SUPPLEMENT TO "ESTIMATING AND TESTING STRUCTURAL CHANGES IN MULTIVARIATE REGRESSIONS"

(*Econometrica*, Vol. 75, No. 2, March, 2007, 459–502)

BY ZHONGJUN QU AND PIERRE PERRON

PROOF OF LEMMA A.1: This lemma was presented by 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

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

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

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

and

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

which, combined with the fact that ξ_i is $\mathcal{F}_i$ measurable, imply that ξ_i is an L^2 mixingale of size $-2-4\delta/((r-2)(r+\delta))$. If we let ψ_j denote the mixing coefficients 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. *Q.E.D.*

PROOF OF LEMMA A.2: Given Assumption A4, parts (a) and (b) follow from Theorem 2 in Eberlain (1986) once we show that $\|E(S_n(\ell)|\mathcal{F}_\ell)\|_2 \le C$ for C independent of ℓ and that uniformly in ℓ,

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

The proof of the foregoing statements proceeds in the same way as in Corradi (1999, pp. 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 an L^2 mixingale with mixing size -2 and bounded fourth moment. For part (d),

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

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

Given that

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

by part (a) and

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

by the law of iterated logarithm (LIL) for vector-valued Brownian motion processes, we have

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

which completes the proof. *Q.E.D.*

PROOF OF LEMMA A.3: To prove this result, we show that under Assumptions 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. 324–329). Following BLS (1998), but using slightly different notation, the quasi-likelihood ratio using the first k observations, evaluated at θ and Σ, can be written as

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

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 in the absense of structural change. *Q.E.D.*

PROPERTY 1: *For each* $\delta \in (0,1]$,

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

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

This statement corresponds to Property 1 of BLS (1998, p. 420). It states that the likelihood function and the maximum likelihood estimate are well behaved in large samples. The uniformity of the bound is important because we need to search over all admissible partitions to find the break points. Bai, Lumsdaine, and Stock (1998) provided a proof for the seemingly unrelated regression model with common regressors, in which case the covariance matrix plays no role in estimating β. Here, our setup is more complicated because the interaction between β and Σ makes the problem nonlinear. The solution is to use an argument based on the minimization of the *Kullback–Leibler* distance, which is from Domowitz and White (1982, Theorem 2.2).

PROOF OF PROPERTY 1: First, simple arguments lead to the result that $E_0(\log \mathcal{L}(1, k; \beta, \Sigma))$ achieves a maximum at $\beta_{(k)} = \beta^0$ and $\Sigma_{(k)} = \Sigma^0 = E_0(k^{-1} \sum_{t=1}^{k} u_t u_t')$, 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 strong law of large numbers (SLLN) and a functional central limit theorem, we have

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

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

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

where

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

In the preceding equation, 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 arbitrarily 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 Hájek–Rényi inequality on $\sum_{t=1}^{k} x_t (\Sigma^0)^{-1} u_t$, together

with the strong consistency of $\hat{\Sigma}_{(k)}$, we have $\sup_{T\delta\le k\le T}\|\hat{\beta}_{(k)}-\beta^0\|=O_p(T^{-1/2})$. 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 Hájek–Rényi inequality, we have $\sup_{T\delta\le k\le T}\|\tilde{\Sigma}_{(k)}-\Sigma^0\|=O_p(T^{-1/2})$. Finally, $\sup_{T\delta\le k\le T}\mathcal{L}(1,k;\hat{\beta}_{(k)},\hat{\Sigma}_{(k)})=O_p(1)$ is a direct implication of the foregoing results, which completes the proof. *Q.E.D.*

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. Second, as before, the interaction between β and Σ makes the problem nonlinear. The solution is to apply the strong approximation theorem and the LIL.

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

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

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

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

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 $\hat{\Sigma}_{(k)}\to^{\text{a.s.}}\Sigma^0$, $\hat{\beta}_{(k)}\to^{\text{a.s.}}\beta^0$, and $k^{-1}\sum_{t=1}^{k}u_tu_t'\to^{\text{a.s.}}\Sigma^0$ as $k\to\infty$, which can be shown by applying a SLLN. Hence, we have

$$\sup_{k\ge k_1}\|\hat{\Sigma}_{(k)}^{-1}-(\Sigma^0)^{-1}\|=O_p(1),$$

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

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

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

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

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

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

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

from Lemma A.2 and $(k^{1/2}\log^{1/2}T)^{-1}\sup_{k\geq k_1}\|W(k)\|=o_{\text{a.s.}}(1)$ 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

