Estimating and Testing Linear Models with Multiple Structural Changes

Jushan Bai; Pierre Perron

ESTIMATING AND TESTING LINEAR MODELS WITH MULTIPLE STRUCTURAL CHANGES

By Jushan Bai and Pierre Perron[1]

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

KEYWORDS: Asymptotic distribution, change point, rate of convergence, model selection.

1. INTRODUCTION

THIS PAPER CONSIDERS ISSUES related to multiple structural changes in the linear regression model estimated by minimizing the sum of squared residuals. Throughout, we treat the dates of the breaks as unknown variables to be estimated. The main aspects considered are the properties of the estimators, including the estimates of the break dates, and the construction of tests that allow inference to be made about the presence of structural change and the number of breaks.

Both the statistics and econometrics literature contains a vast amount of work on issues related to structural change, most of it specifically designed for the case of a single change.[2] The econometric literature has witnessed recently an

[2] See the surveys of Zacks (1983), Krishnaiah and Miao (1988), and Bhattacharya (1994).

upsurge of interest in extending procedures to various models with an unknown change point. With respect to the problem of testing for structural change, recent contributions include the comprehensive treatment of Andrews (1993) and Andrews and Ploberger (1994). This issue has also received a lot of attention in the debate on unit root versus structural change in the trend function of a univariate time series (see Perron (1989)). Issues about the distributional properties of the parameter estimates, in particular those of the break dates, have also been considered (see Bai (1994a, b)).

In comparison, the literature addressing the issue of multiple structural changes is relatively sparse. Recent developments include Andrews, Lee, and Ploberger (1996) who consider optimal tests in the linear model with known variance. Garcia and Perron (1996) study the sup Wald test for two changes in a dynamic time series. In an independent study, Liu, Wu, and Zidek (1997) consider, as we do, multiple shifts in a linear model estimated by least squares. They study the rate of convergence of the estimated break dates, as well as the consistency of a modified Schwarz model selection criterion to determine the number of breaks. Their analysis considers only the so-called pure-structural change case where all the parameters are subject to shifts. Our assumptions are much less restrictive than those of Liu, Wu, and Zidek (1997), and our main idea of argument differs from theirs. Our model allows for general forms of serial correlation and heteroskedasticity in the errors, lagged dependent variables, trending regressors, as well as different distributions for the errors and the regressors across segments. Furthermore, we consider the more general case of a partial structural change model where not all parameters are subject to shifts. A partial change model is useful in allowing potential savings in the number of degrees of freedom, an issue particularly relevant for multiple changes. We obtain the rates of convergence for the estimated break points not only for fixed but also for shrinking magnitudes of shifts. The latter is the basis for the derivation of feasible asymptotic distributions and confidence intervals for the break dates.

Our study considers, in addition, the important problem of testing for multiple structural changes for the case with no trending regressors. To that effect, we present sup Wald type tests for the null hypothesis of no change versus an alternative hypothesis containing an arbitrary number of changes. We also propose a test where the alternative specifies an unknown number of changes up to some maximum and a test of the null hypothesis of, say, l changes versus $l + 1$ changes. The latter is useful for a specific to general modeling strategy to determine the number of changes present. Finally our paper contains a discussion of an estimation strategy for which the locations of the breaks need not be simultaneously determined. Rather our method successively estimates each break point.

The rest of this paper is structured as follows. Section 2 discusses the model and the assumptions imposed on the variables and the errors. Section 3 contains results about the consistency, the rate of convergence, and the asymptotic distribution of the estimates of the break dates (as well as other parameters of

the model). Section 4 proposes test statistics, derives their asymptotic distributions, and presents critical values. Section 5 discusses sequential methods to estimate the model without treating all break points simultaneously. All proofs are collected in an appendix.

2. THE MODEL AND ASSUMPTIONS

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

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

for $j = 1, \ldots, m + 1$ and where we use the convention that $T_0 = 0$ and $T_{m+1} = T$. In this model, y_t is the observed independent variable, x_t ($p \times 1$) and z_t ($q \times 1$) are vectors of covariates, and β and δ_j ($j = 1, \ldots, m + 1$) are the corresponding vectors of coefficients; u_t is the disturbance. The indices $(T_1, \ldots, T_m)$, or the break points, are explicitly treated as unknown. The purpose is to estimate the unknown regression coefficients together with the break points when T observations on (y_t, x_t, z_t) are available. Note that this is a partial structural change model in the sense that β is not subject to shifts and is effectively estimated using the entire sample. When $p = 0$, we obtain a pure structural change model where all the coefficients are subject to change.

The multiple linear regression system (1) may be expressed in matrix form as $Y = X\beta + \overline{Z}\delta + U$, where $Y = (y_1, \ldots, y_T)'$, $X = (x_1, \ldots, x_T)'$, $U = (u_1, \ldots, u_T)'$, $\delta = (\delta_1', \delta_2', \ldots, \delta_{m+1}')'$, and $\overline{Z}$ is the matrix which diagonally partitions Z at the m-partition $(T_1, \ldots, T_m)$, i.e., $\overline{Z} = \text{diag}(Z_1, \ldots, Z_{m+1})$ with $Z_i = (z_{T_{i-1}+1}, \ldots, z_{T_i})'$. Throughout, we denote the true value of a parameter with a 0 superscript. In particular, $\delta^0 = (\delta_1^{0'}, \ldots, \delta_{m+1}^{0'})'$ and $(T_1^0, \ldots, T_m^0)$ are, respectively, the true values of the parameters δ and of the break points. The matrix $\overline{Z}^0$ is the one which diagonally partitions Z at $(T_1^0, \ldots, T_m^0)$. Hence the data-generating process is assumed to be

$$(2) \qquad Y = X\beta^0 + \overline{Z}^0 \delta^0 + U.$$

The goal is first to estimate the unknown coefficients ($\beta^0, \delta_1^0, \ldots, \delta_{m+1}^0, T_1^0, \ldots, T_m^0$), assuming $\delta_i^0 \neq \delta_{i+1}^0$ ($1 \leq k \leq m$). We do not impose the restriction that the regression function is continuous at the turning points. For the latter, readers are referred to Feder (1975) and Gallant and Fuller (1973) for the special case of a polynomial trend regression. This paper focuses on discrete shifts. In general, the number of breaks m can be treated as an unknown variable with true value m^0. However, for now, we treat it as known and discuss methods of estimating it in later sections. We also postpone the problem of testing for the presence of structural change to Section 4.

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

residuals $\sum_{i=1}^{m+1}\sum_{t=T_{i-1}+1}^{T_i}[y_t - x_t'\beta - z_t'\delta_i]^2$. Let $\hat{\beta}(\{T_j\})$ and $\hat{\delta}(\{T_j\})$ denote the resulting estimates. Substituting them in the objective function and denoting the resulting sum of squared residuals as $S_T(T_1,\ldots,T_m)$, the estimated break points $(\hat{T}_1,\ldots,\hat{T}_m)$ are such that

(3) $\left(\hat{T}_1,\ldots,\hat{T}_m\right) = \operatorname{argmin}_{T_1,\ldots,T_m} S_T(T_1,\ldots,T_m),$

where the minimization is taken over all partitions $(T_1,\ldots,T_m)$ such that $T_i - T_{i-1} \geq q$. Thus the break-point estimators are global minimizers of the objective function. Finally, the regression parameter estimates are the associated least-squares estimates at the estimated m-partition $\{\hat{T}_j\}$, i.e. $\hat{\beta} = \hat{\beta}(\{\hat{T}_j\})$ and $\hat{\delta} = \hat{\delta}(\{\hat{T}_j\})$. Note that the break points need not be obtained via an exhaustive grid search. We discuss in Bai and Perron (1996) an efficient algorithm based on the principle of dynamic programming which allows global minimizers to be obtained using a number of sums of squared residuals that is of order $O(T^2)$ for any $m \geq 2$.

The statistical properties of the resulting estimators are studied in the next section under the following set of assumptions. As a matter of notation, we let "$\to_p$" denote convergence in probability, "$\overset{d}{\to}$" convergence in distribution, and "$\Rightarrow$" weak converge in the space $D[0,1]$ under the Skorohod metric (e.g., Pollard (1984)).

ASSUMPTION A1: *Let* $w_t = (x_t', z_t')'$, $W = (w_1,\ldots,w_T)'$, *and* $\overline{W}^0$ *be the diagonal partition of* W *at* $(T_1^0,\ldots,T_m^0)$ *such that* $\overline{W}^0 = \operatorname{diag}(W_1^0,\ldots,W_{m+1}^0)$. *We assume for each* $i = 1,\ldots,m+1$, *with* $T_0^0 = 1$ *and* $T_{m+1}^0 = T$, *that* $W_i^{0'}W_i^0/(T_i^0 - T_{i-1}^0)$ *converges in probability to some nonrandom positive definite matrix not necessarily the same for all* i.

ASSUMPTION A2: *There exists an* $l_0 > 0$ *such that for all* $l > l_0$, *the minimum eigenvalues of* $A_{il} = (1/l)\sum_{T_i^0+1}^{T_i^0+l} w_t w_t'$ *and of* $A_{il}^* = (1/l)\sum_{T_i^0-l}^{T_i^0} w_t w_t'$ *are bounded away from zero* $(i = 1,\ldots,m+1)$.

ASSUMPTION A3: *The matrix* $B_{kl} = \sum_k^l z_t z_t'$ *is invertible for* $l - k \geq q$, *the dimension of* z_t.

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

ASSUMPTION A4(i): *With* $\{\mathscr{F}_i : i = 1,2,\ldots\}$ *a sequence of increasing* σ-fields, *assume that* $\{u_i, \mathscr{F}_i\}$ *forms a* L^r-mixingale *sequence with* $r = 4 + \delta$ *for some* $\delta > 0$ *(McLeish (1975) and Andrews (1988)). That is, there exist nonnegative constants* $\{c_i : i \geq 1\}$ *and* $\{\psi_j : j \geq 0\}$ *such that* $\psi_j \downarrow 0$ *as* $j \to \infty$ *and for all* $i \geq 1$ *and* $j \geq 0$, *we have:* (a) $E|E(u_i|\mathscr{F}_{i-j})|^r \leq c_i^r \psi_j^r$, (b) $E|u_i - E(u_i|\mathscr{F}_{i+j})|^r \leq c_i^r \psi_{j+1}^r$, (c) $\max_i c_i \leq K < \infty$, (d) $\sum_{j=0}^{\infty} j^{1+\kappa}\psi_j < \infty$ *for some* $\kappa > 0$. *We also assume* (e) *that the disturbances* u_t *are independent of the regressors* w_s *for all* t *and* s.

Or:

ASSUMPTION A4(ii): *Let $\mathscr{F}_t^* = \sigma$-field $\{\ldots, w_{t-1}, w_t, \ldots, u_{t-2}, u_{t-1}\}$. We assume* (a) *that $\{u_t\}$ is a martingale difference sequence relative to $\{\mathscr{F}_t^*\}$ and $\sup_t E|u_t|^{4+c}$* $< \infty$ *for some* $c > 0$; (b) $T^{-1}\sum_{t=1}^{[Tv]} z_t z_t' \to_p Q(v)$ *uniformly in* $v \in [0, 1]$, *where* $Q(v)$ *is positive definite for* $v > 0$ *and strictly increasing in* v; (c) *If the disturbances u_t are not independent of the regressors $\{z_s\}$ for all t and s, the minimization problem defined by* (3) *is taken over all possible partitions such that* $T_i - T_{i-1} > \epsilon T$ ($i = 1, \ldots, m+1$) *for some* $\epsilon > 0$.

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

Assumption A1 is standard for multiple linear regressions. Assumption A2 requires that there be enough observations near the true break points so that they can be identified. A2 can be weakened as follows. For some $c > 0$ and $l > l_0$, $\|(\delta_{i+1}^0 - \delta_i^0)'B\| > c\|(\delta_{i+1}^0 - \delta_i^0)\|$ for $B = (1/l)\sum_{T_i^0+1}^{T_i^0+l} z_t z_t'$ and $B = (1/l)\sum_{T_i^0-l}^{T_i^0} z_t z_t'$. In addition, for every $\epsilon > 0$ and $l = [\epsilon T]$, the minimum eigenvalues of $A_{i,l}$ and $A_{i,l}^*$ are bounded away from zero in probability for large T. A3 is imposed because the break points are estimated by a global least-squares search. If the number of observations in each segment is at least some fixed h ($h \geq q$, not depending on T), the invertibility requirement in A3 can be weakened to hold for all combinations (l, k) for which $l - k \geq h$.[3]

The assumptions stated in A4 pertain to two specific cases related to the presence or absence of a lagged dependent variable in w_t. The conditions described in part (i) pertain to the case where no lagged dependent variables are allowed in w_t implied by part (e). In this case, the conditions on the residuals are quite general and allow substantial correlation and heterogeneity. Part (ii) of Assumption A4 considers the case where lagged dependent variables are allowed as regressors. In this case, no serial correlation is permitted in the errors $\{u_t\}$. This extra generality is obtained at the expense of some restrictions on the admissible partitions if a lagged dependent variable is present in the z_t. In such cases, each segment considered to compute global minimizers must contain a positive fraction of the total sample. This is not constraining from a practical point of view since ϵ can be arbitrarily small. Note, however, that this restriction is not necessary if a lagged dependent variable is present only in the x_t's. In both A4(i) and A4(ii), the assumptions are general enough to allow different distributions for both the regressors and the errors in each segment.

The choice between assumptions A4(i) and A4(ii) can be especially interesting in the case of dynamic models when the coefficients associated with the lagged dependent variables are not subject to change. In this case, the investigator can take the dynamic effects into account either in a direct parametric fashion (e.g. introducing lagged dependent variables so as to have uncorrelated residuals) or

---

[3] Note that, for the proof of the consistency, A3 could be dispensed using generalized inverses.

using an indirect nonparametric approach (e.g. leaving the dynamics in the disturbances and applying a nonparametric correction for proper asymptotic inference).

