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

Seong Yeon Chang†  Pierre Perron‡
Xiamen University  Boston University

September 20, 2013; Revised: October 11, 2015

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 (2015) 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: C12, C22

Keywords: Bootstrap; Confidence interval; Dynamic regression models; Inverted likelihood ratio; Non-monotonic power; Serially correlated errors; Structural change

*We are grateful to the associate editor, two anonymous referees, and Yohei Yamamoto for helpful comments. This paper is a revised version of parts of Seong Yeon Chang’s Ph.D. Dissertation at Boston University (Chang, 2014a).

†Wang Yanan Institute for Studies in Economics, MOE Key Laboratory of Econometrics and Fujian Key Laboratory of Statistical Science, Xiamen University, Xiamen, Fujian 361005, China (sychang@xmu.edu.cn).

‡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 linear regression models. 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 (2013) for linear models estimated via band spectral regression and Perron and Yamamoto (2014) for the case of linear models with endogenous regressors estimated by two-stage least squares. 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) showed 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.

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’s (1989) test, that is locally invariant to the magnitude of the break. They showed, via simulations, that the resulting confidence sets 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.

Recently, Eo and Morley (2015) generalized Siegmund’s (1988) method to a system of multivariate regressions, 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 arguing that it provides confidence sets with smallest length compared to other methods.
Another avenue is to use bootstrap procedures. With independent and identically distributed errors, a standard residual bootstrap procedure is possible; with heteroskedastic errors, the wild bootstrap approach of Liu (1988) can be applied; and with serially correlated errors, the sieve bootstrap procedure analyzed by Chang (2014b) 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. Given the limitations of Elliott and Müller’s (2007) method, we consider models with a single break.

Our findings can be summarized as follows. On the basis of achieving an exact coverage rate 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 can exhibit liberal distortions, exact coverage rates below nominal level. It can also be conservative. The bootstrap procedures and the ILR are less prone to size distortions though they occur in some cases. 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. With serially correlated errors and dealing with a change in intercept or a change in the coefficient of a stationary regressors 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 procedures that strike the best balance between coverage rate and length are the bootstrap and the ILR. Bai’s method does 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 the procedures state that the uncertainty about the location of the break date is, say, 50% of the total sample instead of the correct 70% given the liberal size distortions. In either case, the answer is the same: the data are not informative about the location of the break and the estimate is not reliable. Adopting 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 set 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. 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. 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. Perron and Yamamoto (2012) considered the so-called optimal qLL 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 $T_1^0$

$$y_t = x_t'\beta_t + z_t'\gamma + u_t,$$
$$\beta_t = \beta + \delta 1_{t>T_1^0}$$

for $t = 1, \ldots, T$ where $1_A$ 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 $(\beta, \delta, \gamma)$ are the corresponding vectors of coefficients; $u_t$ is the disturbance with $E(u_t^2) = \sigma_t^2$ for $t \leq T_1^0$ and $E(u_t^2) = \sigma_2^2$ for $t > T_1^0$. As usual, we assume that each regime is a fixed proportion of
the sample size so that $T^0_1 = [T\lambda_0]$ for some $\lambda_0 \in (0, 1)$. Note that we allow for a partial structural change model since $\gamma$ 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 $T^0_1$. Below, we review several procedures that have been proposed.

2.1 Bai’s (1997) approach

Bai (1997) considered estimating the break date using

$$
T_1 = \arg \min_{T_1 \in \Lambda} SSR(T_1)
$$

where $SSR(T_1)$ is the sum of squared residuals from the unrestricted regression (1) evaluated at the candidate break date $T_1$ and $\Lambda = T\Lambda_{\varepsilon}$ where $\Lambda_{\varepsilon} = (\varepsilon, 1 - \varepsilon)$ for some small trimming $\varepsilon$. Using $T_1$ as a candidate break date, we can write (1) as

$$
y_t = x_1(T_1)\beta + x_2(T_1)\psi + z_1^*\gamma + u_t
$$

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 $Y = (y_1, \ldots, y_T)'$, $Z = (z_1, \ldots, z_T)'$, and $U = (u_1, \ldots, u_T)'$. $\bar{X}$ is the matrix which diagonally partitions $X$ at $T_1$, i.e., $\bar{X} = \text{diag}(X_1, X_2)$ with $X_1 = (x_1, \ldots, x_{T_1})'$ and $X_2 = (x_{T_1+1}, \ldots, x_T)'$, and $\Gamma = (\Gamma_1', \gamma')' = (\beta', \psi', \gamma')'$. Let $T_1 = [T\lambda]$ with $\lambda \in \Lambda_{\varepsilon}$. 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, in the set of the possible break dates $\Lambda$, $T_1 = \arg \max_{T_1 \in \Lambda} F(T_1)$. 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. 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 the 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 $T_1^0$, let $Q = \lim_{T \to \infty} T^{-1} \sum_{t=1}^T E(x_t x_t')$, $Q_1 = \lim_{T \to \infty} (T_1^0)^{-1} \sum_{t=1}^{T_1^0} E(x_t x_t')$, $Q_2 = \lim_{T \to \infty} (T - T_1^0)^{-1} \sum_{t=T_1^0+1}^T E(x_t x_t')$, $\omega^2 = \lim_{T \to \infty} T^{-1} E[\sum_{t=1}^T x_t u_t][\sum_{t=1}^T x_t' u_t]$, 

$$
\omega_1^2 = \lim_{T \to \infty} (T_1^0)^{-1} E[\sum_{t=1}^{T_1^0} x_t' u_t][\sum_{t=1}^{T_1^0} x_t' u_t]', 
$$

$$
\omega_2^2 = \lim_{T \to \infty} (T - T_1^0)^{-1} E[\sum_{t=T_1^0+1}^T x_t u_t'][\sum_{t=T_1^0+1}^T x_t u_t]' .
$$

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:

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

where $V(s) = W_1(-s) - |s|/2$ if $s \leq 0$ and $V(s) = \phi W_2(s) - \xi |s|/2$ if $s > 0$, where $\xi = \delta' Q_2 \delta / \delta' Q_1 \delta$, and $\phi = \delta^2 \omega_2 \delta / \delta^2 \omega_1 \delta$. 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 $\arg \max_{s \in \mathbb{R}} V(s)$ is derived in Bai (1997). Since $\delta$, $Q_i$, and $\omega_1^2$ 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 $\omega_2^2$ can be constructed using a HAC estimator applied to the vector $\{x_t u_t\}$ and using data over segment $i$ only. As a special case, suppose that $\{x_t, u_t\}$ are 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

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

(5)

This can be evaluated using the estimates $\hat{Q} = T^{-1} \sum_{t=1}^T x_t x_t'$ and $\hat{\sigma}^2 = T^{-1} \sum_{t=1}^T \hat{u}_t^2$. The 100$(1 - a)$% confidence interval is constructed as $[\hat{T}_1 - [q_{1-a/2}/\hat{L}]] - 1, \hat{T}_1 + [q_{1-a/2}/\hat{L}] + 1$ where $q_{1-a/2}$ is the $(1 - a/2)$th quantile of the random variable $\arg \max_{s \in \mathbb{R}} \{W(s) - |s|/2\}$ and $\hat{L} = \hat{\delta} / \hat{\delta}^2; [q_{1-a/2}/\hat{L}]$ is the integer part of $q_{1-a/2}/\hat{L}$. 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

$$
\frac{(\delta' Q \delta)^2}{\delta' \omega_1^2 \delta} (\hat{T}_1 - T_1^0) \Rightarrow \arg \max_{s \in \mathbb{R}} \{W(s) - |s|/2\},
$$

(6)

where $Q = Q_1 = Q_2$ and $\omega^2 = \omega_1^2 = \omega_2^2$. An estimate of $\omega^2$ 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, when the errors are uncorrelated but heteroskedastic and the regressors are identically distributed across segments, \( \xi = 1 \) and \( \phi = \sigma_2^2/\sigma_1^2 \), which can be estimated using \( \hat{\sigma}_1^2 = \hat{T}_1^{-1} \sum_{t=1}^{\hat{T}_1} \hat{u}_t^2 \) and \( \hat{\sigma}_2^2 = (T - \hat{T}_1)^{-1} \sum_{t=\hat{T}_1+1}^{T} \hat{u}_t^2 \), and

\[
\frac{\delta'Q\delta}{\sigma_1^2}(\hat{T}_1 - T_1^0) \Rightarrow \arg \max_{s \in \mathbb{R}} V(s). \tag{7}
\]

2.2 Elliott and Müller’s (2007) approach

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

\[
U_T(T_1) = \frac{1}{T_1^2} \sum_{t=1}^{T_1} \left( \sum_{s=1}^{t} v_s \right)' (\omega_1^2)^{-1} \left( \sum_{s=1}^{t} v_s \right) + \frac{1}{(T - T_1)^2} \sum_{t=T_1+1}^{T} \left( \sum_{s=T_1+1}^{t} v_s \right)' (\omega_2^2)^{-1} \left( \sum_{s=T_1+1}^{t} v_s \right) \tag{8}
\]

where \( v_t = x_t \hat{u}_t \) with \( \hat{u}_t \) the OLS residuals from regression (4). The procedure for constructing the confidence set of the break date is as follows. For every \( T_1 \in \{q+2p+1, \ldots, T-q-2p-1\} \), perform the following steps: 1) Obtain the OLS residuals \( \hat{u}_t \) from regression (4); 2) Construct a consistent estimate of \( \omega_1^2 \) and \( \omega_2^2 \), 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 method of Andrews (1991) or Andrews and Monahan (1992) is recommended; 3) Compute \( \hat{U}_T(T_1) \) from (8) with \( \omega_1^2 \) and \( \omega_2^2 \) replaced by \( \hat{\omega}_1^2 \) and \( \hat{\omega}_2^2 \); 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 \( \hat{U}_T(T_1^0) \) at the significance level \( a \) obtained from the limit distribution \( \int_0^1 B(s)'B(s)ds \), where \( B(s) \) is a \( 2p \times 1 \) vector standard Brownian bridge.

2.3 Bootstrap methods

Bootstrap methods are popular 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 \textit{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. For a pure structural change model, the data-generating process is

\[
y_t = x_t'\beta + x_t'\delta \mathbf{1}_{t>T_1^0} + u_t \quad (t = 1, \ldots, T) \tag{9}
\]

1) \textbf{The case with \textit{i.i.d. errors}.} Consider estimating (9) by OLS with estimated residuals \( \hat{u}_t = y_t - x_t'\hat{\beta} - x_t'\hat{\delta} \mathbf{1}_{t>T_1^0} \) for \( t = 1, \ldots, T \). Define \( \tilde{u} = T^{-1} \sum_{t=1}^{T} \hat{u}_t \) and construct a set \( \tilde{U} = \{ \tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_T \} \). Draw a random sample from \( \tilde{U} \) with replacement and label
Under the assumption that the errors are \(i.i.d\), we can construct a new process \(\{y_t^*\}\) as
\[
y_t^* = x_t^\prime \hat{\theta} + x_t^\prime \hat{\delta} 1_{t>T_1} + \hat{u}_t^*
\]
for \(t = 1, \ldots, T\) and obtain the OLS estimates \(\hat{\beta}_{(b)}, \hat{\delta}_{(b)}, \text{ and } \hat{T}_{1(b)}\) 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 correctly replicate the distribution under the null hypothesis. This approach is also valid in cases 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 \(y_0^* = y_0\) and for \(t = 1, \ldots, T\),
\[
y_t^* = \hat{\gamma}_t y_{t-1}^* + x_t^\prime \hat{\theta} + x_t^\prime \hat{\delta} 1_{t>T_1} + \hat{u}_t^*.
\]
For each bootstrap sample, we estimate the break date \(\hat{T}_{1(b)}^*\) for \(b = 1, \ldots, B\).

