working paper
department
of economics

ESTIMATING AND TESTING LINEAR MODELS WITH MULTIPLE STRUCTURAL CHANGES

Jushan Bai
Pierre Perron

95-17 Sept. 1995

massachusetts
institute of
technology

50 memorial drive
cambridge, mass. 02139

ESTIMATING AND TESTING LINEAR MODELS WITH MULTIPLE STRUCTURAL CHANGES

Jushan Bai
Pierre Perron

95-17 Sept. 1995

Estimating and Testing Linear Models with Multiple Structural Changes[1]

Jushan Bai
Massachusetts Institute of Technology

Pierre Perron
Université de Montréal

First Version: October 1994
This Version: September 1995

[1]This paper has benefited from comments by Maxwell King and Serena Ng as well as from seminar presentations at Harvard/MIT, Northwestern University, McGill University, the University of Copenhagen, the University of Illinois at Urbana-Champaign, the Université de Montréal, the University of Sao Paulo, PUC at Rio de Janeiro, Hitotsubashi University, the 1995 Winter Meeting of the Econometric Society and the 1995 joint Meeting of the Institute of Mathematical Statistics and the Canadian Statistical Association.

Jushan Bai: Department of Economics, E52-247B, Massachusetts Institute of Technology, Cambridge, MA, 02139 . E-mail: jbai@mit.edu. Partial financial support is acknowledged from the National Science Foundation under Grant SBR-9414083.

Pierre Perron: Département de Sciences Economiques, Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montréal, Canada, H3C-3J7. E-mail: perronp@plgcn.umontreal.ca. Support is acknowledged from the Social Sciences and Humanities Research Council of Canada, the Natural Sciences and Engineering Research Council of Canada and the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche du Québec.

Abstract

This paper considers issues related to multiple structural changes, occurring at unknown dates, in the linear regression model estimated by least squares. The main aspects are the properties of the estimators, including the estimates of the break dates, and the construction of tests that allow inference to be made about the presence of structural change and the number of breaks. We consider the general case of a partial structural change model where not all parameters are subject to shifts. We show convergence at rate T of the estimates of the break fractions. We also discuss a procedure that allows one to test the null hypothesis of, say, ℓ changes, versus the alternative hypothesis of $\ell + 1$ changes. This is particularly useful in that it allows a specific to general modeling strategy to consistently determine the appropriate number of changes present. An estimation strategy for which the location of the breaks need not be simultaneously determined is discussed. Instead, our method successively estimates each break point. Empirical applications are presented to illustrate the usefulness of the various procedures.

Keywords: Asymptotic Distribution, Change point, Rate of convergence, Model selection, Dynamic programming.
JEL Classification: C20, C22, C52.

1 Introduction.

This paper considers issues related to multiple structural changes in the linear regression model estimated by minimizing the sum of squared residuals. Throughout, we treat the dates of the breaks as unknown variables to be estimated. The main aspects considered are the properties of the estimators, including the estimates of the break dates, and the construction of tests that allow inference to be made about the presence of structural change and the number of breaks. To that effect we discuss tests of the null hypothesis of no structural change versus an arbitrary number of changes as well as tests of the null hypothesis of, say, ℓ versus $\ell+1$ changes.

Both the statistics and econometrics literature contains a vast amount of work on issues related to structural change, most of it specifically designed for the case of a single change[2]. The econometric literature has witnessed recently an upsurge of interest in extending procedures to various models with an unknown change point, thereby offering serious alternatives to the CUSUM test of Brown, Durbin and Evans (1975).

With respect to the problem of testing for structural change, recent contributions include the comprehensive treatment of Andrews (1993) who considers sup Wald, Likelihood Ratio and Lagrange Multiplier tests. Weighted versions of these tests satisfying some asymptotic optimality criterion are discussed in Andrews and Ploberger (1992). Recent studies also consider econometric models with trending regressors, unit root, cointegrated variables and serial correlation[3]. Methods allowing the investigator to be agnostic about the presence or absence of integrated variables are presented in Perron (1991) and Vogelsang (1993). The issue of structural change has also received a lot of attention in the recent debate on unit root versus structural change in the trend function of a univariate time series[4].Yet, all these recent developments consider only the case of a single structural change.

Issues about the distributional properties of the parameter estimates, in particular those of the break dates, have received somewhat less attention despite their importance. The work of Bai (1994,1995a) contains general results concerning the

[2]For a survey, see Krishnaiah and Miao (1988) and Zacks (1983) as well as the comprehensive treatment of Deshayes and Picard (1986).

[3]See, among others, Christiano (1992), Chu and White (1992), Kim and Sigmund (1989) and Perron (1991) (trending regressors), Kramer, Ploberger and Alt (1988) (serial correlation), Bai, Lumsdaine and Stock (1994) and Hansen (1990, 1992) (models with integrated variables).

[4]See Perron (1989, 1993, 1994), Banerjee, Lumsdaine and Stock (1992), Zivot and Andrews (1992), Perron and Vogelsang (1992) and Gregory and Hansen (1993).

asymptotic distribution of the estimated break date when a single break occurs, in particular the fact that the estimated break fraction converges to its true value at rate T.

In comparison, the literature addressing the issue of multiple structural changes is relatively sparse. Recent developments include Andrews, Lee and Ploberger (1992) who consider optimal tests in the linear model with known variance. Garcia and Perron (1994) study the sup Wald test for two changes in a dynamic time series[5]. To our knowledge, the most comprehensive treatment is that of Liu, Wu and Zidek (1994) who consider, as we do, multiple shifts in a linear model estimated by least squares (in the context of a more general multiple thresholds model). They study the rate of convergence of the estimated break dates, as well as the consistency of a modified Schwarz model selection criterion to determine the number of breaks. Their analysis considers only the so-called pure-structural change case where all the parameters are subject to shifts.

Our assumptions are less restrictive than those of Liu, Wu and Zidek (1994). Furthermore, we consider the more general case of a partial structural change where not all parameters are subject to shifts. Concerning the asymptotic behavior of the estimates of the break dates, we improve on the rate they report, $T/ln^2(T)$, by showing convergence at rate T. We also consider the asymptotic distribution and confidence intervals of the estimates of the break dates.

Our study considers, in addition, the important problem of testing for multiple structural changes. To that effect we present sup Wald type tests for the null hypothesis of no change versus an alternative hypothesis containing an arbitrary number of changes. We also discuss procedures that allow one to test the null hypothesis of, say, ℓ changes, versus the alternative hypothesis of $\ell+1$ changes. This is particularly useful in that it allows a specific to general modeling strategy to consistently determine the appropriate number of changes present (thereby avoiding the use of a model selection criterion which requires the estimation of the model for all possible number of breaks up to some a priori specified maximum). Finally our paper contains a discussion of an

[5]Some contributions include Fu and Curnow (1990) who discuss maximum likelihood estimation of multiple shifts in a somewhat restrictive binomial model. Yao (1988) considers estimating the number of breaks in the mean of a sequence of normal random variables based on the BIC criterion. Yao and Au (1988) treat the estimation of multiple mean breaks in a sequence of random variables and consider estimating the number of breaks using the BIC criterion. Yin (1988) uses the moving-window nonparametric technique to estimate the breaks in a sequence of random variables. Also, Feder (1975) considers estimating the joint points of polynomial type segmented regressions (non-discrete shifts). Other relevant contributions include Kim (1993) and Lumsdaine and Papell (1995).

estimation strategy for which the location of the breaks need not be simultaneously determined. Rather our method successively estimates each break point.

There are many practical advantages arising from the estimation and inference of models with structural changes. To mention a few, we first note that it allows the identification of events that may have fostered the structural changes. For example, an approach often used to examine the effectiveness of policy changes involves dummy variable regressions and inference on the corresponding regression coefficient. An alternative is to compare the estimated break date with the effective date of a policy change (or policy implementation). Another potentially useful aspect is in the field of forecasting. Indeed, if many regimes are present in a given sample, using the most recent regime may lead to better forecasts.

The rest of this paper is structured as follows. Section 2 discusses the model and the assumptions imposed on the variables and the innovations. Section 3 contains results pertaining to the consistency, the rate of convergence and the asymptotic distribution of the estimates of the break dates (as well as the others parameters of the model). Section 4 proposes test statistics, derives their asymptotic distributions and presents critical values. Section 5 discusses sequential methods used to estimate the model without treating all break points simultaneously. Section 6 presents empirical applications. Appendix A contains some mathematical derivations and Appendix B a description of a procedure to obtain estimates with multiple changes based on the principle of dynamic programing .

2 The Model and Assumptions.

Consider the following multiple linear regression with m breaks ($m+1$ regimes):

$$
\begin{array}{ll}
y_t = x_t'\beta + z_t'\delta_1 + u_t, & t = 1, 2, ..., T_1, \\
y_t = x_t'\beta + z_t'\delta_2 + u_t, & t = T_1 + 1, ..., T_2, \\
\vdots & \vdots \\
y_t = x_t'\beta + z_t'\delta_{m+1} + u_t, & t = T_m + 1, ..., T,
\end{array} \tag{1}
$$

where y_t is the observed dependent variable at time t; x_t ($p \times 1$) and z_t ($q \times 1$) are vectors of covariates and β and δ_j ($j = 1, ..., m+1$) are the corresponding vectors of coefficients; u_t is the disturbance at time t. The indices $(T_1, ..., T_m)$, or the break points, are explicitly treated as unknown. The purpose is to estimate the unknown regression coefficients together with the break points when T observations on (y_t, x_t, z_t) are available.

Note that this is a partial structural change model in the sense that the parameter vector β is not subject to shifts and is effectively estimated using the entire sample. When $p = 0$, we obtain a pure structural change model where all the coefficients are subject to change. A partial structural change model is therefore more general and includes the latter as a special case.

To proceed, it is convenient to introduce some terminologies. First, we call an m-partition (or simply a partition) of the integers $(1, ..., T)$, an m-tuple vector of integers $(T_1, ..., T_m)$ such that $1 < T_1 < \cdots T_m < T$. Note also that, throughout, we shall use the convention that $T_0 = 0$ and $T_{m+1} = T$. Second, define the block-diagonal matrix

$$\overline{Z} = \begin{pmatrix} Z_1 & & & \\ & Z_2 & & \\ & & \ddots & \\ & & & Z_{m+1} \end{pmatrix}$$

with $Z_i = (z_{T_{i-1}+1}, ..., z_{T_i})'$. The matrix $\overline{Z}$ is said to diagonally partition $Z = (z_1, ..., z_T)'$ at $(T_1, ..., T_m)$. Using these definitions, the multiple linear regression system (1) may be expressed in matrix form as

$$Y = X\beta + \overline{Z}\delta + U,$$

where $Y = (y_1, ..., y_T)', X = (x_1, ..., x_T)', U = (u_1, ..., u_T)', \delta = (\delta_1', \delta_2', ..., \delta_{m+1}')'$, and $\overline{Z}$ is the matrix which diagonally partitions Z at $(T_1, ..., T_m)$.

Throughout, we denote the true value of a parameter with a 0 superscript or subscript. In particular, $\delta^0 = (\delta_1^{0\prime}, ..., \delta_{m+1}^{0\prime})'$ and $(T_1^0, ..., T_m^0)$ are used to denote, respectively, the true values of the parameters δ and the true break points. The matrix $\overline{Z}_0$ is the one which diagonally partitions Z at $(T_1^0, ..., T_m^0)$. Hence, the data-generating process is assumed to be

$$Y = X\beta^0 + \overline{Z}_0\delta^0 + U.$$

The goal is to estimate the unknown coefficients $(\beta^0, \delta_1^0, ..., \delta_{m+1}^0, T_1^0, ..., T_m^0)$, assuming $\delta_i^0 \neq \delta_{i+1}^0$ $(1 \leq i \leq m)$. In general, the number of breaks m can be treated as an unknown variable with true value m^0. However, for now, we treat it as known and discuss methods of estimating it in later sections. We also postpone the problem of testing for the presence of structural change to Section 4.

The method of estimation considered is that based on the least-squares principle. For each m-partition $(T_1, ..., T_m)$, the associated least-squares estimates of β and δ_j

are obtained by minimizing the sum of squared residuals

$$(Y - X\beta - \overline{Z}\delta)'(Y - X\beta - \overline{Z}\delta) = \sum_{i=1}^{m+1} \sum_{t=T_{i-1}+1}^{T_i} [y_t - x_t'\beta - z_t'\delta_i]^2.$$

Let $\hat{\beta}(\{T_j\})$ and $\hat{\delta}(\{T_j\})$ denote the resulting estimates based on the given m-partition $(T_1, ..., T_m)$ denoted $\{T_j\}$. Substituting these estimates in the objective function and denoting the resulting sum of squared residuals as $S_T(T_1, ..., T_m)$, the estimated break points $(\hat{T}_1, ..., \hat{T}_m)$ are such that

$$(\hat{T}_1, ..., \hat{T}_m) = \text{argmin}_{T_1,...,T_m} S_T(T_1, ..., T_m). \tag{2}$$

where the minimization is taken over all partitions $(T_1, ..., T_m)$ such that $T_i - T_{i-1} \geq q$. Thus the break-point estimators are global minimizers of the objective function. Finally, the regression parameter estimates are obtained using the associated least-squares estimates at the estimated m-partition $\{\hat{T}_j\}$, i.e.

$$\hat{\beta} = \hat{\beta}(\{\hat{T}_j\}), \quad \hat{\delta} = \hat{\delta}(\{\hat{T}_j\}). \tag{3}$$

Since, the break points are discrete parameters and can only take a finite number of values, they can be estimated by grid search. In the case of a pure structural change model, an efficient procedure to obtain global minimizers can be obtained using a dynamic programming approach. This allows the estimates to be calculated using a number of sums of squared residuals (corresponding to the different possible partitions) that is of order $O(T^2)$ for any $m \geq 2$. The calculations needed can be significantly reduced further using standard updating formulae for recursive residuals. In effect, the procedure amounts to computing T sets of recursive residuals and performing pairwise comparisons of the associated sum of squared residuals. The method can easily be extended, in an iterative fashion, to cover the case of a partial structural change model. These issues are discussed in detail in Appendix B[6].

The statistical properties of the resulting estimators are studied in the next section under the following set of assumptions.

A1. Let $w_t = (x_t', z_t')'$, $W = (w_1, ..., w_T)'$ and $\overline{W}^0$ be the diagonal partition of W at the true break points $(T_1^0, ..., T_m^0)$ such that $\overline{W}^0 = diag(W_1^{0'}, ..., W_{m+1}^{0'})$. We assume for each $i = 1, ..., m+1$, that $W_i^{0'}W_i^0/(T_i^0 - T_{i-1}^0)$ converges to a non-random positive

[6]A GAUSS program that calculates global minimizers using this dynamic programing approach is available from the authors upon request. This program also has procedures to compute the other tests and statistics discussed in this paper.

definite matrix (with $T_0^0 = 1$ and $T_{m+1}^0 = T$). The limiting matrices need not be the same for all i.

A2. For large ℓ, the minimum eigenvalues of $\frac{1}{\ell}\sum_{T_i^0+1}^{T_i^0+\ell} w_t w_t'$ and of $\frac{1}{\ell}\sum_{T_i^0-\ell}^{T_i^0} w_t w_t'$ are bounded away from zero ($i = 1, ..., m+1$).

A3. The matrix $A_{k\ell} = \sum_k^\ell z_t z_t'$ is invertible for $\ell - k \geq q$, the dimension of z_t.

A4. The sequence of errors $\{u_t\}$ satisfies either of the following two sets of conditions:

i) Let $\{\mathcal{F}_i : i = \cdots, -2, -1, 0, 1, 2, \cdots\}$ be a sequence of increasing σ-fields. Assume that $\{u_i, \mathcal{F}_i\}$ forms a L^r-mixingale sequence with $r = 4 + \delta$ for some $\delta > 0$ [McLeish (1975) and Andrews (1988)]. That is, there exist nonnegative constants $\{c_i : i \geq 1\}$ and $\{\psi_m : m \geq 0\}$ such that $\psi_m \downarrow 0$ as $m \to \infty$ and for all $i \geq 0$ and $m \geq 0$, we have

(a) $\|E(u_i|\mathcal{F}_{i-m})\|_r \leq c_i\psi_m$,

(b) $\|u_i - E(u_i|\mathcal{F}_{i+m})\|_r \leq c_i\psi_{m+1}$,

where $\|X\|_r = (E|X|^r)^{1/r}$. We assume in addition that

(c) $\max_i c_i \leq K < \infty$,

(d) $\sum_{m=-\infty}^{\infty} \psi_m < \infty$,

and

(e) The disturbance u_t is independent of the regressors $\{z_s, x_s\}$ for all t and s.

ii) Let $\mathcal{F}_t^* = \sigma\text{-field}\{..., w_{t-1}, w_t, ..., u_{t-2}, u_{t-1}\}$.

a) We assume that $\{u_t\}$ is a martingale difference sequence relative to $\{\mathcal{F}_t^*\}$ satisfying $E(u_t|\mathcal{F}_{t-1}^*) = 0$, and $\sup_t E|u_t|^{4+\delta} < \infty$.

b) We have

$$p\lim T^{-1}\sum_{t=1}^{[Tv]} z_t z_t' = Q(v),$$

uniformly in $v \in [0, 1]$, where $Q(v)$ is positive definite for $v > 0$ and strictly increasing in v (i.e. $Q(v) - Q(u)$ is positive definite for $v > u$).

c) If the disturbances u_t are not independent of the regressors $\{z_s\}$ for all t and s, the minimization problem defined by (2) is taken over all possible partitions such that $T_i - T_{i-1} > \epsilon T$ ($i = 1, ..., m+1$) for some $\epsilon > 0$ (note that this is not required under part (i)).

A5. $T_i^0 = [T\lambda_i^0]$, where $0 < \lambda_1^0 < \cdots < \lambda_m^0 < 1$.

Assumption A1 is standard for multiple linear regressions. Assumption A2 requires that there be enough observations near the true break points so that they can be identified. Now consider A3. Because the break points are estimated by a global

least-squares search, we require the sum $\sum_k^\ell z_t z_t'$ to be invertible for $\ell - k \geq q$. In particular, no segment should contain fewer observations than q, as an exact fit is otherwise obtained. If we impose the number of observations in each segment to be at least some fixed number h ($h \geq q$, not depending on T), the invertibility requirement in A3 can be weakened to hold for all combinations (ℓ, k) for which $\ell - k \geq h$. Note that A3 is actually for technical convenience and could be dispensed with. This would require the use of generalized inverses. We assume, for simplicity, the existence of the inverse of $A_{k\ell}$, but the proof goes through with generalized inverses at the expense of a greater technical burden.

The assumptions stated in A4 pertain to two specific cases related to the presence or absence of a lagged dependent variable in z_t. The conditions described in part (i) pertain to the case where no lagged dependent variables are allowed in z_t. In this case, the conditions on the residuals are quite general and allow substantial correlation. A mixingale sequence includes many processes as special cases, such as martingale differences, strong mixing processes, linear processes, and functions of mixing processes, see Andrews (1988) for details. The sequence $\{u_t\}$ need not be stationary but the existence of a uniformly bounded moment of order $4 + \delta$ is required (this can be seen by noting that condition (c) is, in most cases, equivalent to $\sup_i \|u_i\|_r < \infty$). Condition (e) precludes the presence of lagged dependent variables in the regressors.

Part (ii) of Assumption A4 considers the case where lagged dependent variables are allowed as regressors. In this case, no serial correlation is permitted in the errors $\{u_t\}$. The requirement of $\{z_t u_t\}$ forming a martingale difference is to permit weak convergence of the partial sums $T^{-1/2} \sum_{t=[Tu]+1}^{[Tv]} z_t u_t$. This extra generality is obtained at the expense of some restrictions on the admissible partitions. If a lagged dependent varaible is present in the z_t, each segment considered must contain a positive fraction of the total sample. This is not constraining from a practical point of view since ϵ can be arbitrarily small. Note that no such restriction is necessary if a lagged dependent variable is present in the x_t's. In both cases, the assumptions are general enough to allow different distributions for both the regressors and the errors in each segment.

The possibility of lagged dependent variables is potentially quite useful if the parameters associated with the dynamics of the dependent variables are not subject to structural change. In this case, the investigator can take these dynamic effects into account either in a direct parametric fashion (e.g. introducing lagged dependent variables so as to have uncorrelated residuals) or using an indirect nonparametric approach (e.g. leaving the dynamics in the disturbances and applying a nonparametric

correction for proper asymptotic inference). This trade-off can be useful to distinguish gradual from sudden changes the same way a distinction is made between innovational and additive outliers.

Assumption A5 is a standard requirement to permit the development of an asymptotic theory and allows the break points to be asymptotically distinct. It essentially requires the asymptotic experiments to be carried under the assumption that each segments increase proportionately as the sample size increases. We refer to the quantities $(\lambda_1^0, ..., \lambda_m^0)$ as the break fractions and we let $\lambda_0^0 = 0$ and $\lambda_{m+1}^0 = 1$.

