

Monitoring Structural Breaks in Dynamic Regression Models With Bayesian Sequential Probability Test

Haixi Li^{1,1}

^aFreddie Mac, 1551 Park Run Dr., McLean, VA, 22102, USA; Tel: (571)7812981; Email: haixi.li@freddiemac.com

Abstract

Structural breaks are pervasive among macroeconomic and financial time series; consequently, forecast may loss accuracy out of sample, which renders monitoring structural breaks an important practice. We develop a structural break monitoring scheme, Bayesian Sequential Probability Test, for dynamic regression models, which consists of two components: a probabilistic detecting statistics and a sequential stopping procedure. We illustrate that not only the magnitude of deviation from the null signifies the breaks, but the increase in the frequency of such signals also provide important information. The detecting statistics incorporates such information into probabilities through Bayes' theorem. The sequential stopping procedure consists of a sequence of optimal stopping times that are derived by optimizing the loss function sequentially. We show the effectiveness of the BSPT by comparing its performance with that of CUSUM under a variety of DGPs and in a few economic applications.

Keywords: Structural Breaks; Bayesian Decision Theory; Optimal Stopping; Change Point Detection

¹The opinions expressed in this article are those of the author and do not reflect the view of the Freddie Mac. The author thanks Simon Sheng at the American University and Ping.Yu at the University of Hong Kong for helpful comments.

²Latest revision: November 16, 2020

1. Introduction

1.1. *Motivation*

Economy is a complicated, dynamic and evolving system, and relations among economic variables are unstable due to changes in law, policies, technologies and consumer preferences etc. ? documented substantial evidence of parameter instability in their models using 76 representative US monthly time series. However, practitioners of econometrics usually rely on relative stability to build statistical models for forecasts and inferences; when structural breaks happen unacknowledged, forecasts lose accuracy and inferences may be erroneous. Therefore it is of practical importance to detect the structural breaks promptly.

Data arrives sequentially in reality, which raises a concern: is the data generating process of the newly arrived data different from the one of the historical data? More specifically, practitioners need to update inference about the occurrence of a structural break and decide whether it can be concluded or no sufficient evidence to suggest so. Evidently, such decisions will incur two risks: false alarm and delay detection. It is in the practitioners' interest to minimize both risks simultaneously. However, the two risks are conflicting with each other: one risk can only be reduced at the cost of increasing the other. In order to make optimal decisions, practitioners need to strike a balance between these two risks.

To achieve this goal, we are in need of a sequential statistical decision framework that consists of at least two components: the detecting statistics and a decision rule. The detecting statistics should be updated with newly arrived data to convey uncertainty about the occurrence of a structural break; ideally, the proper language to communicate uncertainty is probability; the decision rule should minimize the practitioners' loss function.

Bayesian statistics is naturally sequential and probabilistic: the notion of updating belief in the wake of new information is fundamental to the Bayesian approach(?). Bayesian decision theory based on the Bayesian statistics updated with latest information, combined with decision maker's objective, will help to make the optimal decision. It is not always possible to make statistical inference with regard to the use of such information, as represented as a loss function, but such practice should be encouraged in general (?, p. 1-2); for an early study to incorporate a loss function into statistical analysis, see ?.

In this paper, we develop the Bayesian Sequential Probability Test (BSPT), as a structural break monitoring scheme for dynamic regression models.

1.2. Literature Review

There are extensive studies on structural breaks. However, most existing literature deals with test and identification of structural breaks in data of fixed length. Some classical references include ?, ?, ?, ? and ??????. Such methods are retrospective and can't be applied repeatedly for monitoring purpose. Since the law of iterated logarithm implies that, the empirical size will approach one with a constant boundary function as sample size grows (?). Hence, researchers resort to functional central limit theorem (FCLT) to approximate the boundary crossing probabilities of CUSUM type detecting statistics by those of standard Brownian motions.

? broke the ground of monitoring structural breaks in linear regression models and proposed two monitoring schemes for high frequency data: CUSUM and FL monitoring. The CUSUM detector is a partial sum of OLS recursive residuals as the detecting statistics which will converge to standard Brownian motion under the null of no structural breaks. The detecting statistics of FL monitoring are built on the deviation of updated parameter estimates from the historical estimates. These

statistics will converge to a vector of Brownian bridges under the null. Whenever one element of the detecting statistics exceeds the boundary, the structural break is concluded.

Subsequently, ? proposed an unified framework for monitoring structural breaks, in particular, they used OLS-residuals based CUSUM test, where the regression coefficients are estimated only once and residuals are computed with the estimated coefficients in the monitoring period. If there were structural breaks in the monitoring period, the residuals will deviate from zero consistently. The authors also proposed moving window OLS estimates as another monitoring scheme. The monitoring statistics consists of the differences of moving window parameter estimates and the historical estimate.

Both ? and ? pointed out that the boundary functions in their monitoring scheme are determined for mathematical convenience rather than optimality.