2) The case with serially correlated errors. To account for serial correlation in the errors, whose nature is assumed to remain constant, we consider the sieve bootstrap as suggested by Bühlmann (1997). Let the error term \(u_t\) be generated by the linear process
\[
u_t = \psi(L) \epsilon_t \tag{10}
\]
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 (10) holds if \(u_t\) is a real-valued stationary process and purely stochastic; see Brockwell and Davis (1991). Under the assumption of invertibility, we can express \(u_t\) as a one-sided infinite-order autoregression \(\epsilon_t = \phi(L) u_t\) where \(\phi(z) = \sum_{k=0}^{\infty} \phi_k z^k\). We approximate the process by a finite autoregression of order \(p\), i.e.,
\[
u_t = \phi_{1,T} u_{t-1} + \ldots + \phi_{p,T} u_{t-p} + \epsilon_t.
\]
We estimate \(\phi = (\phi_{1,T}, \ldots, \phi_{p,T})'\) using the Yule-Walker equations to ensure a stationary solution. The estimates are then \(\hat{\phi}_p = (\hat{\phi}_{1,T}, \ldots, \hat{\phi}_{p,T})'\) such that \(\hat{\Gamma}_p \hat{\phi}_p = \hat{\gamma}_p\) where \(\hat{\Gamma}_p = \hat{\gamma}(i-j)]_{i,j=1}^p\), \(\hat{\gamma}_p = (\hat{\gamma}(1), \ldots, \hat{\gamma}(p))', \hat{\gamma}(s) = T^{-1} \sum_{t=1}^{T-s} (\tilde{u}_t - \bar{u})(\tilde{u}_{t+s} - \bar{u})\) and for some \(s \geq 1\), \(\bar{u} = T^{-1} \sum_{t=1}^T \tilde{u}_t\). The sieve bootstrap will be valid under the following conditions.

• Assumption SB: 1) \(u_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}, \psi_0 = 1\) with \(\{\epsilon_t\}\) an \(i.i.d\) sequence, \(E(\epsilon_t) = 0\) and \(E(\epsilon_t^r) < \infty\) for some \(r > 4\); 2) \(\psi(z) \neq 0\) for all \(|z| \leq 1\) and \(\sum_{j=0}^{\infty} |j|^s |\psi_j| < \infty\) for some \(s \geq 1\); 3) \(p = p_T \to \infty\) and \(p_T = o((T/\log T)^{1/2})\) as \(T \to \infty\).
Shibata (1980) showed that the Akaike Information Criterion (AIC) chooses an asymptotically efficient estimate $p_T$ for the optimal order of some projected AR($\infty$). We use the AIC to select the order of the autoregression $p_T$ by minimizing $T(\ln \hat{\sigma} + 1) + 2(p + 1)$ where $\hat{\sigma} = (\hat{\gamma}(0) - \hat{\phi}'\hat{\gamma}_p)^{1/2}$. Under assumption SB, we have $\ln \hat{\sigma} = \ln \sigma + O(\ln(T/T)^{1/2}) + o(p^{-s})$ a.s., as shown by Bühlmann (1995). Therefore, $p_T$ from the AIC satisfies $p_T = o(T^{1/(1+s)})$ a.s., so that the condition in SB (3) holds a.s. since $s > 1$. We can then obtain the estimated residuals, for $t = p_T + 1, \ldots, T$, $\hat{\epsilon}_{t,T} = \sum_{j=0}^{p_T} \hat{\phi}_{j,T} \hat{u}_{t-j} - \hat{\phi}_0,T = 1$. We center the estimated residuals so that their sample mean is zero: for $t = p_T + 1, \ldots, T$, $\hat{\epsilon}_{t,T} = \hat{\epsilon}_{t,T} - (T - p_T)^{-1} \sum_{t=pt+1}^{T} \hat{\epsilon}_{t,T}$ and denote the empirical CDF of $\{\hat{\epsilon}_{t,T}\}_{t=p_T+1}^{T}$ by $F^*(z) = (T - p_T)^{-1} \sum_{t=p_T+1}^{T} 1_{\hat{\epsilon}_{t,T} \leq z}$. We can re-sample, for any $t \in \mathbb{Z}$, $\hat{\epsilon}_t^* i.i.d.$ from $F^*$, and define $\hat{u}_t^*$ by the recursion,

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

(11)

with appropriately chosen $p_T$-initial values of $\hat{u}_t^*$. To that effect, we initially set the initial values to zero and construct the AR($p_T$) process from (11) 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

$$y_t^* = x_t^* \hat{\beta} + x_t^* \hat{\delta} 1_{t > \hat{T}_1} + \hat{u}_t^*$$

and we estimate the break date $\hat{T}_{1(b)}$ for each replication $b = 1, \ldots, B$.

3) The case with heteroskedastic errors. To account for heteroskedasticity, we adopt the wild bootstrap method of Liu (1988). The bootstrap sequence $y_t^*$ is generated by:

$$y_t^* = x_t^* \hat{\beta} + x_t^* \hat{\delta} 1_{t > \hat{T}_1} + \hat{u}_t^*,$$

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

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

so that $\zeta_t$ 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 $b = 1, \ldots, B$. 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-ratio based method

Siegmund (1988) suggested a likelihood-based method to construct a confidence set for a structural break date in a change in mean model with independent normal observations. Recently, Eo and Morley (2015) generalized his 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 allowing the unconditional variance to change across regimes. The parameters are estimated by restricted quasi-maximum likelihood based on the assumption of serially uncorrelated normal errors; see Qu and Perron (2007). The quasi-likelihood function is

\[
L_T(T_1, \gamma, \sigma) = \prod_{t=1}^{T_1} f(y_t|x_t; \gamma_1, \sigma_1) \cdot \prod_{t=T_1+1}^{T} f(y_t|x_t; \gamma_2, \sigma_2)
\]

where for \( j = 1, 2, \)

\[
f(y_t|x_t; \gamma_j, \sigma_j) = \frac{1}{(2\pi)^{1/2}\sigma_j} \exp\left(-\frac{1}{2\sigma_j^2}(y_t - x'_t\gamma_j)^2\right),
\]

\( \gamma_1 = \beta \) and \( \gamma_2 = \beta + \delta \). As a matter of notation, let \( \sigma = (\sigma_1, \sigma_2) \) and \( \gamma = (\gamma_1, \gamma_2) \). The logarithm of the quasi-likelihood function \( L_T(T_1, \gamma, \sigma) \) is given by:

\[
l_T(T_1, \gamma, \sigma) = \ln L_T(T_1, \gamma, \sigma) = \sum_{j=1}^{2} \sum_{t=T_{j-1}+1}^{T_j} \left\{ -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(\sigma_j^2) - \frac{1}{2\sigma_j^2}(y_t - x'_t\gamma_j)^2 \right\}
\]

where we use the convention that \( T_0 = 1 \) and \( T_2 = T \) with a single structural break. The parameters of interest \( (T_1^0, \gamma, \sigma) \) are estimated by maximizing the quasi-likelihood function, i.e., \( \hat{T}_1, \hat{\gamma}, \hat{\sigma} = \arg\max_{(T_1, \gamma, \sigma)} l_T(T_1, \gamma, \sigma) \). Following the notation in Eo and Morley (2015), let \( l_T(T_1) \) denote the logarithm of the profile likelihood function for the break date:

\[
l_T(T_1) = l_T(T_1, \hat{\gamma}(T_1), \hat{\sigma}(T_1)) = \max_{(\gamma, \sigma)} \sum_{t=1}^{T_1} \left\{ -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(\sigma_1^2) - \frac{1}{2\sigma_1^2}(y_t - x'_t\gamma_1)^2 \right\} + \sum_{t=T_1+1}^{T} \left\{ -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(\sigma_2^2) - \frac{1}{2\sigma_2^2}(y_t - x'_t\gamma_2)^2 \right\}.
\]

\[1\] We also considered the flipped confidence interval \( PB_f \equiv [2\hat{T}_1 - q_H, 2\hat{T}_1 - q_L] \). \( PB_f \) has the same length as \( PB \) and could have better coverage properties if the distribution of the statistic of interest is asymmetric. However, \( PB \) shows better finite sample performance in terms coverage rates, hence we only report simulation results for \( PB \) in Section 3.
The 100(1 - \( a \))% confidence sets are constructed by inverting the following \( \alpha \)-level likelihood ratio test of the null hypothesis \( H_0: T_1 = T_1^0 \) sequentially for all admissible break dates \( T_1: LR(T_1) = -2[l_T(T_1) - l_T(T_1)] \), where \( l_T(T_1) = \max_{T_1} l_T(T_1) \) and \( \hat{T}_1 = \arg \max_{T_1} l_T(T_1) \). Now let \( B_1 = (\sigma_2^2 - \sigma_1^2)/(\sigma_2^2; Q_1 = \text{plim}_{T \to \infty}(\sigma_2^2 T_1^0)^{-1} \sum_{t=1}^{T_1^0} x_t x_t'; Q_2 = \text{plim}_{T \to \infty}(\sigma_2^2(1 - T_1^0)^{-1} \sum_{t=T_1^0+1}^{T} x_t x_t') \), \( \Pi_1 = \text{lim}_{T \to \infty} \text{var}\{(T_1^0)^{-1}/2 \sigma_2^2 \sum_{t=1}^{T_1^0} x_t u_t\}, \Pi_2 = \text{lim}_{T \to \infty} \text{var}\{(T - T_1^0)^{-1/2} \sigma_2^{-2} \sum_{t=T_1^0+1}^{T} x_t u_t\}, \Omega_1 = \text{lim}_{T \to \infty} \text{var}\{(T_1^0)^{-1/2} \sum_{t=1}^{T_1^0} (\sigma_1^{-2} u_t^2 - 1)\}, \Omega_2 = \text{lim}_{T \to \infty} \text{var}\{(T - T_1^0)^{-1/2} \sum_{t=T_1^0+1}^{T} (\sigma_2^{-2} u_t^2 - 1)\}, \Gamma_1 = [(1/4) \Omega_1 B_1^2 + \Delta \gamma/\Pi_1 \Delta \gamma]^{1/2}, \Gamma_2 = [(1/4) \Omega_2 B_2^2 + \Delta \gamma/\Pi_2 \Delta \gamma]^{1/2}, \Psi_1 = (1/2) B_1^2 + \Delta \gamma/\Pi_1 \Delta \gamma, \text{ and } \Psi_2 = (1/2) B_2^2 + \Delta \gamma/\Pi_2 \Delta \gamma. \) The magnitudes of the change are assumed to satisfy \( \sigma_2^2 - \sigma_1^2 = v_T \Phi \) and \( \Delta \gamma = \gamma_2 - \gamma_1 = \delta = v_T \delta^* \) where \((\delta^*, \Phi) \neq 0\) independent of \( T \). Further, \( v_T \) is a positive sequence that satisfies \( v_T \to 0 \) and \( T^{1/2} v_T/\text{ln}(T)^2 \to \infty \) as \( T \to \infty \) (see Eo and Morley, 2015, Assumption 7).

