ABSTRACT

Han, Sanggohn. Inference Regarding Multiple Structural Changes in Linear Models Estimated via Two Stage Least Squares. (Under the direction of Dr. Alastair R. Hall)

Bai and Perron(1998) develop methods that are designed to test for structural stability with an unknown number of break points in the sample. Their analysis is in the context of linear regression models estimated via Ordinary Least Squares(OLS). We extend Bai and Perron’s framework for multiple break testing to linear models via Two Stage Least Squares(2SLS). Within our framework, the break points are estimated simultaneously with the regression parameters via minimization of the residual sum of squares on the second step of the 2SLS estimation. We establish the consistency of the resulting estimated break point fractions and obtain the standard convergence rate of break fraction estimators. Based on that convergence rate we derive the limiting distribution of the break point estimators. We prove that the break point estimator have the same limiting distribution of the arg max of two sided Brownian motion process, which is the same distribution considered by Bai and Perron(1998). We also show that various F-statistics for structural instability based on the 2SLS estimator have the same limiting distribution as the analogous statistics for OLS considered by Bai and Perron(1998). This allows us to extend Bai and Perron’s(1998) sequential procedure for selecting the number of break points to the 2SLS setting. Simulation experiment and application to financial market has been implemented.
To my parents, my brother,
and my wife
Biography

Sanggohn Han was born in Tongyoung city located at the southern part of Republic of Korea on September 6, 1965. He attended the Department of Agricultural Economics at Seoul National University during 1984 and 1988-1991 and joined the army during 1986-1988 to serve his country. After the graduation in 1991 with B.E. degree in Economics he worked for Hanhwa Economic Research Institute until 1996 as a research economist. He joined the Department of Economics, North Carolina State University, in 1996 and has been admitted to the co-major program in Economics and Statistics in 2000.
Acknowledgements

I would like to express my sincere gratitude to my academic advisor, Dr. Alastair R. Hall, for his encouragement, guidance, and support throughout my graduate studies. His technical and editorial advice was essential to the completion of this dissertation and has taught me innumerable lessons and insights on the workings of academic research in general.

Dr. David A. Dickey, co-chair of my advisory committee, offered me much-appreciated comments, advice and encouragement on my dissertation. My thanks also go to the other members of my committee, Dr. Atsushi Inoue and Dr. Sastry G. Pantula, for reading previous drafts of this dissertation and providing many valuable comments that improved the presentation and contents of this dissertation. I would like to thank Dr. Denis Pelletier for attending my final oral exam and providing constructive comments on my dissertation.

My parents, Haksam Han and Philson Gong, receive my deepest gratitude and love for their dedication throughout my entire life. I would like to express my special thanks to my parents-in-law and other family members for their support and encouragement. I also would express my gratitude from my heart to my brother, Sanglyeul Han, for his encouragement and financial support throughout the entire graduate studies. His support was in the end what made this dissertation possible. I’m deeply indebted to him.

Last, but not least, I would like to thank my wife Jiyoun Park for her understanding, love, patience, and encouragement. She has always been with me during this exceptional long journey - one that I had never dreamed of. Without her this dissertation could not have been completed.
Contents

List of Tables vii

1 Introduction 1

2 Review on the Estimation Methods in the Presence of Endogenous Variables 9
  2.1 OLS Estimation and Simultaneous Equations Bias 10
  2.2 The Method of Instrumental Variables 14
  2.3 Two Stage Least Squares (2SLS) Estimation Method 17

3 The Consistency of the Break Fraction Estimator 21
  3.1 The Model for 2SLS Break Point Estimation 21
  3.2 Consistency of the Break Fraction Estimator 25

4 The Convergence Rate and Asymptotic Distribution 35
  4.1 Convergence Rate of the Break Fraction Estimator 35
  4.2 Asymptotic Distribution of the Regression Coefficient Estimator 44

5 Asymptotic Distribution of the Break Point Estimators 54
  5.1 Fixed Magnitude of Shift in the Regression Parameters 56
  5.2 Shrinking Magnitude of Shift in the Regression Parameters 70

6 Test Statistics for Multiple Breaks 101
  6.1 SupF Test 102
  6.2 Sequential Test 120

7 Simulation and Application 133
  7.1 Finite Sample Behavior 133
    7.1.1 One Break Model 134
    7.1.2 Two Break Model 136
    7.1.3 No Break Model 138
  7.2 Empirical Application 139

8 Conclusion 145
List of Tables

Table 1: Finite sample behavior of break fraction estimator: One Break Model: \((\beta^0_1, \beta^0_2) = (0.1, -0.1), \rho = 0.5, R^2 = 0.5 \) . . . . . . . . . . . . 156
Table 2: Finite sample behavior of break fraction estimator: One Break Model: \((\beta^0_1, \beta^0_2) = (1, -1), (5, -5), \rho = 0.5, R^2 = 0.5 \) . . . . . . . 157
Table 3: Relative rejection frequencies of \(F\)-statistics: One Break Model: \((\beta^0_1, \beta^0_2) = (0.1, -0.1), \rho = 0.5, R^2 = 0.5 \) . . . . . . . . . . . . 158
Table 4: Relative rejection frequencies of \(F\)-statistics: One Break Model: \((\beta^0_1, \beta^0_2) = (1, -1), \rho = 0.5, R^2 = 0.5 \) . . . . . . . . . . . . 159
Table 5: Relative rejection frequencies of \(F\)-statistics: One Break Model: \((\beta^0_1, \beta^0_2) = (5, -5), \rho = 0.5, R^2 = 0.5 \) . . . . . . . . . . . . 160
Table 6: Empirical distribution of the estimated number of breaks: One Break Model: \((\beta^0_1, \beta^0_2) = (0.1, -0.1), \rho = 0.5, R^2 = 0.5 \) . . . . . . . . . . . . 161
Table 7: Empirical distribution of the estimated number of breaks: One Break Model: \((\beta^0_1, \beta^0_2) = (1, -1), \rho = 0.5, R^2 = 0.5 \) . . . . . . . . . . . . 162
Table 8: Empirical distribution of the estimated number of breaks: One Break Model: \((\beta^0_1, \beta^0_2) = (5, -5), \rho = 0.5, R^2 = 0.5 \) . . . . . . . . . . . . 163
Table 9: Finite sample behavior of break fraction estimator: Two Breaks Model: \((\beta^0_1, \beta^0_2, \beta^0_3) = (-1, 1, -1), (-5, 5, -5), \rho = 0.5, R^2 = 0.5 \) . . . . . . . . . . . . 164
Table 10: Finite sample behavior of break fraction estimator: Two Breaks Model: \((\beta^0_1, \beta^0_2, \beta^0_3) = (-0.1, 0.1, -0.1), \rho = 0.5, R^2 = 0.5 \) . . . . . . . . . . . . 165
Table 11: Relative rejection frequencies of \(F\)-statistics: Two Breaks Model: \((\beta^0_1, \beta^0_2, \beta^0_3) = (-1, 1, -1), \rho = 0.5, R^2 = 0.5 \) . . . . . . . . . . . . 166
Table 12: Relative rejection frequencies of \(F\)-statistics: Two Breaks Model: \((\beta^0_1, \beta^0_2, \beta^0_3) = (-0.1, 0.1, -0.1), \rho = 0.5, R^2 = 0.5 \) . . . . . . . . . . . . 167
Table 13: Relative rejection frequencies of \(F\)-statistics: Two Breaks Model: \((\beta^0_1, \beta^0_2, \beta^0_3) = (-5, 5, -5), \rho = 0.5, R^2 = 0.5 \) . . . . . . . . . . . . 168
Table 14: Empirical distribution of the estimated number of breaks: Two Breaks Model: \((\beta^0_1, \beta^0_2, \beta^0_3) = (-0.1, 0.1, -0.1), \rho = 0.5, R^2 = 0.5 \) . . . . . . . . . . . . 169
Table 15: Empirical distribution of the estimated number of breaks: Two Breaks Model: \((\beta^0_1, \beta^0_2, \beta^0_3) = (-1, 1, -1), \rho = 0.5, R^2 = 0.5 \) . . . . . . . . . . . . 170
Table 16: Empirical distribution of the estimated number of breaks: Two Breaks Model: \((\beta^0_1, \beta^0_2, \beta^0_3) = (-5, 5, -5), \rho = 0.5, R^2 = 0.5 \) . . . . . . . . . . . . 171
Table 17: Relative rejection frequencies of $F$—statistics: No Break Model: 
$(\beta_0^1, \beta_0^2) = (1, 1), \quad \rho = 0.5, \quad R^2 = 0.5$ .................................................. 172

Table 18: Empirical distribution of the estimated number of breaks: No Break Model: 
$(\beta_0^1, \beta_0^2) = (1, 1), \quad \rho = 0.5, \quad R^2 = 0.5$ .................................................. 173

Table 19: Comparison of the estimation results of Consumption based Capital Asset Pricing Model ........................................... 174
Chapter 1

Introduction

Structural change has been a major concern of economics. Because of events like the great depression, oil price shocks, and so on, economic models with constant coefficients have been found to perform poorly for forecasting purposes as well as for the purpose of analyzing the effect of policy changes. Also, the policy regime change in macro-economic policy has been argued to cause the structural instability of parameters in economic models. The main idea underlying this argument is that the changes in policy are likely to affect the expectations of aggregate variables in macro-economic relationships. As a result, shifts in policy are likely to affect those aggregate relationships. In short, if policymakers attempt to take advantage of statistical relationships, effects operating through expectations may cause the relationships to breakdown. This is the famous ‘Lucas critique’.¹ In an empirical study, Bai(1997) found that the response pattern of interest rates to the changes in the discount rates varies over time and the timing of variation is consistent with the timing of changes in the Fed’s operating procedures. It is well known that

¹See David Romer(1996), section 6.4 for general introduction.
the failure to take into account parameter changes may lead to incorrect policy implications and predictions. Not surprisingly, testing parameter constancy has attracted considerable attention in the literature.\textsuperscript{2}

The earliest tests for structural breaks in economic literature are the tests in Chow(1960) which are for stationary variables and a single break. Chow derives two tests. The first of which is known as an analysis of variance test.\textsuperscript{3} It was discussed in Rao(1952) and Kullback and Rosenblatt(1957) among others. The second test is known as a predictive test because it does not test for stability of the coefficients but in fact tests for unbiasedness of predictions. By the 1970s the Chow test had been used extensively in empirical studies to test structural shifts. As long as it is known to which regime each observation belongs, the Chow test can be applied either to timeseries or to cross-sectional data: a historical time point or some threshold value of a key variable may serve as the break point. For example, Hamermesh(1970) estimated wage equations that switch regime at the threshold value of the price index.

There are several limitations to the Chow test. One of which is that, like all conventional $F$ test, it is generally valid only under the rather strong assumptions that the regression error term does not suffer from autocorrelation or heteroscedasticity and the break point is assumed known \textit{a priori}. Toyoda(1974) and Schmidt and Sickles(1977) demonstrated the sensitivity of the Chow test to heteroscedasticity. More specifically, the assumption of homoskedasticity may be particularly implausible when one is testing the equality of two sets of regression parameters,

\textsuperscript{2}For a review of this area, see Maddala and Kim(1998)[chapter 13] and Judge et.al.(1985)[chapter 19]

\textsuperscript{3}In the following dicussion, ‘Chow test’ refers to this first test for parameter stability, which is also called Chow’s first test
since if the coefficient parameter vector differs between two regimes, the variance of
the error terms may well be different as well. A number of papers have addressed
this issue, including Toyoda(1974), Jayatissa(1977), Schmidt and Sickles(1977),
Phillips and McCabe(1983), Ali and Silver(1985), Ohtani and Toyoda(1985), and
Thursby(1992). All of these papers are concerned with the case in which the vari-
ance of the error terms is $\sigma_1^2$ for the first regime and $\sigma_2^2$ for the second regime. An
approach which is often simpler and is valid much more generally is to use a test
statistic that is robust to heteroskedasticity of unknown form(MacKinnon,1989).
Generalization to the dynamic linear or nonlinear simultaneous equations models
is given in Andrews and Fair(1987).

On many occasions, the break point where regime change occurs may not be
known with certainty, and we need to make an inference on it. Quandt(1960)
is one of the earliest persons to propose inference on the break point, the point
at which regimes change. He searched for the point where the likelihood ratio
is the largest. Brown, Durbin, and Evans(1975) suggest the CUSUM test based
on recursive residuals, assuming that the break points were unknown and that all
regressors were independent of the disturbances. This test is of the goodness-of-fit
type in the sense that they seem applicable against a wide variety of alternatives.
Extensions of this test have been made by Ploberger, Kramer, and Alt(1989) and
Kramer, Ploberger, and Alt(1988) for models with lagged dependent variables.
Ploberger and Kramer(1992) showed how the CUSUM test can be carried out
using OLS residuals, thus avoiding the recursive estimation of the regression co-
efficients. A drawback of the CUSUM test is that they have asymptotically low
power against instability in the intercept but not the entire coefficient vector. For the power problem of the CUSUM test, Ploberger, Kramer, and Alt (1989) proposed the fluctuation test based on comparisons between parameter estimates from the partial samples and those of the complete sample. The regressors of the model were assumed stationary.

Andrews (1993) derived the asymptotic null distribution of the sequential likelihood ratio test (Quandt (1960)) of parameter constancy. He also showed that this test has nontrivial local asymptotic power against all alternatives of nonconstant parameters. Andrew and Ploberger (1994) develop tests with stronger optimality properties than those in Andrews (1993) in the context of Maximum Likelihood estimators and this is extended to the GMM framework by Sowell (1996). A predictive test for parameter stability based on GMM estimates is considered by Ghysels and Hall (1990). The test is based on examining whether parameter estimates of one subsample can be used to predict over the other subsample. Ghysels and Hall (1990) shows that the test has equal asymptotic power against local alternatives in both samples regardless of which subsample is conditioned on. Ghysels, Guay and Hall (1997) extends Ghysels and Hall (1990)’s work by allowing for the instability to occur at an unknown point in the sample and shows that under the null hypothesis of no break the predictive test converges to the sum of the square of the standardized tied down Bessel process and the square of the standardized Bessel process. This structure reflects a decomposition of the statistic into a test for parameter variation and a test of the stability of the overidentifying restrictions. But aforementioned predictive tests share the disadvantage of requiring the

\[4\] For the description of CUSUM test and its drawback, refer to Maddala and Kim (1998)[p. 392-394].
prediction subsample\(^5\) to be sufficiently large to allow the application of conventional asymptotic theory. Dufour, Ghysels and Hall(1994) presents a generalized predictive testing procedure for structural stability in nonlinear dynamic simultaneous equations models. Dufour, Ghysels and Hall(1994)'s predictive tests do not require the prediction subsample to be asymptotically large. Therefore, their tests can be applied to test structural stability near the end of the sample. Hall and Sen(1999) proposes new methodology for testing structural stability in models estimated via generalized method of moments. Their approach is based on a decomposition of the null hypothesis into two components involving parameter constancy and the validity of the overidentifying restrictions both before and after the suspected breakpoint.

The alternative to parameter constancy can be either a discrete change in parameters or a continuous one over time. But the smooth continuous change over time can be a more realistic assumption, and include the discrete change as a special case within this more general framework. Bacon and Watts(1971), Maddala(1977), Granger and Terasvirta(1993), and Lin and Terasvirta(1994) discussed the specification and estimation of Smooth Transition Regression model.

Another strand of literature is concerned about the case in which the alternative to constancy is that the parameters are stochastic and fluctuate according to some time series model. LaMotte and McWhorter(1978) assumed that if the null is not true, the parameters follow a random walk, and constructed an exact \( F \) test for testing against that alternative. Assuming the same alternative Nyblom

\(^5\)The subsample used for prediction based on the estimation results from the other subsample which is called “estimation subsample”; see Dufour, et al.(1994)[p.201]
and Makelainen(1983) derived locally most powerful tests. Nyblom(1989) developed the locally most powerful test against a parameter variation in the form of a martingale. The martingale specification has an advantage of covering several types of departure from constancy: for example, a single jump at an unknown time point (change point model) or slow random variation (typically random walk specification of parameters). Goldfeld and Quandt(1973) introduced a Markov Switching Regression (MSR) model in which transition probabilities are explicitly introduced. The transition probabilities may be considered fixed or nonstationary and functions of exogenous variables. In each case the likelihood function is formed and maximized with respect to all relevant variables. Hamilton(1989,1990) extend Goldfeld and Quandt(1973) to Dynamic models with an autoregression. With MSR model there is the question of how many regimes to consider, though most empirical work is on the two-state model. For example, the date of the second switch in the paper by Garcia and Perron(1996) depended on whether the MSR model was considered as a two-state model or a three-state model.

Recently, Bai and Perron(1998) consider issues related to multiple structural changes, occurring at unknown dates, in their linear regression model estimated by ordinary least squares. They examined the main aspects of the structural break models including the consistency of the break fraction estimators, the rate of convergence, and the construction of tests that allow inference to be made about the presence of structural change and the number of breaks. Bai and Perron(2000a,2000b) also consider a simulation study and empirical application of their procedure. They present an efficient algorithm to obtain global minimizers
of the sum of squared residuals (Bai and Perron, 2000b).

This thesis extends Bai and Perron(1998)’s framework to linear models estimated via Two Stage Least Squares (2SLS). Within our framework, the break points are estimated simultaneously with the regression parameters via minimization of the residual sum of squares on the second step of the 2SLS estimation. We establish the consistency of the resulting break point fractions and derive the limiting distribution of the break point estimator. We show that the various F-statistics for testing parameter constancy based on the 2SLS estimator have the same limiting distribution as the analogous statistics for OLS considered by Bai and Perron(1998). This allows us to extend Bai and Perron(1998)’s sequential procedure for selecting the number of break points to the 2SLS setting. We present evidence from a small simulation study that indicates these methods work well in the 2SLS setting. We also discuss an application of our method to the Consumption based Capital Asset Pricing Model (CCAPM) elaborated in Campbell, Lo and Mackinlay(1997) to reestimate their estimation result.

The rest of the thesis is organized as follows. Chapter 2 briefly reviews the method of intrumental variables and 2SLS estimation method. Chapter 3 lays out the model and assumptions for our analysis in the following chapters and establishes the consistency of the break point fraction estimator based on 2SLS estimation method. Chapter 4 proves the $T$-convergence rate of the break point fraction estimator and derives the asymptotic distribution of the regression coefficient estimator. Based on the $T$-convergence rate of the break fraction estimator, chapter 5 presents the limiting distribution of the break point estimator. Chapter
6 presents the limiting distribution of the various $F$-statistics. The simulation evidence and an application of our method to CCAPM is reported at chapter 7, and some concluding remarks are offered in chapter 8.
Chapter 2

Review on the Estimation Methods in the Presence of Endogenous Variables

In many economic models, several variables are jointly determined by the model. Estimating demand and supply equations is an example of this type of formulation; here the price and quantity are jointly determined. Macroeconomic models are also examples of simultaneous equation specification.

In statistical jargon, those jointly determined variables are called endogenous variables compared to the exogenous variables which are determined from outside the system. The exogenous variables are predetermined and independent of the error terms in the model.

In the following section 2.1 we briefly review the inconsistency of the OLS estimator in the presence of simultaneity among the variables. In section 2.2 instrumental variable(IV) estimation method has been introduced to fix the incon-
sistency problem of the OLS estimator, using instrumental variables as a substitute for the endogenous variables that cause simultaneous equation bias. In section 2.3, two stage least squares (2SLS) estimation method is considered. 2SLS estimation method is of the type of IV estimation method where the estimated endogenous variables are used as regressors replacing the endogenous variables.

2.1 OLS Estimation and Simultaneous Equations Bias

A familiar example of a system of simultaneous equations is a model of market equilibrium, consisting of the following:

\[
q_t = \beta_1^0 p_t + \beta_2^0 z_t + u_t \quad \text{demand function}
\]

\[
q_t = \alpha_1^0 p_t + v_t \quad \text{supply function}
\]

where \(z_t\) is assumed to be the exogenous variable determined outside the model and \(q_t\) and \(p_t\) are in log form and are measured in deviations from their means. Thus \(\beta_1^0\) and \(\alpha_1^0\) are the elasticities of demand and supply, respectively. This is an example of the classical linear simultaneous equations model (SEM) with \(M\) endogenous variables, denoted by \(y_{1,t}, \ldots, y_{M,t}\), and \(K\) exogenous variables, denoted by \(x_{1,t}, \ldots, x_{K,t}\), takes the following form.\(^1\)

\[
y_t' \Gamma + x_t' B = \epsilon_t'
\]

or

\[
Yt + XB - \epsilon = 0
\]

\(^1\)It is called “structural form of the model” and expresses the basic structural relationships among economic variables.
where

\[ y_t = [y_{1,t}, y_{2,t}, \ldots, y_{M,t}]' \]
\[ x_t = [x_{1,t}, x_{2,t}, \ldots, x_{K,t}]' \]
\[ \epsilon_t = [\epsilon_{1,t}, \epsilon_{2,t}, \ldots, \epsilon_{M,t}]' \]
\[ Y = [y_1, y_2, \ldots, y_T]' \]
\[ X = [x_1, x_2, \ldots, x_T]' \]
\[ \epsilon = [\epsilon_1, \epsilon_2, \ldots, \epsilon_T]' \]

and \( \Gamma \) and \( B \) are the matrix of fixed parameters of dimension \( M \times M \) and \( K \times M \), respectively.

The system described in (2.1) consists of \( M \) linear equations. Each equation is linear in the components of \( y_t \), with a particular row of \( \Gamma \) and \( B \) containing the coefficients of an equation in the system.

The following assumptions together with (2.1) comprise the classical linear simultaneous equations model:

- \( \Gamma \) is nonsingular. Thus, the model is complete in the sense that we can solve for \( y_t \) in terms of \( x_t \) and \( \epsilon_t \).
- \( x_t \) is exogenous. That is, \( x_t \) and \( \epsilon_t \) are distributed independently of each other.
- The disturbances \( \epsilon_t \), for \( t = 1, 2, \ldots, T \) are uncorrelated and identically distributed with mean zero and positive definite covariance matrix \( \Sigma \).

The structural form of the model in (2.1) can be solved by determining \( y_t \) in
terms of $x_t$ and $\epsilon_t$. The resulting form of the equations is called the “reduced form
of the model” and can be written as

$$y_t' = -x_t' B \Gamma^{-1} + \epsilon_t' \Gamma^{-1}$$

$$= x_t' \Pi + v_t'$$

where

$$\Pi = -B \Gamma^{-1}$$

$$v_t' = \epsilon_t' \Gamma^{-1}$$

The least squares estimator is biased and inconsistent for the parameters of a
structural equation in a simultaneous equation system. To see this, we consider
the $t^{th}$ equation in (2.2) written as

$$Y \Gamma_i + X B_i - \epsilon_i = 0 \quad (2.3)$$

Some elements of $\Gamma_i$ and $B_i$ are generally known to be zero; also it is customary to
select one endogenous variable to appear on the left-hand side of the equation. This
is called selection of a normalization rule and is achieved by setting one coefficient,
say $\gamma_{ii}$, to the value -1. Then, with rearrangement if necessary,

$$y_i = Y_i \gamma_i^0 + X_i \beta_i^0 + \epsilon_i$$

$$= (Y_i \ X_i) \begin{pmatrix} \gamma_i^0 \\ \beta_i^0 \end{pmatrix} + \epsilon_i$$

or compactly as

$$y_i = W_i \alpha_i^0 + \epsilon_i \quad (2.4)$$

where $Y_i$ contains $m_i$ endogenous variables and $X_i$ contains $k_i$ exogenous variables.
The ordinary least squares (OLS) estimator of $\alpha_i$ in (2.4) is

$$\hat{\alpha}_i = (W_i'W_i)^{-1}W_i'y_i \quad (2.5)$$

$$= (W_i'W_i)^{-1}W_i'(W_i\alpha_0^i + \epsilon_i) \quad (2.6)$$

It has expectation

$$E(\hat{\alpha}_i) = E[(W_i'W_i)^{-1}W_i'(W_i\alpha_0^i + \epsilon_i)]$$

$$= \alpha_0^i + E[(W_i'W_i)^{-1}W_i'\epsilon_i] \quad (2.7)$$

The last term in (2.7) does not vanish because $W_i$ contains endogenous variables that are correlated with $\epsilon_i$. Consequently

$$E(\hat{\alpha}_i) \neq \alpha_0^i \quad (2.8)$$

Thus, OLS estimator $\hat{\alpha}_i$ is biased with magnitude $E[(W_i'W_i)^{-1}W_i'\epsilon_i]$. This bias sometimes called “simultaneous equations bias” depends on the model being studied and the particular form of the dependence between $W_i$ and $\epsilon_i$. Unfortunately we can not readily evaluate the expectation of the second term in equation (2.7) since the expectation operator $E$ is a linear operator. But intuitively it should be clear that unless the second term in equation (2.7) is zero, $\hat{\alpha}_i$ is a biased estimator of $\alpha_i^0$.

It is clear from (2.6) that not only $\hat{\alpha}_i$ is biased but it is inconsistent as well. Using the properties of the $plim$, we can express

$$plim(\hat{\alpha}_i) = \alpha_0^i + plim[W_i'W_i/T]^{-1}plim[W_i'\epsilon_i/T] \neq \alpha_0^i$$

In sum, because of the correlation between $W_i$ and $\epsilon_i$, the estimator $\hat{\alpha}_i$ is biased in small samples as well as inconsistent in large samples. So least squares loses
its attractiveness as an estimator. Obviously, we need to explore other estimating methods.

There are several estimation methods leading to better estimates of parameters. The most common are:

- The reduced form method, or indirect least squares (ILS)
- The method of instrumental variables (IV)
- Two stage least squares (2SLS)
- Limited information maximum likelihood (LIML)
- Three stage least squares (3SLS)
- Full information maximum likelihood (FIML)

The first four methods are called single-equation methods because they are applied to one equation of the system at a time. The 3SLS and FIML are called system methods because they are applied to all the equations of the system simultaneously. In the following sections, we restrict our attention to the IV (section 2.2) and 2SLS (section 2.3) estimation methods.²

2.2 The Method of Instrumental Variables

A general method of obtaining consistent estimates of the parameters in simultaneous equations models is the instrumental variable method. Broadly speaking,

²For the general introduction to the above estimation methods, refer to Judge, et al. (1985)[chapter 15] and Greene (1997)[chapter 16].
an instrumental variable is a variable that is uncorrelated with the error term but correlated with the explanatory variables in the equation. For instance, consider the classical model

\[ y_t = x_t' \beta_0 + u_t \] (2.9)

where \( y_t \) is the dependent variable, \( x_t \) is a \( p \times 1 \) explanatory variable. The error term, \( u_t \), is assumed to have a mean of zero and to be correlated with \( x_t \). Then the OLS estimator of \( \hat{\beta} \) is inconsistent due to the correlation between \( x_t \) and \( u_t \). But we can construct a consistent estimator of \( \beta_0 \) by finding a variable \( z_t \) that is uncorrelated with \( u_t \). Suppose that there exists a set of \( q \) variables \( z_t \), where \( q \) is at least as large as \( p \), such that \( z_t \) is correlated with \( x_t \) but not with \( u_t \). In the present context, the zero correlation property \( E(z_t u_t) = 0 \) provides the fundamental basis for estimating \( \beta_0 \). By replacing the expectation \( E(z_t' u_t) \) with the corresponding sample average we obtain

\[
\frac{1}{T} \sum_{t=1}^{T} z_t( y_t - x_t' \beta) = 0
\] (2.10)

This is a system of \( q \) equations in \( p \) unknowns. By noticing that \( Z \) is a \( T \times q \) matrix with its \( t^{th} \) row \( z_t' \), \( X \) is a \( T \times p \) matrix with its \( t^{th} \) row \( x_t' \) and \( y \) is a \( T \times 1 \) vector with its \( t^{th} \) element \( y_t \), we can find the solution to the equations; If \( q = p \), the unique solution is

\[
\tilde{\beta}_T = (Z'X)^{-1} Z'y
\]

provided that \( Z'X \) is nonsingular; If \( q > p \), (2.10) need have no solution, although there may be a value for \( \beta \) that makes \( Z'(y - X\beta) \) “closest” to zero. We can estimate \( \beta_0 \) by finding that value of \( \beta \) which minimizes the quadratic distance.
from zero of $Z'(y - X\beta)$,

$$Q_n(\beta) = (y - X\beta)'Z\hat{P}_TZ'(y - X\beta) \tag{2.11}$$

where $\hat{P}_T$ is a symmetric $q \times q$ positive definite matrix. The first-order conditions for a minimum in (2.11) are

$$-2X'Z\hat{P}_TZ'(y - X\beta) = 0 \tag{2.12}$$

Provided that $X'Z\hat{P}_TZX$ is nonsingular (for which it is necessary that $Z'X$ have full column rank), the resulting solution is

$$\hat{\beta}_T = (X'Z\hat{P}_TZ'X)^{-1}X'Z\hat{P}_TZ'y \tag{2.13}$$

It is called the instrumental variables (IV) estimator (also known as the “method of moments” estimator) and consistent in large samples.\(^3\)

The estimation procedure we have described is known as instrumental variable estimation and the variables in $z_t$ are known as instrumental variables or, more simply, as instruments. The object of instrumental variable estimation is to use the method of moments to generate a consistent estimator by finding instruments $z_t$ that are correlated with the random regressors $x_t$, but uncorrelated with the random error term $u_t$. Of course, in any given situation, several possible instruments could be chosen. Which to choose? One approach is simply to choose $p$ variables among the $q$ in $Z$. But intuition suggests that throwing away the information contained in the remaining $q - p$ columns is inefficient. A better choice is the projection of the columns of $X$ into the column space of $Z$:

$$\hat{X} = Z(Z'Z)^{-1}Z'X$$

\(^3\)For the proof of consistency, see White(1984)[p.24] and Greene(1997)[p.291].
With this choice\footnote{For the virtues of this choice, refer to Greene(1997)[p.293]} of instrumental variables, $\hat{X}$ for $Z$, we have

$$
\hat{\beta} = (\hat{X}'X)^{-1}\hat{X}'y
= [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'y \tag{2.14}
$$

The resulting estimator can also be obtained by substituting $\hat{P}_T = (Z'Z/T)^{-1}$ into (2.13). It is known as Two Stage Least Squares (2SLS) estimator and we’ll look into it further in the section 2.3.

### 2.3 Two Stage Least Squares (2SLS) Estimation Method

In the previous section, it is shown that 2SLS estimator can be derived as a special case of IV estimator. It can also be calculated via two OLS regressions and it is this approach that gives the method its name. In this section, we describe this alternative derivation and show that the resulting estimator is equivalent to the IV estimator in (2.14)

Consider the equation to be estimated:

$$
y = Y_1\gamma_0 + X_1\beta_0 + u \tag{2.15}
$$

where $y$ is $T \times 1$ vector of dependent(endogenous) variable, $Y_1$ is $T \times m_1$ matrix of the other $m_1$ endogenous variables in (2.15), $X_1$ is $T \times (p - m_1)$ matrix of the exogenous variables in (2.15) so that the regressor $[Y_1 \ X_1]$ has dimension $T \times p$, and $u$ is a $T \times 1$ vector of disturbances that is correlated with $Y_1$, but uncorrelated with $X_1$.\footnote{For the virtues of this choice, refer to Greene(1997)[p.293]}
The 2SLS estimation method consists of replacing the endogenous variables on the right hand side of (2.15) by their predicted values from the reduced form equations

\[ Y_1 = Z\Delta_0 + v \]  

where \( Z \) is a matrix of instrumental variables with each row consisted of \( z'_t = (z'_{t,1}, z'_{t,2}, \ldots, z'_{t,q})' \) that is uncorrelated with \( u_t \) for \( q \geq p \) and \( v \) is a matrix of the error terms with each row consisted of \( v'_t \) that is assumed to be uncorrelated with \( z'_t \). \( \Delta_0 \) is a matrix of parameters of the reduced form equation.

To summarize, 2SLS estimation method is described as the following two steps:

1. **Step 1**: Estimate the reduced form equations (2.16) by OLS and obtain the predicted \( \hat{Y}_1 \).
2. **Step 2**: Replace endogenous variables of the right hand side of (2.15) by their predicted vector \( \hat{Y}_1 \) and estimate the resulting equation by OLS.

The name 2SLS arises from the fact that OLS is used in two stages.

The 2SLS estimator can be shown to be the same as the IV estimator as in (2.14). Let \( \hat{Y}_1 \) be the predicted value of \( Y_1 \) from the reduced form equations (2.16) such that

\[ \hat{Y}_1 = Z(Z'Z)^{-1}Z'Y_1 \]

Then, the estimating equations in the second stage of 2SLS estimation method can

---

\(^5\)The necessary condition for identification of (2.15) requires \( q \geq p \).
be written as
\[
\begin{bmatrix}
\hat{Y}'_1 & \hat{Y}'_1 X_1 \\
X'_1 \hat{Y}_1 & X'_1 X_1
\end{bmatrix}
\begin{bmatrix}
\hat{\gamma} \\
\hat{\beta}
\end{bmatrix}
= 
\begin{bmatrix}
\hat{Y}'_1 y \\
X'_1 y
\end{bmatrix}
\tag{2.17}
\]
where \(\hat{\gamma}\) and \(\hat{\beta}\) now denote the 2SLS estimator of \(\gamma_0\) and \(\beta_0\), respectively. The fact that 2SLS estimator is in fact an IV estimator can be shown using the following two relations\(^6\)
\[
\hat{Y}'_1 \hat{Y}_1 = \hat{Y}'_1 Y_1
\tag{2.18}
\]
\[
\hat{Y}'_1 X_1 = Y'_1 X_1
\tag{2.19}
\]
By substituting (2.18)-(2.19) into (2.17) we obtain
\[
\begin{bmatrix}
\hat{Y}'_1 Y_1 & \hat{Y}'_1 X_1 \\
X'_1 Y_1 & X'_1 X_1
\end{bmatrix}
\begin{bmatrix}
\hat{\gamma} \\
\hat{\beta}
\end{bmatrix}
= 
\begin{bmatrix}
\hat{Y}'_1 y \\
X'_1 y
\end{bmatrix}
\tag{2.20}
\]
But the equation (2.20) is the estimating equation for the IV estimator
\[
\begin{bmatrix}
\hat{\gamma} \\
\hat{\beta}
\end{bmatrix}
= (Z'_1 W_1)^{-1} Z'_1 y
\tag{2.21}
\]
where
\[
Z_1 = [\hat{Y}_1 \quad X_1]
\]
\[
W_1 = [Y_1 \quad X_1]
\]
so that \(\hat{Y}_1\) is the set of instruments for \(Y_1\).

Provided a matrix \(Z_1\) can be found such that

- \(\text{plim}(\frac{1}{T} Z'_1 Z_1) = \Sigma_{Z_1 Z_1}\), a finite symmetric positive definite matrix

- \(\text{plim}(\frac{1}{T} Z'_1 W_1) = \Sigma_{Z_1 W_1}\), a finite nonsingular matrix

\(^6\)For the proof, refer to Johnston(1991)[p.473]
\[ \text{plim}(\frac{1}{T}Z_t' u) = 0 \]

the 2SLS estimator is consistent.\(^7\)

It is also well known that under certain conditions such as 
\[ E(z_t u_t) = 0, \ E(u_t^2 | z_t) = \sigma^2, \]
the 2SLS estimator is efficient in the class of all instrumental variable estimators 
using instruments linear in \( z_t \).\(^8\)

\(^7\)see Johnston(1991)[p.478] 
\(^8\)For further details on the asymptotic efficiency of 2SLS estimator, refer to White(1984)[p.78 ∼ 105]
Chapter 3

The Consistency of the Break Fraction Estimator

Bai and Perron (1998) established the consistency of the break fraction estimator using the properties of least squares estimators. This chapter extends their consistency results to the Two Stage Least Squares (2SLS) framework. Section 3.1 lays out the model and assumptions for the following chapters and section 3.2 presents the proof of the consistency of the break fraction estimator based on Two Stage Least Squares (2SLS) estimator.

3.1 The Model for 2SLS Break Point Estimation

For the estimation and testing of structural breaks based on the 2SLS estimator we consider the multiple linear regression model with m breaks (i.e. $m+1$ regimes)

$$
y_t = x_t' \beta_i^0 + u_t, \quad i = 1, \ldots, m+1 \quad t = T_{i-1}^0 + 1, \ldots, T_i^0
$$

(3.1)
where \( T_0^0 = 0 \) and \( T_{m+1}^0 = T \). In this model, \( y_t \) is the dependent variable and \( x_t = (1, x_{t,2}, \cdots, x_{t,p})' \) is a \( p \times 1 \) explanatory variable. The error term, \( u_t \), is assumed to have a mean of zero and to be correlated with \( x_t \). In view of this correlation, OLS estimation of (3.1) would yield inconsistent estimators of the regression parameters. To this end, it is assumed that there exists a \( q \times 1 \) vector of instrumental invariables \( z_t = (z_{t,1}, z_{t,2}, \cdots, z_{t,q})' \) that is uncorrelated with \( u_t \) for \( q \geq p \). The relationship between the explanatory variable and the instrumental variable is assumed to be

\[
x_t' = z'_t \Delta_0 + v_t'
\]

where \( \Delta_0 = (\delta_{1,0}, \delta_{2,0}, \cdots, \delta_{p,0}) \) with dimension \( q \times p \) and each \( \delta_{j,0} \) for \( j = 1, \cdots, p \) has dimension \( q \times 1 \). The error term \( v_t \) is assumed to be uncorrelated with \( z_t \).

Under the assumption that \( E(u_t | z_t) = \sigma^2 \), the optimal IV estimator is the 2SLS estimator. Our analysis is confined to the 2SLS estimator, although we wish to emphasize that the aforementioned conditional homoscedasticity restriction is only imposed in certain parts of the analysis.

To describe the 2SLS estimation of the model, it is assumed that the number of break points \( m \) is known but their locations are not. Therefore the researcher must estimate both the break points and regression parameters. The estimation process proceeds as follows.

On the first stage, the regression equation for \( x_t \) is estimated via OLS using the whole sample to yield the predicted value for \( x_t' \).

\[
\hat{x}_t' = z_t' \hat{\Delta}_T = z_t' (\sum_{t=1}^{T} z_t z_t')^{-1} \sum_{t=1}^{T} z_t x_t'
\]

The second stage of the 2SLS estimation is itself divided into a number of steps be-
cause of the need to estimate both the break points and the regression parameters. 

The first step of the second stage is to estimate the model 

$$y_t = \hat{x}_t'\beta^*_i + \tilde{u}_t, \quad i = 1, \cdots, m + 1 \quad t = T_{i-1} + 1, \cdots, T_i$$ (3.4) 

via OLS for each possible $m-$partition of the sample, denoted by $\{T_j\}_{j=1}^m$, such that $T_i - T_{i-1} \geq q$. The resulting estimates of $\beta^* = (\beta'_1, \beta'_2, \cdots, \beta'_{m+1})'$ are obtained by minimizing the sum of squares of the residuals 

$$S_T(T_1, \cdots, T_m) = \sum_{i=1}^{m+1} \sum_{t=T_{i-1}+1}^{T_i} (y_t - \hat{x}_t'\beta_i)^2$$

with respect to $\beta = (\beta'_1, \beta'_2, \cdots, \beta'_{m+1})'$.

The second step of the second stage involves getting the estimated break points, $(\hat{T}_1, \cdots, \hat{T}_m)$, which can be obtained as 

$$(\hat{T}_1, \cdots, \hat{T}_m) = \arg \min_{T_1, \cdots, T_m} S_T(T_1, \cdots, T_m)$$

The minimization is taken over all partitions, $(T_1, \cdots, T_m)$ such that $T_i - T_{i-1} \geq q$. The 2SLS estimates of the regression parameters, $\hat{\beta}(\{\hat{T}_i\}_{i=1}^m) = (\beta'_1, \beta'_2, \cdots, \beta'_{m+1})'$, is the associated estimates at the estimated partition, $(\{\hat{T}_i\}_{i=1}^m)$.

For our analysis, we impose the following conditions.

**Assumption 1** $T^{-1/2} \sum_{t=1}^{[T_r]} \{(u_t, v_t')\} \otimes z_t \Longrightarrow V^{1/2}B_m(r)$ where $B_m(r)$ is a $m \times 1$ standard Brownian motion with $m = q \times (p + 1)$, $V^{1/2}V^{1/2} = V$, a finite positive definite matrix, and $\Longrightarrow$ denotes weak convergence in the space $D[0, 1]$ under the Skorokhod metric.
Assumption 2 \( T^{-1} \sum_{t=1}^{[Tr]} z_t x_t' \xrightarrow{p} rQ_{ZX}, \) uniformly in \( r \in [0, 1] \) where \( Q_{ZX} \) is a finite matrix with rank equal to \( p \).

Assumption 3 There exists an \( l_0 > 0 \) such that for all \( l > l_0 \), the minimum eigenvalues of \( A_d = (1/l) \sum_{t=T_0+1}^{T_0+l} z_t z_t' \) and of \( A_d^* = (1/l) \sum_{t=T_0-l}^{T_0} z_t z_t' \) are bounded away from zero for all \( i = 1, \ldots, m+1 \).

Assumption 4 \( T_i^0 = [T \lambda_i^0] \), where \( 0 < \lambda_1^0 < \cdots < \lambda_m^0 < 1 \).

Assumption 5 \( T^{-1} \sum_{t=1}^{[Tr]} z_t z_t' \xrightarrow{p} rQ_{ZZ} \) uniformly in \( r \in [0, 1] \) where \( Q_{ZZ} \) is a positive definite matrix of constants.

A few comments on these assumptions are in order. Assumption 1 includes the restriction that \( E(z_t u_t) = 0 \). Assumption 2 includes the standard rank condition for identification in IV estimation in the linear model.\(^1\) This assumption is sufficient for identification. Note that a necessary condition for identification is that the order condition for identification holds, that is, \( q \geq p \). Assumption 3 requires that there is enough observations near the true break points so that they can be identified. Assumption 4 is a standard requirement to allow the development of an asymptotic theory. It says that each segment increases proportionately as the sample size increases. Assumption 5 is standard for multiple linear regressions. It rules out perfect linear dependencies among \( z_t \).

\(^1\)see e.g. Hall(2005)[p.35].
3.2 Consistency of the Break Fraction Estimator

The proof strategy for the consistency of the break fraction estimator is identical to that used by Bai and Perron (1998). The proof builds from the following two properties of the error sum of squares on the second stage of the 2SLS estimation.

• By the definition of 2SLS as the minimizer of the error sum of squares, it follows

\[
(1/T) \sum_{t=1}^{T} \hat{u}_t^2 \leq (1/T) \sum_{t=1}^{T} \tilde{u}_t^2
\]  

(3.5)

where \( \hat{u}_t = y_t - \hat{x}_t' \hat{\beta}_j \) denote the estimated residuals for \( t \in [\hat{T}_j - 1 + 1, \hat{T}_j] \) in the second stage regression of 2SLS estimation procedure and \( \tilde{u}_t = y_t - \tilde{x}_t' \tilde{\beta}_i^0 \) denote the true residuals for \( t \in [T_{i-1}^0 + 1, T_i^0] \).

• Using \( d_t = \tilde{u}_t - \hat{u}_t = \tilde{x}_t' (\tilde{\beta}_j - \beta_i^0) \) over \( t \in [\hat{T}_j - 1 + 1, \hat{T}_j] \cap [T_{i-1}^0 + 1, T_i^0] \), it follows that

\[
(1/T) \sum_{t=1}^{T} \hat{u}_t^2 = (1/T) \sum_{t=1}^{T} \tilde{u}_t^2 + (1/T) \sum_{t=1}^{T} d_t^2 - (2/T) \sum_{t=1}^{T} \tilde{u}_t d_t
\]

(3.6)

Consistency is established by proving that if at least one of the estimated break fractions does not converge in probability to a true break fraction then the results in (3.5) and (3.6) contradict each other. This contradiction is established using the results in the following two lemmas:

• Lemma 1 proves that \( (1/T) \sum_{t=1}^{T} \tilde{u}_t d_t = o_p(1) \)

• Lemma 2 proves that \( (1/T) \sum_{t=1}^{T} d_t^2 \) is bounded away from zero with positive probability if we assume that one of the break points is not consistently estimated.
Lemma 1 Under Assumptions 1~5, we have $(1/T) \sum_{t=1}^{T} \tilde{u}_t d_t = o_p(1)$

proof)

Using the definition of $d_t$, which is $d_t = \tilde{u}_t - \hat{u}_t = (y_t - \hat{x}_t' \beta_1^{(0)}) - (y_t - \hat{x}_t' \hat{\beta}_j) = \hat{x}_t' (\hat{\beta}_j - \beta_0)$ over $t \in [\hat{T}_{j-1} + 1, \hat{T}_j] \cap [T_{i-1} + 1, T_i]$, we have $\tilde{u}_t d_t = \tilde{u}_t \hat{x}_t' (\hat{\beta}_j - \beta_0) = \hat{x}_t' (\hat{\beta}_j - \beta_0) = \hat{x}_t' (\hat{\beta}_j - \beta_0)$. Thus, $\sum_{t=1}^{T} \tilde{u}_t d_t = \sum_{t=1}^{T} \tilde{u}_t \hat{x}_t' (\hat{\beta}_j - \beta_0) = \tilde{u}' \hat{W}^* \hat{\beta} - \tilde{u}' \tilde{W}^0 \beta^0$

in the matrix notation, where $\tilde{W}^*$ and $\tilde{W}^0$ are the diagonal partitions of $W = (\tilde{x}_1, \cdots, \tilde{x}_t)'$ at $(\hat{T}_1, \cdots, \hat{T}_m)$ and $(T^0_1, \cdots, T^0_m)$, respectively.

