Estimating Restricted Structural Change Models*

Pierre Perron[†] **Zhongjun Qu**[‡]

Boston University **Boston University**

January 9, 2004; Revised: October 27, 2004

Abstract

This paper considers issues related to multiple structural changes, occurring at unknown dates, in the linear regression model when restrictions are imposed on the parameters. This includes, for example, the important special case where different non adjacent regimes are the same. The estimates are constructed as global minimizers of the restricted sum of squared residuals and we provide an extension of the algorithm discussed in Bai and Perron (2003) to efficiently compute them. We show that the estimates of the break dates have the same asymptotic properties with or without the restrictions imposed; that is, in large samples, there is no efficiency gain from imposing valid restrictions as far as the estimates of the break dates are concerned. Of course, efficiency gains occur for the other parameters of the model. Simulations show that in small samples, all parameters are more efficiently estimated using the restrictions. We also consider tests of the null hypothesis of no structural change. These are also more powerful when the restrictions are imposed. A Gauss code for all the procedures discussed in this paper is available from the authors.

JEL Classification Number: C20.

Keywords: change-point, segmented regressions, break dates, hypothesis testing, model selection.

*Pierre Perron acknowledges financial support from the National Science Foundation under Grant SES-0078492.

[†]Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (perron@bu.edu).

[‡]Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (qu@bu.edu).

1 Introduction

Both the statistics and econometrics literature contain a vast amount of work on issues related to structural changes with unknown break dates, most of it specifically designed for the case of a single change. However, the problem of multiple structural changes has received considerably less attention. Recently, Bai and Perron (1998, 2003) provided a comprehensive treatment of various issues in the context of multiple structural change models: consistency of estimates of the break dates, tests for structural changes, confidence intervals for the break dates, methods to select the number of breaks and efficient algorithms to compute the estimates. Related contributions include Hawkins (1976) who presents a comprehensive treatment of estimation based on a dynamic programming algorithm. Andrews, Lee and Ploberger (1996) consider optimal tests in the normal linear model with known variance. Garcia and Perron (1996) study the sup Wald test for two changes in a dynamic time series, and Liu, Wu and Zidek (1997) consider multiple structural changes in the context of a general threshold model and propose an information criterion to select the number of changes.

In practice, it is often the case that prior information can be formulated as restrictions on the coefficients of the model. A leading example is the case of a partial structural change model, also analyzed in Bai and Perron (1998, 2003), for which some coefficients are not allowed to change across regimes. But more general forms of restrictions are also common. For example, a model which specifies a specific number of states. With two states, the coefficients are the same in odd and even regimes. Again it could be the case that the value of the parameters in a specific segment is known.

The purpose of this paper is to extend results available for multiple structural change models to the case where arbitrary linear restrictions on the coefficients are available a priori. We consider both theoretical issues related to the limiting distribution of estimates and test statistics, as well as practical issues related to the construction of efficient methods to compute the estimates without having to go through an extensive high dimensional grid search. Our approach builds on the work of Bai and Perron (1998, 2003) and the estimates are based on a global least squares procedure. The extension is, however, nontrivial since the method of localization breaks down with restrictions across regimes. The difficulty is solved by exploring the relationship between the restricted and unrestricted global sum of squared residuals.

We derive the consistency, rate of convergence and asymptotic distribution of the estimates, under both fixed and shrinking shifts. We show that the estimates of the break dates have the same asymptotic properties with or without the restrictions imposed; that is, in large samples, there are no efficiency gains from imposing valid restrictions as far as the estimates of the break dates are concerned. Of course, efficiency gains occur for the other parameters of the model. Simulations show that in small samples, all parameters are more efficiently estimated using the restrictions.

To obtain estimates, we extend the algorithm based on a dynamic programming approach discussed in Bai and Perron (2003). Though, it is only guaranteed to reach a local minimum, departures from the global minimum are rare and occur in cases where the breaks are small to the extent that hypothesis testing would not favor a model with structural change. Hence, the algorithm is reliable in the majority of practical cases.

Concerning hypothesis testing, we extend the popular $\sup F$ type test, i.e., the F test for the null hypothesis that all coefficients are the same across regimes evaluated at the break dates that maximize its value. We derive its limiting distribution. Unfortunately it depends on the nature of the restrictions involved to the extent that the limit distribution is case dependent and we cannot tabulate critical values valid generally. We therefore describe a method to simulate them. A program to implement the procedure involved for arbitrary types of restrictions is made available on the web. A small simulation experiment demonstrates that substantial power gains can be achieved imposing valid restrictions.

The paper is structured as follows. Section 2 discusses the model and assumptions. Section 3 studies the consistency and rate of convergence of the estimates, for both fixed and shrinking magnitude of shifts and Section 4 studies the limiting distribution of the estimates of the break dates for both cases. Section 5 presents a computationally efficient procedure to estimate a restricted structural change model. Section 6 presents simulations to show evidence about the increased efficiency of the estimates. Section 7 describes the test of the null hypothesis of no structural change versus an arbitrary number of changes, derives its limiting distribution and a method to simulate asymptotic critical values, and presents evidence that substantial power gains can be obtained over tests that do no take into account the restrictions. Section 8 offers brief concluding remarks and an appendix contains technical derivations of some results.

2 The model and the assumptions

Consider a general multiple linear regression model with m breaks or $m+1$ regimes. There are T observations and m is assumed known. The break dates occur at $\{T_1, ..., T_m\}$. Let $y = (y_1, ..., y_T)'$ be the dependent variable and Z a T by q matrix of regressors. Define

$\bar{Z} = diag(Z_1, ..., Z_{m+1})$, a T by $(m+1)q$ matrix with $Z_i = (z_{T_{i-1}+1}, ..., z_{T_i})'$ for $i = 1, ..., m+1$, with the convention that $T_0 = 1$ and $T_{m+1} = T$ (each matrix Z_i is a subset of the regressor matrix Z corresponding to regime i). The matrix $\bar{Z}$ is a diagonal partition of Z, the partition being taken with respect to the set of break points $\{T_1, ..., T_m\}$. The vector $u = (u_1, ..., u_T)'$ is the set of disturbances and $\delta = (\delta_1', ..., \delta_{m+1}')'$ is the $(m+1)q$ vector of coefficients.

Our objective is to consider the general pure structural change model with restrictions on the coefficients, i.e.

$$y = \bar{Z}\delta + u \tag{1}$$

where

$$R\delta = r \tag{2}$$

with R a k by $(m+1)q$ matrix with rank k and r a k dimensional vector of constants.

The goal here is to estimate the unknown coefficients whose true values are denoted with a 0 superscript, i.e. $(\delta_1^0, ..., \delta_{m+1}^0, T_1^0, ..., T_m^0)$ using the observables (y, Z) and the restrictions on the coefficients. The method is based on least square principle. More specifically, the estimated break dates are

$$(\widetilde{T}_1, ..., \widetilde{T}_m) = \arg \min_{T_1, ..., T_m} SSR_T^R(T_1, ..., T_m) \tag{3}$$

where $SSR_T^R(T_1, ..., T_m)$ is the sum of square residuals from the restricted OLS regression evaluated at the partition $\{T_1, ..., T_m\}$. We impose the following assumptions on the data, the errors and the break dates.

• Assumption A1: Let $\bar{Z}^0 = diag(Z_1^0, ..., Z_{m+1}^0)$ be the diagonal partition of Z at the true break dates $\{T_1^0, ..., T_m^0\}$. For each $i = 1, ..., m+1$, let $l_i = (T_{i+1}^0 - T_i^0)$, then$(1/l_i)\sum_{t=T_i^0+1}^{T_i^0+[l_i v]} z_t z_t' \to_p Q(v)$ a non-random positive definite matrix uniformly in $v \in [0,1]$.

• Assumption A2: There exists an $l_0 > 0$ such that for all $l > l_0$, the minimum eigenvalues of $(1/l)\sum_{t=T_i^0+1}^{T_i^0+l} z_t z_t'$ and of $(1/l)\sum_{t=T_i^0-l}^{T_i^0} z_t z_t'$ are bounded away from zero $(i = 1, ..., m)$.

• Assumption A3: The matrix $\sum_{t=k}^{l} z_t z_t'$ is invertible for $l - k \geq \epsilon T$ for some $\epsilon > 0$.

• Assumption A4: Let the L_r-norm of a random matrix X be defined by $\|X\|_r = (\sum_i \sum_j E|X_{ij}|^r)^{1/r}$ for $r \geq 1$. (Note that $\|X\|$ is the usual matrix norm or the Euclidean norm of a vector.) With $\{\mathcal{F}_i : i = 1, 2, ..\}$ a sequence of increasing σ-fields, we assume that $\{z_i u_i, \mathcal{F}_i\}$ forms a L^r-mixingale sequence with $r = 2 + \delta$ for some $\delta > 0$. That is, there exist nonnegative constants $\{c_i : i \geq 1\}$ and $\{\psi_j : j \geq 0\}$ such that $\psi_j \downarrow 0$ as $j \to \infty$ and for all $i \geq 1$ and $j \geq 0$, we have: (a) $\|E(z_i u_i|\mathcal{F}_{i-j})\|_r \leq c_i \psi_j$, (b) $\|z_i u_i - E(z_i u_i|\mathcal{F}_{i+j})\|_r \leq c_i \psi_{j+1}$. Also assume (c) $\max_i c_i \leq K < \infty$, (d) $\sum_{j=0}^{\infty} j^{1+k}\psi_j < \infty$, (e) $\|z_i\|_{2r} < M < \infty$ and $\|u_i\|_{2r} < N < \infty$ for some $K, M, N > 0$.

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

• Assumption A6: The minimization problem defined by (3) is taken over all possible partitions such that $T_i - T_{i-1} \geq \epsilon T$ for some $\epsilon > 0$.

Assumption A1 basically rules out unit root regressors; otherwise the particular scaling used is not important and could be relaxed at the expense of substantial technical complications [1]. Assumption A2 imposes restrictions on the regressors in a local neighborhood of the break points. They ensure that there is no local collinearity problem so the break points can be identified. Assumption A3 is a standard invertibility requirement to have well defined estimates. Assumption A4 imposes mild restrictions on the vector $z_t u_t$. They permit a wide class of potential correlation and heterogeneity (including conditional heteroskedasticity) and also allow lagged dependent variables. Finally, Assumption A5 is a standard requirement to have asymptotically distinct break dates and A6 requires that the search for breaks precludes candidates which are too close. This, however, is not constraining in practice since ϵ can be chosen arbitrarily small.

The set of assumptions used differ from those used in Bai and Perron (1998). The main difference is that when heterogeneity and serial correlation is allowed in the errors, the regressors z_t are not assumed to be independent of the errors u_t at all leads and lags. This permits a wider class of models, which has considerable practical advantages. The cost is the need to introduce slightly stronger technical conditions. In particular A1 and A6 are stronger than the requirements in Bai and Perron (1998), though A6 was also imposed there when lagged dependent variables where permitted as regressors. Since we shall make frequent references to the results in Bai and Perron (1998), in the appendix we show that their results go through with this revised set of assumptions.

3 Consistency and rate of convergence

In this section, we discuss the consistency and rate of convergence for models with fixed and shrinking shift magnitudes. As will become clear later considering shrinking shifts is important in that it allows us to obtain a limit distribution that is not dependent upon the true distribution of the errors. As matter of notation, we use " $\rightarrow_p$ " to denote convergence in probability, " $\rightarrow^d$ " converge in distribution, and " $\Rightarrow$ " weak convergence in the space

[1] For example, allowing for trending data would imply a different scaling but it would not change results related to the consistency and rate of convergence of the estimates of the break dates or the fact that the limit distribution of the other estimates is the same as in the known break date case. For the limit distribution of the estimates of the break dates and the test statistics, this assumption matters and will be strenghtenned.

