurs when at least one coefficient from each equation is not restricted to be the same across any two adjacent segments implied by the specified number of breaks.*

Remark 4 *To show how versatile our framework is, consider the case of an ordered break model. Suppose a multi-equation (three equations or more) model with a maximum of 3 breaks allowed, i.e., there are at most four segments. It is supposed that two equations (the first and second without loss of generality) in the system have a single break which is not common. Also, the break for the second equation occurs after that for the first. Then, three*

cases have to be considered: 1) in the first equation the parameters are the same for the last three segments and in the second they are the same for segments 1 and 2 and for segment 3 and 4; 2) the same for the first equation but in the second the parameters are the same in the first three segments; 3) in the first equation the parameters are the same in segments 1 and 2 and 3 and 4 while in the second equation they are the same for segments 1 to 3. One can then evaluate which of the three cases is the one supported by the data by looking at the appropriate objective function, here the likelihood function (see below). The same principle applies to more complex models.

To ease notation, define the n by p matrix x_t by $x'_t = (I \otimes z'_t)S$ so that (1) becomes

$$y_t = x'_t \beta_j + u_t \quad (3)$$

for $T_{j-1} + 1 \leq t \leq T_j$ ($j = 1, \dots, m + 1$). It is useful to express the model in matrix form. Let $Y = (y'_1, \dots, y'_T)'$ be the nT vector of dependent variables, $U = (u'_1, \dots, u'_T)'$ be the error vector and the nT by p matrix of regressors is $X = (x_1, \dots, x_T)'$. For a given partition of the sample using the breaks (T_1, \dots, T_m) , we define the block partition of the matrix X as the nT by $p(m + 1)$ matrix $\bar{X} = \text{diag}(X_1, \dots, X_{m+1})$ where X_j ($j = 1, \dots, m + 1$) is the $n(T_j - T_{j-1})$ by p subset of X that corresponds to observations in regime j . We also define the sub-vector U_j of U in a similar way. Then the regression system (3) can be expressed as

$$Y = \bar{X}\beta + U$$

The true values of the parameters are denoted with a 0 superscript so that the Data Generating Process is assumed to be

$$Y = \bar{X}^0 \beta^0 + U$$

where \bar{X}^0 is the diagonal partition of X using the partition (T_1^0, \dots, T_m^0) . Our analysis is carried under the following set of assumptions.

- Assumption A1: For each $j = 1, \dots, m + 1$ and $l_j \leq T_j^0 - T_{j-1}^0$, $(1/l_j) \sum_{t=T_{j-1}^0+1}^{T_{j-1}^0+l_j} x_t x'_t \xrightarrow{a.s} Q_j^0$ as $l_j \rightarrow \infty$, with Q_j^0 a nonrandom positive definite matrix not necessarily the same for all j .
- Assumption A2: There exists an $l_0 > 0$ such that for all $l > l_0$, the minimum eigenvalues of $(1/l) \sum_{j=T_j^0+1}^{T_j^0+l} x_t x'_t$ and of $(1/l) \sum_{t=T_j^0-l}^{T_j^0} x_t x'_t$ are bounded away from zero ($j = 1, \dots, m$).
- Assumption A3: There exists a $0 < k_0 < \infty$ such that matrix $\sum_{t=k}^l x_t x'_t$ is invertible for $l - k \geq k_0$.

- Assumption A4: Define the L_r -norm of a random matrix X as $\|X\|_r = (\sum_i \sum_j E |X_{ij}|^r)^{1/r}$ for $r \geq 1$ and $\mathcal{F}_t = \sigma\text{-field } \{\dots, x_{t-1}, x_t, \dots, u_{t-2}, u_{t-1}\}$. If $x_t u_t$ is weakly stationary within each segment, then (a) $\{x_t u_t, \mathcal{F}_t\}$ forms a strongly mixing (α -mixing) sequence with size $-4r/(r-2)$ for some $r > 2$, (b) $E(x_t u_t) = 0$ and $\|x_t u_t\|_{2r+\delta} < M < \infty$ for some $\delta > 0$, (c) Let $S_{k,j}(\ell) = \sum_{T_{j-1}^0 + \ell + 1}^{T_{j-1}^0 + \ell + k} x_t u_t$, $j = 1, \dots, m+1$, for each $e \in R^n$ of length 1, $\text{var}(\langle e, S_{k,j}(0) \rangle) \geq v(k)$ for some function $v(k) \rightarrow \infty$ as $k \rightarrow \infty$. If $x_t u_t$ is not weakly stationary within each segment, we assume that (a)-(c) holds, and in addition, that there exists a positive definite matrix $\Omega = [w_{i,s}]$ such that for any $i, s = 1, \dots, p$, we have, uniformly in ℓ , $|k^{-1} E((S_{k,j}(\ell))_i (S_{k,j}(\ell))_s) - w_{i,s}| \leq C_2 k^{-\psi}$, for some $C_2, \psi > 0$.

- Assumption A5: Assumption A4 holds with $x_t u_t$ replaced by u_t or $u_t u_t' - \Sigma_j^0$, for $T_{j-1}^0 < t \leq T_j^0$ ($j = 1, \dots, m+1$).

- Assumption A6: The magnitudes of the shifts satisfy $\beta_{T,j+1}^0 - \beta_{T,j}^0 = v_T \delta_j$ and $\Sigma_{j+1,T}^0 - \Sigma_{j,T}^0 = v_T \Phi_j$, where $(\delta_j, \Phi_j) \neq 0$ and are independent of T . Moreover, v_T is either a positive number independent of T or a sequence of positive numbers satisfying $v_T \rightarrow 0$ and $T^{1/2} v_T / (\log T)^2 \rightarrow \infty$.¹

- Assumption A7: $(\beta^0, \Sigma^0) \in \bar{\Theta}$, with $\bar{\Theta} = \{(\beta, \Sigma) : \|\beta\| \leq c_1, \lambda_{\min}(\Sigma) \geq c_2, \lambda_{\max}(\Sigma) \leq c_3\}$ for some $c_1 < \infty, 0 < c_2 \leq c_3 < \infty$ (where λ_{\min} and λ_{\max} denotes the smallest and largest eigenvalues).

- Assumption A8: $0 < \lambda_1^0 < \dots < \lambda_m^0 < 1$ with $T_i^0 = [T \lambda_i^0]$.

Assumption 1 basically rules out unit root regressors; trending regressors in the form of $(t/T, \dots, (t/T)^\alpha)$ could be permitted at the expense of some technical complications². The assumption allows the regressors to have different distributions in different regimes. This is important since a change in a dynamic model directly leads to distributional change in the moments of the regressors. Assumption A2 imposes restrictions on the regressors in a local neighborhood of the break points. They ensure that there is no local collinearity problem so the break points can be identified. Assumption A3 is a standard invertibility requirement to have well defined estimates.

Assumptions A4 and A5 determine the dependence structure of the processes $x_t u_t$ and u_t . In particular, they imply that $x_t u_t$ and $u_t u_t'$ are short memory processes having bounded forth

¹In order not to over-burden the notation, we will omit the subscript T on $\beta_{T,j}^0$ and $\Sigma_{T,j}^0$, and simply use β_j^0 and Σ_j^0 . This should cause no confusion.

²However, for the limiting distribution of the estimates of break dates and the test statistics, trending regressors are not permitted, as will be made clear later.

moments. The assumptions are imposed to obtain a strong invariance principle, strongly consistent estimates of the coefficients and to have a well behaved likelihood function. The conditions are mild in the sense that they allow for substantial conditional heteroskedasticity and autocorrelation. Also, if no autocorrelation is present, i.e., $\{x_t u_t\}$ and $\{u_t\}$ are martingale difference sequences with respect to the filtration \mathcal{F}_t , then even the weak stationarity assumption can be dropped and u_t allowed to be unconditionally heteroskedastic with bounded fourth moments. Examples of models generated under A4 and A5 are, among others, finite order stationary Vector Autoregressive Model with bounded fourth moment, dynamic panel models with uncorrelated but heteroskedastic errors, and models with only exogenous regressors and stationary short memory errors, such as stationary $ARMA(p, q)$ processes. Note that Assumptions A4 and A5 could be replaced by other sufficient conditions that can yield a strong invariance principle or FCLT, a bounded law of the iterated logarithm, and at the same time, guarantee that a generalized Hajek and Renyi type inequality holds, such as the one in Bai and Perron (1998).³

Assumptions A1-A5 could equally be formulated in terms of the original regressors z_t . This would imply the same conditions for x_t and $x_t u_t$, given that S is a constant matrix with full column rank and that the strong mixing property in A4 is preserved under measurable transformations.

Assumption A6 is standard in the structural change literature, it gives conditions under which the structural changes are asymptotically non-negligible. Assuming a fixed v_T captures the feature of large shifts and a shrinking v_T corresponds to small or intermediate shifts in finite samples. The latter case allows the development of a limiting theory for the break date estimates which does not depend on the exact distributions of the regressors and the errors.

Assumption A7 implies that the data is generated by innovations with a non-degenerate covariance matrix and a finite conditional mean. Assumption A8 is also a standard assumption, which implies asymptotically distinct breaks, i.e., each regime contains a positive fraction of the sample even in the limit.

The set of assumptions used differ from those used in Bai and Perron (1998). The main difference is that when heterogeneity and serial correlation is allowed in the errors, the regressors x_t are not assumed to be independent of the errors u_t at all leads and lags. This permits a wider class of models, which has considerable practical advantages. The cost is the need to introduce slightly stronger technical conditions. In particular A1 is stronger than

³Examples of such sufficient conditions are discussed by Dehling and Philipp (1982), Altissimo and Corradi (2003) and Lavielle and Moulines (2000), among others.

the requirements in Bai and Perron (1998). We also impose that the search for the break dates be done in a set that imposes each segment to be some non-vanishing proportion of the sample, see Assumption A9 below (this was also imposed by Bai and Perron, 1998, but only when lagged dependent variables are permitted as regressors). Our set of assumptions is also considerably more general than that used in Bai, Lumsdaine and Stock (1998) and Bai (2000), where they assume martingale difference errors.

2.1 The estimation method

The method of estimation considered is restricted quasi maximum likelihood (*RQML*) assuming serially uncorrelated Gaussian errors.⁴ Conditional on a given partition of the sample $\mathcal{T} = (T_1, \dots, T_m)$, the Gaussian quasi-likelihood function is

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

where

$$f(y_t | x_t; \beta_j, \Sigma_j) = \frac{1}{(2\pi |\Sigma_j|)^{1/2}} \exp \left\{ -\frac{1}{2} [y_t - x_t' \beta_j]' \Sigma_j^{-1} [y_t - x_t' \beta_j] \right\} \quad (4)$$

And the quasi-likelihood ratio is

$$LR_T = \frac{\prod_{j=1}^{m+1} \prod_{t=T_{j-1}+1}^{T_j} f(y_t | x_t; \beta_j, \Sigma_j)}{\prod_{j=1}^{m+1} \prod_{t=T_{j-1}^0+1}^{T_j^0} f(y_t | x_t; \beta_j^0, \Sigma_j^0)} \quad (5)$$

The aim is to obtain values of $(T_1, \dots, T_m, \beta, \Sigma)$ that maximize LR_T subject to the restrictions

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

Let $lr_T(\cdot)$ denote the log likelihood ratio and $rlr_T(\cdot)$ the restricted log likelihood ratio, then our objective function is

$$rlr_T(\mathcal{T}, \beta, \Sigma) = lr_T(\mathcal{T}, \beta, \Sigma) + \lambda' g(\beta, \text{vec}(\Sigma)) \quad (6)$$

and the estimates are

$$(\tilde{\mathcal{T}}, \tilde{\beta}, \tilde{\Sigma}) = \arg \max_{(T_1, \dots, T_m; \beta; \Sigma)} rlr_T(\mathcal{T}, \beta, \Sigma) \quad (7)$$

Throughout, we also impose the following assumptions on the set of permissible partitions where ε acts as a trimming and imposes a minimal length for each regime.

⁴For an extensive review of the applications of the likelihood principle in structural change problems, see Csörgő and Horváth (1997).

- Assumption A9: The maximization (7) is taken over all partitions $\mathcal{T} = (T_1, \dots, T_m) = (T\lambda_1, \dots, T\lambda_m)$ in the set

$$\Lambda_\varepsilon = \{(\lambda_1, \dots, \lambda_m); |\lambda_{j+1} - \lambda_j| \geq \varepsilon, \lambda_1 \geq \varepsilon, \lambda_m \leq 1 - \varepsilon\}. \quad (8)$$

3 The limiting distributions of the estimates.

In this Section, we consider the limiting distribution of the estimates. We start with the following Lemma.

Lemma 1 *Under Assumptions A1-A9, we have: for $j = 1, \dots, m$,*

$$v_T^2(\tilde{T}_j - T_j^0) = O_p(1)$$

and for $j = 1, \dots, m + 1$,

$$\sqrt{T}(\tilde{\beta}_j - \beta_j^0) = O_p(1) \text{ and } \sqrt{T}(\tilde{\Sigma}_j - \Sigma_j^0) = O_p(1)$$

This Lemma establishes the rate of convergence of the estimates. The results are the same as in most other cases considered in the literature, see Bai (1997, 2000) and Bai and Perron (1998), and imply the following Corollary since the break fractions λ_j are estimated at a rate fast enough not to affect the distribution of $\tilde{\beta}$ asymptotically.

Corollary 1 *Under Assumptions A1-A9, the limiting distribution of $\sqrt{T}(\tilde{\beta} - \beta^0)$ is the same as in the case with known break dates.*

Lemma 1 also allows us to analyze issues related to the limiting distributions by analyzing the behavior of the restricted log likelihood function in a particular compact subset of the parameter space in a neighborhood of the true value. This subset is defined by

$$C_M = \left\{ (\mathcal{T}, \beta, \Sigma) : v_T^2 |T_j - T_j^0| \leq M \text{ for } j = 1, \dots, m, \right. \\ \left. |\sqrt{T}(\beta_j - \beta_j^0)| \leq M, |\sqrt{T}(\Sigma_j - \Sigma_j^0)| \leq M, j = 1, \dots, m + 1 \right\} \quad (9)$$

where M is a fixed positive number. Using Lemma 1, restricting our analysis to values of the parameters in the set C_M is without loss of generality since we can choose M large enough so that the estimates fall in that set with probability arbitrarily close to one. We now state an important result that expresses the restricted likelihood in two parts: one that involves only the break dates and the true values of the coefficients, so that the estimates of the break

dates are not affected by the restrictions imposed on the coefficients; the other involving the parameters of the model, the true values of the break dates and the restrictions, showing that the limiting distributions of these estimates are influenced by the restrictions but not the estimation of the break dates.

Theorem 1 *Under Assumptions A1-A9, we have, for $r = T_j - T_j^0$:*

$$\max_{(\mathcal{T}, \beta, \Sigma) \in C_M} r l r_T = \max_{\mathcal{T} \in C_M, \beta^0, \Sigma^0} \sum_{j=1}^m l r_j^1(r) + \max_{(\beta, \Sigma) \in C_M, \mathcal{T}^0} \left[\sum_{j=1}^{m+1} l r_j^2 + \lambda' g(\beta, \text{vec}(\Sigma)) \right] + o_p(1)$$

where, $l r_j^1(0) = 0$,

$$\begin{aligned} l r_j^1(r) &= \frac{1}{2} \sum_{t=T_j^0+r}^{T_j^0} u_t' ((\Sigma_j^0)^{-1} - (\Sigma_{j+1}^0)^{-1}) u_t - \frac{r}{2} (\log |\Sigma_j^0| - \log |\Sigma_{j+1}^0|) \\ &\quad - \frac{1}{2} \sum_{t=T_j^0+r}^{T_j^0} (\beta_j^0 - \beta_{j+1}^0)' x_t (\Sigma_{j+1}^0)^{-1} x_t' (\beta_j^0 - \beta_{j+1}^0) - \sum_{t=T_j^0+r}^{T_j^0} (\beta_j^0 - \beta_{j+1}^0)' x_t (\Sigma_{j+1}^0)^{-1} u_t \end{aligned}$$

for $r = -1, -2, \dots$,

$$\begin{aligned} l r_j^1(r) &= -\frac{1}{2} \sum_{t=T_j^0+1}^{T_j^0+r} u_t' ((\Sigma_j^0)^{-1} - (\Sigma_{j+1}^0)^{-1}) u_t - \frac{r}{2} (\log |\Sigma_j^0| - \log |\Sigma_{j+1}^0|) \\ &\quad - \frac{1}{2} \sum_{t=T_j^0+1}^{T_j^0+r} (\beta_j^0 - \beta_{j+1}^0)' x_t (\Sigma_j^0)^{-1} x_t' (\beta_j^0 - \beta_{j+1}^0) + \sum_{t=T_j^0+1}^{T_j^0+r} (\beta_j^0 - \beta_{j+1}^0)' x_t (\Sigma_j^0)^{-1} u_t \end{aligned}$$

for $r = 1, 2, \dots$, and

$$\begin{aligned} l r_j^2 &= -\frac{1}{2} \sum_{t=T_{j-1}^0+1}^{T_j^0} (y_t - x_t' \beta_j)' \Sigma_j^{-1} (y_t - x_t' \beta_j) - \frac{T_j^0 - T_{j-1}^0}{2} \log |\Sigma_j| \\ &\quad + \frac{1}{2} \sum_{t=T_{j-1}^0+1}^{T_j^0} (y_t - x_t' \beta_j^0)' (\Sigma_j^0)^{-1} (y_t - x_t' \beta_j^0) + \frac{T_j^0 - T_{j-1}^0}{2} \log |\Sigma_j^0| \end{aligned}$$

This result has strong implications. It says that the optimization problem can be separated into two asymptotically independent components. The first pertains to maximizing with respect to the break dates and does not involve the restrictions on the coefficients but

only the true values (β^0, Σ^0) . The second component pertains to maximizing a term involving the restrictions with respect to (β, Σ) keeping the break dates fixed at their true values T^0 . Hence, the estimates of the break dates \mathcal{T} and the coefficients (β, Σ) are asymptotically independent and the restrictions on the latter do not affect the distribution of the former.

Our result provides theoretical explanation for some simulation findings reported in Bai, Lumsdaine and Stock (1998), which show that the precision of the estimates of the break dates does not improve much even when the sample size is greatly increased. This is clearly an implication of the above result, which states that the estimation of the break dates is not sensitive to the precision of the parameter estimates. The result shows what kind of information can asymptotically improve the efficiency of the breaks estimates. Of relevance are the true values of the parameters (β^0, Σ^0) , in particular the changes in their values across regimes, and the extent of the correlation across errors, as discussed in more details below.

Using this Lemma it becomes easy to derive the limiting distribution of the estimates of the break dates under both fixed and shrinking magnitudes of shifts. For fixed shifts, we have the following Theorem. The proof is immediate and hence omitted.

Theorem 2 *Under Assumptions A1-A9, with v_T a fixed constant and assuming*

$$(\beta_j^0 - \beta_{j+1}^0)' x_t x_t' (\beta_j^0 - \beta_{j+1}^0)' \pm (\beta_j^0 - \beta_{j+1}^0)' x_t (\Sigma_{j+1}^0)^{-1} u_t$$

has a continuous distribution, we have for $j = 1, \dots, m$:

$$\tilde{T}_j - T_j^0 \rightarrow^d \arg \max_r lr_j^1(r).$$

where the maximization is taken over the set of integers.

The Theorem clearly shows that the restrictions on the coefficients do not enter the distribution of the breaks asymptotically. It also shows that restricting the errors to have an homogeneous distribution across regimes, even if true, does not bring efficiency gain asymptotically. In light of the fact that if the error process does undergo changes in distribution, restricting it to be stable implies a misspecified model, this suggests that a robust choice in practice is to always allow the errors to be regime dependent. It is also important to allow the distribution of the regressors to vary across regimes. Otherwise, the confidence interval will not have the correct coverage even asymptotically. The simulations of Bai and Perron (2004), done in a single equation framework, show that it is also preferable to allow it to be regime dependent even if the actual distribution does not change.

The result also says that the error covariance plays an important role in determining the asymptotic distribution. More weights are given on information from the equations with smaller variance. The formula also shows a tension between equations through the correlations in the errors. This suggests that to estimate breaks by minimizing the total sum of squared residuals is likely to result in potentially important efficiency losses even if the different equations all bear the same information to noise ratio.

The drawback of the result in Theorem 2 is that the limiting distribution depends on the exact distribution of the errors. This is a standard problem in the literature and the usual remedy is to consider an asymptotic framework whereby the shifts are shrinking in magnitude as the sample size increases. Stronger assumptions are, however, necessary, in particular to rule out trending regressors. More specifically, we introduce the following assumption.

- Assumption A10: Let $\Delta T_j^0 = T_j^0 - T_{j-1}^0$; for $j = 1, \dots, m$, as $\Delta T_j^0 \rightarrow \infty$, uniformly in $s \in [0, 1]$, $(\Delta T_j^0)^{-1} \sum_{t=T_{j-1}^0+1}^{T_{j-1}^0+[s\Delta T_j^0]} x_t x_t' \rightarrow_p s Q_j^0$ with Q_j^0 a nonrandom positive definite matrix not necessarily the same for all j .

The result pertaining to that case is stated in the next Theorem.

Theorem 3 *Let $\eta_t \equiv (\eta_{t1}, \dots, \eta_{tm}) = (\Sigma_j^0)^{-1/2} u_t$ for $t \in [T_{j-1}^0 + 1, T_j^0]$, and assume that $E[\eta_{tk} \eta_{ul} \eta_{th}] = 0$ for all k, l, h and for every t . Under Assumptions A1-A10, with $v_T \rightarrow 0$ such that $T^{1/2} v_T / (\log T)^2 \rightarrow \infty$ as $T \rightarrow \infty$, and with “ \Rightarrow ” denoting weak convergence under the Skorohod topology, we have, for $j = 1, \dots, m$:*

$$\frac{\Delta_{1,j}^2}{\Gamma_{1,j}^2} (\tilde{T}_j - T_j^0) \Rightarrow \operatorname{argmax}_u \begin{cases} -\frac{|u|}{2} + B_j(u) & \text{for } u \leq 0 \\ -\frac{|u|}{2} \frac{\Delta_{2,j}}{\Delta_{1,j}} + \frac{\Gamma_{2,j}}{\Gamma_{1,j}} B_j(u) & \text{for } u > 0 \end{cases} \quad (10)$$

where

$$\begin{aligned} \Delta_{1,j} &= \frac{1}{2} \operatorname{tr}(A_{1,j}^2) + \delta_j' Q_{1,j} \delta_j \\ \Delta_{2,j} &= \frac{1}{2} \operatorname{tr}(A_{2,j}^2) + \delta_j' Q_{2,j} \delta_j \\ \Gamma_{1,j} &= (\operatorname{vec}(A_{1,j})' \Omega_{1,j}^0 \operatorname{vec}(A_{1,j}) / 4 + \delta_j' \Pi_{1,j} \delta_j)^{1/2} \\ \Gamma_{2,j} &= (\operatorname{vec}(A_{2,j})' \Omega_{2,j}^0 \operatorname{vec}(A_{2,j}) / 4 + \delta_j' \Pi_{2,j} \delta_j)^{1/2} \\ A_{1,j} &= (\Sigma_j^0)^{1/2} (\Sigma_{j+1}^0)^{-1} \Phi_j (\Sigma_j^0)^{-1/2} \\ A_{2,j} &= (\Sigma_{j+1}^0)^{1/2} (\Sigma_j^0)^{-1} \Phi_j (\Sigma_{j+1}^0)^{-1/2} \end{aligned}$$

$$Q_{1,j} = \text{plim}_{T \rightarrow \infty} (T_j^0 - T_{j-1}^0)^{-1} \sum_{t=T_{j-1}^0+1}^{T_j^0} x_t(\Sigma_{j+1}^0)^{-1} x_t'$$

$$Q_{2,j} = \text{plim}_{T \rightarrow \infty} (T_{j+1}^0 - T_j^0)^{-1} \sum_{t=T_j^0+1}^{T_{j+1}^0} x_t(\Sigma_j^0)^{-1} x_t'$$

and

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

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

with $B_j(s)$ a Wiener process defined on the real line and

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

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

Note that our result allows both the marginal distribution of the regressors and of the errors to vary across regimes. Also, the error process is allowed to be autocorrelated as well as conditionally heteroskedastic. Hence, it extends the results of Bai (2000) and Bai, Lumsdaine and Stock (1998).

Remark 5 *The specific form of $\Omega_{1,j}^0$ and $\Omega_{2,j}^0$ depends on some features of the distribution of the errors. When these are identically Normally distributed, $\eta_t \eta_t'$ has a Wishart distribution with*

$$\text{var}(\text{vec}(\eta_t \eta_t')) = (I_{n^2} + K_n)$$

where K_n is the commutation matrix, i.e., the $n^2 \times n^2$ matrix such that

$$K_n \text{vec}(A) = \text{vec}(A')$$

for any $n \times n$ matrix A (see, e.g., Magnus, 1988, p. 35 and p. 164). In this case, if $\eta_t \eta_t'$ are serially uncorrelated, then

$$\Omega_{1,j}^0 = \Omega_{2,j}^0 = \Omega^0 = (I_{n^2} + K_n) \tag{11}$$

Further, using the fact that K_n is an idempotent matrix,

$$\begin{aligned}
\text{vec}(A_{1,j})' \Omega^0 \text{vec}(A_{1,j}) / 4 &= \text{vec}(A_{1,j})' (I_{n^2} + K_n) \text{vec}(A_{1,j}) / 4 \\
&= [\text{vec}(A_{1,j})' \text{vec}(A_{1,j}) + \text{vec}(A'_{1,j})' \text{vec}(A'_{1,j})] / 4 \\
&= \text{vec}(A_{1,j})' \text{vec}(A_{1,j}) / 2
\end{aligned}$$

and $\Pi_{1,j}$ and $\Pi_{2,j}$ simplify accordingly.

Remark 6 Under shrinking magnitude of shifts, $\Sigma_j^0 \rightarrow \Sigma^0$ and $\Omega_j^0 \rightarrow \Omega^0$ as $T \rightarrow \infty$, for $j = 1, \dots, m+1$. So, in principle, the Σ_j^0 and Ω_j^0 in the limiting formula could be replaced with Σ^0 and Ω^0 . However, we keep the original formula since it is expected to yield a better approximation when the change in the variance is not small, and remains valid when the change is small.

