Title
Confidence Sets for the Date of a Single Break in Linear Time Series Regressions

Authors
Elliott, Graham
Muller, Ulrich K.

Publication Date
2004-08-01

Confidence Sets for the Date of a Single Break in Linear Time Series Regressions

Graham Elliott
University of California, San Diego
9500 Gilman Drive
La Jolla, CA, 92093-0508

Ulrich K. Müller
Princeton University
Princeton, NJ, 08544-1021

First Draft: April 2004
This Version: August 2004

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 currently standard method in econometrics to construct 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 highly 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 modified 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 international macroeconomics. Lettau and Ludvigson (2001) and Paye and Timmmermann (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 timing of the break is typically of greater interest. This paper examines a multiple regression model and considers inference about the time of a single break in a subset of the coefficients.

Locating where 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 timing of the break.

The literature on estimation and construction of confidence sets for break dates back to Hinkley (1970), Hawkins (1977), Worsley (1979), Worsley (1986), Bhattacharya (1987) and others—see the reviews by Zacks (1983), Stock (1994) and Bhattacharya (1995) for additional discussion and references. The standard econometric method to construct confidence intervals for the date of breaks relies on work by Bai (1994), which is further developed in Bai (1997a), Bai (1997b), Bai, Lumsdaine, and Stock (1998), Bai and Perron (1998) and Bai (1999). 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 thought experiment that underlies Bai's (1994, 1997b) results is such that the magnitude of the break shrinks, 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 one. In other words, this asymptotic thought experiment focusses on a the part of the parameter space of the magnitude of the break that corresponds to a 'large' break, as the p-values of tests for a break will converge to zero. Inference for the presence of a break becomes trivial for such a 'large' break, although the exact timing of the break remains uncertain. In contrast, one might speak of a 'small' break when both the presence and the timing of the break is uncertain. Analytically, a small break can be represented by an asymptotic thought experiment 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 on the 1% level, for an additional 11 series on the 5% level, and for an additional 6 series on 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 or 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 Volker chairmanship of the US Federal Reserve Board are economically important although they did not test the break. Boivin (2003) finds based on tests and a robustness analysis that a fixed 'Volker' 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 to assessing the conduct of monetary policy. It is also found that instabilities arising through Lucas-critique arguments have been difficult to find empirically (Linde (2001)) and are by implication small, too.

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 growth. Post-war quarterly U.S. real Gross Domestic Product growth has a standard deviation of about unity. Even if growth is 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 of that is 1 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 thought experiment that validates the confidence intervals developed in Bai (1997a). 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.

The question hence arises how to construct valid confidence sets for the date of a break when the break is, at least potentially, small. We suggest basing confidence sets on the inversion of a sequence of tests for a break. 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 hypotheses at the maintained break, 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 sets is given by 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 the tests is correct for any magnitude of the break. By a judicious choice of the efficient test that we suggest to invert, the critical values of the sequence of test statistics does not depend on the maintained break date, at least asymptotically. 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 four evaluates the methods numerically for some standard small sample data generating processes.

2 Properties of Standard Confidence Intervals When Breaks Are Small

This paper considers the linear time series regression model

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

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. We assume 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* $s \leq r_0$ *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* $s \geq r_0$ *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) $\sup_s |T^{-1/2}\sum_{t=1}^{[sT]} Z_t u_t| = O_p(1)$.

(iv) $T^{-1}\sum_{t=[(r_0-s)T]}^{[r_0 T]} Q_t Q_t' \xrightarrow{p} s\Sigma_{Q1} = s\begin{pmatrix} \Sigma_{X1} & \Sigma_{XZ1} \\ \Sigma_{ZX1} & \Sigma_{Z1} \end{pmatrix}$, $T^{-1}\sum_{t=[r_0 T]+1}^{[(r_0+s)T]} Q_t Q_t' \xrightarrow{p} s\Sigma_{XQ2} = s\begin{pmatrix} \Sigma_{X2} & \Sigma_{XZ2} \\ \Sigma_{ZX2} & \Sigma_{Z2} \end{pmatrix}$ *uniformly in* $s \geq 0$, *where* Σ_{Q1} *and* Σ_{Q2} *are full rank.*

In the asymptotic thought experiments considered in this paper, the data that precedes and follows the break are in the fixed proportion $r_0/(1 - r_0)$. This thought experiment is standard in the breaking 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: Compute the break date estimator $\hat{\tau}$ which is given by the value of τ that minimizes the sum of squared residuals of the linear regression (1), where τ_0 in the indicator function is replaced by τ. Denote the coefficient estimate of δ that corresponds to this minimizing choice of τ in the indicator function by $\hat{\delta}$. A level C confidence interval for τ_0 is then given by

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

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), the intervals (2) are asymptotically valid in the thought experiment where $\delta = T^{-1/2+\epsilon} d$ for some $0 < \epsilon < 1/2$ 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 and Ploberger (1994) and Elliott and Müller (2003) have nontrivial local asymptotic power is where $\epsilon = 0$. In other words, in the asymptotic thought experiment that justifies the confidence intervals (2) the p-values of any standard test for breaks converge to zero. With $0 < \epsilon < 1/2$, 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 location.

