Estimation and Inference in Unstable Nonlinear Least Squares Models

Otilia Boldea[*] and Alastair R. Hall[†]

May 3, 2012

Abstract

There is compelling evidence that many macroeconomic and financial variables are not generated by linear models. This evidence is based on testing linearity against either smooth nonlinearity or piece-wise linearity, but there is no framework that encompasses both. This paper provides an econometric framework that allows for both breaks and smooth nonlinearity in-between breaks. We estimate the unknown break-dates simultaneously with other parameters via nonlinear least-squares. Using new central limit results for nonlinear processes, we provide inference methods on break-dates and parameter estimates and several instability tests. We illustrate our methods via simulated and empirical smooth transition models with breaks.

JEL classification: C12, C13, C22

Keywords: Multiple Breaks, Nonlinear Least Squares, Smooth Transition

---

[*]Corresponding author. Tilburg University, Dept. of Econometrics and Operations Research, Warandelaan 2, 5000 LE Tilburg, Netherlands, Email: O.Boldea@uvt.nl, Tel. +31-13-466-3219.

[†]University of Manchester, Email: Alastair.Hall@ manchester.ac.uk

1   Introduction

There is widespread evidence that linear models are not able to capture business cycle asymmetry and other salient features of many macroeconomic and financial time series. Departures from linearity are often modeled via smooth nonlinear regressions (smooth transition, other nonlinear autoregressive models) or piece-wise linear regressions with parameter changes (breaks). For example, some authors find that the New Keynesian Phillips curve exhibits breaks - see e.g. Alogoskoufis and Smith (1991), Kang, Kim, and Morley (2009). On the other hand, some find that the nonlinearity is of the smooth type - Cogley and Sargent (2001), Schaling (2004). Similar findings show that the interest rate reaction function is either smoothly nonlinear or it contains breaks - see e.g. Benati and Surico (2008), Zhang, Osborn, and Kim (2008). These findings extend to exchange rate models - Meese and Rose (1991) and Rossi (2006), asset pricing models - Scheinkman and LeBaron (1989) and Franses and van Dijk (2000), and many more.

While it is possible to have a model with both breaks and otherwise smooth nonlinearity, statistical theories for such models do not exist to our knowledge. As a result, smooth nonlinearity and breaks are treated as competing ways of modeling departures from constant parameter linear models. This is undesirable, as they are not a special case of the other. Moreover, they have different policy implications. For example, if interest rates vary smoothly but asymmetrically in expansions versus recessions, it means that the response of the central bank is asymmetric, but in the same way over each business cycle. However, if there is (also) a break, this indicates that the response is different for the more recent business cycles, and that some part of the sample may no longer be helpful for prediction.

In this paper, we provide a framework that allows for both multiple breaks and smooth nonlinearity in-between these breaks. To that end, we consider a univariate

parametric nonlinear model, whose parameters are allowed to change at multiple locations in the sample. The regression function we consider is very general, and as we discuss in Section 2, our methods apply to a large class of nonlinear parametric models, including nonlinear autoregressive (NLAR) models, autoregressive conditional heteroskedasticity models (ARCH) and smooth transition autoregressive (STAR) models. To our knowledge, the literature so far does not propose tests for more than one break in nonlinear models - *inter alia* Anderson and Mizon (1983), Andrews and Fair (1988), Ghysels and Hall (1990), Andrews (1993), Sowell (1996), Hall and Sen (1999) and Andrews (2003), and no econometric results are derived for estimation and inference for one or multiple break points in otherwise smooth nonlinear models.

We propose several tests for multiple breaks, along with several sequential methods to detect the number of breaks. We also propose nonlinear least-squares (NLS) estimators for the breaks and the other parameters of interest, and provide a comprehensive treatment of their asymptotic properties. While our methods can be viewed as the nonlinear counterparts of the methods in Bai and Perron (1998), Bai (1999) and Perron and Qu (2006), our asymptotic analysis is very different. To facilitate our analysis, we derive a new functional central limit result for nonlinear processes. We believe the methods presented in this paper considerably broaden the scope for empirical analysis of economic time series by, for example, allowing tests for instability in the presence of certain asymmetric responses that linear models are unable to capture. We illustrate this via simulations and an application using logistic smooth transition models with potential breaks.

Smooth transition models have been influential in modeling asymmetric responses over the business cycle for monetary policy, macroeconomic and financial variables, since introduced in economics by Granger and Teräsvirta (1993) and Teräsvirta (1994). Via simulations, we show that our methods for detecting breaks work well in smooth transition models with small samples. We also show that both our partial and pure

3

structural change tests do not confuse breaks with nonlinearity. We illustrate our methods via a smooth transition model for the US interest rate reaction function, in a similar spirit to Kesriyeli, Osborn, and Sensier (2006), but adding more recent data. We find evidence for both breaks and nonlinearity. We show that the breaks are in the phases, but not the transition function of the business cycle. This implies that, for predictive purposes, the whole data set is useful.

The paper is organized as follows: Section 2 introduces the model and estimation framework. Section 3 describes the assumptions. Section 4 gives several examples of nonlinear models that satisfy these assumptions. The asymptotic properties of estimators and tests are presented in Section 5. Section 5 also presents extensions to partial structural change and different sequential testing strategies. Section 6 contains simulations for smooth transition models and an application of these models to the US interest rate reaction function. Section 7 concludes. Sketch proofs of Theorems 1, 2 are relegated to the Appendix, while the detailed proofs of all the Theorems can be found in a Supplemental Appendix that is available from the authors upon request.

2   Model

In this section, we introduce a univariate nonlinear model with m unknown breaks:

$$y_t = f(x_t, \theta_i^0) + u_t \qquad t \in I_i^0 = [T_{i-1}^0 + 1, T_i^0] \qquad i = 1, \ldots m + 1 \qquad (1)$$

where $T_0^0 = 0$ and $T_{m+1}^0 = T$ by convention. Here y_t is the dependent variable, x_t $(q \times 1)$ are the regressors, θ_i^0 $(p \times 1)$ are parameters that change at dates T_i^0, $f : \mathbb{R}^q \times \Theta \to \mathbb{R}$ is a known measurable function on $\mathbb{R}$ for each $\theta \in \Theta$, and T is the sample size. To begin, we consider m to be a known finite positive integer, but we allow for the break dates to be unknown to the researcher; we consider the question of how to estimate m in Section 5. For simplicity, let $f_t(\theta) = f(x_t, \theta)$ and denote by

4

$\bar{T}^m \equiv (T_0 = 0, T_1, \ldots, T_m, T_{m+1} = T)$ any m-partition of the sample interval.[1] To further simplify the notation, we will stack column vectors such as θ_i^0 and θ_i into two corresponding $(m+1)p \times 1$ vectors, θ_0^c and θ^c. For a given sample partition and given parameter values θ^c, denote by $S_T(\bar{T}^m, \theta^c)$ the sum of squares.[2]

One of our main goals is to provide a method for estimating the unknown parameters and change points. As in Bai and Perron (1998), the estimation method we propose is based on the least-squares principle[3] and follows in two steps. First, we obtain the sub-sample NLS estimators for each partition:

$$\hat{\theta}_T^c(\bar{T}^m) = \underset{\theta^c(\bar{T}^m)}{\operatorname{argmin}}\ S_T(\bar{T}^m, \theta^c(\bar{T}^m)). \tag{2}$$

Second, we search over all possible partitions to obtain the break-point estimates. The estimates $\hat{T} = (0, \hat{T}_1, \ldots, \hat{T}_m, T)$ for change-points and $\hat{\theta}_T^c = (\hat{\theta}_1, \ldots, \hat{\theta}_{m+1})$ for parameters are obtained as follows:

$$\hat{T} = \underset{\bar{T}^m}{\operatorname{argmin}}\ S_T(\bar{T}^m, \hat{\theta}_T^c(\bar{T}^m))\ \text{ and }\ \hat{\theta}_T^c = \hat{\theta}_T^c(\hat{T}) \tag{3}$$

The above is an NLS estimation with an appropriate modification to allow for multiple break-points, and can be legitimately performed provided that $E[u_t f_t(\theta_i^0)] = 0$ for each $t \in I_i^0 (i = 1, \ldots m+1)$.

---

[1]We use $T_0 = 0$ rather than $T_0 = 1$ for notational convenience, even though the sample interval is $[1, T]$.

[2]We use superscript c to distinguish between $(m+1)p \times 1$ parameter vectors and the $p \times 1$ parameter vectors at which $f_t(\cdot)$ is evaluated.

[3]Note that an extension to more general settings such as generalized method of moments (GMM) is non-trivial because minimizing a GMM criterion over all possible partitions does not yield consistent estimates of the break-fractions indexing the break-points even for linear models and one break under reasonable conditions, see Hall, Han, and Boldea (2012).

3   Assumptions

To derive the statistical properties of our estimators, we establish a framework that combines elements of asymptotic theory in stable nonlinear models and unstable linear models. As pointed out by Hansen (2000), the marginal distributions of regressors and/or errors may change, possibly at different locations in the sample than the population parameters of the equation of interest. Our framework is designed to achieve as much generality as possible with respect to changes in marginal distributions,[4] as well as with respect to other non-stationarities induced by lagged dependent variables that may enter the model concomitantly with parameter breaks. In dealing with nonlinear asymptotics, besides Assumption 1 which is new, we impose usual smoothness and boundedness assumptions. To deal with instability, we assume uniform convergence of certain quantities, jointly in parameters and a partial sum index.

