Multiple Structural Change Models: A Simulation Analysis*

Jushan Bai[†] Pierre Perron[‡]

Boston College Boston University

April 3, 2001; Revised October 14, 2004

Abstract

In a recent paper, Bai and Perron (1998) considered theoretical issues related to the limiting distribution of estimators and test statistics in the linear model with multiple structural changes. We assess, via simulations, the adequacy of the various methods suggested. These cover the size and power of tests for structural changes, the coverage rates of the confidence intervals for the break dates and the relative merits of methods to select the number of breaks. The various data generating processes considered allow for general conditions on the data and the errors including differences across segments. Various practical recommendations are made.

JEL Classification Number: C20.

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

---

*Jushan Bai acknowledges financial support from the National Science Foundation under Grant SBR9709508. Address for correspondence: Pierre Perron, Department of Economics, Boston University, 270 Bay State Road, Boston, 02215, USA (e-mail: perron@bu.edu).

[†]Department of Economics, Boston College, Chestnut Hill, MA, 02467 (jushan.bai@bc.edu).

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

1 Introduction.

Both the statistics and econometrics literature contain a vast amount of work on issues related to structural change, most of it specifically designed for the case of a single change. However, the problem of multiple structural changes has received considerably less attention. Recently, Bai and Perron (1998, 2003a) provided a comprehensive treatment of various issues in the context of multiple structural change models: consistency of estimates of the break dates, tests for structural changes, confidence intervals for the break dates, methods to select the number of breaks and efficient algorithms to compute the estimates. However, their results are solely asymptotic in nature and the adequacy in finite samples remains to be investigated. In this paper, we intend to partially fill this gap.

We present simulation results pertaining to the behavior of the estimators and tests in finite samples. We consider the problem of forming confidence intervals for the break dates under various hypotheses about the structure of the data and errors across segments. In particular, we may allow the data and errors to have different distributions across segments or impose a common structure. The issue of testing for structural changes is also considered under very general conditions on the data and the errors and the properties of tests, both in the data-generating processes and in the specification of the tests. We also address the issue of estimating the number of breaks. To that effect, we discuss methods based on information criteria and a method based on a sequential testing procedure as suggested in Bai and Perron (1998).

The rest of this paper is structured as follows. Section 2 presents the model and the estimator. Section 3 summarizes the relevant asymptotic results about the construction of confidence intervals for the break dates, the tests for multiple structural changes and methods to estimate the number of breaks. It describes the exact nature of the various tests and procedures upon various specifications about the nature of the errors and data across segments. Section 4 presents the results of simulations analyzing the adequacy of the asymptotic approximations in finite samples, the size and power of the various tests and the relative merits of several methods to estimate the number of structural changes. Some concluding remarks and practical recommendations are contained in Section 5.

2 The Model and Estimators.

For the purpose of the simulation study, we consider the following multiple linear regression with m breaks ($m+1$ regimes):

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

for $j = 1, ..., m+1$. This is a special case of the general model considered in Bai and Perron (1998) corresponding to a pure structural change model. Here, y_t is the observed dependent variable at time t; z_t ($q \times 1$) is a vectors of covariates and δ_j ($j = 1, ..., m+1$) is the corresponding vector of coefficients; u_t is the disturbance at time t. The indices $(T_1, ..., T_m)$, or the break points, are explicitly treated as unknown (we use the convention that $T_0 = 0$ and $T_{m+1} = T$). The purpose is to estimate the unknown regression coefficients together with the break points when T observations on (y_t, z_t) are available.

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

$$S_T(T_1, ..., T_m) = \sum_{i=1}^{m+1} \sum_{t=T_{i-1}+1}^{T_i} [y_t - z_t'\delta_i]^2.$$

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

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

where the minimization is taken over all partitions $(T_1, ..., T_m)$ such that $T_i - T_{i-1} \geq h \geq q$. Thus the break-point estimators are global minimizers of the objective function. Finally, the regression parameter estimates are obtained using the associated least-squares estimates at the estimated m-partition $\{\hat{T}_j\}$, i.e. $\hat{\delta} = \hat{\delta}(\{\hat{T}_j\})$. An efficient algorithm, based on the principle of dynamic programming, to obtain global minimizers of the sum of squared residuals is presented in Bai and Perron (2003a).

Note that, in general, h need not be set to q. Indeed, in many instances the choice of the trimming is made independently of the number of regressors. This is the case, in particular when obtaining estimates for the purpose of constructing test statistics (see Section 3.2 below).

A central result derived in Bai and Perron (1998) concerns the convergence of the break fractions $\hat{\lambda}_i = \hat{T}_i/T$ and the rate of convergence. The results obtained show not only that $\hat{\lambda}_i$ converges to its true value λ_i^0 but that it does so at the fast rate T, i.e. $T(\hat{\lambda}_i - \lambda_i^0) = O_p(1)$ for all i. This convergence result is obtained under a very general set of assumptions allowing a wide variety of models. It, however, precludes integrated variables (with an autoregressive unit root) but permits trending regressors; for example with a trend of the form $g_t = a + b(t/T)$. The assumptions concerning the nature of the errors in relation to the regressors $\{z_t\}$ are of two kinds. First, when no lagged dependent variable is allowed in $\{z_t\}$, the conditions on the residuals are quite general and allow substantial correlation and heteroskedasticity. The second case allows lagged dependent variables as regressors but then, of course, no serial correlation is permitted in the errors $\{u_t\}$. In both cases, the assumptions are general enough to allow different distributions for both the regressors and the errors across segments.

3 Summary of Relevant Asymptotic Results.

3.1 Constructing Confidence Intervals.

To get an asymptotic distribution for the break dates, the strategy considered is to adopt an asymptotic framework where the magnitudes of the shifts converge to zero as the sample size increases. The resulting limiting distribution is then independent of the specific distribution of the pair $\{z_t, u_t\}$. To describe the relevant distributional result, we need to define some notations. For $i = 1, ..., m$, and $\Delta T_i^0 = T_i^0 - T_{i-1}^0$, let

$$\begin{aligned} \Delta_i &= \delta_{i+1}^0 - \delta_i^0, \\ Q_i &= \lim (\Delta T_i^0)^{-1} \sum_{t=T_{i-1}^0+1}^{T_i^0} E(z_t z_t'), \\ \Omega_i &= \lim (\Delta T_i^0)^{-1} \sum_{r=T_{i-1}^0+1}^{T_i^0} \sum_{t=T_{i-1}^0+1}^{T_i^0} E(z_r z_t' u_r u_t). \end{aligned}$$

In the case where the data are non-trending, we have, under various assumptions[1] stated in

[1] The important ones are as follows: the magnitude of the shifts decreases at a suitable rate as the sample size increases, a functional central limit theorem holds for the partial sums of the variables $\{z_t u_t\}$, also $p\lim (\Delta T_i^0)^{-1} \sum_{t=T_{i-1}^0+1}^{T_{i-1}^0+[s\Delta T_i^0]} E(z_t z_t') = sQ_i$ is assumed to exist with Q_i a fixed matrix. The latter precludes trending regressors.

Bai and Perron (1998), the following limiting distribution of the break dates:

$$\frac{(\Delta_i' Q_i \Delta_i)^2}{(\Delta_i' \Omega_i \Delta_i)} (\hat{T}_i - T_i^0) \Rightarrow \arg\max{}_s V^{(i)}(s), \qquad (i = 1, ..., m), \tag{3}$$

where

$$V^{(i)}(s) = \begin{cases} W_1^{(i)}(-s) - |s|/2, & \text{if } s \leq 0, \\ \sqrt{\xi_i}(\phi_{i,2}/\phi_{i,1}) W_2^{(i)}(s) - \xi_i |s|/2, & \text{if } s > 0, \end{cases} \tag{4}$$

and

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

Also, $W_1^{(i)}(s)$ and $W_2^{(i)}(s)$ are independent standard Weiner processes defined on $[0, \infty)$, starting at the origin when $s = 0$. These processes are also independent across i.

The cumulative distribution function of $\text{argmax}_s V^{(i)}(s)$ is derived in Bai (1997a) and all that is needed to compute the relevant critical values are estimates of Δ_i, Q_i, and Ω_i. These are given by

$$\begin{aligned} \hat{\Delta}_i &= \hat{\delta}_{i+1} - \hat{\delta}_i, \\ \hat{Q}_i &= (\Delta \hat{T}_i)^{-1} \sum_{t=\hat{T}_{i-1}+1}^{\hat{T}_i} z_t z_t', \end{aligned}$$

and an estimate of Ω_i can be constructed using the covariance matrix estimator of Andrews (1991) applied to the vector $\{z_t \hat{u}_t\}$ and using data over segment i only. We use the Quadratic Spectral kernel with an AR(1) approximation for each element of the vector $\{z_t \hat{u}_t\}$ to construct the optimal bandwidth (henceforth referred to as a HAC estimator).

In practice, one may want to impose some constraints on this general framework related to the distribution of the errors and regressors across segments. For ease of reference, especially with the simulation results presented later, we shall adopt the following notation. We denote by $cor\_u = 1$ the case where the errors are allowed to be correlated and by $cor\_u = 0$ the case where no correction for serial correlation is made. Similarly, $het\_z = 1$ denotes the case where the regressors are allowed to have heterogenous distributions across segments and by $het\_z = 0$ the case where the distributions are assumed to be homogenous across segments. Finally, $het\_u = 1$ permits heterogenous variances of the residuals across segments and $het\_u = 0$ imposes the same variance throughout. We have the following cases when adding restrictions:

- The regressors z_t are identically distributed across segments ($cor\_u = 1$, $het\_z = 0$, $het\_u = 1$). Then $Q_i = Q_{i+1} = Q$ which can consistently be estimated by $\hat{Q} = T^{-1}\sum_{t=1}^{T} z_t z_t'$. In this case, the limiting result states that

$$\frac{(\hat{\Delta}_i'\hat{Q}\hat{\Delta}_i)^2}{(\hat{\Delta}_i'\hat{\Omega}_i\hat{\Delta}_i)}(\hat{T}_i - T_i^0) \Rightarrow \arg\max{}_s V^{(i)}(s),$$

with $\xi_i = 1$.

- The errors are identically distributed across segments ($cor\_u = 1$, $het\_z = 1$, $het\_u = 0$). Then $\Omega_i = \Omega_{i+1} = \Omega$ which can consistently be estimated using a HAC estimator applied to the variable $\{z_t \hat{u}_t\}$ using data over the whole sample.

