2013

Inference methods for locally ordered and common breaks in multiple regressions

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

INFERENCE METHODS FOR LOCALLY ORDERED AND COMMON BREAKS IN MULTIPLE REGRESSIONS

by

YE LI

B.A., Jilin University, 2003
M.A., Jilin University, 2006

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2013

Approved by

First Reader ____

Pierre Perron, Ph.D.
Professor of Economics

Second Reader ____

Zhongjun Qu, Ph.D.
Assistant Professor of Economics

Third Reader ____

Ivan Fernandez-Val, Ph.D.
Associate Professor of Economics

Acknowledgments

I would like to express my immense gratitude to my advisor, Professor Pierre Perron. Without your constant guidance and support, the work would not have been possible, and I would not have finished my Ph.D. study smoothly. Your wide knowledge in econometrics, great effort and work for me, and generous support encouraged and helped me survive the tough years during my graduate study. I am never the same after meeting with you. Your advice not only enhances my understanding of econometrics but also shapes me to be a better academic scholar.

I also wish to thank to my advisors Zhongjun Qu, Ivan Fernandez and Hiroaki kaido. Thank you for your valuable suggestions and comments on my research, and thank you for giving me so much time and help.

My colleagues at Boston University made this dissertation feasible and made my graduate studies memorable. I appreciate the help from my classmates Jie Hou and Tszkin Chen. Thank you for sharing your econometrics knowledge and skills with me. I wish to thank my classmate Chun Wing Tse as well. Thanks for cheering me up when I am frustrated, and thanks for the detailed comments and suggestions to help me to go through the tough situations all these years. I enjoyed the friendship with Saori Chiba, Yanfei Wang, Jiawen Xu and econometrics group members. All of them made my Ph.D. life colorful and unforgetable.

Last but not least, I thank my parents for the their never ending support to me, which makes my every step of growth possible.

The entire journey of Ph.D. life is full of hard work, a lot of times, joys and tears. I appreciate every person who came into my life during this time. I appreciate every experience I went through during this period. I believe it is the unique and the most valuable chapter in my life. May I once again thank my advisor professor Pierre Perron sincerely.

INFERENCE METHODS FOR LOCALLY ORDERED AND COMMON BREAKS IN MULTIPLE REGRESSIONS

(Order No.)

YE LI

Boston University, Graduate School of Arts and Sciences, 2013

Major Professor: Pierre Perron, Professor of Economics

ABSTRACT

This dissertation consists of two chapters related to inference about locally ordered and common breaks in multiple regressions and one chapter pertains to modeling exchange rate volatility with random level shifts.

The first chapter considers inference about locally ordered breaks in a system of equations. These apply when break dates in different equations are not separated by a positive fraction of the sample size. We extend the results of Qu and Perron (2007) in several directions allowing: a) the covariates to be any mix of trends, stationary or integrated regressors; b) breaks in the variance-covariance matrix of errors; c) an arbitrary number of breaks occurring in a subset of equations. We show that the limit distributions derived provide good approximations to the finite sample distributions and forming confidence intervals in a joint fashion allows more precision.

The second chapter considers testing for common breaks and estimating locally ordered breaks in a multiple system with joined segmented trends. To test for common breaks, we consider a likelihood ratio type test. The null hypothesis is that some subsets of coefficients for slope shifts share some common breaks, while the alternative hypothesis is that the breaks dates are different and possibly locally ordered. To estimate locally ordered breaks, we use a quasi-maximum likelihood estimation method. We show consistency and derive rate of convergence and asymptotic distributions of test statistics and estimates of the break dates. Simulation results show that the asymptotic results provide useful approximations in finite samples.

In the third chapter, we estimate a random level shifts model for the log-absolute returns of

the Dollar-Mark and Dollar-Yen exchange rates, in order to assess whether random level shifts can explain the long memory property. The results show that there are few level shifts, but once they are taken into account, the long memory property disappears. We also provide out-of-sample forecasting comparisons, which show that the random level shifts model outperforms standard fractionally integrated models.

Contents

[1] **This chapter is joint work with Professor Pierre Perron**
[2] **This chapter is joint work with Professor Pierre Perron**
[3] **This chapter is joint work with Professor Pierre Perron**

List of Tables

List of Figures

List of Abbreviations

ARFIMA	Autoregressive Fractional Integration Moving Average
DGP	Data Generating Process
GLS	Genenralized Least Square
MCS	Model Confidence Set
MSFE	Mean Square Forecast Error
RLS	Random Level Shift

Chapter 1

Inference on Locally Ordered Breaks in Multiple Regressions[1]

1.1 Introduction

Issues related to structure breaks have received a lot of attention in 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 structure changes, and Bai and Perron (2003) present an efficient algorithm to obtain the break date estimates, which minimize 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 (2008, 2010) consider issues related to estimation and testing for multiple structure breaks in a single cointegrating equation.

Work related to structural changes in a multiple equation 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 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,

[1] **This chapter is joint work with Professor Pierre Perron**

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 an arbitrary number of 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.

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. Appendix A contains the theoretical derivations, while Appendix B presents the results for cases involving changes in both the coefficients and the covariance-matrix of the errors.

1.2 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 (2008). We have n equations and T observations excluding the initial conditions if lagged dependent variables are used as regressors. The total number of structural changes in the system is m^*. The break dates are denoted by the m^* vector $\mathcal{T} = (T_1, ..., T_{m^*})$ and we use the convention that $T_0 = 0$ and $T_{m^*+1} = T$. A subscript j indexes a regime ($j = 1, ..., m^* + 1$). A subscript t indexes a temporal observation ($t = 1, ..., T$) and a subscript i indexes the equation ($i = 1, ..., n$) to which a scalar dependent variable y_{it} is associated. The parameter q is the number of regressors and $h_{tT} = (h_{1tT}, ..., h_{qtT})'$ is the set that includes the regressors from all equations. Let $y_t = (y_{1t}, ..., y_{nt})'$ and $u_t = (u_{1t}, ..., u_{nt})'$, the model considered is

$$y_t = (h'_{tT} \otimes I_n)S\beta_{(j)} + u_t, \tag{1.1}$$

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 $T_{j-1} + 1 \leq t \leq T_j$ ($j = 1, ..., m^* + 1$). 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

$$z_t = z_{t-1} + u_{zt}, \quad (1.2)$$

$$x_t = \mu_x + u_{xt}, \quad (1.3)$$

where z_0 is assumed, for simplicity, to be 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.

The set of basic parameters in regime j consists of the p vector $\beta_{(j)}$ and the n by n matrix $\Sigma_{(j)}$. The matrix S is of dimension nq by p with full column rank. Though, in principle it is allowed to have entries that are arbitrary constants, it is usually a selection matrix involving elements that are 0 or 1 and, hence, specifies which regressors appear in each equation. We allow for the imposition of a set of r restrictions of the form:

$$g(\beta, vec(\Sigma)) = 0, \quad (1.4)$$

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). To ease notation, define the n by p matrix $X'_{tT} = (h'_{tT} \otimes I_n)S$, so that (1.1) becomes

$$y_t = X'_{tT}\beta_{(j)} + u_t, \quad (1.5)$$

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 $X = (X_{1T}, ..., X_{TT})'$. 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 $\bar{X} = diag(X_1, ..., X_{m^*+1})$ 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 (1.5) 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 $j = 1, \ldots, m^* + 1$.

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

$$0 < \lambda_1^0 < ... < \lambda_l^0 < \lambda_{l+1}^0 \leq ... \leq \lambda_{l+m}^0 < \lambda_{l+m+1}^0 < ... < \lambda_{m^*}^0 < 1. \tag{1.6}$$

This stipulates that the break dates $(T_{l+1}^0, ..., T_{l+m}^0)$ need not be separated by a positive fraction of the sample size T, while the others are. These will be the 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 $(K_1^0, ..., K_m^0) = (T_{l+1}^0, ..., T_{l+m}^0)$. 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 1.1 *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 $(K_1^0, \cdots, K_m^0)$, assumed to each occur in a different equation within a subset of m $(\leq n)$ equations, are said to be locally ordered if $K_1^0 < ... < K_m^0$, with the differences such that 1) $v_T^2(K_s^0 - K_1^0) \leq M_T$ with $M_T \to 0$ as $T \to \infty$; 2) $(K_s^0 - K_1^0)/(\log^2 T) \to \infty$, for $s = 2, ..., m$.*

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).

As a matter of notation, "$\xrightarrow{p}$" denotes convergence in probability, "$\xrightarrow{d}$" convergence in distribution and "$\Rightarrow$" weak convergence in the space $D[0,1]$ 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 1.1 *Let $H = (h_{1T}, \cdots, h_{TT})'$ and $\bar{H}_0$ be the diagonal partition of H at $(T_1^0, \cdots, T_{m^*}^0)$ such that $\bar{H}_0 = diag(H_1^0, ..., H_{m^*+1}^0)$. For each $j = 1, \cdots, m^*+1$, $T^{-1}H_j^{0\prime}H_j^0$ converges to a (possibly) random matrix, not necessarily the same for all j.*

Assumption 1.2 *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 $A_{jl}^* = (1/l)\sum_{t=K_j^0-l}^{K_j^0} h_{tT}h_{tT}'$ are bounded away from zero, for $j = 1, ..., m^*$.*

Assumption 1.3 *The matrix $B_{lk} = \sum_{t=k}^{l} h_{tT}h_{tT}'$ is invertible for $l - k \geq p$.*

Assumption 1.4 *Let $\mathcal{F}_t = \sigma-$field $\{..., \zeta_{t-1}, \zeta_t, ..., u_{t-2}, u_{t-1}\}$. If ζ_t is weakly stationary within each segment, then (a) $\{\zeta_t, \mathcal{F}_t\}$ forms a strongly mixing (α-mixing) sequence with size $-4r/(r-2)$ for some $2 < r < 8$. (b) $E(\zeta_t) = 0$ and $\sup_t ||\zeta_t||_{2r+\delta} < M < \infty$ for some $\delta > 0$ and $M > 0$. (c) Let $S_{k,j}(l) = \sum_{T_{j-1}^0+l+1}^{T_{j-1}^0+l+k}(\zeta_t \otimes u_t)$, $j = 1, ..., m^*+1$, for each $e \in R^n$ of length 1, $var(< e, S_{k,j}(0) >) \geq v(k)$ for some function $v(k) \to \infty$ as $k \to \infty$ (with $<\cdot>$, the usual inner product). If ζ_t is not weakly stationary within each segment, we assume that (a)-(c) hold and, in addition, that there exists a positive definite matrix $\Omega = [\varpi_{i,s}]$ such that for any $i, s = 1, ..., p$, we have, uniformly in l, $|k^{-1}E((S_{k,j}(l))_i\ (S_{k,j}(l))_s)| \leq C_2 k^{-\psi}$, 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 1.5 $E(u_{xt} \otimes u_t) = 0$.

Assumption 1.6 *For $j = 1, ..., m^*+1$, $\beta_{(j+1)}^0 - \beta_{(j)}^0 = v_T\delta_{(j)}$ and $\Sigma_{(j+1)}^0 - \Sigma_{(j)}^0 = v_T\Phi_{(j)}$ for some $\delta_{(j)}$ and $\Phi_{(j)}$ independent of T, with $v_T \to 0$ and $T^{1/2}v_T/(\log^2 T) \to \infty$ as $T \to \infty$.*

Assumption 1.7 *Break fractions $(\lambda_1^0, ..., \lambda_{m^*}^0)$ satisfy (1.6) with $(K_1^0, ..., K_m^0) = (T_{j+1}^0, ..., T_{j+m}^0)$ being locally ordered as stated in Definition 1.*

Assumption 1.8 $T^{-1/2}\sum_{t=1}^{[Tr]} u_{zt} \Rightarrow \Omega_z^{1/2}W_z(r)$ *uniformly in $r \in [0,1]$ with Ω_z positive definite.*

Assumption 1.9 *For all t and s: a) $E[(u_{xt} \otimes u_t)z_s'] = 0$, b) $E[(u_{xt} \otimes u_t)u_s'] = 0$, c) $E[(u_{xt} \otimes u_t)u_{xs}'] = 0$, d) $E[\eta_{tk}\eta_{tl}\eta_{th}] = 0$ for all k, l, h and for every t, e) $E[u_{ztk}\eta_{tl}\eta_{th}] = 0$ 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 ζ_t. In particular, they imply that $\zeta_t \otimes u_t$ and ζ_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 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 (2011), 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.

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 Ω_z 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 $j = 1, ..., m+1$, we have, as $\triangle K_j^0 \to \infty$, uniformly in $s \in [0, 1]$,

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

Also, with $\eta_t = (\Sigma^0_{(j)})^{-1/2} u_t$ for $K^0_{j-1}+1 \le t \le K^0_j$, we have for $j = 1, ..., m+1$,

$$
\begin{array}{lll}
(\triangle K^0_j)^{-1/2} \sum_{t=K^0_{j-1}+1}^{K^0_{j-1}+[s\triangle K^0_j]} \eta_t & \Rightarrow & W^{(1)}_{\eta,j}(s) \quad 0 < s \le 1 \\
(\triangle K^0_j)^{-1/2} \sum_{t=K^0_j+[s\triangle K^0_j]}^{K^0_j} \eta_t & \Rightarrow & W^{(2)}_{\eta,j}(s) \quad -1 \le s < 0
\end{array}
$$

where the weak convergence is in the space $D[0,1]^n$ and $\{W^{(1)}_{\eta,j}(s), W^{(2)}_{\eta,j}(s)\}$ are Brownian motion processes defined on $[0,1]^n$. We define the two-sided Brownian motion $W_{\eta,j}(s)$ satisfying $W_{\eta,j}(s) = W^{(1)}_{\eta,j}(s)$ for $s < 0$, $W_{\eta,j}(s) = 0$ for $s = 0$ and $W_{\eta,j}(s) = W^{(2)}_{\eta,j}(s)$ for $s > 0$. Also

$$
\begin{array}{lll}
(\triangle K^0_j)^{-1/2} \sum_{t=K^0_{j-1}+1}^{K^0_{j-1}+[s\triangle K^0_j]} (\eta_t\eta_t' - I_n) & \Rightarrow & \xi^{(1)}_{(j)}(s) \quad 0 < s \le 1 \\
(\triangle K^0_j)^{-1/2} \sum_{t=K^0_j+[s\triangle K^0_j]}^{K^0_j} (\eta_t\eta_t' - I_n) & \Rightarrow & \xi^{(2)}_{(j)}(s) \quad -1 \le s < 0
\end{array}
$$

where the weak convergence is in the space $D[0,1]^{n^2}$ and where the entries of the $n \times n$ matrices $\xi^{(1)}_{(j)}(s)$ and $\xi^{(2)}_{(j)}(s)$ are Brownian motion processes defined on $[0,1]$. We define the two-sided Brownian motion $\xi_{(j)}(\cdot)$ satisfying $\xi_{(j)}(s) = \xi^{(1)}_{(j)}(s)$ for $s > 0$, $\xi_{(j)}(s) = 0$ for $s = 0$ and $\xi_{(j)}(s) = \xi^{(2)}_{(j)}(s)$ for $s < 0$. We also have

$$
\begin{array}{lll}
(\triangle K^0_j)^{-1/2} \sum_{t=K^0_{j-1}+1}^{K^0_{j-1}+[s\triangle K^0_j]} x_t \otimes u_t & \Rightarrow & M^{1/2}_{x,j} W^{(1)}_{x,j}(s) \quad 0 < s \le 1 \\
(\triangle K^0_j)^{-1/2} \sum_{t=K^0_j+[s\triangle K^0_j]}^{K^0_j} x_t \otimes u_t & \Rightarrow & M^{1/2}_{x,j} W^{(2)}_{x,j}(s) \quad -1 \le s < 0
\end{array}
$$

where

$$
M_{x,j} = \lim_{T\to\infty} var\{(\triangle K^0_j)^{-1/2} \sum_{t=K^0_{j-1}+1}^{K^0_j} x_t \otimes u_t\}
$$

with $W^{(1)}_{x,j}(s)$ and $W^{(2)}_{x,j}(s)$ independent Gaussian processes. We define $W_{x,j}(s)$ as a two-sided standard multivariate Gaussian process such that $W_{x,j}(s) = W^{(1)}_{x,j}(s)$ for $s < 0$, $W_{x,j}(s) = 0$ for $s = 0$ and $W_{x,j}(s) = W^{(2)}_{x,j}(s)$ 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, 2008). 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.

1.3 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). 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 $K_{m+1}^0 = T_{l+m+1}^0$.

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

$$f(y_t|X_{Tt}; \beta_{(j)}, \Sigma_{(j)}) = \frac{1}{(2\pi)^{1/2}|\Sigma_{(j)}|^{1/2}} \exp\{-\frac{1}{2}(y_t - X'_{Tt}\beta_{(j)})'\Sigma_{(j)}^{-1}(y_t - X'_{Tt}\beta_{(j)})\}$$

and the quasi-likelihood ratio is

$$LR_T(K, \beta, \Sigma) = \frac{\prod_{j=1}^{m+1} \prod_{t=K_{j-1}+1}^{K_j} f(y_t|X_{Tt}; \beta_{(j)}, \Sigma_{(j)})}{\prod_{j=1}^{m+1} \prod_{t=K_{j-1}^0+1}^{K_j^0} f(y_t|X_{Tt}; \beta_{(j)}^0, \Sigma_{(j)}^0)}$$

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

$$(K, \beta, \Sigma) = \arg \max_{(K_1,...,K_m,\beta,\Sigma)} rlr_T(K, \beta, \Sigma). \tag{1.7}$$

where the supremum is taken over the set

$$\mathbb{k}_\varepsilon = \{(K_1, ..., K_m) : T_l - K_1 > [T\varepsilon], K_2 - K_1 \geqslant h, ..., K_m - K_{m-1} \geqslant h, T_{l+m+1} - K_m \geq [T\varepsilon]\},$$

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).

1.4 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 1.1 *Let* $\{\hat{K}_j, j = 1, ..., m;\ \hat{\beta}_{(j)}$ *and* $\hat{\Sigma}_{(j)},\ j = 1, ..., m+1\}$ *be defined as the solution to the maximization problem (1.7). Then under Assumptions A1-A9,* $v_T^2(\hat{K}_j - K_j^0) = O_p(1)$, *and* $\sqrt{T}(\hat{\beta}_{(j)} - \beta_{(j)}^0) = O_p(1), \sqrt{T}(\hat{\Sigma}_{(j)} - \Sigma_{(j)}^0) = O_p(1)$.

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. It allows us to analyze the asymptotic distribution of the estimates of the break dates in the following compact neighborhood of the true value:

$$C_M = \{(K, \beta, \Sigma) : v_T^2|K_j - K_j^0| \leqslant M, \sqrt{T}|(\beta_{(j)} - \beta_{(j)}^0)| \leqslant M, \sqrt{T}|(\Sigma_{(j)} - \Sigma_{(j)}^0)| \leqslant M\}.$$

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 trues values. From the definition of locally ordered breaks, the true break date satisfies $v_T^2(K_j^0 - K_{j-1}^0) \leqslant M_T$ 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. This corresponds to the following three cases:

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

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 β^0. The case with changes in both β^0 and the covariance matrix of the errors Σ^0 is more involved and relegated to Appendix B. Note that given the possible cases for the relative positions of the estimates relative to the true break dates, in the maximization problem, we need only consider candidate break dates $(K_1, ..., K_m)$ that satisfy a similar ordering, namely: Case 1: $K_1 < ... < K_m \leq K_1^0 < ... < K_m^0$; Case 2: $K_1^0 < ... < K_m^0 \leq K_1 < ... < K_m$; Case 3: $K_1 < ... < K_b \leq K_1^0 < ... < K_m^0 \leq K_{b+1} < ... < K_m$.

Theorem 1.2 *Under Assumptions A1-A9, with m breaks in β^0 only, we have*

$$\begin{aligned}\max_{(K,\beta,\Sigma)} rlr_T &= \max_{(\beta,\Sigma)\in C_M, K^0} [rlr_T^1(K,\beta,\Sigma) + \lambda' g(\beta, vec(\Sigma))] \\ &\quad + \max_{K\in C_M,(\beta^0,\Sigma^0)} rlr_T^2(K,\beta^0,\Sigma^0) + o_p(1)\end{aligned}$$

where

$$\begin{aligned}rlr_T^1(K^0,\beta,\Sigma) &= -(1/2)[\textstyle\sum_{j=1}^{m+1}\sum_{t=K_{j-1}^0+1}^{K_j^0}(Y_t - X_{Tt}'\beta_{(j)})'(\Sigma)^{-1}(Y_t - X_{Tt}'\beta_{(j)}) \\ &\quad - \textstyle\sum_{j=1}^{m+1}\sum_{t=K_{j-1}^0+1}^{K_j^0}(Y_t - X_{Tt}'\beta_{(j)}^0)'(\Sigma^0)^{-1}(Y_t - X_{Tt}'\beta_{(j)}^0)]\end{aligned}$$

and for Case 1 ($K_1 < ... < K_m \leq K_1^0 < ... < K_m^0$)*:*

$$\begin{aligned}
& rlr_T^2(K_{1,}, ..., K_m, \beta^0, \Sigma^0) \\
= \;& -(1/2)\textstyle\sum_{j=1}^{m}(\beta^0_{(j+1)} - \beta^0_{(j)})'(\sum_{t=K_j+1}^{K_j^0} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(j+1)} - \beta^0_{(j)}) \\
& -\textstyle\sum_{j=1}^{m-1}\sum_{i=j+1}^{m}(\beta^0_{(j+1)} - \beta^0_{(j)})'(\sum_{t=K_i+1}^{K_i^0} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(i+1)} - \beta^0_{(i)}) \\
& +\textstyle\sum_{j=1}^{m}(\beta^0_{(j+1)} - \beta^0_{(j)})'(\sum_{t=K_j+1}^{K_j^0} X_{Tt}(\Sigma^0)^{-1}u_t) + o_p(1)
\end{aligned}$$

For Case 2 ($K_1^0 < ... < K_m^0 \leq K_1 < ... < K_m$)*:*

$$\begin{aligned}
& rlr_T^2(K_{1,...,}K_m, \beta^0, \Sigma^0) \\
= \;& -(1/2)\textstyle\sum_{j=1}^{m}(\beta^0_{(j+1)} - \beta^0_{(j)})'(\sum_{t=K_j^0+1}^{K_j} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(j+1)} - \beta^0_{(j)}) \\
& -\textstyle\sum_{j=1}^{m-1}\sum_{i=j+1}^{m}(\beta^0_{(j+1)} - \beta^0_{(j)})'(\sum_{t=K_j^0+1}^{K_j} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(i+1)} - \beta^0_{(i)}) \\
& -\textstyle\sum_{j=1}^{m}(\beta^0_{(j+1)} - \beta^0_{(j)})'(\sum_{t=K_j^0+1}^{K_j} X_{Tt}(\Sigma^0)^{-1}u_t) + o_p(1)
\end{aligned}$$

and for Case 3 ($K_1 < ... < K_b \leq K_1^0 < ... < K_m^0 \leq K_{b+1} < ... < K_m$, *with* $1 \leq b \leq m$)*:*

$$\begin{aligned}
& rlr_T^2(K_{1,...,}K_m, \beta^0, \Sigma^0) \\
= \;& -(1/2)\textstyle\sum_{j=1}^{b}(\beta^0_{(j+1)} - \beta^0_{(j)})'(\sum_{t=K_j+1}^{K_j^0} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(j+1)} - \beta^0_{(j)}) \\
& -\textstyle\sum_{j=1}^{b-1}\sum_{i=j+1}^{b}(\beta^0_{(j+1)} - \beta^0_{(j)})'(\sum_{t=K_i+1}^{K_i^0} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(i+1)} - \beta^0_{(i)}) \\
& -(1/2)\textstyle\sum_{j=b+1}^{m}(\beta^0_{(j+1)} - \beta^0_{(j)})'(\sum_{t=K_j^0+1}^{K_j} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(j+1)} - \beta^0_{(j)}) \\
& -\textstyle\sum_{j=b+1}^{m-1}\sum_{i=j+1}^{m}(\beta^0_{(j+1)} - \beta^0_{(j)})'(\sum_{t=K_j^0+1}^{K_j} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(i+1)} - \beta^0_{(i)}) \\
& +\textstyle\sum_{j=1}^{b}(\beta^0_{(j+1)} - \beta^0_{(j)})'(\sum_{t=K_j+1}^{K_j^0} X_{Tt}(\Sigma^0)^{-1}u_t) \\
& -\textstyle\sum_{j=b+1}^{m}(\beta^0_{(j+1)} - \beta^0_{(j)})'(\sum_{t=K_j^0+1}^{K_j} X_{Tt}(\Sigma^0)^{-1}u_t) + o_p(1)
\end{aligned}$$

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.

Remark 1.1 *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:*

$$\begin{aligned}
rlr_T^1(K^0, \beta, \Sigma) \;=\; & -(1/2)[\textstyle\sum_{j=1}^{3}\sum_{t=K_{j-1}^0+1}^{K_j^0}(Y_t - X'_{Tt}\beta_{(j)})'(\Sigma)^{-1}(Y_t - X'_{Tt}\beta_{(j)}) \\
& -\textstyle\sum_{j=1}^{3}\sum_{t=K_{j-1}^0+1}^{K_j^0}(Y_t - X'_{Tt}\beta^0_{(j)})'(\Sigma^0)^{-1}(Y_t - X'_{Tt}\beta^0_{(j)})]
\end{aligned}$$

and for Case 1 ($K_1 < K_2 \leq K_1^0 < K_2^0$):

$$\begin{aligned} rlr_T^2(K_1, K_2, \beta^0, \Sigma^0) \;=\; & -(1/2)(\beta^0_{(2)} - \beta^0_{(1)})'(\textstyle\sum_{t=K_1+1}^{K_1^0} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(2)} - \beta^0_{(1)}) \\ & -(1/2)(\beta^0_{(3)} - \beta^0_{(2)})'(\textstyle\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(3)} - \beta^0_{(2)}) \\ & -(\beta^0_{(3)} - \beta^0_{(2)})'(\textstyle\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(2)} - \beta^0_{(1)}) \\ & +(\beta^0_{(2)} - \beta^0_{(1)})'(\textstyle\sum_{t=K_1+1}^{K_1^0} X_{Tt}(\Sigma^0)^{-1}u_t) \\ & +(\beta^0_{(3)} - \beta^0_{(2)})'(\textstyle\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma^0)^{-1}u_t) + o_p(1) \end{aligned}$$

and for Case 2 ($K_1^0 < K_2^0 \leq K_1 < K_2$)

$$\begin{aligned} rlr_T^3(K_1, K_2, \beta^0, \Sigma^0) \;=\; & -(1/2)(\beta^0_{(2)} - \beta^0_{(1)})'(\textstyle\sum_{t=K_1^0+1}^{K_1} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(2)} - \beta^0_{(1)}) \\ & -(1/2)(\beta^0_{(3)} - \beta^0_{(2)})'(\textstyle\sum_{t=K_2^0+1}^{K_2} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(3)} - \beta^0_{(2)}) \\ & -(\beta^0_{(3)} - \beta^0_{(2)})'(\textstyle\sum_{t=K_1^0+1}^{K_1} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(2)} - \beta^0_{(1)}) \\ & -(\beta^0_{(2)} - \beta^0_{(1)})'(\textstyle\sum_{t=K_1^0+1}^{K_1} X_{Tt}(\Sigma^0)^{-1}u_t) \\ & -(\beta^0_{(3)} - \beta^0_{(2)})'(\textstyle\sum_{t=K_2^0+1}^{K_2} X_{Tt}(\Sigma^0)^{-1}u_t) + o_p(1) \end{aligned}$$

while for Case 3 ($K_1 \leq K_1^0 < K_2^0 \leq K_2$)

$$\begin{aligned} rlr_T^2(K_1, K_2, \beta^0, \Sigma^0) \;=\; & -(1/2)(\beta^0_{(2)} - \beta^0_{(1)})'(\textstyle\sum_{t=K_1+1}^{K_1^0} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(2)} - \beta^0_{(1)}) \\ & -(1/2)(\beta^0_{(3)} - \beta^0_{(2)})'(\textstyle\sum_{t=K_2^0+1}^{K_2} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(3)} - \beta^0_{(2)}) \\ & +(\beta^0_{(2)} - \beta^0_{(1)})'(\textstyle\sum_{t=K_1+1}^{K_1^0} X_{Tt}(\Sigma^0)^{-1}u_t) \\ & -(\beta^0_{(3)} - \beta^0_{(2)})'(\textstyle\sum_{t=K_2^0+1}^{K_2} X_{Tt}(\Sigma^0)^{-1}u_t) + o_p(1) \end{aligned}$$

In order to derive the limit distribution of the estimates of the break dates, some additional results are needed. These are stated in the following lemma.

Lemma 1.1 *Under Assumptions A1-A9, we have*

$$\begin{aligned} v_T^2 \textstyle\sum_{t=K_j^0+1}^{K_j^0+[s_j v_T^{-2}]} X_{Tt}(\Sigma^0_{(l)})^{-1}X'_{Tt} \;&\Rightarrow\; S'(s_j D(\lambda_j^0) \otimes (\Sigma^0_{(l)})^{-1})S \qquad s_j > 0 \\ v_T^2 \textstyle\sum_{t=K_j^0+[s_j v_T^{-2}]}^{K_j^0} X_{Tt}(\Sigma^0_{(l)})^{-1}X'_{Tt} \;&\Rightarrow\; S'(|s_j| D(\lambda_j^0) \otimes (\Sigma^0_{(l)})^{-1})S \qquad s_j < 0 \end{aligned}$$

$$\begin{aligned} v_T \textstyle\sum_{t=K_j^0+1}^{K_j^0+[s_j v_T^{-2}]} X_{Tt}(\Sigma^0_{(l)})^{-1}U_t \;&\Rightarrow\; S'(I_q \otimes (\Sigma^0_{(l)})^{-1}(\Sigma^0_{(j+1)})^{1/2})\Omega(\lambda_j^0)W_j(s_j) \quad s_j > 0 \\ v_T \textstyle\sum_{t=K_j^0+[s_j v_T^{-2}]}^{K_j^0} X_{Tt}(\Sigma^0_{(l)})^{-1}U_t \;&\Rightarrow\; S'(I_q \otimes (\Sigma^0_{(l)})^{-1}(\Sigma^0_{(j)})^{1/2})\Omega(\lambda_j^0)W_j(s_j) \quad s_j < 0 \end{aligned}$$

with

$$D(\lambda_j^0) = \begin{pmatrix} \Omega_z^{1/2} W_z(\lambda_j^0) W_z(\lambda_j^0)' \Omega_z^{1/2} & \lambda_j^0 \Omega_z^{1/2} W_z(\lambda_j^0) & \Omega_z^{1/2} W_z(\lambda_j^0) \mu_{x,j}' \\ \lambda_j^0 W_z(\lambda_j^0)' \Omega_z^{1/2} & (\lambda_j^0)^2 & \lambda_j^0 \mu_{x,j}' \\ \mu_{x,j} W_z(\lambda_j^0)' \Omega_z^{1/2} & \lambda_j^0 \mu_{x,j} & Q_{x,j} \end{pmatrix}$$

$$\Omega(\lambda_j^0) = \begin{pmatrix} \Omega_z^{1/2} W_z(\lambda_j^0) \otimes I_n & 0 \\ \lambda_j^0 I_n & 0 \\ 0 & M_{x\eta,j}^{1/2} \end{pmatrix} \quad , \quad W_j(s_j) = \begin{pmatrix} W_{\eta,j}(s_j) \\ W_{x\eta,j}(s_j) \end{pmatrix}$$

where $M_{x\eta,j} = (\Sigma^0)^{-1/2} M_{x,j} (\Sigma^0)^{-1/2}$ *and* $W_{x\eta,j}(\cdot) = (\Sigma^0)^{-1/2} W_{x,j}(\cdot)$.

Remark 1.2 *Without structural breaks in* Σ^0*, Lemma 1 reduces to*

$$\begin{aligned} v_T^2 \textstyle\sum_{t=K_j^0+1}^{K_j^0+[s_j v_T^{-2}]} X_{Tt} (\Sigma^0)^{-1} X_{Tt}' &\Rightarrow S'(s_j D(\lambda_j^0) \otimes (\Sigma^0)^{-1}) S \qquad s_j > 0 \\ v_T^2 \textstyle\sum_{t=K_j^0+[s_j v_T^{-2}]}^{K_j^0} X_{Tt} (\Sigma^0)^{-1} X_{Tt}' &\Rightarrow S'(|s_j| D(\lambda_j^0) \otimes (\Sigma^0)^{-1}) S \qquad s_j < 0 \end{aligned}$$

$$\begin{aligned} v_T \textstyle\sum_{t=K_j^0+1}^{K_j^0+[s_j v_T^{-2}]} X_{Tt} (\Sigma^0)^{-1} U_t &\Rightarrow S'(I_q \otimes (\Sigma^0)^{-1/2}) \Omega(\lambda_j^0) W_j(s_j) \qquad s_j > 0 \\ v_T \textstyle\sum_{t=K_j^0+[s_j v_T^{-2}]}^{K_j^0} X_{Tt} (\Sigma^0)^{-1} U_t &\Rightarrow S'(I_q \otimes (\Sigma^0)^{-1/2}) \Omega(\lambda_j^0) W_j(s_j) \qquad s_j < 0 \end{aligned}$$

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 β^0. The more general case with also changes in Σ^0 is presented in Appendix B.

Theorem 1.3 *Let* $Q(\lambda_j^0) = S'(D(\lambda_j^0) \otimes (\Sigma^0)^{-1}) S$, $\Delta(\lambda_j^0) = S'(I_q \otimes (\Sigma^0)^{-1/2}) \Omega(\lambda_j^0)$, $\Pi_j = \delta_j' Q(\lambda_j^0) \delta_j$, $\Pi_{ij}^i = \delta_i' Q(\lambda_i^0) \delta_j$,$\Pi_{ij}^j = \delta_i' Q(\lambda_j^0) \delta_j$, $\Upsilon_j = \delta_j' \Delta(\lambda_j^0) \Delta(\lambda_j^0)' \delta_j$, $b = \Upsilon_1 / \Pi_1^2$ *and* $s_j = b v_j$. *Under Assumptions A1-A9, with* m *breaks in* β^0 *only, we have:*

$$\Pi_1 v_T^2 (\hat{K}_1 - K_1^0, ..., \hat{K}_m - K_m^0) \Rightarrow \arg\max_{v_1 \leq ... \leq v_m} H(v_1, ..., v_m)$$

where $H(v_1, ..., v_m) = 0$ *if* $v_1 = ... = v_m = 0$*, and for Case 1 with* $v_1 \leq ... \leq v_m \leq 0$*:*

$$H(v_1, ..., v_m) = -\frac{1}{2} \textstyle\sum_{j=1}^m |v_j| \dfrac{\Pi_j}{\Pi_1} + \textstyle\sum_{j=1}^m (\dfrac{\Upsilon_j}{\Pi_1})^{1/2} B_j(v_j) - \textstyle\sum_{j=1}^{m-1} \sum_{i=j+1}^m |v_i| \dfrac{\Pi_{ij}^i}{\Pi_1}$$

For Case 2 with $0 \leq v_1 \leq ... \leq v_m$*:*

$$H(v_1, ..., v_m) = -\frac{1}{2} \textstyle\sum_{j=1}^m |v_j| \dfrac{\Pi_j}{\Pi_1} - \textstyle\sum_{j=1}^m (\dfrac{\Upsilon_j}{\Pi_1})^{1/2} B_j(v_j) - \textstyle\sum_{j=1}^{m-1} \sum_{i=j+1}^m |v_j| \dfrac{\Pi_{ij}}{\Pi_1}$$

and for Case 3 with $v_1 \le ... \le v_b \le 0 \le v_{b+1} \le ... \le v_m$*:*

$$\begin{aligned} H(v_1,...,v_m) &= -\frac{1}{2}\sum_{j=1}^{m}|v_j|\frac{\Pi_j}{\Pi_1} - \sum_{j=1}^{b-1}\sum_{i=j+1}^{b}|v_i|\frac{\Pi_{ij}^{i}}{\Pi_1} - \sum_{j=b+1}^{m-1}\sum_{i=j+1}^{m}|v_j|\frac{\Pi_{ij}^{j}}{\Pi_1} \\ &\quad + \sum_{j=1}^{b}(\frac{\Upsilon_j}{\Pi_1})^{1/2}B_j(v_j) - \sum_{j=b+1}^{m}(\frac{\Upsilon_j}{\Pi_1})^{1/2}B_j(v_j) \end{aligned}$$

Remark 1.3 *For the case with only one pair of locally ordered breaks, Theorem 3 reduces to:*

$$\Pi_1 v_T^2(\hat{K}_1 - K_1^0, \hat{K}_2 - K_2^0) \Rightarrow \arg\max_{v_1 \le v_2} H(v_1, v_2)$$

where $H(v_1, v_2) = 0$ *if* $v_1 = v_2 = 0$*, and for Case 1 with* $v_1 \le v_2 \le 0$*:*

$$H(v_1,v_2) = -\frac{1}{2}|v_1| + (\frac{\Upsilon_1}{\Pi_1})^{1/2}B_1(v_1) - \frac{1}{2}|v_2|\frac{\Pi_2}{\Pi_1} + (\frac{\Upsilon_2}{\Pi_1})^{1/2}B_2(v_2) - |v_2|\frac{\Pi_{12}^2}{\Pi_1},$$

and for Case 2 with $0 \le v_1 \le v_2$*:*

$$H(v_1,v_2) = -\frac{1}{2}|v_1| - (\frac{\Upsilon_1}{\Pi_1})^{1/2}B_1(v_1) - \frac{1}{2}|v_2|\frac{\Pi_2}{\Pi_1} - (\frac{\Upsilon_2}{\Pi_1})^{1/2}B_2(v_2) - |v_1|\frac{\Pi_{12}^1}{\Pi_1},$$

for Case 3 with $v_1 < 0, v_2 > 0$*:*

$$H(v_1,v_2) = -\frac{1}{2}|v_1| + (\frac{\Upsilon_1}{\Pi_1})^{1/2}B_1(v_1) - \frac{1}{2}|v_2|\frac{\Pi_2}{\Pi_1} - (\frac{\Upsilon_2}{\Pi_1})^{1/2}B_2(v_2),$$

The cumulative distribution function does not have an tractable analytical formula. However, the relevant quantiles can be obtained using simulations. First, generate the realizations of H by replacing the true value of 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 $(v_1, v_2, \cdots, v_m) \in [-M, M]$. This is repeated for all possible cases and the overall maximum obtained. These steps are repeated to obtain the relevant quantiles.

Remark 1.4 *For the special cases with a single type of regressors, the limit distribution differs according to the specifications of the matrices* $D(\lambda_j^0)$ *and* $\Omega(\lambda_j^0)$*, and the definitions of* $Q(\lambda_j^0)$*,* $\Delta(\lambda_j^0)$*,* Π_j*,* Π_{ij}^i*,* Π_{ij}^j*, and* Υ_j*. When only stationary regressors are present, we have*

$$D(\lambda_j^0) = Q_{x,j},\ \Omega(\lambda_j^0) = M_{x\eta,j}^{1/2}.$$

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

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

With only integrated regressors,

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

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

$$\begin{aligned} D(\lambda_j^0) &= \begin{pmatrix} \Omega_z^{1/2} W_z(\lambda_j^0) W_z(\lambda_j^0)' \Omega_z^{1/2} & \Omega_z^{1/2} W_z(\lambda_j^0) \mu_{x,j}' \\ \mu_{x,j} W_z(\lambda_j^0)' \Omega_z^{1/2} & Q_{x,j} \end{pmatrix}, \\ \Omega(\lambda_j^0) &= \begin{pmatrix} \Omega_z^{1/2} W_Z(\lambda_j^0) \otimes I_n & 0 \\ 0 & M_{x\eta,j}^{1/2} \end{pmatrix}. \end{aligned}$$

b) with stationary and trend regressors,

$$D(\lambda_j^0) = \begin{pmatrix} (\lambda_j^0)^2 & \lambda_j^0 \mu_{x,j}' \\ \lambda_j^0 \mu_{x,j} & Q_{x,j} \end{pmatrix},\ \Omega(\lambda_j^0) = \begin{pmatrix} \lambda_j^0 I_n & 0 \\ 0 & M_{x\eta,j}^{1/2} \end{pmatrix}.$$

c) with a trend and unit root regressors,

$$\begin{aligned} D(\lambda_j^0) &= \begin{pmatrix} \Omega_z^{1/2} W_z(\lambda_j^0) W_z(\lambda_j^0)' \Omega_z^{1/2} & \lambda_j^0 \Omega_z^{1/2} W_z(\lambda_j^0) \\ \lambda_j^0 W_z(\lambda_j^0)' \Omega_z^{1/2} & (\lambda_j^0)^2 \end{pmatrix}, \\ \Omega(\lambda_j^0) &= \begin{pmatrix} \Omega_z^{1/2} W_z(\lambda_j^0) \otimes I_n \\ \lambda_j^0 I_n \end{pmatrix}. \end{aligned}$$

The results allow us to construct 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 Σ^0. The long-run covariance matrix $M_{x,j}$ 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 $\hat{\delta}_{(j)} = \hat{\beta}_{(j+1)} - \hat{\beta}_{(j)}$.

1.5 Monte Carlo Simulations

In this section, we 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:

Table 1.1: Specification of Data Generating Process

	Equation 1	Equation 2
DGP-1:	$y_{1t} = \beta_{11}^{(j)} x_t + u_{1t}$	$y_{2t} = \beta_{21}^{(j)} x_t + u_{2t}$
DGP-2:	$y_{1t} = \beta_{11}^{(j)} \frac{t}{T} + u_{1t}$	$y_{2t} = \beta_{21}^{(j)} \frac{t}{T} + u_{2t}$
DGP-3:	$y_{1t} = \beta_{11}^{(j)} \frac{z_t}{\sqrt{T}} + u_{1t}$	$y_{2t} = \beta_{21}^{(j)} \frac{z_t}{\sqrt{T}} + u_{2t}$
DGP-4:	$y_{1t} = \beta_{11}^{(j)} x_t + \beta_{12}^{(j)} \frac{t}{T} + +u_{1t}$	$y_{2t} = \beta_{21}^{(j)} x_t + \beta_{22}^{(j)} \frac{t}{T} + u_{2t}$
DGP-5:	$y_{1t} = \beta_{11}^{(j)} x_t + \beta_{12}^{(j)} \frac{z_t}{\sqrt{T}} + u_{1t}$	$y_{2t} = \beta_{21}^{(j)} x_t + \beta_{22}^{(j)} \frac{z_t}{\sqrt{T}} + u_{2t}$
DGP-6:	$y_{1t} = \beta_{11}^{(j)} \frac{t}{T} + \beta_{12}^{(j)} \frac{z_t}{\sqrt{T}} + u_{1t}$	$y_{2t} = \beta_{21}^{(j)} \frac{t}{T} + \beta_{22}^{(j)} \frac{z_t}{\sqrt{T}} + u_{2t}$
DGP-7:	$y_{1t} = \beta_{11}^{(j)} \frac{z_t}{\sqrt{T}} + \beta_{12}^{(j)} \frac{t}{T} + \beta_{13}^{(j)} x_t + u_{1t}$	$y_{2t} = \beta_{21}^{(j)} \frac{z_t}{\sqrt{T}} + \beta_{22}^{(j)} \frac{t}{T} + \beta_{23}^{(j)} x_t + u_{2t}$

for $j = 1, 2$. The values of the coefficients and break dates used are as follows:

Table 1.2: Specification of Data Generating Process

	$\beta_{11}^{(1)}$	$\beta_{11}^{(2)}$	$\beta_{12}^{(1)}$	$\beta_{12}^{(2)}$	$\beta_{13}^{(1)}$	$\beta_{13}^{(2)}$	$\beta_{21}^{(1)}$	$\beta_{21}^{(2)}$	$\beta_{22}^{(1)}$	$\beta_{22}^{(2)}$	$\beta_{23}^{(1)}$	$\beta_{23}^{(2)}$	K_1^0	K_2^0
DGP-1:	1	1.5					1	1.6					80	95
DGP-2:	1	1.7					1	1.8					80	95
DGP-3:	1	2.0					1.5	2.3					80	86
DGP-4:	1	1.5	2	3			2	2.6	3	3.8			80	95
DGP-5:	2	2.5	3	3.7			3	3.6	1.5	2.3			80	95
DGP-6:	1	3	1	1.5			1	2.5	1.5	2.1			80	90
DGP-7:	1	1.5	3	3.3	3	3.6	2	2.6	2	2.2	1.5	2.2	80	95

We set $u_t = (u_{1t}, u_{2t})' \sim i.i.d\ N(0, \Sigma)$ with $var(u_{1t}) = var(u_{2t}) = 1$ 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 1.3: Exact Sizes of the Joint Asymptotic Confidence Intervals

ρ	Nominal Size	DGP1	DGP2	DGP3	DGP4	DGP5	DGP6	DGP7
$\rho = 0$	.90	.88	.92	.93	.85	.88	.92	.87
$\rho = 0$	.95	.92	.97	.95	.91	.93	.94	.92
$\rho = 0.3$	.90	.88	.92	.92	.87	.90	.91	.89
$\rho = 0.3$	.95	.95	1.0	.93	.93	.94	.93	.93

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. 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.

Table 1.4: Joint versus Single Equation Confidence Intervals

	90% CI for K_1	90% CI for K_2	95% CI for K_1	95% CI for K_2
Bai	[74,87]	[90,100]	[70,89]	[88,101]
$\rho = 0$	[74,81]	[93,100]	[71,82]	[92,101]
$\rho = 0.3$	[73,83]	[91,101]	[71,85]	[89,102]

1.6 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.

1.7 Appendix 1

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

$$\limsup_{k\to\infty} (k^{1/2} \log^{1/2} k)^{-1} ||\Omega_k^{-1/2} \textstyle\sum_{i=1}^{k} \zeta_i|| = O_p(1).$$

Lemma 1.3 *Let $\eta_t = (\Sigma^0_{(j)})^{-1/2}u_t$ for $K^0_{j-1} + 1 \leq t \leq K^0_j$. 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*

$$\begin{aligned} for\ s &< 0, v_T \textstyle\sum_{t=K_j^0+[sv_T^{-2}]}^{K_j^0} (\eta_t\eta_t' - I_n) \Rightarrow \xi_j^{(1)}(s) \\ for\ s &> 0, v_T \textstyle\sum_{t=K_j^0+1}^{K_j^0+[sv_T^{-2}]} (\eta_t\eta_t' - I_n) \Rightarrow \xi_j^{(2)}(s) \end{aligned}$$

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

$$\begin{aligned} for\ s &< 0, v_T \textstyle\sum_{t=K_j^0+[sv_T^{-2}]}^{K_j^0} x_t \otimes u_t \Rightarrow M_{x,j}^{1/2} W_{x,j}^{(1)}(s) \\ for\ s &> 0, v_T \textstyle\sum_{t=K_j^0+1}^{K_j^0+[sv_T^{-2}]} x_t \otimes u_t \Rightarrow M_{x,j}^{1/2} W_{x,j}^{(2)}(s) \end{aligned}$$

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 $j = 1, ..., m$.

