ECONOMETRIC REVIEWS
2017, VOL. 36, NOS. 6–9, 667–698
http://dx.doi.org/10.1080/07474938.2017.1307523

The asymptotic behaviour of the residual sum of squares in models with multiple break points

Alastair R. Hall[a], Denise R. Osborn[a,b], and Nikolaos Sakkas[c]

[a]Economics, School of Social Sciences, University of Manchester, Manchester, UK; [b]Tasmanian School of Business and Economics, University of Tasmania, Hobart, Australia; [c]Department of Economics, University of Bath, Bath, England

ABSTRACT
Models with multiple discrete breaks in parameters are usually estimated via least squares. This paper, first, derives the asymptotic expectation of the residual sum of squares and shows that the number of estimated break points and the number of regression parameters affect the expectation differently. Second, we propose a statistic for testing the joint hypothesis that the breaks occur at specified points in the sample. Our analytical results cover models estimated by the ordinary, nonlinear, and two-stage least squares. An application to U.S. monetary policy rejects the assumption that breaks are associated with changes in the chair of the Fed.

KEYWORDS
Linear models; nonlinear models; ordinary least squares; parameter change; two-stage least squares; US monetary policy

JEL CLASSIFICATION
C12; C13; C26; E52

1. Introduction

There has been a considerable literature in econometrics on least square-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 et al., 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 the Akaike information criterion (AIC) 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 the estimation of three mean parameters. Kurozumi and Tuvaandorj (2011) extend Ninomiya's (2005) analysis to systems of linear regressions with exogenous

CONTACT Alastair R. Hall alastair.hall@manchester.ac.uk Economics, SoSS, University of Manchester, Manchester M13 9PL, UK.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lecr.

regressors and heteroscedasticity. Within this general case, the implications for the relative impacts of the parameter estimators and break-estimators are less easily summarized. However, if their results are specialized to the case of a single regression equation with homoscedastic errors, then their analysis reveals the same conclusions as Ninomiya's (2005) regarding the relative impacts of the estimated parameters and estimated breaks on AIC.

The paper makes three contributions. First, 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. For the linear model, this finding reproduces that of Kurozumi and Tuvaandorj (2011), but the extension to nonlinear models is new. In addition, our derivation is different to Kurozumi and Tuvaandorj (2011) and is based on a decomposition of the residual sum of squares that, we believe, provides interesting insights into the derived result. 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. Second, 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 threshold autoregressive (TAR) model, no joint test appears to have been proposed previously in the literature. This statistic can be used to construct confidence sets for the breaks. This issue has recently received some attention in linear models; e.g., see Elliott and Mueller (2007), Eo and Morley (2015), and Chang and Perron (2015). Unlike these previous methods, our approach treats the breaks jointly rather than constructing individual intervals for each break. Our third contribution is to examine breaks in the U.S. monetary policy, for which we shed new light on the common assumption that Volcker taking over as Fed chair marked an immediate policy change [Clarida et al. (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 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 prespecified points in the sample, discusses their use to construct joint confidence sets and an evaluation of the test properties via a simulation study. Section 5 examines breaks in U.S. monetary policy, while Section 6 concludes. All proofs are relegated to a mathematical appendix.

2. RSS with exogenous regressors

Our analysis of the asymptotic expectation of the residual sum of squares cover 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. Although the linear case is already covered by Kurozumi and Tuvaandorj (2011), it is pedagogically convenient to begin by first developing our results in that context for the following reasons. First, our presentation is different from Kurozumi and Tuvaandorj (2011) as it involves a decomposition of the residual sum of squares that we believe provide an interesting insight into why the asymptotic expectation takes the form it does in that model. Second, we apply a similar decomposition in all three models considered here and so an explicit presentation of the result for the linear model with exogenous regressors serves to underscore for the reader the common underlying structure that is present in all three cases. Third, this common structure is exploited to introduce new tests in Section 4 that can be applied in models estimated by OLS, NLS, and 2SLS, and so it is convenient to highlight how this structure is present in all three cases from the outset. Finally, as the linear model with exogenous regressors is the simplest—and a leading—case, it provides the most convenient framework in which to introduce the results.

2.1. Linear models

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

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

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 $0 < \lambda_1^0 < \ldots < \lambda_m^0 < 1$.[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

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

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 $(T_1, \ldots, T_m)$ such that $T_i - T_{i-1} > max\{p-1, \epsilon T\}$ for some $\epsilon > 0$ and $\epsilon < inf_i(\lambda_{i+1}^0 - \lambda_i^0)$.

The parameter ϵ is known as the trimming parameter. Assumption 2 requires that each segment considered contains sufficient observations for estimation of the model with finite T, while containing a positive fraction of the sample asymptotically. The second part of this restriction is motivated by the requirement that large sample statistical theory can be applied to deduce the limiting behavior of the relevant statistics in every subsample considered. In applications with macroeconomic data, ϵ is typically set equal to 0.2, 0.15, or 0.10, with this choice justified by simulation studies calibrated to the sample sizes and time series properties of macroeconomic data.[2] However, outside these settings, other values may be appropriate. If judged by the adequacy of the reliability of asymptotic inference, Bai and Perron (2006) provide simulation evidence that ϵ can be smaller when data are independently and identically correlated than when the data are heteroscedastic and/or serially correlated. Bai and Perron (2006) also suggest that the trimming parameter can be reduced as the sample size increases. Such a scheme might be appropriate with high-frequency data.[3] However, the appropriate choice of ϵ in this context is potentially complicated by the fact that as T increases, it seems desirable to allow for the number of break points also to increase. Killick et al. (2012) provide an algorithm for break point selection that allows the number of possible breaks to increase with the sample size, but they do not provide any formal guidance on the choice of ϵ. Thus, while clearly an important issue, it is to our knowledge an open question as to how the trimming parameter should be chosen as the sample size increases.

The estimates of $\beta^* = (\beta_1^{*\prime}, \ldots, \beta_{m+1}^{*\prime})'$ are obtained by minimizing the sum of squared residuals

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

[1] $[\cdot]$ denotes the integer part of the quantity in brackets.
[2] For example, see Bai and Perron (2006).
[3] We thank a referee for drawing our attention to this issue.

with respect to $\beta = (\beta_1{}', \ldots, \beta_{m+1}{}')'$. 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, $(\hat{T}_1, \ldots, \hat{T}_m)$, are then defined as:

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

where the minimization is taken over all possible partitions, $(T_1, \ldots, T_m)$, and the associated minimized residual sum of squares is denoted as $RSS(\widehat{T}_1, \ldots, \widehat{T}_m) = S_T\left(\widehat{T}_1, \ldots, \widehat{T}_m;\ \hat{\beta}(\{\widehat{T}_i\}_{i=1}^m)\right)$. 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 behaviors of $\hat{\lambda}$ and $\hat{\beta}(\{\hat{T}_i\}_{i=1}^m)$, together with various tests for parameter variation that arises naturally in this context.