- The errors and the data are identically distributed across segments ($cor\_u = 1$, $het\_z = 0$, $het\_u = 0$). Here, we have $\xi_i = 1$, and $\phi_{i,1} = \phi_{i,2}$ and the limiting distribution reduces to

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

which has a density function symmetric about the origin.

- The errors are serially uncorrelated ($cor\_u = 0$, $het\_z = 1$, $het\_u = 1$). In this case $\Omega_i = \sigma_i^2 Q_i$ and $\phi_{i,1}^2 = \phi_{i,2}^2 = \sigma_i^2$ which can be estimated using $\hat{\sigma}_i^2 = (\Delta\hat{T}_i)^{-1}\sum_{t=\hat{T}_{i-1}+1}^{\hat{T}_i} \hat{u}_t^2$. The confidence intervals can then be constructed from the approximation

$$\frac{(\hat{\Delta}_i'\hat{Q}_i\hat{\Delta}_i)}{\hat{\sigma}_i^2}(\hat{T}_i - T_i^0) \Rightarrow \arg\max{}_s V^{(i)}(s). \tag{5}$$

- The errors are serially uncorrelated and the regressors are identically distributed across segments ($cor\_u = 0$, $het\_z = 0$, $het\_u = 1$). Here $\phi_{i,1}^2 = \phi_{i,2}^2 = \sigma_i^2$ and $\xi_i = 1$. The confidence intervals can then be constructed from the approximation

$$\frac{(\hat{\Delta}_i'\hat{Q}\hat{\Delta}_i)}{\hat{\sigma}_i^2}(\hat{T}_i - T_i^0) \Rightarrow \arg\max_s\{W^{(i)}(s) - |s|/2\}. \tag{6}$$

- The errors are serially uncorrelated and identically distributed across segments ($cor\_u = 0$, $het\_z = 1$, $het\_u = 0$). The approximation is the same as (5) with $\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T}\hat{u}_t^2$ instead of $\hat{\sigma}_i^2$.

- The errors are serially uncorrelated and both the data and the errors are identically distributed across segments ($cor\_u = 0$, $het\_z = 0$, $het\_u = 0$). The approximation is the same as (6) with $\hat{\sigma}^2$ instead of $\hat{\sigma}_i^2$.

Since the break dates are integer valued, we consider confidence intervals that are likewise integer-valued by using the highest smaller integer for the lower bound and the smallest higher integer for the upper bound.

3.2 Test Statistics for Multiple Breaks.

3.2.1 A Test of no break versus a fixed number of breaks.

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

$$F_T^*(\lambda_1, ..., \lambda_k; q) = \frac{1}{T}\left(\frac{T-(k+1)q}{kq}\right)\hat{\delta}' R'(R\hat{V}(\hat{\delta})R')^{-1}R\hat{\delta}, \tag{7}$$

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

$$V(\hat{\delta}) = p\lim T(\overline{Z}'\overline{Z})^{-1}\overline{Z}'\Omega\overline{Z}(\overline{Z}'\overline{Z})^{-1}. \tag{8}$$

The statistic F_T^* is simply the conventional F-statistic for testing $\delta_1 = \cdots = \delta_{k+1}$ against $\delta_i \neq \delta_{i+1}$ for some i given the partition $(T_1, ..., T_k)$. The *supF* type test statistic is then defined as

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

where

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

for some arbitrary positive number ϵ. In this general case, allowing for serial correlation in the errors, the $\text{supF}_T^*(k;q)$ may be rather cumbersome to compute. However, one can obtain a much simpler, yet asymptotically equivalent, version by using the estimates of the break dates obtained from the global minimization of the sum of squared residuals. Denote, these estimates by $\hat{\lambda}_i = \hat{T}_i/T$ for $i = 1, ..., k$, the test is then

$$\text{supF}_T(k; q) = F_T^*(\hat{\lambda}_1, ..., \hat{\lambda}_k; q)$$

The estimates $\hat{\lambda}_1, ..., \hat{\lambda}_k$ are equivalently the arguments that maximizes the following F-statistic:

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

and

$$\widetilde{V}(\hat{\delta}) = (\frac{\overline{Z}'\overline{Z}}{T})^{-1},$$

the covariance matrix of $\hat{\delta}$ assuming spherical errors. This procedure is asymptotically equivalent since the break dates are consistent even in the presence of serial correlation. The asymptotic distribution still depends on the specification of the set Λ_ϵ via the imposition of the minimal length h of a segment. Hence, $\epsilon = h/T$.

Various versions of the tests can be obtained depending on the assumptions made with respect to the distribution of the data and the errors across segments. These variations relates to different specifications in the construction of the estimate of the limiting covariance matrix $V(\hat{\delta})$ given by (8). They are the following.

- No serial correlation, different distributions for the data and identical distribution for the errors across segments (*cor_u* $= 0$, *het_z* $= 1$, *het_u* $= 0$). In this base case, the estimate is

$$\hat{V}(\hat{\delta}) = \hat{\sigma}^2 \left(\frac{\overline{Z}'\overline{Z}}{T}\right)^{-1}.$$

- No serial correlation in the errors, different variances of the errors and different distributions of the data across segments (*cor_u* $= 0$, *het_z* $= 1$, *het_u* $= 1$). In this case,

$$\hat{V}(\hat{\delta}) = diag(\hat{V}(\hat{\delta}_1), ..., \hat{V}(\hat{\delta}_{m+1})),$$

  where $\hat{V}(\hat{\delta}_i)$ is the covariance matrix of the estimate $\hat{\delta}_i$ using only data from segment i, i.e. $\hat{V}(\hat{\delta}_i) = \hat{\sigma}_i^2[(\Delta\hat{T}_i)^{-1}\sum_{t=\hat{T}_{i-1}+1}^{\hat{T}_i} z_t z_t']^{-1}$ with $\hat{\sigma}_i^2 = (\Delta\hat{T}_i)^{-1}\sum_{t=\hat{T}_{i-1}+1}^{\hat{T}_i} \hat{u}_t^2$. These are simply the OLS estimates obtained using data from each segment separately.

- Serial correlation in the errors, different distributions for the data and the errors across segments (*cor_u* $= 1$, *het_z* $= 1$, *het_u* $= 1$). Here, we make use of the fact that the errors in different segments are asymptotically independent. Hence, the limiting variance is given by

$$V(\hat{\delta}) = diag(V(\hat{\delta}_1), ..., V(\hat{\delta}_{m+1})),$$

where, for $i = 1, ..., m+1$,

$$V(\hat{\delta}_i) = p\lim(\Delta T_i)(Z_i'Z_i)^{-1}Z_i'\Omega_i Z_i(Z_i'Z_i)^{-1}.$$

This can be consistently estimated, segment by segment, with a HAC estimator of $V(\hat{\delta}_i)$ using only data from segment i.

- Serial correlation in the errors, same distribution for the errors across segments ($cor\_u = 1$, $het\_z = 1$, $het\_u = 0$). In this case the limiting covariance matrix is

$$V(\hat{\delta}) = p\lim T(\overline{Z}'\overline{Z})^{-1}(\Lambda \otimes (Z'\Omega Z))(\overline{Z}'\overline{Z})^{-1},$$

where (using the convention that $\lambda_0 = 0$ and $\lambda_{m+1} = 1$)

$$\Lambda = \begin{pmatrix} \lambda_1 - \lambda_0 & & & \\ & \lambda_2 - \lambda_1 & & \\ & & \ddots & \\ & & & \lambda_{m+1} - \lambda_m \end{pmatrix}.$$

This can be consistently estimated using $\hat{\lambda}_i = \hat{T}_i/T$ and a HAC estimator based on the pair $\{z_t\hat{u}_t\}$ constructed using the full sample. Note that we have an implicit assumption that the regressors z_t have the same distribution across segments since the consistent estimate of $p\lim Z'\Omega Z/T$ is constructed using the full sample. For reasons, discussed below we do not impose that restriction when evaluating $p\lim \overline{Z}'\overline{Z}/T$. That is, we still use $\overline{Z}'\overline{Z}/T$ instead of an estimate of $(\Delta \otimes Q)$ obtained using $\hat{Q} = T^{-1}\sum_{t=1}^{T} z_t z_t'$ based on the full sample.

In the construction of the tests we do not consider imposing the restriction that the distribution of the regressors z_t be the same across segments even if they are (except as they enter in the construction of a HAC estimate involving the pair $\{z_t\hat{u}_t\}$). This might at first sight seem surprising since imposing a valid restriction should lead to more precise estimate. This is, however, not true. Consider the case with no serial correlation in the errors and the same distribution for the errors across segments ($cor\_u = 0$, $het\_u = 0$). Imposing the restriction $het\_z = 0$, leads to the following asymptotic covariance matrix

$$V(\hat{\delta}) = \sigma^2(\Lambda \otimes Q)^{-1},$$

where $Q = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E(z_t z_t')$. Note that a consistent estimate can be obtained using $\hat{Q} = T^{-1}\sum_{t=1}^{T} z_t z_t'$, $\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T} \hat{u}_t^2$ and $\hat{\Lambda}$ constructed using $\hat{\lambda}_i = \hat{T}_i/T$ $(i = 1, ..., m)$. Suppose that the z's are exogenous and the errors have the same variance across segments. Then, for a given partition $(T_1, ..., T_m)$, the exact variance of the estimated coefficients $\hat{\delta}$ is

$$V(\hat{\delta}) = \sigma^2 \left(\frac{\overline{Z}'\overline{Z}}{T}\right)^{-1}.$$

Using the asymptotic version $V(\hat{\delta}) = \sigma^2(\Lambda \otimes Q)^{-1}$ may imply an inaccurate approximation to the exact distribution. This would occur especially if small segments are allowed in which case the exact moment matrix of the regressors may deviate substantially from its full sample analog.

