MPRA
Munich Personal RePEc Archive

Has the Business Cycle Changed in Japan? A Bayesian Analysis Based on a Markov-Switching Model with Multiple Change-Points.

Iiboshi, Hirokuni and Watanabe, Toshiaki

Tokyo Metropolitan University, Hitotsubashi University

July 2005

Online at https://mpra.ub.uni-muenchen.de/93865/
MPRA Paper No. 93865, posted 13 May 2019 16:30 UTC

Has the Business Cycle Changed in Japan?

A Bayesian Analysis Based on a Markov-Switching Model

with Multiple Change-Points.

Hirokuni Uchiyama * and Toshiaki Watanabe *

Abstract

Using a Markov-switching model and Bayesian inference, the turning points of Japanese business cycles are identified from a monthly coincident composite index series, taken over the last thirty years. Ordinarily, in taking such a long-range estimation approach, we would face the following questions: (1) Have there been any structural changes? (2) If so, does the existence of these structural changes prevent the detection of the turning points? (3) How many changes have occurred? (4) When did these changes occur? The Bayesian analysis approach easily provides answers. The estimation results suggest that the Markov-switching model with no changes is unable to identify turning points appropriately, whereas the model with changes selected via the Bayes factor robustly estimates these points for the long period of time considered, and also successfully facilitates estimation of the changes in amplitude that occurred between booms and recessions, as well as in the volatility of business cycles.

Key words or phrases: Bayes factor, Gibbs sampler, marginal likelihood, structural break, turning points of business cycles, unknown change points.

* Faculty of Economics, Tokyo Metropolitan University, 1-1 Minami-Osawa Hachioji Tokyo 192-0397, Japan

1 Introduction

Many methods for detecting business cycle turning points have been formulated in the last half-century, after Burns and Mitchell (1946) proposed a primitive procedure. For instance, Stock and Watson (1991), Hodrick and Prescott (1997), and Baxter and King (1999) attempted to identify turning points by extracting the co-movement from many relevant macroeconomic series, or the cyclical components from the quarterly series of real gross domestic product (GDP), using several filtering methods. Hamilton (1989) also proposed the *Markov mean switching AR model*, in which the mean of the growth rate in each period switches from the mean of the booms to the mean of the recessions, or *vice-versa*, based on their transition probabilities. Many empirical studies have been implemented using these methods.

Recently estimated periods, however, have been as long as four or five decades, and various studies have reported that there were structural breaks in the business cycles of several countries in the post-war period. In the case of the U.S., Kim and Nelson (1999), McConnell and Perez-Quiros (2000), Blanchard and Simon (2001), Stock and Watson (2002), and Kim, Nelson and Piger (2004) showed that a reduction in the volatility of GDP occurred in the mid-eighties. During the same period, European business cycles also experienced a structural break of lower volatility, according to Artis, Krolzig and Toro (2004). These breaks may have changed not only the volatility around the short-run means during booms and recessions, but also the gap between both means of the cycles. In this case, these breaks should prevent the precise detection of turning points, were we to develop analytical methods or models without considering these changes. In order to build a model that takes the presence of these structural changes into account, and to estimate the turning points efficiently, we must first address the following issues: (1) Do structural changes really exist? (2) How many changes have occurred? (3) When did these changes occur? In the last decade, several tests dealing with unknown break points in the classical framework have been proposed, by Andrew (1993), Andrews and Ploberger (1994) and Hansen (1997), *etc.*. But there are some problems with the classical tests of unknown timing. One of these problems is that the test statistics are based on a nonstandard asymptotic distribution, because the estimated change-point becomes a nuisance parameter that exists only under the alternative hypothesis that structural change occurred. Another problem is that, with only a few exceptions (e.g., Bai and Perron, 1998), theoretical studies have dealt with the issue of test statistics only for one-time structural breaks, but have not yet considered the case of multiple breaks. In contrast, a Bayesian approach is very applicable to this problem. Therefore, we adopt a Markov-switching model and use Bayesian inference, following Kim and Nelson (1999); we also extend their models dealing with only one-time change-points to models that consider multiple change points, following the manner of Chib (1998). Koop and Potter (1999) demonstrated that a Bayesian approach is superior to the classical approach for nonlinear models, of which structural break models are a subset. In addition, only a Bayesian approach allows a comparison among models with various numbers of change-points, and a selection of the model with the most appropriate number and timing of such points, using the Bayes factor. Using the model selection procedure, we specify the turning points and evaluate the precision with which the turning points detected via the selected model are estimated.

Our study investigates the last thirty years of the Japanese business cycle, using a coincident composite index that constitutes a monthly series, constructed by the Economic and Social Research Institute (ESRI), from eleven series of coincident business cycles. Although the empirical studies mentioned above identified only a single change in the volatility of real GDP in the economies of both the U.S. and European countries, no such studies exist with regard to Japan, with one exception. Hamori and Bhar (2003) found two breaks in the volatility of the de-meaned quarterly GDP series for the postwar period and divided it into two regimes: a period of low volatility from 1975 through 1990, and a period of high volatility thereafter. However, these authors employed only a two-state Markov-switching model, so they assumed that GDP exhibited only two kinds of volatility. Furthermore, they did not specify the number of breaks or consider the business cycle. Like Hamori and Bhar (2003), we suspect that a one-time-break model is not sufficient to examine the case of Japan. To specify the number and size of the breaks of the cycles, we need to examine various types of models with multiple change-points. This necessity arises from the fact that the Japanese economy has experienced drastic fluctuation, with high growth in the 1960s, a deep recession resulting from the oil crisis in the mid-1970s, mild growth in the early 1980s, and a long depression in the 1990s that occurred in the aftermath of the rapid asset price inflation in the late 1980s. In fact, the empirical results estimated in this paper indicate that there were two structural changes, one in 1975, after the first oil crisis, and the other around 1989, before “*the lost decade*”, as the 1990s are often called; the former change makes the amplitude and volatility of the coincident index shrink, while the latter enlarges both. These results also illustrate that a Markov-switching model that excludes structural changes cannot appropriately be used to detect the turning points of the 1980s, which was a period of mild growth, whereas the model with structural changes selected by means of the Bayes factor isolates these points robustly for the entire period, covering the last thirty years, and also allows successful estimation of the magnitude of the changes in volatility around its means during booms and recessions, as well as the gap between both means in the business cycle.

The organization of this paper is as follows. In Section 2, we introduce the Markov-switching models that take into account the multiple structural changes, which are employed in this study. Section 3 discusses Bayesian inference in the context of this model, using the Gibbs sampler and a model selection procedure that makes use of the posterior probability derived from the Bayes factor and the marginal likelihood. In Section 4, we examine the precision with which the turning points are detected from the coincident composite index using the proposed model. Section 5 concludes the paper.

2 Markov-Switching Model with Multiple Change Points

We now identify the model used to measure the turning points of business cycles with multiple breaks. Following the approach employed by Kim and Nelson (1999), who considered a model with only one break, a Markov-switching model is applied and extended to a model with multiple breaks. First, the Markov-switching model proposed by Hamilton (1989) is defined below.

$$\phi(L)(y_t - \mu_t) = \epsilon_t, \quad \epsilon_t \sim i.i.d.N(0, \sigma_t^2), \tag{2.1}$$

$$\mu_t = \mu_{0t}(1 - S_t) + \mu_{1t} S_t, \tag{2.2}$$

$$\mu_{0t} < \mu_{1t}, \tag{2.3}$$

where y_t is the percentage change in the coincident composite index; μ_{0t} and μ_{1t} are the means of y_t during recessions and booms, respectively; L is a lag operator; roots of $\phi(L) = 0$ lie outside the complex unit circle, so that Eq.(2.1) guarantees stationarity; and S_t is a latent binary indicator that determines the business situation: $S_1 = 1, 0$ denotes a boom and a recession, respectively. S_t switches between zero and one, based on a two-state Markov process, the transition probabilities for which are given below.

$$Pr[S_t = 1 \mid S_{t-1} = 1] = p_{11}, \quad Pr[S_t = 0 \mid S_{t-1} = 1] = 1 - p_{11},$$

$$Pr[S_t = 0 \mid S_{t-1} = 0] = p_{00}, \quad Pr[S_t = 1 \mid S_{t-1} = 0] = 1 - p_{00}, \tag{2.4}$$

$$0 < p_{00} < 1, \quad 0 < p_{11} < 1, \tag{2.5}$$

Eq.(2.4) indicates that the transition probabilities for the business cycle phases, S_t, depend only on those of the previous period, S_{t-1}, but not on other factors, and that these variables are fixed with respect to time. So that these probabilities may be more readily understood, we rewrite them in matrix form, as follows:

$$P = \begin{bmatrix} p_{11} & 1 - p_{00} \\ 1 - p_{11} & p_{00} \end{bmatrix}. \tag{2.6}$$

Eq.(2.6) suggests that each business state is reversible with the other state, e.g., boom with recession, or recession with boom.

