Inference regarding multiple structural changes in linear models estimated via two stage least squares

Hall, Alastair R. and Han, Sanggohn and Boldea, Otilia

North Carolina State University, University of Manchester, UK, Hyundai Research Institute, Tilburg University, Netherlands

20 June 2008

Inference Regarding Multiple Structural Changes in Linear Models Estimated via Two Stage Least Squares[1]

Alastair R. Hall

University of Manchester[2]

Sanggohn Han

Hyundai Research Institute

and

Otilia Boldea

Tilburg University

June 20, 2008

[1]We are grateful to Denise Osborn and Eric Renault for valuable comments, and also for the comments of participants at the presentation of this paper at the World Congress of the Econometric Society, London, August 19-24, 2005, the Triangle Econometrics Conference, RTP, NC, December 2, 2005, CIREQ Conference on Generalized Method of Moments, Montreal, November 16-17, 2007, the ESRC Econometrics group seminar at the Institute of Fiscal Studies, London, the London-Oxbridge Time Series Workshop and at seminars at Erasmus University and the Universities of Birmingham and Warwick. We are also very grateful to Chengsi Zhang and Denise Osborn for providing us with the data used in the empirical example.

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

Abstract

In this paper, we extend Bai and Perron's (1998, *Econometrica*, p.47-78) framework for multiple break testing to linear models estimated via Two Stage Least Squares (2SLS). Within our framework, the break points are estimated simultaneously with the regression parameters via minimization of the residual sum of squares on the second step of the 2SLS estimation. We establish the consistency of the resulting estimated break point fractions. We show that various F-statistics for structural instability based on the 2SLS estimator have the same limiting distribution as the analogous statistics for OLS considered by Bai and Perron (1998). This allows us to extend Bai and Perron's (1998) sequential procedure for selecting the number of break points to the 2SLS setting. Our methods also allow for structural instability in the reduced form that has been identified *a priori* using data-based methods. As an empirical illustration, our methods are used to assess the stability of the New Keynesian Phillips curve.

JEL classification: C12, C13

Keywords: Structural Change, Multiple Break Points, Instrumental Variables Estimation.

1 Introduction

Linear models are widely applied in the analysis of macroeconomic time series. In many cases, at least some of the explanatory variables are correlated with the error and so the model is estimated via Instrumental Variables (IV). While it is routine to assume in estimation that the parameters of these models are constant over time, there are reasons why this assumption may be questionable. In particular, it can be argued that policy changes and/or exogenous shifts may cause realignments in the relationship between economic variables which are reflected in changes in the parameters. Therefore, it is important to develop methods for both detecting parameter instability and also for building models that incorporate this behaviour.

Considerable attention has focused on developing tests for structural instability within the IV or more generally the Generalized Method of Moments (GMM) framework.[1] The majority of this literature has focused on the design of tests against alternatives in which there is structural instability at a single breakpoint in the sample. Although these tests are also shown to have non-trivial power against other alternatives, it is clearly desirable to develop procedures that can discriminate between various forms of instability.

An important step in this direction is taken by Bai and Perron (1998).[2] They develop methods that are designed to test for discrete shifts in the parameters at potentially multiple and unknown break points in the sample. Their analysis is in the context of linear regression models estimated via Ordinary Least Squares (OLS). Within their framework, the break points are estimated simultaneously with the regression parameters via minimization of the residual sum of squares. Bai and Perron (1998) establish the consistency and the limiting distribution of the resulting break point fractions. They also propose a sequential procedure for selecting the number of break points in the sample based on various F-statistics for parameter constancy.

While not the only possible form for structural instability, the model with the discrete shifts at multiple unknown break points has some appeal in macroeconometric applications because it captures the case where relationships change due to changes in policy regime or exogenous shifts. However, since Bai and Perron's (1998) analysis is predicated on the assumption that

[1] See *inter alia* Andrews and Fair (1988), Ghysels and Hall (1990), Andrews (1993), Sowell (1996) and Hall and Sen (1999).

[2] Bai and Perron's (1998) paper also contributes to the literature in statistics on change point estimation in time series. See *inter alia* Picard (1985), Hawkins (1986), Bhattacharya (1987), Yao (1987) and Bai (1994).

all explanatory variables are exogenous, their methods can not be applied to the types of linear macroeconometric models mentioned above.

In this paper, we extend Bai and Perron's (1998) framework to linear models estimated via Two Stage Least Squares (2SLS) and thereby provide a methodology for estimating linear models with endogenous regressors that exhibit discrete shifts in the parameters at multiple unknown points in the sample. Within our framework, the break points are estimated simultaneously with the regression parameters via minimization of the residual sum of squares on the second step of the 2SLS estimation. We establish the consistency of the resulting break point fractions. We show that the various F-statistics for testing parameter constancy based on the 2SLS estimator have the same limiting distribution as the analogous statistics for OLS considered by Bai and Perron (1998). This allows us to extend Bai and Perron's (1998) sequential procedure for selecting the number of break points to the 2SLS setting.

As can be seen from the above summary, our focus is on the stability of the parameters in the second stage regression or, in other words, in the structural equation of interest. However to implement 2SLS, it is necessary in the first stage regression to estimate the reduced form for the endogenous regressors in the structural equation of interest and this, of course, requires an assumption about the constancy or lack thereof of these reduced form parameters. In this paper, we establish the aforementioned results under two scenarios of interest, namely: (i) the parameters in the first stage regression are constant; (ii) the parameters in the first stage regression are subject to discrete shifts within the sample period and the locations of these shifts are estimated *a priori* via a data-based method that satisfies certain conditions. The latter conditions allow the case in which the location of the instability is estimated via an application of Bai and Perron's (1998) methods to the appropriate reduced form equations on an equation by equation basis.

To illustrate our methods, we consider the stability of the New Keynesian Phillips curve (NKPC) estimated using quarterly data for the US over the period 1968.3-2001.4. The NKPC is of considerable theoretical importance in monetary policy analysis and its estimation has received considerable attention in the literature. Zhang, Osborn, and Kim (2007) observe that empirical studies of the NKPC often reach conflicting conclusions about the importance of key variables in the determination of inflation, and argue this may be due to neglected parameter variation. Zhang, Osborn, and Kim (2007) argue that changes in monetary policy regimes may cause changes in the parameters of the NKPC; if true, this would mean that the parameters of the NKPC would exhibit discrete shifts at potentially multiple points in the sample. Zhang, Osborn, and Kim (2007) investigate this issue using a methodology based on uncovering break points in the sample via the maximization of Wald statistics for parameter change associated with 2SLS estimation. However, while their methodology has an intuitive appeal, there is no theoretical justification for their methods as they note; it is, therefore, unclear exactly how to interpret their results. In contrast, our methods can be applied to this model under plausible assumptions about the data. Our analysis indicates that there are shifts in the parameters of both the appropriate reduced forms and also in the NKPC itself.

It is useful to compare our results to two other recent extensions of Bai and Perron's (1998) framework. Qu and Perron (2007) extend Bai and Perron's (1998) framework to systems of regression equations and consider the case in which estimation and inference are based on quasi-maximum likelihood techniques under normality. Perron and Qu (2006) consider the case of a regression equation in which the least squares estimation imposes cross-regime restrictions, such as the equality of parameters in two non-adjacent regimes. While both these papers expand the set of available techniques in important ways, both sets of results are predicated on the assumption that the explanatory variables are uncorrelated with the error(s). To our knowledge, our paper is the first to consider estimation and inference about multiple structural changes in a linear model with endogenous regressors.

An outline of the paper is as follows. Section 2 lays out the model, illustrates it via the NKPC and also explains details of the estimation. Section 3 presents results on the limiting behaviour of the break fraction estimators associated with the 2SLS estimation of the structural equation of interest. It is shown that the break fraction estimators are consistent and deviate from the true break fractions by a term of large order in probability T^{-1}, where T is the sample size. The import of this result is that inference regarding the parameters of the structural equation can be conducted as if the the true break fractions are known *a priori*. In the remainder of the paper, we consider the limiting behaviour of the 2SLS estimator and various associated inference procedures. Section 4 presents the limiting distribution of the 2SLS estimator. Section 5 presents the limiting distributions of the various F-statistics. The simulation evidence is reported in Section 6. Section 7 presents our empirical application and some concluding remarks are offered in Section 8. All proofs are relegated to a mathematical appendix.

2 The Model and The Estimation

2.1 The model

We consider the case in which the equation of interest is a multiple linear regression model with m breaks (*i.e.* $m+1$ regimes), that is

$$y_t \;=\; x_t'\beta_{x,i}^0 \,+\, z_{1,t}'\beta_{z_1,i}^0 \,+\, u_t, \qquad i = 1, ..., m+1, \qquad t = T_{i-1}^0 + 1, ..., T_i^0 \tag{1}$$

where $T_0^0 = 0$ and $T_{m+1}^0 = T$. In this model, y_t is the dependent variable, x_t is a $p_1 \times 1$ vector of explanatory variables that are correlated with the error u_t and $z_{1,t}$ is a $p_2 \times 1$ vector of explanatory variables that are uncorrelated with u_t and includes the intercept. We define $p = p_1 + p_2$. The error term, u_t, is assumed to have a mean of zero.

Following the convention in the literature, we index the break points $\{T_i^0\}$ by break fractions $\{\lambda_i^0\}$. These break fractions must satisfy the following:[3]

Assumption 1 $T_i^0 = [T\lambda_i^0]$, *where* $0 < \lambda_1^0 < ... < \lambda_m^0 < 1$.

Assumption 1 requires the break points to be asymptotically distinct.

In view of the correlation between x_t and u_t, OLS estimation of (1) would yield inconsistent estimators of the regression parameters. We therefore consider the case in which (1) is estimated via 2SLS. To implement 2SLS, it is necessary to specifiy the reduced form for x. As noted in the introduction, we consider two scenarios: (i) the reduced form for x_t is structurally stable; (ii) the reduced form for x_t exhibits parameter variation. We elaborate on these two scenarios in turn.

Scenario (i): stable reduced form.

The reduced form for x_t is assumed to be as follows:

$$x_t' \;=\; z_t'\Delta_0 \,+\, v_t' \tag{2}$$

where $z_t = (z_{t,1}, z_{t,2}, ..., z_{t,q})'$ is a $q \times 1$ vector of instruments that is uncorrelated with both u_t and v_t, $\Delta_0 = (\delta_{1,0}, \delta_{2,0}, ..., \delta_{p_1,0})$ with dimension $q \times p_1$ and each $\delta_{j,0}$ for $j = 1, ..., p_1$ has dimension $q \times 1$. We assume that z_t contains $z_{1,t}$. Under the assumption that $E[{u_t}^2|z_t] = \sigma^2$, the optimal IV estimator is the 2SLS estimator.[4] Our analysis is confined to the 2SLS estimator,

[3] $[\cdot]$ denotes the integer part of the quantity in the brackets.

[4] See, for example, Hall (2005)[p.44].

although we wish to emphasize that the aforementioned conditional homoscedasticity restriction is only imposed in certain parts of the analysis. $\diamond$.

Scenario (ii): unstable reduced form.

The reduced form for x_t is:

$$x_t^{'} \;=\; z_t^{'}\Delta_0^{(i)} \;+\; v_t^{'}, \qquad i = 1, 2, \ldots, h+1, \qquad t = T_{i-1}^* + 1, \ldots, T_i^* \tag{3}$$

where $T_0^* = 0$ and $T_{h+1}^* = T$. The points $\{T_i^*\}$ are assumed to be generated as follows.

Assumption 2 $T_i^* = [T\pi_i^0]$, *where* $0 < \pi_1^0 < \ldots < \pi_h^0 < 1$.

Note that the break fractions $\{\pi_i^0\}$ may or may not coincide with $\{\lambda_i^0\}$. Let $\pi^0 = [\pi_1^0, \pi_2^0, \ldots, \pi_h^0]'$. Within our analysis, it is assumed that the break points in the reduced form are estimated prior to estimation of the structural equation in (1). For our analysis to go through, the estimated break fractions in the reduced form must satisfy certain conditions that are detailed below; these conditions would hold, for instance, if Bai and Perron's (1998) methodology is applied equation by equation to the reduced form.

Equation (3) can be re-written as follows

$$x_t^{'} \;=\; \tilde{z}_t(\pi^0)^{'}\Theta_0 \;+\; v_t^{'}, \qquad t = 1, 2, \ldots, T \tag{4}$$

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 $\mathcal{I}\{t/T \in (0, \pi_1^0]\}$, $h+1^{th}$ element $\mathcal{I}\{t/T \in (\pi_h^0, 1]\}$, k^{th} element $\mathcal{I}\{t/T \in (\pi_{k-1}^0, \pi_k^0]\}$ for $k = 1, 2, \ldots, h$ and $\mathcal{I}\{\cdot\}$ is an indicator variable that takes the value one if the event in the curly brackets occurs. Notice that (4) fits the generic constant parameter form of (2). $\diamond$

To illustrate the potential interest in our framework, we consider the case of the NKPC. For ease of exposition, it suffices here to consider the following stylized version of the NKPC,

$$inf_t \;=\; c_0 \;+\; \alpha_f inf_{t+1|t}^e \;+\; \alpha_b inf_{t-1} \;+\; \alpha_{og} og_t \;+\; u_t \tag{5}$$

where inf_t is inflation in (time) period t, $inf_{t+1|t}^e$ denotes expected inflation in period $t+1$ given information available in period t, og_t is the output gap in period t, u_t is an unobserved error term and $\theta = (c_0, \alpha_f, \alpha_b, \alpha_y)'$ are unknown parameters. The variables $inf_{t+1|t}^e$ and og_t are anticipated to be correlated with the error u_t, and so (5) is commonly estimated via IV; *e.g.*

see Zhang, Osborn, and Kim (2007) and the references therein. Suitable instruments must be both uncorrelated with u_t and correlated with $inf^e_{t+1|t}$ and og_t. In this context, the instrument vector z_t commonly includes such variables as lagged values of expected inflation, the output gap, the short-term interest rate, unemployment, money growth rate and inflation. This model fits within our framework with (5) as the structural equation of interest provided the reduced forms for $inf^e_{t+1|t}$ and og_t are assumed to be given by either (2) or (3). We return to this example in Section 7.

2.2 The estimation

To describe the estimation of the model, it is assumed that the number of break points m is known but their location is not. Therefore the researcher must estimate both the break points and regression parameters. This estimation proceeds as follows. On the first stage, the reduced form for x_t is estimated via OLS using - as appropriate - either (2) or a version of (4) with estimated break fractions substituted for π^0. Let $\hat{x}_t$ denote the resulting predicted value for x_t. The second stage of the 2SLS estimation is itself divided into a number of steps because of the need to estimate both the break points and the regression parameters. The first step of the second stage is to estimate the model

$$y_t \;=\; \hat{x}_t^{'}\beta^*_{x,i} \,+\, z_{1,t}'\beta^*_{z_1,i} \,+\, \tilde{u}_t, \quad i = 1, ..., m+1; \quad t = T_{i-1}+1, ..., T_i \tag{6}$$

via OLS for each possible m-partition of the sample, denoted by $\{T_j\}_{j=1}^m$, such that $T_i - T_{i-1} \geq q$. Letting $\beta_i^{*\prime} = (\beta^{*}_{x,i}{}', \beta^{*}_{z_1,i}{}')'$, the resulting estimates of $\beta^* = (\beta_1^{*\prime}, \beta_2^{*\prime}, ..., \beta^{*}_{m+1}{}')'$ are obtained by minimizing the sum of squares of the residuals

$$S_T(T_1, ..., T_m) = \sum_{i=1}^{m+1} \sum_{t=T_{i-1}+1}^{T_i} (y_t - \hat{x}_t'\beta_{x,i} - z_{1,t}'\beta_{z_1,i})^2 \tag{7}$$

with respect to $\beta = (\beta_1{}', \beta_2{}', ..., \beta_{m+1}{}')'$. We denote these estimators by $\hat{\beta}(\{T_i\}_{i=1}^m)$.

The second step of the second stage involves constructing the minimized sum of squares associated with (6) for each partition, that is

$$S_T(T_1, ..., T_m;\; \hat{\beta}(\{T_i\}_{i=1}^m) \;=\; \sum_{i=1}^{m+1} \sum_{t=T_{i-1}+1}^{T_i} (y_t - \hat{x}_t'\beta_i - z_{1,t}'\beta_{z_1,i})^2\Big|_{\beta=\hat{\beta}(\{T_i\}_{i=1}^m)} \tag{8}$$

The estimates of the break points, $(\hat{T}_1, ..., \hat{T}_m)$, are defined as

$$(\hat{T}_1, ..., \hat{T}_m) = \arg \min_{T_1,...,T_m} S_T(T_1, ..., T_m;\ \hat{\beta}(\{T_i\}_{i=1}^m)) \tag{9}$$

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

3 Limiting behaviour of the break fraction estimators

In this section we analyze the limiting behaviour of the break point fraction estimators $\{\hat{\lambda}_i = \hat{T}_i/T\}$. Two properties are established: consistency and that the estimated break fractions deviate from the true break fractions by an $O_p(T^{-1})$ term. These results are established for both the scenarios regarding the parameters of the reduced form for x_t described in Section 2. We take each of these scenarios in turn.

3.1 Stable reduced form

In this case, the predicted value for x_t is given by

$$\hat{x}_t' \ = \ z_t{}'\hat{\Delta}_T \ = \ z_t{}'(\sum_{t=1}^{T} z_t z_t{}')^{-1} \sum_{t=1}^{T} z_t x_t{}' \tag{10}$$

To facilitate the analysis of this version of the model, we impose the following conditions.

Assumption 3 *Let $b_t = (u_t, v_t')'$ and $\mathcal{F} = \sigma - field\{\ldots, z_{t-1}, z_t, \ldots, b_{t-2}, b_{t-1}\}$. Assume b_t is a martingale difference relative to $\{\mathcal{F}_t\}$ and $sup_t E[\|b_t\|^4] < \infty$.*

Assumption 4 *$rank\{\,[\Delta_0, \Pi]\,\} = p$ 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 5 *There exists an $l_0 > 0$ such that for all $l > l_0$, the minimum eigenvalues of $A_{il} = (1/l)\sum_{t=T_i^0+1}^{T_i^0+l} z_t z_t{}'$ and of $A_{il}^* = (1/l)\sum_{t=T_i^0-l}^{T_i^0} z_t z_t{}'$ are bounded away from zero for all $i = 1, ..., m+1$.*

Assumption 6 *$T^{-1}\sum_{t=1}^{[Tr]} z_t z_t' \xrightarrow{p} Q_{ZZ}(r)$ uniformly in $r \in [0,1]$ where $Q_{ZZ}(r)$ is positive definite for any $r > 0$ and strictly increasing in r.*

Assumption 7 *The minimization in (9) is over all partitions* $(T_1, ..., T_m)$ *such that* $T_i - T_{i-1} > \epsilon T$ *for some* $\epsilon > 0$ *and* $\epsilon < inf_i(\lambda^0_{i+1} - \lambda^0_i)$.

A few comments on these assumptions are in order. Assumption 3 includes the restrictions that b_t is a serially uncorrelated process, and hence the errors in both the structural equation and reduced form exhbit this property. This assumption also includes the restriction that $E[z_t b_t'] = 0_{q\times(p_1+1)}$ which implies both the implicit population moment condition in 2SLS is valid - that is $E[z_t u_t] = 0$ - and also that the conditional mean of the reduced form is correctly specified. However, note that this assumption does allow z_t to contain lagged values of y_t. Assumption 4 implies the standard rank condition for identification in IV estimation in the linear regression model[5] because Assumptions 3, 4 and 6 together imply that

$$T^{-1}\sum_{t=1}^{[Tr]} z_t[x_t', z_{1,t}'] \;\Rightarrow\; Q_{ZZ}(r)[\Delta_0, \Pi] \;=\; Q_{Z,[X,Z_1]}(r) \text{ uniformly in } r \in [0,1]$$

where $Q_{Z,[X,Z_1]}(r)$ has rank equal to p for any $r > 0$. Assumption 5 requires that there be enough observations near the true break points so that they can be identified. This condition is analagous to Bai and Perron's (1998) Assumption A2 and the interested reader is refered to this source for further discussion of this condition. Assumption 7 requires that each segment considered in the minimization contains a positive fraction of the sample asymptotically; in practice ϵ is chosen to be small in the hope that the last part of the assumption is valid.

The proof strategy for consistency is identical to that used by Bai and Perron (1998) in their proof of the corresponding results for OLS estimators. The proof builds from the following two properties of the error sum of squares on the second stage of the 2SLS esimation.

- Since the 2SLS estimators minimize the error sum of squares in (7), it follows that

$$(1/T)\sum_{t=1}^{T}\hat{u}_t^2 \;\le\; (1/T)\sum_{t=1}^{T}\tilde{u}_t^2 \tag{11}$$

  where $\hat{u}_t = y_t - \hat{x}_t'\hat{\beta}_{x,j} - z_{1,t}'\hat{\beta}_{z_1,j}$ denotes the estimated residuals for $t \in [\hat{T}_{j-1}+1, \hat{T}_j]$ in the second stage regression of 2SLS estimation procedure and $\tilde{u}_t = y_t - \hat{x}_t'\beta^0_{x,i} - z_{1,t}'\beta^0_{z_1,i}$ denotes the corresponding residuals evaluated at the true parameter value for $t \in [T^0_{i-1}+1, T^0_i]$.

- Using $d_t = \tilde{u}_t - \hat{u}_t = \hat{x}_t'(\hat{\beta}_{x,j} - \beta^0_{x,i}) - z_{1,t}'(\hat{\beta}_{z_1,j} - \beta^0_{z_1,i})$ over $t \in [\hat{T}_{j-1}+1, \hat{T}_j] \cap [T^0_{i-1}+1, T^0_i]$,

[5] See *e.g.* Hall (2005)[p.35].

it follows that

$$T^{-1}\sum_{t=1}^{T}\hat{u}_t^2 \;=\; T^{-1}\sum_{t=1}^{T}\tilde{u}_t^2 \,+\, T^{-1}\sum_{t=1}^{T}{d_t}^2 \,-\, 2T^{-1}\sum_{t=1}^{T}\tilde{u}_t d_t \tag{12}$$

Consistency is established by proving that if at least one of the estimated break fractions does not converge in probability to a true break fraction then the results in (11)-(12) contradict each other. This conflict is established using the results in the following lemma.

Lemma 1 *Let y_t be generated by (1), x_t be generated by (2), $\hat{x}_t$ be generated by (10) and Assumptions 1, 3-7 hold.*

(i) $T^{-1}\sum_{t=1}^{T}\tilde{u}_t d_t = o_p(1)$.

(ii) If $\hat{\lambda}_j \overset{p}{\nrightarrow} \lambda_j^0$ *for some* j*, then*

$$\limsup_{T\to\infty} P\left(T^{-1}\sum_{t=1}^{T}{d_t}^2 > C\left\{\,\|\Delta_0(\beta^0_{x,j} \,-\, \beta^0_{x,j+1})\|^2 \,+\, \|\beta^0_{z_1,j} \,-\, \beta^0_{z_1,j+1}\|^2\,\right\} \,+\, \xi_T\right) > \bar{\epsilon}$$

for some $C > 0$ *and* $\bar{\epsilon} > 0$*, where* $\xi_T = o_p(1)$.

Using (11)-(12) and Lemma 1, consistency is established along the lines anticipated above.

Theorem 1 *Let y_t be generated by (1), x_t be generated by (2), $\hat{x}_t$ be generated by (10) and Assumptions 1, 3-7 hold, then $\hat{\lambda}_j \xrightarrow{p} \lambda_j^0$ for all $j = 1, 2, ..., m$.*

For the development of inference procedures for determining the number of breaks, it is important to know not only that the break fraction estimators are consistent but also the order of magnitude of their deviation from the true break fraction. This is established in the following theorem.

Theorem 2 *Let y_t be generated by (1), x_t be generated by (2), $\hat{x}_t$ be generated by (10) and Assumptions 1, 3-7 hold then, for every $\eta > 0$, there exists C such that for all large T, $P(T|\hat{\lambda}_j - \lambda_j^0| > C) < \eta$, for $j = 1, ..., m$.*

Therefore, the break fraction estimators deviate from the true break fractions by a term of order in probability T^{-1}.

3.2 Unstable reduced form

Recall that the reduced form exhibits discrete parameter changes at unknown points in the sample and these points are indexed by the break fraction vector, π^0. We suppose that π^0 is estimated by $\hat{\pi}$ and that these estimated break fractions satisfy the following condition.

Assumption 8 $\hat{\pi} \;=\; \pi^0 \,+\, O_p(T^{-1})$

Note that Assumption 8 implies $\hat{\pi}$ is consistent for π^0 and $T(\hat{\pi} - \pi^0)$ is bounded in probability. Such an estimator might be obtained by applying Bai and Perron (1998)'s methodology equation by equation and then pooling the resulting estimates of the break fractions. For our purposes, it only matters that Assumption 8 holds and not how $\hat{\pi}$ is obtained. The latter is, of course, a matter of practical importance but we do not address it here.

These estimated breaks are imposed on the the reduced form for x_t. Let $\hat{\Theta}_T$ be the OLS estimator of Θ_0 from the model

$$x_t' \;=\; \tilde{z}_t(\hat{\pi})'\Theta_0 + error \qquad t = 1, 2, \cdots, T \tag{13}$$

where $\tilde{z}_t(\hat{\pi})$ is defined analogously to $\tilde{z}_t(\pi^0)$, and now define $\hat{x}_t$ to be

$$\hat{x}_t' \;=\; \tilde{z}_t(\hat{\pi})'\hat{\Theta}_T \;=\; \tilde{z}_t(\hat{\pi})'\{\sum_{t=1}^{T} \tilde{z}_t(\hat{\pi})\tilde{z}_t(\hat{\pi})'\}^{-1} \sum_{t=1}^{T} \tilde{z}_t(\hat{\pi})x_t' \tag{14}$$

For the analysis in the case, the regularity conditions need to be altered. Assumption 4 is replaced by:

Assumption 9 *rank* $\left\{ \left[\Delta_0^{(i)}, \Pi\right] \right\} = p$ *for* $i = 1, 2, \cdots, h+1$ *and* Π *is defined in Assumption 4.*

It is also necessary to modify Assumption 7.