Lemma 1.4 *Let $(\zeta_t)_{t\geq 1}$ be a sequence of mean zero R^d-valued random vectors satisfying A4, and* *let ζ_t^* stands for ζ_t or $(t/T)\zeta_t$.Then there exists $L < \infty$ such that for any $\varepsilon > 0$ and $m > 0$:*

$$P(\sup_{m \leq k \leq n} d_k || \sum_{t=1}^{k} \zeta_t^* || > \varepsilon) \leq L\varepsilon^{-2}[md_m^2 + \sum_{t=m+1}^{k} d_t^2]$$

Lemma 1.5 *Under A4, we have uniformly over all $0 < r < s < 1$: a) $\sum_{t=[Tr]}^{[Ts]} \zeta_t = O_p(T^{1/2})$, b) $\sum_{t=[Tr]}^{[Ts]} z_t = O_p(T^{3/2})$, c) $\sum_{t=[Tr]}^{[Ts]} z_t z_t' = O_p(T^2)$, d) $\sum_{t=[Tr]}^{[Ts]} z_t \zeta_t' = O_p(T)$.*

Lemma 1.6 *Let $T_1 = [aT]$ for some $a \in (0,1)$ and ζ_t stands for $(h_{tT} \otimes \eta_t)$ or $(\eta_t \eta_t' - I)$. Under Assumptions A1-A4:*

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

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

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

d) there exists a $A > 0$ such that $\sup_{A v_T^{-2} \leq k \leq T} k^{-1} || \sum_{t=T_1+1}^{T_1+k} \zeta_t || = o_p(v_T)$;

e) $\sup_{1 \leq k \leq A v_T^{-2}} k^{-1} || \sum_{t=T_1+1}^{T_1+k} \zeta_t || = (A v_T^{-2})^{1/2} O_p(1)$ for any constant $A > 0$;

f) $\sup_{1 \leq k \leq \sqrt{T} v_T^{-1}} k^{-1} || \sum_{t=T_1+1}^{T_1+k} \zeta_t || = o_p(v_T)$.

Proof of Theorem 1.1: The proof is based on establishing 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

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

where (β, Σ) are generic values of the parameters and (β^0, Σ^0) are the true values. Since

$$\log f(y_t | X_{Tt}, \beta, \Sigma) = -(1/2)\log(2\pi) - (1/2)\log|\Sigma| - (1/2)||\Sigma^{-1/2}(y_t - X_{Tt}'\beta)||^2,$$

the log-likelihood ratio is

$$\begin{aligned} \log LR_k(\beta, \Sigma) &= -(k/2)\log|(\Sigma^0)^{-1/2}\Sigma(\Sigma^0)^{-1/2}| - (1/2)||\Sigma^{-1/2}(y_t - X_{Tt}'\beta)||^2 \\ &\quad +(1/2)||(\Sigma^0)^{-1/2}u_t||^2. \end{aligned}$$

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

$$P(\sup_{\sqrt{T}v_T^{-1}\leq k\leq T} \sup_{S(\beta,\Sigma)^c} \log LR_k(\beta,\Sigma) > -C\sqrt{T}v_T) < \epsilon$$

where $S(\beta,\Sigma)^c = \{(\beta,\Sigma) : ||\beta-\beta^0|| \geq v_T \text{ or } ||\Sigma-\Sigma^0|| \geq v_T\}$.

Property 4. Let $T_1 = [Ta]$ for some $a \in (0,1]$ and let $T_2 = [\sqrt{T}v_T^{-1}]$ where v_T satisfies Assumption A6. Also let

$$\begin{aligned} y_t &= x_t'\beta_{(1)}^0 + (\Sigma_{(1)}^0)^{1/2}\eta_t & t &= 1,...,T_1 \\ y_t &= x_t'\beta_{(2)}^0 + (\Sigma_{(2)}^0)^{1/2}\eta_t & t &= T_1+1,...,T_1+T_2 \end{aligned}$$

where $||\beta_{(2)}^0 - \beta_{(1)}^0|| \leq Mv_T$ and $||\Sigma_{(2)}^0 - \Sigma_{(1)}^0|| \leq Mv_T$ 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 $\sqrt{T}(\beta_{(N)} - \beta_{(1)}^0) = O_p(1), \sqrt{T}(\Sigma_{(N)} - \Sigma_{(1)}^0) = O_p(1)$.

Property 5. With v_T satisfying Assumption A6, for each β and Σ 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:

$$\begin{aligned} \log LR_k(\beta,\Sigma) &= -(k/2)\log|\Sigma| - (1/2)\textstyle\sum_{t=T_{\rho+1}}^{T_\rho+k}(y_t - X_{Tt}'\beta)'\Sigma^{-1}(y_t - X_{Tt}'\beta) \\ &\quad +(k/2)\log|\Sigma^0| - (1/2)\textstyle\sum_{t=T_{\rho+1}}^{T_\rho+k}(y_t - X_{Tt}'\beta^0)'(\Sigma^0)^{-1}(y_t - X_{Tt}'\beta^0) \end{aligned}$$

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

$$\begin{aligned} L_{1T}(\beta,\Sigma) &= -(k/2)\log|I+\Psi_T| - (1/2)\textstyle\sum_{t=T_{\rho+1}}^{T_\rho+k}\eta_t'(I+\Psi_T)^{-1}\eta_t \\ &\quad +(1/2)\textstyle\sum_{t=T_{\rho+1}}^{T_\rho+k}\eta_t'\eta_t \\ L_{2T}(\beta,\Sigma) &= -(1/2)(\beta-\beta^0)'(\textstyle\sum_{t=T_{\rho+1}}^{T_\rho+k} X_{Tt}\Sigma^{-1}X_{Tt}')(\beta-\beta^0) \\ &\quad +(\beta-\beta^0)'(\textstyle\sum_{t=T_{\rho+1}}^{T_\rho+k} X_{Tt}\Sigma^{-1}(\Sigma^0)^{1/2}\eta_t) \end{aligned}$$

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

$$\begin{aligned} & L_{1T}(\beta + T^{-1/2}\lambda, \Sigma + T^{-1/2}\Xi) - L_{1T}(\beta,\Sigma) \\ = \;& -(k/2)[\log|I+\Psi_T^1| - \log|I+\Psi_T^2|] \\ & -(1/2)\textstyle\sum_{t=T_\rho+1}^{T_\rho+k} \eta_t'((I+\Psi_T^1)^{-1} - (I+\Psi_T^2)^{-1})\eta_t \end{aligned}$$

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

$$\begin{aligned} (I+\Psi_T^1)^{-1} - (I+\Psi_T^2)^{-1} &= (A+T^{-1/2}B)^{-1} - A^{-1} \\ &= (I - T^{-1/2}A^{-1}B + T^{-1}A^{-1}BA^{-1}B)A^{-1} - A^{-1} \\ &= -T^{-1/2}A^{-1}BA^{-1} + O_p(T^{-1}) \end{aligned}$$

then

$$\begin{aligned} & -(1/2)\textstyle\sum_{t=T_\rho+1}^{T_\rho+k} \eta_t'((I+\Psi_T^1)^{-1} - (I+\Psi_T^2)^{-1})\eta_t \\ = \;& (1/2)T^{-1/2}tr(A^{-1}BA^{-1}\textstyle\sum_{t=T_\rho+1}^{T_\rho+k}(\eta_t\eta_t' - I)) \\ & +(k/2)T^{-1/2}tr(A^{-1}BA^{-1}) + o_p(1) \\ = \;& (k/2)T^{-1/2}tr(A^{-1}BA^{-1}) + o_p(1). \end{aligned}$$

Given that

$$\begin{aligned} & -(k/2)[\log|I+\Psi_T^1| - \log|I+\Psi_T^2|] \\ = \;& -(k/2)[\log|A+T^{-1/2}B| - \log|A|] \\ = \;& -(k/2)[\log|A| + tr(A^{-1}T^{-1/2}B) - \\ & tr(T^{-1}A^{-1}BA^{-1}B) + o_p(T^{-1}) - \log|A|] \\ = \;& -(k/2)T^{-1/2}tr(A^{-1}B) + o_p(1), \end{aligned}$$

we have

$$\begin{aligned} & L_{1T}(\beta + T^{-1/2}\lambda, \Sigma + T^{-1/2}\Xi) - L_{1T}(\beta,\Sigma) \\ = \;& (k/2)T^{-1/2}[tr(A^{-1}BA^{-1}) - tr(A^{-1}B)] + o_p(1) \end{aligned}$$

$$= \quad -(k/2)T^{-1/2}tr(A^{-1}BA^{-1}(\Sigma^0)^{-1/2}(\Sigma-\Sigma^0)(\Sigma^0)^{-1/2}) + o_p(1)$$
$$= \quad O_p(1)$$

and

$$\begin{aligned}
& L_{2T}(\beta + T^{-1/2}\lambda, \Sigma + T^{-1/2}\Xi) - L_{2T}(\beta, \Sigma) \\
= \; & -(1/2)(\beta + T^{-1/2}\lambda - \beta^0)'(\textstyle\sum_{t=T_{\rho+1}}^{T_\rho+k} X_{Tt}(\Sigma + T^{-1/2}\Xi)^{-1}X'_{Tt})(\beta + T^{-1/2}\lambda - \beta^0) \\
& +(\beta + T^{-1/2}\lambda - \beta^0)'(\textstyle\sum_{t=T_{\rho+1}}^{T_\rho+k} X_{Tt}(\Sigma + T^{-1/2}\Xi)^{-1}(\Sigma^0)^{1/2}\eta_t \\
& +(1/2)(\beta - \beta^0)'(\textstyle\sum_{t=T_{\rho+1}}^{T_\rho+k} X_{Tt}\Sigma^{-1}X'_{Tt})(\beta - \beta^0) \\
& -(\beta - \beta^0)'\textstyle\sum_{t=T_{\rho+1}}^{T_\rho+k} X_{Tt}\Sigma^{-1}(\Sigma^0)^{1/2}\eta_t \\
= \; & -(1/2)(\beta - \beta^0)'(\textstyle\sum_{t=T_{\rho+1}}^{T_\rho+k} X_{Tt}((\Sigma + T^{-1/2}\Xi)^{-1} - \Sigma^{-1})X'_{Tt}) \\
& (\beta + T^{-1/2}\lambda - \beta^0) \\
& +(\beta - \beta^0)'(\textstyle\sum_{t=T_{\rho+1}}^{T_\rho+k} X_{Tt}((\Sigma + T^{-1/2}\Xi)^{-1} - \Sigma^{-1})(\Sigma^0)^{-1/2}\eta_t \\
& -(1/2)T^{-1}\lambda'(\textstyle\sum_{t=T_{\rho+1}}^{T_\rho+k} X_{Tt}(\Sigma + T^{-1/2}\Xi)^{-1}X'_{Tt})\lambda \\
& -T^{-1/2}\lambda'(\textstyle\sum_{t=T_{\rho+1}}^{T_\rho+k} X_{Tt}(\Sigma + T^{-1/2}\Xi)^{-1}X'_{Tt})(\beta - \beta^0) \\
& +T^{-1/2}\lambda'(\textstyle\sum_{t=T_{\rho+1}}^{T_\rho+k} X_{Tt}(\Sigma + T^{-1/2}\Xi)^{-1}(\Sigma^0)^{1/2}\eta_t \\
= \; & O_p(1)
\end{aligned}$$

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 β and Σ satisfying $||\beta - \beta^0|| \leq Mv_T$ and $||\Sigma - \Sigma^0|| \leq Mv_T$, with $M < \infty$, we have

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

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 1.1 *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 $N = [\sqrt{T} v_T^{-1}]$ and $A_j = \{(K_1, ..., K_m) : |K_j - K_j^0| > N\}$. Then to prove that $(\hat{K}_1, ..., \hat{K}_m) \notin A_j$, it suffices to show that $P(\sup_{(K_1,...,K_m)\in A_j} LR(K_1, ..., K_m) \geqslant 1) \leq \epsilon$. Let $LR(K_1, ..., K_m)$ denote the likelihood ratio evaluated at $(K_1, ..., K_m)$ and note that $LR(K_1, ..., K_m)$

$\geq LR(K_1^0, ..., K_m^0)$

$\geq LR(K_1^0, ..., K_m^0, \beta^0, \Sigma^0)$. We consider the five possible cases separately.

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

Then

$$\begin{aligned} LR(K_1, ..., K_m) &= \textstyle\prod_{l=1}^{j} LR(K_{l-1} + 1, K_l) \cdot LR(K_j + 1, K_{j+1}) \\ &\quad \cdot \textstyle\prod_{l=j+2}^{m+2} LR(K_{l-1} + 1, K_l) \end{aligned}$$

or

$$\begin{aligned} \log LR(K_1, ..., K_m) &= \textstyle\sum_{l=1}^{j} \log LR(K_{l-1} + 1, K_l) \\ &\quad + \log LR(K_j + 1, K_{j+1}) \\ &\quad + \textstyle\sum_{l=j+2}^{m+1} \log LR(K_{l-1} + 1, K_l). \end{aligned}$$

From Properties 2 and 3, there exists constants $C_1 > 0$ and $C_2 > 0$, such that

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

as $T \to \infty$.

Case 2. $K_1 \leq ... \leq K_m \leq K_j^0 - N < K_j^0 < K_j^0 + N$.

Then

$$\begin{aligned} LR(K_1, ..., K_m) &= \textstyle\prod_{l=1}^{m+1} LR(K_{l-1} + 1, K_l) \cdot LR(K_m + 1, K_j^0 - N) \\ &\quad \cdot LR(K_j^0 - N, K_j^0 + N) \cdot LR(K_j^0 + N + 1, T) \end{aligned}$$

or

$$\begin{aligned}\log LR(K_1,...,K_m) &= \textstyle\sum_{l=1}^{m+1}\log LR(K_{l-1}+1,K_l)\\&+\log LR(K_m+1,K_j^0-N)\\&+\log LR(K_j^0-N,K_j^0+N)\\&+\log LR(K_j^0+N+1,T).\end{aligned}$$

Case 3. $K_j^0-N<K_j^0<K_j^0+N\le K_1\le ...\le K_m$.

Then

$$\begin{aligned}LR(K_1,...,K_m) &= LR(1,K_j^0-N)\cdot LR(K_j^0-N+1,K_j^0+N)\\&\cdot LR(K_j^0+N,K_1)\cdot\textstyle\prod_{l=2}^{m+1}LR(K_{l-1}+1,K_l)\end{aligned}$$

or

$$\begin{aligned}\log LR(K_1,...,K_m) &= \log LR(1,K_j^0-N)\\&+\log LR(K_j^0-N+1,K_j^0+N)\\&+\log LR(K_j^0+N,K_1)\\&+\textstyle\sum_{l=2}^{m+1}\log LR(K_{l-1}+1,K_l).\end{aligned}$$

Case 4. There exist $i, i>j$,

such that $K_1\le ...\le K_j\le\cdots\le K_i\le K_j^0-N<K_j^0<K_j^0+N\le K_{i+1}\le ...\le K_m$.

Then

$$\begin{aligned}LR(K_1,...,K_m) &= \textstyle\prod_{l=1}^{i+1}LR(K_{l-1}+1,K_l)\\&\cdot LR(K_i+1,K_j^0-N)\\&\cdot LR(K_j^0-N,K_j^0+N)\\&\cdot LR(K_j^0+N+1,K_{i+1})\\&\cdot\textstyle\prod_{l=i+2}^{m+1}LR(K_{l-1}+1,K_l),\end{aligned}$$

and

$$\begin{aligned}\log LR(K_1,...,K_m) &= \textstyle\sum_{l=1}^{i+1}\log LR(K_{l-1}+1,K_l)\\ &\quad +\log LR(K_i+1,K_j^0-N)\\ &\quad +\log LR(K_j^0-N,K_j^0+N)\\ &\quad +\log LR(K_j^0+N+1,K_{i+1})\\ &\quad +\textstyle\sum_{l=i+2}^{m+1}\log LR(K_{l-1}+1,K_l).\end{aligned}$$

Case 5. The region $[K_j^0-N,K_j^0+N]$ is partitioned into C intervals, in which case there must be at least one sub-interval such that $[K_{l-1}+1,K_l]\geq C^{-1}\sqrt{T}v_T^{-1}$, then

$$\begin{aligned}LR(K_1,...,K_m) &= \textstyle\prod_{l=1}^{j+1}LR(K_{l-1}+1,K_l)\\ &\quad \cdot\textstyle\prod_{l=j+2}^{j+C}LR(K_{l-1}+1,K_l)\\ &\quad \cdot LR(K_{j+C}+1,K_j^0+N)\\ &\quad \cdot LR(K_j^0+N+1,K_{j+C+1})\\ &\quad \cdot\textstyle\prod_{l=j+C+2}^{m+1}LR(K_{l-1}+1,K_l)\end{aligned}$$

and

$$\begin{aligned}\log LR(K_1,...,K_m),K_m) &= \textstyle\sum_{l=1}^{j+1}\log LR(K_{l-1}+1,K_l)\\ &\quad +\textstyle\sum_{l=j+2}^{j+C}\log LR(K_{l-1}+1,K_l)\\ &\quad +\log LR(K_{j+C}+1,K_j^0+N)\\ &\quad +\log LR(K_j^0+N+1,K_{j+C+1})\\ &\quad +\textstyle\sum_{l=j+C+2}^{m+2}\log LR(K_{l-1}+1,K_l).\end{aligned}$$

For all the cases, the log-likelihood function has at least one term with $K_l-K_{l-1}\geq\sqrt{T}v_T^{-1}$, $||\beta-\beta^0||\geq dv_T$ 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

$$P(\sup_{\sqrt{T}v_T^{-1}\leq k\leq T}\ \sup_{S(\beta,\Sigma)^c}\log LR_k(\beta,\Sigma)>-C\sqrt{T}v_T)<\epsilon.$$

Applying Property 2 to the other terms, we have

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

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

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

as $T \to \infty$. This shows that $P(\sup_{(K_1,...,K_m)\in A_j} L(K_1,...,K_m) \ge 1) \le \epsilon$.

Proposition 1.2 *For every $\epsilon > 0$, there exists a $C > 0$, such that $P(|\hat{K}_j - K_j^0| > Cv_T^{-2}) < \epsilon$.*

Proof. Given Proposition 1, it remains to prove that

$$P(\sup_{(K_1,...,k_m)\in A_j(C)} \frac{L(K_1,...,K_m)}{L(K_1^0,...,K_m^0)} \ge 1) \le \epsilon$$

for $A_j(C) = \{(K_1,...,K_m) : |K_j - K_j^0| \le \sqrt{T}v_T^{-1}$ and $|K_j - K_j^0| > Cv_T^{-2}\ \}$. 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 β^0 and Σ^0 follow given the rate of convergence of $\hat{K}_j$.

Proof of Theorem 1.2: The likelihood ratio is

$$\begin{aligned}&LR_T(K_1,...K_m,\beta_{(j)},\Sigma_{(j)}) \\ = &\textstyle\prod_{j=1}^{m+1}\prod_{t=K_{j-1}+1}^{K_j} f(Y_t|X_{Tt},\beta_{(j)},\Sigma_{(j)}) \\ &-\textstyle\prod_{j=1}^{m+1}\prod_{t=K_{j-1}^0+1}^{K_j^0} f(Y_t|X_{Tt},\beta_{(j)}^0,\Sigma^0) \\ = &\textstyle\prod_{j=1}^{m+1}\prod_{t=K_{j-1}+1}^{K_j}(2\pi)^{-1/2}|\Sigma_{(j)}|^{-1/2} \\ &\exp\{-(1/2)(Y_t - X_{Tt}'\beta_{(j)})'(\Sigma_{(j)})^{-1}(Y_t - X_{Tt}'\beta_{(j)})\} \\ &-\textstyle\prod_{j=1}^{m+1}\prod_{t=K_{j-1}^0+1}^{K_j^0}(2\pi)^{-1/2}|\Sigma^0|^{-1/2} \\ &\exp\{-(1/2)(Y_t - X_{Tt}'\beta_{(j)}^0)'(\Sigma^0)^{-1}(Y_t - X_{Tt}'\beta_{(j)}^0)\}.\end{aligned}$$

With the restrictions $g(\beta_{(j)}, vec(\Sigma_{(j)})) = 0$, the restricted log-likelihood function is

$$rlr_T(K_j, \beta_{(j)}, \Sigma_{(j)}) = -(1/2)\sum_{j=1}^{m+1}(K_j - K_{j-1})\log|\Sigma_{(j)}|$$
$$-(1/2)\sum_{j=1}^{m+1}\sum_{t=K_{j-1}+1}^{K_j}(Y_t - X'_{Tt}\beta_{(j)})'(\Sigma_{(j)})^{-1}(Y_t - X'_{Tt}\beta_{(j)})$$
$$+(1/2)\sum_{j=1}^{m+1}(K_j^0 - K_{j-1}^0)\log|\Sigma^0|$$
$$+(1/2)\sum_{j=1}^{m+1}\sum_{t=K_{j-1}+1}^{K_j^0}(Y_t - X'_{Tt}\beta^0_{(j)})'(\Sigma^0)^{-1}(Y_t - X'_{Tt}\beta^0_{(j)})$$
$$+\lambda' g(\beta_{(j)}, vec(\Sigma_{(j)}))$$

Let

$$A = \sum_{j=1}^{m+1}(K_j - K_{j-1})\log|\Sigma_{(j)}| - \sum_{j=1}^{m+1}(K_j^0 - K_{j-1}^0)\log|\Sigma^0|$$

$$\begin{aligned} B \;&=\; \sum_{j=1}^{m+1}\sum_{t=K_{j-1}+1}^{K_j}(Y_t - X'_{Tt}\beta_{(j)})'(\Sigma_{(j)})^{-1}(Y_t - X'_{Tt}\beta_{(j)}) \\ &\quad - \sum_{j=1}^{m+1}\sum_{t=K_{j-1}^0+1}^{K_j^0}(Y_t - X'_{Tt}\beta^0_{(j)})'(\Sigma^0)^{-1}(Y_t - X'_{Tt}\beta^0_{(j)}) \end{aligned}$$

then $lr_T = -A/2 - B/2$ and $rlr_T = -A/2 - B/2 + \lambda' g(\beta_{(j)}, vec(\Sigma_{(j)}))$. We have

$$\begin{aligned} A \;&=\; \sum_{j=1}^{m+1}(K_j - K_{j-1})\log|\Sigma_{(j)}| - \sum_{j=1}^{m+1}(K_j^0 - K_{j-1}^0)\log|\Sigma^0| \\ &=\; \sum_{j=1}^{m+1}(K_j - K_{j-1})\log|\Sigma_{(j)}| - \sum_{j=1}^{m+1}(K_j^0 - K_{j-1}^0)\log|\Sigma_{(j)}| \\ &\quad + \sum_{j=1}^{m+1}(K_j^0 - K_{j-1}^0)\log|\Sigma_{(j)}| - \sum_{j=1}^{m+1}(K_j^0 - K_{j-1}^0)\log|\Sigma^0| \\ &=\; \sum_{j=1}^{m}(K_j - K_j^0)(\log|\Sigma_{(j)}| - \log|\Sigma_{(j+1)}|) \qquad \text{(A.1)} \\ &\quad - \sum_{j=1}^{m+1}(K_j^0 - K_{j-1}^0)(\log|\Sigma_{(j)}| - \log|\Sigma^0|) \qquad \text{(A.2)} \end{aligned}$$

$$\begin{aligned} B \;&=\; \sum_{j=1}^{m+1}\sum_{t=K_{j-1}+1}^{K_j}(Y_t - X'_{Tt}\beta_{(j)})'(\Sigma_{(j)})^{-1}(Y_t - X'_{Tt}\beta_{(j)}) \\ &\quad - \sum_{j=1}^{m+1}\sum_{t=K_{j-1}^0+1}^{K_j^0}(Y_t - X'_{Tt}\beta^0_{(j)})'(\Sigma^0)^{-1}(Y_t - X'_{Tt}\beta^0_{(j)}) \\ &=\; \sum_{j=1}^{m+1}\sum_{t=K_{j-1}+1}^{K_j}(Y_t - X'_{Tt}\beta_{(j)})'(\Sigma_{(j)})^{-1}(Y_t - X'_{Tt}\beta_{(j)}) \\ &\quad - \sum_{j=1}^{m+1}\sum_{t=K_{j-1}^0+1}^{K_j^0}(Y_t - X'_{Tt}\beta_{(j)})'(\Sigma_{(j)})^{-1}(Y_t - X'_{Tt}\beta_{(j)}) \\ &\quad + \sum_{j=1}^{m+1}\sum_{t=K_{j-1}^0+1}^{K_j^0}(Y_t - X'_{Tt}\beta_{(j)})'(\Sigma_{(j)})^{-1}(Y_t - X'_{Tt}\beta_{(j)}) \\ &\quad - \sum_{j=1}^{m+1}\sum_{t=K_{j-1}^0+1}^{K_j^0}(Y_t - X'_{Tt}\beta^0_{(j)})'(\Sigma^0)^{-1}(Y_t - X'_{Tt}\beta^0_{(j)}) \end{aligned}$$

Let

$$\begin{aligned} & rlr_T^1(K_j, \beta_{(j)}, \Sigma_{(j)}) \\ = & -(1/2)(A.2) - (1/2)(B.2) + \lambda' g(\beta_{(j)}, vec(\Sigma_{(j)})) \end{aligned}$$

$$rlr_T^2(K_j, \beta_{(j)}, \Sigma_{(j)}) = -(1/2)(A.1) - (1/2)(B.1)$$

and note that rlr_T^2 will depend on the position of K_j relative to that of K_j^0. 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,

$$\begin{aligned} (\Sigma_{(j)})^{-1} &= (\Sigma^0 + T^{-1/2}\Sigma^*_{(j)})^{-1} \\ &= (\Sigma^0(I + T^{-1/2}(\Sigma^0)^{-1}\Sigma^*_{(j)}))^{-1} \\ &= (I - T^{-1/2}(\Sigma^0)^{-1}\Sigma^*_{(j)} - T^{-1}(\Sigma^0)^{-1}\Sigma^*_{(j)}(\Sigma^0)^{-1}\Sigma^*_{(j)} \\ &\quad + o_p(T^{-1}))(\Sigma^0)^{-1} \\ &= (\Sigma^0)^{-1} - T^{-1/2}(\Sigma^0)^{-1}\Sigma^*_{(j)}(\Sigma^0)^{-1} \\ &\quad - T^{-1}(\Sigma^0)^{-1}\Sigma^*_{(j)}(\Sigma^0)^{-1}\Sigma^*_{(j)}(\Sigma^0)^{-1} + o_p(1) \end{aligned}$$

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

$$\begin{aligned} (A.1) &= \textstyle\sum_{j=1}^2 (K_j - K_j^0)(\log|\Sigma_{(j)}| - \log|\Sigma_{(j+1)}|) \\ &= \textstyle\sum_{j=1}^2 (K_j - K_j^0)(\log|\Sigma^0| - T^{-1/2} tr((\Sigma^0)^{-1}\Sigma^*_{(j)}) \\ &\quad - \log|\Sigma^0| - T^{-1/2} tr((\Sigma^0)^{-1}\Sigma^*_{(j+1)})) + o_p(1) \end{aligned}$$

For $(B.1)$, we need to consider the three cases separately. For Case 1 ($K_1 < K_2 \le K_1^0 < K_2^0$):

$$\begin{aligned} (B.1) = & -\textstyle\sum_{t=K_1+1}^{K_1^0} (Y_t - X'_{Tt}\beta_{(1)})'(\Sigma_{(1)})^{-1}(Y_t - X'_{Tt}\beta_{(1)}) && (1) \\ & -\textstyle\sum_{t=K_1^0+1}^{K_2^0} (Y_t - X'_{Tt}\beta_{(2)})'(\Sigma_{(2)})^{-1}(Y_t - X'_{Tt}\beta_{(2)}) && (2) \\ & +\textstyle\sum_{t=K_1+1}^{K_2} (Y_t - X'_{Tt}\beta_{(2)})'(\Sigma_{(2)})^{-1}(Y_t - X'_{Tt}\beta_{(2)}) && (3) \end{aligned}$$

$$+\textstyle\sum_{t=K_2+1}^{K_1^0}(Y_t - X_{Tt}'\beta_{(3)})'(\Sigma_{(3)})^{-1}(Y_t - X_{Tt}'\beta_{(3)}) \qquad (4)$$

$$+\textstyle\sum_{t=K_1^0+1}^{K_2^0}(Y_t - X_{Tt}'\beta_{(3)})'(\Sigma_{(3)})^{-1}(Y_t - X_{Tt}'\beta_{(3)}) \qquad (5)$$

We analyze each of the five terms separately.

$$\begin{aligned}
(1) \quad = \quad & -\textstyle\sum_{t=K_1+1}^{K_1^0}(u_t + X_{Tt}'\beta_{(1)}^0 - X_{Tt}'\beta_{(1)})' \\
& (\Sigma_{(1)})^{-1}(u_t + X_{Tt}'\beta_{(1)}^0 - X_{Tt}'\beta_{(1)}) \\
= \quad & -\textstyle\sum_{t=K_1+1}^{K_1^0} u_t'(\Sigma^0)^{-1}u_t \\
& +T^{-1/2}\textstyle\sum_{t=K_1+1}^{K_1^0} u_t'(\Sigma^0)^{-1}\Sigma_{(1)}^*(\Sigma^0)^{-1}u_t \\
& +(\beta_{(1)} - \beta_{(1)}^0)'(\textstyle\sum_{t=K_1+1}^{K_1^0} X_{Tt}(\Sigma^0)^{-1}X_{Tt}')(\beta_{(1)} - \beta_{(1)}^0) \\
& +o_p(1) \\
= \quad & -\textstyle\sum_{t=K_1+1}^{K_1^0} u_t'(\Sigma^0)^{-1}u_t \\
& +T^{-1/2}\textstyle\sum_{t=K_1+1}^{K_1^0} u_t'(\Sigma^0)^{-1}\Sigma_{(1)}^*(\Sigma^0)^{-1}u_t + o_p(1)
\end{aligned}$$

$$\begin{aligned}
(2) \quad = \quad & -\textstyle\sum_{t=K_1^0+1}^{K_2^0}(u_t + X_{Tt}'\beta_{(2)}^0 - X_{Tt}'\beta_{(2)})' \\
& (\Sigma_{(2)})^{-1}(u_t + X_{Tt}'\beta_{(2)}^0 - X_{Tt}'\beta_{(2)}) \\
= \quad & -\textstyle\sum_{t=K_1^0+1}^{K_2^0} u_t'(\Sigma^0)^{-1}u_t \\
& +T^{-1/2}\textstyle\sum_{t=K_1^0+1}^{K_2^0} u_t'(\Sigma^0)^{-1}\Sigma_{(2)}^*(\Sigma^0)^{-1}u_t + o_p(1)
\end{aligned}$$

$$\begin{aligned}
(3) \quad = \quad & \textstyle\sum_{t=K_1+1}^{K_2}(u_t + X_{Tt}'\beta_{(1)}^0 - X_{Tt}'\beta_{(2)})' \\
& (\Sigma_{(2)})^{-1}(u_t + X_{Tt}'\beta_{(1)}^0 - X_{Tt}'\beta_{(2)}) \\
= \quad & \textstyle\sum_{t=K_1+1}^{K_2} u_t'(\Sigma^0)^{-1}u_t \\
& -T^{-1/2}\textstyle\sum_{t=K_1+1}^{K_2} u_t'(\Sigma^0)^{-1}\Sigma_{(2)}^*(\Sigma^0)^{-1}u_t \\
& +(\beta_{(2)}^0 - \beta_{(1)}^0)'(\textstyle\sum_{t=K_1+1}^{K_2} X_{Tt}(\Sigma^0)^{-1}X_{Tt}')(\beta_{(2)}^0 - \beta_{(1)}^0) \\
& -2(\beta_{(2)}^0 - \beta_{(1)}^0)'(\textstyle\sum_{t=K_1+1}^{K_2} X_{Tt}(\Sigma^0)^{-1}u_t) + o_p(1)
\end{aligned}$$

$$\begin{aligned}
(4) &= \textstyle\sum_{t=K_2+1}^{K_1^0}(u_t + X_{Tt}'\beta_{(1)}^0 - X_{Tt}'\beta_{(3)})'(\Sigma_{(3)})^{-1}(u_t + X_{Tt}'\beta_{(1)}^0 - X_{Tt}'\beta_{(3)}) \\
&= \textstyle\sum_{t=K_2+1}^{K_1^0} u_t'(\Sigma^0)^{-1}u_t - T^{-1/2}\sum_{t=K_2+1}^{K_1^0} u_t'(\Sigma^0)^{-1}\Sigma_{(3)}^*(\Sigma^0)^{-1}u_t \\
&\quad +(\beta_{(3)}^0 - \beta_{(1)}^0)'(\textstyle\sum_{t=K_2+1}^{K_1^0} X_{Tt}(\Sigma^0)^{-1}X_{Tt}')(\beta_{(3)}^0 - \beta_{(1)}^0) \\
&\quad -2(\beta_{(3)}^0 - \beta_{(1)}^0)'(\textstyle\sum_{t=K_2+1}^{K_1^0} X_{Tt}(\Sigma^0)^{-1}u_t) + o_p(1)
\end{aligned}$$

$$\begin{aligned}
(5) &= \textstyle\sum_{t=K_1^0+1}^{K_2^0}(u_t + X_{Tt}'\beta_{(2)}^0 - X_{Tt}'\beta_{(3)})'(\Sigma_{(3)})^{-1}(u_t + X_{Tt}'\beta_{(2)}^0 - X_{Tt}'\beta_{(3)}) \\
&= \textstyle\sum_{t=K_1^0+1}^{K_2^0} u_t'(\Sigma^0)^{-1}u_t - T^{-1/2}\sum_{t=K_1^0+1}^{K_2^0} u_t'(\Sigma^0)^{-1}\Sigma_{(3)}^*(\Sigma^0)^{-1}u_t \\
&\quad +(\beta_{(3)}^0 - \beta_{(2)}^0)'(\textstyle\sum_{t=K_1^0+1}^{K_2^0} X_{Tt}(\Sigma^0)^{-1}X_{Tt}')(\beta_{(3)}^0 - \beta_{(2)}^0) \\
&\quad -2(\beta_{(3)}^0 - \beta_{(2)}^0)'(\textstyle\sum_{t=K_1^0+1}^{K_2^0} X_{Tt}(\Sigma^0)^{-1}u_t) + o_p(1)
\end{aligned}$$

Hence,

$$\begin{aligned}
& rlr_T^2(K_1, K_2, \beta, \Sigma) \\
= & -(1/2)\textstyle\sum_{j=1}^{2}(K_j - K_j^0)[\log|\Sigma^0| - T^{-1/2}tr((\Sigma^0)^{-1}\Sigma_{(j)}^*) && \text{(C.1)} \\
& -\log|\Sigma^0| - T^{-1/2}tr((\Sigma^0)^{-1}\Sigma_{(j+1)}^*)] \\
& -(1/2)[-\textstyle\sum_{t=K_1+1}^{K_1^0} u_t'(\Sigma^0)^{-1}u_t + T^{-1/2}\sum_{t=K_1+1}^{K_1^0} u_t'(\Sigma^0)^{-1}\Sigma_{(1)}^*(\Sigma^0)^{-1}u_t \\
& -\textstyle\sum_{t=K_1^0+1}^{K_2^0} u_t'(\Sigma^0)^{-1}u_t + T^{-1/2}\sum_{t=K_1^0+1}^{K_2^0} u_t'(\Sigma^0)^{-1}\Sigma_{(2)}^*(\Sigma^0)^{-1}u_t \\
& +\textstyle\sum_{t=K_1+1}^{K_2} u_t'(\Sigma^0)^{-1}u_t - T^{-1/2}\sum_{t=K_1+1}^{K_2} u_t'(\Sigma^0)^{-1}\Sigma_{(2)}^*(\Sigma^0)^{-1}u_t \\
& +\textstyle\sum_{t=K_2+1}^{K_1^0} u_t'(\Sigma^0)^{-1}u_t - T^{-1/2}\sum_{t=K_2+1}^{K_1^0} u_t'(\Sigma^0)^{-1}\Sigma_{(3)}^*(\Sigma^0)^{-1}u_t \\
& +\textstyle\sum_{t=K_1^0+1}^{K_2^0} u_t'(\Sigma^0)^{-1}u_t - T^{-1/2}\sum_{t=K_1^0+1}^{K_2^0} u_t'(\Sigma^0)^{-1}\Sigma_{(3)}^*(\Sigma^0)^{-1}u_t] \\
& -(1/2)[(\beta_{(2)}^0 - \beta_{(1)}^0)'(\textstyle\sum_{t=K_1+1}^{K_2} X_{Tt}(\Sigma^0)^{-1}X_{Tt}')(\beta_{(2)}^0 - \beta_{(1)}^0) && \text{(C.2)} \\
& +(\beta_{(3)}^0 - \beta_{(1)}^0)'(\textstyle\sum_{t=K_2+1}^{K_1^0} X_{Tt}(\Sigma^0)^{-1}X_{Tt}')(\beta_{(3)}^0 - \beta_{(1)}^0) \\
& +(\beta_{(3)}^0 - \beta_{(2)}^0)'(\textstyle\sum_{t=K_1^0+1}^{K_2^0} X_{Tt}(\Sigma^0)^{-1}X_{Tt}')(\beta_{(3)}^0 - \beta_{(2)}^0)] \\
& +[(\beta_{(2)}^0 - \beta_{(1)}^0)'(\textstyle\sum_{t=K_1+1}^{K_2} X_{Tt}(\Sigma^0)^{-1}u_t) && \text{(C.3)} \\
& +(\beta_{(3)}^0 - \beta_{(2)}^0)'(\textstyle\sum_{t=K_1^0+1}^{K_2^0} X_{Tt}(\Sigma^0)^{-1}u_t) \\
& +(\beta_{(3)}^0 - \beta_{(1)}^0)'(\textstyle\sum_{t=K_2+1}^{K_1^0} X_{Tt}(\Sigma^0)^{-1}u_t)] + o_p(1)
\end{aligned}$$

Note that

$$\begin{aligned}
& (K_j - K_j^0)T^{-1/2}tr((\Sigma^0)^{-1}\Sigma^*_{(j)}) + T^{-1/2}\sum_{t=K_j+1}^{K_j^0} u_t'(\Sigma^0)^{-1}\Sigma^*_{(j)}(\Sigma^0)^{-1}u_t \\
= \; & (K_j - K_j^0)T^{-1/2}tr((\Sigma^0)^{-1}\Sigma^*_{(j)}) + T^{-1/2}tr((\Sigma^0)^{-1}\Sigma^*_{(j)}(\Sigma^0)^{-1}\sum_{t=K_j+1}^{K_j^0} u_t u_t') \\
= \; & (K_j - K_j^0)T^{-1/2}tr((\Sigma^0)^{-1}\Sigma^*_{(j)}) + T^{-1/2}tr((\Sigma^0)^{-1}\Sigma^*_{(j)}(\Sigma^0)^{-1}\sum_{t=K_j+1}^{K_j^0}(u_t u_t' - \Sigma^0)) \\
& -(K_j - K_j^0)T^{-1/2}tr((\Sigma^0)^{-1}\Sigma^*_{(j)}) = o_p(1)
\end{aligned}$$

$$\begin{aligned}
v_T \sum_{t=K_j+1}^{K_j^0}(u_t u_t' - \Sigma^0) \quad &= \quad (\Sigma^0)^{1/2}[v_T \sum_{t=K_j+1}^{K_j^0}(\eta_t \eta_t' - I)](\Sigma^0)^{1/2} \\
&\Rightarrow \quad (\Sigma^0)^{1/2}[\xi(s_j)](\Sigma^0)^{1/2}
\end{aligned}$$

$$\begin{gathered}
\sum_{j=1}^{2}(K_j - K_j^0)[-T^{-1/2}tr((\Sigma^0)^{-1}\Sigma^*_{(j)}) - T^{-1/2}tr((\Sigma^0)^{-1}\Sigma^*_{(j+1)})] \\
+\sum_{t=K_1+1}^{K_1^0} u_t'(\Sigma^0)^{-1}u_t + \sum_{t=K_1^0+1}^{K_2^0} u_t'(\Sigma^0)^{-1}u_t - \sum_{t=K_1+1}^{K_2} u_t'(\Sigma^0)^{-1}u_t \\
-\sum_{t=K_2+1}^{K_1^0} u_t'(\Sigma^0)^{-1}u_t - \sum_{t=K_1^0+1}^{K_2^0} u_t'(\Sigma^0)^{-1}u_t = o_p(1)
\end{gathered}$$

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

$$\begin{aligned}
& (C.2) = -(1/2)[(\beta^0_{(2)} - \beta^0_{(1)})'(\sum_{t=K_1+1}^{K_2} X_{Tt}(\Sigma^0)^{-1}X_{Tt}')(\beta^0_{(2)} - \beta^0_{(1)}) \\
& +(\beta^0_{(3)} - \beta^0_{(2)})'(\sum_{t=K_2+1}^{K_1^0} X_{Tt}(\Sigma^0)^{-1}X_{Tt}')(\beta^0_{(3)} - \beta^0_{(2)}) \\
& +(\beta^0_{(2)} - \beta^0_{(1)})'(\sum_{t=K_2+1}^{K_1^0} X_{Tt}(\Sigma^0)^{-1}X_{Tt}')(\beta^0_{(2)} - \beta^0_{(1)}) \\
& +2(\beta^0_{(3)} - \beta^0_{(2)})'(\sum_{t=K_2+1}^{K_1^0} X_{Tt}(\Sigma^0)^{-1}X_{Tt}')(\beta^0_{(2)} - \beta^0_{(1)}) \\
& +(\beta^0_{(3)} - \beta^0_{(2)})'(\sum_{t=K_1^0+1}^{K_2^0} X_{Tt}(\Sigma^0)^{-1}X_{Tt}')(\beta^0_{(3)} - \beta^0_{(2)})] \\
= \; & -(1/2)[(\beta^0_{(2)} - \beta^0_{(1)})'(\sum_{t=K_1+1}^{K_1^0} X_{Tt}(\Sigma^0)^{-1}X_{Tt}')(\beta^0_{(2)} - \beta^0_{(1)}) \\
& +(\beta^0_{(3)} - \beta^0_{(2)})'(\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma^0)^{-1}X_{Tt}')(\beta^0_{(3)} - \beta^0_{(2)}) \\
& +2(\beta^0_{(3)} - \beta^0_{(2)})'(\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma^0)^{-1}X_{Tt}')(\beta^0_{(2)} - \beta^0_{(1)})] + o_p(1)
\end{aligned}$$

$$\begin{aligned}
& (C.3) = (\beta^0_{(2)} - \beta^0_{(1)})'(\sum_{t=K_1+1}^{K_2} X_{Tt}(\Sigma^0)^{-1}u_t) \\
& +(\beta^0_{(2)} - \beta^0_{(1)})'(\sum_{t=K_2+1}^{K_1^0} X_{Tt}(\Sigma^0)^{-1}u_t)
\end{aligned}$$

$$
\begin{aligned}
&+(\beta^0_{(3)}-\beta^0_{(2)})'(\textstyle\sum_{t=K_1^0+1}^{K_2^0} X_{Tt}(\Sigma^0)^{-1}u_t) \\
&+(\beta^0_{(3)}-\beta^0_{(2)})'(\textstyle\sum_{t=K_2+1}^{K_1^0} X_{Tt}(\Sigma^0)^{-1}u_t) \\
= \;&\beta^0_{(2)}-\beta^0_{(1)})'(\textstyle\sum_{t=K_1+1}^{K_1^0} X_{Tt}(\Sigma^0)^{-1}u_t) \\
&+(\beta^0_{(3)}-\beta^0_{(2)})'(\textstyle\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma^0)^{-1}u_t)
\end{aligned}
$$

Finally,

$$
\begin{aligned}
rlr_T^2(K_1,K_2,\beta,\Sigma) \;=\; & -(1/2)(\beta^0_{(2)}-\beta^0_{(1)})'(\textstyle\sum_{t=K_1+1}^{K_1^0} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(2)}-\beta^0_{(1)}) \\
& -(1/2)(\beta^0_{(3)}-\beta^0_{(2)})'(\textstyle\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(3)}-\beta^0_{(2)}) \\
& -(\beta^0_{(3)}-\beta^0_{(2)})'(\textstyle\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(2)}-\beta^0_{(1)}) \\
& +\beta^0_{(2)}-\beta^0_{(1)})'(\textstyle\sum_{t=K_1+1}^{K_1^0} X_{Tt}(\Sigma^0)^{-1}u_t) \\
& +(\beta^0_{(3)}-\beta^0_{(2)})'(\textstyle\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma^0)^{-1}u_t)+o_p(1)
\end{aligned}
$$

Case 2: $K_1^0 < K_2^0 \le K_1 < K_2$. For (B.1), we have

$$
\begin{aligned}
(B.1) \;=\; & \textstyle\sum_{t=K_1^0+1}^{K_2^0}(Y_t-X'_{Tt}\beta_{(1)})'(\Sigma_{(1)})^{-1}(Y_t-X'_{Tt}\beta_{(1)}) && (1.8)\\
& +\textstyle\sum_{t=K_2^0+1}^{K_1}(Y_t-X'_{Tt}\beta_{(1)})'(\Sigma_{(1)})^{-1}(Y_t-X'_{Tt}\beta_{(1)}) && (1.9)\\
& -\textstyle\sum_{t=K_1^0+1}^{K_2^0}(Y_t-X'_{Tt}\beta_{(2)})'(\Sigma_{(2)})^{-1}(Y_t-X'_{Tt}\beta_{(2)}) \\
& -\textstyle\sum_{t=K_2^0+1}^{K_2}(Y_t-X'_{Tt}\beta_{(3)})'(\Sigma_{(3)})^{-1}(Y_t-X'_{Tt}\beta_{(3)}) \\
& +\textstyle\sum_{t=K_1+1}^{K_2}(Y_t-X'_{Tt}\beta_{(2)})'(\Sigma_{(2)})^{-1}(Y_t-X'_{Tt}\beta_{(2)})
\end{aligned}
$$

so that

$$
\begin{aligned}
& rlr_T^2(K_1,K_2,\beta,\Sigma) = \\
& -(1/2)(\beta^0_{(2)}-\beta^0_{(1)})'(\textstyle\sum_{t=K_1^0+1}^{K_1} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(2)}-\beta^0_{(1)}) \\
& -(1/2)(\beta^0_{(3)}-\beta^0_{(2)})'(\textstyle\sum_{t=K_2^0+1}^{K_2} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(3)}-\beta^0_{(2)}) \\
& -(\beta^0_{(3)}-\beta^0_{(2)})'(\textstyle\sum_{t=K_1^0+1}^{K_1} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(2)}-\beta^0_{(1)})
\end{aligned}
$$

$$-(\beta^0_{(2)} - \beta^0_{(1)})' \sum_{t=K_1^0+1}^{K_1} X_{Tt}(\Sigma^0)^{-1}u_t$$
$$-(\beta^0_{(3)} - \beta^0_{(2)})' \sum_{t=K_2^0+1}^{K_2} X_{Tt}(\Sigma^0)^{-1}u_t + o_p(1)$$

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

$$\begin{aligned}(B.1) \quad = \quad & -\sum_{t=K_1+1}^{K_1^0}(Y_t - X'_{Tt}\beta_{(2)})'(\Sigma_{(2)})^{-1}(Y_t - X'_{Tt}\beta_{(2)}) \\ & -\sum_{t=K_1+1}^{K_1^0}(Y_t - X'_{Tt}\beta_{(1)})'(\Sigma_{(1)})^{-1}(Y_t - X'_{Tt}\beta_{(1)}) \\ & +\sum_{t=K_2^0+1}^{K_2}(Y_t - X'_{Tt}\beta_{(2)})'(\Sigma_{(2)})^{-1}(Y_t - X'_{Tt}\beta_{(2)}) \\ & -\sum_{t=K_2^0+1}^{K_2}(Y_t - X'_{Tt}\beta_{(3)})'(\Sigma_{(3)})^{-1}(Y_t - X'_{Tt}\beta_{(3)})\end{aligned}$$

and using arguments similar to those for Case 1,

$$\begin{aligned}& rlr^2_T(K_{1,}K_2, \beta, \Sigma) \\ = \quad & -(1/2)(\beta^0_{(2)} - \beta^0_{(1)})'(\sum_{t=K_1+1}^{K_1^0} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(2)} - \beta^0_{(1)}) \\ & -(1/2)(\beta^0_{(3)} - \beta^0_{(2)})'(\sum_{t=K_2^0+1}^{K_2} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(3)} - \beta^0_{(2)}) \\ & +\beta^0_{(2)} - \beta^0_{(1)})'(\sum_{t=K_1+1}^{K_1^0} X_{Tt}(\Sigma^0)^{-1}u_t) \\ & -(\beta^0_{(3)} - \beta^0_{(2)})'(\sum_{t=K_2^0+1}^{K_2} X_{Tt}(\Sigma^0)^{-1}u_t) + o_p(1)\end{aligned}$$

b) Multiple breaks in coefficients only.

