Discussion Paper Series

Inference Regarding Multiple Structural Changes in Linear Models with Endogenous Regressors

By

Alastair R. Hall, Sanggohn Han, Otilia Boldea

Centre for Growth and Business Cycle Research, Economic Studies, University of Manchester, Manchester, M13 9PL, UK

October 2009
Number 125

Inference Regarding Multiple Structural Changes in Linear Models with Endogenous Regressors[1]

Alastair R. Hall

University of Manchester[2]

Sanggohn Han

SAS Institute

Otilia Boldea

Tilburg University

September 10, 2009

[1]We are grateful to Denise Osborn and Eric Renault for valuable comments and to Nikolaos Sakkas for assistance with the computations. The paper has also benefited from the comments of a co-editor and three anonymous referees. An earlier version of this paper was circulated under the title "Inference regarding multiple structural changes in linear models estimated via 2SLS" which was presented at the World Congress of the Econometric Society, London, August 19-24, 2005, the Triangle Econometrics Conference, RTP, NC, December 2, 2005, CIREQ Conference on Generalized Method of Moments, Montreal, November 16-17, 2007, the ESRC Econometrics group seminar at the Institute of Fiscal Studies, London, the London-Oxbridge Time Series Workshop and at seminars at Erasmus University and the Universities of Birmingham and Warwick. We are also very grateful to Chengsi Zhang and Denise Osborn for providing us with the data used in the empirical example. The first 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

This paper considers estimation and inference within a linear model with endogenous regressors and multiple changes in the parameters at unknown times. It is shown that estimation via a Generalized Method of Moments criterion yields inconsistent estimators of the break fractions under reasonable conditions. In contrast, minimization of the Two Stage Least Squares (2SLS) minimand is shown to yield consistent estimators of the break fractions. We further establish the consistency and asymptotic normality of the 2SLS parameter estimators in this model. We propose and derive the limiting distributions of various tests for structural change, and also propose a method for estimating the number of breaks based on these tests. The analysis covers the cases where the reduced form is either stable or unstable. Simulation evidence validates our methodology in finite samples. The methodology is illustrated via an application to the New Keynesian Phillips curve for the US.

JEL classification: C12, C13

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

1 Introduction

While it is routine to assume in estimation that the parameters of econometric models are constant over time, there are reasons why this assumption may be questionable. In particular, it can be argued that policy changes and/or exogenous shifts may cause realignments in the relationship between economic variables which are reflected in changes in the parameters. Therefore, it is important to develop methods for both detecting parameter instability and also for building models that incorporate this behaviour.

Considerable attention has focused on developing tests for structural instability within the IV or more generally within the Generalized Method of Moments (GMM) framework.[1] The majority of this literature has focused on the design of tests against the alternative of one structural break. Although these tests are also shown to have non-trivial power against other alternatives, it is clearly desirable to develop procedures that can discriminate between various forms of instability, including multiple unknown breaks. An important step in this direction is taken by Bai and Perron (1998).[2] Their analysis is in the context of linear regression models estimated via Ordinary Least Squares (OLS). Within their framework, the break points are estimated simultaneously with the regression parameters via minimization of the residual sum of squares. Bai and Perron (1998) establish the consistency and the limiting distribution of the resulting break point fractions. They also propose a sequential procedure for selecting the number of break points in the sample based on various F-statistics for parameter constancy.

While not the only possible form for structural instability, the model with discrete shifts at multiple unknown break points has some appeal in macroeconometric applications because it captures the case where relationships change due to changes in policy regime or exogenous shifts. However, since Bai and Perron's (1998) analysis is predicated on the assumption that all explanatory variables are exogenous, their methods can not be applied to macroeconometric models where the regressors are correlated with the errors.[3]

In this paper, we consider the extension of Bai and Perron's (1998) framework to linear

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

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

[3]A similar comment applies to the recent extensions of Bai and Perron's (1998) framework by Perron and Qu (2006) and Qu and Perron (2007).

models with endogenous regressors estimated via IV. There are two common approaches to IV estimation in econometrics: GMM and Two Stage Least Squares (2SLS). We begin by exploring the properties of break point and parameter estimators obtained by minimizing a GMM criterion. In the context of a one break model, we show that the GMM estimator of the break fraction (that indexes the break point) is inconsistent in general and provide a set of conditions under which the GMM break fraction estimator has a non-degenerate limiting distribution. Inspection of the proofs indicates that this behaviour stems from construction of the minimand as the square of a sums. This structure allows the opportunity for the effects of the misspecification associated with the selection of the wrong break point to offset in the minimand and confound the estimation. In contrast to GMM, the 2SLS minimand is a sum of squares and thus of a more promising construction. We therefore consider the case in which the break points are estimated simultaneously with the regression parameters via minimization of the residual sum of squares on the second step of the 2SLS estimation. To employ this strategy, it is necessary in the first stage regression to estimate the reduced form for the endogenous regressors in the structural equation of interest and this, of course, requires an assumption about the constancy or lack thereof of these reduced form parameters. In this paper, we consider two scenarios of interest, namely: (i) the parameters in the first stage regression are constant; (ii) the parameters in the first stage regression are subject to discrete shifts within the sample period and the locations of these shifts are estimated *a priori* via a data-based method that satisfies certain conditions. The latter conditions allow the case in which the location of the instability is estimated via an application of Bai and Perron's (1998) methods to each reduced form equation. Under both scenarios for the reduced form, we establish the consistency of the resulting break fractions estimators and both the consistency and asymptotic normality of the parameter estimators of the equation of interest. However, it turns out that the behaviour of the reduced form impacts on the limiting behaviour of test statistics for parameter change. In the case where the reduced form is stable, we show that the various F-statistics and Wald statistics for testing parameter constancy based on the 2SLS estimator have the same limiting distribution as the analogous statistics for OLS considered by Bai and Perron (1998). However, the corresponding results do not hold if the reduced form is unstable. This failure stems from the limiting behaviour of certain sample moments and is similar to that highlighted by Hansen (2000) in his analysis of the sup-F test (Andrews, 1993) when there are changes in the marginal distribution of the regressors.

Nevertheless, we are able to propose a simple methodology for estimating the number of breaks in both scenarios described above.

To illustrate our methods, we consider the stability of the New Keynesian Phillips curve (NKPC) estimated using quarterly data for the US over the period 1968.3-2001.4. The NKPC is of considerable theoretical importance in monetary policy analysis as it is used to identify the forward-looking components of inflation, as well as the trade-off between inflation and unemployment over the cycle. Zhang, Osborn, and Kim (2008) observe that empirical studies of the NKPC often reach conflicting conclusions about the importance of key variables in the determination of inflation, and argue this may be due to neglected parameter variation. Zhang, Osborn, and Kim (2008) argue that changes in monetary policy regimes may cause changes in the parameters of the NKPC; if true, this would mean that the parameters of the NKPC would exhibit discrete shifts at potentially multiple points in the sample. Zhang, Osborn, and Kim (2008) investigate this issue using a methodology based on uncovering break points in the sample via the maximization of Wald statistics for parameter change associated with 2SLS estimation. However, while their methodology has an intuitive appeal, there is no theoretical justification for their methods. In contrast, our methods can be applied to this model under plausible assumptions about the data. Our analysis indicates that there are shifts in the parameters of both the appropriate reduced forms and also in the NKPC itself.

The outline of the paper is as follows. Section 2 considers estimation based on a GMM minimand. Section 3 lays out the basic structure of the 2SLS estimation of the break point and parameter estimators. Section 4 establishes the properties of the estimators and various tests of parameter change when the reduced form is stable, describes an algorithm for estimation of the number of breaks and also validates our procedures in finite samples via simulations. Section 5 establishes the properties of the estimators when the reduced form is unstable, and proposes a methodology for estimating the number of breaks, partly exploiting the results for the stable reduced form. The finite sample performance of these methods is also evaluated using a small simulation study. Section 6 illustrates our methodology in the context of NKPC estimation for the US. Section 7 concludes. The mathematical appendix contains sketch proofs of the results in the paper; more detailed proofs are relegated to a supplemental appendix that is available from the authors upon request.

2 Inference based on the GMM minimand

Consider the following linear model with one break

$$y_t \;=\; x_t'\theta_0^{(i)} \;+\; u_t,\; t = 1, 2 \ldots T \tag{1}$$

where $\theta_0^{(i)} \;=\; \theta_0^{(1)}$ for $t/T \leq \lambda^0$ and $\theta_0^{(i)} = \theta_0^{(2)}$ for $t/T > \lambda^0$, $\lambda^0 \in (0,1)$ and $\theta_0^{(1)} \neq \theta_0^{(2)}$. Let x_t and $\theta_0^{(i)}$ be $p \times 1$. We assume that there exists a $q \times 1$ vector of variables, z_t, that are used as instruments for x_t, where $q > p$. Define $v_t = (x_t', u_t, z_t')'$.

For ease of presentation in this section, we assume that $\{v_t\}$ is an independent sequence but in line with the model in (1), we allow the data generation process for v_t to change (potentially) at $[T\lambda^0]$. These restrictions are embodied in the following assumption.

Assumption 1 *(i) v_t is independently distributed; (ii) $E[z_t x_t'] = M_1$, $t/T \leq \lambda^0$, $E[z_t x_t'] = M_2$, $t/T > \lambda^0$, rank $M_i = p$, $i = 1, 2$ (iii) $E[z_t u_t] = 0$, (iv) $\sup_t E\|v_t\|^4 < \infty$.*

For convenience of notation, we define the matrices:

$$\begin{aligned} N_1(\lambda) \;&=\; min(\lambda,\, \lambda_0)M_1 \;+\; max(\lambda - \lambda_0,\, 0)M_2 \\ N_2(\lambda) \;&=\; max(\lambda_0 - \lambda,\, 0)M_1 \;+\; min(1 - \lambda,\, 1 - \lambda_0)M_2 \end{aligned}$$

If the researcher knows there is a break but is unaware of its location, a natural approach is to estimate the location by minimizing the GMM criterion over all candidate partitions. Following Andrews (1993), GMM estimation of $\theta(\lambda)$ for each candidate break fraction, λ, is based on $E[f(v_t, \theta(\lambda); \lambda)] = 0$ where

$$f(v_t, \theta(\lambda); \lambda) \;=\; \begin{bmatrix} z_t\,\{y_t \;-\; x_t'\theta_1(\lambda)\}\,\mathcal{I}_{t,T}(\lambda) \\ z_t\,\{y_t \;-\; x_t'\theta_2(\lambda)\}\,\{1 \;-\; \mathcal{I}_{t,T}(\lambda)\} \end{bmatrix} \tag{2}$$

where $\theta(\lambda) = (\theta_1(\lambda)', \theta_2(\lambda)')'$, $\theta_i(\lambda) \in \Theta \subset \Re^p$ and $\mathcal{I}_{t,T}(\lambda)$ is an indicator variable that takes the value one if $t/T \;\leq\; \lambda$ and the value zero otherwise. The partial-sum GMM estimators of $[\theta_1(\lambda)', \theta_2(\lambda)']'$ are defined as follows:

$$\hat{\theta}_T(\lambda) \;=\; argmin_{\theta(\lambda)\in\Theta\times\Theta} Q_T(\theta(\lambda); \lambda) \tag{3}$$

where $\hat{\theta}_T(\lambda) = vec[\hat{\theta}_{1,T}(\lambda), \hat{\theta}_{2,T}(\lambda)]$, $Q_T(\theta(\lambda); \lambda) = g_T(\theta(\lambda); \lambda)' W_T(\lambda) g_T(\theta(\lambda); \lambda)$, $g_T(\theta(\lambda); \lambda) = T^{-1}\sum_{t=1}^{T} f(v_t, \theta(\lambda); \lambda)$, $W_T(\lambda) = diag\{W_{1,T}(\lambda), W_{2,T}(\lambda)\}$ and $W_{i,T}(\lambda)$ is a $q \times q$ deterministic matrix. We assume $W_{i,T}(\lambda)$ does not depend on $\theta(\lambda)$ but may depend on T. Thus we are considering a "first-step" GMM estimation in which the weighting matrix is a matrix of constants. The advantage of this restriction is that it considerably simplifies the analysis.[4]

Given a set of GMM estimations over $\lambda \in \Lambda \in (0,1)$, the break point estimator is:

$$\hat{\lambda}_T \;=\; argmin_{\lambda\in\Lambda} argmin_{\theta(\lambda)\in\Theta\times\Theta} Q_T(\theta(\lambda);\lambda) \tag{4}$$

This section shows that $\hat{\lambda}_T$ is not consistent for λ_0 under reasonable conditions. To establish this result, we introduce the following assumptions.

Assumption 2 *$E[f(v_t, \theta(\lambda^0); \lambda^0)] = 0$ for $\theta_0(\lambda^0) = \left(\theta_0^{(1)\prime}, \theta_0^{(2)\prime}\right)'$.*

Assumption 3 *Set $b_t = [z_t u_t, vec\{z_t x_t' - \bar{M}(\lambda)\}]$ for $\bar{M}(\lambda) = \mathcal{I}_{t,T}(\lambda) M_1 + (1 - \mathcal{I}_{t,T}(\lambda)) M_2$. Define $T^{-1/2}\sum_{t=1}^{[Tr]} b_t \Rightarrow \Omega^{1/2} B_m(r)$ where $B_m(r)$ is a $m \times 1$ vector of standard Brownian motions, $m = (p+1)q$ and $\Omega = \Omega^{1/2}\Omega^{1/2\prime}$ is a positive definite (pd) finite matrix.*

Assumption 4 *The minimum eigenvalues of $N_i(\lambda)' N_i(\lambda)$, $i = 1, 2$, are bounded away from zero uniformly in $\lambda \in \Lambda$.*

Assumption 5 *$W_{i,T}(\lambda)$ is a deterministic, positive semi-definite matrix that converges to $W_i(\lambda)$, a positive definite matrix, for all λ and $i = 1, 2$.*

Assumption 2 states that the population moment condition is valid at the true parameter values and at the true break. Assumption 3 states the convergence results needed to underpin the analysis. Assumption 4 ensures that the partial sum GMM estimators defined below are identified; notice it implies M_1 and M_2 are full rank.

Our first result involves the population analog to the GMM minimand. Define $\tilde{Q}_T(\theta(\lambda);\lambda) = E\left[Q_T(\theta(\lambda);\lambda)\right]$ and its limit as $\lim_{T\to\infty} \tilde{Q}_T(\theta(\lambda);\lambda) = \tilde{Q}(\theta(\lambda);\lambda)$.

[4]We note that a similar analysis based on the second-step GMM minimand needs to consider the properties of the long variance matrix estimator employed. Such an analysis is complicated by the issue of centering; for example see Hall, Inoue, and Peixe (2003) for an analysis of the impact of centering in covariance matrix estimation on the overidentifying restrictions test in the presence of structural instability. We anticipate similar problems arise here. Furthermore, it seems reasonable to anticipate that such an analysis of the second-step GMM minimand requires the type of analysis of the first-step estimator presented here, and that if the first-step estimation fails to identify the true break then this will undermine estimation of the break fraction on the second step.

Proposition 1 *If equation (1) and Assumptions 1, 2, 4 and 5 hold then:* $\tilde{Q}(\theta_*(\lambda);\lambda) = 0$ *in the following cases*

(i) $\lambda = \lambda^0$*:* $\theta_*(\lambda) = \theta(\lambda^0) = (\theta_0^{(1)'}, \theta_0^{(2)'})'$*;*

(ii) $\lambda < \lambda^0$*:* $\theta_0^{(1)} - \theta_0^{(2)} \in \mathcal{N}(M_1 - M_2)$*,* $\theta_*(\lambda) = \left[\theta_0^{(1)'}, \theta_*^{(2)}(\lambda)'\right]'$ *where*

$$\theta_*^{(2)}(\lambda) \;=\; \frac{(\lambda^0 \,-\, \lambda)\theta_0^{(1)} \,+\, (1 \,-\, \lambda^0)\theta_0^{(2)}}{1 \,-\, \lambda}$$

and $\mathcal{N}(A)$ *denotes the nullspace of a matrix* A*;*

(iii) $\lambda > \lambda^0$*:* $\theta_0^{(1)} - \theta_0^{(2)} \in \mathcal{N}(M_1 - M_2)$*,* $\theta_*(\lambda) = \left[\theta_*^{(1)}(\lambda)', \theta_0^{(2)'}\right]'$ *where*

$$\theta_*^{(1)}(\lambda) \;=\; \frac{\lambda^0\theta_0^{(1)} \,+\, (\lambda \,-\, \lambda^0)\theta_0^{(2)}}{\lambda}$$

Remark 1: Proposition 1 indicates that under the condition $\theta_0^{(1)} - \theta_0^{(2)} \in \mathcal{N}(M_1 - M_2)$ there is a value of the parameters that sets the population analog to the GMM minimand equal to zero for every choice of λ. Notice this value of θ depends on λ. Thus, the population analog of the GMM minimand does not have a unique minimum in $\theta(\lambda)$ for $\lambda \in (0, 1)$.
Remark 2: One case in which the condition $\theta_0^{(1)} - \theta_0^{(2)} \in \mathcal{N}(M_1 - M_2)$ is trivially satisfied is where $M_1 = M_2$, and thus $E[x_t z_t']$ remains constant throughout the sample. Notice however, that this moment constancy is sufficient but not necessary for the condition to hold.

Given Proposition 1, we have the following result.

Proposition 2 *If Assumptions 1-5 hold and* $\theta_0^{(1)} - \theta_0^{(2)} \in \mathcal{N}(M_1 - M_2)$ *then* $\hat{\theta}_T(\lambda) \xrightarrow{p} \theta_*(\lambda)$ *uniformly in* λ *where* $\theta_*(\lambda)$ *is defined in Proposition 1.*

The next proposition presents the limiting properties of the break fraction estimator under the conditions on the true parameters in Proposition 2.

Proposition 3 *If Assumptions 1-5 hold and* $\theta_0^{(1)} - \theta_0^{(2)} \in \mathcal{N}(M_1 - M_2)$

$$\hat{\lambda}_T \;\Rightarrow\; argmin_{\lambda \in \Lambda}\{Q_1(\lambda;\, \lambda_0) \,+\, Q_2(\lambda;\, \lambda_0)\}$$

where $Q_i(\lambda,\lambda^0) = \xi_i(\lambda)'\Xi_i(\lambda)\xi_i(\lambda)$, $\Xi_i(\lambda) = [I_q - N_i(\lambda)H_i(\lambda)]'W_i(\lambda)[I_q - N_i(\lambda)H_i(\lambda)]$, $H_i(\lambda) = [N_i(\lambda)'W_i(\lambda)N_i(\lambda)]^{-1}N_i(\lambda)'W_i(\lambda)$,

$$\begin{aligned}
\xi_1(\lambda) &= V_{zu}(\lambda) + \{1 - \mathcal{I}_\lambda(\lambda^0)\}\left\{[(\theta_0^{(1)} - \theta_0^{(2)})' \otimes I_q]\left[\frac{(\lambda-\lambda^0)}{\lambda}V_\mu(\lambda^0) - \frac{\lambda^0}{\lambda}[V_\mu(\lambda) - V_\mu(\lambda^0)]\right]\right\}, \\
\xi_2(\lambda) &= V_{zu}(1) - V_{zu}(\lambda) + \{\mathcal{I}_\lambda(\lambda^0)\}\left\{[(\theta_0^{(1)} - \theta_0^{(2)})' \otimes I_q]\left[\frac{(1-\lambda^0)}{(1-\lambda)}[V_\mu(\lambda^0) - V_\mu(\lambda)]\right.\right. \\
&\quad \left.\left. - \frac{(\lambda^0-\lambda)}{(1-\lambda)}[V_\mu(1) - V_\mu(\lambda^0)]\right]\right\},
\end{aligned}$$

$\mathcal{I}_\lambda(\lambda^0)$ *is an indicator variable that takes the value one if* $\lambda \leq \lambda^0$ *and zero otherwise, and* $[V_{zu}(\lambda)', V_\mu(\lambda)']' = \Omega^{1/2}B_m(\lambda)$ *with* $V_{zu}(\lambda)$ *of dimension* $q \times 1$.

Remark 3: Proposition 3 indicates that $\hat{\lambda}_T$ converges to a non-generate random variable and is thus not consistent for λ^0 under the conditions of the proposition.

Remark 4: While we focus on the one break model, the inconsistency result generalizes to the multiple break model under certain conditions. For example, if two adjacent regimes satisfy the conditions of our one break model.