Assumption 10 *The minimization in (9) is over all partitions* $(T_1, ..., T_m)$ *such that* $T_i - T_{i-1} > \epsilon T$ *for some* $\epsilon > 0$ *and* $\epsilon < inf_i(\lambda_{i+1}^0 - \lambda_i^0)$ *and* $\epsilon < inf_j(\pi_{j+1}^0 - \pi_j^0)$.

The following theorem establishes the consistency of the break fraction estimators.

Theorem 3 *If Assumptions 1-3, 5-10 hold,* y_t *is generated via (1),* x_t *is generated via (4) and* $\hat{x}_t$ *is calculated via (14), then*

$$\hat{\lambda}_j \xrightarrow{p} \lambda_j^0 \qquad \text{for all } j = 1, 2, \cdots, m.$$

In order to extend Theorem 2, we impose one final condition.

Assumption 11 *There exists an* $l_* > 0$ *such that for all* $l > l_*$*, the minimum eigenvalues of* $B_{il} = (1/l)\sum_{t=T_i^*+1}^{T_i^*+l} z_t z_t'$ *and of* $B_{il}^* = (1/l)\sum_{t=T_i^*-l}^{T_i^*} z_t z_t'$ *are bounded away from zero for all* $i = 1, ..., h+1$.

Assumption 11 is similar to Asssumption 5 above but refers to the break points in the reduced form. The order in probability of the estimated break fractions is given in the following theorem.

Theorem 4 *If Assumptions 1-3, 5-11 hold,* y_t *is generated via (1),* x_t *is generated via (4) and* $\hat{x}_t$ *is calculated via (14), then, for every* $\eta > 0$*, there exists* C *such that for all large* T*,* $P(T|\hat{\lambda}_j - \lambda_j^0| > C) < \eta$*, for* $j = 1, ..., m$.

3.3 Discussion

At this stage, it is useful to comment on the nature of the foregoing analysis. First consider the case where the reduced form is structurally stable. In this case, Theorems 1-2 establish that the break fraction estimators, $\{\hat{\lambda}_j\}$, are consistent and $\hat{\lambda}_j - \lambda_j^0 = O_p(T^{-1})$. Now consider the case where the reduced form exhibits parameter variation. If the location of the breaks in the reduced form are known *a priori* then, as noted above, the reduced form can be re-written as a structurally stable regression equation involving the augmented parameter vector.[6] Therefore, in this case, the limiting behaviour of the break fraction estimators associated with the structural equation is covered by Theorems 1-2. However, in most cases, the locations of the breaks in the reduced form are unknown and so must be estimated *a priori*. In this case, Theorems 3-4 provide conditions on the estimators of the reduced form break fractions, $\{\hat{\pi}_i\}$, under which the break fraction estimators associated with the structural equation, $\{\hat{\lambda}_j\}$, are consistent and $\hat{\lambda}_j - \lambda_j^0 = O_p(T^{-1})$.

Of the scenarios described above, the most empirically relevant is likely to be the one involving estimation of break fractions in both reduced form and structural equations. Under our assumptions, the estimators of the break fractions in both reduced form and structural equations converge at rate T to the true break fractions. It emerges below that this rate is sufficiently fast that the estimation of the break fractions can be ignored in the asymptotic analysis of the

[6]See equation (4).

2SLS estimators and its associated statistics.[7] In other words, for the purposes of the asymptotic analysis of the 2SLS estimator and its associated statistics, we can essentially proceed as if the break fractions in both equations are known. Since, as noted above, the reduced form with known break points can be rewritten as a constant parameter regression model, we focus exclusively for the remainder of the paper on the case in which the reduced form is structurally stable. The analagous results for the model with parameter variation in the reduced form can be deduced from the results presented with an appropriate redefinition of the regressor vector in the reduced form.

4 The limiting distribution of the 2SLS estimators

Once the break fractions are estimated, it is clearly desirable to perform inference about the structural parameters $\{\beta_i^0\}$. If the break fractions are known *a priori* then standard arguments can be employed to show the root T asymptotic normality of the 2SLS estimator. Since the estimated break fractions converge at rate T, this standard asymptotic distribution theory can be extended to the 2SLS estimates based on the estimated break fractions.

Theorem 5 *Let y_t be generated by (1), x_t be generated by (2), $\hat{x}_t$ be generated by (10) and Assumptions 3-6 hold, then*

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

where $\beta^0 = [\beta_1^{0\prime}, \beta_2^{0\prime}, \ldots, \beta_{h+1}^{0\ \prime}]'$, $\beta_i^0 = [\beta_{x,i}^{0\ \prime}, \beta_{z_1,i}^{0\ \prime}]'$,

$$\begin{aligned} V_\beta &= \begin{pmatrix} V_\beta^{(1,1)} & \cdots & V_\beta^{(1,m+1)} \\ \vdots & \ddots & \vdots \\ V_\beta^{(m+1,1)} & \cdots & V_\beta^{(m+1,m+1)} \end{pmatrix} \\ V_\beta^{(i,j)} &= R_i S_{(i,j)} R_j', \quad \textit{for } i,j = 1,2,\ldots m+1 \\ R_i &= \left(A(1) Q_{ZZ}(1)^{-1} Q_i Q_{ZZ}(1)^{-1} A(1)'\right)^{-1} A(1) Q_{ZZ}(1)^{-1} \end{aligned}$$

and $Q_i = Q_{ZZ}(\lambda_i^0) - Q_{ZZ}(\lambda_{i-1}^0)$, $A(r)' = [Q_{ZX}(r), Q_{Z_1Z}(r)']$, $Q_{Z_1Z}(r)$ is the probability limit of $T^{-1}\sum_{t=1}^{[Tr]} z_{1,t} z_t'$ (defined in Assumption 6), $S_{(i,j)} = lim_{T\to\infty} Cov[T^{-1/2}\sum_{i_0} z_t \tilde{u}_t, T^{-1/2}\sum_{j_0} z_t \tilde{u}_t]$, $\sum_{i_0}$ denotes the summation over $t = [T\lambda_{i-1}^0]+1, \ldots [T\lambda_i^0]$, and we set $\lambda_0^0 = 0$, $\lambda_{m+1}^0 = 1$.

[7] A similar finding is reported by Bai and Perron (1998) in their analysis of OLS estimators.

Note that $S_{(i,j)}$ is non-zero in general because the first stage regression pools observations across regimes and this creates a connection between the aforementioned sums from different regimes. However, if the reduced form is also unstable then the connection across regimes is broken in one leading case. If the breaks in the structural equation also occur in the reduced form then the predictions are only based on the observations in the sub-sample in question and so V_β is block diagonal. Specifically, if $h \geq m$ and $\lambda_i^0 = \pi_j^0$ for some j for each i then

$$V_\beta \; = \; diag(\tilde{V}_\beta^{(1,1)}, \tilde{V}_\beta^{(2,2)}, \ldots, \tilde{V}_\beta^{(m+1,m+1)}) \tag{15}$$

where $\tilde{V}_\beta^{(i,i)} \; = \; \tilde{R}_i \tilde{S}_{(i,i)} \tilde{R}_i'$, $\tilde{R}_i \; = \; \left(A_i Q_i^{-1} A_i'\right)^{-1} A_i Q_i^{-1}$, $A_i \; = \; A(\lambda_i^0) - A(\lambda_{i-1}^0)$, and $\tilde{S}_{(i,i)} \; = \; lim_{T\to\infty} T^{-1} \sum_{i_0} Var[z_t u_t]$. Notice that $\tilde{V}_\beta^{(i,i)}$ is just the variance of the 2SLS estimator based on the i^{th} sub-sample allowing potentially for breaks in the reduced form within that sub-sample.

5 Test statistics for multiple breaks

The sup-F type test of no structural break ($m = 0$) versus the alternative hypothesis that there is $m = 1$ break has been considered by Andrews (1993). Bai and Perron (1998) generalize Andrew's sup-F type test to the hypothesis $m = k$ for linear models estimated via OLS. In this section, we extend Bai and Perron's results to linear models estimated via $2SLS$.

For this part of the analysis, we impose the following restrictions.

Assumption 12 *(i)* $T^{-1} \sum_{t=1}^{[Tr]} z_t z_t' \xrightarrow{p} r Q_{ZZ}$ *uniformly in* $r \in [0,1]$ *where* Q_{ZZ} *is a positive definite matrix of constants;*

(ii) the conditional variance of the errors is independent of t*, that is*

$$Var \left[\left(\begin{array}{c} u_t \\ v_t \end{array} \right) \Big| z_t \right] \; = \; \Omega \; = \; \left[\begin{array}{cc} \sigma^2 & \gamma' \\ \gamma & \Sigma \end{array} \right]$$

where Ω *is a constant, positive definite matrix,* σ^2 *is a scalar and* Σ *is a* $p_1 \times p_1$ *matrix;*

The restrictions in Assumption 12 are analogous to that imposed by Bai and Perron (1998) in their Assumptions A8 and A9 which underpin their analysis of various F-statistics for testing for multiple breaks within the OLS framework.[8]

[8] Although note that the conditional variance restriction in Assumption 12 involves both u_t and v_t whereas Bai and Perron (1998) need only restrict the conditional variance of u_t.

Assumptions 3 and 12 together ensure that a uniform version of the multivariate functional central limit theorem in de Jong and Davidson (2000) holds:

$$T^{-1/2}\sum_{t=1}^{[Tr]}\begin{pmatrix} u_t \\ v_t \end{pmatrix} \otimes z_t \implies (\Omega^{1/2}\otimes Q_{ZZ}^{1/2})B_n(r) \tag{16}$$

where $B_n(r)$ is a $n\times 1$ standard Brownian motion with $n = q\times(p_1+1)$ and "$\Longrightarrow$" denotes weak convergence in the space $D[0,1]$ under the skorohod metric.

The sup-F type test statistic can be defined as follows. Let $(T_1, ..., T_k)$ be a partition such that $T_i = [T\lambda_i]$ $(i = 1, ..., k)$. Define

$$F_T(\lambda_1, ..., \lambda_k; p) \;=\; \left\{\frac{T-(k+1)p}{kp}\right\}\left\{\frac{SSR_0 - SSR_k}{SSR_k}\right\} \tag{17}$$

where SSR_0 and SSR_k are the sum of squared residuals based on the fitted X under null and alternative hypothesis, respectively. Recall from Assumption 7 that the minimization is performed over partitions which are asymptotically large and the size of the partitions is controlled by ϵ, a non-negative constant. Accordingly, we define $\Lambda_\epsilon = \{(\lambda_1, ..., \lambda_k) : |\lambda_{i+1}-\lambda_i| \geq \epsilon, \lambda_1 \geq \epsilon, \lambda_k \leq 1-\epsilon\}$. Finally, the sup-F type test statistic is defined as

$$Sup - F_T(k; p) \;=\; Sup_{(\lambda_1,...,\lambda_k)\in\Lambda_\epsilon} F_T(\lambda_1, .., \lambda_k; p) \tag{18}$$

Theorem 6 *If the data are generated by (1)-(2) with $m = 0$, $\hat{x}_t$ is generated by (10) and Assumptions 1, 3-7 and 12 hold then $Sup - F_T(k;p) \Rightarrow Sup - F_{k,p} \equiv Sup_{(\lambda_1,...,\lambda_k)\in\Lambda_\epsilon} F(\lambda_1, .., \lambda_k; p)$ where*

$$F(\lambda_1, ..., \lambda_k; p) \;\;\equiv\;\; \frac{1}{kp}\sum_{i=1}^{k}\frac{||\lambda_{i+1}W_i - \lambda_i W_{i+1}||^2}{\lambda_i\lambda_{i+1}(\lambda_{i+1}-\lambda_i)}$$

where k is the number of break points under the alternative hypothesis, and $W_i \equiv B_p(\lambda_i)$.

We note that the limiting distribution in Theorem 6 is exactly the same as the one in Bai and Perron's (1998) analogous result for the sup-F test based on OLS estimators when the regressors are exogenous. Percentiles for this distribution can be found in Bai and Perron (1998)[Table I] for $\epsilon = 0.05$ and in Bai and Perron (2001) for other values of ϵ.

The $Sup - F_T(k; p)$ statistic is used to test the null hypothesis of structural stability against the k-break model, and so is designed for the case in which a particular choice of k is of interest. In many circumstances, a researcher is unlikely to know *a priori* the appropriate choice of k

for the alternative hypothesis. To circumvent this problem, Bai and Perron (1998) propose so called "Double Maximum tests" that combine information from the $Sup - F_T(k;p)$ statistics for different values of k running from one to some ceiling K. We consider here only the following example of Double Maximum test,[9]

$$UDmaxF_T(K;p) \; = \; \max_{1\leq k \leq K} \; \sup_{(\lambda_1,...,\lambda_k)\in\Lambda_\epsilon} F_T(\lambda_1,...,\lambda_k;p) \tag{19}$$

The limiting distribution of this statistic follows directly from Theorem 6.

Corollary 1 *Under the conditions of Theorem 6, it follows that*

$$UDmaxF_T(K;p) \implies \max_{1\leq k\leq K}\{Sup - F_{k,p}\}$$

Critical values for the limiting distribution in Corollary 1 are presented in Bai and Perron (1998)[Table 1] for $\epsilon = 0.05$ and in Bai and Perron (2001) for other values of ϵ.

The $Sup - F_T(k;p)$ and $UDmaxF_T(K;p)$ statistics are used to test the null hypothesis of no breaks. It is also of interest to develop statistics for testing the null hypothesis of l breaks against the alternative of $l+1$ breaks. Following Bai and Perron (1998), a suitable statistic can be constructed as follows. For the model with l breaks, the estimated break points, denoted by $\hat{T}_1,...,\hat{T}_l$, are obtained by a global minimization of the sum of the squared residuals as in (9). For the model with $l+1$ breaks, l of the breaks are fixed at $\hat{T}_1,...,\hat{T}_l$ and then the location of the $(l+1)^{th}$ break is chosen by minimizing the residual sum of squares. The test statistic is given by

$$F_T(l+1|l) = \max_{1\leq i\leq l+1}\{\frac{SSR_l(\hat{T}_1,...,\hat{T}_l) - \inf_{\tau\in\Lambda_{i,\eta}} SSR_{l+1}(\hat{T}_1,...,\hat{T}_{i-1},\tau,\hat{T}_i,...,\hat{T}_l)\}}{\hat{\sigma}_i^2}\} \tag{20}$$

where

$$\begin{aligned}
\hat{\sigma}_i^2 &= \sum_{t=\hat{T}_{i-1}+1}^{\hat{T}_i} (y_t - \hat{x}_t'\hat{\beta}_{x,i} - z_{1,t}'\hat{\beta}_{z_1,i})^2/(\hat{T}_i - \hat{T}_{i-1} - p) \\
\Lambda_{i,\eta} &= \{\tau : \hat{T}_{i-1} + (\hat{T}_i - \hat{T}_{i-1})\eta \leq \tau \leq \hat{T}_i - (\hat{T}_i - \hat{T}_{i-1})\eta\}
\end{aligned}$$

and $\hat{\beta}_i$ is the 2SLS estimator calculated using the sample $\hat{T}_{i-1}+1,\ldots,\hat{T}_i$ on the second stage.

The following theorem gives the limiting distribution of this statistic under the null hypothesis of l breaks.

[9] *UDmax* denotes Unweighted Double maximum. Bai and Perron (1998) also consider a *WDmax* statistic in which the the maximum is taken over weighted values of the $Sup - F_T(k;p)$ statistics. Analogous *WDmax* statistics can be developed within our framework, but for brevity we do not explore them here.

Theorem 7 *If the data are generated by (1)-(2) with* $m = l$, $\hat{x}_t$ *is generated by (10) and Assumptions 1, 3-7 and 12 hold then then* $\lim_{T\to\infty} P(F_T(l+1|l) \leq x) = G_{p,\eta}(x)^{l+1}$ *where* $G_{p,\eta}(x)$ *is the distribution function of* $\sup_{\eta\leq\mu\leq 1-\eta} \|W(\mu) - \mu W(1)\|^2/\mu(1-\mu)$ *and* $W(\mu) \equiv B_p(\mu)$.

Once again, the limiting behaviour of the test statistic is the same as that of the analogous statistic proposed by Bai and Perron (1998) for the OLS case. Critical values can be found in Bai and Perron (1998)[Table II] for the case in which calculated with $\eta = .05$ and in Bai and Perron (2001) for other values of η.

Following Bai and Perron (1998), the statistics described in this section can be used to determine the estimated number of breakpoints, $\hat{k}_T$ say, via the following sequential strategy. On the first step, use either $Sup - F_T(1;p)$ or $UDmaxF_T(K,p)$ to test the null hypothesis that there are no breaks. If this null is not rejected then $\hat{k}_T = 0$; else proceed to the next step. On the second step $F_T(2|1)$ is used to test the null hypothesis that there is only one break against the alternative hypothesis of two breaks. If $F_T(2|1)$ is insignificant then $\hat{k}_T = 1$; else proceed to the next step. On the l^{th} step $F_T(l+1|l)$ is used to test the null hypothesis that there are l breaks against the alternative hypothesis of $l+1$ breaks. If $F_T(l+1|l)$ is insignificant then $\hat{k}_T = l$; else proceed to the next step. This sequence is continued until some preset ceiling for the number of breaks, L say, is reached. If all statistics in the sequence are significant then the conclusion is that there are at least L breaks. We evaluate the finite sample performance of this strategy as part of the simulation study reported in the following section.

To conclude our discussion of these F-statistics, we return to the issue of the assumptions on the errors. Assumption 12 requires the errors to be homoscedastic and serially uncorrelated. It is, however, possible to relax this assumption to some extent as we now discuss. Suppose that it is assumed that a regime is characterized by both a change in the regression parameter vector and also a change in the conditional variance matrix of the errors, that is Ω in Assumption 12 is replaced by Ω_i for $t \in \left([T\lambda_{i-1}^0]+1, [T\lambda_i^0]\right)$. Since the calculation of $F_T(l+1|l)$ only involves sub-sample covariance matrix estimators, it follows that the limiting distribution of the test statistic is unaffected by heteroscedasticity of this type. It is therefore possible to use the the test statistics described above to develop a sequential strategy to determine the number of breaks for the case where the no break model is homoscedastic and the l break models involve a conditional error variance that is constant within a regime but varies across regimes.

6 Finite sample behaviour

In this section, we evaluate the finite sample behaviour of the various statistics discussed in the previous sections via a small simulation study. The simulation design involves models with zero, one or two breaks. Since our analysis of the break fractions is premised on the existence of a break, we begin by discussing the one break and two break models. We then conclude the sections by considering the behaviour of the test statistics in the no break model.

6.1 One break model

The data generating process for the structural equation is:

$$\begin{aligned} y_t &= [1, x_t]'\beta_1^0 + u_t, \qquad \text{for } t = 1, \ldots, [T/2] \\ &= [1, x_t]'\beta_2^0 + u_t, \qquad \text{for } t = [T/2]+1, \ldots, T \end{aligned} \tag{21}$$

The reduced form equation for the scalar variable x_t is:

$$x_t = z_t'\delta + v_t, \qquad \text{for } t = 1, \ldots, T \tag{22}$$

where δ is $q \times 1$. The errors are generated as follows: $(u_t, v_t)' \sim IN(0_{2\times 1}, \Omega)$ where the diagonal elements of Ω are equal to one and the off-diagonal elements are equal to 0.5. The instrumental variables, z_t, are generated via: $z_t \sim i.i.d\ N(0_{q\times 1}, I_q)$. The specific parameter values are as follows: (i) $T = 60, 120, 240, 480$; (ii) $(\beta_1^0, \beta_2^0) = ([1, 0.1]', [-1, -0.1]')$; (iii) $q = 2, 4, 8$; (iv) δ is chosen to yield the population $R^2 = 0.5$ for the regression in (22).[10] For each configuration, 1000 simulations are performed.

The results are presented in Tables 1-4. We first consider the behaviour of the break fraction estimator calculated under the assumption that there is only one break. Table 1 reports the proportion of the simulations in which $|\hat{\lambda}_1 - \lambda_1^0| \le c$ for $c = 0.01, 0.02, 0.03, 0.05, 0.1$. It can be seen that in the smallest sample size ($T = 60$) there is some dispersion but the proportions clearly increase with T and exhibit behaviour in line with the consistency result in Theorem 1. Table 2 reports the relative rejection frequencies of $Sup - F_T(k; 1)$ (for $k = 1, 2$), $UDmaxF_T(5; 1)$ and $F_T(l + 1|l)$ (for $l = 1, 2$) statistics where, in both cases the nominal size is 0.05. Notice that the

[10]For this model, $\delta = \sqrt{R^2/(q - q \times R^2)}$; see Hahn and Inoue (2002).

alternative hypothesis is true for the $Sup - F_T(k;1)$ and $UDmaxF_T(5;1)$ statistics and so these relative frequencies are empirical powers for this statistic. Whereas, for $l = 1$, the null hypothesis is correct for $F_T(l+1|l)$ and so the relative frequencies are the empirical size, and for $l = 2$, the null assumes more breaks than there actually are. Both $Sup - F_T(k;1)$ and $UDmaxF_T(5;1)$ reject 100% of the time. The $F_T(2|1)$ statistic is close to its nominal size; $F_T(3|2)$ tends to reject less frequently than the nominal size. Table 3 reports the results from using the sequential strategy based on these statistics that is described in Section 5 with a maximum number of breaks set equal to five. The results indicate that the strategy works well in each case. Table 4 reports the empirical coverage of the large sample confidence intervals based on the limiting distribution in Theorem 5, with all limiting covariances replaced by their empirical counterparts.[11] As can be seen, the empirical coverage is very close to the nominal level in all cases, and is within 3 simulation standard deviations of the nominal level for all confidence levels in all but the smallest sample size.

6.2 Two break model

The data generation process for the structural equation is:

$$\begin{aligned} y_t &= [1,\, x_t]'\beta_1^0 + u_t, & \text{for } t &= 1, \ldots, [T/3] \\ &= [1,\, x_t]'\beta_2^0 + u_t, & \text{for } t &= [T/3]+1, \ldots, [2T/3] \\ &= [1,\, x_t]'\beta_3^0 + u_t, & \text{for } t &= [2T/3]+1, \ldots, T \end{aligned}$$

Two choices for β^0 are considered: $(\beta_1^0, \beta_2^0, \beta_3^0) = (\,[1,\, 0.1]', [-1,\, -0.1]',\, [1,\, 0.1]'\,)$. All other aspects of the design are the same as the one break model.

The results are reported in Tables 5-9. Again, we begin by considering the performance of the estimated break fractions. Table 5 reveals that, as in the one break model, there is some dispersion in the estimates of the break fractions in the smallest sample size but nevertheless the empirical distribution of the break fraction estimator is evidently collapsing on the true fraction

[11] Within this model, it can be shown that $S_{i,i} = (\lambda_i^0 - \lambda_{i-1}^0)\left\{V_{1,1} + (1 + \lambda_{i-1}^0 - \lambda_i^0)\,[(\beta_i^{0'} \otimes I_q)V_{2,2}(\beta_i^0 \otimes I_q) + 2V_{1,2}(\beta_i^0 \otimes I_q)\,]\right\}$ and $S_{(i,j)} = -(\lambda_i^0 - \lambda_{i-1}^0)(\lambda_j^0 - \lambda_{j-1}^0)[V_{1,2}(\beta_j^0 \otimes I_q) + (\beta_i^{0'} \otimes I_q)V_{2,1} + (\beta_i^{0'} \otimes I_q) \times \; V_{2,2}(\beta_j^0 \otimes I_q)]$ where $V = \begin{pmatrix} V_{1,1} & V_{1,2} \\ V_{1,2}' & V_{2,2} \end{pmatrix}$ is the long-run covariance of $T^{-1/2}\sum_{t=1}^{T}(u_t, v_t')' \otimes z_t$, $V_{1,1}$ is $q \times q$ and $V_{2,2}$ is $qp_1 \times qp_1$. Consistent estimators of $S_{i,j}$ are constructed using these formulae in the obvious fashion.

as T increases. Table 6 reports the relative rejection frequencies of $Sup - F_T(k;1)$ (for $k = 1, 2$), $UDmaxF_T(5;1)$ and $F_T(l+1|l)$ (for $l = 1, 2$) statistics where, in both cases the nominal size is 0.05. As in the one break model, the statistics are applied with $k, l = 1, 2$. Notice that the alternative hypothesis is true for the $Sup - F_T(k;1)$, $UDmaxF_T(5;1)$ and $F_T(2|1)$ statistics and so these relative frequencies are empirical powers for this statistic. Whereas, the null hypothesis is correct for $F_T(3|2)$ and so the relative frequencies are the empirical size. From Table 6 it can be seen that, unlike the one break model, there is a difference in the power properties of the tests. While $Sup - F_T(k;2)$ and $UDmaxF_T(5;1)$ reject 100% of the time, $Sup - F_T(k;1)$ only rejects 74% power in the smallest sample size although it does reject 100% of the time in larger sample sizes. The test of one break against two ($F_T(2|1)$) also rejects 100% in every case. The test of two breaks against three ($F_T(3|2)$) is slightly undersized; this contrasts with the results for $F_T(2|1)$ in the one break model and likely reflects the smaller sub-sample sizes in the two break model. Table 7 reports the results using the sequential strategy for estimating the number of breaks. As would be expected given the power results, the sequential strategy starting with $Sup - F_T(k;1)$ has a marked tendency to under estimate the number of breaks in the smallest sample size. In contrast, the sequential strategy starting with $UDmaxF_T(5;1)$ works well at all sample sizes as it never underfits and picks the true order never less than 94% of the time. Tables 8-9 report the empirical coverage of the large sample confidence intervals based on the limiting distribution in Theorem 5. In the smallest sample size ($T = 60$), the coverage is lower than the nominal level and more than three simulation standard errors away from the nominal level; this can be explained by small sizes of the sub-samples in this case. However, the empirical coverage is within three simulation standard errors for all intervals at the other samples ($T = 120, 240, 480$) and very close to the nominal level in the larger samples.

