Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 141 (2007) 1196–1218
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Confidence sets for the date of a single break in linear time series regressions

Graham Elliott[a,*], Ulrich K. Müller[b]

[a]Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508, USA
[b]Department of Economics, Princeton University, Princeton, NJ 08544-1021, USA

Available online 26 March 2007

Abstract

This paper considers the problem of constructing confidence sets for the date of a single break in a linear time series regression. We establish analytically and by small sample simulation that the current standard method in econometrics for constructing such confidence intervals has a coverage rate far below nominal levels when breaks are of moderate magnitude. Given that breaks of moderate magnitude are a theoretically and empirically relevant phenomenon, we proceed to develop an appropriate alternative. We suggest constructing confidence sets by inverting a sequence of tests. Each of the tests maintains a specific break date under the null hypothesis, and rejects when a break occurs elsewhere. By inverting a certain variant of a locally best invariant test, we ensure that the asymptotic critical value does not depend on the maintained break date. A valid confidence set can hence be obtained by assessing which of the sequence of test statistics exceeds a single number.

JEL classification: C22; C12

Keywords: Test inversion; Coverage control; Locally best test

1. Introduction

It is fairly common to find some form of structural instability in time series models. Tests often reject (Stock and Watson, 1996) the stability of bivariate relationships between macroeconomic series. Similar results have been established for data used in finance and

*Corresponding author. Tel.: +1 858 534 4481; fax: +1 858 534 7040.
E-mail address: gelliott@weber.ucsd.edu (G. Elliott).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2007.02.001

international macroeconomics. Lettau and Ludvigson (2001) and Timmermann and Paye (2004), for example, find instabilities in return forecasting models. The next step after finding such instabilities is to document their form. In general, the answer to this question is going to be the evolution of the unstable parameter over time. With the additional assumption that the parameters change only once, the answer boils down to the time and magnitude of the break. Arguably, the date of the break is typically of greater interest. This paper examines a multiple regression model and considers inference about the date of a single break in a subset of the coefficients.

Determining when parameters change is interesting for a number of reasons. First, this is often an interesting question for economics in its own right. Having observed instability in the mean of growth, we may well be interested in determining when this happened in order to trace the causes of the change. Second, such results can be useful for forecasting. When models are subject to a break, better forecasts will typically emerge from putting more (or all) weight on observations after the break (Pesaran and Timmermann, 2002). Finally, from a model building perspective, it is of obvious interest to determine the stable periods, which are determined by the date of the break.

The literature on estimation and construction of confidence sets for break dates goes back to Hinkley (1970), Hawkins (1977), Worsley (1979, 1986), Bhattacharya (1987) and others—see the reviews by Zacks (1983), Stock (1994) and Bhattacharya (1994) for additional discussion and references. The standard econometric method for constructing confidence intervals for the date of breaks relies on work by Bai (1994), which is further developed in Bai (1997a,b, 1999), Bai et al. (1998) and Bai and Perron (1998). For the problem of a single break in a linear time series regression, the main reference is Bai (1997b).

As is standard in time series econometrics, Bai (1997b) relies on asymptotic arguments to justify his method of constructing confidence intervals for the date of a break. The aim of any asymptotic argument is to provide useful small sample approximations. Specifically, for the problem of dating breaks, one would want the asymptotic approximation to be good for a wide range of plausible break magnitudes, such that confidence sets of the break date have approximately correct coverage irrespective of the magnitude of the break.

The asymptotic analysis that underlies Bai's (1994, 1997b) results involves a break of shrinking magnitude, but at a rate that is slow enough such that for a large enough sample size, reasonable tests for breaks will detect the presence of the break with probability close to one. In other words, this asymptotic analysis focusses on the part of the parameter space where the magnitude of the break is 'large' in the sense that p-values of tests for a break converge to zero. Inference for the presence of a break becomes trivial for such a 'large' break, although the exact date of the break remains uncertain. In contrast, one might speak of a 'small' break when both the presence and the date of the break are uncertain. Analytically, a small break can be represented by an asymptotic analysis where the magnitude of the break shrinks at a rate such that tests for a break have nontrivial power that is strictly below one.

In many practical applications, breaks that are of interest are arguably not large in this sense. After all, formal econometric tests for the presence of breaks are employed precisely because there is uncertainty about the presence of a break. From an empirical point of view, the observed p-values are often borderline significant; in the Stock and Watson (1996) study, for instance, the QLR statistic investigated by Andrews (1993) rejects stability of 76 US postwar macroeconomic series for 23 series at the 1% level, for an additional 11 series at the 5% level and for an additional 6 series at the 10% level. In a similar vein, variations in the conduct of monetary policy that some argue are crucial to understand the US postwar period are small enough that a debate has arisen as to both the size and nature of the breaks and whether they are there at all. For example, Orphanides (2004) argues that the relationships are quite stable. Clarida et al. (2000) argue that the economic differences pre and post the Volcker chairmanship of the US Federal Reserve Board are economically important although they did not test for the break. Boivin (2003) finds based on tests and a robustness analysis that a fixed 'Volcker' break does not capture well changes in the Taylor rule relationships. In all, any changes to the relationship are small compared to the variation of the data even though their existence is important for assessing the conduct of monetary policy.

Breaks that are small in this statistical sense are, of course, not necessarily small in an economic sense. As usual, economic and statistical significance are two very distinct concepts. As an example, consider the possibility of a break in the growth of income. Postwar quarterly US real gross domestic product growth measured in percentage points has a standard deviation of about unity. Even if growth is independent and identically distributed (i.i.d.) Gaussian, this variation will make it very difficult to detect, let alone date, a break of mean growth that is smaller than 0.25 percentage points. But, of course, a break that leads to yearly growth being one percentage point higher is a very important event for an economy.

Given the importance of 'small' breaks, one might wonder about the accuracy of the asymptotic approximation that validates the confidence intervals developed in Bai (1997b). As we show below, the coverage rates of these confidence intervals are far below nominal levels for small breaks. This is true even for breaks whose magnitude is such that their presence is picked up with standard tests with very high probability. These findings are consistent with the recent simulation study of Bai and Perron (2006).

The question hence arises of how to construct valid confidence sets for the date of a break when the break is, at least potentially, small. We follow the standard approach to constructing confidence sets by inverting a sequence of tests; see, for instance, Lehmann (1986, p. 90). This approach to confidence set construction has been used before in nonstandard time series problems by Stock (1991) and Hansen (2000).

The idea is to test the sequence of null hypotheses that maintain the break to be at a certain date. The hypotheses are judged by tests that allow for a break under the null hypothesis at the maintained date, but that reject for breaks at other dates. If the maintained break date is wrong, then there is a break at one of these other dates, and the test rejects. The confidence set is the collection of all maintained dates for which the test does not reject. By imposing invariance of the tests to the magnitude of the break at the maintained date, we ensure that coverage of this confidence set remains correct for any magnitude of the break, at least asymptotically. By a judicious choice of the efficient tests we suggest inverting the critical values of the sequence of test statistics does not depend on the maintained break date in the limit. The construction of a valid confidence set for the break date of arbitrary magnitude can hence be generated by comparing a sequence of test statistics with a single critical value.

In the next section we analytically investigate the coverage properties of the popular method of obtaining confidence intervals when the magnitude of the break is small. This motivates the need for a new method. The third section derives the test statistics to be inverted. Section 4 evaluates the methods numerically for some standard small sample data generating processes. Proofs are collected in an Appendix.

2. Properties of standard confidence intervals when breaks are small

This paper considers the linear time series regression model

$$y_t = X_t'\beta + \mathbf{1}[t>\tau_0]X_t'\delta + Z_t'\gamma + u_t, \quad t = 1, \ldots, T, \tag{1}$$

where $\mathbf{1}[\cdot]$ is the indicator function, y_t is a scalar, X_t, β and δ are $k \times 1$ vectors, Z_t and γ are $p \times 1$, $\{y_t, X_t, Z_t\}$ are observed, τ_0, β, δ and γ are unknown and $\{u_t\}$ is a mean zero disturbance. Define $Q_t = (X_t', Z_t')'$. Let '$\xrightarrow{p}$' denote convergence in probability and '$\Rightarrow$' convergence of the underlying probability measures as $T \to \infty$ and let $[\cdot]$ be the greatest smaller integer function. For the asymptotic results, we impose the following regularity condition on model (1):

Condition 1. (i) $\tau_0 = [r_0 T]$ for some $0<r_0<1$.

(ii) $T^{-1/2}\sum_{t=1}^{[sT]} X_t u_t \Rightarrow \Omega_1^{1/2} W(s)$ for $0 \leqslant s \leqslant r_0$ and $T^{-1/2}\sum_{t=\tau_0+1}^{[sT]} X_t u_t \Rightarrow \Omega_2^{1/2}(W(s) - W(r_0))$ for $r_0 \leqslant s \leqslant 1$ with Ω_1 and Ω_2 some symmetric and positive definite $k \times k$ matrices and $W(\cdot)$ a $k \times 1$ standard Wiener process.

(iii) $\sup_{0\leqslant s \leqslant 1} \| T^{-1/2}\sum_{t=1}^{[sT]} Z_t u_t \| = O_p(1)$.

(iv)

$$T^{-1}\sum_{t=1}^{[sT]} Q_t Q_t' \xrightarrow{p} s\Sigma_{Q1} = s\begin{pmatrix} \Sigma_{X1} & \Sigma_{XZ1} \\ \Sigma_{ZX1} & \Sigma_{Z1} \end{pmatrix}$$

uniformly in $0 \leqslant s \leqslant r_0$ and

$$T^{-1}\sum_{t=\tau_0+1}^{[sT]} Q_t Q_t' \xrightarrow{p} (s - r_0)\Sigma_{Q2} = (s - r_0)\begin{pmatrix} \Sigma_{X2} & \Sigma_{XZ2} \\ \Sigma_{ZX2} & \Sigma_{Z2} \end{pmatrix}$$

uniformly in $r_0 \leqslant s \leqslant 1$, where Σ_{Q1} and Σ_{Q2} are full rank.

In the asymptotic analysis considered in this paper, the number of observations that precede and follow the break are in the fixed proportion $r_0/(1 - r_0)$. This is standard in the structural break literature, although recently alternative asymptotics have been considered by Andrews (2003). With $\tau_0 = [r_0 T]$, the data generated by this model necessarily becomes a double array, as τ_0 depends on T, although we do not indicate this dependence on T to enhance readability. Conditions (ii)–(iv) are standard high-level time series conditions, that allow for heterogeneous and autocorrelated $\{u_t\}$ and regressors $\{Q_t\}$. Condition 1 also accommodates regressions with only weakly exogenous regressors. As in Bai (1997b), both the second moment of $\{Q_t\}$ and the long-run variance of $\{Q_t u_t\}$ are allowed to change at the break date τ_0.