To illustrate the nature of the limiting distribution in Proposition 3, we simulate the behaviour of $\hat{\lambda}_T$ in the following model.

One break model: The data generating process for the structural equation is:

$$\begin{aligned}
y_t &= [1, x_t]'\beta_1^0 + u_t, \qquad \text{for } t = 1, \ldots, [T/2] \\
&= [1, x_t]'\beta_2^0 + u_t, \qquad \text{for } t = [T/2]+1, \ldots, T
\end{aligned} \tag{5}$$

The reduced form equation for the scalar variable x_t is:

$$x_t = z_t'\delta + v_t, \qquad \text{for } t = 1, \ldots, T \tag{6}$$

where δ is $q \times 1$. The errors are generated as follows: $(u_t, v_t)' \sim IN(0_{2\times1}, \Omega)$ 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: $z_t \sim i.i.d\ N(0_{q\times1}, I_q)$. The specific parameter values are as follows: (i) $T = 480$; (ii) $(\beta_1^0, \beta_2^0) = ([1, 0.1]', [-1, -0.1]')$; (iii) $q = 4$; (iv) δ is chosen to yield the population $R^2 = 0.5$ for the regression in (6).[5] 1000 simulations are performed.

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

Figure 1 contains a plot of the empirical distribution of $\hat{\lambda}_T$ when $\Lambda = [0.15, 0.85]$. The distribution has mode around the true break fraction, $\lambda_0 = 0.5$, but is also relatively diffuse over Λ.[6]

For purposes of comparison, we also simulated the behaviour of $\hat{\lambda}_T$ in a model with no breaks and $(\beta_1^0, \beta_2^0) = [1, 0.1]'$, that is when it is assumed there is one break but in fact there are none; all other aspects of the design are the same as the one-break model above. As can be seen from Figure 2, the peak at $\lambda = 0.5$ is absent but the distribution of the break fraction estimators is similarly diffuse in the no-break and one-break models.

Propositions 1-3 indicate that a break-point estimation strategy based on the GMM minimand, while intuitively appealing at first sight, is flawed. This leaves us searching for an alternative approach for making valid inference in the multiple-break linear model with endogenous regressors. A way forward is suggested by inspection of the proof of Proposition 1. The source of the inconsistency lies in the structure of the minimand in (4). The minimand is a quadratic form in the sample moments, that is the square of sums. This structure affords the opportunity for the effects of misspecification to offset within the minimand. Such an opportunity is not afforded if the minimand is a sum of squares. Estimation based on a 2SLS minimand has exactly this structure, and in the remainder of this paper we demonstrate that this approach is simple to implement, yields consistent estimators of both the break-fractions and structural parameters and is also a convenient framework for inference within the multiple-break linear model with endogenous regressors.

3 Estimation based on 2SLS

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

$$y_t \;=\; x_t'\beta_{x,i}^0 \,+\, z_{1,t}'\beta_{z_1,i}^0 \,+\, u_t, \qquad i = 1, ..., m+1, \qquad t = T_{i-1}^0 + 1, ..., T_i^0 \tag{7}$$

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

[6] For the record, we note that the distribution looks qualitatively the same at $T = 10,000$. Results available from the authors upon request.

u_t is a mean zero error. We define $p = p_1 + p_2$. Given that some regressors are endogenous, it is plausible that (7) belongs to a system of structural equations and thus, for simplicity, we refer to (7) as the "structural equation".

As usual in the literature, we require the break points to be asymptotically distinct.

Assumption 6 $T_i^0 = [T\lambda_i^0]$, *where* $0 < \lambda_1^0 < ... < \lambda_m^0 < 1$.[7]

To implement 2SLS, it is necessary to specify the reduced form for x_t. As noted in the introduction, we consider scenarios in which the reduced form for x_t is either stable or unstable. In this section, we consider the case in which the reduced form is stable,

$$x_t' \ = \ z_t'\Delta_0 \ + \ v_t' \tag{8}$$

where $z_t = (z_{t,1}, z_{t,2}, ..., z_{t,q})'$ is a $q \times 1$ vector of instruments that is uncorrelated with both u_t and v_t, $\Delta_0 = (\delta_{1,0}, \delta_{2,0}, ..., \delta_{p_1,0})$ with dimension $q \times p_1$ and each $\delta_{j,0}$ for $j = 1, ..., p_1$ has dimension $q \times 1$. We assume that z_t contains $z_{1,t}$. Under the assumption that $E[{u_t}^2|z_t] = \sigma^2$, the optimal IV estimator is the 2SLS estimator.[8] Our analysis is confined to the 2SLS estimator, although note that the aforementioned conditional homoscedasticity restriction is only imposed in certain parts of the analysis.

We propose the following estimation method. On the first stage, the reduced form for x_t is estimated via OLS using (8) and let $\hat{x}_t$ denote the resulting predicted value for x_t, that is

$$\hat{x}_t' \ = \ {z_t}'\hat{\Delta}_T \ = \ {z_t}'(\sum_{t=1}^{T} z_t{z_t}')^{-1}\sum_{t=1}^{T} z_t{x_t}' \tag{9}$$

In the second stage, we first estimate

$$y_t \ = \ \hat{x}_t'\beta_{x,i}^* \ + \ z_{1,t}'\beta_{z_1,i}^* \ + \ \tilde{u}_t, \quad i = 1, ..., m+1; \quad t = T_{i-1}+1, ..., T_i \tag{10}$$

via OLS for each possible m-partition of the sample, denoted by $\{T_j\}_{j=1}^m$. We assume:

Assumption 7 *Equation (10) 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 7 requires that each segment considered in the minimization contains a positive fraction of the sample asymptotically; in practice ϵ is chosen to be small in the hope that the

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

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

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

$$S_T(T_1, ..., T_m; \beta) = \sum_{i=1}^{m+1} \sum_{t=T_{i-1}+1}^{T_i} (y_t - \hat{x}_t' \beta_{x,i} - z_{1,t}' \beta_{z_1,i})^2 \tag{11}$$

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

$$(\hat{T}_1, ..., \hat{T}_m) = \arg\min_{T_1, ..., T_m} S_T(T_1, ..., T_m;\ \hat{\beta}(\{T_i\}_{i=1}^m)) \tag{12}$$

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

4 2SLS based inference when the reduced form is stable

This section is divided into four parts. In part (i), we consider the limiting behaviour of both the break point fraction estimators $\{\hat{\lambda}_i = \hat{T}_i/T\}$ and the estimators of the structural parameters, $\hat{\beta}(\{\hat{T}_i\}_{i=1}^m)$. In part (ii), we propose a number of statistics for testing various hypotheses that naturally arise in models with multiple change points. Part (iii) describes how these test statistics can be used to estimate the number of break points.[9]

(i) Limiting behaviour of the estimators

To facilitate the analysis, we impose the following conditions.

Assumption 8 *(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 $diag[\xi_{T,1}^{-1}, \ldots, \xi_{T,n}^{-1}] = \Xi_T^{-1}$ is $O(T^{-1})$ where Ξ_T is the $n \times n$ diagonal matrix with the eigenvalues $(\xi_{T,1}, \ldots, \xi_{T,n})$ of V_T along the diagonal; (ii) $E[h_{t,i}] = 0$ and, for some $d > 2$, $\|h_{t,i}\|_d < \Gamma < \infty$*

[9] Bai, Chen, Chong, and Wang (2008) present an analysis of the multiple break in models with measurement error. Note that while their orthogonality condition implies stability of a corresponding reduced form, their setting is different from ours since they consider the properties of sequential break-point estimators, while we rely on a global analysis.

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 $\|h_t - E[h_t|\mathcal{G}_{t-m}^{t+m}]\|_2 \leq \nu_m$ with $\nu_m = O(m^{-1/2})$ where $\mathcal{G}_{t-m}^{t+m}$ is a sigma- algebra based on $(g_{t-m}, \ldots, g_{t+m})$; (iii) $\{g_t\}$ is either ϕ-mixing of size $m^{-d/(2(d-1))}$ or α-mixing of size $m^{-d/(d-2)}$.

Assumption 9 *$rank\{\,[\Delta_0,\, \Pi]\,\} = p$ where $\Pi' = [I_{p_2},\, 0_{p_2\times(q-p_2)}]$, I_a denotes the $a \times a$ identity matrix and $0_{a\times b}$ is the $a \times b$ null matrix.*

Assumption 10 *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 11 *$T^{-1}\sum_{t=1}^{[Tr]} z_t z_t' \xrightarrow{p} Q_{ZZ}(r)$ uniformly in $r \in [0,1]$ where $Q_{ZZ}(r)$ is positive definite for any $r > 0$ and strictly increasing in r.*

Assumption 8 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).[10] 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 9 implies the standard rank condition for identification in IV estimation in the linear regression model[11] because Assumptions 8(ii), 9 and 11 together imply that

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

where $Q_{Z,[X,Z_1]}(r)$ has rank equal to p for any $r > 0$. Assumption 10 requires that there are enough observations near the true break points so that they can be identified and is analogous to Bai and Perron's (1998) Assumption A2.

We first establish the consistency of the break fraction estimators via a similar argument to Bai and Perron (1998). The proof builds from the following two properties of the error sum of squares on the second stage of the 2SLS estimation: first, since the 2SLS estimators minimize

[10] 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).

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

the error sum of squares in (11), it follows that

$$(1/T)\sum_{t=1}^{T}\hat{u}_t^2 \;\le\; (1/T)\sum_{t=1}^{T}\tilde{u}_t^2 \tag{13}$$

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

$$T^{-1}\sum_{t=1}^{T}\hat{u}_t^2 \;=\; T^{-1}\sum_{t=1}^{T}\tilde{u}_t^2 \,+\, T^{-1}\sum_{t=1}^{T}{d_t}^2 \,-\, 2T^{-1}\sum_{t=1}^{T}\tilde{u}_t d_t. \tag{14}$$

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 (13)-(14) contradict each other. This conflict is established using the results in the following lemma.

Lemma 1 *Let y_t be generated by (7), x_t be generated by (8), $\hat{x}_t$ be generated by (9) and Assumptions 6-11 hold.*

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

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

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

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

Using (13)-(14) and Lemma 1, consistency is established along the lines anticipated above.

Theorem 1 *Let y_t be generated by (7), x_t be generated by (8), $\hat{x}_t$ be generated by (9) and Assumptions 6-11 hold, then $\hat{\lambda}_j \xrightarrow{p} \lambda^0_j$ for all $j = 1, 2, ..., m$.*

The consistency of the 2SLS-based break point estimator is in sharp contrast to the inconsistency of the GMM-based estimator established in Proposition 3. To illustrate the finite sample differences between the estimators, we simulated the behaviour of 2SLS-based estimator in the one-break model considered in Section 2 and plot the empirical distribution of the break fraction estimator in Figure 1. In contrast to the diffuse distribution of the GMM-based estimator, the distribution of the 2SLS-based estimator is very concentrated around the true break fraction.

For completeness, we also simulated the behaviour of the 2SLS-based estimator in the no-break model when the estimation is performed under the assumption of one break. In this case, the 2SLS-based and GMM-based estimators of the break fraction are similarly diffuse.

To establish asymptotic normality of the parameter estimators, we need to show that the break-fractions are converging faster than the parameters and thus their randomness does not contaminate the limiting distribution of the parameter estimators. This is established in the following result.

Theorem 2 *Let y_t be generated by (7), x_t be generated by (8), $\hat{x}_t$ be generated by (9) and Assumptions 6-11 hold then, for every $\eta > 0$, there exists C such that for all large T, $P(T|\hat{\lambda}_j - \lambda_j^0| > C) < \eta$, for $j = 1, ..., m$.*

Given Theorem 2, it can be shown that the limiting distribution of the 2SLS parameter estimators is the same as if the break-points are known *a priori.*

Theorem 3 *Let y_t be generated by (7), x_t be generated by (8), $\hat{x}_t$ be generated by (9) and Assumptions 6-11 hold, then*

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

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

$$\begin{aligned}
V_\beta &= \begin{pmatrix} V_\beta^{(1,1)} & \cdots & V_\beta^{(1,m+1)} \\ \vdots & \ddots & \vdots \\ V_\beta^{(m+1,1)} & \cdots & V_\beta^{(m+1,m+1)} \end{pmatrix} \\
V_{i,i} &= A_i\{C_iV_iC_i' - Q_iQ_{ZZ}(1)^{-1}D_iV_iC_i' - C_iV_iD_i'Q_{ZZ}(1)^{-1}Q_i + Q_iQ_{ZZ}(1)^{-1}D_iVD_i'Q_{ZZ}(1)^{-1}Q_i\}A_i' \\
V_{i,j} &= A_iQ_iQ_{ZZ}(1)^{-1}D_iVD_j'Q_{ZZ}(1)^{-1}Q_jA_j' - A_iQ_iQ_{ZZ}(1)^{-1}D_iV_jC_j'A_j' \\
&\quad - A_iC_iV_iD_j'Q_{ZZ}(1)Q_jA_j', \qquad \textit{for } i \neq j \\
A_i &= [\Psi'Q_i\Psi]^{-1}\Psi', \qquad \textit{for } i = 1, 2, \ldots m+1 \\
\Psi &= [\Delta_0, \Pi], \qquad C_i = [I_q, \beta_{x,i}^{0\ \prime} \otimes I_q], \qquad D_i = [0_{q\times q}, \beta_{x,i}^{0\ \prime} \otimes I_q] \\
Q_i &= Q_{ZZ}(\lambda_i^0) - Q_{ZZ}(\lambda_{i-1}^0), \qquad V_i = Var\left[T^{-1/2}\sum_{t=[\lambda_{i-1}T]+1}^{[\lambda_iT]} h_t\right], \qquad V = Var\left[T^{-1/2}\sum_{t=1}^{T} h_t\right].
\end{aligned}$$

Note that $V_{(i,j)}$ is non-zero in general because the first stage regression pools observations across regimes and this creates a connection between the 2SLS estimators from different regimes. A consistent estimator of this variance can be constructed in a straightforward fashion by replacing Δ_0, β^0, $Q_{ZZ}(r)$, Q_i, V_i and V by respectively $\hat{\Delta}_T$, $\hat{\beta}(\{\hat{T}_i\}_{i=1}^m)$, $T^{-1}\sum_{t=1}^{[Tr]} z_t z_t'$, and HAC estimators of V_i and V based on $\hat{u}_t = y_t - (x_t', z_{1,t}')\hat{\beta}(\{\hat{T}_i\}_{i=1}^m)$ and $\hat{v}_t = x_t - \hat{\Delta}_T' z_t$.[12]

(ii) Hypothesis Testing:

In this sub-section, we consider three types of hypothesis tests that naturally arise in this class of models: (a) $H_0 : m = 0$ vs $H_1 : m = k$; (b) $H_0 : m = 0$ vs $H_1 : m \leq K$; (c) $H_0 : m = \ell$ vs $H_1 : m = \ell+1$. We consider F-type tests and Wald-type tests for each. To develop both types of tests, we need to impose additional assumptions on the instrument cross-product matrix and long run variance of the instrument-error product vector, h_t. The exact nature of the assumptions depends on the type of statistic and the null hypothesis.

We begin by considering F-type statistics for $H_0 : m = 0$. For this scenario, we impose the following two assumptions.

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

Assumption 13 *Let* $b_t = (u_t, v_t')'$ *and* $\mathcal{F} = \sigma - field\{\ldots, z_{t-1}, z_t, \ldots, b_{t-2}, b_{t-1}\}$. b_t *is a martingale difference relative to* $\{\mathcal{F}_t\}$ *and* $sup_t E[\|b_t\|^4] < \infty$ *and the conditional variance of the errors is independent of* t*, that is* $Var[u_t, v_t'|z_t] = \Omega$*, a constant pd matrix with the conditional variances of* u_t *and* v_t *denoted by* σ^2 *and* Σ *respectively, and the conditional covariance between* u_t *and* v_t *denoted by* γ'*.*

The restrictions in Assumptions 12-13 are analogous to those imposed by Bai and Perron (1998) in their Assumptions A8 and A9 which underpin their analysis of various F-statistics for testing for multiple breaks within the OLS framework.

The sup-F type test of $H_0 : m = 0$ vs $H_A : m = 1$ has been considered by Andrews (1993). The results below are the 2SLS extensions of Bai and Perron's (1998) tests.

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

$$F_T(\lambda_1, ..., \lambda_k; p) \; = \; \left\{ \frac{T - (k+1)p}{kp} \right\} \left\{ \frac{SSR_0 - SSR_k}{SSR_k} \right\} \tag{15}$$

[12]See Andrews (1991) for details of HAC estimators.

where SSR_0 and SSR_k are the sum of squared residuals based on fitted x_t under null and alternative hypothesis, respectively. Recall from Assumption 7 that the minimization is performed over partitions which are asymptotically large and the size of the partitions is controlled by ϵ, a non-negative constant. Accordingly, we define

$$\Lambda_\epsilon = \{(\lambda_1, ..., \lambda_k) : |\lambda_{i+1} - \lambda_i| \geq \epsilon, \lambda_1 \geq \epsilon, \lambda_k \leq 1 - \epsilon\}.$$

Finally, the sup-F test statistic is defined as

$$Sup - F_T(k; p) \; = \; Sup_{(\lambda_1,...,\lambda_k)\in\Lambda_\epsilon} F_T(\lambda_1, .., \lambda_k; p) \tag{16}$$

Theorem 4 *If the data are generated by (7)-(8) with $m = 0$, $\hat{x}_t$ is generated by (9) and Assumptions 6-13 hold then*[13] $Sup - F_T(k; p) \Rightarrow Sup - F_{k,p} \equiv Sup_{(\lambda_1,...,\lambda_k)\in\Lambda_\epsilon} F(\lambda_1, .., \lambda_k; p)$ *where*

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

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

We note that the limiting distribution in Theorem 4 is exactly the same as the one in Bai and Perron's (1998) analogous result for the sup-F test based on OLS estimators when the regressors are exogenous. Percentiles for this distribution can be found in Bai and Perron (1998)[Table I] for $\epsilon = 0.05$ and in Bai and Perron (2001) for other values of ϵ.

The $Sup - F_T(k; p)$ statistic is used to test the null hypothesis of structural stability against the k-break model, and so is designed for the case in which a particular choice of k is of interest. In many circumstances, a researcher is unlikely to know *a priori* the appropriate choice of k for the alternative hypothesis. To circumvent this problem, Bai and Perron (1998) propose so called "Double Maximum tests" that combine information from the $Sup - F_T(k; p)$ statistics for different values of k running from one to some ceiling K. We consider here only the following example of Double Maximum test,[14]

$$UDmaxF_T(K; p) \; = \; \max_{1\leq k\leq K} \; \sup_{(\lambda_1,...,\lambda_k)\in\Lambda_\epsilon} \; F_T(\lambda_1, ..., \lambda_k; p) \tag{17}$$

The limiting distribution of this statistic follows directly from Theorem 4.

[13] "$\Rightarrow$" denotes weak convergence in the space $D[0, 1]$ under the Skorohod metric.

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

Corollary 1 *Under the conditions of Theorem 4, it follows that*

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

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

The $Sup - F_T(k;p)$ and $UDmaxF_T(K;p)$ statistics are used to test the null hypothesis of no breaks. It is also of interest to develop statistics for testing the null hypothesis of l breaks against the alternative of $l+1$ breaks. For this scenario, we relax Assumptions 12 and 13 as follows.

Assumption 14 $T^{-1}\sum_{t=[Tr]+1}^{[Ts]} z_t z_t' \xrightarrow{p} (r-s)Q_{ZZ}^{(i)}$, *where* $\lambda_{i-1}^0 \leq r < s \leq \lambda_i^0$, *uniformly in* $r\times s$ *and* $Q_{ZZ}^{(i)}$ *is a positive definite matrix of constants, not necessarily the same for all* i.

Assumption 15 $Var\left[(u_t, v_t')'\right) \mid z_t\,] = \Omega_i$, *a pd matrix of constants, for* $t \in \left([T\lambda_{i-1}^0]+1, [T\lambda_i^0]\right)$ *and* σ_i^2, Σ_i *and* γ_i *denote the sub-matrices of* Ω_i *relating respectively to the conditional variance of* u_t, *the conditional variance of* v_t *and the conditional covariance of* v_t *and* u_t.

Notice that Assumption 14 only imposes homogeneity of the instrument cross-product matrix within each regime and Assumption 15 allows the conditional error variance to change at the same time as the structural parameters.