3 Consistency and Limiting Distributions.

In this section, we are interested in the consistency property of the estimated break fractions and especially the rate of convergence. The result about the rate of convergence will allow us to derive results about the asymptotic distribution of the estimates of the break dates as well as the estimated regression coefficients. We let $\hat{\lambda} = (\hat{\lambda}_1, ..., \hat{\lambda}_m)$ with corresponding true values $\lambda^0 = (\lambda_1^0, ..., \lambda_m^0)$. We shall first show that $\hat{\lambda}$ is consistent for λ^0 and later that the rate of convergence is T. As a matter of notation, we let "$\xrightarrow{p}$" denote convergence in probability, "$\xrightarrow{d}$" convergence in distribution, and "$\Rightarrow$" weak convergence in the space $D[0,1]$ under the uniform metric, see Pollard (1984, Chapter 5).

3.1 Consistency.

The main result of this section is summarized in the following proposition which states the consistency of $\hat{\lambda}$ for λ^0.

Proposition 1 *Under assumptions A1-A5, the estimated break fractions are consistent. That is, for each $\eta > 0$ and each $\epsilon > 0$, we have, when T is large:*

$$P(|\hat{\lambda}_k - \lambda_k^0| > \eta) < \epsilon, \quad (k = 1, ..., m).$$

We outline the main steps of the proof using a few lemmas that are proved in the appendix. Note that $\hat{T}_i$ need not equal T_i^0 so that estimated segments (or regimes) need not correspond to true regimes and the two partitions $\{\hat{T}_i\}$ and $\{T_i^0\}$ are allowed to overlap. By showing that $\hat{\lambda}_j \to \lambda_j^0$ we, in effect, bound the degree of overlap.

Denote by $\hat{u}_t$ the estimated residual for the t-th observation and by d_t the difference between the fitted regression "line" and the true regression line. That is,

$$\hat{u}_t = y_t - x_t\hat{\beta} - z_t\hat{\delta}_k, \qquad t \in [\hat{T}_{k-1} + 1,\ \hat{T}_k]$$

and

$$d_t = x_t'(\hat{\beta} - \beta^0) + z_t'(\hat{\delta}_k - \delta_j^0), \qquad t \in [\hat{T}_{k-1} + 1,\ \hat{T}_k] \cap [T_{j-1}^0 + 1,\ T_j^0]$$

for $k, j = 1, ..., m + 1$. Note that, in general, d_t is defined over $(m + 1)^2$ different segments corresponding to each of the possible m-partitions $\{\hat{T}_i\}$ and $\{T_i^0\}$. Using elementary properties of projections,

$$\frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2 \leq \frac{1}{T}\sum_{t=1}^{T}u_t^2, \tag{4}$$

and using $\hat{u}_t = u_t - d_t$,

$$\frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2 = \frac{1}{T}\sum_{t=1}^{T}u_t^2 + \frac{1}{T}\sum_{t=1}^{T}d_t^2 - 2\frac{1}{T}\sum_{t=1}^{T}u_t d_t. \tag{5}$$

The proof of Proposition 1 simply uses relations (4) and (5) and the associated limits of $T^{-1}\sum_{t=1}^{T} d_t^2$ and $T^{-1}\sum_{t=1}^{T} u_t d_t$. We start with the latter.

Lemma 1 *Under assumptions A1-A5, we have*

$$\frac{1}{T}\sum_{t=1}^{T}u_t d_t = o_p(1).$$

In the absence of structural changes ($\delta_j = 0$ for all j), $T^{-1}\sum u_t d_t = T^{-1}(\hat{\beta} - \beta)'\sum x_t u_t$, which is readily seen to be $O_p(T^{-1/2})$. In this case, Lemma 1 holds trivially. The proof of this lemma, presented in the appendix, is quite involved when breaks occurring at unknown dates exist. Lemma 1 allows us to state directly the following result applicable in the case of homogeneous disturbances.

Corollary 1 *If $E(u_t^2) = \sigma^2$ for all t, then under Assumptions A1-A5, we have as $T \to \infty$:*

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{1}^{T}\hat{u}_t^2 \xrightarrow{p} \sigma^2.$$

Proof: Inequality (4) implies that $\hat{\sigma}^2 \leq \sigma^2 + o_p(1)$ while (5) implies, by Lemma 1, $\hat{\sigma}^2 = \sigma^2 + \sum d_t^2/T + o_p(1) \geq \sigma^2 + o_p(1)$. □

Lemma 1 together with (4) and (5) implies that, under A1-A5

$$\frac{1}{T}\sum_{t=1}^{T} d_t^2 \xrightarrow{p} 0. \tag{6}$$

The proof of the consistency of $\hat{\lambda}$ for λ^0 follows by showing that (6) implies $\hat{\lambda} \xrightarrow{p} \lambda^0$. More specifically, we prove in the appendix that (6) cannot hold if $\hat{\lambda}_i \not\to_p \lambda_i^0$ for some i. This is stated in the following lemma.

Lemma 2 *Suppose that assumptions A1-A5 are satisfied and that some break date, say λ_j^0, cannot be consistently estimated, then*

$$\lim_{T\to\infty} \sup P\left(\frac{1}{T}\sum_{t=1}^{T} d_t^2 > C \parallel \delta_j^0 - \delta_{j+1}^0 \parallel^2\right) > \epsilon_0,$$

for some constant $C > 0$ and some $\epsilon_0 > 0$.

We are now in the position to prove Proposition 1. Using (5) and Lemmas 1 and 2, and under the supposition that some break date is not consistently estimated, we have the inequality

$$\frac{1}{T}\sum_{1}^{T} \hat{u}_t^2 \geq \frac{1}{T}\sum_{1}^{T} u_t^2 + C\|\delta_j^0 - \delta_{j+1}^0\|^2 + o_p(1)$$

which holds with probability no less than some $\epsilon_0 > 0$. This is in contradiction with the inequality (4), which holds with probability 1 for all T. Hence, all break dates are consistently estimated.

The consistency result for $\hat{\lambda}$ allows us to state consistency results for the parameter estimates $\hat{\beta}$ and $\hat{\delta}_k$ $(k = 1, ..., m)$. The proof of the following proposition, presented in the appendix, uses arguments similar to those involved in the proof of Lemma 2.

Proposition 2 *Under assumptions A1-A5, we have*

$$\begin{aligned} \hat{\beta} - \beta^0 &\xrightarrow{p} 0 \\ \hat{\delta}_k - \delta_k^0 &\xrightarrow{p} 0 \qquad (k = 1, ..., m). \end{aligned} \tag{7}$$

That is, the least squares estimators of the regression coefficients are consistent.

3.2 Rates of convergence.

We now consider the rate of convergence of the estimates. We start by showing that $\hat{\lambda}_k$ converge to its true value at rate T. More precisely, we have:

Proposition 3 *Under assumptions A1-A5, for every $\eta > 0$, there exists a $C < \infty$, such that for all large T,*

$$P(|T(\hat{\lambda}_k - \lambda_k^0)| > C) < \eta, \qquad (k = 1, ..., m).$$

The proof is presented in the appendix. It is important to remark that the rate T convergence pertains to the estimated break fractions $\hat{\lambda}_i$ and not to $\hat{T}_i$, the estimates of the break dates. For the latter, our result states that with high probability the bias is bounded by some constant C that is independent of the sample size T, i.e. with high probability, we have $|\hat{T}_i - T_i^0| < C$.

This result about the rate of convergence of the break fractions allows us to derive the rate of convergence and obtain the asymptotic distribution of the estimated coefficients $\hat{\beta}$ and $\hat{\delta}$. The relevant results are stated in the following proposition whose proof is similar to Corollary 1 of Bai (1994b) and is therefore omitted.

Proposition 4 *Under assumptions A1-A5, the estimates $\hat{\beta}$ and $\hat{\delta}$ in (3) are $\sqrt{T}$ consistent and asymptotically normal such that*

$$\begin{pmatrix} \sqrt{T}(\hat{\beta} - \beta^0) \\ \sqrt{T}(\hat{\delta} - \delta^0) \end{pmatrix} \xrightarrow{d} N(0, V^{-1}\Phi V^{-1}),$$

where

$$V = \text{plim}\, \frac{1}{T} \begin{pmatrix} X'X & X'\overline{Z}_0 \\ \overline{Z}_0'X & \overline{Z}_0'\overline{Z}_0 \end{pmatrix}, \tag{8}$$

$$\Phi = \text{plim} \frac{1}{T}(X, \overline{Z}_0)'\Omega(X, \overline{Z}_0), \tag{9}$$

and

$$\Omega = E(UU').$$

Note that when the errors are serially uncorrelated and homoskedastic we have $\Phi = \sigma^2 V$ and the asymptotic covariance matrix reduces to $\sigma^2 V^{-1}$, which can be consistently estimated using a consistent estimate of σ^2. When serial correlation and/or heteroskedasticity is present, a consistent estimate of Φ can be constructed along the lines of Newey and West (1987) and Andrews (1991). Note that the correction for possible serial correlation can be made assuming identical distributions across segments or allowing the distributions of both the regressors and the errors to differ.

3.3 Limiting Distributions of Break Dates

While the consistency of the estimates of the break fractions depends on the global behavior of the objective function, their limiting distribution depends on the local behavior around the true break dates. For this reason, the technical apparatus required to prove consistency is rather different in the multiple breaks case compared to the single break case. However, the analysis of the limiting distribution is similar in both cases. In the single break case, results concerning the limiting distribution of the break dates are discussed in Bai (1994b). We provide, in this section, the appropriate extensions to the case of multiple changes. Since the methods are similar, we only discuss the relevant results and refer the reader to Bai (1994b) for more details.

We discuss two types of limiting distributions for the estimates of the break points. The first corresponds to shifts of fixed magnitude and the second to shifts of shrinking magnitudes as the sample size increases.

3.3.1 Limiting Distributions with Fixed Shifts.

To study the limiting distribution in the case of fixed shifts, we need the following stronger assumptions.

A6. For regime i, $\{z_t; T_{i-1}^0 + 1 \leq t \leq T_i^0\}$ is a strictly stationary process ($i = 1, ..., m+1$).

We say that z_t follows process i when the time index t belongs to the ith regime. Note that for different regimes, $\{z_t, u_t\}$ may follow different stochastic processes and, hence, piecewise stationarity is allowed. Also, no stationarity assumption need be imposed on the x_t's. To characterize the limiting distribution, we first define a stochastic process $W^{(i)}(k)$ on the set of integers as follows: $W^{(i)}(0) = 0, W^{(i)}(k) = W_1^{(i)}(k)$ for $k < 0$, and $W^{(i)}(k) = W_2^{(i)}(k)$ for $k > 0$ where, for $i = 1, ..., m$:

$$(10) \qquad W_1^{(i)}(k) = -\Delta_i' \sum_{t=k+1}^{0} z_t^{(i)} z_t'^{(i)} \Delta_i + 2\Delta_i' \sum_{t=k+1}^{0} z_t^{(i)} u_t^{(i)}, \qquad k = -1, -2, ...$$

$$(11) \qquad W_2^{(i)}(k) = -\Delta_i' \sum_{t=1}^{k} z_t^{(i+1)} z_t'^{(i+1)} \Delta_i - 2\Delta_i' \sum_{t=1}^{k} z_t^{(i+1)} u_t^{(i+1)}, \qquad k = 1, 2, ...$$

with $\Delta_i = (\delta_{i+1}^0 - \delta_i^0)$ and where $\{z_t^{(i)}, u_t^{(i)}\}$ follows the ith stochastic process for all t. For example, when $(z_t^{(i)}, u_t^{(i)})$ is independent over time, the process $W^{(i)}$ is a two-sided random walk with (stochastic) drift.

Proposition 5 *Under assumptions A1-A6, and assuming that* $[\Delta_i' z_t]^2 \pm \Delta_i' z_t u_t$ *has a continuous distribution (for all t), then*

$$\hat{T}_i - T_i^0 \xrightarrow{d} \arg\max_k W^{(i)}(k) \qquad (i = 1, ..., m).$$

Furthermore, the estimated break points are asymptotically independent of each other and of the estimated regression parameters.

The assumption that $\{[\Delta_i' z_t]^2 \pm \Delta_i' z_t u_t\}$ has a continuous distribution ensures the uniqueness of the maximum of $W^{(i)}$ (with probability 1). Note that the limiting distributions of the break points estimates in the multiple break model are the same as in a single break model. This is basically due to the fact that these limiting distributions are determined by the local behavior of the objective function because of the fast rate of convergence of the estimated break points. Accordingly, one can view the limiting distribution of $\hat{T}_i - T_i^0$ as being solely determined by the segment $[T_{i-1}^0 + 1, T_{i+1}^0]$.

Because the estimates of the break fractions are consistent at rate T, they are essentially determined by a bounded finite number of observations with large probability no matter how large is the sample. This explains the asymptotic independence of the estimated break points since each contributing segments are increasingly apart as the sample size increases. The regression coefficients, on the other hand, are estimated using the entire data set or at least a positive fraction. One can then think of possibly deleting the finite number of observations that determine the distribution of the break points when estimating the regression coefficients without affecting their limiting distribution. Hence, the information determining the distribution of the break points and the remaining coefficients can be viewed as non-overlapping observations from which the asymptotic independence of the break-point estimates and the coefficients estimates follows.

The above result, though of definite theoretical interest, is perhaps of limited practical use because of the dependence of the limiting distribution of $\hat{T}_i - T_i^0$ on the distribution of $\{z_t^{(i)}, u_t^{(i)}\}$ for each segment i (though it is independent of x_t and β) . Hinkley (1971) provides an analytical expression for the probability density function in the case where $z_t = 1$ and u_t i.i.d. normal. In general, if one knows the distribution of $\{z_t^{(i)}, u_t^{(i)}\}$, the stochastic process $W^{(i)}$ and the location of its maximum can be easily simulated.

3.3.2 Limiting Distributions with Shrinking Shifts.

An alternative strategy is to consider an asymptotic framework where the magnitudes of the shifts converge to zero as the sample size increases. Even though the setup is particularly well suited to provide an adequate approximation to the exact distribution when the shifts are small, it remains adequate even for moderate shifts. This alternative asymptotic framework also allows a substantial relaxation of the assumption concerning the distribution of $\{z_t, u_t\}$. Moreover, the resulting limiting distribution is independent of the specific distribution of this pair.

We provide, in this section, a description of the results when the data are not trending. The required conditions are stated in the next assumption.

A7. a) Let $\Delta T_i^0 = T_i^0 - T_{i-1}^0$. The process $\{(z_t, u_t); T_{i-1}^0 + 1 \leq t \leq T_i^0\}$ is such that

$$p\lim(\Delta T_i^0)^{-1} \sum_{T_{i-1}^0+1}^{T_{i-1}^0+[s\Delta T_i^0]} Ez_t z_t' = sQ_i \text{ and } p\lim(\Delta T_i^0)^{-1} \sum_{T_{i-1}^0+1}^{T_{i-1}^0+[s\Delta T_i^0]} Eu_t^2 = s\sigma_i^2,$$

uniformly in $s \in [0,1]$ for $i = 1, ..., m+1$.

b) The following limit exists:

$$p\lim(\Delta T_i^0)^{-1} \sum_{r=T_{i-1}^0+1}^{T_{i-1}^0+[s\Delta T_i^0]} \sum_{t=T_{i-1}^0+1}^{T_{i-1}^0+[s\Delta T_i^0]} E(z_r z_t' u_r u_t) = s\Omega_i \quad (i = 1, ..., m+1).$$

uniformly in $s \in [0,1]$.

c) A functional central limit theorem holds for $\{z_t u_t\}$. That is, as $\Delta T_i^0 \to \infty$:

$$(\Delta T_i^0)^{-1/2} \sum_{t=T_{i-1}^0+1}^{T_{i-1}^0+[s\Delta T_i^0]} z_t u_t \Rightarrow B_i(s) \quad (i = 1, ..., m+1),$$

where $B(s)$ is a multivariate Gaussian process on $[0,1]$ with mean zero and covariance $EB_i(s)B_i(u) = \min\{s,u\}\Omega_i$.

The next assumption concerns the behavior of the shifts as T increases. Its role lies in the fact that if the shifts decrease as T increases more observations are needed around each break point to discern it. Hence, this make possible the application of a central limit theorem.

A8. Let $\Delta_{T,i} = \delta_{T,i+1}^0 - \delta_{T,i}^0$ $(i = 1, ..., m)$. Assume $\Delta_{T,i} = v_T \Delta_i$ for some Δ_i independent of T, where v_T is a scalar satisfying

$$v_T \to 0, \text{ and } T^{1/2-\alpha} v_T \to \infty, \text{ for some } \alpha \in (0, 1/2).$$

For $i = 1, ..., m$, let:

$$\begin{aligned} \xi_i &= \Delta_i' Q_{i+1} \Delta_i / \Delta_i' Q_i \Delta_i, \\ \phi_{i,1}^2 &= \Delta_i' \Omega_i \Delta_i / \Delta_i' Q_i \Delta_i, \\ \phi_{i,2}^2 &= \Delta_i' \Omega_{i+1} \Delta_i / \Delta_i' Q_{i+1} \Delta_i, \end{aligned}$$

and let $W_1^{(i)}(s)$ and $W_2^{(i)}(s)$ be independent standard Wiener processes defined on $[0, \infty)$, starting at the origin when $s = 0$. These processes are also independent across i. Define, for $i = 1, ..., m$:

$$Z^{(i)}(s) = \begin{cases} \phi_{i,1} W_1^{(i)}(-s) - |s|/2, & \text{if } s \leq 0, \\ \sqrt{\xi_i}\phi_{i,2} W_2^{(i)}(s) - \xi_i |s|/2, & \text{if } s > 0. \end{cases} \tag{12}$$

We are now in a position to state the following result.

Proposition 6 *Under assumptions A1-A5, A7-A8,*

$$(\Delta_i' Q_i \Delta_i) v_T^2 (\hat{T}_i - T_i^0) \xrightarrow{d} \arg\max_s Z^{(i)}(s) \quad (i = 1, ..., m) \tag{13}$$

The limiting distribution is the same as that occurring in a single break model. The density function of $\arg\max_s Z^{(i)}(s)$ is derived in Bai (1994b). When the limits Q_i, Ω_i and σ_i^2 are the same for adjacent i's, $\xi_i = 1$, and $\phi_{i,1} = \phi_{i,2} \equiv \phi$, in which case the limiting distribution (13) reduces to:

$$\begin{aligned} (\Delta_i' Q \Delta_i) v_T^2 (\hat{T}_i - T_i^0) &\xrightarrow{d} \arg\max_s \{\phi W^{(i)}(s) - |s|/2\} \\ &\stackrel{d}{=} \phi^2 \arg\max_s \{W^{(i)}(s) - |s|/2\}. \end{aligned}$$

or

$$\frac{(\Delta_i' Q \Delta_i)^2}{(\Delta_i' \Omega \Delta_i)} v_T^2 (\hat{T}_i - T_i^0) \xrightarrow{d} \arg\max_s \{W^{(i)}(s) - |s|/2\}. \tag{14}$$

Finally, when the errors are uncorrelated, we have $\Omega = \sigma^2 Q$ and the limiting result reduces to

$$\frac{(\Delta_i' Q \Delta_i)}{\sigma^2} v_T^2 (\hat{T}_i - T_i^0) \xrightarrow{d} \arg\max_s \{W^{(i)}(s) - |s|/2\}. \tag{15}$$

In this case, the limiting density function is symmetric about the origin. This case has been analyzed by Bhattacharya (1987), Picard (1985) and Yao (1987) for a single

break. The cumulative distribution function of $\arg\max_s\{W^{(i)}(s) - |s|/2\}$ is (see Yao (1987)):

$$H(x) = 1 + (2\pi)^{-1/2}\sqrt{x}e^{-x/8} - \frac{1}{2}(x+5)\Phi(-\sqrt{x}/2) + \frac{3}{2}e^{x}\Phi(-3\sqrt{x}/2),$$

for $x > 0$ and $H(x) = 1 - H(-x)$, where $\Phi(x)$ is the distribution function of a standard normal variable. For instance, the 95% and 97.5% quantiles are 7.7 and 11.0, respectively.

The results discussed above allows easy construction of confidence intervals for the break dates. All that is needed is to construct consistent estimates of the various parameters; $T^{-1}\sum_{t=1}^{T} z_t z_t'$ for Q, $T^{-1}\sum_{t=1}^{T}\hat{u}_t$ for σ^2. The parameters $v_T\Delta_i$ can be consistently estimated by the differences in the coefficient estimates $\hat{\delta}_i - \hat{\delta}_{i-1}$. When serial correlation is present Ω can be estimated using a kernel-based method as discussed in Andrews (1991). Note that when the segments are not homogeneous, obtaining consistent estimates of these quantities is still possible using data over the relevant subsamples only.

