The Asymptotic Behaviour of the Residual Sum of Squares
in Models with Multiple Break Points

Alastair R. Hall
University of Manchester
Denise R. Osborn
University of Manchester
Nikolaos D. Sakkas
University of Bath

March 9, 2014

1The authors gratefully acknowledge the support of the ESRC grant RES-062-23-1351. The paper has benefitted from the comments of participants at the Ken@75 Conference, Warwick University and at a seminar at Universidad Carlos III de Madrid.

2Corresponding author. Economics, SoSS, University of Manchester, Manchester M13 9PL, UK.
Email: alastair.hall@manchester.ac.uk
Abstract

There has been a considerable literature on least squares-based estimation and testing in models with multiple discrete breaks in the parameters, see *inter alia* Bai and Perron (1998) and Hall, Han, and Boldea (2012) and Boldea and Hall (2013). In these contexts, if the model is assumed to have $m$ breaks then the break points (the points at which the parameters change) are estimated by minimizing the residual sum of squares over all possible data partitions involving $m$ breaks. A natural side-product of this estimation is this minimized residual sum of squares and this quantity plays an important role in subsequent inferences about the model. This paper, firstly, derives the asymptotic expectation of the residual sum of squares, the form of which indicates that the number of estimated break points and the number of regression parameters affect this expectation in different ways. Secondly, we propose a statistic for testing the joint hypothesis that the breaks occur at specified fixed break points in the sample. Under its null hypothesis, this statistic is shown to have a limiting distribution that is non-standard but simulatable, being a functional of independent random variables with exponential distributions whose parameters can be consistently estimated. In a special case, it is possible to normalize the statistic to make it pivotal and we provide percentiles for the associated limiting distribution. Our results cover the cases of either the linear or nonlinear regression model with exogenous regressors estimated via Ordinary (or Nonlinear) Least Squares which or a linear model in which some regressors are endogenous and the model is estimated via Two Stage Least Squares.

*JEL classification:* C12, C13, C26

*Keywords:* Linear models, Nonlinear models, Ordinary Least Squares, Two-Stage Least Squares, Parameter Change
1 Introduction

There has been a considerable literature in econometrics on least squares-based estimation and testing in models with discrete breaks in the parameters. The seminal paper by Bai and Perron (1998) developed a framework for estimation and inference in linear regression models estimated via Ordinary Least Squares (OLS) that has served as the template for similar frameworks in more general models, including systems of linear regression models (Perron and Qu, 2006), linear models with endogenous regressors estimated via Two Stage Least Squares (2SLS, Hall, Han, and Boldea, 2012), and nonlinear regression models estimated by Nonlinear Least Squares (NLS, Boldea and Hall, 2013).

Within these models, the key parameters of interest are those indexing the breaks - the break fractions - and the regime specific coefficients. If the model in question is assumed to have \( m \) breaks, then these key parameters are estimated by minimizing the residual sum of squares over all possible data partitions involving \( m \) breaks. The asymptotic analysis then focuses on establishing the consistency of and a limiting distribution theory for these parameters, and also on the development of a limiting distribution theory for statistics for testing hypotheses about the number of breaks. However, relatively little attention has been paid to the minimized residual sum of squares per se, despite its key role in inference for these models.

The first study to examine analytically the consequences of coefficient break point estimation on the residual sum of squares appears to be Ninomiya (2005), who considers breaks in the mean of a Gaussian process with inference on the number of breaks conducted through AIC (the Akaike Information Criterion) viewed as the bias-corrected maximum log-likelihood estimator. Ninomiya (2005) finds the required bias correction implies that estimation of each break fraction parameter has an impact equivalent to three mean parameters on the the maximized log likelihood. This result, namely weighting estimation of each break as three times that of an individual regression coefficient, is used by Hall, Osborn, and Sakkas (2013) to propose a modified penalty term for an information criterion that is employed when the number and dates of breaks are estimated, along with the regression coefficients, in an OLS context. Although the Monte Carlo study of Hall, Osborn, and Sakkas (2013) shows the modified criteria to perform well, their paper provides no formal analytical results.

The present paper examines the asymptotic behaviour of the residual sum of squares in the
types of model listed in the opening paragraph. To capture the realistic situation where the
precise dates of change are unclear, the breaks are assumed to be magnitude that “shrinks” with
the sample size; see Bai (1997).

Firstly, we derive the asymptotic expectation of the residual sum of squares. For linear or
nonlinear regression models with exogenous regressors, this expectation depends on the num-
bers of estimated break points and estimated mean parameters, with estimation of each break
point again having a weight of three relative to each mean parameter. For linear models with
endogenous regressors estimated via 2SLS, the asymptotic bias depends on the same factors as
the exogenous regressor case but also on the degree of overidentification, the variance structure
of the structural and reduced form errors and, if the reduced form is itself subject to discrete
breaks which are consistently estimated as part of the first stage of 2SLS, the relative locations
of the structural and reduced form breaks. Although the impact of coefficient estimation is
modified by the relative locations of structural and reduced form breaks when the latter apply,
evertheless the result that each estimated break date has the same impact on the expectation
as three estimated mean parameters largely carries over also to the 2SLS context.

Secondly, we propose a statistic for testing the joint hypothesis that the breaks occur at
specified fixed break points in the sample. Under its null hypothesis, this statistic is shown to have
a limiting distribution that is non-standard but simulatable, being a functional of independent
random variables with exponential distributions whose parameters can be consistently estimated.
In a special case, it is possible to normalize the statistic to make it asymptotically pivotal and we
provide percentiles for the associated limiting distribution. To our knowledge, no such joint test
has been proposed in the literature. We also discuss how this statistic can be used to construct
confidence sets for the break fractions.

An outline of the remainder of the paper is as follows. Section 2 presents the asymptotic
expectation of the minimized residual sum of squares in regression models with exogenous regres-
sors. Section 3 then examines the case of a model with some endogenous regressors estimated
via 2SLS, where the reduced form may be either stable (with no breaks) or unstable and subject
to breaks that need not coincide with those of the structural form. Section 4 proposes a test for
the joint hypothesis that the breaks occur at certain pre-specified points in the sample. A Monte
Carlo analysis is employed in Section 5 to examine the implications for inference on individual
coefficients. Section 6 concludes. All proofs are relegated to a mathematical appendix.


2 RSS with Exogenous Regressors

Our analysis of the asymptotic expectation of the residual sum of squares covers both linear and nonlinear regression models estimated by least squares. However, since the assumptions differ in some important ways, it is convenient to treat the two cases separately.

2.1 Linear models

Consider the case in which the equation of interest is a linear regression model exhibiting $m$ breaks, such that

$$y_t = x_t' \beta_i + u_t, \quad i = 1, ..., m + 1, \quad t = T^0_{i-1} + 1, ..., T^0_i,$$  \hspace{1cm} (1)

where $T^0_0 = 0$ and $T^0_{m+1} = T$, where $T$ is the total sample size. Thus, $y_t$ is the dependent variable, while $x_t$ is a $p \times 1$ vector of exogenous explanatory variables that includes the constant term, and $u_t$ is a mean zero error. As usual in the literature, we require the true break points to be asymptotically distinct.

Assumption 1 $T^0_i = [T \lambda^0_i], \text{ where } 0 < \lambda^0_1 < ... < \lambda^0_m < 1$. \hspace{1cm} (1)

Suppose now that a researcher knows the number of the breaks but not their location(s). We use $\lambda$ to denote an arbitrary set of $m$ break fractions, with $\lambda = [\lambda_1, \lambda_2, ..., \lambda_m]'$ and $0 < \lambda_1 < \lambda_2 < ... < \lambda_m < 1$, $\lambda_0 = 0$, and $\lambda_{m+1} = 1$. In order to minimize the overall residual sum of squares, the researcher estimates the regression model

$$y_t = x_t' \beta^*_i + e^*_t, \quad i = 1, ..., m + 1, \quad t = T_{i-1} + 1, ..., T_i,$$  \hspace{1cm} (2)

for each possible unique $m$-partition of the sample, where $T_i = [\lambda_i T]$, and $e^*_t$ is an error term. This is embodied in the following assumption:

Assumption 2 Equation (2) is estimated over all partitions $(T_1, ..., T_m)$ such that $T_i - T_{i-1} > \max\{q - 1, \epsilon T\}$ for some $\epsilon > 0$ and $\epsilon < \inf_i (\lambda^0_{i+1} - \lambda^0_i)$.  \hspace{1cm} (2)

Assumption 2 requires that each segment considered in the estimation contains a positive fraction of the sample asymptotically; in practice $\epsilon$ is chosen to be small in the hope that the

\[1\lfloor \cdot \rfloor\] denotes the integer part of the quantity in the brackets.
last part of the assumption is valid. The estimates of $\beta^* = (\beta_1^*, \beta_2^*, ..., \beta_{m+1}^*)^\prime$ are obtained by minimizing the sum of squared residuals

$$S_T(T_1, ..., T_m; \beta) = \sum_{i=1}^{m+1} \sum_{t=T_i-1+1}^{T_i} \{y_t - x_t^\prime \beta_i\}^2$$

with respect to $\beta = (\beta_1^*, \beta_2^*, ..., \beta_{m+1}^*)^\prime$. We denote these estimators by $\hat{\beta}(\{T_i\}_{i=1}^m)$ with $\hat{\beta}_j(\{T_i\}_{i=1}^m)$ being the associated estimator of $\beta_j^*$. The estimators of the break points, $(\hat{T}_1, ..., \hat{T}_m)$, are then defined as

$$(\hat{T}_1, ..., \hat{T}_m) = \arg\min_{T_1, ..., T_m} S_T \left( T_1, ..., T_m; \hat{\beta}(\{T_i\}_{i=1}^m) \right)$$

where the minimization is taken over all possible partitions, $(T_1, ..., T_m)$, and the associated minimized residual sum of squares is denoted $RSS(\hat{T}_1, \hat{T}_2, ..., \hat{T}_m) = S_T \left( \hat{T}_1, ..., \hat{T}_m; \hat{\beta}(\{T_i\}_{i=1}^m) \right)$.

The OLS estimates, $\hat{\beta}(\{\hat{T_i}\}_{i=1}^m)$, are then the regression parameter estimates associated with the estimated partitions. The estimated break fractions are collected in $\hat{\lambda}$, the $m \times 1$ vector with $j^{th}$ element $\hat{T}_j/T$. Bai (1997) and Bai and Perron (1998) derive the large sample behaviour of $\hat{\lambda}$ and $\hat{\beta}(\{\hat{T_i}\}_{i=1}^m)$ and various tests for parameter variation that naturally arise in this context.

Our focus is the large sample behaviour of the minimized residual sum of squares. To this end, we consider the asymptotic expectation of the bias term

$$\xi_T = RSS(\hat{T}_1, \hat{T}_2, ..., \hat{T}_m) - T\sigma^2,$$

where

$$RSS(T_1, ..., T_m) = \sum_{j=1}^m RSS_j(T_1, ..., T_m)$$

$$RSS_j(T_1, ..., T_m) = \sum_{t=T_{j-1}+1}^{T_j} \left\{ y_t - x_t^\prime \hat{\beta}_j(\{T_i\}_{i=1}^m) \right\}^2.$$  

Hence $\xi_T$ defined by (5) is the difference between the (minimized) residual sum of squares in (2) and the expected error sum of squares, $T\sigma^2 = E[\sum_{t=1}^{T} u_t^2]$, in the data generating process (DGP) of (1).

We decompose $\xi_T$ into three components,

$$\xi_T = \sum_{j=1}^3 \xi_{j,T}.$$  

The first component,

$$\xi_{1,T} = RSS(\hat{T}_1, \hat{T}_2, ..., \hat{T}_m) - RSS(T_1^0, T_2^0, ..., T_m^0),$$

4
represents the effect on the residual sums of squares from using the estimated rather than the true break dates. The second component is defined as

\[ \xi_{2,T} = \text{RSS}(T_1^0, T_2^0, \ldots, T_m^0) - \text{ESS}(T_1^0, T_2^0, \ldots, T_m^0), \tag{10} \]

where \( \text{ESS}(T_1^0, T_2^0, \ldots, T_m^0) \) is the error sum of squares for (1) evaluated using the true \( \{\beta_i^0\}_{i=1}^{m+1} \). Hence \( \xi_{2,T} \) is the impact on the residual sum of squares from estimating the coefficients of (1) with known (true) break dates. The final component is

\[ \xi_{3,T} = \text{ESS}(T_1^0, T_2^0, \ldots, T_m^0) - T\sigma^2, \tag{11} \]

and therefore captures the effects of the specific random disturbances \( u_t \).

Let \( AE[\cdot] \) denote the asymptotic expectation operator. To derive the \( AE[\xi_T] \), we make the following assumption about magnitudes of the breaks.

**Assumption 3** \( \beta_{i+1}^0 - \beta_i^0 = \theta_i^0 \sigma T \) where \( \theta_i^0 = T^{-\alpha} \) for some \( \alpha \in (0, 0.5) \) and \( i = 1, 2, \ldots, m \).

Assumption 3 is the so-called “shrinking breaks” case, which is designed to capture the situation in which there is uncertainty about the location of the breaks in moderate sized samples. This assumption, with breaks restricted to shrink at a slower rate than \( T^{-1/2} \), is commonly employed in the literature to deduce a limiting distribution for break-point estimators; see Bai (1997) and Bai and Perron (1998). Similar assumptions are inter alia \( T^{1/2}w_T \to \infty \), in (Bai and Perron 1998) and \( T^{1/2}w_T/(\log T)^2 \to \infty \) in (Qu and Perron 2007).

Assumptions are also imposed about the regressors and errors, as follows.

**Assumption 4** \( T^{-1} \sum_{t=T_{i-1}^0+1}^{T_i^0} x_t x_t' \to_{P}^{r} rQ_i \) uniformly in \( r \in (0, \lambda_i^0 - \lambda_{i-1}^0) \), where \( Q_i \) is a positive definite matrix for \( i = 1, 2, \ldots, m + 1 \).

**Assumption 5** (i) \( E[u_t | \mathcal{F}_t] = 0 \) where \( \mathcal{F}_t \) is the \( \sigma \)-algebra generated by \( \{x_t, u_{t-1}, x_{t-1}, u_{t-2}, \ldots\} \); (ii) \( E[\|h_i\|_d] < H_d < \infty \) for \( t = 1, 2, \ldots \) and some \( d > 2 \), where \( h_{t,i} \) is the \( i^{th} \) element of \( h_t = u_t x_t \); (iii) \( V_{T,i}(r) = \text{Var}[T^{-1/2} \sum_{t=T_{i-1}^0+1}^{T_i^0} h_t] \) is uniformly positive definite for all \( T \) sufficiently large and \( \lim_{T \to \infty} V_{T,i}(r) = rV_i \), uniformly in \( r \in (0, \lambda_i^0 - \lambda_{i-1}^0) \) where \( V_i \) is a positive definite matrix of constants; (iv) \( \sigma_i^2 = E[u_t^2 | \mathcal{F}_t, t/T \in [\lambda_{i-1}^0, \lambda_i^0]] \) is a positive finite constant for all \( i \); (v) \( \sigma_i^2 = \sigma^2, i = 1, 2, \ldots, m + 1 \).