6.3 No break model

The previous two designs involve cases where there is a change in the regression parameters of the structural equation. It is also of interest to explore how the test statistics perform in the case where there is no break and so the model is structurally constant. To this end, data are generated from (21) with $\beta_1^0 = \beta_2^0 = [1, 1]$. All other aspects of the design are the same as the one break model. Table 10 contains the empirical rejection frequencies for $Sup - F_T(k;1)$

$(k = 1, 2)$, $F_T(l+1|l)$ $(l = 1, 2)$ and $UDmaxF_T(5; 1)$ statistics. Note that within this design, the null hypothesis is correct for the $Sup - F_T(1; 1)$, $Sup - F_T(2; 1)$, and $UDmaxF_T(5, 1)$ statistics, and so the rejection frequency equals the empirical size. For $F_T(2|1)$ and $F_T(3|2)$ statistics, the null hypothesis involves more breaks than are present in the data. From Table 11, it can be seen that $Sup - F_T(1; 1)$, $Sup - F_T(2; 1)$, and $UDmaxF_T(5, 1)$ exhibit empirical size close to the nominal level of 0.05; both $F_T(2|1)$ and $F_T(3|2)$ reject less frequently than the size. Table 10 presents the empirical distribution of $\hat{k}_T$ based on the sequential strategies using $Sup - F_T(1; 1)$ and $UDmaxF_T(5, 1)$. Both strategies indicate that no breaks are present in nearly every case.

7 Application

In this section we use our methods to explore the stability of the New Keynesian Phillips curve (NKPC). Zhang, Osborn, and Kim (2007) report that the stylized version of the NKPC in (5) does not have serially uncorrelated errors as required by our Assumption 3, and so we follow their practice and include lagged values of $\Delta inf_t = inf_t - inf_{t-1}$ to remove this dynamic structure from the errors. Accordingly, our analysis is based on

$$inf_t \;=\; c_0 \,+\, \alpha_f inf^e_{t+1|t} \,+\, \alpha_b inf_{t-1} \,+\, \alpha_{og} og_t \,+\, \sum_{i=1}^{3} \alpha_i \Delta inf_{t-i} \,+\, u_t \tag{23}$$

The data is for the US and is quarterly spanning 1968.3-2001.4. The span of the data is slightly longer than Zhang, Osborn, and Kim (2007) but the definitions of the variables are the same and as follows: inf_t is the annualized quarterly growth rate of the GDP deflator, og_t is obtained from the estimates of potential GDP published by the Congressional Budget Office, $inf^e_{t+1|t}$ is the Greenbook one quarter ahead forecast of inflation prepared within the Fed.[12]

Both expected inflation and the output gap are taken to be endogenous and we model their reduced forms as

$$inf^e_{t+1|t} \;=\; z_t'\delta_1 \,+\, v_{1,t} \tag{24}$$

$$og_t \;=\; z_t'\delta_2 \,+\, v_{2,t} \tag{25}$$

where z_t contains all other explanatory variables on the righthand side of (23) along with the

[12] One interesting aspect of Zhang, Osborn, and Kim's (2007) study is that they employ various different inflation forecasts in their estimation. We focus here on just one of their choices for brevity.

first lagged value of each of the short term interest rate, the unemployment rate, and the growth rate of the money aggregate M2.

We first consider the stability of the reduced forms in (24)-(25) using Bai and Perron's (1998) methodology. [13] We assume that the maximum number of breaks is 5 and set $\epsilon = 0.1$. The results are reported in Table 11. First consider the reduced form for $inf^e_{t+1|t}$. There is clear evidence of parameter variation with all the sup-F statistics being significant at the 1% level. Using the sequential testing strategy, we identify two breaks: one at 1975.2 and the other at 1981.1. As a robustness check, we also use BIC to choose the break points and obtain the same estimates.[14] Now consider the reduced form for og_t. Again, there is evidence of parameter variation. The sequential strategy suggests a break at 1975.2. In contrast, BIC favours the model with no breaks. Given our purposes, it seems better to impose this break in our estimation of the reduced form.

We now consider the results for the NKPC. Given the evidence above, the predicted values of expected inflation are constructed allowing for breaks at 1975.2 and 1981.1, and the predicted value for the output gap is constructed allowing for a break at 1975.2. As with the reduced forms, we assume that the maximum number of breaks is 5 and set $\epsilon = 0.1$. The results from the 2SLS estimations of the NKPC are given in Table 12. As with the reduced forms, there is evidence of instability from the sup-F tests. Using the sequential strategy, we estimate there to be only one break located at 1975.1.[15] Parenthetically, we note that if the number of breaks is chosen by minimizing the BIC,

$$BIC(m) \ = \ ln[\min_{T_1,...,T_m} S_T(T_1,...,T_m;\ \hat{\beta}(\{T_i\}_{i=1}^m))/T] \ + \ m(p+1)ln(T)/T$$

then the estimated number is also one and the location is again 1975.1.

The estimated NKPC is as follows (omitting the error and with estimates to 2dp; standard errors in parentheses):

[13] These calculations are made using the code available from http://people.bu.edu/perron/code.html. All hypotheses are tested with F-statistics which are the OLS analogs of those discussed in the text; further details can be found in Bai and Perron (1998).

[14] For ease of presentation, we define the BIC criterion below for 2SLS; the appropriate modification for OLS is then obvious.

[15] We note that it was not possible to calculate the test of the four break model against the five break model because the location of the breaks in the four break model meant certain sub-samples in the five break model were too small.

for 1969.1-1975.1:

$$inf_t \;=\; -\underset{(0.09)}{4.45} + \underset{(0.16)}{0.52}\, inf^e_{t+1|t} + \underset{(0.15)}{1.48}\, inf_{t-1} + \underset{(0.03)}{0.39}\, og_t - \underset{(0.10)}{1.39}\, \Delta inf_{t-1} - \underset{(0.08)}{1.05}\, \Delta inf_{t-2} - \underset{(0.03)}{0.37}\, \Delta inf_{t-3}$$

for 1975.2-2001.4:

$$inf_t \;=\; -\underset{(0.17)}{0.27} + \underset{(0.24)}{0.69}\, inf^e_{t+1|t} + \underset{(0.21)}{0.33}\, inf_{t-1} + \underset{(0.19)}{0.11}\, og_t - \underset{(0.12)}{0.16}\, \Delta inf_{t-1} - \underset{(0.09)}{0.13}\, \Delta inf_{t-2} - \underset{(0.29)}{0.28}\, \Delta inf_{t-3}$$

Of particular interest are the coefficients on expected and lagged inflation as they reflect the degree to which policy is forward or backward looking respectively. One most striking difference between the two periods is in the coefficient on lagged inflation. Our results suggest that this variable plays a far weaker role in the post-1975.1 sample. However, one important caveat is the small size of the pre-1975.1 subsample.

It is interesting to note that our results closely match Zhang, Osborn, and Kim's (2007) findings with regard to both the number of breaks and the location of the break.[16] However, we cannot directly compare our estimates as Zhang, Osborn, and Kim (2007) do not report the specific estimates associated with this sample break.

8 Concluding Remarks

In this paper, we extend Bai and Perron's (1998) framework for multiple break testing to linear models estimated via Two Stage Least Squares (2SLS). Within our framework, the break points are estimated simultaneously with the regression parameters via minimization of the residual sum of squares on the second step of the 2SLS estimation. We establish the consistency of the resulting estimated break point fractions. We show that various F-statistics for structural instability based on the 2SLS estimator have the same limiting distribution as the analogous statistics for OLS considered by Bai and Perron (1998). This allows us to extend Bai and Perron's (1998) sequential procedure for selecting the number of break points to the 2SLS setting.

Our focus is on the stability of the parameters in the structural equation of interest. However to implement 2SLS, it is necessary in the first stage regression to estimate the reduced form for the endogenous regressors in the structural equation of interest and this, of course, requires an assumption about the constancy or lack thereof of these reduced form parameters. In this

[16] We note that with other choices of inflation forecast series, Zhang, Osborn, and Kim (2007) find evidence of breaks at other points in the sample.

paper, we establish the aforementioned results under two scenarios of interest, namely: (i) the parameters in the first stage regression are constant; (ii) the parameters in the first stage regression are subject to discrete shifts within the sample period and the locations of these shifts are estimated *a priori* via a data-based method that satisfies certain conditions. The latter conditions allow the case in which the location of the instability is estimated via an application of Bai and Perron's (1998) methods to the appropriate reduced form equations on an equation by equation basis. We have illustrated the empirical relevance of our framework via an application to the New Keynesian Phillips curve. Most empirical investigations of the NKPC assume the parameters are constant. However, our results indicate that if estimated over 1968-2001 then this relationship is not stable.

In practice, a researcher may also be interested in performing inference about the timing of the structural changes. Hall, Han, and Boldea (2007) provide a distribution theory for the break fraction estimators in the case where the reduced form regression parameters are structurally stable. The extension of this theory to the case in which the reduced form exhibits parameter variation is complicated by the potential dependence on the limiting distribution of the estimated break fractions in the structural equations on that of the estimated break fractions from the reduced form. This extension is work in progress.

In two recent papers, Perron and Qu extend Bai and Perron's (1998) framework in a number of interesting ways. Qu and Perron (2007) consider estimation and inference of multiple structural changes in systems of regression equations, and show that there are efficiency gains from estimation of the system rather than on an equation by equation basis. Perron and Qu (2006) show that there are also efficiency gains from imposing cross-regime restrictions, such as the equality of parameters in two non-adjacent regimes. It would be interesting to explore the potential for such efficiency gains within the context of our 2SLS framework; however, these extensions are beyond the scope of the current paper and are left to future research.

Mathematical Appendix

We begin with an item of terminology. We say that a matrix A, say, is a diagonal partition at $(T_1, T_2, \ldots T_m)$ of the $T \times k$ matrix W whose t^{th} row is $\hat{x}_t'$ if $A = diag(W_{T_1}, ..., W_{T_{m+1}})$ and $W_{T_i} = (\hat{x}_{T_{i-1}+1}, ..., \hat{x}_{T_i})'$.[17]

We write (6) for the true partition (so that $\beta_i^* = \beta_i^0$) as

$$Y \ = \ \bar{W}^0 \beta^0 \ + \ \tilde{U} \tag{26}$$

where $Y = (y_1, ..., y_T)'$, $\bar{W}^0$ is a diagonal partition of W at $(T_1^0, ..., T_{m+1}^0)$, $\tilde{U} = (\tilde{u}_1, ..., \tilde{u}_T)'$, and $\beta^0 = \beta^0(\{T_i^0\}_{i=1}^m) = (\beta_1^{0\prime}, \beta_2^{0\prime}, ..., \beta_{m+1}^{0\prime})'$ with $\beta_i^0 = (\beta_{i,1}^0, \beta_{i,2}^0, ..., \beta_{i,p}^0)'$. We also define: $\bar{W}^*$ to be a diagonal partition of W at $(\hat{T}_1, ..., \hat{T}_m)$; $Z = (z_1, ..., z_T)'$; $V = (v_1, ..., v_T)'$.

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

$$\|A\| = \sup_{x \neq 0} \|Ax\| / \|x\| \tag{27}$$

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

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

$$\|A\| \ \leq \ (trA'A)^{1/2} \tag{28}$$

- For a projection matrix P, we have

$$\|PA\| \leq \|A\| \tag{29}$$

- Let $A : R_1 \rightarrow R_2$ and $B : R_2 \rightarrow R_3$ be linear operators. Then we have[18]

$$\|BA\| \leq \|B\| \|A\| \tag{30}$$

[17] Note that $diag(.)$ stands for block diagonal here.

[18] See Ortega (1987)[p. 93-4].

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

To simplify the presentation, we prove all the desired results for the special case in which $\beta^0_{z_1,i} = 0_{p_2}$ and $z_{1,t}$ is omitted from the structural equation during estimation. It is easily verified that all the desired results extend to the model presented in the main text.

Proof of Lemma 1

Part (i):

Using the definition of d_t, it follows that, for $t \in [\hat{T}_{j-1}+1, \hat{T}_j]$,

$$\tilde{u}_t d_t \;=\; \tilde{u}_t \hat{x}_t' (\hat{\beta}_j - \beta_i^0) \;=\; \tilde{u}_t \hat{x}_t' \hat{\beta}_j \;-\; \tilde{u}_t \hat{x}_t' \beta_i^0$$

and hence that

$$\begin{aligned} \sum_{t=1}^{T} \tilde{u}_t d_t &= \sum_{t=1}^{T} \tilde{u}_t \hat{x}_t' \hat{\beta}(t,T) \;-\; \sum_{t=1}^{T} \tilde{u}_t \hat{x}_t' \beta^0(t,T) \\ &= \tilde{U}' \bar{W}^* \hat{\beta} - \tilde{U}' \bar{W}^0 \beta^0 \end{aligned} \tag{31}$$

where $\hat{\beta}(t,T) \;=\; \sum_{i=1}^m \hat{\beta}_j \,\mathcal{I}\left\{\, t/T \in (\hat{\lambda}_{j-1}, \hat{\lambda}_j] \,\right\}$ and $\beta^0(t,T) \;=\; \sum_{i=1}^m \beta_j^0 \,\mathcal{I}\{\, t/T \in (\lambda_{j-1}, \lambda_j] \,\}$. From (31), it follows that Lemma 1(i) is established if it can be shown that

$$T^{-1}(\tilde{U}' \bar{W}^* \hat{\beta} - \tilde{U}' \bar{W}^0 \beta^0) \;=\; o_p(1) \tag{32}$$

Since the 2SLS estimator based on the partition $(\hat{T}_1, ..., \hat{T}_m)$ is $\hat{\beta} = (\bar{W}^{*'} \bar{W}^*)^{-1} \bar{W}^{*'} Y$, it follows that

$$\begin{aligned} \tilde{U}' \bar{W}^* \hat{\beta} \;-\; \tilde{U}' \bar{W}^0 \beta^0 &= \tilde{U}' \bar{W}^* (\bar{W}^{*'} \bar{W}^*)^{-1} \bar{W}^{*'} Y - \tilde{U}' \bar{W}^0 \beta^0 \\ &= \tilde{U}' P_{\bar{W}^*} (\bar{W}^0 \beta^0 + \tilde{U}) - \tilde{U}' \bar{W}^0 \beta^0 \\ &= \tilde{U}' P_{\bar{W}^*} \bar{W}^0 \beta^0 + \tilde{U}' P_{\bar{W}^*} \tilde{U} - \tilde{U}' \bar{W}^0 \beta^0 \end{aligned} \tag{33}$$

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

We now analyze the terms on the right hand side of (33). It is most convenient to begin by analyzing $\|P_{\bar{W}^*} \tilde{U}\|$. To this end, it is convenient to define $\sum_i$ to denote the summation over observations $t = \hat{T}_i + 1, \hat{T}_i + 2, \ldots, \hat{T}_{i+1}$. We first note $\|P_{\bar{W}^*}\tilde{U}\|^2 = \tilde{U}' P_{\bar{W}^*}\tilde{U}$ is the sum of the $m+1$ terms

$$n_{i,T} \;=\; \left(\sum_i \hat{x}_t \tilde{u}_t\right)' \left(\sum_i \hat{x}_t \hat{x}_t'\right)^{-1} \left(\sum_i \hat{x}_t \tilde{u}_t\right) \tag{34}$$

for $i = 0, 1, ..., m$. and so we can deduce the order of $\|P_{\bar{W}^*}\tilde{U}\|^2$ by considering the behaviour of $\sum_i \hat{x}_t \tilde{u}_t$ and $\sum_i \hat{x}_t \hat{x}_t'$. From (2) and (10), it follows that

$$\hat{x}_t^{'} \;=\; z_t' \Delta_0 + z_t{}'(Z'Z)^{-1} Z'V \tag{35}$$

From (1), it follows that

$$\begin{aligned}
\tilde{u}_t &= y_t - \hat{x}_t' \beta^0(t,T) \\
&= (x_t{}'\beta^0(t,T) + u_t) - \hat{x}_t'\beta^0(t,T) \\
&= u_t + v_t{}'\beta^0(t,T) - z_t{}'[(Z'Z)^{-1}Z'V]\beta^0(t,T)
\end{aligned} \tag{36}$$

It follows from (35)-(36) that

$$\begin{aligned}
\sum_i \hat{x}_t \tilde{u}_t &= \sum_i [\Delta_0{}' z_t + V'Z(Z'Z)^{-1} z_t][u_t + v_t{}'\beta^0(t,T) - z_t'(Z'Z)^{-1}Z'V\beta^0(t,T)] \\
&= \sum_i [\Delta_0{}' z_t u_t + V'Z(Z'Z)^{-1} z_t u_t + \Delta_0{}' z_t v_t{}'\beta^0(t,T) + V'Z(Z'Z)^{-1} z_t v_t{}'\beta^0(t,T) \\
&\quad - \Delta_0{}' z_t z_t'(Z'Z)^{-1}Z'V\beta^0(t,T) - V'Z(Z'Z)^{-1} z_t z_t'(Z'Z)^{-1}Z'V\beta^0(t,T)] \\
&= \Delta_0{}' \sum_i z_t u_t + V'Z(Z'Z)^{-1} \sum_i z_t u_t + \Delta_0{}' \sum_i z_t v_t{}'\beta^0(t,T) \\
&\quad + V'Z(Z'Z)^{-1} \sum_i z_t v_t{}'\beta^0(t,T) - \Delta_0{}' \sum_i z_t z_t' \cdot (Z'Z)^{-1}Z'V\beta^0(t,T) \\
&\quad - V'Z(Z'Z)^{-1} \sum_i z_t z_t'(Z'Z)^{-1}Z'V\beta^0(t,T)
\end{aligned} \tag{37}$$

From (37) and Assumptions 3 and 6, it follows that

$$\sum_i \hat{x}_t \tilde{u}_t \;=\; O_p(T^{1/2}) \tag{38}$$

Now consider $\sum_i \hat{x}_t \hat{x}_t^{'}$. To this end, define $\sum_t$ to denote the summation over observations $t = 1, 2, \ldots, T$. From (2) and (10), it follows that

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

and hence that

$$\begin{aligned}\sum_i \hat{x}_t\hat{x}_t^{'} &= (\sum_t x_t z_t')(\sum_t z_t z_t')^{-1}(\sum_i z_t z_t')(\sum_t z_t z_t')^{-1}(\sum_t z_t x_t') \\ &= (T^{-1}\sum_t x_t z_t')(T^{-1}\sum_t z_t z_t')^{-1}(\sum_i z_t z_t')(T^{-1}\sum_t z_t z_t')^{-1}(T^{-1}\sum_t z_t x_t') \quad (39)\end{aligned}$$

From (39) and Assumptions 3 and 6, it follows that

$$\sum_i \hat{x}_t\hat{x}_t^{'} \;=\; O_p(T) \tag{40}$$

From (34), (38) and (40), it follows that $n_{i,T} = O_p(1)$ and hence that

$$\|P_{\bar{W}^*}\tilde{U}\|^2 \;=\; O_p(1) \tag{41}$$

Therefore, the second term on the right hand side of (33) is $O_p(1)$. Now consider the first term on the right hand side of (33). Using (30), it follows that

$$\|\tilde{U}'P_{\bar{W}^*}\bar{W}^0\beta^0\| \leq \|\tilde{U}'P_{\bar{W}^*}\| \cdot \|\bar{W}^0\beta^0\| \tag{42}$$

Since $W = P_z X$, where X is the original design matrix and $P_Z = Z(Z'Z)^{-1}Z'$ is a projection matrix, it follows from (28)-(29), (2) and Assumptions 3, 4 and 6 that

$$\|\bar{W}^0\| = \|W\| \;=\; \|P_Z X\| \;\leq\; \|X\| \;\leq\; (trX'X)^{1/2} = O_p(T^{1/2}) \tag{43}$$

and hence from (41)-(43) that

$$\|\tilde{U}'P_{\bar{W}^*}\bar{W}^0\beta^0\| \;=\; O_p(T^{1/2}) \tag{44}$$

Finally, consider the third term on the right hand side of (33), $\tilde{U}'\bar{W}^0\beta^0$. Notice that $\tilde{U}'\bar{W}^0$ consists of $m+1$ terms, $\sum_{t=T_{i-1}^0+1}^{T_i^0} \hat{x}_t\tilde{u}_t$. Using a similar argument to the derivation of (38), it can be shown that $\sum_{t=T_{i-1}^0+1}^{T_i^0} \hat{x}_t\tilde{u}_t = O_p(T^{1/2})$ and hence that

$$\|\tilde{U}'\bar{W}^0\beta^0\| \;=\; O_p(T^{1/2}) \tag{45}$$

Combining (33), (41), (44) and (45), it follows that

$$\tilde{U}'\bar{W}^*\hat{\beta} - \tilde{U}'\bar{W}^0\beta^0 \;=\; O_p(T^{1/2})$$

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

Part (ii):

Suppose $\hat{\lambda}_j \overset{p}{\nrightarrow} \lambda_j^0$ for some j. In this case, there exists $\eta > 0$ such that no estimated breaks fall into $[T(\lambda_j^0 - \eta), T(\lambda_j^0 + \eta)]$ with some positive probability ϵ. Suppose further that the interval belongs to the k^{th} estimated regime, then it follows that $\hat{T}_{k-1} < T(\lambda_j^0 - \eta)$ and $T(\lambda_j^0 + \eta) < \hat{T}_k$. Thus it follows that: $d_t = \hat{x}_t'(\hat{\beta}_k - \beta_j^0)$ for $t \in [T(\lambda_j^0 - \eta), T\lambda_j^0]$, and $d_t = \hat{x}_t'(\hat{\beta}_k - \beta_{j+1}^0)$ for $t \in [T\lambda_j^0 + 1, T(\lambda_j^0 + \eta)]$. Using these identities, it follows that

$$\sum_{t=1}^{T} d_t^2 \;\geq\; \sum_1 d_t^2 + \sum_2 d_t^2 \tag{46}$$

where

$$\begin{aligned} \sum_1 d_t^2 &= \left(\hat{\beta}_k - \beta_j^0\right)' \left(\sum_1 \hat{x}_t \hat{x}_t'\right) \left(\hat{\beta}_k - \beta_j^0\right) & (47) \\ \sum_2 d_t^2 &= \left(\hat{\beta}_k - \beta_{j+1}^0\right)' \left(\sum_2 \hat{x}_t \hat{x}_t'\right) \left(\hat{\beta}_k - \beta_{j+1}^0\right) & (48) \end{aligned}$$

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

At this stage, it is necessary to define γ_1 and γ_2 to be the smallest eigenvalue of $\sum_1 z_t z_t'$ and $\sum_2 z_t z_t'$, respectively. Then, since $\sum_i \hat{x}_t \hat{x}_t' = \hat{\Delta}_T' \left(\sum_i z_t z_t'\right) \hat{\Delta}_T$, it follows that[19]

$$\begin{aligned} \sum_1 {d_t}^2 + \sum_2 {d_t}^2 &= (\hat{\beta}_k - \beta_j^0)' \hat{\Delta}_T' \left(\sum_1 z_t z_t'\right) \hat{\Delta}_T (\hat{\beta}_k - \beta_j^0) \\ &\quad + (\hat{\beta}_k - \beta_{j+1}^0)' \hat{\Delta}_T' \left(\sum_2 z_t z_t'\right) \hat{\Delta}_T (\hat{\beta}_k - \beta_{j+1}^0) \\ &= \left(\hat{\Delta}_T(\hat{\beta}_k - \beta_j^0)\right)' \left(\sum_1 z_t z_t'\right) \left(\hat{\Delta}_T(\hat{\beta}_k - \beta_j^0)\right) \\ &\quad + \left(\hat{\Delta}_T(\hat{\beta}_k - \beta_{j+1}^0)\right)' \left(\sum_2 z_t z_t'\right) \left(\hat{\Delta}_T(\hat{\beta}_k - \beta_{j+1}^0)\right) \\ &\geq \gamma_1 \|\hat{\Delta}_T(\hat{\beta}_k - \beta_j^0)\|^2 + \gamma_2 \|\hat{\Delta}_T(\hat{\beta}_k - \beta_{j+1}^0)\|^2 \\ &\geq min\{\gamma_1, \gamma_2\} \cdot \left(\|\hat{\Delta}_T(\hat{\beta}_k - \beta_j^0)\|^2 + \|\hat{\Delta}_T(\hat{\beta}_k - \beta_{j+1}^0)\|^2\right) \\ &\geq (1/2) \cdot min\{\gamma_1, \gamma_2\} \cdot \|\hat{\Delta}_T(\beta_j^0 - \beta_{j+1}^0)\|^2 & (49) \end{aligned}$$

[19] The last inequality exploits: $(n-a)'A(n-a) + (n-b)'A(n-b) \geq (1/2)(a-b)'A(a-b)$ for an arbitrary positive definite matrix A and for all n; see Bai and Perron (1998)[p.69].

Now consider the right hand side of (49). We have

$$\sum_{1} z_t z_t' \; = \; (T\eta)(1/T\eta) \sum_{t=T(\lambda_j^0-\eta)}^{T\lambda_j^0} z_t z_t' \; = \; (T\eta)A_T \tag{50}$$

where $A_T = (1/T\eta)\sum_{t=T(\lambda_j^0-\eta)}^{T\lambda_j^0} z_t z_t'$. From Assumption 5, the smallest eigenvalue of A_T is bounded away from zero. Thus, the smallest eigenvalue of $(T\eta)A_T$ is of order $T\eta$. Similarly, the smallest eigenvalue of $\sum_2 z_t z_t'$ is of order $T\eta$. Using these two order statements in (49), it follows that

$$\sum_{t=1}^{T} {d_t}^2 \; \geq \; \sum_{1} {d_t}^2 + \sum_{2} {d_t}^2 \geq TC \cdot \|\hat{\Delta}_T(\beta_j^0 \, - \, \beta_{j+1}^0)\|^2$$

for some $C > 0$ and hence that

$$T^{-1}\sum_{t=1}^{T} d_t^2 \; \geq \; C\|\hat{\Delta}_T(\beta_j^0 \, - \, \beta_{j+1}^0)\|^2 \tag{51}$$

Under Assumptions 3, 4 and 6 $\hat{\Delta}_T \xrightarrow{p} \Delta_0$ and hence it follows from (51) that

$$T^{-1}\sum_{t=1}^{T} d_t^2 \; \geq \; C\|\Delta_0(\beta_j^0 \, - \, \beta_{j+1}^0)\|^2 \, + \, \xi_T \tag{52}$$

where

$$\xi_T \; = \; C\left\{ \|\hat{\Delta}_T(\beta_j^0 \, - \, \beta_{j+1}^0)\|^2 \, - \, \|\Delta_0(\beta_j^0 \, - \, \beta_{j+1}^0)\|^2 \right\}$$

Given the consistency of $\hat{\Delta}_T$, we have $\xi_T = o_p(1)$. The desired result then follows from (52) upon recalling that the analysis is premised on an event that occurs with probability ϵ.

