MANCHESTER 1824

The University of Manchester

Economics Discussion Paper Series EDP-1504

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

Alastair R. Hall

Denise R. Osborn

Nikolaos Sakkas

March 2015

Economics
School of Social Sciences
The University of Manchester
Manchester M13 9PL

The Asymptotic Behaviour of the Residual Sum of Squares in Models with Multiple Break Points[1]

Alastair R. Hall[2]

University of Manchester

Denise R. Osborn

University of Manchester & University of Tasmania

Nikolaos Sakkas

University of Bath

January 23, 2015

[1]The authors gratefully acknowledge the support of the ESRC grant RES-062-23-1351. The paper has benefited from comments of James Nason, Mardi Dungey and participants at the Ken@75 Conference, Warwick University, the Montreal Econometrics seminar, and seminars at Universidad Carlos III de Madrid, Korea University and the University of York (UK).

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

Abstract

Models with multiple discrete breaks in parameters are usually estimated via least squares. This paper, firstly, derives the asymptotic expectation of the residual sum of squares, the form of which indicates that the number of estimated break points and the number of regression parameters affect the expectation in different ways. Secondly, we propose a statistic for testing the joint hypothesis that the breaks occur at specified points in the sample and show that the statistic has a limiting null distribution that is non-standard but simulatable. In an important special case, the statistic can be normalized to make it pivotal and we provide percentiles for the associated limiting distribution. Our analytical results cover linear and nonlinear regression models with exogenous regressors estimated via Ordinary (or Nonlinear) Least Squares and a linear model in which some regressors are endogenous and the model is estimated via Two Stage Least Squares. An application to US monetary policy rejects the common assumption that identified breaks are associated with changes in the chair of the Fed.

JEL classification: C12, C13, C26, E52

Keywords: Linear models, Nonlinear models, Ordinary Least Squares, Two-Stage Least Squares, Parameter change, US monetary policy

1 Introduction

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

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

The first study to examine analytically the consequences of coefficient break point estimation on the residual sum of squares appears to be Ninomiya (2005), who considers breaks in the mean of a Gaussian process with inference on the number of breaks conducted through AIC (the Akaike Information Criterion) viewed as the bias-corrected maximum log-likelihood estimator. Ninomiya (2005) finds the required bias implies that estimation of each break fraction parameter has an impact on the the maximized log likelihood equivalent to estimation of three mean parameters. This result, namely weighting estimation of each break point as three times that of an individual regression coefficient, is used by Hall, Osborn, and Sakkas (2013) to propose a modified penalty term for an information criterion employed when the number and dates of breaks are estimated (along with the regression coefficients) in the OLS context, with Hall, Osborn, and Sakkas (2015) providing an extension to the 2SLS case. Although these papers include Monte Carlo studies that show the modified criteria to perform well, no formal analytical results are provided to justify the relative weighting of break date versus coefficient estimation.

The present paper fills this gap, by studying the asymptotic behaviour of the residual sum of squares in structural break models estimated by least squares (specifically, OLS, NLS and 2SLS). To capture the realistic situation where the precise dates of change are unclear, the breaks are assumed to be of magnitude that "shrinks" with the sample size; see Bai (1997).

More specifically, the paper makes three contributions. Firstly, we derive the asymptotic expectation of the residual sum of squares in models with breaks in the coefficients at unknown dates. For linear or nonlinear regression models with exogenous regressors, this expectation depends on the numbers of estimated break points and estimated mean parameters, with the former having a weight of three relative to each mean parameter. Although the expression is more complicated in linear models estimated via 2SLS, nevertheless the principal result, namely that each estimated break date has the same impact on the expectation as three estimated mean parameters, carries over to this context. Secondly, we propose a statistic for testing the joint hypothesis that the breaks occur at specified points in the sample. Under its null hypothesis, this statistic is shown to have a limiting distribution that is non-standard but, under certain assumptions, asymptotically pivotal after normalization; percentiles are provided for this limiting distribution. Although the same distribution is obtained by Hansen (2000) (see also Hansen, 1997) in the context of testing the location of the single threshold in a TAR (threshold autoregressive) model, no joint test appears to have been proposed previously in the literature. Our third contribution is to examine breaks in US monetary policy, for which we shed new light on the common assumption that Paul Volcker taking over as Fed chair marked an immediate policy change (see, for example, Clarida, Gali, and Gertler, 2000).

An outline of the remainder of the paper is as follows. Section 2 obtains the asymptotic expectation of the minimized residual sum of squares for linear and nonlinear regression models with exogenous regressors. Section 3 then examines the case of a model with endogenous regressors estimated via 2SLS, where the reduced form may be either stable (with no breaks) or unstable and subject to breaks that need not coincide with those of the structural form. Section 4 proposes our joint test for the hypothesis that breaks occur at certain pre-specified points in the sample. A Monte Carlo analysis in Section 5 illustrates finite sample implications of the preceding analyses. Section 6 examines breaks in US monetary policy, while Section 7 concludes. All proofs are relegated to a mathematical appendix.

2 RSS with Exogenous Regressors

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

2.1 Linear models

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

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

with $T_0^0 = 0$ and $T_{m+1}^0 = T$, where T is the total sample size. Thus, y_t is the dependent variable, while x_t is a $p \times 1$ vector of exogenous explanatory variables that typically includes the constant term, and u_t is a mean zero error. As usual in the literature, we require the true break points to be asymptotically distinct.

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

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

$$y_t \;=\; x_t'\beta_i^* \,+\, e_t^*, \quad i = 1, ..., m+1, \quad t = T_{i-1} + 1, ..., T_i, \tag{2}$$

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

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

Assumption 2 requires that each segment considered contains sufficient observations for estimation with finite T, while containing a positive fraction of the sample asymptotically; in

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

practice ϵ is chosen to be small in the hope that the last part of the assumption is valid. The estimates of $\beta^* = (\beta_1^{*\prime}, ..., \beta_{m+1}^{*\prime})'$ are obtained by minimizing the sum of squared residuals

$$S_T(T_1, ..., T_m;\, \beta) = \sum_{i=1}^{m+1} \sum_{t=T_{i-1}+1}^{T_i} \{y_t \,-\, x_t'\beta_i\}^2 \tag{3}$$

with respect to $\beta = (\beta_1{}', ..., \beta_{m+1}{}')'$. We denote these estimators by $\hat{\beta}(\{T_i\}_{i=1}^m)$ with $\hat{\beta}_j(\{T_i\}_{i=1}^m)$ being the associated estimator of β_j^* relating to segment j. The estimators of the break points, $(\hat{T}_1, ..., \hat{T}_m)$, are then defined as

$$(\hat{T}_1, ..., \hat{T}_m) = \underset{T_1,...,T_m}{\operatorname{argmin}}\, S_T\left(T_1, ..., T_m;\ \hat{\beta}(\{T_i\}_{i=1}^m)\right) \tag{4}$$

where the minimization is taken over all possible partitions, $(T_1, ..., T_m)$, and the associated minimized residual sum of squares is denoted $RSS(\widehat{T}_1, ..., \widehat{T}_m) \,=\, S_T\left(\widehat{T}_1, ..., \widehat{T}_m;\ \hat{\beta}(\{\widehat{T}_i\}_{i=1}^m)\right)$. The OLS estimates, $\hat{\beta}(\{\hat{T}_i\}_{i=1}^m)$, are then the regression parameter estimates associated with the estimated partitions. The estimated break fractions are collected in $\hat{\lambda}$, the $m \times 1$ vector with j^{th} element $\hat{T}_j/T$. Bai (1997) and Bai and Perron (1998) derive the large sample behaviours of $\hat{\lambda}$ and $\hat{\beta}(\{\hat{T}_i\}_{i=1}^m)$, together with various tests for parameter variation that arise naturally in this context.

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

$$\xi_T = RSS(\widehat{T}_1, ..., \widehat{T}_m) \,-\, T\sigma^2, \tag{5}$$

where

$$RSS(T_1, \ldots, T_m) \;=\; \sum_{j=1}^{m+1} RSS_j(T_1, \ldots T_m), \tag{6}$$

$$RSS_j(T_1, \ldots, T_m) \;=\; \sum_{t=T_{j-1}+1}^{T_j} \left\{y_t \,-\, x_t'\,\hat{\beta}_j(\{T_i\}_{i=1}^m\})\right\}^2. \tag{7}$$

Hence ξ_T defined by (5) is the difference between the (minimized) residual sum of squares in (3) and the expected error sum of squares, $T\sigma^2 \,=\, E[\sum_{t=1}^T u_t^2]$, in the data generating process (DGP) of (1).

We decompose ξ_T into three components,

$$\xi_T = \sum_{j=1}^{3} \xi_{j,T}. \tag{8}$$

The first component,

$$\xi_{1,T} = RSS(\widehat{T}_1, ..., \widehat{T}_m) - RSS(T_1^0, ..., T_m^0), \tag{9}$$

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

$$\xi_{2,T} = RSS(T_1^0, ..., T_m^0) - ESS(T_1^0, ..., T_m^0), \tag{10}$$

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

$$\xi_{3,T} = ESS(T_1^0, ..., T_m^0) \, - \, T\sigma^2, \tag{11}$$

and therefore captures the effects of the specific random disturbance sequence $\{u_t\}$.

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

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

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

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

Assumption 4 *$T^{-1}\sum_{t=T_{i-1}^0+1}^{T_{i-1}^0+[rT]} x_t x_t' \xrightarrow{p} rQ_i$ uniformly in $r \in (0, \lambda_i^0 - \lambda_{i-1}^0)$, where Q_i is a positive definite matrix for $i = 1, \ldots m+1$.*

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

[2]That is, there exists γ such that $c'V_T(1)c > \gamma > 0$ for all vectors of constants c such that $\|c\| = 1$.

Assumption 6 *There exists an $l_0 > 0$ such that for all $l > l_0$, the minimum eigenvalues of $A_{il} = (1/l)\sum_{t=T_i^0+1}^{T_i^0+l} x_t x_t'$ and of $\bar{A}_{il} = (1/l)\sum_{t=T_i^0-l}^{T_i^0} x_t x_t'$ are bounded away from zero for all $i = 1, ..., m$.*

Assumption 4 limits the behaviour of the regressor cross product matrix and rules out trending regressors but allows regime specific behaviour. Assumption 5(i)-(iii) ensures $\{x_t u_t\}$ satisfies the Functional Central Limit Theorem within each regime (*e.g.* see White (2001)[Theorem 7.19]). Parts (iv)-(v) place restrictions on V_i; they are stated separately since these are relaxed in some parts of our analysis. Finally, Assumption 6 requires there be enough observations near the true break points so that they can be identified and is analogous to the extension proposed in Bai and Perron (1998) to their Assumption A2.

The component $\xi_{1,T}$ is the focus of much of our analysis. This is closely related to the asymptotic distribution of the estimator for the location of a single break point obtained, under an assumption of a "shrinking" or "small" break, by Yao (1987) for the mean of an *i.i.d.* process and very recently extended to more general linear and nonlinear univariate time series models by Ling (2015). Bai (1997) examines the break point estimator in a regression model, with Hansen (1997, 2000) considering the analogous case of threshold estimation in a single threshold TAR model, while multiple breaks are studied in Bai and Perron (1998). Lemmata 1 to 3, stated below, readily follow from results available in this literature.

Lemma 1 *Under Assumptions 1, 2, 3, 4, 5(i)-(iii) and 6, there exist positive constants K_i, $i = 1, ..., m$, such that for large T, $\Pr\left(\left|T_i - T_i^0\right| > K_i s_T^{-2}\right) < C_i$ for any positive $C_i < \infty$. Then for $k_i \in [-K_i, K_i]$, $i = 1, ..., m$,*

$$\xi_{1,T} \xrightarrow{d} \sum_{i=1}^{m} \min_{k_i} G_i(k_i) \tag{12}$$

where

$$G_i(k_i) = \begin{cases} |k_i|\ a_{i,1} - 2\, c_{i,1}^{1/2} W_{i,1}(-k_i), & \text{if } k_i \leq 0 \\ |k_i|\ a_{i,2} - 2\, c_{i,2}^{1/2} W_{i,2}(k_i), & \text{if } k_i > 0 \end{cases} \tag{13}$$

in which $W_{i,j}(.)$ $(i = 1, ..., m,\ j = 1, 2)$ are independent Brownian motions on $[0, \infty)$ and

$$a_{i,j} = \theta_i^{0\prime} Q_{(i-1)+j}\, \theta_i^0 \tag{14}$$

$$c_{i,j} = \theta_i^{0\prime} V_{(i-1)+j}\, \theta_i^0. \tag{15}$$

Lemma 1 follows from arguments in Bai and Perron (1998) and Bai (1997); see the Appendix below. Clearly, minimization of $G_i(k_i)$ is equivalent to maximization of $\widetilde{G}_i(k_i) = -G_i(k_i)$, namely the maximum of two independent Brownian motion processes with negative drifts. The following Lemmata and Definition provide distributional results relating to this maximum.

Lemma 2 *Let $W(.)$ be a standard Brownian motion on $[0, \infty)$. Then, for $\alpha > 0$, $\gamma > 0$ and $k \in [0, \infty)$*

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

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

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

$$\begin{aligned} F(w; \mu_1, \mu_2) &= (1 - e^{-\mu_1 w})(1 - e^{-\mu_2 w}) \\ &= \int_0^w f\,(b;\, \mu_1,\, \mu_2)\, db \end{aligned}$$

where

$$f\,(b; \mu_1, \mu_2) = \sum_{i=1}^{2} \mu_i e^{-b\mu_i} - \mu e^{-b\mu} \tag{16}$$

for $\mu = \sum_{i=1}^{2} \mu_i$.

Lemma 3 *Let $v_i \sim$ exponential(μ_i) for $i = 1, 2$ and $v_1 \perp v_2$. Then $b = \max\{v_1, v_2\} \sim \mathcal{B}(\mu_1, \mu_2)$ and*

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

Lemma 2, which is stated in Bai (1997)[p.563] and, for $\gamma = 1$, in Stryhn (1996)[Proposition 1], makes clear that the maximum value taken by an individual Brownian motion process with negative drift follows an exponential distribution. Our notation for the distribution of the maximum of two independent processes is given by (16). The result in (17), which is key to our analysis, follows from the mean of an exponential distribution. Although not stated in this form, Ninomiya (2005) uses the result in Lemma 3 in his analysis of the mean shift model.

Having established this background, the following theorem gives the form of $AE[\xi_T]$ for the linear model with exogenous regressors.

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

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

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

2.2 Nonlinear models

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

$$y_t = f(x_t, \beta_i^0) + u_t, \qquad i = 1, ..., m+1, \qquad t = T_{i-1}^0 + 1, ..., T_i^0, \tag{18}$$

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

Compared with the OLS case, the consistency and large sample distribution of $\hat{\lambda}$ and $\hat{\beta}(\{\hat{T}_i\}_{i=1}^m)$ have been established to date in the NLS setting only under more restrictive conditions on the dynamic structure of the data and also the rate of shrinkage between regimes; see Boldea and Hall (2013)[Assumptions 2-8]. These additional restrictions arise because of the inherent nonlinearity of the model; see Boldea and Hall (2013) for further discussion. We impose these conditions, but for brevity relegate some to the Appendix. In addition to (18) replacing (2), Assumption 3 is modified so that $\alpha \in [0.25, 0.5)$, and analogues are required for Assumptions 4 (with x_t

replaced by $F_t(\beta_0) = \partial f_t(\beta)/\partial\beta)|_{\beta=\beta_0})$ and 5 (with h_t replaced by $u_t F_t(\beta_0)$). We note that these assumptions cover a range of models such as smooth transition autoregressive (STAR) and nonlinear ARCH.

