Breaks in US Monetary Policy: An Information Criteria Approach to Inference with Endogenous Regressors [1] [2]

Alastair R. Hall[3]

University of Manchester

Denise R. Osborn

University of Manchester

Nikolaos D. Sakkas

University of Manchester

May 17, 2011

[1]This paper is preliminary and should not be quoted without the permission of the authors.

[2]The authors acknowledge the support of the ESRC grant RES-062-23-1351.

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

ABSTRACT

The presence and implications of breaks in US monetary policy are investigated through a "structural" equation that allows for the endogeneity of inflation and unemployment gap forecasts. The analysis establishes the consistency, for both the number of breaks and their locations within the sample, when inference is conducted through an information criteria approach with an appropriately specified penalty function. When breaks are taken into consideration in the reduced form equations for inflation and unemployment forecasts produced within the US Fed, we find that US monetary policy changes in 1980 and 1986/7. However, despite further changes in the reduced form coefficients, no break in the US monetary policy rule is detected after 1987.

Keywords: US monetary policy, structural breaks, information criteria

JEL codes: C13, C26, E52

1 Introduction

US monetary policy is widely acknowledged to have altered over the last five decades with changes in monetary policy often posited as the key explanation for changes in the properties of inflation and, sometimes, real activity. Studies that explore these issues employ a variety of techniques, with many treating the date(s) of change as known. For example, Boivan and Giannone (2006) estimate multivariate vector autoregressive (VAR) and structural models with their sample split in 1979, reflecting the date at which Paul Volcker became chairman of the US Federal Reserve, while the VAR analysis of Ahmed, Levin, and Wilson (2004) uses sub-samples covering 1960 to 1979 and 1984 to 2002, with 1980 to 1983 omitted due to uncertainty about the potential date of change. Clearly, such studies not only assume that the date of any policy (or other) change is known, at least within a narrow range, but also that no further breaks occur in the overall period under investigation.

Other studies, using a range of techniques, recognise that US monetary policy and related relationships may have undergone multiple changes over the postwar period. Among these, Cogley and Sargent (2001) employ a VAR specification with random parameter variation, while Sims and Zha (2006) use a multivariate Markov switching model. Another strand of literature draws conclusions about the role of changes in monetary policy through formal tests for structural breaks at unknown dates, with these tests typically conducted through an analysis of univariate inflation series; see, for example, Cecchetti and Debelle (2006). A more direct approach is taken by Duffy and Engle-Warnick (2006) who apply such tests to a version of the Fed's monetary policy reaction function, but they note the limitation of their analysis not being able to allow for endogeneity and hence they employ only backward-looking specifications.

The present paper examines breaks in US monetary policy through direct estimation of a 'structural equation'[1] representing the monetary policy of the Fed. Endogeneity is taken into account through the use of a Two Stage Least Squares (2SLS) approach, with the number of breaks determined using an information criterion. Although information criteria are widely applied in other model selection contexts, this paper is (to our knowledge) the first to propose their use for determining the number of breaks for an equation estimated by 2SLS. Methodologically,

[1]Here the term 'structural equation' is used in the sense of an equation from a simultaneous system with endogenous right-hand side regressors. Thus, it not necessarily represent the behavioural response of an optimizing individual economic agent.

therefore, we also establish the consistency of (certain) information criteria in this context.

Over recent years, GMM has become a very popular alternative to instrumental variable estimators (including 2SLS) for an equation with endogenous regressors. However, Hall, Han, and Boldea (2010) [HHB] show that it is not well suited to break point estimation in this context. Specifically, they show that minimizing the sum of partial generalized method of moments minimands over all partitions of the sample fails to yield consistent estimates of the break point in leading cases of interest. This arises because the GMM minimand is the square of sums, which allows effects of misspecification associated with the selection of an incorrect break point to be offset in the minimand, thus confounding the break point estimation. In contrast, the 2SLS minimand is a sum of squares and this construction offers no scope for offsetting effects of such misspecification to apply.

Hall, Han, and Boldea (2010) establish the consistency of break point estimators based on a 2SLS minimand when the true number of breaks is known. Their consistency result covers cases where the reduced form is either stable or subject to instability that takes the form of discrete changes in the parameters. However, HHB also show that the stability, or lack thereof, of the reduced form is crucial for the limiting distribution of parameter variation tests in the structural equation. Therefore, stability of the reduced form should be considered prior to that of the structural relationship. Conditional on reduced form breaks, HHB propose estimating the structural equation via 2SLS, with breaks analyzed using a strategy based on partitioning the total sample into sub-samples within which the reduced form is stable[2].

The analysis in this paper follows a similar strategy to that of HHB, and again relies on prior (and consistent) estimation of breaks in the reduced form equations. However, whereas HHB rely on hypothesis tests to select the number of breaks in the structural relationship, this paper employs an information criterion approach. In practical terms, this combines the Bai and Perron (1998) algorithm for searching for break dates, given the number of breaks, with use of an information criterion to select a preferred specification from among the set of resulting models, each estimated by 2SLS for given break dates.

An outline of the paper is as follows. Section 2 sets out a structural equation of interest in a generic context, with Section 3 then introducing the information criterion approach to break

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

estimation. The statistical properties of the estimators of both the number and location of the breaks are analyzed in this latter section, with proofs relegated to a mathematical appendix. The application to US monetary policy follows in Section 4, while a final section concludes.

2 The Structural Equation

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

$$y_t \;=\; x_t'\beta^0_{x,i} \,+\, z_{1,t}'\beta^0_{z_1,i} \,+\, u_t, \qquad i = 1, ..., m+1, \qquad t = T^0_{i-1}+1, ..., T^0_i \tag{1}$$

where $T_0^0 = 0$ and $T^0_{m+1} = T$, where T is the total sample size. Thus, y_t is the dependent variable, while 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 usual in the literature, we require the break points to be asymptotically distinct.

Assumption 1 $T^0_i = [T\lambda^0_i]$, *where* $0 < \lambda^0_1 < ... < \lambda^0_m < 1$.[3]

As a structural equation, we allow the explanatory variables, x_t, to be correlated with the errors, u_t and x_t requires a reduced form representation to be estimated using appropriate instruments. This estimation is done *a priori* in the first stage of a Two Stage Least Squares (2SLS) procedure. Furthermore, we allow for this reduced form to be subject to discrete shifts in the sample period,

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

where $T^*_0 = 0$ and $T^*_{h+1} = T$. 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 are $\Delta_0^{(i)} = (\delta^{(i)}_{1,0}, \delta^{(i)}_{2,0}, ..., \delta^{(i)}_{p_1,0})$, each with dimension $q \times p_1$, and each $\delta^{(i)}_{j,0}$ is dimension $q \times 1$, for $j = 1, ..., p_1$. The points $\{T^*_i\}$ are assumed to be generated as follows.

Assumption 2 $T^*_i = [T\pi^0_i]$, *where* $0 < \pi^0_1 < \ldots < \pi^0_h < 1$.

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

Note that the break fractions in the reduced form, $\pi^0 = [\pi_1^0, \pi_2^0, \ldots, \pi_h^0]'$, may or may not coincide with the breaks in the structural equation, $\lambda^0 = [\lambda_1^0, \lambda_2^0, \ldots, \lambda_m^0]'$. Also note that (2) can be re-written as follows

$$x_t(\pi^0)' = \tilde{z}_t(\pi^0)'\Theta_0 + v_t', \qquad t = 1, 2, \ldots, T \tag{3}$$

where $\Theta_0 = [\Delta_0^{(1)'}, \Delta_0^{(2)'}, \ldots \Delta_0^{(h+1)'}]'$. $\tilde{z}_t(\pi^0) = \iota(t,T) \otimes z_t$, $\iota(t,T)$ is a $(h+1) \times 1$ vector with first element $\mathcal{I}\{t/T \in (0, \pi_1^0]\}$, $h+1^{th}$ element $\mathcal{I}\{t/T \in (\pi_h^0, 1]\}$, k^{th} element $\mathcal{I}\{t/T \in (\pi_{k-1}^0, \pi_k^0]\}$ for $k = 1, 2, \ldots, h$ and $\mathcal{I}\{\cdot\}$ is an indicator variable that takes the value one if the event in the curly brackets occurs.

Let $\hat{\pi} = [\hat{\pi}_1, \hat{\pi}_2, \ldots, \hat{\pi}_h]'$ denote estimators of π^0. It is assumed these estimators satisfy the following condition.

Assumption 3 $\hat{\pi} = \pi^0 + O_p(T^{-1})$

This condition would be satisfied if, for example, the break dates in the reduced form are estimated by applying Bai and Perron's (1998) methodology equation by equation and then pooling the estimates of the break fractions. Let $\hat{x}_t(\hat{\pi})$ denote the resulting fitted values that is,

$$\hat{x}_t(\hat{\pi})' = \tilde{z}_t(\hat{\pi})'\hat{\Theta}_T(\hat{\pi}) = \tilde{z}_t(\hat{\pi})' \left(\sum_{t=1}^{T} \tilde{z}_t(\hat{\pi})\tilde{z}_t(\hat{\pi})' \right)^{-1} \sum_{t=1}^{T} \tilde{z}_t(\hat{\pi}) x_t' \tag{4}$$

where $\tilde{z}_t(\hat{\pi})$ is defined analogously to $\tilde{z}_t(\pi^0)$ based on the estimator of the true break points in the reduced form.

To facilitate our analysis we impose the following assumptions:

Assumption 4 *(i) $h_t = (u_t, v_t')' \otimes z_t$ is an array of real valued $n \times 1$ random vectors (where $n = (p+1)q$) defined on the probability space $(\Omega, \mathcal{F}, P)$, $V_T = Var[\sum_{t=1}^{T} h_t]$ is such that $diag[\xi_{T,1}^{-1}, \ldots, \xi_{T,n}^{-1}] = \Xi_T^{-1}$ is $O(T^{-1})$ where Ξ_T is the $n \times n$ diagonal matrix with the eigenvalues $(\xi_{T,1}, \ldots, \xi_{T,n})$ of V_T along the diagonal; (ii) $E[h_{t,i}] = 0$ and, for some $d > 2$, $\|h_{t,i}\|_d < \Gamma < \infty$ for $t = 1, 2, \ldots$ and $i = 1, 2, \ldots n$ where $h_{t,i}$ is the i^{th} element of h_t; (iii) $\{h_{t,i}\}$ is near epoch dependent with respect to $\{g_t\}$ such that $\|h_t - E[h_t|\mathcal{G}_{t-m}^{t+m}]\|_2 \leq \nu_m$ with $\nu_m = O(m^{-1/2})$ where $\mathcal{G}_{t-m}^{t+m}$ is a sigma- algebra based on $(g_{t-m}, \ldots, g_{t+m})$; (iii) $\{g_t\}$ is either ϕ-mixing of size $m^{-d/(2(d-1))}$ or α-mixing of size $m^{-d/(d-2)}$; (iv) $V_T(r) = Var[T^{-1/2} \sum_{t=1}^{[Tr]} h_t]$ satisfies $V_T(r) \rightarrow rV$ uniformly in $r \in [0,1]$ where V is a pd matrix.*

Assumption 5 *$Var[u_t] = \sigma_u^2$, $Cov[u_t, v_t] = \Sigma_{uv}$, and $Var[v_t] = \Sigma_v$, for all t.*

Assumption 6 *$rank\{\Upsilon_i^0\} = p$ where $\Upsilon_i^0 = \left[\Delta_0^{(i)}, \Pi\right]$, 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 7 *For $\sharp = 0, *$, there exists an $l_\sharp > 0$ such that for all $l > l_\sharp$, the minimum eigenvalues of $A_{il} = (1/l)\sum_{t=T_i^\sharp+1}^{T_i^\sharp+l} z_t z_t'$ and of $\bar{A}_{il} = (1/l)\sum_{t=T_i^\sharp-l}^{T_i^\sharp} z_t z_t'$ are bounded away from zero for all $i = 1, ..., \nu^\sharp + 1$ where $\nu^0 = m$ and $\nu^* = h$.*