The state-of-the-art econometric method to obtain confidence intervals for τ_0 developed by Bai (1997b) proceeds as follows: minimize the sum of squared residuals of the linear regression (1) over all coefficient vectors and break dates. Denote the minimizing choice for the break magnitude and break date by $\hat{\delta}$ and $\hat{\tau}$, respectively. A level C confidence interval for τ_0 is then constructed as

$$[\hat{\tau} - [\lambda_{(1+C)/2} m] - 1, \hat{\tau} - [\lambda_{(1-C)/2} m] + 1], \tag{2}$$

where $m = \hat{\delta}'\Omega_1\hat{\delta}/(\hat{\delta}'\Sigma_{X1}\hat{\delta})^2$ and λ_c is the 100c percentile of the distribution of an absolutely continuous random variable whose distribution depends on two parameters that can be consistently estimated by $\hat{\delta}'\Omega_2\hat{\delta}/(\hat{\delta}'\Omega_1\hat{\delta})$ and $\hat{\delta}'\Sigma_{X2}\hat{\delta}/(\hat{\delta}'\Sigma_{X1}\hat{\delta})$—see Bai (1997b) for details. In the special case where $\Omega_1 = \Omega_2$ and $\Sigma_{X1} = \Sigma_{X2}$, λ_c is the 100c percentile of the distribution of $\arg\min_s W(s) - |s|/2$. This distribution is symmetric, so that the level C confidence interval becomes $[\hat{\tau} - [\lambda_{(1+C)/2}m] - 1, \hat{\tau} + [\lambda_{(1+C)/2}m] + 1]$ with $m = \hat{\delta}'\Omega\hat{\delta}/(\hat{\delta}'\Sigma_X\hat{\delta})^2$. Typically, Ω_i and Σ_{Xi} for $i = 1, 2$ are unknown, but can be consistently estimated. For expositional ease, we abstract from this additional estimation problem and assume Ω_i and Σ_{Xi} known in the following discussion of the properties of the confidence intervals (2).

As shown by Bai (1997b), intervals (2) are asymptotically valid when $\delta = T^{-1/2+\varepsilon}d$ for some $0<\varepsilon<\frac{1}{2}$ and $d\neq 0$. Although the magnitude of the break δ shrinks under these asymptotics, the generated breaks are still large in the sense that they will be detected with probability one with any reasonable test for breaks: the neighborhood in which the tests of Nyblom (1989), Andrews (1993), Andrews and Ploberger (1994) and Elliott and Müller (2006) have nontrivial local asymptotic power is where $\varepsilon = 0$. In other words, under asymptotics that justify the confidence intervals (2) the p-values of any standard test for breaks converge to zero. With $0<\varepsilon<\frac{1}{2}$, there is ample information about the break in the sense that it is obvious that there is a break, the only question concerns its exact date.

In fact, when $0<\varepsilon<\frac{1}{2}$, $\hat{\tau}/T$ is a consistent estimator of r_0—see Bai (1997b). The break is large enough to pinpoint down exactly its date in terms of the fraction of the sample. The uncertainty that is described by the confidence interval (2) arises only because the break date τ_0 is an order of magnitude larger than r_0, since $\tau_0 = [Tr_0]$.

As argued above, it is unclear whether breaks typically encountered in practice are necessarily large enough for this asymptotic analysis to yield satisfactory approximations. The p-values of tests for breaks are never zero, and quite often indicate only borderline significance. Also from an economic theory standpoint there is typically nothing to suggest that breaks are necessarily large in the sense that their statistical detection is guaranteed. This raises the question as to the accuracy of the approximation that underlies (2) when in fact the break is smaller.

In order to answer this question, we consider the properties of the confidence interval (2) when $\delta = T^{-1/2}d$, i.e. where $\varepsilon = 0$. These asymptotics provide more accurate representations of small samples in which the break size is moderate in the sense that p-values of tests for breaks are typically significant, but not zero. This applies to large breaks in a relatively small sample, or smaller breaks in a large sample. When $\|d\|$ is very large, then the probability of detecting the break is very close to one. One might hence think of asymptotics with $\delta = T^{-1/2}d$ as providing the continuous bridge between a stable linear regression (when $d = 0$) and one with a large break ($\|d\|$ large).

In contrast to the setup with $0<\varepsilon<\frac{1}{2}$, r_0 is not consistently estimable when $\delta = T^{-1/2}d$ for any finite value of $\|d\|$. The reason is simply that if even efficient tests cannot consistently determine whether there is a break (although for $\|d\|$ large enough their power will become arbitrarily close to one), there cannot exist a statistic that consistently estimates a property of that break. In other words, the uncertainty about the break date in asymptotics with $\delta = T^{-1/2}d$ extends to the fraction r_0. It is interesting to note that running regression (1) with τ_0 replaced by $\hat{\tau}$ and ignoring the fact that $\hat{\tau}$ is estimated will therefore not yield asymptotically correct inference about δ and β, in contrast to asymptotics where $\delta = T^{-1/2+\varepsilon}d$ for some $0<\varepsilon<\frac{1}{2}$.

For expositional ease and to reduce the notational burden, the following proposition establishes the asymptotic properties of the confidence interval (2) when $\delta = T^{-1/2}d$ in the special case where $\Omega_1 = \Omega_2 = \Omega$ and $\Sigma_{Q1} = \Sigma_{Q2} = \Sigma_Q = \begin{pmatrix} \Sigma_X & \Sigma_{XZ} \\ \Sigma_{ZX} & \Sigma_Z \end{pmatrix}$.

Proposition 1. *For any* $\frac{1}{2} > \bar{\lambda} > 0$, *define for a standard* $k \times 1$ *Wiener process* $W(\cdot)$,

$$M(s) = \Omega^{1/2}W(s) + \mathbf{1}[s \geqslant r_0](s - r_0)\Sigma_X d,$$

$$G(s) = \frac{M(s)'\Sigma_X^{-1}M(s)}{s} + \frac{(M(1) - M(s))'\Sigma_X^{-1}(M(1) - M(s))}{1 - s}.$$

Then under Condition 1, *when* $G(s)$ *has a unique maximum with probability one on* $[\bar{\lambda}, 1 - \bar{\lambda}]$, $\Omega_1 = \Omega_2 = \Omega$, $\Sigma_{Q1} = \Sigma_{Q2} = \Sigma_Q$ *and* $\delta = T^{-1/2}d$,

$$T^{-1}(\hat{\tau}, m) \Rightarrow \left(\hat{r}_a, \frac{\hat{\delta}_a'\Omega\hat{\delta}_a}{(\hat{\delta}_a'\Sigma_X\hat{\delta}_a)^2}\right),$$

where $\hat{\tau}$ *minimizes the sum of squared residuals in the linear regression* (1) *with* τ_0 *replaced by* τ *over all* $\tau \in (\bar{\lambda}T, (1 - \bar{\lambda})T)$ *and*

$$\hat{r}_a = \arg \max_{\bar{\lambda} \leqslant s \leqslant 1 - \bar{\lambda}} G(s),$$

$$\hat{\delta}_a = \Sigma_X^{-1} \frac{\hat{r}_a M(1) - M(\hat{r}_a)}{\hat{r}_a(1 - \hat{r}_a)}.$$

Several comments can be made regarding Proposition 1. First, in the statement of the proposition, the potential choices of the break date are trimmed away from the endpoints. Such trimming is standard in the literature on tests for breaks (Andrews, 1993; Andrews and Ploberger, 1994).

Second, the margin of error of the confidence intervals (2) is $m \sim T$ (i.e. $m = O_p(T)$ and m is not $o_p(T)$). As discussed, the uncertainty about the break date under these local asymptotics extends to uncertainty about r_0. Although the confidence intervals (2) have not been constructed for this case, they automatically adapt and cover (with probability one) a positive fraction of all possible break dates asymptotically.

Third, note that the asymptotic distribution of $T^{-1}(\hat{\tau}, m)$ is the same for $d = d_0$ and $d = -d_0$ for any d_0, so that asymptotic coverage properties are symmetric in the sign of the break.

Finally, the asymptotic distribution of $(\hat{\tau} - \tau_0)/m$ is no longer given by $\arg\min_s W(s) - |s|/2$, but it depends on r_0, Ω and Σ_X in a complicated way. It is hence not possible to construct asymptotically justified confidence intervals for local asymptotics by adding and subtracting the margin of error m from $\hat{\tau}$. The precise magnitude of the effects and whether the confidence interval (2) undercovers or overcovers are unclear and require a numerical evaluation.

Table 1 depicts the asymptotic coverage rates of nominal 95% confidence intervals (2) for $k = 1$, $\Omega = \Sigma_X = 1$ and various values of d and r_0, along with the asymptotic local power of a 5%-level Nyblom (1989) test for a break in β. The trimming parameter $\bar{\lambda}$ is set to $\bar{\lambda} = 0.05$; smaller values of $\bar{\lambda}$ lead to worse coverage for breaks with $d \leqslant 8$, while leaving results for larger breaks largely unaffected. For $d = 8$, coverage rates are below 88%, and much smaller still for $d = 4$. This is despite the fact that breaks with $d = 8$ have a high probability of being detected with Nyblom's tests for breaks, at least as long as they do not occur close to the beginning or end of the sample. The asymptotic distribution of p-values

Table 1
Local asymptotic properties of Bai's (1997b) CIs

d	$r_0 = 0.5$		$r_0 = 0.35$		$r_0 = 0.2$	
	Cov.	Nybl.	Cov.	Nybl.	Cov.	Nybl.
4	0.711	0.438	0.707	0.375	0.700	0.204
8	0.877	0.953	0.870	0.915	0.840	0.651
12	0.923	1.000	0.918	0.999	0.907	0.956
16	0.936	1.000	0.939	1.000	0.930	0.999

For each r_0, the first column is asymptotic coverage of the confidence intervals (2), and the second column is local asymptotic power of the 5%-level Nyblom (1989) test for the presence of a break. Based on 10,000 replications with 1000 step approximations to continuous time processes.

of the Nyblom test for $d = 4$ and $r_0 = 0.35$ is such that 17% are below 1%, 20% are between 1% and 5% and 13% are between 5% and 10%. This corresponds at least roughly to the distribution of p-values found by Stock and Watson (1996) for the stability of 76 macroseries, although this comparison obviously suffers from the lack of independence of the macroseries. When $d = 16$, which corresponds to a break that is big enough to be almost always detected, the asymptotic approximation that justifies (2) seems to become more accurate, as effective coverage rates become closer to the nominal level.