Besides the literature in econometrics, there is a long history of studies on sequential tests in statistics. Early work dates back at least to ?. ?'s sequential probability ratio test (SPRT) has become a benchmark to sequential tests nowadays. Since its inception , the subject has been extensively studied and applied in quality control (?) and clinical trails (?). We refer to ? for a comprehensive review. ? was the first to adopt the Bayesian approach to solve the change detection problem with detecting statistics derived through Bayes' theorem and the optimal stopping time obtained by minimizing a loss function which is the sum of the probability of false alarm and expected length of delay detection. But, the alternative distribution in ?'s problem is assumed to be known; such assumption is too restrictive for economic applications. We will revive ?'s ingenious work and extend it to cases with unknown alternative distributions to monitor structural breaks in dynamic regression models in this paper.

?

2. Bayesian Monitoring Scheme

2.1. Setup of the Problem

Consider a dynamic regression model in the following form:

$$y_t = x_t' \beta_t + \varepsilon_t, \quad (2.1)$$

where x_t is a k dimensional vector of regressors and β_t is a $k \times 1$ vector of regression coefficients, and the error terms ε_t is i.i.d white noise. Assume in the history $t \in \{1, \dots, n\}$, the coefficients in Equation (2.1) are constant and equal to β_0 . We want to monitor new data from $n+1$ onward to test if there is any structural break happening since. The null hypothesis is

$$\beta_t = \beta_0 \forall t \quad (2.2)$$

against the alternative that from some unknown time $\tau > n$, the coefficients changes to $\beta_1 \neq \beta_0$. To be more precise, we assume the following statistical process:

$$H_0 : y_t = x_t' \beta_0 + \varepsilon_t \text{ for } t = 0, 1, 2, \dots, \tau - 1,$$

$$H_1 : y_t = x_t' \beta_1 + \varepsilon_t \text{ for } t = \tau, \tau + 1, \dots$$

In practice, practitioners estimate the model with historical data, and start monitoring the out of sample performance of the estimated model with newly arrived data. Let $\hat{\beta}^n$ denote the OLS estimates with historical data up to $t = n$. Under the null, the residuals can be approximated as

$$\begin{aligned} \hat{\varepsilon}_t &= \varepsilon_t - x_t'(\hat{\beta}^n - \beta_0) \\ &= \varepsilon_t - x_t' \left(\left(\sum_{i=0}^n x_i' x_i \right)^{-1} \sum_{i=0}^n x_i' \varepsilon_i \right) \\ &\approx \varepsilon_t. \end{aligned}$$

The approximation is reasonable when $\hat{\beta}^n$ is estimated consistently. Under the alternative, the residuals can be approximated as

$$\begin{aligned}
\hat{\varepsilon}_t &= \varepsilon_t - x'_t(\hat{\beta}^n - \beta_1) \\
&= \varepsilon_t - x'_t(\beta_0 - \beta_1 + (\sum_{i=0}^n x'_i x_i)^{-1} \sum_{i=0}^n x'_i \varepsilon_i) \\
&\approx \varepsilon_t - x'_t(\beta_0 - \beta_1) \\
&\approx \varepsilon_t - \mu.
\end{aligned}$$

The approximation from the third equation to the fourth, is taken for mathematical convenience. One of the challenges in the change detection literature is the vast potential alternatives and the scantiness of the data under the alternative. Since delaying is costly, and decisions have to be made with limited data. Given the challenge, we take the approach of simplifying the stochastic process under the alternative, since our primary goal is to detect the break with minimum delay, the specification of the stochastic process under the alternative is of secondary concern. However, we did check the performance of the BSPT with simulation and economic applications to verify the validity of such simplification. To summarize, we assume the OLS residuals follow

$$\hat{\varepsilon}_t \sim \begin{cases} f_0 = \mathcal{N}(0, \sigma^2) & \text{for } t = 1, 2, \dots, \tau - 1 \\ f_1 = \mathcal{N}(\mu, \sigma^2) & \text{for } t = \tau, \tau + 1, \dots \end{cases} \quad (2.3)$$

? documented CUSUM type of detecting statistics with recursive residuals and OLS residuals. Under the null, the detecting statistics will converge to Brownian motion and Brownian bridge respectively. The residuals will have zero mean under the null and none zero mean under the alternative. The boundary function is chosen such that under the null, the empirical size is under the prescribed level. The asymptotically behaviour of CUSUM type of detecting statistics is only characterized under

the null, and the boundary function is determined correspondingly. The only fact under the alternative is the residuals will have none zero mean. In the comments of ?, one suggested to specify the distribution for the null and alternative as f_0 and f_1 respectively and derive the statistics from the likelihood ratio f_0/f_1 . This is inline with our idea which will serve the purpose of monitoring breaks: “to make hard and fast decisions”.¹

There is no assumption regarding the alternative with CUSUM; The Bayesian approach requires so. It is hard to argue which one is more appropriate, we would argue let the users to decide which scheme to serve the purpose better for their problems.

? show that the recursive residuals are i.i.d, $\mathcal{N}(0, \sigma^2)$ under more restrictive assumptions. In this paper, to assume OLS residuals are i.i.d $\mathcal{N}(0, \hat{\sigma}^2)$ means $\hat{\beta}^n$ is estimated consistently with historical data and the residuals can be standardized.¹

2.2. The Bayesian Sequential Probability Test

The Bayesian Sequential Probability Test is a Bayesian sequential statistical decision framework for monitoring structural breaks in dynamic regression models. It consists of two components: the detecting statistics of structural breaks and the decision rule (stopping time) that will minimize the practitioner’s loss function. Please

¹In the comments of ?, the commentator quote the authors:“(the work) should be regarded as yardsticks for the interpretation of data rather than leading to hard and fast decisions”.

