Econometrics Journal (2003), volume **6**, pp. 72–78.

Critical values for multiple structural change tests

JUSHAN BAI[†] AND PIERRE PERRON[‡]

[†]*Department of Economics, New York University, 269 Mercer Street, 7th Floor, New York, NY 10003, USA*
E-mail: jushan.bai@nyu.edu
[‡]*Department of Economics, Boston University, 270 Bay State Rd., Boston, MA 02215, USA*
E-mail: perron@bu.edu

Received: July 2002

Summary Bai and Perron (1998) considered theoretical issues related to the limiting distribution of estimators and test statistics in the linear model with multiple structural changes. The asymptotic distributions of the tests depend on a trimming parameter ϵ and critical values were tabulated for $\epsilon = 0.05$. As discussed in Bai and Perron (2000), larger values of ϵ are needed to achieve tests with correct size in finite samples, when allowing for heterogeneity across segments or serial correlation in the errors. The aim of this paper is to supplement the set of critical values available with other values of ϵ to enable proper empirical applications. We provide response surface regressions valid for a wide range of parameters.

Keywords: *Hypothesis testing, Response surface, Simulations, Change-point, Segmented regressions.*

1. INTRODUCTION

Bai and Perron (1998), henceforth BP, considered estimating multiple structural changes in a linear model. The results are obtained under a general framework of partial structural changes which allows a subset of the parameters not to change.[1] Methods to efficiently compute estimates are discussed in Bai and Perron (2003). BP also addressed the problem of testing for multiple structural changes under very general conditions on the data and the errors: they considered a type test for the null hypothesis of no change vs. a pre-specified number of changes and also vs. an alternative of an arbitrary number of changes (up to some maximum), as well as a procedure that allows one to test the null hypothesis of, say, ℓ changes, vs. the alternative hypothesis of $\ell + 1$ changes. The latter is particularly useful in that it allows a specific to general modeling strategy to consistently determine the appropriate number of changes in the data. The tests can be constructed allowing different serial correlation in the errors, different distribution for the data and the errors across segments or imposing a common structure.

The relevant asymptotic distributions depend on a trimming parameter $\epsilon = h/T$ where T is the sample size and h is the minimal permissible length of a segment. Critical values were provided for $\epsilon = 0.05$. As discussed in Bai and Perron (2000), a trimming as small as 5% of the sample can lead to substantial size distortions when allowing different variances of the

[1]Other studies related to multiple structural changes include Andrews *et al.* (1996), Garcia and Perron (1996) and Liu *et al.* (1997), among others.

errors across segments or when serial correlation is permitted. This is because one is then trying to estimate various quantities using very few observations; for example, if $T = 100$ and $\epsilon = 0.05$, one ends up estimating, for some segments, quantities like the variance of the residuals using only five observations. Similarly, with serial correlation a heteroscedasticity and autocorrelation consistent covariance matrix estimator would need to be applied to very short samples. The estimates are then highly imprecise and the tests accordingly show size distortions. When allowing different variances across segments or serial correlation, a higher value of ϵ should be used.

This paper is structured as follows. Section 2 discusses the relevant framework considered and Section 3, the test statistics. Section 4 presents the summary regression used to obtain the critical values of interest. Brief concluding remarks are presented in Section 4.

2. THE MODEL AND ESTIMATORS

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

$$y_t = x_t'\beta + z_t'\delta_j + u_t, \qquad t = T_{j-1} + 1, \ldots, T_j, \tag{2.1}$$

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

The multiple linear regression system (2.1) may be expressed in matrix form as

$$Y = X\beta + \overline{Z}\delta + U, \tag{2.2}$$

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 $(T_1, \ldots, T_m)$, i.e. $\overline{Z} = \text{diag}(Z_1, \ldots, Z_{m+1})$ with $Z_i = (z_{T_{i-1}+1}, \ldots, z_{T_i})'$. 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. The method of estimation considered is based on the least-squares principle. For each m-partition $(T_1, \ldots, T_m)$, the associated least-squares estimates of β and δ_j are obtained by minimizing the sum of squared residuals

$$(Y - X\beta - \overline{Z}\delta)'(Y - X\beta - \overline{Z}\delta) = \sum_{i=1}^{m+1} \sum_{t=T_{i-1}+1}^{T_i} [y_t - x_t'\beta - z_t'\delta_i]^2. \tag{2.3}$$

