Likelihood-Based Confidence Sets for the Timing of Structural Breaks

Eo, Yunjong and Morley, James C.

Washington University in St. Louis, Washington University in St. Louis

5 September 2008

Likelihood-Based Confidence Sets for the Timing of Structural Breaks *

Yunjong Eo and James C. Morley †

Department of Economics
Washington University in St. Louis
Campus Box 1208, One Brookings Drive
St. Louis, MO 63130-4899

First draft: March 8 2008
This draft: September 5 2008

Abstract

In this paper, we propose a new approach to constructing confidence sets for the timing of structural breaks. This approach involves using Markov-chain Monte Carlo methods to simulate marginal "fiducial" distributions of break dates from the likelihood function. We compare our proposed approach to asymptotic and bootstrap confidence sets and find that it performs best in terms of producing short confidence sets with accurate coverage rates. Our approach also has the advantages of i) being broadly applicable to different patterns of structural breaks, ii) being computationally efficient, and iii) requiring only the ability to evaluate the likelihood function over parameter values, thus allowing for many possible distributional assumptions for the data. In our application, we investigate the nature and timing of structural breaks in postwar U.S. Real GDP. Based on marginal fiducial distributions, we find much tighter 95% confidence sets for the timing of the so-called "Great Moderation" than has been reported in previous studies.

Keywords: Fiducial Inference; Bootstrap Methods; Structural Breaks; Confidence Intervals and Sets; Coverage Accuracy and Expected Length; Markov-chain Monte Carlo;

JEL classification: C15(Simulation Methods); C22 (Time-Series Models)

---

*We thank Jeremy Piger, Werner Ploberger, Junsoo Lee, Tao Zha, and participants at the 2008 North American Summer Meeting of the Econometric Society, at the 2008 Missouri Economic Conference, and in the applied time series research group at Washington University for helpful comments. All errors are our own.

†Contact: Yunjong Eo (eo@wustl.edu) or James C. Morley (morley@wustl.edu).

1 Introduction

In this paper, we propose a new approach to constructing confidence sets for the timing of structural breaks in time series.[1] Our proposed approach involves simulating marginal "fiducial" distributions of break dates from the likelihood function. The practical implementation of this approach utilizes Markov-chain Monte Carlo (MCMC) methods that have been widely used for posterior simulation in the Bayesian literature. Indeed, the resulting confidence sets are the same as Bayesian highest-posterior-density (HPD) credible sets given noninformative priors. However, we take a strictly classical viewpoint with respect to our inferences about the confidence sets by considering their coverage accuracy and expected length across repeated samples.

In order to motivate our use of "fiducial" distributions to construct confidence sets, it is necessary to discuss fiducial inference, which was first developed by Fisher (1930) and involves making probability statements about parameters, with the probabilities being proportional to the likelihood function and having a frequentist interpretation.[2] While the idea of making probability statements directly about parameters is antithetical to the standard classical viewpoint, it has long been understood that there can be, in certain settings, a close relationship between fiducial confidence sets and classical confidence sets. In particular, in the case of a single parameter for which a pivotal test statistic is available, fiducial confidence sets and the classical confidence sets based on inverting the test statistic will be the same, even if their interpretation is different.[3] Thus, our strategy here is to use fiducial distributions as a means of generating a classical estimator for a confidence set, much in the same way as a likelihood function is used to generate a classical estimator for a model parameter.

[1] Because our approach can produce disconnected subsets of possible break dates, we prefer the terminology of "confidence set" to "confidence interval", although we still refer to "length" rather than "size" of a confidence set to make it clear that we are considering inferences about a single parameter, rather than a multi-dimensional confidence region.

[2] For more details about fiducial inference, see Fraser (1961a,b). Barnett (1999) provides an accessible and in-depth discussion of the issues surrounding fiducial inference in his textbook on comparative methods of statistical inference. Recently, Hannig (2006) argues for the fiducial approach as a tool for deriving classical inference procedures, which is the general strategy that we take in this paper.

[3] A "pivotal" test statistic has a known distribution that is independent of the model parameters. For example, a t-statistic in a simple linear regression model with exogenous regressors and serially uncorrelated errors is pivotal because it has a Student-t distribution that only depends on sample size and the number of regressors, not on the values of model parameters.

Our proposed approach can be related to some existing methods for constructing confidence sets. It is most directly analogous to Sims and Zha (1999), who consider Bayesian credibility sets based on noninformative priors as a means of constructing classical "error bands" for impulse response functions in structural Vector Autoregressive (VAR) models. On a more general level, our approach has a similar motivation to bootstrap methods for constructing confidence sets (see, for example, Kilian (1999) and MacKinnon (2002)). Specifically, while bootstrap confidence sets will only be exact when based on a pivotal quantity, they appear to perform well, and notably better than confidence bands based on asymptotic distributions, in settings when a bootstrap distribution approximates finite-sample distributions that are close to being pivotal. Along these lines, we consider the possibility that a marginal fiducial distribution implicitly provides an even better approximation of finite-sample distributions that are even closer to being pivotal. Meanwhile, because our approach uses marginal distributions, it can also be related to the extensive testing literature in which nuisance parameters are integrated out of the likelihood (see, for example, Andrews and Ploberger (1994)).[4] Indeed, our approach is directly motivated by Elliott and Müller (2007), who propose constructing confidence sets for the timing of structural breaks by inverting a test statistic for which nuisance parameters have been integrated out.

In the specific setting of making inferences about the timing of structural breaks, we compare our fiducial distribution (FD) approach to a range of asymptotic and bootstrap methods. In terms of asymptotic methods, Bai (1997) provides the standard approach to constructing a confidence interval for a single break, Bai, Lumsdaine, and Stock (1998) consider multivariate models, and Bai and Perron (2003) allow for multiple structural breaks. Elliott and Müller (2007) show that Bai's approach attains very low coverage rates when structural breaks in parameters are small and they propose an approach, mentioned above, that is based on inverting sequential tests over the parameter space of break dates. Their approach attains more accurate coverage rates than Bai's approach, but at the cost of much

[4] In the statistics literature, Berger, Liseo, and Wolpert (1999) argue for the use of integrated likelihood methods (i.e., marginal fiducial distributions) as a more practical and robust means of eliminating nuisance parameters than considering a profile likelihood in which the nuisance parameters are maximized out of the likelihood function.

longer confidence sets. Meanwhile, in terms of bootstrap methods, we are motivated, in part, by Diebold and Chen (1996)'s comparison of asymptotic and bootstrap methods for testing the existence of structure change. They find that, especially for smaller samples and persistent dynamics, the bootstrap approximation to the finite-sample distributions of these tests is usually more accurate than the asymptotic approximation. Here, we use bootstrap methods to construct confidence sets for the break dates themselves. In terms of bootstrap methods, we consider constructing confidence sets using a "bootstrap percentile" approach, a "bootstrap standard error" approach, a "bootstrap inverted likelihood ratio (LR)" approach, and an approach based on a bootstrap of Bai's (1997) asymptotically pivotal statistic.

In order to compare the various methods of constructing confidence sets for break dates, we conduct Monte Carlo analysis of coverage accuracy and expected length. This analysis suggests that in finite samples with structural breaks of the kind hypothesized for economic time series such U.S. real GDP, the FD and bootstrap inverted LR approaches are admissible in the sense of producing relatively short confidence sets while maintaining accurate coverage rates compared to nominal confidence levels. Indeed, for sample sizes of 320 and 640, which can be compared with 238 observations for postwar quarterly real GDP between 1947:Q1 and 2005:Q2, the confidence sets for these two likelihood-based approaches are approximately from about a half to one fourth as long as those of other methods, while always attaining exact coverage rates.[5] For example, while Elliott and Müller's inverted test approach succeeds in having accurate coverage rates, the average length of confidence sets for the timing of a break in long-run growth is 40 periods for a sample size of 320 versus only 15 periods using the FD approach. Meanwhile, as discussed in Elliott and Müller (2007), Bai's confidence sets attain too low coverage rates compared to the nominal confidence levels.

The Monte Carlo analysis supports the use of likelihood-based confidence sets for the timing of structural change. Beyond this analysis, however, there are additional practical

[5] Perhaps, from a Bayesian perspective, we should not be too surprised that both the FD and bootstrap inverted LR bootstrap approaches have exact coverage in these Monte Carlo experiments. While they are implemented in different ways, both are directly based on the shape of the likelihood function. Thus, they are both directly linked to Bayesian credibility bands given noninformative priors. If the noninformative priors are appropriate from an "objective" Bayesian point of view, such that the posterior probabilities in a given experiment are correct in a "fair bet" sense conditional on the data, then it follows that the frequentist coverage across repeated experiments will also be correct. We discuss this interpretation in more detail in the next section.

reasons to use likelihood-based confidence sets and the FD approach in particular.

First, the FD approach is broadly applicable in the sense that we can easily consider structural breaks in different parameters occurring at different dates. This more complicated pattern of structural breaks has been hypothesized for U.S. real GDP and other macroeconomic time series. In particular, Kim and Nelson (1999) and McConnell and Perez-Quiros (2000) detect the break date of the volatility deduction in 1984:Q1, while Zivot and Andrews (1992) and others find that there is a trend break in 1972:Q2 in the unit root literature. Stock and Watson (1996, 2002) show that most of U.S. macroeconomic data are unstable and have volatility changes. Among the methods considered here, only the FD and bootstrap approaches can construct confidence sets for different break dates of different parameters, such as a mean and a variance changing at different points of the sample. However, in this setting, the bootstrap approaches produce multi-dimensional confidence regions, rather than confidence sets for each of the individual structural breaks. To get a confidence set for a specific break date, we would need to integrate out other break dates, but the integration is generally infeasible for a bootstrap.[6] By contrast, integrating out break dates and other parameters is a straightforward feature of the FD approach.

Second, although bootstrap methods are available for general settings, the FD approach is much more computationally efficient, especially for more complicated structural changes (e.g., respective breaks in mean and variance). Suppose, for example, we have 100 observations for an econometric model with two parameters such as a constant and a variance and both parameters have structural breaks, but at different dates. Implementing bootstrap methods for this model is computationally costly because for each bootstrap data set we would need to consider 10,000 (=100×100) combinations of two break dates and estimate break dates as well as model parameters, a mean and a variance before and after breaks,

[6] In principle, one could consider an inverted LR approach based on the likelihood profile with respect to one break at a time. That is, the likelihood could be calculated for all different possibilities of one break date, while maximizing the likelihood with respect to the other parameters and break dates. However, in the case of multiple break dates, there is a conceptual issue in terms of which possibilities to consider for one break date given the possibilities for the other break dates. For example, suppose the likelihood is maximized with structural breaks in periods 10 and 20. The likelihood profiles for each break date will have two peaks of equal height with the given break occurring in either period 10 or in period 20 and the other break estimated to have occurred in period 20 or period 10. Meanwhile, in the context of the bootstrap inverted LR approach, the calculation of the likelihood profile for every break in every bootstrap sample would be computationally impractical.

simultaneously by maximum likelihood estimation (MLE).[7] In the case of 199 bootstraps, which is a relatively small number of artificial samples, we would need to conduct MLE 1,990,000 (=199×10,000) times. To give a sense of this number, even if it takes just one second per estimation, 1,990,000 cases would take more than 23 days.

Third, the FD approach requires only the ability to evaluate the likelihood function over parameter values and potential break dates. Thus, it can be applied given any econometric model with a specified likelihood function (e.g., models with normal distributions, Student-t distributions, Poisson distributions, and so on). While the bootstrap inverted LR approach is also available, in principle, for any specified likelihood function, computational difficulties become an issue again. This is because numerical optimization over the entire parameter space can be challenging for non-normal models due to irregularities in the likelihood surface. By contrast, the evaluation of the likelihood for the FD approach via MCMC methods can be broken down into more manageable steps. In addition, we only need to evaluate the likelihood for the actual data, as opposed to considering restricted and unrestricted likelihoods for the actual data and for each bootstrap sample in the case of the bootstrap inverted LR approach.

For our application, we examine the nature and timing of structural breaks in postwar U.S. real GDP. First, we apply sequential tests to determine the number and types of structural breaks. Then, for each break, we employ the various methods of constructing confidence sets considered in our Monte Carlo analysis to make inferences about their timing. Under the assumption of a unit root, we find support for the Great Moderation in the form of a variance reduction in quarterly growth rates, with the maximum likelihood estimate of the break in 1984:Q1 and a narrow 95% confidence set of 1982:Q1 to 1984:Q4 based on the FD approach. Notably, this 95% confidence set is smaller than the 67% confidence set based on Bai's (1997) approach reported in Stock and Watson (2002). In our application, the lengths of the confidence sets based on Bai's method are more than twice as long as those based on FD approach. Meanwhile, given a unit root, we find no significant evidence of a break in the long-run growth rate, so we do not construct a confidence set for such a break in this case. On the other hand, under the assumption of trend stationarity, we do find significant

[7] In practice, as discussed in greater detail in the next section, we would need to exclude some portion of the beginning and end of the sample space to avoid a severe distortion in our inferences. However, the number of combinations would still be large.

evidence of a structural break in the form of a reduction in drift, with the maximum likelihood estimate of the break in 1969:Q1. However, the confidence sets for the timing of the drift break are far from precise. For example, the outer bounds of the FD confidence set are 1963:Q2 and 1983:Q3, respectively.[8] Again, we find evidence for the Great Moderation under the assumption of trend stationarity, with maximum likelihood estimate of the break in 1982:Q4 and a narrow FD confidence set from 1982:Q3 to 1984:Q2. Finally, we check the robustness of our inferences for the FD approach by allowing for Student-t distributed errors in our structural break models and we find that the resulting confidence sets for the Great Moderation are very similar to the normal error case.

