

COMPUTATION AND ANALYSIS OF MULTIPLE STRUCTURAL CHANGE MODELS

JUSHAN BAI^a AND PIERRE PERRON^{b*}

^a Department of Economics, Boston College, Chestnut Hill, MA 02467, USA

^b Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215, USA

SUMMARY

In a recent paper, Bai and Perron (1998) considered theoretical issues related to the limiting distribution of estimators and test statistics in the linear model with multiple structural changes. In this companion paper, we consider practical issues for the empirical applications of the procedures. We first address the problem of estimation of the break dates and present an efficient algorithm to obtain global minimizers of the sum of squared residuals. This algorithm is based on the principle of dynamic programming and requires at most least-squares operations of order $O(T^2)$ for any number of breaks. Our method can be applied to both pure and partial structural change models. Second, we consider the problem of forming confidence intervals for the break dates under various hypotheses about the structure of the data and the errors across segments. Third, we address the issue of testing for structural changes under very general conditions on the data and the errors. Fourth, we address the issue of estimating the number of breaks. Finally, a few empirical applications are presented to illustrate the usefulness of the procedures. All methods discussed are implemented in a GAUSS program. Copyright © 2002 John Wiley & Sons, Ltd.

1. INTRODUCTION

Both the statistics and econometrics literature contain a vast amount of work on issues related to structural change, most of it specifically designed for the case of a single change. The problem of multiple structural changes has received considerably less but an increasing attention. Related literature includes Andrews, Lee and Ploberger (1996), Garcia and Perron (1996), Liu, Wu and Zidek (1997), Pesaran and Timmermann ('Model instability and choice of observation window', unpublished manuscript, 1999), Lumsdaine and Papell (1997), and Morimune and Nakagawa (1997). Most of these studies are concerned with issues related to hypothesis testing in the context of multiple changes. Recently, Bai and Perron (1998) considered estimating multiple structural changes in a linear model estimated by least-squares. They derived the rate of convergence and the limiting distributions of the estimated break points. The results are obtained under a general framework of partial structural changes which allows a subset of the parameters not to change (and, of course, includes a pure structural change model as a special case). They also addressed the important problem of testing for multiple structural changes: a sup Wald type tests for the null hypothesis of no change versus an alternative containing an arbitrary number of changes and a procedure that allows one to test the null hypothesis of, say, ℓ changes, versus the alternative

* Correspondence to: Pierre Perron, Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215, USA. E-mail: perron@bu.edu

Contract/grant sponsor: Fonds pour la Formation de Chercheurs et l'Aide à la Recherche du Québec (FCAR).

Contract/grant sponsor: National Science Foundation; Contract/grant numbers: SBR9709508; SES-0078492.

hypothesis of $\ell + 1$ changes. The latter is particularly useful in that it allows a specific to general modelling strategy to consistently determine the appropriate number of changes in the data.

The present study focuses on the empirical implementation of the theoretical results of Bai and Perron (1998), henceforth referred to as BP. We first address the problem of the estimation of the break dates and present an efficient algorithm to obtain global minimizers of the sum of squared residuals based on the principle of dynamic programming which requires at most least-squares operations of order $O(T^2)$ for any number of breaks. Our method can be applied to both pure and partial structural change models. We also consider the problem of forming confidence intervals for the break dates under various hypotheses about the structure of the data and errors across segments. In particular, we may allow the data and errors to have different distributions across segments or impose a common structure. The issue of testing for structural changes is also considered under very general conditions on the data and the errors. We discuss how the tests can be constructed allowing different serial correlation in the errors, different distribution for the data and the errors across segments or imposing a common structure. We also address the issue of estimating the number of breaks. To that effect, we discuss methods based on information criteria and a method based on a sequential testing procedure. Empirical applications are presented to illustrate the usefulness of the procedures. All methods discussed are implemented in a GAUSS program.

The rest of this paper is structured as follows. Section 2 presents the model and the estimator. Section 3 discusses in detail an algorithm, based on the principle of dynamic programming, that allows us to efficiently estimate models with multiple structural changes. Section 4 discusses the construction of confidence intervals for the various parameters, in particular the break dates. Section 5 discusses tests for multiple structural changes, methods to estimate the number of breaks and summarizes practical recommendations based on a simulation study presented in Bai and Perron ('Multiple structural change models: a simulation analysis', unpublished manuscript, 2000). Empirical applications are presented in Section 6. Some conclusions are contained in Section 7.

2. THE MODEL AND ESTIMATORS

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

$$y_t = x_t' \beta + z_t' \delta_j + u_t \quad t = T_{j-1} + 1, \dots, T_j \quad (1)$$

for $j = 1, \dots, m + 1$. In this model, 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 ($j = 1, \dots, m + 1$) are the corresponding vectors of coefficients; u_t is the disturbance at time t . The indices (T_1, \dots, T_m) , or the break points, are explicitly treated as unknown (we use the convention that $T_0 = 0$ and $T_{m+1} = T$). 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. This is a partial structural change model since the parameter vector β is not subject to shifts and is estimated using the entire sample. When $p = 0$, we obtain a pure structural change model where all the coefficients are subject to change. The variance of u_t needs not be constant. Indeed, breaks in variance are permitted provided they occur at the same dates as the breaks in the parameters of the regression.¹

¹ The existence of breaks in the variance could be exploited to increase the precision of the break date estimators. We, however, do not pursue this avenue and instead treat the variance as a nuisance parameter and focus on breaks in the

The multiple linear regression system (1) may be expressed in matrix form as

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

where $Y = (y_1, \dots, y_T)', X = (x_1, \dots, x_T)', U = (u_1, \dots, u_T)', \delta = (\delta'_1, \delta'_2, \dots, \delta'_{m+1})'$, and \bar{Z} is the matrix which diagonally partitions Z at (T_1, \dots, T_m) , i.e. $\bar{Z} = \text{diag}(Z_1, \dots, Z_{m+1})$ with $Z_i = (z_{T_{i-1}+1}, \dots, z_{T_i})'$. We denote the true value of a parameter with a 0 superscript. In particular, $\delta^0 = (\delta_1^0, \dots, \delta_{m+1}^0)'$ and (T_1^0, \dots, T_m^0) are used to denote, respectively, the true values of the parameters δ and the true break points. The matrix \bar{Z}^0 is the one which diagonally partitions Z at (T_1^0, \dots, T_m^0) . Hence, the data-generating process is assumed to be

$$Y = X\beta^0 + \bar{Z}^0\delta^0 + U \quad (2)$$

The method of estimation considered is that based on the least-squares principle. For each m -partition (T_1, \dots, T_m) , the associated least-squares estimates of β and δ_j are obtained by minimizing the sum of squared residuals

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

Let $\hat{\beta}(\{T_j\})$ and $\hat{\delta}(\{T_j\})$ denote the estimates based on the given m -partition (T_1, \dots, T_m) denoted $\{T_j\}$. Substituting these in the objective function and denoting the resulting sum of squared residuals as $S_T(T_1, \dots, T_m)$, the estimated break points $(\hat{T}_1, \dots, \hat{T}_m)$ are such that $(\hat{T}_1, \dots, \hat{T}_m) = \underset{(T_1, \dots, T_m)}{\text{argmin}} S_T(T_1, \dots, T_m)$, where the minimization is taken over all partitions (T_1, \dots, T_m) such that $T_i - T_{i-1} \geq q$.² Thus the break-point estimators are global minimizers of the objective function. The regression parameter estimates are the estimates associated with the m -partition $\{\hat{T}_j\}$, i.e. $\hat{\beta} = \hat{\beta}(\{\hat{T}_j\})$, $\hat{\delta} = \hat{\delta}(\{\hat{T}_j\})$. Since, the break points are discrete parameters and can only take a finite number of values, they can be estimated by a grid search. This method becomes rapidly computationally excessive when $m > 2$. Fortunately, there exists a very efficient method which we now discuss.

3. METHOD TO COMPUTE GLOBAL MINIMIZERS

We now consider 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. This algorithm uses at most least-squares operations of order $O(T^2)$ for any number of structural changes m , unlike a standard grid search procedure which would require least squares operations of order $O(T^m)$. The basis of the method, for specialized cases, is not new and was considered by Guthery (1974), Bellman and Roth (1969) and Fisher (1958). Nevertheless, it seems to have been forgotten, at least in the econometrics literature, and a thorough description appears useful. The

conditional mean of y_t . Hence, while permitting breaks in the variance, we do not make use of it other than to estimate the variance segment by segment when such changes are permitted.

² It is possible to relax the constraint that a segment be at least of length q by making use of generalized inverses. We, however, have not considered this extension in the algorithm presented in Section 3.

original method works only for pure structural change models; we propose a scheme that allows estimating more general partial structural change models.