¹In reality, the OLS residuals are interdependent and heteroskedastic. The assumptions are simplifications for these nuisance parameters: the main purpose is to detect the break with dynamic information; τ is the parameter of interest, and others are of secondary concern. We are relying on the fact that under the null the residuals will fluctuate around zero, and after the break, the residuals will deviate from zero consistently. We are trying to capture these basic facts about the structural break with simplistic assumptions. Please see Section ?? for more detailed discussion.

see Figure(??) for an overview of the framework.

2.2.1. The Detecting Statistics

Consider a random sequence of OLS errors $\{\varepsilon_t, t = 1, 2, 3, \dots\}$. $t = 1$ is the starting time of the monitoring period. Assume that conditional on the unobservable parameter τ , the structural break time, the sequence $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{\tau-1}\}$ being i.i.d with marginal distribution f_0 , and $\{\varepsilon_\tau, \varepsilon_{\tau+1}, \dots\}$ being i.i.d with marginal distribution f_1 . The null distribution f_0 is assumed to be known, and f_1 is parameterized distribution with parameters denoted as $\theta \in \Theta$. θ indexes the alternative distribution f_1 as we assume the alternative distribution is unknown¹.

Assume geometric prior for the structural break time τ :

$$P(\tau = k) = \begin{cases} \pi & \text{if } k = 0 \\ (1 - \pi)\rho(1 - \rho)^{k-1} & \text{if } k = 1, 2, \dots \end{cases} \quad (2.4)$$

Let $\Pi_0(\theta)$ denote the prior distribution over Θ , and $\Pi_t(\theta)$ denote the posterior distribution of the parameters given information \mathcal{I}_t . So the prior includes two components: the geometric prior distribution for τ and prior distribution $\Pi_0(\theta)$ over the parameter space Θ .

Denote $\varepsilon^t = \{\varepsilon_1, \dots, \varepsilon_t\}$ as the history of random variable ε_t and $\mathcal{X}^t = \mathcal{X}_1 \times \mathcal{X}_2 \times \dots \mathcal{X}_t$, where \mathcal{X}_t is the sample space for ε_t . We now define the probability distribution. For each set $A = \{\omega: \varepsilon_1 \leq x_1, \dots, \varepsilon_t \leq x_t\} \in \mathcal{X}^t$, given $\theta \in \Theta$

$$\begin{aligned} P_\theta(A) &= \pi P_\theta^1(A) + (1 - \pi) \sum_{i=0}^{t-1} \rho(1 - \rho)^i P^0\{\varepsilon_1 \leq x_1, \dots, \varepsilon_i \leq x_i\} \\ &\quad \times P_\theta^1\{\varepsilon_{i+1} \leq x_{i+1}, \dots, \varepsilon_t \leq x_t\} + (1 - \pi)(1 - \rho)^t P^0(A) \end{aligned}$$

¹ θ contains the unknown parameters in alternative distribution f_1 , In the numerical exercises in this paper, we assume the variance can be standardized in the alternative.

The P^0 and P_θ^1 are the probability measures underlying the null distribution and the alternative distribution given $\theta \in \Theta$ respectively. More generally, the probability measure given distribution Π over parameter space Θ is defined as:

$$P_\Pi(A) = \int_{\Theta} P_\theta^\pi(A) d\Pi(\theta) \quad (2.5)$$

Definition 2.1. Define the detecting statistics as the posterior probability of structural break happened before time t , and \mathcal{I}_t denotes the information available at time t .

$$\pi_t = P(\tau \leq t | \mathcal{I}_t), \text{ for } t = 1, 2, 3, \dots \quad (2.6)$$

We will describe the evolution of π_t . Given $\theta \in \Theta$, let $\pi_{t,\theta}$ denote the probability of a structural break that has occurred before t conditional on $\theta \in \Theta$:

$$\pi_{t,\theta} = P(\tau \leq t | \mathcal{I}_t, \theta), \quad (2.7)$$

where \mathcal{I}_t is the information set available at time t .

Given π_{t-1} and ε_t , $\pi_{t,\theta}$ evolves according to the following equation:

$$\pi_{t,\theta} = \frac{(\rho(1 - \pi_{t-1}) + \pi_{t-1}) f_1(\varepsilon_t | \theta)}{f(\varepsilon_t | \theta)} \quad (2.8)$$

where

$$f(\varepsilon_t | \theta) = (1 - \rho)(1 - \pi_{t-1})f_0(\varepsilon_t) + (\pi_{t-1} + \rho(1 - \pi_{t-1})) f_1(\varepsilon_t | \theta).$$

Given the likelihood at time t , we can update the posterior at $t - 1$. The posterior distribution over the parameter space is denoted as Π_t , and Π_t evolves as

$$\Pi_t(\theta) = \frac{f(\varepsilon_t | \theta) \Pi_{t-1}(\theta)}{\int f(\varepsilon_t | \theta) d\Pi_{t-1}(\theta)}. \quad (2.9)$$

Lastly, π_t is updated as

$$\pi_t = \int_{\Theta} \pi_{t,\theta} d\Pi_t(\theta) \quad (2.10)$$

Since the parameter θ is unknown, the idea is to average it out with available information. It is trivial to show that given posterior distribution Π_t , the expectation of π_t only dependent on π_{t-1} , that makes value function iteration a possibility, which is key to solve the practitioner's optimization problem.