Assumption 8 *$T^{-1}\sum_{t=1}^{[Tr]} z_t z_t' \overset{p.u}{\to} Q_{ZZ}(r)$ uniformly in $r \in [0,1]$ where $Q_{ZZ}(r)$ is positive definite for any $r > 0$ and strictly increasing in r. $Q_{ZZ}(r) - Q_{ZZ}(s)$ is positive definite for any $r > s$.*

Assumption 4 allows substantial dependence and heterogeneity in $(u_t, v_t')' \otimes z_t$ but at the same time imposes sufficient restrictions to deduce a Functional Central Limit Theorem for $T^{-1/2}\sum_{t=1}^{[Tr]} h_t$; see Wooldridge and White (1988). This assumption also contains the restrictions that the implicit population moment condition in 2SLS is valid - that is $E[z_t u_t] = 0$ - and the conditional mean of the reduced form is correctly specified. Assumption 6 restricts the unconditional variance and covariances of the structural equation and reduced form errors to be constant over time. (We conjecture that this asusmption can be relaxed but this issue is still under investigation.) Assumption 6 implies the standard rank condition for identification in IV estimation in the linear regression model[4] because Assumptions 4(ii), 6 and 8 together imply that

$$T^{-1}\sum_{t=[sT]+1}^{[Tr]} z_t[x_t', z_{1,t}'] \overset{p}{\to} [Q_{ZZ}(r) - Q_{ZZ}(s)]\Upsilon_0 = Q_{Z,[X,Z_1]}(r,s) \text{ uniformly in } r > s+\epsilon, r, s \in [0,1] \tag{5}$$

where $Q_{Z,[X,Z_1]}(r,s)$ has rank equal to p for any r, s (satisfying the above conditions). Note this assumption implies $q \geq p$. Assumption 7 requires that there be enough observations near the true break points in either the structural equation or reduced form so that they can be identified and is analogous to the extension proposed in Bai and Perron (1998) to their Assumption A2.

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

3 Consistency of an Information Criterion

Suppose now that a researcher knows neither the number nor the location of the breaks in the structural equation. Consider the case where n breaks are estimated at $\tau(n) = [\tau_1, \tau_2, \ldots, \tau_n]'$ with $0 < \tau_1 < \tau_2 < \ldots < \tau_n < 1$, $\tau_0 = 0$, and $\tau_{n+1} = 1$. Then, the second stage of 2SLS can begin with the estimation of (1) via OLS for each possible n-partition of the sample that is,

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

where $T_i = [\tau_i T]$, and the regressors x_t are estimated using the fitted values of the first stage of 2SLS, $\hat{x}_t(\hat{\pi})$. We further assume that

Assumption 9 *Equation (6) is estimated over all partitions* $(T_1, ..., T_n)$ *such that* $T_i - T_{i-1} > max\{q-1, \epsilon T\}$ *for some* $\epsilon > 0$ *and* $\epsilon < inf_i(\lambda^0_{i+1} - \lambda^0_i)$ *and* $\epsilon < inf_j(\pi^0_{j+1} - \pi^0_j)$.

Assumption 9 requires that each segment considered in the estimation contains a positive fraction of the sample asymptotically; in practice ϵ is chosen to be small in the hope that the last part of the assumption is valid. Letting $\beta^{*\prime}_i = (\beta^{*}_{x,i}{}', \beta^{*}_{z_1,i}{}')'$, for a given n-partition, the estimates of $\beta^* = (\beta^{*\prime}_1, \beta^{*\prime}_2, ..., \beta^{*}_{n+1}{}')'$ are obtained by minimizing the sum of squared residuals

$$S_T(T_1, ..., T_n; \beta) = \sum_{i=1}^{n+1} \sum_{t=T_{i-1}+1}^{T_i} \left\{ y_t - \hat{x}_t(\hat{\pi})'\beta_{x,i} - z'_{1,t}\beta_{z_1,i} \right\}^2$$

with respect to $\beta = (\beta_1{}', \beta_2{}', ..., \beta_{n+1}{}')'$. We denote these estimators by $\hat{\beta}(\tau(n))$. The estimators of the break points, $(\hat{T}_1, ..., \hat{T}_n)$, are then defined as

$$\hat{\tau}(n) = (\hat{T}_1, ..., \hat{T}_n) = \arg\min_{T_1, ..., T_n} S_T\left(T_1, ..., T_n; \hat{\beta}(\tau(n))\right) \tag{7}$$

where the minimization is taken over all possible partitions, $(T_1, ..., T_n)$. The 2SLS estimates of the regression parameters, $\hat{\beta}(\hat{\tau}(n)) = (\hat{\beta}'_1, \hat{\beta}'_2, ..., \hat{\beta}'_{n+1})'$, are the regression parameter estimates associated with each of the estimated partitions.

The estimators $\hat{\tau}(n)$ and $\hat{\beta}(\hat{\tau}(n))$ are calculated conditional on n. In practice, n is often unknown *a priori*. Hall, Han, and Boldea (2010) propose a method for estimation of n based on the sequential application of certain test statistics for parameter variation. Here we consider an alternative approach based on minimization of the following information criterion (IC),

$$IC\left(\tau(n); n, \hat{\pi}\right) = \ln\left[\hat{\sigma}^2(\tau(n); n, \hat{\pi})\right] + K(n, T), \tag{8}$$

where

$$\hat{\sigma}^2(\tau(n); n, \hat{\pi}) = (T-p)^{-1} RSS(\tau(n); n, \hat{\pi}), \tag{9}$$

$$RSS(\tau(n); n, \hat{\pi}) = \sum_{j=1}^{n} RSS_j(\tau(n); n, \hat{\pi}), \tag{10}$$

$$RSS_j(\tau(n); n, \hat{\pi}) = \sum_{t=[\tau_{j-1}T]+1}^{[\tau_j T]} \left\{ y_t - \hat{x}_t(\hat{\pi})' \hat{\beta}_{x,i} - z'_{1,t} \hat{\beta}_{z_1,i} \right\}^2, \tag{11}$$

and $K(n, T)$ is a deterministic penalty term governed by the following Assumption,

Assumption 10 *$K(n, T) = o(1)$ as $T \to \infty$, it is a strictly increasing function of n, and $TK(n, T) \to \infty$ as $T \to \infty$.*

Then, the estimated number of breaks, denoted $\hat{n}$, is the value that minimizes the IC, that is

$$\hat{n} = argmin_{n \in \mathcal{N}} IC\left(\tau(n); n, \hat{\pi}\right). \tag{12}$$

where $\mathcal{N} = \{0, 1, \ldots, N\}$. The associated estimators of the break locations are $\hat{\tau}(\hat{n})$. N is the maximum number of breaks considered and we assume this is large enough to ensure $m \in \mathcal{N}$:

Assumption 11 *$N \geq m$.*

The proof of consistency of our method rests on the limiting properties of $RSS(\tau(n); n, \hat{\pi})$. The following lemma presents the limiting behaviour of $RSS_j(\tau(n); n, \hat{\pi})$.

Lemma 1 *Let y_t be generated by (1), x_t be generated by (2), $\hat{x}_t(\hat{\pi})$ be generated by (4) and Assumptions 1-9 hold. Then, for segment j of the data, $t = [\tau_{j-1}T] + 1, \ldots, [\tau_j T]$,*

(i) If $\nexists\, i$ such that $\lambda_i^0 \in [\tau_{j-1}, \tau_j]$ then

$$T^{-1} RSS_j(\tau(n); n, \hat{\pi}) \stackrel{p.u}{\to} (\tau_j - \tau_{j-1}) \Gamma_i.$$

(ii) If there exists i and $\kappa > 0$ such that $\lambda_i^0, \lambda_{i+1}^0, \ldots, \lambda_{i+\kappa}^0 \in [\tau_{j-1}, \tau_j]$ then

$$\begin{aligned} T^{-1} RSS_j(\tau(n); n, \hat{\pi}) \stackrel{p.u}{\to} & (\lambda_i^0 - \tau_{j-1}) \Gamma_i + (\lambda_{i+1}^0 - \lambda_i^0) \Gamma_{i+1} + \ldots \\ & + (\lambda_{i+k}^0 - \lambda_{i+\kappa-1}^0) \Gamma_{i+\kappa} + (\tau_j - \lambda_{i+\kappa}^0) \Gamma_{i+\kappa+1} + F. \end{aligned}$$

where $\Gamma_i = \sigma_u^2 + 2\Sigma_{uv}\beta_{x,i}^0 + \beta_{x,i}^{0\prime}\Sigma_v\beta_{x,i}^0$. $\overset{p.u}{\to}$ *denotes limit in probability, that exists uniformly in a segment defined by* $\tau_{j-1}+\epsilon < \tau_j$*, for* $\epsilon > 0$ *and* $\tau_{j-1}, \tau_j \in [0,1]$. F *is a positive constant (that is defined in the proof) which depends on* τ_{j-1}*,* τ_j*, certain limit matrices and the parameters of the model.*

Lemma 1 demonstrates the impact of neglected breaks on the residual sum of squares in segment j. Part (i) states that if there are no neglected breaks then $T^{-1}RSS_j(\tau(n); n, \hat{\pi})$ converges to the (scaled) variance $(\tau_j - \tau_{j-1})\Gamma_i$; part (ii) shows that if there are neglected breaks then $T^{-1}RSS_j(\tau(n); n, \hat{\pi})$ converges to its scaled variance plus a positive constant. Notice that the scaled variance in question is that of $u_t + \beta_{x,i}^{0\prime}v_t$, and this reflects both the error u_t and the measurement error inherent in the substitution of $\hat{x}_t(\hat{\pi})$ for x_t.

Given the additivity of $RSS(\cdot)$ in $RSS_j(\cdot)$, the results in Lemma 1 can be used to deduce the limiting behaviour of $T^{-1}RSS(\cdot)$ for any partition . For any partition with no neglected breaks, $T^{-1}RSS(\cdot)$ converges to $\Gamma = \sum_{i=1}^n \Gamma_i$ and for any partition with at least one neglected break $T^{-1}RSS(\cdot)$ converges to $\Gamma + \xi$, $\xi > 0$. This behaviour, combined with Assumptions 10 and 11 implies the consistency of $[\hat{n}, \hat{\tau}(\hat{n})]$. This is stated formally in the following theorem.

Theorem 1 *Under Assumptions 1-11,*

$$[\hat{n}, \hat{\tau}(\hat{n})] \overset{p.u}{\to} [m, \lambda^0]$$

where $\lambda^0 = [\lambda_1^0, \ldots, \lambda_m^0]'$ *is the collection of the true break fractions in (1).*

Remark 1: To implement the estimation procedure, it is necessary to pick a penalty term that satisfies Assumption 10. A natural choice is $K(n,T) = (n+1)p\,ln(T)/T$, which is associated with BIC (Schwarz (1978)), because this choice has been found to work well in other settings. Another possibility is $K(n,T) = (n+1)p\,ln[ln(T)]/T$ which is the choice associated with HQIC (Hannan and Quinn (1979)). The choice associated with AIC (Akaike (1974)) does not satisfy Assumption 10 and its use would yield an estimator that would have a zero probability of choosing too few breaks but a non-zero probability of choosing too many breaks in the limit.

Remark 2: HHB propose a methodology for estimation of m based on the sequential application of tests for various forms of parameter variation. If these tests are performed with a fixed significance level then the resulting estimator of m has a zero probability of underfitting but a non-zero probability of overfitting in the limit due to the non-zero probability of type one errors inherent in the decision rules for the tests. Simulation results in HHB suggest that the tendency to overfit can be substantially reduced by using 1% significance levels; nevertheless, the resulting estimator of the number of breaks is not consistent. This may be seen as an advantage of the IC approach.

