University of Toronto
Department of Economics

Working Paper 427

Modelling Regime Switching and Structural Breaks with an Infinite Dimension Markov Switching Model

By Yong Song

April 15, 2011

Modelling Regime Switching and Structural Breaks with an Infinite Dimension Markov Switching Model *

Yong Song
University of Toronto
tommy.song@utoronto.ca

March 31, 2011

Abstract

This paper proposes an infinite dimension Markov switching model to accommodate regime switching and structural break dynamics or a combination of both in a Bayesian framework. Two parallel hierarchical structures, one governing the transition probabilities and another governing the parameters of the conditional data density, keep the model parsimonious and improve forecasts. This nonparametric approach allows for regime persistence and estimates the number of states automatically. A global identification algorithm for structural changes versus regime switching is presented. Applications to U.S. real interest rates and inflation compare the new model to existing parametric alternatives. Besides identifying episodes of regime switching and structural breaks, the hierarchical distribution governing the parameters of the conditional data density provides significant gains to forecasting precision.

*I am sincerely thankful to my supervisor, Professor John M. Maheu, for guidance and many helpful comments. I am also grateful to Martin Burda, Wing Chan, John Geweke, Christian Gourieroux, Robert Kohn, Thomas McCurdy, James Morley, Rodney Strachan and seminar attendants at Australian National University, Bank of Canada, CEA conference, CenSoC at University of Technology Sydney, University of Melbourne, University of New South Wales, University of Toronto, and Wilfrid Laurier University.

1 Introduction

This paper contributes to the current literature by accommodating regime switching and structural break dynamics in a unified framework. Current regime switching models are not suitable for capturing instability of dynamics because they assume a finite number of states and that the future is like the past. Structural break models allow the dynamics to change over time, however, they may incur loss in the estimation precision because the past states cannot recur and the parameters in each state are estimated separately. An infinite dimension Markov switching model is proposed to accommodate both types of model and provide much richer dynamics. The paper shows how to globally identify structural breaks versus regime switching. In applications to U.S. real interest rates and inflation, the new model performs better than the alternative parametric regime switching models and the structural break models in terms of in-sample fit and out-of-sample forecasts. The model estimation and forecasting are based on a Bayesian framework.

Regime switching models were first applied by Hamilton (1989) to U.S. GNP data. It is an important methodology to model nonlinear dynamics and is widely applied to economic data including business cycles (Hamilton, 1989), bull and bear markets (Maheu et al., 2010), interest rates (Ang and Bekaert, 2002) and inflation (Evans and Wachtel, 1993). There are two common features of these models. First, past states can recur over time. Second, the number of states is finite (it is usually 2 and at most 4). In the rest of the paper, a regime switching model is assumed to have both features. However, the second feature may cause biased out-of-sample forecasts if sudden changes of the dynamics exist.

In contrast to the regime switching models, structural break models can capture dynamic instability by assuming an infinite or a much larger number of states at the cost of extra restrictions. For example, Koop and Potter (2007) proposed a structural break model with an infinite number of states. If there is a change in the data dynamics, it will be captured by a new state. The restriction in their model is that the parameters in a new state are different from those in the previous ones. This condition is imposed for estimation tractability. However, it prevents the data divided by break points from sharing the same model parameter, and could incur some loss in estimation precision. In the current literature, structural break models such as Chib (1998); Pesaran et al. (2006); Wang and Zivot (2000) and Maheu and Gordon (2008) have the same feature as Koop and Potter (2007); namely that the states cannot recur. In the rest of the paper, a structural break model is assumed to have non-recurring states and an infinite or a large number of states.

As we can see, regime switching and structural break dynamics have different implications for data fitting and forecasting. What is missing in the current literature is

a method to reconcile them. For instance, a common practice is to use one approach or the other in applications to specific problems. Levin and Piger (2004) modelled U.S. inflation as a structural break process while Evans and Wachtel (1993) assumed a two-regime Markov switching model. Which feature is more important for inflation analysis, regime switching, structural breaks or both? Garcia and Perron (1996) used a three-regime Markov switching model for U.S. real interest rates while Wang and Zivot (2000) applied a model with structural breaks in mean and volatility. Did the real interest rates in 1981 have distinct dynamics or return to a historical state with the same dynamics? Existing econometric models have difficulty answering these questions.

This paper provides a solution by proposing an infinite dimension Markov switching model. It incorporates regime switching and structural break dynamics in a unified framework. Recurring states are allowed to improve estimation and forecasting precision. An unknown number of states is embedded in the infinite dimension structure and estimated endogenously to capture the dynamic instability. Different from the Bayesian model averaging methodology, this model combines different dynamics nonlinearly.

The proposed model builds on and extends Fox et al. (2008). They used a Dirichlet process[1] as a prior on the transition probabilities of an infinite hidden Markov switching model. The key innovation in their work is introducing a *sticky* parameter that favours state persistence and avoids the saturation of states. Their model is denoted by FSJW in the rest of the paper. Jochmann (2010) applies FSJW to investigate the structural breaks in the U.S. inflation dynamics.

The contributions of this paper are as follows. First, a second hierarchical structure in addition to FSJW is introduced to allow learning and sharing of information for the parameter of the conditional data density in each state. This approach is labelled as the sticky double hierarchical Dirichlet process hidden Markov model (SDHDP-HMM). Second, I present an algorithm to globally define structural breaks versus regime switching dynamics.[2] This is done by avoiding the label switching problem and focusing on labelling invariant posterior statistics. Lastly, the paper provides a detailed comparison of the new SDHDP-HMM against existing alternative regime switching and structural change models by out-of-sample density forecasting through a simulation study and two empirical applications to U.S. real interest rates and inflation. The results show that the SDHDP-HMM is robust to model uncertainty and superior in forecasting, and the hierarchical structure on the conditional data density parameters improves out-of-sample performance significantly.

In the application to U.S. real interest rates, the SDHDP-HMM is compared to the regime switching model by Garcia and Perron (1996) in a Bayesian framework

[1]The Dirichlet process is a commonly used prior in Bayesian nonparametric models.

[2]Jochmann (2010) proposes to identify structural breaks, but ignores recurring states in the posterior inference.

and the structural break model by Wang and Zivot (2000) with minor modifications. The results of the SDHDP-HMM supports Garcia and Perron's (1996) finding that the switching points occurred at the beginning of 1973 (the oil crisis) and the middle of 1981 (the federal budget deficit) instead of Huizinga and Mishkin's (1986) finding of October 1979 and October 1982 (both are monetary policy changes). The SDHDP-HMM also identifies two of the three turning points found by Wang and Zivot (2000). The model comparison based on the predictive likelihood shows regime switching dynamics dominates structural break dynamics for U.S. real interest rates.

The second application is to U.S. inflation. The SDHDP-HMM is compared to the regime switching model by Evans and Wachtel (1993) in a Bayesian framework and a structural break model by Chib (1998). This application shows that inflation has features of both regime switching and structural breaks. The SDHDP-HMM can capture both features and provide richer dynamics than existing parametric models. The predictive likelihoods further confirm that it is robust to model uncertainty and superior in forecasting.

The rest of the paper is organized as follows: Section 2 introduces the Dirichlet process to make this paper self-contained. Section 3 outlines the sticky double hierarchical Dirichlet process hidden Markov model and discusses its model structure and implications. Section 4 sketches the posterior sampling algorithm, explains how to identify the regime switching and the structural break dynamics, and describes the forecasting method. Section 5 compares the SDHDP-HMM to regime switching and structural break models through simulation. Section 6 studies the dynamics of U.S. real interest rate by revisiting the Markov switching model of Garcia and Perron (1996) in the Bayesian framework and the structural break model of Wang and Zivot (2000) with minor modification, and comparing them to the SDHDP-HMM using an extended data set. Section 7 applies the SDHDP-HMM to U.S. inflation, and compares it to Evans and Wachtel's (1993) Markov switching model in a Bayesian framework, Chib's (1998) structural break model and Fox et al.'s (2008) model. Section 8 concludes.

2 Dirichlet process

Before introducing the Dirichlet process, the definition of the Dirichlet distribution is the following:

Definition The **Dirichlet distribution** is denoted by $\mathbf{Dir}(\alpha)$, where α is a K-dimensional vector of positive values. Each sample x from $\mathbf{Dir}(\alpha)$ is a K-dimentional

vector with $x_i \in (0,1)$ and $\sum_{i=1}^{K} x_i = 1$. The probability density function is:

$$p(x \mid \alpha) = \frac{\Gamma(\sum_{i=1}^{K} \alpha_i)}{\prod_{i=1}^{K} \Gamma(\alpha_i)} \prod_{i=1}^{K} x_i^{\alpha_i - 1}$$

A special case is the **Beta distribution** denoted by $\mathbf{B}(\alpha_1, \alpha_2)$, which is a Dirichlet distribution with $K = 2$. ∎

Define $\alpha_0 = \sum_{i=1}^{K} \alpha_i$, X_i, the ith element of the random vector X from a Dirichlet distribution $\mathbf{Dir}(\alpha)$, has mean $\frac{\alpha_i}{\alpha_0}$ and variance $\frac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(\alpha_0+1)}$. Hence, we can further decompose α into two parts: a shape parameter $G_0 = (\frac{\alpha_1}{\alpha_0}, \cdots, \frac{\alpha_K}{\alpha_0})$ and a concentration parameter α_0. The shape parameter G_0 represents the center of the random vector X and the concentration parameter α_0 controls how close X is to G_0.

The Dirichlet distribution is conjugate to the multimonial distribution in the following sense: if

$$\begin{aligned} X &\sim \mathbf{Dir}(\alpha) \\ \beta = (n_1, \ldots, n_K) \mid X &\sim \mathbf{Mult}(X) \end{aligned}$$

where n_i is the number of occurrences of i in a sample of $n = \sum_{i=1}^{K} n_i$ points from the discrete distribution on $\{1, \cdots, K\}$ defined by X, then

$$X \mid \beta = (n_1, \ldots, n_K) \sim \mathbf{Dir}(\alpha + \beta).$$

This relationship is used in Bayesian statistics to estimate the hidden parameters X, given a collection of n samples. Intuitively, if the prior is represented as $\mathbf{Dir}(\alpha)$, then $\mathbf{Dir}(\alpha + \beta)$ is the posterior following a sequence of observations with histogram β.

The Dirichlet process was introduced by Ferguson (1973) as the extension of the Dirichlet distribution from a finite dimension to an infinite dimension. It is a distribution of distributions and has two parameters: the shape parameter G_0 is a distribution over a sample space Ω, and the concentration parameter α_0 is a positive scalar. They have similar interpretations as their counterparts in the Dirichlet distribution. The formal definition is the following:

Definition The Dirichlet process over a set Ω is a stochastic process whose sample path is a probability distribution over Ω. For a random distribution F distributed according to a Dirichlet process $\mathbf{DP}(\alpha_0, G_0)$, given any finite measurable partition

$A_1, A_2, \cdots, A_K$ of the sample space Ω, the random vector $(F(A_1), \cdots, F(A_K))$ is distributed as a Dirichlet distribution with parameters $(\alpha_0 G_0(A_1), \cdots, \alpha_0 G_0(A_K))$. ∎

Use the results from the Dirichlet distribution, for any measurable set A, the random variable $F(A)$ has mean $G_0(A)$ and variance $\frac{G_0(A)(1-G_0(A))}{\alpha_0+1}$. The mean implies the shape parameter G_0 represents the centre of a random distribution F drawn from a Dirichlet process $\mathbf{DP}(\alpha_0, G_0)$. We define $a_i \sim F$ as an observation drawn from the distribution F. Because by definition $P(a_i \in A \mid F) = F(A)$, we can derive $P(a_i \in A \mid G_0) = E(P(a_i \in A \mid F) \mid G_0) = E(F(A) \mid G_0) = G_0(A)$. Hence, the shape parameter G_0 is also the marginal distribution of an observation a_i. The variance implies the concentration parameter α_0 controls how close the random distribution F is to the shape parameter G_0. The larger α_0 is, the more likely F is close to G_0, and vice versa.

Suppose there are n observations, $a = (a_1, \cdots, a_n)$, drawn from the distribution F. Use $\sum_{i=1}^{n} \delta_{a_i}(A_j)$ to represent the number of a_i in set A_j, where $A_1, \cdots, A_K$ is a measurable partition of the sample space Ω and $\delta_{a_i}(A_j)$ is the Dirac measure, where

$$\delta_{a_i}(A_j) = \begin{cases} 1 & \text{if } a_i \in A_j \\ 0 & \text{if } a_i \notin A_j. \end{cases}$$

Conditional on $(F(A_1), \cdots, F(A_K))$, the vector $\left(\sum_{i=1}^{n} \delta_{a_i}(A_1), \cdots, \sum_{i=1}^{n} \delta_{a_i}(A_K)\right)$ has a multinomial distribution. By the conjugacy of the Dirichlet distribution to the multimomial distribution, the posterior distribution of $(F(A_1), \cdots, F(A_K))$ is still a Dirichlet distribution:

$$(F(A_1), \cdots, F(A_K)) \mid a \sim \mathbf{Dir}\left(\alpha_0 G_0(A_1) + \sum_{i=1}^{n} \delta_{a_i}(A_1), \cdots, \alpha_0 G_0(A_K) + \sum_{i=1}^{n} \delta_{a_i}(A_K)\right)$$

Because this result is valid for any finite measurable partition, the posterior of F is still a Dirichlet process by definition, with new parameters α_0^* and G_0^*, where

$$\alpha_0^* = \alpha_0 + n$$

$$G_0^* = \frac{\alpha_0}{\alpha_0 + n} G_0 + \frac{n}{\alpha_0 + n} \sum_{i=1}^{n} \frac{\delta_{a_i}}{n}$$

The posterior shape parameter, G_0^*, is the mixture of the prior and the empirical distribution implied by observations. As $n \to \infty$, the shape parameter of the posterior converges to the empirical distribution. The concentration parameter $\alpha_0^* \to \infty$ implies the posterior of F converges to the empirical distribution with probability one. Fergu-

son (1973) showed that a random distribution drawn from a Dirichlet process is almost surely discrete, although the shape parameter G_0 can be continuous. Thus, the Dirichlet process can only be used to model continuous distributions with approximation.