We next consider how to insert multiple structural breaks into the model above. Suppose that the two shift parameters, μ_{0t} and μ_{1t}, as well as the variance of ϵ_t, σ_t^2, depend on a parameter D_t, the value of which changes at unknown points in time, $\Upsilon_n = \{\tau_1, \tau_2, \cdots, \tau_n\}$, and remains constant otherwise, and where $\tau_1 > 1$ and $\tau_n < T$. Upon setting

$$D_t = \begin{cases} 1 & \text{for} \quad 1 \le t \le \tau_1 \\ 2 & \text{for} \quad \tau_1 < t \le \tau_2 \\ \vdots & \\ n & \text{for} \quad \tau_{n-1} < t \le \tau_n \\ n+1 & \text{for} \quad \tau_n < t \le T \end{cases} \tag{2.7}$$

$$D_1 = 1, \quad D_T = n + 1,$$

where n is the number of change points and D_t is independent of S_t. Given D_t and n, in order to facilitate the expression of the shift parameters, μ_{0t}, μ_{1t} and σ_t^2, which are bred by multiple breaks, let $\boldsymbol{d_t}$ be $n \times 1$ the vector whose $i-th$ element d_i is one, and otherwise zero, at the period t between the change-points

τ_{i-1} and τ_i.

$$\mathbf{d_t'} = \{d_1, d_2, d_3, \cdots, d_n\},$$

$$d_i = 1, d_{\neq i} = 0 \text{ for } D_t = i, \quad i = 1, 2, \cdots, n, \tag{2.8}$$

where $d_{\neq i}$ are elements, except for the $i - th$ element. Using $\boldsymbol{d_t}$, these three shift parameters can be expressed as

$$\mu_{0t} = \boldsymbol{\mu_0' d_t}, \qquad \boldsymbol{\mu_0'} = \{\mu_{01}, \mu_{02}, \cdots, \mu_{0n}\} \tag{2.9}$$

$$\mu_{1t} = \boldsymbol{\mu_1' d_t} \qquad \boldsymbol{\mu_1'} = \{\mu_{11}, \mu_{12}, \cdots, \mu_{1n}\} \tag{2.10}$$

$$\sigma_t^2 = \boldsymbol{\sigma^{2'} d_t} \qquad \boldsymbol{\sigma^{2'}} = \{\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2\}, \tag{2.11}$$

where the elements μ_{0i} and μ_{1i}, respectively, denote the short-run means of y_t during the contraction phase and the expansion phase, between change-points τ_{i-1} and τ_i. In addition, σ_i^2 represents the variance between change-points τ_{i-1} and τ_i.

Following the approach of Chib (1998), the latent parameter D_t is constructed from the multiple-state Markov switching model, so that D_t shifts from $D_{\tau_m} = m$ to $D_{\tau_{m+1}} = m + 1$ at an unknown change-point τ_m constrained by constant transition probabilities, similar to the unobserved variable S_t, as explained above. However, unlike S_t, the current value of D_t may be only one of two possibilities in the multi-state Markov-switching model. That is, D_t can either stay at its current value or jump to the next higher value, since D_t plays the role of the shift parameter in the change-point model. Accordingly, these transition probabilities are represented as

$$\begin{aligned}
Pr[D_t = 1 \mid D_{t-1} = 1] &= q_{11}, \quad Pr[D_t = 2 \mid D_{t-1} = 1] = 1 - q_{11}, \\
Pr[D_t = 2 \mid D_{t-1} = 2] &= q_{22}, \quad Pr[D_t = 3 \mid D_{t-1} = 2] = 1 - q_{22}, \\
&\vdots \\
Pr[D_t = n-1 \mid D_{t-1} = n-1] &= q_{n-1,n-1}, \quad Pr[D_t = n \mid D_{t-1} = n-1] = 1 - q_{n-1,n-1}, \\
Pr[D_t = n \mid D_{t-1} = n] &= 1. \qquad (2.12)
\end{aligned}$$

Furthermore, with respect to the matrix expression, we can rewrite these probabilities as

$$Q = \begin{bmatrix} q_{11} & 0 & 0 & \cdots & 0 \\ 1-q11 & q_{22} & 0 & \cdots 0 & 0 \\ 0 & 1-q_{22} & q_{33} & \cdots 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 \cdots & q_{n-1,n-1} & 0 \\ 0 & 0 & \cdots & 1-q_{n-1,n-1} & 1 \end{bmatrix}. \tag{2.13}$$

The transition probability matrix Eq.(2.13) indicates that each current state is irreversible with its opposite state, contrary to the situation in Eq. (2.6). For example, conditional on $D_t = 2$, whose elements are located in the second column in Eq.(2.13), two nonzero probabilities always exist; q_{22} that D_{t+1} may stay at 2, and $1 - q_{22}$ that D_{t+1} may jump to 3, but the probability that $D_{t+1} = 1$ or that $D_{t+1} = 4$ should be zero.

Chib (1998) and Kim and Nelson (1999) implemented the inference and model selection procedure for their proposed model, based on Bayesian methods using the Gibbs sampler. The next section discusses these procedures.

3 Bayesian Inference and Model Selection

3.1 Inference via Gibbs Sampling Procedure

The Gibbs sampler is an estimation method that makes use of the property that the conditional distribution of each parameter converges to its marginal distribution under iterated computation (Casella and George 1993). Using this algorithm, the posterior marginal distributions of parameters are derived from their posterior conditional distributions.

For Bayesian inference in the context of the model explained in the last section, we need to derive the joint posterior density, as follows:

$$f(\tilde{\mu}, \tilde{\phi}, \tilde{\sigma^2}, \tilde{D_T}, \tilde{S_T}, P, Q \mid \tilde{Y_T}) = f(\tilde{\mu}, \tilde{\phi}, \tilde{\sigma^2} \mid \tilde{Y_T}, \tilde{S_T}, \tilde{D_T}) f(P \mid \tilde{S_T}) f(Q \mid \tilde{D_T}) f(\tilde{S_T} \mid \tilde{Y_T}) f(\tilde{D_t} \mid \tilde{Y_T}),$$

where $\tilde{\mu} \equiv [\mu_{01}, \cdots, \mu_{0n}, \mu_{11}, \cdots, \mu_{1n}]'$; $\tilde{\phi} \equiv [\phi_1, \cdots, \phi_k]'$; $\tilde{\sigma^2} \equiv [\sigma_1^2, \cdots, \sigma_n^2]'$; $\tilde{D_T} \equiv [D_1, \cdots, D_T]'$; $\tilde{S_T} \equiv [S_1, \cdots, S_T]'$; $\tilde{Y_T} \equiv [y_1, \cdots, y_T]'$. The first term on the right-hand-side (RHS) denotes the posterior density of the parameters $\tilde{\mu}$, $\tilde{\phi}$, and $\tilde{\sigma^2}$ in Eq.(2.1), conditional on $\tilde{Y_T}$ and $\tilde{S_T}$. The second and third terms are the posterior densities of parameters P and Q in Eq.(2.6) and Eq.(2.13). Conditional on $\tilde{S_T}$ and $\tilde{D_T}$, $f(P \mid \tilde{S_T})$ and $f(Q \mid \tilde{D_T})$ are here respectively assumed to be independent of the other parameters and of the data $\tilde{Y_T}$. The fourth and last terms represent the posterior densities of the latent variables $\tilde{S_T}$ and $\tilde{D_T}$, which are independent of each other, conditional on Y_T. Unfortunately, direct computation of these posterior densities is not an attractive option. However, these can instead be inferred via an implementation of the Gibbs sampler, because we may easily derive the posterior distributions of each of the variates, conditional on all of the other variates, as follows:

- $[\ \tilde{\mu} \ \mid\ \theta_{-\mu}, \tilde{Y_T}, \tilde{S_T}, \tilde{D_T}\ ]$
- $[\ \tilde{\phi} \ \mid\ \theta_{-\phi}, \tilde{Y_T}, \tilde{S_T}, \tilde{D_T}\ ]$
- $[\ \tilde{\sigma^2} \ \mid\ \theta_{-\sigma^2}, \tilde{Y_t}, \tilde{S_t}, \tilde{D_t}\ ]$
- $[\ \tilde{S_t} \ \mid\ \tilde{Y_t}, \tilde{D_T}, \theta\ ]$
- $[\ \tilde{D_t} \ \mid\ \tilde{Y_t}, \tilde{S_T}, \theta\ ]$
- $[\ P \ \mid\ \tilde{S_T}\ ]$
- $[\ Q \ \mid\ \tilde{D_T}\ ]$

where $\theta = (\tilde{\mu}, \tilde{\phi}, \tilde{\sigma^2}, P, Q)$, and, for example, $\theta_{-\mu}$ denotes all the parameters in θ, excluding $\tilde{\mu}$. Thus, using arbitrary starting values for the parameters, the conditional distributions above can be simulated, and generating values are sampled as the posterior marginal distributions of the parameters after convergence occurs. The following seven steps describe the generation of the conditional distributions indicated above.