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

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

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).

The bias term of (5) arises from (i) estimating the unknown break dates, (ii) estimating the regime-specific coefficients of (1), and (iii) the random disturbances. Reflecting these, we decompose ξ_T into three components,

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

The first component,

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

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:

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

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

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

and therefore captures the effects of the specific random disturbance sequence $\{u_t\}$. Previous structural break analyses, including Bai (1997) and Bai and Perron (1998), separately considers the roles of break date and coefficient estimation. Therefore, (8) can be viewed as explicitly recognizing a decomposition that has previously been implicit in the literature.

Let $AE[\cdot]$ denote the asymptotic expectation of the term in brackets.[4] To derive the $AE[\xi_T]$, we make the following assumption about the magnitudes of the breaks:

Assumption 3. $\beta^0_{i+1} - \beta^0_i = \theta^0_{T,i} = \theta^0_i s_T$ where $s_T = T^{-\alpha}$ for some $\alpha \in (0, 0.5)$ and $i = 1, \ldots, m$.

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

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

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

Assumption 5. (i) $E[u_t \mid \mathcal{F}_t] = 0$ 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) $V_{T,i}(r) = Var[T^{-1/2}\sum_{t=T^0_{i-1}+1}^{T^0_{i-1}+[rT]} h_t]$ is uniformly positive definite for all T sufficiently large[5], and $\lim_{T\to\infty} V_{T,i}(r) = rV_i$, uniformly in $r \in (0, \lambda^0_i - \lambda^0_{i-1})$ where V_i is a positive definite matrix of constants; (iv) $\sigma^2_i = E[u^2_t \mid \mathcal{F}_t,\ t/T \in [\lambda^0_{i-1}, \lambda^0_i)]$ is a positive finite constant for all i; (v) $\sigma^2_i = \sigma^2$, $i = 1, \ldots, m+1$.

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

Assumption 4 limits the behavior of the regressor cross product matrix and rules out trending regressors but allows regime-specific behavior. As noted by Qu and Perron (2007, p. 471), an assumption of this type is introduced with shrinking breaks, so that the limiting distribution of the break date estimator does not depend on the distribution of u_t in (1). The role of Assumption 5 is to limit the dependence structure of $\{x_t u_t\}$ and $\{u_t\}$. In particular, parts (i)–(iii) of Assumption 5 ensure that $\{x_t u_t\}$ is a short memory and satisfies a functional central limit theorem within each regime (White (2001) [Theorem 7.19]). Parts (iv) and (v) concern $\{u_t\}$ and the parts are stated separately since they are relaxed in some parts of our analysis. Qu and Perron (2007, p. 466) discuss models which satisfy assumptions of this type. Two prominent time series examples relevant to our case are (a) a regression model with (nontrending) exogenous variables whose properties are time-invariant within each regime and with short memory disturbances, such as a stationary $ARMA-GARCH$ process and (b) an individual equation from a first-order VAR system where coefficient breaks occur simultaneously across equations, with the roots of each regime-specific characteristic polynomial lying outside the unit circle. In both cases, relaxing part (v) would allow the disturbance parameters to change at the break dates. 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 1–3 are stated in the Appendix to Kurozumi and Tuvaandorj (2011).

[4]For example, see Amemiya (1985) [p. 94].

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

Lemma 1. *Under Assumptions 1, 2, 3, 4, 5(i)–(iii), and 6, there exist positive constants K_i, $i = 1, \ldots, m$, 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 $k_i \epsilon [-K_i, K_i]$, $i = 1, \ldots, m$,*

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

where

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

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

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

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

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 standard Brownian motion on $[0, \infty)$. Then, for $\alpha > 0$, $\gamma > 0$ and $k \epsilon [0, \infty)$*

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

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 and is stated in Kurozumi and Tuvaandorj (2011) [p. 221]. 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 proposition gives the form of $AE[\xi_T]$ for the linear model with exogenous regressors.

Proposition 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. Proposition 1(i) is a corollary of Kurozumi and Tuvaandorj (2011) Proposition 1.[6]

Remark 2. A comparison of $AE[\xi_{1,T}]$ and $AE[\xi_{2,T}]$ indicates that the break parameters and the regression parameters affect $AE[\xi_T]$ differently. Proposition 1(i) shows that the bias due to estimation of an additional break date increases in absolute value by $3\sigma^2$. From Proposition 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 thrice 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 et al. (2013) in the context of structural break estimation.[7]

2.2. Nonlinear models

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

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

where $f : \mathbb{R}^q \times \mathbf{B} \to \mathbb{R}$ 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 including nonlinear AR, smooth transition autoregressive and nonlinear ARCH. Boldea and Hall (2013, pp. 160–161) provide a detailed discussion.

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 1. *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) and (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 Proposition 1.*

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

---

[6] See Kurozumi and Tuvaandorj (2011) [p. 222].

[7] Under more general conditions than those imposed in Proposition 1, the form of the trade-off between estimated breaks and parameters is different, see Kurozumi and Tuvaandorj (2011) for model selection criteria tailored to this case.

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

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

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). It should be noted that the analysis of this section assumes strong instruments; some comments are made in our Section 6 about extending the analysis to the case of weak instruments.

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 $k = 1, \ldots, h$.

The reduced form of (20) can be rewritten as:

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

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 that 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 [8] $1 - 2\alpha_r > 0$

[8] 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.

and $\hat{\pi}$ is consistent for π^0. Let $\hat{x}_t$ denote the resulting fitted values, i.e.,

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

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

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 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

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

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 $\beta_i^{*\prime} = (\beta_{x,i}^{*\prime}, \beta_{z_1,i}^{*\prime})'$ 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 $\hat{\beta}(\{\hat{T}_i\}_{i=1}^m) = (\hat{\beta}_1', \ldots, \hat{\beta}_{m+1}')'$ 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 $(T_1, \ldots, T_m)$ 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 $\epsilon < inf_k(\pi_{k+1}^0 - \pi_k^0)$, $k = 1, \ldots, h$.

The generalization in Assumption 10 implies that the search for structural form breaks not only cover the relevant structural form intervals but 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) $E[h_{1,t} \mid \mathcal{F}_t] = 0$ 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) $Var[h_{1,t} \mid \mathcal{F}_t, t/T \in [\lambda_{i-1}^0, \lambda_i^0)] = \Omega_i$, 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 $i = 1, \ldots, m$.

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—i.e., $E[z_t u_t] = 0$—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.[9] 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 the structural equation,

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

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

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

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 $\bar{x}_t = E[x_t \,|z_t\,]$ embodying the true reduced form regimes when those coefficients are subject to breaks. Also define

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

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

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

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

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

With known reduced form coefficients, the quantity ρ^2 provides the asymptotic variance of the composite structural form disturbance $\bar{u}_{t,i}$ 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 $S_T\left(\widehat{T}_1, \ldots, \widehat{T}_m;\; \hat{\beta}(\{\widehat{T}_i\}_{i=1}^m)\right)$ as $RSS(\widehat{T}_1, \ldots, \widehat{T}_m)$, we consider $AE[\xi_T]$ where, analogous to (5),

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

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).