The limiting distribution in the case of trending regressors is discussed in Bai (1994b) for a single structural change model. His results remains valid for multiple breaks. We omit the details and refer the reader to that paper for more details.

4 Test Statistics for Multiple Breaks.

4.1 A Test of no break versus some fixed number of breaks.

We consider the sup F type test of no structural break ($m = 0$) versus the alternative hypothesis that there are $m = k$ breaks. Let $(T_1, ..., T_k)$ be a partition such that $T_i = [T\lambda_i]$ $(i = 1, ..., k)$. Again let $\overline{Z}$ denote the matrix which diagonally partitions Z at $(T_1, ..., T_k)$. Let R be the conventional matrix such that $(R\delta)' = (\delta_1' - \delta_2', ..., \delta_k' - \delta_{k+1}')$. Define

$$F_T(\lambda_1, ..., \lambda_k; q) = \left(\frac{T - (k+1)q - p}{kq}\right)\frac{\hat{\delta}'R'(R(\bar{Z}'M_X\bar{Z})^{-1}R')^{-1}R\hat{\delta}}{SSR_k}. \tag{16}$$

Here SSR_k is the sum of squared residuals under the alternative hypothesis, which depends on $(T_1, ..., T_k)$; q is the number of regressors whose coefficients are subject to change. The statistic F_T is simply the conventional F-statistic for testing $\delta_1 = \cdots = \delta_{k+1}$ against $\delta_i \neq \delta_{i+1}$ for some i with given $T_1, ..., T_k$. To carry the asymptotic analysis, we need to impose some restrictions on the possible values of the break dates.

In particular, we need to restrict each break date to be asymptotically distinct and bounded from the boundaries of the sample. To this effect, we define the following set for some arbitrary small positive number ϵ:

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

The sup F type test statistic is then defined as

$$\text{supF}_T(k; q) = \sup_{(\lambda_1,...,\lambda_k)\in\Lambda_\epsilon} F_T(\lambda_1, ..., \lambda_k; q).$$

This test is a generalization of the supF test considered by Andrews (1993) and others for the case k = 1.

The limiting distribution of the test depends on the nature of the regressors and the presence or absence of serial correlation and heterogeneity in the residuals. We consider the case where the following additional assumptions are imposed.

A9. Let $w_t = (x_t', z_t')'$ and Q be some positive definite matrix, we assume that

$$\text{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{[Ts]} w_t w_t' = sQ,$$

uniformly in $s \in [0, 1]$. Note that A9 precludes the presence of trending regressors. Extensions to the general case where $\text{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{[Ts]} w_t w_t' = Q(s)$, which allows trending regressors, are beyond the scope of the present paper. They will be discussed in a separate paper by the authors.

A10. The disturbances $\{u_t\}$ form an array of martingale differences relative to $\{\mathcal{F}_t\}$ where $\mathcal{F}_t = \sigma$-field $\{\cdots, w_{t-1}, w_t, \cdots, u_{t-2}, u_{t-1}\}$. Also, $E[u_t^2] = \sigma^2$ for all t and a functional central limit theorem holds for $\{w_t u_t\}$ such that

$$T^{-1/2} \sum_{t=1}^{[Tr]} w_t u_t \Rightarrow \sigma Q^{1/2} W^*(r),$$

where $W^*(r)$ is a $(p + q)$ vector of independent Wiener processes.

The case where $\{u_t\}$ satisfies the general conditions stated in Assumption A4 is discussed in Section 4.4 below. We show how the results remain valid provided appropriate modifications are made to account for the effect of serial correlation on the asymptotic distributions.

The following proposition, proved in the appendix, relates to the asymptotic distribution of the sup $F_T(k; q)$ test.

Proposition 7 *Under assumptions A9-A10:*

$$\text{supF}_T(k;q) \xrightarrow{d} \text{supF}_{k,q} \stackrel{def}{=} \sup_{(\lambda_1,...,\lambda_k)\in\Lambda_\epsilon} F(\lambda_1,...,\lambda_k;q),$$

where

$$F(\lambda_1,...,\lambda_k;q) = \frac{1}{kq}\sum_{i=1}^{k} \frac{[\lambda_i W_q(\lambda_{i+1}) - \lambda_{i+1}W_q(\lambda_i)]'[\lambda_i W_q(\lambda_{i+1}) - \lambda_{i+1}W_q(\lambda_i)]}{\lambda_i\lambda_{i+1}(\lambda_{i+1}-\lambda_i)},$$

$W_q(\cdot)$ *is a q-vector of independent Wiener processes on* $[0,1]$ *and* $\lambda_{k+1} = 1$.

Note that the asymptotic distribution of the test statistic depends on the value of ϵ in Δ_ϵ. As ϵ converges to zero, the test statistic diverges to infinity. Thus a small positive value instead of zero can improve the power significantly, see Andrews (1993) for further details. In what follows, we have adopted a value $\epsilon = 0.05$. No critical values for supF tests for $k \geq 2$ are available except those of Garcia and Perron (1994) who provide a partial tabulation for $k = 2$ and $q = 1$.

Asymptotic critical values are obtained via simulations, using an approach similar to that in Andrews (1993) and Garcia and Perron (1994). The Wiener process $W_q(\lambda)$ is approximated by the partial sums $n^{-1/2}\sum_{i=1}^{[n\lambda]} e_i$ with e_i i.i.d. $N(0, I_q)$ and $n = 1,000$. The number of replications is 10,000. For each replication, the supremum of $F(\lambda_1,...,\lambda_k;q)$ with respect to $(\lambda_1,...,\lambda_k)$ over the set Λ_ϵ is obtained via a dynamic programming algorithm (see the appendix for further details). We present, in Table 1, critical values covering cases with up to 9 breaks (i.e., up to 10 regimes, $k = 1,...,9$) and up to 10 regressors ($q = 1,...,10$) whose coefficients are the object of the test. The reported values are scaled up by q for comparison purposes. The column corresponding to one break ($k = 1$) can also be found in Andrews (1993).

4.2 A double maximum test.

The test discussed in the previous section requires the specification of the number of breaks, m, under the alternative hypothesis. It is of interest to consider a test of no structural break against an unknown number of breaks given some upper bound M on the maximum number of breaks permitted. A new test, called the *double maximum test,* can now be defined as

$$\text{DmaxF}_T(M,q) = \max_{1\leq m\leq M} \sup_{(\lambda_1,...,\lambda_m)\in\Lambda_\epsilon} F_T(\lambda_1,...,\lambda_m;q). \tag{17}$$

For a fixed m, $F(\lambda_1, ..., \lambda_m; q)$ is the sum of m dependent chi-square random variables with q degrees of freedom, each one divided by m. This scaling by m can be viewed, in some sense, as a prior imposed to account for the fact that as m increases a fixed sample of data becomes less informative about the hypotheses being confronted.

The last column of Table 1 reports the critical values of this test for $M = 5$ and $\epsilon = 0.05$. This should be sufficient for most empirical applications. In any event, the critical values vary little for choices of the upper bound M larger than 5.

4.3 Test of ℓ versus $\ell + 1$ breaks.

This section considers a test of the null hypothesis that ℓ unknown breaks exist against the alternative that an additional break exists. Ideally, one would base the test on the difference between the sum of squared residuals obtained allowing ℓ breaks and that obtained allowing $\ell + 1$ breaks. The limiting distribution of this test statistic is, however, difficult to use. Here we pursue a different strategy. For the model with ℓ breaks, the estimated break points denoted by $\hat{T}_1, ..., \hat{T}_\ell$, are obtained by a global minimization of the sum of squared residuals. Our test strategy proceeds conditional on the ℓ estimated break points under the null hypothesis by testing each $(\ell + 1)$ segments for the presence of an additional break.

The test amounts to the application of $(\ell+1)$ tests of the null hypothesis of no structural change versus the alternative hypothesis of a single change. The test is therefore applied to each segment containing the observations $\hat{T}_{i-1}$ to $\hat{T}_i$ $(i = 1, ..., \ell + 1)$ using again the convention that $\hat{T}_0 = 1$ and $\hat{T}_{\ell+1} = T$. We conclude for a rejection in favor of a model with $(\ell+1)$ breaks if the overall minimal value of the sum of squared residuals (over all segments where an additional break is included) is sufficiently smaller than the sum of squared residuals from the ℓ break model. The break date thus selected is the one associated with this overall minimum. More precisely, the test is defined by:

$$(18)\quad F_T(\ell + 1|\ell) = \{S_T(\hat{T}_1, ..., \hat{T}_\ell) - \min_{1 \le i \le \ell+1} \inf_{\tau \in \Lambda_{i,\eta}} S_T(\hat{T}_1, ..., \hat{T}_{i-1}, \tau, \hat{T}_i, ..., \hat{T}_\ell)\}/\hat{\sigma}^2,$$

where

$$(19)\qquad \Lambda_{i,\eta} = \{\tau; \hat{T}_{i-1} + (\hat{T}_i - \hat{T}_{i-1})\eta \le \tau \le \hat{T}_i - (\hat{T}_i - \hat{T}_{i-1})\eta\},$$

and $\hat{\sigma}^2$ is a consistent estimate of σ^2 under the null hypothesis. Note that for $i = 1$, $S_T(\hat{T}_1, ..., \hat{T}_{i-1}, \tau, \hat{T}_i, ..., \hat{T}_\ell)$ is understood as $S_T(\tau, \hat{T}_1, ..., \hat{T}_\ell)$ and for $i = \ell + 1$ as $S_T(\hat{T}_1, ..., \hat{T}_\ell, \tau)$.

In the case of a pure structural change, an alternative interpretation of the test can be given as follows. Let $D(i,j)$ be the minimized sum of squared residuals for the segment containing observations from $(i+1)$ to j, then the test statistic can be written as

$$(20) \qquad F_T(\ell+1|\ell) = \sup_{1\le i\le \ell+1} \sup_{\tau\in\Delta_{i,\eta}} \{D(\hat{T}_{i-1},\hat{T}_i) - D(\hat{T}_{i-1},\tau) - D(\tau,\hat{T}_i)\}/\hat{\sigma}^2.$$

This follows from $S_T(\hat{T}_1,...,\hat{T}_\ell) = D(0,\hat{T}_1) + D(\hat{T}_1,\hat{T}_2) + \cdots + D(\hat{T}_\ell,T)$ and a similar expression for $S_T(\hat{T}_1,...,\hat{T}_i,k,\hat{T}_{i+1},\cdots,\hat{T}_\ell)$ so that many common terms are canceled out. Under assumptions A9-A10, standard arguments show that,

$$(21)\ \sigma^{-2} \sup_{\tau\in\Delta^0_{i,\eta}} \{D(T^0_{i-1},T^0_i) - D(T^0_{i-1},\tau) - D(\tau,T^0_i)\} \Rightarrow \sup_{\eta\le\mu\le 1-\eta} \frac{\|W_q(\mu) - \mu W_q(1)\|^2}{\mu(1-\mu)},$$

where, as before, $W_q(\cdot)$ is a $q-$vector of independent Wiener processes on $[0,1]$ and $\Delta^0_{i,\eta}$ is as defined in (19) with $\hat{T}_i$ replaced by T^0_i. Under the null hypothesis, Proposition 3 asserts that $\hat{T}_i = T^0_i + O_p(1)$. Using this result, it is not difficult to show that the weak convergence in (21) also holds with T^0_{i-1} and T^0_i replaced by $\hat{T}_{i-1}$ and $\hat{T}_i$, respectively. In addition, because over different regimes $D(\cdot,\cdot)$ are computed using non-overlapping observations, the weak limits in (21) for different i's are independent. Thus the limit of (20) is the maximum of $\ell+1$ independent random variables in the form of (21), and we have the following result:

Proposition 8 *Under assumptions A9-A10:*

$$(22) \qquad \lim_{T\to\infty} P(F_T(\ell+1|\ell) \le x) = G_{q,\eta}(x)^{\ell+1},$$

where $G_{q,\eta}(x)$ is the distribution function of the random variable

$$\sup_{\eta\le\mu\le 1-\eta} \frac{\|W_q(\mu) - \mu W_q(1)\|^2}{\mu(1-\mu)}.$$

The critical values of this test for different values of ℓ can be obtained from the distribution function $G_{q,\eta}(x)$. A partial tabulation of some percentage points can be found in DeLong (1981) and Andrews (1993) (see also the first column of our Table 1). However, the grid presented is not fine enough to allow obtaining the relevant percentage points of $G_{q,\eta}(x)^{\ell+1}$. Accordingly, we provide a full set of critical values in Table 2 calculated with $\eta = .05$. These were obtained using a simulation method similar to that used for the construction of Table 1.

Note that $\hat{\sigma}^2$ is only required to be consistent under the null hypothesis for the validity of the stated asymptotic distribution. The test may, however, have better power if $\hat{\sigma}^2$ is also consistent under the alternative hypothesis. If the latter is true, $\hat{\sigma}^2$ constructed under the null hypothesis will overestimate σ^2. The test statistic is then biased downward, thereby decreasing its power. A consistent estimator under both the null and the alternative hypothesis is given by

$$\hat{\sigma}^2 = \frac{1}{T} S_T(\hat{T}_1, ..., \hat{T}_{\ell+1}),$$

where $\hat{T}_1, ..., \hat{T}_{\ell+1}$ are the estimates of the $\ell+1$ break points. Note, finally, that the results discussed in this section, including (22), hold true for partial structural changes. Also, it is important to note that the results carry through allowing different distributions across segments for the regressors and the errors. That is, Proposition 7 remains valid under A7(a and c) instead of A9-A10, provided $\hat{\sigma}^2$ is replaced by $\hat{\sigma}_i^2$ in (20).

4.4 Extensions to serially correlated errors.

The tests discussed above can be applied without the imposition of serially uncorrelated errors as specified in Assumption A10. In this case, some modifications are necessary to take account of the change in the limiting distribution of the statistics under the null hypothesis. A simple modification is to use the following version of the F-test instead of that specified in (16):

$$F_T^*(\lambda_1, ..., \lambda_k; q) = \frac{1}{T} \left(\frac{T - (k+1)q - p}{kq} \right) \hat{\delta}' R' (R \hat{V}(\hat{\delta}) R')^{-1} R \hat{\delta}, \tag{23}$$

where $\hat{V}(\hat{\delta})$ is an estimate of the variance covariance matrix of $\hat{\delta}$ that is robust to serial correlation and heteroskedasticity; i.e. a consistent estimate of

$$V(\hat{\delta}) = \text{plim} T (\overline{Z}' M_X \overline{Z})^{-1} \overline{Z}' M_X \Omega M_X \overline{Z} (\overline{Z}' M_X \overline{Z})^{-1}. \tag{24}$$

A consistent estimate of $V(\hat{\delta})$ can be obtained using methods such as those suggested by, e.g., Andrews (1991). Again, note that this estimate can be constructed allowing identical or different distributions for the regressors and the errors across segments. In some instances, the form of the statistic reduces in an interesting way. Consider the case of a pure structural change model ($\beta = 0$) where the explanatory variables are such that

$$\text{plim} \frac{1}{T} \overline{Z}' \Omega \overline{Z} = h_u(0) \text{plim} \frac{1}{T} \overline{Z}' \overline{Z} \tag{25}$$

with $h_u(0)$ the spectral density function of the errors u_t evaluated at the zero frequency. In that case,

$$V(\hat{\delta}) = h_u(0)\text{plim}(\frac{\overline{Z}'\overline{Z}}{T})^{-1},$$

and the robust version of the F-test can be constructed as:

$$F_T^*(\lambda_1, ..., \lambda_k; q) = \left(\frac{T-(k+1)q-p}{kq}\right) \hat{\delta}' R'(R(\overline{Z}'\overline{Z})^{-1}R')^{-1}R\hat{\delta}/\hat{h}_u(0).$$

with $\hat{h}_u(0)$ a consistent estimate of $h_u(0)$. In that case, we have the following asymptotically equivalent test

$$F_T^*(\lambda_1, ..., \lambda_k; q) = \frac{\hat{\sigma}^2}{\hat{h}_u(0)} F_T(\lambda_1, ..., \lambda_k; q),$$

with $\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T}\hat{u}_t^2$ a consistent estimate of the variance of the residuals. Hence, the robust version of the test is simply a scaled version of the original statistic. The condition (25) holds, for example, where testing for a change in mean as in Garcia and Perron (1994).

The computation of the robust version of the F-test (23) can be quite computationally involved when considering all combinations of possible break points, especially if a data dependent method is used to construct the robust asymptotic covariance matrix of $\hat{\delta}$. An asymptotically equivalent version is to first take the supremum of the original F-test to obtain the break points, i.e. imposing $\Omega = \sigma^2 I$. This can be done since the break fractions are T-consistent even in the presence of correlated errors. One can then obtain a robust version of the test by evaluating (23) and (24) at these estimated break dates.

The extensions for the $D\max F_T$ test and the sequential $F_T(\ell+1|\ell)$ are similar since they are simply functions of the $\sup F_T(k, q)$ tests.

4.5 Consistency of the tests.

A test is consistent if, under the alternative hypothesis, the associated test statistic diverges to infinity as the sample size increases. Because $\text{supF}_T(1; q) \leq 2\text{supF}_T(2; q) \leq k\,\text{supF}_T(k; q)$, the consistency of the $\text{supF}_T(k; q)$ $(k \geq 2)$ follows immediately from the results of Andrews (1993) who proved that the test based on the statistic $\text{supF}_T(1; q)$ is consistent for various alternatives including multiple breaks. Consequently, the test based on the statistic $\text{Dmax}F_T(k; q)$ is also consistent.

We next argue that the test based on $F_T(\ell+1|\ell)$ is also consistent. If there are more than ℓ breaks and a model with only ℓ breaks is estimated, there must be at least one break that is not estimated. Hence, at least one segment contains a nontrivial break point in the sense that both boundaries of each segment is separated from the true break point by a positive fraction of the total number of observations. For this segment, the $\sup F_T(1;q)$ test statistic converges to infinity as the sample size increases since it is consistent. Accordingly, the statistic $F_T(\ell+1|\ell)$ (computed for $\ell+1$ segments) also converges to infinity. This shows consistency.

5 Sequential Methods.

In this section we discuss issues related to the sequential estimation of the breaks points. We start, in section 5.1 with some results about the limit of break point estimates in underspecified models, i.e. when the regression structure allows for a smaller number of breaks than contained in the data-generating process. An interesting by-product of this analyses is the possibility of a sequential algorithm to estimate models with an unknown number of breaks. This is discussed in section 5.2.

5.1 The Limit of Break Point Estimates in Underspecified Models.

In this section, we show that the estimate of the break fraction in a single structural change regression applied to data that contain two breaks converge to one of the two true break fractions. In independent work, Chong (1994) obtains a similar result (see also Bai (1994c) for an earlier exposition).