In the following arguments, we use matrix notation whenever it is convenient for the derivation of the results. Our second stage regression model (3.4) may be expressed in matrix form as

$$Y = \tilde{W}^0 \beta^0 + \tilde{U}$$

where $Y = (y_1, \cdots, y_T)'$, $\tilde{W} = diag\{W_1, \cdots, W_{m+1}\}$ with $W_i = (\tilde{x}_{i-1}^{T_0} + 1, \cdots, \tilde{x}_i^{T_0})'$, $\tilde{U} = (\tilde{u}_1, \cdots, \tilde{u}_T)'$ and $\beta^0 = \beta^0\{\{T_i\}_{i=1}^{m}\} = (\beta_1^{(0)}, \beta_2^{(0)}, \cdots, \beta_{m+1}^{(0)})'$ with $\beta_i^{(0)} = (\beta_{i1}, \beta_{i2}, \cdots, \beta_{ip}^{(0)})'$ is the regression parameters corresponding to the partition $\{\{T_i\}_{i=1}^{m}\}$.

To prove Lemma 1, we need to show that

$$T^{-1} (\tilde{U}' \tilde{W}^* \hat{\beta} - \tilde{U}' \tilde{W}^0 \beta^0) = o_p(1)$$

Since $\hat{\beta} = (W^*W^*)^{-1}W^*Y$ from the second stage regression of 2SLS estimation procedure, we have

$$\tilde{U}' \tilde{W}^* \hat{\beta} - \tilde{U}' \tilde{W}^0 \beta^0 = \tilde{U}' \tilde{W}^* (\tilde{W}^*\tilde{W}^*)^{-1} \tilde{W}^* Y - \tilde{U}' \tilde{W}^0 \beta^0$$

$$\begin{align*}
&= \tilde{U}' P_{W} (\tilde{W}^0 \beta^0 + \tilde{U}) - \tilde{U}' \tilde{W}^0 \beta^0 \\
&= \tilde{U}' P_{W} \tilde{W}^0 \beta^0 + \tilde{U}' P_{W} \tilde{U} - \tilde{U}' \tilde{W}^0 \beta^0
\end{align*}$$

(3.7)
where $P_{W^*} = \bar{W}^* (\bar{W}^* \bar{W}^*)^{-1} \bar{W}^*$

We first show that

$$\|P_{W^*} \bar{U}\| = O_p(1) \quad (3.8)$$

To this end, we need certain properties of matrix norm and so state these here for convenience. Corresponding to the vector (Euclidean) norm $\|x\| = (\sum_{i=1}^{p} x_i^2)^{1/2}$ we define the matrix (Euclidean) norm as

$$\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}$$

for matrix $A$. Below we use the following properties of this norm:

- $\|A\|$ is equal to the square root of the maximum eigenvalue of $A^\prime A$ and thus,

$$\|A\| \leq (\text{tr}A^\prime A)^{1/2} \quad (3.9)$$

- For a projection matrix $P$, we have

$$\|PA\| \leq \|A\| \quad (3.10)$$

- Let $A : R_1 \rightarrow R_2$ and $B : R_2 \rightarrow R_3$ be linear operators, then we have\(^2\)

$$\|BA\| \leq \|B\| \|A\| \quad (3.11)$$

Finally, for a sequence of matrices, we write $A_T = o_p(1)$ if each of its elements is $o_p(1)$. and likewise for $O_p(1)$.

\(^2\)see Ortega(1987)[p.93-4]
For the proof of (3.8), first notice that \( \|P \tilde{W} \tilde{U}\|^2 = \tilde{U}'P \tilde{W} \tilde{U} \) is the sum of the \( m + 1 \) terms

\[
n_{i,T} = \left( \sum_{T_{i+1}} \hat{x}_t \tilde{u}_t \right)' \left( \sum_{T_{i+1}} \hat{x}_t \hat{x}'_t \right)^{-1} \left( \sum_{T_{i+1}} \hat{x}_t \tilde{u}_t \right)
\]  

(3.12)

for \( i = 0, 1, \ldots, m \). and so we can deduce the order of \( \|P \tilde{W} \tilde{U}\|^2 \) by considering the behavior of \( \sum_{T_{i+1}} \hat{x}_t \tilde{u}_t \) and \( \sum_{T_{i+1}} \hat{x}_t \hat{x}'_t \). From (3.2) and (3.3), it follows that

\[
\hat{x}'_t = z'_i \Delta_T
\]

(3.13)

\[
= z'_i (\Delta_0 + (Z'Z)^{-1}Z'V)
\]

\[
= z'_i \Delta_0 + z'_i (Z'Z)^{-1}Z'V
\]

(3.14)

where \( Z = (z_1, \ldots, z_T)' \) and \( V = (v_1, \ldots, v_T)' \). From (3.1) and (3.4), it follows that

\[
\tilde{u}_t = y_t - \hat{x}'_t \beta^0_i
\]

\[
= (x'_t \beta^0_i + u_t) - \hat{x}'_t \beta^0_i
\]

\[
= u_t + (x_t - \hat{x}_t)' \beta^0_i = u_t + [(v_t + \Delta_0' z_t) - \hat{\Delta}'_T z_t]' \beta^0_i
\]

\[
= u_t + v'_t \beta^0_i + z'_i (\Delta_0 - \hat{\Delta}_T) \beta^0_i
\]

\[
= u_t + v'_t \beta^0_i + z'_i [(Z'Z)^{-1}Z'V] \beta^0_i
\]

(3.15)

For notational convenience we denote \( \Sigma_i \) to be the summation over observations \( t = \hat{T}_i + 1, \ldots, \hat{T}_{i+1} \). From (3.14) and (3.15) it follows that

\[
\sum_i \hat{x}_t \tilde{u}_t = \sum_i [\Delta'_0 z_t + V'Z(Z'Z)^{-1}z_t] \cdot [u_t + v'_t \beta^0_i - z'_i (Z'Z)^{-1}Z'V \beta^0_i]
\]

\[
= \sum_i [\Delta'_0 z_t u_t + V'Z(Z'Z)^{-1} z_t u_t + \Delta'_0 z_t v'_t \beta^0_i + V'Z(Z'Z)^{-1} z_t v'_t \beta^0_i
\]

\[
- \Delta'_0 z_t z'_i (Z'Z)^{-1}Z'V \beta^0_i - V'Z(Z'Z)^{-1} z_t z'_i (Z'Z)^{-1}Z'V \beta^0_i
\]

\[
= \Delta'_0 \sum i z_t u_t + V'Z(Z'Z)^{-1} \sum i z_t u_t + \Delta'_0 \sum i z_t v'_t \beta^0_i
\]

28
\[ + V'Z(Z'Z)^{-1} \sum_i z_t v_t' \cdot \beta_i^0 - \Delta_0' \sum_i z_t z_t' \cdot (Z'Z)^{-1} Z'V \cdot \beta_i^0 \]
\[ - V'Z(Z'Z)^{-1} \sum_i z_t z_t' \cdot (Z'Z)^{-1} Z'V \cdot \beta_i^0 \]

(3.16)

In order to investigate the limiting behavior of \( \sum_i \hat{x}_t \tilde{u}_t \), we need to check the order of the magnitude of each term in (3.16). To this end, we check the order of the magnitude of \( \sum_i z_t u_t \) and \( \sum_i z_t v'_t \) in turn. Consider \( \sum_i z_t u_t \) by defining \( T_i = [Ts] \) and \( \hat{T}_{i+1} = [Tr] \) for some break fractions \( s \) and \( r \) where \( r > s \). It follows that

\[
\left\| T^{-1/2} \sum_{t=\hat{T}_{i+1}}^{\hat{T}_{i+1}} z_t u_t \right\| \leq \sup_{r,s} \left\| T^{-1/2} \sum_{t=\hat{T}_{i+1}}^{[Ts]+1} z_t u_t \right\|
\]
\[
= \sup_{r,s} \left\| T^{-1/2} \sum_{t=1}^{[Tr]} z_t u_t - T^{-1/2} \sum_{t=1}^{[Ts]} z_t u_t \right\|
\]
\[
\leq \sup_{r,s} \left\{ \left\| T^{-1/2} \sum_{t=1}^{[Tr]} z_t u_t \right\| + \left\| T^{-1/2} \sum_{t=1}^{[Ts]} z_t u_t \right\| \right\}
\]
\[
\leq 2 \sup_{r} \left\| T^{-1/2} \sum_{t=1}^{[Tr]} z_t u_t \right\|
\]
\[
= O_p(1) \quad (3.17)
\]

where (3.17) follows from Assumption 1 and the Continuous Mapping Theorem. Therefore,

\[
\sum_i z_t u_t = O_p(T^{1/2}) \quad (3.18)
\]

Similarly, we have

\[
\sum_i z_t v'_t = O_p(T^{1/2}) \quad (3.19)
\]

From (3.16), (3.18), (3.19) and Assumptions 1, 2 and 5, it follows that

\[
\sum_i \hat{x}_t \tilde{u}_t = O_p(T^{1/2}) + O_p(1) + O_p(T^{1/2}) + O_p(1) - O_p(T^{1/2}) - O_p(1)
\]
\[
= O_p(T^{1/2}) \quad (3.20)
\]
Now consider $\sum_i \hat{x}_t \hat{x}_t'$. From (3.13), it follows that

$$
\hat{x}_t \hat{x}_t' = \hat{\Delta}'_T z_t \hat{\Delta}_T
= X'Z(Z'Z)^{-1} z_t (Z'Z)^{-1} Z'X
= (\sum_t x_t z_t') (\sum_t z_t z_t')^{-1} (\sum_t z_t x_t)
$$

and hence that

$$
\sum_i \hat{x}_t \hat{x}_t' = (\sum_t x_t z_t') (\sum_t z_t z_t')^{-1} (\sum_t z_t z_t')^{-1} (\sum_t z_t x_t')
= (T^{-1} \sum_t x_t z_t') (T^{-1} \sum_t z_t z_t')^{-1} (\sum_t z_t z_t')^{-1} (T^{-1} \sum_t z_t x_t')
$$

(3.21)

From (3.21) and Assumptions 1, 2 and 5, it follows that

$$
\sum_i \hat{x}_t \hat{x}_t' = O_p(1) \cdot O_p(1) \cdot O_p(T) \cdot O_p(1) \cdot O_p(1)
= O_p(T)
$$

(3.22)

From (3.12), (3.20) and (3.22), it follows that $n_{i,T} = O_p(1)$ and hence that

$$
\|P_{W} \hat{U}\|^2 = O_p(1)
$$

(3.23)

Therefore, the second term on the right hand side of (3.7) is $O_p(1)$. Now consider the first term on the right hand side of (3.7). Using (3.11), it follows that

$$
\|\hat{U}' P_{W} \hat{W}^{-1} \beta \| \leq \|\hat{U}' P_{W}\| : \|\hat{W}^{-1} \beta\|
$$

(3.24)

Since $W = P_z X$, where $X$ is the original design matrix and $P_Z = Z(Z'Z)^{-1} Z'$ is a projection matrix, it follows from (3.9)-(3.10), (3.2) and Assumptions 1, 2 and 5 that

$$
\|\hat{W}^{-1}\| = \|\hat{W}\| = \|P_Z X\| \leq \|X\| \leq (tr X'X)^{1/2} = O_p(T^{1/2})
$$

(3.25)
and hence from (3.23)-(3.25) that

\[ \| \tilde{U}'P_W\tilde{W}^0\beta^0 \| = O_p(T^{1/2}) \]  

(3.26)

Finally, consider the third term on the right hand side of (3.7), \( \tilde{U}'\tilde{W}^0\beta^0 \). Notice that \( \tilde{U}'\tilde{W}^0 \) consists of \( m + 1 \) terms, \( \sum_{t=T_{j-1}+1}^{T_0} \tilde{x}_t\tilde{u}_t \), corresponding to the true regimes. Using a similar argument to the derivation of (3.20), it can be shown that

\[ \sum_{t=T_{j-1}+1}^{T_0} \tilde{x}_t\tilde{u}_t = O_p(T^{1/2}) \]  

and hence that

\[ \| \tilde{U}'\tilde{W}^0\beta^0 \| = O_p(T^{1/2}) \]  

(3.27)

Combining (3.23), (3.26) and (3.27), it follows that

\[ \tilde{U}'\tilde{W}^*\beta - \tilde{U}'\tilde{W}^0\beta^0 \]

\[ = \tilde{U}'P_W\tilde{W}^0\beta^0 + \tilde{U}'P_W\tilde{U} - \tilde{U}'\tilde{W}^0\beta^0 \]

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

\[ = O_p(T^{1/2}) \]

and hence that \( T^{-1}(\tilde{U}'\tilde{W}^*\beta - \tilde{U}'\tilde{W}^0\beta^0) = O_p(T^{-1/2}) = o_p(1) \) which is the desired result.

**Lemma 2** Under Assumptions 1~5, If the estimated break fraction \( \hat{\lambda}_j \not\xrightarrow{} \lambda_j^0 \) for some \( j \), then

\[ \limsup_{T \to \infty} P\left( T^{-1} \sum_{t=1}^{T} d_t^2 > C\| \hat{\Delta}_T (\beta_j^0 - \beta_{j+1}^0) \| \right) > \epsilon \]

for some \( C > 0 \) and \( \epsilon > 0 \)

**proof**

Denote \( \hat{\lambda} = (\hat{\lambda}_1, \hat{\lambda}_2, \cdots, \hat{\lambda}_m) = (\hat{T}_1/T, \hat{T}_2/T, \cdots, \hat{T}_m/T) \) to be the estimated break
fractions of the true values \( \lambda^0 = (\lambda_1^0, \lambda_2^0, \ldots, \lambda_m^0) \). Suppose \( \hat{\lambda}_j \neq \lambda^0_j \) for some \( j \).

Then there exists \( \eta > 0 \) such that no estimated breaks fall into \([T(\lambda^0_j - \eta), T(\lambda^0_j + \eta)]\) with some positive probability for a subsequence of \( T \) (without loss of generality, assume this subsequence is the same as \( T \)). Suppose further that the interval belongs to the \( k \)th estimated regime, then it follows that \( \hat{T}_{k-1} < T(\lambda^0_j - \eta) \) and \( T(\lambda^0_j + \eta) < \hat{T}_k \). Thus, for \( t \in [T(\lambda^0_j - \eta), T\lambda^0_j] \), we have \( d_t = \hat{x}_t(\hat{\beta}_k - \beta^0_j) \) and for \( t \in [T\lambda^0_j + 1, T(\lambda^0_j + \eta)] \), we have \( d_t = \hat{x}_t(\hat{\beta}_k - \beta^0_{j+1}) \).

Therefore, it follows that

\[
\sum_{t=1}^{T} d_t^2 \geq \sum_1^1 d_t^2 + \sum_2^2 d_t^2
\]

\[
= (\hat{\beta}_k - \beta^0_j)' \left( \sum_1 \hat{x}_t\hat{x}_t' \right) (\hat{\beta}_k - \beta^0_j)
\]

\[
+ (\hat{\beta}_k - \beta^0_{j+1})' \left( \sum_2 \hat{x}_t\hat{x}_t' \right) (\hat{\beta}_k - \beta^0_{j+1})
\]

where \( \sum_1 \) extends over the set \( \{T(\lambda^0_j - \eta) \leq t \leq T\lambda^0_j\} \) and \( \sum_2 \) extends over the set \( \{T\lambda^0_j + 1 \leq t \leq T(\lambda^0_j + \eta)\} \).

Let \( \gamma_1 \) and \( \gamma_2 \) denote the smallest eigenvalue of \( \sum_1 z_t z_t' \) and \( \sum_2 z_t z_t' \), respectively.

Using

\[
\sum_1 \hat{x}_t\hat{x}_t' = \sum_1 \hat{\Delta}_T z_t z_t' \hat{\Delta}_T = \hat{\Delta}_T \left( \sum_1 z_t z_t' \right) \hat{\Delta}_T
\]

and

\[
\sum_2 \hat{x}_t\hat{x}_t' = \hat{\Delta}'_T \left( \sum_2 z_t z_t' \right) \hat{\Delta}_T
\]
it follows that
\[
\sum_1 d_1^2 + \sum_2 d_2^2 = (\hat{\beta}_k - \beta_j^0)' \hat{\Delta}_T' \left( \sum_1 z_t z'_t \right) \hat{\Delta}_T (\hat{\beta}_k - \beta_j^0)
\]
\[
+ (\hat{\beta}_k - \beta_{j+1}^0)' \hat{\Delta}_T' \left( \sum_2 z_t z'_t \right) \hat{\Delta}_T (\hat{\beta}_k - \beta_{j+1}^0)
\]
\[
= (\hat{\Delta}_T (\hat{\beta}_k - \beta_j^0))' \left( \sum_1 z_t z'_t \right) (\hat{\Delta}_T (\hat{\beta}_k - \beta_j^0))
\]
\[
+ (\hat{\Delta}_T (\hat{\beta}_k - \beta_{j+1}^0))' \left( \sum_2 z_t z'_t \right) (\hat{\Delta}_T (\hat{\beta}_k - \beta_{j+1}^0))
\]
\[
\geq \gamma_1 \|\hat{\Delta}_T (\hat{\beta}_k - \beta_j^0)\|^2 + \gamma_2 \|\hat{\Delta}_T (\hat{\beta}_k - \beta_{j+1}^0)\|^2
\]
\[
\geq \min\{\gamma_1, \gamma_2\} \cdot \|\hat{\Delta}_T (\hat{\beta}_k - \beta_j^0)\| + \|\hat{\Delta}_T (\hat{\beta}_k - \beta_{j+1}^0)\|
\]
\[
\geq \frac{1}{2} \cdot \min\{\gamma_1, \gamma_2\} \cdot \|\hat{\Delta}_T (\beta_j^0 - \beta_{j+1}^0)\|^2 \quad (3.28)
\]

Note that in the limit \(\|\hat{\Delta}_T (\beta_j^0 - \beta_{j+1}^0)\|^2 > 0\) with positive probability because \(\hat{\Delta}_T\) converges in probability to the true parameter \(\Delta_0 \neq 0\) which follows from the full rank condition of Assumption 2.

Now consider the right hand side of (3.28). We have
\[
\sum_1 z_t z'_t = (T\eta)(1/T\eta) \sum_{T \lambda_j^0} z_t z'_t = (T\eta)A_T
\]
\[
(3.29)
\]
where \(A_T = (1/T\eta) \sum_{T \lambda_j^0} z_t z'_t\). From Assumption 3, the smallest eigenvalue of \(A_T\) is bounded away from zero. Thus, the smallest eigenvalue of \((T\eta)A_T\) is of order \(T\eta\). Similarly, the smallest eigenvalue of \(\sum_2 z_t z'_t = (T\eta)(1/T\eta) \sum_{T \lambda_{j+1}^0} z_t z'_t\) is of order \(T\eta\). Using these two order statements in (3.28), it follows that
\[
\sum_{t=1}^T d_t^2 \geq \sum_1 d_1^2 + \sum_2 d_2^2 \geq TC \cdot \|\hat{\Delta}_T (\beta_j^0 - \beta_{j+1}^0)\|^2
\]

for some \(C > 0\) and hence that
\[
T^{-1} \sum_{t=1}^T d_t^2 = C \|\hat{\Delta}_T (\beta_j^0 - \beta_{j+1}^0)\|^2 \quad (3.30)
\]

\(3\)The last inequality comes from the fact that \((n - a)'A(a - a) + (n - b)'A(a - b) \geq (1/2)(a - b)'A(a - b)\) for an arbitrary positive definite matrix \(A\) and for all \(n\); see Bai and Perron(1998)[p.69]
The desired result then follows from (3.30) upon recalling that the analysis is premised on an event that occurs with probability $\epsilon$.

**Theorem 1** Under Assumptions 1~5, the estimated break fractions $\hat{\lambda}_j \xrightarrow{p} \lambda_j^0$ for all $j = 1, 2, \cdots, m$

**Proof**

Suppose that $\hat{\lambda}_j \not\xrightarrow{p} \lambda_j^0$ for some $j$ in probability. In this case it follows from (3.6) and Lemmas 1-2 that with probability, $\epsilon > 0$

$$\frac{1}{T} \sum_{t=1}^{T} \hat{u}_t^2 > \frac{1}{T} \sum_{t=1}^{T} \tilde{u}_t^2 + C\|\Delta_T(\beta_j^0 - \beta_{j+1}^0)\|^2 + o_p(1)$$  \tag{3.31}

From Assumptions 1, 2 and 5, it follows that $\hat{\Delta}_T \xrightarrow{p} \Delta_0$. Therefore, it follows from this property and (3.31) that

$$\frac{1}{T} \sum_{t=1}^{T} \hat{u}_t^2 > \frac{1}{T} \sum_{t=1}^{T} \tilde{u}_t^2 + C\|\Delta_0(\beta_j^0 - \beta_{j+1}^0)\|^2 + o_p(1)$$  \tag{3.32}

with probability at least as large as $\epsilon$. Assumptions 2 and 5 imply $\Delta_0$ is full rank and so $\|\Delta_0(\beta_j^0 - \beta_{j+1}^0)\|^2 > 0$. Therefore, (3.32) conflicts with (3.5) which must hold for all $T$ with probability one. Therefore, it must follows $\hat{\lambda}_j \xrightarrow{p} \lambda_j^0$ for all $j$. 


Chapter 4

The Convergence Rate and Asymptotic Distribution

In previous chapter we established the consistency of the break fraction estimator. In this chapter we obtain the convergence rate of the break fraction estimator and derive the asymptotic distribution of the estimators of the regression coefficients. More precisely, section 4.1 proves the $T$ convergence of the break fraction estimator and section 4.2 shows standard root-$T$ asymptotic normality of the estimated coefficients of $\hat{\beta}$.

4.1 Convergence Rate of the Break Fraction Estimator

The rate of convergence not only describes how fast the estimator converges to the true value, it is also necessary in order to derive the limiting distribution. The following Theorem 2 shows that the estimated break fraction $\hat{\lambda}_k$ converges to its true value at rate $T$. It is important to remark that the rate $T$ convergence
pertains to the estimated break fraction \( \hat{\lambda}_k \) and not to \( \hat{T}_i \), the estimated break date. For the latter, Theorem 2 states that with high probability its deviation from the true break is bounded by some constant \( C \) that is independent of \( T \).

**Theorem 2** Under Assumptions 1–5, for every \( \eta > 0 \), there exists a \( C \) such that for all large \( T \), \( P(\vert \hat{\lambda}_k - \lambda_k^0 \vert > C) < \eta \) for \( k = 1, \ldots, m \)

**proof**

The general proof strategy is the same as the one employed in Bai and Perron’s (1998) proof of their Proposition 2 although the specific details are naturally different. Following Bai and Perron (1998), we assume (without loss of generality) that there are only 3 break points \( (m = 3) \) in our population model. We start with the explicit proof of \( T \)-consistency of the second break fraction, \( \hat{\lambda}_2 \). To this end, for each \( \epsilon > 0 \) we define

\[
V_\epsilon = \{(T_1, T_2, T_3) : \vert T_i - T_i^0 \vert \leq \epsilon T, \; i = 1, 2, 3\} \tag{4.1}
\]

Note that Theorem 1 implies

\[
P(\{\hat{T}_1, \hat{T}_2, \hat{T}_3 \in V_\epsilon \}) \rightarrow 1.
\]

Therefore, it suffices to consider the behavior of \( S_T(T_1, T_2, T_3) \) over \( V_\epsilon \) for which \( \vert T_i - T_i^0 \vert < \epsilon T \) for all \( i \). Without loss of generality, we can restrict attention to the case in which \( T_2 < T_2^0 \).

\footnote{Bai and Perron (1998) note that the proof for this case is easily modified to cover the case of \( T_2 > T_2^0 \) using an argument of symmetry}

For \( C > 0 \), we define

\[
V_\epsilon(C) = \{(T_1, T_2, T_3) : \vert T_i - T_i^0 \vert \leq \epsilon T, \; i = 1, 2, 3 \text{ but } T_2 - T_2^0 < -C\} \tag{4.2}
\]

From the above definitions (4.1)-(4.2), we can easily recognize that \( V_\epsilon(C) \subset V_\epsilon \).
Notice that the desired result would be established if it can be shown that for large $C$, $(\hat{T}_1, \hat{T}_2, \hat{T}_3) \not\in V_\epsilon(C)$ - and hence $|\hat{T}_2 - T_2^0| < C$ - with high probability for large $T$. Since $S_T(\hat{T}_1, \hat{T}_2, \hat{T}_3) \leq S_T(\hat{T}_1, T_2^0, \hat{T}_3)$ with probability one as $T \to \infty$, the desired result can be established if it can be shown that for each $\eta > 0$, there exists $C > 0$ and $\epsilon > 0$ such that for large $T$,

$$P(\min\{S_T(T_1, T_2, T_3) - S_T(T_1, T_2^0, T_3)\} \leq 0) < \eta$$

or equivalently

$$P(\min\{[S_T(T_1, T_2, T_3) - S_T(T_1, T_2^0, T_3)]/(T_2^0 - T_2)\} \leq 0) < \eta$$

(4.3)

where the minimum is taken over the set $V_\epsilon(C)$. Therefore, we now prove (4.3).

Define $SSR_1 = S_T(T_1, T_2, T_3)$, $SSR_2 = S_T(T_1, T_2^0, T_3)$, and $SSR_3 = S_T(T_1, T_2, T_2^0, T_3)$. Using these definitions, we have

$$S_T(T_1, T_2, T_3) - S_T(T_1, T_2^0, T_3) = SSR_1 - SSR_2$$

$$= (SSR_1 - SSR_3) - (SSR_2 - SSR_3)$$

(4.4)

This transformation of adding and subtracting $SSR_3$ term helps us in analyzing the given problem in terms of two single structural change problems: the first allowing an additional fourth break at time $T_2^0$ between $T_2$ and $T_3$ and the second an additional fourth break at time $T_2$ between $T_1$ and $T_2^0$. It is then easy to derive exact expressions for (4.4) in terms of estimated coefficients.

Let $(\hat{\beta}_1^*, \hat{\beta}_2^*, \hat{\beta}_\Delta^*, \hat{\beta}_3^*, \hat{\beta}_4^*)$ denote the estimators of $(\beta_1^0, \beta_2^0, \beta_2^0, \beta_3^0, \beta_4^0)$ based on the partition $(T_1, T_2, T_2^0, T_3)$. In particular, $\hat{\beta}_2^*$ is an estimate of $\beta_2^0$ associated with $(0, \cdots, 0, W_{T_1+1}, \cdots, W_{T_2}, 0, \cdots, 0)'$, $\hat{\beta}_\Delta$ is an estimate of $\beta_2^0$ associated with
\(W_{\Delta} = (0, \ldots, 0, W_{T_2+1}, \ldots, W_{T_2^0}, 0, \ldots, 0)'\) and \(\hat{\beta}_3^*\) is an estimate of \(\beta_3^0\) associated with \((0, \ldots, 0, W_{T_2+1}, \ldots, W_{T_3}, 0, \ldots, 0)'\) where \(W = (\bar{x}_1, \ldots, \bar{x}_T)'.\)

Now consider the right hand side of the equation (4.4) term by term. Consider \(SSR_1 - SSR_3\). We have\(^2\)

\[
SSR_1 - SSR_3 = S_T(T_1, T_2, T_3) - S_T(T_1, T_2, T_2^0, T_3) \\
= (\hat{\beta}_3^* - \hat{\beta}_{\Delta})'W_{\Delta}'M_{\bar{W}}W_{\Delta}(\hat{\beta}_3^* - \hat{\beta}_{\Delta})
\]

(4.5)

where \(\bar{W}\) is the diagonal partition of \(W\) at \((T_1, T_2, T_3)\) and \(M_{\bar{W}} = I - \bar{W}(\bar{W}'\bar{W})^{-1}\bar{W}'.\)

Similarly, we have

\[
SSR_2 - SSR_3 = S_T(T_1, T_2^0, T_3) - S_T(T_1, T_2, T_2^0, T_3) \\
= (\hat{\beta}_2^* - \hat{\beta}_{\Delta})'W_{\Delta}'M_{\bar{W}}W_{\Delta}(\hat{\beta}_2^* - \hat{\beta}_{\Delta})
\]

(4.6)

where \(\bar{W}\) is the diagonal partition of \(W\) at \((T_1, T_2^0, T_3)\).

From (4.5)-(4.6), it follows that (4.4) can be written as

\[
SSR_1 - SSR_2 = (\hat{\beta}_3^* - \hat{\beta}_{\Delta})'W_{\Delta}'M_{\bar{W}}W_{\Delta}(\hat{\beta}_3^* - \hat{\beta}_{\Delta}) - (\hat{\beta}_2^* - \hat{\beta}_{\Delta})'W_{\Delta}'M_{\bar{W}}W_{\Delta}
\]

\[
\times (\hat{\beta}_2^* - \hat{\beta}_{\Delta})
\]

(4.7)

Using \(W_{\Delta}'M_{\bar{W}}W_{\Delta} \leq W_{\Delta}'W_{\Delta}\), it follows from (4.7) that

\[
SSR_1 - SSR_2 \geq (\hat{\beta}_3^* - \hat{\beta}_{\Delta})'W_{\Delta}'M_{\bar{W}}W_{\Delta}(\hat{\beta}_3^* - \hat{\beta}_{\Delta}) - (\hat{\beta}_2^* - \hat{\beta}_{\Delta})'W_{\Delta}'W_{\Delta}(\hat{\beta}_2^* - \hat{\beta}_{\Delta})
\]

(4.8)

Substituting for \(M_{\bar{W}}\) in (4.8) and dividing both sides by \(T_2^0 - T_2\), we obtain

\[
(SSR_1 - SSR_2)/(T_2^0 - T_2) \geq [(\hat{\beta}_3^* - \hat{\beta}_{\Delta})'W_{\Delta}'M_{\bar{W}}W_{\Delta}(\hat{\beta}_3^* - \hat{\beta}_{\Delta}) - (\hat{\beta}_2^* - \hat{\beta}_{\Delta})']
\]

\(^2\)See Amemiya(1985)[p.31]

38
\[ \times W'_\Delta W_\Delta(\hat{\beta}^*_2 - \hat{\beta}_\Delta)/(T^0_2 - T_2) \]
\[ = (\hat{\beta}^*_3 - \hat{\beta}_\Delta)'[W'_\Delta W_\Delta/(T^0_2 - T_2)](\hat{\beta}^*_3 - \hat{\beta}_\Delta) \]
\[ - (\hat{\beta}^*_3 - \hat{\beta}_\Delta)'[W'_\Delta \hat{W}/(T_2^0 - T_2)][\hat{W}'\hat{W}/T]^{-1} \]
\[ \times [\hat{W}'W_\Delta/(T^0_2 - T_2)](\hat{\beta}^*_3 - \hat{\beta}_\Delta)' \]
\[ \times [W'_\Delta W_\Delta/(T^0_2 - T_2)](\hat{\beta}^*_3 - \hat{\beta}_\Delta) \]
\[ = N_1 - N_2 - N_3 \quad (4.9) \]

where

\[ N_1 = (\hat{\beta}^*_3 - \hat{\beta}_\Delta)'[W'_\Delta W_\Delta/(T^0_2 - T_2)](\hat{\beta}^*_3 - \hat{\beta}_\Delta) \]
\[ N_2 = (\hat{\beta}^*_3 - \hat{\beta}_\Delta)'[W'_\Delta \hat{W}/(T^0_2 - T_2)][\hat{W}'\hat{W}/T]^{-1}[\hat{W}'W_\Delta/T](\hat{\beta}^*_3 - \hat{\beta}_\Delta) \]
\[ N_3 = (\hat{\beta}^*_3 - \hat{\beta}_\Delta)'[W'_\Delta W_\Delta/(T^0_2 - T_2)](\hat{\beta}^*_3 - \hat{\beta}_\Delta) \]

We now consider the behavior of \( N_1, N_2 \) and \( N_3 \) in turn. Consider \( N_1 \). First, note that by controlling \( \epsilon \) to be small enough, we can control the distance between \( T_i \) and \( T^0_i \) to be small over \( V_\epsilon(C) \). Thus, \( \hat{\beta}^*_3 \) should be close to \( \beta^*_3 \) over \( V_\epsilon(C) \). Note that \( \hat{\beta}_\Delta \) is estimated using observations from \((W_{T_2+1}, \cdots, W_{T^0_2})'\). Over \( V_\epsilon(C) \) where \( T_2 - T^0_2 < -C \), large \( C \) implies that we have more observations for estimating \( \beta^*_2 \) by \( \hat{\beta}_\Delta \). Thus, \( \hat{\beta}_\Delta \) is also close to \( \beta^*_2 \) if \( C \) is large. Hence, for large \( C \), large \( T \), and small \( \epsilon \)

\[ N_1 \geq (1/2)(\beta^*_0 - \beta^*_2)'[W'_\Delta W_\Delta/(T^0_2 - T_2)](\beta^*_0 - \beta^*_2) \quad (4.10) \]

Next, consider \( N_2 \). From the property of LS estimation, \( \hat{\beta}^*_3 \) and \( \hat{\beta}_\Delta \) are \( O_p(1) \) uniformly on \( V_\epsilon(C) \). Also, on \( V_\epsilon(C) \), \((\hat{W}'\hat{W}/T)^{-1} = O_p(1) \) and \( W'_\Delta \hat{W}/(T^0_2 - T_2) = O_p(1) \) because \( W'_\Delta \hat{W} \) is a sum of \( T^0_2 - T_2 \) observations. Furthermore, \(||\hat{W}'W_\Delta/T|| = ||[\hat{W}'W_\Delta/(T^0_2 - T_2)][(T^0_2 - T_2)/T]|| = ||\hat{W}'W_\Delta/(T^0_2 - T_2)||/(T^0_2 - T_2)/T \leq O_p(1)\epsilon \)
over $V_\epsilon(C)$. Thus, we have that

$$N_2 \leq O_p(1) \cdot O_p(1) \cdot O_p(1) \cdot O_p(1) \cdot O_p(1) \epsilon \cdot O_p(1)$$

$$= O_p(1) \epsilon$$

(4.11)

Finally, consider term $N_3$. Since both $\hat{\beta}_2^*$ and $\hat{\beta}_\Delta$ are estimating $\beta_0^2$, it follows that $||\hat{\beta}_2^* - \hat{\beta}_\Delta|| < \rho$ with large probability for every $\rho > 0$, for large $T$, large $C$, and small $\epsilon$. Also, $||W_\Delta W_\Delta/(T_2^0 - T_2)|| = O_p(1)$ uniformly on $V_\epsilon(C)$. Therefore, it follows that

$$N_3 \leq \rho O_p(1)$$

(4.12)

Now combining our results in (4.9)-(4.12), it follows that

$$[(SSR_1 - SSR_2)/(T_2^0 - T_2)] \geq 2^{-1}(\beta_3^0 - \beta_2^0)[W_\Delta W_\Delta/(T_2^0 - T_2)]$$

$$\times (\beta_3^0 - \beta_2^0) - \epsilon O_p(1) - \rho O_p(1)$$

(4.13)

with large probability.

We now show that the first term on the right hand side of (4.13) dominates. Noting that

$$W_\Delta' W_\Delta/(T_2^0 - T_2) = (T_2^0 - T_2)^{-1} \sum_{t=T_2+1}^{T_2^0} \hat{x}_t \hat{x}_t'$$

$$= (T_2^0 - T_2)^{-1} \sum_{t=T_2+1}^{T_2^0} \hat{\Delta}_T' z_t z_t' \hat{\Delta}_T$$

$$= \hat{\Delta}_T' (T_2^0 - T_2)^{-1} \sum_{t=T_2+1}^{T_2^0} z_t z_t' \hat{\Delta}_T$$

and $\hat{\Delta}_T \convergence p \Delta_0$, a matrix of full column rank (from Assumptions 2 and 5), it follows from Assumption 3 that, with large probability, the minimum eigenvalue of $W_\Delta' W_\Delta/(T_2^0 - T_2)$ is bounded away from zero on $V_\epsilon(C)$. Therefore, the first term
on the right hand side of (4.13) dominates the other two terms. This term is positive by Assumption 3. Therefore, 
\[ \frac{(SSR_1 - SSR_2)}{(T_2^0 - T_2)} > 0 \] over \( V_\epsilon(C) \) with large probability. This proves (4.3).

The proof for \( \hat{\lambda}_1, \hat{\lambda}_3 \) is virtually the same as that of \( \hat{\lambda}_2 \). We provide the proof of \( T \)-consistency of \( \hat{\lambda}_1 \). To this end, for \( C > 0 \) we newly define

\[ V_\epsilon(C) = \{(T_1, T_2, T_3) : |T_i - T_i^0| \leq \epsilon T, i = 1, 2, 3 \text{ but } T_1 - T_1^0 < -C\} \quad (4.14) \]

such that \( V_\epsilon(C) \subset V_\epsilon \). Without loss of generality, we assume \( T_1 < T_1^0 \). Since \( S_T(\hat{T}_1, \hat{T}_2, \hat{T}_3) \leq S_T(T_1^0, \hat{T}_2, \hat{T}_3) \) with probability 1, the desired result can be established if it can be shown that for each \( \eta > 0 \), there exists \( C > 0 \) and \( \epsilon > 0 \) such that for large \( T \),

\[ P(\min\{|S_T(T_1, T_2, T_3) - S_T(T_1^0, T_2, T_3)|\} \leq 0) < \eta \quad (4.15) \]

where the minimum is taken over the set \( V_\epsilon(C) \).

Define \( SSR_1 = S_T(T_1, T_2, T_3), SSR_2 = S_T(T_1^0, T_2, T_3) \) and \( SSR_3 = S_T(T_1, T_1^0, T_2, T_3) \). Using these definitions, we have

\[
S_T(T_1, T_2, T_3) - S_T(T_1^0, T_2, T_3) = SSR_1 - SSR_2 \\
= (SSR_1 - SSR_3) - (SSR_2 - SSR_3)
\]

(4.16)

where the first term allows an additional fourth break at time \( T_1^0 \) between \( T_1 \) and \( T_2 \) and the second an additional fourth break at time \( T_1 \) between 0 and \( T_1^0 \).

Let \( (\hat{\beta}_1^*, \hat{\beta}_2^*, \hat{\beta}_3^*, \hat{\beta}_4^*) \) denote the estimators of \( (\beta_1^0, \beta_1^0, \beta_2^0, \beta_3^0, \beta_4^0) \) based on the partition \( (T_1, T_1^0, T_2, T_3) \). In particular, \( \hat{\beta}_1^* \) is an estimate of \( \beta_1^0 \) associated
with \((W_1, \ldots, W_{T_1}, 0, \ldots, 0)\)', \(\hat{\beta}_\Delta\) is an estimate of \(\beta_1^0\) associated with \(W_\Delta = (0, \ldots, 0, W_{T_1+1}, \ldots, W_{T^0}, 0, \ldots, 0)\)' and \(\hat{\beta}_2^\ast\) is an estimate of \(\beta_2^0\) associated with \((0, \ldots, 0, W_{T^0_{T_1+1}}, \ldots, W_{T_2}, 0, \ldots, 0)\)'.

Now we consider the right hand side of the equation (4.16) in turn. Consider \(SSR_1 - SSR_3\). We have
\[
SSR_1 - SSR_3 = S_T(T_1, T_2, T_3) - S_T(T_1, T_1^0, T_2, T_3)
= (\hat{\beta}_2^\ast - \hat{\beta}_\Delta)^TW_\Delta^{}M_WW_\Delta(\hat{\beta}_2^\ast - \hat{\beta}_\Delta) \tag{4.17}
\]
where \(\hat{W}\) is the diagonal partition of \(W\) at \((T_1, T_2, T_3)\). Similarly, we have
\[
SSR_2 - SSR_3 = S_T(T_1^0, T_2, T_3) - S_T(T_1, T_1^0, T_2, T_3)
= (\hat{\beta}_1^\ast - \hat{\beta}_\Delta)^TW_\Delta^{}M_WW_\Delta(\hat{\beta}_1^\ast - \hat{\beta}_\Delta) \tag{4.18}
\]
where \(\hat{W}\) is the diagonal partition of \(W\) at \((T_1^0, T_2, T_3)\). From (4.17)-(4.18), it follows that (4.16) can be written as
\[
SSR_1 - SSR_2 = (\hat{\beta}_2^\ast - \hat{\beta}_\Delta)^TW_\Delta^{}M_WW_\Delta(\hat{\beta}_2^\ast - \hat{\beta}_\Delta) - (\hat{\beta}_1^\ast - \hat{\beta}_\Delta)^TW_\Delta^{}M_WW_\Delta
\times (\hat{\beta}_1^\ast - \hat{\beta}_\Delta)
\geq (\hat{\beta}_2^\ast - \hat{\beta}_\Delta)^TW_\Delta^{}M_WW_\Delta(\hat{\beta}_2^\ast - \hat{\beta}_\Delta) - (\hat{\beta}_1^\ast - \hat{\beta}_\Delta)^TW_\Delta^{}W_\Delta^{}(\hat{\beta}_1^\ast - \hat{\beta}_\Delta)
\tag{4.19}
\]
Substituting \(M_\hat{W}\) into (4.19) and dividing both sides by \((T_1^0 - T_1)\), we obtain
\[
(SSR_1 - SSR_2)/(T_1^0 - T_1) \geq [(\hat{\beta}_2^\ast - \hat{\beta}_\Delta)^TW_\Delta^{}M_WW_\Delta(\hat{\beta}_2^\ast - \hat{\beta}_\Delta) - (\hat{\beta}_1^\ast - \hat{\beta}_\Delta)^T
\times W_\Delta^{}W_\Delta^{}(\hat{\beta}_1^\ast - \hat{\beta}_\Delta)]/(T_1^0 - T_1)
= (\hat{\beta}_2^\ast - \hat{\beta}_\Delta)^T[W_\Delta^{}W_\Delta^{}/(T_1^0 - T_1)](\hat{\beta}_2^\ast - \hat{\beta}_\Delta)
- (\hat{\beta}_2^\ast - \hat{\beta}_\Delta)^T[W_\Delta^{}W_\Delta^{}/(T_1^0 - T_1)][\hat{W}^T\hat{W}/T]^{-1}
\]
\begin{align*}
\times [\tilde{W}'\Delta W_\Delta/(T_1 - T_1)](\hat{\beta}_2 - \hat{\beta}_\Delta) &
\times [W_\Delta^\prime W_\Delta/(T_1^0 - T_1)](\hat{\beta}_1 - \hat{\beta}_\Delta) \\
= M_1 - M_2 - M_3 & \quad (4.20)
\end{align*}

We now consider the limit behavior of $M_1, M_2$ and $M_3$ in turn. Consider term $M_1$. By controlling $\epsilon$ to be small enough, we can control the distance between $T_i$ and $T_i^0$ to be small over $V_\epsilon(C)$. Thus, $\hat{\beta}_2^*$ should be close to $\beta_0^2$ over $V_\epsilon(C)$. Note that $\hat{\beta}_\Delta$ is estimated using observations from $(W_{T_1 + 1}, \cdots, W_{T_0})'$. Over $V_\epsilon(C)$ where $T_1 - T_1^0 < -C$, large $C$ implies that we have more observations for estimating $\beta_1^0$ by $\hat{\beta}_\Delta$. Thus, $\hat{\beta}_\Delta$ is also close to $\beta_1^0$ if $C$ is large. Hence, for large $C$, large $T$, and small $\epsilon$

\begin{align*}
M_1 \geq (1/2)(\beta_2^0 - \beta_1^0)'[W_\Delta^\prime W_\Delta/(T_1^0 - T_1)](\beta_2^0 - \beta_1^0) & \quad (4.21)
\end{align*}

Next, consider term $M_2$. From the property of LS estimation, $\hat{\beta}_2^*$ and $\hat{\beta}_\Delta$ are $O_p(1)$ uniformly on $V_\epsilon(C)$. Also, on $V_\epsilon(C), (\tilde{W}'\tilde{W}/T)^{-1} = O_p(1)$ and $W_\Delta^\prime \tilde{W}/(T_1^0 - T_1) = O_p(1)$ because $W_\Delta^\prime \tilde{W}$ is a sum of $T_1^0 - T_1$ observations. Furthermore, $||\tilde{W}'W_\Delta/T|| = ||[\tilde{W}'W_\Delta/(T_1^0 - T_1)][(T_1^0 - T_1)/T]|| = ||\tilde{W}'W_\Delta/(T_1^0 - T_1)||/(T_1^0 - T_1)/T \leq O_p(1)\epsilon$ over $V_\epsilon(C)$. Thus, we have that

\begin{align*}
M_2 \leq O_p(1) \cdot O_p(1) \cdot O_p(1) \cdot O_p(1)\epsilon \cdot O_p(1) & \quad (4.22)
\end{align*}

Finally, consider term $M_3$. Since both $\hat{\beta}_1^*$ and $\hat{\beta}_\Delta$ are estimating $\beta_1^0$, $||\hat{\beta}_1^* - \hat{\beta}_\Delta|| \leq \rho$ with large probability for every $\rho > 0$, for large $T$, large $C$, and small $\epsilon$. Also, $||W_\Delta^\prime W_\Delta/(T_1^0 - T_1)|| = O_p(1)$ uniformly on $V_\epsilon(C)$. Therefore, it follows that

\begin{align*}
M_3 \leq \rho O_p(1) & \quad (4.23)
\end{align*}
Now combining above results in (4.20)-(4.23), it follows that

\[
[(SSR_1 - SSR_2)/(T^0_1 - T_1)] \geq 2^{-1}(\beta^0_2 - \beta^0_1)[W^\prime_\Delta W_//(T^0_1 - T_1)](\beta^0_2 - \beta^0_1) - \epsilon O_p(1) - \rho O_p(1)
\]

(4.24)

with large probability.