Returning to the example of US GDP growth introduced in the Introduction, suppose one wanted to date a break in mean growth with a sample of $T = 180$ quarterly observations. When quarterly growth is i.i.d. with unit variance (which roughly corresponds to the sample variance), then $d = 12$ corresponds to a break in the quarterly growth rate of $\frac{12}{\sqrt{180}} = 0.89$ percentage points. For the asymptotic approximation underlying (2) to be somewhat accurate, the break in mean growth has hence to be larger than 3.5% on a yearly basis!

This asymptotic analysis suggests that the standard way of constructing confidence intervals based on (2) leads to substantial undercoverage when the magnitude of the break is not very large, but large enough to be detected with high probability by a test for structural stability. A small sample Monte Carlo study in Section 4 below confirms this to be an accurate prediction for some standard data generating processes.

3. Valid confidence sets for small breaks

As shown in the preceding analysis, the standard method for constructing a confidence interval for the date of a break in the coefficient of a linear regression does not control coverage when the break is small. At the same time, small breaks are often plausible from a theoretical point of view and are found to be relevant empirically. This raises the question of how to construct confidence sets that maintain nominal coverage rates when breaks are small or large.

A level C confidence set can be thought of as a collection of parameter values that cannot be rejected with a level $1 - C$ hypothesis test; see, for instance, Lehmann (1986, p. 90). In standard setups, estimators are asymptotically unbiased and Gaussian with a variance that, at least locally, does not depend on the parameter value. If one bases the sequence of tests on this estimator, the set of parameter values for which the test does not reject becomes a symmetric interval around the parameter estimator.

The problem at hand is not standard in this sense, as the asymptotic distribution of the estimator $\hat{r}$ is not Gaussian centered around r_0—see Proposition 1 above. What is more, the asymptotic distribution of $\hat{r}$ depends on r_0 in a highly complicated fashion. Basing valid tests for specific values of r_0 (or equivalently τ_0) on $\hat{r}$ therefore becomes a difficult endeavor. But this does not alter the fact that a valid level C confidence set for τ_0 can be constructed by inverting a sequence of level $(1 - C)$ tests, each maintaining that under the null hypothesis, the true break date τ_0 coincides with the maintained break date τ_{m}, i.e. H_0: $\tau_0 = \tau_{\mathrm{m}}$ for $\tau_{\mathrm{m}} = 1, \ldots, T$. As long as the test with the true null hypothesis has correct level, the resulting confidence set has correct coverage, as the probability of excluding the true value τ_0 is identical to the type I error of the employed significance test. For tests with $\tau_{\mathrm{m}} \neq \tau_0$, the break occurs at a date different from the maintained break. Tests that reject with high probability when faced with a break that occurs at a date other than the maintained break date τ_{m} will result in short confidence sets. The more powerful the tests are against this alternative, the shorter the confidence set becomes on average (cf. Pratt, 1961).

Confidence sets for the break date of the coefficient in a linear regression model hence can be obtained by inverting a sequence of hypothesis tests of the null hypothesis of a maintained break at date τ_{m} against the alternative that the break occurs at some other date

$$\mathrm{H}_0\colon \tau_0 = \tau_{\mathrm{m}} \quad \text{against} \quad \mathrm{H}_1\colon \tau_0 \neq \tau_{\mathrm{m}}. \tag{3}$$

The construction of these tests faces three challenges: (I) Their rejection probability under the null hypothesis must not exceed the level for any value of the break size δ. (II) It is powerful against alternatives where $\tau_0 \neq \tau_{\mathrm{m}}$. (III) A practical (but not conceptual) complication is that the critical value of test statistics of (3) will typically depend on the maintained break date τ_{m}. For the construction of a confidence set, one would hence need to compute T test statistics, and compare them to T different critical values, which is highly cumbersome.

Consider these complications in turn. First, concerning (I), in order to control the rejection probability under the null hypothesis for any value of δ, we impose invariance of the test to transformations of y_t that correspond to varying δ. Specifically, we consider tests that are invariant to transformations of the data

$$\{y_t, X_t, Z_t\} \to \{y_t + X_t' b_0 + \mathbf{1}[t > \tau_{\mathrm{m}}] X_t' d_0 + Z_t' g_0, X_t, Z_t\} \quad \text{for all } b_0, d_0, g_0. \tag{4}$$

When $\{X_t, Z_t\}$ is strictly exogenous, this invariance requirement will make the distribution of the test statistic independent of the values of β, γ and δ under the null hypothesis. But even if $\{X_t, Z_t\}$ is not strictly exogenous, the asymptotic null distribution of the invariant test statistics will still be independent of β, γ and δ under Condition 1, as shown in Proposition 3 below. For a scalar AR(1) process with no Z_t and $X_t = y_{t-1}$, for instance, the requirement of invariance to the transformations $\{y_t, y_{t-1}\} \to \{y_t - b_0 y_{t-1}, y_{t-1}\}$ for all b_0 amounts to the sensible restriction that the stability of the regression of $\{y_t\}$ on $\{y_{t-1}\}$ should not be decided differently than the stability of the regression of $\{\Delta y_t\}$ on $\{y_{t-1}\}$. In practice, the invariance will be achieved by basing tests on OLS residuals of regression (1) with τ_0 replaced by τ_{m}.

Second, in order to ensure that the tests to be inverted are powerful (II), one would like to choose the most powerful test of (3). For the construction of efficient tests based on the Neyman–Pearson lemma, one needs an assumption concerning the distribution of the disturbance $\{u_t\}$ and other properties of model (1).

Condition 2. (i) u_t is i.i.d. $\mathcal{N}(0,\sigma^2)$.
(ii) $\{u_t\}_{t=1}^T$ and $\{Q_t\}_{t=1}^T$ are independent.

Part (i) of the condition specifies the distribution of $\{u_t\}$ to be Gaussian. Only the efficiency of the following test depends on this (often unrealistic) assumption, but not the validity of the resulting test. In fact, the asymptotic local power of the efficient test tailor-made for Gaussian disturbances turns out to be the same for all models with i.i.d. innovations of variance σ^2. The assumption of Gaussianity of $\{u_t\}$ for the construction of efficient tests is least favorable in this sense. If $\{u_t\}$ were serially correlated with known correlation structure, then efficient tests would be constructed from the GLS transformation of the model. This will result in a different small sample optimal test, which will in general have higher power even asymptotically.

Part (ii) of Condition 2 requires Q_t to be strictly exogenous. To the best of our knowledge, all small sample optimality results for invariant tests, such as those derived in Andrews et al. (1996) and Forchini (2002), make this assumption. Again, Condition 2 (ii) is only required for the small sample efficiency of the test derived in Proposition 2 below; the test remains asymptotically valid under much weaker assumptions, which include models with weakly exogenous $\{X_t, Z_t\}$.

Unfortunately, even under Condition 2, a uniformly most powerful test does not exist, as efficient test statistics depend on both the true break date τ_0 and δ, both of which are unknown. In fact, under the invariance requirement (4), the parameter δ that describes the magnitude of the break under the alternative is not identified under the null hypothesis, as the distribution of any maximal invariant to (4) does not depend on δ (at least in the case of strictly exogenous $\{X_t, Z_t\}$). As in Andrews and Ploberger (1994), we therefore consider tests that maximize weighted average power: a test φ is an efficient level α test φ^* of $\tau_0 = \tau_{\mathrm{m}}$ against $\tau_0 \neq \tau_{\mathrm{m}}$ when it maximizes the weighted average power criterion

$$\sum_{t\neq\tau_{\mathrm{m}}} w_t \int \mathrm{P}(\varphi \text{ rejects}|\tau_0 = t, \delta = d)\,\mathrm{d}v_t(d) \tag{5}$$

over all tests which satisfy $\mathrm{P}(\varphi \text{ rejects } |\tau_0 = \tau_{\mathrm{m}}) = \alpha$, where $\{w_t\}_{t=1}^T$ is a sequence of nonnegative real numbers, and $\{v_t\}_{t=1}^T$ is a sequence of nonnegative measures on $\mathbb{R}^k$. The prespecified sequences $\{w_t\}_{t=1}^T$ and $\{v_t\}_{t=1}^T$ direct the power towards alternatives of certain dates τ_0 and break magnitudes, respectively. From a Bayesian perspective, the weights $\{w_t\}$ and $\{v_t\}$, suitably normalized to ensure their integration to one, can be interpreted as probability measures: if τ_0 and δ were random and followed these distributions under the alternative, then φ^* is the most powerful test against this (single) alternative.

The efficient tests depend on the weighting functions $\{w_t\}$ and $\{v_t\}$, so the question is how to make a suitable choice. As demonstrated in Elliott and Müller (2006), however, the power of tests for structural stability does not greatly depend on the specific choice of weights, at least as long as they do not concentrate too heavily on specific values for τ_0 and δ. With power roughly comparable for alternative weighting schemes, ease of computation becomes arguably a relevant guide.

A solution to the final complication (III), the dependence of the critical value of the sequence of tests on the maintained break date, can hence be generated by a judicious choice of the weighting functions with little cost in terms of inadequate power properties. Specifically, consider measures of the break size v_t that are probability measures of mean zero $k \times 1$ Gaussian vector with covariance matrix $b^2 H_t$, where

$$H_t = \begin{cases} \tau_{\mathrm{m}}^{-2}\Omega_1^{-1} & \text{for } t<\tau_{\mathrm{m}} \\ (T-\tau_{\mathrm{m}})^{-2}\Omega_2^{-1} & \text{for } t>\tau_{\mathrm{m}} \end{cases} \quad \text{and} \quad w_t = 1 \quad \forall t \neq \tau_{\mathrm{m}}. \tag{6}$$

This choice of weighting functions puts equal weight on alternative break dates. Furthermore, the direction of the break as measured by the covariance matrix of the measures v_t is proportional to the long-run covariance matrix of $\{X_t u_t\}$ (which depends on whether $t<\tau_{\mathrm{m}}$ or $t>\tau_{\mathrm{m}}$). The magnitude of the potential break is piecewise constant before and after the maintained break date τ_{m}. Even if $\Omega_1 = \Omega_2$, the break size will not be identical, though, but depends on τ_{m}: when τ_{m} is close to T, for instance, then this choice of v_t puts less weight on large breaks that occur prior to τ_{m} compared to those that occur after.

While not altogether indefensible, this choice of weighting scheme is mostly motivated by the fact that the resulting efficient test statistic has an asymptotic distribution that does not depend on τ_{m}. This makes the construction of an (asymptotically) valid confidence set especially simple, as the sequence of test statistics can be compared to a single critical value, as in Hansen (2000).