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

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

$$\xi_{1,T} = RSS(\widehat{T}_1, \dots, \widehat{T}_m; \hat{\pi}) - RSS(T_1^0, \dots, T_m^0; \pi^0) \tag{31}$$

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:

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

where $ESS(T_1^0, \dots, T_m^0)$ 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:

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

where $ESS^e(T_1^0, \dots, T_m^0)$ 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

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

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

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

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

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

Theorem 2. *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 $\pi_k^0 \leq \lambda_{i-1}^0 < \pi_{k+1}^0 < \dots < \pi_{k+\ell_i}^0 < \lambda_i^0 \leq \pi_{k+\ell_i+1}^0$, 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 4. Theorem 2 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 5. The expression for $AE[\xi_{1,T}]$ carries over from Proposition 1 and Theorem 1, 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 6. Theorem 2(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 behavior 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 2:

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 7. 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 2 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 2(ii).

Remark 8. Corollary 1 also clarifies the role of the reduced form in minimization of the 2SLS residual sum of squares in models with no breaks. When conventional 2SLS is applied to a stable structural form $(m = 0, h = 0)$, (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 9. Two further special cases of Theorem 2 are of interest; in both only the numbers of breaks matter, not their locations *per se*. First, 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 2 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 7, 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. Second, 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 10. 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 8, 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.

Remark 11. Hall, et al. (2015) proposes an information criterion for break selection in models estimated by 2SLS in which the penalty function gives the number breaks thrice the weight of the number of parameters. They report simulation evidence that suggests this enhanced weighting of the breaks improves performance over a criteria that weights the estimated breaks and parameters equally in the penalty function.

4. Testing break dates

The discussion of Sections 2 and 3 notes that $AE[\xi_{1,T}]$ exhibits similar behavior in all the models considered, and this is due to the large sample behavior 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

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

with $0 < \overline{\lambda}_1 < \ldots < \overline{\lambda}_m < 1$, against the alternative hypothesis that at least one $\lambda_i^0 \neq \overline{\lambda}_i$ $(i = 1, \ldots, m)$. 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

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

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

Theorem 3. *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 $\mu_{i,j} = 0.5a_{i,j}/c_{i,j}$ 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 12. The limiting distributions in Theorem 3 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), $\hat{Q}_\ell = (\hat{T}_\ell - \hat{T}_{\ell-1})^{-1} \sum_{t=\hat{T}_{\ell-1}+1}^{\hat{T}_\ell} x_t x_t'$, $\hat{V}_\ell = (\hat{T}_\ell - \hat{T}_{\ell-1})^{-1} \sum_{t=\hat{T}_{\ell-1}+1}^{\hat{T}_\ell} \hat{u}^2_{\ell,t} x_t x_t'$, $\hat{u}_{\ell,t} = y_t - x_t'\hat{\beta}_\ell$. 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 $\hat{\sigma}^2_\ell = (\hat{T}_\ell - \hat{T}_{\ell-1})^{-1} \sum_{t=\hat{T}_{\ell-1}+1}^{\hat{T}_\ell} \hat{u}^2_{\ell,t}$. Finally, if Assumption 5(v) holds and the error variance is constant over all regimes, an additional consistent estimator is[10]

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

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

Remark 13. 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

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

This leads to the following corollary to Theorem 3:

Corollary 2. *Under the conditions of Theorem 3, including Assumption 5(iv)–(v), we have $F_\lambda(\overline{\lambda}) \xrightarrow{d} \sum_{i=1}^m \overline{b}_i$ where $\{\overline{b}_i\}_{i=1}^m$ are mutually independent and $\overline{b}_i \sim \mathcal{B}(0.5, 0.5)$.*

Percentiles of this limiting distribution, simulated in MATLAB using 10 million replications, are presented in Table 1. Hansen (1997, 2000) develops a test of 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 $m = 1$.

[10] 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 Proposition 1.

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

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.

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:

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

where $N_\lambda(\bar{\lambda})$ is defined in (40) and $q_{m,1-\alpha}$ is the $100(1-\alpha)^{th}$ quantile of $\sum_{i=1}^{m} \bar{b}_i$ defined in Theorem 3. Clearly, with Assumptions 5(iv) and (v) imposed, the asymptotic critical values of Table 1 can be employed for $q_{m,1-\alpha}$.

The limiting distribution of the multiple break date test statistic in Theorem 3 has not, to the best of our knowledge, been obtained in the previous literature. Nevertheless, under similar assumptions to ours, Yao (1987) and Bai (1997) obtain the marginal distribution of a single break fraction estimator. This special case of the distribution in the theorem is used by Bai (1997) and also Bai and Perron (1998) to construct a confidence interval for the date of each break[11]. Since the m break date distributions are asymptotically independent, a joint test of the null hypothesis (39) could be deduced from these. However, rather than using a confidence interval approach, (44) compares *RSS* at the hypothesized break dates with the overall minimized *RSS*, providing a natural test statistic in the least square context considered here. In common with the approach of Elliott and Mueller (2007), but not that of Yao (1987), the confidence sets in (45) do not imply the dates corresponding to a specific $\bar{\lambda}_i$ are necessarily contiguous. However, unlike our methods, Elliott and Mueller (2007) results only apply to the one-break model.

Eo and Morley (2015) recently propose a method for constructing confidence intervals for individual breaks in systems of linear models with exogenous regressors based on inverting a likelihood ratio statistic and show it yields narrower confidence intervals than those proposed by Bai (1997). It would be interesting to explore an approach that takes the joint approach proposed here within the context of the likelihood ratio test-based inference, but this is beyond the scope of the current paper.

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 (54) and (55), respectively. It therefore follows from Lemmata 2 and 3, together with Definition 1, that the statistic $N_\lambda(\bar{\lambda})$ given by (40) has the limiting distribution for a nonlinear model as given in the first part of Theorem 3. Further, the imposition of Assumption 5 parts (iv) or (iv) and (v) yields the same specializations of $\mu_{i,j}$ as described in Theorem 3.

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

[11]As noted in Subsection 2.1, although they do not explicitly use a decomposition like our Eq. (8), Bai (1997) and Bai and Perron (1998) implicitly do so when considering break date estimation and inference.

(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 $F_\lambda(\bar{\lambda})$ 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 fraction hypothesis of (39) in a nonlinear regression model.

In an analogous way, Theorem 3 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 (56)–(57) or (61)–(62) 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 Subsection 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

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

Under the assumptions listed above for the 2SLS case, including Assumption 11(iv)–(v), and the H_0 of (39), $F_\lambda^{2SLS}(\bar{\lambda})$ 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.