Remark 3: A further difference between HHB's approach and the IC approach is in terms of the assumptions about the limiting behaviour of the instrument cross-product matrix. The theory underlying certain tests employed in HHB's methodology requires the standardized partial sum instrument cross-product matrix to be linear in the sampling fraction within the assumed regimes under the appropriate null that is, $T^{-1}\sum_{t=T_{i-1}^0+1}^{T_{i-1}^0+[rT]} z_t z_t' \xrightarrow{p} rQ_i$, uniformly in $r \in (0, \lambda_i^0 - \lambda_{i-1}^0]$, where Q_i is a pd matrix of constants. This rules out changes in the mean and variance of the instruments at different times from the changes in the structural parameters. This assumption is more restrictive than Assumption 8. Thus the IC approach is potentially more robust to such changes in the behaviour of z_t (in the limit).

4 US Monetary Policy

Breaks in US monetary policy are examined through a "Taylor rule" type of specification. However, since monetary policy is conducted in real time, it is important that such evaluations consider the data as available at the time when monetary policy decisions are taken. This section first discusses the form of the monetary policy rule we employ, including data considerations, before turning to our results.

4.1 Monetary Policy Rules and Data

We examine US monetary policy through a modified version of the "Taylor rule":

$$r_t = \delta_\pi \pi_{t+i|t} + \delta_y \widetilde{y}_{t+i|t} + \delta_1 r_{t-1} + \delta_2 r_{t-2} + c + \varepsilon_t \tag{13}$$

where r_t is the actual Federal Funds rate while $\pi_{t+i|t}$ and $\widetilde{y}_{t+i|t}$ are forecasts of inflation and a proxy for the output gap, respectively. Unfortunately, however, historical real-time output gap forecasts from the Fed are available only from late 1987, yielding a sample that is too short for a structural breaks analysis. Consequently, following Boivin (2006), we employ a real-time unemployment gap measure as a proxy for the output gap in (13).

Our empirical analysis employs real-time Greenbook data, namely data prepared within the Fed in preparation for each meeting of the Federal Open Market Operations Committee (FOMC). Although FOMC meetings are not quarterly, we follow the usual convention of using the meeting closest to the middle of the quarter as relating to the specific quarter. Greenbook inflation and output growth data are currently available from 1968Q4 to 2005Q4, and the series used are based on forecasts of percentage inflation, as measured by the relevant GNP or GDP deflator, and forecasts for unemployment. More explicitly, and as in Boivin (2006), $\widetilde{y}_{t+i|t}$ is measured as the natural rate of unemployment minus the Fed's forecast for quarter $t+i$, where the natural rate is computed as an average of the historical unemployment rate over data as available at t. The interest rate series is the average actual federal Funds rate for the third month of the quarter, with the timing of the third month chosen to ensure that this reflects any monetary policy change effected during that quarter.

Our analysis recognises that $\pi_{t+i|t}$ and $\widetilde{y}_{t+i|t}$ $(i = 1, 2)$ are endogenous to the Fed's monetary policy decision. Although the Greenbook forecasts are produced as input to the interest rate decision, views about the trajectory of the economy are closely bound with these decisions and hence the exogeneity of the forecasts appears *a priori* implausible.

To implement the 2SLS methodology for breaks, the stability of the reduced form equations needs to be be examined. The instrument set used for this purpose includes lags on each of π_t, $\widetilde{y}_t$, and r_t, together with observed GNP/GDP growth (denoted Δy_t) and the interest rate on long-term (ten year) bonds (lr_t). All data are real-time, with the Greenbook data from the Philadelphia Fed's real-time database used for $\pi_{t+i|t}$ and to construct $\widetilde{y}_{t+i|t}$. Figures 1 and 2 show the inflation and unemployment gap forecast series used in our analysis, with Figure 3 illustrating the movements in both short- and long-run interest rates.

4.2 Results

The first step in the analysis is the to examine the stability of the Greenbook inflation and implied unemployment gap forecasts, with results presented in Appendix Tables A.1 and A.2 for the two-quarter ahead forecasts ($\pi_{t+2|t}$ and $\widetilde{y}_{t+2|t}$, respectively). All results, for each reduced form equation and also for the monetary policy rule (13), are obtained allowing a maximum of five breaks and with a minimum of 15% of the total sample observations required to be in each regime, with the HQ criterion used for break detection. Due to the serial correlation that results with overlapping forecast horizons, HAC standard errors are used to compute the t-ratios shown in Appendix Tables A.1 and A.2.

Breaks are detected in the Fed's reduced form inflation forecast equation in 1975 and 1981, with a third break in either 1986 (one quarter ahead) or 1994 (two quarters ahead). Perhaps the most notable feature of the coefficient changes in the inflation forecast equations in Appendix Table A.1 is the greater role played by the long-term bond rate after 1981. This contrasts with the first regime (to 1975), where the only individually significant explanatory variable is the lagged observed inflation value. Breaks in the unemployment gap forecasts are also uncovered in 1975 and 1981, with other breaks in 1991 and (for two quarters ahead forecasts only) at the end of 1999. At least from the mid-1970s, unemployment gap forecasts are explained primarily by lagged gap observations and past output growth, although the relative contributions of these variables apparently changes over time (see Appendix Table A.2).

Our main focus is, however, the monetary policy rule of (13), for which results are presented in Table 1 for both $i = 1$ and 2. The Hannan-Quinn criterion detects breaks in 1980Q3 and in either 1986Q1 or 1987Q2, which coincide well with the middle period being the "non-standard" monetary policy regime when nonborrowed reserves were targeted. Overall, the results are robust to the choice of one or two quarter ahead forecasts, although it is notable that in the final regime (from 1986/7) the estimated Fed responses to future inflation and unemployment are stronger and more significant when the longer forecast horizon is used.

There are a number of interesting features of the estimated coefficients in Table 1. Firstly, interest rate smoothing appears to be a feature primarily of the period after 1986/7, with neither lagged interest rate coefficient individually significant at the 5 percent before this period. In other respects, the pre-1980 and post-1986/7 regimes are broadly similar in that monetary policy reacts to future values of both inflation and the unemployment gap. In the middle regime, however, real activity plays no role.

A further comparison is provided by the implied steady-state monetary policy responses in Panel B of Table 1. While there is evidence that the relative weights of future inflation and the future unemployment gap are interchanged in the period from 1986/7 compared with pre-1980, the largely constant response of US monetary policy to future inflation is striking.

The break dates detected for the monetary policy rule of the US Fed are effectively in line with those detected by Duffy and Engle-Warnick (2006). However, our analysis is a very substantial improvement by allowing for the endogeneity inherent in the use of a forward-looking monetary policy rule. It is also notable that our finding of no break after 1987 in either specification, and despite evidence that the reduced form relationships change after this date, supports the proposition that the monetary policy objectives of the US Fed have remained constant since the middle of the 1980s.

5 Conclusions

This paper examines the nature of changes in US monetary policy, allowing for endogeneity in the forecasts embodied in interest rate decisions and with breaks examined through an information criteria approach applied to both underlying reduced form equations and the monetary policy rule of promary interest. Assumptions about the relative timing of breaks in these relationships are avoided by considering each equation separately. Our findings confirm the assumption of much earlier literature (including, among many examples, Ahmed *et al.*, 2004, and Boivan and Giannone, 2006) that US monetary policy has experienced regime switches over the post war period. Indeed, although they employ different techniques to ours, Duffy and Engle-Warnick (2006) also uncover monetary policy breaks at similar dates to ours. On the other hand, our results do not confirm that monetary policy switches are associated with changes in the inflation process (as suggested by Ahmed *et al.* (2004), Zhang *et al.* (2008) and others). To be more precise, by comparing the break dates found in the reduced and structural form relationships, our findings imply that the inflation forecasting process of the US Fed may have changed separately from changes in the Fed's monetary policy responses.

In order to conduct our analysis, the paper develops an information criteria approach to break detection in an equation with endogenous regressors. This provides an alternative, and potentially more flexible, approach to that of Hall, Han and Boldea (2010) based on hypothesis testing. As with that approach, consistent inference on the number and dates of breaks in the structural relationships of interest relies on breaks in the reduced form equations being detected and taken into account. To our knowledge, information criteria have not been previously explored in the present context. Nevertheless, not all information criteria deliver consistent inference, with BIC (Schwarz (1978)) and HQIC (Hannan and Quinn (1979)) satisfying the condition required on the penalty term for consistency to apply, whereas AIC (Akaike (1974)) does not.

However, this paper also raises a number of issues which are the subject of our on-going research. These include investigation, through Monte Carlo analysis, of the relative performance of information criteria and hypothesis testing approaches to structural breaks in equations with endogenous regressors, and extension of the analysis here to allow the researcher to investigate whether restrictions apply to parameters across regimes.

Table 1. Estimated Monetary Policy Rules

	1968Q4-1980Q2	1980Q3-1985Q4	1986Q1-2005Q4	1968Q4-1980Q2	1980Q3-1987Q1	1987Q2-2005Q4
	One-quarter ahead forecast values			*Two-quarter ahead forecast values*		
A. Estimated coefficients						
$\pi_{t+i\|t}$	0.80 (4.36)	1.66 (11.35)	0.28 (2.17)	0.76 (4.92)	1.83 (9.94)	0.44 (3.65)
$\widetilde{y}_{t+i\|t}$	1.02 (3.27)	0.03 (0.11)	0.18 (2.07)	1.02 (3.31)	-0.48 (0.22)	0.30 (3.71)
r_{t-1}	0.38 (1.29)	-0.19 (1.25)	1.24 (7.46)	0.40 (1.33)	-0.07 (0.38)	1.27 (8.65)
r_{t-2}	0.05 (0.40)	0.09 (0.50)	-0.39 (2.76)	0.15 (0.93)	0.12 (0.62)	-0.49 (3.98)
c	0.88 (1.05)	3.92 (3.48)	0.04 (0.30)	0.54 (0.60)	1.03 (0.97)	-0.03 (0.24)
B. Implied steady-state monetary policy responses						
$\pi_{t+i\|t}$	1.41	1.49	1.86	1.67	1.94	1.99
$\widetilde{y}_{t+i\|t}$	1.80	0.02	1.16	2.25	-0.05	1.35

Notes: Breaks in the monetary policy rule (13) are detected using the Hannan-Quinn information criterion, with a maximum of five breaks and a minmum of 15% of sample observations required to be in each estimated monetary policy regime. Inflation and unemployment gap forecasts are treated as endogenous in the monetary policy rule, with breaks detected separately for each reduced form equation. Figures in parentheses in Panel A are t-ratios. The implied steady-state responses of monetary policy to inflation and the unemployment gap shown n Panel B are obtained from the estimated coefficients assuming constant short-term interest rates ($r_t = r_{t-1} = r_{t-2}$).