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

Theorem 2 *Let y_t be generated by (18) and the following Assumptions hold: 1, 2 with (18) replacing (2), 3 for $\alpha \in [0.25, 0.5)$, 5(i),(iv)-(v) and A.1-A.4 (in the Appendix). Then $AE[\xi_T]$ and $AE[\xi_{i,T}]$ $(i = 1,2,3)$ are given by the respective expressions in Theorem 1.*

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

3 Two Stage Least Squares RSS

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

$$y_t \; = \; x_t'\beta_{x,i}^0 \, + \, z_{1,t}'\beta_{z_1,i}^0 \, + \, u_t, \qquad i = 1,...,m+1, \qquad t = T_{i-1}^0 + 1,...,T_i^0, \tag{19}$$

where $T_0^0 = 0$, $T_{m+1}^0 = T$ and T is the total sample size. Here x_t is a $p_1 \times 1$ vector of endogenous explanatory variables, $z_{1,t}$ is a $p_2 \times 1$ vector of exogenous variables including the intercept, and u_t is a mean zero error. We define $p = p_1 + p_2$. As in the previous section, we assume the location and magnitude of the breaks are governed by Assumptions 1 and 3 respectively.

As (19) is a structural equation, the endogenous explanatory variables, x_t, are (in general) correlated with the errors, u_t, and so 2SLS requires a reduced form representation to be estimated using appropriate instruments. The reduced form is discussed in the first subsection below, before attention is focussed on (19).

3.1 Reduced form model

The reduced form model is

$$x_t^{'} \; = \; z_t^{'}\Delta_k^0 \, + \, v_t^{'}, \qquad k = 1,2,\ldots,h+1, \qquad t = T_{k-1}^{\dagger} + 1,\ldots,T_k^{\dagger}, \tag{20}$$

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

Assumption 7 $T_k^\dagger = [T\pi_k^0]$, *where* $0 < \pi_1^0 < \ldots < \pi_h^0 < 1$.

Note that the reduced form break fractions, $\pi^0 = [\pi_1^0, \ldots, \pi_h^0]'$, may or may not coincide with the breaks in the structural equation, $\lambda^0 = [\lambda_1^0, \ldots, \lambda_m^0]'$. Analogously to the structural form Assumption 3, we assume the breaks in the reduced form are shrinking.

Assumption 8 $\Delta_{k+1}^0 - \Delta_k^0 = A_{T,k}^0 = A_k^0 r_T$ *where* $r_T = T^{-\alpha_r}$, *for* $\alpha_r \,\epsilon\, (0,\ 0.5)$ *and* $k = 1, ..., h$.

The reduced form of (20) can be re-written as

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

where $\Theta^0 = [\Delta_1^{0'}, \ldots \Delta_{h+1}^{0'}]^{'}$, $\tilde{z}_t(\pi^0) = \iota(t,T) \otimes z_t$, $\iota(t,T)$ 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 $k = 2, \ldots, h$ and $\mathcal{I}\{\cdot\}$ is an indicator variable that takes the value one if the event in the curly brackets occurs.

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

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

This condition would be satisfied if, for example, the break dates in the reduced form are estimated by OLS equation by equation and the estimates of the break fractions are then pooled; see Bai and Perron (1998)[Proposition 5] and Bai (1997)[Proposition 1]. Notice that under our assumption[3] $1 - 2\alpha_r > 0$ and $\hat{\pi}$ is consistent for π^0. Let $\hat{x}_t$ denote the resulting fitted values, that is,

$$\hat{x}'_t \;=\; \tilde{z}_t(\hat{\pi})'\hat{\Theta}_T(\hat{\pi}) \;=\; \tilde{z}_t(\hat{\pi})'\left(\sum_{t=1}^{T}\tilde{z}_t(\hat{\pi})\tilde{z}_t(\hat{\pi})'\right)^{-1}\sum_{t=1}^{T}\tilde{z}_t(\hat{\pi})x'_t \tag{22}$$

where $\tilde{z}_t(\hat{\pi})$ is defined analogously to $\tilde{z}_t(\pi^0)$.

[3]We note that this assumption may impose further restrictions upon the data than those assumed below. See Bai and Perron (1998) and Bai (1997) for further details.

In the special case when the reduced form is stable, (20) is replaced by a model with a single regime ($h = 0$), while Assumptions 7 and 8 are redundant. Obviously, (22) then becomes the corresponding OLS expression for $\hat{x}_t'$.

3.2 Structural form RSS

For estimation of (19), the statistic of interest is the minimized residual sum of squares from the second stage estimation. Now suppose that a researcher knows the number of the breaks in (19) but not their locations. As in the previous section, we use λ to denote an arbitrary set of m break fractions in the model of interest. The second stage of 2SLS can begin with the estimation via OLS of

$$y_t \;=\; \hat{x}_t'\beta_{x,i}^* \,+\, z_{1,t}'\beta_{z_1,i}^* \,+\, u_t^*, \quad i = 1, ..., m+1, \quad t = T_{i-1}+1, ..., T_i, \tag{23}$$

for each possible unique m-partition of the sample, where $T_i = [\lambda_i T]$ and u_t^* is an error term. Defining β_i^* for a given partition as $\beta_i^{*\prime} = ({\beta_{x,i}^*}', {\beta_{z_1,i}^*}')'$ and replacing x_t by $\widehat{w}_t = (\widehat{x}_t', z_{1,t}')'$, estimation proceeds by minimizing the residual sum of squares as discussed in Section 2, leading to the 2SLS estimates $\hat{\beta}(\{\hat{T}_i\}_{i=1}^m) = (\hat{\beta}_1', ..., \hat{\beta}_{m+1}')'$ and associated estimated break fractions given by $\hat{\lambda}$, the $m \times 1$ vector with i^{th} element $\hat{T}_i/T$.

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

Assumption 10 *Equation (23) 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* $\epsilon < inf_i(\lambda_{i+1}^0 - \lambda_i^0)$, *and* $\epsilon < inf_k(\pi_{k+1}^0 - \pi_k^0)$, $k = 1, ..., h$.

The generalization in Assumption 10 implies that the search for structural form breaks not only covers the relevant structural form intervals, but is also conducted in all intervals between (true) reduced form breaks. However, when the reduced form is stable, this latter requirement is redundant. For ease of presentation, the following assumptions also redefine some notation used in Section 2.

Assumption 11 *For* $h_{1,t} = (u_t, v_t')'$ *and* $h_{t,i}$ *the* i^{th} *element of* $h_t = h_{1,t} \otimes z_t$*: (i)* $E[h_{1,t} \,|\, \mathcal{F}_t] = 0$ *where* $\mathcal{F}_t$ *is the* σ*-algebra generated by* $\{z_t,\ h_{1,t-1},\ z_{t-1},\ h_{1,t-2},\ \ldots\}$*; (ii)* $E[\|h_{t,i}\|_d] < H_d < \infty$ *for* $t = 1, 2, \ldots$ *and some* $d > 2$*; (iii)* $V_{T,i}(r) = Var[T^{-1/2} \sum_{t=T_{i-1}^0+1}^{T_{i-1}^0+[rT]} h_t]$ *is uniformly positive* *definite for all T sufficiently large and $\lim_{T\to\infty} V_{T,i}(r) = rV_i$, uniformly in $r \in (0, \lambda_i^0 - \lambda_{i-1}^0)$ where V_i is a positive definite matrix of constants; (iv) $Var[h_{1,t} \,|\, \mathcal{F}_t\,, t/T \in [\lambda_{i-1}^0, \lambda_i^0)] = \Omega_i$, where Ω_i is the $(p_1+1)\times(p_1+1)$ positive definite matrix of constants given by*

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

with σ_i^2 a scalar; (v) $\Omega_i = \Omega$, $i = 1, \ldots, m+1$.

Assumption 12 *$rank\{\Upsilon_i^0\} = p$ where $\Upsilon_i^0 = [\Delta_i^0, \Pi]$, for $i = 1, 2, \cdots, h+1$ where $\Pi' = [I_{p_2}, 0_{p_2\times(q-p_2)}]$, I_a denotes the $a \times a$ identity matrix and $0_{a\times b}$ is the $a \times b$ null matrix.*

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

Assumption 14 *(i) $T^{-1}\sum_{t=T_{i-1}^0+1}^{T_{i-1}^0+[rT]} z_t z_t' \xrightarrow{p} rQ_{ZZ}(i)$ uniformly in $r \in (0, \lambda_i^0 - \lambda_{i-1}^0)$, where $Q_{ZZ}(i)$ is a positive definite matrix for $i = 1, \ldots, m+1$, (ii) $Q_{ZZ}(i) = Q_{ZZ}$, $i = 1, \ldots, m+1$.*

Assumption 11 requires $h_{1,t}$ to be a conditionally homoscedastic martingale difference sequence, and imposes sufficient conditions to ensure the analogue of $T^{-1/2}\sum_{t=1}^{[Tr]} h_t$ satisfies a Functional Central Limit Theorem within each regime (see White (2001)[Theorem 7.19]). It also contains the restrictions that the implicit population moment condition for 2SLS is valid - that is, $E[z_t u_t] = 0$ - and the conditional mean of the reduced form is correctly specified. Assumptions 11 and 14 combined imply that $V_i = V = \Omega \otimes Q_{ZZ}$. Assumptions 12 and 14, in conjunction with Assumption 11, imply the standard rank condition for identification in IV estimation of the linear regression model.[4] Note Assumption 12 implies $q \geq p$. Assumption 13 requires there be enough observations near the true break points of the structural equation so that they can be identified.

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

$$y_t = \overline{x}_t'\beta_{x,i}^0 + z_{1,t}'\beta_{z,i}^0 + \overline{u}_{t,i}, \tag{24}$$

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

$$\overline{u}_{t,i} = u_t + v_t^{'}\beta_{x,i}^0, \tag{25}$$

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

which is the composite disturbance that applies in (19) for regime i when the endogenous x_t are substituted by $E[x_t\,|z_t]$ from the reduced form. Therefore, (24) applies when the reduced form coefficients are known, with $\overline{x}_t = E[x_t\,|z_t]$ embodying the true reduced form regimes when those coefficients are subject to breaks. Also define

$$\overline{v}_{t,i} \;=\; (x_t - \overline{x}_t)'\beta^0_{x,i} \;=\; v_t^{'}\beta^0_{x,i}. \tag{26}$$

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

$$\rho_i^2 = Var[\overline{u}_{t,i}] \quad \to \quad \rho^2 = \sigma^2 + 2\gamma'\beta^0_x + \beta^{0\prime}_x\Sigma\beta^0_x, \tag{27}$$

$$\overline{\rho}_i = Cov[\overline{v}_{t,i}, \overline{u}_{t,i}] \quad \to \quad \overline{\rho} = \gamma'\beta^0_x + \beta^{0\prime}_x\Sigma\beta^0_x, \tag{28}$$

$$\omega_i^2 = Var[\overline{v}_{t,i}] \quad \to \quad \omega^2 = \beta^{0\prime}_x\Sigma\beta^0_x. \tag{29}$$

With known reduced form coefficients, the quantity ρ^2 provides the asymptotic variance of the composite structural form disturbance $\overline{u}_{t,i}$ of (25) with shrinking coefficients. Therefore, $T\rho^2$ plays an analogous role in our analysis of the residual sum of squares for 2SLS as does $T\sigma^2$ for the OLS case.

Denoting the 2SLS minimized $S_T\left(\widehat{T}_1, ..., \widehat{T}_m;\ \hat{\beta}(\{\widehat{T}_i\}_{i=1}^m)\right)$ as $RSS(\widehat{T}_1, ..., \widehat{T}_m)$, we consider $AE[\xi_T]$ where, analogous to (5),

$$\xi_T = RSS(\widehat{T}_1, ..., \widehat{T}_m) - T\rho^2 \tag{30}$$

in which $AE[\cdot]$ again denotes the asymptotic expectation operator. Hence ξ_T is the difference between the residual sum of squares in the second-step of 2SLS and the expected error sum of squares in (24).

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

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

The first component

$$\xi_{1,T} = RSS(\widehat{T}_1, ..., \widehat{T}_m; \hat{\pi}) - RSS(T^0_1, ..., T^0_m; \pi^0) \tag{31}$$

represents the effect on the second stage residual sums of squares from estimating the coefficients of (19) within each structural form partition based on the estimated rather than the true break dates in both the structural equation and (if relevant) the reduced form. Both elements of (31) are obtained using $\widehat{x}_t$ from (22). The second component is defined as

$$\xi_{2,T} = RSS(T_1^0, ..., T_m^0; \pi^0) - ESS(T_1^0, ..., T_m^0), \tag{32}$$

where $ESS(T_1^0, ..., T_m^0)$ is the error sum of squares for (19) evaluated using the true $\{\beta_i^0\}_{i=1}^{m+1}$ in conjunction with $\widehat{x}_t$. Hence $\xi_{2,T}$ is the impact on the residual sum of squares from estimating the coefficients of (23) with known (true) break dates and evaluated using the first stage $\widehat{x}_t$ with true break dates. The third component is given by

$$\xi_{3,T} = ESS(T_1^0, ..., T_m^0) - ESS^e(T_1^0, ..., T_m^0), \tag{33}$$

where $ESS^e(T_1^0, ..., T_m^0)$ is the error sum of squares evaluated using the true $\{\beta_i^0\}_{i=1}^{m+1}$ in conjunction with the reduced form $\overline{x}_t = E[x_t \,|z_t]$. Consequently $\xi_{3,T}$ is the effect from using $\widehat{x}_t$ rather than $\overline{x}_t$ for computation of the structural equation error sums of squares. The final component is

$$\xi_{4,T} = ESS^e(T_1^0, ..., T_m^0) - T\rho^2, \tag{34}$$

and hence captures the effects of the composite $\overline{u}_{t,i}$ in the structural equation of (24).

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

$$\delta\lambda_i^0 = \lambda_i^0 - \lambda_{i-1}^0 \qquad for\ i = 1, ..., m+1, \tag{35}$$

with $\lambda_0^0 = 0$ and $\lambda_{m+1}^0 = 1$; $\delta\pi_i^0$ $(i = 1, ..., h+1)$ is defined analogously for the true reduced form regime fractions.

Theorem 3 *Let y_t be generated by (19), x_t be generated by (20), and $\hat{x}_t$ be given by (22). Let Assumptions 1, 3, 7 - 14 hold. Then we have: (i) $AE[\xi_{1,T}] = -3m\rho^2$; (ii) $AE[\xi_{2,T}] = -p(m+1)\rho^2 + p(\rho^2 - \sigma^2)\sum_{i=1}^{m+1} d_i/(\delta\lambda_i^0)$; (iii) $AE[\xi_{3,T}] = -q(h+1)(\rho^2 - \sigma^2)$; (iv) $AE[\xi_{4,T}] = 0$; and so*

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

where

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

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

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

Remark 3: Theorem 3 indicates that $AE[\xi_T]$ depends on: the number of structural form breaks, m, the number of mean parameters in each regime, p, the number of instruments, q, the covariance structure of the composite error $\overline{u}_{t,i}$ through $(\rho^2 - \sigma^2) = 2\gamma'\beta_x^0 + \beta_x^{0\prime}\Sigma\beta_x^0$, and also on the relative locations of the structural and reduced form breaks.