\(^2\)That is there exists \( \gamma \) such that \( c'V_T(1)c > \gamma > 0 \) for all vectors of constants \( c \) such that \( \|c\| = 1 \).
Assumption 6 There exists an $l_0 > 0$ such that for all $l > l_0$, the minimum eigenvalues of $A_{il} = (1/l) \sum_{t=T_0+1}^{T_0+l} x_i x'_t$ and of $\bar{A}_{il} = (1/l) \sum_{t=T_0-l}^{T_0} x_i x'_t$ are bounded away from zero for all $i = 1, ..., m + 1$.

Assumption 6 limits the behaviour of the regressor cross product matrix and rules out trending regressors but allows regime specific behaviour. Assumption 5(i)-(iii) ensures \{x_t u_t\} satisfies the Functional Central Limit Theorem within each regime (e.g. see White (2001)[Theorem 7.19]); parts (iv)-(v) place restrictions on $V_i$, and while both are imposed in this part of the analysis, this will be relaxed in later sections and hence are stated separately in the assumption. Finally, Assumption 6 requires that there be enough observations near the true break points so that they can be identified and is analogous to the extension proposed in Bai and Perron (1998) to their Assumption A2.

The following theorem gives the form of $AE[\xi_T]$ for this model.

**Theorem 1** Let $y_t$ be generated by (3), and Assumptions 1-6 hold. Then we have: (i) $AE[\xi_{1,T}] = -3m\sigma^2$; (ii) $AE[\xi_{2,T}] = -p(m + 1)\sigma^2$; (iii) $AE[\xi_{3,T}] = 0$; and so

$$AE[\xi_T] = -[(p + 3)m + p]\sigma^2.$$ 

**Remark 1:** Note that Theorem 1 implies $T^{-1} RSS(\hat{T}_1, \hat{T}_2, ..., \hat{T}) \xrightarrow{p} \sigma^2$ as previously documented in the literature, for example see Bai (1997) and Bai and Perron (1998).

**Remark 2:** A comparison of $AE[\xi_{1,T}]$ and $AE[\xi_{2,T}]$ indicates that the break parameters and the regression parameters affect $AE[\xi_T]$ differently. Within this model, an additional break increases the number of break fractions by one and the number of regression parameters by $p$ (equal to the number of regression coefficients in the additional regime). From Theorem 1(i), it can be seen that the bias due to estimation of each break increases in absolute value by $3\sigma^2$, thus $3\sigma^2$ times the number of additional breaks in total. From Theorem 1(ii), the asymptotic bias due to estimation of the additional regression parameters increases in absolute value by $p\sigma^2$, thus $\sigma^2$ times the number of additional regression parameters. As noted by Ninomiya (2005) in his analysis of the mean shift model, this can be interpreted as implying estimation of the break fraction has three time the impact of estimation of a regression parameter on the bias, providing
a theoretical motivation for the modified information criteria penalty function proposed by Hall, Osborn, and Sakkas (2013) in the context of structural break estimation.

The limiting distribution of the component $\xi_{1,T}$ plays a key role in the result of Theorem 1 and our subsequent analysis. Using the same arguments as Bai (1997) (see also the Appendix below), for an individual break $i$ we need to consider $T_i = T_{i0}^0 + [k_i, s_T^{-2}]$ for $k_i \epsilon [-K_i, K_i]$ and

$$\xi_{1,T} \xrightarrow{d} \sum_{i=1}^{m} \min_{k_i} G_i(k_i)$$

where

$$G_i(k_i) = \begin{cases} |k_i| a_{i,1} - 2 c_{i,1}^{1/2} W_{i,1}(-k_i), & k_i \leq 0 \\ |k_i| a_{i,2} - 2 c_{i,2}^{1/2} W_{i,2}(k_i), & k_i > 0 \end{cases}$$

in which $W_{i,j}(\cdot)$ are independent Brownian motions on $[0, \infty)$ and

$$a_{i,j} = \theta_i^0 Q_{(i-1)+j} \theta_i^0$$

$$c_{i,j} = \theta_i^0 V_{(i-1)+j} \theta_i^0.$$

Clearly, minimization of $G_i(k_i)$ is equivalent to maximization of $\tilde{G}_i(k_i) = -G_i(k_i)$, namely the maximum of two independent Brownian motion processes with negative drift.

The following Lemmata and Definition provide distributional results for this maximization.

**Lemma 1.** Let $W(\cdot)$ be standard Brownian motion on $[0, \infty)$. Then, for $\alpha > 0$, $\gamma > 0$ and $k \epsilon [0, \infty)$

$$\Pr\left\{ \max_{k} [\gamma W(k) - \alpha k] > m \right\} = \exp(-\mu m)$$

which is the cumulative distribution function (cdf) of the exponential distribution with parameter $\mu = 2\alpha/\gamma^2$.

**Definition 1** Let $B(\mu_1, \mu_2)$ denote the distribution with cdf

$$F(w; \mu_1, \mu_2) = (1 - e^{-\mu_1 w})(1 - e^{-\mu_2 w})$$

$$= \int_0^w f(b; \mu_1, \mu_2) \, db$$

where

$$f(b; \mu_1, \mu_2) = \sum_{i=1}^{2} \mu_i e^{-b \mu_i} - \mu e^{-b \mu}$$

**Lemma 1** is stated in Bai (1997)[p.563] and, for $\gamma = 1$, in Stryhn (1996)[Proposition 1].
for $\mu = \sum_{i=1}^{2} \mu_i$.

**Lemma 2** Let $v_i \sim \text{exponential}(\mu_i)$ for $i = 1, 2$ and $v_1 \perp v_2$. Then $b = \max\{v_1, v_2\} \sim \mathcal{B}(\mu_1, \mu_2)$ and

$$E[b] = \mu_1^{-1} + \mu_2^{-1} - (\mu_1 + \mu_2)^{-1}. \quad (17)$$

Lemma 1 states that the maximum value taken by an individual Brownian motion process with negative drift follows an exponential distribution, with the maximum of two independent processes then having the form of (16). Lemma 2 follows from the mean of an exponential distribution.

If only parts (i)-(iii) of Assumption 5 are imposed then we have

$$b_i = \min_{k_i} G_i(k_i) \sim \mathcal{B}(\mu_{i,1}, \mu_{i,2}), \quad (18)$$

where $\mu_{i,j} = 0.5a_{i,j}/c_{i,j}$. If parts (iv) and (v) of Assumption 5 are also imposed then from (13)-(15), $\mu_{i,j} = 0.5\sigma^{-2}$, so that

$$b_i = \min_{k_i} G_i(k_i) \sim \mathcal{B}(0.5\sigma^{-2}, 0.5\sigma^{-2}) \quad (19)$$

and hence (from Lemma 2) $E[b_i] = 3\sigma^2$ for each break. With $m$ distinct break points, $\xi_{1,T}$ in (12) converges in distribution to the sum of $m$ independent variables, each of the form of (19) and with an expected value of $3\sigma^2$.

### 2.2 Nonlinear models

Analogously to (1), consider a univariate nonlinear model with $m$ unknown breaks:

$$y_t = f(x_t, \beta^0_i) + u_t, \quad i = 1, \ldots, m + 1, \quad t = T^0_{i-1} + 1, \ldots, T^0_i, \quad (20)$$

where $f : \mathbb{R}^q \times \mathbb{B} \to \mathbb{R}$ is a known measurable function on $\mathbb{R}$ for each $\beta \in \mathbb{B}$. For simplicity, let $f_t(\beta) = f(x_t, \beta)$. To avoid excessive notation we redefine the estimators and residual sum of squares analogously to Section 2.1 only replacing $x'_t \beta_i$ by $f_t(\beta_i)$ in (3).

To date in the NLS setting, the consistency and large sample distribution of $\hat{\lambda}$ and $\hat{\beta}(\{(\hat{T}_i)_{i=1}^{m}\})$ have only been established under more restrictive conditions on the dynamic structure of the data and also the rate of shrinkage between regimes; see Boldea and Hall (2013)[Assumptions 2-8]. These additional restrictions arise because of the inherent nonlinearity of the model; see Boldea and Hall (2013) for further discussion. We impose these conditions here but for brevity
relegate a full statement to the Appendix. These assumptions include Assumptions 1, 2 with (20) replacing (2), 3 with \( \alpha \in [0.25, 0.5] \), and the analogues to Assumptions 4 (with \( x_t \) replaced by \( F_t(\beta_0) = \partial f_t(\beta)/\partial \beta |_{\beta=\beta_0} \)) and 5 (with \( h_t \) replaced by \( u_t F_t(\beta_0) \)). We note that these assumptions cover a range of models such as smooth transition autoregressive (STAR) and nonlinear ARCH.

Then, defining \( \xi_T \) and \( \xi_{i,T} \), \( i = 1, 2, 3 \), as in (8)-(11) with the nonlinear regression function \( f(\cdot, \cdot) \) replacing its linear counterpart, we have the following theorem.

**Theorem 2** Let \( y_t \) be generated by (20) and the following Assumptions hold: 1, 3 with (20) replacing 2, 3 for \( \alpha \in [0.25, 0.5] \), 2(i)-(iv)-(v) and A.1-A.4 (in the appendix). Then \( AE[\xi_T] \) and \( AE[\xi_{i,T}] \) \( (i = 1, 2, 3) \) equal the respective values given in Theorem 1.

**Remark 3:** Theorem 2 reveals that \( AE[\xi_T] \) does not depend on the form of \( f(\cdot, \cdot) \), beyond that embodied in the assumptions. Consequently, Remarks 1 and 2 continue to apply in the nonlinear context.

In the next section, we consider analogous results for linear models with endogenous regressors estimated via 2SLS.

### 3 Two Stage Least Squares RSS

Now we consider the case in which the equation of interest is a structural relationship from a simultaneous system, with this equation exhibiting \( m \) breaks, such that

\[
y_t = x_t' \beta_{x,t}^0 + z_{1,t}' \beta_{z_{1,t}}^0 + u_t, \quad i = 1, ..., m + 1, \quad t = T_{i-1}^0 + 1, ..., T_i^0,
\]

where \( T_0^0 = 0 \) and \( T_{m+1}^0 = T \), where \( T \) is the total sample size. Thus, \( y_t \) is the dependent variable, while \( x_t \) is a \( p_1 \times 1 \) vector of endogenous explanatory variables, \( z_{1,t} \) is a \( p_2 \times 1 \) vector of exogenous variables including the intercept, and \( u_t \) is a mean zero error. We define \( p = p_1 + p_2 \).

As in the previous section, we assume the location and magnitude of the breaks are governed by Assumptions 1 and 3 respectively.

As (21) is a structural equation, the endogenous explanatory variables, \( x_t \), are (in general) correlated with the errors, \( u_t \), and so 2SLS requires a reduced form representation to be estimated using appropriate instruments. However, the reduced form coefficients may themselves be subject to breaks, as discussed in the first subsection below, before we focus attention on (21).
3.1 Reduced form model

The reduced form model is

\[ x_t' = z_t' \Delta_0^{(i)} + v_t', \quad i = 1, 2, \ldots, h + 1, \quad t = T_{i-1}^\dagger + 1, \ldots, T_i^\dagger, \]  

(22)

where \( T_0^\dagger = 0 \) and \( T_{h+1}^\dagger = T \). The vector \( z_t = (z_{1,t}', z_{2,t}', \ldots, z_{h+1,t}')' \) is \( q \times 1 \) and contains variables that are uncorrelated with both \( u_t \) and \( v_t \) and are appropriate instruments for \( x_t \) in the first stage of the 2SLS estimation. The parameter matrices \( \Delta_0^{(i)} \) are each of \( q \times p_1 \). In line with Section 2, the number of reduced form breaks, \( h \), is assumed known, with the break points \( \{T_i^\dagger\} \) generated as follows.

Assumption 7 \( T_i^\dagger = [T_{\pi_0}^0] \), where \( 0 < \pi_0^1 < \ldots < \pi_0^h < 1 \).

Note that the break fractions in the reduced form, \( \pi^0 = [\pi_0^1, \pi_0^2, \ldots, \pi_0^h]' \), may or may not coincide with the breaks in the structural equation, \( \lambda^0 = [\lambda_0^0, \lambda_0^0, \ldots, \lambda_0^m]' \). Analogously to the structural form Assumption 3, we assume the breaks in the reduced form are shrinking.

Assumption 8 \( \Delta_0^{j+1} - \Delta_0^j = A_0^{j} r_T \) where \( r_T = T^{-\alpha_r} \), for \( \alpha_r \in (0, 0.5) \) and \( j = 1, \ldots, h \).

The reduced form of (22) can be re-written as

\[ x_t(\pi^0)' = \tilde{z}_t(\pi^0)' \Theta_0 + v_t', \quad t = 1, 2, \ldots, T \]  

(23)

where \( \Theta_0 = [\Delta_0^{(1)'}, \Delta_0^{(2)'}, \ldots, \Delta_0^{(h+1)'}, \tilde{z}_t(\pi^0) = \iota(t, T) \otimes z_t, \iota(t, T) \) is a \( (h + 1) \times 1 \) vector with first element \( I\{t/T \in (0, \pi_0^1]\} \), \( h+1^{th} \) element \( I\{t/T \in (\pi_0^h, 1]\} \), \( k^{th} \) element \( I\{t/T \in (\pi_0^{k-1}, \pi_0^k]\} \) for \( k = 1, 2, \ldots, h \) and \( I\{\cdot\} \) is an indicator variable that takes the value one if the event in the curly brackets occurs.

Let \( \hat{\pi} = [\hat{\pi}_1, \hat{\pi}_2, \ldots, \hat{\pi}_h]' \) denote estimators of \( \pi^0 \). These estimators are not our prime concern and it is assumed they satisfy the following condition.

Assumption 9 \( \hat{\pi} = \pi^0 + O_p(T^{-1-2\alpha_r}) \) for some \( \alpha_r \in (0, 0.5) \).

This condition would be satisfied if, for example, the break dates in the reduced form are estimated by OLS equation by equation and the estimates of the break fractions are then pooled; see Bai and Perron (1998)[Proposition 5] and Bai (1997)[Proposition 1]. Notice that under our
assumptions $1 - 2\alpha_r > 0$ and $\hat{\pi}$ is consistent for $\pi^0$. Let $\hat{x}'_t$ denote the resulting fitted values, that is,
\[
\hat{x}'_t = \hat{z}_t(\hat{\pi}')(\hat{\Theta}(\hat{\pi}))^{-1} \sum_{t=1}^{T} \bar{z}_t(\hat{\pi})x'_t,
\]
where $\bar{z}_t(\hat{\pi})$ is defined analogously to $\bar{z}_t(\pi^0)$ based on the estimator of the true break points in the reduced form.

