

A Comparison of Alternative Methods to Construct Confidence Intervals for the Estimate of a Break Date in Linear Regression Models

Seongyeon Chang*

Pierre Perron[†]

Boston University

Boston University

September 19, 2013

Abstract

This paper considers constructing confidence intervals for the date of a structural break in linear regression models. Using extensive simulations, we compare the performance of various procedures in terms of exact coverage rates and lengths of the confidence intervals. These include the procedures of Bai (1997) based on the asymptotic distribution under a shrinking shift framework, Elliott and Müller (2007) based on inverting a test locally invariant to the magnitude of break, Eo and Morley (2013) based on inverting a likelihood ratio test, and various bootstrap procedures. On the basis of achieving an exact coverage rate that is closest to the nominal level, Elliott and Müller's (2007) approach is by far the best one. However, this comes with a very high cost in terms of the length of the confidence intervals. When the errors are serially correlated and dealing with a change in intercept or a change in the coefficient of a stationary regressor with a high signal to noise ratio, the length of the confidence interval increases and approaches the whole sample as the magnitude of the change increases. The same problem occurs in models with a lagged dependent variable, a common case in practice. This drawback is not present for the other methods, which have similar properties. Theoretical results are provided to explain the drawbacks of Elliott and Müller's (2007) method.

JEL Classification: C22

Keywords: Structural change; Confidence interval; Serially correlated errors; Dynamic regression models; Bootstrap; Inverted likelihood ratio.

*Department of Economics, Boston University, 270 Bay State Rd., Boston, MA 02215 (seongch@bu.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 changes with unknown break dates (see, Perron, 2006, for a detailed review). In this paper, we consider the problem of constructing confidence intervals for the break date in a linear regression model. Important early contributions about the limit distribution of the estimate of the break date, from which confidence intervals can be obtained, include Bai (1994) for a change in mean in a linear process and Bai (1997) for a one time change in some coefficients in a linear regression model. These results have been extended to multiple structural changes by Bai and Perron (1998, 2003) for the linear regression model, Perron and Qu (2006) for the case with restrictions on the parameters (and relaxations of the conditions on the regressors and errors), Perron and Yamamoto (2013a) for the case of linear models with endogenous regressors estimated by two-stage least squares and Perron and Yamamoto (2013b) for linear models estimated via band spectral regression. In all cases, the limit distributions of the estimates of the break dates have a common structure and are obtained using an asymptotic framework whereby the magnitude of the change shrinks as the sample size increases. The limiting distribution is non-standard, but quantiles can be obtained numerically. Simulation results presented by Bai and Perron (2006) show that the coverage rate is adequate for moderate to large breaks. Stock and Watson (2002) applied such a procedure to construct confidence intervals of the break date for a change in parameters in simple regression models for various economic time series.

Recently, Elliott and Müller (2007) considered constructing confidence intervals for the date of a single structural break in linear time series regressions, focusing on breaks with small magnitude. They criticized Bai's (1997) approach on the basis that the empirical coverage rates of the confidence intervals obtained from the asymptotic distribution of the estimate of the break date are below the nominal rates when the magnitude of the break is small. They suggested constructing confidence intervals by inverting a test statistic, labelled \hat{U}_T and related to Nyblom (1989) test, that is locally invariant to the magnitude of break. They showed, via simulations, that the resulting confidence intervals have exact coverage rates very close to the nominal level, for a variety of models including a change in unconditional variance, serial correlation and/or heteroskedasticity in the errors.

More recently, Eo and Morley (2013) generalized Siegmund's (1988) method to a system of multivariate regressions, which is based on inverting the Likelihood Ratio test to obtain a confidence set, labelled ILR. Using results from Qu and Perron (2007), they considered the

problem in the context of a system of multivariate equations. They advocated this approach based on the fact that it provides confidence intervals with smallest length compared to other methods.

Another possible avenue is to use some bootstrap procedures. When the errors are independent and identically distributed, a standard residual bootstrap procedure is possible; when the errors are heteroskedastic, the wild bootstrap approach of Liu (1988) can be applied; and with serially correlated errors, the sieve bootstrap procedure analyzed by Chang and Perron (2013) is applicable.

The goal of this paper is to compare the relative merits of these various procedures to form confidence intervals in terms of exact coverage rates and lengths. To do so, we resort to extensive simulation experiments involving models with a wide variety of features: change in mean, change in the coefficient of a stationary regressor, heteroskedastic errors, serially correlated errors and models with a lagged dependent variable. Our setup builds upon the simulation design of Elliott and Müller (2007) but extends it in several dimensions. Without loss of generality, we consider models with a single break.

Our findings can be summarized as follows. On the basis of achieving an exact coverage rate that is closest to the nominal level, Elliott and Müller's (2007) approach is by far the best one. In all cases, the exact coverage rate is indistinguishable from the nominal level. For small breaks, Bai's method and the ILR can exhibit liberal distortions, exact coverage rates below nominal level. They can also be conservative. The bootstrap procedure are never liberal but can also be conservative. However, this superior performance of Elliott and Müller's (2007) approach in terms of coverage rates comes at very high costs in terms of the length of the confidence intervals. When the errors are serially correlated and dealing with a change in intercept or a change in the coefficient of a stationary regressor with a high signal to noise ratio, the length of the confidence interval increases and approaches the whole sample as the magnitude of the change increases. The same problem occurs in models with a lagged dependent variable, a common case in practice. This drawback is not present for the other methods.

The procedure that strikes the best balance between coverage rate and length is the bootstrap. The ILR and Bai's method do have some liberal size distortions but mostly for small breaks, in which case the lengths are relatively large. The issue then is whether, for all practical purposes, it matters if these procedures state that the uncertainty about the location of the break date is, say, 50% instead of the correct 70% given the liberal size distortions. In either case, the answer is the same from a practical perspective, namely that

the data are not informative about the location of the break and we should not view the estimate as reliable. If one adopts this view, the performances of the ILR, Bai and bootstrap methods are comparable.

For Elliott and Müller's (2007) approach, we show theoretically why the length of the confidence interval approaches the whole sample as the break magnitude increases when dealing with a regression with potentially serially correlated errors or when dealing with a regression with a lagged dependent variable included as regressor. Given that the simulation results are case specific, the theoretical results presented make clear the exact features of the data generating process that give rise to this problem.

The main problem in Elliott and Müller's (2007) approach is the adoption of a local asymptotic framework in which the test being inverted is locally invariant to the magnitude of the break. In the literature, such a framework is prevalent. In finite samples, however, it has been shown that it cannot be a reliable guide under various circumstances. The problem here is related to some earlier studies. Deng and Perron (2008) studied the power functions of the CUSUM and CUSUM of squares tests and showed that the relative properties of those two tests can be different from what the local asymptotic framework used by Ploberger and Krämer (1990) suggested when the errors are serially correlated or a lagged dependent variable is included as a regressor in the model. Kim and Perron (2009) considered the local asymptotic framework used by Andrews and Ploberger (1994). They compared the asymptotic relative efficiency of the LM, Wald and LR based statistics using the criterion of the relative approximate Bahadur slopes of the tests. They showed LM-based procedures to be inefficient. Perron and Yamamoto (2012) considered the so-called optimal $q\hat{L}L$ test of Elliott and Müller's (2006) for general parameter variations in which a local asymptotic framework is also adopted. The power function of this test goes to zero as the magnitude of the break increases under the same circumstances. It was also shown that the sup-Wald test, though not optimal, has better power, unless the magnitude of the break is very small in which case the differences in power are very minor.

The structure of the paper is as follows. Section 2 lays out the model under consideration and reviews the various procedures to construct the confidence intervals. Section 3 presents the results of the simulation experiments. Section 4 provides theoretical results about the properties of Elliott and Müller's (2007) approach. Section 5 offers brief conclusions and an appendix contains technical derivations.

2 The model and procedures

We consider a linear regression model with a single structural break at T_1^0

$$y_t = x_t' \beta_t + z_t' \gamma + u_t, \quad (1)$$

$$\beta_t = \beta + \delta \mathbf{1}_{t > T_1^0} \quad (2)$$

for $t = 1, \dots, T$ where $\mathbf{1}_A$ denotes the indicator function for the event A . Here, y_t is the observed dependent variable, x_t ($p \times 1$) and z_t ($q \times 1$) are vectors of covariates, and (β, δ, γ) are the corresponding vectors of coefficients; u_t is the disturbance. As usual, we assume that each regime is a fixed proportion of the sample size so that $T_1^0 = [T\lambda_0]$ for some $\lambda_0 \in (0, 1)$; $[\cdot]$ is the greatest smaller integer function. Note that we allow for a partial structural change model since γ is not subject to change. A pure structural change model is a special case with $q = 0$.

The issue of interest is to form a confidence interval for the break date T_1^0 . Below, we review several procedures that have been proposed.

2.1 Bai's (1997) approach

Bai (1997) considered estimating the break date using

$$\hat{T}_1 = \arg \min_{T_1} SSR(T_1) \quad (3)$$

where $SSR(T_1)$ is the sum of squared residuals from the unrestricted regression (1) and (2) evaluated at the candidate break date T_1 . Using T_1 as the break date, we can write (1) as

$$y_t = x_1(T_1)'_t \beta + x_2(T_1)'_t \psi + z_t' \gamma + u_t \quad (4)$$

where $x_1(T_1)_t = x_t$ if $t \leq T_1$ and 0 otherwise and $x_2(T_1)_t = x_t$ if $t > T_1$ and 0 otherwise. Here, $\psi = \beta + \delta$. In matrix notations,

$$Y = [\bar{X}, Z]\Gamma + U$$

where $Y = (y_1, \dots, y_T)'$, $Z = (z_1, \dots, z_T)'$, and $U = (u_1, \dots, u_T)'$. \bar{X} is the matrix which diagonally partitions X at T_1 , i.e., $\bar{X} = \text{diag}(X_1, X_2)$ with $X_1 = (x_1, \dots, x_{T_1})'$ and $X_2 = (x_{T_1+1}, \dots, x_T)'$, and $\Gamma = (\Gamma_1', \gamma')' = (\beta', \psi', \gamma')'$. Let $T_1 = [T\lambda]$ with $\lambda \in \Lambda_\epsilon = (\epsilon_1, 1 - \epsilon_2)$. Define

$$F(T_1) = \left(\frac{T - 2p - q}{p} \right) \frac{\hat{\Gamma}_1' H' (H(\bar{X}' M_Z \bar{X})^{-1} H')^{-1} H \hat{\Gamma}_1}{SSR(T_1)}$$

where $\hat{\Gamma}_1 = (\hat{\beta}', \hat{\psi}')'$ is the OLS estimate from (4), H is the conventional matrix such that $(H\Gamma_1)' = (\psi' - \beta') = \delta'$ and $M_Z = I - Z(Z'Z)^{-1}Z'$. In a pure structural break model, M_Z reduces to a $T \times T$ identity matrix. With a single structural break,

$$\hat{T}_1 = \arg \max_{T_1 \in \Lambda} F(T_1).$$

Amemiya (1985) and Bai (1997) showed that the break date that maximizes the Wald test is the same as the break date that minimizes the sum of squared residuals from the unrestricted regression model defined in (3). As noted in Bai (1997), the estimator of the break fraction $\hat{\lambda}$ is T -consistent even with serially correlated errors.