Assumption 1. *Let* $v_t = (x_t', u_t)'$. *Then:*

(i) $\{v_t\}$ *is a piece-wise geometrically ergodic process, i.e. for some finite* $m^* > 0$ *and each sub-sample* $[T_{j-1}^* + 1, T_j^*]$, *where* $T_j^* = [T\lambda_j^*]$, $j = 1, \ldots, m^* + 1$, $\lambda_0^* = 0 < \lambda_1^* < \ldots < \lambda_{m^*}^* < \lambda_{m^*+1}^* = 1$, *there exists a unique stationary distribution* Q_j *such that:*

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

with $0 < \rho < 1$, $A \in \mathcal{F}_{T_{j-1}^*+t}^{T_j^*}$, $B \in \mathcal{F}_{-\infty}^{T_{j-1}^*}$, $\mathcal{F}_k^l$ *is the* σ-*algebra generated by* $(v_k, \ldots, v_l)$, *and* $g_j(\cdot)$ *is a positive uniformly integrable function. If* $\{x_t\}$ *does not contain lagged dependent variables, then (i) holds with* $\{v_t\}$ *augmented by* y_t.

(ii) $\{v_t\}$ *is a* β-*mixing process with exponential decay, i.e. there exists* $N > 0$ *such*

---

[4]Allowing for these types of changes is important in many settings. For example, when estimating a possibly asymmetric (nonlinear) interest rate reaction function, regressors such as output gap or inflation gap may exhibit changes in variance, due to a period of Great Moderation - see e.g. Stock and Watson (2002) - and these changes may occur at different locations than those in the parameters of the equation of interest.

that for $B \in \mathcal{F}^a_{-\infty}$,

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

(iii) $E[u_t f_t(\theta)] = 0$ for each $\theta \in \Theta$.

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

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

Assumption 4. *(i) $S_T(\theta^c) \equiv S_T(\bar{T}^m, \theta^c)$ has a unique global minimum at θ_0^c and $(0, T_1^0, \ldots, T_m^0, T)$; (ii) Let $A_{i,T}(\theta_i^0) = Var \ T^{-1/2} \sum_{t \in I_i^0} u_t F_t(\theta_i^0)$, for $i = 1, \ldots, m+1$, and $A_T(\theta, r) = Var \ T^{-1/2} \sum_{t=1}^{[Tr]} u_t F_t(\theta)$. Then $A_{i,T}(\theta_i^0) \overset{p}{\to} A_i(\theta_i^0)$, and $A_T(\theta, r) \overset{p}{\to} A(\theta, r)$, where the two limits are finite positive definite matrices not depending on T, and the latter convergence holds uniformly in $\theta \times r \in \Theta \times [0,1]$. (iii) Let $D_{i,T}(\theta_i^0) = T^{-1} \sum_{t \in I_i^0} F_t(\theta_i^0) F_t(\theta_i^0)'$ and $D_T(\theta, r) = T^{-1} \sum_{t=1}^{[Tr]} F_t(\theta) F_t(\theta)'$. Then $D_{i,T}(\theta_i^0) \overset{p}{\to} D_i(\theta_i^0)$ and $D_T(\theta, r) \overset{p}{\to} D(\theta, r)$, where the two limits are finite positive definite (p.d.), and the latter convergence holds uniformly in $\theta \times r \in \Theta \times [0,1]$; (iv) $E[f_t(\theta_i^0)] \neq E[f_t(\theta_{i+1}^0)]$, for each $i = 1, 2, \ldots, m$.*

Assumption 5. *$T_i^0 = [T\lambda_i^0]$, where $0 < \lambda_1^0 < \ldots < \lambda_m^0 < 1$.*

Assumption 1(i) can be interpreted as asymptotic stationarity of $\{v_t\}$ within regimes, and it allows for breaks in the marginal distribution of regressors and errors.[5] Additionally, it allows for 'temporary' nonstationary behavior, which is especially useful in the presence of lagged dependent variables in nonlinear models, in

---

[5]Note that m^* as well as λ_j^* are taken as given and are not objects of inference here, unless

which case (1) may induce recurring changes in their marginal distribution. In this case, Assumption 1(i) ensures that even if the process y_t starts in a certain regime at a draw from the nonergodic distribution, it converges to the stable distribution of that regime, so enough homogeneity in the process is preserved to ensure that a uniform central limit theorem still holds in that particular regime.[6]

Assumption 1(ii) ensures that the dependence within and among sub-samples dies out at the same rate as the ergodicity rate. If $m^* = 0$, $\{v_t\}$ admits a Markov chain representation and is geometrically ergodic as in Assumption 1(i), then $\{v_t\}$ is β-mixing with exponential decay, subject to an absolute continuity condition on the starting values - see e.g. Rosenblatt (1971), Mokkadem (1985) - and this connection is often exploited in nonlinear ARCH and general ARCH (GARCH) models - see e.g. Carrasco and Chen (2002). If $\{v_t\}$ is a Markov chain, but $m^* > 0$, then piece-wise geometric ergodicity only implies that the β-mixing coefficients on those sub-samples (thus, for restricted σ-algebras) are exponentially decaying, and we could allow for slower decay across sub-samples. For coherence purposes, we stick to Assumption 1.

Assumption 1(iii) ensures that the model can be estimated via NLS. Assumption 2 and 3 are typical smoothness and boundedness assumptions encountered in nonlinear models. Assumption 4 (i) is the usual NLS identification assumption. Assumptions 4(ii)-(iii) allow substantial heterogeneity in the second moments of regressors and errors. Assumption 4(iv) ensures that the parameter shifts across regimes can be identified. Assumption 5 is a typical assumption for unstable models, allowing the break-fractions to be fixed and hence the break-points to be asymptotically distinct.

all breaks in $\{v_t\}$ either are aligned or coincide with the breaks the parameters of (1), depending on whether $\{x_t\}$ contains lagged dependent variables or not. When the breaks in $\{v_t\}$ are neither aligned nor coincide with the parameter breaks, knowledge of λ_j^* is irrelevant as far as asymptotic distribution results are concerned, but may be of course crucial for both getting consistent estimates of certain asymptotic variances, as well as obtaining the null distribution of stability tests - see Hansen (2000) and Section 5.

[6]In the absence of lagged dependent variables, we need piece-wise ergodicity of $f_t(\theta)$, which we ensure by augmenting $\{v_t\}$ with y_t. Alternatively, one could verify piece-wise ergodicity of y_t on a case by case basis by specifying a functional form for $f_t(\theta)$; we do so in Section 4.

4 Examples

Since Assumptions 2-4 are standard for nonlinear regression functions that are smooth on I_i^0, for $i = 1, \ldots, m+1$, and Assumption 5 is usual for break-point models, we give examples of processes that satisfy Assumption 1. Assumption 1(i)-(ii) implies that in-between the breaks in the regressors T_j^*, the model is geometrically ergodic and β-mixing with exponential decay. For simplicity, we restrict our attention to Markov chains, where geometric ergodicity implies β-mixing with exponential decay, subject to an absolute continuity condition on the first observation. Thus, for the examples below, it suffices to provide conditions for geometric ergodicity on $I_j^* = [T_{j-1}^* + 1, T_j^*]$, with $j = 1, \ldots, m^*$. We give such conditions for several nonlinear models below.

4.1 NLAR Models

Let the model in (1) be an NLAR(p)-model:

$$y_t = f(y_{t-1}, \ldots, y_{t-p}, \theta_i^0) + u_t, \qquad t \in I_i^0 \qquad i = 1, \ldots m+1.$$

Note that the lagged regressors have asymptotically the same break-points as y_t, so geometric ergodicity needs to be verified on I_i^0 only. If $u_t \sim i.i.d$, the model has a Markov-chain representation. Via the usual drift criterion in Meyn and Tweedie (1993), An and Huang (1996) show that y_t is geometrically ergodic, if for some $p \times 1$ vector x, $\sup_{\|x\| \leq K} f(x, \theta) < \infty$ for any K, and $\lim_{\|x\| \to \infty} \frac{f(x,\theta) - \alpha'x}{\|x\|} = 0$, where $\alpha' = (\alpha_1, \ldots, \alpha_p)$ are such that $u^p - \alpha_1 u^{p-1} - \ldots - \alpha_{p-1}u - \alpha_p \neq 0$ for all $|u| \geq 1$. These conditions are satisfied, for example, if $f(\cdot, \cdot)$ is uniformly bounded or an exponential autoregressive function. For more details and examples, see An and Huang (1996).

4.2 Nonlinear ARCH Models

Consider the following nonlinear ARCH (p) model:

$$y_t = h_t^{1/2}(\theta_i^0)\epsilon_t, \ \ h_t(\theta_i^0) = g(Y_{t-1}^2, \theta_i^0), \qquad t \in I_i^0, \qquad i = 1, \dots m+1$$

with $\epsilon_t \sim i.i.d.(0,1)$, and $Y_{t-1}^2 = (y_{t-1}^2, \dots, y_{t-p}^2)'$. Note that $y_t^2 = h_t(\theta_i^0)\epsilon_t^2$, and $y_t^2 = h_t(\theta_i^0) + u_t$, where $u_t = (\epsilon_t^2 - 1)h_t(\theta_i^0)$. The latter can be estimated by NLS when $\sup_{\theta,t} E[h_t^2(\theta)] < \infty$, because then $E[u_t h_t(\theta)] = E[\epsilon_t^2 - 1]E[h_t(\theta_i^0)h_t(\theta)] = 0.$[7] As above, it is sufficient to verify geometric ergodicity on I_i^0. Because $\epsilon_t \sim i.i.d$, (Y_t^2, u_t) is geometrically ergodic on I_i^0 if $h_t(\theta)$ is geometrically ergodic; the conditions are as for the NLAR(p) model above, with $f(\cdot, \cdot)$ replaced by $h(\cdot, \cdot)$. Moment conditions that ensure geometric ergodicity can be found in e.g. Pantula (1988).