In fact, when $0 < \epsilon < 1/2$, $\hat{\tau}/T$ is a consistent estimator of r_0—see Bai (1997b). The break is large enough to pinpoint down exactly its location 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 very much unclear whether breaks typically encountered in practice are necessarily large enough for this asymptotic thought experiment 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 important question about 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) in the asymptotic thought experiment where $\delta = T^{-1/2}d$, i.e. where $\epsilon = 0$. 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. When d is very large, then the probability of detecting the break in this thought experiment 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 set-up with $0 < \epsilon < 1/2$, 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 location in asymptotics with $\delta = T^{-1/2}d$ extends to the fraction r_0.

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 $\Sigma_{Q1} = \Sigma_{Q2} = \Sigma_Q$.

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

$$\begin{aligned} M(s) &= \Omega^{1/2}W(s) + \mathbf{1}[s \geq r_0](s - r_0)\Sigma_X d \\ G(s) &= \frac{M(s)'\Sigma_X^{-1}M(s)}{s} + \frac{(M(1) - M(s))'\Sigma_X^{-1}(M(1) - M(s))}{1 - s} \\ \hat{r}_a &= \arg\max_{\bar{\lambda} \leq s \leq 1-\bar{\lambda}} G(s) \\ \hat{\delta}_a &= \Sigma_X^{-1}\frac{\hat{r}_a M(1) - M(\hat{r}_a)}{\hat{r}_a(1 - \hat{r}_a)}. \end{aligned}$$

Then under Condition 1 and $\delta = T^{-1/2}d$,

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

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

There are several points to make about the result in Proposition 1. First, in the statement of the proposition, the potential choices of the break date are trimmed away from the endpoints. While such trimming is standard in the literature on tests for breaks (Andrews (1993), Andrews and Ploberger (1994)), it is quite innocuous in the statement of the Proposition. The reason is that $W(s)'W(s)/s$ converges to zero as $s \to 0$ almost surely, so that the probability for $\hat{r}_a$ to be close to the bounds is very small even for a very small choice of $\bar{\lambda}$.

Second, the margin of error of the confidence intervals (2) becomes of order $m = O_p(T)$ (and not $o_p(T)$). As discussed, the uncertainty about the break location 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.

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. It is hence not possible to construct asymptotically justified confidence intervals for local asymptotics by adding and subtracting the margin of error m from $\hat{\tau}$.

Table 1 depicts the asymptotic coverage rates of nominal 95% confidence intervals (2) for $k = 1$, $\Omega = \Sigma_X$ and various values of $\Omega^{-1/2}d$ and r_0, along with the asymptotic local power of a 5%-level Nyblom (1989) test for breaks. For $d = 8$, coverage rates are below 87%, and much smaller still for $d = 4$. This is despite the fact that breaks with $d = 8$ have a very high probability of being detected with Nyblom's tests for breaks, at least as long as they do not occur at the very beginning or very end of the sample. The asymptotic distribution of p-values of the Nyblom test for $d = 4$ and $r_0 = 0.35$ is such that 17% are below 1%, 20% are between 1%–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 macro series, although this comparison obviously suffers from the lack of independence of the macro series. 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 Section 1, 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

Table 1: Local Asymptotic Properties of Bai's (1997b) CIs

	$r_0 = 0.5$		$r_0 = 0.35$		$r_0 = 0.2$	
$\Omega^{-1/2}d$	Cov.	Nybl.	Cov.	Nybl.	Cov.	Nybl.
4	0.564	0.438	0.544	0.375	0.484	0.204
8	0.862	0.953	0.844	0.915	0.793	0.651
12	0.927	1.000	0.920	0.999	0.906	0.956
16	0.939	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.

variance), then $d = 12$ corresponds to a break in the quarterly growth rate of $12/\sqrt{180} = 0.89$ 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 percent on a yearly basis!

This asymptotic analysis suggests that the standard way of constructing confidence intervals based on (2) leads to substantial undercoverage in small samples 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 highly relevant empirically. This raises the question of how to construct confidence intervals 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. In standard set-ups, estimators are asymptotically unbiased and Gaussian with a variance that 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 is 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 $(1 - C)$ significance tests, each maintaining that under the null hypothesis, $\tau_0 = \tau$ for $\tau = 1, \cdots, T$. As long as each of these tests has correct level, the resulting confidence set has correct coverage, as the probability of excluding the correct value is identical to the type I error of the employed significance test. When $\tau_0 \neq \tau$, the break will occur at a date different to the maintained break. Tests that reject with high probability for a break that occurs at a date other than the maintained break date τ will result in short confidence sets. The more powerful the sequence of tests, 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 against the alternative that the break occurs at some other date

$$H_0 : \tau_0 = \tau \qquad \text{against} \qquad H_1 : \tau_0 \neq \tau. \tag{3}$$

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 $\tau_0 \neq \tau$. (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. 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

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

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 $\{y_{t-1}\}$.