We now show that the first term on the right hand side of (4.24) dominates. Noting that

\[
W^\prime_\Delta W_//(T^0_1 - T_1) = (T^0_1 - T_1)^{-1} \sum_{t=T_1+1}^{T^0_1} \hat{x}_t \hat{x}'_t
\]

\[
= (T^0_1 - T_1)^{-1} \sum_{t=T_1+1}^{T^0_1} \hat{\Delta}'_T z_t z'_t \hat{\Delta}_T
\]

\[
= \hat{\Delta}'_T (T^0_1 - T_1)^{-1} \sum_{t=T_1+1}^{T^0_1} z_t z'_t \hat{\Delta}_T
\]

and \( \hat{\Delta}_T \rightarrow \Delta_0 \), a matrix of full column rank, it follows from Assumption 3, with large probability, the minimum eigenvalue of \( W^\prime_\Delta W_//(T^0_1 - T_1) \) is bounded away from zero on \( V^C \). Therefore, the first term on the right hand side of (4.24) dominates the other two terms and is positive. This term is positive by Assumption 3. Therefore, \( [(SSR_1 - SSR_2)/(T^0_1 - T_1)] > 0 \) over \( V^C \) with large probability. This proves (4.15) and completes the proof of the Theorem.

4.2 Asymptotic Distribution of the Regression Coefficient Estimator

Once the break fractions are estimated, it is clearly desirable to perform inference about the structural parameters \( \beta^0 \). If the break fractions are known a priori then standard arguments can be employed to show the root \( T \) asymptotic
normality of the 2SLS estimator. Since the estimated break fractions converge at rate $T$, this standard asymptotic distribution theory can be extended to the 2SLS estimates based on the estimated break fractions. The relevant results are stated in the following Theorem.

**Theorem 3** If Assumptions 1∼5 hold, then

$$T^{1/2} \left( \hat{\beta}(\{\hat{T}_i\}_{i=1}^m) - \beta^0 \right) \implies N \left( 0_{p(m+1)\times 1}, V_\beta \right)$$

where

$$V_\beta = \begin{pmatrix} V^{(1,1)}_\beta & \ldots & V^{(1,m+1)}_\beta \\ \vdots & \ddots & \vdots \\ V^{(m+1,1)}_\beta & \ldots & V^{(m+1,m+1)}_\beta \end{pmatrix}$$

$$V^{(i,i)}_\beta = (\lambda^0_i - \lambda^0_{i-1})^{-2} \left( Q_{xz}Q_{zz}^{-1}Q_{zx} \right)^{-1} Q_{xz}Q_{zz}^{-1}S_{(i,i)}Q_{zz}^{-1}Q_{zx} \left( Q_{xz}Q_{zz}^{-1}Q_{zx} \right)^{-1}$$

$$V^{(i,j)}_\beta = [(\lambda^0_i - \lambda^0_{i-1})(\lambda^0_j - \lambda^0_{j-1})]^{-1} \left( Q_{xz}Q_{zz}^{-1}Q_{zx} \right)^{-1} Q_{xz}Q_{zz}^{-1}S_{(i,j)}Q_{zz}^{-1}Q_{zx} \cdot \left( Q_{xz}Q_{zz}^{-1}Q_{zx} \right)^{-1} \text{ for } i \neq j$$

Note that $S_{(i,i)}$ denotes the asymptotic variance of $\frac{1}{\sqrt{T}} \sum_{i_0} z_i \tilde{u}_t$, $S_{(i,j)}$ denotes the asymptotic covariance of $\frac{1}{\sqrt{T}} \sum_{i_0} z_i \tilde{u}_t$ and $\frac{1}{\sqrt{T}} \sum_{j_0} z_j \tilde{u}_t$, and their specific forms are given at the end of the proof.

**proof)**

Let $W^*$ and $W^0$ denote diagonal partitions of $W = (\hat{x}_1, \cdots, \hat{x}_T)'$ at $(\hat{T}_1, \cdots, \hat{T}_m)$ and $(T^0_1, \cdots, T^0_m)$, respectively. Then 2SLS estimator is defined as

$$\hat{\beta} = \left( W^{s'} W^* \right)^{-1} W^{s'} Y$$

(4.25)
where \( Y = \bar{W}^0\beta^0 + \tilde{U} \) which can be rearranged as

\[
Y = \bar{W}^0\beta^0 - \bar{W}^*\beta^0 + \bar{W}^*\beta^0 + \tilde{U} = \bar{W}^*\beta^0 + (\bar{W}^0 - \bar{W}^*)\beta^0 + \tilde{U} = \bar{W}^*\beta^0 + \tilde{U}^* \tag{4.26}
\]

where \( \tilde{U}^* = (\bar{W}^0 - \bar{W}^*)\beta^0 + \tilde{U} \).

By substituting (4.26) into (4.25), we have

\[
\hat{\beta} = (\bar{W}^*\bar{W}^*)^{-1}\bar{W}^* (\bar{W}^0\beta^0 + \tilde{U}^*) = (\bar{W}^*\bar{W}^*)^{-1}\bar{W}^* \bar{W}^*\beta^0 + (\bar{W}^*\bar{W}^*)^{-1}\bar{W}^*\tilde{U}^* = \beta^0 + (\bar{W}^*\bar{W}^*)^{-1}\bar{W}^*\tilde{U}^* \tag{4.27}
\]

Thus,

\[
\hat{\beta} - \beta^0 = (\bar{W}^*\bar{W}^*)^{-1}\bar{W}^*\tilde{U}^* = (\bar{W}^*\bar{W}^*)^{-1}\bar{W}^*[(\bar{W}^0 - \bar{W}^*)\beta^0 + \tilde{U}]
\]

Therefore,

\[
\sqrt{T}(\hat{\beta} - \beta^0) = \left( \frac{1}{T} \bar{W}^*\bar{W}^* \right)^{-1} \frac{1}{\sqrt{T}} \bar{W}^* [\tilde{U} + (\bar{W}^0 - \bar{W}^*)\beta^0] = \left( \frac{1}{T} \bar{W}^*\bar{W}^* \right)^{-1} \left( \frac{1}{\sqrt{T}} \bar{W}^*\tilde{U} + \frac{1}{\sqrt{T}} ([\bar{W}^0 - \bar{W}^*]\beta^0) \right) \tag{4.28}
\]

Since Theorem 2 proved that \( \tilde{T}_i - T^0_i = O_p(1) \) for all \( i \), \( (\bar{W}^*\bar{W}^0 - \bar{W}^*\bar{W}^*) \) in (4.28) involves bounded number of terms with probability one. Therefore,

\[
\frac{1}{\sqrt{T}} \|W^0\bar{W}^* - W^*\bar{W}^*\| = \frac{1}{\sqrt{T}} O_p(1) = o_p(1). \tag{4.29}
\]

46
Thus, we have

\[ \sqrt{T}(\hat{\beta} - \beta^0) = \left( \frac{1}{T} W^s W^* \right)^{-1} \frac{1}{\sqrt{T}} W^s \tilde{U} + o_p(1) \]  \hfill (4.30)

By the addition and subtraction of \((T^{-1} \bar{W}^s W^*)^{-1} T^{-1/2} \bar{W}^0 \tilde{U}\) to the right hand side of (4.30), we obtain

\[ \sqrt{T}(\hat{\beta} - \beta^0) = \left( T^{-1} \bar{W}^s W^* \right)^{-1} T^{-1/2} \bar{W}^0 \tilde{U} + \left( T^{-1} \bar{W}^s W^* \right)^{-1} T^{-1/2} \bar{W}^0 \tilde{U} \]

\[ - \left( T^{-1} \bar{W}^s W^* \right)^{-1} T^{-1/2} \bar{W}^0 \tilde{U} + o_p(1) \]

\[ = \left( T^{-1} \bar{W}^0 \tilde{U} \right)^{-1} T^{-1/2} \bar{W}^0 \tilde{U} + \left( T^{-1} \bar{W}^s W^* \right)^{-1} T^{-1/2} \bar{W}^0 \tilde{U} \]

\[ - \left( T^{-1} \bar{W}^0 \tilde{U} \right)^{-1} T^{-1/2} \bar{W}^0 \tilde{U} + \left( T^{-1} \bar{W}^s W^* \right)^{-1} T^{-1/2} \bar{W}^0 \tilde{U} \]

\[ \times T^{-1/2}(\bar{W}^s - \bar{W}^0) \tilde{U} + o_p(1) \]

\[ = \left( T^{-1} \bar{W}^0 \tilde{U} \right)^{-1} T^{-1/2} \bar{W}^0 \tilde{U} + \left( T^{-1} \bar{W}^0 \tilde{U} \right)^{-1} \]

\[ \times \left( T^{-1} \bar{W}^0 \tilde{U} - T^{-1} \bar{W}^s W^* \right) \left( T^{-1} \bar{W}^s W^* \right)^{-1} T^{-1/2} \bar{W}^0 \tilde{U} \]

\[ + \left( T^{-1} \bar{W}^s W^* \right)^{-1} T^{-1/2}(\bar{W}^s - \bar{W}^0) \tilde{U} + o_p(1) \]  \hfill (4.31)

Using a similar argument as in (4.29), it follows

\[ \|T^{-1} \bar{W}^0 \tilde{U} - T^{-1} \bar{W}^s W^*\| = o_p(1) \]  \hfill (4.32)

\[ \|T^{-1/2}(\bar{W}^s - \bar{W}^0) \tilde{U}\| = o_p(1) \]  \hfill (4.33)

Using the triangle inequality, equations (4.32)-(4.33), Assumptions 1 and 2, the property of the matrix norm inequality given in (3.11), it follows from (4.31) that

\[ \sqrt{T}(\hat{\beta} - \beta^0) = \left( \frac{1}{T} \bar{W}^0 \tilde{U} \right)^{-1} \frac{1}{\sqrt{T}} \bar{W}^0 \tilde{U} + o_p(1) \]
By the block diagonal structure of $\hat{W}^0\hat{W}^0$, the coefficient vector of the $i$–th regime can be written as

$$\sqrt{T} (\hat{\beta}_i - \beta_i^0) = \left( \frac{1}{T} \sum_{t_0} \hat{x}_t \hat{x}_t' \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t_0} \hat{x}_t \tilde{u}_t + o_p(1) \quad (4.34)$$

where $\sum_{t_0}$ denote the summation over observations $t = T_{i-1}^0, \ldots, T_i^0$. To derive the limit distribution of $\sqrt{T} (\hat{\beta}_i - \beta_i^0)$ we investigate the limit behavior of each factor of the nondegenerated term on the right hand side of (4.34). We start with the first factor of it. Using the Assumptions 2 and 5, it follows that

$$\frac{1}{T} \sum_{t_0} \hat{x}_t \hat{x}_t' = (\lambda_i^0 - \lambda_{i-1}^0) \left( \sum_t x_t z_t' \right) \left( \sum_t z_t x_t' \right)^{-1} \left( \frac{1}{T_i^0 - T_{i-1}^0} \sum_{t_0} z_t z_t' \right) \times \left( \sum_t z_t x_t' \right)^{-1} \left( \sum_t z_t z_t' \right)$$

$$\Rightarrow (\lambda_i^0 - \lambda_{i-1}^0) Q_{XZ} Q_{ZZ}^{-1} Q_{ZX} \quad (4.35)$$

Now, consider the second factor of the nondegenerated term in (4.34). Using $\hat{x}_t = X'Z(Z'Z)^{-1}z_t$, we have

$$\frac{1}{\sqrt{T}} \sum_{t_0} \hat{x}_t \tilde{u}_t = X'Z(Z'Z)^{-1} \frac{1}{\sqrt{T}} \sum_{t_0} z_t \tilde{u}_t$$

$$\Rightarrow Q_{XZ} Q_{ZZ}^{-1} \frac{1}{\sqrt{T}} \sum_{t_0} z_t \tilde{u}_t \quad (4.36)$$

Since $\sum_{t_0} z_t \tilde{u}_t$ is a function of $\sum_{t_0} z_t u_t, \sum_{t_0} z_t v_t'$ and $Z'V$ which have asymptotic normal distributions by Assumption 1, it follows that

$$\frac{1}{\sqrt{T}} \sum_{t_0} \hat{x}_t \tilde{u}_t \sim_a N \left( 0, Q_{XZ} Q_{ZZ}^{-1} S_{(i,i)} Q_{ZZ}^{-1} Q_{ZX} \right) \quad (4.37)$$

where $S_{(i,i)}$ denote the asymptotic variance of $\frac{1}{\sqrt{T}} \sum_{t_0} z_t \tilde{u}_t$. 

48
Combining the results in (4.35) and (4.37), we obtain
\[
\sqrt{T} (\hat{\beta}_i - \beta_0^i) \Rightarrow \left( \frac{1}{T} \sum_{i_0} \bar{X}_i \bar{X}_i' \right)^{-1} \frac{1}{\sqrt{T}} \sum_{i_0} \bar{X}_i \bar{u}_t \tag{4.38}
\]
\[
\sim_a N \left( 0, V^{(i,i)}_\beta \right) \tag{4.39}
\]
where \( V^{(i,i)}_\beta = (\lambda_i^0 - \lambda_{i-1}^0)^{-2} [Q_{xz} Q_{zz}^{-1} Q_{zx}]^{-1} Q_{xz} Q_{zz}^{-1} S_{(i,i)} Q_{zz}^{-1} Q_{zx} \) [\( Q_{xz} Q_{zz}^{-1} Q_{zx} \)]^{-1}.

Now, consider the covariance matrix of coefficient vectors between different regimes such as regime \( i \) and regime \( j \). Using (4.35)-(4.36), we have
\[
\begin{align*}
Cov(\sqrt{T} (\hat{\beta}_i - \beta_0^i), \sqrt{T} (\hat{\beta}_j - \beta_0^j)) & \Rightarrow Cov( (\lambda_i^0 - \lambda_{i-1}^0)^{-1} [Q_{xz} Q_{zz}^{-1} Q_{zx}]^{-1} Q_{xz} Q_{zz}^{-1} ) \\
\times T^{-1/2} \sum_{i_0} z_i \bar{u}_t, (\lambda_j^0 - \lambda_{j-1}^0)^{-1} [Q_{xz} Q_{zz}^{-1} Q_{zx}]^{-1} Q_{xz} Q_{zz}^{-1} T^{-1/2} \sum_{j_0} z_j \bar{u}_t ) \\
& = (\lambda_i^0 - \lambda_{i-1}^0)^{-1} [Q_{xz} Q_{zz}^{-1} Q_{zx}]^{-1} Q_{xz} Q_{zz}^{-1} Cov(T^{-1/2} \sum_{i_0} z_i \bar{u}_t, T^{-1/2} \sum_{j_0} z_j \bar{u}_t ) \\
\times (\lambda_j^0 - \lambda_{j-1}^0)^{-1} Q_{zz}^{-1} Q_{zx} [Q_{xz} Q_{zz}^{-1} Q_{zx}]^{-1} \\
& = [(\lambda_i^0 - \lambda_{i-1}^0)(\lambda_j^0 - \lambda_{j-1}^0)]^{-1} [Q_{xz} Q_{zz}^{-1} Q_{zx}]^{-1} Q_{xz} Q_{zz}^{-1} \times S_{(i,j)} Q_{zz}^{-1} Q_{zx} \\
& \times Q_{xz} Q_{zz}^{-1} Q_{zx}^{-1} 
\end{align*}
\]
where \( S_{(i,j)} \) denotes the asymptotic covariance of \( \frac{1}{\sqrt{T}} \sum_{i_0} z_i \bar{u}_t \) and \( \frac{1}{\sqrt{T}} \sum_{j_0} z_j \bar{u}_t \).

Therefore, combining (4.39) and (4.40) we obtain
\[
T^{1/2} (\hat{\beta} - \beta_0^i) \overset{p}{\rightarrow} N (0, V_\beta)
\]
with where
\[
V_\beta = \begin{pmatrix}
V^{(1,1)}_\beta & \cdots & V^{(1,m+1)}_\beta \\
\vdots & \ddots & \vdots \\
V^{(m+1,1)}_\beta & \cdots & V^{(m+1,m+1)}_\beta
\end{pmatrix}
\]
\[
V^{(i,i)}_\beta = (\lambda_i^0 - \lambda_{i-1}^0)^{-2} [Q_{xz} Q_{zz}^{-1} Q_{zx}]^{-1} Q_{xz} Q_{zz}^{-1} S_{(i,i)} Q_{zz}^{-1} Q_{zx} (Q_{xz} Q_{zz}^{-1} Q_{zx})^{-1}
\]
\[
V^{(i,j)}_\beta = [(\lambda_i^0 - \lambda_{i-1}^0)(\lambda_j^0 - \lambda_{j-1}^0)]^{-1} (Q_{xz} Q_{zz}^{-1} Q_{zx})^{-1} Q_{xz} Q_{zz}^{-1} S_{(i,j)} Q_{zz}^{-1} Q_{zx} \\
\times (Q_{xz} Q_{zz}^{-1} Q_{zx})^{-1} \quad \text{for} \ i \neq j
\]

49
Finally, we derive the form of the asymptotic variance and covariance matrices of $T^{-1/2} \sum_{i0} z_t \tilde{u}_t$ and $T^{-1/2} \sum_{j0} z_t \tilde{u}_t$. We begin with the asymptotic variance matrix of $T^{-1/2} \sum_{i0} z_t \tilde{u}_t$, which is denoted by $S_{(i,i)}$. Note that $T^{-1/2} \sum_{i0} z_t \tilde{u}_t$ can be shown to be a function of $\sum_{i0} z_t u_t, \sum_{i0} z_t v_t'$ and $Z'V$. From (3.15) we obtain

\begin{equation}
T^{-1/2} \sum_{i0} z_t \tilde{u}_t = T^{-1/2} \sum_{i0} z_t (u_t + v_t' \beta_i^0 - z_t'(Z'Z)^{-1}Z'V \beta_i^0)
\end{equation}

\begin{equation}
= T^{-1/2} \sum_{i0} z_t u_t + T^{-1/2} \sum_{i0} z_t v_t' \beta_i^0 - T^{-1/2} \sum_{i0} z_t z_t'(Z'Z)^{-1}Z'V \beta_i^0
\end{equation}

(4.41)

So, in order to get the variance matrix of $T^{-1/2} \sum_{i0} z_t \tilde{u}_t$, we need to know the variance and covariance matrices of the terms in (4.41). Notice that under Assumption 1

\begin{equation}
T^{-1/2} \left[ \sum_{t=1}^{T} \begin{pmatrix} u_t \\ v_t \end{pmatrix} \right] \otimes z_t \implies V^{1/2}B(r)
\end{equation}

(4.42)

where

\begin{equation}
V = \begin{pmatrix} V_{1,1} & V_{1,2} \\ V_{2,1} & V_{2,2} \end{pmatrix}
\end{equation}

To simplify the exposition in the following derivations, define

$A_i \equiv T^{-1/2} \sum_{i0} z_t u_t$

$B_i \equiv T^{-1/2} \sum_{i0} z_t v_t' \beta_i^0$

$C_i \equiv T^{-1/2} \sum_{i0} z_t z_t'(Z'Z)^{-1}Z'V \beta_i^0$

and

$A_j \equiv T^{-1/2} \sum_{j0} z_t u_t$
\[ B_j \equiv T^{-1/2} \sum_{j_0} z_t \upsilon_j^0 \beta_j^0 \]

\[ C_j \equiv T^{-1/2} \sum_{j_0} z_t z'_t (Z'Z)^{-1} Z' \beta_j^0 \]

From (4.42), it is clear that the variance matrices are:

\[
\begin{align*}
\text{Var}[A_i] &= (\lambda_i^0 - \lambda_{i-1}^0) V_{1,1} \\
\text{Var}[B_i] &= \text{Var}[T^{-1/2} \sum_{i_0} (\beta_i^{0'} \otimes I_q)(v_t \otimes z_t)] \\
&= (\lambda_i^0 - \lambda_{i-1}^0)(\beta_i^{0'} \otimes I_q)V_{2,2}(\beta_i^{0'} \otimes I_q) \tag{4.44} \\
\text{Var}[C_i] &= \text{Var}[T^{-1} \sum_{i_0} z_t z'_t (T^{-1} Z'Z)^{-1} T^{-1/2} \sum_{t=1}^T (\beta_i^{0'} \otimes I_q)(v_t \otimes z_t)] \\
&= (\lambda_i^0 - \lambda_{i-1}^0)^2(\beta_i^{0'} \otimes I_q)V_{2,2}(\beta_i^{0'} \otimes I_q) \tag{4.45}
\end{align*}
\]

and the covariance matrices are:

\[
\begin{align*}
\text{Cov}[A_i, B_i] &= \text{Cov}[A_i, T^{-1/2} \sum_{i_0} (\beta_i^{0'} \otimes I_q)(v_t \otimes z_t)] \\
&= (\lambda_i^0 - \lambda_{i-1}^0)V_{1,2}(\beta_i^{0'} \otimes I_q) \tag{4.46} \\
\text{Cov}[A_i, C_i] &= (\lambda_i^0 - \lambda_{i-1}^0)^2V_{1,2}(\beta_i^{0'} \otimes I_q) \tag{4.47} \\
\text{Cov}[B_i, C_i] &= \text{Cov}[T^{-1/2} \sum_{i_0} (\beta_i^{0'} \otimes I_q)(v_t \otimes z_t), T^{-1} \sum_{i_0} z_t z'_t (T^{-1} Z'Z)^{-1} T^{-1/2} \sum_{t=1}^T (\beta_i^{0'} \otimes I_q)(v_t \otimes z_t)] \\
&= (\lambda_i^0 - \lambda_{i-1}^0)^2(\beta_i^{0'} \otimes I_q)V_{2,2}(\beta_i^{0'} \otimes I_q) \tag{4.48}
\end{align*}
\]

Now, combining (4.43)-(4.48), we obtain the asymptotic variance of \( T^{-1/2} \sum_{i_0} z_t \tilde{u}_t \)

\[
S_{(i,i)} = \text{Var}[T^{-1/2} \sum_{i_0} z_t \tilde{u}_t] \\
= (\lambda_i^0 - \lambda_{i-1}^0)\{V_{1,1} + (\beta_i^{0'} \otimes I_q)V_{2,2}(\beta_i^{0'} \otimes I_q) + (\lambda_i^0 - \lambda_{i-1}^0)(\beta_i^{0'} \otimes I_q)V_{2,2} \\
\times (\beta_i^{0'} \otimes I_q) + V_{1,2}(\beta_i^{0'} \otimes I_q) + (\beta_i^{0'} \otimes I_q)V_{2,1} - (\lambda_i^0 - \lambda_{i-1}^0)V_{1,2}(\beta_i^{0'} \otimes I_q) \\
- (\lambda_i^0 - \lambda_{i-1}^0)(\beta_i^{0'} \otimes I_q)V_{2,1} - 2(\lambda_i^0 - \lambda_{i-1}^0)(\beta_i^{0'} \otimes I_q)V_{2,2}(\beta_i^{0'} \otimes I_q)\} \\
= (\lambda_i^0 - \lambda_{i-1}^0)\{V_{1,1} + (\beta_i^{0'} \otimes I_q)V_{2,2}(\beta_i^{0'} \otimes I_q) - (\lambda_i^0 - \lambda_{i-1}^0)(\beta_i^{0'} \otimes I_q)V_{2,2} \}
\]

51
To derive asymptotic covariance matrix between the sums over different regimes, we have

\[ \times (\beta_i^0 \otimes I_q) + V_{1,2}(\beta_i^0 \otimes I_q) + (\beta_i^0 \otimes I_q) V_{2,1} - (\lambda_i^0 - \lambda_{i-1}^0) V_{1,2} (\beta_i^0 \otimes I_q) \]

\[ - (\lambda_i^0 - \lambda_{i-1}^0) (\beta_i^0 \otimes I_q) V_{2,1} \}

\[ = (\lambda_i^0 - \lambda_{i-1}^0) \{ V_{1,1} + (1 - \lambda_i^0 + \lambda_{i-1}^0) (\beta_i^0 \otimes I_q) V_{2,2} (\beta_i^0 \otimes I_q) \]

\[ + (1 - \lambda_i^0 + \lambda_{i-1}^0) V_{1,2} (\beta_i^0 \otimes I_q) + (1 - \lambda_i^0 + \lambda_{i-1}^0) (\beta_i^0 \otimes I_q) V_{2,1} \}

\[ = (\lambda_i^0 - \lambda_{i-1}^0) \{ V_{1,1} + (1 - \lambda_i^0 + \lambda_{i-1}^0) ([\beta_i^0 \otimes I_q] V_{2,2} (\beta_i^0 \otimes I_q) \]

\[ + V_{1,2} (\beta_i^0 \otimes I_q) + (\beta_i^0 \otimes I_q) V_{2,1} \}

Now, consider the asymptotic covariance matrix of \( T^{-1/2} \sum_{i=0}^{\infty} z_i \tilde{u}_t \) and \( T^{-1/2} \sum_{j=0}^{\infty} z_j \tilde{u}_t \), which is denoted by \( S_{i,j} \). Using (4.41) we have

\[ S_{i,j} = \text{Cov}(T^{-1/2} \sum_{i=0}^{\infty} z_i \tilde{u}_t, T^{-1/2} \sum_{j=0}^{\infty} z_j \tilde{u}_t) \]

\[ = \text{Cov}(T^{-1/2} \sum_{i=0}^{\infty} z_i (u_t + v_i' \beta_i^0 - z_i' (Z'Z)^{-1} Z'V \beta_i^0), T^{-1/2} \sum_{j=0}^{\infty} z_j (u_t + v_j' \beta_j^0 - z_j' (Z'Z)^{-1} Z'V \beta_j^0) \]

\[ = \text{Cov}(A_i, A_j) + \text{Cov}(A_i, B_j) - \text{Cov}(A_i, C_j) + \text{Cov}(B_i, A_j) \]

\[ + \text{Cov}(B_i, B_j) - \text{Cov}(B_i, C_j) - \text{Cov}(C_i, A_j) - \text{Cov}(C_i, B_j) \]

\[ + \text{Cov}(C_i, C_j) \] (4.49)

To derive asymptotic covariance matrix \( S_{i,j} \), we work on each term in (4.49) in turn. Since the following covariance matrices in (4.50)-(4.53) involve the covariance between the sums over different regimes, we have

\[ \text{Cov}(A_i, A_j) = 0 \] (4.50)

\[ \text{Cov}(A_i, B_j) = 0 \] (4.51)

\[ \text{Cov}(B_i, A_j) = 0 \] (4.52)

\[ \text{Cov}(B_i, B_j) = 0 \] (4.53)
As in (4.42)-(4.48) the other terms can be shown to be:

\[ \text{Cov}(A_i, C_j) = \text{Cov}(A_i, (\lambda_i^0 - \lambda_{i-1}^0)[T(\lambda_j^0 - \lambda_{j-1}^0)]^{-1}\sum_{jo} z_{io}z'_{io}(T^{-1}Z'Z)^{-1} \times T^{-1/2}Z'V(\beta_j^0) \]

\[ = (\lambda_i^0 - \lambda_{i-1}^0)\text{Cov}(A_i, T^{-1/2}Z'V(\beta_j^0)) \]

\[ = (\lambda_i^0 - \lambda_{i-1}^0)\text{Cov}(A_i, T^{-1/2}\sum_{io} z_{io}z'_{io}(\beta_j^0)) \]

\[ = (\lambda_i^0 - \lambda_{i-1}^0)(\lambda_j^0 - \lambda_{j-1}^0)V_{i,2}(\beta_j^0 \otimes I_q) \] \hspace{1cm} (4.54)

\[ \text{Cov}(B_i, C_j) = (\lambda_i^0 - \lambda_{i-1}^0)(\lambda_j^0 - \lambda_{j-1}^0)(\beta_i^{0r} \otimes I_q)V_{2,2}(\beta_j^0 \otimes I_q) \] \hspace{1cm} (4.55)

\[ \text{Cov}(C_i, A_j) = (\lambda_i^0 - \lambda_{i-1}^0)(\lambda_j^0 - \lambda_{j-1}^0)(\beta_i^{0r} \otimes I_q)V_{2,1} \] \hspace{1cm} (4.56)

\[ \text{Cov}(C_i, B_j) = (\lambda_i^0 - \lambda_{i-1}^0)(\lambda_j^0 - \lambda_{j-1}^0)(\beta_i^{0r} \otimes I_q)V_{2,2}(\beta_j^0 \otimes I_q) \] \hspace{1cm} (4.57)

\[ \text{Cov}(C_i, C_j) = (\lambda_i^0 - \lambda_{i-1}^0)(\lambda_j^0 - \lambda_{j-1}^0)(\beta_i^{0r} \otimes I_q)V_{2,2}(\beta_j^0 \otimes I_q) \] \hspace{1cm} (4.58)

Now, combining (4.50)-(4.58), we obtain

\[ S_{(i,j)} = -((\lambda_i^0 - \lambda_{i-1}^0)(\lambda_j^0 - \lambda_{j-1}^0)V_{i,2}(\beta_j^0 \otimes I_q) - (\lambda_i^0 - \lambda_{i-1}^0)(\lambda_j^0 - \lambda_{j-1}^0)(\beta_i^{0r} \otimes I_q) \times V_{2,2}(\beta_j^0 \otimes I_q) - ((\lambda_i^0 - \lambda_{i-1}^0)(\lambda_j^0 - \lambda_{j-1}^0)(\beta_i^{0r} \otimes I_q)V_{2,1}) \]

\[ \times V_{2,2}(\beta_j^0 \otimes I_q)] \]
Chapter 5

Asymptotic Distribution of the Break Point Estimators

Given the rate of convergence, it is natural to consider the limiting distribution of the break point estimators as the following step. For the least squares method Bai (1997) derived the limiting distribution of the break point estimator with a single break point imposed. Bai (1997) considered two frameworks of asymptotic distribution. One is based on the fixed magnitude of shift in regression parameters, the other on a shrinking magnitude of shift.

In general, the limiting distribution of the break point estimators obtained specifying fixed magnitude of shifts in the regression coefficients depends on the exact distribution of regressors and error term, \( \{x_t, u_t\} \). That is, it is highly data dependent and the analytic solution for the limiting distribution is typically hard to obtain in this case.

An alternative asymptotic theory is to consider small shifts in the regression coefficients, assuming the magnitude of shifts converges to zero as the sample size
increases. In this setup, the limiting distribution is invariant to the underlying distribution of \( \{x_t, u_t\} \) and the resulting distribution can be used as an approximation even for moderate shifts.

For fixed magnitude of shift case, Bai(1997) imposed a strong restriction on the joint distribution of \( \{x_t, u_t\} \) by assuming strict stationarity. Even though this additional assumption may restrict the applicability of the model a lot, it makes it easy to derive the limiting distribution of the break points at least. Also, the results may lay the foundation for the case of an alternative shrinking magnitude of shift.

For the shrinking magnitude of shift case, second order stationarity of \( \{x_t, z_t\} \) within each regime has been assumed. Relying on a functional central limit theorem, Bai(1997) proved that the break point estimator has a functional of two independent Brownian motion processes as its limiting distribution. Based on the resulting asymptotic distribution of the break point estimator, a confidence interval can be constructed easily.

In this chapter, we adopt Bai(1997)’s approach to our 2SLS setup to derive the limiting distribution of the break point estimator. In section 5.1, we consider the case of a fixed magnitude shift in the regression parameters. A quite restrictive assumption of strict stationarity in the regressors and errors in our 2SLS regression model is added to our set of assumptions. In section 5.2, we consider the shrinking magnitude of shift case and relax the joint distribution assumption on the regressors from strict stationary to weak stationary within each regime. We obtain the convergence rate of break point estimators. Based on this convergence rate, we derive the limiting distribution for small shifts. This limiting distribution
can then be used as an approximation to the underlying distribution for moderate
shifts, or even large shifts.

With assumptions of strict stationarity or weak stationarity on the regressors
and errors within each regime, each segment is asymptotically distinct and the
analysis of the limiting distribution of the break dates is similar to that in the
single break model. Thus, for the following derivation of the limiting distributions
we just focus on the single break case.

5.1 Fixed Magnitude of Shift in the Regression
Parameters

Consider the following second stage linear regression model with a single break
point at $k_0$.

$$y_t = \hat{x}'_t \beta^0_1 + \tilde{u}_t, \quad i = 1, 2 \quad t = T^0_{i-1} + 1, \ldots, T^0_i$$

That is, by denoting $T^0_i = k_0$ we can write our model for each regime as

$$y_t = \hat{x}'_t \beta^0_1 + \tilde{u}_t, \quad t = 1, \ldots, k_0 \quad (5.1)$$

$$y_t = \hat{x}'_t \beta^0_2 + \tilde{u}_t, \quad t = k_0 + 1, \ldots, T \quad (5.2)$$

For the derivation of the limiting distribution of the break point estimator $\hat{k}$, we
reparametrize our regression model for the second regime. Equation (5.2) can be
rewritten as

$$y_t = \hat{x}'_t \beta^0_1 + \hat{x}'_t (\beta^0_2 - \beta^0_1) + \tilde{u}_t$$

$$= \hat{x}'_t \beta^0_1 + \hat{x}'_t (\beta^0_2 - \beta^0_1) + \tilde{u}_t \quad (5.3)$$
To present the model in a matrix format, we define \( Y = (y_1, \ldots, y_T)' \), \( W = (\hat{x}_1, \ldots, \hat{x}_T)' \), \( W_1 = (\hat{x}_1, \ldots, \hat{x}_k, 0, \ldots, 0)' \), \( W_2 = (0, \ldots, 0, \hat{x}_{k+1}, \ldots, \hat{x}_T)' \) and \( W_0 = (0, \ldots, 0, \hat{x}_{k_0+1}, \ldots, \hat{x}_T)' \). Then, by combining (5.1) and (5.3), our transformed regression model can be written in matrix form as

\[
Y = W \beta_0 + W_0 \theta^0 + \tilde{U}
\]

where \( \theta^0 = \beta_2^0 - \beta_1^0 \)

Least squares is used to estimate the model. Let \( S_T(k) \) denote the sum of squared residuals when regressing \( Y \) on \( W \) and \( W_2 \). The break point estimator \( \hat{k} \) is defined as

\[
\hat{k} = \min_{1 \leq k \leq T} S_T(k)
\]

Bai(1993) showed that (5.4) is equivalent to

\[
\hat{k} = \max_{1 \leq k \leq T} \xi_W(k)
\]

where \( \xi_W \) is Wald statistic which is defined as

\[
\xi_W(k) = \frac{\hat{\theta}(k)'(W_W M_W W_2)\hat{\theta}(k)}{\hat{\sigma}^2(k)}
\]

with \( M_W = I - W(W'W)^{-1}W' \), \( \hat{\sigma}^2(k) = S_T(k)/(T - 2p) \), and \([\hat{\beta}_i(k), \hat{\theta}(k)]\) being the least squares estimator by regressing \( Y \) on \( W \) and \( W_2 \). Note that \( k \in [\pi T, (1 - \pi)T] \) is needed to insure the existence of least square estimators where \( \pi \) is a small positive number.

To show (5.5), let \( \bar{S} \) denote the sum of squared residuals by regressing \( Y \) on \( W \) alone. Then we obtain the identity

\[
\bar{S} - S_T(k) = \hat{\theta}(k)'(W_W M_W W_2)\hat{\theta}(k)
\]

57
Following Bai(1993)’s argument, it can be easily seen that the Wald test statistic is a monotonic transformation of $S_T(k)$. The Wald statistic can be written as

$$\xi_W(k) = \left( \frac{T - 2p}{p} \right) \left( \frac{\bar{S} - S_T(k)}{S_T(k)} \right)$$

Because $\bar{S}$ does not depend on $k$ and the Wald statistic is a strictly decreasing transformation of $S_T(k)$, it follows immediately that

$$\hat{k} = \arg \min_k S_T(k) = \arg \max_k \xi_W(k)$$

or using (5.6)

$$\hat{k} = \arg \max_k \hat{\theta}(k)'(W_2'M_WW_2)\hat{\theta}(k)$$


Let $V_T(k) = \hat{\theta}(k)'(W_2'M_WW_2)\hat{\theta}(k)$, then $\hat{k} = \arg \max_k V_T(k) = \arg \max_k [V_T(k) - V_T(k_0)]$. The LS estimator $\hat{\theta}(k)$ can be expressed as

$$\hat{\theta}(k) = \left( W_2'M_WW_2 \right)^{-1}W_2'M_WY$$

$$= \left( W_2'M_WW_2 \right)^{-1}W_2'M_W \left[ W_0' \beta_0 \right. + W_0' \theta_0 + \bar{U} \left. \right]$$

$$= \left( W_2'M_WW_2 \right)^{-1}W_2'M_WW_2' \beta_0^0 + \left( W_2'M_WW_2 \right)^{-1}W_2'M_WW_0\theta_0$$

$$+ \left( W_2'M_WW_2 \right)^{-1}W_2'M_W\bar{U}$$

$$= \left( W_2'M_WW_2 \right)^{-1}W_2'M_WW_0\theta_0 + \left( W_2'M_WW_2 \right)^{-1}W_2'M_W\hat{U}$$

Similarly, we have

$$\hat{\theta}(k_0) = \left( W_0'M_WW_0 \right)^{-1}W_0'M_WY$$
Thus, it follows that

\[ V_T(k) = \hat{\theta}(k)'(W'_2M_WW_2)\hat{\theta}(k) \]
\[ = ((W'_0M_WM_0)^{-1}W'_0M_WM_0[\theta_0^0 + W_0\theta_0^0 + \tilde{U}]) \]
\[ = (W'_0M_WM_0)^{-1}W'_0M_WM_0\theta_0^0 + (W'_0M_WM_0)^{-1}W'_0M_WM_\tilde{U} \]
\[ = \theta_0^0 + (W'_0M_WM_0)^{-1}W'_0M_WM_\tilde{U}. \]

Thus, it follows that

\[ V_T(k) = \hat{\theta}(k)'(W'_2M_WM_2)\hat{\theta}(k) \]
\[ = [((W'_2M_WM_2))^{-1}W'_2M_WM_2\theta_0^0 + (W'_2M_WM_2)^{-1}W'_2M_WM_\tilde{U}]'(W'_2M_WM_2) \]
\[ \times [((W'_2M_WM_2))^{-1}W'_2M_WM_2\theta_0^0 + (W'_2M_WM_2)^{-1}W'_2M_WM_\tilde{U}] \]
\[ = \theta^0(\theta_0^0 + (W'_2M_WM_2)^{-1}W'_2M_WM_\tilde{U}) \]
\[ + \theta^0(\theta_0^0 + (W'_2M_WM_2)^{-1}W'_2M_WM_\tilde{U}) \]
\[ + \tilde{U}'M_WM_2(W'_2M_WM_2)^{-1}(W'_2M_WM_2)(W'_2M_WM_2)^{-1}W'_2M_WM_\tilde{U} \]
\[ + \tilde{U}'M_WM_2(W'_2M_WM_2)^{-1}(W'_2M_WM_2)(W'_2M_WM_2)^{-1}W'_2M_WM_\tilde{U} \]
\[ = \theta^0(\theta_0^0 + (W'_2M_WM_2)^{-1}W'_2M_WM_\tilde{U}) \]
\[ \times (W'_2M_WM_2)^{-1}(W'_2M_WM_\tilde{U}) + (\tilde{U}'M_WM_2)(W'_2M_WM_2)^{-1}W'_2M_WM_\tilde{U} \]

and

\[ V_T(k_0) = \theta^0(\theta_0^0 + (W'_2M_WM_2)^{-1}W'_2M_WM_\tilde{U}) \]
\[ + 2\theta^0(\theta_0^0 + (W'_2M_WM_2)^{-1}W'_2M_WM_\tilde{U}) \]
\[ \times (W'_2M_WM_2)^{-1}(W'_2M_WM_\tilde{U}) + (\tilde{U}'M_WM_2)(W'_2M_WM_2)^{-1}W'_2M_WM_\tilde{U} \]
\[ = \theta^0(\theta_0^0 + (W'_2M_WM_2)^{-1}W'_2M_WM_\tilde{U}) + (\tilde{U}'M_WM_2)(W'_2M_WM_2)^{-1}W'_2M_WM_\tilde{U} \]

Therefore, we have

\[ V_T(k) - V_T(k_0) = \theta^0[2\theta^0(\theta_0^0 + (W'_2M_WM_2)^{-1}W'_2M_WM_\tilde{U}) + (\tilde{U}'M_WM_2)(W'_2M_WM_2)^{-1}W'_2M_WM_\tilde{U} - 2\theta^0(\theta_0^0 + (W'_2M_WM_2)^{-1}W'_2M_WM_\tilde{U})] \]
\[
\begin{align*}
+ \hat{U}'M_{W_2}(W_{2}'M_{W_2})^{-1}W_{2}'M_{W_2} \hat{U} - \hat{U}'M_{W_0}(W_{0}'M_{W_0})^{-1} \\
\times W_{0}'M_{W_0} \hat{U}
\end{align*}
\]

\[
= \theta^0 \left[ W_{0}'M_{W_2}(W_{2}'M_{W_2})^{-1}W_{2}'M_{W_2}W_0 - W_{0}'M_{W_0} \theta^0 + H_T(k) \right]
\]

\[
\text{where } H_T(k) = 2\theta^0 (W_{0}'M_{W_2})(W_{2}'M_{W_2})^{-1}W_{2}'M_{W_2} \hat{U} - 2\theta^0 (W_{0}'M_{W_0}) + \hat{U}'M_{W_2}
\]

\[
(W_{2}'M_{W_2})^{-1}W_{2}'M_{W_2} \hat{U} - \hat{U}'M_{W_0}(W_{0}'M_{W_0})^{-1}W_{0}'M_{W_0} \hat{U}
\]

Define for \( k \neq k_0 \)

\[
G_T(k) = \frac{\theta^0 \left[ W_{0}'M_{W_0} - W_{0}'M_{W_2}(W_{2}'M_{W_2})^{-1}W_{2}'M_{W_0} \theta^0 \right]}{|k_0 - k|}
\]

For \( k = k_0 \), define \( G_T(k_0) = \theta^0 \theta^0 \). Then, we have the following identity

\[
V_T(k) - V_T(k_0) = -|k_0 - k|G_T(k) + H_T(k) \quad \text{for all } k \quad (5.7)
\]

In Lemma 3, we establish the following two convergence results

\[
|k_0 - k|G_T(k) = \theta^0 W_\Delta \theta^0 + o_p(1)
\]

\[
H_T(k) = 2\theta^0 W_\Delta \hat{U} \cdot \Xi + o_p(1)
\]

to conclude that

\[
V_T(k) - V_T(k_0) = -\theta^0 W_\Delta \theta^0 + 2\theta^0 W_\Delta \hat{U} \cdot \Xi + o_p(1)
\]

\( W_\Delta \) is defined as

\[
W_\Delta = W_2 - W_0 = (0, \cdots, 0, \hat{x}_{k_0+1}, \cdots, \hat{x}_{k_0}, 0, \cdots, 0)' \quad \text{for } k < k_0
\]

\[
W_\Delta = -(W_2 - W_0) = (0, \cdots, 0, \hat{x}_{k_0+1}, \cdots, \hat{x}_k, 0, \cdots, 0)' \quad \text{for } k > k_0
\]

\[
W_\Delta = 0 \quad \text{for } k = k_0
\]

and \( \Xi \) is defined as

\[
\Xi = 1 \quad \text{for } k_0 > k
\]

\[
\Xi = -1 \quad \text{for } k_0 < k
\]
Thus, it follows that

\[ W_2 = W_0 + W_\Delta \cdot \Xi \quad (5.8) \]

Lemma 3 plays the key role in the following Theorem 4 which derives the limiting distribution of the break point estimator under the assumption of fixed magnitude of shift in regression parameters.

