JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **21**: 175–190 (2006)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jae.821

TIMING STRUCTURAL CHANGE: A CONDITIONAL PROBABILISTIC APPROACH

DAVID N. DEJONG,[a] ROMAN LIESENFELD[b] AND JEAN-FRANCOIS RICHARD[a]*

[a] *Department of Economics, University of Pittsburgh, USA*
[b] *Department of Economics, Universität Kiel, Germany*

SUMMARY

We propose a strategy for assessing structural stability in time-series frameworks when potential change dates are unknown. Existing stability tests are effective in detecting structural change, but procedures for identifying timing are imprecise, especially in assessing the stability of variance parameters. We present a likelihood-based procedure for assigning conditional probabilities to the occurrence of structural breaks at alternative dates. The procedure is effective in improving the precision with which inferences regarding timing can be made. We illustrate parametric and non-parametric implementations of the procedure through Monte Carlo experiments, and an assessment of the volatility reduction in the growth rate of US GDP.

1. INTRODUCTION

Following the pioneering work of Andrews (1993), implementation of tests for structural stability in time-series frameworks for cases in which the timing of potential breaks is unknown has become straightforward. Hansen (2001) provides an overview of the development of such tests, and several applications. But while stability tests have proven to be effective in detecting the presence of structural change, existing procedures for identifying timing are highly imprecise, especially when applied to breaks in higher-order moments like the variance. For example, in their analysis of the structural stability of growth in US, GDP, Stock and Watson (2002) report an asymptotic 67% confidence interval for their estimate of the timing of the occurrence of a reduction in the innovation variance of this series. This interval spans 12 quarters: 1982:IV–1985:III. Stock and Watson depart from the conventional use of 95% confidence intervals because, as they note, '... 95% intervals ... are so wide as to be uninformative' (p. 12, fn 4).

Here, we present a likelihood-based procedure designed to pinpoint the timing with which suspected structural changes are most likely to have occurred. Given the occurrence of n regime changes, where n is identified *a priori*, the procedure identifies n break dates as those receiving highest conditional (given the observed data) probabilities relative to the full range of their potential alternatives. The effectiveness of classical tests in detecting the presence of structural instability makes them an attractive tool for identifying the occurrence of regime changes *a priori*, but alternative approaches are also applicable (including inter-ocular trauma).

The procedure is related in spirit to a classification analysis, under which the goal is to classify each element of a set of observations into one of several potential categories (see Anderson, 1984, ch. 6 for an overview and references). Given a model specification with one regime change,

* Correspondence to: Jean-Francois Richard, Department of Economics, University of Pittsburgh, Pittsburgh, PA 15260, USA. E-mail: fantin@pitt.edu

Received 15 May 2003
Revised 23 June 2004

associated observations will correspond with one of two likelihood functions: one prevailing prior to the break date and one after the break date. The procedure identifies the break date as the optimal point at which to divide the sample. Given multiple regime changes, additional issues of implementation (e.g., timing the breaks jointly or one at a time) arise. The full resolution of these issues is beyond the scope of this paper, in which we focus exclusively on the case of a single regime change.

Given one regime change, the specific implementation of the procedure is as follows. First, a set of data points observed at the beginning of the sample (say, the first 15% of the total observations) is used to estimate the likelihood function that prevails under the initial structural regime. Likewise, a set of data points observed at the end of the sample is used to estimate the likelihood function that prevails under the second regime. Each date lying between these beginning- and end-of-sample 'reference subsets' represents a potential break date; the final step is to calculate probabilities associated with these mid-sample dates. To calculate the probability assigned to date j, the subset of mid-sample observations is divided into two sets: one that predates j and one that postdates j. The likelihood of a regime change at date j is then calculated by assigning the former set to the beginning-of-sample likelihood function, and the latter set to the end-of-sample likelihood function. The probability of a regime change at date j is obtained by comparing its relative likelihood with those of all alternative dates. Parametric, semi-parametric or non-parametric approaches can be used to approximate the likelihood functions that prevail across regimes. Also, there are alternative approaches one can use for selecting reference subsets from which likelihood estimates of parameters are obtained. We consider using fixed and sequentially expanding subsets here, and obtain slightly superior performance using a particular implementation of the latter approach; details are provided below.

Given that break date probability calculations are made conditionally on the observed data, our procedure has a Bayesian flavour. However, unlike a fully Bayesian approach to the problem of identifying potential structural change, our procedure does not require the specification of a prior distribution for the parameters of the underlying model specification, nor the specification of a distributional assumption regarding fundamental innovations. It merely involves a modest set of probability calculations in support of a simple model-diagnostic exercise. (Applications of fully Bayesian procedures for evaluating structural stability in time-series frameworks abound. For example, see Wang and Zivot, 2000 for a recent application to a switching-regression framework that allows for multiple breaks in the mean and innovation specifications of autoregressive processes.)

The remainder of the paper is organized as follows. We begin by illustrating problems suffered by standard procedures in pinpointing potential break dates. We do so through an empirical application and a set of Monte Carlo experiments. The empirical application is an analysis of the volatility reduction undergone by the growth rate of US GDP in the mid-1980s. This volatility reduction has been analysed extensively through the use of structural-stability tests (e.g., see Kim and Nelson, 1999; McConnell and Perez-Quiros, 2000; Stock and Watson, 2002). In the Monte Carlo experiments, we generate artificial time series featuring a volatility reduction designed to mimic that undergone by the growth rate of US GDP, and estimate break dates using Andrews' (1993) sup-Wald test statistic and Bai's (1997) least squares procedure. Next, we outline the details of our procedure. We then demonstrate its performance under a parametric implementation by repeating the empirical application and Monte Carlo experiments described above. Finally, we demonstrate its performance under a non-parametric implementation by again repeating the empirical application.

The relative precision with which our procedure identifies break dates in the innovation variance of an autoregression is striking. For example, Bai's (1997) least squares procedure based on auxiliary regressions for squared residuals generates an interval between the 16.5% and 83.5% quantile of the sampling distribution of 16 quarters under our leading experimental design, while our procedure generates an interval of 6 quarters under the parametric implementation we consider, and 7 quarters under the non-parametric implementation. Thus our procedure serves as an attractive complement to existing procedures for structural instability: existing tests are effective in identifying the presence of breaks; and our procedure is effective in identifying the timing with which they occurred.