Remark 4: The expression for $AE[\xi_{1,T}]$ carries over from Theorems 1 and 2, and so the effect of estimating the residual sum of squares of interest is asymptotically the same irrespective of whether the model is a linear or nonlinear equation with exogenous regressors or a linear equation with endogenous regressors and consistently estimated reduced form break dates. We also note that Lemma 3 underlies this result in all cases.

Remark 5: Theorem 3(i) does not require Assumption 14(ii), and so $AE[\xi_{1,T}]$ has the stated form even if the instrument cross product matrix exhibits the regime specific behaviour delineated in part (i) of that assumption.

The special case of a stable reduced form is of particular interest. Using the definition of d_i for the case of no reduced form breaks in the structural form regime i, it immediately follows that a stable reduced form implies $\sum_{i=1}^{m+1} d_i \,/(\delta\lambda_i^0) = \sum_{i=1}^{m+1}(\delta\lambda_i^0) = 1$. The resulting asymptotic expectation of the residual sum of squares in the second stage regression is stated as a Corollary to Theorem 3:

Corollary 1 *Let y_t be generated by (19), with x_t generated by (20) and $\hat{x}_t$ be given by (22), both with $h = 0$. Let Assumptions 1-3, 9, 11 and 12-14 hold. Then we have: (i) $AE[\xi_{1,T}] = -3m\rho^2$; (ii) $AE[\xi_{2,T}] = -p(m\rho^2 + \sigma^2)$; (iii) $AE[\xi_{3,T}] = -q(\rho^2 - \sigma^2)$; (iv) $AE[\xi_{4,T}] = 0$; and so*

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

Remark 6: With a stable reduced form, the expression for $AE[\xi_{2,T}]$ in Corollary 1 can be written as $-p\left\{(m+1)\rho^2 - (\rho^2 - \sigma^2)\right\}$. Ignoring the second term, which is independent of m, the term $-(m+1)p\rho^2$ can be associated with estimation of the $(m+1)p$ structural form coefficients. Combined with $AE[\xi_{1,T}] = -3m\rho^2$, the comment in Remark 1 about the relative impacts of break-fraction and regression parameter estimation in models with exogenous regressors applies equally in models with endogenous regressors estimated via 2SLS with stable reduced forms. When the reduced form is unstable, however, this result is modified in that p enters the second term of $AE[\xi_{2,T}]$ in Theorem 3(ii).

Remark 7: Corollary 1 also clarifies the role of the reduced form in minimization of the 2SLS residual sum of squares in models with no breaks. When conventional 2SLS is applied to a stable structural form $(m = 0, h = 0)$, (30) becomes $\xi_T = RSS - T\rho^2$ and

$$AE[\xi_T] \;=\; -p\,\rho^2 \;-\; (q-p)(\rho^2 - \sigma^2). \tag{36}$$

The result shows that the downward bias in the minimized 2SLS residual sum of squares compared with $E[\overline{u}_t^2]$ depends not only on the number of structural form coefficients estimated, p, but also on the extent of overidentification $(q - p)$ and the additional asymptotic variation induced in the structural form by the use of IV estimation, namely $E[\overline{u}_t^2 - u_t^2] = (\rho^2 - \sigma^2)$. In this context where both the reduced forms and structural forms are stable, Pesaran and Smith (1994) propose a generalized R^2 criterion computed from the second stage regression, and (36) makes clear that the value of this criterion will asymptotically depend on characteristics of the reduced form (including the number of instruments) as well as the goodness-of-fit of the structural form equation itself.

Remark 8: Two further special cases of Theorem 3 are of interest; in both only the numbers of breaks matter, not their locations *per se*. Firstly, when all reduced form breaks coincide with structural form breaks, with possible additional structural form breaks, then $\sum_{i=1}^{m+1} d_i/(\delta\lambda_i^0) = h+1$ (see the proof of Theorem 3 in the Appendix). In this case,

$$AE[\xi_T] = -[(p+3)m\, +\, p]\rho^2 - (h+1)(\rho^2 - \sigma^2)(q-p). \tag{37}$$

This expression has a similar interpretation to that drawn out in Remark 6, with the first term of (37) giving the bias due to estimation of the structural form coefficients and break dates, while the second shows the roles of the additional asymptotic variation from using IV, $(\rho^2 - \sigma^2)$, and the extent of overidentification $(q-p)$, with the number of reduced form regimes $(h+1)$ now magnifying the latter effects. Secondly, when all structural form breaks coincide with the dates of reduced form breaks, with possible additional reduced form breaks, then $\sum_{i=1}^{m} d_i/(\delta\lambda_i^0) = m+1$ (as again seen from the Appendix) and

$$AE[\xi_T] = -[(p+3)m + p]\rho^2 - (\rho^2 - \sigma^2)\left[q(h+1) - p(m+1)\right]. \tag{38}$$

This has a similar interpretation to (37), although overidentification in the second term of (38) appears in the form of a comparison of the total numbers of reduced and structural form coefficients estimated.

Remark 9: For the general case where reduced and structural form break dates do not necessarily coincide, the theorem shows that although $AE[\xi_T]$ depends on the relative locations of structural and reduced form break points, the extent of this dependence is bounded. Consequently, based on the interpretation of (37) and (38) in Remark 7, the quantity $q(h+1)-p\sum_{i=1}^{m+1} d_i\,/(\delta\lambda_i^0)$ might be interpreted more generally as a measure of the extent of overidentification of the structural form parameters in the presence of structural and/or reduced form breaks.

4 Testing Break Dates

The discussion of Sections 2 and 3 notes that $AE[\xi_{1,T}]$ exhibits similar behaviour in all the models considered, and this is due to the large sample behaviour of $\xi_{1,T}$ being governed by a version of Lemma 1, and more specifically (12)-(13), in each case. The current section exploits this structure to propose a statistic for testing

$$H_0 : \lambda_i^0 = \overline{\lambda}_i \quad \text{for } i = 1, ..., m, \tag{39}$$

with $0 < \overline{\lambda}_1 < ... < \overline{\lambda}_m < 1$, against the alternative hypothesis that at least one $\lambda_i^0 \neq \overline{\lambda}_i$ $(i = 1, ..., m)$. In other words, we consider the situation where the researcher knows the number of breaks, and wishes to test a joint hypothesis regarding their locations. Given the common structure underlying $AE[\xi_{1,T}]$, we consider the OLS case in some detail in the first subsection and then note (in subsection 4.2) how the result extends to other models considered above.

4.1 OLS-based tests

In the OLS framework (Section 2.1), consider the statistic

$$N_\lambda(\overline{\lambda}) \; = \; RSS(\overline{T}_1, ..., \overline{T}_m) - \min_{T_1,...,T_m} RSS\,(T_1, ..., T_m) \tag{40}$$

where $\overline{T}_i = [\overline{\lambda}_i T]$ and $\overline{\lambda} = (\overline{\lambda}_1, \overline{\lambda}_2, ..., \overline{\lambda}_m)$. The following theorem gives the limiting distribution of $N_\lambda(\overline{\lambda})$.

Theorem 4 *Let y_t be generated by (1) with H_0 of (39) true and Assumptions 1-5(i)-(iii) and 6 hold. Then, for the statistic (40),*

$$N_\lambda(\overline{\lambda}) \xrightarrow{d} \sum_{i=1}^{m} \overline{b}_i$$

where $\{\overline{b}_i\}_{i=1}^m$ are mutually independent and $\overline{b}_i \sim \mathcal{B}(\mu_{i,1}, \mu_{i,2})$ with $\mu_{i,j} = 0.5 a_{i,j}/c_{i,j}$ for $j = 1, 2$, $a_{i,j}$ and $c_{i,j}$ defined in (14) and (15) respectively, and $\mathcal{B}(\mu_1, \mu_2)$ as in Definition 1. In addition, if Assumption 5(iv) holds then $\mu_{i,j} = 0.5\sigma^2_{i+j-1}$; and if Assumption 5(v) also holds then $\mu_{i,j} = 0.5\sigma^2$.

Remark 10: The limiting distributions in Theorem 4 depend on model parameters. However, asymptotically valid inference can be performed by simulating the null distribution using consistent estimators of $\mu_{i,j}$ under H_0 and then comparing $N_\lambda(\overline{\lambda})$ to the appropriate percentile of this simulated distribution. A consistent estimator for $\mu_{i,j}$ is given by

$$\hat{\mu}_{i,j} \; = \; \frac{\hat{\theta}_i' \hat{Q}_{i+j-1} \hat{\theta}_i}{2\hat{\theta}_i' \hat{V}_{i+j-1} \hat{\theta}_i} \tag{41}$$

where $\hat{\theta}_i = \hat{\beta}_{i+1} - \hat{\beta}_i$, $\hat{\beta}_i = \hat{\beta}_i(\{\hat{T}_\ell\}_{\ell=1}^m)$ (defined in Section 2.1), $\hat{Q}_\ell = (\hat{T}_\ell - \hat{T}_{\ell-1})^{-1} \sum_{t=\hat{T}_{\ell-1}+1}^{\hat{T}_\ell} x_t x_t'$, $\hat{V}_\ell = (\hat{T}_\ell - \hat{T}_{\ell-1})^{-1} \sum_{t=\hat{T}_{\ell-1}+1}^{\hat{T}_\ell} \hat{u}_{\ell,t}^2 x_t x_t'$, $\hat{u}_{\ell,t} = y_t - x_t' \hat{\beta}_\ell$. This provides a heteroscedasticity-consistent estimator. If Assumption 5(iv) holds and homoscedasticity applies within each regime, then an alternative consistent estimator is

$$\hat{\mu}_{i,j} \; = \; 0.5 \hat{\sigma}^2_{i+j-1} \tag{42}$$

where $\hat{\sigma}_\ell^2 = (\hat{T}_\ell - \hat{T}_{\ell-1})^{-1} \sum_{t=\hat{T}_{\ell-1}+1}^{\hat{T}_\ell} \hat{u}_{\ell,t}^2$. Finally, if Assumption 5(v) holds and the error variance is constant over all regimes, an additional consistent estimator is[5]

$$\hat{\mu}_{i,j} \; = \; 0.5 \hat{\sigma}^2 \tag{43}$$

[5] A degrees of freedom correction can be applied in the denominator of $\hat{\sigma}^2$, to allow for estimation of coefficients and also break dates, as suggested by Theorem 1.

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

Remark 11: If Assumption 5(iv)-(v) holds, then it is possible to normalize the statistic to remove nuisance parameters from the limiting distribution. To this end consider the F-type test statistic

$$F_\lambda(\overline{\lambda}) = \frac{RSS(\overline{T}_1, ..., \overline{T}_m) - \min_{T_1,...,T_m} RSS\,(T_1, ..., T_m)}{\widehat{\sigma}^2}. \tag{44}$$

This leads to the following corollary to Theorem 4:

Corollary 2 *Under the conditions of Theorem 4, including Assumption 5(iv)-(v), we have* $F_\lambda(\overline{\lambda}) \overset{d}{\to} \sum_{i=1}^m \overline{b}_i$ *where* $\{\overline{b}_i\}_{i=1}^m$ *are mutually independent and* $\overline{b}_i \sim \mathcal{B}(0.5, 0.5)$.

Percentiles of this limiting distribution, simulated in MATLAB using 10 million replications, are presented in Table 1. Hansen (1997, 2000) develops a test for the null hypothesis of a known threshold value in a single threshold TAR model, with his statistic being a special case of $F_\lambda(\overline{\lambda})$ with $m = 1$. The critical values presented by Hansen (1997, 2000) are effectively identical to those of Table 1 for $m = 1$.

The statistics above can be used to generate confidence sets for the break fractions. For the linear model with exogenous regressors, an approximate $100(1-\alpha)\%$ confidence set for the break fractions is given by:

$$\left\{ \overline{\lambda} \text{ s.t } N_\lambda(\overline{\lambda}) < q_{m,1-\alpha} \right\} \tag{45}$$

where $N_\lambda(\overline{\lambda})$ is defined in (40) and $q_{m,1-\alpha}$ is the $100(1-\alpha)^{th}$ quantile of $\sum_{i=1}^m \overline{b}_i$ defined in Theorem 4. Clearly, with Assumptions 5(iv)-(v) imposed, the asymptotic critical values of Table 1 can be employed for $q_{m,1-\alpha}$.

Under similar assumptions to ours, Yao (1987) and Bai (1997) obtain the marginal distribution of a single break fraction estimator, which is used by Bai (1997) and also Bai and Perron (1998) to construct a confidence interval for the date of each break. Since the m break date distributions are asymptotically independent, a joint test of the null hypothesis (39) could be deduced from these. In contrast, (44) compares RSS at the hypothesized break dates with the overall minimized RSS, providing a natural test statistic in the least squares context considered here. In common with the confidence interval approach of Elliott and Muller (2007), but not that of Yao (1987), the confidence sets in (45) do not imply the dates included corresponding to a specific $\overline{\lambda}_i$ are necessarily contiguous.

4.2 Other models

As shown in the Appendix, Lemma 1 continues to apply for nonlinear regression models that satisfy Assumptions 1, 2 with (18) replacing (2), 3 for $\alpha \in [0.25, 0.5)$, 5(i), and A.1-A.4 (in the Appendix). In the NLS case, however, $a_{i,j}$, $c_{i,j}$ given in (14) and (15) are replaced by the Appendix expressions (58) and (59), respectively. It therefore follows from Lemmata 2 and 3, together with Definition 1, that the statistic $N_\lambda(\overline{\lambda})$ given by (40) has the limiting distribution for a nonlinear model as given in the first part of Theorem 4. Further, the imposition of Assumption 5 parts (iv) or (iv)-(v) yields the same specializations of $\mu_{i,j}$ as described in Theorem 4.

A consistent estimator of $\mu_{i,j}$ for use in simulation of the limiting distribution is given by (41), except that the following changes are required: $\hat{\beta}_i$ now denotes the NLS estimator of the parameter vector in (estimated) regime i; x_t is replaced by $F_t(\hat{\beta}_\ell) = \partial f_t(\beta)/\partial\beta)|_{\beta=\hat{\beta}_\ell}$ in $\hat{Q}_\ell$ and $\hat{V}_\ell$. If Assumption 5(iv) holds then an alternative consistent estimator is given by (42) but with $\hat{u}_{\ell,t}$ being the NLS residual; if Assumption 5(v) holds then a further consistent estimator is given by (43) with the same redefinition of the residual. Similarly, we can define an analogous version of $F_\lambda(\overline{\lambda})$ for this model which has the limiting distribution given in Corollary 2 under all the assumptions made for this model, including Assumption 5(iv)-(v). Therefore, under these assumptions, the critical values of Table 1 can be applied for testing the joint break fractions hypothesis of (39) in a nonlinear regression model.

In an analogous way, Theorem 4 extends to 2SLS models that satisfy Assumptions 1, 3, 7 - 11(i)-(iii), 12-14 with the forms of $a_{i,j}$, $c_{i,j}$ implied, as appropriate, by either (60)-(61) or (65)-(66) of the Appendix.

