Asymptotic distribution theory for break point estimators in models estimated via 2SLS[1]

Otilia Boldea

Tilburg University

Alastair R. Hall[2]

University of Manchester

and

Sanggohn Han

SAS Institute

February 10, 2010

[1]We are grateful to an anonymous referee for valuable comments. The second author acknowledges the support of the ESRC grant RES-062-23-1351.

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

Abstract

In this paper, we present a limiting distribution theory for the break point estimator in a linear regression model with multiple structural breaks obtained by minimizing a Two Stage Least Squares (2SLS) objective function. Our analysis covers both the case in which the reduced form for the endogenous regressors is stable and the case in which it is unstable with multiple structural breaks. For stable reduced forms, we present a limiting distribution theory under two different scenarios: in the case where the parameter change is of fixed magnitude, it is shown that the resulting distribution depends on the distribution of the data and is not of much practical use for inference; in the case where the magnitude of the parameter change shrinks with the sample size, it is shown that the resulting distribution can be used to construct approximate large sample confidence intervals for the break points. For unstable reduced forms, we consider the case where the magnitudes of the parameter changes in both the equation of interest and the reduced forms shrink with the sample size at potentially different rates and not necessarily the same locations in the sample. The resulting limiting distribution theory can be used to construct approximate large sample confidence intervals for the break points. The finite sample performance of these intervals are analyzed in a small simulation study and the intervals are illustrated via an application to the New Keynesian Phillips curve.

JEL classification: C12, C13

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

1 Introduction

Econometric time series models are based on the assumption that the economic relationships, or "structure", in question are stable over time. However, with samples covering extended periods, this assumption is always open to question and this has led to considerable interest in the development of statistical methods for detecting structural instability.[1] In designing such methods, it is necessary to specify how the structure may change over time and a popular specification is one in which the parameters of the model are subject to discrete shifts at unknown points in the sample. This scenario can be motivated by the idea of policy regime changes.[2] Within this type of setting, the main concern is to estimate economic relationships in the different regimes and compare them. However, since not all policy changes may impact the economic relationship of interest, an important precursor to this analysis is the identification of the points in the sample, if any, at which the parameters change. This raises the issue of how to perform inference about the location of the so-called "break points", that is the points in the sample at which the parameters change, and motivates the interest to obtain a limiting distribution theory for break point estimators.[3] It is the latter which is the focus of this paper.

There is a literature in time series on the limiting distribution of break point estimators for estimation of changes in the mean of processes; see Hinckley (1970), Picard (1985), Bhattacharya (1987), Yao (1987), Bai (1994, 1997a). A limiting distribution theory has also been presented in the context of linear regression models estimated via Ordinary Least Squares (OLS). Bai (1997b) considers the case in which there is only one break. He presents two alternative limit theories for the break point estimator. One assumes the magnitude of change between the regimes is fixed; the resulting distribution theory for the break-point turns out to depend on the distribution of the data. The other assumes the magnitude of the parameter change is shrinking with the sample size[4]: this approach leads to practical methods for inference about the location of the

[1]See *inter alia* Andrews and Fair (1988), Ghysels and Hall (1990a,b), Andrews (1993), Andrews and Ploberger (1994), Sowell (1996), Hall and Sen (1999) as well as the other references below.

[2]For example, Bai (1997b) explores the impact of changes in monetary policy on the relationship between the interest rate and the discount factor in the US, and Zhang, Osborn, and Kim (2008) explore the impact of monetary policy changes on the Phillips curve.

[3]The term "change point" is also used in the literature to denote the points in the sample at which the parameter values change.

[4]The assumption of shrinking breaks is a mathematical device designed to produce confidence intervals for the break points whose asymptotic properties provide a reasonable approximation to finite sample behaviour when

break point. Bai and Perron (1998) consider the case of multiple break points that are estimated simultaneously. They present a limiting distribution theory for the break point estimators based on the assumption that the parameter change is shrinking as the sample size increases; this can be used by practitioners to perform inference about the location of the break points.

One maintained assumption in Bai's (1997b) and Bai and Perron's (1998) analyses is that the regressors are uncorrelated with the errors so that OLS is an appropriate method of estimation. This is a leading case, of course, but there are also many cases in econometrics where the regressors are correlated with the errors and so OLS yields inconsistent estimators. Once OLS is rejected as inappropriate, an alternative method of estimation must be chosen. As shown by Hall, Han, and Boldea (2009), minimizing the sum of partial generalized method of moments minimands over all partitions of the sample fails to yield consistent estimates of the break point in leading cases of interest. We thus follow the approach of Hall, Han, and Boldea (2009) and consider the case in which the estimation of the regression parameters and break points is performed by minimizing a Two Stage Least Squares (2SLS) objective function.[5] Hall, Han, and Boldea (2009) establish the consistency of these 2SLS estimators, a limiting distribution theory for the 2SLS estimators of the regression parameters, propose a number of tests for parameter variation and a methodology for estimating the number of break points. However, they do not consider the distribution of the break point estimators.

In this paper, we derive the distribution of the break point estimators based on minimization of the 2SLS objective function. As in Hall, Han, and Boldea (2009), our analysis covers both the case in which the reduced form for the endogenous regressors is stable and the case in which it is unstable with multiple structural breaks.[6]

For stable reduced forms, we present a limiting distribution theory under two different scenarios regarding the magnitude of the parameter change between regimes. First, if the parameter change is of fixed magnitude, the resulting distribution is shown to be the natural extension of

the breaks are of "moderate" size; see Bai and Perron (1998).

[5]There is a considerable literature on the use of Instrumental Variables (IV) and 2SLS in linear models with endogenous regressors in econometrics; see Christ (1994) or Hall (2005)[Chapter 1] for a historical review and examples in which such endogeneity arises.

[6]Note that all breaks in a structural system of equations are either reflected in the structural equation of interest, or in the reduced forms, or both; thus it is important to distinguish between stable and unstable reduced forms.

Bai's (1997b) result for OLS estimators and is consequently dependent on the distribution of the data. Second, if the magnitude of the parameter change shrinks with the sample size, the resulting distribution can be used to construct approximate large sample confidence intervals for the break points. For unstable reduced forms, we consider the case where the magnitude of the parameter changes in both the equation of interest and the reduced form shrink with the sample size at potentially different rates and different locations for the structural equation and reduced form. The resulting limiting distribution theory can be used to construct approximate large sample confidence intervals for the break points. The finite sample performance of these intervals is analyzed in a small simulation study and the intervals are illustrated via an application to the New Keynesian Phillips curve.

An outline of the paper is as follows. Section 2 contains results for the stable reduced form case. Section 3 presents the analysis for the unstable reduced form case and several break point estimators obtained using the methodology described in Hall, Han, and Boldea (2009). Section 4 reports results from a small simulation study and also the empirical application. Section 5 offers some concluding remarks. The mathematical appendix contains proofs of the results in the paper.

2 Stable reduced form case

In this section, we present a limiting distribution theory for the break point estimator based on minimization of the 2SLS objective function in the case where the reduced form is stable. Section 2.1 describes the model and summarizes certain preliminary results. Section 2.2 presents the limiting distribution of the break point estimators in both the fixed-break and shrinking-break cases.

2.1 Preliminaries

Consider the case in which the equation of interest is a linear regression model with m breaks, 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, $z_{1,t}$ is a $p_2 \times 1$ vector of exogenous variables including the intercept, and u_t is a mean zero error. We define $p = p_1 + p_2$. Given that some regressors are endogenous, it is plausible that (1) belongs to a system of structural equations and thus, for simplicity, we refer to (1) as the "structural equation". As is commonly assumed in the literature, we require the break points to be asymptotically distinct.

Assumption 1 *$T_i^0 = [T\lambda_i^0]$, where $0 < \lambda_1^0 < ... < \lambda_m^0 < 1$.*[7]

To implement 2SLS, it is necessary to specify the reduced form for x_t. In this section, we consider the case in which the reduced form is stable,

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

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 $z_{1,t}$.

Hall, Han, and Boldea (2009) (HHB hereafter) propose the following method for estimation of the structural equation based on minimizing a 2SLS objective function. On the first stage, the reduced form for x_t is estimated via OLS using (2) and let $\hat{x}_t$ denote the resulting predicted value for x_t, that is

$$\hat{x}_t' \; = \; z_t{}'\hat{\Delta}_T \; = \; z_t{}'(\sum_{t=1}^{T} z_t z_t{}')^{-1} \sum_{t=1}^{T} z_t x_t{}'. \tag{3}$$

In the second stage, the structural equation,

$$y_t \; = \; \hat{x}_t^{'}\beta_{x,i}^* \, + \, z_{1,t}'\beta_{z_1,i}^* \, + \, \tilde{u}_t, \quad i = 1, ..., m+1; \quad t = T_{i-1}+1, ..., T_i, \tag{4}$$

is estimated via OLS for each possible m-partition of the sample, denoted by $\{T_j\}_{j=1}^m$ or $(T_1, \ldots, T_m)$. We assume:

Assumption 2 *Equation (4) is estimated over all partitions $(T_1, ..., T_m)$ such that $T_i - T_{i-1} > max\{q-1, \epsilon T\}$ for some $\epsilon > 0$ and $\epsilon < inf_i(\lambda_{i+1}^0 - \lambda_i^0)$.*

Assumption 2 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.

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

Letting $\beta_i^* = ({\beta_{x,i}^*}', {\beta_{z_1,i}^*}')'$, for a given m-partition, the estimates of $\beta^* = ({\beta_1^*}', {\beta_2^*}', ..., {\beta_{m+1}^*}')'$ are obtained by minimizing the sum of squared residuals

$$S_T(T_1, ..., T_m; \beta) = \sum_{i=1}^{m+1} \sum_{t=T_{i-1}+1}^{T_i} (y_t - \hat{x}_t' \beta_{x,i} - z_{1,t}' \beta_{z_1,i})^2 \tag{5}$$

with respect to $\beta = ({\beta_1}', {\beta_2}', ..., {\beta_{m+1}}')'$. We denote these estimators by $\hat{\beta}(\{T_i\}_{i=1}^m)$. The estimates of the break points, $(\hat{T}_1, ..., \hat{T}_m)$, are defined as

$$(\hat{T}_1, ..., \hat{T}_m) = \arg \min_{T_1,...,T_m} S_T\left(T_1, ..., T_m;\ \hat{\beta}(\{T_i\}_{i=1}^m)\right) \tag{6}$$

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

HHB focus on inference about the parameters $\beta^0 = ({\beta_1^0}', ..., {\beta_{m+1}^0}')'$, where $\beta_i^0 = ({\beta_{x,i}^0}', {\beta_{z_1,i}^0}')'$. Specifically, they derive the limiting distributions of both $\hat{\beta}$ and also various tests for parameter variation. However, to establish these results, they need to prove certain convergence results regarding the break point estimators. These results are also relevant to our analysis of the limiting distribution of the break point estimator in the fixed-break case, and so we summarize them below in a lemma. To present these results, we must state certain additional assumptions.

Assumption 3 *(i) $h_t = (u_t, v_t')' \otimes z_t$ is an array of real valued $n \times 1$ random vectors (where $n = (p+1)q$) defined on the probability space $(\Omega, \mathcal{F}, P)$, $V_T = Var[\sum_{t=1}^T h_t]$ is such that $diag[\gamma_{T,1}^{-1}, \ldots, \gamma_{T,n}^{-1}] = \Gamma_T^{-1}$ is $O(T^{-1})$ where Γ_T is the $n \times n$ diagonal matrix with the eigenvalues $(\gamma_{T,1}, \ldots, \gamma_{T,n})$ of V_T along the diagonal; (ii) $E[h_{t,i}] = 0$ and, for some $d > 2$, $\|h_{t,i}\|_d < \kappa < \infty$ for $t = 1, 2, \ldots$ and $i = 1, 2, \ldots n$ where $h_{t,i}$ is the i^{th} element of h_t; (iii) $\{h_{t,i}\}$ is near epoch dependent with respect to $\{g_t\}$ such that $\|h_t - E[h_t|\mathcal{G}_{t-m}^{t+m}]\|_2 \le \nu_m$ with $\nu_m = O(m^{-1/2})$ where $\mathcal{G}_{t-m}^{t+m}$ is a sigma- algebra based on $(g_{t-m}, \ldots, g_{t+m})$; (iv) $\{g_t\}$ is either ϕ-mixing of size $m^{-d/(2(d-1))}$ or α-mixing of size $m^{-d/(d-2)}$.*

Assumption 4 *$rank\{\Upsilon_0\} = p$ where $\Upsilon_0 = [\Delta_0, \Pi]$, $\Pi' = [I_{p_2}, 0_{p_2 \times (q-p_2)}]$, I_a denotes the $a \times a$ identity matrix and $0_{a \times b}$ is the $a \times b$ null matrix.*[8]

[8]Note that this notation is convenient for calculations involving the augmented matrix of projected endogenous regressors and observed exogenous regressors in the second stage.

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 (thereafter pd) for any* $r > 0$ *and strictly increasing in* r.

Assumption 3 allows substantial dependence and heterogeneity in $(u_t, v_t')' \otimes z_t$ but at the same time imposes sufficient restrictions to deduce a Central Limit Theorem for $T^{-1/2}\sum_{t=1}^{[Tr]} h_t$; see Wooldridge and White (1988).[9] This assumption also contains the restrictions that the implicit population moment condition in 2SLS is valid - that is $E[z_t u_t] = 0$ - and the conditional mean of the reduced form is correctly specified. Assumption 4 implies the standard rank condition for identification in IV estimation in the linear regression model[10] because Assumptions 3(ii), 4 and 6 together imply that

$$T^{-1}\sum_{t=1}^{[Tr]} z_t[x_t', z_{1,t}'] \Rightarrow Q_{ZZ}(r)\Upsilon_0 = Q_{Z,[X,Z_1]}(r) \text{ uniformly in } r \in [0,1] \tag{7}$$

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 and is analogous to Bai and Perron's (1998) Assumption A2.

Define the break fraction estimators to be $\hat{\lambda}_i = \hat{T}_i/T$, for $i = 1, 2, \ldots m$. HHB[Theorems 1 & 2] establish the following properties of these 2SLS break fraction estimators.

Lemma 1 *Let* y_t *be generated by (1),* x_t *be generated by (2),* $\hat{x}_t$ *be generated by (3) and Assumptions 1-6 hold, then (i)* $\hat{\lambda}_i \xrightarrow{p} \lambda_i^0$*,* $i = 1, 2, \ldots, m$*; (ii) for every* $\eta > 0$*, there exists* C *such that for all large* T*,* $P(T|\hat{\lambda}_i - \lambda_i^0| > C) < \eta$*,* $i = 1, 2, \ldots, m$.

Therefore, the break fraction estimator deviates from the true break fractions by a term of order in probability T^{-1}. While HHB establish the rate of convergence of $\hat{\lambda}_i$, they do not present a limiting distribution theory for these estimators.

[9] This rests on showing that under the stated conditions $\{h_t, \mathcal{G}_{-\infty}^t\}$ is a mixingale of size -1/2 with constants $c_{T,j} = n\xi_{T,j}^{-1/2} max(1, \|b_{t,j}\|_r)$; see Wooldridge and White (1988).

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

2.2 Limiting distribution of break point estimators

In this section, we present a limiting distribution for the break point estimators. We consider two different scenarios for the parameter change across regimes: when it is fixed and when it is shrinking with the sample size. Although the resulting distribution theory in each of these scenarios turns out to be different, part of the derivations are common. It is therefore convenient to present both scenarios within the following single assumption.

Assumption 7 *Let $\beta_{i+1}^0 - \beta_i^0 = \theta_{i,T}^0 = \theta_i^0 s_T$ where $s_T = T^{-\alpha}$ for some $\alpha \in [0, 1/2)$ and $i = 1, 2, \ldots m$.*

Note that under this assumption, if $\alpha = 0$ then we have the fixed break case but if $\alpha \neq 0$ then the parameter change is shrinking with the sample size but at a slower rate than $T^{-1/2}$. It should be noted that the assumption of shrinking breaks at this rate is used as a mathematical device to develop a limiting distribution theory that is designed to provide an approximation to finite sample behaviour in models with moderate-sized changes in the parameters. The simulation results in Section 4.1 provide guidance on the accuracy of this approximation for different magnitudes of parameter change.

The derivation of the limiting distribution theory below is premised on the consistency and the known rate of convergence of the break fraction estimators. These are already presented in Lemma 1 for the fixed-break case. The corresponding results for the shrinking-break case are presented in the following proposition.

Proposition 1 *Let y_t be generated by (1), x_t be generated by (2), $\hat{x}_t$ be generated by (3) and Assumptions 1-7 ($\alpha \neq 0$) hold, then (i) $\hat{\lambda}_i \xrightarrow{p} \lambda_i^0$, $i = 1, 2, \ldots, m$; (ii) for every $\eta > 0$, there exists $C > 0$ such that for all large T, $P(T|\hat{\lambda}_i - \lambda_i^0| > C s_T^{-2}) < \eta$, $i = 1, 2, \ldots, m$.*

Remark 1: Proposition 1(ii) states that the break point estimator converges to the true break point at a rate equal to the inverse of the square of the rate at which the difference between the regimes disappears. Note that this is the same rate of convergence as is exhibited by the corresponding statistic in the case where x_t and u_t are uncorrelated and the model is estimated by OLS; see Bai (1997b)[Proposition 1].

We now turn to the issue of characterizing the limiting distribution of $\hat{T}_i$. To achieve this end, we first present the statistic that determines the large sample behaviour of the break point estimator; see Proposition 2 below. The form of this statistic is the same for both the fixed-break and the shrinking-break cases, but its large sample behaviour is different across the two cases. We therefore consider the form of the limiting distribution in the fixed-break and shrinking-break cases in turn.

From Lemma 1(ii) and Proposition 1(ii), it follows that in considering the limiting behaviour of $\{\hat{T}_i\}_{i=1}^m$ we can confine attention to possible break points within the following set $B = \cup_{i=1}^m B_i$ where $B_i = \{|T_i - T_i^0| \ \leq C_i s_T^{-2}\}$.[11]

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

$$\hat{T}_i - T_i^0 \ = \ argmin_{T_i \in B_i} \begin{cases} \Psi_T(T_i), \ for \ T_i \neq T_i^0 \\ 0, \ for \ T_i = T_i^0 \end{cases} \tag{8}$$

where

$$\begin{aligned} \Psi_T(T_i) \quad &= \quad (-1)^{\mathcal{I}[T_i < T_i^0]} 2\theta_{T,i}^{0\prime} \Upsilon_0' \sum_{t=(T_i \wedge T_i^0)+1}^{T_i \vee T_i^0} z_t \left(u_t \ + \ v_t' \beta_x^0(t,T) \right) \\ &\quad + \theta_{T,i}^{0\prime} \Upsilon_0' \sum_{t=(T_i \wedge T_i^0)+1}^{T_i \vee T_i^0} z_t z_t' \Upsilon_0 \theta_{T,i}^0 \ + \ o_p(1), \ uniformly \ in \ B_i, \end{aligned}$$

$\beta_x^0(t,T) = \beta_{x,i}^0$ for $t = T_{i-1}^0 + 1, T_{i-1}^0 + 2, \ldots, T_i^0$ and $i = 1, 2, \ldots, m+1$, $a \vee b = max\{a, b\}$, $a \wedge b = min\{a, b\}$, and $\mathcal{I}[\cdot]$ is an indicator variable that takes the value one if the event in the square brackets occurs.

We now consider the implications of Proposition 2 for the limiting distribution of the break point estimator in the two scenarios about the magnitude of the break.

(i) Fixed-break case:

If Assumption 7 holds with $\alpha = 0$ then, without further restrictions, the limiting distribution of the random variable on the right-hand side of (8) is intractable. A similar problem is encountered

[11] See Han (2006) or an earlier version of this paper Hall, Han, and Boldea (2007) for a formal proof of this assertion.

by Bai (1997b) in his analysis of the break points in models estimated by OLS. He circumvents this problem by restricting attention to strictly stationary processes.[12] We impose the same restriction here.

Assumption 8 *The process $\{z_t, u_t, v_t\}_{t=-\infty}^{\infty}$ is strictly stationary.*

To facilitate the presentation of the limiting distribution of $\hat{T}_i$, we introduce a stochastic process $R_i^*(s)$ on the set of integers that is defined as follows:

$$R_i^*(s) = \begin{cases} R_1^{(i)}(s) & : \quad s < 0 \\ 0 & : \quad s = 0 \\ R_2^{(i)}(s) & : \quad s > 0 \end{cases}$$

with

$$\begin{aligned} R_1^{(i)}(s) &= \theta_i^{0\prime}\Upsilon_0' \sum_{t=s+1}^{0} z_t z_t' \Upsilon_0 \theta_i^0 - 2\theta_i^{0\prime}\Upsilon_0' \left(\sum_{t=s+1}^{0} z_t u_t + \sum_{t=s+1}^{0} z_t v_t' \beta_{x,i}^0 \right) \\ &\quad \text{for } s = -1, -2, \cdots \\ R_2^{(i)}(s) &= \theta_i^{0\prime}\Upsilon_0' \sum_{t=1}^{s} z_t z_t' \Upsilon_0 \theta_i^0 + 2\theta_i^{0\prime}\Upsilon_0' \left(\sum_{t=1}^{s} z_t u_t + \sum_{t=1}^{s} z_t v_t' \beta_{x,i+1}^0 \right) \\ &\quad \text{for } s = 1, 2, \cdots \end{aligned}$$

We note that if (z_t, u_t, v_t) is independent over t then the process $R_i^*(s)$ is a two-sided random walk with stochastic drifts. It is necessary to impose a restriction on the random variables that drive $R_i^*(s)$.