The remainder of this paper is organized as follows. Section 2 provides details for the various methods of constructing confidence sets for structural break dates. Section 3 reports the results for the Monte Carlo analysis of coverage accuracy and length of the confidence sets for the various methods under consideration. Section 4 presents estimation results for models that capture structural breaks in postwar U.S. real GDP and reports confidence sets for the timing of the structural breaks. Section 5 concludes.

2 Methods

In this section, we provide details of our proposed FD approach to constructing confidence sets for the timing of structural breaks. We also review the asymptotic methods presented in Bai (1997) and Elliott and Müller (2007) and introduce various bootstrapping methods: a bootstrap percentile approach, a bootstrap standard error approach, a bootstrap inverted LR approach, and a bootstrap of Bai's statistic. Previous Monte Carlo studies of bootstrap methods in other settings have shown that no specific approach is always superior in terms of coverage accuracy. For example, see MacKinnon (2002) for confidence intervals of regular parameters and Kilian (1999) for confidence intervals of impulse responses for VAR models. Thus, we consider various methods here in order to determine which ones work best in the context of structural breaks.

[8] The set consists of three disjointed intervals of [1963:2,1980:4], [1981:2,1982:3], and [1983:2,1983:3]. Thus, assuming correct coverage of the confidence set, the outer bounds provide a somewhat conservative inference, although the gaps in the set are relatively short.

In terms of possible patterns of structural breaks, we consider a linear econometric model that allows the variance and coefficients to undergo breaks at different dates.

$$y_t = \sum_{k=1}^{K} X'_{kt}\beta_{kt} + Z'_t\gamma + e_t, \quad e_t \sim i.i.d.\mathcal{N}(0,\sigma_t^2), \quad t = 1,...,T \tag{1}$$

where

$$\sigma_t^2 = \begin{cases} \sigma_0^2 & \text{if } 1 \le t \le \tau_{K+1,1} \\ \sigma_1^2 & \text{if } n_1 < t \le \tau_{K+1,2} \\ \vdots & \\ \sigma_N^2 & \text{if } n_N < t \le T. \end{cases}$$

The model in (1) has K subsets of regressors, X_{kt}'s, $k = 1,...,K$. Each subset has q_k regressors, corresponding to kth group of regressors. That is, each subset of regressors has its own change-point system. In particular, each subset may have a different number of breaks occurring at different dates than those of other groups independently.[9] The coefficient vector β_{kt} is β_{kj} (kth group's $j+1$th regime parameters) if $\tau_{kj} < t \le \tau_{kj+1}, j = 0,..,M_k$. The number of breaks for kth subset is denoted by M_k. Z_t and γ represent regressors and their respective coefficients that do not change. The variance can have N breaks and nth regime's variance is σ_{n-1}^2. In practice, the potential break dates are restricted between the middle $(1-2\lambda)$ portion of the sample period, $\Pi = [\lambda T, (1-\lambda)T]$, to avoid an end of sample distortion. Two consecutive break dates are also at least a distance of λ of the total sample size apart for a similar reason.

2.1 Confidence Sets Based on Marginal Fiducial Distributions

In order to derive the marginal fiducial distributions for break dates, let $f(y|\psi,\tau)$ denote the probability density function (pdf) for a model with structural break date(s) τ and parameters $\psi = (\beta,\gamma,\sigma) \in \Psi$ in (1) evaluated at the observed data $Y = y$. The likelihood function for the model is defined by $L(\psi,\tau|y) = f(y|\psi,\tau)$ since the break date τ can be interpreted as a parameter. Then, briefly ignoring problems of interpretation, a joint pdf for ψ and τ can

[9] For example, consider an AR(3) model with structural breaks. The constant coefficient has one break and the persistence coefficients have 2 breaks. Then, $K = 2$, $X_{1t} = 1$, $X_{2t} = (y_{t-1}, y_{t-2}, y_{t-3})'$, $q_1 = 1$, and $q_2 = 3$.

always be constructed, at least in principle, by simply multiplying the likelihood function by the inverse of its integral (or summation) with respect to model parameters:

$$\pi(\psi, \tau|y) = L(\psi, \tau|y) \times [\sum_{\tau=\lambda T}^{(1-\lambda)T} \int_{\Psi} L(\psi, \tau|y) d\psi]^{-1}. \tag{2}$$

Interpreting these probabilities as having meaning in a frequentist sense is called fiducial inference and is highly controversial in the statistics literature (see Barnett (1999) for a discussion of the controversies surrounding fiducial inference).

In this paper, we do not address the debate over the general coherence of fiducial inference. Our goal here is merely to use fiducial distributions as a means of making more traditional classical inferences. Importantly, we do not directly consider the joint fiducial distribution in (2) because fiducial and classical confidence intervals are generally at odds with each other in multidimensional cases. Instead, we consider the marginal fiducial distributions for the break dates:

$$\pi(\tau|y) = \int_{\Psi} \pi(\psi, \tau|y) d\psi \tag{3}$$

In particular, for a single parameter and given a pivotal test statistic for that parameter, fiducial and classical confidence sets based on inverting the statistic are the same. Thus, if we assume that there exists a test statistic whose distribution is related to the marginal fiducial distribution of a parameter and it is close to being pivotal, even in finite samples, the fiducial confidence intervals will be the similar to classical confidence intervals based on inverting that test statistic. The key point, then, in using fiducial distributions is that we can construct the classical confidence interval without directly having the test statistic.

As mentioned in the introduction, we see our approach as similar to bootstrap analysis. If the finite-sample distribution of the likelihood-based test statistic were pivotal, the confidence set based on the marginal fiducial distributions would be exact. Meanwhile, even if the finite-sample distributions are not strictly pivotal, the FD confidence sets could reflect a better approximation of the finite-sample distributions of the implicit test statistic than asymptotic or bootstrap methods.

The most obvious potential problem with using marginal fiducial distributions to construct confidence sets is finding these distributions in the first place. In particular, it is generally infeasible to use analytical methods to integrate the likelihood function with respect to model parameters in order to get the denominator in (2) and then to integrate the resulting joint fiducial distribution to get the marginal fiducial distribution (3). However, the marginal fiducial distributions of model parameters can be easily simulated via MCMC methods.[10] In particular, suppose the parameters and break date(s) in the model are grouped into two different blocks: ψ and τ, respectively. The parameters in one block can be sampled conditional on data and the parameters in the other block. For model parameters, ψ,

$$\begin{aligned} \boldsymbol{\psi} &\sim \pi(\psi|\tau, \mathbf{y}) \\ &\propto \pi(\psi) f(\mathbf{y}|\psi, \tau) \\ &\propto f(\mathbf{y}|\psi, \tau). \end{aligned}$$

For break dates, τ,

$$\tau \sim \pi(\tau|\psi, \mathbf{y})$$

where

$$\begin{aligned} \pi(\tau|\psi, \mathbf{y}) &= \frac{\pi(\tau) f(\mathbf{y}|\psi, \tau)}{\pi(\mathbf{y}|\psi)} \\ &= \frac{\pi(\tau) f(\mathbf{y}|\psi, \tau)}{\sum_{\tau} \pi(\tau) f(\mathbf{y}|\psi, \tau)} \\ &= \frac{f(\mathbf{y}|\psi, \tau)}{\sum_{\tau} f(\mathbf{y}|\psi, \tau)}. \end{aligned}$$

The MCMC method simulates parameter values from their conditional distributions until the draws behave as if drawn from their joint and marginal distributions. Because the parameters of the denominator in (2) are uniformly integrated, $\pi(\psi)$ and $\pi(\tau)$ should be chosen to be constant over parameter values. For example, $\pi(\tau)$ of one break model is from

[10] Chib and Greenberg (1995) provide a good introduction of MCMC methods by focusing on the Metropolis-Hastings (MH) algorithm. Note that, from a Bayesian perspective, simulation from the likelihood in this way would be considered to be posterior simulation given noninformative priors. This approach is always available, at least in principle, as long as it takes place over the parameter space for which the likelihood is finite.

a uniform distribution on the integers:

$$\pi(\tau) = \begin{cases} \frac{1}{(1-2\lambda)T} & for \quad \tau \in [\lambda T, (1-\lambda)T] \\ 0 & otherwise. \end{cases}$$

Then, the draws behave exactly same as draws from the marginal fiducial distribution in (3) based on integrating the other parameters and break dates out of the joint fiducial distribution in (2). A more detailed MH algorithm used in this paper is presented in an appendix.

Given a simulated marginal fiducial distribution for a break date, we can then construct a confidence set at a $1-\alpha$ level in different ways. In practice, we choose to construct the confidence set using the Bayesian highest-posterior-density (HPD) concept in order to obtain the shortest confidence sets possible for a given confidence level.[11] In particular, the confidence set is

$$S(y) = \{\tau | \pi(\tau|\mathbf{y}) \geq k(1-\alpha)\} \tag{4}$$

where $k(1-\alpha)$ is the largest constant satisfying $\pi(S|\mathbf{y}) \geq 1-\alpha$. Since the break dates have discrete distributions, it is straightforward to find points with highest probability because the simulated marginal fiducial distributions produce different simulated frequencies for each possible break date.[12] When we apply this highest probability concept to constructing the confidence set from the marginal fiducial distribution, the result is a "highest-fiducial density" (HFD) confidence set.

Three issues should be addressed here.

First, in a classical context, the confidence set at any given confidence level $1-\alpha$ can

[11] The notion that lengths of confidence sets with the same confidence levels can differ and shorter ones are to be preferred can be illustrated by the following simple example: Consider T observations of a scalar random variable, $\{x_1, \ldots, x_T\}$, normally distributed with unknown mean μ and known variance 1. We want to construct a confidence interval of the true μ by using the observations. We can think of two different confidence interval estimators that have a 95% confidence level: $[-\infty, \bar{x}+1.65/\sqrt{T}]$ and $[-1.96/\sqrt{T}+\bar{x}, \bar{x}+1.96/\sqrt{T}]$ where $\bar{x}$ is the sample mean. Both confidence intervals would include the true mean μ with probability 0.95 when computed over repeated samples. However, the length of the former is infinite, while that of the latter is only $3.92/\sqrt{T}$. Thus, the former provides much more information about the true μ than the latter in the sense that we can narrow the range of the true value. Thus, we would prefer shorter confidence interval, all else equal.

[12] In the case of a continuous distribution (e.g., a regular parameter in a regression model), we would have to use kernel density estimation for a given bandwidth parameter in order to measure relative densities.

only be justified through its coverage rate in repeated samples. This means that if we could compute confidence sets for infinitely many data sets from population, they would include the true value of the parameter in $100 \times (1-\alpha)\%$ of the data sets. Let τ_0 and $C(Y)$ denote a true break date and any usual classical confidence set estimator, with Y having the distribution $f(y|\tau_0)$ that depends on the true break date τ_0. Then,

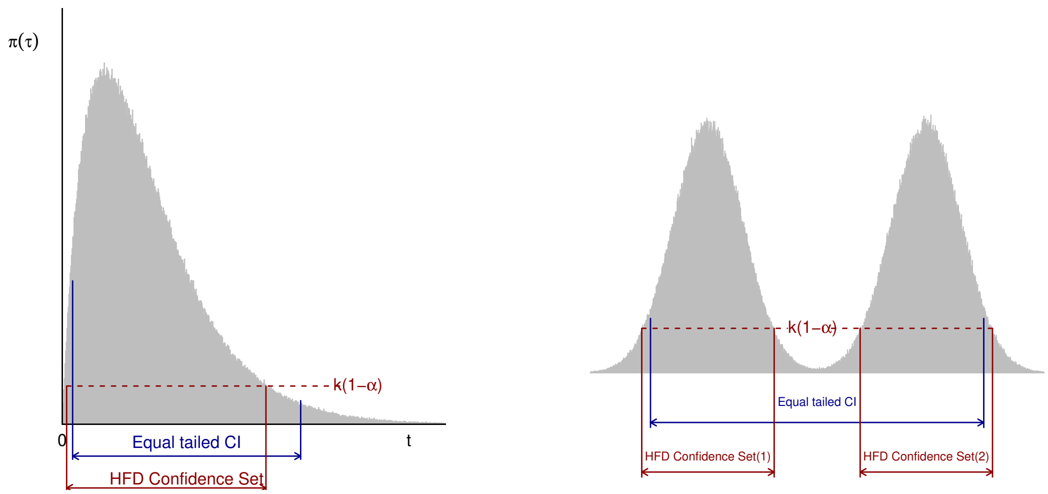
$$\begin{aligned} E[\tau_0 \in C(Y)|\tau_0] &= \int \mathbf{1}[\tau_0 \in C(y)] f(y|\tau_0) dy \\ &= 1 - \alpha \end{aligned}$$

where $\mathbf{1}[\cdot]$ is an indicator function. However, the confidence set estimate for actual data y will not contain the true value with probability $100 \times (1-\alpha)\%$. In a given sample, the confidence set covers the true parameter with probability 0 or 1. By contrast, in the Bayesian context, a posterior distribution of a break date represents a probability distribution conditional on the actual data. Thus, a "credible" set with $1-\alpha$ level has the subjective probability $1-\alpha$ that the true value lies inside of the set. For a Bayesian credible set with noninformative priors, which is equivalent to an FD confidence set,

$$\begin{aligned} E[\tau_0 \in S(Y)|\tau_0] &= \int Pr[\tau_0 \in S(y)] f(y|\tau_0) dy \\ &= \int (1-\alpha) f(y|\tau_0) dy \\ &\quad \text{since } Pr[\tau_0 \in S(y)|\tau_0] = 1-\alpha \text{ for any } y \sim f(y|\tau_0) \\ &= (1-\alpha) \int f(y|\tau_0) dy \\ &= 1-\alpha \end{aligned}$$

Thus, although conceptually different, Bayesian credible sets and, by implication, fiducial confidence sets provide a natural enough means of constructing confidence sets in a classical repeated-sampling sense. Of course, the frequentist performance of Bayesian estimators varies from one setting to another, so the finite-sample coverage rates of these confidence sets for the timing of structural breaks remains an open question to be addressed with our Monte Carlo analysis.

