Inference on Locally Ordered Breaks in Multiple Regressions*

Ye Li‡  Pierre Perron‡
Moody’s Analytics  Boston University

November 3, 2012; Revised: February 2, 2015

Abstract

We consider issues related to inference about locally ordered breaks in a system of equations, as originally proposed by Qu and Perron (2007). These apply when break dates in different equations within the system are not separated by a positive fraction of the sample size. This allows constructing joint confidence intervals of all such locally ordered break dates. We extend the results of Qu and Perron (2007) in several directions. First, we allow the covariates to be any mix of trends and stationary or integrated regressors. Second, we allow for breaks in the variance-covariance matrix of the errors. Third, we allow for multiple locally ordered breaks, each occurring in a different equation within a subset of equations in the system. Via some simulation experiments, we show first that the limit distributions derived provide good approximations to the finite sample distributions. Second, we show that forming confidence intervals in such a joint fashion allows more precision (tighter intervals) compared to the standard approach of forming confidence intervals using the method of Bai and Perron (1998) applied to a single equation. Simulations also indicate that using the locally ordered break confidence intervals yields better coverage rates than using the framework for globally distinct breaks when the break dates are separated by roughly 10% of the total sample size.

Keywords: locally ordered breaks, multiple regressions, change-points, break dates.
JEL Classification: C32, C33.

*We are especially grateful to Zhongjun Qu for helpful comments, as well as three referees.
‡Senior economist, Moody’s Analytics, 121 N. Walnut Street, West Chester, PA, 19380 (liye1983@gmail.com).
‡Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (perron@bu.edu).
1 Introduction

Issues related to structural breaks have received a lot of attention in the statistics and econometrics literature (see Perron, 2006, for a survey). In the last fifteen years, substantial advances have been made in the econometrics literature to cover more general models in the context of estimating and testing structural breaks in both single equation and multiple equations systems. Bai (1997) studies the least squares estimation of a single change point in regressions involving stationary and/or trending regressors. He derives the consistency, rate of convergence and the limiting distributions of change point estimates under general conditions on the regressors and error terms. Bai and Perron (1998) extend the testing and estimation analysis to the case of multiple structural changes, and Bai and Perron (2003) present an efficient algorithm to obtain the break date estimates, which minimizes the global sum of squared residuals. Perron and Qu (2006) considerably relax the conditions used in Bai and Perron (1998) and analyze models in which restrictions within or across regimes are allowed. Kejriwal and Perron (2008a, 2010) consider issues related to estimation and testing for multiple structural breaks in a single cointegrating equation.

Work related to structural changes in a multiple equations system is comparatively scarce. Bai, Lumsdaine, and Stock (1998) consider inference procedures for the estimate of a single break date in multivariate times series. They show that the accuracy of break-point estimators is not much improved with simply having more observations, but can be improved when considering a system of series with common breaks. Also Bai (2000) considers the estimation of multiple structural break points in a VAR system with stationary regressors allowing both the coefficients of the regression model and those of the variance-covariance matrix to change. He derives the consistency, rate of convergence and asymptotic distribution for the estimates of the break dates. Qu and Perron (2007) cover the more general case of multiple structural changes occurring at unknown dates in linear multivariate regression models that include, among others, vector autoregressions, certain linear panel data models, and seemingly unrelated regression. They also introduce a novel structure that was labelled as “locally ordered breaks”. Oka and Perron (2011) address the issue of testing for common breaks across or within equations in multiple equations systems with stationary, trending and unit-root regressors.

With the exception of Qu and Perron (2007) who considered models with regime-wise stationary covariates, the class of models considered so far in the structural break literature consider break dates modelled as being asymptotically distinct in the sense that each regime
is separated by a positive fraction of the sample size. So asymptotically, as the total sample size increases, the number of observations within each regime increases proportionally. This rules out a class of models that may have wide appeal in practice whereby the breaks across equations are close to each other and, hence, cannot be considered as asymptotically distinct so that the estimates can be treated independently when considering inference. In the terminology of Qu and Perron (2007), these are “locally ordered breaks”. This theoretical setup allows constructing joint confidence intervals of all such locally ordered break dates. Qu and Perron (2007) provide appropriate methods for estimation, inference and testing of locally ordered breaks in multiple regression systems with stationary regressors.

The aim of this paper is to extend their analysis in several directions. First, we allow the covariates to be any mix of trends and stationary or integrated regressors (i.e., having an autoregressive unit root). Second, we allow for breaks in the variance-covariance matrix of the errors. Third, we allow for multiple locally ordered breaks, each occurring in a different equation within a subset of equations in the system. In order to do so we adopt the framework and use some results of Oka and Perron (2011).

Via some simulation experiments, we show first that the limit distributions derived provide good approximations to the finite sample distributions. Second, we show that forming confidence intervals in such a joint fashion allows more precision (tighter intervals) compared to the standard approach of forming confidence intervals using the method of Bai and Perron (1998) applied to a single equation. Simulations also indicate that using the locally ordered break confidence intervals yields better coverage rates than using the framework for globally distinct breaks when the break dates are separated by roughly 10% of the total sample size.

The structure of the paper is as follows. Section 2 presents the general framework adopted and the assumptions imposed on the regressors and the errors, as well as some preliminary limit results. Section 3 provides the details about the method of estimation based on quasi-maximum likelihood. In Section 4, we present our main theoretical results pertaining to the consistency, rate of convergence and joint limit distributions of the locally ordered breaks. Section 5 discusses the results obtained from simulations about the adequacy of the asymptotic distributions in providing good approximations in finite samples. Section 6 provides brief concluding remarks. The appendix contains theoretical derivations for the case with no change in the covariance matrix of the errors. A supplementary document available online contains results and derivations for cases involving changes in both the coefficients and the covariance-matrix of the errors.
2 Model and Assumptions

We adopt a framework and assumptions similar to those in Oka and Perron (2011), Qu and Perron (2007) and Kejriwal and Perron (2008a). 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 \( T = (T_1, \ldots, T_{m^*}) \) and we use the convention that \( T_0 = 0 \) and \( T_{m^*+1} = T \). A subscript \( j \) indexes a regime \((j = 1, \ldots, m^*+1)\). A subscript \( t \) indexes a temporal observation \((t = 1, \ldots, T)\) and a subscript \( i \) indexes the equation \((i = 1, \ldots, n)\) to which a scalar dependent variable \( y_{it} \) is associated. The parameter \( q \) is the number of regressors and \( h_{tT} = (h_{1tT}, \ldots, h_{qtT})' \) is the set that includes the regressors from all equations. Let \( y_t = (y_{1t}, \ldots, y_{nt})' \) and \( u_t = (u_{1t}, \ldots, u_{nt})' \), the model considered is

\[
y_t = (h_{tT} \otimes I_n)S\beta_{(j)} + u_t,
\]

where \( I_n \) is a \( n \) by \( n \) identity matrix, \( S \) is a selection matrix, and \( u_t \) is an error term having mean 0 and covariance matrix \( \Sigma_{(j)} \) for \( T_{j-1} + 1 \leq t \leq T_j \) \((j = 1, \ldots, m^* + 1)\). The set of regressors includes integrated processes, trends and stationary ones, specified by

\[
h_{tT}' = (T^{-1/2}z_t', T^{-1}t, x_t'),
\]

where the scaling is introduced so that the order of all components is the same. The \( q_x \times 1 \) vector \( x_t \) contains the stationary regressors, while the \( q_z \times 1 \) vector \( z_t \) the integrated ones, so that \( h_{tT} \) is a \( q \equiv q_x + 1 + q_z \) vector. These are defined by

\[
\begin{align*}
z_t &= z_{t-1} + u_{zt}, \\
x_t &= \mu_x + u_{xt},
\end{align*}
\]

where \( z_0 \) is assumed, for simplicity, to be a vector of either \( O_p(1) \) random variables or fixed finite constants. By labelling the regressors \( x_t \) as \( I(0) \), we mean that the partial sums of the associated noise components satisfy a functional central limit theorem. The conditions imposed are discussed below. We then label a variable as \( I(1) \) if it is the accumulation of an \( I(0) \) process.

As will be made precise below, in all cases \( u_t \) is assumed to be an \( I(0) \) process. Hence, the stochastic properties of the dependent variable \( y_t \) depends of the nature of the regressors included. For instance if the regression includes both \( I(1) \) and \( I(0) \) regressors then \( y_t \) is \( I(1) \) and cointegrated with the \( I(1) \) regressors. An example of this specification is the dynamic OLS regression (Saikkonen, 1991, Kejriwal and Perron, 2008b) whereby the estimate of a
The cointegrating vector is obtained by augmenting the static cointegrating relation with leads and lags of the first-differences of the $I(1)$ right-hand side variables. A trend can also be included in practice.

The set of basic parameters in regime $j$ consists of the $p$ vector $\beta_{(j)}$ and the $n \times n$ matrix $\Sigma_{(j)}$. The matrix $S$ is of dimension $nq \times p$ 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. We allow for the imposition of a set of $r$ restrictions of the form:

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

where $\beta = (\beta_{(1)}, ..., \beta_{(m^*+1)})'$, $\Sigma = (\Sigma_{(1)}, ..., \Sigma_{(m^*+1)})$ 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. For a discussion of how general the framework is, see Qu and Perron (2007). For example, a common set of restrictions is used to have a partial structural change model by imposing equality of some coefficients across regimes. To ease notation, define the $n \times p$ matrix $X'_{IT} = (h'_{IT} \otimes I_n)S$, so that (1) becomes

$$y_t = X'_{IT}\beta_{(j)} + u_t,$$

for $T_{j-1} + 1 \leq t \leq T_j$ ($j = 1, ..., m^* + 1$). It is useful to express the model in matrix form. Let $Y = (y_1', ..., y_T')'$ be the $nT$ vector of dependent variables, $U = (u_1', ..., u_T')'$ be the error vector and the $nT$ by $p$ matrix of regressors is $X = (X_{1T}, ..., X_{m^*+1}T)$. For a given partition of the sample using the break dates $(T_1, ..., T_{m^*})$, we define the block diagonal partition of the matrix $X$ as the $nT$ by $p(m^*+1)$ matrix $\bar{X} = \text{diag}(X_1, ..., X_{m^*+1})$ where $X_j (j = 1, ..., m^*+1)$ is the $n(T_j - T_{j-1})$ by $p$ subset of $X$ that corresponds to observations in regime $j$. Then the regression system (5) 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 $(T_1^0, ..., T_{m^*}^0)$. Also, the true covariance matrix of error terms is denoted by $\Sigma_{(j)}^0$ for each regime $j = 1, ..., m^* + 1$.

Let the break fractions be defined by $T_j^0 = [T\lambda_j^0]$ ($j = 1, ..., m^*$). We assume that

$$0 < \lambda_1^0 < ... < \lambda_t^0 < \lambda_{t+1}^0 \leq ... \leq \lambda_{t+m}^0 < \lambda_{t+m+1}^0 < ... < \lambda_{m^*}^0 < 1.$$  

This stipulates that the $m$ break dates $(T_{t+1}^0, ..., T_{t+m}^0)$ need not be separated by a positive fraction of the sample size $T$, while the others are. These will be the $m$ locally ordered
breaks considered in this paper. Since our main interest lies in the estimates of these locally ordered breaks, for ease of notation, we shall label them as \((K_0^0, \ldots, K_m^0) = (T_{l+1}^0, \ldots, T_{l+m}^0)\). As a matter of convention, we shall also denote \(K_0^0 = T_l^0\) and \(K_{m+1}^0 = T_{l+m+1}^0\). The conditions imposed on these break dates are stated in the following definition.

**Definition 1** Locally Ordered Breaks (LOB): Let \(v_T\) be a sequence of positive numbers that satisfies \(v_T \to 0\) and \(T^{1/2}v_T/(\log^2 T) \to \infty\). The break dates \((K_0^0, \ldots, K_m^0)\), assumed to each occur in a different equation within a subset of \(m (\leq n)\) equations, are said to be locally ordered if \(K_0^0 < \ldots < K_m^0\), with the differences such that 1) \(v_T^2(K_0^0 - K_1^0) \leq M_T\) with \(M_T \to 0\) as \(T \to \infty\); 2) \((K_s^0 - K_1^0)/(\log^2 T) \to \infty\), for \(s = 2, \ldots, m\).

The condition (1) implies that \((K_s^0 - K_1^0)/T \to 0\) and imposes an upper bound on the distance between the break dates. Hence, asymptotically the distances between the break dates become a negligible portion of the sample size. Note, however, that since each of the locally ordered breaks are assumed to belong to distinct equations, each within-equation regime contains a positive fraction of the total sample size. The condition (2) imposes a lower bound of the distance between the break dates so that, asymptotically, the sample size between each regime increases, albeit at a slow rate. The lower bound departs from the definition of locally ordered breaks in Qu and Perron (2007) so that the distance between the break dates increases with the sample size but at a slow enough rate. This allows models with heterogeneity across segments and models with lagged dependent variables, which were not possible in the original treatment of Qu and Perron (2007).

Following Qu and Perron (2007), note that testing for structural changes can be performed by searching for breaks in the set

\[
\Lambda^*_\varepsilon = \{(T_1, \ldots, T_l, K_1, \ldots, K_m, T_{l+m+1}, \ldots, T_m^*): \\
|T_{i+1} - T_i| \geq \varepsilon T \text{ for } i = 0, \ldots, l, l + m + 1, \ldots, m^*; \\
(K_1 - T_l) \geq \varepsilon T, \ (T_{l+m+1} - K_m) \geq \varepsilon T \text{ and } v_T^2(K_j - K_{j-1}) \leq M_T \text{ for } j = 1, \ldots, m, \\
\text{with } M_T \to 0, \ v_T \to 0 \text{ and } T^{1/2}v_T/(\log T)^2 \to \infty \text{ as } T \to \infty \}
\]

The various tests discussed in Qu and Perron (2007) remain valid and will have the same limit distribution as when constructed assuming \(m^* - m\) asymptotically distinct break.

As a matter of notation, \(\overset{p}{\Rightarrow}\) denotes convergence in probability, \(\overset{d}{\Rightarrow}\) convergence in distribution and \(\Rightarrow\) weak convergence in the space \(D[0, 1]\) under the Skorohod topology. Also, define the \(L_r\)-norm of a random matrix \(X\) as \(\|X\|_r = (\sum_{(i)} \sum_{(j)} E|X_{ij}|^r)^{1/r}\) for \(r \geq 1\).
We make the following assumptions on the regressors and the elements of the noise component 
\[ \zeta_t = (u'_t, u'_{zt}, u'_{xt})' \].

**Assumption 1** Let \( H = (h_{1T}, \ldots, h_{TT})' \) and \( \tilde{H}_0 \) be the diagonal partition of \( H \) at \((T_1^0, \ldots, T_{m^*})\) such that \( \tilde{H}_0 = \text{diag}(H_0^0, \ldots, H_{nT}^0) \). For each \( j = 1, \ldots, m^* + 1 \), \( T^{-1}H_j^0 H_j^0 \) converges to a (possibly) random matrix, not necessarily the same for all \( j \).

**Assumption 2** There exists a \( l_0 > 0 \) such that for all \( l > l_0 \), the minimum eigenvalues of \( A_{jl} = (1/l) \sum_{t=T_j^0+1}^{T_j^0+l} h_{tt} h_{tt}' \) and \( A_{jl}^* = (1/l) \sum_{t=K_j^0-l}^{K_j^0-1} h_{tt} h_{tt}' \) are bounded away from zero, for \( j = 1, \ldots, m^* \).

**Assumption 3** The matrix \( B_{ik} = \sum_{t=k}^l h_{tt} h_{tt}' \) is invertible for any \( l > 0 \) and \( k > 0 \) such that \( l - k \geq p \).

**Assumption 4** Let \( \mathcal{F}_t = \sigma - \text{field} \{ \ldots, \zeta_{t-1}, \zeta_t, \ldots, u_{t-2}, u_{t-1} \} \). If \( \zeta_t \) is weakly stationary within each segment, then (a) \( \{ \zeta_t, \mathcal{F}_t \} \) forms a strongly mixing (\( \alpha \)-mixing) sequence with size \( -4r/(r - 2) \) for some \( 2 < r < 8 \). (b) \( E(\zeta_t) = 0 \) and sup \( \| \zeta_t \|_2 \) \( < M \) \( \leq 1 \) \( \zeta_t \) has a positive definite matrix \( \Omega = [\omega_{i,s}] \) such that for any \( i, s = 1, \ldots, p \), we have, uniformly in \( l \), \( |k^{-1} \text{E}(\langle S_{k,j}(l) \rangle_i (S_{k,j}(l))_s) - \omega_{i,s}| \leq C_2 k^{-\psi} \), for some \( C_2, \psi > 0 \). It is also assumed that \( \{u_t u_t' - \Sigma_{(j)}^0\} \) satisfies the conditions stated in this assumption.

**Assumption 5** \( E(u_{zt} \otimes u_t) = 0 \).

**Assumption 6** For \( j = 1, \ldots, m^* + 1 \), \( \beta_{(j+1)}^0 - \beta_{(j)}^0 = \nu_T \delta_{(j)} \) and \( \Sigma_{(j+1)}^0 - \Sigma_{(j)}^0 = \nu_T \Phi_{(j)} \) for some \( \delta_{(j)} \) and \( \Phi_{(j)} \) independent of \( T \), with \( \nu_T \to 0 \) and \( T^{1/2} \nu_T / (\log^2 T) \to \infty \) as \( T \to \infty \).

**Assumption 7** The break fractions \( \{\lambda_1^0, \ldots, \lambda_{m^*}^0\} \) satisfy (6) with \( (K_1^0, \ldots, K_{m^*}^0) = (T_{j+1}^0, \ldots, T_{j+m^*}^0) \) being locally ordered as stated in Definition 1.

**Assumption 8** \( T^{-1/2} \sum_{t=1}^{T_r} u_{zt} \Rightarrow \Omega_z^{1/2} W_z(r) \) uniformly in \( r \in [0, 1] \) with \( \Omega_z \) positive definite.