**Lemma 3** \textit{Under Assumptions 1\textasciitilde 5,}

\[ V_T(k) - V_T(k_0) = -\theta^\prime W_\Delta W_\Delta \theta^0 + 2\theta^\prime W_\Delta \hat{U} \cdot \Xi + o_p(1) \]

\textbf{Proof)

Before we investigate the convergence of \(|k_0 - k|G_T(k)|\), we recall that \( \Xi = \text{sgn}(k_0 - k) \).

\[ |k_0 - k|G_T(k) = \theta^\prime [W'_0 M_W W_0 - W'_0 M_W W_2(W'_2 M_W W_2)^{-1} W'_2 M_W W_0] \theta^0 \]

\[ = \theta^\prime [(W_2 - \Xi W_\Delta)^\prime M_W (W_2 - \Xi W_\Delta) - (W_2 - \Xi W_\Delta)^\prime M_W W_2 \times (W'_2 M_W W_2)^{-1} W'_2 M_W (W_2 - \Xi W_\Delta)] \theta^0 \]

\[ = \theta^\prime [W'_2 M_W W_2 - W'_2 W_\Delta \Xi - \Xi W'_\Delta M_W W_2 + \Xi W'_\Delta M_W W_\Delta \Xi \]

\[ - (W'_2 M_W W_2 - \Xi W'_\Delta M_W W_2)(W'_2 M_W W_2)^{-1}(W'_2 M_W W_2 \]

\[ - W'_2 M_W W_\Delta \Xi)] \theta^0 \]

\[ = \theta^\prime [W'_2 M_W W_2 - W'_2 M_W W_\Delta \Xi - \Xi W'_\Delta M_W W_2 + \Xi W'_\Delta M_W W_\Delta \Xi \]

\[ - W'_2 M_W W_2 + W'_2 M_W W_\Delta \Xi + \Xi W'_\Delta M_W W_2 - \Xi W'_\Delta M_W W_2 \times (W'_2 M_W W_2)^{-1} W'_2 M_W W_\Delta \Xi] \theta^0 \]

\[ = \theta^\prime [W'_\Delta M_W W_\Delta - W'_\Delta M_W W_2(W'_2 M_W W_2)^{-1} W'_2 M_W W_\Delta] \theta^0 \]

\[ (5.9) \]
Now, we investigate (5.9) term by term. First consider the second term of (5.9).

Define \( D(C) = \{ k : |k - k_0| \leq C \} \). In Theorem 2 we have already proved \( P(|\hat{k} - k_0| > C) < \eta \) for every \( \eta > 0 \). Thus it is sufficient to investigate the behavior of \( V_T(k) - V_T(k_0) \) over \( D(C) \) for the establishment of the limiting distribution of the break point estimators. Over the set \( D(C) \), \( W'_\Delta M_W W_2 \) consists of a sum of finite terms. Thus, it is clear that \( ||W'_\Delta M_W W_2|| = |k_0 - k|O_p(1) = O_p(1) \).

Since \( ||W'_2 M_W W_2|| = O_p(T) \), the second term of the equation (5.9) is bounded by \( O_p(1) \cdot O_p(T^{-1}) \cdot O_p(1) = o_p(1) \).

Next, consider the first term \( W'_\Delta M_W W_\Delta = W'_\Delta W_\Delta + W'_\Delta W(W'W)^{-1}W W_\Delta \). We know that \( W'_\Delta W(W'W)^{-1}W W_\Delta = |k_0 - k|O_p(1) \cdot O_p(T^{-1}) \cdot |k_0 - k|O_p(1) = O_p(1) \cdot O_p(T^{-1}) \cdot O_p(1) = o_p(1) \).

Thus, combining the results on two terms, we get

\[ |k_0 - k|G_T(k) = \theta^\theta' W'_\Delta W_\Delta \theta^\theta' + o_p(1) \]

Next, consider \( H_T(k) \).

\[
H_T(k) = 2\theta^\theta'(W'_0 M_W W_2)(W'_2 M_W W_2)^{-1}W'_2 M_W \tilde{U} - 2\theta^\theta'(W'_0 M_W \tilde{U}) \\
+ \tilde{U}' M_W W_2(W'_2 M_W W_2)^{-1}W'_2 M_W \tilde{U} - \tilde{U}' M_W W_0(W'_0 M_W W_0)^{-1}W'_0 M_W \tilde{U} \\
= (A.1) + (A.2)
\]

(5.10)

where we defined

\[
(A.1) = 2\theta^\theta'(W'_0 M_W W_2)(W'_2 M_W W_2)^{-1}W'_2 M_W \tilde{U} - 2\theta^\theta'(W'_0 M_W \tilde{U}) \\
(A.2) = \tilde{U}' M_W W_2(W'_2 M_W W_2)^{-1}W'_2 M_W \tilde{U} - \tilde{U}' M_W W_0(W'_0 M_W W_0)^{-1}W'_0 M_W \tilde{U}
\]

In the following derivation, we show \( (A.2) = o_p(1) \). Thus the limit behavior of
$H_T(k)$ over $D(C)$ is dominated by the limit behavior of (A.1).

First, consider (A.1).

\[(A.1) = 2\theta^\alpha (W_2 - W_\Delta \Xi)' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \hat{U} - 2\theta^\alpha W_0' M_W \hat{U} \]

= \[2\theta^\alpha (W_2' M_W \hat{U} - \Xi W_\Delta' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \hat{U}) - 2\theta^\alpha W_0' M_W \hat{U} \]

= \[2\theta^\alpha W_2' M_W \hat{U} - 2\Xi \theta^\alpha W_\Delta' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \hat{U} - 2\theta^\alpha W_0' M_W \hat{U} \]

= \[2\theta^\alpha (W_2 - W_0)' M_W \hat{U} - 2\Xi \theta^\alpha W_\Delta' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \hat{U} \]

= \[\Xi [2\theta^\alpha W_\Delta' M_W \hat{U} - 2\theta^\alpha W_\Delta' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \hat{U}] \]

(5.11)

Let’s investigate the convergence result of each term of (5.11) in turn. Noticing that $W_\Delta W = O_p(1)$ over $D(C)$ and $(W' W)^{-1} W' \hat{U} = (1/\sqrt{T})(W' W/T)^{-1} W' \hat{U}/\sqrt{T} = T^{-1/2} O_p(1)$, the first term in (5.11) can be written as

\[W_\Delta' M_W \hat{U} = W_\Delta' \hat{U} - W_\Delta W (W' W)^{-1} W' \hat{U} \]

= \[W_\Delta' \hat{U} - O_p(1) \cdot T^{-1/2} O_p(1) \]

= \[W_\Delta' \hat{U} + o_p(1) \quad \text{over } D(C) \]

Similarly, we observe that

\[W_\Delta' M_W W_2 = W_\Delta' W_2 - W_\Delta' W (W' W)^{-1} W' W_2 \]

= \[O_p(1) - O_p(1) \cdot O_p(T^{-1}) \cdot O_p(T) \]

= \[O_p(1) \quad \text{over } D(C) \]

\[(W_2' M_W W_2)^{-1} W_2' M_W \hat{U} = O_p(T^{-1}) \cdot O_p(T^{1/2}) = O_p(T^{-1/2}) \]

Thus, the second term in (5.11) can be written as

\[W_\Delta' M_W W_2 \cdot (W_2' M_W W_2)^{-1} W_2' M_W \hat{U} = O_p(1) \cdot O_p(T^{-1/2}) = o_p(1) \quad \text{over } D(C) \]
Thus, combining the results on the two terms in (5.11), we have

\[(A.1) \quad \Xi[2\theta^0W_\Delta^rM_WM_\Delta\tilde{U} - 2\theta^0W_\Delta^rM_WW_2(W_2'M_WW_2)^{-1}W_2'M_WM_\Delta\tilde{U}] = 2\Xi\theta^0W_\Delta^r\tilde{U} + o_p(1)\]

Now, we prove that \((A.2) = o_p(1)\).

\[(A.2) = \tilde{U}'M_WW_2(W_2'M_WM_\Delta\tilde{U}\tilde{U}'\tilde{M}_W - \tilde{U}'M_WW_0(W_0'M_WM_\Delta\tilde{U})^{-1}W_0'M_WM_\Delta\tilde{U} - \tilde{U}'M_W(W_2 - W_\Delta\Xi)\times[(W_2 - W_\Delta\Xi)'M_WM_\Delta\Xi]^{-1}(W_2 - W_\Delta\Xi)'M_WM_\Delta\tilde{U} = (M) - (N)\]

where for notational convenience \((M)\) and \((N)\) are defined as

\[(M) \equiv \tilde{U}'M_WW_2(W_2'M_WM_\Delta\tilde{U}\tilde{M}_W - \tilde{U}'M_WW_0(W_0'M_WM_\Delta\tilde{U})^{-1}W_0'M_WM_\Delta\tilde{U}\]

\[(N) \equiv \tilde{U}'M_W(W_2 - W_\Delta\Xi)[(W_2 - W_\Delta\Xi)'M_WM_\Delta\Xi]^{-1}(W_2 - W_\Delta\Xi)'M_WM_\Delta\tilde{U}\]

To investigate the limiting behavior of \((A.2)\) over the set \(D(C)\), it is helpful to check the limit behavior of \((N)\) first. What we want to show is that \((N)\) has the same probability limit as \((M)\). So in the limit the two terms cancel out to produce \((A.2) = o_p(1)\).

Using the relationship \(W_0 = W_2 - W_\Delta\Xi\), \((N)\) can be written as

\[(N) = \{\tilde{U}'M_W(W_2 - W_\Delta\Xi)/\sqrt{T}\}[(W_2 - W_\Delta\Xi)'M_WM_\Delta\Xi/T]^{-1}\times\{(W_2 - W_\Delta\Xi)'M_WM_\Delta\tilde{U}/\sqrt{T}\}\]

Let’s check the limit behavior of the first factor of \((N)\).

\[\tilde{U}'M_W(W_2 - W_\Delta\Xi)/\sqrt{T} = \tilde{U}'M_WW_2/\sqrt{T} - \tilde{U}'M_WM_\Delta\Xi/\sqrt{T}\]

\[= \tilde{U}'M_WW_2/\sqrt{T} + o_p(1)\]
The last equality follows from the fact that $\tilde{U}'M_W W_\triangle \Xi$ is the sum of the $|k - k_0|$ terms and the total number of the added terms is bounded over $D(C) = \{k : |k - k_0| \leq C\}$. In other words,

\[
\tilde{U}'M_W W_\triangle = \tilde{U}'(I - W(W'W)^{-1}W')W_\triangle \\
= \tilde{U}'W_\triangle - \tilde{U}'W(W'W)^{-1}W'W_\triangle \\
= O_p(1)\sqrt{|k - k_0|} - O_p(T^{1/2})O_p(T^{-1})O_p(1)|k - k_0| \\
= O_p(1) - O_p(T^{-1/2}) \\
= O_p(1) \quad \text{over } D(C)
\]

Similarly, for the second factor of $(N)$

\[
(W_2 - W_\triangle \Xi)'M_W (W_2 - W_\triangle \Xi)/T = W_2'M_W W_2/T - (1/T)\{W_2'M_W W_\triangle \Xi \\
+ \Xi W_\triangle W_2 - W_\triangle M_W W_\triangle\} \\
= W_2'M_W W_2/T + o_p(1)
\]

Thus, combining the convergence results on the two factors of $(N)$, we have

\[
(N) = \left(\tilde{U}'M_W W_2/\sqrt{T} + o_p(1)\right) (W_2'M_W W_2/T + o_p(1))^{-1} \left(W_2'M_W \tilde{U}/\sqrt{T} + o_p(1)\right)
\]

Therefore,

\[
(A.2) = (M) - (N) \\
= \tilde{U}'M_W W_2(W_2'M_W W_2)^{-1}W_2'M_W \tilde{U} - \left(\tilde{U}'M_W W_2/\sqrt{T} + o_p(1)\right) \\
\times (W_2'M_W W_2/T + o_p(1))^{-1} \left(W_2'M_W \tilde{U}/\sqrt{T} + o_p(1)\right) \\
= \left(\tilde{U}'M_W W_2/\sqrt{T}\right) (W_2'M_W W_2/T)^{-1} \left(W_2'M_W \tilde{U}/\sqrt{T}\right)
\]
\[
- \left( \dot{U}' M W W_2 / \sqrt{T} + o_p(1) \right) \left( W_2' M W W_2 / T + o_p(1) \right)^{-1} \\
\times \left( W_2' M W \ddot{U} / \sqrt{T} + o_p(1) \right) \\
= o_p(1)
\]

Thus

\[
H_T(k) = (A.1) + (A.2) \\
= \left( 2 \Xi \theta \theta' W_\Delta \ddot{U} + o_p(1) \right) + o_p(1) \\
= 2 \Xi \theta \theta' W_\Delta \ddot{U} + o_p(1)
\]

Finally,

\[
V_T(k) - V_T(k_0) = -|k_0 - k| G_T(k) + H_T(k) \\
= \left( -\theta \theta' W_\Delta \theta + o_p(1) \right) + \left( 2 \Xi \theta \theta' W_\Delta \ddot{U} + o_p(1) \right) \\
= -\theta \theta' W_\Delta \theta + 2 \Xi \theta \theta' W_\Delta \ddot{U} + o_p(1)
\]

as claimed in Lemma 3. This ends the proof of Lemma 3.

Now, we are ready to derive the limiting distribution of the break point estimators in the case of fixed magnitude of shift in parameters. Due to the analytical intractability in this case, an additional assumption is necessary to obtain the limiting distribution.

**Assumption 6** The process \( \{x_t, z_t, u_t, v_t\}_{t=-\infty}^{\infty} \) is strictly stationary.
We define a stochastic process $R^*(m)$ on the set of integers as follows:

$$R^*(m) = \begin{cases} 
R_1(m) : & m < 0 \\
0 : & m = 0 \\
R_2(m) : & m > 0 
\end{cases}$$

with

$$R_1(m) = -\theta_0' \Delta_0' \sum_{t=m+1}^{0} z_t z'_t \Delta_0 \theta^0 + 2\theta_0' \Delta_0' \left( \sum_{t=m+1}^{0} z_t u_t + \sum_{t=m+1}^{0} z_t v'_i \beta^0_2 \right)$$

for $m = -1, -2, \ldots$

$$R_2(m) = -\theta_0' \Delta_0' \sum_{t=1}^{m} z_t z'_t \Delta_0 \theta^0 - 2\theta_0' \Delta_0' \left( \sum_{t=1}^{m} z_t u_t + \sum_{t=1}^{m} z_t v'_i \beta^0_2 \right)$$

for $m = 1, 2, \ldots$

In the case of independence for $(z_t, u_t, v_t)$ over $t$, the process $R^*(m)$ is a two-sided random walk with stochastic drifts.

**Theorem 4** Under Assumptions 1∼6, and further assuming that $-(z'_t \Delta_0 \theta^0)^2 \pm 2\theta_0' \Delta_0'(z_t u_t + z_t v'_i \beta^0_2)$ where $i = 1, 2$ has a continuous distribution,

$$\hat{k} - k_0 \rightarrow_d \arg \max_m \ R^*(m)$$

**Proof**

First, we consider the case of $k < k_0$.

$$-\theta_0' W'_{\Delta} W_{\Delta} \theta^0 + 2\theta_0' W'_{\Delta} U \cdot \Xi = -\theta_0' W'_{\Delta} W_{\Delta} \theta^0 + 2\theta_0' W'_{\Delta} \tilde{U}$$

$$= -\theta_0' \sum_{t=k+1}^{k_0} w_t w'_t \theta^0 + 2\theta_0' \sum_{t=k+1}^{k_0} w_t \tilde{u}_t$$

$$= -\theta_0' \sum_{t=k+1}^{k_0} \Delta'_T z_t z'_t \Delta_T \theta^0 + 2\theta_0' \sum_{t=k+1}^{k_0} \Delta'_T z_t \tilde{u}_t$$

(5.12)
By substituting $\bar{u}_t = u_t + v'_t \beta_1^0 - z'_t [(Z'Z)^{-1} Z' V] \beta_1^0$ into (5.12) it follows that

$$-\theta'^0 W'_\Delta W_\Delta \theta^0 + 2\theta'^0 W'_\Delta \hat{U} \cdot \Xi = -\theta'^0 \hat{\Delta}' T \sum_{t=k+1}^{k_0} z_t z'_t \hat{\Delta} T \theta^0 + 2\theta'^0 \hat{\Delta}' T \left( \sum_{t=k+1}^{k_0} z_t u_t \right)$$

$$+ \sum_{t=k+1}^{k_0} z_t v'_t(\beta_1^0) + 2\theta'^0 \hat{\Delta}' T \sum_{t=k+1}^{k_0} z_t z'_t (Z'Z)^{-1} Z' V \beta_1^0$$

$$= -\theta'^0 \Delta'_0 \sum_{t=k+1}^{k_0} z_t z'_t \Delta_0 \theta^0 + 2\theta'^0 \Delta'_0 \left( \sum_{t=k+1}^{k_0} z_t u_t \right)$$

$$+ \sum_{t=k+1}^{k_0} z_t v'_t(\beta_1^0) + o_p(1)$$

where the last equality comes from the following convergence result over $D(C)$

$$\sum_{t=k+1}^{k_0} z_t z'_t (Z'Z)^{-1} Z' V = |k_0 - k| O_p(1) \cdot O_p(T^{-1}) \cdot O_p(T^{1/2}) = O_p(T^{-1/2}) = o_p(1).$$

Under strict stationarity, $\{z_t, u_t, v_t\}^{k_0}_{t=k+1}$ and $\{z_t, u_t, v_t\}^0_{t=k-k_0+1}$ have the same joint distribution. Therefore, (5.13) has the same distribution as $R_1(k - k_0)$ over $D(C)$.

Similarly, for $k > k_0$

$$-\theta'^0 W'_\Delta W_\Delta \theta^0 + 2\theta'^0 W'_\Delta \hat{U} \cdot \Xi = -\theta'^0 W'_\Delta W_\Delta \theta^0 - 2\theta'^0 W'_\Delta \hat{U}$$

$$= -\theta'^0 \Delta'_0 \sum_{t=k_0+1}^{k} z_t z'_t \Delta_0 \theta^0 - 2\theta'^0 \Delta'_0 \left( \sum_{t=k_0+1}^{k} z_t u_t \right)$$

$$+ \sum_{t=k_0+1}^{k} z_t v'_t(\beta_2^0) + o_p(1)$$

which has the same distribution as $R_2(k - k_0)$ over $D(C)$.

Thus, Lemma 3 implies that $V_T(k) - V_T(k_0)$ converges in distribution to $R^*(k - k_0)$ over the bounded set $D(C)$. Let $\hat{k}_C = \arg \max_{|k-k_0| \leq C} V_T(k) - V_T(k_0)$ and $m_C^* = \arg \max_{|m| \leq C} R^*(m)$. The uniform convergence of $V_T(k) - V_T(k_0)$ to $R^*(k - k_0)$ on any bounded set of integers (i.e. the difference $|k - k_0|$ is bounded) implies that $\hat{k}_C - k_0 \rightarrow_d m_C^*$. That is, for large $T$, $|P(\hat{k}_C - k_0 = j) - P(m_C^* = j)| \leq \epsilon_T$ for
j) < \epsilon \text{ for all } |j| \leq C. Actually, this convergence in distribution can be established for the whole range, not just over the bounded set \( D(C) \). By assuming 
\( \left( z_i^0 \Delta_0 \theta^0 \right) + 2 \theta^0 \Delta_0' z_t u_t + z_t v_i^0 \beta^0_0 \) has a continuous distribution, the limit distribution \( R^*(m) \) has a continuous distribution. Thus, the process \( R^*(m) \) has a unique maximum with probability one because \( P(R^*(m) = R^*(m')) = 0 \) for \( m \neq m' \).

Let \( m^* = \arg \max_m R^*(m) \). Since both \( \theta^0 \Delta_0' \sum_{t=m+1}^m z_t \Delta_0 \theta^0 = O_p(m) \) and 
\( \theta^0 \Delta_0' \sum_{t=1}^m z_t z_t' \Delta_0 \theta^0 = O_p(m) \) dominate \( \theta^0 \Delta_0' \sum_{t=m+1}^m z_t u_t + \sum_{t=m+1}^m z_t v_i^0 \beta^0_0 = O_p(m^{1/2}) \) and 
\( \theta^0 \Delta_0' \sum_{t=1}^m z_t u_t + \sum_{t=1}^m z_t v_i^0 \beta^0_0 = O_p(m^{1/2}) \), respectively, we have \( R^*(m) \to -\infty \) with probability tending to 1 as \( |m| \to \infty \). Thus, \( m^* \) is \( O_p(1) \).

Therefore, we have that for every \( \epsilon > 0 \), there exists \( C_1 < \infty \) such that \( P(|m^*| > C_1) < \epsilon \). And we know from Theorem 2 that \( P(|\hat{k} - k_0| > C_2) < \epsilon \).

However, if \( |\hat{k} - k_0| \leq C \) where \( C = \max\{C_1, C_2\} \), then \( \hat{k} = \hat{k}_C \) and if \( |m^*| \leq C \), then \( m^* = m^*_C \).

Thus,

\[
|P(\hat{k} - k_0 = j) - P(m^* = j)| \leq |P(\hat{k}_C - k_0 = j) - P(m^*_C = j)| \\
+ P(|\hat{k} - k_0| > C) + P(|m^*| > C) \\
< 3\epsilon
\]

Note that the first inequality in the above expression can be explained by the following way. First, we consider 3 events the union of which cover the whole sample space. Those are \( \{|\hat{k} - k_0| \leq C \text{ and } |m^*| \leq C\}, \{|\hat{k} - k_0| > C\} \) and \( \{|m^*| > C\} \). But the first event \( \{|\hat{k} - k_0| \leq C \text{ and } |m^*| \leq C\} \) is equivalent to the event \( \{\hat{k} = \hat{k}_C \text{ and } m^* = m^*_C\} \) by the definition of \( \hat{k}_C \) and \( m^*_C \). Thus, when the first event happens, we have the equality \( P(\hat{k} - k_0 = j) - P(m^* = j) = P(\hat{k}_C - k_0 = j) - P(m^*_C = j) \leq 3\epsilon \).
$P(m_C^* = j) - P(m_C^*_C = j)$. The union of other two events is the complement of the first event. Thus, the first inequality follows.

Since $\epsilon$ can be arbitrarily small and $C$ can be arbitrarily large, $P(\hat{k} - k_0 = j) = P(m^* = j)$. Therefore, $\hat{k} - k_0$ has the same limit distribution as $m^* = \arg \max_m R^*(m)$.

### 5.2 Shrinking Magnitude of Shift in the Regression Parameters

The application of the limiting distribution for the fixed shift case has some serious limitations due to its highly data-dependent nature. Analytical limiting distributions are hard to obtain in general. We needed Assumption 6 for the validity of the derivation in section 5.1. Also, because of this data dependence the limiting distribution is of less practical importance. We can avoid this problem by considering an asymptotic framework where the magnitudes of the shifts converge to zero as the sample size increases.

The approach we take here closely follows that of Bai(1997). We find a limiting distribution for small changes. And this limiting distribution can then be used as an approximation to the underlying distribution for moderate shifts.

To fit into the case of shrinking magnitude of shift in parameters, we need to modify the true behavior of regression parameters in our previous second regime model (5.3). We assume that $\theta_T^0 \equiv \beta_{2,T}^0 - \beta_{1}^0 \to 0$ as $T$ increases. The first regime model (5.1) remains unchanged. For the order of magnitude of shift, We assume
Assumption 7  Assuming $\theta_T^0$ depends on $T$

$$\theta_T^0 \to 0 \text{ and } T^{1/2-\alpha}\theta_T^0 \to \infty$$

for some $\alpha \in (0, 1/2)$

That is, $\theta_T^0$ converges to zero at slower rate than $T^{-1/2}$. Now we’ll establish the rate of convergence of the break date under Assumption 7.

In the previous section, we treated $\theta_T^0$ as a constant, not varying with $T$ and established

$$\hat{k} = k_0 + O_p(1)$$

With some modifications, In the following Theorem 5 we can show

$$\hat{k} = k_0 + O_p(\|\theta_T^0\|^{-2})$$

The rate of convergence not only describes how fast the estimator converges to the true value, it is also necessary in order to derive the limiting distribution.

In Theorem 6, with additional assumptions including the weak stationarity of \(\{x_t, z_t, u_t, v_t\}\) we establish the limiting distribution of break dates. As a preliminary step for the proof of the theorems, we need to prove additional lemmas.

Before that we redefine

$$V_T(k) - V_T(k_0) = -|k_0 - k|G_T(k) + H_T(k) \quad (5.14)$$

where

$$G_T(k) = \frac{\theta_T^0[W_0'M_WW_0 - W_0'M_WW_2(W_2'M_WW_2)^{-1}W_2'M_WW_0]\theta_T^0}{|k_0 - k|} \quad \text{for } k \neq k_0$$

$$G_T(k) = \theta_T^0\theta_T^0 \quad \text{for } k = k_0$$

and
\[ H_T(k) = 2\theta_T'(W_0'MWW_2)(W_2'MWW_2)^{-1}W_2'MWW_2 \tilde{U} - 2\theta_T'(W_0'MW\tilde{U}) \]
\[ + \tilde{U}'MW_2(W_2'MWW_2)^{-1}W_2'MW\tilde{U} \]
\[ - \tilde{U}'MW_0(W_0'MWW_0)^{-1}W_0'MWM\tilde{U} \] (5.15)

Lemma 4 is useful to establish positive definiteness of \( G_T(k) \) when \(|k-k_0|\) increases beyond \( C\|\theta_T^2\|^2 \) in the proof of Lemma 5. Lemma 7 investigates the limit behavior of \( V_T(k) - V_T(k_0) \) under Assumption 7. The result of it works as a stepping stone to Theorem 6 which derives the limiting distribution of the break points under Assumption 7.

**Lemma 4** The following two inequalities hold

\[ W_0'MWW_0 - W_0'MWW_2(W_2'MWW_2)^{-1}W_2'MMW_0 \geq W_0'MWW_0 - W_0'MWW_2(W_2'MWW_2)^{-1}W_2'MWW_0 \]
for \( k < k_0 \) (5.16)

\[ W_0'MWW_0 - W_0'MWW_2(W_2'MWW_2)^{-1}W_2'MWW_0 \geq W_0'MWW_0 - W_0'MWW_2(W_2'MWW_2)^{-1}W_2'MWW_0 \]
for \( k \geq k_0 \) (5.17)

**Proof**

Let \( H = (W_2'MW_2)^{-1} - (W'MW)^{-1} \).

For \( k \leq k_0 \)

\[ W_0'MWW_2(W_2'MWW_2)^{-1}W_2'MWW_0 = W_0'MWWW_2(W_2'MWW_2)^{-1} \]
\[ \times W_2'MWW_0 \] (5.18)

since

\[ W_0'MWW_2 = W_0'(I - W(W'M)^{-1}W'M)W_2 \]
Let \( A \) multiplying from the right by \( W'W \) to show

\[
W_0'W_2 - W_0'W(W'W)^{-1}W'W_2 = W_0'W_0 - W_0'W_0(W'W)^{-1}W_2'W_2
\]

\[
= W_0'W_0(W_2'W_2)^{-1}W_2'W_2 - W_0'W_0(W'W)^{-1}W_2'W_2
\]

\[
= W_0'W_0[(W_2'W_2)^{-1} - (W'W)^{-1}]W_2'W_2 = W_0'W_0 HW_2'W_2
\]

and

\[
W_2'M_2W_2 = W_2'(I - W(W'W)^{-1}W')W_2
\]

\[
= W_2'W_2 - W_2'W(W'W)^{-1}W_2'W_2 = W_2'W_2 - W_2'W_2(W'W)^{-1}W_2'W_2
\]

\[
= W_2'W_2(W_2'W_2)^{-1}W_2'W_2 - W_2'W_2(W'W)^{-1}W_2'W_2
\]

\[
= W_2'W_2[(W_2'W_2)^{-1} - (W'W)^{-1}]W_2'W_2 = W_2'W_2 HW_2'W_2
\]

Let \( A = H^{1/2}W_2'W_2 \). Since \( I - A(A'A)^{-1}A' \) is a projection matrix, we have \( I - A(A'A)^{-1}A' \geq 0 \). Multiplying this inequality by \((W_0'W_0)H^{1/2}\) from the left and multiplying from the right by \(H^{1/2}(W_0'W_0)\), we have

\[
(W_0'W_0)H^{1/2}H^{1/2}(W_0'W_0) - (W_0'W_0)H^{1/2}A(A'A)^{-1}A'H^{1/2}(W_0'W_0)
\]

\[
= W_0'W_0 HW_0'W_0 - W_0'W_0 H^{1/2}H^{1/2}W_2'W_2(W_2'W_2 H^{1/2}H^{1/2}W_2'W_2)^{-1}W_2'W_2 H^{1/2}
\]

\[
\times H^{1/2}W_0'W_0
\]

\[
= W_0'W_0 HW_0'W_0 - W_0'W_0 HW_2'W_2(W_2'W_2 HW_2'W_2)^{-1}W_2'W_2 HW_0'W_0
\]

\[
\geq 0
\]

Since the second term above is identical to (5.18), for the proof of (5.16) it suffices to show

\[
W_0'M_2W_0 - W_0'W_0 HW_0'W_0 \geq W_0'W_0 HW_2'W_2(W_2'W_2)^{-1}W_0'W_0
\]

(5.19)

In fact, the equality holds in (5.19) because the left hand side of (5.19) is

\[
W_0'M_2W_0 - W_0'W_0 HW_0'W_0 = W_0'(I - W(W'W)^{-1}W')W_0 - W_0'W_0 HW_0'W_0
\]
Using $W_2'W_2 = W_0'W_0 + W_\Delta W_\Delta$, (5.20) can be written as

$$W_0'M_\Delta W_0 - W_0'W_0HW_0^0W_0 = (W_2'W_2 - W_\Delta W_\Delta)[(W_0'^0W_0)^{-1} - (W_2'^0W_2)^{-1}]W_0'W_0$$

$$= [W_2'W_2(W_0'^0W_0)^{-1} - W_2'^0W_2(W_2'^0W_2)^{-1} - W_\Delta W_\Delta$$

$$\times (W_0'^0W_0)^{-1} + W_\Delta W_\Delta(W_2'^0W_2)^{-1}]W_0'W_0$$

$$= (W_2'W_2 - W_0'W_0 - W_\Delta W_\Delta) + W_\Delta W_\Delta(W_2'^0W_2)^{-1}$$

$$\times W_0'W_0$$

$$= W_\Delta W_\Delta(W_2'^0W_2)^{-1}W_0'W_0$$

Similarly, we can prove the case for $k \geq k_0$.

Define $W_2 = (\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_k, 0, \ldots, 0)'$, $W_0 = (\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_k, 0, \ldots, 0)'$ and $N = (W_2'^0W_2)^{-1} - (W'W)^{-1}$. Then

$$W_0'M_\Delta W_2 = W_0'W_2 - W_0'W(W'W)^{-1}W'W_2$$

$$= W_2'W_2 - W_0'W_0(W'W)^{-1}W_2'W_2$$

$$= W'W - W_2'^0W_2 - (W'W - W_0'^0W_0)(W'W)^{-1}(W'W - W_2'^0W_2)$$

$$= W'W - W_2'^0W_2 - W'W + W_2'^0W_2 + W_0'^0W_0 - W_0'^0W_0(W'W)^{-1}$$

$$\times W_2'^0W_2$$
Thus, we have

\[ W_0^s W_0^* - W_0^s W_0^*(W'W)^{-1} W_2^s W_2^* \]
\[ = W_0^s W_0^*(W_2^s W_2^*)^{-1} W_2^s W_2^* - W_0^s W_0^*(W'W)^{-1} W_2^s W_2^* \]
\[ = W_0^s W_0^*[(W_2^s W_2^*)^{-1} - (W'W)^{-1}] W_2^s W_2^* \]
\[ = W_0^s W_0^* NW_2^s W_2^* \]

and

\[ W'_2 M_W W_2 = W'_2 W_2 - W'_2 W(W'W)^{-1} W_2 \]
\[ = W'_2 W_2 - W'_2 W_2 (W'W)^{-1} W'_2 W_2 \]
\[ = (W'W - W_2^s W_2^*) - (W'W - W_2^s W_2^*)(W'W)^{-1} (W'W - W_2^s W_2^*) \]
\[ = W'W - W_2^s W_2^* - W'_2 W_2 + W_2^s W_2^* + W_2^s W_2^* - W_2^s W_2^* (W'W)^{-1} \]
\[ \times W_2^s W_2^* \]
\[ = W_2^s W_2^* - W_2^s W_2^* (W'W)^{-1} W_2^s W_2^* \]
\[ = W_2^s W_2^*[(W_2^s W_2^*)^{-1} - (W'W)^{-1}] W_2^s W_2^* \]
\[ = W_2^s W_2^* NW_2^s W_2^* \]

Thus, we have

\[ W'_0 M_W W_2 (W'_2 M_W W_2)^{-1} W'_2 M_W W_0 = W_0^s W_0^* NW_2^s W_2^* (W_2^s W_2^* NW_2^s W_2^*)^{-1} \]
\[ \times W_2^s W_2^* NW_2^s W_2^* \]

Let \( B = N^{1/2} W_2^s W_2^* \).

Using the fact, \( I - B(B'B)^{-1} B' \geq 0 \), we have

\[ (W_0^s W_0^*) N^{1/2} N^{1/2} (W_0^s W_0^*) - (W_0^s W_0^*) N^{1/2} B(B'B)^{-1} B' N^{1/2} (W_0^s W_0^*) \]
\[ = W_0^s W_0^* NW_0^s W_0^* - W_0^s W_0^* N^{1/2} N^{1/2} W_2 W_2^* W_2^* (W_2^s W_2^* N^{1/2} N^{1/2} W_2 W_2^*)^{-1} \]
\[ \times W_2^s W_2^* N^{1/2} N^{1/2} W_0^s W_0^* \]

75
\[
W_0^* W_0^* N W_0^* W_0^* - W_0^* W_0^* N W_2^* W_0^* (W_2^* W_0^* N W_2^* W_0^*)^{-1} W_2^* W_2^* N W_0^* W_0^* \\
\geq 0
\]

For the proof of (5.17), it suffices to show
\[
W_0^* M W_0^* W_0^* W_0^* N W_0^* W_0^* \\
\geq W_2^* W_2^* (W'W - W_2^* W_2^*)^{-1} \\
\times (W'W - W_0^* W_0^*) \quad (5.21)
\]

In fact, the equality holds in (5.21) because the left hand side of (5.21) is
\[
W_0^* M W_0^* W_0^* W_0^* N W_0^* W_0^* \\
= W_0^* (I - W(W'W)^{-1} W') W_0^* - W_0^* W_0^* N W_0^* W_0^* \\
\times W_0^* W_0^* \\
= W_0^* W_0^* - W_0^* W_0^* W_0^* W_0^* - W_0^* W_0^* N \\
\times W_0^* W_0^* \\
= (W'W - W_0^* W_0^*) - (W'W - W_0^* W_0^*) (W'W)^{-1} \\
\times (W'W - W_0^* W_0^*) - W_0^* W_0^* N W_0^* W_0^* \\
= W'W - W_0^* W_0^* - W'W + W_0^* W_0^* + W_0^* W_0^* \\
- W_0^* W_0^*(W'W)^{-1} W_0^* W_0^* - W_0^* W_0^* N W_0^* W_0^* \\
= W_0^* W_0^* - W_0^* W_0^* (W'W)^{-1} W_0^* W_0^* - W_0^* W_0^* N \\
\times W_0^* W_0^* \\
= W_0^* W_0^* [(W_0^* W_0^*)^{-1} - (W'W)^{-1} - N] W_0^* W_0^* \\
= W_0^* W_0^* [(W_0^* W_0^*)^{-1} - (W_2' W_2')^{-1}] W_0^* W_0^* \\
\quad (5.22)
\]

Using \( W_2' W_2' = W_0^* W_0^* + W_2' W_2' \), (5.22) can be written as
\[
W_0^* M W_0^* W_0^* W_0^* N W_0^* W_0^* = (W_2' W_2' - W_2' W_2') [(W_0^* W_0^*)^{-1} - (W_2' W_2')^{-1}]
\]
\[ \times W_0^* W_0^* \]
\[ = [W_2^* W_2^* (W_0^* W_0^*)^{-1} - W_2^* W_2^* (W_2^* W_2^*)^{-1} \]
\[ - W_\Delta^* W_\Delta (W_0^* W_0^*)^{-1} + W_\Delta^* W_\Delta (W_2^* W_2^*)^{-1} ] \]
\[ \times W_0^* W_0^* \]
\[ = (W_2^* W_2^* - W_0^* W_0^* - W_\Delta^* W_\Delta) + W_\Delta^* W_\Delta \]
\[ \times (W_2^* W_2^*)^{-1} W_0^* W_0^* \]
\[ = W_\Delta^* W_\Delta (W_2^* W_2^*)^{-1} W_0^* W_0^* \]
\[ = W_\Delta^* W_\Delta (W_2^* W_2^*)^{-1} (W_2^* W_2^* - W_0^* W_0^*) \]

By defining \( K(C) = \{ k : |k - k_0| > C \| \theta_T^0 \|^{-2} \} \) and \( T \eta \leq k \leq (1 - \eta) T \} \) for a small number \( \eta > 0 \), Lemma 5 shows

\[ \inf_{k \in K(C)} G_T(k) \geq \gamma \| \theta_T^0 \|^{-2} \]

with probability at least \( 1 - \epsilon \) for every \( \epsilon > 0 \).

Since it is defined that \( V_T(k) - V_T(k_0) = -|k_0 - k| G_T(k) + H_T(k) \) for all \( k \), \( V_T(k) \geq V_T(k_0) \) is equivalent to \( (H_T(k)/|k_0 - k|) \geq G_T(k) \). Thus, in Theorem 5 we establish \( \hat{k} = k_0 + O_p(\| \theta_T^0 \|^{-2}) \) where \( \hat{k} = \arg \max_k V_T(k) \) by showing that \( P(\sup_{k \in K(C)} V_T(k) \geq V_T(k_0)) \leq \epsilon \). For that it is sufficient to prove

\[ P(\sup_{k \in K(C)} |H_T(k)/(k_0 - k)| \geq \gamma \| \theta_T^0 \|^{-2}) < \epsilon \].

**Lemma 5** Under the Assumptions 2~3, there exists a \( \gamma > 0 \) such that for every \( \epsilon > 0 \) there exists \( C < \infty \) such that

\[ \inf_{k \in K(C)} G_T(k) \geq \gamma \| \theta_T^0 \|^{-2} \]

with probability at least \( 1 - \epsilon \).
Proof)
Suppose $k \leq k_0$. Let $A(k) = (k_0 - k)^{-1}W'_\Delta W_\Delta(W'_2 W_2)^{-1}W'_0 W_0$. Note that $A(k)$ is symmetric and is positive definite when $W'_\Delta W_\Delta$ is invertible. This is shown by rewriting $A(k)$ as

$$A(k) = (k_0 - k)^{-1}W'_\Delta W_\Delta(W'_\Delta W_\Delta + W'_0 W_0)^{-1}W'_0 W_0$$

$$= (k_0 - k)^{-1}W'_\Delta W_\Delta \left\{ (W'_\Delta W_\Delta)^{-1}[(W'_\Delta W_\Delta)^{-1} + (W'_0 W_0)^{-1}]^{-1} (W'_0 W_0)^{-1} \right\}$$

$$= (k_0 - k)^{-1}[(W'_\Delta W_\Delta)^{-1} + (W'_0 W_0)^{-1}]^{-1}$$

Then, by Lemma 4, $G_T(k) \geq \theta_T^0 A(k) \theta_T^0 \geq \gamma_T(k) ||\theta_T||^2$ where $\gamma_T(k)$ is the minimum eigenvalue of $A(k)$.

Now, for the proof of the lemma, it is sufficient to argue that, with probability tending to 1, $\gamma_T(k)$ is bounded away from zero as $k_0 - k$ increases.

For large $k_0 - k$, $W'_\Delta W_\Delta = \sum_{i=k+1}^{k_0} w_i w'_i = \hat{\Delta}'_T \sum_{i=k+1}^{k_0} z_i z'_i \hat{\Delta}_T$ will be positive definite with large probability by Assumption 3.

Now,

$$A(k)^{-1} = (W'_0 W_0)^{-1}W'_2 W_2[(k_0 - k)^{-1}W'_\Delta W_\Delta]^{-1}$$

$$||A(k)^{-1}|| \leq ||(W'_0 W_0)^{-1}W'_2 W_2||||(k_0 - k)^{-1}W'_\Delta W_\Delta||^{-1}$$

But

$$||(W'_0 W_0)^{-1}W'_2 W_2|| \leq ||(W'_0 W_0)^{-1}W'W||$$

$$\leq ||(W'_0 W_0)^{-1}|| ||W'W||$$

$$= ||(\hat{\Delta}'_T Z_0 Z_0 \hat{\Delta}_T)^{-1}|| ||\hat{\Delta}'_T Z' Z \hat{\Delta}_T||$$

where $Z_0 = (0, \ldots, 0, z_{k_0+1}, \ldots, z_T)'$ and $Z = (z_1, \ldots, z_T)'$. Thus, $||(W'_0 W_0)^{-1}W'_2 W_2||$ is bounded by Assumption 2 and 3.
In addition, the minimum eigenvalue of \((k_0 - k)^{-1} W'_{\triangle} W_{\triangle}\) is bounded away from zero by Assumption 3 with large probability. So, \(||[(k_0 - k)^{-1} W'_{\triangle} W_{\triangle}]^{-1}||\) is bounded with large probability for all large \(k_0 - k\).

This implies that the minimum eigenvalue of \(A(k), \gamma_T(k)\), is bounded away from zero as \(k_0 - k\) increases.

Suppose \(k > k_0\). Let \(B(k) = (k_0 - k)^{-1} W'_{\triangle} W_{\triangle} (W'W - W'_2 W_2)^{-1} (W'W - W'_0 W_0) = (k_0 - k)^{-1} W'_{\triangle} W_{\triangle} (W'_2 W'_2)^{-1} W'_0 W'_0\). It is clear that \(B(k)\) is symmetric and is positive definite when \(W'_{\triangle} W_{\triangle}\) is invertible.

By Lemma 4, \(G_T(k) \geq \theta^T_B(k) \theta_T \geq \gamma^*_T(k)||\theta_T||^2\) where \(\gamma^*_T(k)\) is the minimum eigenvalue of \(B(k)\).

Now, for the proof we argue that, with probability tending to one, \(\gamma^*_T(k)\) is bounded away from zero as \(k_0 - k\) increases.

For large \(k_0 - k\), \(W'_{\triangle} W_{\triangle} = \sum_{t=k+1}^{k_0} w_t w_t' = \Delta_T^T \sum_{t=k+1}^{k_0} z_t z_t' \Delta_T\) will be positive definite with large probability by Assumption 3.

Now,

\[
B(k)^{-1} = (W'_0 W'_0)^{-1} W'_2 W'_2 [(k_0 - k)^{-1} W'_{\triangle} W_{\triangle}]^{-1}
\]

\[
||B(k)^{-1}|| \leq \|(W'_0 W'_0)^{-1} W'_2 W'_2\| \cdot \||[(k_0 - k)^{-1} W'_{\triangle} W_{\triangle}]^{-1}||\]

But

\[
\|(W'_0 W'_0)^{-1} W'_2 W'_2\| \leq \|(W'_0 W'_0)^{-1} W'W\|
\]

\[
\leq \|(W'_0 W'_0)^{-1} W'W\|
\]

\[
= \|\Delta_T^T Z'_0 Z_0 ^* \Delta_T\|^{-1} \|\Delta_T^T Z' Z \Delta_T\|
\]

is bounded by Assumption 2 and 3. Here, \(Z_0^* = (Z_1, \ldots, Z_{k_0}, 0, \ldots, 0)'\) defined.
In addition, the minimum eigenvalue of \((k_0 - k)^{-1}W^\prime \triangle W\) is bounded away from zero with large probability. This implies that the minimum eigenvalue of \(B(k), \gamma^*_r(k)\), is bounded away from zero as \(k_0 - k\) increases. This completes the proof of the lemma.

For the proof of the stochastic boundedness of the break point estimator in Theorem 5, we need a maximal inequality result which is stated in the following Lemma 6. In general, maximal inequalities, which place bounds on the extreme behavior of which a sequence is capable over a succession of steps, are essential tools of limit theory. For the validity of Lemma 6, we need Assumption 8 to be satisfied.

**Assumption 8** For some real number \(r > 2\) and constant \(A_r < \infty\)

\[
E \left\| \sum_{t=i}^{j} w_t \tilde{u}_t \right\|^r \leq A_r (j - i)^{r/2} \quad \text{for all } 1 \leq i \leq j \leq T
\]

Furthermore, by defining \(M_m = \max\{|S_1|, \ldots, |S_m|\}\) where \(S_m = \sum_{t=1}^{m} w_t \tilde{u}_t\), Assumption 8 implies the existence of a constant \(K_r < \infty\) such that \(E(M^r_m) \leq K_r m^{r/2}\) for \(m \geq 1\).