Assumption A5 is a standard requirement to permit the development of an asymptotic theory and allows the break points to be asymptotically distinct. It considers the asymptotic experiments under the assumption that each segment increases proportionately as the sample size increases. We refer to the quantities $\lambda_0 = (\lambda_1^0, \ldots, \lambda_m^0)$ as the break fractions and we let $\lambda_0^0 = 0$ and $\lambda_{m+1}^0 = 1$. Finally, we assume that polynomial trending regressors are written in the form of $(t/T)^l$ ($l \geq 0$) or, more generally, written as a continuous function of the time trend, $g(t/T)$ (see Bai (1994b, 1995) for the case of a single break). The consistency and rate of convergence of the estimated break points apply to trending regressors. However, the assumptions in Section 4 for the test statistics rule out trending regressors.

3. CONSISTENCY AND LIMITING DISTRIBUTIONS

In this section, we analyze the consistency of the estimated break fractions and their rate of convergence. The latter allows us to derive results about the asymptotic distribution of the estimates $(\hat{\beta}, \hat{\delta}_1, \ldots, \hat{\delta}_{m+1}, \hat{T}_1, \ldots, \hat{T}_m)$. We let $\hat{\lambda} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_m) = (\hat{T}_1/T, \ldots, \hat{T}_m/T)$ with corresponding true values $\lambda^0 = (\lambda_1^0, \ldots, \lambda_m^0)$. We shall first show that $\hat{\lambda}$ is consistent for λ^0 and later that the rate of convergence is T.

3.1. *Consistency*

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

PROPOSITION 1: *Under A1–A5*: $\hat{\lambda}_k \to_p \lambda_k^0$, $k = 1, \ldots, m$.

We outline the main steps of the proof using a few lemmas that are proved in the appendix. Denote by $\hat{u}_t$ the estimated residuals and by d_t the difference between the fitted and true values. That is, $\hat{u}_t = y_t - x_t\hat{\beta} - z_t\hat{\delta}_k$, for $t \in [\hat{T}_{k-1} + 1, \hat{T}_k]$ and $d_t = x_t'(\hat{\beta} - \beta^0) + z_t'(\hat{\delta}_k - \delta_j^0)$, for $t \in [\hat{T}_{k-1} + 1, \hat{T}_k] \cap [T_{j-1}^0 + 1, T_j^0]$ ($k, j = 1, \ldots, m + 1$). Note that, in general, d_t is defined over $(m + 1)^2$ different segments for each of the possible m-partitions $\{\hat{T}_i\}$ and $\{T_i^0\}$. Using properties of projections,

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

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

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

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

LEMMA 1: *Under A1–A5, we have $T^{-1}\sum_{t=1}^{T} u_t d_t = o_p(1)$.*

Lemma 1 together with (4) and (5) implies that $T^{-1}\sum_{t=1}^{T} d_t^2 \to_p 0$. The proof follows by showing that this implies $\hat{\lambda} \to_p \lambda^0$. More specifically, $T^{-1}\sum_{t=1}^{T} d_t^2 \to_p 0$ cannot hold if $\hat{\lambda}_j \not\to_p \lambda_j^0$ for some j. This is stated in the following lemma.

LEMMA 2: *Assume A1–A5 hold and that $\hat{\lambda}_j \not\to_p \lambda_j^0$ for some j; then*

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

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

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

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

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

3.2. *Rates of Convergence*

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

PROPOSITION 2: *Under A1–A5, for every $\eta > 0$, there exists a $C < \infty$, such that for all large T, $P(|T(\hat{\lambda}_k - \lambda_k^0)| > C) < \eta$ $(k = 1,\ldots,m)$.*

It is important to remark that the rate T convergence pertains to the estimated break fraction $\hat{\lambda}_i$ and not to $\hat{T}_i$, the estimated break date. For the latter, our result states that with high probability its deviation from the true break is bounded by some constant C that is independent of T, i.e. with high probability, we have $|\hat{T}_i - T_i^0| < C$.

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

PROPOSITION 3: *Let* $\hat{\theta} = (\hat{\beta}, \hat{\delta})$ *and* $\theta^0 = (\beta^0, \delta^0)$. *Under* A1–A5, $\sqrt{T}(\hat{\theta} - \theta^0) \overset{d}{\rightarrow} N(0, V^{-1}\Phi V^{-1})$, *with* $V = \text{plim } T^{-1}\overline{W}^{0\prime}\overline{W}^0$, $\Phi = \text{plim } T^{-1}\overline{W}^{0\prime}\Omega\overline{W}^0$, *and* $\Omega = E(UU')$.

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

3.3. *Limiting Distributions of Break Dates*

Note first that, as in the single break case, the usual limiting distribution of the break dates obtained specifying fixed magnitude of changes depends on the exact distribution of the pair $\{z_t, u_t\}$. On the other hand, a strategy that permits obtaining pivotal statistics is to consider an asymptotic framework where the magnitudes of the shifts converge to zero as the sample size increases. Even though the setup is particularly well suited to provide an adequate approximation to the exact distribution when the shifts are small, it remains adequate even for moderate shifts. The required conditions are stated in the next assumptions defined for $i = 1, \ldots, m$.

ASSUMPTION A6: *Let* $\Delta_{T,i} = \delta^0_{T,i+1} - \delta^0_{T,i}$. *Assume* $\Delta_{T,i} = v_T \Delta_i$ *for some* Δ_i *independent of* T, *where* $v_T > 0$ *is a scalar satisfying* $v_T \rightarrow 0$ *and* $T^{(1/2)-\vartheta}v_T \rightarrow \infty$ *for some* $\vartheta \in (0, 1/2)$. *In addition, we assume* $E\|z_t\|^2 < M$ *and* $E|u_t|^{2/\vartheta} < M$ *for some* $M < \infty$ *and all* t.

Note that for a smaller magnitude of shift (small v_T), which corresponds to a smaller ϑ, A6 requires the existence of a higher moment of u_t. When v_T is a fixed constant, we can choose ϑ arbitrarily close to $1/2$. In this case, the requirement of $E|u_t|^{2/\vartheta} < M$ reduces to the existence of $4 + \delta$ moment, as stated in A4.

PROPOSITION 4: *Under Assumptions* A1–A6, *we have for* $k = 1, \ldots, m$: (i) $\hat{\lambda}_k \rightarrow_p \lambda^0_k$; *and* (ii) *for every* $\eta > 0$ *there exists a* $C < \infty$ *such that for all large* T, $P(|Tv_T^2(\hat{\lambda}_k - \lambda^0_k)| > C) < \eta$.

Proposition 4 asserts that the estimated break fractions remain consistent even in the case where the shifts decrease as the sample size increases. The rate of convergence is, of course, no longer T but rather Tv_T^2. This rate is sufficient to establish root-T consistency for the estimated regression parameters. This result will not be presented to save space, and interested readers are referred to

Bai (1994a, 1994b) for the case of a single break. Proposition 4 allows us to study the limiting distribution of the estimated break dates. It asserts that we can restrict the analysis to a "neighborhood" of length C/v_T^2 around the true break dates T_i^0 which makes possible the application of a central limit theorem since this "neighborhood" increases when v_T decreases. With the mixing assumptions on the errors, each segment is asymptotically distinct and the analysis of the limiting distribution of the break dates is similar to that in the single break case as analyzed in Bai (1994a, 1994b). We provide, in the rest of this section, a description of the results when the data are not trending and under the assumption that the following conditions are satisfied.

ASSUMPTION A7: *Let* $\Delta T_i^0 = T_i^0 - T_{i-1}^0$; *we assume, for* $i = 1,\ldots, m+1$, *that as* $\Delta T_i^0 \to \infty$:

(a) $\qquad (\Delta T_i^0)^{-1} \sum_{t=T_{i-1}^0+1}^{T_{i-1}^0+[s\,\Delta T_i^0]} z_t z_t' \to_p s Q_i, \quad (\Delta T_i^0)^{-1} \sum_{t=T_{i-1}^0+1}^{T_{i-1}^0+[s\,\Delta T_i^0]} u_t^2 \to_p s\sigma_i^2$

and

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

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

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

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

Now, define for $i = 1,\ldots, m$: $\xi_i = \Delta_i' Q_{i+1} \Delta_i / \Delta_i' Q_i \Delta_i$, $\phi_{i,1}^2 = \Delta_i' \Omega_i \Delta_i / \Delta_i' Q_i \Delta_i$, $\phi_{i,2}^2 = \Delta_i' \Omega_{i+1} \Delta_i / \Delta_i' Q_{i+1} \Delta_i$, and let $W_1^{(i)}(s)$ and $W_2^{(i)}(s)$ be independent Wiener processes defined on $[0,\infty)$, starting at 0 when $s = 0$. These processes are also independent across i. Also, define $Z^{(i)}(s) = \phi_{i,1} W_1^{(i)}(-s) - |s|/2$, for $s \leq 0$, and $Z^{(i)}(s) = \sqrt{\xi_i}\, \phi_{i,2} W_2^{(i)}(s) - \xi_i |s|/2$, for $s > 0$. We can state the following result.

PROPOSITION 5: *Under* $A1$–$A7$, $(\Delta_i' Q_i \Delta_i) v_T^2(\hat{T}_i - T_i^0) \Rightarrow \text{argmax}_s Z^{(i)}(s)$ ($i = 1,\ldots,m$).

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

(6) $\qquad \dfrac{(\Delta_i' Q \Delta_i)^2}{(\Delta_i' \Omega \Delta_i)} v_T^2 (\hat{T}_i - T_i^0) \Rightarrow \arg \max_s \{W^{(i)}(s) - |s|/2\}$

which is symmetric about the origin and has distribution function (see Yao (1987)):

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

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

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

The limiting distribution in the case of trending regressors is discussed in Bai (1994b, 1995) for a single structural change model. His results remain valid for multiple breaks. We omit the details and refer the reader to those papers.

4. TEST STATISTICS FOR MULTIPLE BREAKS

4.1. *A Test of No Break Versus Some Fixed Number of Breaks*

We consider the sup F type test of no structural break ($m = 0$) versus the alternative hypothesis that there are $m = k$ breaks. Let $(T_1, \ldots, T_k)$ be a partition such that $T_i = [T\lambda_i]$ $(i = 1, \ldots, k)$. Define

(7)
$$F_T(\lambda_1, \ldots, \lambda_k; q) = \left(\frac{T - (k+1)q - p}{kq}\right) \frac{\hat{\delta}'R'\left(R\left(\bar{Z}'M_X\bar{Z}\right)^{-1}R'\right)^{-1}R\hat{\delta}}{SSR_k}$$

where R is the conventional matrix such that $(R\delta)' = (\delta_1' - \delta_2', \ldots, \delta_k' - \delta_{k+1}')$ and $M_X = I - X(X'X)^{-1}X'$. Here SSR_k is the sum of squared residuals under the alternative hypothesis, which depends on $(T_1, \ldots, T_k)$. To carry out the asymptotic analysis, we need to impose some restrictions on the possible values of the break dates. In particular, we need to restrict each break date to be asymptotically distinct and bounded from the boundaries of the sample. To this effect, we define the following set for some arbitrary small positive number ϵ: $\Lambda_\epsilon = \{(\lambda_1, \ldots, \lambda_k); |\lambda_{i+1} - \lambda_i| \geq \epsilon, \lambda_1 \geq \epsilon, \lambda_k \leq 1 - \epsilon\}$. The sup F type statistic is then defined as $\sup F_T(k; q) = \sup_{(\lambda_1, \ldots, \lambda_k) \in \Lambda_\epsilon} F_T(\lambda_1, \ldots, \lambda_k; q)$. It is a generalization of the sup F test considered by Andrews (1993) and others for the case $k = 1$. The limiting distribution of the test depends on the nature of the regressors and the presence or absence of serial correlation and heterogeneity in the residuals. We consider the case where the following assumptions are imposed.

ASSUMPTION A8: $T^{-1}\sum_{t=1}^{[Ts]}w_t w_t' \to_p sQ$, *uniformly in* $s \in [0,1]$, *for Q some positive definite matrix.*

Note that A8 precludes the presence of trending regressors. Extensions to the general case where $\text{plim}_{T\to\infty}T^{-1}\sum_{t=1}^{[Ts]}w_t w_t' = Q(s)$, which allows trending regressors, are beyond the scope of the present paper.

ASSUMPTION A9: *The errors* $\{u_t\}$ *form an array of martingale differences relative to* $\{\mathscr{F}_t\} = \sigma$-*field* $\{\ldots, w_{t-1}, w_t, \ldots, u_{t-2}, u_{t-1}\}$. *Also,* $E[u_t^2] = \sigma^2$ *for all* t *and* $T^{-1/2}\sum_{t=1}^{[Tr]}w_t u_t \Rightarrow \sigma Q^{1/2}W^*(r)$, *with* $W^*(r)$ *a* $(p+q)$ *vector of independent Wiener processes.*

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

PROPOSITION 6: *Let* $W_q(\cdot)$ *be a* q-*vector of independent Wiener processes on* $[0,1]$. *Under A8–A9 and* $m = 0$, $\sup F_T(k; q) \Rightarrow \sup F_{k,q} \stackrel{def}{=} \sup_{(\lambda_1,\ldots,\lambda_k)\in \Lambda_\epsilon}F(\lambda_1, \ldots, \lambda_k; q)$, *with*

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

Note that the asymptotic distribution of the test statistic depends on the value of ϵ in Δ_ϵ. As ϵ converges to zero, the critical values of the limiting random variable of $\sup F_T(k; q)$ diverge to infinity. Because the computed test statistic for a given sample is finite, a small positive value of ϵ can improve the power significantly; see Andrews (1993) for further details. In what follows, we have adopted $\epsilon = 0.05$. No critical values for $k \geq 2$ are available except those of Garcia and Perron (1996) for $k = 2$ and $q = 1$.