3.1. The Triangular Matrix of Sums of Squared Residuals

The basic idea of the approach becomes 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 $O(T^2)$. This is presented in Figure 1 for the special case with $T = 25$ and $m = 2$ where the vertical

	Terminal date																																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25										
1	x^a	x^a	x^a	x^a	•	•	•	•	•	•	•	•	•	•	•	x^b																			
2	x^a	x^a	x^a	x^a	x^c	x^b																													
3		x^a	x^a	x^a	x^a	x^c	x^b																												
4			x^a	x^a	x^a	x^a	x^c	x^b																											
5				x^a	x^a	x^a	x^a	x^c	x^b																										
6					x^a	x^a	x^a	x^a	•	•	•	•	•	•	•	•	•	•	•	•	x^b														
7						x^a	x^a	x^a	x^a	•	•	•	•	•	•	•	•	•	•	•	x^b														
S	8						x^a	x^a	x^a	x^a	•	•	•	•	•	•	•	•	•	•	x^b														
t	9							x^a	x^a	x^a	x^a	•	•	•	•	•	•	•	•	•	x^b														
a	10								x^a	x^a	x^a	x^a	•	•	•	•	•	•	•	•	x^b														
r	11									x^a	x^a	x^a	x^a	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•							
t	12									x^a	x^a	x^a	x^a	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•						
i	13										x^a	x^a	x^a	x^a	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•						
n	14										x^a	x^a	x^a	x^a	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•					
g	15											x^a	x^a	x^a	x^a	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•				
16												x^a	x^a	x^a	x^a	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•				
D	17												x^a	x^a	x^a	x^a	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•			
a	18												x^a	x^a	x^a	x^a	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•			
t	19												x^a	x^a	x^a	x^a	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•			
e	20													x^a	x^a	x^a	x^a	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		
21														x^a	x^a	x^a	x^a	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		
22														x^a	x^a	x^a	x^a	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		
23														x^a	x^a	x^a	x^a	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		
24															x^a	x^a	x^a	x^a	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	
25																x^a	x^a	x^a	x^a	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•

Notes:

The vertical number indicates the initial date of a segment while the horizontal number indicates the terminal date. For example, the entry (4,10) indicates a segment that starts at date 4 and ends at date 10, hence having 7 observations.

x^a indicates a segment not considered since it must be at least of length 5.

x^b indicates a segment not considered since otherwise there would be no place for 3 segments of length 5.

x^c indicates a segment not considered since otherwise there would be no place for a segment of length 5 prior to it.

A • indicates an admissible segment.

Figure 1. Example of the triangular matrix of sums of squared residuals with $T = 25$, $h = 5$ and $m = 2$

axis represents the initial date of a segment and the horizontal axis the ending date. Each entry represents an estimated sum of squared residuals corresponding to the associated segment. The global sum of squared residuals for any m -partition (T_1, \dots, T_m) and for any value of m must necessarily be a particular linear combination of these $T(T+1)/2$ sums of squared residuals. The estimates of the break dates, the m -partition $(\hat{T}_1, \dots, \hat{T}_m)$, correspond to this linear combination with a minimal value. The dynamic programming algorithm can be seen as an efficient way to compare possible combinations (corresponding to different m -partitions) to achieve a minimum global sum of squared residuals.

In practice, less than $T(T+1)/2$ segments are permissible. First, some minimum distance, h , between each break may be imposed, as is done in the construction of the tests discussed in Section 5 (we suppose without loss of generality that $h \geq q$). This implies a reduction in the number of segments to be considered of $(h-1)T - (h-2)(h-1)/2$ (see Figure 1). Other reductions are possible. The largest segment must be short enough to allow m other segments before or after. For example, when the segment starts at a date between 1 and h , the maximal length of this segment is $T - hm$ when m breaks are allowed. This allows a further reduction in the total number of segments considered of $h^2m(m+1)/2$. Finally, a segment cannot start at dates 2 to h , since otherwise no segment of minimal length h could be inserted at the beginning of the sample. This allows a further reduction of $T(h-1) - mh(h-1) - (h-1)^2 - h(h-1)/2$ segments to be considered. We discuss below how this triangular matrix of sums of squared residuals can be constructed and used to obtain global minimizers for both cases of pure and partial structural change models.

3.2. The Case of a Pure Structural Change Model

We first consider the case of a pure structural change model with the regression given by:

$$Y = \bar{Z}\delta + U \quad (3)$$

In such a case, the computation of the estimates $\hat{\delta}$, \hat{u}_t and $S_T(T_1, \dots, T_m)$ can be done applying OLS segment by segment without constraints among them. The computation of the triangular matrix of sums of squared residuals 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. 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 have 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 relevant combinations (i, j) . Note that the number of matrix inversions needed is simply of order $O(T)$.

3.3. The Dynamic Programming Algorithm

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 $SSR(\{T_{r,n}\})$ be the

sum of squared residuals associated with the optimal partition containing r breaks using the first n observations. The optimal partition solves the following recursive problem:

$$SSR(\{T_{m,T}\}) = \min_{mh \leq j \leq T-h} [SSR(\{T_{m-1,j}\}) + SSR(j+1, T)] \quad (4)$$

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 (saved earlier) can be inserted to achieve a minimal sum of squared residuals. 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.

It is important to note that, in practice, this method is very fast using samples of the usual sizes. Indeed, the major computation cost is the construction of the triangular matrix of sums of squared residuals for all possible segments. The search for the optimal m -partition represents a marginal addition to the total computing time. This means that it is only marginally longer to obtain global minimizers with five or ten breaks as it is with two.

3.4. The Case of a Partial Structural Change Model

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 ($p > 0$). This is because we cannot concentrate out the parameters β without knowing the appropriate partition, i.e. the estimate of β associated with a global minimization depends on the optimal partition which we are trying to obtain. Unlike for the pure structural change model for which we can write the regression in the form (3), each element of the triangular matrix of sums of squared residuals depends on the final optimal m -partition that we search.

However, a simple iterative procedure is possible. Let $\theta = (\delta, T_1, \dots, T_m)$, 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.³

We discuss the details of this method in our context with a slight modification that permits a very rapid convergence. Note that the first step, minimizing with respect to θ keeping β fixed,

³ The convergence properties of this scheme are discussed in Sargan (1964). Of course, convergence to the global minimum is not guaranteed and a proper choice of the initial value of β might be important to avoid a local minimum.

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 $\{T^*\} = (T_1^*, \dots, T_m^*)$). The application of Sargan's method suggest that the second step be a simple linear regression with $y_t - z'_t \delta_j^*$ being the dependent variable for t in regime j ($j = 1, \dots, m+1$), the regimes being defined by the partition $\{T^*\}$.

Important efficiency gains can be obtained making a slight modification to the second step. The idea is to only keep $\{T^*\}$ fixed and to maximize again with respect to δ and β simultaneously. Hence, δ is updated at each of the two steps. The reason why this leads to important efficiency gains can be explained as follows. In general, the values $\{T^*\}$ obtained at the first iteration will be quite close to the values $\{\hat{T}\}$ that correspond to the global minimum (unless the initial value of β is very far from its true value β^0). Intuitively this is so because a misspecification in the initial value of β has little effect on the estimates $\{T^*\}$, since the latter depend mostly on the changes in the coefficients δ (associated with the z_t variables) across regimes. Consider a second step which applies the OLS regression $Y = X\beta + \bar{Z}^* \delta + U$, with \bar{Z}^* , the diagonal partition of Z at the m -partition $\{T^*\} = (T_1^*, \dots, T_m^*)$. If the values $\{T^*\}$ are equal to $\{\hat{T}\}$, corresponding to the global minimum, the estimates of β and δ from this second step are then those corresponding to the global minimum. Experiments with real and simulated data showed that, in the majority of cases, a single iteration is sufficient. In a few cases two are required but it was difficult to find examples where three were needed.

To highlight the contrast between the two methods, consider what happens if δ is not re-updated in the second step. This step becomes a simple OLS regression of the form $Y - \bar{Z}^* \delta^* = X\beta + U$. Note that even if $\{T^*\}$ is equal to $\{\hat{T}\}$, corresponding to the global minimum, the estimate of β will not necessarily be close to $\hat{\beta}$ (the value at the global minimum) unless, of course, δ^* is already close to $\hat{\delta}$ at the first iteration (which can only happen with a small probability). Hence, the need for additional iterations; and experiments on real and simulated data have shown that the number of iterations necessary can be high, even in simple models.

The convergence criterion adopted is that the change in the objective function $S_T(T_1, \dots, T_m)$ be smaller than some arbitrary ε . Using the iterative method suggested, it is possible to specify $\varepsilon = 0$ because of the discrete nature of the variables (T_1, \dots, T_m) . Indeed, in most of the experiments performed, the minimum was attained after the first iteration and the second one only verified that there was effectively no change in the objective function.