To carry out the asymptotic analysis, we need to impose some restrictions on the possible values of the break dates. In particular, each break date must be asymptotically distinct and bounded from the boundaries of the sample. To this effect, let $\lambda_i = T_i/T$ $(i = 1, \ldots, m)$ and define the following set for some arbitrary positive number ϵ, a trimming parameter which imposes a minimal length h for a segment, i.e. $\epsilon = h/T$,

$$\Lambda_\epsilon = \{(\lambda_1, \ldots, \lambda_m); |\lambda_{i+1} - \lambda_i| \geq \epsilon, \lambda_1 \geq \epsilon, \lambda_m \leq 1 - \epsilon\}. \tag{2.4}$$

Let $\hat{\beta}(\{T_j\})$ and $\hat{\delta}(\{T_j\})$, denote the estimates based on the given m-partition $(T_1, \ldots, T_m)$, denoted $\{T_j\}$. Substituting these 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

$$(\hat{T}_1, \ldots, \hat{T}_m) = \text{argmin}_{(\lambda_1, \ldots, \lambda_m) \in \Lambda_\epsilon} S_T(T_1, \ldots, T_m), \tag{2.5}$$

i.e. with the minimization taken over all partitions $(T_1, \ldots, T_m)$ such that $T_i - T_{i-1} \geq h = T\epsilon$. The regression parameter estimates are the estimates associated with the m-partition $\{\hat{T}_j\}$. A method, based on a dynamic programming algorithm, to efficiently compute these estimates is presented in Bai and Perron (2003).

3. TEST STATISTICS FOR MULTIPLE BREAKS

3.1. A test of no break vs. a fixed number of breaks

BP considered the sup F type test of no structural break ($m = 0$) vs. $m = k$ breaks. Let R be the conventional matrix such that $(R\delta)' = (\delta_1' - \delta_2', \ldots, \delta_k' - \delta_{k+1}')$. Define

$$F_T(\lambda_1, \ldots, \lambda_k; q) = \frac{1}{T}\left(\frac{T - (k+1)q - p}{kq}\right)\hat{\delta}' R'(R\hat{V}(\hat{\delta})R')^{-1}R\hat{\delta}, \tag{3.6}$$

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

$$V(\hat{\delta}) = p \lim_{T\to\infty} T(\overline{Z}' M_X \overline{Z})^{-1}\overline{Z}' M_X \Omega M_X \overline{Z}(\overline{Z}' M_X \overline{Z})^{-1} \tag{3.7}$$

where $M_X = I - X(X'X)^{-1}X'$. Note that if some restrictions are imposed on the nature of the heterogeneity across segments and/or serial correlation in the errors, a different form of $V(\hat{\delta})$ results. For example, imposing the same variance and no autocorrelation in the errors (though allowing the distribution of the regressors to differ across segments) leads to $V(\hat{\delta}) = p \lim_{T\to\infty} T(\overline{Z}' M_X \overline{Z})^{-1}$.

Following Andrews (1993) and others, the test is

$$\sup F_T(k; q) = F_T(\hat{\lambda}_1, \ldots, \hat{\lambda}_k; q), \tag{3.8}$$

where $(\hat{\lambda}_1, \ldots, \hat{\lambda}_k)$ minimizes the global sum of squared residuals under the specified trimming, which is equivalent to maximizing the F-test assuming spherical errors. This is asymptotically equivalent to, and yet much simpler to construct than, maximizing the F-test (3.6) since the estimated break dates are consistent even in the presence of serial correlation.