4.3 STAR Models

Let the model of interest be a smooth transition model of order p:

$$y_t = x_t'\beta_i \left[1 - G(q_t, \gamma_i, c_i)\right] + x_t'\beta_i^* \left[G(q_t, \gamma_i, c_i)\right] + u_t \qquad (4)$$

$$G(q_t, \gamma_i, c_i) = \frac{exp(\gamma_i(q_t - c_i))}{1 + exp(\gamma_i(q_t - c_i))},$$

for $t \in I_i^0$, (i=1, ..., m+1), with $q_t \perp y_t$ or $q_t = y_{t-1}$, $x_t' = (1, y_{t-1}, \dots, y_{t-p})$ and $\beta_i, \beta_i^*, \gamma_i > 0, c_i$ the unknown parameters. Sufficient conditions for geometric ergodicity can be found in Chan and Tong (1986), Davidson (2002), and require in general that, for all $i = 1, \dots, m+1$, $\sup_{0 \leq \theta \leq 1}\left(\sum_{j=2}^{p+1}|\beta_{i,j} + \theta\beta_{i,j}^*|\right) < 1$, where $\beta_{i,j}, \beta_{i,j}^*$ are the jth elements of β_i, respectively β_i^*. If $p = 1$, then Chan and Tong (1986) show that is sufficient that $\beta_{i,2} < 1$, $\beta_{i,2} + \beta_{i,2}^* < 1$ and $\beta_{i,2}(\beta_{i,2} + \beta_{i,2}^*) < 1$.

---

[7]Note that in general, ARCH and GARCH models only satisfy our framework if the optimization function can be written as a sum of squares, because the proof heavily relies on the properties of least-squares-type criteria.

5 Asymptotics of Nonlinear Models with Breaks

5.1 Estimation

The following theorem states the consistency and rate of convergence of break-fraction estimates for nonlinear models.

Theorem 1. *Under Assumptions 1-5, (i) for each $i = 1, \ldots, m$, let $\hat{\lambda}_i$ be the smallest number such that $\hat{T}_i = [T\hat{\lambda}_i]$. Then $\hat{\lambda}_i \overset{p}{\to} \lambda_i^0$; (ii) for every $\eta > 0$, there exists a finite $C > 0$ such that for all large T, we have $P(\mid T(\hat{\lambda}_k - \lambda_k^0) \mid > C) < \eta$, $(k = 1, \ldots, m)$.*

To see why this theorem requires a new uniform central limit result for nonlinear processes, define $\hat{u}_t = y_t - f_t(\hat{\theta}_k)$, for $t \in \hat{I}_k$ and $d_t = \hat{u}_t - u_t = f_t(\theta_j^0) - f_t(\hat{\theta}_k)$, for $t \in I_j^0 \cap \hat{I}_k$, with $\hat{I}_k = [\hat{T}_{k-1} + 1, \hat{T}_k]$ and $k, j = 1, \ldots, m + 1$. Also, denote $\psi_t(\theta) = u_t f_t(\theta)$. Then:

$$T^{-1} \sum_{t=1}^{T} u_t d_t = T^{-1} \sum_{i=1}^{m+1} \sum_{I_i^0} \psi_t(\theta_i^0) - T^{-1} \sum_{i=1}^{m+1} \sum_{\hat{I}_i} \psi_t(\hat{\theta}_i) = I + II.$$

The proof of consistency crucially rests on showing that $I + II$ is $o_p(1)$. While $I = o_p(1)$ by a simple law of large numbers for nonlinear models, the analysis of II is more complicated as it contains not only sums with random endpoints but summands that depend on the parameter estimators, which in turn depend on the random endpoints. In showing II, we need to derive the following uniform central limit result for nonlinear models.

Lemma 1. *Under Assumptions 1-2 and 3(i)-(ii), $Q_T(\theta, r) = T^{-1/2} \sum_{t=1}^{[Tr]} \psi_t(\theta) = O_p(1)$ uniformly in $\theta \times r \in \Theta \times [0, 1]$.*

Lemma 1 was shown by Caner (2007) under the assumption that $\{\psi_t(\theta)\}$ is a strictly stationary process. In this paper, we relax strict stationarity to piece-wise ergodicity, to make our approach suitable for nonlinear models that contain both

lagged dependent variables and breaks. Even when $m^* = 0$, to our knowledge, no such uniform functional central limit theorem was proven before for a general $f(\cdot, \cdot)$ under geometric ergodicity. From the proof of Lemma 1, it follows that the limiting distribution of $Q_T(\theta, r)$ is for $m^* = 0$ - as in Caner (2007) - a Kiefer process in (θ, r). When $m^* \neq 0$, Lemma 1 indicates that even though $Q_T(\theta, r)$ may not have a unique limit for all r, it is uniformly bounded. As our examples illustrate in Section 4, there are many processes that satisfy this Lemma.

Given the rate of convergence for the break-fraction estimators in Theorem 1, $\hat{\lambda}_i$ can be treated as known in the analysis of $\hat{\theta}_i$, for $i = 1, \ldots, m + 1$. The asymptotic properties of $\hat{\theta}_i$ are given below.

Theorem 2. *Under Assumptions 1-5, $\hat{\theta}_i$ and $\hat{\theta}_j$ are asymptotically independent and $T^{1/2}(\hat{\theta}_i - \theta_i^0) \xrightarrow{d} \mathcal{N}\left(0, \Phi_i(\theta_i^0)\right)$, where $\Phi_i(\theta_i^0) = [D_i(\theta_i^0)]^{-1} A_i(\theta_i^0)[D_i(\theta_i^0)]^{-1}$ for $i, j = 1, \ldots, m + 1, i \neq j$.*[8]

To derive the asymptotic distribution of the break-fractions, we need the following assumption:

Assumption 6. *(i) Assumption 1 holds with $m = m^*, T_i^* = T_i^0, (i = 1, \ldots, m)$ if $\{v_t\}$ does not contain any lagged dependent variables. If v_t contains lags of y_t, then Assumption 1 holds with $m^* = m$ with $T_i^* = T_i^0$, but for $v_t^* = \{y_t, x_t^*\}$ instead of $\{v_t\}$, with x_t^* being all regressors besides the lagged dependent variables; $E[u_t|x_t] = 0$ and $E[u_t u_s | x_k x_l] = 0$ for all $t \neq s$ and all k, l; (ii) Let $D_{T,i}(\theta, r) = T^{-1} \sum_{t=T_{i-1}^0 + 1}^{T_{i-1}^0 + [Tr]} F_t(\theta) F_t(\theta)'$. Then $D_{T,i}(\theta, r) \xrightarrow{p} r D_i(\theta)$, uniformly in $\theta \times r \in \Theta \times [0, \lambda_i^0 - \lambda_{i-1}^0]$, where $D_i(\theta)$ is a p.d. matrix; (iii) Let $A_{T,i}(\theta, r) = Var\, T^{-1} \sum_{t=T_{i-1}^0 + 1}^{T_{i-1}^0 + [Tr]} u_t(\theta) F_t(\theta)$. Then $A_{T,i}(\theta, r) \xrightarrow{p} r A_i(\theta)$, uniformly in $\theta \times r \in \Theta \times [0, \lambda_i^0 - \lambda_{i-1}^0]$, where $A_i(\theta)$ is a p.d. matrix not depending on T, with $A_i(\theta)$ not necessarily the same for all i.*

---

[8]Details on how to construct confidence intervals in different cases can be found in an earlier version of the paper, Boldea and Hall (2010).

This is needed to ensure a non-data dependent asymptotic distribution. For the same purposes, one usually considers parameter shifts that shrink with the sample size - see Bai (1994), Bai (1995) and Bai (1997). Shrinking parameter shifts give rise to an asymptotic approximation to the break-fraction distribution; this approximation is valid in practice for small breaks, that are still large enough to be detectable in the limit. Thus, assume:

Assumption 7. *For* $i = 1, \ldots, m$, $\theta^0_{i+1,T} - \theta^0_{i,T} = \delta_i w_T$, *where* δ_i *are fixed* $p \times 1$ *vectors and* $\{w_T\}$ *is a scalar series such that* $w_T \to 0$ *and* $T^{1/2} w_T^2 \to \infty$ *as* $T \to \infty$.

Similar assumptions are *inter alia* $T^{1/2} w_T \to \infty$, in Bai and Perron (1998) and $T^{1/2} w_T / (log T)^2 \to \infty$ in Qu and Perron (2007). Our assumption allows only shifts of order $T^{-1/4}$ or larger; this assumption cannot be easily relaxed unless the rate in Lemma 1 can be improved upon, and we are not aware of such a result since Lemma 1 itself is new. In practice, our assumption implies that the asymptotic distribution we derive is a reasonable approximation for slightly larger shifts compared to the ones for linear models. Given this assumption, we have:

Theorem 3. *Let* $\phi = \delta'_1 A_2(\theta^0_1)\delta_1 / [\delta'_1 A_1(\theta^0_1)\delta_1]$ *and* $\xi = \delta'_1 D_2(\theta^0_1)\delta_1 / [\delta'_1 D_1(\theta^0_1)\ \delta_1]$. *Under Assumptions 2-7, for* $m = 1$,

$$\frac{[\delta'_1 D_1(\theta^0_1)\delta_1]^2}{\delta'_1 A_1(\theta^0_1)\delta_1} w_T^2 [\hat{k} - k_0] \Rightarrow \underset{v}{\text{argmax}}\, Z(v)$$

where $Z(v) = J_1(-v) - 0.5|v|, v \leq 0$, $Z(v) = \sqrt{\phi} J_2(v) - 0.5\xi|v|, v > 0$, $J_1(v), J_2(v)$ *are two independent standard scalar Gaussian processes defined on* $[0, \infty]$, *and* '$\Rightarrow$' *denotes weak convergence in Skorohod metric.*