3.5. The Choice of the Initial Value for β

The efficiency of the method proposed above to achieve a global minimum (by opposition to a local minimum) depends on an appropriate 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. To be precise, write this pure structural change model as

$$y_t = x'_t \delta_{1,j} + z'_t \delta_{2,j} + u_t \quad t = T_{j-1} + 1, \dots, T_j$$

for $j = 1, \dots, m+1$. The application of the dynamic programming algorithm to this model gives estimates $(\delta_{1,j}^a, \delta_{2,j}^a; j = 1, \dots, m+1)$ and (T_1^a, \dots, T_m^a) . To obtain an initial value of the vector β , we only need to use the OLS regression $Y - \bar{Z}^a \delta_2^a = X\beta + U$, where \bar{Z}^a is the diagonal partition

of Z at the m -partition (T_1^a, \dots, T_m^a) and δ_2^a is the estimate of δ_2 . The estimate so obtained, say β^a , is used to initialize the iteration procedure.

Using this method to choose the initial value of β is justified on the grounds that the estimates $\lambda_j^a = T_j^a/T$ of the break fractions $\lambda_j^0 = T_j^0/T$ are convergent at rate T even when some coefficients do not change across regimes. All that is needed is that at least one coefficient changes at every break date. Hence, the estimate β^a obtained is asymptotically equivalent to the estimate β associated with the global minimum. This permits reaching the global minimum in very few iterations and greatly reduces the risk of reaching a local minimum. Indeed, this later problem did not occur in any of the experiments that we tried.

It may be the case that using this method to initialize the vector β is difficult to implement in practice; for example, when the dimension p of the vector β and/or the number m of changes are large. In such cases, one can always use some fixed initial values. Here, however, the problem of convergence towards a local minimum becomes more important and care should be used by applying some sensitivity analyses.

3.6. Extension to Threshold Models

The algorithm can be adapted to estimate threshold models of the form:

$$y_t = x_t' \beta + z_t' \delta_j + u_t \quad \tau_{j-1} < v_t \leq \tau_j \quad (5)$$

for $j = 1, \dots, m+1$ with the convention that $\tau_0 = -\infty$ and $\tau_{m+1} = +\infty$. Again, 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 ($j = 1, \dots, m+1$) are the corresponding vectors of coefficients; u_t is the disturbance at time t . Here, the functional form of the regression depends on the value of some observable variable v_t . This variable can be an element of the vectors x_t or z_t but need not, and it should be predetermined relative to u_t (e.g. a lagged value of the dependent variable is permitted). There are m threshold points (τ_1, \dots, τ_m) which are unknown and, hence, $m+1$ regimes. The purpose is to estimate the unknown regression coefficients together with the threshold points when T observations on (y_t, x_t, z_t) are available. This is a partial threshold model in the sense that β is not subject to shifts and is effectively estimated using the entire sample. When $p = 0$, we obtain a pure threshold model with all coefficients subject to change.

To describe the estimation method, let $v' = (v_1, \dots, v_T)$ and $v^{*'} = (v_{t_1}, \dots, v_{t_T})$ be the sorted version of v' such that $v_{t_1} \leq v_{t_2} \leq \dots \leq v_{t_T}$. The indices (t_1, \dots, t_T) are a permutation of the time indices $(1, \dots, T)$. Now, for $i = 1, \dots, m$, let T_i be the time index such that $v_{t_j} \leq \tau_i$ for all j such that $j \leq T_i$ and $v_{t_j} > \tau_i$ for all j such that $j > T_i$. The m -partition (T_1, \dots, T_m) is the partition that corresponds to the time indices of the sorted vector $v^{*'}$ when the variables v_{t_j} reach each of the m thresholds. We can write model (5) using all variables sorted according to the partition (T_1, \dots, T_m) . Then, we have, for $j = 1, \dots, T$ and $i = 1, \dots, m+1$:

$$y_{t_j} = x_{t_j}' \beta + z_{t_j}' \delta_i + u_{t_j} \quad j = T_{i-1}, \dots, T_i \quad (6)$$

(using $T_0 = 0$ and $T_{m+1} = T$). This model is in the form of a partial structural change model that we have considered. Note that this change in the time scale maintains the structure of the model even with lagged dependent variables (see Tsay, 1998). One can obtain consistent estimates of the parameters (T_1, \dots, T_m) using the dynamic programming algorithm. Let the estimate of

the partition be denoted by $(\hat{T}_1, \dots, \hat{T}_m)$; the estimates of the thresholds are then recovered as $\hat{\tau}_j = v_{t_r}$ with $r = \hat{T}_j$ for $j = 1, \dots, m$. One can then recover the estimates of β and δ_i from (5) by OLS conditioning on the threshold values.

4. CONSTRUCTING CONFIDENCE INTERVALS

A central result derived in BP concerns the convergence of the break fractions $\hat{\lambda}_i = \hat{T}_i/T$ and the rate of convergence. The results obtained show not only that $\hat{\lambda}_i$ converges to its true value λ_i^0 but that it does so at the fast rate T , i.e. $T(\hat{\lambda}_i - \lambda_i^0) = O_p(1)$ for all i . It is important, however, to note that this rate T convergence pertains to the estimated break fractions $\hat{\lambda}_i$ and not to the break dates \hat{T}_i themselves. For the latter, the result shows that with a probability arbitrarily close to 1, the distance between \hat{T}_i and T_i^0 is, in large samples, bounded by a constant independent of the sample size.

This convergence result is obtained under a very general set of assumptions allowing a wide variety of models. It, however, precludes integrated variables (with an autoregressive unit root) but permits trending regressors. The assumptions concerning the nature of the errors in relation to the regressors $\{x_t, z_t\}$ are of two kinds. First, when no lagged dependent variable is allowed in $\{x_t, z_t\}$, the conditions on the residuals are quite general and allow substantial correlation and heteroscedasticity. The second case allows lagged dependent variables as regressors but then, of course, no serial correlation is permitted in the errors $\{u_t\}$. 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 the two cases described above is potentially quite useful in dynamic models when the parameters associated with the lagged 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 non-parametric approach (e.g. leaving the dynamics in the disturbances and applying a non-parametric 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. Consider, for example, the case of a change in mean for a correlated series. When specifying $z_t = \{1\}$ and $x_t = \{\emptyset\}$, all the dynamics is contained in the error term and does not affect the impact of the change in mean on the level of the series. The change is, hence, abrupt. However, when specifying $z_t = \{1\}$ and $x_t = \{\text{lags of } y_t\}$, a change in the coefficient associated with the constant z_t is related to a change in the level of y_t that varies for the periods following the break. This change depends on the autoregressive dynamics and takes effect gradually.

4.1. Confidence Intervals for the Parameters β and δ

The fact that the quantities $\hat{\lambda}_i$ converge at the fast rate of T is enough to guarantee that the estimation of the break dates has no effect on the limiting distribution of the other parameters of the model. This permits us to recover, for these estimates, the standard \sqrt{T} asymptotic normality. More precisely, let $\hat{\theta} = (\hat{\beta}, \hat{\delta})$ and $\theta^0 = (\beta^0, \delta^0)$, then $\sqrt{T}(\hat{\theta} - \theta^0) \xrightarrow{d} N(0, V^{-1}\Phi V^{-1})$ with $V = \text{plim } T^{-1}\bar{W}^0\bar{W}^0$, $\Phi = \text{plim } T^{-1}\bar{W}^0\Omega\bar{W}^0$, $\Omega = E(UU')$, and where $\bar{W}^0 = \text{diag}(W_1^0, \dots, W_{m+1}^0)$.

is the diagonal partition, at (T_1^0, \dots, T_m^0) , of $W = (w_1, \dots, w_T)$ with $w_t = (x'_t, z'_t)$. Note that when the errors are serially uncorrelated and homoscedastic we have $\Phi = \sigma^2 V$ and the asymptotic covariance matrix reduces to $\sigma^2 V^{-1}$, which can be consistently estimated using an estimate of σ^2 . When serial correlation and/or heteroscedasticity is present, a consistent estimate of Φ can be constructed along the lines of Andrews (1991). In all cases where covariance matrix robust to heteroscedasticity and serial correlation are needed, we use Andrews's (1991) data dependent method with the Quadratic Spectral kernel and an AR(1) approximation to select the bandwidth (henceforth, referred to as the HAC estimator). We also allow in the program using pre-whitening as suggested in Andrews and Monahan (1992). Note that the correction for possible serial correlation can be made allowing the distributions of the regressors and errors to differ.