$D[0,1]$ under the Skorohod topology. Also, $\widetilde{\lambda} = (\widetilde{\lambda}_1, ..., \widetilde{\lambda}_m) = (\widetilde{T}_1/T, ..., \widetilde{T}_m/T)$ where $\{\widetilde{T}_1, ..., \widetilde{T}_m\}$ is the break date partition that corresponds to the minimal value of the restricted Sum of Squared Residuals (SSR). Similarly $\{\hat{T}_1, ..., \hat{T}_m\}$ is the break date partition that corresponds to the minimal value of the unrestricted SSR. The corresponding true values are denoted $\lambda^0 = (\lambda_1^0, ..., \lambda_m^0) = \{T_1^0/T, ..., T_m^0/T\}$. We first consider the case with shifts of fixed magnitudes.

3.1 Shifts of fixed magnitudes

Proposition 1 *Under A1-A6, we have for shifts in the coefficients that are of fixed magnitudes independent of T: a) $\widetilde{\lambda}_k \to_p \lambda_k^0$, $(k = 1, ..., m)$; b) for every $\varepsilon > 0$, there exists a $C < \infty$, such that for large T, $P(|T(\widetilde{\lambda}_k - \lambda_k^0)| > C) < \varepsilon$ for every $k = 1, ..., m$.*

The proof is in the appendix. It is well known that in the unrestricted case, the estimated break fractions converge to their true values at rate T. With restrictions supposed to be valid, we indeed have more information, hence, we should expect the convergence rate to be no smaller than T. What is interesting is that the rate of convergence is not improved. It suggests that the estimates of the break dates may be not sensitive to additional information. This will be confirmed once the asymptotic distribution of the break dates is derived.

3.2 Shifts of shrinking magnitudes

For models with shifts of shrinking magnitudes, we need the following assumption:

• Assumption A7: Let $\Delta_{T,i} = \delta_{T,i+1}^0 - \delta_{T,i}^0$. Assume $\Delta_{T,i} = v_T \Delta_i$, for some Δ_i independent of T where $v_T > 0$ is a scalar satisfying $v_T \to 0$ and $T^{1/2-\eta} v_T \to \infty$ for some $\eta \in (0, 1/2)$.

Proposition 2 *Under assumption A1-A7, we have for $k = 1, ..., m$: a) $\widetilde{\lambda}_k \to_p \lambda_k^0$; b) for every $\varepsilon > 0$, there exists a $C < \infty$, such that for all large T, $P(|T v_T^2(\widetilde{\lambda}_k - \lambda_k^0)| > C) < \varepsilon$.*

This proposition, proved in the appendix, shows that here also the results are the same as in the case of an unrestricted model.

3.3 The limiting distribution of the parameter estimates

The rate T convergence of the estimated break fractions is enough to establish the $\sqrt{T}$ consistency of the estimated parameters. We now consider their limiting distribution and show, among other things, that it is the same as if the break dates were known.

Proposition 3 *Let $\widetilde{\delta}$ be the estimated parameters under restricted OLS. Then, under A1-A6, we have $\sqrt{T}(\widetilde{\delta}-\delta^0) \rightarrow^d N(0,V)$ with V the variance of the restricted OLS estimator evaluated under the true partition:*

$$V = \Phi + Q^{-1}R'D^{-1}R\Phi R'D^{-1}RQ^{-1} - \Phi R'D^{-1}RQ^{-1} - Q^{-1}R'D^{-1}R\Phi$$

where $Q = p\lim_{T\to\infty} T^{-1}\bar{Z}^{0\prime}\bar{Z}^0$, $D = RQ^{-1}R'$, $\Phi = Q^{-1}(p\lim_{T\to\infty} T^{-1}\bar{Z}^{0\prime}\Omega\bar{Z}^0)Q^{-1}$ and $\Omega = E(uu')$.

Remark 1 *As in standard restricted OLS regressions, there is an efficiency gain over the unrestricted OLS in the sense that $(\Phi - V)$ is a positive semi-definite matrix.*

Remark 2 *In the case with uncorrelated and homoskedastic errors, the variance-covariance matrix simplifies to $V = \sigma^2 Q^{-1} - \sigma^2 Q^{-1}R'D^{-1}RQ^{-1}$, which can be consistently estimated using a consistent estimate of σ^2. When serial correlation and/or heteroskedasticity is present, a consistent estimate of $p\lim_{T\to\infty} T^{-1}\bar{Z}^{0\prime}\Omega\bar{Z}^0$ can be constructed using the method of Andrews (1991), among others.*

4 The limiting distribution of the estimated break dates

First we establish the following Lemma:

Lemma 1 *Let $SSR_T^U(T_1,...,T_m)$ be the Sum of Squared Residuals obtained using the partition $\{T_1,...,T_m\}$ without imposing the restrictions, i.e. the unrestricted SSR. Define*

$$\begin{aligned}\widetilde{T}(C) &\equiv (\widetilde{T}_1(C),...,\widetilde{T}_m(C)) \\ &= \arg\min_{|T_k-T_k^0|\le C\|v_T\|^{-2},\ k=1,...,m} SSR_T^R(T_1,...,T_m) - S_T^R(T_1^0,...,T_m^0)\end{aligned}$$

with C a positive number. Under Assumptions A1-A6 and with either shifts of fixed magnitudes or with shrinking shifts satisfying A7, we have

$$\begin{aligned}\widetilde{T}_i(C) &= \arg\min_{|T_k-T_k^0|\le C\|v_T\|^{-2},\ k=1,...,m} \left[SSR_T^U(T_1,...,T_m) - SSR_T^U(T_1^0,...,T_m^0)\right] + o_p(1) \\ &= \arg\min_{|T_i-T_i^0|\le C\|v_T\|^{-2}} G(T_i)\end{aligned}$$

where

$$G(T_i) = -(\delta_i^0-\delta_{i+1}^0)'(\sum_{t=T_i+1}^{T_i^0} z_t z_t')(\delta_i^0-\delta_{i+1}^0) - 2(\delta_i^0-\delta_{i+1}^0)'\sum_{t=T_i+1}^{T_i^0} z_t u_t + o_p(1)$$

for $T_i - T_i^0 < 0$, *and*

$$G(T_i) = -(\delta_i^0 - \delta_{i+1}^0)'(\sum_{t=T_i^0+1}^{T_i} z_t z_t')(\delta_i^0 - \delta_{i+1}^0) + 2(\delta_i^0 - \delta_{i+1}^0)' \sum_{t=T_i^0+1}^{T_i} z_t u_t + o_p(1)$$

for $T_i - T_i^0 \geq 0$.

The importance of this result is that once the break dates are restricted in the set $\{T_k : |T_k - T_k^0| \leq C \|v_T\|^{-2}, k = 1, ..., m\}$ the $\sqrt{T}$ consistency of $\widetilde{\delta}$ is guaranteed and the effect of the restrictions on the estimates of the break dates becomes negligible asymptotically. Hence the restricted minimization problem reduces to the unrestricted minimization problem and the limit distribution of the break dates are the same in both cases. These are stated in the following results for the fixed and shrinking shifts scenarios.

4.1 Fixed magnitude of shifts

As is well known, the limiting distribution of the estimated break dates under fixed magnitudes of shifts depends on the exact distribution of the regressors and the errors in a local neighborhood of the break points. Hence, the result requires the following assumption.

• Assumption A8: The process $\{z_t, u_t\}$ is strictly stationary.

Let $W^*(0) = 0$, $W^*(m) = W_1^*(m)$ for $m < 0$ and $W^*(m) = W_2^*(m)$ for $m > 0$, with

$$\begin{aligned} W_1^*(m) &= -(\delta_i^0 - \delta_{i+1}^0)'(\sum_{t=m+1}^{0} z_t z_t')(\delta_i^0 - \delta_{i+1}^0) - 2(\delta_i^0 - \delta_{i+1}^0)' \sum_{t=m+1}^{0} z_t u_t \\ W_2^*(m) &= -(\delta_i^0 - \delta_{i+1}^0)'(\sum_{t=1}^{m} z_t z_t')(\delta_i^0 - \delta_{i+1}^0) + 2(\delta_i^0 - \delta_{i+1}^0)' \sum_{t=1}^{m} z_t u_t \end{aligned}$$

Proposition 4 *Under assumption A1-A6 and A8, and assuming that* $(\delta_i^{0'} z_t)^2 \pm \delta_i^{0'} z_t' u_t$ *has a continuous distribution,* $\widetilde{T}_i - T_i^0 \to^d \arg\max_m W^*(m)$.

Using Lemma 1 above, the rest of the proof is the same as the proof in Bai (1997) and is thus omitted. What is important about this proposition is that the limiting distribution of the estimated break dates is invariant to the linear restrictions on the coefficients.

4.2 Shrinking magnitude of shift

The main reason to consider a framework whereby the shifts are decreasing in magnitude as the sample size increases is to obtain a limiting distribution which is independent of the exact distribution of the regressors and the errors in a local neighborhood of the break dates. The ensuing distribution is designed to be a good approximation in finite samples for small shifts but simulation results have shown that it remains adequate for moderate to large shifts as well. To state the result, we make the following additional assumption:

- Assumption A9: Let $\Delta T_i^0 = T_i^0 - T_{i-1}^0$; for $i = 1, ..., m$, as $\Delta T_i^0 \to \infty$, uniformly in $s \in [0, 1]$:

1. $(\Delta T_i^0)^{-1} \sum_{t=T_{i-1}^0+1}^{T_{i-1}^0+[s\Delta T_i^0]} z_t z_t' \to_p sQ_i$,

2. $(\Delta T_i^0)^{-1} \sum_{t=T_{i-1}^0+1}^{T_{i-1}^0+[s\Delta T_i^0]} u_t^2 \to_p s\sigma_i^2$,

3. $(\Delta T_i^0)^{-1} \sum_{t=T_{i-1}^0+1}^{T_{i-1}^0+[s\Delta T_i^0]} \sum_{r=T_{i-1}^0+1}^{T_{i-1}^0+[s\Delta T_i^0]} E(z_r z_t' u_r u_t) \to_p s\Omega_i$;

4. $(\Delta T_i^0)^{-1/2} \sum_{t=T_{i-1}^0+1}^{T_{i-1}^0+[s\Delta T_i^0]} z_t u_t \Rightarrow B_i(s)$ with $B_i(s)$ a multivariate Gaussian Process on $[0, 1]$ with mean zero and covariance $E[B_i(s)B_i(u)'] = \min\{s, u\}\,\Omega_i$.

We then have the following result:

Proposition 5 *Under Assumptions A1-A7 and A9 and further assuming that $(\delta_i^{0'} z_t)^2 \pm \delta_i^{0'} z_t' u_t$ has a continuous distribution, we have, for $i = 1, ..., m$:*

$$\frac{(\Delta_i' Q_i \Delta_i)^2}{\Delta_i' \Omega_i \Delta_i}(\widetilde{T}_i - T_i^0) \to^d \arg\max_s V^{(i)}(s)$$

where $V^{(i)}(s) = W_1^{(i)}(-s) - |s|/2$ if $s \le 0$, $V^{(i)}(s) = \sqrt{\xi_i}(\phi_{i,2}/\phi_{i,1})W_2^{(i)}(s) - \xi_i |s|/2$ if $s > 0$, with $\xi_i = \Delta_i' Q_{i+1} \Delta_i / \Delta_i' Q_i \Delta_i$, $\phi_{i,1}^2 = \Delta_i' \Omega_i \Delta_i / \Delta_i' Q_i \Delta_i$ and $\phi_{i,2}^2 = \Delta_i' \Omega_{i+1} \Delta_i / \Delta_i' Q_{i+1} \Delta_i$.