The density of $\text{argmax}_v\, Z(v)$ along with details can be found in Bai (1997). Given this density, a confidence interval can be constructed as follows. Let $\hat{\omega}_{1,i} = (\hat{\theta}_2 - \hat{\theta}_1)' \hat{A}_i(\hat{\theta}_1)(\hat{\theta}_2 - \hat{\theta}_1)$, $\hat{\omega}_{2,i} = (\hat{\theta}_2 - \hat{\theta}_1)' \hat{D}_i(\hat{\theta}_1)(\hat{\theta}_2 - \hat{\theta}_1)$, $\hat{D}_i(\theta) = (\hat{T}_i - \hat{T}_{i-1})^{-1} \sum_{t=\hat{T}_{i-1}+1}^{\hat{T}_i}$

$F_t(\theta)F_t(\theta)'$; $\hat{A}_i(\theta)$ a HAC estimator of the long-run variance $A_i(\theta)$, and $\hat{H} = \hat{\omega}_{2,1}^2/\hat{\omega}_{1,1}$. Also, let $\hat{\xi} = \hat{\omega}_{2,2}/\hat{\omega}_{2,1}$ and $\hat{\phi} = \hat{\omega}_{1,2}/\hat{\omega}_{1,1}$. Then, a $100(1-\alpha)\%$ confidence interval for $\hat{k}$ is $(\hat{k} - [c_1/\hat{H}] - 1, \hat{k} + [c_2/\hat{H}] + 1)$, where c_1 and c_2 are respectively the $(\alpha/2)^{th}$ and $(1 - \alpha/2)^{th}$ quantiles for $\mathrm{argmax}_v \, Z(v)$ which can be calculated using equations (B.2) and (B.3) in Bai (1997).[9]

5.2 Tests for Multiple Breaks

This section is concerned with finding the number of breaks m, so far treated as known. To that end, we propose nonlinear equivalents to the tests in Bai and Perron (1998), Perron and Qu (2006) and Bai (1999). Given the results in Theorem 1-3, we show that their limiting distributions carry over from linear settings. The critical values are tabulated in Bai and Perron (1998), Bai and Perron (2006), or can be calculated along with their respective p-values from Bai (1999) and Hall and Sakkas (2012), respectively.

To build sequential strategies for estimating the number of breaks m, we provide three types of tests: for zero against k breaks, for zero against an unknown number of breaks, and for an additional break. The first test is for the hypotheses: $H_0 : m = 0$ vs. $H_A : m = k$, where $k > 0$ is a fixed finite integer. For this purpose, consider a partition $\bar{T}_k$ of the sample interval such that $T_i = [T\lambda_i]$, for $i = 1, \ldots, k$. We also need to restrict each change point to be asymptotically distinct and bounded away from the end-points of the sample. To this end, define $\Lambda_{\epsilon,k} = \{\bar{\lambda}_k \equiv (\lambda_1, \ldots, \lambda_k) : |\lambda_{i+1} - \lambda_i| \geq \epsilon, \lambda_1 \geq \epsilon, \lambda_k \leq 1 - \epsilon\}$, where ϵ is a small number, in practice ranging from 0.05 to 0.20. Then the test is $sup\, F(0:k) = \sup_{\bar{\lambda}_k \in \Lambda_{\epsilon,k}}[T - (k+1)p](SSR_0/SSR_k - 1)/kp$, where SSR_0 and SSR_k are the sums of squared residuals under the null, respectively

---

[9]Theorem 3 can be extended to yield confidence intervals for the multiple break model, because given Assumption 1, the sample segments are asymptotically independent, allowing for the analysis of the limiting distribution to be carried out as in the one break case. Details on constructing these are available upon request.

under the alternative hypothesis. For this and other tests, we need to assume (within-regime) conditional homoskedasticity.[10]

Assumption 8. *(i) $E[u_t^2 \mid x_t] = \sum_{i=1}^{m+1} \sigma_i^2 \mathbf{1}\{t \in I_i^0\}$ for all t; (ii) Part (i) holds with $\sigma_i = \sigma$ for all $i = 1, \ldots, m+1$.*

To test for an unknown number of breaks, i.e. $H_0 : m = 0$ vs. $H_A : m \leq M$, for a fixed M, define the test $D \max(0 : M) = \max_{1 \leq m \leq M} a_m sup \, F(0 : m)$. Details on picking a_m can be found in Bai and Perron (1998). In the simulations, we set $a_m = 1$, in which case this test is denoted $UDmax(0 : M)$.

The third hypothesis of interest is $H_0 : m = \ell$ vs. $H_A : m = \ell + 1$, and can be tested in two ways. The first strategy we propose is as in Bai and Perron (1998), to estimate ℓ breaks, impose them under H_0 and H_A, and for H_A, to estimate an additional break. The test statistic is:

$$seq \, FI(\ell : \ell+1) = \max_{1 \leq i \leq \ell+1} \frac{1}{\hat{\sigma}_i^2} \left\{ S_T(\hat{T}_1, \ldots, \hat{T}_\ell) - \inf_{\tau \in \Delta_{i,\ell}} S_T(\hat{T}_1, \ldots, \hat{T}_{i-1}, \tau, \hat{T}_i, \ldots, \hat{T}_\ell) \right\}$$

where $S_T(T_1, \ldots, T_m) = S_T(\bar{T}^m, \hat{\theta}^c(\bar{T}^m))$, $\Delta_{i,\ell} = \{\tau : \hat{T}_{i-1} + (\hat{T}_i - \hat{T}_{i-1})\eta \leq \tau \leq \hat{T}_i - (\hat{T}_i - \hat{T}_{i-1})\eta\}$ and $\hat{\sigma}_i^2 \xrightarrow{p} \sigma_i^2$. An alternative strategy is to estimate ℓ breaks under H_0, and re-estimate $\ell+1$ breaks under H_A. The resulting test statistic is the nonlinear counterpart of the test in Bai (1999), $seq \, FII(\ell : \ell+1) = [T-(\ell+2)p](SSR_\ell/SSR_{\ell+1} - 1)$, where $SSR_\ell = S_T(\hat{T}_1, \ldots, \hat{T}_\ell)$ and $SSR_{\ell+1} = S_T(\hat{T}_1, \ldots, \hat{T}_\ell, \hat{T}_{\ell+1})$ are the estimated sum of squared residuals for ℓ, respectively $\ell + 1$ breaks, in $\Lambda_{\epsilon,\ell}$, respectively $\Lambda_{\epsilon,\ell+1}$. The following results are proven in the Supplemental Appendix:

Theorem 4. *Under Assumptions 2-6 and 8(i), the null asymptotic distributions for the tests are: (i) for $sup \, F(0 : k)$, as in Bai and Perron (1998), Proposition 6; (ii) for $D \max(0 : M)$ as in Bai and Perron (1998), Section 4.2; (iii) for seq $FI(\ell : \ell+1)$, as*

---

[10] Extensions to *sup* Wald tests without this assumption can be found in a previous version of the paper, Boldea and Hall (2010).

in Bai and Perron (1998), Proposition 7; (iv) for seqFII ($\ell : \ell+1$) - with Assumption 8(ii) replacing 8(i)- as in Bai (1999), Theorem 1 and Corollary 1.

5.3   Sequential Estimation of the Number of Breaks

Using the test statistics above, we suggest four simple sequential methods for obtaining an estimator, $\hat{m}_T$ say, of the number of breaks.

Specifically, in the first step of the sequential estimation, use either sup $F(0 : 1)$ or $UDmaxF(0 : M)$, to test $H_0 : m = 0$. If this null is not rejected , then $\hat{m}_T = 0$; else proceed to the next step. On the second step, use $seq\,FI(1 : 2)$ or $seqF\,II(1 : 2)$ to test $H_0 : m = 1$ vs. $H_a : m = 2$. If $seqFI(1 : 2)$ or $seqFII(1 : 2)$ does not reject, then $\hat{m}_T = 1$; else proceed to the next step. On the ℓth step, by means of $seq\,FI(\ell : \ell + 1)$ or $seqF\,II(\ell : \ell + 1)$, test $H_0 : m = \ell$ vs. $H_A : m = \ell + 1$ and if H_0 is not rejected, then $\hat{m}_T = \ell$; else proceed to the next step. This sequential procedure stops when M, the ceiling on the number of breaks, is reached. If all statistics in the sequence are significant then $\hat{m}_T \geq M$. Thus, we have four strategies: (1) sup $F(0 : 1)$ first, then $seq\,FI$; (2) sup $F(0 : 1)$ first, then $seq\,FII$; (3) $UDmax\,F(0 : M)$ first, then $seq\,FI$, and (4) $UDmax\,F(0 : M)$ first, then $seq\,FII$.[11]

5.4   Extension to Partial Structural Change

In this section, we provide a test for partial structural change, that is, for breaks in only a subset of parameters. For simplicity, we only consider the test for one break, but the results can be extended to multiple breaks in a similar fashion as for pure structural change. Let $\theta_i^0 = (\beta_i^0, \gamma_i^0)$, $i = 1, 2$, where $\beta_1^0 = \beta_2^0$, and γ_i^0 are $p_1 \times 1$, with $p_1 < p$. The null and alternative hypothesis are: $H_0 : \gamma_1^0 = \gamma_2^0$ vs. $H_A : \gamma_1^0 \neq \gamma_2^0$, and

---

[11]The finite sample properties of strategy (1) and (3) have been investigated for linear models by Bai and Perron (2006), and found to work well in the absence of serial correlation in the errors. In the presence of serial correlation, similar strategies as the ones described in Boldea and Hall (2010), Section 5.3, have also been studied in Bai and Perron (2006) for linear models and found to work well in finite samples.

the test statistic is sup $F_R(0:1) = \sup_{\lambda_1 \in \Lambda_{\epsilon,1}}[T-p](SSR_0/SSR_1^R(\lambda_1) - 1)/p_1$, where $SSR_1^R(\lambda_1)$ is obtained as SSR_1 for candidate break-point $[T\lambda_1]$, but with imposing the restriction $\beta_1 = \beta_2$.[12]

Theorem 5. *Under Assumptions 2-6 and 8(i), the null asymptotic distribution of* sup $F_R(0:1)$ *is as in Perron and Qu (2006), equation (7).*

6   STAR Models with Breaks

In this section, we analyze the small sample performance and the practical implications of a STAR model with potential breaks. STAR models have been defined in economics by Granger and Teräsvirta (1993), and widely used for modeling asymmetric behavior of many macroeconomic variables, such as exchange rates, interest rates, asset prices, unemployment. To discuss the implications of such models in the presence of breaks, consider the simple STAR model with one break in (4), for $m = 1$, and $x_t' = (1, y_{t-1})$. The parameters of interest are $\theta_i^0 = (\beta_i, \beta_i^*, \gamma_i, c_i)$, along with the unknown break T_1^0. For each sub-sample I_i^0, y_t transits in a smooth fashion between two linear regimes: $x_t'\beta_i$ and $x_t'\beta_i^*$, and the transition is characterized here by the logistic function $G(\cdot, \cdot, \cdot)$, taking values in (0,1). The variable q_t is the state variable that governs the transition. The smoothness parameter γ_i governs the smoothness of transition: the larger it is, the faster the transition. The so called "threshold" value c_i governs each observation tending to one regime or the other depending on the value of q_t.