In the special case when the reduced form is stable, (22) is replaced by a model with a single regime ($h = 0$), while Assumptions 7 and 8 are redundant. Obviously, (24) then becomes the corresponding OLS expression for $\hat{x}'_t$.

### 3.2 Structural form RSS

For estimation of (21), the statistic of interest is the minimized residual sum of squares from the second stage estimation. We therefore now suppose that a researcher knows the number of the breaks in (21) but not their locations. As in the previous section, we use $\lambda$ to denote an arbitrary set of $m$ break fractions in the model of interest. The second stage of 2SLS can begin with the estimation via OLS of
\[
y_t = \hat{x}'_i \beta_{x,i}^* + z'_1 \beta_{1,i}^* + u^*_i, \quad i = 1, \ldots, m+1, \quad t = T_{i-1} + 1, \ldots, T_i,
\]
for each possible unique $(m+1)$-partition of the sample, where $T_i = [\lambda, T]$ and $u^*_i$ is an error term. Letting $\beta_i' = (\beta_{x,i}', \beta_{1,i}')'$, for a given partition, the estimates of $\beta^* = (\beta_1^*, \beta_2^*, \ldots, \beta_{m+1}^*)'$ are obtained by minimizing the sum of squared residuals
\[
S_T(T_1, \ldots, T_m; \beta) = \sum_{i=1}^{m+1} \sum_{t=T_{i-1}+1}^{T_i} (y_t - \hat{x}'_i \beta_{x,i} - z'_1 \beta_{1,i})^2
\]
with respect to $\beta = (\beta_1', \beta_2', \ldots, \beta_{m+1}')$. As in the OLS case, we denote these estimators by $\hat{\beta}([T_i]_{i=1}^m)$, while the estimators of the break points are
\[
(\hat{T}_1, \ldots, \hat{T}_m) = \arg \min_{T_1, \ldots, T_m} S_T(T_1, \ldots, T_m; \hat{\beta}([T_i]_{i=1}^m))
\]
where the minimization is taken over all possible partitions, $(T_1, \ldots, T_m)$. The 2SLS estimates $\hat{\beta}([\hat{T}_i]_{i=1}^m) = (\hat{\beta}_1', \hat{\beta}_2', \ldots, \hat{\beta}_{m+1}')$ are then the regression parameter estimates associated with the
estimated partitions. As for the OLS case, the estimated break fractions are denoted \( \hat{\lambda} \), the \( m \times 1 \) vector with \( j^{th} \) element \( \hat{T}_j/T \).

Given the existence of breaks in both structural and reduced form equations, we modify the definition of admissible partitions over which the minimization in (27) is achieved.

**Assumption 10** Equation (27) is estimated over all partitions \((T_1, ..., T_m)\) such that \( T_i - T_{i-1} > \max\{q - 1, \epsilon T\} \) for some \( \epsilon > 0 \) and \( \epsilon < \inf f_j(\lambda^0_{i+1} - \lambda^0_i) \) and \( \epsilon < \inf f_j(\pi^0_{j+1} - \pi^0_j) \).

However, when the reduced form is stable, the partitions are required only to satisfy Assumption 2 as in the OLS case. The following assumptions also involve redefining some notation used in Section 2.

Defining \( h_{1,t} = (u_t, v_t')' \) and \( h_{i,t} \) to be the \( i^{th} \) element of \( h_t = h_{1,t} \otimes z_t \) we impose the following additional assumptions.

**Assumption 11** (i) \( E[h_{1,t} | F_t] = 0 \) where \( F_t \) is the \( \sigma \)-algebra generated by \( \{z_t, h_{1,t-1}, z_{t-1}, h_{1,t-2}, \ldots \} \); (ii) \( E[\|h_{i,t}\|_d] < H_d < \infty \) for \( t = 1, 2, \ldots \) and some \( d > 2 \); (iii) \( V_{T,i}(r) = Var[T^{-1/2} \sum_{t=T_{i-1}^{l+1}}^{T_l} h_t] \) is uniformly positive definite for all \( T \) sufficiently large and \( \lim_{T \to \infty} V_{T,i}(r) = r V_i \), uniformly in \( r \in (0, \lambda^0_i - \lambda^0_{i-1}) \) where \( V_i \) is a positive definite matrix of constants; (iv) \( Var[h_{1,t} | F_t t/T \in [\lambda^0_{i-1}, \lambda^0_i]] = \Omega_i \), where \( \Omega_i \) is the \((p_1 + 1) \times (p_1 + 1)\) positive definite matrix of constants given by

\[
\Omega_i = \begin{bmatrix} \sigma^2_i & \gamma_i' \\ \gamma_i & \Sigma_i \end{bmatrix},
\]

with \( \sigma^2_i \) a scalar; (v) \( \Omega_i = \Omega, i = 1, 2, \ldots m. \)

**Assumption 12** \( \text{rank}\{T^0_i\} = p \) where \( T^0_i = \begin{bmatrix} A_0^{(i)} & \Pi \end{bmatrix} \), for \( i = 1, 2, \ldots, h + 1 \) where \( \Pi = [I_{p_2}, 0_{p_2 \times (q-p_2)}] \), \( I_a \) denotes the \( a \times a \) identity matrix and \( 0_{a \times b} \) is the \( a \times b \) null matrix.

**Assumption 13** There exists an \( l_0 > 0 \) such that for all \( l > l_0 \), the minimum eigenvalues of \( A_{il} = (1/l) \sum_{t=T_{i-1}^{l+1}}^{T_l} z_t z_t' \) and of \( \tilde{A}_{il} = (1/l) \sum_{t=T_{i}^{l+1}}^{T_l} z_t z_t' \) are bounded away from zero for all \( i = 1, \ldots, m + 1 \).

**Assumption 14** (i) \( T^{-1} \sum_{t=T_{i-1}^{l+1}}^{T_l} z_t z_t' \to r Q_{ZZ}(i) \) uniformly in \( r \in (0, \lambda^0_i - \lambda^0_{i-1}) \), where \( Q_{ZZ}(i) \) is a positive definite matrix for \( i = 1, 2, \ldots m + 1 \); (ii) \( Q_{ZZ}(i) = Q_{ZZ}, i = 1, 2, \ldots m + 1 . \)

\(^3\)We redefine some notation for ease of presentation; the definition is clear form the context.
Assumption 11 requires $h_{t,t}$ to be a conditionally homoscedastic martingale difference sequence, and imposes sufficient conditions to ensure $T^{-1/2} \sum_{t=1}^{[Tr]} h_t$ satisfies a Functional Central Limit Theorem within each regime (see White (2001)[Theorem 7.19]). It also contains the restrictions that the implicit population moment condition in 2SLS is valid - that is, $E[z_t u_t] = 0$ - and the conditional mean of the reduced form is correctly specified. Assumptions 11 and 14 combined imply: $V_i = V = \Omega \otimes Q_ZZ$. Assumptions 12 and 14, in conjunction with Assumption 11, imply the standard rank condition for identification in IV estimation in the linear regression model.\footnote{See e.g. Hall (2005)[p.35].} Note Assumption 12 implies $q \geq p$. Assumption 13 requires that there be enough observations near the true break points of the structural equation so that they can be identified.

To facilitate the analysis below, we introduce an alternative version of structural equation,

$$y_t = \bar{x}_t' \beta_{0,i} + z_{1,t}' \beta_{0,i} + \bar{u}_{t,i}, \quad (28)$$

where $\bar{x}_t = E[x_t | z_t]$ and hence

$$\bar{u}_{t,i} = u_t + v_t' \beta_{0,i}, \quad (29)$$

which is the composite disturbance that applies in (21) for regime $i$ when the endogenous $x_t$ are substituted by $E[x_t | z_t]$ from the reduced form. Therefore, (28) applies when the reduced form coefficients are known, with $\bar{x}_t = E[x_t | z_t]$ embodying the true reduced form regimes when those coefficients are subject to breaks. Also define

$$\bar{v}_{t,i} = (x_t - \bar{x}_t)' \beta_{0,i} = v_t' \beta_{0,i}. \quad (30)$$

Applying Assumption 3 to the structural form coefficient vector $\beta_{0,i} = (\beta_{0,i}^{0x}, \beta_{0,i}^{0z})'$, breaks in the structural form coefficients are of asymptotically negligible magnitude, with $\beta_{0,i}^{0x} \rightarrow \beta_{0}^{0x}$, say, for all $i = 1, \ldots, m + 1$. Under this assumption, then we have for all $i = 1, \ldots, m + 1$

$$\rho_{i}^2 = Var[\bar{u}_{t,i}] \rightarrow \rho^2 = \sigma^2 + 2\gamma' \beta_{0}^{0x} + \beta_{0}^{0y} \Sigma_{\beta_{0}^{0x}}, \quad (31)$$

$$\bar{\rho} = Cov[\bar{v}_{t,i}, \bar{u}_{t,i}] \rightarrow \bar{\rho} = \gamma' \beta_{0}^{0x} + \beta_{0}^{0y} \Sigma_{\beta_{0}^{0x}}, \quad (32)$$

$$\omega_{i}^2 = Var[\bar{v}_{t,i}] \rightarrow \omega^2 = \beta_{0}^{0y} \Sigma_{\beta_{0}^{0x}}. \quad (33)$$

With known reduced form coefficients, the quantity $\rho^2$ provides the asymptotic variance of the composite structural form disturbance $\bar{u}_{t,i}$ of (29) with shrinking coefficients. Therefore, $T \rho^2$
plays an analogous role in our analysis of the residual sum of squares for 2SLS as does \( T\sigma^2 \) for the OLS case in (5).

Denoting the minimized \( S_T \left( \hat{T}_1, ..., \hat{T}_m; \hat{\beta}(\{\hat{\mathbf{T}}_i\}_{i=1}^m) \right) \) in (27) as \( \text{RSS}(\hat{T}_1, \hat{T}_2, ..., \hat{T}_m) \), we consider \( AE[\xi_T] \) where, analogous to (5),

\[
\xi_T = \text{RSS}(\hat{T}_1, \hat{T}_2, ..., \hat{T}_m) - T\rho^2
\]  

(34)

in which \( AE[\cdot] \) again denotes the asymptotic expectation operator. Hence \( \xi_T \) defined by (34) is the difference between the residual sum of squares in the second step of 2SLS and the expected error sum of squares in (28).

Generalizing the approach of Section 2 to the 2SLS case requires the role of the reduced form to be recognized and we now decompose \( \xi_T \) into four components,

\[
\xi_T = \sum_{j=1}^{4} \xi_{j,T}.
\]

The first component,

\[
\xi_{1,T} = \text{RSS}(\hat{T}_1, \hat{T}_2, ..., \hat{T}_m; \hat{\pi}) - \text{RSS}(T^0_1, T^0_2, ..., T^0_m; \pi^0),
\]

(35)

represents the effect on the second stage residual sums of squares from estimating the coefficients of (21) within each partition using the estimated rather than the true break dates in both the structural equation and (if relevant) the reduced form. Both elements of (35) are obtained using \( \hat{x}_t \) from (24). The second component is defined as

\[
\xi_{2,T} = \text{ESS}(T^0_1, T^0_2, ..., T^0_m) - \text{ESS}(T^0_1, T^0_2, ..., T^0_m),
\]

(36)

where \( \text{ESS}(T^0_1, T^0_2, ..., T^0_m) \) is the error sum of squares for (21) evaluated using the true \( \{\beta^0_i\}_{i=1}^{m+1} \) in conjunction with \( \hat{x}_t \). Hence \( \xi_{2,T} \) is the impact on the residual sum of squares from estimating the coefficients of (25) using 2SLS with known (true) break dates and evaluated using the first stage \( \hat{x}_t \). The third component is given by

\[
\xi_{3,T} = \text{ESS}(T^0_1, T^0_2, ..., T^0_m) - \text{ESS}^e(T^0_1, T^0_2, ..., T^0_m),
\]

(37)

where \( \text{ESS}^e(T^0_1, T^0_2, ..., T^0_m) \) is the error sum of squares evaluated using the true \( \{\beta^0_i\}_{i=1}^{m+1} \) in conjunction with \( \hat{x}_t = E[x_t | z_t] \) from the reduced form. Consequently \( \xi_{3,T} \) is the effect from using \( \hat{x}_t \) rather than \( \pi_t \) for computation of the structural equation error sums of squares. The
Theorem 3 indicates that

\[ \xi_{4,T} = ESS^e(T_1^0, T_2^0, ..., T_m^0) - T \rho^2, \]  

and hence captures the effects of the composite \( \pi_{t,i} \) in the structural equation of \( \xi_t \), under the shrinking breaks Assumption 3.

Theorem 3 then generalizes the result of Theorem 1 to the 2SLS case, employing the notation

\[ \delta \lambda_i^0 = \begin{cases} \lambda_i^0 & \text{for } i = 1, \\ \lambda_i^0 - \lambda_{i-1}^0 & \text{for } i = 2, ..., m, \\ 1 - \lambda_m^0 & \text{for } i = m + 1, \end{cases} \]  

with \( \delta \pi_i^0 \) defined analogously for the true reduced form regime fractions.

**Theorem 3** Let \( y_t \) be generated by (21), \( x_t \) be generated by (22), and \( \tilde{x}_t \) be given by (24).

Let Assumptions 1, 3, 7 - 14 hold. Then we have: (i) \( AE[\xi_{1,T}] = -3mp^2 \); (ii) \( AE[\xi_{2,T}] = -p(m+1)\rho^2 + p(\rho^2 - \sigma^2) \sum_{i=1}^{m+1} d_i / (\delta \lambda_i^0) \); (iii) \( AE[\xi_{3,T}] = -q(h+1)(\rho^2 - \sigma^2) \); (iv) \( AE[\xi_{4,T}] = 0 \);

and so

\[ AE[\xi_T] = -[(p + 3)m + p(\rho^2 - \sigma^2) \left( q(h+1) - p \sum_{i=1}^{m+1} d_i / (\delta \lambda_i^0) \right)], \]

where

\[ 0 < \sum_{i=1}^{m+1} d_i / (\delta \lambda_i^0) \leq \min[(h + 1), (m + 1)] \]

in which \( d_i \) is defined as follows: if there are no reduced form breaks between \( \lambda_{i-1}^0 \) and \( \lambda_i^0 \), and so \( \pi_k \leq \lambda_{i-1}^0 < \lambda_i^0 \leq \pi_{k+1} \), say, then \( d_i = (\delta \lambda_i^0)^2 / (\delta \pi_{k+1}^0) \); if there are reduced form breaks between \( \lambda_{i-1}^0 \) and \( \lambda_i^0 \) and so \( \pi_k \leq \lambda_{i-1}^0 < \pi_{k+1} < ... < \pi_{k+\ell_i} < \lambda_i^0 \leq \pi_{k+\ell_i+1} \), say, then

\[ d_i = \frac{(\pi_{k+\ell_i+1} - \lambda_i^0)^2}{\delta \pi_{k+\ell_i+1}^0} + \frac{(\lambda_i^0 - \pi_{k+\ell_i+1})^2}{\delta \pi_{k+\ell_i+1}^0} + \pi_{k+\ell_i} - \pi_{k+1}. \]

**Remark 4:** Theorem 3 indicates that \( AE[\xi_T] \) depends on: the number of structural form breaks, \( m \), the number of mean parameters in each regime, \( p \), the number of instruments, \( q \), the co-variance structure of the composite error \( \pi_{t,i} \) through \( (\rho^2 - \sigma^2) = 2(\gamma')^2 + \beta_x^0 \bar{\gamma}_x^0 \), and also on the relative locations of the structural and reduced form breaks. The results are nevertheless compatible with \( T^{-1} \text{RSS}(T_1, ..., T_m) \sim T \rho^2 \), the variance of the composite error in (28).