Assumption 9 $(z_t'\Upsilon_0\theta_i^0)^2 \pm 2\theta_i^{0\prime}\Upsilon_0' z_t(u_t + v_t'\beta_{x,i}^0)$ *has a continuous distribution for $i = 1, 2, \ldots m$, and Assumption 3 (iii),(iv) holds with h_t replaced by z_t.*

Assumption 3 (iii), (iv) for z_t and h_t together ensure that $(z_t'\Upsilon_0\theta_i^0)^2 \pm 2\theta_i^{0\prime}\Upsilon_0' z_t(u_t + v_t'\beta_{x,i}^0)$ is also near-epoch dependent of the same size as h_t, and also satisfies Assumption 3 (iii), (iv), by Theorems 17.8 and 17.12 in Davidson (1994), pp. 267-269. We now present the limiting distribution of the break points in the fixed break case.

Theorem 1 *Let y_t be generated by (1), x_t be generated by (2), $\hat{x}_t$ be generated by (3) and Assumptions 1-6, 7 (with $\alpha = 0$), 8 and 9 hold then:*

$$\hat{T}_i - T_i^0 \xrightarrow{d} \arg\min_s \quad R_i^*(s)$$

for $i = 1, 2, \ldots, m$.

[12]This approach is also pursued by Bhattacharya (1987), Picard (1985) and Yao (1987).

Remark 2: To derive the probability function of the limiting distribution, it is necessary to know both β^0 and the distribution of (z_t', u_t, v_t'). However, under the assumptions of Theorem 1, there are cases in which the distribution of $(z_t'\Upsilon_0\theta_i^0)^2 \pm 2\theta_i^{0\prime}\Upsilon_0' z_t(u_t + v_t'\beta_{x,i}^0)$ can be described through a moment generating function that is known in the literature. For example, if there are no exogenous regressors in the structural equation ($z_t = z_{2,t}$), z_t, u_t, v_t are all scalar random variables, (z_t, u_t, v_t) is independently distributed over t, $z_t \sim \mathcal{N}(0, \sigma_z^2)$, $z_t \perp (u_t, v_t)$, $(u_t, v_t) \sim \mathcal{N}(0, \Omega)$, with Ω a 2×2 covariance matrix with $\Omega_{1,1} = \sigma_u^2$, $\Omega_{1,2} = \sigma_{uv}$, $\Omega_{2,2} = \sigma_v^2$, then the distributions of $R_1^i(s)$ with $i = 1, \ldots, m+1$, can be described by the following moment generating function:

$$\mathcal{M}_1^i(u) = \left(\varrho_i^0 \sigma_z \vartheta_i\right)^{|s|} \times [a_i(u)]^{-|s|/2} \times \exp\left\{ |s| \frac{(\rho_1^2 - \rho_{2,i}^2)u^2 + 2\rho_1\rho_{2,i}u}{2a_i(u)} \right\}$$

where $\varrho_i^0 = \theta_i^0 \Delta_0 \neq 0$, $\rho_1 = \mu_z/\sigma_z$; $\vartheta_i = \sqrt{\sigma_z^2(\varrho_i^0)^2 + \sigma_u^2 + \sigma_v^2(\beta_{i,0})^2 + 2\sigma_{uv}\beta_{i,0}}$; $\rho_{2,i} = \mu_z\varrho_i^0/\vartheta_i$; $r_i = \varrho_i^0\sigma_z/\vartheta_i$ and $a_i(u) = [1 - (1 + r_i u)] \times [1 + (1 - r_i)u]$.[13] The distribution of $R_2^i(s)$ can be described by the same moment generating function above, but with $\beta_{i,0}$ replaced with β_{i+1}^0.

Remark 3: It is interesting to contrast our Proposition 2 with Bai's (1997b)[Proposition 2] in which the limiting distribution of $\hat{T}_i$ is presented for the case in which $m = 1$, x_t and u_t are uncorrelated and (1) is estimated via OLS. In the latter case, Bai (1997b) shows that $\hat{T}_1 - T_1^0 \longrightarrow_d \arg\max_s W^*(s)$ where $W^*(s)$ has the same structure as $R_1^*(s)$ but its behaviour is driven by

$$b(x_t, u_t) \;=\; \theta_1^{0\prime} x_t' x_t \theta_1^0 \pm 2x_t u_t.$$

In contrast, the limiting distribution in Theorem 1 is driven by $b(z_t'\Upsilon_0, u_t + v_t'\beta_{x,i}^0)$. Therefore the limiting distribution in Theorem 1 is the same as would be obtained from Bai's (1997b)[Proposition 2] if y_t is regressed on $E[x_t|z_t]$ and $z_{1,t}$ using OLS.

Remark 4: The form of the limiting distribution of $\hat{T}_i$ is governed by $R_i^*(.)$. Given the assumptions of Theorem 1, the form of $R_i^*(.)$ only depends on i through θ_i^0 and $\beta_{x,i}^0$. In fact, the generic nature of this form follows from Assumptions 1, 3 and 9, implying that $\hat{T}_i$ and $\hat{T}_j$ are asymptotically independent for $i \neq j$.

[13] This result, along with details about the distribution functions and their numerical computation, can be found in Craig (1936). If we further assume that, for some regime, $\varrho_i^0 = 1$ and z_t, respectively $(u_t + v_t\beta_i^0)$ are standard normal variables, then in that regime, $z_t^2 - z_t(u_t + v_t\beta_{x,i}^0)$ is the sum of a χ_1^2 variable and an independently distributed random variable with distribution function $K_0(u)/\pi$, where $K_0(\cdot)$ is the Bessel function of the second kind of a purely imaginary argument of order zero - see e.g. Craig (1936), pp. 1. Thus, the moment generating function of $R_1^i(s)$ simplifies to $\mathcal{M}_1^i(u) = [\sqrt{2}a_i(u)]^{-|s|/2}$, with $r_i = 1/\sqrt{2}$.

In view of Remark 2, without further assumptions, the limiting distribution in Theorem 1 is not useful for inference in general because of its dependence on unknowns. Therefore, we now turn to an alternative framework that does yield practical methods of inference about the break points.

(ii) Shrinking-break case:

Impose Assumption 7 with $\alpha \neq 0$, as well as:

Assumption 10 *$T^{-1}\sum_{t=T_{i-1}^0+1}^{T_{i-1}^0+[rT]} \xrightarrow{p} rQ_i$, uniformly in $r \in (0, \lambda_i^0 - \lambda_{i-1}^0]$, where Q_i is a pd matrix of constants.*

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

$$Var\left[\begin{pmatrix} u_t \\ v_t \end{pmatrix} |z_t\right] = \Omega_i = \begin{pmatrix} \sigma_i^2 & \gamma_i' \\ \gamma_i & \Sigma_i \end{pmatrix}$$

where Ω_i is a constant, pd matrix, σ_i^2 is a scalar and Σ_i is $p_1 \times p_1$ matrix.

Assumption 10 allows the behaviour of the instrument cross product matrix to vary across regimes, but it is more restrictive than Assumption 6. Assumption 11 restricts the error processes to have constant conditional second moments within regime but allows these moments to vary across regimes.

To present the limiting distribution, it is also useful to define $\Omega_i^{1/2}$ and $Q_i^{1/2}$ to be the symmetric matrices satisfying $\Omega_i = \Omega_i^{1/2}\Omega_i^{1/2}$ and $Q_i = Q_i^{1/2}Q_i^{1/2}$. Notice that $\Omega_i^{1/2}$ can be decomposed as $\Omega_i^{1/2} = [N_1^i, N_2^i]$ where N_1^i is a $(p_1+1)\times 1$ vector and N_2^i is $(p_1+1)\times p_1$ so that $N_1^{i\prime}N_1^i = \sigma_i^2$, $N_1^{i\prime}N_2^i = \gamma_i$, $N_2^{i\prime}N_2^i = \Sigma_i$.

Theorem 2 *Under Assumptions 1-5, 7 (with $\alpha \neq 0$), 10 and 11, we have:*

$$\frac{(\theta_{i,T}^{0\prime}\Upsilon_0'Q_i\Upsilon_0\theta_{i,T}^0)^2}{\theta_{i,T}^{0\prime}\Upsilon_0'\Phi_i\Upsilon_0\theta_{i,T}^0}(\hat{T}_i - T_i^0) \xrightarrow{d} \arg\min_c Z_i(c)$$

for $i = 1, 2, \ldots, m$, where

$$\xi_i = \frac{\theta_i^{0\prime}\Upsilon_0'Q_{i+1}\Upsilon_0\theta_i^0}{\theta_i^{0\prime}\Upsilon_0'Q_i\Upsilon_0\theta_i^0} \qquad \phi_i = \frac{\theta_i^{0\prime}\Upsilon_0'\Phi_{i+1}\Upsilon_0\theta_i^0}{\theta_i^{0\prime}\Upsilon_0'\Phi_i\Upsilon_0\theta_i^0}$$

$$\Phi_i = [(N_1^i + N_2^i\beta_x^0)' \otimes Q_i^{1/2}][(N_1^i + N_2^i\beta_x^0)' \otimes Q_i^{1/2}]', \quad \textit{for } i = 1, 2, \ldots m+1$$

$$Z_i(c) = \begin{cases} |c|/2 - W_1^{(i)}(-c) & : \; c \leq 0 \\ \xi_i c/2 - \sqrt{\phi_i}W_2^{(i)}(c) & : \; c > 0 \end{cases},$$

β_x^0 is the limiting common value of $\{\beta_{x,i}^0\}$ under Assumption 7 and $W_j^{(i)}(c)$, $j = 1, 2$, for each i, are two independent Brownian motion processes defined on $[0, \infty)$, starting at the origin when $c = 0$, and $\{W_j^{(i)}(c)\}_{j=1}^2$ is independent of $\{W_j^{(k)}(c)\}_{j=1}^2$ for all $k \neq i$.

Remark 5: It is interesting to compare Theorem 2 with Bai's (1997b) Proposition 3, in which the corresponding distribution is presented for $m = 1$ in the case where x_t and u_t are uncorrelated and the model is estimated by OLS. The two limiting distributions have the same generic structure but the definitions of ξ_1, ϕ_1, and Φ_1 are different as is the scaling factor of $\hat{k} - k_0$. Inspection reveals that the result in Theorem 2 is equivalent to what would be obtained from applying Bai's (1997b) result to the case in which y_t is regressed on $E[x_t|z_t]$ and $z_{1,t}$ with error $u_t + v_t'\beta_{x,i}^0$.

Remark 6: The density of $\arg\min_c Z(c)$ is characterized by Bai (1997b) and he notes it is symmetric only if $\xi_i = 1$ and $\phi_i = 1$. It is possible to identify in our setting one special case in which $\xi_i = \phi_i = 1$, that is where $\Omega_{i+1} = \Omega_1 = \Omega$, $Q_{i+1} = Q_i = Q$.

The distributional result in Theorem 2 can be used to construct confidence intervals for T_i^0. To this end, denote: $\hat{\theta}_i = \hat{\beta}_{i+1} - \hat{\beta}_i$, $\hat{Q}_i = (\hat{T}_i - \hat{T}_{i-1})^{-1} \sum_i z_t z_t'$, where $\sum_i$ denotes sum over $t = \hat{T}_{i-1} + 1, \ldots, \hat{T}_i$, $\hat{\Omega}_i = (\hat{T}_i - \hat{T}_{i-1})^{-1} \sum_i \hat{b}_t \hat{b}_t'$, $\hat{b}_t = [\hat{u}_t, \hat{v}_t']'$, $w_t = [\hat{x}_t', z_{1,t}']'$, $\hat{u}_t = y_t - w_t'\hat{\beta}_i$, for $t = \hat{T}_{i-1} + 1, \ldots, \hat{T}_i$, $i = 1, 2, \ldots m$, $\hat{v}_t = (x_t - \hat{\Delta}_T' z_t)$, $\hat{\Omega}_i^{1/2}$ is the symmetric matrix such that $\hat{\Omega}_i = \hat{\Omega}_i^{1/2} \hat{\Omega}_i^{1/2}$, $\hat{\Omega}_i^{1/2} = [\hat{N}_1^i, \hat{N}_2^i]$ is partitioned conformably with $\Omega_i^{1/2}$,

$$\begin{aligned} \hat{\xi}_i &= \frac{\hat{\theta}_i' \hat{\Upsilon}_T' \hat{Q}_{i+1} \hat{\Upsilon}_T \hat{\theta}_i}{\hat{\theta}_i' \hat{\Upsilon}_T' \hat{Q}_i \hat{\Upsilon}_T \hat{\theta}_i}, \qquad \hat{\phi}_i = \frac{\hat{\theta}_i' \hat{\Upsilon}_T' \hat{\Phi}_{i+1} \hat{\Upsilon}_T \hat{\theta}_i}{\hat{\theta}_i' \hat{\Upsilon}_T' \hat{\Phi}_i \hat{\Upsilon}_T \hat{\theta}_i}, \\ \hat{\Phi}_i &= [(\hat{N}_1^i + \hat{N}_2^i \hat{\beta}_{x,i})' \otimes \hat{Q}_i^{1/2}][(\hat{N}_1^i + \hat{N}_2^i \hat{\beta}_{x,i})' \otimes \hat{Q}_i^{1/2}]', \end{aligned}$$

and $\hat{\Upsilon}_T = [\hat{\Delta}_T, \Pi]$. It then follows that

$$\left(\hat{T}_i - \left[\frac{a_2}{\hat{H}_i} \right] - 1, \hat{T}_i - \left[\frac{a_1}{\hat{H}_i} \right] + 1 \right) \tag{9}$$

is a $100(1-\alpha)$ percent confidence interval for T_i^0 where $[\,\cdot\,]$ denotes the integer part of the term in the brackets,

$$\hat{H}_i = \frac{(\hat{\theta}_i' \hat{\Upsilon}_T' \hat{Q}_i \hat{\Upsilon}_T \hat{\theta}_i)^2}{\hat{\theta}_i' \hat{\Upsilon}_T' \hat{\Phi}_i \hat{\Upsilon}_T \hat{\theta}_i}$$

and a_1 and a_2 are respectively the $\alpha/2^{th}$ and $(1-\alpha/2)^{th}$ quantiles for $\arg\min_s Z(s)$ which can be calculated using equations (B.2) and (B.3) in Bai (1997b). It is worth noting that even though the asymptotic distribution is symmetric, in general its finite sample approximation is not; this is due to the fact that for each i, one estimates β_x^0 by $\hat{\beta}_{x,i}$.

3 Unstable reduced form case

In this section, we present a limiting distribution theory for the break point estimator based on minimization of the 2SLS objective function in the case where the reduced form is unstable. To motivate the results presented, it is necessary to briefly summarize certain results in HHB.

For the unstable reduced form case, HHB propose a methodology for estimation of the break points in which the break points are identified in the reduced form first and then, conditional on these, the structural equation is estimated via 2SLS and analyzed for the presence of breaks using a strategy based on partitioning the sample into sub-samples within which the reduced form is stable.[14] The basic idea is to divide the break points in the structural equation into two types: (i) breaks that occur in the structural equation but not in the reduced form; (ii) breaks that occur simultaneously in both the structural equation and reduced form. HHB's methodology estimates the number and location of the breaks in (i) and (ii) separately in the following two steps.

- *Step 1:* for each sub-sample, the number of breaks in the structural equation are estimated and their locations determined using 2SLS-based methods that assume a stable reduced form.
- *Step 2:* for each break point in the reduced form in turn, a Wald statistic is used to test if this break point is also present in the structural equation. If the evidence suggests the break point is common then the location of the break point in question can be re-estimated from the structural equation.[15]

The number and location of the breaks in the structural equation is then deduced by combining the results from Steps 1 and 2. Within this methodology, two scenarios naturally arise for break point estimators.

- *Scenario 1:* Step 1 involves a scenario in which break point estimators that only pertain to the structural equation are obtained by minimizing a 2SLS criterion that assumes a stable reduced form over sub-samples with potentially random end-points.

[14]This partitioning is crucial for obtaining pivotal statistics and confidence intervals for the break estimators in the structural equation of interest.

[15]There are two options at this point. In addition to the option given in the text, inference about the break point can be based on the reduced form estimation.

- *Scenario 2:* Step 2 involves a scenario in which a single break point is estimated by minimizing a 2SLS criterion that assumes an unstable reduced form over sub-samples with potentially random end-points and with the break points in the reduced form estimated (consistently) *a priori* and imposed in the construction of $\hat{x}_t$.

In this section, we present a distribution theory for both scenarios. To that end, note that HHB develop their analysis under the assumption that the breaks in the reduced form are fixed and $\hat{\pi} = \pi^0 + O_p(T^{-1})$. As part of this analysis, they establish that the consistency and convergence rate results in Lemma 1 extend to the unstable reduced form case. However, the previous section demonstrates that a shrinking-break framework is more fruitful for the development of practical methods of inference. Therefore, we adopt the same framework here and so assume shrinking-breaks in both the structural equation and the reduced form. As part of our analysis, we establish the consistency and rate of convergence for the break point estimator within this framework.

Section 3.1 describes the model and summarizes certain preliminary results. Section 3.2 presents the limiting distribution of the break point estimators.

3.1 Preliminaries

We now consider the case in which the reduced form for x_t is:

$$x_t^{'} = z_t^{'}\Delta_0^{(i)} + v_t^{'}, \qquad i = 1, 2, \ldots, h+1, \qquad t = T_{i-1}^* + 1, \ldots, T_i^* \tag{10}$$

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

Assumption 12 $T_i^* = [T\pi_i^0]$, *where* $0 < \pi_1^0 < \ldots < \pi_h^0 < 1$.

Thus, as with the structural equation, the breaks in the reduced form are assumed to be asymptotically distinct. Note that the break fractions $\{\pi_i^0\}$ may or may not coincide with $\{\lambda_i^0\}$. Let $\pi^0 = [\pi_1^0, \pi_2^0, \ldots, \pi_h^0]'$. Also note that (10) can be re-written as follows

$$x_t^{'} = \tilde{z}_t(\pi^0)^{'}\Theta_0 + v_t^{'}, \qquad t = 1, 2, \ldots, T \tag{11}$$

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.

Within our analysis, it is assumed that π^0 is 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. Once the instability of the reduced form is incorporated into $\hat{x}_t$, the 2SLS estimation is implemented in the fashion described in the preamble to Section 3. However, the presence of this additional source of instability means that it is also necessary to modify Assumption 2.

Assumption 13 *The minimization in (6) is over all partitions* $(T_1, ..., T_m)$ *such that* $T_i - T_{i-1} > max\{q-1, \epsilon T\}$ *for some* $\epsilon > 0$ *and* $\epsilon < inf_i(\lambda^0_{i+1} - \lambda^0_i)$ *and* $\epsilon < inf_j(\pi^0_{j+1} - \pi^0_j)$.

As noted in the preamble, our analysis is premised on shrinking breaks. Thus, in addition to Assumption 7 with $\alpha \neq 0$, we impose the following.

Assumption 14 $\Delta_0^{(i+1)} - \Delta_0^{(i)} = \delta^0_{i,T} \; = \; \delta^0_i s^*_T$ *where* $s^*_T = T^{-\rho}$, $\rho \in (0, 0.5)$.

Note that like Asssumption 7, Assumption 14 implies the breaks are shrinking at a rate slower than $T^{-1/2}$. It is also worth pointing out that our analysis does not require any relationship between α and ρ.

Let $\hat{\Theta}_T$ be the OLS estimator of Θ_0 from the model

$$x_t' \;=\; \tilde{z}_t(\hat{\pi})'\Theta_0 + error \qquad t = 1, 2, \cdots, T \tag{12}$$

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

$$\hat{x}_t' \;=\; \tilde{z}_t(\hat{\pi})'\hat{\Theta}_T \;=\; \tilde{z}_t(\hat{\pi})'\{\sum_{t=1}^{T}\tilde{z}_t(\hat{\pi})\tilde{z}_t(\hat{\pi})'\}^{-1}\sum_{t=1}^{T}\tilde{z}_t(\hat{\pi})x_t' \tag{13}$$

In our analysis we maintain Assumptions 3, 5 and 6 but need to replace the identification condition in Assumption 4 by the following condition.

Assumption 15 $rank\{\Upsilon^0_j\} = p$ *where* $\Upsilon^0_j = \left[\Delta_0^{(j)}, \Pi\right]$, *for* $j = 1, 2, \cdots, h+1$ *for* Π *defined in Assumption 4.*

Using a similar manipulation to (7), it can be shown that Assumption 15 implies that β^0_i is identified.

3.2 Limiting distribution theory for break point estimators

Scenario 1:

Consider the case in which the $j+1^{th}$ regime for the reduced form coincides with $\ell+1$ regimes for the structural equation that is,

Assumption 16 $\pi_j^0 < \lambda_k^0 < \lambda_{k+1}^0 < \ldots < \lambda_{k+\ell}^0 < \pi_{j+1}^0$, *for some k and ℓ such that $k+\ell \leq m$.*

Notice that Assumption 16 does not preclude the possibility that either $\lambda_{k-1}^0 = \pi_j^0$ and/or $\lambda_{k+\ell+1}^0 = \pi_{j+1}^0$, but refers to $\lambda_k^0, \ldots, \lambda_{k+\ell}^0$ as indexing breaks that only pertain to the structural equation of interest.

Let $\hat{\pi}_j$ and $\hat{\pi}_{j+1}$ be the estimators of the π_j^0 and π_{j+1}^0. We consider the estimators of $\{\lambda_i^0\}_{i=k}^{k+\ell}$ based on the sub-sample $t = [T\hat{\pi}_j]+1, \ldots, [T\hat{\pi}_{j+1}]$ that is, $\hat{\lambda}_i = \hat{T}_i/T$ where

$$(\hat{T}_k, ..., \hat{T}_{k+\ell}) = \arg\min_{T_k,\ldots,T_{k+\ell}} S_T^{(j)}\left(T_k, ..., T_{k+\ell};\ \hat{\beta}(\{T_i\}_{i=k}^{k+\ell})\right) \tag{14}$$

and

$$\begin{aligned} S_T^{(j)}(T_k, ..., T_{k+\ell}; \beta) \quad = \quad & \sum_{t=[T\hat{\pi}_j]+1}^{T_k} (y_t - \hat{x}_t'\beta_{x,k} - z_{1,t}'\beta_{z_1,k})^2 \\ & + \sum_{i=k+1}^{k+\ell} \sum_{t=T_{i-1}+1}^{T_i} (y_t - \hat{x}_t'\beta_{x,i} - z_{1,t}'\beta_{z_1,i})^2 \\ & + \sum_{t=T_{k+\ell}+1}^{[T\hat{\pi}_{j+1}]} (y_t - \hat{x}_t'\beta_{x,k+\ell+1} - z_{1,t}'\beta_{z_1,k+\ell+1})^2 \end{aligned} \tag{15}$$

where $\hat{\beta}(\{T_i\}_{i=k}^{k+\ell})$ denote the 2SLS estimators obtained by minimizing $S_T^{(j)}$ for the corresponding partition of $t = [T\hat{\pi}_j]+1, \ldots, [T\hat{\pi}_{j+1}]$.