2.2.2. The Practitioner's Objective Function

The practitioner's objective is to detect the structural break as soon as possible with minimum probability of false alarm. This is a sequential decision problem, where the practitioner will examine a sequence of observations one at a time, and decide after each observation whether to conclude the break or continue sampling.

Such decisions will incur two risks: false alarm and delay detection. When the announcement occurs before the break, it is a false alarm; otherwise, it is delay detection¹. Moreover, the two risks are conflicting with each other: one risk can only be minimized at the cost of the other. The loss function is the weighted sum of the two risks, and the practitioner will need to strike a balance between the two risks in order to achieve optimality.

The expected loss function of stopping time T given parameter θ is given by

$$L(\theta, T) = E_{\theta} \{1(T < \tau) + c(T - \tau)^+\}. \quad (2.11)$$

The first component is the probability of false alarm, and the second component is the expected length of delay detection multiplied by a controlling factor c . The controlling factor reflects the practitioner's penalty on delay detection relative to

¹The definitions are conceptual, since the structural breaks are not observable, so which mistake has been made is only known retrospectively with certain confidence when there are sufficient information.

false alarm. The stopping time T or stopping rule² is a measurable function of information available at time t . The loss function with known alternative was first proposed and solved by ?, but the assumption of known alternative is too restrictive to economic applications in general. Our work was built on ??’s original work and extend it to unknown alternative in the context of applied econometrics. Please refer to ? for exposition with more modern notation.

The Bayes risk function given the prior Π_0 is given by

$$r(\pi, \Pi_0) = \inf_{T \in \mathbf{T}} E^{\Pi_0} L(\theta, T). \quad (2.12)$$

The expectation is taken over the parameter space Θ .

2.2.3. the Solution to the Practitioner’s Problem

The optimization problem in Equation(??) is very complicated. To solve it with ? type of stopping time at time 0 is intractable. This is mainly due to the posterior distribution Π_t that defines the expectation, is evolving. We formulate the solution based on the ground work laid down by ?¹ for Bayesian sequential decision problems: after each new observation is taken, we should consider a new sequential problem with updated prior, i.e. the posterior as the new prior, and minimize the loss function looking forward with available information². To be more specific, consider ε_t is observed, then we obtain the posterior Π_t given ε^t , then we can consider a new

²Stopping time is a more formal notion, it basically means the final time to announce the structural break. It is a random function measurable to information available at the time.

¹? laid down the ground work for Bayesian sequential analysis in Chapter 7, which help us to formulate the solution.

²The problem in general is very complicated and difficult, other approaches in change detection literature include ?. The authors proposed a solution which is asymptotically Bayes in the sense $\rho \rightarrow 0$.

sequential problem with $\Pi_t(\theta)$ as the prior and $\{\varepsilon_{t+1}, \varepsilon_{t+2}, \dots\}$ as the possible observations. We will solve the optimization problem every time when there is new data, hence the optimal stopping time will reflect the latest information available which is represented by the posterior distribution at the time. This is in sharp contrast to the boundary function of CUSUM which is usually a deterministic function of time. We will obtain a sequence of optimal stopping times indexed by the posterior distribution Π_t . We present the solution more formally in the flowing³.

Define the sequential stopping procedure \mathcal{T} as a sequence of stopping times $T_0, T_1(\varepsilon^1), T_2(\varepsilon^2), \dots$: the procedure will stop when the first stopping time declares to stop.

$$\mathcal{T}(\mathcal{X}) = \inf\{t \geq 0 : T_t(\varepsilon^t) = t\} \quad (2.13)$$

At any time t , after ε^t is observed, the practitioner is facing the following problem:

$$r(\pi, \Pi_t) = \inf_{T \in \mathbf{T}} E^{\Pi_t} \{1(T < \tau) + c(T - \tau)^+\}, \quad (2.14)$$

where \mathbf{T} is a collection of stopping times such that $E^{\Pi_t}(T_t) < \infty$. $r(\pi, \Pi_t)$ represents the minimum Bayes risk that can be attained when ε^t has been observed. Given ε^t , the decision is whether to stop or continue sampling. If immediate stop yield less loss than expected loss of continuing, it will be optimal to stop or continue otherwise. It is not difficult to show the loss of immediate stop given π_t , but to show the expected loss of continuing involves quantifying the value function. The value function will be computed through value function iteration with the latest information available through the posterior distribution Π_t ¹.

³To build up the machinery for Bayesian sequential analysis is a complicated task, that requires tremendous amount of notations. Figure (??) is meant to offer an overview of the framework.

¹Other approximation method to attain optimal stopping time has been proposed such as one-step-ahead method.