Step 1 Generate $\tilde{\mu}$ from the conditional posterior distribution, i.e., from the truncated normal distribution, conditional on the given parameters, $\tilde{\sigma}^{2,(j-1)}$ and $\tilde{\phi}^{(j-1)}$, and on the latent variables $\tilde{S_T}^{(j-1)}$ and $\tilde{D_T}^{(j-1)}$, as follows. (The superscript j denotes the $j-th$ iteration.)

$$\tilde{\mu}^{(j)} \mid \tilde{\sigma}^{2,(j-1)}, \tilde{\phi}^{(j-1)}, \tilde{S_T}^{(j-1)}, \tilde{D_T}^{(j-1)}, \tilde{Y_T} \sim \mathcal{N}(M_1, \Sigma_{M1}) \mathcal{I}(\mu_{1i} > \mu_{0i}),$$

$$M_1 = (\Sigma_{M0}^{-1} + \sigma^{-2} Z'Z)^{-1} (\Sigma_{M0}^{-1} M_0 + \sigma^{-2} Z'z)$$

$$\Sigma_{M1} = (\Sigma_{M0}^{-1} + \sigma^{-2} Z'Z)^{-1},$$

where M_1 and Σ_{M1} are the posterior means and variances of the coefficients; M_0 and Σ_{M0} are the corresponding prior values; and z and Z are, respectively, the $\big((T-k)\times 1\big)$ vector composed of the elements $z_t = y_t - \sum_{i=1}^{k} \phi_i y_{t-i}$ and the $\big((T-k)\times(2\times n)\big)$ matrix composed of the $\big(1\times(2\times n)\big)$ vectors $Z_t = (\boldsymbol{d_t'} \otimes [1 - S_t - \sum_{i=1}^{k} \phi_i(1 - S_{t-i})],\ \boldsymbol{d_t'} \otimes [S_t - \sum_{i=1}^{k} \phi_i S_{t-i}])$, where $\otimes$ denotes element-by-element multiplication. $\mathcal{I}(\mu_{1i} > \mu_{0i})$ is an indicator function that returns unity, if $\mu_{1i} > \mu_{0i}$, following Eq.(2.3), and zero otherwise. Furthermore, the prior distribution is also given by the truncated normal distribution

$$\tilde{\mu} \mid \tilde{\sigma^2}, \tilde{\phi} \sim \mathcal{N}(M_0, \Sigma_{M0})\mathcal{I}(\mu_{1i} > \mu_{0i}).$$

Step 2 Generate $\tilde{\phi}$ from the conditional posterior distribution, i.e., from the truncated normal distribution, conditional on the given parameters, $\tilde{\sigma}^{2,(j-1)}$ and $\tilde{\mu}^{(j)}$, and on $\tilde{S_T}^{(j-1)}$ and $\tilde{D_T}^{(j-1)}$, in order to guarantee the stationarity of Eq.(2.1), as follows:

$$\tilde{\phi}^{(j)} \mid \tilde{\sigma}^{2,(j-1)}, \tilde{\mu}^{(j)}, \tilde{S_T}^{(j-1)}, \tilde{D_T}^{(j-1)}, \tilde{Y_T} \sim \mathcal{N}(\Phi_1, \Sigma_{\phi 1})\mathcal{I}(s(\phi)),$$

$$\Phi_1 = (\Sigma_{\phi 0}^{-1} + \sigma^{-2}V'V)^{-1}(\Sigma_{\phi 0}^{-1}\Phi_0 + \sigma^{-2}V'v)$$

$$\Sigma_{\phi 1} = (\Sigma_{\phi 0}^{-1} + \sigma^{-2}V'V)^{-1},$$

where Φ_1 and $\Sigma_{\phi 1}$ are the posterior means and variances of the coefficients; Φ_0 and $\Sigma_{\phi 0}$ are the corresponding prior values; and v and V are, respectively, the $\big((T-k)\times 1\big)$ vector composed of $v_t = y_t - \mu_{0t}(1-S_t) - \mu_{1t}S_t$ and the $\big((T-k)\times k\big)$ matrix composed of the $(1\times k)$ vector $V_t = (v_{t-1}, \cdots, v_{t-k})$ given in Eq.(1). In addition, $\mathcal{I}(s(\phi))$ is an indicator function that returns unity, if the roots of $\phi(L) = 0$ lie outside the unit circle, and zero otherwise. The conjugate prior distribution is as follows:

$$\tilde{\phi} \mid \tilde{\sigma}^2, \tilde{\mu}, \sim \mathcal{N}(\Phi_0, \Sigma_{\phi 0})\mathcal{I}(s(\phi)).$$

Step 3 Generate $\tilde{\sigma^2}$ from the conditional posterior distribution, i.e., from the inverted gamma distribution, conditional on the given parameters $\tilde{\phi}^{(j)}$ and $\tilde{\mu}^{(j)}$, and on $\tilde{S_T}^{(j-1)}$ and $\tilde{D_T}^{(j-1)}$, as follows.

$$\tilde{\sigma}^{2,(j)} \mid \tilde{\phi}^{(j)}, \tilde{\mu}^{(j)}, \tilde{S_T}^{(j-1)}, \tilde{D_T}^{(j-1)}, \tilde{Y_T} \sim \mathcal{IG}(\frac{\nu_1}{2}, \frac{\delta_1}{2}),$$

$$\nu_1 = \nu_0 + T - k,$$

$$\delta_1 = \delta_0 + \sum_{t=k+1}^{T} \Big(y_t - \mu_{0t}(1-S_t) - \mu_{1t}S_t - \sum_{i=1}^{k} \phi_i(y_{t-i} - \mu_{0t-i}(1-S_{t-i}) - \mu_{1t}S_{t-i})\Big)^2,$$

where ν_1 and δ_1 are the posterior variances and degrees of freedom, and ν_0 and δ_0 are the priors. In addition, the conjugate prior distribution is the inverted gamma distribution

$$\sigma^2 \mid \phi, \mu_0, \mu_1 \sim \mathcal{IG}(\frac{\nu_0}{2}, \frac{\delta_0}{2}).$$

Step 4 Generate $\tilde{S}_t$ following the algorithm of the *multi-move sampler* originally motivated by Chib (1996). Here, let $S^{t+1} = (S_{t+1}, \cdots, S_T)$; then, the joint density of $\tilde{S_T}$ can be expressed as

(3.1)

$$p(\tilde{S_T} \mid \tilde{Y_T}, \tilde{D_T}, \tilde{\mu}, \tilde{\phi}, \tilde{\sigma}^2, P) = p(S_T \mid \tilde{Y_T}, \tilde{D_T}, \tilde{\mu}, \tilde{\phi}, \tilde{\sigma}^2, P) \times \cdots \times p(S_t \mid \tilde{Y_T}, S^{t+1}, \tilde{D_T}, \tilde{\mu}, \tilde{\phi}, \tilde{\sigma}^2, P)$$
$$\times \cdots \times p(S_1 \mid \tilde{Y_T}, S^2, \tilde{D_T}, \tilde{\mu}, \tilde{\phi}, \tilde{\sigma}^2, P),$$

The sampling of S_t is implemented based on the term, $p(S_t \mid \tilde{Y_T}, \tilde{D_T}, S^{t+1}, \tilde{\mu}, \tilde{\phi}, \tilde{\sigma}^2, P)$ of the joint density above, from S_1 to S_T. This sampling procedure is summarized below. By Bayes' theorem, the target function can be divided into two terms,

$$p(S_t \mid \tilde{Y_T}, S^{t+1}, \tilde{D_T}, \tilde{\mu}^{(j)}, \tilde{\phi}^{(j)}, \tilde{\sigma}^{2(j)}, P^{(j-1)}) \propto p(S_t \mid \tilde{Y_T}, \tilde{D_T}, \tilde{\mu}^{(j)}, \tilde{\phi}^{(j)}, \tilde{\sigma}^{2(j)}, P) \times p(S_{t+1} \mid S_t, P),$$

and then the second term of the RHS is derived from the Markov transition probabilities in Eq.(2.6), and the first term is obtained by executing the following two steps.

(i) Prediction Step Derive $p(S_t = h \mid \tilde{Y}_{t-1}, \tilde{D}_T, \tilde{\mu}, \tilde{\phi}, \tilde{\sigma}^2, P)$ from Eq.(3.2).

(3.2) $$p(S_t = h \mid \tilde{Y}_{t-1}, \tilde{D_T}, \tilde{\mu}, \tilde{\phi}, \tilde{\sigma}^2, P) = \sum_{l=1}^{h} p_{lh} \times p(S_{t-1} = l \mid \tilde{Y}_{t-1}, \tilde{D_T}, \tilde{\mu}, \tilde{\phi}, \tilde{\sigma}^2, P),$$

for $h = 1, 2$, and where $\tilde{Y}_{t-1} = (y_1, y_2, \cdots, y_t)$, $p(S_{t-1} \mid \tilde{Y}_t, \tilde{D}_t, \tilde{\mu}, \tilde{\phi}, \tilde{\sigma}^2, P)$ is derived from the updated step for period $t-1$, and p_{lh} are the Markov transition probabilities from Eq.(2.6).

(ii) Updating Step Derive $p(S_t \mid \tilde{Y}_t, \tilde{D_T}, \tilde{\mu}, \tilde{\phi}, \tilde{\sigma}^2, P)$ from Eq.(3.3).

(3.3)

$$p(S_t \mid \tilde{Y}_t, \tilde{D_T}, \tilde{\mu}, \tilde{\phi}, \tilde{\sigma}^2, P) \propto p(S_t \mid \tilde{Y}_{t-1}, \tilde{D_T}, \tilde{\mu}, \tilde{\phi}, \tilde{\sigma}^2, P) \times f(y_t \mid \tilde{S}_t, \tilde{Y}_t, \tilde{D_T}, \tilde{\mu}, \tilde{\phi}, \tilde{\sigma}^2, P, Q),$$

where $f(y_t \mid \tilde{S}_t, \tilde{Y}_t, \tilde{D}_t, \tilde{\mu}, \tilde{\phi}, \tilde{\sigma}^2, P, Q)$ is the likelihood function.

Step 5 Generate $\tilde{D}_t$ in a manner similar to the generation undertaken in Step 4. The sampling of D_t is executed based on the function $p(D_t \mid \tilde{Y_T}, \tilde{S_T}, D^{t+1}, \tilde{\mu}, \tilde{\phi}, \tilde{\sigma}^2, Q)$, from D_1 to D_T. This function may also be divided into two terms,

$$p(D_t \mid \tilde{Y_T}, D^{t+1}, \tilde{S_T}, \tilde{\mu}^{(j)}, \tilde{\phi}^{(j)}, \tilde{\sigma}^{2(j)}, Q^{(j-1)}) \propto p(D_t \mid \tilde{Y_T}, \tilde{S_T}, \tilde{\mu}^{(j)}, \tilde{\phi}^{(j)}, \tilde{\sigma}^{2(j)}, Q) \times p(D_{t+1} \mid D_t, Q),$$

where the second term of the RHS is derived from the Markov transition probability in Eq.(2.13), and the first term is obtained from the following two steps.