Following Bai and Perron (1998), a suitable statistic can be constructed as follows. For the model with l breaks, the estimated break points, denoted by $\hat{T}_1, ..., \hat{T}_l$, are obtained by a global minimization of the sum of the squared residuals as in (12). For the model with $l+1$ breaks, l of the breaks are fixed at $\hat{T}_1, ..., \hat{T}_l$ and then the location of the $(l+1)^{th}$ break is chosen by minimizing the residual sum of squares. The test statistic is given by

$$F_T(l+1|l) = \max_{1\leq i\leq l+1}\left\{\frac{SSR_l(\hat{T}_1,...,\hat{T}_l) - \inf_{\tau\in\Lambda_{i,\eta}} SSR_{l+1}(\hat{T}_1,...,\hat{T}_{i-1},\tau,\hat{T}_i,...,\hat{T}_l)}{\hat{\sigma}_i^2}\right\} \quad (18)$$

where

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

and $\hat{\beta}_i' = (\hat{\beta}_{x,i}', \hat{\beta}_{z_1,i}')$ is the 2SLS estimator calculated using the sample $\hat{T}_{i-1}+1, \ldots, \hat{T}_i$ on the second stage. The following theorem gives the limiting distribution of this statistic under the null hypothesis of l breaks.

Theorem 5 *If the data are generated by (7)-(8) with $m = l$, $\hat{x}_t$ is generated by (9) and Assumptions 6- 11, 14 and 15 hold then* $\lim_{T\to\infty} P(F_T(l+1|l) \leq x) = G_{p,\eta}(x)^{l+1}$ *where $G_{p,\eta}(x)$ is the distribution function of* $\sup_{\eta\leq\mu\leq 1-\eta} \|W(\mu) - \mu W(1)\|^2/\mu(1-\mu)$ *and $W(\mu) \equiv B_p(\mu)$.*

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

The restriction on the errors in Assumptions 13 or 15 is satisfied in some applications but rules out many other cases of interest. Unfortunately, it is not simple to modify the F-type statistics to handle more general error processes, and so we also consider statistics based on the Wald principle. For this part of the analysis, the errors are only restricted to satisfy the following:

Assumption 16 *Define* $V_T(r) = Var[T^{-1/2}\sum_{t=1}^{[Tr]} h_t]$ *then* $V_T(r) \rightarrow rV$ *uniformly in* $r \in [0,1]$ *where V is a positive definite matrix.*

Notice that this assumption allows for serial correlation and conditional heteroscedasticity in h_t and, thus, in the errors u_t and v_t. However, note that we maintain Assumption 8(ii) which includes $E[h_t] = 0$, and so if the errors are serially correlated then, in general, z_t must exclude lagged values of y_t or x_t.

To develop the Wald test of $H_0 : m = 0$ versus $H_1 : m = k$, we restate the null and alternative hypotheses in terms of linear restrictions on the parameters. Accordingly, we define $R_k = \tilde{R}_k \otimes I_p$ where $\tilde{R}_k$ is the $k \times (k+1)$ matrix whose $i - j^{th}$ element, $\tilde{R}_k(i,j)$, is given by: $\tilde{R}_k(i,i) = 1$, $\tilde{R}_k(i,i+1) = -1$, $\tilde{R}_k(i,j) = 0$ for $i = 1, 2, \ldots k$ and $j \neq i, i+1$. The null and alternative can then be equivalently stated as: $H_0 : R_k\beta^0(k) = 0$ versus $H_1 : R_k\beta^0(k) \neq 0$ where $\beta^0(k) = (\beta_1^{0\prime}, \beta_2^{0\prime}, \ldots, \beta_k^{0\prime})'$. The test statistic is then:

$$Sup - Wald_T(k,p) \;=\; \sup_{(\lambda_1,\lambda_2\ldots\lambda_k)\in\Lambda_\epsilon} T\hat{\beta}(\bar{T}_k)'R_k'[R_k\hat{V}_W(\bar{T}_k)R_k']^{-1}R_k\hat{\beta}(\bar{T}_k) \tag{19}$$

where $\hat{\beta}(\bar{T}_k)$ is the 2SLS estimator of $\beta^0(k)$ based on k-partition $\bar{T}_k = ([\lambda_1 T], \ldots, [\lambda_k T])$, $\hat{V}_W(\bar{T}_k) = diag\left[\hat{V}_W^{(1)}(\bar{T}_k), \ldots, \hat{V}_W^{(k)}(\bar{T}_k)\right]$,

$$\hat{V}_W^{(i)}(\bar{T}_k) \;=\; \left\{T^{-1}\sum_{t=[\lambda_{i-1}T]+1}^{[\lambda_i T]} \hat{x}_t\hat{x}_t'\right\}^{-1} \hat{H}_i(\bar{T}_k)\left\{T^{-1}\sum_{t=[\lambda_{i-1}T]+1}^{[\lambda_i T]} \hat{x}_t\hat{x}_t'\right\}^{-1}$$

and $\hat{H}_i(\bar{T}_k)$ is a consistent estimator of $H_i = lim_{T\to\infty} Var[T^{-1/2}\sum_{t=[\lambda_{i-1}T]+1}^{[\lambda_i T]} \Delta_0' z_t\{u_t + v_t'\beta^0_{x,i}(k)\}]$. $\hat{H}_i(\bar{T}_k)$ can be constructed using a HAC estimator based on $\hat{\Delta}_T' z_t\{\hat{u}_t + \hat{v}_t'\hat{\beta}_x\}$, $\hat{u}_t = y_t - x_t'\hat{\beta}_x - z_{1,t}'\hat{\beta}_{z_1}$ and $\hat{v}_t = x_t - \hat{\Delta}_T' z_t$, and $\{\hat{\beta}_x, \hat{\beta}_{z_1}\}$ are the 2SLS estimators of the coefficients on x and z_1 obtained under the null hypothesis of no breaks.

An important feature of $\hat{V}_W^{(i)}(\bar{T}_k)$ is that it ignores the dependence across sub-samples noted in the discussion following Theorem 3. The reason for this is as follows: under Assumption 12, $T^{1/2}R_k\hat{\beta}(\bar{T}_k)$ does not involve the terms that create the dependence between estimators from different regimes. The following theorem gives the limiting distribution of the sup-Wald test.

Theorem 6 *If the data are generated by (7)-(8) with $m = 0$, $\hat{x}_t$ is generated by (9) and Assumptions 6-12 and 16 hold then*

$$Sup - Wald_T(k,p) \;\Rightarrow\; \sum_{i=1}^{k} \frac{||\lambda_{i+1}W_i - \lambda_i W_{i+1}||^2}{\lambda_i\lambda_{i+1}(\lambda_{i+1} - \lambda_i)}$$

where k is the number of break points under the alternative hypothesis.

A comparison of Theorems 4 and 6 indicates that $(1/kp)Sup - Wald_T(k,p)$ has the same limiting distribution as $Sup - F_T(k;p)$.

To test $H_0 : m = 0$ vs $H_1 : m \le K$, we define analogously to $UDmaxF_T(K;p)$ the statistic:

$$UDmaxWald_T(K;p) \;=\; \max_{1\le k\le K}(1/kp)Sup - Wald_T(k,p)$$

Corollary 2 *Under the conditions of Theorem 6, it follows that*

$$UDmaxWald_T(K;p) \;\Longrightarrow\; \max_{1\le k\le K}\{Sup - F_{k,p}\}$$

The limiting distribution of $UDmaxWald_T(K;p)$ is identical to that for $UDmaxF_T(K;p)$ given in Corollary 1. Notice that the test statistic involves $Sup - Wald_T(k,p)$ divided by kp; this scaling is employed because the limiting distribution of $Sup - Wald_T(k,p)$ is increasing in k for fixed p and so, without the scaling, the test statistic $\max_{1\le k\le K} Sup - Wald_T(k,p)$ would be equivalent to testing 0 versus K breaks.

To test $H_0 : m = \ell$ vs $H_1 : m = \ell + 1$ via the Wald principle, we proceed as follows. Under the null hypothesis, there are ℓ breaks and hence $\ell + 1$ regimes within which the parameters are constant; under the alternative one of these regimes contains an additional break point at which the parameters change. We can therefore test the null hypothesis by calculating, for each of the

$\ell + 1$ regimes, the Wald statistic for a single break and then basing inference on the supremum of these $\ell + 1$ statistics. Therefore, the test statistic is:

$$Wald_T(\ell + 1|\ell) = \max_{1 \leq i \leq \ell+1} \sup_{\tau \in \Lambda_{i,\eta}} Wald_{T,\ell}(\tau, i; p)$$

where $Wald_{T,\ell}(\tau, i; p)$ is defined to be the Wald statistic for a single break at $t = \hat{T}_{i-1} + \tau$ based on the sub-sample $\Lambda_{i,\eta}$, that is

$$Wald_{T,\ell}(\tau, i; p) \; = \; T\hat{\beta}(\tau; i)' R_1' [R_1 \hat{V}_W(\tau; i) R_1']^{-1} R_1 \hat{\beta}(\tau; i)$$

where $\hat{\beta}(\tau; i) = [\hat{\beta}_1'(\tau; i), \hat{\beta}_2'(\tau; i)]'$, $\hat{\beta}_1(\tau; i)$ are the 2SLS estimators of the parameters in the structural equation based on observations $S_1(\tau, i) = \{\hat{T}_{i-1} + 1, \hat{T}_{i-1} + 2, \ldots, \hat{T}_{i-1} + \tau\}$, $\hat{\beta}_2(\tau; i)$ are the 2SLS estimators of the parameters in the structural equation based on observations $S_2(\tau, i) = \{\hat{T}_{i-1} + \tau + 1, \ldots, \hat{T}_i\}$, $\hat{V}_W(\tau; i) = diag[\hat{V}_W^{(1)}(\tau; i), \hat{V}_W^{(2)}(\tau; i)]$,

$$\hat{V}_W^{(j)}(\tau; i) \; = \; \{T^{-1} \sum_{S_j(\tau,i)} \hat{x}_t \hat{x}_t'\}^{-1} \hat{H}_i^{(j)}(\bar{T}_k) \{T^{-1} \sum_{S_j(\tau,i)} \hat{x}_t \hat{x}_t'\}^{-1},$$

$\sum_{S_j(\tau,i)}$ denotes summation over $t \in S_j(\tau, i)$ for $j = 1, 2$, and $\hat{H}_i^{(j)}$ is a consistent estimator of $lim_{T\to\infty} Var[T^{-1/2} \sum_{S_j(\tau,i)} \Delta_0' z_t \{u_t + v_t' \beta_{x,i}^0\}]$. $\hat{H}_i^{(j)}$ can be constructed using a HAC estimator based on $\hat{\Delta}_T' z_t \{\hat{u}_t + \hat{v}_t' \hat{\beta}_{x,i}\}$, $\hat{u}_t = y_t - x_t' \hat{\beta}_{x,i} - z_{1,t}' \hat{\beta}_{z_1,i}$ and $\hat{v}_t = x_t - \hat{\Delta}_T' z_t$; such an estimator is consistent under H_0. The following theorem gives the limiting distribution of $Wald_T(\ell + 1|\ell)$.

Theorem 7 *If the data are generated by (7)-(8) with* $m = \ell$, $\hat{x}_t$ *is generated by (9) and Assumptions 6 - 11, 14 and 16 hold then* $\lim_{T\to\infty} P(Wald_T(\ell + 1|\ell) \leq x) = G_{p,\eta}(x)^{l+1}$ *where* $G_{p,\eta}(x)$ *is defined in Theorem 5.*

(iii) Estimation of the number of breaks

Following Bai and Perron (1998), the statistics described in this section can be used to determine the estimated number of break points, $\hat{m}_T$ say, via the following sequential strategy (for illustrative purposes we describe the method in terms of the F-type statistics but the same strategy can also be used with the Wald-type tests). On the first step, use either $Sup - F_T(1; p)$ or $UDmaxF_T(K, p)$ to test the null hypothesis that there are no breaks. If this null is not rejected then $\hat{m}_T = 0$; else proceed to the next step. On the second step $F_T(2|1)$ is used to test the null hypothesis that there is only one break against the alternative hypothesis of two breaks. If $F_T(2|1)$ is insignificant then $\hat{m}_T = 1$; else proceed to the next step. On the l^{th} step $F_T(l + 1|l)$

is used to test the null hypothesis that there are l breaks against the alternative hypothesis of $l+1$ breaks. If $F_T(l+1|l)$ is insignificant then $\hat{m}_T = l$; else proceed to the next step. This sequence is continued until some preset ceiling for the number of breaks, L say, is reached. If all statistics in the sequence are significant then the conclusion is that there are at least L breaks.

(iv) Finite sample performance

In this sub-section, we evaluate the finite sample performance of the methods described in this section. We consider in order models with one, two and no breaks.

One break model: We return to the model used in the simulations reported in Section 2, except this time, we report results for $q = 4, 8$ and $T = 120, 240, 480$. Recall that Figure 1 contains a plot of the empirical distribution of $\hat{\lambda}_1$ for the estimation with $m = 1$. It can be seen that this distribution is collapsing toward a point mass of one at $\lambda_1^0 = 0.5$ as T increases in line with Theorem 1. Table 1 reports the coverage probabilities of the 2SLS estimator of β_i^0 based on the asymptotic distribution in Theorem 3.[15] As can be seen, the coverage is close to the nominal levels. Table 2 reports the rejection frequencies for the F-type and Wald-type statistics. Specifically, we report values for: (i) the $Sup - F_T(k;1)$ and $Sup - Wald_T(k;1)$ statistics with $k = 1, 2$, and the $UDmaxF_T(5,1)$ and $UDmaxWald_T(5,1)$; note that the null hypothesis is incorrect for these statistics; (ii) the $F_T(l+1|l)$ and $Wald_T(l+1|l)$ statistics for $l = 1, 2, 3$; note that the null is correct for $l = 1$ but involves more than the true number of breaks for $l > 1$. It can be seen that the sup - type and UDmax - type statistics correctly reject the null with probability one. The $F_T(2|1)$ and $Wald_T(2|1)$ statistics are slightly undersized but close to their nominal size; if l exceeds the true number of breaks then both $F_T(l+1|l)$ and $Wald_T(l+1|l)$ reject very rarely. Table 3 reports the empirical distribution of the estimated number of break points obtained using the sequential strategy in (iii) above with $L = 5$. We first note that the results are identical whether the $Sup - F_T(1;p)$ $(Sup - Wald_T(1;p))$ or the

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

$UDmaxF_T(5,1)$ ($UDmaxWald_T(5,1)$) statistic is used on the first step (and so we only report the latter) although there are some slight differences if the F - type or Wald - type statistic is used. As can be seen, the method estimates the true number with probability never less than 93% and never underfits. Overfitting is confined to picking two breaks (one too many) with a three break model being picked only once in some designs; more than three breaks are never selected.

Two break model: The data generation process for the structural equation is:

$$y_t \; = \; [1,\, x_t]'\beta_i^0 \; + \, u_t,$$

where $\beta_i^0 = (-1)^{i+1}[1, 0.1]$ for $t = [\lambda_{i-1}T + 1, \ldots [\lambda_i T]$, $\lambda_1 = 1/3$, $\lambda_2 = 2/3$. All other aspects of the design are the same as the one break model.

Figure 3 contains plots of the empirical distribution of the break fraction estimators for the estimation with $m = 2$. It can be seen that the distribution for each break fraction estimator is collapsing toward a point mass of one at the appropriate true parameter value (0.33 or 0.66) as T increases in line with Theorem 1. Table 4 reports the coverage probabilities of the 2SLS estimator of β_i^0 based on the asymptotic distribution in Theorem 3. As in the one break model, the coverage probabilities are very close to the nominal levels. Table 5 reports the rejection frequencies for the test statistics. As in the one break model, the null hypothesis is incorrect for the $Sup - F_T(k; 1)$ and $Sup - Wald_T(k; 1)$ statistics with $k = 1, 2$, and the $UDmaxF_T(5,1)$ and $UDmaxWald_T(5,1)$ statistics. However, this time for $F_T(l+1|l)$ and $Wald_T(l+1|l)$, the null is incorrect for $l = 1$ but correct for $l = 2$. It can be seen that the sup - type and UDmax - type statistics, $F_T(2|1)$ and $Wald_T(2|1)$ correctly reject the null with probability one. The $F_T(3|2)$ and $Wald_T(3|2)$ statistics are slightly undersized but close to their nominal size. Table 6 reports the empirical distribution of the estimated number of break points obtained using the sequential strategy in (iii) above with $L = 5$.[16] As can be seen, the method estimates the true number with probability never less than 95% and never underfits. Overfitting is confined to picking three breaks (one too many).

No break model: Data are generated from (5) with $\beta_1^0 = \beta_2^0 = [1, 1]$. All other aspects of the design are the same as the one break model. Table 7 contains the empirical rejection frequencies

[16] As in the one break model, the results are the same whether the $Sup - F_T(1;p)$ ($Sup - Wald_T(1;p)$) or the $UDmaxF_T(5,1)$ ($UDmaxWald_T(5,1)$) statistic is used on the first step.

of the test statistics: note that the null hypothesis is correct for all statistics except $F_T(l+1|l)$ and $Wald_T(l+1|l)$ for which the null involves the assumption of (too many) breaks. It can be seen that the sup- and UDmax- type tests based on the F statistic are close to their nominal size but the corresponding tests based on the Wald statistic tend to be slightly over-sized. Interestingly the sup- type Wald tests are closer to their nominal size than the UDmax-Wald test. This difference has implications for the estimation of the number of breaks: the sequential strategy based on F-statistics selects the true value of m at least 94% of the time, but the strategy based on the Wald statistics only does so at least 90% of the time.

5 Unstable Reduced Form: Model and Estimation

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

$$x_t^{'} \;=\; z_t^{'}\Delta_0^{(i)} \;+\; v_t^{'}, \qquad i = 1, 2, \ldots, h+1, \qquad t = T_{i-1}^{*}+1, \ldots, T_i^{*} \tag{20}$$

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

Assumption 17 $T_i^* = [T\pi_i^0]$, *where* $0 < \pi_1^0 < \ldots < \pi_h^0 < 1$.

Note that the break fractions $\{\pi_i^0\}$ may or may not coincide with $\{\lambda_i^0\}$. Let $\pi^0 = [\pi_1^0, \pi_2^0, \ldots, \pi_h^0]'$. Also note that (20) can be re-written as follows

$$x_t^{'} \;=\; \tilde{z}_t(\pi^0)^{'}\Theta_0 \;+\; v_t^{'}, \qquad t = 1, 2, \ldots, T \tag{21}$$

where $\Theta_0 = [\Delta_0^{(1)'}, \Delta_0^{(2)'}, \ldots, \Delta_0^{(h+1)'}]'$, $\tilde{z}_t(\pi^0) = \iota(t,T) \otimes z_t$, $\iota(t,T)$ is a $(h+1) \times 1$ vector with first element $\mathcal{I}\{t/T \in (0, \pi_1^0]\}$, $h+1^{th}$ element $\mathcal{I}\{t/T \in (\pi_h^0, 1]\}$, k^{th} element $\mathcal{I}\{t/T \in (\pi_{k-1}^0, \pi_k^0]\}$ for $k = 1, 2, \ldots, h$ and $\mathcal{I}\{\cdot\}$ is an indicator variable that takes the value one if the event in the curly brackets occurs. Notice that (21) fits the generic constant parameter form of (8), and this similarity facilitates the analysis of the limiting properties of the estimators below.

Within our analysis, it is assumed that the break points in the reduced form are estimated prior to estimation of the structural equation in (7). 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 Section 3. However, the presence of this additional source of instability means that it is also necessary to modify Assumption 7.

Assumption 18 *The minimization in (12) 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* $\epsilon < inf_i(\lambda_{i+1}^0 - \lambda_i^0)$ *and* $\epsilon < inf_j(\pi_{j+1}^0 - \pi_j^0)$.

The remainder of our discussion focuses on the unstable reduced form case. In part (i), we consider the limiting behaviour of the estimators of the break fraction and the structural parameters, and in part (ii) we consider hypothesis testing and estimation of the number of breaks.

(i) Limiting behaviour of the estimators

We suppose that the vector of true break points in the reduced form, π^0, is estimated by $\hat{\pi}$ which satisfies the following condition.

Assumption 19 $\hat{\pi} \; = \; \pi^0 \; + \; O_p(T^{-1})$

Note that Assumption 19 implies $\hat{\pi}$ is consistent for π^0 and $T(\hat{\pi} - \pi^0)$ is bounded in probability. Such an estimator might be obtained by applying Bai and Perron (1998)'s methodology equation by equation and then pooling the resulting estimates of the break fractions. For our purposes, it only matters that Assumption 19 holds and not how $\hat{\pi}$ is obtained. The latter is, of course, a matter of practical importance but its exploration is beyond the scope of this paper.