Again, with Lemma 1, the rest of the proof is similar to that of Bai (1997). The proposition says that the limiting distribution of the estimated break dates is also invariant to linear restrictions on the coefficients in the shrinking shifts framework. Hence, in large samples, imposing restrictions only improves the efficiency of the estimated regression coefficients. However, the limit distribution is a function of the other parameters of the model which are more precisely estimated with restrictions even in large samples. Hence, imposing the restrictions should allow better inference about the break dates as well.

5 A computationally efficient procedure to estimate a restricted structural break model.

The computation associated with estimating a multiple structural breaks model is not a trivial issue. A standard grid search imply least-squares computations of order $O(T^m)$ which quickly becomes prohibitive. Bai and Perron (2003) describes an efficient estimation procedure based on a dynamic programming algorithm which involves at most least-squares computations of order $O(T^2)$ for any number of breaks (an earlier comprehensive exposition is contained in Hawkins, 1976). We will extend their procedure to provide an estimation method for multiple structural change models with restrictions. Note that a partial structural change model is a special case of the more general structural change model with restrictions (since in this case the restrictions are simply that some coefficients are the same in all regimes). Bai and Perron (2003) considered an iterative procedure to generate the estimates. Our method is equivalent to theirs in this special case.

5.1 The dynamic programming algorithm

In is useful to review the basics of the dynamic programming algorithm discussed in Bai and Perron (2003) given that we make extensive use of it in devising our estimation method. Let $SSR(\{T_{r,n}\})$ be the sum of square residuals associated with the optimal partition containing r breaks using the first n observations. Let $SSR(j+1,n)$ be the SSR obtained by applying OLS to a segment that starts at $j+1$ and ends at date n. Also, let $h = \epsilon T$ be the minimal permissible length of a segment. The optimization procedure is based on solving the following recursive problem

$$SSR(\{T_{r,n}\}) = \min_{rh \leq j \leq n-h} [SSR(\{T_{r-1,j}\}) + SSR(j+1,n)]$$

More specifically, the algorithm is as follows:

1. Compute and save $SSR(i,j)$ for pairs i,j satisfying $i - j \geq h$. Since the number of segments which satisfy this requirement is at most $O_p(T^2)$, this involves computations at most of order $O(T^2)$. The matrix inversions involved is only of order $O(T)$ using recursive residuals to compute the the sum of squared residuals.

2. Compute and store $SSR(\{T_{1,n}\})$ for $2h \leq n \leq T-(m-1)h$ by solving the following problem:

$$SSR(\{T_{1,n}\}) = \min_{h \leq j \leq n-h} [SSR(1,j) + SSR(j+1,n)]$$

3. Sequentially compute and store $SSR(\{T_{r,n}\})$ for $r = 2, ..., m-1$. For each r, n ranges from $(r+1)h$ to $T-(m-r)h$.

4. The estimates of the break dates are then obtained by solving

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

This method cannot be applied directly to models with restrictions since the sum of squared residuals for a segment cannot be computed independently of other segments. Our extension to tackle this problem is discussed next.

5.2 The recursive procedure with restrictions

The estimation procedure proposed is specified by the following algorithm:

1. First use the dynamic programming algorithm described above to estimate an unrestricted model. Store the estimated break dates.

2. Estimate the coefficients using the restrictions conditional on the break dates obtained in part (1). Denote the estimates as $\widetilde{\delta}_k$ $(k = 1, ..., m+1)$.

3. Compute and store the restricted sums of squared residuals $RSSR(\{T_{1,n}\})$, for $2h \leq n \leq T-(m-1)h$. This is done by the following recursive problem

$$RSSR(\{T_{1,n}\}) = \min_{h \leq j \leq n-h} \left[RSSR^1(1,j) + RSSR^2(j+1,n)\right]$$

where $RSSR^1(1,j) = \sum_{t=1}^{j}(y_t - z_t\widetilde{\delta}_1)^2$ and $RSSR^2(j+1,n) = \sum_{t=j+1}^{n}(y_t - z_t\widetilde{\delta}_2)^2$. Then, sequentially compute and store $RSSR(\{T_{r,n}\})$ for $r = 2, ..., m-1$, with n ranging from $(r+1)h$ to $T-(m-r)h$. This is done by solving the recursive problem

$$RSSR(\{T_{r,n}\}) = \min_{rh \leq j \leq n-h} \left[RSSR(\{T_{r-1,j}\}) + RSSR^{r+1}(j+1,n)\right]$$

with $RSSR^{r+1}(j+1,n) = \sum_{t=j+1}^{n}(y_t - z_t\widetilde{\delta}_{r+1})^2$. Finally compute

$$RSSR(\{T_{m,T}\}) = \min_{mh \leq j \leq T-h} \left[RSSR(\{T_{m-1,j}\}) + RSSR^{m+1}(j+1,T)\right]$$

Store the estimated break dates.

4. Repeat steps 2 and 3 until convergence.

The main computation involved in this procedure is in step 1. The others involve marginal computational costs. Experiments with real and simulated data showed that convergence is rapid, seldom requiring more than 3 iterations. So the procedure is computationally efficient.

While this algorithm quickly reaches a local minimum, it is not guaranteed to reach the global minimum of the restricted sum of squared residuals because the iterative scheme makes the overall optimization problem non-linear. This is a common problem for which there is no full proof solution. Simulations have shown that this is indeed the case. The difference is, however, important only when the estimation from step 1 is very imprecise, the restrictions do not help much to estimate the coefficients and the trimming defining the minimal lenght of a segment is small. In such cases, care must be exercised in trying to start the algorithm with values different from those given by estimating the model without restrictions. Note also that the chances of obtaining estimates that do not correspond to the global minimizers of the restricted sum of squared residuals occur when shifts are small, in which case hypothesis testing would likely not indicate evidence in favor of a structural change model. Furtermore, in most cases for which the algorithm did not yield estimates corresponding to global minimizers, these estimates were very far from the true values and most often on the boundaries of the permissible region, e.g., at the very beginning or end of the sample. Hence, most of the cases for which such a problem occurs could be detected by the practitioner. The reader is referred to Qu (2005) for more details.

6 Finite sample properties of the estimates.

In this section we present simulation results to assess what are the efficiency gains from estimating a restricted model for the regression coefficients and the break dates. All results are based on 500 replications. The data generating process we use is a model with two regimes and three breaks. The sample size is 120 and each regime has the same length. More specifically,

$$y_t = \delta_t + e_t$$

with

$$\delta_t = \left\{ \begin{array}{cc} 0.5 & \text{for } t \in \left[1, \frac{T}{4}\right] \cup \left[\frac{T}{2}+1, \frac{3T}{4}\right] \\ 0 & \text{for } t \in \left[\frac{T}{4}+1, \frac{T}{2}\right] \cup \left[\frac{3T}{4}+1, T\right] \end{array} \right\}$$

and $e_t \sim i.i.d.\ N(0,1)$. The unconstrained regression model is

$$y_t = \delta_1 1(t \leq T/4) + \delta_2 1(T/4 < t \leq T/2) + \delta_3 1(T/2 < t \leq 3T/4) + \delta_4 1(3T/4 < t \leq T) + e_t$$

with $1(\cdot)$, the indicator function. Hence, the parameter vector is $\delta = (\delta_1, \delta_2, \delta_3, \delta_4)$. The first set of valid restrictions considered is specified by:

$$R = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}, \quad r = 0 \tag{4}$$

When constructing the estimates, the minimization problems are constrained so that each segment has at least 5 observations (see Bai and Perron, 2001 and 2003, on the importance of the minimal length of a segment). Estimates of the break dates are constructed as global minimizers of the unrestricted SSR obtained using the method of Bai and Perron (2003), and as the global minimizers of the restricted SSR. The cumulative distribution function of the estimates of the break dates $\{T_1, T_3, T_3\}$ are presented in Figure 1. The results clearly shows that the two methods lead to similar distributions for the break dates. However, they also show that the variability of the estimates can be reduced using the restricted model, contrary to what is suggested by the asymptotic results. Indeed, the unrestricted estimates show a fatter tail on the right for the first break, on both sides for the second and on the left for the third break. Hence, in small samples with a relatively large number of breaks, efficiency gains can be obtained for the break dates using restrictions. The probability density functions of the unrestricted estimates of the parameters $(\hat{\delta}_1, \hat{\delta}_2, \hat{\delta}_3, \hat{\delta}_4)$ and the restricted estimates $(\widetilde{\delta}_1 = \widetilde{\delta}_3 = \widetilde{\delta}_a, \widetilde{\delta}_2 = \widetilde{\delta}_4 = \widetilde{\delta}_b)$ are presented in Figure 2. Here, as expected from the asymptotic results, the restricted estimates are clearly more efficient.

The second set of valid restrictions considered are stronger and specified by:

$$R = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad r = 0. \tag{5}$$

The cumulative distribution function of the restricted and unrestricted estimates of the break dates $\{T_1, T_3, T_3\}$ are presented in Figure 3. In this case, since the restrictions are more informative, the restricted estimates of the break dates are substantially more precise (less variance) than the unrestricted estimates. Hence, important efficiency gains can be obtained for the break dates when the restrictions are informative. Figure 4 presents the probability density function of the estimates of the parameters δ. These show the restricted estimates to be far more precise than the unrestricted ones.

7 Testing the null hypothesis of no structural change

We now consider testing the null hypothesis of no structural change versus an alternative hypothesis of, say, k changes imposing the restrictions on the parameters of the model. Following Quandt (1958), Andrews (1993), Bai and Perron (1998) and others, we consider a $\sup F$ type test. Let $(T_1, ..., T_k)$ be a partition such that $T_i = [T\lambda_i]$ $(i = 1, ..., k)$. Let H be the conventional k by $q(k+1)$ matrix such that $(H\delta)' = (\delta_1' - \delta_2', ..., \delta_k' - \delta_{k+1}')$. Define

$$F_T(\lambda_1, ..., \lambda_k; q) = \widetilde{\delta}' H'(H\widetilde{V}(\widetilde{\delta})H')^{-} H\widetilde{\delta}, \tag{6}$$

where $\widetilde{\delta}$ is the estimate of δ obtained using the partition $\{\lambda_1, ..., \lambda_k\}$, and $\widetilde{V}(\widetilde{\delta})$ is an estimate of the variance covariance matrix of $\widetilde{\delta}$ that may be constructed to be robust to heteroskedasticity and serial correlation in the errors. As usual, for a matrix A, A^- denotes the generalized inverse of A. Such a generalized inverse is needed since in general the covariance matrix of $\widetilde{\delta}$ will be singular given that restrictions are imposed.