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

$$\begin{aligned}(A.1) \quad = \quad & \sum_{j=1}^m (K_j - K_j^0)(\log|\Sigma_{(j)}| - \log|\Sigma_{(j+1)}|) \\ = \quad & \sum_{j=1}^m (K_j - K_j^0)(\log|\Sigma^0| - T^{-1/2}tr((\Sigma^0)^{-1}\Sigma^*_{(j)}) \\ & -\log|\Sigma^0| + T^{-1/2}tr((\Sigma^0)^{-1}\Sigma^*_{(j+1)})) + o_p(1)\end{aligned}$$

$$(B.1) \quad = \quad -\sum_{t=K_1+1}^{K_1^0}(Y_t - X'_{Tt}\beta_{(1)})'(\Sigma_{(1)})^{-1}(Y_t - X'_{Tt}\beta_{(1)}) \tag{D.1}$$
$$+\sum_{j=2}^m \sum_{t=K_{j-1}+1}^{K_j}(Y_t - X'_{Tt}\beta_{(j)})'(\Sigma_{(j)})^{-1}(Y_t - X'_{Tt}\beta_{(j)}) \tag{D.2}$$
$$-\sum_{j=2}^m \sum_{t=K_{j-1}^0+1}^{K_j^0}(Y_t - X'_{Tt}\beta_{(j)})'(\Sigma_{(j)})^{-1}(Y_t - X'_{Tt}\beta_{(j)}) \tag{D.3}$$

$$+\sum_{t=K_m+1}^{K_1^0}(Y_t - X_{Tt}'\beta_{(m+1)})'(\Sigma_{(m+1)})^{-1}(Y_t - X_{Tt}'\beta_{(m+1)}) \tag{D.4}$$

$$+\sum_{j=2}^{m}\sum_{t=K_{j-1}^0+1}^{K_j^0}(Y_t - X_{Tt}'\beta_{(m+1)})'(\Sigma_{(m+1)})^{-1}(Y_t - X_{Tt}'\beta_{(m+1)}) \tag{D.5}$$

with

$$\begin{aligned}
(D.1) &= -\sum_{t=K_1+1}^{K_1^0}(u_t + X_{Tt}'\beta_{(1)}^0 - X_{Tt}'\beta_{(1)})' \\
&\quad (\Sigma_{(1)})^{-1}(u_t + X_{Tt}'\beta_{(1)}^0 - X_{Tt}'\beta_{(1)}) \\
&= -\sum_{t=K_1+1}^{K_1^0} u_t'(\Sigma^0)^{-1}u_t \\
&\quad + T^{-1/2}\sum_{t=K_1+1}^{K_1^0} u_t'(\Sigma^0)^{-1}\Sigma_{(1)}^*(\Sigma^0)^{-1}u_t + o_p(1)
\end{aligned}$$

$$\begin{aligned}
(D.2) &= \sum_{j=2}^{m}\sum_{t=K_{j-1}+1}^{K_j}(u_t + X_{Tt}'\beta_{(1)}^0 - X_{Tt}'\beta_{(j)})' \\
&\quad (\Sigma_{(j)})^{-1}(u_t + X_{Tt}'\beta_{(1)}^0 - X_{Tt}'\beta_{(j)}) \\
&= \sum_{j=2}^{m}\sum_{t=K_{j-1}+1}^{K_j} u_t'(\Sigma^0)^{-1}u_t \\
&\quad - T^{-1/2}\sum_{j=2}^{m}\sum_{t=K_{j-1}+1}^{K_j} u_t'(\Sigma^0)^{-1}\Sigma_{(j)}^*(\Sigma^0)^{-1}u_t \\
&\quad + \sum_{j=2}^{m}(\beta_{(j)} - \beta_{(1)}^0)'(\sum_{t=K_{j-1}+1}^{K_j} X_{Tt}(\Sigma^0)^{-1}X_{Tt}')(\beta_{(j)} - \beta_{(1)}^0) \\
&\quad - 2\sum_{j=2}^{m}(\beta_{(j)} - \beta_{(1)}^0)'(\sum_{t=K_{j-1}+1}^{K_j} X_{Tt}(\Sigma^0)^{-1}u_t) + o_p(1)
\end{aligned}$$

$$\begin{aligned}
(D.3) &= -\sum_{j=2}^{m}\sum_{t=K_{j-1}^0+1}^{K_j^0}(u_t + X_{Tt}'\beta_{(j)}^0 - X_{Tt}'\beta_{(j)})' \\
&\quad (\Sigma_{(j)})^{-1}(u_t + X_{Tt}'\beta_{(j)}^0 - X_{Tt}'\beta_{(j)}) \\
&= -\sum_{j=2}^{m}\sum_{t=K_{j-1}^0+1}^{K_j^0} u_t'(\Sigma^0)^{-1}u_t \\
&\quad + T^{-1/2}\sum_{j=2}^{m}\sum_{t=K_{j-1}^0+1}^{K_j^0} u_t'(\Sigma^0)^{-1}\Sigma_{(j)}^*(\Sigma^0)^{-1}u_t + o_p(1)
\end{aligned}$$

$$\begin{aligned}
(D.4) &= \sum_{t=K_m+1}^{K_1^0}(u_t + X_{Tt}'\beta_{(1)}^0 - X_{Tt}'\beta_{(m+1)})' \\
&\quad (\Sigma_{(m+1)})^{-1}(u_t + X_{Tt}'\beta_{(1)}^0 - X_{Tt}'\beta_{(m+1)}) \\
&= \sum_{t=K_m+1}^{K_1^0} u_t'(\Sigma^0)^{-1}u_t \\
&\quad - T^{-1/2}\sum_{t=K_m+1}^{K_1^0} u_t'(\Sigma^0)^{-1}\Sigma_{(m+1)}^*(\Sigma^0)^{-1}u_t
\end{aligned}$$

$$+(\beta^0_{(m+1)} - \beta^0_{(1)})'(\textstyle\sum_{t=K_m+1}^{K_1^0} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(m+1)} - \beta^0_{(1)})$$
$$-2(\beta^0_{(m+1)} - \beta^0_{(1)})'(\textstyle\sum_{t=K_m+1}^{K_1^0} X_{Tt}(\Sigma^0)^{-1}u_t) + o_p(1)$$

$$\begin{aligned}
(D.5) &= \textstyle\sum_{j=2}^{m}\sum_{t=K_{j-1}^0+1}^{K_j^0}(u_t + X'_{Tt}\beta^0_{(j)} - X'_{Tt}\beta_{(m+1)})' \\
&\quad (\Sigma_{(m+1)})^{-1}(u_t + X'_{Tt}\beta^0_{(j)} - X'_{Tt}\beta_{(m+1)}) \\
&= \textstyle\sum_{j=2}^{m}\sum_{t=K_{j-1}^0+1}^{K_j^0} u'_t(\Sigma^0)^{-1}u_t \\
&\quad -T^{-1/2}\textstyle\sum_{j=2}^{m}\sum_{t=K_{j-1}^0+1}^{K_j^0} u'_t(\Sigma^0)^{-1}\Sigma^*_{(m+1)}(\Sigma^0)^{-1}u_t \\
&\quad +\textstyle\sum_{j=2}^{m}(\beta^0_{(m+1)} - \beta^0_{(j)})'(\sum_{t=K_{j-1}^0+1}^{K_j^0} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(m+1)} - \beta^0_{(j)}) \\
&\quad -2\textstyle\sum_{j=2}^{m}(\beta^0_{(m+1)} - \beta^0_{(j)})'(\sum_{t=K_{j-1}^0+1}^{K_j^0} X_{Tt}(\Sigma^0)^{-1}u_t) + o_p(1)
\end{aligned}$$

Hence, $rlr_T^2(K_1, \cdots, K_m, \beta, \Sigma) = -(1/2)((A.1) + (B.1))$ is composed on the following four parts:

$$\begin{aligned}
I &= -(1/2)[\textstyle\sum_{j=1}^{m}(K_j - K_j^0)(\log|\Sigma^0| - \log|\Sigma^0|) \\
&\quad -\textstyle\sum_{t=K_1+1}^{K_1^0} u'_t(\Sigma^0)^{-1}u_t + \sum_{j=2}^{m}\sum_{t=K_{j-1}+1}^{K_j} u'_t(\Sigma^0)^{-1}u_t \\
&\quad -\textstyle\sum_{j=2}^{m}\sum_{t=K_{j-1}^0+1}^{K_j^0} u'_t(\Sigma^0)^{-1}u_t + \sum_{t=K_m+1}^{K_1^0} u'_t(\Sigma^0)^{-1}u_t \\
&\quad +\textstyle\sum_{j=2}^{m}\sum_{t=K_{j-1}^0+1}^{K_j^0} u'_t(\Sigma^0)^{-1}u_t = 0
\end{aligned}$$

$$\begin{aligned}
II &= -(1/2)[-\textstyle\sum_{j=1}^{m} T^{-1/2}(K_j - K_j^0)(tr((\Sigma^0)^{-1}\Sigma^*_{(j)}) \\
&\quad -T^{-1/2}tr((\Sigma^0)^{-1}\Sigma^*_{(j+1)})) \\
&\quad +T^{-1/2}\textstyle\sum_{t=K_1+1}^{K_1^0} u'_t(\Sigma^0)^{-1}\Sigma^*_{(1)}(\Sigma^0)^{-1}u_t \\
&\quad -T^{-1/2}\textstyle\sum_{j=2}^{m}\sum_{t=K_{j-1}+1}^{K_j} u'_t(\Sigma^0)^{-1}\Sigma^*_{(j)}(\Sigma^0)^{-1}u_t \\
&\quad +T^{-1/2}\textstyle\sum_{j=2}^{m}\sum_{t=K_{j-1}^0+1}^{K_j^0} u'_t(\Sigma^0)^{-1}\Sigma^*_{(j)}(\Sigma^0_{(j)})^{-1}u_t \\
&\quad -T^{-1/2}\textstyle\sum_{t=K_m+1}^{K_1^0} u'_t(\Sigma^0)^{-1}\Sigma^*_{(m+1)}(\Sigma^0)^{-1}u_t \\
&\quad -T^{-1/2}\textstyle\sum_{j=2}^{m}\sum_{t=K_{j-1}^0+1}^{K_j^0} u'_t(\Sigma^0)^{-1}\Sigma^*_{(m+1)}(\Sigma^0)^{-1}u_t \\
&= o_p(1)
\end{aligned}$$

$$
\begin{aligned}
III &= -(1/2)[\textstyle\sum_{j=2}^{m}(\beta^0_{(j)}-\beta^0_{(1)})'(\sum_{t=K_{j-1}+1}^{K_j}X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(j)}-\beta^0_{(1)})\\
&\quad +(\beta^0_{(m+1)}-\beta^0_{(1)})'(\textstyle\sum_{t=K_m+1}^{K_1^0}X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(m+1)}-\beta^0_{(1)})\\
&\quad +\textstyle\sum_{j=2}^{m}(\beta^0_{(m+1)}-\beta^0_{(j)})'(\sum_{t=K^0_{j-1}+1}^{K_j^0}X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(m+1)}-\beta^0_{(j)})]\\
&= -(1/2)[\textstyle\sum_{j=1}^{m-1}(\beta^0_{(j+1)}-\beta^0_{(j)})'(\sum_{t=K_j+1}^{K_m}X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(j+1)}-\beta^0_{(j)})\\
&\quad +\textstyle\sum_{j=1}^{m}(\beta^0_{(j+1)}-\beta^0_{(j)})'(\sum_{t=K_m+1}^{K_1^0}X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(j+1)}-\beta^0_{(j)})\\
&\quad +\textstyle\sum_{j=2}^{m}(\beta^0_{(j+1)}-\beta^0_{(j)})'(\sum_{t=K_1^0+1}^{K_j^0}X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(j+1)}-\beta^0_{(j)})\\
&\quad +2\textstyle\sum_{j=1}^{m-2}\sum_{i=j+1}^{m-1}(\beta^0_{(j+1)}-\beta^0_{(j)})'(\sum_{t=K_i+1}^{K_m}X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(i+1)}-\beta^0_{(i)})\\
&\quad +2\textstyle\sum_{j=1}^{m-1}\sum_{i=j+1}^{m}(\beta^0_{(j+1)}-\beta^0_{(j)})'(\sum_{t=K_m+1}^{K_1^0}X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(i+1)}-\beta^0_{(i)})\\
&\quad +2\textstyle\sum_{j=2}^{m-1}\sum_{i=j+1}^{m}(\beta^0_{(j+1)}-\beta^0_{(j)})'(\sum_{t=K_1^0+1}^{K_j^0}X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(i+1)}-\beta^0_{(i)})]\\
&\quad +o_p(1)\\
&= \textstyle\sum_{j=1}^{m}(\beta^0_{(j+1)}-\beta^0_{(j)})'(\sum_{t=K_j+1}^{K_j^0}X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(j+1)}-\beta^0_{(j)})\\
&\quad +2\textstyle\sum_{j=1}^{m-1}\sum_{i=j+1}^{m}(\beta^0_{(j+1)}-\beta^0_{(j)})'(\sum_{t=K_i+1}^{K_i^0}X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(i+1)}-\beta^0_{(i)})\\
&\quad +o_p(1)
\end{aligned}
$$

$$
\begin{aligned}
IV &= -2\textstyle\sum_{j=2}^{m}(\beta_{(j)}-\beta^0_{(1)})'(\sum_{t=K_{j-1}+1}^{K_j}X_{Tt}(\Sigma^0)^{-1}u_t)\\
&\quad -2(\beta^0_{(m+1)}-\beta^0_{(1)})'(\textstyle\sum_{t=K_m+1}^{K_1^0}X_{Tt}(\Sigma^0)^{-1}u_t)\\
&\quad -2\textstyle\sum_{j=2}^{m}(\beta^0_{(m+1)}-\beta^0_{(j)})'(\sum_{t=K^0_{j-1}+1}^{K_j^0}X_{Tt}(\Sigma^0)^{-1}u_t)\\
&= -2\textstyle\sum_{j=1}^{m-1}(\beta_{(j+1)}-\beta^0_{(j)})'(\sum_{t=K_j+1}^{Km}X_{Tt}(\Sigma^0)^{-1}u_t)\\
&\quad -2\textstyle\sum_{j=1}^{m}(\beta_{(j+1)}-\beta^0_{(j)})'(\sum_{t=K_m+1}^{K_1^0}X_{Tt}(\Sigma^0)^{-1}u_t)\\
&\quad -2\textstyle\sum_{j=2}^{m}(\beta_{(j+1)}-\beta^0_{(j)})'(\sum_{t=K_1^0+1}^{K_j^0}X_{Tt}(\Sigma^0)^{-1}u_t)\\
&= -2\textstyle\sum_{j=1}^{m}(\beta_{(j+1)}-\beta^0_{(j)})'(\sum_{t=K_j+1}^{K_j^0}X_{Tt}(\Sigma^0)^{-1}u_t)
\end{aligned}
$$

$$
\begin{aligned}
& rlr_T^2(K_1, ..., K_m, \beta_{(j)}, \Sigma_{(j)}) \\
= \; & -(1/2)\textstyle\sum_{j=1}^m (\beta^0_{(j+1)} - \beta^0_{(j)})'(\sum_{t=K_j+1}^{K_j^0} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(j+1)} - \beta^0_{(j)}) \\
& -\textstyle\sum_{j=1}^{m-1}\sum_{i=j+1}^m (\beta^0_{(j+1)} - \beta^0_{(j)})'(\sum_{t=K_i+1}^{K_i^0} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(i+1)} - \beta^0_{(i)}) \\
& +\textstyle\sum_{j=1}^m (\beta_{(j+1)} - \beta^0_{(j)})'(\sum_{t=K_j+1}^{K_j^0} X_{Tt}(\Sigma^0)^{-1}u_t) + o_p(1)
\end{aligned}
$$

Case 2: $K_1^0 < ... < K_m^0 \leq K_1 < ... < K_m$. We first have,

$$
\begin{aligned}
(B.1) \; = \; & \textstyle\sum_{j=2}^m \sum_{t=K_{j-1}^0+1}^{K_j^0} (Y_t - X'_{Tt}\beta_{(1)})'(\Sigma_{(1)})^{-1}(Y_t - X'_{Tt}\beta_{(1)}) \\
& -\textstyle\sum_{j=2}^m \sum_{t=K_{j-1}^0+1}^{K_j^0} (Y_t - X'_{Tt}\beta_{(j)})'(\Sigma_{(j)})^{-1}(Y_t - X'_{Tt}\beta_{(j)}) \\
& +\textstyle\sum_{j=2}^m \sum_{t=K_{j-1}+1}^{K_j} (Y_t - X'_{Tt}\beta_{(j)})'(\Sigma_{(j)})^{-1}(Y_t - X'_{Tt}\beta_{(j)}) \\
& +\textstyle\sum_{t=K_m^0+1}^{K_m} (Y_t - X'_{Tt}\beta_{(m+1)})'(\Sigma_{(m+1)})^{-1}(Y_t - X'_{Tt}\beta_{(m+1)}) \\
& +\textstyle\sum_{t=K_m^0+1}^{K_1} (Y_t - X'_{Tt}\beta_{(1)})'(\Sigma_{(1)})^{-1}(Y_t - X'_{Tt}\beta_{(1)})
\end{aligned}
$$

Using arguments similar to those for Case 1, we get

$$
\begin{aligned}
& rlr_T^2(K_1, ..., K_m, \beta_{(j)}, \Sigma_{(j)}) \\
= \; & -(1/2)\textstyle\sum_{j=1}^m (\beta^0_{(j+1)} - \beta^0_{(j)})'(\sum_{t=K_j^0+1}^{K_j} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(j+1)} - \beta^0_{(j)}) \\
& -\textstyle\sum_{j=1}^{m-1}\sum_{i=j+1}^m (\beta^0_{(j+1)} - \beta^0_{(j)})'(\sum_{t=K_j^0+1}^{K_j} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(i+1)} - \beta^0_{(i)}) \\
& -\textstyle\sum_{j=1}^m (\beta_{(j+1)} - \beta^0_{(j)})'(\sum_{t=K_j^0+1}^{K_j} X_{Tt}(\Sigma^0)^{-1}u_t) + o_p(1)
\end{aligned}
$$

Case 3: $K_1 < ... < K_b \leq K_1^0 < ... < K_m^0 \leq K_{b+1} < ... < K_m$. In this case,

$$
\begin{aligned}
(B.1) \; = \; & -\textstyle\sum_{t=K_1+1}^{K_1^0} (Y_t - X'_{Tt}\beta_{(1)})'(\Sigma_{(1)})^{-1}(Y_t - X'_{Tt}\beta_{(1)}) \\
& +\textstyle\sum_{j=2}^b \sum_{t=K_{j-1}+1}^{K_j} (Y_t - X'_{Tt}\beta_{(j)})'(\Sigma_{(j)})^{-1}(Y_t - X'_{Tt}\beta_{(j)}) \\
& -\textstyle\sum_{j=2}^m \sum_{t=K_{j-1}^0+1}^{K_j^0} (Y_t - X'_{Tt}\beta_{(j)})'(\Sigma_{(j)})^{-1}(Y_t - X'_{Tt}\beta_{(j)}) \\
& +\textstyle\sum_{t=K_b+1}^{K_{b+1}} (Y_t - X'_{Tt}\beta_{(b+1)})'(\Sigma_{(b+1)})^{-1}(Y_t - X'_{Tt}\beta_{(b+1)}) \\
& -\textstyle\sum_{t=K_m^0+1}^{K_m} (Y_t - X'_{Tt}\beta_{(m+1)})'(\Sigma_{(m+1)})^{-1}(Y_t - X'_{Tt}\beta_{(m+1)}) \\
& +\textstyle\sum_{j=b+2}^m \sum_{t=K_{j-1}+1}^{K_j} (Y_t - X'_{Tt}\beta_{(j)})'(\Sigma_{(j)})^{-1}(Y_t - X'_{Tt}\beta_{(j)})
\end{aligned}
$$

so that

$$
\begin{aligned}
& rlr_T^2(K_{1,\ldots,}K_m, \beta_{(j)}, \Sigma_{(j)}) \\
= \; & -(1/2)\textstyle\sum_{j=1}^{b}(\beta^0_{(j+1)} - \beta^0_{(j)})'(\sum_{t=K_j+1}^{K_j^0} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(j+1)} - \beta^0_{(j)}) \\
& - \textstyle\sum_{j=1}^{b-1}\sum_{i=j+1}^{m}(\beta^0_{(j+1)} - \beta^0_{(j)})'(\sum_{t=K_i+1}^{K_i^0} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(i+1)} - \beta^0_{(i)}) \\
& -(1/2)\textstyle\sum_{j=b+1}^{m}(\beta^0_{(j+1)} - \beta^0_{(j)})'(\sum_{t=K_j^0+1}^{K_j} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(j+1)} - \beta^0_{(j)}) \\
& - \textstyle\sum_{j=b+1}^{m-1}\sum_{i=j+1}^{m}(\beta^0_{(j+1)} - \beta^0_{(j)})'(\sum_{t=K_j^0+1}^{K_j} X_{Tt}(\Sigma^0)^{-1}X'_{Tt})(\beta^0_{(i+1)} - \beta^0_{(i)}) \\
& + \textstyle\sum_{j=1}^{b}(\beta_{(j+1)} - \beta^0_{(j)})'\sum_{t=K_j+1}^{K_j^0} X_{Tt}(\Sigma^0)^{-1}u_t \\
& - \textstyle\sum_{j=b+1}^{m}(\beta_{(j+1)} - \beta^0_{(j)})'\sum_{t=K_j^0+1}^{K_j} X_{Tt}(\Sigma^0)^{-1}u_t + o_p(1)
\end{aligned}
$$

Proof of Lemma 1.1: For the case with $s_j > 0$, we have

$$
\begin{aligned}
v_T^2\textstyle\sum_{t=K_j^0+1}^{K_j^0+[s_jv_T^2]} X_{Tt}(\Sigma^0_{(l)})^{-1}X'_{Tt} &= v_T^2\textstyle\sum_{t=K_j^0+1}^{K_j^0+[s_jv_T^2]} S'(h_{tT}\otimes I_n)(\Sigma^0_{(l)})^{-1}(h'_{tT}\otimes I_n)S \\
&= S'(I_q\otimes(\Sigma^0_{(l)})^{-1})[v_T^2\textstyle\sum_{t=K_j^0+1}^{K_j^0+[s_jv_T^2]} h_{tT}h'_{tT}\otimes I_n]S
\end{aligned}
$$

Note that

$$
h_{tT}h'_{tT} = \begin{pmatrix} T^{-1}z_tz'_t & T^{-3/2}tz_t & T^{-1/2}z_tx'_t \\ T^{-1}tz'_t & T^{-2}t^2 & T^{-1}tx'_t \\ T^{-1/2}z'_tx_t & T^{-1}tx_t & x_tx'_t \end{pmatrix}
$$

The limit distribution for each component is as follows:

$$
\begin{aligned}
v_T^2\textstyle\sum_{t=K_j^0+1}^{K_j^0+[s_jv_T^2]} T^{-3/2}tz_t &= T^{-3/2}v_T^2\textstyle\sum_{t=K_j^0+1}^{K_j^0+[s_jv_T^2]}(t-K_j^0+K_j^0)((z_t-z_{K_j^0})+z_{K_j^0}) \\
&= o_p(1) + T^{-3/2}s_jK_j^0z_{K_j^0} \Rightarrow s_j\lambda_j^0\Omega_z^{1/2}W_z(\lambda_j^0)
\end{aligned}
$$

$$
\begin{aligned}
v_T^2\textstyle\sum_{t=K_j^0+1}^{K_j^0+[s_jv_T^2]} T^{-2}t^2 &= v_T^2\textstyle\sum_{t=K_j^0+1}^{K_j^0+[s_jv_T^2]} T^{-2}(t-K_j^0+K_j^0)^2 \\
&= o_p(1) + s_jT^{-2}(K_j^0)^2 \Rightarrow s_j(\lambda_j^0)^2
\end{aligned}
$$

$$
\begin{aligned}
v_T^2\textstyle\sum_{t=K_j^0+1}^{K_j^0+[s_jv_T^2]} T^{-1}tx'_t &= v_T^2\textstyle\sum_{t=K_j^0+1}^{K_j^0+[s_jV_T^2]} T^{-1}(t-K_j^0+K_j^0)x'_t \\
&= o_p(1) + T^{-1}K_j^0x'_t \Rightarrow s_j\lambda_j^0\mu'_{x,j}
\end{aligned}
$$

$$
v_T^2\textstyle\sum_{t=K_j^0+1}^{K_j^0+[s_jV_T^2]} x_tx'_t \Rightarrow s_jQ_{x,j}.
$$

Let

$$D(\lambda_j^0) = \begin{pmatrix} \Omega_z^{1/2} W_z(\lambda_j^0) W_z(\lambda_j^0)' \Omega_z^{1/2} & \lambda_j^0 \Omega_z^{1/2} W_z(\lambda_j^0) & \Omega_z^{1/2} W_z(\lambda_j^0) \mu_{x,j}' \\ \lambda_j^0 W_z(\lambda_j^0)' \Omega_z^{1/2} & (\lambda_j^0)^2 & \lambda_j^0 \mu_{x,j}' \\ \mu_x W_z(\lambda_j^0)' \Omega_z^{1/2} & \lambda_j^0 \mu_{x,j} & Q_{x,j} \end{pmatrix}$$

then

$$v_T^2 \sum_{t=K_j^0+1}^{K_j^0+[s_j v_T^2]} h_{tT} h_{tT}' \Rightarrow s_j D(\lambda_j^0)$$

and

$$v_T^2 \sum_{t=K_j^0+1}^{K_j^0+[s_j v_T^2]} X_{Tt} (\Sigma_{(l)}^0)^{-1} X_{Tt}' \Rightarrow S'(I_q \otimes (\Sigma_{(l)}^0)^{-1})(s_j D(\lambda_j^0) \otimes I_n) S$$

Consider now

$$\begin{aligned} v_T \sum_{t=K_j^0+1}^{K_j^0+[s_j v_T^2]} X_{Tt} (\Sigma_{(l)}^0)^{-1} U_t &= v_T \sum_{t=K_j^0+1}^{K_j^0+[s_j v_T^2]} S'(h_{tT} \otimes I_n)(\Sigma_{(l)}^0)^{-1} (\Sigma_{(j+1)}^0)^{1/2} \eta_t \\ &= S'(I_q \otimes (\Sigma_{(l)}^0)^{-1} (\Sigma_{(j+1)}^0)^{1/2})[v_T \sum_{t=K_j^0+1}^{K_j^0+[s_j v_T^2]} h_{tT} \otimes \eta_t] \end{aligned}$$

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

$$\begin{aligned} v_T \sum_{t=K_j^0+1}^{K_j^0+[s_j v_T^2]} T^{-1/2} z_t \otimes \eta_t &= v_T \sum_{t=K_j^0+1}^{K_j^0+[s_j v_T^2]} T^{-1/2} (z_t - z_{K_j^0} + z_{K_j^0}) \otimes \eta_t \\ &= o_p(1) + v_T \sum_{t=K_j^0+1}^{K_j^0+[s_j v_T^2]} T^{-1/2} z_{K_j^0} \otimes \eta_t \\ &\Rightarrow \Omega_z^{1/2} W_z(\lambda_j^0) \otimes W_{\eta,j}(s_j) \end{aligned}$$

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

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

Hence,

$$v_T \sum_{t=K_j^0+1}^{K_j^0+[s_j v_T^2]} h_{tT} \otimes \eta_t \Rightarrow \begin{pmatrix} \Omega_z^{1/2} W_z(\lambda_j^0) \otimes W_{\eta,j}(s_j) \\ \lambda_j^0 W_{\eta,j}(s_j) \Omega_z^{1/2} \\ M_{x\eta,j}^{1/2} W_{x\eta,j}(s_j) \end{pmatrix}$$

$$
\begin{aligned}
&= \begin{pmatrix} \Omega_z^{1/2} W_z(\lambda_j^0) \otimes I_n & 0 \\ \lambda_j^0 I_n & 0 \\ 0 & M_{x\eta,j} \end{pmatrix} \begin{pmatrix} W_{\eta,j}(s_j) \\ W_{x\eta,j}(s_j) \end{pmatrix} \\
&= \Omega(\lambda_j^0) W_j(s_j)
\end{aligned}
$$

and

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

Proof of Theorem 1.3 ($m = 2$, breaks in coefficients only).

Let $Q(\lambda_j^0) = S'(I_q \otimes (\Sigma^0)^{-1})(D(\lambda_j^0) \otimes I_n)S$ and $\Delta(\lambda_j^0) = S'(I_q \otimes (\Sigma^0)^{-1/2})\Omega(\lambda_j^0)$, then **Case 1** ($K_1 < K_2 \leq K_1^0 < K_2^0$):

$$
\begin{aligned}
rlr_T^2(K_1, K_2) &\Rightarrow H^1(s_1, s_2) \\
&= -(1/2)|s_1|\delta_1' Q(\lambda_1^0)\delta_1 + \delta_1' \Delta(\lambda_1^0) W_1(s_1) \\
&\quad -(1/2)|s_2|\delta_2' Q(\lambda_2^0)\delta_2 + \delta_2' \Delta(\lambda_2^0) W_1(s_2) \\
&\quad -|s_2|\delta_1' Q(\lambda_2^0)\delta_2
\end{aligned}
$$

Case 2 ($K_1 \leq K_1^0 < K_2^0 \leq K_2$):

$$
\begin{aligned}
rlr_T^2(K_1, K_2) &\Rightarrow H^2(s_1, s_2) \\
&= -(1/2)|s_1|\delta_1' Q(\lambda_1^0)\delta_1 + \delta_1' \Delta(\lambda_1^0) W_1(s_1) \\
&\quad -(1/2)|s_2|\delta_2' Q(\lambda_2^0)\delta_2 - \delta_2' \Delta(\lambda_2^0) W_2(s_2)
\end{aligned}
$$

Case 3 ($K_1^0 < K_2^0 \leq K_1 < K_2$):

$$
\begin{aligned}
rlr_T^2(K_1, K_2) &\Rightarrow H^3(s_1, s_2) \\
&= -(1/2)|s_1|\delta_1' Q(\lambda_1^0)\delta_1 - \delta_1' \Delta(\lambda_1^0) W_1(s_1) \\
&\quad -(1/2)|s_2|\delta_2' Q(\lambda_2^0)\delta_2 - \delta_2' \Delta(\lambda_2^0) W_2(s_2) \\
&\quad -|s_2|\delta_1' Q(\lambda_1^0)\delta_2
\end{aligned}
$$

$$\delta_1'\Delta(\lambda_1^0)W_1(s_1) \stackrel{d}{=} (\delta_1'\Delta(\lambda_1^0)\Delta(\lambda_1^0)'\delta_1)^{1/2}B_1(s_1)$$
$$\delta_2'\Delta(\lambda_2^0)W_2(s_2) \stackrel{d}{=} (\delta_2'\Delta(\lambda_2^0)\Delta(\lambda_2^0)'\delta_2)^{1/2}B_2(s_2)$$

and let $\Pi_1 = \delta_1'Q(\lambda_1^0)\delta_1$, $\Pi_2 = \delta_2'Q(\lambda_2^0)\delta_2$, $\Pi_{12}^1 = \delta_1'Q(\lambda_1^0)\delta_2$, $\Pi_{12}^2 = \delta_1'Q(\lambda_2^0)\delta_2$, $\Upsilon_1 = \delta_1'\Delta(\lambda_1^0)\Delta(\lambda_1^0)'\delta_1$ and $\Upsilon_2 = \delta_2'\Delta(\lambda_2^0)\Delta(\lambda_2^0)'\delta_2$, then

$$\begin{aligned}
H^1(s_1,s_2) &= -(1/2)|s_1|\Pi_1 + \Upsilon_1^{1/2}B_1(s_1) \\
&\quad -(1/2)|s_2|\Pi_2 + \Upsilon_2^{1/2}B_2(s_2) - |s_2|\Pi_{12}^2 \\
H^2(s_1,s_2) &= -(1/2)|s_1|\Pi_1 + \Upsilon_1^{1/2}B_1(s_1) \\
&\quad -(1/2)|s_2|\Pi_2 - \Upsilon_2^{1/2}B_2(s_2) \\
H^3(s_1,s_2) &= -(1/2)|s_1|\Pi_1 - \Upsilon_1^{1/2}B_1(s_1) \\
&\quad -(1/2)|s_2|\Pi_2 - \Upsilon_2^{1/2}B_2(s_2) - |s_1|\Pi_{12}^1
\end{aligned}$$

Let $b = \Upsilon_1/\Pi_1^2$, $s_1 = bv_1$ and $s_2 = bv_2$. Applying a changing variables technique as in Bai (1997), we obtain the following results.

Case 1 ($v_1 \le v_2 \le 0$):

$$\begin{aligned}
\Pi_1 v_T^2(\hat{K}_1 - K_1^0, \hat{K}_2 - K_2^0) &\Rightarrow \arg\max_{v_1\le v_2\le 0} -(1/2)|v_1| + (\frac{\Upsilon_1}{\Pi_1})^{1/2}B_1(v_1) \\
&\quad -(1/2)|v_2|\frac{\Pi_2}{\Pi_1} + (\frac{\Upsilon_2}{\Pi_1})^{1/2}B_2(v_2) - |v_2|\frac{\Pi_{12}^2}{\Pi_1}
\end{aligned}$$

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

$$\begin{aligned}
\Pi_1 v_T^2(\hat{K}_1 - K_1^0, \hat{K}_2 - K_2^0) &\Rightarrow \arg\max_{v_1<0,v_2>0} -(1/2)|v_1| + B_1(v_1) \\
&\quad -(1/2)|v_2|\frac{\Pi_2}{\Pi_1} - (\frac{\Upsilon_2}{\Upsilon_1})^{1/2}B_2(v_2)
\end{aligned}$$

Case 3 ($0 \ge v_1 \ge v_2$)

$$\begin{aligned}\Pi_1 v_T^2(\hat{K}_1 - K_1^0, \hat{K}_2 - K_2^0) &\Rightarrow \arg\max_{0\geq v_1 \geq v_2} -(1/2)|v_1| - B_1(v_1) \\ &\quad -(1/2)|v_2|\frac{\Pi_2}{\Pi_1} - (\frac{\Upsilon_2}{\Upsilon_1})^{1/2} B_2(v_2) - |v_1|\frac{\Pi_{12}}{\Pi_1}\end{aligned}$$

Proof of Theorem 1.3 (multiple breaks in coefficients only):

For **Case 1** ($K_1 < ... < K_m \leq K_1^0 < ... < K_m^0$):

$$rlr_T^2(K_1, ..., K_m) \Rightarrow H^1(s_1, ..., s_m)$$

$$\begin{aligned}rlr_T^2(K_1, ..., K_m) &\Rightarrow H^1(s_1, ..., s_m) \\ &= -(1/2)\textstyle\sum_{j=1}^m |s_j|\delta_j' Q(\lambda_j^0)\delta_j \\ &\quad -\textstyle\sum_{j=1}^{m-1}\sum_{i=j+1}^m |s_i|\delta_j' Q(\lambda_i^0)\delta_i + \sum_{j=1}^m \delta_j'\Delta(\lambda_j^0) W_j(s_j)\end{aligned}$$

For **Case 2** ($K_1^0 < ... < K_m^0 \leq K_1 < ... < K_m$):

$$\begin{aligned}rlr_T^2(K_1, ..., K_m) &\Rightarrow H^2(s_1, ..., s_m) = -(1/2)\textstyle\sum_{j=1}^m |s_j|\delta_j' Q(\lambda_j^0)\delta_j \\ &\quad -\textstyle\sum_{j=1}^{m-1}\sum_{i=j+1}^m |s_j|\delta_j' Q(\lambda_j^0)\delta_i - \sum_{j=1}^m \delta_j'\Delta(\lambda_j^0) W_j(s_j)\end{aligned}$$

For **Case 3** ($K_1 < ... < K_b \leq K_1^0 < ... < K_m^0 < K_{b+1} < ... < K_m$):

$$\begin{aligned}rlr_T^2(K_1, ..., K_m) &\Rightarrow H^3(s_1, ..., s_m) \\ &= -(1/2)\textstyle\sum_{j=1}^b |s_j|\delta_j' Q(\lambda_j^0)\delta_j - \sum_{j=1}^{b-1}\sum_{i=j+1}^b |s_i|\delta_j' Q(\lambda_i^0)\delta_i \\ &\quad +\textstyle\sum_{j=1}^b \delta_j'\Delta(\lambda_j^0) W_j(s_j) - (1/2)\sum_{j=b+1}^b |s_j|\delta_j' Q(\lambda_j^0)\delta_j \\ &\quad -\textstyle\sum_{j=b+1}^{m-1}\sum_{i=j+1}^m |s_j|\delta_j' Q(\lambda_j^0)\delta_i - \sum_{j=b+1}^m \delta_j'\Delta(\lambda_j^0) W_j(s_j)\end{aligned}$$

Let $\Pi_j = \delta_j' Q(\lambda_1^0)\delta_j$, $\Upsilon_j = \delta_j'\Delta(\lambda_1^0)\Delta(\lambda_1^0)'\delta_j$, $\Pi_{ij}^i = \delta_i' Q(\lambda_i^0)\delta_j$,$\Pi_{ij}^j = \delta_i' Q(\lambda_j^0)\delta_j, b = \Upsilon_1/\Pi_1^2$ and $s_j = bv_j$. We then have the following results.