2. EXISTING PROCEDURES FOR ESTIMATING BREAK DATES: AN ILLUSTRATION

As noted, considerable attention has been given to the issue of whether the volatility reduction undergone by the growth rate of US GDP represents a structural break. The top panel of Figure 1 illustrates the post-war behaviour of this series, and summary statistics are provided in Table I. (The series $\{g_t, t : 1 \to T\}$ is computed as logged differences in quarterly GDP measured in chain-weighted 1996 prices, annualized by multiplying by 400, spanning 1947:II through 2002:III.) Prior to 1984:I (the break date identified using the parametric implementation of the probabilistic procedure discussed below), the standard deviation of this series was 4.73, in comparison with a measure of 4.05 computed over the full sample, and 2.13 after 1984:I.

To analyse whether this behaviour constitutes a structural break in the volatility of GDP, we assume an AR(p) specification for g_t, i.e.,

$$g_t = \mu + \phi_1 g_{t-1} + \cdots + \phi_p g_{t-p} + \varepsilon_t, \qquad \mathrm{var}(\varepsilon_t) = \sigma^2 \tag{1}$$

and test for changes of unknown timing in the innovation variance σ^2. (We use an AR(1) specification in our applications, as additional lags are statistically insignificant. Use of additional

Table I. Summary statistics

Real GDP growth	Mean	Std. dev.
Full sample	3.35	4.05
1947:II–1984:I	3.50	4.73
1984:II–2002:III	3.04	2.13

Estimates of $g_t = \mu + \phi_1 g_{t-1} + \varepsilon_t$, $\mathrm{E}(\varepsilon_t^2) = \sigma^2$ without structural breaks, and sup-Wald statistics for a structural break

Parameter	Estimate	Std. error	sup-Wald stat.	Asymptotic p-value	Break date
μ	2.20	0.39	1.99	0.79	none
ϕ_1	0.34	0.07	3.18	0.52	none
σ	3.82		33.70	0.00	1984:I

Least squares estimate of the break date of σ^2

Estimated date	67% conf. interval	95% conf. interval
1983:II	1979:IV–1986:IV (29)	1971:I–1995:III (99)

Note: The numbers of quarters spanned by confidence intervals are given in parentheses.

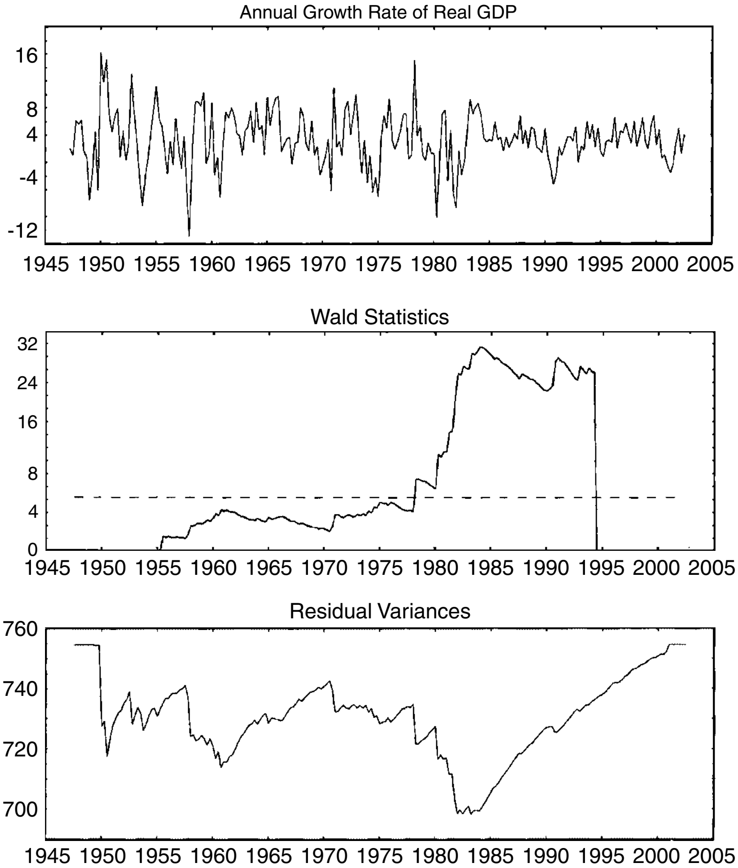

Figure 1. GDP growth measured as logged differences in quarterly GDP, measured in chain-weighted 1996 prices, annualized by multiplying by 400, spanning 1947:I through 2002:II (upper panel); heteroskedasticity-consistent Wald statistic sequence for testing for a structural break in the innovation variance of the AR(1) fitted to the GDP growth rate (solid line, middle panel) and the corresponding asymptotic 5% critical value (dashed line, middle panel); residual variance as a function of potential break dates (bottom panel)

lags yields similar results.) To test for changes in the parameters of specification (1) at unknown dates we use a testing procedure based on the Wald form of the Quandt (1960) statistic with a heteroskedasticity-consistent covariance matrix, following the strategy employed, e.g., by McConnell and Perez-Quiros (2000) and Stock and Watson (2002). This procedure is based on the largest Wald statistic (sup-Wald statistic) for a structural break test over all potential break dates j between dates T_1 and T_2. (Following McConnell and Perez-Quiros, 2000 and Stock and Watson, 2002, we use $T_1 = [0.15 \times T]$ and $T_2 = [0.85 \times T]$.) Asymptotic critical values for the sup-Wald statistic are provided by Andrews (1993), and asymptotic p-values by Hansen (1997).

We use this procedure to test for a break in each of the parameters in (1) separately. For a change in σ^2, the test is based on the sup-Wald statistic for a break in the mean of the squared residuals from the estimated autoregression. In particular, letting e_t denote the residuals associated with OLS estimates of (1), the sup-Wald statistic for a break in σ^2 is $\sup_{\{T_1 \leq j \leq T_2\}}[W(j)]$, where the heteroskedasticity-consistent Wald statistic is given by

$$W(j) = \frac{\frac{j(T-j)}{T}(\hat{\sigma}_2^2 - \hat{\sigma}_1^2)^2}{\left(\frac{T-j}{jT}\right)\sum_{t=1}^{j}(e_t^2 - \hat{\sigma}_1^2)^2 + \left(\frac{j}{(T-j)T}\right)\sum_{t=j+1}^{T}(e_t^2 - \hat{\sigma}_2^2)^2} \tag{2}$$

with $\hat{\sigma}_1^2 = \left(\sum_{t=1}^{j} e_t^2\right)/j$ and $\hat{\sigma}_2^2 = \left(\sum_{t=j+1}^{T} e_t^2\right)/(T-j)$. Consistent with the findings of McConnell and Perez-Quiros (2000), this test yields no evidence of a break in the AR parameters μ and ϕ_1 estimated for g_t, but sharp evidence of a change in σ^2: the asymptotic p-values obtained for μ and ϕ_1 are 0.79 and 0.52, while the p-value obtained for σ^2 is zero (see Table I).