Proof of Theorem 1:

Suppose that $\hat{\lambda}_j \stackrel{p}{\nrightarrow} \lambda_j^0$ for some j in probability. In this case, it follows from (12) and Lemma 1 that

$$(1/T)\sum_{t=1}^{T} \hat{u}_t^2 \; = \; (1/T)\sum_{t=1}^{T} \tilde{u}_t^2 + C \cdot \|\Delta_0(\beta_j^0 - \beta_{j+1}^0)\|^2 + o_p(1) \tag{53}$$

with probability at least as large as $\bar{\epsilon} > 0$. Assumptions 4 states that Δ_0 is full rank and so $\|\Delta_0(\beta_j^0 \, - \, \beta_{j+1}^0)\|^2 > 0$. Therefore, (53) conflicts with (11) which must hold for all T with probability one. Therefore, it must follow $\hat{\lambda}_j \xrightarrow{p} \lambda_j^0$ for all j.

Proof of Theorem 2:

The general proof strategy is the same as the one employed in Bai and Perron's (1998) proof of their Proposition 2 although the specific details are naturally different. Following Bai and Perron (1998), we assume (without loss of generality) that there are only 3 break points, that is $m = 3$. Here we present the proof for the middle break fraction, $\hat{\lambda}_2$. The proof for the end break fractions, $\hat{\lambda}_1$ and $\hat{\lambda}_3$, follows along similar lines and is omitted for brevity.[20]

For each $\epsilon > 0$ define $V_\epsilon = \{(T_1, T_2, T_3) : |T_i - T_i^0| \leq \epsilon T, i = 1, 2, 3\}$. Note that Theorem 1 implies $P(\{\hat{T}_1, \hat{T}_2, \hat{T}_3\} \in V_\epsilon) \longrightarrow 1$ as $T \to \infty$. Therefore, it suffices to consider the behaviour of $S_T(T_1, T_2, T_3)$ over V_ϵ for which $|T_i - T_i^0| < \epsilon T$ for all i. Without loss of genarality, we can restrict attention to the case in which $T_2 < T_2^0$.[21] For $C > 0$, we define

$$V_\epsilon(C) = \{(T_1, T_2, T_3) : |T_i - T_i^0| \leq \epsilon T,\ i = 1, 2, 3 \text{ but } T_2 - T_2^0 < -C\} \tag{54}$$

Note that by definition $V_\epsilon(C) \subset V_\epsilon$. Notice that the desired result would be established if it can be shown that for large C, $(\hat{T}_1, \hat{T}_2, \hat{T}_3) \notin V_\epsilon(C)$ - and hence $|\hat{T}_2 - T_2^0| < C$ - with high probability for large T. Since $S_T(\hat{T}_1, \hat{T}_2, \hat{T}_3) \leq S_T(\hat{T}_1, T_2^0, \hat{T}_3)$ with probability one as $T \to \infty$, the desired result can be established if it can be shown that for each $\eta > 0$, there exists $C > 0$ and $\epsilon > 0$ such that for large T,

$$P(min\{[S_T(T_1, T_2, T_3) - S_T(T_1, T_2^0, T_3)]/(T_2^0 - T_2)\} \ < \ 0) < \eta \tag{55}$$

where the minimum is taken over the set $V_\epsilon(C)$. Therefore, we now prove (55).

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

$$\begin{aligned} S_T(T_1, T_2, T_3) \ - \ S_T(T_1, T_2^0, T_3) &= SSR_1 \ - \ SSR_2 \\ &= (SSR_1 \ - \ SSR_3) \ - \ (SSR_2 \ - \ SSR_3) \end{aligned} \tag{56}$$

Note that: $SSR_1 - SSR_3$ is the difference in the residual sum of squares between breaks at (T_1, T_2, T_3) and when there is a fourth break at time T_2^0 between T_2 and T_3 in addition to those at (T_1, T_2, T_3); $SSR_2 - SSR_3$ is the difference in the residual sum of squares between breaks at (T_1, T_2^0, T_3) and when there is a fourth break at time T_2 between T_1 and T_2^0 in addition to those at (T_1, T_2^0, T_3).

[20] The proof is presented in Han (2006).

[21] Bai and Perron (1998) note that the proof for this case is easily modified to cover the case of $T_2 > T_2^0$ using an argument of symmetry.

To analyze the terms on the right hand side of (56), it is useful to define the 2SLS estimators in the four break model and emphasize the sub-samples upon which certain of these estimators are based. Let $(\hat{\beta}_1^*, \hat{\beta}_2^*, \hat{\beta}_\triangle, \hat{\beta}_3^*, \hat{\beta}_4^*)$ denote the 2SLS estimators of the regression coefficients in the five regimes of the four break model associated with the the partition (T_1, T_2, T_2^0, T_3). Note that $\hat{\beta}_2^*$ is based on observations $T_1+1, \ldots, T_2$; $\hat{\beta}_\triangle$ is based on observations $T_2+1, \ldots, T_2^0$; $\hat{\beta}_3^*$ is based on observations $T_2^0+1, \ldots, T_3$.

Now define $\bar{W}$ to be the diagonal partition of W at (T_1, T_2, T_3), $\tilde{W}$ is the diagonal partition of W at (T_1, T_2^0, T_3), $W_\triangle = (0_{p\times T_2}, \hat{x}_{T_2+1}, ..., \hat{x}_{T_2^0}, 0_{p\times(T-T_2^0)})'$ and $M_{\bar{W}} = I_T - \bar{W}(\bar{W}'\bar{W})^{-1}\bar{W}'$.

Now consider the right hand side of (56). It can be shown that[22]

$$SSR_1 - SSR_3 = (\hat{\beta}_3^* - \hat{\beta}_\triangle)' W_\triangle' M_{\bar{W}} W_\triangle (\hat{\beta}_3^* - \hat{\beta}_\triangle) \tag{57}$$

$$SSR_2 - SSR_3 = (\hat{\beta}_2^* - \hat{\beta}_\triangle)' W_\triangle' M_{\tilde{W}} W_\triangle (\hat{\beta}_2^* - \hat{\beta}_\triangle) \tag{58}$$

From (57)-(58), it follows that (56) can be written as

$$SSR_1 - SSR_2 = (\hat{\beta}_3^* - \hat{\beta}_\triangle)' W_\triangle' M_{\bar{W}} W_\triangle (\hat{\beta}_3^* - \hat{\beta}_\triangle) - (\hat{\beta}_2^* - \hat{\beta}_\triangle)' W_\triangle' M_{\tilde{W}} W_\triangle (\hat{\beta}_2^* - \hat{\beta}_\triangle) \tag{59}$$

Using $W_\triangle' M_{\tilde{W}} W_\triangle \leq W_\triangle' W_\triangle$, it follows from (59) that

$$SSR_1 - SSR_2 \geq (\hat{\beta}_3^* - \hat{\beta}_\triangle)' W_\triangle' M_{\bar{W}} W_\triangle (\hat{\beta}_3^* - \hat{\beta}_\triangle) - (\hat{\beta}_2^* - \hat{\beta}_\triangle)' W_\triangle' W_\triangle (\hat{\beta}_2^* - \hat{\beta}_\triangle) \tag{60}$$

Substituting for $M_{\bar{W}}$ in (60) and dividing both sides by $T_2^0 - T_2$, we obtain

$$\frac{SSR_1 - SSR_2}{T_2^0 - T_2} \geq N_1 - N_2 - N_3 \tag{61}$$

where

$$N_1 = (\hat{\beta}_3^* - \hat{\beta}_\triangle)'[(T_2^0 - T_2)^{-1} W_\triangle' W_\triangle](\hat{\beta}_3^* - \hat{\beta}_\triangle) \tag{62}$$

$$N_2 = (\hat{\beta}_3^* - \hat{\beta}_\triangle)'[(T_2^0 - T_2)^{-1} W_\triangle' \bar{W}][T^{-1}\bar{W}'\bar{W}]^{-1}[T^{-1}\bar{W}' W_\triangle](\hat{\beta}_3^* - \hat{\beta}_\triangle) \tag{63}$$

$$N_3 = (\hat{\beta}_2^* - \hat{\beta}_\triangle)'[(T_2^0 - T_2)^{-1} W_\triangle' W_\triangle](\hat{\beta}_2^* - \hat{\beta}_\triangle) \tag{64}$$

We now consider the behaviour of N_1, N_2 and N_3 in turn.

[22] See Amemiya (1985) equation (1.5.31) or Han (2006).

Consider N_1. First, note that by controlling ϵ to be small enough, we can control the distance between T_i and T_i^0 to be small over $V_\epsilon(C)$. Thus, $\hat{\beta}_3^*$ should be close to β_3^0 over $V_\epsilon(C)$. Second, note that $\hat{\beta}_\triangle$ is estimated using observations from $(T_2+1, ..., T_2^0)$, and that if C is large then this estimation is based a large number of observations and hence $\hat{\beta}_\triangle$ is close to β_2^0 with high probability. Therefore, for large C, large T, and small ϵ, we have $N_1 \geq (1/2)(\beta_3^0 - \beta_2^0)'[W_\triangle' W_\triangle/(T_2^0 - T_2)](\beta_3^0 - \beta_2^0)$ with large probability.

Now consider N_2. From the property of LS estimation, $\hat{\beta}_3^*$ and $\hat{\beta}_\triangle$ are $O_p(1)$ uniformly on $V_\epsilon(C)$. We also have that, on $V_\epsilon(C)$, $(\bar{W}'\bar{W}/T)^{-1} = O_p(1)$ and $W_\triangle'\bar{W}/(T_2^0 - T_2) = O_p(1)$. Furthermore, $||\bar{W}'W_\triangle/T|| = ||[\bar{W}'W_\triangle/(T_2^0 - T_2)] \cdot [(T_2^0 - T_2)/T]|| = ||\bar{W}'W_\triangle/(T_2^0 - T_2)|| \cdot (T_2^0 - T_2)/T \leq O_p(1)\epsilon$ over $V_\epsilon(C)$. therefore, we have that $N_2 \leq O_p(1) \cdot O_p(1) \cdot O_p(1) \cdot O_p(1)\epsilon \cdot O_p(1) = O_p(1)\epsilon$.

Finally, consider N_3. Since both $\hat{\beta}_2^*$ and $\hat{\beta}_\triangle$ are estimating β_2^0, it follows that $||\hat{\beta}_2^* - \hat{\beta}_\triangle|| < \rho$ with large probability for every $\rho > 0$, for large T, large C, and small ϵ. Furthermore, we have $||W_\triangle' W_\triangle/(T_2^0 - T_2)|| = O_p(1)$ uniformly on $V_\epsilon(C)$. Therefore, it follows that $N_3 \leq \rho O_p(1)$.

Combining (61) with our analyses of N_1, N_2 and N_3, it follows that

$$\frac{SSR_1 - SSR_2}{T_2^0 - T_2} \;\geq\; 2^{-1}(\beta_3^0 - \beta_2^0)' \cdot [W_\triangle' W_\triangle/(T_2^0 - T_2)] \cdot (\beta_3^0 - \beta_2^0) - \epsilon O_p(1) - \rho O_p(1) \qquad (65)$$

with large probability. We now show that the first term on the right hand side of (65) dominates. Noting that

$$(T_2^0 - T_2)^{-1} W_\triangle' W_\triangle \;=\; \hat{\Delta}_T'(T_2^0 - T_2)^{-1} \sum_{t=T_2+1}^{T_2^0} z_t z_t' \hat{\Delta}_T \qquad (66)$$

and $\hat{\Delta}_T \xrightarrow{p} \Delta_0$, a matrix of full column rank (from Assumption 4), it follows from Assumption 5 that, with large probability, the minimum eigenvalue of $W_\triangle' W_\triangle/(T_2^0 - T_2)$ is bounded away from zero on $V_\epsilon(C)$. Therefore, the first term on the right hand side of (65) dominates. This term is also positive by Assumption 5. Therefore, $[(SSR_1 - SSR_2)/(T_2^0 - T_2)] > 0$ over $V_\epsilon(C)$ with large probability which proves (55).

Proof of Theorem 3:

Before proving this result, it is useful to present the following lemma regarding the behaviour of the 2SLS based on an arbitrary partition of the data. Accordingly, define $\hat{\beta}(s, r)$ to be the 2SLS based on the observations $t = [Ts]+1, \cdots, [Tr]$, that is

$$\hat{\beta}(s,r) = \{\sum_{t=[Ts]+1}^{[Tr]} \hat{x}_t \hat{x}_t'\}^{-1} \sum_{t=[Ts]+1}^{[Tr]} \hat{x}_t y_t \tag{67}$$

Lemma A.1: Under the conditions of Theorem 3, we have

$$sup_{(s,r)\in(0,1)^2, r>+\epsilon s} \|\hat{\beta}(s,r)\| = O_p(1)$$

where ϵ is defined in Assumption 10.

Proof of Lemma A.1:

Based on an arbitrary partition of the data, 2SLS coefficient estimator can be written as

$$\begin{aligned}
\hat{\beta}(s,r) &= \{\sum_{t=[Ts]+1}^{[Tr]} \hat{x}_t(\hat{\pi})\hat{x}_t(\hat{\pi})'\}^{-1} \sum_{t=[Ts]+1}^{[Tr]} \hat{x}_t(\hat{\pi}) y_t \\
&= \{\sum_{t=[Ts]+1}^{[Tr]} \hat{x}_t(\hat{\pi})\hat{x}_t(\hat{\pi})'\}^{-1} \sum_{t=[Ts]+1}^{[Tr]} \hat{x}_t(\hat{\pi})[\hat{x}_t(\hat{\pi})'\beta^0(t,T) + \tilde{u}_t(\hat{\pi})] \\
&= \{\sum_{t=[Ts]+1}^{[Tr]} \hat{x}_t(\hat{\pi})\hat{x}_t(\hat{\pi})'\}^{-1} \sum_{t=[Ts]+1}^{[Tr]} \hat{x}_t(\hat{\pi})\hat{x}_t(\hat{\pi})'\beta^0(t,T) + \{\sum_{t=[Ts]+1}^{[Tr]} \hat{x}_t(\hat{\pi})\hat{x}_t(\hat{\pi})'\}^{-1} \\
&\quad \times \sum_{t=[Ts]+1}^{[Tr]} \hat{x}_t(\hat{\pi})\tilde{u}_t(\hat{\pi})
\end{aligned} \tag{68}$$

where

$$\begin{aligned}
\hat{x}_t(\hat{\pi}) &= \sum_{t=1}^{T} x_t \tilde{z}_t(\hat{\pi})'\{\sum_{t=1}^{T} \tilde{z}_t(\hat{\pi})\tilde{z}_t(\hat{\pi})'\}^{-1}\tilde{z}_t(\hat{\pi}) & (69)\\
\tilde{u}_t(\hat{\pi}) &= y_t - \hat{x}_t(\hat{\pi})'\beta^0(t,T) \\
&= \tilde{u}_t(\pi^0) + [\tilde{z}_t(\pi^0)'\{\sum_{t=1}^{T} \tilde{z}_t(\pi^0)\tilde{z}_t(\pi^0)'\}^{-1} \sum_{t=1}^{T} \tilde{z}_t(\pi^0)x_t' - \tilde{z}_t(\hat{\pi})'\{\sum_{t=1}^{T} \tilde{z}_t(\hat{\pi})\tilde{z}_t(\hat{\pi})'\}^{-1} \\
&\quad \times \sum_{t=1}^{T} \tilde{z}_t(\hat{\pi})x_t']\beta^0(t,T) & (70)
\end{aligned}$$

and

$$\begin{aligned}
\hat{x}_t(\pi^0) &= \sum_{t=1}^{T} x_t \tilde{z}_t(\pi^0)'\{\sum_{t=1}^{T} \tilde{z}_t(\pi^0)\tilde{z}_t(\pi^0)'\}^{-1}\tilde{z}_t(\pi^0) & (71)\\
\tilde{u}_t(\pi^0) &= y_t - \hat{x}_t(\pi^0)'\beta^0(t,T) & (72)
\end{aligned}$$

From (69) and (70), it follows that

$$\sum_{t=[Ts]+1}^{[Tr]} \hat{x}_t(\hat{\pi})\hat{x}_t(\hat{\pi})' = \sum_{t=1}^{T} x_t\tilde{z}_t(\hat{\pi})'\{\sum_{t=1}^{T}\tilde{z}_t(\hat{\pi})\tilde{z}_t(\hat{\pi})'\}^{-1}\sum_{t=[Ts]+1}^{[Tr]}\tilde{z}_t(\hat{\pi})\tilde{z}_t(\hat{\pi})'\{\sum_{t=1}^{T}\tilde{z}_t(\hat{\pi})\tilde{z}_t(\hat{\pi})'\}^{-1} \times \sum_{t=1}^{T}\tilde{z}_t(\hat{\pi})x_t' \tag{73}$$

and

$$\begin{aligned}\sum_{t=[Ts]+1}^{[Tr]} \hat{x}_t(\hat{\pi})\tilde{u}_t(\hat{\pi}) &= \sum_{t=1}^{T} x_t\tilde{z}_t(\hat{\pi})'\{\sum_{t=1}^{T}\tilde{z}_t(\hat{\pi})\tilde{z}_t(\hat{\pi})'\}^{-1}[\sum_{t=[Ts]+1}^{[Tr]}\tilde{z}_t(\hat{\pi})\tilde{u}_t(\pi^0) + \sum_{t=[Ts]+1}^{[Tr]}\tilde{z}_t(\hat{\pi})\tilde{z}_t(\pi^0)' \\ &\times\{\sum_{t=1}^{T}\tilde{z}_t(\pi^0)\tilde{z}_t(\pi^0)'\}^{-1}\sum_{t=1}^{T}\tilde{z}_t(\pi^0)x_t'\beta^0(t,T) - \sum_{t=[Ts]+1}^{[Tr]}\tilde{z}_t(\hat{\pi}_t)\tilde{z}_t(\hat{\pi})' \\ &\times\{\sum_{t=1}^{T}\tilde{z}_t(\hat{\pi})\tilde{z}_t(\hat{\pi})'\}^{-1}\sum_{t=1}^{T}\tilde{z}_t(\hat{\pi})x_t'\beta^0(t,T)]\end{aligned} \tag{74}$$

Notice that $\sum_{t=[Ts]+1}^{[Tr]}\tilde{z}_t(\hat{\pi})\tilde{u}_t(\hat{\pi})$ depends on $\sum_{t=[Ts]+1}^{[Tr]}\tilde{z}_t(\hat{\pi})\tilde{u}_t(\pi^0)$. Using a similar argument to the derivation of (36), we have

$$\tilde{u}_t(\pi^0) = u_t + v_t'\beta^0(t,T) - \tilde{z}_t(\pi^0)'[\{\sum_{t=1}^{T}\tilde{z}_t(\pi^0)\tilde{z}_t(\pi^0)'\}^{-1}\sum_{t=1}^{T}\tilde{z}_t(\pi^0)v_t^{'}]\beta^0(t,T) \tag{75}$$

and hence

$$\begin{aligned}\sum_{t=[Ts]+1}^{[Tr]}\tilde{z}_t(\hat{\pi})\tilde{u}_t(\pi^0) &= \sum_{t=[Ts]+1}^{[Tr]}\tilde{z}_t(\hat{\pi})u_t + \sum_{t=[Ts]+1}^{[Tr]}\tilde{z}_t(\hat{\pi})v_t'\beta^0(t,T) - \sum_{t=[Ts]+1}^{[Tr]}\tilde{z}_t(\hat{\pi})\tilde{z}_t(\pi^0)' \\ &\times\{\sum_{t=1}^{T}\tilde{z}_t(\pi^0)\tilde{z}_t(\pi^0)\}^{-1}\sum_{t=1}^{T}\tilde{z}_t(\pi^0)v_t'\beta^0(t,T)\end{aligned} \tag{76}$$

We now consider the limiting behaviour of the sums in (73)-(76). From Assumptions 6 and 8, it follows that

$$\begin{aligned}T^{-1}\sum_{t=1}^{T}\tilde{z}_t(\hat{\pi})\tilde{z}_t(\hat{\pi})' &= T^{-1}\sum_{t=1}^{T}\tilde{z}_t(\pi^0)\tilde{z}_t(\pi^0)' + o_p(1) \\ &\xrightarrow{p} Q\end{aligned} \tag{77}$$

where Q is the block diagonal matrix $diag(Q_1, Q_2, \ldots Q_{h+1})$ and $Q_i = Q_{ZZ}(\pi_i^0) - Q_{ZZ}(\pi_{i-1}^0)$ and we set $\pi_0^0 = 0$, $\pi_{h+1}^0 = 1$. From Assumptions 3, 6 and 8, and (3) it follows that

$$T^{-1}\sum_{t=1}^{T}\tilde{z}_t(\hat{\pi})x_t' = T^{-1}\sum_{t=1}^{T}\tilde{z}_t(\pi^0)x_t' + o_p(1) \xrightarrow{p} Q\Theta_0 \tag{78}$$

From Assumptions 6 and 8, it follows that

$$T^{-1}\sum_{t=[Ts]+1}^{[Tr]}\tilde{z}_t(\hat{\pi})\tilde{z}_t(\pi^0)' \xrightarrow{p} \tilde{Q}(s,r) \text{ uniformly in } r, s,\ (r > s+\epsilon) \tag{79}$$

where - assuming $\pi_i^0 < s \leq \pi_{i+1}^0$ and $\pi_{i+\ell}^0 < r \leq \pi_{i+\ell+1}^0$ without loss of generality - $\tilde{Q}(s,r) = [0_{(h+1)q\times iq}, A(s,r), 0_{(h+1)q\times(h-i-\ell-1)q}$ and $A(s,r)$ is the block diagonal matrix $diag(Q_{ZZ}(\pi_{i+1}^0) - Q_{ZZ}(s), Q(i+2), \ldots, Q(i+\ell), Q_{ZZ}(r) - Q_{ZZ}(\pi_{i+\ell}^0)$.

Finally, it follows from Assumption 3, 6 and 8 that $T^{-1/2}\sum_{t=1}^{[Tr]} \tilde{z}_t(\hat{\pi}) \otimes \{(u_t, v_t{}')'\}$ and $T^{-1/2}\sum_{t=1}^{[Tr]} \tilde{z}_t(\pi^0) \otimes \{(u_t, v_t{}')'\}$ satisfy a functional central limit theorem. The latter distributional result combined with (68)-(79) yields the desired result.

Proof of Theorem 3:

The proof follows similar lines to Theorem 1. We first establish the analogs to Lemma 1 (a)-(b) and then use them to deduce the desired result.

Lemma A.2 *Under the conditions of Theorem 3, we have*

(a) $T^{-1}\sum_{t=1}^{T} \tilde{u}_t d_t = o_p(1)$.

(b1) If $\hat{\lambda}_j \overset{p}{\nrightarrow} \lambda_j^0$ *for some* j, *and* $\lambda_j^0 \in (\pi_i^0, \pi_{i+1}^0)$, *then*

$$\limsup_{T\to\infty} P\left(T^{-1}\sum_{t=1}^{T} d_t{}^2 > C\|\Delta_0^{(i+1)}(\beta_j^0 - \beta_{j+1}^0) + \xi_T'\|^2\right) > \bar{\epsilon}$$

for some $C > 0$ *and* $\bar{\epsilon} > 0$, *where* $\xi_T' = o_p(1)$.

(b2) If $\hat{\lambda}_j \overset{p}{\nrightarrow} \lambda_j^0$ *for some* j, *and* $\lambda_j^0 = \pi_i^0$ *for some* i, *then*

$$\limsup_{T\to\infty} P\left(T^{-1}\sum_{t=1}^{T} d_t{}^2 > C\{\|\Delta_0^{(i)}(\hat{\beta}_k - \beta_j^0)\|^2 + \|\Delta_0^{(i+1)}(\hat{\beta}_k - \beta_{j+1}^0)\|^2 + \xi_T{}''\}\right) > \bar{\epsilon}$$

for some $C > 0$ *and* $\bar{\epsilon} > 0$, *where* $\xi_T{}'' = o_p(1)$.

Proof of Lemma A.2:

Part (a):

We first consider the case in which π^0 is known and so $\hat{x}_t$ is calculated via

$$\hat{x}_t(\pi^0)' = \tilde{z}_t(\pi^0)'\{\sum_{t=1}^{T}\tilde{z}_t(\pi^0)\tilde{z}_t(\pi^0)'\}^{-1}\sum_{t=1}^{T}\tilde{z}_t(\pi^0)x_t' \tag{80}$$

and define $d_t(\pi^0) = \hat{x}_t(\pi^0)'\left\{\hat{\beta}(t,T) - \beta^0(t,T)\right\}$. In this case, we have from (75)

$$\begin{aligned}\tilde{u}_t = \tilde{u}_t(\pi^0) &= y_t - \hat{x}_t(\pi^0)'\beta^0(t,T)\\ &= u_t + v_t'\beta^0(t,T) - \tilde{z}_t(\pi^0)'[\{\tilde{Z}(\pi^0)'\tilde{Z}(\pi^0)\}^{-1}\tilde{Z}(\pi^0)'V]\beta^0(t,T)\end{aligned}$$

and from (71),

$$\hat{x}_t(\pi^0) \;=\; z_t(\pi^0)'\Theta_0 \,+\, \tilde{z}_t(\pi^0)'\left(\tilde{Z}(\pi^0)'\tilde{Z}(\pi^0)\right)^{-1}\tilde{Z}(\pi^0)'V \tag{81}$$

where $\tilde{Z}(\pi^0)$ is the $T \times q(n+1)$ matrix with the t^{th} row $\tilde{z}_t(\pi^0)'$.