For **Case 1** ($v_1 \leq ... \leq v_m \leq 0$):

$$\Pi_1 v_T^2(\hat{K}_1 - K_1^0, ..., \hat{K}_m - K_m^0) \Rightarrow \arg\max_{v_1 \leq ... \leq v_m \leq 0} -(1/2)\textstyle\sum_{j=1}^m |v_j|\frac{\Pi_j}{\Pi_1}$$

$$+\sum_{j=1}^{m}(\frac{\Upsilon_j}{\Pi_1})^{1/2}B_j(v_j) - \sum_{j=1}^{m-1}\sum_{i=j+1}^{m}|v_i|\frac{\Pi_{ij}^{i}}{\Pi_1}$$

For **Case 2** ($0 \leq v_1 \leq ... \leq v_m$):

$$\Pi_1 v_T^2(\hat{K}_1 - K_1^0, ..., \hat{K}_m - K_m^0) \quad \Rightarrow \quad \arg\max_{0\leq v_1\leq ...\leq v_m} -(1/2)\sum_{j=1}^{m}|v_j|\frac{\Pi_j}{\Pi_1}$$
$$-\sum_{j=1}^{m}(\frac{\Upsilon_j}{\Pi_1})^{1/2}B_j(v_j) - \sum_{j=1}^{m-1}\sum_{i=j+1}^{m}|v_j|\frac{\Pi_{ij}^{j}}{\Pi_1}$$

For **Case 3** ($v_1 \leq ... \leq v_b \leq 0 \leq v_{b+1} \leq ... \leq v_m$):

$$\Pi_1 v_T^2(\hat{K}_1 - K_1^0, ..., \hat{K}_m - K_m^0)$$
$$\Rightarrow \quad \arg\max_{v_1\leq ...\leq v_b\leq 0\leq v_{b+1}\leq ...\leq v_m} -(1/2)\sum_{j=1}^{m}|v_j|\frac{\Pi_j}{\Pi_1} - \sum_{j=1}^{b-1}\sum_{i=j+1}^{b}|v_i|\frac{\Pi_{ij}^{i}}{\Pi_1}$$
$$-\sum_{j=b+1}^{m-1}\sum_{i=j+1}^{m}|v_j|\frac{\Pi_{ij}^{j}}{\Pi_1} + \sum_{j=1}^{b}(\frac{\Upsilon_j}{\Pi_1})^{1/2}B_j(v_j) - \sum_{j=b+1}^{m}(\frac{\Upsilon_j}{\Pi_1})^{1/2}B_j(v_j)$$

Results for Multiple Breaks in Both Coefficients and Variance-Covariance Matrix

Theorem 1.4 *Under Assumptions A1-A9, we have*

$$\begin{aligned}\max_{(K,\beta,\Sigma)} rlr_T &= \max_{(\beta,\Sigma)\in C_M, K^0}[rlr_T^1(K^0,\beta,\Sigma)+\lambda' g(\beta, vec(\Sigma))] \\ &\quad + \max_{K\in C_M,(\beta^0,\Sigma^0)} rlr_T^2(K,\beta,\Sigma)+o_p(1)\end{aligned}$$

where

$$\begin{aligned}rlr_T^1(K^0,\beta,\Sigma) &= -(1/2)[\textstyle\sum_{j=1}^{m+1}\sum_{t=K_{j-1}^0+1}^{K_j^0}(Y_t - X'_{Tt}\beta_{(j)})'(\Sigma_{(j)})^{-1}(Y_t - X'_{Tt}\beta_{(j)}) \\ &\quad -\textstyle\sum_{j=1}^{m+1}\sum_{t=K_{j-1}^0+1}^{K_j^0}(Y_t - X'_{Tt}\beta^0_{(j)})'(\Sigma^0_{(j)})^{-1}(Y_t - X'_{Tt}\beta^0_{(j)})]\end{aligned}$$

and for Case 1 ($K_1 < ... < K_m \le K_1^0 < ... < K_m^0$):

$$\begin{aligned}&+(1/2)\textstyle\sum_{j=1}^{m-1}(\beta^0_{(j+1)}-\beta^0_{(j)})' \\ &(\textstyle\sum_{l=j+1}^{m}\sum_{t=K_l+1}^{K_l^0}X_{Tt}(\Sigma^0_{(l)})^{-1}X'_{Tt})(\beta^0_{(j+1)}-\beta^0_{(j)}) \\ &-\textstyle\sum_{j=1}^{m-1}\sum_{i=j+1}^{m}(\beta^0_{(j+1)}-\beta^0_{(j)})' \\ &(\textstyle\sum_{l=i}^{m}\sum_{t=K_l+1}^{K_l^0}X_{Tt}(\Sigma^0_{(l+1)})^{-1}X'_{Tt})(\beta^0_{(i+1)}-\beta^0_{(i)}) \\ &+\textstyle\sum_{j=1}^{m-2}\sum_{i=j+1}^{m-1}(\beta^0_{(j+1)}-\beta^0_{(j)})' \\ &(\textstyle\sum_{l=i+1}^{m}\sum_{t=K_l+1}^{K_l^0}X_{Tt}(\Sigma^0_{(l)})^{-1}X'_{Tt})(\beta^0_{(i+1)}-\beta^0_{(i)}) \\ &+\textstyle\sum_{j=1}^{m}(\beta^0_{(j+1)}-\beta^0_{(j)})' \\ &(\textstyle\sum_{l=j}^{m}\sum_{t=K_l+1}^{K_l^0}X_{Tt}(\Sigma^0_{(l+1)})^{-1}u_t) \\ &+\textstyle\sum_{j=1}^{m-1}(\beta^0_{(j+1)}-\beta^0_{(j)})' \\ &(\textstyle\sum_{l=j+1}^{m}\sum_{t=K_{l+1}+1}^{K_{l+1}^0}X_{Tt}(\Sigma^0_{(l)})^{-1}u_t)+o_p(1)\end{aligned}$$

$$\begin{aligned}&rlr_T^2(K_1,...,K_m,\beta_{(j)},\Sigma_{(j)}) \\ =\ &(1/2)\textstyle\sum_{j=1}^{m} tr((\Sigma^0_{(j)})^{1/2}(\Sigma^0_{(j+1)})^{-1}\Phi_{(j)} \\ &(\Sigma^0_{(j)})^{-1/2}(v_T\textstyle\sum_{t=K_j+1}^{K_j^0}\eta_t\eta'_t - I_n)) \\ &-(1/4)\textstyle\sum_{j=1}^{m}|v_T^2(K_j-K_j^0) \\ &|tr((\Sigma^0_{(j+1)})^{-1}\Phi_{(j)}(\Sigma^0_{(j+1)})^{-1}\Phi_{(j)}) \\ &-(1/2)\textstyle\sum_{j=1}^{m}(\beta^0_{(j+1)}-\beta^0_{(j)})' \\ &(\textstyle\sum_{l=j}^{m}\sum_{t=K_l+1}^{K_l^0}X_{Tt}(\Sigma^0_{(l+1)})^{-1}X'_{Tt})(\beta^0_{(j+1)}-\beta^0_{(j)})\end{aligned}$$

For Case 2 ($K_1^0 < ... < K_m^0 \le K_1 < ... < K_m$):

$$
\begin{aligned}
& rlr_T^2(K_{1,...}, K_m, \beta_{(j)}, \Sigma_{(j)}) \\
= \; & -(1/2)\textstyle\sum_{j=1}^m tr((\Sigma_{(j+1)}^0)^{1/2}(\Sigma_{(j)}^0)^{-1}\Phi_{(j)}(\Sigma_{(j+1)}^0)^{-1/2}(v_T\sum_{t=K_j^0+1}^{K_j}\eta_t\eta_t' - I_n)) \\
& -(1/4)\textstyle\sum_{j=1}^m |v_T^2(K_j - K_j^0)| tr((\Sigma_{(j)}^0)^{-1}\Phi_{(j)}(\Sigma_{(j)}^0)^{-1}\Phi_{(j)}) \\
& -(1/2)\textstyle\sum_{j=1}^m(\beta_{(j+1)}^0 - \beta_{(j)}^0)'(\sum_{l=1}^j\sum_{t=K_l^0+1}^{K_l} X_{Tt}(\Sigma_{(l)}^0)^{-1}X_{Tt}')(\beta_{(j+1)}^0 - \beta_{(j)}^0) \\
& +(1/2)\textstyle\sum_{j=2}^m(\beta_{(j+1)}^0 - \beta_{(j)}^0)'(\sum_{l=1}^{j-1}\sum_{t=K_l^0+1}^{K_l} X_{Tt}(\Sigma_{(l+1)}^0)^{-1}X_{Tt}')(\beta_{(j+1)}^0 - \beta_{(j)}^0) \\
& +\textstyle\sum_{j=2}^{m-1}\sum_{i=j+1}^m(\beta_{(j+1)}^0 - \beta_{(j)}^0)'(\sum_{l=1}^{j-1}\sum_{t=K_l^0+1}^{K_l} X_{Tt}(\Sigma_{(l+1)}^0)^{-1}X_{Tt}')(\beta_{(i+1)}^0 - \beta_{(i)}^0) \\
& -\textstyle\sum_{j=1}^{m-1}\sum_{i=j+1}^m(\beta_{(j+1)}^0 - \beta_{(j)}^0)'(\sum_{l=1}^{j}\sum_{t=K_l^0+1}^{K_l} X_{Tt}(\Sigma_{(l)}^0)^{-1}X_{Tt}')(\beta_{(i+1)}^0 - \beta_{(i)}^0) \\
& -\textstyle\sum_{j=1}^m(\beta_{(j+1)}^0 - \beta_{(j)}^0)'(\sum_{l=1}^j\sum_{t=K_l^0+1}^{K_l} X_{Tt}(\Sigma_{(l)}^0)^{-1}u_t) \\
& +\textstyle\sum_{j=2}^m(\beta_{(j+1)}^0 - \beta_{(j)}^0)'(\sum_{l=1}^{j-1}\sum_{t=K_l^0+1}^{K_l} X_{Tt}(\Sigma_{(l+1)}^0)^{-1}u_t) + o_p(1)
\end{aligned}
$$

For Case 3 ($K_1 < ... < K_b \le K_1^0 < ... < K_m^0 \le K_{b+1} < \cdots < K_m$ for some $1 \le b \le m$):

$$
\begin{aligned}
& rlr_T^2(K_{1,...}, K_m, \beta_{(j)}, \Sigma_{(j)}) \\
= \; & (1/2)\textstyle\sum_{j=1}^b tr((\Sigma_{(j)}^0)^{1/2}(\Sigma_{(j+1)}^0)^{-1}\Phi_{(j)} \\
& (\Sigma_{(j)}^0)^{-1/2}(v_T\textstyle\sum_{t=K_j+1}^{K_j^0}\eta_t\eta_t' - I_n)) \\
& -(1/4)\textstyle\sum_{j=1}^b |v_T^2(K_j - K_j^0)| tr((\Sigma_{(j+1)}^0)^{-1}\Phi_{(j)}(\Sigma_{(j+1)}^0)^{-1}\Phi_{(j)}) \\
& -(1/2)\textstyle\sum_{j=b+1}^m tr((\Sigma_{(j+1)}^0)^{1/2}(\Sigma_{(j)}^0)^{-1} \\
& \Phi_{(j)}(\Sigma_{(j+1)}^0)^{-1/2}(v_T\textstyle\sum_{t=K_j^0+1}^{K_j}\eta_t\eta_t' - I_n)) \\
& -(1/4)\textstyle\sum_{j=b+1}^m |v_T^2(K_j - K_j^0)| \\
& tr((\Sigma_{(j)}^0)^{-1}\Phi_{(j)}(\Sigma_{(j)}^0)^{-1}\Phi_{(j)}) \\
& -(1/2)\textstyle\sum_{j=1}^b(\beta_{(j+1)}^0 - \beta_{(j)}^0)'(\sum_{l=j}^b\sum_{t=K_l+1}^{K_l^0} X_{Tt}(\Sigma_{(l+1)}^0)^{-1}X_{Tt}')(\beta_{(j+1)}^0 - \beta_{(j)}^0) \\
& +(1/2)\textstyle\sum_{j=1}^{b-1}(\beta_{(j+1)}^0 - \beta_{(j)}^0)'(\sum_{l=j+1}^b\sum_{t=K_l+1}^{K_l^0} X_{Tt}(\Sigma_{(l)}^0)^{-1}X_{Tt}')(\beta_{(j+1)}^0 - \beta_{(j)}^0) \\
& -\textstyle\sum_{j=1}^{b-1}\sum_{i=j+1}^b(\beta_{(j+1)}^0 - \beta_{(j)}^0)'(\sum_{l=i}^b\sum_{t=K_l+1}^{K_l^0} X_{Tt}(\Sigma_{(l+1)}^0)^{-1}X_{Tt}')(\beta_{(i+1)}^0 - \beta_{(i)}^0) \\
& +\textstyle\sum_{j=1}^{b-2}\sum_{i=j+1}^{b-1}(\beta_{(j+1)}^0 - \beta_{(j)}^0)'
\end{aligned}
$$

Remark 1.5 *With only one pair of locally ordered breaks, Theorem B.1 involves the following simpler expressions.*

For Case 1 ($K_1 < K_2 \leq K_1^0 < K_2^0$):

$$\begin{aligned}
& rlr_T^2(K_1, K_2, \beta_{(j)}, \Sigma_{(j)}) \\
= \quad & (1/2)\sum_{j=1}^2 tr((\Sigma_{(j)}^0)^{1/2}(\Sigma_{(j+1)}^0)^{-1}\Phi_{(j)}(\Sigma_{(j)}^0)^{-1/2}(v_T \sum_{t=K_j+1}^{K_j^0} \eta_t\eta_t' - I_n)) \\
& +(1/4)\sum_{j=1}^2 |v_T^2(K_j - K_j^0)| tr((\Sigma_{(j+1)}^0)^{-1}\Phi_{(j)}(\Sigma_{(j+1)}^0)^{-1}\Phi_{(j)}) \\
& -(1/2)(\beta_{(2)}^0 - \beta_{(1)}^0)'(\sum_{t=K_1+1}^{K_1^0} X_{Tt}(\Sigma_{(2)}^0)^{-1}X_{Tt}')(\beta_{(2)}^0 - \beta_{(1)}^0) \\
& +(1/2)(\beta_{(2)}^0 - \beta_{(1)}^0)'(\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma_{(2)}^0)^{-1}X_{Tt}')(\beta_{(2)}^0 - \beta_{(1)}^0) \\
& -(1/2)(\beta_{(2)}^0 - \beta_{(1)}^0)'(\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma_{(3)}^0)^{-1}X_{Tt}')(\beta_{(2)}^0 - \beta_{(1)}^0) \\
& -(1/2)(\beta_{(3)}^0 - \beta_{(2)}^0)'(\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma_{(3)}^0)^{-1}X_{Tt}')(\beta_{(3)}^0 - \beta_{(2)}^0) \\
& -(\beta_{(3)}^0 - \beta_{(2)}^0)'(\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma_{(3)}^0)^{-1}X_{Tt}')(\beta_{(2)}^0 - \beta_{(1)}^0) \\
& +(\beta_{(2)}^0 - \beta_{(1)}^0)'(\sum_{t=K_1+1}^{K_1^0} X_{Tt}(\Sigma_{(2)}^0)^{-1}u_t) \\
& +(\beta_{(3)}^0 - \beta_{(2)}^0)'(\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma_{(3)}^0)^{-1}u_t) \\
& -(\beta_{(2)}^0 - \beta_{(1)}^0)'(\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma_{(2)}^0)^{-1}u_t) \\
& +(\beta_{(2)}^0 - \beta_{(1)}^0)'(\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma_{(3)}^0)^{-1}u_t) + o_p(1)
\end{aligned}$$

For Case 2 ($K_1^0 < K_2^0 \leq K_1 < K_2$):

$$\begin{aligned}
& rlr_T^2(K_1, K_2, \beta_{(j)}, \Sigma_{(j)}) \\
= \quad & -(1/2)\sum_{j=1}^2 tr((\Sigma_{(j+1)}^0)^{1/2}(\Sigma_{(j)}^0)^{-1}\Phi_{(j)}(\Sigma_{(j+1)}^0)^{-1/2}(v_T \sum_{t=K_j^0+1}^{K_j} \eta_t\eta_t' - I_n)) \\
& -(1/4)\sum_{j=1}^2 |v_T^2(K_j - K_j^0)| tr((\Sigma_{(j)}^0)^{-1}\Phi_{(j)}(\Sigma_{(j)}^0)^{-1}\Phi_{(j)}) \\
& -(1/2)(\beta_{(2)}^0 - \beta_{(1)}^0)'(\sum_{t=K_1^0+1}^{K_1} X_{Tt}(\Sigma_{(1)}^0)^{-1}X_{Tt}')(\beta_{(2)}^0 - \beta_{(1)}^0) \\
& -(1/2)(\beta_{(3)}^0 - \beta_{(2)}^0)'(\sum_{t=K_1^0+1}^{K_1} X_{Tt}(\Sigma_{(1)}^0)^{-1}X_{Tt}')(\beta_{(3)}^0 - \beta_{(2)}^0) \\
& -(1/2)(\beta_{(3)}^0 - \beta_{(2)}^0)'(\sum_{t=K_2^0+1}^{K_2} X_{Tt}(\Sigma_{(2)}^0)^{-1}X_{Tt}')(\beta_{(3)}^0 - \beta_{(2)}^0) \\
& +(1/2)(\beta_{(3)}^0 - \beta_{(2)}^0)'(\sum_{t=K_1^0+1}^{K_1} X_{Tt}(\Sigma_{(2)}^0)^{-1}X_{Tt}')(\beta_{(3)}^0 - \beta_{(2)}^0) \\
& -(\beta_{(3)}^0 - \beta_{(2)}^0)'(\sum_{t=K_1^0+1}^{K_1} X_{Tt}(\Sigma_{(1)}^0)^{-1}X_{Tt}')(\beta_{(2)}^0 - \beta_{(1)}^0) \\
& -(\beta_{(2)}^0 - \beta_{(1)}^0)'(\sum_{t=K_1^0+1}^{K_1} X_{Tt}(\Sigma_{(1)}^0)^{-1}u_t) \\
& -(\beta_{(3)}^0 - \beta_{(2)}^0)'(\sum_{t=K_1^0+1}^{K_1} X_{Tt}(\Sigma_{(1)}^0)^{-1}u_t)
\end{aligned}$$

$$-(\beta^0_{(3)} - \beta^0_{(2)})'(\sum_{t=K_2^0+1}^{K_2} X_{Tt}(\Sigma^0_{(2)})^{-1}u_t)$$
$$+(\beta^0_{(3)} - \beta^0_{(2)})'(\sum_{t=K_1^0+1}^{K_1} X_{Tt}\Sigma^{0-1}_{(2)}u_t) + o_p(1)$$

For Case 3 ($K_1 \leq K_1^0 < K_2^0 \leq K_2$):

$$\begin{aligned}
& rlr_T^2(K_1, K_2, \beta_{(j)}, \Sigma_{(j)}) \\
= \; & (1/2)tr((\Sigma^0_{(1)})^{-1/2}(\Sigma^0_{(2)})^{-1}\Phi_{(1)}(\Sigma^0_{(1)})^{-1/2}(v_T \sum_{t=K_1+1}^{K_1^0} \eta_t\eta_t' - I_n)) \\
& +(1/4)|v_T^2(K_1 - K_1^0)|tr((\Sigma^0_{(2)})^{-1}\Phi_{(1)}(\Sigma^0_{(2)})^{-1}\Phi_{(1)}) \\
& -(1/2)tr((\Sigma^0_{(3)})^{1/2}(\Sigma^0_{(2)})^{-1}\Phi_{(2)}(\Sigma^0_{(3)})^{-1}(v_T \sum_{t=K_2^0+1}^{K_2} \eta_t\eta_t' - I_n)) \\
& -(1/4)|v_T^2(K_2 - K_2^0)|tr((\Sigma^0_{(2)})^{-1}\Phi_{(2)}(\Sigma^0_{(2)})^{-1}\Phi_{(2)}) \\
& -(1/2)(\beta^0_{(2)} - \beta^0_{(1)})'(\sum_{t=K_1+1}^{K_1^0} X_{Tt}(\Sigma^0_{(2)})^{-1}X_{Tt}')(\beta^0_{(2)} - \beta^0_{(1)}) \\
& -(1/2)(\beta^0_{(3)} - \beta^0_{(2)})'(\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma^0_{(3)})^{-1}X_{Tt}')(\beta^0_{(3)} - \beta^0_{(2)}) \\
& +(\beta^0_{(2)} - \beta^0_{(1)})'(\sum_{t=K_1+1}^{K_1^0} X_{Tt}(\Sigma^0_{(2)})^{-1}u_t) \\
& -(\beta^0_{(3)} - \beta^0_{(2)})'(\sum_{t=K_2^0+1}^{K_2} X_{Tt}(\Sigma^0_{(2)})^{-1}u_t) + o_p(1)
\end{aligned}$$

Proof of Theorem 1.4: We first consider a more detailed proof for the case with one pair of locally ordered breaks ($m = 2$) and later outline the main changes for the more general case.

For **Case 1** ($K_1 < K_2 \leq K_1^0 < K_2^0$), we have from the proof of Theorem 2:

$$\begin{aligned}
(A.1) \; = \; & \sum_{j=1}^{2}(K_j - K_j^0)(\log|\Sigma_{(j)}| - \log|\Sigma_{(j+1)}|) \\
= \; & \sum_{j=1}^{2}(K_j - K_j^0)(\log|\Sigma^0_{(j)}| - \log|\Sigma^0_{(j+1)}|) \\
& + \sum_{j=1}^{2}(K_j - K_j^0)(T^{-1/2}tr((\Sigma^0_{(j)})^{-1}\Sigma^*_{(j)}) \\
& -T^{-1/2}tr((\Sigma^0_{(j+1)})^{-1}\Sigma^*_{(j+1)})) + o_p(1)
\end{aligned}$$

$$(B.1) \; = \; -\sum_{t=K_1+1}^{K_1^0}(Y_t - X_{Tt}'\beta_{(1)})'(\Sigma_{(1)})^{-1}(Y_t - X_{Tt}'\beta_{(1)}) \quad \text{(E.1)}$$
$$+\sum_{t=K_1+1}^{K_2}(Y_t - X_{Tt}'\beta_{(2)})'(\Sigma_{(2)})^{-1}(Y_t - X_{Tt}'\beta_{(2)}) \quad \text{(E.2)}$$
$$+\sum_{t=K_2+1}^{K_1^0}(Y_t - X_{Tt}'\beta_{(3)})'(\Sigma_{(3)})^{-1}(Y_t - X_{Tt}'\beta_{(3)}) \quad \text{(E.3)}$$

and

$$+\sum_{t=K_1^0+1}^{K_2^0}(Y_t - X_{Tt}'\beta_{(3)})'(\Sigma_{(3)})^{-1}(Y_t - X_{Tt}'\beta_{(3)}) \tag{E.4}$$
$$-\sum_{t=K_1^0+1}^{K_2^0}(Y_t - X_{Tt}'\beta_{(2)})'(\Sigma_{(2)})^{-1}(Y_t - X_{Tt}'\beta_{(2)}) \tag{E.5}$$

We again expand each term of (B.1).

$$\begin{aligned}
(E.1) &= -\sum_{t=K_1+1}^{K_1^0}(u_t + X_{Tt}'\beta_{(1)}^0 - X_{Tt}'\beta_{(1)})' \\
&\quad (\Sigma_{(1)})^{-1}(u_t + X_{Tt}'\beta_{(1)}^0 - X_{Tt}'\beta_{(1)}) \\
&= -\sum_{t=K_1+1}^{K_1^0} u_t'(\Sigma_{(1)}^0)^{-1}u_t \\
&\quad + T^{-1/2}\sum_{t=K_1+1}^{K_1^0} u_t'(\Sigma_{(1)}^0)^{-1}\Sigma_{(1)}^*(\Sigma_{(1)}^0)^{-1}u_t + o_p(1)
\end{aligned}$$

$$\begin{aligned}
(E.2) &= \sum_{t=K_1+1}^{K_2}(u_t + X_{Tt}'\beta_{(1)}^0 - X_{Tt}'\beta_{(2)})' \\
&\quad (\Sigma_{(2)})^{-1}(u_t + X_{Tt}'\beta_{(1)}^0 - X_{Tt}'\beta_{(2)}) \\
&= \sum_{t=K_1+1}^{K_2} u_t'(\Sigma_{(2)}^0)^{-1}u_t - T^{-1/2}\sum_{t=K_1+1}^{K_2} u_t'(\Sigma_{(2)}^0)^{-1}\Sigma_{(2)}^*(\Sigma_{(2)}^0)^{-1}u_t \\
&\quad + (\beta_{(2)}^0 - \beta_{(1)}^0)'(\sum_{t=K_1+1}^{K_2} X_{Tt}(\Sigma_{(2)}^0)^{-1}X_{Tt}')(\beta_{(2)}^0 - \beta_{(1)}^0) \\
&\quad - 2(\beta_{(2)}^0 - \beta_{(1)}^0)'(\sum_{t=K_1+1}^{K_2} X_{Tt}(\Sigma_{(2)}^0)^{-1}u_t) + o_p(1)
\end{aligned}$$

$$\begin{aligned}
(E.3) &= \sum_{t=K_2+1}^{K_1^0}(u_t + X_{Tt}'\beta_{(1)}^0 - X_{Tt}'\beta_{(3)})' \\
&\quad (\Sigma_{(3)})^{-1}(u_t + X_{Tt}'\beta_{(1)}^0 - X_{Tt}'\beta_{(3)}) \\
&= \sum_{t=K_2+1}^{K_1^0} u_t'(\Sigma_{(3)}^0)^{-1}u_t \\
&\quad - T^{-1/2}\sum_{t=K_2+1}^{K_1^0} u_t'(\Sigma_{(3)}^0)^{-1}\Sigma_{(3)}^*(\Sigma_{(3)}^0)^{-1}u_t \\
&\quad + (\beta_{(3)}^0 - \beta_{(1)}^0)'(\sum_{t=K_2+1}^{K_1^0} X_{Tt}(\Sigma_{(3)}^0)^{-1}X_{Tt}')(\beta_{(3)}^0 - \beta_{(1)}^0) \\
&\quad - 2(\beta_{(3)}^0 - \beta_{(1)}^0)'(\sum_{t=K_2+1}^{K_1^0} X_{Tt}(\Sigma_{(3)}^0)^{-1}u_t) + o_p(1)
\end{aligned}$$

$$\begin{aligned}
(E.5) &= -\sum_{t=K_1^0+1}^{K_2^0}(u_t + X_{Tt}'\beta_{(2)}^0 - X_{Tt}'\beta_{(2)})'(\Sigma_{(2)})^{-1}(u_t + X_{Tt}'\beta_{(2)}^0 - X_{Tt}'\beta_{(2)}) \\
&= -\sum_{t=K_1^0+1}^{K_2^0} u_t'(\Sigma_{(2)}^0)^{-1}u_t + T^{-1/2}\sum_{t=K_1^0+1}^{K_2^0} u_t'(\Sigma_{(2)}^0)^{-1}\Sigma_{(2)}^*(\Sigma_{(2)}^0)^{-1}u_t + o_p(1)
\end{aligned}$$

$$
\begin{aligned}
(E.4) &= \textstyle\sum_{t=K_1^0+1}^{K_2^0}(u_t + X_{Tt}'\beta_{(2)}^0 - X_{Tt}'\beta_{(3)})' \\
&\quad (\Sigma_{(3)})^{-1}(u_t + X_{Tt}'\beta_{(2)}^0 - X_{Tt}'\beta_{(3)}) \\
&= \textstyle\sum_{t=K_1^0+1}^{K_2^0} u_t'(\Sigma_{(3)}^0)^{-1}u_t - T^{-1/2}\sum_{t=K_1^0+1}^{K_2^0} u_t'(\Sigma_{(3)}^0)^{-1}\Sigma_{(3)}^*(\Sigma_{(3)}^0)^{-1}u_t \\
&\quad +(\beta_{(3)}^0 - \beta_{(2)}^0)'(\textstyle\sum_{t=K_1^0+1}^{K_2^0} X_{Tt}(\Sigma_{(3)}^0)^{-1}X_{Tt}')(\beta_{(3)}^0 - \beta_{(2)}^0) \\
&\quad -2(\beta_{(3)}^0 - \beta_{(2)}^0)'(\textstyle\sum_{t=K_1^0+1}^{K_2^0} X_{Tt}(\Sigma_{(3)}^0)^{-1}u_t) + o_p(1)
\end{aligned}
$$

Hence, $rlr_T^2(K_1,K_2,\beta,\Sigma) = -(1/2)((A.1)+(B.1))$ consists of the following four parts, labelled I through IV:

$$
\begin{aligned}
I &= -(1/2)[\textstyle\sum_{j=1}^2(K_j - K_j^0)(\log|\Sigma_{(j)}^0| - \log|\Sigma_{(j+1)}^0|) \\
&\quad + \textstyle\sum_{j=1}^2\sum_{t=K_j+1}^{K_j^0}(u_t'(\Sigma_{(j+1)}^0)^{-1}u_t - u_t'(\Sigma_{(j)}^0)^{-1}u_t)] \\
&= -(1/2)\textstyle\sum_{j=1}^2[(K_j - K_j^0)tr((\Sigma_{(j+1)}^0)^{-1}(\Sigma_{(j)}^0 - \Sigma_{(j+1)}^0)) \\
&\quad +tr((\Sigma_{(j+1)}^0)^{-1}(\Sigma_{(j)}^0 - \Sigma_{(j+1)}^0)(\Sigma_{(j)}^0)^{-1})(\textstyle\sum_{t=K_j+1}^{K_j^0} u_t u_t' - \Sigma_{(j)}^0))] \\
&\quad +(K_j^0 - K_j)tr((\Sigma_{(j+1)}^0)^{-1}(\Sigma_{(j)}^0 - \Sigma_{(j+1)}^0))] \\
&\quad +(1/4)\textstyle\sum_{j=1}^2 v_T^2(K_j - K_j^0)tr((\Sigma_{(j+1)}^0)^{-1}\Phi_{(j)}(\Sigma_{(j+1)}^0)^{-1}\Phi_{(j)}) + o_p(1) \\
&= (1/2)\textstyle\sum_{j=1}^2 tr((\Sigma_{(j)}^0)^{1/2}(\Sigma_{(j+1)}^0)^{-1}\Phi_{(j)}(\Sigma_{(j)}^0)^{-1/2}(v_T\sum_{t=K_j+1}^{K_j^0}\eta_t\eta_t' - I_n)) \\
&\quad +(1/4)\textstyle\sum_{j=1}^2 v_T^2(K_j - K_j^0)tr((\Sigma_{(j+1)}^0)^{-1}\Phi_{(j)}(\Sigma_{(j+1)}^0)^{-1}\Phi_{(j)}) + o_p(1)
\end{aligned}
$$

$$
\begin{aligned}
II &= -(1/2)T^{-1/2}\textstyle\sum_{j=1}^2(K_j - K_j^0)(tr((\Sigma_{(j)}^0)^{-1}\Sigma_{(j)}^*) \\
&\quad -T^{-1/2}tr((\Sigma_{(j+1)}^0)^{-1}\Sigma_{(j+1)}^*)) \\
&\quad -(1/2)T^{-1/2}\textstyle\sum_{j=1}^2\sum_{t=K_j+1}^{K_j^0}(u_t'(\Sigma_{(j)}^0)^{-1}\Sigma_{(j)}^*(\Sigma_{(j)}^0)^{-1}u_t \\
&\quad -u_t'(\Sigma_{(j+1)}^0)^{-1}\Sigma_{(j+1)}^*(\Sigma_{(j+1)}^0)^{-1}u_t \\
&= o_p(1)
\end{aligned}
$$

$$
III = -(1/2)[(\beta_{(2)}^0 - \beta_{(1)}^0)'(\textstyle\sum_{t=K_1+1}^{K_2} X_{Tt}(\Sigma_{(2)}^0)^{-1}X_{Tt}')(\beta_{(2)}^0 - \beta_{(1)}^0)
$$

$$
\begin{aligned}
& +(\beta^0_{(3)}-\beta^0_{(1)})'(\textstyle\sum_{t=K_2+1}^{K_1^0} X_{Tt}(\Sigma^0_{(3)})^{-1}X'_{Tt})(\beta^0_{(3)}-\beta^0_{(1)}) \\
& +(\beta^0_{(3)}-\beta^0_{(2)})'(\textstyle\sum_{t=K_1^0+1}^{K_2^0} X_{Tt}(\Sigma^0_{(3)})^{-1}X'_{Tt})(\beta^0_{(3)}-\beta^0_{(2)})] \\
= \; & -(1/2)(\beta^0_{(2)}-\beta^0_{(1)})'(\textstyle\sum_{t=K_1+1}^{K_1^0} X_{Tt}(\Sigma^0_{(2)})^{-1}X'_{Tt})(\beta^0_{(2)}-\beta^0_{(1)}) \\
& +(1/2)(\beta^0_{(2)}-\beta^0_{(1)})'(\textstyle\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma^0_{(2)})^{-1}X'_{Tt})(\beta^0_{(2)}-\beta^0_{(1)}) \\
& -(1/2)(\beta^0_{(2)}-\beta^0_{(1)})'(\textstyle\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma^0_{(3)})^{-1}X'_{Tt})(\beta^0_{(2)}-\beta^0_{(1)}) \\
& -(1/2)(\beta^0_{(3)}-\beta^0_{(2)})'(\textstyle\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma^0_{(3)})^{-1}X'_{Tt})(\beta^0_{(3)}-\beta^0_{(2)}) \\
& -(\beta^0_{(3)}-\beta^0_{(2)})'(\textstyle\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma^0_{(3)})^{-1}X'_{Tt})(\beta^0_{(2)}-\beta^0_{(1)})
\end{aligned}
$$

$$
\begin{aligned}
IV \;=\; & (\beta^0_{(2)}-\beta^0_{(1)})'(\textstyle\sum_{t=K_1+1}^{K_2} X_{Tt}(\Sigma^0_{(2)})^{-1}u_t) \\
& +(\beta^0_{(3)}-\beta^0_{(2)})'(\textstyle\sum_{t=K_2+1}^{K_1^0} X_{Tt}(\Sigma^0_{(3)})^{-1}u_t) \\
& +(\beta^0_{(2)}-\beta^0_{(1)})'(\textstyle\sum_{t=K_2+1}^{K_1^0} X_{Tt}(\Sigma^0_{(3)})^{-1}u_t) \\
& +(\beta^0_{(3)}-\beta^0_{(2)})'(\textstyle\sum_{t=K_1^0+1}^{K_2^0} X_{Tt}(\Sigma^0_{(3)})^{-1}u_t) \\
=\; & (\beta^0_{(2)}-\beta^0_{(1)})'(\textstyle\sum_{t=K_1+1}^{K_1^0} X_{Tt}(\Sigma^0_{(2)})^{-1}u_t) \\
& +(\beta^0_{(3)}-\beta^0_{(2)})'(\textstyle\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma^0_{(3)})^{-1}u_t) \\
& -(\beta^0_{(2)}-\beta^0_{(1)})'(\textstyle\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma^0_{(2)})^{-1}u_t) \\
& +(\beta^0_{(2)}-\beta^0_{(1)})'(\textstyle\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma^0_{(3)})^{-1}u_t) + o_p(1)
\end{aligned}
$$

Hence,

$$
\begin{aligned}
& rlr_T^2(K_1,,K_2,\beta_{(j)},\Sigma_{(j)}) \\
= \; & (1/2)\textstyle\sum_{j=1}^2 tr((\Sigma^0_{(j)})^{1/2}(\Sigma^0_{(j+1)})^{-1}\Phi_{(j)}(\Sigma^0_{(j)})^{-1/2}(v_T\sum_{t=K_j+1}^{K_j^0}\eta_t\eta_t'-I_n)) \\
& -(1/4)\textstyle\sum_{j=1}^2 |v_T^2(K_j-K_j^0)| tr(\Sigma^{0-1}_{(j+1)}\Phi_{(j)}(\Sigma^0_{(j+1)})^{-1}\Phi_{(j)}) \\
& -(1/2)(\beta^0_{(2)}-\beta^0_{(1)})'(\textstyle\sum_{t=K_1+1}^{K_1^0} X_{Tt}(\Sigma^0_{(2)})^{-1}X'_{Tt})(\beta^0_{(2)}-\beta^0_{(1)}) \\
& +(1/2)(\beta^0_{(2)}-\beta^0_{(1)})'(\textstyle\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma^0_{(2)})^{-1}X'_{Tt})(\beta^0_{(2)}-\beta^0_{(1)}) \\
& -(1/2)(\beta^0_{(2)}-\beta^0_{(1)})'(\textstyle\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma^0_{(3)})^{-1}X'_{Tt})(\beta^0_{(2)}-\beta^0_{(1)})
\end{aligned}
$$

$$-(1/2)(\beta^0_{(3)} - \beta^0_{(2)})'(\textstyle\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma^0_{(3)})^{-1}X'_{Tt})(\beta^0_{(3)} - \beta^0_{(2)})$$
$$-(\beta^0_{(3)} - \beta^0_{(2)})'(\textstyle\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma^0_{(3)})^{-1}X'_{Tt})(\beta^0_{(2)} - \beta^0_{(1)})$$
$$+(\beta^0_{(2)} - \beta^0_{(1)})'(\textstyle\sum_{t=K_1+1}^{K_1^0} X_{Tt}(\Sigma^0_{(2)})^{-1}u_t)$$
$$+(\beta^0_{(3)} - \beta^0_{(2)})'(\textstyle\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma^0_{(3)})^{-1}u_t)$$
$$-(\beta^0_{(2)} - \beta^0_{(1)})'(\textstyle\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma^0_{(2)})^{-1}u_t)$$
$$+(\beta^0_{(2)} - \beta^0_{(1)})'(\textstyle\sum_{t=K_2+1}^{K_2^0} X_{Tt}(\Sigma^0_{(3)})^{-1}u_t) + o_p(1)$$

For **Case 2** ($K_1^0 < K_2^0 \le K_1 < K_2$):

$$\begin{aligned}
(B.1) \quad = \quad & \textstyle\sum_{t=K_1^0+1}^{K_2^0}(Y_t - X'_{Tt}\beta_{(1)})'(\Sigma_{(1)})^{-1}(Y_t - X'_{Tt}\beta_{(1)}) \\
& + \textstyle\sum_{t=K_2^0+1}^{K_1}(Y_t - X'_{Tt}\beta_{(1)})'(\Sigma_{(1)})^{-1}(Y_t - X'_{Tt}\beta_{(1)}) \\
& - \textstyle\sum_{t=K_1^0+1}^{K_2^0}(Y_t - X'_{Tt}\beta_{(2)})'(\Sigma_{(2)})^{-1}(Y_t - X'_{Tt}\beta_{(2)}) \\
& - \textstyle\sum_{t=K_2^0+1}^{K_2}(Y_t - X'_{Tt}\beta_{(3)})'(\Sigma_{(3)})^{-1}(Y_t - X'_{Tt}\beta_{(3)}) \\
& + \textstyle\sum_{t=K_1+1}^{K_2}(Y_t - X'_{Tt}\beta_{(2)})'(\Sigma_{(2)})^{-1}(Y_t - X'_{Tt}\beta_{(2)})
\end{aligned}$$

$$\begin{aligned}
& rlr_T^2(K_1, K_2, \beta_{(j)}, \Sigma_{(j)}) \\
= \quad & -(1/2)\textstyle\sum_{j=1}^2 tr((\Sigma^0_{(j+1)})^{1/2}(\Sigma^0_{(j)})^{-1}\Phi_{(j)}(\Sigma^0_{(j+1)})^{-1/2}(v_T\sum_{t=K_j^0+1}^{K_j}\eta_t\eta'_t - I_n)) \\
& -(1/4)\textstyle\sum_{j=1}^2 |v_T^2(K_j - K_j^0)| tr((\Sigma^0_{(j)})^{-1}\Phi_{(j)}(\Sigma^0_{(j)})^{-1}\Phi_{(j)}) \\
& -(1/2)(\beta^0_{(2)} - \beta^0_{(1)})'(\textstyle\sum_{t=K_1^0+1}^{K_1} X_{Tt}(\Sigma^0_{(1)})^{-1}X'_{Tt})(\beta^0_{(2)} - \beta^0_{(1)}) \\
& -(1/2)(\beta^0_{(3)} - \beta^0_{(2)})'(\textstyle\sum_{t=K_1^0+1}^{K_1} X_{Tt}(\Sigma^0_{(1)})^{-1}X'_{Tt})(\beta^0_{(3)} - \beta^0_{(2)}) \\
& -(1/2)(\beta^0_{(3)} - \beta^0_{(2)})'(\textstyle\sum_{t=K_2^0+1}^{K_2} X_{Tt}(\Sigma^0_{(2)})^{-1}X'_{Tt})(\beta^0_{(3)} - \beta^0_{(2)}) \\
& +(1/2)(\beta^0_{(3)} - \beta^0_{(2)})'(\textstyle\sum_{t=K_1^0+1}^{K_1} X_{Tt}(\Sigma^0_{(2)})^{-1}X'_{Tt})(\beta^0_{(3)} - \beta^0_{(2)}) \\
& -(\beta^0_{(3)} - \beta^0_{(2)})'(\textstyle\sum_{t=K_1^0+1}^{K_1} X_{Tt}(\Sigma^0_{(1)})^{-1}X'_{Tt})(\beta^0_{(2)} - \beta^0_{(1)}) \\
& -(\beta^0_{(2)} - \beta^0_{(1)})'(\textstyle\sum_{t=K_1^0+1}^{K_1} X_{Tt}(\Sigma^0_{(1)})^{-1}u_t) \\
& -(\beta^0_{(3)} - \beta^0_{(2)})'(\textstyle\sum_{t=K_1^0+1}^{K_1} X_{Tt}(\Sigma^0_{(1)})^{-1}u_t)
\end{aligned}$$

$$-(\beta^0_{(3)} - \beta^0_{(2)})'(\textstyle\sum_{t=K_2^0+1}^{K_2} X_{Tt}(\Sigma^0_{(2)})^{-1}u_t)$$
$$+(\beta^0_{(3)} - \beta^0_{(2)})'(\textstyle\sum_{t=K_1^0+1}^{K_1} X_{Tt}(\Sigma^0_{(2)})^{-1}u_t) + o_p(1)$$

For **Case 3** ($K_1 \leq K_1^0 < K_2^0 \leq K_2$), we have

$$\begin{aligned}(B.1) \quad = \quad & -\textstyle\sum_{t=K_1+1}^{K_1^0}(Y_t - X'_{Tt}\beta_{(2)})'(\Sigma_{(2)})^{-1}(Y_t - X'_{Tt}\beta_{(2)}) \\ & -\textstyle\sum_{t=K_1+1}^{K_1^0}(Y_t - X'_{Tt}\beta_{(1)})'(\Sigma_{(1)})^{-1}(Y_t - X'_{Tt}\beta_{(1)}) \\ & +\textstyle\sum_{t=K_2^0+1}^{K_2}(Y_t - X'_{Tt}\beta_{(2)})'(\Sigma_{(2)})^{-1}(Y_t - X'_{Tt}\beta_{(2)}) \\ & -\textstyle\sum_{t=K_2^0+1}^{K_2}(Y_t - X'_{Tt}\beta_{(3)})'(\Sigma_{(3)})^{-1}(Y_t - X'_{Tt}\beta_{(3)})\end{aligned}$$

$$\begin{aligned} & rlr_T^2(K_1, K_2, \beta_{(j)}, \Sigma_{(j)}) \\ = \quad & (1/2)tr((\Sigma^0_{(1)})^{1/2}(\Sigma^0_{(2)})^{-1}\Phi_{(1)}(\Sigma^0_{(1)})^{-1/2}(v_T \textstyle\sum_{t=K_1+1}^{K_1^0}\eta_t\eta_t' - I_n)) \\ & -(1/4)|v_T^2(K_1 - K_1^0)|tr((\Sigma^0_{(2)})^{-1}\Phi_{(1)}(\Sigma^0_{(2)})^{-1}\Phi_{(1)}) \\ & -(1/2)tr((\Sigma^0_{(3)})^{1/2}(\Sigma^0_{(2)})^{-1}\Phi_{(2)}(\Sigma^0_{(3)})^{-1/2}(v_T \textstyle\sum_{t=K_2^0+1}^{K_2}\eta_t\eta_t' - I_n)) \\ & -(1/4)|v_T^2(K_2 - K_2^0)|tr((\Sigma^0_{(2)})^{-1}\Phi_{(2)}(\Sigma^0_{(2)})^{-1}\Phi_{(2)}) \\ & -(1/2)(\beta^0_{(2)} - \beta^0_{(1)})'(\textstyle\sum_{t=K_1+1}^{K_1^0} X_{Tt}(\Sigma^0_{(2)})^{-1}X'_{Tt})(\beta^0_{(2)} - \beta^0_{(1)}) \\ & -(1/2)(\beta^0_{(3)} - \beta^0_{(2)})'(\textstyle\sum_{t=K_2^0+1}^{K_2} X_{Tt}(\Sigma^0_{(2)})^{-1}X'_{Tt})(\beta^0_{(3)} - \beta^0_{(2)}) \\ & +(\beta^0_{(2)} - \beta^0_{(1)})'(\textstyle\sum_{t=K_1+1}^{K_1^0} X_{Tt}(\Sigma^0_{(2)})^{-1}u_t) \\ & -(\beta^0_{(3)} - \beta^0_{(2)})'(\textstyle\sum_{t=K_2^0+1}^{K_2} X_{Tt}(\Sigma^0_{(2)})^{-1}u_t) + o_p(1)\end{aligned}$$

Proof of Theorem 1.4: Multiple breaks.