The analytical formula for the cumulative distribution function of (10) has been derived by Bai (1997). Hence, the relevant quantiles could be easily calculated if we knew the coefficients affecting the distribution. In practice, the true values of the coefficients are unknown. However, they can be consistently estimated and it is easy to show that the coverage rates will be asymptotically valid provided root- T consistent estimates are used instead of the true values. Then the distribution can be simulated and confidence intervals can be constructed using the empirical distribution. The various quantities can be estimated as follows

$$\begin{aligned}
v_T^2 \tilde{\Delta}_{1,j} &= \text{tr}(v_T^2 \tilde{A}_{1,j}^2) / 2 + (\tilde{\beta}_{j+1} - \tilde{\beta}_j)' \tilde{Q}_{1,j} (\tilde{\beta}_{j+1} - \tilde{\beta}_j) \\
v_T^2 \tilde{\Delta}_{2,j} &= \text{tr}(v_T^2 \tilde{A}_{2,j}^2) / 2 + (\tilde{\beta}_{j+1} - \tilde{\beta}_j)' \tilde{Q}_{2,j} (\tilde{\beta}_{j+1} - \tilde{\beta}_j) \\
v_T^2 \tilde{\Gamma}_{1,j} &= [\text{vec}(v_T \tilde{A}_{1,j})' \tilde{\Omega}_{1,j}^0 \text{vec}(v_T \tilde{A}_{1,j}) / 4 + (\tilde{\beta}_{j+1} - \tilde{\beta}_j)' \tilde{\Pi}_{1,j} (\tilde{\beta}_{j+1} - \tilde{\beta}_j)]^{1/2} \\
v_T^2 \tilde{\Gamma}_{2,j} &= [\text{vec}(v_T \tilde{A}_{2,j})' \tilde{\Omega}_{2,j}^0 \text{vec}(v_T \tilde{A}_{2,j}) / 4 + (\tilde{\beta}_{j+1} - \tilde{\beta}_j)' \tilde{\Pi}_{2,j} (\tilde{\beta}_{j+1} - \tilde{\beta}_j)]^{1/2}
\end{aligned}$$

where

$$\begin{aligned}
v_T \tilde{A}_{1,j} &= (\tilde{\Sigma}_j)^{1/2} (\tilde{\Sigma}_{j+1})^{-1} (\tilde{\Sigma}_{j+1} - \tilde{\Sigma}_j) (\tilde{\Sigma}_j)^{-1/2} \\
v_T \tilde{A}_{2,j} &= (\tilde{\Sigma}_{j+1})^{1/2} (\tilde{\Sigma}_j)^{-1} (\tilde{\Sigma}_{j+1} - \tilde{\Sigma}_j) (\tilde{\Sigma}_{j+1})^{-1/2} \\
\tilde{Q}_{1,j} &= (\tilde{T}_j - \tilde{T}_{j-1})^{-1} \sum_{t=\tilde{T}_{j-1}+1}^{\tilde{T}_j} x_t (\tilde{\Sigma}_{j+1})^{-1} x_t' \\
\tilde{Q}_{2,j} &= (\tilde{T}_{j+1} - \tilde{T}_j)^{-1} \sum_{t=\tilde{T}_j+1}^{\tilde{T}_{j+1}} x_t (\tilde{\Sigma}_j)^{-1} x_t'
\end{aligned}$$

and $\tilde{\Pi}_{1,j}, \tilde{\Pi}_{2,j}, \tilde{\Omega}_{1,j}^0, \tilde{\Omega}_{2,j}^0$, the estimates of the long run variance of the corresponding quantities, can be constructed using a method based on a weighted sum of sample autocovariances of the relevant quantities, as discussed in, e.g., Andrews (1991).

Our result is framed to encompass the most general cases. In many instances, however, the limiting formulae and the estimation simplify substantially. For example, if there are only changes in the conditional mean, then all the terms involving the difference in the covariance matrices drop. If, in addition, the distribution of the regressors is assumed to be stable, then

$$Q_{1,j} = Q_{2,j} = \text{plim}_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T x_t (\Sigma^0)^{-1} x_t'$$

for all $j = 1, \dots, m + 1$. Also, if the error process is assumed to form a martingale difference sequence, as is often the case in a dynamic model, then

$$\begin{aligned} \Pi_{1,j} &= \text{plim}_{T \rightarrow \infty} (T_j^0 - T_{j-1}^0)^{-1} \left(\sum_{t=T_{j-1}^0+1}^{T_j^0} x_t (\Sigma_{j+1}^0)^{-1} (\Sigma_j^0) (\Sigma_{j+1}^0)^{-1} x_t' \right) \\ \Pi_{2,j} &= \text{plim}_{T \rightarrow \infty} (T_{j+1}^0 - T_j^0)^{-1} \left(\sum_{t=T_j^0+1}^{T_{j+1}^0} x_t (\Sigma_j^0)^{-1} (\Sigma_{j+1}^0) (\Sigma_j^0)^{-1} x_t' \right) \end{aligned}$$

and, hence, no special estimate is needed to construct the above quantities.

Though only root- 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 even though imposing restrictions does not have a first-order effect on the limiting distribution of the estimates of the break dates.

3.1 Block-partial structural break models

It is of interest to look more deeply into the case where a subset of the equations in the system is restricted not to have any break, i.e., the restrictions imposed are such that all parameters have equal value for all segments. We call this a block-partial structural break model since only a particular block of equations is allowed to be influenced by breaks. A natural question to ask is whether there is any gain, pertaining to the estimates of the break dates, made possible by the inclusion of equations with no break. Our results show that, indeed, important gains are possible depending on the extent of the correlation between the errors in the two blocks of equations. The intuitive explanation is that the precision in the

estimates of the break dates affects the likelihood function through both the variance of that block of equations and the correlation with the other equations. Hence, by including such correlated equations without breaks, additional forces are in play to better pinpoint the break dates.

To make these assertions more precise, suppose we have the following block structure

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = (I_n \otimes z'_t) \begin{pmatrix} \beta_{1t} \\ \beta_2 \end{pmatrix} + u_t$$

where y_{1t} and y_{2t} are vectors of dimensions n_1 and n_2 , respectively ($n_1 + n_2 = n$). Hence, all equations share the same regressors but the coefficients are allowed to change only in the first block of equations. We also assume for simplicity that the distribution of the regressors does not vary across regimes and the covariance matrix of the errors is stable. Denote the asymptotic variance of the estimate of the j^{th} break obtained using the full system as V_j^f , and that obtained from the subset involving only the first n_1 equations as V_j^p . We then have the following relative asymptotic efficiency

$$V_j^f/V_j^p = \left(\frac{\Delta\beta'_{1,j}(\Sigma_{11}^{-1} \otimes Q)\Delta\beta_{1,j}}{\Delta\beta'_{1,j}((\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma'_{12})^{-1} \otimes Q)\Delta\beta_{1,j}} \right)^2 \leq 1 \quad (12)$$

where $Q = \text{plim}_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T z_t z'_t$. This inequality follows since $\Sigma_{12}\Sigma_{22}^{-1}\Sigma'_{12}$ is positive semi-definite. To illustrate the potential gain from using the full system, consider the following simple case with two equations

$$y_{1t} = a_1 1(t \leq T_1) + (a_1 + \lambda) 1(t > T_1) + u_{1t} \quad (13)$$

$$y_{2t} = a_2 + u_{2t} \quad (14)$$

with

$$\begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \sim i.i.d. \left(0, \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix} \right)$$

Here, there is one shift in mean in the first equation and the second equation is stable. From (12), the relative efficiency of using both equations as the system, over using only the first, is $1/(1 - r_{12}^2)$ with r_{12} the correlation coefficient between u_{1t} and u_{2t} . When $r_{12} = 0.5$, we have, in large samples, a reduction in variance of 44% using both equations. When $r_{12} = 0.8$, this reduction reaches 87%. Hence, the gains of using a multivariate system can indeed be quite large.

The fact that correlation in the errors of different equations induces an increase in efficiency when some of the equations have no break (or more generally distinct breaks) is in contrast to the results of Bai, Lumsdaine and Stock (1998). They show that with common breaks the precision increases with the number of equations. But in this case positive correlation in the error may increase the variance and mitigate the beneficial effect of having more equations. The intuition is that with positive correlation in the errors, the additional information about the common breaks becomes less independent, more of a duplicate of the same information. For example, consider two equations with a single common break where y_{1t} is generated by (13) with $\lambda = \sigma_1^2 = \sigma_2^2$, and y_{2t} is generated by the same process with u_{2t} instead of u_{1t} . Then the relative efficiency of using both equations as the system, over using only the first, is $2/(1+r_{12})$. Hence, $r_{12} > 0$ induces a loss of efficiency, though no value can eliminate completely the benefits of including the additional equation. Note that, in the block partial breaks model, the sign of the correlation plays no role in affecting the efficiency. This is not surprising since the break structure of the two sets of equations are different and the positive correlation does not lead to a duplicate of the same information. This insight should carry through to other type of models, such as ordered breaks models, where the breaks are not simultaneous and correlation in the errors is present.

4 An efficient algorithm for estimation

We now discuss how to construct the QMLE based on Normal serially uncorrelated errors. The computation of the estimates in the general framework considered is not a trivial issue. In principle, a grid search can be used but it becomes rapidly impractical since it involves the computation of maximum likelihood estimates of order $O(T^m)$. Our approach is to extend the work of Hawkins (1976) and Bai and Perron (2003a) who advocate a dynamic programming algorithm. The basic idea is as follows. With any possible number of breaks, it is the case that the overall value of the log likelihood function is the sum of the values associated with a particular combination of $m+1$ segments. Hence, if we have the information about the log likelihood values for all possible segments, of which there are at most $T(T+1)/2$, then all that is needed is a method to assess which particular combination of $m+1$ segments leads to the highest value of the likelihood function. This is achieved using a dynamic programming algorithm. The relevant steps are discussed in this section, though we omit details that are discussed more thoroughly in Bai and Perron (2003a).

4.1 The optimization problem without restrictions

Consider first the case with no restriction since then the value of the log likelihood function for different segments can be obtained by estimating the model with data segment by segment. Denote the possible segments by pairs (i, j) with i the starting date and j the ending date. Let $h = \varepsilon T$ be the minimum length of a segment, then i ranges from 1 to $T - h + 1$ and j ranges from h to T (some entries would be redundant but to facilitate the discussion we consider this range, see Bai and Perron, 2003a, for details). In what follows a subscript (i, j) denotes the value of the relevant variable over the segment (i, j) ; for example $Y_{(i,j)}$ is the vector of dependent variables from time i to time j , $X_{(i,j)}$ is the corresponding matrix of regressors and $\Sigma_{(i,j)}$ is the covariance matrix of $U_{(i,j)}$.

Suppose first that $\Sigma_{(i,j)}$ is known. Then the QMLE is equivalent to the GLS estimate and

$$\hat{\beta}_{(i,j)}^{GLS} = \left(X'_{(i,j)} \Sigma_{(i,j)}^{-1} X_{(i,j)} \right)^{-1} X'_{(i,j)} \Sigma_{(i,j)}^{-1} Y_{(i,j)} \quad (15)$$

An estimate that is asymptotically equivalent to the QMLE can be obtained using an iterative procedure that starts with the OLS estimates of $\beta_{(i,j)}$ to get an estimate of $\Sigma_{(i,j)}$. One then substitute in (15) to get a new estimate of $\beta_{(i,j)}$, and iterate until convergence. Denote the resulting estimate by $\hat{\beta}_{(i,j)}$. The log likelihood value over the segment (i, j) is then

$$l(i, j) = -\frac{j-i+1}{2} \left\{ (\log(2\pi) + 1) + \log \det \left(\frac{1}{(j-i+1)} \sum_{t=i}^j \left[y_t - x'_t \hat{\beta}_{(i,j)} \right] \left[y_t - x'_t \hat{\beta}_{(i,j)} \right]' \right) \right\}$$

An efficient way to compute the log likelihood values for the segments (i, j) with i fixed and $j = i + h - 1, \dots, T$ is to use a recursive residuals approach. The relevant updating formulae for the multivariate case have been derived by Tobing and McGilchrist (1992). These are, for a known $\Sigma_{(j+1,j+1)}$,

$$\begin{aligned} \hat{\beta}_{(i,j+1)}^{GLS} &= \hat{\beta}_{(i,j)}^{GLS} + H_{(i,j)}^{-1} X'_{(j+1,j+1)} \left(\Sigma_{(j+1,j+1)} + X_{(j+1,j+1)} H_{(i,j)} X'_{(j+1,j+1)} \right)^{-1} \\ &\quad \times [Y_{(j+1,j+1)} - X_{(j+1,j+1)} \hat{\beta}_{(i,j)}^{GLS}] \end{aligned} \quad (16)$$

and

$$\begin{aligned} H_{(i,j+1)} &= H_{(i,j)} + X'_{(j+1,j+1)} \Sigma_{(j+1,j+1)} X_{(j+1,j+1)} \\ H_{(i,j+1)}^{-1} &= H_{(i,j)}^{-1} - H_{(i,j)}^{-1} X'_{(j+1,j+1)} \left(\Sigma_{(j+1,j+1)} + X_{(j+1,j+1)} H_{(i,j)}^{-1} X'_{(j+1,j+1)} \right)^{-1} X_{(j+1,j+1)} H_{(i,j)}^{-1} \end{aligned}$$

Since $\Sigma_{(j+1,j+1)}$ is unknown, we must iterate on the updating formula (16) starting with

$$\Sigma_{(j+1,j+1)} = (j-i+1)^{-1} [Y_{(i,j)} - X_{(i,j)} \hat{\beta}_{(i,j)}^{GLS}] [Y_{(i,j)} - X_{(i,j)} \hat{\beta}_{(i,j)}^{GLS}]'$$

Upon convergence, we obtain the approximate QMLE $\hat{\beta}_{(i,j+1)}^{GLS}$ and $\hat{\Sigma}_{(i,j+1)}$.

To obtain $\hat{\beta}_{(i,j+1)}$ for $j = i + h, \dots, T - 1$, we use this updating scheme starting with $\hat{\beta}_{(i,i+h-1)}^{GLS}$ and

$$H_{(i,i+h-1)} = X_{(i,i+h-1)} \hat{\Sigma}_{(i,i+h-1)}^{-1} X'_{(i,i+h-1)}$$

with $\hat{\beta}_{(i,i+h-1)}^{GLS}$ and $\hat{\Sigma}_{(i,i+h-1)}^{-1}$ obtained from (15) iterating until convergence. Repeating this process for $i = 2, \dots, T - h + 1$, one obtains the triangular matrix of log likelihood values $l(i, j)$ for $i = 1, \dots, T - h + 1$ and $j = i + h - 1, \dots, T$. Note that the computations involve only inversions of $h \times h$ and $n \times n$ matrices.

Once the triangular matrix of likelihood values is obtained, the next step is to find which combination of $m + 1$ segments has the largest sum, i.e., which partition $\mathcal{T} = (T_1, \dots, T_m)$ is optimal. This is achieved using a dynamic programming algorithm. Denote by $l(\{T_{r,n}\})$ the value of the log likelihood function with the optimal partition associated with a system having r breaks and estimated using the first n observations. Then, the optimal partition for the system with m breaks using the full set of observations is the solution to the following dynamic programming problem (see Bai and Perron, 2003a, for details):

$$l(\{T_{m,T}\}) = \max_{mh \leq j \leq T-h} [l(\{T_{m-1,j}\}) + l(j + 1, T)]$$

Note that the two layers of recursions allow us to solve two problems: the inversion of large matrices and the use of non-linear algorithms to obtain the *MLE*. Convergence of the iterative GLS scheme to obtain the *MLE* should be fast since the initial estimate of the covariance matrix of the errors is that obtained with one less observation. Hence, generally, convergence only takes a few iterations.

4.2 The procedure allowing for restrictions on the coefficients

With restrictions on the coefficients, the algorithm described above is no longer sufficient since the triangular matrix of log likelihood values cannot be constructed using estimation segment by segment. Indeed, the restricted estimates of the parameters in one segment depend on those in the optimal partition that includes that segment, each of which would have an optimal partition that need not include the original one. Fortunately, there is a simple iterative scheme using the same principles outlined above that can yield the desired result. The idea is to first estimate the breaks and the coefficients without restrictions. Then, conditional on the estimated breaks, use all the restrictions to estimate the coefficients. Third, use the updated coefficients to repartition the sample. Finally, repeat the process until convergence. More specifically, the algorithm is as follows.

1. Use the dynamic programming algorithm described above to estimate an unrestricted model with m breaks.
2. Obtain the restricted QMLE of the coefficients conditional on the break dates obtained in part 1. Denote the estimates for each segment as $\tilde{\beta}_j$ and $\tilde{\Sigma}_j$ ($j = 1, \dots, m+1$), and the resulting value of the restricted log likelihood function for a segment (i, j) by $rl^r(i, j)$ when estimates from the r^{th} regime are used, i.e.,

$$rl^r(i, j) = -\frac{j-i+1}{2}[(\log(2\pi) + 1) + \log|\tilde{\Sigma}_r|]$$

3. Compute and store the values of the break dates and the restricted likelihood function corresponding to optimal one break partitions, $rl(\{T_{1,n}\})$, for $2h \leq n \leq T - (m-1)h$. This is done solving the following recursive problem

$$rl(\{T_{1,n}\}) = \max_{h \leq j \leq n-h} [rl^1(1, j) + rl^2(j+1, n)]$$

Then, sequentially compute and store $rl(\{T_{r,n}\})$ for $r = 2, \dots, m-1$, with n ranging from $(r+1)h$ to $T - (m-r)h$. This is done by solving the recursive problem

$$rl(\{T_{r,n}\}) = \max_{rh \leq j \leq n-h} [rl(\{T_{r-1,j}\}) + rl^{r+1}(j+1, n)]$$

Finally compute

$$rl(\{T_{m,T}\}) = \max_{mh \leq j \leq T-h} [rl(\{T_{m-1,j}\}) + rl^{m+1}(j+1, T)]$$

and store the m estimated break dates.

4. Repeat steps 2 and 3 until convergence.

The main computation involved in this procedure is in step 1. The reason is that the estimates of the break dates obtained from the unrestricted model are not only consistent but also asymptotically as efficient compared to estimates obtained imposing the restrictions. This ensures that the estimates of the coefficients in Step 2 are not only \sqrt{T} convergent, but also asymptotically efficient. Hence, convergence is typically fast and simulations have shown that only a few iterations are needed, in most cases.

5 Locally ordered breaks model

In the class of models considered so far, the breaks across equations either happen simultaneously or are distant from each other in the sense that they are separated by a positive fraction of the sample. Accordingly, the estimates of the break dates can be treated independently. This rules out a class of models which may have wide appeal in practice, when breaks across equations are close to each other and, hence, cannot be treated independently. As mentioned in the introduction, we label such cases as “locally ordered breaks” for which the methodology discussed so far does not provide an appropriate framework for estimation and inference. Estimating the model assuming common breaks is not desirable and we should not treat the breaks in the two equations separately since this would imply a loss of efficiency. Our goal is to discuss an algorithm to estimate break dates in such models and construct confidence intervals for the estimates.

To keep things manageable, yet covering most cases of potential applications, we consider a simplified class of models containing two subsets of equations:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = x'_t \begin{pmatrix} \beta_{1t} \\ \beta_{2t} \end{pmatrix} + U_t$$

and the total number of breaks in the system is again m . Restrictions of the form (2) can still be imposed, which in particular can allow some coefficients in each equation not to change across regimes, and so on. But we suppose that all breaks are either common or asymptotically distinct except for the j^{th} one, which is locally ordered in the system. Without loss of generality, we assume that the j^{th} break occurs first in the first subset of equations, at some date $k_{1,j}^0$, and after in the second set, at some date $k_{2,j}^0 \geq k_{1,j}^0$. We now define precisely what is meant by locally ordered breaks.

Definition 1 (*Locally ordered breaks*) Let v_T be a sequence of positive numbers satisfying $v_T \rightarrow 0$ and $T^{1/2}v_T/(\log T)^2 \rightarrow \infty$, breaks across equations occurring at dates $k_{1,j}^0$ and $k_{2,j}^0$ are said to be locally ordered if $k_{1,j}^0 \leq k_{2,j}^0$ and $v_T^2(k_{2,j}^0 - k_{1,j}^0) \leq M_T$ with $M_T \rightarrow 0$ as $T \rightarrow \infty$.

Remark 7 The condition $v_T^2(k_{2,j}^0 - k_{1,j}^0) \rightarrow 0$ implies that $(k_{2,j}^0 - k_{1,j}^0)/T \rightarrow 0$, although $(k_{2,j}^0 - k_{1,j}^0)$ itself is not restricted to be finite as T increases. If $\lim_{T \rightarrow \infty} (k_{2,j}^0 - k_{1,j}^0)/T > 0$, then the two breaks become asymptotically distinct and the usual asymptotic framework discussed above applies. What is important is that asymptotically the distance between the two breaks becomes a negligible proportion of the sample size at a fast enough rate dictated by the condition on the rate of decrease of v_T .

We also impose the following assumption.

- Assumption LOB: a) Assumptions A1 to A10 continue to hold for the all breaks except that A4 is strengthened to require martingale difference errors and for the j^{th} break date A7 is replaced by Definition 1; b) it is also assumed that the covariance matrix of the errors Σ is not changing at the j^{th} break.

5.1 The estimation algorithm

We first explain how to estimate the model using an extension of the algorithm discussed in Section 4. The steps are as follows.

1. Use the dynamic programming algorithm to estimate an unrestricted model. This means that we treat the two sets of equations separately when estimating the breaks, hence information regarding the ordering of the breaks is not used. The estimates will be consistent but not efficient. Store the triangular array of estimated break dates.
2. Estimate the coefficients conditional on the break dates obtained in part 1, imposing restrictions if any. Note that the estimates of the coefficients are asymptotically efficient even though we have not imposed the ordered restrictions on the breaks.
3. Denote the value of the likelihood function for the segment between the $(j-1)^{\text{th}}$ break and $k_{2,j}$, say $(T_{j-1}, k_{2,j})$, which contains a break in the first system occurring at date $k_{1,j}$, as $rl(T_{j-1}, k_{2,j}(k_{1,j}))$. Compute and save

$$rl(T_{j-1}, k_{2,j}) = \sup_{T_{j-1}+h \leq k_{1,j} \leq k_{2,j}} rl(T_{j-1}, k_{2,j}(k_{1,j}))$$

for all $k_{2,j} \leq T - (m-j)h$, where $h = [\varepsilon T]$ is again the minimal length of segments defined by asymptotically distinct break dates.

4. Find the combination of the segments that maximizes the global likelihood function. To do this, the dynamic programming algorithm is used.
5. Repeat step 2 to 4 until convergence.

Remark 8 *We have only considered a class of models with two ordered breaks occurring in two subsets of equations. The estimation algorithm can easily be extended to allow for, say, d such locally ordered breaks occurring in d equations or subsets of equations, i.e., where Definition 1 is extended to have $k_{1,j}^0 \leq k_{2,j}^0 \leq \dots \leq k_{d,j}^0$ with the differences $v_T^2(k_{s,j}^0 - k_{1,j}^0) \rightarrow 0$ for $s = 2, \dots, d$. Only step 3 needs to be modified accordingly.*

5.2 The limiting distribution

We now consider the limiting distribution of the break dates. Note that we only need a separate treatment for the j^{th} break which is locally ordered in the system since all others are asymptotically distinct and the distribution theory discussed in Section 3 applies. Denote the estimates obtained using the algorithm described above as $\hat{k}_{1,j}$ and $\hat{k}_{2,j}$, we have the following result regarding the consistency and rate of convergence of the estimates of the break dates.

Lemma 2 *Under Assumption LOB, with $\tilde{k}_{1,j}$ and $\tilde{k}_{2,j}$ the corresponding estimates, we have for all $i = 1, \dots, j-1, j+1, \dots, m$: $v_T^2(\tilde{T}_i - T_i^0) = O_p(1)$, $v_T^2(\tilde{k}_{1,j} - k_{1,j}^0) = O_p(1)$, $v_T^2(\tilde{k}_{2,j} - k_{2,j}^0) = O_p(1)$.*

Lemma 2 shows that the rate of convergence of the break dates are the same as stated in Lemma 1. Hence the relevant expressions for the likelihood function correspond that of Theorem 1 with $m+1$ breaks, the j^{th} and the $(j+1)^{\text{th}}$ being locally ordered ones. Then, we need to examine $lr_j^1(r_1) + lr_{j+1}^1(r_2)$ in order to derive the joint limiting distribution for the pair $r_1 = \hat{k}_{1,j} - k_{1,j}^0$ and $r_2 = \hat{k}_{2,j} - k_{2,j}^0$. To illustrate how the derivation proceeds, consider the case where $r_1 < 0$ and $r_2 > 0$. Under the assumption LOB, the covariance matrices Σ_j^0 and Σ_{j+1}^0 are the same with a common value, denoted $\Sigma_{j,j+1}^0$, and we have

$$\begin{aligned} & lr_j^1(r_1) + lr_{j+1}^1(r_2) \\ = & (\Delta\beta'_{1,j}, 0) \sum_{t=k_{1,j}^0+r_1}^{k_{1,j}^0} x_t(\Sigma_{j,j+1}^0)^{-1} u_t - \frac{1}{2} (\Delta\beta'_{1,j}, 0) \sum_{t=k_{1,j}^0+r_1}^{k_{1,j}^0} x_t(\Sigma_{j,j+1}^0)^{-1} x_t'(\Delta\beta'_{1,j}, 0)' \\ & - (0, \Delta\beta'_{2,j}) \sum_{t=k_{2,j}^0+r_2}^{k_{2,j}^0} x_t(\Sigma_{j,j+1}^0)^{-1} u_t - \frac{1}{2} (0, \Delta\beta'_{2,j}) \sum_{t=k_{2,j}^0+r_2}^{k_{2,j}^0} x_t(\Sigma_{j,j+1}^0)^{-1} x_t'(0, \Delta\beta'_{2,j})' \end{aligned}$$

Similar expressions obtain for the other possible configurations ($r_1 < 0, r_2 < 0, r_1 - r_2 \leq k_{2,j}^0 - k_{1,j}^0$, $r_1 \geq 0, r_2 \leq 0, r_1 - r_2 \leq k_{2,j}^0 - k_{1,j}^0$, and $r_1 > 0, r_2 > 0, r_1 - r_2 \leq k_{2,j}^0 - k_{1,j}^0$), and the limiting distribution, derived in details in the appendix, is given by

Theorem 4 *Under the Assumption LOB, we have*

$$v_T^2 \Pi_{1,j} \left((\hat{k}_{1,j} - k_{1,j}^0), (\hat{k}_{2,j} - k_{2,j}^0) \right) \Rightarrow \arg \max_{v_1 \leq v_2} Z(v_1, v_2)$$

with $Z(v_1, v_2)$ defined as follows: $Z(v_1, v_2) = 0$ if $v_1 = v_2 = 0$;

$$Z(v_1, v_2) = W_1(v_1) - \frac{|v_1|}{2} + (\Pi_{2,j}/\Pi_{1,j})^{1/2} W_2(v_2) - \frac{|v_2|}{2} (\Pi_{2,j} + 2\Pi_{12,j})/\Pi_{1,j}$$

if $v_1 < 0, v_2 < 0$ and $v_1 \leq v_2$;

$$Z(v_1, v_2) = W_1(v_1) - \frac{|v_1|}{2} + (\Pi_{2,j+1}/\Pi_{1,j})^{1/2}W_3(v_2) - \frac{|v_2|}{2}(\Pi_{2,j+1}/\Pi_{1,j})$$

if $v_1 < 0, v_2 > 0$; and

$$\begin{aligned} Z(v_1, v_2) &= (\Pi_{1,j+1}/\Pi_{1,j})^{1/2}W_4(v_1) - \frac{|v_1|}{2}(\Pi_{1,j+1} + 2\Pi_{12,j+1})/\Pi_{1,j} + (\Pi_{2,j+1}/\Pi_{1,j})^{1/2}W_3(v_2) \\ &\quad - \frac{|v_2|}{2}(\Pi_{2,j+1}/\Pi_{1,j}) \end{aligned}$$

if $v_1 > 0, v_2 > 0$ and $v_1 \leq v_2$. Here, $W_1(\cdot)$, $W_2(\cdot)$, $W_3(\cdot)$ and $W_4(\cdot)$ are independent Wiener processes defined on the real line and

$$Q_j = \text{plim}_{T \rightarrow \infty} (T_j^0 - T_{j-1}^0)^{-1} \sum_{t=T_{j-1}^0+1}^{T_j^0} x_t (\Sigma_{j,j+1}^0)^{-1} x_t'$$

$$Q_{j+1} = \text{plim}_{T \rightarrow \infty} (T_{j+1}^0 - T_j^0)^{-1} \sum_{t=T_j^0+1}^{T_{j+1}^0} x_t (\Sigma_{j,j+1}^0)^{-1} x_t'$$

$\delta'_{1,j} = v_T^{-1}(\beta_{1,j+1}^{0'} - \beta_{1,j}^{0'}, 0)$, $\delta'_{2,j} = v_T^{-1}(0, \beta_{2,j+1}^{0'} - \beta_{2,j}^{0'})$, $\Pi_{1,j} = \delta'_{1,j}Q_j\delta_{1,j}$, $\Pi_{12,j} = \delta'_{1,j}Q_j\delta_{2,j}$, $\Pi_{2,j} = \delta'_{2,j}Q_j\delta_{2,j}$, $\Pi_{1,j+1} = \delta'_{1,j}Q_{j+1}\delta_{1,j}$, $\Pi_{12,j+1} = \delta'_{1,j}Q_{j+1}\delta_{2,j}$ and $\Pi_{2,j+1} = \delta'_{2,j}Q_{j+1}\delta_{2,j}$.