**Assumption 9** For all \( t \) and \( s \): a) \( E(\langle u_{zt} \otimes u_t \rangle_s) = 0 \), b) \( E(\langle u_{zt} \otimes u_t \rangle u_s') = 0 \), c) \( E(\langle u_{zt} \otimes u_t \rangle u_{zst}') = 0 \), d) \( E[u_{tk} u_{lt} u_{th}] = 0 \) for all \( k, l, h \) and for every \( t \), e) \( E[u_{zt} u_{lt} u_{th}] = 0 \) for all \( k, l, h \) and for every \( t \).
Assumption A1 is needed for multiple linear regressions involving both stationary and integrated regressors, it requires that the sample moments of the regressors exists. Assumption A2 ensures that there is no local collinearity problem so that the break dates can be identified. Assumption A3 is the standard invertibility requirement to have well defined estimates. Assumption A4 determines the dependence structure of the processes $\zeta_t \otimes u_t$ and $\zeta_t$. In particular, they imply that $\zeta_t \otimes u_t$ and $\zeta_t$ are short memory processes having bounded fourth moments. The assumptions are imposed to obtain a functional central limit theorem, a generalized Hajek and Renyi (1955) type inequality and a strong law of large numbers that allow us to show the estimates of the break dates are consistent and to derive the rate of convergence. The conditions are mild in the sense that they allow for substantial conditional heteroskedasticity and autocorrelation. Assumption A5 specifies that the stationary regressors are contemporaneously uncorrelated with the errors, which is used to obtain consistent estimates. It can be relaxed by interpreting the coefficients as the pseudo-true values, i.e., as the limit in probability of the inconsistent estimates. As shown in Perron and Yamamoto (2015), this still permits consistent estimation of the break fractions and the confidence intervals for the estimates can be constructed in the usual manner. Assumption A6 implies a shrinking shift asymptotic framework whereby the magnitudes of the shifts converge to zero as the sample size increases. This allows the development of a limiting theory for the break date estimates which does not depend on the exact distributions of regressors and the errors. Note that $v_T$ in Assumption 6 is the same as in Definition 1 (LOB). Since this quantity will not appear in the limit distribution, there is no need to specify its value in practice. It is simply a theoretical device to obtain non-degenerate limit distributions.

For the integrated regressors, things are different and we need to impose a homogenous distribution throughout the sample as stated in Assumption A8. Allowing for heterogeneity in the distribution of the errors underlying the $I(1)$ regressors would be considerably more difficult. Instead of having a limit distribution in terms of standard Wiener processes, we would have time-deformed Wiener processes according to the variance profile of the errors through time; see, e.g., Cavaliere and Taylor (2007). This would lead to important complications given that, as shown below, the limit distribution of the estimates of the break dates depends on the whole time profile of the limit Wiener processes. The requirement that $\Omega_2$ be positive definite rules out cointegration among the $I(1)$ regressors and is needed to ensure a set of regressors that has a positive definite limit. To discuss the conditions imposed by Assumption A9, it is useful to first describe the implied limit distributions of various sample moments. Since our interest is about the estimates of the locally ordered
breaks, we consider results pertaining to segments defined by these locally ordered breaks. Let \( \triangle K_j^0 = K_j^0 - K_{j-1}^0 \) for \( j = 1, \ldots, m + 1 \), we have, as \( \triangle K_j^0 \to \infty \), uniformly in \( s \in [0, 1] \),

\[
(\triangle K_j^0)^{-1} \sum_{t = K_{j-1}^0}^{K_j^0 + [s \triangle K_j^0]} x_t x_t' \overset{p}{\to} sQ_{x,j}
\]

\[
(\triangle K_j^0)^{-1} \sum_{t = K_{j-1}^0}^{K_j^0 + [s \triangle K_j^0]} u_t u_t' \overset{p}{\to} s\Sigma_{x,j}^0
\]

Also, with \( \eta_t = (\Sigma_{x,j}^0)^{-1/2}u_t \) for \( K_{j-1}^0 + 1 \leq t \leq K_j^0 \), we have for \( j = 1, \ldots, m + 1 \),

\[
(\triangle K_j^0)^{-1/2} \sum_{t = K_{j-1}^0 + [s \triangle K_j^0]}^{K_j^0} \eta_t \Rightarrow W_{\eta,j}^{(1)}(s) \quad 0 < s \leq 1
\]

\[
(\triangle K_j^0)^{-1/2} \sum_{t = K_{j-1}^0 + [s \triangle K_j^0]}^{K_j^0} \eta_t \Rightarrow W_{\eta,j}^{(2)}(s) \quad -1 \leq s < 0
\]

where the weak convergence is in the space \( D[0, 1]^n \) and \( \{W_{\eta,j}^{(1)}(s), W_{\eta,j}^{(2)}(s)\} \) are Brownian motion processes defined on \( [0, 1]^n \) with covariance matrix \( (\Sigma_{x,j}^0)^{-1}\Omega_{\eta,j} \) where

\[
\Omega_{\eta,j} = \lim_{T \to \infty} \text{var}\{(\triangle K_j^0)^{-1/2} \sum_{t = K_{j-1}^0 + 1}^{K_j^0} u_t\}
\]

We define the two-sided Brownian motion \( W_{\eta,j}(s) \) satisfying \( W_{\eta,j}(s) = W_{\eta,j}^{(1)}(s) \) for \( s < 0 \), \( W_{\eta,j}(s) = 0 \) for \( s = 0 \) and \( W_{\eta,j}(s) = W_{\eta,j}^{(2)}(s) \) for \( s > 0 \). It will also be useful to define \( W_{\eta,j}^*(s) = (\Sigma_{x,j}^0)^{-1/2}\Omega_{\eta,j}^{-1/2}W_{\eta,j}(s) \). Also

\[
(\triangle K_j^0)^{-1/2} \sum_{t = K_{j-1}^0 + [s \triangle K_j^0]}^{K_j^0} (\eta_t \eta_t' - I_n) \Rightarrow \xi_{\eta,j}^{(1)}(s) \quad 0 < s \leq 1
\]

\[
(\triangle K_j^0)^{-1/2} \sum_{t = K_{j-1}^0 + [s \triangle K_j^0]}^{K_j^0} (\eta_t \eta_t' - I_n) \Rightarrow \xi_{\eta,j}^{(2)}(s) \quad -1 \leq s < 0
\]

where the weak convergence is in the space \( D[0, 1]^{n^2} \) and where the entries of the \( n \times n \) matrices \( \xi_{\eta,j}^{(1)}(s) \) and \( \xi_{\eta,j}^{(2)}(s) \) are Brownian motion processes defined on \([0, 1] \) with covariance matrix

\[
\Omega_{\xi,j} = \lim_{T \to \infty} \text{var}\{(\triangle K_j^0)^{-1/2} \sum_{t = K_{j-1}^0 + 1}^{K_j^0} (\eta_t \eta_t' - I_n)\}
\]

We define the two-sided Brownian motion \( \xi_{\eta,j}(s) \) satisfying \( \xi_{\eta,j}(s) = \xi_{\eta,j}^{(1)}(s) \) for \( s > 0 \), \( \xi_{\eta,j}(s) = 0 \) for \( s = 0 \) and \( \xi_{\eta,j}(s) = \xi_{\eta,j}^{(2)}(s) \) for \( s < 0 \). We also have

\[
(\triangle K_j^0)^{-1/2} \sum_{t = K_{j-1}^0 + [s \triangle K_j^0]}^{K_j^0} x_t \otimes u_t \Rightarrow M_{x,j}^{1/2}W_{x,j}^{(1)}(s) \quad 0 < s \leq 1
\]

\[
(\triangle K_j^0)^{-1/2} \sum_{t = K_{j-1}^0 + [s \triangle K_j^0]}^{K_j^0} x_t \otimes u_t \Rightarrow M_{x,j}^{1/2}W_{x,j}^{(2)}(s) \quad -1 \leq s < 0
\]

where

\[
M_{x,j} = \lim_{T \to \infty} \text{var}\{(\triangle K_j^0)^{-1/2} \sum_{t = K_{j-1}^0 + 1}^{K_j^0} x_t \otimes u_t\}
\]
with \(W_{x,j}(s)\) and \(W_{x,j}(s)\) independent Gaussian processes. We define \(W_{x,j}(s)\) as a two-sided standard multivariate Gaussian process such that \(W_{x,j}(s) = W_{x,j}^{(1)}(s)\) for \(s < 0\), \(W_{x,j}(s) = 0\) for \(s = 0\) and \(W_{x,j}(s) = W_{x,j}^{(2)}(s)\) for \(s > 0\). Also \(M_{x,j}\) is a nonrandom positive definite matrix not necessarily the same for all \(j\).

Assumption A9 restricts somewhat the class of models applicable but is quite mild. Sufficient, though not necessary, conditions for it to hold are: for (a) that the \(I(0)\) regressors are uncorrelated with the errors contemporaneously even conditional on the \(I(1)\) variables; for (b) that the autocovariance structure of the \(I(0)\) regressors be independent of the errors and, similarly, for (c) that the autocovariance structure of the errors be independent of the \(I(0)\) regressors. This assumption is needed to guarantee that \(W_{x,j}(\cdot)\) and \(W_{z}(\cdot)\) are uncorrelated with \(W_{\eta,j}(\cdot)\) and, being Gaussian, are therefore independent (these are the same conditions used in Kejriwal and Perron, 2008a). Without these conditions, the analysis would be much more complex. Similarly, part (d) implies that \(W_{x,j}(s)\) and \(W_{\eta,j}(\cdot)\) are independent of \(\xi_{j}(\cdot)\) (this is the same condition used in Qu and Perron, 2007) and part (e) implies that \(W_{z}(\cdot)\) and \(\xi_{j}(\cdot)\) are independent. Note that \(W_{x,j}(\cdot)\) and \(W_{z}(\cdot)\) can be correlated.

3 Estimation

In the sequel, we shall only consider the locally ordered breaks and ignore those whose break fractions are asymptotically distinct. This will be sufficient to obtain the relevant joint limit distributions of the locally ordered breaks. Inference about the estimates of the non-locally ordered breaks follows using the method developed in Qu and Perron (2007) and Kejriwal and Perron (2008a). Hence, we suppose that the system of equations contains the break dates \((K_{1}^{0}, ..., K_{m}^{0})\) with the convention that \(T_{l}^{0} = K_{0}^{0}\) and \(K_{m+1}^{0} = T_{m+1}^{0}\).

The method of estimation considered here is restricted quasi-maximum likelihood that assumes serially uncorrelated Gaussian errors. Conditional on a given partition of the sample \(K = (K_{1}, ..., K_{m})\), the Gaussian quasi-likelihood function is

\[
L_{T}(K, \beta, \Sigma) = \prod_{j=1}^{m+1} \prod_{t=K_{j-1}+1}^{K_{j}} f(y_{t}|X_{Tt}; \beta_{(j)}, \Sigma_{(j)})
\]

where

\[
f(y_{t}|X_{Tt}; \beta_{(j)}, \Sigma_{(j)}) = \frac{1}{(2\pi)^{1/2}|\Sigma_{(j)}|^{1/2}} \exp\left\{-\frac{1}{2}(y_{t} - X'_{Tt}\beta_{(j)})\Sigma_{(j)}^{-1}(y_{t} - X'_{Tt}\beta_{(j)})\right\}
\]

and the quasi-likelihood ratio is

\[
LR_{T}(K, \beta, \Sigma) = \frac{\prod_{j=1}^{m+1} \prod_{t=K_{j-1}+1}^{K_{j}} f(y_{t}|X_{Tt}; \beta_{(j)}, \Sigma_{(j)})}{\prod_{j=1}^{m+1} \prod_{t=K_{j-1}+1}^{K_{j}} f(y_{t}|X_{Tt}; \beta_{(j)}^{0}, \Sigma_{(j)}^{0})}
\]
We want to obtain the estimates \((\hat{K}_1, ..., \hat{K}_m, \hat{\beta}_{(j)}, \hat{\Sigma}_{(j)})\) as the values of \((K_1, ..., K_m, \beta_{(j)}, \Sigma_{(j)})\) which maximize \(LR_T\) subject to restrictions \(g(\beta, vec(\Sigma)) = 0\). Let \(lr_T(\cdot)\) denotes the log-likelihood ratio and \(rlr_T(\cdot)\) denotes the restricted log-likelihood ratio, the objective function is then

\[
rlr_T(K, \beta, \Sigma) = lr_T(K, \beta, \Sigma) + \lambda g(\beta, vec(\Sigma))
\]

and the estimates are

\[
(K, \beta, \Sigma) = \arg \max_{(K_1, ..., K_m, \beta, \Sigma)} rlr_T(K, \beta, \Sigma).
\] (8)

where the supremum is taken over the set

\[
k_\varepsilon = \{(K_1, ..., K_m) : T_i - K_1 > [T\varepsilon], K_2 - K_1 \geq h, ..., K_m - K_{m-1} \geq h, T_{i+m+1} - K_m \geq [T\varepsilon]\},
\]

where \(\varepsilon > 0\) is an arbitrarily small number and \(h\) is at least as large as the maximum number of parameters to be estimated within a regime. What makes this estimation problem different from those analyzed previously is that the search over candidate break dates does not impose a partition such that the break dates are separated by a positive fraction of the sample size (e.g., Qu and Perron, 2007).

4 Limiting Distribution of the Estimates

We start with the following result about the consistency and rate of convergence of the estimates of the break dates.

**Theorem 1** Let \(\{\hat{K}_j, j = 1, ..., m; \hat{\beta}_{(j)} \text{ and } \hat{\Sigma}_{(j)}, j = 1, ..., m + 1\}\) be defined as the solution to the maximization problem (8). Then under Assumptions A1-A9, \(v_T^2(\hat{K}_j - K^0_j) = O_p(1)\), \(\sqrt{T}(\hat{\beta}_{(j)} - \beta^0_{(j)}) = O_p(1)\) and \(\sqrt{T}(\hat{\Sigma}_{(j)} - \Sigma^0_{(j)}) = O_p(1)\).

Theorem 1 shows the rate of convergence of the estimates of the break dates are the same as in the case with asymptotically distinct break fractions and the maximization problem being taken over a partition defined by asymptotically distinct break fractions. This is of interest in its own right and, in particular, generalizes the result of Qu and Perron (2007) to the multiple breaks case with integrated regressors and/or trends as well as stationary regressors. It allows us to analyze the asymptotic distribution of the estimates of the break dates in the following compact neighborhood of the true value:

\[
C_M = \{(K, \beta, \Sigma) : v_T^2|K_j - K^0_j| \leq M, \sqrt{T}|(\beta_{(j)} - \beta^0_{(j)})| \leq M, \sqrt{T}|(\Sigma_{(j)} - \Sigma^0_{(j)})| \leq M\}.
\]
Since we can choose $M$ large enough, the estimates will be in this set with probability arbitrarily close to 1.

Before proceeding, it is important to discuss the possible ordering of the estimates of the breaks dates relative to the true values. From the definition of locally ordered breaks, the true break date satisfies $v_T^2(K_{j}^0 - K_{j-1}^0) \leq M_T$ with $M_T \to 0$ as $T \to \infty$, so that $(K_{j}^0 - K_{j-1}^0) = o_p(v_T^{-2})$. It also follows that $(K_{m}^0 - K_{1}^0) = o_p(v_T^{-2})$. On the other hand, from Theorem 1, $v_T^2(\hat{K}_j - K_{j}^0) = O_p(1)$. Hence, in large samples, with probability arbitrarily close to one, the values of all the true break dates will either: 1) occur before any of the estimates; 2) occur after any of the estimates; or 3) all occur between two estimates. In other words, since the true locally ordered break dates are “closer to each other” than the estimates, each increasing as $T$ increases, there cannot be an overlap between the estimates and the true values of the break dates. Hence, we have the following three cases:

- Case 1: $\hat{K}_1 < \ldots < \hat{K}_m \leq K_1^0 < \ldots < K_m^0$;
- Case 2: $K_1^0 < \ldots < K_m^0 \leq \hat{K}_1 < \ldots < \hat{K}_m$;
- Case 3: for some $1 \leq b \leq m$, $\hat{K}_1 < \ldots < \hat{K}_b \leq K_1^0 < K_2^0 < \ldots < K_m^0 \leq \hat{K}_{b+1} < \ldots < \hat{K}_m$.

The relevant limits to be derived will need to allow for these three different scenarios. Following Qu and Perron (2007), the next step is to decompose the likelihood function in two components: 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. The relevant result is stated in the following theorem for the case with only changes in the regression parameters $\beta^0$. The case with changes in both $\beta^0$ and the covariance matrix of the errors $\Sigma^0$ is more involved and relegated to the Supplementary Material available online. Note that given the possible cases for the positions of the estimates relative to the true break dates, in the maximization problem, we need only consider candidate break dates $(K_1, \ldots, K_m)$ that satisfy a similar ordering, namely: Case 1: $K_1 < \ldots < K_m \leq K_1^0 < \ldots < K_m^0$; Case 2: $K_1^0 < \ldots < K_m^0 \leq K_1 < \ldots < K_m$; Case 3: $K_1 < \ldots < K_b \leq K_1^0 < \ldots < K_m^0 \leq K_{b+1} < \ldots < K_m$.