For a random distribution $F \sim \mathbf{DP}(\alpha_0, G_0)$, because F is almost surely discrete, it can be represented by two parts: different values for θ_i and their corresponding probabilities p_i, where $i = 1, 2, \cdots$. Sethuraman (1994) found the stick breaking representation of the Dirichlet process by writing $F \equiv (\theta, p)$, where $\theta \equiv (\theta_1, \theta_2, \cdots)'$, $p \equiv (p_1, p_2, \cdots)'$ with $p_i > 0$ and $\sum_{i=1}^{\infty} p_i = 1$. The $F \sim \mathbf{DP}(\alpha_0, G_0)$ can be generated by

$$V_i \overset{iid}{\sim} \mathbf{Beta}(1, \alpha_0) \tag{1}$$

$$p_i = V_i \prod_{j=1}^{i-1} (1 - V_j) \tag{2}$$

$$\theta_i \overset{iid}{\sim} G_0 \tag{3}$$

where $i = 1, 2, \cdots$. In this representation, p and θ are generated independently. The process generating p, (1) and (2), is called the stick breaking process. The name comes after the p_i's generation. For each i, the remaining probability, $1 - \sum_{j=1}^{i-1} p_j$, is sliced by a proportion of V_i and given to p_i. It's like breaking a stick an infinite number of times. This paper uses the notation $p \sim \mathbf{SBP}(\alpha_0)$ for this process.

The Dirichlet process was not widely used for continuous random variables until West et al. (1994) and Escobar and West (1995) proposed the Dirichlet process mixture model (DPM). A simple DPM model assumes the distribution of the random variable y is an infinite mixture of different distributions.

$$p \sim \mathbf{SBP}(\alpha_0) \tag{4}$$

$$\theta_i \overset{iid}{\sim} G_0 \quad \text{for } i = 1, 2, \cdots \tag{5}$$

$$g(y) = \sum_{i=1}^{\infty} p_i f(y \mid \theta_i) \tag{6}$$

where $g(y)$ is the probability density function of y and $f(y \mid \theta_i)$ is some probability density function depending on θ_i. For example, if $f(y \mid \theta_i)$ is the normal distribution density function and θ_i represents the mean and variance, y is distributed as an infinite mixture of normal distributions. Hence, continuous random variables can be modelled non-parametrically by the DPM model.

3 Sticky double hierarchical Dirichlet process hidden Markov model

The DPM model is used for cross sectional data in West et al. (1994), Escobar and West (1995) and Shahbaba and Neal (2009) because of the exchangeability of the observations. However, it is not appropriate for time series modelling because of its lack of state persistence. This paper extends the work of Fox et al. (2008) to propose the sticky double hierarchical Dirichlet process hidden Markov model as follows:

$$\pi_0 \sim \mathbf{SBP}(\gamma) \tag{7}$$

$$\pi_i \mid \pi_0 \sim \mathbf{DP}\left(c,\ (1-\rho)\pi_0 + \rho\delta_i\right) \tag{8}$$

$$\lambda \sim \mathcal{G} \tag{9}$$

$$\theta_i \overset{iid}{\sim} G_0(\lambda) \tag{10}$$

$$s_t \mid s_{t-1} = i \sim \pi_i \tag{11}$$

$$y_t \mid s_t = j, Y_{t-1} \sim f(y_t \mid \theta_j, Y_{t-1}) \tag{12}$$

where $i, j = 1, 2, \cdots$, and $Y_t = (y_1, \cdots, y_t)'$ represents the data up to time t.

(7) and (8) comprise the *first* hierarchical structure which governs the transition probabilities. π_0 is the hierarchical distribution drawn from the stick breaking process with parameter γ and represents a discrete distribution with support on the natural numbers. Each infinite dimensional vector π_i is drawn from a Dirichlet process with the concentration parameter c and the shape parameter $(1-\rho)\pi_0 + \rho\delta_i$, which is a convex combination of the hierarchical distribution π_0 and a degenerate distribution at integer j. There are three points worth noticing for clarity. First, because the shape parameter $(1-\rho)\pi_0 + \rho\delta_i$ has support only on natural numbers and each number is associated with non-zero probability, the random distribution π_i can only take values of the natural numbers and each value will receive positive probability by the stick breaking representation. When combining the same values and sorting them in ascending order, each π_i will have π_{ij} representing the probability of taking integer j. So we can use the vector $\pi_i = (\pi_{i1}, \pi_{i2}, \cdots)'$ to represent a distribution drawn from $\mathbf{DP}(c, (1-\rho)\pi_0 + \rho\delta_i)$. Second, π_i is the infinite dimension vector of transition probabilities given the past state $s_{t-1} = i$ by (11); the probability of transition from state i to state j is π_{ij}. Stacking π_is to construct the infinite dimensional transition matrix $P = (\pi_1', \pi_2', \cdots)'$ gives the hidden Markov model representation. Lastly, if ρ is larger, π_i is expected to have a larger probability at integer i. This implies s_t, the state at time t, is more likely to be the same as s_{t-1}. Hence, ρ captures state persistence. In the rest of the paper, ρ is referred as the sticky coefficient.

(9) and (10) comprise the *second* hierarchical structure which governs the parameters of the conditional data density. $G_0(\lambda)$ is the hierarchical distribution from which the state dependent parameter θ_i is drawn independently; $\mathcal{G}$ is the prior of λ. This structure provides a way of learning λ from past values of θ_i to improve estimation and forecasting. If a new state is born, the conditional data density parameter θ_{new} is drawn from $G_0(\lambda)$. Without the *second* hierarchical structure, the new draw θ_{new} depends on some assumed prior. Pesaran et al. (2006) argued the importance of modelling the hierarchical distribution for the conditional data density parameters in the presence of structural breaks. This paper adopts their method to estimate the hierarchical distribution $G_0(\lambda)$.

In comparison to the SDHDP-HMM, FSJW is comprised of (7)-(8) and (10)-(12). The stick breaking representation of the Dirichlet process is not fully explored by FSJW, since it has only one hierarchical structure on the transition probabilities. In fact, the stick breaking representation (1)-(3) decomposes the generation of a distribution F from a Dirichlet process into two independent parts: the probabilities are generated from a stick breaking process and the parameter values are independently generated from the shape parameter. The SDHDP-HMM takes fuller advantage of this structure than FSJW by modelling two parallel hierarchical structures.

The SDHDP-HMM can be summarized as an infinite dimension Markov switching model with a specific prior. Conditional on the hierarchical distribution π_0 and the sticky coefficient ρ, the mean of the transition matrix is

$$E(P \mid \pi_0, \rho) = (1-\rho) \cdot \begin{bmatrix} \pi_{01} & \pi_{02} & \pi_{03} & \cdots \\ \pi_{01} & \pi_{02} & \pi_{03} & \cdots \\ \pi_{01} & \pi_{02} & \pi_{03} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} + \rho \cdot \begin{bmatrix} 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

The sticky coefficient ρ captures the state persistence by adding weights to the diagonal elements of the transition matrix. The concentration parameter c controls how close P is to $E(P \mid \pi, \rho)$.

The common practice of setting the prior on the transition matrix of a Markov switching model assumes each row of the transition matrix is drawn from a Dirichlet distribution independently. If extended to the infinite dimension, each row π_i should be drawn from a stick breaking process. However, Teh et al. (2006) argued this prior may have an overparametrization problem without a hierarchical structure similar to (7) and (8), because it precludes each π_i from sharing information between each other. In terms of parsimony, the SDHDP-HMM only needs one stick breaking process for the hierarchical distribution π_0, instead of assuming an infinite number of the stick breaking processes for the whole transition matrix P. In other words, the hierarchical structure

on the transition probabilities collapses setting the prior on the infinite dimension matrix P to the infinite dimension vector π_0.

The SDHDP-HMM is also related to the DPM model (4)-(6), because (12) can be replaced by

$$y_t \mid s_{t-1} = i, Y_{t-1} \sim \sum_{j=1}^{\infty} \pi_{ij} f(y_t \mid \theta_j, Y_{t-1}). \tag{13}$$

On one hand, the DPM representation implies the SDHDP-HMM is nonparametric. On the other hand, in contrast to the DPM model, the mixture probability π_{ij} is state dependent. This feature allows the SDHDP-HMM to capture time varying dynamics.

In summary, the SDHDP-HMM is an infinite state space Markov switching model with a specific form of prior to capture state persistence. Two parallel hierarchical structures are proposed to provide parsimony and improve forecasting. It preserves the nonparametric methodology of the DPM model but has state dependent probabilities in its mixture components.

4 Estimation, inference and forecasting

In the following simulation study and applications, the conditional dynamics $y_t \mid \theta_j, Y_{t-1}$ in (12) is set as a Gaussian AR(q) process:

$$y_t \mid \theta_j, Y_{t-1} \sim \mathbf{N}(\phi_{j0} + \phi_{j1} y_{t-1} + \cdots + \phi_{jq} y_{t-q},\ \sigma_j^2).$$

By definition, the conditional data density parameter is $\theta_i = (\phi_i', \sigma_i)'$ with $\phi_i = (\phi_{i0}, \phi_{i1}, \cdots, \phi_{iq})'$.

The hierarchical distribution $G_0(\lambda)$ in (10) is assumed as the regular normal-gamma distribution in the Bayesian literature.[3] The conditional data density parameter θ_i is generated as follows:

$$\sigma_i^{-2} \sim \mathbf{G}\left(\chi/2, \nu/2\right), \quad \phi_i \mid \sigma_i \sim \mathbf{N}\left(\phi, \sigma_i^2 H^{-1}\right) \tag{14}$$

By definition, $\lambda = (\phi, H, \chi, \nu)$. ϕ is a $(q+1) \times 1$ vector, H is a $(q+1) \times (q+1)$ positive definite matrix, and χ and ν are positive scalars. It is a standard conjugate prior for linear models. Precision parameter σ_i^{-2} is drawn from a gamma distribution with degree of freedom $\nu/2$ and multiplier $\chi/2$. Given the hierarchical distribution parameter λ, the conditional mean and variance of σ_i^{-2} are ν/χ and $2\nu/(\chi)^2$, respectively. And $\phi_i \mid \sigma_i$ is drawn from a multivariate normal distribution with mean ϕ and covariance

[3]For example, see Geweke (2009).

matrix $\sigma_i^2 H^{-1}$.

The prior on the hierarchical parameters λ in (9) follows Pesaran et al. (2006):

$$H \sim \mathbf{W}(A_0, a_0) \tag{15}$$

$$\phi \mid H \sim \mathbf{N}(m_0, \tau_0 H^{-1}) \tag{16}$$

$$\chi \sim \mathbf{G}(d_0/2, c_0/2) \tag{17}$$

$$\nu \sim \mathbf{Exp}(\rho_0). \tag{18}$$

H is drawn from a Wishart distribution with parameters of a $(q+1) \times (q+1)$ positive definite matrix A_0 and a positive scalar a_0. Samples from this distribution are positive definite matrices. The expected value of H is $A_0 a_0$. The variance of H_{ij}, the ith row and jth colomn element of H, is $a_0(A_{ij}^2 + A_{ii}A_{ij})$, where A_{ij} is the ith row and jth column element of A_0. m_0 is a $(q+1) \times 1$ vector representing the mean of ϕ, and τ_0 is a positive scalar, which controls the prior belief of the dispersion of ϕ. χ is distributed as a gamma distribution with the multiplier $d_0/2$ and the degree of freedom $c_0/2$. ν has an exponential distribution with parameter ρ_0,

The posterior sampling is based on Markov chain Monte Carlo (MCMC) methods. Fox et al. (2008) showed the sampling scheme of FSJW. The block sampler based on approximation by a finite number of states is more efficient than the individual sampler.[4]

In order to apply the block sampler following Fox et al. (2008), the SDHDP-HMM is approximated by a finite number of states proposed as follows:

$$\pi_0 \sim \mathbf{Dir}\left(\frac{\gamma}{L}, \cdots, \frac{\gamma}{L}\right) \tag{19}$$

$$\pi_i \mid \pi_0 \sim \mathbf{Dir}\left((1-\rho)c\pi_{01}, ..., (1-\rho)c\alpha\pi_{0i} + \rho c, \cdots, (1-\rho)c\pi_{0L}\right) \tag{20}$$

$$\lambda \sim \mathcal{G} \tag{21}$$

$$\theta_i \overset{iid}{\sim} G_0(\lambda) \tag{22}$$

$$s_t \mid s_{t-1} = i \sim \pi_i \tag{23}$$

$$y_t \mid s_t = j, Y_{t-1} \sim \mathbf{N}(\phi_{j0} + \phi_{j1} y_{t-1} + \cdots + \phi_{jq} y_{t-q},\ \sigma_j^2) \tag{24}$$

where L is the maximal number of states in the approximation and $i = 1, 2, \cdots, L$. The hierarchical distribution $G_0(\lambda)$ and its prior are set as (14) and (15)-(18), respectively.

From the empirical point of view, the essence of the SDHDP-HMM is not only its infinite dimension, but also its sensible hierarchical structure of the prior. If L is large enough, the finite approximation (19)-(24) is equivalent to the original model (7)-(12)

[4]Consistency of the approximation was proved by Ishwaran and Zarepour (2000), and Ishwaran and Zarepour (2002). Ishwaran and James (2001) compared the individual sampler with the block sampler and found the latter to be more efficient in terms of mixing.

in practice.

4.1 Estimation

Appendix A shows the detailed posterior sampling algorithm. The parameter space is partitioned into four parts: (S, I), (Θ, P, π_0), (ϕ, H, χ) and ν. S, I and Θ are the collections of s_t, a binary auxiliary variable I_t and θ_i, respectively.[5] Each part is sampled conditional on the other parts and the data Y as follows:

1. Sample $(S, I) \mid \Theta, P, Y$
   (a) Sample $S \mid \Theta, P, Y$ by the forward and backward smoother in Chib (1996).
   (b) Sample $I \mid S$ by a Polya Urn scheme.
2. Sample $(\Theta, P, \pi_0) \mid S, I, Y$
   (a) Sample $\Theta \mid S, Y$ by regular linear model result.
   (b) Sample $\pi_0 \mid I$ by a Dirichlet distribution.
   (c) Sample $P \mid \pi_0, S$ by Dirichlet distributions.
3. Sample $(\phi, H, \chi) \mid S, \Theta, \nu$
   (a) Sample $(\phi, H) \mid S, \Theta$ by conjugacy of the Normal-Wishart distribution.
   (b) Sample $\chi \mid \nu, S, \Theta$ by a gamma distribution.
4. Sample $\nu \mid \chi, S, \Theta$ by a Metropolis-Hastings algorithm.