In Bai (1997), the limiting distribution of the estimate of the break date was derived under various assumptions. If the magnitude of break is fixed, the limiting distribution depends on the exact distributions of both the regressors and the errors which are unknown in general. To avoid this problem, a common approach is to use an asymptotic framework with shrinking magnitudes of shifts. To describe the limit distribution that applies in this case, we need to define some notations. For the true structural break date T_1^0 , $Q = p \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T E(x_t x_t')$, $Q_1 = p \lim_{T \rightarrow \infty} (T_1^0)^{-1} \sum_{t=1}^{T_1^0} E(x_t x_t')$, $Q_2 = p \lim_{T \rightarrow \infty} (T - T_1^0)^{-1} \sum_{t=T_1^0+1}^T E(x_t x_t')$,

$$\begin{aligned} \omega^2 &= \lim_{T \rightarrow \infty} T^{-1} E \left[\sum_{t=1}^T x_t' u_t \right] \left[\sum_{t=1}^T x_t' u_t \right]', \\ \omega_1^2 &= \lim_{T \rightarrow \infty} (T_1^0)^{-1} E \left[\sum_{t=1}^{T_1^0} x_t' u_t \right] \left[\sum_{t=1}^{T_1^0} x_t' u_t \right]', \\ \omega_2^2 &= \lim_{T \rightarrow \infty} (T - T_1^0)^{-1} E \left[\sum_{t=T_1^0+1}^T x_t' u_t \right] \left[\sum_{t=T_1^0+1}^T x_t' u_t \right]'. \end{aligned}$$

Let \Rightarrow denote weak convergence under the Skorohod topology. Under some conditions, Bai (1997) showed that the limit distribution of the estimate of the break date is given by:

$$\frac{(\delta' Q_1 \delta)^2}{\delta' \omega_1^2 \delta} (\hat{T}_1 - T_1^0) \Rightarrow \arg \max_{s \in \mathbb{R}} V(s) \quad (5)$$

where $V(s) = W_1(-s) - |s|/2$ if $s \leq 0$ and $V(s) = \sqrt{\phi} W_2(s) - \xi |s|/2$ if $s > 0$, where $\xi = \delta' Q_2 \delta / \delta' Q_1 \delta$ and $\phi = \delta' \omega_2^2 \delta / \delta' \omega_1^2 \delta$. Also, $W_i(s)$, $i = 1, 2$, are two independent standard Wiener processes defined on $[0, \infty)$, starting at the origin when $s = 0$. The cumulative distribution function of $\arg \max V(s)$ is derived in Bai (1997) (see Appendix B therein).

Since δ , Q_i , and ω_i^2 for $i = 1, 2$ are unknown, consistent estimates are needed. These are given by $\hat{\delta} = \hat{\psi} - \hat{\beta}$, $\hat{Q}_1 = \hat{T}_1^{-1} \sum_{t=1}^{\hat{T}_1} x_t x_t'$, $\hat{Q}_2 = (T - \hat{T}_1)^{-1} \sum_{t=\hat{T}_1+1}^T x_t x_t'$ and an estimate of ω_i can be constructed using a HAC estimator applied to the vector $\{x_t \hat{u}_t\}$ and using data over segment i only.

As a special case, suppose that $\{x_t, u_t\}$ is second-order stationary for the whole sample and the errors are uncorrelated, then $Q_1 = Q_2 = Q$, $\omega_1^2 = \omega_2^2 = \sigma^2 Q$ and

$$\frac{\delta' Q \delta}{\sigma^2} (\hat{T}_1 - T_1^0) \Rightarrow \arg \max_{s \in \mathbb{R}} |W(s) - |s|/2|.$$

This can be evaluated using the estimates $\hat{Q} = T^{-1} \sum_{t=1}^T x_t x_t'$ and $\hat{\sigma}^2 = T^{-1} \sum_{t=1}^T \hat{u}_t^2$. A $100(1 - a)\%$ confidence interval is constructed as

$$\left[\hat{T}_1 - \left[\zeta / \hat{L} \right] - 1, \hat{T}_1 + \left[\zeta / \hat{L} \right] + 1 \right] \quad (6)$$

where ζ is the $(1 - a/2)$ th quantile of the random variable $\arg \max |W(s) - |s|/2|$ and $\hat{L} = \hat{\delta}' \hat{Q} \hat{\delta} / \hat{\sigma}^2$; $[\zeta / \hat{L}]$ is the integer part of ζ / \hat{L} . The asymptotic distributions are also available under various assumptions on both the regressors and the disturbances, see Bai (1997) and Bai and Perron (1998, 2003, 2006). When the errors are serially correlated and the regressors are identically distributed across segments, $\xi = \phi = 1$, and

$$\frac{(\delta' Q \delta)^2}{\delta' \omega^2 \delta} (\hat{T}_1 - T_1^0) \Rightarrow \arg \max_{s \in \mathbb{R}} |W(s) - |s|/2|.$$

where $Q = Q_1 = Q_2$ and $\omega = \omega_1 = \omega_2$. An estimate of ω can be constructed using a HAC estimator applied to the vector $\{x_t \hat{u}_t\}$ using data over the whole sample. On the other hand, if the errors are uncorrelated but heteroskedastic and the regressors are identically distributed across segments, then

$$\frac{\delta' Q \delta}{\sigma_1^2} (\hat{T}_1 - T_1^0) \Rightarrow \arg \max_{s \in \mathbb{R}} |W(s) - |s|/2|,$$

where σ_1^2 can be estimated using $\hat{\sigma}_1^2 = \hat{T}_1^{-1} \sum_{t=1}^{\hat{T}_1} \hat{u}_t^2$.

2.2 Elliott and Müller's (2007) approach

The method proposed by Elliott and Müller (2007) is based on inverting the following test statistic:

$$\begin{aligned} U_T(T_1) = & T_1^{-2} \sum_{t=1}^{T_1} \left(\sum_{s=1}^t v_s \right)' \omega_1^{-2} \left(\sum_{s=1}^t v_s \right) \\ & + (T - T_1)^{-2} \sum_{t=T_1+1}^T \left(\sum_{s=T_1+1}^t v_s \right)' \omega_2^{-2} \left(\sum_{s=T_1+1}^t v_s \right) \end{aligned} \quad (7)$$

where $v_t = x_t \hat{u}_t$ with \hat{u}_t the OLS residuals from the regression (4). The procedure for constructing the confidence interval of the break date is as follows. For every $T_1 \in \{2p + 1, \dots, T - 2p - 1\}$, perform the following steps:

- Step 1: Obtain the OLS residuals \hat{u}_t from the regression (4).
- Step 2: Construct a consistent estimate of ω_1^2 and ω_2^2 , the $p \times p$ long-run covariance matrices of $v_t = x_t \hat{u}_t$, for each sub-samples defined by T_1 . For serially correlated errors, the automatic bandwidth estimator of Andrews (1991) or Andrews and Monahan (1992) is recommended.
- Step 3: Compute $\hat{U}_T(T_1)$ as in equation (7) with ω_1^2 and ω_2^2 replaced by $\hat{\omega}_1^2$ and $\hat{\omega}_2^2$, respectively.
- Step 4: Test the null hypothesis $H_0 : T_1 = T_1^0$. T_1 is included in the $100(1 - a)\%$ confidence set if $\hat{U}_T(T_1) < cv_a$ and is excluded otherwise, where cv_a is the asymptotic critical value of \hat{U}_T at the significance level a . The relevant limit distribution is given by

$$\hat{U}_T(T_1^0) \Rightarrow \int_0^1 B(s)' B(s) ds$$

where $B(s)$ is a $2p \times 1$ vector standard Brownian bridge.

2.3 Bootstrap Methods

Bootstrap methods are popular alternatives to approximate the exact distribution of estimators and can be used to construct confidence intervals. We consider three variants depending on the regressors and assumptions about the errors: 1) a simple residual-based method applicable when the errors are assumed to be *i.i.d.*; 2) a sieve bootstrap applicable when the errors are potentially serially correlated; 3) a wild bootstrap method to account for heteroskedasticity in the errors. The data-generating process is again

$$y_t = x_t' \beta + x_t' \delta \mathbf{1}_{t > T_1^0} + u_t. \quad (8)$$

for $t = 1, \dots, T$.

1) The case with *i.i.d.* errors. Consider estimating (8) by OLS with estimated residuals $\hat{u}_t = y_t - x_t' \hat{\beta} - x_t' \hat{\delta} \mathbf{1}_{t > \hat{T}_1}$ for $t = 1, \dots, T$. Define $\bar{\hat{u}} = T^{-1} \sum_{t=1}^T \hat{u}_t$ and construct a set $\hat{U} = \{\hat{u}_1 - \bar{\hat{u}}, \hat{u}_2 - \bar{\hat{u}}, \dots, \hat{u}_T - \bar{\hat{u}}\}$. Draw a random sample from \hat{U} with replacement and label

it as $\hat{U}_{(b)}^* = \{\hat{u}_1^*, \dots, \hat{u}_T^*\}$ for $b = 1, \dots, B$, where B is the number of bootstrap replications. Under the assumption that the errors are *i.i.d.*, we can construct a new process $\{y_t^*\}$ as

$$y_t^* = x_t' \hat{\beta} + x_t' \hat{\delta} \mathbf{1}_{t > \hat{T}_1} + \hat{u}_t^*$$

for $t = 1, \dots, T$ and obtain the OLS estimates $\hat{\beta}_{(b)}^*$, $\hat{\delta}_{(b)}^*$, and $\hat{T}_{1(b)}^*$ associated with each bootstrap sample $\hat{U}_{(b)}^*$. Note that we allow for the change to estimate the residuals used to generate the bootstrap samples. This allows replicating the distribution under the null hypothesis of no change whether a change is present or not. If the change is not allowed and one occurs, the estimated residuals will be contaminated and the procedure will not replicate correctly the distribution under the null hypothesis.

This approach is also valid in the case with a lagged dependent variable included as a regressor. We again re-sample \hat{u}_t^* for \hat{U}^* and construct y_t^* recursively, using

$$y_t^* = \hat{\alpha} y_{t-1}^* + x_t' \hat{\beta} + x_t' \hat{\delta} \mathbf{1}_{t > \hat{T}_1} + \hat{u}_t^*$$

for $t = 1, \dots, T$ and $y_0^* = y_0$. For each bootstrap sample, we estimate the break date $\hat{T}_{1(b)}^*$, for $b = 1, \dots, B$.