In the 2SLS case, however, the construction of a consistent estimator of $\mu_{i,j}$ for use in simulation of the limiting distribution depends on the location of the i^{th} break in the structural equation relative to the reduced form breaks. If $\hat{\pi}_{k-1} < \hat{\lambda}_i < \hat{\pi}_k$ for some k, then a consistent estimator of $\mu_{i,j}$ is given by:

$$\hat{\mu}_{i,j} = \frac{\hat{\theta}_i'\hat{\Upsilon}_k'\hat{Q}_{ZZ}(i+j-1)\,\hat{\Upsilon}_k\,\hat{\theta}_i}{2\hat{\theta}_i'\hat{\Upsilon}_k'\hat{\Phi}(i+j-1)\,\hat{\Upsilon}_k\,\hat{\theta}_i} \tag{46}$$

for $j = 1, 2$, where $\hat{\theta}_i = \hat{\beta}_{i+1} - \hat{\beta}_i$, $\hat{\beta}_i = (\hat{\beta}_{x,i}', \hat{\beta}_{z_1,i}')'$ are the 2SLS estimators of the structural equation coefficients in the estimated i^{th} regime (as defined in Section 3.2), $\hat{\Upsilon}_k = [\hat{\Delta}_k, \Pi]$ where $\hat{\Delta}_k$ are the OLS estimators of the reduced form parameters in the k^{th} estimated reduced form regime, $\hat{Q}_{ZZ}(\ell) = (\hat{T}_\ell - \hat{T}_{\ell-1})^{-1}\sum_{t=\hat{T}_{\ell-1}+1}^{\hat{T}_\ell} z_t z_t'$, $\hat{\Phi}(\ell) = \hat{C}_\ell\hat{V}_\ell\hat{C}_\ell'$, $\hat{C}_\ell = \hat{\nu}_\ell' \otimes I_q$, $\hat{\nu}_\ell = [1, \hat{\beta}_{x,\ell}']$, $\hat{V}_\ell =$

$(\hat{T}_\ell - \hat{T}_{\ell-1})^{-1} \sum_{t=\hat{T}_{\ell-1}+1}^{\hat{T}_\ell} \hat{h}_t \hat{h}_t'$, $\hat{h}_t = \hat{h}_{1,t} \otimes z_t$, $\hat{h}_{1,t} = [\hat{u}_t, \hat{v}_t']'$, $\hat{u}_t = y_t - (x_t', z_{1,t}) \sum_{i=1}^{m+1} \hat{\beta}_i \mathcal{I}\{t/T \in (\hat{\lambda}_{i-1}, \hat{\lambda}_i]\}$, $\hat{v}_t' = x_t' - z_t' \sum_{k=1}^{h+1} \hat{\Delta}_k \mathcal{I}\{t \in (\hat{\pi}_{k-1}, \hat{\pi}_k]\}$. If $\hat{\pi}_{k-1} = \hat{\lambda}_i$ for some k then a consistent estimator of $\mu_{i,j}$ is given by

$$\hat{\mu}_{i,j} \ = \ \frac{\hat{\theta}_i' \hat{\Upsilon}_{k+j-1}' \hat{Q}_{ZZ}(i+j-1)\, \hat{\Upsilon}_{k+j-1}\, \hat{\theta}_i}{2\hat{\theta}_i' \hat{\Upsilon}_{k+j-1}' \hat{\Phi}(i+j-1)\, \hat{\Upsilon}_{k+j-1}\, \hat{\theta}_i} \tag{47}$$

and all other definitions remain the same.

Regardless of the relative positions of the structural and reduced form breaks, if in addition Assumption 11(iv) holds then a consistent estimator for $\mu_{i,j}$ is provided by

$$\hat{\mu}_{i,j} \ = \ 0.5\hat{\rho}_{i+j-1}^2 \tag{48}$$

where $\hat{\rho}_\ell^2 \ = \ (\hat{T}_\ell - \hat{T}_{\ell-1})^{-1} \hat{\nu}_\ell \{\sum_{t=\hat{T}_{\ell-1}+1}^{\hat{T}_\ell} \hat{h}_{1,t} \hat{h}_{1,t}'\} \hat{\nu}_\ell'$. Further, if Assumptions 11(iv)-(v) hold then an alternative consistent estimator for $\mu_{i,j}$ is:

$$\hat{\mu}_{i,j} \ = \ 0.5\hat{\rho}^2 \tag{49}$$

where $\hat{\rho}^2 = T^{-1} \sum_{\ell=1}^{m+1} \hat{\nu}_\ell \{\sum_{t=\hat{T}_{\ell-1}+1}^{\hat{T}_\ell} \hat{h}_{1,t} \hat{h}_{1,t}'\} \hat{\nu}_\ell'$. In this last case, the dependence of the limiting distribution on model parameters can be removed by using

$$F_{\lambda}^{2SLS}(\overline{\lambda}) = \frac{RSS(\overline{T}_1, ..., \overline{T}_m) - \min\limits_{T_1, ..., T_m} RSS\,(T_1, ..., T_m)}{\widehat{\rho}^2}. \tag{50}$$

Under the assumptions listed above for the 2SLS case, including Assumption 11(iv)-(v), and the H_0 of (39), $F_{\lambda}^{2SLS}(\overline{\lambda})$ converges to the limiting distribution in Corollary 2. This enables the critical values of Table 1 to be employed also for testing break dates in a structural model estimated by 2SLS.

As discussed for the linear model with exogenous regressors in the previous subsection, the hypothesis tests for break dates can be inverted to obtain joint confidence intervals for the dates of the m breaks in the models of this subsection.

5 Simulation Results

A Monte Carlo analysis is undertaken in this section in order to illustrate two implications of our results, namely: (i) estimation of the error variance; (ii) hypothesis testing about the break fraction.

5.1 Error variance estimation

We examine the behaviour of three estimators of the error variance in the linear model with exogenous regressors of Section 2.1. The model with m breaks is given by (1) and the three variance estimators take the generic form:

$$\hat{\sigma}_k^2 \ = \ RSS(\widehat{T}_1, ..., \widehat{T}_m)/(T - m(p+1) - km), \ k = 0, 1, 3 \tag{51}$$

where $RSS(\,\cdot\,)$ is defined in (6)-(7), but the estimators differ in the degrees of freedom correction made for break fraction estimation, namely km. Clearly, $k = 0$ makes no such degrees of freedom correction and is employed in Bai and Perron (1998); $k = 1$ effectively treats break point and mean parameter estimation symmetrically in the degrees of freedom correction and is used by Yao (1988); $k = 3$ is suggested by Theorem 1. The finite sample performance of these three estimators is investigated in two ways: bias and the coverage probabilities of confidence intervals for the mean parameters based on the three versions of (51). These confidence intervals take the generic form:

$$\hat{\beta}_{i,j} \ \pm z_{\alpha/2}\hat{\sigma}_k\sqrt{D_{i,j,T}}, \tag{52}$$

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

We consider a linear model with exogenous regressors with $m = 1$ and $p = 2$, with the data generating process (DGP) taking the form:

$$y_t = \begin{cases} \mu_1 + \gamma_1 w_t + u_t \text{ if } t \leq [0.5T] \\ \mu_2 + \gamma_2 w_t + u_t \text{ if } t > [0.5T] \end{cases}$$

where u_t is a sequence of $i.i.d.$ $N(0,1)$ random variables and w_t is a scalar $i.i.d.$ $N(1,1)$ random variable that is uncorrelated with u_t. Thus, in terms of the notation in Section 2.1, $x_t = [1, w_t]'$ and $\beta_i^0 = [\mu_i, \gamma_i]'$. Since Theorem 1 assumes shrinking breaks (Assumption 3), we fix $\mu_2 = \gamma_2 = 1$ and report results for $\mu_1 = \gamma_1 = 1 - (0.3 \times 50^\alpha/T^\alpha)$, for $\alpha = 0.0, 0.1, 0.2, 0.3, 0.4, 0.49$ ($\alpha = 0$ being the fixed breaks case) and sample sizes $T = 120, 240, 360, 480$.

As in the analysis of Section 2, estimation is performed imposing the true number of breaks. The break dates are estimated as defined in (4) except that in practice regimes are restricted to contain at least $[\epsilon T]$ observations. The parameter ϵ, often referred to as the trimming parameter, is set at $\epsilon = 0.1$ throughout our Monte Carlo simulations. The efficient search algorithm of Bai and Perron (2003) is employed in our analysis. Each DGP is replicated 5000 times and within each replication the same random observations are employed across all methods. All simulations are performed in MATLAB.

The results are reported in Table 2. For each parameter configuration, we report the percentage bias of the error variance estimator, together with the coverage probabilities (expressed as percentages) of the coefficient confidence intervals that are closest to (C_{90}) and furthest away (F_{90}) from the nominal level of 90% where the comparison is over all parameters in all regimes. Thus, for a given k, the coverage probability of the confidence intervals in (52) for any i, j, $covp_{i,j}$ say, satisfies

$$|C_{90} - .9| \leq |covp_{i,j} - .9| \leq |F_{90} - .9|.$$

Perhaps not surprisingly, no adjustment in (51) for break date estimation leads to underestimation of the true variance, with our simulations showing a downward bias of more than 2% with $T = 120$, reducing to around 0.50% to 0.75% with $T = 480$. Effectively counting the break estimation as equivalent to a coefficient by using $k = 1$ reduces the extent of the bias in all cases examined, but the variance continues to be underestimated. On the other hand, $k = 3$ leads to modest overestimation except when T is relatively large. However, except for $T = 120$ and $\alpha = 0$, the absolute bias is always less that 0.5% when this higher break weight is used. Indeed, the magnitude of the bias shows a clear ranking of the error variance estimators with $k = 3$ dominating $k = 1$ dominating $k = 0$.

The coverage probabilities indicate that the confidence intervals for individual coefficients are always conservative, covering the true coefficients with empirical probability less than the nominal 90%. Although the simulated coverage probabilty improves with T, it deteriorates as α increases, namely as the magnitude of the break declines, leading to the empirical coverage being around 10 percentage points below the nominal level when $\alpha = 0.49$. Although $\alpha \in (0, 0.5)$ in Assumption 3 ensures that break fraction estimation asymptotically converges at a faster rate than regression coefficient estimation, our results suggest that imprecision in break date estimation may contaminate the distribution of coefficient estimates in finite samples when breaks are relatively small in magnitude.[6] Nevertheless, for all cases examined in Table 2, taking

[6]Hansen (1997) finds a corresponding result in the analogous context of a TAR model, leading him to suggest

account of break date estimation through the use of $k = 3$ when estimating the error variance unambiguously improves the coverage of confidence intervals for the coefficients, compared with the alternative options of $k = 0$ or $k = 1$.

5.2 Break fraction hypothesis tests

Again taking the case of the linear model with exogenous regressors for illustrative purposes, we now examine the performance of the normalized statistic $F_\lambda(\bar{\lambda})$ of (44). Based on the results of the previous subsection, and as suggested by Theorem 1, $\widehat{\sigma}^2$ is computed as in (51) with $k = 3$.

Using the same design as in subsection 5.1, with $m = 1$ and $p = 2$, we consider tests of $H_0 : \overline{\lambda}_1 = 0.5 + \kappa$ for $\kappa = 0, 0.02, 0.04, \ldots, 0.2$. Since $\lambda_1^0 = 0.5$ in our DGP, $\kappa = 0$ corresponds to the case in which the null is true, and as κ increases the distance between the hypothesized value and the truth increases. The calculated test statistic is compared to the critical value in Table 1 for a 5% significance level. Power curves are plotted in Figure 1 for $\alpha = 0.0, 0.1, 0.2, 0.3, 0.4, 0.49$ and κ from 0 to 0.2, representing values of $\overline{\lambda}_1$ between 0.5 and 0.7. The results are again based on 5,000 replications for each case.

As expected, power increases with κ for each T and α, with power inversely related to α for given T and κ. For example, with $\alpha = 0.1$, power is more than 0.95 when $T = 480$ and $\kappa = 20$ ($\overline{\lambda}_1 = 0.7$), but reaches little more than 0.5 for these T and κ values when $\alpha = 0.4$.[7] Clearly, it is difficult to detect deviations from the hypothesized location when the break is small. On the other hand, although developed under the shrinking breaks assumption, the test performs well when the break magnitude is fixed ($\alpha = 0$).

The test also exhibits good size performance overall. It is generally a little under-sized for small values of α, is well-sized (with empirical sizes between 0.041 and 0.057) when $\alpha = 0.2$ and is typically modestly oversized for larger α, although it remains marginally under-sized for $T = 120$ even with $\alpha = 0.4$. Perhaps not surprisingly, the greatest size distortion across the cases considered occurs for the small breaks that apply with $\alpha = 0.49$ and $T = 480$, where the empirical size is 0.085.

---

that empirical confidence intervals for the slope coefficients can be improved by taking account of the imprecision of threshold estimation. Pursuing this line of research is, however, beyond the scope of the present study.

[7] The former case represents a change in each of the two coefficients of magnitude 0.24 compared to 0.1 for the latter, both to be considered in relation to $\sigma^2 = 1$.

6 US Monetary Policy

US monetary policy is widely acknowledged to have undergone change since the 1970s, with many arguing that this provides a key explanation for changes in the properties of inflation and (sometimes) real activity. Studies that explore these issues typically either treat the date(s) of change as known, or employ essentially *ad hoc* approaches to deal with the issue. For example, Boivan and Giannone (2006) split their sample in 1979, reflecting the date at which Paul Volcker became chairman of the US Federal Reserve, while Ahmed, Levin, and Wilson (2004) use sub-samples covering 1960 to 1979 and 1984 to 2002, with 1980 to 1983 omitted due to uncertainty about potential dates of change. In a similar vein, the seminal study of Clarida, Gali, and Gertler (2000) adopts the 1979 change date, but also acknowledges uncertainty about breaks and examines interest rate reaction functions estimated over the individual subsamples implied by the periods of office of the four Fed chairmen within their overall sample period, and also consider a possible post-1982 sample. Although the literature largely accepts that a new monetary policy regime commenced immediately on Volcker becoming chairman in 1979Q3, Duffy and Engle-Warnick (2006) throw some doubt on this finding, since their application of the sequential test procedure of Bai and Perron (1998) in a dynamic monetary policy model finds a 1980Q3 break rather than one a year or more earlier. Nevertheless, the tests available to Duffy and Engle-Warnick (2006) do not allow for endogeneity and they employ only backward-looking specifications.

We examine hypotheses about breaks in US monetary policy using the forward-looking dynamic model

$$r_t = \beta_\pi \pi_{t+1|t} + \beta_x \widetilde{x}_{t+1|t} + \beta_1 r_{t-1} + \beta_2 r_{t-2} + c + u_t \qquad (53)$$

where r_t is the actual Federal Funds rate, while $\pi_{t+1|t}$ and $\widetilde{x}_{t+1|t}$ are forecasts of inflation and a proxy for the output gap, respectively, and c is a constant. We follow Orphanides (2004), who revisits the analysis of Clarida, Gali, and Gertler (2000), by employing real-time data and, more specifically, Greenbook forecasts prepared by Fed staff for meetings of the Federal Open Market Operations Committee (FOMC).[8] The Greenbook provides forecasts of key variables, including inflation, output and unemployment, which informs FOMC interest rate decisions. Although,

[8] All real-time data we use, including the Greenbook forecasts, were downloaded from the website of the Federal Reserve Bank of Philadelphia.

for simplicity, our specification in (53) assumes that policymakers focus on forecasts for the following quarter, Orphanides (2004) finds results to be largely unaffected for horizons between 1 and 4 quarters. Our sample period is 1968Q4 to 2005Q4, which is appropriate for our purpose of examining implicit hypotheses made in the literature about changes in US monetary policy responses.

Although FOMC meetings are held more frequently (and sometimes irregularly), we follow the usual convention of treating them as quarterly by employing forecasts made for the meeting closest to the middle of the quarter. As Greenbook output gap forecasts are available only from late 1987, we follow Boivan (2006) and employ a real-time unemployment gap measure as a proxy in (53). More explicitly, as in Boivan (2006), $\widetilde{x}_{t+1|t}$ is measured as the natural rate of unemployment minus the Fed's forecast, where the natural rate is computed as an average of the historical unemployment rate over data as available at t. The inflation forecasts $\pi_{t+1|t}$ relate to the GNP or GDP price deflator (as appropriate), and are given in the Greenbook as quarter on quarter growth rates, expressed as annualized percentage points. The interest rate series is the average actual Federal Funds rate for the third month of the quarter, with the third month used to ensure that r_t reflects any monetary policy change effected during that quarter.