Define the optimal stopping time

$$T_t^* = \arg \inf_{T \in \mathbf{T}} E^{\Pi_t} \{1 (T < \tau) + c (T - \tau)^+\}. \quad (2.15)$$

Hence the optimal sequential stopping procedure will consist of a sequence of optimal stopping times:

$$\mathcal{T}^* = \{T_1^*, T_2^*, \dots, T_t^*, \dots\}. \quad (2.16)$$

Now, we solve the optimal stopping problem for any generic prior distribution Π ¹:

Lemma 2.1. *Suppose $E^{\Pi}(\tau) < \infty$ and $E^{\Pi}(T) < \infty$ and define the sequence π_t by*

$$\pi_t = P(\tau \leq t \mid \mathcal{I}_t), \text{ for } t = 0, 1, 2, \dots \quad (2.17)$$

Then for $T \in \mathbf{T}$, we can write

$$E^{\Pi} \{1 (T < \tau) + c (T - \tau)^+\} = E^{\Pi}(1 - \pi_T + c \sum_{i=0}^{T-1} \pi_i) \quad (2.18)$$

Now the Bayes risk function is transformed as

$$r(\pi, \Pi) = \inf_{T \in \mathbf{T}} \left\{ E^{\Pi}(1 - \pi_T + c \sum_{t=0}^{T-1} \pi_t) \right\} \quad (2.19)$$

Such a result may appear surprising at first, but after looking at it closely, it is very intuitive: the loss is the sum of probability of false alarm ($1 - \pi_T$) and expected delay ($c \sum_{t=0}^{T-1} \pi_t$) at the time of stopping, c is the loss of per period of delay. ? first showed this transformation for known alternative distribution.

Since we have shown that π_t is a Markov process given expectation E^{Π} defined with prior distribution Π , we can apply the result of Markov optimal stopping theory. The optimization problem became:

$$v(\pi) = \min \left\{ 1 - \pi, c\pi + \int_{\Theta} v(\pi'_\theta \mid \pi) d\Pi(\theta) \right\}, \quad (2.20)$$

¹If the prior distribution Π is degenerate, it is then the problem showed by ?.

where π'_θ given π is obtained by Equation (??), and the function $v(\pi)$ can be obtained through value function iteration. The stopping time

$$T^* = \inf \{t \geq 0 : v(\pi_t) = (1 - \pi_t)\} \quad (2.21)$$

is the optimal stopping time. This result follows from Markov optimal stopping theory¹. The intuition is evident: to stop will incur loss $1 - \pi$, which is the probability of false alarm; to continue will incur one period of expected loss of delay $c\pi$ plus the expected loss of next period on. The first time that stopping yields less loss, then it is the time to stop. The controlling factor c is the unit loss of delay.²

The function in Equation (??) defines two regions of the state space π : the continuation region defined as $\{\pi : v(\pi) > 1 - \pi\}$; the rejection region defined as $\{\pi : v(\pi) < 1 - \pi\}$. It is optimal to stop when π cross from continuation region into rejection region. Continue from Equation(??), we have

$$T^* = \inf \{t \geq 0 \mid \pi_t \geq \pi^*\}. \quad (2.22)$$

The Equation (??) defines the optimal stopping time as a threshold π^* such that $v(\pi) = 1 - \pi$. When the detecting statistics $\pi_t \geq \pi^*$ at time t , the structural break should be concluded.

Figure 1: Iteration of value function until it converges. The value function crosses the line $1 - \pi$ at π^* .

¹Please see Corollary 3.11 on page 59 of ? for more detailed explanation.

²Equation(??) is essentially value function iteration, π is the state variable, which is more familiar to economist. Figure (??) illustrates the process of value function iteration.

3. Simulation

We present some simulation results in this section. Following ? and ?, the data are generated from i.i.d $\mathcal{N}(0, 1)$ and $\mathcal{N}(0.8, 1)$ for null and alternative distribution respectively³. The monitoring period starts at 100 and the break times are $\tau \in \{110, 150, 200, 300\}$. The number of simulation is 10,000. The mean of the alternative distribution is unknown to the practitioner. The results are summarized in Table(??). Figure(??) provide an illustration of the BSPT. From the above table, one can see that with higher value of control factor c , the expected delay is shorter and probability of false alarm is higher. This is due to the trade-off between the two risks that is weighted by control factor.

Figure 2: Demonstration of BSPT with Monte Carlo simulation. The mean of the simulated process changes from 0 to 0.8 at 150. When the detecting statistics exceed the decision threshold: $\pi \geq \mathcal{T}^*$ at 160, the structural break is announced.

³As a well known fact that the bigger the difference between null and alternative, the easier the detection. We choose 0.8 that is in line with previous studies.

Control Factor	Structure Break Time	Expected Delay(sd)	Prob of False Alarm
$c = 0.005$	110	27.64 (14.70)	0.28%
	150	29.27 (21.43)	0.52%
	200	30.68 (27.75)	1.09%
	300	33.75 (38.77)	2.61%
$c = 0.008$	110	25.65 (14.24)	0.40%
	150	27.31 (20.37)	0.81%
	200	28.57 (25.84)	1.71%
	300	31.02 (35.49)	4.19%
$c = 0.01$	110	24.88 (13.97)	0.48%
	150	26.41 (20.04)	1.04%
	200	27.56 (24.63)	2.16%
	300	29.49 (32.63)	5.28%

Table 1: Summary of Simulation Results with Structural Break in the Mean

We also compare the performance of BSPT with that of CUSUM. The results are presented in Table (??). The alternative process is simulated with the following equation: $y_t = \alpha + \gamma y_{t-1} + \epsilon_t$, where ϵ_t is white noise. This exercise also exams the BSPT's ability to deal with intertemporally dependent alternative processes. In general, BSPT can detect the break with much less average delay under similar empirical size. More importantly, BSPT is much more robust against early and late structural breaks. It is well known that the CUSUM type of detecting statistics is powerful for the early break and loss its power gradually as monitoring continues.