Proposition 2. *Under Condition* 2, *the locally best test with respect to* b^2 *of* (3) *that is invariant to* (4) *and that maximizes the weighted average power* (5) *with weighting functions* (6) *rejects for large values of the statistic*

$$\mathrm{U}_T(\tau_{\mathrm{m}}) = \tau_{\mathrm{m}}^{-2}\sum_{t=1}^{\tau_{\mathrm{m}}}\left(\sum_{s=1}^{t} v_s\right)'\Omega_1^{-1}\left(\sum_{s=1}^{t} v_s\right) + (T-\tau_{\mathrm{m}})^{-2}\sum_{t=\tau_{\mathrm{m}}+1}^{T}\left(\sum_{s=\tau_{\mathrm{m}}+1}^{t} v_s\right)'\Omega_2^{-1}\left(\sum_{s=\tau_{\mathrm{m}}+1}^{t} v_s\right), \tag{7}$$

where $v_t = X_t e_t$ *and* e_t *are the residuals of the ordinary least-squares regression* (1) *with* τ_0 *replaced by* τ_{m}.

Busetti and Harvey (2001) and Kurozumi (2002) suggest a specialized version of $\mathrm{U}_T(\tau_{\mathrm{m}})$ for constant and trending $\{X_t\}$ as a test statistic for the null of stationarity under a maintained break at date τ_{m}, although they do not derive optimality properties. The locally best test against martingale variation in the coefficients of a linear regression model has been derived by Nyblom (1989). Specialized to the test of a single break of random magnitude and occurring at a random time (which results in a martingale for the now random coefficient), the usual Nyblom statistic applied to a stable linear regression model puts equal probability on the break occurring at all dates, and the covariance matrix of the break size is constant. It is possible to apply the Nyblom statistic to the breaking regression model (1) with τ_0 replaced by the maintained break date τ_{m}, although one would not recover the asymptotic distribution derived by Nyblom (1989), as the regressor $\{\mathbf{1}[t>\tau_{\mathrm{m}}]X_t\}$ does not satisfy the necessary regularity conditions.

From this perspective, the weighting scheme (6) can be understood as yielding the sum of two Nyblom statistics, at least when there is no Z_t: one for the regression for $t = 1, \ldots, \tau_{\mathrm{m}}$ and one for the regression $t = \tau_{\mathrm{m}} + 1, \ldots, T$. This makes perfect intuitive sense: when the maintained break τ_{m} is not equal to the true break date τ_0, there is one break either prior or after τ_{m}. One way to test this is to use a Nyblom statistic for the (under the null hypothesis stable) standard regression model for $t = 1, \ldots, \tau_{\mathrm{m}}$, and another Nyblom statistic for the (under the null hypothesis also stable) standard regression model for $t = \tau_{\mathrm{m}} + 1, \ldots, T$. Proposition 2 shows that this procedure does not only make intuitive sense, but is also optimal for the weighting scheme (6).

As described in Proposition 2, the test statistic $\mathrm{U}_T(\tau_{\mathrm{m}})$ is not a feasible statistic, as Ω_1 and Ω_2 are typically unknown. But under the null hypothesis of $\tau_0 = \tau_{\mathrm{m}}$, under weak regularity conditions on X_t and u_t, Ω_1 and Ω_2 can typically be consistently estimated by any standard long-run variance estimator applied to $\{v_t\}_{t=1}^{\tau_{\mathrm{m}}}$ and $\{v_t\}_{t=\tau_{\mathrm{m}}+1}^{T}$—for primitive conditions see, for instance, Newey and West (1987) or Andrews (1991). Denote by $\hat{\mathrm{U}}_T(\tau_{\mathrm{m}})$ the statistic $\mathrm{U}_T(\tau_{\mathrm{m}})$ with Ω_1 and Ω_2 replaced by such estimators $\hat{\Omega}_1$ and $\hat{\Omega}_2$.

Proposition 3. *If $\hat{\Omega}_1 \xrightarrow{p} \Omega_1$ and $\hat{\Omega}_2 \xrightarrow{p} \Omega_2$, then under Condition* 1

$$\hat{\mathrm{U}}_T(\tau_0) \Rightarrow \int_0^1 B(s)'B(s)\,\mathrm{d}s,$$

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

The distribution of the integral of a squared Brownian bridge has been studied by MacNeill (1978) and Nabeya and Tanaka (1988). For convenience, critical values of $\hat{\mathrm{U}}_T(\tau_{\mathrm{m}})$ for $k = 1, \ldots, 6$ are reproduced in Table 2.

As required, the asymptotic null distribution of $\hat{\mathrm{U}}_T(\tau_{\mathrm{m}})$ does not depend on δ. For any size of break δ, the collection of values of $\tau_{\mathrm{m}} = 1, \ldots, T$ for which the test $\hat{\mathrm{U}}_T(\tau_{\mathrm{m}})$ does not exceed its asymptotic critical value of significance level $(1 - C)$ hence has asymptotic coverage C, i.e. is a valid confidence set: the only way the true value is excluded from this confidence set is when $\hat{\mathrm{U}}_T(\tau_{\mathrm{m}}) = \hat{\mathrm{U}}_T(\tau_0)$ exceeds the critical value. Note that this in particular implies that the confidence set is valid under asymptotics with $\delta = T^{-1/2}d$ for some fixed d, in contrast to the confidence interval (2).

In detail, one proceeds as follows:

- For any $\tau_{\mathrm{m}} = p + 2k + 1, \ldots, T - p - 2k - 1$, compute the least-squares regression of $\{y_t\}_{t=1}^{T}$ on $\{X_t, \mathbf{1}[t > \tau_{\mathrm{m}}]X_t, Z_t\}_{t=1}^{T}$.
- Construct $\{v_t\}_{t=1}^{T} = \{X_t e_t\}_{t=1}^{T}$, where e_t are the residuals from this regression.

Table 2
Critical values of $\hat{\mathrm{U}}_T(\tau_{\mathrm{m}})$

k (%)	1	2	3	4	5	6
10	0.600	1.063	1.482	1.895	2.293	2.692
5	0.745	1.238	1.674	2.117	2.537	2.951
1	1.067	1.633	2.118	2.570	3.036	3.510

Based on 50,000 replications and 1000 step approximations to continuous time processes.

- Compute the long-run variance estimators $\hat{\Omega}_1$ and $\hat{\Omega}_2$ of $\{v_t\}_{t=1}^{\tau_m}$ and $\{v_t\}_{t=\tau_m+1}^{T}$, respectively. An attractive choice is to use the automatic bandwidth estimators of Andrews (1991) or Andrews and Monahan (1992). If it is known that $\Omega_1 = \Omega_2$, then it is advisable to rely instead on a single long-run variance estimator $\hat{\Omega}$ based on $\{v_t\}_{t=1}^{T}$.
- Compute $\hat{\mathrm{U}}_T(\tau_m)$ as in (7) with Ω_1 and Ω_2 replaced by $\hat{\Omega}_1$ and $\hat{\Omega}_2$, respectively.
- Include τ_m in the level C confidence set when $\hat{\mathrm{U}}_T(\tau_m) < \mathrm{cv}_{1-C}$ and exclude it otherwise, where cv_{1-C} is the level $(1-C)$ critical value of the statistic $\hat{\mathrm{U}}_T(\tau_m)$ from Table 2.

There is no guarantee that this method yields contiguous confidence sets. The reason for this is straightforward. The confidence set construction procedure looks for dates that are compatible with no breaks elsewhere. When the break is small, there may be a number of possible regions for dates that appear plausible candidates for the break. The confidence set includes all these regions. Note that this is not a sign that there are multiple breaks, but rather that the exact date of one break is difficult to determine. A confidence set with good coverage properties will reflect this uncertainty.

It is also possible that the confidence set is empty—this will happen when the test rejects for each possible break date. When the model contains multiple large breaks, this will

Table 3
Empirical small sample coverage and length of confidence sets: model (M1): constant regressor, i.i.d. disturbances

	$d=4$		$d=8$		$d=12$		$d=16$	
	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.
$r_0 = 0.5$								
$\hat{\mathrm{U}}_T(\tau_m)$.eq	0.949	77.7	0.949	42.4	0.949	22.1	0.949	15.1
$\hat{\mathrm{U}}_T(\tau_m)$.neq	0.950	77.2	0.950	42.3	0.950	22.7	0.950	15.8
Bai.eq	0.698	54.5	0.890	33.1	0.940	17.0	0.959	10.5
Bai.het	0.698	54.5	0.890	33.1	0.940	17.0	0.959	10.5
Bai.hneq	0.686	53.0	0.882	33.1	0.938	17.0	0.956	10.5
Nyblom	0.428		0.948		1.000		1.000	
$r_0 = 0.35$								
$\hat{\mathrm{U}}_T(\tau_m)$.eq	0.952	79.0	0.952	44.3	0.952	22.5	0.952	15.0
$\hat{\mathrm{U}}_T(\tau_m)$.neq	0.954	78.7	0.954	44.1	0.954	23.1	0.954	15.7
Bai.eq	0.692	52.5	0.878	32.8	0.937	17.1	0.962	10.5
Bai.het	0.692	52.5	0.878	32.8	0.937	17.1	0.962	10.5
Bai.hneq	0.676	50.8	0.873	32.6	0.932	17.1	0.959	10.5
Nyblom	0.366		0.902		0.999		1.000	
$r_0 = 0.2$								
$\hat{\mathrm{U}}_T(\tau_m)$.eq	0.949	83.2	0.949	55.7	0.949	27.1	0.949	15.3
$\hat{\mathrm{U}}_T(\tau_m)$.neq	0.951	83.3	0.951	56.1	0.951	27.9	0.951	16.2
Bai.eq	0.660	46.8	0.851	31.3	0.926	17.4	0.955	10.7
Bai.het	0.660	46.8	0.851	31.3	0.926	17.4	0.955	10.7
Bai.hneq	0.631	44.4	0.832	30.2	0.914	17.0	0.947	10.5
Nyblom	0.189		0.617		0.939		0.997	

The model is $y_t = \beta + dT^{-1/2}\mathbf{1}[t > [r_0 T]] + u_t$, $u_t \sim iid\, \mathcal{N}(0,1)$, $T = 100$. Cov. and Lgth. refer to the coverage probability and average number of dates in the confidence sets of the various methods described in the text. Nyblom indicates the rejection probability of 5%-level Nyblom (1989) test for stability of β, using the asymptotic critical value. Based on 10,000 replications.

happen asymptotically with probability one. In practice then one would take this as a signal that the maintained model of a single break is not appropriate for the data. The converse situation, where there are no breaks, will result in confidence sets that suggest a break could be anywhere and so for models without a break most dates will be included in the confidence set. The reason for this is that the test, looking for a break in the sample away from the maintained break date, will fail to reject with probability equal to one minus the level of the test. Also this property makes sense. If there is weak to no evidence of a break, then a procedure that tries to locate the break finds it could be anywhere.