The following proposition establishes the consistency and convergence rate of $\hat{\lambda}_i$, for $i = k, k+1, \ldots k+\ell$.

Proposition 3 *Let y_t be generated by (1), x_t be generated by (2), $\hat{x}_t$ be generated by (13) and $\hat{\lambda}_i = \hat{T}_i/T$ with $\hat{T}_i$ defined in (14). If Assumptions 1-5, 7 (with $\alpha \neq 0$), 10, 12-16 hold, then for $i = k, k+1, \ldots k+\ell$ we have: (i) $\hat{\lambda}_i \xrightarrow{p} \lambda_i^0$; (ii) for every $\eta > 0$, there exists $C > 0$ such that for all large T, $P(T|\hat{\lambda}_i - \lambda_i^0| > Cs_T^{-2}) < \eta$.*

Remark 7: A comparison of Propositions 1 and 3 indicates that consistency and the rate of convergence are the same irrespective of whether the sample end-points are fixed or estimated breaks from the reduced forms.

Remark 8: While Proposition 3 holds irrespective of whether $\lambda^0_{k-1} = \pi^0_j$ and/or $\lambda^0_{k+\ell+1} = \pi^0_{j+1}$, we note that if either of these conditions holds then it does impact on the limiting behaviour of certain statistics considered in the proof of the proposition.[16]

Theorem 3 *Let y_t be generated by (1), x_t be generated by (2), $\hat{x}_t$ be generated by (13) and $\hat{\lambda}_i = \hat{T}_i/T$ with $\hat{T}_i$ defined in (14). If Assumptions 1-5, 7 (with $\alpha \neq 0$), 10, 12-16 hold, then for $i = k, k+1, \ldots k+\ell$ we have:*

$$\frac{(\theta^{0\prime}_{i,T}\Upsilon'_0\, Q_i \Upsilon_0 \theta^0_{i,T})^2}{\theta^{0\prime}_{i,T}\Upsilon'_0\, \Phi_i \Upsilon_0 \theta^0_{i,T}}(\hat{T}_i - T^0_i) \xrightarrow{d} \arg\min_c Z_i(c)$$

where

$$\xi_i = \frac{\theta^{0\prime}_i \Upsilon'_0 Q_{i+1} \Upsilon_0 \theta^0_i}{\theta^{0\prime}_i \Upsilon'_0 Q_i \Upsilon_0 \theta^0_i}, \qquad \phi_i = \frac{\theta^{0\prime}_i \Upsilon'_0\, \Phi_{i+1} \Upsilon_0 \theta^0_i}{\theta^{0\prime}_i \Upsilon'_0 \Phi_i \Upsilon_0 \theta^0_i},$$

Υ_0 is the common limiting value of $\{\Upsilon^0_j\}$ under Assumption 14, Φ_i is defined as in Theorem 2 and $Z_i(c)$ is defined as in Theorem 2 but with the ξ_i and ϕ_i stated here.

Remark 9: A comparison of the limiting distributions in Theorems 2 and 3 reveals that they are qualitatively the same. Thus, under the assumptions stated, the random end-points of the estimation sub-sample do not impact on the limiting distribution of the break point estimator.

The distributional result in Theorem 3 can be used to construct confidence intervals for T^0_i. To this end, we introduce the following definitions: $\hat{\theta}_i = \hat{\beta}_{i+1} - \hat{\beta}_i$, $\hat{Q}_i = (\hat{T}_i - \hat{T}_{i-1})^{-1}\sum_i z_t z'_t$, where $\sum_k$ denotes sum over $t = [\hat{\pi}_j T]+1, [\hat{\pi}_j T]+2, \ldots, \hat{T}_k$, $\sum_i$ denotes sum over $t = \hat{T}_{i-1}+1, \hat{T}_{i-1}+2, \ldots, \hat{T}_i$, for i=$k+1, \ldots k+\ell$, $\sum_{k+\ell+1}$ denotes sum over $t = \hat{T}_{k+\ell}+1, \hat{T}_{k+\ell}+2, \ldots, [\hat{\pi}_{j+1}T]$, $\hat{\Omega}_i = (\hat{T}_i - \hat{T}_{i-1})^{-1}\sum_i \hat{b}_t \hat{b}'_t$, $\hat{b}_t = [\hat{u}_t, \hat{v}'_t]'$, $w_t = [\hat{x}'_t, z'_{1,t}]'$, $\hat{u}_t = y_t - w'_t\hat{\beta}_k$, for $t = [\hat{\pi}_j T]+1, [\hat{\pi}_j T]+2, \ldots, \hat{T}_{k+1}$, $\hat{u}_t = y_t - w'_t \hat{\beta}_i$ for $t = \hat{T}_{i-1}+1, \hat{T}_{i-1}+2, \ldots, \hat{T}_i$ and $i = k+1, \ldots k+\ell$, $\hat{u}_t = y_t - w'_t\hat{\beta}_{k+\ell+1}$ for $t = \hat{T}_{k+\ell}+1, \hat{T}_{k+\ell}+2, \ldots, [\hat{\pi}_{j+1}T]$, $\hat{v}_t = (x_t - \hat{\Delta}'_j z_t)$, $\hat{\Delta}_j$ is the estimator of $\Delta^{(j)}_0$ from (13), $\hat{\Omega}^{1/2}_i$ is the symmetric matrix such that $\hat{\Omega}_i = \hat{\Omega}^{1/2}_i \hat{\Omega}^{1/2}_i$, $\hat{\Omega}^{1/2}_i = [\hat{N}^i_1, \hat{N}^i_2]$ is partitioned conformably with $\Omega^{1/2}_i$,

$$\hat{\xi}_i = \frac{\hat{\theta}'_i \hat{\Upsilon}'_{j+1} \hat{Q}_{i+1} \hat{\Upsilon}_{j+1} \hat{\theta}_i}{\hat{\theta}'_i \hat{\Upsilon}'_{j+1} \hat{Q}_i \hat{\Upsilon}_{j+1} \hat{\theta}_i}, \qquad \hat{\phi}_i = \frac{\hat{\theta}'_i \hat{\Upsilon}'_{j+1} \hat{\Phi}_{i+1} \hat{\Upsilon}_{j+1} \hat{\theta}_i}{\hat{\theta}'_i \hat{\Upsilon}'_{j+1} \hat{\Phi}_i \hat{\Upsilon}_{j+1} \hat{\theta}_i},$$

$$\hat{\Phi}_i = [(\hat{N}^i_1 + \hat{N}^i_2 \hat{\beta}_{x,i})' \otimes \hat{Q}^{1/2}_i][(\hat{N}^i_1 + \hat{N}^i_2 \hat{\beta}_{x,i})' \otimes \hat{Q}^{1/2}_i]',$$

[16]For brevity, we only present in the appendix a proof for the case in which $\lambda^0_{k-1} \neq \pi^0_j$ and $\lambda^0_{k+\ell+1} \neq \pi^0_{j+1}$. A supplemental appendix (available from the authors upon request) contains the proof for the case in which $\lambda^0_{k-1} = \pi^0_j$ and/or $\lambda^0_{k+\ell+1} = \pi^0_{j+1}$.

and $\hat{\Upsilon}_{j+1} = [\hat{\Delta}_{j+1}, \Pi]$. It then follows that

$$\left(\hat{T}_i - \left[\frac{a_2}{\hat{H}_i}\right] - 1, \hat{T}_i - \left[\frac{a_1}{\hat{H}_i}\right] + 1\right) \tag{16}$$

is a $100(1-\alpha)$ percent confidence interval for T_i^0 where $[\cdot]$ denotes the integer part of the term in the brackets,

$$\hat{H}_i = \frac{(\hat{\theta}_i' \hat{\Upsilon}_{j+1}' \hat{Q}_i \hat{\Upsilon}_{j+1} \hat{\theta}_i)^2}{\hat{\theta}_i' \hat{\Upsilon}_{j+1}' \hat{\Phi}_i \hat{\Upsilon}_{j+1} \hat{\theta}_i}$$

and a_1 and a_2 are defined as in (9).

Scenario 2:

Consider the case in which

Assumption 17 $\pi_{j-1}^0 \leq \lambda_{k-1}^0 < \pi_j^0 = \lambda_k^0 < \lambda_{k+1}^0 \leq \pi_{j+1}^0$ *for some* j *and* k.[17]

Let $\hat{\pi}_j$ be the estimator of π_j^0 obtained from the reduced form, and $\hat{\lambda}_{k-1}$, $\hat{\lambda}_{k+1}$ be estimators of λ_{k-1}^0, λ_{k+1}^0 obtained via the method described in Scenario 1 above.

We consider the estimators of λ_k^0 based on the sub-sample $t = [T\hat{\lambda}_{k-1}] + 1, \ldots, [T\hat{\lambda}_{k+1}]$ that is, $\hat{\lambda}_k = \hat{T}_k / T$ where

$$(\hat{T}_k) = \arg\min_{T_k} S_T^{(*k)}(T_k;\ \hat{\beta}(T_k)) \tag{17}$$

and

$$S_T^{(*k)}(T_k;\ \beta) = \sum_{t=[T\hat{\lambda}_{k-1}]+1}^{T_k} (y_t - \hat{x}_t'\beta_{x,k} - z_{1,t}'\beta_{z_1,k})^2 + \sum_{t=T_k+1}^{[T\hat{\lambda}_{k+1}]} (y_t - \hat{x}_t'\beta_{x,k+1} - z_{1,t}'\beta_{z_1,k+1})^2, \tag{18}$$

where $\hat{\beta}(T_k)$ denote the 2SLS obtained by minimizing $S_T^{(*k)}$ for the given partition of $t = [T\hat{\lambda}_{k-1}] + 1, \ldots, [T\hat{\lambda}_{k+1}]$.

Proposition 4 *Let* y_t *be generated by (1),* x_t *be generated by (2),* $\hat{x}_t$ *be generated by (13) and* $\hat{\lambda}_k = \hat{T}_k/T$ *with* $\hat{T}_k$ *defined in (17). If Assumptions 1-5, 7 (with* $\alpha \neq 0$*), 10, 12-17 hold, then we have: (i)* $\hat{\lambda}_k \xrightarrow{p} \lambda_k^0$*; (ii) for every* $\eta > 0$*, there exists* $C > 0$ *such that for all large* T*,* $P(T|\hat{\lambda}_k - \lambda_k^0| > Cs_T^{-2}) < \eta$.

Remark 10: A comparison of Propositions 1, 3 and 4 indicates that consistency and the rate of convergence properties are the same in all three cases covered.

[17] Note that this case can be extended to multiple common break points in the same fashion as in Section 3.2, Scenario 1.

Theorem 4 *Let y_t be generated by (1), x_t be generated by (2), $\hat{x}_t$ be generated by (13) and $\hat{\lambda}_k = \hat{T}_k/T$ with $\hat{T}_k$ defined in (17). If Assumptions 1-5, 7 (with $\alpha \neq 0$), 10, 12-17 hold, then we have:*

$$\frac{(\theta^{0\prime}_{k,T}\Upsilon_0'\, Q_k \Upsilon_0 \theta^0_{k,T})^2}{\theta^{0\prime}_{k,T}\Upsilon_0'\, \Phi_k \Upsilon_0 \theta^0_{k,T}}(\hat{T}_k - T^0_k) \xrightarrow{d} \arg\min_c Z_k(c)$$

where $Z_k(c)$, Υ_0, ξ_k, and ϕ_k are defined as in Theorem 2.

Remark 11: A comparison of the distributions in Theorems 2, 3 and 4 reveals that the limiting distributions are qualitatively the same.

The distributional result in Theorem 4 can be used to construct a confidence interval for T^0_k. The form of this interval is essentially the same as implied by (16) but with $\hat{\Upsilon}_{i+1}$ replaced by $\hat{\Upsilon}_i$ in the denominators of $\hat{\xi}_i$ and $\hat{\phi}_i$, where $i = k$ here.

4 Simulation study and empirical application

4.1 Simulations

Here, we report results of a small simulation study designed to gain insight into the accuracy of the limiting distribution approximation in both the stable and the unstable reduced form cases. The data generation process for the structural equation is taken as:

$$y_t = [1,\, x_t]'\beta^0_i \,+\, u_t, \qquad \text{for } t \,=\, [T\lambda^0_{i-1}] + 1, \ldots, [T\lambda^0_i]$$

where $i = 1, \ldots, m+1$, $\lambda^0_0 = 0, \lambda^0_{m+1} = T$ by convention.

(i) Cases I-II: Stable reduced form

In the stable reduced form setting, we consider: Case I: $m = 1$, $\lambda^0_1 = 0.5$ and Case II: $m = 2$, $\lambda^0_1 = 1/2; \lambda^0_2 = 2/3$, with scalar reduced form:

$$x_t \;=\; [1, z_t]'\delta \,+\, v_t, \qquad \text{for } t = 1, \ldots, T \tag{19}$$

when reduced form is stable, and 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: $z_t \sim i.i.d\; N(0_{(q-1)\times 1}, I_{q-1})$, and we set $T = 60, 120, 240, 480$; (β_1^0, β_2^0)=$([c, 0.1]', [-c, -0.1]')$, for $c = 0.3, 0.5, 1$; $q - 1 = 2, 4, 8$ and δ to yield the population $R^2 = 0.5$ for the regression in (19).[18] For each configuration, 1000 simulations are performed.

Table 1 reports the empirical coverage of the 90%, 95% and 99% confidence intervals based on (9), for Case I, and reveals that the magnitude of c impacts on the quality of the approximation. If $c = 0.3$ then the confidence intervals are mostly undersized, although the empirical coverage is close to the nominal level at the largest sample for which $T = 480$; if $c = 0.5$ then the confidence intervals are undersized for $T = 60, 120$ but close to nominal level for $T = 240, 480$; if $c = 1$ then the empirical coverage exceeds the nominal level for the 90% and 95% nominal intervals for $T \geq 60$. For $c = 1$, closer inspection of the empirical distribution of the break point reveals that most of its probability mass is either at the true break point or one observation off (only very rarely two or three data points off). Since, by construction, the break point confidence intervals contain at least three points, if the break point estimator is one data point off its true value, the confidence interval will necessarily contain the true value. Hence, over-coverage is unavoidable. Finally we note that the number of instruments has no discernable impact on the empirical coverage.

For the two-break case, Case II, the results are presented in Table 2 and exhibit similar patterns to the single break case, although it is important to remember when making a comparison between the two models that in the two-break model the sub-samples are inevitably smaller. Thus, coverage for $c = 0.3$ is inevitably smaller even though it improves with sample size, and for $c = 1$ we observe again patterns of over-coverage for the same reason stated for Case I.

(ii) Case III: Unstable reduced form with distinct breaks

This case pertains to Scenario 1 of Section 3.2. All aspects of the design are the same as for the stable reduced form with $m = 1$, except that $\lambda_1^0 = 0.6$, and the scalar reduced form is:

$$x_t \;=\; [1, z_t]'\delta_i \;+\; v_t, \qquad \text{for } t = [T\pi_{i-1}^0], \ldots, [T\pi_i^0] \tag{20}$$

with $i = 1, 2$, $\pi_0^0 = 1, \pi_2^0 = T$ by convention, and $\pi_1^0 = 0.5$. Thus, we have a reduced form with a break that occurs earlier than the break in the structural equation. Table 3 reports the results

[18]For this model, $\{\delta\}_j = (q-1)^{-1}\sqrt{R^2/(1-R^2)}$, with $\{\delta\}_j$ denoting the j^{th} element of δ, $j = 1, \ldots, q$; see Hahn and Inoue (2002).

from estimating the break in the structural equation from a sub-sample $[\hat{k}_1^{rf} + 1, T]$, where $\hat{k}_1^{rf}$ is the OLS break point estimator of $[T\pi_1^0]$ from the reduced form. The results are reported only for samples $T = 120, 240, 480$, to avoid small-sample issues related to not having enough observations between $[T\pi_1^0]$ and $[T\lambda_1^0]$. All patterns are similar to Case I.

(iii) Case IV: Unstable reduced form with one common break

This case pertains to Scenario 2 of Section 3.2, where the stable reduced form is as in Case III but the structural equation has two breaks: $m = 2$, with $\lambda_1^0 = 0.5$, a break common to the reduced form, and $\lambda_2^0 = 0.6$, a break pertaining only to the structural equation. We apply the same principle as in Case III to estimate λ_2^0 by $\hat{\lambda}_2$, then, as described in Section 3.2, we estimate λ_1^0 in interval $[1, [T\hat{\lambda}_2]]$ but with the reduced form calculated using the break estimate from the reduced form, $\hat{\pi}_1$. From Table 4, it is evident that using random end-points as well as pre-imposing $\hat{\pi}_1$ in the reduced form before estimation of the break in the structural equation does not affect the empirical coverage. In fact, it is interesting to note that most coverage levels are higher than in Case III.

Overall, the results suggest that the limiting distribution theory based on shrinking shifts can provide a reasonable approximation in the types of sample sizes encountered with macroeconomic data for which the amount of change is moderate but not too small. It would be interesting to develop a better understanding of the scenarios for which these intervals are appropriate but this is left to future research.

4.2 Application to the New Keynesian Phillips curve

In this sub-section, we assess the stability of the New Keynesian Phillips curve (NKPC), as formulated in Zhang, Osborn, and Kim (2008). This version of the NKPC is a linear model with regressors, some of which are anticipated to be correlated with the error. One contribution of their study is to raise the question of whether monetary policy changes have caused changes in the parameters in the NKPC. To investigate this issue, Zhang, Osborn, and Kim (2008) estimate the NKPC via Instrumental Variables and use informal methods to assess whether the parameters have exhibited discrete changes at any points in the sample. However, they provide no theoretical justification for their methods. As can be recognized from the description, the scenario above fits our framework, and in the sub-section we re-investigate the stability of the NKPC using the methods in HHB. Our resuts indicate that there is instability in the NKPC, and we use the theory developed in Section 3 to provide confidence intervals for the break point.

The data is quarterly from the US, spanning 1969.1-2005.4. The definitions of the variables are the same as theirs: 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, and $inf^e_{t+1|t}$ is taken from the Michigan inflation expectations survey.[19] With this notation, the structural equation of interest is:

$$inf_t \;=\; c_0 \;+\; \alpha_f inf^e_{t+1|t} \;+\; \alpha_b inf_{t-1} \;+\; \alpha_{og} og_t \;+\; \sum_{i=1}^{3} \alpha_i \Delta inf_{t-i} \;+\; u_t \tag{21}$$

where inf_t is inflation in (time) period t, $inf^e_{t+1|t}$ 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 $\theta = (c_0, \alpha_f, \alpha_b, \alpha_{og}, \alpha_1, \alpha_2, \alpha_3)'$ are unknown parameters. The variables $inf^e_{t+1|t}$ and og_t are anticipated to be correlated with the error u_t, and so (21) is commonly estimated via IV; *e.g.* see Zhang, Osborn, and Kim (2008) 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.[20] Hence, the reduced forms are:

$$inf^e_{t+1|t} \;=\; z_t'\delta_1 \;+\; v_{1,t} \tag{22}$$

$$og_t \;=\; z_t'\delta_2 \;+\; v_{2,t} \tag{23}$$

where:

$$z_t' = [1, inf_{t-1}, \Delta inf_{t-1}, \Delta inf_{t-2}, \Delta inf_{t-3}, inf^e_{t|t-1}, og_{t-1}, r_{t-1}, \mu_{t-1}, u_{t-1}]$$

with μ_t, r_t and u_t denoting respectively the M2 growth rate, the three-month Treasury Bill rate and the unemployment rate at time t.

Our sample comprises $T = 148$ observations. Consistent with the methodology proposed in HHB, we first need to account for any instability in the reduced forms. Using equation by

[19] While Zhang, Osborn, and Kim (2008) consider inflation expectations from different surveys as well, we focus for brevity on the Michigan survey only.