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

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

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

and

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

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

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

where the last equality follows because

$$\sup_{k\geq k_1}\left\| A_k'\left(\sum_{t=1}^{k} x_t(\Sigma^0)^{-1}u_t\right)\right\|$$

$$\leq \sup_{k\geq k_1}\left\| k^{-1/2}\left(\sum_{t=1}^{k} x_t(\Sigma^0)^{-1}u_t\right)\right\| O_p(\log^{1/2} T)$$
$$= O_p(\log T),$$
$$\sup_{k\geq k_1}\left\| A_k'\left(\sum_{t=1}^{k} x_t(\Sigma^0)^{-1}x_t'\right)A_k\right\|$$
$$\leq \sup_{k\geq k_1}\|A_k'\| \sup_{k\geq k_1}\left\|\sum_{t=1}^{k} x_t(\Sigma^0)^{-1}x_t'\right\| \sup_{k\geq k_1}\|A_k\|$$
$$= O_p(\log T).$$

REMARK S.1: 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 (S.2) is also $O_p(\log T)$. To see this, note that

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

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

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

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

$$\Phi_k = \left(k^{-1/2}\sum_{t=1}^{k} x_t(\Sigma^0)^{-1}u_t\right)'\left[k^{-1}\sum_{t=1}^{k} x_t(\Sigma^0)^{-1}x_t'\right]^{-1} \times \left(k^{-1/2}\sum_{t=1}^{k} x_t(\Sigma^0)^{-1}u_t\right).$$

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

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 is used.

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

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

PROOF: We first consider the behavior of the likelihood function over the compact set

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

where $\lambda_{\min}$ and $\lambda_{\max}$ denote 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) is an inner point of $\bar{\Theta}_2$. We first want to show that

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

Following BLS (1998, pp. 422–424), 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

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

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 $\mathcal{L}_{2T}$ 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

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

We then need to show that

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

and

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

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

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

Also,

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

Using Lemma A.2, we have, for any fixed $r>0$,

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

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

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

which implies

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

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) showed that

$$\begin{aligned}\text{(S.6)}\qquad &\sup_{k\geq T\delta}\sup_{S_{1,T}\cap\bar{\Theta}_2}\mathcal{L}_{1T} \\ &\leq-\frac{k}{2}\left[\sum_{i=1}^{n}\left(\log(1+\lambda_i)+\left(\frac{1}{1+\lambda_i}-1\right)(1+\operatorname{sign}(\lambda_i)ab_T)\right)\right]\end{aligned}$$

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

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

We still need to show that the preceding 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

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

that is, 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 $(\hat{\theta}_{(k)}, \hat{\Sigma}_{(k)})$ that satisfies $\inf(\lambda_{\min}(\hat{\Sigma}_{(k)})) \to d_l > 0$ with $d_l < d_2$ or $\sup(\lambda_{\max}(\hat{\Sigma}_{(k)})) \to d_u < \infty$ with $d_u > d_3$; (2) with positive probability, there is a sequence $(\hat{\theta}_{(k)}, \hat{\Sigma}_{(k)})$ with $\inf(\lambda_{\min}(\hat{\Sigma}_{(k)})) \to 0$ or $\sup(\lambda_{\max}(\hat{\Sigma}_{(k)})) \to \infty$; (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 $\hat{\theta}_{(k)}$ that lie on the boundary of S_T will yield a larger value of the likelihood function almost surely. This completes the proof. *Q.E.D.*

PROPERTY 4: *A property that corresponds 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 $\liminf_{T\to\infty}(h_T d_T^2)/T \geq h > 0$.

PROPERTY 5: *Let* h_T *and* d_T *be positive sequences such that* h_T *is nondecreasing,* $d_T \to +\infty$, *and* $(h_T d_T^2)/T \to h < \infty$. *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 that satisfy* $p_1 < \infty, 0 < p_2 \leq p_3 < \infty$. *Define* $S_T = \{(\beta,\Sigma); \|\beta - \beta^0\| \geq T^{-1/2}\log T$ *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*,

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

PROOF: 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

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

or

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

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

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

where

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

Because $l^{-1} \sum_{t=1}^{l} z_t z_t' \to^{\text{a.s.}} Q_z$, 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

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

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

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

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

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

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

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 A.1, we have

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

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

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

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

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

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

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

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

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

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

PROPERTY 7: *Under the conditions of the Property* 6, *we have*

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

PROPERTY 8: *We have*

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

where the supremum with respect to β^*, Σ^*, λ, and Ξ 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*

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

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

PROPERTY 10: *Assume the same setup 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 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, that is, $v_T^2(T_j - T_j^0) \le M$. 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

$$\begin{aligned} lr_T = &\underbrace{\frac{1}{2}\sum_{j=1}^{m}\sum_{t=T_j}^{T_j^0}(y_t - x_t'\beta_j)'\Sigma_j^{-1}(y_t - x_t'\beta_j)}_{\text{(I)}} \\ &\underbrace{-\frac{1}{2}\sum_{j=1}^{m}\sum_{t=T_j}^{T_j^0}(y_t - x_t'\beta_{j+1})'\Sigma_{j+1}^{-1}(y_t - x_t'\beta_{j+1})}_{\text{(II)}} \\ &\underbrace{+\sum_{j=1}^{m}\frac{T_j^0 - T_j}{2}(\log|\Sigma_j| - \log|\Sigma_{j+1}|)}_{\text{(III)}} + \sum_{j=1}^{m+1} lr_j^2 \end{aligned}$$

with lr_j^2 as defined in Theorem 1. Hence, we need to show that the sum of the terms (I)–(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