(i) Prediction Step Derive $p(D_t = g \mid \tilde{Y}_{t-1}, \tilde{S}_T, \tilde{\mu}, \tilde{\phi}, \tilde{\sigma}^2, Q)$ from Eq.(3.4).

(3.4) $$p(D_t = g \mid \tilde{Y}_{t-1}, \tilde{S_T}, \tilde{\mu}, \tilde{\phi}, \tilde{\sigma}^2, Q) = \sum_{l=g-1}^{g} q_{lg} \times p(D_{t-1} = l \mid \tilde{Y}_{t-1}, \tilde{D_T}, \tilde{\mu}, \tilde{\phi}, \tilde{\sigma}^2, Q),$$

for $g = 1, 2, \cdots, n$, where $\tilde{Y}_{t-1} = (y_1, y_2, \cdots, y_t)$, $p(D_{t-1} \mid \tilde{Y}_t, \tilde{S_T}, \tilde{\mu}, \tilde{\phi}, \tilde{\sigma}^2, Q)$ is available from the updating step of period $t-1$, and where q_{lq} is the Markov transition probability from Eq.(2.13).

(ii) Updating Step Derive $p(D_t \mid \tilde{Y}_t, \tilde{S_T}, \tilde{\mu}, \tilde{\phi}, \tilde{\sigma}^2, Q)$ from Eq.(3.5).

(3.5)

$$p(D_t \mid \tilde{Y}_t, \tilde{S_T}, \tilde{\mu}, \tilde{\phi}, \tilde{\sigma}^2, Q) \propto p(D_t \mid \tilde{Y}_{t-1}, \tilde{S_T}, \tilde{\mu}, \tilde{\phi}, \tilde{\sigma}^2, Q) \times f(y_t \mid \tilde{D}_t, \tilde{Y}_t, \tilde{S_T}, \tilde{\mu}, \tilde{\phi}, \tilde{\sigma}^2, P, Q),$$

where $f(y_t \mid \tilde{D}_t, \tilde{Y}_t, \tilde{S_T}, \tilde{\mu}, \tilde{\phi}, \tilde{\sigma}^2, P, Q)$ is the likelihood function.

Step 6 Generate P from the conditional posterior distribution, i.e., from the beta distribution, conditional on $\tilde{S}_t$, since the posterior distribution of p_{ii} is assumed to be independent of that of p_{jj}, $j \neq i$.

$$p_{ii}^{(j)} \mid \tilde{S_T}^{(j)} \sim beta(u_{ii} + n_{ii}, u_{ij} + n_{ij}),$$

where u_{ii} and u_{ij}, $i, j = 0, 1$ are prior hyperparameters and n_{ij} denotes the number of transitions from state i to state j that may easily be counted, for a given $\tilde{S_T}$. In this case, the prior distribution is

$$p_{ii} \sim beta(u_{ii}, u_{ij}).$$

Step 7 Generate Q in a manner similar to that used in Step 6. The posterior distribution is given by

$$q_{ii}^{(j)} \mid \tilde{D_T}^{(j)} \sim beta(r_{ii} + m_{ii}, r_{ij} + 1),$$

where r_{ii} and r_{ij}, $i, j = 0, 1$ are prior hyperparameters and m_{ii} denotes the number of transitions for which the state i does not change. In the second term of the beta distribution, one is added to the prior parameters as the number of one-step transitions from state i to state $i+1$. In addition, the prior distribution is

$$q_{ii} \sim beta(r_{ii}, r_{ij}).$$

3.2 Model Selection by Bayes Factor

Similar to the use of the *AIC (Akaike Information Criterion)* in classical inference, in Bayesian inference, we compare the marginal density of the sample data y (equivalently, *the marginal likelihood*) under the model M_k , $p(y \mid M_k)$, to those of others, when we evaluate which model is superior. We can easily capture this marginal likelihood from the simulation result of the Gibbs sampler, following Chib (1995). Furthermore, it is convenient to use the *posterior probability*, derived from the Bayes factor, and the marginal likelihood in the model selection process.

We may define the marginal log-likelihood for each model as given below.

$$\log p(y \mid M_k) = \log p(y \mid \theta^*, M_k) + \log p(\theta^* \mid M_k) - \log p(\theta^* \mid y, M_k),$$

where $\log p(y \mid \theta^*, M_k)$ is the log-likelihood function and $\log p(\theta^* \mid M_k)$ and $\log p(\theta^* \mid y, M_k)$ are the prior and posterior densities of each parameter θ^*. Here, the mean of the sample resulting from the Gibbs sampler is adopted as θ^*, after the fashion of Chib (1995).

The log-likelihood function and the log of the prior density at $\theta = \theta^*$ may be evaluated relatively easily. First, the log-likelihood function is given by

$$\begin{aligned} \ln f(\tilde{Y}_t \mid \tilde{\theta}^*) = \sum_{t=k+1}^{T} \ln \Big(& \sum_{V_t=1}^{2\times(n+1)} \cdots \sum_{V_{t-k}=1}^{2\times(n+1)} \\ & \times p(V_t, \cdots, V_{t_k} \mid \tilde{Y}_{t-1}, \tilde{\theta}^*) \\ & \times f(y_t \mid \tilde{Y}_{t-1}, V_t, \cdots, V_{t-k}, \tilde{\theta}^*) \Big), \end{aligned} \tag{3.6}$$

where $V_t = 1 \quad$ if $S_t = 0$ and $D_t = 0$;

$V_t = 2 \quad$ if $S_t = 0$ and $D_t = 1$;

$\vdots$

$V_t = n+1 \quad$ if $S_t = 0$ and $D_t = n+1$;

$V_t = n+2 \quad$ if $S_t = 1$ and $D_t = 0$;

$V_t = n+3 \quad$ if $S_t = 1$ and $D_t = 1$;

$\vdots$

$V_t = 2(n+1) \quad$ if $S_t = 1$ and $D_t = n+1$.

Secondly, the log of the prior density is given by

$$\ln \pi(\tilde{\theta}^*) = \ln \pi(\tilde{\mu}^*) + \ln \pi(\tilde{\theta}^*) + \ln \pi(\tilde{\sigma}^{2*}) + \ln \pi(P^*, Q^*), \tag{3.7}$$

where it is assumed a priori that $\tilde{\mu}^*$, $\tilde{\theta}^*$, $\tilde{\sigma}^{2*}$, P^* and Q^* are independent of one another.

Evaluation of the posterior density at $\tilde{\theta} = \tilde{\theta}^*$ is more demanding, but we can take advantage of the approach proposed by Chib (1995). For this purpose, consider the following decomposition of the posterior density:

$$\begin{aligned} \pi(\tilde{\theta}^* \mid \tilde{Y_T}) &= \pi(\tilde{\mu}^* \mid \tilde{Y_T})\pi(\tilde{\phi}^* \mid \tilde{\mu}^*)\pi(\tilde{\sigma}^{2*} \mid \tilde{\phi}^*, \tilde{Y}_T) \\ &\times \pi(P^*, Q^* \mid \tilde{\mu}^*, \tilde{\phi}^*, \tilde{\sigma}^2, \tilde{Y}_T). \end{aligned} \tag{3.8}$$

The above decomposition of the posterior density suggests that $\pi(\tilde{\mu}^* \mid \tilde{Y_T})$ may be calculated based on draws from the full Gibbs run, and that $\pi(\tilde{\phi}^* \mid \tilde{\mu}^*)$, $\pi(\tilde{\sigma}^{2*} \mid \tilde{\phi}^*, \tilde{Y}_T)$ and $\pi(P^*, Q^* \mid \tilde{\mu}^*, \tilde{\phi}^*, \tilde{\sigma}^2, \tilde{Y}_T)$ may be calculated based on draws from the reduced Gibbs runs. The following explains how each of these may be calculated based on output from appropriate Gibbs runs:

$$\begin{aligned} &\pi(\tilde{\mu}^* \mid \tilde{Y_T}) \\ &\quad = \frac{1}{G}\sum_{g=1}^{G} \pi(\tilde{\mu}^* \mid \tilde{\phi}^g, \tilde{\sigma}^{2g}, \tilde{D_T}^g, P^g, Q^g, \tilde{Y}_T), \end{aligned} \tag{3.9}$$

$$\begin{aligned} &\pi(\tilde{\phi}^* \mid \tilde{\mu}^*) \\ &\quad = \frac{1}{G}\sum_{g=1}^{G} \pi(\tilde{\phi}^* \mid \ \tilde{\mu}^*, \tilde{\sigma}^{2g1}, \tilde{S_T}^{g1}, \tilde{D_T}^{g1}, P^{g1}, Q^{g1}, \tilde{Y}_T), \end{aligned} \tag{3.10}$$

$$\begin{aligned} &\pi(\tilde{\sigma}^{2*} \mid \tilde{\phi}^*, \tilde{Y}_T) \\ &\quad = \frac{1}{G}\sum_{g=1}^{G} \pi(\tilde{\sigma}^* \mid \ \tilde{\mu}^*, \tilde{\phi}^*, \tilde{S_T}^{g2}, \tilde{D_T}^{g2}, P^{g2}, Q^{g2}, \tilde{Y}_T), \end{aligned} \tag{3.11}$$

$$\begin{aligned} &\pi(P^*, Q^* \mid \tilde{\mu}^*, \tilde{\phi}^*, \tilde{\sigma}^2, \tilde{Y}_T) \\ &\quad = \frac{1}{G}\sum_{g=1}^{G} \pi(P^*, Q^* \mid \ \tilde{\mu}^*, \tilde{\phi}^*, \tilde{\sigma}^* \tilde{S_T}^{g3}, \tilde{D_T}^{g3}, \tilde{Y}_T), \end{aligned} \tag{3.12}$$

where the superscript g refers to the $g-th$ draw from the full Gibbs run, and the superscript g_i, for $i = 1, 2, 3,$, refers to the $g_i - th$ draw from the appropriate, reduced Gibbs run. Thus, apart from the usual G iterations for the full Gibbs run, we need $3 \times G$ additional iterations for the appropriate, reduced Gibbs run.