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

$$x_t' \;=\; \tilde{z}_t(\hat{\pi})'\Theta_0 + error \qquad t = 1, 2, \cdots, T \tag{22}$$

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

$$\hat{x}_t' \;=\; \tilde{z}_t(\hat{\pi})'\hat{\Theta}_T \;=\; \tilde{z}_t(\hat{\pi})'\{\sum_{t=1}^{T} \tilde{z}_t(\hat{\pi})\tilde{z}_t(\hat{\pi})'\}^{-1} \sum_{t=1}^{T} \tilde{z}_t(\hat{\pi})x_t' \tag{23}$$

In our analysis we maintain Assumptions 8, 10 and 11 but need to replace the identification condition in Assumption 9 by the following condition.

Assumption 20 $rank\left\{ \left[\Delta_0^{(i)}, \Pi\right] \right\} = p$ *for* $i = 1, 2, \cdots, h+1$ *for* Π *defined in Assumption 9.*

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

Theorem 8 *If Assumptions 6, 8, 10, 11, 17-20 hold,* y_t *is generated via (7),* x_t *is generated via (21) and* $\hat{x}_t$ *is calculated via (23), then*

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

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

Assumption 21 *There exists an $l_* > 0$ such that for all $l > l_*$, the minimum eigenvalues of $B_{il} = (1/l)\sum_{t=T_i^*+1}^{T_i^*+l} z_t z_t{}'$ and of $B_{il}^* = (1/l)\sum_{t=T_i^*-l}^{T_i^*} z_t z_t{}'$ are bounded away from zero for all $i = 1, ..., h+1$.*

Assumption 21 is similar to Assumption 10 above but refers to the break points in the reduced form. The order in probability of the estimated break fractions is given in the following theorem.

Theorem 9 *If Assumptions 6, 8, 10, 11, 17-21 hold, y_t is generated via (7), x_t is generated via (21) and $\hat{x}_t$ is calculated via (23), then, for every $\eta > 0$, there exists C such that for all large T, $P(T|\hat{\lambda}_j - \lambda_j^0| > C) < \eta$, for $j = 1, ..., m$.*

We now consider the limiting distribution of the structural parameter estimators.

Theorem 10 *If Assumptions 6, 8, 10, 11, 17-21 hold, y_t is generated via (7), x_t is generated via (21) and $\hat{x}_t$ is calculated via (23), then*

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

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

$$\begin{aligned}
V_\beta &= \begin{pmatrix} V_\beta^{(1,1)} & \cdots & V_\beta^{(1,m+1)} \\ \vdots & \ddots & \vdots \\ V_\beta^{(m+1,1)} & \cdots & V_\beta^{(m+1,m+1)} \end{pmatrix} \\
V_{i,i} &= \tilde{A}_i\{\tilde{C}_i\tilde{V}_i\tilde{C}_i' - \tilde{Q}_i\tilde{Q}_{ZZ}(1)^{-1}\tilde{D}_i\tilde{V}_i\tilde{C}_i' - \tilde{C}_i\tilde{V}_i\tilde{D}_i'\tilde{Q}_{ZZ}(1)^{-1}\tilde{Q}_i + \tilde{Q}_i\tilde{Q}_{ZZ}^{-1}(1)\tilde{D}_i\tilde{V}\tilde{D}_i'\tilde{Q}_{ZZ}(1)^{-1}\tilde{Q}_i\}\tilde{A}_i' \\
V_{i,j} &= \tilde{A}_i\tilde{Q}_i\tilde{Q}_{ZZ}(1)^{-1}\tilde{D}_i\tilde{V}\tilde{D}_j'\tilde{Q}_{ZZ}(1)^{-1}\tilde{Q}_j\tilde{A}_j' - \tilde{A}_i\tilde{Q}_i\tilde{Q}_{ZZ}(1)^{-1}\tilde{D}_i\tilde{V}_j\tilde{C}_j'\tilde{A}_j' \\
&\quad - \tilde{A}_i\tilde{C}_i\tilde{V}_i\tilde{D}_j'\tilde{Q}_{ZZ}(1)^{-1}\tilde{Q}_j\tilde{A}_j', \qquad \textit{for } i \neq j \\
\tilde{A}_i &= [\tilde{\Psi}'\tilde{Q}_i\tilde{\Psi}]^{-1}\tilde{\Psi}', \qquad \textit{for } i = 1, 2, \ldots m+1 \\
\tilde{\Psi}' &= [\Psi_1', \Psi_2', \ldots, \Psi_{h+1}'], \qquad \Psi_i = [\Delta_0^{(i)}, \Pi], \\
\tilde{C}_i &= [I_{\tilde{q}}, \beta_{x,i}^{0}{}' \otimes I_{\tilde{q}}], \qquad D_i = [0_{\tilde{q}\times\tilde{q}}, \beta_{x,i}^{0}{}' \otimes I_{\tilde{q}}], \qquad \tilde{q} = q(h+1), \\
\tilde{Q}_i &= \tilde{Q}_{ZZ}(\lambda_i^0) - \tilde{Q}_{ZZ}(\lambda_{i-1}^0), \qquad \tilde{Q}_{ZZ}(\lambda) = plim\, T^{-1}\sum_{t=1}^{[\lambda T]}\tilde{z}_t(\pi^0)\tilde{z}_t(\pi^0)' \\
\tilde{V}_i &= Var\left[T^{-1/2}\sum_{t=[\lambda_{i-1}T]+1}^{[\lambda_i T]}\tilde{h}_t\right], \qquad \tilde{V} = Var\left[T^{-1/2}\sum_{t=1}^{T}\tilde{h}_t\right], \qquad \tilde{h}_t = (u_t, v_t') \otimes \tilde{z}_t(\pi^0).
\end{aligned}$$

In general, the form of the covariance matrix depends on the relative locations of the breaks in the structural equation and the reduced form. However, it is worth noting that certian simplifications are possible in cases that may be of empirical relevance. First, if all the breaks in the structural and reduced form equations coincide then we have the following result.

Corollary 3 *Under the conditions of Theorem 10, if* $m = h$ *and* $\lambda_i^0 = \pi_i^0$ *for all* $i = 1, 2, \ldots m$ *then* $V_\beta = diag[V_{1,1}, V_{2,2}, \ldots V_{m+1,m+1}]$ *where* $V_{i,i} = \bar{A}_i \bar{H}_i \bar{A}_i'$ *where* $\bar{A}_i = [\Delta_0^{(i)\prime} Q_i \Delta_0^{(i)}]^{-1} \Delta_0^{(i)\prime}$ *and* $\bar{H}_i = lim_{T\to\infty} Var\left[T^{-1/2} \sum_{t=[\lambda_{i-1}^0 T]+1}^{[\lambda_i^0 T]} z_t u_t\right]$.

The intuition behind this result is that in this case the terms involving the reduced form error cancel out asymptotically in $T^{-1/2} \sum_{t=[\lambda_{i-1}^0 T]+1}^{[\lambda_i^0 T]} z_t \tilde{u}_t$. Second, if there are more breaks in the reduced form than in the structural equation but all the breaks in the structural equation coincide with a corresponding break in the reduced form then we have the following result.

Corollary 4 *Under the conditions of Theorem 10, if* $m < h$ *and* $\lambda_i^0 = \pi_{j(i)}^0$ *for all* $i = 1, 2, \ldots m$ *and some* $j(i)$ *then* $V_\beta = diag[V_{1,1}, V_{2,2}, \ldots V_{m+1,m+1}]$ *where* $V_{i,i}$ *is defined in Theorem 10.*

The intuition behind this result is that the pattern of the breaks means that there is no correlation asymptotically between the 2SLS estimators in different regimes.

(ii) Hypothesis Testing and Estimation of the Number of Breaks

In the case where the reduced form is stable, it is possible to develop statistics with the distributions tabulated in Bai and Perron (1998). Unfortunately, these statistics do not appear to extend directly to the unstable reduced form case. For while the unstable reduced form in (20) can be re-written as a "stable reduced form" involving augmented parameter and instrument vectors, it does not satisfy the assumptions imposed in the derivation of the tests in Section 4 above. To illustrate this issue, consider the assumed behaviour of the instrument cross-product matrix, $T^{-1} \sum_{t=1}^{[Tr]} z_t z_t'$. Under Assumption 12, the limit of this matrix is rQ_{ZZ} and is thus linear in r. However, if we consider the augmented instrument cross-product matrix $T^{-1} \sum_{t=1}^{[Tr]} \tilde{z}_t(\pi^0) \tilde{z}_t(\pi^0)'$ then the limit of this matrix cannot be linear in r. In fact, if Assumption 12 holds and $\pi_{i-1}^0 < r < \pi_i^0$ for some i then

$$\begin{aligned} T^{-1} \sum_{t=1}^{[Tr]} \tilde{z}_t(\pi^0) \tilde{z}_t(\pi^0)' &\xrightarrow{p} (\pi_1^0, \pi_2^0 - \pi_1^0, \ldots, \pi_{i-1}^0 - \pi_{i-2}^0, r - \pi_{i-1}^0, 0_{1\times(h+1-i)}) \otimes Q_{ZZ} \\ &\neq rM, \text{ for some matrix } M \end{aligned}$$

A similar problem arises with the long run variance matrix $lim_{T\to\infty} Var\left[T^{-1/2}\sum_{t=1}^{[Tr]}\tilde{h}_t\right]$.[17]

However, it is possible to develop fixed break point tests within this setting and in this sub-section we show that such tests can be combined with those derived for the stable reduced form case to produce a method for estimation of m. This method turns out to be quite simple and thus has an appeal for practitioners. We first outline the method for estimation of m and then present the necessary fixed break point test statistic.

Methodology for estimation of m

1. Estimate reduced form and test for multiple changes in parameters using, for example, the methods in Bai and Perron (1998).

2.(a) If the reduced form is judged stable then use the methodology described in Section 4 (iii) to estimate m.

2.(b) If the reduced form is unstable then estimate h using, for example, the methods in Bai and Perron (1998). Let $\hat{h}$ be the number of breaks, and collect the estimates into the $\hat{h} \times 1$ vector $\hat{\pi}$.

   (i) Divide the sample into $\hat{h}+1$ sub-samples: $\mathcal{T}_j = \{t \in [\hat{\tau}_{j-1}+1, \ldots \hat{\tau}_j]\}$, where $\hat{\tau}_j = [\hat{\pi}_j T]$, $\hat{\pi}_0 = 0$ and $\hat{\pi}_{h+1} = 1$.

   (ii) Apply the methodology described in Section 4 (iii) to estimate the number of breaks in the structural equation for $\mathcal{T}_j$.[18] Let $\hat{m}(j)$ be the number of breaks on this segment and denote the location of these breaks by $\hat{\lambda}_i(j)$ for $i = 1, 2, \ldots \hat{m}(j)$.

   (iii) Define $\mathcal{L} = \{\hat{\lambda}_i(j);\ i = 1, 2, \ldots \hat{m}(j);\ j = 1, 2, \ldots \hat{h}\}$. Conditional on breaks in $\mathcal{L}$, test whether is a break in the structural equation at $\hat{\tau}_j$ for $j = 1, 2 \ldots \hat{h}$ individually using the test statistic $Wald_T(j)$ defined below. Define $\mathcal{L}_\pi = \{\hat{\pi}_j,$ for which $Wald_T(j)$ is significant $;j = 1, 2, \ldots \hat{h}\}$.[19]

   (iv) Estimated set of break points is $\mathcal{L} \cup \mathcal{L}_\pi$, and the estimated number of break points, $\hat{m}$, is the cardinality of $\mathcal{L} \cup \mathcal{L}_\pi$.

[17]The consequences of the nonlinearity of such limits has been explored in the context of single break point tests by Hansen (2000).

[18]In calculating the tests, the sub-sample $\mathcal{T}_j$ is treated as the entire sample and so the sample size is $\hat{\tau}_j - \hat{\tau}_{j-1}$.

[19]See the discussion following Theorem 11.

We now present the formula for $Wald_T(j)$ and its limiting distribution. Suppose we wish to test the null hypotheses that there is a break in the structural equation at $\hat{\tau}_j$ conditional on the breaks in $\mathcal{L}$. In this case, we can confine attention to the sample $t = [\hat{\lambda}_{m(j-1)}(j-1)T]+1,\ldots,[\hat{\lambda}_1(j)T]$ and employ the Wald test for a single (fixed) break at $\hat{\tau}_j$. To facilitate the exposition, we write the structural equation as:

$$\begin{aligned} y_t &= (x_t', z_{1,t}')b_1(j) + u_t, \quad \text{for } t = [\hat{\lambda}_{m(j-1)}(j-1)T]+1,\ldots\hat{\tau}_j \\ &= (x_t', z_{1,t}')b_2(j) + u_t, \quad \text{for } t = \hat{\tau}_j + 1,\ldots[\hat{\lambda}_1(j)T] \end{aligned}$$

Let $\{\hat{b}_1(j),\, \hat{b}_2(j)\}$ be the 2SLS estimators of $\{b_1(j),\, b_2(j)\}$; then, the appropriate Wald statistic is

$$Wald_T(j) = T\left\{\hat{b}_1(j) - \hat{b}_2(j)\right\}' \{\bar{V}(j)\}^{-1}\left\{\hat{b}_1(j) - \hat{b}_2(j)\right\} \tag{24}$$

where

$$\begin{aligned} \bar{V}(j) &= \bar{V}_1(j) + \bar{V}_2(j), \qquad \bar{V}_k(j) = \bar{A}_k\left\{\bar{C}_kV\bar{C}_k' + \bar{D}_kV\bar{D}_k' + c_1\left(\bar{C}_kV\bar{D}_k' + \bar{D}_kV\bar{C}_k'\right)\right\}\bar{A}_k' \\ \bar{A}_1 &= (\Psi_j'Q_{ZZ}\Psi_j)^{-1}\Psi_j', \qquad \bar{C}_1 = (\pi_j^0 - \nu_0)^{-1/2}[I_q, b_x(j)'\otimes I_q], \qquad \nu_0 = \lambda_l(j) = plim\hat{\lambda}_{m(j-1)}(j-1) \\ \bar{D}_1 &= (\pi_j^0 - \pi_{j-1}^0)^{-1/2}[0_{q\times q}, b_x(j)'\otimes I_q], \qquad c_1 = (\pi_j^0 - \nu_0)^{1/2}(\pi_j^0 - \pi_{j-1}^0)^{-1/2} \\ \bar{A}_2 &= (\Psi_{j+1}'Q_{ZZ}\Psi_{j+1})^{-1}\Psi_{j+1}', \qquad \bar{C}_2 = (\nu_1 - \pi_j^0)^{-1/2}[I_q, b_x(j)'\otimes I_q], \qquad \nu_1 = \lambda_u(j) = plim\hat{\lambda}_1(j) \\ \bar{D}_2 &= (\pi_{j+1}^0 - \pi_j^0)^{-1/2}[0_{q\times q}, b_x(j)'\otimes I_q], \qquad c_2 = (\nu_1 - \pi_j^0)^{1/2}(\pi_{j+1}^0 - \pi_j^0)^{-1/2}, \end{aligned}$$

and $b(j) = [b_x(j)', b_{z_1}(j)']'$ is the common value of $\{\beta_i(j),\ i = 1,2\}$ under H_0.

Theorem 11 *If Assumptions 6, 8, 10, 11, 17-21 hold, y_t is generated via (7), x_t is generated via (21) and $\hat{x}_t$ is calculated via (23) then under $H_0 : b_1(j) = b_2(j)$, we have $Wald_T(j) \xrightarrow{d} \chi_p^2$.*

There may be strong reasons to suppose that a break in the reduced form is either present in the structural equation or it is not, and thus the outcome of the Wald test is sufficient to distinguish between these two states of the world. However, since the Wald test has power against other break points, it may be advisable to re-estimate the structural equation on $t = [\hat{\lambda}_{m(j-1)}(j-1)T]+1,\ldots[\hat{\lambda}_1(j)T]$ to determine the location of the break.

(iii) Finite sample performance:

We now investigate the finite sample properties of the Wald statistic and the methodology for estimation of m discussed above. Data are generated from the structural equation,

$$y_t = [1, x_t]\beta^{(i)} + u_t$$

where $i = 1$ if $t/T \leq \lambda^0$, and $i = 2$ else, and the reduced form

$$x_t \;=\; z_t'\delta^{(j)} \,+\, v_t$$

where $j = 1$ if $t/T \leq \pi^0$, and $j = 2$ else. The vector z_t is 5×1 and includes the intercept. The reduced form parameters are: $\delta^{(i)} = (-1)^{i+1}[1, d]$, for $i = 1, 2$, and d is chosen to ensure the population $R^2 = 0.5$; see footnote 5. We consider three scenarios of interest: *Case I*, no breaks in the structural but a break in the reduced form, ($\lambda^0 = 0$), $\beta^{(i)} = [1, 0.1]'$, $i = 1, 2$; $\pi^0 = 0.5$; *Case II*, a coincident break in the structural equation and the reduced form, $\lambda^0 = \pi^0 = 0.5$, $\beta^{(i)} = (-1)^{i+1}[1, 0.1]'$; *Case III*, a break in both equations but at distinct points in the sample, $\lambda^0 = 0.6$, $\pi^0 = 0.4$, $\beta^{(i)} = (-1)^{i+1}[1, 0.1]'$. All other aspects of the data generation process for the reduced form are the same as in the stable reduced form case. Experimentation revealed that Bai and Perron's (1998) methodology did such a good job of estimating and locating the break in the reduced form that we are able to treat this break as fixed in the estimation to reduce the computational burden without compromising the results. A maximum of three breaks is allowed in each sub-sample.

The results are presented in Table 9. We report results using both 5% and 1% significance levels for all tests. Overall, the methodology for estimating the number of breaks works well: if a 5% significance level is used then the true number of breaks in the structural equation is estimated at least 94% of the time; if a 1% level is used then the minimum is at least 98% of the time. In Case III where the breaks do not coincide, the methodology yields reliable estimators of the location of the break in the structural equation with 99% of the replications yielding an estimator within .03 of the true break fraction at $T = 240$ and .013 at $T = 480$. Given the basis in hypothesis testing, there is a chance of overfitting and this explains why the true value of m is not selected 100% of the time. In this design, the 1% significance level appears preferable. Further work is needed to explore the properties of the methodology in other settings. Nevertheless, these initial results are encouraging.

6 Empirical Application

In this section, we use our methods to explore the stability of the New Keynesian Phillips curve (NKPC) model for US data. Zhang, Osborn, and Kim (2008) report that the stylized version of the NKPC does not have serially uncorrelated errors, so we follow their practice and include lagged values of the change in inflation $\Delta inf_t = inf_t - inf_{t-1}$ to remove this dynamic structure from the errors. Accordingly, our analysis is based on the following NKPC version:

$$inf_t \;=\; c_0 \,+\, \alpha_f inf^e_{t+1|t} \,+\, \alpha_b inf_{t-1} \,+\, \alpha_{og} og_t \,+\, \sum_{i=1}^{3} \alpha_i \Delta inf_{t-i} \,+\, u_t \tag{25}$$

Whether in equation (25) the usual output gap measure or a real marginal cost measure should be used to study the trade-off between inflation and unemployment over the cycle is an issue at the center of a current debate.[20] Gali and Gertler (1999) attribute the usual findings of negative α_{og} to measurement error in potential output, and argue that real marginal cost better accounts for direct productivity gains on inflation. On the other hand, real marginal cost is also unobserved, and other authors, e.g. Rudd and Whelan (2005) argue that the current practice of replacing marginal cost with average unit labor cost has little theoretical foundations. In our framework, we find - for the sub-samples with enough observations - evidence of a trade-off between inflation and unemployment (to the extent that output gap reflects employment), and a measure that would more directly reflect productivity gains on inflation would only be expected to strengthen our result.

We use quarterly US data spanning 1968.3-2001.4. The span of the data is slightly longer than Zhang, Osborn, and Kim (2008) but the definitions of the variables are the same: inf_t is the annualized quarterly growth rate of the GDP deflator, og_t is obtained from the estimates of potential GDP published by the Congressional Budget Office, $inf^e_{t+1|t}$ is the Greenbook one quarter ahead forecast of inflation prepared within the Fed.[21]

Both expected inflation and output gap are endogenous, with reduced forms:

$$inf^e_{t+1|t} \;=\; z_t'\delta_1 \,+\, v_{1,t} \tag{26}$$

$$og_t \;=\; z_t'\delta_2 \,+\, v_{2,t} \tag{27}$$

where z_t contains all other explanatory variables on the righthand side of (25) along with the first lagged value of each of the short term interest rate, the unemployment rate, and the growth rate of the money aggregate M2.