4.3. Simulation evidence

A Monte Carlo analysis is undertaken in this section in order to examine the performance of the normalized statistic $F_{\lambda}(\bar{\lambda})$ of (44). We consider a linear model with exogenous regressors given by (1) with $m = 1$ and $p = 2$, with the 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 the analysis 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 $T = 120, 240, 360, 480$.

Following the implication of Proposition 1 that the impact of the break fraction estimation is thrice that of a regression parameter, we add $-3m$ as an additional correction in the degrees of freedom for the variance estimator that takes the, otherwise generic, form:

$$\hat{\sigma}^2 = RSS(\widehat{T}_1, \ldots, \widehat{T}_m)/(T - m(p + 1) - 3m),$$

where $RSS(\cdot)$ is defined in (6)–(7).

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 $\epsilon = 0.1$. The efficient search algorithm of Bai and Perron (2003) is employed in our analysis.

We examine the power of the test for a range of null hypotheses given by $H_0 : \bar{\lambda}_1 = 0.5 + \kappa$ 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 $\bar{\lambda}_1$ between 0.5 and 0.7. The results are based on 5,000 replications of the DGP for each case with the simulations performed in MATLAB.

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 $\kappa = 20$ ($\bar{\lambda}_1 = 0.7$) but reaches little more than 0.5 for these T and κ values when $\alpha = 0.4$.[12] 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 break 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.

5. U.S. monetary policy

The U.S. 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 those 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)

[12] 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 $\sigma^2 = 1$.

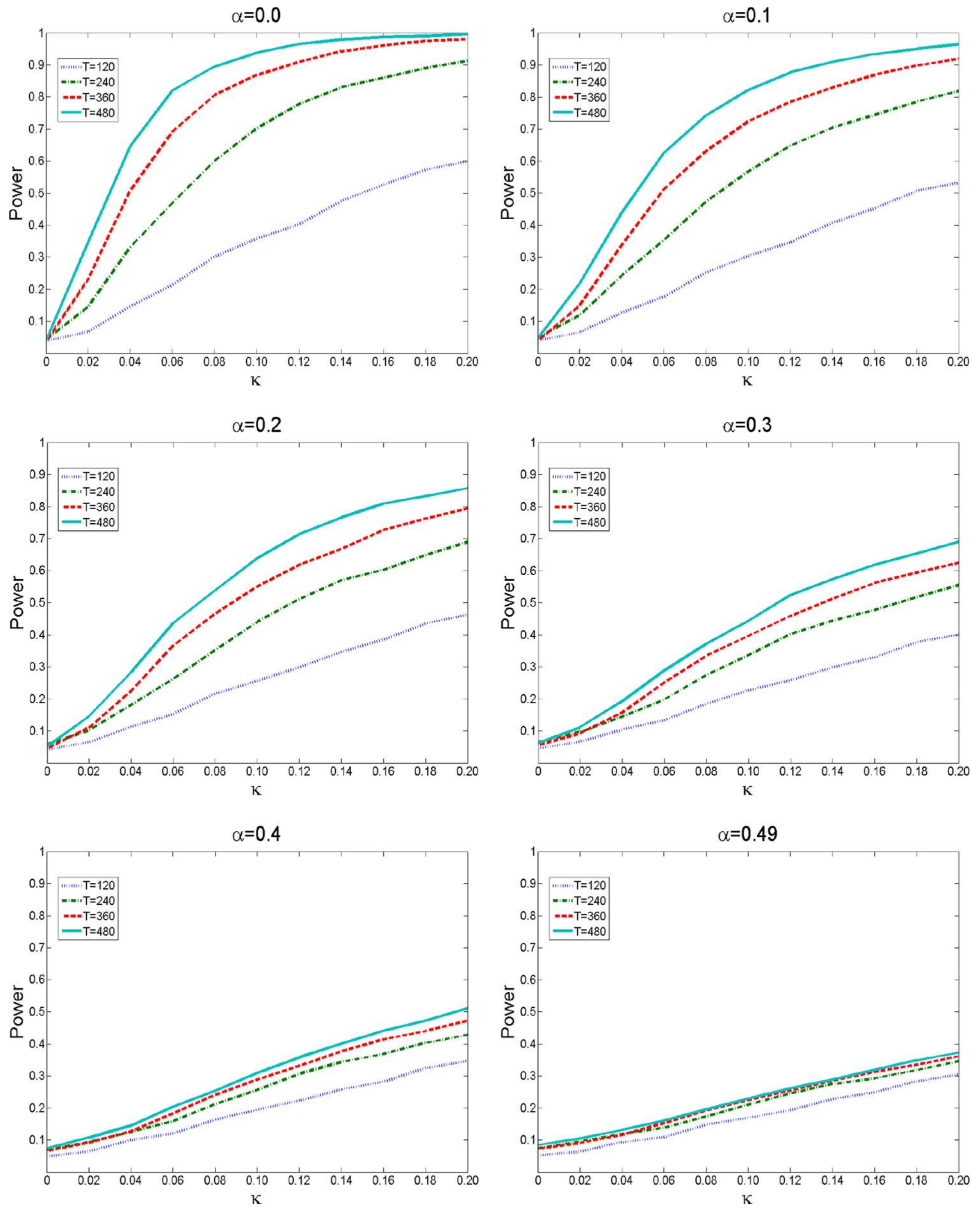

Figure 1. Power of $F_{\lambda}(\bar{\lambda})$ statistics for $H_0 : \bar{\lambda}_1 = 0.5 + \kappa$. The true break is at $\lambda_1^0 = 0.5$. Power is shown for $\kappa = 0, 0.02, \ldots, 0.2$ where $\kappa = 0$ corresponds to the true break and higher values to null hypotheses that are further away 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 $0.3 \times 50^{\alpha}/T^{\alpha}$.

split their sample in 1979, reflecting the date at which Volcker became chairman of the U.S. Federal Reserve, while Ahmed et al. (2004) use subsamples covering 1960–1979 and 1984–2002, with 1980–1983 omitted due to uncertainty about potential dates of change. In a similar vein, the seminal study of Clarida et al. (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 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 the U.S. monetary policy using the forward-looking dynamic model

$$r_t = \beta_\pi \pi_{t+1|t} + \beta_x \widetilde{x}_{t+1|t} + \beta_1 r_{t-1} + \beta_2 r_{t-2} + \beta_0 + u_t \tag{51}$$

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. We follow Orphanides (2004), who revisits the analysis of Clarida et al. (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).[13] The Greenbook provides forecasts of key variables, including inflation, output, and unemployment, which informs FOMC interest rate decisions. Although, for simplicity, our specification in (51) 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 U.S. 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 (51). 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 Gross National Product (GNP) or Gross Domestic Product (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 (51), since Ellison and Sargent (2012) argue that the FOMC may doubt the accuracy of these staff forecasts and instead favor 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 (51) 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 (10 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. Lagged inflation and growth rates are employed as instruments in line with new Keynesian models in which expectations in (51) are formed from past observations on output growth, inflation and interest rates. There is both theoretical and empirical evidence that the interest rate spread contains useful information for monetary policymakers [see, for example, Ellingsen and Söderström (2001) and Rudebusch and Wu (2008)] and hence is also included.[14]

Based on the analyses of Hall et al. (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 (51). Specifically, we employ Bayesian Information Criterion (BIC) and Hannan-Quinn Information Criterion (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 Proposition 1. [15] The maximum

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

[14] Standard diagnostics reject the null hypothesis of weak identification with our data set.