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 a sequence of independent and mean zero Gaussian variates of variance σ^2.*

(ii) The conditional distribution of Q_t given $\{Q_{t-1}, Q_{t-2}, \cdots, y_{t-1}, y_{t-2}, ...\}$ is independent of $\{u_t\}$, and it does not depend on β, γ, δ and τ_0. Furthermore, the unconditional distribution of $\{Q_t\}$ does not depend on δ and τ_0.

Part (i) of the condition specifies the distribution of $\{u_t\}$ to be Gaussian. Only the efficiency of the following tests depends on this (often unrealistic) assumption, but not the validity of the resulting tests. 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.

Part (ii) of Condition 2 requires the conditional distribution of Q_t given past values of Q_t and y_t not to depend on β, γ, δ and τ_0, which is the assumption of weak exogeneity as described in detail by Engle, Hendry and Richard (1983). This assumption will allow a factorization of the likelihood of $\{y_t, Q_t\}_{t=1}^T$ into two pieces, one capturing the contribution to the likelihood of $u_t = y_t - X_t'\beta - \mathbf{1}[t > \tau_0]X_t'\delta - Z_t'\gamma$ and the other the contribution of Q_t conditional on $\{Q_{t-1}, Q_{t-2}, \cdots, y_{t-1}, y_{t-2}, ...\}$. The independence of the latter piece of the parameters ensures that it cancels in the ratio of the likelihoods of the null and alternative hypothesis, making the resulting optimal statistic independent of the conditional distribution of Q_t. In addition, the break parameters δ and τ_0 are assumed strictly exogenous to the regressors, which rules out certain breaks of coefficients of weakly exogenous regressors. Again, the requirement of this strict exogeneity only affects the small sample optimality of the test statistic (7) below, tests remain asymptotically valid as long as Condition 1 holds.

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 $\tau_0 = \tau$ against $\tau_0 \neq \tau$ when it maximizes the weighted average power criterion

$$\sum_{t \neq \tau} w_t \int P(\varphi \text{ rejects} | \tau_0 = t, \, \delta = d) d\nu_t(d) \tag{5}$$

over all tests which satisfy $P(\varphi \text{ rejects} | \tau_0 = \tau) = \alpha$, where $\{w_t\}_{t=1}^T$ is a sequence of nonnegative real numbers, and $\{\nu_t\}_{t=1}^T$ is a sequence of nonnegative measures on $\mathbb{R}^k$. The prespecified sequences $\{w_t\}_{t=1}^T$ and $\{\nu_t\}_{t=1}^T$ direct the power against alternatives of certain dates τ_0 and break magnitudes, respectively. From a Bayesian perspective, the weights $\{w_t\}$ and $\{\nu_t\}$, 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 depends on the weighting functions $\{w_t\}$ and $\{\nu_t\}$, so that the question arises how to make a suitable choice. As demonstrated in Elliott and Müller (2003), 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 ν_t that are probability measures of mean zero $k \times 1$ Gaussian vector with covariance matrix $b^2 H_t$, where

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

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 ν_t corresponds to the long-run covariance matrix of $\{X_t u_t\}$ (which depends on whether $t < \tau$ or $t > \tau$). The magnitude of the potential break is piecewise constant before and after the maintained break date τ. Even if $\Omega_1 = \Omega_2$, the break size will not be identical, though, but depends on τ: When τ is close to T, for instance, then this choice of ν_t puts less weight on large breaks that occur prior to τ 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. 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.

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*

$$U_T(\tau) = \tau^{-2} \sum_{t=1}^{\tau} \left(\sum_{s=1}^{t} v_s \right)' \Omega_1^{-1} \left(\sum_{s=1}^{t} v_s \right) + (T-\tau)^{-2} \sum_{t=\tau+1}^{T} \left(\sum_{s=\tau+1}^{T} v_s \right)' \Omega_2^{-1} \left(\sum_{s=\tau+1}^{T} v_s \right) \tag{7}$$

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.

Busetti and Harvey (2001) and Kurozumi (2002) suggest a specialized version of $U_T(\tau)$ for constant and trending $\{X_t\}$ as a test statistic for the null of stationarity under a maintained break at date τ, 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 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, although one would not recover the usual asymptotic distribution, as the regressor $\{\mathbf{1}[t > \tau_0]X_t\}$ 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 $t = 1, \cdots, \tau$ and one for the regression $t = \tau + 1, \cdots, T$. This makes perfect intuitive sense: When the maintained break τ is not equal to the true break date τ_0, there is one break either prior or after τ. One way to test this is to use a Nyblom statistic for the (under the null hypothesis stable) standard regression model for $t = 1, \cdots, \tau$, and another Nyblom statistic for the (under the null hypothesis also stable) standard regression model for $t = \tau+1, \cdots, T$. Proposition 2 shows that this 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 $U_T(\tau)$ is not a feasible statistic, as Ω_1 and Ω_2 are typically unknown. But under the null hypothesis of $\tau_0 = \tau$, 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 $\{v_t\}_{t=1}^{\tau}$ and $\{v_t\}_{t=\tau+1}^{T}$—for primitive conditions see, for instance, Newey and West (1987) or Andrews (1991). Denote by $\hat{U}_T(\tau)$ the statistic $U_T(\tau)$ 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 and $\tau = \tau_0$*

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

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 $\hat{U}_T(\tau_0)$ for $k = 1, \cdots, 6$ are reproduced in Table 2.