**Theorem 2** Under Assumptions A1-A9, with $m$ breaks in $\beta^0$ only, we have

\[
\max_{(K, \Sigma)} rlr_T = \max_{(\beta, \Sigma) \in C_M, K^0} [rlr_T^1(K, \beta, \Sigma) + \lambda'(\beta, vec(\Sigma))] + \max_{K \in C_M(\beta^0, \Sigma^0)} rlr_T^2(K, \beta^0, \Sigma^0) + o_p(1)
\]
where
\[
rlr_T^1(K^0, \beta, \Sigma) = -(1/2)\left[ \sum_{j=1}^{m+1} \sum_{t=K^0_j+1}^{K^0_j} (Y_t - X'_t(\beta(j)))' \left( \Sigma^{-1}(Y_t - X'_t(\beta(j))) \right) \right]
- \sum_{j=1}^{m+1} \sum_{t=K^0_j+1}^{K^0_j} (Y_t - X'_t(\beta(j)))' \left( \Sigma^{-1}(Y_t - X'_t(\beta(j))) \right)
\]
and for Case 1 ($K_1 < \ldots < K_m \leq K^0_1 < \ldots < K^0_m$):
\[
rlr_T^2(K_1, \ldots, K_m, \beta^0, \Sigma^0) = -(1/2)\sum_{j=1}^{m} (\beta^0_{j+1} - \beta^0_j)' \left( \sum_{t=K^0_j+1}^{K^0_j} X_T(t)^{-1} X'_t(\beta(j)) \right) (\beta^0_{j+1} - \beta^0_j)
- \sum_{j=1}^{m} \sum_{i=j+1}^{m} (\beta^0_{j+1} - \beta^0_j)' \left( \sum_{t=K^0_i+1}^{K^0_j} X_T(t)^{-1} X'_t(\beta(j)) \right) (\beta^0_{i+1} - \beta^0_i)
+ \sum_{j=1}^{m} (\beta^0_{j+1} - \beta^0_j)' \left( \sum_{t=K^0_j+1}^{K^0_j} X_T(t)^{-1} u_t \right) + o_p(1)
\]
For Case 2 ($K^0_1 < \ldots < K^0_m \leq K_1 < \ldots < K_m$):
\[
rlr_T^2(K_1, \ldots, K_m, \beta^0, \Sigma^0) = -(1/2)\sum_{j=1}^{m} (\beta^0_{j+1} - \beta^0_j)' \left( \sum_{t=K^0_j+1}^{K^0_j} X_T(t)^{-1} X'_t(\beta(j)) \right) (\beta^0_{j+1} - \beta^0_j)
- \sum_{j=1}^{m} \sum_{i=j+1}^{m} (\beta^0_{j+1} - \beta^0_j)' \left( \sum_{t=K^0_i+1}^{K^0_j} X_T(t)^{-1} X'_t(\beta(j)) \right) (\beta^0_{i+1} - \beta^0_i)
- \sum_{j=1}^{m} (\beta^0_{j+1} - \beta^0_j)' \left( \sum_{t=K^0_j+1}^{K^0_j} X_T(t)^{-1} u_t \right) + o_p(1)
\]
and for Case 3 ($K_1 < \ldots < K_b \leq K^0_1 < \ldots < K^0_m \leq K_{b+1} < \ldots < K_m$, with $1 \leq b \leq m$):
\[
rlr_T^2(K_1, \ldots, K_m, \beta^0, \Sigma^0) = -(1/2)\sum_{j=b+1}^{m} (\beta^0_{j+1} - \beta^0_j)' \left( \sum_{t=K^0_j+1}^{K^0_j} X_T(t)^{-1} X'_t(\beta(j)) \right) (\beta^0_{j+1} - \beta^0_j)
- \sum_{j=b+1}^{m} \sum_{i=j+1}^{m} (\beta^0_{j+1} - \beta^0_j)' \left( \sum_{t=K^0_i+1}^{K^0_j} X_T(t)^{-1} X'_t(\beta(j)) \right) (\beta^0_{i+1} - \beta^0_i)
- \sum_{j=b+1}^{m} (\beta^0_{j+1} - \beta^0_j)' \left( \sum_{t=K^0_j+1}^{K^0_j} X_T(t)^{-1} u_t \right) + o_p(1)
\]
This result has strong implications. In particular, it implies that to analyze the asymptotic distribution of the estimates of the break dates, we need only consider the component $rlr_T^2$. 

12
Remark 1 In the leading case of interest with only one pair of locally ordered breaks in
coefficients only, the various components in Theorem 2 reduce to:

\[
rlr_T^{1}(K^0, \beta, \Sigma) = -(1/2)[\sum_{j=1}^{3} \sum_{t=K^0_{j-1}+1}^{K^0_j} (Y_t - X'_T \beta_{(j)}') (\Sigma)^{-1} (Y_t - X'_T \beta_{(j)})]
\]

and for Case 1 \((K_1 = K_2 \leq K^1_0 < K^2_0)\):

\[
rlr_T^{2}(K_1, K_2, \beta^0, \Sigma^0) = -(1/2)(\beta^0_{(2)} - \beta^0_{(1)})' (\sum_{t=K^0_{1}+1}^{K^0_{2}} X_T (\Sigma^0)^{-1} X'_T) (\beta^0_{(2)} - \beta^0_{(1)})
\]

and for Case 2 \((K^1_0 < K^2_0 \leq K_1 = K_2)\):

\[
rlr_T^{2}(K_1, K_2, \beta^0, \Sigma^0) = -(1/2)(\beta^0_{(2)} - \beta^0_{(1)})' (\sum_{t=K^0_{1}+1}^{K^0_{2}} X_T (\Sigma^0)^{-1} X'_T) (\beta^0_{(2)} - \beta^0_{(1)})
\]

while for Case 3 \((K_1 = K^1_0 < K^2_0 \leq K_2)\):

\[
rlr_T^{2}(K_1, K_2, \beta^0, \Sigma^0) = -(1/2)(\beta^0_{(2)} - \beta^0_{(1)})' (\sum_{t=K^0_{1}+1}^{K^0_{2}} X_T (\Sigma^0)^{-1} X'_T) (\beta^0_{(2)} - \beta^0_{(1)})
\]

Lemma 1 Under Assumptions A1-A9, we have

\[
\psi_T^2 \sum_{t=K^0+1}^{K^0+[s_j \psi_T^2]} X_T (\Sigma^0)^{-1} X'_T \Rightarrow S'(s_j D(\lambda^0_0) \otimes (\Sigma^0)^{-1}) S \quad s_j > 0
\]

\[
\psi_T^2 \sum_{t=K^0+1}^{K^0+[s_j \psi_T^2]} X_T (\Sigma^0)^{-1} X'_T \Rightarrow S'(|s_j| D(\lambda^0_0) \otimes (\Sigma^0)^{-1}) S \quad s_j < 0
\]
\[
v_T \sum_{t=K_j^0+1}^{K_j^0+[s_jv_t^{-2}]} X_T(t)(\Sigma_t^0)^{-1}U_t \Rightarrow S'(I_q \otimes (\Sigma_t^0)^{-1}(\Sigma_t^0)^{1/2})\Omega(\lambda_j^0)W_j(s_j) \quad s_j > 0
\]
\[
v_T \sum_{t=K_j^0+[s_jv_t^{-2}]}^{K_j^0} X_T(t)(\Sigma_t^0)^{-1}U_t \Rightarrow S'(I_q \otimes (\Sigma_t^0)^{-1}(\Sigma_t^0)^{1/2})\Omega(\lambda_j^0)W_j(s_j) \quad s_j < 0
\]

with
\[
D(\lambda_j^0) = \begin{pmatrix}
\Omega_z^{1/2}W_z(\lambda_j^0)&\lambda_j^0\Omega_z W_z(\lambda_j^0) \\
\lambda_j^0W_z(\lambda_j^0)\Omega_z^{1/2} & (\lambda_j^0)^2 \\
\mu_{x,j}W_z(\lambda_j^0)\Omega_z^{1/2} & \lambda_j^0\mu_{x,j} \\
\end{pmatrix}
\]
\[
\Omega(\lambda_j^0) = \begin{pmatrix}
\Omega_z^{1/2}W_z(\lambda_j^0)\otimes(\Sigma_j^0)^{-1/2}(\Omega_{\eta(j)})^{1/2} & 0 \\
\lambda_j^0(\Sigma_j^0)^{-1/2}(\Omega_{\eta(j)})^{1/2} & 0 \\
0 & M_{x_{\eta,j}}^{1/2}
\end{pmatrix}, \quad W_j(s_j) = \begin{pmatrix}
W_{x,\eta,j}(s_j) \\
W_{x,\eta,j}(s_j)
\end{pmatrix}
\]

where \(M_{x_{\eta,j}} = (\Sigma^0)^{-1/2}M_{x,j}(\Sigma^0)^{-1/2}, W_{x,\eta,j}(\cdot) = (\Sigma^0)^{-1/2}W_{x,j}(\cdot)\) and \(Q_{x,j}\) as defined in (7).

**Remark 2** Without structural breaks in \(\Sigma_0\), Lemma 1 reduces to
\[
v_T^2 \sum_{t=K_j^0+1}^{K_j^0+[s_jv_t^{-2}]} X_T(t)(\Sigma_t^0)^{-1}X_T' \Rightarrow S'(I_q \otimes (\Sigma_t^0)^{-1})S \quad s_j > 0
\]
\[
v_T^2 \sum_{t=K_j^0+[s_jv_t^{-2}]}^{K_j^0} X_T(t)(\Sigma_t^0)^{-1}X_T' \Rightarrow S'(I_q \otimes (\Sigma_t^0)^{-1})S \quad s_j < 0
\]

We are now in a position to state the main result of this paper about the joint limit distribution of the locally ordered breaks. Again, we present in the text the case with only changes in the regression coefficients \(\beta^0\). The more general case with also changes in \(\Sigma^0\) is presented in the Supplementary Material available online.

**Theorem 3** Let \(\delta_{\eta(j)} = \beta_{\eta(j+1)}^0 - \beta_{\eta(j)}^0\), \(Q(\lambda_j^0) = S'(D(\lambda_j^0) \otimes (\Sigma^0)^{-1})S\),
\[
\Omega(\lambda_j^0) = \begin{pmatrix}
\Omega_z^{1/2}W_z(\lambda_j^0)\otimes(\Sigma_j^0)^{-1/2}(\Omega_{\eta})^{1/2} & 0 \\
\lambda_j^0(\Sigma_j^0)^{-1/2}(\Omega_{\eta})^{1/2} & 0 \\
0 & M_{x_{\eta,j}}^{1/2}
\end{pmatrix}
\]
For the case with only one pair of locally ordered breaks, Theorem 3 reduces to:

\[ H(v_1, v_2) = -\frac{1}{2} |v_1| + \left( \frac{\gamma_1}{\Pi_1} \right)^{1/2} B_1(v_1) - \frac{1}{2} |v_2| \frac{\Pi_2}{\Pi_1} + \left( \frac{\gamma_2}{\Pi_1} \right)^{1/2} B_2(v_2) - |v_2| \frac{\Pi_2^2}{\Pi_1}, \]

for Case 1 with \( v_1 \leq v_2 \leq 0 \):

where \( H(v_1, v_2) = 0 \) if \( v_1 = v_2 = 0 \), and for Case 2 with \( 0 \leq v_1 \leq v_2 \):

\[ H(v_1, v_2) = -\frac{1}{2} |v_1| - \left( \frac{\gamma_1}{\Pi_1} \right)^{1/2} B_1(v_1) - \frac{1}{2} |v_2| \frac{\Pi_2}{\Pi_1} - \left( \frac{\gamma_2}{\Pi_1} \right)^{1/2} B_2(v_2) - |v_1| \frac{\Pi_2^2}{\Pi_1}, \]

for Case 3 with \( v_1 < 0, v_2 > 0 \):

\[ H(v_1, v_2) = -\frac{1}{2} |v_1| + \left( \frac{\gamma_1}{\Pi_1} \right)^{1/2} B_1(v_1) - \frac{1}{2} |v_2| \frac{\Pi_2}{\Pi_1} - \left( \frac{\gamma_2}{\Pi_1} \right)^{1/2} B_2(v_2), \]
The cumulative distribution function does not have a tractable analytical formula. However, the relevant quantiles can be obtained using simulations. First, generate the realizations of $H$ by replacing the true value of the parameters with consistent estimates and simulating the Brownian motions over a reasonable range, i.e. $[-M, M]$. Then, apply a dynamic programming algorithm to find the global maximizers of $H$ over $(v_1, v_2, ..., v_m) \in [-M, M]$. This is repeated for all possible cases and the overall maximum obtained. These steps are repeated to obtain the relevant quantiles.

**Remark 4** For the special cases with a single type of regressors and martingale difference errors $u_t$ so that $\Sigma^0 = \Omega_j$, the limit distribution differs according to the specifications of the matrices $D(\lambda^0_j)$ and $\Omega(\lambda^0_j)$, and the definitions of $Q(\lambda^0_j)$, $\Delta(\lambda^0_j)$, $\Pi_j$, $\Pi^1_j$, $\Pi^2_j$, and $\Upsilon_j$. When only stationary regressors are present, we have:

$$D(\lambda^0_j) = Q_{x,j}, \Omega(\lambda^0_j) = M^{1/2}_{x\eta,j}.$$  

On the other hand, with only a trend as the regressor:

$$D(\lambda^0_j) = (\lambda^0_j)^2, \Omega(\lambda^0_j) = \lambda^0_j.$$  

With only integrated regressors,

$$D(\lambda^0_j) = \Omega_z^{1/2}W_z(\lambda^0_j)\Omega_z^{1/2}, \quad \Omega(\lambda^0_j) = \Omega_z^{1/2}W_z(\lambda^0_j) \otimes I_n.$$  

If only two types of regressors are involved, we have: a) with stationary and unit root regressors,

$$D(\lambda^0_j) = \begin{pmatrix} \Omega_z^{1/2}W_z(\lambda^0_j)\Omega_z^{1/2} & \Omega_z^{1/2}W_z(\lambda^0_j)\mu^\prime_{x,j} \\ \mu_{x,j}W_z(\lambda^0_j)\Omega_z^{1/2} & Q_{x,j} \end{pmatrix}, \quad \Omega(\lambda^0_j) = \begin{pmatrix} \Omega_z^{1/2}W_z(\lambda^0_j) \otimes I_n & 0 \\ 0 & M^{1/2}_{x\eta,j} \end{pmatrix}.$$  

b) with stationary and trend regressors,

$$D(\lambda^0_j) = \begin{pmatrix} (\lambda^0_j)^2 & \lambda^0_j\mu^\prime_{x,j} \\ \lambda^0_j\mu_{x,j} & Q_{x,j} \end{pmatrix}, \quad \Omega(\lambda^0_j) = \begin{pmatrix} \lambda^0_jI_n & 0 \\ 0 & M^{1/2}_{x\eta,j} \end{pmatrix}.$$  

c) with a trend and unit root regressors,

$$D(\lambda^0_j) = \begin{pmatrix} \Omega_z^{1/2}W_z(\lambda^0_j)\Omega_z^{1/2} & \lambda^0_j\Omega_z^{1/2}W_z(\lambda^0_j) \\ \lambda^0_jW_z(\lambda^0_j)\Omega_z^{1/2} & (\lambda^0_j)^2 \end{pmatrix}, \quad \Omega(\lambda^0_j) = \begin{pmatrix} \Omega_z^{1/2}W_z(\lambda^0_j) \otimes I_n \\ \lambda^0_jI_n \end{pmatrix}.$$  

16
The results allow constructing joint confidence intervals for the estimates of the break dates, all we need is to have consistent estimates of the various parameters. From Theorem 1, quasi-maximum likelihood estimation will provide consistent estimates of the break fractions, even though the estimates of break dates are not consistent per se. The method will also deliver consistent estimates of the coefficients $\beta_{(j)}^0$ and the variance-covariance matrix $\Sigma^0$. The long-run covariance matrices $M_{x,j}$ and $\Omega_\eta$ can be estimated using kernel-based methods, as suggested by Andrews (1991). Also, with homogeneity across segments: $\hat{Q}_x = T^{-1} \sum_{t=1}^{T} x_t x'_t$, $\hat{\mu}_x = T^{-1} \sum_{t=1}^{T} x_t$, $\hat{\Sigma} = T^{-1} \sum_{t=1}^{T} \hat{u}_t \hat{u}'_t$ and $\delta \beta(j) = \hat{\beta}(j+1) - \hat{\beta}(j)$.

5 Monte Carlo Simulations

We now provide simulation evidence to address the following three issues: 1) the exact coverage rate under a variety of Data Generating Processes (DGP); 2) the effect of serially correlation across equations on the length of the confidence intervals; 3) practical guidelines regarding when to use a locally ordered breaks framework as opposed to one that assumes globally disjoint breaks.

5.1 The exact coverage rates for various DGPs.

We first provide Monte Carlo simulation results to examine the accuracy of the asymptotic distribution to the corresponding finite sample distribution. We conducted several simulation experiments with different combinations of regressors. For all cases there is only one pair of locally ordered breaks and two equations. Throughout $T = 200$ and 500 replications are used. The data generating processes considered are the following, which cover all possible combinations of the types of regressors:

<table>
<thead>
<tr>
<th></th>
<th>Equation 1</th>
<th>Equation 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGP-1:</td>
<td>$y_{1t} = \beta_{11}^{(j)} x_t + u_{1t}$</td>
<td>$y_{2t} = \beta_{21}^{(j)} x_t + u_{2t}$</td>
</tr>
<tr>
<td>DGP-2:</td>
<td>$y_{1t} = \beta_{11}^{(j)} \frac{t}{T} + u_{1t}$</td>
<td>$y_{2t} = \beta_{21}^{(j)} \frac{t}{T} + u_{2t}$</td>
</tr>
<tr>
<td>DGP-3:</td>
<td>$y_{1t} = \beta_{11}^{(j)} \frac{z_t}{\sqrt{T}} + u_{1t}$</td>
<td>$y_{2t} = \beta_{21}^{(j)} \frac{z_t}{\sqrt{T}} + u_{2t}$</td>
</tr>
<tr>
<td>DGP-4:</td>
<td>$y_{1t} = \beta_{11}^{(j)} x_t + \beta_{12}^{(j)} \frac{t}{T} + u_{1t}$</td>
<td>$y_{2t} = \beta_{21}^{(j)} x_t + \beta_{22}^{(j)} \frac{t}{T} + u_{2t}$</td>
</tr>
<tr>
<td>DGP-5:</td>
<td>$y_{1t} = \beta_{11}^{(j)} x_t + \beta_{12}^{(j)} \frac{z_t}{\sqrt{T}} + u_{1t}$</td>
<td>$y_{2t} = \beta_{21}^{(j)} x_t + \beta_{22}^{(j)} \frac{z_t}{\sqrt{T}} + u_{2t}$</td>
</tr>
<tr>
<td>DGP-6:</td>
<td>$y_{1t} = \beta_{11}^{(j)} \frac{t}{T} + \beta_{12}^{(j)} \frac{z_t}{\sqrt{T}} + u_{1t}$</td>
<td>$y_{2t} = \beta_{21}^{(j)} \frac{t}{T} + \beta_{22}^{(j)} \frac{z_t}{\sqrt{T}} + u_{2t}$</td>
</tr>
<tr>
<td>DGP-7:</td>
<td>$y_{1t} = \beta_{11}^{(j)} \frac{z_t}{\sqrt{T}} + \beta_{12}^{(j)} \frac{t}{T} + \beta_{13}^{(j)} x_t + u_{1t}$</td>
<td>$y_{2t} = \beta_{21}^{(j)} \frac{z_t}{\sqrt{T}} + \beta_{22}^{(j)} \frac{t}{T} + \beta_{23}^{(j)} x_t + u_{2t}$</td>
</tr>
</tbody>
</table>
for $j = 1, 2$. The values of the coefficients and break dates used are as follows, chosen more or less arbitrarily to obtain non-degenerate distributions that are neither too tight nor too dispersed:

\begin{align*}
\beta_{11}^{(1)} &= 1, & \beta_{11}^{(2)} &= 1.5, & \beta_{12}^{(1)} &= 1, & \beta_{12}^{(2)} &= 1.6, & \beta_{13}^{(1)} &= 1, & \beta_{13}^{(2)} &= 1.5, \\
\beta_{21}^{(1)} &= 1, & \beta_{21}^{(2)} &= 2.3, & \beta_{22}^{(1)} &= 1, & \beta_{22}^{(2)} &= 2.3, & \beta_{23}^{(1)} &= 1, & \beta_{23}^{(2)} &= 2.3. \\
K_0 &= 80, & K_1 &= 95.
\end{align*}

DGP-1: $\begin{bmatrix} 1 & 1.5 & 1 & 1.6 & 80 & 95 \end{bmatrix}$

DGP-2: $\begin{bmatrix} 1 & 1.7 & 1.8 & 80 & 95 \end{bmatrix}$

DGP-3: $\begin{bmatrix} 1 & 2.0 & 2.3 & 80 & 86 \end{bmatrix}$

DGP-4: $\begin{bmatrix} 1 & 1.5 & 2 & 3 & 80 & 95 \end{bmatrix}$

DGP-5: $\begin{bmatrix} 2 & 2.5 & 3 & 3.7 & 80 & 95 \end{bmatrix}$

DGP-6: $\begin{bmatrix} 1 & 3 & 1.5 & 80 & 95 \end{bmatrix}$

DGP-7: $\begin{bmatrix} 1 & 1.5 & 3 & 3.3 & 80 & 95 \end{bmatrix}$

We set $u_t = (u_{1t}, u_{2t})' \sim i.i.d. N(0, \Sigma)$ with $\text{var}(u_{1t}) = \text{var}(u_{2t}) = 1$ and $E(u_{1t}u_{2t}) = \rho$. The stationary regressor $x_t$ is specified as the sum of a constant term $\mu = 2$ and a normal error $\epsilon_t \sim N(0, 1)$. The $I(1)$ regressors are the partial sums of $i.i.d. N(0, 1)$ random variables. We consider two cases with $\rho = 0, 0.3$. We first consider the exact sizes of the joint confidence intervals constructed using the asymptotic distributions, for nominal sizes of 90% and 95%. The results are presented in Table 1. They show that, in general, the finite sample coverages are close to the nominal ones.