[20] See Zhang, Osborn, and Kim (2008) for evidence that such instruments are not weak in our context.

equation the methods proposed in Bai and Perron's (1998), we find two breaks in the reduced form for $inf^e_{t+1|t}$, with estimated locations 1975:2 and 1980:4, and one break in the reduced form for og_t, with estimated location 1975:1; the corresponding 95% confidence intervals are $[1974:4, 1975:3]$, $[1980:3, 1981:4]$, and $[1974:4, 1976:1]$ respectively.[21]

Following Hall, Han, and Boldea (2009), we first test for additional breaks over the subsample $[1981:1, 2005:4]$ for which the reduced form is estimated to be stable. Table 5 reports both sup-F and sup-Wald-type instability tests, with a cut-off of $\epsilon = 0.15$[22]; all results provide evidence for no additional breaks. Next, as proposed in Hall, Han, and Boldea (2009), we use Wald tests to test the structural equation over [1969:1,1980:4] for a known break at 1975 : 1, 1975 : 2, and over [1975:2,2005:4] for a known break at 1980 : 4. The Wald tests have p-values 0.0389, 0.0014, and 0.9184 respectively, indicating that only the first (true) break is common to the structural equation and the reduced forms, and that the NKPC has a break toward the end of 1974 or early 1975 but its precise location is unclear. Therefore, we re-estimate the NKPC allowing for a single unknown break in the structural equation, imposing the breaks in the reduced forms.[23] The proposed methodology in Section 3.2 indicates the break to be at 1974 : 4, with corresponding parameter estimates:

for 1969:1-1974:4

$$\begin{aligned} inf_t \;=\; & -\underset{(1.77)}{4.75} + \underset{(0.22)}{0.39}\, inf^e_{t+1|t} + \underset{(0.47)}{1.58}\, inf_{t-1} + \underset{(0.21)}{0.32}\, og_t - \underset{(0.56)}{1.48}\, \Delta inf_{t-1} - \underset{(0.46)}{1.16}\, \Delta inf_{t-2} \\ & - \underset{(0.25)}{0.42}\, \Delta inf_{t-3} \end{aligned}$$

for 1975:1-2005:4

$$\begin{aligned} inf_t \;=\; & -\underset{(0.27)}{0.84} + \underset{(0.10)}{0.51}\, inf^e_{t+1|t} + \underset{(0.08)}{0.55}\, inf_{t-1} + \underset{(0.05)}{0.06}\, og_t - \underset{(0.07)}{0.33}\, \Delta inf_{t-1} - \underset{(0.08)}{0.25}\, \Delta inf_{t-2} \\ & - \underset{(0.09)}{0.29}\, \Delta inf_{t-3} \end{aligned}$$

[21] Estimating the reduced forms jointly via the methods in Qu and Perron (2007) is not required in our framework, but may be desirable for increasing the efficiency of the break point estimates from the reduced forms and also for testing whether the reduced forms for og_t and $inf^e_{t+1|t}$ share a common break. This was not possible due to the fact that while a 20% cut-off is computationally necessary for the method in Qu and Perron (2007) to deliver sensible results for our data, the same 20% cut-off leads to excluding both break candidates 1975:1 and 1975:2.

[22] Smaller cut-offs yield similar results, indicating that the tests most likely do not suffer from end-of-sample problems.

[23] According to HHB, we should also test in [1969:1,1975:1] and [1975:2,1980:4] for an unknown break, but both the samples are too small for obtaining meaningful results.

The coefficient on output gap is insignificant, a common finding in the literature, see e.g. Gali and Gertler (1999)[24]. As Zhang, Osborn, and Kim (2008), we find that the forward looking component of inflation has become more important in recent years.[25]

Based on the result in Theorem 4, the 99%, 95% and 90% confidence intervals are all estimated to be $[1974:3, 1975:1]$.[26] It is interesting to compare our results on the breaks with those obtained in Zhang, Osborn, and Kim (2008). They report evidence of a break in the NKPC in 1974-1975 and also find evidence of break in 1980 : 4. However, their methods make no attempt to distinguish breaks in a structural equation of interest from those coming from other parts of the system that cause breaks in at least one reduced form. In contrast, our analysis does distinguish between these two types of breaks and we find evidence of a break in NKPC only at the end of 1974 with the break in 1980 being present only in one of the reduced forms. Thus our results refute evidence for 1980 : 4 as a break in the NKPC beyond the implied change it induces in the conditional mean of the expected inflation.

5 Concluding remarks

In this paper, we present a limiting distribution theory for the break point estimators in a linear regression model with multiple breaks, estimated via Two Stage Least Squares under two different scenarios: stable and unstable reduced forms. For stable reduced forms, we consider first the case where the parameter change is of fixed magnitude; in this case the resulting distribution depends on the distribution of the data and is not of much practical use for inference. Secondly, we consider the case where the magnitude of the parameter change shrinks with the sample size; in this case, the resulting distribution can be used to construct approximate large sample confidence intervals for the break point.

Due to the failure of the fixed-shifts framework to deliver pivotal statistics that can be used

[24] While measures of real marginal cost instead of output gap, as advocated by Gali and Gertler (1999), are not explored here, partly due to the still ongoing debate whether real marginal cost is accurately measured by proxies such as average unit labor cost - see Rudd and Whelan (2005), such proxies are only bound to strengthen our results.

[25] Note that the backward looking coefficient estimate is not 0.55, but 0.55-0.33=0.22, thus much smaller than the forward looking component.

[26] Note that the confidence intervals do not coincide before employing the integer part operator as in equation (9).

for the construction of approximate confidence intervals, in the unstable reduced form scenario we focus on shrinking shifts. As pointed out in Hall, Han, and Boldea (2009), handling break point estimators for the structural equation requires pre-estimating the breaks in the reduced form. In this paper, we show that pre-partitioning the sample with break points estimated from the reduced form instead of the true ones does not impact the limiting distribution of the break points that are specific to the structural equation only. Using the latter break point estimators to re-partition the sample into regions of only common breaks, we derive the limiting distribution of a newly proposed estimator for the common break point. Both scenarios allow for the magnitude of the breaks to differ across equations.

The finite sample performance of the proposed confidence intervals are illustrated via simulations and an application to the New Keynesian Phillips curve.

Our results add to the literature on break point distributions. Previous contributions have concentrated on level shifts in univariate time series models or on parameter shifts in linear regression models estimated via OLS in which the regressors are uncorrelated with the errors. Within our framework, the regressors of the linear regression model are allowed to be correlated with the error and the shifts are allowed to be nearly weakly identified at different rates across equations, encompassing a large number of applications in macroeconomics.

Mathematical Appendix

The proof of Proposition 1 rests on certain results that are presented together in Lemma A.1.

(a) Statement and proof of Lemma A.1:

Lemma A.1 If Assumptions 1-7 hold then for $w_t = [\hat{x}_t', z_{1,t}']'$ we have: (i) $\sum_{t=1}^{[Tr]} w_t \tilde{u}_t = O_p(T^{1/2})$ uniformly in $r \in [0,1]$; (ii) $\sum_{t=1}^{[Tr]} w_t w_t' = O_p(T)$ uniformly in $r \in [0,1]$.

Proof of (i): First note that: $\tilde{u}_t = u_t + (x_t - \hat{x}_t)'\beta_x^0(t,T)$, where $\beta_x^0(t,T) \equiv \beta_{x,i}^0$ for $t \in [T_{i-1}^0 + 1, \ldots, T_i^0]$, $i = 1, 2, \ldots, m$; $(x_t - \hat{x}_t)' = v_t' - z_t'(Z'Z)^{-1}Z'V$ where Z is the $T \times q$ matrix with t^{th} row z_t' and V is the $T \times p_1$ matrix with t^{th} row v_t'; $w_t = \hat{\Upsilon}_T' z_t$ where $\hat{\Upsilon}_T = [\hat{\Delta}_T, \Pi]$. Using these identities, it follows that

$$\begin{aligned}
\sum_{t=1}^{[Tr]} w_t \tilde{u}_t &= \hat{\Upsilon}_T' \sum_{t=1}^{[Tr]} z_t [u_t + v_t'\beta_x^0(t,T)] \\
&\quad - \hat{\Upsilon}_T' \left(T^{-1} \sum_{t=1}^{[Tr]} z_t z_t' \right) \left(T^{-1} \sum_{t=1}^{T} z_t z_t' \right)^{-1} \sum_{t=1}^{T} z_t v_t' \beta_x^0(t,T). \qquad (24)
\end{aligned}$$

Assumption 7 states that $\beta_x^0(t,T) = \beta_1^0 + O(s_T)$. Using this result along with Assumption 6, it follows from (24) that

$$\begin{aligned}
T^{-1/2} \sum_{t=1}^{[Tr]} w_t \tilde{u}_t &= \Upsilon_0' \Bigg\{ T^{-1/2} \sum_{t=1}^{[Tr]} z_t u_t \; + \; (I_q - A) T^{-1/2} \sum_{t=1}^{[Tr]} z_t v_t' \beta_1^0 \; - \; A T^{-1/2} \sum_{t=[Tr]+1}^{T} z_t v_t' \beta_1^0 \\
&\quad + (I_q - A) T^{-1/2} \sum_{t=1}^{[Tr]} z_t v_t' O(s_T) \; - \; A T^{-1/2} \sum_{t=[Tr]+1}^{T} z_t v_t' O(s_T) \Bigg\} \\
&\quad + o_p(1) \qquad (25)
\end{aligned}$$

where $A = Q_{ZZ}(r) Q_{ZZ}(1)^{-1}$. Under Assumption 3, it follows from Wooldridge and White (1988)[Theorem 2.11] that: $T^{-1/2} \sum_{t=1}^{[Tr]} z_t u_t = O_p(1)$ uniformly in r, and $T^{-1/2} \sum_{t=1}^{[Tr]} z_t v_t' = O_p(1)$ uniformly in r. Therefore it follows from (25) that under our assumptions part (i) holds.

Proof of (ii): We have

$$\begin{aligned}
\| T^{-1} \sum_{t=1}^{[Tr]} w_t w_t' \| &= \| \hat{\Upsilon}_T' T^{-1} \sum_{t=1}^{[Tr]} z_t z_t' \hat{\Upsilon}_T \| \; = \; \| \Upsilon_0' Q_{ZZ}(r) \Upsilon_0 \; + \; o_p(1) \| \\
&\leq \| \Upsilon_0' Q_{ZZ}(r) \Upsilon_0 \| \; + \; o_p(1) \\
&= O_p(1), \text{ uniformly in } r
\end{aligned}$$

where the last equality follows from Assumption 6. $\diamond$

(b) Proof of Proposition 1:

Part (i): The basic proof strategy is the same as that for Lemma 1 (see HHB for details) and builds from the following two properties of the error sum of squares on the second stage of the 2SLS estimation: first, since the 2SLS estimators minimize the error sum of squares in (5), it follows that

$$(1/T)\sum_{t=1}^{T}\hat{u}_t^2 \;\leq\; (1/T)\sum_{t=1}^{T}\tilde{u}_t^2 \tag{26}$$

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 $\tilde{u}_t = y_t - \hat{x}_t^{'}\beta_{x,i}^0 - z_{1,t}'\beta_{z_1,i}^0$ denotes the corresponding residuals evaluated at the true parameter value for $t \in [T_{i-1}^0+1, T_i^0]$; and second, using $d_t = \tilde{u}_t - \hat{u}_t = \hat{x}_t^{'}(\hat{\beta}_{x,j} - \beta_{x,i}^0) - z_{1,t}^{'}(\hat{\beta}_{z_1,j} - \beta_{z_1,i}^0)$ over $t \in [\hat{T}_{j-1}+1, \hat{T}_j] \cap [T_{i-1}^0+1, T_i^0]$, it follows that

$$T^{-1}\sum_{t=1}^{T}\hat{u}_t^2 \;=\; T^{-1}\sum_{t=1}^{T}\tilde{u}_t^2 \;+\; T^{-1}\sum_{t=1}^{T}{d_t}^2 \;-\; 2T^{-1}\sum_{t=1}^{T}\tilde{u}_t d_t. \tag{27}$$

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 (26)-(27) contradict each other.

From Hall, Han, and Boldea (2009) equation (60) it follows that

$$\sum_{t=1}^{T}\tilde{u}_t d_t \;=\; \tilde{U}'P_{\bar{W}^*}(\bar{W}^* - \bar{W}^0)\beta^0 \;+\; \tilde{U}'P_{\bar{W}^*}\tilde{U} \;-\; \tilde{U}'(\bar{W}^* - \bar{W}^0)\beta^0 \tag{28}$$

where P_S denotes the projection matrix of S, i.e. $P_S = S(S'S)^{-1}S'$ for any matrix S, $\bar{W}^*$ is the diagonal partition of W at $[\hat{T}_1, \hat{T}_2, \ldots, \hat{T}_m]$, W is the $T \times p$ matrix with t^{th} row $w_t' = [\hat{x}_t', z_{1,t}']$, $\bar{W}^0$ is the diagonal partition of W at $[T_1^0, T_2^0, \ldots, T_m^0]$, $\tilde{U} = [\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_T]$.

For ease of presentation, we assume $m = 2$ but the proof generalizes in a straightforward manner. Using Lemma A.1 and Assumption 7, it follows that[27]

$$\begin{aligned} \|\bar{W}^{*'}(\bar{W}^* - \bar{W}^0)\beta^0\| &\leq \|\sum_{t=(\hat{T}_1\wedge T_1^0)+1}^{\hat{T}_1\vee T_1^0} w_t w_t'(\beta_2^0 - \beta_1^0)\| + \|\sum_{t=(\hat{T}_2\wedge T_2^0)+1}^{\hat{T}_2\vee T_2^0} w_t w_t'(\beta_3^0 - \beta_2^0)\| \\ &= O_p(Ts_T), \end{aligned} \tag{29}$$

$$\begin{aligned} \|\tilde{U}'(\bar{W}^* - \bar{W}^0)\beta^0\| &\leq \|\sum_{t=(\hat{T}_1\wedge T_1^0)+1}^{\hat{T}_1\vee T_1^0} \tilde{u}_t w_t'(\beta_2^0 - \beta_1^0)\| + \|\sum_{t=(\hat{T}_2\wedge T_2^0)+1}^{\hat{T}_2\vee T_2^0} \tilde{u}_t w_t'(\beta_3^0 - \beta_2^0)\| \\ &= O_p(T^{1/2}s_T). \end{aligned} \tag{30}$$

[27] The symbols $\vee$ and $\wedge$ are defined in Proposition 2.

From (28)-(30), it follows that $\sum_{t=1}^{T} \tilde{u}_t d_t = O_p(T^{1/2} s_T)$; notice that this holds irrespective of the relationship between $\{\hat{T}_i\}$ and $\{T_i^0\}$.

Now consider $\sum_{t=1}^{T} d_t^2$. Repeating the steps in the proof of HHB[Lemma1(ii)], it follows that under the assumptions here, if one of the break fraction estimators does not converge to the true value then $\sum_{t=1}^{T} d_t^2 = O_p(T s_T)$. Thus if one of the break fraction estimators does not converge to the true value then $\sum_{t=1}^{T} d_t^2 >> \sum_{t=1}^{T} \tilde{u}_t d_t$[28] which implies (26) and (27) contradict. This establishes the desired result.

Part (ii): Without loss of generality, we assume $m = 2$ and focus on $\hat{T}_2$. Using a similar logic to HHB's proof of their Theorem 2, it follows that the desired result is established if it can be shown that for each $\eta > 0$, there exists $C > 0$ and $\epsilon > 0$ such that for large T,

$$P\left(min\{[S_T(T_1, T_2) - S_T(T_1, T_2^0)]/(T_2^0 - T_2)\} < 0\right) < \eta \qquad (31)$$

where the minimum is taken over $V_\epsilon(C) = \{|T_i^0 - T_i| \leq \epsilon T,\ i = 1, 2;\ T_2^0 - T_2 > C s_T^{-2}\}$ and we have suppressed the dependence of the residual sum of squares on the regression parameter estimators for ease of presentation. Again by similar logic to HHB, it can be shown that

$$\frac{S_T(T_1, T_2) - S_T(T_1, T_2^0)}{T_2^0 - T_2} \geq N_1 - N_2 - N_3 \qquad (32)$$

where

$$\begin{aligned} N_1 &= (\hat{\beta}_3^* - \hat{\beta}_\Delta)' \left(\frac{W_\Delta' W_\Delta}{T_2^0 - T_2}\right) (\hat{\beta}_3^* - \hat{\beta}_\Delta) \\ N_2 &= (\hat{\beta}_3^* - \hat{\beta}_\Delta)' \left(\frac{W_\Delta' \bar{W}}{T_2^0 - T_2}\right) \left(\frac{\bar{W}'\bar{W}}{T}\right)^{-1} \left(\frac{\bar{W}' W_\Delta}{T}\right) (\hat{\beta}_3^* - \hat{\beta}_\Delta) \\ N_3 &= (\hat{\beta}_2^* - \hat{\beta}_\Delta)' \left(\frac{W_\Delta' W_\Delta}{T_2^0 - T_2}\right) (\hat{\beta}_2^* - \hat{\beta}_\Delta) \end{aligned}$$

where $\hat{\beta}_2^*$ is the 2SLS estimator of the regression parameter based on $t = T_1 + 1, \ldots, T_2$, $\hat{\beta}_\Delta$ is the 2SLS estimator of the regression parameter based on $t = T_2 + 1, \ldots, T_2^0$, $\hat{\beta}_3^*$ is the 2SLS estimator of the regression parameter based on $t = T_2^0 + 1, \ldots, T$, $W_\Delta = [0_{p \times T_2}, w_{T_2+1}, \ldots, w_{T_2^0}, 0_{p \times (T - T_2^0)}]'$ and $\bar{W}$ is the diagonal partition of W at $[T_1, T_2]$.

Since $(T_2^0 - T_2)^{-1} W_\Delta' \bar{W} = O_p(1)$ for large enough C and $T^{-1} \bar{W}'\bar{W} = O_p(1)$ from Lemma A.1(ii), it follows that

$$\left\| \frac{W_\Delta' \bar{W}}{T} \right\| = \left\| \frac{T_2^0 - T_2}{T} \left(\frac{W_\Delta' \bar{W}}{T_2^0 - T_2}\right) \right\| \leq \epsilon \left\| \frac{W_\Delta' \bar{W}}{T_2^0 - T_2} \right\| = \epsilon O_p(1)$$

[28] Here, the symbol '>>' denotes 'of a larger order in probability'.

and so N_1 dominates N_2 for large T, small ϵ. To show that N_1 also dominates N_3, we must consider the behaviour of $\hat{\beta}_2^*$, $\hat{\beta}_\Delta$ and $\hat{\beta}_3^*$. It can be shown that $\hat{\beta}_\Delta = \beta_2^0 + O_p(T^{-1/2})$ for large C, and $\hat{\beta}_3^* = \beta_3^0 + O_p(T^{-1/2})$. For $\hat{\beta}_2^*$, we note that

$$\hat{\beta}_2^* \;=\; \beta_2^0 + \left(\sum_{t=T_1+1}^{T_2} w_t w_t' \right)^{-1} \sum_{t=T_1+1}^{T_2} w_t \tilde{u}_t + \left(\sum_{t=T_1+1}^{T_2} w_t w_t' \right)^{-1} \sum_{t=(T_1 \wedge T_1^0)+1}^{T_1 \vee T_1^0} w_t w_t' \left(\beta_1^0 - \beta_2^0\right) \mathcal{I}[T_1 < T_1^0]$$

where $\mathcal{I}[\cdot]$ is an indicator variable that takes the value one if the event in the bracket occurs. Therefore, using Lemma A.1 and Assumption 7, we have $\hat{\beta}_2^* = \beta_2^0 + O_p(T^{-1/2}) + \epsilon O_p(s_T) = \beta_2^0 + \epsilon O_p(s_T)$. Combining these results, we have $\hat{\beta}_3^* - \hat{\beta}_\Delta = \theta_{T,2}^0 + O_p(T^{-1/2})$ and $\hat{\beta}_2^* - \hat{\beta}_\Delta = \epsilon O_p(s_T)$. Therefore, it follows that

$$\begin{aligned} N_1 &= \theta_{T,2}^{0\prime} \left(\frac{W_\Delta' W_\Delta}{T_2^0 - T_2} \right) \theta_{T,2}^0 \;+\; o_p(1) \;=\; O_p(s_T^2) \\ N_3 &= \epsilon O_p(s_T) \left(\frac{W_\Delta' W_\Delta}{T_2^0 - T_2} \right) \epsilon O_p(s_T) \;=\; \epsilon^2 O_p(s_T^2) \end{aligned}$$

and so $N_1 >> N_3$ for small enough ϵ. Furthermore, we note that

$$\frac{W_\Delta' W_\Delta}{T_2^0 - T_2} \;=\; \hat{\Upsilon}_T' \left(\frac{\sum_{t=T_2+1}^{T_2^0} z_t z_t'}{T_2^0 - T_2} \right) \hat{\Upsilon}_T$$

has eigenvalues that are non-negative by construction and, by Assumptions 3 - 5, bounded away from zero for large C with large probability. This implies that for small ϵ and large C and large T, (31) holds. $\diamond$

Proof of Proposition 2

For ease of presentation we focus on the case with two breaks; the proof generalizes in a straightforward fashion to $m > 2$.

We can equivalently define the break point estimators via

$$(\hat{T}_1, \hat{T}_2) \;=\; argmin_{(T_1,T_2)\in B} \left[SSR(T_1,T_2) - SSR(T_1^0, T_2^0)\right] \tag{33}$$

where $SSR(T_1, T_2)$ denotes the residual sum of squares from the second-step regression in 2SLS of the structural equation assuming breaks at (T_1, T_2).