[20] We thank an anonymous referee for pointing out this issue.

[21] One interesting aspect of Zhang, Osborn, and Kim's (2008) study is that they employ various different inflation forecasts in their estimation. We focus here on just one of their choices for brevity.

We first assess the stability of the reduced forms in (26)-(27) via Bai and Perron's (1998) methodology.[22] We assume that the maximum number of breaks is 5 and set $\epsilon = 0.1$. The results are reported in Table 10. First consider the reduced form for $inf^e_{t+1|t}$. There is clear evidence of parameter variation with all the sup-F statistics being significant at the 1% level. Using the sequential testing strategy, we identify two breaks: one at 1975.2 and the other at 1981.1. As a robustness check, we also use BIC to choose the break points and obtain the same estimates.[23] Now consider the reduced form for og_t. Again, there is evidence of parameter variation. The sequential strategy suggests a break at 1975.2. In contrast, BIC favours the model with no breaks. As pointed out in the sequential strategy of Section 5, for our purposes, it does not matter whether the break at 1975.2 occurs in both reduced forms or not; only the union of all breaks in the reduced forms counts.

This union is $\{1975.2, 1981.1\}$, thus there are three sub-samples, each with stable reduced forms. According to the methodology described in Section 5, we test each of the sub-samples for additional unknown breaks in the structural equation, possibly present because of other structural parts of the economy not modeled here. The outcomes of sup-F tests and sup-Wald tests - robust to heteroskedasticity - all proposed in Section 4 - are reported in Table 11. In this table, we define the BIC for a certain number of breaks m as:

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

The first two sub-samples are quite small, so we test for maximum one break in the first two sub-samples and maximum two breaks in the last. The results for all samples, coupled with BIC, suggest no further evidence of breaks. Next, we use fixed break-point tests to test whether the breaks in the reduced form coincide with those in the structural equation. The p-values for F tests and Wald tests are respectively: $0.001, 0.003$ for a break at 1975.2 and 0.000, 0.000 for a break at 1981.1, indicating that the structural equation features both breaks.

The predicted values for NKPC for the period 1981.1-2001.4[24] are as follows (standard errors

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

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

[24] The results for the first two samples are omitted because these samples are quite small in relation to the number of parameters.

in parentheses):

$$inf_t = -\underset{(0.04)}{0.23} + \underset{(0.19)}{0.60} inf^e_{t+1|t} + \underset{(0.18)}{0.22} inf_{t-1} + \underset{(0.05)}{0.06} og_t$$
$$- \underset{(0.16)}{0.20} \Delta inf_{t-1} - \underset{(0.14)}{0.20} \Delta inf_{t-2} - \underset{(0.10)}{0.22} \Delta inf_{t-3}$$

Our results suggest that the forward-looking component of inflation dominates the backward-looking component, in accordance to Zhang, Osborn, and Kim (2008). Our results also closely match Zhang, Osborn, and Kim's (2008) findings with regard to the location of first break, but we find evidence of a second break at 1981.1.[25]

7 Concluding remarks

In this paper, we propose a simple methodology for estimation and inference in linear regression models with endogenous regressors and multiple breaks. We first show that an approach based on minimizing a GMM criterion over all possible partitions does not yield, in general, consistent estimates of the break-fractions and parameters; in contrast, methods based on 2SLS do deliver consistent estimates due to a more promising construction of the minimand. The methods we propose are based on a sequential strategy in which the reduced form is first tested for breaks and if breaks are present then this information is incorporated into the estimation of the structural equation. We illustrate our methods via simulations and an empirical application to the NKPC for US. We show that the NKPC over the period of study is subject to instability, confirming findings such as in Zhang, Osborn, and Kim (2008).

An interesting aspect of our analysis is that we show the limiting distribution of various tests for structural stability is not invariant to the nature of the reduced form. Specifically, if the reduced form is stable then we show that the tests based on our 2SLS estimators have the same limiting distribution derived by Bai and Perron (1998) for the analogous tests based on OLS estimators in a linear model with exogenous regressors. However, if the reduced form is unstable then the limiting distribution is different. This highlights the importance of assessing the structural stability of the reduced form prior to analyzing the structural equation.

[25] We note that with other choices of inflation forecast series, Zhang, Osborn, and Kim (2008) find evidence of breaks at other points in the sample.

Mathematical Appendix

Appendix 1: Results involving GMM

Proof of Proposition 1

Since $W_T(\lambda)$ is deterministic, we replace it by its limit in the proof without loss of generality. Given the form of W, for $u_t(\theta) = y_t - x_t'\theta$ and $f_{t,i}(\lambda) = u_t(\theta_i(\lambda))z_t$, $(i = 1, 2)$, we have:

$$\begin{aligned}\tilde{Q}_T(\theta(\lambda); \lambda) &= E\left[T^{-2}\sum_{t,s=1}^{[\lambda T]} f_{t,1}(\lambda)'\, W_1\, f_{t,1}(\lambda)\right] + E\left[T^{-2}\sum_{t,s=[\lambda T]+1}^{T} f_{t,2}(\lambda)'\, W_2\, f_{t,2}(\lambda)\right] \\ &\equiv A_{1,T}(\lambda) + A_{2,T}(\lambda) \end{aligned} \quad (28)$$

Case 1: $\lambda = \lambda^0$. Set $T_1 = [\lambda^0 T]$ and $T_2 = T - T_1$. Using similar arguments to Han and Phillips (2006),

$$\begin{aligned}\tilde{Q}_T(\theta(\lambda^0); \lambda^0) = \frac{T_1(T_1-1)}{T^2} E_1[f_{t,1}(\lambda^0)]'\, W_1\, E_1[f_{t,1}(\lambda^0)] + \frac{1}{T^2}\sum_{t=1}^{T_1} tr\{W_1\, E_1[f_{t,1}(\lambda^0) f_{t,1}(\lambda^0)']\} \\ + \frac{T_2(T_2-1)}{T^2} E_2[f_{t,2}(\lambda^0)]'\, W_2\, E_2[f_{t,2}(\lambda^0)] + \frac{1}{T^2}\sum_{t=T_1+1}^{T} tr\{W_2\, E_2[f_{t,2}(\lambda^0) f_{t,2}(\lambda^0)']\} \end{aligned} \quad (29)$$

From (29) and Assumption 1, it follows that $\tilde{Q}_T(\theta(\lambda^0); \lambda^0) \xrightarrow{p} \tilde{Q}(\theta(\lambda^0))$, with

$$\tilde{Q}(\theta(\lambda^0)) = (\lambda^0)^2 E_1[f_{t,1}(\lambda^0)]'\, W_1\, E_1[f_{t,1}(\lambda^0)] + (1 - \lambda^0)^2 E_2[f_{t,2}(\lambda^0)]'\, W_2\, E_2[f_{t,2}(\lambda^0)] \quad (30)$$

Substituting $\theta(\lambda^0) = \theta_0(\lambda^0)$ in (30), it follows that $\tilde{Q}(\theta(\lambda^0)) = 0$.

Case 2: $\lambda < \lambda^0$. Set $T_1 = [\lambda T]$, $T_2 = [\lambda^0 T]$, and $T_* = T_2 - T_1$. Since $\lambda < \lambda^0$, we have

$$A_{1,T}(\lambda) = \frac{T_1(T_1-1)}{T^2} E_1[f_{t,1}(\lambda)]'\, W_1\, E_1[f_{t,1}(\lambda)] + \frac{1}{T^2}\sum_{t=1}^{T_1} tr\{W_1\, E_1[f_{t,1}(\lambda) f_{t,1}(\lambda)']\} \quad (31)$$

From (31) and Assumption 1, it follows that $A_{1,T}(\lambda) \rightarrow \lambda^2 E_1[f_{t,1}(\lambda)]'\, W_1\, E_1[f_{t,1}(\lambda)]$.

Now consider $A_{2,T}(\lambda)$. We have

$$\begin{aligned}A_{2,T} = E\left[T^{-2}\sum_{t=T_1}^{T_2} f_{t,2}(\lambda)'\, W_2 \sum_{t=T_1}^{T_2} f_{t,2}(\lambda)\right] + E\left[T^{-2}\sum_{t=T_2+1}^{T} f_{t,2}(\lambda)'\, W_2 \sum_{t=T_2+1}^{T} f_{t,2}(\lambda)\right] \\ + 2E\left[T^{-2}\sum_{t=T_1+1}^{T_2} f_{t,2}(\lambda)'\, W_2 \sum_{t=T_2+1}^{T} f_{t,2}(\lambda)\right] = a_{1,T} + a_{2,T} + 2a_{3,T}, \text{ respectively.} \end{aligned} \quad (32)$$

Under our assumptions we have:

$$
\begin{aligned}
a_{1,T} &\rightarrow (\lambda^0 - \lambda)^2 E_1[f_{t,2}(\lambda)'] \, W_2 \, E_1[f_{t,2}(\lambda)], && (33)\\
a_{2,T} &\rightarrow (1 - \lambda^0)^2 E_2[f_{t,2}(\lambda)'] \, W_2 \, E_2[f_{t,2}(\lambda)], && (34)\\
a_{3,T} &\rightarrow (\lambda^0 - \lambda)(1 - \lambda^0) E_1[f_{t,2}(\lambda)'] \, W_2 \, E_2[f_{t,2}(\lambda)] && (35)
\end{aligned}
$$

Combining (31)-(35), yields $\tilde{Q}_T(\theta(\lambda);\lambda) \rightarrow \tilde{Q}(\theta(\lambda);\lambda)$, where

$$
\begin{aligned}
\tilde{Q}(\theta(\lambda);\lambda) = \; & \lambda^2 E_1[f_{t,1}(\lambda)]' \, W_1 \, E_1[f_{t,1}(\lambda)] + (\lambda^0 - \lambda)^2 E_1[f_{t,2}(\lambda)'] \, W_2 \, E[f_{t,2}(\lambda)] \\
& + (1 - \lambda^0)^2 E_2[f_{t,2}(\lambda)'] \, W_2 \, E_2[f_{t,2}(\lambda)] + 2(\lambda^0 - \lambda)(1 - \lambda^0) E_1[f_{t,2}(\lambda)'] \, W_2 \, E_2[f_{t,2}(\lambda)]
\end{aligned}
$$

We now evaluate the expectations above. Since $u_t(\theta) = u_t + x_t'(\theta_0 - \theta)$, it follows that $E_i[f_{t,j}(\lambda)] = M_i(\theta_0^{(i)} - \theta_j(\lambda))$ and so,

$$
\begin{aligned}
\tilde{Q}(\theta(\lambda);\lambda) \;=\; & \lambda^2 \{\theta_0^{(1)} - \theta_1(\lambda)\}' M_1' W_1 M_1 \{\theta_0^{(1)} - \theta_1(\lambda)\} \\
& + (\lambda^0 - \lambda)^2 \{\theta_0^{(1)} - \theta_2(\lambda)\}' D_1' D_1 \{\theta_0^{(1)} - \theta_2(\lambda)\} \\
& + (1 - \lambda^0)^2 \{\theta_0^{(2)} - \theta_2(\lambda)\}' D_2' D_2 \{\theta_0^{(2)} - \theta_2(\lambda)\} \\
& + 2(\lambda^0 - \lambda)(1 - \lambda^0) \{\theta_0^{(1)} - \theta_2(\lambda)\}' D_1' D_2 \{\theta_0^{(2)} - \theta_2(\lambda)\} \qquad (36)
\end{aligned}
$$

where $D_i = C_2 M_i$, and $W_i = C_i' C_i$ (where nonsingular C_i exists via Assumption 5). Now notice that for $\theta(\lambda) = (\theta_0^{(1)'}, \theta_*^{(2)'})'$, we have

$$
\tilde{Q}(\theta(\lambda);\lambda) \;=\; \left\{ \frac{(\lambda^0 - \lambda)^2 (1 - \lambda^0)^2}{(1 - \lambda)^2} \right\} \xi' \xi
$$

where $\xi = C_2(M_1 - M_2)(\theta_0^{(1)} - \theta_0^{(2)})$. The result then follows immediately upon noting that C_2 is pd by definition.

Case 3: $\lambda > \lambda^0$. This case can be handled similarly to Case 2 and is omitted for simplicity.

Proof of Proposition 2:

Define $Z_1(\lambda) = [z_1, z_2, \ldots, z_{[\lambda T]}]'$, $Z_2(\lambda) = [z_{[\lambda T]+1}, z_{[\lambda T]+2}, \ldots, z_T]'$, $X_1(\lambda) = [x_1, x_2, \ldots, x_{[\lambda T]}]'$, $X_2(\lambda) = [x_{[\lambda T]+1}, x_{[\lambda T]+2}, \ldots, x_T]'$, $y_1(\lambda) = [y_1, y_2, \ldots, y_{[\lambda T]}]'$, $y_2(\lambda) = [y_{[\lambda T]+1}, y_{[\lambda T]+2}, \ldots, y_T]'$. Since the model is linear, it follows by similar arguments to, for example, Hall (2005)[Chap. 2.2] that

$$\begin{bmatrix} \hat{\theta}_{1,T}(\lambda) \\ \hat{\theta}_{2,T}(\lambda) \end{bmatrix} = \begin{bmatrix} H_{1,T}Z_1(\lambda)'y_1(\lambda) \\ H_{2,T}(\lambda)Z_2(\lambda)'y_2(\lambda) \end{bmatrix} \tag{37}$$

where $H_{i,T}(\lambda) = [X_i(\lambda)'Z_i(\lambda)W_{i,T}(\lambda)Z_i(\lambda)'X_i(\lambda)]^{-1}X_i(\lambda)'Z_i(\lambda)W_{i,T}(\lambda)$ for $i = 1, 2$. First consider $\hat{\theta}_{1,T}(\lambda)$. From Assumption 3, it follows that, uniformly in λ:

$$T^{-1}X_1(\lambda)'Z_1(\lambda) \xrightarrow{p} N_1(\lambda), \tag{38}$$

$$T^{-1}Z_1(\lambda)'y_1(\lambda) \xrightarrow{p} \left\{ \begin{array}{c} \lambda M_1\theta_0^{(1)}, \qquad \text{for } \lambda \leq \lambda^0 \\ \lambda^0 M_1\theta_0^{(1)} + (\lambda - \lambda^0)M_2\theta_0^{(2)}, \text{ for } \lambda > \lambda^0 \end{array} \right\}, \tag{39}$$

$$T^{-1}X_2(\lambda)'Z_2(\lambda) \xrightarrow{p} N_2(\lambda), \tag{40}$$

$$T^{-1}Z_2(\lambda)'y_2(\lambda) \xrightarrow{p} \left\{ \begin{array}{c} (\lambda^0 - \lambda)M_1\theta_0^{(1)} + (1 - \lambda^0)M_2\theta_0^{(2)}, \qquad \text{for } \lambda \leq \lambda^0 \\ (1 - \lambda)M_2\theta_0^{(2)}, \text{ for } \lambda > \lambda^0 \end{array} \right\}. \tag{41}$$

Therefore, (37)-(41) yield $\hat{\theta}_T(\lambda) \xrightarrow{p} \tilde{\theta}(\lambda) = [\tilde{\theta}_1(\lambda)', \tilde{\theta}_2(\lambda)']'$ uniformly in λ where

$$\tilde{\theta}_1(\lambda) = \theta_0^{(1)}\mathcal{I}_\lambda(\lambda^0) + \{1 - \mathcal{I}_\lambda(\lambda^0)\}\bar{\theta}_*^{(1)}(\lambda) \tag{42}$$

$$\tilde{\theta}_2(\lambda) = \bar{\theta}_*^{(2)}(\lambda)\mathcal{I}_\lambda(\lambda^0) + \{1 - \mathcal{I}_\lambda(\lambda^0)\}\theta_0^{(2)} \tag{43}$$

where $\mathcal{I}_\lambda(\lambda^0)$ is the indicator function defined in the statement of Proposition 3, and

$$\bar{\theta}_*^{(1)}(\lambda) = \{N_1(\lambda)'W_1N_1(\lambda)\}^{-1} N_1(\lambda)'W_1[\lambda^0 M_1\theta_0^{(1)} + (\lambda - \lambda^0)M_2\theta_0^{(2)}] \tag{44}$$

$$\bar{\theta}_*^{(2)}(\lambda) = \{N_2(\lambda)'W_2N_2(\lambda)\}^{-1} N_2(\lambda)'W_2[(\lambda^0 - \lambda)M_1\theta_0^{(1)} + (1 - \lambda^0)M_2\theta_0^{(2)}] \tag{45}$$

From (44)-(45), it follows that if $\theta_0^{(1)} - \theta_0^{(2)} \in \mathcal{N}(M_1 - M_2)$ then $\bar{\theta}_*^{(i)}(\lambda) = \theta_*^{(i)}(\lambda)$ for $i = 1, 2$.[26]

To prove Proposition 3, we need the following Lemma, whose proof is relegated to the Supplemental Appendix.

Lemma A.1 *If Assumptions 1-5 hold and* $\theta_0^{(1)} - \theta_0^{(2)} \in \mathcal{N}(M_1 - M_2)$*, then:*

$$\begin{bmatrix} T^{1/2}\left(\hat{\theta}_{1,T}(\lambda) - \theta_{*,1}(\lambda)\right) \\ T^{1/2}\left(\hat{\theta}_{2,T}(\lambda) - \theta_{*,2}(\lambda)\right) \end{bmatrix} \Rightarrow \begin{bmatrix} H_1(\lambda) & 0_{p\times p} \\ 0_{p\times p} & H_2(\lambda) \end{bmatrix} \begin{bmatrix} \xi_1(\lambda) \\ \xi_2(\lambda) \end{bmatrix}$$

[26]This can be verified as follows. Consider $\bar{\theta}_*^{(1)}$: for $\lambda \leq \lambda^0$, the result is trivial; for $\lambda > \lambda^0$, add and subtract the term $N_1(\lambda)\theta_*^{(1)}(\lambda)$ inside the brackets in (44) and then rearrange the terms.

where $H_i(\lambda)$ and $\xi_i(\lambda)$ are defined in Proposition 3.