Second, given any confidence level, the HPD credible set has the shortest length while



(a) Asymmetric distribution

(b) Bimodal distribution

Figure 1: Equal-tailed and Highest-Fiducial-Density confidence sets

maintaining the same expected coverage rate as the specified confidence level. The confidence set in (4) might seem odd from a classical viewpoint. Because a Bayesian credible set is constructed from the probability distribution of τ conditional on the observed data, one can directly minimize its length for a given confidence level. In the same way, we are able to minimize the length for FD confidence sets with the HFD concept. Consider, for example, the case in which the fiducial distribution of a parameter of interest is asymmetric and unimodal. Figure 1(a) illustrates the case of an asymmetric distribution. If the distribution of the parameter is asymmetric but $\alpha/2$ and $1-\alpha/2$ quantiles of the distribution are taken as bounds of a confidence set (equal tailed), many unlikely (in a Bayesian or fiducial sense) parameter values would be included. Because the distribution is asymmetric, the density at the quantile of $\alpha/2$ is greater than that at $1-\alpha/2$ in Figure 1(a). In this case, we would have an unnecessarily wide confidence set. By contrast, we can make the set consist of most likely points with the HFD confidence set in Figure 1(a). Meanwhile, if the posterior distribution is bimodal and there are some values where densities between two local modes are very low, then the standard symmetric credible set might include unnecessary area. This bimodal case is shown in Figure 1(b). Again, the HFD confidence set is shorter, although it consists of two disjointed areas.

Third, the MCMC approach for constructing confidence sets can be applied to any type of structural break model, as long as we can specify the likelihood function for the model. This means that variables need not follow a normal distribution- for example, we might consider Poisson distributions, Student-t distributions, and so on. This approach can also be used for more complicated models such that each group of parameters is allowed to have breaks at different dates. For example, suppose there are two parameters of interest, such as when we regress U.S. real GDP growth on a constant and we allow it to have structural breaks in long-run growth and volatility, respectively. Deriving asymptotic distributions for this case is very complicated or infeasible. For example, in order to use Bai's (1997) asymptotic approach, error terms and explanatory variables should be covariance-stationary within each regime. If the volatility break date is different from the long-run growth break date, as might be hypothesized for postwar U.S. real GDP, the errors will not be covariance-stationary within one of the structural regimes for long-run growth, because the variance changes. By contrast, because it is easy to make a pdf conditional on different types of structural breaks at different break dates, it is straightforward to construct a likelihood function for the model and, therefore, use the FD approach.

2.2 Confidence Sets Based on Asymptotic Methods

In the time series econometrics literature, there have been many attempts to construct confidence intervals for structural change. For example, Bai (1997) derives the limiting distribution of a single break date in univariate linear regression models with normal errors and stochastic regressors and/or a disjointed time trend. Bai, Lumsdaine, and Stock (1998) consider multivariate models and Bai and Perron (2003) consider multiple structural changes. Recently, Elliott and Müller (2007) have pointed out that Bai's approach has low coverage rates relative to the nominal confidence level when changes in coefficients are small in magnitude. They propose an alternative approach that involves inverting a sequence of tests. Under the null hypothesis of each break date among candidate dates, a test is performed. Given a nominal level, if the test cannot reject the null, then the null hypothesis break date is included in the confidence set. Although Monte Carlo analysis by Elliott and Müller (2007) shows that their approach performs well in terms of producing coverage rates close to the nominal confidence level, it produces much longer confidence sets than Bai's approach.

2.3 Confidence Sets Based on Bootstrap Methods

In order to discuss bootstrap methods, we explicitly consider the case of only one structural break. This is done for ease of presentation only, as it is straightforward, at least conceptually, to consider multiple breaks. Also, throughout this paper, we consider parametric bootstrap methods, but it should be noted that there is nothing prohibiting the use of non-parametric or semi-parametric methods.

2.3.1 Confidence Sets Based on a "Bootstrap Percentile" Approach

Let τ_0 and $q(\cdot)$ denote the true break date and a quantile function for the difference between an estimator of the break date and the true break date, $\hat{\tau} - \tau_0$, respectively. Then,

$$\begin{aligned} 1 - \alpha &= Pr[q(\alpha/2) \leq \hat{\tau} - \tau_0 \leq q(1 - \alpha/2)] \\ &= Pr[-q(1 - \alpha/2) \leq \tau_0 - \hat{\tau} \leq -q(\alpha/2)] \\ &= Pr[-q(1 - \alpha/2) + \hat{\tau} \leq \tau_0 \leq -q(\alpha/2) + \hat{\tau}] \qquad (5) \end{aligned}$$

Because we do not know the true quantile function $q(\cdot)$ and it is unlikely to be fixed across different τ_0 in a finite sample, the quantile values are calculated based on a bootstrap under the null hypothesis of the estimated break date $\hat{\tau}$. In particular, the estimated break date $\hat{\tau}$ is regarded as the true break date in bootstrap samples.[13]

In the first step, we find a break date and values of model parameters with which the likelihood function is maximized in (1):

$$\{\hat{\tau}, \hat{\psi}\} = \underset{\tau \in \Pi, \psi \in \Psi}{argmax} \;\; L(\tau, \psi | y). \qquad (6)$$

In the second step, we generate B bootstrap samples based on the bootstrap data generating

[13] For a complete discussion of the percentile bootstrap approach, see Efron and Tibshirani (1994) Ch.13.

process (DGP) using $\{\hat{\tau}, \hat{\psi}\}$ from (6),

$$\{\mathbf{y^1}, ..., \mathbf{y^i}, ..., \mathbf{y^B}\}.$$

For each bootstrap sample, we detect the break date denoted by $\tau^{*(b)}$ from bth bootstrap sample by using MLE as in (6). That is, we store estimated break dates for the bootstrap samples,

$$\{\tau^{*(1)}, ..., \tau^{*(b)}, ..., \tau^{*(B)}\}, \tag{7}$$

sort break dates in (7) in ascending order and find $(\alpha/2)(B+1)$th and $(1-\alpha/2)(B+1)$th break dates:

$$\{\tau^*_{\alpha/2}, \tau^*_{1-\alpha/2}\}. \tag{8}$$

We replace the quantile values in (5) by the bootstrapped quantile values, $\{\tau^*_{\alpha/2} - \hat{\tau}, \tau^*_{1-\alpha/2} - \hat{\tau}\}$ from (8). Thus, the bootstrap percentile confidence set is

$$[-(\tau^*_{1-\alpha/2} - \hat{\tau}) + \hat{\tau}, \quad -(\tau^*_{\alpha/2} - \hat{\tau}) + \hat{\tau}]$$

$$= [2\hat{\tau} - \tau^*_{1-\alpha/2}, \quad 2\hat{\tau} - \tau^*_{\alpha/2}] \tag{9}$$

Note that the confidence set (9) is different from that in Efron (1979) because the quantile values are "flipped." The unflipped confidence interval could have poor coverage properties under an asymmetric distribution for the statistic of interest (see MacKinnon (2002) on this point). Of course, if the distribution of statistic in interest is symmetric (i.e. $q(1-\alpha/2) = -q(\alpha/2)$), then there is no difference between Efron's and the flipped percentile confidence set.

2.3.2 Confidence Sets Based on a "Bootstrap Standard Error" Approach

We follow the same initial steps as in the calculation of the percentile bootstrap confidence set and compute the standard error of bootstrapped break dates in (7). The standard error of bootstrapped break dates is

$$s.e.(\tau^*) = \sqrt{\frac{1}{B}\sum_{b=1}^{B}(\tau^{*(b)} - \overline{\tau}^*)^2}$$

where

$$\overline{\tau}^* = \frac{1}{B}\sum_{b=1}^{B}\tau^{*(b)}.$$

A standard error confidence set at 95% confidence level is

$$[\hat{\tau} - 1.96 \times s.e.(\tau^*), \hat{\tau} + 1.96 \times s.e.(\tau^*)]$$

Although there is no reason to believe the distribution of the break date estimator can be approximated by a normal distribution in a finite sample, we arbitrarily choose 1.96 for a 95% confidence level. Note that this confidence set is a contiguous and symmetric interval around the estimated break date. Its main benefit in practice is a relative computational simplicity. It also appears to work surprisingly well in some other settings (for example, see MacKinnon (2006)).

2.3.3 Confidence Sets Based on a "Bootstrap Inverted LR" Approach

To use a likelihood ratio statistic for constructing a confidence set, we need to know its distribution. However, the distribution is nonstandard in this setting of testing an estimated break date against a null date and depends on model parameters. Thus, we approximate the distribution based on bootstrapping. Given an estimated break date $\hat{\tau}$ from data, we compute the likelihood ratio value conditional on $\hat{\tau}$ and $\tau^{*(b)}$ from the bth bootstrap data set:

$$LR^{*(b)} = -2[logL(\hat{\psi}^{*(b)}, \hat{\tau}|\mathbf{y}^{(b)}) - logL(\psi^{*(b)}, \tau^{*(b)}|\mathbf{y}^{(b)})].$$

We store the log-likelihood ratio values from bootstraps,

$$\{LR^{*(1)}, ..., LR^{*(b)}, ..., LR^{*(B)}\}, \tag{10}$$

and sort them to determine the $\alpha(B+1)$th LR value, $LR^*(\tau^*_\alpha)$ as the critical value at $1-\alpha$ confidence level. Then, a bootstrapped inverted LR confidence set is

$$S_{LR} = \{\tau | LR(\tau) \leq LR^*(\tau^*_\alpha)\}$$

where $LR(\tau)$ is calculated from the data over each date, $\tau \in [\lambda T, (1-\lambda)T]$. Note that because the same critical value is applied to both tails of the inverted LR statistic, a bootstrapped inverted confidence set could be asymmetric and disjointed. This is directly analogous to the calculation of a confidence set based on the HFD concept. Here, it is a "highest-relative-likelihood" concept that is used for including break dates in the confidence set. Also, because the bootstrap inverted LR approach is directly based on the shape of the likelihood, the resulting confidence set should be similar to that of the FD approach, although the practical method of calculation is quite different.

2.3.4 Confidence Sets Based on a Bootstrap of Bai's Asymptotically Pivotal Statistic

For any bootstrap distribution to be exact, the distribution should not depend on any unknown parameters. At least, in order to approximate the asymptotic distribution in large samples, it should be asymptotically pivotal.[14] Bai (1997) derives an asymptotic distribution for constructing a confidence set. This distribution is asymptotically pivotal only if errors are second-order stationary and serially uncorrelated and explanatory variables are second-order stationary. Because Bai (1997)'s approach considers only one break of one regressor group in a linear regression, $K = 1$ for the number of groups and $M_1 = 1$ for the number of breaks in (1), and we can drop the subscript k for indicating group. Bai (1997) constructs the following statistic,

$$\frac{(\delta' Q \delta)^2}{\sigma^2}(\hat{\tau} - \tau_0), \tag{11}$$

where

$$Q = E[X_t X_t'].$$

[14] If it were pivotal in finite sample, a bootstrap would be an exact Monte Carlo simulation of a distribution.

Under standard conditions (see Bai (1997) for details), the statistic in (11) converges asymptotically in distribution to a non-standard distribution. We bootstrap the non-standard distribution and construct confidence intervals by using equal tailed quantile values in (11) as

$$[\hat{\tau} - \frac{\hat{\sigma}^2}{(\hat{\delta}'\hat{Q}\hat{\delta})^2} \times q(1-\alpha/2), \hat{\tau} - \frac{\hat{\sigma}^2}{(\hat{\delta}'\hat{Q}\hat{\delta})^2} \times q(\alpha/2)].$$

3 Monte Carlo Analysis

In order to investigate the performance of the various methods for constructing confidence sets discussed in the previous section, we perform various Monte Carlo experiments: a break in mean, a break in variance, and a break in drift, multiple breaks in mean and/or variance, and a break in mean for serially correlated data. Each experiment examines the coverage accuracy and expected length of the confidence sets based on different methods. The coverage rate is measured as the percentage frequency that confidence sets of different methods include the true break date across 5,000 simulations and its accuracy is based on comparing it to a specified nominal confidence level. The expected length of the confidence sets is measured by the average length across the Monte Carlo simulations. The nominal confidence level is 95% and the sample sizes are set to 40, 80, 160, 320, and 640, respectively. We use 199 bootstrap samples for each bootstrap method. For the FD approach, we employ the MH algorithm with a multivariate Student-t proposal distribution. The marginal fiducial distributions of parameters of interest are constructed using 2,000 draws from the joint fiducial distribution after a burn-in sample of 500 draws. The trimming value for possible break dates, λ, is 0.15. For comparison, we also consider constructing confidence sets using the asymptotic methods developed by Bai (1997) and Elliott and Müller (2007). Because their methods are based on some restrictive assumptions in terms of regressor distributions - e.g., covariance stationarity within each regime - and Elliott and Müller's approach is not usable for confidence sets of a change in variance, they are included in the Monte Carlo experiments only when applicable. Readers are referred to the original articles for the practical details of implementing these asymptotic methods. The results of the Monte Carlo experiments are presented in the next subsections.