Clearly the case of $T_i = T_i^0$, $i = 1,2$ is trivial and so we concentrate on $T_i \neq T_i^0$ for at least one $i = 1,2$. Define $\hat{\beta}_i = \hat{\beta}_i(T_1, T_2)$ and $\tilde{\beta}_i = \hat{\beta}_i(T_1^0, T_2^0)$, for $i = 1,2$.[29] We first show

[29] This involves an abuse of notation with respect to the definition of $\hat{\beta}_i$ in Section 2.1 but the interpretation is clear from the context.

that $T^{1/2}(\hat{\beta}_i - \tilde{\beta}_i) = o_p(1)$, for $i = 1, 2, 3$, $u.B.$ where $u.B$ stands for "uniformly in B". We concentrate on the case for $i = 1$; the proof is easily extended to the other two cases. We have

$$T^{1/2}(\hat{\beta}_1 - \beta_1^0) = \left(T^{-1}\sum_{t=1}^{T_1} w_t w_t'\right)^{-1}\left(T^{-1/2}\sum_{t=1}^{T_1} w_t \tilde{u}_t\right) + \left(T^{-1}\sum_{t=1}^{T_1} w_t w_t'\right)^{-1} \times \left(T^{-1/2}\sum_{t=(T_1\wedge T_1^0)+1}^{T_1\vee T_1^0} w_t w_t'\right)(\beta_2^0 - \beta_1^0)\mathcal{I}[T_1 > T_1^0], \tag{34}$$

and

$$T^{1/2}(\tilde{\beta}_1 - \beta_1^0) = \left(T^{-1}\sum_{t=1}^{T_1^0} w_t w_t'\right)^{-1}\left(T^{-1/2}\sum_{t=1}^{T_1^0} w_t \tilde{u}_t\right) \tag{35}$$

To analyze $T^{1/2}(\hat{\beta}_1 - \tilde{\beta}_1)$, note that:[30]

$$\begin{aligned}\left(T^{-1}\sum_{t=1}^{T_1} w_t w_t'\right)^{-1} &= \left(T^{-1}\sum_{t=1}^{T_1^0} w_t w_t'\right)^{-1} + \\ &\quad \left(T^{-1}\sum_{t=1}^{T_1^0} w_t w_t'\right)^{-1}\left(T^{-1}\sum_{t=(T_1\wedge T_1^0)+1}^{T_1\vee T_1^0} w_t w_t'(-1)^{\mathcal{I}[T_1<T_1^0]}\right)\left(T^{-1}\sum_{t=1}^{T_1} w_t w_t'\right)^{-1} \\ &= \left(T^{-1}\sum_{t=1}^{T_1^0} w_t w_t'\right)^{-1} + O_p(T^{-1}s_T^{-2}),\end{aligned} \tag{36}$$

and

$$\begin{aligned}T^{-1/2}\sum_{t=(T_1\wedge T_1^0)+1}^{T_1\vee T_1^0} w_t\tilde{u}_t &= \Upsilon_T'\left(T^{-1/2}\sum_{t=(T_1\wedge T_1^0)+1}^{T_1\vee T_1^0} z_t[u_t + v_t'\beta_x^0(t,T)]\right. \\ &\quad \left. + T^{-1/2}\sum_{t=(T_1\wedge T_1^0)+1}^{T_1\vee T_1^0} z_t z_t'(\Delta_0 - \hat{\Delta})\beta_x^0(t,T)\right) \\ &= O_p(T^{-1/2}s_T^{-1}) + O_p(T^{-1}s_T^{-2}) = O_p(T^{-1/2}s_T^{-1}).\end{aligned} \tag{37}$$

From (34)-(37), it follows that $T^{1/2}(\hat{\beta}_1 - \tilde{\beta}_1) = O_p(T^{-1/2}s_T^{-1})$. Similar arguments yield $T^{1/2}(\hat{\beta}_i - \tilde{\beta}_i) = O_p(T^{-1/2}s_T^{-1})$ for $i = 2, 3$.

Now consider $SSR(T_1, T_2) - SSR(T_1^0, T_2^0)$. Using $\hat{u}_t(\beta) = \tilde{u}_t + w_t'[\beta^0(t,T) - \beta]$, we have

$$\hat{u}_t(\beta)^2 = \tilde{u}_t + 2[\beta^0(t,T) - \beta]' w_t \tilde{u}_t + [\beta^0(t,T) - \beta]' w_t w_t' [\beta^0(t,T) - \beta]$$

and so

$$SSR(T_1, T_2) - SSR(T_1^0, T_2^0) = \sum_{t=1}^{T} a_t + 2\sum_{t=1}^{T} c_t = A + 2C, \text{ say,} \tag{38}$$

[30]The first identity uses $A^{-1} = B^{-1} + B^{-1}(B - A)A^{-1}$.

where

$$a_t \;=\; [\tilde{\beta}(t,T) \,-\, \hat{\beta}(t,T)]' w_t w_t' \{[\beta^0(t,T) - \tilde{\beta}(t,T)] \,+\, [\beta^0(t,T) - \hat{\beta}(t,T)]\}, \tag{39}$$

$$c_t \;=\; [\tilde{\beta}(t,T) \,-\, \hat{\beta}(t,T)]' w_t \tilde{u}_t, \tag{40}$$

$$\hat{\beta}(t,T) \;=\; \hat{\beta}_i, \text{ for } t \in [T_{i-1}+1, \ldots, T_i], \; i = 1,2,3, \; T_0 = 1, \; T_3 = T,$$

$$\tilde{\beta}(t,T) \;=\; \tilde{\beta}_i, \text{ for } t \in [T^0_{i-1}+1, \ldots, T^0_i], \; i = 1,2,3, \; T^0_0 = 1, \; T^0_3 = T.$$

Define $B^c \equiv [1,T] \setminus (B_1 \cup B_2)$. Then

$$A \;=\; \sum_{B_1} a_t \,+\, \sum_{B_2} a_t \,+\, \sum_{B^c} a_t \tag{41}$$

where $\sum_{B_i}$ denotes sum over $t \in B_i$ and $\sum_{B^c}$ denotes sum over $t \in B^c$. On B^c, we have $T^{1/2}[\tilde{\beta}(t,T) \,-\, \hat{\beta}(t,T)] = T^{1/2}[\tilde{\beta}_i - \hat{\beta}_i] = O_p(T^{-1/2} s_T^{-1}) = o_p(1)$. On B_1, we have

$$\begin{aligned} T^{1/2}[\tilde{\beta}(t,T) \,-\, \hat{\beta}(t,T)] \;&=\; T^{1/2}(\tilde{\beta}_1 - \hat{\beta}_2)\,\mathcal{I}[T_1 < T^0_1] \,+\, T^{1/2}(\tilde{\beta}_2 - \hat{\beta}_1)\,\mathcal{I}[T_1 > T^0_1] \\ &=\; \{T^{1/2}(\tilde{\beta}_1 - \beta^0_1) - T^{1/2}(\hat{\beta}_2 - \beta^0_2) + T^{1/2}(\beta^0_1 - \beta^0_2)\}\,\mathcal{I}[T_1 < T^0_1] \\ &\quad +\{T^{1/2}(\tilde{\beta}_2 - \beta^0_2) - T^{1/2}(\hat{\beta}_1 - \beta^0_1) + T^{1/2}(\beta^0_2 - \beta^0_1)\}\mathcal{I}[T_1 > T^0_1] \\ &=\; (-1)^{\mathcal{I}[T_1 < T^0_1]} T^{1/2} \theta^0_{T,1} \,+\, O_p(1), \end{aligned}$$

where the last identity uses (35) to deduce $T^{1/2}(\tilde{\beta}_i - \beta^0_i) \,=\, O_p(1)$ and then the latter result in conjunction with $T^{1/2}(\hat{\beta}_i - \tilde{\beta}_i) \,=\, o_p(1)$ (shown above) to deduce $T^{1/2}(\hat{\beta}_i - \beta^0_i) \,=\, O_p(1)$. Similarly, we have on B_2: $T^{1/2}[\tilde{\beta}(t,T) \,-\, \hat{\beta}(t,T)] \,=\, (-1)^{\mathcal{I}[T_2 < T^0_2]} T^{1/2} \theta^0_{T,2} + O_p(1)$. Therefore, we have for $i = 1,2$,

$$\begin{aligned} \sum_{B_i} a_t \;&=\; [T^{1/2}\theta^0_{T,i} \,+\, O_p(1)]' T^{-1} \sum_{t=(T_i \wedge T^0_i)+1}^{T_i \vee T^0_i} w_t w_t' [T^{1/2}\theta^0_{T,i} \,+\, O_p(1)] \\ &=\; \theta^{0\prime}_{T,i} \Upsilon_0' \sum_{t=(T_i \wedge T^0_i)+1}^{T_i \vee T^0_i} z_t z_t' \theta^0_{T,i} \,+\, o_p(1), \; u.B \end{aligned} \tag{42}$$

In contrast, we have:

$$\sum_{B^c} a_t \;=\; O_p(T^{-1} s_T^{-1}) O_p(T) O_p(T^{-1/2}) \;=\; o_p(1), \; u.B \tag{43}$$

From (42)-(43), it follows that

$$A \;=\; \sum_{i=1}^{2} \left\{ \theta^{0\prime}_{T,i} \Upsilon_0' \sum_{t=(T_i \wedge T^0_i)+1}^{T_i \vee T^0_i} z_t z_t' \Upsilon_0 \theta^0_{T,i} \right\} \,+\, o_p(1), \; u.B \tag{44}$$

By similar arguments, we have for:

$$C \;=\; \sum_{B_1} c_t \;+\; \sum_{B_2} c_t \;+\; \sum_{B^c} c_t, \tag{45}$$

where

$$\begin{aligned}\sum_{B^c} c_t \;&=\; T^{1/2}[\tilde{\beta}(t,T) \;-\; \hat{\beta}(t,T)] \left(T^{-1/2} \sum_{B^c} w_t \tilde{u}_t \right) \;=\; O_p(T^{-1/2} s_T^{-1}) O_p(1) = o_p(1),\; u.B \\ \sum_{B_i} c_t \;&=\; (-1)^{\mathcal{I}[T_i < T_i^0]} \theta_{T,i}^{0\prime} (\sum_{B_i} w_t \tilde{u}_t)\end{aligned}$$

From (37), we have for $i = 1, 2$[31]

$$\sum_{B_i} w_t \tilde{u}_t \;=\; \Upsilon_0' \sum_{t=(T_i \wedge T_i^0)+1}^{T_i \vee T_i^0} z_t [u_t + v_t' \beta_x^0(t,T)] \;+\; o_p(1),\; u.B.$$

Substituting these results into (45), we obtain

$$C \;=\; \sum_{i=1}^{2} \left\{ (-1)^{\mathcal{I}[T_i < T_i^0]} \theta_{T,i}^{0\prime} \Upsilon_0' \sum_{t=(T_i \wedge T_i^0)+1}^{T_i \vee T_i^0} z_t [u_t + v_t' \beta_x^0(t,T)] \right\} + o_p(1),\; u.B. \tag{46}$$

The proof is completed by combining (38), (41), (44) and (46), and noting that by Assumption 3, the segments $[(T_1 \wedge T_1^0) + 1, T_1 \vee T_1^0]$ and $[(T_2 \wedge T_2^0) + 1, T_2 \vee T_2^0]$ are asymptotically independent. $\diamond$

Proof of Theorem 1

From Assumption 8 it follows that $\{z_t, u_t, v_t\}_{t=k+1}^{k_0}$ and $\{z_t, u_t, v_t\}_{t=k-k_0+1}^{0}$ have the same joint distribution, and so $\Psi_T(T_i)$ has the same distribution as $\Psi_T(T_i - T_i^0) = R_i^*(s)$. The result then follows from Proposition 2. $\diamond$

Proof of Theorem 2:

Define the rescaled Brownian motions $W_j^{(i)}(c)$ with $c \in [0, \infty]$, $j = 1, 2$, as in Theorem 2. As the generic form of the limiting distribution is the same for each i, we prove the limiting distribution has this form for $m = 1$.[32] Since $m = 1$, we simplify the notation by setting $\hat{k} = \hat{T}_1$, $k_0 = T_1^0$, $\theta_T^0 = \theta_{1,T}^0$, $\theta_1^0 = \theta^0$, $W_j = W_j^{(1)}$, for $j = 1, 2$.

From Proposition 1(ii), it follows that in considering the limiting behaviour of $\hat{k}$ we can confine attention to possible break points within the following set $B = \{|k - k_0| \;\leq C s_T^{-2}\}$. Therefore, it suffices to consider the behaviour of $\Psi_T(k) \equiv \Psi_T(T_1)$ for $k = k_0 + [c s_T^{-2}]$ and $c \in [-C, C]$.

[31] We can repeat the steps preceeding (37) to deduce the analagous result for $\sum_{t=(T_2 \wedge T_2^0)+1}^{T_2 \vee T_2^0} w_t \tilde{u}_t$.

[32] The result generalizes straightforwardly to $m > 1$.

We first consider $c \leq 0$ (that is $k \leq k_0$). We have

$$s_T^2 \sum_{t=k+1}^{k_0} z_t z_t' \implies |c| Q_1 \tag{47}$$

$$s_T \sum_{t=k+1}^{k_0} z_t(u_t + v_t' \beta_{x,1}^0) \implies \left[(N_1^1 + N_2^1 \beta_{x,1}^0)' \otimes Q_1^{1/2}\right] W_1(-c) \tag{48}$$

It follows from (47)-(48) that, for $c \leq 0$,

$$\Psi_T(k) \Rightarrow |c| \theta^{0\prime} \Upsilon_0' Q_1 \Upsilon_0 \theta^0 - 2(\theta^{0\prime} \Upsilon_0' \Phi_1 \Upsilon_0 \theta_0)^{1/2} W_1(-c) \tag{49}$$

Similarly, for $c > 0$, we have[33]

$$\Psi_T(k) \Rightarrow |c| \theta^{0\prime} \Upsilon_0' Q_2 \Upsilon_0 \theta^0 - 2(\theta^{0\prime} \Upsilon_0' \Phi_2 \Upsilon_0 \theta^0)^{1/2} W_2(c) \tag{50}$$

where $W_2(\cdot)$ is another Brownian motion process on $[0, \infty)$. The two processes W_1 and W_2 are independent because they are the limiting processes corresponding to the asymptotically independent regimes.

Thus, we have from the Continuous Mapping Theorem that

$$s_T^2(\hat{k} - k_0) \xrightarrow{d} \arg\min_c G(c) \tag{51}$$

where

$$G(c) \equiv \begin{cases} |c| \theta^{0\prime} \Upsilon_0' Q_1 \Upsilon_0 \theta^0 - 2(\theta^{0\prime} \Upsilon_0' \Phi_1 \Upsilon_0 \theta^0)^{1/2} W_1(-c) & : \quad c \leq 0 \\ |c| \theta^{0\prime} \Upsilon_0' Q_2 \Upsilon_0 \theta^0 - 2(\theta^{0\prime} \Upsilon_0' \Phi_2 \Upsilon_0 \theta^0)^{1/2} W_2(c) & : \quad c > 0 \end{cases}$$

We now show that (51) implies the desired result. By a change of variable $c = bv$ with

$$b = \frac{\theta^{0\prime} \Upsilon_0' \Phi_1 \Upsilon_0 \theta^0}{(\theta^{0\prime} \Upsilon_0' Q_1 \Upsilon_0 \theta^0)^2}$$

it can be shown that

$$\arg\min_c G(c) = b \cdot \arg\min_v Z(v). \tag{52}$$

We now establish (52).

[33] Note we use $W_2(c) \overset{d}{=} -W_2(c)$.

For $c \leq 0$

$$\begin{aligned}
G(c) &= |c|\theta^{0\prime}\Upsilon_0'Q_1\Upsilon_0\theta^0 - 2(\theta^{0\prime}\Upsilon_0'\Phi_1\Upsilon_0\theta^0)^{1/2}W_1(-c) \\
&= |bv|\cdot\theta^{0\prime}\Upsilon_0'Q_1\Upsilon_0\theta^0 - 2(\theta^{0\prime}\Upsilon_0'\Phi_1\Upsilon_0\theta^0)^{1/2}W_1(-bv) \\
&= |v|b\cdot\theta^{0\prime}\Upsilon_0'Q_1\Upsilon_0\theta_0 - 2(\theta^{0\prime}\Upsilon_0'\Phi_1\Upsilon_0\theta^0)^{1/2}\sqrt{b}\cdot W_1(-v) \\
&= |v|\frac{\theta^{0\prime}\Upsilon_0'\Phi_1\Upsilon_0\theta^0}{(\theta^{0\prime}\Upsilon_0'Q_1\Upsilon_0\theta^0)^2}\cdot\theta^{0\prime}\Upsilon_0'Q_1\Upsilon_0\theta^0 - 2(\theta^{0\prime}\Upsilon_0'\Phi_1\Upsilon_0\theta^0)^{1/2}\frac{(\theta^{0\prime}\Upsilon_0'\Phi_1\Upsilon_0\theta_0)^{1/2}}{\theta^{0\prime}\Upsilon_0'Q_1\Upsilon_0\theta^0}W_1(-v) \\
&= |v|\frac{\theta^{0\prime}\Upsilon_0'\Phi_1\Upsilon_0\theta^0}{\theta^{0\prime}\Upsilon_0'Q_1\Upsilon_0\theta^0} - 2\frac{\theta^{0\prime}\Upsilon_0'\Phi_1\Upsilon_0\theta^0}{\theta^{0\prime}\Upsilon_0'Q_1\Upsilon_0\theta^0}W_1(-v)
\end{aligned}$$

Thus, it follows that

$$\begin{aligned}
\arg\min_c G(c) &= \arg\min_v \left\{|v|\left[\frac{\theta^{0\prime}\Upsilon_0'\Phi_1\Upsilon_0\theta^0}{\theta^{0\prime}\Upsilon_0'Q_1\Upsilon_0\theta^0}\right] - 2\left[\frac{\theta^{0\prime}\Upsilon_0'\Phi_1\Upsilon_0\theta^0}{\theta^{0\prime}\Upsilon_0'Q_1\Upsilon_0\theta^0}\right]W_1(-v)\right\} \\
&= \arg\min_v \left\{\frac{|v|}{2} - W_1(-v)\right\}\frac{\theta^{0\prime}\Upsilon_0\Phi_1\Upsilon_0\theta^0}{\theta^{0\prime}\Upsilon_0'Q_1\Upsilon_0\theta^0} \\
&= \arg\min_v \left\{\frac{|v|}{2} - W_1(-v)\right\}
\end{aligned}$$

Similarly, for $c > 0$, we have that

$$\begin{aligned}
G(c) &= v\frac{\theta^{0\prime}\Upsilon_0'\Phi_1\Upsilon_0\theta^0}{(\theta^{0\prime}\Upsilon_0'Q_1\Upsilon_0\theta^0)^2}\theta^{0\prime}\Upsilon_0'Q_2\Upsilon_0\theta_0 - 2(\theta^{0'}\Upsilon_0'\Phi_2\Upsilon_0\theta^0)^{1/2}\frac{(\theta^{0\prime}\Upsilon_0'\Phi_1\Upsilon_0\theta^0)^{1/2}}{\theta^{0\prime}\Upsilon_0'Q_1\Upsilon_0\theta^0}W_2(v) \\
&= \frac{\theta^{0\prime}\Upsilon_0'\Phi_1\Upsilon_0\theta^0}{\theta^{0\prime}\Upsilon_0'Q_1\Upsilon_0\theta^0}\left[\frac{\theta^{0\prime}\Upsilon_0'Q_2\Upsilon_0\theta^0}{\theta^{0\prime}\Upsilon_0'Q_1\Upsilon_0\theta^0}v - 2\left(\frac{\theta^{0\prime}\Upsilon_0'\Phi_2\Upsilon_0\theta^0}{\theta^{0\prime}\Upsilon_0'\Phi_1\Upsilon_0\theta^0}\right)^{1/2}W_2(v)\right] \\
&= \frac{\theta^{0\prime}\Upsilon_0'\Phi_1\Upsilon_0\theta^0}{\theta^{0\prime}\Upsilon_0'Q_1\Upsilon_0\theta^0}\left[\xi v - 2\sqrt{\phi}W_2(v)\right]
\end{aligned}$$

Thus, we have

$$\begin{aligned}
\arg\min_c G(c) &= \arg\min_v \left\{-\frac{\xi v}{2} + \sqrt{\phi}W_2(v)\right\}\frac{\theta^{0\prime}\Upsilon_0'\Phi_1\Upsilon_0\theta^0}{\theta^{0\prime}\Upsilon_0'Q_1\Upsilon_0\theta^0} \\
&= \arg\min_v \left\{-\frac{\xi v}{2} + \sqrt{\phi}W_2(v)\right\}
\end{aligned}$$

Finally, the statement in Theorem 2 can be established in the following way. Since $\Psi_T(k) \Rightarrow G(s)$ and $\arg\min_c G(c) = b\cdot\arg\min_v Z(v)$, we have $b^{-1}v_T^2(\hat{k}-k_0) \xrightarrow{d} \arg\min_v Z(v)$. Using Assumption 7, we have $b^{-1}v_T^2 = (\theta_T^{0\prime}\Upsilon_0'Q_1\Upsilon_0\theta_T^0)^2/(\theta_T^{0\prime}\Upsilon_0'\Phi_1\Upsilon_0\theta_T^0)$ and thus, the desired result follows. $\diamond$

Proof of Proposition 3:

For ease of presentation, we focus on the following model with $m = h = 1$,[34]

$$y_t \;=\; \begin{cases} (x_t', z_{1,t}')\beta_1^0 \,+\, u_t, & t \leq T_1^0 \\ (x_t', z_{1,t}')\beta_2^0 \,+\, u_t, & t > T_1^0 \end{cases} \tag{53}$$

$$x_t' \;=\; \begin{cases} z_t'\Delta_1^0 \,+\, v_t, & t \leq T_1^* \\ z_t'\Delta_2^0 \,+\, v_t, & t > T_1^* \end{cases} \tag{54}$$

with $\pi_1^0 < \lambda_1^0$. For ease of notation, we set $k_1 = [T\pi_1]$, $k_1^0 = [T\pi_1^0]$, $k_2 = [T\lambda_1]$, $k_2^0 = [T\lambda_1^0]$. Also let $\hat{k}_1^{rf}$ denote the estimator of k_1^0 based on estimation of (54) that is, $\hat{k}_i^{rf} = [T\hat{\pi}_1]$. From Bai (1997b) or Bai and Perron (1998), it follows that in the shrinking-break case we have $\hat{k}_1^{rf} \in B^* = \{k_1 : \; |k_1 - k_1^0| \leq C^*(s_T^*)^{-2}\}$ for some $C^* > 0$. We now consider the properties of $\hat{k}_2 = [T\hat{\lambda}_1]$ where $\hat{\lambda}_1$ is obtained by minimizing the 2SLS objective function for (53) using the sub-sample $[\hat{k}_1^{rf} + 1, T]$.