**Lemma 1 (Eo and Morley, 2015, Proposition 1)** With \( W(\cdot) \) a standard Wiener process, the likelihood ratio statistic for the break date satisfies

\[
LR(T_1^0) \Rightarrow \Xi = \max_{\nu} \left\{ \begin{array}{ll}
w_1(-|\nu| + 2W(\nu)) & \text{for } \nu \in (-\infty, 0], \\
w_2(-|\nu| + 2W(\nu)) & \text{for } \nu \in (0, \infty)
\end{array} \right.
\]

where \( w_1 = \Gamma_1^2/\Psi_1 \) and \( w_2 = \Gamma_2^2/\Psi_2 \). The distribution function of \( \Xi \) is

\[
P(\Xi \leq z) = \left(1 - \exp\left(-\frac{z}{2w_1}\right)\right) \left(1 - \exp\left(-\frac{z}{2w_2}\right)\right).
\]  

(12)

With the level \( \alpha \) critical value \( \kappa \) obtained from (12), the 100(1 - \( a \))% confidence set is constructed by \( ILR = \{T_1: LR(T_1) \leq \kappa\} \).

**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 considered in Elliott and Müller (2007) are revisited and expanded. For regression models without a lagged dependent variable, the data generating process (DGP) is specified as:

\[
y_t = x_t' \beta + x_t' \delta 1_{t > T_1^0} + z_t' \gamma + u_t
\]  

(13)

for \( t = 1, \ldots, T \) with \( T = 100 \). The true break date is \( T_1^0 = \lfloor T \lambda_0 \rfloor \), \( \lambda_0 = 0.5 \), and the trimming \( \varepsilon = 0.15 \). The magnitude of change is \( \delta = dT^{-1/2} \) with \( d \in \{4, 8, \ldots, 48\} \), or equivalently \( \delta \in \{0.4, 0.8, \ldots, 4.8\} \). 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 the variance of $u_t$ that quadruples at $T_0^0$;

• M3: same as M1, but with AR(1) errors, $u_t = 0.3u_{t-1} + \epsilon_t$, $\epsilon_t \sim i.i.d.N(0,0.49)$;

• M3-1: same as M3, but with persistent AR(1) errors, $u_t = 0.8u_{t-1} + \epsilon_t$, $\epsilon_t \sim i.i.d.N(0,0.04)$;

• 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., $x_t = 0.5x_{t-1} + v_t$, $v_t \sim i.i.d.N(0,0.75)$, $z_t = 1$, $\gamma = 1$, 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, $x_t = \mu + 0.5x_{t-1} + v_t$ with $\mu = 5$ and $v_t \sim i.i.d.N(0,0.75)$, $z_t = 1$, $\gamma = 1$, and AR(1) errors $u_t = 0.3u_{t-1} + \epsilon_t$, $\epsilon_t \sim i.i.d.N(0,0.49)$;

• M7-1: Same as M7, but with persistent AR(1) errors, $x_t = \mu + 0.5x_{t-1} + v_t$ with $\mu = 5$ and $v_t \sim i.i.d.N(0,0.75)$, $z_t = 1$, $\gamma = 1$, and $u_t = 0.8u_{t-1} + \epsilon_t$, $\epsilon_t \sim i.i.d.N(0,0.04)$;

• M8: Same as M7, but with MA(1) errors, $x_t = \mu + 0.5x_{t-1} + v_t$ with $\mu = 5$ and $v_t \sim i.i.d.N(0,0.75)$, $z_t = 1$, $\gamma = 1$, $u_t = \epsilon_t - 0.3\epsilon_{t-1}$, $\epsilon_t \sim i.i.d.N(0,2.04)$.

• M9: AR(1) model, $y_t = \delta(1-\alpha)1_{t>T_0^0} + \alpha y_{t-1} + \epsilon_t$, $\epsilon_t \sim i.i.d.N(0,0.49)$, and $\alpha = 0.3$;

• M9-1: Same as M9, but $\epsilon_t \sim i.i.d.N(0,0.04)$ with a persistent AR(1) coefficient $\alpha = 0.8$;

For M7, M7-1, and M8, $E(x_t) = 2\mu$, which contrasts to M5 with $E(x_t) = 0$. This specification implies that the signal to noise ratio $|E(x_t)/\text{var}(x_t)^{1/2}|$ is high. Note that M9 is equivalent to M3 but with the dynamics in the regression instead of the errors. M9 can be used to assess the effect of serial correlation modeled parametrically or non-parametrically on the confidence intervals. We also consider a dynamic regression, given by

$$y_t = \alpha y_{t-1} + \delta 1_{t>T_0^0} + \epsilon_t, \quad \epsilon_t \sim i.i.d.N(0,1),$$

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 consider 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 quantitatively similar, we only report results for $\hat{U}_T.neq$. For all cases, 3,000 replications are used.
and to construct the percentile bootstrap confidence intervals (PB) the number of bootstrap samples is set to $B = 399$. We use the limiting distribution (5) of Bai and the residual bootstrap with i.i.d. errors for M1, M5, M9, M9-1, D1, and D2; the limiting distribution (6) and the sieve bootstrap with serially correlated errors for M3, M3-1, M4, M7, M7-1, and M8. With heteroskedastic errors (M2 and M6), we adopt the limiting distribution (7) and the wild bootstrap. For the ILR approach, we use the distribution function of $\Xi$ in (12) where $w_i, i = 1, 2$ can be replaced by consistent estimates to construct confidence sets. In models with i.i.d. errors, we have $w_i^2 = \sigma^2_i (i = 1, 2)$ and we estimate $\sigma^2_i$ by the sample variance of the estimated residuals within each regime. For models with serially correlated or heteroskedastic errors, we use Andrews and Mohanan’s (1992) AR(1) prewhitened two-stage procedure to select the bandwidth with a quadratic spectral kernel.

The results for 95% coverage rates are presented in Tables 1-14 for the various DGPs. Each table corresponds to a different DGP and presents, as a function of $\delta$, the exact coverage rate, the average length and the power of the sup-Wald test. The latter provides some information about how “big” the change is.

Consider first the coverage rates. Elliott and Møller’s (2007) (henceforth 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 (coverage rates above 95%) in some cases, e.g., M3, M3-1, M4, M6, M7, M7-1, M8, M9-1, D1. 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, M4, M5, M8, D2 and M3, M3-1, M6, M7, M7-1, M9, M9-1, D1 for large breaks. It is worthwhile noting that the bootstrap methods show liberal distortions (coverage rates below 95%) in some cases (M2 for medium breaks, M3-1, M6 and M7-1 for small breaks). The method of Bai exhibits liberal distortions in some cases when the break is small. This occurs for M1, M3, M3-1, M4, M5, M6, M9, M9-1, D1 and D2. The liberal distortions can remain to some extent even for large break in the case of M2, M5 and M9-1. The ILR also has cases with liberal coverage rates (M9-1, D1 for small breaks). It can also be conservative (M1, M2, M3, M5, M7, M7-1, D2 and most cases for large breaks). Bai’s method is more liberal than ILR for small breaks with M9-1 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 bootstrap method yields more often smaller lengths (M2, M3-1, M7-1, M8), but Bai’s method yields smaller lengths in some other cases (M4, M9-1, D1). The ILR delivers confidence intervals with lengths that are, in general, a close second relative to the best of the bootstrap and Bai’s method in most cases while it provides the shortest confidence intervals for M1,
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 some practically relevant 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 and M3-1 (change in mean model with serial correlation in the errors), M7, M7-1 and M8 (models with serial correlation and a random regressors such that the signal to noise ratio is high), M9 and M9-1 (AR(1) models corresponding to M3 and M3-1, respectively), and D1 (model with a lagged dependent included). It also occurs for M4 and D2 with larger values of $\delta$ from unreported simulations. The length can be more than half the whole sample even for moderate breaks in the case of M8. For Models M3-1, M7-1, M9-1, and D1, with an autoregressive coefficient 0.8, common in practice, the average length of the confidence intervals is the whole sample for any break size.

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 procedures that strike the best balance between coverage rate and length are the bootstrap and ILR. Bai’s method does 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% of the total sample 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. Adopting this view, the performances of the Bai, ILR and bootstrap methods are comparable, though for small breaks the bootstrap and ILR achieve a better balance between coverage rates and lengths.

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:

\[ y_t = x_t' \beta_t + z_t' \gamma + u_t; \quad \beta_t = \beta + \delta \mathbf{1}_{t>T_0} \tag{14} \]

where \( T_0 = [T \lambda_0] \) for some \( \lambda_0 \in (0, 1) \). We impose the following assumptions:

**Assumption A**: (1) \( u_t \) is stationary and ergodic with \( E(u_t) = 0, \ E(u_t^2 | F_t) = \sigma_u^2 \) and \( E(u_t^4) < \infty \), where \( F_t \) is the \( \sigma \)-field generated by \( \{ x_{t-s}, y_{t-s-1}, z_{t-s}, u_{t-s-1} | s \geq 0 \} \); (2) \( x_t \) satisfies \( \text{plim}_{T \to \infty} T^{-1} \sum_{t=1}^{T} x_t = c \), some constant \( p \times 1 \) vector and \( \text{plim}_{T \to \infty} T^{-1} \sum_{t=1}^{T} x_t x_t' = R(r) \), where \( R(r) \) is a nonsingular non-random \( p \times p \) matrix (with \( R(1) \equiv R \)); (3) \( \sup_{0 \leq s \leq 1} \| T^{-1/2} \sum_{t=1}^{T} z_t u_t \| = O_p(1) \), and \( \text{plim}_{T \to \infty} T^{-1} \sum_{t=1}^{T} z_t x_t' = s \Sigma_{zx} \), \( \text{plim}_{T \to \infty} T^{-1} \sum_{t=1}^{T} z_t z_t' = s \Sigma_{zz} \) uniformly in \( 0 \leq s \leq 1 \) where \( \Sigma_{zx} \) and \( \Sigma_{zz} \) are full rank.