A natural candidate for an estimate of the date of the occurrence of a break in σ^2 often used in the literature is the date j that yields the largest Wald statistic $W(j)$, i.e., $j_{sw}^* = \arg\sup_{\{T_1 \leq j \leq T_2\}}[W(j)]$. For the GDP series we analyse, the estimate is 1984:I. However, note from the second panel of Figure 1 that there is considerable uncertainty associated with the identification of this specific date. In particular, Wald statistics become significant at the 5% level as early as 1978, and the sequence of Wald statistics is fairly flat between 1982 and 1995. Furthermore, it is well known that break date estimates based upon sup-Wald type statistics are good estimates only when the homoskedastic form of the Wald statistic is used (e.g., see Hansen, 2001).

An alternative estimator for the timing of the break is provided by Bai's (1997) least squares (LS) procedure. Applied to the break date for σ^2 (under the assumption that the remaining AR parameters are constant), the procedure involves splitting squared residuals obtained from (1) at each possible break date, regressing them for both subsamples separately on a constant, and storing the resulting sum of squared residuals of this auxiliary regression. The estimated break date is that which minimizes the full-sample residual variance of the auxiliary regressions. Hence, the LS estimator for the break date of σ^2 has the following form: $j_{ls}^* = \arg\min_{\{T_1 \leq j \leq T_2\}}[S(j)]$, where

$$S(j) = \sum_{t=1}^{j}(e_t^2 - \hat{\sigma}_1^2)^2 + \sum_{t=j+1}^{T}(e_t^2 - \hat{\sigma}_2^2)^2 \tag{3}$$

Bai (1997) derives the asymptotic distribution of this estimator, and shows how it can be used to construct associated confidence intervals. It can be shown that the residual variance $S(j)$ is a monotonic transformation of the corresponding homoskedastic form of the Wald statistic for a break in j (e.g., see Bai, 1997). Thus the date that minimizes $S(j)$ is equal to the date that maximizes the homoskedastic form of the Wald statistic.

The sequence of residual variances $\{S(j), j : T_1 \rightarrow T_2\}$ obtained for GDP growth is illustrated in the bottom panel of Figure 1. This sequence attains its minimum at 1983:II; the associated 67% asymptotic confidence interval of this estimate is 1979:IV–1986:IV, a rather wide interval spanning seven years.

To illustrate some small-sample characteristics of these break date estimators, we conducted a series of six Monte Carlo experiments under which we repeatedly applied the sup-Wald and LS procedures to artificial data designed to mimic the behaviour of US GDP. Specifically, we fed artificial realizations of $\{\varepsilon_t\}$ into (1) (parameterized using the estimates reported in Table I) to produce artificial realizations of $\{g_t\}$ of length 222 (the number of observations of the actual series).

Under Experiment 1, for each realization, the first 147 values of $\{\varepsilon_t\}$ were drawn from a normal distribution with standard deviation $\sigma_1 = 4.46$ (the estimated standard deviation of AR(1) residuals of the subsample before the estimated break date 1984:I); likewise, the standard deviation used to obtain the remaining 75 values was $\sigma_1 = 2.00$ (the estimated standard deviation after the estimated break date 1984:I). We generated 10 000 artificial samples of $\{g_t\}$ in this manner, and applied the break date estimation procedures to each realization. Results of Experiment 1 are reported in the top set of entries in Table II and Figure 2.

The top panel of Figure 2 presents a histogram of the 10 000 break date estimates obtained using the sup-Wald procedure. (The rate at which these peak values exceeded their associated 5% asymptotic critical value, indicating a rejection of the null hypothesis of no break, exceeded 99.5% in each of the six sets of experiments we conducted, indicating outstanding power in detecting the presence of a break.) The bottom panel is analogous for the LS break date estimates. Both histograms have distinct peaks at the actual break date of 1984:I, but have thick tails and are heavily skewed: the histogram of the sup-Wald based estimates to the right, and that of the LS estimates to the left. Corresponding 67% quantile intervals (between the 16.5% and the 83.5% quantile) and 95% quantile intervals (between the 2.5% and the 97.5% quantile) are, according to Table II [1983:IV, 1990:I] and [1982:III, 1994:I] for the estimate based on the sup-Wald statistic and [1980:II, 1984:I] and [1964:IV, 1984:II] for the LS procedure. Thus, despite the dramatic

Table II. Monte Carlo experiments with existing break date estimators

	Mean	Std. dev.	67% interval	95% interval
	Experiment 1: true break date 84:I with $\sigma_1 = 4.46, \sigma_2 = 2.00$			
sup-Wald procedure	86.06	13.23	83:IV–90:I(26)	82:III–94:I(47)
LS procedure	81.60	20.44	80:II–84:I(16)	64:IV–84:II(79)
	Experiment 2: true break date 75:I with $\sigma_1 = 4.46, \sigma_2 = 2.00$			
sup-Wald procedure	78.38	23.21	74:IV–84:II(39)	73:III–93:IV(82)
LS procedure	73.00	15.36	71:IV–75:I(14)	60:IV–75:II(59)
	Experiment 3: true break date 66:I with $\sigma_1 = 4.46, \sigma_2 = 2.00$			
sup-Wald procedure	70.67	34.53	65:III–79:I(55)	64:III–93:IV(118)
LS procedure	64.17	12.14	62:IV–66:I(14)	54:I–66:II(50)
	Experiment 4: true break date 84:I with $\sigma_1 = 2.00, \sigma_2 = 4.46$			
sup-Wald procedure	79.07	35.68	69:III–84:II(60)	56:I–85:IV(120)
LS procedure	85.84	12.33	84:I–87:II(14)	83:III–96:II(52)
	Experiment 5: true break date 75:I with $\sigma_1 = 2.00, \sigma_2 = 4.46$			
sup-Wald procedure	71.51	23.68	65:III–75:II(40)	56:I–76:III(83)
LS procedure	77.04	15.85	75:I–78:III(15)	74:IV–90:II(63)
	Experiment 6: true break date 66:I with $\sigma_1 = 2.00, \sigma_2 = 4.46$			
sup-Wald procedure	63.84	13.59	59:IV–66:II(27)	55:IV–67:III(48)
LS procedure	68.44	20.72	66:I–69:IV(16)	65:IV–85:II(79)

Note: The 67% interval is the interval between the 16.5% and the 83.5% quantile, and the 95% interval is the interval between the 2.5% and the 97.5% quantile. The numbers of quarters spanned by the intervals are given in parentheses. Means and standard deviations are given in decimal notation, where 0.00, 0.25, 0.50 and 0.75 represent quarters I, II, III and IV.