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>Nominal Size</th>
<th>DGP1</th>
<th>DGP2</th>
<th>DGP3</th>
<th>DGP4</th>
<th>DGP5</th>
<th>DGP6</th>
<th>DGP7</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho = 0$</td>
<td>.90</td>
<td>.88</td>
<td>.92</td>
<td>.93</td>
<td>.85</td>
<td>.88</td>
<td>.92</td>
<td>.87</td>
</tr>
<tr>
<td>$\rho = 0$</td>
<td>.95</td>
<td>.92</td>
<td>.97</td>
<td>.95</td>
<td>.91</td>
<td>.93</td>
<td>.94</td>
<td>.92</td>
</tr>
<tr>
<td>$\rho = 0.3$</td>
<td>.90</td>
<td>.88</td>
<td>.92</td>
<td>.92</td>
<td>.87</td>
<td>.90</td>
<td>.91</td>
<td>.89</td>
</tr>
<tr>
<td>$\rho = 0.3$</td>
<td>.95</td>
<td>.95</td>
<td>1.0</td>
<td>.93</td>
<td>.93</td>
<td>.94</td>
<td>.93</td>
<td>.93</td>
</tr>
</tbody>
</table>

We now assess the benefits of constructing confidence intervals in a joint fashion using the locally ordered break framework relative to using the method of Bai (1997) one equation at a time. To that effect, we adopt the specifications of DGP-1 with $\rho = 0$. The results are presented in Table 2. They show indeed that the length of the confidence intervals are smaller using the locally ordered joint approach, especially for the first break date.
### 5.2 The effect of serial correlation on the length of the confidence intervals

The data generating process is

\[
y_{1t} = \beta_1 x_t + u_{1t} \\
y_{2t} = \beta_2 x_t + u_{2t}
\]

with \( \beta_1 = 1 \) for \( 1 \leq t \leq K_1^0 \), \( \beta_1 = 1.4, \beta_2 = 1 \) for \( K_1^0 + 1 \leq t \leq K_2^0 \) and \( \beta_1 = 1.4, \beta_2 = 1 \) for \( K_2^0 + 1 \leq t \leq T \). We set \( T = 200, K_1^0 = 80 \) and \( K_2^0 = 100 \). The regressor is generated by \( x_t = \mu_x + u_{xt} \) with \( \mu_x = 2 \) and \( u_{xt} \sim N(0, 1) \). We set \( u_t = (u_{1t}, u_{2t})' \sim i.i.d \ N(0, \Sigma) \) with \( \text{var}(u_{1t}) = \text{var}(u_{2t}) = 1 \) and \( E(u_{1t}u_{2t}) = \rho \). The number of simulations is 200. Table 3 presents the average length of the confidence intervals for both break dates for nominal coverage rates of 90% and 95% with \( \rho = 0, 0.3, 0.5 \) and 0.8. The results show that, as expected, the length of the confidence intervals decrease noticeably as \( \rho \) increases, especially for the second break date.

### Table 3: Length of the Confidence Intervals as a Function of \( \rho \)

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>90% CI ( K_1 )</th>
<th>90% CI ( K_2 )</th>
<th>95% CI ( K_1 )</th>
<th>95% CI ( K_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>[68, 91]</td>
<td>[70, 91]</td>
<td>[73, 88]</td>
<td>[74, 85]</td>
</tr>
<tr>
<td>0.3</td>
<td>[92, 107]</td>
<td>[93, 106]</td>
<td>[94, 105]</td>
<td>[95, 104]</td>
</tr>
<tr>
<td>0.5</td>
<td>[64, 96]</td>
<td>[64, 95]</td>
<td>[67, 94]</td>
<td>[77, 84]</td>
</tr>
<tr>
<td>0.8</td>
<td>[90, 110]</td>
<td>[90, 108]</td>
<td>[91, 107]</td>
<td>[97, 102]</td>
</tr>
</tbody>
</table>

### 5.3 Locally ordered versus globally distinct

We now present some results that can yield some guidelines as to when to use a locally ordered break framework as opposed to the globally distinct framework of Bai (1997). The DGP used is the same as in Section 5.2 with \( \rho = 0 \). The difference is that we consider two sample sizes and combinations of break dates that are separated by 5%, 10% and 15% of the total sample. Hence, with \( T = 200 \), we set: a) \( K_1^0 = 80, K_2^0 = 90 \) (5%), b) \( K_1^0 = 80, K_2^0 = 100 \) (10%).
(10%) and c) $K_1^0 = 80, K_2^0 = 110$ (15%). With $T = 400$, wet set: a) $K_1^0 = 160, K_2^0 = 180$ (5%), b) $K_1^0 = 160, K_2^0 = 200$ (10%) and c) $K_1^0 = 160, K_2^0 = 220$ (15%). The results show that the exact size is closer to the nominal size for the locally ordered method when the break dates are separated by 5% of the sample. The reverse holds when the break dates are separated by 15% of the sample, while when they are separated by 10% of the sample both methods are (roughly) equally good. Hence, the recommendation is to use the locally ordered break framework when the break dates are separated by approximately 10% or less of the sample.

<table>
<thead>
<tr>
<th>$T = 200$</th>
<th>Nominal Size</th>
<th>$K_1^0 = 80, K_2^0 = 90$</th>
<th>$K_1^0 = 80, K_2^0 = 100$</th>
<th>$K_1^0 = 80, K_2^0 = 110$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Method</td>
<td>0.95</td>
<td>1</td>
<td>0.955</td>
<td>0.945</td>
</tr>
<tr>
<td>Locally Ordered</td>
<td>0.95</td>
<td>0.965</td>
<td>0.965</td>
<td>0.935</td>
</tr>
<tr>
<td>Standard Method</td>
<td>0.90</td>
<td>0.95</td>
<td>0.915</td>
<td>0.88</td>
</tr>
<tr>
<td>Locally Ordered</td>
<td>0.90</td>
<td>0.93</td>
<td>0.895</td>
<td>0.885</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$T = 400$</th>
<th>$K_1^0 = 160, K_2^0 = 180$</th>
<th>$K_1^0 = 160, K_2^0 = 200$</th>
<th>$K_1^0 = 160, K_2^0 = 220$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Method</td>
<td>0.950</td>
<td>0.985</td>
<td>0.940</td>
</tr>
<tr>
<td>Locally Ordered</td>
<td>0.950</td>
<td>0.955</td>
<td>0.955</td>
</tr>
<tr>
<td>Standard Method</td>
<td>0.900</td>
<td>0.910</td>
<td>0.900</td>
</tr>
<tr>
<td>Locally Ordered</td>
<td>0.900</td>
<td>0.910</td>
<td>0.925</td>
</tr>
</tbody>
</table>

6 Conclusion

We studied the problem of multiple locally ordered breaks occurring at unknown dates in a multiple regression system with integrated, trends and stationary regressors. We analyzed cases with shifts in both the coefficients and the covariance matrix of the errors. Theoretical results concerning the consistency, rate of convergence and asymptotic distributions of the break dates were obtained. Through simulations, we showed that the asymptotic distribution derived provides good approximations in finite samples and allow more precise inference compared to single equation methods treating each break date in isolation. Our results are of practical interest given that breaks across equations can often be treated as ordered in applications.
References


Appendix

They key ingredients in the proofs are a generalized Hajek-Renyi type 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) applicable under the stated assumptions. We first state a few lemmas due to Qu and Perron (2007) and Oka and Perron (2011).

Lemma A.1 Let \((\zeta_i)_{i \geq 1}\) be a sequence of mean zero \(R^d\)-valued random vectors satisfying A4. Define \(S_k(l) = \sum_{i=1}^{l+k} \zeta_i\), then, (a) (SAT) the covariance of \(k^{-1/2}S_k(l)\), \(\Omega_k\) converge, with the limit denoted by \(\Omega\), and there exists a Brownian Motion \((W(t))_{t \geq 0}\) with covariance matrix \(\Omega\) such that \(\sum_{i=1}^{l} \zeta_i - W(t) = O_{a.s.}(t^{1/2-\kappa})\) for some \(\kappa > 0\); (b) (FCLT) \(T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \zeta_t \Rightarrow \Omega^{1/2}W^*(r)\), where \(W^*(r)\) is a \(d\)-vector of independent Wiener processes and \(\Rightarrow\) denotes weak convergence under the Skorohod topology; (c) (SLLN) \(k^{-1} \sum_{i=1}^{k} \zeta_i \Rightarrow 0\) as \(k \to \infty\); (d) (LIL)

\[
\lim \sup_{k \to \infty} (k^{1/2} \log^{1/2} k)^{-1} ||\Omega_k^{-1/2} \sum_{i=1}^{k} \zeta_i|| = O_p(1).
\]

Lemma A.2 Let \(\eta_t = (\Sigma^0_{(j)})^{-1/2} u_t\) for \(K_{j-1}^0 + 1 \leq t \leq K_j^0\). Under Assumption A4, with \(v_T\) a sequence of positive numbers satisfying \(v_T \to 0\) and \(T^{1/2}v_T/(\log^2 T) \to \infty\), we have

for \(s < 0\), \(v_T \sum_{t=K_j^0+[sv_T^2]}^{K_j^0} (\eta_t \eta'_t - I_n) \Rightarrow \xi_j^{(1)}(s)\)

for \(s > 0\), \(v_T \sum_{t=K_j^0+[sv_T^2]}^{K_j^0+1} (\eta_t \eta'_t - I_n) \Rightarrow \xi_j^{(2)}(s)\)

where the weak convergence is in the space \(D[0,\infty)^n\) and where the entries of the \(n \times n\) matrices \(\xi_j^{(1)}(s)\) and \(\xi_j^{(2)}(s)\) are Brownian processes defined on the real line with covariance matrix

\[\Omega_{x(j)} = \lim_{T \to \infty} \text{var}\{(\Delta K_j^0)^{-1/2} \sum_{t=K_j^0}^{K_j} (\eta_t \eta'_t - I_n)\}.
\]

The two-sided Brownian motion \(\xi_j(\cdot)\) satisfies \(\xi_j(s) = \xi_j^{(1)}(s)\) for \(s < 0\), \(\xi_j(s) = 0\) for \(s = 0\), and \(\xi_j(s) = \xi_j^{(2)}(s)\) for \(s > 0\). Furthermore,

for \(s < 0\), \(v_T \sum_{t=K_j^0+[sv_T^2]}^{K_j^0} x_t \otimes u_t \Rightarrow M_{x,j}^{1/2}W_{x,j}^{(1)}(s)\)

for \(s > 0\), \(v_T \sum_{t=K_j^0+[sv_T^2]}^{K_j^0+1} x_t \otimes u_t \Rightarrow M_{x,j}^{1/2}W_{x,j}^{(2)}(s)\)

where the weak convergence is in the space \(D[0,\infty)^p\) and where the entries of the \(p\) vectors \(W_{x,j}^{(1)}(s)\) and \(W_{x,j}^{(2)}(s)\) are independent Wiener processes defined on the real line. Let \(W_{x,j}(\cdot)\) denote the two-sided Brownian motion such that \(W_{x,j}(s) = W_{x,j}^{(1)}(s)\) for \(s < 0\), \(W_{x,j}(s) = 0\) for \(s = 0\) and \(W_{x,j}(s) = W_{x,j}^{(2)}(s)\) for \(s > 0\). Also \(W_{x,j}^{(1)}(s)\) and \(W_{x,j}^{(2)}(s)\) (resp., \(\xi_j^{(1)}(s)\) and \(\xi_j^{(2)}(s)\)) are different independent copies for \(j = 1, \ldots, m\).
Lemma A.3 Let \((\zeta_t)_{t \geq 1}\) be a sequence of mean zero \(\mathbb{R}^d\)-valued random vectors satisfying \(A_4\), and let \(\zeta_t^r\) stands for \(\zeta_t\) or \((t/T)\zeta_t\). Then there exists \(L < \infty\) such that for any \(\varepsilon > 0\) and \(m > 0\):

\[
P \left( \sup_{m < k \leq n} \frac{1}{k} \sum_{t=1}^{k} \| \zeta_t^r \| > \varepsilon \right) \leq L e^{-2 \varepsilon^2 m^2 + \sum_{t=m+1}^{k} d_t^2}
\]

Lemma A.4 Under \(A_4\), we have uniformly over all \(0 < r < s < 1\):

a) \(\sum_{t=[rT]}^{[sT]} \zeta_t = O_p(T^{1/2})\),

b) \(\sum_{t=[rT]}^{[sT]} z_t = O_p(T^{3/2})\),

c) \(\sum_{t=[rT]}^{[sT]} z_t^r = O_p(T^2)\),

d) \(\sum_{t=[rT]}^{[sT]} z_t^r = O_p(T)\).

Lemma A.5 Let \(T_1 = [aT]\) for some \(a \in (0, 1)\) and \(\zeta_t\) stands for \((h_t \otimes \eta_t)\) or \((\eta_t' \eta_t - I)\). Under Assumptions A1-A4:

a) there exists a positive constant \(k_0\) such that \(\sup_{k_0 \leq k \leq T}(\sqrt{k} \log k)^{-1} \| \sum_{t=T_1}^{T_1+k} \zeta_t \| = o_p(1)\);

b) \(\sup_{ST \leq k \leq T} k^{-1} \| \sum_{t=T_1}^{T_1+k} \zeta_t \| = o_p(vT / \sqrt{\log T})\) for any \(\delta \in (0, 1)\);

c) \(\sup_{\sqrt{T}v \leq k \leq T} k^{-1} \| \sum_{t=T_1}^{T_1+k} \zeta_t \| = o_p(vT / \sqrt{\log T})\);

Proof of Theorem 1: The proof is based on estimating several properties of the sequential likelihood functions with \(k\) observations free from structure change. To that effect let \(T_\rho = [T \rho]\) for some \(0 < \rho < 1\). For a sample involving \(k\) observations from \(T_\rho + 1\) to \(T_\rho + k\), the likelihood ratio is

\[
LR_k(\beta, \Sigma) = \frac{\prod_{t=T_\rho+1}^{T_\rho+k} f(y_t | X_{T_\rho}, \beta, \Sigma)}{\prod_{t=T_\rho+1}^{T_\rho+k} f(y_t | X_{T_\rho}, \beta^0, \Sigma^0)}
\]

where \((\beta, \Sigma)\) are generic values of the parameters and \((\beta^0, \Sigma^0)\) are the true values. Since

\[
\log f(y_t | X_{T_\rho}, \beta, \Sigma) = -(1/2) \log(2\pi) - (1/2) \log |\Sigma| - (1/2) \| \Sigma^{-1/2}(y_t - X_{T_\rho}' \beta) \|^2,
\]

the log-likelihood ratio is

\[
\log LR_k(\beta, \Sigma) = -(k/2) \log \| (\Sigma^0)^{-1/2}(\Sigma^0)^{-1/2} - (1/2) \| \Sigma^{-1/2}(y_t - X_{T_\rho}' \beta) \|^2 + (1/2) \| (\Sigma^0)^{-1/2} u_t \|^2.
\]

The following three properties of the likelihood ratio are proved in Oka and Perron (2011).

Property 1. For any \(\delta \in (0, 1)\) and for \(T\) large enough, \(\sup_{\delta T \leq k \leq T} LR_k(\beta, \Sigma) = O_p(1)\).

Property 2. For any \(\epsilon > 0\), there exists a \(B > 0\) such that \(P(\sup_{1 \leq k \leq T} \log LR_k(\beta, \Sigma) > B(\log T)^2) < \epsilon\) for sufficiently large \(T\).

Property 3. Let \(b_T = dv_T\) for some \(d > 0\). For any \(\epsilon > 0\), there exists \(C > 0\), such that

\[
P \left( \sup_{\sqrt{T}v \leq k \leq T} \sup_{S(\beta, \Sigma) \geq \epsilon} \log LR_k(\beta, \Sigma) > -C \sqrt{T}v_T \right) < \epsilon
\]

where \(S(\beta, \Sigma) = \{(\beta, \Sigma) : ||\beta - \beta^0|| \geq \varepsilon \text{ or } ||\Sigma - \Sigma^0|| \geq \varepsilon\}\).
**Property 4.** Let $T_1 = [Ta]$ for some $a \in (0, 1]$ and let $T_2 = \lfloor \sqrt{T}v_T^{-1} \rfloor$ where $v_T$ satisfies Assumption A6. Also let

$$y_t = x_t'\theta(1) + (\Sigma(1))^{1/2} \eta_t \quad t = 1, ..., T_1$$
$$y_t = x_t'\theta(2) + (\Sigma(2))^{1/2} \eta_t \quad t = T_1 + 1, ..., T_1 + T_2$$

where $||\beta(2) - \beta(1)|| \leq Mv_T$ and $||\Sigma(2) - \Sigma(1)|| \leq Mv_T$ for some $M < \infty$. Let $N = T_1 + T_2$ be the size of the pooled sample and let $(\beta(N), \Sigma(N))$ be the associated estimates. Then $\sqrt{T}(\beta(N) - \beta(1)) = O_p(1)$, $\sqrt{T}(\Sigma(N) - \Sigma(1)) = O_p(1)$.