Assumptions A (1-3) are standard high level assumptions in the literature. They are general enough to allow for serial correlation in both the regressors and the errors, as well as lagged dependent variables. Note that we consider the case of equal variance across segments, i.e., \( \omega_1^2 = \omega_2^2 \), 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: 

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

with \( \hat{\gamma}_j = T^{-1} \sum_{t=j+1}^{T} (\hat{u}_t - \bar{u})(\hat{u}_{t-j} - \bar{u}) \), where \( \hat{u}_t \) are the OLS residuals from regression (14), \( \bar{u} \) is their sample average, \( k(j, m) \) is some kernel function with the bandwidth \( m \). 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 \( C(\delta) = 4 \tilde{\rho}(\delta)^2 / (1 - \tilde{\rho}(\delta)^2)^2 \) with \( \tilde{\rho}(\delta) \) 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-sample. It is useful to define the following condition on the regressors, given that the theoretical results depend on whether it holds or not.

**Condition C1**: \( \delta' R_1 \delta / \delta' R \delta \to 1 \) as \( \| \delta \| \to \infty \) with \( R_j \equiv \text{plim}_{T \to \infty} T^{-1} \sum_{t=j+1}^{T} x_t x_{t-j}' \) \( (R = R_0) \).
As shown in Perron (1991), condition C1 induces a bias of \( \hat{\rho}(\delta) \) towards one. The larger the magnitude of a change \( ||\delta|| \) gets, the faster \( \hat{\rho}(\delta) \) goes to one. Since the bandwidth \( m \) is proportional to \( C(\delta) = 4\hat{\rho}(\delta)^2/(1 - \hat{\rho}(\delta)^2) \), it is an increasing function of \( ||\delta|| \) so that more covariance terms are included to estimate the long-run variance \( \omega^2 \). Hence, \( \hat{\omega}^2 \) is an increasing function of \( ||\delta|| \) and the value of \( \hat{U}_T \) eventually goes to zero as the magnitude of the change increases. Note that C1 holds when the only regressor subject to change is a constant. For the stochastic regressors \( x_t \) (\( p \times 1 \)), we define the signal to noise ratio as \( \Gamma_x^{-1/2}||\mu_x|| \) where \( \Gamma_x = E[(x_t - \mu_x)(x_t - \mu_x)^T] \) and \( \mu_x = E(x_t) \). Condition C1 holds approximately with a stochastic regressor (\( p = 1 \)) as the signal to noise ratio increases.\(^2\) In the multivariate case, condition C1 can hold approximately, for instance when the change in the coefficient of a regressor with a strong signal to noise ratio dominates the changes in the coefficients associated with regressors having a weak signal to noise ratio. As an example, Musso, Stracca and van Dick (2009) study the instability and nonlinearity in the Euro Area Phillips curve. A specification they considered is the simple model and van Dick (2009) study the instability and nonlinearity in the Euro Area Phillips curve.

The main results, proved in the Appendix, are stated in the following theorem.

**Theorem 1** Let the data be generated by (14) satisfying assumption A 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,

\[
T_{\delta-1}^1\hat{U}_T(T_1) = \frac{\lambda_0^2(\lambda_0 - \lambda)^2}{3\lambda^3} \delta' R' \delta \sigma_p(\|\delta\|^2 + 2) + o_p(1) \quad \text{for} \lambda > \lambda_0,
\]

and

\[
T_{\delta-1}^1\hat{U}_T(T_1) = \frac{(1 - \lambda_0)^2(\lambda_0 - \lambda)^2}{3(1 - \lambda)^3} \delta' R' \delta \sigma_p(\|\delta\|^2 + 2) + o_p(1) \quad \text{for} \lambda < \lambda_0
\]

(ii) if C1 is not satisfied,

\[
T_{\delta-1}^1\hat{U}_T(T_1) = \frac{\lambda_0^2(\lambda_0 - \lambda)^2}{3\lambda^3} \delta' R' \delta \sigma_p(\|\delta\|^2) + o_p(1) \quad \text{for} \lambda > \lambda_0,
\]

and

\[
T_{\delta-1}^1\hat{U}_T(T_1) = \frac{(1 - \lambda_0)^2(\lambda_0 - \lambda)^2}{3(1 - \lambda)^3} \delta' R' \delta \sigma_p(\|\delta\|^2) + o_p(1) \quad \text{for} \lambda < \lambda_0
\]

\(^2\)Suppose that \( x_t \) is a stationary process as in DGP M7 defined in Section 3. By a Weak Law of Large Numbers, we have \( R_0 = \lim_{T \to \infty} T^{-1} \sum_{t=1}^T x_t^2 = E(x_t^2) = \text{var}(x_t) + [E(x_t)]^2 = \sigma_x^2/(1 - \rho_x^2) + (\mu/(1 - \rho_x))^2 \), and similarly, \( R_1 = \lim_{T \to \infty} T^{-1} \sum_{t=2}^T x_t x_{t-1} = E(x_t x_{t-1}) = \text{cov}(x_t, x_{t-1}) + E(x_t) E(x_{t-1}) = (\sigma_x^2 \rho_x)/(1 - \rho_x^2) + (\mu/(1 - \rho_x))^2 \). After some algebra, we have \( R_1/R = 1 - (1 - \rho_x)/(1 + E(x_t)^2/\text{var}(x_t))^{-1} \). As the signal to noise ratio \( E(x_t)/\text{var}(x_t) \) increases, \( R_1/R \to 1 \); hence condition C1 holds approximately. In the case of a constant regressor, \( \text{var}(x_t) = 0 \) so that the signal to noise ratio is infinity.

15
where in all cases $o_p(1)$ is a remainder term that converges to 0 as $T \to \infty$ uniformly in $\|\delta\|$.

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, for any fixed $T$, the limit of $\hat{U}_T$ will be zero as the magnitude of the break, $\|\delta\|$, increases if condition C1 is satisfied, see equations (15) and (16). Hence, in this case the length of the confidence interval will be the whole sample. This provides a theoretical explanation for 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 (Models M3 and M3-1) or in the case of testing for a change in the coefficient of a stationary regressor when the signal to noise ratio is high (Models M7, M7-1, 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 $\hat{\rho}(\delta)$ towards unity and more so as $\|\delta\|$ 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 $\|\delta\|$, so that $\hat{\omega}^2$ is an increasing function of $\|\delta\|$. 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 estimate $\hat{\omega}_1^2$ and $\hat{\omega}_2^2$ for each candidate sub-sample, since one or the other will be contaminated by the unaccounted shift.

When condition C1 does not hold, for a fixed $T$, as $\|\delta\|$ increases, the limit of the test statistic is bounded above zero and the power depends on various factors, see equation (17) and (18). In general, the power of the test will approach one as $\|\delta\|$ increases and the length of the confidence interval is accordingly not the whole sample.

### 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:

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

for $t = 1, \ldots, T$, where $y_0 = 0$ for simplicity. Note that no change in the coefficient of the lagged dependent variable, $\alpha$, is allowed without loss of generality. 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 assumption A. Then, for a candidate break date \( T_1 = [T \lambda] \) with \( \lambda \in (0, 1) \) such that \( \lambda \neq \lambda_0 \): (i) for \( \lambda > \lambda_0 \),

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

uniformly in \( \|\delta\| \), where \( R = T^{-1} \sum_{t=1}^T x_t x_t' \); (ii) for \( \lambda < \lambda_0 \), uniformly in \( \|\delta\| \),

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

Note that the value of \( T^{-1} \hat{U}_T(T_1) \) 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 consistent for any fixed \( \|\delta\| \). 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 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 as \( \|\delta\| \) increases 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 values of \( \hat{U}_T(T_1) \) for a variety of cases.

In Figure 1, the rejection probabilities of \( \hat{U}_T(T_1) \) are plotted for \( \alpha \in \{0.3, 0.5, 0.8\} \). Panels (a) and (b) show the rejection probabilities of \( \hat{U}_T(T_1) \) for two candidate break dates \( T_1 = [0.3T] \) and \( [0.7T] \), respectively. The test \( \hat{U}_T(T_1) \) has some power with small changes, but power is reduced to zero as the magnitude of the change increases for all values of \( \alpha \). This implies that the candidate break dates \( T_1 = [0.3T] \) and \( [0.7T] \) will be included in the confidence set when the change is large. Of importance is the fact that the power function decreases substantially as \( \alpha \) increases. When \( \alpha = 0.8 \), power is less than size for all values of the break magnitude. Hence, the length of the confidence set will approach the whole sample more quickly as \( \|\delta\| \) increases, the closer \( \alpha \) is to one. 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 2 presents the graphs of the values of \( \hat{U}_T(T_1) \) as a function of \( \alpha \), 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 \( \lambda \) 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 \( \alpha \) approaches one.

17
Figure 3 shows the values 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 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 $\alpha$ and $T$. While the value of $\hat{U}_T(T_1)$ 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 $\alpha$ 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 set 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 $\hat{\alpha}$ towards unity and more so as $||\delta||$ increases. This makes the structural change appear 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 a 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. As argued in Perron (2006), one should also abandon partial-sums type tests. 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

Note that all limit statements are taken as $T \to \infty$ and $\xrightarrow{p}$ denotes convergence in probability.

**Proof of Theorem 1:** The data-generating process is defined by (14). In matrix notation,

$$
Y = \begin{bmatrix}
x'_{i_0} & 0 & z'_{i_0} \\
\vdots & \vdots & \vdots \\
x'_{T_1} & 0 & z'_{T_1} \\
x'_{T_1+1} & x'_{T_1+1} & z'_{T_1+1} \\
\vdots & \vdots & \vdots \\
x'_{T} & x'_{T} & z'_{T}
\end{bmatrix}
\begin{bmatrix}
[\beta] \\
[\delta] \\
[\gamma]
\end{bmatrix}
+ U = \begin{bmatrix} X_0 & Z \end{bmatrix}
\begin{bmatrix}
[\beta] \\
[\delta] \\
[\gamma]
\end{bmatrix}
+ U = M_0 \Gamma + U.
$$

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 $\Gamma$ is given by

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

and the OLS residuals are

$$
\hat{U} = Y - \hat{Y} = U + M_0 \Gamma - M \hat{\Gamma}
= U + M \Gamma - M \hat{\Gamma} - M \hat{\Gamma} + M_0 \Gamma = U - M (\hat{\Gamma} - \Gamma) - (M - M_0) \Gamma
= U + M (M' M)^{-1} M' (M - M_0) \Gamma - M (M' M)^{-1} M' U - (M - M_0) \Gamma
= [I - M (M' M)^{-1} M'] [U - (M - M_0) \Gamma].
$$

(A.1)

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

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

where $\tilde{Z} = (\tilde{z}_1, ..., \tilde{z}_T)$ with $\tilde{z}_t$ the OLS residuals from a regression of $z'_t$ on $\{x'_{t}, x'_{t:1:T} \}$. We can then write (A.1) as:

$$
\hat{U} = (I - X_1 (X'_{1} X_1)^{-1} X'_1 - \tilde{Z} (\tilde{Z}' \tilde{Z})^{-1} \tilde{Z}') U
- (I - X_1 (X'_{1} X_1)^{-1} X'_1 - \tilde{Z} (\tilde{Z}' \tilde{Z})^{-1} \tilde{Z}')(M - M_0) \Gamma.
$$

(A.2)

Using a Weak Law of Large Number (WLLN), it is sufficient to consider the second term in (A.2). Consider first the case $T_1 < T_1^0$, i.e., $\lambda < \lambda_0$. Note that $(M - M_0) \Gamma = (0, \ldots, 0, (x'_{T_1+1} \delta)'$,
where \( A = (\sum_{t=1}^{T_1} x_t'x_t)^{-1} \) and \( B = (\sum_{t=T_1+1}^{T} x_t'x_t)^{-1} \). For \( 0 < s \leq \lambda \), \( T^{-1} \sum_{t=1}^{[Ts]} x_t \hat{u}_t = o_p(1) \).