3.1 A Structural Break in Mean

For a break in mean, the model in (1) can be simplified as follows:

$$y_t = \mu_0 + \mu_1 \mathbf{1}[t > rT] + e_t, \quad e_t \sim i.i.d.\mathcal{N}(0, \sigma^2), \tag{12}$$

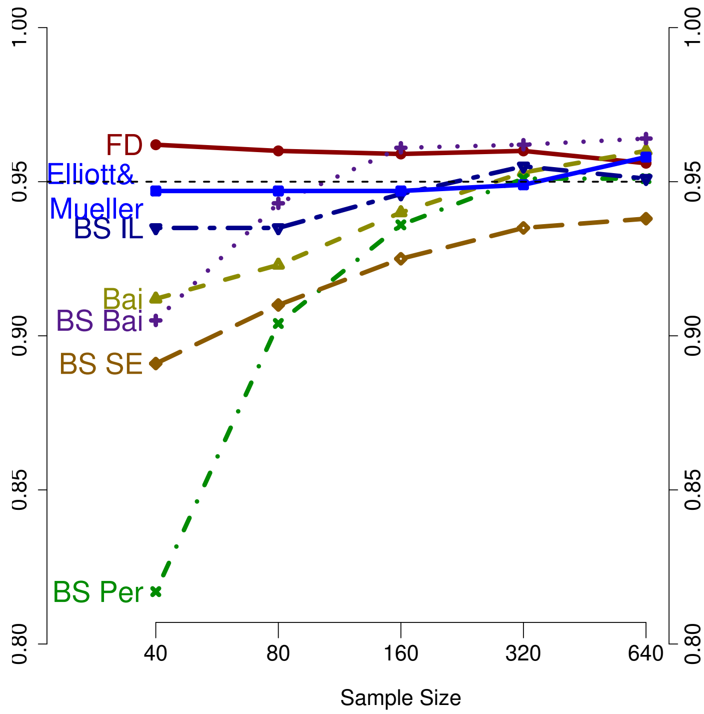
where $\mathbf{1}[\cdot]$ is an indicator function. We set $\mu_0 = 1$, $\mu_1 = -0.5$, and $\sigma = 0.5$. For the experiment, the true break point fraction r is 0.5. The results are presented in Figures 2(a) and 2(b).

The results show that confidence sets for the Elliott and Müller approach, the bootstrapped inverted LR approach, and the FD approach have good coverage accuracy across different sample sizes. However, the expected length of the Elliott and Müller confidence set is greater than the expected length for the FD or bootstrap inverted LR approaches. Also, note that the ratio of lengths of the Elliott and Müller confidence sets to those of the likelihood based methods gets larger as the sample size increases. Strikingly, when the sample size is 640, the Elliott and Müller confidence set is more than three times as long as that of likelihood-based confidence sets. Bootstrap percentile and standard error confidence sets have good coverage accuracy for large sample sizes such as 320 and 640, but they are also longer than the FD and bootstrap inverted LR confidence sets. Meanwhile, the bootstrap percentile and bootstrap standard error confidence sets do not have good coverage accuracy for smaller sample sizes. Bai's confidence intervals achieve relatively low coverage rates and longer lengths, whereas the approach based on a bootstrap of Bai's statistic brings out much longer lengths without much improvement in terms of coverage rates.

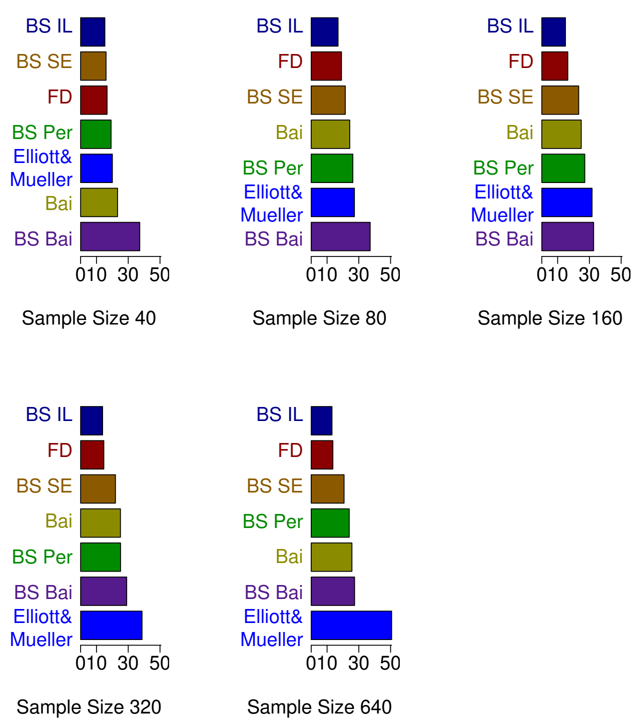
3.2 A Structural Break in Variance

For a break in the variance of the error term, the model can be written simply as

$$y_t = \mu + e_t, \quad e_t \sim i.i.d.\mathcal{N}(0, \sigma_t^2) \tag{13}$$



(a) Coverage rates of confidence sets



(b) Average lengths of confidence sets

Figure 2: Coverage rates and average lengths of confidence sets for the timing of a break in mean

where

$$\sigma_t^2 = \begin{cases} \sigma_0^2 & \text{if } 1 \le t \le rT \\ \sigma_1^2 & \text{if } rT < t \le T, \end{cases}$$

We set $\mu = 1$, $\sigma_0 = 1$, and $\sigma_1 = 0.5$. The true break point fraction r is 0.5. The results are presented in Figures 3(a) and 3(b).

Since an asymptotic distribution for a structural break in variance is not rigorously developed in the literature, we employ Bai's method and construct confidence sets of break dates in mean of the absolute values of errors from the regression model in (13) as in Stock and Watson (2002) and Sensier and van Dijk (2004):

$$\sqrt{\frac{\pi}{2}}|e_t| = \sigma_0(1 - \mathbf{1}[t > rT]) + \sigma_1\mathbf{1}[t > rT] + \epsilon_t.$$

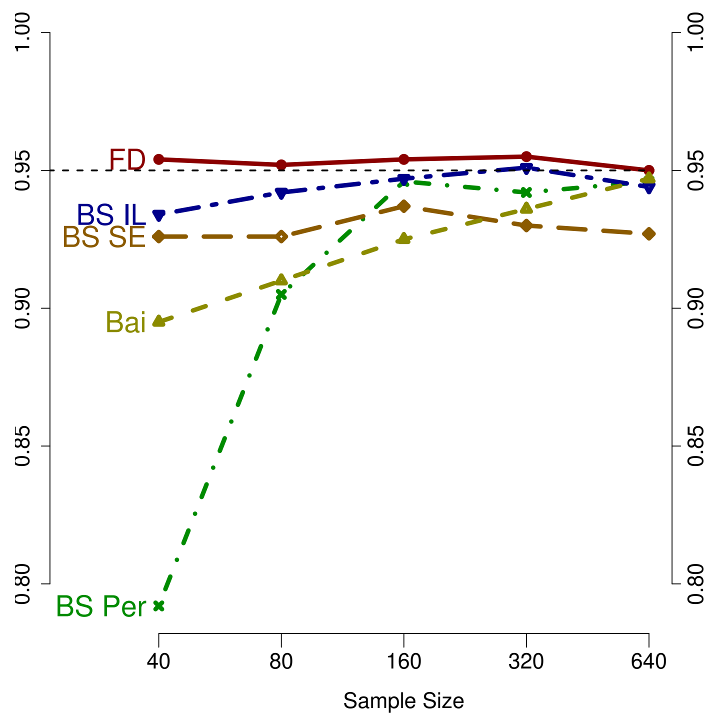
The coverage rates of FD confidence sets are close to 95% over different sample sizes. The bootstrap percentile and bootstrap inverted LR confidence sets have good coverage accuracy for sample sizes larger than 160. However, the bootstrap percentile confidence sets become much longer than those of the bootstrap inverted LR and FD methods as the sample size increases. Meanwhile, the bootstrap standard error approach attains coverage rates well below 95%, with relatively longer confidence sets compared to the FD and bootstrap inverted LR approaches. The coverage rates of confidence sets based on Bai's approach get close to 95% as the sample size increases, but the lengths are greater than those of confidence sets based on other approaches, including being approximately twice as long as the confidence sets based on the FD and bootstrap inverted LR approaches.

3.3 A Structural Break in Drift

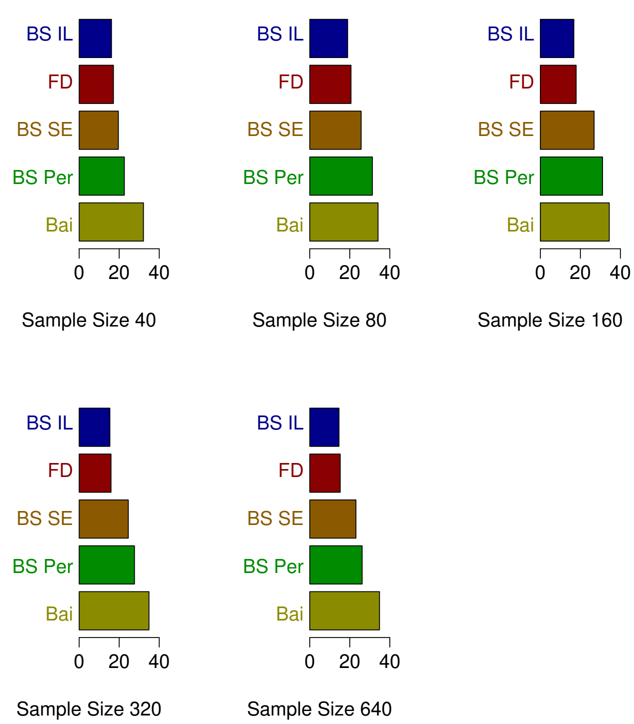
For a break in drift in an otherwise trend stationary process, the model can be written as

$$y_t = \alpha + \beta_0(t/T) + \beta_1(t/T)\mathbf{1}[t > rT] + e_t, \quad e_t \sim i.i.d.\mathcal{N}(0, \sigma^2)$$

We set $\alpha = 1$, $\beta_0 = 2$, $\beta_1 = -0.5$, and $\sigma = 0.3$. The true break point fraction r is 0.5. The results are presented in Figures 4(a) and 4(b).

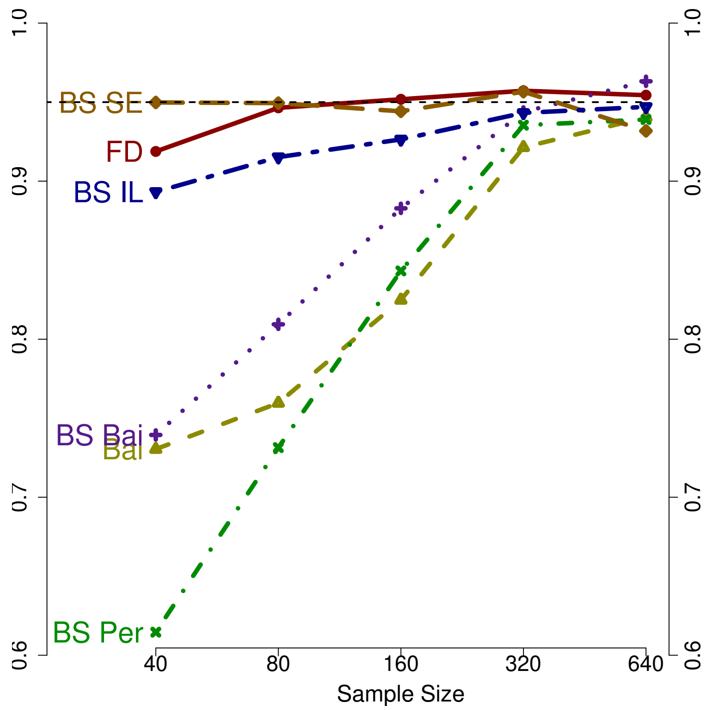


(a) Coverage rates of confidence sets

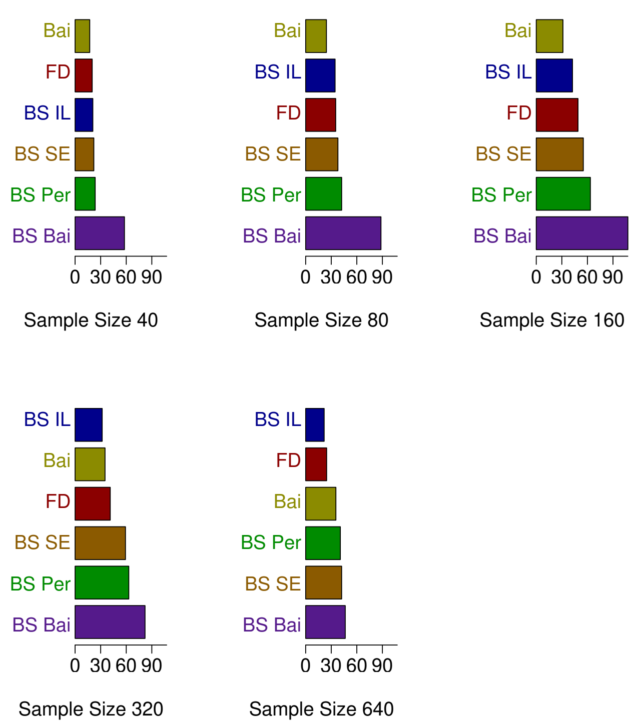


(b) Average lengths of confidence sets

Figure 3: Coverage rates and average lengths of confidence sets for the timing of a break in variance



(a) Coverage rates of confidence sets



(b) Average lengths of confidence sets

Figure 4: Coverage rates and average lengths of confidence sets for the timing of a break in drift

The bootstrap standard error approach and the likelihood-based approaches work well with good coverage accuracy. The FD confidence sets have relatively short lengths for each sample size. The confidence sets based on Bai's asymptotic approach, the bootstrap of Bai's statistic, and the bootstrap percentile approach have coverage rates far below the nominal 95% level, with levels around 60% to 80% in sample sizes of 40 to 160. The confidence sets based on the bootstrap of Bai's statistic have relatively large expected lengths for each sample size. The bootstrap standard error approach achieves quite stable coverage rates close to 95%, but it also produces confidence sets with relatively large expected lengths. The bootstrap inverted LR approach has just below 95% coverage rate with relatively short confidence sets.