[15] Proposition 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.

Table 2. 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)

Breaks in the monetary policy rule (51) are detected using the BIC and HQIC information criteria with $m(p+3)$ effective parameters, a maximum of five breaks and a minimum of 10% of sample observations are 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 ($r_t = r_{t-1} = r_{t-2}$).

number of breaks is set to 5 in each case, with 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. Bai and Perron (2006) provide a discussion and some evidence on the choice of the trimming parameter for structural break tests, while the simulations of Hall et al. (2013) consider the choice for information criteria approaches. We use $\epsilon = 0.10$ in (51) since there are relatively few coefficients to be estimated and the disturbances are uncorrelated.

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 (51), both criteria then indicate that two breaks occur in the U.S. monetary policy. The search algorithm estimates the break dates as 1980Q3 and 1985Q3. [16]

The estimated monetary policy reaction functions are presented in Table 2, 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 into 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 U.S. 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 et al. (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 the U.S. monetary policy, including Clarida et al. (2000) and Orphanides (2004), assume that a break occurs in 1979Q2, with a new regime applying when Volcker took up appointment as the Fed chairman in the following quarter. Indeed, Clarida et al. (2000) take this further and informally investigate whether monetary 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 Greenspan taking up office in August that year.[17] Applying the tools of Section 4, we therefore test the

[16] 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 $\epsilon = 0.10$.

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

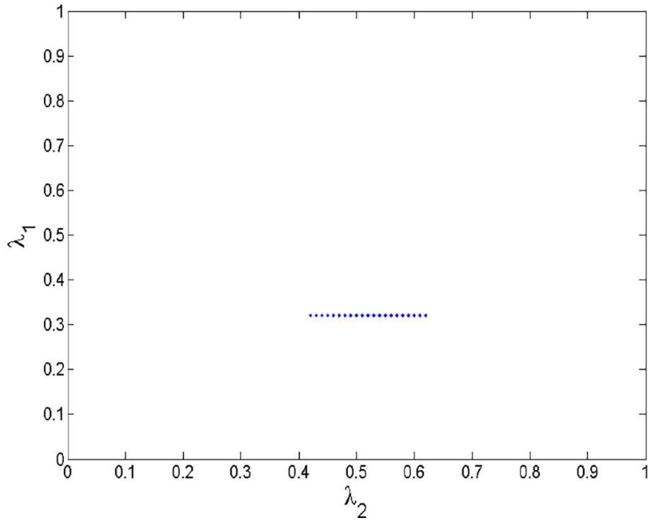

Figure 2. 99% break fraction confidence set for monetary policy application. The confidence set shows the break fraction pairs (λ_1, λ_2) for which the statistic $F_\lambda(\bar{\lambda})$ does not reject the corresponding joint null hypothesis at the 1% level, when applied to each permissible null hypothesis 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.

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. [18] 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 the U.S. 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. Although detailed analysis of the evidence is beyond the scope of this paper, it is notable that the policies now referred to as "Reaganomics," and introduced after his election as U.S. President in November 1980, included a focus on the control of inflation. Therefore, it may be that the monetary policy regime often associated with Volcker as Fed Chairman is for practical purposes very similar to that of Reagan, making the two difficult to distinguish empirically. As to the second break, our results support other studies, including Clarida et al. (2000), who suggest that the date

[18] 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.

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.

6. Concluding remarks

A considerable literature now exists concerned with least square-based estimation and testing in models with multiple discrete breaks in the parameters, see *inter alia* Bai and Perron (1998), Hall et al. (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, first, 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. Second, 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 nonstandard 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 the U.S. monetary policy. Such breaks are widely acknowledge in the literature, but are usually assumed to coincide with changes in the chair of the Federal Reserve; see, for example, Clarida et al. (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 rejected in favor of a break in late 1980. Nevertheless, we also note that monetary policy under both Volcker, as Fed Chairman, and President Reagan focused on inflation, and the start of the new regime may be difficult to determine from the data. 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 1969–2005.

Our analysis of the 2SLS case assumes that the instruments are strong and that the difference between the regimes is shrinking. It would be interesting to extend our analysis to the case where the relationship between the instruments and endogenous regressors is allowed to diminish with the sample size in the fashion of nearly weak or weak instruments. Recently, Antoine and Boldea (2016) introduce a framework in which the strength of the instruments is potentially regime dependent and implicitly controls the rate at which the breaks in the reduced form are shrinking. Interestingly, they show that if a break is located between two regimes, in one of which the instruments are nearly weak and in the other of which the instruments are weak, then the break fraction can still be consistently estimated. It would be interesting to examine the behavior of the statistics considered in our paper within Antoine and Boldea (2016) framework, however this is left to future research.

Appendix

Mathematical appendix

Proof of Proposition 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

$$AE[\xi_{1,T}] = -3m\sigma^2. \tag{52}$$

Part (ii): Using standard least square algebra,

$$\begin{aligned}\xi_{2,T} &= RSS(T_1^0, \ldots, T_m^0) - ESS(T_1^0, \ldots, T_m^0)\\ &= \sum_{i=1}^{m+1}\sum_{t=T_{i-1}^0+1}^{T_i^0}(y_t - x_t'\widehat{\beta}_i)^2 - \sum_{i=1}^{m+1}\sum_{t=T_{i-1}^0+1}^{T_i^0}(y_t - x_t'\beta_i^0)^2\\ &= -\sum_{i=1}^{m+1}(\widehat{\beta}_i - \beta_i^0)'(X_i'X_i)(\widehat{\beta}_i - \beta_i^0) \end{aligned}\tag{53}$$

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,

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

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 $E[u_t^2] = \sigma^2$.

Proof of Theorem 1. 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 piecewise geometrically ergodic process, i.e., for each subsample $[T_{j-1}^0+1, T_j^0]$, 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 $B \in \mathcal{F}^a_{-\infty}$,

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

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 $f^{(2)}_{t,i,j}$. 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 $i,j = 1,\dots p$, $\sup_{t,\beta} E|u_t f^{(2)}_{t,i,j}(\beta)|^s < \infty$.