$j = 1, \ldots, m+1$), and note that both are $O_p(1)$. Term (I) is then equivalent to

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

The last equality follows because

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

and

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

Furthermore, using the fact that

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

we deduce that

$$\text{(S.11)}\qquad \text{(I)} = \frac{1}{2}\sum_{j=1}^{m}\sum_{t=T_j}^{T_j^0}\operatorname{tr}\bigl(((\Sigma_j^0)^{-1} - T^{-1/2}(\Sigma_j^0)^{-1}\Sigma_j^*(\Sigma_j^0)^{-1})u_t u_t'\bigr) + o_p(1).$$

For term (II), note that, with $\Delta\beta_j^0 \equiv \beta_{j+1}^0 - \beta_j^0$,

$$\begin{aligned}\text{(S.12)}\qquad & -\frac{1}{2}\sum_{j=1}^{m}\sum_{t=T_j}^{T_j^0}\left(u_t - x_t'\Delta\beta_j^0 - \frac{x_t'\beta_{j+1}^*}{\sqrt{T}}\right)'\left(\Sigma_{j+1}^0 + \frac{\Sigma_{j+1}^*}{\sqrt{T}}\right)^{-1}\\ &\qquad\times\left(u_t - x_t'\Delta\beta_j^0 - \frac{x_t'\beta_{j+1}^*}{\sqrt{T}}\right)\\ &= -\frac{1}{2}\sum_{j=1}^{m}\sum_{t=T_j}^{T_j^0}(u_t - x_t'\Delta\beta_j^0)'\left(\Sigma_{j+1}^0 + \frac{\Sigma_{j+1}^*}{\sqrt{T}}\right)^{-1}(u_t - x_t'\Delta\beta_j^0) + o_p(1)\\ &= -\frac{1}{2}\sum_{j=1}^{m}\sum_{t=T_j}^{T_j^0}(u_t - x_t'\Delta\beta_j^0)'\left[(\Sigma_{j+1}^0)^{-1} - (\Sigma_{j+1}^0)^{-1}\frac{\Sigma_{j+1}^*}{\sqrt{T}}(\Sigma_{j+1}^0)^{-1}\right]\\ &\qquad\qquad\times(u_t - x_t'\Delta\beta_j^0) + o_p(1)\\ &= -\frac{1}{2}\sum_{j=1}^{m}\sum_{t=T_j}^{T_j^0}\operatorname{tr}\bigl(((\Sigma_{j+1}^0)^{-1} - T^{-1/2}(\Sigma_{j+1}^0)^{-1}\Sigma_{j+1}^*(\Sigma_{j+1}^0)^{-1})u_t u_t'\bigr)\\ &\quad -\frac{1}{2}\sum_{j=1}^{m}\sum_{t=T_j}^{T_j^0}\Delta\beta_j^{0\prime}x_t(\Sigma_{j+1}^0)^{-1}x_t'\Delta\beta_j^0 + \sum_{j=1}^{m}\sum_{T_j}^{T_j^0}\Delta\beta_j^{0\prime}x_t(\Sigma_{j+1}^0)^{-1}u_t\\ &\quad + o_p(1).\end{aligned}$$

For term (III),

$$\begin{aligned}\text{(S.13)}\qquad & \frac{1}{2}\sum_{j=1}^{m}(T_j^0 - T_j)(\log|\Sigma_j| - \log|\Sigma_{j+1}|)\\ &= \frac{1}{2}\sum_{j=1}^{m}(T_j^0 - T_j)\bigl(\log|\Sigma_j^0| + T^{-1/2}(\Sigma_j^0)^{-1}\Sigma_j^* - \log|\Sigma_{j+1}^0|\\ &\qquad\qquad - T^{-1/2}(\Sigma_{j+1}^0)^{-1}\Sigma_{j+1}^* + O_p(T^{-1})\bigr)\end{aligned}$$