3.4 Multiple Structural Breaks

For multiple structural breaks, we consider two cases: (i) two breaks in mean and (ii) one break in mean and one break in variance, possibly at different break points. We examine the performance of the FD approach only, as the Monte Carlo simulations for bootstrap methods are computationally infeasible and Bai's method has already been shown to perform poorly in the case of only one break. In terms of implementation, the two break dates are sampled from different blocks of the MH algorithm.

For case (i), the true DGP is as follows:

$$y_t = \mu_0 + \mu_1 \mathbf{1}[t > r_1 T] + \mu_2 \mathbf{1}[t > r_2 T] + e_t, \quad e_t \sim i.i.d.\mathcal{N}(0, \sigma^2),$$

We set $\mu_0 = 1$, $\mu_1 = -0.5$, $\mu_2 = 0.5$, and $\sigma = 0.5$. For the experiment, the true breakpoint fractions r_1 and r_2 are 0.3 and 0.7, respectively. The results of Monte Carlo simulation are summarized in Table 1. In finite samples, the true break dates are over-covered since the minimum distance between two break dates increases the relative probabilities of allowable break dates from what they would otherwise be. However, this effect disappears and the coverage rates converge to the nominal 95% when the sample size is larger.

For case (ii), the magnitudes of breaks in mean and variance are the same as in the case of one break in mean or variance. The true DGP is as follows:

Table 1: Coverage rates and average lengths of confidence sets for the timing of two structural breaks in mean from Fiducial distribution approach based on Monte Carlo simulations

Fiducial Dist.	40	80	160	320	640
First Break	0.981	0.971	0.966	0.956	0.944
	[16.09]	[28.39]	[45.46]	[59.37]	[65.53]
Second Break	0.984	0.973	0.958	0.949	0.947
	[16.03]	[28.26]	[45.84]	[61.14]	[64.70]

Notes: The coverage rates for confidence sets are constructed using a nominal 95% confidence level, with average lengths of the confidence sets reported in square brackets.

Table 2: Coverage rates and average lengths of confidence sets for the timing of one break in mean and one break in variance from Fiducial distribution approach based on Monte Carlo simulations

Fiducial Dist.		Sample size 40		
Breakpoint fraction	r_1	0.3	0.4	0.5
	r_2	0.7	0.6	0.5
Break in mean		0.964	0.955	0.948
		[22.82]	[21.93]	[21.10]
Break in variance		0.941	0.947	0.961
		[20.51]	[18.88]	[17.92]

Notes: The coverage rates for confidence sets are constructed using a nominal 95% confidence level, with average lengths of the confidence sets reported in square brackets.

$$y_t = \mu_0 + \mu_1 \mathbf{1}[t > r_1 T] + e_t, \quad e_t \sim i.i.d.\mathcal{N}(0, \sigma_t^2)$$

where

$$\sigma_t^2 = \begin{cases} \sigma_0^2 & \text{if } 1 \leq t \leq r_2 T \\ \sigma_1^2 & \text{if } r_2 T < t \leq T, \end{cases}$$

We set $\mu_0 = 1$, $\mu_1 = -0.5$, $\sigma_0 = 1$, and $\sigma_1 = 0.5$. We perform three experiments by varying true breakpoint fractions: a) 0.3 and 0.7, b) 0.4 and 0.6, and c) 0.5 and 0.5.[15] The first break fraction in each experiment is for break in mean and the second is for break in variance. We consider the sample size 40 only since the results of Monte Carlo simulations are similar to those from one break in mean or variance. The results are summarized in Table 2. The coverage rates are close to the nominal level 95%. Thus, when using the FD approach, the existence of a break in mean does not materially affect the inference about a break in variance and vice versa.

3.5 Robustness

Next, we examine whether the performance of FD approach is robust to the location of break date or the presence of serially correlation.

First, we vary the location of true breakpoint as a fraction of the sample from $r = 0.3$ to 0.5 in (12) for the case of a break in mean.[16] The sample size is 40. Figure 5 presents the results. The coverage rates of confidence sets are robust to locations of true break dates. FD, bootstrap inverted LR, and Elliott and Müller's approaches result in coverage rates close to 95%.

Second, we generate serially correlated data with a structural break in mean. The true DGP is as follows:

$$y_t = \mu_0 + \mu_1 \mathbf{1}[t > rT] + \rho y_{t-1} + e_t, \quad e_t \sim i.i.d.\mathcal{N}(0, \sigma^2),$$

[15] Even when the true break dates are the same, we think of the breaks as being independent in the sense that they can be different in our estimation. Imposing a common break date for the two parameters would, of course, increase power and reduce the expected length of the confidence set in the case of c)

[16] In the cases of $r > 0.5$, the Monte Carlo simulation results should be symmetric to the results when $r < 0.5$. Thus, we consider $r \leq 0.5$ only.

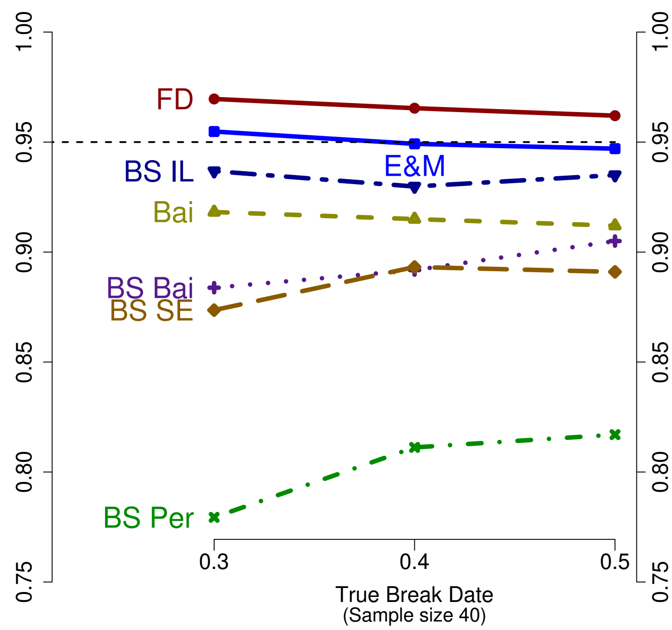


Figure 5: Coverage rates of confidence sets for the different timing of a break in mean

where $\mu_0 = 1$, $\mu_1 = -0.5$, $\sigma = 0.5$, $r = 0.5$ and $\rho = 0.3$. Figure 6 presents the results. The coverage rates based on all the methods converge to 95% as the sample size increases. However, in finite samples such as 40 and 80, only the FD and Elliott and Müller confidence sets attain 95% coverage rates.

3.6 Summary of Monte Carlo Results

The main findings for the Monte Carlo analysis can be summarized as follows.

First, in most cases, likelihood-based confidence sets - i.e., those based on the FD and bootstrap inverted LR approaches - produce the most accurate coverage rates and shortest expected lengths. Furthermore, the ratios of the lengths of confidence sets based on other approaches to lengths of the likelihood-based confidence sets become larger as the sample size increases. Thus, for confidence set estimators $S(Y)$ given a nominal confidence level $1-\alpha$, suppose we define a loss function which increases in the length of confidence set and the absolute value of difference between the expected coverage rate and the nominal confidence

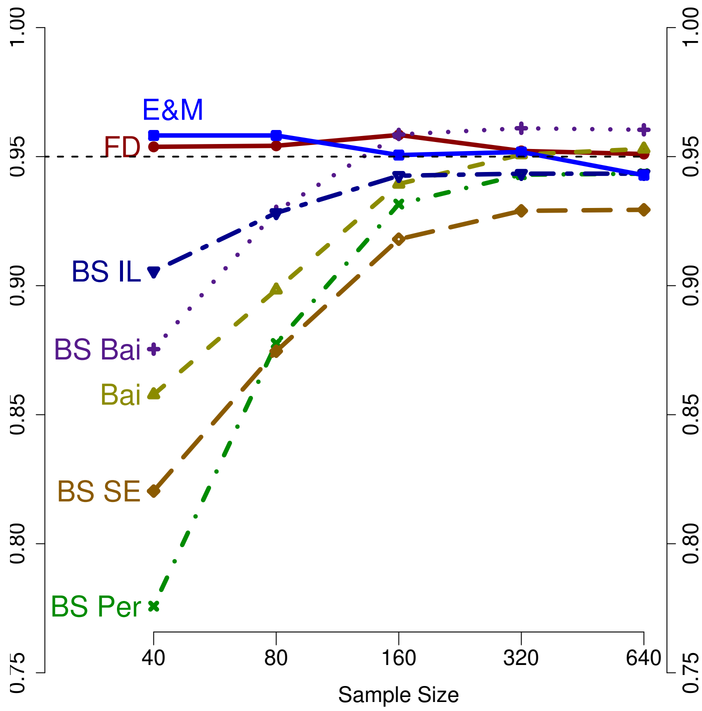


Figure 6: Coverage rates of confidence sets for a break in mean from serially correlated data

level:

$$\mathcal{L}(S(Y); 1-\alpha) = \mathcal{L}(d(S(Y)), |Pr[\tau_0 \in S(Y)] - (1-\alpha)|)$$

where $d(S(Y))$ is the expected length of confidence set, $\mathcal{L}(0,0) = 0$, and $inf \mathcal{L} = 0$. Then, given this loss function, the likelihood-based confidence set estimators will be admissible in the sense that their coverage rates converge to the nominal confidence level and the lengths of their confidence sets are shorter than those of other methods as the sample size increases. This result is consistent with Siegmund (1988), who analytically calculates and compares expected lengths of confidence sets for various methods. For analytical tractability, he assumes that the distributions before and after the break are $\mathcal{N}(0,1)$ and $\mathcal{N}(\delta,1)$ and the true parameter values are assumed to be known. In such a case, the expected lengths of confidence sets from likelihood-based approaches are asymptotically equal and they are half of the length of confidence sets based on a pivotal quantity $\hat{\tau} - \tau$ such as Bai and bootstrap percentile methods in the present paper. Our Monte Calro simulation results with the unknown parameters confirm his results in the more realistic setting of unknown parameter values.

Second, the FD approach constructs confidence sets based on marginal distributions of break dates. The fact that the coverage is so accurate directly implies that, in some way, the marginal fiducial distributions provide good approximations of finite-sample distributions that are close to being pivotal. Part of this accuracy could be because the nuisance parameters other than breakpoint parameters of our interest are integrated out of the likelihood (for example, see Andrews and Ploberger (1994), Berger, Liseo, and Wolpert (1999), and Elliott and Müller (2007)). However, the deeper question is what is it about the marginal fiducial distributions that is essentially pivotal. Perhaps surprisingly, additional Monte Carlo analysis reveals that various features of the marginal fiducial distributions are quite sensitive to different parameter values. For example, we find that the relative height of the mode of the fiducial distribution to the cut-off, $k(1-\alpha)$, for including break dates in the confidence set (a feature of the fiducial distribution that can be thought of as analogous to an LR statistic), has a distribution across repeated samples that is highly dependent on the size of the structural break. Instead, what appears to be more stable and explains the success of the FD approach is the relative value of the cut-off $k(1-\alpha)$ to the fiducial probability of the true break date $Pr[\tau = \tau_0]$. In particular, the distribution of $k(1-\alpha) - Pr[\tau = \tau_0]$ appears to always have an $(1-\alpha)$ quantile equal to zero. To the extent that this quantile is zero and higher quantiles are strictly positive, the FD confidence sets will have exactly correct coverage. Of course, if the $(1-\alpha)$ quantile were positive, the FD confidence sets would under-cover because the cut-off $k(1-\alpha)$ would be above the $Pr[\tau = \tau_0]$ in more than α samples. On the other hand, if the $(1-\alpha)$ quantile is zero, but some higher quantiles are also zero, then the FD confidence sets would over-cover. Figure 7 reports Monte Carlo results for the distributions of the difference $k(1-\alpha) - Pr[\tau = \tau_0]$ for different size structural breaks in mean and for $1-\alpha = 0.95, 0.80$, and 0.68. We find that, regardless of the size of the break, the $(1-\alpha)$ quantile is always zero. When structural breaks are large, some higher quantiles are also zero, so there is a tendency to over-cover in such cases. However, over-coverage in the case of very large structural changes is not particularly troubling given that the only way to attain correct coverage in such a setting would be to have empty confidence sets in α samples.