Asymptotic critical values are obtained via simulations. The Wiener process $W_q(\lambda)$ is approximated by the partial sums $n^{-1/2}\sum_{i=1}^{[n\lambda]}e_i$ with e_i i.i.d. $N(0, I_q)$ and $n = 1,000$. The number of replications is 10,000. For each replication, the supremum of $F(\lambda_1, \ldots, \lambda_k; q)$ with respect to $(\lambda_1, \ldots, \lambda_k)$ over the set Λ_ϵ is obtained via a dynamic programming algorithm. We present, in Table I, critical values covering cases with up to 9 breaks ($k = 1, \ldots, 9$) and up to 10 regressors ($q = 1, \ldots, 10$) whose coefficients are the object of the test. The values reported are scaled up by q for comparison purposes. The column corresponding to $k = 1$ can also be found in Andrews (1993). Because $\sup F_T(1; q) \leq 2\sup F_T(2; q) \leq k\sup F_T(k; q)$, the consistency of the $\sup F_T(k; q)$ ($k \geq 2$) follows from Andrews (1993) who proved the consistency of $\sup F_T(1; q)$ for various alternatives including multiple breaks.

TABLE I

ASYMPTOTIC CRITICAL VALUES OF THE MULTIPLE-BREAK TEST.
THE ENTRIES ARE QUANTILES x SUCH THAT $P(\sup F_{k,q} \leq x/q) = \alpha$.

q	α	Number of Breaks, k									UDmax	WDmax
		1	2	3	4	5	6	7	8	9		
1	.90	8.02	7.87	7.07	6.61	6.14	5.74	5.40	5.09	4.81	8.78	9.14
	.95	9.63	8.78	7.85	7.21	6.69	6.23	5.86	5.51	5.20	10.17	10.91
	.975	11.17	9.81	8.52	7.79	7.22	6.70	6.27	5.92	5.56	11.52	12.53
	.99	13.58	10.95	9.37	8.50	7.85	7.21	6.75	6.33	5.98	13.74	15.02
2	.90	11.02	10.48	9.61	8.99	8.50	8.06	7.66	7.32	7.01	11.69	12.33
	.95	12.89	11.60	10.46	9.71	9.12	8.65	8.19	7.79	7.46	13.27	14.19
	.975	14.53	12.64	11.20	10.29	9.69	9.10	8.64	8.18	7.80	14.69	16.04
	.99	16.64	13.78	12.06	11.00	10.28	9.65	9.11	8.66	8.22	16.79	18.11
3	.90	13.43	12.73	11.76	11.04	10.49	10.02	9.59	9.21	8.86	14.05	14.76
	.95	15.37	13.84	12.64	11.83	11.15	10.61	10.14	9.71	9.32	15.80	16.82
	.975	17.17	14.91	13.44	12.49	11.75	11.13	10.62	10.14	9.72	17.36	18.79
	.99	19.25	16.27	14.48	13.40	12.56	11.80	11.22	10.67	10.19	19.38	20.81
4	.90	15.53	14.65	13.63	12.91	12.33	11.79	11.34	10.93	10.55	16.17	16.95
	.95	17.60	15.84	14.63	13.71	12.99	12.42	11.91	11.49	11.04	17.88	19.07
	.975	19.35	16.85	15.44	14.43	13.64	13.01	12.46	11.94	11.49	19.51	20.89
	.99	21.20	18.21	16.43	15.21	14.45	13.70	13.04	12.48	12.02	21.25	22.81
5	.90	17.42	16.45	15.44	14.69	14.05	13.51	13.02	12.59	12.18	17.94	18.85
	.95	19.50	17.60	16.40	15.52	14.79	14.19	13.63	13.16	12.70	19.74	20.95
	.975	21.47	18.75	17.26	16.13	15.40	14.75	14.19	13.66	13.17	21.57	23.04
	.99	23.99	20.18	18.19	17.09	16.14	15.34	14.81	14.26	13.72	24.00	25.46
6	.90	19.38	18.15	17.17	16.39	15.74	15.18	14.63	14.18	13.74	19.92	20.89
	.95	21.59	19.61	18.23	17.27	16.50	15.86	15.29	14.77	14.30	21.90	23.27
	.975	23.73	20.80	19.15	18.07	17.21	16.49	15.84	15.29	14.78	23.83	25.22
	.99	25.95	22.18	20.29	18.93	17.97	17.20	16.54	15.94	15.35	26.07	27.63
7	.90	21.23	19.93	18.75	17.98	17.28	16.69	16.16	15.69	15.24	21.79	22.81
	.95	23.50	21.30	19.83	18.91	18.10	17.43	16.83	16.28	15.79	23.77	25.02
	.975	25.23	22.54	20.85	19.68	18.79	18.03	17.38	16.79	16.31	25.46	26.92
	.99	28.01	24.07	21.89	20.68	19.68	18.81	18.10	17.49	16.96	28.02	29.57
8	.90	22.92	21.56	20.43	19.58	18.84	18.21	17.69	17.19	16.70	23.53	24.55
	.95	25.22	23.03	21.48	20.46	19.66	18.97	18.37	17.80	17.30	25.51	26.83
	.975	27.21	24.20	22.41	21.29	20.39	19.63	18.98	18.34	17.78	27.32	28.98
	.99	29.60	25.66	23.44	22.22	21.22	20.40	19.66	19.03	18.46	29.60	31.32
9	.90	24.75	23.15	21.98	21.12	20.37	19.72	19.13	18.58	18.09	25.19	26.40
	.95	27.08	24.55	23.16	22.08	21.22	20.49	19.90	19.29	18.79	27.28	28.78
	.975	29.13	25.92	24.14	22.97	21.98	21.28	20.59	19.98	19.39	29.20	30.82
	.99	31.66	27.42	25.13	24.01	23.06	22.18	21.35	20.63	19.94	31.72	33.32
10	.90	26.13	24.70	23.48	22.57	21.83	21.16	20.57	20.03	19.55	26.66	27.79
	.95	28.49	26.17	24.59	23.59	22.71	21.93	21.34	20.74	20.17	28.75	30.16
	.975	30.67	27.52	25.69	24.47	23.45	22.71	21.95	21.34	20.79	30.84	32.46
	.99	33.62	29.14	26.90	25.58	24.44	23.49	22.75	22.09	21.47	33.86	35.47

Notes: 1. The test UDmax is defined as $\max_{1 \leq k \leq 5}\sup_{(\lambda_1,\ldots,\lambda_k)\in\Lambda_\epsilon} F(\lambda_1,\ldots,\lambda_k;q)$ multiplied by q. 2. The test WDmax is given in (9) multiplied by q, and M is chosen to be 5.

4.2. *A Double Maximum Test*

The test discussed above requires the specification of the number of breaks, m, under the alternative hypothesis. It is of interest to consider tests of no structural break against an unknown number of breaks given some upper bound

M. Consider the following new class of tests, called the *double maximum tests*:

$$(8) \qquad \text{Dmax} F_T(M, q, a_1, \ldots, a_M) = \max_{1 \leq m \leq M} a_m \sup_{(\lambda_1, \ldots, \lambda_m) \in \Lambda_\epsilon} F_T(\lambda_1, \ldots, \lambda_m; q),$$

defined for some fixed weights $\{a_1, \ldots, a_M\}$. Note that the asymptotic distribution of this class of tests is easily obtained from Proposition 6. Indeed, we have

$$\text{Dmax} F_T(M, q, a_1, \ldots, a_M) \Rightarrow \max_{1 \leq m \leq M} a_m \sup_{(\lambda_1, \ldots, \lambda_m) \in \Lambda_\epsilon} F(\lambda_1, \ldots, \lambda_m; q).$$

The weights may reflect the imposition of some priors on the likelihood of various numbers of breaks. Apart from such considerations, precise theoretical guidelines about their choice remain an open question. An obvious candidate is to set all weights equal to unity and we label this version of the test as UDmax $F_T(M, q) = \max_{1 \leq m \leq M} \sup_{(\lambda_1, \ldots, \lambda_m) \in \Lambda_\epsilon} F_T(\lambda_1, \ldots, \lambda_m; q)$. For a fixed m, $F(\lambda_1, \ldots, \lambda_m; q)$ is the sum of m dependent chi-square random variables with q degrees of freedom, each one divided by m. This scaling by m can be viewed, in some sense, as a prior imposed to account for the fact that as m increases a fixed sample of data becomes less informative about the hypotheses being confronted. Since for any fixed q the critical values of the individual tests $\sup_{(\lambda_1, \ldots, \lambda_m) \in \Lambda_\epsilon} F_T(\lambda_1, \ldots, \lambda_m; q)$ decrease as m increases, this implies that the marginal p-values decrease with m and may lead to a test with low power if the number of breaks is large. One way to alleviate this problem is to consider a set of weights such that the marginal p-values are equal across values of m. This implies weights that depend on q and the significance level of the test, say α. To be more precise, let $c(q, \alpha, m)$ be the asymptotic critical value of the test $\sup_{(\lambda_1, \ldots, \lambda_m) \in \Lambda_\epsilon} F_T(\lambda_1, \ldots, \lambda_m; q)$ for a significance level α. The weights are then defined as $a_1 = 1$ and for $m > 1$ as $a_m = c(q, \alpha, 1)/c(q, \alpha, m)$. This version is denoted

$$(9) \qquad \text{WDmax } F_T(M, q) = \max_{1 \leq m \leq M} \frac{c(q, \alpha, 1)}{c(q, \alpha, m)}$$
$$\times \sup_{(\lambda_1, \ldots, \lambda_m) \in \Lambda_\epsilon} F_T(\lambda_1, \ldots, \lambda_m; q).$$

The last two columns of Table I report the asymptotic critical values of both tests for $M = 5$ and $\epsilon = 0.05$. This should be sufficient for most empirical applications. In any event, the critical values vary little for choices of the upper bound M larger than 5. The consistency of the tests follows directly from the consistency of $\sup F_T(k; q)$.

4.3. *Test of l versus l + 1 Breaks*

This section considers a test of the null hypothesis of l breaks against the alternative that an additional break exists. Ideally, one would base the test on the difference between the sum of squared residuals obtained with l breaks and that obtained with $l + 1$ breaks. The limiting distribution of this test statistic is,

however, difficult to obtain. Here, we pursue a different strategy. For the model with l breaks, the estimated break points, denoted by $\hat{T}_1, \ldots, \hat{T}_l$, are obtained by a global minimization of the sum of squared residuals. Our strategy proceeds by testing each $(l + 1)$ segment (obtained using the estimated partition $\hat{T}_1, \ldots, \hat{T}_l$) for the presence of an additional break. We assume the magnitude of shifts is fixed (nonshrinking) in this section.

The test amounts to the application of $(l + 1)$ tests of the null hypothesis of no structural change versus the alternative hypothesis of a single change. It is applied to each segment containing the observations $\hat{T}_{i-1} + 1$ to $\hat{T}_i$ $(i = 1, \ldots, l + 1)$ using again the convention that $\hat{T}_0 = 0$ and $\hat{T}_{l+1} = T$. We conclude for a rejection in favor of a model with $(l + 1)$ breaks if the overall minimal value of the sum of squared residuals (over all segments where an additional break is included) is sufficiently smaller than the sum of squared residuals from the l break model. The break date thus selected is the one associated with this overall minimum. More precisely, the test is defined by

$$(10) \qquad F_T(l + 1 | l) = \left\{ S_T\left(\hat{T}_1, \ldots, \hat{T}_l\right) \right.$$

$$\left. - \min_{1 \leq i \leq l+1} \inf_{\tau \in \Lambda_{i,\eta}} S_T\left(\hat{T}_1, \ldots, \hat{T}_{i-1}, \tau, \hat{T}_i, \ldots, \hat{T}_l\right) \right\} / \hat{\sigma}^2,$$

where

$$(11) \qquad \Lambda_{i,\eta} = \left\{ \tau; \hat{T}_{i-1} + \left(\hat{T}_i - \hat{T}_{i-1}\right)\eta \leq \tau \leq \hat{T}_i - \left(\hat{T}_i - \hat{T}_{i-1}\right)\eta \right\}$$

and $\hat{\sigma}^2$ is a consistent estimate of σ^2 under the null hypothesis. Note that for $i = 1$, $S_T(\hat{T}_1, \ldots, \hat{T}_{i-1}, \tau, \hat{T}_i, \ldots, \hat{T}_l)$ is understood as $S_T(\tau, \hat{T}_1, \ldots, \hat{T}_l)$ and for $i = l + 1$ as $S_T(\hat{T}_1, \ldots, \hat{T}_l, \tau)$. We have the following result, proved in the Appendix:

PROPOSITION 7: *Under Assumptions A8–A9 and $m = l$: $\lim_{T \to \infty} P(F_T(l + 1 | l) \leq x) = G_{q,\eta}(x)^{l+1}$ with $G_{q,\eta}(x)$ the distribution function of $\sup_{\eta \leq \mu \leq 1-\eta} \|W_q(\mu) - \mu W_q(1)\|^2 / (\mu(1 - \mu))$.*

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

Note that $\hat{\sigma}^2$ is only required to be consistent under the null hypothesis for the validity of the stated asymptotic distribution. The test may, however, have better power if $\hat{\sigma}^2$ is also consistent under the alternative hypothesis. Also, it is important to note that the results carry through allowing different distributions across segments for the regressors and the errors. That is, Proposition 7 remains

TABLE II

ASYMPTOTIC CRITICAL VALUES OF THE SEQUENTIAL TEST $F_T(l+1|l)$.
THE ENTRIES ARE THE QUANTILES x SUCH THAT $G_{q,\eta}(x)^{l+1} = \alpha$.