Case 1 ($K_1 < ... < K_m \leq K_1^0 < ... < K_m^0$):

$$(B.1) \quad = \quad -\textstyle\sum_{t=K_1+1}^{K_1^0}(Y_t - X'_{Tt}\beta_{(1)})'(\Sigma_{(1)})^{-1}(Y_t - X'_{Tt}\beta_{(1)}) \tag{F.1}$$
$$+\textstyle\sum_{j=2}^{m}\sum_{t=K_{j-1}+1}^{K_j}(Y_t - X'_{Tt}\beta_{(j)})'(\Sigma_{(j)})^{-1}(Y_t - X'_{Tt}\beta_{(j)}) \tag{F.2}$$
$$-\textstyle\sum_{j=2}^{m}\sum_{t=K_{j-1}^0+1}^{K_j^0}(Y_t - X'_{Tt}\beta_{(j)})'(\Sigma_{(j)})^{-1}(Y_t - X'_{Tt}\beta_{(j)}) \tag{F.3}$$
$$+\textstyle\sum_{t=K_m+1}^{K_1^0}(Y_t - X'_{Tt}\beta_{(m+1)})'(\Sigma_{(m+1)})^{-1}(Y_t - X'_{Tt}\beta_{(m+1)}) \tag{F.4}$$
$$+\textstyle\sum_{j=2}^{m}\sum_{t=K_{j-1}^0+1}^{K_j^0}(Y_t - X'_{Tt}\beta_{(m+1)})'(\Sigma_{(m+1)})^{-1}(Y_t - X'_{Tt}\beta_{(m+1)}) \tag{F.5}$$

We consider the development of each term.

$$
\begin{aligned}
(F.1) &= -\textstyle\sum_{t=K_1+1}^{K_1^0}(u_t + X'_{Tt}\beta^0_{(1)} - X'_{Tt}\beta_{(1)})'(\Sigma_{(1)})^{-1}(u_t + X'_{Tt}\beta^0_{(1)} - X'_{Tt}\beta_{(1)}) \\
&= -\textstyle\sum_{t=K_1+1}^{K_1^0} u'_t(\Sigma^0_{(1)})^{-1}u_t + T^{-1/2}\sum_{t=K_1+1}^{K_1^0} u'_t(\Sigma^0_{(1)})^{-1}\Sigma^*_{(1)}(\Sigma^0_{(1)})^{-1}u_t + o_p(1)
\end{aligned}
$$

$$
\begin{aligned}
(F.2) &= \textstyle\sum_{j=2}^{m}\sum_{t=K_{j-1}+1}^{K_j}(u_t + X'_{Tt}\beta^0_{(1)} - X'_{Tt}\beta_{(j)})'(\Sigma_{(j)})^{-1}(u_t + X'_{Tt}\beta^0_{(1)} - X'_{Tt}\beta_{(j)}) \\
&= \textstyle\sum_{j=2}^{m}\sum_{t=K_{j-1}+1}^{K_j} u'_t(\Sigma^0_{(j)})^{-1}u_t - T^{-1/2}\sum_{j=2}^{m}\sum_{t=K_{j-1}+1}^{K_j} u'_t(\Sigma^0_{(j)})^{-1}\Sigma^*_{(j)}(\Sigma^0_{(j)})^{-1}u_t \\
&\quad + \textstyle\sum_{j=2}^{m}(\beta_{(j)} - \beta^0_{(1)})'(\sum_{t=K_{j-1}+1}^{K_j} X_{Tt}(\Sigma^0_{(j)})^{-1}X'_{Tt})(\beta_{(j)} - \beta^0_{(1)}) \\
&\quad -2\textstyle\sum_{j=2}^{m}(\beta_{(j)} - \beta^0_{(1)})'(\sum_{t=K_{j-1}+1}^{K_j} X_{Tt}(\Sigma^0_{(j)})^{-1}u_t) + o_p(1)
\end{aligned}
$$

$$
\begin{aligned}
(F.3) &= -\textstyle\sum_{j=2}^{m}\sum_{t=K^0_{j-1}+1}^{K_j^0}(u_t + X'_{Tt}\beta^0_{(j)} - X'_{Tt}\beta_{(j)})' \\
&\quad (\Sigma_{(j)})^{-1}(u_t + X'_{Tt}\beta^0_{(j)} - X'_{Tt}\beta_{(j)}) \\
&= -\textstyle\sum_{j=2}^{m}\sum_{t=K^0_{j-1}+1}^{K_j^0} u'_t(\Sigma^0_{(j)})^{-1}u_t \\
&\quad +T^{-1/2}\textstyle\sum_{j=2}^{m}\sum_{t=K^0_{j-1}+1}^{K_j^0} u'_t(\Sigma^0_{(j)})^{-1}\Sigma^*_{(j)}(\Sigma^0_{(j)})^{-1}u_t + o_p(1)
\end{aligned}
$$

$$
\begin{aligned}
(F.4) &= \textstyle\sum_{t=K_m+1}^{K_1^0}(u_t + X'_{Tt}\beta^0_{(1)} - X'_{Tt}\beta_{(m+1)})'(\Sigma_{(m+1)})^{-1}(u_t + X'_{Tt}\beta^0_{(1)} - X'_{Tt}\beta_{(m+1)}) \\
&= \textstyle\sum_{t=K_m+1}^{K_1^0} u'_t(\Sigma^0_{(m+1)})^{-1}u_t - T^{-\frac{1}{2}}\sum_{t=K_m+1}^{K_1^0} u'_t(\Sigma^0_{(m+1)})^{-1}\Sigma^*_{(m+1)}(\Sigma^0_{(m+1)})^{-1}u_t \\
&\quad +(\beta^0_{(m+1)} - \beta^0_{(1)})'(\textstyle\sum_{t=K_m+1}^{K_1^0} X_{Tt}(\Sigma^0_{(m+1)})^{-1}X'_{Tt})(\beta^0_{(m+1)} - \beta^0_{(1)}) \\
&\quad -2(\beta^0_{(m+1)} - \beta^0_{(1)})'(\textstyle\sum_{t=K_m+1}^{K_1^0} X_{Tt}(\Sigma^0_{(m+1)})^{-1}u_t) + o_p(1)
\end{aligned}
$$

$$
\begin{aligned}
(F.5) &= \textstyle\sum_{j=2}^{m}\sum_{t=K^0_{j-1}+1}^{K_j^0}(u_t + X'_{Tt}\beta^0_{(j)} - X'_{Tt}\beta_{(m+1)})' \\
&\quad (\Sigma_{(m+1)})^{-1}(u_t + X'_{Tt}\beta^0_{(j)} - X'_{Tt}\beta_{(m+1)}) \\
&= \textstyle\sum_{j=2}^{m}\sum_{t=K^0_{j-1}+1}^{K_j^0} u'_t(\Sigma^0_{(m+1)})^{-1}u_t \\
&\quad -T^{-1/2}\textstyle\sum_{j=2}^{m}\sum_{t=K^0_{j-1}+1}^{K_j^0} u'_t(\Sigma^0_{(m+1)})^{-1}\Sigma^*_{(m+1)}(\Sigma^0_{(m+1)})^{-1}u_t \\
&\quad +\textstyle\sum_{j=2}^{m}(\beta^0_{(m+1)} - \beta^0_{(j)})'(\sum_{t=K^0_{j-1}+1}^{K_j^0} X_{Tt}(\Sigma^0_{(m+1)})^{-1}X'_{Tt})(\beta^0_{(m+1)} - \beta^0_{(j)}) \\
&\quad -2\textstyle\sum_{j=2}^{m}(\beta^0_{(m+1)} - \beta^0_{(j)})'(\sum_{t=K^0_{j-1}+1}^{K_j^0} X_{Tt}(\Sigma^0_{(m+1)})^{-1}u_t) + o_p(1)
\end{aligned}
$$

Hence, $rlr^2_T(K_1, K_2, \beta, \Sigma) = -(1/2)((A.1) + (B.1))$ contains four parts as follows:

$$\begin{aligned}
I \;=\; & -(1/2)\textstyle\sum_{j=1}^{m}(K_j - K_j^0)(\log|\Sigma^0_{(j)}| - \log|\Sigma^0_{(j+1)}|) \\
& +(1/2)\textstyle\sum_{t=K_1+1}^{K_1^0} u_t'(\Sigma^0_{(1)})^{-1}u_t - (1/2)\sum_{j=2}^{m}\sum_{t=K_{j-1}+1}^{K_j} u_t'(\Sigma^0_{(j)})^{-1}u_t \\
& +(1/2)\textstyle\sum_{j=2}^{m}\sum_{t=K_{j-1}^0+1}^{K_j^0} u_t'(\Sigma^0_{(j)})^{-1}u_t - (1/2)\sum_{t=K_m+1}^{K_1^0} u_t'(\Sigma^0_{(m+1)})^{-1}u_t \\
& -(1/2)\textstyle\sum_{j=2}^{m}\sum_{t=K_{j-1}^0+1}^{K_j^0} u_t'(\Sigma^0_{(m+1)})^{-1}u_t] \\
=\; & (1/2)\textstyle\sum_{j=1}^{m} tr((\Sigma^0_{(j)})^{1/2}(\Sigma^0_{(j+1)})^{-1}\Phi_{(j)}(\Sigma^0_{(j)})^{-1/2}(v_T\sum_{t=K_j+1}^{K_j^0}\eta_t\eta_t' - I_n)) \\
& +(1/4)\textstyle\sum_{j=1}^{m} v_T^2(K_j - K_j^0)tr((\Sigma^0_{(j+1)})^{-1}\Phi_{(j)}(\Sigma^0_{(j+1)})^{-1}\Phi_{(j)}) + o_p(1)
\end{aligned}$$

$$\begin{aligned}
II \;=\; & -(1/2)[-T^{-1/2}\textstyle\sum_{j=1}^{m}(K_j - K_j^0)(tr((\Sigma^0_{(j)})^{-1}\Sigma^*_{(j)}) - tr((\Sigma^0_{(j+1)})^{-1}\Sigma^*_{(j+1)})) \\
& +T^{-1/2}\textstyle\sum_{j=1}^{m}\sum_{t=K_j+1}^{K_j^0}(u_t'(\Sigma^0_{(j)})^{-1}\Sigma^*_{(j)}(\Sigma^0_{(j)})^{-1}u_t - u_t'(\Sigma^0_{(j+1)})^{-1}\Sigma^*_{(j+1)}(\Sigma^0_{(j+1)})^{-1}u_t)] \\
=\; & o_p(1)
\end{aligned}$$

$$\begin{aligned}
III \;=\; & -(1/2)[\textstyle\sum_{j=2}^{m}(\beta^0_{(j)} - \beta^0_{(1)})'(\sum_{t=K_{j-1}+1}^{K_j} X_{Tt}(\Sigma^0_{(j)})^{-1}X_{Tt}')(\beta^0_{(j)} - \beta^0_{(1)}) \\
& +(\beta^0_{(m+1)} - \beta^0_{(1)})'(\textstyle\sum_{t=K_m+1}^{K_1^0} X_{Tt}(\Sigma^0_{(m+1)})^{-1}X_{Tt}')(\beta^0_{(m+1)} - \beta^0_{(1)}) \\
& +\textstyle\sum_{j=2}^{m}(\beta^0_{(m+1)} - \beta^0_{(j)})'(\sum_{t=K_{j-1}^0+1}^{K_j^0} X_{Tt}(\Sigma^0_{(m+1)})^{-1}X_{Tt}')(\beta^0_{(m+1)} - \beta^0_{(j)})] \\
=\; & -(1/2)\textstyle\sum_{j=1}^{m}(\beta^0_{(j+1)} - \beta^0_{(j)})'(\sum_{l=j}^{m}\sum_{t=K_l+1}^{K_l^0} X_{Tt}(\Sigma^0_{(l+1)})^{-1}X_{Tt}')(\beta^0_{(j+1)} - \beta^0_{(j)}) \\
& +(1/2)\textstyle\sum_{j=1}^{m-1}(\beta^0_{(j+1)} - \beta^0_{(j)})'(\sum_{l=j+1}^{m}\sum_{t=K_l+1}^{K_l^0} X_{Tt}(\Sigma^0_{(l)})^{-1}X_{Tt}')(\beta^0_{(j+1)} - \beta^0_{(j)}) \\
& -\textstyle\sum_{j=1}^{m-1}\sum_{i=j+1}^{m}(\beta^0_{(j+1)} - \beta^0_{(j)})'(\sum_{l=i}^{m}\sum_{t=K_l+1}^{K_l^0} X_{Tt}(\Sigma^0_{(l+1)})^{-1}X_{Tt}')(\beta^0_{(i+1)} - \beta^0_{(i)}) \\
& +\textstyle\sum_{j=1}^{m-2}\sum_{i=j+1}^{m-1}(\beta^0_{(j+1)} - \beta^0_{(j)})'(\sum_{l=i+1}^{m}\sum_{t=K_l+1}^{K_l^0} X_{Tt}(\Sigma^0_{(l)})^{-1}X_{Tt}')(\beta^0_{(i+1)} - \beta^0_{(i)})
\end{aligned}$$

$$\begin{aligned}
IV \;=\; & \textstyle\sum_{j=2}^{m}(\beta^0_{(j)} - \beta^0_{(1)})'(\sum_{t=K_{j-1}+1}^{K_j} X_{Tt}(\Sigma^0_{(j)})^{-1}u_t) \\
& +(\beta^0_{(m+1)} - \beta^0_{(1)})'(\textstyle\sum_{t=K_m+1}^{K_1^0} X_{Tt}(\Sigma^0_{(m+1)})^{-1}u_t) \\
& +\textstyle\sum_{j=2}^{m}(\beta^0_{(m+1)} - \beta^0_{(j)})'(\sum_{t=K_{j-1}^0+1}^{K_j^0} X_{Tt}(\Sigma^0_{(m+1)})^{-1}u_t) \\
=\; & \textstyle\sum_{j=1}^{m}(\beta^0_{(j+1)} - \beta^0_{(j)})'\sum_{l=j}^{m}\sum_{t=K_l+1}^{K_l^0} X_{Tt}(\Sigma^0_{(l+1)})^{-1}u_t \\
& +\textstyle\sum_{j=1}^{m-1}(\beta^0_{(j+1)} - \beta^0_{(j)})'\sum_{l=j+1}^{m}\sum_{t=K_l+1}^{K_l^0} X_{Tt}(\Sigma^0_{(l)})^{-1}u_t
\end{aligned}$$

$$
\begin{aligned}
& rlr_T^2(K_1,...,K_m,\beta_{(j)},\Sigma_{(j)}) \\
= \;& (1/2)\textstyle\sum_{j=1}^m tr((\Sigma_{(j)}^0)^{1/2}(\Sigma_{(j+1)}^0)^{-1}\Phi_{(j)}(\Sigma_{(j)}^0)^{-1/2}(v_T\sum_{t=K_j+1}^{K_j^0}\eta_t\eta_t'-I_n)) \\
& +(1/4)\textstyle\sum_{j=1}^m v_T^2(K_j-K_j^0)tr((\Sigma_{(j+1)}^0)^{-1}\Phi_{(j)}(\Sigma_{(j+1)}^0)^{-1}\Phi_{(j)}) \\
& -(1/2)\textstyle\sum_{j=1}^m(\beta_{(j+1)}^0-\beta_{(j)}^0)'(\sum_{l=j}^m\sum_{t=K_l+1}^{K_l^0}X_{Tt}(\Sigma_{(l+1)}^0)^{-1}X_{Tt}')(\beta_{(j+1)}^0-\beta_{(j)}^0) \\
& +(1/2)\textstyle\sum_{j=1}^{m-1}(\beta_{(j+1)}^0-\beta_{(j)}^0)'(\sum_{l=j+1}^m\sum_{t=K_l+1}^{K_l^0}X_{Tt}(\Sigma_{(l)}^0)^{-1}X_{Tt}')(\beta_{(j+1)}^0-\beta_{(j)}^0) \\
& -\textstyle\sum_{j=1}^{m-1}\sum_{i=j+1}^m(\beta_{(j+1)}^0-\beta_{(j)}^0)'(\sum_{l=i}^m\sum_{t=K_l+1}^{K_l^0}X_{Tt}(\Sigma_{(l+1)}^0)^{-1}X_{Tt}')(\beta_{(i+1)}^0-\beta_{(i)}^0) \\
& +\textstyle\sum_{j=1}^{m-2}\sum_{i=j+1}^{m-1}(\beta_{(j+1)}^0-\beta_{(j)}^0)'(\sum_{l=i+1}^m\sum_{t=K_l+1}^{K_l^0}X_{Tt}(\Sigma_{(l)}^0)^{-1}X_{Tt}')(\beta_{(i+1)}^0-\beta_{(i)}^0)] \\
& +\textstyle\sum_{j=1}^m(\beta_{(j+1)}^0-\beta_{(j)}^0)'(\sum_{l=j}^m\sum_{t=K_l+1}^{K_l^0}X_{Tt}(\Sigma_{(l+1)}^0)^{-1}u_t) \\
& +\textstyle\sum_{j=1}^{m-1}(\beta_{(j+1)}^0-\beta_{(j)}^0)'(\sum_{l=j+1}^m\sum_{t=K_l+1}^{K_l^0}X_{Tt}(\Sigma_{(l)}^0)^{-1}u_t)+o_p(1)
\end{aligned}
$$

Case 2: $K_1^0<...<K_m^0\le K_1<...<K_m$:

$$
\begin{aligned}
(B.1) \;=\; & \textstyle\sum_{j=2}^m\sum_{t=K_{j-1}^0+1}^{K_j^0}(Y_t-X_{Tt}'\beta_{(1)})'(\Sigma_{(1)})^{-1}(Y_t-X_{Tt}'\beta_{(1)}) \\
& -\textstyle\sum_{j=2}^m\sum_{t=K_{j-1}^0+1}^{K_j^0}(Y_t-X_{Tt}'\beta_{(j)})'(\Sigma_{(j)})^{-1}(Y_t-X_{Tt}'\beta_{(j)}) \\
& +\textstyle\sum_{j=2}^m\sum_{t=K_{j-1}+1}^{K_j}(Y_t-X_{Tt}'\beta_{(j)})'(\Sigma_{(j)})^{-1}(Y_t-X_{Tt}'\beta_{(j)}) \\
& -\textstyle\sum_{t=K_m^0+1}^{K_m}(Y_t-X_{Tt}'\beta_{(m+1)})'(\Sigma_{(m+1)})^{-1}(Y_t-X_{Tt}'\beta_{(m+1)}) \\
& +\textstyle\sum_{t=K_m^0+1}^{K_1}(Y_t-X_{Tt}'\beta_{(1)})'(\Sigma_{(1)})^{-1}(Y_t-X_{Tt}'\beta_{(1)})
\end{aligned}
$$

Following developments for Case 1, we have:

$$
\begin{aligned}
& rlr_T^2(K_1,...,K_m,\beta_{(j)},\Sigma_{(j)}) \\
= \;& -(1/2)\textstyle\sum_{j=1}^m tr((\Sigma_{(j+1)}^0)^{1/2}(\Sigma_{(j)}^0)^{-1}\Phi_{(j)}(\Sigma_{(j+1)}^0)^{-1/2}(v_T\sum_{t=K_j^0+1}^{K_j}\eta_t\eta_t'-I_n)) \\
& -(1/4)\textstyle\sum_{j=1}^m|v_T^2(K_j-K_j^0)|tr((\Sigma_{(j)}^0)^{-1}\Phi_{(j)}(\Sigma_{(j)}^0)^{-1}\Phi_{(j)}) \\
& -(1/2)\textstyle\sum_{j=1}^m(\beta_{(j+1)}^0-\beta_{(j)}^0)'(\sum_{l=1}^j\sum_{t=K_l^0+1}^{K_l}X_{Tt}(\Sigma_{(l)}^0)^{-1}X_{Tt}')(\beta_{(j+1)}^0-\beta_{(j)}^0) \\
& +(1/2)\textstyle\sum_{j=2}^m(\beta_{(j+1)}^0-\beta_{(j)}^0)'(\sum_{l=1}^{j-1}\sum_{t=K_l^0+1}^{K_l}X_{Tt}(\Sigma_{(l+1)}^0)^{-1}X_{Tt}')(\beta_{(j+1)}^0-\beta_{(j)}^0) \\
& +\textstyle\sum_{j=2}^{m-1}\sum_{i=j+1}^m(\beta_{(j+1)}^0-\beta_{(j)}^0)'(\sum_{l=1}^{j-1}\sum_{t=K_l^0+1}^{K_l}X_{Tt}(\Sigma_{(l+1)}^0)^{-1}X_{Tt}')(\beta_{(i+1)}^0-\beta_{(i)}^0)
\end{aligned}
$$

$$- \sum_{j=1}^{m-1} \sum_{i=j+1}^{m} (\beta^0_{(j+1)} - \beta^0_{(j)})' (\sum_{l=1}^{j} \sum_{t=K_l^0+1}^{K_l} X_{Tt} (\Sigma^0_{(l)})^{-1} X'_{Tt}) (\beta^0_{(i+1)} - \beta^0_{(i)})$$
$$- \sum_{j=1}^{m} (\beta^0_{(j+1)} - \beta^0_{(j)})' \sum_{l=1}^{j} \sum_{t=K_l^0+1}^{K_l} X_{Tt} (\Sigma^0_{(l)})^{-1} u_t$$
$$+ \sum_{j=2}^{m} (\beta^0_{(j+1)} - \beta^0_{(j)})' \sum_{l=1}^{j-1} \sum_{t=K_l^0+1}^{K_l} X_{Tt} (\Sigma^0_{(l+1)})^{-1} u_t + o_p(1)$$

Case 3: $K_1 < ... < K_b \leq K_1^0 < ... < K_m^0 \leq K_{b+1} < ... < K_m$, for some $1 \leq b \leq m$:

$$\begin{aligned}
& rlr_T^2(K_1,..., K_m, \beta_{(j)}, \Sigma_{(j)}) \\
= \; & (1/2) \sum_{j=1}^{b} tr((\Sigma^0_{(j)})^{1/2} (\Sigma^0_{(j+1)})^{-1} \Phi_{(j)} (\Sigma^0_{(j)})^{-1/2} (v_T \sum_{t=K_j+1}^{K_j^0} \eta_t \eta_t' - I_n)) \\
& -(1/4) \sum_{j=1}^{b} |v_T^2 (K_j - K_j^0)| tr((\Sigma^0_{(j+1)})^{-1} \Phi_{(j)} (\Sigma^0_{(j+1)})^{-1} \Phi_{(j)}) \\
& -(1/2) \sum_{j=b+1}^{m} tr((\Sigma^0_{(j+1)})^{1/2} (\Sigma^0_{(j)})^{-1} \Phi_{(j)} (\Sigma^0_{(j+1)})^{-1/2} (v_T \sum_{t=K_j^0+1}^{K_j} \eta_t \eta_t' - I_n)) \\
& -(1/4) \sum_{j=b+1}^{m} |v_T^2 (K_j - K_j^0)| tr((\Sigma^0_{(j)})^{-1} \Phi_{(j)} (\Sigma^0_{(j)})^{-1} \Phi_{(j)}) \\
& -(1/2) \sum_{j=1}^{b} (\beta^0_{(j+1)} - \beta^0_{(j)})' (\sum_{l=j}^{b} \sum_{t=K_l+1}^{K_l^0} X_{Tt} (\Sigma^0_{(l+1)})^{-1} X'_{Tt}) (\beta^0_{(j+1)} - \beta^0_{(j)})
\end{aligned}$$

$$\begin{aligned}
& +(1/2) \sum_{j=1}^{b-1} (\beta^0_{(j+1)} - \beta^0_{(j)})' (\sum_{l=j+1}^{b} \sum_{t=K_l+1}^{K_l^0} X_{Tt} (\Sigma^0_{(l)})^{-1} X'_{Tt}) (\beta^0_{(j+1)} - \beta^0_{(j)}) \\
& - \sum_{j=1}^{b-1} \sum_{i=j+1}^{b} (\beta^0_{(j+1)} - \beta^0_{(j)})' (\sum_{l=i}^{b} \sum_{t=K_l+1}^{K_l^0} X_{Tt} (\Sigma^0_{(l+1)})^{-1} X'_{Tt}) (\beta^0_{(i+1)} - \beta^0_{(i)}) \\
& + \sum_{j=1}^{b-2} \sum_{i=j+1}^{b-1} (\beta^0_{(j+1)} - \beta^0_{(j)})' (\sum_{l=i+1}^{b} \sum_{t=K_l+1}^{K_l^0} X_{Tt} (\Sigma^0_{(l)})^{-1} X'_{Tt}) (\beta^0_{(i+1)} - \beta^0_{(i)}) \\
& -(1/2) \sum_{j=b+1}^{m} (\beta^0_{(j+1)} - \beta^0_{(j)})' (\sum_{l=b+1}^{j} \sum_{t=K_l^0+1}^{K_l} X_{Tt} (\Sigma^0_{(l)})^{-1} X'_{Tt}) (\beta^0_{(j+1)} - \beta^0_{(j)}) \\
& +(1/2) \sum_{j=b+2}^{m} (\beta^0_{(j+1)} - \beta^0_{(j)})' (\sum_{l=b+1}^{j-1} \sum_{t=K_l^0+1}^{K_l} X_{Tt} (\Sigma^0_{(l+1)})^{-1} X'_{Tt}) (\beta^0_{(j+1)} - \beta^0_{(j)}) \\
& - \sum_{j=b+1}^{m-1} \sum_{i=j+1}^{m} (\beta^0_{(j+1)} - \beta^0_{(j)})' (\sum_{l=b+1}^{j} \sum_{t=K_l^0+1}^{K_l} X_{Tt} (\Sigma^0_{(l)})^{-1} X'_{Tt}) (\beta^0_{(i+1)} - \beta^0_{(i)}) \\
& + \sum_{j=b+2}^{m-1} \sum_{i=j+1}^{m} (\beta^0_{(j+1)} - \beta^0_{(j)})' (\sum_{l=b+1}^{j-1} \sum_{t=K_l^0+1}^{K_l} X_{Tt} (\Sigma^0_{(l+1)})^{-1} X'_{Tt}) (\beta^0_{(i+1)} - \beta^0_{(i)}) \\
& + \sum_{j=1}^{b} (\beta^0_{(j+1)} - \beta^0_{(j)})' (\sum_{l=j}^{b} \sum_{t=K_l+1}^{K_l^0} X_{Tt} (\Sigma^0_{(l+1)})^{-1} u_t) \\
& + \sum_{j=1}^{b-1} (\beta^0_{(j+1)} - \beta^0_{(j)})' (\sum_{l=j+1}^{b} \sum_{t=K_l+1}^{K_l^0} X_{Tt} (\Sigma^0_{(l)})^{-1} u_t) \\
& - \sum_{j=b+1}^{m} (\beta^0_{(j+1)} - \beta^0_{(j)})' (\sum_{l=b+1}^{j} \sum_{t=K_l^0+1}^{K_l} X_{Tt} (\Sigma^0_{(l)})^{-1} u_t) \\
& + \sum_{j=b+2}^{m} (\beta^0_{(j+1)} - \beta^0_{(j)})' (\sum_{l=b+1}^{j} \sum_{t=K_l^0+1}^{K_l} X_{Tt} (\Sigma^0_{(l+1)})^{-1} u_t) + o_p(1)
\end{aligned}$$

Limit Distribution of the Locally Ordered Break Dates with Breaks in both Coefficients and Variance-covariance Matrix; $m = 2$.

For **Case 1** ($K_1 < K_2 \le K_1^0 < K_2^0$), let $A_{(j)} = (\Sigma^0_{(j+1)})^{-1}\Phi_{(j)}$, then using Lemma 1, we get

$$
\begin{aligned}
& rlr_T^2(K_1, K_2, \beta_{(j)}, \Sigma_{(j)}) \Rightarrow H_T^1(s_1, s_2) \\
= \quad & (1/2)\textstyle\sum_{j=1}^2 tr(A_{(j)}\xi_j(s_j)) - (1/4)\sum_{j=1}^2 |s_j| tr(A_{(j)}^2) \\
& -(1/2)|s_1|[\delta_1' S'(I_q \otimes (\Sigma^0_{(2)})^{-1})(D(\lambda_1^0) \otimes I_n)S\delta_1] \\
& +(1/2)|s_2|[\delta_1' S'(I_q \otimes (\Sigma^0_{(2)})^{-1})(D(\lambda_2^0) \otimes I_n)S\delta_1 \\
& -\delta_1' S'(I_q \otimes (\Sigma^0_{(3)})^{-1})(D(\lambda_2^0) \otimes I_n)S\delta_1 \\
& -\delta_2' S'(I_q \otimes (\Sigma^0_{(3)})^{-1})(D(\lambda_2^0) \otimes I_n)S\delta_1] \\
& -|s_2|[\delta_1' S'(I_q \otimes (\Sigma^0_{(3)})^{-1})(D(\lambda_2^0) \otimes I_n)S]\delta_2 \\
& +[\delta_1' S'(I_q \otimes (\Sigma^0_{(2)})^{-1})(\Sigma^0_{(1)})^{1/2}\Omega(\lambda_1^0)]W_1(s_1) \\
& +[\delta_2' S'(I_q \otimes (\Sigma^0_{(3)})^{-1})(\Sigma^0_{(2)})^{1/2}\Omega(\lambda_2^0) - \delta_1' S'(I_q \otimes (\Sigma^0_{(2)})^{-1})(\Sigma^0_{(2)})^{1/2}\Omega(\lambda_2^0) \\
& +\delta_1' S'(I_q \otimes (\Sigma^0_{(3)})^{-1})(\Sigma^0_{(2)})^{1/2}\Omega(\lambda_2^0)]W_2(s_2)
\end{aligned}
$$

Note that

$$tr(A_{(j)}\xi_j(s_j)) = vec(A_{(j)})' vec(\xi_j(s_j)) \stackrel{d}{=} vec(A_{(j)})' \Omega vec(A_{(j)}) U_j(s_j)$$

for $j = 1, 2$, where $\Omega = E(vec(\xi(1))vec(\xi(1))')$. Also let $A = \delta_1' S'(I_q \otimes (\Sigma^0_{(2)})^{-1})(\Sigma^0_{(1)})^{1/2}\Omega(\lambda_1^0)$,
$B = \delta_2' S'(I_q \otimes (\Sigma^0_{(3)})^{-1})(\Sigma^0_{(2)})^{1/2}\Omega(\lambda_2^0) - \delta_1' S'(I_q \otimes (\Sigma^0_{(2)})^{-1})(\Sigma^0_{(2)})^{1/2}\Omega(\lambda_2^0)$
$+\delta_1' S'(I_q \otimes (\Sigma^0_{(3)})^{-1})(\Sigma^0_{(2)})^{1/2}\Omega(\lambda_2^0)$, then

$$
\begin{aligned}
AW_1(s_1) &\stackrel{d}{=} (A'A)^{1/2} V_1(s_1) \\
BW_s(s_s) &\stackrel{d}{=} (B'B)^{1/2} V_2(s_2)
\end{aligned}
$$

where $U_1(\cdot)$, $U_2(\cdot)$, $V_1(\cdot)$ and $V_2(\cdot)$ are standard two-sided Brownian motions. The independence implies

$$
\begin{aligned}
& (1/2)tr(A_{(1)}\xi_1(s_1)) + \delta_1' \Delta_{(3)}(\lambda_1^0) W_1(s_1) \\
& \stackrel{d}{=} [(1/4)vec(A_{(1)})' \Omega vec(A_{(1)}) + A'A]^{1/2} B_1(S_1),
\end{aligned}
$$

$$
\begin{aligned}
& (1/2)tr(A_{(2)}\xi_2(s_2)) + BW_2(s_2) \\
& \stackrel{d}{=} [(1/4)vec(A_{(2)})' \Omega vec(A_{(2)}) + B'B]^{1/2} B_2(s_2)
\end{aligned}
$$

Then

$$\begin{aligned} H^1(s_1,s_2) &= [(1/4)vec(A_{(1)})'\Omega vec(A_{(1)}) + A'A]^{1/2}B_1(S_1) \\ &\quad [(1/4)vec(A_{(2)})'\Omega vec(A_{(2)}) + B'B]^{1/2}B_2(s_2) \\ &\quad -(1/2)|s_1|[(1/2)tr(A_1^2) + \delta_1'S'(I_q \otimes (\Sigma_{(2)}^0)^{-1})(D(\lambda_1^0) \otimes I_n)S\delta_1] \\ &\quad -(1/2)|s_2|[(1/2)tr(A_2^2) + \delta_1'S'(I_q \otimes (\Sigma_{(2)}^0)^{-1})(D(\lambda_2^0) \otimes I_n)S\delta_1 \\ &\quad -\delta_1'S'(I_q \otimes (\Sigma_{(3)}^0)^{-1})(D(\lambda_2^0) \otimes I_n)S\delta_1 \\ &\quad -\delta_2'S'(I_q \otimes (\Sigma_{(3)}^0)^{-1})(D(\lambda_2^0) \otimes I_n)S\delta_1] \\ &\quad -|s_2|[\delta_1'S'(I_q \otimes (\Sigma_{(3)}^0)^{-1})(D(\lambda_2^0) \otimes I_n)S]\delta_2 \end{aligned}$$

$$\begin{aligned} \Pi_1 &= (1/2)tr(A_1^2) + \delta_1'S'(I_q \otimes (\Sigma_{(2)}^0)^{-1})(D(\lambda_1^0) \otimes I_n)S\delta_1 \\ \Pi_2 &= (1/2)tr(A_2^2) + \delta_1'S'(I_q \otimes (\Sigma_{(2)}^0)^{-1})(D(\lambda_2^0) \otimes I_n)S\delta_1 \\ &\quad -\delta_1'S'(I_q \otimes (\Sigma_{(3)}^0)^{-1})(D(\lambda_2^0) \otimes I_n)S\delta_1 \\ &\quad -\delta_2'S'(I_q \otimes (\Sigma_{(3)}^0)^{-1})(D(\lambda_2^0) \otimes I_n)S\delta_1] \\ \Pi_{12} &= \delta_1'S'(I_q \otimes (\Sigma_{(3)}^0)^{-1})(D(\lambda_2^0) \otimes I_n)S\delta_2 \\ \Upsilon_1 &= (1/4)vec(A_{(1)})'\Omega vec(A_{(1)}) + A'A \\ \Upsilon_2 &= (1/4)vec(A_{(2)})'\Omega vec(A_{(2)}) + B'B \end{aligned}$$

Let $s_1 = bv_1$, $s_2 = bv_2$ and $b = \Upsilon_1/\Pi_1^2$, then

$$\Pi_1 v_T^2(\hat{K}_1 - K_1^0, \hat{K}_2 - K_2^0) \Rightarrow \arg\max_{v_1 \le v_2 \le 0} H^1(v_1,v_2)$$

where

$$\begin{aligned} H^1(v_1,v_2) &= -(1/2)|v_1| - (1/2)|v_2|\frac{\Pi_2}{\Pi_1} - |v_2|\frac{\Pi_{12}}{\Pi_1} \\ &\quad +(\frac{\Upsilon_1}{\Pi_1})^{1/2}B_1(v_1) + (\frac{\Upsilon_2}{\Pi_1})^{1/2}B_2(v_2) \end{aligned}$$

For **Case 2** ($K_1^0 < K_2^0 \le K_1 < K_2$),

let $A = \delta_1'S'(I_q \otimes (\Sigma_{(1)}^0)^{-1})(\Sigma_{(1)}^0)^{1/2}\Omega(\lambda_1^0) + \delta_2'S'(I_q \otimes (\Sigma_{(1)}^0)^{-1})(\Sigma_{(1)}^0)^{1/2}\Omega(\lambda_1^0)$
$-\delta_1'S'(I_q \otimes (\Sigma_{(2)}^0)^{-1})(\Sigma_{(1)}^0)^{1/2}\Omega(\lambda_1^0)$,

$B = \delta_2' S'(I_q \otimes (\Sigma^0_{(2)})^{-1})(\Sigma^0_{(3)})^{1/2}\Omega(\lambda^0_2)$,

$$\begin{aligned}
\Pi_1 &= (1/2)tr(A_1^2) + \delta_1' S'(I_q \otimes (\Sigma^0_{(1)})^{-1})(D(\lambda^0_1) \otimes I_n)S\delta_1 \\
&\quad +\delta_2' S'(I_q \otimes (\Sigma^0_{(1)})^{-1})(D(\lambda^0_1) \otimes I_n)S\delta_2 \\
&\quad -\delta_2' S'(I_q \otimes (\Sigma^0_{(2)})^{-1})(D(\lambda^0_1) \otimes I_n)S\delta_2] \\
\Pi_2 &= (1/2)tr(A_2^2) \\
&\quad +\delta_2' S'(I_q \otimes (\Sigma^0_{(2)})^{-1})(D(\lambda^0_2) \otimes I_n)S\delta_2 \\
\Pi_{12} &= \delta_1' S'(I_q \otimes (\Sigma^0_{(1)})^{-1})(D(\lambda^0_1) \otimes I_n)S\delta_2 \\
\Upsilon_1 &= (1/4)vec(A_{(1)})'\Omega vec(A_{(1)}) + A'A \\
\Upsilon_2 &= (1/4)vec(A_{(2)})'\Omega vec(A_{(2)}) + B'B
\end{aligned}$$

Then

$$\Pi_1 v_T^2(\hat{K}_1 - K_1^0, \hat{K}_2 - K_2^0) \Rightarrow \arg\max_{0 \le v_1 \le v_2} H^2(v_1, v_2)$$

where

$$\begin{aligned}
H^3(v_1, v_2) &= -(1/2)|v_1| - (1/2)|v_2|\frac{\Pi_2}{\Pi_1} - |v_1|\frac{\Pi_{12}}{\Pi_1} \\
&\quad -(\frac{\Upsilon_1}{\Pi_1})^{1/2}B_1(v_1) - (\frac{\Upsilon_2}{\Pi_1})^{1/2}B_2(v_2).
\end{aligned}$$

For **Case 3** ($K_1 \le K_1^0 < K_2^0 \le K_2$), let $A = \delta_1' S'(I_q \otimes (\Sigma^0_{(2)})^{-1})(\Sigma^0_{(1)})^{1/2}\Omega(\lambda^0_1)$,
$B = \delta_2' S'(I_q \otimes (\Sigma^0_{(2)})^{-1})(\Sigma^0_{(3)})^{1/2}\Omega(\lambda^0_2)$,

$$\begin{aligned}
\Pi_1 &= (1/2)tr(A_1^2) + \delta_1' S'(I_q \otimes (\Sigma^0_{(2)})^{-1})(D(\lambda^0_1) \otimes I_n)S\delta_1 \\
\Pi_2 &= (1/2)tr(A_2^2) + \delta_2' S'(I_q \otimes (\Sigma^0_{(3)})^{-1})(D(\lambda^0_2) \otimes I_n)S\delta_2 \\
\Pi_{12} &= \delta_1' S'(I_q \otimes (\Sigma^0_{(1)})^{-1})(D(\lambda^0_1) \otimes I_n)S\delta_2 \\
\Upsilon_1 &= (1/4)vec(A_{(1)})'\Omega vec(A_{(1)}) + A'A \\
\Upsilon_2 &= (1/4)vec(A_{(2)})'\Omega vec(A_{(2)}) + B'B
\end{aligned}$$

then

$$\Pi_1 v_T^2(\hat{K}_1 - K_1^0, \hat{K}_2 - K_2^0) \Rightarrow \arg\max_{v_1 < 0, v_2 > 0} H^3(v_1, v_2)$$

where

$$\begin{aligned} H^2(v_1, v_2) &= -(1/2)|v_1| - (1/2)|v_2|\frac{\Pi_2}{\Pi_1} \\ &+(\frac{\Upsilon_1}{\Pi_1})^{1/2} B_1(v_1) - (\frac{\Upsilon_2}{\Pi_1})^{1/2} B_2(v_2) \end{aligned}$$

In summary:

Theorem 1.5 *Under Assumptions A1-A9, with two locally ordered breaks in β^0 and/or Σ^0, we have:*

$$\Pi_1 v_T^2(\hat{K}_1 - K_1^0, \hat{K}_2 - K_2^0) \Rightarrow \arg\max_{v_1 \leq v_2} H(v_1, v_2)$$

where $H(v_1, v_2) = 0$ if $v_1 = v_2 = 0$, and for Case 1 with $v_1 \leq v_2 \leq 0$: $H(v_1, v_2) = H^1(v_1, v_2)$; for Case 2 with $0 \leq v_1 \leq v_2$: $H(v_1, v_2) = H^2(v_1, v_2)$ and for Case 3 with $v_1 < 0$ and $v_2 > 0$: $H(v_1, v_2) = H^3(v_1, v_2)$.

Limit Distribution of the Locally Ordered Break Dates with Breaks in both Coefficients and Variance-covariance Matrix; arbitrary number of breaks.

For this most general case, we present only the main results. The derivations are similar as for simpler case, though much more tedious.

Case 1: $K_1 < ... < K_m \leq K_1^0 < ... < K_m^0$. We have

$$\begin{aligned} & rlr_T^2(K_{1,...}, K_m, \beta_{(j)}, \Sigma_{(j)}) \Rightarrow H_T^1(s_1, \cdots, s_m) \\ = \; & (1/2)\textstyle\sum_{j=1}^m tr(A_{(j)}\xi_j(s_j)) - (1/4)\sum_{j=1}^m |s_j| tr(A_{(j)}^2) \\ & -(1/2)\textstyle\sum_{j=1}^m \delta_j'(\sum_{l=j}^m |s_l| S'(I_q \otimes (\Sigma_{(l+1)}^0)^{-1})(D(\lambda_l^0) \otimes I_n)S)\delta_j \\ & +(1/2)\textstyle\sum_{j=1}^m \delta_j'(\sum_{l=j+1}^m |s_l| S'(I_q \otimes (\Sigma_{(l)}^0)^{-1})(D(\lambda_l^0) \otimes I_n)S)\delta_j \\ & -\textstyle\sum_{j=1}^{m-1}\sum_{i=j+1}^m \delta_j' \sum_{l=i}^m |s_l| S'(I_q \otimes (\Sigma_{(l+1)}^0)^{-1})(D(\lambda_l^0) \otimes I_n)S)\delta_i \end{aligned}$$

Let

$$\begin{aligned} Q_{j+1}(\lambda_j^0) &= S'(I_q \otimes (\Sigma_{(j+1)}^0)^{-1})(D(\lambda_j^0) \otimes I_n)S), \\ Q_j(\lambda_j^0) &= S'(I_q \otimes (\Sigma_{(j)}^0)^{-1})(D(\lambda_j^0) \otimes I_n)S) \\ \Delta_{j+1}(\lambda_j^0) &= S'(I_q \otimes (\Sigma_{(j+1)}^0)^{-1})(\Sigma_{(j)}^0)^{1/2} \\ \Delta_j(\lambda_j^0) &= S'(I_q \otimes (\Sigma_{(j)}^0)^{-1})(\Sigma_{(j)}^0)^{1/2} \end{aligned}$$

Then

$$
\begin{aligned}
H_T^1(s_1,\cdots,s_m) &= -(1/2)\textstyle\sum_{j=1}^m tr(A_{(j)}\xi_j(s_j)) \\
&\quad -(1/4)\textstyle\sum_{j=1}^m |s_j| tr(A_{(j)}^2) \\
&\quad -(1/2)\textstyle\sum_{j=1}^m |s_j|(\sum_{l=1}^j \delta_l' Q_{j+1}(\lambda_j^0)\delta_l \\
&\quad +(1/2)\textstyle\sum_{j=2}^m |s_j|\sum_{l=1}^{j-1} \delta_l' Q_j(\lambda_j^0)\delta_l \\
&\quad -\textstyle\sum_{j=2}^m |s_j|(\sum_{l=1}^{j-1}\sum_{i=l+1}^j \delta_l' Q_{j+1}(\lambda_j^0)\delta_i \\
&\quad +\textstyle\sum_{j=3}^m |s_j|(\sum_{l=1}^{j-1}\sum_{i=l+1}^j \delta_l' Q_j(\lambda_j^0)\delta_l \\
&\quad +\textstyle\sum_{j=1}^m (\sum_{l=1}^j \delta_l' \Delta_{j+1}(\lambda_j^0)W_j(s_j) \\
&\quad +\textstyle\sum_{j=2}^m (\sum_{l=1}^{j-1} \delta_l' \Delta_j(\lambda_j^0)W_j(s_j) \\
&= -(1/2)|s_1|((1/2)tr(A_{(1)}^2) + \delta_1' Q_2(\lambda_1^0)\delta_1) \\
&\quad -|s_2|\delta_1' Q_3(\lambda_2^0)\delta_2 - (1/2)\textstyle\sum_{j=2}^m |s_j|[(1/2)tr(A_{(j)}^2) \\
&\quad +\textstyle\sum_{l=1}^j \delta_l' Q_{j+1}(\lambda_j^0)\delta_l - \sum_{l=1}^{j-1}\delta_l' Q_j(\lambda_j^0)\delta_l] \\
&\quad -\textstyle\sum_{j=3}^m |s_j|[\sum_{l=1}^{j-1}\sum_{i=l+1}^j \delta_l' Q_{j+1}(\lambda_j^0)\delta_i \\
&\quad -\textstyle\sum_{l=1}^{j-2}\sum_{i=l+1}^{j-1} \delta_l' Q_j(\lambda_j^0)\delta_i] \\
&\quad +(1/2)tr(A_{(1)}\xi_1(s_1)) + \delta_1'\Delta_1(\lambda_1^0)W_1(s_1) \\
&\quad +(1/2)\textstyle\sum_{j=2}^m [tr(A_{(j)}\xi_j(s_j)) \\
&\quad +(\textstyle\sum_{l=1}^j \delta_l'\Delta_{j+1}(\lambda_j^0) + \sum_{l=1}^{j-1}\delta_l'\Delta_j(\lambda_j^0))W_j(s_j)]
\end{aligned}
$$

Applying similar arguments as before, let

$$
\begin{aligned}
\Pi_1 &= (1/2)tr(A_{(1)}^2) + \delta_1' Q_2(\lambda_1^0)\delta_1, \\
\Upsilon_1 &= (1/4)vec(A_{(1)})'\Omega vec(A_{(1)}) + \delta_1'\Delta_1(\lambda_1^0)\Delta_1(\lambda_1^0)'\delta_1,
\end{aligned}
$$

then we have

$$
\Pi_1 v_T^2(\hat{K}_1 - K_1^0, ..., \hat{K}_m - K_m^0) \Rightarrow \arg\max_{v_1\le\cdots\le v_m\le 0} H^1(v_1, ..., v_m)
$$

with

$$
\begin{aligned}
& H^1(v_1,...,v_m) \\
= \quad & -(1/2)|v_1| + (\frac{\Upsilon_1}{\Pi_1})^{1/2} B_1(v_1) - \frac{1}{\Pi_1}|v_2|\delta_1' Q_3(\lambda_2^0)\delta_2 \\
& -(1/2)\frac{1}{\Pi_1}\textstyle\sum_{j=2}^m |v_j|[(1/2)tr(A_{(j)}^2) \\
& +\textstyle\sum_{l=1}^j \delta_l' Q_{j+1}(\lambda_j^0)\delta_l - \sum_{l=1}^{j-1} \delta_l' Q_j(\lambda_j^0)\delta_l] \\
& -\frac{1}{\Pi_1}\textstyle\sum_{j=3}^m |v_j|[\sum_{l=1}^{j-1}\sum_{i=l+1}^j \delta_l' Q_{j+1}(\lambda_j^0)\delta_i \\
& -\textstyle\sum_{l=1}^{j-2}\sum_{i=l+1}^{j-1} \delta_l' Q_j(\lambda_j^0)\delta_i] \\
& +\textstyle\sum_{j=2}^m (\frac{1}{\Pi_1})^{1/2}[(1/4)vec(A_{(j)})'\Omega vec(A_{(j)}) \\
& +(\textstyle\sum_{l=1}^j \delta_l'\Delta_{j+1}(\lambda_j^0) + \sum_{l=1}^{j-1}\delta_l'\Delta_j(\lambda_j^0))' \\
& (\textstyle\sum_{l=1}^j \delta_l'\Delta_{j+1}(\lambda_j^0) + \sum_{l=1}^{j-1}\delta_l'\Delta_j(\lambda_j^0))]^{1/2} B_j(v_j)
\end{aligned}
$$

Using similar arguments, we obtain the following for

Case 2 ($K_1^0 < ... < K_m^0 \leq K_1 < ... < K_m$).

Let $\Pi_1 = ((1/2)tr(A_1^2) + \sum_{l=1}^m \delta_l' Q_j(\lambda_1^0)\delta_l - \sum_{l=2}^m \delta_l' Q_{j+1}(\lambda_1^0)\delta_l$,

$\Upsilon_1 = (1/4)vec(A_1)'\Omega vec(A_1) + (\sum_{l=1}^m \delta_l'\Delta_{j+1}(\lambda_1^0) - \sum_{l=2}^{jm} \delta_l'\Delta_{j+1}(\lambda_1^0))$

$(\sum_{l=1}^m \delta_l'\Delta_{j+1}(\lambda_1^0) - \sum_{l=2}^{jm} \delta_l'\Delta_{j+1}(\lambda_1^0))'$, then

$$\Pi_1 v_T^2(\hat{K}_1 - K_1^0, ..., \hat{K}_m - K_m^0) \Rightarrow \arg \max_{0 \leq v_1 \leq ... \leq v_m} H^2(v_1,...,v_m)$$

where

$$
\begin{aligned}
H^2(v_1,...,v_m) \quad = \quad & -(1/2)|v_1| - (\frac{\Upsilon_1}{\Pi_1})^{1/2} B_1(v_1) \\
& -(\frac{1}{\Pi_1})^{1/2}[(1/4)vec(A_{(m)})'\Omega vec(A_{(m)}) \\
& +\delta_m'\Delta_m(\lambda_m^0)\Delta_m(\lambda_m^0)'\delta_m]^{1/2} B_m(v_m) \\
& -(1/2)\frac{1}{\Pi_1}\textstyle\sum_{j=2}^{m-1} |v_j|[(1/2)tr(A_{(j)}^2) \\
& +\textstyle\sum_{l=j}^m \delta_l' Q_j(\lambda_j^0)\delta_l - \sum_{l=j+1}^m \delta_l' Q_{j+1}(\lambda_j^0)\delta_l] \\
& -(1/2)\frac{1}{\Pi_1}|v_m|[(1/2)tr(A_{(m)}^2) + \delta_m' Q_m(\lambda_m^0)\delta_m] \\
& -|v_{m-1}|\frac{1}{\Pi_1}[\delta_{m-1}' Q_{m-1}(\lambda_{m-1}^0)\delta_m]
\end{aligned}
$$

$$-\frac{1}{\Pi_1}\sum_{j=1}^{m-2}|v_j|[\sum_{l=j}^{m-1}\sum_{i=j+1}^{m}\delta_l'Q_j(\lambda_j^0)\delta_i$$
$$-\sum_{l=j+1}^{m-1}\sum_{i=j+1}^{m}\delta_l'Q_{j+1}(\lambda_j^0)\delta_i]$$
$$-\sum_{j=2}^{m-1}(\frac{1}{\Pi_1})^{1/2}[(1/4)vec(A_{(j)})'\Omega vec(A_{(j)})(\sum_{l=j}^{m}\delta_l'\Delta_j(\lambda_j^0)$$
$$-\sum_{l=j+1}^{m}\delta_l'\Delta_{j+1}(\lambda_j^0))$$
$$(\sum_{l=j}^{m}\delta_l'\Delta_j(\lambda_j^0)-\sum_{l=j+1}^{m}\delta_l'\Delta_{j+1}(\lambda_j^0))']^{1/2}B_j(v_j)$$

Case 3 ($K_1 < ... < K_b \le K_1^0 \le ... \le K_m^0 \le K_{b+1} < ... < K_m$, for some $1 \le b \le m$). Let

$$\begin{aligned}\Pi_1 &= (1/2)tr(A_1^2)+\delta_1'Q_2(\lambda_1^0)\delta_1,\\ \Upsilon_1 &= (1/4)vec(A_1)'\Omega vec(A_1)+\delta_1'\Delta_1(\lambda_1^0)\Delta_1'(\lambda_1^0)\delta_1.\end{aligned}$$

Then

$$\Pi_1 v_T^2(\hat{K}_1-K_1^0,...,\hat{K}_m-K_m^0)\Rightarrow \arg\max_{v_1\le...\le v_b\le 0\le v_{b+1}\le...\le v_m} H^3(v_1,...,v_m)$$

where

$$\begin{aligned}H^3(v_1,...,v_m) &= -(1/2)|v_1|+B_1(v_1)-|s_2|\frac{1}{\Pi_1}\delta_1'Q_3(\lambda_1^0)\delta_2\\
&-(1/2)\frac{1}{\Pi_1}\sum_{j=2}^{b}|v_j|[\frac{1}{2}tr(A_{(j)}^2)\\
&+\sum_{l=1}^{j}\delta_l'Q_{j+1}(\lambda_j^0)\delta_l-\sum_{l=1}^{j-1}\delta_l'Q_j(\lambda_j^0)\delta_l]\\
&-\frac{1}{\Pi_1}\sum_{j=3}^{b}|v_j|(\sum_{l=1}^{j-1}\sum_{i=l+1}^{j}\delta_l'Q_{j+1}(\lambda_j^0)\delta_i\\
&-\sum_{l=1}^{j-2}\sum_{i=l+1}^{j-1}\delta_l'Q_j(\lambda_j^0)\delta_i)\\
&+\sum_{j=2}^{b}(\frac{1}{\Pi_1})^{1/2}[(1/4)vec(A_{(j)})'\Omega vec(A_{(j)})\\
&+(\sum_{l=1}^{j}\delta_l'\Delta_{j+1}(\lambda_j^0)\\
&+\sum_{l=1}^{j-1}\delta_l'\Delta_j(\lambda_j^0))(\sum_{l=1}^{j}\delta_l'\Delta_{j+1}(\lambda_j^0)\\
&+\sum_{l=1}^{j-1}\delta_l'\Delta_j(\lambda_j^0))']^{1/2}B_j(v_j)\\
&(1/2)\frac{1}{\Pi_1}\sum_{j=b+1}^{m-1}|v_j|[(1/2)tr(A_{(j)}^2)\\
&+\sum_{l=j}^{m}\delta_l'Q_j(\lambda_j^0)\delta_l-\sum_{l=j+1}^{m}\delta_l'Q_{j+1}(\lambda_j^0)\delta_l]\end{aligned}$$

$$- \sum_{j=b+1}^{m-1} (\frac{1}{\Pi_1})^{1/2}[(1/4)vec(A_{(j)})^{-1}\Omega vec(A_{(j)})$$
$$+(\sum_{l=j}^{m} \delta_l' \Delta_j(\lambda_j^0) - \sum_{l=j+1}^{m} \delta_l' \Delta_{j+1}(\lambda_j^0))(\sum_{l=j}^{m} \delta_l' \Delta_j(\lambda_j^0)$$
$$- \sum_{l=j+1}^{m} \delta_l' \Delta_{j+1}(\lambda_j^0))']^{1/2} B_j(v_j)$$
$$-(\frac{1}{\Pi_1})^{1/2}[(1/4)vec(A_{(m)})'\Omega vec(A_{(m)})$$
$$+\delta_m' \Delta_m(\lambda_j^0)\Delta_m(\lambda_j^0)'\delta_m]^{1/2} B_m(v_m)$$
$$-(1/2)\frac{1}{\Pi_1}|v_m|[(1/2)tr(A_{(m)}^2) + \delta_m' Q_m(\lambda_j^0)\delta_m]$$
$$-\frac{1}{\Pi_1}|v_{m-1}|\delta_{m-1}' Q_{m-1}(\lambda_j^0)\delta_m$$
$$-\frac{1}{\Pi_1}\sum_{j=b+1}^{m-2} |v_j|[\sum_{l=j}^{m-1} \sum_{i=j+1}^{m} \delta_l' Q_j(\lambda_j^0)\delta_i$$
$$- \sum_{l=j+1}^{m-1} \sum_{i=j+1}^{m} \delta_l' Q_{j+1}(\lambda_j^0)\delta_i]$$

In summary:

Theorem 1.6 *Under Assumptions A1-A9, with m locally ordered breaks in β^0 and/or Σ^0:*

$$\Pi_1 v_T^2(\hat{K}_1 - K_1^0, ..., \hat{K}_m - K_m^0) \Rightarrow \arg \max_{v_1 \leq ... \leq v_m} H(v_1, ..., v_m)$$

where $H(v_1, ..., v_m) = 0$ if $v_1 = ... = v_m = 0$, and for Case 1 with $v_1 \leq ... \leq v_m \leq 0$: $H(v_1, ..., v_m) = H^1(v_1, ..., v_m)$; for Case 2 with $0 \leq v_1 \leq ... \leq v_m$: $H(v_1, ..., v_m) = H^2(v_1, ..., v_m)$ and for Case 3 with $v_1 \leq ... \leq v_b \leq 0 \leq v_{b+1} \leq ... \leq v_m$: $H(v_1, ..., v_m) = H^3(v_1, ..., v_m)$.