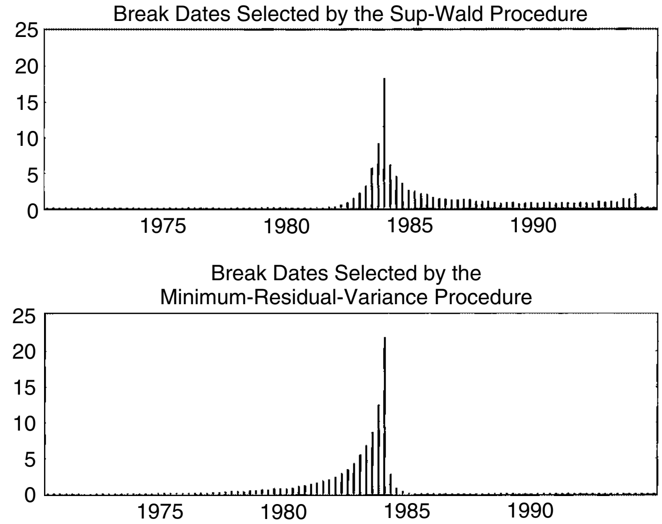

Figure 2. Histogram of break date estimates obtained using the sup-Wald procedure (upper panel) and least squares procedure (lower panel) for 10 000 artificial realizations of $\{g_t\}$. The true break point is at 1984:I, and the standard deviation of the AR(1) innovations is 4.46 before and 2.00 after the break

reduction in innovation variance σ^2 built into our experimental design, the dispersion of the small-sample distributions of the break date estimators we obtain is enormous. Also note that the sup-Wald procedure has a small positive bias while the LS estimator is slightly downward biased.

We conducted five additional experiments (Experiments 2–6) in order to explore two potential sources of the skewness exhibited by the distributions of these statistics. The two sources are the location of the break and the temporal ordering of the high- and low-variance regions. To explore the potential influence of the location of the break, we preserved the temporal ordering of Experiment 1 (an initial standard deviation of 4.46, followed by 2.00), and considered two alternative break dates: one at date 1975:I (the mid-point of the sample) under Experiment 2; the other at date 1966:I (the mirror image of the 1984:I date) under Experiment 3. Resulting intervals (reported in Table II) indicate the same pattern of skewness obtained under Experiment 1: the sup-Wald distributions remain skewed to the right, and the LS distributions remain skewed to the left. The right-skewness of the sup-Wald distributions is illustrated by the close proximity of the lower bounds of its intervals with the mean break dates, and the extreme distance of the upper bounds. For example, lower bounds of the 67% quantile intervals are within 16 and 18 quarters of the mean break dates in Experiments 2 and 3, while upper bounds differ by 24 and 34 quarters. The left-skewness of the LS distributions is reflected by the mirror image of this relationship. Thus break location is not an important factor in determining the skewness of these distributions.

To explore the potential influence of the temporal ordering of the high- and low-variance regions, we conducted three additional experiments (Experiments 4–6, summarized in Table II) featuring the three break dates considered above, but with a reversal of break orderings from low- to high-variance regimes. Here the pattern of skewness is exactly reversed: the sup-Wald distributions are skewed to the left in this case, and the LS distributions to the right. This is illustrated once again by comparing the relative proximity of upper and lower bounds of the quantile intervals with mean break dates. In addition, Figure 3 illustrates both distributions obtained under Experiment 4 (featuring a break date at 1984:I). Thus it appears that the skewness of these distributions depends on the temporal ordering of high- and low-variance regions: sup-Wald distributions are skewed towards low-variance regions, and LS distributions towards high-variance regions. And both distributions feature enormous dispersions, making them ineffective in pinpointing break dates in the variance of AR-innovations.

The reason for the skewness of the LS distributions towards the high-variance region is the fact that the probability of obtaining small e_t^2 realizations under a distribution with high variance σ^2 is higher than the corresponding probability of obtaining large e_t^2 realizations under a distribution with low variance. Hence, based upon the objective function (3), the likelihood of classifying a date as a break date towards a low-variance regime is higher when the date is in a high-variance regime than when it is in a low-variance regime. Interestingly, this effect is overcompensated by the heteroskedasticity correction incorporated in the objective function (2) of the sup-Wald procedure. This overcompensation leads to skewness in the sup-Wald distributions in the opposite direction of the skewness exhibited by the LS distributions.

As to the rather low precision of the break date estimators, note that their objective functions (2) and (3) are based upon fourth-order sample moments; this generates an enormous dispersion in the objective functions, and hence in the resulting estimates. To address this particular issue one could use regressions of $|e_t|$ as a measure for σ (instead of e_t^2 as a measure for σ^2) to construct corresponding LS-break date estimators. Indeed, reconducting Experiment 1 for the LS procedure with auxiliary regressions based upon $|e_t|$ reveals a significant reduction in the dispersion of the corresponding break date estimates relative to those based upon e_t^2. In particular, the 67% and 95% quantile intervals are [1982:II, 1984:I] and [1976:IV, 1985:I] in this case, compared with intervals of [1980:II, 1984:I] and [1964:IV, 1984:II] obtained using e_t^2.

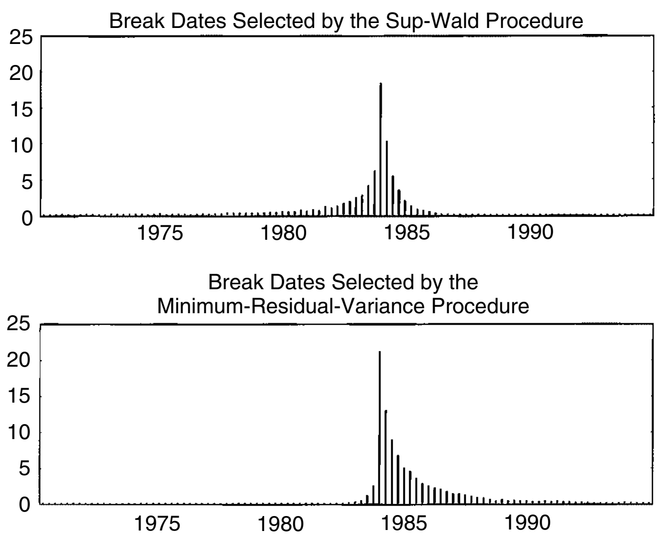

Figure 3. Histogram of break date estimates obtained using the sup-Wald procedure (upper panel) and least squares procedure (lower panel) for 10 000 artificial realizations of $\{g_t\}$. The true break point is at 1984:I, and the standard deviation of the AR(1) innovations is 2.00 before and 4.46 after the break

Yet another important feature of the break date estimators responsible for their imprecision is that the variance estimates $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are based upon incorrect classifications of the e_t^2 observations into high- and low-variance regimes for all potential break dates except for the true date. Hence the objective functions (2) and (3) are based upon inconsistent parameter estimates. This tends to flatten the objective functions relative to their hypothetical counterparts based upon the true values of σ_1^2 and σ_2^2, and also relative to functions based upon consistent estimates. In the following section we present an alternative approach to identifying break dates that meets these objections, and so may be more suitable in pinpointing break dates.