If q_t is a business cycle proxy, we can interpret the parameters β_i and β_i^* as governing the "phase" of the business cycle (expansion or recession), and parameters γ_i, c_i govern the transition between phases. If $\theta_1^0 \neq \theta_2^0$, we have a break at T_1^0, and it can be interpreted as a change in the data generating process for y_t with recent business

---

[12]The extension to allowing the rest of the $p - p_1$ parameters to change under the null and alternative, at the same break-point that the other $p - p_1$ parameters, requires using the break-point as a pre-estimate. This may make the distribution of such a test complicated to derive, and we leave this to future research.

cycles, after T_1^0. There is considerable evidence that this is the case for example, for interest rates, see Alogoskoufis and Smith (1991), Kang, Kim, and Morley (2009). On the other hand, there is considerable evidence that the interest rate reaction function is smoothly asymmetric over each business cycle - see e.g. Schaling (2004), Dolado, María-Dolores, and Ruge-Murcia (2004), Kim, Osborn, and Sensier (2005), Kesriyeli, Osborn, and Sensier (2006). In our framework, we can allow for both features jointly.

We provide both simulations for a simple STAR model, and an application to the Federal funds interest rates. We allow for both pure and partial structural change. The distinction is important in practice; for example, if the linear parameters β_i, β_i^* change, for interest rates this can be interpreted as monetary policy changes in the expansion/recession phases. If γ_i, c_i change, the transition of monetary policy within the business cycle is different than it used to be. Moreover, unlike pure structural change, partial structural change implies that the whole dataset needs to be used for prediction.

6.1   Simulations

All of our simulations, unless mentioned otherwise, are done in the context the one-transition STAR model in (4), similar to the models in Franses and van Dijk (2000). Even without breaks, STAR models are computationally challenging for two reasons. First, because they require large samples to ensure enough variation in the transition function to identify the linear parameters. Second, because the parameter estimates are usually computed in two steps: an inner loop where the concentrated nonlinear objective function is maximized and the nonlinear parameters conditional on the linear ones are obtained, and another step where a back-and-forth iteration between nonlinear and linear parameter optimization is performed. The first step is computationally challenging as the concentrated nonlinear objective function is usually flat in small samples.

18

Although for linear models, Bai and Perron (2003) show that one need not perform more than $T(T+1)/2$ operations to find the estimated partition, for nonlinear models, each of this operation requires a new nonlinear optimization with two steps.[13] To make sure that the computational errors in each step are small, we use a KNITRO optimization function written for MATLAB, that is specifically designed to concomitantly monitor the relative size of the computation and optimization errors in different steps.[14] Despite computational complexity, we are able to show that, even for small sample sizes, our method is accurate in detecting reasonably large breaks.

For the data generating process (DGP), unless mentioned otherwise, we use the STAR as in (4), with $x_t' = (1, y_{t-1})$, $c_i = 0$, and m breaks, rewritten as:

$$y_t = x_t' \beta_i^1 + x_t' \beta_i^2 \, G(q_t, \gamma_i, c_i) + u_t \qquad \text{for } t \in I_i^0 \tag{5}$$

where $\beta_i^1 = \beta_i$ and $\beta_i^2 = \beta_i^* - \beta_i$ in (4). Let $u_t \perp q_t$, $u_t \sim i.i.d. \, N(0, 0.05^2)$, and $q_t \sim i.i.d. \, N(1, 1)$. We initialize the time series y at 0, and apply a burn-in of 100 observations before considering our sample T. Let $\theta^{i\prime} = (\beta_i^{1\prime}, \beta_i^{2\prime}, \gamma_i')$ and the cut-off $\epsilon = 0.15$. All results are for 1000 simulations.

We first analyze the small sample properties of our testing procedures for pure structural change. Consider the following DGPs: (A I) $T = 100$, $m = 0$, $\theta_1' = (0, 0.5, 0, -0.5, 0.5)$; (A II) $T = 100$, $m = 1$, $T_1^0 = 50$, $\theta_1' = (0, 0.5, 0, -0.5, 0.5)$, $\theta_2' = (-0.5, -0.5, 0.5, 0.5, 1)$; (A III) $T = 150$, $m = 2$, $T_1^0 = 50$, $T_2^0 = 100$, $\theta_1' = \theta_3' = (0, 0.5, 0, -0.5, 0.5)$, $\theta_2' = (-0.5, -0.5, 0.5, 0.5, 1)$.

The results for $sup \, F$, $seq \, FI(\ell : \ell + 1)$, $seq \, FII(\ell : \ell + 1)$ and $UDmax \, F(0 : 2)$ are presented in Table 1. The critical values for $UDmax \, F(0 : 2)$ were not available in

---

[13]Note that unlike for linear models, for nonlinear models there is no exact updating formula for the sum of squared residuals, and it is not clear whether approximate updating rules such as the "unscented" Kalman filter work in the presence of breaks.

[14]For STAR estimation, our starting point is the code in McAleer and Medeiros (2008) that is available online at *https://sites.google.com/site/marcelocmedeiros/Home/codes*.

the literature and are computed via additional simulations.[15]. The critical values for $seq\,FII(\ell : \ell+1)$ are computed from Bai (1999), following his suggestion by assuming that the second order term of the cdf on pp. 305 is negligible; the rest are available from Bai and Perron (2003). Table 1 reports the rejection frequency of all these tests at different nominal levels. The tests sizes are close to nominal for one and two breaks. All tests have large power, and this is expected for large enough breaks; see Table 4 for smaller breaks. Although it is known that parameter estimates have a large in-sample bias for most STAR models in small samples,[16] we show that despite this bias, the tests have reasonable size and power even for samples sizes of $T = 100$, respectively $T = 150$. The tests for A I are undersized, likely related to the true DGP having no intercept - Table 3 for DGP B I shows improvement.

Table 2 reports the empirical distribution of $\hat{m}_T$ for A I - A III defined above, detected via the four sequential strategies described in Section 5.3, at the 5% level. All strategies seem to perform well; the ones based on $seqFI(\ell : \ell + 1)$ work slightly better. If the number of breaks is detected accurately, we find that the estimated breaks for DGPs A II and A III are either equal or one observation away from the true break. This implies that the empirical coverage of the break-point confidence intervals is 100% at any significance level; this is expected because the break-point estimation seems to be very accurate for large enough breaks.

Next, we compare tests for pure and partial structural change. We consider partial structural changes where either γ_i or $\beta_i' = (\beta_i^{1\prime}, \beta_i^{2\prime})$ change. These tests are useful to identify whether the break-point is in the transition function (γ_i changes across i) or the linear regimes (β_i changes across i).

For partial structural change, one needs an initial estimate for the break - see Perron and Qu (2006), pp. 10, followed by iterations to obtain the (implicitly defined

---

[15]We thank Nikolaos Sakkas for his help.

[16]To save space, parameter estimates simulations are not provided, but they are available upon request.

in the tests) break-point estimate of the restricted structural change models. As suggested in Perron and Qu (2006), we use a pure structural change break-point estimate to initialize the iterations. Table 3 reports the size and power of both pure and partial structural change for one break, $T = 100$, under the following DGPs: (B I - no change) $m = 0$, $\theta'_1 = (-0.5, -0.5, 0.5, 0.5, 1)$; (B II - change in β_i only): $m = 1$, $\beta'_1 = (0, 0.5, 0, -0.5)$, $\beta'_2 = (-0.5, -0.5, 0.5, 0.5)$, $\gamma_1 = \gamma_2 = 1$; (B III - change in γ_i only): $m = 1$, $\beta'_1 = \beta'_2 = (0, 0.5, 0, -0.5)$, $\gamma_1 = 0.1$, $\gamma_2 = 1$; (B IV - $m = 0$) but the DGP is:

$$y_t = x'_t \beta^1 + x'_t \beta^2 G(q_t, \gamma_1, 0) + x'_t \beta^3 G(q_t, \gamma_2, 0) + u_t$$

with $\beta^1 = (0, 0.5)'$, $\beta^2 = (0, -0.5)'$, $\beta^3 = (0.5, 0.5)'$, $\gamma_1 = 0.1$, $\gamma_2 = 1$. The latter is a two-transition model with no breaks, which we use to check whether our tests reject in the presence of misspecified nonlinearity. Table 3 indicates that all tests have 100% power. The test for changes in β_i only seems oversized. This is expected since the parameters β_i^2 are estimated less accurately when the variation in $G(q_t, \gamma_i)$ is small. Thus, we believe that this is not a problem of our test, and may rather be related to poor identification of STAR models in small samples. This problem is less evident for the last DGP. The results for the DGP B IV suggest that all structural change tests, pure or partial, designed for structural change specifically, don't have power against additional nonlinearity. This is reassuring as it indicates that our tests will not confuse breaks with nonlinearity.