The cumulative distribution function of the random variable $\arg \max_{v_1 \leq v_2} Z(v_1, v_2)$ does not have an tractable analytical formula. However, the relevant quantiles can be obtained using simulations. To do this, first generate the a realization of $Z(v_1, v_2)$ by replacing the true value of the parameters with consistent estimates and simulating the Brownian motion processes $W_1(\cdot)$, $W_2(\cdot)$, $W_3(\cdot)$ and $W_4(\cdot)$ over an reasonable range, say, $[-M, M]$. Then, apply the dynamic programming to find the global maximum of the function $Z(v_1, v_2)$ over $v_1, v_2 \in [-M, M]$ with the restriction $v_1 \leq v_2$. Finally, repeat the above two steps to obtain the relevant quantiles of the distribution.

6 Testing for structural change

We now consider testing for structural changes. Our setup is quite general in that we shall consider tests that allow for changes in the coefficients of the conditional mean or in the variance of the error term or both. Also, we can allow only a subset of coefficients to change across regimes, hence partial structural break and block partial structural break models are

permitted. We first consider using a likelihood ratio test for the null hypothesis of no change in any of the coefficients versus an alternative hypothesis with a pre-specified number of changes, say m .

In order to derive the limiting distribution of the test under the null hypothesis of no structural change, we impose the following additional assumptions on the data generating.

- Assumption A11: $T^{-1} \sum_{t=1}^{[Ts]} x_t x_t' \rightarrow_p sQ$, uniformly in $s \in [0, 1]$, for Q some positive definite matrix.
- Assumption A12: The errors $\{u_t\}$ form an array of martingale differences relative to $\mathcal{F}_t = \sigma\text{-field} \{ \dots, x_{t-1}, x_t, \dots, u_{t-2}, u_{t-1} \}$, and, additionally, $E(u_t u_t') = \Sigma^0$ for all t and $T^{-1/2} \sum_{t=1}^{[Ts]} x_t u_t \Rightarrow \Phi^{1/2} W(s)$, where $\Phi = \text{plim}_{T \rightarrow \infty} T^{-1} X' (I_n \otimes \Sigma^0) X$ and $W(s)$ is a vector of independent Wiener processes. Also, with $\eta_t \equiv (\eta_{t1}, \dots, \eta_{tn})' = (\Sigma^0)^{-1/2} u_t$, we have $T^{-1/2} \sum_{t=1}^{[Ts]} (\eta_t \eta_t' - I_n) \Rightarrow \xi(s)$ where $\xi(s)$ is an $n \times n$ matrix of Brownian motion processes with $\Omega = \text{var}(\text{vec}(\xi(1)))$.

Assumption A11 rules out trending regressors and imposes the requirement that the limit moment matrix of the regressors be homogeneous throughout the sample. Hence, we avoid the case where the marginal distribution of the regressors may change while the coefficients do not (see, e.g., Hansen, 2000). This follows from our basic premise that regimes are defined by changes in some coefficients. Similarly, Assumption A12 rules out instability in the error process and states that a basic functional central limit theorem holds for the weighted partial sums of the errors and their products. Note that A12 assumes no serial correlation in the errors u_t . This will be relaxed later.

6.1 The specification of the alternative hypothesis

We start with the case where, under the alternative hypothesis, we allow only a subset of the coefficients to change across regimes and there are otherwise no other restrictions. Hence, for a given equation, say the i^{th} one, and a given m -partition $\mathcal{T} = (T_1, \dots, T_m)$, we have

$$y_{it} = x'_{iat} \beta_{ia} + x'_{ibt} \beta_{ibj} + u_{it} \quad \text{for } T_{j-1} + 1 \leq t \leq T_j \quad (j = 1, \dots, m + 1)$$

where x_{iat} and β_{ia} are vectors of dimensions p_{ia} while x_{ibt} and β_{ibj} are vectors of dimensions p_{ib} , $p_{ia} + p_{ib} = p_i$, the number of included (or non-zero) regressors in the i^{th} equation. Hence, for this equation only p_{ib} coefficients are allowed to change. Note that this framework is such that any coefficient which is allowed to change does so simultaneously. The case where

different coefficients could change at different dates will be considered later. Stacking the system equation by equation, we have

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

where x_{at} and x_{bt} are $p_a \times n$ and $p_b \times n$, respectively, matrices of regressors, while β_a and β_{bj} are p_a and p_b dimensional vectors, respectively. Here, p_a is the total number of coefficients that are not subject to change, while p_b is the total number of coefficients that are allowed to change. Hence, x_{at} may contain zero elements corresponding to regressors that are allowed to change in one equation but not in at least one other. Similarly, x_{bt} may contain zero elements corresponding to regressors that are not allowed to change in one equation but are allowed to do so in at least one other. Note also that zero elements can occur simply because some regressors are altogether excluded from an equation.

The specification of the covariance matrix of the errors is such that

$$E(u_t u'_t) = \Sigma_j = \begin{pmatrix} \sigma_{11j} & \sigma_{12j} & \dots & \sigma_{1nj} \\ \sigma_{21j} & \sigma_{22j} & \dots & \sigma_{2nj} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1j} & \sigma_{n2j} & \dots & \sigma_{nnj} \end{pmatrix} \quad \text{for } T_{j-1} + 1 \leq t < T_j \quad (j = 1, \dots, m + 1)$$

where for, say, the i^{th} equation, we have $n - n_{bi}$ entries such that $\sigma_{ikj} = \sigma_{ik}$, i.e., we have n_{bi} entries whose coefficients are allowed to change. What will be important is the total number of coefficients that are allowed to change across regimes, which we denote by n_b ($\leq n(n + 1)/2$). As a matter of notation, let H be a full row rank matrix of dimension $n_b^* \times n^2$ such that $H\text{vec}(\Sigma)$ is the $n_b^* = n_{bd} + 2n_{bo}$ dimensional vector of the coefficients allowed to change (including both upper and lower triangle covariance entries). Here, n_{bd} is the number of diagonal entries of Σ (variances) allowed to change and n_{bo} is the number of covariances allowed to change (number of entries in the upper triangle of Σ). The total number of independent entries of Σ allowed to change is $n_b = n_{bd} + n_{bo}$. For example, in a two equation system with only the covariance allowed to change

$$H = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Under the null hypothesis of no change in structure the estimates are the values $\tilde{\beta}' = (\tilde{\beta}'_a, \tilde{\beta}'_b)$

and $\tilde{\Sigma}$ that jointly solve the following system of equations

$$\begin{aligned}\tilde{\Sigma} &= \frac{1}{T} \sum_{t=1}^T \left(y_t - x_t' \tilde{\beta} \right) \left(y_t - x_t' \tilde{\beta} \right)' \\ \tilde{\beta} &= \left(\sum_{t=1}^T x_t \tilde{\Sigma}^{-1} x_t' \right)^{-1} \sum_{t=1}^T x_t \tilde{\Sigma}^{-1} y_t\end{aligned}$$

with the resulting value of the log-likelihood function being

$$\log \tilde{L}_T = -\frac{T}{2} (\log 2\pi + 1) - \frac{T}{2} \log |\tilde{\Sigma}|.$$

For a given partition $\mathcal{T} = (T_1, \dots, T_m) = ([T\lambda_1], \dots, [T\lambda_m])$, the class of models described above can be estimated by quasi maximum likelihood using the appropriate restrictions. Denote the log-likelihood value by $\log \hat{L}_T(T_1, \dots, T_m)$. The test considered is the maximal value of the likelihood ratio test over all admissible partitions in the set Λ_ε defined by Assumption A9, i.e.,

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

where the estimates $(\hat{T}_1, \dots, \hat{T}_m)$ are the QMLE obtained considering only those partitions in Λ_ε . The parameter ε acts as a truncation which imposes a minimal length for each segment and will affect the limiting distribution of the test. It is also useful to describe the exact form of the log likelihood value and the estimates of the coefficients for some leading cases, when constructed using a given partition $\mathcal{T} = (T_1, \dots, T_m)$.

Example 1 (*Pure structural change model in the conditional mean*) When $p_{ai} = 0$ for all $i = 1, \dots, n$, and $n_b = 0$, we have the case of a pure structural change in the conditional mean for which the coefficients of every regressors in all equations are allowed to change across regimes. Then p_b is the number of regressors in the system whose coefficient is allowed to change. The log likelihood function under the alternative hypothesis is

$$\log \hat{L}_T(T_1, \dots, T_m) = -\frac{T}{2} (\log 2\pi + 1) - \frac{T}{2} \log |\hat{\Sigma}|$$

and the QMLE jointly solve the equations

$$\begin{aligned}\hat{\Sigma} &= \frac{1}{T} \sum_{j=1}^{m+1} \sum_{t=T_{j-1}+1}^{T_j} (y_t - x_t' \hat{\beta}_j)(y_t - x_t' \hat{\beta}_j)' \\ \hat{\beta}_j &= \left(\sum_{t=T_{j-1}+1}^{T_j} x_t \hat{\Sigma}^{-1} x_t' \right)^{-1} \sum_{t=T_{j-1}+1}^{T_j} x_t \hat{\Sigma}^{-1} y_t\end{aligned}$$

Example 2 (*Pure structural change model in the covariance matrix of the errors*) When $p_b = 0$ and $n_b = n(n+1)/2$, none of the coefficients of the conditional mean equations are allowed to change while all parameters of the covariance matrix of the errors are allowed to do so. The log likelihood function under the alternative hypothesis is

$$\log \hat{L}_T(T_1, \dots, T_m) = -\frac{T}{2} (\log 2\pi + 1) - \sum_{j=1}^{m+1} \frac{T_j - T_{j-1}}{2} \log |\hat{\Sigma}_j| \quad (17)$$

and the QMLE jointly solve the equations

$$\begin{aligned} \hat{\Sigma}_j &= \frac{1}{T_j - T_{j-1}} \sum_{t=T_{j-1}+1}^{T_j} (y_t - x_t' \hat{\beta})(y_t - x_t' \hat{\beta})' \\ \hat{\beta} &= \left(\sum_{j=1}^{m+1} \sum_{t=T_{j-1}+1}^{T_j} x_t \hat{\Sigma}_j^{-1} x_t' \right)^{-1} \sum_{j=1}^{m+1} \sum_{t=T_{j-1}+1}^{T_j} x_t \hat{\Sigma}_j^{-1} y_t \end{aligned}$$

Example 3 (*Complete pure structural change model*) When $p_{ai} = 0$ for all $i = 1, \dots, n$, and $n_b = n(n+1)/2$, all coefficients are allowed to change. The log likelihood function under the alternative hypothesis is still (17) with the QMLE jointly solving the equations

$$\begin{aligned} \hat{\Sigma}_j &= \frac{1}{T_j - T_{j-1}} \sum_{t=T_{j-1}+1}^{T_j} (y_t - x_t' \hat{\beta}_j)(y_t - x_t' \hat{\beta}_j)' \\ \hat{\beta}_j &= \left(\sum_{t=T_{j-1}+1}^{T_j} x_t \hat{\Sigma}_j^{-1} x_t' \right)^{-1} \sum_{t=T_{j-1}+1}^{T_j} x_t \hat{\Sigma}_j^{-1} y_t \end{aligned}$$

6.2 The limiting distribution of the test.

We now consider the limiting distribution of the $\sup LR_T(m, p_b, n_{bd}, n_{bo}, \varepsilon)$ test under the null hypothesis in the context of the class of models described in the previous subsection. The following result is proved in the appendix:

Theorem 5 *Under Assumptions A7 and A11-A12, with the $\sup LR_T(m, p_b, n_{bd}, n_{bo}, \varepsilon)$ test constructed for an alternative hypothesis in the class of models described in Section 6.1, we have, as $T \rightarrow \infty$,*

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

with

$$LR_j(\lambda, p_b, n_b^*) = \frac{\|\lambda_j W_{p_b}(\lambda_{j+1}) - \lambda_{j+1} W_{p_b}(\lambda_j)\|^2}{(\lambda_{j+1} - \lambda_j) \lambda_j \lambda_{j+1}} + \frac{1}{2} \frac{(\lambda_j W_{n_b^*}(\lambda_{j+1}) - \lambda_{j+1} W_{n_b^*}(\lambda_j))' H \Omega H' (\lambda_j W_{n_b^*}(\lambda_{j+1}) - \lambda_{j+1} W_{n_b^*}(\lambda_j))}{(\lambda_{j+1} - \lambda_j) \lambda_j \lambda_{j+1}}$$

where $\lambda = (\lambda_1, \dots, \lambda_m)$, $W_{p_b}(\cdot)$ and $W_{n_b^*}(\cdot)$ are p_b and $n_b^* = n_{bd} + 2n_{bo}$ dimensional vectors of independent Wiener processes, and $n_b^* = \text{rank}(H)$.

Note that the limiting distribution of the test statistic depends on a) p_b , the number of regressors whose coefficient is allowed to change; b) n_b^* the number of coefficients of the covariance matrix allowed to change; c) the matrix H that specifies which coefficients of the covariance matrix are allowed to change, and d) the distribution of the errors via the matrix Ω . In the analysis below, whenever an element of the covariance matrix of the errors is allowed to change, we shall assume Normality of the errors and use the specification (11) for Ω .

When no change in the covariance matrix of the errors is allowed, the limiting distribution of the test is:

$$\sup LR_T(m, p_b, 0, 0, \varepsilon) \Rightarrow \sup_{(\lambda_1, \dots, \lambda_m) \in \Lambda_\varepsilon} \sum_{j=1}^m \frac{\|\lambda_j W_{p_b}(\lambda_{j+1}) - \lambda_{j+1} W_{p_b}(\lambda_j)\|^2}{(\lambda_{j+1} - \lambda_j) \lambda_j \lambda_{j+1}} \quad (18)$$

which is of the same form as the expression in Proposition 6 of Bai and Perron (1998). It depends only on (m, p_b, ε) and tabulated critical values can be found in Bai and Perron (1998, 2003b).

When no change in the coefficients β is allowed and only elements of the covariance matrix are allowed to change the limiting distribution of the test is

$$\sup_{(\lambda_1, \dots, \lambda_m) \in \Lambda_\varepsilon} \frac{1}{2} \sum_{j=1}^m \frac{(\lambda_j W_{n_b^*}(\lambda_{j+1}) - \lambda_{j+1} W_{n_b^*}(\lambda_j))' H \Omega H' (\lambda_j W_{n_b^*}(\lambda_{j+1}) - \lambda_{j+1} W_{n_b^*}(\lambda_j))}{(\lambda_{j+1} - \lambda_j) \lambda_j \lambda_{j+1}}$$

and the distribution depends, in particular on the matrix H that specifies which coefficient of the covariance matrix is allowed to change. The distribution, however, simplifies considerably for cases of interest. To see this note that, since H is a selection matrix applied to $\text{vec}(\Sigma)$ any row that selects the (i, k) element of Σ can be written as an n^2 vector of the form $e'_{n,i} \otimes e'_{n,k}$ where $e_{n,i}$ is an $n \times 1$ vector with a 1 in the i^{th} position and 0 elsewhere. Hence, assuming Normality, any element of $H \Omega H'$ involving the selection of the (i, k) and (l, m) element of

Σ , can be written, in view of (11), as (with the second equality following from, e.g., Magnus, 1988, exercise 3.3).

$$\begin{aligned}
(H\Omega H')_{(i,k),(l,m)} &= (e'_{n,i} \otimes e'_{n,k}) (I_{n^2} + K_n) (e_{n,l} \otimes e_{n,m}) \\
&= (e'_{n,i} \otimes e'_{n,k}) (e_{n,l} \otimes e_{n,m}) + (e'_{n,i} \otimes e'_{n,k}) (e_{n,m} \otimes e_{n,l}) \\
&= \begin{cases} 2 & \text{if } i = k = l = m \\ 1 & \text{if } i = l \neq k = m \text{ or } i = m \neq k = l \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

This result greatly simplifies the form of the limiting distribution. In particular, the matrix $H\Omega H'$ is such that inference about changes in any one element of Σ is independent of changes in any other independent element (i.e., not the two entries for a covariance term). The value of an entry differs, however, when a variance or a covariance is allowed to change. Suppose that only n_b diagonal elements of Σ (i.e., variances) are allowed to change. Then $(H\Omega H') = 2I_{n_b}$ and the limiting distribution of $\sup LR_T(m, p_b, n_{db}, 0, \varepsilon)$ is also of the form (18) with n_b instead of p_b . When only n_b independent off-diagonal elements of Σ are allowed to change, $(H\Omega H') = ii'$ where i is a $2n_b \times 1$ vector of ones, i.e., $(H\Omega H')$ is a $2n_b \times 2n_b$ matrix of ones. Then, straightforward algebra reveals that the limiting distribution is still given by (18). Hence we have the following result under Normal errors.

Corollary 2 *With p_b elements of β and $n_b = n_{bd} + n_{bo}$ elements of Σ allowed to change, and under the conditions of Theorem 5 with the added assumption that the errors are Normally distributed, the limiting distribution of $\sup LR_T(m, p_b, n_{bd}, n_{bo}, \varepsilon)$ is*

$$\sup_{(\lambda_1, \dots, \lambda_m) \in \Lambda_\varepsilon} \sum_{j=1}^m \frac{\|\lambda_j W_{p_b+n_b}(\lambda_{j+1}) - \lambda_{j+1} W_{p_b+n_b}(\lambda_j)\|^2}{(\lambda_{j+1} - \lambda_j) \lambda_j \lambda_{j+1}}$$

which of the same form as the expression in Proposition 6 of Bai and Perron (1998). It depends only on m , $p_b + n_b$ and ε , where $n_b = n_{bd} + n_{bo}$, and tabulated critical values can be found in Bai and Perron (1998, 2003b).

6.3 Test of 1 versus 1+1 breaks

It is often the case that we do not know the number of changes in the system, and a statistical procedure to determine it is needed. For this purpose, information criteria such as those proposed by Liu, Wu and Zidek (1997) and Bai (2000) are possible. But as argued by

Perron (1997), these perform rather poorly, especially in models involving lagged dependent variables. Hence, it is useful to have a complementary test-based procedure. Following Bai and Perron (1998), we consider a sequential testing procedure based on the estimates of the break dates obtained from a global maximization of the likelihood function.

Consider a model with ℓ breaks, with estimated break dates denoted by $(\hat{T}_1, \dots, \hat{T}_\ell)$, which are obtained by a global maximization of the likelihood function. The procedure to test the null hypothesis of ℓ breaks versus the alternative hypothesis of $\ell + 1$ breaks is to perform a one break test for each of the $(\ell + 1)$ segments defined by the partition $(\hat{T}_1, \dots, \hat{T}_\ell)$ and to assess whether the maximum of the tests is significant. More precisely, the test is defined by

$$\sup SEQ_T(\ell + 1|\ell) = \max_{1 \leq j \leq \ell+1} \sup_{\tau \in \Lambda_{j,\varepsilon}} lr_T(\hat{T}_1, \dots, \hat{T}_{j-1}, \tau, \hat{T}_j, \dots, \hat{T}_\ell) - lr_T(\hat{T}_1, \dots, \hat{T}_\ell)$$

where

$$\Lambda_{j,\varepsilon} = \left\{ \tau; \hat{T}_{j-1} + (\hat{T}_j - \hat{T}_{j-1})\varepsilon \leq \tau \leq \hat{T}_j - (\hat{T}_j - \hat{T}_{j-1})\varepsilon \right\}$$

Note that this is different from a purely sequential procedure since for each value of ℓ the estimates of the break dates are re-estimated to get those that correspond to the global maximizers of the likelihood function. We have the following result for the limiting distribution of the test.

Theorem 6 *Under Assumptions A1-A9, A11-A12 and $m=\ell$, with the test $SEQ_T(\ell + 1|\ell)$ constructed for an alternative hypothesis in the class of models described in Section 6.1, we have, as $T \rightarrow \infty$,*

$$\lim_{T \rightarrow \infty} P(SEQ_T(\ell + 1|\ell) \leq x) = G_\varepsilon(x)^{\ell+1}$$

with $G_\varepsilon(x)$ the distribution function of $\sup_{(\lambda_1, \dots, \lambda_m) \in \Lambda_\varepsilon} LR_1(\lambda, p_b, n_b^*)$ defined in Theorem 5.

The proof is a straightforward extension of the proof of Theorem 5 and is therefore omitted. Note that the limiting distribution of the test simplifies as described above for the special cases considered.

6.4 Double maximum tests

As in Bai and Perron (1998), we also consider a test of the null hypothesis of no break versus the alternative hypothesis of some unknown number of breaks between 1 and some upper bound M . These are called double maximum tests since they are based on the maximum of the (possibly weighted) individual tests for the null of no break versus m breaks ($m = 1, \dots, M$). These are particularly useful to determine whether some structural change is

present since a sequential testing procedure can be unreliable for particular forms of multiple changes (see, Bai and Perron, 2004). More precisely, the test and its limiting distribution are given by

$$\begin{aligned} D \max LR_T(M) &= \max_{1 \leq m \leq M} a_m \sup LR_T(m, p_b, n_{bd}, n_{bo}, \varepsilon) \\ &\Rightarrow \max_{1 \leq m \leq M} a_m \sup_{(\lambda_1, \dots, \lambda_m) \in \Lambda_\varepsilon} \sum_{j=1}^m LR_j(\lambda, p_b, n_b^*) \end{aligned}$$

with $LR_j(\lambda, p_b, n_b^*)$ as defined in Theorem 5. We consider an equally weighted version defined by $a_m = 1$, denoted $UD \max LR_T(M)$, and a second version that applies weights to the individuals tests such that the marginal p-values are equal across values of m , denoted $WD \max LR_T(M)$. More precisely, $a_1 = 1$ and for $m > 1$, $a_m = c(\alpha, 1)/c(\alpha, m)$ where $c(\alpha, m)$ is the asymptotic critical value of the test $\sup LR_T(m, p_b, n_{bd}, n_{bo}, \varepsilon)$ at significance level α . Again, the limiting distribution of the test simplifies as described above for the special cases considered.

In practice, we suggest to use the following procedure to determine the number of structural breaks. First, use either $UD \max LR_T(M)$ or $WD \max LR_T(M)$ to test whether at least one break is present. If the test rejects, then apply the test $SEQ_T(\ell + 1|\ell)$ sequentially, for $\ell = 1, 2, \dots$, until the test fails to reject the null hypothesis of no additional structural break.

6.5 Extensions to serially correlated errors

When the error process u_t are serially correlated, the likelihood ratio type tests for changes in the coefficients of the conditional mean depend on nuisance parameters and would be hard to implement in practice. In such a case, structural changes in the regression coefficients can still be tested using Wald type statistics taking into account the presence of serial correlation. For example, the robust version corresponding to the test $\sup LR_T(m, p_b, 0, 0, \varepsilon)$ is

$$\sup F_T(m, p_b, \varepsilon) = \sup_{(\lambda_1, \dots, \lambda_m) \in \Lambda_\varepsilon} \left(\frac{T - (m + 1)p_b - p_a}{m} \right) \tilde{\beta}'_b R' \left(R \tilde{V}(\tilde{\beta}_b) R' \right)^{-1} R \tilde{\beta}_b \quad (19)$$

where $\tilde{\beta}_b$ is the quasi-likelihood estimate of the coefficients that are subject to change, under a given partition of the sample, R is the conventional matrix such that $(R\beta_b)' = (\beta'_{b,1} - \beta'_{b,2}, \dots, \beta'_{b,m} - \beta'_{b,m+1})$ and $\tilde{V}(\tilde{\beta}_b)$ is an estimate of the variance covariance matrix of $\tilde{\beta}_b$ that is robust to serial correlation and heteroskedasticity, i.e, a consistent estimate of

$$V(\tilde{\beta}_b) = \text{plim}_{T \rightarrow \infty} T^{-1} (\bar{X}_b^* \bar{X}_b^*)^{-1} \Omega_{\bar{X}_b^*} (\bar{X}_b^* \bar{X}_b^*)^{-1}$$

where $\bar{X}_b^* = (I_T \otimes \Sigma^0)^{-1/2} M_{X_a} \bar{X}_b$, $U_b^* = (I_T \otimes \Sigma^0)^{-1/2} M_{X_a} U$, $X_a = (x_{a1}, \dots, x_{aT})'$, $\bar{X}_b = \text{diag}(X_{b,1}, \dots, X_{b,m+1})$ where $X_{b,j}$ ($j = 1, \dots, m+1$) is the $n(T_j - T_{j-1})$ by p_b subset of X_b that corresponds to observations in regime j , $\Sigma^0 = \text{plim}_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T u_t u_t'$ and $\Omega_{\bar{X}_b^*} = E(\bar{X}_b^{*'} U_b^* U_b^{*' } \bar{X}_b^*)$.