After initiate the parameter values, the algorithm is applied iteratively many times to obtain a large sample of the model parameters. The first block of samples is discarded to remove dependence on the initial values. The rest of the sample, $\{S^{(i)}, \Theta^{(i)}, P^{(i)}, \pi_0^{(i)}, \phi^{(i)}, H^{(i)}, \chi^{(i)}, \nu^{(i)}\}_{i=1}^{N}$ are used for inferences as if they were drawn from the posterior distribution. Simulation consistent posterior statistics are computed as sample averages. For example, the posterior mean of ϕ, $E(\phi \mid Y)$, is calculated by $\frac{1}{N}\sum_{i=1}^{N}\phi^{(i)}$.

Fox et al. (2008) did not consider the label switching problem, which is an issue in mixture models.[6] For example, switching the values of (θ_j, π_j) and (θ_k, π_k), swapping the values of state s_t for $s_t = j, k$, while keeping the other parameters unchanged, will result in the same likelihood value in the finite approximation of the SDHDP-HMM. Inferences on a label dependent statistic such as θ_j are misleading without extra constraints. Geweke (2007) showed that inappropriate constraints can also result in misleading inferences. To identify regime switching and structural breaks, this paper uses label invariant statistics. So the posterior sampling algorithm can be implemented without modification as suggested by Geweke (2007).

[5] I_t is an auxiliary variable for sampling of π_0. The details are in the appendix A.
[6] See Celeux et al. (2000), Frühwirth-Schnatter (2001) and Geweke (2007)

4.2 Identification of regime switching and structural breaks

A heuristic illustration of how an SDHDP-HMM nests different dynamics, including regime switching and structural breaks, is plotted in figure 1. Each path comprised by arrows is one sample path of state S in an SDHDP-HMM. Figure 1a represents the no state change case (the Gaussian AR(q) model from the assumption). Figures 1b-1d are the regime switching, structural break and frequent parameter change cases, respectively. Figure 1e captures more complicated dynamics, in which some states are only visited for one consecutive period while others are not.

The current literature does not study the identification of regime switching and structural breaks in infinite dimension Markov switching models. This paper proposes a global identification algorithm to identify regime switching and structural breaks based on whether a state is recurrent or not. In detail, if a state only appears for one consecutive period, it is classified as a non-recurrent state. Otherwise, it is defined as recurrent. The starting time of a recurrent (non-recurrent) state is identified as a regime switching (structural break) point. In figure 2, states 1 and 4, marked with circles, are non-recurrent states and the starting points of these two segments are identified as structural breaks. States 2 and 3, marked with triangles, are recurrent states. The starting time of each consecutive period is identified as a regime switching point.

In detail, if there exist time t_0 and t_1 (without loss of generality, let $t_0 \leq t_1$) such that $s_t = j$ if and only if $t_0 \leq t \leq t_1$, then state j is non-recurrent and t_0 is identified as a break point. On the other hand, if $s_{t_0} \neq s_{t_0-1}$ and t_0 is not a break point, then t_0 is identified as a regime switching point.

This identifcation criteria is simply because, in general, states are recurrent in the regime switching models but non-recurrent in structural break models. There are two points worth noticing. First, in terms of mathematical statistics, a recurrent (non-recurrent) state in a Markov chain is defined as a state which will be visited with probability one (less than one) in the future. This paper defines the recurrence (non-recurrence) as a statistic on one realized posterior sample path of the state variable S. Because the mathematical definition is not applicable to the estimation with a finite sample size, there should be no confusion between these two concepts. Second, a true path of states from a regime switching model can have non-recurrent states because of randomness or a small sample size. For example, states 2, 3 and 4 in figure 2 can be generated from a three-regime switching model. The algorithm identifies state 4 as a non-recurrent states, and its starting point is classified as a break point. Hence, this identification approach may label a switching point of a regime switching model as a structural break even if the true states were observed. However, this is simply accidental. As more data are observed, an embedded regime switching model will have

all its states identified as recurrrent.

More importantly, the purpose of the identification is not to decompose the infinite dimension Markov switching model into several regime switching and structural break sub-models (there is no unique way even if we wanted to), but to study the richer dynamics which allow recurrent states while accommodating structural breaks. Even if a non-recurrent state was generated from a regime switching model, it usually has different implication from the recurrent states of the same model.

Hence, separating the recurrent and non-recurrent states is both empirically reasonable and theoretically consistent with the definition of regime switching and structural breaks of the existing respective models. In the rest of the paper, the SDHDP-HMM associates regime switching and structural breaks to recurrent and non-recurrent states.

4.3 Forecast and model comparison

Predictive likelihood is used to compare the SDHDP-HMM to the existing regime switching and structural break models. It is similar to the marginal likelihood by Kass and Raftery (1995). Conditional on an initial data set Y_t, the predictive likelihood of $Y_{t+1}^T = (y_{t+1}, \cdots, y_T)$ by model M_i is calculated as

$$p(Y_{t+1}^T \mid Y_t, M_i) = \prod_{\tau=t+1}^{T} p(y_\tau \mid Y_{\tau-1}, M_i). \tag{25}$$

It is equivalent to the marginal likelihood $p(Y_T \mid M_i)$ if $t = 0$.

The calculation of one-period predictive likelihood of model M_i, $p(y_t \mid Y_{t-1}, M_i)$, is

$$\hat{p}(y_t \mid Y_{t-1}, M_i) = \frac{1}{N} \sum_{i=1}^{N} f(y_t \mid \Upsilon^{(i)}, Y_{t-1}, M_i) \tag{26}$$

where $\Upsilon^{(i)}$ is one sample of parameters from the posterior distribution conditional on the historical data Y_{t-1}. For the SDHDP-HMM, (26) is

$$\hat{p}(y_t \mid Y_{t-1}) = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{L} \pi_{jk}^{(i)} f(y_t \mid \theta_k^{(i)}, s_{t-1}^{(i)} = j, Y_{t-1}).$$

After the calculation of the one-period predictive likelihood, $\hat{p}(y_t \mid Y_{t-1})$, the data is updated by adding one observation, y_t, and the model is re-estimated for the prediction of the next period. This is repeated until the last predictive likelihood, $\hat{p}(y_T \mid Y_{T-1})$, is obtained.

Kass and Raftery (1995) compared model M_i and M_j by the difference of their log marginal likelihood: $\log(BF_{ij}) = \log(Y \mid M_i) - \log(Y \mid M_j)$. They suggested

interpreting the evidence for M_i versus M_j as: not worth more than a bare mention for $0 \leq \log(BF_{ij}) < 1$; positive for $1 \leq \log(BF_{ij}) < 3$; strong for $3 \leq \log(BF_{ij}) < 5$; and very strong for $\log(BF_{ij}) \geq 5$. BF_{ij} is referred as the Bayes factor of M_i versus M_j. This paper uses this criteria for model comparison by predictive likelihood. Geweke and Amisano (2010) showed the interpretation is the same as Kass and Raftery (1995) if we regard the initial data Y_t as a training sample.

5 Simulation evidence

To investigate how the SDHDP-HMM reconciles the regime switching and the structural break models, this section provides some simulation evidence based on three models: the SDHDP-HMM, a finite Markov switching model, and a structural break model. Each model simulates a data set of 1000 observations, and all three data sets are estimated by a SDHDP-HMM with the same prior. First, I plot the posterior means of the conditional data density parameters $E(\theta_{s_t} \mid Y_T)$ and the true values θ_{s_t} over time. If the SDHDP-HMM fits the model well, the posterior means should be close to the true ones. Second, more rigorous study is based on the predictive likelihoods. Each of the three models are estimated on each of the three simulated data sets. The last 100 observations are used to calculate the predictive likelihood. If the SDHDP-HMM is able to accommodate the other two models, its predictive likelihood based on the data simulated from the alternative model should be close to the predictive likelihood estimated by the true model; and if the SDHDP-HMM provides richer dynamics than the other two models, its predictive likelihood based on the data simulated from the SDHDP-HMM should strongly dominates the predictive likelihoods calculated by the other two models.

The parameters of the SDHDP-HMM in the simulation are set as: $\gamma = 3, c = 10, \rho = 0.9, \chi = 2, \nu = 2, \phi = \mathbf{0}$ and $H = \mathbf{I}$. The number of AR lags is set as 2. The simulation is done through the Polya-Urn scheme without approximation as in Fox et al. (2009). The simulated data are plotted in figure 3.

The first competitor is a K-state Markov switching model as follows:

$$(p_{i1}, \cdots, p_{iK}) \sim \mathbf{Dir}(a_{i1}, \cdots, a_{iK}) \tag{27}$$

$$(\phi_i, \sigma_i) \overset{iid}{\sim} G_0 \tag{28}$$

$$Pr(s_t = j \mid s_{t-1} = i) = p_{ij} \tag{29}$$

$$y_t \mid s_t = j, Y_{t-1} \sim \mathbf{N}(\phi_{j0} + \phi_{j1} y_{t-1} + \cdots + \phi_{jq} y_{t-q}, \sigma_j^2) \tag{30}$$

where $i, j = 1, \cdots, K$. Each AR process uses 2 lags as in the SDHDP-HMM. The number of states, K, is set as 3. Conditional data density parameters are $\phi_1 = (0, 0.8, 0)$,

$\phi_2 = (1, -0.5, 0.2)$, $\phi_3 = (2, 0.1, 0.3)$ and $(\sigma_1, \sigma_2, \sigma_3) = (1, 0.5, 2)$. The transition matrix is set as $P = \begin{bmatrix} 0.96 & 0.02 & 0.02 \\ 0.02 & 0.96 & 0.02 \\ 0.02 & 0.02 & 0.96 \end{bmatrix}$. The simulated data are plotted in figure 5. The prior of each row of the transition matrix, $(p_{j1}, \cdots, p_{jK})$, is set as independent Dirichlet distribution $\mathbf{Dir}(1, \cdots, 1)$. The prior of the conditional data density parameters G_0 is set as the normal-gamma distribution, where $\sigma_i^{-2} \sim \mathbf{G}(1, 1)$ and $\phi_i \mid \sigma_i \sim \mathbf{N}\left(\mathbf{0}, \sigma_i^2 \mathbf{I}\right)$.

The second competitor is a K-state structural break model from Chib (1998):

$$p \sim \mathbf{B}(a_p, b_p) \tag{31}$$

$$Pr(s_t = i \mid s_{t-1} = i) = \begin{cases} p & \text{if } i < K \\ 1 & \text{if } i = K \end{cases} \tag{32}$$

$$Pr(s_t = i + 1 \mid s_{t-1} = i) = 1 - p \quad \text{if } i < K \tag{33}$$

$$(\phi_i, \sigma_i) \overset{iid}{\sim} G_0 \quad \text{for } i = 1, \cdots, K \tag{34}$$

$$y_t \mid s_t = i, Y_{t-1} \sim \mathbf{N}(\phi_{i0} + \phi_{i1} y_{t-1} + \cdots + \phi_{iq} y_{t-q}, \sigma_k^2) \tag{35}$$

where $i = 1, 2, \cdots, K$ is the state indicator. The break probability $1 - p$ and the number of AR lags are set as 0.003 and 2, respectively. In the simulation, the $K = 4$ and the parameters of the conditional data density are $\phi_1 = (0, 0.8, 0)$, $\phi_2 = (1, -0.5, 0.2)$, $\phi_3 = (0.5, 0.1, 0.3)$, $\phi_4 = (0, 0.5, 0.2)$ and $(\sigma_1, \sigma_2, \sigma_3, \sigma_4) = (1, 0.5, 1, 0.5)$. The simulated data are plotted in figure 8. $K = 5$ is used in the estimation to nest the true data generating process. The prior of p is set as a beta distribution $\mathbf{B}(9, 1)$, and G_0 is set in the same way as the Markov switching model of (28).

All of the three simulated data sets are estimated by the SDHDP-HMM. The parameters γ, c, ρ and the number of AR lags are set in the same way as in the SDHDP-HMM used in the simulation. The maximal number of states, L, is assumed as 10. The priors on the other parameters are weakly informative as follows: $H \sim \mathbf{W}(0.2\mathbf{I}, 5)$, $\phi \mid H \sim \mathbf{N}(\mathbf{0}, H^{-1})$, $\chi \sim \mathbf{G}(0.5, 0.5)$ and $\nu \sim \mathbf{Exp}(1)$.

The intercept, the persistence parameter (sum of AR coefficients), the standard deviation and the cumulative number of active states of the simulated data from the SDHDP-HMM over time are plotted in figure 4 using solid lines. The posterior means of those parameters from the estimation are also plotted for comparison in the same figure using dashed lines. It is not surprising that the estimated values tracks the true ones closely and sharply identifies the change points. Because the estimation is based on the finite approximation, while the simulation is based on the true data generating process, the results support the validity of the block sampler.

Figure 6 plots the true values of the intercept, the persistence, the volatility and the cumulative number of switching of the simulated data from the Markov switching

model over time using solid lines. It also includes the posterior means of these parameters estimated from the SDHDP-HMM marked with dashed lines. Figure 7 plots the true and the posterior mean of the regime switching and structural break probabilities implied by the SDHDP-HMM. The SDHDP-HMM sharply identifies almost all the switching points. From the middle panel, the global identification does not find prominent strauctural breaks.

Figure 9 plots the true parameters from the data simulated from the structural break model using solid lines and the posterior mean of those parameters estimated from the SDHDP-HMM using dashed lines. Again, the SDHDP-HMM tracks different parameters closely. Figure 10 plots the true and the posterior mean of the structural break and regime switching probabilities. The SDHDP-HMM identifies all the break points. The bottom panel shows some small probabilities of regime switching around the structural break points. Those values are very small compared to the structural break probabilities.