Now we consider the derivation of the Bayes factor and the posterior probabilities. Suppose there are two hypotheses, M_k and M_l. With the prior odds, defined as $p(M_k)/p(M_l)$, which are often taken to be equal to one when it is unclear which hypothesis is correct, we can compute the posterior odds, which are given by

$$\frac{P(M_k \mid y)}{P(M_l \mid y)} = \frac{p(M_k)}{p(M_l)} \cdot \frac{p(y \mid M_k)}{p(y \mid M_l)},$$

where $p(y \mid M_k)/p(y \mid M_l)$ is referred to as the Bayes factor, which is derived from the marginal likelihood. When several models are being considered, the posterior odds yield posterior probabilities. Suppose that N models with $M_1, M_2, \cdots, M_N$ are being considered, and that each of the hypotheses, $M_1, M_2, \cdots, M_{N-1}$, is compared with M_N. Then, the posterior probability for model M_k is

$$PosteriorProbability(M_k \mid y) = \frac{P(M_k \mid y)/P(M_N \mid y)}{\sum_{l=1}^{N}[P(M_l \mid y)/P(M_N \mid y)]},$$

where Posterior Odds $P(M_N \mid y)/P(M_N \mid y) = 1$. In the next section, we evaluate and select a suitable model from among the candidates by using posterior probabilities and setting the prior odds equal to one.

4 Evidence of Multiple Changes of Business Cycle

In this section, we consider whether change-points exist and, if so, we specify their number and magnitude for the Japanese business cycle over the last three decades; we also identify the turning points, taking the change-points into consideration. To this end, we estimate four types of Markov-switching models: one without change-points, and models with one, two, and three such points. We then identify their number by selecting the most appropriate model from among the candidates using the Bayes factor. The data employed are the monthly series, containing 357 observations, of the ”Coincident Composite Index,” which is constructed by the ESRI from eleven coincident series, selected from macroeconomic variables between May 1974 and January 2004. We then take logs of, and first-difference, this variable, which we then multiply by 100 for expression as the percentage growth rate. The lag order, k, is chosen based on the Bayesian Schwarz Criterion (BSC) for the model with no changes, and is set at $k = 2$.

Bayesian inference requires specification of the prior distributions of the model parameters. Here, the orthodox priors are adopted, as in the following Tables. All inferences are implemented from the sampling of 10,000 iterations by Gibbs simulation, after discarding 5,000 iterations to mitigate the effects of initial conditions. Using these 10,000 draws for each of the parameters, we calculate the posterior means, the standard deviations of the posterior means, the 95 % bands, and the convergence diagnostic statistics proposed by Geweke (1992). The posterior means are calculated by averaging the simulated draws. The standard deviations of the posterior means are computed using a Parzen window with a bandwidth of 1000. The 95 % bands are calculated using the 2.5-th and 97.5-th percentiles of the simulated draws. The convergence of the Gibbs sampler can be assessed using the method proposed by Geweke (1992). He suggested a comparison of those values occurring early in the sequence with those occurring later in the sequence. Let $\theta^{(j)}$ be the $j - th$ draw of a parameter from among the recorded 10,000 draws, and let $\bar{\theta}_A = 1/n_A \sum_{j=1}^{n_A} \theta^{(j)}$ and $\bar{\theta}_B = 1/n_B \sum_{j=10,000-n_B}^{10,000} \theta^{(j)}$. Using these values, Geweke (1992) proposed the following statistic, called *a convergence diagnostic* (CD).

$$CD = \frac{\bar{\theta}_A - \bar{\theta}_B}{\sqrt{\hat{\sigma}_A^2/n_A + \hat{\sigma}_B^2/n_B}}, \tag{4.1}$$

where $\sqrt{\hat{\sigma}_A^2/n_A}$ and $\sqrt{\hat{\sigma}_B^2/n_B}$ are the standard deviations of $\bar{\theta}_A$ and $\bar{\theta}_B$. If the sequence of $\theta^{(j)}$ is stationary, it converges in distribution to the standard normal. We set $n_A = 1,000$ and $n_B = 5,000$ and compute σ_A^2 and σ_B^2, using Parzen windows with bandwidths of 100 and 500, respectively. In calculating the marginal likelihood, we set the number of iterations to 1,000 for the evaluation of the posterior densities.

The estimation result from each of the models is summarized in Tables 1 through 4. In addition, the posterior distributions of the change-points of each model, as well as both the average growth rates of the expansions and contractions and the probabilities of the contractions, are drawn in Figures 2 through 4. First, we evaluate the performance of the model without changes. Fig.1 (a) and (b) illustrate the two kinds of average growth rates during expansions and contractions, and the probabilities of contraction, respectively. For the entire sample period, i.e., from May 1974 to Jan 2004, the average growth rates per month during contractions and expansions, μ_{01} and μ_{11}, were about -1.2 % and 0.4 %, respectively. In addition, the variance σ_1^2 was about 1.04 (Table 1). Fig.1 (a) and (b) also illustrate the contraction periods reported by ESRI. Since the probabilities of contraction do not coincide well with the report of the ESRI, the recessions are not successfully isolated by the model, from 1977 to 1987, although they are so between 1974 and 1975, and after 1990, as can be seen from Fig.1 (b).

When the model has a single change-point, the change-point may be located near 1975, as the histogram in Fig.2 (b) illustrates. In particular, March 1975 is most likely the date on which the change occurred, since the unimodal posterior distribution indicates that its peak was located in March 1975. At this break, the mean of the monthly growth rate during recessions increased from -2.3 % to -0.9 %, an increase of 1.4 points. In addition, the mean during booms increased from -0.56 % to 0.55 %. In other words, the level of each of the two means seems to have jumped at the break point, while keeping the size of the gap between the means in booms and recessions unchanged. The variances, σ_1^2 and σ_2^2, leveled off around 1.0 (Table 2). Finally, Fig.2.(b) shows that, with the addition of the single change-point to the model, the estimated recessions can be isolated between 1977 and 1987, in contrast to the results for the model with no change.

In the case of the model with two points, the posterior distribution of the first change-point was bimodal; one mode was in March 1975, and the other was in March 1976, as illustrated in Fig.3 (b). The posterior distribution of the second point had quite a wide range, and its mode was located in March 1989. As can be seen from Table 3, the gap between the means of the expansions and contractions was the widest, out of the three regimes, (about 2.0 points) before the first change-point. This gap shrank by half, and the level of both means rose, in the second regime (Apr. 1976 through Mar. 1989). Following this, the gap again widened, to around 1.7 points, in the third regime (Apr. 1989 through Jan. 2004). The changes in volatility σ_i^2 were similar to those in the case of the means; σ_1^2, σ_2^2, and σ_3^2 were nearly 1.2, 0.5, and 1.1, respectively (Table 3). Fig.3 (b) also indicates that the probabilities of contraction estimated in the model including two change-points coincide well with the report of the ESRI for the whole of the sample period.

Fig.4 presents the results for the model with three points, in which the posterior distribution of the first point is located between 1975 and 1977, the second one overlaps that of the first, and the third point is distributed around 1989. The aspects of these distributions allow us to infer that only two change-points existed over the last thirty years. Since three change-points cannot be specified by the model, the range of estimation of the shift parameters, $\tilde{\mu}$, $\tilde{\sigma^2}$, becomes much wider than in the previous models. In particular, the standard deviations and 95 % band of σ_i^2, for $i = 1, 2, 3, 4$, are tremendous as Table 4. The recessions indicated by this model are similar to those suggested by the previous model, although the probabilities estimated are somewhat lower.

Table 5 presents the log-likelihood, the marginal log-likelihood, and the posterior probability, as explained in the previous section, for each model. The log-likelihood of the model with three points is the highest and is very close to that of the model with two points, whereas the likelihood of the other model is far away from these. However, the model with two points, instead of that with three, may be selected, with as much as 90 % of the posterior probability derived from the marginal probability. Consequently, we can specify that there were two change points in the business cycle during the whole period considered, from the point of view of Bayesian inference. Like Hamori and Bhar (2003), who used quarterly GDP series in their estimation, we also find two structural breaks, the first in 1975, after the first oil crisis, and the second around 1989, before the so-called "lost decade" of the 1990s, using the monthly coincident composite index series for the last three decades. In the first regime (May 1974 through Mar. 1976), the gap between booms and recessions, and the volatility around the means, were the largest for the three regimes. Between Apr. 1976 and Mar. 1989, the Japanese economy experienced a period of stable growth, during which this gap and the volatility shrank by half. After Apr. 1989, the economy was again unstable. Both the gap and the volatility increased at the close of the second regime.

Table 6 compares the precision with which the turning points are measured in each of the models. The dates of peaks (troughs) are derived from the date immediately before the probability of recession reaches more (less) than 50 %. When we do not take into consideration a sufficient number of change-points in the model, the turning points cannot be sufficiently established for the stable economic growth periods in the 1980s, in which there was a small gap between the growth rates of the expansions and contractions, as can be seen from the results of the models without changes and with one change. Meanwhile, in the case that the model includes too many change-points, the recession is likely to be as well identified as in the case with the appropriate number of points.