For \( \lambda < s \leq \lambda_0 \),

\[
T^{-1} \sum_{t=[T\lambda]+1}^{[Ts]} x_t \hat{u}_t = -T^{-1} \sum_{t=[T\lambda]+1}^{[Ts]} x_t x_t' \delta
\]

\[
+ \left( T^{-1} \sum_{t=[T\lambda]+1}^{[Ts]} x_t x_t' \right) \left( T^{-1} \sum_{t=[T\lambda]+1}^{T} x_t x_t' \right)^{-1} \left( T^{-1} \sum_{t=[T\lambda]+1}^{[Ts]} x_t x_t' \delta \right)
\]

\[
+ \left( T^{-1} \sum_{t=[T\lambda]+1}^{[Ts]} x_t z_t' \right) \left( T^{-1} \sum_{t=1}^{T} \tilde{z}_t \tilde{z}_t' \right)^{-1} \left( T^{-1} \sum_{t=[T\lambda]+1}^{[Ts]} \tilde{z}_t x_t' \delta \right). \tag{A.3}
\]

Consider the third term in (A.3). For \( t \leq T_1 = [Ts] \), \( \tilde{z}_t = z_t = z_t - (\sum_{s=1}^{[Ts]} z_s x_s')(\sum_{s=1}^{[Ts]} x_s x_s')^{-1} x_t \). Similarly, for \( t > T_1 \), \( \tilde{z}_t = z_t - (\sum_{s=[T\lambda]+1}^{T} z_s x_s')(\sum_{s=[T\lambda]+1}^{[Ts]} x_s x_s')^{-1} x_t \). 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

\[
\sup_{T_1} \left| \left( \sum_{t=[T\lambda]+1}^{[Ts]} x_t z_t' \right) \left( \sum_{t=1}^{T} \tilde{z}_t \tilde{z}_t' \right)^{-1} \left( \sum_{t=[T\lambda]+1}^{[Ts]} \tilde{z}_t x_t' \delta \right) - \left( \sum_{t=[T\lambda]+1}^{[Ts]} x_t z_t' \right) \left( \sum_{t=1}^{T} \tilde{z}_t \tilde{z}_t' \right)^{-1} \left( \sum_{t=[T\lambda]+1}^{[Ts]} \tilde{z}_t x_t' \delta \right) \right| \to 0
\]

where \( \tilde{z}_t = z_t - \Sigma_{xx} R^{-1} x_t \) and \( \Sigma_{xx} = \text{plim}_{T \to \infty} T^{-1} \sum_{t=1}^{T} z_t x_t' \). Note that \( \tilde{z}_t \) does not depend on \( T_1 \) and the third term in (A.3) is \( o_p(1) \) as \( T \to \infty \). Therefore,

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

21
Similarly, for \( \lambda_0 < s \leq 1 \),
\[
T^{-1} \sum_{t=[T\lambda]+1}^{[Ts]} x_t u_t = -T^{-1} \sum_{t=[T\lambda]+1}^{[T\lambda_0]} x_t x_t' \delta \\
+ \left( T^{-1} \sum_{t=[T\lambda]+1}^{[Ts]} x_t x_t' \right) \left( T^{-1} \sum_{t=[T\lambda]+1}^{T} x_t x_t' \right)^{-1} \left( T^{-1} \sum_{t=[T\lambda]+1}^{[T\lambda_0]} x_t x_t' \delta \right) + o_p(1) \]
\[
\overset{p}{\rightarrow} (s - 1) \left( \frac{\lambda_0 - \lambda}{1 - \lambda} \right) R \delta.
\]

Next, consider the case where \( T_1 > T_1^0 \), that is, \( \lambda > \lambda_0 \). Note that \( (M - M_0)\Gamma = (0, \ldots, 0, -(x'_{T+1} \delta)', \ldots, -(x'_{T} \delta)', 0, \ldots, 0)' \). For \( \lambda < s \leq 1 \), \( T^{-1} \sum_{t=[T\lambda]+1}^{[Ts]} x_t u_t = o_p(1) \).

On the other hand, for \( 0 < s \leq \lambda_0 \), \( T^{-1} \sum_{t=1}^{[Ts]} x_t u_t \overset{p}{\rightarrow} -s(\lambda - \lambda_0)/\lambda \), and for \( \lambda_0 < s \leq \lambda \), \( T^{-1} \sum_{t=1}^{[Ts]} x_t u_t \overset{p}{\rightarrow} (s - \lambda_0)\lambda_0/\lambda \). It is easy to show that
\[
\hat{\Gamma}_0 = T^{-1} \hat{U}' \hat{U} \overset{p}{\rightarrow} \gamma_0 + (\lambda_0 - \lambda) \frac{1 - \lambda_0}{1 - \lambda} \delta' R \delta
\]
because \( T^{-1} U'M = o_p(1) \) and \( T^{-1} U'M_0 = o_p(1) \) by a WLLN. Also,
\[
\hat{\gamma}_1 = T^{-1} \sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1} \overset{p}{\rightarrow} \gamma_1 + (\lambda_0 - \lambda) \frac{1 - \lambda_0}{1 - \lambda} \delta' R_1 \delta.
\]

Using these results, the estimate of the first-order autocorrelation of the errors is such that
\[
\hat{\rho}(\delta) = \frac{T^{-1} \sum_{t=2}^{T} (\hat{u}_t - \bar{u})(\hat{u}_{t-1} - \bar{u})}{T^{-1} \sum_{t=2}^{T} (\hat{u}_t - \bar{u})^2} \overset{p}{\rightarrow} \frac{\gamma_1 + (\lambda_0 - \lambda) \frac{1 - \lambda_0}{1 - \lambda} \delta' R_1 \delta}{\gamma_0 + (\lambda_0 - \lambda) \frac{1 - \lambda_0}{1 - \lambda} \delta' R \delta}.
\]

The test statistic evaluated at a candidate break date \( T_1 = [T\lambda] \) can be expressed as:
\[
\hat{U}_T(T_1) = \frac{T^2}{[T\lambda]^2} \sum_{t=1}^{[T\lambda]} \left( T^{-1} \sum_{s=1}^{t} x_s \hat{u}_s \right)' \left( \hat{\omega}^2 \right)^{-1} \left( T^{-1} \sum_{s=1}^{t} x_s \hat{u}_s \right) \\
+ \frac{T^2}{(T - [T\lambda])^2} \sum_{t=[T\lambda]+1}^{T} \left( T^{-1} \sum_{s=[T\lambda]+1}^{t} x_s \hat{u}_s \right)' \left( \hat{\omega}^2 \right)^{-1} \left( T^{-1} \sum_{s=[T\lambda]+1}^{t} x_s \hat{u}_s \right).
\]
When using Andrews’s AR(1) approximation, \( m \propto (C(\delta)T)^{1/\theta} \) 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 \( \sum_{j=1}^{T-1} k(j, m) = O(m) \),

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

\[
= \left( \gamma_0 + 2 \sum_{j=1}^{T-1} k(j, m) \gamma_j \right) + (\lambda_0 - \lambda) \frac{1 - \lambda_0}{1 - \lambda} \left( \delta' R \delta + 2 \sum_{j=1}^{T-1} k(j, m) \delta' R_j \delta \right) + o_p(1)
\]

\[
= h(0) + (\lambda_0 - \lambda) \frac{1 - \lambda_0}{1 - \lambda} \left( \delta' R \delta + 2 \sum_{j=1}^{T-1} k(j, m) \delta' R_j \delta \right) + o_p(1)
\]

\[
\leq h(0) + (\lambda_0 - \lambda) \frac{1 - \lambda_0}{1 - \lambda} \delta' R \delta O(m) + o_p(\|\delta\|^2)
\]

where \( h(0) = \lim_{T \to \infty} \text{var}(T^{-1/2} \sum_{t=1}^{T} u_t) \), 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 condition C1 holds, then \( \hat{\rho}(\delta) \xrightarrow{p} 1 \) in (A.4) 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 \( \hat{\omega}^2 = O_p(\|\delta\|^{2+2/\theta} T^{1/\theta}) \). If C1 does not hold, \( \hat{\rho}(\delta) \xrightarrow{p} \rho^* < 1 \), say, as \( \|\delta\| \) increases such that \( C(\delta) = O_p(1), m = O_p(T^{1/\theta}) \), and \( \hat{\omega}^2 = O_p(\|\delta\|^2 T^{1/\theta}) \). 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

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

First, assume that \( T_1 < T_1^0 \). Then,

\[
T^{-1} \hat{U}' \hat{U} = T^{-1} U' U + \delta' \left( T^{-1} \sum_{t=T_1+1}^{T_1^0} x_t x_t' \right) \delta
\]

\[
- \delta' \left( T^{-1} \sum_{t=T_1+1}^{T_1^0} x_t x_t' \right) \left( T^{-1} \sum_{t=T_1+1}^{T_1^0} x_t x_t' \right)^{-1} \left( T^{-1} \sum_{t=T_1+1}^{T_1^0} x_t x_t' \right) \delta + o_p(1)
\]

\[
\xrightarrow{p} \sigma_u^2 + (\lambda_0 - \lambda) \delta ' R \delta - (\lambda_0 - \lambda)^2 \delta ' R \delta = \sigma_u^2 + (1 + \lambda - \lambda_0)(\lambda_0 - \lambda) \delta ' R \delta.
\]
On the other hand, if $T_1 > T_0^1$, then
\[
T^{-1} \hat{U}' \hat{U} = T^{-1} U'U + \delta' \left( T^{-1} \sum_{t=T_0^1}^{T_1} x_t' x_t \right) \delta
- \delta' \left( T^{-1} \sum_{t=T_0^1+1}^{T_1} x_t' x_t \right) \left( T^{-1} \sum_{t=1}^{T_1} x_t' x_t \right)^{-1} \left( T^{-1} \sum_{t=T_0^1+1}^{T_1} x_t x_t \right) \delta + o_p(1)
\overset{p}{\to} \sigma_u^2 + (\lambda - \lambda_0) \delta'R\delta - (\lambda - \lambda_0)^2 \delta'R\delta = \sigma_u^2 + (1 - \lambda + \lambda_0)(\lambda - \lambda_0) \delta'R\delta.
\]