Note that Assumptions 3 and 6 imply

$$T^{-1/2}\sum_{t=1}^{[Tr]}\tilde{z}_t(\pi^0)\otimes\{(u_t, v_t')'\} \;=\; O_p(1) \tag{82}$$

and Assumption 6 implies

$$T^{-1}\sum_{t=1}^{[Tr]}\tilde{z}_t(\pi^0)\tilde{z}_t(\pi^0)' \;=\; O_p(1) \tag{83}$$

Using (82)-(83) and similar arguments to the proof of Lemma 1(a), it is straightforward to show that $T^{-1}\sum_{t=1}^{T}\tilde{u}_t(\pi^0)d_t(\pi^0) = o_p(1)$.

Now consider the case where $\hat{x}_t'$ is calculated via (14), which we now denote by $\hat{x}_t(\hat{\pi})'$ for emphasis, and $d_t(\hat{\pi}) = \hat{x}_t(\hat{\pi})'\left\{\hat{\beta}(t,T) - \beta^0(t,T)\right\}$. In this case, we have from (69)-(70),

$$\begin{aligned}
\tilde{u}_t(\hat{\pi}) \;&=\; y_t - \hat{x}_t(\hat{\pi})'\beta^0(t,T) \\
&=\; \tilde{u}_t(\pi^0) \,-\, \tilde{z}_t(\pi^0)'\{[\tilde{Z}(\hat{\pi})'\tilde{Z}(\hat{\pi})]^{-1}\tilde{Z}(\hat{\pi})'V \,-\, [\tilde{Z}(\pi^0)'\tilde{Z}(\pi^0)]^{-1}\tilde{Z}(\pi^0)'V\}\beta^0(t,T) \\
&\quad + [\tilde{z}_t(\pi^0) \,-\, \tilde{z}_t(\hat{\pi})]'[\tilde{Z}(\hat{\pi})'\tilde{Z}(\hat{\pi})]^{-1}\tilde{Z}(\hat{\pi})'V\beta^0(t,T) \,+\, \tilde{z}_t(\pi^0)'\{I \,-\, [\tilde{Z}(\hat{\pi})'\tilde{Z}(\hat{\pi})]^{-1} \\
&\quad \times \tilde{Z}(\hat{\pi})'\tilde{Z}(\pi^0)\}\Theta_0\beta^0(t,T) \,+\, [\tilde{z}_t(\pi^0) \,-\, \tilde{z}_t(\hat{\pi})]'[\tilde{Z}(\hat{\pi})'\tilde{Z}(\hat{\pi})]^{-1}\tilde{Z}(\hat{\pi})'\tilde{Z}(\pi^0)\Theta_0\beta^0(t,T)
\end{aligned}$$

and

$$\begin{aligned}
\hat{x}_t(\hat{\pi}) \;&=\; \hat{x}_t(\pi^0) \,-\, \tilde{z}_t(\pi^0)'\{I \,-\, [\tilde{Z}(\hat{\pi})'\tilde{Z}(\hat{\pi})]^{-1}\tilde{Z}(\hat{\pi})'\tilde{Z}(\pi^0)\}\Theta_0 \,+\, [\tilde{z}_t(\hat{\pi}) \,-\, \tilde{z}_t(\pi^0)]'[\tilde{Z}(\hat{\pi})'\tilde{Z}(\hat{\pi})]^{-1} \\
&\quad \times \tilde{Z}(\hat{\pi})'\tilde{Z}(\pi^0)\Theta_0 \,+\, [\tilde{z}_t(\hat{\pi}) \,-\, \tilde{z}_t(\pi^0)]'[\tilde{Z}(\pi^0)'\tilde{Z}(\pi^0)]^{-1}\tilde{Z}(\pi^0)'V \,+\, \tilde{z}_t(\hat{\pi})'\{[\tilde{Z}(\hat{\pi})'\tilde{Z}(\hat{\pi})]^{-1} \\
&\quad \times \tilde{Z}(\hat{\pi})'V \,-\, [\tilde{Z}(\pi^0)'\tilde{Z}(\pi^0)]^{-1}\tilde{Z}(\pi^0)'V\}
\end{aligned}$$

It follows from Assumptions 3, 6 and 8 that: $T^{-1}\tilde{Z}(\hat{\pi})'\tilde{Z}(\pi^0) \,=\, T^{-1}\tilde{Z}(\hat{\pi})'\tilde{Z}(\hat{\pi}) + o_p(1) \,=\, T^{-1}\tilde{Z}(\pi^0)'\tilde{Z}(\pi^0) \,+\, o_p(1)$, $T^{-1/2}\tilde{Z}(\hat{\pi})'V = T^{-1/2}\tilde{Z}(\pi^0)'V \,+\, o_p(1) = O_p(1)$ and $T^{-1/2}\tilde{Z}(\hat{\pi})'U = T^{-1/2}\tilde{Z}(\pi^0)'U + o_p(1) \,=\, O_p(1)$. Hence, it follows that $T^{-1}\sum_{t=1}^{T}\tilde{u}_t(\hat{\pi})d_t(\hat{\pi}) \,=\, T^{-1}\sum_{t=1}^{T}\tilde{u}_t(\pi^0)d_t(\pi^0) + o_p(1) = o_p(1)$ which gives the desired result.

Part (b1):

Again we begin by considering the case in which π^0 is known and so $\hat{x}_t(\pi^0)$ - defined in (80) - is used to predict x_t.

Since $\lambda_j^0 \in (\pi_i^0, \pi_{i+1}^0)$, we can choose $\eta > 0$ such that there is no estimated break in $[T(\lambda_j^0 - \eta), T(\lambda_j^0 + \eta)]$ with some positive probability ϵ and $\pi_i^0 < \lambda_j^0 - \eta < \lambda_j^0 + \eta < \pi_{i+1}^0$. As in the proof of Lemma 1(ii) assume $\hat{T}_{k-1} < T(\lambda_j^0 - \eta)$ and $T(\lambda_j^0 + \eta) < \hat{T}_k$.

Define

$$\begin{aligned} d_t(\pi^0) &= \hat{x}_t(\pi^0)'(\hat{\beta}_k - \beta_j^0) \qquad \text{for}\, t \in [T(\lambda_j^0 - \eta),\, T\lambda_j^0] \\ &= \hat{x}_t(\pi^0)'(\hat{\beta}_k - \beta_{j+1}^0) \qquad \text{for}\, t \in [T\lambda_j^0 + 1,\, T(\lambda_j^0 + \eta)] \end{aligned} \tag{84}$$

We have

$$\sum_{t=1}^{T} \{d_t(\pi^0)\}^2 \geq \sum_1 \{d_t(\pi^0)\}^2 + \sum_2 \{d_t(\pi^0)\}^2 \tag{85}$$

where

$$\begin{aligned} \sum_1 \{d_t(\pi^0)\}^2 &= (\hat{\beta}_k - \beta_j^0)' \sum_1 \hat{x}_t(\pi^0)\hat{x}_t(\pi^0)'(\hat{\beta}_k - \beta_j^0) \\ \sum_2 \{d_t(\pi^0)\}^2 &= (\hat{\beta}_k - \beta_{j+1}^0)' \sum_2 \hat{x}_t(\pi^0)\hat{x}_t(\pi^0)'(\hat{\beta}_k - \beta_{j+1}^0) \end{aligned}$$

and (as before) $\sum_1$ extends over $\{T(\lambda_j^0 - \eta) \leq t \leq T\lambda_j^0\}$ and $\sum_2$ extends over $\{T\lambda_j^0 + 1 \leq t \leq T(\lambda_j^0 + \eta)\}$.

Now, since $\pi_i^0 < \lambda_j^0 - \eta < \lambda_j^0 + \eta < \pi_{i+1}^0$,

$$\sum_i \hat{x}_t(\pi^0)\hat{x}_t(\pi^0)' = \hat{\Delta}_{T,i+1} \sum_i z_t z_t' \hat{\Delta}_{T,i+1}'$$

We can therefore follow the same argument as in the proof of Lemma 1(ii) to deduce that, for some $C > 0$

$$T^{-1} \sum_1 \{d_t(\pi^0)\}^2 + T^{-1} \sum_2 \{d_t(\pi^0)\}^2 \geq C \| \Delta_0^{(i+1)} (\beta_j^0 - \beta_{j+1}^0) \|^2 + \xi_T^*$$

where $\xi_T^* = C\{\| \hat{\Delta}_{T,i+1}(\beta_j^0 - \beta_{j+1}^0) \| - \| \Delta_0^{(i+1)}(\beta_j^0 - \beta_{j+1}^0) \|\}$. From Assumptions 3 and 6, and (3) it follows that $\hat{\Delta}_{T,i+1} \xrightarrow{p} \Delta_0^{(i+1)}$ and hence that $\xi_T = o_p(1)$.

Now consider the case in which π^0 is unknown and estimated via $\hat{\pi}$. Define

$$d_t(\hat{\pi}) = \hat{x}_t(\hat{\pi})(\hat{\beta}_k - \beta_j^0) \qquad \text{for} \quad t \in [T(\lambda_j^0 - \eta),\, T\lambda_j^0] \tag{86}$$

$$\quad = \hat{x}_t(\hat{\pi})(\hat{\beta}_k - \beta_{j+1}^0) \qquad \text{for} \quad t \in [T\lambda_j^0 + 1,\, T(\lambda_j^0 + \eta)] \tag{87}$$

Since $\hat{x}_t(\hat{\pi}) = z_t \hat{\Delta}(t, T)$ where $\hat{\Delta}(t, T) = \sum_{i=1}^{n+1} \hat{\Delta}_T^{(i)} \mathcal{I}\{t/T \in (\hat{\pi}_{i-1}, \hat{\pi}_i]\}$, we have

$$T^{-1} \sum_{t=1}^{T} \{d_t(\hat{\pi})\}^2 \geq T^{-1} \sum_1 \{d_t(\hat{\pi})\}^2 + T^{-1} \sum_2 \{d_t(\hat{\pi})\}^2 \tag{88}$$

where

$$T^{-1}\sum_{1}\{d_t(\hat{\pi})\}^2 \;=\; (\hat{\beta}_k - \beta_j^0)'\{T^{-1}\sum_{1}\hat{\Delta}(t,T)'z_t z_t'\hat{\Delta}(t,T)\}(\hat{\beta}_k - \beta_j^0) \tag{89}$$

$$\begin{aligned} T^{-1}\sum_{2}\{d_t(\hat{\pi})\}^2 \;&=\; (\hat{\beta}_k - \beta_{j+1}^0)'\{T^{-1}\sum_{2}\hat{\Delta}(t,T)'z_t z_t'\hat{\Delta}(t,T)\} \\ &\quad \times (\hat{\beta}_k - \beta_{j+1}^0) \end{aligned} \tag{90}$$

From Assumptions 3, 6 and 8, and (3) $\hat{\Delta}_{T,i} \xrightarrow{p} \Delta_0^{(i)}$ and $\hat{\pi} \xrightarrow{p} \pi^0$ and so using $\pi_i^0 < \lambda_j^0 - \eta < \lambda_j^0 + \eta < \pi_{i+1}^0$ we have

$$(1/T)\sum_{1}\{d_t(\hat{\pi})\}^2 \;=\; T^{-1}\sum_{1}\{d_t(\pi^0)\}^2 \,+\, \xi_{1,T}' \tag{91}$$

where

$$\begin{aligned} \xi_{1,T}' \;&=\; (\hat{\beta}_k - \beta_j^0)'\{T^{-1}\sum_{1}\hat{\Delta}(t,T)'z_t z_t'\hat{\Delta}(t,T) - \Delta_0^{(i+1)'}T^{-1}\sum_{1}z_t z_t' \\ &\quad \times \Delta_0^{(i+1)}\}(\hat{\beta}_k - \beta_j^0) \end{aligned}$$

and

$$T^{-1}\sum_{2}\{d_t(\hat{\pi})\}^2 \;=\; T^{-1}\sum_{2}\{d_t(\pi^0)\}^2 \,+\, \xi_{2,T}' \tag{92}$$

where

$$\begin{aligned} \xi_{2,T}' \;&=\; (\hat{\beta}_k - \beta_{j+1}^0)'\{T^{-1}\sum_{2}\hat{\Delta}(t,T)'z_t z_t'\hat{\Delta}(t,T) \,-\, \Delta_0^{(i+1)'}T^{-1}\sum_{2}z_t z_t' \\ &\quad \times \Delta_0^{(i+1)}\}(\hat{\beta}_k - \beta_{j+1}^0) \end{aligned}$$

Under our assumptions, $\xi_{i,T}' \,=\, o_p(1)$ - note that using Lemma A.1 we have $\|\hat{\beta}_k\| \leq sup_{(s,r)}\,\|\hat{\beta}(s,r)\| = O_p(1)$. Combining (88), (91) and (92), we have

$$\begin{aligned} (1/T)\sum_{t=1}^{T}\{d_t(\hat{\pi})\}^2 \;&\geq\; (1/T)\sum_{1}\{d_t(\hat{\pi})\}^2 \,+\, (1/T)\sum_{2}\{d_t(\hat{\pi})\}^2 \\ &=\; (1/T)\sum_{1}\{d_t(\pi^0)\}^2 \,+\, (1/T)\sum_{2}\{d_t(\pi^0)\}^2 \,+\, \xi_{1,T}' \,+\, \xi_{2,T}' \\ &\geq\; C\|\Delta_0^{(i+1)}(\beta_j^0 - \beta_{j+1}^0)\|^2 \,+\, \xi_T^* \,+\, \xi_{1,T}' \,+\, \xi_{2,T}' \end{aligned}$$

Recalling that this analysis is premised on an event that occurs with probability $\bar{\epsilon}$, it follows that

$$\limsup_{T\to\infty} P\left(T^{-1}\sum_{t=1}^{T}\{d_t(\hat{\pi})\}^2 \,>\, C\|\Delta_0^{(i+1)}(\beta_j^0 - \beta_{j+1}^0)\|^2 \,+\, \xi_T'\right) \,>\, \bar{\epsilon}$$

where $\xi'_T = \xi^*_T + \xi'_{1,T} + \xi'_{2,T} = o_p(1)$.

Part (b2):

As for part (b1), we assume that λ^0_j lies in the k^{th} estimated regime, and we choose η so that there is no estimated break in $[T(\lambda^0_j - \eta), T(\lambda^0_j + \eta)]$ with some positive probability ϵ. This time $\lambda^0_j = \pi^0_i$ but we can choose η such that $\pi^0_{i-1} < \lambda^0_j - \eta,\ \lambda^0_j + \eta < \pi^0_{i+1}$.

Again we begin by considering the case in which π^0 is known and so the predicted value of x_t is $\hat{x}_t(\pi^0)$ in (71). Define $d_t(\pi^0)$ as in (84). By definition, (85) holds but this time as $\lambda^0_j = \pi^0_i$.

$$\sum_1 \{d_t(\pi^0)\}^2 = (\hat{\beta}_k - \beta^0_j)'\hat{\Delta}_{T,i}\sum_1 z_t z_t'\hat{\Delta}_{T,i}(\hat{\beta}_k - \beta^0_j) \tag{93}$$

$$\sum_2 \{d_t(\pi^0)\}^2 = (\hat{\beta}_k - \beta^0_{j+1})'\hat{\Delta}_{T,i+1}\sum_2 z_t z_t'\hat{\Delta}_{T,i+1}(\hat{\beta}_k - \beta^0_{j+1}) \tag{94}$$

By repeating the steps in the proof of Lemma 1(ii), we have

$$\sum_1 \{d_t(\pi^0)\}^2 + \sum_2 \{d_t(\pi^0)\}^2 \geq \min\{\gamma_1, \gamma_2\}\{\|\hat{\Delta}_{T,i}(\hat{\beta}_k - \beta^0_j)\|^2 + \|\hat{\Delta}_{T,i+1}(\hat{\beta}_k - \beta^0_{j+1})\|^2\}$$

By similar arguments to Lemma 1(ii), we can then deduce that

$$\begin{aligned} T^{-1}\sum_1 \{d_t(\pi^0)\}^2 + T^{-1}\sum_2 \{d_t(\pi^0)\}^2 &\geq \{\|\hat{\Delta}_{T,i}(\hat{\beta}_k - \beta^0_j)\|^2 + \|\hat{\Delta}_{T,i+1}(\hat{\beta}_k - \beta^0_{j+1})\|^2\} \\ &= C\{\|\Delta^{(i)}_0(\hat{\beta}_k - \beta^0_j)\|^2 + \|\Delta^{(i+1)}_0(\hat{\beta}_k - \beta^0_{j+1})\|^2\} \\ &\quad + \xi^{**}_T \end{aligned} \tag{95}$$

where

$$\begin{aligned} \xi^{**}_T &= C\{\|\hat{\Delta}_{T,i}(\hat{\beta}_k - \beta^0_j)\|^2 - \|\Delta^{(i)}_0(\hat{\beta}_k - \beta^0_j)\|^2 \\ &\quad + \|\hat{\Delta}_{T,i+1}(\hat{\beta}_k - \beta^0_j)\|^2 - \|\Delta^{(i+1)}_0(\hat{\beta}_k - \beta^0_{j+1})\|^2\} \end{aligned}$$

Note that under our assumptions $\xi^{**}_T = o_p(1)$.

Now consider the case in which π^0 is unknown and estimated via $\hat{\pi}$. Define $d_t(\hat{\pi})$ as in (86)-(87). Following similar steps to the proof of part (b1), we have that (88)-(90) hold. From Assumptions 3, 6 and 8, and (3) $\hat{\Delta}_T(i) \xrightarrow{p} \Delta^{(i)}_0$, $\forall i$ and $\hat{\pi} \xrightarrow{p} \pi^0$ and so

$$T^{-1}\sum_1 \{d_t(\hat{\pi})\}^2 = T^{-1}\sum_1 \{d_t(\pi^0)\}^2 + \xi''_{1,T} \tag{96}$$

where $T^{-1}\sum_1 \{d_t(\pi^0)\}^2$ is defined in (93) and $\xi''_{1,T} = (\hat{\beta}_k - \beta^0_j)'\{T^{-1}\sum_1 \hat{\Delta}(t,T)'z_t z_t'\hat{\Delta}(t,T) - \Delta^{(i+1)\prime}_0 T^{-1}\sum_1 z_t z_t'\Delta^{(i+1)}_0\}(\hat{\beta}_k - \beta^0_j)$ and $\xi''_{1,T} = o_p(1)$.

Similarly,

$$T^{-1}\sum_2 \{d_t(\hat{\pi})\}^2 \;=\; T^{-1}\sum_2 \{d_t(\pi^0)\}^2 + \xi_{2,T}^{''} \tag{97}$$

where $T^{-1}\sum_2 \{d_t(\pi^0)\}^2$ is defined in (94) and $\xi_{2,T}^{''} \;=\; (\hat{\beta}_k - \beta_{j+1}^0)'\{T^{-1}\sum_2 \hat{\Delta}(t,T)' z_t z_t' \hat{\Delta}(t,T) - \Delta_0^{(i+1)'} T^{-1}\sum_2 z_t z_t' \Delta_0^{(i+1)}\}(\hat{\beta}_k \;-\; \beta_{j+1}^0)$ and under our assumptions $\xi_{2,T}^{''} \;=\; o_p(1)$.

Combining (85), (93), (94), (95), (96) and (97), we have

$$T^{-1}\sum \{d_t(\hat{\pi})\}^2 \quad\geq\quad C\{\|\Delta_0^{(i+1)}(\hat{\beta}_k - \beta_j^0)\|^2 \;+\; \|\Delta_0^{(i+1)}(\hat{\beta}_k \;-\; \beta_{j+1}^0)\|^2 \;+\; \xi_T^{''}\}$$

where $\xi_T{''} \;=\; \xi_T^* + \xi_{1,T}^{''} + \xi_{2,T}^{''}$ which because the analysis is premised on an event with probability $\bar{\epsilon}$ yields the desired result.

Proof of Theorem 3:

Suppose that $\hat{\lambda}_j \stackrel{p}{\not\to} \lambda_j^0$ for some j. In this case it follows from (12) and Lemma A.2 that with probability $\bar{\epsilon} > 0$:

- Case 1: If for some i, $\pi_i^0 < \lambda_j^0 < \pi_{i+1}^0$

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

- Case 2: If $\pi_i^0 \;=\; \lambda_j^0$ for some i

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

Thus, we have

- Case 1: Assumption 9 and $\beta_j^0 \neq \beta_{j+1}^0$ implies $\|\Delta_0^{(i+1)}(\beta_j^0 - \beta_{j+1}^0)\|^2 > 0$, which gives the result as in the proof of Theorem 1.

- Case 2: Now as $\beta_j^0 \neq \beta_{j+1}^0$ and $\Delta_0^{(i)}$, $\Delta_0^{(i+1)}$ are rank p from Assumption 9, it must follow that $\|\Delta_0^{(i)}(\hat{\beta}_k - \beta_j^0)\|^2 \;+\; \|\Delta_0^{(i+1)}(\hat{\beta}_k - \beta_{j+1}^0)\|^2 > 0$ with probability one, which gives the result via the same argument as in Theorem 1.

Proof of Theorem 4:

The general proof strategy is the same as that for Theorem 2. Again, we assume (without loss of generality) that there are only 3 break points, that is $m = 3$, and present the proof for the middle break fraction, $\hat{\lambda}_2$.

Define V_ϵ and $V_\epsilon(C)$ as in the proof of theorem 2. Using the same logic as the proof of Theorem 2, it suffices to consider the behaviour of $S_T(T_1, T_2, T_3)$ over V_ϵ for which $|T_i - T_i^0| < \epsilon T$ for all i. As before, we restrict attention to the case in which $T_2 < T_2^0$. The desired result can be established if it can be shown that for each $\eta > 0$, there exists $C > 0$ and $\epsilon > 0$ such that for large T,

$$P(min\{[S_T(T_1, T_2, T_3) - S_T(T_1, T_2^0, T_3)]/(T_2^0 - T_2)\} \; < \; 0) < \eta \tag{98}$$

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

It is possible to follow the same steps as in the proof of Theorem 2 to show that

$$\frac{SSR_1 - SSR_2}{T_2^0 - T_2} \;\geq\; 2^{-1}(\beta_3^0 - \beta_2^0)' \left[W_\triangle' W_\triangle/(T_2^0 - T_2)\right] (\beta_3^0 - \beta_2^0) - \epsilon O_p(1) - \rho O_p(1) \tag{99}$$

with large probability. We now show that the first term on the right hand side of (65) dominates. Using $\hat{\Delta}(t, T) \;=\; \sum_{i=1}^{m+1} \hat{\Delta}_T^{(i)} \mathcal{I}\{t/T \in (\hat{\pi}_{i-1}, \hat{\pi}_i]\}$, we have

$$(T_2^0 - T_2)^{-1} W_\triangle' W_\triangle \;=\; (T_2^0 - T_2)^{-1} \sum_{t=T_2+1}^{T_2^0} \hat{\Delta}(t, T)' z_t z_t' \hat{\Delta}(t, T) \tag{100}$$

To facilitate the proof, we assume that $T\pi_i^0 \in (T_2,\, T_2^0)$ but $T\pi_k^0 \notin (T_2,\, T_2^0)$ for all $k \neq i$.[23] From Assumptions 3, 6, 8 and 9, it follows that

$$\begin{aligned}(T_2^0 - T_2)^{-1} \sum_{t=T_2+1}^{T_2^0} \hat{\Delta}(t, T)' z_t z_t' \hat{\Delta}(t, T) \;\;=\;\; & (T_2^0 - T_2)^{-1} \Delta_0^{(i)'} \sum_1 z_t z_t' \Delta_0^{(i)} \\ & + (T_2^0 - T_2)^{-1} \Delta_0^{(i+1)'} \sum_2 z_t z_t' \Delta_0^{(i+1)} \\ & + o_p(1)\end{aligned} \tag{101}$$

where $\sum_1$ extends over $\{T_2 + 1 \;\leq\; t \;\leq\; T_i^*\}$ and $\sum_2$ extends over $\{T_i^* + 1 \;\leq\; t \;\leq\; T_2^0\}$.

Now, it follows from Assumptions 9 and 11 that the first and second terms on the right hand side of (101) are bounded away from zero as follows:

$$(T_2^0 - T_2)^{-1} \Delta_0^{(i)'} \sum_1 z_t z_t' \Delta_0^{(i)} \;\;\geq\;\; \alpha_1 \gamma_1 \|\Delta_0^{(i)}\|^2 \;>\; 0 \tag{102}$$

$$\Delta_0^{(i+1)'} \sum_2 z_t z_t' \Delta_0^{(i+1)} \;\;\geq\;\; \alpha_2 \gamma_2 \|\Delta_0^{(i+1)}\|^2 \;>\; 0 \tag{103}$$

where γ_1 and γ_2 are the smallest eigenvalues of $(T_i^* - T_2)^{-1} \sum_1 z_t z_t'$ and $(T_2^0 - T_i^*)^{-1} \sum_2 z_t z_t'$, respectively, and $\alpha_1 = (T_i^* - T_2)/(T_2^0 - T_2)$, $\alpha_2 = (T_2^0 - T_i^*)/(T_2^0 - T_2)$. Therefore, combining

[23] The proof is easily modified to handle other scenarios regarding the location of the break points.

(101)-(103), we obtain

$$
\begin{aligned}
(T_2^0 - T_2)^{-1} \sum_{t=T_2+1}^{T_2^0} \hat{\Delta}(t,T)' z_t z_t' \hat{\Delta}(t,T) \;&\geq\; \alpha_1\gamma_1 \|\Delta_0^i\|^2 \,+\, \alpha_2\gamma_2 \|\Delta_0^{(i+1)}\|^2 \,+\, o_p(1) \\
&\geq\; \min\{\alpha_1\gamma_1,\, \alpha_2\gamma_2\} \left(\|\Delta_0^{(i)}\|^2 \,+\, \|\Delta_0^{(i+1)}\|^2 \right) \\
&\quad + o_p(1) \qquad (104)
\end{aligned}
$$

From Assumptions 9 and 11, it follows that the first term on the right hand side of (104) is bounded away from zero on $V_\epsilon(C)$ with large probability. Therefore, the first term on the right hand side of (99) dominates and is positive for large C, small ϵ and large T which in turn proves (98).