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

valid under A7 instead of A8–A9, provided $\hat{\sigma}^2$ is replaced by $\hat{\sigma}_i^2$ in (44) of the Appendix.

We next argue that the test based on $F_T(l+1|l)$ is also consistent. If there are more than l breaks and a model with only l breaks is estimated, there must be at least one break that is not estimated. Hence, at least one segment contains

a nontrivial break point in the sense that both boundaries of each segment are separated from the true break point by a positive fraction of the total number of observations. For this segment, the $\sup F_T(1; q)$ test statistic diverges to infinity as the sample size increases since it is consistent. Accordingly, the statistic $F_T(l + 1|l)$ (computed for $l + 1$ segments) also diverges to infinity. This shows consistency.

4.4. *Extensions to Serially Correlated Errors*

The tests discussed above can be applied without the imposition of serially uncorrelated errors as specified in Assumption A9. A simple modification is to use the following version of the F test instead of that specified in (7):

$$(12) \qquad F_T^*(\lambda_1, \ldots, \lambda_k; q) = \left(\frac{T - (k + 1)q - p}{kq} \right) \hat{\delta}' R' (R\hat{V}(\hat{\delta})R')^{-1} R\hat{\delta},$$

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

$$(13) \qquad V(\hat{\delta}) = \text{plim } T \left(\overline{Z}'M_X\overline{Z} \right)^{-1} \overline{Z}'M_X \Omega M_X \overline{Z} \left(\overline{Z}'M_X\overline{Z} \right)^{-1}.$$

Note that it can be constructed allowing identical or different distributions for the regressors and the errors across segments. In some instances, the form of the statistic reduces in an interesting way. For example, consider a pure structural change model ($\beta = 0$) where the explanatory variables are such that $\text{plim } T^{-1}\overline{Z}'\Omega\overline{Z} = h_u(0)\text{plim } T^{-1}\overline{Z}'\overline{Z}$ with $h_u(0)$ the spectral density function of the errors u_t evaluated at the zero frequency. In that case, we have the asymptotically equivalent test

$$F_T^*(\lambda_1, \ldots, \lambda_k; q) = \left(\hat{\sigma}^2/\hat{h}_u(0) \right) F_T(\lambda_1, \ldots, \lambda_k; q),$$

with $\hat{\sigma}^2 = T^{-1}\sum_{t=1}^T \hat{u}_t^2$ and $\hat{h}_u(0)$ a consistent estimate of $h_u(0)$. Hence, the robust version of the test is simply a scaled version of the original statistic. This is the case, for instance, when testing for a change in mean as in Garcia and Perron (1996).

The computation of the robust version of the F test (12) can be involved especially if a data dependent method is used to construct the robust asymptotic covariance matrix of $\hat{\delta}$. Since the break fractions are T-consistent even with correlated errors, an asymptotically equivalent version is to first take the supremum of the original F test to obtain the break points, i.e. imposing $\Omega = \sigma^2 I$. The robust version of the test is obtained by evaluating (12) and (13) at these estimated break dates.

5. SEQUENTIAL METHODS

In this section we discuss issues related to the sequential estimation of the break points. We start, in Section 5.1 with results about the limit of break point

estimates in underspecified models. An interesting by-product is a sequential algorithm to estimate models with an unknown number of breaks discussed in Section 5.2.

5.1. *The Limit of Break Point Estimates in Underspecified Models*

In this section, we show that the estimate of the break fraction in a single structural change regression applied to data that contain two breaks converges to one of the two true break fractions. In independent work, Chong (1994) obtains a similar result (see also Bai (1994c) for an earlier exposition). To present our arguments, we consider a simple three-regime model:

$$(14) \qquad y_t = \mu_j + \epsilon_t, \quad \text{if} \quad [T\lambda_{j-1}] + 1 \le t \le [T\lambda_j],$$

for $j = 1, 2, 3$ and with $\epsilon_t \sim$ i.i.d.$(0, \sigma_\epsilon^2)$. Assume $\mu_1 \ne \mu_2$, $\mu_2 \ne \mu_3$, and $\lambda_1 < \lambda_2$, so there are two break points in the model. Let $\hat{T}_a$ denote the estimated single shift point. Our aim is to show that $\hat{T}_a/T$ is consistent for either λ_1 or λ_2 depending on the relative magnitudes of the shifts and the spell of each regime. To verify this claim, we examine the global behavior of $S(\tau)$, the limit of $T^{-1}S_T([T\tau])$. We let $S_T(0)$ and $S_T(T)$ be the sum of squared residuals for the full sample without a break and $S_T([T\tau])$ is then well defined for all $\tau \in [0, 1]$. It is not difficult to show that the convergence of $T^{-1}S_T([T\tau])$ to $S(\tau)$ is uniform in $\tau \in [0, 1]$. In particular

$$(15) \qquad \frac{1}{T}S_T([T\lambda_1]) \to_p S(\lambda_1) = \sigma_\epsilon^2 + \frac{(1 - \lambda_2)(\lambda_2 - \lambda_1)}{1 - \lambda_1}(\mu_2 - \mu_3)^2,$$

$$(16) \qquad \frac{1}{T}S_T([T\lambda_2]) \to_p S(\lambda_2) = \sigma_\epsilon^2 + \frac{\lambda_1}{\lambda_2}(\lambda_2 - \lambda_1)(\mu_1 - \mu_2)^2.$$

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

LEMMA 3: *Suppose that the data are generated by* (14) *and that* $S(\lambda_1) < S(\lambda_2)$; *the estimated single break point* $\hat{T}_a/T$ *is consistent for* λ_1.

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

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

5.2. *Sequential Estimation of the Break Points*

The arguments in Section 5.1 showed that $\hat{T}_a/T$ is consistent for one of the true break points, the one that allows the greatest reduction in the sum of squared residuals. Suppose, as above, that this break point is λ_1, which, in general, may not be known. In that case, we choose one break point either in the intervals $[1, \hat{T}_a]$ or $[\hat{T}_a, T]$, such that the sum of squared residuals for all observations $[1, T]$ is minimized. Let $\hat{\tau}$ be this estimator. With probability tending to 1 as T increases, it is easy to show that the estimated break point $\hat{\tau}$ will be in the interval $[\hat{T}_a, T]$. Similarly, if $\hat{T}_a$ is actually consistent for λ_2 (this will be true if $S(\lambda_1) > S(\lambda_2)$), the second estimated break point will be in $[1, \hat{T}_a]$. Generally, let $(\hat{N}_1, \hat{N}_2)$ be the ordered version of $(\hat{T}_a, \hat{\tau})$ such that $\hat{N}_1 < \hat{N}_2$. Then $(\hat{N}_1/T, \hat{N}_2/T)$ is consistent for (λ_1, λ_2). The preceding argument implies that we can obtain consistent estimates of λ_1 and λ_2 in a sequential way.

5.2.1. *Sequential Estimation with a Known Number of Break Points*

The above analysis suggests a straightforward sequential algorithm for estimating models with multiple break points. Consider first the case of a known number of break points, say m. Once the first break point is identified, the sample is split into two subsamples separated by this first estimated break point. For each subsample, a one break model is estimated and the second break point is chosen as that break point (of the two obtained) which allows the greatest reduction in the sum of squared residuals. The sample is then partitioned in three regimes and a third break point is selected as the estimate from three estimated one-break models that allows the greatest reduction in the sum of squared residuals. This process is continued until the m break points are selected. It yields consistent estimates of the break points though the estimates are not guaranteed to be identical to those obtained by global minimization. Interestingly, it allows the estimation of models with any fixed number of structural changes using least-squares operations that are only of order $O(T)$.

5.2.2. *Sequential Estimation with an Unknown Number of Breaks*

Consider now the case of an unknown number of breaks which is likely to be of particular relevance in practice. A standard problem is that an improvement

in the objective function is always possible by allowing more breaks. This naturally leads to the consideration of a penalty factor for the increased dimension of a model. Yao (1988) suggests the use of the Bayesian Information Criterion and Liu, Wu, and Zidek (1997) suggest a modified Schwarz' criterion; see also Yao and Au (1989). We propose an alternative method directly related to the sequential procedure outlined above. Start by estimating a model with a small number of breaks that are thought to be necessary (or start with no break). Then perform parameter-constancy tests for every subsample (those obtained by cutting off at the estimated breaks), adding a break to a subsample associated with a rejection using the test $F_T(l + 1|l)$. This process is repeated by increasing l sequentially until the test fails to reject the null hypothesis of no additional structural changes. A distinct advantage of such a model selection device over those based on information criteria is that it can easily allow and take into account the effect of possible serial correlation in the errors.

Note that the application of the test $F_T(l + 1|l)$ in this sequential context is different from that discussed earlier. Indeed, the result of Proposition 7 is based on having the first l breaks obtained as global minimizers of the sum of squared residuals assuming l breaks. The limiting distribution of the $F_T(l + 1|l)$ test in the sequential setup is the same because rate T convergence still holds, as shown in Bai (1997), when the break points are obtained sequentially.

With probability approaching 1 as the sample size increases, the number of breaks determined this way will be no less than the true number. The procedure does not provide a consistent estimate of the true number of breaks, say m_0, since it implies a nonzero probability of rejection under the null hypothesis given by the level of the test, say α. However, the asymptotic probability of selecting a model with a larger number of breaks, say $m_0 + j$, is given by α^j which decreases rapidly. Hence, there is no need (with large probability) to estimate models with more than the true number of breaks. The sequential procedure could be made consistent by adopting a significance level for the test $F_T(l + 1|l)$ that decreases to zero, at a suitable rate, as the sample size increases. A result to that effect is presented in the next proposition whose proof is similar to that of Hosoya (1989) and is, therefore, omitted.

PROPOSITION 8: *Let $\hat{m}$ be the number of breaks obtained using the sequential method based on the statistic $F_T(l + 1|l)$ applied with some size α_T, and let m_0 be the true number of breaks. If α_T converges to 0 slowly enough (for the test based on $F_T(l + 1|l)$ to remain consistent), then, under Assumptions A1–A5, $P(\hat{m} = m_0)$ → 1, as $T \to \infty$.*

6. CONCLUSIONS

Our analysis has presented a comprehensive treatment of issues related to the estimation of linear models with multiple structural changes, to tests for the presence of multiple structural changes and to the determination of the number of changes present. Our results being asymptotic in nature, there is certainly a

need to evaluate the quality of the approximations and the power of the tests in finite samples via simulations. We present such a simulation study in a companion paper, Bai and Perron (1996). Among the topics to be investigated, an important one appears to be the relative merits of different methods to select the number of structural changes. There are, of course, many other issues on the agenda: for instance, extensions of the test procedures to include tests that are optimal with respect to some criteria and extensions to nonlinear models. In addition, while the consistency and rate of convergence for the estimated break points apply to trending regressors, the limiting distributions of the various tests for structural change remain to be studied in the presence of trending regressors.

Dept. of Economics, E52-247B, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A; jbai@mit.edu

and

Dept. of Economics, Boston University, 270 Bay State Rd., Boston, MA 02215, U.S.A.; perron@bu.edu

MATHEMATICAL APPENDIX

As a matter of notation, for a sequence of matrices B_T, we write $B_T = o_p(1)$ if each of its elements is $o_p(1)$ and likewise for $O_p(1)$. For a matrix A, $M_A = I - P_A$ with $P_A = A(A'A)^{-1}A'$. We use $\|\cdot\|$ to denote the Euclidean norm, i.e. $\|x\| = (\sum_1^p x_i^2)^{1/2}$ for $x \in R^p$. For a matrix A, we use the vector-induced norm, i.e. $\|A\| = \sup_{x \neq 0} \|Ax\|/\|x\|$. Note that the norm of A is equal to the square root of the maximum eigenvalue of $A'A$, and thus $\|A\| \leq [\text{tr}(A'A)]^{1/2}$. Also, for a projection matrix P, $\|PA\| \leq \|A\|$. Limits are taken as T, the sample size, increases to infinity. We start with a series of lemmas that will be used subsequently. Assumption A5 is assumed throughout.

LEMMA A.1: *Let S and V be two matrices having the same number of rows. Then the matrix $S'M_V S$ is nondecreasing as more rows are added to the matrix (S, V).*

PROOF: Write $S = (S_1', S_2')'$ and $V = (V_1', V_2')'$. We need to show that for an arbitrary vector α (having the same dimension as the number of rows of S and V) $\alpha' S' M_V S \alpha \geq \alpha' S_1' M_{V_1} S_1 \alpha$. Note that $\alpha' S' M_V S \alpha$ ($\alpha' S_1' M_{V_1} S_1 \alpha$) is the sum of squares of the residuals from a projection of $S\alpha$ ($S_1\alpha$) on the space spanned by V (V_1). The inequality is verified using the fact that the sum of squared residuals is nondecreasing as the number of observations increases (here the number of rows of S_1 and S). See, e.g., Brown, Durbin, and Evans (1975).   Q.E.D.