3. A PROBABILISTIC APPROACH

As above, let T denote the total number of observations of the variable g_t. We begin by dividing these observations into three groups. The first contains observations in the range of dates $[1, T_1]$, the second contains observations in the range $[T_1 + 1, T_2]$, and the third contains observations in the range $[T_2 + 1, T]$. By assumption, the first group is known to belong to the regime prevailing before the break, and the last to the regime prevailing after the break. The goal of the exercise involves dividing the observations in the middle group into their respective regimes. This is accomplished by comparing probabilities associated with each possible division of these observations.

To calculate these probabilities, let $L_1(\cdot|\theta_1)$ denote the likelihood function prevailing under the initial regime, and $L_2(\cdot|\theta_2)$ the likelihood function prevailing under the final regime (the θ's representing vectors of parameters). Finally, let j_0 denote the true unknown break date; following convention, j_0 is defined as the last period of the first regime. The conditional likelihood given the occurrence of a break at date j, $L(\{g_t\}|j_0 = j, \theta_1, \theta_2)$, is given by

$$L(\{g_t\}|j_0 = j, \theta_1, \theta_2) = L_1(\{g_t\}_{t=1}^{j}|\theta_1) \cdot L_2(\{g_t\}_{t=j+1}^{T}|\theta_2) \tag{4}$$

The conditional probability associated with the occurrence of a break at date j, $p(j_0 = j|\{g_t\}, \theta_1, \theta_2)$, is given by its likelihood value relative to the likelihood values associated with each potential break date:

$$p(j_0 = j|\{g_t\}, \theta_1, \theta_2) = \frac{L(\{g_t\}|j_0 = j, \theta_1, \theta_2)}{\sum_{\tau=T_1+1}^{T_2} L(\{g_t\}|j_0 = \tau, \theta_1, \theta_2)} \tag{5}$$

Our point estimate of j_0 is simply the date $j \in [T_1 + 1, T_2]$ receiving the highest conditional probability, i.e.,[1]

$$j_p^* = \arg \max_{\{T_1 \le j \le T_2\}} [p(j_0 = j|\{g_t\}, \theta_1, \theta_2)] = \arg \max_{\{T_1 \le j \le T_2\}} [L(\{g_t\}|j_0 = j, \theta_1, \theta_2)] \tag{6}$$

Further, by accumulating probabilities, we can compute the CDF of j_p^*, from which we can calculate the conditional probability that j_0 falls within any given interval. In turn, for any given nominal coverage probability (e.g., 67% or 95%), one can construct associated coverage interval(s).

[1] Note that the calculation of (5) and (6) requires parameter estimates $(\hat{\theta}_1, \hat{\theta}_2)$; alternative strategies for obtaining such estimates are discussed below. Note further that θ_1 and θ_2 can potentially contain subsets of elements that are identical, which would enhance the efficiency with which estimates of θ_1 and θ_2 may be obtained.

Here, approximate lower bounds of the nominal $x\%$ intervals we construct guarantee that the CDFs evaluated at that point are at least as large as $(100 - x)/200$; approximate upper bounds are constructed analogously. Note that coverage intervals are distinct from quantile intervals of the sampling distribution: the former reflect uncertainty regarding the break date conditional on the observed sample; the latter reflect uncertainty one would encounter in facing repeated samples. As above, small-sample quantile intervals are computed through Monte Carlo experimentation.

This probabilistic approach is closely related to allocation rules used in classification analysis to partition data into disjoint groups or clusters typically using the objectives' probabilities of belonging to particular groups (see, e.g., Seber, 1984). Note that in our case all observations are automatically and simultaneously allocated into their corresponding group by fixing the break date, given the temporal ordering of the observations. Taking this special feature into account, our probabilistic approach is equivalent to standard assignment rules in classification analysis which minimize the probability of classification errors (e.g., see Seber, 1984, ch. 6.2). Also notice that if the prior probability of $j_0 = j$ is equal for all $j \in [T_1 + 1, T_2]$ the probability (5) corresponds (for known parameter values θ_1 and θ_2) to the posterior probability of $j = j_0$. This adds a Bayesian flavour to the probabilistic approach. However, a fully Bayesian approach, which is beyond the scope of this paper, would additionally require a prior distribution for the parameters θ_1 and θ_2. Finally, note that if the distributions of g_t under the two regimes are *completely known* and *Gaussian*, Bai's (1997) LS estimate for a break date in the *mean process* of g_t is identical to that estimated by the probabilistic approach: in this case the objective function of the probabilistic approach $L(\{g_t\}|j_0 = j, \theta_1, \theta_2)$ is a monotonic transformation of the corresponding sum of squared errors which are to be minimized by the LS estimate. However, the probabilistic approach generalizes more naturally to the case of breaks in higher-order moments and to the case of multiple breaks. In particular, for the case of n breaks with true break dates $j_0^{(1)}, \ldots, j_0^{(n)}$ one would maximize the corresponding probability $p(j_0^{(1)} = j^{(1)}, \ldots, j_0^{(n)} = j^{(n)}|\{g_t\}, \theta_1, \ldots, \theta_{n+1})$.

Practical implementation of the probabilistic procedure requires specifications of the likelihood functions and estimates of the parameters appearing in (5). Likelihood functions may be specified parametrically, semi-parametrically or non-parametrically. In the empirical applications below, we first employ a parametric implementation, and then outline a non-parametric implementation. Also, there are various ways in which available information may be used to obtain parameter estimates. We consider two approaches here. In the first (hereafter, the one-shot approach), θ_1 is estimated (e.g., by maximum likelihood) using the observations in the range $[1, T_1]$, and remains fixed for all subsequent probability calculations according to (5) (and likewise for θ_2). This ensures that the parameter estimates are based only on observations which are by assumption correctly allocated to their corresponding regimes and, hence, leads to an objective function based on consistent parameter estimates. However, this approach ignores potentially useful information provided by observations between $T_1 + 1$ and T_2. Accordingly, in the second (hereafter, the sequential approach), θ_1 and θ_2 are estimated initially as just described, but are then updated by sequentially assigning initially unallocated observations to the two regimes. This is accomplished as follows. Given the initial estimates of θ_1 and θ_2, we calculate the probability that $j_0 = T_1 + 1$, and then the probability that $j_0 = T_2$. If the former probability is smaller than the latter, we increase T_1 by one period (thus classifying its associated observation as belonging to regime 1), and re-estimate θ_1. Otherwise we decrease T_2 by one period, and re-estimate θ_2. This process is then repeated, and continues until all observations have been classified. At that stage, the last date associated with regime 1 is the estimated break date. Note that this algorithm ensures that the allocation in each step is based upon consistent parameter estimates given the allocation decision in the preceding step. Associated coverage intervals are constructed using the CDF obtained using the final estimates of $(\theta_1,\ \theta_2)$.[2]

4. EMPIRICAL APPLICATION, PARAMETRIC IMPLEMENTATION

As in Section 2, our interest in this application is in identifying the timing of the break in the innovation variance σ^2 associated with (1) for GDP growth. This involves assigning the residuals associated with (1) to separate likelihood functions. We do this here under the assumption that innovations to (1) are Gaussian, implying normal likelihood functions for the residuals.