Proof of Part (i): The basic proof strategy is the same as that Proposition 1 (i). For ease of notation, set $\hat{k}_1 = \hat{k}_1^{rf}$. By similar arguments to (28), we have

$$\sum_{t=\hat{k}_1}^{T} \tilde{u}_t d_t \;=\; \tilde{U}' P_{\bar{W}^*}(\bar{W}^* - \bar{W}^0)\beta^0 \,+\, \tilde{U}' P_{\bar{W}^*}\tilde{U} \,-\, \tilde{U}'(\bar{W}^* - \bar{W}^0)\beta^0 \tag{55}$$

where $\bar{W}^*$ is now a diagonal partition of W at $\hat{k}_2$, $W = [w_{\hat{k}_1+1}, w_{\hat{k}_1+2}, \ldots, w_T]'$, $\bar{W}^0$ is now the diagonal partition of W at k_2^0, $\tilde{U} = [\tilde{u}_{\hat{k}_1+1}, \tilde{u}_{\hat{k}_1+2}, \ldots, \tilde{u}_T]$.

We consider the terms in (55) in turn. First consider $\bar{W}^{*\prime}\bar{W}^*$. To this end, define $\hat{\delta}(t,T) = \Delta^0(t,T) - \hat{\Delta}(t,T)$, where $\Delta^0(t,T) = \Delta_1^0\mathcal{I}\{t \leq k_1^0\} + \Delta_2^0\mathcal{I}\{t > k_1^0\}$, $\hat{\Delta}(t,T) = \hat{\Delta}_1\mathcal{I}\{t \leq \hat{k}_1\} + \hat{\Delta}_2\mathcal{I}\{t > \hat{k}_1\}$, and hence, for $t \in [\hat{k}_1 + 1, T]$:

$$\hat{\delta}(t,T) \;=\; \Delta_2^0 - \hat{\Delta}_2 \,+\, (\Delta_1^0 \,-\, \Delta_2^0)\mathcal{I}\{\hat{k}_1 \leq k_1^0, t \leq k_1^0\} \tag{56}$$

Since $\hat{k}_1 \in B^*$, it follows by standard arguments that $\hat{\Delta}_2 = \Delta_2^0 + O_p(T^{-1/2})$ and this property combined with Assumption 14 yields

$$\hat{\delta}(t,T) \;=\; O_p(T^{-1/2}) \,+\, O(s_T^*)\mathcal{I}\{\hat{k}_1 \leq k_1^0, t \leq k_1^0\} \tag{57}$$

It therefore follows that

$$\bar{W}^{*\prime}\bar{W}^* \;=\; \sum_{t=\hat{k}_1+1}^{T} w_t w_t' \;=\; \hat{\Upsilon}_2' \sum_{t=\hat{k}_1+1}^{T} z_t z_t' \hat{\Upsilon}_2 \;=\; O_p(1)O_p(T)O_p(1) \;=\; O_p(T) \tag{58}$$

[34] It is apparent from the proofs that the results extend to both end-points of the sample being random and the multiple break models under Assumption 3. See the Supplementary Appendix for the proof in which there is also a break in the structural equation at k_1^0.

where $\hat{\Upsilon}_2 = [\hat{\Delta}_2, \Pi]$.

Now consider $\bar{W}^{*\prime}\tilde{U}$. Since $\tilde{u}_t = u_t + v_t'\beta_x^0(t,T) + z_t'\hat{\delta}(t,T)\beta_x^0(t,T)$, it follows that

$$\bar{W}^{*\prime}\tilde{U} \;=\; \hat{\Upsilon}_2' \sum_{t=\hat{k}_1+1}^{T} z_t[u_t + v_t'\beta_x^0(t,T)] \;+\; \hat{\Upsilon}_2' \sum_{t=\hat{k}_1+1}^{T} z_t z_t'\hat{\delta}(t,T)\beta_x^0(t,T) \tag{59}$$

Now, we have

$$\begin{aligned}\sum_{t=\hat{k}_1+1}^{T} z_t[u_t + v_t'\beta_x^0(t,T)] &= \sum_{t=k_1^0+1}^{k_2^0} z_t[u_t + v_t'\beta_{x,1}^0] + \sum_{t=k_2^0+1}^{T} z_t[u_t + v_t'\beta_{x,2}^0] \\ &\quad + (-1)^{\mathcal{I}\{\hat{k}_1 > k_1^0\}} \sum_{t=(\hat{k}_1 \wedge k_1^0)+1}^{\hat{k}_1 \vee k_1^0} z_t[u_t + v_t'\beta_{x,1}^0] \\ &= O_p(T^{1/2}) \;+\; (-1)^{\mathcal{I}\{\hat{k}_1 > k_1^0\}} \sum_{B^*} z_t[u_t + v_t'\beta_{x,1}^0] \\ &= O_p(T^{1/2}) \;+\; O_p([s_T^*]^{-1}) \;=\; O_p(T^{1/2}).\end{aligned} \tag{60}$$

In addition, we have

$$\begin{aligned}\sum_{t=\hat{k}_1+1}^{T} z_t z_t'\hat{\delta}(t,T)\beta_x^0(t,T) &= \sum_{t=\hat{k}_1+1}^{T} z_t z_t'(\Delta_2^0 - \hat{\Delta}_2)\beta_x^0(t,T) \;+\; \sum_{t=\hat{k}_1+1}^{T} z_t z_t'(\Delta_1^0 - \Delta_2^0)\mathcal{I}\{\hat{k}_1 \le k_1^0, t \le k_1^0\}\beta_x^0(t,T) \\ &= O_p(T^{1/2}) \;+\; O_p([s_T^*]^{-1}) \;=\; O_p(T^{1/2}).\end{aligned} \tag{61}$$

Combining (59)-(61), we have that

$$\bar{W}^{*\prime}\tilde{U} \;=\; O_p(T^{1/2}). \tag{62}$$

For $\bar{W}^{*\prime}(\bar{W}^* - \bar{W}^0)\beta^0$, we have

$$\left\|\bar{W}^{*\prime}(\bar{W}^* - \bar{W}^0)\beta^0\right\| \;=\; \left\|\sum_{t=(k_2 \wedge k_2^0)+1}^{k_2 \vee k_2^0} w_t w_t'(\beta_2^0 - \beta_1^0)\right\| \;=\; O_p(T s_T) \tag{63}$$

For $\tilde{U}'(\bar{W}^* - \bar{W}^0)\beta^0$, we have

$$\left\|\tilde{U}'(\bar{W}^* - \bar{W}^0)\beta^0\right\| \;=\; \left\|\sum_{t=(k_2 \wedge k_2^0)+1}^{k_2 \vee k_2^0} \tilde{u}_t w_t'(\beta_2^0 - \beta_1^0)\right\| \tag{64}$$

and

$$\begin{aligned}\sum_{t=(k_2 \wedge k_2^0)+1}^{k_2 \vee k_2^0} w_t \tilde{u}_t &= \hat{\Upsilon}_2' \sum_{t=(k_2 \wedge k_2^0)+1}^{k_2 \vee k_2^0} z_t[u_t + v_t'\beta_x^0(t,T)] \;+\; \hat{\Upsilon}_2' \sum_{t=(k_2 \wedge k_2^0)+1}^{k_2 \vee k_2^0} z_t z_t'\hat{\delta}(t,T)\beta_x^0(t,T) \\ &= O_p(T^{1/2})\end{aligned} \tag{65}$$

Combining (64)-(65), we obtain

$$\left\|\tilde{U}'(\bar{W}^* - \bar{W}^0)\beta^0\right\| \;=\; O_p(T^{1/2}s_T) \tag{66}$$

From (55), (58), (62), (63), (66), it follows that $\sum_{t=\hat{k}_1+1}^{T} \tilde{u}_t d_t \;=\; O_p(T^{1/2}s_T)$. Using similar arguments, it can be shown that $\sum_{t=\hat{k}_1+1}^{T} d_t^2 \;=\; O_p(Ts_T)$. The result then follows by similar arguments to the proof of Proposition 1 (i).

Proof of Part (ii): The general proof strategy is similar to Proposition 1 (ii). Define $V_\epsilon(C) = \{k_2 : |k_2 - k_2^0| < \epsilon T, k_2^0 - k_2 > Cs_T^{-2}\}$, SSR_1 to be the residual sum of squares from 2SLS estimation of the structural equation based on sample $[\hat{k}_1 + 1, T]$ with a break at k_2, SSR_2 to be the residual sum of squares from 2SLS estimation of the structural equation based on sample $[\hat{k}_1 + 1, T]$ with a break at k_2^0, SSR_3 to be the residual sum of squares from 2SLS estimation of the structural equation based on sample $[\hat{k}_1 + 1, T]$ with breaks at k_2 and k_2^0. By similar arguments to the proof of Proposition 1 (ii), we have

$$\frac{SSR_1 \;-\; SSR_2}{k_2^0 \;-\; k_2} \;\geq\; N_1 \;-\; N_2 \;-\; N_3 \tag{67}$$

where

$$\begin{aligned}
N_1 &= (\hat{\beta}_2^* - \hat{\beta}_\Delta)'\left(\frac{W_\Delta' W_\Delta}{k_2^0 - k_2}\right)(\hat{\beta}_2^* - \hat{\beta}_\Delta) \\
N_2 &= (\hat{\beta}_2^* - \hat{\beta}_\Delta)'\left(\frac{W_\Delta' \bar{W}}{k_2^0 - k_2}\right)\left(\frac{\bar{W}'\bar{W}}{T}\right)^{-1}\left(\frac{\bar{W}' W_\Delta}{T}\right)(\hat{\beta}_2^* - \hat{\beta}_\Delta) \\
N_3 &= (\hat{\beta}_1^* - \hat{\beta}_\Delta)'\left(\frac{W_\Delta' W_\Delta}{k_2^0 - k_2}\right)(\hat{\beta}_1^* - \hat{\beta}_\Delta)
\end{aligned}$$

where $\hat{\beta}_1^*$ is the 2SLS estimator of the regression parameter based on $t = \hat{k}_1+1, \ldots, k_2$, $\hat{\beta}_\Delta$ is the 2SLS estimator of the regression parameter based on $t = k_2+1, \ldots, k_2^0$, $\hat{\beta}_2^*$ is the 2SLS estimator of the regression parameter based on $t = k_2^0+1, \ldots, T$, $W_\Delta = [0_{p\times(k_2-\hat{k}_1)}, w_{k_2+1}, \ldots, w_{k_2^0}, 0_{p\times(T-k_2^0)}]'$ and $\bar{W}$ is the diagonal partition of W at k_2.

It is straightforward to show that N_1 dominates N_2 for small ϵ. Therefore, we focus on showing that N_1 dominates N_3 for small ϵ, large C and N_1 is positive with large probability. To this end, we start by considering the properties of the parameter estimators in N_1 and N_3. For large C, we have $\hat{\beta}_\Delta = \beta_1^0 + O_p(T^{-1/2})$ because it is based on a large sub-sample for which β_1^0 is the true parameter in the structural equation. Also we have $\hat{\beta}_2^* = \beta_2^0 + O_p(T^{-1/2})$ as it is an estimator of β_2^0 obtained from a model with the correct break imposed. Now consider $\hat{\beta}_1^*$. By definition

$$\hat{\beta}_1^* \;=\; \beta_1^0 \,+\, \left(\sum_{t=\hat{k}_1+1}^{k_2} w_t w_t'\right)^{-1} \sum_{t=\hat{k}_1+1}^{k_2} w_t \tilde{u}_t \tag{68}$$

From Assumption 10, it follows that $\sum_{t=\hat{k}_1+1}^{k_2} w_t w_t' = O_p(T)$ uniformly in $V_\epsilon(C)$. Now consider $\sum_{t=\hat{k}_1+1}^{k_2} w_t \tilde{u}_t$. We have

$$\begin{aligned}\sum_{t=\hat{k}_1+1}^{k_2} w_t \tilde{u}_t \;&=\; \hat{\Upsilon}_2' \sum_{t=\hat{k}_1+1}^{k_2} z_t(u_t + v_t'\beta_{x,1}^0) \,+\, \hat{\Upsilon}_2' \sum_{t=\hat{k}_1+1}^{k_2} z_t z_t'(\Delta_2^0 - \hat{\Delta}_2)\beta_{x,1}^0 \\ &\quad + \hat{\Upsilon}_2' \sum_{t=\hat{k}_1+1}^{k_2} z_t z_t' \mathcal{I}\{\hat{k}_1 \le k_1^0, t \le k_1^0\}\,(\Delta_1^0 - \Delta_2^0)\beta_{x,1}^0.\end{aligned}$$

Examining each term in turn, we have

$$\begin{aligned}\sum_{t=\hat{k}_1+1}^{k_2} z_t(u_t + v_t'\beta_{x,1}^0) \;&=\; O_p([s_T^*]^{-1})\,\mathcal{I}\{k_2 \le k_1^0, \hat{k}_1 \le k_1^0\} \,+\, O_p(T^{1/2})\,(1 \,-\, \mathcal{I}\{k_2 \le k_1^0, \hat{k}_1 \le k_1^0\}) \\ &=\; O_p(T^{1/2}), \\ \sum_{t=\hat{k}_1+1}^{k_2} z_t z_t'(\Delta_2^0 - \hat{\Delta}_2)\beta_{x,1}^0 \;&=\; O_p(T^{-1/2}[s_T^*]^{-2})\,\mathcal{I}\{k_2 \le k_1^0, \hat{k}_1 \le k_1^0\} \,+\, O_p(T^{1/2})\,(1 \,-\, \mathcal{I}\{k_2 \le k_1^0, \hat{k}_1 \le k_1^0\}) \\ &=\; O_p(T^{1/2})\end{aligned}$$

and

$$\hat{\Upsilon}_2' \sum_{t=\hat{k}_1+1}^{k_2} z_t z_t' \mathcal{I}\{\hat{k}_1 \le k_1^0, t \le k_1^0\}(\Delta_1^0 - \Delta_2^0)\beta_{x,1}^0 \;=\; O_p([s_T^*]^{-1}).$$

Therefore it follows that $\hat{\beta}_1^* = \beta_1^0 + O_p(T^{-1/2})$. Using the derived properties of the estimators, it follows that $\hat{\beta}_1 - \hat{\beta}_\Delta = O_p(T^{-1/2})$ and $\hat{\beta}_2 - \hat{\beta}_\Delta = \beta_2^0 - \beta_1^0 + O_p(T^{-1/2}) = O_p(s_T)$. Using these results in the formulae for N_1 and N_3, it is clear that N_1 dominates N_3. Furthermore,

$$\begin{aligned}N_1 \;&=\; (\beta_2^0 - \beta_1^0)' \frac{1}{k_2^0 - k_2} \sum_{t=k_2+1}^{k_2^0} w_t w_t'(\beta_2^0 - \beta_1^0) \,+\, o_p(1) \\ &=\; (\beta_2^0 - \beta_1^0)'\Upsilon_2^{0\prime} Q_2 \Upsilon_2^0(\beta_2^0 - \beta_1^0) \,+\, o_p(1)\end{aligned}$$

for large C and large T. Since Q_2 is pd and $\beta_2^0 - \beta_1^0 \neq 0$ for large but finite T, the required result then follows by similar arguments to the proof of Proposition 1. The case of $k_2 > k_2^0$ can be handled in a similar way and thus is omitted. $\diamond$.

Proof of Theorem 3

Consider again the model used in the proof of Proposition 3. Define $\hat{\beta}_1$ to be the 2SLS estimator based on $t \in [\hat{k}_1+1, k_2]$, $\hat{\beta}_2$ to be the 2SLS estimator based on $t \in [k_2+1, T]$, $\tilde{\beta}_1$ to be the 2SLS estimator based on $t \in [\hat{k}_1+1, k_2^0]$, and $\tilde{\beta}_2$ to be the 2SLS estimator based on $t \in [k_2^0+1, T]$. To facilitate the proof we must first consider the properties of these estimators. Note that from Proposition 3 (ii) it follows that we need to consider only $k_2 \in B_2 = \{k_2 : \ |k_2 - k_2^0| < C_2 s_T^{-2}\}$.

We have

$$T^{1/2}(\tilde{\beta}_1 \, - \, \beta_1^0) \; = \; \left(T^{-1} \sum_{t=\hat{k}_1+1}^{k_2^0} w_t w_t' \right)^{-1} T^{-1/2} \sum_{t=\hat{k}_1+1}^{k_2^0} w_t \tilde{u}_t. \qquad (69)$$

Let $[1, \hat{k}_2] \setminus (B^* \cup B_2) \equiv \bar{B}$ and note that $B^* \cap B_2 = \emptyset$. We have

$$\begin{aligned} T^{-1} \sum_{t=\hat{k}_1+1}^{k_2^0} w_t w_t' \;&=\; T^{-1} \sum_{t=k_1^0+1}^{k_2^0} w_t w_t' \, + \, (-1)^{\mathcal{I}\{\hat{k}_1 > k_1^0\}} T^{-1} \sum_{t=(\hat{k}_1 \wedge k_1^0)+1}^{\hat{k}_1 \vee k_1^0} w_t w_t' \\ &=\; O_p(1) \, + \, (-1)^{\mathcal{I}\{\hat{k}_1 > k_1^0\}} T^{-1} \sum_{B^*} w_t w_t' \\ &=\; O_p(1) \, + \, O_p(T^{-1}[s_T^*]^{-2}) \; = \; O_p(1). \end{aligned}$$

Similarly, we have

$$\begin{aligned} T^{-1/2} \sum_{t=\hat{k}_1+1}^{k_2^0} w_t \tilde{u}_t \;&=\; T^{-1/2} \sum_{t=k_1^0+1}^{k_2^0} w_t \tilde{u}_t \, + \, (-1)^{\mathcal{I}\{\hat{k}_1 > k_1^0\}} T^{-1/2} \sum_{t=(\hat{k}_1 \wedge k_1^0)+1}^{\hat{k}_1 \vee k_1^0} w_t \tilde{u}_t \\ &=\; O_p(1) \, + \, (-1)^{\mathcal{I}\{\hat{k}_1 > k_1^0\}} T^{-1/2} \sum_{B^*} w_t \tilde{u}_t \\ &=\; O_p(1) \, + \, O_p(T^{-1/2}[s_T^*]^{-1}) \; = \; O_p(1). \end{aligned}$$

Thus, it follows from (69) that $\tilde{\beta}_1 = \beta_1^0 + O_p(T^{-1/2})$. Now consider $\hat{\beta}_1 - \tilde{\beta}_1$. By definition, we have

$$\begin{aligned} T^{1/2}(\hat{\beta}_1 - \tilde{\beta}_1) \;&=\; \left(T^{-1} \sum_{t=\hat{k}_1+1}^{k_2} w_t w_t' \right)^{-1} T^{-1/2} \sum_{t=\hat{k}_1+1}^{k_2} w_t \tilde{u}_t \\ &\quad - \left(T^{-1} \sum_{t=\hat{k}_1+1}^{k_2^0} w_t w_t' \right)^{-1} T^{-1/2} \sum_{t=\hat{k}_1+1}^{k_2^0} w_t \tilde{u}_t. \qquad (70) \end{aligned}$$

We have[35]

$$\left(T^{-1}\sum_{t=\hat{k}_1+1}^{k_2} w_t w_t'\right)^{-1} = \left(T^{-1}\sum_{t=\hat{k}_1+1}^{k_2^0} w_t w_t'\right)^{-1} + \left(T^{-1}\sum_{t=\hat{k}_1+1}^{k_2^0} w_t w_t'\right)^{-1} \times$$
$$(-1)^{\mathcal{I}\{k_2>k_2^0\}} T^{-1} \sum_{t=(k_2\wedge k_2^0)+1}^{k_2\vee k_2^0} w_t w_t' \left(T^{-1}\sum_{t=\hat{k}_1+1}^{k_2} w_t w_t'\right)^{-1} \quad (71)$$
$$= \left(T^{-1}\sum_{t=\hat{k}_1+1}^{k_2^0} w_t w_t'\right)^{-1} + o_p(1), \qquad \text{uniformly in } B_2,$$

and

$$T^{-1/2}\sum_{t=\hat{k}_1+1}^{k_2} w_t\tilde{u}_t - T^{-1/2}\sum_{t=\hat{k}_1+1}^{k_2^0} w_t\tilde{u}_t = (-1)^{\mathcal{I}\{k_2<k_2^0\}}\left[T^{-1/2}\hat{\Upsilon}_2' \sum_{t=(k_2\wedge k_2^0)+1}^{k_2\vee k_2^0} z_t[u_t + v_t'\beta_x^0(t,T)]\right.$$
$$\left. + T^{-1/2}\hat{\Upsilon}_2' \sum_{t=(k_2\wedge k_2^0)+1}^{k_2\vee k_2^0} z_t z_t'(\Delta_2^0 - \hat{\Delta}_2)\beta_x^0(t,T)\right] \quad (72)$$
$$= (-1)^{\mathcal{I}\{k_2<k_2^0\}} T^{-1/2}\hat{\Upsilon}_2' \sum_{t=(k_2\wedge k_2^0)+1}^{k_2\vee k_2^0} z_t[u_t + v_t'\beta_x^0(t,T)]$$
$$+ O_p(T^{-1}s_T^{-2}), \qquad \text{uniformly in } B_2$$
$$= O_p(T^{-1/2}s_T^{-1}), \qquad \text{uniformly in } B_2.$$

Therefore, using these results in (70), we obtain $T^{1/2}(\hat{\beta}_1 - \tilde{\beta}_1) = O_p(T^{-1/2}s_T^{-1})$ uniformly in B_2.