Proposition 3.1. *Let $W_q(\cdot)$ be a q-vector of independent Wiener processes on* [0, 1]. *Under A*4 *and A*8 *of BP and* $m = 0$, $\sup F_T(k; q) \Rightarrow \sup F_{k,q} = \sup_{(\lambda_1, \ldots, \lambda_k) \in \Lambda_\epsilon} F(\lambda_1, \ldots, \lambda_k; q)$, *with*

$$F(\lambda_1, \ldots, \lambda_k; q) = \frac{1}{kq} \sum_{i=1}^{k} \frac{[\lambda_i W_q(\lambda_{i+1}) - \lambda_{i+1} W_q(\lambda_i)]'[\lambda_i W_q(\lambda_{i+1}) - \lambda_{i+1} W_q(\lambda_i)]}{\lambda_i \lambda_{i+1}(\lambda_{i+1} - \lambda_i)}. \quad (3.9)$$

Various versions of the tests can be obtained depending on the assumptions made with respect to the distribution of the data and the errors across segments. These relate to different specifications in the construction of the estimate of $V(\hat{\delta})$ given by (3.7). As discussed in the introduction, the simulation results of Bai and Perron (2000) indicate that a large value of ϵ is needed when heterogeneity across segments or correlation in the errors is involved. For that purpose, we obtained critical values for $\epsilon = 0.10$, 0.15, 0.20 and 0.25 in addition to those pertaining to $\epsilon = 0.05$ presented in BP. These were obtained via simulations. The Wiener processes $W_q(\lambda)$ are approximated by the partial sums $n^{-1/2} \sum_{i=1}^{[n\lambda]} e_i$ with $e_i \sim i.i.d.\ N(0, I_q)$ and $n = 1000$. 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 (see Bai and Perron (2003)). The critical values obtained correspond to the 0.90, 0.95, 0.975 and 0.99 quantiles which allow tests with significance levels 0.10, 0.05, 0.025 and 0.01, respectively. The number of regressors q varies from one to 10. Note that when $\epsilon = 0.10$ the maximum number of breaks considered, k, is eight since allowing nine breaks imposes the estimates to be exactly $\hat{\lambda}_1 = 0.1$, $\hat{\lambda}_2 = 0.2$ up to $\hat{\lambda}_9 = 0.9$. For similar reasons, the maximum number of breaks allowed is five when $\epsilon = 0.15$, three when $\epsilon = 0.20$ and two when $\epsilon = 0.25$. When $\epsilon = 0.05$, the maximal value of k is set to 9.

3.2. Double maximum tests

Often, an investigator wishes not to pre-specify a particular number of breaks to make inference. To allow this BP have introduced two tests of the null hypothesis of no structural break against an unknown number of breaks given some upper bound M. These are called the *double maximum tests*. The first is an equally weighted version defined by $UD\max F_T(M, q) = \max_{1 \le m \le M} F_T(\hat{\lambda}_1, \ldots, \hat{\lambda}_m; q)$, where $\hat{\lambda}_j = \hat{T}_j / T$ $(j = 1, \ldots, m)$ are the estimates of the break points obtained using the global minimization of the sum of squared residuals assuming segments of minimal length $h = \epsilon T$. The limiting distribution of this test is given by $\max_{1 \le m \le M} \sup_{(\lambda_1, \ldots, \lambda_m) \in \Lambda_\epsilon} F(\lambda_1, \ldots, \lambda_m; q)$.

The second test applies weights to the individuals tests such that the marginal p-values are equal across values of m and is denoted $WD\max F_T(M, q)$. 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 $F_T(\hat{\lambda}_1, \ldots, \hat{\lambda}_m; q)$ 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

$$WD\max F_T(M, q) = \max_{1 \le m \le M} \frac{c(q, \alpha, 1)}{c(q, \alpha, m)} F_T(\hat{\lambda}_1, \ldots, \hat{\lambda}_m; q) \quad (3.10)$$

$$\Rightarrow \max_{1 \le m \le M} \frac{c(q, \alpha, 1)}{c(q, \alpha, m)} \sup_{(\lambda_1, \ldots, \lambda_m) \in \Lambda_\epsilon} F(\lambda_1, \ldots, \lambda_m; q). \quad (3.11)$$

Critical values were provided for $M = 5$ and $\epsilon = 0.05$ in BP. Additional critical values for $\epsilon = 0.10$ $(M = 5)$, 0.15 $(M = 5)$, 0.20 $(M = 3)$ and 0.25 $(M = 2)$ were simulated (again with q ranging from 1 to 10).