Under normality, θ_1 and θ_2 consist of a single parameter: the innovation variances σ_1^2 and σ_2^2. The break date estimate is according to (5)–(7) obtained by maximizing the following objective function:

$$Q(j) = -\frac{1}{2}\sum_{t=1}^{j}\left[\ln(\tilde{\sigma}_1^2) + \frac{e_t^2}{\tilde{\sigma}_1^2}\right] - \frac{1}{2}\sum_{t=j+1}^{T}\left[\ln(\tilde{\sigma}_2^2) + \frac{e_t^2}{\tilde{\sigma}_2^2}\right] \tag{7}$$

where $\tilde{\sigma}_1^2$ and $\tilde{\sigma}_2^2$ are estimates for the innovation variances obtained using sample variances of the residuals e_t from autoregression (1) computed over the relevant ranges of the sample. As above, initial and final ranges used to obtain these estimates for the one-shot implementation and to obtain initial estimates for the sequential implementation correspond to the first and last 15% of the total observations. Results obtained using the one-shot and sequential estimation procedures are presented in Table III; break-point probability estimates and corresponding CDFs are plotted in Figure 4.

Note from Table III that point estimates obtained using both procedures are as identified above: 1984:I. In addition, 67% and 95% coverage intervals obtained using the one-shot procedure are [1982:III, 1984:I] (7 quarters) and [1981:IV, 1984:III] (12 quarters); intervals obtained using the sequential procedure are [1983:I, 1984:I] (5 quarters) and [1982:I, 1984:IV] (12 quarters). The similarity of results obtained using these alternative implementations of the probabilistic procedure is underscored in Figure 4 by the similarity of the probability estimates they produce. This similarity reflects the fact that the estimates of σ_1 and σ_2 obtained under the one-shot

Table III. Break date estimates based upon the parametric probabilistic approach

	Estimate	67% coverage interval	95% coverage interval
One-shot	84:I	82:III–84:I(7)	81:IV–84:III(12)
Sequential	84:I	83:I–84:I(5)	82:I–84:IV(12)

Note: The numbers of quarters covered by coverage intervals are given in parentheses.

[2] Yet another possible implementation would be a probabilistic version of the k-means algorithm proposed, e.g., by Sebestyen (1962) in order to partition data into clusters: in the first step the break date is estimated according to the one-shot approach based on parameter estimates using the observations in the intervals $[1, T_1]$ and $[T_2 + 1, T]$; in the second step parameters are re-estimated based on the partition of the observations obtained in the first step. The steps are repeated until no further increase in the probability (5) occurs. This algorithm could also be the basis for an implementation for estimating multiple breaks. The implementation of this algorithm to the case of one and multiple breaks is left for future research.

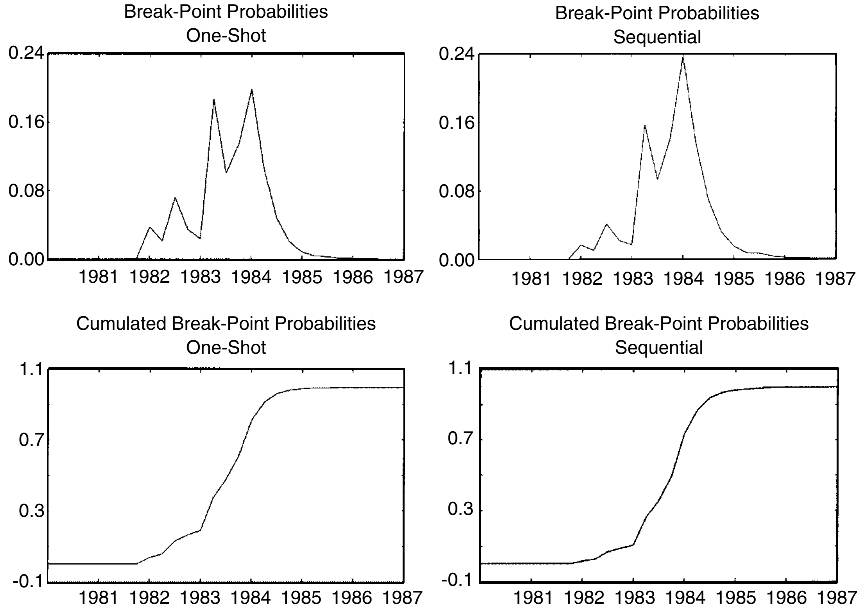

Figure 4. Estimated break date probabilities for a break in the innovation variance associated with an AR(1) for GDP growth obtained using the parametric one-shot implementation (upper left panel) and the parametric sequential implementation (upper right panel); with corresponding cumulated break-point probabilities (lower panels)

procedure, 5.04 and 2.08, are close to the ultimate estimates 4.46 and 2.00 obtained under the sequential procedure.

We now turn to an analysis of the repeated-sample performance of our procedures, which can be used to produce small-sample quantile intervals. Once again, this is facilitated via Monte Carlo experimentation, under which we applied both versions of our procedure to data generated using the DGPs employed in Experiments 1–6 above. Results of these experiments are presented in Table IV.