Appendix

Mathematical Appendix

Proof of Lemma 1

Case(i) Assume that (1) is stable for $t = [\tau_{j-1}T] + 1, \ldots, [\tau_j T]$, where τ_i denotes the estimated break fraction, so that for some i,

$$y_t = x_t'\beta_{x,i}^0 + z_{1,t}'\beta_{z_1,i}^0 + u_t \quad t = [\tau_{j-1}T] + 1, \ldots, [\tau_j T]. \tag{14}$$

Let $\hat{\beta}_j$ be the 2SLS estimator of $\beta_i^0 = [\beta_{x,i}^{0\prime}, \beta_{z_1,i}^{0\prime}]'$ based on (14) using $\hat{x}_t(\hat{\pi})$ defined in (4), and define $w_t(\pi) = [\hat{x}_t(\pi)', z_{1,t}']'$. Then, we have

$$\hat{\beta}_j \;=\; \left(\sum_j w_t(\hat{\pi})w_t(\hat{\pi})'\right)^{-1} \sum_j w_t(\hat{\pi})y_t \;=\; \beta_i^0 + \left(\sum_j w_t(\hat{\pi})w_t(\hat{\pi})'\right)^{-1} \sum_j w_t(\hat{\pi})\tilde{u}_t(\hat{\pi}),$$

where $\sum_j$ denotes $\sum_{[\tau_{j-1}T]+1}^{[\tau_j T]}$ and

$$\tilde{u}_t(\hat{\pi}) \;=\; y_t \;-\; w_t(\hat{\pi})'\beta_i^0. \tag{15}$$

To facilitate the analysis of $RSS_j(\tau(n); n, \hat{\pi})$ (henceforth RSS_j), we consider $y_t - w_t(\hat{\pi})'\hat{\beta}_j$. Defining $\tilde{u}_t(\pi^0)$ and $\hat{x}_t(\pi^0)$ analogously to $\tilde{u}_t(\hat{\pi})$ and $\hat{x}_t(\hat{\pi})$, it can shown that (15) implies

$$\begin{aligned} y_t - w_t(\hat{\pi})'\hat{\beta}_j \;&=\; \tilde{u}_t(\pi^0) + \left[\hat{x}_t(\pi^0)' - \hat{x}_t(\hat{\pi})'\right]\beta_{x,i}^0 - w_t(\hat{\pi})'\left(\sum_j w_t(\hat{\pi})w_t(\hat{\pi})'\right)^{-1} \\ &\quad \times \sum_j w_t(\hat{\pi})\tilde{u}_t(\hat{\pi}). \end{aligned} \tag{16}$$

From (16), it follows that

$$T^{-1}RSS_j \;=\; T^{-1}\sum_j (A_t + B_t - C_t)^2, \tag{17}$$

where $A_t \;=\; \tilde{u}_t(\pi^0)$, $B_t \;=\; \left[\hat{x}_t(\pi^0)' - \hat{x}_t(\hat{\pi})'\right]\beta_{x,i}^0$, and

$$C_t \;=\; w_t(\hat{\pi})'\left(\sum_j w_t(\hat{\pi})w_t(\hat{\pi})'\right)^{-1} \sum_{t=[Ts]+1}^{[Tr]} w_t(\hat{\pi})\tilde{u}_t(\hat{\pi}).$$

. We now consider in turn the terms obtained by multiplying out the quadratic in (17).

First for A_t^2: using (14) and substituting for $\hat{x}_t(\pi^0)$ from (3), we have

$$\begin{aligned} T^{-1}\sum_j \tilde{u}_t(\pi^0)^2 &= T^{-1}\sum_j \left\{y_t - w_t(\pi^0)'\beta_i^0\right\}^2 \\ &= T^{-1}\sum_j \left\{u_t + [x_t - \hat{x}_t(\pi^0)]'\beta_{x,i}^0\right\}^2 \\ &= T^{-1}\sum_j \left\{u_t + v_t'\beta_{x,i}^0 - \tilde{z}_t(\pi^0)'\left[\hat{\Theta}_T(\pi_0) - \Theta_0\right]\beta_{x,i}^0\right\}^2. \end{aligned} \tag{18}$$

$\hat{\Theta}_T(\pi_0)$ is the (infeasible) OLS estimator constructed using the true reduced form break fractions $\{\pi^0\}$ as the break dates, and as such, may be decomposed as

$$\hat{\Theta}_T(\pi_0) = \Theta_0 + \left(\sum_{t=1}^T \tilde{z}_t(\pi^0)\tilde{z}_t(\pi^0)'\right)^{-1}\sum_{t=1}^T \tilde{z}_t(\pi^0)v_t'.$$

Substituting this formula into (18) we obtain

$$T^{-1}\sum_j A_t^2 = T^{-1}\sum_j \left\{a_t + b_t - c_t\right\}^2, \tag{19}$$

where $a_t = u_t$, $b_t = v_t'\beta_{x,i}^0$, and $c_t = \tilde{z}_t(\pi^0)'\left(\sum_{t=1}^T \tilde{z}_t(\pi^0)\tilde{z}_t(\pi^0)'\right)^{-1}\sum_{t=1}^T \tilde{z}_t(\pi^0)v_t'\beta_{x,i}^0$.

By Assumptions 4 and 5, it follows that for the terms a_t^2, b_t^2, and $2a_tb_t$ in (19), respectively we have,

$$\begin{aligned} T^{-1}\sum_j u_t^2 &\overset{p.u}{\to} (\tau_j - \tau_{j-1})\sigma_u^2, \\ T^{-1}\sum_j \beta_{x,i}^{0'}v_tv_t'\beta_{x,i}^0 &\overset{p.u}{\to} (\tau_j - \tau_{j-1})\beta_{x,i}^0\Sigma_v\beta_{x,i}^0, \\ T^{-1}2\sum_j u_tv_t'\beta_{x,i}^0 &\overset{p.u}{\to} (\tau_j - \tau_{j-1})2\Sigma_{uv}\beta_{x,i}^0. \end{aligned}$$

For the remaining terms in (19), using Assumptions 2 and 8 we have $T^{-1}\sum_{t=[Ts]+1}^{[Tr]} z_tz_t' \overset{p.u}{\to} Q_{ZZ}(r) - Q_{ZZ}(s) = M_{ZZ}(s,r)$ for $r > s + \epsilon$ is also *pd* and monotonically increasing. Also by Assumptions 2 and 8, it follows that

$$T^{-1}\sum_{t=1}^T \tilde{z}_t(\pi^0)\tilde{z}_t(\pi^0)' \overset{p.u}{\to} \tilde{Q}_{ZZ}(1),$$

also *pd*, where $\tilde{Q}_{ZZ}(1)$ is the block diagonal matrix $diag(Q_1, Q_2, \ldots Q_{h+1})$ and $Q_i = Q_{ZZ}(\pi_i^0) - Q_{ZZ}(\pi_{i-1}^0)$ and we set $\pi_0^0 = 0$, $\pi_{h+1}^0 = 1$. Then, for a segment of the data $t = [\tau_{j-1}T] + 1, \ldots, [\tau_jT]$, it follows that

$$T^{-1}\sum_{t=[\tau_{j-1}T]+1}^{[\tau_jT]} \tilde{z}_t(\pi^0)\tilde{z}_t(\pi^0)' \overset{p.u}{\to} \tilde{Q}(\tau_{j-1},\tau_j) \text{ in } \tau_{j-1},\tau_j,\ (\tau_j > \tau_{j-1} + \epsilon) \text{ and } pd, \tag{20}$$

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

Furthermore, it follows from Assumptions 2 and 4, that $T^{-1/2}\sum_{t=1}^{[Tr]} \tilde{z}_t(\pi^0) \otimes \{(u_t, v_t{}')'\}$ is $O_p(1)$ via a central limit theorem. The above suffice to show that for the remaining terms in (19) we have, $T^{-1}\sum_j a_t c_t \overset{p.u}{\to} 0$, $T^{-1}\sum_j b_t c_t \overset{p.u}{\to} 0$ and $T^{-1}\sum_j c_t^2 \overset{p.u}{\to} 0$.

Combining these results regarding the term A_t^2 in (17) it follows that

$$T^{-1}\sum_j A_t^2 \overset{p.u}{\to} (\tau_j - \tau_{j-1})\Gamma_i \tag{21}$$

with Γ_i defined in Lemma 1.

The term involving B_t^2 in (17) can be written as

$$\begin{aligned} T^{-1}\sum_j B_t^2 &= T^{-1}\sum_j \beta_{x,i}^{0\prime}\left[\hat{x}_t(\pi^0) - \hat{x}_t(\hat{\pi})\right]\left[\hat{x}_t(\pi^{0\prime}) - \hat{x}_t(\hat{\pi})'\right]\beta_{x,i}^0 \\ &= \beta_{x,i}^{0\prime}\left\{T^{-1}\sum_j\left[\hat{x}_t(\pi^0)\hat{x}_t(\pi^0)' + \hat{x}_t(\hat{\pi})\hat{x}_t(\hat{\pi})' - 2\hat{x}_t(\pi^0)\hat{x}_t(\hat{\pi})'\right]\right\}\beta_{x,i}^0. \end{aligned} \tag{22}$$

The following results will determine the probability limit of (22). From Assumptions 3 and 8 it follows that

$$\begin{aligned} &T^{-1}\sum_{t=1}^T \tilde{z}_t(\hat{\pi})\tilde{z}_t(\hat{\pi})'^{-1}\sum_{t=1}^T \tilde{z}_t(\pi^0)\tilde{z}_t(\pi^0)' + o_p(1) \\ &\qquad \overset{p.u}{\to} \quad \tilde{Q}_{ZZ}(1) \end{aligned} \tag{23}$$

and also,

$$T^{-1}\sum_j \tilde{z}_t(\hat{\pi})\tilde{z}_t(\pi^0)' \overset{p.u}{\to} \tilde{Q}(\tau_{j-1},\tau_j). \tag{24}$$

From Assumptions 3, 4, and 8 it follows that

$$T^{-1}\sum_{t=1}^T \tilde{z}_t(\hat{\pi})x_t'^{-1}\sum_{t=1}^T \tilde{z}_t(\pi^0)x_t' + o_p(1) \overset{p.u}{\to} \tilde{Q}_{ZZ}(1)\Theta_0 \tag{25}$$

By (23), (25), and (4),

$$\begin{aligned} T^{-1}\sum_j \hat{x}_t(\hat{\pi})\hat{x}_t(\hat{\pi})' &= T^{-1}\sum_{t=1}^T x_t\tilde{z}_t(\hat{\pi})'\left(\sum_{t=1}^T \tilde{z}_t(\hat{\pi})\tilde{z}_t(\hat{\pi})'\right)^{-1}\sum_j \tilde{z}_t(\hat{\pi})\tilde{z}_t(\hat{\pi})' \\ &\quad \times\left(\sum_{t=1}^T \tilde{z}_t(\hat{\pi})\tilde{z}_t(\hat{\pi})'\right)^{-1}\sum_{t=1}^T \tilde{z}_t(\hat{\pi})x_t' \\ &\overset{p.u}{\to} \Theta_0'\tilde{Q}_{ZZ}'(1)\tilde{Q}_{ZZ}^{-1}(1)\tilde{Q}(\tau_{j-1},\tau_j)\tilde{Q}_{ZZ}^{-1}(1)\tilde{Q}_{ZZ}(1)\Theta_0 \\ &= \Theta_0'\tilde{Q}(\tau_{j-1},\tau_j)\Theta_0 \end{aligned} \tag{26}$$

which is *pd* by the construction of $\tilde{Q}(\tau_{j-1}, \tau_j)$. Similarly, we have

$$T^{-1}\sum_j \hat{x}_t(\pi^0)\hat{x}_t(\pi^0)^{'} \overset{p.u}{\to} \Theta_0^{'}\tilde{Q}(\tau_{j-1}, \tau_j)\Theta_0. \tag{27}$$

For the last term in (22), we combine (23), (24), (25), and (4), and use Assumption 2 to deduce that

$$\begin{aligned} 2T^{-1}\sum_j \hat{x}_t(\pi^0)\hat{x}_t(\hat{\pi})^{'} &= 2T^{-1}\sum_{t=1}^{T} x_t\tilde{z}_t(\pi_0)^{'}\left(\sum_{t=1}^{T}\tilde{z}_t(\pi_0)\tilde{z}_t(\pi_0)^{'}\right)^{-1}\sum_j \tilde{z}_t(\pi_0)\tilde{z}_t(\hat{\pi})^{'} \\ &\quad \times \left(\sum_{t=1}^{T}\tilde{z}_t(\hat{\pi})\tilde{z}_t(\hat{\pi})^{\prime}\right)^{-1}\sum_{t=1}^{T}\tilde{z}_t(\hat{\pi})x_t^{\prime} \\ &\overset{p.u}{\to} 2\Theta_0^{'}\tilde{Q}(\tau_{j-1}, \tau_j)\Theta_0. \end{aligned} \tag{28}$$

Combining (26), (27), and (28), the probability limit of (22) is

$$T^{-1}\sum_j B_t^2 \overset{p.u}{\to} 0. \tag{29}$$

Now consider the terms involving C_t in (17). Start by considering $\sum_j w_t(\hat{\pi})u_t(\hat{\pi})$. If we expand $\tilde{u}_t(\hat{\pi})$ similarly to (16), and substitute for $\hat{x}_t(\hat{\pi})$, then from (4) we obtain

$$\sum_j w_t(\hat{\pi})\tilde{u}_t(\hat{\pi}) \;=\; \sum_j w_t(\hat{\pi})\left\{\tilde{u}_t(\pi^0) + \left[\hat{x}_t(\pi^0)^{'} - \hat{x}_t(\hat{\pi})^{'}\right]\beta_{x,i}^0\right\}.$$