2) The case with serially correlated errors. To account for potential serial correlation in the errors, whose nature is assumed to remain constant throughout the sample, we consider the sieve bootstrap as suggested by Bühlmann (1997). Suppose that the error term u_t is generated by the linear process

$$u_t = \psi(L) \epsilon_t \tag{9}$$

where L is the usual lag operator and $\psi(z) = \sum_{k=0}^{\infty} \psi_k z^k$. By Wold's theorem, a one-sided infinite order MA representation of the form (9) holds if u_t is a real-valued stationary process with expectation $Eu_t = \mu_u$, and is purely stochastic; see Brockwell and Davis (1989). In this case, the process is invertible and we can express u_t as a one-sided infinite-order autoregression

$$\epsilon_t = \phi(L) u_t \tag{10}$$

where $\phi(z) = \sum_{k=0}^{\infty} \phi_k z^k$. We approximate the process by a finite autoregression of order p , i.e.,

$$u_t = \phi_{1,T} u_{t-1} + \dots + \phi_{p,T} u_{t-p} + \epsilon_t \tag{11}$$

We estimate $\phi = (\phi_{1,T}, \dots, \phi_{p,T})'$ using the Yule-Walker equations since this method yields a stationary process by construction. These are defined by $\Gamma_p \phi = \gamma_p$ and $\sigma^2 = \gamma(0) - \phi' \gamma_p$,

where $\Gamma_p = [\gamma(i-j)]_{i,j=1}^p$ and $\gamma_p = (\gamma(1), \dots, \gamma(p))'$. Hence, we can estimate $\hat{\phi}$ and $\hat{\sigma}^2$ from $\hat{\gamma}(1), \dots, \hat{\gamma}(p)$, defined by $\hat{\gamma}(s) = T^{-1} \sum_{t=s+1}^T \hat{u}_t \hat{u}_{t-s}$. The sieve bootstrap will be valid under the following conditions.

- **Assumption SB1** $u_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$, $\psi_0 = 1$ ($t \in \mathbb{Z}$) where $\{\epsilon_t\}_{t \in \mathbb{Z}}$ is an *i.i.d.* sequence with $E(\epsilon_t) = 0$, $E|\epsilon_t|^r < \infty$ for some $r > 4$.
- **Assumption SB2** $\psi(z) \neq 0$ for all $|z| \leq 1$, and $\sum_{j=0}^{\infty} |j|^s |\psi_j| < \infty$ for some $s \geq 1$.
- **Assumption SB3** Let $p = p_T \rightarrow \infty$ and $p_T = o((T/\log T)^{1/2})$ as $T \rightarrow \infty$ and $\hat{\phi}_p = (\hat{\phi}_{1,T}, \dots, \hat{\phi}_{p,T})'$ satisfy the empirical Yule-Walker equations

$$\hat{\Gamma}_p \hat{\phi}_p = \hat{\gamma}_p$$

where $\hat{\Gamma}_p = [\hat{\gamma}(i-j)]_{i,j=1}^p$ and $\hat{\gamma}_p = (\hat{\gamma}(1), \dots, \hat{\gamma}(p))'$.

We use the Akaike Information Criterion (AIC) to select the order of the autoregression p_T since it is compatible with Assumption SB3. We can then obtain the estimated residuals as

$$\hat{\epsilon}_{t,T} = \sum_{j=0}^{p_T} \hat{\phi}_{j,T} \hat{u}_{t-j}$$

where $\hat{\phi}_{0,T} = 1$. We center the estimated residuals so that their sample mean is zero:

$$\tilde{\epsilon}_{t,T} = \hat{\epsilon}_{t,T} - (T-p)^{-1} \sum_{t=p+1}^T \hat{\epsilon}_{t,T}$$

and denote the empirical CDF of $\{\tilde{\epsilon}_{t,T}\}_{t=p+1}^T$ by

$$F^*(z) = (T-p)^{-1} \sum_{t=p+1}^T \mathbf{1}_{\tilde{\epsilon}_{t,T} \leq z} \quad (12)$$

We can re-sample, for any $t \in \mathbb{Z}$, ϵ_t^* *i.i.d.* from F^* , and define \hat{u}_t^* by recursion,

$$\sum_{j=0}^{p_T} \hat{\phi}_{j,T} \hat{u}_{t-j}^* = \epsilon_t^* \quad (13)$$

with approximately chosen p -initial values of u_t^* . To that effect, we initially set the initial values to zero and construct the $AR(p_T)$ process from (13) for a sufficiently long period to

ensure stationarity of the process. We then discard the initial values to have a sample of size $T + p_T$. Then, y_t^* is constructed as

$$y_t^* = x_t' \hat{\beta} + x_t' \hat{\delta} \mathbf{1}_{t > \hat{T}_1} + \hat{u}_t^*$$

and can estimate the break date, $\hat{T}_{1(b)}^*$ for each replications $b = 1, \dots, B$.

3) The case with heteroskedastic errors. To account for potential heteroskedasticity, we adopt the wild bootstrap method of Liu (1988). The bootstrap sequence $\{y_t^*; t = 1, \dots, T\}$ is generated by:

$$y_t^* = x_t' \hat{\beta} + x_t' \hat{\delta} \mathbf{1}_{t > \hat{T}_1} + \hat{u}_t^*,$$

where $\hat{u}_t^* = f_t(\hat{u}_t) \epsilon_t$, with

$$\epsilon_t = \begin{cases} \frac{-(\sqrt{5}-1)}{2} & \text{with prob. } p = (\sqrt{5} + 1)/(2\sqrt{5}) \\ \frac{(\sqrt{5}+1)}{2} & \text{with prob. } 1 - p, \end{cases}$$

so that ϵ_t is a random variable with mean zero and variance one. Also, we set $f_t(\hat{u}_t) = (T/(T - 2))^{1/2} \hat{u}_t$. Again, we estimate the break date $\hat{T}_{1(b)}$ from the bootstrap samples $\{y_t^*, x_t\}$ for each replication $b = 1, \dots, B$. For details about the wild bootstrap, see Liu (1988) and Davidson and Flachaire (2008).

In all cases the “percentile bootstrap confidence interval” is constructed as follows. First, sort the estimated break dates $\hat{T}_{1(b)}^*$ from every bootstrap sample in ascending order (the estimate \hat{T}_1 should be included in the sorted set). Denote the quantiles of interest, $a/2$ and $(1 - a/2)$, for equal-tailed probability intervals, by q_L and q_H . The $100(1 - a)\%$ percentile bootstrap confidence interval (PB) is defined as $PB \equiv [q_L, q_H]$.

2.4 Likelihood-based Method

Siegmund (1988) suggested a likelihood-based method to construct a confidence set for a structural break date. He considered a change in mean model with independent normal observations. Recently, Eo and Morley (2013) generalized Siegmund’s (1988) method to a system of multivariate regressions. We review the so-called Inverted Likelihood Ratio (ILR) confidence set for the case of a linear model with no serial correlation in the errors but allowing the variance to change across regimes. The reader is referred to Eo and Morley (2013) for details about the case with serial correlation in the errors.

The parameters are estimated by restricted quasi-maximum likelihood based on the assumption of *i.i.d.* errors; see Qu and Perron (2007). The quasi-likelihood function is

$$L_T(T_1, \beta, \delta, \sigma_{u,1}, \sigma_{u,2}) = \prod_{t=1}^{T_1} f(y_t|x_t; \beta, \delta, \sigma_{u,1}) + \prod_{t=T_1+1}^T f(y_t|x_t; \beta, \delta, \sigma_{u,2})$$

where for $j = 1, 2$, where

$$f(y_t|x_t; \beta, \delta, \sigma_{u,j}) = \frac{1}{\sqrt{2\pi}\sigma_{u,j}} \exp\left(-\frac{1}{2\sigma_{u,j}^2}(y_t - x_t'\beta - x_t'\mathbf{1}_{t>T_1}\delta)^2\right).$$

Then, the likelihood ratio test statistic is given by

$$\begin{aligned} \ln LR_T(T_1, \beta, \delta, \sigma_{u,1}, \sigma_{u,2}) &= \left(\sum_{t=1}^{T_1} \ln f(y_t|x_t; \beta, \delta, \sigma_{u,1}) + \sum_{t=T_1+1}^T \ln f(y_t|x_t; \beta, \delta, \sigma_{u,2}) \right) \\ &\quad - \left(\sum_{t=1}^{T_1^0} \ln f(y_t|x_t; \beta, \delta, \sigma_{u,1}) + \sum_{t=T_1^0+1}^T \ln f(y_t|x_t; \beta, \delta, \sigma_{u,2}) \right) \\ &= lr(T_1 - T_1^0) \end{aligned}$$

where, letting $r = T_1 - T_1^0$,

$$lr(r) = 0 \quad \text{for } r = 0,$$

$$lr(r) = -r(\ln \sigma_{u,2} - \ln \sigma_{u,1}) - \frac{1}{2} \sum_{t=T_1+1}^{T_1^0} \{(y_t - x_t'\beta - x_t'\delta)^2 \sigma_{u,2}^{-2} - (y_t - x_t')^2 \sigma_{u,1}^{-2}\} \quad \text{for } r < 0,$$

$$lr(r) = -r(\ln \sigma_{u,1} - \ln \sigma_{u,2}) - \frac{1}{2} \sum_{t=T_1+1}^{T_1^0} [(y_t - x_t')^2 \sigma_{u,1}^{-2} - (y_t - x_t'\beta - x_t'\delta)^2 \sigma_{u,2}^{-2}] \quad \text{for } r > 0.$$

When there is a structural break in either the conditional mean or the unconditional variance, under some assumptions,

$$lr(\hat{T}_1 - T_1^0) \Rightarrow \max_{\nu} \left\{ -\frac{1}{2}|\nu| + W(\nu) \right\} \quad \text{for } \nu \in (-\infty, \infty)$$

where $W(\cdot)$ is a standard Wiener process; see Corollaries 1 and 2 in Eo and Morley (2013). The ILR confidence set can be obtained as $ILR = \{T_1 | lr(\hat{T}_1 - T_1) < \kappa\}$ where κ denotes the critical value of a $100(1-a)\%$ ILR confidence set. Bhattacharya and Brockwell (1976) showed that $\Xi_1 = \sup_{\nu \leq 0} \{-|\nu|/2 + W(\nu)\}$ and $\Xi_2 = \sup_{\nu > 0} \{-|\nu|/2 + W(\nu)\}$ are *i.i.d.* exponential random variables, and $P(\Xi_1 \leq z) = P(\Xi_2 \leq z) = 1 - \exp(-z)$, $z > 0$. Hence,

$$\begin{aligned} P\left(\sup_{\nu} \{-|\nu|/2 + W(\nu)\} \leq \kappa\right) &= P(\max\{\Xi_1, \Xi_2\} \leq \kappa) = P(\Xi_1 \leq \kappa)P(\Xi_2 \leq \kappa) \\ &= (1 - \exp(-\kappa))^2 = 1 - a \end{aligned}$$

and the critical value of the likelihood ratio test is given by $\kappa = -\ln(1 - (1 - a)^{1/2})$.