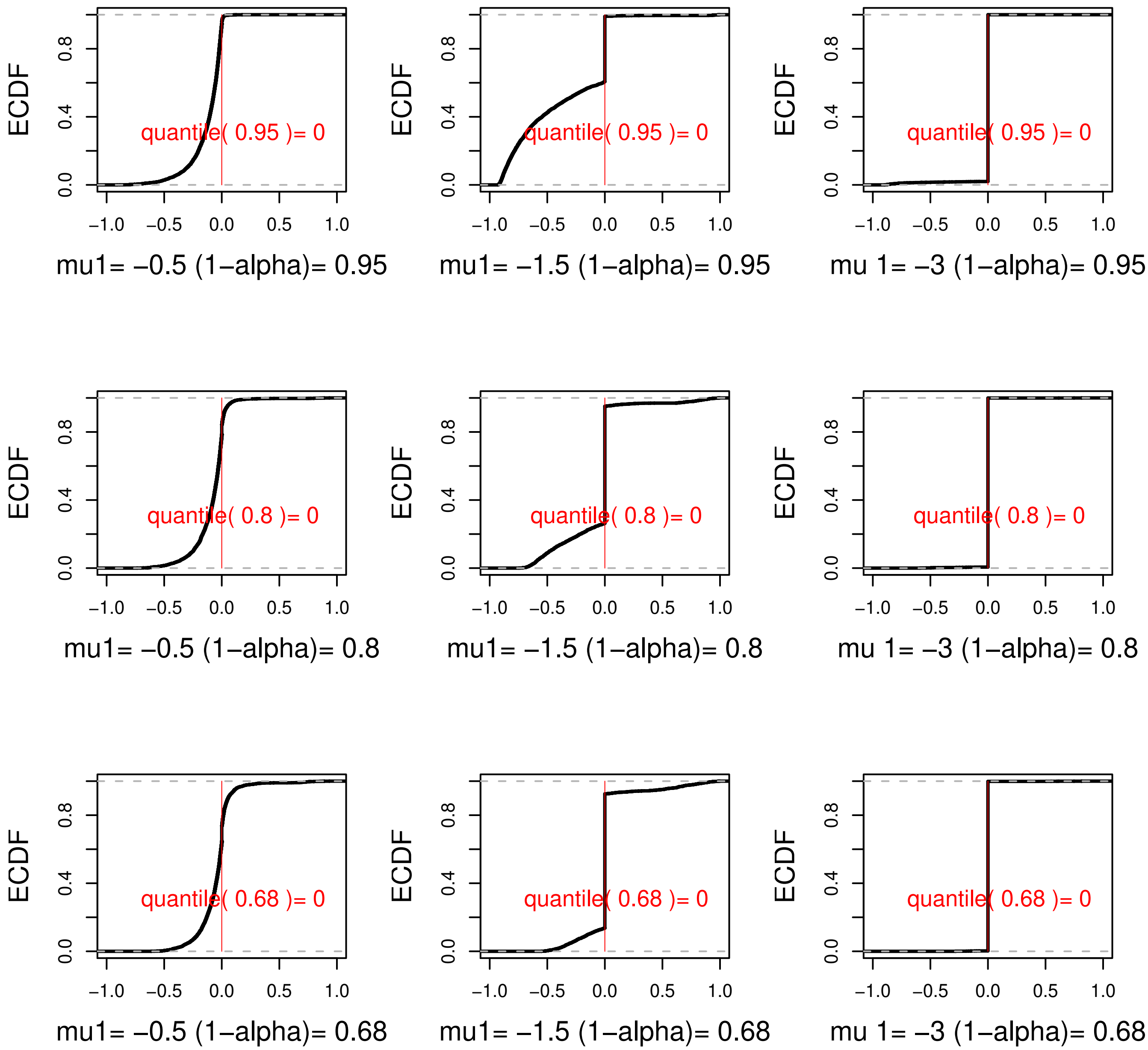


Figure 7: Empirical cumulative distribution function of $k(1-\alpha) - Pr[\tau = \tau_0]$

Third, some bootstrap methods perform very well in a sense of exact coverage in large sample sizes. However, when the sample sizes are small or the estimated break dates for the actual data are closer to the beginning or end of sample, the flipping in the percentile bootstrap method produces very low coverage rates. This is because the parameter space of interest (i.e., the possible break dates) is limited to 100 $\times(1-2\lambda)\%$ centered dates by trimming as $\Pi = [\lambda T, (1-\lambda)T]$ and some dates among potential break dates might not be considered in the procedure to construct confidence set. In particular, suppose the true break date is close to the first possible break date considered, $\lambda T \in \Pi$. The number of periods between the true break date and the first possible break date is very small and the number of periods between the true break date and the ending break date $(1-\lambda)T$ is large. Then, the two subsample periods before and after the estimated break date will be asymmetric. By flipping, break dates that are estimated in the first subsample period area from bootstrapped data sets will be used for bootstrapped distribution in the second subsample period, so that the first subsample period cannot fully cover the second sub-interval, which is longer than the first-sample period. By contrast, the second subsample period is flipped to cover the first sub-interval, but some portion of the second subsample period might be out of the bounds for possible break dates in Π. Thus, the bootstrap percentile methods may produce very low coverage rates.[17] When the sample size is very small, this problem arises even for the structural change at the middle point of the sample, $r = 0.5$.

Fourth, because Elliott and Müller (2007) use sequential tests to attain an exact coverage rate, their test statistic is constructed in order to focus on producing the exact size of the test rather than having high power to reject false break dates. Thus, their confidence sets may be unnecessarily long.[18] By contrast, since the FD approach relies on break date distributions conditional on the actual data, the confidence set for the FD method can be constructed

[17] To illustrate, suppose the sample size is 40, the estimated break ($\hat{\tau}$) is 10, and we have a 15% trimming rule for candidate break dates. Because we exclude the first 6 and last 6 points based on trimming, Π=[7,34]. Suppose the the lower and upper bootstrap quantiles for $\tau^{*b} - \hat{\tau}$ are -3 and 24, respectively. The lower quantile is determined by the trimming rule. Then, the confidence set based on flipped quantiles would be [-14,13]. Thus, the trimming implicitly means that [14,34] will also be excluded from Π in practice and the coverage of the confidence sets based on flipping will be far below the nominal confidence level.

[18] In methods of inverting sequential tests, having a test with the exact size equal to the nominal significance level means having a confidence set estimator with the exact coverage rate equal to the nominal confidence level.

to be as short as possible, while maintaining an accurate coverage rate. In addition, if the magnitude of the parameter change is very large, Elliott and Müller's test easily rejects the false hypotheses of break dates. Thus, in order to maintain correct coverage, all hypotheses of break dates including the true break date will be rejected in α samples. That is, their method can produces empty confidence sets, which are not particularly helpful when found or reported in a given empirical application.

4 Application to Postwar U.S. Real GDP

4.1 Model Specification

We estimate break dates and model parameters by maximum likelihood estimation and apply the various methods presented in section 2 in order to construct confidence sets for structural breaks in postwar U.S. real GDP. In terms of our general model specification, we consider the two possibilities that real GDP follows i) a nonstationary process with a unit root and ii) a trend stationary process.

Under the assumption of a unit root, the model for M growth breaks and N variance breaks is given as follows:

$$\Delta y_t = c_0 + \sum_{m=1}^{M} c_m D(T_m) + \sum_{j=1}^{p} \phi_j \Delta y_{t-j} + e_t,\ e_t \sim i.i.d.\mathcal{N}(0, \sigma_t^2) \tag{14}$$

where

$$\sigma_t^2 = \begin{cases} \sigma_0^2 & \text{if } 1 \leq t \leq \tau_{\sigma_1^2} \\ \sigma_1^2 & \text{if } \tau_{\sigma_1^2} < t \leq \tau_{\sigma_2^2} \\ \vdots & \\ \sigma_N^2 & \text{if } \tau_{\sigma_N^2} < t \leq T, \end{cases}$$

and $D(T_m) = 1$ if $t > \tau_{c_m}$ and 0 otherwise and τ_{c_m} is mth break point for constant. The lagged first differences account for serial correlation. We determine the number of first difference terms, p, by the Schwarz Information Criterion (SIC).

We consider sequential asymptotic tests to determine the number and types of structural breaks.[19] The testing proceeds as follows: We begin with the null hypothesis of no structural breaks and the alternative of one break in constant or variance. If we can reject the null hypothesis of no break, we assume one more break in the constant or variance and the alternative hypothesis in the previous test becomes the null hypothesis in the new test. The estimates of break dates in the alternative hypothesis from the previous test are maintained and given in the null and the alternative for the new test and an additional break at an unknown date is considered in the alternative. We keep performing these tests sequentially until a null hypothesis cannot be rejected. In order to determine the number of breaks and model specification, we consider LR statistics based on maximum likelihood estimates under the null and alternative.[20] The LR statistic is calculated from the log-likelihood values for the regression in (14). For example, suppose the null hypothesis regression has M breaks for constant and N breaks for variance and the alternative has $M+1$ breaks and N breaks, respectively. Then, given the number of break dates for the constant and variance parameters, the particular break dates are chosen to make the LR test statistic from (14) as large as possible and we reject the null hypothesis if this supLR test statistic is larger than a specified critical value. Thus,

$$
\begin{aligned}
& LR(\tau_{c_0}, \ldots, \tau_{c_{M+1}}, \tau_{\sigma_0^2}, \ldots, \tau_{\sigma_N^2}) \\
= -2 \times [\quad & logL(c_0, \ldots, c_M, \sigma_0^2, \ldots, \sigma_N^2, \phi_1, ..., \phi_p | y) \\
& -logL(c_0, \ldots, c_{M+1}, \sigma_0^2, \ldots, \sigma_N^2, \phi_1, ..., \phi_p | y) \quad]
\end{aligned}
$$

$$
\begin{aligned}
& \{\hat{\tau}_{c_1}, ..., \hat{\tau}_{c_{M+1}}, \hat{\tau}_{\sigma_1^2}, \ldots, \hat{\tau}_{\sigma_N^2}\} \\
= & \underset{\tau_{c_1}, ..., \tau_{c_{M+1}} \in \Pi, \quad \tau_{\sigma_1^2}, ..., \tau_{\sigma_N^2} \in \Pi}{\arg \mathrm{Max}} LR(\tau_{c_1}, \ldots, \tau_{c_{M+1}}, \tau_{\sigma_1^2}, \ldots, \tau_{\sigma_N^2}) \\
& \text{supLR statistic} = \text{val Max} \;\; LR(\hat{\tau}_{c_1}, \ldots, \hat{\tau}_{c_{M+1}}, \hat{\tau}_{\sigma_1^2}, \ldots, \hat{\tau}_{\sigma_N^2})
\end{aligned}
$$

[19] In principle, we could consider Bayesian model selection to determine the number and types of structural breaks. Levin and Piger (2007) use Monte Carlo analysis to show that this approach has good frequentist properties in finite samples. However, for simplicity, we focus on new techniques for making inferences about the timing of structural breaks, rather than inferences about their number and type.

[20] Bai (1999) proposed an LR-type test for multiple structural breaks. However, while his test can be applied to partial structural changes with some parameters changing and others remaining constant, all of the parameters that change are restricted to have same break dates. By contrast, our approach allows for different parameters to break at different break dates.

Andrews (1993, 2003) provides tables of critical values for a supLR test of an unknown break date. We use these critical values for our test, although we note that our test is generally more complicated than the simple case of a one-time structural break for which the critical values were derived. Once a structural break model is selected, we re-estimate break dates simultaneously, rather than conditioning on some breaks as was done when testing a null break versus alternatives with additional breaks. Thus, the estimated break dates could be different from those in the tests for the number of breaks. We adopt this kind of two-step procedure in order to be able to use Andrews' critical values in the first step, as the critical values would only be strictly correct if we knew and imposed the break date under the null when estimating the model under the null and alternative.

Under the assumption of trend stationarity, our model includes a constant, a drift, a lagged level, and p lagged first differences. The model for M breaks in drift and N breaks in variance is as follows.

$$y_t = \alpha + \beta_0(t/T) + \sum_{m=1}^{M} \beta_i DT_m + \rho y_{t-1} + \sum_{j=1}^{p} \phi_j \Delta y_{t-j} + e_t,\ e_t \sim i.i.d.\mathcal{N}(0, \sigma_t^2) \tag{15}$$

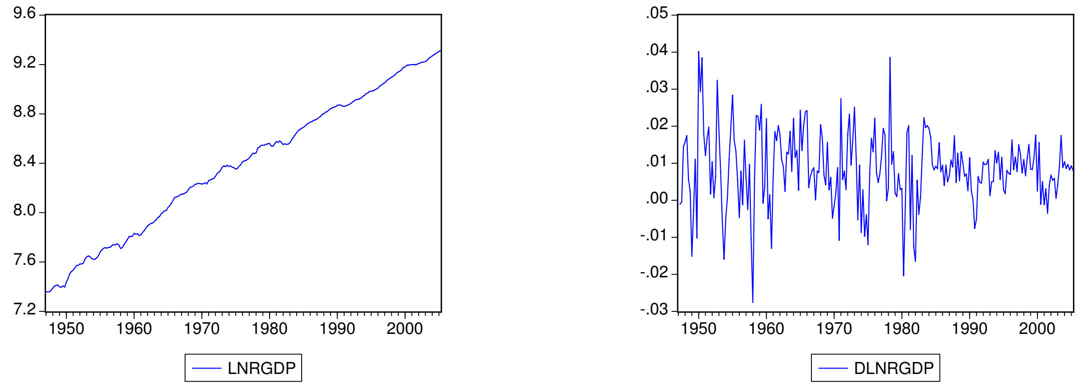
where $DT_m = (t - \tau_{\beta_m})/T$ if $t > \tau_{\beta_m}$ and 0 otherwise, and τ_{β_m} and $\tau_{\sigma_n^2}$ are mth and nth break points for the drift and for the variance, respectively. Including DT_m allows for two segmented trends to be connected.[21] For the model (15), the number of extra first difference regressors, p, is again determined by SIC. The remaining steps (finding break dates and calculating the test statistic) are also exactly the same as in the previous case.