LEMMA A.2: *Under $A1$, $\sup_{T_1, \ldots, T_m} (X'M_{\bar{Z}}X/T)^{-1} = O_p(1)$, where the supremum is taken over all possible partitions such that $|T_{i-1} - T_i| \geq q$ ($i = 1, \ldots, m+1$).*

PROOF: We have the identity $X'M_{\bar{Z}}X = X_1'M_{Z_1}X_1 + \cdots + X_{m+1}'M_{Z_{m+1}}X_{m+1}$. Each partition $(T_1, \ldots, T_m)$ leaves at least one true regime intact. In other words, there exists an i such that (X_i, Z_i) contains (X_i^0, Z_i^0) as a submatrix. We have $X_i'M_{Z_i}X_i \geq X_i^{0'}M_{Z_i^0}X_i^0$ using Lemma A.1. Hence $(X'M_{\bar{Z}}X/T)^{-1} \leq (X_i^{0'}M_{Z_i^0}X_i^0/T)^{-1}$. This implies $\|(X'M_{\bar{Z}}X/T)^{-1}\| \leq \max_i \|(X_i^{0'}M_{Z_i^0}X_i^0/T)^{-1}\|$ for all partitions. The lemma now follows from Assumption A.1.   Q.E.D.

LEMMA A.3: *Under* $A1$, $\sup_{T_1,\ldots,T_m} X' M_{\bar{Z}} \bar{Z}^0 = O_p(T)$.

PROOF: Because $M_{\bar{Z}}$ is a projection matrix, we have $\|X' M_{\bar{Z}} \bar{Z}^0\| \le \|X\| \|M_{\bar{Z}} \bar{Z}^0\| \le \|X\| \|\bar{Z}^0\|$ uniformly over all partitions. The lemma follows from $\|X\| \le [\mathrm{tr}(X'X)]^{1/2} = O_p(T^{1/2})$ and similarly $\|\bar{Z}^0\| = O_p(T^{1/2})$. $\hspace{2cm}$ Q.E.D.

LEMMA A.4: *Under* $A4$, *there exists* $\alpha < 1/2$ *such that* $\sup_{T_1,\ldots,T_m} \|P_{\bar{Z}} U\| = O_p(T^\alpha)$, *where the supremum with respect to* $(T_1,\ldots,T_m)$ *is taken over all possible partitions such that* $|T_{i-1} - T_i| \ge q$ $(i = 1,\ldots,m+1)$ *under Assumption* $A4(i)$ *and over partitions such that* $|T_{i-1} - T_i| \ge \epsilon T$ *for some* $\epsilon > 0$ *under Assumption* $A4(ii)$.

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

$$\left(\sum_{T_i+1}^{T_{i+1}} z_t u_t \right)' \left(\sum_{T_i+1}^{T_{i+1}} z_t z_t' \right)^{-1} \left(\sum_{T_i+1}^{T_{i+1}} z_t u_t \right),$$

for $i = 0,\ldots,m$. Thus it suffices to prove that

$$(17) \qquad \sup_{1 \le k < l \le T} \left\| \sum_{t=k}^{l} \xi_t \right\| = O_p(T^\alpha)$$

with $l - k \ge q$ and $\xi_t = \xi_t(k,l) = (A_{kl})^{-1/2} z_t u_t$ with $A_{kl} = \sum_{j=k}^{l} z_j z_j'$. Now

$$(18) \qquad P\left(\sup_{1 \le k < l \le T} \left\| \sum_{t=k}^{l} \xi_t \right\| > T^\alpha \right) \le \sum_{k=1}^{T} \sum_{l=k+q}^{T} P\left(\left\| \sum_{t=k}^{l} \xi_t \right\| > T^\alpha \right)$$

$$\le T^{-2\alpha s} \sum_{k=1}^{T} \sum_{l=k+q}^{T} E\left\| \sum_{t=k}^{l} \xi_t \right\|^{2s}.$$

By the mixingale property, we can write $u_t = \sum_{j=-\infty}^{\infty} u_{jt}$, with $u_{jt} = E(u_t|\mathscr{F}_{t-j}) - E(u_t|\mathscr{F}_{t-j-1})$ and for each j, $\{u_{jt},\mathscr{F}_{t-j}\}$ is a sequence of martingale differences. Hence, we have $\sum_{t=k}^{l} \xi_t = \sum_{j=-\infty}^{\infty} \sum_{t=k}^{l} \xi_{jt}$, where $\xi_{jt} = (A_{kl})^{-1/2} z_t u_{jt}$. By Minkowski's inequality,

$$(19) \qquad E\left\| \sum_{t=k}^{l} \xi_t \right\|^{2s} \le \left(\sum_{j=-\infty}^{\infty} \left[E\left\| \sum_{t=k}^{l} \xi_{jt} \right\|^{2s} \right]^{1/2s} \right)^{2s}.$$

A key point is that for fixed j, k, and l, $\{\xi_{jt},\mathscr{F}_{t-j}\}$ $(t = k,\ldots,l)$ form a sequence of martingale differences. Thus by Burkholder's inequality (Hall and Heyde (1980, p. 23)) there exists a $C > 0$, only depending on q and s, such that

$$(20) \qquad E\left\| \sum_{t=k}^{l} \xi_{jt} \right\|^{2s} \le CE\left(\sum_{t=k}^{l} \|\xi_{jt}\|^2 \right)^s \le C\left(\sum_{t=k}^{l} (E\|\xi_{jt}\|^{2s})^{1/s} \right)^s,$$

where the second step follows by Minkowski's inequality. Now $\|\xi_{jt}\|^2 = z_t'(A_{kl})^{-1} z_t u_{jt}^2$. Thus $(E\|\xi_{jt}\|^{2s})^{1/2} = z_t'(A_{kl})^{-1} z_t (E|u_{jt}|^{2s})^{1/s}$. By A4(a), for $r = 2s$, we can show (see Hansen (1991)) $(E|u_{jt}|^{2s})^{1/2s} \le 2c_t \psi_{|j|} \le 2(\max_i c_i) \psi_{|j|} \le K \psi_{|j|}$ for all j. It follows that $(E\|\xi_{jt}\|^{2s})^{1/s} \le z_t'(A_{kl})^{-1} z_t K^2 \psi_{|j|}^2$. Thus from (20),

$$(21) \qquad E\left\| \sum_{t=k}^{l} \xi_{jt} \right\|^{2s} \le C\left(\sum_{t=k}^{l} z_t'(A_{kl})^{-1} z_t K^2 \psi_{|j|}^2 \right)^s = C(K\psi_{|j|})^{2s} q^s$$

where we have used the fact that $\sum_{t=k}^{l} z_t'(A_{kl})^{-1} z_t = \text{tr}((A_{kl})^{-1}\sum_k z_t z_t') = \text{tr}(I) = q$. Using (21) and (19), we have $E\|\sum_{t=k}^{l}\xi_t\|^{2s} \le Cq^s K^{2s}(\sum_{j=-\infty}^{\infty}\psi_{|j|})^{2s} < \infty$. Since the bound does not depend on k and l, this implies, in view of (18), that with $l - k \ge q$, $P(\sup_{1 \le k < l \le T}\|\sum_{t=k}^{l}\xi_t\| > T^\alpha) \le C_1 T^{-2\alpha s+2}$ for some $C_1 > 0$. Let $s = 2 + c/2$ (the moment of order $4 + c$ of u_t exists by A4(i)); we can choose an $\alpha \in (0, 2/(4 + c))$ such that $T^{-2\alpha s+2} \to 0$. This proves (17) and hence the lemma when A4(i) holds.

Consider now the case where A4(ii) is assumed to hold. Here, $T^{-1}\sum_{t=[Tu]+1}^{[Tv]} z_t z_t' \to Q(v) - Q(u)$, and hence $(T^{-1}\sum_{t=[Tu]+1}^{[Tv]} z_t z_t')^{-1} \to (Q(v) - Q(u))^{-1}$ uniformly in v and u such that $v - u > \epsilon > 0$. Also, $T^{-1/2}\sum_{t=[Tu]+1}^{[Tv]} z_t u_t = O_p(1)$ uniformly using a functional central limit theorem for martingale differences. Accordingly, $|U'P_{\bar{Z}}U| = O_p(1)$ uniformly in $T_1, \ldots, T_m$ and the statement of the lemma holds with $\alpha = 0$. $Q.E.D.$

LEMMA A.5: *Under* $A1-A4$, *for some* $\alpha < 1/2$, (a) $\sup_{T_1,\ldots,T_m} X'P_{\bar{Z}}U = O_p(T^{\alpha+1/2})$; (b) $\sup_{T_1,\ldots,T_m}\bar{Z}^{0\prime}P_{\bar{Z}}U = O_p(T^{\alpha+1/2})$.

PROOF: This follows from Lemma A.4, $\|X\| = O_p(T^{1/2})$, and $\|X'P_{\bar{Z}}U\| \le \|X\|\|P_{\bar{Z}}U\|$. Similar arguments apply for part (b). $Q.E.D.$

PROOF OF LEMMA 1: By the definition of d_t, $\sum_t^T u_t d_t = U'X(\hat{\beta} - \beta^0) + U'\bar{Z}^*\hat{\delta} - U'\bar{Z}^0\delta^0$ where $\bar{Z}^*$ is the diagonal partition of Z at $(\hat{T}_1, \ldots, \hat{T}_m)$. To prove the lemma, it suffices to show $T^{-1}U'X(\hat{\beta} - \beta^0) = o_p(1)$ and $T^{-1}(U'\bar{Z}^*\hat{\delta} - U'\bar{Z}^0\delta^0) = o_p(1)$. We shall prove a stronger result. Let $(T_1, \ldots, T_m)$ be an arbitrary partition and $\bar{Z}$ be the associated diagonal partition of Z. Also let $\hat{\beta}(\{T_j\})$ and $\hat{\delta}(\{T_j\})$ be, respectively, the estimates of β and δ corresponding to this same partition. We shall prove

$$(22) \qquad \sup_{T_1,\ldots,T_m} \frac{1}{T}|U'X(\hat{\beta}(\{T_j\}) - \beta^0)| = O_p(T^{-1/2}) = o_p(1),$$

$$(23) \qquad \sup_{T_1,\ldots,T_m} \frac{1}{T}|U'\bar{Z}\hat{\delta}(\{T_j\}) - U'\bar{Z}^0\delta^0| = O_p(T^{\alpha-1/2}) = o_p(1),$$

where α is given in Lemma A.4. First consider (22). We can rewrite

$$(24) \qquad \hat{\beta}(\{T_j\}) - \beta_0 = (X'M_{\bar{Z}}X)^{-1}X'M_{\bar{Z}}\bar{Z}^0\delta^0 + (X'M_{\bar{Z}}X)^{-1}X'M_{\bar{Z}}U.$$

The first term is $O_p(1)$ uniformly over all partitions by Lemmas A.2 and A.3. The second term is $O_p(T^{\alpha-1/2}) = o_p(1)$ by Lemmas A.2 and A.4 (note that $X'M_{\bar{Z}}U = O_p(T^{\alpha+1/2})$). Thus $\hat{\beta}(\{T_j\}) - \beta^0 = O_p(1)$ uniformly over all partitions. This implies (22) since $U'X = O_p(T^{1/2})$. Next, from $\hat{\delta}(\{T_j\}) = (\bar{Z}M_X\bar{Z})^{-1}\bar{Z}'M_X Y$, $M_X X = 0$, and (2), we obtain

$$(25) \qquad U'\bar{Z}\hat{\delta}(\{T_j\}) - U'\bar{Z}^0\delta^0$$

$$= U'\bar{Z}(\bar{Z}'M_X\bar{Z})^{-1}\bar{Z}'M_X\bar{Z}^0\delta^0 + U'\bar{Z}(\bar{Z}'M_X\bar{Z})^{-1}\bar{Z}'M_X U - U'\bar{Z}^0\delta^0.$$

From the identity

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

the first term of (25) is equal to

$$(26) \qquad U'\bar{Z}(\bar{Z}'M_X\bar{Z})^{-1}\bar{Z}'M_X\bar{Z}^0\delta^0 = U'P_{\bar{Z}}[M_X + X(X'M_{\bar{Z}}X)^{-1}X'P_{\bar{Z}}M_X]\bar{Z}^0\delta^0.$$

Because $P_{\bar{Z}}$ and M_X are projection matrices, $\|M_X\bar{Z}^0\| \le \|\bar{Z}^0\| = O_p(T^{1/2})$ and $\|X'P_{\bar{Z}}M_X\bar{Z}^0\| \le \|X\|\|\bar{Z}^0\| = O_p(T)$. Hence, this term is $O_p(T^{\alpha+1/2})$ uniformly over all partitions using Lemmas A.2, A.4, and A.5. Similar arguments show that the second term of (25) is $O_p(T^{2\alpha})$. The last term of (25) is $O_p(T^{1/2})$. Combining these results and noting that $2\alpha < \alpha + 1/2$, we have $U'\bar{Z}\hat{\delta}(\{T_j\}) - U'\bar{Z}^0\delta^0 = O_p(T^{\alpha+1/2})$. This implies (23). $Q.E.D.$

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

$$(27) \qquad \sum_{t=1}^T d_t^2 \geq \sum_1 d_t^2 + \sum_2 d_t^2$$

$$= \begin{pmatrix} \hat{\beta} - \beta^0 \\ \hat{\delta}_k - \delta_j^0 \end{pmatrix}' \begin{pmatrix} \sum_1 x_t x_t' & \sum_1 x_t z_t' \\ \sum_1 z_t x_t' & \sum_1 z_t z_t' \end{pmatrix} \begin{pmatrix} \hat{\beta} - \beta^0 \\ \hat{\delta}_k - \delta_j^0 \end{pmatrix}$$

$$+ \begin{pmatrix} \hat{\beta} - \beta^0 \\ \hat{\delta}_k - \delta_{j+1}^0 \end{pmatrix}' \begin{pmatrix} \sum_2 x_t x_t' & \sum_2 z_t z_t' \\ \sum_2 z_t x_t' & \sum_2 z_t z_t' \end{pmatrix} \begin{pmatrix} \hat{\beta} - \beta^0 \\ \hat{\delta}_k - \delta_{j+1}^0 \end{pmatrix},$$

where Σ_1 extends over the set $T(\lambda_j^0 - \eta) \leq t \leq T\lambda_j^0$ and Σ_2 extends over the set $T\lambda_j^0 + 1 \leq t \leq T(\lambda_j^0 + \eta)$. Let γ_T and γ_T^* be the smallest eigenvalue of the first and second matrices in (27). Then

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

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

$$\geq (1/2) \min\{\gamma_T, \gamma_T^*\} \| \delta_j^0 - \delta_{j+1}^0 \|^2.$$

The last inequality follows from

$$(x - a)' A(x - a) + (x - b)' A(x - b) \geq (1/2)(a - b)' A(a - b)$$

for an arbitrary positive definite matrix A and for all x. Now the first matrix in (27) can be written as $(T\eta)(1/T\eta) \sum_{T(\lambda_j^0 - \eta)}^{T\lambda_j^0} w_t w_t' \equiv (T\eta) A_T$, say. By A2, the smallest eigenvalue of A_T is bounded away from zero. Thus the smallest eigenvalue of $(T\eta) A_T$, γ_T, is of the order $T\eta$. The same can be said for γ_T^*. Therefore, $\sum_1^T d_t^2 > TC \| \delta_j^0 - \delta_{j+1}^0 \|^2$ for some $C > 0$ with probability no less than $\epsilon_0 > 0$. Q.E.D.