3.3. A test of ℓ vs. $\ell + 1$ breaks

BP proposed a test for ℓ vs. $\ell + 1$ breaks, labelled $\sup F_T(\ell + 1 \mid \ell)$. The test amounts to the application of $(\ell + 1)$ tests of the null hypothesis of no structural change vs. the alternative hypothesis of a single change. It is applied to each segment containing the observations $\hat{T}_{i-1} + 1$ to $\hat{T}_i (i = 1, \ldots, \ell + 1)$ using the convention that $\hat{T}_0 = 0$ and $\hat{T}_{\ell+1} = T$. We conclude for a rejection in favor of a model with $(\ell + 1)$ breaks if the overall maximal value of the $\sup F_T(1; q)$ (over all segments where an additional break is included) is sufficiently large. The break date thus selected is the one associated with this overall maximum. The limiting distribution is stated in the following proposition.

Proposition 3.2. *Under A4 and A8 of BP and* $m = \ell$: $\lim_{T\to\infty} P(F_T(\ell + 1 \mid \ell) \leq x) = G_{q,\epsilon}(x)^{\ell+1}$, *with* $G_{q,\epsilon}(x)$ *the distribution function of* $\sup_{\epsilon\leq\mu\leq1-\epsilon} \|W_q(\mu) - \mu W_q(1)\|^2/(\mu(1-\mu))$.

Note that the estimates $\hat{T}_i$ need not be the global minimizers of the sum of squared residuals; one can also use sequential one at a time estimates (see Bai (1997)). Asymptotic critical values were provided by BP for a trimming $\epsilon = 0.05$ for q ranging from 1 to 10 and ℓ ranging from 0 to 9. Simulated critical values were obtained for $\epsilon = 0.10$, 0.15, 0.20 and 0.25 in addition to the value $\epsilon = 0.05$ presented in BP. Note that, unlike for the $\sup F_T(k; q)$ test, we do not need to impose similar restrictions on the number of breaks for different values of the trimming ϵ.[2]

4. RESPONSE SURFACE REGRESSIONS

In order to present the results more succinctly, we use response surface regressions.[3] We experimented with a variety of specifications and we opted for a class of nonlinear regressions of the form

$$y_i = (\beta_1' x_{1i}) \exp(\beta_2' x_{2i}) + e_i \tag{4.12}$$

where y_i is the simulated critical value. The choice of regressors to include was dictated by overall significance subject to the requirement that the R^2 of the regression be no smaller than 0.999. The list of regressors we retained are

- for the $\sup F_T(k; q)$ test: $x_1 = \{1, q, q^2, k, \epsilon, q/\epsilon\}$, $x_2 = \{1/k, 1/(\epsilon * k)\}$;
- for the *UD* max $F_T(M; q)$ and *WD* max $F_T(M; q)$ tests: $x_1 = \{1, q, \epsilon\}$, $x_2 = \{q, \epsilon\}$;
- for the $\sup F_T(\ell \mid \ell + 1)$ test: $x_1 = \{1, q, q^2, \ell + 1, 1/(\ell + 1), \epsilon\}$, $x_2 = \{q, q^2\}$.

The same specifications apply to all significance levels. The results are presented in Table 1. These permit a quick computation of the approximate critical values as the fitted values of the regression for particular choices of q, k or ℓ, and ϵ. We computed the difference between the fitted values and the actual simulated values and these were found to be very small. Hence, little distortion in size will occur using this set of results to compute critical values.

[2]However, considering more than int$[1/\epsilon] - 2$ breaks implies changing ϵ as one progresses through the sequential procedure. For example, one could use a trimming $\epsilon = 0.05$ and find six breaks in the first half of the sample, then switch to a trimming of $\epsilon = 0.20$ to test for a seventh break.

[3]The full sets of critical values are available in tabulated format at `http:/econ.bu.edu/perron`.

Table 1. Estimated response surface regressions (all statistics multiplied by q).