Computing the statistic $\sup \mathrm{F}_T(\lambda_1, ..., \lambda_k; q)$ where the supremum is taken over all possible partitions such that $|\lambda_i - \lambda_{i-1}| < \epsilon$ can be very cumbersome. We therefore consider an asymptotically equivalent test given by $F_T(\widetilde{\lambda}_1, ..., \widetilde{\lambda}_k; q)$ where $\widetilde{\lambda}_1$, ..., $\widetilde{\lambda}_k$ minimize the global restricted sum of squared residuals. This is equivalent to maximizing the F-test assuming spherical errors. It is asymptotically equivalent to, and yet much simpler to construct than, maximizing the F-test (6) since the estimated break dates are consistent even in the presence of serial correlation and heteroskedasticity. Note that the statistic still depends on a trimming parameter via the imposition of the minimal length h of a segment, namely $\epsilon = h/T$. To derive the limit distribution under the null hypothesis, we need the following assumption.

- Assumption A10: a) $T^{-1} \sum_{t=1}^{[Ts]} z_t z_t' \to_p sQ$, uniformly in $s \in [0, 1]$, for Q some positive definite matrix; b) $E(u_t^2) = \sigma^2$ for all t and $T^{-1/2} \sum_{t=1}^{[Ts]} z_t u_t \Rightarrow \sigma Q^{1/2} W_q(s)$ where $W_q(s)$ is a q vector of independent Wiener processes.

Assumption A10 rules out trending regressors. It also imposes an homogenous distribution throughout the sample. It is important to note that this assumption pertains only to the null hypothesis of no structural change in which case only one segment occurs. Under the alternative hypothesis of say k breaks, the distribution of the data and the errors can be regime dependent, i.e., σ^2 and Q can vary across the $k+1$ segments. The following result, proved in the appendix, summarizes the asymptotic distribution of the test under the null hypothesis.

Proposition 6 *With A1-A6 and under the null hypothesis of no structural change, in which* *case Assumption A10 applies, we have:*

$$F_T(\widetilde{\lambda}_1, ..., \widetilde{\lambda}_k; q) \Rightarrow \sup_{|\lambda_i - \lambda_{i-1}| > \epsilon} F(\lambda_1, ..., \lambda_k; q) \tag{7}$$

with

$$F(\lambda_1, ..., \lambda_k; q) = W^{*\prime} S[S'(\Lambda \otimes I_q)S]^{-1} S' H' [HS(S'(\Lambda \otimes I_q)S)^{-1} H' S']^{-} HS[S'(\Lambda \otimes I_q)S]^{-1} S' W^{*}. \tag{8}$$

and where the $q(k+1)$ vector W^* is defined by

$$W^* = [W_q(\lambda_1), W_q(\lambda_2) - W_q(\lambda_1), ..., W_q(1) - W_q(\lambda_k)]$$

The matrix Λ is a $(k+1)$ by $(k+1)$ matrix defined by $\Lambda = diag(\lambda_1, \lambda_2 - \lambda_1, ..., 1 - \lambda_k)$, and the $q(k+1)$ by d matrix S is defined by the following reparameterization of the restrictions: $\delta = S\theta + s$, with θ a d-vector of basic parameters and s a $q(k+1)$ vector of constants.

Note that our result generalizes that of Bai and Perron (1998). If there are no restrictions, $S = I_{(k+1)q}$ and the model is one of pure structural change with q regressors. In this case, (8) reduces to the expression in Bai and Perron (1998, Proposition 6) scaled by kq. Also, for a partial structural change model, with p coefficients not allowed to change, (8) reduces similarly (with $q - p$ instead of q). It is also of interest to note that the limit distribution of the test simplifies when the restrictions take the form $R\delta_i = r$ for all $i = 1, ..., m+1$ (i.e., the restrictions affect the coefficients the same way across all regimes and there are no cross regimes restrictions). In this case, simple algebra shows that (8) again reduces to the re-scaled expression in Bai and Perron (1998) with $q - rank(R)$ instead of q.

To show that, in general, the limit distribution depends on the specific structure of the matrix S, consider the two examples in the previous section. In this case, $q = 1$ and $k = 3$. For the first set of restrictions,

$$S' = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}.$$

Straightforward algebra shows that (8) reduces to

$$\frac{([W(\lambda_1) - \lambda_1 W(1)] - [W(\lambda_2) - \lambda_2 W(1)] + [W(\lambda_3) - \lambda_3 W(1)])^2}{(\lambda_1 - \lambda_2 + \lambda_3)[(\lambda_2 - \lambda_1) - (1 - \lambda_3)]}$$

For the second set of restrictions, $S' = [\ 1 \quad 0 \quad 1 \quad 0\ ]$ and (8) reduces to

$$[W(\lambda_1) + W(\lambda_3) - W(\lambda_2)]^2 / (\lambda_1 + \lambda_3 - \lambda_2).$$

7.1 Simulating the asymptotic critical values

The limit distribution of the test depends on q, k and ϵ as in Bai and Perron (1998) but also on the specific structure of the matrix S. This makes a tabulation of the critical values infeasible. But the limit distribution can be simulated. The idea is to use a design involving a finite sample of data, say T^*, such that the resulting F test has, if T^* is large enough, a distribution close to that of the limit functional $\sup \mathrm{F}(\lambda_1, ..., \lambda_k; q, S)$ as specified by (7). We then replicate the experiment to get N realizations from which we can extract the relevant quantiles. The steps involved for a particular replication are described below.

First note that the break points that maximize the test $F_T(\lambda_1, ..., \lambda_k; q, S)$ are asymptotically the same as those that minimize the global sum of squared residuals imposing the restrictions specified by the matrix S. Under the null hypothesis of no structural change, these estimates will have the same asymptotic distribution as the maximizers of $F(\lambda_1, ..., \lambda_k; q, S)$ specified by (8). Hence, for a particular replication, the first thing to do is to obtain estimates of the break dates minimizing the sum of squared residuals subject to restrictions. But to do so, one needs to specify some regressors. The fact is that the limit distribution does not depend on the nature of the regressors, all that is required is that they satisfy (A.8) and (A.9) defined in the appendix. A choice of regressors that will lead the finite sample distribution to be a good approximation to the limit one is to specify the dependent variable and the k regressors as independent $N(0, 1)$ random variables. With these data, we can then obtain a realization for the estimates of the break dates using the dynamic programming algorithm discussed in Section 5.2.

The next step is to simulate the functional $F(\lambda_1, ..., \lambda_k; q, S)$ as specified by (8) evaluated at the estimated break dates obtained above. Here, we first need to simulate a vector of independent Wiener processes and the vector W^*. Again, to ensure a good approximation, we use the partial sum of a sequence of independent $N(0, 1)$ random variables, i.e. $W_q(\lambda_i) \sim T^{*-1/2} \sum_{t=1}^{[T^* \lambda_i]} e_{q,t}$ where $e_{q,t}$ is a sequence of independent $N(0, I_q)$. This is done for $\lambda_i = \hat{T}_i^*/T^*$ $(i = 1, ..., k)$ with $\hat{T}_i^*$ the estimates of the break dates obtained in the first step. With these and the matrices H and S, we can then construct $F(\lambda_1, ..., \lambda_k; q, S)$.

A code written in the Gauss language is available from the authors. It allows the simulations of the 1%, 2.5%, 5% and 10% quantiles for user specified values of k, q, S, T^* and N. We have experimented with this code in cases for which asymptotic critical values are available and found it to provide reliable estimates.

7.2 Power simulations

To illustrate how performing hypothesis testing taking into account valid restrictions can be advantageous, we performed simple experiments related to the models discussed in Section 6.1. We used tests with a nominal size 10% and the critical values were simulated using the algorithm described in the previous section with $T^* = 1000$ and $N = 1000$. For both models, the sample size is $T = 120$ and we set $\delta_2 = \delta_4 = 0$ and $\delta_1 = \delta_3 = c$ and varied c from 0 to 2 in steps of 0.1 (size and power are evaluated at 21 points). Recall that model 1 specifies segments 1 and 3 and segments 2 and 4 to be the same, respectively. Model 2 adds the further restriction that the value of the parameters in segments 2 and 4 be 0. For each case, we also evaluated the power function of tests for 1, 2 or 3 breaks without imposing the restrictions (as discussed in Bai and Perron, 2003) to gauge the extent of the power gains. The results are presented in Figure 10, Panel (a) for model 1 and Panel (b) for model 2. They clearly show that imposing restrictions can increase power substantially.

8 Conclusions

We have presented a comprehensive treatment of estimation and inference in a single equation with multiple structural changes subject to restrictions. Our framework offers many new applications of interest including the following examples. Bekaert et al. (2002) argue that for economies with emerging stock markets, net equity flows should increase when the process of emergence starts and should decrease (presumably back to the initial level) once the transition phase is completed. Their estimates allowing two break dates indeed point in that direction, especially for Latin American countries. However, the testing procedure they apply cannot conclude for the presence of two breaks. Our procedure would allow to impose the same value for the mean of net equity flows prior to emergence and after the transition, yielding tests with higher power and better point estimates. Another case where a similar problem occurs is discussed by Sensier and van Dijk (2004). They argue that volatility of price time series show an increase in the early seventies and a decrease of roughly similar absolute magnitude in the early 1980s. An example involving additional restrictions on coefficients (though usually applied in the context of $I(1)$ series) is discussed by Martin (2000) and concerns the sustainability of government budget deficits. With R_t real tax revenues and G_t real debt interest inclusive expenditure, a fiscal policy is considered sustainable if $R_t = \alpha + \beta G_t + u_t$ with u_t a stationary process. It is strongly sustainable if $\beta = 1$ and weakly so if $0 < \beta < 1$. They argue that structural changes to this basic relationship have occurred. Interestingly, despite temporary deviations, a strongly sustainable fiscal policy would prevail with a coefficient $\beta = 1$ in the first and last segments provided they are large enough compared to the middle segments. Estimation could be performed imposing such a restriction, though inference would require an extension of our work to $I(1)$ variables, a topic of interest for further research. Also of interest are extensions related to multivariate systems. Bai, Lumsdaine and Stock (1998) have analyzed such a system when the break date is assumed to be the same across equations. They show important efficiency gains. This is, however, a special type of restrictions. The analysis of multi-equations systems with general restrictions is the object of ongoing research by the authors.

References

[1] Andrews, D.W.K., 1988. Laws of large numbers for dependent non-identically distributed random variables, Econometric Theory 4, 458-467.

[2] Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation, Econometrica 59, 817-858.

[3] Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change point, Econometrica 61, 821-856.

[4] Andrews, D.W.K., Lee, L., Ploberger, W., 1996. Optimal changepoint tests for normal linear regression, Journal of Econometrics 70, 9-38.

[5] Bai, J., 1997. Estimation of a change point in multiple regression models', Review of Economic and Statistics 79, 551-563.

[6] Bai, J., Lumsdaine, R.L., Stock, J.H., 1998. Testing for and dating common breaks in multivariate time series, Review of Economic Studies 65, 395-432.

[7] Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes, Econometrica 66, 47-78.

[8] Bai, J., Perron, P., 2001. Multiple structural change models: a simulation analysis, unpublished manuscript, Department of Economics, Boston University.

[9] Bai, J., Perron, P., 2003. Additional critical values for multiple structural changes tests, Econometrics Journal 6, 72-78.

[10] Bai, J., Perron, P., 2003. Computation and analysis of multiple structural change models, Journal of Applied Econometrics 18, 1-22.

[11] Bekaert, G., Harvey, C.R., Lumsdaine, R.L., 2002. The dynamics of emerging market equity flows, Journal of International Money and Finance 21, 295-350.