Assumption A.4. (i) $S_T(T_1,\dots,T_m;\beta)$ has a unique global minimum at β_0 and $(T_1^0,\dots,T_m^0)$; (ii) Let $V^*_{T,i}(\beta,r) = \text{Var}\ T^{-1/2}\sum_{t=T^0_{i-1}+1}^{T^0_{i-1}+[Tr]} u_t(\beta)F_t(\beta)$. Then $V^*_{T,i}(\beta,r) \xrightarrow{p} rV^*_i(\beta)$, uniformly in $\beta\times r \in \mathbf{B}\times[0,\lambda^0_i - \lambda^0_{i-1}]$, where $V^*_i(\beta)$ is a positive definite (PD) matrix not depending on T, with $V^*_i(\beta)$ not necessarily the same for all i; (iii) Let $Q^*_{T,i}(\beta,r) = T^{-1}\sum_{t=T^0_{i-1}+1}^{T^0_{i-1}+[Tr]} F_t(\beta)F_t(\beta)'$. Then $Q^*_{T,i}(\beta,r) \xrightarrow{p} rQ^*_i(\beta)$, uniformly in $\beta\times r \in \mathbf{B}\times[0,\lambda^0_i-\lambda^0_{i-1}]$, where $Q^*_i(\beta)$ is a PD matrix; (iv) $E[f_t(\beta^0_i)] \ne E[f_t(\beta^0_{i+1})]$, for each $i = 1,\dots,m$.

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

$$a_{i,j} = \theta^{0\prime}_i Q^*_{i+j-1}(\beta^0_{i+j-1})\,\theta^0_i \tag{54}$$

$$c_{i,j} = \theta^{0\prime}_i V^*_{i+j-1}(\beta^0_{i+j-1})\,\theta^0_i \tag{55}$$

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

Proof of Theorem 2. Part (i): From the principle of least squares, $\xi_{1,T}$ as defined for 2SLS by (31) can be written as

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

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 et al. (2012), we consider scenarios where some breaks occur in the structural form but not the reduced form and where at least some breaks are common to both; the former includes the special case of a stable reduced form. These scenarios can be represented as follows.

Scenario 1: $\pi^0_j < \lambda^0_{k+1} < \dots < \lambda^0_{k+\ell} < \pi^0_{j+1}$

Scenario 2: $\pi^0_{j-1} \le \lambda^0_k < \pi^0_j = \lambda^0_{k+1} < \dots < \lambda^0_{k+\ell} \le \pi^0_{j+1}$

Scenario 1

Consider, first, a single reduced form break and m structural form breaks, with $0 \le \pi^0_1 < \lambda^0_1 < \dots < \lambda^0_m < T$, so that

$$y_t = (x_t',\ z_{1,t}')\beta^0_i + u_t, \qquad i = 1,\dots,m,\ t = T^0_{i-1}+1,\dots,T^0_i$$

$$x_t' = \begin{cases} z_t'\Delta^0_1 + v_t & t \le T^\dagger_1 \\ z_t'\Delta^0_2 + v_t & t > T^\dagger_1 \end{cases}$$

As in Boldea et al. (2012), proof of Theorem 3, the relevant intervals for the limiting behavior 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, $i = 1, \ldots, m$. Then, from Boldea et al. (2012) [Proposition 2], the minimization implies that $\xi_{1,T}$ can be written as:

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

with

$$A_i(T_i) = \theta_{T,i}^{0\,\prime} \Upsilon_2^{0\prime} \sum_{t=(T_i \wedge T_i^0)+1}^{T_i \vee T_i^0} z_t z_t' \,\Upsilon_2^0\, \theta_{T,i}^0$$

$$C_i(T_i) = (-1)^{\mathcal{I}\{T_i < T_i^0\}}\, \theta_{T,i}^{0\,\prime}\, \Upsilon_2^{0\prime} \sum_{t=(T_i \wedge T_i^0)+1}^{T_i \vee T_i^0} z_t \overline{u}_{t,i}$$

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 $k_i \epsilon [-K_i, K_i]$. Using the same arguments as Boldea et al. (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

$$a_{i,j} \;=\; \theta_i^{0\,\prime} \Upsilon_2^{0\prime} Q_{ZZ}(i+j-1)\, \Upsilon_2^0\, \theta_i^0 \tag{56}$$

$$c_{i,j} = \theta_i^{0\,\prime} \Upsilon_2^{0\prime} \Phi(i+j-1) \Upsilon_2^0\, \theta_i^0 \tag{57}$$

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

$$c_{i,j} \;=\; \rho_i^2 a_{i,j} \rightarrow \rho^2 a_{i,j} \tag{58}$$

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

Therefore, applying Lemmata 1 and 2, we have

$$\min_{|k_i|} G(|k_i|) \sim \mathcal{B}(a_{i,j}/2c_{i,j},\; a_{i,j}/2c_{i,j}) = \mathcal{B}(0.5\rho^{-2},\; 0.5\rho^{-2}), \tag{59}$$

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

$$AE[\xi_{1,T}] = -3m\rho^2. \tag{60}$$

Under the shrinking breaks Assumption 8, and with distinct reduced and structural form breaks such that $\pi_j^0 < \lambda_{k+1}^0 < \ldots < \lambda_{k+\ell}^0 < \pi_{j+1}^0$, the result immediately extends to the case where the number of reduced form breaks is $h > 1$. It also immediately 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 et al. (2012), and since the m breaks at $T_1^0, \ldots, T_m^0$ 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 (56) and (57), respectively, for $i = 2, \ldots, m$, but $a_{1,j}$ and $c_{1,j}$ are as follows:

$$a_{1,j} = \theta_1^{0\,\prime} \Upsilon_j^{0\prime} Q_{ZZ}(j)\, \Upsilon_j^0\, \theta_1^0, \qquad j = 1, 2 \tag{61}$$

$$c_{1,j} = \theta_1^{0\,\prime} \Upsilon_j^{0\prime} \Phi(j)\, \Upsilon_j^0\, \theta_1^0, \qquad j = 1, 2. \tag{62}$$

Under our assumptions, therefore, (58) applies and consequently (59) holds for a break that is common to the reduced and structural forms. Therefore (60) holds under Scenario 2.