		CUSUM			BSPT		
	Break Time	Exp Delay(sd)	type I	type II	Exp Delay(sd)	type I	type II
$\alpha = 0.8$	110	30(19)	0.55%	0%	26(14)	0.36%	0%
	150	58(34)	1.90%	0%	27(20)	0.86%	0%
$\gamma = 0$	200	90(52)	2.52%	0%	28(25)	1.57%	0%
	300	151(84)	2.71%	0.12%	31(35)	4.40%	0%
$\alpha = 0.3$	110	61(74)	0.48%	0.18%	34(32)	0.36%	0%
	150	108(99)	1.99%	0.91%	37(35)	0.86%	0%
$\gamma = 0.5$	200	153(120)	2.64%	2.29%	38(38)	1.57%	0%
	300	223(134)	3.41%	8.50%	41(45)	4.40%	0%
$\alpha = 0.1$	110	86(129)	0.55%	8.76%	33(36)	0.36%	0%
	150	152(162)	1.92%	19.66%	34(35)	0.86%	0%
$\gamma = 0.7$	200	199(168)	2.64%	30.52%	34(34)	1.57%	0%
	300	240(158)	3.16%	48.43%	35(36)	4.40%	0%
$\alpha = 0.0$	110	48(80)	0.57%	6.80%	22(23)	0.36%	0%
	150	96(119)	1.89%	20.39%	23(24)	0.86%	0%
$\gamma = 0.8$	200	134(138)	2.45%	36.58%	24(23)	1.57%	0%
	300	172(139)	2.99%	59.24%	24(24)	4.40%	0%

Table 2: We compare the performance of CUSUM and BSPT under the same alternative process: $y_t = \alpha + \gamma y_{t-1} + \epsilon_t$, where ϵ_t is white noise. The results of CUSUM are obtained with OLS-CUSUM method in the R package strucchange by ? with 10,000 simulations. The control factor in BSPT is set at 0.008 and the prior is presented in Table (??). Type II error refers to the break is not detected before 900 periods.

This result can be attributed to a few advantages of BSPT: First, the sequential stopping procedure consists of a sequence of optimal stopping times that reflect dynamic information in the data arriving sequentially. Second, it is commonly known in the literature that large deviation from the null is evidence in favour of the alternative; but the increase in the frequency of such deviation is also very important for detection. The detecting statistics of BSPT is keen to incorporate the increase in the frequency and convert it to the posterior probability through Bayesian updating. Third, the trade off between false alarm and delay detection is optimized dynamically

with latest information. In contrast to the CUSUM type of statistics, the decision rule usually is a deterministic function of time.

4. Applications of BSPT

We demonstrate the ability of BSPT with a few economic applications including German M1 money demand, US labor productivity and US HPI. We use the same set of prior as in Table (??) and $\rho = 0.01$.

Prior Distribution in Simulation	
μ	$[-2, -0.6] \cup [0.6, 2]$
π	0.5
ρ	0.01

Table 3: Prior Distribution For the Parameters

4.1. The Boom-Bust Cycles and Forecast Performance of House Price Model

The recent financial crisis in 2008 has revived the research interest in housing price forecast. Particularly the challenge posed by the boom-bust cycles of house price to forecast. When the house price gets into a boom cycle, forecast tends to be underestimated; when the house price gets into a bust cycle, forecast tends to be overestimated. Following ?, we use AR(2) model for one period ahead forecast for five states with diverse house price dynamics: California, Florida, Massachusetts, Ohio and Texas. The house price of theses states display a variety of different growth patterns. We use FHFA HPI quarterly purchase-only indexes and focus on the boom-bust cycle in the 2000s.

Figure 3: The Annual Growth Rate of HPI for Five States

We use data from 1991 Q2 to 2000 Q4 as the historical sample to estimate the AR(2) model for the five states, and start monitoring the one period ahead forecast errors from 2000 Q1.

For FL, we first identify a structural break in the forecast errors with positive mean in 2004 Q1. and then we update the model estimation with data from 1991 to 2005, and then start monitoring from 2004. We identify forecast errors with negative mean at 2006 Q1. Please see Figure(??,??) for the graphical presentation of the exercise.

Figure 4: The Boom of House Price in FL

Figure 5: The Bust of House Price in FL

As for CA, we detect a structural break in the forecast errors with negative mean in 2005 Q3. As it can be seen from Figure(??). Similar to CA, we also identify bust at 2006 Q1 for OH and 2010 Q2 for TX. For MA, we did not find sufficient evidence to reject the null with zero mean forecast errors. Please see Figure(??,??,?? and ??).