**Lemma 6** Under Assumption 8, there exists a \(B < \infty\) such that

for every \(\zeta > 0\) and \(m > 0\)

\[
P \left( \sup_{k \geq m} \frac{1}{k} \left\| \sum_{t=1}^{k} w_t \tilde{u}_t \right\| > \zeta \right) \leq \frac{B}{\zeta^4 m^2}
\]

\(^1\)see the equation (7.10) in Serfling(1970) [p.1244].
Proof)

This is essentially proved by Serfling(1970) (Theorm 5.1). It says that under Assumption 8, for each \( \zeta > 0 \) there exists a constant \( C_\zeta < \infty \) (depending on \( A_r \) and \( K_r \)) such that

\[
P \left( \sup_{k \geq m} \left\| \frac{S_k}{k} \right\| > \zeta \right) \leq C_\zeta \cdot m^{-r/2} \quad \text{for all } m \geq 1
\]

where \( C_\zeta = (A_r + K_r)(\frac{\zeta}{2})^{-r}(1 - 2^{-r/2})^{-1} \)

Thus,

\[
P \left( \sup_{k \geq m} \frac{1}{k} \left\| \sum_{t=1}^{k} w_i \tilde{u}_t \right\| > \zeta \right) \leq C_\zeta \cdot m^{-r/2}
\]

By letting \( r = 4, B = (A_r + K_r)(1/2)^{-r}(1 - 2^{-r/2})^{-1} \), we get the desired maximal inequality

\[
P \left( \sup_{k \geq m} \frac{1}{k} \left\| \sum_{t=1}^{k} w_i \tilde{u}_t \right\| > \zeta \right) \leq \frac{B}{\zeta^4 m^2}
\]

**Theorem 5** Under the Assumptions 1 ~ 5 and 8

If \( \theta_I^0 \xrightarrow{p} 0 \) but satisfying Assumption 7, then

\[
\hat{k} = k_0 + O_p(||\theta_I^0||^{-2})
\]

Proof)

By definition, \( \hat{k} = \arg \max_k V_T(k) \). Thus, \( V_T(\hat{k}) \geq V_T(k_0) \). Therefore, for the proof of the Proposition, it suffices to show that for each \( \epsilon > 0 \), there exists \( C > 0 \) such that

\[
P \left( \sup_{k \in K(C)} V_T(k) \geq V_T(k_0) \right) < \epsilon \quad (5.23)
\]

From the previous identity (5.14)

\[
V_T(k) - V_T(k_0) = -|k_0 - k|G_T(k) + H_T(k) \quad \text{for all } k
\]
$V_T(k) \geq V_T(k_0)$ is equivalent to

$$(H_T(k)/|k_0 - k|) \geq G_T(k)$$

By Lemma 5, it suffices to prove that

$$P \left( \sup_{k \in K(C)} \left| \frac{H_T(k)}{k_0 - k} \right| \geq \gamma ||\theta_T^0||^2 \right) < \epsilon \quad (5.24)$$

$H_T(k)$ defined at (5.15) can be decomposed into two parts such that

$$H_T(k) = (B.1) + (B.2)$$

where

$$(B.1) = 2\theta_T^0 W_0' M W W_2 (W_2' M W W_2)^{-1} W_2' M W \tilde{U} - 2\theta_T^0 W_0' M W \tilde{U}$$

and

$$(B.2) = \tilde{U}' M W W_2 (W_2' M W W_2)^{-1} W_2' M W \tilde{U} - \tilde{U}' M W W_0 (W_0' M W W_0)^{-1} W_0' M W \tilde{U}$$

First consider the expression (B.1). Recalling the definition $W_0 = W_2 - W_\Delta \Xi$ from (5.8), (B.1) can be transformed into

$$(B.1) = \Xi [2\theta_T^0 W_\Delta M W \tilde{U} - 2\theta_T^0 W_\Delta' M W W_2 (W_2' M W W_2)^{-1} W_2' M W \tilde{U}] \quad (5.25)$$

following the same derivation at (5.11) except that \(\theta^0\) in (5.11) has been replaced by \(\theta_T^0\) in the current shrinking magnitude of shift case.

The first term within the square bracket can be expanded as

$$W'_{\Delta} M W \tilde{U} = W'_{\Delta} \tilde{U} - W'_{\Delta} W (W' W)^{-1} W' \tilde{U}$$

It is clear that $W'_{\Delta} W = |k_0 - k| O_p(1)$ and $(W' W)^{-1} W' \tilde{U} = (1/\sqrt{T})(W' W/T)^{-1} \times W' \tilde{U}/\sqrt{T} = T^{-1/2} O_p(1)$.
Thus, the first term at (5.25) can be written as

\[ \theta_T^{0\prime}W_{\Delta}^\prime M_W \bar{U} = \theta_T^{0\prime}W_{\Delta}^\prime \bar{U} - ||\theta_T^{0\prime}|| \cdot |k_0 - k|O_p(1) \cdot T^{-1/2}O_p(1) \]

\[ = \theta_T^{0\prime}W_{\Delta}^\prime \bar{U} - |k_0 - k|T^{-1/2}||\theta_T^{0\prime}||O_p(1) \quad (5.26) \]

Now, we consider each factor in the second term within the square bracket at (5.25).

\[ W_{\Delta}^\prime M_W W_2 = W_{\Delta}^\prime W_2 - W_{\Delta}^\prime W(WW')^{-1}WW' \]

\[ = |k_0 - k|O_p(1) - |k_0 - k|O_p(1) \cdot O_p(T^{-1}) \cdot O_p(T) \]

\[ = |k_0 - k|O_p(1) \]

\[ (W_2' M_W W_2)^{-1} W_2' M_W \bar{U} = O_p(T^{-1}) \cdot O_p(T^{1/2}) = O_p(T^{-1/2}) \]

Thus, the second factor at (5.25)

\[ 2 \Xi \theta_T^{0\prime} W_{\Delta}^\prime M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \bar{U} = 2 ||\theta_T^{0\prime}|| \cdot |k_0 - k|O_p(1) \cdot O_p(T^{-1/2}) \Xi \]

\[ \quad (5.27) \]

Therefore, combining the results in (5.26) and (5.27) over the set \( K(C) \)

\[ (B.1) \quad = \Xi \left( 2 \theta_T^{0\prime} W_{\Delta}^\prime \bar{U} - 2|k_0 - k|T^{-1/2}||\theta_T^{0\prime}||O_p(1) - 2|k_0 - k|T^{-1/2}||\theta_T^{0\prime}||O_p(1) \right) \]

\[ = 2 \theta_T^{0\prime} W_{\Delta}^\prime \bar{U} \Xi + |k_0 - k|T^{-1/2}||\theta_T^{0\prime}||O_p(1) \Xi \]

\[ \quad (5.28) \]

Now, consider the expression

\[ (B.2) \quad = \bar{U}'M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \bar{U} - \bar{U}'M_W (W_2 - W_{\Delta} \Xi) \]

\[ \times [(W_2 - W_{\Delta} \Xi)'M_W (W_2 - W_{\Delta} \Xi)]^{-1}(W_2 - W_{\Delta} \Xi)'M_W \bar{U} \]
The first term of (B.2) is easily seen to be bounded as

\[ \tilde{U}'MW_2(W'_2MW_2)^{-1}W'_2MW\tilde{U} = O_p(T^{1/2}) \cdot O_p(T^{-1}) \cdot O_p(T^{1/2}) = O_p(1) \]  \hspace{1cm} (5.29)

The second term of (B.2) can be expanded further

\[ \tilde{U}'MW_2(W_2 - W_\Delta\Xi) \cdot [(W_2 - W_\Delta\Xi)'MW_2(W_2 - W_\Delta\Xi)]^{-1}(W_2 - W_\Delta\Xi)'MW\tilde{U} = (\tilde{U}'MW_2 - \tilde{U}'MW_\Delta\Xi) [W'_2MW_2 - W'_2MW_\Delta\Xi - \Xi W'_2MW_2 \]

\[ + W'_2MWW_\Delta]^{-1} (W'_2MW\tilde{U} - W'_2MW\tilde{U}\Xi) \]

Investigating the limit behavior of each term over the set \( K(C) \), we have the following convergence results.

\[ \tilde{U}'MW_2 = O_p(T^{1/2}) \]

\[ \tilde{U}'MW_\Delta = \tilde{U}'W_\Delta - \tilde{U}'W(W'W)^{-1}W'W_\Delta \]

\[ = \sqrt{|k_0 - k|} \cdot O_p(1) - O_p(T^{1/2}) \cdot O_p(T^{-1}) \cdot |k_0 - k| O_p(1) \]

\[ = \sqrt{|k_0 - k|} O_p(1) - |k_0 - k| O_p(T^{-1/2}) \]

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

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

\[ W'_2MWW_\Delta = O_p(1) \cdot |k_0 - k| = O_p(T) \]

\[ W'_2MW_2 = O_p(T) \]

and

\[ W'_\Delta MW_\Delta = W'_\Delta W_\Delta - W'_\Delta W(W'W)^{-1}W'W_\Delta \]

\[ = |k_0 - k| O_p(1) - |k_0 - k| O_p(T^{-1}) \cdot |k_0 - k| O_p(1) \]

\[ = |k_0 - k| O_p(1) - |k_0 - k|^2 O_p(T^{-1}) \]
Thus, the second term of (B.2) is

\[
\left( O_p(T^{1/2}) - O_p(T^{1/2}) \right) \left[ O_p(T) - O_p(T) + O_p(T) \right]^{-1} \\
\times \left( O_p(T^{1/2}) - O_p(T^{1/2}) \right) \\
= O_p(T^{1/2})O_p(T^{-1})O_p(T^{1/2}) \\
= O_p(1) \tag{5.30}
\]

Finally, combining (5.28), (5.29) and (5.10)

\[
H_T(k) = (B.1) + (B.2) \\
= \left( 2\theta_T^0 W_{\Delta} \tilde{U} \Xi + |k_0 - k|T^{-1/2}||\theta_T^0||O_p(1)\Xi \right) + (O_p(1) + O_p(1)) \\
= \left( 2\theta_T^0 W_{\Delta} \tilde{U} \Xi + |k_0 - k|T^{-1/2}||\theta_T^0||O_p(1)\Xi \right) + O_p(1)
\]

Thus, over the set \( K(C) \)

\[
\frac{H_T(k)}{|k_0 - k|} = \frac{2}{|k_0 - k|} \theta_T^0 W_{\Delta} \tilde{U} \Xi + T^{-1/2}||\theta_T^0||O_p(1) + \frac{O_p(1)}{|k_0 - k|} \tag{5.31}
\]

We can now prove (5.24) using the above equation (5.31). For that, we investigate the probability limit behavior of the supremum of each term in (5.31) over \( K(C) \) and using the triangle inequality we show that

\[
P \left( \sup_{k \in K(C)} \left| \frac{H_T(k)}{k_0 - k} \right| > \gamma||\theta_T^0||^2 \right) \leq P \left( \sup_{k \in K(C)} \frac{2}{k_0 - k} |\theta_T^0 W_{\Delta} \tilde{U}| > \frac{\gamma||\theta_T^0||^2}{3} \right) \\
+ P \left( |T^{-1/2}||\theta_T^0||O_p(1)| > \frac{\gamma||\theta_T^0||^2}{3} \right) + P \left( \sup_{k \in K(C)} \left| \frac{O_p(1)}{k_0 - k} \right| > \frac{\gamma||\theta_T^0||^2}{3} \right) \\
< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon
\]
By the symmetry of the argument, we only consider the case for $k < k_0$. First,

$$P \left( \sup_{k \in K(C)} \left| \frac{2}{k_0 - k} \theta_T^0 W_{\Delta} \tilde{U} \right| > \frac{\gamma \|\theta_T^0\|^2}{3} \right) \leq P \left( \sup_{k \leq k_0} \left| \frac{1}{k_0 - k} \sum_{t=k+1}^{k_0} w_t \tilde{u}_t \right| > \frac{\gamma \|\theta_T^0\|}{6} \right)$$

Applying Lemma 6 with $\zeta = \frac{\gamma \|\theta_T^0\|}{6}$, $m = C \|\theta_T^0\|^{-2}$, the right-hand side above is bounded by

$$B \frac{6^4}{\gamma^4 \|\theta_T^0\|^4} \cdot \frac{1}{C^2 \|\theta_T^0\|^{-4}} = B \frac{1296}{\gamma^4 C^2} < \frac{\epsilon}{3}$$

for large C.

Second,

$$P \left( \frac{|O_p(1)|}{T^{1/2} \|\theta_T^0\|} > \frac{\gamma \|\theta_T^0\|^2}{3} \right) = P \left( \frac{|O_p(1)|}{T^{1/2} \|\theta_T^0\|} > \frac{\gamma}{3} \right) < \frac{\epsilon}{3}$$

because $(T^{1/2} \|\theta_T^0\|)^{-1} \to 0$.

Finally, consider the limit behavior of the supremum of the third term in (5.31). By imposing the restriction $k < k_0$ to the set $K(C)$, we get $k \leq k_0 - C \|\theta_T^0\|^{-2}$ which implies

$$\left| \frac{1}{k_0 - k} \right| \leq \frac{1}{C \|\theta_T^0\|^2}$$

Thus, for $k < k_0$

$$P \left( \sup_{k \in K(C)} \left| \frac{O_p(1)}{k_0 - k} \right| > \frac{\gamma \|\theta_T^0\|^2}{3} \right) < P \left( \sup_{k \in K(C)} \|\theta_T^0\|^2 \left| \frac{O_p(1)}{C} \right| > \frac{\gamma \|\theta_T^0\|^2}{3} \right)$$

$$= P \left( \left| \frac{O_p(1)}{C} \right| > \frac{\theta}{3} \right)$$

$$< \frac{\epsilon}{3}$$

for large C.

Combining the results on the three terms of (5.31), we conclude

$$P \left( \sup_{k \in K(C)} \left| \frac{H_T(k)}{k_0 - k} \right| > \gamma \|\theta_T^0\|^2 \right) < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon$$
which is the expression (5.24).
This completes the proof of the Theorem 5.

Lemma 7 investigates the limit behavior of $V_T(k) - V_T(k_0)$ under Assumption 7. It is used in the following Theorem 6 which derives the limiting distribution of the break points. The proof of Lemma 7 is almost identical to that of Lemma 3 for the fixed magnitude of shift case.

**Lemma 7** Under Assumptions 1~5, and 7

\[
V_T(k) - V_T(k_0) = -\theta_T' W_\Delta W_\Delta \theta_T^0 + 2\theta_T' W_\Delta \tilde{U} \cdot \Xi + o_p(1)
\]

**Proof)**
Recalling $V_T(k) - V_T(k_0) = -|k_0 - k|G_T(k) + H_T(k)$ from (5.14), we study the convergence of each term in the sum in turn. We start with

\[
|k_0 - k|G_T(k) = \theta_T'[W_0'M_WW_0 - W_0'M_WW_2(W_2'M_WW_2)^{-1}W_2'M_WW_0]\theta_T^0
\]

\[
= \theta_T'[(W_2 - \Xi W_\Delta)'M_W(W_2 - \Xi W_\Delta) - (W_2 - \Xi W_\Delta)'M_WW_2
\times (W_2'M_WW_2)^{-1}W_2'M_W(W_2 - \Xi W_\Delta)]\theta_T^0
\]

\[
= \theta_T'[W_\Delta'M_WW_\Delta - W_\Delta'M_WW_2(W_2'M_WW_2)^{-1}W_2'M_WW_\Delta]\theta_T^0
\]

(5.32)

From Theorem 5 we have $P(|\hat{k} - k_0| > C||\theta_T^0||^{-2}) < \eta$ for every $\eta > 0$. Thus it is sufficient to investigate the behavior of $V_T(k) - V_T(k_0)$ over $D(C) \equiv \{k : |k - k_0| \leq C||\theta_T^0||^{-2}\}$ for the derivation of the limiting distribution of break points.

First, consider the second term in (5.32). Since $\|W_\Delta'M_WW_2\| = O_p(1)||\theta_T^0||^{-2}$ and $\|(W_2'M_WW_2)^{-1}\| = O_p(T^{-1})$ over the set $D(C)$, the second term in the equa-
tion (5.32) is bounded by \(\|\theta_T^0\|^2 O_p(1)\|\theta_T^0\|^2 \cdot O_p\left(\frac{1}{T}\right) = O_p\left(\frac{1}{T}\right)\|\theta_T^0\|^2 = O_p(1)\) under Assumption 7.

Next, consider the first term \(\theta_T^0 W_{\Delta} M_W W_{\Delta} \theta_T^0 = \theta_T^0 W_{\Delta} W_{\Delta} \theta_T^0 + \theta_T^0 W_{\Delta} W (W'W)^{-1} W' W_{\Delta} \theta_T^0\). We know that \(\theta_T^0 W_{\Delta} W (W'W)^{-1} W' W_{\Delta} \theta_T^0 = \|\theta_T^0\|^2 O_p(1)\) \times \|\theta_T^0\|^2 = O_p(1)\) under Assumption 7.

Thus, combining the results on two terms, we get

\[
|k_0 - k| G_T(k) = \theta_T^0 W_{\Delta} W_{\Delta} \theta_T^0 + o_p(1)
\]

Next, consider \(H_T(k)\).

\[
H_T(k) = 2\theta_T^0 (W_0^0 M_W W_2) (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} - 2\theta_T^0 (W_0^0 M_W \tilde{U})
\]

\[
+ \tilde{U}' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} - \tilde{U}' M_W W_0 (W_0' M_W W_0)^{-1} W_0' M_W \tilde{U}
\]

\[
= (B.1) + (B.2)
\]

where we defined

\[
(B.1) = 2\theta_T^0 (W_0^0 M_W W_2) (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} - 2\theta_T^0 (W_0^0 M_W \tilde{U})
\]

\[
(B.2) = \tilde{U}' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} - \tilde{U}' M_W W_0 (W_0' M_W W_0)^{-1} W_0' M_W \tilde{U}
\]

In the following derivation, we show \((B.2) = o_p(1)\). Thus the limit behavior of \(H_T(k)\) over \(D(C)\) is dominated by the limit behavior of \((B.1)\).

First, consider \((B.1)\).

\[
(B.1) = 2\theta_T^0 (W_2 - W_{\Delta} \Xi)' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} - 2\theta_T^0 W_0^0 M_W \tilde{U}
\]

\[
= 2\theta_T^0 (W_2' M_W \tilde{U} - \Xi W_{\Delta} M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U}) - 2\theta_T^0 W_0^0 M_W \tilde{U}
\]

\[
= \Xi[2\theta_T^0 W_{\Delta} M_W \tilde{U} - 2\theta_T^0 W_{\Delta} M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U}]
\]

88
Let’s investigate the convergence result of each term in turn. Noticing that $W'_\Delta W = O_p(1)\|\theta'_T\|^2$ over $D(C)$ and $(WW)^{-1}W'\tilde{U} = (1/\sqrt{T}) \cdot (W'W/T)^{-1}W'\tilde{U}/\sqrt{T} = T^{-1/2}O_p(1)$, the first term in the last equality of (B.1) can be written as

$$\theta'^0 W'_\Delta M W \tilde{U} = \theta'^0 W'_\Delta \tilde{U} - \theta'^0 W'_\Delta W (W'W)^{-1}W' \tilde{U}$$

$$= \theta'^0 W'_\Delta \tilde{U} - \|\theta_T\| \cdot O_p(1)\|\theta'_T\|^2 \cdot T^{-1/2}O_p(1)$$

$$= \theta'^0 W'_\Delta \tilde{U} - O_p(1)/(T^{1/2}\|\theta'_T\|)$$

$$= \theta'^0 W'_\Delta \tilde{U} + o_p(1) \quad \text{over } D(C)$$

Similarly, we observe that

$$W'_\Delta M W W_2 = W'_\Delta W_2 - W'_\Delta W (W'W)^{-1}W' W_2$$

$$= O_p(1)\|\theta'_T\|^2 - O_p(1)\|\theta'_T\|^2 \cdot O_p(T^{-1}) \cdot O_p(T)$$

$$= O_p(1)\|\theta'_T\|^2 \quad \text{over } D(C)$$

$$(W'_2 M W W_2)^{-1}W'_2 M W \tilde{U} = O_p(T^{-1}) \cdot O_p(T^{1/2}) = O_p(T^{-1/2})$$

Thus, the second term in the last equality of (B.1) can be written as

$$\theta'^0 W'_\Delta M W W_2 \cdot (W'_2 M W W_2)^{-1}W'_2 M W \tilde{U} = \|\theta'_T\| \cdot O_p(1)\|\theta'_T\|^2 \cdot O_p(T^{-1/2})$$

$$= O_p(1)/(T^{1/2}\|\theta'_T\|)$$

$$= o_p(1) \quad \text{over } D(C)$$

Thus, combining the results on the two terms of (B.1), we have

$$(B.1) = \Xi[2\theta'^0 W'_\Delta M W \tilde{U} - 2\theta'^0 W'_\Delta M W W_2 (W'_2 M W W_2)^{-1}W'_2 M W \tilde{U}]$$

$$= 2\Xi\theta'^0 W'_\Delta \tilde{U} + o_p(1)$$
Now, we prove that \((B.2) = o_p(1)\) in the limit.

\[
(B.2) = \hat{U}'M_WW_2(W_2'M_WW_2)^{-1}W_2'M_W\hat{U} - \hat{U}'M_WW_0(W_0'M_WW_0)^{-1}W_0'M_W\hat{U}
\]

\[
= \hat{U}'M_WW_2(W_2'M_WW_2)^{-1}W_2'M_W\hat{U} - \hat{U}'M_W(W_2 - W_\triangle \Xi)
\times [(W_2 - W_\triangle \Xi)'M_W(W_2 - W_\triangle \Xi)]^{-1}(W_2 - W_\triangle \Xi)'M_W\hat{U}
\]

\[
\equiv (M) - (N)
\]

where \((M)\) and \((N)\) are defined as

\[
(M) \equiv \hat{U}'M_WW_2(W_2'M_WW_2)^{-1}W_2'M_W\hat{U}
\]

\[
(N) \equiv \hat{U}'M_W(W_2 - W_\triangle \Xi)[(W_2 - W_\triangle \Xi)'M_W(W_2 - W_\triangle \Xi)]^{-1}(W_2 - W_\triangle \Xi)'M_W\hat{U}
\]

To investigate the limiting behavior of \((B.2)\) over the set \(D(C)\), it is helpful to check the limit behavior of \((N)\) first. What we want to show is that \((N)\) has the same probability limit as \((M)\). So in the limit the two terms cancel out to produce \((B.2) = o_p(1)\).

Using the relationship \(W_0 = W_2 - W_\triangle \Xi\), \((N)\) can be written as

\[
(N) = \left\{ \hat{U}'M_W(W_2 - W_\triangle \Xi)/\sqrt{T} \right\} \cdot [(W_2 - W_\triangle \Xi)'M_W(W_2 - W_\triangle \Xi)/T]^{-1}
\times \left\{ (W_2 - W_\triangle \Xi)'M_W\hat{U}/\sqrt{T} \right\}
\]

Let’s check the limit behavior of the first factor of \((N)\).

\[
\hat{U}'M_W(W_2 - W_\triangle \Xi)/\sqrt{T} = \hat{U}'M_WW_2/\sqrt{T} - \hat{U}'M_WW_\triangle \Xi/\sqrt{T}
\]

\[
= \hat{U}'M_WW_2/\sqrt{T} + o_p(1)
\]

The last equality follows from the fact that

\[
\hat{U}'M_WW_\triangle \Xi/\sqrt{T} = W_\triangle \hat{U}\Xi/\sqrt{T} + o_p(1)
\]
\[ O_p(1) \sqrt{\|\theta_T\|^{-2}/T} + o_p(1) \]
\[ = o_p(1) + o_p(1) \]
\[ = o_p(1) \quad \text{over } D(C) \equiv \{ k : |k - k_0| \leq C\|\theta_T\|^{-2} \} \]

Similarly, for the second factor of \((N)\)

\[
(W_2 - W_\triangle \Xi)'M_W(W_2 - W_\triangle \Xi)/T = \frac{W_2' M_W W_2}{T} - \left\{ \frac{W_2' M_W W_\triangle \Xi}{T} + \Xi W_\triangle^\prime M_W W_2 - W_\triangle^\prime M_W W_\triangle \right\}
\]
\[ = W_2' M_W W_2/T + o_p(1) \]

Thus, combining the convergence results on the two factors of \((N)\), we have

\[
(N) = \left( \tilde{U}'M_W W_2 / \sqrt{T} + o_p(1) \right) \left( W_2' M_W W_2/T + o_p(1) \right)^{-1} \left( W_2' M_W \tilde{U} / \sqrt{T} + o_p(1) \right)
\]

Therefore,

\[
(B.2) = (M) - (N)
\]
\[ = \tilde{U}'M_W W_2(W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} - \left( \tilde{U}'M_W W_2 / \sqrt{T} + o_p(1) \right)
\]
\[ \times \left( W_2' M_W W_2/T + o_p(1) \right)^{-1} \left( W_2' M_W \tilde{U} / \sqrt{T} + o_p(1) \right)
\]
\[ = \left( \tilde{U}'M_W W_2 / \sqrt{T} \right) \left( W_2' M_W W_2/T \right)^{-1} \left( W_2' M_W \tilde{U} / \sqrt{T} \right)
\]
\[ - \left( \tilde{U}'M_W W_2 / \sqrt{T} + o_p(1) \right) \left( W_2' M_W W_2/T + o_p(1) \right)^{-1}
\]
\[ \times \left( W_2' M_W \tilde{U} / \sqrt{T} + o_p(1) \right)
\]
\[ = o_p(1) \]

Thus

\[
H_T(k) = (B.1) + (B.2)
\]
\[ = \left( 2\Xi \theta_T^0 W_\triangle^\prime \tilde{U} + o_p(1) \right) + o_p(1) \]
\[ = 2\Xi \theta_T^0 W_\triangle^\prime \tilde{U} + o_p(1) \]

91
Finally,

\[ V_T(k) - V_T(k_0) = -|k_0 - k|G_T(k) + H_T(k) \]

\[ = \left( -\theta_T^0 W'_\Delta W_\Delta \theta_T^0_T + o_p(1) \right) + \left( 2\Xi \theta_T^0 W'_\Delta \tilde{U} + o_p(1) \right) \]

\[ = -\theta_T^0 W'_\Delta W_\Delta \theta_T^0_T + 2\Xi \theta_T^0 W'_\Delta \tilde{U} + o_p(1) \]

as claimed in Lemma 7. This ends the proof of Lemma 7.

Based on the convergence rate in Theorem 5, we are now ready to derive the limiting distribution of the given break point estimators. Contrary to the fixed magnitude of shift case, the shrinking magnitude of shift case doesn’t require a strong assumption of strict stationarity of regressors and errors to control the joint distribution of neighbors in the sequence. The weaker assumption of covariance stationarity within each regime should be enough for the well-behaved limiting distribution of breaks. We also modify the previous Assumption 7 to Assumption 10 just for easier asymptotic argument.

**Assumption 9** \( \{z_t, x_t\} \) is second-order stationary within each regime such that \( Ez_tz'_t = Q_1 \) for \( t \leq k_0 \) and \( Ez_tz'_t = Q_2 \) for \( t > k_0 \).

**Assumption 10** \( \theta_T^0 = \theta_0 v_T \), where \( v_T \) is a positive number such that \( v_T \rightarrow 0 \) and \( T^{(1/2)-\alpha} v_T \rightarrow \infty \) for some \( \alpha \in (0, 1/2) \) and \( \theta_0 \neq 0 \).

In Theorem 5, we showed that \( \hat{k} = k_0 + O_p(\|\theta_T^0\|^2) \). So, the limiting distribution of \( \hat{k} \) may be obtained by the local weak convergence of \( V_T(k) - V_T(k_0) \) in
conjunction with the continuous mapping theorem for the argmax functional. $k$ may vary in a neighborhood of $k_0$ such that $k = k_0 + [sv_T^2]$ where $v_T^2 = O(\|\theta_0\|^2)$ and $s$ is a real number in a compact set.

In Theorem 6, we start by deriving the limiting process

$$V_T(k_0 + [sv_T^2]) - V_T(k_0) \Rightarrow G(s)$$

for $s \in [-C, C]$ where $C > 0$ and “$\Rightarrow$” denotes weak convergence in the space $D[-C, C]$ under the Skorokhod metric. Then the continuous mapping theorem can be invoked to obtain

$$v_T^2 \overset{d}{\rightarrow} \arg \max_s G(s)$$

which can be transformed into

$$b^{-1}v_T^2(\hat{k} - k_0) \overset{d}{\rightarrow} \arg \max_s Z(s)$$

In order to understand this final form, we define some notations used in the proof.

Let $B_i(r) = \left[ B_i^1(r)', B_i^2(r)', \ldots, B_i^{p+1}(r)' \right]'$ be a $[q \times (p+1)] \times 1$ standard Brownian motion corresponding to the $i$th regime where $B_i^j(r)'$ is $q \times 1$ for $j = 1, \ldots, p+1$. And simultaneously $B_i^{mat}$ is defined as $vec(B_i^{mat}) = B_i(r)$. Standard Brownian motion is merely the leading member of an extensive family of a.s. continuous process on $[0, 1]$. To deal with partial sum process of $\{z_t u_t\}$ and $\{z_t v_t\}$ whose limit processes are not standard Brownian motion processes, we need to specify the variance structure of them. So, we impose the following conditional moment assumption.

**Assumption 11** For regime $i, i = 1, 2$, the errors $\{u_t, v_t\}$ satisfy

$$Var \left[ \begin{pmatrix} u_t \\ v_t \end{pmatrix} | z_t \right] = \Omega_i = \begin{bmatrix} \sigma_i^2 & \gamma_i' \\ \gamma_i & \Sigma_i \end{bmatrix}$$
where $\Omega_i$ is a constant, positive definite matrix. $\sigma_i^2$ is a scalar and $\Sigma_i$ is $p \times p$ matrix.

$\Omega_i^{1/2}$ and $Q_i^{1/2}$ are defined as the symmetric matrices satisfying $\Omega_i = \Omega_i^{1/2} \Omega_i^{1/2}$ and $Q_i = Q_i^{1/2} Q_i^{1/2}$. And $\Omega_i^{1/2}$ can be decomposed as

$$\Omega_i^{1/2} = \begin{bmatrix} N_1' \\ N_2' \end{bmatrix}$$

where $N_1'$ is a $1 \times (p + 1)$ vector and $N_2'$ is $p \times (p + 1)$. Note that, since $\Omega_i^{1/2}$ is symmetric

$$\Omega_i = \begin{bmatrix} N_1' N_1 & N_1' N_2 \\ N_2' N_1 & N_2' N_2 \end{bmatrix} = \begin{bmatrix} \sigma_i^2 & \gamma_i' \\ \gamma_i & \Sigma_i \end{bmatrix}$$

Define

$$\xi = \frac{\delta_0' \Delta_0' Q_2 \Delta_0 \delta_0}{\delta_0' \Delta_0' Q_1 \Delta_0 \delta_0} \text{ and } \phi = \frac{\delta_0' \Delta_0' \Phi_2 \Delta_0 \delta_0}{\delta_0' \Delta_0' \Phi_1 \Delta_0 \delta_0}$$

where $\Phi_i = [(N_i' + N_2' \beta_i') \otimes Q_i^{1/2}] [(N_i' + N_2' \beta_i') \otimes Q_i^{1/2}]'$ for $i = 1, 2$.

Let $W_i(s)^{vec} \equiv \sqrt{s} B_i(1)$ and $W_i(s)$ for $i = 1, 2$ to be independent vector and scale Brownian motion processes, respectively, defined on $[0, \infty)$, starting at the origin when $s = 0$.

Let

$$Z(s) = \begin{cases} W_1(-s) - |s|/2 & : s \leq 0 \\
\sqrt{s} W_2(s) - \xi s/2 & : s > 0 \end{cases} \quad (5.34)$$
Theorem 6  Under Assumptions 1~5 and 9~11

\[ \frac{(\theta_T^0 \Delta_0^t Q_1 \Delta_0^t \theta_T^0)^2}{\theta_T^0 \Delta_0^t \Phi_1 \Delta_0^t \theta_T^0} (\hat{k} - k_0) \xrightarrow{d} \arg \max_s Z(s) \]

Proof

Theorem 5 proved that \( \hat{k} = k_0 + O_p(\|\theta_T^0\|^2) \). Since \( \theta_T^0 = \theta_0 v_T \) under Assumption 10, we have \( \hat{k} = k_0 + O_p(v_T^{-2}) \).

For the proof of the current Proposition, we will derive the limiting process of \( V_T(k) - V_T(k_0) \) for \( k = k_0 + [sv_T^{-2}] \) and \( s \in [-C, C] \). Then, the continuous mapping theorem can be invoked for the final step of the proof. That is, the limiting distribution of \( \hat{k} \) may be obtained by the local weak convergence of \( V_T(k) - V_T(k_0) \) in conjunction with the continuous mapping theorem for the arg max functional.

We first consider \( s \leq 0 \). That is, \( k \leq k_0 \).

From Lemma 7,

\[
V_T(k) - V_T(k_0) = -\theta_0^t W_{\Delta} W_{\Delta} \theta_0^t + 2\theta_0^t W_{\Delta} \tilde{U} \Xi + o_p(1)
\]

\[
= -\theta_0^t (v_T^2 \sum_{t=k+1}^{k_0} w_t w_t^t) \theta_0 + 2\theta_0^t (v_T \sum_{t=k+1}^{k_0} w_t \tilde{u}_t) + o_p(1)
\]

\[
= -\theta_0^t \Delta_0^t (v_T^2 \sum_{t=k+1}^{k_0} z_t z_t^t) \Delta_0 \theta_0 + 2\theta_0^t \Delta_0^t (v_T \sum_{t=k+1}^{k_0} z_t \tilde{u}_t) + o_p(1)
\]

\[(5.35)\]

For the first term in (5.35) we know that it involves \( [sv_T^{-2}] \) (the absolute value of \( [sv_T^{-2}] \)) observations of \( z_t \). Thus, by Assumptions 2, 5 and 9,

\[
v_T^2 \sum_{t=k+1}^{k_0} z_t z_t^t \Rightarrow |s|Q_1 \quad (5.36)
\]

The second term in (5.35) can be further expanded to

\[
v_T \sum_{t=k+1}^{k_0} z_t \tilde{u}_t = v_T \sum_{t=k+1}^{k_0} z_t u_t + v_T \sum_{t=k+1}^{k_0} z_t v_t' \beta_1^0
\]

95
\[-v_T \sum_{t=k+1}^{k_0} z_t z'_t [(Z'Z)^{-1} Z'V] \beta_1^0 \quad (5.37)\]

We investigate the limit behavior of each term in (5.37) one by one. Let’s consider the first term, \(v_T \sum_{t=k+1}^{k_0} z_t u_t\). By the functional central limit theorem of Assumption 1,

\[-sv_T^{-2} \sum_{t=k+1}^{k_0} z_t u_t \Rightarrow (N_1' \otimes Q_1^{1/2}) B_1(1)\]

Denoting the rescaled version of the standard Brownian motion process by \(W_{vec}(t) = \sqrt{t} B_1(1)\), we can rewrite above convergence result as

\[-v_T \sum_{t=k+1}^{k_0} z_t u_t \Rightarrow \sqrt{-s} (N_1' \otimes Q_1^{1/2}) B_1(1)\]

\[= (N_1' \otimes Q_1^{1/2}) W_{vec}(-s)\]

Similarly, by denoting \(W_{max}(t) = \sqrt{t} B_{1max}(1)\) the limit process of the second term of (5.37) can be shown to be

\[v_T \sum_{t=k+1}^{k_0} z_t v'_t \Rightarrow Q_1^{1/2} W_{max}(-s) N_2\]

Finally, the last term of (5.37) disappears in the limit.

\[v_T \sum_{t=k+1}^{k_0} z_t z'_t [(Z'Z)^{-1} Z'V] \beta_1^0 = T^{-1/2} v_T^{-1} s \cdot (sv_T^{-2})^{-1} \sum_{t=k+1}^{k_0} z_t z'_t \times [(T^{-1} Z'Z)^{-1} T^{-1/2} Z'V] \beta_1^0\]

\[= T^{-1/2} v_T^{-1} O_p(1) = o_p(1)\]

Thus, combining the results on (5.37), we obtain

\[v_T \sum_{t=k+1}^{k_0} z_t \tilde{u}_t \Rightarrow (N_1' \otimes Q_1^{1/2}) W_1(-s) + Q_1^{1/2} W_{max}(-s) N_2^1 \beta_1^0\]
We can further simplify the expression of the limit process by using \( \text{vec}(A_1 A_2 A_3) = (A'_3 \otimes A_1) \text{vec}(A_2) \). In other words, we can rewrite

\[
Q_{1/2} W_1^{\text{mat}} (-s) N_2^{1/2} \beta_1^0 = \text{vec}(Q_{1/2} W_1^{\text{mat}} (-s) N_2^{1/2} \beta_1^0)
\]

\[
= (\beta_1^{0'} N_2^{1'} \otimes Q_{1/2}^{1/2}) \text{vec}(W_1^{\text{mat}}(-s))
\]

\[
= (\beta_1^{0'} N_2^{1'} \otimes Q_{1/2}^{1/2}) W_1^{\text{vec}}(-s)
\]

Thus,

\[
v_T \sum_{t=k+1}^{k_0} z_t \tilde{u}_t \implies [(N_1^{1'} \otimes Q_{1/2}^{1/2}) + (\beta_1^{0'} N_2^{1'} \otimes Q_{1/2}^{1/2})] W_1^{\text{vec}}(-s)
\]

\[
= [(N_1^{1'} + \beta_1^{0'} N_2^{1'}) \otimes Q_{1/2}^{1/2}] W_1^{\text{vec}}(-s)
\]

\[
= [(N_1 + N_2^{1'}) \beta_1^{0'} \otimes Q_{1/2}^{1/2}] W_1^{\text{vec}}(-s)
\]

(5.38)

Let \( \Phi_1 = [(N_1^{1'} + N_2^{1'}) \beta_1^{0'} \otimes Q_{1/2}^{1/2}] [(N_1^{1'} + N_2^{1'}) \beta_1^{0'} \otimes Q_{1/2}^{1/2}]' \). Then \( \theta_0' \Delta_0' (v_T \sum_{t=k+1}^{k_0} z_t \tilde{u}_t) \)

has an asymptotic distribution of \( (\theta_0' \Delta_0' \Phi_1 \Delta_0 \theta_0)^{1/2} W_1(-s) \) where \( W_1(\cdot) \) is a rescaled Brownian motion process defined on \([0, \infty)\).

Thus, it follows that for \( s \leq 0 \),

\[
V_T(k) - V_T(k_0) = V_T(k_0 + \lceil sv_T^2 \rceil) - V_T(k_0)
\]

\[
\Rightarrow -s|\theta_0' \Delta_0' Q_1 \Delta_0 \theta_0 + 2\theta_0' \Delta_0' [(N_1^{1'} + N_2^{1'}) \beta_1^{0'} \otimes Q_{1/2}^{1/2}] W_1^{\text{vec}}(-s)
\]

\[
d \equiv -s|\theta_0' \Delta_0' Q_1 \Delta_0 \theta_0 + 2(\theta_0' \Delta_0' \Phi_1 \Delta_0 \theta_0)^{1/2} W_1(-s)
\]

where \( \equiv \) denotes “distributed as”.

Similarly, for \( s > 0 \)

\[
V_T(k_0 + \lceil sv_T^2 \rceil) - V_T(k_0)
\]

97
\[ -|s|\theta_0'\Delta_0'Q_2\Delta_0\theta_0 + 2(\theta_0'\Delta_0'\Phi_2\Delta_0\theta_0)^{1/2}W_2(s) \]

where \( W_2(\cdot) \) is another Brownian motion process on \([0, \infty)\). The two processes \( W_1 \) and \( W_2 \) are independent because they are the limiting processes corresponding to the asymptotically independent regimes.

In summary,

\[ \begin{align*}
V_T(k_0 + [sv_T^{-2}]) - V_T(k_0) & \implies G(s) \\
& \equiv \begin{cases} 
-|s|\theta_0'\Delta_0'Q_1\Delta_0\theta_0 + 2(\theta_0'\Delta_0'\Phi_1\Delta_0\theta_0)^{1/2}W_1(-s) & : s \leq 0 \\
-|s|\theta_0'\Delta_0'Q_2\Delta_0\theta_0 + 2(\theta_0'\Delta_0'\Phi_2\Delta_0\theta_0)^{1/2}W_2(s) & : s > 0
\end{cases}
\end{align*} \]

Now, we can invoke the continuous mapping theorem to conclude

\[ v_T^2(\hat{k} - k_0) \longrightarrow_d \arg \max_s G(s) \]

However, by a change of variable \( s = bv \) with

\[ b = \frac{\theta_0'\Delta_0'\Phi_1\Delta_0\theta_0}{(\theta_0'\Delta_0'Q_1\Delta_0\theta_0)^2} \]

it can be shown that \( \arg \max_s G(s) = b \cdot \arg \max_v Z(v) \) where \( Z(v) \) is defined in equation (5.34). In the following, we show the validity of this argument.

For \( s \leq 0 \)

\[ G(s) = -|s|\theta_0'\Delta_0'Q_1\Delta_0\theta_0 + 2(\theta_0'\Delta_0'\Phi_1\Delta_0\theta_0)^{1/2}W_1(-s) \]

\[ = -|bv| \cdot \theta_0'\Delta_0'Q_1\Delta_0\theta_0 + 2(\theta_0'\Delta_0'\Phi_1\Delta_0\theta_0)^{1/2}W_1(-bv) \]

\[ = -|v|b \cdot \theta_0'\Delta_0'Q_1\Delta_0\theta_0 + 2(\theta_0'\Delta_0'\Phi_1\Delta_0\theta_0)^{1/2} \sqrt{b} \cdot W_1(-v) \]

\[ = -|v| \frac{\theta_0'\Delta_0'\Phi_1\Delta_0\theta_0}{(\theta_0'\Delta_0'Q_1\Delta_0\theta_0)^2} \cdot \theta_0'\Delta_0'Q_1\Delta_0\theta_0 + 2(\theta_0'\Delta_0'\Phi_1\Delta_0\theta_0)^{1/2} \]

98
Finally, the statement in Theorem 6 can be established in the following way. Since

Thus,

\[
\arg \max_s G(s) = \arg \max_v \left\{ -|v| \frac{\theta'_0 \Delta'_0 \Phi_1 \Delta_0 \theta_0}{\theta'_0 \Delta'_0 Q_1 \Delta_0 \theta_0} + 2 \frac{\theta'_0 \Delta'_0 \Phi_1 \Delta_0 \theta_0}{\theta'_0 \Delta'_0 Q_1 \Delta_0 \theta_0} W_1(-v) \right\}
\]

\[
= \arg \max_v \left\{ -|v| \frac{\theta'_0 \Delta'_0 \Phi_1 \Delta_0 \theta_0}{\theta'_0 \Delta'_0 Q_1 \Delta_0 \theta_0} \right\}
\]

Thus,

Similarly, for \( s > 0 \)

\[
G(s) = -s \cdot \theta'_0 \Delta'_0 Q_2 \Delta_0 \theta_0 + 2(\theta'_0 \Delta'_0 \Omega_2 \Delta_0 \theta_0)^{1/2} W_2(s)
\]

\[
= -b \cdot \theta'_0 \Delta'_0 Q_2 \Delta_0 \theta_0 + 2(\theta'_0 \Delta'_0 \Omega_2 \Delta_0 \theta_0)^{1/2} \sqrt{b} W_2(v)
\]

\[
= -v \frac{\theta'_0 \Delta'_0 \Phi_1 \Delta_0 \theta_0}{(\theta'_0 \Delta'_0 Q_1 \Delta_0 \theta_0)^2} \theta'_0 \Delta'_0 Q_2 \Delta_0 \theta_0 + 2(\theta'_0 \Delta'_0 \Omega_2 \Delta_0 \theta_0)^{1/2}
\]

\[
\times \left( \frac{\theta'_0 \Delta'_0 \Phi_1 \Delta_0 \theta_0}{\theta'_0 \Delta'_0 Q_1 \Delta_0 \theta_0} \right)^{1/2} W_2(v)
\]

\[
= \frac{\theta'_0 \Delta'_0 \Phi_1 \Delta_0 \theta_0}{\theta'_0 \Delta'_0 Q_1 \Delta_0 \theta_0} \left[ -\theta'_0 \Delta'_0 Q_2 \Delta_0 \theta_0 \frac{v}{\theta'_0 \Delta'_0 Q_1 \Delta_0 \theta_0} + 2 \left( \frac{\theta'_0 \Delta'_0 \Phi_1 \Delta_0 \theta_0}{\theta'_0 \Delta'_0 Q_1 \Delta_0 \theta_0} \right)^{1/2} W_2(v) \right]
\]

\[
= \frac{\theta'_0 \Delta'_0 \Phi_1 \Delta_0 \theta_0}{\theta'_0 \Delta'_0 Q_1 \Delta_0 \theta_0} \left[ -\xi v + 2 \sqrt{\phi} W_2(v) \right]
\]

Thus,

\[
\arg \max_s G(s) = \arg \max_v \left\{ -\frac{\xi v}{2} + \sqrt{\phi} W_2(v) \right\} \frac{\theta'_0 \Delta'_0 \Phi_1 \Delta_0 \theta_0}{\theta'_0 \Delta'_0 Q_1 \Delta_0 \theta_0}
\]

\[
= \arg \max_v \left\{ -\frac{\xi v}{2} + \sqrt{\phi} W_2(v) \right\}
\]

Finally, the statement in Theorem 6 can be established in the following way. Since
\( V_T(k_0 + [sv_T^2]) - V_T(k_0) \Rightarrow G(s) \) and \( \arg \max_s G(s) = b \cdot \arg \max_v Z(v) \), we have \( b^{-1}v_T^2(\hat{k} - k_0) \longrightarrow_d \arg \max_v Z(v) \). But

\[
\begin{align*}
b^{-1}v_T^2 &= \frac{(\theta_0' \Delta_0 Q_1 \Delta_0 \theta_0)^2}{\theta_0' \Delta_0 \Phi_1 \Delta_0 \theta_0} v_T^2 \\
&= \frac{(v_T^{-1}\theta_T^\prime \Delta_0 Q_1 \Delta_0 v_T^{-1} \theta_T^0)^2}{v_T^{-1}\theta_T^\prime \Delta_0 \Phi_1 \Delta_0 v_T^{-1} \theta_T^0} v_T^2 \\
&= \frac{(\theta_T^\prime \Delta_0 Q_1 \Delta_0 \theta_T^0)^2}{\theta_T^\prime \Delta_0 \Phi_1 \Delta_0 \theta_T^0}
\end{align*}
\]

where the second equality comes from Assumption 10.