Part (ii): From standard least square algebra,

$$\begin{aligned}\xi_{2,T} &= RSS(T_1^0, T_2^0, \ldots, T_m^0; \pi^0) - ESS(T_1^0, T_2^0, \ldots, T_m^0)\\ &= \sum_{i=1}^{m} \sum_{t=T_{i-1}^0+1}^{T_i^0} (y_t - \widehat{x}_t(\pi^0)'\widehat{\beta}_{x,i} - z'_{1,t}\widehat{\beta}_{z,i})^2 - \sum_{i=1}^{m} \sum_{t=T_{i-1}^0+1}^{T_i^0} (y_t - \widehat{x}_t(\pi^0)'\beta^0_{x,i} - z'_{1,t}\beta^0_{z,i})^2\\ &= -\sum_{i=1}^{m} (\widehat{\beta}_i - \beta_i^0)'(\widehat{W}'_i \widehat{W}_i)(\widehat{\beta}_i - \beta_i^0) \end{aligned} \tag{63}$$

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 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 $\pi_\ell^0 < \lambda \leq \pi_{\ell+1}^0$, then it follows from Assumption 14 that

$$\widetilde{Q}_{ZZ}(\lambda) = \phi(\lambda) \otimes Q_{ZZ} \tag{64}$$

where

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

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

$$\widetilde{Q}_i = \phi_i^{(1)} \otimes Q_{ZZ} \tag{65}$$

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

$$\phi_i^{(1)} = diag[0, \ldots, 0, \delta\lambda_i^0, 0, \ldots, 0]; \tag{66}$$

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

$$\phi_i^{(1)} = diag\left[0, \ldots, 0, (\pi_{k+1}^0 - \lambda_{i-1}^0), \delta\pi_{k+2}^0, \ldots, \delta\pi_{k+\ell_i}^0, (\lambda_i^0 - \pi_{k+\ell_i}^0), 0, \ldots, 0\right]. \tag{67}$$

For later reference, it is also useful to note that

$$\widetilde{Q}_{ZZ}(1) = \phi_0 \otimes Q_{ZZ} \tag{68}$$

where

$$\phi_0 = \phi(1) = diag[\delta\pi_1^0, \delta\pi_2^0 \ldots, \delta\pi_{h+1}^0]. \tag{69}$$

We now return to the proof. From the proof of Hall et al. (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 $\widetilde{\Upsilon}' = [\Upsilon_1^{0\prime}, \Upsilon_2^{0\prime}, \ldots, \Upsilon_{h+1}^{0\prime}]$. From Hall et al. (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 et al. (2012)[Theorem 8],

$$V_{i,i}^\beta = \widetilde{A}_i \left\{\widetilde{C}_i\widetilde{V}_i\widetilde{C}'_i - \widetilde{E}_i\widetilde{D}_i\widetilde{V}_i\widetilde{C}'_i - \widetilde{C}_i\widetilde{V}_i\widetilde{D}'_i\widetilde{E}'_i + \widetilde{E}_i\widetilde{D}_i\widetilde{V}_i\widetilde{D}'_i\widetilde{E}'_i\right\}\widetilde{A}'_i \tag{70}$$

and

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

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 (66) or (67), as appropriate. Also using (64),

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

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

Now consider each of the terms of (70) in turn. First, 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 (66) and $\pi_k^0 \leq \lambda_{i-1}^0 < \lambda_i^0 \leq \pi_{k+1}^0$ then

$$\widetilde{\Upsilon}'\widetilde{C}_i\widetilde{V}_i\widetilde{C}_i'\widetilde{\Upsilon} = \rho_i^2(\delta\lambda_i^0)\Upsilon_{k+1}^{0\prime}Q_{ZZ}\Upsilon_{k+1}^0 \rightarrow (\delta\lambda_i^0)\rho^2\,\Upsilon^{0\prime}\,Q_{ZZ}\,\Upsilon^0 \tag{71}$$

under Assumption 8. If $\phi_1^{(i)}$ is given by (67) then we have

$$\begin{aligned}\widetilde{\Upsilon}'\widetilde{C}_i\widetilde{V}_i\widetilde{C}_i'\widetilde{\Upsilon} &= \rho_i^2\left\{(\pi_{k+1}^0 - \lambda_{i-1}^0)\Upsilon_{k+1}^{0\prime}Q_{ZZ}\Upsilon_{k+1}^0 + \delta\pi_{k+2}^0\Upsilon_{k+2}^{0\prime}Q_{ZZ}\Upsilon_{k+2}^0 + \ldots\right.\\ &\quad \left. + (\lambda_i - \pi_{k+\ell_i}^0)\Upsilon_{k+\ell_i+1}^{0\prime}Q_{ZZ}\Upsilon_{k+\ell_i+1}^0\right\}\\ &\rightarrow (\delta\lambda_i^0)\rho^2\,\Upsilon^{0\prime}Q_{ZZ}\,\Upsilon^0\end{aligned} \tag{72}$$

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

$$\widetilde{\Upsilon}'\widetilde{E}_i\widetilde{D}_i\widetilde{V}_i\widetilde{C}_i'\widetilde{\Upsilon} \rightarrow \overline{\rho}\,d_i\Upsilon^{0\prime}Q_{ZZ}\,\Upsilon^0 \tag{73}$$

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 (66) then

$$d_i = \frac{(\delta\lambda_i^0)^2}{\delta\pi_{k+1}^0}, \tag{74}$$

and if $\phi_i^{(1)}$ is given by (67) then

$$d_i = \frac{(\pi_{k+1}^0 - \lambda_{i-1}^0)^2}{\delta\pi_{k+1}^0} + \frac{(\lambda_i^0 - \pi_{k+\ell_i}^0)^2}{\delta\pi_{k+\ell_i+1}^0} - \pi_{k+1}^0 + \pi_{k+\ell_i}^0. \tag{75}$$

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 (69), then

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

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

$$\widetilde{\Upsilon}'\widetilde{E}_i\widetilde{D}_i\widetilde{V}_i\widetilde{D}_i'\widetilde{E}_i'\widetilde{\Upsilon} \;\to\; \omega^2 d_i \Upsilon^{0\prime} Q_{ZZ}\,\Upsilon^0. \tag{76}$$

Substituting from (72), (73), and (76) into (70) 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

$$AE[\xi_{2,T}] = -\sum_{i=1}^{m+1} tr\left[(V_{i,i}^{\beta})^{1/2}M_{ww}^{(i)}(V_{i,i}^{\beta})^{1/2}\right] = -\sum_{i=1}^{m+1} tr[\,M_{ww}^{(i)}V_{i,i}^{\beta}]$$
$$= -p(m+1)\rho^2 + p\sum_{i=1}^{m+1}\frac{d_i}{\delta\lambda_i^0}(2\beta_x^{0\prime}\gamma + \beta_x^{0\prime}\Sigma\beta_x^0)$$
$$= -p(m+1)\rho^2 + p(\rho^2 - \sigma^2)\sum_{i=1}^{m+1}\frac{d_i}{\delta\lambda_i^0} \tag{77}$$

where the last expression is obtained using (27).