The same problem occurs in the case with no serial correlation in the errors and different variance for the residuals across segments ($cor\_u = 0$, $het\_u = 1$). Imposing $het\_z = 0$ gives the limiting variance

$$V(\hat{\delta}) = (\Lambda^* \otimes Q)^{-1}$$

where

$$\Lambda^* = \begin{pmatrix} \sigma_1^2(\lambda_1 - \lambda_0) & & & \\ & \sigma_2^2(\lambda_2 - \lambda_1) & & \\ & & \ddots & \\ & & & \sigma_{m+1}^2(\lambda_{m+1} - \lambda_m) \end{pmatrix},$$

which can be consistently estimated using $\hat{Q}$, $\hat{\lambda}_i = \hat{T}_i/T$ and $\hat{\sigma}_i^2 = (\Delta\hat{T}_i)^{-1}\sum_{t=\hat{T}_{i-1}+1}^{\hat{T}_i} \hat{u}_t^2$. Again, in finite samples, imposing the constraint that $Z_i'Z_i/(\Delta\hat{T}_i)$ be approximated by $\hat{Q}$ over all segments may imply a poor approximation in finite samples. We have found, in these two cases, that imposing a common distribution for the regressors across segments leads to tests with worse properties even when the data indeed have an invariant distribution. These distortions becomes less important, however, when the sample size is large and/or the trimming ϵ is large.

The relevant asymptotic distribution has been derived in Bai and Perron (1998) and critical values can be found in Bai and Perron (1998) for a trimming $\epsilon = .05$ and values of k from 1 to 9 and values of q from 1 to 10. As the simulation experiments will show, a trimming as small as 5% of the total sample can lead to tests with substantial size distortions when allowing different variances of the errors across segments or when serial correlation is permitted. This is because one is then trying to estimate various quantities using very few observations; for example, if $T = 100$ and $\epsilon = .05$, one ends up estimating, for some segments, quantities like the variance of the residuals using only 5 observations. Similarly, with serial correlation a HAC estimator would need to be applied to very short samples. The estimates are then highly imprecise and the tests accordingly show size distortions. When allowing different variances across segments or serial correlation a higher value of ϵ should be used.

Hence, the case ($cor\_u = 0$, $het\_z = 1$, $het\_u = 0$) should be considered the base case in which the tests can be constructed using an arbitrary small trimming ϵ. For all other cases, care should be exercised in the choice of ϵ and larger values should be considered. Critical values for trimming parameter $\epsilon = .10$, .15, .20 and .25 can be found in Bai and Perron (2003b). Note that when $\epsilon = .10$ the maximum number of break considered is 8 since allowing 9 breaks impose the estimates to be exactly $\hat{\lambda}_1 = .1$, $\hat{\lambda}_2 = .2$ up to $\hat{\lambda}_9 = .9$. For similar reasons, the maximum number of breaks allowed is 5 when $\epsilon = .15$, 3 when $\epsilon = .20$ and 2 when $\epsilon = .25$.

3.2.2 Double maximum tests.

Often, an investigator wishes not to pre-specify a particular number of breaks to make inference. To allow this Bai and Perron (1998) have introduced two tests of the null hypothesis of no structural break against an unknown number of breaks given some upper bound M. These are called the *double maximum tests.* The first is an equal weighted version defined by

$$UD\max F_T^*(M,q) = \max_{1\leq m\leq M} \sup_{(\lambda_1,\ldots,\lambda_m)\in\Lambda_\epsilon} F_T^*(\lambda_1,...,\lambda_m;q).$$

We use the asymptotically equivalent version

$$UD\max F_T(M,q) = \max_{1\leq m\leq M} F_T(\hat{\lambda}_1,...,\hat{\lambda}_m;q),$$

where $\hat{\lambda}_j = \hat{T}_j/T$ $(j = 1,..,m)$ are the estimates of the break points obtained using the global minimization of the sum of squared residuals.

The second test applies weights to the individuals tests 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 $\sup_{(\lambda_1,\ldots,\lambda_m)\in\Lambda_\epsilon} F_T(\lambda_1,...,\lambda_m;q)$ 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

$$WD\max F_T^*(M,q) = \max_{1\leq m\leq M} \frac{c(q,\alpha,1)}{c(q,\alpha,m)} \sup_{(\lambda_1,\ldots,\lambda_m)\in\Lambda_\epsilon} F_T^*(\lambda_1,...,\lambda_m;q). \tag{9}$$

Again, we use the asymptotically equivalent version

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

Note that, unlike the $UD\max F_T(M,q)$ test, the value of the $WD\max F_T(M,q)$ depends on the significance level chosen since the weights themselves depend on α. Critical values can be found in Bai and Perron (1998, 2003b) for $\epsilon = .05$ $(M = 5)$, $\epsilon = .10$ $(M = 5)$, .15 $(M = 5)$, .20 $(M = 3)$ and .25 $(M = 2)$.

3.2.3 A test of ℓ versus $\ell + 1$ breaks.

Bai and Perron (1998) proposed a test for ℓ versus $\ell + 1$ breaks. This test is labelled $\sup F_T(\ell+1|\ell)$. The method amounts to the application of $(\ell+1)$ tests of the null hypothesis of no structural change versus the alternative hypothesis of a single change. The test is applied to each segment containing the observations $\hat{T}_{i-1}$ to $\hat{T}_i$ $(i = 1,...,\ell+1)$. The estimates $\hat{T}_i$ need not be the global minimizers of the sum of squared residuals, all that is required is that the break fractions $\hat{\lambda}_i = \hat{T}_i/T$ converge to their true value at rate T. We conclude for a rejection in favor of a model with $(\ell + 1)$ breaks if the overall minimal value of the sum of squared residuals (over all segments where an additional break is included) is sufficiently smaller than the sum of squared residuals from the ℓ breaks model. The break date thus selected is the one associated with this overall minimum.

Asymptotic critical values were provided by Bai and Perron (1998, 2003b) for q ranging from 1 to 10, and for trimming values ε of .05, .10, .15, .20 and .25. Of course, all the same options are available as for the previous tests concerning the potential specifications of the nature of the distributions for the errors and the data across segments.

3.3 Estimating the number of breaks.

A common procedure to select the dimension of a model is to consider an information criterion. Yao (1988) suggests the use of the Bayesian Information Criterion (BIC) defined as

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

where $p^* = (m+1)q + m + p$, and $\hat{\sigma}^2(m) = T^{-1}S_T(\hat{T}_1,...,\hat{T}_m)$. He showed that the number of breaks can be consistently estimated (at least for normal sequence of random variables with shifts in mean). An alternative proposed by Liu, Wu and Zidek (1997) is a modified

Schwarz' criterion that takes the form:

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

They suggest using $\delta_0 = 0.1$ and $c_0 = 0.299$. Perron (1997) presented a simulation study of the behavior of the these two information criteria and of the AIC in the context of estimating the number of changes in the trend function of a series in the presence of serial correlation. The results first showed the AIC to perform very badly and, hence, this criterion will not be considered any further. The BIC and LWZ perform reasonably well when no serial correlation in the errors is present but imply choosing a number of breaks much higher than the true value when serial correlation is present. When no serial correlation is present in the errors but a lagged dependent variable is present, the BIC performs badly when the coefficient on the lagged dependent variable is large (and more so as it approaches unity). In such cases, the LWZ performs better under the null of no break but underestimate the number of breaks when some are present.

The method suggested by Bai and Perron (1998) is based on the sequential application of the sup $F_T(\ell+1|\ell)$ test. The procedure to estimate the number of breaks is the following. Start by estimating a model with a small number of breaks that are thought to be necessary (or start with no break). Then perform parameter-constancy tests for each subsamples (those obtained by cutting off at the estimated breaks), adding a break to a subsample associated with a rejection with the test sup $F_T(\ell+1|\ell)$. This process is repeated increasing ℓ sequentially until the test sup $F_T(\ell+1|\ell)$ fails to reject the null hypothesis of no additional structural changes. The limiting distribution of the test is the same when using global minimizers for the estimates of the break dates or sequential one-at-a-time estimates since both imply break fractions that converge at rate T (see Bai (1997b)). The final number of breaks is thus equal to the number of rejections obtained with the parameter constancy tests plus the number of breaks used in the initial round.

A distinct advantage of model selection procedures based on hypothesis testing is that, unlike information criteria, they can directly take into account the possible presence of serial correlation in the errors and non-homogeneous variances across segments.

4 Simulation Experiments.

In this section, we present the results of simulation experiments to analyze the size and power of the tests, the coverage rates of the confidence intervals for the break dates and the adequacy of the various methods to select the number of structural changes. A wide variety of data generating processes are considered allowing different variances for the residuals and different distributions for the regressors across segments as well as serial correlation. All computations are performed in GAUSS using a computer program that is available on request for non-profit academic use (see Bai and Perron (2003a) for a thorough description of the features of this program).

4.1 The case with no break.

We start with the case where the data generating processes exhibit no structural change and, hence, analyze the size of the tests and how well the methods to select the number of break points actually select none. Throughout $\{e_t\}$ denotes a sequence of *i.i.d.* $N(0,1)$ random variables, $\{\Psi_t\}$ is a sequence of *i.i.d.* $N(1,1)$ random variables uncorrelated with $\{e_t\}$. We use sample sizes of $T = 120$ and $T = 240$. The values of the trimming ϵ and the maximum number of breaks (M) considered are: $\epsilon = .05$ and $M = 5$, $\epsilon = .10$ and $M = 5$, $\epsilon = .15$ and $M = 5$, $\epsilon = .20$ and $M = 3$, $\epsilon = .25$ and $M = 2$. In all cases, 2,000 replications are used.

The data generating processes and the corresponding regressors used are:

- DGP-1: $y_t = e_t$ and $z_t = \{1\}$ $(q = 1)$;
- DGP-2: $y_t = \Psi_t + e_t$ and $z_t = \{1, \Psi_t\}$ $(q = 2)$;
- DGP-3: $y_t = 0.5y_{t-1} + e_t$ and $z_t = \{1, y_{t-1}\}$ $(q = 2)$.
- DGP-4: $y_t = v_t$ with $v_t = 0.5v_{t-1} + e_t$ and $z_t = \{1\}$ $(q = 1)$;
- DGP-5: $y_t = v_t$ with $v_t = e_t + 0.5e_{t-1}$ and $z_t = \{1\}$ $(q = 1)$;
- DGP-6: $y_t = v_t$ with $v_t = e_t - 0.3e_{t-1}$ and $z_t = \{1\}$ $(q = 1)$;

The DGP-1 with *i.i.d.* data is a base case to assess the basic properties of the tests and methods to select the number of breaks. It is useful to assess the effect of allowing different variances of the errors across segments and/or serial correlation when these features are not present. The DGP-2 is a variation which includes an exogenous regressor. DGP-3 is one where serial correlation is taken into account parametrically. DGPs 4 to 6 are used to assess the effect of serial correlation in the errors and how well the corrections for its presence leads to tests with adequate sizes.