[12] Garcia, R., Perron, P., 1996. An analysis of the real interest rate under regime shifts, Review of Economics and Statistics 78, 111-125.

[13] Hall, P., Heyde, C.C., 1980. Martingale limit theory and its applications, New York: Academic Press.

[14] Hawkins, D. M., 1976. Point estimation of the parameters of a piecewise regression model, Applied Statistics 25, 51-57

[15] Horn, R.A., Johnson, C.R., 1985. Matrix analysis, New York; Cambridge University Press.

[16] Liu, J., Wu, S., Zidek J.V., 1997. On segmented multivariate regressions', Statistica Sinica 7, 497-525.

[17] Magnus, J., Neudecker, H., 1999. Matrix differential calculus with application in statistics and econometrics, Revised Edition, New York: Wiley.

[18] Martin, G.M., 2000. US deficit sustainability: a new approach based on multiple endogenous breaks, Journal of Applied Econometrics 15, 83-105.

[19] McLeish, D.L., 1975., A maximal inequality and dependent strong laws, The Annals of Probability 5, 829-839.

[20] Qu, Z., 2005. Essays on structural change, long memory and cointegration, unpublished PhD Dissertation, Department of Economics, Boston University.

[21] Quandt, R.E., 1958. The estimation of the parameters of a linear regression system obeying two separate regimes, Journal of the American Statistical Association 53, 873-880.

[22] Sensier, M., van Dijk, D., 2004. Testing for volatility changes in U.S. macroeconomic time series, The Review of Economics and Statistics 86, 833-839.

[23] Tobing, H., McGilchrist, C.A., 1992. Recursive residuals for multivariate regression models, Australian Journal of Statistics 34, 217-232.

Appendix

Throughout, we use the Euclidean norm for vectors. For matrices, we use the vector induced norm, i.e. $\|A\| = \sup_{x\neq 0} \|Ax\| / \|x\|$. Note that $\|A\|$ is then equivalent to the square of maximum eigenvalue of $A'A$ and thus $\|A\| \leq [tr(A'A)]^{1/2}$. Also, $\|PA\| \leq \|A\|$, for a projection matrix P. We shall make use of the fact that for an n by n matrix A, $\|A'A\| = \|A\|^2$ (e.g., Horn and Johnson (1985), p. 312).

We first need to establish that under our set of assumptions, the results of Bai and Perron (1998) remain valid. The main difference is in the corresponding Assumption A4. In Bai and Perron (1998), this assumption is used to establish Lemma A.4 and Lemma A.6. We now show that an equivalent result holds with our version of Assumption A4, the strenghtenning of A1 and the addition of A6.

First note that Assumption A4 guarantees that $\sum_{i=1}^{k} z_i u_i$ satisfies the generalized Hajek and Renyi inequality for mixingales (c.f. Lemma A.6 of Bai and Perron (1998)), i.e., there exists an $L > 0$ such that for every $c > 0$ and $m > 0$,

$$P\left(\sup_{k\geq m} \frac{1}{k} \left\| \sum_{i=1}^{k} z_i u_i \right\| > c\right) \leq \frac{L}{c^2 m}$$

With this, we have the following Lemma which is actually stronger than Lemma A.4 of Bai and Perron (1998).

Lemma A.1 *Let $\bar{Z}$ denote the partitioned matrix of regressors using the m-partition $(T_1, ..., T_m)$ and $P_{\bar{Z}}u$ denote the projection of the error process on the space spanned by $\bar{Z}$, then*

$$\sup_{T_1,\dots,T_m} (\|P_{\bar{Z}}u\|) = O_p\left(\log^{1/2} T\right)$$

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

Proof: We shall prove that $\|u'P_{\bar{Z}}u\| = O_p(\log T)$ uniformly in $(T_1, ..., T_m)$. Note that $u'P_{\bar{Z}}u$ is the summation of the $m+1$ terms

$$\left(\sum_{t=T_i+1}^{T_{i+1}} z_t u_t\right)' \left(\sum_{t=T_i+1}^{T_{i+1}} z_t z_t'\right)^{-1} \left(\sum_{t=T_i+1}^{T_{i+1}} z_t u_t\right)$$

for $i = 0, ..., m$. Define $l = T_{i+1} - T_i$, then it suffices to show that

$$\left\| \left(l^{-1} \sum_{t=T_i+1}^{T_{i+1}} z_t z_t'\right)^{-1} \right\| = O_p(1) \tag{A.1}$$

and

$$\left\| l^{-1/2} \left(\sum_{t=T_i+1}^{T_{i+1}} z_t u_t\right) \right\| = O_p\left(\log^{1/2} T\right) \tag{A.2}$$

uniformly in $(T_1, ..., T_m)$ with $|T_i - T_{i-1}| \geq \epsilon T$.

First consider (A.1) and note that the summation may either contain data from only one regime, say regime k, or may contain data from multiple regimes, say, from regimes k and $k+1$. In either case, it always contains at least $[\epsilon T/2]$ data points from one particular regime and, without loss of generality, we assume the first $[\epsilon T/2]$ data points are from regime k. Then

$$l^{-1} \sum_{t=T_i+1}^{T_{i+1}} z_t z_t' = l^{-1} \sum_{t=T_k^0+1}^{T_k^0+[\epsilon T/2]} z_t z_t' + l^{-1} \sum_{t=T_k^0+[\epsilon T/2]+1}^{T_{i+1}} z_t z_t'$$

and

$$\left(l^{-1} \sum_{t=T_i+1}^{T_{i+1}} z_t z_t' \right)^{-1} \leq \left(l^{-1} \sum_{t=T_k^0+1}^{T_k^0+[\epsilon T/2]} z_t z_t' \right)^{-1}$$

in the sense that the difference between the two matrices is positive semi-definite. (c.f., Magnus and Neudecker, 1999, page 239, exercise 23). Accordingly,

$$\left\| \left(l^{-1} \sum_{t=T_i+1}^{T_{i+1}} z_t z_t' \right)^{-1} \right\| \leq \left\| \left(l^{-1} \sum_{t=T_k^0+1}^{T_k^0+[\epsilon T/2]} z_t z_t' \right)^{-1} \right\| = \frac{1}{\epsilon} O_p(1)$$

Now, consider the term (A.2), under A4, $\sum_{t=T_i+1}^{T_{i+1}} z_t u_t$ satisfies the generalized Hajek and Renyi inequality stated above, hence

$$\begin{aligned} P\left(\sup_{l \geq \epsilon T} \left\| l^{-1/2} \left(\sum_{t=T_i+1}^{T_{i+1}} z_i u_i \right) \right\| > \log^{1/2} T \right) &= P\left(\sup_{l \geq \epsilon T} \left\| l^{-1} \left(\sum_{t=T_i+1}^{T_{i+1}} z_i u_i \right) \right\| > l^{-1/2} \log^{1/2} T \right) \\ &\leq \frac{T}{\epsilon T \log T} \to 0 \end{aligned}$$

which implies

$$\left\| l^{-1/2} \left(\sum_{t=T_i+1}^{T_{i+1}} z_t u_t \right) \right\| = O_p\left(\log^{1/2} T \right)$$

Combining these results, we have

$$\left(\sum_{t=T_i+1}^{T_{i+1}} z_t u_t \right)' \left(\sum_{t=T_i+1}^{T_{i+1}} z_t z_t' \right)^{-1} \left(\sum_{t=T_i+1}^{T_{i+1}} z_t u_t \right) = O_p(\log T)$$

Proof of Proposition 1(a): Denote the restricted SSR under the optimal partition $\{\widetilde{T}_1, ..., \widetilde{T}_m\}$ as $\widetilde{u}'\widetilde{u}$, the unrestricted SSR under the optimal partition $\{\hat{T}_1, ..., \hat{T}_m\}$ as $\hat{u}'\hat{u}$ and the restricted SSR under the true partition $\{T_1^0, ..., T_m^0\}$ as $\widetilde{u}^{0\prime}\widetilde{u}^0$. First note that $\widetilde{u}'\widetilde{u} \leq \widetilde{u}^{0\prime}\widetilde{u}^0$ and, assuming the restrictions hold, $\widetilde{u}^{0\prime}\widetilde{u}^0 \leq u'u$. Hence, we have

$$T^{-1}\widetilde{u}'\widetilde{u} \leq T^{-1}u'u \tag{A.3}$$

Now, from Bai and Perron (1998, Lemma 2), when at least one break is not consistently estimated, we have that

$$T^{-1}\hat{u}'\hat{u} \geq T^{-1}u'u + C\left\|\delta_j^0 - \delta_{j+1}^0\right\|^2 + o_p(1)$$

holds with positive probability. Using the fact that $\hat{u}'\hat{u} \leq \widetilde{u}'\widetilde{u}$, we also have that

$$T^{-1}\widetilde{u}'\widetilde{u} \geq T^{-1}u'u + C\left\|\delta_j^0 - \delta_{j+1}^0\right\|^2 + o_p(1) \tag{A.4}$$

holds with positive probability. Comparing (A.3) and (A.4), we get a contradiction if at least one break fraction is not consistently estimated, which completes the proof.

Proof of Proposition 1(b): The difficulty is that with restrictions across regimes, the estimates cannot be analyzed using results obtained segment by segment. The solution is to analyze directly the relationship between the Sum of Squared Residuals from the restricted (SSR_T^R) and unrestricted (SSR_T^U) models. We have

$$SSR_T^R(T_1, ..., T_m) = SSR_T^U(T_1, ..., T_m) + (r - R\hat{\delta})'[R(\bar{Z}'\bar{Z})^{-1}R']^{-1}(r - R\hat{\delta})$$

Define the set

$$V_{\varepsilon,i}(C) = \{(T_1, ..., T_m) : |T_i - T_i^0| < \varepsilon T,\ |T_i - T_i^0| > C,\ \text{ for some } i\}$$

Note that we are using the fact that the break fraction is consistent. To prove the result, all we need to show is that for every $\eta > 0$, there exists a $C < \infty$ and $\varepsilon > 0$, such that for $(T_1, ...T_m) \in V_{\varepsilon,i}(C)$, we have

$$P(\min(SSR_T^R(T_1, ..., T_i, ..., T_m) - SSR_T^R(T_1, ..., T_i^0, ...T_m))/(T_i^0 - T_i)) \leq 0) < \eta$$

for T large enough. As a matter of notation, let the sequence $\{T_1, ..., T_i, ..., T_m\}$ be a partition with associated unrestricted Sum of Squared Residuals SSR_1^U, partitioned regressor matrix $\bar{Z}_1$ and estimated coefficients $\hat{\delta}_1$. Similarly let the sequence $\{T_1, ..., T_i^0, ...T_m\}$ be a partition with associated unrestricted Sum of Squared Residuals SSR_0^U, partitioned regressor matrix $\bar{Z}_0$ and estimated coefficients $\hat{\delta}_0$. We also use a superscript 0 for the true value. We first state a Lemma that will be useful in subsequent developments.