To present our arguments, we consider a simple three-regime model:

$$
\begin{aligned}
y_t &= \mu_1 + \epsilon_t, \quad \text{if } t \leq [T\lambda_1] \\
y_t &= \mu_2 + \epsilon_t, \quad \text{if } [T\lambda_1]+1 \leq t \leq [T\lambda_2] \\
y_t &= \mu_3 + \epsilon_t, \quad \text{if } [T\lambda_2]+1 \leq t \leq T.
\end{aligned} \tag{26}
$$

with $\epsilon_t \sim i.i.d.(0,\sigma_\epsilon^2)$. Assume $\mu_1 \neq \mu_2$, $\mu_2 \neq \mu_3$, and $\lambda_1 < \lambda_2$, so there are two break points in the model. Let $\hat{T}_a$ denote the estimated single shift point. Our aim is to show that despite the misspecification of the number of regimes, $\hat{T}_a/T$ is consistent for either λ_1 or λ_2 depending on the relative magnitudes of the shifts and the spell of

each regime. To verify this claim, we examine the global behavior of $S(\tau)$, defined as the limit of $T^{-1}S_T([T\tau])$. We define $S_T(0)$ and $S_T(T)$ as the sum of squared residuals for the full sample without a break (i.e. $\sum_1^T (y_t - \bar{y})^2$, where $\bar{y}$ is the sample mean). In this way, $S_T([T\tau])$ is well defined for all $\tau \in [0,1]$. It is not difficult to show that the convergence of $T^{-1}S_T([T\tau])$ to $S(\tau)$ is uniform in $\tau \in [0,1]$. In particular

$$\frac{1}{T}S_T([T\lambda_1]) \xrightarrow{p} S(\lambda_1) = \sigma_\epsilon^2 + \frac{(1-\lambda_2)(\lambda_2-\lambda_1)}{1-\lambda_1}(\mu_2-\mu_3)^2 \tag{27}$$

and

$$\frac{1}{T}S_T([T\lambda_2]) \xrightarrow{p} S(\lambda_2) = \sigma_\epsilon^2 + \frac{\lambda_1}{\lambda_2}(\lambda_2-\lambda_1)(\mu_1-\mu_2)^2. \tag{28}$$

Without loss of generality we consider the case where $S(\lambda_1) < S(\lambda_2)$, our claim is stated in the following lemma

Lemma 3 *Suppose that the data are generated by (26) and that $S(\lambda_1) < S(\lambda_2)$, the estimated single break point $\hat{T}_a/T$ is consistent for λ_1.*

Proof: This can be proved by showing that $S(\tau)$ for $\tau \in [0,1]$ has a unique minimum at λ_1. The function $S(\tau)$ has different expressions over $[0,1]$. Some algebra reveals that

$$S(\tau) - S(\lambda_1) = \frac{\lambda_1 - \tau}{(1-\tau)(1-\lambda_1)}[(1-\lambda_1)(\mu_1-\mu_2) + (1-\lambda_2)(\mu_2-\mu_3)]^2, \quad \tau \le \lambda_1,$$

which is nonnegative. Under the assumption that $S(\lambda_1) < S(\lambda_2)$, the expression in the bracket above is nonzero, so $S(\tau) - S(\lambda_1)$ is strictly positive for $\tau < \lambda_1$. By symmetry (regarded as reversing the data order), $S(\tau) - S(\lambda_2)$ is nonnegative for $\tau > \lambda_2$. Thus for $\tau \in [\lambda_2, 1]$, $S(\tau) - S(\lambda_1) = S(\tau) - S(\lambda_2) + S(\lambda_2) - S(\lambda_1) \ge S(\lambda_2) - S(\lambda_1) > 0$. It remains to consider the case where $\tau \in (\lambda_1, \lambda_2)$. Again, simple algebra shows

$$\begin{aligned}
S(\tau) - S(\lambda_1) &= (\tau-\lambda_1)[\frac{\lambda_1}{\tau}(\mu_2-\mu_1)^2 - \frac{(1-\lambda_2)^2}{(1-\tau)(1-\lambda_1)}(\mu_3-\mu_2)^2] \\
&= (\tau-\lambda_1)\frac{\lambda_2}{\tau}[\frac{\lambda_1}{\lambda_2}(\mu_2-\mu_1)^2 - \frac{\tau(1-\lambda_2)(1-\lambda_2)}{\lambda_2(1-\tau)(1-\lambda_1)}(\mu_3-\mu_2)^2] \\
&\ge (\tau-\lambda_1)\frac{\lambda_2}{\tau}[\frac{\lambda_1}{\lambda_2}(\mu_2-\mu_1)^2 - \frac{(1-\lambda_2)}{(1-\lambda_1)}(\mu_3-\mu_2)^2] \\
&\ge (\tau-\lambda_1)[S(\lambda_2) - S(\lambda_1)]
\end{aligned}$$

where the first inequality follows from $\frac{\tau(1-\lambda_2)}{\lambda_2(1-\tau)} \le 1$ and the second inequality follows from $\frac{\lambda_2}{\tau} \le 1$. Thus $S(\tau) - S(\lambda_1)$ is strictly positive for $\tau \in (\lambda_1, \lambda_2)$. Thus we have

shown that $S(\tau)$ has a unique global minimum at λ_1 when $S(\lambda_1) < S(\lambda_2)$. Now because $S_T(\hat{T}_a) \leq S_T([T\lambda_1])$, it follows that $\hat{T}_a/T$ is consistent for λ_1.□

The assumption that $S(\lambda_1) < S(\lambda_2)$ implies that the first break point is more pronounced or dominating in terms of the relative magnitude of shifts and the regime spells. The above lemma implies that only when the dominating break is identified can the sum of squared residuals be reduced the most. Given that $\hat{T}_a/T$ is consistent for λ_1, one can use the subsample $[\hat{T}_a, T]$ to estimate another break point such that the sum of squared residuals is minimized for this subsample. The resulting estimate is then consistent for λ_2. This follows from the same type of argument as in the preceding paragraph because only λ_2 can be the dominating break in the sample $[\hat{T}_a, T]$, even if $\hat{T}_a < [T\lambda_1]$. Hence, if one knows that $\hat{T}_a/T$ is consistent for λ_1, a consistent estimator for λ_2 can also be obtained.

It is relatively straightforward to extend the argument to the case where a one-break model is fitted to a relationship that exhibits more than two breaks. The principle is the same and the estimate of the break fraction converges to one of the true break fraction, namely the one which allows a greatest reduction in the sum of squared residuals. It is also conjectured that a similar result holds when, say, an m_1 break model is fitted to a relationship that has m_2 breaks (with $m_2 > m_1$). Such a general result is not, however, needed for the arguments that follow.

5.2 Sequential estimation of the break points.

The arguments in section 5.1 showed that $\hat{T}_a/T$ is consistent for one of the true break point, the one that allows the greatest reduction in the sum of squared residuals. Suppose, as above, that this break point is λ_1, which, in general, may not be known (i.e., we do not know if the other break is before or after). In that case, we choose one break point either in the interval $[1, \hat{T}_a]$ or in the interval $[\hat{T}_a, T]$, such that the sum of squared residuals for all observations $[1, T]$ is minimized. With probability tending to 1 as the sample size increases, the estimated break point will be in the interval $[\hat{T}_a, T]$. This follows since, in the absence of a break in the interval $[1, \hat{T}_a]$, allowing one more break in that segment will not significantly reduce the sum of squared residuals. On the other hand, allowing one more break in the interval $[\hat{T}_a, T]$ will permit a large reduction given the presence of a break. Notationally, with probability tending to 1,

$$\min\{ \inf_{\tau} S_T(\tau, \hat{T}_a),\ \inf_{\tau} S_T(\hat{T}_a, \tau) \} = \inf_{\tau} S_T(\hat{T}_a, \tau) = S_T(\hat{T}_a, \hat{\tau})$$

where $\hat{\tau}$ is equivalently obtained by minimizing the sum of squared residuals for the subsample $[\hat{T}_a, T]$, and thus $\hat{\tau}/T$ is consistent for λ_2. The preceding argument implies that we can obtain consistent estimates of λ_1 and λ_2 in a sequential way.

Similarly, if $\hat{T}_a$ is actually consistent for λ_2 (this will be true if $S(\lambda_1) > S(\lambda_2)$), the second estimated shift point will be in $[1, \hat{T}_a]$. Generally, let $(\hat{N}_1, \hat{N}_2)$ be the ordered version of $(\hat{T}_a, \hat{\tau})$ such that $\hat{N}_1 < \hat{N}_2$. Then $(\hat{N}_1/T, \hat{N}_2/T)$ is consistent for (λ_1, λ_2).

5.2.1 Sequential estimation with a known number of break points.

The above analysis suggests a straightforward algorithm for estimating models with multiple break points whether the number of break points is known or unknown. Consider first the case of a known number of break points, say m. The idea is to estimate the breaks sequentially rather than simultaneously. Once the first break point is identified, the sample is split into two subsamples separated by this first estimated break point. For each subsample, a one break model is estimated and the second break point is chosen as that break point (of the two obtained) which allows the greatest reduction in the sum of squared residuals. The sample is then partitioned in three regimes and again a third break point is selected as one of the three estimates from three estimated one-break model that allow the greatest reduction in the sum of squared residuals. The process is continued until the m break points are selected.

The procedure is simple to implement using existing least squares routines with minor modifications. It yields consistent estimates of the break points; though the estimates are not guaranteed to be identical to those obtained by global minimization. Interestingly, it allows the estimation of models with any number of structural changes using a number of least-squares estimation that is only of order $O(T)$.

5.2.2 Sequential estimation with an unknown number of breaks.

Consider now the case of an unknown number of breaks which is likely to be of particular relevance in practice. A standard problem with any estimation procedure is that an improvement in the objective function is always possible by allowing more breaks. This naturally leads to consider a penalty factor for the increased dimension of a model. Yao (1988) suggests the use of the Bayesian Information Criterion (BIC) defined as

$$BIC(m) = \ln \hat{\sigma}^2(m) + p^* \ln(T)/T,$$

where $p^* = (m+1)q + m + p$, and $\hat{\sigma}^2(m) = T^{-1}S_T(\hat{T}_1, ..., \hat{T}_m)$. He showed that the number of breaks can be consistently estimated. Many other criteria such as the adjusted R^2, the prediction error criterion and Mallows C_p can be used as well. The AIC criterion is not recommended because, as is widely known, it has a tendency to overestimate the dimension of a model. An alternative proposed by Liu, Wu and Zidek (1994) is a modified Schwarz' criterion that takes the form:

$$MIC(m) = \ln(S_T(\hat{T}_1, ..., \hat{T}_m)/(T - p^*)) + (p^*/T)c_0(\ln(T))^{2+\delta_0}.$$

They suggest using $\delta_0 = 0.1$ and $c_0 = 0.299$.

We propose an alternative method to determine the number of breaks. The approach is directly related to the sequential procedure outlined above. Start by estimating a model with a small number of breaks that are thought to be necessary (or start with no break). Then perform parameter-constancy tests for each subsamples (those obtained by cutting off at the estimated breaks), adding a break to a subsample associated with a rejection with the test $F_T(\ell + 1|\ell)$. This process is repeated increasing ℓ sequentially until the test $F_T(\ell + 1|\ell)$ fails to reject the null hypothesis of no additional structural changes. The final number of breaks is thus equal to the number of rejections obtained with the parameter constancy tests plus the number of breaks used in the initial round.

It is important to note that the application of the test $F_T(\ell+1|\ell)$ in this sequential context is rather different from that discussed earlier. Indeed, the result of Proposition 8 is based on having the first ℓ breaks obtained simultaneously, i.e. as global minimizers of the sum of squared residuals assuming ℓ breaks. The reason for this is that the stated limiting distribution of the test requires convergence of the estimates of the break factions at rate T. Fortunately, this rate T convergence extends to the case where the break points are obtained sequentially one at a time. This last result is proved in a very recent study by Bai (1995b). Hence, the limiting distribution of the $F_T(\ell+1|\ell)$ test in the current sequential setup is the same as that stated in Proposition 8.

With probability approaching 1 as the sample size increases, the number of breaks determined this way will be no less than the true number. The procedure does not provide a consistent estimate of the true number of breaks, say m_0. This is because the sequential method is based on a test procedure which implies a non-zero probability of rejection under the null hypothesis given by the level of the test, say α. However, the asymptotic probability of selecting a model with a larger number of breaks, say m_0+j,

is given by α^j which decreases rapidly. Hence, there is no need (with large probability) to estimate models with more than the true number of breaks, as is necessary when using a model selection approach based on an information criterion.

The sequential procedure could be made consistent by adopting a significance level for the test $F_T(\ell+1|\ell)$ that decreases to zero, at a suitable rate, as the sample size increases. A result to that effect is presented in the next proposition whose proof is provided in the appendix.

Proposition 9 *Let $\hat{m}$ be the number of breaks obtained using the sequential method based on the statistic $F_T(\ell+1|\ell)$ applied with some size α_T, and let m_0 be the true number of breaks. If α_T converges to 0 slowly enough (for the test based on $F_T(\ell+1|\ell)$ to remain consistent), then, under assumptions A1-A5,*

$$P(\hat{m} = m_0) \to 1,$$

as $T \to \infty$. That is, the estimated number of breaks is consistent for the true number.

6 Empirical Applications.

In this section, we discuss two empirical applications of the procedures presented in this paper. The first analyzes the U.S. ex-post real interest rate series considered by Garcia and Perron (1994). The second reevaluates some findings of Alogoskoufis and Smith (1991) who analyze the issue of changes in the persistence of inflation and the corresponding shifts in an expectations-augmented Phillips curve resulting from such changes in persistence.

6.1 The U.S Ex-Post Real Interest Rate.

Garcia and Perron (1994) considered the time series properties of the U.S. Ex-Post real interest rate (constructed from the three-month treasury bill rate deflated by the CPI inflation rate taken from the Citibase data base). The data are quarterly and the sample is 1961:1-1986:3. Figure 1 presents a graph of the series. The issue of interest is the presence of structural changes in the mean of the series. To that effect we apply our procedure with only a constant as regressor (i.e. $z_t = \{1\}$) and take into account potential serial correlation via non-parametric adjustments. In the implementation of the procedure, we allowed up to 5 breaks and each segment was constrained to have at least 7 observations. The results are presented in Table 3.

The first issue to be considered is the determination of the number of breaks. Here the sup $F_T(k)$ tests are all significant for k between 1 and 5. So at least one break is present. The sup $F_T(2|1)$ test takes value 34.32 and is therefore highly significant. The sequential procedure (using a 1% significance level), BIC and the modified Schwarz criterion of Liu, Wu and Zidek (1994) all select two breaks. Hence, we conclude in favor of the presence of two breaks. Of direct interest are the estimates obtained under global minimization. The break dates are estimated at 1972:3 and 1980:3. The first date has a rather large confidence interval (between 1971:2 and 1973:4 at the 95% level). The second break date is, however, precisely estimated since the 95% confidence interval covers only one quarter before and after. The differences in the estimated means over each segment are significant and point to a decrease of 3.16% in late 1972 and a large increase of 7.44% in late 1980. These results confirm those of Garcia and Perron (1994).

6.2 Changes in the Persistence of Inflation and the Phillips Curve.

Alogoskoufis and Smith (1991) consider the following version of an expectations-augmented Phillips curve:

$$\Delta w_t = \alpha_1 + \alpha_2 E(\Delta p_t | I_{t-1}) + \alpha_3 \Delta u_t + \alpha_4 u_{t-1} + \xi_t,$$

where w_t is the log of nominal wages, p_t is the log of the Consumer Price Index, and u_t is the unemployment rate. They posit that inflation is an $AR(1)$ so that

$$E(\Delta p_t | I_{t-1}) = \delta_1 + \delta_2 \Delta p_{t-1}. \tag{29}$$

Hence, upon substitution, the Phillips curve is:

$$\Delta w_t = \gamma_1 + \gamma_2 \Delta p_{t-1} + \gamma_3 \Delta u_t + \gamma_4 u_{t-1} + \xi_t, \tag{30}$$

where $\gamma_2 = \alpha_2 \delta_2$. Here, a parameter of importance is δ_2 which is interpreted as measuring the persistence of inflation. Using post-war annual data from the United Kingdom and the United States, Alogoskoufis and Smith (1991) argue that the process describing inflation exhibits a one-time structural change from 1967 to 1968, whereby the autoregressive parameter δ_2 is significantly higher in the second period. This is interpreted as evidence that the abandonment of the Bretton Woods system relaxed the discipline imposed by the gold standard and created higher persistence in inflation.

They also argue that the parameter γ_2 in the Phillips curve equation (30) exhibit a similar increase at the same time, thereby lending support to the empirical significance of the Lucas critique.

Using the methods presented in this paper, we reevaluate Alogoskoufis and Smith (1991) claims using post-war annual data for the United Kingdom[7]. Consider first the structural stability of the $AR(1)$ representation of inflation whose series is depicted in Figure 2. Details of the estimation results are contained in Table 4. When applying a one break model, we indeed find the same results, namely a structural change in 1967 with δ_2 increasing from .274 to .739 while δ_1 remains constant. The estimate of the break is, however, imprecisely estimated with a 95% confidence interval covering the period 1961 – 1973. More importantly, the sup $F_T(1)$ test is not significant at any conventional level indicating that the data does not support a one break model. The sup $F_T(2)$ test is, however, significant at the 5% level and the sup $F_T(2|1)$ test is significant at the 10% level. The sup $F_T(\ell+1|\ell)$ test is not significant for any $\ell \geq 2$. BIC selects two breaks and MIC selects one suggesting that the latter may impose too strong a penalty when the sample size is small. Overall, the tests support a two break model.

The estimates of a two breaks model reveal a rather different interpretation of the data. The first break date is not linked to the end of the Bretton Woods system but rather with the first oil price shock in 1973. The second break is located in 1980. The coefficient estimates point to the importance of shifts in the level of inflation rather than changes in persistence. Indeed, the coefficient δ_1 varies from .021 to .130 in the period 1973 – 1980, and back to .011 after 1980. If anything, the data suggests a significant decrease in the persistence of inflation in the period 1973 – 1980, while the estimates of the autoregressive parameters are not significantly different in the first and last segments.

Since, there indeed appears to be structural changes in the inflation process, it is of interest to see if the Phillips curve equation underwent similar changes in accordance with the Lucas critique. The results are presented in Table 5. Here the evidence point to a single structural change. The sup $F_T(k)$ tests are significant for all k while the sup $F_T(\ell+1|\ell)$ tests are not significant for any $\ell \geq 1$. Furthermore, both the sequential procedure and the criterion of Liu, Wu and Zidek (1994) select a one break model (only BIC chooses two breaks). Again, the estimate of the break is associated with the first

[7]The data are the same as in Alogoskoufis and Smith (1991) and were kindly provided by George Alogoskoufis. We refer the reader to their paper for details on the definition and source of each series.

oil price shock (1973) and not with the end of the Bretton Woods system (the 95% confidence interval is small). The estimate of γ_2 indeed shows a marked decrease similar to the decrease in the persistence of inflation (the parameter estimate even becomes negative but not significantly so). Given the changes in the estimates of the parameter γ_3 and γ_4, the data suggest that the Phillips curve itself underwent a structural change in 1973. The data, however, do not support any adjustment of the Phillips curve following the change in the inflation process in 1980.

Since BIC selects two breaks, we also present results for this specification. These lend even less empirical support to the Lucas critique since the breaks dates are 1966 and 1975 and do not correspond to those of the inflation process.

7 Conclusions.

Our analysis has presented a rather comprehensive treatment of issues related to the estimation of linear models with multiple structural changes, to tests for the presence of multiple structural changes and to the determination of the number of changes present. Our results being asymptotic in nature, there is certainly a need to evaluate the quality of the approximations and the power of the tests in finite samples via simulations. We intend to present such a simulation study in a subsequent paper. Among the topics to be investigated, an important one appears to be the relative merits of different methods to select the number of structural changes. There are, of course, many other issues on the agenda. For instance, extensions of the test procedures to include tests that are optimal with respect to some criteria, extensions to nonlinear models and the derivation of tests that are valid in the presence of trending regressors.

A Mathematical Appendix.

As a matter of notation, we let $o_p(1)$ and $O_p(1)$ denote, respectively, a sequence of random variables converging to zero in probability and one that is stochastically bounded. Unless indicated otherwise, all convergence are taken as T, the sample size, increases to infinity. For a sequence of matrices B_T, we write $B_T = o_p(1)$ if each of its elements is $o_p(1)$ and likewise for $O_p(1)$. For a matrix A, M_A denotes the orthogonal projection matrix, $I - A(A'A)^{-1}A'$. We use $\|\cdot\|$ to denote the Euclidean norm, i.e. $\|x\| = (\sum_1^p x_i^2)^{1/2}$ for $x \in R^p$. For a matrix A, we use the vector-induced norm, i.e. $\|A\| = \sup_{x\neq 0} \|Ax\|/\|x\|$. We note that the norm of A is equal to the square root of the maximum eigenvalue of $A'A$, and thus $\|A\| \leq [tr(A'A)]^{1/2}$. Also, for a projection matrix P, $\|PA\| \leq \|A\|$. Finally, $[a]$ represents the integer part of a.

Proof of Lemma 1: We start with a series of lemmas that will be used subsequently. Assumption A5 is assumed throughout.

Lemma A.1 *Let S and V be two matrices having the same number of rows. Then the matrix $S'M_VS$ is non decreasing as more observations (rows) are added to the matrix (S, V).*

Proof: Write $S = (S_1', S_2')'$ and $V = (V_1', V_2')'$. We need to show that for an arbitrary vector α (having the same dimension as the number of rows of S and V)

$$\alpha' S' M_V S \alpha \geq \alpha' S_1' M_{V_1} S_1 \alpha.$$

Note that $\alpha' S' M_V S \alpha$ is the sum of squares of the residuals from a projection of $S\alpha$ on the space spanned by V. Similarly, $\alpha' S_1' M_{V_1} S_1 \alpha$ is the sum of squared residuals from a projection of αS_1 on V_1. The inequality is verified using the fact that the sum of squared residuals is non-decreasing as the number of observations increases (here the number of rows of S_1 and S). See, e.g., Brown, Durbin and Evans (1975).□

Lemma A.2 *Under assumption A1,*

$$\sup_{T_1,...,T_m} \left(\frac{X' M_{\overline{Z}} X}{T} \right)^{-1} = O_p(1).$$

where the supremum with respect to $(T_1, ..., T_m)$ is taken over all possible partitions such that $|T_{i-1} - T_i| \geq q$ $(i = 1, ..., m+1)$; the matrix $\overline{Z}$ diagonally partitions Z at $(T_1, ..., T_m)$.