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

Collecting the results in (S.11), (S.12), and (S.13), the sum of terms (I)–(III) is

$$
\begin{aligned}
&\frac{1}{2}\sum_{j=1}^{m}\sum_{t=T_j}^{T_j^0} u_t'((\Sigma_j^0)^{-1} - (\Sigma_{j+1}^0)^{-1})u_t - \frac{1}{2}\sum_{j=1}^{m}\sum_{t=T_j}^{T_j^0}\Delta\beta_j^{0\prime}x_t(\Sigma_{j+1}^0)^{-1}x_t'\Delta\beta_j^0 \\
&\quad + \frac{1}{2}\sum_{j=1}^{m}(T_j^0 - T_j)(\log|\Sigma_j^0| - \log|\Sigma_{j+1}^0|) + \sum_{j=1}^{m}\sum_{t=T_j}^{T_j^0}\Delta\beta_j^{0\prime}x_t(\Sigma_{j+1}^0)^{-1}u_t \\
&\quad - \frac{1}{2}T^{-1/2}\operatorname{tr}\left(((\Sigma_j^0)^{-1}\Sigma_j^*(\Sigma_j^0)^{-1})\sum_{j=1}^{m}\sum_{t=T_j}^{T_j^0}(u_tu_t' - \Sigma_j^0)\right) \\
&\quad + \frac{1}{2}T^{-1/2}\operatorname{tr}\left(((\Sigma_{j+1}^0)^{-1}\Sigma_{j+1}^*(\Sigma_{j+1}^0)^{-1})\sum_{j=1}^{m}\sum_{t=T_j}^{T_j^0}(u_tu_t' - \Sigma_{j+1}^0)\right) \\
&\quad + o_p(1).
\end{aligned}
$$

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

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

and

$$T^{-1/2}\operatorname{tr}\left(((\Sigma_{j+1}^0)^{-1}\Sigma_{j+1}^*(\Sigma_{j+1}^0)^{-1})\sum_{j=1}^{m}\sum_{t=T_j}^{T_j^0}(u_t u_t' - \Sigma_{j+1}^0)\right)$$

$$= T^{-1/2}\operatorname{tr}\left(((\Sigma_{j+1}^0)^{-1}\Sigma_{j+1}^*(\Sigma_{j+1}^0)^{-1})\right.$$

$$\left.\times\left\{\sum_{j=1}^{m}\sum_{t=T_j}^{T_j^0}(u_t u_t' - \Sigma_j^0) + \sum_{j=1}^{m}\sum_{t=T_j}^{T_j^0}(\Sigma_j^0 - \Sigma_{j+1}^0)\right\}\right)$$

$$= o_p(1). \qquad \textit{Q.E.D.}$$

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

LEMMA S.1: *Let* $\eta_t = (\Sigma_j^0)^{-1/2}u_t$. *Under Assumptions* A4 *and* A5, *with* v_T *a sequence of positive numbers that satisfy* $v_T \to 0$ *and* $T^{1/2}v_T/(\log T)^2 \to \infty$, *we have*

$$\textit{for} \quad s < 0, \quad v_T \sum_{t=T_j^0+[s/v_T^2]}^{T_j^0} (\eta_t\eta_t' - I) \Rightarrow \xi_{1,j}(s),$$

$$\textit{for} \quad s > 0, \quad v_T \sum_{t=T_j^0}^{T_j^0+[s/v_T^2]} (\eta_t\eta_t' - I) \Rightarrow \xi_{2,j}(s),$$

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*

$$\textit{for} \quad s < 0, \quad v_T \sum_{t=T_j^0+[s/v_T^2]}^{T_j^0} x_t(\Sigma_{j+1}^0)^{-1}u_t \Rightarrow (\Pi_{1,j})^{1/2}\zeta_{1,j}(s),$$

$$\textit{for} \quad s > 0, \quad v_T \sum_{t=T_j^0}^{T_j^0+[s/v_T^2]} x_t(\Sigma_j^0)^{-1}u_t \Rightarrow (\Pi_{2,j})^{1/2}\zeta_{2,j}(s),$$

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 $j = 1, \ldots, m$. 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 jth 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

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

For the second term, we have

$$
\begin{aligned}
&-\frac{r}{2}(\log|\Sigma_j^0| - \log|\Sigma_{j+1}^0|) \\
&\quad= -\frac{r}{2}\log|(\Sigma_j^0 - \Sigma_{j+1}^0 + \Sigma_{j+1}^0)(\Sigma_{j+1}^0)^{-1}|
\end{aligned}
$$

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

Then the sum of the first two terms is

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

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

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

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

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

with

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

Combining the foregoing results, we have, for $s < 0$,

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

Now, $\operatorname{vec}(A_{1,j})'\operatorname{vec}(\xi_{1,j}(s)) \stackrel{d}{=} (\operatorname{vec}(A_{1,j})'\Omega_{1,j}^0\operatorname{vec}(A_{1,j}))^{1/2}V_{1,j}(s)$, where $V_{1,j}(s)$ is a standard Wiener process. Similarly, $\delta_j'(\Pi_{1,j})^{1/2}\zeta_{1,j}(s) \stackrel{d}{=} (\delta_j'(\Pi_{1,j})\delta_j)^{1/2} \times U_{1,j}(s)$, 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

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

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

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

The proof for the case $s > 0$ is similar:

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

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

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

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

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

where $B_j(s) = B_{1,j}(s)$ for $s \le 0$ 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 3. *Q.E.D.*