Under Assumptions A1-A4, A10 and the additional assumptions under which a consistent estimate of $\Omega_{\bar{X}_b^*}$ can be obtained using kernel based method as in Andrews (1991), the limiting distribution of $\sup F_T(m, p_b, \varepsilon)$ is the same as in the case with martingale difference errors, i.e,

$$\sup F_T(m, p_b, \varepsilon) \Rightarrow \sup_{(\lambda_1, \dots, \lambda_m) \in \Lambda_\varepsilon} \sum_{j=1}^m \frac{\|\lambda_j W_{p_b}(\lambda_{j+1}) - \lambda_{j+1} W_{p_b}(\lambda_j)\|^2}{(\lambda_{j+1} - \lambda_j) \lambda_j \lambda_{j+1}}$$

where p_b denotes the total number of coefficients that are subject to change.

The robust version of $UD \max LR_T(M)$, $WD \max LR_T(M)$ and $SEQ_T(\ell + 1|\ell)$ can be constructed in a similar manner. In practice, the computation of the above tests could be very involved, especially if a data dependent method is used to construct the robust asymptotic covariance, $V(\tilde{\beta}_b)$. Following Bai and Perron (1998), we suggest first to use the dynamic programming algorithm to get the break points corresponding to the global maximization of the likelihood function, then plug the estimates into (19) to construct the test. This will not affect the consistency of the test since the break fractions are consistently estimated under the alternative.

6.6 Tests allowing for locally ordered breaks

Suppose that, in the testing procedure, we want to allow for locally ordered breaks as defined in Section 5. Consider the case where one of the m breaks under the alternative, the j^{th} one, is locally ordered across two subsets of equations. To allow for this the set of permissible partitions is no longer Λ_ε defined by (8). Let the value of the j^{th} break be $k_{1,j}$ for the first subset of equations and $k_{2,j}$ for the second, the permissible set is

$$\begin{aligned} \Lambda_\varepsilon^* &= \{(T_1, \dots, k_{1,j}, k_{2,j}, \dots, T_m); |T_{i+1} - T_i| \geq \varepsilon T \text{ for } i = 0, \dots, j-1, j+1, \dots, m; \\ &|T_{j\pm 1} - k_{1,j}| \geq \varepsilon T \text{ and } v_T^2(k_{2,j} - k_{1,j}) \leq M_T, \\ &\text{with } M_T \rightarrow 0, v_T \rightarrow 0 \text{ and } T^{1/2}v_T/(\log T)^2 \rightarrow \infty \text{ as } T \rightarrow \infty\} \end{aligned}$$

We then have the following result.

Theorem 7 *Let the $\sup LR_T(m, p_b, n_{bd}, n_{bo}, \varepsilon)$ test be constructed with estimates of the break dates restricted to be from the set of admissible partitions defined by Λ_ε^* . Its limiting distribution remains as stated in Theorem 5.*

This result has important practical implications. If it is believed that two subsets of equations share locally ordered breaks, one can then search over admissible partitions defined by the set Λ_ε^* , where a break can occur at slightly different dates across equations, and the limiting distribution of the test is still that corresponding to m distinct globally disjoint breaks. Hence, permitting breaks across equations to be locally ordered can lead to tests with increased power since under the alternative hypothesis the likelihood function will be increased.

Note that the same result holds true if more than one break are locally ordered across equations, or if the local ordering of a break applies to more than two subsets of equations. All that needs to be done is to change the permissible set of partitions accordingly. Also the same applies to the other tests discussed above.

6.7 Structural change tests with restrictions on the coefficients

Greater power can be achieved by imposing relevant restrictions on the coefficients when constructing structural change tests, as shown in Perron and Qu (2004) for the single equation case. In the multivariate regressions setup the problem is, however, substantially more complex. A comprehensive treatment of testing with general restrictions is beyond the scope of this paper and will be reported elsewhere. In the following, we discuss the test for two practically important special cases, namely globally ordered breaks and switching regimes.

Consider first the test for globally ordered breaks models in a system that includes two subsets of equations. We suppose that in the first subset p_{b1} coefficients are subject to a break followed by a break in p_{b2} coefficients in the second subset, but we remain agnostic about the time lag between the breaks. More specifically, we test the null hypothesis of no change against changes specified by the following admissible partitions of the sample,

$$\Lambda_\varepsilon^G = \{(\lambda_1, \lambda_2); \varepsilon \leq \lambda_1 \leq \lambda_2 \leq 1 - \varepsilon\}$$

where $\lambda_1 = k_1/T$ stands for the first break and $\lambda_2 = k_2/T$ for the second. When only changes in the regression coefficients are allowed, the test is

$$\sup LR_T^G(k_1, k_2, p_{b1}, p_{b2}, \varepsilon) = \sup_{(\lambda_1, \lambda_2) \in \Lambda_\varepsilon^G} [2 \log \hat{L}_T(k_1, k_2) - 2 \log \tilde{L}_T]$$

where $\log \tilde{L}_T$ denotes the maximized log likelihood function under the hypothesis of no break, and $\log \hat{L}_T(k_1, k_2)$ denotes the maximized log likelihood function under the alternative hypothesis, with restrictions of the type discussed in Section 6.1 imposed on the coefficients,

if applicable. Since no change in the covariance matrix of the errors is allowed, the latter is estimated using the full sample. The following Theorem, proved in the appendix, states the limiting distribution of the test.

Theorem 8 *Let $W_{p_{b1}}(\cdot)$ and $W_{p_{b2}}(\cdot)$ be p_{b1} and p_{b2} vectors of independent Wiener processes on $[0, 1]$. Then under Assumptions A11-A12,*

$$\sup LR_T^G \Rightarrow \text{Sup}_{(\lambda_1, \lambda_2) \in \Lambda_\varepsilon} \frac{\|\lambda_1 W_{p_{b1}}(1) - W_{p_{b1}}(\lambda_1)\|^2}{(1 - \lambda_1) \lambda_1} + \frac{\|\lambda_2 W_{p_{b2}}(1) - W_{p_{b2}}(\lambda_2)\|^2}{(1 - \lambda_2) \lambda_2}$$

The distribution is different from the common breaks case. However, it can easily be simulated for prespecified values of p_{b1} , p_{b2} and ε .

Now, consider another situation, where the system switches from regime 1 to regime 2, then switches back to regime 1. This type of phenomenon has been noticed by Sensier and van Dijk (2004), who argue that the volatility of the time series of aggregate price indices show an increase in the early 1970s and a decrease of roughly similar absolute magnitude in the early 1980s. Let p_b and n_b denote the number of regressors and of independent entries of the covariance matrix of the errors, respectively, subject to change. The test is then,

$$\sup LR_T^S(k_1, k_2, p_b, n_b, \varepsilon) = \sup_{(\lambda_1, \lambda_2) \in \Lambda_\varepsilon} [2 \log \hat{L}_T(k_1, k_2) - 2 \log \tilde{L}_T]$$

where $\log \tilde{L}_T$ denotes the log likelihood function estimated under the null hypothesis of no change, and $\log \hat{L}_T(k_1, k_2)$ denotes the the maximized value of the likelihood function under the alternative hypothesis of two changes but imposing the restriction that the first and the third regime are the same, and imposing that the maximization is taken over the following set of admissible partitions

$$\Lambda_\varepsilon = \{(\lambda_1, \lambda_2); \lambda_1 \geq \varepsilon, \lambda_2 - \lambda_1 \geq \varepsilon, \lambda_2 \leq 1 - \varepsilon\}$$

The limiting distribution of the test is presented in the next Theorem, whose proof is straightforward and omitted.

Theorem 9 *Let $W_{p_b+n_b}(\cdot)$ be a $p_b + n_b$ vector of independent Wiener processes on $[0, 1]$. Then under Assumptions A11-A12 (assuming Normal errors when allowing changes in the covariance matrix of the errors),*

$$\sup LR_T^S \Rightarrow \text{Sup}_{(\lambda_1, \lambda_2) \in \Lambda_\varepsilon} \frac{\| [W_{p_b+n_b}(\lambda_2) - \lambda_2 W_{p_b+n_b}(1)] - [W_{p_b+n_b}(\lambda_1) - \lambda_1 W_{p_b+n_b}(1)] \|^2}{(\lambda_2 - \lambda_1) (1 - \lambda_2 + \lambda_1)}$$

These two examples, show how the limit distribution of the test vary on a case by case basis when imposing restrictions on the coefficients of the system.

7 Conclusions

We have presented a comprehensive treatment of issues related to the estimation, inference and computation in linear multivariate regressions models with multiple structural changes. Our study provides a unified treatment of models with common breaks, partial structural breaks models, block partial breaks models and locally ordered breaks models, among others. The latter is novel and of particular interest given potential applicability in contexts related to the Lucas critique. Our results being asymptotic in nature, there is certainly a need to evaluate the quality of the approximations. An important question to answer is how our results depend on the time span and the cross section dimension of the system. It is also important to examine the robustness of the procedure, i.e, how sensitive are the properties of the estimates of the break dates to certain misspecifications of the system. Such investigations, and others, are the object of the ongoing research.

References

- [1] Altissimo, F. and V. Corradi (2003): “Strong Rules for Detecting the Number of Breaks in a Time Series,” *Journal of Econometrics*, 117, 207-244.
- [2] Alogoskoufis, G.S. and R. Smith (1991): “The Phillips Curve, the Persistence of Inflation, and the Lucas Critique,” *American Economic Review*, 81, 1254-1275.
- [3] Andrews, D.W.K. (1991): “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation,” *Econometrica*, 59, 817-858.
- [4] ——— (1993): “Testing for Parameter Instability and Structural Change with Unknown Change Point,” *Econometrica*, 61, 821-856.
- [5] Bai, J. (1997): “Estimation of a Change Point in Multiple Regression Models,” *Review of Economics and Statistics*, 79, 551-563.
- [6] ——— (1999): “Likelihood Ratio Tests for Multiple Structural Changes,” *Journal of Econometrics*, 91, 299-323.
- [7] ——— (2000): “Vector Autoregressive Models with Structural Changes in Regression Coefficients and in Variance-Covariance Matrices,” *Annals of Economics and Finance*, 1, 303-339.
- [8] Bai, J., R.L. Lumsdaine and J.H. Stock (1998): “Testing for and Dating Breaks in Multivariate Time Series,” *Review of Economic Studies*, 65, 395-432.
- [9] Bai, J. and P. Perron (1998): “Estimating and Testing Linear Models with Multiple Structural Changes,” *Econometrica*, 66, 47-78.
- [10] ——— (2003a): “Computation and Analysis of Multiple Structural Change Models,” *Journal of Applied Econometrics*, 18, 1-22.
- [11] ——— (2003b): “Critical Values for Multiple Structural Change Tests,” *Econometrics Journal*, 6, 72-78.
- [12] ——— (2004): “Multiple Structural Change Models: A Simulation Analysis,” manuscript, Department of Economics, Boston University.
- [13] Corradi, V. (1999): “Deciding between $I(0)$ and $I(1)$ via FLIL-based Bounds,” *Econometric Theory*, 15, 643-663.
- [14] Csörgő, M and L. Horváth (1997): *Limit Theorems in Change-point Analysis*, Wiley Series in Probability and Statistics, New York: John Wiley.
- [15] Dehling, H. and W. Philipp (1982): “Almost Sure Invariance Principles for Weakly Dependent Vector-Valued Random Variables,” *The Annals of Probability*, 10, 689-701.

- [16] Domowitz, I. and H. White (1982): “Misspecified Models with Dependent Observations,” *Journal of Econometrics*, 20, 35-58.
- [17] Eberlein, E. (1986): “On Strong Invariance Principles under Dependent Assumptions,” *The Annals of Probability*, 14, 260-270.
- [18] Hansen, B.E. (1991): “Strong Laws for Dependent Heterogeneous Processes,” *Econometric Theory*, 7, 213-221.
- [19] ———(2000): “Testing for Structural Change in Conditional Models,” *Journal of Econometrics*, 97, 93-115.
- [20] Hansen, P.R. (2003): “Structural Changes in the Cointegrated Vector Autoregressive Model,” *Journal of Econometrics*, 114, 261-295.
- [21] Hawkins, D.M. (1976): “Point Estimation of the Parameters of a Piecewise Regression Model,” *Applied Statistics*, 25, 51-57.
- [22] Ibragimov, I.A. (1962): “Some Limit Theorems for Stationary Processes,” *Theory of Probability and its Applications*, 7, 349-82.
- [23] Liu, J., S. Wu and J.V. Zidek (1997): “On Segmented Multivariate Regressions,” *Statistica Sinica*, 7, 497-525.
- [24] Levielle, V. and E. Moulines (2000): “Least-squares Estimation of Unknown Number of Shifts in a Time Series,” *Journal of Time Series Analysis*, 21, 33-59.
- [25] Magnus, J.R. (1988): *Linear Structures*, London: Charles Griffin & Co.
- [26] Ng, S. and T.J. Vogelsang (2002): “Analysis of Vector Autoregressions in the Presence of Shifts in Mean,” *Econometric Reviews*, 21, 353-381.
- [27] Perron, P. (1997): “L’Estimation de Modèles avec Changements Structurels Multiples,” *Actualité Économique*, 73, 457-505
- [28] Perron, P. and Z. Qu (2004): “Estimating Restricted Structural Change Models,” unpublished manuscript, Department of Economics, Boston University.
- [29] Sensier, M., D. van Dijk, (2004): “Testing for Volatility Changes in U.S. Macroeconomic Time Series,” *The Review of Economics and Statistics*, 86, 833-839.
- [30] Tobing, H. and C.A. McGilchrist (1992): “Recursive Residuals for Multivariate Regression Models,” *Australian Journal of Statistics*, 34, 217-232.
- [31] Yao, Y.-C. (1988): “Estimating the Number of Change-Points via Schwarz’ Criterion,” *Statistics and Probability Letters*, 6, 181-189.

Appendix

It is useful to first state a few key results that will be used throughout the proofs, namely a generalized Hajek-Renyi inequality, a Strong Law of Large Numbers (SLLN), a Functional Central Limit Theorem (FCLT), a Strong Approximation Theorem (SAT) and a Law of Iterated Logarithm (LIL). These are stated in the following two Lemmas.

Lemma A.1 (*Generalized Hajek-Renyi inequality*) *Let $(\xi_i)_{i \geq 1}$ be a sequence of mean zero R^d -valued random vectors. Define \mathcal{F}_k as an increasing σ -field generated by $(\xi_i)_{i \leq k}$. Suppose $(\xi_i)_{i \geq 1}$ satisfies Assumption A4 with $x_i u_i$ replaced by ξ_i , then there exists an $L < \infty$ such that, for every $c > 0$ and $m > 0$, $P(\sup_{k \geq k_0} k^{-1} \|\sum_{i=1}^k \xi_i\| > c) \leq (L/c^2 k_0)$*

Proof: This Lemma is presented in Bai and Perron (1998) under a set of mixingale conditions. We show that Assumption A4 implies them. By the mixingale inequality of Ibragimov (1962), we have

$$\|E(\xi_{i+k} | \mathcal{F}_i)\|_2 \leq 2(\sqrt{2} + 1) \alpha_k^{1/2-1/s} \|\xi_{i+k}\|_s$$

for any $s \geq 2$. In particular, let $s = r + \delta$, with r and δ defined in Assumption A4, we then have

$$\|E(\xi_{i+k} | \mathcal{F}_i)\|_2 \leq 6M \alpha_k^{1/2-1/s}$$

and

$$\alpha_k^{1/2-1/s} = k^{-\frac{4r}{r-2}(1/2-1/(r+\delta))} = k^{-2-\frac{4\delta}{(r-2)(r+\delta)}}$$

which, combined with the fact that ξ_i is \mathcal{F}_i measurable, imply that ξ_i is a L^2 mixingale of size $-2 - 4\delta/((r-2)(r+\delta))$. If we let ψ_j denote the mixing coefficient and define $\kappa = 2\delta/((r-2)(r+\delta))$, then $\sum_{j=0}^{\infty} j^{1+\kappa} \psi_j < \infty$. Hence the mixingale conditions required in Bai and Perron (1998) are satisfied and the Lemma holds.

Lemma A.2 *Let $(\xi_i)_{i \geq 1}$ be a sequence of mean zero R^d -valued random vectors satisfying A4. Define $S_k(\ell) = \sum_{i=1}^{\ell+k} \xi_i$, then, (a) (SAT) the covariances 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 $\sum_{i=1}^t \xi_i - W(t) = O_{a.s.}(t^{1/2-\kappa})$ for some $\kappa > 0$; (b) (FCLT) $T^{-1/2} \sum_{t=1}^{[Tr]} \xi_t \Rightarrow \Omega W^*(r)$ where $W^*(r)$ is a d -vector of independent Wiener processes and \Rightarrow denotes weak convergence under the Skorohod topology; (c) (SLLN) $k^{-1} \sum_{i=1}^k \xi_i \xrightarrow{a.s.} 0$ as $k \rightarrow \infty$; (d) (LIL) $\limsup_{k \rightarrow \infty} (k^{1/2} \log^{1/2} k)^{-1} \|\Omega_k^{-1/2} \sum_{i=1}^k \xi_i\| = O_p(1)$.*

Proof: Given assumption A4, parts (a) and (b) follow from Theorem 2 in Eberlain (1986), once we show that $\|E(S_n(\ell) | \mathcal{F}_\ell)\|_2 \leq C$ for C independent of ℓ and that uniformly in ℓ ,

$$\|E(S_k(\ell) S_k(\ell)' | \mathcal{F}_m) - E(S_k(\ell) S_k(\ell)')\|_1 = O(k^{1-\theta})$$

The proof of the above proceeds in the same way as in Corradi (1999, p. 651-652). Hence, details are omitted. Part (c) can be proved by applying Corollary 2 of Hansen (1991), using

the fact that ξ_i is a L^2 mixingale with mixing size -2 and bounded fourth moment. For part (d),

$$\begin{aligned} & \frac{1}{k^{1/2} \log^{1/2} k} \left\| \Omega_k^{-1/2} \sum_{i=1}^k \xi_i \right\| \\ &= \frac{1}{k^{1/2} \log^{1/2} k} \left\| \Omega_k^{-1/2} \sum_{i=1}^k \xi_i - \Omega^{-1/2} W(k) + \Omega^{-1/2} W(k) \right\| \\ &\leq \frac{1}{k^{1/2} \log^{1/2} k} \left\| \Omega_k^{-1/2} \sum_{i=1}^k \xi_i - \Omega^{-1/2} W(k) \right\| + \frac{1}{k^{1/2} \log^{1/2} k} \left\| \Omega^{-1/2} W(k) \right\| \end{aligned}$$

Since

$$\frac{1}{k^{1/2} \log^{1/2} k} \left\| \Omega_k^{-1/2} \sum_{i=1}^k \xi_i - \Omega^{-1/2} W(k) \right\| = o_{a.s.}(1)$$

by part (a), and

$$\limsup_{k \rightarrow \infty} \frac{1}{k^{1/2} \log^{1/2} k} \left\| \Omega^{-1/2} W(k) \right\| = O_p(1)$$

by the LIL for vector valued Brownian motion processes, we have

$$\limsup_{k \rightarrow \infty} \frac{1}{k^{1/2} \log^{1/2} k} \left\| \Omega_k^{-1/2} \sum_{i=1}^k \xi_i \right\| = O_p(1)$$

which completes the proof.

We first establish the consistency and rate of convergence of the estimates of the break fractions under A1-A9 when no restrictions are imposed and the estimates are obtained by maximizing the unrestricted quasi-likelihood ratio.

Lemma A.3 *Let $(\hat{T}_1, \dots, \hat{T}_m)$, $(\hat{\beta}_j, \hat{\Sigma}_j)$ be the solution to $\arg \max_{(T_1, \dots, T_m; \beta; \Sigma)} lr_T(\mathcal{T}, \beta, \Sigma)$ with $lr_T(\mathcal{T}, \beta, \Sigma)$ defined by the log of (5) and (4), then under A1-A9: for all $j = 1, \dots, m$, $v_T^2(\hat{T}_j - T_j^0) = O_p(1)$, and for $j = 1, \dots, m+1$, $\sqrt{T}(\hat{\beta}_j - \beta_j^0) = O_p(1)$ and $\sqrt{T}(\hat{\Sigma}_j - \Sigma_j^0) = O_p(1)$.*

Proof of Lemma A3: To prove this result, we show that under A1-A9, slightly different versions of the 10 properties of the quasi-likelihood ratio discussed in Bai, Lumsdaine and Stock (1998), henceforth BLS (1998), and Bai (2000) are satisfied under our set of assumptions. Once these are established, the proof proceeds as in Bai (2000, pp. 324-9). Following BLS (1998), but using slightly different notations, the quasi-likelihood ratio using the first k observations, evaluated at θ and Σ , can be written as

$$\mathcal{L}(1, k; \beta, \Sigma) = \frac{\prod_{t=1}^k f(y_t | x_t, \dots, \beta, \Sigma)}{\prod_{t=1}^k f(y_t | x_t, \dots, \beta^0, \Sigma^0)}$$

Let $\hat{\beta}_{(k)}$ and $\hat{\Sigma}_{(k)}$ denote the values of β and Σ that correspond to the maximum of $\mathcal{L}(1, k; \beta, \Sigma)$. We have the following property about the magnitude of the parameter estimates and the likelihood function.

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

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

$$\sup_{T\delta \leq k \leq T} \left(\|\hat{\beta}_{(k)} - \beta^0\| + \|\hat{\Sigma}_{(k)} - \Sigma^0\| \right) = O_p(T^{-1/2})$$

This corresponds to Property 1 of BLS (1998, p. 420). It states that the likelihood function and the MLE are well behaved in large samples. The uniformity of the bound is important since we need to search over all admissible partitions to find the break points. BLS provides a proof for SUR model with common regressors, in which case the covariance matrix plays no role in estimating β . Here, our setup is more complicated since the interaction between β and Σ makes the problem non-linear. The solution is to use an argument based on the minimization of the *Kullback-Leibler* distance, which is due to Domowitz and White (1982, Theorem 2.2).

Proof of Property 1: First, simple arguments leads to the result that $E_0(\log \mathcal{L}(1, k; \beta, \Sigma))$ achieves a maximum at $\beta_{(k)} = \beta^0$ and $\Sigma^0 = E_0(k^{-1} \sum_{t=1}^k u_t u_t')$ where E_0 denotes the expectation taken over the true density. Let Θ_1 denote an open sphere that contains (β^0, Σ^0) and let $\bar{\Theta}_1$ be its closure constructed in such a way that it excludes values of Σ such that $|\Sigma| = 0$. Then, applying a SLLN and a FCLT, we have

$$|\log \mathcal{L}(1, k; \beta, \Sigma) - E_0(\log \mathcal{L}(1, k; \beta, \Sigma))| \xrightarrow{a.s} 0$$

uniformly over $\bar{\Theta}_1$. Using the continuous mapping theorem, we have

$$\|\tilde{\beta}_{(k)} - \beta^0\| + \|\tilde{\Sigma}_{(k)} - \Sigma^0\| \xrightarrow{a.s} 0$$

where

$$(\tilde{\beta}_{(k)}, \tilde{\Sigma}_{(k)}) = \arg \max_{(\beta, \Sigma) \in \bar{\Theta}_1} \log \mathcal{L}(1, k; \beta, \Sigma)$$

In the above, the maximization is taken over a compact set. We now show that the strong consistency still holds when that restriction is dropped. Notice that for large k , with probability arbitrary close to 1, $\log \mathcal{L}(1, k; \beta, \Sigma)$ is continuous and strictly concave at $(\tilde{\beta}_{(k)}, \tilde{\Sigma}_{(k)})$, an inner point of $\bar{\Theta}_1$. Under the assumption that the likelihood function does not have multiple maxima, we can conclude that for large k , $(\tilde{\beta}_{(k)}, \tilde{\Sigma}_{(k)})$ coincides with $(\hat{\beta}_{(k)}, \hat{\Sigma}_{(k)})$, which is the unique solution obtained by solving the first order condition of the quasi-maximum likelihood without directly imposing $(\beta, \Sigma) \in \bar{\Theta}_1$. Hence the strong consistency of $(\hat{\beta}_{(k)}, \hat{\Sigma}_{(k)})$ is proved.

Now, using the fact that $\hat{\beta}_{(k)} - \beta^0 = (\sum_{t=1}^k x_t \hat{\Sigma}_{(k)}^{-1} x_t')^{-1} \sum_{t=1}^k x_t \hat{\Sigma}_{(k)}^{-1} u_t$ and applying the generalized Hajek-Renyi inequality on $\sum_{t=1}^k x_t (\Sigma^0)^{-1} u_t$, together with the strong consistency of $\hat{\Sigma}_{(k)}$, we have $\sup_{T\delta \leq k \leq T} \|\hat{\beta}_{(k)} - \beta^0\| = O_p(T^{-1/2})$. For $\hat{\Sigma}_{(k)} - \Sigma^0$, we use the fact that

$$\hat{\Sigma}_{(k)} - \Sigma^0 = \frac{1}{k} \sum_{t=1}^k (u_t - x_t'(\hat{\beta}_{(k)} - \beta^0))(u_t - x_t'(\hat{\beta}_{(k)} - \beta^0))' - \Sigma^0$$

and again applying the generalized Hajek-Renyi inequality, we have $\sup_{T\delta \leq k \leq T} \|\tilde{\Sigma}_{(k)} - \Sigma^0\| = O_p(T^{-1/2})$. Finally, $\sup_{T\delta \leq k \leq T} \mathcal{L}(1, k; \hat{\beta}_{(k)}, \hat{\Sigma}_{(k)}) = O_p(1)$ is a directly implication of the above results which completes the proof.

The following property corresponds to Property 2 of BLS (1998), which provides a bound for the sequential quasi-likelihood function in small samples. Two additional complications appear in our context. First, we allow a general dependence structure for the errors and the regressors. Secondly, as before, the interaction between β and Σ makes the problem non-linear. The solution is to apply the SAT and the LIL.