Proof: We have the identity $X'M_{\overline{Z}}X = X_1'M_{Z_1}X_1 + \cdots + X_{m+1}'M_{Z_{m+1}}X_{m+1}$. An m-break model has $m+1$ regimes. Each partition $(T_1, ..., T_m)$ leaves at least one true regime uncut. In other words, there exists an i such that (X_i, Z_i) contains (X_i^0, Z_i^0) as a sub matrix. By Lemma A.1, $X_i'M_{Z_i}X_i \geq X_i^{0\prime}M_{Z_i^0}X_i^0$. Hence $(X'M_{\overline{Z}}X/T)^{-1} \leq (X_i^{0\prime}M_{Z_i^0}X_i^0/T)^{-1}$. This implies $\|(X'M_{\overline{Z}}X/T)^{-1}\| \leq \max_i \|(X_i^{0\prime}M_{Z_i^0}X_i^0/T)^{-1}\|$ for all partitions. The lemma now follows from assumption A1. □.

It should be pointed out that $(\overline{Z}'M_X\overline{Z}/T)^{-1}$ is not uniformly stochastically bounded over all possible partitions. In fact, it is easy to see that this matrix is $O_P(T)$ for some partitions.

Lemma A.3 *Under assumption A1,*

$$\sup_{T_1,...,T_m} X'M_{\overline{Z}}\overline{Z}_0 = O_p(T).$$

Proof: Because $M_{\overline{Z}}$ is a projection matrix, we have $\|X'M_{\overline{Z}}\overline{Z}_0\| \leq \|X\|\|M_{\overline{Z}}\overline{Z}_0\| \leq \|X\|\|\overline{Z}_0\|$ uniformly over all partitions. The lemma follows from $\|X\| \leq \sqrt{tr(X'X)} = O_p(\sqrt{T})$ and similarly $\|\overline{Z}_0\| = O_p(\sqrt{T})$. □

Lemma A.4 *The following identity holds:*

$$(\overline{Z}'M_X\overline{Z})^{-1} = (\overline{Z}'\overline{Z})^{-1} + (\overline{Z}'\overline{Z})^{-1}(\overline{Z}'X)(X'M_{\overline{Z}}X)^{-1}X'\overline{Z}(\overline{Z}'\overline{Z})^{-1}.$$

Proof: Follows from direct verification. □.

Lemma A.5 *Under assumption A4, there exists $\alpha < 1/2$ such that*

$$\sup_{T_1,...,T_m} \|\overline{Z}(\overline{Z}'\overline{Z})^{-1}\overline{Z}'U\| = O_p(T^\alpha),$$

where the supremum with respect to $(T_1, ..., T_m)$ is taken over all possible partitions such that $|T_{i-1} - T_i| \geq q$ $(i = 1, ..., m+1)$ under assumption A4(i) and over partitions such that $|T_{i-1} - T_i| \geq \epsilon T$ for some $\epsilon > 0$ under assumption A4(ii).

Proof: Consider first the case where part (i) of assumption A4 is assumed to hold. Because of the independence assumption between z_s and u_t, we can treat the z_t's as nonstochastic, otherwise conditional arguments can be used. Let $P_{\overline{Z}} = \overline{Z}(\overline{Z}'\overline{Z})^{-1}\overline{Z}'$. We shall prove that $|U'P_{\overline{Z}}U| = O_p(T^{2\alpha})$ uniformly in $T_1, ..., T_m$. However, $U'P_{\overline{Z}}U$ is the summation of the $m+1$ terms

$$(\sum_{T_i+1}^{T_{i+1}} z_t u_t)'(\sum_{T_i+1}^{T_{i+1}} z_t z_t')^{-1}(\sum_{T_i+1}^{T_{i+1}} z_t u_t), \quad (i = 0, ..., m)$$

Thus it suffices to prove that

$$\sup_{1\le k<\ell\le T} \|\sum_{t=k}^{\ell}\xi_t\| = O_p(T^\alpha) \tag{31}$$

with $\ell - k \ge q$ and where ξ_t is a $q \times 1$ vector defined by $\xi_t = \xi_t(k,\ell) = (A_{k\ell})^{-1/2} z_t u_t$ with $A_{k\ell} = \sum_{j=k}^{\ell} z_j z_j'$. For notational simplicity, the dependence of ξ_t on k and ℓ will not be exhibited. Now

$$\begin{aligned} P\left(\sup_{1\le k<\ell\le T} \|\sum_{t=k}^{\ell}\xi_t\| > T^\alpha\right) &\le \sum_{k=1}^{T}\sum_{\ell=k+q}^{T} P\left(\|\sum_{t=k}^{\ell}\xi_t\| > T^\alpha\right) \\ &\le T^{-2\alpha s}\sum_{k=1}^{T}\sum_{\ell=k+q}^{T} E\left\|\sum_{t=k}^{\ell}\xi_t\right\|^{2s} \end{aligned} \tag{32}$$

By the mixingale property, we can write u_t as

$$u_t = \sum_{j=-\infty}^{\infty} u_{jt}, \quad \text{with } u_{jt} = E(u_t|\mathcal{F}_{t-j}) - E(u_t|\mathcal{F}_{t-j-1})$$

and for each j, $\{u_{jt}, \mathcal{F}_{t-j}\}$ is a sequence of martingale differences. Using this decomposition, we have

$$\sum_{t=k}^{\ell}\xi_t = \sum_{j=-\infty}^{\infty}\sum_{t=k}^{\ell}\xi_{jt}.$$

where $\xi_{jt} = (A_{k\ell})^{-1/2} z_t u_{jt}$. By Minkowski's inequality,

$$E\left\|\sum_{t=k}^{\ell}\xi_t\right\|^{2s} \le \left(\sum_{j=-\infty}^{\infty}\left[E\left\|\sum_{t=k}^{\ell}\xi_{jt}\right\|^{2s}\right]^{1/2s}\right)^{2s}. \tag{33}$$

A key point is that for fixed j, k, and ℓ, $\{\xi_{jt}, \mathcal{F}_{t-j}\}$ $(t = k, ..., \ell)$ form a sequence of martingale differences. Thus by Burkholder's inequality (Hall and Heyde 1981, p.23) there exists a $C > 0$, only depending on q and s, such that

$$E\left\|\sum_{t=k}^{\ell}\xi_{jt}\right\|^{2s} \le CE\left(\sum_{t=k}^{\ell}\|\xi_{jt}\|^2\right)^s \le C\left(\sum_{t=k}^{\ell}(E\|\xi_{jt}\|^{2s})^{1/s}\right)^s. \tag{34}$$

where the second step follows by Minkowski's inequality. Now $\|\xi_{jt}\|^2 = z_t'(A_{k\ell})^{-1} z_t u_{jt}^2$. Thus $(E\|\xi_{jt}\|^{2s})^{1/s} = z_t'(A_{k\ell})^{-1} z_t (E|u_{jt}|^{2s})^{1/s}$. By A4(a), for $r = 2s$, we can show (see Hansen 1991)

$$(E|u_{jt}|^{2s})^{1/2s} \le 2c_t\psi_j \le 2(\max_i c_i)\psi_j \le K\psi_j \quad \text{for all } j.$$

It follows that $(E\|\xi_{jt}\|^{2s})^{1/s} \leq z_t'(A_{k\ell})^{-1}z_t K^2\psi_j^2$. Thus from (34),

$$E\left\|\sum_{t=k}^{\ell}\xi_{jt}\right\|^{2s} \leq C\left(\sum_{t=k}^{\ell} z_t'(A_{k\ell})^{-1}z_t K^2\psi_j^2\right)^s = C(K\psi_j)^{2s}q^s \tag{35}$$

where we have used the fact that $\sum_{t=k}^{\ell} z_t'(A_{k\ell})^{-1}z_t = trace((A_{k\ell})^{-1}\sum_k^{\ell} z_t z_t') = trace(I) = q$. Combining (35) and (33), we have

$$E\left\|\sum_{t=k}^{\ell}\xi_t\right\|^{2s} \leq Cq^s K^{2s}\left(\sum_{j=-\infty}^{\infty}\psi_j\right)^{2s} < \infty.$$

Note that the right hand side above does not depend on k and ℓ. This implies, in view of (32), that with $\ell - k \geq q$,

$$P\left(\sup_{1\leq k<\ell\leq T}\|\sum_{t=k}^{\ell}\xi_t\| > T^{\alpha}\right) \leq C_1 T^{-2\alpha s+2}$$

for some $C_1 > 0$. Let $s = 2+\delta/2$ (the moment of order $4+\delta$ of u_t exists by assumption A4), we can choose $\alpha \in (0,1/2)$ such that $T^{-2\alpha s+2} \to 0$. This proves (31) and hence the lemma.

Consider now the case where part (ii) of assumption A4 is assumed to hold. In this case

$$T^{-1}\sum_{t=[Tu]+1}^{[Tv]} z_t z_t' \to Q(v) - Q(u),$$

and hence $(T^{-1}\sum_{t=[Tu]+1}^{[Tv]} z_t z_t')^{-1} \to (Q(v)-Q(u))^{-1}$ uniformly in v and u such that $v - u > \epsilon > 0$. Also,

$$T^{-1/2}\sum_{t=[Tu]+1}^{[Tv]} z_t u_t = O_p(1)$$

uniformly using a functional central limit theorem for martingales differences. Accordingly, $|U'P_{\overline{Z}}U| = O_p(1)$ uniformly in $T_1,...,T_m$ and the statement of the lemma holds with $\alpha = 0$. □

Lemma A.6 *Under assumptions A1-A4, we have for some $\alpha < 1/2$,*

$$\sup_{T_1,...,T_m} X'\overline{Z}(\overline{Z}'\overline{Z})^{-1}\overline{Z}'U = O_p(T^{\alpha+1/2}). \tag{36}$$

Proof: This follows from Lemma A.5, $\|X\| = O_p(T^{1/2})$ and $\|X'\overline{Z}(\overline{Z}'\overline{Z})^{-1}\overline{Z}'U\| \leq \|X\|\,\|\overline{Z}(\overline{Z}'\overline{Z})^{-1}\overline{Z}'U\|$. □

Note that the same argument leads to

$$\sup_{T_1,...,T_m} \overline{Z}_0'\overline{Z}(\overline{Z}'\overline{Z})^{-1}\overline{Z}U' = O_p(T^{\alpha+1/2}). \tag{37}$$

Proof of Lemma 1: By the definition of d_t,

$$\sum_1^T u_t d_t = U'X(\hat{\beta} - \beta^0) + U'\overline{Z}^*\hat{\delta} - U'\overline{Z}_0\delta^0$$

where $\overline{Z}^*$ diagonally partitions Z at $(\hat{T}_1, ..., \hat{T}_m)$. Because $U'X\beta^0 = O_p(T^{1/2})$ and $U'\overline{Z}_0 = O_p(T^{1/2})$ (these terms do not depend on $(\hat{T}_1, ..., \hat{T}_m)$), to prove the lemma, it suffices to show $T^{-1}U'X\hat{\beta} = o_p(1)$ and $T^{-1}U'\overline{Z}^*\hat{\delta} = o_p(1)$. We shall prove a stronger result. Let $(T_1, ..., T_m)$ be an arbitrary partition and $\overline{Z}$ be the associated diagonal partition of Z. Also let $\hat{\beta}(\{T_j\})$ and $\hat{\delta}(\{T_j\})$ be, respectively, the estimates of β and δ corresponding to this same partition. We shall prove

$$\sup_{T_1,...,T_m} \frac{1}{T}U'X\hat{\beta}(\{T_j\}) = o_p(1), \tag{38}$$

$$\sup_{T_1,...,T_m} \frac{1}{T}U'\overline{Z}\hat{\delta}(\{T_j\}) = o_p(1), \tag{39}$$

where the supremum with respect to $(T_1, ..., T_m)$ is taken over the same partitions as those in Lemma 5. First consider (38). The estimator $\hat{\beta}(\{T_j\})$ can be written as

$$\begin{aligned} \hat{\beta}(\{T_j\}) &= (X'M_{\overline{Z}}X)^{-1}X'M_{\overline{Z}}Y \\ &= (X'M_{\overline{Z}}X)^{-1}X'M_{\overline{Z}}\overline{Z}_0\delta^0 + (X'M_{\overline{Z}}X)^{-1}X'M_{\overline{Z}}U. \end{aligned} \tag{40}$$

Using the argument of Lemma A.3, we deduce that $X'M_{\overline{Z}}U = O_p(T)$. This together with lemmas A.2 and A.3 implies that $\hat{\beta}(\{T_j\}) = O_p(1)$ uniformly over all partitions. Hence $T^{-1}U'X\hat{\beta}(\{T_j\}) = O_p(T^{-1/2})$ uniformly over all partitions, obtaining (38).

Next, consider (39). From $\hat{\delta}(\{T_j\}) = (\overline{Z}'M_X\overline{Z})^{-1}\overline{Z}'M_XY$ and $M_XX = 0$, we obtain

$$\begin{aligned} U'\overline{Z}\hat{\delta}(\{T_j\}) &= U'\overline{Z}(\overline{Z}'M_X\overline{Z})^{-1}\overline{Z}'M_X\overline{Z}_0\delta^0 \\ &\quad + U'\overline{Z}(\overline{Z}'M_X\overline{Z})^{-1}\overline{Z}'M_XU \\ &= (I) + (II). \end{aligned} \tag{41}$$

By Lemma A.4,

$$\begin{aligned} (I) &= U'\overline{Z}(\overline{Z}'\overline{Z})^{-1}\overline{Z}'M_X\overline{Z}_0\delta^0 \\ &\quad + U'\overline{Z}(\overline{Z}'\overline{Z})^{-1}\overline{Z}'X(X'M_{\overline{Z}}X)^{-1}X'P_{\overline{Z}}M_X\overline{Z}_0\delta^0. \end{aligned} \tag{42}$$

Because $P_{\overline{Z}}$ and M_X are projection matrices, $\|M_X\overline{Z}_0\| \leq \|\overline{Z}_0\| = O_p(T^{1/2})$ and $\|X'P_{\overline{Z}}M_X\overline{Z}_0\| \leq \|X\|\|\overline{Z}_0\| = O_p(T)$. Now, we have $(I) = O_p(T^{\alpha+1/2})$ uniformly over all partitions using Lemmas A.2, A.5, and A.6.

It remains to derive a bound for (II). Note that (I) and (II) differ only in that $\overline{Z}_0\delta^0$ is replaced by U. Similar argument shows that (II) is of a lower order of magnitude, more specifically $(II) = O_p(T^{2\alpha})$ (the details are omitted to avoid repetition). Because $2\alpha < \alpha + 1/2$, we have $U'\overline{Z}\hat{\delta}(\{T_j\}) = O_p(T^{\alpha+1/2})$. Equivalently, $T^{-1}U'\overline{Z}\hat{\delta}(\{T_j\}) = O_p(T^{\alpha-1/2}) = o_p(1)$ uniformly over all partitions. This proves (39) and, hence, the proof of Lemma 1 is complete. □

Proof of Lemma 2: If there exists a break, say, λ_j^0, which cannot be consistently estimated, then with some positive probability $\epsilon_0 > 0$ there exists a positive number $\eta > 0$ such that no estimated break falls in the interval $[T(\lambda_j^0 - \eta), T(\lambda_j^0 + \eta)]$ for a subsequence of T (without loss of generality, assume this subsequence is the same as T). Suppose this interval is classified into the k-th regime, namely, $\hat{T}_{k-1} \leq T(\lambda_j^0 - \eta)$ and $T(\lambda_j^0 + \eta) \leq \hat{T}_k$. Then $d_t = x_t'(\hat{\beta} - \beta^0) + z_t'(\hat{\delta}_k - \delta_j^0)$ for $t \in [T(\lambda_j^0 - \eta), T\lambda_j^0]$ and $d_t = x_t'(\hat{\beta} - \beta^0) + z_t'(\hat{\delta}_k - \delta_{j+1}^0)$ for $t \in [T\lambda_j^0 + 1, T(\lambda_j^0 + \eta)]$. We have

$$
\begin{aligned}
\sum_{t=1}^{T} d_t^2 &\geq \sum_1 d_t^2 + \sum_2 d_t^2 \\
&= \begin{pmatrix} \hat{\beta} - \beta^0 \\ \hat{\delta}_k - \delta_j^0 \end{pmatrix}' \begin{pmatrix} \sum_1 x_t x_t' & \sum_1 x_t z_t' \\ \sum_1 z_t x_t' & \sum_1 z_t z_t' \end{pmatrix} \begin{pmatrix} \hat{\beta} - \beta^0 \\ \hat{\delta}_k - \delta_j^0 \end{pmatrix} \qquad (43) \\
&+ \begin{pmatrix} \hat{\beta} - \beta^0 \\ \hat{\delta}_k - \delta_{j+1}^0 \end{pmatrix}' \begin{pmatrix} \sum_2 x_t x_t' & \sum_2 x_t z_t' \\ \sum_2 z_t x_t' & \sum_2 z_t z_t' \end{pmatrix} \begin{pmatrix} \hat{\beta} - \beta^0 \\ \hat{\delta}_k - \delta_{j+1}^0 \end{pmatrix},
\end{aligned}
$$

where $\sum_1$ extends over the set $T(\lambda_j^0 - \eta) \leq t \leq T\lambda_j^0$ and $\sum_2$ extends over the set $T\lambda_j^0 + 1 \leq t \leq T(\lambda_j^0 + \eta)$. Let γ_T be the smallest eigenvalue of the first matrix in (43) and γ_T^* be the smallest eigenvalue of the second matrix in (43). Then

$$
\sum_1 d_t^2 + \sum_2 d_t^2 \geq \gamma_T \left[\|\hat{\beta} - \beta^0\|^2 + \|\hat{\delta}_k - \delta_j^0\|^2\right] + \gamma_T^* \left[\|\hat{\beta} - \beta^0\|^2 + \|\hat{\delta}_k - \delta_{j+1}^0\|^2\right]
$$

$$
\geq \min\{\gamma_T, \gamma_T^*\} \left(\|\hat{\delta}_k - \delta_j^0\|^2 + \|\hat{\delta}_k - \delta_{j+1}^0\|^2\right) \geq \frac{1}{2} \min\{\gamma_T, \gamma_T^*\} \|\delta_j^0 - \delta_{j+1}^0\|^2.
$$

The last inequality follows from $(x-a)'A(x-a)+(x-b)'A(x-b) \geq (1/2)(a-b)'A(a-b)$ for an arbitrary positive definite matrix A and for all x. Now the first matrix in (43) can be written as $(T\eta)\frac{1}{T\eta}\sum_{T(\lambda_j^0-\eta)}^{T\lambda_j^0} w_t w_t' \equiv (T\eta)A_T$, say. By assumption A2, the smallest eigenvalue of A_T is bounded away from zero. Thus the smallest eigenvalue of $(T\eta)A_T$, γ_T, is of the order $T\eta$. The same can be said for γ_T^*. Therefore, $\sum_1^T d_t^2 >$

$T\eta C_1\|\delta_j^0 - \delta_{j+1}^0\|^2 \equiv TC\|\delta_j^0 - \delta_{j+1}^0\|^2$ for some $C = \eta C_1 > 0$ with probability no less than $\epsilon_0 > 0$. □

Proof of Proposition 2: Because $\hat{\lambda}_k$ $(k = 1, ..., m)$ is consistent by Proposition 1, with large probability, $\hat{\delta}_k$ is estimated using at least a positive fraction of the observations from $[T_{k-1}^0, T_k^0]$, say using $t \in [T(\lambda_{k-1}^0 + \epsilon), T(\lambda_k^0 - \epsilon)]$. Over this interval, $d_t = x_t(\hat{\beta} - \beta^0) + z_t'(\hat{\delta}_k - \delta_k^0)$. Hence $\sum_1^T d_t^2 \geq \sum_* d_t^2$ with $\sum_*$ extending over the same interval. If either $\hat{\beta}$ or $\hat{\delta}_k$ is not consistent, then, with some positive probability, either $\|\hat{\beta} - \beta\| > a$ or $\|\hat{\delta}_k - \delta_k^0\| > a$ will be true for some $a > 0$. Similar argument as in the proof of proposition 1 leads to $\sum_* d_t^2 > T(\lambda_k^0 - \lambda_{k-1}^0 - 2\epsilon)a^2C$ for some $C > 0$, with some positive probability. This again gives rise to a contradiction with (4) in view of (5) and Lemma 1. □

Proof of Proposition 3: Without loss of generality, we assume there are only three breaks ($m = 3$) and provide an explicit proof of T-consistency for $\hat{\lambda}_2$ only. The analysis for $\hat{\lambda}_1$ and $\hat{\lambda}_3$ is virtually the same (and actually simpler) and is thus omitted.