**Remark 5:** The expression for \( AE[\xi_{1,T}] \) carries over from Theorems 1 and 2 and so the effect of estimating the residual sum of squares of interest is asymptotically the same irrespective of
whether the model is a linear or nonlinear equation with exogenous regressors or a linear equation with endogenous regressors and consistently estimated reduced form break dates. We also note that Lemma 2 underlies this result in all cases.

Remark 6: Theorem 3(i) does not require Assumption 14(ii), and so \( AE[\xi_{1,T}] \) has the stated form even if the instrument cross product matrix exhibits the regime specific behaviour delineated in part (i) of that assumption.

The asymptotic residual sum of squares in the second stage regression for the important special case of a stable reduced form is stated as a Corollary to Theorem 3:

**Corollary 1** Let \( y_t \) be generated by (21), with \( x_t \) generated by (22) and \( \hat{x}_t \) be given by (24), both with with \( h = 0 \). Let Assumptions 1-3, 9, 11 and 12-14 hold. Then we have: (i) \( AE[\xi_{1,T}] = -3m\rho^2 \); (ii) \( AE[\xi_{2,T}] = -p(m\rho^2 + \sigma^2) \); (iii) \( AE[\xi_{3,T}] = -q(\rho^2 - \sigma^2) \); (iv) \( AE[\xi_{4,T}] = 0 \); and so

\[
AE[\xi_T] = -[(p + 3)m + p]\rho^2 - (q - p)(\rho^2 - \sigma^2).
\]

Remark 7: With a stable reduced form, the term \( AE[\xi_{2,T}] \) that captures the impact of estimation of the mean parameters in Corollary 1 can be written as \(-p\{(m + 1)\rho^2 - (\rho^2 - \sigma^2)\}\), and contributes a factor of \(-(m + 1)p\) (ignoring the term that is independent of \( m \)). Combined with \( AE[\xi_{1,T}] = -3m\rho^2 \), the comment in Remark 2 about the differing impacts of the break-fraction and regression parameters in models with exogenous regressors applies equally in models with endogenous regressors estimated via 2SLS with stable reduced forms. When the reduced form is unstable, however, this result is modified by the second term of Theorem 3(ii).

Remark 8: Corollary 1 also clarifies the role of the reduced form in minimization of the 2SLS residual sum of squares in models with no breaks. When conventional 2SLS is applied to a stable structural form \( (m = 0) \), (34) becomes \( \xi_T = \text{RSS} - T\rho^2 \) with

\[
AE[\xi_T] = -p\rho^2 - (q - p)(\rho^2 - \sigma^2).
\]

The result shows that the downward bias in the minimized 2SLS residual sum of squares compared with \( E[u_t^2] \) depends not only on the number of structural form coefficients estimated, namely \( p \), but also on the extent of overidentification \( (q - p) \) and the additional asymptotic variation induced in the structural form by the use of IV estimation, namely \( E[u_t^2 - u_t^2] = (\rho^2 - \sigma^2) \).

In this context where both the reduced forms and structural forms are stable, Pesaran and
Smith (1994) propose a generalized $R^2$ criterion computed from the second stage regression, analogous to (25), and (40) makes clear that the value of this criterion will asymptotically depend on characteristics of the reduced form (including the number of instruments) as well as the goodness-of-fit of the structural form equation itself.

**Remark 9:** Two further special cases of Theorem 3 are of interest; in both only the numbers of breaks matter, not their locations *per se*. Firstly, when all reduced form breaks coincide with structural form breaks, but with possible additional structural form breaks, then
\[ \sum_{i=1}^{m+1} \frac{d_i}{\delta \lambda_0^i} = h + 1 \] (see the proof of Theorem 3 in the Appendix). In this case,

\[ AE[\xi_T] = -[(p + 3)m + p] \rho^2 - (h + 1)(\rho^2 - \sigma^2)(q - p). \quad (41) \]

This expression has a similar interpretation to that drawn out in Remark 7, with the first term of (41) giving the bias due to estimation of the structural form coefficients and break dates, while the second shows the roles of the additional asymptotic variation from using IV, $\rho^2 - \sigma^2$, and the extent of overidentification $(q - p)$, with the number of reduced form regimes $(h + 1)$ now magnifying the latter effects. Secondly, when all structural form breaks coincide with the dates of reduced form breaks, with possible additional reduced form breaks, then
\[ \sum_{i=1}^{m} \frac{d_i}{\delta \lambda_0^i} = m + 1 \] (as again seen from the Appendix proof) and

\[ AE[\xi_T] = -[(p + 3)m + p] \rho^2 - (\rho^2 - \sigma^2) [q(h + 1) - (m + 1)]. \quad (42) \]

This has a similar interpretation to that just discussed for (41), although overidentification in the second term of (42) appears in the form of a comparison of the total numbers of reduced and structural form coefficients estimated.

**Remark 10:** For the general case where reduced and structural form break dates do not necessarily coincide, then the theorem shows that although $AE[\xi_T]$ can depend on the relative locations of structural and reduced form break points, the extent of this dependence is bounded. Based on the interpretation of (41) and (42) in Remark 8, the quantity $q(h+1)-p \sum_{i=1}^{m+1} \frac{d_i}{\delta \lambda_0^i}$ might be interpreted more generally as a measure of the extent of overidentification of the structural form parameters in the presence of breaks. From the bound for the summation term given in the theorem, the maximum extent of overidentification occurs when $\sum_{i=1}^{m+1} \frac{d_i}{\delta \lambda_0^i} = \min[(h + 1), (m + 1)]$, and hence the breaks in either the reduced or structural form coincide with those in the other (with possible additional breaks).
4 Testing Break Dates

It has been noted above that $\text{AE}[\xi_1,T]$ exhibits similar behaviour in all the models considered, and this stems from the fact that the large sample behaviour of $\xi_1,T$ is governed by a version of (13) in each case. In this section, we exploit this structure to propose a method for testing

$$H_0 : \lambda_i^0 = \bar{\lambda}_i \text{ for } i = 1, \ldots, m,$$

(43)

for $0 < \bar{\lambda}_1 < \ldots < \bar{\lambda}_m < 1$, against the alternative hypothesis that at least one $\lambda_i^0 \neq \bar{\lambda}_i$ ($i = 1, \ldots, m$). Given this common underlying structure, we consider explicitly the OLS case and then note how the result extends to the other models.

Therefore, in the OLS framework (Section 2.1), consider the following statistic

$$N_\lambda(\bar{\lambda}) = \text{RSS}(T_1, \ldots, T_m) - \min_{T_1, \ldots, T_m} \text{RSS}(T_1, \ldots, T_m)$$

(44)

where $T_i = [\lambda_i, \bar{\lambda}]$ and $\bar{\lambda} = (\bar{\lambda}_1, \bar{\lambda}_2, \ldots, \bar{\lambda}_m)$. The following theorem gives the limiting distribution of $N_\lambda(\bar{\lambda})$.

**Theorem 4** Let $y_t$ be generated by (1) with $H_0$ of (43) true and Assumptions [1,6] (i)-(iii), and [6] hold. Then, for the statistic (44),

$$N_\lambda(\bar{\lambda}) \xrightarrow{d} \sum_{i=1}^{m} b_i$$

where $\{b_i\}_{i=1}^{m}$ are mutually independent and $b_i \sim B(\mu_{i,1}, \mu_{i,2})$ with $\mu_{i,j} = 0.5a_{i,j}/c_{i,j}$ for $j = 1, 2$, $a_{i,j}$ and $c_{i,j}$ defined in (14) and (15) respectively, and $B(\mu_1, \mu_2)$ as in Definition 1. In addition, if Assumption [7] (iv) holds then $\mu_{i,j} = 0.5\sigma^2_{i,j-1}$; and if Assumption [7] (v) also holds then $\mu_{i,j} = 0.5\sigma^2$.

**Remark 11:** The limiting distributions in Theorem 4 depend on model parameters. However, asymptotically valid inference can be performed by simulating the null distribution using consistent estimators of $\mu_{i,j}$ under $H_0$ and then comparing $N_\lambda(\bar{\lambda})$ to the appropriate percentile of this simulated distribution. A consistent estimator for $\mu_{i,j}$ is given by

$$\hat{\mu}_{i,j} = \frac{\hat{\theta}_i \hat{Q}_{i,j-1} \hat{\theta}_j}{2\hat{\theta}_i \hat{V}_{i,j-1} \hat{\theta}_j}$$

(45)

where $\hat{\theta}_i = \hat{\beta}_{i+1} - \hat{\beta}_i$, $\hat{\beta}_i = \hat{\beta}_i([\hat{T}_i]_{i=1}^{m})$ (defined in Section 2.1), $\hat{Q}_t = (\hat{T}_t - \hat{T}_{t-1})^{-1} \sum_{i=\hat{T}_{t-1}+1}^{\hat{T}_t} x_t x'_t$, $\hat{V}_t = (\hat{T}_t - \hat{T}_{t-1})^{-1} \sum_{i=\hat{T}_{t-1}+1}^{\hat{T}_t} \hat{u}_{t,i} x_t x'_t$, $\hat{u}_{t,i} = y_t - x'_t \hat{\beta}_t$. If Assumption [7] (iv) holds then an alternative consistent estimator is

$$\hat{\mu}_{i,j} = 0.5\hat{\sigma}^2_{i,j-1}$$

(46)
where \( \hat{\sigma}^2 = (\hat{T}_\ell - \hat{T}_{\ell-1})^{-1} \sum_{t=\hat{T}_{\ell-1}+1}^{\hat{T}_{\ell}} \hat{u}_{t,\ell}^2 \); and if Assumption 5(v) holds then a additional consistent estimator is

\[
\hat{\mu}_{i,j} = 0.5\hat{\sigma}^2 \tag{47}
\]

where \( \hat{\sigma}^2 = T^{-1} \sum_{\ell=1}^{m+1} \sum_{t=\hat{T}_{\ell-1}+1}^{\hat{T}_{\ell}} \hat{u}_{t,\ell}^2 \).

Remark 12: If Assumption 5(iv)-(v) holds as well then it is possible to normalize the statistic to remove the nuisance parameter from the limiting distribution. To this end consider the \( F \)-type test statistic

\[
F_\lambda(\lambda) = \frac{RSS(T_1, \ldots, T_m) - \min_{T_1, \ldots, T_m} RSS(T_1, \ldots, T_m)}{\hat{\sigma}^2} \
\]

where \( T_i = [\overline{X}_i, T] \). Then we have the following corollary to Theorem 4:

**Corollary 2** Under the conditions of Theorem 4 (including Assumption 3(iv)-(v)), we have

\[
F_\lambda(\lambda) \overset{d}{\rightarrow} \sum_{i=1}^{m} b_i \text{ where } \{b_i\}_{i=1}^{m} \text{ are mutually independent and } b_i \sim \mathcal{B}(0.5, 0.5).
\]

Percentiles of this limiting distribution are presented in Table 1. The distribution was simulating in MATLAB using 10 million iterations.

Remark 13: Theorem 4 extends to the nonlinear regression models that satisfy Assumptions 1, 2 with (20) replacing (2), (3) for \( \alpha \in [0.25, 0.5) \), (5(i)), and A.1-A.4 (in the appendix), with \( a_{i,j}, c_{i,j} \) given by (62), (63) respectively. Further, the imposition of Assumption 5 parts (iv) or (iv)-(v) yield the same specializations of \( \mu_{i,j} \) as described in Theorem 4. A consistent estimators of \( \mu_{i,j} \) for use in simulation of the limiting distribution is given by (45) but with the following changes: \( \hat{\beta}_i \) now denotes the NLS estimator of the parameter in (estimated) regime \( i \); in \( \hat{Q}_t \) and \( \hat{V}_t \), \( x_t \) is replaced by \( F_t(\hat{\beta}_t) \). If Assumption 5(iv) holds then an alternative consistent estimator is given by (46) but with \( \hat{u}_{t,\ell} \) being the NLS residual; if Assumption 5(v) holds then an alternative consistent estimator is given by (47) with the same redefinition of the residual. Similarly, we can define an analogous version of \( F_\lambda(\lambda) \) for this model which has the limiting distribution given in Corollary 2 under all the assumptions mentioned in this remark.

Remark 14: Theorem 4 extends to the 2SLS models that satisfy Assumptions 1, 3, 7-11(i)-(iii), 12-14 with the forms of \( a_{i,j}, c_{i,j} \) implied, as appropriate, by either (67)-(68) or (71)-(74). The construction of a consistent estimator of \( \mu_{i,j} \) for use in simulation of the limiting distribution depends on the location of the \( i^{th} \) break in the structural equation relative to the reduced form.