3 Simulation Experiments

In this section, we present simulation results in a non-local perspective on the coverage rate and the length of the confidence interval for the various procedures. All models in the simulations reported in Elliott and Müller (2007) are revisited and expended. For the regression models without a lagged dependent variable, the data generating process (DGP) is specified as:

$$y_t = x_t' \beta + x_t' \delta \mathbf{1}_{t > T_1^0} + z_t' \gamma + u_t \quad (14)$$

for $t = 1, \dots, T$ with $T = 100$. The true break date is $T_1^0 = [T\lambda_0]$ and $\lambda_0 = 0.5$. The magnitude of a change, δ , is defined by $\delta \equiv dT^{-1/2}$ where $d \in \{4, 8, \dots, 50\}$, or equivalently $\delta \in \{0.4, 0.8, \dots, 5\}$. Without loss of generality, we set $\beta = 0$. The different DGPs considered are the following:

- M1: a change in mean, $x_t = 1$, $z_t = 0$, and $u_t \sim i.i.d.N(0, 1)$;
- M2: same as M1, but with a change in unconditional variance that quadruples at T_1^0 ;
- M3: same as M1, but with AR(1) errors, $u_t = 0.3u_{t-1} + \epsilon_t$, $\epsilon_t \sim i.i.d.N(0, 0.49)$;
- M4: same as M1, but with MA(1) errors, $u_t = \epsilon_t - 0.3\epsilon_{t-1}$, $\epsilon_t \sim i.i.d.N(0, 2.04)$;
- M5: x_t is a stationary Gaussian AR(1) process, i.e., $x_t = 0.5x_{t-1} + \xi_t$, $\xi_t \sim i.i.d.N(0, 0.75)$, and $z_t = 1$, and $\{u_t\} \sim i.i.d.N(0, 1)$ independent of $\{x_t\}$;
- M6: Same as (M5), but with heteroskedastic errors such that $u_t = \epsilon_t |x_t|$, where $\epsilon_t \sim i.i.d.N(0, 0.333)$ and independent of x_t ;
- M7: Random regressor, $x_t = \mu + 0.5x_{t-1} + \xi_t$ with $\mu = 5$ and $\xi_t \sim i.i.d.N(0, 0.75)$, and AR(1) errors $u_t = 0.3u_{t-1} + \epsilon_t$, $\epsilon_t \sim i.i.d.N(0, 0.49)$;
- M8: Random regressor, $x_t = \mu + 0.5x_{t-1} + \xi_t$ with $\mu = 5$ and $\xi_t \sim i.i.d.N(0, 0.75)$, and MA(1) errors $u_t = \epsilon_t - 0.3\epsilon_{t-1}$, $\epsilon_t \sim i.i.d.N(0, 2.04)$.

We also consider a dynamic regression model, given by

$$y_t = \alpha y_{t-1} + \delta \mathbf{1}_{t > T_1^0} + \epsilon_t, \quad \epsilon_t \sim i.i.d.N(0, 1),$$

where $y_0 = 0$. Model D1 sets $\alpha = 0.8$ and for Model D2 $\alpha = 0$, the latter corresponding to the case in which an irrelevant lagged dependent variable is included.

We considered two versions of \hat{U}_T : $\hat{U}_{T.eq}$ which imposes the long-run variances pre- and post- break to be the same and $\hat{U}_{T.neq}$ which allows $\omega_1^2 \neq \omega_2^2$. Since the results are qualitatively similar, we only report results for $\hat{U}_{T.neq}$.

The results for 95% coverage rates are presented in Figures 1 for the various DGPs. Consider first the coverage rates for the various methods. The Elliott and Müller's (2007) (EM) method yields, overall, exact coverage rates that are the closest to 95% across all methods. It is never liberal, though it can be conservative in some cases (e.g., M3, M4). The bootstrap methods also have exact coverage rates close to 95% but the extent to which they can be conservative is greater, e.g., M1, M3, M4, M8 and M5 through M7, D1, D2 for large breaks. The method of Bai and the Inverted Likelihood Ratio (ILR) exhibit liberal distortions (coverage rates below 95%) in some cases when the break is small. For Bai's method, this occurs for M1, M2, M5, M6, D1 and D2. The liberal distortions can remain to some extent even for large break in the case of M5 and D1. The ILR also has cases with liberal coverage rates (M1, M3, M5, M6 for very small breaks, D1 and D2). The distortions are especially severe for M3 with small breaks. It can also be conservative (M4, M8, and most cases for large breaks). Overall, for small breaks, the ILR delivers more liberal coverage rates compared to Bai (M1, M3, M5, M7, D1, D2). Bai's method is more liberal than ILR for small breaks with M2, M6 and for large break with M5 and D1.

Consider now the length of the confidence intervals. For large breaks, the differences between those delivered by Bai, ILR and the bootstrap are minor. For small breaks, the ILR has more often smaller length (M1, especially M2, M3, D2) but Bai's method yields smaller lengths in some cases (M4, M6). The bootstrap method delivers confidence intervals with lengths that are, in general, a close second relative to the best of ILR and Bai.

The most striking feature of the results is the length of the confidence intervals delivered by EM. They are, in almost all cases and for any values of the magnitude of the break, the widest amongst all methods. More importantly, in many cases the average length increases as the magnitude of the break increases and even covers the whole sample for large breaks. This occurs for M3 (change in mean model with serial correlation in the errors), M7 and M8 (models with serial correlation and a random regressors such that the signal to noise ratio is high), and D1 and D2 (models with a lagged dependent included). It also occurs for M4 with larger values of δ from unreported simulations. The length can be more than half the whole sample even for moderate breaks in the case of M7. For Model D1, with an autoregressive

coefficient taking value 0.8, common in practice, the length of the confidence interval is the whole sample for any value of the magnitude of the break.

On the basis of the simulations, which procedure is to be recommended? If one insists on having a coverage rate that is closest to the nominal level, then EM's approach is clearly the preferred one. But the price in terms of the average lengths that it delivers is way too high. It always has the longest confidence interval, which in important practical cases can increase to be the whole sample as the magnitude of the break increases. The procedure that strikes the best balance between coverage rate and length is the bootstrap. It is, on the other hand, obviously more computationally intensive. The ILR and Bai's method do have some liberal size distortions but most often for small breaks only, in which case the lengths are relatively large given the fact that there is little information in the data. The issue then is whether, for all practical purposes, it matters if these procedures state that the uncertainty about the location of the break date is, say, 50% instead of the correct 70% given the liberal size distortions. In either case, the answer is the same from a practical perspective, namely that the data are not informative about the location of the break and we should not view the estimate as reliable. If one adopts this view, the performances of the ILR, Bai and bootstrap methods are comparable, though for small breaks the bootstrap achieves a better balance between coverage rates and lengths of the confidence intervals.

It remains to understand why EM's method performs so badly, in important cases, in terms the length of the confidence interval. This is addressed theoretically in the next section.

4 Theoretical Results about Elliott and Müller's (2007) approach

We now provide theoretical explanations for some of the simulation results pertaining to the properties of Elliott and Müller's (2007) approach to construct confidence intervals. Of interest is to show theoretically why the length of the confidence interval approaches the whole sample as the break magnitude increases when dealing with a regression with potentially serially correlated errors or when dealing with a regression with a lagged dependent variable included as regressor. Since the simulation results are case specific, the theoretical results will make clear the exact features of the DGP that give rise to this problem.

4.1 Static Regression with Serially Correlated Errors

The data generating process (DGP) is defined as follows:

$$y_t = x_t' \beta_t + z_t' \gamma + u_t; \quad \beta_t = \beta + \delta \mathbf{1}_{t > T_1^0} \quad (15)$$

where $T_1^0 = [T\lambda_0]$ for some $\lambda_0 \in (0, 1)$. We impose the following assumptions on the errors and the regressors:

- **Assumption A1:** The errors u_t are stationary and ergodic with $E(u_t|\mathcal{F}_t) = 0$, $E(u_t^2|\mathcal{F}_t) = \sigma_u^2$ and $E(u_t^4) < \infty$, where \mathcal{F}_t is the σ -field generated by $\{x_{t-s}, z_{t-s}, u_{t-s-1} | s \geq 0\}$.
- **Assumption A2:** The regressors x_t satisfy $p \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T x_t = c$, some constant $p \times 1$ vector and $p \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^{[Tr]} x_t x_t' = R(r)$, where $R(r)$ is a nonsingular non-random $p \times p$ matrix (as a matter of notation, $R(1) \equiv R$).
- **Assumption A3:** The regressors z_t satisfy $\sup_{0 \leq s \leq 1} \|T^{-1/2} \sum_{t=1}^{[Ts]} z_t u_t\| = O_p(1)$ and $T^{-1} \sum_{t=1}^{[Ts]} z_t x_t' = s \Sigma_{ZX}$ uniformly in $0 \leq s \leq 1$.

Assumptions A1-A3 are standard high level assumptions in the literature. They are general enough to allow for heteroskedasticity and serial correlation in both the regressors and the errors, as well as lagged dependent variables provided the errors are a martingale difference sequence. Note that we consider the case of equal variance across segments, for simplicity as none of the results depend on it.

To account for potentially serially correlated errors, we use the following estimate of the long-run variance, based on a weighted sum of autocovariances,

$$\hat{\omega}^2 = \hat{\gamma}_0 + 2 \sum_{j=1}^{T-1} k(j, m) \hat{\gamma}_j \quad (16)$$

with $\hat{\gamma}_j = T^{-1} \sum_{t=j+1}^T (\hat{u}_t - \bar{\hat{u}})(\hat{u}_{t-j} - \bar{\hat{u}})$, where \hat{u}_t are the OLS residuals from regression (15), $\bar{\hat{u}}$ is their sample average, $k(j, m)$ is some kernel function and m is the bandwidth. In the simulations, we use the Bartlett kernel $k(j, m) = 1 - j/m$ if $j \leq m$ and 0 otherwise. Following Andrews (1991), we adopt an AR(1) approximation for u_t so that the data-dependent rule for the bandwidth is such that $m = (C(\delta)T)^{1/\theta}$ where, e.g., $\theta = 3$ for the Bartlett kernel and $C(\delta) = 4\hat{\rho}(\delta)^2/(1 - \hat{\rho}(\delta)^2)^2$ with $\hat{\rho}(\delta)$ the OLS estimate from a regression of \hat{u}_t on \hat{u}_{t-1} . In the general case, the kernel function $k(j, m)$ satisfies the condition $\sum_{j=1}^{T-1} |k(j, m)| = O(m)$. Note that using an estimate that assumes $\omega_1^2 = \omega_2^2$ is inconsequential as explained below. The same results will hold using a different estimate for each candidate sub-samples.