A more rigorous model comparison can be found in table 1. It shows the log predictive likelihoods of the last 100 observations estimated by all of the above three models on all of the three simulated data sets. The SDHDP-HMM is robust to model misspecification because it is not strongly rejected against the true model by the log predictive likelihoods. For example, if the true data generating process is the Markov switching model, the log predictive likelihoods computed by the true model and the SDHDP-HMM are -208.10 and -208.32, respectively. The difference is only $-208.10-(-208.32) = 0.22 < 1$, which is not worth more than a bare mention. On the other hand, both the Markov switching model and the structural break model are strongly rejected if the other one is the true model. For example, if the structural break model is the data generating process, the log predictive likelihoods calculate by the true model and the Markov switching model are -178.41 and -187.26. Their difference is $-178.41-(-187.26) = 8.85 > 5$, which is very strong against the misspecified model.

In addition to its robustness, the SDHDP-HMM is also able to capture more complicated dynamics than the Markov switching model and the structural break model. If the SDHDP-HMM is the true data generating process, the Markov switching model and the structural break model are both rejected strongly. The log predictive likelihood of the SDHDP-HMM is 12.75 larger than the Markov switching model and 91.4 larger than the structural break model. Both values are greater than 5.

In summary, the simulation evidence shows the SDHDP-HMM is robust to model uncertainty. Both of the Markov switching model and the structural break model can be tracked closely. Meanwhile, SDHDP-HMM provides richer dynamics than the other two types of models.

6 Application to U.S. real interest rate

The first application is to U.S. real interest rates. Previous studies by Fama (1975); Rose (1988) and Walsh (1987) tested the stability of their dynamics. While Fama (1975) found the *ex ante* real interest rate as a constant, Rose (1988) and Walsh (1987) cannot reject the existence of an integrated component. Garcia and Perron (1996) reconciled these results using a three-regime Markov switching model and found switching points at the beginning of 1973 (the oil crisis) and the middle of 1981 (the federal budget deficit) using quarterly U.S. real interest rates of Huizinga and Mishkin (1986) from 1961Q1-1986Q3. The real interest rate dynamics in each state are characterized by an Gaussian AR(2) process. Wang and Zivot (2000) used the same data to investigate structural breaks and found support of four states (3 breaks) by Bayes factors.

This paper constructs U.S. quarterly real interest rates in the same way as Huizinga and Mishkin (1986) and extends their data set to a total of 252 observations from 1947Q1 to 2009Q4. The last 200 observations are used for predictive likelihood calculation. Alternative models for comparison include the Markov switching model of Garcia and Perron (1996) put in a Bayesian framework, the structural break model of Wang and Zivot (2000) with minor modifications and linear AR models. All but the linear model have the Gaussian AR(2) process in each state as in Garcia and Perron (1996) and Wang and Zivot (2000).

The priors of the SDHDP-HMM are set as follows:

$$
\begin{aligned}
\pi_0 &\sim \mathbf{Dir}(1/L, \cdots, 1/L) \\
\pi_i \mid \pi_0 &\sim \mathbf{Dir}(\pi_{01}, \cdots, \pi_{0i} + 9, \cdots, \pi_{0L}) \\
H &\sim \mathbf{W}(0.2\ \mathbf{I}, 5) \\
\phi \mid H &\sim \mathbf{N}(\mathbf{0}, H^{-1}) \\
\chi &\sim \mathbf{G}(0.5, 2.5) \\
\nu &\sim \mathbf{Exp}(5)
\end{aligned}
$$

where $i = 1, \cdots, L$. The block sampler uses the truncation of $L = 10$.[7]

The Markov switching model used is (27)-(30). Garcia and Perron (1996) estimated the model in the classical approach and this paper revisits their paper in the Bayesian framework. The prior of each row of the transition matrix, $(p_{i1}, \cdots, p_{iK})$, is set as $\mathbf{Dir}(1, \cdots, 1)$. The priors of ϕ_i and σ_i are $\sigma_i^{-2} \sim \mathbf{G}(2.5, 0.5)$ and $\phi_i \mid \sigma_i \sim \mathbf{N}(\mathbf{0}, \sigma_i^2 \cdot \mathbf{I})$.

The structural break model is (31)-(35). The model proposed in this paper allows simultaneous breaks of the intercept, the AR coefficients and the volatility, while Wang

[7] $L = 10$ is chosen to represent a potentially large number of states and keep a reasonable amount of computation. Some larger L's are also tried and produce similar results.

and Zivot (2000) only allowed the intercept and the volatility to change. The prior of p is a beta distribution $\mathbf{B}(9,1)$, and parameters ϕ_i and σ_i have the same priors as the Markov switching model.

A linear AR model is applied as a benchmark for model comparison:

$$(\phi, \sigma) \sim G_0 \tag{36}$$

$$y_t \mid Y_{t-1} \sim \mathbf{N}(\phi_0 + \phi_1 y_{t-1} + \cdots + \phi_q y_{t-q}, \sigma^2) \tag{37}$$

where the prior of σ is set in the same way as in the Markov switching model and the structural break model. The prior of $\phi \mid \sigma$ is $\mathbf{N}(\mathbf{0}, \sigma^2 \cdot \mathbf{I})$, where the dimension of vector $\mathbf{0}$ and the identity matrix $\mathbf{I}$ depends on the number of lags q in the AR model.

Table 2 shows the log predictive likelihoods of the different models. Firstly, the table shows that all linear models are dominated by nonlinear models. Secondly, the log predictive likelihoods strongly support the Markov switching models against the structural break models. The log predictive likelihood of the four-regime or five-regime Markov switching model is larger than that of any K-regime structural break models by more than 5, which is very strong based on Kass and Raftery (1995). Lastly, although the SDHDP-HMM does not strongly dominate the Markov switching models, it still performs the best among all the models. This is consistent with the simulation evidence that the SDHDP-HMM can provide robust forecasts by optimally combining regime switching and structural breaks in the Bayesian framework.

The whole sample is estimated by the SDHDP-HMM with the same prior as in the predictive likelihood calculation. Figure 11 plots the posterior mean of different parameters over time, including the regime switching and structural break probabilities. There is no sign of structural breaks from the bottom panel, so the regime switching dynamics prevail over the structural break dynamics, which is consistent with table 2 based on the predictive likelihoods. Three important regimes are found in the figure: one has high volatility and high persistence, one has low volatility and intermediate persistence and the last one has intermediate volatility and low persistence.

Figure 12 plots the posterior mean of the cumulative number of active states over time. A state is defined as active if it is occupied by data. The posterior mean of the total number of active states is 3.4. Compared to the truncation of $L = 10$ in the estimation, this value implies that the finite truncation restriction is not binding, so the nonparametric flavor is preserved.

Garcia and Perron (1996) found switching points at the beginning of 1973 and the middle of 1981. In the SDHDP-HMM, the probability of regime switching in 1973Q1 is 0.39, which is consistent with their finding. From 1980Q2 to 1981Q1, the probabilities of regime switching are 0.18, 0.13, 0.32 and 0.19, respectively. There are many uncertainties in the switching point identification at these times. However, it

is quite likely that the state changed in one of these episodes, which is only slightly earlier than in Garcia and Perron (1996). On the other hand, Huizinga and Mishkin (1986) identified October 1979 and October 1982 as the turning points. Probabilities of regime switching or structural breaks in 1979Q3 and Q4 are less than 0.02 and 0.04 respectively, while in 1982Q3 and 1982Q4 they are both less than 0.01. Thus, the SDHDP-HMM supports Garcia and Perron (1996) against Huizinga and Mishkin (1986).

As an attempt to locate potential state changing points, I define a time with the sum of regime switching and structural break probability greater than 0.3 as a candidate turning point. There are 9 points in total: 1952Q1, 1952Q3, 1956Q2, 1958Q2, 1973Q1, 1980Q4, 1986Q2, 2002Q1, and 2005Q3. Among those points, 1973Q1 and 1980Q4 are consistent with Garcia and Perron (1996). Wang and Zivot (2000) found 1970Q3, 1980Q2 and 1985Q4 as structural break points. 1980Q4 and 1986Q2 are close to their finding. However, the SDHDP-HMM does not identify late 1970 as neither a break nor a switching point, which contradicts their result.

In summary, by using a larger sample, U.S. real interest rates are better described by a regime switching model than a structural break one. The robustness of the SDHDP-HMM to model uncertainty is supported by the predictive likelihoods. The SDHDP-HMM performs better than all the parametric alternatives in forecasting.

7 Application to U.S. inflation

The second application is to the U.S. inflation. Ang et al. (2007) studied the performance of different methods including time series models, Phillips curve based models, asset pricing models and surveys. The regime switching model is the best in their most recent sub-sample. Evans and Wachtel (1993) applied a two-regime Markov switching model to explain consistent inflation forecast bias. Their model incorporated a random walk model of Stock and Watson (1991) in one regime and a stationary AR(1) model in another. Structural breaks in inflation were studied by Groen et al. (2009); Levin and Piger (2004) and Duffy and Engle-Warnick (2006). Application of the SDHDP-HMM can reconcile these two types of models and provide more description of the inflation dynamics.

Monthly inflation rates are constructed from U.S. Bureau of Labor Statistics based on CPI-U. There are 1152 observations from Feb 1914 to Jan 2010. They are computed as annualized monthly CPI-U growth rates scaled by 100. The alternative models for comparison include the FSJW, the regime switching model of Evans and Wachtel (1993), a structural break model from Chib (1998) and linear Gaussian AR(q) models.

For the SDHDP-HMM, each state has Gaussian AR(1) dynamics. $L = 10$ and the

priors are:

$$\begin{aligned}
\pi_0 &\sim \mathbf{Dir}(1/L, \cdots, 1/L) \\
\pi_i \mid \pi_0 &\sim \mathbf{Dir}(\pi_{01}, \cdots, \pi_{0i} + 9, \cdots, \pi_{0L}) \\
H &\sim \mathbf{W}(0.2\ \mathbf{I},\ 5) \\
\phi \mid H &\sim \mathbf{N}(\mathbf{0},\ H^{-1}) \\
\chi &\sim \mathbf{G}(0.5,\ 2.5) \\
\nu &\sim \mathbf{Exp}(5)
\end{aligned}$$

with $i = 1, \cdots, L$.

In FSJW, each state has Gaussian AR(1) dynamics and the number of states $L = 10$, as in the SDHDP-HMM, to use the block sampler. The priors of the transition probabilities are the same as in the SDHDP-HMM. The prior on the parameters of conditional data density is normal-gamma: $\sigma_i^{-2} \sim \mathbf{G}(0.5, 0.5)$ and $\phi_i \mid \sigma_i \sim \mathbf{N}(\mathbf{0}, \sigma_i^2\ \mathbf{I})$.

For comparison, the structural break model of (31)-(35) is also applied with the number of the AR lags equal to 1. The prior of p is a beta distribution $\mathbf{B}(9, 1)$; and the priors of ϕ_i and σ_i are the same as in FSJW.

Another alternative model is the regime switching model of Evans and Wachtel (1993):

$$\begin{aligned}
P(s_t = i \mid s_{t-1} = i) &= p_i \\
(\phi_0, \sigma_0) &\sim G_0 \\
\sigma_1 &\sim G_1 \\
y_t \mid s_t = 0, Y_{t-1} &\sim \mathbf{N}(\phi_{00} + \phi_{01} y_{t-1}, \sigma_0^2) \\
y_t \mid s_t = 1, Y_{t-1} &\sim \mathbf{N}(y_{t-1}, \sigma_1^2)
\end{aligned}$$

where $i = 1, 2$. The prior of the self-transition probability, p_i, is a beta distribution $\mathbf{B}(9, 1)$. ϕ_0, σ_0, and σ_1 have the same priors as FSJW and the structural break model.

The linear AR model of (36) and (37) is applied as a benchmark for model comparison. The prior of σ is set the same as in FSJW, the Markov switching model and the structural break model. The prior of $\phi \mid \sigma$ is $\mathbf{N}(\mathbf{0}, \sigma^2 \cdot \mathbf{I})$, where the dimension of the vector $\mathbf{0}$ and the identity matrix $\mathbf{I}$ depends on the number of lags q in the AR model.

The last 200 observations are used to calculate the log predictive likelihoods. The results are shown in table 3. First, the linear models are strongly dominated by the nonlinear models. Second, the regime switching model of Evans and Wachtel (1993) strongly dominates the structural break models. Third, FSJW strongly dominates all the other parametric alternatives including the regime switching model. The difference

between the log predictive likelihoods of FSJW and the regime switching model is $-82.45-(-92.50)=6.05$, which implies heuristically FSJW is $\exp(6.05)\approx 424$ times better than the Evans and Wachtel (1993) model. Last, The SDHDP-HMM is the best model in terms of the log predictive likelihood. The difference of the log likelihoods of the SDHDP-HMM and FSJW is $-74.07-(-82.45)=8.38$, which implies the SDHDP-HMM is $\exp(8.34)\approx 4188$ times better than FSJW. Because the SDHDP-HMM nest the parametric alternatives, its dominance can be attributed to the fact that both the regime switching and the structural break dynamics are important for inflation, and each single type of the parametric model alone can not capture its dynamics.

The models are estimated on the whole sample. The posterior summary statistics are located in table 4. The posterior mean of the persistence parameter is 0.97 with a 95% density interval of $(0.742, 1.199)$, which implies the inflation dynamics are likely to be persistent in a new state. On the other hand, FSJW draws the parameters of the conditional data density for each new state from the prior assumption. This key difference contributes to the superior forecasting ability of the SDHDP-HMM to FSJW.

The smoothed means of conditional data density parameters, break probabilities and switching probabilities over time for the SDHDP-HMM are in figure 13. The instability of the dynamics is consistent with Jochmann (2010). The last panel plots the structural breaks and regime switching probabilities at different times. There are two major breaks at 1920-07 and 1930-05. The structural break and regime switching probabilities of 1920-07 are 0.3 and 0.5, respectively. There is quite a large chance for this time to have unique dynamics different from other periods. For 1930-05, the structural break and regime switching probabilities are 0.13 and 0.09. This implies that if the state changed at this time, it would be more likely to be a structural break.