4.2. Confidence Intervals for the Break Dates

To get an asymptotic distribution for the break dates, the strategy considered is to adopt an asymptotic framework where the magnitudes of the shifts converge to zero as the sample size increases. The resulting limiting distribution is then independent of the specific distribution of the pair $\{z_t, u_t\}$. To describe the relevant distributional result, we need to define some notations. For $i = 1, \dots, m$, and $\Delta T_i^0 = T_i^0 - T_{i-1}^0$, let $\Delta_i = \delta_{i+1}^0 - \delta_i^0$, $Q_i = p \lim(\Delta T_i^0)^{-1} \sum_{t=T_{i-1}^0+1}^{T_i^0} E(z_t z'_t)$, and $\Omega_i = p \lim(\Delta T_i^0)^{-1} \sum_{r=T_{i-1}^0+1}^{T_i^0} \sum_{t=T_{i-1}^0+1}^{T_i^0} E(z_r z'_t u_r u_t)$. In the case where the data are non-trending, we have, under various assumptions stated in BP, the following limiting distribution of the break dates:

$$\frac{(\Delta_i' Q_i \Delta_i)^2}{(\Delta_i' \Omega_i \Delta_i)} (\hat{T}_i - T_i^0) \Rightarrow \arg \max_s V^{(i)}(s) \quad (i = 1, \dots, m) \quad (7)$$

where $V^{(i)}(s) = W_1^{(i)}(-s) - |s|/2$, if $s \leq 0$, $V^{(i)}(s) = \sqrt{\xi_i} (\phi_{i,2}/\phi_{i,1}) W_2^{(i)}(s) - \xi_i |s|/2$, if $s > 0$, and $\xi_i = \Delta_i' Q_{i+1} \Delta_i / \Delta_i' Q_i \Delta_i$, $\phi_{i,1}^2 = \Delta_i' \Omega_i \Delta_i / \Delta_i' Q_i \Delta_i$, $\phi_{i,2}^2 = \Delta_i' \Omega_{i+1} \Delta_i / \Delta_i' Q_{i+1} \Delta_i$. Also, $W_1^{(i)}(s)$ and $W_2^{(i)}(s)$ are independent standard Weiner processes defined on $[0, \infty)$, starting at the origin when $s = 0$. These processes are also independent across i . The cumulative distribution function of $\arg \max_s V^{(i)}(s)$ is derived in Bai (1997a) and all that is needed to compute the relevant critical values are estimates of Δ_i , Q_i , and Ω_i . These are given by $\hat{\Delta}_i = \hat{\delta}_{i+1} - \hat{\delta}_i$, $\hat{Q}_i = (\hat{\Delta}_i)^{-1} \sum_{t=\hat{T}_{i-1}^0+1}^{\hat{T}_i} z_t z'_t$, and an estimate of Ω_i can be constructed using a HAC estimator applied to the vector $\{z_t \hat{u}_t\}$ and using data over segment i only.

In practice, one may want to impose some constraints related to the distribution of the errors and regressors across segments. We then have the following cases:

- The regressors z_t are identically distributed across segments. Then $Q_i = Q_{i+1} = Q$ which can be estimated by $\hat{Q} = T^{-1} \sum_{t=1}^T z_t z'_t$. In this case, the limiting result states that $[(\hat{\Delta}_i' \hat{Q} \hat{\Delta}_i)^2 / (\hat{\Delta}_i' \hat{\Omega}_i \hat{\Delta}_i)] (\hat{T}_i - T_i^0) \Rightarrow \arg \max_s V^{(i)}(s)$, with $\xi_i = 1$.
- The errors are identically distributed across segments. Then $\Omega_i = \Omega_{i+1} = \Omega$ which can be estimated using a HAC estimator applied to $\{z_t \hat{u}_t\}$ using data over the whole sample.
- The errors and the data are identically distributed across segments. Here, we have $\xi_i = 1$, and $\phi_{i,1} = \phi_{i,2}$ and the limiting distribution reduces to $[(\hat{\Delta}_i' \hat{Q} \hat{\Delta}_i)^2 / (\hat{\Delta}_i' \hat{\Omega} \hat{\Delta}_i)] (\hat{T}_i - T_i^0) \Rightarrow \arg \max_s \{W^{(i)}(s) - |s|/2\}$, which has a density function symmetric about the origin. Here, $W^{(i)}(s)$ denotes a two-sided Brownian motion defined on \Re .

- The errors are serially uncorrelated. In this case $\Omega_i = \sigma_i^2 Q_i$ and $\phi_{i,1}^2 = \phi_{i,2}^2 = \sigma_i^2$ which can be estimated using $\hat{\sigma}_i^2 = (\Delta\hat{T}_i)^{-1} \sum_{t=\hat{T}_{i-1}+1}^{\hat{T}_i} \hat{u}_t^2$. The confidence intervals can then be constructed from the approximation

$$\frac{(\hat{\Delta}'_i \hat{Q}_i \hat{\Delta}_i)}{\hat{\sigma}_i^2} (\hat{T}_i - T_i^0) \Rightarrow \arg \max_s V^{(i)}(s) \quad (8)$$

- The errors are serially uncorrelated and the regressors are identically distributed across segments. Here $\phi_{i,1}^2 = \phi_{i,2}^2 = \sigma_i^2$ and $\xi_i = 1$. The confidence intervals can then be constructed from the approximation

$$\frac{(\hat{\Delta}'_i \hat{Q}_i \hat{\Delta}_i)}{\hat{\sigma}_i^2} (\hat{T}_i - T_i^0) \Rightarrow \arg \max_s \{W^{(i)}(s) - |s|/2\} \quad (9)$$

- The errors are serially uncorrelated and identically distributed across segments. The approximation is the same as (8) with $\hat{\sigma}^2 = T^{-1} \sum_{t=1}^T \hat{u}_t^2$ instead of $\hat{\sigma}_i^2$.
- The errors are serially uncorrelated and both the data and the errors are identically distributed across segments. The approximation is the same as (9) with $\hat{\sigma}^2$ instead of $\hat{\sigma}_i^2$.

All the cases discussed above are allowed as options in the accompanying computer program. Since the break dates are integer valued, we consider confidence intervals that are likewise integer-valued by using the highest smaller integer for the lower bound and the smallest higher integer for the upper bound.

4.3. The Case with Trending Regressors

Simple modifications can be applied to deal with the case of trending regressors. Suppose that we have regressor z_t of the form $z_t = [g_1(t/T), \dots, g_q(t/T)]$, with $g_i(t/T)$ having bounded derivatives on $[0,1]$. For example, in the case of a polynomial trend function, $g_i(t/T) = (t/T)^i$. Then, (see Bai, 1997a)

$$\frac{\hat{\Delta}'_i g(\hat{\lambda}_i) g(\hat{\lambda}_i)' \hat{\Delta}_i}{\psi_i} \Rightarrow \begin{cases} W_1^{(i)}(-s) - |s|/2 & \text{if } s \leq 0 \\ \sqrt{\psi_{i+1}/\psi_i} W_2^{(i)}(s) - |s|/2 & \text{if } s > 0 \end{cases}$$

where $g(\hat{\lambda}) = [g_1(\hat{T}_i/T), \dots, g_q(\hat{T}_i/T)]$ and $\psi_i = p \lim(\Delta T_i^0)^{-1} \sum_{r=T_{i-1}^0+1}^{T_i^0} \sum_{t=T_{i-1}^0+1}^{T_i^0} E(u_r u_t)$. If the errors have the same distribution across segments, we have $[\hat{\Delta}'_i g(\hat{\lambda}_i) g(\hat{\lambda}_i)' \hat{\Delta}_i / f_u(0)] \Rightarrow \arg \max_s \{W^{(i)}(s) - |s|/2\}$, where $f_u(0)$ is (2π times) the spectral density function of u_t at frequency zero which can be consistently estimated using standard kernel methods. If u_t is uncorrelated, $f_u(0)$ is replaced by $\sigma^2 = p \lim T^{-1} \sum_{t=1}^T E(u_t^2)$ estimated by $\hat{\sigma}^2 = T^{-1} \sum_{t=1}^T \hat{u}_t^2$.

5. TEST STATISTICS FOR MULTIPLE BREAKS

5.1. A Test of No Break versus a Fixed Number of Breaks

We consider the supF type test of no structural break ($m = 0$) versus $m = k$ breaks. Let (T_1, \dots, T_k) be a partition such that $T_i = [T\lambda_i]$ ($i = 1, \dots, k$). Let R be the conventional matrix

such that $(R\delta)' = (\delta'_1 - \delta'_2, \dots, \delta'_k - \delta'_{k+1})$. Define

$$F_T(\lambda_1, \dots, \lambda_k; q) = \frac{1}{T} \left(\frac{T - (k + 1)q - p}{kq} \right) \hat{\delta}' R'(R\hat{V}(\hat{\delta})R')^{-1} R\hat{\delta} \quad (10)$$

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

$$V(\hat{\delta}) = p \lim T(\bar{Z}' M_X \bar{Z})^{-1} \bar{Z}' M_X \Omega M_X \bar{Z} (\bar{Z}' M_X \bar{Z})^{-1} \quad (11)$$

Following Andrews (1993) and others, the test is $\sup F_T(k; q) = F_T(\hat{\lambda}_1, \dots, \hat{\lambda}_k; q)$ where $\hat{\lambda}_1, \dots, \hat{\lambda}_k$ minimize the global sum of squared residuals which is equivalent to maximizing the F-test assuming spherical errors. This is asymptotically equivalent to, and yet much simpler to construct than, maximizing the F-test (10) since the estimated break dates are consistent even in the presence of serial correlation. The asymptotic distribution depends on a trimming parameter via the imposition of the minimal length h of a segment, namely $\varepsilon = h/T$.