By the consistency result of proposition 1, for each $\epsilon > 0$ and T large, we have $|\hat{T}_k - T_k^0| < \epsilon T$, with high probability. Therefore we only need to examine the behavior of the sum of squared residuals, $S_T(T_1, T_2, T_3)$, for those T_i that are close to the true breaks such that $|T_i - T_i^0| < \epsilon T$ for all i. Also using an argument of symmetry, we can, without loss of generality, consider the case $T_2 < T_2^0$. For $C > 0$, define

$$T_\epsilon(C) = \{(T_1, T_2, T_3); |T_i - T_i^0| < \epsilon T, 1 \leq i \leq 3, T_2 - T_2^0 < -C\}.$$

Because $S_T(\hat{T}_1, \hat{T}_2, \hat{T}_3) \leq S_T(\hat{T}_1, T_2^0, \hat{T}_3)$ with probability 1, to prove the proposition it is enough to show that for each $\eta > 0$, there exist $C > 0$ and $\epsilon > 0$ such that for large T,

$$P(\min\{S_T(T_1, T_2, T_3) - S_T(T_1, T_2^0, T_3)\} \leq 0) < \eta, \tag{44}$$

where the minimum is taken over the set $T_\epsilon(C)$. Such a relation would imply that for a large C, global optimization cannot be achieved on $T_\epsilon(C)$. Thus with large probability, $\hat{T}_2 - T_2^0| \leq C$. Now denote

$$SSR_1 = S_T(T_1, T_2, T_3),$$

$$SSR_2 = S_T(T_1, T_2^0, T_3),$$

and introduce

$$SSR_3 = S_T(T_1, T_2, T_2^0, T_3).$$

By definition, we have

$$(45) \qquad S_T(T_1,T_2,T_3) - S_T(T_1,T_2^0,T_3) = (SSR_1 - SSR_3) - (SSR_2 - SSR_3).$$

This latter relation is useful because it allows us to carry the analysis in terms of two problems involving a single structural change. Indeed, note that $SSR_1 - SSR_3$ is the difference in the sums of squared residuals allowing an additional fourth break at time T_2^0 between the breaks T_2 and T_3. Similarly, $SSR_2 - SSR_3$ is the difference in the sums of squared residuals allowing an additional fourth break at time T_2 between the breaks T_1 and T_2^0. Hence, in each case SSR_1 and SSR_2 can be viewed as the sum of squared residuals from a constrained version of a more general model whose sum of squared residuals is SSR_3.

It is then easy to derive exact expressions for the two components on the right hand side of (45) in terms of estimated coefficients. Consider first $SSR_1 - SSR_3$, we have (e.g., Amemiya 1985, p. 31):

$$S_T(T_1,T_2,T_3) - S_T(T_1,T_2,T_2^0,T_3) = (\hat{\delta}_3^* - \hat{\delta}_\Delta)' Z_\Delta' M_{\overline{W}} Z_\Delta (\hat{\delta}_3^* - \hat{\delta}_\Delta),$$

where $\overline{W} = (X, \overline{Z})$, with $\overline{Z}$ the diagonal partition of Z at (T_1,T_2,T_3), $\hat{\delta}_3^*$ is the vector of estimated coefficients associated with the regressors $(0,...,0,z_{T_2^0+1},...,z_{T_3},0,...,0)'$, and $\hat{\delta}_\Delta$ is the vector of estimated coefficients associated with the regressors $Z_\Delta = (0,...,0,z_{T_2+1},...,z_{T_2^0},0,...,0)'$ (see Figure 3). Similarly, we have for $SSR_2 - SSR_3$:

$$S_T(T_1,T_2^0,T_3) - S_T(T_1,T_2,T_2^0,T_3) = (\hat{\delta}_2 - \hat{\delta}_\Delta)' Z_\Delta' M_{\tilde{W}} Z_\Delta (\hat{\delta}_2 - \hat{\delta}_\Delta),$$

where $\tilde{W} = (X, \tilde{Z})$ with $\tilde{Z}$ the diagonal partition of Z at (T_1,T_2^0,T_3), and $\hat{\delta}_2$ is the vector of estimated coefficients associated with the regressors $(0,...,0,z_{T_1+1},...,z_{T_2},0,...,0)'$ (again, see Figure 3). Thus

$$(46) \quad SSR_1 - SSR_2 = (\hat{\delta}_3^* - \hat{\delta}_\Delta)' Z_\Delta' M_{\overline{W}} Z_\Delta (\hat{\delta}_3^* - \hat{\delta}_\Delta) - (\hat{\delta}_2 - \hat{\delta}_\Delta)' Z_\Delta' M_{\tilde{W}} Z_\Delta (\hat{\delta}_2 - \hat{\delta}_\Delta)$$

The second term on the right is bounded by $(\hat{\delta}_2 - \hat{\delta}_\Delta)' Z_\Delta' Z_\Delta (\hat{\delta}_2 - \hat{\delta}_\Delta)$ because $Z_\Delta' M_{\tilde{W}} Z_\Delta \leq Z_\Delta' Z_\Delta$, $M_{\tilde{W}}$ being a projection matrix. Expanding the first term on the right hand side of (46), we have

$$\begin{aligned} & SSR_1 - SSR_2 \\ (47) \qquad & \geq (\hat{\delta}_3^* - \hat{\delta}_\Delta)' Z_\Delta' Z_\Delta (\hat{\delta}_3^* - \hat{\delta}_\Delta) - (\hat{\delta}_3^* - \hat{\delta}_\Delta) Z_\Delta' \overline{W} (\overline{W}'\overline{W})^{-1} \overline{W}' Z_\Delta' (\hat{\delta}_3^* - \hat{\delta}_\Delta) \\ & \quad -(\hat{\delta}_2 - \hat{\delta}_\Delta)' Z_\Delta' Z_\Delta (\hat{\delta}_2 - \hat{\delta}_\Delta) \equiv (I) - (II) - (III). \end{aligned}$$

Consider the limiting behavior of term (I). Note first that the estimates $\hat{\delta}_3^*$ will be close to δ_3^0 given that, on the set $T_\epsilon(C)$, the distance between T_i and T_i^0 can be controlled and made small by choosing a small ϵ. Noting that $\hat{\delta}_\Delta$ is estimated using observations from the second true regime only, $\hat{\delta}_\Delta$ is close to δ_2^0 for a large enough C, on $T_\epsilon(C)$. Hence, for large C, large T and small ϵ, expression (I) is no less than $(1/2)(\delta_3^0 - \delta_2^0)' Z_\Delta' Z_\Delta (\delta_3^0 - \delta_2^0)$ with large probability.

Next consider term (II). By the strong law of large numbers, it is easy to argue that on $T_\epsilon(C)$, $\hat{\delta}_3^*$ and $\hat{\delta}_\Delta$ are $O_p(1)$ uniformly. Also on $T_\epsilon(C)$, $(\overline{W}'\overline{W}/T)^{-1} = O_p(1)$ and $Z_\Delta'\overline{W}/C = O_p(1)$ (because $Z_\Delta'\overline{W}$ involves no more than C observations). Thus (II) is no larger than $(1/T)C^2 O_p(1)$.

Consider finally (III). Because both $\hat{\delta}_2$ and $\hat{\delta}_\Delta$ are close to δ_2^0, $\|\hat{\delta}_2 - \hat{\delta}_\Delta\| < \rho$ with large probability for any given small number $\rho > 0$ (this is true for large T, large C and small ϵ). Thus (III) is no larger than $\rho \iota' Z_\Delta' Z_\Delta \iota$, with ι a vector of 1's.

In summary, we have that the inequality

$$(48) \quad SSR_1 - SSR_2 \geq (1/2)(\delta_3^0 - \delta_2^0)' Z_\Delta' Z_\Delta (\delta_3^0 - \delta_2^0) - \frac{1}{T} C^2 O_p(1) - \rho \iota' Z_\Delta' Z_\Delta \iota,$$

holds with large probability. The first term on the right hand side of (48) is of the same order of magnitude as C since the smallest eigenvalue of $Z_\Delta' Z_\Delta / C$ is bounded away from zero by A2. The other two terms are dominated by the first term and thus with large probability, $SSR_1 - SSR_2 > 0$. This proves (44) and thus the proposition. □

Proof of Proposition 7: Note that we can write:

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

where SSR_0 and SSR_k are the sum of squared residuals under the null hypothesis and under the alternative allowing k breaks, respectively. We have $(T - (k+1)q - p)^{-1} SSR_k \to_p \sigma^2$. Hence, we concentrate on the limit of $F_T^* = SSR_0 - SSR_k$. Now, let $D^U(i,j)$ ($D^R(i,j)$, resp.) be the sum of squared residuals from the unrestricted (restricted, resp.) model using data from segments i to j (inclusively), i.e. from observation $T_{i-1} + 1$ to T_j (these notations have different meanings from the $D(i,j)$ defined in Section 4.3, where i and j refer to the numbering of the observations not the numbering of segments). We can then write:

$$F_T^* = D^R(1, k+1) - \sum_{i=1}^{k+1} D^U(i,i),$$

or

$$(49)\ F_T^* = \sum_{i=1}^{k} \left[D^R(1,i+1) - D^R(1,i) - D^U(i+1,i+1) \right] + D^R(1,1) - D^U(1,1).$$

Consider first the estimate of the coefficients on the x's. Let $\hat{\beta}^U$ and $\hat{\beta}^R$ be the estimate of β in the unrestricted and restricted models, respectively. We have $\hat{\beta}^U = (X'M_{\overline{Z}}X)^{-1}X'M_{\overline{Z}}Y$ and $\hat{\beta}^R = (X'M_ZX)^{-1}X'M_ZY$ where $Z = (z_1', ..., z_T')'$. We need to introduce further notations. Let $Y_{1,j}$, $U_{1,j}$, $X_{1,j}$ and $Z_{1,j}$ denote the corresponding vectors or matrices containing elements belonging to the partition from segment 1 to segment j (inclusively). Also, let Y_j, U_j, X_j and Z_j be the vectors or matrices containing elements from segment j only. Now, let $\hat{\delta}_{1,j}^R$ be the estimate of δ using data on the z's from segments 1 to j in the restricted model. Also, let $\hat{\delta}_j^U$ be the estimate of δ_j using data on the z's from segment j only in the unrestricted model. We have

$$\begin{aligned} \hat{\delta}_{1,j}^R &= (Z_{1,j}'Z_{1,j})^{-1}Z_{1,j}'(Y_{1,j} - X_{1,j}\hat{\beta}^R), \\ \hat{\delta}_j^U &= (Z_j'Z_j)^{-1}Z_j'(Y_j - X_j\hat{\beta}^U). \end{aligned}$$

Using the fact that, under the null hypothesis, $Y = X\beta + Z\delta + U = X\beta + \overline{Z}\bar{\delta} + U$ and $Y_j = X_j\beta + Z_j\delta + U_j$ (with $\bar{\delta} = (\delta', ..., \delta')'$ a $(k+1)$ vector with δ defined by $\delta_1 = \delta_2 = ... = \delta_{k+1} \equiv \delta$), straightforward algebra yields

$$\begin{aligned} D^R(1,j) &= \| (I - P_{Z_{1,j}})(U_{1,j} - X_{1,j}A_T) \|^2, \\ D^U(j,j) &= \| (I - P_{Z_j})(U_j - X_j\bar{A}_T) \|^2, \end{aligned}$$

where

$$\begin{aligned} A_T &= (X'M_ZX)^{-1}X'M_ZU, \\ \bar{A}_T &= (X'M_{\overline{Z}}X)^{-1}X'M_{\overline{Z}}U. \end{aligned}$$

Consider the ith element in the summation defining F_T^* in (49), we have

$$\begin{aligned} F_{T,i} &= D^R(1,i+1) - D^R(1,i) - D^U(i+1,i+1) \\ &= \| (I - P_{Z_{1,i+1}})(U_{1,i+1} - X_{1,i+1}A_T) \|^2 - \| (I - P_{Z_{1,i}})(U_{1,i} - X_{1,i}A_T) \|^2 \\ &\quad - \| (I - P_{Z_{i+1}})(U_{i+1} - X_{i+1}\bar{A}_T) \|^2 \end{aligned}$$

To simplify the exposition, we introduce the notation $S_j = Z_{1,j}'U_{1,j}$, $H_j = Z_{1,j}'Z_{1,j}$, $K_j = Z_{1,j}'X_{1,j}$, $L_j = X_{1,j}'X_{1,j}$, and $M_j = X_{1,j}'U_{1,j}$. Noting that

$$\begin{aligned} U_{1,i+1}'U_{1,i+1} &= U_{1,i}'U_{1,i} + U_{i+1}'U_{i+1}, \\ X_{1,i+1}'X_{1,i+1} &= X_{1,i}'X_{1,i} + X_{i+1}'X_{i+1}, \\ U_{1,i+1}'X_{1,i+1} &= U_{1,i}'X_{1,i} + U_{i+1}'X_{i+1}, \end{aligned}$$

we deduce that

$$\begin{aligned} F_{T,i} &= -S'_{i+1}H^{-1}_{i+1}S_{i+1} + S'_iH^{-1}_iS_i + (S_{i+1}-S_i)'[H_{i+1}-H_i]^{-1}(S_{i+1}-S_i) \\ &\quad +2S'_{i+1}H^{-1}_{i+1}K_{i+1}A_T - 2S'_iH^{-1}_iK_iA_T \\ &\quad -2(S_{i+1}-S_i)'[H_{i+1}-H_i]^{-1}(K_{i+1}-K_i)\bar{A}_T \\ &\quad +2(M_{i+1}-M_i)'(\bar{A}_T-A_T) + (\bar{A}_T-A_T)'(L_{i+1}-L_i)(\bar{A}_T-A_T). \end{aligned}$$

Using the stated assumptions, we have the following basic convergence results:

1) $T^{-1/2}(X_{1,j}, Z_{1,j})'U_{1,j} \Rightarrow \sigma(B_1(\lambda_j), B_2(\lambda_j))' \equiv \sigma B(\lambda_j)'$ where $B(r)$ is a $(q+p)$ dimensional vector Brownian motion with covariance matrix

$$Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}.$$

2) $T^{-1}(X_{1,j}, Z_{1,j})'(X_{1,j}, Z_{1,j}) \rightarrow^p \sigma^2\lambda_jQ$.

From these two limits, we deduce easily the following results

a) $T^{-1/2}S_j \Rightarrow \sigma B_2(\lambda_j)$;

b) $T^{-1}H_j \rightarrow^p \sigma^2\lambda_jQ_{22}$;

c) $T^{-1}K_j \rightarrow^p \sigma^2\lambda_jQ_{21}$;

d) $T^{-1}L_j \rightarrow^p \sigma^2\lambda_jQ_{11}$;

e) $T^{-1/2}M_j \Rightarrow \sigma B_1(\lambda_j)$;

f)

$$\begin{aligned} T^{1/2}A_T &= (T^{-1}X'X - T^{-1}X'Z(T^{-1}Z'Z)^{-1}T^{-1}Z'X)^{-1} \\ &\quad \times(T^{-1/2}X'U - T^{-1}X'Z(T^{-1}Z'Z)^{-1}T^{-1/2}Z'U) \\ &\Rightarrow \sigma^{-1}(Q_{11}-Q_{12}Q_{22}^{-1}Q_{21})^{-1}(B_1(1)-Q_{12}Q_{22}^{-1}B_2(1)) \equiv A^*. \end{aligned}$$

It remains to consider the limit of $T^{1/2}\bar{A}_T$. Let $\Lambda = diag\{\lambda_1, \lambda_2-\lambda_1, ..., 1-\lambda_k\}$, a $(k+1)$ by $(k+1)$ diagonal matrix. We deduce that

i) $T^{-1}\overline{Z}'\overline{Z} \rightarrow^p \sigma^2(\Lambda \otimes Q_{22})$;

ii) $T^{-1}X'\overline{Z} \rightarrow^p \sigma^2(e'\Lambda \otimes Q_{12})$ where $e' = (1, 1, ..., 1)$, a $(k+1)$ vector;

iii) $T^{-1/2}\overline{Z}'U \Rightarrow \sigma(B_2(\lambda_1), B_2(\lambda_2-\lambda_1), ..., B_2(1-\lambda_k))' \equiv B^*$.

We then obtain

$$\begin{aligned} T^{1/2}\bar{A}_T &= [T^{-1}X'X - T^{-1}X'\overline{Z}(T^{-1}\overline{Z}'\overline{Z})^{-1}T^{-1}\overline{Z}'X]^{-1} \\ &\quad \times[T^{-1/2}X'U - T^{-1}X'\overline{Z}(T^{-1}\overline{Z}'\overline{Z})^{-1}T^{-1/2}\overline{Z}'U] \\ &\Rightarrow \sigma^{-1}[Q_{11} - (e'\Lambda \otimes Q_{12})(\Lambda \otimes Q_{22})^{-1}(\Lambda e \otimes Q_{21})]^{-1} \end{aligned}$$

$$
\begin{aligned}
&\times[B_1(1) - (e'\Lambda \otimes Q_{12})(\Lambda \otimes Q_{22})^{-1}B^*] \\
= \;& \sigma^{-1}[Q_{11} - (e'\Lambda e \otimes Q_{12}Q_{22}^{-1}Q_{21})]^{-1}[B_1(1) - (e' \otimes Q_{12}Q_{22}^{-1})B^*] \\
= \;& \sigma^{-1}[Q_{11} - Q_{12}Q_{22}^{-1}Q_{21}]^{-1}[B_1(1) - Q_{12}Q_{22}^{-1}B_2(1)] \equiv A^*.
\end{aligned}
$$

The second equality follows since $e'\Lambda e = 1$ and $(e' \otimes Q_{12}Q_{22}^{-1})B^* = Q_{12}Q_{22}^{-1}B_2(1)$.

Using the results stated above we easily deduce that

$$(M_{i+1} - M_i)'(\bar{A}_T - A_T) \Rightarrow 0,$$

$$(\bar{A}_T - A_T)'(L_{i+1} - L_i)(\bar{A}_T - A_T) \Rightarrow 0,$$

$$
\begin{aligned}
&S'_{i+1}H_{i+1}^{-1}K_{i+1}A_T - S'_iH_i^{-1}K_iA_T - (S_{i+1} - S_i)'[H_{i+1} - H_i]^{-1}(K_{i+1} - K_i)\bar{A}_T \\
&\Rightarrow \sigma B_2(\lambda_{i+1})Q_{22}^{-1}Q_{21}A^* - \sigma B_2(\lambda_i)Q_{22}^{-1}Q_{21}A^* \\
&\quad -\sigma(B_2(\lambda_{i+1}) - B_2(\lambda_i))Q_{22}^{-1}Q_{21}A^* = 0.
\end{aligned}
$$

Hence, we are left with

$$F_{T,i} = -S'_{i+1}H_{i+1}^{-1}S_{i+1} + S'_iH_i^{-1}S_i + (S_{i+1} - S_i)'[H_{i+1} - H_i]^{-1}(S_{i+1} - S_i) + o_p(1),$$

and we deduce, using the fact that $B_2(\lambda_j) = \sigma Q_{22}^{1/2}W(\lambda_j)$ with $W(\lambda_j)$ a vector of q independent standard Wiener processes,

$$
\begin{aligned}
F_{T,i} \;\Rightarrow\; & -B_2(\lambda_{i+1})'Q_{22}^{-1}B_2(\lambda_{i+1})/\lambda_{i+1} + B_2(\lambda_i)'Q_{22}^{-1}B_2(\lambda_i)/\lambda_i \\
& +(B_2(\lambda_{i+1}) - B_2(\lambda_i))'Q_{22}^{-1}(B_2(\lambda_{i+1}) - B_2(\lambda_i))/(\lambda_{i+1} - \lambda_i) \\
= \;& -\sigma^2 \parallel W(\lambda_{i+1}) \parallel^2 /\lambda_{i+1} + \sigma^2 \parallel W(\lambda_i) \parallel^2 /\lambda_i \\
& +\sigma^2 \parallel W(\lambda_{i+1}) - W(\lambda_i) \parallel^2 /(\lambda_{i+1} - \lambda_i) \\
= \;& \sigma^2 \parallel \lambda_i W(\lambda_{i+1}) - \lambda_{i+1}W(\lambda_i) \parallel^2 /\lambda_{i+1}\lambda_i(\lambda_{i+1} - \lambda_i).
\end{aligned}
$$

Finally, note that

$$
\begin{aligned}
D^R(1,1) - D^U(1,1) &= \parallel (I - P_{Z_1})(U_1 - X_1A_T) \parallel^2 - \parallel (I - P_{Z_1})(U_1 - X_1\bar{A}_T) \parallel^2 \\
&= 2U_1'(I - P_{Z_1})X_1(\bar{A}_T - A_T) \\
&\quad +A_T'X_1'(I - P_{Z_1})X_1A_T - \bar{A}_T'X_1'(I - P_{Z_1})X_1\bar{A}_T \\
&\Rightarrow 0.
\end{aligned}
$$

Hence

$$F_T^* \Rightarrow \sigma^2 \sum_{i=1}^{k} \frac{\parallel \lambda_i W(\lambda_{i+1}) - \lambda_{i+1}W(\lambda_i) \parallel^2}{\lambda_{i+1}\lambda_i(\lambda_{i+1} - \lambda_i)},$$

and the result of Proposition 7 follows. □

Proof of Proposition 9: Let $c_{T,\ell} = c(\alpha_T, \ell, T)$ be the critical value of the test $F_T(\ell+1|\ell)$ corresponding to a given size α_T. By definition,

$$P\Big(F_T(\ell+1|\ell) > c_{T,\ell} \text{| conditional on } \ell \text{ breaks}\Big) \leq \alpha_T, \tag{50}$$

for all ℓ. Now suppose the true number of breaks is m_0. By the consistency of the sequential test, we have

$$P\Big(F_T(\ell+1|\ell) > c_{T,\ell} \text{| conditional on } m_0 \text{ breaks}\Big) \to 1, \quad \text{for } \ell = 0, 1, ..., m_0 - 1, \tag{51}$$

since $c_{T,\ell}$ increases slowly enough given the assumption on the rate of decrease of α_T. Let $\hat{m}$ be the number of breaks estimated by the sequential procedure. The event $\{\hat{m} < m_0\}$ satisfies

$$\{\hat{m} < m_0\} \subset \cup_{k=0}^{m_0-1} A_{T,k},$$

where $A_{T,k} = \{F_T(k+1|k) \leq c_{T,k}\}$. That is, in order to obtain $\hat{m} < m_0$, it must be the case that the hypothesis of k against $k+1$ breaks cannot be rejected for some $k < m_0$. By (51), $P(A_{T,k}|m_0) \to 0$ as $T \to \infty$ for each $k < m_0$. It follows that

$$P(\hat{m} < m_0) \leq \sum_{k=0}^{m_0-1} P(A_{T,k}|m_0) \to 0, \text{ as } T \to \infty. \tag{52}$$

Next, consider the event $\{\hat{m} > m_0\}$. In order to obtain $\hat{m} > m_0$, it must be the case that the hypothesis of m_0 breaks against m_0+1 breaks is rejected. This happens, in large samples, with probability no more than α_T, given m_0 breaks. Formally, by (50) with $\ell = m_0$,

$$P(\hat{m} > m_0) \leq P\Big(F_T(m_0+1|m_0) > c_{T,m_0}|m_0\Big) \leq \alpha_T. \tag{53}$$

Combining (52) and (53), we see that $P(\hat{m} \neq m_0) \to 0$ as $T \to \infty$. That is, the estimate of the number of breaks obtained using the sequential test is consistent. □

B Computational Appendix.