**Property 5.** With $v_T$ satisfying Assumption A6, for each $\beta$ and $\Sigma$ satisfying $||\beta - \beta^0|| \leq Mv_T$ and $||\Sigma - \Sigma^0|| \leq Mv_T$, with $M < \infty$, we have

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

**Proof:**

\[
\log LR_k(\beta, \Sigma) = -(k/2) \log |\Sigma| - (1/2) \sum_{t=T_{p+1}}^{T_p+k} (y_t - X_t'\beta)'\Sigma^{-1}(y_t - X_t'\beta) + (k/2) \log |\Sigma^0| - (1/2) \sum_{t=T_{p+1}}^{T_p+k} (y_t - X_t'\beta^0)'(\Sigma^0)^{-1}(y_t - X_t'\beta^0)
\]

Let $\Psi_T = (\Sigma^0)^{-1/2}(\Sigma - \Sigma^0)(\Sigma^0)^{-1/2}$ and

\[
L_{1T}(\beta, \Sigma) = -(k/2) \log |I + \Psi_T| - (1/2) \sum_{t=T_{p+1}}^{T_p+k} \eta_t'(I + \Psi_T)^{-1}\eta_t + (1/2) \sum_{t=T_{p+1}}^{T_p+k} \eta_t'\eta_t
\]

\[
L_{2T}(\beta, \Sigma) = -(1/2)(\beta - \beta^0)'(\sum_{t=T_{p+1}}^{T_p+k} X_t\Sigma^{-1}X_t')(\beta - \beta^0) + (\beta - \beta^0)'(\sum_{t=T_{p+1}}^{T_p+k} X_t\Sigma^{-1}(\Sigma^0)^{-1/2}\eta_t)
\]

Then $\log LR_k(\beta, \Sigma) = L_{1T}(\beta, \Sigma) + L_{2T}(\beta, \Sigma)$. First, for $L_{1T}(\beta, \Sigma)$,

\[
L_{1T}(\beta + T^{-1/2}\lambda, \Sigma + T^{-1/2}\Xi) - L_{1T}(\beta, \Sigma) = -(k/2) \log |I + \Psi_T^1| - \log |I + \Psi_T^2| - (1/2) \sum_{t=T_{p+1}}^{T_p+k} \eta_t'((I + \Psi_T^1)^{-1} - (I + \Psi_T^2)^{-1})\eta_t
\]

with $I + \Psi_T^1 = (\Sigma^0)^{-1/2}(\Sigma - \Sigma^0)^{-1/2} + T^{-1/2}(\Sigma^0)^{-1/2}\Xi(\Sigma^0)^{-1/2}$, $I + \Psi_T^2 = (\Sigma^0)^{-1/2}(\Sigma - \Sigma^0)^{-1/2}$. Let $A = (\Sigma^0)^{-1/2}(\Sigma - \Sigma^0)^{-1/2}$, $B = (\Sigma^0)^{-1/2}\Xi(\Sigma^0)^{-1/2}$, then $I + \Psi_T^1 = A + T^{-1/2}B$, $I + \Psi_T^2 = A$. Since

\[
(I + \Psi_T^1)^{-1} - (I + \Psi_T^2)^{-1} = (A + T^{-1/2}B)^{-1} - A^{-1}
\]

\[
= (I - T^{-1/2}A^{-1}B + T^{-1/2}A^{-1}BA^{-1}B)A^{-1} - A^{-1}
\]

\[
= -T^{-1/2}A^{-1}BA^{-1} + O_p(T^{-1})
\]

then

\[
-(1/2) \sum_{t=T_{p+1}}^{T_p+k} \eta_t'((I + \Psi_T^1)^{-1} - (I + \Psi_T^2)^{-1})\eta_t
\]

\[
= (1/2)T^{-1/2}tr(A^{-1}BA^{-1} - \sum_{t=T_{p+1}}^{T_p+k} (\eta_t\eta_t' - I)) + (k/2)T^{-1/2}tr(A^{-1}BA^{-1} - o_p(1))
\]

\[
= (k/2)T^{-1/2}tr(A^{-1}BA^{-1} - o_p(1)).
\]
Given that

\[-(k/2)[\log |I + \Psi_T^1| - \log |I + \Psi_T^2|]\]

\[= -(k/2)[\log |A + T^{-1/2}B| - \log |A|]\]

\[= -(k/2)[\log |A| + tr(A^{-1}T^{-1/2}B) - tr(T^{-1}A^{-1}BA^{-1}B) + o_p(T^{-1}) - \log |A|]\]

\[= -(k/2)T^{-1/2}tr(A^{-1}B) + o_p(1),\]

we have

\[L_{1T}(\beta + T^{-1/2}\lambda, \Sigma + T^{-1/2}\Xi) - L_{1T}(\beta, \Sigma)\]

\[= (k/2)T^{-1/2}[tr(A^{-1}BA^{-1} - tr(A^{-1}B)] + o_p(1)\]

\[= -(k/2)T^{-1/2}tr(A^{-1}BA^{-1}(\Sigma_0)^{-1/2}(\Sigma - \Sigma_0)(\Sigma_0)^{-1/2}) + o_p(1) = O_p(1)\]

and

\[L_{2T}(\beta + T^{-1/2}\lambda, \Sigma + T^{-1/2}\Xi) - L_{2T}(\beta, \Sigma)\]

\[= -(1/2)(\beta + T^{-1/2}\lambda - \beta^0)^t(\sum_{t=T_{p+1}}^{T_{p+k}} X_T t((\Sigma + T^{-1/2}\Xi)^{-1} - \Sigma^{-1})X_T t)(\beta + T^{-1/2}\lambda - \beta^0)\]

\[+ (\beta + T^{-1/2}\lambda - \beta^0)^t(\sum_{t=T_{p+1}}^{T_{p+k}} X_T t((\Sigma + T^{-1/2}\Xi)^{-1} - \Sigma^{-1})X_T t)(\beta - \beta^0)\]

\[+ (1/2)(\beta - \beta^0)^t(\sum_{t=T_{p+1}}^{T_{p+k}} X_T t((\Sigma + T^{-1/2}\Xi)^{-1} - \Sigma^{-1})X_T t)(\beta - \beta^0)\]

\[= -(1/2)(\beta - \beta^0)^t(\sum_{t=T_{p+1}}^{T_{p+k}} X_T t((\Sigma + T^{-1/2}\Xi)^{-1} - \Sigma^{-1})X_T t)(\beta - \beta^0)\]

\[+ (\beta - \beta^0)^t(\sum_{t=T_{p+1}}^{T_{p+k}} X_T t((\Sigma + T^{-1/2}\Xi)^{-1} - \Sigma^{-1})X_T t)(\beta - \beta^0)\]

\[= O_p(1)\]

since except the fourth term, which is \(O_p(1)\), all other terms are \(o_p(1)\). Hence, \(\log LR_k(\beta + T^{-1/2}\lambda, \Sigma + T^{-1/2}\Xi) - \log LR_k(\beta, \Sigma) = O_p(1)\).

**Property 6.** With \(v_T\) satisfying Assumption A6, for each \(\beta\) and \(\Sigma\) satisfying \(||\beta - \beta^0|| \leq M v_T\) and \(||\Sigma - \Sigma_0|| \leq M v_T\), with \(M < \infty\), we have

\[
\sup_{1 \leq k \leq M v_T^{-2}} \sup_{\lambda, \Xi} \frac{LR_k(\beta + T^{-1/2}\lambda, \Sigma + T^{-1/2}\Xi)}{LR_k(\beta, \Sigma)} = o_p(1)
\]

The proof is quite similar to the proof of Property 5, except for two terms whose order are \(o_p(1)\) instead of \(O_p(1)\). Given these properties, we are in a position to prove results about the consistency and rate of convergence of the estimates of the break dates. We start with the following proposition.

**Proposition 1** Under Assumption A1-A9, we have for every \(\epsilon > 0\), \(P(\|\hat{K}_j - K_j^0\| > \sqrt{T}v_T^{-1}) < \epsilon\) \((j = 1, ..., m)\).
Proof. Let \( N = \sqrt{TV_T^{-1}} \) and \( A_j = \{ (K_1, ..., K_m) : |K_j - K_0^j| > N \} \). Then to prove that \( (\overline{K}_1, ..., \overline{K}_m) \notin A_j \), it suffices to show that \( P(\sup_{(K_1, ..., K_m) \in A_j} LR(K_1, ..., K_m) \geq 1) \leq \epsilon \). Let \( LR(K_1, ..., K_m) \) denote the likelihood ratio evaluated at \( (K_1, ..., K_m) \) and note that \( LR(K_1, ..., K_m) \geq LR(K_1^0, ..., K_m^0, \beta^0, \Sigma^0) \). We consider the five possible cases separately.

Case 1. \( K_1 \leq \cdots \leq K_j \leq K_j^0 - N < K_j^0 < K_{j+1}^0 + N < K_{j+1} < \cdots , K_m \). Then

\[
LR(K_1, ..., K_m) = \prod_{l=1}^{j} LR(K_{l-1} + 1, K_l) \cdot LR(K_j + 1, K_{j+1}) \cdot \prod_{l=j+2}^{m+2} LR(K_{l-1} + 1, K_l)
\]
or

\[
\log LR(K_1, ..., K_m) = \sum_{l=1}^{j} \log LR(K_{l-1}+1, K_l) + \log LR(K_{j+1}+1, K_{j+1}) + \sum_{l=j+2}^{m+1} \log LR(K_{l-1}+1, K_l).
\]

From Properties 2 and 3, there exists constants \( C_1 > 0 \) and \( C_2 > 0 \), so that

\[
\log LR(K_1, ..., K_m) < C_1 \log^2 T - C_2 \sqrt{TV_T} = C_1 \log^2 T[1 - C_2 \sqrt{TV_T}/(\log^2 T)] \to -\infty
\]
as \( T \to \infty \).

Case 2. \( K_1 \leq \cdots \leq K_m \leq K_j^0 - N < K_j^0 < K_j^0 + N \). Then

\[
LR(K_1, ..., K_m) = \prod_{l=1}^{m+1} LR(K_{l-1}+1, K_l) \cdot LR(K_m+1, K_{j}^0-N) \cdot LR(K_j^0-N, K_j^0+N) \cdot LR(K_j^0+N+1, T)
\]
or

\[
\log LR(K_1, ..., K_m) = \sum_{l=1}^{m+1} \log LR(K_{l-1}+1, K_l) + \log LR(K_m+1, K_{j}^0-N) + \log LR(K_j^0-N, K_j^0+N) + \log LR(K_j^0+N+1, T).
\]

Case 3. \( K_j^0 - N < K_j^0 < K_j^0 + N \leq K_1 \leq \cdots \leq K_m \). Then

\[
LR(K_1, ..., K_m) = LR(1, K_j^0-N) \cdot LR(K_j^0-N+1, K_j^0+N) \cdot LR(K_j^0+N, K_1) \cdot \prod_{l=2}^{m+1} LR(K_{l-1}+1, K_l)
\]
or

\[
\log LR(K_1, ..., K_m) = \log LR(1, K_j^0-N) + \log LR(K_j^0-N+1, K_j^0+N) + \log LR(K_j^0+N, K_1) + \sum_{l=2}^{m+1} \log LR(K_{l-1}+1, K_l).
\]

Case 4. There exist \( i, i > j \), such that \( K_1 \leq \cdots \leq K_j \leq \cdots \leq K_i \leq K_i^0 - N < K_i^0 < K_j^0 + N \leq K_{i+1} \leq \cdots \leq K_m \). Then

\[
LR(K_1, ..., K_m) = \prod_{l=1}^{i+1} LR(K_{l-1}+1, K_l) \cdot LR(K_i+1, K_j^0-N) \cdot LR(K_j^0-N, K_j^0+N)
\]
\[
\cdot LR(K_j^0+N+1, K_{i+1}) \cdot \prod_{l=i+2}^{m+1} LR(K_{l-1}+1, K_l),
\]
and

\[
\log LR(K_1, ..., K_m) = \sum_{l=1}^{i+1} \log LR(K_{l-1}+1, K_l) + \log LR(K_i+1, K_j^0-N) + \log LR(K_j^0-N, K_j^0+N)
\]
\[
+ \log LR(K_j^0+N+1, K_{i+1}) + \sum_{l=i+2}^{m+1} \log LR(K_{l-1}+1, K_l).
\]
Case 5. The region $[K_0^0 - N, K_j^0 + N]$ is partitioned into $C$ intervals, in which case there must be at least one sub-interval such that $[K_{t-1} + 1, K_t] \geq C^{-1} \sqrt{T}v_T^{-1}$, then

$$LR(K_1, ..., K_m) = \prod_{j=1}^{m+1} LR(K_{j-1} + 1, K_j) \cdot \prod_{i=j+2}^{m+1} LR(K_{i-1} + 1, K_i) \cdot LR(K_{j+1} + 1, K_j)$$

and

$$\log LR(K_1, ..., K_m) = \sum_{i=1}^{j} \log LR(K_{i-1} + 1, K_i) + \sum_{i=j+2}^{j+C} \log LR(K_{i-1} + 1, K_i)$$

For all the cases, the log-likelihood function has at least one term with $K_t - K_{t-1} \geq \sqrt{T}v_T^{-1}$, $||\beta - \beta^0|| \geq dv_T$ or $||\Sigma - \Sigma^0|| \geq dv_T$ (given that a structural break happens). Hence, applying Property 3, there is at least one term in the log-likelihood function such that

$$P(\sup_{\beta, \Sigma} \sup_{k \leq T} \log LR_k(\beta, \Sigma) > -C\sqrt{T}v_T) < \epsilon.$$ 

Applying Property 2 to the other terms, we have

$$P(\sup_{1 \leq k \leq T} \log LR_k(\beta, \Sigma) > B(\log T)^2) < \epsilon.$$ 

Hence, for the whole log-likelihood function, there exist constants $C_1 > 0$ and $C_2 > 0$, such that

$$\log LR(K_1, ..., K_m) < C_1 \log^2 T - C_2 \sqrt{T}v_T = C_1 \log^2 T[1 - C_2 \sqrt{T}v_T/(\log^2 T)] \rightarrow -\infty$$

as $T \rightarrow \infty$. This shows that $P(\sup_{(K_1, ..., K_m) \in A_j} L(K_1, ..., K_m) \geq 1) \leq \epsilon$.

**Proposition 2** For every $\epsilon > 0$, there exists a $C > 0$, such that $P(||\hat{K}_j - K_j^0|| > C\sqrt{T}^{-2}) < \epsilon$.

**Proof.** Given Proposition 1, it remains to prove that

$$P(\sup_{(K_1, ..., K_m) \in A_j(C)} L(K_1, ..., K_m) \geq 1) \leq \epsilon$$

for $A_j(C) = \{(K_1, ..., K_m) : |K_j - K_j^0| \leq \sqrt{T}v_T^{-1}$ and $|K_j - K_j^0| > C\sqrt{T}^{-2}\}$. Using Property 5, we can show that the estimates of the coefficients are $\sqrt{T}$ consistent, then applying arguments as in Bai (2000, pp. 327-328) to the five cases analysed above, we obtain the desired result. This establishes Theorem 1. The results pertaining to the estimate of $\beta^0$ and $\Sigma^0$ follow given the rate of convergence of $\hat{K}_j$.

**Proof of Theorem 2:** The likelihood ratio is

$$LR_T(K_1, ..., K_m, \beta(j), \Sigma(j))$$

$$= \prod_{j=1}^{m+1} \prod_{t=K_{j-1} + 1}^{K_j} f(Y_t | X_{Tt}, \beta(j), \Sigma(j)) - \prod_{j=1}^{m+1} \prod_{t=K_{j-1} + 1}^{K_j} f(Y_t | X_{Tt}, \beta(j), \Sigma(j))^{-1} \exp\left\{-(1/2)(Y_t - X_{Tt}^\top \beta(j))'(\Sigma(j))^{-1}(Y_t - X_{Tt}^\top \beta(j))\right\}$$

$$- \prod_{j=1}^{m+1} \prod_{t=K_{j-1} + 1}^{K_j} (2\pi)^{-1/2}|\Sigma(j)|^{-1/2} \exp\left\{-(1/2)(Y_t - X_{Tt}^\top \beta(j))'(\Sigma(j))^{-1}(Y_t - X_{Tt}^\top \beta(j))\right\}.$$
With the restrictions $g(\beta_j, \text{vec}(\Sigma_j)) = 0$, the restricted log-likelihood function is
\[
rlr_T(K_j, \beta_j, \Sigma_j) = -(1/2) \sum_{j=1}^{m+1} (K_j - K_{j-1}) \log |\Sigma_j| \\
- (1/2) \sum_{j=1}^{m+1} \sum_{t=K_{j-1}+1}^{K_j} (Y_t - X_T'\beta_j)'(\Sigma_j)^{-1}(Y_t - X_T'\beta_j) + (1/2) \sum_{j=1}^{m+1} (K_j^0 - K_{j-1}^0) \log |\Sigma^0| \\
+ (1/2) \sum_{j=1}^{m+1} \sum_{t=K_{j-1}+1}^{K_j^0} (Y_t - X_T'\beta_j^0)'(\Sigma^0)^{-1}(Y_t - X_T'\beta_j^0) + \lambda'g(\beta_j, \text{vec}(\Sigma_j))
\]