The results are presented in Table 1. Consider first, the base case represented by DGP-1 where the series is white noise. With the specification $cor\_u = 0$ and $het\_u = 0$ all tests have the right size for any value of the trimming ε. As expected, the sequential procedure chooses no break around 95% of the time. The BIC between 94% and 98 % (depending on ε) and the LWZ 100% of the time. When different variances of the residuals are allowed across segments, we see substantial size distortions when the trimming ϵ is small. These, however, disappear when ε reaches .15 or .20. The sequential procedure is somewhat biased when $\varepsilon = .05$ but this bias disappears quickly as soon as ε reaches .10. Similar size distortions occur when allowing serial correlation in the errors ($cor\_u = 1$). These are somewhat more severe if, in addition, different variances are allowed. When $het\_u = 0$, the sequential procedure shows no size distortion at any values of ε. However, if $het\_u = 1$, the sequential procedure is adequate only if ε is at least .15.

A similar picture emerges for DGP-2 where a random regressors is included. If $cor\_u = het\_u = 0$, all tests have the right size. However, allowing for either different variances and/or serial correlation in the residuals induces substantial size distortions unless ε is large. When no serial correlation is allowed, the procedures have the right size if ε is at least .15; when serial correlation is allowed a larger value is needed.

The results for DGP-3, which is an AR(1), shows that if one is testing against a large number of breaks (or using the $WD\max$ test) there are some distortions even if $cor\_u = het\_u = 0$ when ε is small. The sequential procedure remains, however, adequate for any values of ε. If different variances are allowed size distortions occur unless ε is at least .20.

The DGPs 4 to 6 are cases where serial correlation is present in the residuals. As expected, if $cor\_u = 0$, all procedures show substantial size distortions (with positive correlation the tests are liberal and with negative correlation they are conservative). It is therefore important to correct for serial correlation. This, however, can be done adequately only if a large trimming is used, .15 or .20 depending on the cases. An interesting feature, however, is that the sequential procedure works very well for any values of ε when the variances are constrained to be the same ($het\_u = 0$). In particular, it performs much better than the information criterion BIC (and also LWZ in the case of positive AR errors).

In summary, if no serial correlation is present and allowed for, all procedures work well for any values of the trimming ε when the specification $cor\_u = het\_u = 0$ is used. If serial correlation is present a larger value of the trimming is needed when constructing the tests using the specification $cor\_u = 1$. This is also the case if different variances are allowed across segments. Also, the results show the sequential procedure to perform quite well for any values of the trimming provided one is correcting for serial correlation when needed and not correcting for it when it is not needed.

4.2 The case with one break.

The basic data generating process considered is (Case 1):

$$\begin{aligned} y_t &= \mu_1 + \gamma_1 \Psi_t + e_t, \qquad \text{if } t \leq [0.5T], \\ y_t &= \mu_2 + \gamma_2 \Psi_t + e_t, \qquad \text{if } t > [0.5T], \end{aligned}$$

where $\Psi_t \sim i.i.d\ N(1,1)$ and $e_t \sim i.i.d\ N(0,1)$ and both are uncorrelated. Since, no serial correlation is present in the errors and no change in the distribution of the data or the errors is allowed, we use the specification $cor\_u = het\_u = 0$ and $\varepsilon = .05$. For the tests, we use $het\_z = 1$ and to construct the confidence intervals on the break dates, we use $het\_z = 0$. We consider three types of shifts: a) a change in intercept only ($\gamma_1 = \gamma_2 = 1$), b) a change in slope only ($\mu_1 = \mu_2 = 0$), and c) a simultaneous change in slope and intercept.

We also consider a variation without the regressor Ψ_t^* with errors that are serially correlated:

- Case 2: $\gamma_1 = \gamma_2 = 0$, and e_t replaced by $v_t = 0.5v_{t-1} + e_t$. Here $z_t = \{1\}$.

In this second case, we use the specifications $cor\_u = 1$, $het\_u = 0$ and $\varepsilon = .20$. Again, for the tests, we use $het\_z = 1$ and to construct the confidence intervals on the break dates, we use $het\_z = 0$. The experiments are performed for $T = 120$ and $T = 240$ and again 2,000 replications are used.

The results are presented in Table 2. Row (a) presents a case with a small change in intercept only. Here the power of the test is rather low and the coverage rate of the break date is imprecise. We shall use this base case to investigate what increases power. There are, nevertheless, some features of interest. First, the power of the $\sup F(k)$ test is decreasing as k increases (more so as k reaches 5; not shown). However, both $D\max$ tests have power as high as the case with $k = 1$ (which gives the highest power). Also, of the three methods to select the number of breaks, the sequential methods works best. The criterion LWZ is quite inaccurate since it chooses no break 98% of the times. Row (b) considers the same specifications but doubling the sample size to 240. The power of the tests increases, the sequential method selects 1 break more often and the coverage rate is better but not to a great extent. For comparisons, row (c) keeps $T = 120$ but doubles the size of the shift in intercept. Here power increases a lot, the sequential procedure chooses $m = 1$ 95% of the time and the exact coverage rate is close to the nominal 95%. Hence, we can conclude that what is important is not the size of the sample but the size of the break.

Row (d) presents the case of a mild change in slope. Again, the power of the $supF(k)$ decreases as k increases but the $D\max$ tests have as high power as the $\sup F(1)$ test. Also, the sequential procedure is best to select the correct value $m = 1$ while the LWZ is very inaccurate. Row (e) considers merging the small shifts in intercept and slope. We see that the simultaneous occurrence of two shifts at the same dates increases considerably the power of the tests and the precision of the selected number of breaks, as well as the coverage rate of the break date (much more than an increase in sample size). Rows (e) and (f) consider a larger change in slope only and larger simultaneous changes, respectively. Here, the power of the tests is one. In such cases, the coverage rates are accurate and all methods select the correct number of breaks accurately.

Rows (h) to (k) consider case 2 of a change in mean with serially correlated errors. We see that the presence of serial correlation decreases the power of the test substantially. Here, for a given shift, doubling the sample size induces a negligible increase in power and in the accuracy of the selection methods or coverage rates. Nevertheless, the coverage rates are quite accurate which shows that the non-parametric correction for the presence of serial correlation seems to be effective.

4.3 The case with two breaks.

For Case 1, the basic structure is similar except that now the data generating process is:

$$
\begin{aligned}
y_t &= \mu_1 + \gamma_1 \Psi_t^* + e_t^*, && \text{if } 1 < t \leq [T/3], \\
y_t &= \mu_2 + \gamma_2 \Psi_t^* + e_t^*, && \text{if } [T/3] < t \leq [2T/3], \\
y_t &= \mu_3 + \gamma_3 \Psi_t^* + e_t^*, && \text{if } [2T/3] < t < T,
\end{aligned}
$$

where

$$
\begin{aligned}
\Psi_t^* &\sim i.i.d\ N(\varsigma_1, 1), && \text{if } 1 < t \leq [T/3], \\
\Psi_t^* &\sim i.i.d\ N(\varsigma_2, 1), && \text{if } [T/3] < t \leq [2T/3], \\
\Psi_t^* &\sim i.i.d\ N(\varsigma_3, 1), && \text{if } [2T/3] < t \leq T,
\end{aligned}
$$

and

$$
\begin{aligned}
e_t^* &\sim i.i.d\ N(0, \sigma_1^2), && \text{if } 1 < t \leq [T/3], \\
e_t^* &\sim i.i.d\ N(0, \sigma_2^2), && \text{if } [T/3] < t \leq [2T/3], \\
e_t^* &\sim i.i.d\ N(0, \sigma_3^2), && \text{if } [2T/3] < t \leq T.
\end{aligned}
$$

For Case 2, we have only changes in mean with serially correlated errors. That is

$$\begin{aligned} y_t &= \mu_1 + v_t, \quad &\text{if } 1 < t \leq [T/3], \\ y_t &= \mu_2 + v_t, \quad &\text{if } [T/3] < t \leq [2T/3], \\ y_t &= \mu_3 + v_t, \quad &\text{if } [2T/3] < t \leq T, \end{aligned}$$

where $v_t = 0.5v_{t-1} + e_t$ with $e_t \sim i.i.d.\ N(0, 1)$.

We first consider Case 1 where the data and errors are identically distributed across segments, that is $\sigma_1^2 = \sigma_2^2 = \sigma_3^2$ and $\varsigma_1 = \varsigma_2 = \varsigma_3$. Results are first presented in Table 3 for cases where the shifts involve either only the intercept (rows (a) to (h)) or in the slope (rows (j) to (o)). In all cases $T = 120$, $T_1 = 40$, $T_2 = 80$, $\varepsilon = .05$, $cor\_u = 0$, $het\_u = 0$, and $het\_z = 1$ for the construction of the tests and $het\_z = 0$ for the construction of the confidence intervals for the break dates.

We start with a case where the detection of the number of breaks is notoriously difficult. Here, the intercept increases by some value at $T_1 = 40$ and goes back to its original value at $T_2 = 80$. Row (a) considers the case where this change is .5. The power is, indeed, very low and all methods basically select no break. The case where the change is 1 (row (b)) is very instructive about the usefulness of the $D\max$ tests and the $\sup F(\ell + 1|\ell)$ test to determine the number of breaks. Here the power of the $\sup F(1)$ test is very low and, hence, the sequential procedure selects 2 breaks only 31% of the time. However, the $UD\max$ and $WD\max$ tests have high power (82% and 88%, respectively). The sup(2|1) test also has high power (73%). Hence, a useful strategy is to fist decide that some break is present based on the $D\max$ test. Then look at the $\sup F(\ell+1|\ell)$ to see if more than one is present. In the example of row (b) this would lead to selecting 2 breaks 64% of the time. Another example of the usefulness of this strategy is presented in row (k). Here there is a change in slope from 1 to 2 then back to 1. The sequential procedure chooses 2 breaks only 69% of the time. However, the strategy discussed above would lead to select 2 breaks almost 100% of the times since the $D\max$ tests have 99% power and the $supF(2|1)$ has 98% power.