Thus,

\[
b^{-1}v_T^2(\hat{k} - k_0) = \frac{(\theta_T^\prime \Delta_0 Q_1 \Delta_0 \theta_T^0)^2}{\theta_T^\prime \Delta_0 \Phi_1 \Delta_0 \theta_T^0} (\hat{k} - k_0) \longrightarrow_d \arg \max_s Z(s)
\]

as claimed.
Chapter 6

Test Statistics for Multiple Breaks

Much of the literature concerned with estimation and inference from a sample of economic data deals with a situation when the statistical model is correctly specified. Consequently, it is customary to assume that the parameterized linear statistical model used for purposes of inference is consistent with the sampling process from which the sample observations were generated. However, the possibilities for model misspecification are numerous and false statistical models are most likely the rule rather than the exception. Several diagnostic tests have been developed to assess the validity of model specification. The most popular test among them is the over-identifying restriction test developed by Sargan(1958) for instrumental variable estimation in linear models and it has been subsequently generalized to GMM by Hansen(1982). But this strand of tests possessed low local power against many misspecifications of interest. This lack of power suggests the importance of complementing the over-identifying restriction test with tests specifically designed to test for structural instability. Bai and Perron(1998) develop methods that are designed to test for structural stability with an unknown
number of break points in the sample. Their analysis is in the context of linear regression models estimated via Ordinary Least Squares (OLS). We extend Bai and Perron’s framework for multiple break testing to linear models via Two Stage Least Squares (2SLS).

6.1 SupF Test

The sup F type test of no structural break \( (m = 0) \) versus the alternative hypothesis that there is \( m = 1 \) break has been considered by Andrews (1993). Bai and Perron (1998) generalized Andrew’s sup F type test to the hypothesis \( m = k \). i.e. multiple breaks. In their generalization, they relied on the least-squares principle for the estimation of the unknown regression coefficients. In this chapter, we extend Bai and Perron’s OLS results to the 2SLS setup where the endogenous explanatory variables in our regression model are correlated with the regression residuals.

For the test of null hypothesis of no break \( (m = 0) \) versus the alternative hypothesis of \( k \) breaks \( (m = k) \) in the population model, the sup F type test statistic can be defined as follows:

Let \( (T_1, \cdots, T_k) \) be a partition such that \( T_i = [T \lambda_i] \ (i = 1, \cdots, k) \). Define

\[
F_T(\lambda_1, \cdots, \lambda_k; p) = \left( \frac{T - (k + 1)p}{kp} \right) \frac{SSR_0 - SSR_k}{SSR_k}
\]

where \( SSR_0 \) and \( SSR_k \) are the sum of squared residuals based on the fitted \( X \) under null and alternative hypothesis, respectively.

For asymptotic analysis, we need to impose some restrictions on the possible
values of the break dates. Specifically, each break date should be asymptotically distinct and bounded from the boundaries of the sample. Thus, we define
\[ \Lambda_\epsilon = \{(\lambda_1, \cdots, \lambda_k) : |\lambda_{i+1} - \lambda_i| \geq \epsilon, \lambda_1 \geq \epsilon, \lambda_k \leq 1 - \epsilon\} \]
for some arbitrary small positive number \( \epsilon \).

Finally, the sup F type test statistic is defined as
\[ \sup F_T(k; p) = \sup_{(\lambda_1, \cdots, \lambda_k) \in \Lambda_\epsilon} F_T(\lambda_1, \cdots, \lambda_k; p) \]

The limiting distribution of the test depends on the nature of the regressors and the residuals. In the following we impose conditional homoscedasticity assumption (Assumption 11).

**Theorem 7** If the data are generated by (3.1)-(3.2) with \( m = 0 \) and Assumptions 1-5 and 11 hold then
\[ \sup F_T(k; p) \Rightarrow \sup F_{k,p} \equiv \sup_{(\lambda_1, \cdots, \lambda_k) \in \Lambda_\epsilon} F(\lambda_1, \cdots, \lambda_k; p) \]

with
\[ F(\lambda_1, \cdots, \lambda_k; p) = \frac{1}{kp} \sum_{i=1}^{k} \frac{||\lambda_{i+1}W(\lambda_i) - \lambda_iW(\lambda_{i+1})||^2}{\lambda_i\lambda_{i+1}(\lambda_{i+1} - \lambda_i)} \]
where \( k \) is the number of break points under the alternative hypothesis, and \( W(\lambda) \) is \( p \times 1 \) vector standard Brownian motion process.

**proof**

By defining \( SSR_0 \) and \( SSR_k \) as the sum of squared residuals based on the fitted \( X \) under null and alternative hypothesis, respectively, we can write
\[ F_T(\lambda_1, \cdots, \lambda_k; p) = (SSR_0 - SSR_k) / [kp(T - (k + 1)p)^{-1}SSR_k] \quad (6.1) \]
First, we consider the limit behavior of the numerator, $F_T^* \equiv SSR_0 - SSR_k$, of
$F_T(\lambda_1, \ldots, \lambda_k; p)$. To this end, we define $D^R(i, j)$ to be sum of squared residuals
from the restricted model using data from segments $i$ to $j$, i.e., from observation
$T_{i-1} + 1$ to $T_j$. Similarly, $D^U(i, i)$ for the unrestricted model.

Using this notation, we can write $F_T^*$ as follows:

$$F_T^* = D^R(1, k + 1) - \sum_{i=1}^{k+1} D^U(i, i)$$

$$= \sum_{i=1}^{k} [D^R(1, i + 1) - D^R(1, i) - D^U(i + 1, i + 1)] + D^R(1, 1) - D^U(1, 1)$$

$$= \sum_{i=1}^{k} [D^R(1, i + 1) - D^R(1, i) - D^U(i + 1, i + 1)] \quad (6.2)$$

$$= \sum_{i=1}^{k} F_{T,i} \text{ say.} \quad (6.3)$$

To analyze the limit behavior of the terms on the right hand side of (6.2), we
introduce the following notations for the least squares estimators.

The restricted estimator based on the full sample is

$$\hat{\beta}^R = (W'W)^{-1}W'Y$$

The unrestricted estimator based on the full sample is

$$\hat{\beta}^U = (\bar{W}'\bar{W}^*)^{-1}\bar{W}'Y$$

where $\bar{W}^*$ is the diagonal partition of $W = (\hat{x}_1, \ldots, \hat{x}_T)'$ at $(\hat{T}_1, \ldots, \hat{T}_m)$.

The least squares estimator of the common regression parameter under $H_0$ based
on segment 1 through $j$ of the partition,

$$\hat{\beta}_{1,j}^R = (W'_{1,j}W_{1,j})^{-1}W'_{1,j}Y_{1,j}$$
where \( Y_{1,j}, \tilde{U}_{1,j}, W_{1,j} \) denote the matrices (vectors) consisting of the rows 1 through \( T_j \) of \( Y, \tilde{U}, W \), respectively.

The least squares estimator based on the observations in the \( j^{th} \) segment of the partition,

\[
\hat{\beta}_j^U = (W_j'W_j)^{-1}W_j'Y_j
\]

where \( Y_j, U_j, W_j \) are the matrices (vectors) containing row \( T_{j-1} + 1 \) through \( T_j \) of \( Y, \tilde{U}, W \), respectively.

Note that under the null hypothesis that \( \beta_0 = \bar{\beta}_0 \) in (3.1) for \( i = 1, 2, \ldots, k+1 \), we have

\[
Y = W\bar{\beta}_0 + \tilde{U}
\]

\[
= \bar{W}^0(\iota_{k+1} \otimes \bar{\beta}_0) + \tilde{U}
\]

and

\[
Y_j = W_j\bar{\beta}_0 + \tilde{U}_j \quad (6.4)
\]

where \( \iota_{k+1} \) is a \((k+1) \times 1\) vector of ones. Thus, we have the sum of squared residuals

\[
D^R(1,j) = ||(I - P_{W_{1,j}})\tilde{U}_{1,j}||^2 \quad (6.5)
\]

\[
D^U(j,j) = ||(I - P_{W_j})\tilde{U}_j||^2 \quad (6.6)
\]

where \( P_{W_{1,j}} = W_{1,j}(W_{1,j}'W_{1,j})^{-1}W_{1,j}' \) and \( P_{W_j} = W_j(W_j'W_j)^{-1}W_j' \).

Now Consider \( F_{T,i} \) in (6.3). Using (6.2), (6.5) and (6.6), we have

\[
F_{T,i} = D^R(1, i + 1) - D^R(1, i) - D^U(i + 1, i + 1)
\]

\[
= ||(I - P_{W_{i+1}})\tilde{U}_{i,i+1}||^2 - ||(I - P_{W_{1,i}})\tilde{U}_{1,i}||^2 - ||(I - P_{W_{i+1}})\tilde{U}_{i+1}||^2
\]
To simplify the exposition in the following derivation, define

\[ S_j \equiv W_{1,j}^t \tilde{U}_{1,j}, \quad H_j \equiv W_{1,j}^t W_{1,j}. \]

Then \( F_{T,i} \) can be rewritten as

\[
F_{T,i} = \tilde{U}_{1,i+1}^t (I - P_{W_{1,i+1}}) \tilde{U}_{1,i+1} - \tilde{U}_{1,i}^t (I - P_{W_{1,i}}) \tilde{U}_{1,i} - \tilde{U}_{1,i+1}^t (I - P_{W_{1,i+1}}) \tilde{U}_{1,i+1} \\
= \tilde{U}_{1,i+1}^t \tilde{U}_{1,i+1} - \tilde{U}_{1,i}^t \tilde{U}_{1,i+1} P_{W_{1,i+1}} \tilde{U}_{1,i+1} + \tilde{U}_{1,i}^t P_{W_{1,i}} \tilde{U}_{1,i+1} \\
- \tilde{U}_{1,i+1}^t \tilde{U}_{i+1} + \tilde{U}_{i+1}^t P_{W_{1,i+1}} \tilde{U}_{i+1} \\
= - \tilde{U}_{1,i+1}^t P_{W_{1,i+1}} \tilde{U}_{1,i+1} + \tilde{U}_{1,i}^t P_{W_{1,i}} \tilde{U}_{1,i} + \tilde{U}_{i+1}^t P_{W_{1,i+1}} \tilde{U}_{i+1} \\
= - S_{i+1}^t H_{i+1}^{-1} S_{i+1} + S_i^t H_i^{-1} S_i + (S_{i+1} - S_i)^t (H_{i+1} - H_i)^{-1} (S_{i+1} - S_i)
\]

(6.7)

The limiting behavior of \( F_{T,i} \) is deduced from the limiting behavior of \( S_j \) and \( H_j \).

To proceed further it is useful to explore the implications of Assumptions 1 and 67. Let \( B(r) = [B_1(r)', B_2(r)', \ldots, B_{p+1}(r)']' \) where \( B_i(r)' \) is \( q \times 1 \), and

\[
\Omega^{1/2} = \begin{bmatrix} N_1' \\ N_2' \end{bmatrix}
\]

(6.8)

with \( N_1' \) is a \( 1 \times (p + 1) \) vector whose \( i^{th} \) element is \( N_{1,i} \), and \( N_2' \) is \( p \times (p + 1) \).

Then under Assumptions 1 and 67

\[
T^{-1/2} \sum_{t=1}^{[T]} \begin{pmatrix} u_t \\ v_t \end{pmatrix} \otimes z_t \Longrightarrow (\Omega^{1/2} \otimes Q_{ZZ}^{1/2}) B(r)
\]

(6.9)

where \( \Omega^{1/2} \) and \( Q_{ZZ}^{1/2} \) are the symmetric matrices satisfying \( \Omega = \Omega^{1/2} \Omega^{1/2} \) and \( Q_{ZZ} = Q_{ZZ}^{1/2} Q_{ZZ}^{1/2} \). Note that

\[
\Omega = \begin{bmatrix} N_1' N_1 & N_1' N_2 \\ N_2' N_1 & N_2' N_2 \end{bmatrix} = \begin{bmatrix} \sigma & \gamma' \\ \gamma & \Sigma \end{bmatrix}
\]

(6.10)
where the second and third matrices are partitioned conformably. It follows from (6.9) and (6.8) that

\[
T^{-1/2} \sum_{t=1}^{[Tr]} z_t u_t = (N'_1 \otimes Q_{ZZ}^{1/2}) B(r) = \sum_{i=1}^{p+1} N_{1,i} Q_{ZZ}^{1/2} B_i(r) = Q_{ZZ}^{1/2} \sum_{i=1}^{p+1} N_{1,i} B_i(r) = Q_{ZZ}^{1/2} \tilde{D}^*(r), \text{ say}
\]

and

\[
T^{-1/2} \sum_{t=1}^{[Tr]} v_t \otimes z_t = T^{-1/2} \sum_{t=1}^{[Tr]} \text{vec}(z_t v'_t) = (N'_2 \otimes Q_{ZZ}^{1/2}) B(r) \tag{6.11}
\]

Note that (6.11) implies

\[
T^{-1/2} \sum_{t=1}^{[Tr]} z_t v'_t \Rightarrow Q_{ZZ}^{1/2} B^{\text{mat}}(r) N_2 = Q_{ZZ}^{1/2} \tilde{D}^*(r), \text{ say} \tag{6.12}
\]

where \( \text{vec}(B^{\text{mat}}) = B(r) \).

To deduce the limiting behavior of \( S_j \) with \( \beta^0_i = \bar{\beta}_0 \) we note that:

\[
T^{-1/2} S_j = T^{-1/2} W_{1,j} \tilde{U}_{1,j} = T^{-1/2} \sum_{t=1}^{[T \lambda_j]} \hat{x}_t \tilde{u}_t
\]

\[
= T^{-1/2} [\Delta_0' \sum_{t=1}^{[T \lambda_j]} z_t u_t + V'Z(Z'Z)^{-1} \sum_{t=1}^{[T \lambda_j]} z_t u_t + \Delta_0' \sum_{t=1}^{[T \lambda_j]} z_t v'_t \bar{\beta}_0 + V'Z(Z'Z)^{-1} \sum_{t=1}^{[T \lambda_j]} z_t v'_t \bar{\beta}_0 - \Delta_0' \sum_{t=1}^{[T \lambda_j]} z_t (Z'Z)^{-1} Z'V \bar{\beta}_0 - V'Z(Z'Z)^{-1} \sum_{t=1}^{[T \lambda_j]} z_t (Z'Z)^{-1} Z'V \bar{\beta}_0 + V'Z(Z'Z)^{-1} \sum_{t=1}^{[T \lambda_j]} z_t (Z'Z)^{-1} Z'V \bar{\beta}_0] \]

107
we have
\[ T^{-1}H_j = T^{-1}W_{i,j}W_{1,j} = T^{-1} \sum_{t=1}^{\lceil T \lambda \rceil} \hat{x}_t \hat{x}_t' \]
\[ = T^{-1} \sum_{t=1}^{\lceil T \lambda \rceil} \hat{\Delta}_T z_t (\hat{\Delta}_T z_t)' \]
\[ = \hat{\Delta}_T T^{-1} \sum_{t=1}^{\lceil T \lambda \rceil} z_t z_t' \hat{\Delta}_T \]
\[ \Rightarrow \Delta_0' (\lambda_j Q_{ZZ}) \Delta_0 \] (6.14)

We now use (6.13) and (6.14) to deduce the limiting behavior of the terms on the right hand side of (6.7). First consider \( S'_{i+1} H_{i+1}^{-1} S_{i+1} \). From (6.13) and (6.14), we have

\[ S'_{i+1} H_{i+1}^{-1} S_{i+1} \Rightarrow (\Delta_0' Q_{ZZ}^{1/2} \tilde{D}^* (\lambda_{i+1}) + \Delta_0' Q_{ZZ}^{1/2} D^* (\lambda_{i+1}) \tilde{b}_0 - \Delta_0' \lambda_{i+1} Q_{ZZ}^{1/2} D^* (1) \tilde{b}_0)' \]
\[ \times (\Delta_0' \lambda_{i+1} Q_{ZZ} \Delta_0)^{-1} (\Delta_0' Q_{ZZ}^{1/2} \tilde{D}^* (\lambda_{i+1}) + \Delta_0' Q_{ZZ}^{1/2} D^* (\lambda_{i+1}) \tilde{b}_0 - \Delta_0' \lambda_{i+1} Q_{ZZ}^{1/2} D^* (1) \tilde{b}_0) \]
\[ = (\Delta_0' Q_{ZZ}^{1/2} \tilde{D}^* (\lambda_{i+1}) + D^* (\lambda_{i+1}) \tilde{b}_0 - \lambda_{i+1} D^* (1) \tilde{b}_0)' (\Delta_0' \lambda_{i+1} Q_{ZZ} \Delta_0)^{-1} (\Delta_0' Q_{ZZ}^{1/2} \tilde{D}^* (\lambda_{i+1}) + \Delta_0' Q_{ZZ}^{1/2} D^* (\lambda_{i+1}) \tilde{b}_0 - \Delta_0' \lambda_{i+1} Q_{ZZ}^{1/2} D^* (1) \tilde{b}_0) \]
\[
\times \Delta_0^{-1}(\Delta_0'Q_{ZZ}'D^*(\lambda_{i+1}) + D^*(\lambda_{i+1})\beta_0 - \lambda_{i+1}D^*(1)\beta_0)
\]

\[
= \lambda_{i+1}^{-1}[\tilde{D}^*(\lambda_{i+1}) + D^*(\lambda_{i+1})\beta_0 - \lambda_{i+1}D^*(1)\beta_0]'
\times (\Delta_0'Q_{ZZ}'(\Delta_0'Q_{ZZ}\Delta_0)^{-1}(\Delta_0'Q_{ZZ}'))
\times [\tilde{D}^*(\lambda_{i+1}) + D^*(\lambda_{i+1})\beta_0 - \lambda_{i+1}D^*(1)\beta_0]
\]

(6.15)

Since \((\Delta_0'Q_{ZZ}^{-1/2})'(\Delta_0'Q_{ZZ}\Delta_0)^{-1}(\Delta_0'Q_{ZZ}^{-1/2})\) is a projection matrix, we have

\[
(\Delta_0'Q_{ZZ}^{-1/2})'(\Delta_0'Q_{ZZ}\Delta_0)^{-1}(\Delta_0'Q_{ZZ}^{-1/2}) = C'\Lambda C = C'\Lambda'\Lambda C
\]

\[
= (\Lambda C)'\Lambda C
\]

(6.16)

where \(C\) is an orthogonal matrix whose columns consist of the orthogonal, normalized eigenvectors of the matrix of LHS expression and \(\Lambda\) is a diagonal matrix, \(p\) of whose diagonal elements are ones with the remaining \(q-p\) equal to zero. Note that from the properties of a projection matrix, which are symmetric and idempotent, we have

\[
\text{rank}[(\Delta_0'Q_{ZZ}^{-1/2})'(\Delta_0'Q_{ZZ}\Delta_0)^{-1}(\Delta_0'Q_{ZZ}^{-1/2})] = \text{tr}[(\Delta_0'Q_{ZZ}^{-1/2})'(\Delta_0'Q_{ZZ}\Delta_0)^{-1}(\Delta_0'Q_{ZZ}^{-1/2})]
\]

\[
= \text{tr}[(\Delta_0'Q_{ZZ}\Delta_0)^{-1}(\Delta_0'Q_{ZZ}^{-1/2}Q_{ZZ}^{1/2})]
\]

\[
= \text{tr}[(\Delta_0'Q_{ZZ}\Delta_0)^{-1}(\Delta_0'Q_{ZZ}\Delta_0)]
\]

\[
= \text{tr}[I_p] = p
\]

Thus, using (6.16) we have

\[
S_{i+1}'H_{i+1}^{-1}S_{i+1} \Rightarrow \lambda_{i+1}^{-1}(\sigma\tilde{D}^*(\lambda_{i+1}) + [D^*(\lambda_{i+1}) - \lambda_{i+1}D^*(1)]\beta_0)'C'\Lambda C
\]

\[
\times (\tilde{D}^*(\lambda_{i+1}) + [D^*(\lambda_{i+1}) - \lambda_{i+1}D^*(1)]\beta_0)
\]

\[
= \lambda_{i+1}^{-1}(\tilde{D}^*(\lambda_{i+1}) + [D^*(\lambda_{i+1}) - \lambda_{i+1}D^*(1)]\beta_0)'(\Lambda C)'\Lambda C
\]

\[
\times (\tilde{D}^*(\lambda_{i+1}) + [D^*(\lambda_{i+1}) - \lambda_{i+1}D^*(1)]\beta_0)
\]

109
\[ \lambda_{i+1}^{-1}(\Delta \hat{D}^*(\lambda_{i+1}) + \Lambda C[D^*(\lambda_{i+1}) - \lambda_{i+1}D^*(1)]\bar{\beta}_0)'^{i} \times (\Delta \hat{D}^*(\lambda_{i+1}) + \Lambda C[D^*(\lambda_{i+1}) - \lambda_{i+1}D^*(1)]\bar{\beta}_0) \]

Similarly, we have

\[ S_i' H_i^{-1} S_i \Rightarrow \lambda_i^{-1}(\Delta \hat{D}^*(\lambda_i) + \Lambda C[D^*(\lambda_i) - \lambda_iD^*(1)]\bar{\beta}_0)'^{i} \times (\Delta \hat{D}^*(\lambda_i) + \Lambda C[D^*(\lambda_i) - \lambda_iD^*(1)]\bar{\beta}_0) \]

Finally, consider \( A_i = (S_{i+1} - S_i)'(H_{i+1} - H_i)^{-1}(S_{i+1} - S_i) \). Using (6.13) and (6.14), it follows that

\[ A_i \Rightarrow [\Delta_0' Q_{\bar{Z}Z}^{1/2}((\hat{D}^*(\lambda_{i+1}) - \hat{D}^*(\lambda_i)) + \Delta_0' Q_{\bar{Z}Z}^{1/2}(D^*(\lambda_{i+1}) - D^*(\lambda_i))\bar{\beta}_0
\]
\[ - \Delta_0' (\lambda_{i+1} - \lambda_i)Q_{\bar{Z}Z}^{1/2}D^*(1)\bar{\beta}_0)'[\Delta_0' (\lambda_{i+1} - \lambda_i)Q_{\bar{Z}Z}^{1/2}\Lambda_0^{-1}[\Delta_0' Q_{\bar{Z}Z}^{1/2}((\hat{D}^*(\lambda_{i+1}) - \hat{D}^*(\lambda_i))
\]
\[ - \hat{D}^*(\lambda_i)) + \Delta_0' Q_{\bar{Z}Z}^{1/2}(D^*(\lambda_{i+1}) - D^*(\lambda_i))\bar{\beta}_0 - \Delta_0' (\lambda_{i+1} - \lambda_i)Q_{\bar{Z}Z}^{1/2}D^*(1)\bar{\beta}_0]
\]

Finally, consider \( A_i = (S_{i+1} - S_i)'(H_{i+1} - H_i)^{-1}(S_{i+1} - S_i) \). Using (6.13) and (6.14), it follows that

\[ A_i \Rightarrow [\Delta_0' Q_{\bar{Z}Z}^{1/2}((\hat{D}^*(\lambda_{i+1}) - \hat{D}^*(\lambda_i)) + (D^*(\lambda_{i+1}) - D^*(\lambda_i))\bar{\beta}_0 - (\lambda_{i+1} - \lambda_i)D^*(1)\bar{\beta}_0]
\]

Finally, consider \( A_i = (S_{i+1} - S_i)'(H_{i+1} - H_i)^{-1}(S_{i+1} - S_i) \). Using (6.13) and (6.14), it follows that

\[ A_i \Rightarrow [\Delta_0' Q_{\bar{Z}Z}^{1/2}((\hat{D}^*(\lambda_{i+1}) - \hat{D}^*(\lambda_i)) + (D^*(\lambda_{i+1}) - D^*(\lambda_i))\bar{\beta}_0 - (\lambda_{i+1} - \lambda_i)D^*(1)\bar{\beta}_0]
\]
We now use (6.17)-(6.19) to deduce the limiting behavior of $F_{T,i}$. To this end, we now write $D_i = \Lambda C D^*(\lambda_i)$, $\tilde{D}_i = \Lambda C \tilde{D}^*(\lambda_i)$, and $D_1 = \Lambda C D^*(1)$.

From (6.17)-(6.19) it follows that

$$
F_{T,i} \Rightarrow \lambda_i^{-1}[\tilde{D}_i + (D_i - \lambda_i D_1)\tilde{\beta}_0]'[\tilde{D}_i + (D_i - \lambda_i D_1)\tilde{\beta}_0] \\
- \lambda_{i+1}^{-1}[\tilde{D}_{i+1} + (D_{i+1} - \lambda_{i+1} D_1)\tilde{\beta}_0]'[\tilde{D}_{i+1} + (D_{i+1} - \lambda_{i+1} D_1)\tilde{\beta}_0] \\
+ (\lambda_{i+1} - \lambda_i)^{-1}[(\tilde{D}_{i+1} - \tilde{D}_i) + (D_{i+1} - \lambda_{i+1} D_1 - D_i + \lambda_i D_1)\tilde{\beta}_0]' \\
\times [(\tilde{D}_{i+1} - \tilde{D}_i) + (D_{i+1} - \lambda_{i+1} D_1 - D_i + \lambda_i D_1)\tilde{\beta}_0] \\
(6.20)
$$

Now we decompose the limit terms of $F_{T,i}$ into

$$
F_{T,i} \Rightarrow L_i^{(s)} + L_i^{(c)}
$$

where $L_i^{(s)}$ contains square terms in $\tilde{D}_i$, $D_i$, $D_1$, etc. and $L_i^{(c)}$ contains cross product terms in $\tilde{D}_i$, $D_i$, $D_1$, etc. Rearranging the terms in (6.20), we obtain

$$
L_i^{(s)} = \lambda_i^{-1}[\tilde{D}_i'\tilde{D}_i + \tilde{D}_i D_i \tilde{\beta}_0 + (D_i \tilde{\beta}_0)'\tilde{D}_i + (D_i \tilde{\beta}_0)'D_i \tilde{\beta}_0] \\
+ \lambda_i^2(D_i \tilde{\beta}_0)'D_i \tilde{\beta}_0 - \lambda_{i+1}^{-1}[\tilde{D}_{i+1}'\tilde{D}_{i+1} + \tilde{D}_{i+1} D_{i+1} \tilde{\beta}_0] \\
+ (D_{i+1} \tilde{\beta}_0)'\tilde{D}_{i+1} + (D_{i+1} \tilde{\beta}_0)'D_{i+1} \tilde{\beta}_0 + \lambda_{i+1}^2(D_{i+1} \tilde{\beta}_0)'D_{i+1} \tilde{\beta}_0] \\
+ (\lambda_{i+1} - \lambda_i)^{-1}[\tilde{D}_{i+1}'\tilde{D}_{i+1} + \tilde{D}_{i+1} D_{i+1} \tilde{\beta}_0 + \tilde{D}_{i+1}'D_{i+1} + \tilde{D}_{i+1}'D_{i+1} \tilde{\beta}_0] \\
+ (D_{i+1} \tilde{\beta}_0)'\tilde{D}_{i+1} + (D_{i+1} \tilde{\beta}_0)'D_{i+1} \tilde{\beta}_0 + \lambda_{i+1}^2(D_{i+1} \tilde{\beta}_0)'D_{i+1} \tilde{\beta}_0] \\
- \lambda_{i+1} \lambda_i (D_i \tilde{\beta}_0)'D_i \tilde{\beta}_0 + (B_i \tilde{\beta}_0)'\tilde{D}_i + (D_i \tilde{\beta}_0)'D_i \tilde{\beta}_0 \\
- \lambda_i \lambda_{i+1} (D_i \tilde{\beta}_0)'D_i \tilde{\beta}_0 + \lambda_i^2(D_i \tilde{\beta}_0)'D_i \tilde{\beta}_0]
$$
\[
\begin{align*}
&= (\lambda_i\lambda_{i+1}(\lambda_{i+1} - \lambda_i))^{-1}\{\lambda_{i+1}^2[\tilde{D}_i'\tilde{D}_i + \tilde{D}_i'D_i\bar{\beta}_0 + (D_i\bar{\beta}_0)'\tilde{D}_i

\begin{align*}
+ (D_i\bar{\beta}_0)'D_i\bar{\beta}_0 + \lambda_i^2(D_i\bar{\beta}_0)'D_i\bar{\beta}_0 - \lambda_{i+1}\lambda_i[\tilde{D}_i'\tilde{D}_i + \tilde{D}_i'D_i\bar{\beta}_0

\begin{align*}
+ (D_i\bar{\beta}_0)'\tilde{D}_i + (D_i\bar{\beta}_0)'D_i\bar{\beta}_0 + \lambda_i^2(D_i\bar{\beta}_0)'D_i\bar{\beta}_0

\begin{align*}
- \lambda_i\lambda_{i+1}[\tilde{D}_i'\tilde{D}_i + \tilde{D}_i'D_i\bar{\beta}_0 + (D_i\bar{\beta}_0)'\tilde{D}_i

\begin{align*}
+ (D_i\bar{\beta}_0)'D_i\bar{\beta}_0 + \lambda_i^2(D_i\bar{\beta}_0)'D_i\bar{\beta}_0

\begin{align*}
+ \lambda_i^2(D_i\bar{\beta}_0)'D_i\bar{\beta}_0 - \lambda_i\lambda_{i+1}(D_i\bar{\beta}_0)'D_i\bar{\beta}_0 + \lambda_i^2(D_i\bar{\beta}_0)'D_i\bar{\beta}_0

\begin{align*}
+ \tilde{D}_i'\tilde{D}_i + (D_i\bar{\beta}_0)'\tilde{D}_i + (D_i\bar{\beta}_0)'D_i\bar{\beta}_0

\begin{align*}
+ \lambda_i^2(D_i\bar{\beta}_0)'D_i\bar{\beta}_0

\begin{align*}
= \{\lambda_i\lambda_{i+1}(\lambda_{i+1} - \lambda_i))^{-1}\{\lambda_{i+1}^2[\tilde{D}_i'\tilde{D}_i + \tilde{D}_i'D_i\bar{\beta}_0 + (D_i\bar{\beta}_0)'\tilde{D}_i

\begin{align*}
+ (D_i\bar{\beta}_0)'D_i\bar{\beta}_0 + \lambda_i^2(D_i\bar{\beta}_0)'D_i\bar{\beta}_0

\begin{align*}
+ \lambda_i^2[D_i'\tilde{D}_i] + D_i\bar{\beta}_0]'[\tilde{D}_i + D_i\bar{\beta}_0]

\begin{align*}
+ \lambda_i^2[D_i'\tilde{D}_i + D_i\bar{\beta}_0]'[\tilde{D}_i + D_i\bar{\beta}_0]

\begin{align*}
and

\begin{align*}
L_{(c)}^i &= (\lambda_i\lambda_{i+1}(\lambda_{i+1} - \lambda_i))^{-1}\{\lambda_{i+1}^2[\tilde{D}_i'\tilde{D}_i + \tilde{D}_i'D_i\bar{\beta}_0 - \lambda_i(D_i\bar{\beta}_0)'D_i\bar{\beta}_0

\begin{align*}
- \lambda_i(D_i\bar{\beta}_0)'\tilde{D}_i - \lambda_i(D_i\bar{\beta}_0)'D_i\bar{\beta}_0 - \lambda_{i+1}\lambda_i[-\lambda_i\tilde{D}_i'\tilde{D}_i\bar{\beta}_0

\begin{align*}
- \lambda_i(D_i\bar{\beta}_0)'D_i\bar{\beta}_0 - \lambda_i(D_i\bar{\beta}_0)'\tilde{D}_i - \lambda_i(D_i\bar{\beta}_0)'D_i\bar{\beta}_0

\begin{align*}
- \lambda_i\lambda_{i+1}[\tilde{D}_i'\tilde{D}_i + \tilde{D}_i'D_i\bar{\beta}_0 - \lambda_{i+1}(D_i\bar{\beta}_0)'D_i\bar{\beta}_0 - \lambda_{i+1}(D_i\bar{\beta}_0)'\tilde{D}_i

\begin{align*}
- \lambda_{i+1}(D_i\bar{\beta}_0)'D_i\bar{\beta}_0 + \lambda_i^2[-\lambda_i\tilde{D}_i'\tilde{D}_i + D_i\bar{\beta}_0] - \lambda_{i+1}(D_i\bar{\beta}_0)'D_i\bar{\beta}_0

\begin{align*}
- \lambda_{i+1}(D_i\bar{\beta}_0)'D_i\bar{\beta}_0 - \lambda_{i+1}(D_i\bar{\beta}_0)'D_i\bar{\beta}_0 + \lambda_i\lambda_{i+1}[\tilde{D}_i'\tilde{D}_i

\begin{align*}
112
From (6.21)-(6.22) it follows that

\[ F_{T,i} \Rightarrow L^{(s)}_i + L^{(c)}_i \]

\[ = \{\lambda_i \lambda_{i+1}(\lambda_{i+1} - \lambda_i)\}^{-1}(\lambda_{i+1}^2[D_i + D_i \bar{\beta}_0][D_i + D_i \bar{\beta}_0] + \lambda_i \lambda_{i+1}(\lambda_{i+1} - \lambda_i))^{-1} \]

\[ \times (-\lambda_{i+1}D_i + D_{i+1}\bar{\beta}_0)D_{i+1}D_i\bar{\beta}_0 + \bar{\beta}_0')D_{i+1}D_i\bar{\beta}_0 + (D_{i+1}\bar{\beta}_0)'D_{i+1}D_i\bar{\beta}_0 \]

\[ + (D_{i+1}\bar{\beta}_0)'D_{i+1}D_i\bar{\beta}_0 + (D_{i+1}\bar{\beta}_0)'D_{i+1}D_i\bar{\beta}_0 \]

\[ \Rightarrow \{\lambda_i \lambda_{i+1}(\lambda_{i+1} - \lambda_i)\}^{-1}(\lambda_{i+1}^2[D_i + D_i \bar{\beta}_0] - \lambda_i[D_i + D_i \bar{\beta}_0])' \]

\[ \times (\lambda_{i+1}[D_i + D_i \bar{\beta}_0] - \lambda_i[D_i + D_i \bar{\beta}_0]) \]

\[ = \{\lambda_i \lambda_{i+1}(\lambda_{i+1} - \lambda_i)\}^{-1}([\lambda_{i+1}D_i - \lambda_iD_{i+1}] + [\lambda_{i+1}D_i - \lambda_iD_{i+1}]\bar{\beta}_0)' \]

\[ \times ([\lambda_{i+1}D_i - \lambda_iD_{i+1}] + [\lambda_{i+1}D_i - \lambda_iD_{i+1}]\bar{\beta}_0) \]

\[ = \{\lambda_i \lambda_{i+1}(\lambda_{i+1} - \lambda_i)\}^{-1}||[\lambda_{i+1}D_i - \lambda_iD_{i+1}] + [\lambda_{i+1}D_i - \lambda_iD_{i+1}]\bar{\beta}_0||^2 \]

113
Recall our test statistic

\[ F_T(\lambda_1, \cdots, \lambda_k; p) = (SSR_0 - SSR_k) / [kp(T - (k + 1)p)^{-1}SSR_k]. \]

It follows from (6.23) that the limiting behavior of the numerator of \( F_T(\lambda_1, \cdots, \lambda_k; p) \) is given by

\[ SSR_0 - SSR_k = F^*_T = \sum_{i=1}^k F_{T,i} \]

\[ \Rightarrow \sum_{i=1}^k \{\lambda_i \lambda_{i+1}(\lambda_{i+1} - \lambda_i)\}^{-1}||[\lambda_{i+1} \hat{D}_i - \lambda_i \hat{D}_{i+1}] + [\lambda_{i+1} D_i - \lambda_i D_{i+1}] \tilde{\beta}_0||^2 \]

(6.24)

Now, consider the denominator of \( F_T(\lambda_1, \cdots, \lambda_k; p) \). Using (6.4) and (6.6), it can be shown that

\[ SSR_k = \sum_{i=1}^k D^U(i, i) = \sum_{i=1}^k ||(I - P_{W_i}) \tilde{U}_i||^2 \]

\[ = \sum_{i=1}^k \hat{U}'_i (I - P_{W_i}) \hat{U}_i \]

\[ = \sum_{i=1}^k \hat{U}'_i \hat{U}_i - \sum_{i=1}^k \hat{U}'_i P_{W_i} \hat{U}_i \]

Thus, it follows that

\[ (T - (k + 1)p)^{-1}SSR_k = (T - (k + 1)p)^{-1} \sum_{i=1}^k \hat{U}'_i \hat{U}_i - (T - (k + 1)p)^{-1} \sum_{i=1}^k \hat{U}'_i P_{W_i} \hat{U}_i \]

(6.25)

In the following, we show that the first term in (6.25) dominates and converges to \( \sigma^2 + 2\gamma' \bar{\beta}_0 + \bar{\beta}_0' \Sigma \bar{\beta}_0 \) in probability and the second term in (6.25) is \( o_p(1) \).
First consider the second term. Since

$$
\tilde{U}_i P_{\tilde{W}} \tilde{U}_i = (S_i - S_{i-1})' (H_i - H_{i-1})^{-1} (S_i - S_{i-1})
$$

it follows from (6.19) that

$$
\tilde{U}_i P_{\tilde{W}} \tilde{U}_i \Rightarrow (\lambda_{i+1} - \lambda_i)^{-1} [\lambda C(\hat{D}^*(\lambda_{i+1}) - \hat{D}^*(\lambda_i)) + \lambda C(D^*(\lambda_{i+1}) - D^*(\lambda_i)) - \lambda_{i+1}D^*(1) + \lambda_i D^*(1)] [\lambda C(\hat{D}^*(\lambda_{i+1}) - \hat{D}^*(\lambda_i)) + \lambda_{i+1}D^*(1) + \lambda_i D^*(1)] \\
= (\lambda_i - \lambda_{i-1})^{-1} [\hat{D}_i - \hat{D}_{i-1} + (D_i - D_{i-1} - \lambda_i D_1 + \lambda_{i-1} D_1) \beta_0'] \\
\times [\hat{D}_i - \hat{D}_{i-1} + (D_i - D_{i-1} - \lambda_i D_1 + \lambda_{i-1} D_1) \beta_0] (6.26)
$$

and hence that \((T - (k + 1)p)^{-1} \sum_{i=1}^{k} \tilde{U}_i P_{\tilde{W}} \tilde{U}_i = o_p(1)\).

Now consider the first term of (6.25). From (3.15), it follows that under the null hypothesis of no breaks,

$$
\tilde{U}_i \tilde{U}_i = \sum_{t=\lfloor T \lambda_i \rfloor + 1}^{\lceil T \lambda_i \rceil} \tilde{u}_t^2
$$

$$
= \sum_i \left[(u_i + v_i' \beta_0) - z_i' (Z'Z)^{-1} Z' V \beta_0 \right]^2
$$

$$
= \sum_i \left[(u_i + v_i' \beta_0)^2 + (z_i' (Z'Z)^{-1} Z' V \beta_0)^2 - 2(u_i + v_i' \beta_0) z_i' (Z'Z)^{-1} Z' V \beta_0 \right]
$$

where we denote \( \sum_i = \sum_{t=\lfloor T \lambda_i \rfloor + 1}^{\lceil T \lambda_i \rceil} \) for notational convenience.

Since

$$
\sum_i (z_i' (Z'Z)^{-1} Z' V \beta_0)^2 = \sum \beta_0' V' Z (Z'Z)^{-1} z_i z_i' (Z'Z)^{-1} Z' V \beta_0
$$

$$
= \beta_0' V' Z (Z'Z)^{-1} \sum_i z_i z_i' (Z'Z)^{-1} Z' V \beta_0
$$

$$
= \beta_0' T^{-1/2} V' Z (T^{-1} Z' Z)^{-1} (T^{-1} \sum_i z_i z_i') (T^{-1} Z' Z)^{-1}
$$

$$
\times T^{-1/2} Z' V \beta_0
$$

$$
= O_p(1)
$$

115
\[ \sum_i (u_t + v_t' \beta_0) z_t'(Z'Z)^{-1} Z'V \beta_0 = \sum_i u_t z_t'(Z'Z)^{-1} Z'V \beta_0 + \sum_i v_t' \beta_0 z_t'(Z'Z)^{-1} Z'V \beta_0 \]
\[ = T^{-1/2} \sum_i u_t z_t'(T^{-1} Z'Z)^{-1} T^{-1/2} Z'V \beta_0 + T^{-1/2} \sum_i v_t' \beta_0 z_t'(T^{-1} Z'Z)^{-1} T^{-1/2} Z'V \beta_0 \]
\[ = O_p(1) \]

it follows that
\[ (T - (k + 1)p)^{-1} \sum_{i=1}^k \tilde{U}_i' \tilde{U}_i = (T - (k + 1)p)^{-1} \sum_{i=1}^k \sum_{[T \lambda_i]} (u_t + v_t' \beta_0)^2 + o_p(1) \]
\[ = (T - (k + 1)p)^{-1} \sum_{i=1}^k \sum_{[T \lambda_i]} (u_t^2 + 2u_t v_t' \beta_0)
+ \beta_0^0 v_t v_t' \beta_0 + o_p(1) \]
\[ \Rightarrow \sigma^2 + 2\gamma' \beta_0 + \beta_0^0 \Sigma \beta_0 \quad (6.27) \]

From (6.25)-(6.26) and (6.27), it follows that
\[ (T - (k + 1)p)^{-1} SSR_k \Rightarrow \sigma^2 + 2\gamma' \beta_0 + \beta_0^0 \Sigma \beta_0 \quad (6.28) \]

Combining (6.1), (6.24), and (6.28), we obtain
\[ F_T(\lambda_1, \cdots, \lambda_k; p) = \frac{SSR_0 - SSR_k}{[kp(T - (k + 1)p)^{-1} SSR_k]} \]
\[ \Rightarrow \frac{1}{kp} \sum_{i=1}^k \frac{||[\lambda_{i+1} \tilde{D}_i - \lambda_i \tilde{D}_{i+1}] + [\lambda_{i+1} D_i - \lambda_i D_{i+1}] \beta_0||^2}{\lambda_i \lambda_{i+1}(\lambda_{i+1} - \lambda_i) \sigma^2 + 2\gamma' \beta_0 + \beta_0^0 \Sigma \beta_0} \quad (6.29) \]

We now show that this limiting distribution has the alternative representation given in Theorem 7. First notice that the limit distribution on the right hand side
of (6.29), \( Q \) say, can be written as

\[
Q = \frac{1}{kp} \sum_{i=1}^{k} \left[ \lambda_i \lambda_i b(\lambda_i) - \lambda_i b(\lambda_i+1) \right] \frac{||\lambda_{i+1} b(\lambda_i) - \lambda_i b(\lambda_i+1)||}{\lambda_i \lambda_i + 1} \sum_{i=1}^{2} \left[ \lambda_i b(\lambda_i) - \lambda_i b(\lambda_{i+1}) \right] (6.30)
\]

where \( b(\lambda_i) = [\tilde{D}_i + D_i/\hat{\beta}_0] \).