Let
\[
A = \sum_{j=1}^{m+1} (K_j - K_{j-1}) \log |\Sigma_j| - \sum_{j=1}^{m+1} (K_j^0 - K_{j-1}^0) \log |\Sigma^0|
\]
\[
B = \sum_{j=1}^{m+1} \sum_{t=K_{j-1}+1}^{K_j} (Y_t - X_T'\beta_j)'(\Sigma_j)^{-1}(Y_t - X_T'\beta_j) \\
- \sum_{j=1}^{m+1} \sum_{t=K_{j-1}+1}^{K_j^0} (Y_t - X_T'\beta_j^0)'(\Sigma^0)^{-1}(Y_t - X_T'\beta_j^0)
\]
then $lr_T = -A/2 - B/2$ and $rlr_T = -A/2 - B/2 + \lambda'g(\beta_j, \text{vec}(\Sigma_j))$. We have
\[
A = \sum_{j=1}^{m+1} (K_j - K_{j-1}) \log |\Sigma_j| - \sum_{j=1}^{m+1} (K_j^0 - K_{j-1}^0) \log |\Sigma^0| \\
= \sum_{j=1}^{m+1} (K_j - K_{j-1}) \log |\Sigma_j| - \sum_{j=1}^{m+1} (K_j^0 - K_{j-1}^0) \log |\Sigma^0| \\
+ \sum_{j=1}^{m+1} (K_j^0 - K_{j-1}^0) \log |\Sigma^0| - \sum_{j=1}^{m+1} (K_j^0 - K_{j-1}^0) \log |\Sigma^0| \\
= \sum_{j=1}^{m+1} (K_j - K_j^0)(\log |\Sigma_j| - \log |\Sigma_j+1|) \\
- \sum_{j=1}^{m+1} (K_j^0 - K_{j-1}^0)(\log |\Sigma_j| - \log |\Sigma^0|)
\tag{A.1}
\]
\[
B = \sum_{j=1}^{m+1} \sum_{t=K_{j-1}+1}^{K_j} (Y_t - X_T'\beta_j)'(\Sigma_j)^{-1}(Y_t - X_T'\beta_j) \\
- \sum_{j=1}^{m+1} \sum_{t=K_{j-1}+1}^{K_j^0} (Y_t - X_T'\beta_j^0)'(\Sigma^0)^{-1}(Y_t - X_T'\beta_j^0)
\]
\[
= \sum_{j=1}^{m+1} \sum_{t=K_{j-1}+1}^{K_j} (Y_t - X_T'\beta_j)'(\Sigma_j)^{-1}(Y_t - X_T'\beta_j) \\
- \sum_{j=1}^{m+1} \sum_{t=K_{j-1}+1}^{K_j^0} (Y_t - X_T'\beta_j^0)'(\Sigma^0)^{-1}(Y_t - X_T'\beta_j^0) \\
+ \sum_{j=1}^{m+1} \sum_{t=K_{j-1}+1}^{K_j^0} (Y_t - X_T'\beta_j^0)'(\Sigma^0)^{-1}(Y_t - X_T'\beta_j^0) \\
- \sum_{j=1}^{m+1} \sum_{t=K_{j-1}+1}^{K_j^0} (Y_t - X_T'\beta_j^0)'(\Sigma^0)^{-1}(Y_t - X_T'\beta_j^0)
\tag{B.1}
\]
Let
\[
rlr_T^1(K_j, \beta_j, \Sigma_j) = -(1/2)(A.2) - (1/2)(B.2) + \lambda'g(\beta_j, \text{vec}(\Sigma_j))
\]
\[
rlr_T^2(K_j, \beta_j, \Sigma_j) = -(1/2)(A.1) - (1/2)(B.1)
\]
and note that \( r l r^2 \) will depend on the position of \( K_j \) relative to that of \( K^0_j \). Using the fact that \( \sqrt{T} (\Sigma_j - \Sigma^0) = o_p(1) \), then \( \Sigma_j = \Sigma^0 + T^{-1/2} \Sigma_j^* \) where \( \Sigma_j^* = \sqrt{T} (\Sigma_j - \Sigma^0) \). Hence,

\[
(\Sigma_j)^{-1} = (\Sigma^0 + T^{-1/2} \Sigma_j^*)^{-1} = (\Sigma^0 (I + T^{-1/2} (\Sigma^0)^{-1} \Sigma_j^*))^{-1} = (I - T^{-1/2} (\Sigma^0)^{-1} \Sigma_j^* - T^{-1} (\Sigma^0)^{-1} \Sigma_j^* (\Sigma^0)^{-1} \Sigma_j^*) + o_p(T^{-1}))(\Sigma^0)^{-1} = (\Sigma^0)^{-1} - T^{-1/2} (\Sigma^0)^{-1} \Sigma_j^* (\Sigma^0)^{-1} - T^{-1} (\Sigma^0)^{-1} \Sigma_j^* (\Sigma^0)^{-1} \Sigma_j^* (\Sigma^0)^{-1} + o_p(1)
\]

We start with the simplest case and then expend the results to more general cases.

a) \( m = 2 \), no break in variance-covariance matrix. In this case

\[
(A.1) = \sum_{j=1}^{2} (K_j - K^0_j)(\log |\Sigma_j| - \log |\Sigma_{j+1}|) = \sum_{j=1}^{2} (K_j - K^0_j)(\log |\Sigma^0| - T^{-1/2}tr((\Sigma^0)^{-1} \Sigma_j^*)) - \log |\Sigma^0| - T^{-1/2}tr((\Sigma^0)^{-1} \Sigma_j^*)) + o_p(1)
\]

For (B.1), we need to consider the three cases separately.

For Case 1 \( K_1 < K_2 \leq K^0_1 < K^0_2 \):

\[
(B.1) = -\sum_{t=K_1+1}^{K^0_1} (u_t + X_t' \beta^0_1') (\Sigma^0)^{-1} (Y_t - X_t' \beta_1) + \sum_{t=K_2+1}^{K^0_2} (u_t + X_t' \beta^0_2') (\Sigma^0)^{-1} (Y_t - X_t' \beta_2)
\]

We analyze each of the five terms separately.

1. \( \sum_{t=K_1+1}^{K^0_1} (u_t + X_t' \beta^0_1') (\Sigma^0)^{-1} (Y_t - X_t' \beta_1) \)

2. \( \sum_{t=K_2+1}^{K^0_2} (u_t + X_t' \beta^0_2') (\Sigma^0)^{-1} (Y_t - X_t' \beta_2) \)

3. \( \sum_{t=K_1+1}^{K_2} (u_t + X_t' \beta_1' - 2\Sigma^0)^{-1} (u_t + X_t' \beta^0_1') (\Sigma^0)^{-1} (Y_t - X_t' \beta_1) + o_p(1) \)

4. \( \sum_{t=K_2+1}^{K^0_2} (u_t + X_t' \beta_2' - 2\Sigma^0)^{-1} (u_t + X_t' \beta^0_2') (\Sigma^0)^{-1} (Y_t - X_t' \beta_2) + o_p(1) \)

5. \( \sum_{t=K_1+1}^{K_2} (u_t + X_t' \beta_1' - 2\Sigma^0)^{-1} (u_t + X_t' \beta^0_1') (\Sigma^0)^{-1} (Y_t - X_t' \beta_1) + o_p(1) \)

A-8
(4) \[ r \text{tr}_T^2(K_1, K_2, K_j, K_0, \beta, \Sigma) = \sum_{t=K_2+1}^{K_0} (u_t + X_T^t \beta_{(1)} - X_T^t \beta_{(3)})^t (\Sigma_{(3)})^{-1} (u_t + X_T^t \beta_{(1)} - X_T^t \beta_{(3)}) \]
\[ = \sum_{t=K_2+1}^{K_0} u_t^t (\Sigma_{(1)})^{-1} u_t - T^{-1/2} \sum_{t=K_2+1}^{K_0} u_t^t (\Sigma_{(2)})^{-1} \Sigma_{(2)}^{-1} u_t + (\beta_{(3)} - \beta_{(1)})^t (\Sigma_{(3)})^{-1} X_T^t (\Sigma_{(3)})^{-1} (\beta_{(3)} - \beta_{(1)}) - 2(\beta_{(3)} - \beta_{(1)})^t (\Sigma_{t=K_2+1}^{K_0}) X_T^t (\Sigma_{(3)})^{-1} u_t + o_p(1) \]

(5) \[ r \text{tr}_T^2(K_1, K_2, K_j, K_0, \beta, \Sigma) = \sum_{t=K_2+1}^{K_0} (u_t + X_T^t \beta_{(2)} - X_T^t \beta_{(3)})^t (\Sigma_{(3)})^{-1} (u_t + X_T^t \beta_{(2)} - X_T^t \beta_{(3)}) \]
\[ = \sum_{t=K_2+1}^{K_0} u_t^t (\Sigma_{(1)})^{-1} u_t - T^{-1/2} \sum_{t=K_2+1}^{K_0} u_t^t (\Sigma_{(2)})^{-1} \Sigma_{(2)}^{-1} u_t + (\beta_{(3)} - \beta_{(2)})^t (\Sigma_{(3)})^{-1} X_T^t (\beta_{(3)} - \beta_{(2)}) - 2(\beta_{(3)} - \beta_{(2)})^t (\Sigma_{t=K_2+1}^{K_0}) X_T^t (\Sigma_{(3)})^{-1} u_t + o_p(1) \]

Hence,

\[ r \text{tr}_T^2(K_1, K_2, \beta, \Sigma) = -(1/2) \sum_{j=1}^{2} (K_j - K_j^0) \left[ \log |\Sigma_{(j)}| - T^{-1/2} \text{tr}((\Sigma_{(j)}^{-1} \Sigma_{(j)})^{-1} \Sigma_{(j+1)}) \right] - (1/2) \left[ - \sum_{t=K_1+1}^{K_2+1} u_t^t (\Sigma_{(1)})^{-1} u_t + T^{-1/2} \sum_{t=K_1+1}^{K_2+1} u_t^t (\Sigma_{(2)})^{-1} \Sigma_{(2)}^{-1} u_t \right.
\[ - \sum_{t=K_1+1}^{K_2+1} u_t^t (\Sigma_{(3)})^{-1} u_t - T^{-1/2} \sum_{t=K_1+1}^{K_2+1} u_t^t (\Sigma_{(3)})^{-1} \Sigma_{(3)}^{-1} u_t \]
\[ - (1/2) [(\beta_{(2)} - \beta_{(1)})^t (\Sigma_{(3)})^{-1} X_T^t ((\beta_{(2)} - \beta_{(1)})) + (\beta_{(3)} - \beta_{(1)})^t (\Sigma_{(3)})^{-1} X_T^t (\beta_{(3)} - \beta_{(1)}) + (\beta_{(3)} - \beta_{(2)})^t (\Sigma_{(3)})^{-1} X_T^t (\beta_{(3)} - \beta_{(2)}) + (\beta_{(3)} - \beta_{(3)})^t (\Sigma_{(3)})^{-1} X_T^t (\beta_{(3)} - \beta_{(3)})] \]
\[ + (\beta_{(3)} - \beta_{(1)})^t (\Sigma_{(3)})^{-1} X_T^t (\beta_{(3)} - \beta_{(1)}) + (\beta_{(3)} - \beta_{(2)})^t (\Sigma_{(3)})^{-1} X_T^t (\beta_{(3)} - \beta_{(2)}) + (\beta_{(3)} - \beta_{(3)})^t (\Sigma_{(3)})^{-1} X_T^t (\beta_{(3)} - \beta_{(3)})] \]
\[ + o_p(1) \]

Note that

\[ (K_j - K_j^0) T^{-1/2} \text{tr}((\Sigma_{(j)}^{-1} \Sigma_{(j)})^{-1} \Sigma_{(j)}) = T^{-1/2} \sum_{t=K_1+1}^{K_0} u_t^t (\Sigma_{(j)}^{-1} \Sigma_{(j)})^{-1} u_t \]
\[ = (K_j - K_j^0) T^{-1/2} \text{tr}((\Sigma_{(j)}^{-1} \Sigma_{(j)})^{-1} \Sigma_{(j)}) + T^{-1/2} \text{tr}((\Sigma_{(j)}^{-1} \Sigma_{(j)})^{-1} \sum_{t=K_1+1}^{K_0} u_t u_t) \]
\[ = (K_j - K_j^0) T^{-1/2} \text{tr}((\Sigma_{(j)}^{-1} \Sigma_{(j)})^{-1} \sum_{t=K_1+1}^{K_0} u_t u_t - \Sigma_{(j)}) - (K_j - K_j^0) T^{-1/2} \text{tr}((\Sigma_{(j)}^{-1} \Sigma_{(j)})^{-1} \sum_{t=K_1+1}^{K_0} u_t u_t - \Sigma_{(j)}) = o_p(1) \]
\[ v_T \sum_{t=K_1+1}^{K_0} (u_t u_t' - \Sigma^0) = (\Sigma^0)^{1/2} [v_T \sum_{t=K_1+1}^{K_0} (\eta_t' \eta_t - I)] (\Sigma^0)^{1/2} \Rightarrow (\Sigma^0)^{1/2} [\xi(s_j)] (\Sigma^0)^{1/2} \]

\[
\sum_{j=1}^{2} (K_j - K_0) [- T^{-1/2} tr((\Sigma^0)^{-1} \Sigma_j) - T^{-1/2} tr((\Sigma^0)^{-1} \Sigma_{j+1})] \\
+ \sum_{t=K_1+1}^{K_0} u_t' (\Sigma^0)^{-1} u_t + \sum_{t=K_1+1}^{K_0} u_t' (\Sigma^0)^{-1} u_t - \sum_{t=K_1+1}^{K_0} u_t' (\Sigma^0)^{-1} u_t \\
- \sum_{t=K_2+1}^{K_0} u_t' (\Sigma^0)^{-1} u_t - \sum_{t=K_1+1}^{K_0} u_t' (\Sigma^0)^{-1} u_t = o_p(1)
\]

So (C.1) becomes \( o_p(1) \). For (C.2) and (C.3), we have

\[
(C.2) = - (1/2) [(\beta_0 - \beta_0)'] (\sum_{t=K_1+1}^{K_0} X_T (\Sigma^0)^{-1} X_t') (\beta_0 - \beta_0) \\
+ (\beta_0 - \beta_0) (\sum_{t=K_1+1}^{K_0} X_T (\Sigma^0)^{-1} X_t') (\beta_0 - \beta_0) \\
+ (\beta_0 - \beta_0) (\sum_{t=K_1+1}^{K_0} X_T (\Sigma^0)^{-1} X_t') (\beta_0 - \beta_0) \\
+ 2(\beta_0 - \beta_0) (\sum_{t=K_1+1}^{K_0} X_T (\Sigma^0)^{-1} X_t') (\beta_0 - \beta_0) \\
+ (\beta_0 - \beta_0) (\sum_{t=K_1+1}^{K_0} X_T (\Sigma^0)^{-1} X_t') (\beta_0 - \beta_0) \\
= - (1/2) [(\beta_0 - \beta_0)'] (\sum_{t=K_1+1}^{K_0} X_T (\Sigma^0)^{-1} X_t') (\beta_0 - \beta_0) \\
+ (\beta_0 - \beta_0) (\sum_{t=K_1+1}^{K_0} X_T (\Sigma^0)^{-1} X_t') (\beta_0 - \beta_0) \\
+ 2(\beta_0 - \beta_0) (\sum_{t=K_1+1}^{K_0} X_T (\Sigma^0)^{-1} X_t') (\beta_0 - \beta_0) \\
+ (\beta_0 - \beta_0) (\sum_{t=K_1+1}^{K_0} X_T (\Sigma^0)^{-1} X_t') (\beta_0 - \beta_0) \\
+ (\beta_0 - \beta_0) (\sum_{t=K_1+1}^{K_0} X_T (\Sigma^0)^{-1} X_t') (\beta_0 - \beta_0) \\
= \beta_0 (\sum_{t=K_1+1}^{K_0} X_T (\Sigma^0)^{-1} X_t') (\beta_0 - \beta_0) \\
+ (\beta_0 - \beta_0) (\sum_{t=K_1+1}^{K_0} X_T (\Sigma^0)^{-1} X_t') (\beta_0 - \beta_0) \\
= \beta_0 (\sum_{t=K_1+1}^{K_0} X_T (\Sigma^0)^{-1} X_t') (\beta_0 - \beta_0) \\
+ (\beta_0 - \beta_0) (\sum_{t=K_1+1}^{K_0} X_T (\Sigma^0)^{-1} X_t') (\beta_0 - \beta_0) \\
\]

Finally,

\[
r h_T^2 (K_1, K_2, \beta, \Sigma) = - (1/2) [(\beta_0 - \beta_0)'] (\sum_{t=K_1+1}^{K_0} X_T (\Sigma^0)^{-1} X_t') (\beta_0 - \beta_0) \\
- (1/2) [(\beta_0 - \beta_0)'] (\sum_{t=K_1+1}^{K_0} X_T (\Sigma^0)^{-1} X_t') (\beta_0 - \beta_0) \\
- (\beta_0 - \beta_0) (\sum_{t=K_1+1}^{K_0} X_T (\Sigma^0)^{-1} X_t') (\beta_0 - \beta_0) \\
+ (\beta_0 - \beta_0) (\sum_{t=K_1+1}^{K_0} X_T (\Sigma^0)^{-1} X_t') (\beta_0 - \beta_0) \\
+ (\beta_0 - \beta_0) (\sum_{t=K_1+1}^{K_0} X_T (\Sigma^0)^{-1} X_t') (\beta_0 - \beta_0) \\
+ (\beta_0 - \beta_0) (\sum_{t=K_1+1}^{K_0} X_T (\Sigma^0)^{-1} X_t') (\beta_0 - \beta_0) \\
+ (\beta_0 - \beta_0) (\sum_{t=K_1+1}^{K_0} X_T (\Sigma^0)^{-1} X_t') (\beta_0 - \beta_0) + o_p(1)
\]
Case 2: $K_1^0 < K_2^0 \leq K_1 < K_2$. For (B.1), we have

\begin{align}
(B.1) &= \sum_{t=K_1^0+1}^{K_2^0} (Y_t - X_{Tt} \beta(1))' (\Sigma(1)^{-1}) (Y_t - X_{Tt} \beta(1)) \\
&\quad + \sum_{t=K_1^0+1}^{K_2^0} (Y_t - X_{Tt} \beta(2))' (\Sigma(2)^{-1}) (Y_t - X_{Tt} \beta(2)) \\
&\quad - \sum_{t=K_1^0+1}^{K_2^0} (Y_t - X_{Tt} \beta(3))' (\Sigma(3)^{-1}) (Y_t - X_{Tt} \beta(3)) \\
&\quad + \sum_{t=K_1^0+1}^{K_2^0} (Y_t - X_{Tt} \beta(2))' (\Sigma(2)^{-1}) (Y_t - X_{Tt} \beta(2))
\end{align}

so that

\begin{align}
rlr_T^2(K_1, K_2, \beta, \Sigma) &= -(1/2) (\beta(2) - \beta(1))' (\sum_{t=K_1^0+1}^{K_2^0} X_{Tt} (\Sigma(0)^{-1}) X_{Tt}) (\beta(2) - \beta(1)) \\
&\quad - (1/2) (\beta(3) - \beta(2))' (\sum_{t=K_1^0+1}^{K_2^0} X_{Tt} (\Sigma(0)^{-1}) X_{Tt}) (\beta(3) - \beta(2)) \\
&\quad - (\beta(2) - \beta(1))' (\sum_{t=K_1^0+1}^{K_2^0} X_{Tt} (\Sigma(0)^{-1}) X_{Tt}) (\beta(2) - \beta(1)) \\
&\quad - (\beta(3) - \beta(2))' (\sum_{t=K_1^0+1}^{K_2^0} X_{Tt} (\Sigma(0)^{-1}) X_{Tt}) (\beta(3) - \beta(2)) \\
&\quad - (\beta(3) - \beta(2))' (\sum_{t=K_1^0+1}^{K_2^0} X_{Tt} (\Sigma(0)^{-1}) u_t + o_p(1))
\end{align}

Case 3: $K_1 \leq K_1^0 < K_2^0 \leq K_2$. For (B.1), we have,

\begin{align}
(B.1) &= -\sum_{t=K_1+1}^{K_2} (Y_t - X_{Tt} \beta(2))' (\Sigma(2)^{-1}) (Y_t - X_{Tt} \beta(2)) \\
&\quad - \sum_{t=K_1+1}^{K_2} (Y_t - X_{Tt} \beta(1))' (\Sigma(1)^{-1}) (Y_t - X_{Tt} \beta(1)) \\
&\quad + \sum_{t=K_2+1}^{K_2+1} (Y_t - X_{Tt} \beta(2))' (\Sigma(2)^{-1}) (Y_t - X_{Tt} \beta(2)) \\
&\quad - \sum_{t=K_2+1}^{K_2+1} (Y_t - X_{Tt} \beta(3))' (\Sigma(3)^{-1}) (Y_t - X_{Tt} \beta(3))
\end{align}

and using arguments similar to those for Case 1,

\begin{align}
rlr_T^2(K_1, K_2, \beta, \Sigma) &= -(1/2) (\beta(2) - \beta(1))' (\sum_{t=K_1+1}^{K_2} X_{Tt} (\Sigma(0)^{-1}) X_{Tt}) (\beta(2) - \beta(1)) \\
&\quad - (1/2) (\beta(3) - \beta(2))' (\sum_{t=K_1+1}^{K_2} X_{Tt} (\Sigma(0)^{-1}) X_{Tt}) (\beta(3) - \beta(2)) + (\beta(2) - \beta(1))' (\sum_{t=K_1+1}^{K_2} X_{Tt} (\Sigma(0)^{-1}) u_t) \\
&\quad - (\beta(3) - \beta(2))' (\sum_{t=K_1+1}^{K_2} X_{Tt} (\Sigma(0)^{-1}) u_t) + o_p(1)
\end{align}

b) Multiple breaks in coefficients only.