4.2 Empirical Results for Postwar U.S. quarterly Real GDP

4.2.1 Estimating Break Dates

To estimate structural breaks, we use postwar quarterly U.S. real GDP data from 1947:Q1 to 2005:Q2, corresponding to 234 observations. Thus, based on a symmetric 15% trimming

[21] Note that the kinked (or connected) time trend specification in our application is different from the disconnected one considered in Bai (1997). For the reference for a kinked time trend break, see Perron (1989) and Zivot and Andrews (1992).



(a) Logarithm of postwar U.S. real GDP

(b) Annualized growth rate of postwar U.S. real GDP (based on log differences)

Figure 8: Postwar U.S. real GDP

value, the candidate break dates are from 1956:Q1 to 1996:Q2. The data are seasonally adjusted and based on Billions of Chained 2000 Dollars from U.S. Department of Commerce: Bureau of Economic Analysis.

Figure 8(a) depicts the natural logarithm of postwar U.S. real GDP. It may be hard to identify a drift or variance break from a casual glance. Figure 8(b) shows growth rates of real GDP. In this case, the volatility decline seems fairly apparent, even if the exact timing of the break is unclear.

First, we consider the assumption that real GDP has a unit root. Based on SIC, we find that one lag of first differences is sufficient to capture any serial correlation in growth rates. Detailed results are given in Table 3. Based on our testing, we find support for a model of real GDP growth with one variance break, which is estimated to occur in 1984:Q1. The estimated ratio of standard deviations of shocks after the break compared to before the break is 0.43. This supLR test statistic against the null of no structural breaks is 62.20. Based on Andrews (2003)'s test statistics for one unknown break of a parameter over the middle 70% of a sample period, the critical value at 5% significance level is 8.68. Meanwhile, the null hypothesis of no break in growth model cannot be rejected against the alternative hypothesis of the break in mean at a 5% significance level. Likewise, the null hypothesis of one break in variance cannot be rejected against the alternatives of an additional break in

Table 3: Structural break estimates for the growth rate of U.S. real GDP

	Break Dates	c_0	c_1	Growth rate	ϕ	σ_0^2	σ_1^2	Vol.Ratio(SD)	lnL
No Break		2.26		3.38	0.33	14.00			-635.34
		(0.32)			(0.06)	(1.30)			
One Br. in Con.	1966:Q1	2.81	-0.77	4.14/3.01	0.32	13.87			-634.27
		(0.49)	(0.13)		(0.06)	(1.29)			
One Br. in Var.	1984:Q1	2.19		3.21	0.32	20.03	3.63	0.43	-604.23
		(0.27)			(0.06)	(2.34)	(0.56)		
One Br. in Con. &	1987:Q1	2.58	-0.64	3.89/2.93	0.34	20.31	3.66	0.42	-603.14
One Br. in Var.	1983:Q2	(0.37)	(0.10)		(0.06)	(2.40)	(0.55)		

the mean or an additional break in variance.[22]

Second, we consider the assumption that real GDP is trend stationary. In this case, based on SIC, one lagged difference is, again, necessary to capture serial correlation. The results for the trend stationary model are reported in Table 4. The estimated break date for a one-time break in drift is 1968:Q2 and its test statistic is 9.37. Thus, we can reject the null hypothesis of no structural change at the 5% level. The estimated variance break is 1983:Q2 and the test statistic is 62.06. Thus, the null hypothesis of no breaks is also rejected for the alternative of one break of variance. Because the null hypothesis of no breaks is rejected in favor of two different alternatives, we proceed to test for an additional break in drift or variance, with each previous alternative serving as a new null. The null of one break in variance cannot be rejected against the alternative of an additional break in variance, even without a restriction on the date of the null break. In this case, the supLR statistic is 4.26. However, the null of one break in variance can be rejected against the alternative of additional break in drift given the variance break date of 1983:Q2, with a supLR statistic of 10.11. Without the restriction of the variance break date, the LR statistic would have been 11.88 with the estimated drift break date of 1969:Q1 and variance break date of 1982:Q4.

It should be noted that the estimated drift break date of 1969:Q1 based on simultaneous estimation differs from the drift break date of 1968:Q2 based on restricted estimation. Meanwhile, the coefficient of the drift decreases by about 22% (0.37 to 0.29) from before and after the break and the estimated ratio of standard deviations of shocks after the break

[22] Consistent with the fact that the model with one break in mean and one break in variance is not chosen by our test procedure, the confidence set for a break in mean for such an alternative model covers about 90% of the parameter space of possible break dates. This wide confidence set reflects the lack of empirical support for a sudden break in the long-run growth rate.

Table 4: Structural break estimates for the level of U.S. real GDP

	Break Dates	β_0	β_1	β_2	ρ	ϕ	σ_0^2	σ_1^2	σ_3^2	Vol.Ratio(SD)	lnL
No Break		0.14			0.96	0.35	13.47				-630.88
		(0.07)			(0.02)	(0.06)					
One Br. in Drift	1968:Q2	0.36	-0.07		0.91	0.37	12.94				-626.20
		(0.09)	(0.02)		(0.02)	(0.06)					
One Br. in Var.	1983:Q2	0.14			0.96	0.36	19.54	3.62		0.43	-599.85
		(0.05)			(0.01)	(0.06)					
One Br. in Drift &	1969:Q1	0.37	-0.08		0.90	0.37	19.00	3.45		0.43	-593.91
One Br. in Var.	1982:Q4	(0.08)	(0.02)		(0.02)	(0.06)					
Two Br.s in Drift	1958:Q1	0.34	0.14	-0.15	0.89	0.37	12.71				-624.08
	1966:Q4	(0.09)	(0.05)	(0.04)	(0.02)	(0.06)					
Two Br.s in Var.	1960:Q4	0.14			0.96	0.35	26.29	15.69	3.59	0.77	-597.72
	1982:Q2	(0.05)			(0.01)	(0.06)				0.48	

compared to before the break is 0.43, which is exactly same as in the unit root case. Also, the null of one break in drift cannot be rejected against the alternative of an additional break in drift, with a supLR statistic 4.24. Thus, under trend stationarity, we choose a model with one break in drift, estimated to be in 1969:Q1, and one break in variance, estimated to be in 1982:Q4.

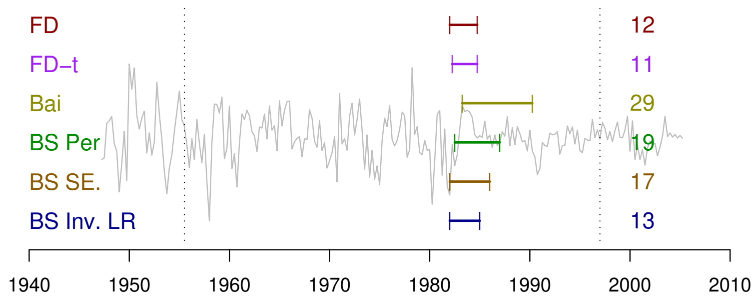
To sum up, for postwar real GDP data, there is strong evidence of structural break in volatility in the early 1980s, but more mixed evidence for a structural break in long-run growth that may have occurred in the late 1960s. The timing of the possible break in long-run growth in the late 1960s is, perhaps, surprising given the emphasis on a possible break in 1973 in the unit root literature (e.g., see Perron (1989) and Zivot and Andrews (1992)). However, it is consistent with the timing found by Bai, Lumsdaine, and Stock (1998) when they estimate break dates in a variety of multivariate models using real per capita demand components related to consumption and investment. The reason for different estimates of break dates is that, in the unit root testing literature, a break date is chosen based on maximizing the evidence against a unit root, while here it is chosen based on maximizing the likelihood function with respect to the break date.

4.2.2 Confidence Sets for Break Dates in Postwar U.S. real GDP

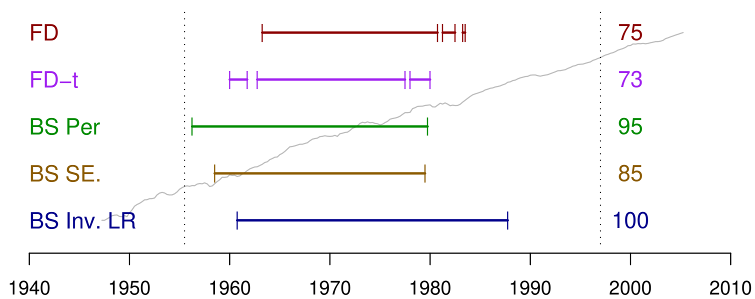
Based on the various methods for constructing confidence sets discussed in section 2, we find confidence sets for break dates in U.S. real GDP. It should be mentioned that all the confidence sets based on bootstrap methods require the computation of likelihood values. In particular, because we consider structural breaks in drift and variance respectively at different dates, it is necessary to compute likelihood values through maximization simultaneously with respect to coefficients and variance, rather than estimating models via OLS. However, bootstrapping with two different break dates is computationally infeasible given the present technology. Thus, we construct confidence sets for one break date by fixing the other break date at its maximum likelihood estimate. The number of bootstraps is also limited to 199. It is important to emphasize, however, that multiple breaks do not pose any problem for the FD approach because MCMC simulation from the likelihood requires only the addition of one more block in the sampler for an additional break. Thus, this problem for the bootstrap methods nicely illustrates the computational benefits of the FD approach.

The confidence sets for a break date in variance for the unit root and trend stationarity assumptions are very similar. These results are displayed in Figures 9(a) and 9(c). Across assumptions and methods, the break dates range between 1981 to 1990. The confidence sets under the assumption of trend stationarity are relatively tighter, with their lengths about one year shorter than those of confidence sets under the assumption of a unit root. Because likelihood-based confidence sets were shown to be the most accurate in the Monte Carlo analysis, we focus on them. The FD confidence sets include, at a 95% confidence level, dates between 1982:Q1 to 1984:Q4 under the unit root assumption and between 1982:Q3 to 1984:Q2 under the trend stationary assumption. The bootstrap inverted LR method produces confidence sets between 1982:Q1 to 1985:Q1 under the unit root assumption and between 1982:Q3 to 1984:Q1 under the trend stationarity assumption. Notably, these narrow confidence sets are as short or shorter than the 67% confidence interval of 1982:Q4 to 1985:Q3 reported in Stock and Watson (2002) that was based on Bai's method.[23] When we apply Bai's method to our dataset, the confidence set ranges from 1983:Q2 to 1990:Q2. The length is more than twice as long as that of the confidence sets for the likelihood-based approaches. Meanwhile, the narrowness of the confidence sets for the likelihood-based approaches supports the idea that the volatility reduction was sudden rather than gradual, as

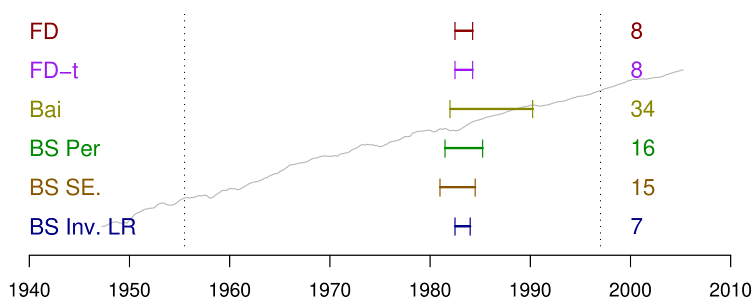
[23] Stock and Watson (2002) consider four quarter growth of U.S. real GDP, rather than annualized quarterly growth, as considered here. They discuss that because they use Bai's (1997) method by regressing the absolute value of residuals from an autoregression of real GDP growth on a constant and allowing a break in the constant from the auxiliary regression, the break estimator has a non-normal and heavy-tailed distribution, and 95% confidence interval would be very wide, hence their reporting of the 67% interval.



(a) Variance break date in growth



(b) Drift break date in levels



(c) Variance break date in levels

Figure 9: Confidence Sets for the Timing of Structural Breaks in Postwar U.S. real GDP

Note: The vertical dotted lines denote the outside bounds on potential break dates, $\Pi = [\lambda T, (1-\lambda)T]$ and the number in the right side of each confidence set is its length.

a gradual decline would presumably have corresponded to a relatively flat likelihood with respect to break dates and, therefore, would have resulted in a wide confidence set rather than a narrow one.

Under the assumption of trend stationarity, the confidence sets for the break in drift range as wide as the late 1950s to late 1980s, as displayed in Figure 9(b). The bootstrap percentile and bootstrap standard error confidence sets are shifted relatively earlier in the sample. For the bootstrap percentile approach, the confidence set covers 1956:Q2 to 1979:Q4. For the bootstrap standard errors approach, the confidence set covers 1958:Q3 to 1979:Q3. The bootstrapped inverted LR confidence set is the widest from 1960:Q4 to 1987:Q4.[24] Meanwhile, the FD confidence set consists of three disconnected intervals, which are 1963:Q2 to 1980:Q4, 1981:Q2 to 1982:Q3, and 1983:Q2 to 1983:Q3, although the gaps between the disconnected periods are very short, so we could, without too much loss of power, consider a slightly conservative 95% confidence set of 1963:Q2 to 1983:Q3. Just as the narrow confidence sets for the variance suggested that the structural change was sudden rather than gradual, these wide confidence sets for the drift lead to doubts about whether any structural change in long-run growth was sudden.