Lemma A.2 *(Tobing and McGilchrist, 1992) Let A and B be two symmetric n by n matrices related by*

$$B = A + ZF^{-1}Z'$$

where F is symmetric of full rank p and Z is a n by p matrix. If $Rank\,(B) = Rank\,(A)$, then generalized inverses B^- and A^- exist satisfying

$$B^- = A^- - A^- Z D^{-1} Z' A^-$$

where

$$D = F + Z' A^- Z$$

Using this Lemma, on the set $V_{\varepsilon,i}(C)$ for C large enough, we can deduce that

$$(\bar{Z}_1' \bar{Z}_1)^{-1} = (\bar{Z}_0' \bar{Z}_0)^{-1} + O_p(\frac{T_i - T_i^0}{T^2}) \qquad \text{(A.5)}$$

and

$$[R(\bar{Z}_1' \bar{Z}_1)^{-1} R']^{-1} = [R(\bar{Z}_0' \bar{Z}_0)^{-1} R']^{-1} + O_p(T_i - T_i^0) \qquad \text{(A.6)}$$

Now, define $\hat{\delta}_\Delta = \hat{\delta}_1 - \hat{\delta}_0$. We have

$$\begin{aligned} \hat{\delta}_\Delta &= (\bar{Z}_1' \bar{Z}_1)^{-1} \bar{Z}_1' (\bar{Z}^0 \delta^0 + u) - (\bar{Z}_0' \bar{Z}_0)^{-1} \bar{Z}_0' (Z^0 \delta^0 + u) \\ &= (\bar{Z}_0' \bar{Z}_0)^{-1} \left(\bar{Z}_1' - \bar{Z}_0'\right) \bar{Z}^0 \delta^0 + (\bar{Z}_0' \bar{Z}_0)^{-1} \left(\bar{Z}_1' - \bar{Z}_0'\right) u + \left|T_i - T_i^0\right| O_p(T^{-1}) \\ &= (\bar{Z}_0' \bar{Z}_0)^{-1/2} A_T \end{aligned}$$

with

$$A_T = (\bar{Z}_0' \bar{Z}_0)^{-1/2} \left(\bar{Z}_1' - \bar{Z}_0'\right) \bar{Z}^0 \delta^0 + (\bar{Z}_0' \bar{Z}_0)^{-1/2} \left(\bar{Z}_1' - \bar{Z}_0'\right) u + \left|T_i - T_i^0\right| O_p(T^{-1/2})$$

using (A.5). Now, $\left(\bar{Z}_1' - \bar{Z}_0'\right) \bar{Z}^0 = |T_i - T_i^0|\, O_p(1)$, hence

$$(\bar{Z}_0' \bar{Z}_0)^{-1/2} \left(\bar{Z}_1' - \bar{Z}_0'\right) \bar{Z}^0 \delta^0 = |T_i - T_0|\, O_p(T^{-1/2}).$$

For the second term, $\left(\bar{Z}_1' - \bar{Z}_0'\right) u = |T_i - T_0|\, O_p(1)$. Hence, $A_T = |T_i - T_i^0| O_p(T^{-1/2})$ and $\hat{\delta}_\Delta = |T_i - T_i^0| O_p(T^{-1})$. Then, using the fact that $\hat{\delta}_\Delta' R' = |T_i - T_i^0| O_p(T^{-1})$ and $(r - \hat{\delta}_1 R)' = |T_i - T_i^0| O_p(T^{-1})$,

$$\begin{aligned} & SSR_T^R(T_1, ..., T_i, ..., T_m) - SSR_T^R(T_1, ..., T_i^0, ..., T_m) \\ = \; & \left[SSR_1^U - SSR_0^U\right] + \\ & (r - R\hat{\delta}_1)' [R(\bar{Z}_1' \bar{Z}_1)^{-1} R']^{-1} (r - R\hat{\delta}_1) - (r - R\hat{\delta}_0)' [R(\bar{Z}_0' \bar{Z}_0)^{-1} R']^{-1} (r - R\hat{\delta}_0) \\ = \; & \left[SSR_1^U - SSR_0^U\right] + (r - R\hat{\delta}_1)' [R(\bar{Z}_0' \bar{Z}_0)^{-1} R']^{-1} (r - R\hat{\delta}_1) \\ & - (r - R\hat{\delta}_0)' [R(\bar{Z}_0' \bar{Z}_0)^{-1} R']^{-1} (r - R\hat{\delta}_0) + \left|T_i - T_i^0\right|^2 O_p(T^{-1}) \\ = \; & \left[SSR_1^U - SSR_0^U\right] + (\hat{\delta}_0 + \hat{\delta}_\Delta)' R' [R(\bar{Z}_0' \bar{Z}_0)^{-1} R']^{-1} R(\hat{\delta}_0 + \hat{\delta}_\Delta) \\ & - \hat{\delta}_0' R' [R(\bar{Z}_0' \bar{Z}_0)^{-1} R']^{-1} R\hat{\delta}_0 - \left[2r' [R(\bar{Z}_0' \bar{Z}_0)^{-1} R']^{-1} R(\hat{\delta}_1 - \hat{\delta}_0)\right] + \left|T_i - T_i^0\right|^2 O_p(T^{-1}) \end{aligned}$$

$$
\begin{aligned}
= & \left[SSR_1^U - SSR_0^U\right] + 2\hat{\delta}_\Delta' R'[R(\bar{Z}_0'\bar{Z}_0)^{-1}R']^{-1}R[(\bar{Z}_0'\bar{Z}_0)^{-1}\bar{Z}_0'u] \\
& + \hat{\delta}_\Delta' R'[R(\bar{Z}_0'\bar{Z}_0)^{-1}R']^{-1}R\hat{\delta}_\Delta + 2\hat{\delta}_\Delta' R'[R(\bar{Z}_0'\bar{Z}_0)^{-1}R']^{-1}[R(\bar{Z}_0'\bar{Z}_0)^{-1}\bar{Z}_0'\bar{Z}^0\delta^0 - r] \\
& + \left|T_i - T_i^0\right|^2 O_p(T^{-1}) \\
= & \left[SSR_1^U - SSR_0^U\right] \quad (T1) \\
& + 2\hat{\delta}_\Delta' R'[R(\bar{Z}_0'\bar{Z}_0)^{-1}R']^{-1}R[(\bar{Z}_0'\bar{Z}_0)^{-1}\bar{Z}_0'u] \quad (T2) \\
& + \hat{\delta}_\Delta' R'[R(\bar{Z}_0'\bar{Z}_0)^{-1}R']^{-1}R\hat{\delta}_\Delta \quad (T3) \\
& + 2\hat{\delta}_\Delta' R'[R(\bar{Z}_0'\bar{Z}_0)^{-1}R']^{-1}R(\bar{Z}_0'\bar{Z}_0)^{-1}\bar{Z}_0'(\bar{Z}^0 - \bar{Z}_0)\delta^0 \quad (T4) \\
& + \left|T_i - T_i^0\right|^2 O_p(T^{-1})
\end{aligned}
$$

We consider the order of each term. For (T1),

$$
\frac{SSR_1^U - SSR_0^U}{T_i - T_i^0} \geq 2^{-1}(\delta_{i+1}^0 - \delta_i^0)' \frac{(Z_1 - Z_0)'(Z_1 - Z_0)}{|T_i - T_i^0|}(\delta_{i+1}^0 - \delta_i^0) - \varepsilon O_p(1) - \rho O_p(1)
$$

where ε and ρ can be made arbitrarily small by choosing a small ε and a large T. Now $(Z_1 - Z_0)'(Z_1 - Z_0)$ has $|T_i - T_i^0|$ terms and has minimum eigenvalue bounded away from zero. So with large probability it is the dominating component and the term (T1) is positive with large probability. Consider now term (T2).

$$
\begin{aligned}
& \left\| \hat{\delta}_\Delta' R'[R(\bar{Z}_0'\bar{Z}_0)^{-1}R']^{-1}R[(\bar{Z}_0'\bar{Z}_0)^{-1}\bar{Z}_0'u] \right\| \\
= & \left\| A_T' P_{(\bar{Z}_0'\bar{Z}_0)^{-1/2}R'}[(\bar{Z}_0'\bar{Z}_0)^{-1/2}\bar{Z}_0'u] \right\| \leq \|A_T'\| \left\|(\bar{Z}_0'\bar{Z}_0)^{-1/2}\bar{Z}_0'u\right\|
\end{aligned}
$$

Since $A_T' = |T_i - T_i^0|O_p(T^{-1/2})$ and $(\bar{Z}_0'\bar{Z}_0)^{-1/2}\bar{Z}_0'u = O_p(\log T)$,the term (T2) divided by $|T_k - T_k^0|$ is $o_p(1)$. Now, term (T3) divided by $|T_k - T_k^0|$ is proportional to $|T_k - T_k^0|O_p(T^{-1}) = \varepsilon O_p(1)$, which can again be made arbitrarily small by choosing ε small enough. The key to analyze term (T4) is to note that $(\bar{Z}^0 - \bar{Z}_0)$ has at most $2\varepsilon T$ non zero terms in each column. So $||(\bar{Z}^0 - \bar{Z}_0)|| = (2\varepsilon)^{1/2}O_p(T^{1/2})$ and $\bar{Z}_0'(\bar{Z}^0 - \bar{Z}_0) = 2\varepsilon O_p(T)$. Hence, term (T4) divided by $|T_i - T_i^0|$ is of order $2\varepsilon O_p(1)$ which can again be made arbitrarily small by choosing ε small enough. Summarizing, the term (T1) dominates all others and is positive with large probability for large T. This says that, for large T, we have with large probability

$$
[SSR_T^R(T_1, ..., T_i, ..., T_m) - S_T^R(T_1, ..., T_i^0, ..., T_m)]/|T_i - T_i^0| > 0. \tag{A.7}
$$

Thus with large probability the estimated break date $\hat{T}_i$ cannot be in the set $V_{\varepsilon,i}(C)$ and, hence, $|\hat{T}_i - T_i^0| < C$ when T is large. This completes the proof.

Proof of Proposition 2: The proof of consistency uses Proposition 4(i) of Bai and Perron (1998) and proceeds in the same way as the case of breaks with fixed magnitudes. The details are, hence, omitted. To derive the rate of convergence, define the set

$$V_{\varepsilon,i}(C) = \{(T_1, ...T_m) : |T_i - T_i^0| < \varepsilon T, |T_i - T_i^0| > C/v_T^2 \text{ for some } i\}.$$

We want to show that for every $\eta > 0$, there exists a $C < \infty$ and $\varepsilon > 0$ such that for $(T_1, ..., T_m) \in V_{\varepsilon,i}(C)$

$$P(\min(SSR_T^R(T_1, ..., T_i, ..., T_m) - SSR_T^R(T_1, ..., T_i^0, ...T_m))/(T_i^0 - T_i)) \leq 0) < \eta$$

for large T. Bai and Perron (1998) show that the term (T1) divided by $|T_i - T_i^0|$ is positive with probability arbitrarily close to one when T is large. For the term (T2) divided by $|T_i - T_i^0|$, we have, using argument similar to those used in the proof of Proposition 1(b) for the term (T2),

$$2\hat{\delta}_\Delta' R'[R(\bar{Z}_0'\bar{Z}_0)^{-1}R']^{-1}R[(\bar{Z}_0'\bar{Z}_0)^{-1}\bar{Z}_0 u]/|T_i - T_i^0| = o_p(1)$$

which can then be omitted. The term (T3) is always non negative. For the term (T4), we already showed that $\hat{\delta}_\Delta' = |T_i - T_i^0| \, O_p(T^{-1})$. The key element here is that $(\bar{Z}^0 - \bar{Z}_0)'\delta = O_p(T^{1/2}v_T)$. Thus the term (T4) divided by $|T_i - T_i^0|$ is of order $O_p(v_T) = o_p(1)$ and can thus be omitted. To summarize, a break date which minimizes SSR must be outside the set $V_{\varepsilon,i}(C)$. Hence, $|\hat{T}_i - T_i^0| > C/v_T^2$, for large T, which completes the proof.