Since $\tilde{\beta}_2$ is based on an estimation with the correct break imposed, it follows by standard arguments that $\tilde{\beta}_2 = \beta_2^0 + O_p(T^{-1/2})$. Now consider $\hat{\beta}_2 - \tilde{\beta}_2$. We have

$$T^{1/2}(\hat{\beta}_2 - \tilde{\beta}_2) = \left(T^{-1}\sum_{t=k_2+1}^{T} w_t w_t'\right)^{-1} T^{-1/2}\sum_{t=k_2+1}^{T} w_t\tilde{u}_t - \left(T^{-1}\sum_{t=k_2^0+1}^{T} w_t w_t'\right)^{-1} T^{-1/2}\sum_{t=k_2^0+1}^{T} w_t\tilde{u}_t$$
$$+\mathcal{I}\{k_2<k_2^0\}\left(T^{-1}\sum_{t=k_2+1}^{T} w_t w_t'\right)^{-1}\left(T^{-1}\sum_{t=(k_2\wedge k_2^0)+1}^{k_2\vee k_2^0} w_t w_t'\right) T^{1/2}(\beta_1^0 - \beta_2^0). \quad (73)$$

By similar arguments to (71), we have

$$\left(T^{-1}\sum_{t=k_2+1}^{T} w_t w_t'\right)^{-1} = \left(T^{-1}\sum_{t=k_2^0+1}^{T} w_t w_t'\right)^{-1} + o_p(1), \qquad \text{uniformly in } B_2,$$

and by similar arguments to (72),

$$T^{-1/2}\sum_{t=k_2+1}^{T} w_t\tilde{u}_t - T^{-1/2}\sum_{t=k_2^0+1}^{T} w_t\tilde{u}_t = (-1)^{\mathcal{I}\{k_2>k_2^0\}} T^{-1/2}\sum_{t=(k_2\wedge k_2^0)+1}^{k_2\vee k_2^0} w_t\tilde{u}_t$$
$$= O_p(T^{-1/2}s_T^{-1}), \qquad \text{uniformly in } B_2.$$

[35] Using $A^{-1} - B^{-1} = B^{-1}(B-A)A^{-1}$.

Therefore, we have

$$\left(T^{-1}\sum_{t=k_2+1}^{T} w_t w_t'\right)^{-1}\left(T^{-1}\sum_{t=(k_2\wedge k_2^0)+1}^{k_2\vee k_2^0} w_t w_t'\right) T^{1/2}(\beta_1^0-\beta_2^0) \;=\; O_p(T^{-1/2}s_T^{-1})$$

and so $T^{1/2}(\hat{\beta}_2-\tilde{\beta}_2)=O_p(T^{-1/2}s_T^{-1})$.

With this background, we now consider the distribution of $\hat{k}_2$, where

$$\hat{k}_2 \;=\; argmin_{k_2\in B_2}[SSR(\hat{k}_1,k_2)-SSR(\hat{k}_1,k_2^0)]$$

and $SSR(k_1,k_2)$ denotes the residual sum of squares in interval $[k_1+1,T]$ with partition at k_2. Obviously if $k_2=k_2^0$ then the minimand is zero, and so we concentrate on the case in which $k_2\neq k_2^0$.

Define $\hat{\beta}(t,T)=\hat{\beta}_1\mathcal{I}\{t\leq k_2\}+\hat{\beta}_2\mathcal{I}\{t>k_2\}$ and $\tilde{\beta}(t,T)=\tilde{\beta}_1\mathcal{I}\{t\leq k_2^0\}+\tilde{\beta}_2\mathcal{I}\{t>k_2^0\}$. Notice that from our previous results we have:

$$\begin{aligned}
T^{1/2}[\tilde{\beta}(t,T)\,-\,\hat{\beta}(t,T)] \;&=\; T^{1/2}(\tilde{\beta}_1\,-\,\hat{\beta}_1)\mathcal{I}\{t\leq(k_2\wedge k_2^0)\}\,+\,T^{1/2}(\tilde{\beta}_2\,-\,\hat{\beta}_2)\mathcal{I}\{t>(k_2\vee k_2^0)\}\\
&\quad+\mathcal{I}\{(k_2\wedge k_2^0)+1\leq t\leq(k_2\vee k_2^0)\}\Big[T^{1/2}(\tilde{\beta}_1\,-\,\hat{\beta}_2)\mathcal{I}\{k_2<k_2^0\}\\
&\quad+T^{1/2}(\tilde{\beta}_2\,-\,\hat{\beta}_1)\mathcal{I}\{k_2>k_2^0\}\Big]\\
&=\; O_p(T^{-1/2}s_T^{-1})\,+\,T^{1/2}s_T\theta_1^0(-1)^{\mathcal{I}\{k_2<k_2^0\}}\,+\,O_p(1).
\end{aligned}$$

Let $\bar{B}_2=[\hat{k}_1+1,T]\setminus[(k_2\wedge k_2^0)+1,k_2\vee k_2^0]$, then using similar arguments to the derivation of (38) we have

$$SSR(\hat{k}_1,k_2)-SSR(\hat{k}_1,k_2^0) \;=\; \sum_{t=\hat{k}_1+1}^{T} a_t \,+\,2\sum_{t=\hat{k}_1+1}^{T} c_t \;=\; A+2C \tag{74}$$

where $a_t=T^{1/2}[\tilde{\beta}(t,T)-\hat{\beta}(t,T)]T^{-1}w_tw_t'\Big\{T^{1/2}[\beta^0(t,T)-\hat{\beta}(t,T)]+T^{1/2}[\beta^0(t,T)-\tilde{\beta}(t,T)]\Big\}$, and $c_t=T^{1/2}[\tilde{\beta}(t,T)\,-\,\hat{\beta}(t,T)]T^{-1/2}w_t\tilde{u}_t$. Consider A and C in turn. We have $A=\sum_{B_2}a_t+\sum_{\bar{B}_2}a_t$, and

$$\begin{aligned}
\sum_{\bar{B}_2} a_t \;&=\; O_p(T^{-1/2}s_T^{-1})T^{-1}\sum_{\bar{B}_2} w_tw_t'\,O_p(1)=o_p(1),\text{ uniformly in } B_2,\\
\sum_{B_2} a_t \;&=\; T^{1/2}s_T\theta_1^{0\prime}(-1)^{\mathcal{I}\{k_2<k_2^0\}}\left(T^{-1}\sum_{B_2} w_tw_t'\right)\,\times\\
&\quad\Big\{T^{1/2}[\beta^0(t,T)-\hat{\beta}(t,T)]\,+\,T^{1/2}[\beta^0(t,T)-\tilde{\beta}(t,T)]\Big\}.
\end{aligned}$$

If $t \in B_2$ then we have

$$
\begin{aligned}
T^{1/2}[\beta^0(t,T) - \hat{\beta}(t,T)] &= T^{1/2}[\beta_1^0 - \hat{\beta}_2]\mathcal{I}\{k_2 < k_2^0\} + [\beta_2^0 - \hat{\beta}_1]\mathcal{I}\{k_2 > k_2^0\} \\
T^{1/2}[\beta^0(t,T) - \tilde{\beta}(t,T)] &= T^{1/2}[\beta_1^0 - \tilde{\beta}_1]\mathcal{I}\{k_2 < k_2^0\} + T^{1/2}[\beta_2^0 - \tilde{\beta}_2]\mathcal{I}\{k_2 > k_2^0\}
\end{aligned}
$$

and so setting $d_T = T^{1/2}[\beta^0(t,T) - \tilde{\beta}(t,T)] + T^{1/2}[\beta^0(t,T) - \hat{\beta}(t,T)]$, we have

$$
\begin{aligned}
d_T &= \left\{T^{1/2}[\beta_1^0 - \tilde{\beta}_1] + T^{1/2}[\beta_2^0 - \hat{\beta}_2] + T^{1/2}(\beta_1^0 - \beta_2^0)\right\}\mathcal{I}\{k_2 < k_2^0\} \\
&\quad + \left\{T^{1/2}[\beta_2^0 - \tilde{\beta}_2] + T^{1/2}[\beta_1^0 - \hat{\beta}_1] + T^{1/2}(\beta_2^0 - \beta_1^0)\right\}\mathcal{I}\{k_2 > k_2^0\} \\
&= T^{1/2}\theta_{T,1}^0(-1)^{\mathcal{I}\{k_2<k_2^0\}} + O_p(1), \qquad \text{uniformly in } B_2.
\end{aligned}
$$

Hence, we have

$$
\sum_{B_2} a_t = \theta_{T,1}^{0\prime}\sum_{B_2} w_t w_t' \theta_{T,1}^0 = \theta_{T,1}^{0\prime}\Upsilon_2^{0\prime}Q_2\Upsilon_2^0\theta_{T,1}^0|k_2^0 - k_2| + o_p(1), \qquad \text{uniformly in } B_2.
$$

Recalling that $A = \sum_{B_2} a_t + \sum_{\bar{B}_2} a_t$, we obtain from the above results that

$$
A = \theta_{T,1}^{0\prime}\Upsilon_2^{0\prime}Q_2\Upsilon_2^0\theta_{T,1}^0|k_2^0 - k_2| + o_p(1), \qquad \text{uniformly in } B_2. \tag{75}
$$

Similarly, we have $C = \sum_{B_2} c_t + \sum_{\bar{B}_2} c_t$ where

$$
\begin{aligned}
\sum_{\bar{B}_2} c_t &= \sum_{\bar{B}_2} T^{1/2}[\tilde{\beta}(t,T) - \hat{\beta}(t,T)]T^{-1/2}w_t\tilde{u}_t = O_p(T^{-1/2}s_T^{-1})O_p(1) = o_p(1), \qquad \text{uniformly in } B_2, \\
\sum_{B_2} c_t &= \sum_{B_2} T^{1/2}[\tilde{\beta}(t,T) - \hat{\beta}(t,T)]T^{-1/2}w_t\tilde{u}_t = [(-1)^{\mathcal{I}\{k_2<k_2^0\}}\theta_{T,1}^{0\prime} + O_p(1)]\sum_{B_2} w_t\tilde{u}_t \\
&= [(-1)^{\mathcal{I}\{k_2<k_2^0\}}\theta_{T,1}^{0\prime} + O_p(1)]\left\{\Upsilon_2^{0\prime}T^{-1/2}\sum_{t=(k_2\wedge k_2^0)+1}^{k_2\vee k_2^0} z_t[u_t + v_t'\beta_x^0(t,T)] + O_p(T^{-1}s_T^{-2})\right\} \\
&= [(-1)^{\mathcal{I}\{k_2<k_2^0\}}\theta_{T,1}^{0\prime}]\left\{\Upsilon_2^{0\prime}T^{-1/2}\sum_{t=(k_2\wedge k_2^0)+1}^{k_2\vee k_2^0} z_t[u_t + v_t'\beta_x^0(t,T)]\right\} + o_p(1), \quad \text{uniformly in } B_2.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
A + 2C &= 2(-1)^{\mathcal{I}\{k_2<k_2^0\}}\theta_{T,1}^{0\prime}\Upsilon_2^{0\prime}T^{-1/2}\sum_{t=(k_2\wedge k_2^0)+1}^{k_2\vee k_2^0} z_t[u_t + v_t'\beta_x^0(t,T)] \\
&\quad + |k_2 - k_2^0|\theta_{T,1}^{0\prime}\Upsilon_2^{0\prime}Q_2\Upsilon_2^0\theta_{T,1}^0 + o_p(1), \qquad \text{uniformly in } B_2.
\end{aligned} \tag{76}
$$

It can be recognized that (76) has the same basic structure as (8) and so the rest of the proof follows by similar arguments to the proof of Proposition 2. $\diamond$

Proof of Proposition 4

Consider the following model with $m = 2$ and $h = 1$.

$$y_t \;=\; \begin{cases} (x_t', z_{1,t}')\beta_1^0 \,+\, u_t, & t \le T_1^0 \\ (x_t', z_{1,t}')\beta_2^0 \,+\, u_t, & T_1^0 + 1 \le t < T_2^0 \\ (x_t', z_{1,t}')\beta_3^0 \,+\, u_t, & t > T_2^0 \end{cases} \tag{77}$$

$$x_t' \;=\; \begin{cases} z_t'\Delta_1^0 \,+\, v_t, & t \le T_1^0 \\ z_t'\Delta_2^0 \,+\, v_t, & t > T_1^0 \end{cases} \tag{78}$$

with $\pi_1^0 = \lambda_1^0$, thus $T_1^* = T_1^0$ in the notation of Section 3.2. For ease of notation, we set $\kappa = [T\pi_1]$, $k_i = [T\lambda_i]$, $k_i^0 = [T\lambda_i^0]$. Also let $\hat{\kappa}$ denote the estimator of k_1^0 from the reduced form, that is, $\hat{\kappa} = [T\hat{\pi}_1]$. As in the proof of Proposition 3, we have $\hat{\kappa} \in B^* = \{\kappa : \; |\kappa - k_1^0| \le C_1^*[s_T^*]^{-2}\}$, for some $C_1 > 0$, and from that proposition we also need only consider $\hat{k}_2 \in B_2 = \{k_2 : \; |k_2 - k_2^0| \le C_2 s_T^{-2}\}$ for some $C_2 > 0$. We now consider the properties of $\hat{k}_1 = [T\hat{\lambda}_1]$ where $\hat{\lambda}_1$ is defined in (17) with $\hat{\lambda}_{k-1} = 1$ and $\hat{\lambda}_{k+1} = \hat{k}_2$.

Proof of part (i): The basic proof strategy is the same as that Proposition 1 (i). By similar arguments to (28), we have

$$\sum_{t=1}^{\hat{k}_2} \tilde{u}_t d_t \;=\; \tilde{U}' P_{\bar{W}^*}(\bar{W}^* - \bar{W}^0)\beta^0 \;+\; \tilde{U}' P_{\bar{W}^*}\tilde{U} \;-\; \tilde{U}'(\bar{W}^* - \bar{W}^0)\beta^0 \tag{79}$$

where $\bar{W}^*$ is now a diagonal partition of W at $\hat{k}_1$, $W = [w_1, w_2, \ldots, w_{\hat{k}_2}]'$, $\bar{W}^0$ is now the diagonal partition of W at k_1^0, $\tilde{U} = [\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_{\hat{k}_2}]$.

We consider the terms in (79) in turn. First consider $\bar{W}^{*\prime}\bar{W}^*$. To this end, define $\hat{\delta}(t,T) = \Delta^0(t,T) - \hat{\Delta}(t,T)$, where $\Delta^0(t,T) = \Delta_1^0\mathcal{I}\{t \le k_1^0\} + \Delta_2^0\mathcal{I}\{t > k_1^0\}$, $\hat{\Delta}(t,T) = \hat{\Delta}_1\mathcal{I}\{t \le \hat{\kappa}\} + \hat{\Delta}_2\mathcal{I}\{t > \hat{\kappa}\}$, therefore

$$\hat{\delta}(t,T) \;=\; \begin{cases} \Delta_1^0 - \hat{\Delta}_1, & t \le \hat{\kappa} \wedge k_1^0 \\ \Delta_2^0 - \hat{\Delta}_2, & t > \hat{\kappa} \vee k_1^0 \\ (\Delta_1^0 - \hat{\Delta}_2)\mathcal{I}\{\hat{\kappa} < k_1^0\} \,+\, (\Delta_2^0 - \hat{\Delta}_1)\mathcal{I}\{\hat{\kappa} > k_1^0\}, & t \in B^* \end{cases} \tag{80}$$

Letting $\bar{B}^* = (B^*)^c$, the complement of B^* on $[1, \hat{k}_2]$, we then have: $\hat{\delta}(t,T) = O_p(T^{-1/2})$ for $t \in \bar{B}^*$; $\hat{\delta}(t,T) = O_p(s_T^*)$ for $t \in B^*$. It then follows that

$$\left\|\bar{W}^{*\prime}\bar{W}^*\right\| \;=\; \|\sum_{t=1}^{\hat{k}_2} w_t w_t'\| \;\le\; \|\hat{\Upsilon}(t,T)'\hat{\Upsilon}(t,T)\|\,\|\sum_{t=1}^{\hat{k}_2} z_t z_t'\| \;=\; O_p(T) \tag{81}$$

where $\hat{\Upsilon}(t,T) = [\hat{\Delta}(t,T), \Pi]$.

Now consider $\bar{W}^{*\prime}\tilde{U}$. We have

$$\|\bar{W}^{*\prime}\tilde{U}\| \;=\; \|\sum_{t\in B^*} w_t\tilde{u}_t \;+\; \sum_{t\in\bar{B}^*} w_t\tilde{u}_t\| \;\leq\; \|\sum_{t\in B^*} w_t\tilde{u}_t\| + \|\sum_{t\in\bar{B}^*} w_t\tilde{u}_t\| \tag{82}$$

Now,

$$\begin{aligned}\|\sum_{t\in B^*} w_t\tilde{u}_t\| &\leq \|\sum_{t\in B^*}\hat{\Upsilon}(t,T)'z_t[u_t + v_t'\beta_x^0(t,T)]\| + \|\sum_{t\in B^*}\hat{\Upsilon}(t,T)'z_t z_t'\hat{\delta}(t,T)\beta_x^0(t,T)\| \\ &\leq O_p([s_T^*]^{-1}) \;+\; O_p([s_T^*]^{-2})O_p(s_T^*)O_p(1) \;=\; O_p([s_T^*]^{-1}),\end{aligned}$$

and

$$\begin{aligned}\|\sum_{t\in\bar{B}^*} w_t\tilde{u}_t\| &\leq \|\sum_{t\in\bar{B}^*}\hat{\Upsilon}(t,T)'z_t[u_t + v_t'\beta_x^0(t,T)]\| + \|\sum_{t\in\bar{B}^*}\hat{\Upsilon}(t,T)'z_t z_t'\hat{\delta}(t,T)\beta_x^0(t,T)\| \\ &= O_p(T^{1/2}).\end{aligned}$$

Thus it follows from (82) that

$$\bar{W}^{*\prime}\tilde{U} \;=\; O_p(T^{1/2}) \;+\; O_p([s_T^*]^{-1}) \;=\; O_p(T^{1/2}). \tag{83}$$

For $\bar{W}^{*\prime}(\bar{W}^* - \bar{W}^0)\beta^0$, we have

$$\left\|\bar{W}^{*\prime}(\bar{W}^* - \bar{W}^0)\beta^0\right\| \;=\; \left\|\sum_{t=(\hat{k}_1\wedge k_1^0)+1}^{\hat{k}_1\vee k_1^0} w_t w_t'(\beta_2^0 - \beta_1^0)\right\| \;=\; O_p(Ts_T), \tag{84}$$

and for $\tilde{U}'(\bar{W}^* - \bar{W}^0)\beta^0$, we have

$$\left\|\tilde{U}'(\bar{W}^* - \bar{W}^0)\beta^0\right\| \;=\; \left\|\sum_{t=(\hat{k}_1\wedge k_1^0)+1}^{\hat{k}_1\vee k_1^0} \tilde{u}_t w_t'(\beta_2^0 - \beta_1^0)\right\| \;\leq\; O_p(T^{1/2}s_T). \tag{85}$$

Combining (79), (81) and (83)-(85), we obtain $\sum_{t=1}^{\hat{k}_2}\tilde{u}_t d_t = O_p(T^{1/2}s_T)$. The desired result then follows by similar arguments to the proof of Proposition 1 (i).

Proof of part (ii): The general proof strategy is similar to Proposition 1 (ii). Define $V_\epsilon(C) = \{k_1 : |k_1 - k_1^0| < \epsilon T, k_1^0 - k_1 > C_1 s_T^{-2}\}$[36], for some $C_1 > 0$, SSR_1 to be the residual sum of squares from 2SLS estimation of the structural equation based on sample $[1, \hat{k}_2]$ with a break at k_1, SSR_2 to be the residual sum of squares from 2SLS estimation of the structural equation

[36] The case $k_1 > k_1^0$ can be handled in a similar fashion.

based on sample $[1, \hat{k}_2]$ with a break at k_1^0, SSR_3 to be the residual sum of squares from 2SLS estimation of the structural equation based on sample $[1, \hat{k}_2]$ with breaks at k_1 and k_1^0. By similar arguments to the proof of Proposition 1 (ii), we have

$$\frac{SSR_1 - SSR_2}{k_1^0 - k_1} \geq N_1 - N_2 - N_3 \tag{86}$$

where

$$\begin{aligned}
N_1 &= (\hat{\beta}_2^* - \hat{\beta}_\Delta)' \left(\frac{W_\Delta' W_\Delta}{k_1^0 - k_1} \right) (\hat{\beta}_2^* - \hat{\beta}_\Delta) \\
N_2 &= (\hat{\beta}_2^* - \hat{\beta}_\Delta)' \left(\frac{W_\Delta' \bar{W}}{k_1^0 - k_1} \right) \left(\frac{\bar{W}' \bar{W}}{T} \right)^{-1} \left(\frac{\bar{W}' W_\Delta}{T} \right) (\hat{\beta}_2^* - \hat{\beta}_\Delta) \\
N_3 &= (\hat{\beta}_1^* - \hat{\beta}_\Delta)' \left(\frac{W_\Delta' W_\Delta}{k_1^0 - k_1} \right) (\hat{\beta}_1^* - \hat{\beta}_\Delta)
\end{aligned}$$

where $\hat{\beta}_1^*$ is the 2SLS estimator of the regression parameter based on $t = 1, 2, \ldots, \hat{k}_1$, $\hat{\beta}_\Delta$ is the 2SLS estimator of the regression parameter based on $t = k_1 + 1, \ldots, k_1^0$, $\hat{\beta}_2^*$ is the 2SLS estimator of the regression parameter based on $t = k_1^0 + 1, \ldots, \hat{k}_2$, $W_\Delta = [0_{p \times k_1}, w_{k_1+1}, \ldots, w_{k_1^0}, 0_{p \times (\hat{k}_2 - k_1^0)}]'$ and $\bar{W}$ is the diagonal partition of $W = [w_1, \ldots, w_{\hat{k}_2}]$ at k_1.