We can rewrite the OLS residuals $\hat{U}$ as follows:
\[
\hat{U} = [I - X_1(X'_1 X_1)^{-1} X'_1 - \tilde{Y}_{-1} (\tilde{Y}'_{-1} \tilde{Y}_{-1})^{-1} \tilde{Y}'_{-1}] U
- [I - X_1(X'_1 X_1)^{-1} X'_1 - \tilde{Y}_{-1} (\tilde{Y}'_{-1} \tilde{Y}_{-1})^{-1} \tilde{Y}'_{-1}] (M - M_0) \Gamma
\]
where $\tilde{Y}_{-1} = (\tilde{y}_0, ..., \tilde{y}_{T-1})$ with $\tilde{y}_{T-1}$ the OLS residuals from a regression of $y_{t-1}$ on $\{x'_t, x'_t 1_{t>T_1}\}$. Using a WLLN, it is sufficient to consider the second term in $\hat{U}$. First, consider the case where $T_1 < T_0$. For $0 < s < \lambda$, $T^{-1} \sum_{t=1}^{[Ts]} x_t \hat{u}_t = o_p(1)$. Moreover, under assumption A, we can show that $T^{-1} \sum_{t=[T\lambda]+1}^{[Ts]} x_t \hat{u}_t \overset{p}{\to} s \Sigma_{yx}$. Hence, for $\lambda < s \leq \lambda_0$, $T^{-1} \sum_{t=[T\lambda]+1}^{[Ts]} x_t \hat{u}_t \overset{p}{\to} (s - \lambda)((-1+\lambda_0)/(1-\lambda)) R\delta$, and for $\lambda_0 < s \leq 1$, $T^{-1} \sum_{t=[T\lambda]+1}^{[Ts]} x_t \hat{u}_t \overset{p}{\to} (s-1)((\lambda_0 - \lambda)/(1-\lambda)) R\delta$. Therefore,
\[
T^{-1} \hat{U}_T(T_1) = \frac{1}{1-\lambda^2} \left\{ \left( \frac{1-\lambda^0}{1-\lambda} \right)^2 \int_0^{\lambda^0} (s - \lambda)^2 ds + \left( \frac{-\lambda^0}{1-\lambda} \right)^2 \int_0^{1} (s-1)^2 ds \right\} \delta'R'R\delta
\sigma_u^2 + (1 + \lambda - \lambda_0)(\lambda_0 - \lambda) \delta'R\delta + o_p(1).
\]

Next, consider the case where $T_1 > T_0^1$. As shown in the proof of Theorem 1, for $\lambda < s \leq 1$, $T^{-1} \sum_{t=[T\lambda]+1}^{[Ts]} x_t \hat{u}_t = o_p(1)$. Moreover, for $0 < s \leq \lambda_0$, $T^{-1} \sum_{t=1}^{[Ts]} x_t \hat{u}_t \overset{p}{\to} -s(\lambda - \lambda_0)/\lambda$, and for $\lambda_0 < s \leq \lambda$, $T^{-1} \sum_{t=1}^{[Ts]} x_t \hat{u}_t \overset{p}{\to} (s - \lambda_0) \lambda_0/\lambda$. Hence,
\[
T^{-1} \hat{U}_T(T_1) = \frac{1}{\lambda^2} \left\{ \left( \frac{1-\lambda^0}{\lambda} \right)^2 \int_0^{\lambda^0} s^2 ds + \left( \frac{\lambda^0}{\lambda} \right)^2 \int_0^{\lambda} (s-\lambda_0)^2 ds \right\} \delta'R'R\delta
\sigma_u^2 + (1 - \lambda + \lambda_0)(\lambda - \lambda_0) \delta'R\delta + o_p(1),
\]
which completes the proof.
References


### Table 1: DGP M1

<table>
<thead>
<tr>
<th>$\delta = dT^{1/2}$</th>
<th>0.4</th>
<th>0.8</th>
<th>1.2</th>
<th>1.6</th>
<th>2.0</th>
<th>2.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U_T\neq$</td>
<td>0.948</td>
<td>76.9</td>
<td>0.955</td>
<td>41.4</td>
<td>0.953</td>
<td>22.6</td>
</tr>
<tr>
<td>Bai</td>
<td>0.874</td>
<td>72.1</td>
<td>0.906</td>
<td>35.3</td>
<td>0.938</td>
<td>15.9</td>
</tr>
<tr>
<td>PB</td>
<td>0.984</td>
<td>59.6</td>
<td>0.960</td>
<td>37.1</td>
<td>0.963</td>
<td>17.8</td>
</tr>
<tr>
<td>ILR</td>
<td>0.960</td>
<td>59.4</td>
<td>0.965</td>
<td>30.4</td>
<td>0.972</td>
<td>12.5</td>
</tr>
<tr>
<td>sup $W$</td>
<td>0.345</td>
<td>90.7</td>
<td>0.997</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

$U_T\neq$ refers to the rejection probability of the $sup$-Wald test using the 5% asymptotic critical value. The number of simulations is 3,000.

Note: $y_t = \delta 1_{t>0.5T} + u_t$, $u_t \sim i.i.d. N(0, 1)$, and $T = 100$. Cov. and Lght. denote the coverage probability and average length of the confidence intervals (or average number of dates in the confidence set). $sup W$ refers to the rejection probability of the sup-Wald test using the 5% asymptotic critical value. The number of simulations is 3,000.

### Table 2: DGP M2

<table>
<thead>
<tr>
<th>$\delta = dT^{1/2}$</th>
<th>0.4</th>
<th>0.8</th>
<th>1.2</th>
<th>1.6</th>
<th>2.0</th>
<th>2.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U_T\neq$</td>
<td>0.949</td>
<td>84.5</td>
<td>0.946</td>
<td>67.3</td>
<td>0.955</td>
<td>44.7</td>
</tr>
<tr>
<td>Bai</td>
<td>0.829</td>
<td>60.6</td>
<td>0.841</td>
<td>33.9</td>
<td>0.868</td>
<td>17.2</td>
</tr>
<tr>
<td>PB</td>
<td>0.934</td>
<td>52.0</td>
<td>0.869</td>
<td>43.1</td>
<td>0.817</td>
<td>32.7</td>
</tr>
<tr>
<td>ILR</td>
<td>0.989</td>
<td>68.1</td>
<td>0.973</td>
<td>58.1</td>
<td>0.984</td>
<td>37.7</td>
</tr>
<tr>
<td>sup $W$</td>
<td>0.482</td>
<td>0.860</td>
<td>0.985</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

$U_T\neq$ refers to the rejection probability of the $sup$-Wald test using the 5% asymptotic critical value. The number of simulations is 3,000.

Note: $y_t = \delta 1_{t>0.5T} + u_t$, $u_t = (1 + 1_{t>0.5T})\epsilon_t$, $\epsilon_t \sim i.i.d. N(0, 1)$, and $T = 100$. See notes to Table 1.
Table 3: DGP M3 (AR coefficient = 0.3)

<table>
<thead>
<tr>
<th>$\delta = dT^{1/2}$</th>
<th>0.4</th>
<th>0.8</th>
<th>1.2</th>
<th>1.6</th>
<th>2.0</th>
<th>2.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U T\neq$</td>
<td>0.977</td>
<td>89.8</td>
<td>0.973</td>
<td>82.8</td>
<td>0.977</td>
<td>74.9</td>
</tr>
<tr>
<td>Bai</td>
<td>0.821</td>
<td>68.9</td>
<td>0.802</td>
<td>35.0</td>
<td>0.850</td>
<td>19.1</td>
</tr>
<tr>
<td>PB</td>
<td>0.983</td>
<td>57.2</td>
<td>0.948</td>
<td>33.6</td>
<td>0.945</td>
<td>15.7</td>
</tr>
<tr>
<td>ILR</td>
<td>0.950</td>
<td>57.8</td>
<td>0.965</td>
<td>29.3</td>
<td>0.973</td>
<td>13.3</td>
</tr>
<tr>
<td>sup $W$</td>
<td>0.464</td>
<td>0.936</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

$^U T\neq$ | 2.8 | 3.2 | 3.6 | 4.0 | 4.4 | 4.8 |

$U T\neq$ | 0.971 | 68.6 | 0.975 | 71.9 | 0.974 | 76.4 | 0.974 | 80.0 | 0.970 | 82.8 | 0.974 | 85.3 |
| Bai       | 0.968 | 8.6  | 0.983 | 7.4  | 0.988 | 7.6  | 0.996 | 6.9  | 0.995 | 6.9  | 0.996 | 7.2  |
| PB        | 0.952 | 1.2  | 0.970 | 0.4  | 0.987 | 0.1  | 0.994 | 0.0  | 0.999 | 0.0  | 0.999 | 0.0  |
| ILR       | 0.997 | 2.4  | 1.000 | 1.9  | 1.000 | 1.4  | 1.000 | 1.2  | 1.000 | 1.1  | 1.000 | 1.1  |
| sup $W$   | 1     | 1    | 1     | 1     | 1     | 1     | 1     | 1     |

Note: $y_t = \delta 1_{t>[0.57]} + u_t, u_t = 0.3u_{t-1} + \varepsilon_t, \varepsilon_t \sim i.i.d. N(0, 0.49), T = 100$. See notes to Table 1.

Table 4: DGP M3-1 (AR coefficient = 0.8)

<table>
<thead>
<tr>
<th>$\delta = dT^{1/2}$</th>
<th>0.4</th>
<th>0.8</th>
<th>1.2</th>
<th>1.6</th>
<th>2.0</th>
<th>2.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U T\neq$</td>
<td>0.990</td>
<td>91.5</td>
<td>0.990</td>
<td>91.2</td>
<td>0.984</td>
<td>92.2</td>
</tr>
<tr>
<td>Bai</td>
<td>0.822</td>
<td>68.9</td>
<td>0.890</td>
<td>45.4</td>
<td>0.964</td>
<td>35.0</td>
</tr>
<tr>
<td>PB</td>
<td>0.911</td>
<td>46.4</td>
<td>0.898</td>
<td>18.5</td>
<td>0.941</td>
<td>4.1</td>
</tr>
<tr>
<td>ILR</td>
<td>0.942</td>
<td>53.9</td>
<td>0.983</td>
<td>30.2</td>
<td>0.999</td>
<td>15.3</td>
</tr>
<tr>
<td>sup $W$</td>
<td>0.678</td>
<td>0.984</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

$^U T\neq$ | 2.8 | 3.2 | 3.6 | 4.0 | 4.4 | 4.8 |

$U T\neq$ | 0.987 | 92.9 | 0.990 | 93.0 | 0.985 | 92.8 | 0.987 | 92.7 | 0.988 | 92.5 | 0.992 | 92.5 |
| Bai       | 0.998 | 15.3 | 0.997 | 13.1 | 0.992 | 11.4 | 0.986 | 11.4 | 0.975 | 8.8  | 0.971 | 8.9  |
| PB        | 1.000 | 0.0  | 1.000 | 0.0  | 1.000 | 0.0  | 1.000 | 0.0  | 1.000 | 0.0  | 1.000 | 0.0  |
| ILR       | 0.997 | 2.4  | 1.000 | 1.8  | 1.000 | 1.5  | 1.000 | 1.3  | 1.000 | 1.3  | 1.000 | 1.2  |
| sup $W$   | 1     | 1    | 1     | 1     | 1     | 1     | 1     |