PROOF OF THEOREM 5: As a matter of notation, let

$$\tilde{\Sigma}_{1,j} = \frac{1}{T_j}\sum_{t=1}^{T_j}(y_t - x'_{at}\tilde{\beta}_a - x'_{bt}\tilde{\beta}_{b1,j})(y_t - x'_{at}\tilde{\beta}_a - x'_{bt}\tilde{\beta}_{b1,j})' \tag{S.14}$$

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,

$$\hat{\Sigma}_j = \frac{1}{T_j - T_{j-1}}\sum_{t=T_{j-1}+1}^{T_j}(y_t - x'_{at}\hat{\beta}_a - x'_{bt}\hat{\beta}_{bj})(y_t - x'_{at}\hat{\beta}_a - x'_{bt}\hat{\beta}_{bj})' \tag{S.15}$$

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

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

For a given partition of the sample, we have

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

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

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

$$\log|\tilde{\Sigma}_{1,j}| = \log|\Sigma^0| + \operatorname{tr}((\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j} - \Sigma^0)) - \frac{1}{2}\operatorname{tr}((\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j} - \Sigma^0)(\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j} - \Sigma^0)) + o_p(T^{-1}),$$

$$\log|\hat{\Sigma}_{j+1}| = \log|\Sigma^0| + \operatorname{tr}((\Sigma^0)^{-1}(\hat{\Sigma}_{j+1} - \Sigma^0)) - \frac{1}{2}\operatorname{tr}((\Sigma^0)^{-1}(\hat{\Sigma}_{j+1} - \Sigma^0)(\Sigma^0)^{-1}(\hat{\Sigma}_{j+1} - \Sigma^0)) + o_p(T^{-1}).$$

Hence,

$$F_T^j \equiv F_{1,T}^j + F_{2,T}^j$$

$$= \operatorname{tr}\big(T_{j+1}(\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j+1} - \Sigma^0) - T_j(\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j} - \Sigma^0) - (T_{j+1} - T_j)(\Sigma^0)^{-1}(\hat{\Sigma}_{j+1} - \Sigma^0)\big) \tag{S.16}$$

$$- \frac{1}{2}\operatorname{tr}\big(T_{j+1}[(\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j+1} - \Sigma^0)]^2 - T_j[(\Sigma^0)^{-1}(\tilde{\Sigma}_{1,j} - \Sigma^0)]^2 - (T_{j+1} - T_j)[(\Sigma^0)^{-1}(\hat{\Sigma}_{j+1} - \Sigma^0)]^2\big). \tag{S.17}$$

Note that the first term in (S.16), denoted $F_{1,T}^j$, will be nonvanishing when allowance is made for changes in β^0, while the second term, denoted $F_{2,T}^j$, will be nonvanishing 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 are 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

$$\tilde{\beta}_a = [W'M_ZW]^{-1}W'M_ZY^d, \tag{S.18}$$

$$\tilde{\beta}_{b1,j} = (Z_{1,j}'Z_{1,j})^{-1}Z_{1,j}'(Y_{1,j}^d - W_{1,j}\tilde{\beta}_a), \tag{S.19}$$

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

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

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

$$\hat{\beta}_a = [\bar{W}' M_{\bar{Z}} \bar{W}]^{-1} \bar{W}' M_{\bar{Z}} \bar{Y}^d, \tag{S.20}$$

$$\hat{\beta}_{b,j+1} = (\bar{Z}'_{j+1}\bar{Z}_{j+1})^{-1}\bar{Z}'_{j+1}(\bar{Y}^d_{j+1} - \bar{W}_{j+1}\hat{\beta}_a). \tag{S.21}$$

In (S.18)–(S.21), 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 $F^j_{1,T}$ (the analysis for the second is identical):

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

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

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

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

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

$$
F^j_{1,T} = U^{d\prime}_{1,j} P_{Z_{1,j}} U^d_{1,j} + U^{d\prime}_{j+1} P_{\bar{Z}_{j+1}} U^d_{j+1} - U^{d\prime}_{1,j+1} P_{Z_{1,j+1}} U^d_{1,j+1} + o_p(1).
$$

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

$$
U^{d\prime}_{1,j+1} P_{Z_{1,j+1}} U^d_{1,j+1} \Rightarrow [W_{p_b}(\lambda_{j+1})' W_{p_b}(\lambda_{j+1})]/\lambda_{j+1}.
$$

Using similar arguments

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

and

$$
\begin{aligned}
&U^{d\prime}_{j+1} P_{\bar{Z}_{j+1}} U^d_{j+1} \\
&\quad \Rightarrow (W_{p_b}(\lambda_{j+1}) - W_{p_b}(\lambda_j))' (W_{p_b}(\lambda_{j+1}) - W_{p_b}(\lambda_j))/(\lambda_{j+1} - \lambda_j).
\end{aligned}
$$