5 Summary and Conclusion

Using a Markov-switching model and Bayesian inference, the turning points of Japanese business cycles are measured from a monthly coincident composite index series, taken over the last thirty years. Ordinarily, in taking such a long-range estimation approach, we would face the following questions: (1) Have there been any structural changes? (2) If so, does the existence of these structural changes prevent detection of the turning points? (3) How many changes occurred? (4) When did these changes occur? While the classical approach is unable to deal with these problems, the Bayesian approach can easily provide answers. We specified the number and timing of the change-points by selecting the appropriate model, using the Bayes factor, from among four candidates: a model without points, and models with one, two, and three points.

The estimation results suggest that the Markov-switching model with no changes is unable to identify the turning points appropriately, whereas the model with changes selected via the Bayes factor robustly estimates these points for the long period of time considered, and also successfully facilitates estimation of the changes in amplitude that occurred between booms and recessions, as well as the volatility of business cycles. In the Japanese economy, there were two structural changes; these changes occurred in 1975, after the first oil crisis, and around 1989, before the so-called “lost decade” of the 1990s. The former break makes the amplitude and the volatility of the coincident index shrink, while the latter enlarges both.

Reference

[1] Andrew, D.W.K. (1993) “Tests for Parameter Instability and Structural Change With Unknown Change Point,” *Econometrica*, 61, 821-856.

[2] Andrews, D.W.K. and W. Ploberger (1994) “Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative,” *Journal of Econometrics*, 70, 9-38.

[3] Artis, M., H.-M. Krolzig and J. Toro (2004) “The European Business Cycle, ” *Oxford Economic Papers*, 56, 1-44.

[4] Bai, J. and P. Perron (1998) “ Estimating and Testing Linear Models with Multiple Structural Changes,” *Econometrica*, 66, 1, 47-78.

[5] Baxter, M. and R. King (1999) “Measuring Business Cycles: Approximate Band-Pass Filters For Economic Time Series, ” *Review of Economics and Economic Statistics* 81, 4, 575-593.

[6] Blanchard, O. and J. Simon (2001) “The Long and Large Decline in U.S. Output Volatility,” *Brooking Papers on Economic Activity*, 1, 135-174.

[7] Burns, A.F. and Mitchell, W.C. (1946) ”Measuring Business Cycles,” *NBER*, New York.

[8] Casella, G. and E. I. George (1993) ”Explaining the Gibbs Sampler,” *The American Statistician*, 46(3), 167-174.

[9] Chib, S. (1995) “Marginal Likelihood From the Gibbs Output”, *Journal of the American Statistical Asociation*, 90, 1313-1321.

[10] Chib, S. (1996) “Calculating Posterior Distributions and Model Estimates in Markov Mixture Models,” *Journal of Econometrics*, 75, 79-98.

[11] Chib, S. (1998) “Estimation and Comparison of Multiple Change-Point Models,” *Journal of Econometrics*, 86, 221-241.

[12] Geweke, J. (1992). ”Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments,” in J.M.Bernardo, J.O.Berger, A.P.Dawid and A.F.M.Smith (eds.) *Bayesian Statistics 4*, Oxford University Press, Oxford.

[13] Hamilton, J. D. (1989). ”A New Approach to the Economic Analysis of Nonstationary Time Series and Business Cycle,” *Econometrica* 57(2) 357-384.

[14] Hamori, S. and R. Bhar (2003) “Alternative Characterisation of Volatility in the Growth Rate of Real GDP,” *Japan and World Economy*, 15, 223-231.

[15] Hansen, B.E. (1997) “Approximate Asymptotic P Values for Structural-Change Tests,” *Journal of Business and Economic Statistics*, 15, 1, 60-67.

[16] Hodrick, R. and E. Prescott (1997) “ Post-War US Business Cycles: an Empirical Investigation, ” *Journal of Money, Credit, and Banking* , 29, 1-16.

[17] Kim, C.-J. and Nelson, C.R., (1999). "Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle, ," *Review of Economics and Economic Statistics* 81, 608-616.

[18] Kim, C.-J., C.R. Nelson and J. Piger (2004) “The Less-Volatile U.S. Economy: A Bayesian Investigation of Timing, Breadth, and Potential Explanations, ” *Journal of Business and Economic Statistics*, 22, 1, 80-93.

[19] Koop, G. and S.M. Potter (1999) “Bayes Factors and Nonlinearity: Evidence From Economic Time Series,” *Journal of Econometrics* 88, 251-281.

[20] McConnell, M.M. and G.P. Perez-Quiros (2000) “ Output Fluctuations in the United States: What Has Changed Since the Early 1980s? ” *American Economic Review*, 90, 1464-1476.

[21] Stock, J.H. and M.W. Watson (1991) “A New Index of Coincident Economic Indicators,” in K.Lahili and G.H. Moore (eds), *Leading Economic Indicators: New Approaches and Forecasting Records*, Cambridge University Press, Cambrodge, MA, 63-89.

[22] Stock, J.H. and M.W. Watson (2002) “Has the Business Cycle Changed and Why?” *NBER Macroeconomics Annual*, 17, 159-218.

Table 1. Model with No Change Point [a]

parameter[b]	Prior mean	Prior S.D.	Posterior Distribution mean.[c]	S.D.[d]	95%band[e]		CD[f]
ϕ_1	0	2	-0.078	0.061	(-0.193	0.043)	0.783
ϕ_2	0	2	0.200	0.062	(0.081	0.319)	0.501
μ_{01}	0	2	-1.212	0.242	(-1.716	-0.802)	1.435
μ_{11}	0	2	0.409	0.093	(0.232	0.593)	0.397
σ_1^2	1	1	1.037	0.092	(0.879	1.226)	0.101
p_{00}	0.8	0.163	0.887	0.055	(0.770	0.960)	-0.582
p_{11}	0.8	0.163	0.968	0.018	(0.935	0.989)	-1.247

Note [a] The first 5,000 iterations of Gibbs sampler are discarded to guarantee convergence and then the next 10,000 iterations are used for calculating the posterior means, the standard deviations (S.D.) of the posterior means, 95% band, and the convergence diagnostic (CD) satistics proposed by Geweke (1992).

[b] See eq.(1), eq.(9),eq.(10) and eq.(11) for the notations of parameters.

[c] The posterior means are computed by averaging the simulated draws.

[d] the standard deviations of the posterior means are computed using a Parzen window with a bandwidth of 1000.

[e] 95 % bands refers to 95 % posterior probability bands. This bands are calculated using the 2.5-th and 97.5-th percentiles of the simulated draws.

[f] CD is computed using eq.(26), where we set $n_A = 1,000$ and $n_B = 5,000$ and compute σ_A^2 and σ_B^2 using a Parzen window with bandwidths of 100 and 500, respectively.

[g] Prior distributions employed;

$\tilde{\phi} \sim N((0,0)', 4I_2)$, $\tilde{\mu} \sim N((0,0)', 4I_2)$, $1/\sigma_1^2 \sim Gamma(1,1)$, $p_{00} \sim Beta(4,1)$, $p_{11} \sim Beta(4,1)$,

Table 2. Model with One Change Point

	Prior		Posterior Distribution				
parameter	mean	S.D.	mean	S.D.	95%band		CD
ϕ_1	0	2	-0.185	0.063	(-0.306	-0.061)	-0.799
ϕ_2	0	2	0.088	0.063	(-0.032	0.212)	-0.910
μ_{01}	0	2	-2.384	0.490	(-3.603	-1.645)	0.610
μ_{02}	0	2	-0.860	0.132	(-1.151	-0.629)	0.890
μ_{11}	0	2	-0.504	1.171	(-2.359	1.490)	-0.258
μ_{12}	0	2	0.552	0.079	(0.388	0.701)	0.927
σ_1^2	1	1	1.185	0.809	(0.354	3.015)	0.832
σ_2^2	1	1	0.933	0.082	(0.789	1.107)	-1.222
p_{00}	0.8	0.163	0.916	0.030	(0.847	0.964)	-0.187
p_{11}	0.8	0.163	0.960	0.016	(0.923	0.985)	-0.323
q_{11} [a]	0.989	0.033	0.953	0.046	(0.831	0.998)	1.232

Note see the note of Table I.

[a] Prior distribution employed;

$q_{11} \sim Beta(9, 0.1)$, which is followed by Kim and Nelson (1999).

Table 3. Model with Two Change Points

parameter	Prior mean	Prior S.D.	Posterior mean	Posterior S.D.	95%band		CD
ϕ_1	0	2	-0.260	0.062	(-0.381	-0.137)	-1.526
ϕ_2	0	2	0.015	0.061	(-0.104	0.139)	-0.863
μ_{01}	0	2	-2.219	0.299	(-2.844	-1.670)	-0.921
μ_{02}	0	2	-0.367	0.135	(-0.629	-0.095)	-0.370
μ_{03}	0	2	-1.127	0.127	(-1.384	-0.887)	1.661
μ_{11}	0	2	0.212	0.918	(-2.009	1.270)	-0.956
μ_{12}	0	2	0.708	0.110	(0.487	0.928)	1.198
μ_{13}	0	2	0.527	0.081	(0.366	0.686)	0.058
σ_1^2	1	1	1.160	0.561	(0.384	2.501)	-1.068
σ_2^2	1	1	0.549	0.094	(0.392	0.758)	0.431
σ_3^2	1	1	1.094	0.128	(0.869	1.367)	-0.768
p_{00}	0.8	0.163	0.921	0.027	(0.860	0.965)	1.061
p_{11}	0.8	0.163	0.952	0.016	(0.915	0.979)	0.187
q_{11} [a]	0.989	0.033	0.961	0.037	(0.860	0.998)	-1.721
q_{22} [a]	0.989	0.033	0.993	0.007	(0.976	0.999)	0.011

Note see the note of Table I.