Property 2 *For each $\epsilon > 0$, there exists a $B > 0$, such that for all large T*

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

Proof of Property 2: The likelihood function evaluated at $\hat{\beta}_{(k)}$ and $\hat{\Sigma}_{(k)}$ can be written as

$$\log \mathcal{L}(1, k; \hat{\beta}_{(k)}, \hat{\Sigma}_{(k)}) = -\frac{k}{2} (\log |\hat{\Sigma}_{(k)}| - \log |\Sigma^0|) + \frac{1}{2} \left(\sum_{t=1}^k u_t' (\Sigma^0)^{-1} u_t - kn \right)$$

Denote $A_k = \hat{\beta}_{(k)} - \beta^0$, then

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

and

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

Note that $\hat{\Sigma}_{(k)} \xrightarrow{a.s} \Sigma^0$, $\hat{\beta}_{(k)} \xrightarrow{a.s} \beta^0$ and $k^{-1} \sum_{t=1}^k u_t u_t' \xrightarrow{a.s} \Sigma^0$ as $k \rightarrow \infty$ which can be shown applying a SLLN. Hence, we have

$$\sup_{k \geq k_1} \left\| \hat{\Sigma}_{(k)}^{-1} - (\Sigma^0)^{-1} \right\| = O_p(1), \quad \sup_{k \geq k_1} \left\| \left(\frac{1}{k} \sum_{t=1}^k x_t (\Sigma^0)^{-1} x_t' \right)^{-1} \right\| = O_p(1)$$

and

$$\sup_{k \geq k_1} \left\| \left(\frac{1}{k} \sum_{t=1}^k x_t \hat{\Sigma}_{(k)}^{-1} x_t' \right)^{-1} - \left(\frac{1}{k} \sum_{t=1}^k x_t (\Sigma^0)^{-1} x_t' \right)^{-1} \right\| = O_p(1)$$

for some fixed k_1 . Since $\sup_{k \leq k_1} \mathcal{L}(1, k; \hat{\beta}_{(k)}, \hat{\Sigma}_{(k)}) = O_p(1)$, without loss of generality, we may assume $k \geq k_1$. Then

$$\begin{aligned} \sup_{k \geq k_1} \|A_k\| &\leq \sup_{k \geq k_1} k^{-1/2} \left\| \left(k^{-1} \sum_{t=1}^k x_t \hat{\Sigma}_{(k)}^{-1} x_t' \right)^{-1} \right\| \sup_{k \geq k_1} \left\| k^{-1/2} \sum_{t=1}^k x_t \hat{\Sigma}_{(k)}^{-1} u_t \right\| \\ &= \sup_{k \geq k_1} \left\| k^{-1/2} \sum_{t=1}^k x_t \hat{\Sigma}_{(k)}^{-1} u_t \right\| O_p(1) \end{aligned}$$

Now, let $\Omega_k^0 = \text{var}(\sum_{t=1}^k x_t (\Sigma^0)^{-1} u_t)$, we have

$$\begin{aligned} & \frac{1}{k^{1/2} \log^{1/2} T} \sup_{k \geq k_1} \left\| (\Omega_k^0)^{-1/2} \sum_{t=1}^k x_t (\Sigma^0)^{-1} u_t \right\| \\ \leq & \frac{1}{k^{1/2} \log^{1/2} T} \sup_{k \geq k_1} \left\| (\Omega_k^0)^{-1/2} \sum_{t=1}^k x_t (\Sigma^0)^{-1} u_t - W(k) \right\| + \frac{1}{k^{1/2} \log^{1/2} T} \sup_{k \geq k_1} \|W(k)\| \end{aligned}$$

where $W(k)$ is a vector-valued Wiener process. Hence,

$$\frac{1}{k^{1/2} \log^{1/2} T} \sup_{k \geq k_1} \left\| (\Omega_k^0)^{-1/2} \sum_{t=1}^k x_t (\Sigma^0)^{-1} u_t - W(k) \right\| = o_{a.s.}(1)$$

from Lemma A2, and $(k^{1/2} \log^{1/2} T)^{-1} \sup_{k \geq k_1} \|W(k)\| = o_{a.s.}(1)$ using a LIL for a vector valued Wiener process. This shows that $\sup_{k \geq k_1} \|A_k\| = O_p(\log^{1/2} T)$. Now we use this result to obtain a bound for the likelihood function. Applying a Taylor series expansion, we have

$$\begin{aligned} \log \mathcal{L}(1, k; \hat{\beta}_{(k)}, \hat{\Sigma}_{(k)}) &= -\frac{k}{2} \text{tr}(\hat{\Sigma}_{(k)}(\Sigma^0)^{-1} - I) + \frac{1}{2} \left(\sum_{t=1}^k u_t' (\Sigma^0)^{-1} u_t - kn \right) \\ &\quad + \frac{k}{4} \text{tr}\{(\hat{\Sigma}_{(k)}(\Sigma^0)^{-1} - I)^2\} + O_p(1) \end{aligned}$$

where the remainder term is $O_p(1)$ uniformly in k . We shall now show that

$$-\frac{k}{2} \text{tr}(\hat{\Sigma}_{(k)}(\Sigma^0)^{-1} - I) + \frac{1}{2} \left(\sum_{t=1}^k u_t' (\Sigma^0)^{-1} u_t - kn \right) = O_p(\log T) \quad (\text{A.1})$$

and

$$\frac{k}{4} \text{tr}\{(\hat{\Sigma}_{(k)}(\Sigma^0)^{-1} - I)^2\} = O_p(\log T) \quad (\text{A.2})$$

uniformly in k . First, for (A.1),

$$\begin{aligned} & -\frac{k}{2} \text{tr}(\hat{\Sigma}_{(k)}(\Sigma^0)^{-1} - I) + \frac{1}{2} \left(\sum_{t=1}^k u_t' (\Sigma^0)^{-1} u_t - kn \right) \\ = & -\frac{k}{2} \text{tr} \left(\frac{1}{k} \sum_{t=1}^k ((\Sigma^0)^{-1} u_t - (\Sigma^0)^{-1} x_t' A_k)' (u_t - x_t' A_k) - I \right) + \frac{1}{2} \left(\sum_{t=1}^k u_t' (\Sigma^0)^{-1} u_t - kn \right) \\ = & -\frac{1}{2} \text{tr} \left\{ A_k' \left(\sum_{t=1}^k x_t (\Sigma^0)^{-1} x_t' \right) A_k \right\} + \frac{1}{2} \text{tr} \left\{ A_k' \left(\sum_{t=1}^k x_t (\Sigma^0)^{-1} u_t \right) \right\} \\ & + \frac{1}{2} \text{tr} \left\{ \sum_{t=1}^k (u_t' (\Sigma^0)^{-1} x_t) A_k \right\} \\ = & O_p(\log T) \end{aligned}$$

where the last equality follows since

$$\begin{aligned} \sup_{k \geq k_1} \left\| A'_k \left(\sum_{t=1}^k x_t (\Sigma^0)^{-1} u_t \right) \right\| &\leq \sup_{k \geq k_1} \left\| k^{-1/2} \left(\sum_{t=1}^k x_t (\Sigma^0)^{-1} u_t \right) \right\| O_p(\log^{1/2} T) = O_p(\log T) \\ \sup_{k \geq k_1} \left\| A'_k \left(\sum_{t=1}^k x_t (\Sigma^0)^{-1} x'_t \right) A_k \right\| &\leq \sup_{k \geq k_1} \|A'_k\| \sup_{k \geq k_1} \left\| \sum_{t=1}^k x_t (\Sigma^0)^{-1} x'_t \right\| \sup_{k \geq k_1} \|A_k\| = O_p(\log T) \end{aligned}$$

Remark 9 *If the process $x_t(\Sigma^0)^{-1}u_t$ is assumed to be strictly stationary, then Theorem 5.5 and Corollary 5.4 of Hall and Heyde (1980, p. 145) says that a LIL holds for the process if it has uniformly bounded second moments and satisfies*

$$\sum_{\ell=1}^{\infty} \left\| E(x_t(\Sigma^0)^{-1}u_t | \mathcal{F}_{t-\ell}) \right\|_2 + \sum_{\ell=1}^{\infty} \left\| x_t(\Sigma^0)^{-1}u_t - E(x_t(\Sigma^0)^{-1}u_t | \mathcal{F}_{t+\ell}) \right\|_2 < \infty$$

In this case, the LIL could be applied directly without first resorting to the strong invariance principle.

What remains to be shown is that (A.2) is also $O_p(\log T)$. To see this, note that,

$$\frac{k}{4} \text{tr} \left\{ (\hat{\Sigma}_{(k)}(\Sigma^0)^{-1} - I)^2 \right\} \equiv \frac{k}{4} \text{tr} \left\{ [\Psi_1 + \Psi_2 - \Psi_3]^2 \right\} \leq \frac{3k}{4} \text{tr} \left\{ \Psi_1^2 + \Psi_2^2 + \Psi_3^2 \right\}$$

where

$$\Psi_1 = \frac{1}{k} \sum_{t=1}^k (u_t(\Sigma^0)^{-1}u'_t - I), \quad \Psi_2 = \frac{1}{k} \sum_{t=1}^k x'_t A_k A'_k x_t (\Sigma^0)^{-1}$$

and

$$\Psi_3 = \frac{1}{k} \sum_{t=1}^k (u_t A'_k x_t (\Sigma^0)^{-1} + x'_t A_k u'_t (\Sigma^0)^{-1})$$

For the first term, $(3k/4)\text{tr}(\Psi_1^2) = O_p(\log T)$ after applying the strong invariance principle and the LIL on the Brownian motion process. For the second term

$$\frac{3k}{4} \text{tr}(\Psi_2^2) \leq \frac{3k}{4} (\text{tr}(\Psi_2))^2 = \frac{3}{4} (\text{tr}(k^{-1/2}\Phi_k) + o_p(1))^2 = O_p(\log T)$$

where the inequality follows because Ψ_2 is a symmetric positive definite matrix and Φ_k is defined as

$$\Phi_k = (k^{-1/2} \sum_{t=1}^k x_t(\Sigma^0)^{-1}u_t)' [k^{-1} \sum_{t=1}^k x_t(\Sigma^0)^{-1}x'_t]^{-1} (k^{-1/2} \sum_{t=1}^k x_t(\Sigma^0)^{-1}u_t)$$

For the third term, following the same arguments, we have $(3k/4)\text{tr}(\Psi_3^2) = O_p(\log T)$ which completes the proof.

The following property states that the value of the likelihood ratio is arbitrarily small for large T when the parameters are evaluated away from zero, assuming a positive fraction of the observations are used.

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

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

Proof of Property 3: We first consider the behavior of the likelihood function over the following compact set

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

where λ_{\min} and λ_{\max} denotes the smallest and largest eigenvalues and the finite constants d_1 , d_2 and d_3 are chosen in such a way that (β^0, Σ^0) an inner point of $\bar{\Theta}_2$. We first want to show that

$$\Pr \left(\sup_{k \geq T^\delta} \sup_{(\beta, \Sigma) \in S_T \cap \bar{\Theta}_2} T^D \mathcal{L}(1, k; \beta, \Sigma) > 1 \right) < \epsilon$$

Following BLS (1998, pp. 422-4), we decompose the sequential log-likelihood as

$$\log \mathcal{L}(1, k; \beta, \Sigma) = \mathcal{L}_{1,T} + \mathcal{L}_{2,T}$$

where

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

and

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

with $\beta^* = \beta - \beta^0$, $\Sigma^* = \Sigma - \Sigma^0$, $\eta_t = (\Sigma^0)^{-1} u_t$, and $\Psi_T = (\Sigma^0)^{-1/2} \Sigma^* (\Sigma^0)^{-1/2}$. Note that only $\mathcal{L}_{2,T}$ depends on β^* . Now, let $S_T = S_{1,T} \cup S_{2,T}$, with

$$S_{1,T} = \{(\beta, \Sigma); \|\Sigma - \Sigma^0\| \geq T^{-1/2} \log T, \beta \text{ arbitrary}\}$$

and

$$S_{2,T} = \{(\beta, \Sigma); \|\beta - \beta^0\| \geq T^{-1/2} \log T \text{ and } \|\Sigma - \Sigma^0\| \leq T^{-1/2} \log T\}$$

We then need to show that

$$\Pr \left(\sup_{k \geq T^\delta} \sup_{(\beta, \Sigma) \in S_{1,T} \cap \bar{\Theta}_2} T^D \mathcal{L}(1, k; \beta, \Sigma) > 1 \right) < \epsilon \quad (\text{A.3})$$

and

$$\Pr \left(\sup_{k \geq T^\delta} \sup_{(\beta, \Sigma) \in S_{2,T} \cap \bar{\Theta}_2} T^D \mathcal{L}(1, k; \beta, \Sigma) > 1 \right) < \epsilon \quad (\text{A.4})$$

The proof of (A.4) proceeds exactly as in BLS (1998) and, hence, is omitted. It remains to show (A.3), or

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

First note that, on $S_{1,T}$, \mathcal{L}_{2T} is a quadratic function of β^* and has maximum value

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

Applying a SLLN,

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

Also,

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

Using Lemma A2, we have, for any fixed $r > 0$,

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

while $\|S' (\Sigma^{-1} \otimes I_n)\| = \sum_{i=1}^n (1 + \lambda_i)^{-1} O_p(1)$ with λ_i ($i = 1, \dots, n$) the eigenvalues of $(\Sigma^0)^{-1/2} \Sigma^* (\Sigma^0)^{-1/2}$. Hence,

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

which implies

$$\sup_{k \geq T\delta} \sup_{S_{1,T} \cap \bar{\Theta}_2} \mathcal{L}_{2T} \leq \frac{k}{2} \sum_{i=1}^n \frac{1}{1 + \lambda_i} r^2 b_T^2 \quad (\text{A.5})$$

where $b_T = T^{-1/2} \log T$ with the inequality holding with probability arbitrarily close to 1 for large T . For \mathcal{L}_{1T} , BLS (1998) show that

$$\sup_{k \geq T\delta} \sup_{S_{1,T} \cap \bar{\Theta}_2} \mathcal{L}_{1T} \leq -\frac{k}{2} \left[\sum_{i=1}^n \left(\log(1 + \lambda_i) + \left(\frac{1}{1 + \lambda_i} - 1 \right) (1 + \text{sign}(\lambda_i) a b_T) \right) \right] \quad (\text{A.6})$$

with probability arbitrarily close to 1 for large T , where a is a fixed positive number which can be made arbitrarily small. Combining the above two inequalities, and using arguments as in BLS (1998), we can show that

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

We still need to show that the above sup-bound remains valid even if the maximization problem is carried over an unrestricted parameter space. To this end, it is sufficient to show that

$$\lim_{k \rightarrow \infty} \Pr \left(\arg \max_{(\theta, \Sigma) \in S_T} \mathcal{L}(1, k; \beta, \Sigma) \in \bar{\Theta}_2 \right) = 1$$

i.e, the sequence of global maximizers of the quasi-likelihood function, $(\hat{\theta}_{(k)}, \hat{\Sigma}_{(k)})$ eventually falls into the compact set $\bar{\Theta}_2$ almost surely. Suppose this is not so, then there are three possibilities: 1) with positive probability, there is a sequence $(\hat{\theta}_{(k)}, \hat{\Sigma}_{(k)})$ satisfying $\inf(\lambda_{\min}(\hat{\Sigma}_{(k)})) \rightarrow d_l > 0$, with $d_l < d_2$ or $\sup(\lambda_{\max}(\hat{\Sigma}_{(k)})) \rightarrow d_u < \infty$, with $d_u > d_3$; 2) with positive probability, there is a sequence $(\hat{\theta}_{(k)}, \hat{\Sigma}_{(k)})$ with $\inf(\lambda_{\min}(\hat{\Sigma}_{(k)})) \rightarrow 0$ or $\sup(\lambda_{\max}(\hat{\Sigma}_{(k)})) \rightarrow \infty$; 3) with positive probability, there is a sequence $(\hat{\theta}_{(k)}, \hat{\Sigma}_{(k)})$ with $\inf(\lambda_{\min}(\hat{\Sigma}_{(k)})) \geq d_2$ and $\sup(\lambda_{\max}(\hat{\Sigma}_{(k)})) \leq d_3$ but $\limsup(\hat{\theta}_{(k)}) > d_1$.

The first case is ruled out by the asymptotic identifiability condition and the uniform almost sure convergence of the likelihood function over a compact set. Indeed, in this case, by definition, $(\hat{\theta}_{(k)}, \hat{\Sigma}_{(k)}) = \arg \max_{(\theta, \Sigma) \in S_T} \mathcal{L}(1, k; \beta, \Sigma)$, and if $\hat{\Sigma}$ has bounded eigenvalues for large k , then $\hat{\Sigma}_{(k)}$ must lie on the boundary of S_T , an inner point of $\bar{\Theta}_2$. The second case is also impossible because the log likelihood function would then diverge to minus infinity. The third case is ruled out again, because values of $\hat{\theta}_{(k)}$ lying on the boundary of S_T will yield a larger value of the likelihood function almost surely. This completes the proof.

Property 4 *A Property corresponding to Property 4 of BLS (1998) is not needed.*

The next property concerns the value of the likelihood ratio when no positive fraction of the observations is involved. It is slightly different from that of BLS (1998), in the sense that the maximum is taken over a restricted set. The restriction simplifies the proof and is also what is needed for the intended application. Also, as pointed by Bai (2000), the existence of a limit for $(h_T d_T^2)/T$ is not necessary. It is sufficient to have $\liminf_{T \rightarrow \infty} (h_T d_T^2)/T \geq h > 0$.

Property 5 *Let h_T and d_T be positive sequences such that h_T is non-decreasing, $d_T \rightarrow +\infty$, and $(h_T d_T^2)/T \rightarrow h < \infty$. Define $\bar{\Theta}_3 = \{(\beta, \Sigma) : \|\beta\| \leq p_1, \lambda_{\min}(\Sigma) \geq p_2, \lambda_{\max}(\Sigma) \leq p_3\}$, where p_1, p_2 and p_3 are arbitrary constants satisfying $p_1 < \infty, 0 < p_2 \leq p_3 < \infty$. Define $S_T = \{(\beta, \Sigma) : \|\beta - \beta^0\| \geq T^{-1/2} \log T \text{ or } \|\Sigma - \Sigma^0\| \geq T^{-1/2} \log T\}$. Then, for any $\epsilon > 0$, there exists an $A > 0$, such that when T is large*

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

Proof of Property 5: As in the proof of Property 3, we only need to look at the behavior of \mathcal{L}_{2T} over $S_{1,T} \cap \bar{\Theta}_3$, the rest of the proof is the same as BLS (1998). In other words, we need to show

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

or

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

Define $b_T = T^{-1/2}d_T$. Upon showing that (A.5) and (A.6) hold, all the arguments in the previous proof go through. The proof of (A.6) is the same as in BLS (1998) with only minor modifications and, hence, omitted. For (A.5), we have,

$$\sup_{S_{1,T}} \mathcal{L}_{2T} = \frac{k}{2} \left(\frac{1}{k} \sum_{t=1}^k x_t \Sigma^{-1} u_t \right)' \left(\frac{1}{k} \sum_{t=1}^k x_t \Sigma^{-1} x_t' \right)^{-1} \left(\frac{1}{k} \sum_{t=1}^k x_t \Sigma^{-1} u_t \right) \quad (\text{A.7})$$

where

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

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

$$\Pr \left(\sup_{k \geq k_1} \left\| \frac{1}{k} \sum_{t=1}^k z_t z_t' - Q_z \right\| > \epsilon \right) < \epsilon$$

Define $Q_\Delta = k^{-1} \sum_{t=1}^k z_t z_t' - Q_z$, then

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

where $A = S'(\Sigma^{-1} \otimes Q_z) S$ and $B = S'(\Sigma^{-1} \otimes Q_\Delta) S$. Since Σ^{-1} has uniformly bounded eigenvalues and $k^{-1} \sum_{t=1}^k z_t z_t'$ is positive definite for large k , A^{-1} and B^{-1} have bounded eigenvalues. Since B is uniformly small, $-A^{-1} B (A + B)^{-1}$ is uniformly small for large k . More precisely, $[S'(\Sigma^{-1} \otimes k^{-1} \sum_{t=1}^k z_t z_t') S]^{-1} - [S'(\Sigma^{-1} \otimes Q_z) S]^{-1} = o_{a.s}(1)$ as $k \rightarrow \infty$. Given the fact that there exists an $M > 0$ such that

$$\sup_{(\beta, \Sigma) \in S_{1,T} \cap \bar{\Theta}_3} \left\| [S'(\Sigma^{-1} \otimes Q_z) S]^{-1} \right\| < M$$

we have, for any $\epsilon > 0$, that there exists an $A > 0$ such that

$$\Pr \left(\sup_{k \geq Ah_T} \sup_{(\beta, \Sigma) \in S_{1,T} \cap \bar{\Theta}_3} \left\| \left(\frac{1}{k} \sum_{t=1}^k x_t \Sigma^{-1} x_t' \right)^{-1} \right\| > 2M \right) < \epsilon \quad (\text{A.8})$$

Now,

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

Using Lemma A1, we have

$$\Pr \left(\sup_{k \geq Ah_T} \left\| \frac{1}{k} \sum_{t=1}^k (I_n \otimes z_t) u_t \right\| > ab_T \right) \leq \frac{C_1}{Ah_T a^2 b_T} < \frac{2C_1}{Aa^2 h} \quad (\text{A.9})$$

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

$$\|S' (\Sigma^{-1} \otimes I_n)\| \leq nC_2 \sum_{i=1}^n \frac{1}{1 + \lambda_i} \quad (\text{A.10})$$

for some $0 < C_2 < \infty$, which depends on the matrix S . Now, combining (A.8)-(A.10), we have, for any $\epsilon > 0$ that there exists an $\bar{A} > 0$, such that with probability no less than $1 - \epsilon$,

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

with $G = 2n^3 C_2^2 M / p_2$, a finite constant depending on the dimension of the system, the limit moment matrix of the regressors and the property of the compact space $\bar{\Theta}_3$. Since a^2 can be made arbitrarily small by choosing a large A so can γ^2 , this establishes (A.5). The rest of the proof is essentially the same as that of Property 3, hence omitted.

The next Properties are the same as Lemmas 6 to 10 of Bai (2000). Since the proofs are similar, they are omitted.

Property 6 *With v_T satisfying Assumption A6, for each β and Σ such that $\|\beta - \beta^0\| \leq Mv_T$ and $\|\Sigma - \Sigma^0\| \leq Mv_T$, with $M < \infty$, we have*

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

where the supremum with respect to λ and Ξ is taken over a compact set such that $\|\lambda\| \leq M$ and $\|\Xi\| \leq M$.

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

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

Property 8

$$\sup_{T\delta \leq k \leq T} \sup_{\beta^*, \Sigma^*, \lambda, \Xi} \log \frac{\mathcal{L}(1, k; \beta^0 + T^{-1/2}\beta^* + T^{-1}\lambda, \Sigma^0 + T^{-1/2}\Sigma^* + T^{-1}\Xi)}{\mathcal{L}(1, k; \beta^0 + T^{-1/2}\beta^*, \Sigma^0 + T^{-1/2}\Sigma^*)} = o_p(1)$$

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

Property 9 Let $T_1 = [T\alpha]$ for some $\alpha \in (0, 1]$, and let $T_2 = [\sqrt{T}v_T^{-1}]$, where v_T satisfies Assumption A6. Consider

$$\begin{aligned} y_t &= x'_t \beta_1^0 + \Sigma_1^0 \eta_t & t &= 1, \dots, T_1 \\ y_t &= x'_t \beta_2^0 + \Sigma_2^0 \eta_t & t &= T_1 + 1, \dots, T_1 + T_2 \end{aligned}$$

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

Property 10 Assume the same set up as in Property 9 but with $T_2 = [Mv_T^{-2}]$. Then, $\hat{\beta}_n - \beta_1^0 = O_p(T^{-1/2})$, $\hat{\Sigma}_n - \Sigma_1^0 = O_p(T^{-1/2})$, $\hat{\beta}_n - \hat{\beta}_1 = O_p(T^{-1})$ and $\hat{\Sigma}_n - \hat{\Sigma}_1 = O_p(T^{-1})$.

Proof of Lemma 1: Consider first the consistency of the estimates of the break fractions $\tilde{\lambda}_j$. We first show that for any $\epsilon > 0$, as $T \rightarrow \infty$

$$P(|\tilde{T}_j - T_j^0| > \sqrt{T}v_T) < \epsilon, \quad j = 1, \dots, m, \quad (\text{A.11})$$

Define $A_1 = \{(T_1, \dots, T_m) \in \Lambda_\epsilon, |T_j - T_j^0| \leq \sqrt{T}v_T \text{ for all } j = 1, \dots, m\}$. Showing (A.11) is equivalent to proving that

$$P\left(\sup_{(T_1, \dots, T_m) \notin A_1} rlr_T(T_1, \dots, T_m) > \epsilon\right) < \epsilon \quad (\text{A.12})$$

as $T \rightarrow \infty$ for any $\epsilon > 0$. Now using the fact that the restricted likelihood is always less than or equal to the unrestricted likelihood, to prove (A.12) it is sufficient to show that

$$P\left(\sup_{(T_1, \dots, T_m) \notin A_1} lr_T(T_1, \dots, T_m) > \epsilon\right) < \epsilon$$

which follows from the proof of Lemma A3, more specifically, using Properties 1, 2, 3, and proceeding in the same way as in Bai (2000, pp. 324-26). Then, using (A.11), we want to show that for any $\epsilon > 0$, there exists a $C > 0$, such that, as $T \rightarrow \infty$:

$$P(|\tilde{T}_j - T_j^0| > Cv_T^{-2}) < \epsilon, \quad j = 1, \dots, m.$$

Define the set $A_2(C) = \{(T_1, \dots, T_m) \in A_1, |T_j - T_j^0| > Cv_T^2 \text{ for some } j\}$. It is sufficient to show that

$$P\left(\sup_{(T_1, \dots, T_m) \in A_2(C)} lr_T(T_1, \dots, T_m) > \epsilon\right) < \epsilon.$$

which again follows from the proof of Lemma A3, in particular, by using Properties 1, 5, 6, 9 and proceeding in the same way as in Bai (2000, pp. 326-28). The result about the rate of convergence of the estimates of the other parameters follow directly from the stated result about the rate of convergence of $\tilde{\lambda}_j$ (see Bai, 1997, and Bai and Perron, 1998).