Note: $y_t = \delta 1_{t>[0.57]} + u_t, u_t = 0.8u_{t-1} + \varepsilon_t, \varepsilon_t \sim i.i.d. N(0, 0.04), T = 100$. See notes to Table 1.
Table 5: DGP M4

<table>
<thead>
<tr>
<th>$\delta = dT^{1/2}$</th>
<th>0.4</th>
<th>0.8</th>
<th>1.2</th>
<th>1.6</th>
<th>2.0</th>
<th>2.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{U}_T:neq$</td>
<td>0.977</td>
<td>81.8</td>
<td>0.976</td>
<td>54.4</td>
<td>0.977</td>
<td>33.9</td>
</tr>
<tr>
<td>Bai</td>
<td>0.882</td>
<td>76.7</td>
<td>0.868</td>
<td>38.6</td>
<td>0.913</td>
<td>19.1</td>
</tr>
<tr>
<td>PB</td>
<td>0.999</td>
<td>63.7</td>
<td>0.997</td>
<td>41.9</td>
<td>0.996</td>
<td>19.8</td>
</tr>
<tr>
<td>ILR</td>
<td>0.953</td>
<td>65.4</td>
<td>0.947</td>
<td>57.5</td>
<td>0.947</td>
<td>44.6</td>
</tr>
<tr>
<td>sup $W$</td>
<td>0.276</td>
<td>0.877</td>
<td>0.999</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2.8</td>
<td>3.2</td>
<td>3.6</td>
<td>4.0</td>
<td>4.4</td>
<td>4.8</td>
</tr>
</tbody>
</table>

Note: $y_t = \delta 1_{t>[0.5T]} + u_t, \ u_t = \epsilon_t - 0.3\epsilon_{t-1}, \ \epsilon_t \sim i.i.d. \ N(0,2.04), \ T = 100$. See notes to Table 1.

Table 6: DGP M5

<table>
<thead>
<tr>
<th>$\delta = dT^{1/2}$</th>
<th>0.4</th>
<th>0.8</th>
<th>1.2</th>
<th>1.6</th>
<th>2.0</th>
<th>2.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{U}_T:neq$</td>
<td>0.957</td>
<td>79.5</td>
<td>0.956</td>
<td>51.1</td>
<td>0.954</td>
<td>31.8</td>
</tr>
<tr>
<td>Bai</td>
<td>0.865</td>
<td>68.3</td>
<td>0.894</td>
<td>34.5</td>
<td>0.886</td>
<td>16.5</td>
</tr>
<tr>
<td>PB</td>
<td>0.993</td>
<td>59.2</td>
<td>0.968</td>
<td>36.7</td>
<td>0.954</td>
<td>19.4</td>
</tr>
<tr>
<td>ILR</td>
<td>0.968</td>
<td>60.9</td>
<td>0.970</td>
<td>34.2</td>
<td>0.977</td>
<td>16.5</td>
</tr>
<tr>
<td>sup $W$</td>
<td>0.361</td>
<td>0.897</td>
<td>0.997</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2.8</td>
<td>3.2</td>
<td>3.6</td>
<td>4.0</td>
<td>4.4</td>
<td>4.8</td>
</tr>
</tbody>
</table>

Note: $y_t = x_t\delta 1_{t>[0.5T]} + \gamma + u_t, \ x_t = 0.5x_{t-1} + v_t, \ v_t \sim i.i.d. \ N(0,0.75), \ \gamma = 1, \ u_t \sim i.i.d. \ N(0,1)$ independent of $\{x_t\}, \ T = 100$. See notes to Table 1.
Table 7: DGP M6

<table>
<thead>
<tr>
<th>$\delta = dT^{1/2}$</th>
<th>0.4</th>
<th>0.8</th>
<th>1.2</th>
<th>1.6</th>
<th>2.0</th>
<th>2.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U_{t, \neq}$</td>
<td>0.962</td>
<td>77.9</td>
<td>0.962</td>
<td>45.8</td>
<td>0.958</td>
<td>28.3</td>
</tr>
<tr>
<td>Bai</td>
<td>0.846</td>
<td>65.6</td>
<td>0.864</td>
<td>29.1</td>
<td>0.945</td>
<td>14.1</td>
</tr>
<tr>
<td>PB</td>
<td>0.945</td>
<td>54.3</td>
<td>0.852</td>
<td>26.1</td>
<td>0.866</td>
<td>10.0</td>
</tr>
<tr>
<td>ILR</td>
<td>0.946</td>
<td>56.8</td>
<td>0.933</td>
<td>29.9</td>
<td>0.974</td>
<td>16.2</td>
</tr>
<tr>
<td>sup $W$</td>
<td>0.707</td>
<td>0.976</td>
<td>0.990</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\Upsilon_{t, \neq}$</td>
<td>0.963</td>
<td>14.5</td>
<td>0.963</td>
<td>13.3</td>
<td>0.960</td>
<td>12.6</td>
</tr>
<tr>
<td>Bai</td>
<td>0.987</td>
<td>3.2</td>
<td>0.989</td>
<td>2.6</td>
<td>0.992</td>
<td>2.4</td>
</tr>
<tr>
<td>PB</td>
<td>0.996</td>
<td>0.8</td>
<td>1.000</td>
<td>0.4</td>
<td>1.000</td>
<td>0.3</td>
</tr>
<tr>
<td>ILR</td>
<td>0.997</td>
<td>4.2</td>
<td>1.000</td>
<td>3.3</td>
<td>1.000</td>
<td>3.1</td>
</tr>
<tr>
<td>sup $W$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: $y_t = x_t \delta 1_{t>[0.5T]} + \gamma + u_t, \quad u_t = \epsilon_t|x_t|, \quad \epsilon_t \sim i.i.d. N(0,0.333), \quad x_t = 0.5x_{t-1} + v_t$ where $v_t \sim i.i.d. N(0,0.75), \gamma = 1, \quad T = 100$. See notes to Table 1.

Table 8: DGP M7 (AR coefficient = 0.3)

<table>
<thead>
<tr>
<th>$\delta = dT^{1/2}$</th>
<th>0.4</th>
<th>0.8</th>
<th>1.2</th>
<th>1.6</th>
<th>2.0</th>
<th>2.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U_{t, \neq}$</td>
<td>0.973</td>
<td>72.7</td>
<td>0.976</td>
<td>88.9</td>
<td>0.974</td>
<td>89.6</td>
</tr>
<tr>
<td>Bai</td>
<td>1.000</td>
<td>2.9</td>
<td>1.000</td>
<td>2.1</td>
<td>1.000</td>
<td>2.0</td>
</tr>
<tr>
<td>PB</td>
<td>0.950</td>
<td>8.1</td>
<td>0.986</td>
<td>0.4</td>
<td>0.997</td>
<td>0.0</td>
</tr>
<tr>
<td>ILR</td>
<td>0.999</td>
<td>2.0</td>
<td>1.000</td>
<td>1.1</td>
<td>1.000</td>
<td>1.0</td>
</tr>
<tr>
<td>sup $W$</td>
<td>0.281</td>
<td>0.466</td>
<td>0.574</td>
<td>0.669</td>
<td>0.719</td>
<td>0.749</td>
</tr>
<tr>
<td>$\Upsilon_{t, \neq}$</td>
<td>0.974</td>
<td>89.7</td>
<td>0.971</td>
<td>89.9</td>
<td>0.976</td>
<td>89.7</td>
</tr>
<tr>
<td>Bai</td>
<td>1.000</td>
<td>2.0</td>
<td>1.000</td>
<td>2.0</td>
<td>1.000</td>
<td>2.0</td>
</tr>
<tr>
<td>PB</td>
<td>1.000</td>
<td>0.0</td>
<td>1.000</td>
<td>0.0</td>
<td>1.000</td>
<td>0.0</td>
</tr>
<tr>
<td>ILR</td>
<td>1.000</td>
<td>1.0</td>
<td>1.000</td>
<td>1.0</td>
<td>1.000</td>
<td>1.0</td>
</tr>
<tr>
<td>sup $W$</td>
<td>0.819</td>
<td>0.827</td>
<td>0.885</td>
<td>0.874</td>
<td>0.903</td>
<td>0.928</td>
</tr>
</tbody>
</table>

Note: $y_t = x_t \delta 1_{t>[0.5T]} + \gamma + u_t, \quad u_t = 0.3u_{t-1} + \epsilon_t, \quad \epsilon_t \sim i.i.d. N(0,0.49), \quad x_t = \mu + 0.5x_{t-1} + v_t, \quad v_t \sim i.i.d. N(0,0.75), \gamma = 1, \mu = 5, \quad T = 100$. See notes to Table 1.
Table 9: DGP M7-1 (AR coefficient = 0.8)

<table>
<thead>
<tr>
<th>( \delta = dT^{1/2} )</th>
<th>0.4</th>
<th>0.8</th>
<th>1.2</th>
<th>1.6</th>
<th>2.0</th>
<th>2.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{U}._{T,:neq} )</td>
<td>0.983</td>
<td>91.1</td>
<td>0.981</td>
<td>90.6</td>
<td>0.985</td>
<td>90.4</td>
</tr>
<tr>
<td>Bai</td>
<td>0.940</td>
<td>11.6</td>
<td>0.990</td>
<td>3.8</td>
<td>1.000</td>
<td>2.6</td>
</tr>
<tr>
<td>PB</td>
<td>0.934</td>
<td>5.1</td>
<td>1.000</td>
<td>0.1</td>
<td>1.000</td>
<td>0.0</td>
</tr>
<tr>
<td>ILR</td>
<td>0.968</td>
<td>13.6</td>
<td>1.000</td>
<td>4.8</td>
<td>1.000</td>
<td>2.9</td>
</tr>
<tr>
<td>sup( W )</td>
<td>0.081</td>
<td>0.161</td>
<td>0.223</td>
<td>0.280</td>
<td>0.333</td>
<td>0.418</td>
</tr>
</tbody>
</table>

Note: \( y_t = x_t \delta_1 I_{t>0.5T} + \gamma + u_t, \ u_t = 0.8u_{t-1} + \epsilon_t, \ \epsilon_t \sim i.i.d. N(0,0.04), \ x_t = \mu + 0.5x_{t-1} + v_t, \ v_t \sim i.i.d. N(0,0.75), \gamma = 1, \mu = 5, T = 100 \). See notes to Table 1.