Proof of Theorem 5:

For notational brevity, set $\hat{\beta} = \hat{\beta}(\{\hat{T}_i\}_{i=1}^m)$. By definition, the 2SLS estimator is

$$
\hat{\beta} \;=\; \left(\bar{W}^{*'} \bar{W}^* \right)^{-1} \bar{W}^{*'} Y \qquad (105)
$$

From (26), it follows that

$$
Y \;=\; \bar{W}^* \beta^0 + U^* \qquad (106)
$$

where $U^* = (\bar{W}^0 - \bar{W}^*)\beta^0 + \tilde{U}$. Substituting in (105) for Y from (106), we obtain

$$
\hat{\beta} \;=\; \beta^0 + (\bar{W}^{*'} \bar{W}^*)^{-1} \bar{W}^{*'} U^*
$$

and hence that

$$
\hat{\beta} - \beta^0 \;=\; (\bar{W}^{*'} \bar{W}^*)^{-1} \bar{W}^{*'} [(\bar{W}^0 - \bar{W}^*)\beta^0 + \tilde{U}] \qquad (107)
$$

From (107) it follows that

$$
T^{1/2}(\hat{\beta} - \beta^0) \;=\; \left(T^{-1} \bar{W}^{*'} \bar{W}^* \right)^{-1} T^{-1/2} \bar{W}^{*'} [\tilde{U} + (\bar{W}^0 - \bar{W}^*)\beta^0] \qquad (108)
$$

Theorem 2 implies that $\hat{T}_i - T_i^0 = O_p(1)$ for all i. Therefore, the summation $\bar{W}^{*'} \bar{W}^0 - \bar{W}^{*'} \bar{W}^*$ involves a bounded number of terms with probability one, and so

$$
\bar{W}^{*'} \bar{W}^0 - \bar{W}^{*'} \bar{W}^* \;=\; O_p(1) \qquad (109)
$$

Hence, it follows that

$$
T^{-1/2} \| \bar{W}^{*'} \bar{W}^0 - \bar{W}^{*'} \bar{W}^* \| \;=\; o_p(1) \qquad (110)
$$

and taken together (108)-(110) imply that

$$T^{1/2}(\hat{\beta}-\beta^0) \;=\; \left(T^{-1}\bar{W}^{*\prime}\bar{W}^{*}\right)^{-1} T^{-1/2}\bar{W}^{*\prime}\tilde{U} \;+\; o_p(1) \tag{111}$$

The addition and subtraction of $\left(T^{-1}\bar{W}^{0\prime}\bar{W}^{0}\right)^{-1} T^{-1/2}\bar{W}^{0\prime}\tilde{U}$ to the right hand side of (111) and some rearrangement yields

$$\begin{aligned} T^{1/2}(\hat{\beta}-\beta^0) \;\;=\;\; & \left(T^{-1}\bar{W}^{0\prime}\bar{W}^{0}\right)^{-1} T^{-1/2}\bar{W}^{0\prime}\tilde{U} \\ & + \left(T^{-1}\bar{W}^{0\prime}\bar{W}^{0}\right)^{-1}\left(T^{-1}\bar{W}^{0\prime}\bar{W}^{0} \;-\; T^{-1}\bar{W}^{*\prime}\bar{W}^{*}\right)\left(T^{-1}\bar{W}^{*\prime}\bar{W}^{*}\right)^{-1} T^{-1/2}\bar{W}^{0\prime}\tilde{U} \\ & + \left(T^{-1}\bar{W}^{*\prime}\bar{W}^{*}\right)^{-1} T^{-1/2}(\bar{W}^{*\prime} \;-\; \bar{W}^{0\prime})\tilde{U} \;+\; o_p(1) \end{aligned} \tag{112}$$

Using a similar argument to (109), it follows that

$$\|T^{-1}\bar{W}^{0\prime}\bar{W}^{0} \;-\; T^{-1}\bar{W}^{*\prime}\bar{W}^{*}\| \;\;=\;\; o_p(1) \tag{113}$$

$$\|T^{-1/2}(\bar{W}^{*\prime} \;-\; \bar{W}^{0\prime})\tilde{U}\| \;\;=\;\; o_p(1) \tag{114}$$

Using the Triangle inequality, equations (113)-(114), Assumptions 4 and 6, and the property of the matrix norm given in (30), it follows from (112) that

$$T^{1/2}(\hat{\beta}-\beta^0) \;=\; \left(T^{-1}\bar{W}^{0\prime}\bar{W}^{0}\right)^{-1} T^{-1/2}\bar{W}^{0\prime}\tilde{U} \;+\; o_p(1) \tag{115}$$

Given the block diagonal structure of $\bar{W}^{0\prime}\bar{W}^{0}$, the coefficient vector of the $i-th$ regime can be written as

$$T^{1/2}\left(\hat{\beta}_i-\beta_i^0\right) \;=\; \left(\frac{1}{T}\sum_{i_0}\hat{x}_t\hat{x}_t'\right)^{-1} T^{-1/2}\sum_{i_0}\hat{x}_t\tilde{u}_t + o_p(1) \tag{116}$$

The result then follows directly from (116), $\hat{x}_t = X'Z(Z'Z)^{-1}z_t$ and Assumptions 3 and 6.

Proof of Theorem 6:

The F-statistic can then be written as

$$F_T(\lambda_1,...,\lambda_k;p) \;=\; F_T^*/[kp(T-(k+1)p)^{-1}SSR_k] \tag{117}$$

where $F_T^* = SSR_0 - SSR_k$. We first consider the limiting behaviour of F_T^*. To this end, we define: $D^R(i,j)$ to be the sum of the squared residuals from the restricted model using data from segments i to j, that is the observations from $T_{i-1}+1$ to T_j; $D^U(i,j)$ to be the corresponding for the unrestricted model. Using this notation, we can write F_T^* as follows:[24]

$$
\begin{aligned}
F_T^* &= D^R(1,k+1) - \sum_{i=1}^{k+1} D^U(i,i) \\
&= \sum_{i=1}^{k}[D^R(1,i+1) - D^R(1,i) - D^U(i+1,i+1)] + D^R(1,1) - D^U(1,1) \\
&= \sum_{i=1}^{k}[D^R(1,i+1) - D^R(1,i) - D^U(i+1,i+1)] \qquad (118) \\
&= \sum_{i=1}^{k} F_{T,i}, \text{ say.} \qquad (119)
\end{aligned}
$$

To analyze the behaviour of the terms on the right hand side of (118), it is necessary to introduce the following leasts squares estimators in addition to (what can now be termed) the unrestricted estimator given in (105):

- The restricted estimator based on the full sample is

$$\hat{\beta}^R = (W'W)^{-1}W'Y \qquad (120)$$

- The least squares estimator of the common regression parameter under H_0 based on segments 1 through j of the partition,

$$\hat{\beta}^R_{1,j} = (W'_{1,j}W_{1,j})^{-1}W'_{1,j}Y_{1,j} \qquad (121)$$

  where $Y_{1,j}$, $\tilde{U}_{1,j}$, $W_{1,j}$ denote the matrices (vectors) consisting of the rows 1 through T_j of Y, $\tilde{U}$, W, respectively.

- The least squares estimator based on the observations in the j^{th} segment of the partition,

$$\hat{\beta}^U_j = (W'_jW_j)^{-1}W'_jY_j \qquad (122)$$

  where Y_j, U_j, W_j be the matrices (vectors) containing rows $T_{j-1}+1$ through T_j of Y, $\tilde{U}$, W, respectively.

Note that under the null hypothesis that $\beta_i^0 = \bar{\beta}_0$ in (1) for $i = 1, 2, \ldots, k+1$, we have

$$
\begin{aligned}
Y &= W\bar{\beta}_0 + \tilde{U} \qquad (123) \\
&= \bar{W}^0(\iota_{k+1} \otimes \bar{\beta}_0) + \tilde{U} \qquad (124) \\
Y_j &= W_j\bar{\beta}_0 + \tilde{U}_j \qquad (125)
\end{aligned}
$$

[24]Note that the unrestricted and restricted models are the same on segment (i,i) for any i.

where ι_{k+1} is a $(k+1)\times 1$ vector of ones. Using (120)-(125), it can be shown that

$$D^R(1,j) = ||(I-P_{W_{1,j}})\tilde{U}_{1,j}||^2 \tag{126}$$

$$D^U(j,j) = ||(I-P_{W_j})\tilde{U}_j||^2 \tag{127}$$

where $P_{W_{1,j}} = W_{1,j}(W_{1,j}^{'}W_{1,j})^{-1}W_{1,j}^{'}$ and $P_{W_j} = W_j(W_j^{'}W_j)^{-1}W_j^{'}$. Now consider $F_{T,i}$ in (119). From (118), (126) and (127), it follows that

$$\begin{aligned} F_{T,i} &= ||(I-P_{W_{1,i+1}})\tilde{U}_{1,i+1}||^2 - ||(I-P_{W_{1,i}})\tilde{U}_{1,i}||^2 - ||(I-P_{W_{i+1}})\tilde{U}_{i+1}||^2 \\ &= -S_{i+1}'H_{i+1}^{-1}S_{i+1} + S_i'H_i^{-1}S_i + (S_{i+1}-S_i)'(H_{i+1}-H_i)^{-1}(S_{i+1}-S_i) \end{aligned} \tag{128}$$

where $S_j = W_{1,j}'\tilde{U}_{1,j}$ and $H_j = W_{1,j}'W_{1,j}$.

The limiting behaviour of $F_{T,i}$ is deduced from the limiting behaviour of S_j and H_j. To proceed further, it is useful to explore further the implications of (16). Let $B(r) = [B_1(r)^{'}, B_2(r)^{'}, \ldots, B_{p+1}(r)^{'}]^{'}$ where $B_i(r)^{'}$ is $q \times 1$, and

$$\Omega^{1/2} = \begin{bmatrix} N_1^{'} \\ N_2^{'} \end{bmatrix} \tag{129}$$

where $N_1^{'}$ is a $1\times(p+1)$ vector whose i^{th} element is $N_{1,i}$, and $N_2^{'}$ is $p\times(p+1)$. Note that, since $\Omega^{1/2}$ is symmetric,

$$\Omega = \begin{bmatrix} N_1^{'}N_1 & N_1^{'}N_2 \\ N_2^{'}N_1 & N_2^{'}N_2 \end{bmatrix} = \begin{bmatrix} \sigma^2 & \gamma^{'} \\ \gamma & \Sigma \end{bmatrix} \tag{130}$$

where the second and third matrices are partitioned conformably. It follows from (16) and (129) that

$$T^{-1/2}\sum_{t=1}^{[Tr]} z_t u_t \Longrightarrow (N_1^{'} \otimes Q_{ZZ}^{1/2})B(r) \tag{131}$$

$$= \sum_{i=1}^{p+1} N_{1,i}Q_{ZZ}^{1/2}B_i(r)$$

$$= Q_{ZZ}^{1/2}\sum_{i=1}^{p+1} N_{1,i}B_i(r) \tag{132}$$

$$= Q_{ZZ}^{1/2}\tilde{D}^*(r), \text{ say} \tag{133}$$

and

$$T^{-1/2}\sum_{t=1}^{[Tr]} v_t \otimes z_t = T^{-1/2}\sum_{t=1}^{[Tr]} vec(z_t v_t^{'})$$

$$\Longrightarrow (N_2^{'} \otimes Q_{ZZ}^{1/2})B(r) \tag{134}$$

Note that (134) implies

$$T^{-1/2}\sum_{t=1}^{[Tr]} z_t v_t^{'} \implies Q_{ZZ}^{1/2} B^{mat}(r) N_2 \tag{135}$$

$$= Q_{ZZ}^{1'2} D^{*}(r), \text{ say} \tag{136}$$

where $vec(B^{mat}(r)) = B(r)$.

To deduce the limiting behaviour of S_j, we note that (37) (with $\beta_i^0 = \bar{\beta}_0$) implies:

$$\begin{aligned} T^{-1/2}S_j &= \Delta_0{}' T^{-1/2}\sum_{t=1}^{[T\lambda_j]} z_t u_t + V'Z(Z'Z)^{-1}T^{-1/2}\sum_{t=1}^{[T\lambda_j]} z_t u_t + \Delta_0{}' T^{-1/2}\sum_{t=1}^{[T\lambda_j]} z_t v_t{}' \bar{\beta}_0 \\ &\quad + V'Z(Z'Z)^{-1}T^{-1/2}\sum_{t=1}^{[T\lambda_j]} z_t v_t{}'\bar{\beta}_0 - \Delta_0{}' \sum_{t=1}^{[T\lambda_j]} z_t z_t'(Z'Z)^{-1}T^{-1/2}Z'V\bar{\beta}_0 \\ &\quad - T^{-1/2}V'Z(Z'Z)^{-1}\sum_{t=1}^{[T\lambda_j]} z_t z_t'(Z'Z)^{-1}Z'V\bar{\beta}_0 \end{aligned} \tag{137}$$

Under Assumptions 3-6 and 12, it follows from (133), (136) and (137) that[25]

$$T^{-1/2}S_j \implies \Delta_0{}' Q_{ZZ}^{1/2}\tilde{D}^{*}(\lambda_j) + \Delta_0' Q_{ZZ}^{1/2} D^{*}(\lambda_j)\bar{\beta}_0 - \Delta_0{}'\lambda_j Q_{ZZ}^{1/2} D^{*}(1)\bar{\beta}_0 \tag{138}$$

Similarly, we have

$$\begin{aligned} T^{-1}H_j &= T^{-1}W_{1,j}'W_{1,j} = T^{-1}\sum_{t=1}^{[T\lambda_j]} \hat{x}_t \hat{x}_t' \\ &= T^{-1}\sum_{t=1}^{[T\lambda_j]} \hat{\Delta}_T' z_t (\hat{\Delta}_T' z_t)' \end{aligned} \tag{139}$$

Under Assumptions 3-6, it follows from (139) that

$$T^{-1}H_j \xrightarrow{p} \Delta_0{}'(\lambda_j Q_{ZZ})\Delta_0 \tag{140}$$

We now use (138)-(140) to deduce the limiting behaviour of the terms on the right hand side

[25] See Han (2006).

of (128). First consider $S'_{i+1}H^{-1}_{i+1}S_{i+1}$. From (138)-(140), we have

$$\begin{aligned}
S'_{i+1}H^{-1}_{i+1}S_{i+1} \quad &\Longrightarrow \quad (\Delta_0{}'Q^{1/2}_{ZZ}\tilde{D}^*(\lambda_{i+1}) + \Delta_0{}'Q^{1/2}_{ZZ}D^*(\lambda_{i+1})\bar{\beta}_0 - \Delta_0{}'\lambda_{i+1}Q^{1/2}_{ZZ}D^*(1)\bar{\beta}_0)' \times \\
&\quad\quad (\Delta_0{}'\lambda_{i+1}Q_{ZZ}\Delta_0)^{-1} \times \\
&\quad\quad (\Delta_0{}'Q^{1/2}_{ZZ}\tilde{D}^*(\lambda_{i+1}) + \Delta_0{}'Q^{1/2}_{ZZ}D^*(\lambda_{i+1})\bar{\beta}_0 - \Delta_0{}'\lambda_{i+1}Q^{1/2}_{ZZ}D^*(1)\bar{\beta}_0) \\
&= \quad (\Delta_0{}'Q^{1/2}_{ZZ}[\tilde{D}^*(\lambda_{i+1}) + D^*(\lambda_{i+1})\bar{\beta}_0 - \lambda_{i+1}D^*(1)\bar{\beta}_0])'(\Delta_0{}'\lambda_{i+1}Q_{ZZ}\Delta_0)^{-1} \\
&\quad\quad \times(\Delta_0{}'Q^{1/2}_{ZZ}[\tilde{D}^*(\lambda_{i+1}) + D^*(\lambda_{i+1})\bar{\beta}_0 - \lambda_{i+1}D^*(1)\bar{\beta}_0]) \\
&= \quad \lambda^{-1}_{i+1}[\tilde{D}^*(\lambda_{i+1}) + D^*(\lambda_{i+1})\bar{\beta}_0 - \lambda_{i+1}D^*(1)\bar{\beta}_0]'(\Delta_0{}'Q^{1/2}_{ZZ})'(\Delta_0{}'Q_{ZZ}\Delta_0)^{-1} \\
&\quad\quad \times(\Delta_0{}'Q^{1/2}_{ZZ})[\tilde{D}^*(\lambda_{i+1}) + D^*(\lambda_{i+1})\bar{\beta}_0 - \lambda_{i+1}D^*(1)\bar{\beta}_0] \qquad (141)
\end{aligned}$$

To simplify (141) note that $(\Delta_0{}'Q^{1/2}_{ZZ})'(\Delta_0{}'Q_{ZZ}\Delta_0)^{-1}(\Delta_0{}'Q^{1/2}_{ZZ})$ is a projection matrix which, from Assumptions 4 and 6 is of rank p. It follows

$$\begin{aligned}
(\Delta_0{}'Q^{1/2}_{ZZ})'(\Delta_0{}'Q_{ZZ}\Delta_0)^{-1}(\Delta_0{}'Q^{1/2}_{ZZ}) \quad &= \quad C'\Lambda C = C'\Lambda'\Lambda C \qquad (142) \\
&= \quad (\Lambda C)'\Lambda C \qquad (143)
\end{aligned}$$

where C is an orthogonal matrix and Λ is a diagonal matrix, p of whose diagonal elements are one with the remaining $q-p$ equal to zero. Substituting (142) in (141) and using (143), we obtain

$$\begin{aligned}
S'_{i+1}H^{-1}_{i+1}S_{i+1} \quad &\Longrightarrow \quad \lambda^{-1}_{i+1}(\tilde{D}^*(\lambda_{i+1}) + [D^*(\lambda_{i+1}) - \lambda_{i+1}D^*(1)]\bar{\beta}_0)'C'\Lambda C \\
&\quad\quad \times(\tilde{D}^*(\lambda_{i+1}) + [D^*(\lambda_{i+1}) - \lambda_{i+1}D^*(1)]\bar{\beta}_0) \\
&= \quad \lambda^{-1}_{i+1}(\tilde{D}^*(\lambda_{i+1}) + [D^*(\lambda_{i+1}) - \lambda_{i+1}D^*(1)]\bar{\beta}_0)'(\Lambda C)'\Lambda C \\
&\quad\quad \times(\tilde{D}^*(\lambda_{i+1}) + [D^*(\lambda_{i+1}) - \lambda_{i+1}D^*(1)]\bar{\beta}_0) \\
&= \quad \lambda^{-1}_{i+1}(\Lambda C\tilde{D}^*(\lambda_{i+1}) + \Lambda C[D^*(\lambda_{i+1}) - \lambda_{i+1}D^*(1)]\bar{\beta}_0)' \\
&\quad\quad \times(\Lambda C\tilde{D}^*(\lambda_{i+1}) + \Lambda C[D^*(\lambda_{i+1}) - \lambda_{i+1}D^*(1)]\bar{\beta}_0) \qquad (144)
\end{aligned}$$

Similar logic yields

$$\begin{aligned}
S'_iH^{-1}_iS_i \quad &\Longrightarrow \quad \lambda^{-1}_i(\Lambda C\tilde{D}^*(\lambda_i) + \Lambda C[D^*(\lambda_i) - \lambda_iD^*(1)]\bar{\beta}_0)' \\
&\quad\quad \times(\Lambda C\tilde{D}^*(\lambda_i) + \Lambda C[D^*(\lambda_i) - \lambda_iD^*(1)]\bar{\beta}_0) \qquad (145)
\end{aligned}$$

Now consider $A_i = (S_{i+1} - S_i)'(H_{i+1} - H_i)^{-1}(S_{i+1} - S_i)$. Using (138)-(140), it follows that

$$\begin{aligned}
A_i \quad \Longrightarrow \quad & [\Delta_0{}' Q_{ZZ}^{1/2}(\tilde{D}^*(\lambda_{i+1}) - \tilde{D}^*(\lambda_i)) + \Delta_0{}' Q_{ZZ}^{1/2}(D^*(\lambda_{i+1}) - D^*(\lambda_i))\bar{\beta}_0 \\
& -\Delta_0{}'(\lambda_{i+1} - \lambda_i) Q_{ZZ}^{1/2} D^*(1)\bar{\beta}_0]'[\Delta_0{}'(\lambda_{i+1} - \lambda_i) Q_{ZZ}\Delta_0]^{-1}[\Delta_0{}' Q_{ZZ}^{1/2}(\tilde{D}^*(\lambda_{i+1}) - \tilde{D}^*(\lambda_i)) \\
& +\Delta_0{}' Q_{ZZ}^{1/2}(D^*(\lambda_{i+1}) - D^*(\lambda_i))\bar{\beta}_0 - \Delta_0{}'(\lambda_{i+1} - \lambda_i) Q_{ZZ}^{1/2} D^*(1)\bar{\beta}_0] \\
= \quad & (\Delta_0{}' Q_{ZZ}^{1/2}[\tilde{D}^*(\lambda_{i+1}) - \tilde{D}^*(\lambda_i) + (D^*(\lambda_{i+1}) - D^*(\lambda_i))\bar{\beta}_0 - (\lambda_{i+1} - \lambda_i) D^*(1)\bar{\beta}_0])' \\
& \times[\Delta_0{}'(\lambda_{i+1} - \lambda_i) Q_{ZZ}\Delta_0]^{-1}(\Delta_0{}' Q_{ZZ}^{1/2}[\tilde{D}^*(\lambda_{i+1}) - \tilde{D}^*(\lambda_i) + (D^*(\lambda_{i+1}) - D^*(\lambda_i))\bar{\beta}_0 \\
& -(\lambda_{i+1} - \lambda_i) D^*(1)\bar{\beta}_0]) \\
= \quad & [\tilde{D}^*(\lambda_{i+1}) - \tilde{D}^*(\lambda_i) + (D^*(\lambda_{i+1}) - D^*(\lambda_i))\bar{\beta}_0 - (\lambda_{i+1} - \lambda_i) D^*(1)\bar{\beta}_0]'(\Delta_0{}' Q_{ZZ}^{1/2})' \\
& \times(\lambda_{i+1} - \lambda_i)^{-1}[\Delta_0{}' Q_{ZZ}\Delta_0]^{-1}(\Delta_0{}' Q_{ZZ}^{1/2})[\tilde{D}^*(\lambda_{i+1}) - \tilde{D}^*(\lambda_i) + (D^*(\lambda_{i+1}) - D^*(\lambda_i))\bar{\beta}_0 \\
& -(\lambda_{i+1} - \lambda_i) D^*(1)\bar{\beta}_0] \\
= \quad & (\lambda_{i+1} - \lambda_i)^{-1}[\tilde{D}^*(\lambda_{i+1}) - \tilde{D}^*(\lambda_i) + (D^*(\lambda_{i+1}) - D^*(\lambda_i))\bar{\beta}_0 - (\lambda_{i+1} - \lambda_i) D^*(1)\bar{\beta}_0]' \\
& \times C'\Lambda C[\tilde{D}^*(\lambda_{i+1}) - \tilde{D}^*(\lambda_i) + (D^*(\lambda_{i+1}) - D^*(\lambda_i))\bar{\beta}_0 - (\lambda_{i+1} - \lambda_i) D^*(1)\bar{\beta}_0] \\
= \quad & (\lambda_{i+1} - \lambda_i)^{-1}[\Lambda C(\tilde{D}^*(\lambda_{i+1}) - \tilde{D}^*(\lambda_i)) + \Lambda C(D^*(\lambda_{i+1}) - D^*(\lambda_i))\bar{\beta}_0 \\
& -(\lambda_{i+1} - \lambda_i)\Lambda C D^*(1)\bar{\beta}_0]'[\Lambda C(\tilde{D}^*(\lambda_{i+1}) - \tilde{D}^*(\lambda_i)) + \Lambda C(D^*(\lambda_{i+1}) - D^*(\lambda_i))\bar{\beta}_0 \\
& -(\lambda_{i+1} - \lambda_i)\Lambda C D^*(1)\bar{\beta}_0] \\
= \quad & (\lambda_{i+1} - \lambda_i)^{-1}[\Lambda C(\tilde{D}^*(\lambda_{i+1}) - \tilde{D}^*(\lambda_i)) + \Lambda C(D^*(\lambda_{i+1}) - D^*(\lambda_i) - \lambda_{i+1} D^*(1) \\
& +\lambda_i D^*(1))\bar{\beta}_0]'[\Lambda C(\tilde{D}^*(\lambda_{i+1}) - \tilde{D}^*(\lambda_i)) + \Lambda C(D^*(\lambda_{i+1}) - D^*(\lambda_i) - \lambda_{i+1} D^*(1) \\
& + \lambda_i D^*(1))\bar{\beta}_0] \qquad (146)
\end{aligned}$$

We now use (144)-(146) to deduce the limiting behaviour of $F_{T,i}$. To this end, we now write $D_i = \Lambda C D^*(\lambda_i)$, $\tilde{D}_i = \Lambda C \tilde{D}^*(\lambda_i)$ and $D_1 = \Lambda C D^*(1)$.