Remark 1.6 *Under shrinking magnitudes of shifts, $\Sigma_{(j)}^0 \to \Sigma^0$ as $T \to \infty$, for $j = 1, ..., m+1$. Hence, $\Sigma_{(j)}^0$ could be replaced by Σ^0. However, using the results with $\Sigma_{(j)}^0$ is likely to provide better approximations.*

Chapter 2

Inference Related to Locally Ordered and Common Breaks in a Multivariate System with Joined Segmented Trends[1]

2.1 Introduction

Issues related to structural breaks have received a lot of attention in the statistics and econometrics literature (see Perron, 2006, for a survey). Substantial advances have been made to cover general models in the context of estimating and testing structural breaks in both single and multiple equations systems. In the single equation case, 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 the errors. Bai and Perron (1998) extend the testing and estimation analysis to the case of multiple structure changes, while Bai and Perron (2003) present an efficient algorithm to obtain the break date estimates as the global minimizers of the overall sum of squared residuals.

Much of the work in the literature concentrated on the case where the regressors and errors are stationary. Nevertheless, issues related to structural changes are also important in the context of trending regressors and non-stationary time series. Perron and Zhu (2005) consider a linear trend function subject to a one-time change in the parameters. They analyze the consistency, rate of convergence and limiting distributions of the parameters with errors that can be stationary or have an autoregressive unit root. They consider three different models: a "joint broken trend", a "local disjoint broken trend" and a "global disjoint broken trend". They show that each case involves different asymptotic results, in particular pertaining to the rate of convergence and the asymptotic distribution of the estimates of the break dates. The model we consider in this paper

[1] **This chapter is joint work with Professor Pierre Perron**

is the "joint broken trend" model, whereby the slope of the trend changes and the series is joined at the time of the break.

Advances have also been made for structural change problems in the context of a multiple equations system. Bai, Lumsdaine and Stock (1998) develop methods to construct a confidence interval for the estimate of a single break date in multivariate system, assuming a priori the break date to be common across all equations. Bai (2000) analyses multiple structural changes in both the regression coefficients and the variance-covariance matrix in vector autoregressive models. Qu and Perron (2007) provide a comprehensive treatment of issues related to estimation, inference and computation with multiple structural changes that occur at unknown dates in linear multivariate regression models, including VAR, certain linear panel data models, and seemingly unrelated regression. They also introduce a novel structure labelled as "locally ordered breaks". These occur when one has prior knowledge of which coefficient is subject to the first break and when the subsequent break date is local in the sense that the distance is not a positive fraction of the sample size (and can even be a fixed small number), as usually assumed in the structural change literature. They provide appropriate methods for estimation, inference and testing for such locally ordered breaks in the context of a bivariate system with stationary regressors. As for testing in a multivariate system, Oka and Perron (2011) provide a general framework to test for common breaks across or within equations in a multivariate system with stationary, trending and/or unit root regressors. The null hypothesis is that some subsets of the regressors share some common breaks, with the break dates separated by a positive fraction of the sample size, while the alternative hypothesis is that the break dates are not the same and need not to be asymptotically distinct. Li and Perron (2012), building on the work of Oka and Perron (2001), extend the analysis of locally ordered breaks to cover systems with stationary, trending and/or unit root regressors.

A problem with the analysis of Oka and Perron (2011) and Li and Perron (2012) is that, while trends are permitted, the trend function cannot be restricted to be joined at the time of the break. Such breaking trends are very relevant in practice as evidenced by many series in macroeconomics, finance and even climate change (e.g., Estrada, Perron and Martinez, 2013). The latter case is indeed the motivation behind this paper as global and hemispheric temperatures as well as radiative forcings (e.g., greenhouse gases) are well approximated by a linear trend with a one-time change in slope near 1955 with the noise component being stationary. As shown in

Perron and Zhu (2005), the limit results with joined segmented trends are very different from locally or globally disjoined trends. Hence, the need for a separate treatment. The aim of this paper is to provide the relevant results to test for common breaks in a bivariate system of trending series, as in Oka and Perron (2011), as well as to extend the results of Qu and Perron (2007) and Li and Perron (2012) for the analysis of locally ordered breaks in such systems.

The structure of this paper is as follows. Section 2 presents the model and results about the test for common breaks. Section 3 discusses inference about locally ordered break dates. Section 4 provides some simulation results to assess the adequacy of the limit distributions in providing useful approximations in finite sample. Section 5 offers brief concluding remarks and technical derivations are contained in an appendix.

2.2 Testing Common Breaks

We adopt a framework similar to that of Perron and Zhu (2005), extended to have multiple breaks and considering a bivariate system. Each variable is represented by a linear trend with multiple changes in slope such that the trend function is joined at each break date. There are m_1 breaks in the trend of the first variable and m_2 breaks for the second variable. More specifically, the bivariate system is, for $t = 1, ..., T$, with T the sample size:

$$\begin{aligned} y_{1t} &= \mu_1^0 + \beta_1^0 t + \textstyle\sum_{j_1=1}^{m_1} \delta_{1j_1}^0 1(t \geqslant K_{1j_1}^0)(t - K_{1j_1}^0) + u_{1t} \\ y_{2t} &= \mu_2^0 + \beta_2^0 t + \textstyle\sum_{j_2=1}^{m_2} \delta_{2j_2}^0 1(t \geqslant K_{2j_2}^0)(t - K_{2j_2}^0) + u_{2t} \end{aligned}$$

where $K_{1j_1}^0$ $(j_1 = 1, ..., m_1)$ are the break dates for the changes (with magnitudes $\delta_{1j_1}^0$) in the trend of the first variable and $K_{2j_2}^0$ $(j_2 = 1, ..., m_2)$ are break dates for the changes (with magnitudes $\delta_{2j_2}^0$) in the trend of the second variable. In matrix form, we have, with $y_t = (y_{1t}, y_{2t})'$ and $U_t = (u_{1t}, u_{2t})'$,

$$y_t = X_t^{0\prime}\theta^0 + U_t$$

with

$$
\begin{aligned}
X_t^{0\prime} &= \begin{bmatrix} 1 & t & 1(t \geqslant K_{11}^0)(t-K_{11}^0) & \cdots & 1(t \geqslant K_{1m_1}^0)(t-K_{1m_1}^0) \\ 0 & & \cdots & & 0 \end{bmatrix} \\
&\quad ... \begin{bmatrix} 0 & \cdots & & & 0 \\ 1 & t & 1(t \geqslant K_{11}^0)(t-K_{11}^0) & \cdots & 1(t \geqslant K_{2m_2}^0)(t-K_{2m_2}^0) \end{bmatrix}
\end{aligned}
$$

and

$$
\theta^{0\prime} = \begin{bmatrix} \mu_1^0 & \beta_1^0 & \delta_{11}^0 & \cdots & \delta_{1m_1}^0 & \mu_2^0 & \beta_2^0 & \delta_{21}^0 & \cdots & \delta_{2m_2}^0 \end{bmatrix}.
$$

It is assumed that U_t has mean 0 and covariance matrix Σ^0 and long-run covariance Ω^0, i.e., $\lim_{T\to\infty} E(T^{-1}\sum_{t=1}^{T} U_t U_t') = \Omega^0$.

For the bivariate system, it is assumed that there are m breaks, denoted by $T_1^0, ..., T_m^0$. We are interested in testing whether a break date for one variable is common to one for the other variable. Hence, the null and alternative hypotheses are:

$$
\begin{aligned}
H_0 &: K_{1l_1}^0 = K_{2l_2}^0 = T_l^0 \\
H_1 &: K_{1l_1}^0 \neq K_{2l_2}^0
\end{aligned}
$$

for some l_1 and l_2. We consider testing only for a single common break date, allowing the possibility of other common breaks which are, however, not the object of testing. The analysis could be extended to cover the more general case of testing for multiple common breaks. We do not pursue this extension here as the case of more practical interest is that for which there is a single common break.

To make the notation clear consider the following simple example. There are 4 breaks i.n the first equation and 2 breaks in the second equation. The order of the break dates are as follows: the first equation has the first break at K_{11}^0, the second break is common to both equations, so that $K_{12}^0 = K_{21}^0$. The next break is in the first equation at K_{13}^0 followed by a break in the second equation at K_{22}^0. The last break is in the first equation at K_{14}^0. Hence, there are 5 breaks in total given by $T_1^0 = K_{11}^0, T_2^0 = K_{12}^0 = K_{21}^0$, $T_3^0 = K_{13}^0$, $T_4^0 = K_{22}^0$ and $T_5^0 = K_{14}^0$. In this example, $m = m_1 + m_2 - 1$.

The estimation method we consider is Quasi-Maximum Likelihood assuming serially uncorrelated Gaussian errors. We define the sets $K_1^0 = \{K_{11}^0, ..., K_{1m_1}^0\}$ and $K_2^0 = \{K_{21}^0, ..., K_{2m_2}^0\}$. Also

$\tau^0 = \{T_1^0, ..., T_m^0\}$ is the set of true break dates for the bivariate system. The candidate break dates for the two equations are $K_1 = \{K_{11}, ..., K_{1m_1}\}$ and $K_2 = \{K_{21}, ..., K_{2m_2}\}$. Under H_0, the set of candidate break dates are

$$\tau = \{T_1, ..., T_m; K_{1l_1} = K_{2l_2} \text{ for some } l_1 \text{ and } l_2\},$$

while under H_1 the set is unrestricted.

Let $\widehat{LR}_T$ be the likelihood ratio under H_1, $\widetilde{LR}_T$ be the likelihood ratio under H_0, $\widehat{lr}_T$ be the log-likelihood ratio under H_1 and $\widetilde{lr}_T$ be the log-likelihood ratio under H_0. Also, $\widetilde{X}_t$ denotes the regressor matrix and $(\widetilde{\theta}, \widetilde{\Sigma})$ the estimates of the coefficients and the variance-covariance matrix under H_0. Under H_1 we use the notation $\widehat{X}_t$ and $(\widehat{\theta}, \widehat{\Sigma})$. Then,

$$\begin{aligned} \widehat{LR}_T &= \frac{\prod_{t=1}^T f(Y_t|\widehat{X}_t, \widehat{\theta}, \widehat{\Sigma})}{\prod_{t=1}^T f(Y_t|X_t^0, \theta^0, \Sigma^0)} \\ \widetilde{LR}_T &= \frac{\prod_{t=1}^T f(Y_t|\widetilde{X}_t, \widetilde{\theta}, \widetilde{\Sigma})}{\prod_{t=1}^T f(Y_t|X_t^0, \theta^0, \Sigma^0)} \end{aligned}$$

and

$$\begin{aligned} \widehat{lr}_T &= \textstyle\sum_{t=1}^T \log f(Y_t|\widehat{X}_t, \widehat{\theta}, \widehat{\Sigma}) - \sum_{t=1}^T \log f(Y_t|X_t^0, \theta^0, \Sigma^0) \\ \widetilde{lr}_T &= \textstyle\sum_{t=1}^T \log f(Y_t|\widetilde{X}_t, \widetilde{\theta}, \widetilde{\Sigma}) - \sum_{t=1}^T \log f(Y_t|X_t^0, \theta^0, \Sigma^0) \end{aligned}$$

The test statistic is the likelihood ratio test that compares the values of the likelihood function under the null hypothesis of a common break date and the alternative hypothesis of distinct break dates. It is defined as:

$$CB_T = \max_{T_1, ..., T_m} \widehat{lr}_T(\widehat{X}_t, \widehat{\theta}, \widehat{\Sigma}) - \max_{T_1, ..., T_m \in \tau} \widetilde{lr}_T(\widetilde{X}_t, \widetilde{\theta}, \widetilde{\Sigma})$$

2.2.1 Theoretical Results

We now consider the limit distribution of the test statistic for common breaks. We first state the assumptions needed to obtain the required results.

Assumption 2.1 $0 < \lambda_1^0 < ... < \lambda_m^0 < 1$, *with* $T_j^0 = [T\lambda_j^0]$.

Assumption 2.2 $\delta_{1j_1}^0 \neq 0$ *and* $\delta_{2j2}^0 \neq 0$ *for* $j_1 = 1, ..., m_1$ *and* $j_2 = 1, ..., m_2$.

Assumption 2.3 *Let* $\mathcal{F}_t = \sigma-$*field* $\{\cdots, u_{t-2}, u_{t-1}\}$. *If* u_t *is weakly stationary within each segment, then (a)* $\{u_t, \mathcal{F}_t\}$ *forms a strongly mixing (*α*-mixing) sequence with size* $-4r/(r-2)$ *for* *some* $2 < r < 8$. *(b)* $E(u_t) = 0$ *and* $\sup_t ||u_t||_{2r+\delta} < M < \infty$ *for some* $\delta > 0$ *and* $M > 0$, *(c) let* $S_{k,j}(l) = \sum_{K_{j-1}^0+l+1}^{K_{j-1}^0+l+k} u_t, j = 1, ..., m+1$, *for each* $e \in R^n$ *of length 1,* $var(< e, S_{k,j}(0) >) \geqslant v(k)$ *for some function* $v(k) \to \infty$ *as* $k \to \infty$ *(with* $<\cdot>$*, the usual inner product). If* u_t *is not weakly stationary within each segment, we assume that (a)-(c) holds, and in addition, that there exists a positive definite matrix* $\Omega^* = [w_{i,s}]$ *such that for any* $i, s = 1, ..., p$*, we have, uniformly in* l, $|k^{-1}E((S_{k,j}(l))_i \ (S_{k,j}(l))_s) - w_{i,s}| \leq C_2 k^{-\psi}$*, for some* $C_2, \psi > 0$. *It is also assumed that* $\{u_t u_t' - \Sigma^0\}$ *satisfies the conditions stated in this assumption.*

Assumptions 1 and 2 are standard and simply state that the break dates are asymptotically distinct (i.e., each regime increases proportionally with the sample size T) and the changes in the parameters are non-zero at the break dates. Assumption A3 determines the dependence structure of the processes u_t. In particular, they imply that u_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. They are the same as those in Oka and Perron (2011) and Li and Perron (2012), so that we can use some of their results for trending series. Note, in particular, that under assumption A3, we have for $\eta_t = (\Sigma^0)^{-1/2} u_t$,

$$\begin{aligned} T^{-1/2} \textstyle\sum_{t=1}^{T_j^0} \eta_t &\Rightarrow W(\lambda_j^0) = (W_1(\lambda_j^0), W_2(\lambda_j^0))' \\ &= (\Sigma^0)^{-1/2} (\Omega^0)^{1/2} W^*(\lambda_j^0) = (W_1^*(\lambda_j^0), W_2^*(\lambda_j^0))' \end{aligned}$$

where $W(\cdot)$ is a bivariate vector of Weiner process with vovariance matrix $(\Sigma^0)^{-1}\Omega^0$, $W^*(\cdot)$ is a bivariate vector of independent Wiener processes and "$\Rightarrow$" denotes weak convergence under the Skorohod topology.

We start with some preliminary results about the rate of convergence of the various estimates, whose proofs follow the developments in Perron and Zhu (2005), hence details are omitted.

Proposition 2.1 ***a) (Rate of Convergence Under*** H_1*) Under Assumptions 1-3 and assuming the break dates not to be common, the following quantities are all* $O_p(1)$*:* $\sqrt{T}(\hat{T}_i - T_i^0)$ *(*$i = 1, ..., m$*),* $\sqrt{T}(\hat{\mu}_i - \mu_i^0)$ *(*$i = 1, 2$*),* $T^{3/2}(\hat{\beta}_i - \beta_i^0)$ *(*$i = 1, 2$*),* $T^{3/2}(\hat{\delta}_{1j_1} - \delta_{1j_1}^0)$ *(*$j_1 = 1, ..., m_1$*),* $T^{3/2}(\hat{\delta}_{2j_2} - \delta_{2j_2}^0)$ *(*$j_2 = 1, ..., m_2$*); b)* ***(Rate of Convergence Under*** H_0*) Under Assumptions 1-3 and assuming the break dates to be common, the same results hold for the restricted estimates defined with a "$\sim$"instead of a "ˆ".*

The results in Proposition 1 allow us to analyze the properties of the common break test statistic under the following compact sets:

$$C_M^{(1)} = \{\tilde{\mu}_i, \tilde{\beta}_i, \tilde{\delta}_i, \tilde{T}_i : \sqrt{T}(\tilde{\mu}_i - \mu_i^0) \leq M, T^{3/2}(\tilde{\beta}_i - \beta_i^0) \leq M$$
$$T^{3/2}(\tilde{\delta}_{ij} - \delta_{ij}^0) \leq M, \sqrt{T}(\tilde{T}_i - T_i^0) \leq M\}$$

and

$$C_M^{(2)} = \{\hat{\mu}_i, \hat{\beta}_i, \hat{\delta}_i, \hat{T}_i : \sqrt{T}(\hat{\mu}_i - \mu_i^0) \leq M, T^{3/2}(\hat{\beta}_i - \beta_i^0) \leq M$$
$$T^{3/2}(\hat{\delta}_{ij} - \delta_{ij}^0) \leq M, \sqrt{T}(\hat{T}_i - T_i^0) \leq M\}.$$

Then, within the sets $C_M^{(1)}$ and $C_M^{(2)}$, the common break test statistic is:

$$CB_T = \max_{T_1,\ldots,T_m} \widehat{lr}_T^2(\hat{X}_t, \theta^0, \Sigma^0) - \max_{T_1,\ldots,T_m \in \tau} \widetilde{lr}_T^2(\tilde{X}_t, \theta^0, \Sigma^0)$$

with

$$\begin{aligned} \widehat{lr}_T^2(\hat{X}_t, \theta^0, \Sigma^0) &= -\frac{1}{2}[\textstyle\sum_{t=1}^T \theta^{0\prime}(X_t^0 - \hat{X}_t)(\Sigma^0)^{-1}(X_t^0 - \hat{X}_t)'\theta^0 \\ &\quad +2\textstyle\sum_{t=1}^T \theta^{0\prime}(X_t^0 - \hat{X}_t)(\Sigma^0)^{-1}U_t] \end{aligned}$$

and

$$\begin{aligned} \widetilde{lr}_T^2(\tilde{X}_t, \theta^0, \Sigma^0) &= -\frac{1}{2}[\textstyle\sum_{t=1}^T \theta^{0\prime}(X_t^0 - \tilde{X}_t)(\Sigma^0)^{-1}(X_t^0 - \tilde{X}_t)'\theta^0 \\ &\quad +2\textstyle\sum_{t=1}^T \theta^{0\prime}(X_t^0 - \tilde{X}_t)(\Sigma^0)^{-1}U_t] \end{aligned}$$

The limit distribution of the test statistic is stated in the following Theorem proved in the appendix.

Theorem 2.1 *Under the Assumptions 1-3, we have:*

$$CB_T \Rightarrow \max_{s_1,s_2} \hat{H}(s_1, s_2) - \max_{s_1,s_2 \text{ with } s_{1l_1} = s_{2l_2}} \tilde{H}(s_1, s_2)$$

with

$$
\begin{aligned}
\hat{H}(s_1,s_2) &= \max_{s_1,s_2} -(1/2)[2(\textstyle\sum_{j=1}^{l} tr((\Omega^0)^{1/2}\hat{A}_1) \\
&\quad + \textstyle\sum_{j=l+1}^{m+1} tr((\Omega^0)^{1/2}\hat{A}_2)) \\
&\quad + \textstyle\sum_{j=1}^{l} tr(\hat{B}_1) + \sum_{j=l+1}^{m+1} tr(\hat{B}_2))]
\end{aligned}
$$

where $s_1 = (s_{11}, ..., s_{1m_1})$ *and* $s_2 = (s_{21}, ..., s_{2m_2})$. *Also,*

$$
\begin{aligned}
\tilde{H}(s_1,s_2) &= -(1/2)[2\textstyle\sum_{j=1}^{m+1} tr((\Sigma^0)^{-1}\tilde{A} \\
&\quad + \textstyle\sum_{j=1}^{m+1}(\lambda_j^0 - \lambda_{j-1}^0) tr((\Sigma^0)^{-1}(\Omega^0)^{1/2}\tilde{B})]
\end{aligned}
$$

Also, the elements of the 2×2 *matrices* $\tilde{A}$, $\tilde{B}$, $\hat{A}_1$, $\hat{A}_2$, $\hat{B}_1$ *and* $\hat{B}_2$ *are defined by:*

$$
\begin{aligned}
\tilde{A}(1,1) &= (\textstyle\sum_{i_1=1}^{j_1-1}\delta_{1i_1}^0 s_{1i_1})^2 \\
\tilde{A}(1,2) &= \tilde{A}(2,1) = (\textstyle\sum_{i_1=1}^{j_1-1}\delta_{1i_1}^0 s_{1i_1})(\sum_{i_2=1}^{j_2-1}\delta_{2i_2}^0 s_{2i_2}) \\
\tilde{A}(2,2) &= (\textstyle\sum_{i_2=1}^{j_2-1}\delta_{2i_2}^0 s_{2i_2})^2
\end{aligned}
$$

$$
\begin{aligned}
\tilde{B}(1,1) &= (\textstyle\sum_{i_1=1}^{j_1-1}\delta_{1i_1}^0 s_{1i_1})(W_1^*(\lambda_j^0) - W_1^*(\lambda_{j-1}^0)) \\
\tilde{B}(1,2) &= (\textstyle\sum_{i_2=1}^{j_2-1}\delta_{2i_2}^0 s_{2i_2})(W_1^*(\lambda_j^0) - W_1^*(\lambda_{j-1}^0)) \\
\tilde{B}(2,1) &= (\textstyle\sum_{i_1=1}^{j_1-1}\delta_{1i_1}^0 s_{1i_1})(W_2^*(\lambda_j^0) - W_2^*(\lambda_{j-1}^0)) \\
\tilde{B}(2,2) &= (\textstyle\sum_{i_2=1}^{j_2-1}\delta_{2i_2}^0 s_{2i_2})(W_2^*(\lambda_j^0) - W_2^*(\lambda_{j-1}^0))
\end{aligned}
$$

$$
\begin{aligned}
\hat{A}_1(1,1) &= (\textstyle\sum_{i_1=1}^{l_1-1}\delta_{1i_1}^0 s_{1i_1})(W_1^*(\lambda_j^0) - W_1^*(\lambda_{j-1}^0)) \\
\hat{A}_1(1,2) &= (\textstyle\sum_{i_2=1}^{l_2-1}\delta_{2i_2}^0 s_{2i_2})(W_1^*(\lambda_j^0) - W_1^*(\lambda_{j-1}^0)) \\
\hat{A}_1(2,1) &= (\textstyle\sum_{i_1=1}^{l_1-1}\delta_{1i_1}^0 s_{1i_1})(W_2^*(\lambda_j^0) - W_2^*(\lambda_{j-1}^0)) \\
\hat{A}_1(2,2) &= (\textstyle\sum_{i_2=1}^{l_2-1}\delta_{2i_2}^0 s_{2i_2})(W_2^*(\lambda_j^0) - W_2^*(\lambda_{j-1}^0))
\end{aligned}
$$

$$
\begin{aligned}
\hat{A}_2(1,1) &= (\textstyle\sum_{i_1=1}^{l_1-1}\delta_{1i_1}^0 s_{1i_1} + \delta_{1l_1}^0 s_{1l_1} \\
&\quad + \textstyle\sum_{i_1=l_1+1}^{j_1-1}\delta_{1i_1}^0 s_{1i_1})(W_1^*(\lambda_j^0) - W_1^*(\lambda_{j-1}^0)) \\
\hat{A}_2(1,2) &= (\textstyle\sum_{i_2=1}^{l_2-1}\delta_{2i_2}^0 s_{2i_2} + \delta_{2l_2}^0 s_{2l_2} \\
&\quad + \textstyle\sum_{i_2=l_2+1}^{j_2-1}\delta_{2i_2}^0 s_{2i_2})(W_1^*(\lambda_j^0) - W_1^*(\lambda_{j-1}^0)) \\
\hat{A}_2(2,1) &= (\textstyle\sum_{i_1=1}^{l_1-1}\delta_{1i_1}^0 s_{1i_1} + \delta_{1l_1}^0 s_{1l_1} \\
&\quad + \textstyle\sum_{i_1=l_1+1}^{j_1-1}\delta_{1i_1}^0 s_{1i_1})(W_2^*(\lambda_j^0) - W_2^*(\lambda_{j-1}^0)) \\
\hat{A}_2(2,2) &= (\textstyle\sum_{i_2=1}^{l_2-1}\delta_{2i_2}^0 s_{2i_2} + \delta_{2l_2}^0 s_{2l_2} \\
&\quad + \textstyle\sum_{i_2=l_2+1}^{j_2-1}\delta_{2i_2}^0 s_{2i_2})(W_2^*(\lambda_j^0) - W_2^*(\lambda_{j-1}^0))
\end{aligned}
$$

$$\begin{aligned}
\hat{B}_1(1,1) &= (\lambda_j^0 - \lambda_{j-1}^0)(\textstyle\sum_{i_1=1}^{l_1-1} \delta_{1i_1}^0 s_{1i_1})^2 \\
\hat{B}_1(1,2) &= (\lambda_j^0 - \lambda_{j-1}^0)(\textstyle\sum_{i_1=1}^{l_1-1} \delta_{1i_1}^0 s_{1i_1})(\textstyle\sum_{i_2=1}^{l_2-1} \delta_{2i_2}^0 s_{2i_2}) \\
\hat{B}_1(2,1) &= (\lambda_j^0 - \lambda_{j-1}^0)(\textstyle\sum_{i_1=1}^{l_1-1} \delta_{1i_1}^0 s_{1i_1})(\textstyle\sum_{i_2=1}^{l_2-1} \delta_{2i_2}^0 s_{2i_2}) \\
\hat{B}_1(2,2) &= (\lambda_j^0 - \lambda_{j-1}^0)(\textstyle\sum_{i_2=1}^{l_2-1} \delta_{2i_2}^0 s_{2i_2})^2
\end{aligned}$$

$$\begin{aligned}
\hat{B}_2(1,1) &= (\lambda_j^0 - \lambda_{j-1}^0)(\textstyle\sum_{i_1=1}^{l_1-1} \delta_{1i_1}^0 s_{1i_1} \\
&\quad +\delta_{1l_1}^0 s_{1l_1} + \textstyle\sum_{i_1=l_1+1}^{j_1-1} \delta_{1i_1}^0 s_{1i_1}) \\
\hat{B}_2(1,2) &= \widehat{B}_2(2,1) \\
&= (\lambda_j^0 - \lambda_{j-1}^0)(\textstyle\sum_{i_1=1}^{l_1-1} \delta_{1i_1}^0 s_{1i_1} \\
&\quad +\delta_{1l_1}^0 s_{1l_1} + \textstyle\sum_{i_1=l_1+1}^{j_1-1} \delta_{1i_1}^0 s_{1i_1}) \\
&\quad \times(\textstyle\sum_{i_2=1}^{l_2-1} \delta_{1i_2}^0 s_{1i_2} + \delta_{2l_2}^0 s_{2l_2} \\
&\quad + \textstyle\sum_{i_2=l_2+1}^{j_2-1} \delta_{2i_2}^0 s_{2i_2}) \\
\hat{B}_2(2,2) &= (\lambda_j^0 - \lambda_{j-1}^0)(\textstyle\sum_{i_2=1}^{l_2-1} \delta_{2i_2}^0 s_{2i_2} \\
&\quad +\delta_{2l_2}^0 s_{2l_2} + \textstyle\sum_{i_2=l_2+1}^{j_2-1} \delta_{2i_2}^0 s_{2i_2})^2
\end{aligned}$$

Remark 2.1 *In the leading case of interest when testing for one common break in two equations, we have the DGP:*

$$\begin{aligned}
y_{1t} &= \mu_1^0 + \beta_1^0 t + \delta_1^0 1(t \geqslant K_1^0)(t - K_1^0) + u_{1t} \\
y_{2t} &= \mu_2^0 + \beta_2^0 t + \delta_2^0 1(t \geqslant K_2^0)(t - K_2^0) + u_{2t}
\end{aligned}$$

and we are interested in testing:

$$\begin{aligned}
H_0 &: K_1^0 = K_2^0 = T_1^0 \\
H_1 &: K_1^0 \neq K_2^0
\end{aligned}$$

The limit distributin of the test statitic reduces to:

$$\begin{aligned}
CB_T &\Rightarrow \arg\max_{s_1,s_2} \hat{H}(s_1,s_2) - \arg\max_{s_1} \tilde{H}(s_1) \\
&= \arg\max_{s_1,s_2}[-tr((\Omega^0)^{1/2} \\
&\quad \begin{pmatrix} \delta_1^0 s_1(W_1^*(1) - W_1^*(\lambda^0)) & \delta_2^0 s_2(W_1^*(1) - W_1^*(\lambda^0)) \\ \delta_1^0 s_1(W_2^*(1) - W_2^*(\lambda^0)) & \delta_2^0 s_2(W_2^*(1) - W_2^*(\lambda^0)) \end{pmatrix}) \\
&\quad -\frac{1-\lambda^0}{2} tr(\begin{pmatrix} \delta_1^{02} s_1^2 & \delta_1^0\delta_2^0 s_1 s_2 \\ \delta_1^0\delta_2^0 s_1 s_2 & \delta_2^{02} s_2^2 \end{pmatrix})]
\end{aligned}$$

$$-\arg\max_{s_1}[-tr((\Omega^0)^{1/2}$$
$$\begin{pmatrix} \delta_1^0 s_1(W_1^*(1)-W_1^*(\lambda^0)) & \delta_2^0 s_1(W_1^*(1)-W_1^*(\lambda^0)) \\ \delta_1^0 s_1(W_2^*(1)-W_2^*(\lambda^0)) & \delta_2^0 s_1(W_2^*(1)-W_2^*(\lambda^0)) \end{pmatrix})$$
$$-\frac{1-\lambda^0}{2}tr(\begin{pmatrix} \delta_1^{02}s_1^2 & \delta_1^0\delta_2^0 s_1^2 \\ \delta_1^0\delta_2^0 s_1^2 & \delta_2^{02}s_1^2 \end{pmatrix})]$$

2.3 Inference about Locally Ordered Breaks

We now consider the problem of forming a confidence interval for a pair of locally ordered breaks for changes in the slope of a trend function joined at the time of the break. Again, we consider a bivariate system with a single break in each equation. We also omit other possible breaks that are not locally ordered since confidence intervals for these can be constructed in the usual manner (see, Qu and Perron, 2007, and Li and Perron, 2012). Accordingly, the model is:

$$\begin{aligned} y_{1t} &= \mu_1^0 + \beta_1^0 t + \delta_1^0 1(t \geqslant K_1^0)(t-K_1^0) + u_{1t} \\ y_{2t} &= \mu_2^0 + \beta_2^0 t + \delta_2^0 1(t \geqslant K_2^0)(t-K_2^0) + u_{2t} \end{aligned}$$

Again, $U_t = (u_{1t}, u_{2t})$ has mean 0, variance Σ^0, long-run variance Ω^0 and satisfies Assumption A3.

Definition 2.1 *Locally Ordered Breaks (LOB): Let v_T be a sequence of positive numbers that satisfies $v_T \to 0$ and $T^{1/2}v_T/(\log^2 T) \to \infty$. K_1^0 and K_2^0 are said to be locally ordered if $K_1^0 \leqslant K_2^0$, and $v_T^2(K_2^0 - K_1^0) \leqslant M_T$ with $M_T \to 0$ as $T \to \infty$.*

Remark 2.2 *The condition $v_T^2(K_2^0 - K_1^0) \leqslant M_T$ with $M_T \to 0$ implies that $(K_2^0 - K_1^0)/T \to 0$. Hence, asymptotically the distance between the break dates becomes a negligible portion of the sample size. If $lim_{T\to\infty}(K_2^0 - K_1^0)/T > 0$, then the two break dates are asymptotically distinct and the usual asymptotic distribution theory applies.*

The method of estimation considered is restricted quasi-maximum likelihood that assumes serially uncorrelated Gaussian errors. Conditional on a given partition of the sample $K = (K_1, K_2)$, the Gaussian quasi-likelihood function is

$$L_T(K,\beta,\Sigma) = \textstyle\prod_{t=1}^T f(y_t|X_t;\theta,\Sigma)$$

$$f(y_t|X_t;\theta,\Sigma) = \frac{1}{(2\pi)^{1/2}|\Sigma|^{1/2}}\exp\{-\frac{1}{2}(y_t - X_t'\theta)'\Sigma^{-1}(y_t - X_t'\theta)\}$$

and the quasi-likelihood ratio is

$$LR_T(K,\beta,\Sigma) = \frac{\prod_{t=1}^{T} f(y_t|X_t;\theta,\Sigma)}{\prod_{t=1}^{T} f(y_t|X_t;\theta^0,\Sigma^0)}.$$

We wish to obtain the values of $(K_1, K_2,, \theta, \Sigma)$,which maximizes LR_T subject to restrictions $g(\theta^0, vec(\Sigma^0)) = 0$, if applicable. Let $lr_T(\cdot)$ denote the log-likelihood ratio and $rlr_T(\cdot)$ the restricted log-likelihood ratio, the objective function is then

$$rlr_T(K,\beta,\Sigma) = lr_T(K,\theta,\Sigma) + \lambda' g(\theta^0, vec(\Sigma^0))$$

and the estimates are

$$(\hat{K},\hat{\theta},\hat{\Sigma}) = \arg \max_{(K_1,K_2,\theta,\Sigma)} rlr_T(K,\theta,\Sigma).$$

The supremum with respect to (K_1, K_2) is taken over a restricted set of partitions $\Bbbk_\varepsilon$. For a small number $\varepsilon > 0$, $\Bbbk_\varepsilon = \{(K_1, K_2) : K_1 \geq [T\epsilon], K_2 - K_1 \geqslant h, T - K_2 \geq [T\epsilon]\}$, where h is at least as large as the number of parameters to be estimated in each regime. We do not impose the restriction that the number of observations in a single regime increases as the sample size increases, as is common when dealing with non-local break dates.

2.3.1 Theoretical Results

We again start with some preliminary results about the rate of convergence of the estimates. Throughout, we assume that δ_1^0 and δ_2^0 are non-zero and that Assumption 3 holds.

Proposition 2.2 *Under the stated conditions, the following quantities are $O_p(1)$: $\sqrt{T}(\hat{K}_1 - K_1^0)$, $\sqrt{T}(\hat{K}_2 - K_2^0)$, $\sqrt{T}(\hat{\mu}_1 - \mu_1^0)$, $\sqrt{T}(\hat{\mu}_2 - \mu_2^0)$, $T^{3/2}(\hat{\beta}_1 - \beta_1^0)$, $T^{3/2}(\hat{\beta}_2 - \beta_2^0)$, $T^{3/2}(\hat{\delta}_1 - \delta_1^0)$, $T^{3/2}(\hat{\delta}_2 - \delta_2^0)$.*

The results follow using arguments similar to those in Perron and Zhu (2005), hence details are omitted. Given the rates of convergence, we can analyze the likelihood function under the compact set

$$C_M = \{\mu_i, \beta_i, \delta_i, K_i : \sqrt{T}(\mu_i - \mu_i^0) \leq M, T^{3/2}(\beta_i - \beta_i^0) \leq M$$
$$T^{3/2}(\delta_i - \delta_i^0) \leq M, \sqrt{T}(K_i - K_i^0) \leq M\}$$

The log-likelihood function is given by:

$$\begin{aligned} lr_T &= -\frac{T}{2}[\log|\Sigma| - \log|\Sigma^0|] \\ &\quad -\frac{1}{2}[\textstyle\sum_{t=1}^T[(Y_t - X_t^{0\prime}\theta)'\Sigma^{-1}(Y_t - X_t^{0\prime}\theta) \\ &\quad - \textstyle\sum_{t=1}^T(Y_t - X_t^{0\prime}\theta^0)'\Sigma^{-1}(Y_t - X_t^{0\prime}\theta^0)] \\ &\quad -\frac{1}{2}[\textstyle\sum_{t=1}^T \theta^{0\prime}(X_t^0 - X_t)(\Sigma^0)^{-1}(X_t^0 - X_t)'\theta^0 \\ &\quad +2\textstyle\sum_{t=1}^T \theta^{0\prime}(X_t^0 - X_t)(\Sigma^0)^{-1}U_t] + o_p(1) \end{aligned}$$

Then, within the set C_M, this log-likelihood function can be separated in two parts (except for terms that converge in probability to 0), namely,

$$\begin{aligned} lr_T^1 &= -\frac{T}{2}[\log|\Sigma| - \log|\Sigma^0|] \\ &\quad -\frac{1}{2}[\textstyle\sum_{t=1}^T[(Y_t - X_t^{0\prime}\theta)'\Sigma^{-1}(Y_t - X_t^{0\prime}\theta) \\ &\quad - \textstyle\sum_{t=1}^T(Y_t - X_t^{0\prime}\theta^0)'\Sigma^{-1}(Y_t - X_t^{0\prime}\theta^0)] \end{aligned}$$

$$\begin{aligned} lr_T^2 &= -\frac{1}{2}[\textstyle\sum_{t=1}^T \theta^{0\prime}(X_t^0 - X_t)(\Sigma^0)^{-1}(X_t^0 - X_t)\theta^0 \\ &\quad +2\textstyle\sum_{t=1}^T \theta^{0\prime}(X_t^0 - X_t)(\Sigma^0)^{-1}U_t] \end{aligned}$$

so that to derive the asymptotic distribution of the estimates of the break dates, we need only focus on lr_T^2. Given the rates of convergence and the definition of locally ordered breaks, we have the following six cases for the position of the estimates of the break dates relative to the true values; Case 1: $\hat{K}_1 \leq K_1^0 \leq \hat{K}_2 \leq K_2^0$, Case 2: $\hat{K}_1 \leq K_1^0 \leq K_2^0 \leq \hat{K}_2$, Case 3: $\hat{K}_1 \leq \hat{K}_2 \leq K_1^0 \leq K_2^0$, Case 4: $K_1^0 \leq \hat{K}_1 \leq K_2^0 \leq \hat{K}_2$,Case 5: $K_1^0 \leq \hat{K}_1 \leq \hat{K}_2 \leq K_2^0$, Case 6: $K_1^0 \leq K_2^0 \leq \hat{K}_1 \leq \hat{K}_2$. It turns out, however, that the limit distribution of the estimates of the break dates is the same in all cases, as stated in the following theorem.

Theorem 2.2 *Under the stated conditions,*

$$\begin{aligned} \begin{bmatrix} T^{1/2}(\hat{K}_1 - K_1^0) \\ T^{1/2}(\hat{K}_2 - K_2^0) \end{bmatrix} &\Rightarrow \arg\max_{s_1,s_2} H(s_1, s_2) \\ &= -\frac{1}{2}\{tr(\begin{bmatrix} (1-\lambda_2^0)\delta_1^{02}s_1^2 & (1-\lambda_2^0)\delta_1^0\delta_2^0 s_1 s_2 \\ (1-\lambda_2^0)\delta_1^0\delta_2^0 s_1 s_2 & (1-\lambda_2^0)\delta_2^{02}s_2^2 \end{bmatrix}) \end{aligned}$$

$$+2tr((\Omega^0)^{1/2}$$
$$\begin{bmatrix} \delta_1^0 s_1(W_1^*(1) - W_1^*(\lambda_2^0)) & \delta_2^0 s_2(W_1^*(1) - W_1^*(\lambda_2^0)) \\ \delta_1^0 s_1(W_2^*(1) - W_2^*(\lambda_2^0)) & \delta_2^0 s_2(W_2^*(1) - W_2^*(\lambda_2^0)) \end{bmatrix})\}$$

Remark 2.3 *In the case with*

$$\Omega^0 = \begin{pmatrix} \omega_1^2 & 0 \\ 0 & \omega_2^2 \end{pmatrix}$$

we have

$$\begin{aligned} \begin{bmatrix} T^{1/2}(\hat{K}_1 - K_1^0) \\ T^{1/2}(\hat{K}_2 - K_2^0) \end{bmatrix} &\Rightarrow \arg\max_{s_1,s_2} H(s_1, s_2) \\ &= -(1/2)(1-\lambda_2^0)((\delta_1^0)^2 s_1^2 + (\delta_2^0)^2 s_2^2) \\ &\quad -((\omega_1\delta_1^0)s_1(W_1^*(1) - W_1^*(\lambda_2^0)) \\ &\quad +(\omega_1\delta_2^0)s_2(W_2^*(1) - W_2^*(\lambda_2^0))) \end{aligned}$$

The first order conditions yield,

$$\begin{aligned} -(1-\lambda_2^0)(\delta_1^0)^2 s_1 &= \omega_1\delta_1^0(W_1^*(1) - W_1^*(\lambda_2^0)) \\ -(1-\lambda_2^0)(\delta_2^0)^2 s_2 &= \omega_2\delta_2^0(W_2^*(1) - W_2^*(\lambda_2^0)) \end{aligned}$$

So that,

$$\begin{aligned} s_1 &= -\omega_1 \frac{W_1^*(1) - W_1^*(\lambda_2^0)}{(1-\lambda_2^0)\delta_1^0} \\ s_2 &= -\omega_2 \frac{W_2^*(1) - W_2^*(\lambda_2^0)}{(1-\lambda_2^0)\delta_2^0}. \end{aligned}$$

These imply that

$$\begin{bmatrix} T^{1/2}(\hat{K}_1 - K_1^0) \\ T^{1/2}(\hat{K}_2 - K_2^0) \end{bmatrix} \Rightarrow N(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{\omega_1^2}{(1-\lambda_2^0)\delta_1^{02}} & 0 \\ 0 & \frac{\omega_2^2}{(1-\lambda_2^0)\delta_2^{02}} \end{bmatrix})$$

which corresponds to the result obtained by Perron and Zhu (2005).

2.4 Simulation Result

In this section, we provide Monte Carlo simulation results to assess the adequacy of the asymptotic distribution in providing useful approximations to the finite sample distribution of the locally ordered break dates. We also we provide simulation results concerning the size and power of the common breaks test statistic. All results are based on 1,000 replications and the Wiener process are approximate by the partial sums of a sequence of *i.i.d.* $N(0,1)$ random variable of length $1,000$.