To illustrate the dominance of the regime switching dynamics over the structural break dynamics, figure 14 plots the probabilities of past states to be the same as the last period, Jan 2010, or $p(z_\tau = z_{201001} \mid Y)$. Most of the positive probabilities are before 1955. This emphasizes the importance of modelling recurrent states in forecasting. Structural break models perform worse than the SDHDP-HMM and the regime switching model because they drop much useful information.

Figure 15 plots the smoothed regression coefficients, standard deviations and break probabilities over time estimated by the structural break model with $K = 10$. Structural breaks happened in the first half of the sample, therefore the recent regime switching implied by the SDHDP-HMM is not identified.

Figure 16 plots the smoothed probabilities of the random walk state and the smoothed volatility estimated by the regime switching model of Evans and Wachtel (1993) over time. The random walk dynamics dominate after 1953. In recent times, inflation dynamics entered into the stationary AR(1) state. This is consistent with the

SDHDP-HMM evidence shown in figure 14 that the most recent episodes are associated with data before 1955. In another word, there is a regime switching back to the same state in the past.

In figure 17, all regime switching and structural break probabilities are plotted for comparison. The first panel is the regime switching model; the second panel is the structural break model and the last is the SDHDP-HMM. Two features can be summarized from the figure. First, the change points identified by the structural break model and the regime switching model are associated with the change points identified by the SDHDP-HMM. Second, the SDHDP-HMM estimates more turning points than each of the alternative models. This implies it captures some dynamics that can not be identified by the regime switching or the structural break models alone. Together with the log predictive likelihood results in table 3, the inflation shows both regime switching and structural break features.

In summary, the regime switching and the structural break dynamics are both important for inflation modelling and forecasting. The SDHDP-HMM is able to capture both of these features. In the SDHDP-HMM, the parameters of the conditional data density in each state can provide information for the learning of the hierarchical distribution $G_0(\lambda)$ and significantly improve forecasting.

8 Conclusion

This paper proposes to apply an infinite dimension Markov switching model labelled as the sticky double hierarchical Dirichlet process hidden Markov model (SDHDP-HMM) to accommodate regime switching and structural break dynamics. Two parallel hierarchical structures, one governing the transition probabilities and the other governing the parameters of the conditional data density, are imposed for parsimony and to improve forecasts. An algorithm for the global identification of regime switching and structural breaks is proposed based on label invariant statistics. A simulation study shows the SDHDP-HMM is robust to model uncertainty and able to capture more complicated dynamics than the regime switching and the structural break models.

Applications to U.S. real interest rates and inflation show the SDHDP-HMM is robust to model uncertainty and provides better forecasts than regime switching and structural break models. The second hierarchical structure on the data density parameters provides significant improvement in inflation forecasting. From both the predictive likelihood results and the posterior probabilities of regime switching and structural breaks, U.S. real interest rates are better described by a regime switching model while inflation has both features of regime switching and structural breaks.

A Block sampling

A.1 Sample $(S, I) \mid \Theta, P, Y$

$S \mid \Theta, P, Y$ is sampled by the forward and backward smoother in Chib (1996).

I is introduced to facilitate the π_0 sampling. From (19) and (20), the filtered distribution of π_i conditional on $S_t = (s_1, \cdots, s_t)$ and π_0 is a Dirichlet distribution:

$$\pi_i \mid S_t, \pi_0 \sim \mathbf{Dir}\left(c(1-\rho)\pi_{01} + n_{i1}^{(t)}, \cdots, c(1-\rho)\pi_{0i} + c\rho + n_{ii}^{(t)}, \cdots, c(1-\rho)\pi_{0L} + n_{iL}^{(t)}\right)$$

where $n_{ij}^{(t)}$ is the number of $\{\tau \mid s_\tau = j, s_{\tau-1} = i, \tau \leq t\}$. Integrate out π_i, the conditional distribution of s_{t+1} given S_t and π_0 is:

$$p(s_{t+1} = j \mid s_t = i, S_t, \pi_0) \propto c(1-\rho)\pi_{0j} + c\rho\delta_i(j) + n_{ij}^{(t)}$$

Construct a variable I_t with a Bernoulli distribution:

$$p(I_{t+1} \mid s_t = i, S_t, \pi_0) \propto \begin{cases} c\rho + \sum_{j=1}^{L} n_{ij}^{(t)} & \text{if } I_{t+1} = 0 \\ c(1-\rho) & \text{if } I_{t+1} = 1 \end{cases}$$

Construct the conditional distribution:

$$p(s_{t+1} = j \mid I_{t+1} = 0, s_t = i, S_t, \beta) \propto n_{ij}^{(t)} + c\rho\delta_i(j)$$

$$p(s_{t+1} = j \mid I_{t+1} = 1, s_t = i, S_t, \beta) \propto \pi_{0j}$$

This construction preserves the same conditional distribution of s_{t+1} given S_t and π_0. To sample $I \mid S$, use the Bernoulli distribution:

$$I_{t+1} \mid s_t = i, s_{t+1} = j, \pi_0 \sim \mathbf{Ber}\left(\frac{c(1-\rho)\pi_{0j}}{n_{ij}^{(t)} + c\rho\delta_i(j) + c(1-\rho)\pi_{0j}}\right).$$

A.2 Sample $(\Theta, P, \pi_0) \mid S, I, Y$

After sampling I and S, write $m_i = \sum_{s_t = i} I_t$. By construction, the conditional posterior of π_0 given S and I only depends on I and is a Dirichlet distribution by conjugacy:

$$\pi_0 \mid S, I \sim \mathbf{Dir}\left(\frac{\gamma}{L} + m_1, \ldots, \frac{\gamma}{L} + m_L\right)$$

This approach of sampling π_0 is simpler than Fox et al. (2009).

Conditional on π_0 and S, the sampling of π_i is straightforward by conjugacy:

$$\pi_i \mid \pi_0, S \sim \mathbf{Dir}(c(1-\rho)\pi_{01} + n_{i1}, \cdots, c(1-\rho)\pi_{0i} + c\rho + n_{ii}, \cdots, c(1-\rho)\pi_{0L} + n_{iL})$$

where n_{ij} is the number of $\{\tau \mid s_\tau = j, s_{\tau-1} = i\}$.

Sampling $\Theta \mid S, Y$ uses the results of regular linear models. The prior is:

$$(\phi_i, \sigma_i^{-2}) \sim \mathbf{N} - \mathbf{G}(\phi, H, \chi, \nu).$$

By conjugacy, the posterior is:

$$(\phi_i, \sigma_i^{-2}) \mid S, Y \sim \mathbf{N} - \mathbf{G}(\overline{\phi}_i, \overline{H}_i, \overline{\chi}_i, \overline{\nu}_i)$$

with:

$$\overline{\phi}_i = \overline{H}_i^{-1}(H\phi + X_i' Y_i)$$

$$\overline{H}_i = H + X_i' X_i$$

$$\overline{\chi}_i = \chi + Y_i' Y_i + \phi' H \phi - \overline{\phi}' \overline{H} \overline{\phi}$$

$$\overline{\nu}_i = \nu + n_i$$

where Y_i is the collection of y_t in state i. $x_t = (1, y_{t-1}, \cdots, y_{t-q})$ is the regressor in the AR(q) model. X_i and n_i are the collection of x_t and the number of observations in state i, respectively.

A.3 Sample $(\phi, H, \chi) \mid S, \Theta, \nu$

The conditional posterior is:

$$\phi, H \mid \{\phi_i, \sigma_i\}_{i=1}^K \sim \mathbf{N} - \mathbf{W}(m_1, \tau_1, A_1, a_1)$$

where K is the number of active states. ϕ_i and σ_i are the parameters associated with these states:

$$m_1 = \frac{1}{\tau_0^{-1} + \sum_{i=1}^K \sigma_i^{-2}} \left(\tau_0^{-1} m_0 + \sum_{i=1}^K \sigma_i^{-2} \phi_i \right)$$

$$\tau_1 = \frac{1}{\tau_0^{-1} + \sum_{i=1}^K \sigma_i^{-2}}$$

$$A_1 = \left(A_0^{-1} + \sum_{i=1}^{K} \sigma_i^{-2} \phi_i \phi_i' + \tau_0^{-1} m_0 m_0' - \tau_1^{-1} m_1 m_1' \right)^{-1}$$
$$a_1 = a_0 + K.$$

The conditional posterior of χ is:

$$\chi \mid \nu, \{\sigma_i\}_{i=1}^K \sim \mathbf{G}(d_1/2, c_1/2)$$

with $d_1 = d_0 + \sum_{i=1}^{K} \sigma_i^{-2}$ and $c_1 = c_0 + K\nu$.

A.4 Sample $\nu \mid \chi, S, \Theta$

The conditional posterior of ν has no regular density form:

$$p(\nu \mid \chi, \{\sigma_i\}_{i=1}^K) \propto \left(\frac{(\chi/2)^{\nu/2}}{\Gamma(\nu/2)} \right)^K \left(\prod_{i=1}^{K} \sigma_i^{-2} \right)^{\nu/2} \exp\{-\frac{\nu}{\rho_0}\}.$$

The Metroplolis-Hastings method is applied to sample ν. Draw a new ν from a proposal distribution:

$$\nu \mid \nu' \sim \mathbf{G}(\frac{\zeta_\nu}{\nu'}, \zeta_\nu)$$

with acceptance probability $\min \left\{ 1, \frac{p(\nu|\chi,\{\sigma_i\}_{i=1}^K) f_G(\nu'; \frac{\zeta_\nu}{\nu}, \zeta_\nu)}{p(\nu'|\chi,\{\sigma_i\}_{i=1}^K) f_G(\nu; \frac{\zeta_\nu}{\nu'}, \zeta_\nu)} \right\}$, where ν' is the value from the previous sweep. ζ_ν is fine tuned to produce a reasonable acceptance rate around 0.5, as suggested by Roberts et al. (1997) and Müller (1991).

References

Ang, A. and Bekaert, G. Regime switches in interest rates. *Journal of Business & Economic Statistics*, 20(2):163–182, 2002.

Ang, A., Bekaert, G., and Wei, M. Do macro variables, asset markets, or surveys forecast inflation better? *Journal of Monetary Economics*, 54(4):1163–1212, 2007.

Celeux, G., Hurn, M., and Robert, C.P. Computational and Inferential Difficulties with Mixture Posterior Distributions. *Journal of the American Statistical Association*, 95 (451), 2000.

Chib, S. Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics*, 75(1):79–97, 1996.

Chib, S. Estimation and comparison of multiple change-point models. *Journal of Econometrics*, 86(2):221–241, 1998.

Duffy, J. and Engle-Warnick, J. Multiple regimes in US monetary policy? A nonparametric approach. *Journal of Money Credit and Banking*, 38(5):1363, 2006.

Escobar, MD and West, M. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90, 1995.

Evans, M. and Wachtel, P. Inflation regimes and the sources of inflation uncertainty. *Journal of Money, Credit and Banking*, pages 475–511, 1993.

Fama, E.F. Short-term interest rates as predictors of inflation. *The American Economic Review*, 65(3):269–282, 1975.

Ferguson. A bayesian analysis of some nonparametric problem. *The Annals of Statistics*, 1(2):209–230, 1973.

Fox, E.B., Sudderth, E.B., Jordan, M.I., and Willsky, A.S. An HDP-HMM for systems with state persistence. In *Proceedings of the 25th international conference on Machine learning*, pages 312–319. ACM, 2008.

Fox, E.B., Sudderth, E.B., Jordan, M.I., and Willsky, A.S. The Sticky HDP-HMM: Bayesian Nonparametric Hidden Markov Models with Persistent States. *Arxiv preprint arXiv:0905.2592*, 2009.

Frühwirth-Schnatter, S. Markov Chain Monte Carlo Estimation of Classical and Dynamic Switching and Mixture Models. *Journal of the American Statistical Association*, 96(453), 2001.

Garcia, R. and Perron, P. An analysis of the real interest rate under regime shifts. *The Review of Economics and Statistics*, 78(1):111–125, 1996.

Geweke, J. Interpretation and inference in mixture models: Simple MCMC works. *Computational Statistics & Data Analysis*, 51(7):3529–3550, 2007.

Geweke, J. *Complete and Incomplete Econometric Models*. Princeton Univ Pr, 2009.

Geweke, J. and Amisano, G. Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting*, 2010.

Groen, J.J.J., Paap, R., and Ravazzolo, F. Real-time inflation forecasting in a changing world. *http://hdl.handle.net/1765/16709*, 2009.

Hamilton, J.D. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, 57(2): 357–384, 1989.

Huizinga, J. and Mishkin, F.S. Monetary policy regime shifts and the unusual behavior of real interest rates, 1986.

Ishwaran, H. and James, L.F. Gibbs Sampling Methods for Stick-Breaking Priors. *Journal of the American Statistical Association*, 96(453), 2001.

Ishwaran, H. and Zarepour, M. Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika*, 87(2):371, 2000.

Ishwaran, H. and Zarepour, M. Dirichlet prior sieves in finite normal mixtures. *Statistica Sinica*, 12(3):941–963, 2002.

Jochmann, M. Modeling U S Inflation Dynamics: A Bayesian Nonparametric Approach. *Working Paper Series*, 2010.

Kass, R.E. and Raftery, A.E. Bayes factors. *Journal of the American Statistical Association*, 90(430), 1995.

Koop, G. and Potter, S.M. Estimation and forecasting in models with multiple breaks. *Review of Economic Studies*, 74(3):763, 2007.

Levin, A.T. and Piger, J.M. Is inflation persistence intrinsic in industrial economies? 2004.

Maheu, J.M. and Gordon, S. Learning, forecasting and structural breaks. *Journal of Applied Econometrics*, 23(5):553–584, 2008.

Maheu, J.M., McCurdy, T.H., and Song, Y. Components of bull and bear markets: bull corrections and bear rallies. *Working Papers*, 2010.

Müller, P. A generic approach to posterior integration and Gibbs sampling. *Rapport technique*, pages 91–09, 1991.