From (144)-(146) it follows that

$$\begin{aligned}
F_{T,i} \quad \Longrightarrow \quad & \lambda_i^{-1}[\tilde{D}_i + (D_i - \lambda_i D_1)\bar{\beta}_0]'[\tilde{D}_i + (D_i - \lambda_i D_1)\bar{\beta}_0] \\
& -\lambda_{i+1}^{-1}[\tilde{D}_{i+1} + (D_{i+1} - \lambda_{i+1} D_1)\bar{\beta}_0]'[\tilde{D}_{i+1} + (D_{i+1} - \lambda_{i+1} D_1)\bar{\beta}_0] \\
& +(\lambda_{i+1} - \lambda_i)^{-1}[(\tilde{D}_{i+1} - \tilde{D}_i) + (D_{i+1} - \lambda_{i+1} D_1 - D_i + \lambda_i D_1)\bar{\beta}_0]' \\
& \times[(\tilde{D}_{i+1} - \tilde{D}_i) + (D_{i+1} - \lambda_{i+1} D_1 - D_i + \lambda_i D_1)\bar{\beta}_0] \qquad (147)
\end{aligned}$$

Multiplying out (147) and rearranging terms, we obtain[26]

$$F_{T,i} \implies \{\lambda_i \lambda_{i+1}(\lambda_{i+1} - \lambda_i)\}^{-1} ||[\lambda_{i+1}\tilde{D}_i - \lambda_i \tilde{D}_{i+1}] + [\lambda_{i+1} D_i - \lambda_i D_{i+1}]\bar{\beta}_0||^2 \tag{148}$$

It follows from (148) that the limiting behaviour of the numerator of $F_T(\lambda_1, ..., \lambda_k; p)$ is given by:

$$F_T^* \implies \sum_{i=1}^{k} \{\lambda_i \lambda_{i+1}(\lambda_{i+1} - \lambda_i)\}^{-1} ||[\lambda_{i+1}\tilde{D}_i - \lambda_i \tilde{D}_{i+1}] + [\lambda_{i+1} D_i - \lambda_i D_{i+1}]\bar{\beta}_0||^2 \tag{149}$$

Now, consider the denominator of $F_T(\lambda_1, ..., \lambda_k; p)$. Using (124)-(127), it can be shown that

$$\begin{aligned} SSR_k &= \sum_{i=1}^{k} D^U(i,i) \\ &= \sum_{i=1}^{k} ||(I - P_{W_i})\tilde{U}_i||^2 \\ &= \sum_{i=1}^{k} \tilde{U}_i' \tilde{U}_i - \sum_{i=1}^{k} \tilde{U}_i' P_{W_i} \tilde{U}_i \end{aligned} \tag{150}$$

From (150) it follows that

$$(T - (k+1)p)^{-1} SSR_k = (T - (k+1)p)^{-1} \sum_{i=1}^{k} \tilde{U}_i' \tilde{U}_i - (T - (k+1)p)^{-1} \sum_{i=1}^{k} \tilde{U}_i' P_{W_i} \tilde{U}_i \tag{151}$$

We now consider the limiting behaviour of the terms on the right hand side of (151) in turn. Since

$$\tilde{U}_i' P_{W_i} \tilde{U}_i = (S_i - S_{i-1})'(H_i - H_{i-1})^{-1}(S_i - S_{i-1}) \tag{152}$$

it follows from (146) that

$$\begin{aligned} \tilde{U}_i' P_{W_i} \tilde{U}_i \implies & (\lambda_i - \lambda_{i-1})^{-1}[\tilde{D}_i - \tilde{D}_{i-1} + (D_i - D_{i-1} - \lambda_i D_1 + \lambda_{i-1} D_1)\bar{\beta}_0]' \\ & \times [\tilde{D}_i - \tilde{D}_{i-1} + (D_i - D_{i-1} - \lambda_i D_1 + \lambda_{i-1} D_1)\bar{\beta}_0] \end{aligned} \tag{153}$$

and hence that $(T - (k+1)p)^{-1} \sum_{i=1}^{k} \tilde{U}_i' P_{W_i} \tilde{U}_i = o_p(1)$.

Now consider $(T - (k+1)p)^{-1} \sum_{i=1}^{k} \tilde{U}_i' \tilde{U}_i$. From (36), it follows that under the null hypothesis of no breaks,

$$\begin{aligned} \tilde{U}_i' \tilde{U}_i &= \sum_{t=[T\lambda_{i-1}]+1}^{[T\lambda_i]} \tilde{u}_t^2 \\ &= \sum_i [(u_t + v_t{}'\bar{\beta}_0) - z_t'(Z'Z)^{-1} Z' V \bar{\beta}_0]^2 \\ &= \sum_i \{(u_t + v_t{}'\bar{\beta}_0)^2 + (z_t'(Z'Z)^{-1} Z' V \bar{\beta}_0)^2 - 2(u_t + v_t{}'\bar{\beta}_0) z_t'(Z'Z)^{-1} Z' V \bar{\beta}_0\} \end{aligned} \tag{154}$$

[26] See Han (2006).

Since

$$\sum_i (z_t'(Z'Z)^{-1}Z'V\bar{\beta}_0)^2 \;=\; \bar{\beta}_0' T^{-1/2}V'Z(T^{-1}Z'Z)^{-1}T^{-1}\sum_i z_t z_t'(T^{-1}Z'Z)^{-1}T^{-1/2}Z'V\bar{\beta}_0$$
$$= O_p(1) \tag{155}$$

and

$$\sum_i (u_t + v_t{}'\bar{\beta}_0)z_t'(Z'Z)^{-1}Z'V\bar{\beta}_0 \;=\; T^{-1/2}\sum_i u_t z_t'(T^{-1}Z'Z)^{-1}T^{-1/2}Z'V\bar{\beta}_0$$
$$+T^{-1/2}\sum_i v_t{}'\bar{\beta}_0 z_t'(T^{-1}Z'Z)^{-1}T^{-1/2}Z'V\bar{\beta}_0$$
$$= O_p(1) \tag{156}$$

it follows that

$$(T-(k+1)p)^{-1}\sum_{i=1}^{k}\tilde{U}_i'\tilde{U}_i \;=\; (T-(k+1)p)^{-1}\sum_{i=1}^{k}\sum_{[T\lambda_{i-1}]+1}^{[T\lambda_i]}(u_t + v_t{}'\bar{\beta}_0)^2 \;+\; o_p(1)$$
$$= \sigma^2 + 2\gamma'\bar{\beta}_0 + \bar{\beta}_0'\Sigma\bar{\beta}_0 \;+\; o_p(1) \tag{157}$$

From (151)-(153) and (157), it follows that

$$(T-(k+1)p)^{-1}SSR_k \xrightarrow{p} \sigma^2 + 2\gamma'\bar{\beta}_0 + \bar{\beta}_0'\Sigma\bar{\beta}_0 \tag{158}$$

Combining (117), (149) and (158), we obtain

$$F_T(\lambda_1,...,\lambda_k;p) \;\Longrightarrow\; \frac{1}{kp}\sum_{i=1}^{k}\frac{||[\lambda_{i+1}\tilde{D}_i - \lambda_i\tilde{D}_{i+1}] + [\lambda_{i+1}D_i - \lambda_i D_{i+1}]\bar{\beta}_0||^2}{\lambda_i\lambda_{i+1}(\lambda_{i+1}-\lambda_i)[\sigma^2 + 2\gamma'\bar{\beta}_0 + \bar{\beta}_0'\Sigma\bar{\beta}_0]} \tag{159}$$

We now show that this limiting distribution has the alternative representation given in Theorem 6. First notice that the limit distribution on the right hand side of (159), a_i say, can be written as

$$a_i \;=\; \frac{1}{kp}\sum_{i=1}^{k}\frac{||\lambda_{i+1}b(\lambda_i) - \lambda_i b(\lambda_{i+1})||^2}{\lambda_i\lambda_{i+1}(\lambda_{i+1}-\lambda_i)[\sigma^2 + 2\gamma'\bar{\beta}_0 + \bar{\beta}_0'\Sigma\bar{\beta}_0]} \tag{160}$$

where $b(\lambda_i) = [\tilde{D}_i + D_i\bar{\beta}_0]$.

Therefore the desired result will be established if it can be shown that

$$b(\lambda_i) \;\stackrel{d}{=}\; [\sigma^2 + 2\gamma'\bar{\beta}_0 + \bar{\beta}_0'\Sigma\bar{\beta}_0]^{1/2}\begin{bmatrix} W_i \\ 0_{(q-p)\times 1}\end{bmatrix} \tag{161}$$

where W_i is a $p \times 1$ vector of standard Brownian motion process and $0_{(q-p)\times 1}$ is a $(q-p)\times 1$ null vector and $\stackrel{d}{=}$ denotes "distributed as". We now show that (161) holds. Without loss of generality, we assume the rows of $(\Delta_0{}'Q_{ZZ}^{1/2})'(\Delta_0{}'Q_{ZZ}\Delta_0)^{-1}(\Delta_0{}'Q_{ZZ}^{1/2})$ are arranged such that $\Lambda = diag(\iota_p', 0'_{(q-p)\times 1})$ in (142) where ι_p is $p \times 1$ vector of ones. It therefore follows that

$$
\begin{aligned}
\tilde{D}_i &= \Lambda C \tilde{D}^*(\lambda_i) \;=\; \Lambda C \sum_{j=1}^{p+1} N_{1,j} B_j(\lambda_i) \;=\; \sum_{j=1}^{p+1} N_{1,j} \Lambda C B_j(\lambda_i) \\
&\overset{d}{=} \sum_{j=1}^{p+1} N_{1,j} \Lambda B_j(\lambda_i) \\
&= \begin{bmatrix} \sum_{j=1}^{p+1} N_{1,j} B_{j,1:p}(\lambda_i) \\ 0_{(q-p)\times 1} \end{bmatrix} \\
&= \begin{bmatrix} (N_1^{'} \otimes I_p) B_{1:p}(\lambda_i) \\ 0_{(q-p)\times 1} \end{bmatrix} \qquad (162)
\end{aligned}
$$

where $B_{j,1:p}(.)$ is a $p \times 1$ vector containing the first p rows of $B_j(.)$ and $B_{1:p}(.) = [B_{1,1:p}(.)^{'}, B_{2,1:p}(.)^{'}, \ldots, B_{p+1,1:p}(.)^{'}]^{'}$. It also follows using similar arguments that

$$
\begin{aligned}
D_i \bar{\beta}_0 &= \Lambda C B^{mat}(\lambda_i) N_2 \bar{\beta}_0 \\
&= (\bar{\beta}_0^{'} N_2^{'} \otimes \Lambda C) B(\lambda_i) \\
&\overset{d}{=} (\bar{\beta}_0^{'} N_2^{'} \otimes I_q) \begin{bmatrix} B_{1,1:p}(\lambda_i) \\ 0_{(q-p)\times 1} \\ B_{2,1:p}(\lambda_i) \\ 0_{(q-p)\times 1} \\ \vdots \\ B_{p+1,1:p}(\lambda_i) \\ 0_{(q-p)\times 1} \end{bmatrix} \\
&= \begin{bmatrix} (\bar{\beta}_0^{'} N_2^{'} \otimes I_p) B_{1:p}(\lambda_i) \\ 0_{(q-p)\times 1} \end{bmatrix} \qquad (163)
\end{aligned}
$$

It follows from (162) and (163) that

$$
\begin{aligned}
b(\lambda_i) &= \begin{bmatrix} \left\{ (N_1^{'} + \bar{\beta}_0^{'} N_2^{'}) \otimes I_p \right\} B_{1:p}(\lambda_i) \\ 0_{(q-p)\times 1} \end{bmatrix} \qquad (164) \\
&= \begin{bmatrix} b_1(\lambda_i) \\ 0_{(q-p)\times 1} \end{bmatrix}, \text{ say,} \qquad (165)
\end{aligned}
$$

Equation (164) proves (161) for the lower $q-p$ elements. For the remining elements, note that it follows from (164)-(165) that:

- $b_1(0) = 0_{p\times 1}$;
- For any dates $0 \le \lambda_1 < \lambda_2 < \ldots, < \lambda_n \le 1$, the changes $b_1(\lambda_2)-b_1(\lambda_1)$, $b_1(\lambda_3)-b_1(\lambda_2)$,…, $b_1(\lambda_n)-b_1(\lambda_{n-1})$ are independent multivariate Gaussian with

$$b_1(\lambda_i) - b_1(\lambda_{i-1}) \sim N\left(0_{p\times 1}, \left(\sigma^2 \;+\; 2\gamma'\bar{\beta}_0 \;+\; \bar{\beta}_0'\Sigma\bar{\beta}_0\right)(\lambda_i - \lambda_{i-1})I_p\right)$$

- For any given realization, $b_1(\lambda)$ is continuous in λ with probability one.

It follows from these three properties that[27]

$$b_1(\lambda_i) \overset{\mathrm{d}}{=} [\sigma^2 + 2\gamma'\bar{\beta}_0 + \bar{\beta}_0'\Sigma\bar{\beta}_0]^{1/2}W_i \tag{166}$$

which completes the proof.

Proof of Theorem 7:

Consider first

$$\tilde{F}_T(i;l) \;=\; \frac{SSR_l(\hat{T}_1,...,\hat{T}_l) - \inf_{\tau\in\Lambda_{i,\eta}} SSR_{l+1}(\hat{T}_1,...,\hat{T}_{i-1},\tau,\hat{T}_i,...,\hat{T}_l)\}}{\hat{\sigma}_i^2} \tag{167}$$

for a given i. Defining $S_T(i,j)$ to be the minimized sum of squared residuals for the segment containing observations from i to j, we can write

$$\tilde{F}_T(i;l) = \sup_{\tau\in\Lambda_{i,\eta}} \frac{\{S_T(\hat{T}_{i-1}+1,\hat{T}_i) - S_T(\hat{T}_{i-1}+1,\tau) - S_T(\tau+1,\hat{T}_i)\}}{\hat{\sigma}_i^2} \tag{168}$$

By similar arguments to (158), it follows that

$$\hat{\sigma}_i^2 \;\xrightarrow{p}\; \sigma^2 + 2\gamma'\beta_i^0 + \beta_i^{0\prime}\Sigma\beta_i^0 \tag{169}$$

Furthermore, from Theorem 2, we have that $\hat{T}_i = T_i^0 + O_p(1)$, and so using (169) it follows that

$$\tilde{F}_T(i;l) \;=\; \sup_{\tau\in\Lambda_{i,\eta}^0}\left\{\frac{S_T(T_{i-1}^0+1,T_i^0) - S_T(T_{i-1}^0+1,\tau) - S_T(\tau+1,T_i^0)}{\sigma^2 + 2\gamma'\beta_i^0 + \beta_i^{0\prime}\Sigma\beta_i^0}\right\} \;+\; o_p(1) \tag{170}$$

where $\Lambda_{i,\eta}^0 = \{\tau : T_{i-1}^0 + (T_i^0 - T_{i-1}^0)\eta \le \tau \le T_i^0 - (T_i^0 - T_{i-1}^0)\eta\}$.

[27] See, *inter alia*, Hamilton (1994)[p.544] for a definition of Brownian motion.

Therefore, we investigate the limiting behaviour of the expression inside the curly bracket in (170). Define

$$G_{T,i} \equiv S_T(T_{i-1}^0+1, T_i^0) - S_T(T_{i-1}^0+1, \tau) - S_T(\tau+1, T_i^0) \tag{171}$$

Let $P_{W_{i-1,i}}$ denote the projection matrix onto the column space of $[\hat{x}_{T_{i-1}^0+1}, \ldots, \hat{x}_{T_i^0}]'$ and $\tilde{U}_{i-1,i}$ denote $[\tilde{u}_{T_{i-1}^0+1}, \ldots, \tilde{u}_{T_i^0}]'$; let $P_{W_{i-1,\tau}}$ denote the projection matrix onto the column space of $[\hat{x}_{T_{i-1}^0+1}, \ldots, \hat{x}_{\tau}]'$ and $\tilde{U}_{i-1,\tau} = [\tilde{u}_{T_{i-1}^0+1}, \ldots, \tilde{u}_{\tau}]'$; $P_{W_{\tau,i}}$ denote the projection matrix onto the column space of $[\hat{x}_{\tau+1}, \ldots, \hat{x}_{T_i^0}]'$ and $\tilde{U}_{\tau,i} = [\tilde{u}_{\tau+1}, \ldots, \tilde{u}_{T_i^0}]'$. Using these definitions, (171) can be rewritten as

$$\begin{aligned}
G_{T,i} &= \|(I-P_{W_{i-1,i}})\tilde{U}_{i-1,i}\|^2 - \|(I-P_{W_{i-1,\tau}})\tilde{U}_{i-1,\tau}\|^2 - \|(I-P_{W_{\tau,i}})\tilde{U}_{\tau,i}\|^2 \\
&= \tilde{U}'_{i-1,i}(I-P_{W_{i-1,i}})\tilde{U}_{i-1,i} - \tilde{U}'_{i-1,\tau}(I-P_{W_{i-1,\tau}})\tilde{U}_{i-1,\tau} - \tilde{U}'_{\tau,i}(I-P_{W_{\tau,i}})\tilde{U}_{\tau,i} \\
&= \tilde{U}'_{i-1,i}\tilde{U}_{i-1,i} - \tilde{U}'_{i-1,i}P_{W_{i-1,i}}\tilde{U}_{i-1,i} - \tilde{U}'_{i-1,\tau}\tilde{U}_{i-1,\tau} + \tilde{U}'_{i-1,\tau}P_{W_{i-1,\tau}}\tilde{U}_{i-1,\tau} \\
&\quad -\tilde{U}'_{\tau,i}\tilde{U}_{\tau,i} + \tilde{U}'_{\tau,i}P_{W_{\tau,i}}\tilde{U}_{\tau,i} \\
&= -\tilde{U}'_{i-1,i}P_{W_{i-1,i}}\tilde{U}_{i-1,i} + \tilde{U}'_{i-1,\tau}P_{W_{i-1,\tau}}\tilde{U}_{i-1,\tau} + \tilde{U}'_{\tau,i}P_{W_{\tau,i}}\tilde{U}_{\tau,i} \\
&= -S'_{i-1,i}H_{i-1,i}^{-1}S_{i-1,i} + S'_{i-1,\tau}H_{i-1,\tau}^{-1}S_{i-1,\tau} + (S_{i-1,i} - S_{i-1,\tau})' \\
&\quad \times (H_{i-1,i} - H_{i-1,\tau})^{-1}(S_{i-1,i} - S_{i-1,\tau})
\end{aligned} \tag{172}$$

where $S_{i-1,\tau} = W'_{i-1,\tau}\tilde{U}_{i-1,\tau}$, $S_{i-1,i} = W'_{i-1,i}\tilde{U}_{i-1,i}$, $H_{i-1,\tau} = W'_{i-1,\tau}W_{i-1,\tau}$, and $H_{i-1,i} = W'_{i-1,i}W_{i-1,i}$. The limiting behavior of $G_{T,i}$ is deduced from the limiting behavior of $S_{i-1,i}, S_{i-1,\tau}, H_{i-1,\tau}$ and $H_{i-1,i}$. To this end, let $\Delta T_i^0 = T_i^0 - T_{i-1}^0$ and note that under our assumptions we have

$$\begin{aligned}
(\Delta T_i^0)^{-1/2} \sum_{t=T_{i-1}^0+1}^{T_{i-1}^0+\Delta T_i^0 \mu} z_t u_t &\Rightarrow (N_1' \otimes Q_{ZZ}^{1/2})B^{(i)}(\mu) && (173) \\
&= Q_{ZZ}^{1/2}\tilde{G}^{*(i)}(\mu), \text{ say} && (174)
\end{aligned}$$

where $B^{(i)}(\mu) = B(\lambda_{i-1}^0 + \mu) - B(\lambda_{i-1}^0)$ and $B(.)$ is defined in (131); and

$$\begin{aligned}
(\Delta T_i^0)^{-1/2} \sum_{t=T_{i-1}^0+1}^{T_{i-1}^0+\Delta T_i^0 \mu} z_t v_t' &\Rightarrow Q_{ZZ}^{1/2}B^{mat(i)}(\mu)N_2 && (175) \\
&= Q_{ZZ}^{1/2}G^{*(i)}(\mu) && (176)
\end{aligned}$$

where $B^{mat(i)}(\mu) = B^{mat}(\lambda_{i-1}^0 + \mu) - B^{mat}(\lambda_{i-1}^0)$ and $B^{mat}(.)$ is defined in (135).

First consider $(\Delta T_i^0)^{-1/2} S_{i-1,\tau}$. Letting $\sum_{t=T_{i-1}^0+1}^{T_{i-1}^0+\Delta T_i^0 \mu} \equiv \sum_1$, we have

$$
\begin{aligned}
(\Delta T_i^0)^{-1/2} S_{i-1,\tau} &= (\Delta T_i^0)^{-1/2} W'_{i-1,\tau} \tilde{U}_{i-1,\tau} \\
&= (\Delta T_i^0)^{-1/2} \sum_1 \hat{x}_t \tilde{u}_t \qquad (177)\\
&= (\Delta T_i^0)^{-1/2} [\Delta_0{}' \sum_1 z_t u_t + V'Z(Z'Z)^{-1} \sum_1 z_t u_t \\
&\quad + \Delta_0{}' \sum_1 z_t v_t{}' \beta_i^0 + V'Z(Z'Z)^{-1} \sum_1 z_t v_t{}' \beta_i^0 \\
&\quad - \Delta_0{}' \sum_1 z_t z_t' (Z'Z)^{-1} Z'V \beta_i^0 - V'Z(Z'Z)^{-1} \sum_1 z_t z_t' \\
&\quad \times (Z'Z)^{-1} Z'V \beta_i^0] \\
&= \Delta_0{}' (\Delta T_i^0)^{-1/2} \sum_1 z_t u_t + V'Z(Z'Z)^{-1} (\Delta T_i^0)^{-1/2} \sum_1 z_t u_t \\
&\quad + \Delta_0{}' (\Delta T_i^0)^{-1/2} \sum_1 z_t v_t{}' \beta_i^0 + V'Z(Z'Z)^{-1} (\Delta T_i^0)^{-1/2} \\
&\quad \times \sum_1 z_t v_t{}' \beta_i^0 - \Delta_0{}' \sum_1 z_t z_t' (Z'Z)^{-1} (\Delta T_i^0)^{-1/2} Z'V \beta_i^0 \\
&\quad - (\Delta T_i^0)^{-1/2} V'Z(Z'Z)^{-1} \sum_1 z_t z_t' (Z'Z)^{-1} Z'V \beta_i^0 \qquad (178)
\end{aligned}
$$

Using Assumptions 3 and 6, (174) and (176), it follows that

$$
\begin{aligned}
(\Delta T_i^0)^{-1/2} S_{i-1,\tau} &= \Delta_0{}' (\Delta T_i^0)^{-1/2} \sum_1 z_t u_t + \Delta_0{}' (\Delta T_i^0)^{-1/2} \sum_1 z_t v_t{}' \beta_i^0 \\
&\quad - \Delta_0' \sum_1 z_t z_t' (Z'Z)^{-1} (\Delta T_i^0)^{-1/2} Z'V \beta_i^0 + o_p(1) \\
&\Rightarrow \Delta_0' Q_{ZZ}^{1/2} \tilde{G}^{*(i)}(\mu) + \Delta_0' Q_{ZZ}^{1/2} G^{*(i)}(\mu) \beta_i^0 - \Delta_0' (\lambda_{i+1}^0 - \lambda_i^0)^{1/2} \\
&\quad \times \mu Q_{ZZ}^{1/2} \sum_{j=1}^{l+1} (\lambda_j - \lambda_{j-1})^{1/2} G^{*(j)}(1) \beta_i^0 \\
&= \Delta_0' Q_{ZZ}^{1/2} \tilde{G}^{*(i)}(\mu) + \Delta_0' Q_{ZZ}^{1/2} G^{*(i)}(\mu) \beta_i^0 - \mu (\lambda_{i+1}^0 - \lambda_i^0)^{1/2} \Delta_0' Q_{ZZ}^{1/2} \\
&\quad \times \sum_{j=1}^{l+1} (\lambda_j - \lambda_{j-1})^{1/2} G^{*(j)}(1) \beta_i^0 \\
&\qquad (179)
\end{aligned}
$$

Now, consider $(\Delta T_i^0)^{-1} H_{i-1,\tau}$. Using Assumption 6 and the consistency of $\hat{\Delta}_T$, it follows that

$$
\begin{aligned}
(\Delta T_i^0)^{-1} H_{i-1,\tau} &= (\Delta T_i^0)^{-1} W'_{i-1,\tau} W_{i-1,\tau} \\
&= (\Delta T_i^0)^{-1} \sum_1 \hat{x}_t \hat{x}_t' \\
&= \mu \hat{\Delta}_T' (\Delta T_i^0 \mu)^{-1} \sum_1 z_t z_t' \hat{\Delta}_T \\
&\Rightarrow \mu \Delta_0' Q_{ZZ} \Delta_0 \qquad (180)
\end{aligned}
$$

It follows from (179) and (180) that[28]

$$
\begin{aligned}
S'_{i-1,i}H^{-1}_{i-1,i}S_{i-1,i} \quad &\Rightarrow \quad (\Delta'_0 Q^{1/2}_{ZZ}\tilde{G}^{*(i)}(1) + \Delta'_0 Q^{1/2}_{ZZ}G^{*(i)}(1)\beta^0_i - (\lambda^0_{i+1} - \lambda^0_i)^{1/2} \\
&\quad \times \Delta'_0 Q^{1/2}_{ZZ}\sum_{j=1}^{l+1}(\lambda_j - \lambda_{j-1})^{1/2}G^{*(j)}(1)\beta^0_i)'(\Delta'_0 Q_{ZZ}\Delta_0)^{-1} \\
&\quad \times (\Delta'_0 Q^{1/2}_{ZZ}\tilde{G}^{*(i)}(1) + \Delta'_0 Q^{1/2}_{ZZ}G^{*(i)}(1)\beta^0_i - (\lambda^0_{i+1} - \lambda^0_i)^{1/2} \\
&\quad \times \Delta'_0 Q^{1/2}_{ZZ}\sum_{j=1}^{l+1}(\lambda_j - \lambda_{j-1})^{1/2}G^{*(j)}(1)\beta^0_i) \\
&= \quad (\Delta'_0 Q^{1/2}_{ZZ}[\tilde{G}^{*(i)}(1) + G^{*(i)}(1)\beta^0_i - (\lambda^0_{i+1} - \lambda^0_i)^{1/2} \\
&\quad \times \sum_{j=1}^{l+1}(\lambda_j - \lambda_{j-1})^{1/2}G^{*(j)}(1)\beta^0_i])'(\Delta'_0 Q_{ZZ}\Delta_0)^{-1}(\Delta'_0 Q^{1/2}_{ZZ}[\tilde{G}^{*(i)}(1) \\
&\quad + G^{*(i)}(1)\beta^0_i - (\lambda^0_{i+1} - \lambda^0_i)^{1/2}\sum_{j=1}^{l+1}(\lambda_j - \lambda_{j-1})^{1/2}G^{*(j)}(1)\beta^0_i]) \\
&= \quad (\Lambda C\tilde{G}^{*(i)}(1) + \Lambda C[G^{*(i)}(1) - (\lambda^0_{i+1} - \lambda^0_i)^{1/2} \\
&\quad \times \sum_{j=1}^{l+1}(\lambda_j - \lambda_{j-1})^{1/2}G^{*(j)}(1)]\beta^0_i)'(\Lambda C\tilde{G}^{*(i)}(1) + \Lambda C[G^{*(i)}(1) \\
&\quad - (\lambda^0_{i+1} - \lambda^0_i)^{1/2}\sum_{j=1}^{l+1}(\lambda_j - \lambda_{j-1})^{1/2}G^{*(j)}(1)]\beta^0_i)
\end{aligned} \tag{181}
$$

Now, define $D_\mu = \Lambda C G^{*(i)}(\mu), \tilde{D}_\mu = \Lambda C\tilde{G}^{*(i)}(\mu), D_1 = \Lambda C G^{*(i)}(1), \tilde{D}_1 = \Lambda C\tilde{G}^{*(i)}(1)$ and $D_i = (\lambda^0_{i+1} - \lambda^0_i)^{1/2}\Lambda C\sum_{j=1}^{l+1}(\lambda_j - \lambda_{j-1})^{1/2}G^{*(j)}(1)$, we have

$$
S'_{i-1,i}H^{-1}_{i-1,i}S_{i-1,i} \Rightarrow (\tilde{D}_1 + [D_1 - D_i]\beta^0_i)'(\tilde{D}_1 + [D_1 - D_i]\beta^0_i) \tag{182}
$$

Similarly, using the results in (179) and (180) we have

$$
S'_{i-1,\tau}H^{-1}_{i-1,\tau}S_{i-1,\tau} \quad \Rightarrow \quad \mu^{-1}(\tilde{D}_\mu + [D_\mu - \mu D_i]\beta^0_i)'(\tilde{D}_\mu + [D_\mu - \mu D_i]\beta^0_i) \tag{183}
$$

[28] Note that we use the spectral decomposition of $(\Delta_0{}'Q^{1/2}_{ZZ})'(\Delta_0{}'Q_{ZZ}\Delta_0)^{-1}(\Delta_0{}'Q^{1/2}_{ZZ})$ as in the proof of Theorem 6.