Proof of Lemma 1: Given the rate of convergence for v_T fixed or satisfying assumption A7, we restrict the search for breaks in the set $\{T_k : |T_k - T_k^0| \leq C \|v_T\|^{-2}, k = 1, ..., m\}$. Note that

$$SSR_T^R(T_1, ..., T_m) = SSR_T^U(T_1, ..., T_m) + (r - R\hat{\delta})'[R(\bar{Z}'\bar{Z})^{-1}R']^{-1}(r - R\hat{\delta})$$

Hence,

$$\begin{aligned}
& SSR_T^R(T_1, ..., T_i, ..., T_m) - SSR_T^R(T_1^0, ..., T_i^0, ..., T_m^0) \\
= \; & \left[SSR_T^U(T_1, ..., T_i, ..., T_m) - SSR_T^U(T_1^0, ..., T_i^0, ..., T_m^0)\right] \\
& + (r - R\hat{\delta})'[R(\bar{Z}'\bar{Z})^{-1}R']^{-1}(r - R\hat{\delta}) - (r - R\hat{\delta}^0)'[R(\bar{Z}^{0\prime}\bar{Z}^0)^{-1}R']^{-1}(r - R\hat{\delta}^0).
\end{aligned}$$

The key is to show that the second term is $o_p(1)$. First, note that $T^{-1}\bar{Z}'\bar{Z} = T^{-1}\bar{Z}^{0\prime}\bar{Z}^0 + o_p(1)$. To see this, consider the case with one break and assume $T_k - T_k^0 < 0$. Then

$$\begin{aligned}
\left\|T^{-1}\bar{Z}'\bar{Z} - T^{-1}\bar{Z}^{0\prime}\bar{Z}^0\right\| &= T^{-1}\left\|\sum_{t=\hat{T}_k+1}^{T_k^0} z_t z_t' \otimes \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}\right\| \\
&\leq \sqrt{2}\frac{1}{T\|v_T\|^2}\left[tr(\sum_{t=\hat{T}_k+1}^{T_k^0} z_t z_t' \|v_T\|^2)^2\right]^{1/2} = o_p(1),
\end{aligned}$$

The last equality follows because $\sum_{t=\hat{T}_i+1}^{T_i^0} z_t z_t'$ involves at most $O_p(\|v_T\|^{-2})$ terms. The case with multiple breaks can be handled similarly. Second, we have

$$T^{-1}[R(\bar{Z}'\bar{Z})^{-1}R']^{-1} = T^{-1}[R(\bar{Z}^{0\prime}\bar{Z}^0)^{-1}R']^{-1} + o_p(1)$$

using (A.6). Third, and most importantly, $\sqrt{T}(\hat{\delta}-\delta^0) = \sqrt{T}(\hat{\delta}^0-\delta^0)+o_p(1)$, since $\sqrt{T}(\hat{\delta}-\delta^0)$ and $\sqrt{T}(\hat{\delta}^0-\delta^0)$ have the same limiting distribution. Hence

$$\begin{aligned}
& (r-R\hat{\delta})'[R(\bar{Z}'\bar{Z})^{-1}R']^{-1}(r-R\hat{\delta}) - (r-R\hat{\delta}^0)'[R(\bar{Z}^{0\prime}\bar{Z}^0)^{-1}R']^{-1}(r-R\hat{\delta}^0) \\
= \; & (r-R\hat{\delta})'\left\{[R(\bar{Z}^{0\prime}\bar{Z}^0)^{-1}R']^{-1}+o_p(1)\right\}(r-R\hat{\delta}) - (r-R\hat{\delta}^0)'[R(\bar{Z}^{0\prime}\bar{Z}^0)^{-1}R']^{-1}(r-R\hat{\delta}^0) \\
= \; & \sqrt{T}(\delta^0-\hat{\delta})'R'[R(T^{-1}\bar{Z}^{0\prime}\bar{Z}^0)^{-1}R']^{-1}R\sqrt{T}(\delta^0-\hat{\delta}) \\
& -\sqrt{T}(\delta^0-\hat{\delta}^0)'R'[R(T^{-1}\bar{Z}^{0\prime}\bar{Z}^0)^{-1}R']^{-1}\sqrt{T}R(\delta^0-\hat{\delta}^0) + o_p(1) = o_p(1)
\end{aligned}$$

Thus

$$\begin{aligned}
\widetilde{T}(c) &= \arg \min_{|T_k-T_k^0|\le C\|v_T\|^{-2},k=1,\ldots,m} SSR_T^R(T_1,...,T_m) - SSR_T^R(T_1^0,...,T_m^0) \\
&= \arg \min_{|T_k-T_k^0|\le C\|v_T\|^{-2},k=1,\ldots,m} \left[SSR_T^U(T_1,...,T_m) - SSR_T^U(T_1^0,...,T_m^0)\right] + o_p(1)
\end{aligned}$$

Given this result, the problem reduces to one which involves searching for a global minimum without restrictions for which results are available (e.g., Bai and Perron, 1998).

Proof of Proposition 7: For a given partition $(T_1,...,T_k)$ and $\bar{Z}$ the associated diagonal partition of the regressor matrix Z, the restricted estimate of δ allowing for breaks is:

$$\widetilde{\delta} = S[S'\bar{Z}'\bar{Z}S]^{-1}S'\bar{Z}'(y-\bar{Z}s)+s$$

with variance

$$V(\widetilde{\delta}) = S[S'\bar{Z}'\bar{Z}S]^{-1}S'\bar{Z}'\Omega\bar{Z}S[S'\bar{Z}'\bar{Z}S]^{-1}S'$$

In what follows, we assume, for simplicity, spherical errors so that $u_t \sim i.i.d.(0,\sigma^2)$ and $V(\widetilde{\delta}) = \sigma^2 S[S'\bar{Z}'\bar{Z}S]^{-1}S'$. All results remain valid in the general case. Under the null hypothesis of no structural change, we have

$$H\widetilde{\delta} = HS[S'\bar{Z}'\bar{Z}S]^{-1}S'\bar{Z}'u.$$

With $\widetilde{\sigma}^2 = T^{-1}\sum_{t=1}^T \widetilde{u}_t^2$ and $\widetilde{u}_t$ the residuals, we can express the F test as

$$F_T(\widetilde{\lambda}_1,...,\widetilde{\lambda}_k;q) = \frac{u'\bar{Z}S[S'\bar{Z}'\bar{Z}S]^{-1}S'H'(HS[S'\bar{Z}'\bar{Z}S]^{-1}S'H')^{-}HS[S'\bar{Z}'\bar{Z}S]^{-1}S'\bar{Z}'u}{\widetilde{\sigma}^2}$$

From Bai and Perron (1998), we have

$$T^{-1}\bar{Z}'\bar{Z} \;\to\; {}_p\Lambda\otimes Q \qquad \text{(A.8)}$$

$$T^{-1/2}\bar{Z}'u \;\Rightarrow\; \sigma(I_{k+1}\otimes Q^{1/2})W^* \qquad \text{(A.9)}$$

Using these results and the fact that $\widetilde{\sigma}^2 \to_p \sigma^2$, the result of Proposition 7 follows directly.

Figure 1: Distribution of the estimates of the break dates

(model 1, break=0.5, h=5, T=120; percentage deviation relative to T)

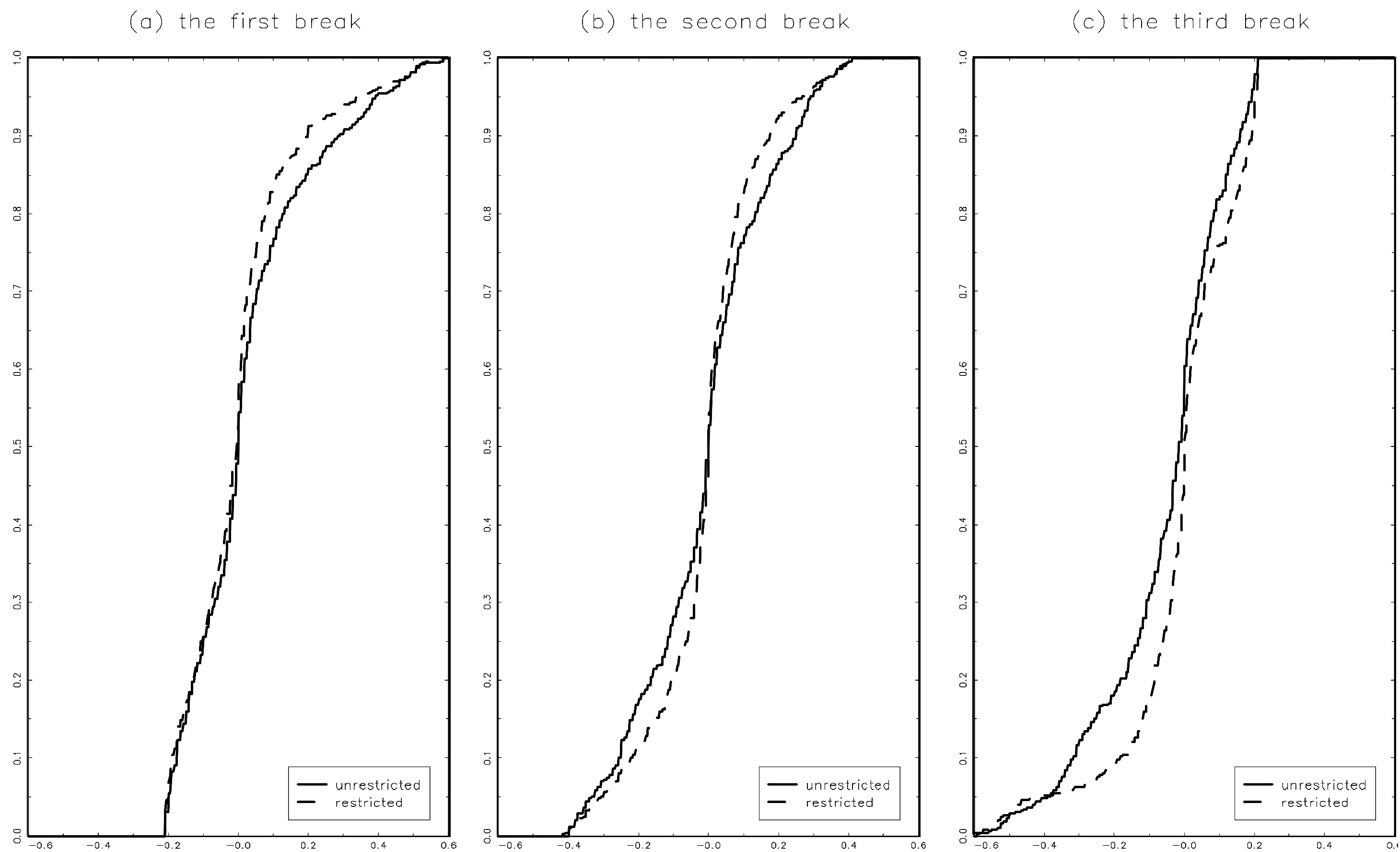

Figure 2: Distribution of the parameter estimates

(model 1, break=0.5, h=5, T=120)