Various versions of the tests can be obtained depending on the assumptions made with respect to the distribution of the data and the errors across segments. These relate to different specifications in the construction of the estimate of $V(\hat{\delta})$ given by (11). In the case of a partial structural change model ($p \neq 0$), we consider three specifications.

- Allowing for serial correlation, different distributions for the data across segments and the same distribution for the errors across segments. The estimate is then $\hat{V}(\hat{\delta}) = (T^{-1} \bar{Z}' M_X \bar{Z})^{-1} \hat{K}_T (T^{-1} \bar{Z}' M_X \bar{Z})^{-1}$ where \hat{K}_T is the HAC estimator of the $(m + 1)q$ vector $\{z_t^* \hat{u}_t\}$ where z_t^* are the elements of the matrix $M_X \bar{Z}$.
- Serially uncorrelated errors, different variances of the errors and different distributions of the data across segments. The estimate is $\hat{V}(\hat{\delta}) = T(\bar{Z}' M_X \bar{Z})^{-1} \hat{\Upsilon} (\bar{Z}' M_X \bar{Z})^{-1}$, with $\hat{\Upsilon} = \sum_{i=1}^{m+1} \hat{\sigma}_i^2 \sum_{t=\hat{T}_{i-1}+1}^{\hat{T}_i} z_t^* z_t^{*\prime}, \hat{\sigma}_i^2 = (\Delta \hat{T}_i)^{-1} \sum_{t=\hat{T}_{i-1}+1}^{\hat{T}_i} \hat{u}_t^2, Z^{*\prime} = (z_1^*, \dots, z_T^*)$ with $Z^* = M_X \bar{Z}$.
- Serially uncorrelated errors, different distributions for the data across segments and the same distribution for the errors across segments. In this case, $V(\hat{\delta}) = \sigma^2 (T^{-1} \bar{Z}' M_X \bar{Z})^{-1}$ which can be estimated using $\hat{\sigma}^2 = T^{-1} \sum_{t=1}^T \hat{u}_t^2$.

In the case of a pure structural change model, we consider more possible specifications on how to estimate the relevant asymptotic covariance matrix. They are the following:

- No serial correlation, different distributions for the data and identical distribution for the errors across segments. In this base case, the estimate is $\hat{V}(\hat{\delta}) = \hat{\sigma}^2 (T^{-1} \bar{Z}' \bar{Z})^{-1}$.
- No serial correlation in the errors, different variances of the errors and different distributions of the data across segments. In this case, $\hat{V}(\hat{\delta}) = \text{diag}(\hat{V}(\hat{\delta}_1), \dots, \hat{V}(\hat{\delta}_{m+1}))$, where $\hat{V}(\hat{\delta}_i)$ uses only data from segment i , i.e. $\hat{V}(\hat{\delta}_i) = \hat{\sigma}_i^2 [(\Delta \hat{T}_i)^{-1} \sum_{t=\hat{T}_{i-1}+1}^{\hat{T}_i} z_t z_t^{*\prime}]^{-1}$ with $\hat{\sigma}_i^2 = (\Delta \hat{T}_i)^{-1} \sum_{t=\hat{T}_{i-1}+1}^{\hat{T}_i} \hat{u}_t^2$. These are simply the OLS estimates obtained using data from each segment separately.
- Serial correlation in the errors, different distributions for the data and the errors across segments. Here, we make use of the fact that the errors in different segments are asymptotically independent. Hence, the limiting variance is given by $V(\hat{\delta}) = \text{diag}(V(\hat{\delta}_1), \dots, V(\hat{\delta}_{m+1}))$, where $V(\hat{\delta}_i) = p \lim (\Delta \hat{T}_i) (Z_i' Z_i)^{-1} Z_i' \Omega_i Z_i (Z_i' Z_i)^{-1}$. This can be consistently estimated, segment by segment, with a HAC estimator of $V(\hat{\delta}_i)$ using only data from segment i .

- Serial correlation in the errors, same distribution for the errors across segments. In this case the limiting covariance matrix is $V(\hat{\delta}) = p \lim T(\bar{Z}'\bar{Z})^{-1}(\Lambda \otimes (Z'\Omega Z))(\bar{Z}'\bar{Z})^{-1}$ where (using the convention that $\lambda_0 = 0$ and $\lambda_{m+1} = 1$) $\Lambda = \text{diag}(\lambda_1 - \lambda_0, \dots, \lambda_{m+1} - \lambda_m)$. This can be consistently estimated using $\hat{\lambda}_i = \hat{T}_i/T$ and a HAC estimator of $Z'\Omega Z$ based on $\{z_t \hat{u}_t\}$ constructed using the full sample.

In the construction of the tests we do not consider imposing the restriction that the distribution of the regressors z_t be the same across segments even if they are (except as they enter in the construction of a HAC estimate involving the pair $\{z_t \hat{u}_t\}$). This might seem surprising since imposing a valid restriction should lead to more precise estimate. This is, however, not the case. Consider the case with no serial correlation in the errors and the same distribution for the errors across segments. Imposing the restriction that the distribution of the regressors z_t be the same across segments leads to the asymptotic covariance matrix $V(\hat{\delta}) = \sigma^2(\Lambda \otimes Q)^{-1}$, where $Q = \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T E(z_t z_t')$. Note that a consistent estimate can be obtained using $\hat{Q} = T^{-1} \sum_{t=1}^T z_t z_t'$, $\hat{\sigma}^2 = T^{-1} \sum_{t=1}^T \hat{u}_t^2$ and $\hat{\Lambda}$ constructed using $\hat{\lambda}_i = \hat{T}_i/T$ ($i = 1, \dots, m$). Suppose that the z 's are exogenous and the errors have the same variance across segments. Then, for a given partition (T_1, \dots, T_m) , the exact variance of $\hat{\delta}$ is $V(\hat{\delta}) = \sigma^2(\bar{Z}'\bar{Z}/T)^{-1}$. Using the asymptotic version may imply an inaccurate approximation especially if small segments are allowed, in which case the exact moment matrix of the regressors may deviate substantially from its full-sample analogue.

The same problem occurs in the case with no serial correlation in the errors and different variances for the residuals across segments. Imposing the restriction that the distribution of the regressors z_t be the same across segments gives the limiting variance $V(\hat{\delta}) = (\Lambda^* \otimes Q)^{-1}$ where $\Lambda^* = \text{diag}(\sigma_1^2(\lambda_1 - \lambda_0), \dots, \sigma_{m+1}^2(\lambda_{m+1} - \lambda_m))$ which can be consistently estimated using \hat{Q} , $\hat{\lambda}_i = \hat{T}_i/T$ and $\hat{\sigma}_i^2 = (\Delta \hat{T}_i)^{-1} \sum_{t=\hat{T}_{i-1}+1}^{\hat{T}_i} \hat{u}_t^2$. Again, imposing the constraint that $Z_i'Z_i/(\Delta \hat{T}_i)$ be approximated by \hat{Q} over all segments may imply a poor approximation in finite samples. We have found, in these two cases, that imposing a common distribution for the regressors across segments leads to tests with worse properties even when the data indeed have an invariant distribution. These distortions becomes less important, however, when the sample size is large and/or the trimming ε is large.

The relevant asymptotic distribution has been derived in BP and critical values were provided for a trimming $\varepsilon = 0.05$ and values of k from 1 to 9 and values of q from 1 to 10. As discussed in Bai and Perron (2000), a trimming as small as 5% of the total sample can lead to tests with substantial size distortions when allowing different variances of the errors across segments or when serial correlation is permitted. This is because one is then trying to estimate various quantities using very few observations; for example, if $T = 100$ and $\varepsilon = 0.05$, one ends up estimating, for some segments, quantities like the variance of the residuals using only 5 observations. Similarly, with serial correlation a HAC estimator would need to be applied to very short samples. The estimates are then highly imprecise and the tests accordingly show size distortions. When allowing different variances across segments or serial correlation, a higher value of ε should be used. Hence, the case with no serial correlation and homogenous errors should be considered the base case in which the tests can be constructed using an arbitrary small trimming ε . For all other cases, care should be exercised in the choice of ε and larger values should be considered. For that purpose, we supplemented the critical values tabulated in BP with similar ones for $\varepsilon = 0.10, 0.15, 0.20$ and 0.25 . The results are presented in Bai and Perron ('Additional critical values for multiple structural

changes tests', unpublished manuscript, 2001). Note that when $\varepsilon = 0.10$ the maximum number of breaks considered is 8 since allowing 9 breaks impose the estimates to be exactly $\hat{\lambda}_1 = 0.1$, $\hat{\lambda}_2 = 0.2$ up to $\hat{\lambda}_9 = 0.9$. For similar reasons, the maximum number of breaks allowed is 5 when $\varepsilon = 0.15$, 3 when $\varepsilon = 0.20$ and 2 when $\varepsilon = 0.25$.