As already noted, Greenbook forecasts are prepared by Fed staff in advance of FOMC meetings and they are, in principle, conditional on interest rate policy remaining unchanged over the forecast horizon. However, it may not be appropriate to treat these as exogenous in (53), since Ellison and Sargent (2012) argue that the FOMC may doubt the accuracy of these staff forecasts and instead favour a "worst case" scenario. Consequently the Greenbook forecasts may be measured with error in relation to the forecasts of the FOMC itself, with the measurement errors correlated with interest rate decisions. To guard against this possibility, our analysis of breaks in (53) employs a 2SLS approach. The instruments used are π_{t-i}, $\widetilde{x}_{t-i}$, r_{t-i}, for lags $i = 1, 2$, GNP/GDP growth (as appropriate at t) and the interest rate spread between long-term (ten year) bonds and the short-term Federal Funds rate, also for the two quarters prior to t, with all variables real-time as at t.

Based on the analyses of Hall, Osborn, and Sakkas (2013, 2015), we use an information criteria approach to inference in both the reduced form equations for $\pi_{t+1|t}$ and $\widetilde{x}_{t+1|t}$ and in the structural form (53). Specifically, we employ BIC and HQIC, with the penalty function in each case taking account of coefficient and break estimation by counting the number of effective parameters estimated as $(p + 3)m$, as suggested by Theorem 1. [9] The maximum number of breaks is set to 5 in each case, with (in view of the numbers of coefficients estimated) trimming parameters set to $\epsilon = 0.15$ (15% of the total sample) for each reduced form equation and $\epsilon = 0.10$ for the structural form.

Both criteria find the reduced form equation for $\widetilde{x}_{t+1|t}$ to be stable over the sample period, but three breaks are indicated in the $\pi_{t+1|t}$ equation, dated at 1974Q4, 1980Q4 and 1986Q3. Using the reduced form predictions (with breaks taken into account) rather than observations for $\pi_{t+1|t}$ and $\widetilde{x}_{t+1|t}$ in (53), both criteria then indicate that two breaks occur in US monetary policy. The search algorithm estimates the break dates as 1980Q3 and 1985Q3.[10]

The estimated monetary policy reaction functions are presented in Table 3, the first column of which shows the 2SLS estimated coefficients under the assumption that the reduced and structural forms are stable, with the remaining three columns taking account of reduced form and structural form breaks. Under the assumption of stability, the equation is poorly determined, with no individual coefficient significant. On the other hand, allowing for breaks shows US monetary policy to react significantly to forecasts for both the unemployment gap and inflation until 1980Q3, followed by a period to 1985Q3 where the response appears to be targeted strongly to inflation. The final regime, from 1985Q4 to 2005Q4 is one of low inflation and relative stability (the so-called Great Moderation), during which responses appear to be dominated by interest rate dynamics. It is notable that, nevertheless, the implied steady-state monetary policy responses to inflation are effectively constant over the whole sample period. This finding contrasts with Clarida, Gali, and Gertler (2000), who argue that the monetary policy response to inflation was stronger after Volcker became Fed chairman than previously, but agrees with the real-time analysis of Orphanides (2004).

As discussed above, many studies of US monetary policy, including Clarida, Gali, and Gertler (2000) and Orphanides (2004), assume that a break occurs in 1979Q2, with a new regime applying when Paul Volcker took up appointment as the Fed chairman in the following quarter. Indeed, Clarida, Gali, and Gertler (2000) take this further and informally investigate whether monetary

[9] Theorem 1 suggests $(m + 1)p + 3m$ parameters but in the context of model comparison using information criteria we omit the p parameters that are common across all m break models.

[10] It might be noted that, with 149 observations available for estimation, the estimated break dates do not lie at the margin of the search interval given by $\epsilon > 0.10$.

policy changes with each Fed chairman. In terms of our analysis, this would imply that the true date of the second break we detect is 1987Q2, with Alan Greenspan taking up office in August that year.[11] Applying the tools of Section 4, we therefore test the joint null hypothesis

$$H_0 : T_1^0 = 1979Q2, T_2^0 = 1987Q2.$$

Under the assumption of homoscedasticity, the test statistic of (50) is $F_{\lambda}^{2SLS}(\overline{\lambda}) = 37.29$, which strongly rejects the null hypothesis at the 1% level in relation to the critical values of Table 1.[12] Relaxing the homoscedasticity assumption by using the 2SLS analogue of (40) leads to a statistic of $N_{\lambda}^{2SLS}(\overline{\lambda}) = 49.73$, which also leads to rejection at the 1% level whether the null distribution of $\overline{b}_i \sim \mathcal{B}(\mu_{i,1}, \mu_{i,2})$ is simulated under the assumption of regime-dependent variances as in (48) or allowing more general heteroscedasticity as in (46). Indeed, $N_{\lambda}^{2SLS}(\overline{\lambda})$ always rejects the joint null hypothesis at this level for any hypothesized $T_1^0 \neq 1980Q3$. On the other hand, there is substantial uncertainty about the second break date, with a 99% joint confidence set including all dates from 1984Q2 (the lower bound of the search interval in combination with 1980Q3) to 1991Q4, inclusive, while reducing the confidence level to 90% brings forward the latter date by only two quarters. Figure 2 illustrates the 99% joint confidence set graphically in terms of the break fractions, with the horizontal line emphasizing the relative uncertainty about λ_2 in contrast to λ_1.

These results shed new light on the timing of changes in US monetary policy. In particular, the widely accepted break date of 1979Q2 is not supported, with our results strongly pointing to the break occurring 1980Q3. Interestingly, Duffy and Engle-Warnick (2006) also find evidence of a break at this later date in a dynamic monetary policy model. While the literature generally associates a monetary policy change with Volcker alone, the intriguing suggestion from our findings is that the election of Ronald Reagan as US President on 4 November 1980 may have heralded the beginning of a new regime. Although detailed analysis of the evidence is beyond the scope of this paper, it is notable that the policies now referred to as 'Reaganomics' included

[11] Our sample also covers the chairmanship of Arthur Burns (1970-1978) and William Miller (1978-1979), but our results do not indicate any change over the first sub-period and the second is too short to be analyzed as a separate regime with these techniques.

[12] Here $\widehat{\rho}^2$ is calculated using a scaling of T^{-1}, as in the expression immediately under (49). Applying a degrees of freedom correction for coefficient and break estimation through a scaling of $(T-7)^{-1}$ yields a statistic of 35.54, which does not affect the substantive conclusions.

a focus on the control of inflation. As to the second break, our results support other studies, including Clarida, Gali, and Gertler (2000), who suggest that the date of change is unclear. However, we go further than previous authors in the sense that our 90% confidence set includes dates into the early 1990s.

7 Concluding Remarks

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

The paper also illustrates the usefulness of the results through an application to breaks in US monetary policy. Such breaks are widely acknowledged in the literature, but are usually assumed to coincide with changes in the chair of the Federal Reserve; see, for example, Clarida, Gali, and Gertler (2000). When subjected to test, we reject this hypothesis on the coincidence of change. In particular, the widely assumed break date of 1979Q2 associated with the end of the pre-Volcker era is strongly rejected in favour of a break in late 1980. An important side-product of our analysis is the joint confidence set we obtain for two dates of change detected in monetary policy over the period 1969 to 2005.

Appendix

Mathematical Appendix

Proof of Lemma 1

From the principle of least squares, (9) is given by

$$\xi_{1,T} = \min_{(T_1,...,T_m)} RSS(T_1, ..., T_m) - RSS(T_1^0, ..., T_m^0). \tag{54}$$

From Bai and Perron (1998)[Proposition 4], for the limiting behaviour of $\{\widehat{T}_i\}_{i=1}^m$ we need to consider possible break dates T_i only within intervals close to each of the true breaks, given by $B = \bigcup_{i=1}^m B_i$ where $B_i = \left\{\left|T_i - T_i^0\right| \le K_i s_T^{-2}\right\}$ for positive constants K_i, $i = 1, ..., m$. That is, for an individual break i we need to consider $T_i = T_i^0 + [k_i\, s_T^{-2}]$ for $k_i \in [-K_i, K_i]$. Using the same arguments as Bai (1997)[equations (8)-(9)], extended to the multi-break case as in Bai and Perron (1998)[Section 3.3], it follows that

$$\xi_{1,T} = \sum_{i=1}^m \min_{T_i} \{A_i(T_i) + 2C_i(T_i)\} + o_p(1), \quad \text{uniformly in } B \tag{55}$$

where

$$\begin{aligned} A_i(T_i) &= \theta_{T,i}^{0\,\prime} \sum_{t=(T_i \wedge T_i^0)+1}^{T_i \vee T_i^0} x_t x_t' \, \theta_{T,i}^0 \\ C_i(T_i) &= (-1)^{\mathcal{I}\{T_i < T_i^0\}} \, \theta_{T,i}^{0\,\prime} \sum_{t=(T_i \wedge T_i^0)+1}^{T_i \vee T_i^0} x_t u_t \end{aligned}$$

for $\theta_{T,i}^0$ defined in Assumption 3, and $a \vee b = \max\{a, b\}$, $a \wedge b = \min\{a, b\}$ and $\mathcal{I}\{.\}$ is an indicator function which takes the value unity when the condition in curly brackets is satisfied. By construction, the summations used in defining $A_i(T_i)$ and $C_i(T_i)$ include the true break date T_i^0. Under Assumptions 4 and 5(i)-(iii), and using the arguments of Bai (1997), equations (12) to (15) of the text then follow. $\diamond$

Proof of Theorem 1

Part (i): The limiting distribution of $\xi_{1,T}$ is given by Lemma 1. Now consider the maximization of $\widetilde{G}_i(k_i) = -G_i(k_i)$ for a single break i. From (13) and Lemma 2, each $\max_{|k_i|} \left\{ 2\, c_{i,j}^{1/2} W_{i,j}(-k_i) - |k_i|\ a_{i,j} \right\}$ for $j = 1, 2$ is exponential with parameter $\mu_j = -0.5 a_{i,j}/c_{i,j}$. Using Assumptions 4 and 5 in (14)-(15) implies that $a_{i,j}/c_{i,j} = 0.5\sigma^{-2}$ and application of Lemma 3 then yields

$$E\left[\max_{k_i} \widetilde{G}_i(k_i)\right] = -E\left[\min_{k_i} G_i(k_i)\right] = 3\sigma^2$$

for each of the m breaks. Since these breaks can be considered separately, we have

$$AE[\xi_{1,T}] = -3m\sigma^2. \tag{56}$$

Part (ii): Using standard least squares algebra,

$$\begin{aligned}
\xi_{2,T} &= RSS(T_1^0, ..., T_m^0) - ESS(T_1^0, ..., T_m^0) \\
&= \sum_{i=1}^{m+1} \sum_{t=T_{i-1}^0+1}^{T_i^0} (y_t - x_t'\widehat{\beta}_i)^2 - \sum_{i=1}^{m+1} \sum_{t=T_{i-1}^0+1}^{T_i^0} (y_t - x_t'\beta_i^0)^2 \\
&= -\sum_{i=1}^{m+1} (\widehat{\beta}_i - \beta_i^0)'(X_i' X_i)(\widehat{\beta}_i - \beta_i^0) \qquad (57)
\end{aligned}$$

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

Under Assumption 4,

$$T^{-1} X_i' X_i \xrightarrow{p} (\delta\lambda_i^0) Q_i \ = \ M_i, \text{ say}$$

where $\delta\lambda_i^0 = \lambda_i^0 - \lambda_{i-1}^0$. From Bai and Perron (1998)[Proposition 3], we have under our assumptions that

$$T^{1/2}\left(\widehat{\beta} - \beta^0\right) \Rightarrow N(0, V_\beta)$$

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

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

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

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

Proof of Theorem 2

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

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

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

$$\sup_A |P(A|B) - P_j(A)| \leq g_j(B)\rho^t$$

with $0 < \rho < 1$, $A \in \mathcal{F}_{T_{j-1}^0+t}^{T_j^0}$, $B \in \mathcal{F}_{-\infty}^{T_{j-1}^0}$, $\mathcal{F}_k^l$ is the σ-algebra generated by $(v_k, \ldots, v_l)$, and $g_j(\cdot)$ is a positive uniformly integrable function.

(ii) $\{v_t\}$ is a β-mixing process with exponential decay, i.e. there exists $N > 0$ such that for $B \in \mathcal{F}_{-\infty}^a$,

$$\beta_t = \sup_a \beta(\mathcal{F}_{-\infty}^a, \mathcal{F}_{a+t}^\infty) \leq N\rho^t, \quad \text{with } \beta(\mathcal{F}_{-\infty}^a, \mathcal{F}_{a+t}^\infty) = \sup_{A \in \mathcal{F}_{a+t}^\infty} E|P(A|B) - P(A)|$$

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

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

Assumption A.4 (i) $S_T(T_1, \ldots, T_m; \beta)$ has a unique global minimum at β_0 and $(T_1^0, \ldots, T_m^0)$; (ii) Let $V_{T,i}^*(\beta, r) =$ Var $T^{-1/2} \sum_{t=T_{i-1}^0+1}^{T_{i-1}^0+[Tr]} u_t(\beta)F_t(\beta)$. Then $V_{T,i}^*(\beta, r) \xrightarrow{p} rV_i^*(\beta)$, uniformly in $\beta \times r \in \mathbf{B} \times [0, \lambda_i^0 - \lambda_{i-1}^0]$, where $V_i^*(\beta)$ is a positive definite (p.d.) matrix not depending on T, with $V_i^*(\beta)$ not necessarily the same for all i; (iii) Let $Q_{T,i}^*(\beta, r) = T^{-1} \sum_{t=T_{i-1}^0+1}^{T_{i-1}^0+[Tr]} F_t(\beta)F_t(\beta)'$. Then $Q_{T,i}^*(\beta, r) \xrightarrow{p} rQ_i^*(\beta)$, uniformly in $\beta \times r \in \mathbf{B} \times [0, \lambda_i^0 - \lambda_{i-1}^0]$, where $Q_i^*(\beta)$ is a p.d. matrix; (iv) $E[f_t(\beta_i^0)] \neq E[f_t(\beta_{i+1}^0)]$, for each $i = 1, \ldots, m$.

The proof follows similar lines to that of Theorem 1. From the arguments of Boldea and Hall (2013), it follows that (12) and (13) continue to apply, but now with

$$a_{i,j} = \theta_i^{0\prime} Q_{i+j-1}^*(\beta_{i+j-1}^0)\,\theta_i^0 \tag{58}$$

$$c_{i,j} = \theta_i^{0\prime} V_{i+j-1}^*(\beta_{i+j-1}^0)\,\theta_i^0 \tag{59}$$

for $j = 1, 2$. The result for $\xi_{1,T}$ then follows using arguments as for the proofs of Lemma 1 and Theorem 1. For $\xi_{2,T}$, the proof again follows the same argument as Theorem 1 using $T^{1/2}(\hat{\beta}_i - \beta_i^0) \overset{d}{\to} \mathcal{N}\left(0, \sigma^2[Q_i^*(\beta_i^0)]^{-1}\right)$ (under our conditions) from analogous arguments to Boldea and Hall (2013)[Theorem 2], while (iii) follows from $E[u_t^2] = \sigma^2$. $\diamond$

Proof of Theorem 3

Part (i): From the principle of least squares, $\xi_{1,T}$ as defined for 2SLS by (31) has an analogous interpretation to (54). There are then two scenarios of interest for the general case of an unstable reduced form with $h > 0$ in (20), namely whether the (true) reduced form and structural breaks are common or not. To be more precise, and following Boldea, Hall, and Han (2012), we consider scenarios where some breaks occur in the structural form but not the reduced form and where at least some breaks are common to both; the former includes the special case of a stable reduced form. These scenarios can be represented as follows.