The case discussed above clearly show the usefulness of considering tests for multiple structural changes. As shown in Andrews (1993) a test for a single change is consistent against an alternative hypothesis of multiple changes. However, as shown here, in finite samples its power can be quite low while tests against more than one change can have much higher power. This also suggests that a mechanical application of a specific to general sequential testing procedure to select the number of breaks can be sub-optimal. Indeed, in practice it is advisable to look at the double maximum tests first to avoid such cases where it is difficult to distinguish between no break and a single break while it is easy to distinguish between no break and more than on break.

The other cases of Table 3 show various configurations for changes in intercept or slope. The results can be summarized as follows. First, intercept changes of the form $\mu_1 = 0$, $\mu_2 = 1$, $\mu_3 = 2$ (increasing steps) are also difficult cases where most procedures fail to select two breaks (the same is true for slope changes of the same form). In general, when the magnitude of the change is small (or difficult to identify) the coverage rates for the break dates are too small (e.g. rows (a,b,j,l,o)). If the changes are very large (e.g., row (h) or row (f, second break)) they are too wide. However, in most cases where the number of breaks is well identified the coverage rates are adequate.

Table 4 first considers Case 1 with simultaneous changes in intercept and slope. Row (a) shows that very little gain in power or accuracy of the coverage rates is gained when two shifts that are very difficult to identify individually occur simultaneously. However, rows (b) and (c) shows that important gains can be obtained in other cases (in particular compare row (b) of Table 4 with row (c) of Table 3).

The other parts of Table 4 consider Case 2 with intercept shifts and serially correlated errors with the specification $cor\_u = 1$. Rows (d) to (k) consider the difficult cases where the mean return to its old value at the second break. Here power is low when the change is .5 and even 1. Hence, serial correlation induces a loss in power. The coverage rates are adequate and we conclude that the non-parametric correction for the presence of serial correlation works well. Also, we see that for given changes in mean, an increase in the sample size has some effect on power, probably due to the fact that, for given trimming ϵ, a larger number of observations allows more precise estimates of nuisance parameters related to correlation in the residuals. When the change in mean is larger, say 2 or 4 (see rows (h) to (k)) the power of the $\sup F(1)$ test is low but the power of the $supF(2)$ and $\sup F(2|1)$ tests are high. Hence, a model selection strategy based on these statistic would conclude basically 100% of the times that 2 breaks are present.

Tables 5.a and 5.b consider cases where the distribution of the errors and the data are heterogenous across segments. The goal is to see if applying the required corrections lead to tests, model selections and coverage rates that are better. Table 5.a considers data generated by the two breaks model with $\gamma_1 = 1, \gamma_2 = 1.5, \gamma_3 = 0.5$ and $\mu_1 = 0, \mu_2 = 1.5, \mu_3 = .5$. Table 5.b considers data generated by the two breaks model with $\gamma_1 = 1, \gamma_2 = 1.5, \gamma_3 = 2$ and $\mu_1 = 0, \mu_2 = .5, \mu_3 = 1$. In all cases, $\sigma_1^2 = \sigma_3^2 = 1$, $\varsigma_1 = \varsigma_3 = 1$ and we vary σ_2^2 and ς_2. To ensure tests with adequate sizes, we set $\varepsilon = .15$ for the cases in Table 5.a and we consider $\varepsilon = .20$ for the cases in Table 5.b. We compare the properties of the procedures using the uncorrected versions ($het\_z = 1$ and $het\_u = 0$ in the construction of the tests, $het\_z = het\_u = 0$ in the construction of the confidence intervals) and the corrected versions ($het\_z = het\_u = 1$ in the construction of the tests and in the construction of the confidence intervals). The relevant columns are the $\sup F(2|1)$ test, the probabilities of selecting 2 breaks and the coverage rates of the break dates (note that for the selection procedures based on the BIC and LWZ, only the uncorrected version is presented since these methods cannot be modified to account for heterogeneity across segments).

The results show that important gains in the power of the tests can be obtained when allowing for different distribution of the errors across segments. In almost all cases, the power of the sup(2|1) test is higher when corrected. For example, in Table 5.b when the variance of the errors is four times higher in the middle segment (and the mean of the regressors is also 4 times higher) and $T = 120$ (row(g)), the power of the uncorrected version is .53 while it is .78 when allowing for different variances. This also translates into a higher probability of selecting two breaks, 76% instead of 52% making the sequential procedure more adequate to select the number of breaks than the BIC. Even stronger comparisons obtain with the second case presented in Table 5.b. For example, in row (g) we see an increase in the power of the $\sup F(2|1)$ test and the probability of choosing 2 breaks rising from 22% to 60%. The results also show that correcting for heterogeneity in the data improves the coverage rates of the confidence intervals of the break dates.

5 Summary and Practical Recommendations.

The simulations have shown the tests, model selection procedures and the construction of the confidence intervals for the break dates to be useful tools to analyze models with multiple breaks. However, care must be taken when using particular specifications. We make the following recommendations.

- First, ensure that the specifications are such that the size of the tests are adequate under the hypothesis of no break. If serial correlation and/or heterogeneity in the data or errors across segments are not allowed in the estimated regression model (and not present in the DGP), using any value of the trimming ε will lead to tests with adequate sizes. However, if such features are allowed, a higher trimming is needed. The simulations show that, with a sample of $T = 120$, $\varepsilon = .15$ should be enough for heterogeneity in the errors or the data. If serial correlation is allowed, $\varepsilon = .20$ may be needed. These could be reduced if larger sample sizes are available.

- Overall, selecting the break point using the BIC works well when breaks are present but less so under the null hypothesis, especially if serial correlation is present. The method based on the LWZ criterion works better under the null hypothesis (even with serial correlation) by imposing a higher penalty. However, this higher penalty translates into a very bad performance when breaks are present. Also, model selection procedures based on information criteria cannot take into account potential heterogeneity across segments unlike the sequential method. Overall, the sequential procedure works best in selecting the number of breaks.

- There are important instances where the performance of the sequential procedure can be improved. A useful strategy is to first look at the $UD\max$ or $WD\max$ tests to see if at least a break is present. Then the number of breaks can be decided based upon an examination of the $\sup F(\ell+1|\ell)$ statistics constructed using estimates of the break dates obtained from a global minimization of the sum of squared residuals. This is, in our opinion, the preferred strategy.

- The power of the $UD\max$ or $WD\max$ tests is almost as high as the power of a test of no change versus an alternative hypothesis that specifies the true number of changes. This provides added justifications for its use in practice.

- The coverage rates for the break dates are adequate unless the break is either too small (so small as not to be detected by the tests) or too big. This is, from a practical point of view, however, an encouraging result. The confidence intervals are inadequate (in that they miss the true break value too often) exactly in those cases where it would be quite difficult to conclude that a break is present (in which case they would not be used anyway). When the breaks are very large the confidence intervals do contain the true values but are quite wide leading to a conservative assessment of the accuracy of the estimates. It was found that correcting for heterogeneity in the data and/or errors across segments yields improvements over a more straightforward uncorrected interval. Correcting for serial correlation also does lead to substantial improvements.

- Correcting for heterogeneity in the distribution of the data or the errors and for serial correlation also improves the power of the tests and the accuracy in the selection of the number of breaks.

References

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

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

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

[4] Bai, J. (1997a): "Estimation of a Change Point in Multiple Regression Models," *Review of Economic and Statistics* 79, 551-563.

[5] Bai, J. (1997b): "Estimating Multiple Breaks one at a Time," *Econometric Theory* 13, 315-352.

[6] Bai, J. and P. Perron (1998): "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica* 66, 47-78.

[7] Bai, J. and P. Perron (2003a): "Computation and Analysis of Muliple Structural Change Models," *Journal of Applied Econometrics* 18, 1-22.

[8] Bai, J. and P. Perron (2003b): "Critical Values for Multiple Structural Change Tests," *Econometrics Journal* 6, 72-78.

[9] Liu, J., S. Wu, and J.V. Zidek (1997): "On Segmented Multivariate Regressions," *Statistica Sinica* 7, 497-525.

[10] Perron, P. (1997): "L'estimation de modèles avec changements structurels multiples," *Actualité Économique* 73, 457-505.