\footnote{The degrees of freedom correction suggested by Theorem 1 can also be applied in the denominator of \( \hat{\sigma}^2 \).}
suitable normalization. To this end, we define

\[
\hat{\mu}_{i,j} = \frac{\hat{\beta}' \hat{Y}_{i,j} \hat{Q}_{ZZ}(i+j-1) \hat{Y}_k \hat{\theta}_i}{2\hat{\theta}' \hat{Y}_{k,j-1} \hat{Q}(i+j-1) \hat{Y}_k \hat{\theta}_i}
\]

for \( j = 1, 2 \), where \( \hat{\theta}_i = \hat{\beta}_{i+1} - \hat{\beta}_i \), \( \hat{\beta}_i = (\hat{\beta}_{x,i}, \hat{\beta}'_{z_{1,i}})' \) are the 2SLS estimators of the mean parameters in the estimated \( i^{th} \) regime (as defined in Section 3.2), \( \hat{Y}_k = [\Delta_k, \Pi] \) where \( \Delta_k \) are the OLS estimators of the reduced form parameters in the \( k^{th} \) estimated (reduced form) regime,

\[
\hat{Q}_{ZZ}(\ell) = (\hat{T}_k - \hat{T}_{k-1})^{-1} \sum_{t=T_{k-1}+1}^{T_k} z_t z'_t, \quad \hat{\Phi}(\ell) = \hat{C}_t \hat{C}'_t, \quad \hat{\varphi}_t = [1, \hat{\beta}'_x, \hat{\nu}_t, \hat{\varphi}_t] = \begin{bmatrix} 1, \hat{\beta}'_x, \hat{\beta}'_z \end{bmatrix}, \quad \hat{V}_t = (\hat{T}_k - \hat{T}_{k-1})^{-1} \sum_{t=T_{k-1}+1}^{T_k} \hat{h}_t \hat{h}'_t, \quad \hat{v}_t = \hat{h}_t \hat{y}_t \quad \forall t \in \{ \hat{\pi}_{k-1}, \hat{\pi}_k \}.
\]

If \( \hat{\pi}_{k-1} = \hat{\lambda}_i \) for some \( k \) then a consistent estimator of \( \mu_{i,j} \) is given by

\[
\hat{\mu}_{i,j} = 0.5\hat{\rho}_{i,j-1}^2
\]

and all other definition remain the same. Irregardless, if in addition Assumption 11(iv) holds then an alternative consistent estimator for \( \mu_{i,j} \) is:

\[
\hat{\mu}_{i,j} = 0.5\hat{\rho}^2
\]

where \( \hat{\rho}^2 = (\hat{T}_k - \hat{T}_{k-1})^{-1} \sum_{t=T_{k-1}+1}^{T_k} \hat{v}_t \{ \hat{h}_t \hat{h}'_t + \hat{v}_t \} \hat{v}'_t \); and if Assumptions 11(iv)-(v) hold then an alternative consistent estimator for \( \mu_{i,j} \) is:

\[
\hat{\mu}_{i,j} = 0.5\hat{\rho}^2
\]

where \( \hat{\rho}^2 = T^{-1} \sum_{t=1}^{T} \hat{v}_t \{ \hat{h}_t \hat{h}'_t + \hat{v}_t \} \hat{v}'_t \). In this last case (i.e. Assumption 11(iv)-(v) hold) then the dependence of the limiting distribution on model parameters can be removed via suitable normalization. To this end, we define

\[
F_{\lambda}^{2SLS}(\bar{X}) = \frac{RSS(T_1, ..., T_m) - \min_{T_1, ..., T_m} RSS(T_1, ..., T_m)}{\hat{\rho}^2}
\]

Under the assumptions listed in this remark and the \( H_0 \) of (43), \( F_{\lambda}^{2SLS}(\bar{X}) \) converges to the limiting distribution in Corollary 2.

Remark 15: The statistics above can be used to generate confidence sets for the break fractions. To illustrate, again consider OLS. An approximate 100(1 - \( \alpha \))% confidence set for the break fractions is given by:

\[
\left\{ \bar{X} \ s.t \ N_\lambda(\bar{X}) < q_{m,1-\alpha} \right\}
\]
where \( N_\lambda(\bar{\lambda}) \) is defined in (44) and \( q_{m,1-\alpha} \) is the 100(1 - \( \alpha \))\textsuperscript{th} quantile of \( \sum_{i=1}^{m} \hat{b}_i \) defined in Theorem 4.

Remark 16: Under similar assumptions, Yao (1987) and Bai (1997) obtain the marginal distribution of a single break fraction estimator, which is used by Bai (1997) and also Bai and Perron (1998) to construct a confidence interval for the date of each break. Since the \( m \) break date distributions are asymptotically independent, a joint test of the null hypothesis (43) could be deduced from these. In contrast, (48) compares RSS at the hypothesised break dates with the overall minimized RSS, providing a natural test statistic in the least squares contexts considered here. In common with the confidence interval approach of Elliott and Muller (2007), but not that of Yao (1987), the confidence sets in (54) do not imply the set of dates included need be contiguous.

5 Simulation results

In this section, we report simulation results to illustrate two applications of our results in linear regression models with exogenous regressors: (i) estimation of the error variance; (ii) hypothesis testing about the break fraction.

(i) Estimation of the error variance

We consider the behaviour of three estimators of the error variance in the model defined in Section 2. These three estimators differ in degrees of freedom correction and take the generic form:

\[
\hat{\sigma}_k^2 = \frac{RSS(\hat{T}_1, \hat{T}_2, \ldots, \hat{T}_m)}{(T - m(p + 1) - c_k)}, \quad k = 1, 2, 3
\]

(55)

where RSS(\( \cdot \)) is defined in (6)-(7). The constant \( c_i \) is defined as follows: \( c_1 = 0 \), this choice makes no degrees of freedom correction to take account of the break estimation and is employed in Bai and Perron (1998); \( c_2 = m \), this choice treats break point and the mean parameters symmetrically in the degrees of freedom correction and is used by Yao (1988); \( c_3 = 3m \), this choice is suggested by Theorem 1. We investigate the finite sample performance of these three estimators in two ways: bias and the coverage probabilities of confidence intervals for the mean parameters based. For the model in Section 2, these confidence intervals take the generic form

\[
\hat{\beta}_{i,j} \pm z_{\alpha/2} \hat{\sigma}_k \sqrt{D_{i,j,T}},
\]

(56)
where \( \hat{\beta}_{i,j} \) is the \( j \)th element the OLS estimator of the mean parameters in the estimated regime \( i \), \( D_{i,j,T} \) is the \( j \)th main diagonal element of \( D_{i,T} = (R_i'R_i/r_{i,T})^{-1}, r_{i,T} = T_i - \bar{T}_{i-1} \), \( R_i \) is \( r_{i,T} \times p \) data matrix for the estimated \( i \)'th regime, with typical row \( x_t' \), and \( z_{a/2} \) is the 100(1 - \( a/2 \))th percentile of the standard normal distribution.

We consider the one break model with \( p = 2 \), that takes the form:

\[
y_t = \begin{cases} 
\mu_1 + w_t'\gamma_1 + u_t & \text{if } t \leq [0.5T] \\
\mu_2 + w_t'\gamma_2 + u_t & \text{if } t > [0.5T] 
\end{cases}
\]

where \( u_t \) is a sequence of \( i.i.d. \) \( N(0,1) \) random variables and \( w_t \) is a scalar \( i.i.d. \) \( N(1,1) \) random variable that are uncorrelated with \( u_t \). Thus, in terms of the notation in Section 2, \( x_t = [1, w_t']' \) and \( \beta_0^i = [\mu_i, \gamma_i]' \). Since Theorem 1 assumes shrinking breaks (Assumption 3), we fix \( \mu_2 = \gamma_2 = 1 \) and report results for \( \mu_1 = \gamma_2 = 1-(0.2 \times 360^\alpha/T^\alpha) \), for \( \alpha = 0.0, 0.1, 0.2, 0.3, 0.4, 0.49 \), and sample sizes \( T = 120, 240, 360, 480 \).

As in Section 2, the estimation is performed imposing the true number of breaks on the DGP. The break dates are estimated as defined in (4) except that in practice regimes are restricted to contain at least \( \lfloor \epsilon T \rfloor \) observations. The parameter \( \epsilon \) is specified by the researcher, and this is often referred to as the trimming parameter. We use \( \epsilon = 0.1 \) throughout this paper. An efficient search algorithm is discussed in some detail by Bai and Perron (2003), and is employed in our analysis below. Each DGP is replicated 5000 times and for each different parameter setting the same pseudo-random number generator seed is employed. All simulations are performed in MATLAB.

The results are reported in Table 2. For each parameter configuration, we report the absolute bias of the error variance estimator and the coverage probabilities of the confidence intervals that are closest to \( (C_{90}) \) and furthest away \( (F_{90}) \) from the nominal level of 90% where the comparison is over all parameters in all regimes. Thus, for given choice of \( k \), the coverage probability of the confidence intervals in (56) for any \( i,j \), \( cov_{i,j} \) say, satisfies

\[
|C_{90} - .9| \leq |cov_{i,j} - .9| \leq |F_{90} - .9|.
\]

It can be seen that, by the metrics reported and for this design, there is a clear ranking of the error variance estimators with \( k = 3 \) dominating \( k = 2 \) dominating \( k = 1 \). We note that the simulated coverage probability improves with \( T \) but is still below the nominal level by a couple of percent at the largest sample size. We attribute this to the estimation of break location and
the range of break magnitudes we are imposing.

(ii) Hypothesis testing about the break fraction

We now consider the performance of $F_\lambda(\bar{\lambda})$ in (48). Using the same design as above, we consider the case where it is desired to test $H_0 : \lambda_0 = 0.5 + \kappa$ for $\kappa = 0, 0.01, 0.02, \ldots, 0.1$. Since $\lambda_0 = 0.5$ in our design, $\kappa = 0$ corresponds to the case in which the null is true, and as $\kappa$ increases the distance between the hypothesized value and the truth increases. The calculated test statistic is compared to the critical value in Table 1 for a 5% significance level. The power curves are plotted in Figure 1 for the different values of $\alpha$. As would be expected, the power increases with $\kappa$ for each $T$ and $\alpha$, the power is inversely related to $\alpha$ for each $T$ and $\kappa$. Finally, we note the empirical size is close to the nominal level.

6 Concluding remarks

There has been a considerable literature on least squares-based estimation and testing in models with multiple discrete breaks in the parameters, see inter alia Bai and Perron (1998) and Hall, Han, and Boldea (2012) and Boldea and Hall (2013). In these contexts, if the model is assumed to have $m$ breaks then the break points (the points at which the parameters change) are estimated by minimizing the residual sum of squares over all possible data partitions involving $m$ breaks. A natural side-product of this estimation is this minimized residual sum of squares and this quantity plays an important role in subsequent inferences about the model. This paper, firstly, derives the asymptotic expectation of the residual sum of squares, the form of which indicates that the number of estimated break points and the number of regression parameters affect this expectation in different ways. Secondly, we propose a statistic for testing the joint hypothesis that the breaks occur at specified fixed break points in the sample. Under its null hypothesis, this statistic is shown to have a limiting distribution that is non-standard but simulatable, being a functional of independent random variables with exponential distributions whose parameters can be consistently estimated. In a special case, it is possible to normalize the statistic to make it pivotal and we provide percentiles for the associated limiting distribution. Our results cover the cases of either the linear or nonlinear regression model with exogenous regressors estimated...
via Ordinary (or Nonlinear) Least Squares which or a linear model in which some regressors are endogenous and the model is estimated via Two Stage Least Squares.

Appendix

Mathematical Appendix

Proof of Theorem 1

Part (i): From the principle of least squares,

\[ \xi_{1,T} = \min_{(T_1,...,T_m)} RSS(T_1,T_2,...,T_m) - RSS(T_1^0,T_2^0,...,T_m^0). \]  
(57)

From Bai and Perron (1998)[Proposition 4], for the limiting behaviour of \{\hat{T}_i\}_{i=1}^m we need to consider possible break dates \( T_i \) only within intervals close to each of the true breaks, as given by \( B = \bigcup_{i=1}^m B_i \) where \( B_i = \{ |T_i - T_i^0| \leq K_i s_{T_i}^{-2} \} \) for positive constants \( K_i, i = 1,...,m \). That is, for an individual break \( i \) we need to consider \( T_i = T_i^0 + [k_i s_{T_i}^{-2}] \) for \( k_i \epsilon [-K_i,K_i] \).

Then, it follows from (57) - see Bai (1997)[equations (8)-(9)] - that,

\[ \xi_{1,T} = \sum_{i=1}^m \min_{T_i} \{ A_i(T_i) + 2C_i(T_i) \} + o_p(1), \] uniformly in \( B \)  
(58)

where

\[ A_i(T_i) = \theta_{T_i}^0 \sum_{t=(T_i \wedge T_i^0)+1}^{T_i \vee T_i^0} x_t \theta_{T_i}^0 \]

\[ C_i(T_i) = (-1)^{I(T_i<T_i^0)} \theta_{T_i}^0 \sum_{t=(T_i \wedge T_i^0)+1}^{T_i \vee T_i^0} x_t u_t \]

for \( \theta_{T_i}^0 \) defined in Assumption 3 and \( a \vee b = \max\{a,b\}, a \wedge b = \min\{a,b\} \) and \( I(\cdot) \) is again an indicator function which takes the value unity when the condition in curly brackets is satisfied.

Equations (12) to (15) of the text then follow, using the same arguments as Bai (1997). The result

\[ AE[\xi_{1,T}] = -3m\sigma^2 \]  
(59)

\*Bai (1997) considers the case where \( m = 1 \) but his arguments extend to the multi-break case and lie behind the analysis in Bai and Perron (1998)[Section 3.3] in which they present the limiting distribution of the break fraction when \( m > 1 \).
then follows from Lemma 2, in combination with (12) and $B = \bigcup_{i=1}^{m} B_i$.

**Part (ii):** Using standard least squares algebra,

$$
\xi_{2,T} = \text{RSS}(T_1^0, \ldots, T_m^0) - \text{ESS}(T_1^0, \ldots, T_m^0)
$$

$$
= \sum_{i=1}^{m+1} \left( \sum_{t=T_{i-1}^0+1}^{T_i^0} (y_t - x_t'\hat{\beta}_i)^2 \right) - \sum_{i=1}^{m+1} \left( \sum_{t=T_{i-1}^0+1}^{T_i^0} (y_t - x_t'\beta_0^i)^2 \right)
$$

$$
= -\sum_{i=1}^{m+1} (\hat{\beta}_i - \beta_0^i)'(X_i'X_i)(\hat{\beta}_i - \beta_0^i)
$$

(60)

in which $X_i$ is the $(T_i^0 - T_{i-1}^0) \times p$ data matrix for the $i^{th}$ regime, with typical row $x_t'$, and the OLS estimates $\hat{\beta} = [\hat{\beta}_1', \hat{\beta}_2', \ldots, \hat{\beta}_{m+1}']$ are obtained imposing the correct break-points.

Under Assumption 4

$$
T^{-1}X_i'X_i \overset{\text{P}}{\rightarrow} (\delta\lambda_0^i)Q_i = M_i, \quad \text{say}.
$$

From Bai and Perron (1998) [Proposition 3], we have under our assumptions that

$$
\text{T}^{1/2} \begin{pmatrix} \hat{\beta} - \beta_0^i \end{pmatrix} \Rightarrow N(0, V_\beta)
$$

where $V_\beta = \sigma^2 \text{diag}[M_1^{-1}, M_2^{-1}, \ldots, M_{m+1}^{-1}]$. Therefore, it follows that

$$
-\xi_{2,T} \overset{\text{d}}{\rightarrow} \sum_{i=1}^{m+1} \kappa_i
$$

where $\kappa_i \sim \sigma^2 \chi_p^2$ and $\kappa_i, \kappa_j$ are independent for $i \neq j$. Consequently, $AE[\xi_{2,T}] = -p(m+1)\sigma^2$.