These results imply that the first component in (S.16) has the limit

$$F_{1,T}^{j} \Rightarrow \frac{(\lambda_j W_{p_b}(\lambda_{j+1}) - \lambda_{j+1} W_{p_b}(\lambda_j))'(\lambda_j W_{p_b}(\lambda_{j+1}) - \lambda_{j+1} W_{p_b}(\lambda_j))}{(\lambda_{j+1} - \lambda_j)\lambda_j \lambda_{j+1}}. \tag{S.22}$$

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

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

Let $((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I)^F$ be the matrix whose entries are those of $((\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - I)$ for the corresponding entries of Σ^0 that are not allowed to vary across regimes; the remaining entries are filled with zeros. We use the superscript F because the nonzero elements are estimates constructed using the full sample, that is,

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

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 zeros. We use the superscript S because the nonzero elements are estimates constructed using the relevant segments, that is,

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

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 (S.17),

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

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

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

and

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

$$\Rightarrow \frac{(\xi_n(\lambda_{j+1}) - \xi_n(\lambda_j))^S (\xi_n(\lambda_{j+1}) - \xi_n(\lambda_j))^S}{\lambda_{j+1} - \lambda_j},$$

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

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

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

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

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

$$\begin{aligned}
\text{(S.23)} \quad \sum_{j=1}^{m} F_{2,T}^j \Rightarrow & -\frac{1}{2} \Bigg[\frac{W_{n_b^*}(\lambda_{j+1})' H' \Omega H W_{n_b^*}(\lambda_{j+1})}{\lambda_{j+1}} - \frac{W_{n_b^*}(\lambda_j)' H' \Omega H W_{n_b^*}(\lambda_j)}{\lambda_j} \\
& - \frac{(W_{n_b^*}(\lambda_{j+1}) - W_{n_b^*}(\lambda_j))' H' \Omega H (W_{n_b^*}(\lambda_{j+1}) - W_{n_b^*}(\lambda_j))}{\lambda_{j+1} - \lambda_j} \Bigg] \\
= & (\lambda_j W_{n_b^*}(\lambda_{j+1}) - \lambda_{j+1} W_{n_b^*}(\lambda_j))' H' \Omega H \\
& \times (\lambda_j W_{n_b^*}(\lambda_{j+1}) - \lambda_{j+1} W_{n_b^*}(\lambda_j)) \\
& / (\lambda_j \lambda_{j+1} (\lambda_{j+1} - \lambda_j)).
\end{aligned}$$

It remains to show that the limiting distribution of the test is given by (S.22) when only changes in β are allowed and is given by (S.23) when only changes in β are allowed. We have

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

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 yields

$$\begin{aligned} LR_T(T_1,\ldots,T_m) &= \operatorname{tr}(T\Sigma_0^{-1}(\tilde{\Sigma}-\hat{\Sigma})) + \frac{T}{2}\operatorname{tr}\big([(\Sigma^0)^{-1}(\hat{\Sigma}-\Sigma^0)]^2\big) \\ &\quad - \frac{T}{2}\operatorname{tr}\big([(\Sigma^0)^{-1}(\tilde{\Sigma}-\Sigma^0)]^2\big) + o_p(T^{-1}). \end{aligned}$$

Consider first the third term

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

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

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

Hence, the likelihood ratio simplifies to

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

$$= \operatorname{tr}\left((\Sigma^0)^{-1} \sum_{j=1}^{m} (T_{j+1}\tilde{\Sigma}_{1,j+1} - T_j\tilde{\Sigma}_{1,j} - (T_{j+1} - T_j)\hat{\Sigma}_{j+1}) \right) + o_p(1)$$

$$= \sum_{j=1}^{m} F_{1,T}^{j} + o_p(1).$$

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

$$T \log|\tilde{\Sigma}| - \sum_{j=0}^{m} (T_{j+1} - T_i) \log|\hat{\Sigma}_{j+1}|$$

$$= \sum_{j=1}^{m} (T_{j+1} \log|\tilde{\Sigma}_{1,j+1}| - T_j \log|\tilde{\Sigma}_{1,j}| - (T_{j+1} - T_j) \log|\hat{\Sigma}_{j+1}|)$$

$$+ T_1(\log|\tilde{\Sigma}_{1,1}| - \log|\hat{\Sigma}_1|)$$

$$\equiv \sum_{j=1}^{m} LR_T^j + T_1(\log|\tilde{\Sigma}_{1,1}| - \log|\hat{\Sigma}_1|),$$

where

$$\tilde{\Sigma}_{1,j} = T_j^{-1} \sum_{t=1}^{T_j} (y_t - x_t'\tilde{\beta})(y_t - x_t'\tilde{\beta})',$$

$$\hat{\Sigma}_j = (T_j - T_{j-1})^{-1} \sum_{t=T_{j-1}+1}^{T_j} (y_t - x_t'\hat{\beta})(y_t - x_t'\hat{\beta})'$$

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 yields

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

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

$$\operatorname{tr}(T_{j+1}(\Sigma^0)^{-1}\tilde{\Sigma}_{1,j+1} - T_j(\Sigma^0)^{-1}\tilde{\Sigma}_{1,j} - (T_{j+1} - T_j)(\Sigma^0)^{-1}\hat{\Sigma}_{i+1})$$

$$= (\Sigma^0)^{-1} \sum_{t=T_j+1}^{T_{j+1}} [(y_t - x_t'\tilde{\beta})(y_t - x_t'\tilde{\beta})' - (y_t - x_t'\hat{\beta})(y_t - x_t'\hat{\beta})']$$

$$= o_p(1).$$

Also,

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

Hence,

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

Q.E.D.