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

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

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

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

and

$$ESS^e(T_1^0,\ldots,T_m^0) = \sum_{i=1}^{m+1}\sum_{t=T_{i-1}^0+1}^{T_i^0}\overline{u}_{t,i}^2, \tag{78}$$

it immediately follows that

$$\xi_{3,T} = \sum_{i=1}^{m+1}\left\{\sum_{t=T_{i-1}^0+1}^{T_i^0}\beta_{x,i}^{0\prime}(\overline{x}_t - \widehat{x}_t)(\overline{x}_t - \widehat{x}_t)'\beta_{x,i}^0 + 2\sum_{t=T_{i-1}^0+1}^{T_i^0}\overline{u}_{t,i}(\overline{x}_t - \widehat{x}_t)'\beta_{x,i}^0\right\}$$
$$= \sum_{i=1}^{m+1}(E_{2i} + 2E_{3i}) \tag{79}$$

where (obviously)

$$E_{2i} = \sum_{t=T_{i-1}^0+1}^{T_i^0} \beta_{x,i}^{0\prime}(\overline{x}_t - \widehat{x}_t)(\overline{x}_t - \widehat{x}_t)'\beta_{x,i}^0 \tag{80}$$

$$E_{3i} = \sum_{t=T_{i-1}^0+1}^{T_i^0} \overline{u}_{t,i}(\overline{x}_t - \widehat{x}_t)'\beta_{x,i}^0. \tag{81}$$

From (21) and (22),

$$\begin{aligned} \overline{x}_t' - \widehat{x}_t' &= \tilde{z}_t'(\Theta^0 - \widehat{\Theta}_T) \\ &= -\tilde{z}_t' \left\{ \sum_{t=1}^{T} \tilde{z}_t\, \tilde{z}_t' \right\}^{-1} \sum_{t=1}^{T} \tilde{z}_t\, v_t' \end{aligned} \tag{82}$$

where it is understood that $\tilde{z}_t = \tilde{z}_t(\pi^0)$. Substituting (82) into (80) and using (26), we can write

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

From (65) and (68), 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 (82) in the definition of (81) 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

$$AE[\xi_{3,T}] \to q[\omega^2 - 2\overline{\rho}] \sum_{i=1}^{m+1} b_i. \tag{83}$$

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 (78), it immediately follows from (27) that $E[\xi_{4,T}] = 0$.Simple algebra then yields the result given for $AE[\xi_T]$ in Theorem 2.

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 $\delta\lambda_i^0$ $(i = 1, \ldots, m+1)$ 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, $\pi_k^0 \leq \lambda_{i-1}^0 < \pi_{k+1}^0 < \ldots < \pi_{k+\ell_i}^0 < \lambda_i^0 \leq \pi_{k+\ell_i+1}^0$, say, with $\ell_i \geq 1$, then

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

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, $\lambda_i^0 \leq \pi_{j-1}^0 < \lambda_{i+1}^0 < \ldots < \lambda_{i+\ell_j}^0 < \pi_j^0 \leq \lambda_{i+\ell_j+1}^0$, then

$$d_j^* = \frac{(\lambda_{i+1}^0 - \pi_{j-1}^0)^2}{\delta\lambda_{i+1}^0 \times \delta\pi_j^0} + \sum_{s=2}^{\ell_j} \frac{\delta\lambda_{i+s}^0}{\delta\pi_j^0} + \frac{(\pi_j^0 - \lambda_{i+\ell_j}^0)^2}{\delta\lambda_{i+\ell_j+1}^0 \times \delta\pi_j^0}. \tag{84}$$

From these definitions, it follows that each $d_j^* \leq 1$; this is obvious for the case of no intermediate structural form breaks, while (84) 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 $(\lambda_{i+1}^0 - \pi_{j-1}^0) \leq \delta\lambda_{i+1}^0$ and $(\pi_j^0 - \lambda_{i+\ell_j}^0) \leq \delta\lambda_{i+\ell_j+1}^0$. Also note that $d_j^* = 1$ in (84) when $\pi_{j-1}^0 = \lambda_i^0$ and $\pi_j^0 = \lambda_{i+\ell_j+1}^0$. Further, since $\lambda_0^0 = \pi_0^0 = 0$ and $\lambda_{m+1}^0 = \pi_{h+1}^0 = 1$, it also follows that $\sum_{i=1}^{m+1} d_i/(\delta\lambda_i^0) = \sum_{j=1}^{h+1} d_j^* \leq (h+1)$, thereby establishing the required result.

Proof of Theorem 3. 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

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

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

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

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

Acknowledgments

The paper has benefited from comments of James Nason, Mardi Dungey and participants at the Ken@75 Conference, Warwick University, and the Montreal Econometrics seminar, and seminars at Universidad Carlos III de Madrid, Korea University, the University of Durham and the University of York. The authors are also grateful to the reviewers and editors of this special issue for their constructive comments on an earlier draft of the paper.

Funding

The authors gratefully acknowledge the support of the ESRC grant RES-062-23-1351.

References

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

Amemiya, T. (1985). *Advanced Econometrics*. Cambridge, MA, USA: Harvard University Press.

Antoine, B., Boldea, O. (2016). Efficient inference with time-varying information and the New Keynesian Phillips curve. Discussion paper, Department of Economics, Simon Fraser University, Vancouver BC, Canada.

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

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

Bai, J., Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics* 18:1–22.

Bai, J., Perron, P. (2006). Multiple structural change models: A simulation analysis. In: Corbae, D., Durlauf, S. N., Hansen, B. E., eds., *Econometric Theory and Practice: Frontiers of Analysis and Applied Research*, New York: Cambridge University Press, pp. 212–237.

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

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

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

Chang, S. Y., Perron, P. (2015). A comparison of alternative methods to construct confidence intervals for the estimate of a break date in linear regression models. Discussion paper, Department of Economics, Boston University, Boston, MA.

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

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

Ellingsen, T., Söderström, U. (2001). Monetary Policy and Market Interest Rates. *American Economic Review* 91:1594–1607.

Elliott, G., Mueller, 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., Sargent, T. J. (2012). A defence of the FOMC. *International Economic Review* 53:1047–1065.

Eo, Y., Morley, J. (2015). Likelihood-ratio-based confidence sets for the timing of structural breaks. *Quantitative Economics* 6:463–497.

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

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

Hall, A. R., Osborn, D. R., Sakkas, N. (2015). Structural break inference using information criteria in models estimated by two stage least squares. *Journal of Time Series Analysis* 36:741–762.

Hansen, B. E. (1997). Inference in TAR models. *Studies in Nonlinear Dynamics and Econometrics* 2(1) https://www.degruyter.com/view/j/snde.1997.2.1/snde.1997.2.1.1024/snde.1997.2.1.1024.xml.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica* 68:575–603.

Killick, R., Fearnhead, P., Eckley, I. A. (2012). Optimal Detection of Changepoints With a Linear Computational Cost. *Journal of the American Statistical Association* 107:1590–1598.

Kurozumi, E., Tuvaandorj, P. (2011). Model selection criteria in multivariate models with multiple structural changes. *Journal of Econometrics* 164:218–238.

Ling, S. (2015). Estimation of change points in linear and nonlinear time series models. *Econometric Theory*, 32:402–430. doi:10.1017/S0266466614000863.

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

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

Qu, Z., Perron, P. (2007). Estimating and Testing Structural Changes in Multivariate Regressions. Econometrica 75:459–502. doi:10.1111/j.1468-0262.2006.00754.x.

Rudebusch, G. D., Wu, T. (2008). A macro-finance model of the term structure, monetary policy and the economy. *The Economic Journal* 118:906–926.

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*. New York, NY, USA: Academic Press (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.