It may be worth noting that because, from a Bayesian perspective, the fiducial distribution for a break date is equivalent to its posterior distribution given noninformative priors, we could easily compute the Bayesian probability that the break exists before a specific date $\tau^\star$:

$$Pr(\tau < \tau^\star|\mathbf{y}) = \frac{1}{G}\sum_{g=1}^{G}\mathbf{1}[\tau^{(g)} < \tau^\star]$$

where G is the number of draws in MCMC procedure and $\tau^{(g)}$ is gth sampled break date. For example, it might be of interest to calculate the probability that the drift break happened before the 1973 oil price shock since our estimated break date is earlier. The probability that the break date is before 1973:Q1 based on drift model with one drift break and one variance break is $Pr(\tau < 1973 : Q1|\mathbf{y}) = 68.78\%$. The median date $\tau^{0.5}$ in the distribution such as $Pr(\tau \leq \tau^{0.5}|\mathbf{y}) = 50\%$ is 1970:Q4. Thus, the computed probability suggests that, at least from a Bayesian perspective, the source of drift break is less likely the first oil price shock than is typically assumed in the literature (e.g., Perron (1989)). Of course, from the classical perspective, 1973 lies within our confidence sets, so we cannot reject the oil shock

[24] Our Monte Carlo analysis in section 3 showed that the bootstrapped inverted LR confidence set generally produced shorter average lengths, while this particular confidence set for a break in drift in postwar U.S. real GDP data turns out to be the longest. However, in the Monte Carlo simulations, about 10% of simulated data sets resulted in the bootstrap inverted LR approach producing the longest confidence set. Thus, the results for the U.S. data might be thought of as an example of one of these 10% cases.

hypothesis for the productivity growth slowdown.

Finally, as described in the section 2, our FD approach can be applied given any distributional assumption. Thus, we also construct confidence sets under the assumption that the model errors follow a Student-t distribution. In order to consider the degrees of freedom in Student-t distribution, we simply add one more block for the degrees of freedom parameter in the MCMC simulation. Our results are robust in the sense that confidence sets from Student-t errors are exactly same as or slightly tighter than those based on normal errors. Notice that the confidence set for the break in drift under the assumption of trend stationarity is shifted 3 years earlier in the sample, although its length is quite similar to the length of confidence set from normal errors.

5 Conclusion

In this paper, we have proposed a fiducial distribution (FD) approach to constructing confidence sets for the timing of structural breaks in economic time series. In terms of practical implementation, the FD approach employes Markov chain Monte Carlo (MCMC) methods to simulate marginal fiducial distributions from the likelihood function. From a Bayesian perspective, this approach is equivalent to constructing Bayesian credible sets with noninformative priors. However, we take a classical interpretation of the confidence sets and evaluate them based on their performance in repeated samples. In particular, our criteria for evaluating confidence sets are that they should be short and have accurate coverage rates given a confidence level.

We have also considered various bootstrap approaches to constructing confidence sets for the timing of structural breaks. In particular, we discussed the implementation of constructing confidence sets using a bootstrap percentile approach, a bootstrap standard error approach, a bootstrap inverted likelihood ratio (LR) approach, and a bootstrap of Bai's (1997) statistic. We conducted Monte Carlo analysis of the coverage accuracy and expected length of confidence sets for different methods.

Our Monte Carlo analysis shows that the FD approach and the bootstrap inverted LR approach generally perform best in terms of accurate coverage rates and shorter expected length of confidence sets under a given confidence level. Notably, both methods are based on the likelihood function, although they are implemented in very different ways and can produce different inferences in a given setting.

We argue for using the FD approach in practice because it has several practical advantages in addition to coverage accuracy and shorter length. First, the FD approach is more broadly applicable in the sense that we can consider different parameters undergoing structural breaks at different break dates, as has been hypothesized for U.S. real GDP and other macroeconomic time series. Second, although bootstrap methods are available for complicated models of structural breaks, the FD approach is more computationally efficient, especially in settings such as respective breaks in mean and variance. Third, the FD approach requires only the ability to evaluate likelihood function over parameter values and potential break dates. Thus, it can be applied to any econometric model with a fully specified likelihood function (i.e. models with Student-t distributions, Poisson distributions, and so on).

Based on the various methods proposed in this paper, we examine the nature and timing of structural breaks in U.S. real GDP. We find support for the Great Moderation, with the maximum likelihood estimate for the break date of 1984:Q1 and a narrow confidence set from 1982:Q1 to 1984:Q4 for the FD approach under the assumption of a unit root. Meanwhile, there is no evidence of a break in the long-run growth rate under a unit root. By contrast, under the assumption of trend stationarity, we find evidence of a break in drift, with 1969:Q1 as the estimated break date. However, the confidence sets for the timing of a break in drift are quite wide, even for the FD approach, ranging from the early 1960s to the early 1980s. Meanwhile, if we take a Bayesian interpretation of the FD approach as providing a posterior under noninformative priors, the probability that a break occurred before the often hypothesized date of 1973:Q1 (e.g., Perron (1989)) is about 69% given normal errors. In terms of the Great Moderation under the assumption of trend stationarity, it is estimated to have occurred in 1982:Q4 and the FD confidence set is quite similar to the confidence set under assumption of a unit root.

References

Andrews, D. W. K. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61(4), 821–56.

——— (2003): "Tests for Parameter Instability and Structural Change with Unknown Change Point: A Corrigendum," *Econometrica*, 71(1), 395–397.

Andrews, D. W. K., and W. Ploberger (1994): "Optimal Tests When a Nuisance Parameter Is Present Only under the Alternative," *Econometrica*, 62(6), 1383–1414.

Bai, J. (1997): "Estimation Of A Change Point In Multiple Regression Models," *The Review of Economics and Statistics*, 79(4), 551–563.

——— (1999): "Likelihood Ratio Tests for Multiple Structural Changes," *Journal of Econometrics*, 91(2), 299–323.

Bai, J., R. L. Lumsdaine, and J. H. Stock (1998): "Testing for and Dating Common Breaks in Multivariate Time Series," *Review of Economic Studies*, 65(3), 395–432.

Bai, J., and P. Perron (2003): "Computation and Analysis of Multiple Structural Change Models," *Journal of Applied Econometrics*, 18(1), 1–22.

Barnett, V. (1999): *Comparative Statistical Inference*. Wiley, 3rd edn.

Berger, J. O., B. Liseo, and R. L. Wolpert (1999): "Integrated Likelihood Methods for Eliminating Nuisance Parameters," *Statistical Science*, 14(1), 1–28.

Chib, S., and E. Greenberg (1995): "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, 49(4), 327–335.

Diebold, F. X., and C. Chen (1996): "Testing StructuralStability with Endogenous Breakpoint: A Size Comparison of Analytic and Bootstrap Procedures," *Journal of Econometrics*, 70(1), 221–241.

Efron, B. (1979): "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, 7 (1), 1–26.

Efron, B., and R. J. Tibshirani (1994): *An Introduction to the Bootstrap*. Chapman & Hall/CRC.

Elliott, G., and U. Müller (2007): "Confidence Sets for the Date of a Single Break in Linear Time Series Regressions," *Journal of Econometrics*, 141, 1196–1218.

Fisher, R. A. (1930): "Inverse Probability," *Proceedings of the Cambridge Philosophical Society*, 28, 257–261.

Fraser, D. A. S. (1961a): "The Fiducial Method and Invariance," *Biometrika*, 48, 261–280.

——— (1961b): "On Fiducial Inference," *The Annals of Mathematical Statistics*, 32, 661–676.

Hannig, J. (2006): "On Fiducial Inference the good, the bad, and the ugly," Working Paper, Colorado State University.

Kilian, L. (1999): "Finite-Sample Properites of Percentile and Percentile-t Bootstrap Confidence Intervals for Impulse Resposes," *The Review of Economics and Statistics*, 81(4), 652–660.

Kim, C.-J., and C. R. Nelson (1999): "Has The U.S. Economy Become More Stable? A Bayesian Approach Based On A Markov-Switching Model Of The Business Cycle," *The Review of Economics and Statistics*, 81(4), 608–616.

Levin, A. T., and J. M. Piger (2007): "Bayesian Model Selection for Structural Break Models," Working Paper.

MacKinnon, J. G. (2002): "Bootstrap Inference in Econometrics," *Canadian Journal of Economics*, 35(4), 615–645.

——— (2006): "Bootstrap Methods in Econometrics," *The Economic Record*, 82(SI), S2–S18.

McConnell, M. M., and G. Perez-Quiros (2000): "Output Fluctuations in the United States: What Has Changed since the Early 1980's?," *American Economic Review*, 90(5), 1464–1476.

Perron, P. (1989): "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis," *Econometrica*, 57(6), 1361–1401.

Sensier, M., and D. van Dijk (2004): "Testing for Volatility Changes in U.S. Macroeconomic Time Series," *The Review of Economics and Statistics*, 86(3), 833–839.

Siegmund, D. (1988): "Confidence Set in Change-point Problems," *International Statistical Review*, 56(1), 31–48.

Sims, C. A., and T. Zha (1999): "Error Bands for Impulse Responses," *Econometrica*, 67(5), 1113–1156.

Stock, J. H., and M. W. Watson (1996): "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business & Economic Statistics*, 14(1), 11–30.

——— (2002): "Has the Business Cycle Changed, and Why?," *NBER Macroeconomics Annual*, 17, 159–218.

Zivot, E., and D. W. K. Andrews (1992): "Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis," *Journal of Business & Economic Statistics*, 10(3), 251–70.

Appendix

MCMC algorithm for a fiducial distribution

For the purpose of illustration, we explain a MCMC algorithm for a model with two groups of parameters that have breaks at different dates. For simulation from the likelihood function, we use the Metropolis-Hastings algorithm. In order to generate model parameters ψ, we maximize the likelihood function conditional on break dates τ_1 and τ_2 at every iteration. This reoptimization gives us mean and variance-covariance for an independent chain proposal density based on a multivariate Student t-distribution.

Step 1: Choose initial values for $\psi^{(0)}$, $\tau_1^{(0)}$, and $\tau_1^{(0)}$.

Step 2: Repeat for $j = 1, ..., M$. In the $(j+1)$th iteration,

1. For generating $\psi^{(j+1)}$,

(i) Propose

$$\boldsymbol{\psi'} \sim q(\psi^{(j)}, \psi' | \tau_1^{(j)}, \tau_2^{(j)}) = t_{|\tau_1^{(j)}, \tau_2^{(j)}}(\hat{\psi}, \hat{\Sigma}, \nu)$$

where $\hat{\psi} = argmaxL(\psi|\tau_1^{(j)}, \tau_2^{(j)})$ and $\hat{\Sigma} = -[\frac{\partial^2 lnL(\psi|\tau_1^{(j)}, \tau_2^{(j)})}{\partial\psi\partial\psi'}|_{\psi=\hat{\psi}}]^{-1}$.

(ii) Calculate

$$\alpha(\psi^{(j)}, \psi' | \tau_1^{(j)}, \tau_2^{(j)}) = min\{\frac{\pi(\psi', \tau_1^{(j)}, \tau_2^{(j)}) q(\psi', \psi^{(j)} | \tau_1^{(j)}, \tau_2^{(j)})}{\pi(\psi^{(j)}, \tau_1^{(j)}, \tau_2^{(j)}) q(\psi^{(j)}, \psi' | \tau_1^{(j)}, \tau_2^{(j)})}, 1\}$$

(iii) Draw U_0 from $\mathcal{U}(0,1)$. If

$$\begin{aligned} U_0 \leq \alpha(\psi^{(j)}, \psi' | \tau_1^{(j)}, \tau_2^{(j)}), \qquad & \psi^{(j+1)} = \psi'. \\ \text{Otherwise}, \qquad & \psi^{(j+1)} = \psi^{(j)}. \end{aligned}$$

2. For generating $\tau_1^{(j+1)}$,

$$\tau_1^{(j+1)} \sim \pi(\tau_1 | \psi^{(j+1)}, \tau_2^{(j)}, \mathbf{y})$$

where

$$\begin{aligned} \pi(\tau_1 | \psi^{(j+1)}, \tau_2^{(j)}, \mathbf{y}) &= \frac{\pi(\tau_1) f(\mathbf{y} | \psi^{(j+1)}, \tau_1, \tau_2^{(j)})}{\pi(\mathbf{y} | \psi^{(j+1)}, \tau_2^{(j)})} \\ &= \frac{\pi(\tau_1) f(\mathbf{y} | \psi^{(j+1)}, \tau_1, \tau_2^{(j)})}{\sum_{\tau_1} \pi(\tau_1) f(\mathbf{y} | \psi^{(j+1)}, \tau_1, \tau_2^{(j)})} \\ &= \frac{f(\mathbf{y} | \psi^{(j+1)}, \tau_1, \tau_2^{(j)})}{\sum_{\tau_1} f(\mathbf{y} | \psi^{(j+1)}, \tau_1, \tau_2^{(j)})} \end{aligned}$$

Since the break date τ is generated from a full conditional density, a proposal density and a target density are exactly same and the acceptance rate for the break date proposal, $\alpha(\tau_1^{(j+1)}, \tau_1^{(j)}|\psi^{(j+1)}, \tau_2^{(j)})$, is always 1.

3. For generating $\tau_2^{(j+1)}$, similarly

$$\tau_2^{(j+1)} \sim \pi(\tau_2|\psi^{(j+1)}, \tau_1^{(j+1)}, \mathbf{y})$$