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

Table 1: Size of the tests and probabilities of selecting breaks

	DGP-1					DGP-2					DGP-3				
ϵ	.05	.10	.15	.20	.25	.05	.10	.15	.20	.25	.05	.10	.15	.20	.25
$cor\_u = 0, het\_u = 0$															
$\sup F(1)$	.05	.04	.05	.04	.04	.05	.04	.05	.05	.05	.05	.06	.07	.05	.06
$\sup F(2)$	.05	.05	.05	.04	.04	.04	.04	.04	.05	.05	.06	.07	.08	.07	.06
$\sup F(3)$	.05	.05	.04	.03		.05	.05	.04	.05		.09	.09	.08	.07	
$\sup F(4)$	.06	.05	.04			.07	.06	.04			.12	.11	.08		
$\sup F(5)$	.06	.05	.04			.08	.07	.03			.15	.12	.07		
$UDMAX$	.05	.04	.05	.04	.04	.05	.04	.05	.05	.05	.06	.06	.07	.06	.07
$WDMAX$	.06	.05	.04	.04	.04	.06	.05	.05	.05	.05	.10	.09	.09	.07	.06
$Sequa - Pr[m = 0]$	.95	.96	.95	.96	.96	.95	.96	.95	.95	.95	.95	.95	.94	.95	.94
$Sequa - Pr[m = 1]$	.05	.04	.05	.04	.04	.05	.04	.05	.05	.05	.05	.05	.06	.05	.06
$Sequa - Pr[m = 2]$	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
$cor\_u = 0, het\_u = 1$															
$\sup F(1)$	.10	.06	.06	.05	.04	.16	.08	.07	.06	.05	.18	.10	.10	.07	.07
$\sup F(2)$	.24	.11	.08	.06	.06	.35	.14	.09	.06	.06	.40	.22	.14	.10	.08
$\sup F(3)$	.24	.11	.07	.05		.42	.17	.08	.07		.49	.26	.14	.11	
$\sup F(4)$	.29	.11	.07			.48	.19	.08			.59	.29	.15		
$\sup F(5)$	.31	.12	.06			.53	.18	.07			.65	.30	.13		
$UDMAX$	.27	.10	.06	.05	.04	.46	.14	.08	.06	.06	.51	.17	.13	.09	.08
$WDMAX$	.33	.12	.07	.06	.05	.57	.19	.10	.07	.06	.66	.27	.16	.10	.08
$Sequa - Pr[m = 0]$	.90	.94	.94	.95	.96	.85	.92	.93	.94	.95	.82	.90	.90	.93	.93
$Sequa - Pr[m = 1]$	.09	.06	.06	.05	.04	.14	.08	.07	.06	.05	.16	.09	.09	.07	.07
$Sequa - Pr[m = 2]$	.01	.00	.00	.00	.00	.01	.00	.00	.00	.00	.02	.01	.01	.00	.00
$cor\_u = 1, het\_u = 0$															
$\sup F(1)$	.06	.06	.06	.05	.05	.08	.06	.06	.07	.07					
$\sup F(2)$	.08	.08	.07	.06	.06	.10	.09	.09	.08	.08					
$\sup F(3)$	.11	.10	.08	.05		.14	.12	.10	.08						
$\sup F(4)$	.15	.12	.08			.18	.16	.10							
$\sup F(5)$	.21	.14	.07			.23	.20	.10							
$UDMAX$	.08	.07	.07	.06	.05	.12	.09	.08	.08	.07					
$WDMAX$	.14	.11	.08	.06	.05	.21	.17	.11	.09	.07					
$Sequa - Pr[m = 0]$	.94	.95	.94	.95	.95	.92	.94	.93	.93	.93					
$Sequa - Pr[m = 1]$	.06	.05	.06	.05	.05	.07	.06	.07	.07	.07					
$Sequa - Pr[m = 2]$	.00	.00	.00	.00	.00	.01	.00	.00	.00	.00					
$cor\_u = 1, het\_u = 1$															
$\sup F(1)$	.12	.08	.07	.05	.05	.25	.14	.11	.10	.08					
$\sup F(2)$	.29	.14	.10	.07	.07	.54	.31	.19	.13	.10					
$\sup F(3)$	.32	.15	.10	.07		.65	.39	.22	.15						
$\sup F(4)$	.37	.16	.09			.75	.44	.25							
$\sup F(5)$	.39	.16	.09			.81	.48	.24							
$UDMAX$	.36	.14	.09	.07	.05	.77	.35	.18	.12	.09					
$WDMAX$	.43	.17	.10	.07	.06	.86	.49	.24	.15	.10					
$Sequa - Pr[m = 0]$	.88	.92	.93	.95	.95	.75	.86	.89	.90	.92					
$Sequa - Pr[m = 1]$	.11	.08	.07	.05	.05	.21	.13	.11	.10	.08					
$Sequa - Pr[m = 2]$	.01	.00	.00	.00	.00	.04	.01	.00	.00	.00					
$BIC - Pr[m = 0]$	.94	.96	.97	.98	.98	.97	.98	.99	.99	.99	.97	.98	.98	.98	.99
$BIC - Pr[m = 1]$	.04	.03	.03	.02	.02	.03	.02	.01	.01	.01	.03	.02	.02	.02	.01
$BIC - Pr[m = 2]$	.02	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
$LWZ - Pr[m = 0]$	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
$LWZ - Pr[m = 1]$	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
$LWZ - Pr[m = 2]$	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00

Table 1 (cont'd): Size of the tests and probabilities of selecting breaks

	DGP-4					DGP-5					DGP-6				
ϵ	.05	.10	.15	.20	.25	.05	.10	.15	.20	.25	.05	.10	.15	.20	.25
$cor\_u = 0, het\_u = 0$															
$\sup F(1)$	.52	.52	.50	.44	.44	.24	.24	.24	.21	.20	.00	.00	.00	.00	.00
$\sup F(2)$	.82	.77	.69	.59	.53	.42	.38	.34	.29	.24	.00	.00	.00	.00	.00
$\sup F(3)$	.90	.84	.74	.60		.51	.44	.39	.30		.00	.00	.00	.00	
$\sup F(4)$	.94	.88	.75			.60	.49	.39			.00	.00	.00		
$\sup F(5)$	.95	.89	.71			.67	.53	.36			.00	.00	.00		
$UDMAX$	.81	.73	.64	.53	.49	.38	.33	.31	.25	.23	.00	.00	.00	.00	.00
$WDMAX$	.92	.85	.73	.59	.53	.55	.46	.38	.28	.24	.00	.00	.00	.00	.00
$Sequa - Pr[m = 0]$	.48	.48	.50	.56	.56	.76	.76	.76	.79	.80	1.0	1.0	1.0	1.0	1.0
$Sequa - Pr[m = 1]$	.29	.31	.33	.31	.35	.18	.18	.21	.18	.18	.00	.00	.00	.00	.00
$Sequa - Pr[m = 2]$	.16	.15	.14	.12	.09	.05	.05	.03	.03	.02	.00	.00	.00	.00	.00
$cor\_u = 0, het\_u = 1$															
$\sup F(1)$	.61	.58	.52	.46	.44	.35	.30	.28	.22	.21	.01	.00	.00	.00	.00
$\sup F(2)$	.92	.84	.72	.62	.56	.70	.51	.41	.31	.27	.04	.00	.00	.00	.00
$\sup F(3)$	.96	.89	.77	.64		.76	.57	.45	.33		.02	.00	.00	.00	
$\sup F(4)$	.98	.92	.78			.84	.63	.46			.03	.00	.00		
$\sup F(5)$	.99	.93	.76			.87	.65	.43			.02	.00	.00		
$UDMAX$	.96	.84	.70	.57	.51	.77	.49	.38	.27	.24	.04	.00	.00	.00	.00
$WDMAX$	.98	.92	.78	.62	.55	.85	.62	.46	.31	.27	.04	.00	.00	.00	.00
$Sequa - Pr[m = 0]$	.39	.42	.48	.55	.56	.65	.70	.73	.78	.79	.99	1.0	1.0	1.0	1.0
$Sequa - Pr[m = 1]$	.27	.32	.33	.32	.35	.23	.23	.23	.19	.19	.01	.00	.00	.00	.00
$Sequa - Pr[m = 2]$	.21	.18	.15	.12	.09	.09	.06	.04	.03	.02	.00	.00	.00	.00	.00
$cor\_u = 1, het\_u = 0$															
$\sup F(1)$	.07	.08	.08	.07	.07	.09	.08	.10	.08	.08	.03	.03	.03	.03	.03
$\sup F(2)$	.12	.14	.12	.10	.09	.15	.16	.14	.12	.11	.05	.04	.04	.03	.03
$\sup F(3)$	.22	.21	.16	.10		.25	.22	.17	.12		.07	.06	.04	.03	
$\sup F(4)$	.34	.28	.17			.38	.29	.18			.11	.08	.05		
$\sup F(5)$	.46	.32	.15			.47	.32	.17			.15	.09	.04		
$UDMAX$	.18	.14	.11	.08	.08	.18	.14	.12	.10	.08	.04	.03	.03	.03	.03
$WDMAX$	.35	.27	.15	.11	.08	.36	.26	.17	.12	.10	.09	.06	.05	.03	.03
$Sequa - Pr[m = 0]$	.93	.92	.92	.93	.93	.91	.92	.91	.92	.93	.97	.97	.97	.97	.97
$Sequa - Pr[m = 1]$	.06	.08	.08	.06	.07	.08	.08	.09	.08	.07	.03	.03	.03	.03	.03
$Sequa - Pr[m = 2]$	.01	.00	.00	.00	.00	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00
$cor\_u = 1, het\_u = 1$															
$\sup F(1)$	.22	.15	.13	.08	.08	.18	.13	.11	.09	.08	.04	.02	.02	.02	.02
$\sup F(2)$	.56	.36	.23	.14	.12	.51	.28	.19	.13	.11	.11	.04	.03	.02	.02
$\sup F(3)$	.64	.41	.26	.15		.56	.31	.21	.13		.10	.03	.02	.02	
$\sup F(4)$	.76	.47	.28			.65	.34	.21			.11	.03	.02		
$\sup F(5)$	.82	.51	.28			.70	.37	.19			.10	.03	.02		
$UDMAX$	.75	.39	.24	.11	.09	.62	.26	.17	.11	.09	.11	.03	.02	.02	.02
$WDMAX$	.84	.52	.31	.15	.11	.72	.36	.21	.13	.11	.14	.04	.03	.02	.02
$Sequa - Pr[m = 0]$	.78	.85	.87	.92	.93	.82	.87	.89	.92	.92	.96	.98	.98	.98	.98
$Sequa - Pr[m = 1]$	.17	.14	.12	.08	.07	.15	.12	.11	.08	.08	.04	.02	.02	.02	.02
$Sequa - Pr[m = 2]$	.04	.01	.01	.00	.00	.03	.01	.00	.00	.00	.00	.00	.00	.00	.00
$BIC - Pr[m = 0]$	.21	.33	.45	.58	.63	.62	.71	.77	.84	.87	1.0	1.0	1.0	1.0	1.0
$BIC - Pr[m = 1]$	.11	.20	.25	.23	.26	.13	.13	.15	.12	.11	.00	.00	.00	.00	.00
$BIC - Pr[m = 2]$	.21	.24	.22	.17	.11	.15	.12	.07	.04	.02	.00	.00	.00	.00	.00
$LWZ - Pr[m = 0]$	.82	.84	.87	.90	.92	.97	.98	.98	.98	.99	1.0	1.0	1.0	1.0	1.0
$LWZ - Pr[m = 1]$	.12	.10	.10	.08	.07	.03	.02	.02	.02	.01	.00	.00	.00	.00	.00
$LWZ - Pr[m = 2]$	.05	.05	.03	.01	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00

Table 2: Power of the tests and break selection when $m = 1$.