In this section, we discuss an algorithm based on the principle of dynamic programming that allows the computation of estimates of the break points as global minimizers of the sum of squared residuals. The method is directly applicable to the case of a pure structural change model. Useful references include Guthery (1974), Bellman and Roth (1969) and Fisher (1958).

A standard grid search procedure to obtain global minimizers with m breaks would require least squares operations of order $O(T^m)$. The dynamic programming approach provides an efficient method that requires least squares operations of order $O(T^2)$ at most for any number of breaks; hence substantial savings in computations can be achieved when estimating a model with more than two breaks (the method also allows some savings in the case of two breaks). The basic reason for this possible reduction in computation is fairly intuitive once it is realized that, with a sample of size T, the total number of possible segments is at most $T(T+1)/2$ and is therefore of order $O(T^2)$. The dynamic programming algorithm can be seen as an efficient way to compare possible combinations of these partitions to achieve a minimum sum of squared residuals.

In practice, less than $T(T+1)/2$ segments are permissible. First, some minimum distance between each break may be imposed, as is done in the construction of the tests discussed in Section 4. Let this minimum distance be denoted by h. Note that $h < q$ is possible in which case the sum of squared residuals is zero; for simplicity we suppose without loss of generality that $h \geq q$. This implies a reduction in the number of segments to be considered of $(h-1)T - (h-2)(h-1)/2$. Now the largest possible segment must be such as to allow m other segments before or after. For example, when the segment starts at a date between 1 and $h-1$, the maximal length of this segment is $T - hm$ when m breaks are allowed (i.e., $m+1$ regimes). This allows a further reduction in the total number of segments considered of $h^2m(m+1)/2 - mh$. Hence all the relevant information can be obtained from the examination of the sums of squared residuals associated with

$$T(T+1)/2 - (h-1)T + (h-2)(h-1)/2 - h^2m(m+1)/2 + mh$$

segments. We therefore need to evaluate the sum of squared residuals associated with

segments having the following starting and ending dates:

starting date	ending date
$i = 1, ..., hm - 1$	$j = h + i - 1, ..., T - hm$
$i = \ell h, ..., (\ell + 1)h - 1$	$j = h + i - 1, ..., T - (m - \ell)h$ $(\ell = 1, ..., m - 1)$
$i = hm + 2, ..., T - hm + 1$	$j = h + i - 1, ..., T$

This can be achieved using standard updating formulae to calculate recursive residuals. Indeed, all the relevant information can be calculated from $T - hm + 1$ sets of recursive residuals and the fact that the sum of squared residuals using, say, t observations is the sum of squared residuals using $t - 1$ observations plus the square of the recursive residual at time t. Hence, the number of matrix inversions needed is simply of an order $O(T)$. To be precise, let $v(i, j)$ be the recursive residual at time j obtained using a sample that starts at date i, and let $SSR(i, j)$ be the sum of squared residuals obtained by applying least-squares to a segment that starts at date i and ends at date j. We note the following recursive relation (e.g., Brown, Durbin and Evans (1975)):

$$SSR(i, j) = SSR(i, j - 1) + v(i, j)^2.$$

All the relevant information is contained in the values $SSR(i, j)$ for the combinations (i, j) indicated above.

Once the sums of squared residuals of the relevant segments have been computed and stored, a dynamic programming approach can be used to evaluate which partition achieves a global minimization of the overall sum of squared residuals. This method essentially proceeds via a sequential examination of optimal one-break (or two segments) partitions. Let $SSR(\{\hat{T}_{r,n}\})$ denote the sum of squared residuals associated with the optimal partition containing r breaks using the first n observations. The optimal partition can be obtained solving the following recursive problem:

$$SSR(\{T_{m,T}\}) = \min_{mh \le j \le T-h} [SSR(\{T_{m-1,j}\}) + SSR(j + 1, T)]. \tag{54}$$

It is instructive to write (54) in the following way:

$$\begin{gathered} SSR(\{T_{m,T}\}) = \\ \min_{mh \le j_1 \le T-h} [SSR(j_1 + 1, T) + \\ \min_{(m-1)h \le j_2 \le j_1 - h} [SSR(j_2 + 1, j_1) + \\ \min_{(m-2)h \le j_3 \le j_2 - h} [SSR(j_3 + 1, j_2) + \\ \vdots \\ \min_{h \le j_m \le j_{m-1} - h} [SSR(1, j_m) + SSR(j_m + 1, j_{m-1})] ...]]] \end{gathered}$$

Looking at the last displayed minimization problem, we see that the procedure starts by evaluating the optimal one-break partition for all sub-samples that allow a possible break ranging from observations h to $T-mh$. Hence, the first step is to store a set of $T-(m+1)h+1$ optimal one break partitions along with their associated sum of squared residuals. Each of the optimal partitions correspond to subsamples ending at dates ranging from $2h$ to $T-(m-1)h$. Consider now the next step which proceeds in a search for optimal partitions with two breaks. Such partitions have ending dates ranging from $3h$ to $T-(m-2)h$. For each of these possible ending dates, the procedure looks at which one-break partition can be inserted to achieve a minimal sum of squared residual. The outcome is a set of $T-(m+1)h+1$ optimal two breaks (or three segments) partitions. The method continues sequentially until a set of $T-(m+1)h+1$ optimal $(m-1)$ breaks partitions are obtained with ending dates ranging from $(m-1)h$ to $T-2h$. The final step is to see which of these optimal $(m-1)$ breaks partitions yields an overall minimal sum of squared residuals when combined with an additional segment. The method can therefore be viewed as a sequential updating of $T-(m+1)h+1$ segments into optimal one, two and up to $m-1$ breaks partitions (or into two, three and up to m sub-segments); the last step simply creating a single optimal m breaks (or $m+1$ segments) partition.

This dynamic programming method to obtain global minimizers of the sum of squared residuals cannot be applied directly to the case of a partial structural change model. This is basically due to the fact that we cannot concentrate out the parameters β without knowing the appropriate partition, i.e. the estimate of β associated with a global minimization depend on the optimal partition which we are trying to obtain. However, a simple iterative procedure is available. Let $\theta=(\delta,T_1,...,T_m)$, we can write the sum of squared residuals as a function of the vectors β and θ, i.e. $SSR(\beta,\theta)$. As discussed in Sargan (1964), we can minimize $SSR(\beta,\theta)$ in an iterative fashion as follows. First minimize with respect to θ keeping β fixed and then minimize with respect to β keeping θ fixed, and iterate. Each iteration assures a decrease in the objective function. The convergence properties of this scheme are discussed in Sargan (1964).

Note that the first step, minimizing with respect to θ keeping β fixed, amounts to applying the dynamic programming algorithm discussed above with $y_t-x_t'\beta$ as the dependent variable. Since β is fixed this is, indeed, a step involving a pure structural change model. Let $\theta^*=(\delta^*,\{T^*\})$ be the associated estimate from this first stage (with $\{T^*\}=(T_1^*,...,T_m^*)$). The second step is a simple linear regression with dependent

variable $y_t - z_t'\delta_j^*$ for t in regime j ($j = 1, ..., m+1$) (the regimes being defined by the partition $\{T^*\}$) and independent variable x_t.

An issue that remains is the choice of the initial value of the vector β to start the iteration. We suggest the following procedure. First apply the dynamic programming algorithm treating all coefficients as subject to change, i.e. treat the model as one of pure structural change and let $\theta^a = (\delta^a, \{T^a\})$ be the estimates of δ and $(T_1, ..., T_m)$ that then minimizes the sum of squared residuals. The initial value of the vector β is taken as the OLS estimate in a regression of $y_t - z_t'\delta_j^a$ on x_t for t in regime j ($j = 1, ..., m+1$), the regimes being defined by the partition $\{T^a\}$. The reason for such a choice of the startup value is that the estimates of the break fractions are consistent even when some of the coefficient of the parameter vector $\delta = (\delta_1, ..., \delta_q)$ do not change across regimes provided at least one does change at each break date. Hence, this choice of the starting value for β is asymptotically equivalent to the estimate associated with the global minimization and should, accordingly, allow obtaining global minimizers with respect to all the parameters in a few iterations.

References

[1] Alogoskoufis, G.S. and R. Smith (1991): "The Phillips Curve, The Persistence of Inflation, and the Lucas Critique: Evidence from Exchange Rate Regimes," *American Economic Review* 81, 1254-1275.

[2] Amemiya, T. (1985): *Advanced Econometrics*, Cambridge: Harvard University Press.

[3] Andrews, D.W.K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59, 817-858.

[4] Andrews, D.W.K. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica* 61, 821-856.

[5] Andrews, D.W.K. and J.C. Monahan (1992): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica* 60, 953-966.

[6] Andrews, D.W.K., I. Lee and W. Ploberger (1992): "Optimal Changepoint Tests for Normal Linear Regression," Cowles Foundation Discussion Paper No. 1016, Yale University.

[7] Andrews, D.W.K. and W. Ploberger (1992): "Optimal Tests of Parameter Constancy," Cowles Foundation Discussion Paper No. 1015, Yale University.

[8] Bai, J. (1994a): "Least Squares Estimation of a Shift in Linear Processes," *Journal of Time Series Analysis* 15, 453-472.

[9] Bai, J. (1994b): "Estimation of Structural Change based on Wald Type Statistics," Working Paper No. 94-6, Department of Economics, M.I.T.

[10] Bai, J. (1994c): "GMM Estimation of Multiple Structural Changes," NSF Proposal.

[11] Bai, J. (1995a): "Least Absolute Deviation Estimation of a Shift," *Econometric Theory* 11, 403-436.

[12] Bai, J. (1995b): "Estimating Multiple Breaks one at a Time," manuscript, Massachusetts Institute of Technology.

[13] Bai, J., R.L. Lumsdaine, and J.H. Stock (1994): "Testing for and Dating Breaks in Integrated and Cointegrated Time Series," manuscript, Kennedy School of Government, Harvard University.

[14] Banerjee, A., R.L. Lumsdaine, and J.H. Stock (1992): "Recursive and Sequential Tests of the Unit Root and Trend Break Hypothesis: Theory and International Evidence," *Journal of Business and Economic Statistics* 10, 271-287.

[15] Bellman, R. and R. Roth (1969): "Curve Fitting by Segmented Straight Lines," *Journal of the American Statistical Association* 64, 1079-1084.

[16] Bhattacharya, P.K. (1987): "Maximum Likelihood Estimation of a Change-point in the Distribution of Independent Random Variables: General Multiparameter Case," *Journal of Multivariate Analysis* 23, 183-2

[17] Brown, R.L., J. Durbin, and J.M. Evans (1975): "Techniques for Testing the Constancy of Regression Relationships Over Time," *Journal of the Royal Statistical Society*, Series B, 37, 149-192.

[18] Chong, T. T-L. (1994): "Consistency of Change-Point Estimators When the Number of Change-Points in Structural Models is Underspecified," manuscript, Department of Economics, University of Rochester.

[19] Christiano, L.J. (1992): "Searching for a Break in GNP", *Journal of Business & Economic Statistics* 10, 237-250.

[20] Chu, C-S. J. and H. White (1992): "A Direct Test for Changing Trend," *Journal of Business and Economic Statistics* 10, 289-300.

[21] Deshayes, J. and D. Picard (1986): "Off-line Statistical Analysis of Change Point Models Using Non Parametric and Likelihood Methods," in M. Basseville and A. Benveniste (eds.), *Detection of Abrupt Changes in Signals and Dynamical Systems*, Springer Verlag Lecture Notes in Control and Information Sciences No. 77, 103-168.

[22] DeLong, D.M. (1981): "Crossing Probabilities for a Square Root Boundary by a Bessel Process," *Communications in Statistics-Theory and Methods*, A10(21), 2197-2213.

[23] Feder, P.I. (1975): "On Asymptotic Distribution Theory in Segmented Regression Problem: Identified Case," *Annals of Statistics* 3, 49-83.

[24] Fisher, W.D. (1958): "On Grouping for Maximum Homogeneity," *Journal of the American Statistical Association* 53, 789-798.

[25] Fu, Y-X. and R.N. Curnow (1990): "Maximum Likelihood Estimation of Multiple Change Points," *Biometrika* 77, 563-573.

[26] Garcia, R. and P. Perron (1994): "An Analysis of the Real Interest Rate under Regime Shifts," forthcoming in *Review of Economics and Statistics.*

[27] Gregory, A.W. and B. Hansen (1993): "Testing for Cointegration with Regime Shifts," forthcoming in *Journal of Econometrics.*

[28] Guthery, S.B. (1974): "Partition Regression," *Journal of the American Statistical Association* 69, 945-947.

[29] Hall, P. and Heyde C. C. (1980): *Martingale Limit Theory and its Applications*, Academic Press, New York.

[30] Hansen, B.E. (1991): "Strong Laws for Dependent Heterogeneous Processes," *Econometric Theory* 7, 213-221.

[31] Hansen, B.E. (1992): "Tests for Parameter Instability in Regressions with I(1) Processes," *Journal of Business and Economic Statistics* 10, 321-335.

[32] Hawkins, D.L. (1987): "A Test for a Change Point in a Parametric Model Based on a Maximal Wald-type Statistic," *Sankhya* 49, 368-376.

[33] Kim, H.J. and D. Siegmund (1989): "The Likelihood Ratio Test for a Change Point in Simple Linear Regressions," *Biometrika* 76, 409-423.

[34] Kim, J.-Y. (1993): "Analysis of Structural Change in Economic Time Series - Detection and Estimation," manuscript, University of Minnesota.

[35] Kramer, W.,W. Ploberger, and R. Alt (1988): "Testing for Structural Changes in Dynamic Models," *Econometrica* 56, 1355-1370.

[36] Krishnaiah, P.R. and B.Q. Miao (1988): "Review about Estimation of Change Points," in *Handbook of Statistics*, vol. 7, P.R. Krishnaiah and C.R. Rao (eds.). New York: Elsevier.

[37] Liu, J., S. Wu, and J.V. Zidek (1994): "On Segmented Multivariate Regressions," manuscript, Department of Statistics, University of British Columbia.

[38] Lumsdaine, R.L. and D.H. Papell (1995): "Multiple Trend Breaks and the Unit Root Hypothesis," manuscript, Princeton University.

[39] Newey, W.K. and K.D. West (1987): "A Simple Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55, 703-708.

[40] Picard, D. (1985): "Testing and Estimating Change-point in Time Series," *Advances in Applied Probability* 17, 841-867.

[41] Perron, P. (1989): "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis," *Econometrica* 57, 1361-1401.

[42] Perron, P. (1991): "A Test for Changes in a Polynomial Trend Function for a Dynamic Time Series," manuscript, Department of Economics, Princeton University.

[43] Perron. P. (1993): "Trend, Unit Root and Structural Change in Macroeconomic Time Series," in *Cointegration for the Applied Economists*, B.B. Rao (ed.), Basingstoke: MacMillan Press, 113-146.

[44] Perron, P. (1994): "Further Evidence of Breaking Trend Functions in Macroeconomic Variables," manuscript, Département de Sciences Economiques, Université de Montréal.

[45] Perron, P. and T.J. Vogelsang (1992): "Nonstationarity and Level Shifts with an Application to Purchasing Power Parity," *Journal of Business and Economic Statistics* 10, 321-335.

[46] Pollard, D. (1984): *Convergence of Stochastic Processes*, New York: Springer-Verlag.

[47] Sargan, J.D. (1964): "Wages and Prices in the United Kingdom: A Study in Econometric Methodology," in P.E. Hart, G. Mills and J.K. Whitaker (eds.), *Econometric Analysis for National Economic Planning*, London: Butterworths, 25-54.

[48] Vogelsang, T.J. (1993): "Wald-type Tests for Detecting Shifts in the Trend Function of a Dynamic Time Series," manuscript, Cornell University.

[49] White, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test of Heteroskedasticity," *Econometrica* 48, 817-838.

[50] Yao, Y.-C.(1987): "Approximating the Distribution of the ML Estimate of the Change-point in a Sequence of Independent r.v.'s," *Annals of Statistics* 3, 1321-1328.

[51] Yao, Y-C. (1988): "Estimating the Number of Change-Points via Schwarz' Criterion," *Statistics and Probability Letters* 6, 181-189.

[52] Yao, Y.-C. and S.T. Au (1989): "Least Squares Estimation of a Step Function," *Sankhyā* 51 Ser. A, 370-381.

[53] Yin, Y.Q. (1988): "Detection of the Number, Locations and Magnitudes of Jumps," *Communications in Statistics-Stochastic Models* 4, 445-455.

[54] Zacks, S. (1983): "Survey of Classical and Bayesian Approaches to the Change-Point Problem: Fixed and Sequential Procedures of Testing and Estimation," in *Recent Advances in Statistics*, M.H. Rivzi, J.S. Rustagi, and D. Sigmund (eds.). New York: Academic Press.

[55] Zivot, E. and D.W.K. Andrews (1992), "Further Evidence on the Great Crash, the Oil Price Shock and the Unit Root Hypothesis," *Journal of Business and Economic Statistics* 10, 251-271.

Table 1: Asymptotic Critical Values of the Multiple-Break Test.
The entries are the quantiles x such that $P(\text{supF}_{k,q} \leq x/q) = \alpha$.