It is useful to define the following condition on the regressors, given that the theoretical results crucially depend on whether it holds or not.

• **Condition C1:**

$$\frac{\delta' R_1 \delta}{\delta' R_0 \delta} \rightarrow 1 \quad \text{as} \quad \|\delta\| \rightarrow \infty$$

where $R_j \equiv p \lim_{T \rightarrow \infty} T^{-1} \sum_{t=j+1}^T x_t x'_{t-j}$.

Note that C1 holds when the only regressor subject to change is a constant. It also holds approximately as the signal to noise ratio increases. The main results, proved in the appendix, are stated in the following theorem.

Theorem 1 *Let the data be generated by (15) satisfying A1-A3 and consider a candidate break date $T_1 = [T\lambda]$ with $\lambda \in (0, 1)$ such that $\lambda \neq \lambda_0$ (i) if C1 is satisfied:*

$$T^{\frac{1}{\theta}-1} \hat{U}_T(T_1) = \frac{\frac{(1-\lambda_0)^2(\lambda_0-\lambda)^2}{3(1-\lambda)^4} \delta' R' R \delta}{O_p(\|\delta\|^{\frac{4}{\theta}+2})} + o_p(1), \quad (17)$$

where $\theta = 3$ for the Bartlett kernel and 5 for the quadratic kernel and others. (ii) if C1 is not satisfied,

$$T^{\frac{1}{\theta}-1} \hat{U}_T(T_1) = \frac{\frac{(1-\lambda_0)^2(\lambda_0-\lambda)^2}{3(1-\lambda)^3} \delta' R' R \delta}{O_p(\|\delta\|^2)} + o_p(1). \quad (18)$$

This theorem is informative about the power property of \hat{U}_T and, hence, about the width of the confidence interval obtained by inverting it. The main result of interest is that the limit of \hat{U}_T will be zero as the magnitude of the break, δ , increases if condition C1 is satisfied, see equation (17). Hence, in this case the length of the confidence interval will be the whole sample. This provides a theoretical explanation of the simulation results presented in Section 3, which showed the length of the confidence interval to increase in the case of testing for a change in mean with serially correlated errors (Model M3) or in the case of testing for a change in the coefficient of a stationary regressor when the signal to noise ratio is large (Models M7 and M8).

The intuition behind this result follows the analysis in Perron (1990). For any candidate break date T_1 not equal to the true break date, the estimated residuals \hat{u}_t used to construct \hat{U}_T are contaminated by the shift. Accordingly, a change in mean will induce a bias of $\hat{\rho}(\delta)$ toward unity and more so as δ increases. In this context, this biased estimate $\hat{\rho}(\delta)$ affects the bandwidth selected via the data-dependent method. Since the bandwidth is proportional to $C(\delta) = 4\hat{\rho}(\delta)^2/(1-\hat{\rho}(\delta)^2)^2$, it gets larger as $\hat{\rho}(\delta)$ approaches one. As a result, the estimate $\hat{\omega}^2$ of the long-run variance includes more covariances that are proportional to the magnitude of the change δ , so that $\hat{\omega}^2$ is an increasing function of δ . Therefore, the value of the test \hat{U}_T

goes to zero eventually as the magnitude of the change increases. Note that the same result holds if one uses separate estimates $\hat{\omega}_1^2$ and $\hat{\omega}_2^2$ for each candidate sub-samples, since one or the other will be contaminated by the unaccounted shift.

When condition C1 does not hold, the limit of the test statistic is bounded above zero and the power in the limit will depend on various factors, see equation (18). In general, the power of the test will approach one as δ increases and the length of the confidence interval is accordingly not the whole sample, in accordance with the simulations results (e.g., Models M1, M2, M5, M6).

4.2 Dynamic Regression

Often lagged dependent variables are introduced as regressors to account for serial correlation in the dependent variable. In this section, we consider the simple case in which a single lagged dependent variable is included, so that the regression is defined as follows:

$$y_t = \alpha y_{t-1} + x_t' \beta_t + u_t, \quad \beta_t = \beta + \delta \mathbf{1}_{t > T_1^0} \quad (19)$$

for $t = 1, \dots, T$, where $y_0 = 0$ for simplicity. Note that no change in the coefficient of the lagged dependent variable, α , is allowed without loss of generality. The errors u_t are independent and identically distributed (*i.i.d.*) with variance σ_u^2 . The main result, proved in the appendix, is presented in the following theorem.

Theorem 2 *Suppose the data are generated by (19) with the regressors x_t satisfying A1-A3 and $u_t \sim i.i.d. (0, \sigma_u^2)$. Then, for a candidate break date $T_1 = [T\lambda]$ with $\lambda \in (0, 1)$ such that $\lambda \neq \lambda_0$,*

$$T^{-1} \hat{U}_T(T_1) = \frac{\frac{(1-\lambda_0)^2(\lambda_0-\lambda)^2}{3(1-\lambda)^3} \delta' R' R \delta}{\sigma_u^2 + (1 + \lambda - \lambda_0)(\lambda_0 - \lambda) \delta' R \delta} + o_p(1) = O_p(1)$$

uniformly in $\|\delta\|$, where $R = T^{-1} \sum_{t=1}^T x_t x_t'$.

Note that the limit value of $\hat{U}_T(T_1)$ is uniformly bounded in $\|\delta\|$. For such a dynamic regression model, Perron and Yamamoto (2012) showed that the sup-Wald test ($\sup W$) is an increasing function of $\|\delta\|$, that is, $\sup W = O_p(\|\delta\|^2)$. The two tests diverge with T and are accordingly consistent for any fixed δ . However, for any fixed T , the expansion in Theorem 2 suggests that the power of the test $\hat{U}_T(T_1)$ will be very different from that of the $\sup W$ as $\|\delta\|$ increases. The power of $\hat{U}_T(T_1)$ will depend on whether the limit value exceeds the critical value used. If for some configuration of parameters, the limit value is below the

critical value for all T_1 , the length of the confidence interval will be the whole sample. To analyze this issue, we present the limit value as stated in Theorem 2 for a variety of cases.

In Figure 2, the limit rejection probabilities of $\hat{U}_T(T_1)$ are plotted for three different values of $\alpha \in \{0.3, 0.5, 0.8\}$. Panels (a) and (b) show the rejection probabilities of $\hat{U}_T(T_1)$ for two candidate break dates $T_1 = [0.3T]$ and $T_1 = [0.7T]$, respectively. The test $\hat{U}_T(T_1)$ has some power with small changes, but power is reduced to zero as the magnitude of a change increases for all values of α . This implies that the candidate break dates $T_1 = [0.3T]$ and $T_1 = [0.7T]$ will be included in the confidence set when the magnitude of the change is large. Since power is increasing as T_1 gets further away from the null value $T_1^0 = [0.5T]$, this implies that the length of the confidence interval will be at least $[0.3T]$ to $[0.7T]$. Of importance is the fact that the power function decreases substantially when α increases. For instance, when $\alpha = 0.8$, power is less than size for all values of the break magnitude. Hence, the length of the confidence interval will approach the whole sample more quickly as $\|\delta\|$ increases, the closer α is to one. Finally, note that even if the included lagged dependent variable is redundant ($\alpha = 0$), the same problem remains albeit, as expected, with the power reversal occurring for larger values of $\|\delta\|$.

Figure 3 presents the graphs of the limit of $\hat{U}_T(T_1)$ as a function of α , for a break magnitude of $\delta \in \{0.4, 1.6, 3.2, 5.0\}$ in each panel. The values reported are the median from 3,000 replications. Again, we consider two cases with $\lambda = \{0.3, 0.7\}$ whose results are broadly similar. Consider the case with α close to zero. For a candidate break date $[0.3T]$ the test $\hat{U}_T(T_1)$ would reject when $\delta = 1.6$, but not when $\delta = 5$, which confirms the non-monotonic power of the test. Again, the power of $\hat{U}_T(T_1)$ decreases as α approaches one.

Figure 4 shows the limits of $\hat{U}_T([T\lambda])$ when $\lambda = 0.3$ for various combinations of $\alpha \in \{0.3, 0.5, 0.8\}$ and $T \in \{100, 250, 500, 1000\}$. The limit value of $\hat{U}_T([T\lambda])$ increases with the sample size T as explained in Theorem 2. It is clear, however, that the power of the test is non-monotonic for every combination of α and T . While the limit value of the $\hat{U}_T(T_1)$ initially increases with $\|\delta\|$, it quickly reverts back to stabilize at a small value which is below the relevant critical value (0.745). This again explains why the length of the confidence interval approaches the whole sample in a model with a lagged dependent variable as $\|\delta\|$ increases, and faster when α is closer to one. When $\alpha = 0.8$, a common value in practice, the procedure is virtually uninformative for any sample size considered yielding a confidence interval that is essentially the whole sample available.

The intuition behind this result follows again from the analysis in Perron (1990). For any candidate break date T_1 not equal to the true break date, the estimated residuals \hat{u}_t used

to construct \hat{U}_T are contaminated by the shift. Accordingly, a change will induce a bias of $\hat{\alpha}$ toward unity and more so as δ increases. This makes the structural change appears as an outlier and, hence, difficult to detect.

5 Conclusion

This paper considered constructing confidence intervals for the date of a structural break in linear regression models. Using extensive simulations, we compared the performance of various procedures in terms of exact coverage rates and lengths of the confidence intervals. On the basis of achieving an exact coverage rate that is closest to the nominal level, Elliott and Müller's (2007) approach is by far the best one. However, this comes with a very high cost in terms of the length of the confidence intervals. When the errors are serially correlated and dealing with a change in intercept or a change in the coefficient of a stationary regressor with a high signal to noise ratio, the length of the confidence interval increases and approaches the whole sample as the magnitude of the change increases. The same problem occurs in models with a lagged dependent variable, a common case in practice. This drawback is not present for the other methods, which have similar properties.