Figure 6: The Bust of House Price in CA

Figure 7: The Bust of House Price in TX

Figure 8: The Bust of House Price in OH

Figure 9: The House Price Forecast Errors in MA

4.2. *US Labor Productivity*

It has been asserted that the US labor productivity has experienced a speedup in the second half of 1990s. ? investigated US labor productivity in the manufacturing/durable sector, and found strong evidence of a structural break sometime between 1992 and 1996, and weaker evidence of structural break in the 1960s and early 1980s. We use data ranging from Jan 1964 to Dec 1980 as the historical sample to estimate an AR(1) model. Standardized forecast errors are obtained from Jan 1980 to Dec 1997. The result is presented in Figure(??). We use the same data as in ?, and detect a structural break at Aug 1997. The BSPT yields a little delayed result than ?, this is because ?'s test statistics is derived from the whole sample, while the stopping time in BSPT only utilizes the information available at the time of the decision.

As can be seen in Figure(??), the forecast errors start to deviate from zero and become positive since early 1992. The detecting statistics starts to pick up, as the information favouring the alternative is not strong, the detecting statistics stays way below the threshold before 1994. After such evidence that favouring the alternative keep flowing in, the information is picked up through likelihood function and update the detecting statistics. This illustrates an important aspect of the BSPT: not only the magnitude of the information favouring the alternative is important to detection, but the increase in the frequency of such information is also important. The Bayesian updating algorithm is keen to pick up such signal. Such intuition is reflected in Equation(??).

Figure 10: The upper panel is the forecast error of US labor productivity. The lower one consists of the detecting statistics π and the sequential stopping procedure \mathcal{T}^* . The forecast errors deviate from zero since 1992, and the probability starts to pick up as more evidence favouring the alternative flows in. Eventually it crosses the threshold at Aug 1997.

4.3. German M1 Money Demand

? studied the stability and linearity of German M1 money demand function. They found clear evidence of structural instability after the monetary unification of Germany happened since 1990 Q3. We use the same linear model and data as in ? and detect the structural break at 1990 Q4 from monitoring perspective. Figure(??) demonstrates the monitoring process.

Figure 11: The German M1 Money Demand

5. Discussion

The monitoring scheme of structural breaks suffers from huge potential alternatives. To make the matter worse, the nature of the problem requires the practitioner to make decisions with limited post-change data because delay is costly. ? studied one-time end-of-sample instability test; and proposed S test that relies on the fact that under the null, the errors are centered around zero, while the post-change errors will deviate from zero, and a large deviation from zero is evidence against the null. We rely on the same notion that under the alternative, the errors will deviate from zero consistently. Also, the primary interest of monitoring is the structural break time τ , all other parameters are of secondary concern.

Given the above reasons, we assume the errors under the alternative to be i.i.d. with constant variance and non-zero mean. The process of unknown mean plus standard errors is the simplest process that can capture the consistent nonzero errors under the alternative. In practice, intertemporally dependent errors are conceivable due to structural breaks. We did simulations to test the performance of BSPT given AR(1) residuals under the alternative. The results are summarized in Table(??).

We choose the prior for the mean of the alternative distribution to be uniform distribution over the set $[-2, -0.6] \cap [0.6, 2]$ ¹. Uniform prior is the most uninformative prior. The prior chosen is used through out the exercises in this paper. This also rises

¹We can include variance as an unknown parameter. Given the limited post-break data and the variance is also a nuisance parameter, we choose standard variance for all the exercises in this paper. We also did simulation with variance as the unknown parameter that can be provided per request. We can also think of the prior as the specification of the composite alternative, since the detection process is essentially sequential hypothesis testing.

another interesting intuition: the purpose of monitoring is to identify the break as soon as possible with minimum risk of false alarm. If the practitioner has knowledge about the alternative, that should be incorporated into the prior. The better the alternative is known, the quicker the break can be identified². The special case is when the alternative is known. Following ?, we also assume geometric prior for the breaking time τ in Equation (??), higher value of ρ means earlier break in the detecting period. This challenge to specify ρ is not unique to the Bayesian approach that the decision about the likely timing of future break. It is well known that the boundary in ? perform well for early breaks, but gets increasingly insensitive to late breaks. ? proposed different boundaries to distribute the empirical size more evenly. The authors also acknowledge that it is desirable to have more flexible and less heuristic instrument to select boundaries according to a specified prior distribution of the timing of the break. If there is information regarding the likely timing of the future break, it can be incorporated as the prior. We choose $\rho = 0.01$ in all our exercises. The simulation results of other specifications can be provided per request.

The controlling factor c is the cost of one period of delay. Higher value of the controlling factor suggests higher punishment for delay. We tabulated different value of the controlling factor and corresponding probability of false alarm and expected length of delay in Table(??). More tabulation can be provided by request.

As we can see from Figure(??), the decision threshold is decreasing since 150. This is because the information favouring the alternative starts to accumulate after 150. the evolution of the posterior distribution reflect this: the probability mass start to

²In certain situation, such prior knowledge is applicable. For example, if the structural break is caused by business cycle, that in normal time, the growth rate is positive and in down time, the growth rate will change from positive to negative.

accumulate in the positive region(see Figure(??)). As such information accumulates, there is less probability of false alarm; in order to minimize delay, we can afford to lower the threshold. Hence the threshold reflect the inflow of the information dynamically.