Proof of Proposition 3:

Note that $J_T(\lambda) = TQ_T(\hat{\theta}_T(\lambda);\lambda) = T^{1/2}g_T(\hat{\theta}_T(\lambda);\lambda)'W_T(\lambda)T^{1/2}g_T\left(\hat{\theta}_T(\lambda);\lambda\right)$. Also,

$$g_T\left(\hat{\theta}_T(\lambda);\lambda\right) = \left[\begin{array}{c} T^{-1/2}\sum_{t=1}^{[\lambda T]} z_t\left(y_t - x_t'\hat{\theta}_{1,T}(\lambda)\right) \\ T^{-1/2}\sum_{t=[\lambda T]+1}^{T} z_t\left(y_t - x_t'\hat{\theta}_{2,T}(\lambda)\right)\end{array}\right] = \left[\begin{array}{c} c_{1,T}(\lambda) \\ c_{2,T}(\lambda)\end{array}\right]. \tag{46}$$

Noting that for $i = 1, 2$,

$$\begin{aligned} y_t - x_t'\hat{\theta}_{i,T}(\lambda) &= u_t - x_t'\left(\hat{\theta}_{i,T}(\lambda) - \theta_0^{(1)}\right), \text{ for } t/T \le \lambda^0 &(47)\\ &= u_t - x_t'\left(\hat{\theta}_{i,T}(\lambda) - \theta_0^{(2)}\right), \text{ for } t/T > \lambda^0, &(48)\end{aligned}$$

it follows that for $\lambda \le \lambda^0$,

$$c_{1,T}(\lambda) = T^{-1/2}\sum_{t=1}^{[\lambda T]} z_t u_t - T^{-1}\sum_{t=1}^{[\lambda T]} z_t x_t' T^{1/2}\left(\hat{\theta}_{1,T}(\lambda) - \theta_0^{(1)}\right), \tag{49}$$

and for $\lambda > \lambda^0$, $c_{1,T}(\lambda)$ is given by[27]

$$\begin{aligned} &T^{-1/2}\sum_{t=1}^{[\lambda T]} z_t u_t - T^{-1}\sum_{t=1}^{[\lambda^0 T]} z_t x_t' T^{1/2}\left(\hat{\theta}_{1,T}(\lambda) - \theta_0^{(1)}\right) - T^{-1}\sum_{t=[\lambda^0 T]+1}^{[\lambda T]} z_t x_t' T^{1/2}\left(\hat{\theta}_{1,T}(\lambda) - \theta_0^{(2)}\right)\\ &= T^{-1/2}\sum_{t=1}^{[\lambda T]} z_t u_t - N_1(\lambda)T^{1/2}\left(\hat{\theta}_{1,T}(\lambda) - \theta_*^{(1)}(\lambda)\right) - T^{-1/2}\sum_{t=1}^{[\lambda^0 T]}(z_t x_t' - M_1)[\hat{\theta}_{1,T}(\lambda) - \theta_0^{(1)}]\\ &- T^{-1/2}\sum_{t=[\lambda^0 T]+1}^{[\lambda T]}(z_t x_t' - M_2)[\hat{\theta}_{1,T}(\lambda) - \theta_0^{(2)}] \end{aligned} \tag{50}$$

Now consider $c_{2,T}(\lambda)$. Using (47)-(48) it follows that for $\lambda \le \lambda^0$, $c_{2,T}(\lambda)$ is given by[28]

$$\begin{aligned} &T^{-1/2}\sum_{t=[\lambda T]+1}^{T} z_t u_t - T^{-1}\sum_{t=[\lambda T]+1}^{[\lambda^0 T]} z_t x_t' T^{1/2}\left(\hat{\theta}_{2,T}(\lambda) - \theta_0^{(1)}\right)\\ &- T^{-1}\sum_{t=[\lambda^0 T]+1}^{T} z_t x_t' T^{1/2}\left(\hat{\theta}_{2,T}(\lambda) - \theta_0^{(2)}\right) = T^{-1/2}\sum_{t=[\lambda T]+1}^{T} z_t u_t - N_2(\lambda)T^{1/2}\left(\hat{\theta}_{2,T}(\lambda) - \theta_*^{(1)}\right)\\ &- T^{-1/2}\sum_{t=[\lambda^0 T]+1}^{T}(z_t x_t' - M_1)[\hat{\theta}_{2,T}(\lambda) - \theta_0^{(2)}] - T^{-1/2}\sum_{t=[\lambda^0 T]+1}^{T}(z_t x_t' - M_2)[(\hat{\theta}_{2,T}(\lambda) - \theta_0^{(2)}] \end{aligned} \tag{51}$$

[27] The equality uses: $N_1(\lambda)\theta_*^{(1)} = \lambda^0 M_1\theta_0^{(1)} + (\lambda - \lambda^0)M_2\theta_2^{(2)}$.

[28] The equality uses: $N_2(\lambda)\theta_*^{(2)} = (\lambda^0 - \lambda)M_1\theta_0^{(1)} + (1 - \lambda)M_2\theta_0^{(2)}$.

and for $\lambda > \lambda^0$

$$c_{2,T}(\lambda) \;=\; T^{-1/2} \sum_{t=[\lambda T]+1}^{T} z_t u_t \;-\; T^{-1} \sum_{t=[\lambda T]+1}^{T} z_t x_t' T^{1/2} \left(\hat{\theta}_{2,T}(\lambda) \;-\; \theta_0^{(2)} \right). \tag{52}$$

The result then follows from equations (46)-(52), Proposition 2, Lemma A.1 and Assumptions 3-5.

Appendix 2: Results involving 2SLS

We begin with an item of terminology. We say that a matrix A, say, is a diagonal partition at $(T_1, T_2, \ldots T_m)$ of the $T \times k$ matrix W whose t^{th} row is $\hat{x}_t'$ if $A = diag(W_{T_1}, ..., W_{T_{m+1}})$ and $W_{T_i} = (\hat{x}_{T_{i-1}+1}, ..., \hat{x}_{T_i})'$.[29] Also, we write (10) for the true partition (so that $\beta_i^* = \beta_i^0$) as

$$Y \;=\; \bar{W}^0 \beta^0 \;+\; \tilde{U} \tag{53}$$

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

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

$$\|A\| = \sup_{x \neq 0} \|Ax\| / \|x\| \tag{54}$$

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

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

$$\|A\| \;\leq\; (trA'A)^{1/2} \tag{55}$$

- For a projection matrix P, we have

$$\|PA\| \leq \|A\| \tag{56}$$

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

- Let $A : R_1 \to R_2$ and $B : R_2 \to R_3$ be linear operators. Then we have[30]

$$\|BA\| \le \|B\|\|A\| \tag{57}$$

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

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

Proof of Lemma 1

Part (i): Using the definition of d_t, it follows that, for $t \in [\hat{T}_{j-1}+1, \hat{T}_j]$, $\tilde{u}_t d_t = \tilde{u}_t \hat{x}_t'(\hat{\beta}_j - \beta_i^0) = \tilde{u}_t \hat{x}_t' \hat{\beta}_j - \tilde{u}_t \hat{x}_t' \beta_i^0$ and hence that

$$\sum_{t=1}^{T} \tilde{u}_t d_t = \sum_{t=1}^{T} \tilde{u}_t \hat{x}_t' \hat{\beta}(t,T) - \sum_{t=1}^{T} \tilde{u}_t \hat{x}_t' \beta^0(t,T) = \tilde{U}'\bar{W}^*\hat{\beta} - \tilde{U}'\bar{W}^0\beta^0 \tag{58}$$

where $\hat{\beta}(t,T) = \sum_{i=1}^m \hat{\beta}_j \mathcal{I}\left\{ t/T \in (\hat{\lambda}_{j-1}, \hat{\lambda}_j] \right\}$ and $\beta^0(t,T) = \sum_{i=1}^m \beta_j^0 \mathcal{I}\{ t/T \in (\lambda_{j-1}, \lambda_j] \}$. From (58), it follows that Lemma 1(i) is established if it can be shown that

$$T^{-1}(\tilde{U}'\bar{W}^*\hat{\beta} - \tilde{U}'\bar{W}^0\beta^0) = o_p(1) \tag{59}$$

Since the 2SLS estimator based on the partition $(\hat{T}_1, ..., \hat{T}_m)$ is $\hat{\beta} = (\bar{W}^{*\prime}\bar{W}^*)^{-1}\bar{W}^{*\prime}Y$, it follows that

$$\tilde{U}'\bar{W}^*\hat{\beta} - \tilde{U}'\bar{W}^0\beta^0 = \tilde{U}'P_{\bar{W}^*}\bar{W}^0\beta^0 + \tilde{U}'P_{\bar{W}^*}\tilde{U} - \tilde{U}'\bar{W}^0\beta^0 \tag{60}$$

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

We now analyze the terms on the right hand side of (60). It is most convenient to begin by analyzing $\|P_{\bar{W}^*}\tilde{U}\|$. To this end, we define $\sum_i$ as the summation over observations $t = \hat{T}_i + 1, \hat{T}_i + 2, \ldots, \hat{T}_{i+1}$. First, note $\|P_{\bar{W}^*}\tilde{U}\|^2 = \tilde{U}'P_{\bar{W}^*}\tilde{U}$ is the sum of the $m+1$ terms

$$n_{i,T} = \left(\sum_i \hat{x}_t \tilde{u}_t\right)' \left(\sum_i \hat{x}_t \hat{x}_t'\right)^{-1} \left(\sum_i \hat{x}_t \tilde{u}_t\right) \tag{61}$$

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

for $i = 0, 1, ..., m$. Using Assumptions 8 and 11, it follows that $\sum_i \hat{x}_t \tilde{u}_t = O_p(T^{1/2}) \sum_i \hat{x}_t \hat{x}_t^{'} = O_p(T)$ and hence that

$$\|P_{\bar{W}^*}\tilde{U}\|^2 \;=\; O_p(1) \tag{62}$$

Now consider the first term on the right hand side of (60). Using (57), it follows that

$$\|\tilde{U}' P_{\bar{W}^*}\bar{W}^0\beta^0\| \leq \|\tilde{U}' P_{\bar{W}^*}\| \cdot \|\bar{W}^0\beta^0\| \tag{63}$$

Since $W = P_z X$, where X is the original design matrix and $P_Z = Z(Z'Z)^{-1}Z'$ is a projection matrix, it follows from (55)-(56), (8) and Assumptions 8, 9 and 11 that

$$\|\bar{W}^0\| = \|W\| \;=\; \|P_Z X\| \;\leq\; \|X\| \;\leq\; (trX'X)^{1/2} = O_p(T^{1/2}) \tag{64}$$

and hence from (62)-(64) that

$$\|\tilde{U}' P_{\bar{W}^*}\bar{W}^0\beta^0\| \;=\; O_p(T^{1/2}) \tag{65}$$

Finally, consider the third term on the right hand side of (60), $\tilde{U}'\bar{W}^0\beta^0$. Notice that $\tilde{U}'\bar{W}^0$ consists of $m+1$ terms, $\sum_{t=T_{i-1}^0+1}^{T_i^0} \hat{x}_t\tilde{u}_t$. It can be shown that $\sum_{t=T_{i-1}^0+1}^{T_i^0} \hat{x}_t\tilde{u}_t = O_p(T^{1/2})$ and hence that

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

Combining (60), (62), (65) and (66), it follows that $\tilde{U}'\bar{W}^*\hat{\beta} - \tilde{U}'\bar{W}^0\beta^0 \;=\; O_p(T^{1/2})$ and hence that $T^{-1}(\tilde{U}'\bar{W}^*\hat{\beta} - \tilde{U}'\bar{W}^0\beta^0) = O_p(T^{-1/2}) = o_p(1)$, which is the desired result.

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

$$\sum_{t=1}^{T} d_t^2 \;\geq\; \sum_1 d_t^2 \;+\; \sum_2 d_t^2 \tag{67}$$

where

$$\sum_1 d_t^2 \;=\; \left(\hat{\beta}_k \;-\; \beta_j^0\right)' \left(\sum_1 \hat{x}_t\hat{x}_t'\right) \left(\hat{\beta}_k \;-\; \beta_j^0\right) \tag{68}$$

$$\sum_2 d_t^2 \;=\; \left(\hat{\beta}_k \;-\; \beta_{j+1}^0\right)' \left(\sum_2 \hat{x}_t\hat{x}_t'\right) \left(\hat{\beta}_k \;-\; \beta_{j+1}^0\right) \tag{69}$$

and $\sum_1$ extends over the set $\{T(\lambda_j^0 - \eta) \leq t \leq T\lambda_j^0\}$ and $\sum_2$ extends over the set $\{T\lambda_j^0 + 1 \leq t \leq T(\lambda_j^0 + \eta)\}$. At this stage, define γ_1 and γ_2 to be the smallest eigenvalue of $\sum_1 z_t z_t'$ and $\sum_2 z_t z_t'$, respectively. Then, since $\sum_i \hat{x}_t \hat{x}_t' = \hat{\Delta}_T' \left(\sum_i z_t z_t'\right) \hat{\Delta}_T$, it follows that[31]

$$\begin{aligned}
\sum_1 {d_t}^2 + \sum_2 {d_t}^2 &= \left(\hat{\Delta}_T(\hat{\beta}_k - \beta_j^0)\right)' \left(\sum_1 z_t z_t'\right) \left(\hat{\Delta}_T(\hat{\beta}_k - \beta_j^0)\right) \\
&\quad + \left(\hat{\Delta}_T(\hat{\beta}_k - \beta_{j+1}^0)\right)' \left(\sum_2 z_t z_t'\right) \left(\hat{\Delta}_T(\hat{\beta}_k - \beta_{j+1}^0)\right) \\
&\geq \gamma_1 \|\hat{\Delta}_T(\hat{\beta}_k - \beta_j^0)\|^2 + \gamma_2 \|\hat{\Delta}_T(\hat{\beta}_k - \beta_{j+1}^0)\|^2 \\
&\geq (1/2) \cdot min\{\gamma_1, \gamma_2\} \cdot \|\hat{\Delta}_T(\beta_j^0 - \beta_{j+1}^0)\|^2 \qquad (70)
\end{aligned}$$

Now consider the right hand side of (70). We have

$$\sum_1 z_t z_t' = (T\eta)(1/T\eta) \sum_{t=T(\lambda_j^0 - \eta)}^{T\lambda_j^0} z_t z_t' = (T\eta) A_T \qquad (71)$$

where $A_T = (1/T\eta) \sum_{t=T(\lambda_j^0 - \eta)}^{T\lambda_j^0} z_t z_t'$. From Assumption 10, the smallest eigenvalue of A_T is bounded away from zero. Thus, the smallest eigenvalue of $(T\eta)A_T$ is of order $T\eta$. Similarly, the smallest eigenvalue of $\sum_2 z_t z_t'$ is of order $T\eta$. Using these two order statements in (70), it follows that

$$\sum_{t=1}^{T} {d_t}^2 \geq \sum_1 {d_t}^2 + \sum_2 {d_t}^2 \geq TC \cdot \|\hat{\Delta}_T(\beta_j^0 - \beta_{j+1}^0)\|^2$$

for some $C > 0$ and hence, using $\hat{\Delta}_T \xrightarrow{p} \Delta_0$, that

$$T^{-1} \sum_{t=1}^{T} d_t^2 \geq C \|\Delta_0(\beta_j^0 - \beta_{j+1}^0)\|^2 + \xi_T \qquad (72)$$

where $\xi_T = C\left\{ \|\hat{\Delta}_T(\beta_j^0 - \beta_{j+1}^0)\|^2 - \|\Delta_0(\beta_j^0 - \beta_{j+1}^0)\|^2 \right\} = o_p(1)$. The desired result then follows from (72) upon recalling that the analysis is premised on an event that occurs with probability ϵ.

Proof of Theorem 1:

Suppose that $\hat{\lambda}_j \overset{p}{\nrightarrow} \lambda_j^0$ for some j in probability. In this case, it follows from (14) and Lemma 1 that

$$(1/T) \sum_{t=1}^{T} \hat{u}_t^2 = (1/T) \sum_{t=1}^{T} \tilde{u}_t^2 + C \cdot \|\Delta_0(\beta_j^0 - \beta_{j+1}^0)\|^2 + o_p(1) \qquad (73)$$

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

with probability at least as large as $\bar{\epsilon} > 0$. Assumption 9 states that Δ_0 is full rank and so $\|\Delta_0(\beta_j^0 - \beta_{j+1}^0)\|^2 > 0$. Therefore, (73) conflicts with (13) which must hold for all T with probability one. Therefore, it must follow $\hat{\lambda}_j \xrightarrow{p} \lambda_j^0$ for all j.

Proof of Theorem 2:

The general proof strategy is the same as the one employed in Bai and Perron's (1998) proof of their Proposition 2, although the specific details are naturally different. Following Bai and Perron (1998), we assume (without loss of generality) that there are only 3 break points, that is $m = 3$. Here we present the proof for the middle break fraction, $\hat{\lambda}_2$. The proof for the end break fractions, $\hat{\lambda}_1$ and $\hat{\lambda}_3$, follows along similar lines and is omitted for brevity.[32]

The desired result can be established if it can be shown that for each $\eta > 0$, there exists $C > 0$ and $\epsilon > 0$ such that for large T,

$$P(min\{[S_T(T_1, T_2, T_3) - S_T(T_1, T_2^0, T_3)]/(T_2^0 - T_2)\} < 0) < \eta \tag{74}$$

where the minimum is taken over the set $V_\epsilon(C) = \{(T_1, T_2, T_3) : |T_i - T_i^0| \leq \epsilon T,\ i = 1, 2, 3 \text{ but } T_2 - T_2^0 < -C\}$. Define $SSR_1 = S_T(T_1, T_2, T_3)$, $SSR_2 = S_T(T_1, T_2^0, T_3)$ and $SSR_3 = S_T(T_1, T_2, T_2^0, T_3)$. Using these definitions, we have

$$S_T(T_1, T_2, T_3) - S_T(T_1, T_2^0, T_3) = (SSR_1 - SSR_3) - (SSR_2 - SSR_3) \tag{75}$$

To analyze the terms on the right hand side of (75), it is useful to define the 2SLS estimators in the four break model and emphasize the sub-samples upon which certain of these estimators are based. Let $(\hat{\beta}_1^*, \hat{\beta}_2^*, \hat{\beta}_\triangle, \hat{\beta}_3^*, \hat{\beta}_4^*)$ denote the 2SLS estimators of the regression coefficients in the five regimes of the four break model associated with the the partition (T_1, T_2, T_2^0, T_3). Note that $\hat{\beta}_2^*$ is based on observations $T_1 + 1, \ldots, T_2$; $\hat{\beta}_\triangle$ is based on observations $T_2 + 1, \ldots, T_2^0$; $\hat{\beta}_3^*$ is based on observations $T_2^0 + 1, \ldots, T_3$. Now define $\bar{W}$ to be the diagonal partition of W at (T_1, T_2, T_3), $\tilde{W}$ is the diagonal partition of W at (T_1, T_2^0, T_3), $W_\triangle = (0_{p \times T_2}, \hat{x}_{T_2+1}, ..., \hat{x}_{T_2^0}, 0_{p \times (T - T_2^0)})'$ and $M_{\bar{W}} = I_T - \bar{W}(\bar{W}'\bar{W})^{-1}\bar{W}'$.

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

It can be shown that[33]

$$SSR_1 - SSR_3 = (\hat{\beta}_3^* - \hat{\beta}_\triangle)' W_\triangle' M_{\bar{W}} W_\triangle (\hat{\beta}_3^* - \hat{\beta}_\triangle) \tag{76}$$

$$SSR_2 - SSR_3 = (\hat{\beta}_2^* - \hat{\beta}_\triangle)' W_\triangle' M_{\tilde{W}} W_\triangle (\hat{\beta}_2^* - \hat{\beta}_\triangle) \tag{77}$$

From (76)-(77) and $W_\triangle' M_{\tilde{W}} W_\triangle \leq W_\triangle' W_\triangle$, it follows that

$$SSR_1 - SSR_2 \geq (\hat{\beta}_3^* - \hat{\beta}_\triangle)' W_\triangle' M_{\bar{W}} W_\triangle (\hat{\beta}_3^* - \hat{\beta}_\triangle) - (\hat{\beta}_2^* - \hat{\beta}_\triangle)' W_\triangle' W_\triangle (\hat{\beta}_2^* - \hat{\beta}_\triangle) \tag{78}$$

Substituting for $M_{\bar{W}}$ in (78) and dividing both sides by $T_2^0 - T_2$, we obtain

$$\frac{SSR_1 - SSR_2}{T_2^0 - T_2} \geq N_1 - N_2 - N_3 \tag{79}$$

where

$$N_1 = (\hat{\beta}_3^* - \hat{\beta}_\triangle)'[(T_2^0 - T_2)^{-1} W_\triangle' W_\triangle](\hat{\beta}_3^* - \hat{\beta}_\triangle) \tag{80}$$

$$N_2 = (\hat{\beta}_3^* - \hat{\beta}_\triangle)'[(T_2^0 - T_2)^{-1} W_\triangle' \bar{W}][T^{-1}\bar{W}'\bar{W}]^{-1}[T^{-1}\bar{W}' W_\triangle](\hat{\beta}_3^* - \hat{\beta}_\triangle) \tag{81}$$

$$N_3 = (\hat{\beta}_2^* - \hat{\beta}_\triangle)'[(T_2^0 - T_2)^{-1} W_\triangle' W_\triangle](\hat{\beta}_2^* - \hat{\beta}_\triangle) \tag{82}$$

It can be shown that under our assumptions N_1 is the dominant term and, as a consequence, that $[(SSR_1 - SSR_2)/(T_2^0 - T_2)] > 0$ over $V_\epsilon(C)$ with large probability which proves (74).