Note that the asymptotic theory for these tests in BP is valid only for the case of non-trending data. The case with trending data, discussed in Bai (1999), yields different asymptotic distributions. However, the asymptotic distributions in the two cases are fairly similar, especially in the tail where critical values are obtained. Hence, one can safely use the same critical values. Using simulations, we found the size distortions to be minor.

5.2. Double Maximum Tests

Often, an investigator wishes not to pre-specify a particular number of breaks to make inference. To allow this BP have introduced two tests of the null hypothesis of no structural break against an unknown number of breaks given some upper bound M . These are called the *double maximum tests*. The first is an equal weighted version defined by $UD \max F_T(M, q) = \max_{1 \leq m \leq M} F_T(\hat{\lambda}_1, \dots, \hat{\lambda}_m; q)$, where $\hat{\lambda}_j = \hat{T}_j/T$ ($j = 1, \dots, m$) are the estimates of the break points obtained using the global minimization of the sum of squared residuals. The second test applies weights to the individual tests such that the marginal p -values are equal across values of m and is denoted $WD \max F_T(M, q)$; see BP for details. Critical values were provided for $M = 5$ and $\varepsilon = 0.05$ in BP. A value $M = 5$ 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. Bai and Perron (2001) provide additional critical values for $\varepsilon = 0.10$ ($M = 5$), 0.15 ($M = 5$), 0.20 ($M = 3$) and 0.25 ($M = 2$).

5.3. A Test of ℓ versus $\ell + 1$ Breaks

BP proposed a test for ℓ versus $\ell + 1$ breaks, labelled $\sup F_T(\ell + 1 | \ell)$. The method 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 applied to each segment containing the observations \hat{T}_{i-1} to \hat{T}_i ($i = 1, \dots, \ell + 1$). We conclude for a rejection in favour 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 breaks model. The break date thus selected is the one associated with this overall minimum. The estimates \hat{T}_i need not be the global minimizers of the sum of squared residuals, one can also use sequential one at a time estimates which allows the construction of a sequential procedure to select the number of breaks (see Bai, 1997b).

Asymptotic critical values were provided by BP for a trimming $\varepsilon = 0.05$ for q ranging from 1 to 10, and Bai and Perron (2001) present additional critical values for $\varepsilon = 0.10, 0.15, 0.20$ and 0.25. Note that, unlike for the $\sup F_T(k; q)$ test, we do not need to impose similar restrictions on the number of breaks for different values of the trimming ε .⁴ Of course, all the same options

⁴ However, considering more than $\text{int}[1/\varepsilon] - 2$ breaks implies changing ε as one progresses through the sequential procedure. For example, one could use a trimming $\varepsilon = 0.05$ and find 6 breaks in the first half of the sample, then switch to a trimming of $\varepsilon = 0.20$ to test for a seventh break. The accompanying computer program does not incorporate the possibility of such switch and, hence, in this case the same constraints as for the $\sup F_T(k; q)$ test on the maximum number of breaks apply.

are available as for the previous tests concerning the potential specifications of the nature of the distributions for the errors and the data across segments.

5.4. Estimating the Number of Breaks

A common procedure to select the dimension of a model is to consider an information criterion. Yao (1988) suggests the use of the Bayesian Information Criterion (BIC) while Liu *et al.* (1997) propose a modified Schwarz criterion (LWZ). Perron (1997) presented a simulation study of the behaviour of these two information criteria and of the AIC in the context of estimating the number of changes in the trend function of a series in the presence of serial correlation. The results first showed the AIC to perform very badly. The BIC and LWZ perform reasonably well in the absence of serial correlation in the errors but chooses a much higher value than the true one in the presence of serial correlation. When no serial correlation is present in the errors but a lagged dependent variable is present, the BIC performs badly when the coefficient on the lagged dependent variable is large. In such cases, the LWZ performs better under the null of no break but underestimate the number of breaks when some are present. The method suggested by BP is based on the sequential application of the $\sup F_T(\ell + 1|\ell)$ test using the sequential estimates of the breaks.

5.5. Summary and Practical Recommendations

Bai and Perron (2000) present an extensive simulation analysis pertaining to the size and power of the tests, the accuracy of the asymptotic approximations for the confidence intervals and the relative merits of different methods to estimate the number of breaks. The methods are shown to be adequate, in general, but care must be taken when using particular specifications. The following recommendations are made:

- First, ensure that the specifications are such that the size of the test is adequate under the hypothesis of no break. If serial correlation and/or heterogeneity in the data or errors across segments are not allowed in the estimated regression model (and not present in the DGP), using any value of the trimming ε will lead to tests with adequate sizes. However, if such features are allowed, a higher trimming is needed. With a sample of $T = 120$, $\varepsilon = 0.15$ should be enough for heterogeneity in the errors or the data. If serial correlation is allowed, $\varepsilon = 0.20$ may be needed. These could be reduced if larger sample sizes are available.
- Overall, selecting the break point using the BIC works well when breaks are present but less so under the null hypothesis of no break, especially if serial correlation is present. The method based on the LWZ criterion works better under the null hypothesis (even with serial correlation) by imposing a higher penalty. However, this higher penalty translates into a very bad performance when breaks are present. Also, model selection procedures based on information criteria cannot take into account potential heterogeneity across segments unlike the sequential method. Overall, the sequential procedure works better.
- There are instances where the sequential procedure can be improved. The problem is that, in the presence of multiple breaks, certain configurations of changes are such that it is difficult to reject the null hypothesis of 0 versus 1 break but it is not difficult to reject the null hypothesis of 0 versus a higher number of breaks (this occurs, for example, when 2 changes are present and the value of the coefficient returns to its original value after the second break). In such

cases, the sequential procedure breaks down. A useful strategy is to first look at the *UD max* or *WD max* tests to see if at least one break is present. If these indicate the presence of at least one break, then the number of breaks can be decided based upon a sequential examination of the $\sup F(\ell + 1|\ell)$ statistics constructed using global minimizers for the break dates (i.e. ignore the test $F(1|0)$ and select m such that the tests $\sup F(\ell + 1|\ell)$ are insignificant for $\ell \geq m$). This method leads to the best results and is recommended for empirical applications. Its usefulness is illustrated in Section 6.2.

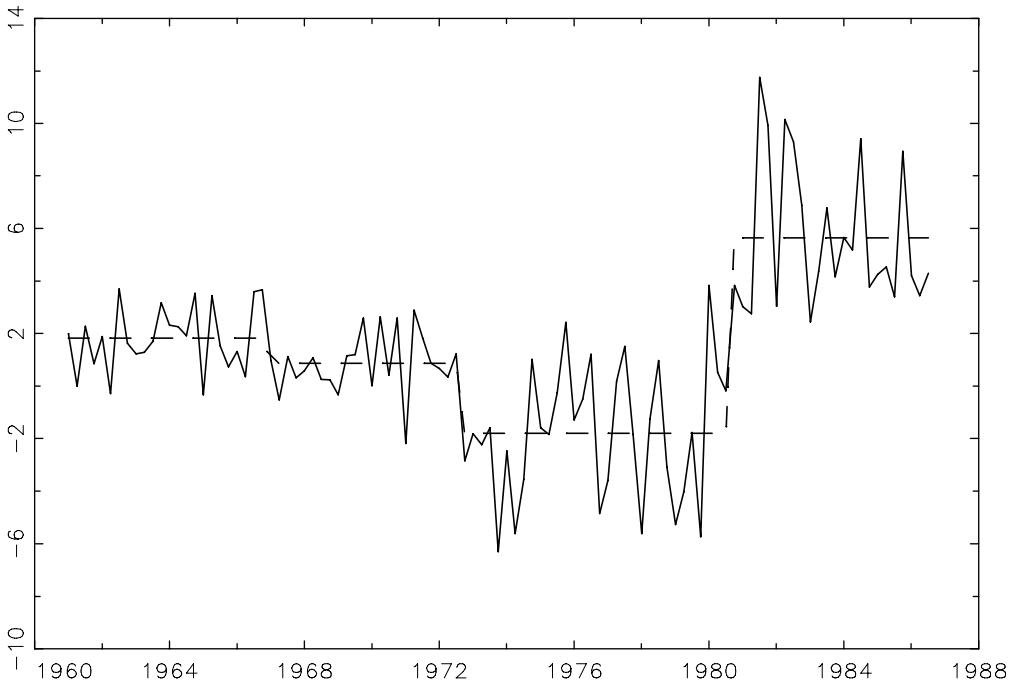
- In general, non-symmetric confidence intervals for break dates provide better coverage rates than symmetric confidence intervals when data are non-stationary (the second moment of regressors are non-constant). This case includes structural changes in autoregressive models because a change in the intercept or a change in the AR coefficient implies non-constant second moment of the observable variables. Indeed, Monte Carlo simulations show asymmetric distributions for the estimated break points for AR models. These asymmetric distributions are well approximated by the asymptotic distributions discussed in Section 4.2.
- The coverage rates for the break dates are adequate unless the break is either too small (so small as not to be detected by the tests) or too large. This is, from a practical point of view, an encouraging result. The confidence intervals are inadequate (in that they miss the true break value too often) exactly in those cases where it would be quite difficult to conclude that a break is present (in which case they would not be used anyway). When the breaks are very large the confidence intervals do contain the true values but lead to a conservative assessment of the accuracy of the estimates. It was found that correcting for heterogeneity in the data and/or errors across segments yields improvements over a more straightforward uncorrected interval. Correcting for serial correlation also does lead to substantial improvements.