[a] Prior distribution employed;

$q_{11} \sim Beta(9, 0.1)$, $q_{22} \sim Beta(9, 0.1)$, which are followed by Kim and Nelson (1999).

Table 4. Model with Three Change Points

	Prior		Posterior Distribution				
parameter	mean	S.D.	mean	S.D.	95%band		CD
ϕ_1	0	2	-0.249	0.100	(-0.391	0.087)	-0.297
ϕ_2	0	2	0.034	0.094	(-0.101	0.335)	-0.505
μ_{01}	0	2	-2.058	0.531	(-2.777	-0.302)	-0.806
μ_{02}	0	2	-1.348	1.602	(-4.332	1.179)	1.589
μ_{03}	0	2	-0.404	0.881	(-2.533	0.832)	-0.543
μ_{04}	0	2	-0.990	0.583	(-1.403	0.318)	0.225
μ_{11}	0	2	0.198	0.817	(-1.806	1.648)	0.764
μ_{12}	0	2	1.623	1.576	(-2.073	4.534)	-1.328
μ_{13}	0	2	0.834	1.047	(0.029	3.655)	1.377
μ_{14}	0	2	0.630	0.639	(0.347	2.167)	0.856
σ_1^2	1	1	1.034	1.424	(0.322	2.621)	1.040
σ_2^2	1	1	3.069	29.290	(0.035	12.600)	-0.830
σ_3^2	1	1	1.487	21.504	(0.230	3.615)	1.243
σ_4^2	1	1	1.212	5.571	(0.823	1.536)	-0.005
p_{00}	0.8	0.163	0.925	0.029	(0.864	0.981)	-2.215
p_{11}	0.8	0.163	0.948	0.039	(0.864	0.979)	1.559
q_{11} [a]	0.989	0.033	0.957	0.044	(0.832	0.999)	-0.565
q_{22} [a]	0.989	0.033	0.927	0.075	(0.724	0.998)	1.916
q_{33} [a]	0.989	0.033	0.984	0.038	(0.866	0.999)	-1.684

Note see the note of Table I.

[a] Prior distribution employed;

$q_{11} \sim Beta(9, 0.1)$, $q_{22} \sim Beta(9, 0.1)$, $q_{33} \sim Beta(9, 0.1)$ which are followed by Kim and Nelson (1999).

Table 5.

Model Selection.

	$Model$	Log $Likelihood$	$Log\ Marginal$ $Likelihood$	$Posterior$ $Probability$
(a).	model with no change	-536.684	-555.993	0.0001
(b).	model with one change point	-529.607	-548.512	0.0907
(c).	model with two change points	-511.992	-546.208	0.9085
(d).	model with three change points	-511.055	-553.369	0.0007

Table 6 . Business Cycle Turning Points

$ESRI$	no change point model	one change point model	two change points model	three change points model
		$Peaks$		
77/1	-	-	-2	-2
80/2	+3	0	0	0
85/6	-	0	-7	-2
91/2	+1	-2	-1	0
97/5	+2	+1	+1	+1
00/10	+2	+2	+2	+2
		$Troughs$		
75/3	0	0	0	0
77/10	-	-	-2	-3
83/2	-27	-4	0	0
86/11	-	-3	0	0
93/10	+1	+2	+2	+2
99/1	-5	-2	-4	-5
02/1	-2	-1	-2	0

Note: The dates of peaks (troughs) are derived from the date immediately before the probability of recession cross over more (less) than 50 %.

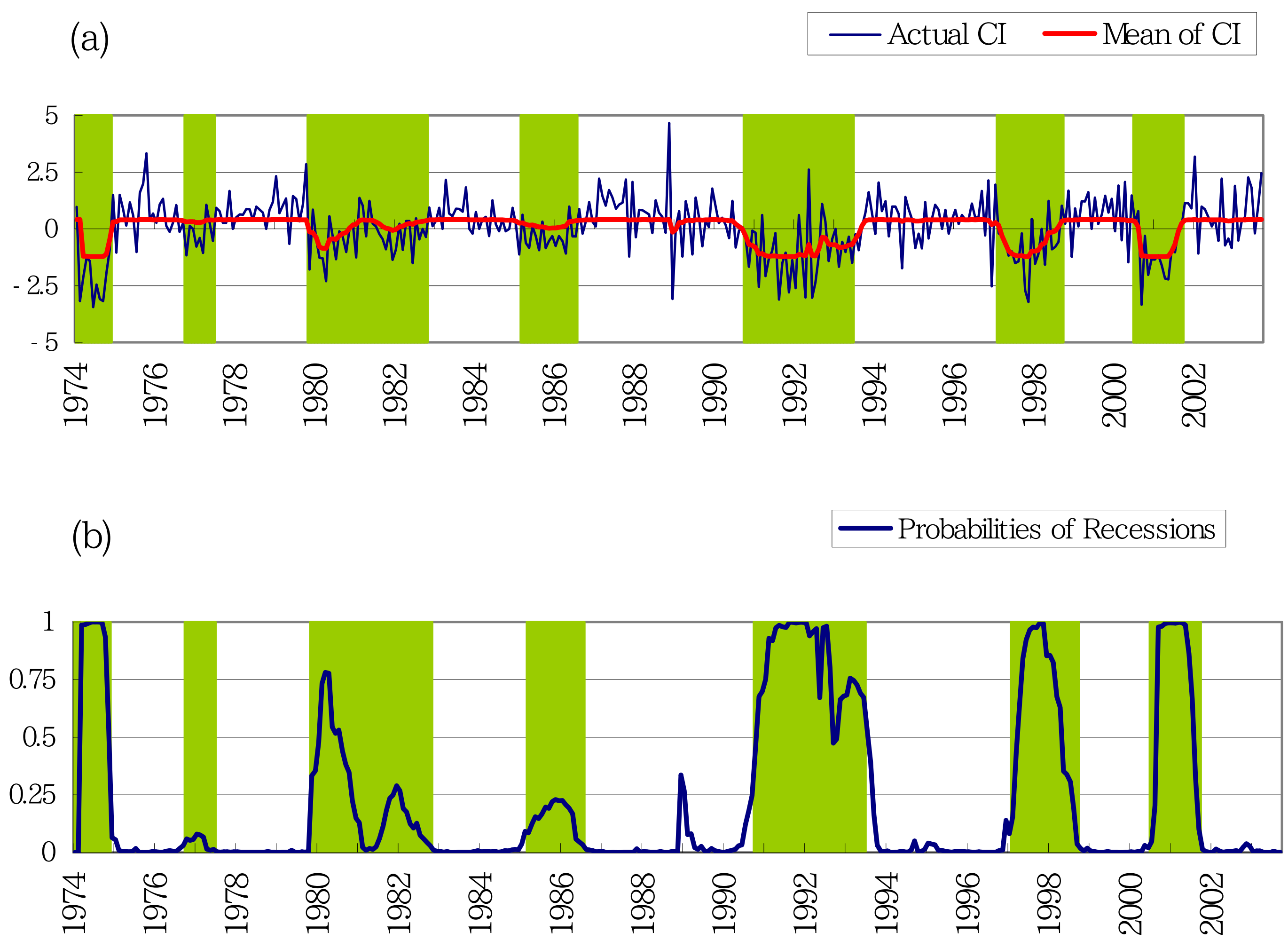

Fig. 1. Markov Switching Model with No Change

Note ; The shade represents recessions reported by ESRI.