Proof of Theorem 1: Given the result of Lemma 1, we can confine the maximization problem to the compact set C_M , defined by (9), for M large enough. Also, without loss of generality, we assume that the candidate estimates of the break dates occur before the true break dates, i.e., $v_T^2(T_j - T_j^0) \leq M$. The log likelihood ratio is defined by

$$\begin{aligned} lr_T &= -\frac{1}{2} \sum_{j=1}^{m+1} \sum_{t=T_{j-1}+1}^{T_j} (y_t - x_t' \beta_j)' \Sigma_j^{-1} (y_t - x_t' \beta_j) - \sum_{j=1}^{m+1} \frac{T_j - T_{j-1}}{2} \log |\Sigma_j| \\ &+ \frac{1}{2} \sum_{j=1}^{m+1} \sum_{t=T_{j-1}^0+1}^{T_j^0} (y_t - x_t' \beta_j^0)' (\Sigma_j^0)^{-1} (y_t - x_t' \beta_j^0) + \sum_{j=1}^{m+1} \frac{T_j^0 - T_{j-1}^0}{2} \log |\Sigma_j^0| \end{aligned}$$

Simple algebra applied to the first two terms reveals that

$$lr_T = \frac{1}{2} \sum_{j=1}^m \sum_{t=T_j}^{T_j^0} (y_t - x_t' \beta_j)' \Sigma_j^{-1} (y_t - x_t' \beta_j) \tag{I}$$

$$- \frac{1}{2} \sum_{j=1}^m \sum_{t=T_j}^{T_j^0} (y_t - x_t' \beta_{j+1})' \Sigma_{j+1}^{-1} (y_t - x_t' \beta_{j+1}) \tag{II}$$

$$+ \sum_{j=1}^m \frac{T_j^0 - T_j}{2} (\log |\Sigma_j| - \log |\Sigma_{j+1}|) \tag{III}$$

$$+ \sum_{j=1}^{m+1} lr_j^2$$

with lr_j^2 as defined in Theorem 1. Hence, we need to show that the sum of the terms (I), (II) and (III) is asymptotically equivalent to $\sum_{j=1}^m lr_j^1 (T_j - T_j^0)$ on the set C_M . Define the following variables, $\beta_j^* = \sqrt{T} |\beta_j - \beta_j^0|$ and $\Sigma_j^* = \sqrt{T} |\Sigma_j - \Sigma_j^0|$ (for $j = 1, \dots, m+1$) and

note that both are $O_p(1)$. Term (I) is then equivalent to

$$\begin{aligned}
& \frac{1}{2} \sum_{j=1}^m \sum_{t=T_j}^{T_j^0} (y_t - x_t' \beta_j^0 - T^{-1/2} x_t' \beta_j^*)' (\Sigma_j^0 + T^{-1/2} \Sigma_j^*)^{-1} (y_t - x_t' \beta_j^0 - T^{-1/2} x_t' \beta_j^*) \\
&= \frac{1}{2} \sum_{j=1}^m \sum_{t=T_j}^{T_j^0} (u_t - T^{-1/2} x_t' \beta_j^*)' (\Sigma_j^0 + T^{-1/2} \Sigma_j^*)^{-1} (u_t - T^{-1/2} x_t' \beta_j^*) \\
&= \frac{1}{2} \sum_{j=1}^m \sum_{t=T_j}^{T_j^0} u_t' (\Sigma_j^0 + T^{-1/2} \Sigma_j^*)^{-1} u_t + o_p(1)
\end{aligned}$$

The last equality follows because

$$\begin{aligned}
T^{-1/2} \sum_{t=T_j}^{T_j^0} (x_t' \beta_j^*)' (\Sigma_j^0 + T^{-1/2} \Sigma_j^*)^{-1} u_t &= T^{-1/2} v_T^{-1} (v_T \sum_{t=T_j}^{T_j^0} (x_t' \beta_j^*)' (\Sigma_j^0 + T^{-1/2} \Sigma_j^*)^{-1} u_t) \\
&= T^{-1/2} v_T^{-1} O_p(1) = o_p(1)
\end{aligned}$$

and

$$\begin{aligned}
& \sum_{t=T_j}^{T_j^0} T^{-1/2} (x_t' \beta_j^*)' (\Sigma_j^0 + T^{-1/2} \Sigma_j^*)^{-1} T^{-1/2} (x_t' \beta_j^*) \\
&= T^{-1} v_T^{-2} (v_T^2 \sum_{t=T_j}^{T_j^0} (x_t' \beta_j^*)' (\Sigma_j^0 + T^{-1/2} \Sigma_j^*)^{-1} (x_t' \beta_j^*)) = T^{-1} v_T^{-2} O_p(1) = o_p(1)
\end{aligned}$$

Further, using the fact that

$$\begin{aligned}
u_t' (\Sigma_j^0 + T^{-1/2} \Sigma_j^*)^{-1} u_t &= \text{tr}((I + T^{-1/2} (\Sigma_j^0)^{-1} \Sigma_j^*)^{-1} (\Sigma_j^0)^{-1} u_t u_t') \\
&= \text{tr}((I - T^{-1/2} (\Sigma_j^0)^{-1} \Sigma_j^* + O_p(T^{-1})) (\Sigma_j^0)^{-1} u_t u_t')
\end{aligned}$$

we deduce that

$$(I) = \frac{1}{2} \sum_{j=1}^m \sum_{t=T_j}^{T_j^0} \text{tr}(((\Sigma_j^0)^{-1} - T^{-1/2} (\Sigma_j^0)^{-1} \Sigma_j^* (\Sigma_j^0)^{-1}) u_t u_t') + o_p(1) \quad (\text{A.13})$$

For term (II), note that

$$\begin{aligned}
& (II) \tag{A.14} \\
&= -\frac{1}{2} \sum_{j=1}^m \sum_{t=T_j}^{T_j^0} (u_t + x_t'(\beta_j^0 - \beta_{j+1}^0) - T^{-1/2}x_t'\beta_{j+1}^*)'(\Sigma_{j+1}^0 + T^{-1/2}\Sigma_{j+1}^*)^{-1} \\
&\quad \times (u_t + x_t'(\beta_j^0 - \beta_{j+1}^0) - T^{-1/2}x_t'\beta_{j+1}^*) \\
&= -\frac{1}{2} \sum_{j=1}^m \sum_{t=T_j}^{T_j^0} (u_t + x_t'(\beta_j^0 - \beta_{j+1}^0))'(\Sigma_{j+1}^0 + T^{-1/2}\Sigma_{j+1}^*)^{-1} \\
&\quad \times (u_t + x_t'(\beta_j^0 - \beta_{j+1}^0)) + o_p(1) \\
&= -\frac{1}{2} \sum_{j=1}^m \sum_{t=T_j}^{T_j^0} (u_t + x_t'(\beta_j^0 - \beta_{j+1}^0))'[(\Sigma_{j+1}^0)^{-1} - T^{-1/2}(\Sigma_{j+1}^0)^{-1}\Sigma_{j+1}^*(\Sigma_{j+1}^0)^{-1}] \\
&\quad \times (u_t + x_t'(\beta_j^0 - \beta_{j+1}^0)) + o_p(1) \\
&= -\frac{1}{2} \sum_{j=1}^m \sum_{t=T_j}^{T_j^0} \text{tr}(((\Sigma_{j+1}^0)^{-1} - T^{-1/2}(\Sigma_{j+1}^0)^{-1}\Sigma_{j+1}^*(\Sigma_{j+1}^0)^{-1})u_t u_t') \\
&\quad - \frac{1}{2} \sum_{j=1}^m \sum_{t=T_j}^{T_j^0} (\beta_j^0 - \beta_{j+1}^0)'x_t(\Sigma_{j+1}^0)^{-1}x_t'(\beta_j^0 - \beta_{j+1}^0) \\
&\quad - \sum_{i=1}^m \sum_{T_j}^{T_j^0} (\beta_j^0 - \beta_{j+1}^0)'x_t(\Sigma_{j+1}^0)^{-1}u_t + o_p(1)
\end{aligned}$$

For term (III),

$$\begin{aligned}
& \frac{1}{2} \sum_{j=1}^m (T_j^0 - T_j)(\log |\Sigma_j| - \log |\Sigma_{j+1}|) \tag{A.15} \\
&= \frac{1}{2} \sum_{j=1}^m (T_j^0 - T_j)(\log |\Sigma_j^0| + T^{-1/2}(\Sigma_j^0)^{-1}\Sigma_j^* - \log |\Sigma_{j+1}^0| - T^{-1/2}(\Sigma_{j+1}^0)^{-1}\Sigma_{j+1}^* + O_p(T^{-1})) \\
&= \frac{1}{2} \sum_{j=1}^m (T_j^0 - T_j)(\log |\Sigma_j^0| - \log |\Sigma_{j+1}^0|) + \frac{1}{2}T^{-1/2} \sum_{j=1}^m (T_j^0 - T_j)(\Sigma_j^0)^{-1}\Sigma_j^* \\
&\quad - \frac{1}{2}T^{-1/2} \sum_{j=1}^m (T_j^0 - T_j)(\Sigma_{j+1}^0)^{-1}\Sigma_{j+1}^* + o_p(1)
\end{aligned}$$

Collecting the results in (A.13), (A.14) and (A.15), we have

$$\begin{aligned}
& (I) + (II) + (III) \\
= & \frac{1}{2} \sum_{j=1}^m \sum_{t=T_j}^{T_j^0} \text{tr}(((\Sigma_j^0)^{-1} - T^{-1/2}(\Sigma_j^0)^{-1}\Sigma_j^*(\Sigma_j^0)^{-1})u_t u_t') \\
& - \frac{1}{2} \sum_{j=1}^m \sum_{t=T_j}^{T_j^0} \text{tr}(((\Sigma_{j+1}^0)^{-1} - T^{-1/2}(\Sigma_{j+1}^0)^{-1}\Sigma_{j+1}^*(\Sigma_{j+1}^0)^{-1})u_t u_t') \\
& - \frac{1}{2} \sum_{j=1}^m \sum_{t=T_j}^{T_j^0} (\beta_j^0 - \beta_{j+1}^0)' x_t (\Sigma_{j+1}^0)^{-1} x_t' (\beta_j^0 - \beta_{j+1}^0) - \sum_{j=1}^m \sum_{t=T_j}^{T_j^0} (\beta_j^0 - \beta_{j+1}^0)' x_t (\Sigma_{j+1}^0)^{-1} u_t \\
& + \frac{1}{2} \sum_{j=1}^m (T_j^0 - T_j) (\log |\Sigma_j^0| - \log |\Sigma_{j+1}^0|) + \frac{1}{2} T^{-1/2} \sum_{i=1}^m (T_j^0 - T_j) (\Sigma_j^0)^{-1} \Sigma_j^* \\
& - \frac{1}{2} T^{-1/2} \sum_{j=1}^m (T_j^0 - T_j) (\Sigma_{j+1}^0)^{-1} \Sigma_{j+1}^* + o_p(1) \\
= & \frac{1}{2} \sum_{j=1}^m \sum_{t=T_j}^{T_j^0} u_t' ((\Sigma_j^0)^{-1} - (\Sigma_{j+1}^0)^{-1}) u_t - \frac{1}{2} \sum_{j=1}^m \sum_{t=T_j}^{T_j^0} (\beta_j^0 - \beta_{j+1}^0)' x_t (\Sigma_{j+1}^0)^{-1} x_t' (\beta_j^0 - \beta_{j+1}^0) \\
& + \frac{1}{2} \sum_{j=1}^m (T_j^0 - T_j) (\log |\Sigma_j^0| - \log |\Sigma_{j+1}^0|) - \sum_{j=1}^m \sum_{t=T_j}^{T_j^0} (\beta_j^0 - \beta_{j+1}^0)' x_t (\Sigma_{j+1}^0)^{-1} u_t \\
& - \frac{1}{2} T^{-1/2} \text{tr}(((\Sigma_j^0)^{-1} \Sigma_j^* (\Sigma_j^0)^{-1}) \sum_{j=1}^m \sum_{t=T_j}^{T_j^0} (u_t u_t' - \Sigma_j^0)) \\
& + \frac{1}{2} T^{-1/2} \text{tr}(((\Sigma_{j+1}^0)^{-1} \Sigma_{j+1}^* (\Sigma_{j+1}^0)^{-1}) \sum_{j=1}^m \sum_{t=T_j}^{T_j^0} (u_t u_t' - \Sigma_{j+1}^0)) + o_p(1)
\end{aligned}$$

The result follows using the fact that the last two terms are $o_p(1)$, since

$$\begin{aligned}
& T^{-1/2} \text{tr}(((\Sigma_j^0)^{-1} \Sigma_j^* (\Sigma_j^0)^{-1}) \sum_{j=1}^m \sum_{t=T_j}^{T_j^0} (u_t u_t' - \Sigma_j^0)) \\
= & T^{-1/2} v_T^{-1} \text{tr}(((\Sigma_j^0)^{-1} \Sigma_j^* (\Sigma_j^0)^{-1}) v_T \sum_{j=1}^m \sum_{t=T_j}^{T_j^0} (u_t u_t' - \Sigma_j^0)) \\
= & T^{-1/2} v_T^{-1} O_p(1) = o_p(1)
\end{aligned}$$

and

$$\begin{aligned}
& T^{-1/2} \text{tr}(((\Sigma_{j+1}^0)^{-1} \Sigma_{j+1}^* (\Sigma_{j+1}^0)^{-1}) \sum_{j=1}^m \sum_{t=T_j}^{T_j^0} (u_t u_t' - \Sigma_{j+1}^0)) \\
= & T^{-1/2} \text{tr}(((\Sigma_{j+1}^0)^{-1} \Sigma_{j+1}^* (\Sigma_{j+1}^0)^{-1}) \sum_{j=1}^m \sum_{t=T_j}^{T_j^0} (u_t u_t' - \Sigma_j^0)) \\
& + T^{-1/2} \text{tr}(((\Sigma_{j+1}^0)^{-1} \Sigma_{j+1}^* (\Sigma_{j+1}^0)^{-1}) \sum_{j=1}^m \sum_{t=T_j}^{T_j^0} (\Sigma_j^0 - \Sigma_{j+1}^0)) = o_p(1)
\end{aligned}$$

The proof of Theorem 3 requires the following Lemma whose proof is direct and hence omitted.

Lemma A.4 *Let $\eta_t = (\Sigma_j^0)^{-1/2} u_t$, under Assumptions A4-A5, with v_T a sequence of positive numbers satisfying $v_T \rightarrow 0$ and $T^{1/2} v_T / (\log T)^2 \rightarrow \infty$, we have*

$$\begin{aligned}
& \text{for } s < 0, v_T \sum_{t=T_j^0 + [s/v_T^2]}^{T_j^0} (\eta_t \eta_t' - I) \Rightarrow \xi_{1,j}(s) \\
& \text{for } s > 0, v_T \sum_{t=T_j^0}^{T_j^0 + [s/v_T^2]} (\eta_t \eta_t' - I) \Rightarrow \xi_{2,j}(s)
\end{aligned}$$

where the weak convergence is in the space $D[0, \infty)^{n^2}$ and where the entries of the $n \times n$ matrices $\xi_{1,j}(s)$ and $\xi_{2,j}(s)$ are Brownian motion processes defined on the real line. Also

$$\begin{aligned}
& \text{for } s < 0, v_T \sum_{t=T_j^0 + [s/v_T^2]}^{T_j^0} x_t (\Sigma_{j+1}^0)^{-1} u_t \Rightarrow (\Pi_{1,j})^{1/2} \zeta_{1,j}(s) \\
& \text{for } s > 0, v_T \sum_{t=T_j^0}^{T_j^0 + [s/v_T^2]} x_t (\Sigma_j^0)^{-1} u_t \Rightarrow (\Pi_{2,j})^{1/2} \zeta_{2,j}(s)
\end{aligned}$$

where the weak convergence is in the space $D[0, \infty)^p$ and where the entries of the p vectors $\zeta_{1,j}(s)$ and $\zeta_{2,j}(s)$ are independent Wiener processes defined on the real line. Also, $\zeta_{1,j}(s)$ and $\zeta_{2,j}(s)$ (resp., $\xi_{1,j}(s)$ and $\xi_{2,j}(s)$) are different independent copies for $j = 1, \dots, m$. Note that $\zeta_{1,j}(s)$ (resp., $\zeta_{2,j}(s)$) and $\xi_{1,j}(s)$ (resp., $\xi_{2,j}(s)$) are not necessarily independent unless $E[\eta_{tk} \eta_{tl} \eta_{th}] = 0$ for all k, l, h and for every t .

Proof of Theorem 3: Without loss of generality, consider the j^{th} break date and start with the case where the candidate estimate is before the true break date. We obtain an expansion for $lr_j^1([s/v_T^2])$ as defined in Theorem 1. Note that s is implicitly defined by $s = v_T^2(T_i - T_i^0) = rv_T^2$. We deal with each term separately. For the first term, we have

$$\begin{aligned}
& \frac{1}{2} \sum_{t=T_j^0+[s/v_T^2]}^{T_j^0} u_t'((\Sigma_j^0)^{-1} - (\Sigma_{j+1}^0)^{-1})u_t \\
&= \frac{1}{2} \sum_{t=T_j^0+[s/v_T^2]}^{T_j^0} tr(((\Sigma_j^0)^{-1} - (\Sigma_{j+1}^0)^{-1})(u_t u_t' - \Sigma_j^0 + \Sigma_{j+1}^0)) \\
&= \frac{1}{2} \sum_{t=T_j^0+[s/v_T^2]}^{T_j^0} tr(((\Sigma_j^0)^{-1} - (\Sigma_{j+1}^0)^{-1})(u_t u_t' - \Sigma_j^0)) - \frac{r}{2} tr(((\Sigma_j^0)^{-1} - (\Sigma_{j+1}^0)^{-1})\Sigma_j^0) \\
&= \frac{1}{2} tr((\Sigma_j^0)^{1/2}(\Sigma_{j+1}^0)^{-1}(\Sigma_{j+1}^0 - \Sigma_j^0)(\Sigma_j^0)^{-1/2} \sum_{t=T_j^0+[s/v_T^2]}^{T_j^0} (\eta_t \eta_t' - I)) - \frac{r}{2} tr((\Sigma_{j+1}^0)^{-1}(\Sigma_{j+1}^0 - \Sigma_j^0)) \\
&= \frac{1}{2} tr((\Sigma_j^0)^{1/2}(\Sigma_{j+1}^0)^{-1}\Phi_j(\Sigma_j^0)^{-1/2}v_T \sum_{t=T_j^0+[s/v_T^2]}^{T_j^0} (\eta_t \eta_t' - I)) - \frac{r}{2} v_T tr((\Sigma_{j+1}^0)^{-1}\Phi_j)
\end{aligned}$$

For the second term, we have:

$$\begin{aligned}
-\frac{r}{2}(\log |\Sigma_j^0| - \log |\Sigma_{j+1}^0|) &= -\frac{r}{2} \log |(\Sigma_j^0 - \Sigma_{j+1}^0 + \Sigma_{j+1}^0)(\Sigma_{j+1}^0)^{-1}| \\
&= -\frac{r}{2} \log |I + (\Sigma_j^0 - \Sigma_{j+1}^0)(\Sigma_{j+1}^0)^{-1}| \\
&= \frac{r}{2} tr((\Sigma_{j+1}^0 - \Sigma_j^0)(\Sigma_{j+1}^0)^{-1}) + \frac{r}{4} tr([(\Sigma_{j+1}^0 - \Sigma_j^0)(\Sigma_{j+1}^0)^{-1}]^2) \\
&= \frac{r}{2} v_T tr(\Phi_j(\Sigma_{j+1}^0)^{-1}) + \frac{r}{4} v_T^2 tr([\Phi_j(\Sigma_{j+1}^0)^{-1}]^2)
\end{aligned}$$

Then the sum of the first two terms is

$$\begin{aligned}
& \frac{1}{2} \sum_{T_j^0+[s/v_T^2]}^{T_j^0} u_t'((\Sigma_j^0)^{-1} - (\Sigma_{j+1}^0)^{-1})u_t - \frac{r}{2}(\log |\Sigma_j^0| - \log |\Sigma_{j+1}^0|) \\
&= \frac{1}{2} tr((\Sigma_j^0)^{1/2}(\Sigma_{j+1}^0)^{-1}\Phi_j(\Sigma_j^0)^{-1/2}v_T \sum_{T_j^0+[s/v_T^2]}^{T_j^0} (\eta_t \eta_t' - I)) + \frac{r}{4} v_T^2 tr([\Phi_j(\Sigma_{j+1}^0)^{-1}]^2) \\
&\Rightarrow \frac{1}{2} tr((\Sigma_j^0)^{1/2}(\Sigma_{j+1}^0)^{-1}\Phi_j(\Sigma_j^0)^{-1/2}\xi_{1,j}(s)) + \frac{s}{4} tr([\Sigma_{j+1}^0]^{-1}\Phi_j]^2) \\
&= \frac{1}{2} tr(A_{1,j}\xi_{1,j}(s)) + \frac{s}{4} tr(A_{1,j}^2)
\end{aligned}$$

where $\xi_{1,j}(s)$ is a non-standard Brownian motion process, with $\text{var} [\text{vec}(\xi_{1,j}(s))] = \Omega_{1,j}^0$. Then, for the third term,

$$-\frac{1}{2} \sum_{t=T_j^0+[s/v_T^2]}^{T_j^0} (\beta_j^0 - \beta_{j+1}^0)' x_t (\Sigma_{j+1}^0)^{-1} x_t' (\beta_j^0 - \beta_{j+1}^0) \rightarrow_p \frac{1}{2} s \delta_j' Q_{1,j} \delta_j$$

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

$$- \sum_{T_j^0+[s/v_T^2]}^{T_j^0} (\beta_j^0 - \beta_{j+1}^0)' x_t (\Sigma_{j+1}^0)^{-1} u_t \Rightarrow \delta_j' (\Pi_{1,j})^{1/2} \zeta_{1,j}(s)$$

with

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

Combining the above results, we have for $s < 0$:

$$lr_j^1([s/v_T^2]) \Rightarrow -\frac{|s|}{2} \left[\frac{1}{2} \text{tr}(A_{1,j}^2) + \delta_j' Q_{1,j} \delta_j \right] + \frac{1}{2} \text{vec}(A_{1,j})' \text{vec}(\xi_{1,j}(s)) + \delta_j' (\Pi_{1,j})^{1/2} \zeta_{1,j}(s)$$

Now, $\text{vec}(A_{1,j})' \text{vec}(\xi_{1,j}(s)) \stackrel{d}{=} (\text{vec}(A_{1,j})' \Omega_{1,j}^0 \text{vec}(A_{1,j}))^{1/2} V_{1,j}(s)$ where $V_{1,j}(s)$ is a standard Wiener process. Similarly, $\delta_j' (\Pi_{1,j})^{1/2} \zeta_{1,j}(s) \stackrel{d}{=} (\delta_j' (\Pi_{1,j}) \delta_j)^{1/2} U_{1,j}(s)$ where $U_{1,j}(s)$ is a standard Wiener process. With the stated conditions, $V_{1,j}(s)$ and $U_{1,j}(s)$ are independent. Then

$$\begin{aligned} & (\text{vec}(A_{1,j})' \Omega_{1,j}^0 \text{vec}(A_{1,j}) / 4)^{1/2} V_{1,j}(s) + (\delta_j' (\Pi_{1,j}) \delta_j)^{1/2} U_{1,j}(s) \\ & \stackrel{d}{=} (\text{vec}(A_{1,j})' \Omega_{1,j}^0 \text{vec}(A_{1,j}) / 4 + \delta_j' (\Pi_{1,j}) \delta_j)^{1/2} B_{1,j}(s) \equiv \Gamma_{1,j} B_{1,j}(s) \end{aligned}$$

where $B(s)$ is a unit Wiener process. Hence, with $\Delta_{1,j} = \text{tr}(A_{1,j}^2)/2 + \delta_j Q_{1,j} \delta_j$, we have

$$lr_j^1([s/v_T^2]) \Rightarrow -\frac{|s|}{2} \Delta_{1,j} + \Gamma_{1,j} B_{1,j}(s)$$

The proof for the case $s > 0$ is similar in which case

$$lr_j^1([s/v_T^2]) \Rightarrow -\frac{|s|}{2} \Delta_{2,j} + \Gamma_{2,j} B_{1,j}(s)$$

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

$$\Gamma_{2,j} = [\text{vec}(A_{2,j})' \Omega_{2,j}^0 \text{vec}(A_{2,j}) / 4 + \delta_j' (\Pi_{2,j}) \delta_j]^{1/2}$$

We also have by definition $lr_j^1(0) = 0$. Now since $s = v_T^2(T_j - T_j^0)$, the $\arg\max$ yields the scaled estimate $v_T^2(\hat{T}_j - T_j^0)$. The result follows since we can take the $\arg\max$ over the compact set C_M and with the use of Lemma this is equivalent to taking the $\arg\max$ in an unrestricted set since with probability arbitrarily close to one the estimates will be contained in C_M . Hence,

$$v_T^2(\hat{T}_j - T_j^0) \Rightarrow \arg\max_s \begin{cases} -\frac{|s|}{2}\Delta_{1,j} + \Gamma_{1,j}B_j(s) & \text{for } s \leq 0 \\ -\frac{|s|}{2}\Delta_{2,j} + \Gamma_{2,j}B_j(s) & \text{for } s > 0 \end{cases}$$

where $B_j(s) = B_{1,j}(s)$ for $s \leq 0$ and $B_j(s) = B_{2,j}(s)$ for $s > 0$. Multiplying by $\Delta_{1,j}/\Gamma_{1,j}^2$ and applying a change of variable with $u = (\Delta_{1,j}^2/\Gamma_{1,j}^2)s$, we obtain Theorem 4.

Proof of Lemma 2. For simplicity, we assume the model only contains two breaks, which are locally ordered and, accordingly, drop the subscript j . We shall need the following additional property, whose proof is omitted.

Lemma A.5 *Consider the following data generating process*

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = x'_t \begin{pmatrix} \beta_{1,1}^0 \mathbf{1}_{(t \leq k_1^0)} + \beta_{1,2}^0 \mathbf{1}_{(t > k_1^0)} \\ \beta_{2,1}^0 \mathbf{1}_{(t \leq k_2^0)} + \beta_{2,2}^0 \mathbf{1}_{(t > k_2^0)} \end{pmatrix} + U_t$$

where (k_1^0, k_2^0) satisfies Assumption LOB and $\|\beta_{1,1}^0 - \beta_{1,2}^0\| \leq Mv_T$, $\|\beta_{2,1}^0 - \beta_{2,2}^0\| \leq Mv_T$ for some $M > 0$. Let $\hat{\beta}(k_1, k_2)$ and $\hat{\Sigma}(k_1, k_2)$ denote the estimates under the partition (k_1, k_2) . If $|k_1 - k_1^0| \leq \sqrt{T}v_T^{-1}$ and $|k_2 - k_2^0| \leq \sqrt{T}v_T^{-1}$, then

$$\begin{aligned} \hat{\beta}(k_1, k_2) - \beta^0 &= O_p(T^{-1/2}) \\ \hat{\Sigma}(k_1, k_2) - \Sigma^0 &= O_p(T^{-1/2}) \end{aligned}$$

where the bounds are uniform in (k_1, k_2) .