The first result to note from Table IV regards the invariance of the performance of our procedure to the location of breaks and temporal ordering of the high- and low-variance regions. The distributions of estimated break dates obtained under all six DGPs are roughly symmetric under both versions of our procedure. This is in contrast to the behaviour of the distributions associated with the sup-Wald and LS procedures. Furthermore, in contrast to sup-Wald and LS, the probabilistic estimators seem to be unbiased. Moreover, the dispersions of the distributions associated with both versions of our procedure are roughly stable across the six DGPs, and much tighter. For example, under Experiment 1, 67% and 95% quantile intervals obtained using the one-shot procedure are [1983:I, 1984:II] (6 quarters) and [1980:IV, 1987:II] (27 quarters); and for the sequential procedure are [1983:I, 1984:II] (6 quarters) and [1981:I, 1986:IV] (24 quarters). In contrast, recall that the 67% and 95% intervals obtained using the sup-Wald procedure spanned 26 and 47 quarters, those obtained using the LS procedure based on squared residuals spanned 16 and 79 quarters and based on absolute residuals 8 and 34 quarters. Also note that standard deviations are significantly reduced: from 13.23 (sup-Wald), 20.44 (LS based on squared residuals)

Table IV. Monte Carlo experiments with the parametric probabilistic approach

	Mean	Std. dev.	67% interval	95% interval
	Experiment 1: true break date 84:I with $\sigma_1 = 4.46$, $\sigma_2 = 2.00$			
One-shot	83.81	6.15	83:I–84:II(6)	80:IV–87:II(27)
Sequential	83.79	5.65	83:I–84:II(6)	81:I–86:IV(24)
	Experiment 2: true break date 75:I with $\sigma_1 = 4.46$, $\sigma_2 = 2.00$			
One-shot	74.91	7.42	74:I–75:III(7)	72:I–79:I(29)
Sequential	74.82	6.50	74:I–75:II(6)	72:II–77:III(23)
	Experiment 3: true break date 66:I with $\sigma_1 = 4.46$, $\sigma_2 = 2.00$			
One-shot	65.92	8.52	65:I–66:II(6)	62:IV–70:I(30)
Sequential	65.78	7.09	65:I–66:II(6)	63:I–68:III(23)
	Experiment 4: true break date 84:I with $\sigma_1 = 2.00$, $\sigma_2 = 4.46$			
One-shot	84.05	8.83	83:III–85:I(7)	80:I–87:I(29)
Sequential	84.20	7.29	83:IV–85:I(6)	81:IV–87:I(22)
	Experiment 5: true break date 75:I with $\sigma_1 = 2.00$, $\sigma_2 = 4.46$			
One-shot	75.10	7.67	74:III–76:I(7)	71:I–78:II(30)
Sequential	75.17	6.67	74:IV–76:I(6)	72:II–78:I(24)
	Experiment 6: true break date 66:I with $\sigma_1 = 2.00$, $\sigma_2 = 4.46$			
One-shot	66.17	6.40	65:IV–67:I(6)	62:II–69:I(28)
Sequential	66.19	5.81	65:IV–67:I(6)	62:IV–69:I(26)
	Bootstrap from the actual sample with a break date at 84:I			
One-shot	83.50	6.02	82:IV–84:II(7)	80:I–86:I(25)
Sequential	83.50	5.43	82:IV–84:I(6)	80:II–85:III(22)

Note: The 67% interval is the interval between the 16.5% and the 83.5% quantile and the 95% interval is the interval between the 2.5% and the 97.5% quantile. The numbers of quarters spanned by the intervals are given in parentheses. Means and standard deviations are given in decimal notation, where .00, .25, .50 and .75 represent quarters I, II, III and IV.

and 10.07 (LS based on absolute residuals) to 6.15 (one-shot) and 5.65 (sequential). Finally, it should be mentioned that under all experiments the sequential approach performs slightly better than the one-shot implementation.

The explanation for the improved precision of the probabilistic approach relative to the sup-Wald and the LS approach based on squared residuals is that the objective function of the probabilistic approach (7) is based on second-order sample moments of e_t rather than fourth-order moments, while the objective functions (3) and (4) are based on fourth-order sample moments. Moreover, as discussed above, the probabilistic approach employs consistent (one-shot) or 'step-wise' consistent (sequential) parameter estimates of σ_1 and σ_2, while the sup-Wald and LS procedures employ inconsistent estimates. This is particularly relevant when comparing the probabilistic approach with the LS procedure based on absolute residuals. Regarding the superior performance of the sequential implementation relative to the one-shot approach, note that the former exploits information from observations between T_1 and T_2 as soon as they are allocated to their regimes, while such information is ignored by the latter.

We conclude with an additional experiment, which is a bootstrap version of Experiment 1. Specifically, here we modified the DGP used in Experiment 1 so that, instead of obtaining artificial realizations of innovations from normal distributions, we obtained them by sampling with replacement from the residuals obtained by estimating (1) with a break date at 1984:I. All other aspects of the DGP are unchanged. We considered this experiment in order to investigate the sensitivity of our results to a departure from the normality assumption. Results of this experiment are also given in Table IV, and match closely those obtained under Experiment 1. For example, the mean of the break date estimates is now 1983:III rather than 1984:I; and quantile interval ranges differ by no more than 2 quarters. The similarity of these results suggests that the assumption of normality does not play a critical role in this application.

5. EMPIRICAL APPLICATION, NON-PARAMETRIC IMPLEMENTATION

As in any likelihood-based analysis, one can always replace distributional assumptions used to specify likelihood functions with non-parametric approximations. Typically, this involves a tradeoff of efficiency versus robustness. Here, we demonstrate the performance of our procedure using a particular non-parametric implementation, applied once again to an identification of the timing of the break in the innovation variance σ^2 associated with (1) for GDP growth. In this case, instead of assigning the residuals associated with (1) to separate likelihood functions specified parametrically, the likelihood functions are approximated non-parametrically. As in the parametric implementation, likelihood approximations are computed using the one-shot and sequential methods for defining reference subsamples.

The implementation considered here is based on the use of a Gaussian kernel. For a given sequence of observations $\{g_t : t : 1 \to N\}$ the pdf associated for a given value of g, $f(g)$, is approximated using

$$f(g) \propto \frac{1}{h \cdot N} \sum_{t=1}^{N} \exp\left\{-\frac{1}{2}\left(\frac{g - g_t}{h}\right)^2\right\} \tag{8}$$

where the parameter h is a bandwidth parameter that determines the smoothness of the approximated pdf.

In the present context, we must estimate two densities, one for each regime. If our objective were that of producing estimates of these component densities with desirable asymptotic properties, we would select bandwidths accordingly (e.g., see Härdle, 1989). Here, however, our objective is that of producing accurate (finite-sample) estimates of the break date, not of the regime densities themselves. Therefore, we proceeded by selecting the bandwidths of the two regimes (h_1 and h_2) so as to minimize the standard deviation of the sequential estimates of the break date. This 'calibration' exercise is conducted by Monte Carlo simulation. We considered two simulations: a Gaussian calibration, using the parameter and break date estimates obtained using our one-shot parametric procedure; and a bootstrap calibration, using bootstrap simulations of the corresponding two subsamples. The former yielded $h_1 = 1.737$, $h_2 = 1.026$; the latter yielded $h_1 = 1.515$, $h_2 = 0.777$.