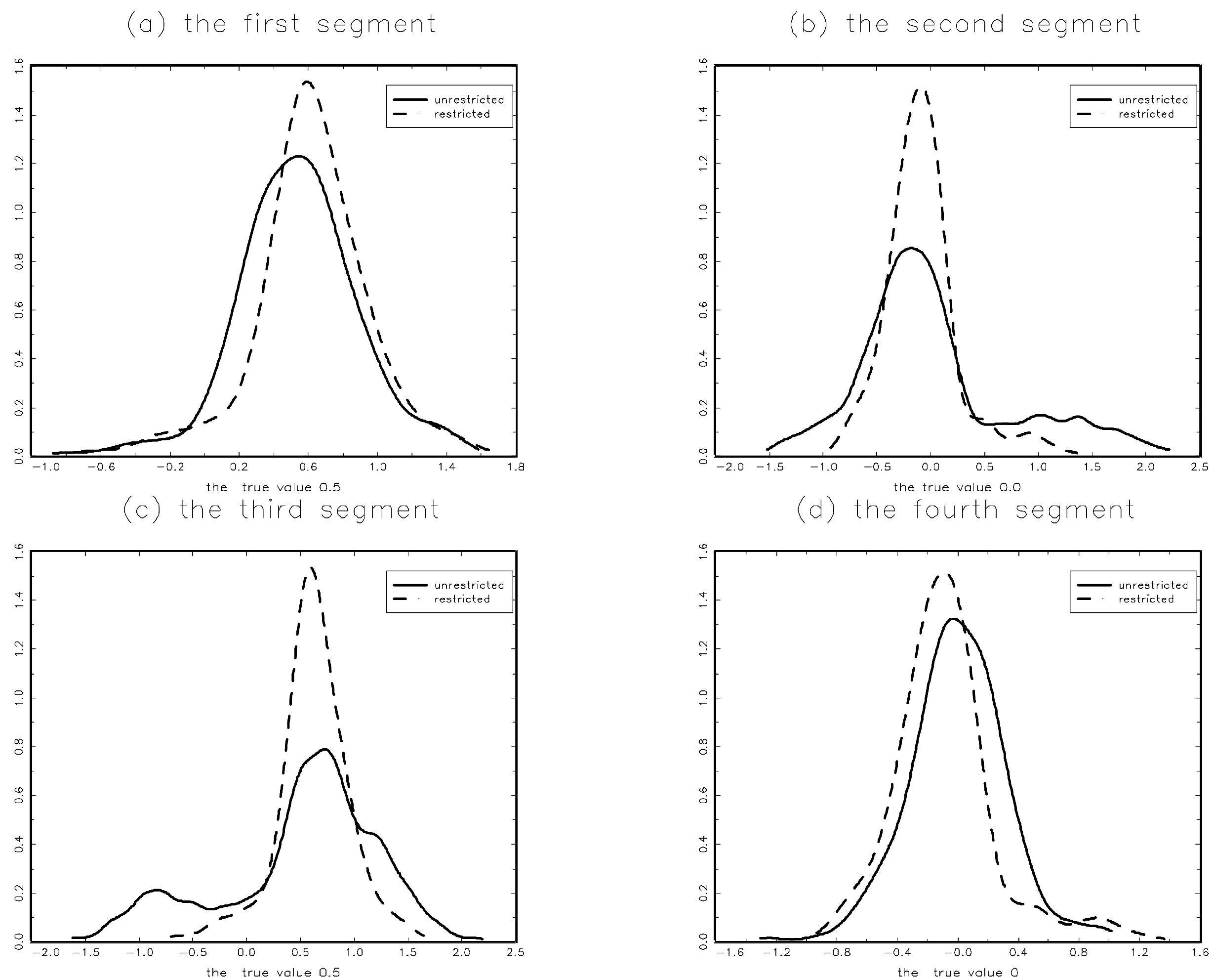

Figure 3: Distribution of the estimates of the break dates

(model 2, break=0.5, h=5, T=120; percentage deviation relative to T)

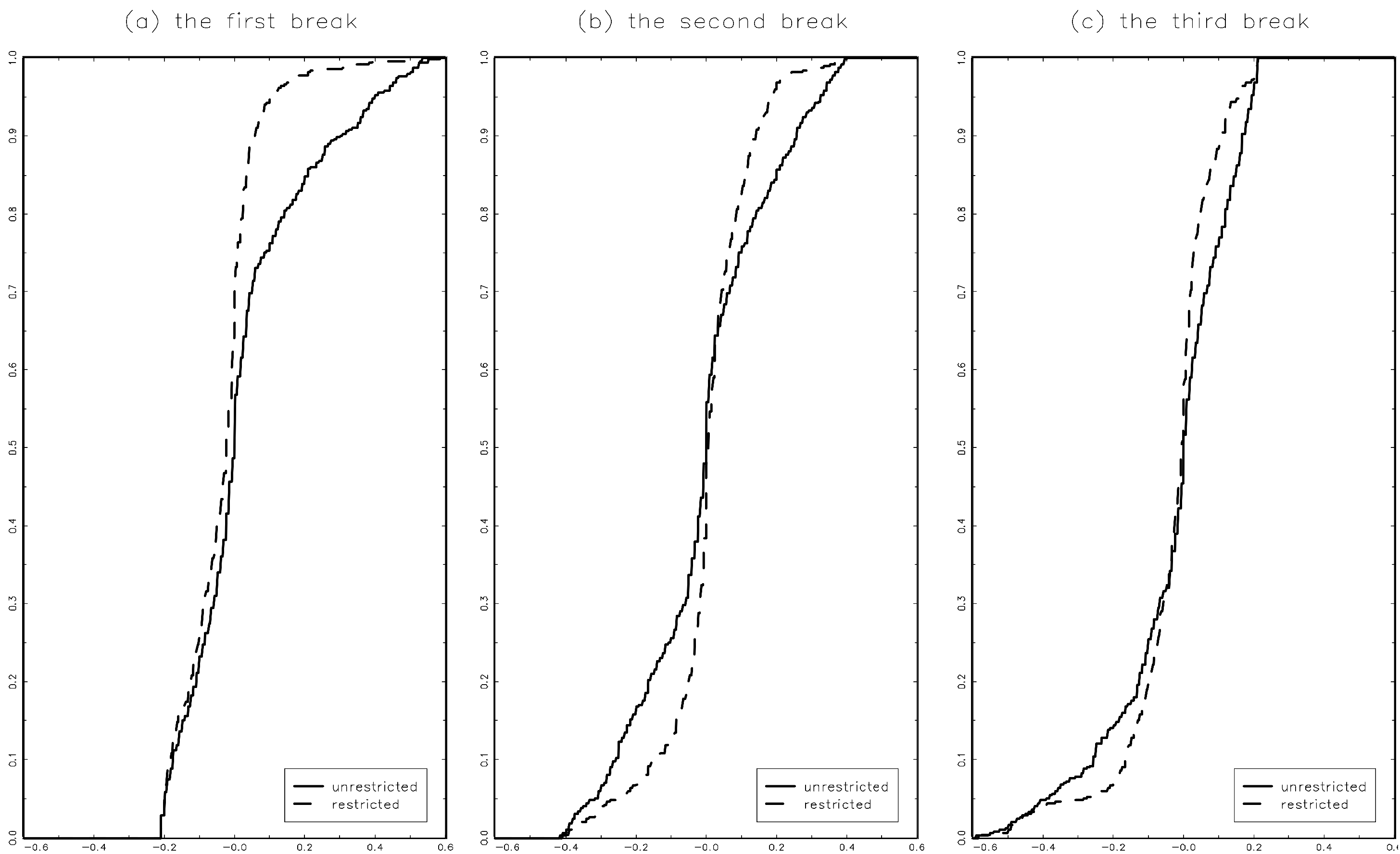

Figure 4: Distribution of the parameter estimates

(model 2, break=0.5, h=5, T=120)

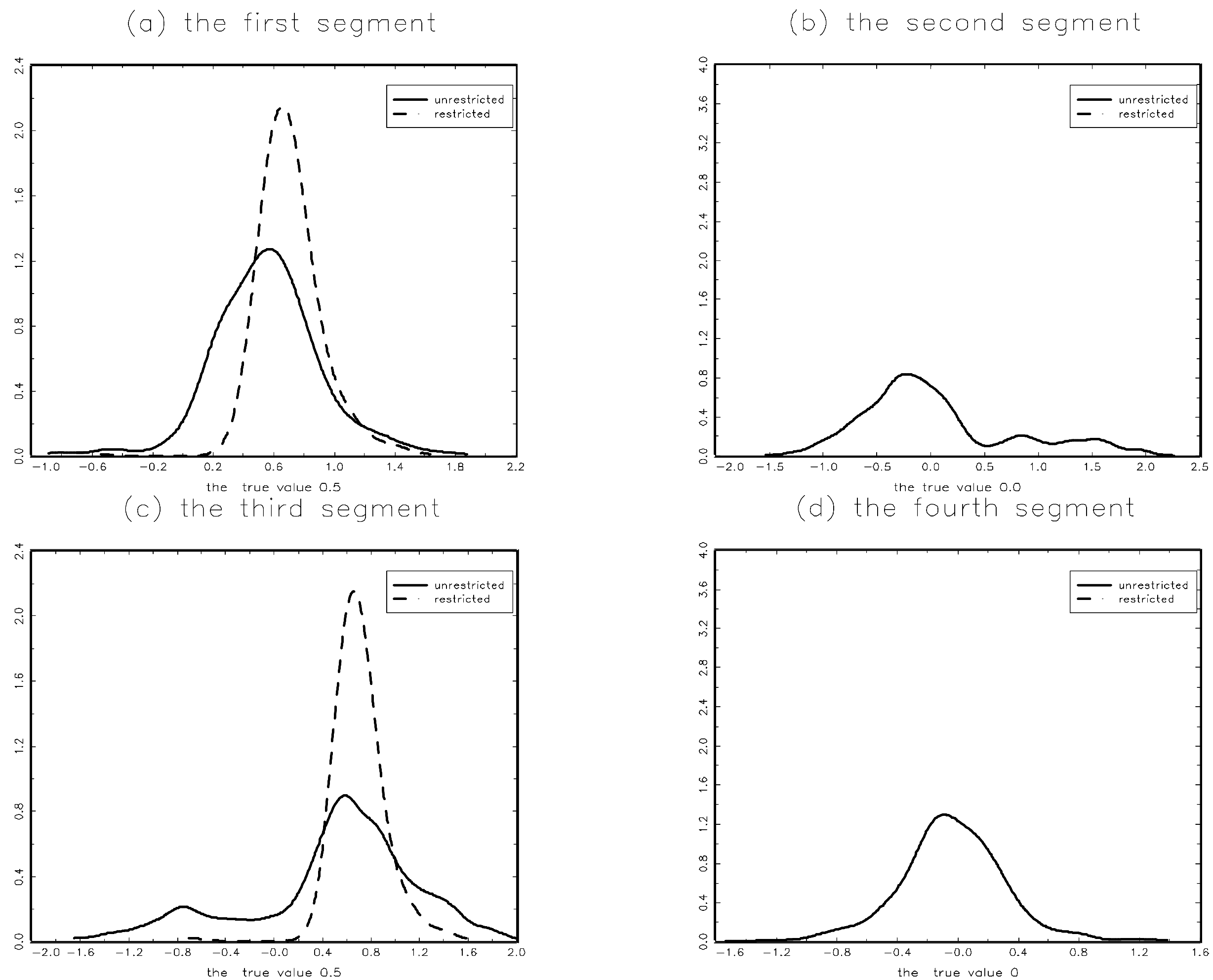

Figure 5: Power comparison of structural change tests