Proof of Theorem 3:

For notational brevity, set $\hat{\beta} = \hat{\beta}(\{\hat{T}_i\}_{i=1}^m)$. It can be shown that

$$T^{1/2}(\hat{\beta} - \beta^0) = \left(T^{-1}\bar{W}^{*'}\bar{W}^*\right)^{-1} T^{-1/2}\bar{W}^{*'}[\tilde{U} + (\bar{W}^0 - \bar{W}^*)\beta^0] \tag{83}$$

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

$$T^{1/2}(\hat{\beta} - \beta^0) = \left(T^{-1}\bar{W}^{*'}\bar{W}^*\right)^{-1} T^{-1/2}\bar{W}^{*'}\tilde{U} + o_p(1) \tag{84}$$

The addition and subtraction of $\left(T^{-1}\bar{W}^{0'}\bar{W}^0\right)^{-1} T^{-1/2}\bar{W}^{0'}\tilde{U}$ to the right hand side of (84) and some rearrangement yields

$$\begin{aligned} T^{1/2}(\hat{\beta} - \beta^0) &= \left(T^{-1}\bar{W}^{0'}\bar{W}^0\right)^{-1} T^{-1/2}\bar{W}^{0'}\tilde{U} \\ &+ \left(T^{-1}\bar{W}^{0'}\bar{W}^0\right)^{-1}\left(T^{-1}\bar{W}^{0'}\bar{W}^0 - T^{-1}\bar{W}^{*'}\bar{W}^*\right)\left(T^{-1}\bar{W}^{*'}\bar{W}^*\right)^{-1} T^{-1/2}\bar{W}^{0'}\tilde{U} \\ &+ \left(T^{-1}\bar{W}^{*'}\bar{W}^*\right)^{-1} T^{-1/2}(\bar{W}^{*'} - \bar{W}^{0'})\tilde{U} + o_p(1) \end{aligned} \tag{85}$$

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

Since $\hat{T}_i - T_i^0 = O_p(1)$ for all i, it follows from (85) using Assumptions 9 and 11 that

$$T^{1/2}(\hat{\beta} - \beta^0) \; = \; \left(T^{-1}\bar{W}^{0'}\bar{W}^0\right)^{-1} T^{-1/2}\bar{W}^{0'}\tilde{U} \; + \; o_p(1) \tag{86}$$

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

$$T^{1/2}\left(\hat{\beta}_i - \beta_i^0\right) \; = \; \left(\frac{1}{T}\sum_{i_0}\hat{x}_t\hat{x}_t'\right)^{-1} T^{-1/2}\sum_{i_0}\hat{x}_t\tilde{u}_t + o_p(1) \tag{87}$$

The result then follows from (87) under our assumptions.

Proof of Theorem 4:

The F-statistic can be written as

$$F_T(\lambda_1, ..., \lambda_k; p) \; = \; F_T^*/[kp\,(T-(k+1)p)^{-1}SSR_k] \tag{88}$$

where $F_T^* = SSR_0 - SSR_k$. We first consider the limiting behaviour of F_T^*. To this end, we define $D^R(i,j)$ to be the sum of squared residuals from the restricted model using observations from i to j, that is, from $T_{i-1}+1$ to T_j, and $D^U(i,j)$ to be the corresponding sum of squared residuals for the unrestricted model. Using this notation, we can write F_T^* as follows:[34]

$$\begin{aligned} F_T^* &= D^R(1,k+1) - \sum_{i=1}^{k+1} D^U(i,i) \; = \; \sum_{i=1}^{k}[D^R(1,i+1) - D^R(1,i) - D^U(i+1,i+1)] && (89)\\ &= \sum_{i=1}^{k} F_{T,i}, \text{ say.} && (90) \end{aligned}$$

It can be shown that

$$\begin{aligned} F_{T,i} &= ||(I-P_{W_{1,i+1}})\tilde{U}_{1,i+1}||^2 \; - \; ||(I-P_{W_{1,i}})\tilde{U}_{1,i}||^2 - ||(I-P_{W_{i+1}})\tilde{U}_{i+1}||^2 \\ &= -S_{i+1}'H_{i+1}^{-1}S_{i+1} + S_i'H_i^{-1}S_i + A_i && (91) \end{aligned}$$

where $S_j = W_{1,j}'\tilde{U}_{1,j}$, $H_j = W_{1,j}'W_{1,j}$ and $A_i = (S_{i+1}-S_i)'(H_{i+1}-H_i)^{-1}(S_{i+1}-S_i)$.

Assumptions 8, 12 and 13 together ensure that the following version of the uniform version of the multivariate functional central limit theorem (FCLT) in Wooldridge and White (1988) holds, that is $T^{-1/2}\sum_{t=1}^{[Tr]} h_t \implies (\Omega^{1/2}\otimes Q_{ZZ}^{1/2})B_n(r)$ where $B_n(r)$ is a $n\times 1$ standard Brownian motion with $n = q\times(p_1+1)$. To explore the implications of this distributional result further, let $B(r) = [B_1(r)', B_2(r)', \ldots, B_{p+1}(r)']'$ where $B_i(r)'$ is $q\times 1$, and $\Omega^{1/2} = [N_1, \, N_2]'$ where N_1'

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

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

$$\Omega \ = \ \left[\begin{array}{cc} N_1^{'}N_1 & N_1^{'}N_2 \\ N_2^{'}N_1 & N_2^{'}N_2 \end{array} \right] = \left[\begin{array}{cc} \sigma^2 & \gamma^{'} \\ \gamma & \Sigma \end{array} \right] \tag{92}$$

where the second and third matrices are partitioned conformably. It follows from the FCLT above that $T^{-1/2}\sum_{t=1}^{[Tr]} z_t u_t \implies (N_1^{'} \otimes Q_{ZZ}^{1/2})B(r) \ = \ Q_{ZZ}^{1/2}\tilde{D}^*(r)$, say and $T^{-1/2}\sum_{t=1}^{[Tr]} z_t v_t^{'} \implies Q_{ZZ}^{1/2}B^{mat}(r)N_2 \ = \ Q_{ZZ}^{1'2}D^*(r)$, say where $vec(B^{mat}(r)) \ = \ B(r)$. Further note $(\Delta_0{}'Q_{ZZ}^{1/2})'\times (\Delta_0{}'Q_{ZZ}\Delta_0)^{-1}(\Delta_0{}'Q_{ZZ}^{1/2}) = C'\Lambda C$ where C is an orthogonal matrix and Λ is a diagonal matrix, whose first p diagonal elements are one and the remaining $q-p$ equal to zero. Using these definitions, it can be shown that

$$\begin{aligned} S_{i+1}'H_{i+1}^{-1}S_{i+1} \quad &\implies \quad \lambda_{i+1}^{-1}(\Lambda C\tilde{D}^*(\lambda_{i+1}) + \Lambda C[D^*(\lambda_{i+1}) - \lambda_{i+1}D^*(1)]\bar{\beta}_0)' \\ &\quad\quad \times(\Lambda C\tilde{D}^*(\lambda_{i+1}) + \Lambda C[D^*(\lambda_{i+1}) - \lambda_{i+1}D^*(1)]\bar{\beta}_0) \\ A_i \quad &\implies \quad (\lambda_{i+1} - \lambda_i)^{-1}[\Lambda C(\tilde{D}^*(\lambda_{i+1}) - \tilde{D}^*(\lambda_i)) + \Lambda C(D^*(\lambda_{i+1}) - D^*(\lambda_i) \\ &\quad\quad -\lambda_{i+1}D^*(1) + \lambda_i D^*(1))\bar{\beta}_0]'[\Lambda C(\tilde{D}^*(\lambda_{i+1}) - \tilde{D}^*(\lambda_i)) \\ &\quad\quad +\Lambda C(D^*(\lambda_{i+1}) - D^*(\lambda_i) - \lambda_{i+1}D^*(1) + \ \lambda_i D^*(1))\bar{\beta}_0] \end{aligned}$$

Now define $D_i = \Lambda C D^*(\lambda_i)$, $\tilde{D}_i = \Lambda C\tilde{D}^*(\lambda_i)$ and $D_1 = \Lambda C D^*(1)$. Then it can be shown that

$$F_{T,i} \implies \{\lambda_i\lambda_{i+1}(\lambda_{i+1} - \lambda_i)\}^{-1}||[\lambda_{i+1}\tilde{D}_i - \lambda_i\tilde{D}_{i+1}] + [\lambda_{i+1}D_i - \lambda_i D_{i+1}]\bar{\beta}_0||^2 \tag{93}$$

It can be shown that $(T-(k+1)p)^{-1}SSR_k \xrightarrow{p} \sigma^2 + 2\gamma^{'}\bar{\beta}_0 + \bar{\beta}_0^{'}\Sigma\bar{\beta}_0$. The desired result then follows after some additional manipulations.

Proof of Theorem 5:

Consider first

$$\tilde{F}_T(i;l) \ = \ \frac{SSR_l(\hat{T}_1,...,\hat{T}_l) - \inf_{\tau\in\Lambda_{i,\eta}} SSR_{l+1}(\hat{T}_1,...,\hat{T}_{i-1},\tau,\hat{T}_i,...,\hat{T}_l)\}}{\hat{\sigma}_i^2} \tag{94}$$

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

$$\tilde{F}_T(i;l) = \sup_{\tau\in\Lambda_{i,\eta}} \frac{\{S_T(\hat{T}_{i-1}+1,\hat{T}_i) - S_T(\hat{T}_{i-1}+1,\tau) - S_T(\tau+1,\hat{T}_i)\}}{\hat{\sigma}_i^2} \tag{95}$$

Under our assumptions, it can be shown that $\hat{\sigma}_i^2 \xrightarrow{p} \sigma_i^2 + 2\gamma_i'\beta_i^0 + \beta_i^{0\prime}\Sigma_i\beta_i^0$. Using the latter and Theorem 2, it follows that

$$\tilde{F}_T(i;l) \;=\; \sup_{\tau\in\Lambda_{i,\eta}^0}\left\{\frac{S_T(T_{i-1}^0+1,T_i^0) - S_T(T_{i-1}^0+1,\tau) - S_T(\tau+1,T_i^0)}{\sigma^2 + 2\gamma'\beta_i^0 + \beta_i^{0\prime}\Sigma\beta_i^0}\right\} + o_p(1) \qquad (96)$$

where $\Lambda_{i,\eta}^0 = \{\tau : T_{i-1}^0 + (T_i^0 - T_{i-1}^0)\eta \le \tau \le T_i^0 - (T_i^0 - T_{i-1}^0)\eta\}$. After some manipulations, it can be shown that $\tilde{F}_T(i;l) \implies \sup_{\eta\le\mu<1-\eta}\|W(\mu) - \mu W(1)\|^2 / \mu(1-\mu)$, and then the result follows.

Proof of Theorem 6:

We start by considering the limiting behaviour of the constituents of

$$Wald_T \;=\; T\hat{\beta}(\bar{T}_k)'R_k'[R_k\hat{V}_W(\bar{T}_k)R_k']^{-1}R_k\hat{\beta}(\bar{T}_k)$$

It can be shown that

$$\begin{aligned}T^{1/2}(\hat{\beta}_i - \bar{\beta}_0) \;&=\; \{(\lambda_i - \lambda_{i-1})\Delta_0'Q_{ZZ}\Delta_0\}^{-1}\Delta_0'T^{-1/2}\sum_i z_t(u_t + v_t'\bar{\beta}_0)\\ &\quad - (\Delta_0'Q_{ZZ}\Delta_0)^{-1}\Delta_0'T^{-1/2}\sum_{t=1}^{T} z_t v_t'\bar{\beta}_0 + o_p(1) \qquad (97)\end{aligned}$$

where $\bar{\beta}_0$ is the common value of $\{\beta_i^0;\ i = 1, 2, \ldots k+1\}$ under H_0. Assumptions 8 and 16 together ensure that the following version of the uniform version of the multivariate functional central limit theorem in Wooldridge and White (1988) holds, that is $T^{-1/2}\sum_{t=1}^{[Tr]} h_t \implies V^{1/2}B_n(r)$ where $B_n(r)$ is a $n\times 1$ standard Brownian motion with $n = q\times(p_1+1)$. Partition $V^{1/2} = [\tilde{N}_1, \tilde{N}_2]'$ where $\tilde{N}_1$ is a $(p+1)\times 1$ vector, and $\tilde{N}_2'$ is $(p+1)\times p$. It then follows that

$$\begin{aligned}T^{1/2}(\hat{\beta}_i - \bar{\beta}_0) \;\Rightarrow\; & (\Delta_0'Q_{ZZ}\Delta_0)^{-1}\Big\{(\lambda_i - \lambda_{i-1})^{-1}\Delta_0'[\tilde{N}_1' + (\bar{\beta}_0'\otimes I_q)\tilde{N}_2'][B_n(\lambda_i) - B_n(\lambda_{i-1})]\\ & - \Delta_0'(\bar{\beta}_0'\otimes I_q)\tilde{N}_2']B_n(1)\Big\}\end{aligned}$$

and hence,

$$\begin{aligned}T^{1/2}(\hat{\beta}_{i+1} - \hat{\beta}_i) \;\Rightarrow\; & (\Delta_0'Q_{ZZ}\Delta_0)^{-1}\Big\{(\lambda_{i+1} - \lambda_i)^{-1}\Delta_0'[\tilde{N}_1' + (\bar{\beta}_0'\otimes I_q)\tilde{N}_2'][B_n(\lambda_{i+1}) - B_n(\lambda_i)]\\ & - (\lambda_i - \lambda_{i-1})^{-1}\Delta_0'[\tilde{N}_1' + (\bar{\beta}_0'\otimes I_q)\tilde{N}_2'][B_n(\lambda_i) - B_n(\lambda_{i-1})]\Big\} \qquad (98)\end{aligned}$$

From (98), it follows that

$$T^{1/2}R_m\hat{\beta} \Rightarrow (\tilde{R}_m\otimes I_p)\{C_m^{-1}\otimes(\Delta_0'Q_{ZZ}\Delta_0)^{-1}A\}\bar{B} \qquad (99)$$

where $A = \Delta_0'[I_q, \bar{\beta}_0' \otimes I_q][\tilde{N}_1, \tilde{N}_2]'$, $C_m = diag[\lambda_1, \lambda_2 - \lambda_1, \ldots, \lambda_m - \lambda_{m-1}, 1 - \lambda_m]$, $\bar{B} = [B(\lambda_1)', \{B(\lambda_2) - B(\lambda_1)\}', \ldots, \{B(\lambda_m) - B(\lambda_{m-1})\}', \{B(1) - B(\lambda_m)\}']'$. Under our conditions, we have

$$\hat{V}_W(i) \xrightarrow{p} (\lambda_i - \lambda_{i-1})^{-1}(\Delta_0' Q_{ZZ} \Delta_0)^{-1} H (\Delta_0' Q_{ZZ} \Delta_0)^{-1} \tag{100}$$

and so

$$\hat{V}_W \xrightarrow{p} (\tilde{R}_m \otimes I_p)\{C_m^{-1} \otimes (\Delta_0' Q_{ZZ} \Delta_0)^{-1} H (\delta_0' Q_{ZZ} \Delta_0)^{-1}\}(\tilde{R}_m' \otimes I_p) \tag{101}$$

If we write $\tilde{A} = A'(\Delta_0' Q_{ZZ} \Delta_0)^{-1}$ then it follows from (99), (101) and $H = AA'$ that

$$\begin{aligned} Wald_T &\Rightarrow \bar{B}'\{C_m^{-1}\tilde{R}_m'[\tilde{R}_m C_m^{-1} \tilde{R}_m']^{-1} \tilde{R}_m C_m^{-1} \otimes \tilde{A}(\tilde{A}'\tilde{A})^{-1}\tilde{A}'\}\bar{B} \\ &\sim \bar{B}'\{C_m^{-1}\tilde{R}_m'[\tilde{R}_m C_m^{-1} \tilde{R}_m']^{-1} \tilde{R}_m C_m^{-1} \otimes I_p\}\bar{B} \end{aligned} \tag{102}$$

The result then follows from the Continuous Mapping Theorem and (102).

Proof of Theorem 7

Define $\tau = T_{i-1}^0 + \mu(T_i^0 - T_{i-1}^0)$. Using similar arguments to the proof of Theorem 6 and imposing the null hypothesis, we obtain

$$T^{1/2} R_1 \left(\hat{\beta}_1(\tau; i) - \hat{\beta}_2(\tau; i)\right) \Rightarrow [\mu(1-\mu)]^{-1} \, [\Delta_0' Q_{ZZ}^{(i)} \Delta_0]^{-1} A_i [B(\mu) - \mu B(1)]$$

where $A_i = \Delta_0'[I_q, \beta_i^{0\prime} \otimes I_q][\tilde{N}_1, \tilde{N}_2]'$, and $\hat{V}_W(\tau; i) \xrightarrow{p} [\mu(1-\mu)]^{-1} \, [\Delta_0' Q_{ZZ}^{(i)} \Delta_0]^{-1} A_i A_i' [\Delta_0' Q_{ZZ}^{(i)} \Delta_0]^{-1}$. The result then follows by similar arguments to the proof of Theorem 5.

Proof of Theorem 8:

The proof follows similar lines to Theorem 1. We first state the analogs to Lemma 1 (a)-(b), the proof of which can be found in the Supplemental Appendix, and then use them to deduce the desired result.

Lemma A.1 *Under the conditions of Theorem 8, we have*

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

(b1) If $\hat{\lambda}_j \not\xrightarrow{p} \lambda_j^0$ *for some* j, *and* $\lambda_j^0 \in (\pi_i^0, \pi_{i+1}^0)$, *then*

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

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

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

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

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

We now use this result to prove Theorem 8. Suppose that $\hat{\lambda}_j \overset{p}{\nrightarrow} \lambda_j^0$ for some j. In this case it follows from (14) and Lemma A.1 that with probability $\bar{\epsilon} > 0$:

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

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

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

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

Thus, we have

- Case 1: Assumption 20 and $\beta_j^0 \neq \beta_{j+1}^0$ implies $\|\Delta_0^{(i+1)}(\beta_j^0 - \beta_{j+1}^0)\|^2 > 0$, which gives the result as in the proof of Theorem 1.
- Case 2: Now as $\beta_j^0 \neq \beta_{j+1}^0$ and $\Delta_0^{(i)}$, $\Delta_0^{(i+1)}$ are rank p from Assumption 20, it must follow that $\|\Delta_0^{(i)}(\hat{\beta}_k - \beta_j^0)\|^2 + \|\Delta_0^{(i+1)}(\hat{\beta}_k - \beta_{j+1}^0)\|^2 > 0$ with probability one, which gives the result via the same argument as in Theorem 1.

Proof of Theorem 9:

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

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

$$P(min\{[S_T(T_1,T_2,T_3) - S_T(T_1,T_2^0,T_3)]/(T_2^0 - T_2)\} \; < \; 0) < \eta \tag{103}$$

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

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

$$\frac{SSR_1 - SSR_2}{T_2^0 - T_2} \; \geq \; 2^{-1}(\beta_3^0 - \beta_2^0)' \, [W_\triangle' W_\triangle/(T_2^0 - T_2)] \, (\beta_3^0 - \beta_2^0) - \epsilon O_p(1) - \rho O_p(1) \tag{104}$$

with large probability. It can be shown that the first term on the right hand side of (104) dominates, and that

$$(T_2^0 - T_2)^{-1} W_\triangle' W_\triangle \; \geq \; \min\{\alpha_1\gamma_1, \, \alpha_2\gamma_2\} \left(\|\Delta_0^{(i)}\|^2 \, + \, \|\Delta_0^{(i+1)}\|^2 \right) \, + \, o_p(1) \tag{105}$$

where γ_1 and γ_2 are the smallest eigenvalues of $(T_i^* - T_2)^{-1} \sum_1 z_t z_t'$ and $(T_2^0 - T_i^*)^{-1} \sum_2 z_t z_t'$, respectively, and $\alpha_1 = (T_i^* - T_2)/(T_2^0 - T_2)$, $\alpha_2 = (T_2^0 - T_i^*)/(T_2^0 - T_2)$. From Assumptions 20 and 21, it follows that the first term on the right hand side of (105) is bounded away from zero on $V_\epsilon(C)$ with large probability. Therefore, the first term on the right hand side of (104) dominates and is positive for large C, small ϵ and large T which in turn proves (103).

Proof of Theorem 10

It can be shown that

$$T^{1/2}\left(\hat{\beta}_i - \beta_i^0\right) \; = \; \left(T^{-1} \sum_{t=[\lambda_{i-1}^0 T]+1}^{[\lambda_i^0 T]} \hat{x}_t(\pi^0)\hat{x}_t(\pi^0)' \right)^{-1} T^{-1/2} \sum_{t=[\lambda_{i-1}^0 T]+1}^{[\lambda_i^0 T]} \hat{x}_t(\pi^0)\tilde{u}_t(\pi^0) \; + \; o_p(1)$$

and

$$T^{-1} \sum_{t=[\lambda_{i-1}^0 T]+1}^{[\lambda_i^0 T]} \hat{x}_t(\pi^0)\hat{x}_t(\pi^0)' \;\; \xrightarrow{p} \;\; \tilde{\Psi}'\tilde{Q}_i\tilde{\Psi}$$

$$T^{-1/2} \sum_{t=[\lambda_{i-1}^0 T]+1}^{[\lambda_i^0 T]} \hat{x}_t(\pi^0)\tilde{u}_t(\pi^0) \;\; = \;\; \tilde{\Psi}'\tilde{C}_i T^{-1/2} \sum_{t=[\lambda_{i-1}^0 T]+1}^{[\lambda_i^0 T]} \tilde{h}_t \; - \; \tilde{Q}_i\tilde{Q}_{ZZ}(1)^{-1}\tilde{D}_i \sum_{t=1}^{T} \tilde{h}_t \; + \; o_p(1)$$

The result then follows after some manipulations.