Case 1: $K_1 < \ldots < K_m \leq K_1^0 < \ldots < K_m^0$. In this more general case, we have:

\begin{align}
(A.1) &= \sum_{j=1}^{m} (K_j - K_j^0) \left( \log |\Sigma(j)| - \log |\Sigma(j+1)| \right) \\
&= \sum_{j=1}^{m} (K_j - K_j^0) \left( \log |\Sigma(0)| - T^{-1/2} tr((\Sigma(0)^{-1}) (\Sigma(j)^*)^{-1}) + \log |\Sigma(0)| + T^{-1/2} tr((\Sigma(0)^{-1}) (\Sigma(j+1)^*)^{-1})) + o_p(1) \right)
\end{align}
\[(B.1) = - \sum_{t=K_t+1}^{K_p} (Y_t - X_{T_t}^\prime \beta(1))' (\Sigma(1))^{-1} (Y_t - X_{T_t}^\prime \beta(1)) \quad \text{(D.1)}
\]
+ \sum_{j=2}^{m} \sum_{t=K_{j-1}+1}^{K_j} (Y_t - X_{T_t}^\prime \beta(j))' (\Sigma(j))^{-1} (Y_t - X_{T_t}^\prime \beta(j))
\quad \text{(D.2)}
- \sum_{j=2}^{m} \sum_{t=K_{j-1}+1}^{K_j} (Y_t - X_{T_t}^\prime \beta(j))' (\Sigma(j))^{-1} (Y_t - X_{T_t}^\prime \beta(j))
\quad \text{(D.3)}
+ \sum_{t=K_m+1}^{K_0} (Y_t - X_{T_t}^\prime \beta(m+1))' (\Sigma(m+1))^{-1} (Y_t - X_{T_t}^\prime \beta(m+1))
\quad \text{(D.4)}
+ \sum_{j=2}^{m} \sum_{t=K_{j-1}+1}^{K_j} (Y_t - X_{T_t}^\prime \beta(m+1))' (\Sigma(m+1))^{-1} (Y_t - X_{T_t}^\prime \beta(m+1))
\quad \text{(D.5)}
\]

with

\[(D.1) = - \sum_{t=K_t+1}^{K_p} (u_t + X_{T_t}^\prime \beta(1))' (\Sigma(1))^{-1} (u_t + X_{T_t}^\prime \beta(1) - X_{T_t}^\prime \beta(1))
\quad = - \sum_{t=K_t+1}^{K_p} u_t' (\Sigma(0))^{-1} u_t + T^{-1/2} \sum_{t=K_t+1}^{K_p} u_t' (\Sigma(0))^{-1} \Sigma(1) (\Sigma(0))^{-1} u_t + o_p(1)
\]

\[(D.2) = \sum_{j=2}^{m} \sum_{t=K_{j-1}+1}^{K_j} (u_t + X_{T_t}^\prime \beta(1))' (\Sigma(1))^{-1} (u_t + X_{T_t}^\prime \beta(1) - X_{T_t}^\prime \beta(1))
\quad = \sum_{j=2}^{m} \sum_{t=K_{j-1}+1}^{K_j} u_t' (\Sigma(0))^{-1} u_t - T^{-1/2} \sum_{j=2}^{m} \sum_{t=K_{j-1}+1}^{K_j} u_t' (\Sigma(0))^{-1} \Sigma(1) (\Sigma(0))^{-1} u_t
\quad + \sum_{j=2}^{m} \sum_{t=K_{j-1}+1}^{K_j} T_t (\Sigma(0))^{-1} X_{T_t}^\prime (\beta(j) - \beta(1))
\quad - 2 \sum_{j=2}^{m} \sum_{t=K_{j-1}+1}^{K_j} T_t (\Sigma(0))^{-1} u_t + o_p(1)
\]

\[(D.3) = - \sum_{j=2}^{m} \sum_{t=K_{j-1}+1}^{K_j} (u_t + X_{T_t}^\prime \beta(j))' (\Sigma(j))^{-1} (u_t + X_{T_t}^\prime \beta(j) - X_{T_t}^\prime \beta(j))
\quad = - \sum_{j=2}^{m} \sum_{t=K_{j-1}+1}^{K_j} u_t' (\Sigma(0))^{-1} u_t + T^{-1/2} \sum_{j=2}^{m} \sum_{t=K_{j-1}+1}^{K_j} u_t' (\Sigma(0))^{-1} \Sigma(1) (\Sigma(0))^{-1} u_t + o_p(1)
\]

\[(D.4) = \sum_{t=K_m+1}^{K_0} (u_t + X_{T_t}^\prime \beta(m+1))' (\Sigma(m+1))^{-1} (u_t + X_{T_t}^\prime \beta(m+1) - X_{T_t}^\prime \beta(m+1))
\quad = \sum_{t=K_m+1}^{K_0} u_t' (\Sigma(0))^{-1} u_t - T^{-1/2} \sum_{t=K_m+1}^{K_0} u_t' (\Sigma(0))^{-1} \Sigma(1) (\Sigma(0))^{-1} u_t
\quad + \sum_{t=K_m+1}^{K_0} T_t (\Sigma(0))^{-1} X_{T_t}^\prime (\beta(m+1) - \beta(1))
\quad - 2 \sum_{t=K_m+1}^{K_0} T_t (\Sigma(0))^{-1} u_t + o_p(1)
\]

\[(D.5) = \sum_{j=2}^{m} \sum_{t=K_{j-1}+1}^{K_j} (u_t + X_{T_t}^\prime \beta(m+1))' (\Sigma(m+1))^{-1} (u_t + X_{T_t}^\prime \beta(m+1) - X_{T_t}^\prime \beta(m+1))
\quad = \sum_{j=2}^{m} \sum_{t=K_{j-1}+1}^{K_j} u_t' (\Sigma(0))^{-1} u_t - T^{-1/2} \sum_{j=2}^{m} \sum_{t=K_{j-1}+1}^{K_j} u_t' (\Sigma(0))^{-1} \Sigma(1) (\Sigma(0))^{-1} u_t
\quad + \sum_{j=2}^{m} \sum_{t=K_{j-1}+1}^{K_j} T_t (\Sigma(0))^{-1} X_{T_t}^\prime (\beta(m+1) - \beta(1))
\quad - 2 \sum_{j=2}^{m} \sum_{t=K_{j-1}+1}^{K_j} T_t (\Sigma(0))^{-1} u_t + o_p(1)
\]
Hence, \( r \ell r^2_T(K_1, \cdots, K_m, \beta, \Sigma) = -(1/2)((A.1) + (B.1)) \) is composed on the following four parts:

\[
I = -(1/2)[\sum_{j=1}^m (K_j - K^0_j)(\log |\Sigma^0| - \log |\Sigma^0|) - \sum_{t=K_{j+1}}^{K_j} u'_t(\Sigma^0)^{-1} u_t + \sum_{j=2}^m \sum_{t=K_{j-1}+1}^{K_j} u'_t(\Sigma^0)^{-1} u_t - \sum_{j=2}^m \sum_{t=K_{j-1}+1}^{K_j} u'_t(\Sigma^0)^{-1} u_t + \sum_{t=K_{m+1}}^{K_j} u'_t(\Sigma^0)^{-1} u_t = 0
\]

\[
II = -(1/2)[-\sum_{j=1}^m T^{-1/2}(K_j - K^0_j)(tr((\Sigma^0)^{-1} \Sigma^*) - T^{-1/2}tr((\Sigma^0)^{-1} \Sigma^*)) + T^{-1/2} \sum_{t=K_{j+1}}^{K_j} u'_t(\Sigma^0)^{-1} \Sigma^*(\Sigma^0)^{-1} u_t - T^{-1/2} \sum_{j=2}^m \sum_{t=K_{j-1}+1}^{K_j} u'_t(\Sigma^0)^{-1} \Sigma^*(\Sigma^0)^{-1} u_t + T^{-1/2} \sum_{j=2}^m \sum_{t=K_{j-1}+1}^{K_j} u'_t(\Sigma^0)^{-1} \Sigma^*(\Sigma^0)^{-1} u_t - T^{-1/2} \sum_{t=K_{m+1}}^{K_j} u'_t(\Sigma^0)^{-1} \Sigma^*(\Sigma^0)^{-1} u_t - T^{-1/2} \sum_{j=2}^m \sum_{t=K_{j-1}+1}^{K_j} u'_t(\Sigma^0)^{-1} \Sigma^*(\Sigma^0)^{-1} u_t = o_p(1)
\]

\[
III = -(1/2)[\sum_{j=1}^m (\beta^0_j - \beta^0_{(1)})'((\sum_{t=K_{j-1}+1}^{K_j} X_{Tt}(\Sigma^0)^{-1} X_{Tt}')(\beta^0_j - \beta^0_{(1)}) + (\beta^0_{(m+1)} - \beta^0_{(1)})'((\sum_{t=K_{m+1}+1}^{K_j} X_{Tt}(\Sigma^0)^{-1} X_{Tt}')(\beta^0_{(m+1)} - \beta^0_{(1)}) + \sum_{j=2}^m (\beta^0_{(m+1)} - \beta^0_{(1)})'((\sum_{t=K_{j-1}+1}^{K_j} X_{Tt}(\Sigma^0)^{-1} X_{Tt}')(\beta^0_{(m+1)} - \beta^0_{(1)})]
\]

\[
= -(1/2)[\sum_{j=1}^m (\beta^0_{(j+1)} - \beta^0_{(j)})'((\sum_{t=K_{j-1}+1}^{K_j} X_{Tt}(\Sigma^0)^{-1} X_{Tt}')(\beta^0_{(j+1)} - \beta^0_{(j)}) + \sum_{j=1}^m (\beta^0_{(j+1)} - \beta^0_{(j)})'((\sum_{t=K_{j+1}+1}^{K_j} X_{Tt}(\Sigma^0)^{-1} X_{Tt}')(\beta^0_{(j+1)} - \beta^0_{(j)}) + \sum_{j=1}^m (\beta^0_{(j+1)} - \beta^0_{(j)})'((\sum_{t=K_{j+1}+1}^{K_j} X_{Tt}(\Sigma^0)^{-1} X_{Tt}')(\beta^0_{(j+1)} - \beta^0_{(j)}) + 2\sum_{j=1}^m (\beta^0_{(j+1)} - \beta^0_{(j)})'((\sum_{t=K_{j+1}+1}^{K_j} X_{Tt}(\Sigma^0)^{-1} X_{Tt}')(\beta^0_{(j+1)} - \beta^0_{(j)})]
\]

\[
= \sum_{j=1}^m (\beta^0_{(j+1)} - \beta^0_{(j)})'((\sum_{t=K_{j+1}+1}^{K_j} X_{Tt}(\Sigma^0)^{-1} X_{Tt}')(\beta^0_{(j+1)} - \beta^0_{(j)}) + 2\sum_{j=1}^m (\beta^0_{(j+1)} - \beta^0_{(j)})'((\sum_{t=K_{j+1}+1}^{K_j} X_{Tt}(\Sigma^0)^{-1} X_{Tt}')(\beta^0_{(j+1)} - \beta^0_{(j)}) + 3\sum_{j=1}^m (\beta^0_{(j+1)} - \beta^0_{(j)})'((\sum_{t=K_{j+1}+1}^{K_j} X_{Tt}(\Sigma^0)^{-1} X_{Tt}')(\beta^0_{(j+1)} - \beta^0_{(j)})]
\]

\[
= \sum_{j=1}^m (\beta^0_{(j+1)} - \beta^0_{(j)})'((\sum_{t=K_{j+1}+1}^{K_j} X_{Tt}(\Sigma^0)^{-1} X_{Tt}')(\beta^0_{(j+1)} - \beta^0_{(j)}) + 2\sum_{j=1}^m (\beta^0_{(j+1)} - \beta^0_{(j)})'((\sum_{t=K_{j+1}+1}^{K_j} X_{Tt}(\Sigma^0)^{-1} X_{Tt}')(\beta^0_{(j+1)} - \beta^0_{(j)}) + 3\sum_{j=1}^m (\beta^0_{(j+1)} - \beta^0_{(j)})'((\sum_{t=K_{j+1}+1}^{K_j} X_{Tt}(\Sigma^0)^{-1} X_{Tt}')(\beta^0_{(j+1)} - \beta^0_{(j)})]
\]

\[
IV = -2\sum_{j=2}^m (\beta^0_{(j+1)} - \beta^0_{(j)})'((\sum_{t=K_{j-1}+1}^{K_j} X_{Tt}(\Sigma^0)^{-1} u_t) - 2(\beta^0_{(m+1)} - \beta^0_{(1)})'((\sum_{t=K_{m+1}+1}^{K_j} X_{Tt}(\Sigma^0)^{-1} u_t) - 2\sum_{j=2}^m (\beta^0_{(m+1)} - \beta^0_{(1)})'((\sum_{t=K_{j-1}+1}^{K_j} X_{Tt}(\Sigma^0)^{-1} u_t)
\]

A-13
\[= \sum_{j=1}^{m} K_{j} \beta_{j+1} - \sum_{t=K_{j}+1}^{m} X_{Tt}(\Sigma_{0})^{-1} u_{t} \]

Hence,

\[r_{T}^{2}(K_{1}, ..., K_{m}, \beta_{j}, \Sigma_{j}) = \sum_{t=K_{j}+1}^{m} X_{Tt}(\Sigma_{0})^{-1} u_{t} + o_{p}(1)\]

**Case 2:** \(K_{1}^{0} < ... < K_{b}^{0} < ... < K_{m}^{0}\). We first have,

\[B.1 = \sum_{j=2}^{b} X_{Tt}(\Sigma_{j})^{-1} (Y_{t} - X_{Tt}^{0} \beta_{(1)})^{0}(\Sigma_{(1)})^{-1} (Y_{t} - X_{Tt}^{0} \beta_{(1)})^{-1} u_{t} + o_{p}(1)\]

Using arguments similar to those for Case 1, we get

\[r_{T}^{2}(K_{1}, ..., K_{m}, \beta_{j}, \Sigma_{j}) = \sum_{t=K_{j}+1}^{m} X_{Tt}(\Sigma_{0})^{-1} u_{t} + o_{p}(1)\]

**Case 3:** \(K_{1} < ... < K_{b} < ... < K_{m}^{0} < K_{b+1} < ... < K_{m}\). In this case,

\[B.1 = \sum_{j=b+1}^{m} X_{Tt}(\Sigma_{j})^{-1} (Y_{t} - X_{Tt}^{0} \beta_{(1)})^{0}(\Sigma_{(1)})^{-1} (Y_{t} - X_{Tt}^{0} \beta_{(1)})^{-1} u_{t} + o_{p}(1)\]
so that
\[
rbr_t^2(K_1, \ldots, K_m, \beta^{(j)}, \Sigma^{(j)}) = -(1/2) \sum_{j=1}^{b} (\beta^{(j)}_{j+1} - \beta^{(j)}_0)' (\sum_{t=K+1}^{K\theta} X_{T_t} (\Sigma^{(j)})^{-1} X_{T_t}') (\beta^{(j)}_{j+1} - \beta^{(j)}_0) \\
- (1/2) \sum_{j=1}^{b} (\beta^{(j)}_{j+1} - \beta^{(j)}_0)' (\sum_{t=K+1}^{K\theta} X_{T_t} (\Sigma^{(j)})^{-1} X_{T_t}' t) (\beta^{(j)}_{j+1} - \beta^{(j)}_0) \\
- \sum_{j=1}^{b} (\beta^{(j)}_{j+1} - \beta^{(j)}_0)' (\sum_{t=K+1}^{K\theta} X_{T_t} (\Sigma^{(j)})^{-1} X_{T_t}' t) (\beta^{(j)}_{j+1} - \beta^{(j)}_0) \\
+ \sum_{j=1}^{b} (\beta^{(j)}_{j+1} - \beta^{(j)}_0)' (\sum_{t=K+1}^{K\theta} X_{T_t} (\Sigma^{(j)})^{-1} u_t) \\
- \sum_{j=1}^{b} (\beta^{(j)}_{j+1} - \beta^{(j)}_0)' (\sum_{t=K+1}^{K\theta} X_{T_t} (\Sigma^{(j)})^{-1} u_t) + o_p(1)
\]