6. EMPIRICAL APPLICATIONS

We discuss two applications of the procedures presented. The first analyses the US *ex-post* real interest rate series considered by Garcia and Perron (1996). The second re-evaluates some findings of Alogoskoufis and Smith (1991) who analysed changes in the persistence of inflation and shifts in an expectations-augmented Phillips curve resulting from such changes.

6.1. The US *Ex-post* Real Interest Rate

Garcia and Perron (1996) considered the US *ex-post* real interest rate (the three-month treasury bill rate deflated by the CPI inflation rate taken from the Citibase data bank). The data are quarterly and the sample is 1961:1–1986:3. Figure 2 presents a graph of the series. Of interest is the presence of abrupt 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 account for potential serial correlation via non-parametric adjustments (see the discussion in Section 4). We allowed up to 5 breaks and used a trimming $\varepsilon = 0.15$, hence each segment has at least 15 observations. We also allowed serial correlation in the errors and different variances of the residuals across segments. The results are presented in Table I.

Figure 2. US *ex-post* real interest rate 1961:1–1986:3

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 the value 34.31 and is therefore highly significant. The $F_T(3|2)$ test has value 14.32 which is also significant at the 5% level. The sequential procedure (using a 5% significance level) selects 3 breaks while the BIC and the modified Schwarz criterion of Liu *et al.* (1997) selects two breaks. Given the documented facts that the information criteria are biased downward and that the sequential procedure and the $F_T(\ell + |\ell)$ perform better in this case, we conclude in favour of the presence of three breaks.

Of direct interest are the estimates obtained under global minimization. The break dates are estimated at 1966:4, 1972:3 and 1980:3. The first date has a rather large confidence interval (between 1964:1 and 1969:2 at the 95% significance level). The other break dates are, however, precisely estimated since the 95% confidence intervals cover only a few quarters before and after. The differences in the estimated means over each segment are significant and point to a decrease of 0.95% in 1966:3, another decrease of 2.67% in late 1972 and a large increase of 7.44% in late 1980. These results contrasts with those of Garcia and Perron (1996) who found only two breaks. This points to the fact that our procedure may be more powerful than the regime switching method they used to detect abrupt changes in level. In particular, the difference in results is largely due to the fact that allowance is made for different error structures across segments.⁵

⁵ Garcia and Perron (1996) only allowed the mean of the series and the variance of the errors to be state dependent, the autoregressive parameters being constrained to be fixed across regimes. If no allowance is made for the possibility

Table I. Empirical results: US *ex-post* real interest rate (1961:1–1986:3)

$z_t = \{1\}$	$q = 1$	$p = 0$	Specifications			
			$h = 15$	$M = 5$		
			Tests ¹			
Sup $F_T(1)$ 57.91*	Sup $F_T(2)$ 43.01*	Sup $F_T(3)$ 33.22*	Sup $F_T(4)$ 24.77*	Sup $F_T(5)$ 18.33*	UDmax 57.91*	WDmax 57.91*
Sup $F(2 1)$ 33.93*	sup $F(3 2)$ 14.72*	sup $F(4 2)$ 0.03				
Number of breaks selected ²						
Sequential	3					
LWZ	2					
BIC	2					
Estimates with Three Breaks ³						
$\hat{\delta}_1$ 1.82 (0.19)	$\hat{\delta}_2$ 0.87 (0.16)	$\hat{\delta}_3$ −1.80 (0.51)	$\hat{\delta}_4$ 5.64 (0.60)	\hat{T}_1 66:4 (64:1–69:2)	\hat{T}_2 72:3 (70:3–72:4)	\hat{T}_3 80:3 (79:4–81:1)

Notes:

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 heteroscedasticity 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 5% size for the sequential test sup $F_T(\ell + 1|\ell)$.

³ In parentheses are the standard errors (robust to serial correlation) for $\hat{\delta}_i$ ($i = 1, \dots, 4$) and the 95% confidence intervals for \hat{T}_i ($i = 1, 2, 3$).

* Significance at the 5% level.

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

Alogoskoufis and Smith (1991) considered the 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

$$E(\Delta p_t | I_{t-1}) = \delta_1 + \delta_2 \Delta p_{t-1} \quad (12)$$

Hence, upon substitution, the Phillips curve is:

$$\Delta w_t = \gamma_1 + \gamma_2 \Delta p_{t-1} + \gamma_3 \Delta u_t + \gamma_4 u_{t-1} + \xi_t \quad (13)$$

where $\gamma_1 = \alpha_1 \delta_1$ and $\gamma_2 = \alpha_2 \delta_2$. 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

of changes in the structure of the correlation in the errors, applying our method leads to the same conclusion as in Garcia and Perron (1996), namely two breaks in 1972:3 and 1980:3. Allowing the structure of the correlation to differ across segments show that the errors are negatively correlated in the 60's and basically *i.i.d.* in the 1970s and 1980s. This negative correlation in the 1960s allows a more precise estimate of the break date by reducing its variance and, accordingly, the tests also have higher power.

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 (13) exhibits 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 re-evaluate Alogoskoufis and Smith's (1991) claim using post-war annual data for the United Kingdom.⁶ Consider first the structural stability of the AR(1) representation of the inflation series whose graph is depicted in Figure 3. When applying a one-break model (not reported), we indeed find the same results, namely a structural change in 1967 with δ_2 increasing from 0.274 to 0.739 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 do not support a one-break model. A feature of substantial importance is that a look at the graph of the inflation series suggests different variability in different periods. To that effect, we have investigated the stability of the inflation process allowing different variances for the residuals across segments. Details of the estimation results are contained in Table II. Again, the sup $F_T(1)$ test is not significant at any conventional level, but the sup $F_T(2)$ test is, however, significant at

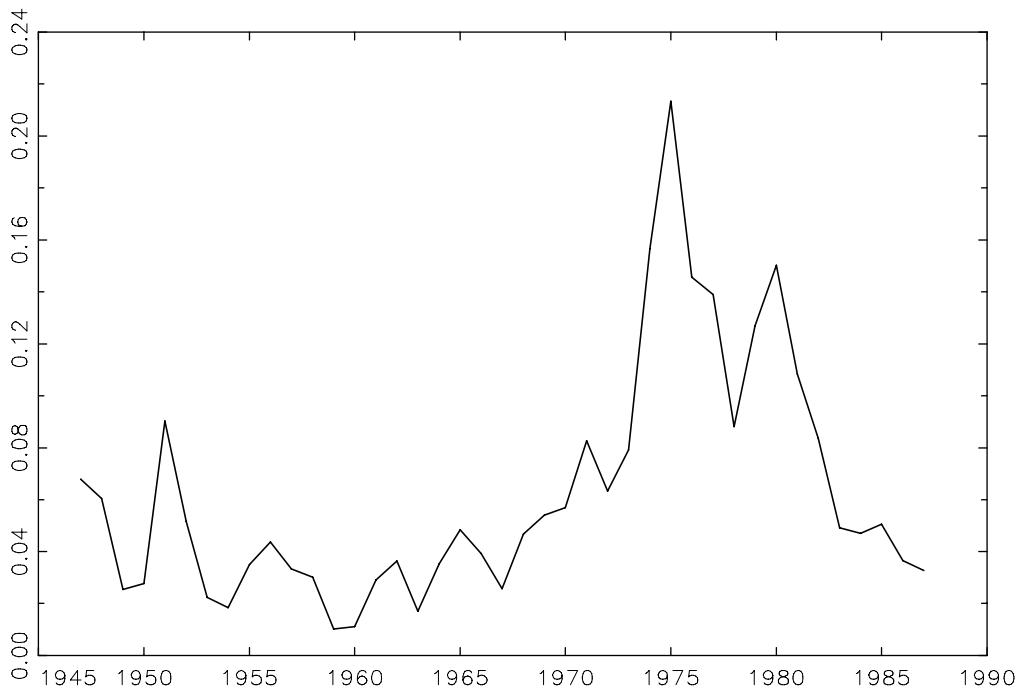


Figure 3. POST-war UK inflation rate 1947–1987

⁶ 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.