As required, the asymptotic null distribution of $\hat{U}_T(\tau)$ does not depend on δ. For any size of break δ, the collection of values of $\tau = 1, \cdots, T$ for which the test $\hat{U}_T(\tau)$ does not exceed its asymptotic critical value of significance level $(1 - C)$ hence has asymptotic coverage C, i.e. is a valid confidence set. Note that this in particular implies that the confidence set is valid under asymptotic thought experiments where $\delta = T^{-1/2}d$ for some fixed d, in contrast to the confidence interval (2).

In detail, one proceeds as follows:

Table 2: Critical Values of $\hat{U}_T(\tau)$

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.

- For any $\tau = p + 2k + 1, \cdots, T - p - 2k - 1$, compute the least squares regression of $\{y_t\}_{t=1}^T$ on $\{X_t, \mathbf{1}[t > \tau]X_t, Z_t\}_{t=1}^T$.
- Construct $\{v_t\}_{t=1}^T = \{X_t e_t\}_{t=1}^T$, where e_t are the residuals from this regression.
- Compute the long-run variance estimators $\hat{\Omega}_1$ and $\hat{\Omega}_2$ of $\{v_t\}_{t=1}^\tau$ and $\{v_t\}_{t=\tau+1}^T$, 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 $\hat{U}_T(\tau)$ as in (7) with Ω_1 and Ω_2 replaced by $\hat{\Omega}_1$ and $\hat{\Omega}_2$, respectively.
- Include τ in the level C confidence set when $\hat{U}_T(\tau) < \text{cv}_{1-C}(k)$ and exclude it otherwise, where cv_{1-C} is the level $(1 - C)$ critical value of the statistic $\hat{U}_T(\tau)$ 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 location 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 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 result 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 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 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 $\hat{U}_T(\tau)$ 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 $\hat{U}_T(\tau)$ are extremely reliable and thus can be interpreted in the usual way.

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

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

with $T = 100$. Specifically, we consider four 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) as model M1, but with disturbances that are independent Gaussian with a variance that quadruples at the break date; (M3) X_t a Gaussian, stationary mean zero first-order autoregressive process with coefficient 0.5 and $Z_t = 1$ with i.i.d. Gaussian disturbances $\{u_t\}$ independent of $\{X_t\}$; (M4) a heteroskedastic version of M3, where the disturbances $\{u_t\}$ are given by $\{\tilde{u}_t|X_t|\}$, where $\{\tilde{u}_t\}$ 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.

We consider a version of $\hat{U}_T(\tau)$ that imposes equivalence of the long-run variances of $\{X_t u_t\}$ prior to and after the break, $\Omega_1 = \Omega_2$, denoted $\hat{U}_T(\tau)$.eq, and one that does not, denoted $\hat{U}_T(\tau)$.neq. While $\hat{U}_T(\tau)$ 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 and M2, of course, Bai.eq=Bai.het.

Tables 3 through 6 show the empirical coverage rates and average confidence set lengths for the confidence interval (2) and confidence sets constructed by inverting the test statistics $\hat{U}_T(\tau)$ 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 $[r_0 T]$, where $r_0 = 0.5$, 0.35 and 0.2. Empirical rejection probabilities are estimated based on 10,000 replications. The Tables also include the rejection probability of a 5%-level Nyblom test for the presence of a break (based on the asymptotic critical value, although unreported simulations show size control to be very good).

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 as long as the break is small. For all four 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 Model M2 with $r_0 = 0.35$ and $d = 8$ the Nyblom test rejects for half of the samples, yet confidence intervals (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 M3 with $r_0 = 0.35$ and $d = 12$ the Nyblom test rejects the null 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 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 $\hat{U}_T(\tau)$ control coverage remarkably well. For the case where both the Bai method and the