Scenario 1: $\pi_j^0 < \lambda_{k+1}^0 < ... < \lambda_{k+\ell}^0 < \pi_{j+1}^0$

Scenario 2: $\pi_{j-1}^0 \leq \lambda_k^0 < \pi_j^0 = \lambda_{k+1}^0 < ... < \lambda_{k+\ell}^0 \leq \pi_{j+1}^0$

Scenario 1

Consider, firstly, a single reduced form break and m structural form breaks, with $0 \leq \pi_1^0 < \lambda_1^0 < ... < \lambda_m^0 < T$, so that

$$\begin{aligned} y_t &= (x_t',\ z_{1,t}')\beta_i^0 + u_t, \qquad i = 1, ..., m,\ t = T_{i-1}^0 + 1, ..., T_i^0 \\ x_t' &= \begin{cases} z_t'\Delta_1^0 + v_t & t \leq T_1^\dagger \\ z_t'\Delta_2^0 + v_t & t > T_1^\dagger \end{cases} \end{aligned}$$

As in Boldea, Hall, and Han (2012), proof of Theorem 3, the relevant intervals for the limiting behaviour of $\{\widehat{T}_i\}_{i=1}^m$ in (31) for 2SLS are again $B = \bigcup_{i=1}^m B_i$ where $B_i = \left\{\left|T_i - T_i^0\right| \leq K_i s_T^{-2}\right\}$ for positive constants K_i, $i = 1, ..., m$. Then, from Boldea, Hall, and Han (2012)[Proposition 2], the minimization implies that $\xi_{1,T}$ can be written as in (55), now with

$$\begin{aligned} A_i(T_i) &= \theta_{T,i}^{0\,\prime}\Upsilon_2^{0\prime} \sum_{t=(T_i \wedge T_i^0)+1}^{T_i \vee T_i^0} z_t z_t' \,\Upsilon_2^0\, \theta_{T,i}^0 \\ C_i(T_i) &= (-1)^{\mathcal{I}\{T_i < T_i^0\}}\, \theta_{T,i}^{0\,\prime}\, \Upsilon_2^{0\prime} \sum_{t=(T_i \wedge T_i^0)+1}^{T_i \vee T_i^0} z_t \overline{u}_{t,i} \end{aligned}$$

for $\theta_{T,i}^0$ and Υ_k^0 $(k = 1, 2)$ defined in Assumptions 3 and 12, respectively and $\overline{u}_{t,i}$ defined in (25).

For break i consider $T_i = T_i^0 + [k_i\, s_T^{-2}]$ for $k_i \in [-K_i, K_i]$. Using the same arguments as Boldea, Hall, and Han (2012) in the proof of their Theorem 2, it follows that the limiting distribution of $\xi_{1,T}$ is given by (12) and (13) as in Lemma 1, but with (from Assumption 14 and Assumption 11(iii)), (14)-(15) replaced by

$$a_{i,j} \quad = \quad \theta_i^{0\prime}\Upsilon_2^{0\prime} Q_{ZZ}(i+j-1)\, \Upsilon_2^0\, \theta_i^0 \tag{60}$$

$$c_{i,j} \quad = \quad \theta_i^{0\prime}\Upsilon_2^{0\prime} \Phi(i+j-1) \Upsilon_2^0\, \theta_i^0 \tag{61}$$

where $\Phi(\ell) = C_\ell V_\ell C_\ell'$, $C_\ell = \nu_\ell' \otimes I_q$, $\nu_\ell = [1, \beta_{x,\ell}^{0\prime}]$. Under Assumption 11(iv) $\Phi(\ell) = \nu_\ell \Omega_\ell \nu_\ell' \otimes Q_{ZZ}(\ell)$, and with the addition of Assumption 11(v), we have $\Phi(\ell) = \nu_\ell \Omega \nu_\ell' \otimes Q_{ZZ}(\ell)$. Thus, under our assumptions

$$c_{i,j} \ = \ \rho_i^2 a_{i,j} \to \rho^2 a_{i,j} \tag{62}$$

where ρ^2 is defined in (27) and Assumption 3 is imposed.

Therefore, applying Lemmata 1 and 2, we have

$$\min_{|k_i|} G(|k_i|) \sim \mathcal{B}(a_{i,j}/2c_{i,j},\ a_{i,j}/2c_{i,j}) = \mathcal{B}(0.5\rho^{-2},\ 0.5\rho^{-2}), \tag{63}$$

and so, as we can consider the breaks separately, it follows from Lemma 3 that

$$AE[\xi_{1,T}] = -3m\rho^2. \tag{64}$$

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

Scenario 2

Under this scenario, consider $h = 1$ in the case where the first of the m structural breaks coincides with the single reduced form break. Hence the data generation process is identical to Scenario 1, except that $T_1^\dagger = T_1^0$ and, consequently, $\pi_1^0 = \lambda_1^0$.

From Boldea, Hall, and Han (2012), and since the m breaks at $T_1^0, \dots, T_m^0$ can be considered separately, the limiting distribution of $\xi_{1,T}$ applies as for scenario 1, with $a_{i,j}$ and $c_{i,j}$ as given by (60) and (61), respectively, for $i = 2, \dots, m$, but $a_{1,j}$ and $c_{1,j}$ are as follows:

$$a_{1,j} \quad = \quad \theta_1^{0\prime}\Upsilon_j^{0\prime} Q_{ZZ}(j)\, \Upsilon_j^0\, \theta_1^0, \qquad\qquad j = 1, 2 \tag{65}$$

$$c_{1,j} \quad = \quad \theta_1^{0\prime}\Upsilon_j^{0\prime} \Phi(j)\, \Upsilon_j^0\, \theta_1^0, \qquad\qquad j = 1, 2. \tag{66}$$

Under our assumptions, therefore, (62) applies and consequently (63) holds for a break that is common to the reduced and structural forms. Therefore (64) holds under Scenario 2.

Part (ii): From standard least squares algebra,

$$\begin{aligned} \xi_{2,T} &= RSS(T_1^0, T_2^0, ..., T_m^0; \pi^0) - ESS(T_1^0, T_2^0, ..., T_m^0) \\ &= \sum_{i=1}^{m} \sum_{t=T_{i-1}^0+1}^{T_i^0} (y_t - \widehat{x}_t(\pi^0)'\widehat{\beta}_{x,i} - z_{1,t}'\widehat{\beta}_{z,i})^2 - \sum_{i=1}^{m} \sum_{t=T_{i-1}^0+1}^{T_i^0} (y_t - \widehat{x}_t(\pi^0)'\beta_{x,i}^0 - z_{1,t}'\beta_{z,i}^0)^2 \\ &= -\sum_{i=1}^{m} (\widehat{\beta}_i - \beta_i^0)'(\widehat{W}_i'\widehat{W}_i)(\widehat{\beta}_i - \beta_i^0) \end{aligned} \tag{67}$$

in which $\widehat{W}_i$ is the $(T_i^0 - T_{i-1}^0) \times p$ data matrix for the i^{th} structural form regime, with typical row $(\widehat{x}_t(\pi^0)', z_{1t}')$, and $\widehat{\beta}_i = (\widehat{\beta}_{x,i}', \widehat{\beta}_{z,i}')'$ are obtained using the true reduced form break fractions of π^0.

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

$$\widetilde{Q}_{ZZ}(\lambda) = \phi(\lambda) \otimes Q_{ZZ} \tag{68}$$

where

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

and $\delta\pi_j^0 = \pi_j^0 - \pi_{j-1}^0$ ($\pi_0^0 = 0$, $\pi_{h+1}^0 = 1$). Therefore, we have

$$\widetilde{Q}_i = \phi_i^{(1)} \otimes Q_{ZZ} \tag{69}$$

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

$$\phi_i^{(1)} \;=\; diag[0, \ldots, 0, \delta\lambda_i^0, 0, \ldots, 0]; \tag{70}$$

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

$$\phi_i^{(1)} = diag\left[0, ..., 0, (\pi_{k+1}^0 - \lambda_{i-1}^0), \delta\pi_{k+2}^0, ..., \delta\pi_{k+\ell_i}^0, (\lambda_i^0 - \pi_{k+\ell_i}^0), 0, ..., 0\right]. \tag{71}$$

For later reference it is also useful to note that

$$\widetilde{Q}_{ZZ}(1) = \phi_0 \otimes Q_{ZZ} \tag{72}$$

where

$$\phi_0 = \phi(1) = diag[\delta\pi_1^0, \delta\pi_2^0 ..., \delta\pi_{h+1}^0]. \tag{73}$$

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

$$T^{-1}\widehat{W}_i' \, \widehat{W}_i = \widehat{M}_{ww}^{(i)} \xrightarrow{p} M_{ww}^{(i)} = \widetilde{\Upsilon}'\widetilde{Q}_i \, \widetilde{\Upsilon}$$

where $\widetilde{\Upsilon}' = [\Upsilon_1^{0\prime}, \Upsilon_2^{0\prime}, ..., \Upsilon_{h+1}^{0\prime}]$. From Hall, Han, and Boldea (2012)[Theorem 3], we have that

$$T^{1/2}\left(\widehat{\beta}_i - \beta_i^0\right) \Rightarrow N(0, V_{i,i}^{\beta})$$

where $V_{i,i}^{\beta}$, as in Hall, Han, and Boldea (2012)[Theorem 8], is

$$V_{i,i}^{\beta} = \widetilde{A}_i \left\{ \widetilde{C}_i \widetilde{V}_i \widetilde{C}_i' - \widetilde{E}_i \widetilde{D}_i \widetilde{V}_i \widetilde{C}_i' - \widetilde{C}_i \widetilde{V}_i \widetilde{D}_i' \widetilde{E}_i' + \widetilde{E}_i \widetilde{D}_i \widetilde{V}_i \widetilde{D}_i' \widetilde{E}_i' \right\} \widetilde{A}_i' \tag{74}$$

and

$$\begin{aligned}
\widetilde{A}_i &= [\widetilde{\Upsilon}'\widetilde{Q}_i \, \widetilde{\Upsilon}]^{-1}\widetilde{\Upsilon}' \\
\widetilde{C}_i &= (1, \beta_{x,i}^{0\prime}) \otimes I_{\widetilde{q}}, \quad \widetilde{D}_i = (0, \beta_{x,i}^{0\prime}) \otimes I_{\widetilde{q}}, \quad \widetilde{q} = q(h+1) \\
\widetilde{E}_i &= \widetilde{Q}_i \widetilde{Q}_{ZZ}(1)^{-1} \\
\widetilde{V}_i &= Var\left[T^{-1/2} \sum_{t=[\lambda_{i-1}^0 T]+1}^{[\lambda_i^0 T]} \widetilde{h}_t \right], \quad \widetilde{h}_t = \begin{pmatrix} u_t \\ v_t \end{pmatrix} \otimes \widetilde{z}_t(\pi_0).
\end{aligned}$$

Under Assumption 11, we have

$$\widetilde{V}_i = \phi_i^{(1)} \otimes V = \phi_i^{(1)} \otimes (\Omega \otimes Q_{ZZ})$$

where $\phi_i^{(1)}$ is defined by (70) or (71), as appropriate. Also using (68),

$$\widetilde{E}_i \; = \; \phi_i^{(2)} \otimes I_q,$$

where $\phi_i^{(2)} = \phi_i^{(1)}\{\phi(1)\}^{-1}$.

Now consider each of the terms of (74) in turn. Firstly, since $(1, \beta_{x,i}^{0\prime})\Omega(1, \beta_{x,i}^{0\prime})' = \rho_i^2$ in (27), then

$$\widetilde{C}_i \widetilde{V}_i \widetilde{C}_i' = \rho_i^2(\phi_i^{(1)} \otimes Q_{ZZ}).$$

If $\phi_1^{(i)}$ is given by (70) and $\pi_k^0 \leq \lambda_{i-1}^0 < \lambda_i^0 \leq \pi_{k+1}^0$ then

$$\widetilde{\Upsilon}'\widetilde{C}_i \widetilde{V}_i \widetilde{C}_i' \widetilde{\Upsilon} \; = \; \rho_i^2(\delta\lambda_i^0)\Upsilon_{k+1}^{0\prime} Q_{ZZ} \Upsilon_{k+1}^0 \; \to \; (\delta\lambda_i^0)\rho^2 \, \Upsilon^{0\prime} \, Q_{ZZ} \, \Upsilon^0 \tag{75}$$

under Assumption 8. If $\phi_1^{(i)}$ is given by (71) then we have

$$\begin{aligned}\widetilde{\Upsilon}'\widetilde{C}_i\widetilde{V}_i\widetilde{C}_i'\widetilde{\Upsilon} &= \rho_i^2\left\{(\pi_{k+1}^0-\lambda_{i-1}^0)\Upsilon_{k+1}^{0\prime}Q_{ZZ}\Upsilon_{k+1}^0+\delta\pi_{k+2}^0\Upsilon_{k+2}^{0\prime}Q_{ZZ}\Upsilon_{k+2}^0+...\right.\\ &\quad\left.+(\lambda_i-\pi_{k+\ell_i}^0)\Upsilon_{k+\ell_i+1}^{0\prime}Q_{ZZ}\Upsilon_{k+\ell_i+1}^0\right\}\\ &\to (\delta\lambda_i^0)\rho^2\,\Upsilon^{0\prime}Q_{ZZ}\,\Upsilon^0 \qquad (76)\end{aligned}$$

under Assumption 8. By similar arguments, $\widetilde{D}_i\widetilde{V}_i\widetilde{C}_i' = (\phi_i^{(1)}\otimes Q_{ZZ})\overline{\rho}_i$ and hence

$$\begin{aligned}\widetilde{E}_i\widetilde{D}_i\widetilde{V}_i\widetilde{C}_i' &= (\phi_i^{(2)}\otimes I_q)(\phi_i^{(1)}\otimes Q)\overline{\rho}_i\\ &= \overline{\rho}_i(\phi_i^{(3)}\otimes Q_{ZZ})\end{aligned}$$

where $\phi_i^{(3)} = \phi_i^{(1)}\phi_i^{(2)}$. Using Assumption 8, it follows that

$$\widetilde{\Upsilon}'\widetilde{E}_i\widetilde{D}_i\widetilde{V}_i\widetilde{C}_i'\widetilde{\Upsilon} \;\to\; \overline{\rho}\,d_i\Upsilon^{0\prime}Q_{ZZ}\,\Upsilon^0 \qquad (77)$$

where $d_i=\sum_{j=1}^{h+1}\{\phi_i^{(3)}\}_{j,j}$ and $\{\phi_i^{(3)}\}_{j,j}$ is the $(j,j)^{th}$ element of $\{\phi_i^{(3)}\}$. Note that if $\phi_i^{(1)}$ is given by (70) then

$$d_i \;=\; \frac{(\delta\lambda_i^0)^2}{\delta\pi_{k+1}^0}, \qquad (78)$$

and if $\phi_i^{(1)}$ is given by (71) then

$$d_i=\frac{(\pi_{k+1}^0-\lambda_{i-1}^0)^2}{\delta\pi_{k+1}^0}+\frac{(\lambda_i^0-\pi_{k+\ell_i}^0)^2}{\delta\pi_{k+\ell_i+1}^0}-\pi_{k+1}^0+\pi_{k+\ell_i}^0. \qquad (79)$$