Table 10: DGP M8

<table>
<thead>
<tr>
<th>( \delta = dT^{1/2} )</th>
<th>0.4</th>
<th>0.8</th>
<th>1.2</th>
<th>1.6</th>
<th>2.0</th>
<th>2.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{U}._{T,:neq} )</td>
<td>0.982</td>
<td>90.3</td>
<td>0.980</td>
<td>90.4</td>
<td>0.987</td>
<td>90.6</td>
</tr>
<tr>
<td>Bai</td>
<td>1.000</td>
<td>2.0</td>
<td>1.000</td>
<td>2.0</td>
<td>1.000</td>
<td>2.0</td>
</tr>
<tr>
<td>PB</td>
<td>1.000</td>
<td>0.0</td>
<td>1.000</td>
<td>0.0</td>
<td>1.000</td>
<td>0.0</td>
</tr>
<tr>
<td>ILR</td>
<td>1.000</td>
<td>1.3</td>
<td>1.000</td>
<td>1.3</td>
<td>1.000</td>
<td>1.4</td>
</tr>
<tr>
<td>sup( W )</td>
<td>0.046</td>
<td>0.499</td>
<td>0.527</td>
<td>0.546</td>
<td>0.590</td>
<td>0.592</td>
</tr>
</tbody>
</table>

Note: \( y_t = x_t \delta_1 I_{t>0.5T} + \gamma + u_t, \ u_t = \epsilon_t - 0.3\epsilon_{t-1}, \ \epsilon_t \sim i.i.d. N(0,2.04), \ x_t = \mu + 0.5x_{t-1} + v_t, \ v_t \sim i.i.d. N(0,0.75), \gamma = 1, \mu = 5, T = 100 \). See notes to Table 1.
Table 11: DGP M9 (AR coefficient = 0.3)

<table>
<thead>
<tr>
<th>$\delta = dT^{1/2}$</th>
<th>0.4</th>
<th>0.8</th>
<th>1.2</th>
<th>1.6</th>
<th>2.0</th>
<th>2.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U_{T, neq}$</td>
<td>0.948</td>
<td>78.9</td>
<td>0.951</td>
<td>53.0</td>
<td>0.954</td>
<td>36.9</td>
</tr>
<tr>
<td>Bai</td>
<td>0.858</td>
<td>63.9</td>
<td>0.859</td>
<td>29.4</td>
<td>0.992</td>
<td>14.7</td>
</tr>
<tr>
<td>PB</td>
<td>0.988</td>
<td>58.9</td>
<td>0.948</td>
<td>38.8</td>
<td>0.961</td>
<td>21.1</td>
</tr>
<tr>
<td>ILR</td>
<td>0.947</td>
<td>57.4</td>
<td>0.950</td>
<td>30.2</td>
<td>0.961</td>
<td>13.7</td>
</tr>
<tr>
<td>sup W</td>
<td>0.444</td>
<td>0.944</td>
<td>0.999</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2.8</td>
<td>3.2</td>
<td>3.6</td>
<td>4.0</td>
<td>4.4</td>
<td>4.8</td>
</tr>
<tr>
<td>$U_{T, neq}$</td>
<td>0.938</td>
<td>59.7</td>
<td>0.949</td>
<td>75.7</td>
<td>0.955</td>
<td>86.2</td>
</tr>
<tr>
<td>Bai</td>
<td>0.981</td>
<td>3.6</td>
<td>0.983</td>
<td>2.8</td>
<td>0.988</td>
<td>2.3</td>
</tr>
<tr>
<td>PB</td>
<td>0.973</td>
<td>3.0</td>
<td>0.977</td>
<td>2.1</td>
<td>0.970</td>
<td>1.4</td>
</tr>
<tr>
<td>ILR</td>
<td>0.990</td>
<td>2.4</td>
<td>0.992</td>
<td>2.0</td>
<td>0.989</td>
<td>1.7</td>
</tr>
<tr>
<td>sup W</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: $y_t = \delta(1 - \alpha)1_{t \geq [0.5T]} + \alpha y_{t-1} + \epsilon_t, \epsilon_t \sim i.i.d. N(0, 0.49), \alpha = 0.3, T = 100$. See notes to Table 1.

Table 12: DGP M9-1 (AR coefficient = 0.8)

<table>
<thead>
<tr>
<th>$\delta = dT^{1/2}$</th>
<th>0.4</th>
<th>0.8</th>
<th>1.2</th>
<th>1.6</th>
<th>2.0</th>
<th>2.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U_{T, neq}$</td>
<td>0.946</td>
<td>87.4</td>
<td>0.965</td>
<td>89.0</td>
<td>0.957</td>
<td>92.4</td>
</tr>
<tr>
<td>Bai</td>
<td>0.740</td>
<td>47.7</td>
<td>0.779</td>
<td>20.1</td>
<td>0.819</td>
<td>10.6</td>
</tr>
<tr>
<td>PB</td>
<td>0.955</td>
<td>54.9</td>
<td>0.944</td>
<td>37.5</td>
<td>0.946</td>
<td>23.3</td>
</tr>
<tr>
<td>ILR</td>
<td>0.887</td>
<td>51.3</td>
<td>0.917</td>
<td>30.6</td>
<td>0.943</td>
<td>17.0</td>
</tr>
<tr>
<td>sup W</td>
<td>0.655</td>
<td>0.978</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2.8</td>
<td>3.2</td>
<td>3.6</td>
<td>4.0</td>
<td>4.4</td>
<td>4.8</td>
</tr>
<tr>
<td>$U_{T, neq}$</td>
<td>0.956</td>
<td>93.1</td>
<td>0.963</td>
<td>92.8</td>
<td>0.965</td>
<td>92.2</td>
</tr>
<tr>
<td>Bai</td>
<td>0.835</td>
<td>3.2</td>
<td>0.865</td>
<td>2.6</td>
<td>0.882</td>
<td>2.3</td>
</tr>
<tr>
<td>PB</td>
<td>0.972</td>
<td>3.3</td>
<td>0.977</td>
<td>2.1</td>
<td>0.971</td>
<td>1.3</td>
</tr>
<tr>
<td>ILR</td>
<td>0.979</td>
<td>2.9</td>
<td>0.991</td>
<td>2.3</td>
<td>0.997</td>
<td>1.7</td>
</tr>
<tr>
<td>sup W</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: $y_t = \delta(1 - \alpha)1_{t \geq [0.5T]} + \alpha y_{t-1} + \epsilon_t, \epsilon_t \sim i.i.d. N(0, 0.04), \alpha = 0.8, T = 100$. See notes to Table 1.
### Table 13: DGP D1

<table>
<thead>
<tr>
<th>$\delta = dT^{1/2}$</th>
<th>0.4</th>
<th>0.8</th>
<th>1.2</th>
<th>1.6</th>
<th>2.0</th>
<th>2.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{U}_{T, \neq}$</td>
<td>0.955</td>
<td>88.1</td>
<td>0.959</td>
<td>89.2</td>
<td>0.959</td>
<td>92.0</td>
</tr>
<tr>
<td>Bai</td>
<td>0.739</td>
<td>53.5</td>
<td>0.789</td>
<td>24.3</td>
<td>0.849</td>
<td>13.9</td>
</tr>
<tr>
<td>PB</td>
<td>0.955</td>
<td>54.9</td>
<td>0.944</td>
<td>37.6</td>
<td>0.946</td>
<td>23.3</td>
</tr>
<tr>
<td>ILR</td>
<td>0.891</td>
<td>51.5</td>
<td>0.932</td>
<td>28.1</td>
<td>0.937</td>
<td>17.0</td>
</tr>
<tr>
<td>sup $W$</td>
<td>0.431</td>
<td>0.826</td>
<td>0.942</td>
<td>0.963</td>
<td>0.985</td>
<td>0.991</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{U}_{T, \neq}$</td>
<td>0.964</td>
<td>93.3</td>
<td>0.968</td>
<td>92.8</td>
<td>0.970</td>
<td>92.1</td>
</tr>
<tr>
<td>Bai</td>
<td>0.975</td>
<td>3.8</td>
<td>0.977</td>
<td>2.9</td>
<td>0.994</td>
<td>2.6</td>
</tr>
<tr>
<td>PB</td>
<td>0.972</td>
<td>3.3</td>
<td>0.977</td>
<td>2.2</td>
<td>0.971</td>
<td>1.3</td>
</tr>
<tr>
<td>ILR</td>
<td>0.984</td>
<td>2.7</td>
<td>0.982</td>
<td>2.2</td>
<td>0.995</td>
<td>1.7</td>
</tr>
<tr>
<td>sup $W$</td>
<td>0.995</td>
<td>0.999</td>
<td>0.998</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: $y_t = \alpha y_{t-1} + \delta 1_{t>[0.5T]} + \epsilon_t$, $\epsilon_t \sim i.i.d. N(0, 1)$, $\alpha = 0.8$, $T = 100$. See notes to Table 1.

### Table 14: DGP D2

<table>
<thead>
<tr>
<th>$\delta = dT^{1/2}$</th>
<th>0.4</th>
<th>0.8</th>
<th>1.2</th>
<th>1.6</th>
<th>2.0</th>
<th>2.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{U}_{T, \neq}$</td>
<td>0.957</td>
<td>78.9</td>
<td>0.956</td>
<td>48.8</td>
<td>0.957</td>
<td>28.3</td>
</tr>
<tr>
<td>Bai</td>
<td>0.847</td>
<td>69.7</td>
<td>0.873</td>
<td>34.1</td>
<td>0.922</td>
<td>15.7</td>
</tr>
<tr>
<td>PB</td>
<td>0.987</td>
<td>59.3</td>
<td>0.959</td>
<td>38.4</td>
<td>0.952</td>
<td>19.6</td>
</tr>
<tr>
<td>ILR</td>
<td>0.968</td>
<td>59.1</td>
<td>0.956</td>
<td>31.2</td>
<td>0.957</td>
<td>13.1</td>
</tr>
<tr>
<td>sup $W$</td>
<td>0.346</td>
<td>0.893</td>
<td>0.996</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{U}_{T, \neq}$</td>
<td>0.954</td>
<td>16.3</td>
<td>0.951</td>
<td>17.3</td>
<td>0.949</td>
<td>19.0</td>
</tr>
<tr>
<td>Bai</td>
<td>0.988</td>
<td>3.0</td>
<td>0.996</td>
<td>2.4</td>
<td>0.992</td>
<td>2.1</td>
</tr>
<tr>
<td>PB</td>
<td>0.966</td>
<td>2.8</td>
<td>0.970</td>
<td>2.0</td>
<td>0.975</td>
<td>1.4</td>
</tr>
<tr>
<td>ILR</td>
<td>0.992</td>
<td>2.5</td>
<td>0.996</td>
<td>1.9</td>
<td>0.994</td>
<td>1.6</td>
</tr>
<tr>
<td>sup $W$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: $y_t = \delta 1_{t>[0.5T]} + \epsilon_t$, $\epsilon_t \sim i.i.d. N(0, 1)$, $T = 100$. See notes to Table 1.
Figure 1: Power Functions of $\hat{U}_T([T\lambda])$ in Dynamic Regression Models. $T = 100$, $\lambda_0 = 0.5$

Figure 2: $\hat{U}_T([T\lambda])$ in Dynamic Regression Models: $T = 100$, $\lambda_0 = 0.5$
Figure 3: $\hat{U}_T([T\lambda])$ in Dynamic Regression Models: $\lambda_0 = 0.5, \lambda = 0.3$