The selection of bandwidths unavoidably involves calibration. However, our experience in this case suggests that this calibration need not be particularly fine, since moderate departures (e.g., 10%–20%) from the 'optimal' values we have produced yield very similar results. This relative robustness suggests that, in general, one ought to be able to construct good rules of thumb for bandwidth selection using, e.g., the 'response surface' techniques of Hendry (1984). In the present case, one such rule of thumb involves fixing the bandwidth ratio h_1/h_2 equal to that of the ratio

Table V. The non-parametric probabilistic approach

	Break date estimates for the actual data			
	Estimated date		67% coverage interval	95% coverage interval
One-shot	83:II		83:I–84:II(6)	81:IV–85:III(16)
Sequential	84:I		83:I–84:II(6)	82:I–85:I(13)
	MC Experiment 1: true break date 84:I with $\sigma_1 = 4.46$, $\sigma_2 = 2.00$			
	Mean	Std. dev.	67% interval	95% interval
One-shot	83.50	6.04	82:III–84:I(7)	80:I–86:I(25)
Sequential	83.50	5.63	82:III–84:I(7)	79:IV–85:II(23)

Note: Non-parametric density estimates for the residuals are obtained by using a Gaussian density kernel with bandwidth parameter $h_1 = 1.737$ for the range $[1, T_1]$ and $h_2 = 1.026$ for the range $[T_2 + 1, T]$. These values were obtained using a Gaussian Monte Carlo calibration used to minimize the variance of the break date estimator. The 67% interval is the interval between the 16.5% and the 83.5% quantile, and the 95% interval is the interval between the 2.5% and the 97.5% quantile. The numbers of quarters spanned by the intervals are given in parentheses. Means and standard deviations are given in decimal notation, where 0.00, 0.25, 0.50 and 0.75 represent quarters I, II, III and IV.

of standard deviations σ_1/σ_2. This reduces the dimensionality of the calibration exercise (response surface) to one dimension without involving a significant loss of performance. Here, use of the two pairs of values we produced yielded similar results, thus we report below only those results obtained using the former pair; this is done in Table V.

Application of this procedure to the actual data yielded point estimates of 1983:II in the one-shot case, and 1984:I in the sequential case (recall both point estimates are 1984:I in the parametric implementations). Also, three of the four sets of coverage intervals obtained in this case closely match their parametric counterparts (the exception being the 95% interval obtained under the one-shot implementation, which is four quarters wider than its parametric counterpart). This similarity of results once again suggests that the assumption of normality does not play a critical role in this application.

Finally, we concluded our analysis by reconducting Experiment 1 using these non-parametric implementations; results are also presented in Table V. Once again, means, standard deviations and quantile intervals are similar to their parametric counterparts. Means of the estimated break dates are 1983:III under both the one-shot and sequential implementations, as opposed to 1983:IV in the parametric case; 67% quantile intervals span 7 quarters here, as opposed to 6 quarters in the parametric case; and 95% intervals are 25 and 23 quarters here (one-shot and sequential implementations), compared with 27 and 24 quarters under the parametric implementations.

Taken together, the full range of results we have obtained using the various implementations of our procedure speak to its general effectiveness in pinpointing the timing of a break even in higher-order moments of a process like the variance.

6. CONCLUSION

We have proposed a two-step strategy for assessing structural stability in time-series frameworks when potential change dates are unknown. The first step involves the use of existing tests for structural stability in detecting the presence of structural change. The second step involves the use of the likelihood-based procedure we have proposed for assigning conditional probabilities to the occurrence of structural breaks at alternative dates.

The procedure is designed to classify each element of a set of observations into one of several potential categories. Given a model specification with one regime change identified *a priori*, associated observations will correspond with one of two likelihood functions: one prevailing prior to the break date and one after the break date. The procedure identifies the break date as the optimal point at which to divide the sample.

We have illustrated the performance of our procedure through a series of Monte Carlo experiments, and an assessment of the volatility reduction in the growth rate of US GDP. Using both parametric and non-parametric implementations, we have found the procedure to be effective in improving the precision with which inferences regarding timing can be made. Notably, under our leading experimental design, Bai's (1997) least squares procedure based on squared residuals generates a finite-sample standard deviation of 20.44 quarters and based on absolute residuals a standard deviation of 10.07 quarters, while our procedure generates a standard deviation of 6.15 quarters under the parametric implementation we employ, and 5.65 quarters under the non-parametric implementation.

ACKNOWLEDGEMENTS

We gratefully acknowledge helpful comments provided by Dick van Dijk and two anonymous referees on an early version of this paper.

REFERENCES

Anderson TW. 1984. *An Introduction to Multivariate Statistical Analysis*, 2nd edn. John Wiley & Sons: New York.

Andrews DWK. 1993. Tests for parameter instability and structural change with unknown change point. *Econometrica* **61**: 821–856.

Bai J. 1997. Estimation of a change point in multiple regression models. *Review of Economics and Statistics* **74**: 551–563.

Hansen BE. 1997. Approximate asymptotic *p*-values for structural change tests. *Journal of Business and Economic Statistics* **15**: 60–67.

Hansen BE. 2001. The new econometrics of structural change: dating breaks in U.S. labor productivity. *Journal of Economic Perspectives* **15**: 117–128.

Härdle W. 1989. *Applied Nonparametric Regression*. Econometric Society Monographs No. 19. Cambridge University Press: Cambridge.

Hendry DF. 1984. Monte Carlo experimentation in econometrics. In *Handbook of Econometrics*, Griliches Z, Intriligator MD (eds). North-Holland: Amsterdam.

Kim CJ, Nelson CR. 1999. Has the U.S. economy become more stable? A Bayesian approach based on a Markov-switching model of business cycle. *Review of Economics and Statistics* **81**: 1–10.

McConnell MM, Perez-Quiros G. 2000. Output fluctuations in the United States: what has changed since the early 1980's? *American Economic Review* **90**: 1464–1476.

Quandt R. 1960. Tests of the hypothesis that a linear regression obeys two separate regimes. *Journal of the American Statistical Association* **55**: 324–330.

Seber GAF. 1984. *Multivariate Observations*. John Wiley & Sons: New York.

Sebestyen GS. 1962. *Decision Making Process in Pattern Recognition*. Macmillan: New York.

Stock JH, Watson MW. 2002. Has the business cycle changed and why? Manuscript, Harvard University, Department of Economics.

Wang J, Zivot E. 2000. A Bayesian time series model of multiple structural changes in level, trend and variances. *Journal of Business and Economic Statistics* **18**: 374–386.