Our results are related to other studies dealing with the power of tests for structural changes. The basic underlying reason for the drawbacks of Elliott and Müller's (2007) approach is the fact that the test which is inverted to obtain the confidence intervals is motivated by optimal properties under a local asymptotic framework whereby the test is locally invariant to the magnitude of break. It is also a partial-sums type test for which only a model restricted to satisfy the null hypothesis of no change is used. It has been shown that tests based on such features have serious non-monotonic power problems; see, e.g., Deng and Perron (2008), Kim and Perron (2009), Perron and Yamamoto (2012). In the context of structural change tests, one should be skeptical of the use of some local asymptotic frameworks whereby the breaks are local to zero to devise testing procedures. These types of frameworks do not yield useful predictions about the finite sample properties of tests. Second, as argued in Perron (2006), one should also abandon partial-sums type tests for which only a model restricted to satisfy the null hypothesis of no change is used. These include the CUSUM, LM and $q\hat{L}L$ (Elliott and Müller, 2006) tests, among many others. These tests are plagued by the problem of a non-monotonic power function such that the power of the test can go to zero as the magnitude of change increases. The $\hat{U}_T(T_1)$ test of Elliott and Müller's (2007) is another example of such tests, which upon inversion yields confidence intervals with poor properties.

Appendix

Proof of Theorem 1: The data generating process is defined as follows. For $t = 1, \dots, T$,

$$y_t = x'_t \beta + x'_t \delta \mathbf{1}_{t > T_1^0} + z'_t \gamma + u_t,$$

or, in matrix notation,

$$Y = \begin{bmatrix} x'_1 & 0 & z'_1 \\ \vdots & \vdots & \vdots \\ x'_{T_1^0} & 0 & z'_{T_1^0} \\ x'_{T_1^0+1} & x'_{T_1^0+1} & z'_{T_1^0+1} \\ \vdots & \vdots & \vdots \\ x'_T & x'_T & z'_T \end{bmatrix} \begin{bmatrix} \beta \\ \delta \\ \gamma \end{bmatrix} + U = \begin{bmatrix} X_0 & Z \end{bmatrix} \begin{bmatrix} \beta \\ \delta \\ \gamma \end{bmatrix} + U = M_0 \Gamma + U.$$

Let X_1 denote X_0 replacing T_1^0 by an arbitrary break date T_1 and define $M = [X_1, Z]$. The OLS estimator of Γ is given by

$$\begin{aligned} \hat{\Gamma} &= (M' M)^{-1} M' Y = (M' M)^{-1} M' (M_0 \Gamma + U) \\ &= (M' M)^{-1} M' (M \Gamma - M \Gamma + M_0 \Gamma + U) \\ &= \Gamma - (M' M)^{-1} M' (M - M_0) \Gamma + (M' M)^{-1} M' U. \end{aligned}$$

and the OLS residuals are

$$\begin{aligned} \hat{U} &= Y - \hat{Y} = U + M_0 \Gamma - M \hat{\Gamma} \\ &= U + M \Gamma - M \hat{\Gamma} - M \Gamma + M_0 \Gamma = U - M(\hat{\Gamma} - \Gamma) - (M - M_0) \Gamma \\ &= U + M(M' M)^{-1} M' (M - M_0) \Gamma - M(M' M)^{-1} M' U - (M - M_0) \Gamma \\ &= [I - M(M' M)^{-1} M'] [U - (M - M_0) \Gamma]. \end{aligned} \tag{A.1}$$

By partitioned inversion applied to $M' M$, we have

$$M(M' M)^{-1} M' = X_1(X'_1 X_1)^{-1} X'_1 + \tilde{Z}(\tilde{Z}' \tilde{Z})^{-1} \tilde{Z}'$$

where $\{\tilde{z}_t\}$ are the OLS residuals from the regression of z'_t on $\{x'_t, x'_t \mathbf{1}_{t > T_1}\}$. We can then write (A.1) as:

$$\begin{aligned} \hat{U} &= (I - X_1(X'_1 X_1)^{-1} X'_1 - \tilde{Z}(\tilde{Z}' \tilde{Z})^{-1} \tilde{Z}') U \\ &\quad - (I - X_1(X'_1 X_1)^{-1} X'_1 - \tilde{Z}(\tilde{Z}' \tilde{Z})^{-1} \tilde{Z}')(M - M_0) \Gamma. \end{aligned} \tag{A.2}$$

Using a Weak Law of Large Numbers (WLLN), it is sufficient to consider the second term in (A.2). Note that $(M - M_0)\Gamma = (0, \dots, 0, (x'_{T_1+1}\delta)', \dots, (x'_{T_1^0}\delta)', 0, \dots, 0)'$ and,

$$X_1(X'_1 X_1)^{-1} X'_1 = \begin{bmatrix} x'_1 A x_1 & \dots & x'_1 A x_{T_1} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x'_{T_1} A x_1 & \dots & x'_{T_1} A x_{T_1} & 0 & \dots & 0 \\ 0 & \dots & 0 & x'_{T_1+1} B x_{T_1+1} & \dots & x'_{T_1+1} B x_T \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & x'_T B x_{T_1+1} & \dots & x'_T B x_T \end{bmatrix}$$

where $A = (\sum_{t=1}^{T_1} x_t x'_t)^{-1}$ and $B = (\sum_{t=T_1+1}^T x_t x'_t)^{-1}$. Let $T_1 = [T\lambda]$ and $T_1^0 = [T\lambda_0]$. For $0 < s \leq \lambda$, $T^{-1} \sum_{t=1}^{[Ts]} x_t \hat{u}_t = o_p(1)$. For $\lambda < s \leq \lambda_0$,

$$\begin{aligned} T^{-1} \sum_{t=[T\lambda]+1}^{[Ts]} x_t \hat{u}_t &= -T^{-1} \sum_{t=[T\lambda]+1}^{[Ts]} x_t x'_t \delta \\ &+ \left(T^{-1} \sum_{t=[T\lambda]+1}^{[Ts]} x_t x'_t \right) \left(T^{-1} \sum_{t=[T\lambda]+1}^T x_t x'_t \right)^{-1} \left(T^{-1} \sum_{t=[T\lambda]+1}^{[T\lambda_0]} x_t x'_t \delta \right) \\ &+ \left(T^{-1} \sum_{t=[T\lambda]+1}^{[Ts]} x_t \tilde{z}'_t \right) \left(T^{-1} \sum_{t=[T\lambda]+1}^T \tilde{z}_t \tilde{z}'_t \right)^{-1} \left(T^{-1} \sum_{t=[T\lambda]+1}^{[T\lambda_0]} \tilde{z}_t x'_t \delta \right). \end{aligned}$$

From the uniform convergence of $T^{-1} \sum_{t=1}^{[Ts]} x_t z'_t$ and $T^{-1} \sum_{t=1}^{[Ts]} x_t x'_t$ in s , we have

$$\begin{aligned} \sup_{\lambda} \left| \left(\sum_{t=[T\lambda]+1}^{[Ts]} x_t \tilde{z}'_t \right) \left(\sum_{t=[T\lambda]+1}^T \tilde{z}_t \tilde{z}'_t \right)^{-1} \left(\sum_{t=[T\lambda]+1}^{[T\lambda_0]} \tilde{z}_t x'_t \delta \right) \right. \\ \left. - \left(\sum_{t=[T\lambda]+1}^{[Ts]} x_t \tilde{z}'_t \right) \left(\sum_{t=[T\lambda]+1}^T \tilde{z}_t \tilde{z}'_t \right)^{-1} \left(\sum_{t=[T\lambda]+1}^{[T\lambda_0]} \tilde{z}_t x'_t \delta \right) \right| \xrightarrow{p} 0 \end{aligned}$$

where $\tilde{z}_t = z_t - \Sigma_{zx} R^{-1} x_t$ and $\Sigma_{zx} = T^{-1} \sum_{t=1}^T z_t x'_t$. Note that \tilde{z}_t does not depend on T_1 . Therefore,

$$\begin{aligned} T^{-1} \sum_{t=[T\lambda]+1}^{[Ts]} x_t \hat{u}_t &\xrightarrow{p} -(s - \lambda) R \delta + (s - \lambda) R (1 - \lambda)^{-1} R^{-1} (\lambda_0 - \lambda) R \delta \\ &= (s - \lambda) \left(\frac{\lambda_0 - 1}{1 - \lambda} \right) R \delta. \end{aligned}$$

Similarly, for $\lambda_0 < s \leq 1$,

$$\begin{aligned}
T^{-1} \sum_{t=[T\lambda]+1}^{[Ts]} x_t \hat{u}_t &= -T^{-1} \sum_{t=[T\lambda]+1}^{[T\lambda_0]} x_t x'_t \delta \\
&+ \left(T^{-1} \sum_{t=[T\lambda]+1}^{[Ts]} x_t x'_t \right) \left(T^{-1} \sum_{t=[T\lambda]+1}^T x_t x'_t \right)^{-1} \left(T^{-1} \sum_{t=[T\lambda]+1}^{[T\lambda_0]} x_t x'_t \delta \right) \\
&\xrightarrow{p} (s-1) \left(\frac{\lambda_0 - \lambda}{1 - \lambda} \right) R \delta.
\end{aligned}$$

It is easy to show that

$$\begin{aligned}
\hat{U}' \hat{U} &= [U - (M - M_0) \Gamma]' [I - M(M' M)^{-1} M'] [U - (M - M_0) \Gamma] \\
&= U' U - U' M (M' M)^{-1} M' U - U' (M - M_0) \Gamma + U' M (M' M)^{-1} M' (M - M_0) \Gamma \\
&\quad - \Gamma' (M - M_0)' U + \Gamma' (M - M_0)' M (M' M)^{-1} M' U \\
&\quad + \Gamma' (M - M_0)' [I - M(M' M)^{-1} M'] (M - M_0) \Gamma.
\end{aligned}$$

Hence,

$$\hat{\gamma}_0 = T^{-1} \hat{U}' \hat{U} \xrightarrow{p} \gamma_0 + (\lambda_0 - \lambda) \frac{1 - \lambda_0}{1 - \lambda} \delta' R \delta,$$

and

$$\hat{\gamma}_1 = T^{-1} \sum_{t=2}^T \hat{u}_t \hat{u}_{t-1} \xrightarrow{p} \gamma_1 + (\lambda_0 - \lambda) \frac{1 - \lambda_0}{1 - \lambda} \delta' R_1 \delta$$

where $R_j = p \lim_{T \rightarrow \infty} T^{-1} \sum_{t=j+1}^T x_t x'_{t-j}$ and $R_0 \equiv R$. Using these results, the estimate of the first-order autocorrelation of the errors is such that

$$\hat{\rho}(\delta) = \frac{T^{-1} \sum_{t=2}^T (\hat{u}_t - \bar{\hat{u}})(\hat{u}_{t-1} - \bar{\hat{u}})}{T^{-1} \sum_{t=2}^T (\hat{u}_{t-1} - \bar{\hat{u}})^2} \xrightarrow{p} \frac{\gamma_1 + (\lambda_0 - \lambda) \frac{1 - \lambda_0}{1 - \lambda} \delta' R_1 \delta}{\gamma_0 + (\lambda_0 - \lambda) \frac{1 - \lambda_0}{1 - \lambda} \delta' R \delta}.$$