PROOF OF PROPOSITION 2: Without loss of generality, we assume there are only three breaks ($m = 3$) and provide an explicit proof of T-consistency for $\hat{\lambda}_2$ only. The analysis for $\hat{\lambda}_1$ and $\hat{\lambda}_3$ is virtually the same (and actually simpler) and is thus omitted. For each $\epsilon > 0$, let $V_\epsilon = \{(T_1, T_2, T_3); |T_i - T_i^0| \leq \epsilon T\}$. From Proposition 1, $P(\{\hat{T}_1, \hat{T}_2, \hat{T}_3\} \in V_\epsilon) \to 1$. Therefore we only need to examine the behavior of the sum of squared residuals, $S_T(T_1, T_2, T_3)$, for those T_i such that $|T_i - T_i^0| < \epsilon T$ for all i. Also using an argument of symmetry, we can, without loss of generality, consider the case $T_2 < T_2^0$. For $C > 0$, define

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

Thus, $V_\epsilon(C) \subset V_\epsilon$. Because $S_T(\hat{T}_1, \hat{T}_2, \hat{T}_3) \leq S_T(\hat{T}_1, T_2^0, \hat{T}_3)$ with probability 1, it is enough to show that for each $\eta > 0$, there exist $C > 0$ and $\epsilon > 0$ such that for large T, $P(\min\{S_T(T_1, T_2, T_3) - S_T(T_1, T_2^0, T_3)\} \leq 0) < \eta$, or equivalently,

$$(28) \qquad P(\min\{[S_T(T_1, T_2, T_3) - S_T(T_1, T_2^0, T_3)]/(T_2^0 - T_2)\} \leq 0) < \eta,$$

where the minimum is taken over the set $V_\epsilon(C)$. Such a relation would imply that for a large C, global optimization cannot be achieved on $V_\epsilon(C)$. Thus with large probability, $|\hat{T}_2 - T_2^0| \le C$. Now denote $SSR_1 = S_T(T_1, T_2, T_3)$, $SSR_2 = S_T(T_1, T_2^0, T_3)$, and introduce $SSR_3 = S_T(T_1, T_2, T_2^0, T_3)$. By definition, we have

$$(29) \qquad S_T(T_1, T_2, T_3) - S_T(T_1, T_2^0, T_3) = (SSR_1 - SSR_3) - (SSR_2 - SSR_3).$$

This latter relation is useful because it allows us to carry the analysis in terms of two problems involving a single structural change, the first allowing an additional fourth break at time T_2^0 between T_2 and T_3 and the second an additional fourth break at time T_2 between the T_1 and T_2^0. It is then easy to derive exact expressions for (29) in terms of estimated coefficients. Let $(\hat{\delta}_1^*, \hat{\delta}_2^*, \hat{\delta}_\Delta, \hat{\delta}_3^*, \hat{\delta}_4^*)$ denote the estimator of $(\delta_1^0, \delta_2^0, \delta_2^0, \delta_3^0, \delta_4^0)$ based on the partition (T_1, T_2, T_2^0, T_3) (note δ_2^0 is repeated once). In particular, $\hat{\delta}_2^*$ is an estimate of δ_2^0 associated with the regressor $(0, \dots, 0, z_{T_1+1}, \dots, z_{T_2}, 0, \dots, 0)'$, $\hat{\delta}_\Delta$ is an estimate of δ_2^0 associated with the regressor $Z_\Delta = (0, \dots, 0, z_{T_2+1}, \dots, z_{T_2^0}, 0, \dots, 0)'$, and $\hat{\delta}_3^*$ is an estimate of δ_3^0 associated with the regressor $(0, \dots, 0, z_{T_2^0+1}, \dots, z_{T_3}, 0, \dots, 0)'$. Now consider $SSR_1 - SSR_3$; we have (e.g., Amemiya (1985, p. 31)),

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

where $\overline{W} = (X, \overline{Z})$, with $\overline{Z}$ the diagonal partition of Z at (T_1, T_2, T_3). Similarly, we have for $SSR_2 - SSR_3$,

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

where $\tilde{W} = (X, \tilde{Z})$ with $\tilde{Z}$ the diagonal partition of Z at (T_1, T_2^0, T_3). Thus

$$(30) \qquad SSR_1 - SSR_2 = (\hat{\delta}_3^* - \hat{\delta}_\Delta)' Z_\Delta' M_{\overline{W}} Z_\Delta (\hat{\delta}_3^* - \hat{\delta}_\Delta) - (\hat{\delta}_2^* - \hat{\delta}_\Delta)' Z_\Delta' M_{\tilde{W}} Z_\Delta (\hat{\delta}_2^* - \hat{\delta}_\Delta)$$
$$\ge (\hat{\delta}_3^* - \hat{\delta}_\Delta)' Z_\Delta' M_{\overline{W}} Z_\Delta (\hat{\delta}_3^* - \hat{\delta}_\Delta) - (\hat{\delta}_2^* - \hat{\delta}_\Delta)' Z_\Delta' Z_\Delta (\hat{\delta}_2^* - \hat{\delta}_\Delta).$$

The inequality is due to $Z_\Delta' M_{\tilde{W}} Z_\Delta \le Z_\Delta' Z_\Delta$. From the definition of $M_{\overline{W}}$, we have

$$(31) \qquad (SSR_1 - SSR_2)/(T_2^0 - T_2)$$
$$\ge (\hat{\delta}_3^* - \hat{\delta}_\Delta)' [Z_\Delta' Z_\Delta / (T_2^0 - T_2)] (\hat{\delta}_3^* - \hat{\delta}_\Delta)$$
$$\quad - (\hat{\delta}_3^* - \hat{\delta}_\Delta)' [Z_\Delta' \overline{W} / (T_2^0 - T_2)][\overline{W}' \overline{W}/T]^{-1} [\overline{W}' Z_\Delta / T](\hat{\delta}_3^* - \hat{\delta}_\Delta)$$
$$\quad - (\hat{\delta}_2^* - \hat{\delta}_\Delta)' \{Z_\Delta' Z_\Delta / (T_2^0 - T_2)\} (\hat{\delta}_2^* - \hat{\delta}_\Delta)$$
$$\equiv (I) - (II) - (III).$$

Consider term (I). Note first that $\hat{\delta}_i^*$ is close to δ_i^0 given that, on the set $V_\epsilon(C)$, the distance between T_i and T_i^0 can be controlled and made small by choosing a small ϵ. Noting that $\hat{\delta}_\Delta$ is estimated using observations from the second true regime only, $\hat{\delta}_\Delta$ is close to δ_2^0 for a large enough C, on $V_\epsilon(C)$. Hence, for large C, large T and small ϵ, (I) is no less than $(1/2)(\delta_3^0 - \delta_2^0)'[Z_\Delta' Z_\Delta / (T_2^0 - T_2)](\delta_3^0 - \delta_2^0)$ with large probability. Next consider term (II). It is easy to show that on $V_\epsilon(C)$, $\hat{\delta}_3^*$ and $\hat{\delta}_\Delta$ are $O_p(1)$ uniformly. Also on $V_\epsilon(C)$, $(\overline{W}' \overline{W}/T)^{-1} = O_p(1)$ and $Z_\Delta' \overline{W} / (T_2^0 - T_2) = O_p(1)$ (because $Z_\Delta' \overline{W}$ involves no more than $T_2^0 - T_2$ observations). Furthermore,

$$\|\overline{W}' Z_\Delta / T\| = \|\overline{W}' Z_\Delta / (T_2^0 - T_2)\| (T_2^0 - T_2)/T \le \epsilon O_p(1).$$

Thus (II) is no larger than $\epsilon O_p(1)$. Consider finally (III). Because both $\hat{\delta}_2^*$ and $\hat{\delta}_\Delta$ are close to δ_2^0, $\|\hat{\delta}_2 - \hat{\delta}_\Delta\| < \rho$ with large probability for every $\rho > 0$ (this is true for large T, large C, and small ϵ). Also, because $\|Z_\Delta' Z_\Delta / (T_2^0 - T_2)\| = O_p(1)$ uniformly on $V_\epsilon(C)$, term (III) is no larger than $\rho O_p(1)$. Hence, the inequality

$$(32) \qquad \frac{SSR_1 - SSR_2}{T_2^0 - T_2} \ge 2^{-1}(\delta_3^0 - \delta_2^0)' \frac{Z_\Delta' Z_\Delta}{T_2^0 - T_2} (\delta_3^0 - \delta_2^0) - \epsilon O_p(1) - \rho O_p(1)$$

holds with large probability. By A2,

$$Z'_\Delta Z_\Delta / (T_2^0 - T_2) = \frac{1}{T_2^0 - T_2} \sum_{t=T_2+1}^{T_2^0} z_t z'_t$$

has its minimum eigenvalue bounded away from zero on $V_\epsilon(C)$. Thus the first term on the right-hand side of (32) is positive and dominates the other two terms. It follows that with large probability, $(SSR_1 - SSR_2)/(T_2^0 - T_2) > 0$. This proves (28) and the proposition. $\qquad$ Q.E.D.

PROOF OF PROPOSITION 4(i): The structure of the proof is similar to that of Proposition 1 but modifications are necessary in view of the fact that $T^{-1} \sum_{t=1}^T d_t^2 \to 0$ even supposing a break is not consistently estimated when the shifts are shrinking. Using (4) and (5) (without dividing T on both sides), we can arrive at the desired contradiction if we can show that $\sum_{t=1}^T d_t^2 > 2 \sum_{t=1}^T u_t d_t$ in the limit as $T \to \infty$. To do this we show that $\sum_{t=1}^T d_t^2$ diverges at a faster rate than $\sum_{t=1}^T u_t d_t$.

We will make use of v_T being small to strengthen the result of (22) and (23). We shall drop the subscript T in $\delta_{T,i}^0$. From $\delta_i^0 - \delta_{i+1}^0 = O(v_T)$ under A6, by adding and subtracting terms, we have $\delta_i^0 - \delta_j^0 = O(v_T)$ for all i and j. Now consider (24). The first term on the right-hand side can be rewritten as $(X'M_{\bar{Z}}X)^{-1} X'M_{\bar{Z}}(\bar{Z}^0 - \bar{Z})\delta^0$ because $M_{\bar{Z}}\bar{Z} = 0$. A key to the proof lies in the fact that $(\bar{Z}^0 - \bar{Z})\delta^0$ depends on changes in the parameters (i.e. $\delta_i^0 - \delta_j^0$). In the case of a single change point, for example, assume $T_1 < T_1^0$; then

$$(\bar{Z}^0 - \bar{Z})\delta^0 = (0, \ldots, 0, z_{T_1+1}, \ldots, z_{T_1^0}, 0, \ldots, 0)'(\delta_1^0 - \delta_2^0).$$

This implies that $X'M_{\bar{Z}}(\bar{Z}^0 - \bar{Z})\delta^0$ is at most $O_p(T)v_T$. By Lemma A.2,

$$(X'M_{\bar{Z}}X)^{-1} X'M_{\bar{Z}}(\bar{Z}^0 - \bar{Z})\delta^0 = O_p(v_T).$$

This implies that, from (24),

$$\hat{\beta}(\{T_j\}) - \beta^0 = O_p(v_T) + O_p(T^{\alpha-1/2}).$$

Thus

$$(33) \qquad U'X(\hat{\beta}(\{T_j\}) - \beta^0) = O_p(T^{1/2}v_T) + O_p(T^\alpha)$$

over all partitions. Next, we combine the first and the third terms of (25) and rewrite them as

$$(34) \qquad U'\bar{Z}(\bar{Z}'M_X\bar{Z})^{-1} \bar{Z}'M_X(\bar{Z}^0 - \bar{Z})\delta^0 + U'(\bar{Z} - \bar{Z}^0)\delta^0.$$

Each term above involves $(\bar{Z} - \bar{Z}^0)\delta^0$. Using the argument in proving (26) and (33), the first term of (34) is $O_p(T^{\alpha+1/2}v_T)$ and the second term is $O_p(T^{1/2}v_T)$, which is dominated by the first. Since the middle term of (25) is $O_p(T^{2\alpha})$, we have

$$(35) \qquad U'\bar{Z}\hat{\delta}(\{T_j\}) - U'\bar{Z}^0\delta^0 = O_p(T^{\alpha+1/2}v_T) + O_p(T^{2\alpha}).$$

Noting that (35) dominates (33), we have $\sum_{t=1}^T u_t d_t = O_p(T^{\alpha+1/2}v_T + T^{2\alpha})$.