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
				$r_0=0.5$				
$\hat{U}_T(\tau)$.eq	0.949	77.7	0.949	42.4	0.949	22.1	0.949	15.1
$\hat{U}_T(\tau)$.neq	0.950	77.2	0.950	42.3	0.950	22.7	0.950	15.8
Bai.eq	0.698	53.5	0.890	32.1	0.940	16.0	0.959	9.5
Bai.het	0.698	53.5	0.890	32.1	0.940	16.0	0.959	9.5
Bai.hneq	0.686	52.0	0.882	32.1	0.938	16.0	0.956	9.5
Nyblom	0.428		0.948		1.000		1.000	
				$r_0=0.35$				
$\hat{U}_T(\tau)$.eq	0.952	79.0	0.952	44.3	0.952	22.5	0.952	15.0
$\hat{U}_T(\tau)$.neq	0.954	78.7	0.954	44.1	0.954	23.1	0.954	15.7
Bai.eq	0.692	51.5	0.878	31.8	0.937	16.1	0.962	9.5
Bai.het	0.692	51.5	0.878	31.8	0.937	16.1	0.962	9.5
Bai.hneq	0.676	49.8	0.873	31.6	0.932	16.1	0.959	9.5
Nyblom	0.366		0.902		0.999		1.000	
				$r_0=0.2$				
$\hat{U}_T(\tau)$.eq	0.949	83.2	0.949	55.7	0.949	27.1	0.949	15.3
$\hat{U}_T(\tau)$.neq	0.951	83.3	0.951	56.1	0.951	27.9	0.951	16.2
Bai.eq	0.660	45.8	0.851	30.3	0.926	16.4	0.955	9.7
Bai.het	0.660	45.8	0.851	30.3	0.926	16.4	0.955	9.7
Bai.hneq	0.631	43.4	0.832	29.2	0.914	16.0	0.947	9.5
Nyblom	0.189		0.617		0.939		0.997	

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
				$r_0 = 0.5$				
$\hat{U}_T(\tau)$.eq	0.936	85.1	0.936	68.8	0.936	46.0	0.936	29.1
$\hat{U}_T(\tau)$.neq	0.950	85.4	0.950	67.5	0.950	44.6	0.950	28.6
Bai.eq	0.572	53.2	0.735	48.5	0.846	33.9	0.894	21.5
Bai.het	0.572	53.2	0.735	48.5	0.846	33.9	0.894	21.5
Bai.hneq	0.614	52.5	0.762	46.5	0.869	33.8	0.918	22.0
Nyblom	0.204		0.613		0.922		0.996	
				$r_0 = 0.35$				
$\hat{U}_T(\tau)$.eq	0.963	87.5	0.963	74.5	0.963	53.6	0.963	34.9
$\hat{U}_T(\tau)$.neq	0.954	86.9	0.954	71.4	0.954	48.6	0.954	30.7
Bai.eq	0.562	59.1	0.735	54.7	0.856	39.8	0.906	25.6
Bai.het	0.562	59.1	0.735	54.7	0.856	39.8	0.906	25.6
Bai.hneq	0.584	52.7	0.747	44.9	0.866	33.5	0.916	22.5
Nyblom	0.135		0.469		0.834		0.983	
				$r_0 = 0.2$				
$\hat{U}_T(\tau)$.eq	0.978	90.2	0.978	83.2	0.978	69.0	0.978	49.9
$\hat{U}_T(\tau)$.neq	0.951	89.3	0.951	80.6	0.951	64.4	0.951	44.4
Bai.eq	0.550	61.4	0.694	54.7	0.829	42.4	0.904	29.6
Bai.het	0.550	61.4	0.694	54.7	0.829	42.4	0.904	29.6
Bai.hneq	0.552	53.6	0.681	42.5	0.814	31.1	0.897	22.3
Nyblom	0.071		0.196		0.442		0.717	

Table 5: Empirical Small Sample Coverage and Length of Confidence Sets

Model M3: Stochastic regressor, i.i.d. disturbances.

	$d=4$		$d=8$		$d=12$		$d=16$	
	Cov.	Lgth	Cov.	Lgth	Cov.	Lgth	Cov.	Lgth
				$r_0=0.5$				
$\hat{U}_T(\tau)$.eq	0.954	79.7	0.954	51.7	0.954	31.0	0.954	21.9
$\hat{U}_T(\tau)$.neq	0.955	79.1	0.955	51.2	0.955	32.0	0.955	23.5
Bai.eq	0.699	53.7	0.856	32.7	0.899	16.5	0.902	9.7
Bai.het	0.682	50.8	0.842	30.9	0.889	15.8	0.893	9.4
Bai.hneq	0.647	48.6	0.819	31.2	0.873	16.3	0.886	9.7
Nyblom	0.373		0.882		0.994		1.000	
				$r_0=0.35$				
$\hat{U}_T(\tau)$.eq	0.953	80.5	0.953	54.0	0.953	32.1	0.953	22.2
$\hat{U}_T(\tau)$.neq	0.954	80.2	0.954	53.9	0.954	33.3	0.954	23.9
Bai.eq	0.693	52.1	0.856	32.3	0.896	16.6	0.903	9.8
Bai.het	0.671	49.3	0.841	30.6	0.885	15.9	0.896	9.5
Bai.hneq	0.639	46.7	0.820	30.6	0.866	16.3	0.880	9.7
Nyblom	0.313		0.803		0.978		0.998	
				$r_0=0.2$				
$\hat{U}_T(\tau)$.eq	0.954	83.3	0.954	63.6	0.954	41.4	0.954	27.3
$\hat{U}_T(\tau)$.neq	0.958	83.8	0.958	65.1	0.958	44.0	0.958	30.4
Bai.eq	0.666	47.4	0.819	30.8	0.881	17.0	0.900	10.0
Bai.het	0.650	45.1	0.803	29.2	0.870	16.2	0.893	9.7
Bai.hneq	0.601	41.6	0.760	27.8	0.832	16.2	0.865	9.8
Nyblom	0.169		0.505		0.782		0.914	

Table 6: Empirical Small Sample Coverage and Length of Confidence Sets
Model M4: Stochastic regressor, heterskedastic disturbances.