First, we derive the rate of convergence for the case where no restriction is imposed. The result for the restricted model then follows using the same argument as in the proof of Lemma 1. We want to prove that

$$\Pr(|\hat{k}_i - k_i^0| > \sqrt{T}v_T^{-1}) < \epsilon \tag{A.16}$$

for $i = 1, 2$. For this, it suffices to prove that, for large T ,

$$\Pr\left(\sup_{(k_1, k_2) \in \tilde{\Lambda}_\epsilon} \mathcal{L}_T^*(k_1, k_2) > \epsilon\right) < \epsilon$$

where

$$\tilde{\Lambda}_\epsilon = \left\{ (k_1, k_2) : \epsilon T \leq k_1 \leq k_2 \leq (1 - \epsilon)T; |k_1 - k_1^0| > \sqrt{T}v_T^{-1} \text{ or } |k_2 - k_2^0| > \sqrt{T}v_T^{-1} \right\}$$

and $\mathcal{L}_T^*(k_1, k_2)$ is the value of the likelihood ratio under the ordered breaks parametrization. Define the following subset of $\tilde{\Lambda}_\varepsilon$

$$\tilde{\Lambda}_{\varepsilon,1} = \left\{ (k_1, k_2) : \varepsilon T \leq k_1 \leq k_2 \leq (1 - \varepsilon)T; |k_1^0 - k_1| > \sqrt{T}v_T^{-1} \right\}$$

We want to show that for large T , $\Pr(\sup_{(k_1, k_2) \in \tilde{\Lambda}_{\varepsilon,1}} \mathcal{L}_T^*(k_1, k_2) > \varepsilon) < \varepsilon$. With this result, the behavior of the likelihood ratio on $(\tilde{\Lambda}_\varepsilon \setminus \tilde{\Lambda}_{\varepsilon,1})$ can be analyzed similarly. Note that the set $\tilde{\Lambda}_{\varepsilon,1}$ consists the following four cases: Case 1: $k_1^0 - k_1 > \sqrt{T}v_T^{-1}$, $0 < k_1^0 - k_2 < \sqrt{T}v_T^{-1}/2$; Case 2: $k_1^0 - k_1 > \sqrt{T}v_T^{-1}$, $k_1^0 - k_2 \geq \sqrt{T}v_T^{-1}/2$; Case 3: $k_1^0 - k_1 > \sqrt{T}v_T^{-1}$, $k_1^0 \leq k_2 \leq k_2^0$; and Case 4: $k_1^0 - k_1 > \sqrt{T}v_T^{-1}$, $k_2 > k_2^0$. We will analyze the function $\mathcal{L}_T^*(k_1, k_2)$ case by case. For Case 1, the likelihood ratio under a given partition, (k_1, k_2, k_1^0, k_2^0) can be written as the product of 5 terms and, using Property 2, we have

$$\mathcal{L}_T^*(k_1, k_2) = \mathcal{L}(k_1 + 1, k_2) \mathcal{L}(k_2^0 + 1, T) O_p(T^{3B})$$

for some $B > 0$. Hence, under Case 1,

$$\sup_{(k_1, k_2) \in \tilde{\Lambda}_{\varepsilon,1}} \mathcal{L}_T^*(k_1, k_2) = \left[\sup_{(k_1, k_2) \in \tilde{\Lambda}_{\varepsilon,1}} \mathcal{L}(k_1 + 1, k_2) \mathcal{L}(k_2^0 + 1, T) \right] O_p(T^{3B})$$

Using arguments as in Bai (2000, pp. 325-6), we have

$$\sup_{(k_1, k_2) \in \tilde{\Lambda}_{\varepsilon,1}} [\mathcal{L}(k_1 + 1, k_2) \mathcal{L}(k_2^0 + 1, T)] = O_p(T^{-A} \log T)$$

holding for every $A > 0$. Hence,

$$\sup_{(k_1, k_2) \in \tilde{\Lambda}_{\varepsilon,1}} \mathcal{L}_T^*(k_1, k_2) = O_p(T^{3B-A} \log T) = o_p(1)$$

for A large enough. For Case 2, the proof follows immediately using the fact that $\mathcal{L}_T^*(k_1, k_2) = \mathcal{L}(k_2 + 1, k_1^0) \mathcal{L}(k_2^0 + 1, T) O_p(T^{3B})$ and

$$\sup_{(k_1, k_2) \in \tilde{\Lambda}_{\varepsilon,1}} [\mathcal{L}(k_2 + 1, k_1^0) \mathcal{L}(k_2^0 + 1, T)] = O_p(T^{-A} \log T)$$

For Case 3, we use $\mathcal{L}_T^*(k_1, k_2) = \mathcal{L}(k_1 + 1, k_1^0) \mathcal{L}(k_2^0 + 1, T) O_p(T^{3B})$ and

$$\sup_{(k_1, k_2) \in \tilde{\Lambda}_{\varepsilon,1}} [\mathcal{L}(k_1 + 1, k_1^0) \mathcal{L}(k_2^0 + 1, T)] = O_p(T^{-A} \log T)$$

For Case 4, we use $\mathcal{L}_T^*(k_1, k_2) = \mathcal{L}(k_1 + 1, k_1^0) \mathcal{L}(k_2 + 1, T) O_p(T^{3B})$ and

$$\sup_{(k_1, k_2) \in \tilde{\Lambda}_{\varepsilon,1}} [\mathcal{L}(k_1 + 1, k_1^0) \mathcal{L}(k_2 + 1, T)] = O_p(T^{-A} \log T)$$

Hence, (A.16) is proved. It still remains to show that for any $\epsilon > 0$, there exists a $C > 0$, such that

$$\Pr\left(\sup_{(k_1, k_2) \in \tilde{\Lambda}_\epsilon^c} \frac{\mathcal{L}_T^*(k_1, k_2)}{\mathcal{L}_T^*(k_1^0, k_2^0)} > \epsilon\right) < \epsilon \quad (\text{A.17})$$

where

$$\tilde{\Lambda}_\epsilon^c = \{(k_1, k_2) : (k_1, k_2) \notin \tilde{\Lambda}_\epsilon, |k_1 - k_1^0| > cv_T^{-2} \text{ or } |k_2 - k_2^0| > Cv_T^{-2}\}$$

Define the following subset of $\tilde{\Lambda}_\epsilon^c$

$$\tilde{\Lambda}_{\epsilon,1}^c = \{(k_1, k_2) : (k_1, k_2) \notin \tilde{\Lambda}_\epsilon, |k_1^0 - k_1| > Cv_T^{-2}\}$$

To prove (A.16), it is sufficient to show that

$$\Pr\left(\sup_{(k_1, k_2) \in \tilde{\Lambda}_{\epsilon,1}^c} \frac{\mathcal{L}_T^*(k_1, k_2)}{\mathcal{L}_T^*(k_1^0, k_2^0)} > \epsilon\right) < \epsilon$$

Noting that

$$\frac{\mathcal{L}_T^*(k_1, k_2)}{\mathcal{L}_T^*(k_1^0, k_2^0)} = \frac{\mathcal{L}_T^*(k_1, k_2)}{\mathcal{L}_T^*(k_1^0, k_2)} \frac{\mathcal{L}_T^*(k_1^0, k_2)}{\mathcal{L}_T^*(k_1^0, k_2^0)}$$

it is then sufficient to show that

$$\Pr\left(\sup_{(k_1, k_2) \in \tilde{\Lambda}_{\epsilon,1}^c} \frac{\mathcal{L}_T^*(k_1, k_2)}{\mathcal{L}_T^*(k_1^0, k_2)} > \epsilon\right) < \epsilon \quad (\text{A.18})$$

and

$$\Pr\left(\sup_{(k_1, k_2) \in \tilde{\Lambda}_{\epsilon,1}^c} \frac{\mathcal{L}_T^*(k_1^0, k_2)}{\mathcal{L}_T^*(k_1^0, k_2^0)} > \epsilon\right) < \epsilon \quad (\text{A.19})$$

Here, we only give the outline of the proof for (A.18). The details are omitted since they only involve repeating the same arguments. First, using Property 11, we can show that the estimates of the coefficients are \sqrt{T} consistent. Then, we can apply the arguments of Bai (2000, pp. 327-8) to the three cases analyzed before, with $\sqrt{T}v_T^{-1}$ replaced by Cv_T^{-2} , and get the desired result. Similar arguments lead to (A.19).

Proof of Theorem 4: Here the relevant expressions for the likelihood function correspond that of Theorem 2 with $m + 1$ breaks, the j^{th} and the $(j + 1)^{\text{th}}$ being the locally ordered ones. Hence, we need to examine $lr_j^1(r_1) + lr_{j+1}^1(r_2)$ in order to derive the joint limiting distribution for the pair $r_1 = \hat{k}_{1,j} - k_{1,j}^0$ and $r_2 = \hat{k}_{2,j} - k_{2,j}^0$. Depending on the locations of the candidate breaks relative to the true breaks, the following four cases need to be considered: Case 1: $r_1 < 0, r_2 < 0, r_1 - r_2 \leq k_{2,j}^0 - k_{1,j}^0$; Case 2: $r_1 < 0, r_2 > 0$; Case 3: $r_1 > 0, r_2 > 0, r_1 - r_2 \leq k_{2,j}^0 - k_{1,j}^0$; Case 4: $r_1 \geq 0, r_2 \leq 0, r_1 - r_2 \leq k_{2,j}^0 - k_{1,j}^0$. For Case 4, $lr_j^1(r_1) + lr_{j+1}^1(r_2) = o_p(1)$ using the assumption LOB. For the other three cases, define $H_{T,j}(r_1, r_2) = lr_j^1(r_1) + lr_{j+1}^1(r_2)$, we have,

Case 1: $r_1 < 0, r_2 < 0, r_1 - r_2 \leq k_{2,j}^0 - k_{1,j}^0$, then,

$$\begin{aligned}
& H_{T,j}(r_1, r_2) \\
&= (\Delta\beta'_{1,j}, 0) \sum_{t=k_{1,j}^0+r_1+1}^{k_{1,j}^0} x_t(\Sigma_{j,j+1}^0)^{-1} u_t - \frac{1}{2} (\Delta\beta'_{1,j}, 0) \sum_{t=k_{1,j}^0+r_1+1}^{k_{1,j}^0} x_t(\Sigma_{j,j+1}^0)^{-1} x'_t (\Delta\beta'_{1,j}, 0)' \\
&\quad + (0, \Delta\beta'_{2,j}) \sum_{t=k_{2,j}^0+r_2+1}^{k_{2,j}^0} x_t(\Sigma_{j,j+1}^0)^{-1} u_t - \frac{1}{2} (0, \Delta\beta'_{2,j}) \sum_{t=k_{2,j}^0+r_2+1}^{k_{2,j}^0} x_t(\Sigma_{j,j+1}^0)^{-1} x'_t (0, \Delta\beta'_{2,j})' \\
&\quad - (\Delta\beta'_{1,j}, 0) \sum_{t=k_{2,j}^0+r_2+1}^{k_{2,j}^0} x_t(\Sigma_{j,j+1}^0)^{-1} x'_t (0, \Delta\beta'_{2,j})' + o_p(1)
\end{aligned}$$

where $\Delta\beta'_{1,j} = (\beta_{1,j+1}^0 - \beta_{1,j}^0)'$, $\Delta\beta'_{2,j} = (\beta_{2,j+1}^0 - \beta_{2,j}^0)'$ and the reminder term is $o_p(1)$ under the Assumption LOB. Let $(\Delta\beta'_{1,j}, 0) = \delta'_{1,j} v_T$, $(0, \Delta\beta'_{2,j}) = \delta'_{2,j} v_T$, $Q_j = \text{plim}_{T \rightarrow \infty} (T_j^0 - T_{j-1}^0)^{-1} \sum_{t=T_{j-1}^0+1}^{T_j^0} x_t(\Sigma_{j,j+1}^0)^{-1} x'_t$ and $\Pi_{12,j} = \delta'_{1,j} Q_j \delta_{2,j}$. Using a Functional Central Limit Theorem,

$$\begin{aligned}
& H_{T,j}([s_1 v_T^{-2}], [s_2 v_T^{-2}]) \\
&\Rightarrow \delta'_{1,j} Q_j^{1/2} W_1(s_1) - \frac{1}{2} |s_1| \delta'_{1,j} Q_j \delta_{1,j} + \delta'_{2,j} Q_j^{1/2} W_2(s_2) - \frac{1}{2} |s_2| \delta'_{2,j} Q_j \delta_{2,j} - |s_2| \delta'_{1,j} Q_j \delta_{2,j} \\
&\stackrel{d}{=} (\Pi_{1,j})^{1/2} W_1(s_1) - \frac{|s_1|}{2} \Pi_{1,j} + (\Pi_{2,j})^{1/2} W_2(s_2) - \left(\frac{1}{2} \Pi_{2,j} + \Pi_{12,j}\right) |s_2| \\
&\equiv H_j^1(s_1, s_2)
\end{aligned}$$

since

$$\begin{aligned}
\delta'_{1,j} Q_j^{1/2} W_1(s_1) &\stackrel{d}{=} (\delta'_{1,j} Q_j \delta_{1,j})^{1/2} W_1(s_1) \equiv (\Pi_{1,j})^{1/2} W_1(s_1) \\
\delta'_{2,j} Q_j^{1/2} W_2(s_2) &\stackrel{d}{=} (\delta'_{2,j} Q_j \delta_{2,j})^{1/2} W_2(s_2) \equiv (\Pi_{2,j})^{1/2} W_2(s_2)
\end{aligned}$$

Case 2: $r_1 < 0, r_2 > 0$. We have

$$\begin{aligned}
& H_{T,j}(r_1, r_2) \\
&= (\Delta\beta'_{1,j}, 0) \sum_{t=k_{1,j}^0+r_1}^{k_{1,j}^0} x_t(\Sigma_{j,j+1}^0)^{-1} u_t - \frac{1}{2} (\Delta\beta'_{1,j}, 0) \sum_{t=k_{1,j}^0+r_1}^{k_{1,j}^0} x_t(\Sigma_{j,j+1}^0)^{-1} x'_t (\Delta\beta'_{1,j}, 0)' \\
&\quad - (0, \Delta\beta'_{2,j}) \sum_{t=k_{2,j}^0+1}^{k_{2,j}^0+r_2} x_t(\Sigma_{j,j+1}^0)^{-1} u_t - \frac{1}{2} (0, \Delta\beta'_{2,j}) \sum_{t=k_{2,j}^0+1}^{k_{2,j}^0+r_2} x_t(\Sigma_{j,j+1}^0)^{-1} x'_t (0, \Delta\beta'_{2,j})'
\end{aligned}$$

and

$$\begin{aligned} H_{T,j}([s_1 v_T^{-2}], [s_2 v_T^{-2}]) &\Rightarrow (\Pi_{1,j})^{1/2} W_1(s_1) - \frac{|s_1|}{2} \Pi_{1,j} + (\Pi_{2,j+1})^{1/2} W_3(s_2) - \frac{|s_2|}{2} \Pi_{2,j+1} \\ &\equiv H_j^2(s_1, s_2) \end{aligned}$$

with $Q_{j+1} = \text{plim}_{T \rightarrow \infty} (T_{j+1}^0 - T_j^0)^{-1} \sum_{t=T_j^0+1}^{T_{j+1}^0} x_t(\Sigma_{j,j+1}^0)^{-1} x_t'$, and $\Pi_{2,j+1} = \delta_{2,j}' Q_{j+1} \delta_{2,j}$.

Case 3: $r_1 > 0, r_2 > 0, r_1 - r_2 \leq k_2^0 - k_1^0$. We have

$$\begin{aligned} &H_{T,j}(r_1, r_2) \\ &= -(\Delta\beta'_{1,j}, 0) \sum_{t=k_{1,j}^0+1}^{k_{1,j}^0+r_1} x_t(\Sigma_{j,j+1}^0)^{-1} u_t - \frac{1}{2} (\Delta\beta'_{1,j}, 0) \sum_{t=k_{1,j}^0+1}^{k_{1,j}^0+r_1} x_t(\Sigma_{j,j+1}^0)^{-1} x_t' (\Delta\beta'_{1,j}, 0)' \\ &\quad - (0, \Delta\beta'_{2,j}) \sum_{t=k_{2,j}^0+1}^{k_{2,j}^0+r_2} x_t(\Sigma_{j,j+1}^0)^{-1} u_t - \frac{1}{2} (0, \Delta\beta'_{2,j}) \sum_{t=k_{2,j}^0+1}^{k_{2,j}^0+r_2} x_t(\Sigma_{j,j+1}^0)^{-1} x_t' (0, \Delta\beta'_{2,j})' \\ &\quad - (\Delta\beta'_{1,j}, 0) \sum_{t=k_{1,j}^0+1}^{k_{1,j}^0+r_1} x_t(\Sigma_{j,j+1}^0)^{-1} x_t' (0, \Delta\beta'_{2,j})' + o_p(1) \end{aligned}$$

and

$$\begin{aligned} &H_{T,j}([s_1 v_T^{-2}], [s_2 v_T^{-2}]) \\ &\Rightarrow (\Pi_{1,j+1})^{1/2} W_4(s_1) - \left(\frac{1}{2} \Pi_{1,j+1} + \Pi_{12,j+1}\right) |s_1| + (\Pi_{2,j+1})^{1/2} W_3(s_2) - \frac{|s_2|}{2} \Pi_{2,j+1} \\ &\equiv H_j^3(s_1, s_2) \end{aligned}$$

with $\Pi_{1,j+1} = \delta_{1,j}' Q_{j+1} \delta_{1,j}$ and $\Pi_{12,j+1} = \delta_{1,j}' Q_{j+1} \delta_{2,j}$.

Finally, define $H_j(s_1, s_2) = 0$ if $s_1 = 0, s_2 = 0$; $H_j(s_1, s_2) = H_j^1(s_1, s_2)$ if $s_1 < 0, s_2 < 0$ and $s_1 \leq s_2$; $H_j(s_1, s_2) = H_j^2(s_1, s_2)$ if $s_1 < 0$ and $s_2 > 0$; and $H_j(s_1, s_2) = H_j^3(s_1, s_2)$ if $s_1 > 0, s_2 > 0$ and $s_1 \leq s_2$. The limiting distribution of the estimated break dates is then

$$\left(v_T^2 (\hat{k}_{1,j} - k_{1,j}^0), v_T^2 (\hat{k}_{2,j} - k_{2,j}^0) \right) \Rightarrow \arg \max_{s_1 \leq s_2} H_j(s_1, s_2)$$

or, with a change of variable technique as in Bai (1997),

$$v_T^2 \Pi_{1,j} \left((\hat{k}_{1,j} - k_{1,j}^0), (\hat{k}_{2,j} - k_{2,j}^0) \right) \Rightarrow \arg \max_{v_1 \leq v_2} Z(v_1, v_2)$$

with $Z(v_1, v_2)$ equal to 0 if $v_1 = v_2 = 0$;

$$Z(v_1, v_2) = W_1(v_1) - \frac{|v_1|}{2} + (\Pi_{2,j}/\Pi_{1,j})^{1/2} W_2(v_2) - \frac{|v_2|}{2} (\Pi_{2,j} + 2\Pi_{12,j})/\Pi_{1,j}$$

if $v_1 < 0, v_2 < 0$ and $v_1 \leq v_2$;

$$Z(v_1, v_2) = W_1(v_1) - \frac{|v_1|}{2} + (\Pi_{2,j+1}/\Pi_{1,j})^{1/2}W_3(v_2) - \frac{|v_2|}{2}(\Pi_{2,j+1}/\Pi_{1,j})$$

if $v_1 < 0, v_2 > 0$; and

$$\begin{aligned} Z(v_1, v_2) &= (\Pi_{1,j+1}/\Pi_{1,j})^{1/2}W_4(v_1) - \frac{|v_1|}{2}(\Pi_{1,j+1} + 2\Pi_{12,j+1})/\Pi_{1,j} + (\Pi_{2,j+1}/\Pi_{1,j})^{1/2}W_3(v_2) \\ &\quad - \frac{|v_2|}{2}(\Pi_{2,j+1}/\Pi_{1,j}) \end{aligned}$$

if $v_1 > 0, v_2 > 0$ and $v_1 \leq v_2$. This completes the proof.

Proof of Theorem 5. As a matter of notation, let

$$\tilde{\Sigma}_{1,j} = \frac{1}{T_j} \sum_{t=1}^{T_j} (y_t - x'_{at}\tilde{\beta}_a - x'_{bt}\tilde{\beta}_{b1,j})(y_t - x'_{at}\tilde{\beta}_a - x'_{bt}\tilde{\beta}_{b1,j})' \quad (\text{A.20})$$

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

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

Also

$$\hat{\Sigma}_j = \frac{1}{T_j - T_{j-1}} \sum_{t=T_{j-1}+1}^{T_j} (y_t - x'_{at}\hat{\beta}_a - x'_{bt}\hat{\beta}_{bj})(y_t - x'_{at}\hat{\beta}_a - x'_{bt}\hat{\beta}_{bj})' \quad (\text{A.21})$$

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

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

For a given partition of the sample, we have

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

Consider a second-order Taylor series expansion of each term:

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

Hence,

$$\begin{aligned}
F_T^j &\equiv F_{1,T}^j + F_{2,T}^j \\
&= \text{tr} \left(T_{j+1}(\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j+1} - \Sigma^0) - T_j(\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j} - \Sigma^0) \right) \tag{A.22}
\end{aligned}$$

$$\begin{aligned}
&\quad - (T_{j+1} - T_j)(\Sigma^0)^{-1}(\hat{\Sigma}_{j+1} - \Sigma^0) \\
&\quad - \frac{1}{2} \text{tr} \left(T_{j+1}[(\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j+1} - \Sigma^0)]^2 \right) \tag{A.23} \\
&\quad - T_j[(\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j} - \Sigma^0)]^2 - (T_{j+1} - T_j)[(\Sigma^0)^{-1}(\hat{\Sigma}_{j+1} - \Sigma^0)]^2
\end{aligned}$$

Note that the first term in (A.22), denoted $F_{1,T}^j$, will be non-vanishing when allowance is made for changes in β^0 , while the second term, denoted $F_{2,T}^j$, will be non-vanishing when allowance is made for changes in Σ^0 .

We first consider $F_{1,T}^j$ and write the regression in matrix form to simplify the derivation. Under the null hypothesis, we have

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

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

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

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

$$\tilde{\beta}_a = [W' M_Z W]^{-1} W' M_Z Y^d \tag{A.24}$$

$$\tilde{\beta}_{b1,j} = (Z'_{1,j} Z_{1,j})^{-1} Z'_{1,j} (Y_{1,j}^d - W_{1,j} \tilde{\beta}_a) \tag{A.25}$$

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

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

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

$$\hat{\beta}_a = [\bar{W}'M_{\bar{Z}}\bar{W}]^{-1}\bar{W}'M_{\bar{Z}}\bar{Y}^d \quad (\text{A.26})$$

$$\hat{\beta}_{b,j+1} = (\bar{Z}'_{j+1}\bar{Z}_{j+1})^{-1}\bar{Z}'_{j+1}(\bar{Y}_{j+1}^d - \bar{W}_{j+1}\hat{\beta}_a) \quad (\text{A.27})$$

In (A.24)-(A.27), the choice of the estimate of the covariance matrix will have no effect provided a consistent one is used. We now analyze the first component of $F_{1,T}^j$ (the analysis for the second is identical).

$$\begin{aligned} & T_{j+1} \text{tr}((\Sigma^0)^{-1}\tilde{\Sigma}_{j+1}) \\ &= \text{tr}((\Sigma^0)^{-1} \sum_{t=1}^{T_{j+1}} (y_t - x'_{at}\tilde{\beta}_a + x'_{bt}\tilde{\beta}_{b1,j+1})(y_t - x'_{at}\tilde{\beta}_a + x'_{bt}\tilde{\beta}_{b1,j+1})') \\ &= \text{tr}\left(\sum_{t=1}^{T_{j+1}} (y_t - x'_{at}\tilde{\beta}_a + x'_{bt}\tilde{\beta}_{b1,j+1})'(\Sigma^0)^{-1}(y_t - x'_{at}\tilde{\beta}_a + x'_{bt}\tilde{\beta}_{b1,j+1})\right) \\ &= (Y_{1,j+1} - X_{a1,j+1}\tilde{\beta}_a - X_{b1,j+1}\tilde{\beta}_{b1,j+1})'(I_T \otimes (\Sigma^0)^{-1})(Y_{1,j+1} - X_{a1,j+1}\tilde{\beta}_a - X_{b1,j+1}\tilde{\beta}_{b1,j+1}) \\ &= (U_{1,j+1} + X_{a1,j+1}(\beta_a - \tilde{\beta}_a) + X_{b1,j+1}(\beta_b - \tilde{\beta}_{b1,j+1}))'(I_T \otimes (\Sigma^0)^{-1}) \\ &\quad (U_{1,j+1} + X_{a1,j+1}(\beta_a - \tilde{\beta}_a) + X_{b1,j+1}(\beta_b - \tilde{\beta}_{b1,j+1})) \\ &= (X_{a1,j+1}(\beta_a - \tilde{\beta}_a) + X_{b1,j+1}(\beta_b - \tilde{\beta}_{b1,j+1}))'(I_T \otimes (\Sigma^0)^{-1}) \\ &\quad (X_{a1,j+1}(\beta_a - \tilde{\beta}_a) + X_{b1,j+1}(\beta_b - \tilde{\beta}_{b1,j+1})) \\ &\quad + 2(X_{a1,j+1}(\beta_a - \tilde{\beta}_a) + X_{b1,j+1}(\beta_b - \tilde{\beta}_{b1,j+1}))'(I_T \otimes (\Sigma^0)^{-1})U_{1,j+1} \\ &\quad + U'_{1,j+1}(I_T \otimes (\Sigma^0)^{-1})U_{1,j+1} \\ &= (W_{1,j+1}(\beta_a - \tilde{\beta}_a) + Z_{1,j+1}(\beta_b - \tilde{\beta}_{b1,j+1}))'(W_{1,j+1}(\beta_a - \tilde{\beta}_a) + Z_{1,j+1}(\beta_b - \tilde{\beta}_{b1,j+1})) \\ &\quad + 2(W_{1,j+1}(\beta_a - \tilde{\beta}_a) + Z_{1,j+1}(\beta_b - \tilde{\beta}_{b1,j+1}))'U_{1,j+1}^d + U'_{1,j+1}(I_T \otimes (\Sigma^0)^{-1})U_{1,j+1} + o_p(1) \\ &= (M_{Z_{1,j+1}}W_{1,j+1}(\beta_a - \tilde{\beta}_a) - P_{Z_{1,j+1}}U_{1,j+1}^d)'(M_{Z_{1,j+1}}W_{1,j+1}(\beta_a - \tilde{\beta}_a) - P_{Z_{1,j+1}}U_{1,j+1}^d) \\ &\quad + 2(M_{Z_{1,j+1}}W_{1,j+1}(\beta_a - \tilde{\beta}_a) - P_{Z_{1,j+1}}U_{1,j+1}^d)'U_{1,j+1}^d + U'_{1,j+1}(I_T \otimes (\Sigma^0)^{-1})U_{1,j+1} + o_p(1) \\ &= (M_{Z_{1,j+1}}W_{1,j+1}A_T + P_{Z_{1,j+1}}U_{1,j+1}^d)'(M_{Z_{1,j+1}}W_{1,j+1}A_T + P_{Z_{1,j+1}}U_{1,j+1}^d) \\ &\quad - 2(M_{Z_{1,j+1}}W_{1,j+1}A_T + P_{Z_{1,j+1}}U_{1,j+1}^d)'U_{1,j+1}^d + U'_{1,j+1}(I_T \otimes (\Sigma^0)^{-1})U_{1,j+1} + o_p(1) \\ &= A'_T W'_{1,j+1} M_{Z_{1,j+1}} W_{1,j+1} A_T - U_{1,j+1}^d P_{Z_{1,j+1}} U_{1,j+1}^d \\ &\quad - 2(M_{Z_{1,j+1}} W_{1,j+1} A_T)' U_{1,j+1}^d + U'_{1,j+1} (I_T \otimes (\Sigma^0)^{-1}) U_{1,j+1} + o_p(1) \end{aligned}$$

where $A_T = [W'M_Z W]^{-1} W'M_Z U^d$. For the third component of $F_{1,T}^j$, we have, using similar arguments,