Pesaran, M.H., Pettenuzzo, D., and Timmermann, A. Forecasting time series subject to multiple structural breaks. *Review of Economic Studies*, 73(4):1057–1084, 2006.

Roberts, GO, Gelman, A., and Gilks, WR. Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, 7(1): 110–120, 1997.

Rose, A.K. Is the real interest rate stable? *Journal of Finance*, 43(5):1095–1112, 1988.

Sethuraman, J. A constructive definition of dirichlet priors. *Statistica Sinica*, 4:639–650, 1994.

Shahbaba, B. and Neal, R.M. Nonlinear models using dirichlet process mixtures. *Journal of Machine Learning Research*, 10:1829–1850, 2009.

Stock, J.H. and Watson, M.W. A probability model of the coincident economic indicators. *Leading Economic indicators: new approaches and forecasting records*, 66, 1991.

Teh, Y.W., Jordan, M.I., Beal, M.J., and Blei, D.M. Hierarchical dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581, 2006.

Walsh, C.E. Three questions concerning nominal and real interest rates. *Economic Review*, (Fall):5–19, 1987.

Wang, J. and Zivot, E. A Bayesian time series model of multiple structural changes in level, trend, and variance. *Journal of Business & Economic Statistics*, 18(3): 374–386, 2000.

West, M., Müller, P., and Escobar, M.D. Hierarchical priors and mixture models, with application in regression and density estimation. *Aspects of uncertainty: A Tribute to DV Lindley*, pages 363–386, 1994.

Table 1: Log predictive likelihoods in simulation study

DGP	Estimated Model		
	SDHDP-HMM	MS	SB
SDHDP-HMM	-170.55	-183.30	-264.65
MS	-208.32	-208.10	-212.07
SB	-179.51	-187.26	-178.41

The SDHDP-HMM is (7)-(12); the MS is the 3-state markov switching model of (27)-(30); and the SB is the 4-state structural break model of (31)-(35). 1000 observations are simulated from each model and the last 100 are used to calculate the predictive likelihoods. The first column shows the names of the data generating processes. The first row shows the names of the estimated models.

Table 2: Log predictive likelihoods of U.S. real interest rates

AR(q)	q=2	q=3	q= 4			
	-457.62	-451.07	-455.97			
MS(K)[b]	K=3	K=4	K=5			
	-433.09	-426.62	-424.51			
SB(K)[c]	K=3	K=4	K=5	K=10	K=15	K=20
	-450.82	-451.62	-437.28	-433.50	-432.69	-434.24
SDHDP-HMM[e]		-423.50				

There are 252 observations from 1947Q1 to 2009Q4 for U.S. quarterly real interest rate. The last 200 observations are used to calculate the predictive likelihoods. MS(K) is the K-state Markov switching model of (27)-(30) and SB(K) is the K-state structural break model of (31)-(35). For the SDHDP-HMM, MS(K) and SB(K), each state has Gaussian AR(2) dynamics.

Table 3: Log predictive likelihoods of U.S. inflation

AR(q)	q=1	q=2	q= 3
	-185.06	-173.17	-173.42
MS [b]		-92.50	
SB(K)[c]	K=3	K=5	K=10
	-125.50	-98.69	-101.18
FSJW[d]		-82.45	
SDHDP-HMM[e]		-74.07	

There are 1153 observations from Feb 1914 to Jan 2010 for U.S. monthly inflation rate. The last 200 observations are used to calculate the predictive likelihoods. MS is the 2-state Markov switching model of Evans and Wachtel (1993); SB(K) is the K-state structural break model of (31)-(35); and the FSJW is Fox et al.'s (2008) model (or the SDHDP-HMM without the hierarchical structure of G_0 on the conditional data density parameters). For the SDHDP-HMM, FSJW, MS and SB(K), each state has Gaussian AR(1) dynamics.

Table 4: Posterior summary of the SDHDP-HMM parameters estimated from U.S. inflation

	mean	Std	95% DI
ϕ_0	0.03	0.20	(-0.376, 0.432)
ϕ_1	0.97	0.11	(0.742, 1.199)
H_{00}	0.77	0.42	(0.225, 1.788)
H_{01}	0.02	0.35	(-0.692, 0.734)
H_{11}	2.06	0.84	(0.768, 4.047)
χ	0.19	0.12	(0.034, 0.488)
ν	1.21	0.50	(0.496, 2.414)

There are 1153 observations from Feb 1914 to Jan 2010 for U.S. monthly inflation rate. Each state has Gaussian AR(1) dynamics:
$y_t = \phi_{s_t 0} + \phi_{s_t 1} y_{t-1} + \sigma_{s_t} \varepsilon_t$. The parameters ϕ_i and σ_i are drawn from the hierarchical distribution:
$\sigma_i^{-1} \sim \mathbf{G}(\chi/2, \nu/2)$ and
$\phi_i \mid \sigma_i \sim \mathbf{N}(\phi, \sigma_i H^{-1})$.

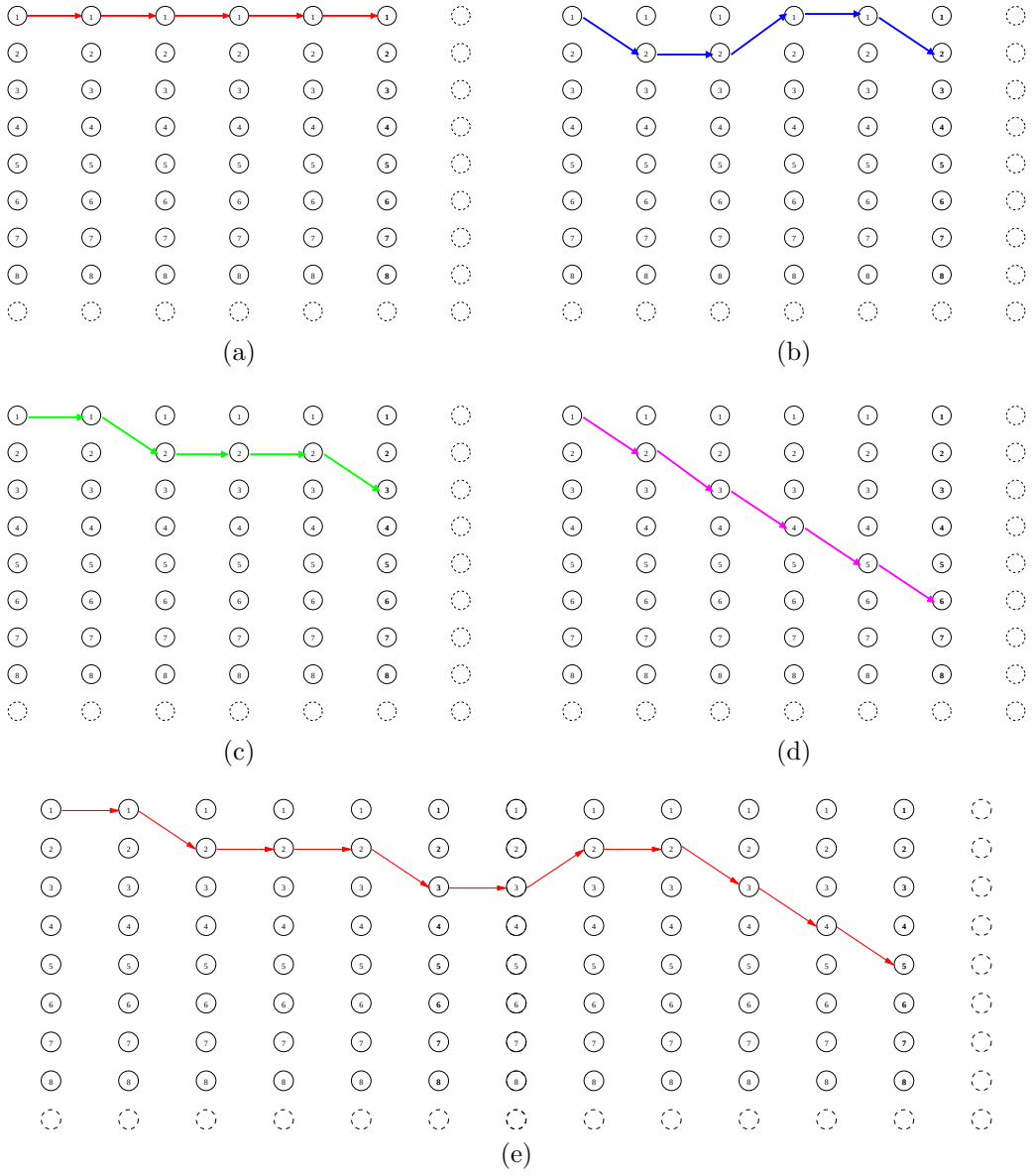

Figure 1: The horizontal dimension from left to right represents time and the vertical circles represent different states. The numbers in the circles are the labels of the states and they are exchangeable. The SDHDP-HMM nests: (a) no state change, (b) regime switching, (c) structural breaks, (d) frequent parameter change and (e) regime switching and structural breaks.

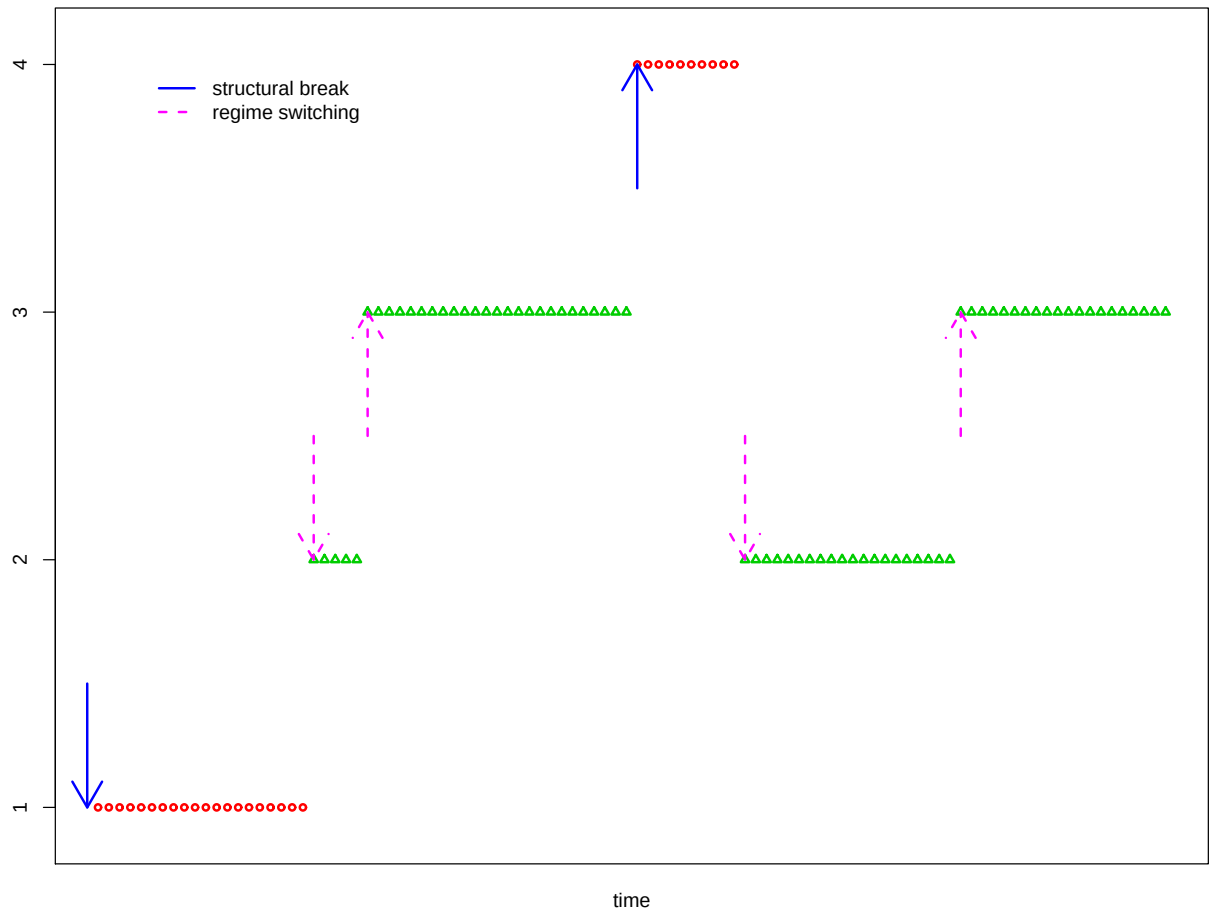

Figure 2: Example of the global identification of regime switching and structural breaks. All the points represent one sample of the states $(s_1, \cdots, s_T)$ from the posterior samples. The circles are non-recurrent states which only appear for one consecutive period and the triangles are recurrent states. The solid arrows point to the break points and the dashed arrows point to the switching points.