	$d=4$		$d=8$		$d=12$		$d=16$	
	Cov.	Lgth	Cov.	Lgth	Cov.	Lgth	Cov.	Lgth
$r_0 = 0.5$								
$\hat{U}_T(\tau)$.eq	0.959	78.6	0.959	47.5	0.959	27.7	0.959	20.0
$\hat{U}_T(\tau)$.neq	0.964	77.7	0.964	46.5	0.964	28.5	0.964	21.4
Bai.eq	0.547	26.4	0.745	11.7	0.857	5.9	0.921	3.8
Bai.het	0.742	52.9	0.879	28.0	0.938	14.0	0.969	8.4
Bai.hneq	0.674	47.1	0.849	26.6	0.923	13.6	0.958	8.1
Nyblom	0.413		0.922		0.996		1.000	
$r_0 = 0.35$								
$\hat{U}_T(\tau)$.eq	0.959	79.5	0.959	49.6	0.959	28.9	0.959	20.5
$\hat{U}_T(\tau)$.neq	0.964	78.8	0.964	48.6	0.964	29.3	0.964	21.7
Bai.eq	0.544	26.2	0.742	11.8	0.848	6.0	0.922	3.8
Bai.het	0.736	51.7	0.878	28.0	0.939	14.1	0.970	8.4
Bai.hneq	0.665	45.0	0.843	26.0	0.916	13.4	0.958	8.1
Nyblom	0.349		0.857		0.987		0.999	
$r_0 = 0.2$								
$\hat{U}_T(\tau)$.eq	0.956	82.7	0.956	59.9	0.956	36.9	0.956	24.4
$\hat{U}_T(\tau)$.neq	0.965	82.9	0.965	59.9	0.965	37.9	0.965	26.1
Bai.eq	0.515	26.6	0.712	12.5	0.844	6.2	0.914	3.9
Bai.het	0.716	49.1	0.852	27.8	0.930	14.5	0.965	8.6
Bai.hneq	0.629	40.4	0.795	23.5	0.900	12.8	0.947	7.7
Nyblom	0.191		0.556		0.825		0.934	

method suggested here result in confidence sets of correct coverage, however, it is seen that the Bai method delivers the smaller set.

When the break in the regression coefficient is accompanied by a break in the long-run 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 M4, the Bai method that fails to account 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 M4 compared to the homoskedastic model M3. This is less surprising than it seems: The normalization of the variance of $\{u_t\}$—in order to ensure a long-run variance of $\{X_t u_t\} = \{|X_t| X_t \tilde{u}_t\}$ equal to unity—makes the disturbance variance of model M4 smaller than in model M3.

Overall, the small sample experiments are highly encouraging for constructing reliable confidence sets for the break date by inverting a sequence of tests based on $\hat{U}_T(\tau)$. 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 location 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.

As a remedy, we suggest 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 to invert 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.

6 Appendix

Proof of Proposition 1:

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

$$\begin{aligned} S(\tau) = & \sum_{t=1}^{T} \nu_t^2 - \left(\sum_{t=1}^{\tau} X_t \nu_t\right)' \left(\sum_{t=1}^{\tau} X_t X_t'\right)^{-1} \sum_{t=1}^{\tau} X_t \nu_t \\ & - \left(\sum_{t=\tau+1}^{T} X_t \nu_t\right)' \left(\sum_{t=\tau+1}^{T} X_t X_t'\right)^{-1} \sum_{t=\tau+1}^{T} X_t \nu_t - \left(\sum_{t=1}^{T} \tilde{Z}_t \nu_t\right)' \left(\sum_{t=1}^{T} \tilde{Z}_t \tilde{Z}_t'\right)^{-1} \sum_{t=1}^{T} \tilde{Z}_t \nu_t \end{aligned}$$

For $t < \tau = [lT]$, $\tilde{Z}_t = Z_t - \left(\sum_{s=1}^{\tau} Z_s X_s'\right)\left(\sum_{s=1}^{\tau} X_s X_s'\right)^{-1} X_t$, and similarly, for $t > \tau$, $\tilde{Z}_t = Z_t - \left(\sum_{s=\tau+1}^{T} Z_s X_s'\right)\left(\sum_{s=\tau+1}^{T} X_s X_s'\right)^{-1} X_t$. From the uniform convergence of $T^{-1}\sum_{t=1}^{[sT]} X_t Z_t'$ and $T^{-1}\sum_{t=1}^{[sT]} X_t X_t'$ in s and $\sup_{0<s<1} |T^{-1/2}\sum_{t=1}^{[sT]} X_t \nu_t| = O_p(1)$ and $\sup_{0<s<1} |T^{-1/2}\sum_{t=1}^{[sT]} Z_t \nu_t| = O_p(1)$ we find

$$\sup_{\tau \in (\bar{\lambda}T, (1-\bar{\lambda})T)} \left| \left(\sum_{t=1}^{T} \tilde{Z}_t \nu_t\right)' \left(\sum_{t=1}^{T} \tilde{Z}_t \tilde{Z}_t'\right)^{-1} \sum_{t=1}^{T} \tilde{Z}_t \nu_t - \left(\sum_{t=1}^{T} \check{Z}_t \nu_t\right)' \left(\sum_{t=1}^{T} \check{Z}_t \check{Z}_t'\right)^{-1} \sum_{t=1}^{T} \check{Z}_t \nu_t \right| = o_p(1)$$