Therefore the desired result will be established if it can be shown that

\[
b(\lambda_i) \stackrel{d}{=} \left[\sigma^2 + 2\gamma' \hat{\beta}_0 + \hat{\beta}_0' \Sigma \hat{\beta}_0 \right]^{1/2} \begin{bmatrix} W_i \\ 0_{(q-p)\times 1} \end{bmatrix} (6.31)
\]

where \( W_i \) is a \( p \times 1 \) vector of standard Brownian motion process and \( 0_{(q-p)\times 1} \) null vector and \( \stackrel{d}{=} \) denotes “distributed as”. We now show that (6.31) holds. Without loss of generality, we assume the rows of \((\Delta_0'Q_{2Z}^{-1}')(\Delta_0'Q_{ZZ}\Delta_0)^{-1}(\Delta_0'Q_{Z2}^{-1/2})\) are arranged such that \( \Lambda = diag(t_p, 0_{(q-p)\times 1}) \) in (6.16) where \( t_p \) is \( p \times 1 \) vector of ones. It therefore follows that

\[
\tilde{B}_i = \Lambda C \tilde{D}^*(\lambda_i) = \Lambda C \sum_{i=1}^{p+1} N_{1,i} B_i(\lambda_i) = \sum_{i=1}^{p+1} N_{1,i} \Lambda C B_i(\lambda_i)
\]

\[
\stackrel{d}{=} \sum_{i=1}^{p+1} N_{1,i} \Lambda B_i(\lambda_i)
\]

\[
= \begin{bmatrix} \sum_{i=1}^{p+1} N_{1,i} \Lambda B_{i,1:p}(\lambda_i) \\ 0_{(q-p)\times 1} \end{bmatrix}
\]

\[
= \begin{bmatrix} (N_1' \otimes I_p) B_{1:p}(\lambda_i) \\ 0_{(q-p)\times 1} \end{bmatrix} (6.32)
\]

where \( B_{i,1:p}(\cdot) \) is a \( p \times 1 \) vector containing the first \( p \) rows of \( B_i(\cdot) \) and \( B_{1:p}(\cdot) = [B_{1,1:p}(\cdot)', B_{2,1:p}(\cdot)', \ldots, B_{p+1,1:p}(\cdot)']' \).

It also follows using similar arguments that

\[
D_i \tilde{\beta}_0 = \Lambda C D_{mat}(\lambda_i) N_2 \tilde{\beta}_0
\]

\[
= (\tilde{\beta}_0' N_2' \otimes \Lambda C) B(\lambda_i)
\]
\[
\begin{bmatrix}
B_{1,1:p}(\lambda_i) \\
0_{(q-p)\times 1} \\
B_{2,1:p}(\lambda_i) \\
0_{(q-p)\times 1} \\
\vdots \\
B_{p+1,1:p}(\lambda_i) \\
0_{(q-p)\times 1}
\end{bmatrix}
= \begin{bmatrix}
(\bar{\beta}'_0 N'_2 \otimes I_q) B_{1:p}(\lambda_i) \\
0_{(q-p)\times 1}
\end{bmatrix}
\] (6.33)

It follows from (6.32) and (6.33) that

\[
b(\lambda_i) = \begin{bmatrix}
\{ (N'_1 + \bar{\beta}'_0 N'_2) \otimes I_p \} B_{1:p}(\lambda_i) \\
0_{(q-p)\times 1}
\end{bmatrix}
\] (6.34)

\[
b(\lambda_i) = \begin{bmatrix}
b_1(\lambda_i) \\
0_{(q-p)\times 1}
\end{bmatrix} \text{, say} \] (6.35)

Equation (6.34) proves (6.31) for the lower \( q-p \) elements. For the remaining elements, note that it follows from (6.34)-(6.35) that:

- \( b_1(0) = 0_{p\times 1} \);

- For any dates \( 0 \leq \lambda_1 < \lambda_2 < \cdots < \lambda_n \leq 1 \), the changes \( b_1(\lambda_2) - b_1(\lambda_1), b_1(\lambda_3) - b_1(\lambda_2), \cdots b_1(\lambda_n) - b_1(\lambda_{n-1}) \) are independent multivariate Gaussian with

\[
b_1(\lambda_i) - b_1(\lambda_{i-1}) \sim N \left( 0_{p\times 1}, \left( \sigma^2 + 2\gamma' \bar{\beta}_0 + \bar{\beta}'_0 \Sigma \bar{\beta}_0 \right) (\lambda_i - \lambda_{i-1})I_p \right) \]

- For any given realization, \( b_1(\lambda) \) is continuous in \( \lambda \) with probability one.

It follows from these three properties that

\[
b_1(\lambda_i) \triangleq \left[ \sigma^2 + 2\gamma' \bar{\beta}_0 + \bar{\beta}'_0 \Sigma \bar{\beta}_0 \right]^{1/2} W_i
\]
which completes the proof.

We note that the limiting distribution in Theorem 7 is exactly the same as the one in Bai and Perron’s (1998) analogous result for the sup-F test based on OLS estimators when the regressors are exogenous. Percentiles for this distribution can be found in Bai and Perron (1998) [Table I] for $\epsilon = 0.05$ and in Bai and Perron (2001) for other values of $\epsilon$.

The $Sup \ F_T(k; p)$ test statistic requires the specification of the number of breaks, $k$, under the alternative hypothesis. In many circumstances, a researcher is unlikely to know a priori the appropriate choice of $k$. To circumvent this problem, Bai and Perron (1998) propose so called “Double Maximum tests” that combine information from the sup $F_T(k; p)$ statistics for different values of $k$ running from one to some ceiling $K$. We consider here only the following example of Double Maximum test,

$$UDmaxF_T(K; p) = \max_{1 \leq k \leq K} \sup_{(\lambda_1, \cdots, \lambda_k) \in \Lambda} F_T(\lambda_1, \cdots, \lambda_k; p)$$

The limiting distribution of this statistic follows directly from Theorem 7.

**Corollary 1** Under the conditions of Theorem 7, it follows that

$$UDmaxF_T(K; p) \Rightarrow \max_{1 \leq k \leq K} \{Sup \ F_k(p)\}$$

Critical values for the limiting distribution in Corollary 1 are presented in Bai and Perron (1998) [Table 1] for $\epsilon = 0.05$ and Bai and Perron (2001) for other values of $\epsilon$.  

119
6.2 Sequential Test

Now we turn our attention to the sequential test of the break dates. It leads to a procedure for determining the number of breaks, which is assumed unknown in the first place.

The natural candidate for a test of the null hypothesis of \(l\) breaks against the alternative that an additional break exists is based on the difference between the sum of squared residuals obtained with \(l\) breaks and that obtained with \(l+1\) breaks but it is difficult to get the limit distribution of the above mentioned natural test. So, we pursue a different strategy to test the null hypothesis, as is suggested by Bai and Perron(1998).

First, \(\text{Sup } F_T(k;p)\) and \(\text{UDmax } F_T(K;p)\) statistics are used to test the null hypothesis of no breaks. For the test of the null hypothesis of \(l\) breaks against the alternative of \(l + 1\) breaks, the break point estimates, denoted by \(\hat{T}_1, \ldots, \hat{T}_l\), are obtained for the \(l\) breaks model by a global minimization of the sum of the squared residuals. Second, for the \(l + 1\) regimes obtained from the estimated partition, \(\{\hat{T}_1, \ldots, \hat{T}_l\}\), we test each of the \(l + 1\) segments for the presence of an additional break. Third, we conclude for a rejection in favor of a model with \(l + 1\) breaks if the overall minimal value of the sum of squared residuals (over all segments where an additional break is included) is sufficiently smaller than the sum of squared residuals from the \(l\) break model.\(^1\) Therefore, the test is defined by

\[ F_T(l + 1|l) \]

\(^1\)The estimates \(\hat{T}_i\) need not be the global minimizers of the sum of squared residuals, one can use sequential one-at-a-time estimates which allows the construction of a sequential procedure to select the number of breaks(see Bai(1997b)).
\[
\frac{SSR_l(\hat{T}_1, \cdots, \hat{T}_l) - \min_{1 \leq i \leq l+1} \inf_{\tau \in \Lambda_{i,\eta}} SSR_{l+1}(\hat{T}_1, \cdots, \hat{T}_{i-1}, \tau, \hat{T}_i, \cdots, \hat{T}_l)}{SSR_l(\hat{T}_1, \cdots, \hat{T}_l)/(T - (l + 1)p)}
\]

where
\[
\Lambda_{i,\eta} = \{ \tau : \hat{T}_{i-1} + (\hat{T}_i - \hat{T}_{i-1})\eta \leq \tau \leq (\hat{T}_i - \hat{T}_{i-1})\eta \}
\]

**Theorem 8** Under Assumptions 1\sim 5, 11 and \(m = l\)

\[
\lim_{T \to \infty} P(F_T(l + 1 | l) \leq x) = G_{p,\eta}(x)^{l+1}
\]

where \(G_{p,\eta}(x)\) is the distribution function of
\[
\sup_{\eta \leq \mu < 1-\eta} \| W(\mu) - \mu W(1) \|^2 / \mu (1 - \mu)
\]

**Proof**

By defining \(S_T(T_i + 1, T_j)\) to be the minimized sum of squared residuals for the segment containing observations from \(T_i + 1\) to \(T_j\), we can write

\[
F_T(l + 1 | l) = \sup_{1 \leq i \leq l+1} \sup_{\tau \in \Lambda_{i,\eta}} \frac{\{S_T(\hat{T}_{i-1} + 1, \hat{T}_i) - S_T(\hat{T}_{i-1} + 1, \tau) - S_T(\tau + 1, \hat{T}_i)\}}{\hat{\sigma}_1^2}
\]

\[\text{(6.36)}\]

where \(\hat{\sigma}_1^2 = SSR_l(\hat{T}_1, \cdots, \hat{T}_l)/(T - (l + 1)p) \rightarrow_p \sigma^2 + 2\sigma \rho' \Sigma^{1/2} \bar{\beta}_0 + \beta' \Sigma \bar{\beta}_0\).

First, under Assumptions 9 \sim 10 we can derive

\[
\sup_{\tau \in \Lambda_{i,\eta}} \frac{\{S_T(T^0_{i-1} + 1, T^0_i) - S_T(T^0_{i-1} + 1, \tau) - S_T(\tau + 1, T^0_i)\}}{[\sigma^2 + 2\sigma \rho' \Sigma^{1/2} \bar{\beta}_0 + \beta' \Sigma \bar{\beta}_0]}
\]

\[
\Rightarrow \sup_{\eta \leq \mu < 1-\eta} \frac{\| W(\mu) - \mu W(1) \|^2}{\mu (1 - \mu)}
\]

\[\text{(6.37)}\]

where \(\Lambda_{i,\eta}^0 = \{ \tau : T^0_{i-1} + (T^0_i - T^0_{i-1})\eta \leq \tau \leq T^0_i - (T^0_i - T^0_{i-1})\eta \}, \tau = T^0_{i-1} + (T^0_i - T^0_{i-1})\mu, S_T(T^0_{i-1} + 1, T^0_i)\) equals sum of squared residuals from the restricted model using observations from \(T^0_{i-1} + 1\) to \(T^0_i\), and similarly for \(S_T(T^0_{i-1} + 1, \tau)\) and
and where bracket which is defined to be

\[ \tau, i \]

\[ \frac{1}{\tau, i} \] respectively. Using these definitions, (6.38) can be further written as

\[
G_{T,i} \equiv S_T(T_{i-1}^0 + 1, T_i^0) - S_T(T_{i-1}^0 + 1, \tau) - S_T(\tau + 1, T_i^0)
\]

(6.38)

We define \( P_{W_{i-1},i} \) and \( \tilde{U}_{i-1,i} \) to be the projection matrix of \( W_{i-1,i} \) and the corresponding error vector from the observations \( T_{i-1}^0 + 1 \) to \( T_i^0 \), respectively. Similarly, \( P_{W_{i-1},i} \), \( \tilde{U}_{i-1,i} \) and \( P_{W_{i-1},i} \), \( \tilde{U}_{i-1,i} \) are defined to be the projection matrices and the corresponding error vectors from observations \( T_{i-1}^0 + 1 \) to \( \tau \) and \( \tau + 1 \) to \( T_i^0 \), respectively. Using these definitions, (6.38) can be further written as

\[
G_{T,i} = \|(I - P_{W_{i-1},i})\tilde{U}_{i-1,i}\|^2 - \|(I - P_{W_{i-1},i})\tilde{U}_{i-1,\tau}\|^2 - \|(I - P_{W_{i-1},i})\tilde{U}_{\tau,i}\|^2
\]

\[ = \tilde{U}_{i-1,i}'(I - P_{W_{i-1},i})\tilde{U}_{i-1,i} - \tilde{U}_{i-1,\tau}'(I - P_{W_{i-1},i})\tilde{U}_{i-1,\tau} - \tilde{U}_{\tau,i}'(I - P_{W_{i-1},i})\tilde{U}_{\tau,i}
\]

\[ = \tilde{U}_{i-1,i}'\tilde{U}_{i-1,i} - \tilde{U}_{i-1,i}'P_{W_{i-1},i}\tilde{U}_{i-1,i} - \tilde{U}_{i-1,\tau}'\tilde{U}_{i-1,\tau} + \tilde{U}_{i-1,\tau}'P_{W_{i-1},i}\tilde{U}_{i-1,\tau}
\]

\[ - \tilde{U}_{\tau,i}'\tilde{U}_{\tau,i} + \tilde{U}_{\tau,i}'P_{W_{i-1},i}\tilde{U}_{\tau,i}
\]

\[ = -S_{\tau,i}'H_{\tau,i}^{-1}S_{\tau,i} + S_{i-1,\tau}'H_{i-1,\tau}^{-1}S_{i-1,\tau} + (S_{\tau,i} - S_{i-1,\tau})'(H_{\tau,i} - H_{i-1,\tau})^{-1}
\]

\[ \times (S_{\tau,i} - S_{i-1,\tau})
\]

(6.39)

where \( S_{i-1,\tau} = W_{i-1,i}'\tilde{U}_{i-1,\tau}, S_{\tau,i} = W_{\tau,i}'\tilde{U}_{\tau,i}, H_{i-1,\tau} = W_{i-1,i}'W_{i-1,\tau}, \) and \( H_{\tau,i} = W_{\tau,i}'W_{\tau,i} \).

The limiting behavior of \( G_{T,i} \) is deduced from the limiting behavior of \( S_{\tau,i}, S_{i-1,\tau}, H_{i-1,\tau} \) and \( H_{\tau,i} \). In the following we derive the limiting distribution of each term in turn.
To deduce the limiting behavior of $S_{r,i}$ we note that:

$$(\Delta T_i^0)^{-1/2} S_{r,i} = (\Delta T_i^0)^{-1/2} W_{i-1,i}^{\prime} \tilde{U}_{i-1,i}$$

$$(\Delta T_i^0)^{-1/2} \sum_{t=T_{i-1}^0+1}^{T_i^0} w_t \tilde{u}_t$$

$$(\Delta T_i^0)^{-1/2} \sum_{t=T_{i-1}^0+1}^{T_i^0} \hat{x}_t \tilde{u}_t$$  \hspace{1cm} (6.40)

where $\Delta T_i^0 = T_i^0 - T_{i-1}^0$. We now expand (6.40) further to investigate the limit behavior of $S_{r,i}$. To this end, it is convenient to define $\sum_{i_0}$ to denote the summation over observations $t = T_{i-1}^0 + 1, \ldots, T_i^0$. That is, $\sum_{i_0} \equiv \sum_{t=T_{i-1}^0+1}^{T_i^0}$. By using the suggested notation (6.40) can be written as

$$$(\Delta T_i^0)^{-1/2} \sum_{i_0} \hat{x}_t \tilde{u}_t = (\Delta T_i^0)^{-1/2} [\Delta_{0}^{t} \sum_{i_0} z_t u_t + V^t Z (Z^t Z)^{-1} \sum_{i_0} z_t u_t$$

$$+ \Delta_0^{t} \sum_{i_0} z_t v_t^{t} \cdot \tilde{\beta}_0 + V^t Z (Z^t Z)^{-1} \sum_{i_0} z_t v_t^{t} \cdot \tilde{\beta}_0$$

$$- \Delta_0^{t} \sum_{i_0} z_t z_t^{t} (Z^t Z)^{-1} Z^t V \cdot \tilde{\beta}_0$$

$$- V^t Z (Z^t Z)^{-1} \sum_{i_0} z_t z_t^{t}$$

$$\times (Z^t Z)^{-1} Z^t V \cdot \tilde{\beta}_0]$$

$$= \Delta_0^{t} \cdot (\Delta T_i^0)^{-1/2} \sum_{i_0} z_t u_t + V^t Z (Z^t Z)^{-1} \cdot (\Delta T_i^0)^{-1/2} \sum_{i_0} z_t u_t$$

$$+ \Delta_0^{t} \cdot (\Delta T_i^0)^{-1/2} \sum_{i_0} z_t v_t^{t} \cdot \tilde{\beta}_0 + V^t Z (Z^t Z)^{-1}$$

$$\times (\Delta T_i^0)^{-1/2} \sum_{i_0} z_t v_t^{t} \cdot \tilde{\beta}_0 - \Delta_0^{t} \sum_{i_0} z_t z_t^{t} (Z^t Z)^{-1}$$

$$\times (\Delta T_i^0)^{-1/2} Z^t V \cdot \tilde{\beta}_0$$

$$\Rightarrow (\Delta T_i^0)^{-1/2} \sum_{i_0} z_t z_t^{t} (Z^t Z)^{-1} Z^t V \cdot \tilde{\beta}_0$$

$$= \Delta_0^{t} \sum_{i_0} z_t z_t^{t} (Z^t Z)^{-1} (T(\lambda^0_{i+1} - \lambda^0_i))^{-1/2} Z^t V \cdot \tilde{\beta}_0$$

$$= (\lambda^0_{i+1} - \lambda^0_i)^{-1/2} \Delta_0^{t} \sum_{i_0} z_t z_t^{t} (Z^t Z)^{-1}$$

$$\times T^{-1/2} Z^t V \cdot \tilde{\beta}_0$$

123
\[ V'Z(Z'Z)^{-1} (\Delta T_i^0)^{-1/2} \sum_{i_0} z_{t} u_{t} = o_p(1) \]
\[ V'Z(Z'Z)^{-1} (\Delta T_i^0)^{-1/2} \sum_{i_0} z_{t} v_{t} \beta_0 = o_p(1) \]

and
\[
(\Delta T_i^0)^{-1/2} V'Z(Z'Z)^{-1} \sum_{i_0} z_t z'_t (Z'Z)^{-1} Z'V \beta_0 = (T(\lambda_{i+1}^0 - \lambda_i^0))^{-1/2} V'Z(Z'Z)^{-1} \\
\times \sum_{i_0} z_t z'_t (Z'Z)^{-1} Z'V \beta_0 \\
= (\lambda_{i+1}^0 - \lambda_i^0)^{-1/2} T^{-1/2} V'Z(Z'Z)^{-1} \\
\times \sum_{i_0} z_t z'_t (Z'Z)^{-1} Z'V \beta_0 \\
= o_p(1)
\]

the previous equation (6.41) can be written as
\[
(\Delta T_i^0)^{-1/2} \sum_{i_0} \bar{x}_t \bar{u}_t = \Delta_0' (\Delta T_i^0)^{-1/2} \sum_{i_0} z_t u_t + \Delta_0' (\Delta T_i^0)^{-1/2} \sum_{i_0} z_t v'_t \beta_0 \\
- (\lambda_{i+1}^0 - \lambda_i^0)^{-1/2} \Delta_0' \sum_{i_0} z_t z'_t (Z'Z)^{-1} T^{-1/2} Z'V \beta_0 + o_p(1) \\
\Rightarrow \Delta_0' Q_{ZZ}^{1/2} D^*(1) + \Delta_0' Q_{ZZ}^{1/2} D^*(1) \beta_0 - (\lambda_{i+1}^0 - \lambda_i^0)^{-1/2} \\
\times \Delta_0' (\lambda_{i+1}^0 - \lambda_i^0) Q_{ZZ}^{1/2} D^*(1) \beta_0 \\
= \Delta_0' Q_{ZZ}^{1/2} D^*(1) + \Delta_0' Q_{ZZ}^{1/2} D^*(1) \beta_0 - (\lambda_{i+1}^0 - \lambda_i^0)^{1/2} \Delta_0' Q_{ZZ}^{1/2} \\
\times D^*(1) \beta_0 \\
= \Delta_0' Q_{ZZ}^{1/2} D^*(1) + \Delta_0' Q_{ZZ}^{1/2} D^*(1) \beta_0 - \Delta_0' Q_{ZZ}^{1/2} D^*(\lambda_{i+1}^0 - \lambda_i^0) \beta_0 \\
(6.42)
\]

where the third line expression uses the fact that
\[
\sum_{i_0} z_t z'_t (Z'Z)^{-1} \cdot T^{-1/2} Z'V \cdot \beta_0 = (\sum_{i_0} z_t z'_t / T)(Z'Z/T)^{-1} T^{-1/2} Z'V \cdot \beta_0 \\
= (\lambda_{i+1}^0 - \lambda_i^0) Q_{ZZ} Q_{ZZ}^{1/2} Q_{ZZ}^{1/2} D^*(1) \beta_0 \\
= (\lambda_{i+1}^0 - \lambda_i^0) Q_{ZZ}^{1/2} D^*(1) \beta_0
\]
To derive the limiting behavior of \((\Delta T^0_i)^{-1/2}S_{i-1,\tau}\), we can follow the same steps as those in \((\Delta T^0_i)^{-1/2}S_{r,j}\).

\[
(\Delta T^0_i)^{-1/2}S_{i-1,\tau} = (\Delta T^0_i)^{-1/2}W'_{i-1,\tau}U_{i-1,\tau}
\]
\[
= (\Delta T^0_i)^{-1/2} \sum_{t=T^0_{i-1}+\Delta_0^0\mu} \sum_{t=T^0_{i-1}+1} w_t \tilde{u}_t
\]
\[
= (\Delta T^0_i)^{-1/2} \sum_{t=T^0_{i-1}+1} \tilde{x}_t \tilde{u}_t
\]
(6.43)

By denoting \(\sum_{t=T^0_{i-1}+\Delta_0^0\mu} \equiv \sum_1\), (6.43) can be simplified in notation to

\[
(\Delta T^0_i)^{-1/2} \sum_1 \tilde{x}_t \tilde{u}_t = (\Delta T^0_i)^{-1/2} \left| \Delta_0' \sum_1 z_t u_t + V'Z(Z'Z)^{-1} \sum_1 z_t u_t \right|
\]
\[
+ \Delta_0' \sum_1 z_t v'_t \bar{\beta}_0 + V'Z(Z'Z)^{-1} \sum_1 z_t v'_t \bar{\beta}_0
\]
\[
- \Delta_0' \sum_1 z_t z'_t (Z'Z)^{-1} Z'V \bar{\beta}_0 - V'Z(Z'Z)^{-1} \sum_1 z_t z'_t (Z'Z)^{-1}
\]
\[
\times Z'V \bar{\beta}_0
\]
\[
= \Delta_0' (\Delta T^0_i)^{-1/2} \sum_1 z_t u_t + V'Z(Z'Z)^{-1} (\Delta T^0_i)^{-1/2} \sum_1 z_t u_t
\]
\[
+ \Delta_0' (\Delta T^0_i)^{-1/2} \sum_1 z_t v'_t \bar{\beta}_0 + V'Z(Z'Z)^{-1} (\Delta T^0_i)^{-1/2}
\]
\[
\times \sum_1 z_t v'_t \bar{\beta}_0 - \Delta_0' \sum_1 z_t z'_t (Z'Z)^{-1} (\Delta T^0_i)^{-1/2} Z'V \bar{\beta}_0
\]
\[
- (\Delta T^0_i)^{-1/2} V'Z(Z'Z)^{-1} \sum_1 z_t z'_t (Z'Z)^{-1} Z'V \bar{\beta}_0
\]
(6.44)

To deduce the limiting behavior of \((\Delta T^0_i)^{-1/2} \sum_1 \tilde{x}_t \tilde{u}_t\), we investigate that of each term in (6.44) in turn.

\[
(\Delta T^0_i)^{-1/2} \sum_1 z_t u_t = \mu^{1/2}(\Delta T^0_i)^{-1/2} \sum_1 z_t u_t
\]
\[
\Rightarrow \mu^{1/2} Q_{zz}^{1/2} \tilde{D}^*(1)
\]
(6.45)

\[
(\Delta T^0_i)^{-1/2} \sum_1 z_t v'_t \bar{\beta}_0 = \mu^{1/2}(\Delta T^0_i)^{-1/2} \sum_1 z_t v'_t \bar{\beta}_0
\]
\[
\Rightarrow \mu^{1/2} Q_{zz}^{1/2} D^*(1) \bar{\beta}_0
\]
(6.46)
\[ \sum_1 z_t z_t' \cdot (Z'Z)^{-1}(\Delta T_t^0)^{-1/2}Z'V \bar{\beta}_0 = (\sum_1 z_t z_t' / T)(Z'Z / T)^{-1}(\Delta T_t^0)^{-1/2}Z'V \bar{\beta}_0 = (\lambda_{t+1}^0 - \lambda_t^0) \mu \sum_1 z_t z_t' \left( \frac{Z'Z}{T} \right)^{-1} \times [T(\lambda_{t+1}^0 - \lambda_t^0)]^{-1/2}Z'V \bar{\beta}_0 = (\lambda_{t+1}^0 - \lambda_t^0) \mu \left( \sum_1 z_t z_t' \frac{Z'Z}{\Delta T_t^n \mu} \left( \frac{Z'Z}{T} \right)^{-1} \times T^{-1/2}Z'V \bar{\beta}_0 \right. \\
\Rightarrow (\lambda_{t+1}^0 - \lambda_t^0) \mu Q_{ZZ} Q_{ZZ}^{1/2} D^*(1) \bar{\beta}_0 = (\lambda_{t+1}^0 - \lambda_t^0) \mu Q_{ZZ}^{1/2} D^*(1) \bar{\beta}_0 \] (6.47)

\[ (\Delta T_t^0)^{-1/2} V'Z(Z'Z)^{-1} \sum_1 z_t z_t' (Z'Z)^{-1} Z'V \bar{\beta}_0 = \frac{V'Z}{[T(\lambda_{t+1}^0 - \lambda_t^0)]^{1/2}} \left( \frac{Z'Z}{T} \right)^{-1} \times \sum_1 z_t z_t' \left( \frac{Z'Z}{T} \right)^{-1} \frac{Z'V}{T} \bar{\beta}_0 = \alpha_p(1) \] (6.48)

Combining the convergence results in (6.45)-(6.48), we have

\[ (\Delta_t^0)^{-1/2} S_{t-1,\tau} = \Delta_0' (\Delta T_t^0)^{-1/2} \sum_1 z_t u + \Delta_0' \cdot (\Delta T_t^0)^{-1/2} \sum_1 z_t v' \bar{\beta}_0 - \Delta_0 \sum_1 z_t z_t' (Z'Z)^{-1}(\Delta T_t^0)^{-1/2}Z'V \bar{\beta}_0 + o_p(1) \]

\[ \Rightarrow \Delta_0^{1/2} Q_{ZZ}^{1/2} D^*(1) + \Delta_0 \mu^{1/2} Q_{ZZ}^{1/2} D^*(1) \bar{\beta}_0 - \Delta_0 (\lambda_{t+1}^0 - \lambda_t^0)^{1/2} \times \mu Q_{ZZ}^{1/2} D^*(1) \bar{\beta}_0 \\
= \Delta_0 Q_{ZZ}^{1/2} D^*(\mu) + \Delta_0 Q_{ZZ}^{1/2} D^*(\mu) \bar{\beta}_0 - \mu \Delta_0 Q_{ZZ}^{1/2} D^*(\lambda_{t+1}^0 - \lambda_t^0) \bar{\beta}_0 \] (6.49)

Now, we'll consider $(\Delta T_t^0)^{-1} H_{t-1,\tau}$. Using the following convergence results based on the assumption 5 and the consistency of $\hat{\Delta}_T$

\[ (\Delta T_t^0 \mu)^{-1} \sum_1 z_t z_t' \Rightarrow Q_{ZZ} \]

\[ \Delta_T \Rightarrow \Delta_0 \] (6.50)
we have

\[
(\Delta T_0^{-1})^{-1} H_{i-1,\tau} = (\Delta T_0^{-1})^{-1} W'_{i-1,\tau} W_{i-1,\tau} \\
= (\Delta T_0^{-1})^{-1} \sum_i w_i w'_i = (\Delta T_0^{-1})^{-1} \sum_i \hat{x}_i \hat{x}'_i \\
= \mu(\Delta T_i^{-1} \mu)^{-1} \sum_i \hat{x}_i \hat{x}'_i \\
= \mu(\Delta T_i^{-1} \mu)^{-1} \sum_i \hat{\Delta}'_T z_i z'_i \hat{\Delta}_T \\
= \mu \hat{\Delta}'_T (\Delta T_i^{-1} \mu)^{-1} \sum z_i z'_i \hat{\Delta}_T \\
\Rightarrow \mu \hat{\Delta}'_0 Q ZZ \Delta_0 \\
\text{(6.51)}
\]

Finally, consider \((\Delta T_0^{-1})^{-1} H_{\tau,i}\). Using the convergence results (6.50) and \((\Delta T_0^{-1})^{-1} \sum z_i z'_i \Rightarrow Q ZZ\), we have

\[
(\Delta T_0^{-1})^{-1} H_{\tau,i} = (\Delta T_0^{-1})^{-1} W'_{\tau-1,i} W_{\tau-1,i} = (\Delta T_0^{-1})^{-1} \sum_{i_0} w_{i_0} w'_{i_0} \\
= (\Delta T_0^{-1})^{-1} \sum_{i_0} \hat{x}_{i_0} \hat{x}'_{i_0} \\
= (\Delta T_0^{-1})^{-1} \sum_{i_0} \hat{\Delta}'_T z_{i_0} z'_i \hat{\Delta}_T \\
= \hat{\Delta}'_T (\Delta T_0^{-1})^{-1} \sum z_{i_0} z'_i \hat{\Delta}_T \\
\Rightarrow \hat{\Delta}'_0 Q ZZ \Delta_0 \\
\text{(6.52)}
\]

Combining the results in (6.42), (6.49), (6.51) and (6.52) from above, we consider the behavior of the terms in (6.39). First consider \(S'_{\tau,i} H_{\tau,i}^{-1} S_{\tau,i}\). Using (6.42) and (6.52), we obtain

\[
S'_{\tau,i} H_{\tau,i}^{-1} S_{\tau,i} \Rightarrow (\Delta' Q_{ZZ}^{1/2} \tilde{D}^*(1) + \Delta' Q_{ZZ}^{1/2} D^*(1) \tilde{\beta}_0 - \Delta' Q_{ZZ}^{1/2} D^*(\lambda_{i+1}' - \lambda_i'))' \\
\times (\Delta' Q_{ZZ} \Delta_0)^{-1} (\Delta' Q_{ZZ}^{1/2} \tilde{D}^*(1) + \Delta' Q_{ZZ}^{1/2} D^*(1) \tilde{\beta}_0 - \Delta' Q_{ZZ}^{1/2} \\
\times D^*(\lambda_{i+1}' - \lambda_i')) \tilde{\beta}_0) \\
= (\Delta' Q_{ZZ} [\tilde{D}^*(1) + D^*(1) \tilde{\beta}_0 - D^*(\lambda_{i+1}' - \lambda_i') \tilde{\beta}_0])'
\]

127
\[ \times (\Delta_0'Q_{ZZZ}\Delta_0)^{-1}(\Delta_0'Q_{ZZZ}^{1/2}[\tilde{D}^*(1) + D^*(1)\tilde{\beta}_0 - D^*(\lambda^0_{i+1} - \lambda^0_i)\tilde{\beta}_0]) \]

\[ = (\Lambda C \tilde{D}^*(1) + \Lambda C[D^*(1) - D^*(\lambda^0_{i+1} - \lambda^0_i)]\tilde{\beta}_0)' \]

\[ \times (\Lambda C \tilde{D}^*(1) + \Lambda C[D^*(1) - D^*(\lambda^0_{i+1} - \lambda^0_i)]\tilde{\beta}_0) \quad (6.53) \]

Now, by denoting \( D_\mu = \Lambda C D^*(\mu) \), \( \tilde{D}_\mu = \Lambda C \tilde{D}^*(\mu) \), \( D_1 = \Lambda C D^*(1) \), \( \tilde{D}_1 = \Lambda C \tilde{D}^*(1) \) and \( D_i = \Lambda C D^*(\lambda^0_{i+1} - \lambda^0_i) \), we have

\[ S'_{\tau,i}H_{\tau,i}^{-1}S_{\tau,i} = (\tilde{D}_1 + [D_1 - D_i]\tilde{\beta}_0)'(\tilde{D}_1 + [D_1 - D_i]\tilde{\beta}_0) \quad (6.54) \]

Similarly, using the results in (6.49) and (6.51) we have

\[ S'_{i-1,\tau}H_{i-1,\tau}^{-1}S_{i-1,\tau} \Rightarrow (\Delta_0'Q_{ZZZ}^{1/2}[\tilde{D}^*(\mu) + \mu D^*(\lambda^0_{i+1} - \lambda^0_i)\tilde{\beta}_0])'(\mu \Delta_0') \]

\[ \times D^*(\lambda^0_{i+1} - \lambda^0_i)\tilde{\beta}_0 - \mu \Delta_0'Q_{ZZZ}^{1/2}[\tilde{D}^*(\mu) + \mu D^*(\lambda^0_{i+1} - \lambda^0_i)\tilde{\beta}_0] \]

\[ = \mu^{-1}(\Lambda C \tilde{D}^*(\mu) + \Lambda C[D^*(\mu) - \mu D^*(\lambda^0_{i+1} - \lambda^0_i)]\tilde{\beta}_0)' \]

\[ \times (\Lambda C \tilde{D}^*(\mu) + \Lambda C[D^*(\mu) - \mu D^*(\lambda^0_{i+1} - \lambda^0_i)]\tilde{\beta}_0) \quad (6.55) \]

Finally, using the results in (6.42), (6.49), (6.51) and (6.52), the last term in (6.39) can be seen as

\[ (S_{\tau,i} - S_{i-1,\tau})'((H_{\tau,i} - H_{i-1,\tau})^{-1}(S_{\tau,i} - S_{i-1,\tau}) \]

\[ \Rightarrow [\Delta_0'Q_{ZZZ}^{1/2}(\tilde{D}^*(1) - D^*(\mu))\Delta_0'Q_{ZZZ}^{1/2}(D^*(1) - D^*(\mu))\tilde{\beta}_0 - \Delta_0'Q_{ZZZ}^{1/2}(1 - \mu) \]

\[ \times D^*(\lambda^0_{i+1} - \lambda^0_i)\tilde{\beta}_0]'(\Delta_0'Q_{ZZZ}\Delta_0(1 - \mu))^{-1}[\Delta_0'Q_{ZZZ}^{1/2}(\tilde{D}^*(1) - \tilde{D}^*(\mu)) \]

\[ + \Delta_0'Q_{ZZZ}^{1/2}(D^*(1) - D^*(\mu))\tilde{\beta}_0 - \Delta_0'Q_{ZZZ}^{1/2}(1 - \mu)D^*(\lambda^0_{i+1} - \lambda^0_i)\tilde{\beta}_0] \]

128
Now we decompose the limit term of
\[ M_i = \alpha^i_{l+1} \lambda_i^0 (1 - \mu) D^*(1 - D^*) \bar{\beta}_0 \]

\[ \times ((1 - \mu) \lambda_i^0 (1 - \mu) D^*(1 - D^*) \bar{\beta}_0 ) \]

Thus, combining results (6.54), (6.55) and (6.56), it follows that

\[ G_{T,i} = -(\hat{D}_1 + [D_1 - D_{l}] \bar{\beta}_0)'(\hat{D}_1 + [D_1 - D_{l}] \bar{\beta}_0) + \mu^{-1}(\hat{D}_1 + [D_1 - D_{l}] \bar{\beta}_0)' \]

\[ \times (\hat{D}_1 + [D_1 - D_{l}] \bar{\beta}_0) + (1 - \mu)^{-1}[(\hat{D}_1 - \hat{D}_1) + (D_1 - D_{l}) \bar{\beta}_0 ] \]

\[ - (1 - \mu) D_{l} \bar{\beta}_0'[(\hat{D}_1 - \hat{D}_1) + (D_1 - D_{l}) \bar{\beta}_0 ] \]

\[ = -(\hat{D}_1 + [D_1 - D_{l}] \bar{\beta}_0)'(\hat{D}_1 + [D_1 - D_{l}] \bar{\beta}_0) + \mu^{-1}(\hat{D}_1 + [D_1 - D_{l}] \bar{\beta}_0)' \]

\[ \times (\hat{D}_1 + [D_1 - D_{l}] \bar{\beta}_0) + (1 - \mu)^{-1}[(\hat{D}_1 - \hat{D}_1) + (D_1 - D_{l}) \bar{\beta}_0 ] \]

\[ - (1 - \mu) D_{l} \bar{\beta}_0'[(\hat{D}_1 - \hat{D}_1) + (D_1 - D_{l}) \bar{\beta}_0 ] \]

Now we decompose the limit term of \( G_{T,i} \) into

\[ G_{T,i} = M_i^{(s)} + M_i^{(c)} \]

where \( M_i^{(s)} \) contains square terms in \( \hat{D}_1, D_1, D_{l} \), etc. and \( M_i^{(c)} \) contains cross product terms in \( \hat{D}_1, D_1, D_{l} \), etc. Rearranging the terms in (6.57), we obtain

\[ M_i^{(s)} = \mu^{-1}(\hat{D}_1 \bar{\beta}_0 + D_1 \bar{\beta}_0 + \bar{\beta}_0 D_1) \]

\[ - (\hat{D}_1 \bar{\beta}_0 + D_1 \bar{\beta}_0 + \bar{\beta}_0 D_1) \]
\[ + (1 - \mu)^{-1}(\tilde{D}'_1 \tilde{D}_1 + \tilde{D}'_1 D_i \tilde{\beta}_0 + \tilde{D}'_\mu D_\mu \tilde{\beta}_0 + \tilde{\beta}_0' D_i \tilde{D}_1 + \tilde{\beta}_0' D'_1 D_1 \tilde{\beta}_0 \\
\quad + (1 - \mu)^2 \tilde{\beta}_0' D'_1 D_1 \tilde{\beta}_0) \]

\[ = \frac{1}{\mu(1 - \mu)}[(1 - \mu)\tilde{D}'_\mu \tilde{D}_\mu + (1 - \mu)\tilde{D}'_\mu D_\mu \tilde{\beta}_0 + (1 - \mu)\tilde{\beta}_0' D'_\mu \tilde{D}_\mu + (1 - \mu)\tilde{\beta}_0' \times D'_\mu D_\mu \tilde{\beta}_0 + (1 - \mu)\mu^2 \tilde{\beta}_0' D'_i D_i \tilde{\beta}_0 - (\mu - \mu^2)\tilde{D}'_1 \tilde{D}_1 - (\mu - \mu^2)\tilde{D}'_1 D_1 \tilde{\beta}_0 \\
\quad - (\mu - \mu^2)\tilde{\beta}_0' D'_i \tilde{D}_1 - (\mu - \mu^2)\tilde{\beta}_0' D'_1 D_1 \tilde{\beta}_0 - (1 - \mu)\tilde{\beta}_0' D'_i \tilde{D}_1 + \mu \tilde{\beta}_0' D'_i \tilde{D}_1 + (1 - \mu)\tilde{\beta}_0' D'_i D_1 \tilde{\beta}_0 \\
\quad + (1 - \mu)^2 \tilde{\beta}_0' D'_i D_1 \tilde{\beta}_0 + \mu \tilde{\beta}_0' D'_i D_1 \tilde{\beta}_0 + \mu \tilde{\beta}_0' D'_\mu D_\mu \tilde{\beta}_0 + \mu \tilde{\beta}_0' D'_\mu D_\mu \tilde{\beta}_0 \\
\quad + (1 - \mu)^2 \tilde{\beta}_0' D'_i D_1 \tilde{\beta}_0 + \mu \tilde{\beta}_0' D'_i D_1 \tilde{\beta}_0 + \mu \tilde{\beta}_0' D'_\mu D_\mu \tilde{\beta}_0]
\]

\[ = \frac{1}{\mu(1 - \mu)}[(\tilde{D}'_\mu + D_\mu \tilde{\beta}_0)'(\tilde{D}_\mu + D_\mu \tilde{\beta}_0) + \mu^2(\tilde{D}_1 + D_1 \tilde{\beta}_0)'(\tilde{D}_1 + D_1 \tilde{\beta}_0)]
\]

(6.58)

and

\[ M^{(c)}_1 = \mu^{-1}(-\tilde{D}'_\mu D_\mu \tilde{\beta}_0 - \tilde{D}'_\mu D_\mu D_\mu \tilde{\beta}_0 - \mu \tilde{\beta}_0' D'_\mu \tilde{D}_\mu - \tilde{\beta}_0' \mu D'_\mu D_\mu \tilde{\beta}_0) \\
\quad - (-\tilde{D}'_i D_i \tilde{\beta}_0 - \tilde{D}'_i D_i \tilde{D}_1 - \tilde{D}'_i D_i \tilde{\beta}_0 - \tilde{\beta}_0' D'_i \tilde{D}_1 \tilde{\beta}_0) \\
\quad + (1 - \mu)^{-1}(-\tilde{D}'_i \tilde{D}_i - D_i \tilde{D}_1 - (1 - \mu)\tilde{D}'_i D_i \tilde{\beta}_0 - \tilde{D}'_i \tilde{D}_1) \\
\quad - \tilde{D}'_\mu D_1 \tilde{\beta}_0 + (1 - \mu)\tilde{D}'_\mu D_i \tilde{\beta}_0 - \tilde{\beta}_0' D'_\mu D_\mu - \tilde{\beta}_0' D'_\mu D_\mu \tilde{\beta}_0 \\
\quad - (1 - \mu)\tilde{\beta}_0' D'_i D_i \tilde{\beta}_0 - \tilde{\beta}_0' D'_i D_i \tilde{\beta}_0 + \tilde{\beta}_0' D'_i D_i \tilde{\beta}_0 - (1 - \mu)\tilde{\beta}_0' D'_i D_i \tilde{\beta}_0 \\
\quad - (1 - \mu)\tilde{\beta}_0' D'_i D_i \tilde{\beta}_0 + (1 - \mu)\tilde{\beta}_0' D'_i \tilde{D}_\mu - (1 - \mu)\tilde{\beta}_0' D'_i \tilde{D}_\mu \\
\quad + (1 - \mu)\tilde{\beta}_0' D'_i D_\mu \tilde{\beta}_0) \\
\quad = \frac{-1}{(1 - \mu)^2}[(\mu - \mu^2)\tilde{D}'_\mu D_\mu \tilde{\beta}_0 + (\mu - \mu^2)\tilde{D}'_\mu D_\mu \tilde{\beta}_0 + (\mu - \mu^2)\tilde{D}'_\mu \tilde{D}_\mu \\
\quad + (\mu - \mu^2)\tilde{D}'_i D_i \tilde{\beta}_0 - (\mu - \mu^2)\tilde{D}'_i D_i \tilde{\beta}_0 - (\mu - \mu^2)\tilde{D}'_i D_i \tilde{\beta}_0 \\
\quad - (\mu - \mu^2)\tilde{D}'_i D_i \tilde{D}_1 - (\mu - \mu^2)\tilde{D}'_i D_i \tilde{\beta}_0 + \mu \tilde{D}'_i \tilde{D}_i + \mu \tilde{D}'_i D_\mu \tilde{\beta}_0 \\
\quad + \mu \tilde{D}'_i D_\mu \tilde{\beta}_0] \\
\quad \text{130}
\]
\[ + (\mu - \mu^2)\tilde{D}_i^T D_i \tilde{\beta}_0 + \mu \tilde{D}_i^T \tilde{D}_1 + \mu \tilde{D}_i^T D_1 \tilde{\beta}_0 - (\mu - \mu^2)\tilde{D}_i^T D_i \tilde{\beta}_0 + \mu \tilde{\beta}_0^T D_i^T \tilde{D}_i + \mu \tilde{\beta}_0^T D_1 \tilde{\beta}_0 \]

\[ + \mu \tilde{\beta}_0^T D_i^T D_i \tilde{\beta}_0 + (\mu - \mu^2)\tilde{\beta}_0^T D_i^T D_i \tilde{\beta}_0 \]

\[ = \frac{-1}{(1 - \mu)\mu}[(\tilde{D}_\mu + D_\mu \tilde{\beta}_0) - \mu(\tilde{D}_1 + D_1 \tilde{\beta}_0)][(\tilde{D}_\mu + D_\mu \tilde{\beta}_0) - \mu(\tilde{D}_1 + D_1 \tilde{\beta}_0)] \]

From (6.58)-(6.59) it follows that

\[ G_{T,i} \Rightarrow \frac{1}{\mu(1 - \mu)}[(\tilde{D}_\mu + D_\mu \tilde{\beta}_0) - \mu(\tilde{D}_1 + D_1 \tilde{\beta}_0)] \]

\[ = \frac{1}{\mu(1 - \mu)}[(\tilde{D}_\mu - \mu \tilde{D}_1) + (D_\mu - \mu D_1) \tilde{\beta}_0][(\tilde{D}_\mu - \mu \tilde{D}_1) + (D_\mu - \mu D_1) \tilde{\beta}_0] \]

\[ = \frac{1}{\mu(1 - \mu)}\|\tilde{D}_\mu - \mu \tilde{D}_1 + (D_\mu - \mu D_1) \tilde{\beta}_0\| \]

\[ = \frac{1}{\mu(1 - \mu)}\|\tilde{D}_\mu + D_\mu \tilde{\beta}_0\| - \mu(\tilde{D}_1 + D_1 \tilde{\beta}_0) \]

where \( b(\mu) = \tilde{D}_\mu + D_\mu \tilde{\beta}_0 \) and \( b(1) = \tilde{D}_1 + D_1 \tilde{\beta}_0 \).