It is straightforward to show that N_1 dominates N_2 for small ϵ. Therefore, we focus on showing that N_1 dominates N_3 for small ϵ, large C and N_1 is positive with large probability. Since $\hat{\beta}_1^*$ and $\hat{\beta}_\Delta$ are sub-sample estimators of β_1^0, it follows by standard arguments that $\hat{\beta}_1^* = \beta_1^0 + O_p(T^{-1/2})$ and $\hat{\beta}_\Delta = \beta_1^0 + O_p(T^{-1/2})$ for C and T large. On the other hand, since

$$\hat{\beta}_2^* = \beta_2^0 + \left(\sum_{t=k_1^0+1}^{\hat{k}_2} w_t w_t' \right)^{-1} \sum_{t=k_1^0+1}^{\hat{k}_2} w_t \tilde{u}_t$$

it follows that $\hat{\beta}_2^* = \beta_2^0 + O_p(T^{-1/2})$. Using these results we obtain $\hat{\beta}_2^* - \hat{\beta}_\Delta = (\beta_2^0 - \beta_1^0) + O_p(T^{-1/2})$ and $\hat{\beta}_1^* - \hat{\beta}_\Delta = O_p(T^{-1/2})$. Since, for large C, we have $W_\Delta' W_\Delta / (k_1^0 - k_1) = O_p(1)$, it follows from the results above that $N_1 = O_p(s_T^2)$ and $N_3 = O_p(T^{-1})$. Therefore, N_1 dominates N_3. Finally for large C, $W_\Delta' W_\Delta / (k_1^0 - k_1)$ is p.d. and so $N_1 > 0$ with large probability. $\diamond$.

Proof of Theorem 4

Consider the model used above in the proof of Proposition 4. Define $\hat{\beta}_1$ to be the 2SLS estimator based on $t \in [1, k_1]$, $\hat{\beta}_2$ to be the 2SLS estimator based on $t \in [k_1 + 1, \hat{k}_2]$, $\tilde{\beta}_1$ to be the 2SLS estimator based on $t \in [1, k_1^0]$, and $\tilde{\beta}_2$ to be the 2SLS estimator based on $t \in [k_1^0 + 1, \hat{k}_2]$. To facilitate the proof we must first consider the properties of these estimators. Note that from Proposition 4 (ii) it follows that we need consider only $k_1 \in B_1 = \{k_1 : |k_1 - k_1^0| < C_1 s_T^{-2}\}$.

Consider first $\hat{\beta}_1$. We have

$$\begin{aligned}\hat{\beta}_1 &= \beta_1^0 + \left(\sum_{t=1}^{k_1} w_t w_t'\right)^{-1} \sum_{t=1}^{k_1} w_t \tilde{u}_t \\ &\quad + \left(\sum_{t=1}^{k_1} w_t w_t'\right)^{-1} \left(\sum_{t=(k_1 \wedge k_1^0)+1}^{k_1 \vee k_1^0} w_t w_t'\right) (\beta_2^0 - \beta_1^0)\mathcal{I}\{k_1 > k_1^0\} \\ &= \beta_1^0 + O_p(T^{-1/2}) + O_p(T^{-1}s_T^{-1}) = \beta_1^0 + O_p(T^{-1/2}) \text{ uniformly in } B_1. \qquad (87)\end{aligned}$$

Also we have

$$\begin{aligned}\hat{\beta}_2 &= \beta_2^0 + \left(\sum_{t=k_1+1}^{\hat{k}_2} w_t w_t'\right)^{-1} \sum_{t=k_1+1}^{\hat{k}_2} w_t \tilde{u}_t \\ &\quad + \left(\sum_{t=k_1+1}^{\hat{k}_2} w_t w_t'\right)^{-1} \left(\sum_{t=(k_1 \wedge k_1^0)+1}^{k_1 \vee k_1^0} w_t w_t'\right) (\beta_1^0 - \beta_2^0)\mathcal{I}\{k_1 < k_1^0\} \\ &\quad + \left(\sum_{t=k_1+1}^{\hat{k}_2} w_t w_t'\right)^{-1} \left(\sum_{t=(\hat{k}_2 \wedge k_2^0)+1}^{\hat{k}_2 \vee k_2^0} w_t w_t'\right) (\beta_3^0 - \beta_2^0)\mathcal{I}\{\hat{k}_2 > k_2^0\} \\ &= \beta_2^0 + O_p(T^{-1/2}) \text{ uniformly in } B_1. \qquad (88)\end{aligned}$$

Now consider $\tilde{\beta}_1$. We have

$$\tilde{\beta}_1 = \beta_1^0 + \left(\sum_{t=1}^{k_1^0} w_t w_t'\right)^{-1} \sum_{t=1}^{k_1^0} w_t \tilde{u}_t = \beta_1^0 + O_p(T^{-1/2}). \qquad (89)$$

For $\tilde{\beta}_2$, we have

$$\begin{aligned}\tilde{\beta}_2 &= \beta_2^0 + \left(\sum_{t=k_1+1}^{\hat{k}_2} w_t w_t'\right)^{-1} \sum_{t=k_1+1}^{\hat{k}_2} w_t \tilde{u}_t \\ &\quad + \left(\sum_{t=k_1}^{\hat{k}_2} w_t w_t'\right)^{-1} \left(\sum_{t=(\hat{k}_2 \wedge k_2^0)+1}^{\hat{k}_2 \vee k_2^0} w_t w_t'\right) (\beta_3^0 - \beta_2^0)\mathcal{I}\{\hat{k}_2 > k_2^0\} \\ &= \beta_2^0 + O_p(T^{-1/2}) \text{ uniformly in } B_1. \qquad (90)\end{aligned}$$

Now consider $\hat{\beta}_1 - \tilde{\beta}_1$. From the formulae above, it follows that

$$T^{1/2}(\hat{\beta}_1 - \tilde{\beta}_1) = \left(\sum_{t=1}^{k_1} w_t w_t'\right)^{-1} \sum_{t=1}^{k_1} w_t \tilde{u}_t - \left(\sum_{t=1}^{k_1^0} w_t w_t'\right)^{-1} \sum_{t=1}^{k_1^0} w_t \tilde{u}_t + o_p(1) \qquad (91)$$

After some manipulations, it follows from (91) that

$$\|T^{1/2}(\hat{\beta}_1 - \tilde{\beta}_1)\| = \left\| \left(\sum_{t=1}^{k_1} w_t w_t'\right)^{-1} \right\| \left\| \sum_{t=(k_1 \wedge k_1^0)+1}^{k_1 \vee k_1^0} w_t \tilde{u}_t \right\| + o_p(1). \qquad (92)$$

Now, $\left\|\sum_{t=(k_1\wedge k_1^0)+1}^{k_1\vee k_1^0} w_t\tilde{u}_t\right\|$ is the same order as $\left\|\sum_{t\in B_1} w_t\tilde{u}_t\right\|$ and

$$\left\|\sum_{t\in B_1} w_t\tilde{u}_t\right\| \;\leq\; \left\|\sum_{t\in B_1\cap B^*} w_t\tilde{u}_t\right\| + \left\|\sum_{t\in B_1\cap \bar{B}^*} w_t\tilde{u}_t\right\|. \tag{93}$$

Since

$$\begin{aligned}
\left\|\sum_{t\in B_1\cap B^*} w_t\tilde{u}_t\right\| &\leq \left\|\sum_{t\in B_1\cap B^*} \hat{\Upsilon}(t,T)'z_t(u_t+v_t'\beta_x^0(t,T)\right\| + \left\|\sum_{t\in B_1\cap B^*} \hat{\Upsilon}(t,T)'z_tz_t'\hat{\delta}(t,T)\beta_x^0(t,T)\right\| \\
&= O_p([s_T\vee s_T^*]^{-1}) \,+\, O_p([s_T\vee s_T^*]^{-2})O_p(s_T^*) \\
&= O_p([s_T\vee s_T^*]^{-1})\left\{O_p(1) \,+\, O_p\left(\frac{s_T^*}{s_T\vee s_T^*}\right)\right\} \\
&= O_p([s_T\vee s_T^*]^{-1}) \;=\; O_p(s_T^{-1}\wedge [s_T^*]^{-1}),
\end{aligned}$$

and

$$\begin{aligned}
\left\|\sum_{t\in B_1\cap \bar{B}^*} w_t\tilde{u}_t\right\| &\leq \left\|\sum_{t\in B_1\cap \bar{B}^*} \hat{\Upsilon}(t,T)'z_t[u_t+v_t'\beta_x^0(t,T)]\right\| + \left\|\sum_{t\in B_1\cap \bar{B}^*} \hat{\Upsilon}(t,T)'z_tz_t'\hat{\delta}(t,T)\beta_x^0(t,T)\right\| \\
&= O_p(s_T^{-1}) \,+\, O_p(s_T^{-2}T^{-1/2}) \;=\; O_p(s_T^{-1}),
\end{aligned}$$

it follows from (93) that $\|\sum_{t\in B_1} w_t\tilde{u}_t\| \leq O_p(s_T^{-1}\wedge[s_T^*]^{-1})+O_p(s_T^{-1}) = O_p(s_T^{-1})$ and hence from (92) we have $T^{1/2}(\hat{\beta}_1-\tilde{\beta}_1) = o_p(1)$. A similar argument can be used to show that $T^{1/2}(\hat{\beta}_2-\tilde{\beta}_2) = o_p(1)$.

With this background, we now consider the distribution of $\hat{k}_1$, where

$$\hat{k}_1 \;=\; argmin_{k_1\in B_1}[SSR(k_1,\hat{k}_2) - SSR(k_1^0,\hat{k}_2)]$$

It is easily established[37] that

$$SSR(k_1,\hat{k}_2) - SSR(k_1^0,\hat{k}_2) \;=\; \sum_{t=1}^{\hat{k}_2} a_t \,+\, 2\sum_{t=1}^{\hat{k}_2} c_t \;=\; A+2C \tag{94}$$

where a_t and c_t are defined as below (74) in the proof of Theorem 3 but with $\hat{\beta}(t,T) = \hat{\beta}_1\mathcal{I}\{t\leq k_1\} + \hat{\beta}_2\mathcal{I}\{t>k_1\}$, $\tilde{\beta}(t,T) = \tilde{\beta}_1\mathcal{I}\{t\leq k_1^0\} + \tilde{\beta}_2\mathcal{I}\{t>k_1^0\}$. Define $I_2 = [1,\hat{k}_2]$ and $\bar{B}_1 = I_2\setminus B_1$. For A, we have $\sum_{t=1}^{\hat{k}_2} a_t = \sum_{t\in B_1} a_t + \sum_{t\in\bar{B}_1} a_t$ and $\sum_{t\in\bar{B}_1} a_t = O_p(T^{-1}s_T^{-2})O_p(1)O_p(1) = o_p(1)$ and

$$\sum_{t\in B_1} a_t \;=\; |k_1 \,-\, k_1^0|\theta_{T,1}^{0\prime}\left(\Upsilon_1^{0\prime}Q_1\Upsilon_1^0\mathcal{I}\{k_1<k_1^0\} \,+\, \Upsilon_2^{0\prime}Q_2\Upsilon_2^0\mathcal{I}\{k_1>k_1^0\}\right)\theta_{T,1} \,+\, o_p(1)$$

[37] By a similar argument to the derivation of (74).

Therefore, we obtain

$$A = |k_1 - k_1^0|\theta_{T,1}^{0\prime}\left(\Upsilon_1^{0\prime}Q_1\Upsilon_1^0\mathcal{I}\{k_1 < k_1^0\} + \Upsilon_2^{0\prime}Q_2\Upsilon_2^0\mathcal{I}\{k_1 > k_1^0\}\right)\theta_{T,1} + o_p(1) \text{ uniformly in} B_1. \tag{95}$$

Now consider C. We have $\sum_{\bar{B}_1} c_t = o_p(1)$, uniformly in B_1 and

$$\sum_{B_1} c_t = [(-1)^{\mathcal{I}\{k_1<k_1^0\}}\theta_{T,1}^{0\prime}]\left\{\left[\Upsilon_1^{0\prime}\mathcal{I}\{k_1 < k_1^0\} + \Upsilon_2^{0\prime}\mathcal{I}\{k_1 > k_1^0\}\right] T^{-1/2}\sum_{t=(k_1\wedge k_1^0)+1}^{k_1\vee k_1^0} z_t[u_t + v_t'\beta_x^0(t,T)]\right\} + o_p(1), \text{ uniformly in } B_1.$$

It follows that

$$\begin{aligned} A + 2C &= |k_1 - k_1^0|\theta_{T,1}^{0\prime}\left(\Upsilon_1^{0\prime}Q_1\Upsilon_1^0\mathcal{I}\{k_1 < k_1^0\} + \Upsilon_2^{0\prime}Q_2\Upsilon_2^0\mathcal{I}\{k_1 > k_1^0\}\right)\theta_{T,1} \\ &\quad +2[(-1)^{\mathcal{I}\{k_1<k_1^0\}}\theta_{T,1}^{0\prime}]\left\{\left[\Upsilon_1^{0\prime}\mathcal{I}\{k_1 < k_1^0\} + \Upsilon_2^{0\prime}\mathcal{I}\{k_1 > k_1^0\}\right] T^{-1/2}\sum_{t=(k_1\wedge k_1^0)+1}^{k_1\vee k_1^0} z_t[u_t + v_t'\beta_x^0(t,T)]\right\} \\ &\quad + o_p(1), \qquad \text{uniformly in } B_1. \end{aligned} \tag{96}$$

It can be recognized that (96) has the same basic structure as (8) and so the rest of the proof follows by similar arguments to the proof of Proposition 2. ⋄.

Table 1: Empirical coverage of break point confidence intervals

Case I, one break model with $(\beta_1^0; \beta_2^0)$*=(c ,0.1; -c,-0.1)*

$q-1$	T	Confidence Interval								
		$c=0.3$			$c=0.5$			$c=1$		
		99 %	95 %	90 %	99 %	95 %	90 %	99 %	95 %	90 %
2	60	.90	.82	.75	.95	.90	.86	.99	.97	.96
	120	.95	.89	.85	.97	.93	.89	.99	.97	.96
	240	.97	.92	.87	.98	.95	.92	1.00	.98	.97
	480	.99	.94	.89	.99	.97	.92	1.00	.99	.98
4	60	.90	.80	.74	94	.88	.83	.99	.98	.96
	120	.93	.86	.80	.97	.93	.90	1.00	.98	.97
	240	.96	.92	.87	.99	.93	.90	1.00	.98	.98
	480	.98	.94	.90	.99	.95	.91	1.00	.99	.98
8	60	.91	.80	.74	.94	.89	.85	.99	.97	.96
	120	.94	.86	.81	.97	.93	.88	.99	.98	.96
	240	.97	.90	.86	.98	.95	.91	.99	.98	.96
	480	.98	.93	.89	.99	.96	.92	.99	.98	.96

Notes: Here $q-1$ is the number of instruments (excluding the intercept), and the column headed $100a\%$ gives the percentage of times (in 1000 simulations) the $100a\%$ confidence intervals for the break points contain the corresponding true values.

Table 2: Empirical coverage of break point confidence intervals

Case II, two break model with $(\beta_1^0; \beta_2^0, \beta_3^0)$*=(c,0.1; -c,-0.1; c,0.1)*

$q-1$	T	Confidence Interval																	
		$c = 0.3$						$c = 0.5$						$c = 1$					
		1^{st} break			2^{nd} break			1^{st} break			2^{nd} break			1^{st} break			2^{nd} break		
		99 %	95 %	90 %	99 %	95 %	90 %	99 %	95 %	90 %	99 %	95 %	90 %	99 %	95 %	90 %	99 %	95 %	90 %
2	60	.91	.75	.66	.93	.81	.71	.94	.86	.79	.94	.87	.84	.98	.95	.94	.98	.96	.94
	120	.94	.82	.76	.95	.86	.78	.96	.91	.89	.97	.92	.88	.99	.98	.96	.99	.98	.97
	240	.97	.88	.81	.97	.92	.86	.98	.95	.91	.98	.94	.90	1.00	.98	.97	1.00	.99	.98
	480	.98	.94	.88	.98	.93	.88	.99	.95	.92	.99	.96	.92	1.00	.98	.97	.99	.98	.97
4	60	.92	.76	.68	.90	.78	.70	.94	.85	.78	.94	.87	.82	.99	.96	.94	.99	.96	.94
	120	.94	.84	.76	.94	.86	.78	.97	.91	.86	.98	.92	.87	.99	.97	.96	.99	.97	.96
	240	.95	.87	.82	.97	.88	.82	.98	.94	.90	.99	.94	.89	.99	.97	.96	1.00	.99	.98
	480	.98	.93	.88	.98	.93	.88	.99	.96	.92	.99	.95	.91	1.00	.98	.96	.99	.97	.96
8	60	.92	.78	.70	.90	.79	.70	.95	.85	.78	.95	.88	.82	.99	.96	.95	.99	.96	.93
	120	.95	.83	.75	.94	.84	.76	.97	.90	.86	.97	.91	.86	1.00	.98	.96	.98	.97	.96
	240	.96	.88	.81	.97	.88	.83	.98	.93	.89	.98	.94	.89	1.00	.98	.96	1.00	.98	.96
	480	.97	.92	.86	.98	.92	.88	.99	.95	.92	.99	.97	.94	1.00	.98	.98	.99	.98	.97

Notes: For definitions see Table 1.

Table 3: Empirical coverage of break point confidence intervals

Case III, one break model with $(\beta_1^0; \beta_2^0)$*=(c ,0.1; -c,-0.1)*

$q-1$	T	Confidence Interval								
		$c=0.3$			$c=0.5$			$c=1$		
		99 %	95 %	90 %	99 %	95 %	90 %	99 %	95 %	90 %
2	120	.89	.80	.73	.95	.88	.83	.98	.95	.92
	240	.93	.86	.82	.95	.90	.85	.98	.93	.91
	480	.97	.90	.85	.98	.92	.86	.99	.96	.93
4	120	.89	.80	.74	.94	.88	.83	.98	.94	.91
	240	.92	.86	.80	.97	.91	.87	.98	.96	.93
	480	.97	.91	.86	.98	.93	.88	.99	.97	.94
8	120	.89	.80	.73	.94	.86	.82	.97	.92	.90
	240	.94	.89	.82	.97	.93	.88	.99	.96	.93
	480	.98	.93	.88	.98	.92	.87	.99	.97	.95

Notes: For definitions see Table 1.

Table 4: Empirical coverage of break point confidence intervals

Case IV, one break model with $(\beta_1^0;\beta_2^0)$*=(c ,0.1; -c,-0.1)*

$q-1$	T	Confidence Interval								
		$c=0.3$			$c=0.5$			$c=1$		
		99 %	95 %	90 %	99 %	95 %	90 %	99 %	95 %	90 %
2	120	.93	.86	.82	.95	.91	.89	.99	.98	.98
	240	.96	.85	.81	.96	.93	.90	1.00	1.00	1.00
	480	.94	.88	.85	.99	.97	.95	1.00	1.00	1.00
4	120	.93	.87	.84	.95	.92	.90	.99	.99	.99
	240	.94	.88	.85	.98	.96	.94	1.00	1.00	1.00
	480	.97	.93	.90	.99	.99	.97	1.00	1.00	1.00
8	120	.93	.88	.82	.95	.92	.89	1.00	.99	.99
	240	.95	.90	.86	.99	.97	.95	1.00	1.00	.99
	480	.97	.94	.91	1.00	.98	.96	1.00	1.00	.99

Notes: For definitions see Table 1.

Table 5: NKPC - stability statistics for structural equation

k	$q\times$ sup-F	F(k+1:k)	sup-Wald	Wald(k+1:k)	BIC
0	-	-	-	-	-0.092
1	15.02	12.06	17.02	8.32	0.066
2	13.78	10.22	12.50	11.07	0.247
3	16.09	9.72	20.29	12.95	0.354

Notes: $q\times$ sup-F and sup-Wald denote the statistics for testing $H_0 : m = 0$ vs. $H_1 : m = k$, the first statistic being multiplied by q; F(k+1:k) and Wald(k+1:k) are the statistics for testing $H_0 : m = k$ vs. $H_1 : m = k + 1$; BIC is the BIC criterion; see Hall, Han, and Boldea (2009) for further details. The percentiles for the statistics are for $k = 1, 2, \ldots$ respectively: (i) $q\times$ sup-F and sup-Wald: (10%, 1%) significance level = (19.70, 26.71), (17.67, 21.87), (16.04, 19.42), (14.55, 17.44), (12.59,15.02); (ii) F(k+1:k) and Wald(k+1:k): (10%, 1%) significance level =(21.79, 28.36), (22.87, 29.30), (24.06,29.86), (24.68, 30.52).

References

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.

Andrews, D. W. K., and Ploberger, W. (1994). 'Optimal tests when a nuisance parameter is present only under the alternative', *Econometrica*, 62: 1383–1414.

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

——— (1997a). 'Estimating multiple breaks one at a time', *Econometric Theory*, 13: 315–352.

——— (1997b). 'Estimation of a change point in multiple regression models', *Review of Economics and Statistics*, 79: 551–563.

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

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.

Christ, C. F. (1994). 'The Cowles Commission's contributions to econometrics at Chicago, 1939-1955', *Journal of Economic Literature*, 32: 30–59.

Craig, C. C. (1936). 'On the frequency function of xy', *Annals of Mathematical Statistics*, 7: 1–15.

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

Gali, J., and Gertler, M. (1999). 'Inflation dynamics: a structural econometric analysis', *Journal of Monetary Economics*, 44: 195–222.

Ghysels, E., and Hall, A. R. (1990a). 'Are consumption based intertemporal asset pricing models structural?', *Journal of Econometrics*, 45: 121–139.

——— (1990b). '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.

——— (2009). 'Inference Regarding Multiple Structural Changes in Linear Models with Endogenous Regressors', Discussion paper, Economics, School of Social Studies, University of Manchester, Manchester, UK.

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.

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.

Hinckley, D. (1970). 'Inference about the change points in a sequence of random variables', *Biometrika*, 57: 1–17.

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.

Rudd, J., and Whelan, K. (2005). 'Does labor's share drive inflation?', *Journal of Money, Credit and Banking*, 37: 297–312.

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

Wooldridge, J., and White, H. (1988). 'Some invariance principles and central limt theorems for dependent heterogeneous processes', *Econometric Theory*, 4: 210–230.

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

Zhang, C., Osborn, D., and Kim, D. (2008). 'The new Keynesian Phillips curve: from sticky inflation to sticky prices', *Journal of Money, Credit and Banking*, 40: 667–699.