**Part (iii):** This follows directly from $E[u_t^2] = \sigma^2$. ◦

**Proof of Theorem 2**

We first state the assumptions listed in the Theorem not stated in the main text.

**Assumption A.1** Define $v_t$ as follows: if $x_t$ contains no lagged values of $y_t$ then $v_t = (x_t', u_t, y_t)'$; if $x_t$ contains lagged values of $y_t$ then $v_t = (x_t', y_t)'$ where $x_t'$ contains all elements of $x_t$ besides the lagged values of $y_t$. Then:

(i) $\{v_t\}$ is a piece-wise geometrically ergodic process, i.e. for each sub-sample $[T_{j-1}^0 + 1, T_j^0]$, there exists a unique stationary distribution $P_j$ such that:

$$
\sup_A |P(A|B) - P_j(A)| \leq g_j(B)\rho
$$
with \(0 < \rho < 1\), \(A \in \mathcal{F}_{T_i}^0\), \(B \in \mathcal{F}_{-\infty}^0\), \(\mathcal{F}_k^i\) is the \(\sigma\)-algebra generated by \((v_k, \ldots, v_l)\), and \(g_j(\cdot)\) is a positive uniformly integrable function.

(ii) \(\{v_i\}\) is a \(\beta\)-mixing process with exponential decay, i.e. there exists \(N > 0\) such that for \(B \in \mathcal{F}_N^0\),

\[
\beta_t = \sup_{\alpha} \beta(\mathcal{F}_t^0, \mathcal{F}_t^0) \leq N \rho^t, \quad \text{with} \quad \beta(\mathcal{F}_t^0, \mathcal{F}_t^0) = \sup_{A \in \mathcal{F}_t^0} \|P(A|B) - P(A)\|
\]

**Assumption A.2** The function \(f_t(\cdot)\) is a known measurable function, twice continuously differentiable in \(\theta\) for each \(t\).

**Assumption A.3** Let \(F_i(\beta) = \partial f_i(\beta)/\partial \beta\), a \(p \times 1\) vector and \(f_i^{(2)}(\beta)\), a \(p \times p\) matrix of second derivatives, i.e. \(f_i^{(2)}(\beta) = \partial^2 f_i(\beta)/(\partial \beta \partial \beta')\), with \((i, j)^{th}\) element \(f_i^{(2)}(\beta)\). Also denote by \(\| \cdot \|\) the Euclidean norm. Then (i) the common parameter space \(B\) is a compact subset of \(\mathbb{R}^p\); for some \(s > 2\), we have: (ii) \(\sup_{t, \beta} E\|u_t f_i(\beta)\|^s < \infty\); (iii) \(\sup_{t, \beta} E\|u_t F_i(\beta)\|^s < \infty\); (iv) For \(i, j = 1, \ldots, p\), \(\sup_{t, \beta} E\|u_t f_i^{(2)}(\beta)\|^s < \infty\).

**Assumption A.4** (i) \(S_T(T_1, \ldots, T_m; \beta)\) has a unique global minimum at \(\beta_0\) and \((T_1^0, \ldots, T_m^0)\); (ii) Let \(A_{T,i}(\beta, r) = \text{Var} T^{-1/2} \sum_{t = T_{i-1}^0 + 1}^{T_i^0} u_i(\beta) F_i(\beta)\). Then \(A_{T,i}(\beta, r) \xrightarrow{P} r A_i(\beta)\), uniformly in \(\beta \times r \in \mathcal{B} \times [0, \lambda_i^0 - \lambda_{i-1}^0]\), where \(A_i(\beta)\) is a positive definite (p.d.) matrix not depending on \(T\), with \(A_i(\beta)\) not necessarily the same for all \(i\); (iii) Let \(D_{T,i}(\beta, r) = T^{-1} \sum_{t = T_{i-1}^0 + 1}^{T_i^0} F_i(\beta) F_i(\beta)'\). Then \(D_{T,i}(\beta, r) \xrightarrow{P} r D_i(\beta)\), uniformly in \(\beta \times r \in \mathcal{B} \times [0, \lambda_i^0 - \lambda_{i-1}^0]\), where \(D_i(\beta)\) is a p.d. matrix; (iv) \(E[f_i(\beta_i^0)] \neq E[f_i(\beta_{i+1}^0)]\), for each \(i = 1, 2, \ldots, m\).

The proof follows similar lines to that of Theorem 1. From Boldea and Hall (2013) equation (39), it follows that (12) applies, with

\[
G_i(k_i) = \left\{ \begin{array}{ll}
|k_i| a_{i,1} - 2c_{i,1}^1 W_{i,1}(-k_i), & k_i \leq 0 \\
|k_i| a_{i,2} - 2c_{i,2}^1 W_{i,2}(k_i), & k_i > 0
\end{array} \right.
\]

and, for \(j = 1, 2\),

\[
a_{i,j} = \theta_i^{(j)} D_{i+j-1}(\beta_{i+j-1}^0) \theta_i^{(j)}
\]

\[
c_{i,j} = \sigma^2 \theta_i^{(j)} A_{i+j-1}(\beta_{i+j-1}^0) \theta_i^{(j)}.
\]

The result for \(\xi_{1,T}\) then follows by the corresponding argument in the proof of Theorem 1. For \(\xi_{2,T}\), the proof again follows the same argument as Theorem 1 using \(T^{1/2}(\hat{\beta}_i - \beta_i^0) \xrightarrow{d}\)
\( \mathcal{N}(0, \sigma^2[D_i(\beta^0_i)]^{-1}) \) (under our conditions) from analogous arguments to Boldea and Hall (2013)[Theorem 2].

**Proof of Theorem 3**

**Part (i):** From the principle of least squares,

\[
\xi_{1,T} = \min_{(T_1, \ldots, T_m)} \text{RSS}(T_1, T_2, \ldots, T_m) - \text{RSS}(T^0_1, T^0_2, \ldots, T^0_m).
\]  

(64)

There are two scenarios of interest for the general case of an unstable reduced form with \( h > 0 \) in (22), namely whether the (true) reduced form and structural breaks are common or not. To be more precise, and following Boldea, Hall, and Han (2012), we consider scenarios where some breaks occur in the structural form but not the reduced form and where at least some breaks are common to both; the former includes the special case of a stable reduced form. These scenarios can be represented as follows.

**Scenario 1:** \( \pi^0_j < \lambda^0_k < \ldots < \lambda^0_{k+\ell} < \pi^0_{j+1} \)

**Scenario 2:** \( \pi^0_{j-1} \leq \lambda^0_{k-1} < \pi^0_j = \lambda^0_k < \lambda^0_{j+\ell} \leq \pi^0_{j+1} \)

**Scenario 1**

Consider, firstly, a single reduced form break and \( m \) structural form breaks, with \( \pi^0_1 < \lambda^0_1 < \ldots < \lambda^0_{m+1} < T \), so that

\[
y_t = (x'_t, z'_t)\beta^0_i + u_t, \quad i = 1, 2, \ldots, m, \ t = T^0_{i-1} + 1, \ldots, T^0_i
\]

\[
x'_t = \begin{cases} 
  z'_t\Delta_0^{(1)} + v_t & t \leq T^*_1 \\
  z'_t\Delta_0^{(2)} + v_t & t > T^*_1 
\end{cases}
\]

Let \( \kappa^*_1 = [T\pi^0_1], \kappa^*_{10} = [T\pi^0_1] \), and define \( \hat{\kappa}^*_1 \) to be the estimator of \( \kappa^*_{10} \) based on the reduced form. As in Boldea, Hall, and Han (2012), proof of Theorem 3, the relevant intervals for the limiting behaviour of \( \{\hat{T}_i\}_{i=1}^m \) in (64) are \( B = \bigcup_{i=1}^m B_i \) where \( B_i = \{|T_i - T^0_i| \leq K_is_T^{-2}\} \) for positive constants \( K_i, \ i = 1, \ldots, m \).

Then, from Boldea, Hall, and Han (2012)[Proposition 2], (64) implies \( \xi_{1,T} \) can be written as
in (58), now with
\[ A_i(T_i) = \theta_{T,i}^0 \Upsilon_2^0 T_i \sum_{t=(T_i \lor T^0_i)+1} z_t z'_{t,i} \Upsilon_2^0 \theta_{T,i}^0, \]
\[ C_i(T_i) = (-1)^2 (T_i < T^0_i) \theta_{T,i}^0 \Upsilon_2^0 T_i \sum_{t=(T_i \lor T^0_i)+1} z_t z'_{t,i} \]
for \( \theta_{T,i}^0 \) and \( \Upsilon_0 \) defined in Assumptions 3 and 12, respectively and \( z_{t,i} \) defined in (29).

For break \( i \) consider \( T_i = T^0_i + [k_i s_{T}^{-2}] \) for \( k_i \epsilon [-K_i, K_i] \). Using the same arguments as Boldea, Hall, and Han (2012) in the proof of their Theorem 2, it follows that
\[ \xi_{1,T} \overset{d}{=} \sum_{i=1}^m \min_{k_i} G_i(k_i) \]
where
\[ G_i(k_i) = \begin{cases} |k_i| a_{i,j} - 2 c_{i,1} W_{i,1}(-k_i), & k_i \leq 0 \\ |k_i| a_{i,j} - 2 c_{i,2} W_{i,2}(k_i), & k_i > 0 \end{cases} \]
in which \( W_{i,j}(. \) are independent Brownian motions on \([0, \infty)\) and, using Assumption 11, we have
\[ a_{i,j} = \theta_i^0 \Upsilon_2^0 Q_Z Z(i + j - 1) \Upsilon_2^0 \theta_i^0 \]
\[ c_{i,j} = \theta_i^0 \Upsilon_2^0 \Phi(i + j - 1) \Upsilon_2^0 \theta_i^0 \]
for \( j = 1, 2 \), where \( \Phi(\ell) = C_\ell V_\ell C'_\ell \), \( C_\ell = \nu_\ell^\prime \otimes I_q \), \( \nu_\ell = [1, \nu_{x,\ell}] \). Under Assumption 11(iv) we have \( \Phi(\ell) = \nu_\ell \Omega_\ell \nu_\ell^\prime \otimes Q_{ZZ}(\ell) \), and with the addition of Assumption 11(v), we have \( \Phi(\ell) = \nu_\ell \Omega_\ell^\prime \otimes Q_{ZZ}(\ell) \). Thus, we have under our assumptions that
\[ c_{i,j} \rightarrow \rho^2 a_{i,j} \]
where \( \rho^2 \) is defined in (31) and Assumption 3 is imposed.

Since (65) implies that the breaks can be considered separately, while \( c_{i,j} = \rho^2 a_{i,j} \) in (69) under the shrinking breaks Assumption 3 then analogously to (19), we have
\[ \max_{|k_i|} \widetilde{G}([k_i]) \sim B(a_{i,j}/2c_{i,j}, a_{i,j}/2c_{i,j}) = B(0.5\rho^{-2}, 0.5\rho^{-2}) \]
Recalling that \( \widetilde{G}(k_i) = -\widetilde{G}(k_i) \), then from Lemma 2 and for \( m \) breaks, it follows that
\[ AE[\xi_{1,T}] = -3m\rho^2. \]

Under the shrinking breaks Assumption 8 and with distinct reduced and structural form breaks such that \( \pi_j^0 < \lambda_k^0 < \ldots < \lambda_{k+\ell}^0 < \pi_{j+1}^0 \), the result immediately extends to the case where
the number of reduced form breaks is \( h > 1 \). It also immediately specialises to the case of a stable reduced form.

**Scenario 2**

Under this scenario, consider \( h = 1 \) in the case where the first of the \( m \) structural breaks coincides with the single reduced form break. Hence the data generation process is identical to Scenario 1, except that \( T^1 = T^0_1 \) and, consequently, \( \pi_1^0 = \lambda_1^0 \).

From Boldea, Hall, and Han (2012) equation (96), and since the \( m \) breaks at \( T^0_0, ..., T^0_m \) can be considered separately, (65) again applies, with the modification that now

\[
a_{1,j} = \theta_1^0 (T^0_0) Q_{zz}(j) \gamma_j^0 \theta_1^0, \quad j = 1, 2 \tag{71}
\]

and

\[
a_{i,j} = \theta_i^0 (T^0_2) Q_{zz}(i + j - 1) \gamma_i^0 \theta_i^0, \quad i = 2, ..., m, \tag{72}
\]

Under our assumptions, we again have

\[
c_{i,j} = \rho_i^2 a_{i,j} \rightarrow \rho^2 a_{i,j} \quad j = 1, 2 \quad i = 1, ..., m \tag{75}
\]

under shrinking breaks. Applying Lemmata 1 and 2 for a single structural form break leads to

\[
\max |k_1| (k_1) \sim B(a_{1,1}/2c_{1,1}, \ a_{1,2}/2c_{1,2}) = B(0.5\rho^{-2}, \ 0.5\rho^{-2}).
\]

Since this holds for a break that is common to the reduced and structural forms, and also (from Scenario 1) when \( \pi_1^0 < \lambda_k^0 < ... < \lambda_{k+\ell}^0 < \pi_{j+1}^0 \), then (70) also holds under Scenario 2.

**Part (ii):** From standard least squares algebra,

\[
\xi_{2,T} = RSS(T^0_1, T^0_2, ..., T^0_m; \pi^0) - ESS(T^0_1, T^0_2, ..., T^0_m)
\]

\[
= \sum_{i=1}^{m} \sum_{t=T^0_{i-1}+1}^{T^0_i} (y_t - \hat{x}_i(\pi^0)\beta_{x,i} - \hat{z}_t^{0,i}\beta_{z,i})^2 - \sum_{i=1}^{m} \sum_{t=T^0_{i-1}+1}^{T^0_i} (y_t - \hat{x}_i(\pi^0)\beta_{x,i} - \hat{z}_t^{0,i}\beta_{z,i})^2
\]

\[
= -\sum_{i=1}^{m} (\hat{\beta}_i - \beta_i^0)'(\hat{W}_i'\hat{W}_i)(\hat{\beta}_i - \beta_i^0) \tag{76}
\]
in which $\tilde{W}_i$ is the $(T_i - T_{i-1}) \times p$ data matrix for the $i^{th}$ structural form regime, with typical row $(\tilde{x}_t(\pi^0)', \tilde{z}_{t,i}')$, and $\tilde{\beta}_i = (\tilde{\beta}_{x,i}', \tilde{\beta}_{z,i}')'$ are obtained using the true reduced form break fractions $\pi^0$.