PROOF OF COROLLARY 2: Note that, because H is a selection matrix applied to $\operatorname{vec}(\Sigma)$, 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 ith position and 0 elsewhere. Hence, assuming Normality, any element of $H\Omega H'$ that involves the selection of the (i,k) and (l,m) element of Σ can be written, in view of (12), 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 (19) 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 1's, that is, $(H\Omega H')$ is a $2n_b \times 2n_b$ matrix of 1's. Then straightforward algebra reveals that the limiting distribution is still given by (19). *Q.E.D.*

PROOF OF THEOREM 7: Assume, without loss of generality, no disjoint break are allowed under the alternative hypothesis and all regression coefficients are 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

$$\Lambda_\varepsilon^* = \big\{(k_1, k_2);\, \varepsilon T \le k_1 \le k_2 \le (1-\varepsilon)T \text{ and } v_T^2(k_2 - k_1) \le M_T \text{ with } M_T \to 0, v_T \to 0, \text{ and } T^{1/2}v_T/(\log T)^2 \to \infty \text{ as } T \to \infty\big\}.$$

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 the 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,

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

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,

$$\begin{aligned} T\log|\hat{\Sigma}^*| - T\log|\hat{\Sigma}| &= \operatorname{tr}(T\Sigma_0^{-1}(\hat{\Sigma}^* - \hat{\Sigma})) + \frac{T}{2}\operatorname{tr}\big([(\Sigma^0)^{-1}(\hat{\Sigma} - \Sigma^0)]^2\big) \\ &\quad - \frac{T}{2}\operatorname{tr}\big([(\Sigma^0)^{-1}(\hat{\Sigma}^* - \Sigma^0)]^2\big) + o_p(T^{-1}) \\ &= \operatorname{tr}(T\Sigma_0^{-1}(\hat{\Sigma}^* - \hat{\Sigma})) + o_p(1), \end{aligned}$$

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

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

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

uniformly in $(k_1, k_2) \in \Lambda_\varepsilon^*$. Let $\hat{\beta}_t$ be the estimate under the locally ordered break model: $\hat{\beta}_t = (\hat{\beta}_{1,1}', \hat{\beta}_{2,1}')'$ if $t \le k_1$, $\hat{\beta}_t = (\hat{\beta}_{1,2}', \hat{\beta}_{2,1}')'$ if $k_1 < t \le k_2$, and $\hat{\beta}_t = (\hat{\beta}_{1,2}', \hat{\beta}_{2,2}')'$ if $t > k_2$. Also let $\hat{\beta}_t^*$ be the estimate under the common break model: $\hat{\beta}_t^* = (\hat{\beta}_{1,1}^{*\prime}, \hat{\beta}_{2,1}^{*\prime})'$ if $t \le k_1$ and $\hat{\beta}_t^* = (\hat{\beta}_{1,2}^{*\prime}, \hat{\beta}_{2,2}^{*\prime})'$ if $t > k_1$. Then, for a given partition $(k_1, k_2) \in \Lambda_\varepsilon^*$, simple arguments lead to $\hat{\beta}_{1,j} - \hat{\beta}_{1,j}^* = O_p((Tv_T)^{-1}\log v_T^{-2})$ and $\hat{\beta}_{2,j} - \hat{\beta}_{2,j}^* = O_p((Tv_T)^{-1}\log v_T^{-2})$ for $j = 1, 2$, which further implies

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

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 $(k_1, k_2) \in \Lambda_\varepsilon^*$. Using the fact that $(k_2 - k_1)/T \to 0$, we have $\operatorname{tr}(\sum_{t=k_1+1}^{k_2} x_t'(\hat{\beta}_t - \beta^0)(\hat{\beta}_t - \beta^0)'x_t) = o_p(1)$, $\operatorname{tr}(\sum_{t=k_1+1}^{k_2} u_t(\hat{\beta}_t - \beta^0)'x_t) = o_p(1)$, $\operatorname{tr}(\sum_{t=k_1+1}^{k_2} x_t'(\hat{\beta}_t^* - \beta^0)(\hat{\beta}_t^* - \beta^0)'x_t) = o_p(1)$, and $\operatorname{tr}(\sum_{t=k_1+1}^{k_2} u_t(\hat{\beta}_t^* - \beta^0)'x_t) = o_p(1)$. 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. *Q.E.D.*