Next, consider $\sum_{t=1}^T d_t^2$. The proof of Lemma 2 is not changed under shrinking shifts. If there exists a change point that cannot be consistently estimated, then

$$\sum_{t=1}^T d_t^2 > TC\|\delta_j^0 - \delta_{j+1}^0\|^2 > TC'v_T^2$$

for some $C' > 0$. Thus $\sum_{t=1}^T d_t^2 > 2\sum_{t=1}^T u_t d_t$ if $Tv_T^2/(T^{\alpha+1/2}v_T + T^{2\alpha}) \to \infty$. This is the case if $T^{(1/2)-\alpha}v_T \to \infty$. Under $E|u_t|^{2/\vartheta} < \infty$ of A6, we can choose α such that $\alpha < \vartheta$ in Lemma A.4. Thus, $T^{(1/2)-\alpha}v_T \ge T^{(1/2)-\vartheta}v_T \to \infty$ by A6. $\qquad$ Q.E.D.

To prove Proposition 4(ii), we first prove a lemma, which generalizes the Hajek and Renyi inequality to mixingales.

LEMMA A.6: *Let $\{\xi_t, \mathscr{F}_t\}$ be a $q \times 1$ L^2 mixingale satisfying* (a)–(e) *of A4(i) with u_i replaced by ξ_i and $|\cdot|$ replaced by $\|\cdot\|$. Then there exists an $L < \infty$ such that, for every $c > 0$ and $m > 0$,*

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

PROOF: Let $\xi_{jt} = E(\xi_t|\mathscr{F}_{t-j}) - E(\xi_t|\mathscr{F}_{t-j-1})$. Then $\xi_t = \sum_{j=-\infty}^{\infty} \xi_{jt}$, and so $\sum_{t=1}^{k} \xi_t = \sum_{j=-\infty}^{\infty}\sum_{t=1}^{k} \xi_{jt}$. Thus, for each $N > 0$,

$$P\left(\sup_{N \geq k \geq m} \frac{1}{k}\left\|\sum_{t=1}^{k} \xi_t\right\| > c\right) \leq P\left(\sum_{j=-\infty}^{\infty} \sup_{N \geq k \geq m} \frac{1}{k}\left\|\sum_{t=1}^{k} \xi_{jt}\right\| > c\right).$$

For each j, $\{\xi_{jt}, \mathscr{F}_{t-j}\}$ forms a sequence of martingale differences. Let $a_j > 0$ for all j such that $\sum_{j=-\infty}^{\infty} a_j = 1$. The right-hand side above is bounded by

$$\sum_{j=-\infty}^{\infty} P\left(\sup_{N \geq k \geq m} \frac{1}{k}\left\|\sum_{t=1}^{k} \xi_{jt}\right\| > a_j c\right)$$

$$\leq \frac{1}{c^2} \sum_{j=-\infty}^{\infty} a_j^{-2}\left(m^{-2}\sum_{i=1}^{m} E\|\xi_{ji}\|^2 + \sum_{i=m+1}^{N} i^{-2} E\|\xi_{ji}\|^2\right);$$

the latter bound is due to Hajek and Renyi's inequality for martingale differences. From the definition of a mixingale and A4(i)(c), $E\|\xi_{ji}\|^2 \leq 4c_i^2\psi_{|j|}^2 \leq 4K^2\psi_{|j|}^2$. Thus the above is bounded by $c^{-2}4K^2(\sum_{j=-\infty}^{\infty} a_j^{-2}\psi_{|j|}^2)(m^{-1} + \sum_{i=m+1}^{\infty} i^{-2})$. Since $\sum_{i=m+1}^{\infty} i^{-2} \leq 2m^{-1}$, if we let $L = 12K^2(\sum_{j=-\infty}^{\infty} a_j^{-2}\psi_{|j|}^2)$, then the desired upper bound is obtained for a fixed N. Since the bound does not depend on N, the lemma is obtained by letting $N \rightarrow \infty$. It remains to choose appropriate a_j's such that $\sum_j a_j^{-2}\psi_{|j|}^2$ is bounded. Let $\nu_0 = 1$ and $\nu_j = j^{-1-\kappa}(j \geq 1)$, where $\kappa > 0$ as given in A4(i)(d). Let $a_j = \nu_j/(1 + 2\sum_{l=1}^{\infty}\nu_l)$ and $a_{-j} = a_j$ for $j \geq 0$. Then $\sum_j a_j = 1$. By Assumption A4(i)(d),

$$\sum_j a_j^{-2}\psi_{|j|}^2 = \left(\psi_0^2 + 2\sum_{j=1}^{\infty} j^{2+2\kappa}\psi_j^2\right)\left(1 + 2\sum_{j=1}^{\infty} j^{-2-2\kappa}\right) < \infty. \qquad Q.E.D.$$

PROOF OF PROPOSITION 4(ii): We shall maintain all the notations in the proof of Proposition 2. Define a new set

$$V_\epsilon^*(C) = \{(T_1, T_2, T_3); |T_i - T_i^0| \leq \epsilon T, 1 \leq i \leq 3, T_2 - T_2^0 < -C/v_T^2\},$$

which is a subset of V_ϵ. We only need to show that (28) holds when the minimum is taken over $V_\epsilon^*(C)$. We can prove that, uniformly on the set $V_\epsilon^*(C)$,

$$\hat{\delta}_i^* - \delta_i^0 = \epsilon O_p(v_T) + O_p(T^{-1/2}) \qquad (i = 1, \ldots, 4), \tag{36}$$

$$\hat{\delta}_\Delta - \delta_2^0 = (Z_\Delta' Z_\Delta)^{-1} Z_\Delta U + \epsilon O_p(v_T) + O_p(T^{-1/2}). \tag{37}$$

The above is easily seen to be true in the case of pure structural changes. In this case, for example, $\hat{\delta}_\Delta - \delta_2^0$ is given exactly by $(Z_\Delta' Z_\Delta)^{-1} Z_\Delta U$. In the case of partial structural changes, the proof of (36) and (37) is more complicated. We shall omit the details and give a brief explanation instead. A detailed proof is available upon request. The term $\epsilon O_p(v_T)$ on the right-hand side of (36) is due to misspecification in the sense that T_i may not be the same as T_i^0. However, this misspecification is

controlled by choosing a small ϵ. The dependence on v_T is explained in part (i). The term $O_p(T^{-1/2})$ is related to disturbances and this specific rate is due to the fact that each δ_i^* is estimated with a positive fraction of the sample for partitions in $V_\epsilon^*(C)$. The last two terms of (37) are due to spillover from the misspecification via partial structural changes.

Using (36) and (37), expression (I) in (31) is no smaller than

$$(\delta_3^0 - \delta_2^0)'[Z_\Delta'Z_\Delta/(T_2^0 - T_2)](\delta_3^0 - \delta_2^0) - \epsilon O_p(v_T^2) - O_p(T^{-1/2}v_T) - O_p(T^{-1}).$$

Because the minimum eigenvalue of $Z_\Delta'Z_\Delta/(T_2^0 - T_2)$ is bounded away from zero on $V_\epsilon^*(C)$ for all large C and $\delta_3^0 - \delta_2^0 = O(v_T)$, (I) is no smaller than $Av_T^2 - \epsilon O_p(v_T^2)$, where A is a positive constant. Note that $\epsilon O_p(v_T^2)$ dominates $O_p(T^{-1/2}v_t)$ and $O_p(T^{-1})$. The latter two terms also appear in (II) and (III) of (31) and will be absorbed into $\epsilon O_p(v_T^2)$. Expression (II) in (31) is bounded by $\epsilon O_p(v_T^2)$. Expression (III) is bounded by

$$(T_2^0 - T_2)^{-1}U'Z_\Delta(Z_\Delta'Z_\Delta)^{-1}Z_\Delta'U + \epsilon O_p(v_T^2).$$

From $[Z_\Delta'Z_\Delta/(T_2^0 - T_2)]^{-1} = O_p(1)$ on $V_\epsilon(C)$ for large C, (III) is further bounded by

$$O_p(1)\|(T_2^0 - T_2)^{-1}Z_\Delta U\|^2 + \epsilon O_p(v_T^2).$$

In summary,

(38)
$$\frac{SSR_1 - SSR_2}{T_2^0 - T_2} \geq Av_T^2 + \left\|\frac{Z_\Delta U}{T_2^0 - T_2}\right\|^2 O_p(1) - \epsilon O_p(v_T^2).$$

For every $\eta > 0$, we can choose a small $\epsilon > 0$ such that $P(\epsilon O_p(v_T^2) > Av_T^2/2) < \eta$ and we can also choose $B < \infty$ such that $P(|O_p(1)| > B) < \nu$. Thus

$$P(\min\{(SSR_1 - SSR_2)/(T_2^0 - T_2)\} \leq 0)$$

$$\leq 2\eta + P\left(\max\left\{B\|(T_2^0 - T_2)^{-1}Z_\Delta U\|^2\right\} > Av_T^2/2\right)$$

$$= 2\eta + P\left(\max_{T_2 < T_2^0 - Cv_T^{-2}}(T_2^0 - T_2)^{-1}\left\|\sum_{t=T_2+1}^{T_2^0}z_t u_t\right\| > [A/(2B)]^{1/2}v_T\right).$$

By Lemma A.6 with $\xi_t = z_t u_t$, $c = [A/(2B)]^{1/2}v_T$, and $m = Cv_T^{-2}$ (applied with data order reversed, i.e. treating T_2^0 as the first observation), the above probability is bounded by

$$2\eta + (2B/A)L(v_T^2 Cv_T^{-2})^{-1} = 2\eta + (2B/A)LC^{-1} < 3\eta$$

for large C.   Q.E.D.

PROOF OF PROPOSITION 6: Note that we can write

$$F_T(\lambda_1, \ldots, \lambda_k; q) = (SSR_0 - SSR_k)/[kq(T - (k+1)q - p)^{-1}SSR_k],$$

where SSR_0 and SSR_k are the sum of squared residuals under the null and alternative hypotheses, respectively. We have

$$(T - (k+1)q - p)^{-1}SSR_k \to_p \sigma^2.$$

Hence, we concentrate on the limit of $F_T^* = SSR_0 - SSR_k$. Now, let $D^U(i,j)$ ($D^R(i,j)$, resp.) be the sum of squared residuals from the unrestricted (restricted, resp.) model using data from segments i to j (inclusively), i.e. from observation $T_{i-1} + 1$ to T_j. We can write $F_T^* = D^R(1, k+1) - \sum_{i=1}^{k+1}D^U(i,i)$, or

(39)
$$F_T^* = \sum_{i=1}^k [D^R(1, i+1) - D^R(1, i) - D^U(i+1, i+1)] + D^R(1,1) - D^U(1,1).$$

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

$$\hat{\delta}_{1,j}^R = (Z_{1,j}'Z_{1,j})^{-1}Z_{1,j}'(Y_{1,j} - X_{1,j}\hat{\beta}^R), \quad \text{and}$$

$$\hat{\delta}_j^U = (Z_j'Z_j)^{-1}Z_j'(Y_j - X_j\hat{\beta}^U).$$

Using the fact that, under the null hypothesis,

$$Y = X\beta + Z\delta + U = X\beta + \overline{Z}\overline{\delta} + U \quad \text{and}$$

$$Y_j = X_j\beta + Z_j\delta + U_j$$

(with $\overline{\delta} = (\delta, \ldots, \delta)$ a $q(k+1)$ vector with δ defined by $\delta_1 = \delta_2 = \cdots = \delta_{k+1} \equiv \delta$), straightforward algebra yields

$$D^R(1,j) = \|(I - P_{Z_{1,j}})(U_{1,j} - X_{1,j}A_T)\|^2 \quad \text{and}$$

$$D^U(j,j) = \|(I - P_{Z_j})(U_j - X_j\overline{A}_T)\|^2,$$

where $A_T = (X'M_ZX)^{-1}X'M_ZU$, and $\overline{A}_T = (X'M_{\overline{Z}}X)^{-1}X'M_{\overline{Z}}U$. Consider the ith element in the summation defining F_T^* in (39); we have

(40) $F_{T,i} = D^R(1,i+1) - D^R(1,i) - D^U(i+1,i+1)$

$$= \|(I - P_{Z_{1,i+1}})(U_{1,i+1} - X_{1,i+1}A_T)\|^2 - \|(I - P_{Z_{1,i}})(U_{1,i} - X_{1,i}A_T)\|^2$$

$$- \|(I - P_{Z_{i+1}})(U_{i+1} - X_{i+1}\overline{A}_T)\|^2.$$

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

$$U_{1,i+1}'U_{1,i+1} = U_{1,i}'U_{1,i} + U_{i+1}'U_{i+1},$$

$$X_{1,i+1}'X_{1,i+1} = X_{1,i}'X_{1,i} + X_{i+1}'X_{i+1},$$

and

$$U_{1,i+1}'X_{1,i+1} = U_{1,i}'X_{1,i} + U_{i+1}'X_{i+1},$$

we deduce that

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

$$+ 2S_{i+1}'H_{i+1}^{-1}K_{i+1}A_T - 2S_i'H_i^{-1}K_iA_T$$

$$- 2(S_{i+1} - S_i)'[H_{i+1} - H_i]^{-1}(K_{i+1} - K_i)\overline{A}_T$$

$$+ 2(M_{i+1} - M_i)'(\overline{A}_T - A_T) + (\overline{A}_T - A_T)'(L_{i+1} - L_i)(\overline{A}_T - A_T).$$

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

(i) $T^{-1/2}(X_{1,j}, Z_{1,j})'U_{1,j} \Rightarrow \sigma(B_1(\lambda_j), B_2(\lambda_j))' \equiv \sigma B(\lambda_j)'$

where $B(r)$ is a $(q + p)$ dimensional vector Brownian motion with covariance matrix