It is useful to first consider the form of $\tilde{Q}_i = \tilde{Q}_{ZZ}(\lambda_i^0) - \tilde{Q}_{ZZ}(\lambda_{i-1}^0)$ where $\tilde{Q}_{ZZ}(r)$ is the uniform in $r \in (0,1]$ limit of $T^{-1} \sum_{t=1}^{T r} \tilde{z}_t(\pi^0)\tilde{z}_t(\pi^0)'$. If, without loss of generality, we assume that $\pi_i^0 < \lambda \leq \pi_{i+1}^0$, then it follows from Assumption 14 that

$$\tilde{Q}_{ZZ}(\lambda) = \phi(\lambda) \otimes Q_{ZZ}$$

(77)

where

$$\phi(\lambda) = diag[\delta \pi_1^0, ..., \delta \pi_i^0, \lambda - \pi_i^0, 0, ..., 0].$$

Therefore, we have

$$\tilde{Q}_i = \phi_i^{(1)} \otimes Q_{ZZ}$$

(78)

where $\phi_i^{(1)} = \phi(\lambda_i^0) - \phi(\lambda_{i-1}^0)$. We note there are two scenarios for $\phi_i^{(1)}$: if there are no reduced form breaks between $\lambda_{i-1}^0$ and $\lambda_i^0$ then

$$\phi_i^{(1)} = diag[0, ..., 0, \delta \lambda_i^0, 0, ..., 0];$$

(79)

if there are reduced form breaks between $\lambda_{i-1}^0$ and $\lambda_i^0$, say $\pi_k^0 < \lambda_{i-1}^0 < \pi_{k+1}^0 < ... < \pi_{k+\ell_i}^0 < \lambda_i^0$, then

$$\phi_i^{(1)} = diag[0, ..., 0, (\pi_{k+1}^0 - \lambda_{i-1}^0), \delta \pi_{k+2}^0, ..., \delta \pi_{k+\ell_i}^0, (\lambda_i^0 - \pi_{k+\ell_i}^0), 0, ..., 0].$$

(80)

For later reference it is also useful to note that

$$\tilde{Q}_{ZZ}(1) = \phi_0 \otimes Q_{ZZ}$$

(81)

where

$$\phi_0 = \phi(1) = diag[\delta \pi_1^0, \delta \pi_2^0, ..., \delta \pi_{h+1}^0].$$

(82)

We now return to the proof. From the proof of Hall, Han, and Boldea (2012)[Theorem 8], we have that

$$T^{-1} \tilde{W}_i' \tilde{W}_i = \tilde{M}_{uw}^{(i)} \tilde{P} \tilde{M}_{uw}^{(i)} = \tilde{Y}' \tilde{Q}_i \tilde{Y}$$

where

$$\tilde{Y}' = [\Upsilon_1^0, \Upsilon_2^0, ..., \Upsilon_{h+1}^0].$$
Using Assumption 3 and either equation (79) or (80) as appropriate, it follows that
\[ T^{-1}\tilde{W}_i' \tilde{W}_i \overset{p}{\rightarrow} (\delta \lambda_0^0)' \Upsilon_0 Q ZZ \ Upsilon_0 = (\delta \lambda_0^0) M_0, \]
where \( M_0 = \Upsilon_0 Q ZZ \ Upsilon_0 \) and \( \Upsilon_0 = [\Delta_0, I]. \) From Hall, Han, and Boldea (2012)[Theorem 3], we have that
\[ T^{1/2} \left( \tilde{\beta}_i - \beta_i^0 \right) \Rightarrow N(0, V_{i,i}^\beta) \]
where \( V_{i,i}^\beta \) is in Hall, Han, and Boldea (2012)[Theorem 8],
\[ V_{i,i}^\beta = \tilde{A}_i \left\{ \tilde{C}_i \tilde{V}_i \tilde{C}_i' - \tilde{E}_i \tilde{D}_i \tilde{V}_i \tilde{D}_i' \tilde{E}_i' + \tilde{E}_i \tilde{D}_i \tilde{V}_i \tilde{D}_i' \tilde{E}_i' \right\} \tilde{A}_i \quad (83) \]
and
\[ \tilde{A}_i = [\tilde{\Upsilon}' \tilde{Q}, \tilde{\Upsilon}]^{-1} \tilde{\Upsilon}' \]
\[ \tilde{C}_i = (1, \beta_{x,i}^{0'}) \otimes I_{\tilde{q}}, \quad \tilde{D}_i = (0, \beta_{x,i}^{0'}) \otimes I_{\tilde{q}}, \quad \tilde{q} = q(h + 1) \]
\[ \tilde{E}_i = \tilde{Q}_i Q ZZ(1)^{-1} \]
\[ \tilde{V}_i = \text{Var} \left[ \int T^{-1/2} \sum_{t=[\lambda_{i-t}^0 T]+1}^{[\lambda_i^0 T]} \tilde{\lambda}_t \right], \quad \tilde{\lambda}_t = \left( \begin{array}{c} u_t \\ v_t \end{array} \right) \otimes \tilde{z}_t(\pi_0). \]
Under Assumption 11 we have
\[ \tilde{V}_i = \phi_i^{(1)} \otimes V = \phi_i^{(1)} \otimes (\Omega \otimes Q ZZ) \]
where \( \phi_i^{(1)} \) is defined by either (79) or (80). Also using (77),
\[ \tilde{E}_i = \phi_i^{(2)} \otimes I_{\tilde{q}}, \]
where \( \phi_i^{(2)} = \phi_i^{(1)} \{ \phi(1) \}^{-1}. \)

Now consider each of the terms of (83) in turn. Firstly, since \((1, \beta_{x,i}^{0'})\Omega(1, \beta_{x,i}^{0'})' = \rho_i^2 \) in (31), then
\[ \tilde{C}_i \tilde{V}_i \tilde{C}_i' = \rho_i^2 (\phi_i^{(1)} \otimes Q ZZ). \]
If \( \phi_i^{(i)} \) is given by (79) and \( \pi_k \leq \lambda_i^0 \leq \pi_{k+1} \) then
\[ \tilde{\Upsilon}' \tilde{C}_i \tilde{V}_i \tilde{C}_i' \tilde{\Upsilon} = \rho_i^2 \delta \lambda_i^0 \Upsilon_{k+1}^0 Q ZZ \ Upsilon_{k+1}^0 \rightarrow (\delta \lambda_i^0) \rho_i^2 \Upsilon_{k+1}^0 \ Upsilon_{k+1}^0 \ Upsilon_0 \]
under Assumption 8. If \( \phi_i^{(i)} \) is given by (80) then we have
\[ \tilde{\Upsilon}' \tilde{C}_i \tilde{V}_i \tilde{C}_i' \tilde{\Upsilon} = \rho_i^2 \left\{ (\pi_{k+1}^0 - \lambda_i^0) \Upsilon_{k+1}^0 Q ZZ \ Upsilon_{k+1}^0 + \delta \pi_{k+2}^0 \Upsilon_{k+2}^0 Q ZZ \ Upsilon_0 + \ldots \right\} \]
\[ \rightarrow (\delta \lambda_i^0) \rho_i^2 \Upsilon_{k+1}^0 \ Upsilon_{k+1}^0 \ Upsilon_0 \quad (85) \]
under Assumption $\square$ By similar arguments, $\bar{D}_i\bar{V}_i\bar{C}_i = (\phi_i^{(1)} \otimes Q_{ZZ})\bar{p}_i$ and hence

$$\bar{E}_i \bar{D}_i \bar{V}_i \bar{C}_i' = (\phi_i^{(2)} \otimes I_q)(\phi_i^{(1)} \otimes Q)\bar{p}_i$$

$$= \bar{p}_i(\phi_i^{(3)} \otimes Q_{ZZ})$$

where $\phi_i^{(3)} = \phi_i^{(1)} \phi_i^{(2)}$. Using Assumption $\square$ it follows that

$$\bar{Y}' \bar{E}_i \bar{D}_i \bar{V}_i \bar{C}_i' \bar{Y} \rightarrow \bar{p}_i d_i \bar{Y}^0 Q_{ZZ} \bar{Y}^0$$

(86)

where $d_i = \sum_{j=1}^{h+1}\{\phi_i^{(3)}\}_{j,j}$ and $\{\phi_i^{(3)}\}_{j,j}$ is the $(j,j)\text{th}$ element of $\{\phi_i^{(3)}\}$. Note that if $\phi_i^{(1)}$ is given by (79) then

$$d_i = \{\delta \lambda_0^0\}^2$$

(87)

and if $\phi_i^{(1)}$ is given by (80) then

$$d_i = \left(\frac{\alpha_0^{k+1} - \lambda_0^{k+1}}{\delta \pi_{k+1}^0}\right)^2 + \left(\frac{\lambda_0^{k+1} - \pi_{k+1}^0}{\delta \pi_{k+1}^0}\right)^2 \pi_{k+1}^0 + \pi_{k+1}^0$$

(88)

Finally, since $\bar{D}_i \bar{V}_i \bar{D}_i' = \omega_2^2(\phi_0 \otimes Q_{ZZ})$ where $\phi_0$ is defined in (82), then

$$\bar{E}_i \bar{D}_i \bar{V}_i \bar{D}_i' \bar{E}_i' = \omega_2^4(\phi_i^{(2)} \otimes I_q)(\phi(1) \otimes Q_{ZZ})(\phi_i^{(2)} \otimes I_q)$$

$$= \omega_2^4(\phi_i^{(3)} \otimes Q_{ZZ})$$

since $\phi_i^{(2)}(\phi(1) = \phi_i^{(1)}$ and $\phi_i^{(1)} \phi_i^{(2)} = \phi_i^{(3)}$. Consequently, under Assumption $\square$ we have

$$\bar{Y}' \bar{E}_i \bar{D}_i \bar{V}_i \bar{D}_i' \bar{E}_i' \bar{Y} \rightarrow \omega^2 d_i \bar{Y}^0 Q_{ZZ} \bar{Y}^0.$$ 

(89)

Substituting from (85), (86) and (89) into (83) yields

$$V_{i,i}^\beta \rightarrow \{M_{ww}^{(i)}\}^{-1}\{(\delta \lambda_0^0)^2 - 2\bar{p}_i \rho^2 + \omega_2^4 d_i \bar{Y}^0 Q_{ZZ} \bar{Y}^0\{M_{ww}^{(i)}\}^{-1}.$$ 

Since $\bar{Y}^0 Q_{ZZ} \bar{Y}^0\{M_{ww}^{(i)}\}^{-1} = (\delta \lambda_0^0)^{-1} I_p$, and further using (32) and (33),

$$\begin{align*}
M_{ww}^{(i)} V_{i,i}^\beta &\rightarrow \left\{\rho^2 - 2\beta_x^0 \gamma \frac{d_i}{\delta \lambda_i^0} - \beta_x^0 \Sigma \beta_x^0 \frac{d_i}{\delta \lambda_i^0}\right\} I_p
\end{align*}$$

and hence

$$AE[\xi_{2,T}] = -\sum_{i=1}^{m+1} \text{tr} \left[(V_{i,i}^\beta)^{1/2} M_{ww}^{(i)} (V_{i,i}^\beta)^{1/2}\right]$$

$$= -\sum_{i=1}^{m+1} \text{tr}[M_{ww}^{(i)} V_{i,i}^\beta]$$

$$= -p(m+1)\rho^2 + p \sum_{i=1}^{m+1} \frac{d_i}{\delta \lambda_i^0} (2\beta_x^0 \gamma + \beta_x^0 \Sigma \beta_x^0)$$

$$= -p(m+1)\rho^2 + p(\rho^2 - \sigma^2) \sum_{i=1}^{m+1} \frac{d_i}{\delta \lambda_i^0}$$

(90)
where the last expression is obtained using (31).

**Part (iii):** For $\xi_{3,T}$ defined by (37), consider the regime-specific errors

$$\begin{align*}
y_t - \tilde{x}_t^0 \beta_{x,i}^0 - z_{1,t}^0 \beta_{z,i}^0 &= (y_t - \pi_t^0 \beta_{x,i}^0 - z_{1,t}^0 \beta_{z,i}^0) + (\pi_t - \tilde{x}_t^0) \beta_{x,i}^0 \\
&= \pi_t - \tilde{x}_t^0 \beta_{x,i}^0
\end{align*}$$

where $\tilde{x}_t$ is obtained using the true reduced form break dates. Since

$$ESS(T_1^0, T_2^0, ..., T_m^0) = \sum_{i=1}^{m+1} \sum_{t=T_{i-1}^0+1}^{T_i^0} [\pi_{t,i} + (\pi_t - \tilde{x}_t^0) \beta_{x,i}^0]^2$$

and

$$ESS'(T_1^0, T_2^0, ..., T_m^0) = \sum_{i=1}^{m+1} \sum_{t=T_{i-1}^0+1}^{T_i^0} \pi_{t,i}^2,$$

it immediately follows that

$$\begin{align*}
\xi_{3,T} &= \sum_{i=1}^{m+1} \left\{ \sum_{t=T_{i-1}^0+1}^{T_i^0} \beta_{x,i}^0 (\pi_t - \tilde{x}_t^0)(\pi_t - \tilde{x}_t^0) \beta_{x,i}^0 + 2 \sum_{t=T_{i-1}^0+1}^{T_i^0} \pi_{t,i} (\pi_t - \tilde{x}_t^0) \beta_{x,i}^0 \right\} \\
&= \sum_{i=1}^{m+1} (E_{2i} + 2E_{3i})
\end{align*}$$

where (obviously)

$$\begin{align*}
E_{2i} &= \sum_{t=T_{i-1}^0+1}^{T_i^0} \beta_{x,i}^0 (\pi_t - \tilde{x}_t^0)(\pi_t - \tilde{x}_t^0) \beta_{x,i}^0 \\
E_{3i} &= \sum_{t=T_{i-1}^0+1}^{T_i^0} \pi_{t,i} (\pi_t - \tilde{x}_t^0) \beta_{x,i}^0.
\end{align*}$$

From (23) and (24),

$$\begin{align*}
\pi_t - \tilde{x}_t^0 &= \tilde{z}_t^0 (\Theta_0 - \hat{\Theta}_T) \\
&= -\tilde{z}_t^0 \left\{ \sum_{t=1}^{T} \tilde{z}_t \tilde{z}_t^0 \right\}^{-1} \sum_{t=1}^{T} \tilde{z}_t v_t^0
\end{align*}$$

where it is understood that $\tilde{z}_t = \tilde{z}_t(\pi^0)$. Substituting (95) into (93), we obtain

$$\begin{align*}
E_{2i} &= \sum_{t=T_{i-1}^0+1}^{T_i^0} \beta_{x,i}^0 v_t \sum_{t=1}^{T} \tilde{z}_t \tilde{z}_t^0 \left\{ \sum_{t=1}^{T} \tilde{z}_t \tilde{z}_t^0 \right\}^{-1} \sum_{t=1}^{T} \tilde{z}_t v_t^0 \beta_{x,i}^0 \\
&= \sum_{t=1}^{T} \beta_{x,i}^0 v_t \sum_{t=1}^{T} \tilde{z}_t \tilde{z}_t^0 \left\{ \sum_{t=1}^{T} \tilde{z}_t \tilde{z}_t^0 \right\}^{-1} \sum_{t=1}^{T} \tilde{z}_t v_t^0 \beta_{x,i}^0.
\end{align*}$$