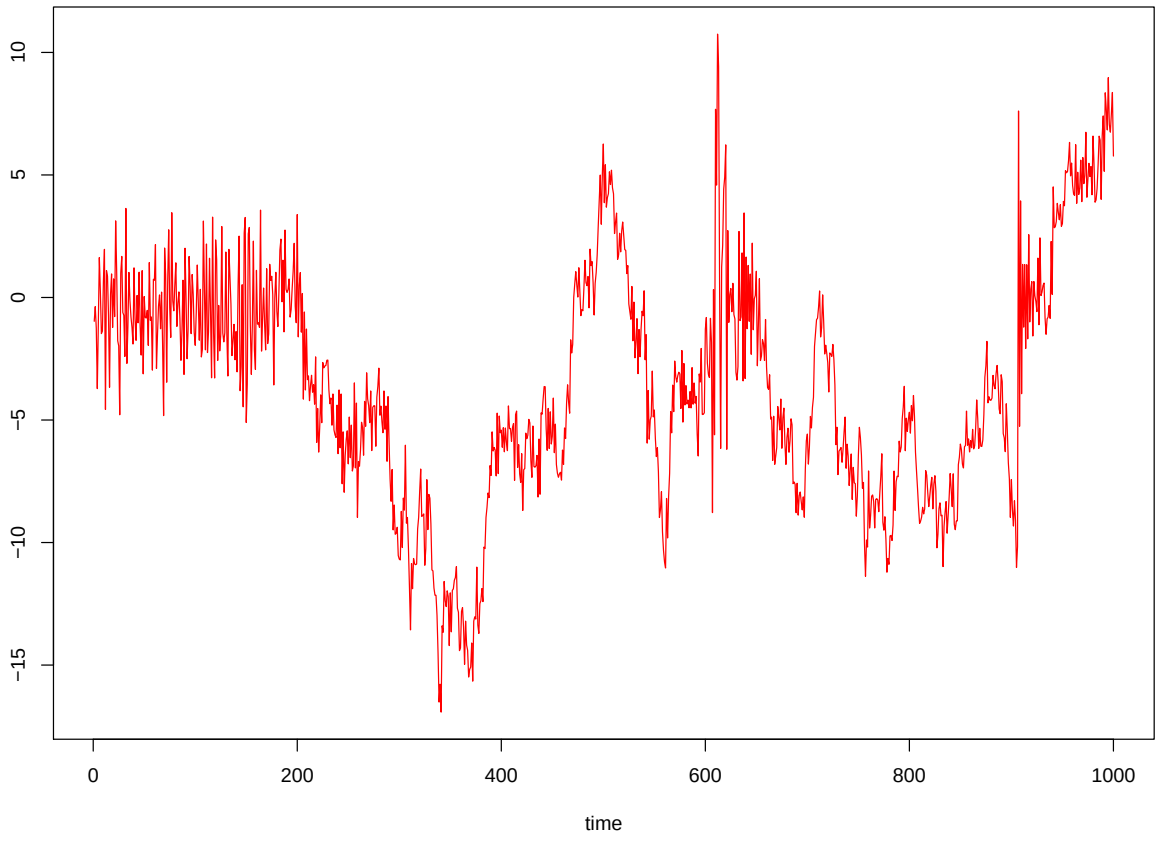

Figure 3: Data simulated by a SDHDP-HMM. Each state has Gaussian AR(2) dynamics: $y_t = \phi_{s_t 0} + \phi_{s_t 1} y_{t-1} + \phi_{s_t 2} y_{t-2} + \sigma_{s_t} \varepsilon_t$.

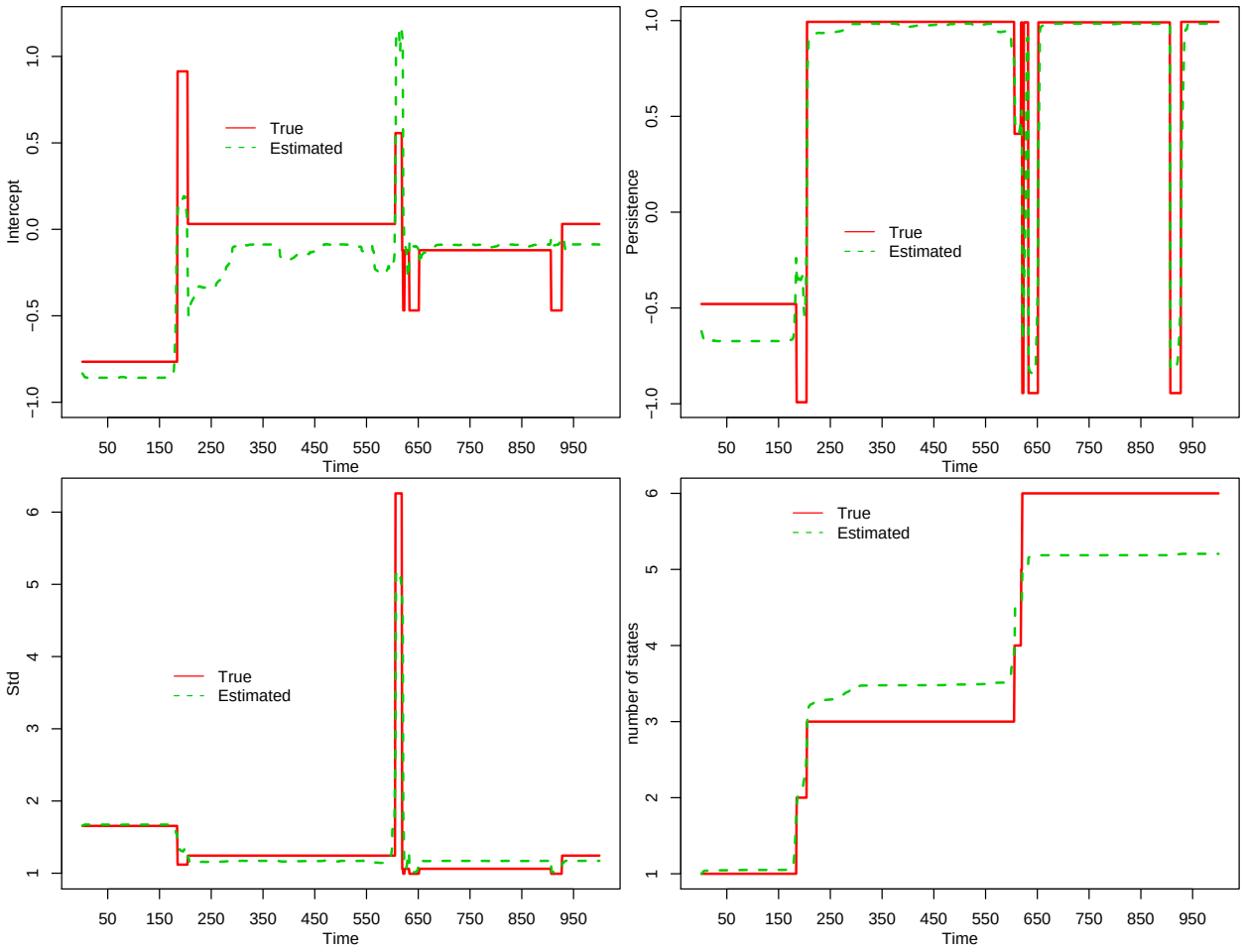

Figure 4: The SDHDP-HMM estimates the data from figure 3. Each state has Gaussian AR(2) dynamics: $y_t = \phi_{s_t0} + \phi_{s_t1} y_{t-1} + \phi_{s_t2} y_{t-2} + \sigma_{s_t} \varepsilon_t$. The solid lines are the true values and the dashed lines are the posterior means of those values estimated by the SDHDP-HMM. The top-left panel plots the intercepts ϕ_{s_t0}; the top-right panel plots the persistence parameters $\phi_{s_t1} + \phi_{s_t2}$; the bottom-left plots the conditional standard deviations σ_{s_t} and the bottom-right plots the cumulative number of the active states (active state means it has been visited at least once).

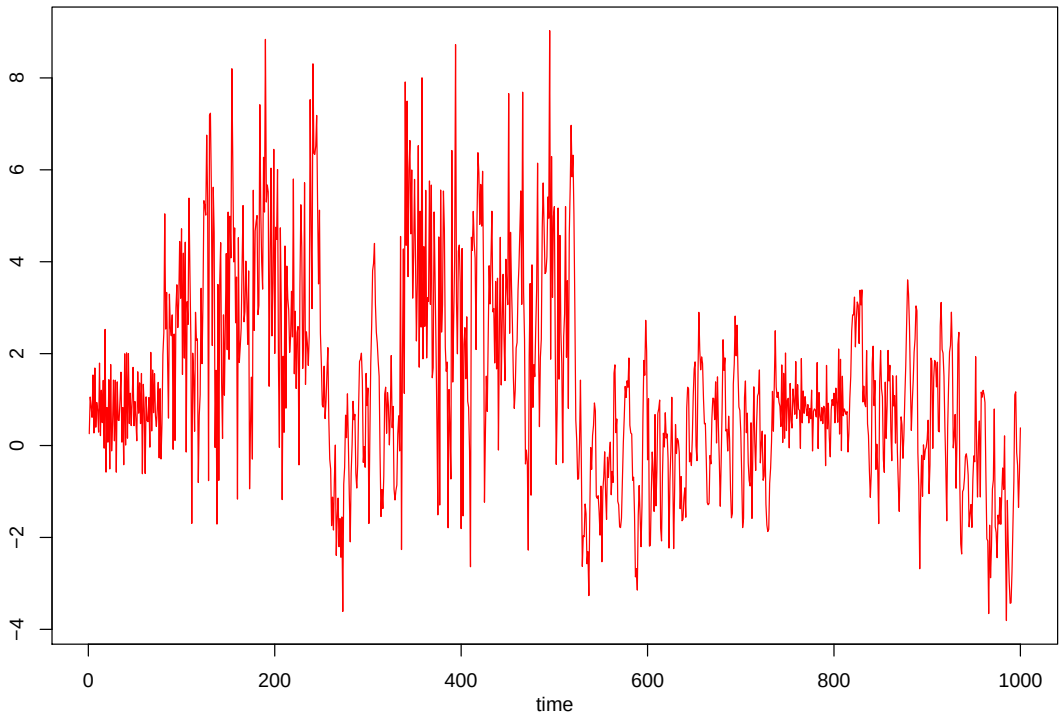

Figure 5: Data simulated by a 3-state Markov switching model of (27)-(30). Each state has Gaussian AR(2) dynamics: $y_t = \phi_{s_t 0} + \phi_{s_t 1} y_{t-1} + \phi_{s_t 2} y_{t-2} + \sigma_{s_t} \varepsilon_t$.

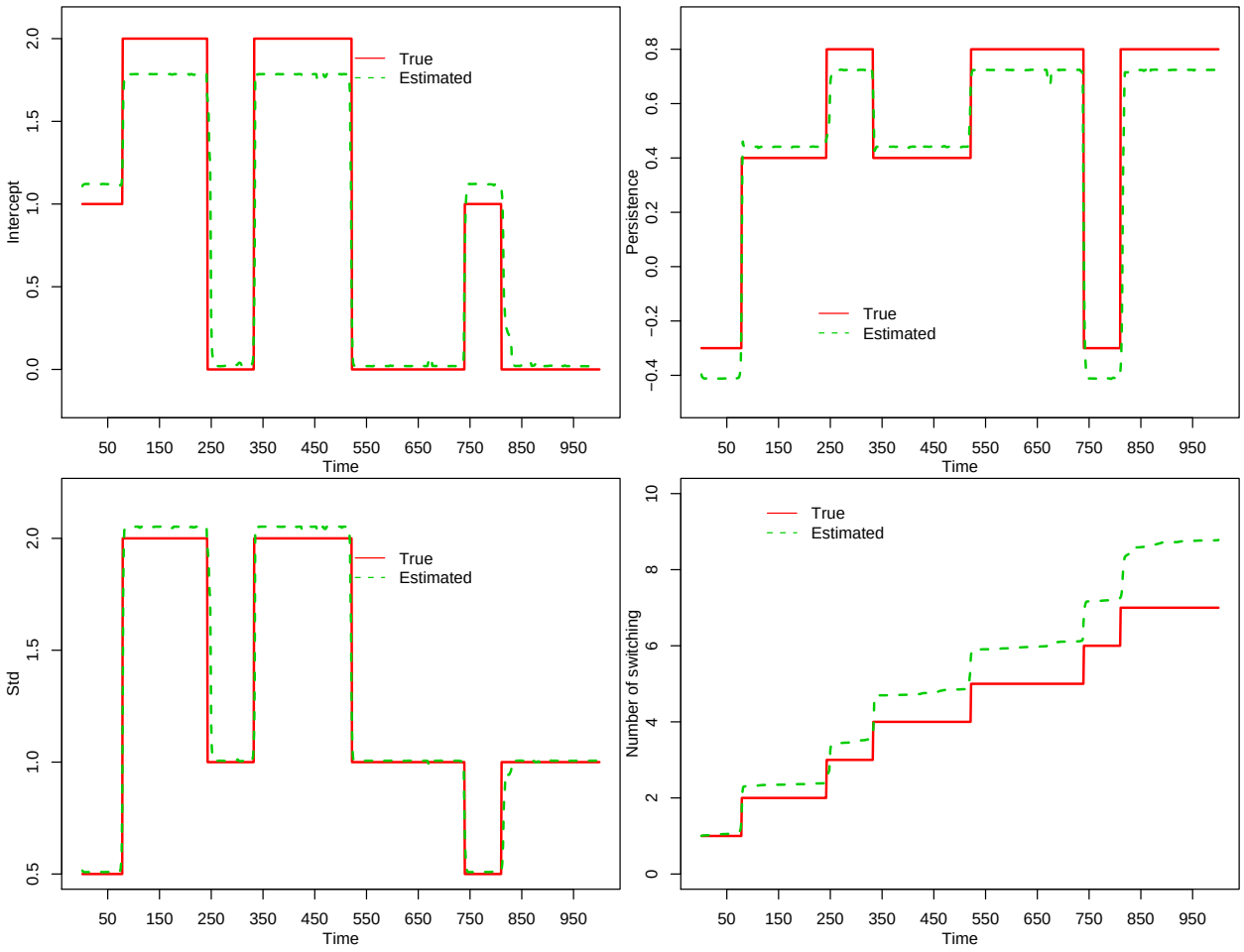

Figure 6: The SDHDP-HMM estimates the data in figure 5. Each state has Gaussian AR(2) dynamics: $y_t = \phi_{s_t0} + \phi_{s_t1} y_{t-1} + \phi_{s_t2} y_{t-2} + \sigma_{s_t} \varepsilon_t$. The solid lines are the true values and the dashed lines are the posterior means of those values estimated by the SDHDP-HMM. The top-left panel plots the intercepts ϕ_{s_t0}; the top-right panel plots the persistence parameters $\phi_{s_t1} + \phi_{s_t2}$; the bottom-left plots the conditional standard deviations σ_{s_t} and the bottom-right plots the cumulative number of regime switching.