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

(ii) $T^{-1}(X_{1,j}, Z_{1,j})'(X_{1,j}, Z_{1,j}) \to_p \sigma^2 \lambda_j Q.$

From these two limits, we deduce easily the following results:

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

(b) $T^{-1} H_j \to_p \sigma^2 \lambda_j Q_{22};$

(c) $T^{-1} K_j \to_p \sigma^2 \lambda_j Q_{21};$

(d) $T^{-1} L_j \to_p \sigma^2 \lambda_j Q_{11};$

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

(f) $T^{1/2} A_T \Rightarrow \sigma^{-1}(Q_{11} - Q_{12} Q_{22}^{-1} Q_{21})^{-1}(B_1(1) - Q_{12} Q_{22}^{-1} B_2(1)) \equiv A^*.$

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

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

(ii) $T^{-1} X' \bar{Z} \to_p \sigma^2(e'\Lambda \otimes Q_{12})$ where $e' = (1, 1, \ldots, 1),$

a $(k + 1)$ vector;

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

We then obtain

(42) $T^{1/2} \bar{A}_T = [T^{-1} X'X - T^{-1} X' \bar{Z}(T^{-1} \bar{Z}' \bar{Z})^{-1} T^{-1} \bar{Z}' X]^{-1}$

$$\times [T^{-1/2} X'U - T^{-1} X' \bar{Z}(T^{-1} \bar{Z}' \bar{Z})^{-1} T^{-1/2} \bar{Z}' U]$$

$$\Rightarrow \sigma^{-1}[Q_{11} - (e'\Lambda \otimes Q_{12})(\Lambda \otimes Q_{22})^{-1}(\Lambda e \otimes Q_{21})]^{-1}$$

$$\times [B_1(1) - (e'\Lambda \otimes Q_{12})(\Lambda \otimes Q_{22})^{-1} B^*]$$

$$= \sigma^{-1}[Q_{11} - (e'\Lambda e \otimes Q_{12} Q_{22}^{-1} Q_{21})]^{-1}[B_1(1) - (e' \otimes Q_{12} Q_{22}^{-1}) B^*]$$

$$= \sigma^{-1}[Q_{11} - Q_{12} Q_{22}^{-1} Q_{21}]^{-1}[B_1(1) - Q_{12} Q_{22}^{-1} B_2(1)] \equiv A^*.$$

The second equality follows since $e'\Lambda e = 1$ and $(e' \otimes Q_{12} Q_{22}^{-1}) B^* = Q_{12} Q_{22}^{-1} B_2(1)$. Using the results stated above we easily deduce that $(M_{i+1} - M_i)'(\bar{A}_T - A_T) \Rightarrow 0$, $(\bar{A}_T - A_T)'(L_{i+1} - L_i)(\bar{A}_T - A_T) \Rightarrow 0$, and

$$S'_{i+1} H_{i+1}^{-1} K_{i+1} A_T - S'_i H_i^{-1} K_i A_T - (S_{i+1} - S_i)'[H_{i+1} - H_i]^{-1}(K_{i+1} - K_i) \bar{A}_T$$

$$\Rightarrow \sigma B_2(\lambda_{i+1}) Q_{22}^{-1} Q_{21} A^* - \sigma B_2(\lambda_i) Q_{22}^{-1} Q_{21} A^*$$

$$- \sigma(B_2(\lambda_{i+1}) - B_2(\lambda_i)) Q_{22}^{-1} Q_{21} A^* = 0.$$

Hence, we are left with

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

and we deduce, using the fact that $B_2(\lambda_j) = \sigma Q_{22}^{1/2} W_q(\lambda_j)$,

$$(43) \qquad F_{T,i} \Rightarrow -B_2(\lambda_{i+1})'Q_{22}^{-1}B_2(\lambda_{i+1})/\lambda_{i+1} + B_2(\lambda_i)'Q_{22}^{-1}B_2(\lambda_i)/\lambda_i$$

$$+ (B_2(\lambda_{i+1}) - B_2(\lambda_i))'Q_{22}^{-1}(B_2(\lambda_{i+1}) - B_2(\lambda_i))/(\lambda_{i+1} - \lambda_i)$$

$$= -\sigma^2\|W_q(\lambda_{i+1})\|^2/\lambda_{i+1} + \sigma^2\|W_q(\lambda_i)\|^2/\lambda_i$$

$$+ \sigma^2\|W_q(\lambda_{i+1}) - W_q(\lambda_i)\|^2/(\lambda_{i+1} - \lambda_i)$$

$$= \sigma^2\|\lambda_i W_q(\lambda_{i+1}) - \lambda_{i+1}W_q(\lambda_i)\|^2/\lambda_{i+1}\lambda_i(\lambda_{i+1} - \lambda_i).$$

Finally, it is easy to verify that $D^R(1,1) - D^U(1,1) \Rightarrow 0$. Note that this convergence result holds jointly for $i = 1, \ldots, k$; hence

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

and the result of Proposition 6 follows. $Q.E.D.$

PROOF OF PROPOSITION 7: For simplicity, we present the arguments in the case of a pure structural change. Let $SSR(i,j)$ be the minimized sum of squared residuals for the segment containing observations from $(i + 1)$ to j; then we can write

$$(44) \qquad F_T(l+1|l) = \sup_{1 \le i \le l+1} \sup_{\tau \in \Delta_{i,\eta}} \{SSR(\hat{T}_{i-1}, \hat{T}_i) - SSR(\hat{T}_{i-1}, \tau) - SSR(\tau, \hat{T}_i)\}/\hat{\sigma}^2.$$

Under Assumptions A8–A9, arguments as in the proof of Proposition 6 show that

$$(45) \qquad \sigma^{-2} \sup_{\tau \in \Delta_{i,\eta}^0} \{SSR(T_{i-1}^0, T_i^0) - SSR(T_{i-1}^0, \tau) - SSR(\tau, T_i^0)\}$$

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

where $\Delta_{i,\eta}^0$ is as defined in (11) with $\hat{T}_i$ replaced by T_i^0. Under the null hypothesis, Proposition 2 asserts that $\hat{T}_i = T_i^0 + O_p(1)$. Using this result, we can show that (45) also holds with T_{i-1}^0 and T_i^0 replaced by $\hat{T}_{i-1}$ and $\hat{T}_i$, respectively. In addition, because over different regimes $SSR(\cdot, \cdot)$ are computed using nonoverlapping observations, the weak limits in (45) for different i's are independent. Thus the limit of (44) is the maximum of $l + 1$ independent random variables in the form of (45).

PROOF OF LEMMA 3: We show that $S(\tau)$ for $\tau \in [0, 1]$ has a unique minimum at λ_1. The function $S(\tau)$ has different expressions over $[0, 1]$. Some algebra reveals that

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

which is nonnegative. Under the assumption that $S(\lambda_1) < S(\lambda_2)$, the expression in brackets is nonzero, so $S(\tau) - S(\lambda_1)$ is strictly positive for $\tau < \lambda_1$. By symmetry (regarded as reversing the data order), $S(\tau) - S(\lambda_2)$ is nonnegative for $\tau > \lambda_2$. Thus for $\tau \in [\lambda_2, 1]$,

$$S(\tau) - S(\lambda_1) = S(\tau) - S(\lambda_2) + S(\lambda_2) - S(\lambda_1) \ge S(\lambda_2) - S(\lambda_1) > 0.$$

It remains to consider the case where $\tau \in (\lambda_1, \lambda_2)$. Again, simple algebra shows

$$S(\tau) - S(\lambda_1) = (\tau - \lambda_1)\frac{\lambda_2}{\tau}\left[\frac{\lambda_1}{\lambda_2}(\mu_2 - \mu_1)^2 - \frac{\tau(1-\lambda_2)(1-\lambda_2)}{\lambda_2(1-\tau)(1-\lambda_1)}(\mu_3 - \mu_2)^2\right]$$

$$\geq (\tau - \lambda_1)\frac{\lambda_2}{\tau}\left[\frac{\lambda_1}{\lambda_2}(\mu_2 - \mu_1)^2 - \frac{(1-\lambda_2)}{(1-\lambda_1)}(\mu_3 - \mu_2)^2\right]$$

$$\geq (\tau - \lambda_1)[S(\lambda_2) - S(\lambda_1)]$$

where the first inequality follows from $[\tau(1-\lambda_2)]/[\lambda_2(1-\tau)] \leq 1$ and the second inequality follows from $\lambda_2/\tau \geq 1$. Thus $S(\tau) - S(\lambda_1)$ is strictly positive for $\tau \in (\lambda_1, \lambda_2)$ and we have shown that $S(\tau)$ has a unique global minimum at λ_1 when $S(\lambda_1) < S(\lambda_2)$. Because $S_T(\hat{T}_a) \leq S_T([T\lambda_1])$, it follows that $\hat{T}_a/T$ is consistent for λ_1. Q.E.D.

REFERENCES

AMEMIYA, T. (1985): *Advanced Econometrics*. Cambridge: Harvard University Press.
ANDREWS, D. W. K. (1988): "Laws of Large Numbers for Dependent Nonidentically Distributed Random Variables," *Econometric Theory*, 4, 458–467.
——— (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.
——— (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856.
ANDREWS, D. W. K., I. LEE, AND W. PLOBERGER (1996): "Optimal Changepoint Tests for Normal Linear Regression," *Journal of Econometrics*, 70, 9–38.
ANDREWS, D. W. K., AND W. PLOBERGER (1994): "Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative," *Econometrica*, 62, 1383–1414.
BAI, J. (1994a): "Least Squares Estimation of a Shift in Linear Processes," *Journal of Time Series Analysis*, 15, 453–472.
——— (1994b): "Estimation of Structural Change Based on Wald Type Statistics," Working Paper No. 94-6, Department of Economics, M.I.T., forthcoming in *Review of Economics and Statistics*.
——— (1994c): "GMM Estimation of Multiple Structural Changes," NSF Proposal.
——— (1995): "Least Absolute Deviation Estimate of a Shift," *Econometric Theory*, 11, 403–436.
——— (1997): "Estimating Multiple Breaks One at a Time," *Econometric Theory*, 13, 315–352.
BAI, J., AND P. PERRON (1996): "Computation and Analysis of Multiple Structural Change Models," Manuscript in Preparation, Université de Montréal.
BHATTACHARYA, P. K. (1994): "Some Aspects of Change-point Analysis," *Change Point Problems*, IMS Lecture Notes—Monograph Series, Vol. 23, ed. by E. Carlstein, H.-G, Müller, and D. Siegmund. Hayward, CA: Institute of Mathematical Statistics, 28–56.
BROWN, R. L., J. DURBIN, AND J. M. EVANS (1975): "Techniques for Testing the Constancy of Regression Relationships Over Time," *Journal of the Royal Statistical Society, Series B*, 37, 149–192.
CHONG, T. T-L. (1994): "Consistency of Change-Point Estimators When the Number of Change-Points in Structural Models is Underspecified," Manuscript, Department of Economics, University of Rochester.
DELONG, D. M. (1981): "Crossing Probabilities for a Square Root Boundary by a Bessel Process," *Communications in Statistics-Theory and Methods*, A10(21), 2197–2213.
FEDER, P. I. (1975): "On Asymptotic Distribution Theory in Segmented Regression Problem: Identified Case," *Annals of Statistics*, 3, 49–83.
GALLANT, A. R., AND W. A. FULLER (1973): "Fitting Segmented Polynomial Regression Models whose Join Points have to be Estimated," *Journal of the American Statistical Association*, 68, 144–147.
GARCIA, R., AND P. PERRON (1996): "An Analysis of the Real Interest Rate under Regime Shifts," *Review of Economics and Statistics*, 78, 111–125.

HALL, P., AND C. C. HEYDE (1980): *Martingale Limit Theory and its Applications*. New York: Academic Press.

HANSEN, B. E. (1991): "Strong Laws for Dependent Heterogeneous Processes," *Econometric Theory*, 7, 213–221.

HOSOYA, Y. (1989): "Hierarchical Statistical Models and a Generalized Likelihood Ratio Test," *Journal of the Royal Statistical Society, Series B*, 51, 435–447.

KRISHNAIAH, P. R., AND B. Q. MIAO (1988): "Review about Estimation of Change Points," in *Handbook of Statistics*, Vol. 7, ed. by P. R. Krishnaiah and C. R. Rao. New York: Elsevier.

LIU, J., S. WU, AND J. V. ZIDEK (1997): "On Segmented Multivariate Regressions," *Statistica Sinica*, 7, 497–525.

MCLEISH, D. L. (1975): "A Maximal Inequality and Dependent Strong Laws," *The Annals of Probability*, 5, 829–839.

PERRON, P. (1989): "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis," *Econometrica*, 57, 1361–1401.

POLLARD, D. (1984): *Convergence of Stochastic Processes*. New York: Springer-Verlag.

YAO, Y.-C. (1987): "Approximating the Distribution of the ML Estimate of the Change-Point in a Sequence of Independent r.v.'s," *Annals of Statistics*, 3, 1321–1328.

——— (1988): "Estimating the Number of Change-Points via Schwarz' Criterion," *Statistics and Probability Letters*, 6, 181–189.

YAO, Y.-C., AND S. T. AU (1989): "Least Squares Estimation of a Step Function," *Sankhyā*, 51, Ser. A, 370–381.

ZACKS, S. (1983): "Survey of Classical and Bayesian Approaches to the Change-Point Problem: Fixed and Sequential Procedures of Testing and Estimation," in *Recent Advances in Statistics*, ed. by M. H. Rivzi, J. S. Rustagi, and D. Sigmund. New York: Academic Press, 245–269.