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

$$\begin{aligned} \arg \min_{\tau \in (\bar{\lambda}T\tau \leq (1-\bar{\lambda})} S(\tau) &= \arg \min_{\bar{\lambda} \leq l \leq 1-\bar{\lambda}} S([lT]) \\ &= \arg \max_{\bar{\lambda} \leq l \leq 1-\bar{\lambda}} \left(\sum_{t=1}^{\tau} X_t \nu_t\right)' \left(\sum_{t=1}^{\tau} X_t X_t'\right)^{-1} \sum_{t=1}^{\tau} X_t \nu_t \\ &\quad + \left(\sum_{t=\tau+1}^{T} X_t \nu_t\right)' \left(\sum_{t=\tau+1}^{T} X_t X_t'\right)^{-1} \sum_{t=\tau+1}^{T} X_t \nu_t + R_T(l) \\ &\Rightarrow \arg \max_{\bar{\lambda} \leq s \leq 1-\bar{\lambda}} G(s) \end{aligned}$$

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

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

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

Now $\tilde{X}_t = X_t - (\sum_{t=\tau}^{T} X_t Q_t')(\sum_{t=1}^{T} Q_t Q_t')^{-1} Q_t$. With $T^{-1}\sum_{t=\tau}^{T} Q_t Q_t' \xrightarrow{p} l\Sigma_Q$ uniformly in l from Condition 1,

$$\begin{aligned} T^{-1}\sum_{t=1}^{T} \tilde{X}_t \tilde{X}_t' &= T^{-1}\sum_{t=\tau+1}^{T} \tilde{X}_t X_t' \\ &= T^{-1}\sum_{t=\tau+1}^{T} X_t X_t' - T^{-1}(\sum_{t=\tau+1}^{T} X_t Q_t')(\sum_{t=1}^{T} Q_t Q_t')^{-1}(\sum_{t=\tau+1}^{T} Q_t X_t') \\ &\xrightarrow{p} l(1-l)\Sigma_X \end{aligned}$$

and also

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

uniformly in l, since $T^{-1/2}\sum_{t=1}^{T} Z_t v_t = O_p(1)$. The application of the CMT hence yields the result for $\hat{\delta}_a$.

Proof of Proposition 2:

Let B_Q be the matrix $T \times (T-2k-p)$ matrix that satisfies $B_Q' B_Q = I_{T-2k-p}$ and $B_Q B_Q' = M_Q$, where M_Q is the projection matrix off the column space spanned by $\{X_t, \mathbf{1}[t > \tau]X_t, Z_t\}$. Let $y = (y_1, \cdots, y_T)'$ and $Q = (Q_1, \cdots, Q_T)'$. Then $(B_Q y, Q)$ is a maximal invariant to the group of transformations (4). By the Neyman-Pearson Lemma, Fubini's Theorem and the likelihood structure in Condition 2, an efficient invariant test for $b^2 > 0$ of (3) maximizing (5) can hence be based on (see Elliott and Müller (2003) for development)

$$LR_T = \sum_{t=1,\cdots,\tau-1,\tau+1,\cdots,T} w_t \int \exp\left[\sigma^{-2} y' M_Q \Xi(t) d - \tfrac{1}{2}\sigma^{-2} d'\Xi(t)' M_Q \Xi(t) d\right] d\nu_t(d)$$

where $\Xi(t)$ is a $T \times k$ matrix with rows X_s' when $s > t$ and a $1 \times k$ zero row vector otherwise.