Finally, we present simulations with smaller breaks. We pick in our view the worst case scenario DGP, in (5) but with only the nonlinear parameters changing. Thus, $\beta'_1 = \beta'_2 = (-0.5, -0.5, 0.5, 0.5)$, and $\gamma_2 = 1$, $c^* \equiv 1 - \gamma_1$ is the magnitude of parameter change, and c^* takes the values in Table 4. We see that both the pure and the partial structural change tests have power 1 for $c^* = 0.9$, and still quite large power for medium breaks $c^* = 0.5$. It seems that we need $c^* = 0.3$ or lower for the power of the tests to dramatically deteriorate. We also see that, as expected, the power of the

21

pure structural change tests in the presence of partial structural change deteriorates faster as the breaks get smaller.

6.2 Application to the US Interest Rate Reaction Function

As previously discussed, there is considerable evidence of either parameter change or smooth nonlinearity in the US interest rate function, but not both.[17]

For example, Kesriyeli, Osborn, and Sensier (2006) find that the US interest rate reaction function between 1984-2005 exhibits smooth transition in two state variables: a quarterly lagged difference in the interest rates and time. They find a large smoothness parameter (1082) for the time transition function, which may be indicative of a break, but because of the absence of an econometric framework, they cannot test for breaks explicitly.[18]

Our framework can be used to test for breaks in the presence of smooth nonlinearity. To that end, we consider a simplified version of the US interest rate reaction in Kesriyeli, Osborn, and Sensier (2006), with m potential breaks, that is, for $i = 1, \ldots, m + 1$,

$$r_t = x'_t \beta_i [1 - G(\Delta_3 r_{t-1}; \gamma_i, c_i)] + x'_t \beta_i^* \, G(\Delta_3 r_{t-1}; \gamma_i, c_i) + u_t, \qquad t \in I_i^0$$

Here, r_t is the federal funds rate, $\Delta_3 r_{t-1} = r_{t-1} - r_{t-4}$ is the same transition variable as in Kesriyeli, Osborn, and Sensier (2006), and $x'_t = (1, r_{t-1}, r_{t-2}, y_{t-1}, \pi_{t-1})$, where y_t and π_t are the output and inflation gap, respectively. Our data comprises US monthly observations from 1984:1-2010:6; the federal funds rate and CPI inflation

---

[17]Note that most of the literature - with the exception of a few studies, see e.g. Bec, Salem, and Collard (2002) - does not model Federal funds interest rates via threshold or Markov switching models, because while the Fed may change interest rate reactions infrequently (in the form of breaks), it is unlikely that it would drastically changes interest rates frequently (as thresholds and switching models would imply) so as not to generate large volatility in the markets.

[18]The tests they use, based on Eitrheim and Teräsvirta (1996), have some power against breaks, but they are not designed against breaks explicitly.

rates are computed from the dataset FRED2, and the output (real GDP) at monthly frequency is the Stock-Watson proxy. From output and inflation, we construct gap data using the Hodrick-Prescott filter with constant equal to 126400, as usual for data at monthly frequency.

The instability tests[19] reported for a cut-off $\epsilon = 0.20$, are in Table 1.

For all strategies 1-4 described in Section 5.3, we find evidence of one break, at 1990:9, with 95% confidence interval [1990:8, 1990:10]. This implies that post-1990, the interest rate reaction function has changed. Indeed, after 1990, the interest rates were raised more often than before, implying that the Fed reacted more drastically in expansions (phases of the business cycle). The tests for partial structural change in $\beta_i = (\beta_i', \beta_i^{*\prime})'$ and in γ_i are 74.06, and 0.93 respectively, with p-values 0.00 and 1.00. Thus, the change is in the phases of the business cycle, rather than in the transition between phases.

The parameter estimates in Table 6 indicate that the interest rate reacts more to the inflation gap compared to the output gap changes, it is asymmetric over the business cycle, and has changed in recent periods. Because of partial structural change, our results imply that the whole data should be used for prediction.

7 Conclusions

In this paper, using new empirical process results, we develop a framework for estimating and testing in NLS models with multiple breaks. By construction, our method nests nonlinearities and breaks, and is useful in practice for jointly modeling breaks and nonlinearity, without considering them as competing alternatives.

Our method can be a powerful tool for empirical macroeconomic and financial modeling. For example, we show that the interest rate reaction function has both

---

[19]Since we have $p = 12$ parameters, for which critical values are not available, we use the methods in Hall and Sakkas (2012) to compute p-values for our tests.

smooth asymmetric behavior, and it varies with more recent data. Our method can also be applied to many macroeconomic variables for which there is evidence of either breaks or nonlinearity, but the econometric analysis so far has not allowed for both.

A number of interesting questions remain, such as the extension of our methods to (i) multivariate models; (ii) more general forms of nonlinearity including threshold models and Markov switching models which our assumptions do not cover; (iii) time series that are not necessarily piece-wise geometrically ergodic. We leave these issues to future research.

8   Appendix

This Appendix only contains sketch proofs of Lemma 1 and Theorems 1,2. The full proofs along with the ones for Theorems 3-5 are relegated to a Supplemental Appendix, available from the authors upon request. Regarding notation, we use $\| \cdot \|$ for both the Euclidean and the matrix norm $\|A\| = [\text{tr}(A'A)]^{1/2}$, and let $\psi_t(\theta) \equiv u_t f_t(\theta)$, $\Psi_t(\theta) \equiv u_t F_t(\theta)$.

Proof of Lemma 1*.*
Consider the cases $m^* = 0$ and $m^* = 1$; the extension to $m^* > 1$ is immediate and omitted for simplicity.

Case $m^* = 0$. In this case, we need to show that for any $\epsilon > 0$, there exists a $\eta_\epsilon > 0$ and a $T_\epsilon > 0$ such that for any $T \geq T_\epsilon$, $P\left(\sup_{\theta \times r} |Q_T(\theta, r)| > \eta_\epsilon\right) < \epsilon$.

This was shown under Assumptions 1,2,3(i)-(ii) by Caner (2007) for strictly stationary processes; here we show that the difference between the distribution function of $Q_T(\theta, r)$ started at $\psi_0(\theta)$ and the distribution function of same process started at its stationary distribution, say $Q(\cdot)$, is $o(1)$ uniformly in $\theta \times r$. To that end, define a sequence $\{b_T\}$ of positive integers such that $b_T \to \infty$ and $b_T/\sqrt{T} \to 0$, and let

$D(i, j, \theta) = |T^{-1/2} \sum_{t=i}^{j} \psi_t(\theta)|$. Then $P(\sup_{\theta \times r} D_T(\theta, r) > \eta_\epsilon)$ is less than:

$$P(\sup_{\theta} D(1, b_T, \theta) > \eta_\epsilon/2) + Q(\sup_{\theta \times r} D(b_T + 1, [Tr], \theta) > \eta_\epsilon/2)$$

$$+ \{|P(\sup_{\theta \times r} D(b_T + 1, [Tr], \theta) > \eta_\epsilon/2) - Q(\sup_{\theta \times r} D(b_T + 1, [Tr], \theta) > \eta_\epsilon/2)|\}$$

$$= I + II + \{III\}.$$

Note that $I \leq P(\sup_{\theta, t} \frac{b_T}{\sqrt{T}} |\psi_t(\theta)| > \frac{\eta_\epsilon}{2}) = o(1)$, uniformly in $\theta \times r$, by Assumption 3(ii). By Caner (2007), Lemma 1, pp. 37, $II = o(1) + \epsilon$, for any $\epsilon > 0$ and $T \geq T_\epsilon$, with the $o(1)$ term uniform in $\theta \times r$ (u.θ.r.). It remains to show that $III = o(1)$ u.θ.r. To that end, in Assumption 1(i), let $\mu(A) = |P(A|B) - Q(A)|$. Since P and Q are probability measures, $P - Q$ is a signed measure μ_*, and by the Hahn-Jordan decomposition, there exist two positive measures μ_*^+ and μ_*^- such that $\mu_* = \mu_*^+ - \mu_*^-$. Hence, $\mu = |\mu_*| = \mu_*^+ + \mu_*^-$. Since $\mu(\varnothing) = 0$ it follows that μ is a measure, therefore sub-additivity holds. Let $E_1 = [\sup_{\theta \times r} D(b_T + 1, [Tr], \theta) > \eta_\epsilon/2]$, $E_2 = [\sup_r \sum_{t=b_T+1}^{[Tr]} \sup_\theta |\psi_t(\theta)| > \eta_\epsilon \sqrt{T}/2]$, $E_3 = [\sum_{t=b_T+1}^{T} \sup_\theta |\psi_t(\theta)| > \eta_\epsilon \sqrt{T}/2]$ and $E_4 = \cup_{t=b_T+1}^{T} [\sup_\theta |\psi_t(\theta)| > \eta_\epsilon \sqrt{T}/[2(T - b_T)]]$. Letting the superscript c denote the complement of a set, we have $E_1 \subseteq E_2 \subseteq E_3 = (E_3 \cap E_4) \cup (E_3 \cap E_4^c) = E_3 \cap E_4 \subseteq E_4$. Using sub-additivity of μ, and noting that $A_t = [\sup_\theta |\psi_t(\theta)| > \frac{\eta_\epsilon \sqrt{T}}{2(T - b_T)}] \in \mathcal{F}_t^\infty$ is an event started at $B \in \mathcal{F}_{-\infty}^0$, one can show that $III = \mu(E_1) \leq \mu(E_4) = o(1)$, u.$\theta$.r.