4. Small sample evaluation

This section explores the small sample properties of the confidence sets suggested here and those derived in Bai (1997b). We find that the analytic results of Section 2 accurately predict the performance of Bai's (1997b) confidence intervals, as they tend to substantially and systematically undercover when the break magnitude is not very large. In most practical applications this renders these intervals uninterpretable. Since we do not know a priori the size of the break, we cannot tell whether the intervals provide an accurate idea as

Table 4
Empirical small sample coverage and length of confidence sets: model (M2): constant regressor, disturbances with breaking variance

	$d=4$		$d=8$		$d=12$		$d=16$	
	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.
$r_0=0.5$								
$\hat{\mathrm{U}}_T(\tau_{\mathrm{m}})$.eq	0.936	85.1	0.936	68.8	0.936	46.0	0.936	29.1
$\hat{\mathrm{U}}_T(\tau_{\mathrm{m}})$.neq	0.950	85.4	0.950	67.5	0.950	44.6	0.950	28.6
Bai.eq	0.572	54.2	0.735	49.5	0.846	34.9	0.894	22.5
Bai.het	0.572	54.2	0.735	49.5	0.846	34.9	0.894	22.5
Bai.hneq	0.614	53.5	0.762	47.5	0.869	34.8	0.918	23.0
Nyblom	0.204		0.613		0.922		0.996	
$r_0=0.35$								
$\hat{\mathrm{U}}_T(\tau_{\mathrm{m}})$.eq	0.963	87.5	0.963	74.5	0.963	53.6	0.963	34.9
$\hat{\mathrm{U}}_T(\tau_{\mathrm{m}})$.neq	0.954	86.9	0.954	71.4	0.954	48.6	0.954	30.7
Bai.eq	0.562	60.1	0.735	55.7	0.856	40.8	0.906	26.6
Bai.het	0.562	60.1	0.735	55.7	0.856	40.8	0.906	26.6
Bai.hneq	0.584	53.7	0.747	45.9	0.866	34.5	0.916	23.5
Nyblom	0.135		0.469		0.834		0.983	
$r_0=0.2$								
$\hat{\mathrm{U}}_T(\tau_{\mathrm{m}})$.eq	0.978	90.2	0.978	83.2	0.978	69.0	0.978	49.9
$\hat{\mathrm{U}}_T(\tau_{\mathrm{m}})$.neq	0.951	89.3	0.951	80.6	0.951	64.4	0.951	44.4
Bai.eq	0.550	62.4	0.694	55.7	0.829	43.4	0.904	30.6
Bai.het	0.550	62.4	0.694	55.7	0.829	43.4	0.904	30.6
Bai.hneq	0.552	54.6	0.681	43.5	0.814	32.1	0.897	23.3
Nyblom	0.071		0.196		0.442		0.717	

The model is $y_t=\beta+dT^{-1/2}\mathbf{1}[t>[r_0T]]+u_t$, $u_t=(1+\mathbf{1}[t>[r_0T]])\varepsilon_t$, $\varepsilon_t \sim iid\,\mathcal{N}(0,1)$, $T=100$. The notes of Table 3 apply.

to the uncertainty in the data over the break date. A comparison of confidence set lengths reveals that confidence sets constructed by inverting the sequence of tests based on $\hat{\mathrm{U}}_T(\tau_\mathrm{m})$ tend to be somewhat longer even for breaks that are large enough for Bai's (1997b) method to yield adequate coverage. At the same time, effective coverage rates of confidence sets constructed by inverting the tests $\hat{\mathrm{U}}_T(\tau_\mathrm{m})$ are very reliable and thus can be interpreted in the usual way.

The small sample data generating processes we consider are special cases of model (1)

$$y_t = X_t'\beta + \mathbf{1}[t > \tau_0]X_t'\delta + Z_t'\gamma + u_t, \quad t = 1, \ldots, T \tag{8}$$

with $T = 100$. Specifically, we consider six models: (M1) a break in the mean, such that $X_t = 1$ and there is no Z_t, and i.i.d. Gaussian disturbances $\{u_t\}$; (M2) same as model (M1), but with disturbances that are independent Gaussian with a variance that quadruples at the break date; (M3) same as model (M1), but with disturbances that are a mean zero stationary Gaussian AR(1) with coefficient 0.3; (M4) same as model (M1), but with disturbances that are a mean zero stationary Gaussian MA(1) with coefficient −0.3; (M5) $\{X_t\}$ a mean zero stationary Gaussian AR(1) with coefficient 0.5 and unit variance, $Z_t = 1$ and i.i.d. Gaussian disturbances $\{u_t\}$ independent of $\{X_t\}$; (M6) a heteroskedastic version

Table 5
Empirical small sample coverage and length of confidence sets: model (M3): constant regressor, AR(1) disturbances

	$d = 4$		$d = 8$		$d = 12$		$d = 16$	
	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.
$r_0 = 0.5$								
$\hat{\mathrm{U}}_T(\tau_\mathrm{m})$.eq	0.965	82.6	0.965	58.2	0.965	38.4	0.965	30.2
$\hat{\mathrm{U}}_T(\tau_\mathrm{m})$.neq	0.971	81.2	0.971	57.9	0.971	42.3	0.971	37.9
Bai.eq	0.734	55.8	0.890	32.2	0.942	16.5	0.962	10.2
Bai.het	0.734	55.8	0.890	32.2	0.942	16.5	0.962	10.2
Bai.hneq	0.708	53.7	0.874	33.0	0.930	17.1	0.952	10.6
Nyblom	0.359		0.884		0.994		1.000	
$r_0 = 0.35$								
$\hat{\mathrm{U}}_T(\tau_\mathrm{m})$.eq	0.963	83.2	0.963	59.7	0.963	39.1	0.963	30.3
$\hat{\mathrm{U}}_T(\tau_\mathrm{m})$.neq	0.970	82.3	0.970	60.8	0.970	45.1	0.970	40.5
Bai.eq	0.727	54.2	0.880	31.9	0.937	16.6	0.965	10.3
Bai.het	0.727	54.2	0.880	31.9	0.937	16.6	0.965	10.3
Bai.hneq	0.691	51.2	0.860	32.2	0.922	17.1	0.953	10.6
Nyblom	0.298		0.810		0.982		0.998	
$r_0 = 0.2$								
$\hat{\mathrm{U}}_T(\tau_\mathrm{m})$.eq	0.965	85.8	0.965	68.7	0.965	47.8	0.965	34.6
$\hat{\mathrm{U}}_T(\tau_\mathrm{m})$.neq	0.970	85.9	0.970	72.9	0.970	59.9	0.970	53.9
Bai.eq	0.710	49.7	0.857	31.0	0.928	17.0	0.961	10.5
Bai.het	0.710	49.7	0.857	31.0	0.928	17.0	0.961	10.5
Bai.hneq	0.644	44.0	0.815	29.4	0.896	16.9	0.942	10.6
Nyblom	0.149		0.455		0.762		0.905	

The model is $y_t = \beta + dT^{-1/2}\mathbf{1}[t > [r_0 T]] + u_t$, $u_t = 0.3u_{t-1} + \varepsilon_t$, $\varepsilon_t \sim iid\,\mathcal{N}(0, 0.49)$, $T = 100$. The notes of Table 3 apply.

of (M5), where the disturbances $\{u_t\}$ are given by $\{\varepsilon_t|X_t|\}$, where $\{\varepsilon_t\}$ are i.i.d. Gaussian independent of $\{X_t\}$. The variance of the disturbances is normalized throughout such that the long-run variance Ω_1 of $\{X_t u_t\}$ prior to the break is unity (that is, the spectral density of the stationary process $\{X_t u_t\}$ of the prebreak data generating process evaluated at zero is $1/(2\pi)$).

In models with uncorrelated $\{X_t u_t\}$, i.e. (M1), (M2), (M5) and (M6), we estimate variances rather than long-run variances of $\{X_t u_t\}$. For models (M3) and (M4), we employ in all methods the Andrews and Monahan (1992) AR(1) prewhitened second stage automatic bandwidth quadratic spectral estimator, where the bandwidth selection is based on an AR(1) model. We consider a version of $\hat{\mathrm{U}}_T(\tau_\mathrm{m})$ that imposes equivalence of the long-run variances of $\{X_t u_t\}$ prior to and after the break, $\Omega_1 = \Omega_2$, denoted by $\hat{\mathrm{U}}_T(\tau_\mathrm{m})$.eq, and one that does not, denoted by $\hat{\mathrm{U}}_T(\tau_\mathrm{m})$.neq. While $\hat{\mathrm{U}}_T(\tau_\mathrm{m})$ is automatically robust against heteroskedasticity, this is not the case for the basic Bai confidence set (2). We therefore compute three versions of Bai confidence sets: one imposing both $\Omega_1 = \Omega_2$ and homoskedasticity (Bai.eq), one imposing $\Omega_1 = \Omega_2$ but allowing for heteroskedasticity (Bai.het) and one allowing for both $\Omega_1 \neq \Omega_2$ and heteroskedasticity (Bai.hneq). In models (M1)–(M4), of course, Bai.eq = Bai.het.