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

$$
\begin{aligned}
F(t) &= \int \exp\left[\sigma^{-2} y' M_Q \Xi(t) d - \tfrac{1}{2}\sigma^{-2} d'\Xi(t)' M_Q \Xi(t) d\right] d\nu_t(d) \\
&= \int (2\pi)^{-k/2} |b^2\tau^{-2}\Omega_1^{-1}|^{-1/2} \exp\left[\sigma^{-2} d' \sum_{s=t}^{T} X_s e_s - \tfrac{1}{2}\sigma^{-2} d'\Xi(t)' M_Q \Xi(t) d - \tfrac{1}{2} b^{-2}\tau^2 d'\Omega_1 d\right] dd \\
&= |b^2\tau^{-2}\Omega_1^{-1}|^{-1/2} |b^{-2}\tau^2\Omega_1 + \sigma^{-2}\Xi(t)' M_Q \Xi(t)|^{-1/2} \\
&\quad \times \exp\left[\tfrac{1}{2}\sigma^{-4} \left(\sum_{s=t}^{T} X_s e_s\right)' \left(b^{-2}\tau^2\Omega_1 + \sigma^{-2}\Xi(t)' M_Q \Xi(t)\right)^{-1} \left(\sum_{s=t}^{T} X_s e_s\right)\right] \\
&= |I_k + b^2\sigma^{-2}\tau^{-2}\Omega_1^{-1}\Xi(t)' M_Q \Xi(t)|^{-1/2} \\
&\quad \times \exp\left[\tfrac{1}{2}\sigma^{-4} b^2 \left(\sum_{s=t}^{\tau} X_s e_s\right)' \left(\tau^2\Omega_1 + \sigma^{-2} b^2 \Xi(t)' M_Q \Xi(t)\right)^{-1} \left(\sum_{s=t}^{\tau} X_s e_s\right)\right]
\end{aligned}
$$

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

$$
\begin{aligned}
2(F(t) - 1) = \sigma^{-4} b^2 \tau^{-2} \left(\sum_{s=t}^{T} X_s e_s\right)' \Omega_1^{-1} \left(\sum_{s=t}^{T} X_s e_s\right) \\
- b^2\sigma^{-2}\tau^{-2} \operatorname{tr} \Omega_1^{-1}\Xi(t)' M_Q \Xi(t) + o(b^2).
\end{aligned}
$$

Proceeding analogously for $t > \tau$ and collecting terms whose distribution depends on δ and τ_0 yields the result.

Proof of Proposition 3:

Proceed similarly to the proof of Proposition 1 to show that under Condition 1, for $s \leq r_0$

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

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.

——— (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856.

——— (2003): "End-of-Sample Instability Tests," *Econometrica*, 71, 1661–1694.

ANDREWS, D., AND J. MONAHAN (1992): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60, 953–966.

ANDREWS, D., AND W. PLOBERGER (1994): "Optimal Tests When a Nuisance Parameter Is Present Only under the Alternative," *Econometrica*, 62, 1383–1414.

BAI, J. (1994): "Least Squares Estimation of a Shift in Linear Processes," *Journal of Time Series Analysis*, 15, 453–470.

——— (1997a): "Estimating Multiple Breaks One at a Time," *Econometric Theory*, 13, 315–352.

——— (1997b): "Estimation of a Change Point in Multiple Regressions," *Review of Economics and Statistics*, 79, 551–563.

——— (1999): "Likelihood Ratio Tests for Multiple Structural Changes," *Journal of Econometrics*, 91, 299–323.

BAI, J., R. LUMSDAINE, AND J. STOCK (1998): "Testing for and Dating Common Breaks in Multivariate Time Series," *Review of Economic Studies*, 65, 395–432.

BAI, J., AND P. PERRON (1998): "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66, 47–78.

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.

——— (1995): "Some Aspects of Change-Point Analysis," in *Change-Point Analysis*, pp. 28–56. Institute of Mathematical Statistics, Lecture Notes, Monograph Series Vol. 9, Hayward, CA.

Boivin, J. (2003): "Has U.S. Monetary Policy Changed? Evidence from Drifting Coefficients and Real Time Data," *mimeo, Columbia University.*

Busetti, F., and A. Harvey (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., J. Gali, and M. Gertler (2000): "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics*, 115, 147–180.

Elliott, G., and U. Müller (2003): "Optimally Testing General Breaking Processes in Linear Time Series Models," *UCSD Working Paper 2003-07.*

Engle, R., D. Hendry, and J.-F. Richard (1983): "Exogeneity," *Econometrica*, 51, 277–304.

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., and D. Pollard (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.

Lettau, M., and S. Ludvigson (2001): "Consumption, Aggregate Wealth and Expected Stock Returns," *Journal of Finance*, 56, 815–849.

Linde, J. (2001): "Testing for the Lucas-Critique: A Quantitative Investigation," *American Economic Review*, 91, 986–1005.

MacNeill, I. (1978): "Properties of Sequences of Partial Sums of Polynomial Regression Residuals with Applications to Test for Change of Regression at Unknown Times," *Annals of Statistics*, 6, 422–433.

Nabeya, S., and K. Tanaka (1988): "Asymptotic Theory of a Test for Constancy of Regression Coefficients Against the Random Walk Alternative," *Annals of Statistics*, 16, 218–235.

Newey, W., and K. West (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., and A. Timmermann (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. (1994): "Unit Roots, Structural Breaks and Trends," in *Handbook of Econometrics*, ed. by R. Engle, and D. McFadden, vol. 4, pp. 2740–2841. North Holland, New York.

Stock, J., and M. Watson (1996): "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business and Economic Statistics*, 14, 11–30.

Timmermann, A., and B. Paye (2004): "Instability of Return Prediction Models," *mimeo, UCSD*.

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.

——— (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 *Recent Advances in Statistics*, ed. by M. R. et Al., pp. 245–269. Academic Press, New York.