2.4.1 Simulation Result for Testing Common Breaks

We start by considering the exact size and power of the common breaks test. The Data Generating Process (DGP) is specified by:

$$\begin{aligned} y_{1t} &= \mu_1^0 + \beta_1^0 t + \delta_1^0 1(t \geqslant K_1^0)(t - K_1^0) + u_{1t} \\ y_{2t} &= \mu_2^0 + \beta_2^0 t + \delta_2^0 1(t \geqslant K_2^0)(t - K_2^0) + u_{2t} \end{aligned}$$

The sample size is set to $T = 100$, and we use the following parameter values

$$\begin{aligned} \mu_1^0 &= 1, \beta_1^0 = 0.5, \delta_1^0 = 0.2 \\ \mu_2^0 &= 0.8, \beta_2^0 = 0.5, \delta_2^0 = 0.3 \end{aligned}$$

with

$$U_t = (u_1, u_{2t})' \sim i.i.d.\ N(0, I_2)$$

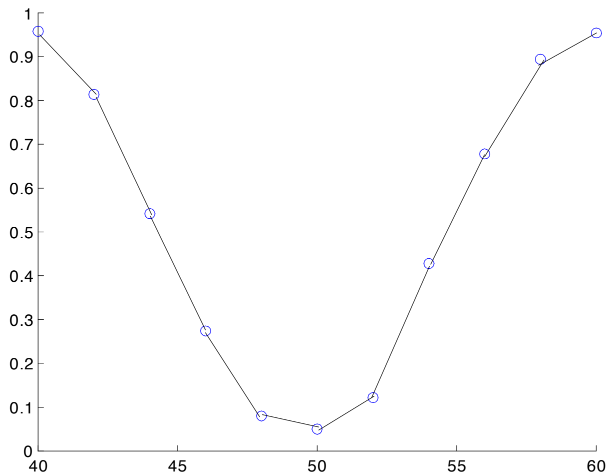
Under the null hypothesis, there is one common break at mid-sample, i.e., $K_1^0 = K_2^0 = 50$. To obtain the quantiles of the asymptotic distribution. We impose $\Omega^0 = I$ and, hence, the limit distribution depends only on δ_1^0, δ_2^0 and λ^0. In order to better assess the reliability of the asymptotic distribution we use the true values. Hence, this permits us to avoid finite sample effects due to the estimation of the parameters. The nominal and exact quantiles of the asymptotic distribution of the test are presented below. The results show a close correspondence so that the test has no size distortions.

Table 2.1: Size Comparision of Test Statistics

Asymptotic	95%	90%	85%	80%	75%	70%
Exact	96%	89.9%	83%	77.4%	73.7%	68.9%

To assess the power of the test, we set $K_1^0 = 50$ and vary the date of the break in the second equation with the following values: $K_2^0 = 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60$. We consider a test with size 5%. The power of the test is shown in the following figure. The results show that power increases rapidly as the distance between the break dates increases, even with a relatively small sample size.

Figure 2.1: Power of Test Statistics



2.4.2 Simulation Result for Locally Ordered Breaks

We now consider the adequacy of the limit distribution for the estimates of locally ordered break dates. The DGP used is:

$$\begin{aligned} y_{1t} &= \mu_1^0 + \beta_1^0 t + \delta_1^0 1(t \geqslant K_1^0)(t - K_1^0) + u_{1t} \\ y_{2t} &= \mu_2^0 + \beta_2^0 t + \delta_2^0 1(t \geqslant K_2^0)(t - K_2^0) + u_{2t} \end{aligned}$$

The sample size is set to $T = 100$, and we use the following parameter values:

$$\begin{aligned} \mu_1^0 &= 1, \beta_1^0 = 0.5, \delta_1^0 = 0.2 \\ \mu_2^0 &= 1, \beta_2^0 = 0.5, \delta_2^0 = 0.4 \\ K_1^0 &= 45, K_2^0 = 50 \end{aligned}$$

with

$$U_t = (u_1, u_{2t})' \sim i.i.d.\ N(0, \Sigma)$$

where

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

We compare the nominal size and real size, with different values of ρ . The result is as follows:

Table 2.2: Size Comparision

Nominal Size	$\rho = 0.9$	$\rho = 0.85$	$\rho = 0.75$
95%	87.9%	83.9%	80.6%
75%	63.7%	61.8%	59.3%
50%	63.7%	33%	36.1%

Though, the experiments are quite limited they indicate that the limit distributions derived can provide tests with reliable size in finite samples and decent power. Also, the limit distribution of the estimates of the locally ordered breaks provide a good approximation to the finite sample distribution. This shows the usefulness of the results derived.

2.5 Conclusions

In this paper, we considered the issues of testing for common breaks in a bivariate system described by a linear trend function with changes in slope such that the series are joined at the time of the breaks. We also considered the problem of performing joint inference about the estimates of the break dates in the same setup when the break dates are locally ordered. We provided the rate of convergence, the asymptotic distribution for the estimates of the locally ordered breaks and the asymptotic distribution of the common break test statistic. Limited simulation results showed that the theoretical results derived deliver good approximations in finite samples.

2.6 Appendix 2

Proof of Theorem 2.1 (asymptotic distribution of the common break test statistic). For $t \in [T^0_{j-1}+1, T^0_j]$, the regressor matrix in regime j is:

$$X^0_{tj} = \begin{bmatrix} 1 & t & t-K^0_{11} & \cdots & t-K^0_{1j_1-1} & 0 & \cdots & 0 & & & & & & & & 0 \\ & & & & & & & 1 & t & t-K^0_{21} & \cdots & t-K^0_{2j_2-1} & 0 & \cdots & 0 \end{bmatrix}$$

so that the data for the first equation is in its j_1-th regime and that for the second equation is in its j_2-th regime. Under H_0, for the jth regime with $t \in [T_{j-1}+1, T_j]$:

$$\tilde{X}_{tj} = \begin{bmatrix} 1 & t & t-\tilde{K}_{11} & \cdots & t-\tilde{K}_{1j_1-1} & 0 & \cdots & 0 & & & & & & & & & 0 \\ & & & & & & & 1 & t & t-\tilde{K}_{21} & \cdots & t-\tilde{K}_{2j_2-1} & 0 & \cdots & 0 \end{bmatrix}$$

so that

$$X_{tj}^0 - \tilde{X}_{tj} =$$

$$\begin{bmatrix} 0 & \begin{matrix}\tilde{K}_{11}^0 \\ -K_{11}^0\end{matrix} & \cdots & \begin{matrix}\tilde{K}_{1j_1-1} \\ -K_{1j_1-1}^0\end{matrix} & 0 & \cdots & & & & & & 0 \\ & & & & & & 0 & \begin{matrix}\tilde{K}_{21} \\ -K_{21}^0\end{matrix} & \cdots & \begin{matrix}\tilde{K}_{2j_2-1}^0 \\ -K_{2j_2-1}^0\end{matrix} & 0 & \cdots & 0 \end{bmatrix}$$

For simplicity, assume that $T_j < T_j^0$, then for $t \in [T_j+1, T_j^0]$, depending on which equation has a change at T_j^0, the form of $\tilde{X}_{tj+1}$ could be either: a) with the first equation having a break at T_j^0, so that:

$$\tilde{X}_{tj+1} = \begin{bmatrix} 1 & t & \begin{matrix} t- \\ \tilde{K}_{11}\end{matrix} & \cdots & \begin{matrix} t- \\ \tilde{K}_{1j_1-1}\end{matrix} & \begin{matrix} t- \\ \tilde{K}_{1j_1}\end{matrix} & \cdots & 0 & & & & & & & 0 \\ & & & & & & & 1 & t & \begin{matrix} t- \\ \tilde{K}_{21}\end{matrix} & \cdots & \begin{matrix} t- \\ \tilde{K}_{2j_2-1}\end{matrix} & 0 & \cdots & 0 \end{bmatrix}$$

b) with the second equation having a break at T_j^0, in which case

$$\tilde{X}_{tj+1} = \begin{bmatrix} 1 & t & t-\tilde{K}_{11} & \cdots & t-\tilde{K}_{1j_1-1} & 0 & \cdots & 0 & & & & & & & 0 \\ & & & & & & & 1 & t & \begin{matrix} t- \\ \tilde{K}_{21}\end{matrix} & \cdots & \begin{matrix} t- \\ \tilde{K}_{2j_2-1}\end{matrix} & \begin{matrix} t- \\ \tilde{K}_{2j_2}\end{matrix} & \cdots & 0 \end{bmatrix}$$

c) with both equations having a break at T_j^0, in which case

$$\tilde{X}_{tj+1} = \begin{bmatrix} 1 & t & t-\tilde{K}_{11} & \cdots & t-\tilde{K}_{1j_1} & \cdots & 0 & 0 & & & & & & 0 \\ & & & & & & & 1 & t & t-\tilde{K}_{21} & \cdots & t-\tilde{K}_{2j_2} & \cdots & 0 \end{bmatrix}$$

We first have

$$
\begin{aligned}
& \textstyle\sum_{t=1}^{T} \theta^{0\prime}(X_t^0 - \tilde{X}_t)(\Sigma^0)^{-1}U_t \\
= & \textstyle\sum_{j=1}^{m+1} \sum_{t=T_{j-1}^0+1}^{T_j^0} \theta^{0\prime}(X_{tj}^0 - \tilde{X}_{tj})(\Sigma^0)^{-1}U_t \\
& \textstyle+\sum_{j=1}^{m} \sum_{t=T_j+1}^{T_j^0} \theta^{0\prime}(X_{tj}^0 - \tilde{X}_{tj+1})(\Sigma^0)^{-1}U_t \\
& \textstyle-\sum_{j=1}^{m} \sum_{t=T_j+1}^{T_j^0} \theta^{0\prime}(X_{tj}^0 - \tilde{X}_{tj})(\Sigma^0)^{-1}U_t \\
= & \textstyle\sum_{j=1}^{m+1} \sum_{t=T_{j-1}^0+1}^{T_j^0} \theta^{0\prime}(X_{tj}^0 - \tilde{X}_{tj})(\Sigma^0)^{-1}U_t + o_p(1) \\
= & \textstyle\sum_{j=1}^{m+1} \sum_{t=T_{j-1}^0+1}^{T_j^0} [\sum_{i_1=1}^{j_1-1} \delta_{1i_1}^0 (K_{1i_1} - K_{1i_1}^0) + \\
& \textstyle\sum_{i_2=1}^{j_2-1} \delta_{1i_2}^0 (K_{1i_2} - K_{1i_2}^0) \;\;](\Sigma^0)^{-1}U_t \\
= & \textstyle\sum_{j=1}^{m+1} tr(\sum_{t=T_{j-1}^0+1}^{T_j^0} (\Sigma^0)^{-1/2}\eta_t \\
& \textstyle[\sum_{i_1=1}^{j_1-1} \delta_{1i_1}^0 (K_{1i_1} - K_{1i_1}^0) \;\;\; \sum_{i_2=1}^{j_2-1} \delta_{1i_2}^0 (K_{1i_2} - K_{1i_2}^0) \;])
\end{aligned}
$$

$$
\begin{aligned}
= & \sum_{j=1}^{m+1} tr((\Sigma^0)^{-1/2} \textstyle\sum_{t=T_{j-1}^0+1}^{T_j^0} \eta_t \\
& \textstyle[\sum_{i_1=1}^{j_1-1} \delta_{1i_1}^0 (K_{1i_1} - K_{1i_1}^0) \;\;\; \sum_{i_2=1}^{j_2-1} \delta_{1i_2}^0 (K_{1i_2} - K_{1i_2}^0) \;]) \\
= & \sum_{j=1}^{m+1} tr((\Sigma^0)^{-1/2} \\
& \begin{bmatrix} \sum\limits_{t=T_{j-1}^0+1}^{T_j^0} \eta_{1t}(\sum\limits_{i_1=1}^{j_1-1} \delta_{1i_1}^0 (K_{1i_1} - K_{1i_1}^0)\,) & \sum\limits_{t=T_{j-1}^0+1}^{T_j^0} \eta_{1t}(\sum\limits_{i_2=1}^{j_2-1} \delta_{2i_2}^0 (K_{2i_2} - K_{2i_2}^0)\,) \\ \sum\limits_{t=T_{j-1}^0+1}^{T_j^0} \eta_{2t}(\sum\limits_{i_1=1}^{j_1-1} \delta_{1i_1}^0 (K_{1i_1} - K_{1i_1}^0)\,) & \sum\limits_{t=T_{j-1}^0+1}^{T_j^0} \eta_{2t}(\sum\limits_{i_2=1}^{j_2-1} \delta_{2i_2}^0 (K_{2i_2} - K_{2i_2}^0)\,) \end{bmatrix} \\
\Rightarrow & \sum_{j=1}^{m+1} tr((\Sigma^0)^{-1/2} \\
& \begin{bmatrix} (\sum\limits_{i_1=1}^{j_1-1} \delta_{1i_1}^0 s_{1i_1})(W_1(\lambda_j^0) - W_1(\lambda_{j-1}^0)) & (\sum\limits_{i_2=1}^{j_2-1} \delta_{2i_2}^0 s_{2i_2})(W_1(\lambda_j^0) - W_1(\lambda_{j-1}^0)) \\ (\sum\limits_{i_1=1}^{j_1-1} \delta_{1i_1}^0 s_{1i_1})(W_2(\lambda_j^0) - W_2(\lambda_{j-1}^0)) & (\sum\limits_{i_2=1}^{j_2-1} \delta_{2i_2}^0 s_{2i_2})(W_2(\lambda_j^0) - W_2(\lambda_{j-1}^0)) \end{bmatrix} \\
\equiv & \sum_{j=1}^{m+1} tr((\Sigma^0)^{-1/2}\tilde{B})
\end{aligned}
$$

with

$$\tilde{B} = \begin{bmatrix} (\sum_{i_1=1}^{j_1-1}\delta^0_{1i_1}s_{1i_1})(W_1(\lambda^0_j)-W_1(\lambda^0_{j-1})) & (\sum_{i_2=1}^{j_2-1}\delta^0_{2i_2}s_{2i_2})(W_1(\lambda^0_j)-W_1(\lambda^0_{j-1})) \\ (\sum_{i_1=1}^{j_1-1}\delta^0_{1i_1}s_{1i_1})(W_2(\lambda^0_j)-W_2(\lambda^0_{j-1})) & (\sum_{i_2=1}^{j_2-1}\delta^0_{2i_2}s_{2i_2})(W_2(\lambda^0_j)-W_2(\lambda^0_{j-1})) \end{bmatrix}$$

Note that in the derivations above and below, we use the fact that on the compact sets $C_M^{(1)}$ and $C_M^{(2)}$, we have $s_{1i_1} = \sqrt{T}(K_{1i_1}-K^0_{1i_1})$ and $s_{2i_2} = \sqrt{T}(K_{2i_2}-K^0_{2i_2})$ are bounded in probability. In what follows, we provide details for case (a) for which the first equation has a break at T_j^0 and indicate the changes needed for the two other cases. Now,

$$\begin{aligned} & \textstyle\sum_{t=1}^{T}\theta^{0\prime}(X_t^0-\tilde{X}_t)(\Sigma^0)^{-1}(X_t^0-\tilde{X}_t)'\theta^0 \\ = & \textstyle\sum_{j=1}^{m+1}\sum_{t=T^0_{j-1}+1}^{T^0_j}\theta^{0\prime}(X^0_{tj}-\tilde{X}_{tj})(\Sigma^0)^{-1}(X^0_{tj}-\tilde{X}_{tj})'\theta^0 \\ & -\textstyle\sum_{j=1}^{m}\sum_{t=T_j+1}^{T^0_j}\theta^{0\prime}(X^0_{tj}-\tilde{X}_{tj})(\Sigma^0)^{-1}(X^0_{tj}-\tilde{X}_{tj})'\theta^0 \\ & +\textstyle\sum_{j=1}^{m}\sum_{t=T_j+1}^{T^0_j}\theta^{0\prime}(X^0_{tj}-\tilde{X}_{tj+1})(\Sigma^0)^{-1}(X^0_{tj}-\tilde{X}_{tj+1})'\theta^0 \end{aligned}$$

$$\begin{aligned} & \textstyle\sum_{t=1}^{T}\theta^{0\prime}(X_t^0-\tilde{X}_t)(\Sigma^0)^{-1}(X_t^0-\tilde{X}_t)'\theta^0 \\ = & \textstyle\sum_{j=1}^{m+1}\sum_{t=T^0_{j-1}+1}^{T^0_j}\theta^{0\prime}(X^0_{tj}-\tilde{X}_{tj})(\Sigma^0)^{-1}(X^0_{tj}-\tilde{X}_{tj})'\theta^0 \\ & -\textstyle\sum_{j=1}^{m}\sum_{t=T_j+1}^{T^0_j}\theta^{0\prime}(X^0_{tj}-\tilde{X}_{tj})(\Sigma^0)^{-1}(X^0_{tj}-\tilde{X}_{tj})'\theta^0 \\ & +\textstyle\sum_{j=1}^{m}\sum_{t=T_j+1}^{T^0_j}\theta^{0\prime}(X^0_{tj}-\tilde{X}_{tj+1})(\Sigma^0)^{-1}(X^0_{tj}-\tilde{X}_{tj+1})'\theta^0 \\ = & \textstyle\sum_{j=1}^{m+1}\sum_{t=T^0_{j-1}+1}^{T^0_j}[\sum_{i_1=1}^{j_1-1}\delta^0_{1i_1}(K_{1i_1}-K^0_{1i_1}) \quad \sum_{i_2=1}^{j_2-1}\delta^0_{1i_2}(K_{1i_2}-K^0_{1i_2})\ ] \\ & \times(\Sigma^0)^{-1}\begin{bmatrix}\sum_{i_1=1}^{j_1-1}\delta^0_{1i_1}(K_{1i_1}-K^0_{1i_1}) \\ \sum_{i_2=1}^{j_2-1}\delta^0_{1i_2}(K_{1i_2}-K^0_{1i_2})\end{bmatrix} \\ & +\textstyle\sum_{j=1}^{m}\sum_{t=T_j+1}^{T^0_j}[\displaystyle\sum_{i_1=1}^{j_1-1}\delta^0_{1i_1}(K_{1i_1}-K^0_{1i_1})+\delta^0_{1j_1}(t-K_{1j_1}) \ \sum_{i_2=1}^{j_2-1}\delta^0_{1i_2}(K_{1i_2}-K^0_{1i_2})\ ] \\ & \times(\Sigma^0)^{-1}\begin{bmatrix}\displaystyle\sum_{i_1=1}^{j_1-1}\delta^0_{1i_1}(K_{1i_1}-K^0_{1i_1})+\delta^0_{1j_1}(t-K_{1j_1}) \\ \displaystyle\sum_{i_2=1}^{j_2-1}\delta^0_{1i_2}(K_{1i_2}-K^0_{1i_2})\end{bmatrix} \end{aligned} \tag{2.1}$$

$$= \sum_{j=1}^{m+1} \sum_{t=T_{j-1}^0+1}^{T_j^0} [\sum_{i_1=1}^{j_1-1} \delta_{1i_1}^0 (K_{1i_1} - K_{1i_1}^0) \ \sum_{i_2=1}^{j_2-1} \delta_{1i_2}^0 (K_{1i_2} - K_{1i_2}^0) \ ]$$

$$(\Sigma^0)^{-1} \begin{bmatrix} \sum_{i_1=1}^{j_1-1} \delta_{1i_1}^0 (K_{1i_1} - K_{1i_1}^0) \\ \sum_{i_2=1}^{j_2-1} \delta_{1i_2}^0 (K_{1i_2} - K_{1i_2}^0) \end{bmatrix} + o_p(1)$$

$$= \sum_{j=1}^{m+1} (\lambda_j^0 - \lambda_{j-1}^0) tr((\Sigma^0)^{-1}$$

$$\begin{bmatrix} (\sum_{i_1=1}^{j_1-1} \delta_{1i_1}^0 s_{1i_1})^2 & (\sum_{i_1=1}^{j_1-1} \delta_{1i_1}^0 s_{1i_1})(\sum_{i_2=1}^{j_2-1} \delta_{2i_2}^0 s_{2i_2}) \\ (\sum_{i_1=1}^{j_1-1} \delta_{1i_1}^0 s_{1i_1})(\sum_{i_2=1}^{j_2-1} \delta_{2i_2}^0 s_{2i_2}) & (\sum_{i_2=1}^{j_2-1} \delta_{2i_2}^0 s_{2i_2})^2 \end{bmatrix})$$

$$= \sum_{j=1}^{m+1} (\lambda_j^0 - \lambda_{j-1}^0) tr((\Sigma^0)^{-1} \tilde{A})$$

with

$$\tilde{A} = \begin{bmatrix} (\sum_{i_1=1}^{j_1-1} \delta_{1i_1}^0 s_{1i_1})^2 & (\sum_{i_1=1}^{j_1-1} \delta_{1i_1}^0 s_{1i_1})(\sum_{i_2=1}^{j_2-1} \delta_{2i_2}^0 s_{2i_2}) \\ (\sum_{i_1=1}^{j_1-1} \delta_{1i_1}^0 s_{1i_1})(\sum_{i_2=1}^{j_2-1} \delta_{2i_2}^0 s_{2i_2}) & (\sum_{i_2=1}^{j_2-1} \delta_{2i_2}^0 s_{2i_2})^2 \end{bmatrix}$$

Remark 2.4 *For case (b), with the second equation having a break at T_j^0, the term (2.1) is replaced by*

$$\sum_{j=1}^{m} \sum_{t=T_j+1}^{T_j^0} [\sum_{i_1=1}^{j_1-1} \delta_{1i_1}^0 (K_{1i_1} - K_{1i_1}^0) + \delta_{1j_1}^0 (t - K_{1j_1}) \ \sum_{i_2=1}^{j_2-1} \delta_{1i_2}^0 (K_{1i_2} - K_{1i_2}^0) + \delta_{2j_1}^0 (t - K_{2j_1})]$$
$$\times (\Sigma^0)^{-1} \begin{bmatrix} \sum_{i_1=1}^{j_1-1} \delta_{1i_1}^0 (K_{1i_1} - K_{1i_1}^0) + \delta_{1j_1}^0 (t - K_{1j_1}) \\ \sum_{i_2=1}^{j_2-1} \delta_{1i_2}^0 (K_{1i_2} - K_{1i_2}^0) + \delta_{2j_1}^0 (t - K_{2j_1}) \end{bmatrix}$$

and for case (c) for which both equations have a break at T_j^0, it is replaced by

$$\sum_{j=1}^{m} \sum_{t=T_j+1}^{T_j^0} [\sum_{i_1=1}^{j_1-1} \delta_{1i_1}^0 (K_{1i_1} - K_{1i_1}^0) \ \sum_{i_2=1}^{j_2-1} \delta_{1i_2}^0 (K_{1i_2} - K_{1i_2}^0) + \delta_{2j_1}^0 (t - K_{2j_1})]$$
$$(\Sigma^0)^{-1} \begin{bmatrix} \sum_{i_1=1}^{j_1-1} \delta_{1i_1}^0 (K_{1i_1} - K_{1i_1}^0) \\ \sum_{i_2=1}^{j_2-1} \delta_{1i_2}^0 (K_{1i_2} - K_{1i_2}^0) + \delta_{2j_1}^0 (t - K_{2j_1}) \end{bmatrix}$$
$$- \sum_{j=1}^{m} \sum_{t=T_j+1}^{T_j^0} [\sum_{i_1=1}^{j_1-1} \delta_{1i_1}^0 (K_{1i_1} - K_{1i_1}^0) \ \sum_{i_2=1}^{j_2-1} \delta_{1i_2}^0 (K_{1i_2} - K_{1i_2}^0) \ ]$$
$$\times (\Sigma^0)^{-1} \begin{bmatrix} \sum_{i_1=1}^{j_1-1} \delta_{1i_1}^0 (K_{1i_1} - K_{1i_1}^0) \\ \sum_{i_2=1}^{j_2-1} \delta_{1i_2}^0 (K_{1i_2} - K_{1i_2}^0) \end{bmatrix}$$

Hence,

$$\widetilde{lr}_T^{\,2} \Rightarrow \max_{s_1,s_2 \text{ with } s_{1l_1}=s_{2l_2}} -(1/2)[2\textstyle\sum_{j=1}^{m+1} tr((\Sigma^0)^{-1/2}\tilde{B}) + \sum_{j=1}^{m+1}(\lambda_j^0 - \lambda_{j-1}^0) tr((\Sigma^0)^{-1/2}\tilde{A})]$$

Under H_1, we have similar results as under H_0 for regimes other than the lth one. However, for the lth regime things are different given that we estimate K_{1l_1} and K_{2l_2}. Now suppose we have the case $T_{l-1}^0 < T_{1l} < T_{2l} < T_l^0$ (note the subscripts 1 and 2 for T_{1l} and T_{2l} are not related to the number of equation, but rather to their order of occurrence). Then a) for $t \in [T_{l-1}^0 + 1, T_l^0]$:

$$X_{tl}^0 = \begin{bmatrix} 1 & t & t-K_{11}^0 & \cdots & t-K_{1l_1-1}^0 & 0 & \cdots & 0 & & & & & & & & 0 \\ & & & & & & & 1 & t & t-K_{21}^0 & \cdots & t-K_{21_2-1}^0 & 0 & \cdots & & 0 \end{bmatrix}$$

b) for $t \in [T_{l-1}^0 + 1, T_{1l}]$:

$$\hat{X}_{t1l} = \begin{bmatrix} 1 & t & t-K_{11}^0 & \cdots & t-\hat{K}_{1l_1-1} & 0 & \cdots & 0 & & & & & & & 0 \\ & & & & & & & 1 & t & t-K_{21}^0 & \cdots & t-\hat{K}_{21_2-1} & 0 & \cdots & 0 \end{bmatrix}$$

c) for $t \in [T_{1l} + 1, T_{2l}]$, if the first break is in the first equation

$$\hat{X}_{t1l} = \begin{bmatrix} 1 & t & \begin{matrix} t- \\ K_{11}^0 \end{matrix} & \cdots & \begin{matrix} t- \\ \hat{K}_{1l_1-1} \end{matrix} & \begin{matrix} t- \\ \hat{K}_{1l_1} \end{matrix} & \cdots & 0 & & & & & & & 0 \\ & & & & & & & 1 & t & \begin{matrix} t- \\ K_{21}^0 \end{matrix} & \cdots & \begin{matrix} t- \\ \hat{K}_{21_2-1} \end{matrix} & 0 & \cdots & 0 \end{bmatrix}$$

while if the first break is in the second equation

$$\hat{X}_{t2l} = \begin{bmatrix} 1 & t & t-K_{11}^0 & \cdots & t-\hat{K}_{1l_1-1} & \cdots & 0 & & & & & & & 0 \\ & & & & & & 1 & t & t-K_{21}^0 & \cdots & t-\hat{K}_{21_2-1} & t-\hat{K}_{21_2} & \cdots & 0 \end{bmatrix}$$

d) for $t \in [T_{2l} + 1, T_l^0]$:

$$\hat{X}_{tl+1} =$$

$$\begin{bmatrix} 1 & t & t-K_{11}^0 & \cdots & t-\hat{K}_{1l_1-1} & t-\hat{K}_{1l_1} & \cdots & 0 & & & & & & & 0 \\ & & & & & & & 1 & t & t-K_{21}^0 & \cdots & t-\hat{K}_{21_2-1} & t-\hat{K}_{21_2} & \cdots & 0 \end{bmatrix}$$

Using these results, we have

$$
\begin{aligned}
& \textstyle\sum_{t=1}^{T} \theta^{0\prime}(X_t^0 - \hat{X}_t)(\Sigma^0)^{-1}U_t \\
= \; & \textstyle\sum_{j=1}^{l} \sum_{t=T_{j-1}^0+1}^{T_j^0} [\sum_{i_1=1}^{l_1-1} \delta_{1i_1}^0 (K_{1i_1} - K_{1i_1}^0) \\
& \textstyle\sum_{i_2=1}^{l_2-1} \delta_{2i_2}^0 (K_{2i_2} - K_{2i_2}^0)](\Sigma^0)^{-1}U_t \\
& \textstyle + \sum_{j=l+1}^{m+1} \sum_{t=T_{j-1}^0+1}^{T_j^0} [\sum_{i_1=1}^{l_1-1} \delta_{1i_1}^0 (K_{1i_1} - K_{1i_1}^0) + \delta_{1l_1}^0 (K_{1l_1} - K_{1l_1}^0) \\
& \textstyle + \sum_{i_1=l_1+1}^{j_1-1} \delta_{1i_1}^0 (K_{1i_1} - K_{1i_1}^0) \\
& \textstyle + \sum_{i_2=1}^{l_2-1} \delta_{2i_2}^0 (K_{2i_2} - K_{2i_2}^0) + \delta_{2l_2}^0 (K_{2l_2} - K_{2l_2}^0) \\
& \textstyle + \sum_{i_2=l_2+1}^{j_2-1} \delta_{2i_2}^0 (K_{2i_2} - K_{2i_2}^0)](\Sigma^0)^{-1}U_t \\
\Rightarrow \; & \textstyle\sum_{j=1}^{l} tr((\Sigma^0)^{-1/2}\hat{A}_1) + \sum_{j=l+1}^{m+1} tr((\Sigma^0)^{-1/2}\hat{A}_2)
\end{aligned}
$$

with

$$
\hat{A}_1 =
\begin{bmatrix}
(\sum_{i_1=1}^{l_1-1} \delta_{1i_1}^0 s_{1i_1})(W_1(\lambda_j^0) - W_1(\lambda_{j-1}^0)) & \begin{array}{c}(\sum_{i_2=1}^{l_2-1} \delta_{2i_2}^0 s_{2i_2})* \\ (W_1(\lambda_j^0) - W_1(\lambda_{j-1}^0))\end{array} \\
(\sum_{i_1=1}^{l_1-1} \delta_{1i_1}^0 s_{1i_1})(W_2(\lambda_j^0) - W_2(\lambda_{j-1}^0)) & \begin{array}{c}(\sum_{i_2=1}^{l_2-1} \delta_{2i_2}^0 s_{2i_2})* \\ (W_1(\lambda_j^0) - W_1(\lambda_{j-1}^0))\end{array}
\end{bmatrix}
$$

and the elements of $\hat{A}_2$ are given by:

$$
\begin{aligned}
\hat{A}_2(1,1) &= (\textstyle\sum_{i_1=1}^{l_1-1} \delta_{1i_1}^0 s_{1i_1} + \delta_{1l_1}^0 s_{1l_1} + \sum_{i_1=l_1+1}^{j_1-1} \delta_{1i_1}^0 s_{1i_1}) \\
&\quad (W_1(\lambda_j^0) - W_1(\lambda_{j-1}^0)) \\
\hat{A}_2(1,2) &= (\textstyle\sum_{i_2=1}^{l_2-1} \delta_{2i_2}^0 s_{2i_2} + \delta_{2l_2}^0 s_{2l_2} + \sum_{i_2=l_2+1}^{j_2-1} \delta_{2i_2}^0 s_{2i_2}) \\
&\quad (W_1(\lambda_j^0) - W_1(\lambda_{j-1}^0)) \\
\hat{A}_2(2,1) &= (\textstyle\sum_{i_1=1}^{l_1-1} \delta_{1i_1}^0 s_{1i_1} + \delta_{1l_1}^0 s_{1l_1} + \sum_{i_1=l_1+1}^{j_1-1} \delta_{1i_1}^0 s_{1i_1}) \\
&\quad (W_2(\lambda_j^0) - W_2(\lambda_{j-1}^0)) \\
\hat{A}_2(2,2) &= (\textstyle\sum_{i_2=1}^{l_2-1} \delta_{2i_2}^0 s_{2i_2} + \delta_{2l_2}^0 s_{2l_2} + \sum_{i_2=l_2+1}^{j_2-1} \delta_{2i_2}^0 s_{2i_2}) \\
&\quad (W_2(\lambda_j^0) - W_2(\lambda_{j-1}^0))
\end{aligned}
$$

Using similar arguments, we have:

$$
\begin{aligned}
& \textstyle\sum_{t=1}^{T} \theta^{0\prime}(X_t^0 - \hat{X}_t)(\Sigma^0)^{-1}(X_t^0 - \hat{X}_t)'\theta^0 \\
= & \textstyle\sum_{j=1}^{l} \sum_{t=T_{j-1}^0+1}^{T_j^0} [\sum_{i_1=1}^{l_1-1} \delta_{1i_1}^0 (K_{1i_1} - K_{1i_1}^0)\ \ \sum_{i_2=1}^{l_2-1} \delta_{2i_2}^0 (K_{2i_2} - K_{2i_2}^0)] \\
& \times(\Sigma^0)^{-1} \begin{bmatrix} \sum_{i_1=1}^{l_1-1} \delta_{1i_1}^0 (K_{1i_1} - K_{1i_1}^0) \\ \sum_{i_2=1}^{l_2-1} \delta_{2i_2}^0 (K_{2i_2} - K_{2i_2}^0) \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
& + \textstyle\sum_{j=l+1}^{m+1} \sum_{t=T_{j-1}^0+1}^{T_j^0} [\sum_{i_1=1}^{l_1-1} \delta_{1i_1}^0 (K_{1i_1} - K_{1i_1}^0) + \delta_{1l_1}^0 (K_{1l_1} - K_{1l_1}^0) \\
& + \textstyle\sum_{i_1=l_1+1}^{j_1-1} \delta_{1i_1}^0 (K_{1i_1} - K_{1i_1}^0)\ + \sum_{i_2=1}^{l_2-1} \delta_{2i_2}^0 (K_{2i_2} - K_{2i_2}^0) \\
& + \delta_{2l_2}^0 (K_{2l_2} - K_{2l_2}^0) + \textstyle\sum_{i_2=l_2+1}^{j_2-1} \delta_{2i_2}^0 (K_{2i_2} - K_{2i_2}^0)] \\
& \times(\Sigma^0)^{-1}
\end{aligned}
$$

$$
\begin{aligned}
& \begin{bmatrix} \sum_{i_1=1}^{l_1-1} \delta_{1i_1}^0 (K_{1i_1} - K_{1i_1}^0) + \delta_{1l_1}^0 (K_{1l_1} - K_{1l_1}^0) + \sum_{i_1=l_1+1}^{j_1-1} \delta_{1i_1}^0 (K_{1i_1} - K_{1i_1}^0) \\ \sum_{i_2=1}^{l_2-1} \delta_{2i_2}^0 (K_{2i_2} - K_{2i_2}^0) + \delta_{2l_2}^0 (K_{2l_2} - K_{2l_2}^0) + \sum_{i_2=l_2+1}^{j_2-1} \delta_{2i_2}^0 (K_{2i_2} - K_{2i_2}^0) \end{bmatrix} \\
\Rightarrow\ & \textstyle\sum_{j=1}^{l} tr((\Sigma^0)^{-1}\hat{B}_1) + \sum_{j=l+1}^{m+1} tr((\Sigma^0)^{-1}\hat{B}_2)
\end{aligned}
$$

where

$$
\hat{B}_1 = \begin{bmatrix} (\lambda_j^0 - \lambda_{j-1}^0)(\sum_{i_1=1}^{l_1-1} \delta_{1i_1}^0 s_{1i_1})^2 & \begin{matrix} (\lambda_j^0 - \lambda_{j-1}^0)* \\ (\sum_{i_1=1}^{l_1-1} \delta_{1i_1}^0 s_{1i_1})(\sum_{i_2=1}^{l_2-1} \delta_{2i_2}^0 s_{2i_2}) \end{matrix} \\ (\lambda_j^0 - \lambda_{j-1}^0)(\sum_{i_1=1}^{l_1-1} \delta_{1i_1}^0 s_{1i_1})(\sum_{i_2=1}^{l_2-1} \delta_{2i_2}^0 s_{2i_2}) & (\lambda_j^0 - \lambda_{j-1}^0)(\sum_{i_2=1}^{l_2-1} \delta_{2i_2}^0 s_{2i_2})^2 \end{bmatrix}
$$

and the elements of $\hat{B}_2$ are:

$$
\begin{aligned}
\hat{B}_2(1,1) &= (\lambda_j^0 - \lambda_{j-1}^0)(\textstyle\sum_{i_1=1}^{l_1-1} \delta_{1i_1}^0 s_{1i_1} + \delta_{1l_1}^0 s_{1l_1} + \sum_{i_1=l_1+1}^{j_1-1} \delta_{1i_1}^0 s_{1i_1})^2 \\
\hat{B}_2(2,2) &= (\lambda_j^0 - \lambda_{j-1}^0)(\textstyle\sum_{i_2=1}^{l_2-1} \delta_{2i_2}^0 s_{2i_2} + \delta_{2l_2}^0 s_{2l_2} + \sum_{i_2=l_2+1}^{j_2-1} \delta_{2i_2}^0 s_{2i_2})^2
\end{aligned}
$$

$$\begin{aligned}\hat{B}_2(1,2) &= \widehat{B}_2(2,1)\\ &= (\lambda_j^0 - \lambda_{j-1}^0)(\textstyle\sum_{i_1=1}^{l_1-1}\delta_{1i_1}^0 s_{1i_1} + \delta_{1l_1}^0 s_{1l_1} + \sum_{i_1=l_1+1}^{j_1-1}\delta_{1i_1}^0 s_{1i_1})\\ &\quad \times(\textstyle\sum_{i_2=1}^{l_2-1}\delta_{1i_2}^0 s_{1i_2} + \delta_{2l_2}^0 s_{2l_2} + \sum_{i_2=l_2+1}^{j_2-1}\delta_{2i_2}^0 s_{2i_2})\end{aligned}$$

Collecting results:

$$\begin{aligned}\widehat{lr}_T^2 &\Longrightarrow \max_{s_1,s_2} tr(2(\textstyle\sum_{j=1}^{l} tr((\Sigma^0)^{-1/2}\hat{A}_1) + \sum_{j=l+1}^{m+1} tr((\Sigma^0)^{-1/2}\hat{A}_2))\\ &\quad + \textstyle\sum_{j=1}^{l} tr((\Sigma^0)^{-1}\hat{B}_1) + \sum_{j=l+1}^{m+1} tr((\Sigma^0)^{-1}\hat{B}_2))\end{aligned}$$

The result stated in Theorem 1 follows using the fact that $W(\cdot) = (\Sigma^0)^{-1/2}\Omega^{1/2}W^*(\cdot)$ and the fact that the term Σ^0 is common to all components, hence can be dropped.

Proof of Theorem 2.2 (asymptotic distribution of the estimates of locally ordered breaks): The limit distribution of the estimates can be obtained from the maximization of

$$\begin{aligned}lr_T^2 &= -(1/2)[\textstyle\sum_{t=1}^{T}\theta^{0\prime}(X_t^0 - X_t)(\Sigma^0)^{-1}(X_t^0 - X_t)\theta^0\\ &\quad +2\textstyle\sum_{t=1}^{T}\theta^{0\prime}(X_t^0 - X_t)(\Sigma^0)^{-1}U_t]\end{aligned}$$

Given the rates of convergence and the definition of locally ordered breaks, we have 6 possible cases for the position of the candidate values for the estimates of the break dates relative to the true values; Case 1: $K_1 \le K_1^0 \le K_2 \le K_2^0$, Case 2: $K_1 \le K_1^0 \le K_2^0 \le K_2$, Case 3: $K_1 \le K_2 \le K_1^0 \le K_2^0$, Case 4: $K_1^0 \le K_1 \le K_2^0 \le K_2$, Case 5: $K_1^0 \le K_1 \le K_2 \le K_2^0$, Case 6: $K_1^0 \le K_2^0 \le K_1 \le K_2$. We derive the results only for Case 1, the arguments being similar for the others. We first have,

$$\begin{aligned}&\textstyle\sum_{t=1}^{T}\theta^{0\prime}(X_t^0 - X_t)(\Sigma^0)^{-1}U_t\\ =\ &\textstyle\sum_{t=K_1^0+1}^{K_2^0}[\delta_1^0(K_1 - K_1^0)\ \ 0](\Sigma^0)^{-1}U_t\\ &+\textstyle\sum_{t=K_2^0+1}^{T}[\delta_1^0(K_1 - K_1^0)\ \ \delta_2^0(K_2 - K_2^0)](\Sigma^0)^{-1}U_t\end{aligned}$$

For the other term,

$$
\begin{aligned}
& \textstyle\sum_{t=1}^{T} \theta^{0\prime}(X_t^0 - X_t)(\Sigma^0)^{-1}(X_t^0 - X_t)\theta^0 \\
= \; & \textstyle\sum_{t=K_1^0+1}^{K_2^0} [\delta_1^0(K_1 - K_1^0) \;\; 0](\Sigma^0)^{-1} \begin{bmatrix} \delta_1^0(K_1 - K_1^0) \\ 0 \end{bmatrix} \\
& + \textstyle\sum_{t=K_2^0+1}^{T} [\delta_1^0(K_1 - K_1^0) \;\; \delta_2^0(K_2 - K_2^0)](\Sigma^0)^{-1} \\
& \begin{bmatrix} \delta_1^0(K_1 - K_1^0) \\ \delta_2^0(K_2 - K_2^0) \end{bmatrix}
\end{aligned}
$$

Collecting terms,

$$
\begin{aligned}
lr_T^2 \;=\; & -\frac{1}{2}\{\textstyle\sum_{t=K_1^0+1}^{K_2^0} [\delta_1^0(K_1 - K_1^0) \; 0](\Sigma^0)^{-1} \\
& \begin{bmatrix} \sum_{t=K_1^0+1}^{K_2^0} \delta_1^0(K_1 - K_1^0) \\ 0 \end{bmatrix} \\
& + \textstyle\sum_{t=K_2^0+1}^{T} [\delta_1^0(K_1 - K_1^0) \; \delta_2^0(K_2 - K_2^0)](\Sigma^0)^{-1} \\
& \begin{bmatrix} \sum_{t=K_1^0+1}^{K_2^0} \delta_1^0(K_1 - K_1^0) \\ \delta_2^0(K_2 - K_2^0) \end{bmatrix}] \\
& -[\textstyle\sum_{t=K_1^0+1}^{K_2^0} [\delta_1^0(K_1 - K_1^0) \; 0](\Sigma^0)^{-1} U_t \\
& + \textstyle\sum_{t=K_2^0+1}^{T} [\delta_1^0(K_1 - K_1^0) \; \delta_2^0(K_2 - K_2^0)](\Sigma^0)^{-1} U_t\}
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow \; & -\frac{1}{2}\{tr((\Sigma^0)^{-1} \\
& \begin{bmatrix} (1-\lambda_2^0)\delta_1^{02} s_1^2 & (1-\lambda_2^0)\delta_1^0\delta_2^0 s_1 s_2 \\ (1-\lambda_2^0)\delta_1^0\delta_2^0 s_1 s_2 & (1-\lambda_2^0)\delta_2^{02} s_2^2 \end{bmatrix}) \\
& +2tr((\Sigma^0)^{-1/2} \\
& \begin{bmatrix} \delta_1^0 s_1(W_1(1) - W_1(\lambda_2^0)) & \delta_2^0 s_2(W_1(1) - W_1(\lambda_2^0)) \\ \delta_1^0 s_1(W_2(1) - W_2(\lambda_2^0)) & \delta_2^0 s_2(W_2(1) - W_2(\lambda_2^0)) \end{bmatrix})\}
\end{aligned}
$$

This implies that

$$\begin{aligned}
\begin{bmatrix} T^{1/2}(\hat{K}_1 - K_1^0) \\ T^{1/2}(\hat{K}_2 - K_2^0) \end{bmatrix} &\Rightarrow \arg\max_{s_1,s_2} H(s_1, s_2) \\
&= -\frac{1}{2}\{tr((\Sigma^0)^{-1} \\
&\quad \begin{bmatrix} (1-\lambda_2^0)\delta_1^{02}s_1^2 & (1-\lambda_2^0)\delta_1^0\delta_2^0 s_1 s_2 \\ (1-\lambda_2^0)\delta_1^0\delta_2^0 s_1 s_2 & (1-\lambda_2^0)\delta_2^{02}s_2^2 \end{bmatrix}) \\
&\quad +2tr((\Sigma^0)^{-1/2} \\
&\quad \begin{bmatrix} \delta_1^0 s_1(W_1(1) - W_1(\lambda_2^0)) & \delta_2^0 s_2(W_1(1) - W_1(\lambda_2^0)) \\ \delta_1^0 s_1(W_2(1) - W_2(\lambda_2^0)) & \delta_2^0 s_2(W_2(1) - W_2(\lambda_2^0)) \end{bmatrix})\}
\end{aligned}$$

The result stated in Theorem 1 follows using the fact that $W(\cdot) = (\Sigma^0)^{-1/2}\Omega^{1/2}W^*(\cdot)$ and the fact that the term Σ^0 is common to all components, hence can be dropped.

Chapter 3

Modeling Exchange Rate Volatility with Random Level Shifts[1]

3.1 Introduction

A vast literature has documented that various measures of the volatility of asset returns display features akin to those of a long-memory process. This is also the case for the volatility of exchange rate series; see, e.g., Anderson et al. (2001) and Anderson and Bollerslev (1997), among others. On the other hand, it has been suggested that the long-memory features present in the data could be due to occasional level shifts; see, e.g. Diebold and Inoue (2001). This follows from similar arguments used in Perron (1989, 1990) who showed that changes in level and/or slope of the trend function of a series causes the estimate of the sum of the autoregressive parameters to be biased towards one, suggesting non-stationarity.

Some recent papers have tried to assess whether random level shifts are indeed responsible for this long-memory feature and not simply a theoretical curiosity. Early attempts to that effect include Stărică and Granger (2005) and Granger and Hyung (2003), who argued that for the volatility of stock market indices the evidence for long-memory is weaker when level shifts are taken into account. Stărică and Granger (2005) presented evidence that log-absolute returns of the S&P 500 index is an *i.i.d.* series affected by occasional shifts in the unconditional variance and showed that this specification has better forecasting performance than the more traditional GARCH(1,1) model and its fractionally integrated counterpart. Perron and Qu (2007) analyzed the time and spectral domain properties of a stationary short memory process affected by random level shifts. Perron and Qu (2010) showed that, when applied to daily S&P 500 log-absolute returns over the period 1928-2002, the level shift model explains both the shape of the autocorrelations and the path of log periodogram estimates as a function of the number of frequency ordinates used.

[1] **This chapter is joint work with Professor Pierre Perron**

Qu and Perron (2012) estimated a stochastic volatility model with level shifts using a Bayesian approach with daily data on returns from the S&P 500 and NASDAQ indices over the period 1980.1-2005.12. They showed that the level shifts account for most of the variation in volatility, that their model provides a better in-sample fit than alternative models and that its forecasting performance is better for the NASDAQ and just as good for the S&P 500 as standard short or long-memory models without level shifts. Lu and Perron (2010) considered a random level shift model for which the series of interest is the sum of a short memory process and a jump or level shift component, modeled as the cumulative sum of a process which is 0 with some probability $(1-\alpha)$ and is a random variable with probability α. They applied it to the logarithm of daily absolute returns for the S&P 500, AMEX, Dow Jones and NASDAQ stock market return indices. The point estimates obtained imply few level shifts for all series. But once these are taken into account, there is little evidence of serial correlation in the remaining noise and, hence, no evidence of long-memory. Once the estimated shifts are introduced to a standard GARCH model applied to the returns series, any evidence of GARCH effects disappears. They also considered rolling out-of-sample forecasts of squared returns. In most cases, the simple random level shifts model clearly outperforms a standard GARCH(1,1) model and, in many cases, it also provides better forecasts than a fractionally integrated GARCH model. Varneskov and Perron (2011) extended the analysis to introduce random level shifts in a general ARFIMA (autoregressive fractionally integrated moving-average) model. They showed that random level shifts are an essential component to model adequately the volatility of various series, whether from daily data or from realized volatility series constructed using high frequency data. From a forecasting perspective, they showed that the random level shift model is the only one that consistently belong to the 10% Model Confidence Set of Hansen et. al. (2011).