Thus, we have

$$\begin{aligned} \sum_j \hat{x}_t(\hat{\pi})\tilde{u}_t(\hat{\pi}) &= \sum_{t=1}^{T} x_t\tilde{z}_t(\hat{\pi})^{'}\left(\sum_{t=1}^{T}\tilde{z}_t(\hat{\pi})\tilde{z}_t(\hat{\pi})^{'}\right)^{-1}\left[\sum_j \tilde{z}_t(\hat{\pi})\tilde{u}_t(\pi_0)\right. \\ &+ \sum_j \tilde{z}_t(\hat{\pi})\tilde{z}_t(\pi_0)^{'}\left(\sum_{t=1}^{T}\tilde{z}_t(\pi_0)\tilde{z}_t(\pi_0)^{'}\right)^{-1}\sum_{t=1}^{T}\tilde{z}_t(\pi_0)x_t^{'}\beta_{x,i}^0 \\ &- \left.\sum_j \tilde{z}_t(\hat{\pi})\tilde{z}_t(\hat{\pi})^{'}\left(\sum_{t=1}^{T}\tilde{z}_t(\hat{\pi})\tilde{z}_t(\hat{\pi})^{'}\right)^{-1}\sum_{t=1}^{T}\tilde{z}_t(\hat{\pi})x_t^{'}\beta_{x,i}^0\right]. \end{aligned} \tag{30}$$

From (23), (24), and (25), the last two terms inside the brackets in (30) cancel out asymptotically. The same equations and also Assumption 4, after expanding $\tilde{u}_t(\pi_0)$ similarly to (19), give

$$\begin{aligned} T^{-1}\sum_j \tilde{z}_t(\hat{\pi})\tilde{u}_t(\pi_0) &= T^{-1}\sum_j \tilde{z}_t(\hat{\pi})u_t + T^{-1}\sum_j \tilde{z}_t(\hat{\pi})v_t^{'}\beta_{x,i}^0 \\ &\quad -T^{-1}\sum_j \tilde{z}_t(\hat{\pi})\tilde{z}_t(\pi_0)\left(\sum_{t=1}^{T}\tilde{z}_t(\pi_0)\tilde{z}_t(\pi_0)^{'}\right)^{-1}\sum_{t=1}^{T}\tilde{z}_t(\pi_0)v_t^{'}\beta_{x,i}^0 \\ &\overset{p.u}{\to} 0 \end{aligned} \tag{31}$$

and therefore it follows from (30) that,

$$T^{-1}\sum_{j}\hat{x}_t(\hat{\pi})\tilde{u}_t(\hat{\pi})\overset{p.u}{\to}0. \tag{32}$$

It can also be shown via similar arguments that

$$T^{-1}\sum_{j}z_{1,t}\tilde{u}_t(\hat{\pi})\overset{p.u}{\to}0. \tag{33}$$

Using (32)-(33), together with (26) and (28) it follows that the limiting behaviour of the terms involving C_t in $T^{-1}RSS_j$ are:

$$\begin{aligned} T^{-1}\sum_{j}C_t^2 &\overset{p.u}{\to} 0 && (34)\\ 2T^{-1}\sum_{j}A_tC_t &\overset{p.u}{\to} 0 && (35)\\ 2T^{-1}\sum_{j}B_tC_t &\overset{p.u}{\to} 0 && (36)\end{aligned}$$

The last remaining term of $T^{-1}RSS_j$ involves $2A_tB_t$,

$$\begin{aligned} T^{-1}\sum_{j}A_tB_t &= T^{-1}\sum_{j}\tilde{u}_t(\pi^0)\left[\hat{x}_t(\pi^0)^{'}-\hat{x}_t(\hat{\pi})^{'}\right]\beta_{x,i}^0\\ &= T^{-1}\sum_{j}\left[\tilde{u}_t(\pi^0)\hat{x}_t(\pi^0)^{'}-\tilde{u}_t(\pi^0)\hat{x}_t(\hat{\pi})^{'}\right]\beta_{x,i}^0\end{aligned}$$

Using (19), and (4) the first term inside the summation can be expanded as

$$\begin{aligned} T^{-1}\sum_{j}\tilde{u}_t(\pi^0)\hat{x}_t(\pi^0)^{'} &= T^{-1}\sum_{j}\left[u_t+v_t^{'}\beta_{x,i}^0-\tilde{z}_t(\pi^0)^{'}\left(\sum_{t=1}^{T}\tilde{z}_t(\pi^0)\tilde{z}_t(\pi^0)^{'}\right)^{-1}\sum_{t=1}^{T}\tilde{z}_t(\pi^0)v_t^{'}\beta_{x,i}^0\right]\\ &\quad\times\tilde{z}_t(\hat{\pi})'\left(\sum_{t=1}^{T}\tilde{z}_t(\hat{\pi})\tilde{z}_t(\hat{\pi})'\right)^{-1}\sum_{t=1}^{T}\tilde{z}_t(\hat{\pi})x_t'\\ &\overset{p.u}{\to} 0\end{aligned}$$

since by Assumption 4, $T^{-1}\sum_{t=[Ts]+1}^{[Tr]}u_t\tilde{z}_t(\pi^0)\overset{p.u}{\to}0$ and $T^{-1}\sum_{t=[Ts]+1}^{[Tr]}v_t^{'}\tilde{z}_t(\pi^0)\overset{p.u}{\to}0$. Also using the same arguments and (23), and (25), $T^{-1}\sum_j\tilde{u}_t(\pi^0)\hat{x}_t(\hat{\pi})^{'}\overset{p.u}{\to}0$ as well, resulting in

$$T^{-1}\sum_{j}A_tB_t\overset{p.u}{\to}0. \tag{37}$$

Collecting the results regarding the limit of $T^{-1}RSS_j$ in (17), found in (21), (29), (**??**), (**??**), (34), and (37), it follows that

$$T^{-1}RSS_j\overset{p.u}{\to}(\tau_j-\tau_{j-1})\Gamma_i. \tag{38}$$

Case (ii): we first consider the case where segment j contains one neglected break and then discuss how the argument extends to more than one neglected break. Let the neglected break be at λ_i^0 so that the model is

$$\begin{aligned} y_t &= x_t^{'}\beta_{x,i}^0 + z_{1,t}^{'}\beta_{z_1,i} + u_t, \qquad t = [\tau_{j-1}T]+1, ..., [\lambda_i^0 T] \\ y_t &= x_t^{'}\beta_{x,i+1}^0 + z_{1,t}^{'}\beta_{z_1,i} + u_t, \qquad t = [\lambda_i^0 T]+1, ..., [\tau_j T]. \end{aligned} \tag{39}$$

Then the residual sum of squares in this segment, RSS_j, may be decomposed as

$$\begin{aligned} T^{-1}RSS_j &= T^{-1}\sum_{t=[\tau_{j-1}T]+1}^{[\lambda_i^0 T]} \left[y_t - w_t(\hat{\pi})^{'}\hat{\beta}_j\right]^2 + T^{-1}\sum_{t=[\lambda_i^0 T]+1}^{[\tau_j T]} \left[y_t - w_t(\hat{\pi})'\hat{\beta}_j\right]^2 \\ &= \xi_1 + \xi_2, \text{ say, respectively.} \end{aligned} \tag{40}$$

We focus on ξ_1. Substituting for y_t from (15), we have

$$\begin{aligned} \xi_1 &= T^{-1}\sum_{t=[\tau_{j-1}T]+1}^{[\lambda_i^0 T]} \left[w_t(\hat{\pi})^{'}\beta_i^0 + \tilde{u}_t(\hat{\pi}) - w_t(\hat{\pi})^{'}\hat{\beta}_j\right]^2 \\ &= T^{-1}\sum_{t=[\tau_{j-1}T]+1}^{[\lambda_i^0 T]} \left[\tilde{u}_t(\hat{\pi}) - w_t(\hat{\pi})^{'}\left(\hat{\beta}_j - \beta_i^0\right)\right]^2 \\ &= T^{-1}\sum_{t=[\tau_{j-1}T]+1}^{[\lambda_i^0 T]} \left[\tilde{u}_t(\hat{\pi})^2 - 2\tilde{u}_t(\hat{\pi})w_t(\hat{\pi})^{'}\left(\hat{\beta}_j - \beta_i^0\right)\right. \\ &\qquad \left. + \left(\hat{\beta}_j - \beta_i^0\right)^{'} w_t(\hat{\pi})w_t(\hat{\pi})^{'}\left(\hat{\beta}_j - \beta_i^0\right)\right]. \end{aligned} \tag{41}$$

The first term in this sum can be written as, $\tilde{u}_t(\hat{\pi})^2 = \left[\tilde{u}_t(\pi^0) + \left(\hat{x}_t(\hat{\pi})^{'} - \hat{x}_t(\pi^0)^{'}\right)\beta_{x,i}^0\right]^2$. Using the results in the proof of Case (i) above for the limits of the terms involving sums of A_t^2, B_t^2, and A_tB_t, found in (21), (29), and (37), it can be shown that,

$$T^{-1}\sum_{t=[\tau_{j-1}T]+1}^{[\lambda_i^0 T]} \tilde{u}_t(\hat{\pi})^2 \overset{p.u}{\to} (\lambda_i^0 - \tau_{j-1})\Gamma_i. \tag{42}$$

To proceed we need to derive $plim\left(\hat{\beta}_j - \beta_i^0\right)$ where

$$\hat{\beta}_j = \left(\sum_j w_t(\hat{\pi})w_t(\hat{\pi})'\right)^{-1} \sum_j w_t(\hat{\pi})y_t. \tag{43}$$

Using similar arguments to (26) and (32) we have that

$$T^{-1}\sum_{j}\hat{x}_t(\hat{\pi})\hat{x}_t(\hat{\pi})' \overset{p.u}{\to} \bar{\Upsilon}_0'\tilde{Q}(\tau_{j-1},\tau_j)\bar{\Upsilon}_0 \tag{44}$$

where $\bar{\Upsilon}_0=[\Theta_0,\bar{\Pi}]$, $\bar{\Pi}=\imath_{h+1}\otimes\Pi$, and

$$\begin{aligned}\sum_{j}w_t(\hat{\pi})y_t &= T^{-1}\sum_{t=[\tau_{j-1}T]+1}^{[\lambda_i^0T]}w_t(\hat{\pi})\left[w_t(\hat{\pi})'\beta_i^0+\tilde{u}_t(\hat{\pi})\right]\\ &\quad +T^{-1}\sum_{t=[\lambda_i^0T]+1}^{[\tau_jT]}w_t(\hat{\pi})\left[w_t(\hat{\pi})'\beta_{i+1}^0+\tilde{u}_t(\hat{\pi})\right]\\ &\overset{p.u}{\to} \bar{\Upsilon}_0'\tilde{Q}(\tau_{j-1},\lambda_i^0)\bar{\Upsilon}_0\beta_i^0+\bar{\Upsilon}_0'\tilde{Q}(\lambda_i^0,\tau_j)\bar{\Upsilon}_0\beta_{i+1}^0.\end{aligned}\tag{45}$$

Combining (43), (44), and (45) results in

$$plim\left(\hat{\beta}_j-\beta_i^0\right)\overset{p.u}{\to}\left\{\bar{\Upsilon}_0'\tilde{Q}(\tau_{j-1},\tau_j)\bar{\Upsilon}_0\right\}^{-1}\left\{\bar{\Upsilon}_0'\tilde{Q}(\tau_{j-1},\lambda_i^0)\bar{\Upsilon}_0\beta_i^0+\bar{\Upsilon}_0'\tilde{Q}(\lambda_i^0,\tau_j)\bar{\Upsilon}_0\beta_{i+1}^0\right\}-\beta_i^0. \tag{46}$$

Furthermore, β_i^0 can be written as,

$$\begin{aligned}\beta_i^0 &= \left\{\bar{\Upsilon}_0'\tilde{Q}(\tau_{j-1},\tau_j)\bar{\Upsilon}_0\right\}^{-1}\left\{\bar{\Upsilon}_0'\tilde{Q}(\tau_{j-1},\tau_j)\bar{\Upsilon}_0\right\}\beta_i^0\\ &= \left\{\bar{\Upsilon}'\tilde{Q}(\tau_{j-1},\tau_j)\bar{\Upsilon}_0\right\}^{-1}\left\{\bar{\Upsilon}_0'\tilde{Q}(\tau_{j-1},\lambda_i^0)\bar{\Upsilon}_0\beta_i^0+\bar{\Upsilon}_0'\tilde{Q}(\lambda_i^0,\tau_j)\bar{\Upsilon}_0\beta_i^0\right\}.\end{aligned}$$

Substituting this into (46) and after some algebra it is shown that

$$plim\left(\hat{\beta}_j-\beta_i^0\right)\overset{p.u}{\to}\left\{\bar{\Upsilon}_0'\tilde{Q}(\tau_{j-1},\tau_j)\bar{\Upsilon}_0\right\}^{-1}\bar{\Upsilon}_0'\tilde{Q}(\lambda_i^0,\tau_j)\bar{\Upsilon}_0\left(\beta_{i+1}^0-\beta_i^0\right)=P_1. \tag{47}$$

P_1 is ensured to be non-zero because $\tilde{Q}(r,s)$ is a block diagonal matrix and each block is positive definite via Assumption 9, and also $\beta_{i+1}^0\neq\beta_i^0$.