		Number of Breaks, k									
q	α	1	2	3	4	5	6	7	8	9	DmaxF
1	.90	8.02	7.87	7.07	6.61	6.14	5.74	5.40	5.09	4.81	8.78
	.95	9.63	8.78	7.85	7.21	6.69	6.23	5.86	5.51	5.20	10.17
	.975	11.17	9.81	8.52	7.79	7.22	6.70	6.27	5.92	5.56	11.52
	.99	13.58	10.95	9.37	8.50	7.85	7.21	6.75	6.33	5.98	13.74
2	.90	11.02	10.48	9.61	8.99	8.50	8.06	7.66	7.32	7.01	11.69
	.95	12.89	11.60	10.46	9.71	9.12	8.65	8.19	7.79	7.46	13.27
	.975	14.53	12.64	11.20	10.29	9.69	9.10	8.64	8.18	7.80	14.69
	.99	16.64	13.78	12.06	11.00	10.28	9.65	9.11	8.66	8.22	16.79
3	.90	13.43	12.73	11.76	11.04	10.49	10.02	9.59	9.21	8.86	14.05
	.95	15.37	13.84	12.64	11.83	11.15	10.61	10.14	9.71	9.32	15.80
	.975	17.17	14.91	13.44	12.49	11.75	11.13	10.62	10.14	9.72	17.36
	.99	19.25	16.27	14.48	13.40	12.56	11.80	11.22	10.67	10.19	19.38
4	.90	15.53	14.65	13.63	12.91	12.33	11.79	11.34	10.93	10.55	16.17
	.95	17.60	15.84	14.63	13.71	12.99	12.42	11.91	11.49	11.04	17.88
	.975	19.35	16.85	15.44	14.43	13.64	13.01	12.46	11.94	11.49	19.51
	.99	21.20	18.21	16.43	15.21	14.45	13.70	13.04	12.48	12.02	21.25
5	.90	17.42	16.45	15.44	14.69	14.05	13.51	13.02	12.59	12.18	17.94
	.95	19.50	17.60	16.40	15.52	14.79	14.19	13.63	13.16	12.70	19.74
	.975	21.47	18.75	17.26	16.13	15.40	14.75	14.19	13.66	13.17	21.57
	.99	23.99	20.18	18.19	17.09	16.14	15.34	14.81	14.26	13.72	24.00
6	.90	19.38	18.15	17.17	16.39	15.74	15.18	14.63	14.18	13.74	19.92
	.95	21.59	19.61	18.23	17.27	16.50	15.86	15.29	14.77	14.30	21.90
	.975	23.73	20.80	19.15	18.07	17.21	16.49	15.84	15.29	14.78	23.83
	.99	25.95	22.18	20.29	18.93	17.97	17.20	16.54	15.94	15.35	26.07
7	.90	21.23	19.93	18.75	17.98	17.28	16.69	16.16	15.69	15.24	21.79
	.95	23.50	21.30	19.83	18.91	18.10	17.43	16.83	16.28	15.79	23.77
	.975	25.23	22.54	20.85	19.68	18.79	18.03	17.38	16.79	16.31	25.46
	.99	28.01	24.07	21.89	20.68	19.68	18.81	18.10	17.49	16.96	28.02
8	.90	22.92	21.56	20.43	19.58	18.84	18.21	17.69	17.19	16.70	23.53
	.95	25.22	23.03	21.48	20.46	19.66	18.97	18.37	17.80	17.30	25.51
	.975	27.21	24.20	22.41	21.29	20.39	19.63	18.98	18.34	17.78	27.32
	.99	29.60	25.66	23.44	22.22	21.22	20.40	19.66	19.03	18.46	29.60
9	.90	24.75	23.15	21.98	21.12	20.37	19.72	19.13	18.58	18.09	25.19
	.95	27.08	24.55	23.16	22.08	21.22	20.49	19.90	19.29	18.79	27.28
	.975	29.13	25.92	24.14	22.97	21.98	21.28	20.59	19.98	19.39	29.20
	.99	31.66	27.42	25.13	24.01	23.06	22.18	21.35	20.63	19.94	31.72
10	.90	26.13	24.70	23.48	22.57	21.83	21.16	20.57	20.03	19.55	26.66
	.95	28.49	26.17	24.59	23.59	22.71	21.93	21.34	20.74	20.17	28.75
	.975	30.67	27.52	25.69	24.47	23.45	22.71	21.95	21.34	20.79	30.84
	.99	33.62	29.14	26.90	25.58	24.44	23.49	22.75	22.09	21.47	33.86

Table 2: Asymptotic Critical Values of the Sequential Test $F_T(\ell+1|\ell)$. The entries are the quantiles x such that $G_{q,\eta}(x)^{\ell+1} = \alpha$.

		ℓ									
q	α	0	1	2	3	4	5	6	7	8	9
1	.90	8.02	9.56	10.45	11.07	11.65	12.07	12.47	12.70	13.07	13.34
	.95	9.63	11.14	12.16	12.83	13.45	14.05	14.29	14.50	14.69	14.88
	.975	11.17	12.88	14.05	14.50	15.03	15.37	15.56	15.73	16.02	16.39
	.99	13.58	15.03	15.62	16.39	16.60	16.90	17.04	17.27	17.32	17.61
2	.90	11.02	12.79	13.72	14.45	14.90	15.35	15.81	16.12	16.44	16.58
	.95	12.89	14.50	15.42	16.16	16.61	17.02	17.27	17.55	17.76	17.97
	.975	14.53	16.19	17.02	17.55	17.98	18.15	18.46	18.74	18.98	19.22
	.99	16.64	17.98	18.66	19.22	20.03	20.87	20.97	21.19	21.43	21.74
3	.90	13.43	15.26	16.38	17.07	17.52	17.91	18.35	18.61	18.92	19.19
	.95	15.37	17.15	17.97	18.72	19.23	19.59	19.94	20.31	21.05	21.20
	.975	17.17	18.75	19.61	20.31	21.33	21.59	21.78	22.07	22.41	22.73
	.99	19.25	21.33	22.01	22.73	23.13	23.48	23.70	23.79	23.84	24.59
4	.90	15.53	17.54	18.55	19.30	19.80	20.15	20.48	20.73	20.94	21.10
	.95	17.60	19.33	20.22	20.75	21.15	21.55	21.90	22.27	22.63	22.83
	.975	19.35	20.76	21.60	22.27	22.84	23.44	23.74	24.14	24.36	24.54
	.99	21.20	22.84	24.04	24.54	24.96	25.36	25.51	25.58	25.63	25.88
5	.90	17.42	19.38	20.46	21.37	21.96	22.47	22.77	23.23	23.56	23.81
	.95	19.50	21.43	22.57	23.33	23.90	24.34	24.62	25.14	25.34	25.51
	.975	21.47	23.34	24.37	25.14	25.58	25.79	25.96	26.39	26.60	26.84
	.99	23.99	25.58	26.32	26.84	27.39	27.86	27.90	28.32	28.38	28.39
6	.90	19.38	21.51	22.81	23.64	24.19	24.59	24.86	25.27	25.53	25.87
	.95	21.59	23.72	24.66	25.29	25.89	26.36	26.84	27.10	27.26	27.40
	.975	23.73	25.41	26.37	27.10	27.42	28.02	28.39	28.75	29.13	29.44
	.99	25.95	27.42	28.60	29.44	30.18	30.52	30.64	30.99	31.25	31.33
7	.90	21.23	23.41	24.51	25.07	25.75	26.30	26.74	27.06	27.46	27.70
	.95	23.50	25.17	26.34	27.19	27.96	28.25	28.64	28.84	28.97	29.14
	.975	25.23	27.24	28.25	28.84	29.14	29.72	30.41	30.76	31.09	31.43
	.99	28.01	29.14	30.61	31.43	32.56	32.75	32.90	33.25	33.25	33.85
8	.90	22.92	25.15	26.38	27.09	27.77	28.15	28.61	28.90	29.19	29.49
	.95	25.22	27.18	28.21	28.99	29.54	30.05	30.45	30.79	31.29	31.75
	.975	27.21	29.01	30.09	30.79	31.80	32.50	32.81	32.86	33.20	33.60
	.99	29.60	31.80	32.84	33.60	34.23	34.57	34.75	35.01	35.50	35.65
9	.90	24.75	26.99	28.11	29.03	29.69	30.18	30.61	30.93	31.14	31.46
	.95	27.08	29.10	30.24	30.99	31.48	32.46	32.71	32.89	33.15	33.43
	.975	29.13	31.04	32.48	32.89	33.47	33.98	34.25	34.74	34.88	35.07
	.99	31.66	33.47	34.60	35.07	35.49	37.08	37.12	37.23	37.47	37.68
10	.90	26.13	28.40	29.68	30.62	31.25	31.81	32.37	32.78	33.09	33.53
	.95	28.49	30.65	31.90	32.83	33.57	34.27	34.53	35.01	35.33	35.65
	.975	30.67	32.87	34.27	35.01	35.86	36.32	36.65	36.90	37.15	37.41
	.99	33.62	35.86	36.68	37.41	38.20	38.70	38.91	39.09	39.11	39.12

Table 3: Empirical Results: U.S. Ex-Post Real Interest Rate

Specifications				
$z_t = \{1\}$	$q = 1$	$p = 0$	$h = 7$	M=5
Tests[1]				
Sup$F_T(1)$	Sup$F_T(2)$	Sup$F_T(3)$	Sup$F_T(4)$	Sup$F_T(5)$
58.53	44.16	53.76	51.88	44.76
Number of Breaks Selected[2]				
Sequential Procedure	2			
LWZ	2			
BIC	2			
Parameter Estimates with Two Breaks[3]				
$\hat{\delta}_1$	= 1.36 (.16)			
$\hat{\delta}_2$	= -1.80 (.50)			
$\hat{\delta}_3$	= 5.64 (.59)			
$\hat{T}_1$	= 72:3 (71:2–73:4)			
$\hat{T}_2$	= 80:3 (80:2–80:4)			

[1] The sup$F_T(k)$ tests and the reported standard errors and confidence intervals allow for the possibility of serial correlation in the disturbances. The heteroskedasticity and autocorrelation consistent covariance matrix is constructed following Andrews (1991) and Andrews and Monahan (1992) using a quadratic kernel with automatic bandwidth selection based on an AR(1) approximation. The residuals are pre-whitened using a VAR(1).

[2] We use a 1% size for the sequential test sup$F_T(\ell + 1|\ell)$.

[3] In parentheses are the standard errors (robust to serial correlation) for $\hat{\delta}_i$ $(i = 1, 2, 3)$ and the 95% confidence intervals for $\hat{T}_1$ and $\hat{T}_2$.

Table 4: Empirical Results: U.K. CPI Inflation Rate 1948-1987

Specifications				
$z_t = \{1, y_{t-1}\}$	$q = 2$	$p = 0$	$h = 5$	M=5
Tests				
Sup$F_T(1)$	Sup$F_T(2)$	Sup$F_T(3)$	Sup$F_T(4)$	Sup$F_T(5)$
5.34	12.69	12.74	10.76	8.58
Sup$F_T(2\|1)$	Sup$F_T(3\|2)$	Sup$F_T(4\|3)$	Sup$F_T(5\|4)$	Sup$F_T(6\|5)$
10.70	7.47	4.58	1.71	1.66
Number of Breaks Selected				
Sequential Procedure	0			
LWZ	1			
BIC	3			
Parameter Estimates with one Break				
$\hat{\delta}_{1,1}$	= .025 (.013)	$\hat{\delta}_{1,2}$	= .024 (.013)	
$\hat{\delta}_{2,1}$	= .274 (.316)	$\hat{\delta}_{2,2}$	= .739 (.125)	
$\hat{T}_1$	= 1967 (1961-1973)			
Parameter Estimates with Two Breaks				
$\hat{\delta}_{1,1}$ = .021 (.010)	$\hat{\delta}_{1,2}$ = .130 (.029)	$\hat{\delta}_{1,3}$ = .011 (.019)		
$\hat{\delta}_{2,1}$ = .488 (.220)	$\hat{\delta}_{2,2}$ = .115 (.206)	$\hat{\delta}_{2,3}$ = .633 (.223)		
$\hat{T}_1$	= 1973 (1972-1974)			
$\hat{T}_2$	= 1980 (1979-1981)			

Table 5: Empirical Results: Phillips Curve Equation

Specifications				
$y_t = \{\Delta w_t\}$	$z_t = \{1, \Delta p_{t-1}\, \Delta u_t, u_{t-1}\}$	$q = 4 \quad p = 0$	$h = 5$	M=5
Tests				
Sup$F_T(1)$	Sup$F_T(2)$	Sup$F_T(3)$	Sup$F_T(4)$	Sup$F_T(5)$
32.41	30.90	26.80	27.98	29.10
Sup$F_T(2\|1)$	Sup$F_T(3\|2)$	Sup$F_T(4\|3)$	Sup$F_T(5\|4)$	Sup$F_T(6\|5)$
9.13	6.34	2.58	2.58	2.40
Number of Breaks Selected				
Sequential Procedure	1			
LWZ	1			
BIC	2			
Parameter Estimates with one Break				
$\hat{\gamma}_{1,1}$	= .008 (.013)	$\hat{\gamma}_{1,2}$	= .220 (.040)	
$\hat{\gamma}_{2,1}$	= .452 (.248)	$\hat{\gamma}_{2,2}$	= -.174 (.236)	
$\hat{\gamma}_{3,1}$	= -.937 (1.337)	$\hat{\gamma}_{3,2}$	= .942 (.800)	
$\hat{\gamma}_{4,1}$	= 2.488 (.857)	$\hat{\gamma}_{4,2}$	= -1.167 (.269)	
$\hat{T}_1$	= 1973 (1972-1974)			
Parameter Estimates with Two Breaks				
$\hat{\gamma}_{1,1}$ = .061 (.029)	$\hat{\gamma}_{1,2}$ = .111 (.061)	$\hat{\gamma}_{1,3}$ = .218 (.058)		
$\hat{\gamma}_{2,1}$ = .289 (.258)	$\hat{\gamma}_{2,2}$ = 1.456 (.297)	$\hat{\gamma}_{2,3}$ = -.217 (.285)		
$\hat{\gamma}_{3,1}$ = -2.899 (1.48)	$\hat{\gamma}_{3,2}$ = -2.232 (2.14)	$\hat{\gamma}_{3,3}$ = .738 (.705)		
$\hat{\gamma}_{4,1}$ = -.961 (1.73)	$\hat{\gamma}_{4,2}$ = -3.292 (2.69)	$\hat{\gamma}_{4,3}$ = -1.098 (.397)		
$\hat{T}_1$	= 1966 (1965-1967)			
$\hat{T}_2$	= 1975 (1974-1976)			

Figure 1: U.S. Ex-Post Real Interest Rate, 1961:1-1986:3

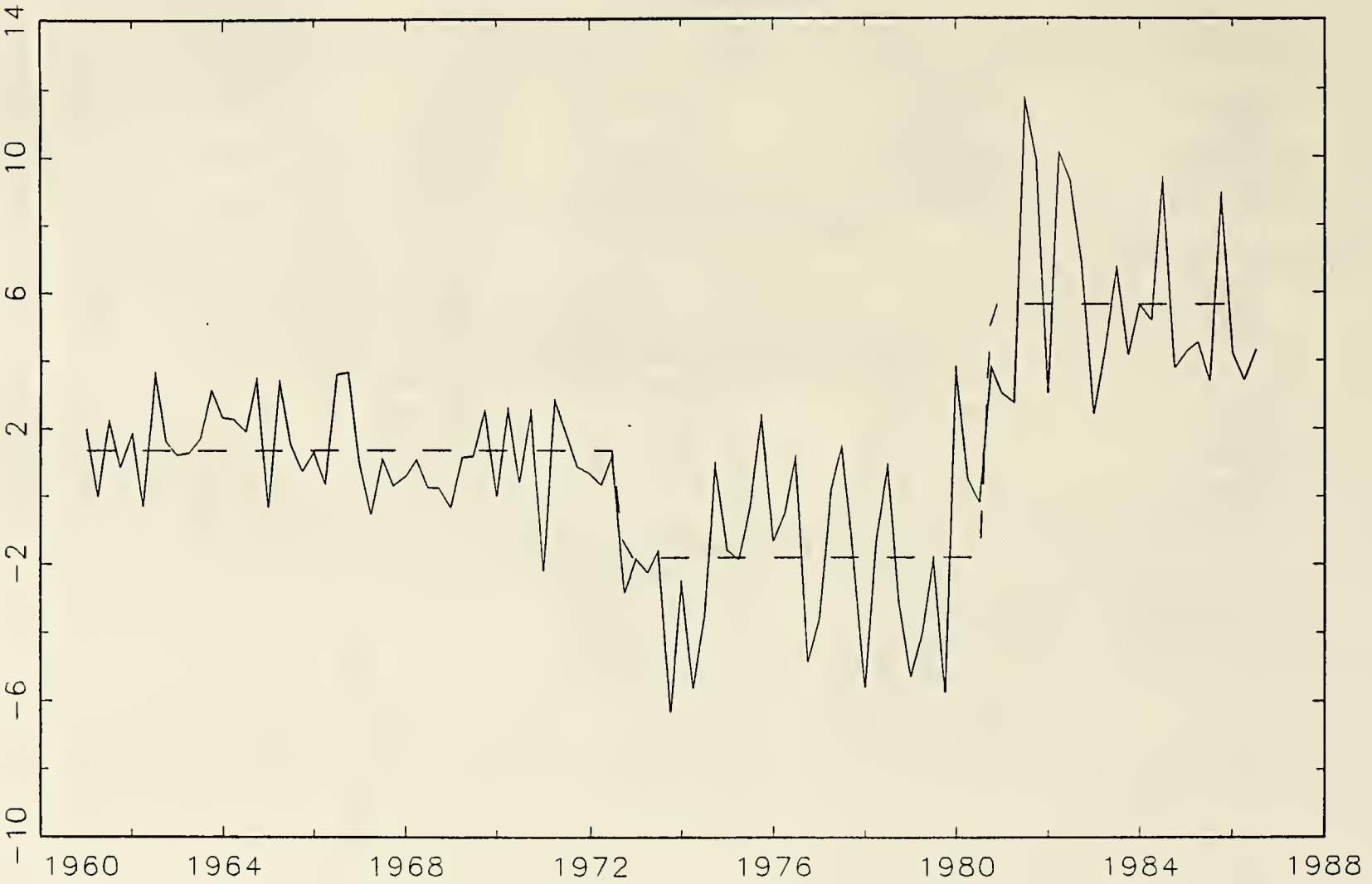

Figure 2: UK CPI Inflation Rate 1946–1987

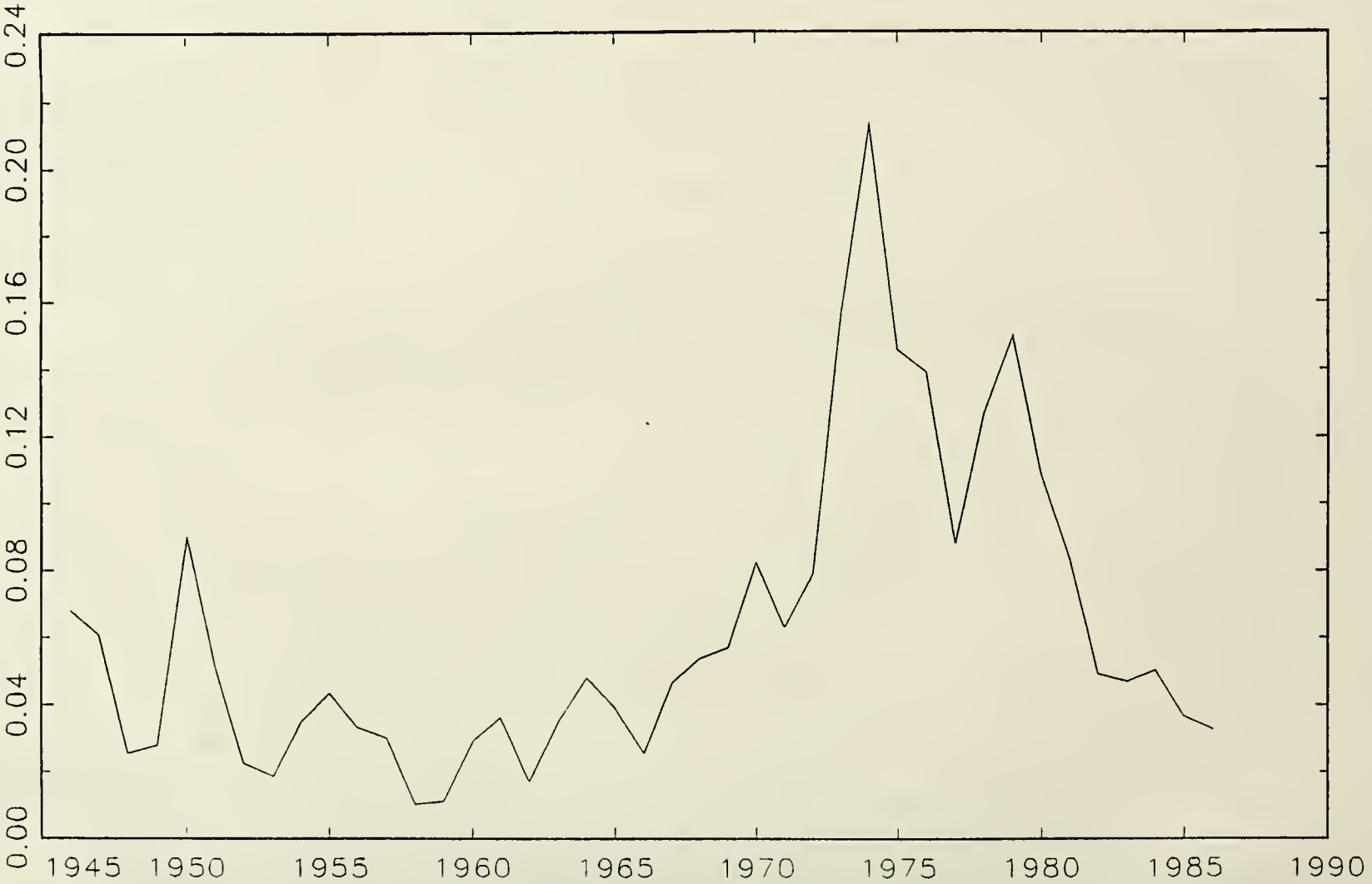

Figure 3: A particular configuration of (T_1, T_2, T_3) in the set $T_\epsilon(C)$ defined in the proof of Proposition 3