The test statistic evaluated at a candidate break date $T_1 = [T\lambda]$ can be expressed as:

$$\begin{aligned}
\hat{U}_T(T_1) &= \frac{T^2}{[T\lambda]^2} \sum_{t=1}^{[T\lambda]} \left(T^{-1} \sum_{s=1}^t x_s \hat{u}_s \right)' \hat{\omega}^{-2} \left(T^{-1} \sum_{s=1}^t x_s \hat{u}_s \right) \\
&\quad + \frac{T^2}{(T - [T\lambda])^2} \sum_{t=[T\lambda]+1}^T \left(T^{-1} \sum_{s=[T\lambda]+1}^t x_s \hat{u}_s \right)' \hat{\omega}^{-2} \left(T^{-1} \sum_{s=[T\lambda]+1}^t x_s \hat{u}_s \right).
\end{aligned}$$

A popular data-dependent method to select the bandwidth m for a kernel function is proposed by Andrews (1991). For simplicity, we use the Bartlett kernel, $k(j, m) = 1 - j/m$ if $j \leq m$ and 0 otherwise, and Andrews's AR(1) approximation so that $m \propto (C(\delta)T)^{1/3}$ where

$C(\delta) = 4\hat{\rho}(\delta)^2/(1 - \hat{\rho}(\delta))^2$ with $\hat{\rho}(\delta)$ the OLS estimate from a regression of \hat{u}_t on \hat{u}_{t-1} . Using the fact that $\sum_{j=1}^{T-1} k(j, m) = O(m)$,

$$\begin{aligned}
\hat{\omega}^2 &= \hat{\gamma}_0 + 2 \sum_{j=1}^{T-1} k(j, m) \hat{\gamma}_j \\
&= \left(\gamma_0 + 2 \sum_{j=1}^{T-1} k(j, m) \gamma_j \right) + (\lambda_0 - \lambda) \frac{1 - \lambda_0}{1 - \lambda} \left(\delta' R \delta + 2 \sum_{j=1}^{T-1} k(j, m) \delta' R_j \delta \right) + o_p(1) \\
&= h(0) + (\lambda_0 - \lambda) \frac{1 - \lambda_0}{1 - \lambda} \left(\delta' R \delta + 2 \sum_{j=1}^{T-1} k(j, m) \delta' R_j \delta \right) + o_p(1) \\
&\leq h(0) + (\lambda_0 - \lambda) \frac{1 - \lambda_0}{1 - \lambda} \left(\delta' R \delta + 2 \sum_{j=1}^{T-1} k(j, m) \delta' R \delta \right) + o_p(1) \\
&\leq h(0) + (\lambda_0 - \lambda) \frac{1 - \lambda_0}{1 - \lambda} \delta' R \delta O(m) + o_p(\|\delta\|^2),
\end{aligned}$$

where $h(0) = \lim_{T \rightarrow \infty} \text{var}(T^{-1/2} \sum_{t=1}^T u_t)$, which is equivalent to $(2\pi \text{ times})$ the spectral density at frequency zero of u_t when the latter is a stationary process. If the condition C1 holds, then $\hat{\rho}(\delta) \xrightarrow{p} 1$ as $\|\delta\|$ increases such that $C(\delta) = O_p(\|\delta\|^4)$ since $1 - \hat{\rho}(\delta)^2 = O_p(\|\delta\|^{-2})$, $m = O_p(\|\delta\|^{4/\theta} T^{1/\theta})$, and $\hat{\omega}^2 = O_p(\|\delta\|^{\frac{4}{\theta} + 2} T^{\frac{1}{\theta}})$ where $\theta = 3$ for the Bartlett kernel and 5 for the quadratic spectral and others. If C1 does not hold, $\hat{\rho}(\delta) \xrightarrow{p} \kappa < 1$, say, as $\|\delta\|$ increases such that $C(\delta) = O_p(1)$, $m = O_p(T^{1/\theta})$, and $\hat{\omega}^2 = O_p(\|\delta\|^2 T^{\frac{1}{\theta}})$. This completes the proof.

Proof of Theorem 2: In matrix notation, (19) can be written as

$$Y = \begin{bmatrix} X_0 & Y_{-1} \end{bmatrix} \begin{bmatrix} \beta \\ \delta \\ \alpha \end{bmatrix} + U = M_0 \Gamma + U.$$

with X_0 as defined in the proof of Theorem 1. We know that

$$\hat{\Gamma} - \Gamma = (M' M)^{-1} M' U - (M' M)^{-1} M' (M - M_0) \Gamma.$$

Hence,

$$\begin{aligned}
T^{-1} \hat{U}' \hat{U} &= T^{-1} U' U + \delta' \left(T^{-1} \sum_{t=T_1+1}^{T_1^0} x_t x_t' \right) \delta \\
&\quad - \delta' \left(T^{-1} \sum_{t=T_1+1}^{T_1^0} x_t x_t' \right) \left(T^{-1} \sum_{t=T_1+1}^T x_t x_t' \right)^{-1} \left(T^{-1} \sum_{t=T_1+1}^{T_1^0} x_t x_t' \right) \delta + o_p(1) \\
&\xrightarrow{p} \sigma_u^2 + (\lambda_0 - \lambda) \delta' R \delta - (\lambda_0 - \lambda)^2 \delta' R \delta \\
&= \sigma_u^2 + (1 + \lambda - \lambda_0)(\lambda_0 - \lambda) \delta' R \delta.
\end{aligned}$$

On the other hand,

$$\begin{aligned}\hat{U} &= [I - X_1(X_1'X_1)^{-1}X_1' - \tilde{Y}_{-1}(\tilde{Y}_{-1}'\tilde{Y}_{-1})^{-1}\tilde{Y}_{-1}']U - \\ &\quad - [I - X_1(X_1'X_1)^{-1}X_1' - \tilde{Y}_{-1}(\tilde{Y}_{-1}'\tilde{Y}_{-1})^{-1}\tilde{Y}_{-1}'](M - M_0)\Gamma.\end{aligned}$$

Using a WLLN, it is sufficient to consider the second term in \hat{U} . For $0 < s < \lambda$, $T^{-1} \sum_{t=1}^{[Ts]} x_t \hat{u}_t = o_p(1)$. Moreover, under Assumptions A2, we can show that $T^{-1} \sum_{t=1}^{[Ts]} y_{t-1} x_t' \xrightarrow{p} s \Sigma_{yx}$, given y_0 . Hence, for $\lambda < s \leq \lambda_0$,

$$T^{-1} \sum_{t=[T\lambda]+1}^{[Ts]} x_t \hat{u}_t \xrightarrow{p} (s - \lambda) \left(\frac{-1 + \lambda_0}{1 - \lambda} \right) R\delta,$$

and for $\lambda_0 < s \leq 1$,

$$T^{-1} \sum_{t=[T\lambda]+1}^{[Ts]} x_t \hat{u}_t \xrightarrow{p} (s - 1) \left(\frac{\lambda_0 - \lambda}{1 - \lambda} \right) R\delta.$$

Therefore,

$$T^{-1} \hat{U}_T(T_1) = \frac{\frac{1}{(1-\lambda)^2} \left\{ \left(\frac{1-\lambda_0}{1-\lambda} \right)^2 \int_{\lambda}^{\lambda_0} (s - \lambda)^2 ds + \left(\frac{\lambda_0 - \lambda}{1-\lambda} \right)^2 \int_{\lambda_0}^1 (s - 1)^2 ds \right\} (\delta' R' R \delta)}{\sigma_u^2 + (1 + \lambda - \lambda_0)(\lambda_0 - \lambda)(\delta' R \delta)} + o_p(1),$$

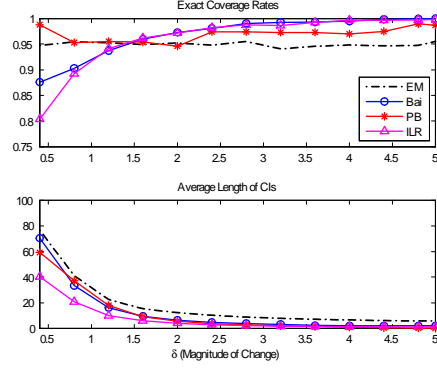
which completes the proof.

References

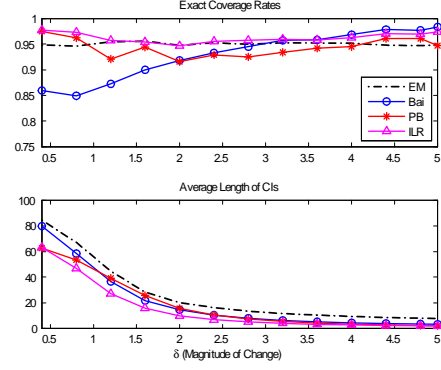
- Amemiya, T., 1985. *Advanced Econometrics*, Harvard University Press.
- Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica*, 59, 817-858.
- Andrews, D.W.K. and Monahan, J.C., 1992. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator, *Econometrica*, 60, 953-966.
- Andrews, D.W.K. and Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica*, 62, 1383-1414.
- Bai, J., 1994. Least squares estimation of a shift in linear processes, *Journal of Time Series Analysis*, 15, 453-472.
- Bai, J., 1997. Estimation of a change point in multiple regressions, *Review of Economics and Statistics*, 79, 551-563.
- Bai, J. and Perron, P., 1998. Estimating and testing linear models with multiple structural changes, *Econometrica*, 66, 47-78.
- Bai, J. and Perron, P., 2003. Computation and analysis of multiple structural change models, *Journal of Applied Econometrics*, 18, 1-22.
- Bai, J. and Perron, P., 2006. Multiple structural change models: a simulation analysis, *Econometric Theory and Practice: Frontiers of Analysis and Applied Research*, Corbae, D., Durlauf, S.N., Hansen, B.E. (Eds.), Cambridge Press, 212-237.
- Bhattacharya, P. and Brockwell, P., 1976. The minimum of an additive process with applications to signal estimation and storage theory, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 37, 51-75.
- Brockwell, P. and Davis, R.A., 1991. *Time Series: Theory and Methods* (2nd ed.), Springer.
- Bühlmann, P., 1997. Sieve bootstrap for time series, *Bernoulli*, 3, 123-148.
- Chang, S. and Perron, P., 2013. Inference on the date of a structural change with bootstrap methods, Unpublished working paper, Boston university.
- Davidson, R. and Flachaire, 2008. The wild bootstrap, tamed at last, *Journal of Econometrics*, 146, 162-169.
- Deng, A. and Perron, P., 2008. A non-local perspective on the power properties of the CUSUM and CUSUM of squares tests for structural change, *Journal of Econometrics*, 142, 212-240.
- Elliott, G. and Müller, U., 2006. Efficient tests for general persistent time variation in regression coefficients, *Review of Economic Studies*, 73, 907-940.