Finally, using the results in (179) and (180), we have

$$\begin{aligned}
&(S_{i-1,i}-S_{i-1,\tau})'(H_{i-1,i}-H_{i-1,\tau})^{-1}(S_{i-1,i}-S_{i-1,\tau})\\
\Rightarrow\quad & [\Delta_0' Q_{ZZ}^{1/2}(\tilde{G}^{*(i)}(1)-\tilde{G}^{*(i)}(\mu))+\Delta_0' Q_{ZZ}^{1/2}(G^{*(i)}(1)-G^{*(i)}(\mu))\beta_i^0-\Delta_0' Q_{ZZ}^{1/2}\\
&\times(1-\mu)(\lambda_{i+1}^0-\lambda_i^0)^{1/2}\sum_{j=1}^{l+1}(\lambda_j-\lambda_{j-1})^{1/2}G^{*(j)}(1)\beta_i^0]'(\Delta_0' Q_{ZZ}\Delta_0(1-\mu))^{-1}\\
&\times[\Delta_0' Q_{ZZ}^{1/2}(\tilde{G}^{*(i)}(1)-\tilde{G}^{*(i)}(\mu))+\Delta_0' Q_{ZZ}^{1/2}(G^{*(i)}(1)-G^{*(i)}(\mu))\beta_i^0-\Delta_0' Q_{ZZ}^{1/2}\\
&\times(1-\mu)(\lambda_{i+1}^0-\lambda_i^0)^{1/2}\sum_{j=1}^{l+1}(\lambda_j-\lambda_{j-1})^{1/2}G^{*(j)}(1)\beta_i^0]\\
=\quad & (1-\mu)^{-1}[(\tilde{D}_1-\tilde{D}_\mu)+(D_1-D_\mu)\beta_i^0-(1-\mu)D_i\beta_i^0]'[(\tilde{D}_1-\tilde{D}_\mu)+(D_1-D_\mu)\beta_i^0\\
&-(1-\mu)D_i\beta_i^0] \qquad (184)
\end{aligned}$$

Thus, combining results in (182), (183) and (184), it follows that

$$\begin{aligned}
G_{T,i}\quad\Rightarrow\quad & -(\tilde{D}_1+[D_1-D_i]\beta_i^0)'(\tilde{D}_1+[D_1-D_i]\beta_i^0)+\mu^{-1}(\tilde{D}_\mu+[D_\mu-\mu D_i]\beta_i^0)'\\
&\times(\tilde{D}_\mu+[D_\mu-\mu D_i]\beta_i^0)+(1-\mu)^{-1}[(\tilde{D}_1-\tilde{D}_\mu)+(D_1-D_\mu)\beta_i^0\\
&-(1-\mu)D_i\beta_i^0]'[(\tilde{D}_1-\tilde{D}_\mu)+(D_1-D_\mu)\beta_i^0-(1-\mu)D_i\beta_i^0]\\
=\quad & -(\tilde{D}_1+[D_1-D_i]\beta_i^0)'(\tilde{D}_1+[D_1-D_i]\beta_i^0)+\mu^{-1}(\tilde{D}_\mu+[D_\mu-\mu D_i]\beta_i^0)'\\
&\times(\tilde{D}_\mu+[D_\mu-\mu D_i]\beta_i^0)+(1-\mu)^{-1}[(\tilde{D}_1-\tilde{D}_\mu)+(D_1-D_\mu\\
&-(1-\mu)D_i)\beta_i^0]'[(\tilde{D}_1-\tilde{D}_\mu)+(D_1-D_\mu-(1-\mu)D_i)\beta_i^0] \qquad (185)
\end{aligned}$$

After some tedious algebra, it can be shown that

$$G_{T,i}\ \Rightarrow\ \frac{1}{\mu(1-\mu)}\|b(\mu)-\mu b(1)\|^2 \qquad (186)$$

where $b(\mu)=\tilde{D}_\mu+D_\mu\beta_i^0$ and $b(1)=\tilde{D}_1+D_1\beta_i^0$. By similar arguments to the proof of Theorem 6[29], it can be shown that

$$b(\mu)\overset{\mathrm{d}}{=}[\sigma^2+2\gamma'\beta_i^0+\beta_i^{0\prime}\Sigma\beta_i^0]^{1/2}\begin{bmatrix} W(\mu)\\ 0_{(q-p)\times 1}\end{bmatrix} \qquad (187)$$

It follows from (170), (171), (186) and (187) that

$$\tilde{F}_T(i;l)\ \Longrightarrow\ \sup_{\eta\le\mu<1-\eta}\frac{\|W(\mu)-\mu W(1)\|^2}{\mu(1-\mu)} \qquad (188)$$

Therefore, the limiting distribution of $F_T(l+1|l)$ is that of the maximum of $l+1$ independent random variables of the form in (188) which is the desired result.

[29]See (161) and subsequent argument.

References

Amemiya, T. (1985). *Advanced Econometrics.* Harvard University Press, Cambridge, MA, U.S.A.

Andrews, D. W. K. (1993). 'Tests for parameter instability and structural change with unknown change point', *Econometrica*, 61: 821–856.

Andrews, D. W. K., and Fair, R. (1988). 'Inference in econometric models with structural change', *Review of Economic Studies*, 55: 615–640.

Bai, J. (1994). 'Least squares estimation of a shift in linear processes', *Journal of Time Series Analysis*, 15: 453–472.

Bai, J., and Perron, P. (1998). 'Estimating and testing linear models with multiple structural changes', *Econometrica*, 66: 47–78.

——— (2001). 'Additional critical values for multiple structural change tests', Discussion paper, Department of Economics, Boston University, Boston, MA.

Bhattacharya, P. K. (1987). 'Maximum Likelihood estimation of a change-point in the distribution of independent random variables: general multiparameter case', *Journal of Multivariate Analysis*, 23: 183–208.

de Jong, R. M., and Davidson, J. (2000). 'THE Functional Central Limit Theorem and Weak Convergence to Stochastic Integrals I', *Econometric Theory*, 16: 621–642.

Ghysels, E., and Hall, A. R. (1990). 'A test for structural stability of Euler condition parameters estimated via the Generalized Method of Moments', *International Economic Review*, 31: 355–364.

Hahn, J., and Inoue, A. (2002). 'A Monte Carlo comparison of various asymptotic approximations to the distribution of instrumental variables estimators', *Econometric Reviews*, 21: 309–336.

Hall, A. R. (2005). *Generalized Method of Moments.* Oxford University Press, Oxford, U.K.

Hall, A. R., Han, S., and Boldea, O. (2007). 'A distribution theory for change point estimators in models estimated by Two Stage Least Squares', Discussion paper, Department of Economics, North Carolina State University, Raleigh, NC.

Hall, A. R., and Sen, A. (1999). 'Structural stability testing in models estimated by Generalized Method of Moments', *Journal of Business and Economic Statistics*, 17: 335–348.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton, NJ, U. S. A.

Han, S. (2006). 'Inference regarding multiple structural changes in linear models estimated via Instrumental Variables', Ph.D. thesis, Department of Economics, North Carolina State University, Raleigh, NC.

Hawkins, D. L. (1986). 'A simple least square method for estimating a change in mean', *Communications in Statistics - Simulation*, 15: 655–679.

Ortega, J. M. (1987). *Matrix Theory: a Second Course*. Plenum Press, New York, NY, U.S.A.

Perron, P., and Qu, Z. (2006). 'Estimating restricted structural change models', *Journal of Econometrics*, 134: 373–399.

Picard, D. (1985). 'Testing and estimating change points in time series', *Journal of Applied Probability*, 20: 411–415.

Qu, Z., and Perron, P. (2007). 'Estimating and testing structural changes in multivariate regressions', *Econometrica*, 75: 459–502.

Sowell, F. (1996). 'Optimal tests of parameter variation in the Generalized Method of Moments framework', *Econometrica*, 64: 1085–1108.

Yao, Y.-C. (1987). 'Approximating the distribution of the ML estimate of the chnage point in a sequence of independent r.v.'s', *Annals of Statistics*, 4: 1321–1328.

Zhang, C., Osborn, D. R., and Kim, D. H. (2007). 'The new Keynesian Phillips curve: from sticky inflation to sticky prices', *Journal of Money, Credit and Banking*, forthcoming.

Table 1: Finite sample behavior of break fraction estimator

one break model with $(\beta_1^0, \beta_2^0) = (\,[1, 0.1]', [-1, -0.1]'\,)$

q	T	Deviation from the True Break Fraction				
		1 %	2 %	3 %	5 %	10 %
2	60	.66	.89	.89	.98	.99
	120	.88	.95	.97	.99	1.00
	240	.96	.99	1.00	1.00	1.00
	480	.99	1.00	1.00	1.00	1.00
4	60	.67	.87	.87	.97	.99
	120	.89	.96	.98	1.00	1.00
	240	.93	.98	1.00	1.00	1.00
	480	.99	1.00	1.00	1.00	1.00
8	60	.66	.85	.85	.95	.99
	120	.87	.95	.97	1.00	1.00
	240	.95	.99	1.00	1.00	1.00
	480	.99	1.00	1.00	1.00	1.00

Notes: The column headed $100a\%$ gives the proportion of the simulations in which $|\hat{\lambda}_1 - \lambda_1^0| \leq a$; q is the number of instruments; T is the sample size.

Table 2: Relative rejection frequencies of F-statistics

one break model: $(\beta_1^0, \beta_2^0) = (\,[1, 0.1]', [-1, -0.1]'\,)$

q	T	supF(k)		supF(l+1:l)		UDmax
		1	2	2:1	3:2	
2	60	1.00	1.00	.07	.01	1.00
	120	1.00	1.00	.06	0	1.00
	240	1.00	1.00	.06	0	1.00
	480	1.00	1.00	.06	0	1.00
4	60	1.00	1.00	.08	.02	1.00
	120	1.00	1.00	.08	0	1.00
	240	1.00	1.00	.07	0	1.00
	480	1.00	1.00	.05	0	1.00
8	60	1.00	1.00	.05	.01	1.00
	120	1.00	1.00	.05	0	1.00
	240	1.00	1.00	.05	.01	1.00
	480	1.00	1.00	.06	0	1.00

Notes: supF(k) denotes the statistic $Sup - F_T(k; 1)$ and the second tier column heading under it denotes k; F(l+1:l) denotes the statistic $F_T(l + 1|l)$ and the second tier column beneath it denotes $l + 1 : l$; UDmax denotes the statistic $UDmaxF_T(5, 1)$; q is the number of instruments; T is the sample size.

Table 3: Empirical distribution of the estimated number of breaks

one break model: $(\beta_1^0, \beta_2^0) = (\,[1, 0.1]', [-1, -0.1]'\,)$

q	T	supF(1)				UDmax			
		0	1	2	3,4,5	0	1	2	3,4,5
2	60	0	.96	.04	0	0	.96	.04	0
	120	0	.97	.03	0	0	.97	.03	0
	240	0	.97	.03	0	0	.97	.03	0
	480	0	.96	.04	0	0	.96	.04	0
4	60	0	.95	.05	0	0	.95	.05	0
	120	0	.96	.04	0	0	.96	.04	0
	240	0	.97	.03	0	0	.97	.03	0
	480	0	.97	.03	0	0	.97	.03	0
8	60	0	.95	.05	0	0	.95	.05	0
	120	0	.97	.03	0	0	.97	.03	0
	240	0	.97	.03	0	0	.97	.03	0
	480	0	.98	.02	0	0	.98	.02	0

Notes: The figures in the block headed supF(1) give the empirical distribution of the estimated number of breaks, $\hat{k}_T$, obtained via the sequential strategy using $Sup - F_T(1;1)$ on the first step with the maximum number of breaks set equal to five. The figures in the block UDmax give the empirical distribution of the estimated number of breaks, $\hat{k}_T$, obtained via the sequential strategy using $UDmaxF_T(5,1)$ on the first step with the maximum number of breaks set equal to five.

Table 4: Empirical coverage of parameter confidence intervals

one break model with $(\beta_1^0, \beta_2^0) = (\,[1, 0.1]', [-1, -0.1]'\,)$

q	T		Confidence Intervals					
			intercept			slope		
			99%	95 %	90 %	99%	95 %	90 %
2	60	1^{st} regime	.99	.93	.88	.98	.92	.88
		2^{nd} regime	.98	.93	.88	.98	.93	.88
	120	1^{st} regime	.99	.95	.90	.99	.96	.90
		2^{nd} regime	.99	.95	.90	.99	.95	.90
	240	1^{st} regime	.99	.93	.88	.99	.94	.88
		2^{nd} regime	.99	.94	.89	.99	.93	.88
	480	1^{st} regime	.99	.94	.89	.99	.94	.88
		2^{nd} regime	.99	.95	.89	.99	.95	.90
4	60	1^{st} regime	.98	.94	.88	.99	.93	.88
		2^{nd} regime	.98	.93	.88	.99	.93	.88
	120	1^{st} regime	.99	.96	.90	.99	.95	.89
		2^{nd} regime	.99	.94	.89	.99	.94	.89
	240	1^{st} regime	.99	.94	.89	.99	.95	.90
		2^{nd} regime	.98	.94	.89	.99	.95	.91
	480	1^{st} regime	.98	.94	.89	.99	.96	.92
		2^{nd} regime	.99	.95	.88	.99	.95	.89
8	60	1^{st} regime	.98	.93	.87	.98	.92	.85
		2^{nd} regime	.98	.92	.86	.98	.92	.84
	120	1^{st} regime	.99	.94	.90	.99	.94	.88
		2^{nd} regime	.99	.94	.89	.98	.93	.88
	240	1^{st} regime	.99	.95	.91	.99	.95	.89
		2^{nd} regime	.99	.96	.91	.98	.93	.88
	480	1^{st} regime	.99	.95	.90	.99	.94	.88
		2^{nd} regime	.99	.95	.89	.99	.95	.88

Notes: The column headed $100a\%$ gives the percentage of times the confidence intervals contain the corresponding true parameter values.

Table 5: Finite sample behavior of break fraction estimator

two break model: $(\beta_1^0, \beta_2^0, \beta_3^0) = ([1,\, 0.1]', [-1,\, -0.1]', [1,\, 0.1]')$

q	T	i-th Break	Deviation from the True Break Fraction 1 %	2 %	3 %	5 %	10 %
2	60	1st	.65	.86	.86	.96	.99
		2nd	.66	.86	.86	.96	.98
	120	1st	.89	.94	.98	1.00	1.00
		2nd	.90	.96	.98	1.00	1.00
	240	1st	.96	.99	1.00	1.00	1.00
		2nd	.96	.99	1.00	1.00	1.00
	480	1st	.99	1.00	1.00	1.00	1.00
		2nd	.99	1.00	1.00	1.00	1.00
4	60	1st	.65	.84	.84	.94	.99
		2nd	.65	.85	.85	.96	.99
	120	1st	.86	.94	.97	1.00	1.00
		2nd	.88	.95	.97	1.00	1.00
	240	1st	.95	.98	1.00	1.00	1.00
		2nd	.95	.99	1.00	1.00	1.00
	480	1st	.99	1.00	1.00	1.00	1.00
		2nd	.98	1.00	1.00	1.00	1.00
8	60	1st	.62	.86	.86	.95	.99
		2nd	.64	.82	.82	.95	.98
	120	1st	.86	.94	.98	1.00	1.00
		2nd	.87	.93	.96	.99	1.00
	240	1st	.94	.99	1.00	1.00	1.00
		2nd	.94	.99	1.00	1.00	1.00
	480	1st	.99	1.00	1.00	1.00	1.00
		2nd	.98	1.00	1.00	1.00	1.00

Notes: See Table 1 for definitions.

Table 6: Relative rejection frequencies of F-statistics

two break model: $(\beta_1^0, \beta_2^0, \beta_3^0) = (\,[1,\, 0.1]', [-1,\, -0.1]',\, [1,\, 0.1]'\,)$

q	T	supF(k)		supF(l+1:l)		UDmax
		1	2	2:1	3:2	
2	60	.74	1.00	1.00	.03	1.00
	120	1.00	1.00	1.00	.03	1.00
	240	1.00	1.00	1.00	.02	1.00
	480	1.00	1.00	1.00	.02	1.00
4	60	.71	1.00	1.00	.04	1.00
	120	1.00	1.00	1.00	.02	1.00
	240	1.00	1.00	1.00	.02	1.00
	480	1.00	1.00	1.00	.01	1.00
8	60	.70	1.00	1.00	.05	1.00
	120	1.00	1.00	1.00	.02	1.00
	240	1.00	1.00	1.00	.02	1.00
	480	1.00	1.00	1.00	.02	1.00

Notes: See Table 2 for definitions.

Table 7: Empirical distribution of the estimated number of breaks

two break model: $(\beta_1^0, \beta_2^0, \beta_3^0) = (\,[1,\,0.1]', [-1,\,-0.1]',\,[1,\,0.1]'\,)$

q	T	supF(1)				UDmax			
		0	1	2	3	0	1	2	3
2	60	.26	0	.71	.03	0	0	.94	.06
	120	0	0	.96	.04	0	0	.96	.04
	240	0	0	.98	.02	0	0	.98	.02
	480	0	0	.98	.02	0	0	.98	.02
4	60	.29	0	.67	.04	0	0	.94	.06
	120	0	0	.96	.04	0	0	.96	.04
	240	0	0	.98	.02	0	0	.98	.02
	480	0	0	.98	.02	0	0	.98	.02
8	60	.30	0	.65	.05	0	0	.94	.06
	120	0	0	.96	.04	0	0	.96	.04
	240	0	0	.98	.02	0	0	.98	.02
	480	0	0	.98	.02	0	0	.98	.02

Notes: See Table 3 for definitions.

Table 8: Empirical coverage of parameter confidence intervals

two break model: $(\beta_1^0, \beta_2^0, \beta_3^0) = ([1,\, 0.1]', [-1,\, -0.1]', [1,\, 0.1]')$

q	T		Confidence Intervals					
			intercept			slope		
			99%	95 %	90 %	99%	95 %	90 %
2	60	1^{st} regime	.97	.92	.86	.98	.91	.85
		2^{nd} regime	.98	.92	.86	.98	.93	.86
		3^{rd} regime	.97	.92	.85	.98	.93	.86
	120	1^{st} regime	.98	.92	.87	.98	.93	.89
		2^{nd} regime	.98	.93	.87	.99	.95	.90
		3^{rd} regime	.98	.94	.89	.99	.94	.89
	240	1^{st} regime	.99	.96	.91	.99	.95	.91
		2^{nd} regime	.99	.93	.88	.99	.95	.90
		3^{rd} regime	.99	.94	.90	.98	.94	.89
	480	1^{st} regime	.99	.96	.90	1.00	.96	.91
		2^{nd} regime	.99	.94	.89	.99	.95	.90
		3^{rd} regime	.99	.93	.88	.99	.95	.91
4	60	1^{st} regime	.98	.93	.86	.98	.92	.87
		2^{nd} regime	.97	.91	.84	.98	.92	.86
		3^{rd} regime	.98	.92	.86	.98	.92	.86
	120	1^{st} regime	.99	.92	.87	.98	.94	.88
		2^{nd} regime	.98	.93	.86	.99	.95	.88
		3^{rd} regime	.99	.94	.89	.99	.95	.88
	240	1^{st} regime	.99	.94	.90	.99	.94	.88
		2^{nd} regime	.99	.93	.88	.99	.95	.88
		3^{rd} regime	.98	.93	.90	.98	.94	.87
	480	1^{st} regime	.99	.94	.89	.99	.95	.90
		2^{nd} regime	.99	.94	.88	.98	.94	.89
		3^{rd} regime	.99	.95	.89	.99	.95	.91

Notes: See Table 4 for definitions.

Table 9: Empirical coverage of parameter confidence intervals ctd.

two break model: $(\beta_1^0, \beta_2^0, \beta_3^0) = (\,[1,\, 0.1]', [-1,\, -0.1]',\, [1,\, 0.1]'\,)$

q	T		Confidence Intervals					
			intercept			slope		
			99%	95 %	90 %	99%	95 %	90 %
8	60	1^{st} regime	.98	.93	.86	.98	.92	.85
		2^{nd} regime	.97	.91	.85	.96	.90	.84
		3^{rd} regime	.98	.92	.86	.98	.92	.85
	120	1^{st} regime	.99	.95	.88	.99	.94	.88
		2^{nd} regime	.98	.92	.87	.98	.92	.87
		3^{rd} regime	.99	.94	.89	.99	.94	.88
	240	1^{st} regime	.99	.96	.90	.99	.94	.88
		2^{nd} regime	.99	.93	.89	.99	.93	.88
		3^{rd} regime	.98	.95	.90	.99	.95	.89
	480	1^{st} regime	.99	.95	.90	.99	.94	.88
		2^{nd} regime	.99	.95	.90	.99	.95	.90
		3^{rd} regime	.99	.95	.91	.99	.94	.89

Notes: See Table 4 for definitions.

Table 10: Relative rejection frequencies of F-statistics

no break model: β^0=(1,1)

q	T	supF(k)		supF(l+1:l)		UDmax
		1	2	2:1	3:2	
2	60	.05	.07	.03	.01	.06
	120	.05	.06	.03	0	.05
	240	.05	.05	.02	0	.05
	480	.05	.05	.02	0	.05
4	60	.06	.07	.04	.01	.07
	120	.05	.04	.02	0	.05
	240	.05	.04	.02	0	.04
	480	.05	.06	.03	0	.06
8	60	.06	.06	.03	0	.06
	120	.04	.05	.02	0	.04
	240	.05	.05	.02	0	.04
	480	.05	.04	.02	0	.05

Notes: See Table 2 for definitions.

Table 11: Empirical distribution of the estimated number of breaks

no break model: β^0*=(1,1)*

q	T	supF(1)				UDmax			
		0	1	2	3,4,5	0	1	2	3,4,5
2	60	.95	.5	0	0	.94	.06	0	0
	120	.95	.05	0	0	.95	.05	0	0
	240	.95	.05	0	0	.95	.05	0	0
	480	.95	.05	0	0	.95	.05	0	0
4	60	.94	.06	0	0	.93	.06	.01	0
	120	.95	.05	0	0	.95	.05	0	0
	240	.95	.05	0	0	.96	.04	0	0
	480	.95	.05	0	0	.94	.05	.01	0
8	60	.94	.06	0	0	.94	.06	0	0
	120	.96	.04	0	0	.96	.03	.01	0
	240	.95	.05	0	0	.96	.04	0	0
	480	.95	.05	0	0	.96	.04	0	0

Notes: See Table 3 for definitions.

Table 12: Application to NKPC - stability statistics for the reduced forms

Dep.var	k	sup-F	F(k+1:k)	BIC
$inf^e_{t+1\|t}$	0			-0.615
	1	43.6	41.7	-0.623
	2	67.0	10.4	-0.680
	3	176.5	34.3	-0.649
	4	80.5	46.8	-0.452
	5	70.2		-0.369
og_t	0			-0.663
	1	50.0	30.53	-0.552
	2	40.1	23.1	-0.497
	3	40.	11.3	-0.276
	4	34.91	11.3	-0.046
	5	31.9		0.255

Notes: Dep. Var. denotes the dependent variable in the reduced form; sup-F denotes the statistic for testing $H_0 : m = 0$ vs. $H_1 : m = k$; F(k+1:k) is the statistic for testing $H_0 : m = k$ vs. $H_1 : m = k + 1$; BIC is the BIC criterion. The percentiles for the statistics are for $k = 1, 2, \ldots$ respectively: (i) sup-F: (10%, 1%) significance level = (25.29, 32.8), (23.33, 28.24), (21.89, 25.63), (20.71, 23.83), (19.63,22.32); (ii) F(k+1:k): (10%, 1%) significance level =(25.29, 32.8), (27.59,34.81), (28.75, 36.32), (29.71,36.65).

Table 13: Application to NKPC - stability statistics for structural equation

k	sup-F	F(k+1:k)	BIC
0			0.021
1	41.3	9.55	0.017
2	25.0	7.83	0.240
3	21.4	12.8	0.427
4	17.4		0.664
5	13.4		0.942

Notes: Sup-F denotes the statistic for testing $H_0 : m = 0$ vs. $H_1 : m = k$; F(k+1:k) is the statistic for testing $H_0 : m = k$ vs. $H_1 : m = k + 1$; BIC is the BIC criterion. The percentiles for the statistics are for $k = 1, 2, \ldots$ respectively: (i) sup-F: (10%, 1%) significance level = (19.7, 26.71), (17.67, 21.87), (16.04, 19.42), (14.55, 17.44), (12.59,15.02); (ii) F(k+1:k): (10%, 1%) significance level =(19.7, 26.71), (21.79, 28.36), (22.87, 29.30).