Table 6
Empirical small sample coverage and length of confidence sets: model (M4): constant regressor, MA(1) disturbances

	$d=4$		$d=8$		$d=12$		$d=16$	
	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.
$r_0 = 0.5$								
$\hat{\mathrm{U}}_T(\tau_\mathrm{m})$.eq	0.975	82.7	0.975	55.0	0.975	31.2	0.975	21.1
$\hat{\mathrm{U}}_T(\tau_\mathrm{m})$.neq	0.978	81.8	0.978	54.5	0.978	32.6	0.978	23.3
Bai.eq	0.685	57.2	0.893	38.7	0.950	20.6	0.967	12.6
Bai.het	0.685	57.2	0.893	38.7	0.950	20.6	0.967	12.6
Bai.hneq	0.684	57.0	0.887	39.4	0.943	21.0	0.959	12.9
Nyblom	0.291		0.875		0.998		1.000	
$r_0 = 0.35$								
$\hat{\mathrm{U}}_T(\tau_\mathrm{m})$.eq	0.976	83.4	0.976	57.0	0.976	31.8	0.976	21.1
$\hat{\mathrm{U}}_T(\tau_\mathrm{m})$.neq	0.977	82.7	0.977	56.9	0.977	33.7	0.977	23.7
Bai.eq	0.674	54.4	0.886	37.8	0.948	20.6	0.966	12.6
Bai.het	0.674	54.4	0.886	37.8	0.948	20.6	0.966	12.6
Bai.hneq	0.674	54.3	0.880	38.3	0.939	21.0	0.957	12.9
Nyblom	0.235		0.797		0.992		1.000	
$r_0 = 0.2$								
$\hat{\mathrm{U}}_T(\tau_\mathrm{m})$.eq	0.972	86.2	0.972	67.6	0.972	40.3	0.972	22.8
$\hat{\mathrm{U}}_T(\tau_\mathrm{m})$.neq	0.976	86.0	0.976	69.1	0.976	44.9	0.976	28.4
Bai.eq	0.654	48.0	0.845	34.2	0.931	20.5	0.956	12.8
Bai.het	0.654	48.0	0.845	34.2	0.931	20.5	0.956	12.8
Bai.hneq	0.640	48.0	0.830	34.4	0.917	20.8	0.946	13.0
Nyblom	0.113		0.433		0.820		0.978	

The model is $y_t = \beta + dT^{-1/2}\mathbf{1}[t > [r_0 T]] + u_t$, $u_t = \varepsilon_t - 0.3\varepsilon_{t-1}$, $\varepsilon_t \sim iid\,\mathcal{N}(0, 2.04)$, $T = 100$. The notes of Table 3 apply.

Table 7
Empirical small sample coverage and length of confidence sets: model (M5): stochastic regressor, i.i.d. disturbances

	$d=4$		$d=8$		$d=12$		$d=16$	
	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.
$r_0=0.5$								
$\hat{\mathrm{U}}_T(\tau_\mathrm{m})$.eq	0.954	79.7	0.954	51.7	0.954	31.0	0.954	21.9
$\hat{\mathrm{U}}_T(\tau_\mathrm{m})$.neq	0.955	79.1	0.955	51.2	0.955	32.0	0.955	23.5
Bai.eq	0.699	54.7	0.856	33.7	0.899	17.5	0.902	10.7
Bai.het	0.682	51.8	0.842	31.9	0.889	16.8	0.893	10.4
Bai.hneq	0.647	49.6	0.819	32.2	0.873	17.3	0.886	10.7
Nyblom	0.373		0.882		0.994		1.000	
$r_0=0.35$								
$\hat{\mathrm{U}}_T(\tau_\mathrm{m})$.eq	0.953	80.5	0.953	54.0	0.953	32.1	0.953	22.2
$\hat{\mathrm{U}}_T(\tau_\mathrm{m})$.neq	0.954	80.2	0.954	53.9	0.954	33.3	0.954	23.9
Bai.eq	0.693	53.1	0.856	33.3	0.896	17.6	0.903	10.8
Bai.het	0.671	50.3	0.841	31.6	0.885	16.9	0.896	10.5
Bai.hneq	0.639	47.7	0.820	31.6	0.866	17.3	0.880	10.7
Nyblom	0.313		0.803		0.978		0.998	
$r_0=0.2$								
$\hat{\mathrm{U}}_T(\tau_\mathrm{m})$.eq	0.954	83.3	0.954	63.6	0.954	41.4	0.954	27.3
$\hat{\mathrm{U}}_T(\tau_\mathrm{m})$.neq	0.958	83.8	0.958	65.1	0.958	44.0	0.958	30.4
Bai.eq	0.666	48.4	0.819	31.8	0.881	18.0	0.900	11.0
Bai.het	0.650	46.1	0.803	30.2	0.870	17.2	0.893	10.7
Bai.hneq	0.601	42.6	0.760	28.8	0.832	17.2	0.865	10.8
Nyblom	0.169		0.505		0.782		0.914	

The model is $y_t = X_t\beta + dT^{-1/2}X_t\mathbf{1}[t>[r_0T]] + \gamma + u_t$, $u_t \sim iid\,\mathcal{N}(0,1)$, $X_t = 0.5X_{t-1}+\xi_t$, $\xi_t \sim iid\,\mathcal{N}(0,0.75)$, $T=100$. The notes of Table 3 apply.

Tables 3–8 show the empirical coverage rates and average confidence set lengths for the confidence interval (2) and confidence sets constructed by inverting the test statistics $\hat{\mathrm{U}}_T(\tau_\mathrm{m})$ as described in Section 3, based on 10,000 replications. In all experiments, we consider confidence sets of 95% nominal coverage, and breaks that occur at date $[r_0T]$, where $r_0 = 0.5$, 0.35 and 0.2. The tables also include the rejection probability of a 5%-level Nyblom test for the presence of a break in β, i.e. based on the test statistic $\mathrm{Ny} = T^{-2}\sum_{t=1}^{T}(\sum_{s=1}^{t}X_s\hat{u}_s)'\hat{\Omega}^{-1}(\sum_{s=1}^{t}X_s\hat{u}_s)$, where $\hat{u}_t$ are the residuals of a regression of $\{y_t\}$ on $\{X_t, Z_t\}$, $\hat{\Omega} = T^{-1}\sum_{t=1}^{T}\hat{u}_t^2X_tX_t'$ in models without autocorrelation and in models (M3) and (M4), $\hat{\Omega}$ is Andrews and Monahan's (1992) long-run variance estimator of $\{X_t\hat{u}_t\}$. The Nyblom test is based on the asymptotic critical value; unreported results show size control to be very reasonable.

Overall, the small sample results confirm the asymptotic results of Section 2: the Bai method fails to cover the true break date with the correct probability so long as the break is small. For all six models and three break dates, the usual method for constructing confidence intervals has coverage far below nominal coverage whenever the break is small enough for the Nyblom statistic to have power substantially below one. For example, in

Table 8
Empirical small sample coverage and length of confidence sets: model (M6): stochastic regressor, heteroskedastic disturbances

	$d=4$		$d=8$		$d=12$		$d=16$	
	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.
$r_0=0.5$								
$\hat{\mathrm{U}}_T(\tau_{\mathrm{m}})$.eq	0.959	78.6	0.959	47.5	0.959	27.7	0.959	20.0
$\hat{\mathrm{U}}_T(\tau_{\mathrm{m}})$.neq	0.964	77.7	0.964	46.5	0.964	28.5	0.964	21.4
Bai.eq	0.547	27.4	0.745	12.7	0.857	6.9	0.921	4.8
Bai.het	0.742	53.9	0.879	29.0	0.938	15.0	0.969	9.4
Bai.hneq	0.674	48.1	0.849	27.6	0.923	14.6	0.958	9.1
Nyblom	0.413		0.922		0.996		1.000	
$r_0=0.35$								
$\hat{\mathrm{U}}_T(\tau_{\mathrm{m}})$.eq	0.959	79.5	0.959	49.6	0.959	28.9	0.959	20.5
$\hat{\mathrm{U}}_T(\tau_{\mathrm{m}})$.neq	0.964	78.8	0.964	48.6	0.964	29.3	0.964	21.7
Bai.eq	0.544	27.2	0.742	12.8	0.848	7.0	0.922	4.8
Bai.het	0.736	52.7	0.878	29.0	0.939	15.1	0.970	9.4
Bai.hneq	0.665	46.0	0.843	27.0	0.916	14.4	0.958	9.1
Nyblom	0.349		0.857		0.987		0.999	
$r_0=0.2$								
$\hat{\mathrm{U}}_T(\tau_{\mathrm{m}})$.eq	0.956	82.7	0.956	59.9	0.956	36.9	0.956	24.4
$\hat{\mathrm{U}}_T(\tau_{\mathrm{m}})$.neq	0.965	82.9	0.965	59.9	0.965	37.9	0.965	26.1
Bai.eq	0.515	27.6	0.712	13.5	0.844	7.2	0.914	4.9
Bai.het	0.716	50.1	0.852	28.8	0.930	15.5	0.965	9.6
Bai.hneq	0.629	41.4	0.795	24.5	0.900	13.8	0.947	8.7
Nyblom	0.191		0.556		0.825		0.934	

The model is $y_t = X_t\beta + dT^{-1/2}X_t\mathbf{1}[t>[r_0T]] + \gamma + u_t$, $u_t = \varepsilon_t|X_t|$, $\varepsilon_t \sim iid\,\mathcal{N}(0, 0.333)$, $X_t = 0.5X_{t-1} + \xi_t$, $\xi_t \sim iid\,\mathcal{N}(0, 0.75)$, $T = 100$. The notes of Table 3 apply.

model (M2) with $r_0 = 0.35$ and $d = 8$, the Nyblom test rejects for half of the samples, yet confidence intervals based on (2) have coverage below 75%. When power of the test for a break gets closer to one, coverage of these confidence intervals is closer but not necessarily at the nominal 95% rate. For example, in model (M5) with $r_0 = 0.35$ and $d = 12$, the Nyblom test rejects the null hypothesis of no break 98% of the time, yet coverage for these confidence intervals is still below 90%. It is only when the breaks are large enough to be essentially always detected that empirical coverage of the Bai confidence intervals equals nominal coverage.

For the cases where coverage is not controlled, there is no way of comparing the average lengths of the confidence sets. However, it is clear from the experiments that the undercoverage translates into confidence intervals (2) that are relatively short, giving a misleading impression as to the uncertainty over the break date. In contrast, confidence sets based on inverting $\hat{\mathrm{U}}_T(\tau_{\mathrm{m}})$ control coverage remarkably well. For the case where both the Bai method and the method suggested here result in confidence sets of correct coverage, however, it is seen that the Bai method delivers the smaller set. This effect is especially pronounced in models (M3) and (M4) that yield autocorrelated $\{X_tu_t\}$. Pronounced autocorrelations of the underlying disturbances render Nyblom (1989)-type tests ill behaved, with size and power of these tests strongly dependent on the long-run variance estimator employed—see Müller (2005). In addition, as pointed out by Vogelsang (1999) and Crainiceanu and Vogelsang (2002), long-run variance estimation adversely affects the power of stationarity tests, as the low-frequency component of the time varying deterministics are mistakenly attributed to low-frequency dynamics. This latter effect increases the length of the confidence sets based on $\hat{\mathrm{U}}_T(\tau_\mathrm{m})$, but has no effect on coverage.