Finally, since $\widetilde{D}_i\widetilde{V}_i\widetilde{D}_i' = \omega_i^2(\phi_0\otimes Q_{ZZ})$ where ϕ_0 is defined in (73), then

$$\begin{aligned}\widetilde{E}_i\widetilde{D}_i\widetilde{V}_i\widetilde{D}_i'\widetilde{E}_i' &= \omega_i^2(\phi_i^{(2)}\otimes I_q)(\phi(1)\otimes Q_{ZZ})(\phi_i^{(2)}\otimes I_q)\\ &= \omega_i^2(\phi_i^{(3)}\otimes Q_{ZZ})\end{aligned}$$

since $\phi_i^{(2)}\phi(1)=\phi_i^{(1)}$ and $\phi_i^{(1)}\phi_i^{(2)}=\phi_i^{(3)}$. Consequently, under Assumption 8, we have

$$\widetilde{\Upsilon}'\widetilde{E}_i\widetilde{D}_i\widetilde{V}_i\widetilde{D}_i'\widetilde{E}_i'\widetilde{\Upsilon} \;\to\; \omega^2 d_i\Upsilon^{0\prime}Q_{ZZ}\,\Upsilon^0. \qquad (80)$$

Substituting from (76), (77) and (80) into (74) yields

$$V_{i,i}^{\beta}\to\{M_{ww}^{(i)}\}^{-1}\{(\delta\lambda_i^0)\rho^2-2\overline{\rho}\,d_i+\omega^2 d_i\}\Upsilon^{0\prime}Q_{ZZ}\,\Upsilon^0\{M_{ww}^{(i)}\}^{-1}.$$

Since $\Upsilon^{0\prime}Q_{ZZ}\,\Upsilon^0\{M_{ww}^{(i)}\}^{-1}=(\delta\lambda_i^0)^{-1}I_p$, and further using (28) and (29),

$$M_{ww}^{(i)}V_{i,i}^{\beta}\to\left\{\rho^2-2\beta_x^{0\prime}\gamma\frac{d_i}{\delta\lambda_i^0}-\beta_x^{0\prime}\Sigma\beta_x^0\frac{d_i}{\delta\lambda_i^0}\right\}I_p$$

and hence

$$
\begin{aligned}
AE[\xi_{2,T}] &= -\sum_{i=1}^{m+1} tr\left[(V_{i,i}^{\beta})^{1/2} M_{ww}^{(i)} (V_{i,i}^{\beta})^{1/2}\right] \\
&= -\sum_{i=1}^{m+1} tr[\, M_{ww}^{(i)} V_{i,i}^{\beta}] \\
&= -p(m+1)\rho^2 + p\sum_{i=1}^{m+1} \frac{d_i}{\delta\lambda_i^0}(2\beta_x^{0\prime}\gamma + \beta_x^{0\prime}\Sigma\beta_x^0) \\
&= -p(m+1)\rho^2 + p(\rho^2 - \sigma^2)\sum_{i=1}^{m+1} \frac{d_i}{\delta\lambda_i^0} \qquad (81)
\end{aligned}
$$

where the last expression is obtained using (27).

Part (iii): For $\xi_{3,T}$ defined by (33), consider the regime-specific errors

$$
\begin{aligned}
y_t - \widehat{x}_t'\beta_{x,i}^0 - z_{1,t}'\beta_{z,i}^0 &= (y_t - \overline{x}_t'\beta_{x,i}^0 - z_{1,t}'\beta_{z,i}^0) + (\overline{x}_t - \widehat{x}_t)'\beta_{x,i}^0 \\
&= \overline{u}_{t,i} + (\overline{x}_t - \widehat{x}_t)'\beta_{x,i}^0
\end{aligned}
$$

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

$$
ESS(T_1^0, ..., T_m^0) = \sum_{i=1}^{m+1} \sum_{t=T_{i-1}^0+1}^{T_i^0} [\overline{u}_{t,i} + (\overline{x}_t' - \widehat{x}_t')\beta_{x,i}^0]^2
$$

and

$$
ESS^e(T_1^0, ..., T_m^0) = \sum_{i=1}^{m+1} \sum_{t=T_{i-1}^0+1}^{T_i^0} \overline{u}_{t,i}^2, \qquad (82)
$$

it immediately follows that

$$
\begin{aligned}
\xi_{3,T} &= \sum_{i=1}^{m+1} \left\{ \sum_{t=T_{i-1}^0+1}^{T_i^0} \beta_{x,i}^{0\prime}(\overline{x}_t - \widehat{x}_t)(\overline{x}_t - \widehat{x}_t)'\beta_{x,i}^0 + 2\sum_{t=T_{i-1}^0+1}^{T_i^0} \overline{u}_{t,i}(\overline{x}_t - \widehat{x}_t)'\beta_{x,i}^0 \right\} \\
&= \sum_{i=1}^{m+1} (E_{2i} + 2E_{3i}) \qquad (83)
\end{aligned}
$$

where (obviously)

$$
\begin{aligned}
E_{2i} &= \sum_{t=T_{i-1}^0+1}^{T_i^0} \beta_{x,i}^{0\prime}(\overline{x}_t - \widehat{x}_t)(\overline{x}_t - \widehat{x}_t)'\beta_{x,i}^0 \qquad (84) \\
E_{3i} &= \sum_{t=T_{i-1}^0+1}^{T_i^0} \overline{u}_{t,i}(\overline{x}_t - \widehat{x}_t)'\beta_{x,i}^0. \qquad (85)
\end{aligned}
$$

From (21) and (22),

$$
\begin{aligned}
\overline{x}_t' - \widehat{x}_t' &= \tilde{z}_t'(\Theta^0 - \widehat{\Theta}_T) \\
&= -\tilde{z}_t' \left\{ \sum_{t=1}^{T} \tilde{z}_t \, \tilde{z}_t' \right\}^{-1} \sum_{t=1}^{T} \tilde{z}_t \, v_t' \qquad (86)
\end{aligned}
$$

where it is understood that $\tilde{z}_t = \tilde{z}_t(\pi^0)$. Substituting (86) into (84) and using (26), we can write

$$
\begin{aligned}
E_{2i} &= T^{-1/2} \sum_{t=1}^{T} \overline{v}_{t,i} \, \tilde{z}_t' \left\{ T^{-1} \sum_{t=1}^{T} \tilde{z}_t \, \tilde{z}_t' \right\}^{-1} T^{-1} \sum_{t=T_{i-1}^0+1}^{T_i^0} \tilde{z}_t \, \tilde{z}_t' \\
&\quad \times \left\{ T^{-1} \sum_{t=1}^{T} \tilde{z}_t \, \tilde{z}_t' \right\}^{-1} T^{-1/2} \sum_{t=1}^{T} \tilde{z}_t \overline{v}_{t,i}.
\end{aligned}
$$

From (69) and (72), it follows that

$$
AE[E_{2i}] = tr \left\{ (\phi_i^{(4)} \otimes Q_{ZZ}^{-1} \right\} lim_{T\to\infty} E \left[\left(T^{-1/2} \sum_{t=1}^{T} \tilde{z}_t \overline{v}_{t,i} \right) \left(T^{-1/2} \sum_{t=1}^{T} \tilde{z}_t \overline{v}_{t,i} \right)' \right]
$$

where $\phi_i^{(4)} = \phi_i^{(2)} \phi_0^{-1}$ and, using $\omega_i^2 = Var[\overline{v}_{t,i}]$ from (29),

$$
lim_{T\to\infty} E \left[\left(T^{-1/2} \sum_{t=1}^{T} \tilde{z}_t \overline{v}_{t,i} \right) \left(T^{-1/2} \sum_{t=1}^{T} \tilde{z}_t \overline{v}_{t,i} \right)' \right] = \omega_i^2 (\phi_0 \otimes Q_{ZZ}).
$$

Therefore

$$
\begin{aligned}
AE[E_{2i}] &= tr \left\{ \phi_0 \phi_i^{(4)} \otimes I_q \right\} \omega_i^2 \\
&= tr \left\{ \phi_i^{(2)} \otimes I_q \right\} \omega_i^2 \\
&= q \omega_i^2 b_i
\end{aligned}
$$

where $b_i = \sum_1^{h+1} \{\phi_i^{(2)}\}_{j,j}$ and $\{\phi_i^{(2)}\}_{j,j}$ is the $(j,j)^{th}$ element of $\phi_i^{(2)}$.

Also substituting (86) in the definition of (85) yields

$$
\begin{aligned}
E_{3i} &= - \sum_{t=T_{i-1}^0+1}^{T_i^0} \overline{u}_{t,i} \tilde{z}_t' \left\{ \sum_{t=1}^{T} \tilde{z}_t \, \tilde{z}_t' \right\}^{-1} \sum_{t=1}^{T} \tilde{z}_t \, v_t' \beta_{x,i}^0 \\
&= -T^{-1/2} \sum_{t=T_{i-1}^0+1}^{T_i^0} \overline{u}_{t,i} \tilde{z}_t' \left\{ \widetilde{Q}_{zz}(1) \right\}^{-1} T^{-1/2} \sum_{t=1}^{T} z_t \overline{v}_{t,i} + o_p(1).
\end{aligned}
$$

Applying similar arguments to those for E_{2i}, we obtain

$$
AE[E_{3i}] = -q \overline{\rho}_i b_i
$$

where $\overline{\rho}_i = Cov[\overline{v}_{t,i}, \overline{u}_{t,i}]$. Therefore, under Assumption 3, we have

$$AE[\xi_{3,T}] \to q[\omega^2 - 2\overline{\rho}] \sum_{i=1}^{m+1} b_i. \tag{87}$$

To complete the proof note that $\sum_{i=1}^{m+1} b_i = h+1$ and $\omega^2 - 2\overline{\rho} = -(\rho^2 - \sigma^2)$ from (28) to (29).

Part (iv): Using the definition of $\xi_{4,T}$ in (34) and also (82), it immediately follows from (27) that $E[\xi_{4,T}] = 0$. Simple algebra then yields the result given for $AE[\xi_T]$ in Theorem 3.

To establish $0 < \sum_{i=1}^{m+1} d_i \,/(\delta\lambda_i^0) \leq \min[(h+1),(m+1)]$, note first that d_i and $\delta\lambda_i^0$ $(i = 1, ..., m+1)$ are strictly positive, by definition. For a structural form regime with no intermediate reduced form breaks, $\pi_k^0 \leq \lambda_{i-1}^0 < \lambda_i^0 \leq \pi_{k+1}^0$, say, it immediately follows that $d_i/(\delta\lambda_i^0) = \{\delta\lambda_i^0\}^2/\{\delta\pi_{k+1}^0 \times \delta\lambda_i^0\} = (\delta\lambda_i^0)/(\delta\pi_{k+1}^0) \leq 1$, with equality holding if and only if $\pi_k^0 = \lambda_{i-1}^0$ and $\lambda_i^0 = \pi_{k+1}^0$. With intermediate reduced form breaks, $\pi_k^0 \leq \lambda_{i-1}^0 < \pi_{k+1}^0 < ... < \pi_{k+\ell_i}^0 < \lambda_i^0 \leq \pi_{k+\ell_i+1}^0$, say, with $\ell_i \geq 1$, then

$$\begin{aligned} d_i &= \frac{(\pi_{k+1}^0 - \lambda_{i-1}^0)^2}{\delta\pi_{k+1}^0} + \frac{(\lambda_i^0 - \pi_{k+\ell_i}^0)^2}{\delta\pi_{k+\ell_i+1}^0} + \pi_{k+\ell_i}^0 - \pi_{k+1}^0 \\ &< \pi_{k+1}^0 - \lambda_{i-1}^0 + \lambda_i^0 - \pi_{k+\ell_i}^0 + \pi_{k+\ell_i}^0 - \pi_{k+1}^0 = \delta\lambda_i^0 \end{aligned}$$

since $\pi_{k+1}^0 - \lambda_{i-1}^0 \leq \delta\pi_{k+1}^0$ and $\lambda_i^0 - \pi_{k+\ell_i}^0 \leq \delta\pi_{k+\ell_i+1}^0$, with equality if both $\lambda_{i-1}^0 = \pi_k^0$ and $\pi_{k+\ell_i+1}^0 = \lambda_i^0$. Therefore, $d_i/\delta\lambda_i^0 \leq 1$ also in this case. Summed over all $m+1$ structural form regimes, it immediately follows that

$$0 < \sum_{i=1}^{m+1} d_i \,/(\delta\lambda_i^0) \leq m+1.$$

From the perspective of the reduced form regimes, define d_j^* as follows: If reduced form regime j contains no structural form breaks, so that $\lambda_i^0 \leq \pi_{j-1}^0 < \pi_j^0 \leq \lambda_{i+1}^0$, $d_j^* = \delta\pi_j^0/\delta\lambda_i$; if reduced form regime j includes ℓ_j structural form breaks, $\lambda_i^0 \leq \pi_{j-1}^0 < \lambda_{i+1}^0 < ... < \lambda_{i+\ell_j}^0 < \pi_j^0 \leq \lambda_{i+\ell_j+1}^0$, then

$$d_j^* = \frac{(\lambda_{i+1}^0 - \pi_{j-1}^0)^2}{\delta\lambda_{i+1}^0 \times \delta\pi_j^0} + \sum_{s=2}^{\ell_j} \frac{\delta\lambda_{i+s}^0}{\delta\pi_j^0} + \frac{(\pi_j^0 - \lambda_{i+\ell_j}^0)^2}{\delta\lambda_{i+\ell_j+1}^0 \times \delta\pi_j^0}. \tag{88}$$

From these definitions, it follows that each $d_j^* \leq 1$; this is obvious for the case of no intermediate structural form breaks, while (88) implies that

$$d_j^* \leq \frac{\lambda_{i+1}^0 - \pi_{j-1}^0}{\delta\pi_j^0} + \sum_{s=2}^{\ell_j} \frac{\delta\lambda_{i+s}^0}{\delta\pi_j^0} + \frac{\pi_j^0 - \lambda_{i+\ell_j}^0}{\delta\pi_j^0} = \frac{\delta\pi_j^0}{\delta\pi_j^0} = 1$$

since $(\lambda^0_{i+1} - \pi^0_{j-1}) \leq \delta\lambda^0_{i+1}$ and $(\pi^0_j - \lambda^0_{i+\ell_j}) \leq \delta\lambda^0_{i+\ell_j+1}$. Also note that $d^*_j = 1$ in (88) when $\pi^0_{j-1} = \lambda^0_i$ and $\pi^0_j = \lambda^0_{i+\ell_j+1}$. Further, since $\lambda^0_0 = \pi^0_0 = 0$ and $\lambda^0_{m+1} = \pi^0_{h+1} = 1$, it also follows that $\sum_{i=1}^{m+1} d_i / (\delta\lambda^0_i) = \sum_{j=1}^{h+1} d^*_j \leq (h+1)$, thereby establishing the required result. $\diamond$

Proof of Theorem 4

Under H_0,

$$N_\lambda(\overline{\lambda}) = -\xi_{1,T}.$$

From (12) and (13),

$$-\xi_{1,T} \xrightarrow{d} \sum_{i=1}^{m} \max_{|k_i|} H_i(|k_i|)$$

where

$$H_i(|k_i|) = \begin{cases} -|k_i|\ \overline{a}_{i,1} + 2\,\overline{c}_{i,1}^{1/2} W_{i,1}(|k_i|), & k_i \leq 0 \\ -|k|\ \overline{a}_{i,2} + 2\,\overline{c}_{i,2}^{1/2} W_{i,2}(|k_i|), & k_i > 0 \end{cases}$$

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

$$\max_{|k_i|} H_i(|k_i|) = \overline{b}_i \sim \mathcal{B}(\mu_{i,1},\ \mu_{i,2}).$$

The desired result follows because Assumptions 1 and 3 imply independence of $\overline{b}_i$ and $\overline{b}_j$ for $i \neq j$. $\diamond$

References

Ahmed, S., Levin, A., and Wilson, B. A. (2004). 'Recent U,S. macroeconomic stability: Good policies, good practices, or good luck?', *Review of Economics and Statistics*, 86: 824–832.