Figure 12: This figure shows the evolution of posterior distribution of the parameter in the alternative distribution. The structural break happens at 150, and is detected at 160. As more information supporting the alternative accumulates, the posterior distribution amasses around 0.8.

6. Concluding Remarks

Relationships among economic variables that are known to be stable historically might change due to structural breaks out-of-sample. Consequently, it is important to monitor structural breaks. Monitoring is a process of sequential statistical decision making. To solve such a problem, we start with formulating the practitioner's objective function as weighted sum of two risks: the probability of false alarm and the expected length of delay detection. We derive the decision rule by optimizing the objective function sequentially, that results in a sequence of optimal stopping times. It is important to note that the magnitude of deviation from the null is sure an evidence favouring the alternative, but the increase in the frequency of such signals is also very important for detecting purpose. The probabilistic detecting statistics of BSPT is keen to pick up such signals through Bayesian updating. We demonstrate the effectiveness of the framework with Monte Carlo simulations and economic applications.

Monitoring is a process of sequential hypothesis testing, and Bayesian statistics is inherently sequential. We believe this is a promising line of research. Future research might include a wider range of unknown parameters; recursively estimated OLS parameters etc.

7. Appendix section

7.1. Figures and Tables

8. Proof

Proof of Lemma ??.

Remark. *The proof with known alternative was done first by ? then ? presented more modern approach. In the case of parameterized distribution, it is assumed that the parameters of the alternative distribution is known a priori. We extend the proof to unknown alternative. The expectation and probability is taken given the uncertainty in the problem, which the parameterized alternative distribution is unknown. Hence the expectation is taken over the parameter space with posterior distribution informed with information available at time t .*

For any stopping time T , we have the following

$$P(T < \tau) = E^\Pi(1(T < \tau)) = E^\Pi(1 - \pi_T) \quad (8.23)$$

The second equality follows from the definition of π_t . Now we need to show

$$E^\Pi(T - \tau)^+ = E^\Pi\left(\sum_{t=0}^{T-1} \pi_t\right) \quad (8.24)$$

For any $t > 0$,

$$\begin{aligned}
E^\Pi(t - \tau)^+ &= \sum_{k=0}^t (t - k)P(\tau = k \mid \mathcal{I}_t) \\
&= \sum_{k=0}^{t-1} P(\tau \leq k \mid \mathcal{I}_t) \\
&= \sum_{k=0}^{t-1} [P(\tau \leq k \mid \mathcal{I}_t) - P(\tau \leq k \mid \mathcal{I}_k)] + \sum_{k=0}^{t-1} P(\tau \leq k \mid \mathcal{I}_k) \\
&= \sum_{k=0}^{t-1} [P(\tau \leq k \mid \mathcal{I}_t) - P(\tau \leq k \mid \mathcal{I}_k)] + \sum_{k=0}^{t-1} \pi_k
\end{aligned}$$

Define

$$\begin{aligned}
M_t &= \sum_{k=0}^{t-1} [P(\tau \leq k \mid \mathcal{I}_t) - P(\tau \leq k \mid \mathcal{I}_k)] \\
&= - \sum_{k=0}^{t-1} [P(\tau > k \mid \mathcal{I}_t) - P(\tau > k \mid \mathcal{I}_k)]
\end{aligned}$$

We can show that M_t is a regular martingale:

$$E^\Pi(M_{t+1} \mid \mathcal{I}_t) = M_t. \quad (8.25)$$

Define

$$\begin{aligned}
M &= \lim_{t \rightarrow \infty} \left\{ \sum_{k=0}^{t-1} [P(\tau \leq k \mid \mathcal{I}_t) - P(\tau \leq k \mid \mathcal{I}_k)] \right\} \\
|M| &\leq \lim_{t \rightarrow \infty} \sum_{k=0}^{\infty} [P(\tau > k \mid \mathcal{I}_t) + \sum_{k=0}^{\infty} P(\tau > k \mid \mathcal{I}_k)].
\end{aligned}$$

And also by assumption, we have

$$E^\Pi \sum_{k=0}^{\infty} [P(\tau > k \mid \mathcal{I}_t)] = \sum_{k=0}^{\infty} P(\tau > k) = E^\Pi(\tau) < \infty$$

and

$$E^\Pi \sum_{k=0}^{\infty} P(\tau > k \mid \mathcal{I}_k) = E^\Pi(\tau) < \infty.$$

Hence, M is integrable. Now we want to show M is regular.

$$\begin{aligned}
E^\Pi(M \mid \mathcal{I}_m) &= \sum_{k=0}^{\infty} E^\Pi[P(\tau \leq k \mid \mathcal{I}_t) - P(\tau \leq k \mid \mathcal{I}_k) \mid \mathcal{I}_m] \\
&= \sum_{k=0}^{m-1} [P(\tau \leq k \mid \mathcal{I}_m) - P(\tau \leq k \mid \mathcal{I}_k)] \\
&= M_m
\end{aligned}$$

The second equation follows since for any $m < k$, we have $I_m \subset I_k$ and the first term will be canceled by the second term in the equation due to iteration property of expectation. Hence M_t is regular. By optional sampling, it implies

$$E^\Pi(M_T) = E^\Pi(M_0) = 0. \tag{8.26}$$

Then the Lemma follows. □

Figure 13: The Skeleton of Bayesian Sequential Probability Test