Proof of Lemma 1: For the case with \( s_j > 0 \), we have
\[
v_T^2 \sum_{t=K+1}^{K\theta} X_{T_t} (\Sigma^{(0)})^{-1} X_{T_t}' = v_T^2 \sum_{t=K+1}^{K\theta} S'(h_{TT} \otimes I_n) (\Sigma^{(0)})^{-1} (h_{TT} \otimes I_n) S
\]
\[
n_{TT} h_{TT}' = \begin{pmatrix}
T^{-1} z_t z_t' & T^{-3/2} t z_t & T^{-1/2} z_t x_t' \\
T^{-1} t z_t' & T^{-2} t t & T^{-1} t x_t' \\
T^{-1/2} z_t' x_t & T^{-1} t x_t & x_t x_t'
\end{pmatrix}
\]
The limit distribution for each component is as follows:
\[
v_T^2 \sum_{t=K+1}^{K\theta} X_{T_t} (\Sigma^{(0)})^{-1} X_{T_t}' = T^{-3/2} v_T^2 \sum_{t=K+1}^{K\theta} (t - K_0 + K_0') (z_t - z_{K_0} + z_{K_0'}) + o_p(1)
\]
\[
v_T^2 \sum_{t=K+1}^{K\theta} t^{-2/3} z_t = T^{-2/3} v_T^2 \sum_{t=K+1}^{K\theta} (t - K_0 + K_0') (z_t - z_{K_0} + z_{K_0'}) + o_p(1)
\]
\[
v_T^2 \sum_{t=K+1}^{K\theta} t^{-2} t = v_T^2 \sum_{t=K+1}^{K\theta} (t - K_0 + K_0')^2 + o_p(1) + s_j T^{-2} (K_0')^2 \Rightarrow s_j (\lambda_0')^2
\]
\[
v_T^2 \sum_{t=K+1}^{K\theta} t^{-1} x_t = v_T^2 \sum_{t=K+1}^{K\theta} (t - K_0 + K_0') x_t' + o_p(1) + T^{-1} K_0' x_t' \Rightarrow s_j (\lambda_0')^2 x_t'
\]
\[
v_T^2 \sum_{t=K+1}^{K\theta} x_t x_t' = s_j Q_{x,t}
\]
Let
\[
D(\lambda_0) = \begin{pmatrix}
\Omega_z^{1/2} W_z (\lambda_0)^{1/2} & (\lambda_0')^2 & \lambda_0 W_z (\lambda_0)^{1/2} & \Omega_z^{1/2} W_z (\lambda_0)^{1/2} & \mu W_z (\lambda_0)^{1/2} \\
(\lambda_0') W_z (\lambda_0)^{1/2} & \Omega_z^{1/2} & \mu W_z (\lambda_0)^{1/2} & \Omega_z^{1/2} W_z (\lambda_0)^{1/2} & \mu W_z (\lambda_0)^{1/2} \\
\lambda_0 W_z (\lambda_0)^{1/2} & \Omega_z^{1/2} W_z (\lambda_0)^{1/2} & \mu W_z (\lambda_0)^{1/2} & \Omega_z^{1/2} W_z (\lambda_0)^{1/2} & \mu W_z (\lambda_0)^{1/2} \\
\mu W_z (\lambda_0)^{1/2} & (\lambda_0') W_z (\lambda_0)^{1/2} & \mu W_z (\lambda_0)^{1/2} & \Omega_z^{1/2} W_z (\lambda_0)^{1/2} & \mu W_z (\lambda_0)^{1/2} \\
\mu W_z (\lambda_0)^{1/2} & (\lambda_0') W_z (\lambda_0)^{1/2} & \mu W_z (\lambda_0)^{1/2} & \mu W_z (\lambda_0)^{1/2} & \mu W_z (\lambda_0)^{1/2}
\end{pmatrix}
\]
\[
\begin{align*}
&v_T^2 \sum_{t=K_j^0+1}^{K_j^0+|s_jv_j^2|} h_{tT} h'_{tT} \Rightarrow s_j D(\lambda_j^0) \\
\text{and} \\
v_T^2 \sum_{t=K_j^0+1}^{K_j^0+|s_jv_j^2|} X_{Tt}(\Sigma_{(l)}^0)^{-1} X_{Tt} \Rightarrow S'(I_q \otimes (\Sigma_{(l)}^0)^{-1})(s_j D(\lambda_j^0) \otimes I_n)S
\end{align*}
\]

Consider now

\[
v_T \sum_{t=K_j^0+1}^{K_j^0+|s_jv_j^2|} X_{Tt}(\Sigma_{(l)}^0)^{-1} U_t = v_T \sum_{t=K_j^0+1}^{K_j^0+|s_jv_j^2|} S'(h_{tT} \otimes I_n)(\Sigma_{(l)}^0)^{-1}(\Sigma_{(j+1)}^0)^{1/2} \eta_t
\]

\[
= S'(I_q \otimes (\Sigma_{(l)}^0)^{-1}(\Sigma_{(j+1)}^0)^{1/2})[v_T \sum_{t=K_j^0+1}^{K_j^0+|s_jv_j^2|} h_{tT} \otimes \eta_t]
\]

Note that \(h_{tT} \otimes \eta_t = (T^{-1/2} z_t \otimes \eta_t, T^{-1} t_{tT}, x_t \otimes \eta_t)'\) with the limit for each component given by

\[
v_T \sum_{t=K_j^0+1}^{K_j^0+|s_jv_j^2|} T^{-1/2} z_t \otimes \eta_t = v_T \sum_{t=K_j^0+1}^{K_j^0+|s_jv_j^2|} T^{-1/2}(z_t - z_{K_j^0} + z_{K_j^0}) \otimes \eta_t
\]

\[
= o_p(1) + v_T \sum_{t=K_j^0+1}^{K_j^0+|s_jv_j^2|} T^{-1/2} z_{K_j^0} \otimes \eta_t \Rightarrow \Omega_{1/2} W_z(\lambda_j^0) \otimes W_{\eta,j}(s_j)
\]

\[
v_T \sum_{t=K_j^0+1}^{K_j^0+|s_jv_j^2|} T^{-1} t_{tT} \otimes \eta_t = v_T \sum_{t=K_j^0+1}^{K_j^0+|s_jv_j^2|} T^{-1}(t - K_j^0 + K_j^0) \eta_t
\]

\[
= o_p(1) + T^{-1} K_j^0 - v_T \sum_{t=K_j^0+1}^{K_j^0+|s_jv_j^2|} \eta_t \Rightarrow \lambda_j^0 W_{\eta,j}(s_j)
\]

\[
v_T \sum_{t=K_j^0+1}^{K_j^0+|s_jv_j^2|} x_t \otimes \eta_t \Rightarrow M_{x_{\eta,j}} W_{x_{\eta,j}}(s_j)
\]

Hence,

\[
v_T \sum_{t=K_j^0+1}^{K_j^0+|s_jv_j^2|} h_{tT} \otimes \eta_t \Rightarrow \begin{pmatrix} \Omega_{1/2} W_z(\lambda_j^0) \otimes W_{\eta,j}(s_j) \\ \lambda_j^0 W_{\eta,j}(s_j) \\ M_{x_{\eta,j}} W_{x_{\eta,j}}(s_j) \end{pmatrix} = \begin{pmatrix} \Omega_{1/2} W_z(\lambda_j^0) \otimes (\Sigma_{(j)}^0)^{-1/2}(\Omega_{\eta(j)})^{1/2} & 0 \\ \lambda_j^0(\Sigma_{(j)}^0)^{-1/2}(\Omega_{\eta(j)})^{1/2} & 0 \\ 0 & M_{x_{\eta,j}} \end{pmatrix} \begin{pmatrix} W_{\eta,j}(s_j) \\ W_{x_{\eta,j}}(s_j) \end{pmatrix}
\]

and

\[
v_T \sum_{t=K_j^0+1}^{K_j^0+|s_jv_j^2|} X_{Tt}(\Sigma_{(l)}^0)^{-1} U_t \Rightarrow S'(I_q \otimes (\Sigma_{(l)}^0)^{-1}(\Sigma_{(j+1)}^0)^{1/2})\Omega(\lambda_j^0) W_j(s_j)
\]

**Proof of Theorem 3** (m = 2, breaks in coefficients only). Let \(Q(\lambda_j^0) = S'(I_q \otimes (\Sigma_{(l)}^0)^{-1})(D(\lambda_j^0) \otimes I_n)S\) and \(\Delta(\lambda_j^0) = S'(I_q \otimes (\Sigma_{(l)}^0)^{-1/2}\Omega(\lambda_j^0)\), then Case 1 (\(K_1 < K_2 \leq K_1^0 < K_2^0\)):

\[
rlr_T^2(K_1, K_2) \Rightarrow H^1(s_1, s_2) = -(1/2)|s_1|\delta_1^0 Q(\lambda_1^0) \delta_1 + \delta_1^0 \Delta(\lambda_1^0) W_1(s_1)
\]

\[
- (1/2)|s_2|\delta_2^0 Q(\lambda_2^0) \delta_2 + \delta_2^0 \Delta(\lambda_2^0) W_1(s_2) - |s_2|\delta_2^0 Q(\lambda_2^0) \delta_2
\]

A-16
Case 2 \((K_1 \leq K_1^0 < K_2^0 \leq K_2):\)
\[
rl_{r,T}^2(K_1, K_2) \Rightarrow H^2(s_1, s_2) = -(1/2)|s_1|\delta_1' Q(\lambda_1^0)\delta_1 + \delta_1' \Delta(\lambda_1^0)W_1(s_1) \\
-(1/2)|s_2|\delta_2' Q(\lambda_2^0)\delta_2 - \delta_2' \Delta(\lambda_2^0)W_2(s_2)
\]

Case 3 \((K_1^0 < K_2^0 \leq K_1 < K_2):\)
\[
rl_{r,T}^2(K_1, K_2) \Rightarrow H^3(s_1, s_2) = -(1/2)|s_1|\delta_1' Q(\lambda_1^0)\delta_1 - \delta_1' \Delta(\lambda_1^0)W_1(s_1) \\
-(1/2)|s_2|\delta_2' Q(\lambda_2^0)\delta_2 - \delta_2' \Delta(\lambda_2^0)W_2(s_2) - |s_2|\delta_1' Q(\lambda_1^0)\delta_2
\]

Note that
\[
\delta_1' \Delta(\lambda_1^0)W_1(s_1) = (\delta_1' \Delta(\lambda_1^0)\Delta(\lambda_1^0)\delta_1)^{1/2} B_1(s_1) \\
\delta_2' \Delta(\lambda_2^0)W_2(s_2) = (\delta_2' \Delta(\lambda_2^0)\Delta(\lambda_2^0)\delta_2)^{1/2} B_2(s_2)
\]
and let \(\Pi_1 = \delta_1' Q(\lambda_1^0)\delta_1, \Pi_2 = \delta_2' Q(\lambda_2^0)\delta_2, \Pi_{12} = \delta_1' Q(\lambda_1^0)\delta_2, \Pi_{21} = \delta_1' Q(\lambda_1^0)\delta_1, \Pi_{22} = \delta_1' Q(\lambda_1^0)\delta_2, \Pi_{12} = \delta_1' \Delta(\lambda_1^0)\Delta(\lambda_1^0)\delta_1\)
and \(\Pi_{22} = \delta_2' \Delta(\lambda_2^0)\Delta(\lambda_2^0)\delta_2, \delta_1' \Delta(\lambda_1^0)\Delta(\lambda_1^0)\delta_1\)

\[
H^1(s_1, s_2) = -(1/2)|s_1|\Pi_1 + \Pi_{11}^{1/2} B_1(s_1) - (1/2)|s_2|\Pi_2 + \Pi_{22}^{1/2} B_2(s_2) - |s_2|\Pi_{12}
\]
\[
H^2(s_1, s_2) = -(1/2)|s_1|\Pi_1 + \Pi_{11}^{1/2} B_1(s_1) - (1/2)|s_2|\Pi_2 - \Pi_{22}^{1/2} B_2(s_2)
\]
\[
H^3(s_1, s_2) = -(1/2)|s_1|\Pi_1 - \Pi_{11}^{1/2} B_1(s_1) - (1/2)|s_2|\Pi_2 + \Pi_{22}^{1/2} B_2(s_2) - |s_1|\Pi_{12}
\]

Let \(b = \frac{\Pi_1}{\Pi_2}, s_1 = bv_1\) and \(s_2 = bv_2\). Applying a changing variables technique as in Bai (1997), we obtain the following results. Case 1 \((v_1 \leq v_2 \leq 0):\)

\[
\Pi_1 v_{r,T}^2(\hat{K}_1 - K_1^0, \hat{K}_2 - K_2^0) = \arg \max_{v_1 \leq v_2 \leq 0} -(1/2)|v_1| + (\frac{\Pi_1}{\Pi_2})^{1/2} B_1(v_1) - (1/2)|v_2| \frac{\Pi_2}{\Pi_1} + \frac{\Pi_{12}^2}{\Pi_1}\frac{\Pi_{21}^2}{\Pi_2} \frac{\Pi_{22}^2}{\Pi_1}
\]

Case 2 \((v_1 < 0, v_2 > 0):\)

\[
\Pi_1 v_{r,T}^2(\hat{K}_1 - K_1^0, \hat{K}_2 - K_2^0) = \arg \max_{v_1 < 0, v_2 > 0} -(1/2)|v_1| + B_1(v_1) - (1/2)|v_2| \frac{\Pi_2}{\Pi_1} - (\frac{\Pi_{22}^2}{\Pi_1})^{1/2} B_2(v_2)
\]

Case 3 \((0 \geq v_1 \geq v_2):\)

\[
\Pi_1 v_{r,T}^2(\hat{K}_1 - K_1^0, \hat{K}_2 - K_2^0) = \arg \max_{0 \geq v_1 \geq v_2} -(1/2)|v_1| - B_1(v_1) - (1/2)|v_2| \frac{\Pi_2}{\Pi_1} - (\frac{\Pi_{22}^2}{\Pi_1})^{1/2} B_2(v_2) - |v_1| \frac{\Pi_{12}}{\Pi_1}
\]

Proof of Theorem 3 (multiple breaks in coefficients only): For Case 1 \((K_1 < ... < K_m \leq K_1^0 < ... < K_m^0):\)
\[
rl_{r,T}^2(K_1, ..., K_m) \Rightarrow H^1(s_1, ..., s_m) = -(1/2) \sum_{j=1}^{m} s_j |s_j|\delta_j' Q(\lambda_j^0)\delta_j \\
- \sum_{j=1}^{m-1} \sum_{i=j+1}^{m} s_i |s_i|\delta_j' Q(\lambda_j^0)\delta_i + \sum_{j=1}^{m} \delta_j' \Delta(\lambda_j^0)W_j(s_j)
\]
For Case 2 ($K_1^0 < \ldots < K_m^0 \leq K_1 < \ldots < K_m$):
\[
rlr_2^2(K_1, \ldots, K_m) \Rightarrow H^2(s_1, \ldots, s_m) = -(1/2) \sum_{j=1}^{m} |s_j| \delta_j^0 Q(\lambda_j^0) \delta_j - \sum_{j=1}^{m-1} \sum_{i=j+1}^{m} |s_j| \delta_j^0 Q(\lambda_i^0) \delta_i - \sum_{j=1}^{m} \delta_j^0 \Delta(\lambda_j^0) W_j(s_j)
\]

For Case 3 ($K_1 < \ldots < K_b \leq K_1^0 < \ldots < K_m^0 < K_{b+1} < \ldots < K_m$):
\[
rlr_2^2(K_1, \ldots, K_m) \Rightarrow H^3(s_1, \ldots, s_m) = -(1/2) \sum_{j=1}^{b} |s_j| \delta_j^0 Q(\lambda_j^0) \delta_j - \sum_{j=1}^{b-1} \sum_{i=j+1}^{b} |s_j| \delta_j^0 Q(\lambda_i^0) \delta_i + \sum_{j=1}^{b} \delta_j^0 \Delta(\lambda_j^0) W_j(s_j) - (1/2) \sum_{j=b+1}^{m} |s_j| \delta_j^0 Q(\lambda_j^0) \delta_j - \sum_{j=b+1}^{m} \delta_j^0 \Delta(\lambda_j^0) W_j(s_j)
\]

Let $\Pi_j = \delta_j^0 Q(\lambda_j^0) \delta_j$, $\Upsilon_j = \delta_j^0 \Delta(\lambda_j^0) \Delta(\lambda_j^0) \delta_j$, $\Pi_j^i = \delta_i^0 Q(\lambda_i^0) \delta_j$, $\Pi_j^i = \delta_i^0 Q(\lambda_i^0) \delta_j$, $b = \Upsilon_1/\Pi_1^2$ and $s_j = bv_j$. We then have the following results. For Case 1 ($v_1 \leq \ldots \leq v_m \leq 0$):
\[
\Pi_1 v_1^2(K_1 - K_1^0, \ldots, K_m - K_m^0) \Rightarrow \arg \max_{v_1 \leq \ldots \leq v_m \leq 0} -(1/2) \sum_{j=1}^{m} |v_j| \frac{\Pi_j}{\Pi_1} + \sum_{j=1}^{m} \left( \frac{\Upsilon_j}{\Pi_1} \right)^{1/2} B_j(v_j) - \sum_{j=1}^{m-1} \sum_{i=j+1}^{m} |v_i| \frac{\Pi_j^i}{\Pi_1}
\]

For Case 2 ($0 \leq v_1 \leq \ldots \leq v_m$):
\[
\Pi_1 v_1^2(K_1 - K_1^0, \ldots, K_m - K_m^0) \Rightarrow \arg \max_{0 \leq v_1 \leq \ldots \leq v_m} -(1/2) \sum_{j=1}^{m} |v_j| \frac{\Pi_j}{\Pi_1} - \sum_{j=1}^{m-1} \sum_{i=j+1}^{m} |v_i| \frac{\Pi_j^i}{\Pi_1}
\]

For Case 3 ($v_1 \leq \ldots \leq v_b \leq 0 \leq v_{b+1} \leq \ldots \leq v_m$):
\[
\Pi_1 v_1^2(K_1 - K_1^0, \ldots, K_m - K_m^0) \Rightarrow \arg \max_{v_1 \leq \ldots \leq v_b \leq 0 \leq v_{b+1} \leq \ldots \leq v_m} -(1/2) \sum_{j=1}^{m} |v_j| \frac{\Pi_j}{\Pi_1} - \sum_{j=1}^{b-1} \sum_{i=j+1}^{b} |v_i| \frac{\Pi_j^i}{\Pi_1} - \sum_{j=b+1}^{m} \left( \frac{\Upsilon_j}{\Pi_1} \right)^{1/2} B_j(v_j) - \sum_{j=b+1}^{m} \sum_{i=j+1}^{m} |v_i| \frac{\Pi_j^i}{\Pi_1}
\]

A-18