Table II. Empirical results: UK CPI inflation rate 1947–1987

$z_t = \{1, y_{t-1}\}$	$q = 2$	$p = 0$	$h = 8$	$M = 3$
$het_u = 1$				$\varepsilon = 0.20$
Sup $F_T(1)$	Sup $F_T(2)$	Sup $F_T(3)$	Sup $F_T(2 1)$	Sup $F_T(3 2)$
8.50	9.88 ^a	6.74 ^b	10.22 ^b	1.25
UD max	WD max (10%)	WD max (5%)		
9.88 ^b	11.71 ^b	12.08		
Tests				
Sequential procedure	Number of breaks selected			
LWZ	0			
BIC	0			
Parameter Estimates with Two Breaks				
$\hat{\delta}_{1,1}$	$\hat{\delta}_{1,2}$	$\hat{\delta}_{1,3}$	\hat{T}_1	\hat{T}_2
0.024	0.00	0.018	1967	1975
(0.008)	(0.020)	(0.016)	(1964–1968)	(1969–1981)
$\hat{\delta}_{2,1}$	$\hat{\delta}_{2,2}$	$\hat{\delta}_{2,3}$		
0.274	1.34	0.684		
(0.200)	(0.250)	(0.136)		

^{a,b} A statistic significant at the 5% and 10% levels, respectively.

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 $\ell \geq 2$. Since the sup $F_T(1)$ test is not significant, it is not surprising that the sequential procedure selects zero break; the BIC and LWZ also select zero break. However, the sup $F_T(2)$, the UDmax, the WDmax and the sup $F_T(2|1)$ tests being all significant, the results, overall, suggest a model with two breaks.

Nevertheless, the estimates of a two-breaks model reveal a similar picture as that suggested by Alogoskoufis and Smith (1991). The first break date is the same as in the one-break model, namely 1967, which is linked to the end of the Bretton Woods system. The second break is located in 1975. The coefficient estimates point to the importance of shifts in the persistence of inflation. Indeed, the coefficient δ_2 varies from 0.274 to 1.34 in 1967. It is, however, back to 0.684 after 1975 suggesting that the effect of the abandonment of the Bretton Woods system was short-lasting.

Since there are 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. Here, the setup involves a partial structural change model since changes in the inflation process should only affect the coefficients γ_1 and γ_2 with no effect on the coefficients γ_3 and γ_4 . The results are presented in Table III. The evidence points strongly to a two-breaks model with exactly the same break dates as for the inflation process (1967 and 1975). The sup $F_T(2|1)$ is significant as well as the sup $F_T(k)$ tests for all k . The sequential method, the BIC and the LWZ all select 2 as the number of breaks. Finally, the UD max and WD max tests are also highly significant. Furthermore, the coefficient γ_2 (associated with the lagged inflation) move in the same direction as the persistence of inflation; in particular there is a substantial increase in this coefficient in 1967 from 0.094 to 1.23 (following a change in persistence from 0.274 to 1.34). In 1975, γ_2 shows

Table III. Empirical results: Phillips curve equation

$y_t = \{\Delta w_t\}$	$q = 2$	$p = 2$	$z_t = \{1, \Delta p_{t-1}\}$	$x_t = \{\Delta u_t, u_{t-1}\}$
$het_u = 0$	$\varepsilon = 0.10$	$h = 4$	$M = 5$	$\varepsilon = 0.10$
$SupF_T(1)$ 22.84 ^a	$SupF_T(2)$ 25.77 ^a	$SupF_T(3)$ 20.76 ^a	$SupF_T(4)$ 17.19 ^a	
$SupF_T(2 1)$ 24.39 ^a	$SupF_T(3 2)$ 4.98	$SupF_T(4 3)$ 4.98	$UD\ max$ 25.77 ^a	$WD\ max(1\%)$ 32.34 ^a
Number of breaks selected				
Sequential procedure	2			
LWZ	2			
BIC	2			
Parameter estimates with two breaks				
$\hat{\gamma}_{1,1}$ 0.066 (0.012)	$\hat{\gamma}_{1,2}$ 0.062 (0.019)	$\hat{\gamma}_{1,3}$ 0.181 (0.054)	\hat{T}_1 1967 (1965–1968)	\hat{T}_2 1975 (1973–1976)
$\hat{\gamma}_{2,1}$ 0.094 (0.240)	$\hat{\gamma}_{2,2}$ 1.23 (0.205)	$\hat{\gamma}_{2,3}$ 0.015 (0.257)		
$\hat{\gamma}_3$ −0.141 (0.581)	$\hat{\gamma}_4$ −0.877 (0.373)			

^aA statistic significant at the 1% level.

a substantial decrease in agreement with the decrease in the persistence of inflation. Overall, the results confirm the conclusions of Alogoskoufis and Smith (1991) and provide support for the Lucas critique.

7. CONCLUSIONS

This paper has presented a comprehensive treatment of practical issues arising in the analysis of models with multiple structural changes. Of considerable interest is a dynamic programming algorithm which makes possible efficient computation of the estimates of the break points as global minimizers of the sum of squared residuals. We have also discussed methods to construct confidence intervals for the break dates, test statistics and model selection procedures. All procedures discussed are available as options in a GAUSS program available for non-profit academic purposes.

ACKNOWLEDGEMENTS

This paper was in part the basis of the 1998 Jacob Marschak Lecture of the Econometric Society presented by Perron at the Latin American Econometric Society Meeting in Lima, Peru on 14 August 1998. The code discussed in this paper is available for non-profit academic use at the *Journal of Applied Econometrics*' Data Archive and also at <http://econ.bu.edu/perron>. Perron

acknowledges financial support from the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche du Québec (FCAR). We acknowledge financial support from the National Science Foundation under Grants SBR9709508 (Bai) and SES-0078492 (Perron). We wish to thank Maxwell King for useful comments.

REFERENCES

- Alogoskoufis GS, Smith R. 1991. The Phillips curve, the persistence of inflation, and the Lucas critique: evidence from exchange rate regimes. *American Economic Review* **81**: 1254–1275.
- Andrews DWK. 1991. Heteroscedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**: 817–858.
- Andrews DWK. 1993. Tests for parameter instability and structural change with unknown change point. *Econometrica* **61**: 821–856.
- Andrews DWK, Monahan JC. 1992. An improved heteroscedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* **60**: 953–966.
- Andrews DWK, Lee I, Ploberger W. 1996. Optimal changepoint tests for normal linear regression. *Journal of Econometrics* **70**: 9–38.
- Bai J. 1997a. Estimation of a change point in multiple regression models. *Review of Economic and Statistics* **79**: 551–563.
- Bai J. 1997b. Estimating multiple breaks one at a time. *Econometric Theory* **13**: 315–352.
- Bai J. 1999. Likelihood ratio tests for multiple structural changes. *Journal of Econometrics* **91**: 299–323.
- Bai J, Perron P. 1998. Estimating and testing linear models with multiple structural changes. *Econometrica* **66**: 47–78.
- Bellman R, Roth R. 1969. Curve fitting by segmented straight lines. *Journal of the American Statistical Association* **64**: 1079–1084.
- Brown RL, Durbin J, Evans JM. 1975. Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society, Series B* **37**: 149–192.
- Fisher WD. 1958. On grouping for maximum homogeneity. *Journal of the American Statistical Association* **53**: 789–798.
- Garcia R, Perron P. 1996. An analysis of the real interest rate under regime shifts. *Review of Economics and Statistics* **78**: 111–125.
- Guthery SB. 1974. Partition regression. *Journal of the American Statistical Association* **69**: 945–947.
- Liu J, Wu S, Zidek JV. 1997. On segmented multivariate regressions. *Statistica Sinica* **7**: 497–525.
- Lumsdaine RL, Papell DH. 1997. Multiple trend breaks and the unit root hypothesis. *Review of Economics and Statistics* **79**: 212–218.
- Morimune K, Nakagawa M. 1997. Unit root tests which allow for multiple breaks. Discussion Paper No. 457, Kyoto Institute of Economic Research, Kyoto University.
- Perron P. 1997. L'estimation de modèles avec changements structurels multiples. *Actualité Économique* **73**: 457–505.
- Sargan JD. 1964. Wages and prices in the United Kingdom: a study in econometric methodology. In *Econometric Analysis for National Economic Planning*, Hart PE, Mills G, Whitaker JK (eds). Butterworths: London; 25–54.
- Tsay RS. 1998. Testing and modeling multivariate threshold models. *Journal of the American Statistical Association* **93**: 1188–1202.
- Yao Y-C. 1988. Estimating the number of change-points via Schwarz' criterion. *Statistics and Probability Letters* **6**: 181–189.