- Elliott, G. and Müller, U., 2007. Confidence sets for the date of a single break in linear time series regressions, *Journal of Econometrics*, 141, 1196-1218.
- Eo, Y. and Morley, J., 2013. Likelihood-based confidence sets for the timing of structural breaks, Unpublished Working Paper, University of Sydney.
- Kim, D. and Perron, P., 2009. Assessing the relative power of structural break tests using a framework based on the approximate Bahadur slope, *Journal of Econometrics*, 149, 26-51.
- Liu, R.Y., 1988. Bootstrap procedures under some non-I.I.D models, *The Annals of Statistics*, 16, 1696-1708.
- Nyblom, J., 1989. Testing for the constancy of parameters over time, *Journal of the American Statistical Association*, 89, 223-230.
- Perron, P., 1990. Testing for a unit root in a time series with a changing mean, *Journal of Business and Economic Statistics*, 8, 153-162.
- Perron, P., 2006. Dealing with structural breaks. In *Palgrave Handbook of Econometrics*, vol. 1: *Econometric Theory*, K. Patterson and T.C. Mills (eds.), Palgrave Macmillan, 278-352.
- Perron, P. and Qu, Z., 2006. Estimating restricted structural change models, *Journal of Econometrics* 134, 373-399.
- Perron, P. and Yamamoto, Y., 2012. On the usefulness or lack thereof of optimality criteria for structural change tests, forthcoming in *Econometric Reviews*.
- Perron, P. and Yamamoto, Y., 2013a. A note on estimating and testing for multiple structural changes in models with endogenous regressors via 2SLS, forthcoming in *Econometric Theory*.
- Perron, P. and Yamamoto, Y., 2013b. Estimating and testing multiple structural changes in linear models using band spectral regressions, forthcoming in *Econometrics Journal*.
- Ploberger, W. and Krämer, W., 1990. The local power of the CUSUM and CUSUM of squares tests, *Econometric Theory*, 6, 335-347.
- Qu, Z. and Perron, P., 2007. Estimating and testing structural changes in multivariate regressions, *Econometrica*, 75, 459-502.
- Siegmund, D., 1988. Confidence set in change-point problems, *International Statistical Review*, 56, 31-48
- Stock, J. and Watson, M., 2002. Has the business cycle changed and why?, *NBER Macroeconomics Annual*, 17, 159-218.

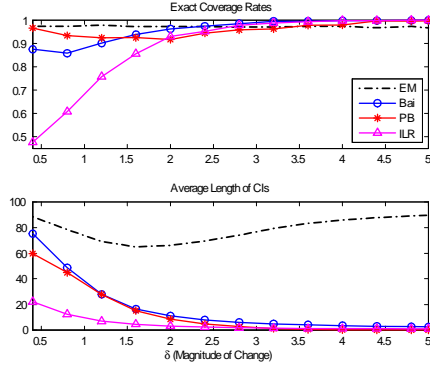
Figure 1: Coverage Rate and Average Length of Confidence Intervals



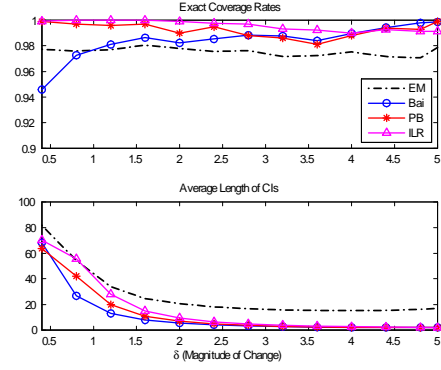
(a) DGP M1



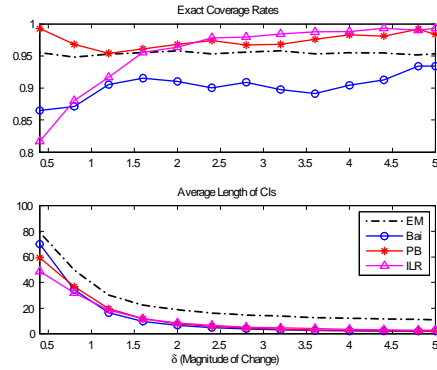
(b) DGP M2



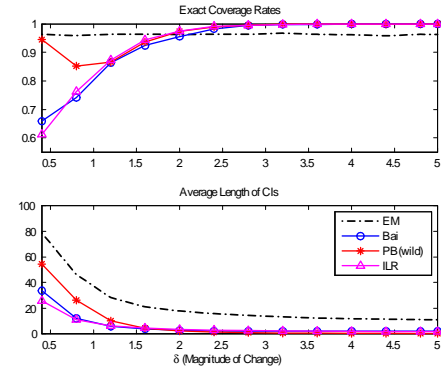
(c) DGP M3



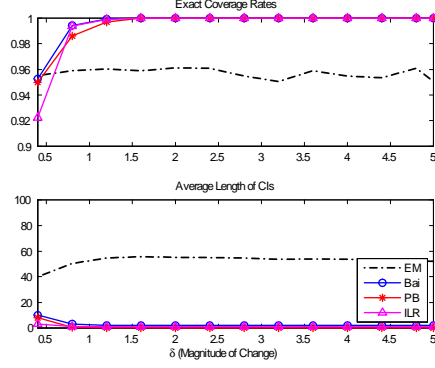
(d) DGP M4



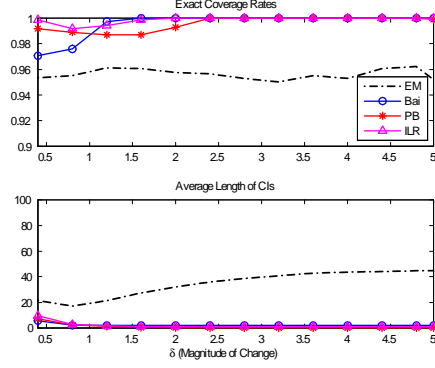
(e) DGP M5



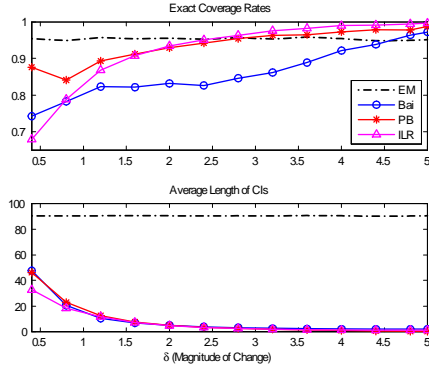
(f) DGP M6



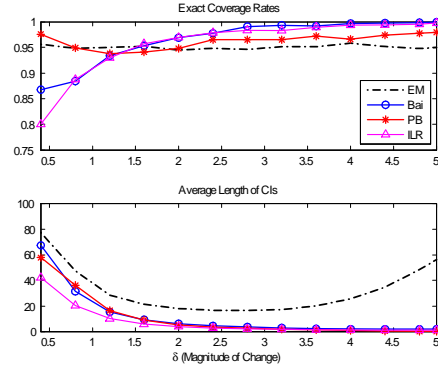
(g) DGP M7



(h) DGP M8



(i) DGP D1



(j) DGP D2

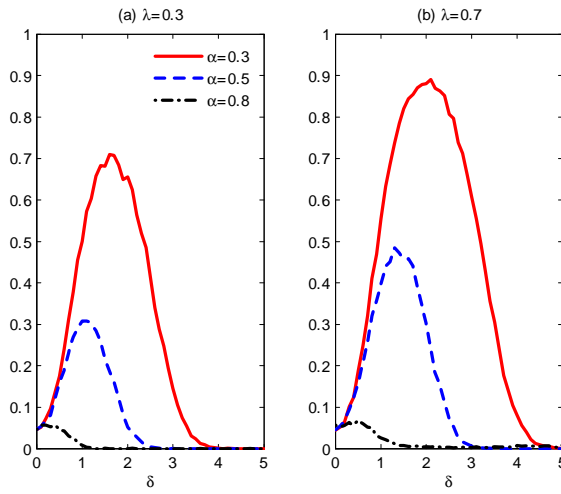


Figure 2: Power Functions of $\hat{U}_T([T\lambda])$ in Dynamic Regression Models. $T = 100, \lambda_0 = 0.5$

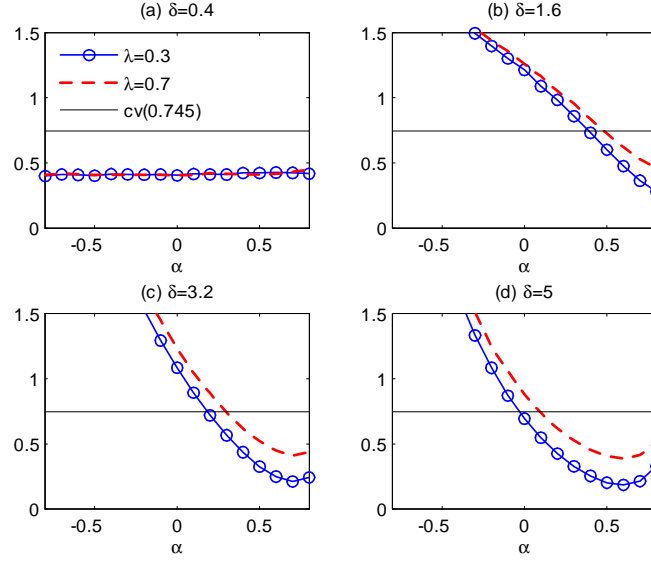


Figure 3: Limits of $\hat{U}_T([T\lambda])$ in Dynamic Regression Models. $T = 100, \lambda = 0.5$

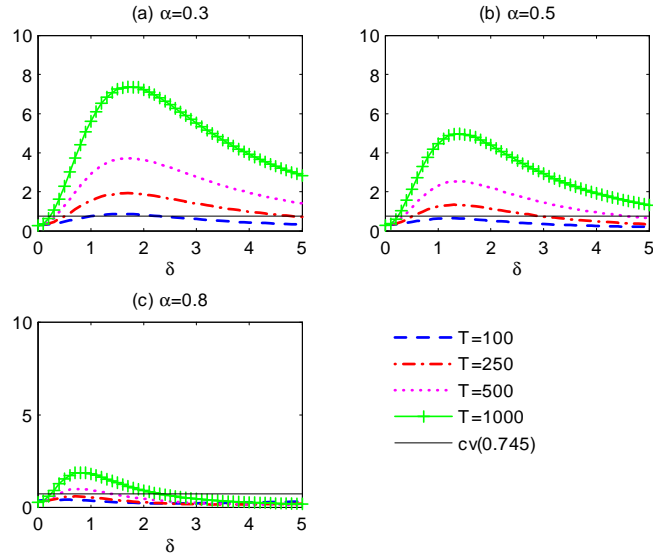


Figure 4: Limits of $\hat{U}_T([T\lambda])$ in Dynamic Regression Models. $\lambda_0 = 0.5, \lambda = 0.3$