				Tests (probability of rejection)							Probability of selecting k breaks									Coverage
				$\sup F(k)$			$\sup F(\ell+1\|\ell)$		$D\max$		Sequa			BIC			LWZ			Rate
	Case	Values	Specifications[1]	1	2	3	2|1	3|2	U	W	0	1	2	0	1	2	0	1	2	95%
a)	1 $\varepsilon = .05$ $T = 120$	$\gamma_1 = \gamma_2 = 1$ $\mu_1 = 0, \mu_2 = .5$	$cor\_u = 0$	.43	.35	.34	.03	.01	.42	.42	.57	.42	.02	.66	.32	.02	.98	.02	.00	.74
b)	1 $\varepsilon = .05$ $T = 240$	$\gamma_1 = \gamma_2 = 1$ $\mu_1 = 0, \mu_2 = .5$	$cor\_u = 0$	.66	.53	.50	.02	.01	.65	.62	.34	.65	.01	.57	.43	.00	.99	.01	.00	.80
c)	1 $\varepsilon = .05$ $T = 120$	$\gamma_1 = \gamma_2 = 1$ $\mu_1 = 0, \mu_2 = 1$	$cor\_u = 0$	.99	.97	.96	.04	.02	.99	.99	.01	.95	.04	.02	.95	.03	.36	.64	.00	.93
d)	1 $\varepsilon = .05$ $T = 120$	$\gamma_1 = 1, \gamma_2 = 1.5$ $\mu_1 = \mu_2 = 0$	$cor\_u = 0$	.79	.69	.66	.03	.01	.78	.77	.21	.77	.02	.28	.68	.03	.87	.13	.00	.83
e)	1 $\varepsilon = .05$ $T = 120$	$\gamma_1 = 1, \gamma_2 = 1.5$ $\mu_1 = 0, \mu_2 = .5$	$cor\_u = 0$	1.0	.99	.98	.04	.02	1.0	1.0	.00	.96	.04	.01	.96	.03	.18	.82	.00	.94
f)	1 $\varepsilon = .05$ $T = 120$	$\gamma_1 = 1, \gamma_2 = 2$ $\mu_1 = \mu_2 = 0$	$cor\_u = 0$	1.0	1.0	1.0	.04	.02	1.0	1.0	.00	.96	.04	.00	.97	.03	.02	.98	.00	.93
g)	1 $\varepsilon = .05$ $T = 120$	$\gamma_1 = 1, \gamma_2 = 2$ $\mu_1 = 0, \mu_2 = 1$	$cor\_u = 0$	1.0	1.0	1.0	.04	.02	1.0	1.0	.00	.96	.04	.00	.97	.03	.00	1.0	.00	.96
h)	2 $\varepsilon = .20$ $T = 120$	$\mu_1 = 0, \mu_2 = .5$	$cor\_u = 1$	.25	.27	.30	.04	.00	.30	.31	.75	.24	.01	.32	.48	.18	.68	.29	.03	.93
i)	2 $\varepsilon = .20$ $T = 240$	$\mu_1 = 0, \mu_2 = .5$	$cor\_u = 1$	.38	.34	.31	.02	.00	.39	.38	.62	.37	.01	.21	.58	.19	.64	.35	.01	.91
j)	2 $\varepsilon = .20$ $T = 120$	$\mu_1 = 0, \mu_2 = 1$	$cor\_u = 1$	.66	.61	.61	.03	.00	.68	.69	.34	.63	.02	.05	.71	.22	.23	.74	.04	.89
k)	2 $\varepsilon = .20$ $T = 240$	$\mu_1 = 0, \mu_2 = 1$	$cor\_u = 1$	.91	.85	.82	.03	.00	.91	.90	.09	.88	.03	.01	.74	.23	.07	.90	.02	.91

Note: [1] In all cases $het\_u = 0$. When constructing the tests, $het\_z = 1$ and when constructing the confidence intervals $het\_z = 0$.

Table 3: Power of the tests and break selection when $m = 2$.

Case 1, $T = 120$, $\varepsilon = .05$, $cor\_u = 0$, $het\_u = 0$.[1]

		Tests (probability of rejection)							Probability of selecting k breaks									Coverage	
		sup $F(k)$			sup $F(\ell+1\|\ell)$		$D\max$		Sequa			BIC			LWZ			Rate	95%
	Values	1	2	3	2|1	3|2	U	W	0	1	2	0	1	2	0	1	2	#1	#2
a)	$\gamma_1 = \gamma_2 = \gamma_3 = 1$ $\mu_1 = \mu_3 = 0, \mu_2 = .5$	.13	.23	.26	.11	.01	.18	.25	.87	.11	.02	.90	.06	.04	1.0	.00	.00	.51	.49
b)	$\gamma_1 = \gamma_2 = \gamma_3 = 1$ $\mu_1 = \mu_3 = 0, \mu_2 = 1$	.41	.89	.89	.73	.03	.82	.88	.59	.08	.31	.31	.05	.62	.98	.00	.02	.87	.85
c)	$\gamma_1 = \gamma_2 = \gamma_3 = 1$ $\mu_1 = 0, \mu_2 = 1, \mu_3 = 2$	1.0	1.0	1.0	.56	.03	1.0	1.0	.00	.44	.54	.00	.38	.59	.00	.96	.04	.88	.86
d)	$\gamma_1 = \gamma_2 = \gamma_3 = 1$ $\mu_1 = 0, \mu_2 = 1, \mu_3 = -1$	1.0	1.0	1.0	.86	.04	1.0	1.0	.00	.14	.82	.00	.13	.83	.02	.67	.31	.89	.96
f)	$\gamma_1 = \gamma_2 = \gamma_3 = 1$ $\mu_1 = 0, \mu_2 = -1, \mu_3 = 2$	1.0	1.0	1.0	.86	.05	1.0	1.0	.00	.14	.82	.00	.13	.82	.00	.71	.29	.88	.99
g)	$\gamma_1 = \gamma_2 = \gamma_3 = 1$ $\mu_1 = 0, \mu_2 = 1, \mu_3 = 3$	1.0	1.0	1.0	.83	.06	1.0	1.0	.00	.17	.77	.00	.15	.80	.00	.75	.25	.88	.96
h)	$\gamma_1 = \gamma_2 = \gamma_3 = 1$ $\mu_1 = 0, \mu_2 = 2, \mu_3 = -1$	1.0	1.0	1.0	1.0	.05	1.0	1.0	.00	.00	.95	.00	.00	.96	.00	.00	1.0	.98	.99
j)	$\gamma_1 = \gamma_3 = 1, \gamma_2 = 1.5$ $\mu_1 = \mu_2 = \mu_3 = 0$	.22	.49	.52	.28	.03	.40	.50	.78	.14	.07	.75	.08	.16	1.0	.00	.00	.67	.66
k)	$\gamma_1 = \gamma_3 = 1, \gamma_2 = 2$ $\mu_1 = \mu_2 = \mu_3 = 0$	.77	1.0	1.0	.98	.04	.99	.99	.23	.01	.69	.02	.01	.93	.61	.01	.38	.92	.93
l)	$\gamma_1 = 1, \gamma_2 = 1.5, \gamma_3 = 2$ $\mu_1 = \mu_2 = \mu_3 = 0$	1.0	1.0	1.0	.14	.02	1.0	.99	.00	.85	.13	.01	.85	.13	.24	.76	.00	.68	.68
m)	$\gamma_1 = 1, \gamma_2 = 2, \gamma_3 = 3$ $\mu_1 = \mu_2 = \mu_3 = 0$	1.0	1.0	1.0	.97	.05	1.0	1.0	.00	.03	.88	.00	.02	.94	.00	.36	.64	.92	.92
o)	$\gamma_1 = 1, \gamma_2 = .5, \gamma_3 = -.5$ $\mu_1 = \mu_2 = \mu_3 = 0$	1.0	1.0	1.0	.41	.04	1.0	1.0	.00	.59	.39	.00	.57	.40	.00	.98	.02	.72	.87

Note: [1]For the construction of the tests, we use $het\_z = 1$ and for the construction of the confidence intervals of the break dates, we use $het\_z = 0$.

Table 4: Power of the tests and break selection when $m = 2$ (cont'd).

				Tests (probability of rejection)							Probability of selecting k breaks									Coverage	
				sup $F(k)$			sup $F(\ell+1\|\ell)$		$D\max$		Sequa			BIC			LWZ			Rate	95%
	Case	Values	Specifications[1]	1	2	3	2|1	3|2	U	W	0	1	2	0	1	2	0	1	2	#1	#2
a)	1 $\varepsilon = .05$ $T = 120$	$\gamma_1 = 1, \gamma_2 = .5, \gamma_3 = 1$ $\mu_1 = \mu_3 = 0, \mu_2 = .5$	$cor\_u = 0$	.12	.21	.24	.08	.01	.17	.25	.88	.10	.02	.90	.06	.03	1.0	.00	.00	.46	.46
b)	1 $\varepsilon = .05$ $T = 240$	$\gamma_1 = 1, \gamma_2 = 1.5, \gamma_3 = 2$ $\mu_1 = 0, \mu_3 = 1, \mu_2 = 2$	$cor\_u = 0$	1.0	1.0	1.0	1.0	.05	1.0	1.0	.00	.00	.89	.00	.00	.95	.00	.12	.88	.95	.95
c)	1 $\varepsilon = .05$ $T = 120$	$\gamma_1 = 1, \gamma_2 = 2, \gamma_3 = 1$ $\mu_1 = 0, \mu_2 = 1, \mu_3 = 2$	$cor\_u = 0$	1.0	1.0	1.0	.82	.04	1.0	1.0	.00	.19	.78	.00	.17	.79	.00	.72	.28	.95	.85
d)	2 $\varepsilon = .20$ $T = 120$	$\mu_1 = \mu_3 = 0, \mu_2 = .5$	$cor\_u = 1$	.14	.28	.25	.09	.00	.23	.28	.86	.13	.01	.39	.24	.37	.79	.12	.08	.96	.95
e)	2 $\varepsilon = .20$ $T = 240$	$\mu_1 = \mu_3 = 0, \mu_2 = .5$	$cor\_u = 1$	.18	.32	.29	.11	.00	.26	.31	.82	.15	.03	.32	.20	.46	.82	.11	.07	.96	.95
f)	2 $\varepsilon = .20$ $T = 120$	$\mu_1 = \mu_3 = 0, \mu_2 = 1$	$cor\_u = 1$	.25	.58	.53	.29	.00	.48	.55	.75	.15	.09	.13	.12	.73	.51	.12	.36	.94	.94
g)	2 $\varepsilon = .20$ $T = 240$	$\mu_1 = \mu_3 = 0, \mu_2 = 1$	$cor\_u = 1$	.43	.83	.74	.55	.00	.71	.78	.57	.16	.26	.03	.04	.90	.33	.11	.56	.93	.93
h)	2 $\varepsilon = .20$ $T = 120$	$\mu_1 = \mu_3 = 0, \mu_2 = 2$	$cor\_u = 1$	.47	.97	.94	.86	.00	.94	.96	.53	.06	.41	.00	.00	.95	.02	.01	.96	.93	.93
i)	2 $\varepsilon = .20$ $T = 240$	$\mu_1 = \mu_3 = 0, \mu_2 = 2$	$cor\_u = 1$	.91	1.0	1.0	1.0	.00	1.0	1.0	.09	.00	.90	.00	.00	.97	.00	.00	1.0	.95	.94
j)	2 $\varepsilon = .20$ $T = 120$	$\mu_1 = \mu_3 = 0, \mu_2 = 4$	$cor\_u = 1$	.37	1.0	1.0	1.0	.00	1.0	1.0	.63	.00	.37	.00	.00	1.0	.00	.00	1.0	.99	.99
k)	2 $\varepsilon = .20$ $T = 240$	$\mu_1 = \mu_3 = 0, \mu_2 = 4$	$cor\_u = 1$	.96	1.0	1.0	1.0	.00	1.0	1.0	.04	.00	.96	.00	.00	1.0	.00	.00	1.0	.99	.99

Note: [1] In all cases $het\_u = 0$. When constructing the tests, $het\_z = 1$ and when constructing the confidence intervals $het\_z = 0$.