Going back to ξ_1, it follows that from (32), (33), (47) and (44) that the last two terms in (41) have the following probability limits

$$T^{-1}\sum_{t=[\tau_{j-1}T]+1}^{[\lambda_i^0T]}\tilde{u}_t(\hat{\pi})w_t(\hat{\pi})'\left(\hat{\beta}_j-\beta_i^0\right)\overset{p.u}{\to}0 \tag{48}$$

and

$$\left(\hat{\beta}_j-\beta_i^0\right)'\sum_{t=[\tau_{j-1}T]+1}^{[\lambda_i^0T]}w_t(\hat{\pi})w_t(\hat{\pi})'\left(\hat{\beta}_j-\beta_i^0\right)\overset{p.u}{\to}P_1'\bar{\Upsilon}_0'\tilde{Q}(\tau_{j-1},\lambda_i^0)\bar{\Upsilon}_0P_1\ >\ 0, \tag{49}$$

since $\tilde{Q}(r,s)$ is positive definite and $P_1\neq 0$.

Collecting the results from (42), (48), and (49),

$$\xi_1 \overset{p.u}{\to} (\lambda_i^0 - \tau_{j-1})\Gamma_i + P_1^{'}\bar{\Upsilon}_0^{\prime}\tilde{Q}(\tau_{j-1}, \lambda_i^0)\bar{\Upsilon}_0 P_1 \ > \ (\lambda_i^0 - \tau_{j-1})\Gamma_i. \tag{50}$$

Analogously for ξ_2, we have

$$\xi_2 \overset{p.u}{\to} (\tau_j - \lambda_i^0)\Gamma_{i+1} + P_2^{'} bar \Upsilon_0^{\prime}\tilde{Q}(\lambda_i^0, \tau_j)\bar{\Upsilon}_0 P_2 \ > \ (\tau_j - \lambda_i^0)\Gamma_{i+1} \tag{51}$$

where

$$P_2 \ = \ \left\{\bar{\Upsilon}_0^{\prime}\tilde{Q}(\tau_{j-1}, \tau_j)\bar{\Upsilon}_0\right\}^{-1} \bar{\Upsilon}_0^{\prime}\tilde{Q}(\tau_{j-1}, \lambda_i^0)\bar{\Upsilon}_0 \left(\beta_i^0 - \beta_{i+1}^0\right) \ \neq \ 0.$$

$$T^{-1}RSS_j \overset{p.u}{\to} (\lambda_i^0 - \tau_{j-1})\Gamma_i + (\tau_j - \lambda_i^0)\Gamma_{i+1} \ + \ F_1 \tag{52}$$

where

$$F_1 \ = \ P_1^{'}\Theta_0^{'}\tilde{Q}(\tau_{j-1}, \lambda_i^0)\Theta_0 P_1 \ + \ P_2^{'}\Theta_0^{'}\tilde{Q}(\lambda_i^0, \tau_j)\Theta_0 P_2 \ > \ 0. \tag{53}$$

This line of argument extends to more than one neglected break. We now show how two neglected breaks in a segment of the structural equation extend the case discussed above. To do this, we must evaluate the limiting distribution of RSS_j in a segment where there are two neglected breaks, denoted λ_i^0, and λ_{i+1}^0. Therefore, the data generation process is

$$\begin{aligned} y_t \ &= \ x_t^{\prime}\beta_{x,i}^0 \ + \ z_{1,t}^{'}\beta_{z_1,i}^0 \ + \ u_t, \qquad t = [\tau_{j-1}T]+1, ..., [\lambda_i^0 T] \\ y_t \ &= \ x_t^{\prime}\beta_{x,i+1}^0 \ + \ z_{1,t}^{'}\beta_{z_1,i}^0 \ + \ u_t, \qquad t = [\lambda_i^0 T]+1, ..., [\lambda_{i+1}^0 T] \\ y_t \ &= \ x_t^{\prime}\beta_{x,i+2}^0 \ + \ z_{1,t}^{'}\beta_{z_1,i}^0 \ + \ u_t, \qquad t = [\lambda_{i+1}^0 T]+1, ..., [\tau_j T]. \end{aligned} \tag{54}$$

The RSS for this segment can be decomposed as,

$$\begin{aligned} T^{-1}RSS_j \ &= \ T^{-1}\sum_{t=[\tau_{j-1}T]+1}^{[\lambda_i^0 T]} \left(y_t - w_t(\hat{\pi})^{'}\hat{\beta}_j\right)^2 + T^{-1}\sum_{t=[\lambda_i^0 T]+1}^{[\lambda_{i+1}^0 T]} \left(y_t - w_t(\hat{\pi})^{'}\hat{\beta}_j\right)^2 \\ &\quad + T^{-1}\sum_{t=[\lambda_{i+1}^0 T]+1}^{[\tau_j T]} \left(y_t - w_t(\hat{\pi})^{'}\hat{\beta}_j\right)^2 \\ &= \ \xi_1 \ + \ \xi_2 \ + \ \xi_3. \end{aligned} \tag{55}$$

Focusing on ξ_1, as in (41), this term can be written as

$$\begin{aligned} \xi_1 \ &= \ T^{-1}\sum_{t=[\tau_{j-1}T]+1}^{[\lambda^0 T]} \left[\tilde{u}_t(\hat{\pi})^2 - 2\tilde{u}_t(\hat{\pi})w_t(\hat{\pi})^{'}\left(\hat{\beta}_j - \beta_i^0\right)\right. \\ &\quad \left. + \ \left(\hat{\beta}_j - \beta_i^0\right)^{'} w_t(\hat{\pi})w_t(\hat{\pi})^{'}\left(\hat{\beta}_j - \beta_i^0\right)\right]. \end{aligned} \tag{56}$$

where $\hat{\beta}_j$ is defined as in (43). To analyse the limit of $\hat{\beta}_j$, note that (44) holds within the scenario considered here. However, this time we have but in this case,

$$\begin{aligned}\sum_j w_t(\hat{\pi})y_t &= T^{-1}\sum_{t=[\tau_{j-1}T]+1}^{[\lambda_i^0 T]} w_t(\hat{\pi})\left[w_t(\hat{\pi})^{'}\beta_i^0 + \tilde{u}_t(\hat{\pi})\right] \\ &+ T^{-1}\sum_{t=[\lambda_i^0 T]+1}^{[\lambda_{i+1}^0 T]} w_t(\hat{\pi})\left[w_t(\hat{\pi})'\beta_{i+1}^0 + \tilde{u}_t(\hat{\pi})\right] \\ &+ T^{-1}\sum_{t=[\lambda_{i+1}^0 T]+1}^{[\tau_j T]} w_t(\hat{\pi})\left[w_t(\hat{\pi})^{'}\beta_{i+2}^0 + \tilde{u}_t(\hat{\pi})\right]\end{aligned}$$

and so

$$\begin{aligned}\sum_j w_t(\hat{\pi})y_t \overset{p.u}{\to}\ & \bar{\Upsilon}_0'\tilde{Q}(\tau_{j-1},\lambda_i^0)\bar{\Upsilon}_0\beta_i^0 + \bar{\Upsilon}_0'\tilde{Q}(\lambda_i^0,\lambda_{i+1})\bar{\Upsilon}_0\beta_{i+1}^0 \\ &+ \bar{\Upsilon}_0'\tilde{Q}(\lambda_{i+1}^0,\tau_j)\bar{\Upsilon}_0\beta_{i+2}^0.\end{aligned}$$

By a similar argument like the one that lead to (46), for $plim\left(\hat{\beta}_{x,j} - \beta_{x,i}^0\right)$ here it follows that

$$\begin{aligned}plim\left(\hat{\beta}_j - \beta_i^0\right) \overset{p.u}{\to}\ & \left\{\bar{\Upsilon}_0'\tilde{Q}(\tau_{j-1},\tau_j)\bar{\Upsilon}_0\right\}^{-1}\left\{\bar{\Upsilon}_0'\tilde{Q}(\tau_{j-1},\lambda_i^0)\bar{\Upsilon}_0\beta_i^0 + \bar{\Upsilon}_0'\tilde{Q}(\lambda_i^0,\lambda_{i+1}^0)\bar{\Upsilon}_0\beta_{i+1}^0\right. \\ &\left. + \bar{\Upsilon}_0'\tilde{Q}(\lambda_{i+1}^0,\tau_j)\bar{\Upsilon}_0\beta_{i+2}^0\right\} - \beta_i^0. \qquad (57)\end{aligned}$$

We can rewrite β_i^0 as

$$\begin{aligned}\beta_i^0 &= \left\{\bar{\Upsilon}_0'\tilde{Q}(\tau_{j-1},\tau_j)\bar{\Upsilon}_0\right\}^{-1}\left\{\bar{\Upsilon}_0'\tilde{Q}(\tau_{j-1},\tau_j)\bar{\Upsilon}_0\right\}\beta_i^0 \\ &= \left\{\bar{\Upsilon}_0'\tilde{Q}(\tau_{j-1},\tau_j)\bar{\Upsilon}_0\right\}^{-1}\left\{\bar{\Upsilon}_0'\tilde{Q}(\tau_{j-1},\lambda_i^0)\bar{\Upsilon}_0 + \bar{\Upsilon}_0'\tilde{Q}(\lambda_i^0,\lambda_{i+1}^0)\bar{\Upsilon}_0\right. \\ &\quad\left. +\bar{\Upsilon}_0'\tilde{Q}(\lambda_{i+1}^0,\tau_j)\bar{\Upsilon}_0\right\}\beta_i^0\end{aligned}$$

Then, after substituting this equation into (57) and rearranging terms, we obtain

$$\begin{aligned}plim\left(\hat{\beta}_j - \beta_i^0\right) \overset{p.u}{\to}\ & \left\{\bar{\Upsilon}_0'\tilde{Q}(\tau_{j-1},\tau_j)\bar{\Upsilon}_0\right\}^{-1} \\ & \left\{\bar{\Upsilon}_0'\tilde{Q}(\lambda_i^0,\lambda_{i+1}^0)\bar{\Upsilon}_0(\beta_{i+1}^0 - \beta_i^0)\right. \\ & \left. + \bar{\Upsilon}_0'\tilde{Q}(\lambda_{i+1}^0,\tau_j)\bar{\Upsilon}_0(\beta_{i+2}^0 - \beta_i^0)\right\} = K_1, \text{ say.} \qquad (58)\end{aligned}$$