33
Using (30), we have
\[
E_{2i} = T^{-1/2} \sum_{t=1}^{T} \bar{u}_{t,i} \bar{z}_t' \left\{ T^{-1} \sum_{t=1}^{T} \bar{z}_t \bar{z}_t' \right\}^{-1} T^{-1} \sum_{t=1}^{T^0} \tilde{z}_t \tilde{z}_t' \times \left\{ T^{-1} \sum_{t=1}^{T} \tilde{z}_t \tilde{z}_t' \right\}^{-1} T^{-1/2} \sum_{t=1}^{T} \bar{z}_t \bar{u}_{t,i}.
\]

From (78) and (81), it follows that
\[
AE[E_{2i}] = tr \left\{ \phi_i^{(4)} \otimes Q_{ZZ} \right\} lim_{T \to \infty} E \left[ \left( T^{-1/2} \sum_{t=1}^{T} \bar{z}_t \bar{v}_{t,i} \right) \left( T^{-1/2} \sum_{t=1}^{T} \bar{z}_t \bar{v}_{t,i} \right)' \right]
\]
where \( \phi_i^{(4)} = \phi_i^{(2)} \phi_0^{-1} \) and
\[
lim_{T \to \infty} E \left[ \left( T^{-1/2} \sum_{t=1}^{T} \bar{z}_t \bar{v}_{t,i} \right) \left( T^{-1/2} \sum_{t=1}^{T} \bar{z}_t \bar{v}_{t,i} \right)' \right] = \omega_i^2 (\phi_0 \otimes Q_{ZZ})
\]
where \( \omega_i^2 = Var[\bar{v}_{t,i}] \). Therefore
\[
AE[E_{2i}] = tr \left\{ \phi_0 \phi_i^{(4)} \otimes I_q \right\} \omega_i^2
\]
\[
= tr \left\{ \phi_i^{(2)} \otimes I_q \right\} \omega_i^2
\]
\[
= q \omega_i^2 b_i
\]
where \( b_i = \sum_{h+1}^{T^0} \phi_i^{(2)}_{j,j} \) and \( \{ \phi_i^{(2)}_{j,j} \} \) is the \((j,j)\)th element of \( \phi_i^{(2)} \).

In a similar way, substituting (95) in the definition of (93) yields
\[
E_{3i} = - T^{-1/2} \sum_{t=T^0+1}^{T} \bar{u}_{t,i} \bar{z}_t' \left\{ T^{-1} \sum_{t=1}^{T} \bar{z}_t \bar{z}_t' \right\}^{-1} T^{-1} \sum_{t=1}^{T^0} \tilde{z}_t \tilde{u}_{t,i} \tilde{z}_{t} + o_p(1).
\]

Applying similar arguments to those for \( E_{2i} \), we obtain
\[
AE[E_{3i}] = -q \tilde{p}_i b_i
\]
where \( \tilde{p}_i = Cov[\bar{v}_{t,i}, \bar{u}_{t,i}] \). Therefore, under Assumption 3 we have
\[
AE[\xi_{3,T}] \to q [\omega^2 - 2 \tilde{p}] \sum_{i=1}^{m+1} b_i.
\]

To complete the proof note that \( \sum_{i=1}^{m+1} b_i = h + 1 \) and \( \omega^2 - 2 \tilde{p} = - (\rho^2 - \sigma^2) \) from (32) to (33).
Part (iv): Using the definition of $\xi_{4,T}$ in (38) and also (91), it immediately follows from (31) that
\[ E[\xi_{4,T}] = 0. \] (97)
Simple algebra then yields the result given for $AE[\xi_T]$ in Theorem 3.

To establish $0 < \sum_{i=1}^{m+1} d_i / (\delta \lambda^0_i) \leq \min[(h+1),(m+1)]$, note first that $d_i$ and $\delta \lambda^0_i$ ($i = 1, \ldots, m + 1$) are strictly positive, by definition. Then for a structural form regime with no intermediate reduced form breaks, $\pi^0_i \leq \lambda^0_i < \lambda^0_0 \leq \pi^0_{k+1}$, say, it immediately follows that $d_i / (\delta \lambda^0_i) = \{\delta \lambda^0_i\}^2 / \{\delta \pi^0_i \times \delta \lambda^0_i\} = (\delta \lambda^0_i) / (\delta \pi^0_i) \leq 1$, with equality holding if and only if $\pi^0_i = \lambda^0_i$ and $\lambda^0_i = \pi^0_{k+1}$. With intermediate reduced form breaks, $\pi^0_k \leq \lambda^0_i < \pi^0_{k+1} < \ldots < \pi^0_{k+\ell_i} < \lambda^0_i$, say, with $\ell_i \geq 1$, then
\[ d_i = \frac{(\pi^0_{k+1} - \lambda^0_i)}{\delta \lambda^0_i} + \frac{(\lambda^0_i - \pi^0_{k+\ell_i})}{\delta \pi^0_{k+\ell_i}} + \lambda^0_i - \pi^0_{k+\ell_i} \]
\[ < \frac{\pi^0_{k+1} - \lambda^0_i}{\delta \lambda^0_i} + \frac{\lambda^0_i - \pi^0_{k+\ell_i}}{\delta \pi^0_{k+\ell_i}} + \frac{\pi^0_{k+\ell_i} - \pi^0_{k+1}}{\delta \pi^0_{k+1}} = \delta \lambda^0_i \]
since $\pi^0_{k+1} - \lambda^0_i \leq \delta \pi^0_{k+1}$ and $\lambda^0_i - \pi^0_{k+\ell_i} \leq \delta \pi^0_{k+\ell_i}$, with equality holding if both $\lambda^0_i = \pi^0_k$ and $\pi^0_{k+\ell_i} = \lambda^0_i$. Therefore, $d_i / (\delta \lambda^0_i) \leq 1$ also in this case. Summed over all $m + 1$ structural form regimes, it immediately follows that
\[ 0 < \sum_{i=1}^{m+1} d_i / (\delta \lambda^0_i) \leq m + 1. \]

From the perspective of the reduced form regimes, define $d^*_j$ as follows: If reduced form regime $j$ contains no structural form breaks, so that $\lambda^0_j \leq \pi^0_{j-1} < \pi^0_j \leq \lambda^0_{i+1}$, $d^*_j = \delta \pi^0_j / \delta \lambda^0_j$; if reduced form regime $j$ includes $\ell_j$ structural form breaks, $\lambda^0_j \leq \pi^0_{j-1} < \pi^0_{j+1} < \ldots < \lambda^0_{i+\ell_j} < \pi^0_j \leq \lambda^0_{i+\ell_j+1}$, then
\[ d^*_j = \frac{(\lambda^0_{i+1} - \pi^0_{j-1})^2}{\delta \lambda^0_{i+1} \times \delta \pi^0_j} + \sum_{s=2}^{\ell_j} \frac{\delta \lambda^0_{i+s}}{\delta \pi^0_j} + \frac{(\pi^0_j - \lambda^0_{i+\ell_j})^2}{\delta \pi^0_{i+\ell_j+1} \times \delta \pi^0_j}. \] (98)
From these definitions, it follows that each $d^*_j \leq 1$: this is obvious for the case of no intermediate structural form breaks, while (98) implies that
\[ d^*_j \leq \frac{\lambda^0_{i+1} - \pi^0_{j-1}}{\delta \pi^0_j} + \sum_{s=2}^{\ell_j} \frac{\delta \lambda^0_{i+s}}{\delta \pi^0_j} + \frac{\pi^0_j - \lambda^0_{i+\ell_j}}{\delta \pi^0_{i+\ell_j+1}} = \frac{\delta \pi^0_j}{\delta \pi^0_j} = 1 \]
since $(\lambda^0_{i+1} - \pi^0_{j-1}) \leq \delta \lambda^0_{i+1}$ and $(\pi^0_j - \lambda^0_{i+\ell_j}) \leq \delta \lambda^0_{i+\ell_j+1}$. Also note that $d^*_j = 1$ in (98) when $\pi^0_{j-1} = \lambda^0_j$ and $\pi^0_j = \lambda^0_{i+\ell_j+1}$. Further, since $\lambda^0_0 = \pi^0_0 = 0$ and $\lambda^0_{m+1} = \pi^0_{n+1} = 1$, it also follows
that \( \sum_{i=1}^{m+1} d_i / (\delta \lambda_i^0) = \sum_{j=1}^{h+1} d_j^* \leq (h + 1) \), thereby establishing the required result. \( \Diamond \)

**Proof of Theorem 4**

Under \( H_0 \),

\[
N_{\lambda}(\lambda) = -\xi_{1,T}.
\]

From (12) and (13),

\[
\xi_{1,T} \rightarrow^d \sum_{i=1}^{m} \max_{|k_i|} H_i(|k_i|)
\]

where

\[
H_i(|k_i|) = \begin{cases} 
-|k_i| a_{i,1} + 2 c_{i,1}^{1/2} W_{i,1}(|k_i|), & k_i \leq 0 \\
-|k_i| a_{i,2} + 2 c_{i,2}^{1/2} W_{i,2}(|k_i|), & k_i > 0 
\end{cases}
\]

where \( a_{i,j}, c_{i,j} \) defined in (14), (15). From Lemmata 1 and 2,

\[
\max_{|k_i|} H_i(|k_i|) = \bar{b}_i \sim B(\mu_{i,1}, \mu_{i,2}).
\]

The desired result follows because Assumptions 1 and 3 imply independence of \( \bar{b}_i \) and \( \bar{b}_j \) for \( i \neq j \). \( \Diamond \)
References


Figure 1: Power of the $F_\lambda$ statistics
Table 1: Critical values for test based on $F_\lambda(\bar{\lambda})$

<table>
<thead>
<tr>
<th>m</th>
<th>10%</th>
<th>5%</th>
<th>1%</th>
<th>m</th>
<th>10%</th>
<th>5%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5.94153</td>
<td>7.35814</td>
<td>10.58452</td>
<td>6</td>
<td>25.28185</td>
<td>27.91964</td>
<td>33.34145</td>
</tr>
<tr>
<td></td>
<td>10.21645</td>
<td>11.98347</td>
<td>15.85402</td>
<td>7</td>
<td>28.85278</td>
<td>31.64847</td>
<td>37.36935</td>
</tr>
<tr>
<td></td>
<td>14.16656</td>
<td>16.20431</td>
<td>20.55913</td>
<td>8</td>
<td>32.38591</td>
<td>35.32267</td>
<td>41.32152</td>
</tr>
<tr>
<td></td>
<td>17.96262</td>
<td>20.22017</td>
<td>24.98770</td>
<td>9</td>
<td>35.88423</td>
<td>38.95839</td>
<td>45.21370</td>
</tr>
<tr>
<td></td>
<td>21.65748</td>
<td>24.11753</td>
<td>29.22647</td>
<td>10</td>
<td>39.35408</td>
<td>42.56136</td>
<td>49.06494</td>
</tr>
</tbody>
</table>

Critical values at the 10%, 5%, and 1% significance level of the limiting distribution of $F_\lambda(\bar{\lambda})$ in Corollary 2, for models with $m$ number of breaks.
Table 2: OLS - one break model

| T   | α  | $|bias|$ | $C_{90}$ | $F_{90}$ | $|bias|$ | $C_{90}$ | $F_{90}$ | $|bias|$ | $C_{90}$ | $F_{90}$ |
|-----|----|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 120 | .0 | .0251   | .8056   | .7936   | .0166   | .8074   | .7948   | .0088   | .8122   | .7994   |
|     | .1 | .0239   | .8182   | .8038   | .0154   | .8200   | .8052   | .0120   | .8236   | .8102   |
|     | .2 | .0226   | .8288   | .8172   | .0141   | .8314   | .8190   | .0102   | .8354   | .8258   |
|     | .3 | .0212   | .8414   | .8310   | .0127   | .8438   | .8332   | .0086   | .8488   | .8382   |
|     | .4 | .0198   | .8550   | .8426   | .0113   | .8578   | .8454   | .0068   | .8612   | .8482   |
|     | .49| .0185   | .8622   | .8524   | .0100   | .8644   | .8544   | .0055   | .8672   | .8586   |
| 240 | .0 | .0113   | .8434   | .8360   | .0071   | .8446   | .8366   | .0013   | .8476   | .8392   |
|     | .1 | .0111   | .8490   | .8400   | .0069   | .8496   | .8408   | .0017   | .8518   | .8438   |
|     | .2 | .0108   | .8550   | .8460   | .0066   | .8556   | .8470   | .0019   | .8574   | .8492   |
|     | .3 | .0106   | .8590   | .8482   | .0064   | .8598   | .8488   | .0022   | .8610   | .8516   |
|     | .4 | .0103   | .8606   | .8506   | .0061   | .8614   | .8514   | .0024   | .8632   | .8534   |
|     | .49| .0101   | .8650   | .8530   | .0059   | .8660   | .8534   | .0026   | .8682   | .8550   |
| 360 | .0 | .0075   | .8576   | .8558   | .0047   | .8580   | .8564   | .0010   | .8600   | .8574   |
|     | .1 | *       | *       | *       | *       | *       | *       | *       | *       | *       |
|     | .2 | *       | *       | *       | *       | *       | *       | *       | *       | *       |
|     | .3 | *       | *       | *       | *       | *       | *       | *       | *       | *       |
|     | .4 | *       | *       | *       | *       | *       | *       | *       | *       | *       |
|     | .49| *       | *       | *       | *       | *       | *       | *       | *       | *       |
| 480 | .0 | .0059   | .8758   | .8676   | .0038   | .8764   | .8684   | .0004   | .8770   | .8694   |
|     | .1 | .0060   | .8744   | .8660   | .0039   | .8750   | .8666   | .0003   | .8756   | .8678   |
|     | .2 | .0060   | .8718   | .8638   | .0040   | .8724   | .8644   | .0003   | .8740   | .8650   |
|     | .3 | .0061   | .8700   | .8610   | .0040   | .8712   | .8620   | .0002   | .8718   | .8622   |
|     | .4 | .0062   | .8674   | .8582   | .0041   | .8684   | .8592   | .0001   | .8694   | .8600   |
|     | .49| .0062   | .8662   | .8580   | .0042   | .8670   | .8586   | .0001   | .8684   | .8592   |

Notes: $k$ indexes the degrees of freedom correction in the error variance estimator as described in Section 4; $|bias| = |\hat{\sigma}_{k}^2 - 1|$ where the true error variance is 1; $C_{90}$ ($F_{90}$) denotes the empirical coverage probability closest to (furthest from) the nominal value of .90 over the intervals in (56) for $\mu_1, \gamma_1, \mu_2$, and $\gamma_2$. 

41