Table 5.a: Power of the tests and break selection when $m = 2$.

Different distributions for the errors and data across segments; $cor\_u = 0$, $\varepsilon = .15$.

Case 1 with $\gamma_1 = 1, \gamma_2 = 1.5, \gamma_3 = .5$ and $\mu_1 = 0, \mu_2 = .5, \mu_3 = -.5$.

			Tests (probability of rejection)					Probability of selecting k breaks									Coverage	
			sup $F(k)$			sup $F(\ell+1\|\ell)$		Sequa			BIC			LWZ			Rate	95%
	Values	Specifications[1]	1	2	3	2|1	3|2	0	1	2	0	1	2	0	1	2	#1	#2
a)	$T=120$	uncorrected	1.0	1.0	1.0	.91	.02	.00	.09	.89	.00	.08	.89	.01	.57	.42	.90	.96
	$\sigma_1^2=1,\sigma_2^2=2,\sigma_3^2=1$																	
	$\varsigma_1=1,\varsigma_2=2,\varsigma_3=1$	corrected	1.0	1.0	1.0	.94	.01	.00	.06	.92							.89	.96
b)	$T=240$	uncorrected	1.0	1.0	1.0	1.0	.02	.00	.00	.98	.00	.00	.98	.00	.20	.80	.93	.96
	$\sigma_1^2=1,\sigma_2^2=2,\sigma_3^2=1$																	
	$\varsigma_1=1,\varsigma_2=2,\varsigma_3=1$	corrected	1.0	1.0	1.0	1.0	.02	.00	.00	.98							.93	.97
c)	$T=120$	uncorrected	1.0	1.0	1.0	.78	.02	.00	.22	.77	.00	.21	.76	.01	.75	.23	.83	.94
	$\sigma_1^2=1,\sigma_2^2=2,\sigma_3^2=1$																	
	$\varsigma_1=1,\varsigma_2=4,\varsigma_3=1$	corrected	1.0	1.0	1.0	.89	.02	.00	.11	.87							.89	.96
d)	$T=240$	uncorrected	1.0	1.0	1.0	.99	.02	.00	.01	.98	.00	.02	.97	.00	.44	.56	.87	.97
	$\sigma_1^2=1,\sigma_2^2=2,\sigma_3^2=1$																	
	$\varsigma_1=1,\varsigma_2=4,\varsigma_3=1$	corrected	1.0	1.0	1.0	1.0	.01	.00	.00	.99							.92	.98
e)	$T=120$	uncorrected	1.0	1.0	1.0	.70	.02	.00	.30	.68	.00	.22	.73	.09	.68	.23	.84	.93
	$\sigma_1^2=1,\sigma_2^2=4,\sigma_3^2=1$																	
	$\varsigma_1=1,\varsigma_2=2,\varsigma_3=1$	corrected	1.0	1.0	1.0	.76	.02	.00	.24	.75							.87	.94
f)	$T=240$	uncorrected	1.0	1.0	1.0	.97	.02	.00	.03	.96	.00	.03	.93	.00	.54	.46	.88	.94
	$\sigma_1^2=1,\sigma_2^2=4,\sigma_3^2=1$																	
	$\varsigma_1=1,\varsigma_2=2,\varsigma_3=1$	corrected	1.0	1.0	1.0	.98	.02	.00	.02	.97							.90	.96
g)	$T=120$	uncorrected	1.0	1.0	1.0	.54	.03	.00	.46	.53	.00	.36	.59	.09	.80	.11	.79	.93
	$\sigma_1^2=1,\sigma_2^2=4,\sigma_3^2=1$																	
	$\varsigma_1=1,\varsigma_2=4,\varsigma_3=1$	corrected	1.0	1.0	1.0	.80	.04	.00	.19	.79							.86	.94
h)	$T=240$	uncorrected	1.0	1.0	1.0	.91	.02	.00	.09	.90	.00	.09	.86	.00	.77	.23	.88	.97
	$\sigma_1^2=1,\sigma_2^2=4,\sigma_3^2=1$																	
	$\varsigma_1=1,\varsigma_2=4,\varsigma_3=1$	corrected	1.0	1.0	1.0	.98	.03	.00	.01	.96							.90	.97

Note: [1] Uncorrected means using $het\_z = 1$ and $het\_u = 0$ in the construction of the tests and $het\_z = 0$, $het\_u = 0$ in the construction of the confidence intervals. Corrected means that $het\_z = 1$ and $het\_u = 1$ for the construction of the tests and the confidence intervals.

Table 5.b: Power of the tests and break selection when $m = 2$.

Different distributions for the errors and data across segments; $cor\_u = 0, \varepsilon = .20$.

Case 1 with $\gamma_1 = 1, \gamma_2 = 1.5, \gamma_3 = 2$ and $\mu_1 = 0, \mu_2 = .5, \mu_3 = 1$

			Tests (probability of rejection)					Probability of selecting k breaks									Coverage	
			sup $F(k)$			sup $F(\ell+1\|\ell)$		Sequa			BIC			LWZ			Rate	95%
	Values	Specifications[1]	1	2	3	2|1	3|2	0	1	2	0	1	2	0	1	2	#1	#2
a)	$T = 120$	Uncorrected	1.0	1.0	1.0	.73	.00	.00	.27	.73	.00	.28	.72	.01	.91	.09	.92	.91
	$\sigma_1^2 = 1, \sigma_2^2 = 2, \sigma_3^2 = 1$																	
	$\varsigma_1 = 1, \varsigma_2 = 2, \varsigma_3 = 1$	Corrected	1.0	1.0	1.0	.78	.00	.00	.22	.77							.91	.90
b)	$T = 240$	Uncorrected	1.0	1.0	1.0	1.0	.00	.00	.01	.98	.00	.01	.99	.00	.46	.54	.94	.94
	$\sigma_1^2 = 1, \sigma_2^2 = 2, \sigma_3^2 = 1$																	
	$\varsigma_1 = 1, \varsigma_2 = 2, \varsigma_3 = 1$	Corrected	1.0	1.0	1.0	1.0	.00	.00	.00	.98							.94	.94
c)	$T = 120$	Uncorrected	1.0	1.0	1.0	.63	.00	.00	.37	.63	.00	.39	.61	.00	.95	.05	.83	.85
	$\sigma_1^2 = 1, \sigma_2^2 = 2, \sigma_3^2 = 1$																	
	$\varsigma_1 = 1, \varsigma_2 = 4, \varsigma_3 = 1$	Corrected	1.0	1.0	1.0	.79	.00	.00	.21	.79							.90	.91
d)	$T = 240$	Uncorrected	1.0	1.0	1.0	.99	.00	.00	.02	.98	.00	.03	.97	.00	.69	.31	.88	.89
	$\sigma_1^2 = 1, \sigma_2^2 = 2, \sigma_3^2 = 1$																	
	$\varsigma_1 = 1, \varsigma_2 = 4, \varsigma_3 = 1$	Corrected	1.0	1.0	1.0	.99	.00	.00	.01	.98							.93	.92
e)	$T = 120$	Uncorrected	1.0	1.0	1.0	.28	.00	.00	.72	.28	.00	.68	.32	.08	.91	.01	.90	.91
	$\sigma_1^2 = 1, \sigma_2^2 = 4, \sigma_3^2 = 1$																	
	$\varsigma_1 = 1, \varsigma_2 = 2, \varsigma_3 = 1$	Corrected	1.0	1.0	1.0	.40	.02	.00	.60	.40							.92	.92
f)	$T = 240$	Uncorrected	1.0	1.0	1.0	.87	.00	.00	.13	.87	.00	.15	.85	.00	.94	.06	.92	.91
	$\sigma_1^2 = 1, \sigma_2^2 = 4, \sigma_3^2 = 1$																	
	$\varsigma_1 = 1, \varsigma_2 = 2, \varsigma_3 = 1$	Corrected	1.0	1.0	1.0	.91	.00	.00	.09	.90							.92	.93
g)	$T = 120$	Uncorrected	1.0	1.0	1.0	.22	.01	.00	.78	.22	.00	.71	.28	.06	.93	.01	.84	.84
	$\sigma_1^2 = 1, \sigma_2^2 = 4, \sigma_3^2 = 1$																	
	$\varsigma_1 = 1, \varsigma_2 = 4, \varsigma_3 = 1$	Corrected	1.0	1.0	1.0	.60	.02	.00	.40	.60							.90	.90
h)	$T = 240$	Uncorrected	1.0	1.0	1.0	.82	.00	.00	.18	.81	.00	.23	.77	.00	.97	.03	.88	.89
	$\sigma_1^2 = 1, \sigma_2^2 = 4, \sigma_3^2 = 1$																	
	$\varsigma_1 = 1, \varsigma_2 = 4, \varsigma_3 = 1$	Corrected	1.0	1.0	1.0	.98	.01	.00	.02	.96							.91	.91

Note: [1] Uncorrected means using $het\_z = 1$ and $het\_u = 0$ in the construction of the tests and $het\_z = 0$, $het\_u = 0$ in the construction of the confidence intervals. Corrected means that $het\_z = 1$ and $het\_u = 1$ for the construction of the tests and the confidence intervals.