Using this expression together with (42) and (**??**), it follows that the limit of ξ_1

$$\xi_1 \overset{p.u}{\to} (\lambda_i^0 - \tau_{j-1})\Gamma_i + K_1^{'}\bar{\Upsilon}_0'\tilde{Q}(\tau_{j-1},\lambda_i^0)\bar{\Upsilon}_0 K_1. \qquad (59)$$

Analogously, we have

$$\xi_2 \overset{p.u}{\to} (\lambda_{i+1}^0 - \lambda_i^0)\Gamma_{i+1} + K_2^{'}\bar{\Upsilon}_0'\tilde{Q}(\lambda_i^0,\lambda_{i+1}^0)\bar{\Upsilon}_0 K_2 \qquad (60)$$

with

$$K_2 \;=\; \left\{\bar{\Upsilon}_0'\tilde{Q}(\tau_{j-1},\tau_j)\bar{\Upsilon}_0\right\}^{-1}\left\{\bar{\Upsilon}_0'\tilde{Q}(\tau_{j-1},\lambda_i^0)\bar{\Upsilon}_0(\beta_i^0-\beta_{i+1}^0)+\bar{\Upsilon}_0'\tilde{Q}(\lambda_{i+1}^0,\tau_j)\bar{\Upsilon}_0(\beta_{i+2}^0-\beta_{i+1}^0)\right\}$$

and,

$$\xi_3 \overset{p.u}{\to} (\tau_j-\lambda_{i+1}^0)\Gamma_{i+2}+K_3'\bar{\Upsilon}_0'\tilde{Q}(\lambda_{i+1}^0,\tau_j)\bar{\Upsilon}_0K_3 \tag{61}$$

with

$$K_3 \;=\; \left\{\bar{\Upsilon}_0'\tilde{Q}(\tau_{j-1},\tau_j)\bar{\Upsilon}_0\right\}^{-1}\left\{\bar{\Upsilon}_0'\tilde{Q}(\tau_{j-1},\lambda_i^0)\bar{\Upsilon}_0(\beta_i^0-\beta_{i+2}^0)+\bar{\Upsilon}_0'\tilde{Q}(\lambda_i^0,\lambda_{i+1}^0)\bar{\Upsilon}_0(\beta_{i+1}^0-\beta_{i+2}^0)\right\}.$$

Combining the above (59), (60), and (61) concludes that

$$T^{-1}RSS_j \overset{p.u}{\to} (\lambda_i^0-\tau_{j-1})\Gamma_i+(\lambda_{i+1}^0-\lambda_i^0)\Gamma_{i+1}+(\tau_j-\lambda_{i+1}^0)\Gamma_{i+2}+F_2 \tag{62}$$

where

$$\begin{aligned} F_2 \quad &= \quad K_1'\bar{\Upsilon}_0'\tilde{Q}(\tau_{j-1},\lambda_i^0)\bar{\Upsilon}_0K_1 \;+\; K_2'\bar{\Upsilon}_0'\tilde{Q}(\lambda_i^0,\lambda_{i+1}^0)\bar{\Upsilon}_0K_2 \\ &\quad + K_3'\bar{\Upsilon}_0'\tilde{Q}(\lambda_{i+1}^0,\tau_j)\bar{\Upsilon}_0K_3 \;>\; 0. \end{aligned} \tag{63}$$

To see that F_2 is positive definite consider the following. Since $\tilde{Q}(r,s)$ is positive definite for any $r>s$ and $\bar{\Upsilon}_0$ is full rank, it suffices to show that not all K_1, K_2, and K_3 can be zero. By (58), K_1 is defined as the $plim\left(\hat{\beta}_j-\beta_i^0\right)$ and analogously K_2 and K_3 are $plim\left(\hat{\beta}_j-\beta_{i+1}^0\right)$, and $plim\left(\hat{\beta}_j-\beta_{i+2}^0\right)$ respectively. From the solution for, say K_1, given in (58), it can be deduced that there can exist a combination of break locations and parameter values for which K_1 is zero. The intuition for this is that $\hat{\beta}_j$, that is estimated over $t=[\tau_{j-1}T]+1,\ldots,[\tau_jT]$, happens to converge to β_i^0. But if this is the case then at least one of K_2 and K_3 will be non zero since $\beta_i^0\neq\beta_{i+1}^0\neq\beta_{i+2}^0$ by the assumption that segment j has two neglected breaks. Therefore, the sum of terms involving those three in (63) will be strictly positive.

The same argument to the case of two neglected breaks in the segment extends to cases with more than two neglected breaks but the proofs are supressed here for brevity. Instead, we present the general form of RSS_j, for κ neglected breaks, that is

$$\begin{aligned} T^{-1}RSS_j(\tau(n);n,\hat{\pi}) \overset{p.u}{\to} &(\lambda_i^0-\tau_{j-1})\Gamma_i \;+\; (\lambda_{i+1}^0-\lambda_i^0)\Gamma_{i+1} \;+\ldots \\ &+(\lambda_{i+k}^0-\lambda_{i+k-1}^0)\Gamma_{i+k} \;+\; (\tau_j-\lambda_{i+k}^0)\Gamma_{i+k+1} \;+\; F. \end{aligned}$$

where F is a positive definite matrix defined as,

$$F \;=\; K_1'\bar{\Upsilon}_0'\tilde{Q}(\tau_{j-1},\lambda_i^0)\bar{\Upsilon}_0K_1 \;+K_2'\bar{\Upsilon}_0'\tilde{Q}(\lambda_i^0,\lambda_{i+1}^0)\bar{\Upsilon}_0K_2 \;+\ldots+\; K_\kappa'\bar{\Upsilon}_0'\tilde{Q}(\lambda_i^0,\lambda_{i+1}^0)$$

where K_ς is defined as

$$K_\varsigma \;=\; \left\{\bar{\Upsilon}_0^{'}\tilde{Q}(\tau_{j-1},\tau_j)\bar{\Upsilon}_0\right\}^{-1}\left\{\bar{\Upsilon}_0^{'}\tilde{Q}(\tau_{j-1},\lambda_i^0)\bar{\Upsilon}_0(\beta_i^0-\beta_{i+\varsigma-1}^0) \,+\, \bar{\Upsilon}_0^{'}\tilde{Q}(\lambda_i^0,\lambda_{i+1}^0)\bar{\Upsilon}_0(\beta_{i+1}^0-\beta_{i+\varsigma-1}^0)+\dots \right.$$
$$\left. +\bar{\Upsilon}_0^{'}\tilde{Q}(\lambda_{i+\kappa-1}^0,\tau_j)\bar{\Upsilon}_0(\beta_{i+\kappa}^0-\beta_{i+\varsigma-1}^0)\right\}$$

for $\varsigma = 1,2,\dots,\kappa+1$.

Proof of Theorem 1

Lemma 1 can be used to establish a proof for the consistency of the information criterion $I\left(\tau(n);n,\hat{\pi}\right)$ in selecting the true number of breaks. This can be achieved by considering the possible cases where the 2SLS procedure may over-fit, under-fit or correctly identify the true number of breaks (m) in the model. Firstly denote

$$\Gamma(\lambda^0,m,\beta^0) \;=\; \sum_{j=1}^{m+1}(\lambda_j-\lambda_{j-1})\left(\sigma_u^2 \,+\, 2\Sigma_{uv}\beta_j^0 \,+\, \beta_j^{0'}\Sigma_v\beta_j^0\right)$$

where with $\lambda^0=(\lambda_1^0,\lambda_2^0,\dots,\lambda_m^0)'$ and $\beta^0=(\beta_1^{0'},\beta_2^{0'},\dots,\beta_{m+1}^{0'})'$. $\Gamma(\lambda^0,m,\beta^0)$ is then the sum of the Γ_i (21) across all segments of the data. In the cases where there are neglected breaks in one or more segments, using the results of *Case (ii)* in (52) and (62) we can show that by adding the terms involving $\Gamma_i,\ i=1,2,\dots,m+1$ across segments, the break fractions of the incorrectly estimated breaks (τ_j,τ_{j-1}) will cancel out and so the limit of all terms involving $Var[u_t,v_t^{'}|z_t]$ will be $\Gamma(\lambda^0,m,\beta^0)$. To illustrate this, consider the case where for only one segment j *s.t.* $[\tau_{j-1}T]+1,\dots,[\tau_jT]$ there is one neglected break λ_j (as shown in (52)). Then,

$$\begin{aligned}
\Gamma(\tau(n),n;\lambda^0,\beta^0) \;&=\; (\lambda_1^0-0)\Gamma_1 \,+\dots+\, (\tau_{j-1}-\lambda_{j-1}^0)\Gamma_j \,+\, (\lambda_j^0-\tau_{j-1})\Gamma_j \,+ (\tau_j-\lambda_j^0)\Gamma_{j+1} \,+\, F_1\\
&\quad+(\lambda_{j+1}^0-\tau_j)\Gamma_{j+1} \,+\dots+\, (1-\lambda_m^0)\Gamma_{m+1}\\
&=\; (\lambda_1^0-\lambda_0^0)\Gamma_1 \,+\dots+\, (\lambda_j^0-\lambda_{j-1}^0)\Gamma_j \,+\, (\lambda_{j+1}^0-\lambda_j^0)\Gamma_{j+1} \,+\dots+\, (1-\lambda_m^0)\Gamma_{m+1} \,+\, F_1\\
&=\; \Gamma(\lambda^0,m,\beta^0)+F_1
\end{aligned}$$

since the true vectors of coefficients β_j^0 are stable in each segment j.

Also, it follows directly from the analysis of *Case (ii)* that a straight forward generalization to the case of a segment with more than two neglected breaks will result in a limit function with the basic characteristics of (62). Denote $F(\tau(n),\lambda^0)$ the collection of any terms of the form of F_1

(53), F_2 (63), or the equivalent of the general case of more than two neglected breaks, that will exist when any number of segments $[\tau_{j-1}T]+1,\ldots,[\tau_j T]$ include one, two, or more neglected breaks. As shown in Lemma 1(ii), all these terms will be strictly positive.

Then, the behaviour of the information criterion can be examined in the following cases:

(1) $if\ n = m$. The estimation procedure has identified the correct number of breaks. The following two scenarios are possible,

(1.1) $if\ \tau(n) = \lambda^0$ then there will not be neglected breaks in any segment and by $Case\ (i)$,

$$I\left(\tau(n); n, \hat{\pi}\right) \overset{p.u}{\to} \Gamma(\lambda^0, m, \beta^0)$$

(1.2) $if\ \tau(n) \neq \lambda^0$ then there must exist $j\ s.t.\ [\tau_{j-1}T]+1,\ldots,[\tau_j T]$ contains at least one neglected break, and therefore

$$I\left(\tau(n); n, \hat{\pi}\right) \overset{p.u}{\to} \Gamma(\lambda^0, m, \beta^0)\ +\ F(\tau(n), \lambda^0)$$

where $F(\tau(n), \lambda^0)\ >\ 0$

(2) $if\ n < m$. The estimation procedure has under-fitted the model. Then there must exist a segment $j\ s.t.\ [\tau_{j-1}T]+1,\ldots,[\tau_j T]$ contains at least one neglected break, and

$$I\left(\tau(n); n, \hat{\pi}\right) \overset{p.u}{\to} \Gamma(\lambda^0, m, \beta^0)\ +\ F(\tau(n), \lambda^0)$$

where $F(\tau(n), \lambda^0)\ >\ 0$

(3) $if\ n > m$. Then the following two scenarios are possible

(3.1) if $\tau(n)$ does not contain λ^0 then there must exist $j\ s.t.\ [\tau_{j-1}T]+1,\ldots,[\tau_j T]$ includes at least one λ_i^0 and

$$I\left(\tau(n); n, \hat{\pi}\right) \overset{p.u}{\to} \Gamma(\lambda^0, m, \beta^0)\ +\ F(\tau(n), \lambda^0)$$

where $F(\tau(n), m)\ >\ 0$

(3.2) if $\tau(n)$ contains λ^0 consider

$$D_T\ =\ \left\{I\left(\tau(n); n, \hat{\pi}\right) - I\left(\lambda^0; m\right)\right\}$$

and

$$\begin{aligned} D_T &= Tln\left\{\hat{\Gamma}(\tau(n); n, \hat{\beta})/\hat{\Gamma}(\lambda^0; m, \beta^0)\right\} + T\left\{K(q,n,T) - K(q,m,T)\right\} \\ &= -QLR_T + T\left\{K(q,n,T) - K(q,m,T)\right\} \end{aligned}$$

where QLR_T is the quasi likelihood ratio test for $H_0 : \tau(n) = \lambda^0$ which is a nested test as $\tau(n) \in \lambda^0$. Under its H_0 by standard arguments $QLR_T = Op(1)$, and since $T\{K(q,n,T) - K(q,m,T)\} \overset{p.u}{\to} +\infty$ it follows that

$$D_T \to \infty.$$

Taken together, cases (1), (2), and (3) imply desired result.