When the break in the regression coefficient is accompanied by a break in the variance of $\{X_t u_t\}$, as in model (M2), the methods that account for that possibility perform somewhat better in terms of coverage and confidence set lengths. As one might expect, in the presence of heteroskedasticity as in model (M6), the Bai method that fails to account for heteroskedasticity does not do well. The effective coverage rates of the asymptotically robust versions of the Bai statistic get closer to the nominal level in model (M6) compared to the homoskedastic model (M5). The reason for this is that the normalization of the variance of $\{u_t\}$—in order to ensure a long-run variance of $\{X_t u_t\} = \{|X_t| X_t \varepsilon_t\}$ equal to unity—makes the disturbance variance of model (M6) smaller than in model (M5).

Overall, the small sample experiments are encouraging for constructing reliable confidence sets for the break date by inverting a sequence of tests based on $\hat{\mathrm{U}}_T(\tau_\mathrm{m})$. Empirical coverage rates are very close to nominal coverage rates for all data generating processes considered here, making the method developed in this paper an attractive choice for applied work.

5. Conclusion

It is more difficult to determine the date of a break than it is to distinguish between models with and without breaks. In practice, breaks that can be detected reasonably well with hypothesis tests are often difficult to date and standard methods of constructing confidence intervals for the break date fail to deliver an accurate description of this uncertainty. It may be possible to use subsampling or bootstrap techniques to account for these difficulties.

The approach taken in this paper is to use an alternative method of constructing a confidence set by inverting a sequence of tests. Each of the tests maintains the null hypothesis that the break occurs at a certain date. By imposing an invariance requirement, the tests control coverage for any magnitude of the break. The confidence sets so obtained hence control coverage also for a small break. In addition, the test statistics that are inverted have an (asymptotic) critical value that does not depend on the maintained break date. The confidence set can hence be computed relatively easily by comparing a sequence of T test statistics with a single critical value, where T is the sample size.

Acknowledgments

The authors would like to thank three anonymous referees and the editor, Peter Robinson, for helpful comments on a previous version. The authors also benefitted from discussions at Aarhus, ANU, Bank of Italy, Columbia, Duke, ECB, FRB at St. Louis, John Hopkins, UT Austin and UTS. Müller gratefully acknowledges financial support through NSF Grant SES 0518036.

Appendix

Proof of Proposition 1. For $\tau_{\rm m} \in [\bar{\lambda}T, (1-\bar{\lambda})T]$, let $l = \tau_{\rm m}/T$. Define $\eta_t = u_t + T^{-1/2}\mathbf{1}[t > \tau_0]X_t'd$, and let $\{\tilde{Z}_t\}$ be the least-squares residuals of a regression of $\{Z_t\}$ on $\{X_t, \mathbf{1}[t>\tau_{\rm m}]X_t\}$. By standard linear regression algebra, the sum of squared residuals of an OLS regression of $\{\eta_t\}$ on $\{X_t, \mathbf{1}[t>\tau_{\rm m}]X_t, Z_t\}$ is given by

$$\begin{aligned} S_T(\tau_{\rm m}) = \sum_{t=1}^{T}\eta_t^2 &- \left(\sum_{t=1}^{\tau_{\rm m}} X_t\eta_t\right)'\left(\sum_{t=1}^{\tau_{\rm m}} X_tX_t'\right)^{-1}\sum_{t=1}^{\tau_{\rm m}} X_t\eta_t \\ &- \left(\sum_{t=\tau_{\rm m}+1}^{T} X_t\eta_t\right)'\left(\sum_{t=\tau_{\rm m}+1}^{T} X_tX_t'\right)^{-1}\sum_{t=\tau_{\rm m}+1}^{T} X_t\eta_t \\ &- \left(\sum_{t=1}^{T} \tilde{Z}_t\eta_t\right)'\left(\sum_{t=1}^{T} \tilde{Z}_t\tilde{Z}_t'\right)^{-1}\sum_{t=1}^{T} \tilde{Z}_t\eta_t. \end{aligned}$$

For $t \leqslant \tau_{\rm m} = [lT]$, $\tilde{Z}_t = Z_t - (\sum_{s=1}^{\tau_{\rm m}} Z_sX_s')(\sum_{s=1}^{\tau_{\rm m}} X_sX_s')^{-1}X_t$ and, similarly, for $t > \tau_{\rm m}$, $\tilde{Z}_t = Z_t - (\sum_{s=\tau_{\rm m}+1}^{T} Z_sX_s')(\sum_{s=\tau_{\rm m}+1}^{T} X_sX_s')^{-1}X_t$. From the uniform convergence of $T^{-1}\sum_{t=1}^{[sT]} X_tZ_t'$ and $T^{-1}\sum_{t=1}^{[sT]} X_tX_t'$ in s and $\sup_{0<s<1}\|T^{-1/2}\sum_{t=1}^{[sT]} X_t\eta_t\| = {\rm O_p}(1)$ and $\sup_{0<s<1}\|T^{-1/2}\sum_{t=1}^{[sT]} Z_t\eta_t\| = {\rm O_p}(1)$ we find

$$\begin{aligned} \sup_{\bar{\lambda}T \leqslant \tau_{\rm m} \leqslant (1-\bar{\lambda})T} \Bigg| &\left(\sum_{t=1}^{T} \tilde{Z}_t\eta_t\right)'\left(\sum_{t=1}^{T} \tilde{Z}_t\tilde{Z}_t'\right)^{-1}\sum_{t=1}^{T} \tilde{Z}_t\eta_t \\ &- \left(\sum_{t=1}^{T} \check{Z}_t\eta_t\right)'\left(\sum_{t=1}^{T} \check{Z}_t\check{Z}_t'\right)^{-1}\sum_{t=1}^{T} \check{Z}_t\eta_t \Bigg| \xrightarrow{\rm p} 0, \end{aligned}$$

where $\check{Z}_t = Z_t - \Sigma_{ZX}\Sigma_X^{-1}X_t$. Note that $\check{Z}_t$ does not depend on $\tau_{\rm m}$. Furthermore, $T^{-1}\sum_{t=1}^{\tau_{\rm m}} X_tX_t' \xrightarrow{\rm p} l\Sigma_X$, $T^{-1/2}\sum_{t=1}^{\tau_{\rm m}} X_t\eta_t \Rightarrow M(l)$, $T^{-1}\sum_{t=\tau_{\rm m}+1}^{T} X_tX_t' \xrightarrow{\rm p} (1-l)\Sigma_X$ and $T^{-1/2}\sum_{t=\tau_{\rm m}+1}^{T} X_t\eta_t \Rightarrow M(1) - M(l)$ uniformly in $0 \leqslant l \leqslant 1$. Hence,

$$\begin{aligned} \arg\min_{\bar{\lambda}T \leqslant \tau_{\rm m} \leqslant (1-\bar{\lambda})T} S_T(\tau_{\rm m}) &= \arg\min_{\bar{\lambda} \leqslant l \leqslant 1-\bar{\lambda}} S_T([lT]) \\ &= \arg\max_{\bar{\lambda} \leqslant l \leqslant 1-\bar{\lambda}} \left(\sum_{t=1}^{[lT]} X_t\eta_t\right)'\left(\sum_{t=1}^{[lT]} X_tX_t'\right)^{-1}\sum_{t=1}^{[lT]} X_t\eta_t \\ &\quad + \left(\sum_{t=[lT]+1}^{T} X_t\eta_t\right)'\left(\sum_{t=[lT]+1}^{T} X_tX_t'\right)^{-1}\sum_{t=[lT]+1}^{T} X_t\eta_t + R_T(l) \\ &\Rightarrow \arg\max_{\bar{\lambda} \leqslant l \leqslant 1-\bar{\lambda}} G(l), \end{aligned}$$

where $\sup_{\bar{\lambda} \leqslant l \leqslant 1-\bar{\lambda}} |R_T(l)| = {\rm o_p}(1)$ and the last line follows from the continuous mapping theorem. The continuous mapping theorem is applicable due to the arguments put forward in Kim and Pollard (1990), as an application of their Theorem 2.7.

Let $\hat{\delta}(\tau_m)$ be the least-squares estimator of δ with τ_0 replaced by $\tau_m = [lT]$, $0 \leqslant l \leqslant 1$, in (1), and let $\{\tilde{X}_t\}$ be the residuals of a regression of $\{\mathbf{1}[t > \tau_m]X_t\}$ on $\{Q_t\}$. Then,

$$\hat{\delta}(\tau_m) = \left(\sum_{t=1}^{T} \tilde{X}_t \tilde{X}_t'\right)^{-1} \sum_{t=1}^{T} \tilde{X}_t \eta_t.$$

Now $\tilde{X}_t = \mathbf{1}[t > \tau_m]X_t - (\sum_{s=\tau_m+1}^{T} X_s Q_s')(\sum_{s=1}^{T} Q_s Q_s')^{-1} Q_t$. With $T^{-1}\sum_{t=\tau_m+1}^{T} Q_t Q_t' \xrightarrow{p} (1-l)\Sigma_Q$ uniformly in $0 \leqslant l \leqslant 1$ from Condition 1,

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T} \tilde{X}_t \tilde{X}_t' &= T^{-1}\sum_{t=1}^{T} \tilde{X}_t \mathbf{1}[t > \tau_m] X_t' \\
&= T^{-1}\sum_{t=\tau_m+1}^{T} X_t X_t' - T^{-1}\left(\sum_{t=\tau_m+1}^{T} X_t Q_t'\right)\left(\sum_{t=1}^{T} Q_t Q_t'\right)^{-1} \sum_{t=\tau_m+1}^{T} Q_t X_t' \\
&\xrightarrow{p} l(1-l)\Sigma_X,
\end{aligned}$$

and also

$$\begin{aligned}
T^{-1/2}\sum_{t=1}^{T} \tilde{X}_t \eta_t &= T^{-1/2}\sum_{t=\tau_m+1}^{T} X_t \eta_t - T^{-1/2}\left(\sum_{t=\tau_m+1}^{T} X_t Q_t'\right)\left(\sum_{t=1}^{T} Q_t Q_t'\right)^{-1} \sum_{t=1}^{T} Q_t \eta_t \\
&\Rightarrow lM(1) - M(l),
\end{aligned}$$

since $T^{-1/2}\sum_{t=1}^{T} Z_t \eta_t = O_p(1)$. The application of the continuous mapping theorem now yields the result for m. $\square$