Case $m^* = 1$. Similarly to $m^* = 0$, $P(\sup_{\theta \times r} D(1, [Tr], \theta) > \eta_\epsilon)$ is less than:

$$P(\sup_{\theta \times (0 \leq r \leq \lambda_1^*)} D(b_T + 1, [Tr], \theta) > \eta_\epsilon/2) + P(\sup_{\theta \times (\lambda_1^* < r \leq 1)} D(b_T + 1, [Tr], \theta) > \eta_\epsilon/2)$$

$$+ o(1) = IV + V + o(1), u.\theta.r.$$

Let $B_i = Q_i \left(\sup_{\theta \times (0 \leq r \leq \lambda_1^*)} D_T(\theta, r) > \eta_\epsilon/8 \right)$. Then $IV \leq o(1) + B_1$, and $V \leq o(1) + B_1 + B_2, u.\theta.r$, for $T \geq T_\epsilon$; this implies the desired result for $m^* = 1$. $\square$

Proof of Theorem 1.

Part (i). From Lemma 1, $T^{-1} \sum_{t=1}^{T} u_t d_t = o_p(1)$ uniformly in $\theta \times r$. Also via Lemma 1, if one break-fraction does not converge to its true value, then one can show that $T^{-1} \sum_{t=1}^{T} d_t^2 > C$ with probability $> \epsilon$ for large T uniformly in θ, which cannot hold by the definition of sum of squared residuals (SSR) given that $T^{-1} \sum_{t=1}^{T} u_t d_t = o_p(1)$. By contradiction, $\hat{\lambda}_i \overset{p}{\to} \lambda_i^0$.

Part (ii). As in Bai and Perron (1998), without loss of generality, assume only three breaks, and prove Theorem 1 (ii) for $\hat{\lambda}_2$ and $\hat{T}_2 < T_2^0$. For any $\epsilon > 0$, define $V_\epsilon = \{(T_1, T_2, T_3) : \mid T_i - T_i^0 \mid \leq \epsilon T \ (i = 1, 2, 3)\}$. Since $\hat{\lambda}_i \overset{p}{\to} \lambda_i^0$, $\lim P\{(\hat{T}_1, \hat{T}_2, \hat{T}_3) \in V_\epsilon\} = 1$. Hence, we need only examine the behavior of break-points contained in V_ϵ. For $C > 0$, define: $V_\epsilon(C) = \{(T_1, T_2, T_3) : \mid T_i - T_i^0 \mid \leq \epsilon T \ (i = 1, 2, 3); \ T_2^0 - T_2 > C\}$, and show that the probability that the break-points are contained in $V_\epsilon(C)$ is small. To that end, denote by $S_T(T_1, T_2, T_3)$ the minimized SSR for a given 3-break-partition $(0, T_1, T_2, T_3, T)$. By definition, $S_T(\hat{T}_1, \hat{T}_2, \hat{T}_3) \leq S_T(\hat{T}_1, T_2^0, \hat{T}_3)$. Let $\Delta_2 = T_2 - T_2^0$. We will show that for any $\eta > 0$, we can pick ϵ and C such that on $V_\epsilon(C)$, we have:

$$P \left\{ \min_{V_\epsilon(C)} (\Delta_2)^{-1} [S_T(T_1, T_2, T_3) - S_T(T_1, T_2^0, T_3)] < 0 \right\} < \eta, \ \text{for } T \geq T(\eta). \quad (6)$$

Equation (6) implies that for large T, with probability $\geq 1 - \eta$, $S_T(\hat{T}_1, \hat{T}_2, \hat{T}_3) > S_T(\hat{T}_1, T_2^0, \hat{T}_3)$, contradicting the sum of squares minimization definition; thus, $\hat{T}_2 \notin V_\epsilon(C)$, completing the proof. We prove (6) by locating the dominating terms in $(\Delta_2)^{-1} [S_T(T_1, T_2, T_3) - S_T(T_1, T_2^0, T_3)]$ and showing that they are positive on $V_\epsilon(C)$ with large probability for large T; to that end, we make extensive use of Lemma 1 - see Supplemental Appendix for details. $\square$

Proof of Theorem 2.

As usual, we need to show uniform convergence of the minimand, and use uniqueness to establish consistency of parameter estimates. Let some partition of the sample

interval be $(0, T_1, \ldots, T_m, T)$. Let $S_{T, \hat{I}_i}(\theta) = T^{-1} \sum_{t=T_{i-1}}^{T_i} u_t^2(\theta)$ be the partial sum of squares in interval $\hat{I}_i = [\hat{T}_{i-1} + 1, \hat{T}_i]$. Moreover, let $\hat{I}_i \triangledown I_i^0 = (\hat{I}_i \setminus I_i^0) \cup (I_i^0 \setminus \hat{I}_i)$, and define as indicator function $\iota_i : \hat{I}_i \triangledown I_i^0 \to \{-1, 1\}$, where $\iota_i(t) = \iota_{i,t} = 1$, if $t \in \hat{I}_i \setminus I_i^0$, and $\iota_{i,t} = -1$, if $t \in I_i^0 \setminus \hat{I}_i$. Then $S_{T, \hat{I}_i}(\theta) - S_{T, I_i^0}(\theta)$ is equal to $\sum_{\hat{I}_i \triangledown I_i^0} \iota_{i,t} [T^{-1} u_t^2] + \sum_{\hat{I}_i \triangledown I_i^0} \iota_{i,t} [T^{-1} d_t^2(\theta, \theta_i^0)] + \sum_{\hat{I}_i \triangledown I_i^0} \iota_{i,t} [T^{-1} 2 u_t \, d_t(\theta, \theta_i^0)]$. By Theorem 1, there are $\leq 2C$ integers in $\hat{I}_i \triangledown I_i^0$. By Lemma 1, $S_{T, \hat{I}_i}(\theta) - S_{T, I_i^0}(\theta) = o_p(1)$; then standard NLS asymptotics yields, under Assumptions 1-4, $\hat{\theta}_i \xrightarrow{p} \theta_i^0$. Since also mean value expansions $T^{1/2} \, \partial S_{T, \hat{I}_i} / \partial \theta$ around θ_i^0 are uniformly within $o_p(1)$ of those using the true break-points, $\hat{\theta}_i$ have the distribution in Theorem 2. Asymptotic independence of $\hat{\theta}_i$ and $\hat{\theta}_j$ for $i \neq j$ follows from Assumption 1. $\qquad \square$

Proof of Theorems 3-5. See Supplemental Appendix. $\qquad \square$

References

Alogoskoufis, G., and Smith, R. (1991). 'The Phillips Curve, the Persistence of Inflation, and the Lucas Critique: Evidence from Exchange-rate Regimes', *American Economic Review*, pp. 1254–1275.

An, H., and Huang, F. (1996). 'The Geometrical Ergodicity of Nonlinear Autoregressive Models', *Statistica Sinica*, 6: 943–956.

Anderson, G. J., and Mizon, G. E. (1983). 'Parameter Constancy Tests: Old and New', Discussion Paper 8325, Economics Department, University of Southampton.

Andrews, D. W. K. (1993). 'Tests for Parameter Instability and Structural Change with Unknown Change Point', *Econometrica*, 61: 821–856.

——— (2003). 'End-of-Sample Instability Tests', *Econometrica*, 71: 1661–1694.

Andrews, D. W. K., and Fair, R. C. (1988). 'Inference in Nonlinear Econometric Models with Structural Change', *The Review of Economic Studies*, 55: 615–639.

Bai, J. (1994). 'Least Squares Estimation of a Shift in Linear Processes', *Journal of Time Series Analysis*, 15: 453–472.

——— (1995). 'Least Absolute Deviation Estimation of a Shift', *Econometric Theory*, 11: 403–436.

——— (1997). 'Estimation of a Change Point in Multiple Regression Models', *Review of Economics and Statistics*, 79: 551–563.

——— (1999). 'Likelihood Ratio Tests for Multiple Structural Changes', *Journal of Econometrics*, 91: 299–323.

Bai, J., and Perron, P. (1998). 'Estimating and Testing Linear Models with Multiple Structural Changes', *Econometrica*, 66: 47–78.

———— (2003). 'Critical Values for Multiple Structural Change Tests', *The Econometrics Journal*, 6: 72–78.

———— (2006). 'Multiple Structural Changes: A Simulation Analysis', in D. Corbae, S. N. Durlauf, and B. Hansen (eds.), *Econometric Theory and Practice: Frontiers of Analysis and Applied Research*, pp. 212–238. Cambridge University Press, Cambridge, UK.

Bec, F., Salem, M., and Collard, F. (2002). 'Asymmetries in Monetary Policy Reaction Function: Evidence for the US, French and German Central Banks', *Studies in Nonlinear Dynamics and Econometrics*, 6, Issue 2, Art. 3.

Benati, L., and Surico, P. (2008). 'Evolving US Monetary Policy and the Decline of Inflation Predictability', *Journal of the European Economic Association*, 6: 634–646.

Boldea, O., and Hall, A. R. (2010). 'Estimation and Inference in Unstable Nonlinear Least Squares Models', *Centre for Growth and Business Cycle Research Discussion Paper Series*, 126.

Caner, M. (2007). 'Boundedly Pivotal Structural Change Tests in Continuous Updating GMM with Strong, Weak identification and Completely Unidentified Cases', *Journal of Econometrics*, 137: 28–67.

Carrasco, M., and Chen, X. (2002). 'Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models', *Econometric Theory*, 18: 17–39.

Chan, K. S., and Tong, H. (1986). 'On Estimating Thresholds in Autoregressive Models', *Journal of Time Series Analysis*, 7: 179–190.

Cogley, T., and Sargent, T. (2001). 'Evolving post-world war II US inflation dynamics', in *NBER Macroeconomics Annual*. MIT Press.