		Level regressors (x_1)								Exponentiated regressors (x_2)					
Test	Quantile	1	q	q^2	k	$\ell+1$	$1/(\ell+1)$	ϵ	q/ϵ	q	q^2	$1/k$	ϵ	$1/(\epsilon k)$	R^2
$\sup F_T(k;q)*q$	0.90	7.551	1.718	−0.041	−0.610			−15.846	0.025			0.338		−0.014	0.9989
$n=270$	0.95	8.238	1.756	−0.043	−0.659			−15.436	0.025			0.389		−0.013	0.9990
	0.975	8.968	1.788	−0.045	−0.715			−15.255	0.025			0.426		−0.013	0.9991
	0.99	9.879	1.771	−0.042	−0.777			−14.551	0.025			0.471		−0.012	0.9992
UD max $F_T(M;q)*q$	0.90	6.917	2.930					−9.275		−0.028			−0.406		0.9999
$n=50$	0.95	8.228	3.095					−9.644		−0.029			−0.291		0.9999
	0.975	9.436	3.304					−9.301		−0.030			−0.259		0.9999
	0.99	11.211	3.366					−7.279		−0.027			−0.268		0.9999
WD max $F_T(M;q)*q$	0.90	7.316	3.128					−8.624		−0.029			−0.412		0.9999
$n=50$	0.95	9.039	3.318					−9.969		−0.030			−0.327		0.9999
	0.975	10.703	3.465					−11.119		−0.031			−0.250		0.9999
	0.99	13.189	3.346					−11.870		−0.026			−0.173		0.9999
$\sup F_T(\ell\|\ell+1)*q$	0.90	8.397	3.702	−0.209		0.317	−3.736	−11.596		−0.027	0.005				0.9999
$n=500$	0.95	9.879	4.086	−0.232		0.322	−3.687	−11.931		−0.039	0.006				0.9999
	0.975	11.424	4.435	−0.261		0.300	−3.700	−12.292		−0.047	0.006				0.9999
	0.99	13.073	4.954	−0.317		0.285	−3.419	−12.452		−0.058	0.008				0.9998

Note: The functional form of the regression is $y_i = (\beta_1' x_{1i}) \exp(\beta_2' x_{2i}) + e_i$ where y_i is the simulated critical value. The table lists the estimated coefficient associated with the stated regressors. All estimates are highly significant, hence standard errors are not reported. n refers to the number of observations used to estimate each regression.

5. CONCLUSIONS

Bai and Perron (2000) presents an extensive simulation analysis pertaining to the size and power of the tests, among other things. A useful strategy is to first look at the *UD* max or *WD* max tests to see if at least one break is present. If these indicate the presence of at least one break, then the number of breaks can be decided based upon a sequential examination of the sup$F(\ell + 1 \mid \ell)$ statistics constructed using global minimizers for the break dates (i.e. ignore the test $F(1 \mid 0)$ and select m such that the tests sup$F(\ell + 1 \mid \ell)$ are insignificant for $\ell \geq m$). This method leads to the best results and is recommended for empirical applications.

The methods are shown to be adequate, in general, but care must be taken when using particular specifications. With respect to testing, the following recommendations are made. First, ensure that the specifications are such that the size of the test is adequate under the hypothesis of no break. If serial correlation and/or heterogeneity in the data or errors across segments are not allowed in the estimated regression model (and not present in the data generating process), using any value of the trimming ϵ will lead to tests with adequate sizes. However, if such features are allowed, a higher trimming is needed. With a sample of $T = 120$, $\epsilon = 0.15$ should be enough for heterogeneity in the errors or the data. If serial correlation is allowed, $\epsilon = 0.20$ may be needed. These could be reduced if larger sample sizes are available. Hence, the critical values provided here should be useful in practice.

ACKNOWLEDGEMENTS

We thank Zhongjun Qu for research assistance and we acknowledge financial support from the National Science Foundation under Grants SBR9709508 (Bai) and SES-0078492 (Perron). We also wish to thank the editor, Karim Abadir, for useful suggestions.

REFERENCES

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica 61*, 821–56.

Andrews, D. W. K., I. Lee and W. Ploberger (1996). Optimal changepoint tests for normal linear regression. *Journal of Econometrics 70*, 9–38.

Bai, J. (1997). Estimating multiple breaks one at a time. *Econometric Theory 13*, 315–52.

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica 66*, 47–78.

Bai, J. and P. Perron (2000). Multiple structural change models: a simulation analysis, unpublished manuscript, Department of Economics, Boston University.

Bai, J. and P. Perron (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics 18*, 1–22.

Garcia, R. and P. Perron (1996). An analysis of the real interest rate under regime shifts. *Review of Economics and Statistics 78*, 111–25.

Liu, J., S. Wu and J. V. Zidek (1997). On segmented multivariate regressions. *Statistica Sinica 7*, 497–525.