Proof of Theorem 11

Using similar arguments to the proof of Theorem 10, it follows that under H_0 we have

$$\begin{aligned} T^{1/2}[\hat{b}_1(j) \,-\, b(j)] &= \bar{A}_1\{(\pi_j^0 - \nu_0)^{-1}T^{-1/2}\sum_0 z_t[u_t + v_t' b_x(j)] \\ &\quad - (\pi_j^0 - \pi_{j-1}^0)^{-1}T^{-1/2}\sum_{r_0} z_t v_t' b_x(j) \,+\, o_p(1) \\ T^{1/2}[\hat{b}_2(j) \,-\, b(j)] &= \bar{A}_2\{(\nu_1 - \pi_j^0)^{-1}T^{-1/2}\sum_1 z_t[u_t + v_t' b_x(j)] \\ &\quad - (\pi_{j+1}^0 - \pi_j^0)^{-1}T^{-1/2}\sum_{r_1} z_t v_t' b_x(j) \,+\, o_p(1) \end{aligned}$$

where $\sum_0$ denotes summation over $t = [\lambda_l(j)T] + 1, \ldots \tau_j$, $\sum_1$ denotes summation over $t = [\pi_j^0 T] + 1, \ldots [\lambda_u(j)T]$, $\sum_{r_0}$ denotes summation over $t = [\pi_{j-1}^0 T] + 1, \ldots [\pi_j^0 T]$, and $\sum_{r_1}$ denotes summation over $t = [\pi_j^0 T] + 1, \ldots [\pi_{j+1}^0 T]$.

References

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

Andrews, D. W. K. (1991). 'Heteroscedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica*, 59: 817–858.

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

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

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

Bai, J., Chen, H., Chong, T., and Wang, S. (2008). 'Generic Consistency of the Break-Point Estimators under Specification Errors in a Multiple-Break Model', *Econometrics Journal*, 11: 287–307.

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

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

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

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

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

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

Hall, A. R., Inoue, A., and Peixe, F. P. M. (2003). 'Covariance estimation and the limiting behaviour of the overidentifying restrictions test in the presence of neglected structural instability', *Econometric Theory*, 19: 962–983.

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, C., and Phillips, P. C. B. (2006). 'GMM with many moment conditions', *Econometrica*, 74: 147–192.

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.

Hansen, B. E. (2000). 'Testing for structural change in conditional models', *Journal of Econometrics*, 97: 93–115.

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

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

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

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

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

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 chnage 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.

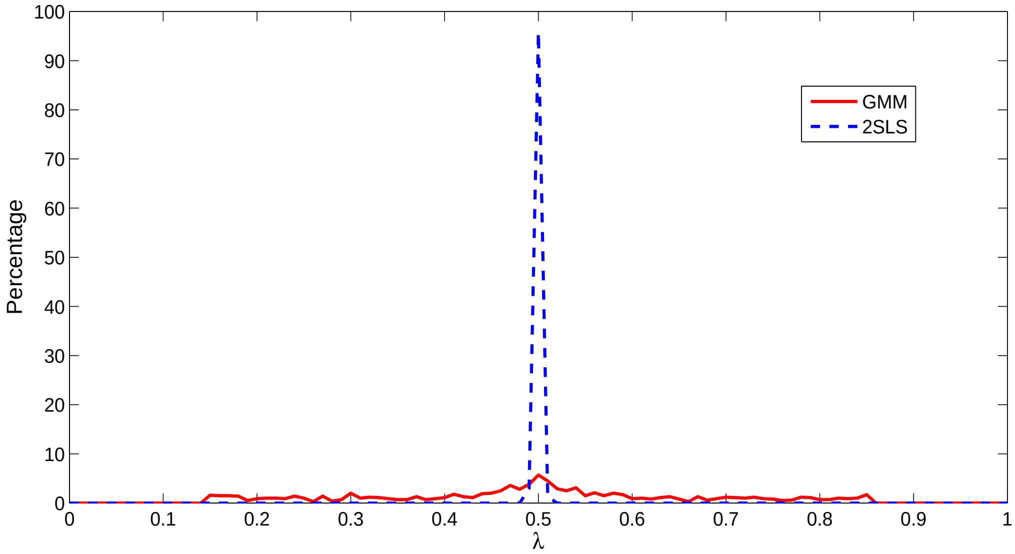

Figure 1: Distribution of estimated break fractions in the one break model

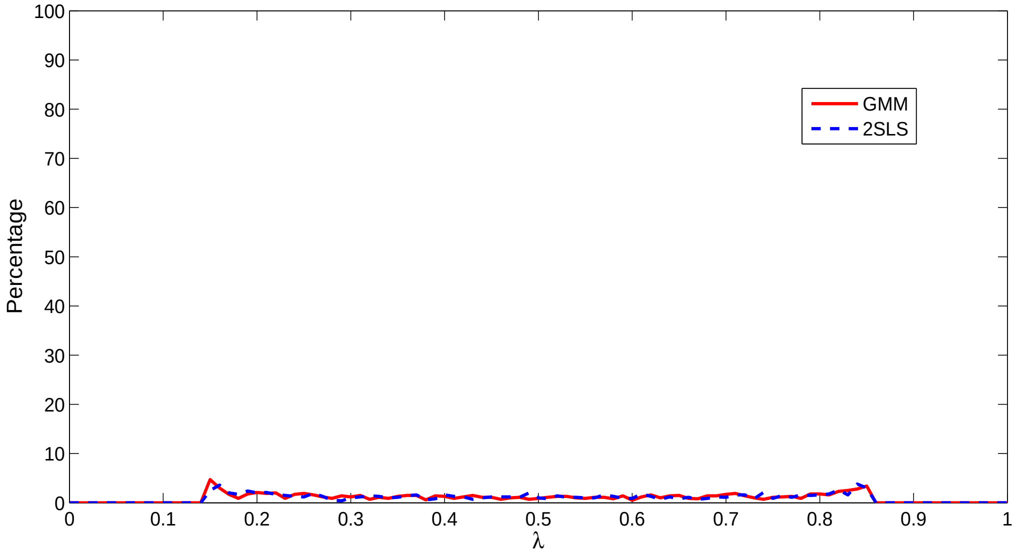

Figure 2: Distribution of estimated break fractions in the no break model

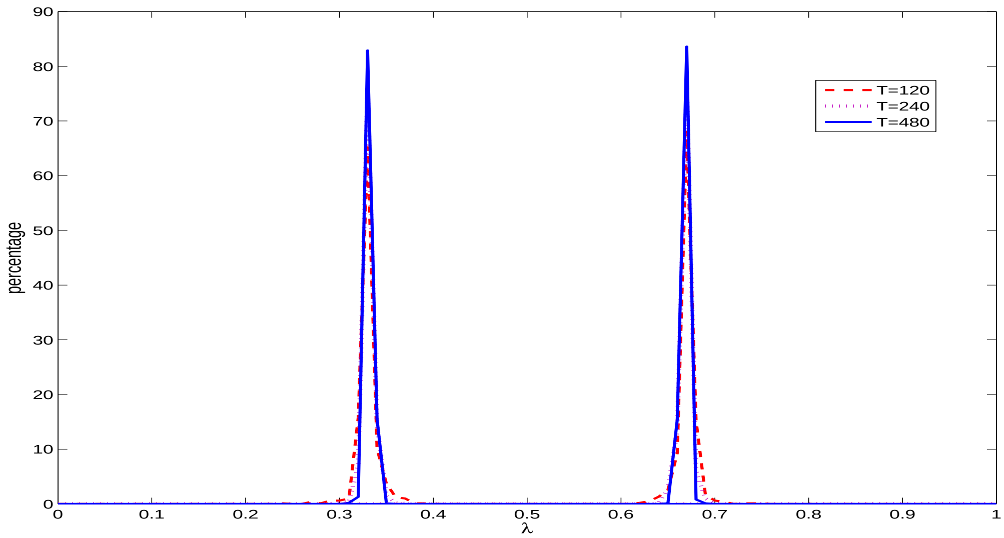

Figure 3: Distribution of estimated break fractions in the two break model

Table 1: Empirical coverage of parameter confidence intervals

one break model with stable reduced form

q	T		Confidence Intervals					
			intercept			slope		
			99%	95 %	90 %	99%	95 %	90 %
2	60	1^{st} regime	.98	.93	.89	.98	.93	.87
		2^{nd} regime	.99	.95	.89	.99	.95	.90
	120	1^{st} regime	.99	.95	.90	.99	.96	.89
		2^{nd} regime	.99	.94	.87	.98	.93	.87
	240	1^{st} regime	.99	.94	.89	.99	.94	.89
		2^{nd} regime	.98	.93	.89	.98	.94	.88
	480	1^{st} regime	.99	.96	.91	.99	.95	.90
		2^{nd} regime	.99	.96	.89	.99	.94	.89
4	60	1^{st} regime	.99	.95	.89	.98	.94	.90
		2^{nd} regime	.98	.93	.87	.98	.94	.90
	120	1^{st} regime	.99	.94	.90	.99	.96	.89
		2^{nd} regime	.99	.94	.88	.98	.94	.88
	240	1^{st} regime	.99	.96	.92	.99	.96	.91
		2^{nd} regime	.98	.94	.89	.99	.94	.89
	480	1^{st} regime	.99	.94	.90	.99	.95	.91
		2^{nd} regime	.98	.95	.90	.99	.95	.88
8	60	1^{st} regime	.99	.94	.88	.99	.94	.89
		2^{nd} regime	.98	.93	.87	.98	.94	.87
	120	1^{st} regime	.99	.96	.92	.99	.95	.89
		2^{nd} regime	.98	.94	.88	.99	.93	.87
	240	1^{st} regime	.99	.95	.91	.99	.96	.90
		2^{nd} regime	.98	.93	.87	.99	.92	.87
	480	1^{st} regime	.99	.96	.90	1.00	.95	.91
		2^{nd} regime	.98	.94	.90	.99	.95	.90

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

Table 2: Relative rejection frequencies of test statistics

one break model with stable reduced form

q	T	$supF(k)$		$F(l+1\|l)$			$F-UDmax$
		1	2	2:1	3:2	4:3	
4	120	1.00	1.00	.023	.001	0.001	1.00
	240	1.00	1.00	.032	.001	0	1.00
	480	1.00	1.00	.033	.001	0	1.00
8	120	1.00	1.00	.044	.003	0	1.00
	240	1.00	1.00	.029	.003	0	1.00
	480	1.00	1.00	.035	.003	0	1.00
		$supWald(k)$		$Wald(l+1\|l)$			$W-UDmax$
		1	2	2:1	3:2	4:3	
4	120	1.00	1.00	.042	.003	0	1.00
	240	1.00	1.00	.051	.004	0	1.00
	480	1.00	1.00	.043	.002	0	1.00
8	120	1.00	1.00	.064	.006	.002	1.00
	240	1.00	1.00	.037	.003	0	1.00
	480	1.00	1.00	.043	.004	0	1.00

Notes: $supF(k)$ denotes the statistic $Sup - F_T(k;1)$; $F(l+1|l)$ denotes the statistic $F_T(l+1|l)$ and the second tier column beneath it denotes $l+1:l$; F-UDmax denotes the statistic $UDmaxF_T(5,1)$; $supWald(k)$ denotes the statistic $Sup - Wald_T(k;1)$; $Wald(l+1|l)$ denotes the statistic $Wald_T(l+1|l)$ and the second tier column beneath it denotes $l+1:l$; W-UDmax denotes the statistic $UDmaxWald_T(5,1)$; the second tier column under the *sup* tests denotes either k or $l+1:l$ as appropriate; q is the number of instruments; T is the sample size.

Table 3: Empirical distribution of the estimated number of breaks

one break model with stable reduced form

q	T	$F-UDmax$				$W-UDmax$			
		0	1	2	3	0	1	2	3
4	120	0	.977	.023	0	0	.958	.042	0
	240	0	.968	.031	.001	0	.949	.050	.001
	480	0	.967	.032	.001	0	.957	.043	0
8	120	0	.956	.044	0	0	.936	.063	.001
	240	0	.971	.029	0	0	.963	.037	0
	480	0	.965	.034	.001	0	.957	.043	0

Notes: The figures in the block headed $F-UDmax$ ($W-UDmax$) give the empirical distribution of the estimated number of breaks, $\hat{m}_T$, obtained via the sequential strategy using $UDmaxF_T(5,1)$ ($UDmaxWald_T(5,1)$). In each case, L (the maximum number of breaks) is set equal to five and all tests are performed with a nominal 5% significance level; $\hat{m}_T > 3$ in none of the simulations.

Table 4: Empirical coverage of parameter confidence intervals

two break model with stable reduced form

q	T		Confidence Intervals					
			intercept			slope		
			99%	95 %	90 %	99%	95 %	90 %
4	120	1^{st} regime	.99	.93	.89	.99	.94	.88
		2^{nd} regime	.98	.93	.89	.98	.92	.97
		3^{rd} regime	.98	.92	.88	.98	.95	.89
	240	1^{st} regime	.99	.95	.90	.99	.95	.89
		2^{nd} regime	.99	.94	.89	.98	.95	.90
		3^{rd} regime	.98	.94	.89	.99	.95	.89
	480	1^{st} regime	.99	.96	.89	.99	.95	.91
		2^{nd} regime	.99	.94	.88	.99	.95	.89
		3^{rd} regime	.99	.95	.91	1.00	.96	.91
8	120	1^{st} regime	.99	.95	.90	.99	.94	.89
		2^{nd} regime	.98	.93	.87	.98	.93	.88
		3^{rd} regime	.99	.93	.88	.99	.95	.88
	240	1^{st} regime	.99	.96	.92	.99	.95	.90
		2^{nd} regime	.98	.93	.87	.99	.94	.88
		3^{rd} regime	.99	.95	.89	.98	.93	.87
	480	1^{st} regime	.99	.95	.90	.99	.95	.90
		2^{nd} regime	.99	.94	.89	.99	.93	.89
		3^{rd} regime	.99	.95	.90	.99	.95	.90

Notes: See Table 1 for definitions.

Table 5: Relative rejection frequencies of test statistics

two break model with stable reduced form

q	T	$supF(k)$		$F(l+1\|l)$		$F-UDmax$
		1	2	2:1	3:2	
4	120	1.00	1.00	1.00	.019	1.00
	240	1.00	1.00	1.00	.013	1.00
	480	1.00	1.00	1.00	.015	1.00
8	120	1.00	1.00	1.00	.016	1.00
	240	1.00	1.00	1.00	.007	1.00
	480	1.00	1.00	1.00	.010	1.00
		$supWald(k)$		$Wald(l+1\|l)$		$W-UDmax$
		1	2	2:1	3:2	
4	120	1.00	1.00	1.00	.032	1.00
	240	1.00	1.00	1.00	.013	1.00
	480	1.00	1.00	1.00	.011	1.00
8	120	1.00	1.00	1.00	.029	1.00
	240	1.00	1.00	1.00	.013	1.00
	480	1.00	1.00	1.00	.013	1.00

Notes: See Table 2 for definitions.

Table 6: Empirical distribution of the estimated number of breaks

two break model with stable reduced form

q	T	$F-UDmax$				$W-UDmax$			
		0	1	2	3	0	1	2	3
4	120	0	0	.961	.039	0	0	.951	.049
	240	0	0	.981	.019	0	0	.976	.024
	480	0	0	.986	.014	0	0	.988	.012
8	120	0	0	.965	.035	0	0	.952	.048
	240	0	0	.981	.019	0	0	.974	.026
	480	0	0	.986	.014	0	0	.984	.016

Notes: See Table 3 for definitions.

Table 7: Relative rejection frequencies of test statistics

no break model

q	T	$supF(k)$					$F(l+1\|l)$		$F-UDmax$
		1	2	3	4	5	2:1	3:2	
4	120	.051	.058	.050	.051	.045	.023	.001	.050
	240	.052	.054	.047	.043	.037	.013	.003	.058
	480	.060	.059	.058	.068	.057	.008	.001	.060
8	120	.043	.042	.053	.049	.045	.014	0	.045
	240	.052	.039	.042	.042	.039	.005	0	.049
	480	.058	.057	.058	.052	.050	.017	.001	.062
		$supWald(k)$					$Wald(l+1\|l)$		$W-UDmax$
		1	2	3	4	5	2:1	3:2	
4	120	.074	.093	.083	.077	.075	.018	.007	.105
	240	.072	.079	.071	.065	.058	.011	.004	.083
	480	.060	.063	.064	.071	.063	.010	0	.064
8	120	.064	.083	.090	.085	.077	.024	.007	.089
	240	.073	.068	.072	.070	.057	.008	.001	.086
	480	.066	.075	.070	.065	.062	.014	0	.075

Notes: See Table 2 for definitions.

Table 8: Empirical distribution of the estimated number of breaks

no break model

q	T	$F - UDmax$				$W - UDmax$			
		0	1	2	3	0	1	2	3
4	120	.947	.048	.005	0	.907	.087	.006	0
	240	.942	.053	.005	0	.917	.079	.003	.001
	480	.940	.056	0.004	0	.936	.059	.005	0
8	120	.955	.039	.006	0	.911	.078	.011	0
	240	.951	.046	.003	0	.914	.084	.002	0
	480	.938	.055	.007	0	.925	.071	.004	0

Notes: See Table 3 for definitions.

Table 9: Distribution of estimated number of breaks with unstable reduced form

				Relative frequency of $\hat{m}$			
Case	T	α	$Wald$	0	1	2	3
I	240	.05	0	.943	.056	.001	0
	240	.01	0	.989	.011	0	0
	480	.05	.001	.954	.046	0	0
	480	.01	0	.993	.007	0	0
II	240	.05	.996	0	.941	.059	0
	240	.01	.997	0	.990	.010	0
	480	.05	1.000	0	.957	.043	0
	480	.01	1.000	0	.991	.009	0
III	240	.05	.005	0	.964	.031	.005
	240	.01	.002	0	.989	.009	.002
	480	.05	.012	0	.969	.020	.011
	480	.01	.002	0	.998	.000	.002

Notes: Case I: no breaks in structural equation, one in the reduced form; Case II: coincident break in structural equations and reduced form; Case III: distinct breaks in structural equation and reduced form. α denotes the nominal significance level of all tests. $Wald$ denotes the rejection frequency of the Wald test in (24). $\hat{m}$ is estimated number of breaks using the methodology in Section 5.

Table 10: Application to NKPC - stability statistics for the reduced forms

Dep.var	k	$sup - F$	$F(k+1\|k)$	BIC
$inf^e_{t+1\|t}$	0			-0.615
	1	43.6	41.7	-0.623
	2	67.0	10.4	-0.680
	3	176.5	34.3	-0.649
	4	80.5	46.8	-0.452
	5	70.2		-0.369
og_t	0			-0.663
	1	50.0	30.53	-0.552
	2	40.1	23.1	-0.497
	3	40.0	11.3	-0.276
	4	34.9	11.3	-0.046
	5	31.9		0.255

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

Table 11: Application to NKPC - stability statistics for structural equation

	$sup-F$		$UD-F$	$sup-Wald$		$UD-Wald$	BIC		
Period	0:1	1:2	0:2	0:1	1:2	0:2	$m=0$	$m=1$	$m=2$
1968.4-1975.2	4.15	-	-	23.94	-	-	0.12	3.52	-
1975.3-1981.1	0.98	-	-	0.69	-	-	0.17	0.73	-
1981.2-2001.4	9.86	34.60	20.39	16.68	18.40	31.54	-1.08	-0.84	-0.84

Notes: The sign "-" indicates tests have not been performed due to not enough observations in sub-samples, $(0 : k)$ is the statistic for testing $H_0 : m = 0$ vs. $H_1 : m = k$; $(k : k + 1)$ is the statistic for testing $H_0 : m = k$ vs. $H_1 : m = k + 1$; UD indicates UDmax tests. The percentiles for both F- and Wald- type statistics are at (10%, 1%) significance level respectively: (i) (0:1) = (19.7, 26.71); (ii) (1:2) = (21.79, 28.36); (iii) UDmax(0:2) = (20.00, 26.75).