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

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

——— (2003). 'Computation and analysis of multiple structural change models', *Journal of Applied Econometrics*, 18: 1–22.

Boivan, J. (2006). 'Has US monetary policy changed? Evidence from drifting coefficients and real-time data', *Journal of Money, Credit and Banking*, 38: 1149–1173.

Boivan, J., and Giannone, M. P. (2006). 'Has monetary policy become more effective?', *Review of Economics and Statistics*, 88: 445–462.

Boldea, O., and Hall, A. R. (2013). 'Estimation and inference in unstable nonlinear least squares models', *Journal of Econometrics*, 172: 158–167.

Boldea, O., Hall, A. R., and Han, S. (2012). 'Asymptotic distribution theory for break point estimators in models estimated via 2SLS', *Econometric Reviews*, 31: 1–33.

Clarida, R., Gali, J., and Gertler, M. (2000). 'Monetary policy rules and macroeconomic stability: Evidence and some theory', *Quarterly Journal of Economics*, 105: 147–180.

Duffy, J., and Engle-Warnick, J. (2006). 'Multiple regimes in US monetary policy? A nonparametric approach', *Journal of Money, Credit and Banking*, 38: 1363–1377.

Elliott, G., and Muller, U. K. (2007). 'Confidence sets for the date of a single break in linear time series regressions', *Journal of Econometrics*, 141: 1196–1218.

Ellison, M., and Sargent, T. J. (2012). 'A defence of the FOMC', *International Economic Review*, 53: 1047–1065.

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

Hall, A. R., Han, S., and Boldea, O. (2012). 'Inference regarding multiple structural changes in linear models with endogenous regressors', *Journal of Econometrics*, 170: 281–302.

Hall, A. R., Osborn, D. R., and Sakkas, N. (2013). 'Inference on structural breaks using information criteria', *Manchester School*, 81: Supplement S3, 54–81.

——— (2015). 'Structural break inference using information criteria in models estimated by two stage least squares', *Journal of Time Series Analysis*, forthcoming.

Hansen, B. E. (1997). 'Inference in TAR models', *Studies in Nonlinear Dynamics and Econometrics*, 2(1).

——— (2000). 'Sample splitting and threshold estimation', *Econometrica*, 68: 575–603.

Ling, S. (2015). 'Estimation of change points in linear and nonlinear time series models', *Econometric Theory*, forthcoming.

Ninomiya, Y. (2005). 'Information criterion for Gaussian change-point model', *Statistics & Probability Letters*, 72: 237–247.

Orphanides, A. (2004). 'Monetary policy rules, macroeconomic stability and inflation: A view from the trenches', *Journal of Money, Credit and Banking*, 36: 151–175.

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

Pesaran, M. H., and Smith, R. J. (1994). 'A generalized R2 criterion for regression models estimated by instrumental variables', *Econometrica*, 62: 705–710.

Stryhn, H. (1996). 'The location of the maximum of asymmetric two-sided Brownian motion with triangular drift', *Statistics & Probability Letters*, 29: 279–284.

White, H. (2001). *Asymptotic Theory for Econometricians*. Academic Press, New York, NY, USA, revised edn.

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

Yao, Y. C. (1988). 'Estimating the number of change-points via Schwarz' criterion', *Statistics & Probability Letters*, 6: 181–9.

Table 1: Critical values for test based on $F_\lambda(\bar{\lambda})$

m	10%	5%	1%	m	10%	5%	1%
1	5.9415	7.3581	10.5845	6	25.2819	27.9196	33.3415
2	10.2164	11.9835	15.8540	7	28.8528	31.6485	37.3694
3	14.1666	16.2043	20.5591	8	32.3859	35.3227	41.3215
4	17.9626	20.2202	24.9877	9	35.8842	38.9584	45.2137
5	21.6575	24.1175	29.2265	10	39.3541	42.5614	49.0649

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

Table 2: Performance of OLS variance estimators

		k=0			k=1			k=3		
T	α	*bias*	C_{90}	F_{90}	*bias*	C_{90}	F_{90}	*bias*	C_{90}	F_{90}
120	0.0	-2.02	85.10	83.94	-1.17	85.38	84.10	0.58	85.76	84.42
	0.1	-2.14	83.90	83.00	-1.28	84.24	83.18	0.46	84.64	83.66
	0.2	-2.25	83.00	81.86	-1.40	83.22	82.06	0.35	83.66	82.58
	0.3	-2.35	82.12	80.84	-1.50	82.38	81.00	0.24	82.80	81.52
	0.4	-2.45	81.38	79.84	-1.60	81.66	80.02	0.14	82.00	80.50
	0.49	-2.54	80.06	79.10	-1.69	80.28	79.22	0.05	80.74	79.62
T	α	*bias*	C_{90}	F_{90}	*bias*	C_{90}	F_{90}	*bias*	C_{90}	F_{90}
240	0.0	-0.89	87.78	87.02	-0.47	87.88	87.14	0.39	87.96	87.34
	0.1	-0.98	86.94	85.74	-0.56	87.00	85.86	0.30	87.18	86.06
	0.2	-1.07	85.54	84.70	-0.65	85.60	84.80	0.20	85.78	85.02
	0.3	-1.17	83.72	82.98	-0.75	83.78	83.04	0.10	84.00	83.34
	0.4	-1.25	82.10	81.38	-0.83	82.22	81.40	0.02	82.36	81.62
	0.49	-1.32	80.52	79.88	-0.90	80.62	79.98	-0.05	80.76	80.22
T	α	*bias*	C_{90}	F_{90}	*bias*	C_{90}	F_{90}	*bias*	C_{90}	F_{90}
360	0.0	-0.62	88.62	88.20	-0.34	88.62	88.32	0.23	88.66	88.40
	0.1	-0.68	87.46	87.28	-0.40	87.50	87.32	0.17	87.58	87.48
	0.2	-0.74	85.96	85.78	-0.46	86.04	85.82	0.10	86.16	85.96
	0.3	-0.82	84.14	83.72	-0.54	84.14	83.76	0.03	84.34	83.90
	0.4	-0.89	82.02	81.22	-0.61	82.06	81.34	-0.05	82.16	81.46
	0.49	-0.95	79.88	79.38	-0.67	79.92	79.40	-0.11	80.06	79.50
T	α	*bias*	C_{90}	F_{90}	*bias*	C_{90}	F_{90}	*bias*	C_{90}	F_{90}
480	0.0	-0.50	89.24	88.26	-0.29	89.32	88.26	0.13	89.50	88.36
	0.1	-0.55	88.40	87.54	-0.34	88.52	87.60	0.08	88.62	87.70
	0.2	-0.60	87.24	86.48	-0.39	87.30	86.58	0.03	87.42	86.66
	0.3	-0.66	85.44	84.02	-0.45	85.48	84.04	-0.03	85.54	84.20
	0.4	-0.72	83.10	81.36	-0.51	83.14	81.46	-0.09	83.16	81.58
	0.49	-0.77	80.80	78.88	-0.56	80.80	78.92	-0.14	80.80	79.02

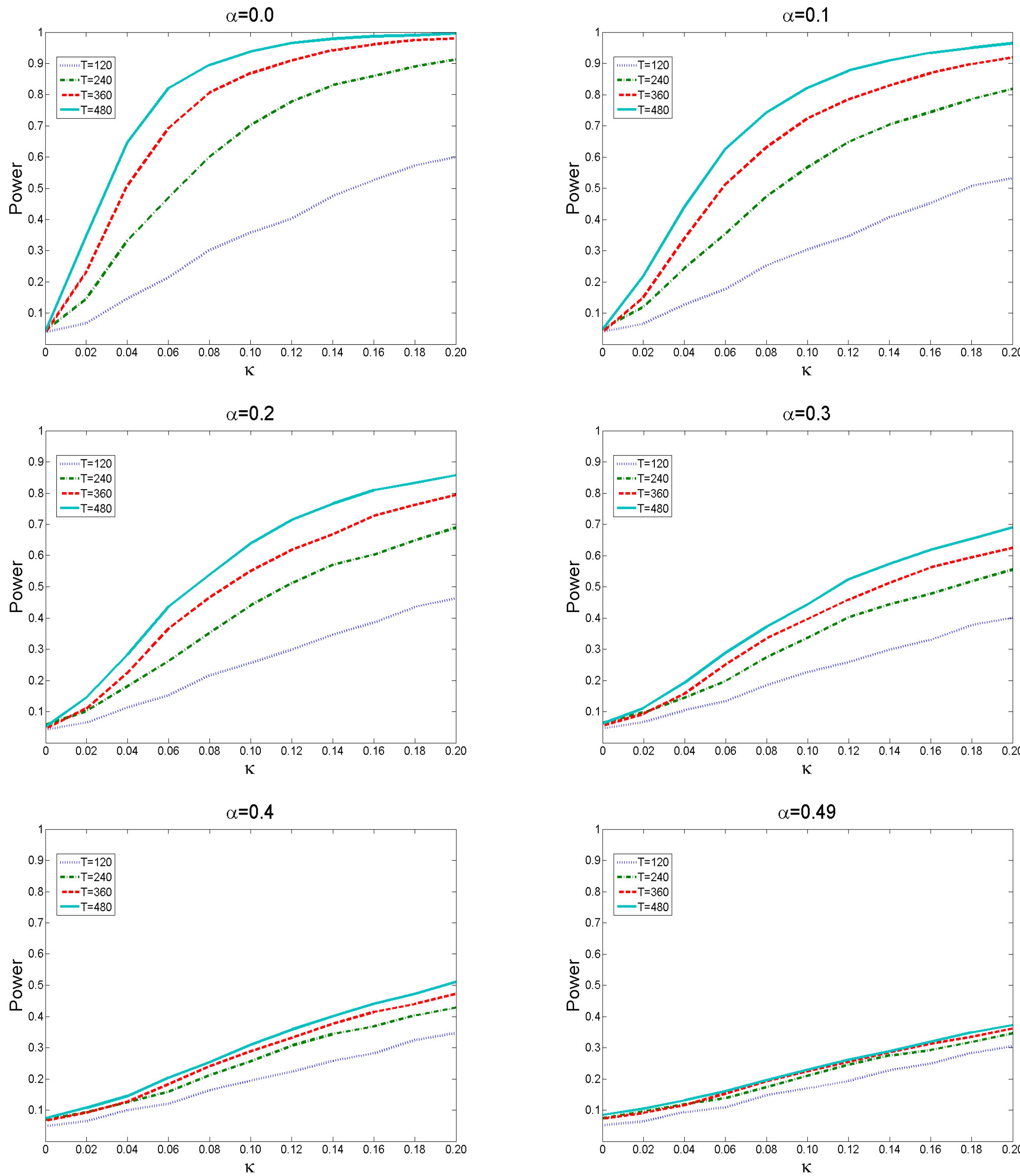
Notes: k indexes the degrees of freedom correction in the error variance estimator as described in Section 5; $bias = \hat{\sigma}_k^2 - 1$ where the true error variance is 1; C_{90} (F_{90}) denotes the empirical coverage probability closest to (furthest from) the nominal value of .90 over the intervals in (52) for μ_1, γ_1, μ_2, and γ_2.The biases and coverage probabilities are given in percentage terms.

Table 3: Estimated Monetary Policy Rules

	1968Q4-2005Q4	1968Q4-1980Q3	1980Q4-1985Q3	1985Q4-2005Q4
A. Estimated coefficients				
$\pi_{t+1\|t}$	0.41 (1.10)	0.80 (3.63)	1.68 (10.53)	0.16 (1.13)
$\widetilde{x}_{t+1\|t}$	0.23 (0.77)	0.94 (2.77)	0.01 (0.04)	0.09 (0.96)
r_{t-1}	0.65 (1.64)	0.37 (1.08)	-0.20 (1.27)	1.33 (8.54)
r_{t-2}	0.11 (0.31)	0.03 (0.17)	0.09 (0.53)	-0.43 (3.17)
c	0.12 (0.21)	1.03 (1.04)	3.75 (2.91)	0.09 (0.59)
B. Implied steady-state monetary policy responses				
$\pi_{t+1\|t}$	1.74 (1.01)	1.33 (1.53)	1.51 (5.73)	1.59 (0.74)
$\widetilde{x}_{t+1\|t}$	1.00 (0.93)	1.56 (1.12)	0.01 (0.04)	0.91 (0.43)

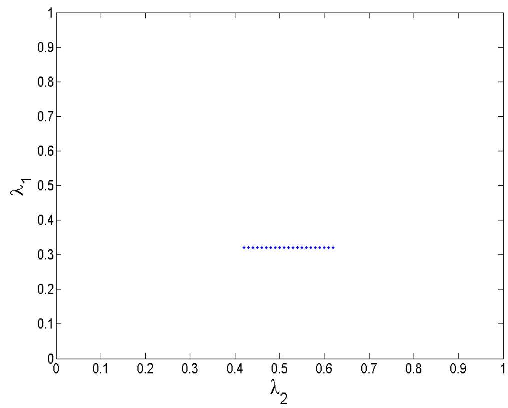
Notes: Breaks in the monetary policy rule (53) are detected using the BIC and HQIC information criteria with $(p+3)m$ effective parameters, a maximum of five breaks and a minimum of 10% of sample observations required to be in each estimated monetary policy regime. Inflation and unemployment gap forecasts are treated as endogenous in the monetary policy rule, with breaks detected separately for each reduced form equation. Figures in parentheses are t-ratios. The implied steady-state responses of monetary policy to inflation and the unemployment gap shown in Panel B are obtained from the estimated coefficients assuming constant short-term interest rates ($r_t = r_{t-1} = r_{t-2}$).

Figure 1: Power of $F_\lambda(\bar{\lambda})$ statistics for $H_0 : \bar{\lambda}_1 = 0.5 + \kappa$



Notes: The true break is at $\lambda_1^0 = 0.5$. Power is shown for $\kappa = 0, 0.02, ..., 0.2$ where $\kappa = 0$ corresponds to the true break and higher values to null hypotheses that are further away from the true break. The tests are conducted at the 5% significance level. Higher α values correspond to smaller magnitudes of the break determined by $\beta_2^0 - \beta_1^0 = [1, 1]' \times 0.3 \times 50^\alpha / T^\alpha$.

Figure 2: 99% break fraction confidence set for monetary policy application



Notes: The confidence set shows the break fraction pairs (λ_1, λ_2) for which the statistic $F_\lambda(\bar{\lambda})$ does not reject the corresponding joint null hypothesis at the 1% level, when applied to each permissible null hypothesis subject to a 15 observation minimum segment ($\epsilon = 0.10$). The $\lambda_1 = 0.32$ break fraction corresponds to 1980Q3 and is the only date of a first break that does not reject the null while λ_2 can take any value from 0.42 to 0.62, or 1984Q2 to 1991Q4.