Davidson, J. (2002). 'Establishing Conditions for the Functional Central Limit Theorem in Nonlinear and Semiparametric Time Series Processes', *Journal of Econometrics*, 106: 179–190.

Dolado, J., María-Dolores, R., and Ruge-Murcia, M. (2004). 'Nonlinear Monetary Policy Rules: Some New Evidence from US', *Studies in Nonlinear Dynamics and Econometrics*, 8, Issue 3, Art. 2.

Eitrheim, Ø., and Teräsvirta, T. (1996). 'Testing the Adequacy of the Smooth Transition Autoregressive Models', *Journal of Econometrics*, 74: 59–75.

Franses, P. H., and van Dijk, D. (2000). *Nonlinear time series models in empirical finance*. Cambridge Univ Pr, Cambridge, UK.

Ghysels, E., and Hall, A. R. (1990). 'A Test for Structural Stability of Euler Conditions Parameters Estimated via the Generalized Method of Moments Estimator', *International Economic Review*, 31: 355–364.

Granger, C., and Teräsvirta, T. (1993). *Modelling Non-Linear Economic Relationships*. Oxford University Press, Oxford, UK.

Hall, A. R., Han, S., and Boldea, O. (2012). 'Inference Regarding Multiple Structural Changes in Linear Models with Endogenous Regressors', *Journal of Econometrics, forthcoming*.

Hall, A. R., and Sakkas, N. (2012). 'Approximate p-values of Certain Tests Involving Hypotheses about Multiple Breaks', *Journal of Econometric Methods, forthcoming*.

Hall, A. R., and Sen, A. (1999). 'Structural Stability Testing in Models Estimated by Generalized Method of Moments', *Journal of Business & Economic Statistics*, 17: 335–348.

Hansen, B. E. (2000). 'Testing for Structural Change in Conditional Models', *Journal of Econometrics*, 97: 93–115.

Kang, K., Kim, C., and Morley, J. (2009). 'Changes in US inflation persistence', *Studies in Nonlinear Dynamics & Econometrics*, 13, Issue 4, Art. 1.

Kesriyeli, M., Osborn, D., and Sensier, M. (2006). 'Nonlinearity and Structural Change in Interest Rate Reaction Functions for the US, UK and Germany', in C. Milas, D. van Dijk, and P. Rothman (eds.), *Nonlinear Time Series Analysis of Business Cycles*, pp. 283–310.

Kim, D., Osborn, D., and Sensier, M. (2005). 'Nonlinearity in the Fed's Monetary Policy Rule', *Journal of Applied Econometrics*, 20: 621–639.

McAleer, M., and Medeiros, M. (2008). 'A Multiple Regime Smooth Transition Heterogeneous Autoregressive Model for Long Memory and Asymmetries', *Journal of Econometrics*, 147: 104–119.

Meese, R., and Rose, A. (1991). 'An Empirical Assessment of Non-Linearities in Models of Exchange Rate Determination', *The Review of Economic Studies*, 58: 603–619.

Meyn, S., and Tweedie, R. (1993). *Markov chains and Stochastic stability*, Communications and Control Engineering Series. Springer-Verlag, London, UK.

Mokkadem, A. (1985). 'Le Modele Non Linéaire AR(1) Général. Ergodicité et Ergodicité Geometrique', *Comptes Rendus de l'Académie des Sciences. Série I. Mathématique*, 331: 889–892.

Pantula, S. (1988). 'Estimation of Autoregressive Models with ARCH Errors', *Sankhyā: The Indian Journal of Statistics*, 50: 119–138.

Perron, P., and Qu, Z. (2006). 'Estimating Restricted Structural Change Models', *Journal of Econometrics*, 134: 373–399.

Qu, Z., and Perron, P. (2007). 'Estimating and Testing Multiple Structural Changes in Multivariate Regressions', *Econometrica*, 75: 459–502.

Rosenblatt, M. (1971). *Markov processes: Structure and Asymptotic Behavior.* Springer, New York.

Rossi, B. (2006). 'Are Exchange Rates Really Random Walks? Some Evidence Robust to Parameter Instability', *Macroeconomic Dynamics*, 10: 20–38.

Schaling, E. (2004). 'The Nonlinear Phillips Curve and Inflation Forecast Targeting: Symmetric versus Asymmetric Monetary Policy Rules', *Journal of Money, Credit and Banking*, 36: 361–386.

Scheinkman, J. A., and LeBaron, B. (1989). 'Nonlinear Dynamics and Stock Returns', *The Journal of Business*, 62: pp. 311–337.

Sowell, F. (1996). 'Optimal Tests for Parameter Instability in the Generalized Method of Moments Framework', *Econometrica*, 64: 1085–1107.

Stock, J. H., and Watson, M. W. (2002). 'Has the Business Cycle Changed and Why?', *NBER Macroeconomics Annual*, pp. 159–218.

Teräsvirta, T. (1994). 'Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models', *Journal of the American Statistical Association*, 89: 208–218.

Zhang, C., Osborn, D., and Kim, D. (2008). 'The New Keynesian Phillips curve: From Sticky Inflation to Sticky Prices', *Journal of Money, Credit and Banking*, 40: 667–699.

Table 1: Relative rejection frequencies of F-statistics for pure structural change

DGP	α	$\sup F$		$seq\,FI$	$seq\,FII$	$UDmax\,F$
		0:1	0:2	1:2	1:2	
A I	.10	0	.04	.02	.03	.01
	.05	0	.02	.01	0	0
	.025	0	.01	0	0	0
	.01	0	0	0	0	0
A II	.10	1	1	.06	.07	1
	.05	1	1	.04	.04	1
	.025	1	1	.024	.021	1
	.01	1	1	.01	.01	1
A III	.10	1	1	.85	.83	1
	.05	1	1	.89	.88	1
	.025	1	1	.925	.915	1
	.01	1	1	.95	.94	1

Here, α is the nominal significance level, and the columns denote the empirical rejection frequency at level α.

Table 2: Empirical distribution of the estimated number of breaks at 5%

DGP	Strategy	$\hat{m}_T = 0$	$\hat{m}_T = 1$	$\hat{m}_T = 2$	Strategy	$\hat{m}_T = 0$	$\hat{m}_T = 1$	$\hat{m}_T = 2$
A I ($m=0$)	1	1	0	0	3	.99	.01	0
	2	1	0	0	4	.99	.01	0
A II ($m=1$)	1	0	.94	.06	3	0	.94	.06
	2	0	.93	.07	4	0	.93	.07
A III ($m=2$)	1	0	.11	.89	3	0	.11	.89
	2	0	.12	.88	4	0	.12	.88

The rows indicate the empirical frequency of a certain estimate $\hat{m}_T$ in 1000 simulations.

Table 3: Rejection Frequency of Tests for Pure and Partial Structural Change

DGP			sup F-test	Significance level			
	Break	In β_i or γ_i	0:1	.10	.05	.025	.01
B I	NO	-	β_i and γ_i	.05	.02	.017	.01
			β_i	.13	.07	.037	.02
			γ_i	.06	.02	.007	0
B II	YES	β_i	β_i and γ_i	1	1	1	1
			β_i	1	1	1	1
B III	YES	γ_i	β_i and γ_i	1	1	1	1
			γ_i	1	1	1	1
B IV	NO	-	β_i and γ_i	.05	.02	.017	.01
			β_i	.05	.02	.007	0
			γ_i	.05	.02	.014	.01

Table 4: Power of the tests for small, medium and large breaks

DGP	$sup F(0:1)$				$sup F_R(0:1)$			
	10%	5%	2.5%	1%	10%	5%	2.5%	1%
$c^*=.9$	1	1	1	1	1	1	1	1
$c^*=.5$	.89	.96	.97	.99	.94	.98	.99	1
$c^*=.3$	.33	.59	.63	.82	.49	.70	.73	.88

Here, the $sup F_R(0:1)$ test is for $H_0 : \gamma_1 = \gamma_2$.

Table 5: Stability Tests and P-Values

	$sup F(0:k)$			$UDmax F$	$seq F I$		$seq F II$	
	0:1	0:2	0:3	0:3	1:2	2:3	1:2	2:3
test	68.44	29.99	34.79	68.44	18.19	13.61	20.09	14.41
p-value	0.00	0.00	0.00	0.00	0.86	1.00	0.15	1.00

Table 6: Estimates for one break, with pure and partial structural change

	Pure structural change		Partial structural change	
	1984:5-1990:9	1990:10-2010:06	1984:5-1990:9	1990:10-2010:06
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
$[1 - G_t]$	1.723***	-0.092*	0.139	-0.037
$r_{t-1}[1 - G_t]$	1.088***	1.356**	1.000	0.9628
$r_{t-2}[1 - G_t]$	-0.313***	-0.363***	-0.050	-0.044
$\pi_{t-1}[1 - G_t]$	0.256**	-0.097**	-0.303	0.023
$y_{t-1}[1 - G_t]$	0.000	0.000	0.001	-0.001
	$\hat{\beta}_1^*$	$\hat{\beta}_2^*$	$\hat{\beta}_1^*$	$\hat{\beta}_2^*$
$[G_t]$	-0.362***	0.327***	1.018	0.066
$r_{t-1}[G_t]$	0.920	0.748**	1.362***	1.962***
$r_{t-2}[G_t]$	0.127**	0.223**	-0.338**	-0.969**
$\pi_{t-1}[G_t]$	0.183	-0.005	0.1383	-0.111
$y_{t-1}[G_t]$	0.002	0.001**	0.001	0.001
$\hat{\gamma}_i$	17.123	4.421**	1.268	1.268
$\hat{c}_i$	0.507***	0.208	-0.839	-0.839

Here, $G_t \equiv G(x_t, \hat{\gamma}_i, \hat{c}_i)$, the first column stands for the regressors to which the parameter estimates correspond, and *, ** and *** indicate significance of parameter estimates at 10%, 5% and 1% respectively.