References

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

Akaike, H. (1974). 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control*, 19: 716–723.

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

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

Cecchetti, S. G., and Debelle, G. (2006). 'Has the inflation process changed?', *Economic Policy*, 21: 311–352.

Cogley, T., and Sargent, T. J. (2001). 'Evolving post-World War II US inflation dynamics?', *NBER Macroeconomics Annual*, 16: 331–373.

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

Hall, A. R., Han, S., and Boldea, O. (2010). 'Inference regarding multiple structural changes in linear models with endogenous regressors', *Forthcoming in the Journal of Econometrics.*

Hannan, E. J., and Quinn, B. G. (1979). 'The Determination of the Order of an Autoregression', *Journal of the Royal Statistical Society*, 41: 190–195.

Schwarz, G. E. (1978). 'Estimating the dimension of a model', *Annals of Statistics*, 6: 461–464.

Sims, C. A., and Zha, T. (2006). 'Where there regime switches in US monetary policy?', *American Economic Review*, 96: 54–81.

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

Appendix Table A.1. Two-Quarter Ahead Inflation Forecast

Reduced Form Estimates

	1968Q4 - 1975Q2	1975Q3 - 1981Q1	1981Q2 - 1994Q1	1994Q2 - 2005Q4
r_{t-1}	0.042 (0.20)	0.169 (1.81)	0.005 (0.08)	0.348 (2.23)
r_{t-2}	0.086 (0.62)	0.426 (3.17)	0.193 (3.24)	-0.167 (1.11)
π_{t-1}	0.276 (3.20)	0.163 (1.32)	0.088 (1.49)	0.057 (0.94)
π_{t-2}	0.199 (1.65)	0.171 (2.47)	0.268 (6.04)	0.195 (2.47)
$\widetilde{y}_{t-1}$	0.634 (0.87)	0.697 (2.23)	0.061 (0.27)	-0.207 (1.03)
$\widetilde{y}_{t-2}$	-0.935 (1.53)	-0.915 (3.12)	0.013 (0.06)	-0.038 (0.17)
Δy_{t-1}	0.027 (0.57)	-0.144 (2.24)	0.016 (0.74)	-0.058 (2.38)
Δy_{t-2}	0.030 (0.63)	-0.048 (2.68)	0.043 (1.60)	-0.051 (1.88)
lr_{t-1}	0.467 (1.13)	-0.086 (0.40)	0.283 (3.17)	0.416 (3.45)
lr_{t-2}	-0.219 (0.45)	-0.672 (2.23)	-0.326 (4.43)	-0.275 (2.43)
constant	-0.775 (0.27)	6.84 (3.81)	1.227 (3.39)	0.374 (1.37)
R^2	0.881	0.956	0.921	0.801
$\widehat{\sigma}$	0.646	0.487	0.397	0.283

Notes: Break dates are detected using the Hannan-Quinn information criterion, allowing a maximum of five breaks with a minimum of 15% of the sample in each regime. Values in parentheses are Newey-West heteroscedasticity and autocorrelation robust t-ratios (in absolute value), which are computed conditional on the detected breaks.

Appendix Table A.2. Two-Quarter Ahead Unemployment Gap Forecast

Reduced Form Estimates

	1968Q4 - 1975Q3	1975Q4 - 1981Q1	1981Q2 - 1990Q4	1991Q1 - 1999Q4	2000Q1 - 2005Q4
r_{t-1}	0.299 (3.34)	-0.207 (1.73)	0.065 (1.59)	-0.009 (0.09)	0.153 (1.66)
r_{t-2}	-0.220 (3.69)	-0.232 (2.52)	-0.120 (2.54)	-0.065 (0.80)	-0.288 (3.50)
π_{t-1}	-0.097 (1.56)	0.144 (2.69)	-0.004 (0.10)	0.058 (1.32)	-0.066 (0.95)
π_{t-2}	-0.166 (2.43)	-0.005 (0.08)	0.005 (0.18)	0.111 (1.90)	0.065 (1.58)
$\widetilde{y}_{t-1}$	0.584 (2.08)	0.768 (3.67)	1.148 (5.24)	0.715 (4.51)	0.386 (2.61)
$\widetilde{y}_{t-2}$	-0.119 (0.59)	0.094 (0.33)	-0.314 (1.45)	0.422 (2.90)	0.865 (3.90)
Δy_{t-1}	-0.024 (0.67)	0.011 (0.22)	0.095 (4.49)	0.064 (2.57)	0.131 (6.08)
Δy_{t-2}	-0.010 (0.48)	0.063 (2.04)	0.033 (1.90)	0.017 (0.85)	0.117 (5.94)
lr_{t-1}	0.013 (0.13)	0.295 (0.90)	-0.185 (2.51)	0.051 (0.71)	-0.078 (0.70)
lr_{t-2}	-0.467 (1.68)	0.118 (0.52)	0.178 (2.11)	-0.030 (0.31)	-0.111 (0.95)
constant	3.335 (1.86)	-1.700 (0.95)	-0.002 (0.01)	-0.284 (0.83)	0.267 (0.59)
R^2	0.958	0.893	0.979	0.980	0.977
$\widehat{\sigma}$	0.364	0.390	0.276	0.181	0.147

Notes: See Table A.1.

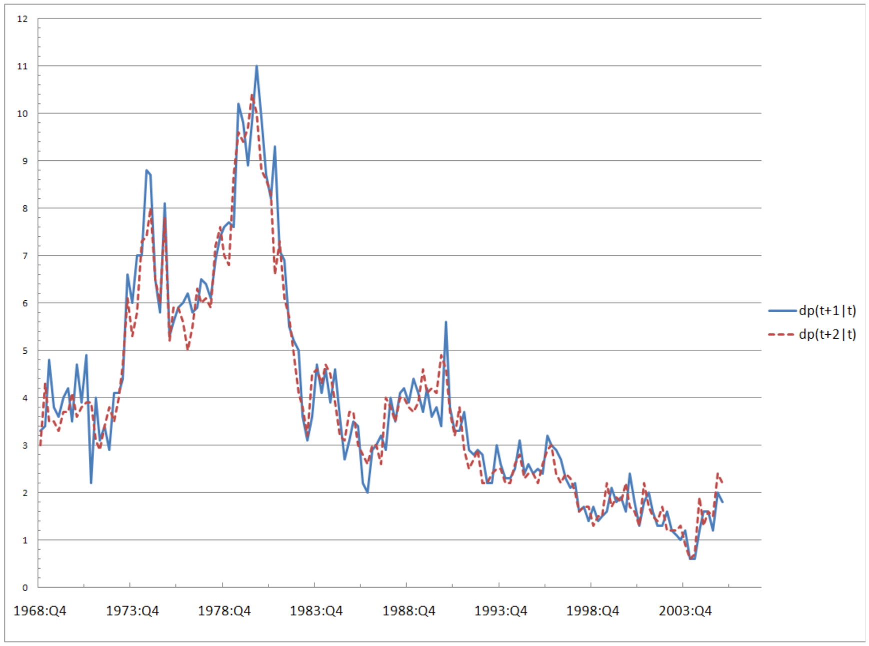

Figure 1. One and two quarter ahead inflation forecasts.

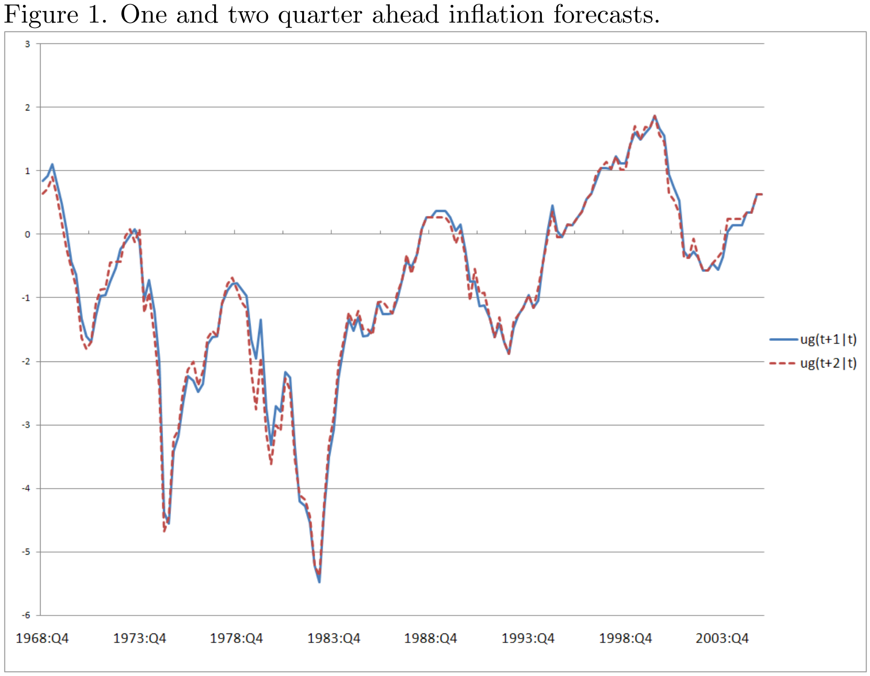

Figure 2. One and two quarter ahead unemployment gap forecasts.

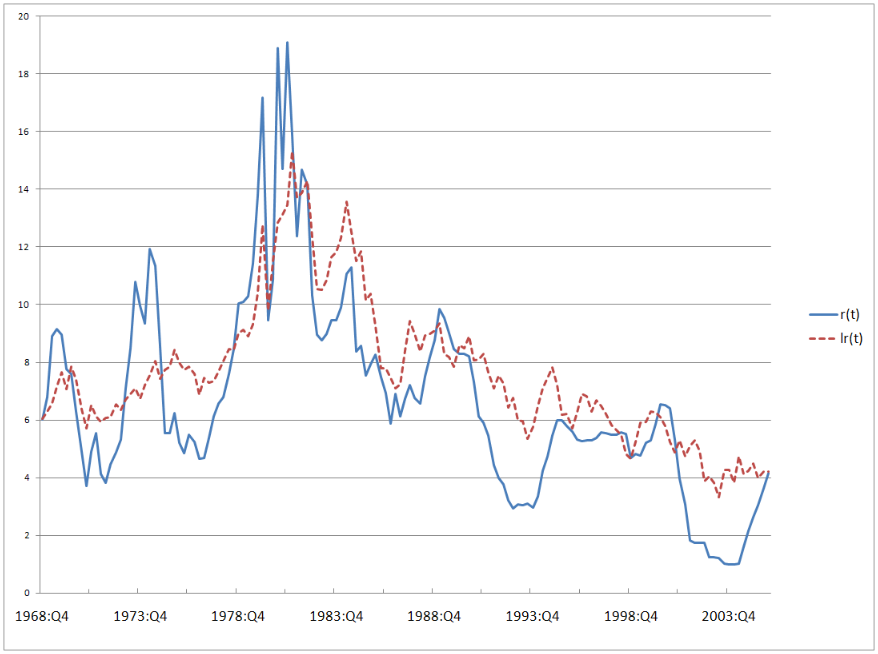

Figure 3. Short and long term interest rates