Hence, there is growing evidence that a random level shift model is indeed a genuine contender to explain the long-memory features of volatility. However, most of the results so far pertain to stock market return indices. Little evidence is available concerning the properties of the volatility of exchange rate series. Our goal is to use some of the methodologies recently developed to address this issue. One exception is Morana and Beltratti (2004) who considered structural changes in the realized variance processes of the DM/U.S.$ and Yen/U.S.$. exchange rates. Their results show that the volatility of DM/U.S.$ and Yen/U.S.$ exchange rates show clear evidence of genuine long memory and that the structural changes can only partially explain the long memory features. Their forecasting exercises indicate that for short-term forecasting neglecting the structural changes is not that important, but that accounting for them provides substantial improvements for long-term forecasting. However, as noted by Varneskov and Perron (2011), the results obtained are very different when considering historical spans of daily returns compared to shorter spans of realized volatility series constructed from high frequency data.

In this paper, we follow the approach of Lu and Perron (2010). We consider historical series of daily exchange rates for the Yen/U.S.$ and DM/U.S.$. We estimate a random level shifts model for log absolute return series, adopting the specification that the series is the sum of a short memory process and level shift component. The level shift component is specified as a mixture model which takes value 0 with probability α and is some random variable with probability $1-\alpha$. To estimate the model, we cast it into a generalized state space framework with a mixture of normal distributions and use the estimation method developed by Perron and Wada (2009). We also evaluate the forecasting performance of the random level shifts model relative to the popular ARFIMA model. We show that the former indeed provides improved forecasts in most cases. Also, we document that though few level shifts are present, once they are taken into account any evidence of long-memory disappears and what is left is a noise component that is essentially white noise.

The structure of the paper is as follows. Section 2 presents the model and the specifications adopted. Section 3 discusses the method of estimation. The empirical results obtained from estimating the model are presented in Section 4 along with evidence that the level shifts account for all the long-memory features. The forecasting evaluations and comparisons are presented in Section 5. Section 6 offers brief concluding remarks.

3.2 Model

The random level shift model considered is specified by

$$y_t = a + \tau_t + c_t$$

where a is constant term, τ_t is the random level shift component and c_t is a short memory process to model the remaining noise. The level shift component is given by:

$$\tau_t = \tau_{t-1} + \delta_t$$

where

$$\delta_t = \pi_t \eta_t$$

with π_t a binomial variable, which takes the value 1 with probability α and value 0 with probability $1-\alpha$. When $\pi_t = 1$, a random level shift η_t happens having a distribution $\eta_t \sim i.i.d\ N(0, \sigma_\eta^2)$. Furthermore, π_t, η_t and c_t are mutually independent. For the short memory component, in general c_t can be defined by the process $c_t = C(L)e_t$ with $e_t \sim\ i.i.d.\ N(0, \sigma_e^2)$, where $C(L) = \Sigma_{i=0}^{\infty} c_i L^i$, $\sum_{i=0}^{\infty} i|c_i| < \infty$ and $C(1) \neq 0$. However, as will be shown for the series considered, once the level shifts are accounted for, barely any serial correlation remains. Accordingly, we shall simply specify c_t as an AR(1) process. Hence, the model to be used is:

$$y_t = a + \tau_t + c_t$$
$$\tau_t = \tau_{t-1} + \delta_t$$
$$c_t = \phi c_{t-1} + e_t$$
$$\delta_t = \pi_t \eta_t$$

Note that we can write $\delta_t = \pi_t \eta_{1t} + (1-\pi_t)\eta_{2t}$, with $\eta_{it} \sim i.i.d\ N(0, \sigma_{\eta i}^2)$ and $\sigma_{\eta 1}^2 = \sigma_\eta^2$, $\sigma_{\eta 2}^2 = 0$. This allows us to cast the model into a state space framework. More specifically, with the error term being a mixture of two normal distributions, where

$$\Delta y_t = c_t - c_{t-1} + \delta_t$$

and

$$\begin{aligned} \delta_t &= \pi_t \eta_{1t} + (1-\pi_t)\eta_{2t} \\ c_t &= \phi c_{t-1} + e_t \end{aligned}$$

In matrix form,

$$\begin{aligned}\Delta y_t &= HX_t + \delta_t \\ X_t &= FX_{t-1} + U_t\end{aligned}$$

In general, when c_t follows an AR(p) process, then

$$X_t = [c_t, c_{t-1}, \cdots, c_{t-p}]'$$

$$F = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \cdots & \phi_p \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}$$

$H = [1, -1, \cdots, 0]$, U_t is a p-dimensional normally distributed random vector with zero mean and covariance matrix

$$Q = \begin{pmatrix} \sigma_e^2 & 0_{1\times(p-1)} \\ 0_{(p-1)\times 1} & 0_{(p-1)\times(p-1)} \end{pmatrix}$$

Comparing this model with the standard state space model, the difference is that the error term is a mixture of two normal distributions.

3.3 Estimation Method

We apply the estimation method proposed by Perron and Wada (2009), see also Wada and Perron (2006). In their paper, they generalized the trend cycle decomposition framework based on unobserved components with errors that are mixtures of normal distributions, thereby allowing shifts in the slope and level of the trend functions. The main ingredient that underlies the estimation procedure is that the model can be written as a state space model with normal errors occurring in two different possible states. These states can be described by the combined values of the Bernoulli random variables. From this we can generate the log likelihood function from the decomposition of the prediction errors to obtain estimates. Let $Y_t = [\Delta y_1, \Delta y_2, \cdots, \Delta y_t]$ represents the observations available at time t and $\theta = [\sigma_\eta^2, \alpha, \sigma_e^2, \phi_1, \cdots, \phi_q]$ be the parameter vector to be estimated. The log-likelihood function is

$$\ln(L) = \sum_{t=1}^{T} \ln f(\Delta y_t | Y_{t-1}, \theta),$$

where,

$$f(\Delta y_t | Y_{t-1}, \theta) = \sum_{i=1}^{2} \sum_{j=1}^{2} f(\Delta y_t | s_{t-1} = i, s_t = j, Y_{t-1}, \theta) Pr(s_{t-1} = i, s_t = j | Y_{t-1}, \theta)$$

Here, s_t is an indicator to represent whether or not a random level shift occurs. That is, when $s_t = 1$, then $\pi_t = 1$ and a random level shift happens; when $s_t = 2$, $\pi_t = 0$ and there is no level shift. Let

$$\varpi_t^{ij} = f(\Delta y_t | s_{t-1} = i, s_t = j, Y_{t-1}, \theta), \quad i, j \in 1, 2$$

and

$$\begin{aligned}
\tilde{\varepsilon}_{t|t-1}^{ij} &= Pr(s_{t-1} = i, s_t = j | Y_{t-1}, \theta) \\
&= Pr(s_t = j) \sum_{k=1}^{2} Pr(s_{t-2} = k, s_{t-1} = i | Y_{t-1}, \theta) \\
&= Pr(s_t = j) \widetilde{\varepsilon}_{t-1|t-1}, \quad i,\ j \in 1, 2
\end{aligned}$$

where,

$$\begin{aligned}
\widetilde{\varepsilon}_{t|t-1}^{ki} &= Pr(s_{t-2} = k, s_{t-1} = i | Y_{t-1}, \theta) \\
&= \frac{f(\Delta y_{t-1} | s_{t-2} = k, s_{t-1} = i, Y_{t-2}, \theta) Pr(s_{t-2} = k, s_{t-1} = i, Y_{t-2}, \theta)}{f(\Delta y_{t-1} | Y_{t-2}, \theta)}
\end{aligned}$$

So we have,

$$\widetilde{\varepsilon}_{t+1|t}^{ki} = Pr(s_{t+1} = i, s_t = k | Y_t, \theta) = Pr(s_{t+1} = i) \sum_{j=1}^{2} \widetilde{\varepsilon}_{t|t}^{jk}.$$

with

$$\widetilde{\varepsilon}^{11}_{t+1|t} = \alpha \sum_{j=1}^{2} \widetilde{\varepsilon}^{j1}_{t|t} = \alpha[\widetilde{\varepsilon}^{11}_{t|t} + \widetilde{\varepsilon}^{21}_{t|t}]$$

$$\widetilde{\varepsilon}^{21}_{t+1|t} = \alpha \sum_{j=1}^{2} \widetilde{\varepsilon}^{j2}_{t|t} = \alpha[\widetilde{\varepsilon}^{12}_{t|t} + \widetilde{\varepsilon}^{22}_{t|t}]$$

$$\widetilde{\varepsilon}^{12}_{t+1|t} = (1-\alpha) \sum_{j=1}^{2} \widetilde{\varepsilon}^{j1}_{t|t} = (1-\alpha)[\widetilde{\varepsilon}^{11}_{t|t} + \widetilde{\varepsilon}^{21}_{t|t}]$$

$$\widetilde{\varepsilon}^{22}_{t+1|t} = (1-\alpha) \sum_{j=1}^{2} \widetilde{\varepsilon}^{j2}_{t|t} = (1-\alpha)[\widetilde{\varepsilon}^{12}_{t|t} + \widetilde{\varepsilon}^{22}_{t|t}]$$

In matrix form,

$$\begin{pmatrix} \widetilde{\varepsilon}^{11}_{t+1|1} \\ \widetilde{\varepsilon}^{21}_{t+1|t} \\ \widetilde{\varepsilon}^{12}_{t+1|t} \\ \widetilde{\varepsilon}^{22}_{t+1|t} \end{pmatrix} = \begin{pmatrix} \alpha & \alpha & 0 & 0 \\ 0 & 0 & \alpha & \alpha \\ 1-\alpha & 1-\alpha & 0 & 0 \\ 0 & 0 & 1-\alpha & 1-\alpha \end{pmatrix} \begin{pmatrix} \widetilde{\varepsilon}^{11}_{t|t} \\ \widetilde{\varepsilon}^{21}_{t|t} \\ \widetilde{\varepsilon}^{12}_{t|t} \\ \widetilde{\varepsilon}^{22}_{t|t} \end{pmatrix}$$

The conditional likelihood function for Δy_t is:

$$f(\Delta y_t | s_{t-1} = i, s_t = j, Y_{t-1}, \theta) = \frac{1}{\sqrt{2\pi}} |f_t^{ij}|^{-1/2} \exp(-\frac{v_t^{ij'} (f_t^{ij})^{-\frac{1}{2}} v_t^{ij}}{2})$$

where v_t^{ij} is the prediction error,

$$v_t^{ij} = \Delta y_t - \Delta y^{ij}_{t|t-1} = \Delta y_t - E[\Delta y_t | s_t = i, s_{t-1} = j, Y_{t-1}, \theta]$$

and $f_t^{ij} = E(v_t^{ij} v_t^{ij'})$ is the prediction error variance. The prediction $\Delta y_{t|t-1}$ based on past information does not depend on the state of time t, but Δy_t does. The basic inputs are predictions for the state variables and their variances, which are

$$X^i_{t|t-1} = F X_{t|t-1}$$

$$P^i_{t|t-1} = F P^i_{t|t-1} F' + Q$$

The prediction error is $v_t^{ij} = \Delta y_t - H X^{ij}_{t|t-1}$, so that $f_t^{ij} = H P^i_{t|t-1} H' + R_j$, where R_j is the variance of the error term, which takes two possible values: $R_j = \sigma_\eta^2$ with probability α when $\pi_t = 1$, $R_j = 0$ with probability $(1-\alpha)$ when $\pi_t = 0$. Applying the updating formula, given

$s_t = j, s_{t-1} = i$, we obtain:

$$X^{ij}_{t|t} = X^{i}_{t|t} - P^{i}_{t|t} H'(HP^{i}_{t|-1}H' + R_j)^{-1}(\Delta y_t - HX^{i}_{t|t-1})$$
$$P^{ij}_{t|t} = P^{i}_{t|t-1} - P^{i}_{t|t-1} H'(HP^{i}_{t|t-1}H' + R_j)^{-1} HP^{i}_{t|t-1}.$$

As in Perron and Wada (2009), to reduce the dimension of the estimation problem, we adopt the re-collapsing procedure of Harrison and Stevens (1976), given by:

$$X^{j}_{t|t} = \frac{\sum_{i=1}^{2} Pr(s_{t-1} = i, s_t = j|Y_t, \theta) X^{ij}_{t|t}}{Pr(s_t = j|Y_t, \theta)}$$

and

$$P^{j}_{t|t} = \frac{\sum_{i=1}^{2} Pr(s_{t-1} = i, s_t = j|Y_t, \theta)[P^{ij}_{t|t} + (X^{j}_{t|t} - X^{ij}_{t|t})(X^{j}_{t|t} - X^{ij}_{t|t})']}{Pr(s_t = j|Y_t, \theta)}.$$

3.4 Empirical Results for Exchange Rate Returns

We consider the random level shift model for two daily exchange rate returns series, the Yen/U.S.$ and DM/U.S.$ (both daily series from 10/11/1983 to 7/30/2010; 6,994 observations; obtained from the CRSP database). We apply our level shift model to log-absolute returns since they do not suffer from a non-negativity constraint as do, say, absolute or squared returns. There is also no loss relative to using squared returns in identifying level shift since log-absolute returns are a monotonic transformation. Since we wish to identify the probability of shifts and their locations, the fact that log-absolute returns are quite noisy is not problematic since our methods are robust to the presence of noise. Another reason is the fact that for many asset returns, a log-absolute transformation yields a series that is closer to being normally distributed (see, e.g., Andersen, Bollerslev, Diebold and Labys, 2001). When returns are zero or close to it, the log absolute value transformation implies extreme negative values. Using our method, these outliers would be attributed to the level shift component and thus bias the probability of shifts upward. To avoid this problem, we bound absolute returns away from zero by adding a small constant, i.e., we use $\log(|r_t| + 0.001)$, a technique introduced to the stochastic volatility literature by Fuller (1996).

We first discuss some features of the series. Figure 1 presents the autocorrelation functions up to 100 lags. They show the autocorrelations to decay slowly, a feature typical of a long-memory process. To provide further evidence about this long-memory feature, we estimate the long-memory parameter d using the log-periodogram estimator of Geweke and Porter-Hudak (1983) with a trimming value of $m = T^{1/2}$. The estimates obtained were 0.32 (s.e.=0.03) for the Yen/U.S.$ and 0.34 (s.e.=0.03) for the DM/U.S.$, strongly suggesting the presence of long-memory. The results are qualitatively the same using other values for the bandwidth.

Our aim is to assess whether this long-memory feature is genuine or caused by the presence of level shifts. To that effect, we estimate the RLS model. For the specification of the short memory component, we consider two cases: 1) $c_t = e_t$, so that the parameters to be estimated are $(\sigma_e^2, \sigma_\eta^2, \alpha)$; 2) $c_t = \phi_1 c_{t-1} + e_t$, so that ϕ_1 is an additional parameter to be estimated. The initial value for the state vector is $X'_{0|0} = (0, 0)'$ and the initial value for the covariance matrix is set to

$$P_{0|0} = \begin{pmatrix} \sigma_e^2 & 0 \\ 0 & 0 \end{pmatrix}.$$

The estimates are presented in Table 1. The first feature of interest is that in the specification with an AR(1) component, the estimate of ϕ_1 is very close to zero, indicating the near absence of serial correlation in the noise component once level shifts are accounted for. Hence, in what follows, we concentrate on the case with *i.i.d.* errors. The probability of shifts is small and imprecisely estimated. However, even if it is not statistically significant, as we shall see, it is practically very significant. One way to see this is that the standard deviation of the shifts component is larger than the standard deviation of the noise. Hence, it plays an important role. Given the point estimate of the probability of shifts, one can deduce an implied estimate of the number of shifts occurring in the sample. For the DM/U.S.$ it is 8 and for the Yen/U.S.$ it is 5. As we shall see, even when such few shifts are taken into account the properties of the remaining noise is dramatically altered; a feature we discuss next.

3.4.1 The Effect of Level Shifts on the Long Memory Property

Given the estimation results, we seek to assess whether or not the random level shift component can explain the long memory property of the exchange rate returns. The strategy we adopt is the following. Given the estimated number of shifts, we estimate the break dates and regime specific means using the method of Bai and Perron (2003). Once these are obtained, we estimate the noise component as the difference between the original series and the fitted level shift process.

To be more specific, let m be the number of breaks (8 for the DM/U.S.$ and 5 for the

Yen/U.S.$), T_i $(i = 1, \cdots, m)$ be the break dates (with the convention that $T_0 = 0, T_{m+1} = T$), and $\{u_i; i = 1, ..., m+1\}$ be the means within each regime. The method of Bai and Perron (2003) allows obtaining estimates of the break dates $\{\hat{T}_i; i = 1, ..., m\}$ and regime-specific means $\{\hat{u}_i; i = 1, ..., m+1\}$ as global minimizers of the objective function:

$$\sum_{i=1}^{m+1} \sum_{t=T_{i-1}+1}^{T_i} (y_t - u_i)^2$$

The noise component, say $\hat{c}_t$ is then obtained as $\hat{c}_t = y_i - \sum_{i=1}^{m+1} \hat{u}_i DU_{i,t}$, where $DU_{i,t} = 1$ if $\hat{T}_{i-1} < t \leq \hat{T}_i$ and 0, otherwise. To get a better view of the implied level shift process and its relation to the volatilty of the exchange rate series, Figure 2 presents a graphs of the fitted level shift process in conjunction with a smoothed estimate of the log-absolute returns, obtained using a standard Gaussian kernel.

The autocorrelation function of $\hat{c}_t$ is presented in Figure 3, which shows that basically no serial correlation is left once the level shifts are taken into account. We estimated the long-memory parameter d using the log-periodogram estimator of Geweke and Porter-Hudak (1983) with the same trimming value of $m = T^{1/2}$, applied to the noise component $\hat{c}_t$. The estimates obtained were -0.05 (s.e.=0.03) for the Yen/U.S.$ and -0.02 (s.e.=0.04) for the DM/U.S.$, reinforcing the conclusion that the long-memory feature in the data disappears once the level shifts are taken into account. Even if the level shifts are few in number, they can fully explain and account for the long-memory features of the exchange rate series.

3.5 Forecasting

We now consider the performance of the random level shift model with white noise errors in forecasting volatility proxied by squared returns relative to the ARFIMA model. The reason to make the comparisons with the ARFIMA model is that it is generally perceived as the best forecasting model for asset volatility. In order to assess the robustness of the results, we adopt two different designs; one follows Stărică and Granger (2005) and Lu and Perron (2010), the other follows the framework of Varneskov and Perron (2010).

For the first experiment, we start forecasting at observations 2,000, and re-estimate the models every 20 days, at which point forecasts of up to 200 days are constructed. Since the proxy of realized squared returns are quite noisy, to reduce the effect of sampling variability we follow

Stărică and Granger (2005) in the construction of a metric to gauge the relative performance. Let $\hat{\sigma}^2_{t,p}$ be a p-step ahead forecast of σ^2_{t+p}, the variance of returns r_t at time $t+p$, proxied by the squared demeaned returns. Let n be the number of forecasts produced, then the estimated MSE is constructed as:

$$MSE(p) = \frac{1}{n}\sum_{t=1}^{n}(\overline{r}^2_{t,p} - \sigma^2_{t,p})$$

where $\sigma^2_{t,p} = \sum_{k=1}^{p} \widehat{\sigma}^2_{t+k}$ and $\overline{r}^2_{t,p} = \sum_{k=1}^{p} r^2_{t+k}$ is the realized volatility over the interval $[t+1, t+p]$. The relative forecasting performance of two models is evaluated by the ratio of their MSEs.

For the random level shift model, since the noise component is serially uncorrelated and the level shifts are unforecastable, the best predictor is the mean of the last regime. The issue then becomes how to obtain a good estimate of the current mean of a regime at a given date at which the forecasts are made, without using information after that date. For reasons discussed in Lu and Perron (2010), we resort to using a backwards CUSUM procedure, as in Pesaran and Timmerman (1999). At each forecasting period, we use the CUSUM test of Brown, Durbin and Evans (1975). We determine the cutoff point to get the mean to forecast as the first time the CUSUM statistic crosses one of the critical lines, determined by the criterion that the probability of at least one of the last 1,000 cumulative sums of standardized recursive residuals crossing a line is 10%. The CUSUM is a procedure that effectively indicates the date at which a forecast failure occurs and is, accordingly, the best suited from a forecasting perspective (see, e.g., Pesaran and Timmerman, 1999). The results are presented in Figure 4. They show that the RLS model provides better forecasts than the ARFIMA model for a wide range of forecast horizons.

To assess whether the increased forecast accuracy is statistically significant we consider tests based on a Mincer and Zarnowitz (1969) type of regression given by

$$r^2_{t+p} - r^2_t = b_1(f^{LS}_{t,p} - r^2_t) + b_2(f^{ARFIMA}_{t,p} - r^2_t) + u_t$$

where $f^{LS}_{t,p}$ denotes the p-step ahead forecast of σ^2_{t+p} from the random level shift model and $f^{ARFIMA}_{t,p}$ denotes the p-step ahead forecast from the ARFIMA model. The goal of this regression is to see if the forecasts from the ARFIMA model are uncorrelated with the forecast errors from the random level shift model. This is done by testing the null hypothesis $H^A_0 : (b_1, b_2) = (0, 1)$ using a standard Wald test with an asymptotic chi-square distribution. We can also test if the forecast errors from the random level shift model are uncorrelated with the forecasts from ARFIMA models. This is done by testing $H_0^B : (b_1, b_2) = (1, 0)$. The p-values of such tests are presented in Table 2. For the DM/U.S.$ series, the results are clear and informative. For most horizons, the null hypothesis H_0^A cannot be rejected at the 5% significance level, while the null hypothesis H_0^B can be rejected. In all cases, the p-values of the test for H_0^A are higher than the p-values for the test H_0^B. The results therefore provide evidence that the RLS model provides statistically significant improvements in forecasting relative to the ARFIMA model. For the Yen/U.S.$ series, the results are somewhat ambiguous. Again, the p-values of the test for H_0^A are higher than the p-values for the test H_0^B for all forecasting horizons. However, in most cases both tests lead to a rejection at the 5% level, indicating that both models fail to yield accurate forecasts. Nevertheless, for very long horizons (160 and 200 steps), the RLS model performs significantly better.

The second design to assess the relative forecasting performance follows the method adopted by Varneskov and Perron (2011). The first main difference is that for random level shift model, we obtain τ-step ahead forecasts directly from the filtered estimates obtained when estimating the state-space model. The τ-step ahead forecast is then given by:

$$\widehat{y}_{t|t+\tau} = y_t + F(\sum_{i=0}^{1}\sum_{j=0}^{1} \Pr(s_{t+1} = j) \Pr(s_t = i|Y_t))H_{t|t}^{ij}$$

The second difference pertains to the method to compare the forecasts. We consider out-of-sample forecasting of the last 900 ($T^{out} \in [1, 900]$) days of the two exchange rate series, which are also divided equally into three subperiods to assess the robustness of the results, $T^{out} \in [1, 300]$, $T^{out} \in [301, 600]$ and $T^{out} \in [601, 900]$. We compare three models, the Random Level Shift, $ARFIMA(1, d, 1)$ and $ARFIMA(0, d, 0)$ and consider direct τ-step ahead forecasting for three different horizons $\tau = (1, 5, 10)$. The τ-step ahead forecasts are defined as $\overline{y}_{t+\tau|t} = \sum_{s=1}^{\tau} \widehat{y}_{t+s|t}$. Similarly the cumulative volatility proxy is defined by $\overline{\sigma}_{t,\tau}^2 = \sum_{s=1}^{\tau} y_{t+s}$. We use the mean square forecast error (MSFE) criterion defined as:

$$MSFE_\tau = \frac{1}{T^f} \sum_{t=1}^{T^f} (\overline{\sigma}_{t,\tau}^2 - \overline{y}_{t+\tau|t})^2$$

where T^f is the total number of forecasts produced. The forecasts are evaluated and compared using the 10% Model Confidence Set (MCS) of Hansen et al. (2011). The MSFE's and accompanying MCS p-values (in parenthesis) for multiple and pairwise comparisons of the three models are shown in Table 3. In 19 out of 24 cases, the RLS model belongs to the 10% Model Confidence Set for all and pairwise comparisons as well as all pairwise comparisons. Again, the evidence about the superiority of the RLS model is strongest in the case of the DM/U.S.$ exchange rate. Hence, in general, the evidence points to the fact that is many cases the RLS model provides statistically better forecasts, while for some others the difference is not statistically significant.

3.6 Conclusion

We considered historical series of daily exchange rates for the Yen/U.S.$ and DM/U.S.$. We estimated a random level shifts model for the log absolute return series, adopting the specification that the series is the sum of a short memory process and level shift component. We documented that though few level shifts are present once they are taken into account any evidence of long-memory disappears and what is left is a noise component that is essentially white noise. We also evaluated the forecasting performance of the random level shift model relative to the popular ARFIMA model. We showed that the former indeed provides improved forecasts in most cases. Our paper therefore adds to the recent literature that considered the volatility of stock market indices, by showing that a random level shift model is indeed a serious contender to explaining the long-memory features of the volatility of exchange rate series.

3.7 Appendix 3

Table 3.1: Estimates of the RLS model with $c_t = e_t$ and $c_t = \phi c_{t-1} + e_t$

	σ_η	α	σ_e	ϕ
Yen/U.S.$	1.043 (0.046)	0.0007 (0.00025)	0.7522 (0.007)	
	1.132 (0.247)	0.0005 (0.00019)	0.7533 (0.00657)	0.0309 (0.0136)
DM/U.S.$	0.678 (0.2418)	0.0012 (0.00078)	0.748 (0.0065)	
	0.2289 (4.377)	0.0089 (1.704)	0.7488 (0.007)	0.009 (0.022)

Table 3.2: Comparison of forecasting performance between the RLS and ARFIMA models

Horizon	p-values for the Wald Statistic			
p	Dollar-Mark		Dollar-Yen	
(days)	H_0^A	H_0^B	H_0^A	H_0^B
20	0.0726	0.0081	0.0024	0.0001
40	0.1524	0.0214	0.0016	0.0001
60	0.0514	0.0030	0.0000	0.0000
80	0.3888	0.0604	0.0000	0.0000
100	0.0198	0.0007	0.0254	0.0015
120	0.2739	0.0484	0.0216	0.0014
140	0.0262	0.0023	0.0000	0.0000
160	0.0020	0.0005	0.0809	0.0067
180	0.6402	0.1731	0.0286	0.0039
200	0.9325	0.3196	0.2084	0.0253

Table 3.3: Forecast evaluations of the RLS, ARFIMA(0,d,0) and ARFIMA(1,d,1) models

a) Dollar-Mark exchange rates

	T-out ∈ [1, 300]			T-out ∈ [301, 600]		
	1-step	5-step	10-step	1-step	5-step	10-step
RLS	0.442$^{(a)(c)}$ (1.000)	3.881 (0.052)	5.299$^{(a)(c)}$ (1.000)	0.699$^{(a)(c)}$ (1.000)	3.878$^{(a)(c)}$ (1.000)	8.978$^{(a)(c)}$ (1.000)
ARFIMA(0,d,0)	0.470 (0.0012)	3.903 (0.0069)	8.664 (0.000)	0.738 $^{(a)}$ (0.1099)	4.925 (0.0005)	13.551 (0.000)
ARFIMA(1,d,1)	0.445 $^{(a)}$ (0.6595)	3.695$^{(a)}$ (1.000)	5.805 $^{(a)}$ (0.1418)	0.712$^{(a)}$ (0.199)	4.177 (0.076)	10.406 (0.0169)
	T-out ∈ [601, 900]			**T-out ∈ [1, 900]**		
	1-step	5-step	10-step	1-step	5-step	10-step
RLS	0.641$^{(a)(c)}$ (0.610)	3.066 (0.042)	8.012 (0.0116)	0.573$^{(a)(c)}$ (1.000)	4.522$^{(a)(c)}$ (0.221)	7.430$^{(a)(c)}$ (1.000)
ARFIMA(0,d,0)	0.632$^{(a)}$ (1.000)	2.917$^{(a)}$ (0.756)	7.216$^{(a)}$ (1.000)	0.594 (0.0036)	4.659 (0.0001)	9.810 0.000
ARFIMA(1,d,1)	0.634$^{(a)}$ (0.741)	2.891$^{(a)}$ (1.000)	7.341$^{(a)}$ (0.626)	0.575$^{(a)}$ (0.657)	4.401$^{(a)}$ (1.000)	7.851 (0.0856)

b) Dollar-Yen exchange rates

	T-out ∈ [1, 300]			T-out ∈ [301, 600]		
	1-step	5-step	10-step	1-step	5-step	10-step
RLS	0.538$^{(a)(c)}$ (0.245)	3.881 (0.052)	10.460 (0.0465)	0.691$^{(a)(c)}$ (0.716)	5.675$^{(a)(c)}$ (0.863)	14.624$^{(a)(c)}$ (1.000)
ARFIMA(0,d,0)	0.534$^{(a)}$ (0.533)	3.903 (0.007)	10.651 (0.0002)	0.701$^{(a)}$ (0.208)	6.208 (0.003)	17.520 (0.000)
ARFIMA(1,d,1)	0.530$^{(a)}$ (1.000)	3.695$^{(a)}$ (1.000)	9.752$^{(a)}$ (1.000)	0.686$^{(a)}$ (1.00)	5.629$^{(a)}$ (1.000)	14.784$^{(a)}$ (0.8489)
	T-out ∈ [601, 900]			**T-out ∈ [1, 900]**		
	1-step	5-step	10-step	1-step	5-step	10-step
RLS	0.641$^{(a)(c)}$ (0.610)	4.008 $^{(a)(c)}$ (0.4487)	10.22$^{(a)(c)}$ (0.295)	0.623$^{(a)(c)}$ (0.101)	4.522$^{(a)(c)}$ (0.221)	11.768$^{(a)(c)}$ (0.2855)
ARFIMA(0,d,0)	0.632$^{(a)}$ (1.000)	3.868$^{(a)}$ (1.000)	9.584$^{(a)}$ (1.000)	0.622$^{(a)}$ (0.255)	4.659 (0.0001)	12.585 (0.00)
ARFIMA(1,d,1)	0.634$^{(a)}$ (0.741)	3.8802$^{(a)}$ (0.9034)	9.714$^{(a)}$ (0.673)	0.617$^{(a)}$ (1.00)	4.401$^{(a)}$ (1.000)	11.417$^{(a)}$ (1.000)

(^{a}) indicates that model belongs to 10% MCS using all comparisions, (^{c}) indicates that the RLS model belongs to the 10% MCS using all pairwise comparisions

Figure 3.1: Autocorrelations of log absolute returns

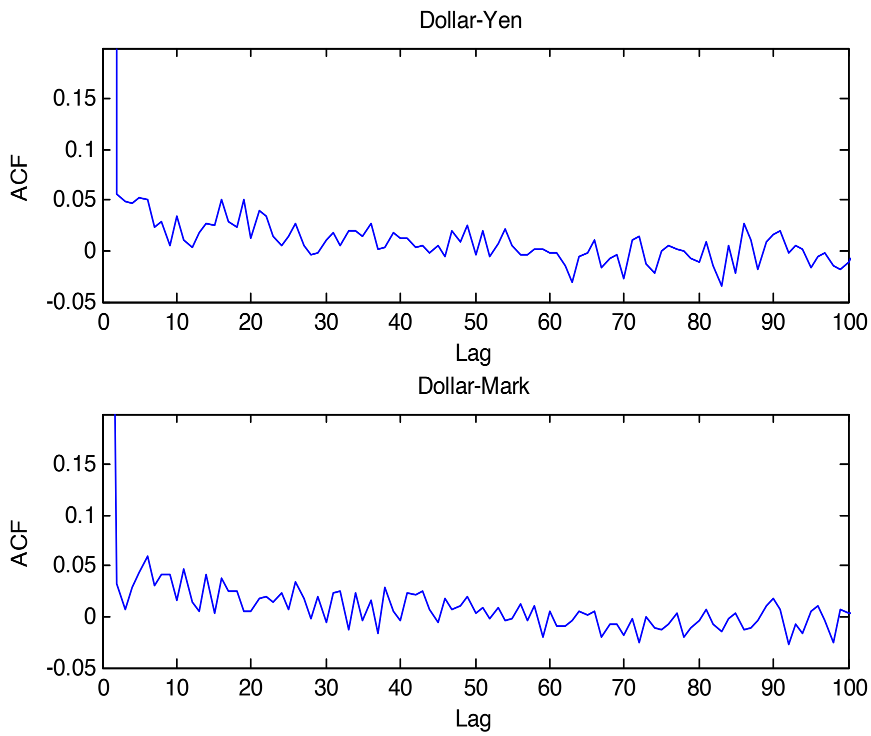


Figure 3.2: Level shift component and smoothed volatility series

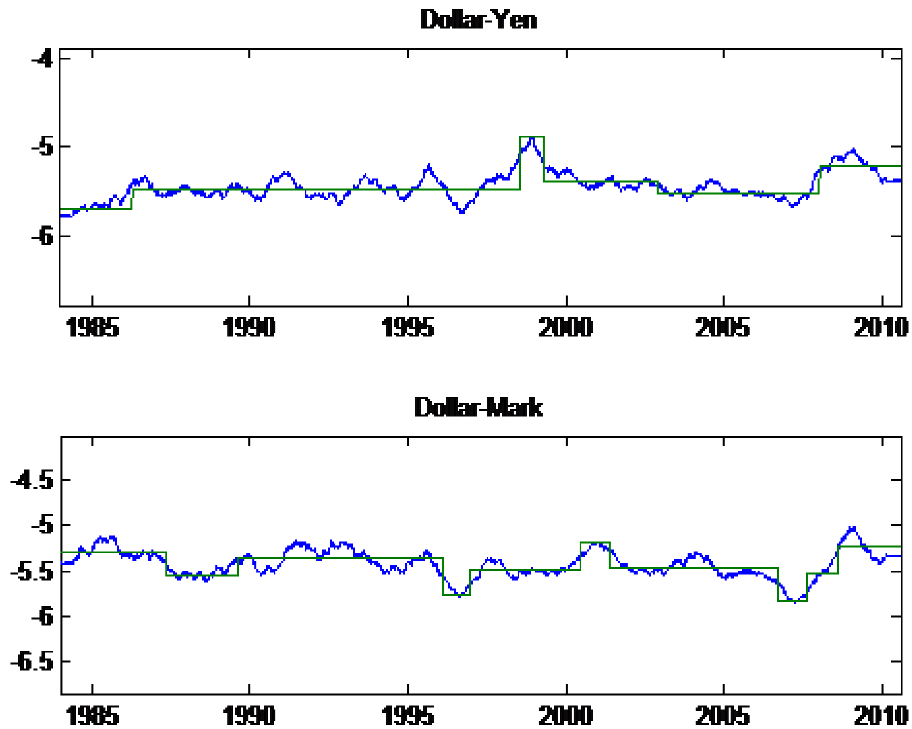


Figure 3.3: Autocorrelations of the residuals from the fitted level shift process

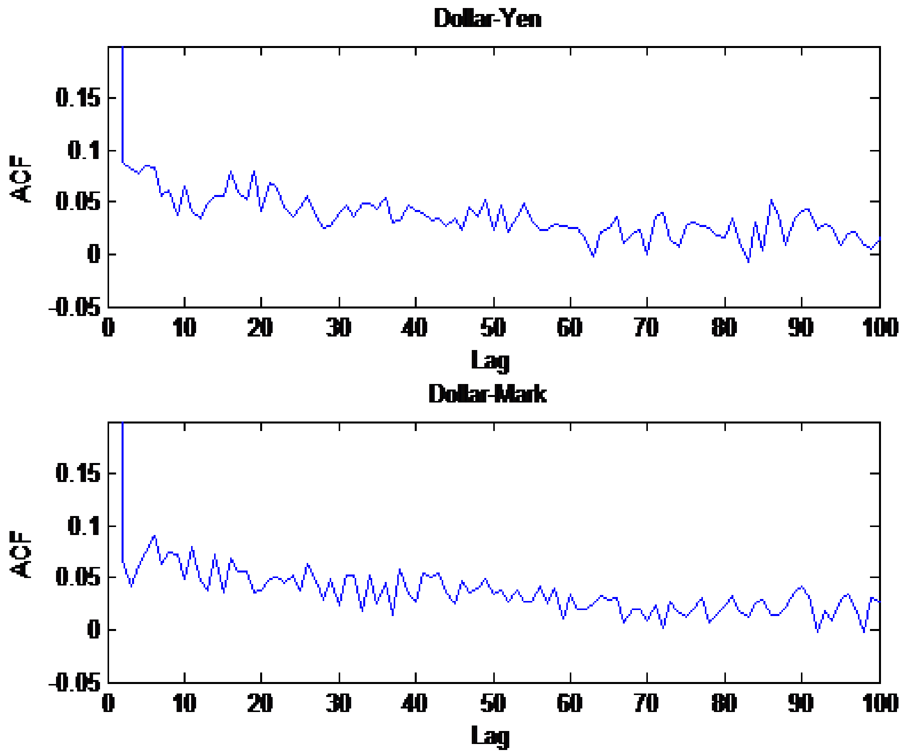
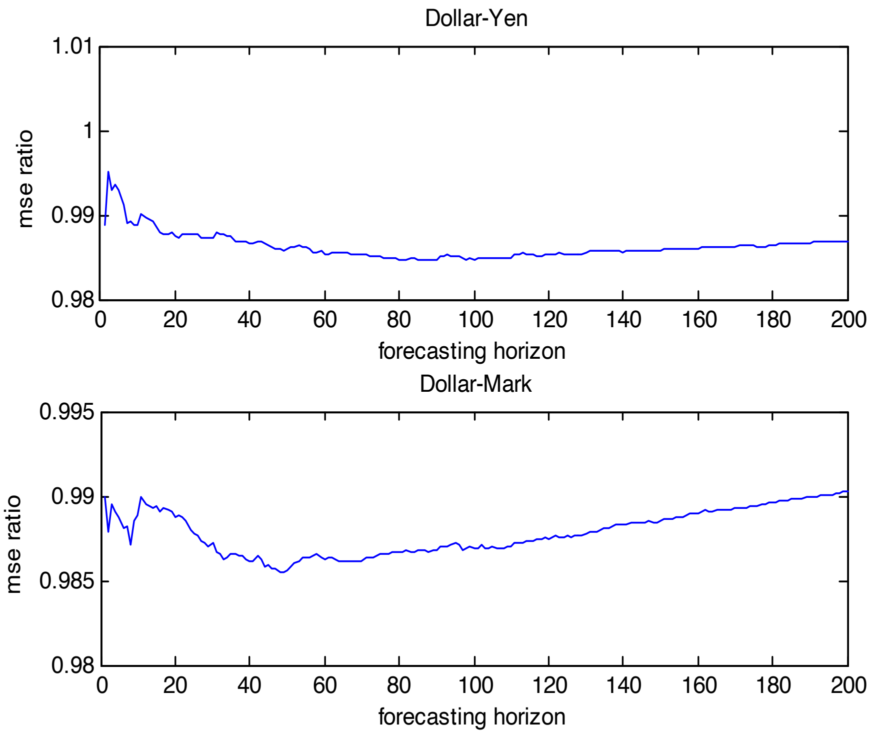


Figure 3.4: MSE ratio



References

Anderson, T.G. and T. Bollerslev (1997), "Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-run in High Frequency Data," *The Journal of Finance* 52, 975-1005.

Anderson, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2001), "The Distribution of Realized Exchange Rate Volatility," *Journal of American Statistical Association* 96, 42-55.

Andrews, D.W.K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59, 816-858.

Bai, J. (1997), "Estimation of a Change Point in Multiple Regression Models," *Review of Economics and Statistics* 79, 551-63.

Bai, J. (2000), "Vector Autoregressive Models with Structural Change in Regression Coefficients and in Variance-Covariance Matrix," *Annals of Economics and Finance* 1, 303-339.

Bai, J. and P. Perron (2003), "Computation and Analysis of Multiple Structural Change Models," *Journal of Applied Econometrics* 18, 1-22.

Bai, J. Lumsdaine, R.L., and J.H. Stock (1998), "Testing for and Dating Common Breaks in Multivariate Time Series," *Review of Economic Studies* 65, 395-432.

Bai, J. and P. Perron (1998), "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica* 66, 47-78.

Bai, J. and P. Perron (2003), "Computation and Analysis of Multiple Structural Change Models," *Journal of Applied Econometrics* 18, 1-22.

Brown, R.L., J. Durbin and J.M. Evans (1975), "Techiques for Testing the Constancy of Regression Relationships over Time,"*Journal of the Royal Statistical Society* 37, 149-192.

Cavaliere, G. and A.M.R.Taylor,(2007), "Testing for Unit Roots in Time Series Models with Non-Stationary Volatility," *Journal of Econometrics* 140, 919-947.

Davidson, J. (1994): Stochastic Limit Theory. Oxford University Press: Oxford.

Deng, A. and P. Perron (2008), "The Limit Distribution of The Cusum of Squares Test Under General Mixing Conditions," *Econometric Theory* 24, 809-822.

Diebold, F.X. and A. Inoue (2001), "Long Memory and Regime Switching," *Journal of Econometrics* 105, 131-159.

Estrada, F., Perron, P. and B. Martinez (2013), "Statistical Evidence about Human Influence on the Climate System," Unpublished Manuscript, Department of Economics, Boston University.

Fuller, W. A. (1996), Introduction to Time Series (2nd ed.), New York: John Wiley.

Geweke, J. and S. Porter-Hudak (1983), "The Estimation and Application of Long Memory Time Series Models," *Journal of Time Series Analysis* 4, 221-238.

Granger, C.W.J. and N. Hyung (2003), "Occasional Structural Breaks and Long Memory with an Application to the S&P 500 Absolute Stock Returns," *Journal of Empirical Finance* 11, 399-421.

Hansen, P. R. A. Lunde and J.M. Nason (2011), "The Model Confidence Set," *Econometrica* 79, 453-497.

Harrison, P.J. and C.F. Stevens (1976), "Bayesian Forecasting," *Journal of the Royal Statistical Society Series B* 38, 205-247.

Ibragimov, I.A. (1962), "Some Limit Theorems for Stationary Processes," *Theory of Probability and Its Applications* 7, 349-382.

Kejriwal, M. and P. Perron (2008), "The Limit Distribution of the Estimates in Cointegrated Regression Models with Multiple Structural Changes," *Journal of Econometrics* 146, 59-73.

Kejriwal, M. and P. Perron (2010), "Testing for Multiple Structural Changes in Cointegrated Regression Models," *Journal of Business and Economic Statistics* 28, 503-522.

Lu, Y.K. and P. Perron (2010), "Modeling and Forecasting Stock Return Volatility Using a Random Level Shift Model," *Journal of Empirical Finance* 17, 138-156.

Mincer, J. and V. Zarnowitz (1969), "The Evaluation of Economic Forecasts," in *Economic Forecasts and Expectations*, ed. by J. Mincer. New York: National Bureau of Economic Research.

Morana, C. and A. Beltratti (2004), "Structural Change and Long-Range Dependence in Volatility of Exchange Rates: Either, Neither or Both?" *Journal of Empirical Finance* 11, 629-658.

Oka, T. and P. Perron (2011), "Testing for Common Breaks in a Multiple Equation System," Boston University Working Paper.

Pesaran, M.H. and A. Timmerman (1999), "Model Instability and Choice of Observation Window," Discussion Paper 99-19, Department of Economics, University of California, San Diego.

Perron, P. (1989), "The Great Crash, The Oil Price Shock, and The Unit Root Hypothesis," *Econometrica* 57, 1361-1401.

Perron, P. (1990), "Testing for a Unit Root in a Time Series with a Changing Mean," *Journal of Business & Economic Statistics* 8, 153-162.

Perron, P. (2006), "Dealing with Structural Breaks," in Palgrave Handbook of Econometrics, Vol.1: *Econometric Theory*, K. Patterson and T.C. Mills (eds.), Palgrave Macmillam, 278-352.

Perron, P. and Z. Qu (2006), "Estimating Restricted Structural Change Models," *Journal of Econometrics* 134, 373-399.

Perron, P. and Z. Qu (2007), "An Analytical Evaluation of the Log-periodogram Estimate in the Presence of Level Shifts," Boston University Working Paper.

Perron, P. and Z. Qu (2010), "Long-Memory and Level Shifts in the Volatility of Stock Market Return Indices," *Journal of Business and Economic Statistics* 28, 275-290.

Perron,P. and X. Zhu (2005), "Structural Breaks with Deterministic and Stochastic Trends," *Journal of Econometrics* 129, 65-119.

Perron, P. and T. Wada (2009), "Let's Take a Break: Trends and Cycles in U.S. Real GDP," *Journal of Monetary Economics* 56, 749-765.

Perron, P. and Y. Yamamoto (2011), "Using OLS to Estimate and Test for Structural Changes in Models with Endogenous Regressors," Unpublished manuscript, Department of Economics, Boston University.

Qu, Z., and P. Perron (2007), "Estimating and Testing Structural Changes in Multivariate Regressions," *Econometrica* 75, 459-502.

Qu, Z. and P. Perron (2012), "A Stochastic Volatility Model with Random Level Shifts: Theory and Applications to S&P 500 and NASDAQ Return Indices," forthcoming in the *Econometrics Journal*.

Stărică, C. and C.W.J. Granger (2005), "Nonstationarities in Stock Returns," *The Review of Economics and Statistics* 87, 503-522.

Varneskov, R.T. and P. Perron (2011), "Combining Long Memory and Level Shifts in Modeling and Forecasting the Volatility of Asset Returns," Boston University Working Paper.

Wada, T. and P. Perron (2006), "An Alternative Trend-Cycle Decomposition using a State Space Model with Mixtures of Normals: Specifications and Applications to International Data," Boston University Working Paper.

Curriculum Vitae

[redacted]

[redacted]