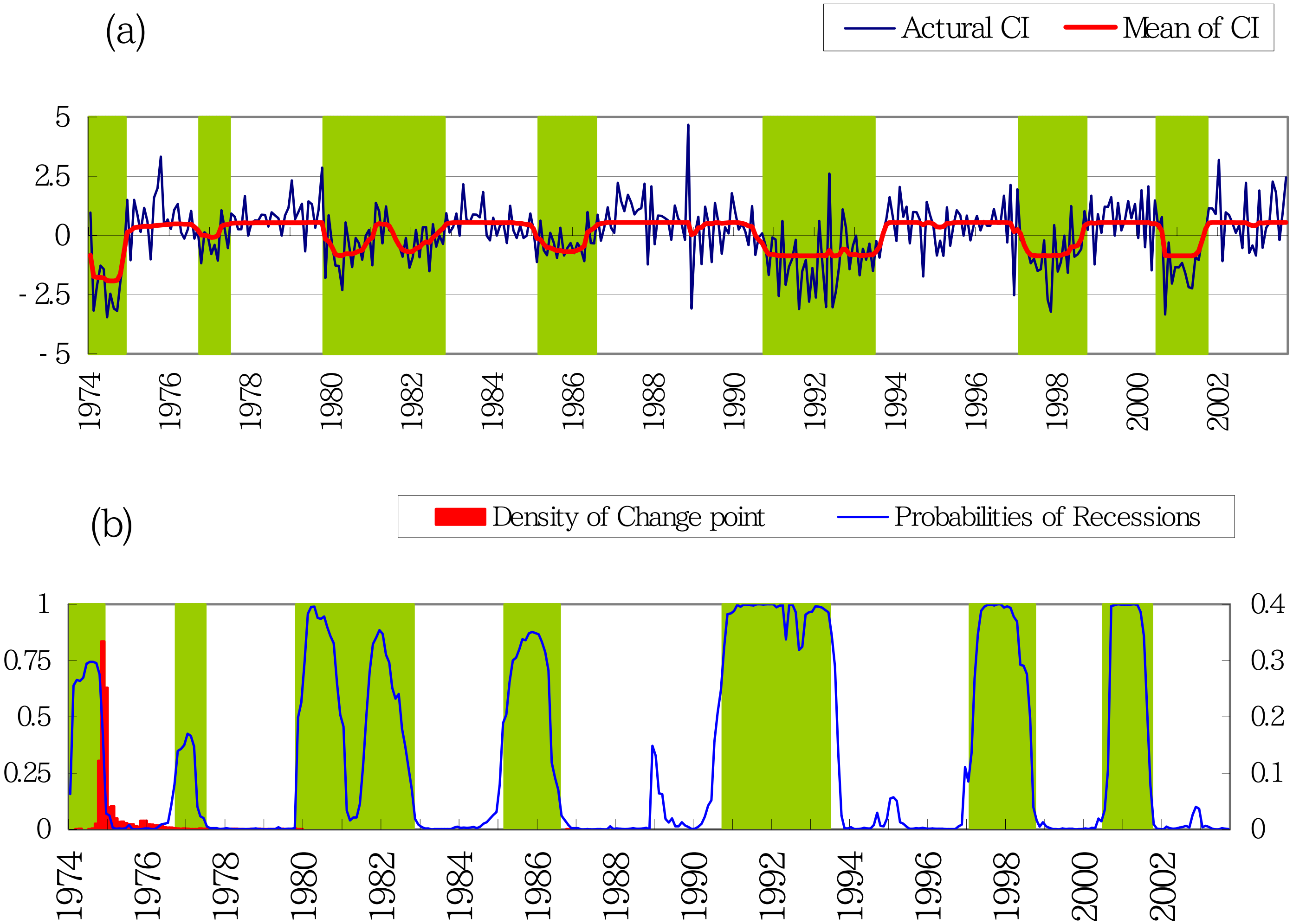

Fig. 2. Markov Switching Model with One Change

Note ; The shade represents recessions reported by ESRI.

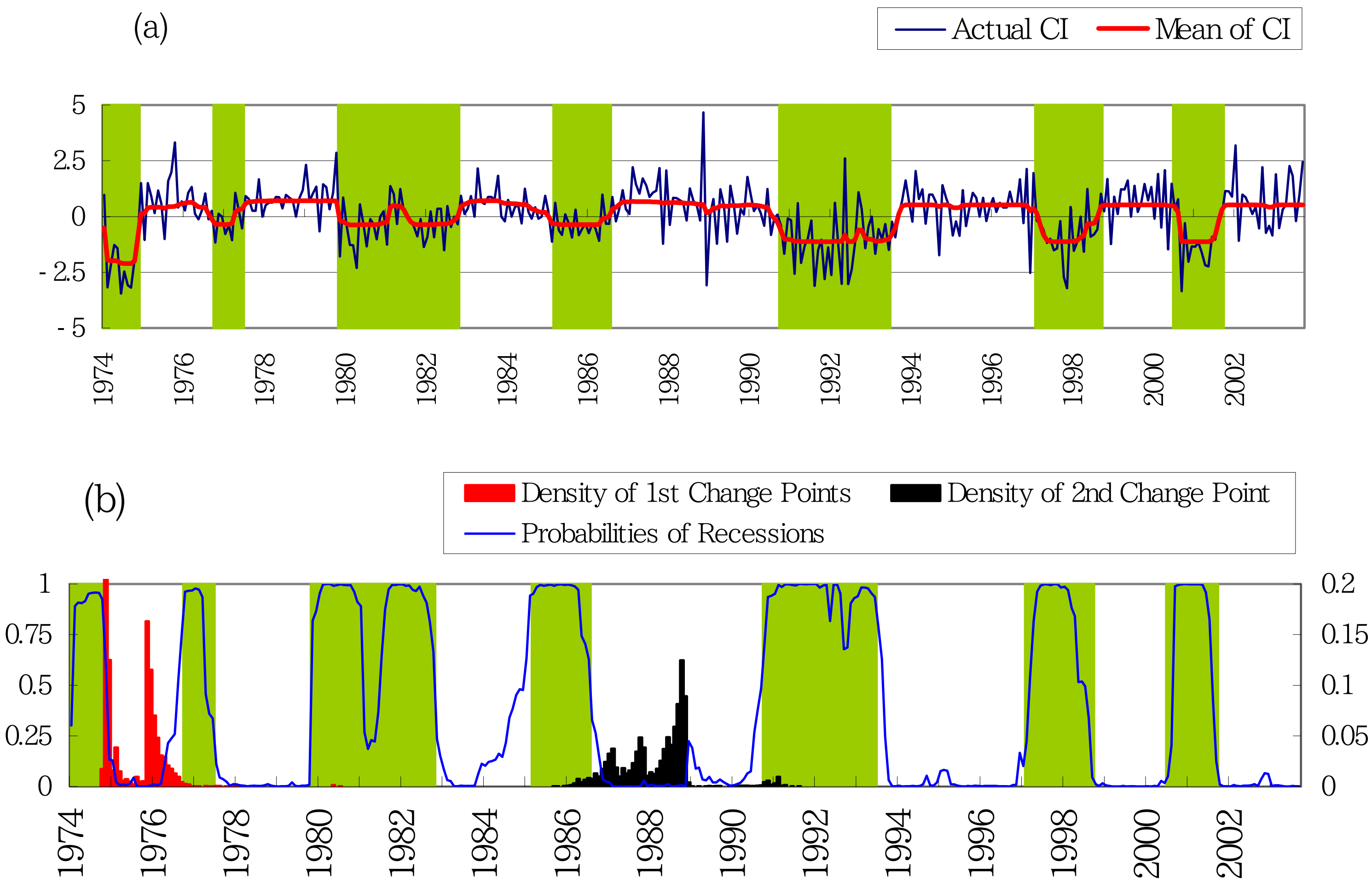

Fig. 3. Markov Switching Model with Two Changes

Note ; The shade represents recessions reported by ESRI.

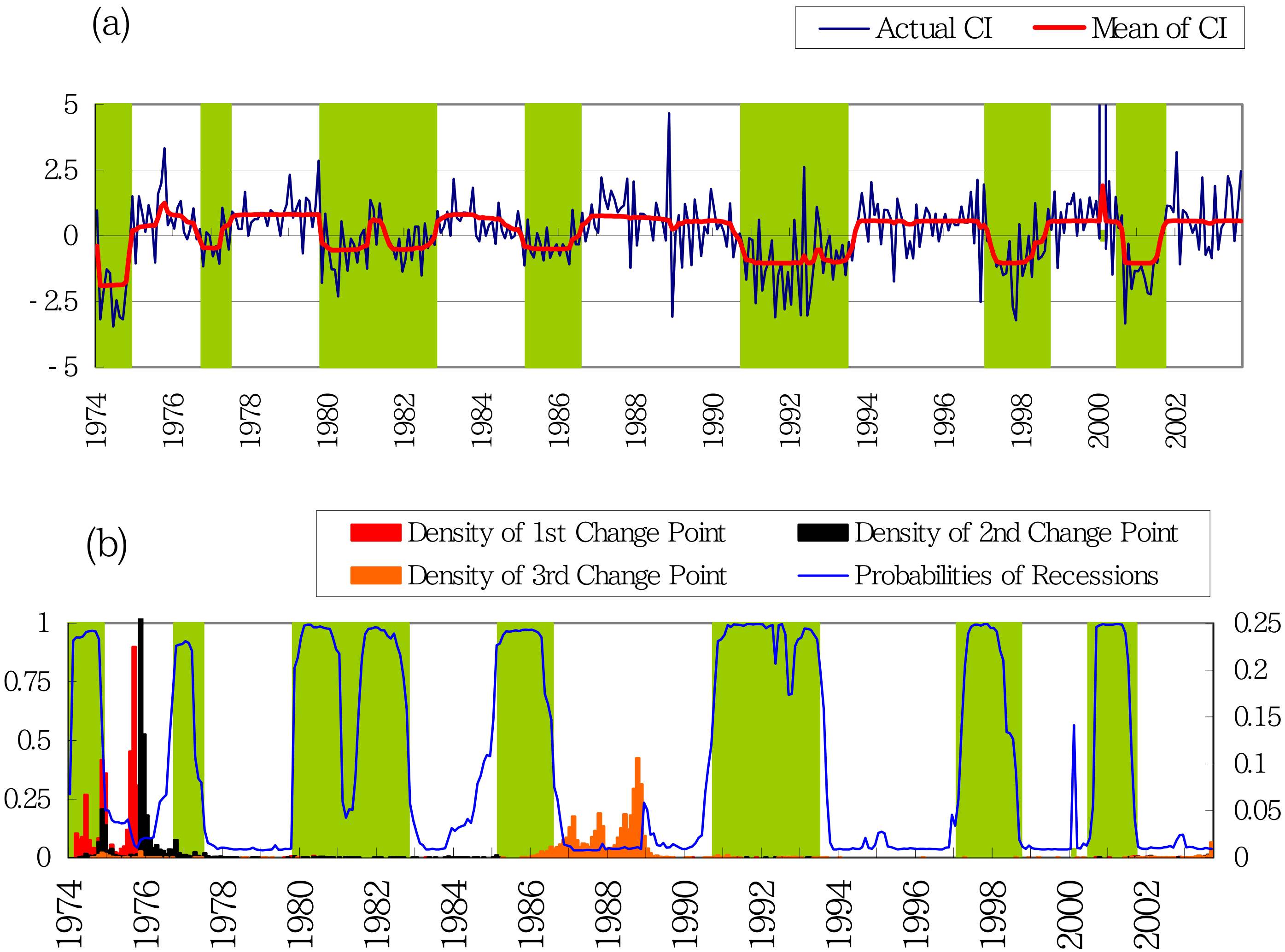

Fig. 4. Markov Switching Model with Three Changes

Note ; The shade represents recessions reported by ESRI.