Thus, the desired result will be established if it can be shown that

\[ b(\mu) \overset{d}{=} \sigma^2 + 2\rho' \tilde{\beta}_0 + \tilde{\beta}_0' \Sigma \tilde{\beta}_0 \]

\[ = \left[ W(\mu) \right] \]

which we already have shown in the proof of Theorem 4.2

Under the null hypothesis, \( \hat{T}_i = T_i^0 + O_p(1) \) by Theorem 2 and we know that the sum of additional finite number of terms doesn’t make difference in limit distribution. Thus, (6.37) also holds with \( T_i^0 \) and \( T_i^0 \) replaced by \( \hat{T}_{i-1} \) and \( \hat{T}_i \), respectively.

---

2Refer to (6.31) and following derivations
In addition, weak limits in (6.37) for different \( \hat{\imath} \)'s are independent. Thus, the limit of (6.36) is the maximum of \( l + 1 \) independent random variables in the form of (6.37). This completes the proof of the proposition.

The critical value of \( F_T(l + 1|l) \) for different values of \( l \) can be obtained from the distribution function \( G_{p,\eta} \). Bai and Perron(1998) provided a full set of critical values in their Table II calculated with \( \eta = .05^3 \) and also showed that the test statistic \( F_T(l + 1|l) \) is consistent as sample size increases.

\footnote{The critical values for other values of \( \eta \) can be found at Bai and Perron(2001).}
Chapter 7

Simulation and Application

In previous chapters we proved the consistency of break fraction estimator and introduced several sup F-type test statistics to detect the structural breaks. In this chapter we conduct Monte Carlo simulation experiment and show the validity of the analytical results of previous chapters. We also discuss an application of the procedure to the Consumption-based Capital Asset Pricing Model (CCAPM) elaborated at Campbell, Lo and Mackinlay (1997).

7.1 Finite Sample Behavior

In this section, we lay out our Monte Carlo experiment design and evaluate the finite sample behavior of the various statistics discussed in the previous chapters via simulation. The simulation design involves models with zero, one or two breaks. The tables at the Appendix include our simulation results on the finite sample behavior of break fraction estimator, relative rejection frequencies of various F-statistics and empirical distribution of the estimated number of breaks. We begin
by discussing the one break and two break models and conclude this section by considering the behavior of the test statistics in the no break model.

7.1.1 One Break Model

The basic data generating process with one break can be described as follows.

\[ y_t = \beta_1^0 x_t + u_t = \beta_1^0 (z_t^\prime \Delta) + w_t \quad \text{for } t = 1, \cdots, [T/2]. \quad (7.1) \]
\[ y_t = \beta_2^0 x_t + u_t = \beta_2^0 (z_t^\prime \Delta) + w_t \quad \text{for } t = [T/2] + 1, \cdots, T. \quad (7.2) \]

The reduced form equation for the scalar variable \( x_t \) is:

\[ x_t = z_t^\prime \Delta + v_t \quad \text{for } t = 1, \cdots, T. \quad (7.3) \]

where \( w_t = \beta_i^0 v_t + u_t, \) for \( i = 1, 2 \) and \( \Delta \) is a \( q \)-dimensional vector consisting of identical elements, say \( \delta. \) The exact value of \( \delta \) comes from a given theoretical \( R^2 \) (Hahn and Inoue(2002)).

\[ R^2 = \frac{\Delta^\prime E(z_t z_t^\prime) \Delta}{1 + \Delta^\prime E(z_t z_t^\prime) \Delta} = \frac{q\delta^2}{1 + q\delta^2} \]

or

\[ \delta = \sqrt{R^2/(q - qR^2)} \quad (7.4) \]

Here, \( R^2 \) is the coefficient of determination between \( x_t \) and \( z_t, \) and \( q \) is the dimension of instrumental variables, \( z_t. \) In the following simulation experiments we consider overidentified regression models. that is, the dimension of the instrument variables is greater than that of scale regressor, \( x_t. \)

The errors are generated from a bivariate normal distribution

\[ (u_t, v_t)^\prime \sim i.i.d \quad N(0, \Omega) \]

134
where

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

and $\rho$ is the correlation coefficient between $u_t$ and $v_t$. Instrumental variables, $z_t$, are generated from a normal distribution.

$$z_t \sim i.i.d \ N(0, I_q)$$

The specific parameter values of simulation parameters are as follows: (i) $T = 60, 120, 240, 480$; (ii) $q = 2, 4, 8$; (iii) $(\beta_0^1, \beta_0^2) = (0.1, -0.1), (1, -1), (5, -5)$; (iv) $\rho = 0.5$; (v) $\delta$ is chosen to yield the population $R^2 = 0.5$ for the regression in (7.3) and (7.5). For each configuration, 1000 simulations are performed.

The results are presented in Tables 1-8. We first consider the behavior of the break fraction estimator calculated under the assumption that there is only one break. Tables 1-2 report the proportion of the simulations in which $|\hat{\lambda}_1 - \lambda_1^0| \leq c$ for $c = 0.01, 0.02, 0.03, 0.05, 0.1$. In Table 1, the change in the regression coefficient parameters is only $\beta_1^0 - \beta_2^0 = 0.2$, and this small change appears to make the detection of the true break point difficult. However, the proportions clearly increases with $T$ and exhibit the behavior in line with the consistency result in Theorem 1. For the other two parameter settings, the change in the regression parameters is much larger, and the estimated break fractions are within 0.01 of the true break fraction in all replications.

Tables 3-5 tabulate the relative rejection frequencies of $Sup F_T(k; 1)$ (for $k = 1, 2$), $UDmax F_T(5; 1)$ and $F_T(l + 1|l)$ (for $l = 1, 2$) statistics. The nominal size
of test statistics is set to 0.05. Notice that the alternative hypothesis is true for the \( \text{Sup } F_T(k; 1) \) statistic and so the corresponding relative rejection frequencies are empirical power for this statistic. Whereas, for \( l = 1 \), the null hypothesis is correct for \( F_T(l + 1|l) \) and so the relative rejection frequencies are the empirical size, and for \( l = 2 \), the null assumes more breaks than there actually are. The simulation results in Tables 3-5 suggest that the statistics \( \text{Sup } F_T(k; 1) \) and \( UDmax F_T(5; 1) \) share similar power properties and have reasonable power even in the case where \((\beta_0^1, \beta_0^2) = (0.1, -0.1)\). The \( F_T(2|1) \) is close to its nominal size; \( F_T(3|2) \) tends to reject less frequently than the nominal size 0.05.

Tables 6-8 report the results from using the sequential strategy based on these statistics that is described in chapter 6 with a maximum number of breaks set equal to five. The results show that the strategy works well in most cases except only one case where \((\beta_0^1, \beta_0^2) = (0.1, -0.1), q = 2 \) and \( T = 60 \). In this case we can observe significant mass at zero. In all but two other cases, the estimated number of breaks has no mass at zero. Also, the choice of statistic on the first step appears to have little impact on the empirical distribution of estimated number of break points \( \hat{k}_T \).

### 7.1.2 Two Break Model

The data generating process with two breaks is almost the same. The only difference is that there are now two breaks in the main regression model:

\[
\begin{align*}
  y_t &= \beta_1^0 x_t + u_t = \beta_1^0 (z'_t \Delta) + w_t \quad \text{for } t = 1, \cdots, [T/3], \\
  y_t &= \beta_2^0 x_t + u_t = \beta_2^0 (z'_t \Delta) + w_t \quad \text{for } t = [T/3] + 1, \cdots, [2T/3].
\end{align*}
\]
\[ y_t = \beta_0^3 x_t + u_t = \beta_0^3 (z_t' \Delta) + w_t \quad \text{for } t = \lfloor 2T/3 \rfloor + 1, \ldots, T. \]

and the reduced form equation for the scalar variable \( x_t \) is:

\[ x_t = z_t' \Delta + v_t \quad \text{for } t = 1, \ldots, T. \]  

(7.5)

Three choices of \( \beta^0 \) are considered: \( (\beta_0^0_1, \beta_0^0_2, \beta_0^0_3) = (-0.1, 0.1, -0.1), (-1, 1, -1), (-5, 5, -5) \). All other aspects of the design are the same as the one break model.

Tables 9-16 report the simulation results for the two break model. We start with the performance of the estimated break fractions which is tabulated at Tables 9-10. The finite sample behavior of the break fraction estimator reveals the same pattern as in the one break model: The proportions increase with \( T \) and the empirical distribution of the break fraction estimator does appear to be collapsing on the true fraction as \( T \) increases. Tables 11-13 report the relative rejection frequencies of \( \text{Sup } F_T(k; 1), F_T(l + 1|l) \), and \( UDmaxF_T(5; 1) \) statistics where the nominal size is set at 0.05. As in the one break model, the statistics are applied with \( k, l = 1, 2 \). Notice that the alternative hypothesis is true for the \( \text{Sup } F_T(k, 1), UDmaxF_T(5; 1) \) and \( F_T(2|1) \) statistics and so these relative rejection frequencies are empirical powers of these statistics. Whereas, the null hypothesis is true for the \( F_T(3|2) \) statistic and so the relative rejection frequencies are the empirical size. Table 11 reports the results for the case in which there is the smallest change between regimes, that is \( (\beta_0^0_1, \beta_0^0_2, \beta_0^0_3) = (-0.1, 0.1, -0.1) \). It can be seen that the \( UDmaxF_T(5, 1) \) statistic has markedly higher power than \( \text{Sup } F_T(1; 1) \) in this case, although both can have low power in small samples, a property shared with \( F_T(2|1) \). However, in the other two parameter configurations, all these statistics reject 100% of the time. The \( F_T(3|2) \) statistic is close to its nominal size in most cases.
Tables 14-16 report the results from using the sequential strategy for estimating the number of breaks. As in the one break model, the results indicate that the strategy works well in the two parameter configurations with relatively more change, that is $\beta_1^0, \beta_2^0, \beta_3^0 = (-1, 1, -1), (-5, 5, -5)$ but less well in the case with the smallest change, $(\beta_1^0, \beta_2^0, \beta_3^0) = (-0.1, 0.1, -0.1)$. In the latter case, there is a non-negligible tendency to underfit at sample sizes $T = 60, 120$. This problem stems from the lower power properties of the statistics $\text{Sup } F_T (1; 1)$ and $\text{UDmax } (5; 1)$ that are used on the first step of the sequential strategy. Recall that $\text{UDmax } (5; 1)$ has better power properties and this translates to a reduced tendency to underfit although there is still a non-negligible mass for $\hat{k}_T$ at zero in the cases where $(q, T) = (2, 60), (2, 120), (4, 60)$.

### 7.1.3 No Break Model

The previous two designs involve cases where there is a change in the regression parameters of the structural equation. It is also of interest to explore how the test statistics perform in the case where there is no break and so the model is structurally constant. To this end, data are generated from (7.1)-(7.2) with $\beta_1^0 = \beta_2^0 = 1$. All other aspects of the design are the same as the one break model.

Tables 17-18 report the simulation results for the no break model. Table 17 contains the empirical rejection frequencies for $\text{Sup } F_T (k; 1)(k = 1, 2), F_T (l + 1|l)(l = 1, 2)$ and $\text{UDmax } F_T (5; 1)$ statistics. Note that within this design, the null hypothesis is correct for the $\text{Sup } F_T (1; 1), \text{Sup } F_T (2; 1)$ and $\text{UDmax } F_T (5, 1)$ statistics, and so the rejection frequency equals the empirical size. For $F_T (2|1)$ and $F_T (3|2)$ statistics, the null hypothesis involves more breaks than are present
in the data. From Table 17, it can be seen that $Sup F_T(1; 1), Sup F_T(2; 1)$ and $UDmax F_T(5, 1)$ exhibit empirical size close to the nominal level of 0.05; both $F_T(2|1)$ and $F_T(3|2)$ reject less frequently than the size. Table 18 presents the empirical distribution of $\hat{k}_T$ based on the sequential strategies using $Sup F_T(1; 1)$ and $UDmax F_T(5; 1)$. Both strategies indicate that no breaks are present in nearly every case.

7.2 Empirical Application

We consider an application of the procedure involving the Consumption-based Capital Asset Pricing Model (CCAPM). Campbell, Lo and Mackinlay (1997) presented instrumental variables regression results for the asset returns and consumption growth. In this section, we briefly review their results and apply our estimation method for the structural breaks.\footnote{For further details on CCAPM, refer to Campbell, Lo and MacKinlay (1997).}

In this model the investigator who can trade freely in asset $i$ maximizes the expectation of a time-separable utility function

$$\max_{E_t} \left[ \sum_{j=0}^{\infty} \delta^j U(C_{t+j}) \right]$$

where $\delta$ is the time discount factor, $C_{t+j}$ is the investor’s consumption in period $t+j$, and $U(C_{t+j})$ is the period utility of consumption at $t+j$. The subscript $t$ in the Expectation operator represents conditional expectation based on the information set at time $t$.

Euler equations describing the investor’s optimal consumption and portfolio
plan is
\[ U'(C_t) = \delta E_t[(1 + R_{i,t+1})U''(C_{t+1})] \] (7.6)

where \( R_{i,t+1} \) is time \( T+1 \) asset \( i \) return. The left-hand side of (7.6) is the marginal utility cost of consuming one real dollar at time \( t \); the right-hand side is the expected marginal utility benefit from investing the dollar in asset \( i \) at time \( t \), selling it at time \( t + 1 \) for \((1 + R_{i,t+1})\) and consuming the proceeds. So, (7.6) describes the optimum.

The following time-separable power utility function has been assumed.
\[ U(C_t) = \frac{C_t^{1-\gamma} - 1}{1 - \gamma} \] (7.7)

where \( \gamma \) is the coefficient of relative risk aversion. The power utility function has the desirable property that if different investors in the economy have the same power utility function and can freely trade all the risks they face, then even if they have different wealth levels they can be aggregated into a single representative investor with the same utility function as the individual investors. This provides some justification for the use of aggregate consumption, rather than individual consumption, in the CCAPM.

Taking the derivative of (7.7) with respect to consumption, we find that marginal utility \( U''(C_t) = C_t^{-\gamma} \). Substituting into (7.6) we get
\[ 1 = E_t \left[ (1 + R_{i,t+1})\delta \left( \frac{C_{t+1}}{C_t} \right)^{-\gamma} \right] \] (7.8)

A typical objective of empirical research is to estimate the coefficient of relative risk aversion \( \gamma \) and to test the restrictions imposed by (7.8). It is easiest to do this if one assumes that asset returns and aggregate consumption are jointly homoskedastic and lognormal.
With joint conditional lognormality and homoskedasticity of asset returns and consumption, we can take the logarithm of (7.8), use the notational convention that lower case letters denote logs, and obtain

\[ 0 = E_t[r_{i,t+1}] + \log \delta - \gamma E_t[\Delta C_{t+1}] + \frac{1}{2} \left[ \sigma_i^2 + \gamma^2 \sigma_c^2 - 2\gamma \sigma_{ic} \right] \]  

(7.9)

Here, the notation \( \sigma_{xy} \) denotes the unconditional covariance of innovations \( \text{Cov}(x_{t+1} - E_t[x_{t+1}], y_{t+1} - E_t[y_{t+1}]) \) and \( \sigma^2_x \equiv \sigma_{xx} \). Equation (7.9) which was first derived by Hansen and Singleton(1983) gives a relation between rational expectations of asset returns and rational expectations of consumption growth. It implies that expected asset returns are perfectly correlated with expected consumption growth, but the standard deviation of expected asset returns is \( \gamma \) times as large as the standard deviation of expected consumption growth.

By defining an error term \( \eta_{i,t+1} \equiv r_{i,t+1} - E_t[r_{i,t+1}] - \gamma(\Delta C_{t+1} - E_t[\Delta C_{t+1}]) \), we can rewrite the equation (7.9) as a regression equation

\[ r_{i,t+1} = \mu_i + \gamma \Delta c_{t+1} + \eta_{i,t+1} \]  

(7.10)

In general, the error term \( \eta_{i,t+1} \) will be correlated with realized consumption growth. Thus, Hansen and Singleton(1983) have proposed instrumental variables(IV) regression for the estimation of \( \gamma \).

Campbell, Lo and MacKinlay(1997) estimated the regression model (7.10) using 1889-1994 US annual data. To estimate the coefficient of relative risk aversion \( \gamma \), they used the lagged variables of real commercial paper rate, the real consumption growth rate, and the log dividend-price ratio because \( \eta_{i,t+1} \) is uncorrelated with any variables in the information set at time \( t \).
The 3rd and 4th rows of Table 19 illustrate their two-stage least squares estimation result. In this table the asset returns \((r_{i,t+1})\) are the real commercial paper rate \((CP)\) and the instruments are either one lag (marked 1 in the second column “Lag”) or one and two lags (marked 2), of the real commercial paper rate, the real consumption growth rate, and the log dividend-price ratio. The consumption growth \((\Delta c_{t+1})\) in the regression model is the annual growth rate of real nondurables and services consumption.

Their IV estimates of \(\gamma\) are negative rather than positive as implied by the underlying theory, but they are not significantly different from zero. This presents weak evidence against the standard consumption CAPM with power utility.

Using the estimation method presented in this paper, we reevaluate Campbell, Lo and MacKinlay(1997)’s estimation result. Our regression model with \(m\) breaks can be written as

\[
    r_{i,t+1} = \mu_{i,j} + \gamma_j \Delta c_{t+1} + \eta_{i,t+1} \quad j = 1, \ldots, m+1, \quad t = T_{j-1}, \ldots, T_j
\]

The formulation of the first stage regression model for the endogenous regressor depends on how many lags of the real commercial paper rate \((r_t)\), the real consumption growth rate \((\Delta c_t)\), and the log dividend-price ratio \((d_t)\) are used as the instruments. For the case where instrumental variables consist of one lag of each variable, it has the form of

\[
    \Delta c_{t+1} = \alpha_0 + \alpha_1 r_{i,t} + \alpha_2 \Delta c_t + \alpha_3 d_t + \epsilon_{t+1} \quad (7.11)
\]

On the other hand, if lags one and two are included, the form of the first stage
regression model is
\[
\Delta c_{t+1} = \omega_0 + \omega_1 r_{i,t} + \omega_2 r_{i,t-1} + \omega_3 \Delta c_t + \omega_4 \Delta c_{t-1} + \omega_5 d_t + \omega_6 d_{t-1} + \epsilon_{t+1}
\]
(7.12)

For the comparison of estimation results, the same data set with time span of 1889-1994 has been used. Real commercial paper rate (6-month commercial paper, bought in January and rolled over in July) is used as a dependent variable and real consumption growth rate ($\Delta c_{t+1}$) is an explanatory variable.

$UD_{\text{max}} F_T(5; 1)$ is employed as an initial test for the selection of the number of breaks in the sequential procedure. If the $UD_{\text{max}} F_T(5; 1)$ test is significant, then the $F_T(l + 1|l)$ test statistic is used to test whether there are further breaks in the sample. All the tests have 5% significance level.

When lag one variables are introduced as instruments (model (7.11)), Our 2SLS estimation method detects 1 break at the observation 23 which corresponds to the year 1914. $UD_{\text{max}} F_T(5; 1)$ has p-value within the range (0.01, 0.025) and $\sup F_T(2|1)$ has p-value larger than 0.1. The corresponding regression coefficients and their estimated standard errors are given in Table 19 at the Lag 1 column of the block labeled “Structural break model”. As Campbell, Lo and MacKinlay(1997) mentioned already over their estimation results, the regression coefficients are not significantly different from zero.

We may reach the same kind of conclusions for the lag two variables case. The p-value of $UD_{\text{max}} F_T(5; 1)$ test statistic is within the range of (0.01, 0.025) which corresponds to the same range when lag 1 variables are used as instruments. But this time our 2SLS estimation method detects 3 break points at the observations 29
(year 1920), 42 (year 1931) and 60 (year 1949). The regression coefficient estimates are all negative contrary to the underlying theory but are not significantly different from zero. So, the sample data doesn’t seem to support the standard consumption CAPM with power utility.
Chapter 8

Conclusion

Bai and Perron (1998) consider issues related to multiple structural changes, occurring at unknown dates, in their linear regression model estimated by least squares. They examined the main aspects of the structural break models including the consistency of the break fraction estimators, the rate of convergence, and the construction of tests that allow inference to be made about the presence of structural change and the number of breaks. This thesis extends Bai and Perron’s (1998) results of LS estimation method to the 2SLS estimation method by employing the instrumental variables.

By adopting the proof technique of Bai and Perron (1998), we established the consistency of the break fraction estimator in our 2SLS regression model and showed that it had T-convergence rate. If the break fractions are known a priori then standard arguments can be employed to show the root-T asymptotic normality of 2SLS coefficient estimator. Since the estimated break fractions converge at rate T, this standard asymptotic distribution theory can be extended to the 2SLS estimates based on the estimated fractions. Theorem 3 proved the standard
asymptotic normality of the estimates of the coefficients.

Given the rate of the convergence, it is natural to consider the limiting distribution of the break point estimators as the following step. We considered two frameworks of asymptotic distribution. One is based on the fixed magnitude of shift in the regression parameters, the other on a shrinking magnitude of shift. The application of the limiting distribution for the fixed shift case has some serious limitations due to its highly data-dependent nature. The analytical limiting distribution is hard to obtain in general. So, in Theorem 6 we consider the case of small shifts in the regression coefficients, assuming the magnitude of shifts converges to zero as the sample size increases. In this setup, the limiting distribution is invariant to the underlying distribution of the regressors and error and the resulting distribution can be used as an approximation even for moderate shifts. By assuming weak stationarity of the regressors and conditional homoskedasticity of the error, we showed that the break point estimator converges in distribution to the arg max of two-sided Brownian motion process, $Z(s)$. This is the same distribution which Bai and Perron (1998) obtained under the OLS framework.

The earliest test for the structural break in economic literature is the Chow test which was used extensively in empirical studies in the 1970s. But the Chow test is based on the rather strong assumptions that the regression error term is homoscedastic and the break point is assumed known a priori. To overcome these drawbacks several tests have been proposed including MacKinnon (1989), a robust test statistic for unknown forms of heteroskedasticity and Andrews and Fair (1987) for the generalization to the dynamic linear or nonlinear simultaneous equations models. Recently, Bai and Perron (1998) generalized Andrew (1993)’s sup F type
test to the hypothesis of multiple breaks. By extending Bai and Perron (1998)’s OLS results to the 2SLS setup where the endogenous variables in our regression model are correlated with the regression residuals, we obtained the same limit distribution of the $supF$ test as the one in Bai and Perron (1998)’s analogous result based on OLS estimators when the regressors are exogenous. The sequential F test to determine the number of break points has the same limit distribution as in Bai and Perron (1998).

To verify the analytic results on the break fraction estimators and the test statistics for the structural breaks, a Monte Carlo simulation experiment has been conducted. Through the simulation experiment we verified that the break fraction estimator is consistent and that the relative rejection frequencies of various F statistics under the null hypothesis are close to their nominal size of 0.05. Also we observed that under alternative hypothesis the various F statistics possessed reasonable empirical powers. Finally, in most of the cases the estimated number of breaks coincided with the true number of breaks in the simulation model.

The results in this thesis can be extended in several directions which might compose of my future research agenda. First, we assumed that the first stage regression model was not subject to the structural changes. But it is unnatural to assume parameter constancy in the first regression model, considering widespread effect of economic structural change. Thus, we should incorporate the structural breaks in the first regression model too. Second, we know that OLS estimator in the 2SLS setup is biased. So it might be of interest to investigate the magnitude of bias when we apply OLS procedure when the the endogenous variables in our regression model are correlated with the regression residuals. Third, we might
extend our results on the linear model into nonlinear models such as GMM.
References


Appendix

Table 1: Finite sample behavior of break fraction estimator:

*One Break Model:* $(\beta_1^0, \beta_2^0) = (0.1, -0.1), \ rho = 0.5, \ R^2 = 0.5$

<table>
<thead>
<tr>
<th>$q$</th>
<th>T</th>
<th>Deviation from the True Break Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1 %</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>.19</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>.38</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>.51</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>.66</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>.27</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>.49</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>.66</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>.82</td>
</tr>
<tr>
<td>8</td>
<td>60</td>
<td>.37</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>.66</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>.78</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>.92</td>
</tr>
</tbody>
</table>

*Notes:* The column headed $100a\%$ gives the proportion of the simulations in which $|\hat{\lambda}_1 - \lambda_1^0| \leq a$; $q$ is the number of instruments; $T$ is the sample size.
Table 2: Finite sample behavior of break fraction estimator:

One Break Model: \((\beta_1^0, \beta_2^0) = (1, -1), (5, -5)\), \(\rho = 0.5\), \(R^2 = 0.5\)

<table>
<thead>
<tr>
<th>(q)</th>
<th>(T)</th>
<th>Deviation from the True Break Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 %</td>
<td>2 %</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>1.00</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>1.00</td>
</tr>
<tr>
<td>8</td>
<td>60</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Notes: See Table 1 for definitions.
Table 3: Relative rejection frequencies of $F$–statistics:

One Break Model: $(\beta_1^0, \beta_2^0) = (0.1, -0.1), \rho = 0.5, R^2 = 0.5$

| $q$ | $T$ | $supF(k)$ | $F(l + 1|l)$ | $UDmax$ |
|-----|-----|-----------|--------------|---------|
|     |     | $1$       | $2$          | $2|1$    | $3|2$    |
| 2   | 60  | .68       | .53          | .04     | 0       | .67     |
|     | 120 | .96       | .89          | .04     | 0       | .95     |
|     | 240 | 1.00      | 1.00         | .05     | .01     | 1.00    |
|     | 480 | 1.00      | 1.00         | .03     | 0       | 1.00    |
| 4   | 60  | .93       | .87          | .06     | .01     | .92     |
|     | 120 | 1.00      | 1.00         | .05     | .01     | 1.00    |
|     | 240 | 1.00      | 1.00         | .06     | .01     | 1.00    |
|     | 480 | 1.00      | 1.00         | .06     | .01     | 1.00    |
| 8   | 60  | 1.00      | .99          | .07     | .01     | 1.00    |
|     | 120 | 1.00      | 1.00         | .07     | .01     | 1.00    |
|     | 240 | 1.00      | 1.00         | .08     | .01     | 1.00    |
|     | 480 | 1.00      | 1.00         | .06     | .01     | 1.00    |

Notes: $supF(k)$ denotes the statistic $Sup F_T(k; 1)$ and the second tier column heading under it denotes $k$; $F(l + 1|l)$ denotes the statistic $F_T(l + 1|l)$ and the second tier column beneath it denotes $l + 1|l$; $q$ is the number of instruments; $T$ is the sample size.
Table 4: Relative rejection frequencies of $F$–statistics:

*One Break Model:* $(\beta_1^0, \beta_2^0) = (1, -1), \rho = 0.5, R^2 = 0.5$

| $q$ | $T$ | $\sup F(k)$ | $F(l + 1|l)$ | $U D_{\max}$ |
|-----|-----|-------------|--------------|-------------|
|     |     | 1           | 2            | 2|1 | 3|2 |
| 2   | 60  | 1.00       | 1.00        | .07         | .02         | 1.00 |
|     | 120 | 1.00       | 1.00        | .04         | .01         | 1.00 |
|     | 240 | 1.00       | 1.00        | .05         | .01         | 1.00 |
|     | 480 | 1.00       | 1.00        | .04         | .01         | 1.00 |
| 4   | 60  | 1.00       | 1.00        | .09         | .03         | 1.00 |
|     | 120 | 1.00       | 1.00        | .08         | .02         | 1.00 |
|     | 240 | 1.00       | 1.00        | .07         | .02         | 1.00 |
|     | 480 | 1.00       | 1.00        | .06         | .01         | 1.00 |
| 8   | 60  | 1.00       | 1.00        | .07         | .03         | 1.00 |
|     | 120 | 1.00       | 1.00        | .05         | .02         | 1.00 |
|     | 240 | 1.00       | 1.00        | .08         | .02         | 1.00 |
|     | 480 | 1.00       | 1.00        | .06         | .01         | 1.00 |

*Notes:* See Table 3 for definitions.
Table 5: Relative rejection frequencies of $F$–statistics:

One Break Model: $(\beta_1, \beta_2) = (5, -5)$, $\rho = 0.5$, $R^2 = 0.5$

<table>
<thead>
<tr>
<th>$q$</th>
<th>$T$</th>
<th>$supF(k)$</th>
<th>$F(l + 1/l)$</th>
<th>$UD_{max}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>1.00</td>
<td>1.00</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>1.00</td>
<td>1.00</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>1.00</td>
<td>1.00</td>
<td>.03</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>1.00</td>
<td>1.00</td>
<td>.03</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>1.00</td>
<td>1.00</td>
<td>.07</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>1.00</td>
<td>1.00</td>
<td>.06</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>1.00</td>
<td>1.00</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>1.00</td>
<td>1.00</td>
<td>.05</td>
</tr>
<tr>
<td>8</td>
<td>60</td>
<td>1.00</td>
<td>1.00</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>1.00</td>
<td>1.00</td>
<td>.07</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>1.00</td>
<td>1.00</td>
<td>.07</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>1.00</td>
<td>1.00</td>
<td>.07</td>
</tr>
</tbody>
</table>

Notes: See Table 3 for definitions.
Table 6: Empirical distribution of the estimated number of breaks:

One Break Model: $(\beta_1^0, \beta_2^0) = (0.1, -0.1), \rho = 0.5, R^2 = 0.5$

<table>
<thead>
<tr>
<th>$q$</th>
<th>$T$</th>
<th>$\text{supF}(1)$</th>
<th>$\text{UDmax}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$0$</td>
<td>$1$</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>.32</td>
<td>.66</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>.05</td>
<td>.94</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>0</td>
<td>.97</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>0</td>
<td>.99</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>.07</td>
<td>.90</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>0</td>
<td>.97</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>0</td>
<td>.97</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>0</td>
<td>.97</td>
</tr>
<tr>
<td>8</td>
<td>60</td>
<td>0</td>
<td>.96</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>0</td>
<td>.96</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>0</td>
<td>.96</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>0</td>
<td>.96</td>
</tr>
</tbody>
</table>

Notes: The figures in the block headed $\text{supF}(1)$ give the empirical distribution of the estimated number of breaks, $\hat{k}_T$, obtained via the sequential strategy using $\text{Sup F}_T(1; 1)$ on the first step with the maximum number of breaks equal to five. The figures in the block headed $\text{UDmax}$ give the empirical distribution of the estimated number of breaks, $\hat{k}_T$, obtained via the sequential strategy using $\text{UDmaxF}_T(5, 1)$ on the first step with the maximum number of breaks equal to five.
### Table 7: Empirical distribution of the estimated number of breaks:

**One Break Model**: \((\beta_0^1, \beta_0^2) = (1, -1), \ \rho = 0.5, \ R^2 = 0.5\)

<table>
<thead>
<tr>
<th>(q)</th>
<th>(T)</th>
<th>(supF(1))</th>
<th>(UDmax)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>0</td>
<td>.96</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>0</td>
<td>.98</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>0</td>
<td>.97</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>0</td>
<td>.98</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>0</td>
<td>.94</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>0</td>
<td>.95</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>0</td>
<td>.96</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>0</td>
<td>.97</td>
</tr>
<tr>
<td>8</td>
<td>60</td>
<td>0</td>
<td>.96</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>0</td>
<td>.97</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>0</td>
<td>.95</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>0</td>
<td>.97</td>
</tr>
</tbody>
</table>

*Notes*: See Table 6 for definitions.
Table 8: Empirical distribution of the estimated number of breaks:

*One Break Model*: $(\beta_1^0, \beta_2^0) = (5, -5), \quad \rho = 0.5, \quad R^2 = 0.5$

<table>
<thead>
<tr>
<th>q</th>
<th>$T$</th>
<th>$supF(1)$</th>
<th>$UDmax$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>0</td>
<td>.98</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>0</td>
<td>.98</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>0</td>
<td>.98</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>0</td>
<td>.99</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>0</td>
<td>.96</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>0</td>
<td>.97</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>0</td>
<td>.97</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>0</td>
<td>.98</td>
</tr>
<tr>
<td>8</td>
<td>60</td>
<td>0</td>
<td>.94</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>0</td>
<td>.97</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>0</td>
<td>.96</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>0</td>
<td>.96</td>
</tr>
</tbody>
</table>

*Notes*: See Table 6 for definitions.
Table 9: Finite sample behavior of break fraction estimator:

*Two Breaks Model:* \((\beta_1^0, \beta_2^0, \beta_3^0) = (-0.1, 0.1, -0.1), \ \rho = 0.5, \ R^2 = 0.5*

<table>
<thead>
<tr>
<th>(q)</th>
<th>(T)</th>
<th>(i-th) Break</th>
<th>Deviation from the True Break Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 %</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>1st</td>
<td>.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>.19</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>1st</td>
<td>.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>.37</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>1st</td>
<td>.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>.51</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>1st</td>
<td>.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>.68</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>1st</td>
<td>.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>.26</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>1st</td>
<td>.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>.50</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>1st</td>
<td>.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>.64</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>1st</td>
<td>.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>.82</td>
</tr>
<tr>
<td>8</td>
<td>60</td>
<td>1st</td>
<td>.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>.34</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>1st</td>
<td>.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>.61</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>1st</td>
<td>.79</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>.78</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>1st</td>
<td>.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>.91</td>
</tr>
</tbody>
</table>

*Notes:* See Table 1 for definitions.
Table 10: Finite sample behavior of break fraction estimator:

Two Breaks Model: \((\beta_0^0, \beta_0^1, \beta_0^2) = (-1, 1, -1), (-5, 5, -5), \ \rho = 0.5, \ R^2 = 0.5\)

<table>
<thead>
<tr>
<th>q</th>
<th>T</th>
<th>i - th Break</th>
<th>Deviation from the True Break Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 %</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>1st</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>1st</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>1st</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>1st</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>1.00</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>1st</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>1st</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>1st</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>1st</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>1.00</td>
</tr>
<tr>
<td>8</td>
<td>60</td>
<td>1st</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>1st</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>1st</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>1st</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Notes: See Table 1 for definitions.
Table 11: Relative rejection frequencies of $F$–statistics:

Two Breaks Model: $(\beta_0^1, \beta_0^2, \beta_0^3) = (-0.1, 0.1, -0.1), \ \rho = 0.5, \ R^2 = 0.5$

| $q$ | $T$ | $supF(k)$ | $F(l + 1|l)$ | $UD_{max}$ |
|-----|-----|-----------|--------------|------------|
|     |     | 1   | 2   | 2|1 | 3|2 |
| 2   | 60  | .18 | .48 | .38 | .02 | .38 |
|     | 120 | .42 | .84 | .71 | .01 | .76 |
|     | 240 | .87 | 1.00 | .99 | .01 | 1.00 |
|     | 480 | 1.00 | 1.00 | 1.00 | .02 | 1.00 |
| 4   | 60  | .30 | .77 | .65 | .03 | .65 |
|     | 120 | .73 | .99 | .95 | .02 | .98 |
|     | 240 | .99 | 1.00 | 1.00 | .03 | 1.00 |
|     | 480 | 1.00 | 1.00 | 1.00 | .03 | 1.00 |
| 8   | 60  | .49 | .96 | .89 | .05 | .93 |
|     | 120 | .97 | 1.00 | 1.00 | .05 | 1.00 |
|     | 240 | 1.00 | 1.00 | 1.00 | .03 | 1.00 |
|     | 480 | 1.00 | 1.00 | 1.00 | .04 | 1.00 |

Notes: See Table 3 for definitions.
Table 12: Relative rejection frequencies of $F$–statistics:

Two Breaks Model: $(\beta_0^1, \beta_2^1, \beta_0^2) = (-1, 1, -1), \ \rho = 0.5, \ R^2 = 0.5$

| $q$ | $T$ | $supF(k)$ | $F(l + 1|l)$ | $UDmax$ |
|-----|-----|-----------|-------------|---------|
|     |     | 1 | 2 | 2|1 | 3|2 |     |     |
| 2   |  60 | 1.00 | 1.00 | 1.00 | .06 | 1.00 |
|     | 120 | 1.00 | 1.00 | 1.00 | .04 | 1.00 |
|     | 240 | 1.00 | 1.00 | 1.00 | .03 | 1.00 |
|     | 480 | 1.00 | 1.00 | 1.00 | .03 | 1.00 |
| 4   |  60 | 1.00 | 1.00 | 1.00 | .05 | 1.00 |
|     | 120 | 1.00 | 1.00 | 1.00 | .04 | 1.00 |
|     | 240 | 1.00 | 1.00 | 1.00 | .04 | 1.00 |
|     | 480 | 1.00 | 1.00 | 1.00 | .04 | 1.00 |
| 8   |  60 | 1.00 | 1.00 | 1.00 | .07 | 1.00 |
|     | 120 | 1.00 | 1.00 | 1.00 | .05 | 1.00 |
|     | 240 | 1.00 | 1.00 | 1.00 | .05 | 1.00 |
|     | 480 | 1.00 | 1.00 | 1.00 | .04 | 1.00 |

Notes: See Table 3 for definitions.
Table 13: Relative rejection frequencies of $F$–statistics:

**Two Breaks Model:** $(\beta_{10}, \beta_{20}, \beta_{30}) = (-5, 5, -5), \quad \rho = 0.5, \quad R^2 = 0.5$

| $q$ | $T$ | $supF(k)$ | $F(l + 1|l)$ | $UDmax$ |
|-----|-----|-----------|--------------|---------|
|     |     | 1 | 2 | 2|1 | 3|2 |       |
| 2   | 60  | 1.00 | 1.00 | 1.00 | .05 | 1.00 |       |
|     | 120 | 1.00 | 1.00 | 1.00 | .02 | 1.00 |       |
|     | 240 | 1.00 | 1.00 | 1.00 | .02 | 1.00 |       |
|     | 480 | 1.00 | 1.00 | 1.00 | .01 | 1.00 |       |
| 4   | 60  | 1.00 | 1.00 | 1.00 | .06 | 1.00 |       |
|     | 120 | 1.00 | 1.00 | 1.00 | .04 | 1.00 |       |
|     | 240 | 1.00 | 1.00 | 1.00 | .03 | 1.00 |       |
|     | 480 | 1.00 | 1.00 | 1.00 | .03 | 1.00 |       |
| 8   | 60  | 1.00 | 1.00 | 1.00 | .05 | 1.00 |       |
|     | 120 | 1.00 | 1.00 | 1.00 | .05 | 1.00 |       |
|     | 240 | 1.00 | 1.00 | 1.00 | .04 | 1.00 |       |
|     | 480 | 1.00 | 1.00 | 1.00 | .04 | 1.00 |       |

*Notes:* See Table 3 for definitions.
Table 14: Empirical distribution of the estimated number of breaks:

Two Breaks Model: \((\beta_1^0, \beta_2^0, \beta_3^0) = (-0.1, 0.1, -0.1), \ \rho = 0.5, \ R^2 = 0.5 \)

<table>
<thead>
<tr>
<th>q</th>
<th>( T )</th>
<th>( supF(1) )</th>
<th>( UDmax )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>.83</td>
<td>.11</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>.58</td>
<td>.13</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>.13</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>.70</td>
<td>.10</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>.27</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>.01</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>60</td>
<td>.51</td>
<td>.07</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>.04</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Notes: See Table 6 for definitions.
Table 15: Empirical distribution of the estimated number of breaks:

Two Breaks Model: \((\beta_0^1, \beta_1^1, \beta_3^0) = (-1, 1, -1), \ \rho = 0.5, \ R^2 = 0.5\)

<table>
<thead>
<tr>
<th>q</th>
<th>T</th>
<th>(supF(1))</th>
<th>(UDmax)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.95</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.95</td>
<td>0.05</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>0.97</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>0.98</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>0.98</td>
<td>0.02</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>0.98</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>0.97</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>0.97</td>
<td>0.03</td>
</tr>
<tr>
<td>8</td>
<td>60</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>0.96</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>0.97</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>0.97</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Notes: See Table 6 for definitions.
Table 16: Empirical distribution of the estimated number of breaks:

Two Breaks Model: \((\beta_1^0, \beta_2^0, \beta_3^0) = (-5, 5, -5), \ \rho = 0.5, \ R^2 = 0.5\)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>(\sup F(1))</th>
<th></th>
<th>(UD_{max})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>0</td>
<td>0</td>
<td>.96</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>0</td>
<td>0</td>
<td>.99</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>0</td>
<td>0</td>
<td>.99</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>0</td>
<td>0</td>
<td>.99</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>0</td>
<td>0</td>
<td>.95</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>0</td>
<td>0</td>
<td>.97</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>0</td>
<td>0</td>
<td>.98</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>0</td>
<td>0</td>
<td>.98</td>
</tr>
<tr>
<td>8</td>
<td>60</td>
<td>0</td>
<td>0</td>
<td>.96</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>0</td>
<td>0</td>
<td>.96</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>0</td>
<td>0</td>
<td>.97</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>0</td>
<td>0</td>
<td>.97</td>
</tr>
</tbody>
</table>

Notes: See Table 6 for definitions.
Table 17: Relative rejection frequencies of $F$–statistics:

No Break Model: $(\beta_1^0, \beta_2^0) = (1, 1), \rho = 0.5, R^2 = 0.5$

| $q$ | $T$ | $supF(k)$ | $F(l + 1|l)$ | $UD_{max}$ |
|-----|-----|-----------|--------------|------------|
|     |     | 1 | 2 | 2|1 | 3|2 |
| 2   | 60  | .03 | .03 | .03 | 0 | .03 |
|     | 120 | .03 | .02 | .03 | 0 | .03 |
|     | 240 | .03 | .02 | .02 | 0 | .02 |
|     | 480 | .02 | .01 | .02 | 0 | .02 |
| 4   | 60  | .03 | .03 | .04 | .01 | .04 |
|     | 120 | .04 | .03 | .04 | 0  | .04 |
|     | 240 | .04 | .03 | .03 | 0  | .03 |
|     | 480 | .04 | .04 | .03 | 0  | .03 |
| 8   | 60  | .04 | .04 | .04 | .01 | .04 |
|     | 120 | .04 | .05 | .04 | 0  | .04 |
|     | 240 | .05 | .05 | .04 | 0  | .04 |
|     | 480 | .06 | .05 | .04 | 0  | .06 |

Notes: $supF(k)$ denotes the statistic $Sup \ F_T(k;1)$ and the second tier column heading under it denotes $k$; $F(l + 1|l)$ denotes the statistic $F_T(l + 1|l)$ and the second tier column beneath it denotes $l + 1|l$; $q$ is the number of instruments; $T$ is the sample size.
Table 18: Empirical distribution of the estimated number of breaks:

*No Break Model:* $(\beta_1^0, \beta_2^0) = (1, 1), \rho = 0.5, R^2 = 0.5$

<table>
<thead>
<tr>
<th>$q$</th>
<th>$T$</th>
<th>$supF(1)$</th>
<th>$UDmax$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>.97</td>
<td>.03</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>.97</td>
<td>.03</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>.97</td>
<td>.03</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>.98</td>
<td>.02</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>.97</td>
<td>.03</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>.96</td>
<td>.04</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>.96</td>
<td>.04</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>.96</td>
<td>.04</td>
</tr>
<tr>
<td>8</td>
<td>60</td>
<td>.96</td>
<td>.04</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>.96</td>
<td>.04</td>
</tr>
<tr>
<td></td>
<td>240</td>
<td>.96</td>
<td>.04</td>
</tr>
<tr>
<td></td>
<td>480</td>
<td>.94</td>
<td>.05</td>
</tr>
</tbody>
</table>

Notes: The figures in the block headed $supF(1)$ give the empirical distribution of the estimated number of breaks, $\hat{k}_T$, obtained via the sequential strategy using $Sup F_T(1; 1)$ on the first step with the maximum number of breaks equal to five. The figures in the block headed $UDmax$ give the empirical distribution of the estimated number of breaks, $\hat{k}_T$, obtained via the sequential strategy using $UDmax F_T(5, 1)$ on the first step with the maximum number of breaks equal to five.
Table 19: Comparison of the estimation results of Consumption based Capital Asset Pricing Model

<table>
<thead>
<tr>
<th>Lag</th>
<th>( r_{i,t+1} )</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1.98</td>
<td>-0.95</td>
</tr>
<tr>
<td>2</td>
<td>1.31</td>
<td>0.56</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Campbell Model</th>
<th>( \hat{\gamma} )</th>
<th>( \hat{\gamma} ) s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.95</td>
<td>0.56</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Structural Break Model</th>
<th>Estimated number of breaks</th>
<th>1</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimated breaks</td>
<td>23</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td></td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>( \hat{\gamma}_j )</td>
<td>1.45</td>
<td>-0.34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-2.92</td>
<td>-1.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.36</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.94</td>
<td></td>
</tr>
<tr>
<td>( \hat{\gamma}_j )</td>
<td></td>
<td>3.86</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.26</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.76</td>
<td></td>
</tr>
<tr>
<td>( \hat{\gamma}_j )</td>
<td></td>
<td>1.76</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.70</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Regression equation for Campbell Model is \( r_{i,t+1} = \mu_i + \gamma \Delta c_{t+1} + \eta_{i,t+1} \) and regression equation of \( j^{th} \) regime for Structural Break Model is \( r_{i,t+1} = \mu_{i,j} + \gamma_j \Delta c_{t+1} + \eta_{i,t+1} \)