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

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

$$F_{1,T}^j = U_{1,j}^{d'}P_{Z_{1,j}}U_{1,j}^d + U_{j+1}^{d'}P_{Z_{j+1}}U_{j+1}^d - U_{1,j+1}^{d'}P_{Z_{1,j+1}}U_{1,j+1}^d + o_p(1)$$

Now, $T^{-1/2}Z'_{1,j}U_{1,j}^d \Rightarrow Q_b^{1/2}W_{p_b}(\lambda_j)$ and $T^{-1}\sum_{t=1}^{T_j}x_{bt}(\Sigma^0)^{-1}x'_{bt} \rightarrow^p \lambda_j Q_b$ with $W_{p_b}(\lambda_j)$ a p_b vector of independent Wiener processes defined on $[0, 1]$ and where Q_b is the appropriate submatrix of Q corresponding to the elements of x_{bt} . Hence,

$$U_{1,j+1}^{d'}P_{Z_{1,j+1}}U_{1,j+1}^d \Rightarrow [W_{p_b}(\lambda_{j+1})'W_{p_b}(\lambda_{j+1})]/\lambda_{j+1}$$

Using similar arguments

$$U_{1,j}^{d'}P_{Z_{1,j}}U_{1,j}^d \Rightarrow [W_{p_b}(\lambda_j)'W_{p_b}(\lambda_j)]/\lambda_j$$

and

$$U_{j+1}^{d'}P_{Z_{j+1}}U_{j+1}^d \Rightarrow (W_{p_b}(\lambda_{j+1}) - W_{p_b}(\lambda_j))'(W_{p_b}(\lambda_{j+1}) - W_{p_b}(\lambda_j))/(\lambda_{j+1} - \lambda_j)$$

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

$$F_{1,T}^j \Rightarrow \frac{(\lambda_j W_{p_b}(\lambda_{j+1}) - \lambda_{j+1} W_{p_b}(\lambda_j))'(\lambda_j W_{p_b}(\lambda_{j+1}) - \lambda_{j+1} W_{p_b}(\lambda_j))}{(\lambda_{j+1} - \lambda_j)\lambda_j\lambda_{j+1}} \quad (\text{A.28})$$

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

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

Let $((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I)^F$ be the matrix whose entries are those of $((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I)$ for the corresponding entries of Σ^0 that are not allowed to vary across regimes, the remaining entries being filled with 0s. We use the subscript F since the non-zero elements are estimates constructed using the full sample, i.e.,

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

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

0s. We use the superscript S since the non-zero elements are estimates constructed using the relevant segments, i.e.,

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

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

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

and, in view of (A.23),

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

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

$$\begin{aligned} &T_{j+1} ((\Sigma^0)^{-1} \tilde{\Sigma}_{1,j+1} - I)^S ((\Sigma^0)^{-1} \tilde{\Sigma}_{1,j+1} - I)^S \\ &= \frac{T}{T_{j+1}} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T_{j+1}} [(\Sigma^0)^{-1} u_t u_t' - I] \right)^S \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T_{j+1}} [(\Sigma^0)^{-1} u_t u_t' - I] \right)^S + o_p(1) \\ &\Rightarrow \xi_n (\lambda_{j+1})^S \xi_n (\lambda_{j+1})^S / \lambda_{j+1} \end{aligned}$$

and

$$\begin{aligned} &T_j ((\Sigma^0)^{-1} \tilde{\Sigma}_{1,j} - I)^S ((\Sigma^0)^{-1} \tilde{\Sigma}_{1,j} - I)^S \\ &= \frac{T}{T_j} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T_j} [(\Sigma^0)^{-1} u_t u_t' - I] \right)^S \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T_j} [(\Sigma^0)^{-1} u_t u_t' - I] \right)^S + o_p(1) \\ &\Rightarrow \xi_n (\lambda_j)^S \xi_n (\lambda_j)^S / \lambda_j \end{aligned}$$

and

$$\begin{aligned} &(T_{j+1} - T_j) ((\Sigma^0)^{-1} \hat{\Sigma}_{j+1}^S - I)^S ((\Sigma^0)^{-1} \hat{\Sigma}_{j+1}^S - I)^S \\ &= \frac{T}{T_{j+1} - T_j} \left(\frac{1}{\sqrt{T}} \sum_{t=T_{j+1}}^{T_{j+1}} [(\Sigma^0)^{-1} u_t u_t' - I] \right)^S \left(\frac{1}{\sqrt{T}} \sum_{t=T_{j+1}}^{T_{j+1}} [(\Sigma^0)^{-1} u_t u_t' - I] \right)^S + o_p(1) \\ &\Rightarrow (\xi_n (\lambda_{j+1}) - \xi_n (\lambda_j))^S (\xi_n (\lambda_{j+1}) - \xi_n (\lambda_j))^S / (\lambda_{j+1} - \lambda_j) \end{aligned}$$

where $\xi(\cdot)$ is an $n \times n$ matrix whose elements are (non-standard) Brownian motions defined on $[0, 1]$ such that $\text{var}(\text{vec}(\xi(1))) = \Omega$. Hence,

$$\begin{aligned} \sum_{j=1}^m F_{2,T}^j &\Rightarrow -\frac{1}{2} \text{tr} \left(\xi_n(\lambda_{j+1})^S \xi_n(\lambda_{j+1})^S / \lambda_{j+1} - \xi_n(\lambda_j)^S \xi_n(\lambda_j)^S / \lambda_j \right. \\ &\quad \left. + (\xi_n(\lambda_{i+1}) - \xi_n(\lambda_i))^S (\xi_n(\lambda_{i+1}) - \xi_n(\lambda_i))^S / (\lambda_{i+1} - \lambda_i) \right) \\ &= -\frac{1}{2} \left[\text{vec}(\xi_n(\lambda_{j+1})^S)' \text{vec}(\xi_n(\lambda_{j+1})^S) / \lambda_{j+1} - \text{vec}(\xi_n(\lambda_j)^S)' \text{vec}(\xi_n(\lambda_j)^S) / \lambda_j \right. \\ &\quad \left. (\text{vec}(\xi_n(\lambda_{j+1})^S) - \text{vec}(\xi_n(\lambda_j)^S))' (\text{vec}(\xi_n(\lambda_{j+1})^S) - \text{vec}(\xi_n(\lambda_j)^S)) / (\lambda_{i+1} - \lambda_i) \right] \end{aligned}$$

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

$$\begin{aligned} \text{vec}(\xi_n(\lambda_{j+1})^S)' \text{vec}(\xi_n(\lambda_{j+1})^S) &= \text{vec}(\xi_n(\lambda_{j+1}))' H' H \text{vec}(\xi_n(\lambda_{j+1})) \\ &\stackrel{d}{=} W_{n_b^*}(\lambda_{j+1})' H \Omega H' W_{n_b^*}(\lambda_{j+1}) \end{aligned}$$

where $W_{n_b^*}(\cdot)$ is an n_b^* vector of independent standard Wiener processes. Hence, we have

$$\begin{aligned} &\sum_{j=1}^m F_{2,T}^j \\ \Rightarrow &-\frac{1}{2} \left[W_{n_b^*}(\lambda_{j+1})' H' \Omega H W_{n_b^*}(\lambda_{j+1}) / \lambda_{j+1} - W_{n_b^*}(\lambda_j)' H' \Omega H W_{n_b^*}(\lambda_j) / \lambda_j \right. \\ &\quad \left. - (W_{n_b^*}(\lambda_{j+1}) - W_{n_b^*}(\lambda_j))' H' \Omega H (W_{n_b^*}(\lambda_{j+1}) - W_{n_b^*}(\lambda_j)) / (\lambda_{j+1} - \lambda_j) \right] \\ = &\frac{(\lambda_j W_{n_b^*}(\lambda_{j+1}) - \lambda_{j+1} W_{n_b^*}(\lambda_j))' H' \Omega H (\lambda_j W_{n_b^*}(\lambda_{j+1}) - \lambda_{j+1} W_{n_b^*}(\lambda_j))}{\lambda_j \lambda_{j+1} (\lambda_{j+1} - \lambda_j)} \quad (\text{A.29}) \end{aligned}$$

It remains to show that the limiting distribution of the test is given by (A.28) when only changes in β are allowed, and is given by (A.29) when only changes in β are allowed. We have

$$LR_T(T_1, \dots, T_m) = T \log |\tilde{\Sigma}| - T \log |\hat{\Sigma}|$$

where $\hat{\Sigma}$ and $\tilde{\Sigma}$ denote the covariance matrix of the errors estimated under the null and alternative hypotheses, respectively. Taking a second order Taylor expansion,

$$\begin{aligned} &LR_T(T_1, \dots, T_m) \\ = &\text{tr}(T \Sigma_0^{-1} (\tilde{\Sigma} - \hat{\Sigma})) + \frac{T}{2} \text{tr}([\Sigma_0^{-1} (\hat{\Sigma} - \Sigma_0^{-1})]^2) - \frac{T}{2} \text{tr}([\Sigma_0^{-1} (\tilde{\Sigma} - \Sigma_0^{-1})]^2) + o_p(T^{-1}) \end{aligned}$$

Consider first the third term

$$\begin{aligned}
[(\Sigma^0)^{-1}(\tilde{\Sigma} - \Sigma^0)]^2 &= [(\Sigma^0)^{-1}(T^{-1} \sum_{t=1}^T (y_t - x_t' \tilde{\beta})(y_t - x_t' \tilde{\beta})' - \Sigma^0)]^2 \\
&= [(\Sigma^0)^{-1}(T^{-1} \sum_{t=1}^T (u_t + x_t'(\beta^0 - \tilde{\beta}))(u_t + x_t'(\beta^0 - \tilde{\beta}))' - \Sigma^0)]^2 \\
&= [(\Sigma^0)^{-1}(T^{-1} \sum_{t=1}^T u_t u_t' - \Sigma^0)]^2 + O_p(T^{-3/2})
\end{aligned}$$

where the last equality follows since $\beta^0 - \tilde{\beta} = O_p(T^{-1/2})$. Similarly, we can show that

$$[(\Sigma^0)^{-1}(\hat{\Sigma} - \Sigma^0)]^2 = [(\Sigma^0)^{-1}(T^{-1} \sum_{t=1}^T u_t u_t' - \Sigma^0)]^2 + o_p(T^{-1})$$

Hence, the likelihood ratio simplifies to

$$\begin{aligned}
lr_T(T_1, \dots, T_k) &= T \text{tr}((\Sigma^0)^{-1}(\tilde{\Sigma} - \hat{\Sigma})) + o_p(1) \\
&\equiv \text{tr}((\Sigma^0)^{-1}(T\tilde{\Sigma} - \sum_{j=0}^m (T_{j+1} - T_j) \hat{\Sigma}_{j+1})) + o_p(1) \\
&= \text{tr}((\Sigma^0)^{-1} \sum_{j=1}^m (T_{j+1} \tilde{\Sigma}_{1,j+1} - T_j \tilde{\Sigma}_{1,j} - (T_{j+1} - T_j) \hat{\Sigma}_{j+1})) + o_p(1) \\
&= \sum_{j=1}^m F_{1,T}^j + o_p(1)
\end{aligned}$$

Now, when only changes in Σ occurs, we have, assuming without loss of generality that all elements of the covariance matrix are allowed to change:

$$\begin{aligned}
&T \log |\tilde{\Sigma}| - \sum_{j=0}^m (T_{j+1} - T_j) \log |\hat{\Sigma}_{j+1}| \\
&= \sum_{j=1}^m (T_{j+1} \log |\tilde{\Sigma}_{1,j+1}| - T_j \log |\tilde{\Sigma}_{1,j}| - (T_{j+1} - T_j) \log |\hat{\Sigma}_{j+1}|) + T_1 (\log |\tilde{\Sigma}_{1,1}| - \log |\hat{\Sigma}_1|) \\
&\equiv \sum_{j=1}^m LR_T^j + T_1 (\log |\tilde{\Sigma}_{1,1}| - \log |\hat{\Sigma}_1|)
\end{aligned}$$

where

$$\begin{aligned}
\tilde{\Sigma}_{1,j} &= T_j^{-1} \sum_{t=1}^{T_j} (y_t - x_t' \tilde{\beta})(y_t - x_t' \tilde{\beta})' \\
\hat{\Sigma}_j &= (T_j - T_{j-1})^{-1} \sum_{t=T_{j-1}+1}^{T_j} (y_t - x_t' \hat{\beta})(y_t - x_t' \hat{\beta})'
\end{aligned}$$

with $\tilde{\beta}$ and $\hat{\beta}$ the estimates under the null and alternative hypotheses, respectively. Now, taking a second order Taylor expansion of LR_T^j ,

$$\begin{aligned} LR_T^j &= tr(T_{j+1}(\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - T_j(\Sigma^0)^{-1}\tilde{\Sigma}_{1,j} - (T_{j+1} - T_j)(\Sigma^0)^{-1}\hat{\Sigma}_{j+1}) \\ &\quad + \frac{T_{j+1}}{2}tr[(\Sigma^0)^{-1}(\hat{\Sigma}_{j+1} - \Sigma^0)]^2 + \frac{T_j}{2}tr[(\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j} - \Sigma^0)]^2 \\ &\quad - \frac{(T_{j+1} - T_j)}{2}tr[(\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j+1} - \Sigma^0)]^2 + o_p(T^{-1}) \end{aligned}$$

Both $\tilde{\beta}$ and $\hat{\beta}$ are regime independent and we also have $\tilde{\beta} - \hat{\beta} = o_p(T^{-1/2})$. Hence,

$$\begin{aligned} &tr(T_{j+1}(\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - T_j(\Sigma^0)^{-1}\tilde{\Sigma}_{1,j} - (T_{j+1} - T_j)(\Sigma^0)^{-1}\hat{\Sigma}_{j+1}) \\ &= (\Sigma^0)^{-1} \sum_{t=T_j+1}^{T_{j+1}} [(y_t - x_t'\tilde{\beta})(y_t - x_t'\tilde{\beta})' - (y_t - x_t'\hat{\beta})(y_t - x_t'\hat{\beta})'] = o_p(1) \end{aligned}$$

Also,

$$\begin{aligned} &T_1(\log |\tilde{\Sigma}_{1,1}| - \log |\hat{\Sigma}_1|) \\ &= T_1 tr((\Sigma^0)^{-1}(\tilde{\Sigma}_{1,1} - \hat{\Sigma}_1)) + \frac{1}{2}T_1 tr([(\Sigma^0)^{-1}(\hat{\Sigma}_1 - \Sigma^0)]^2 - [(\Sigma^0)^{-1}(\tilde{\Sigma}_{1,1} - \Sigma^0)]^2) = o_p(1) \end{aligned}$$

Hence,

$$\begin{aligned} LR_T(T_1, \dots, T_m) &= \sum_{j=1}^m \frac{1}{2}tr(T_{j+1}[(\Sigma^0)^{-1}(\hat{\Sigma}_{j+1} - \Sigma^0)]^2 + T_j[(\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j} - \Sigma^0)]^2 \\ &\quad - (T_{j+1} - T_j)[(\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j+1} - \Sigma^0)]^2) + o_p(1) \\ &= \sum_{j=1}^m F_{2,T}^j + o_p(1) \end{aligned}$$

Proof of Theorem 7: Assume, without loss of generality, no disjoint break allowed under the alternative hypothesis and all regression coefficients allowed to change. For the general case, the proof extends straightforwardly. Hence, using the convention that $T_2 = T$ and $T_0 = 1$, the set of admissible partitions is

$$\begin{aligned} \Lambda_\varepsilon^* &= \{(k_1, k_2); \varepsilon T \leq k_1 \leq k_2 \leq (1 - \varepsilon)T \text{ and } v_T^2(k_2 - k_1) \leq M_T \\ &\quad \text{with } M_T \rightarrow 0, v_T \rightarrow 0 \text{ and } T^{1/2}v_T/(\log T)^2 \rightarrow \infty \text{ as } T \rightarrow \infty\} \end{aligned}$$

For a given partition, the likelihood ratio statistic is defined as

$$LR_T(k_1, k_2; \varepsilon) = T \log |\tilde{\Sigma}| - T \log |\hat{\Sigma}|$$

where $\hat{\Sigma}$ and $\tilde{\Sigma}$ denote the covariance matrix estimated under the null and alternative hypotheses, respectively. Now, consider the likelihood function under the common break alternative, which imposes $k_2 = k_1$, and denote the corresponding estimate of the covariance matrix as $\hat{\Sigma}^*$. Then,

$$LR_T(k_1, k_2; \varepsilon) = (T \log |\tilde{\Sigma}| - T \log |\hat{\Sigma}^*|) + (T \log |\hat{\Sigma}^*| - T \log |\hat{\Sigma}|)$$

Hence,

$$\sup_{(k_1, k_2) \in \Lambda_\varepsilon^*} LR_T(k_1, k_2; \varepsilon) = \sup_{(k_1, k_2) \in \Lambda_\varepsilon^*} \{(T \log |\tilde{\Sigma}| - T \log |\hat{\Sigma}^*|) + (T \log |\hat{\Sigma}^*| - T \log |\hat{\Sigma}|)\}$$

The proof is complete if we can show that

$$\sup_{(k_1, k_2) \in \Lambda_\varepsilon^*} (T \log |\hat{\Sigma}^*| - T \log |\hat{\Sigma}|) = o_p(1)$$

To prove this, apply a second order Taylor expansion,

$$\begin{aligned} & T \log |\hat{\Sigma}^*| - T \log |\hat{\Sigma}| \\ &= tr(T \Sigma_0^{-1} (\hat{\Sigma}^* - \hat{\Sigma})) + \frac{T}{2} tr([\Sigma_0^{-1} (\hat{\Sigma} - \Sigma_0)]^2) - \frac{T}{2} tr([\Sigma_0^{-1} (\hat{\Sigma}^* - \Sigma_0)]^2) + o_p(T^{-1}) \\ &= tr(T \Sigma_0^{-1} (\hat{\Sigma}^* - \hat{\Sigma})) + o_p(1) \end{aligned}$$

where the $o_p(1)$ term is uniform in $(k_1, k_2) \in \Lambda_\varepsilon^*$ and where the last equality holds because

$$\begin{aligned} [(\Sigma_0^{-1} (\hat{\Sigma} - \Sigma_0))]^2 &= [(\Sigma_0^{-1} (T^{-1} \sum_{t=1}^T u_t u_t' - \Sigma_0))]^2 + o_p(T^{-1}) \\ [(\Sigma_0^{-1} (\hat{\Sigma}^* - \Sigma_0))]^2 &= [(\Sigma_0^{-1} (T^{-1} \sum_{t=1}^T u_t u_t' - \Sigma_0))]^2 + o_p(T^{-1}) \end{aligned}$$

uniformly in $(k_1, k_2) \in \Lambda_\varepsilon^*$. Now, let $\hat{\beta}_t$ denote the estimate under the locally ordered break alternative,

$$\hat{\beta}_t = \begin{cases} (\hat{\beta}'_{1,1}, \hat{\beta}'_{2,1})' & \text{if } t \leq k_1 \\ (\hat{\beta}'_{1,2}, \hat{\beta}'_{2,1})' & \text{if } k_1 < t \leq k_2 \\ (\hat{\beta}'_{1,2}, \hat{\beta}'_{2,2})' & \text{if } t > k_2 \end{cases}$$

and let $\hat{\beta}_t^*$ denote the estimate under the common break alternative,

$$\hat{\beta}_t^* = \begin{cases} (\hat{\beta}^{*'}_{1,1}, \hat{\beta}^{*'}_{2,1})' & \text{if } t \leq k_1 \\ (\hat{\beta}^{*'}_{1,2}, \hat{\beta}^{*'}_{2,2})' & \text{if } t > k_1 \end{cases}$$

Then, for a given partition $(k_1, k_2) \in \Lambda_\varepsilon^*$, simple arguments lead to the result that, for $j = 1, 2$,

$$\hat{\beta}_{1,j} - \hat{\beta}_{1,j}^* = O_p((Tv_T)^{-1} \log v_T^{-2}); \quad \hat{\beta}_{2,j} - \hat{\beta}_{2,j}^* = O_p((Tv_T)^{-1} \log v_T^{-2})$$

which further implies

$$\begin{aligned} \text{tr}(T(\hat{\Sigma}_{1,1} - \hat{\Sigma}_{1,2})) &= \text{tr}\left(\sum_{t=1}^T (y_t - x'_t \hat{\beta}_t)(y_t - x'_t \hat{\beta}_t)' - \sum_{t=1}^T (y_t - x'_t \hat{\beta}_t^*)(y_t - x'_t \hat{\beta}_t^*)'\right) \\ &= \text{tr}\left(\sum_{t=k_1+1}^{k_2} (y_t - x'_t \hat{\beta}_t)(y_t - x'_t \hat{\beta}_t)' - \sum_{t=k_1+1}^{k_2} (y_t - x'_t \hat{\beta}_t^*)(y_t - x'_t \hat{\beta}_t^*)'\right) + o_p(1) \\ &= \text{tr}\left(\sum_{t=k_1+1}^{k_2} (u_t - x'_t(\hat{\beta}_t - \beta^0))(u_t - x'_t(\hat{\beta}_t - \beta^0))'\right) \\ &\quad - \text{tr}\left(\sum_{t=k_1+1}^{k_2} (u_t - x'_t(\hat{\beta}_t^* - \beta^0))(u_t - x'_t(\hat{\beta}_t^* - \beta^0))'\right) + o_p(1) \end{aligned}$$

Now, from Lemma A.5 $\hat{\beta}_t - \beta^0 = O_p(T^{-1/2})$ and $\hat{\beta}_t^* - \beta^0 = O_p(T^{-1/2})$ uniformly in $(k_1, k_2) \in \Lambda_\varepsilon^*$. Hence, using the fact that $(k_2 - k_1)/T \rightarrow 0$, we have

$$\begin{aligned} \text{tr}\left(\sum_{t=k_1+1}^{k_2} x'_t(\hat{\beta}_t - \beta^0)(\hat{\beta}_t - \beta^0)'x_t\right) &= o_p(1) \\ \text{tr}\left(\sum_{t=k_1+1}^{k_2} u_t(\hat{\beta}_t - \beta^0)'x_t\right) &= o_p(1) \\ \text{tr}\left(\sum_{t=k_1+1}^{k_2} x'_t(\hat{\beta}_t^* - \beta^0)(\hat{\beta}_t^* - \beta^0)'x_t\right) &= o_p(1) \end{aligned}$$

and

$$\text{tr}\left(\sum_{t=k_1+1}^{k_2} u_t(\hat{\beta}_t^* - \beta^0)'x_t\right) = o_p(1).$$

Hence $\hat{\Sigma}_{1,1} - \hat{\Sigma}_{1,2} = o_p(1)$, the bound being uniform in $(k_1, k_2) \in \Lambda_\varepsilon^*$. This completes the proof.

Proof of Theorem 8: Without loss of generality, assume all the coefficients are subjected to change. Let $\hat{\beta}_t$ denote the coefficients estimates under the globally ordered breaks alternative, then, for a given admissible partition $(\lambda_1, \lambda_2) \in \Lambda_\varepsilon^G$, we have

$$\hat{\beta}_t = \begin{cases} (\hat{\beta}'_{1,1}, \hat{\beta}'_{2,1})' & \text{if } t \leq k_1 \\ (\hat{\beta}'_{1,2}, \hat{\beta}'_{2,1})' & \text{if } k_1 < t \leq k_2 \\ (\hat{\beta}'_{1,2}, \hat{\beta}'_{2,2})' & \text{if } t > k_2 \end{cases}$$

with the corresponding log-likelihood function being

$$\begin{aligned}\mathcal{L}_T^G(k_1, k_2) &= -\frac{T}{2}(\log 2\pi + 1) - \frac{T}{2} \log |\hat{\Sigma}|, \\ \text{with } \hat{\Sigma} &= \frac{1}{T} \sum_{t=1}^T (y_t - x_t' \hat{\beta}_t)(y_t - x_t' \hat{\beta}_t)'\end{aligned}$$

Consider a related model in which only the coefficients in the second set of equations are allowed to change. Let $\tilde{\beta}_t$ denote the corresponding estimates, then, under the same partition as before, we have,

$$\tilde{\beta}_t = \begin{cases} (\tilde{\beta}'_1, \tilde{\beta}'_{2,1})' & \text{if } t \leq k_2 \\ (\tilde{\beta}'_1, \tilde{\beta}'_{2,2})' & \text{if } t > k_2 \end{cases}$$

with the corresponding likelihood function being

$$\begin{aligned}\mathcal{L}_T^G(1, k_2) &= -\frac{T}{2}(\log 2\pi + 1) - \frac{T}{2} \log |\tilde{\Sigma}_2|, \\ \text{with } \tilde{\Sigma}_2 &= \frac{1}{T} \sum_{t=1}^T (y_t - x_t' \tilde{\beta}_t)(y_t - x_t' \tilde{\beta}_t)'\end{aligned}$$

Hence, the likelihood ratio under the given partition can be expressed as

$$LR_T^G(k_1, k_2, p_{b1}, p_{b2}, \varepsilon) = T(\log |\tilde{\Sigma}_1| - \log |\hat{\Sigma}|) + T(\log |\tilde{\Sigma}| - \log |\tilde{\Sigma}_1|)$$

Hence the likelihood ratio is sum of two components, each involves only one break, with some coefficients restricted not to change. Given this, the rest of the proof follows that of Theorem 5, i.e, the analysis of $F_{1,T}^j$.