PROOF OF THEOREM 8: Without loss of generality, assume all the coefficients are subject 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

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

with the corresponding log-likelihood function being

$$\mathcal{L}_T^G(k_1, k_2) = -\frac{T}{2}(\log 2\pi + 1) - \frac{T}{2}\log|\hat{\Sigma}|,$$

with

$$\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}(y_t - x_t'\hat{\beta}_t)(y_t - x_t'\hat{\beta}_t)'.$$

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

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

with the corresponding likelihood function being

$$\mathcal{L}_T^G(1, k_2) = -\frac{T}{2}(\log 2\pi + 1) - \frac{T}{2}\log|\tilde{\Sigma}_2|,$$

with

$$\tilde{\Sigma}_2 = \frac{1}{T}\sum_{t=1}^{T}(y_t - x_t'\tilde{\beta}_t)(y_t - x_t'\tilde{\beta}_t)'.$$

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

$$\begin{aligned} &LR_T^G(k_1, k_2, p_{b1}, p_{b2}, \varepsilon) \\ &\quad = T(\log|\tilde{\Sigma}_2| - \log|\hat{\Sigma}|) + T(\log|\tilde{\Sigma}| - \log|\tilde{\Sigma}_2|). \end{aligned}$$

The likelihood ratio is the sum of two components, each involving only one break, with some coefficients restricted not to change. Given this, the rest of the proof follows that of Theorem 5, that is, the analysis of $F_{1,T}^j$. *Q.E.D.*

THE LIMIT DISTRIBUTION OF THE STRUCTURAL CHANGE TEST IN THE CASE OF SWITCHING REGIMES

Consider a situation where the system switches from regime 1 to regime 2, then switches back to regime 1. This type of phenomenon was noticed by Sensier and van Dijk (2004), who argued that the volatility of the time series of aggregate price indices showed 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 maximized value of the likelihood function under the alternative hypothesis of two changes but imposing the restriction that the first and the third regimes are the same, and imposing that the maximization is taken over the 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 is omitted.

THEOREM S.1: *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 *and* A12 (*assuming Normal errors when allowing changes in the covariance matrix of the errors*),

$$\begin{aligned}\sup LR_T^S \Rightarrow \sup_{(\lambda_1,\lambda_2)\in\Lambda_\varepsilon} &\big\| [W_{p_b+n_b}(\lambda_2) - \lambda_2 W_{p_b+n_b}(1)] \\ &- [W_{p_b+n_b}(\lambda_1) - \lambda_1 W_{p_b+n_b}(1)] \big\|^2 \\ &/((\lambda_2 - \lambda_1)(1 - \lambda_2 + \lambda_1)).\end{aligned}$$

Dept. of Economics, University of Illinois at Urbana–Champaign, 1206 S. Sixth St., Champaign, IL 61820, U.S.A.; zqu@uiuc.edu

and

Dept. of Economics, Boston University, 270 Bay State Rd., Boston, MA 02215, U.S.A.; perron@bu.edu.

REFERENCES

BAI, J. R. (2000): "Vector Autoregressive Models with Structural Changes in Regression Coefficient and in Variance–Covariance Matrices," *Annals of Economics and Finance*, 1, 303–339. [2,11,14]

BAI, J., R. L. LUMSDAINE, AND J. H. STOCK (1998): "Testing for and Dating Common Breaks in Multivariate Time Series," *Review of Economic Studies*, 65, 395–432. [2-4,8-12]

BAI, J., AND P. PERRON (1998): "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometica*, 66, 44–48. [1,26]

CORRADI, V. (1999): "Deciding Between I(0) and I(1) via FLIL-Based Bounds," *Econometric Theory*, 15, 643–663. [1]

DOMOWITZ, I., AND H. WHITE (1982): "Misspecified Models with Dependent Observations," *Journal of Econometrics*, 20, 35–58. [3]

EBERLEIN, E. (1986): "On Strong Invariance Principles under Dependent Assumptions," *The Annals of Probability*, 14, 260–270. [1]

HALL, P., AND C. C. HEYDE (1980): *Martingale Limit Theory and Its Applications*. New York: Academic Press. [7]

HANSEN, B. E. (1991): "Strong Laws for Dependent Heterogeneous Processes," *Econometric Theory*, 7, 213–221. [1]

IBRAGIMOV, I. A. (1962): "Some Limit Theorems for Stationary Processes," *Theory of Probability and Its Applications*, 7, 349–382. [1]

MAGNUS, J. R. (1988): *Linear Structures.* London: Charles Griffin & Co. [32]

SENSIER, M., AND D. VAN DIJK (2004): "Testing for Volatility Changes in U.S. Macroeconomic Time Series," *The Review of Economics and Statistics*, 86, 833–839. [36]