Proof of Proposition 2. Let B be the $T \times (T - 2k - p)$ matrix that satisfies $B'B = I_{T-2k-p}$ and $BB' = M_R$, where M_R is the projection matrix off the column space spanned by $\{R_t\}$, where $R_t = (X_t', \mathbf{1}[t > \tau_m]X_t', Z_t')'$. Let $y = (y_1, \ldots, y_T)'$, and denote by $\{e_t\}$ the OLS residuals from a regression of $\{y_t\}$ on $\{R_t\}$. Then $(B'y, Q)$ is a maximal invariant to the group of transformations (4). Furthermore, conditional on Q, $B'y \sim \mathcal{N}(B'\Xi(\tau_0)\delta, \sigma^2 I_{T-2k-p})$, where $\Xi(t)$ is a $T \times k$ matrix with sth row X_s' when $s > t$ and a $1 \times k$ zero row vector otherwise. By the Neyman–Pearson lemma, Fubini's theorem and the likelihood structure in Condition 2, an efficient invariant test of (3) maximizing (5) can hence be based on

$$\begin{aligned}
LR_T &= \sum_{t \neq \tau_m} w_t \int \exp\left[\sigma^{-2} y' BB' \Xi(t) f - \frac{1}{2}\sigma^{-2} f' \Xi(t)' BB' \Xi(t) f\right] \mathrm{d}\nu_t(f) \\
&= \sum_{t \neq \tau_m} w_t F(t).
\end{aligned}$$

Under the choice of weight functions (6), we compute for $t < \tau_m$,

$$\begin{aligned}
F(t) &= \int \exp\left[\sigma^{-2} y' M_R \Xi(t) f - \frac{1}{2}\sigma^{-2} f' \Xi(t)' M_R \Xi(t) f\right] \mathrm{d}\nu_t(f) \\
&= \int (2\pi)^{-k/2} |b^2 \tau_m^{-2} \Omega_1^{-1}|^{-1/2} \\
&\quad \times \exp\left[\sigma^{-2} f' \sum_{s=t}^{T} X_s e_s - \frac{1}{2}\sigma^{-2} f' \Xi(t)' M_R \Xi(t) f - \frac{1}{2} b^{-2} \tau_m^2 f' \Omega_1 f\right] \mathrm{d}f
\end{aligned}$$

$$
\begin{aligned}
&= |b^2\tau_{\mathrm{m}}^{-2}\Omega_1^{-1}|^{-1/2}|b^{-2}\tau_{\mathrm{m}}^2\Omega_1 + \sigma^{-2}\Xi(t)'M_R\Xi(t)|^{-1/2} \\
&\quad \times \exp\left[\frac{1}{2}\sigma^{-4}\left(\sum_{s=t}^{T} X_s e_s\right)'(b^{-2}\tau_{\mathrm{m}}^2\Omega_1 + \sigma^{-2}\Xi(t)'M_R\Xi(t))^{-1}\left(\sum_{s=t}^{T} X_s e_s\right)\right] \\
&= |I_k + b^2\sigma^{-2}\tau_{\mathrm{m}}^{-2}\Omega_1^{-1}\Xi(t)'M_R\Xi(t)|^{-1/2} \\
&\quad \times \exp\left[\frac{1}{2}\sigma^{-4}b^2\left(\sum_{s=t}^{\tau_{\mathrm{m}}} X_s e_s\right)'(\tau_{\mathrm{m}}^2\Omega_1 + \sigma^{-2}b^2\Xi(t)'M_R\Xi(t))^{-1}\left(\sum_{s=t}^{\tau_{\mathrm{m}}} X_s e_s\right)\right]
\end{aligned}
$$

since $\sum_{s=\tau_{\mathrm{m}}+1}^{T} X_s e_s = 0$. By a one-term Taylor expansion around $b^2 = 0$,

$$
\begin{aligned}
2(F(t) - 1) = \sigma^{-4}b^2\tau_{\mathrm{m}}^{-2}&\left(\sum_{s=t}^{\tau_{\mathrm{m}}} X_s e_s\right)'\Omega_1^{-1}\left(\sum_{s=t}^{\tau_{\mathrm{m}}} X_s e_s\right) \\
&- b^2\sigma^{-2}\tau_{\mathrm{m}}^{-2}\,\mathrm{tr}(\Omega_1^{-1}\Xi(t)'M_R\Xi(t)) + \mathrm{o}(b^2).
\end{aligned}
$$

Proceeding analogously for $t > \tau_{\mathrm{m}}$ and collecting terms whose distribution depends on δ and τ_0 yield the result. □

Proof of Proposition 3. Proceed similarly as in the proof of Proposition 1 to show that under Condition 1, for $s \leqslant r_0$,

$$
\begin{aligned}
T^{-1/2}\sum_{t=1}^{[sT]} X_t e_t &= T^{-1/2}\sum_{t=1}^{[sT]} X_t u_t - \left(\sum_{t=1}^{[sT]} X_t X_t'\right)\left(\sum_{t=1}^{[r_0T]} X_t X_t'\right)^{-1}\left(T^{-1/2}\sum_{t=1}^{[r_0T]} X_t u_t\right) \\
&\quad - \left(\sum_{t=1}^{[sT]} X_t \tilde{Z}_t'\right)\left(\sum_{t=1}^{T} \tilde{Z}_t\tilde{Z}_t'\right)^{-1}\left(T^{-1/2}\sum_{t=1}^{T} \tilde{Z}_t u_t\right) \\
&\Rightarrow \Omega_1^{1/2}\left(W(s) - \frac{s}{r_0}W(r_0)\right).
\end{aligned}
$$

With the analogous result for $s > r_0$, the proposition becomes a consequence of the continuous mapping theorem. □

References

Andrews, D., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.

Andrews, D., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61, 821–856.

Andrews, D., 2003. End-of-sample instability tests. Econometrica 71, 1661–1694.

Andrews, D., Monahan, J., 1992. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. Econometrica 60, 953–966.

Andrews, D., Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative. Econometrica 62, 1383–1414.

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

Bai, J., 1994. Least squares estimation of a shift in linear processes. Journal of Time Series Analysis 15, 453–470.

Bai, J., 1997a. Estimating multiple breaks one at a time. Econometric Theory 13, 315–352.

Bai, J., 1997b. Estimation of a change point in multiple regressions. Review of Economics and Statistics 79, 551–563.

Bai, J., 1999. Likelihood ratio tests for multiple structural changes. Journal of Econometrics 91, 299–323.

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

Bai, J., Perron, P., 2006. Multiple structural change models: a simulation analysis. In: Corbea, D., Durlauf, S., Hansen, B. (Eds.), Econometric Theory and Practice: Frontiers of Analysis and Applied Research. Cambridge University Press, Cambridge, pp. 212–237.

Bai, J., Lumsdaine, R., Stock, J., 1998. Testing for and dating common breaks in multivariate time series. Review of Economic Studies 65, 395–432.

Bhattacharya, P., 1987. Maximum likelihood estimation of a change-point in the distribution of independent random variables: general multiparameter case. Journal of Multivariate Analysis 23, 183–208.

Bhattacharya, P., 1994. Some aspects of change-point analysis. In: Carlstein, E., Müller, H., Siegmund, D. (Eds.), Change-Point Problems. Institute of Mathematical Statistics, Hayward, pp. 28–56.

Boivin, J., 2003. Has U.S. monetary policy changed? Evidence from drifting coefficients and real time data. Mimeo, Columbia University.

Busetti, F., Harvey, A., 2001. Testing for the presence of a random walk in a series with structural breaks. Journal of Time Series Analysis 22, 127–150.

Clarida, R., Gali, J., Gertler, M., 2000. Monetary policy rules and macroeconomic stability: evidence and some theory. Quarterly Journal of Economics 115, 147–180.

Crainiceanu, C., Vogelsang, T., 2002. Spectral density bandwidth choice: source of nonmonotonic power for tests of a mean shift in a time series. Mimeo, Cornell University.

Elliott, G., Müller, U., 2006. Efficient tests for general persistent time variation in regression coefficients. Review of Economic Studies 73, 907–940.

Forchini, G., 2002. Optimal similar tests for structural change for the linear regression model. Econometric Theory 18, 853–867.

Hansen, B., 2000. Sample splitting and threshold estimation. Econometrica 68, 575–603.

Hawkins, D., 1977. Testing a sequence of observations for a shift in location. Journal of the American Statistical Association 72, 180–186.

Hinkley, D., 1970. Inference about the change-point in a sequence of random variables. Biometrika 57, 1–17.

Kim, J., Pollard, D., 1990. Cube root asymptotics. The Annals of Statistics 18, 191–219.

Kurozumi, E., 2002. Testing for stationarity with a break. Journal of Econometrics 108, 63–99.

Lehmann, E., 1986. Testing Statistical Hypotheses. Wiley, New York.

Lettau, M., Ludvigson, S., 2001. Consumption, aggregate wealth and expected stock returns. Journal of Finance 56, 815–849.

MacNeill, I., 1978. Properties of sequences of partial sums of polynomial regression residuals with applications to test for change of regression at unknown times. The Annals of Statistics 6, 422–433.

Müller, U., 2005. Size and power of tests of stationarity in highly autocorrelated time series. Journal of Econometrics 128, 195–213.

Nabeya, S., Tanaka, K., 1988. Asymptotic theory of a test for constancy of regression coefficients against the random walk alternative. The Annals of Statistics 16, 218–235.

Newey, W., West, K., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Nyblom, J., 1989. Testing for the constancy of parameters over time. Journal of the American Statistical Association 84, 223–230.

Orphanides, A., 2004. Monetary policy rules, macroeconomic stability, and inflation: a view from the trenches. Journal of Money, Credit, and Banking 36, 151–175.

Pesaran, M., Timmermann, A., 2002. Market timing and return prediction under model instability. Journal of Empirical Finance 9, 495–510.

Pratt, J., 1961. Length of confidence intervals. Journal of the American Statistical Association 56, 549–567.

Stock, J., 1991. Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series. Journal of Monetary Economics 28, 435–459.

Stock, J., 1994. Unit roots, structural breaks and trends. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, vol. 4. North-Holland, New York, pp. 2740–2841.

Stock, J., Watson, M., 1996. Evidence on structural instability in macroeconomic time series relations. Journal of Business and Economic Statistics 14, 11–30.

Timmermann, A., Paye, B., 2004. Instability of return prediction models. Mimeo, UCSD.

Vogelsang, T., 1999. Sources of nonmonotonic power when testing for a shift in mean of a dynamic time series. Journal of Econometrics 88, 283–299.

Worsley, K., 1979. On the likelihood ratio test for a shift in locations of normal populations. Journal of the American Statistical Association 72, 180–186.

Worsley, K., 1986. Confidence regions and tests for a change-point in a sequence of exponential family random variables. Biometrika 73, 91–104.

Zacks, S., 1983. Survey of classical and Bayesian approaches to the change-point problem: fixed sample and sequential procedures of testing and estimation. In: Rizivi, M., Rustagi, J., Siegmund, D. (Eds.), Recent Advances in Statistics. Academic Press, New York, pp. 245–269.