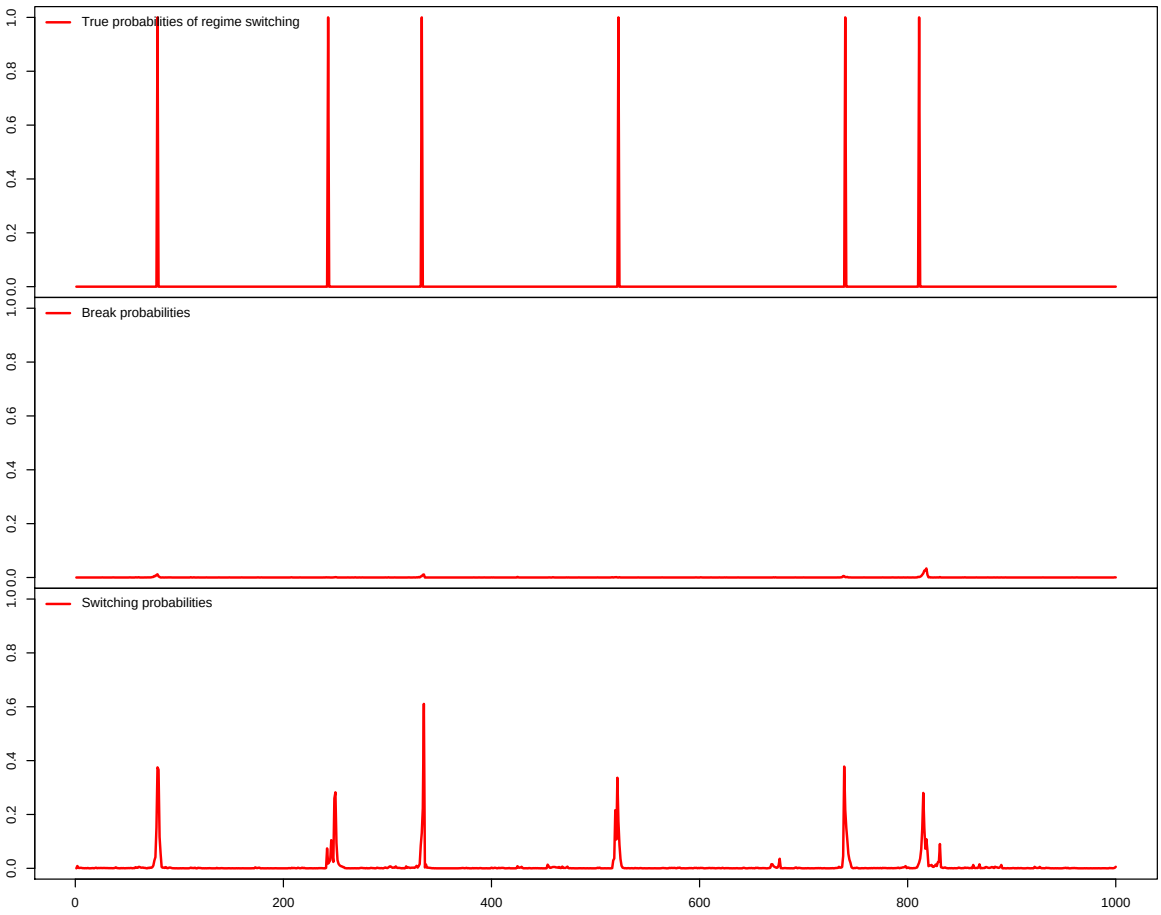

Figure 7: Globally identified smoothed probabilities of structural breaks and regime switching. The data is in figure 5, which is simulated by a 3-state Markov switching model and estimated by the SDHDP-HMM. The top panel is the switching points; the middle panel is the probabilites of structural breaks and the bottom panel is the probabilities of regime switching.

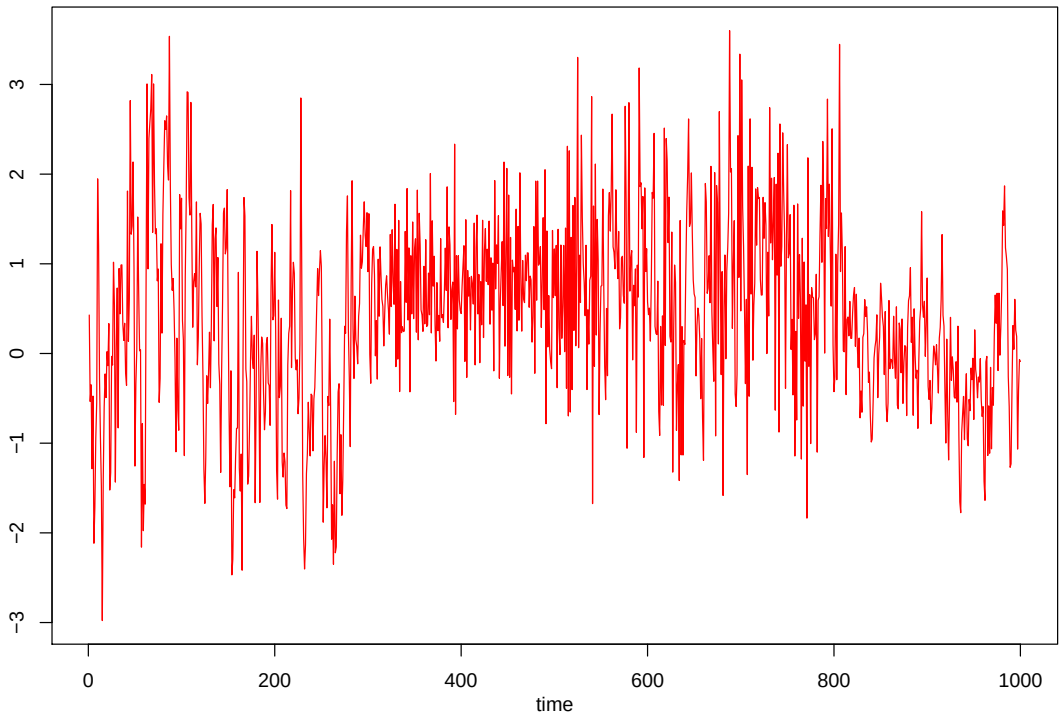

Figure 8: Data simulated by a structural break model of (31)-(35). Each state has Gaussian AR(2) dynamics: $y_t = \phi_{s_t 0} + \phi_{s_t 1} y_{t-1} + \phi_{s_t 2} y_{t-2} + \sigma_{s_t} \varepsilon_t$.

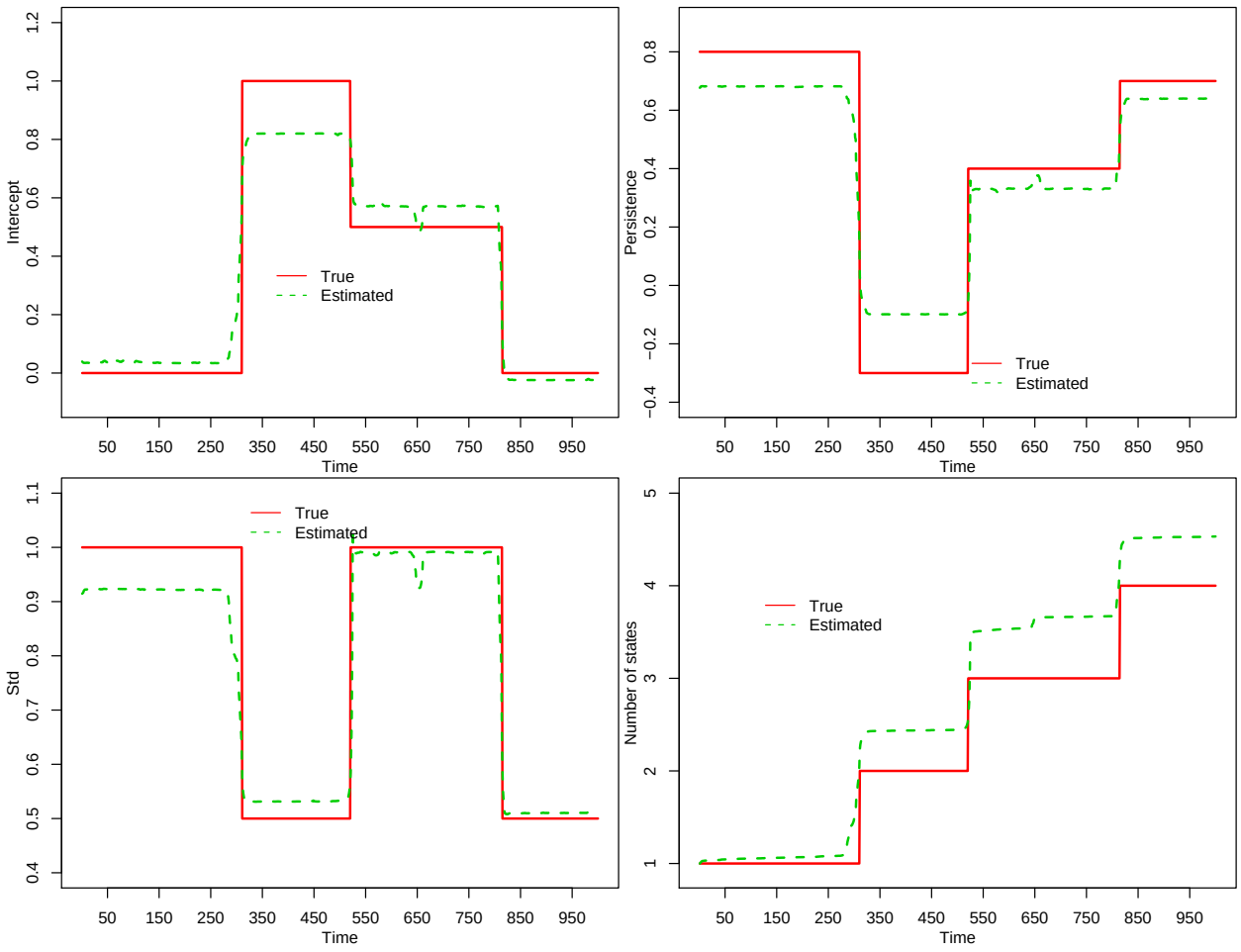

Figure 9: The SDHDP-HMM estimates the data in figure 8. Each state has Gaussian AR(2) dynamics: $y_t = \phi_{s_t0} + \phi_{s_t1} y_{t-1} + \phi_{s_t2} y_{t-2} + \sigma_{s_t} \varepsilon_t$. The solid lines are the true values and the dashed lines are the posterior means of those values estimated by the SDHDP-HMM. The top-left panel plots the intercepts ϕ_{s_t0}; the top-right panel plots the persistence parameters $\phi_{s_t1} + \phi_{s_t2}$; the bottom-left plots the conditional standard deviations σ_{s_t} and the bottom-right plots the cumulative number of states.

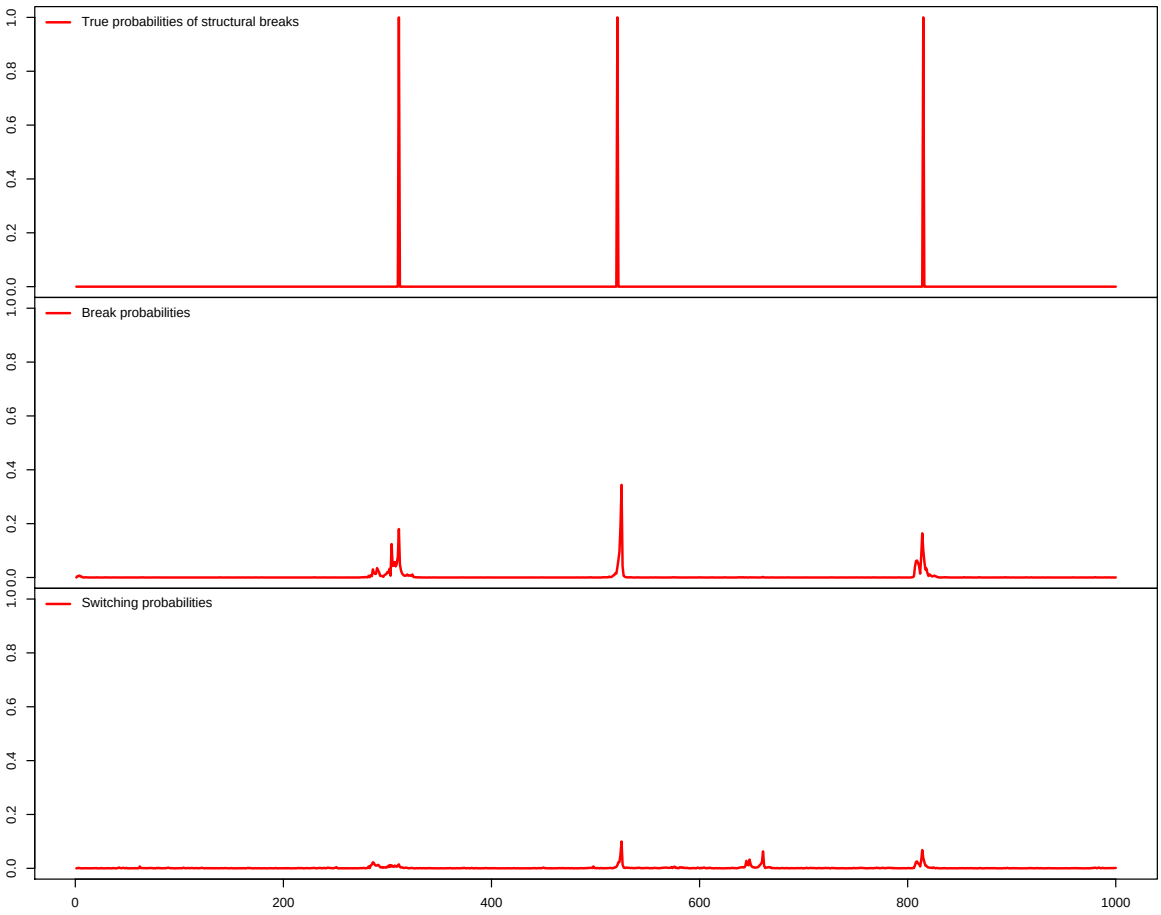

Figure 10: Globally identified smoothed probabilities of structural breaks and regime switching. The data is in figure 8, which is simulated by a 4-state structural break model and estimated by the SDHDP-HMM. The top panel is the switching points; the middle panel is the probabilites of structural breaks and the bottom panel is the probabilities of regime switching.

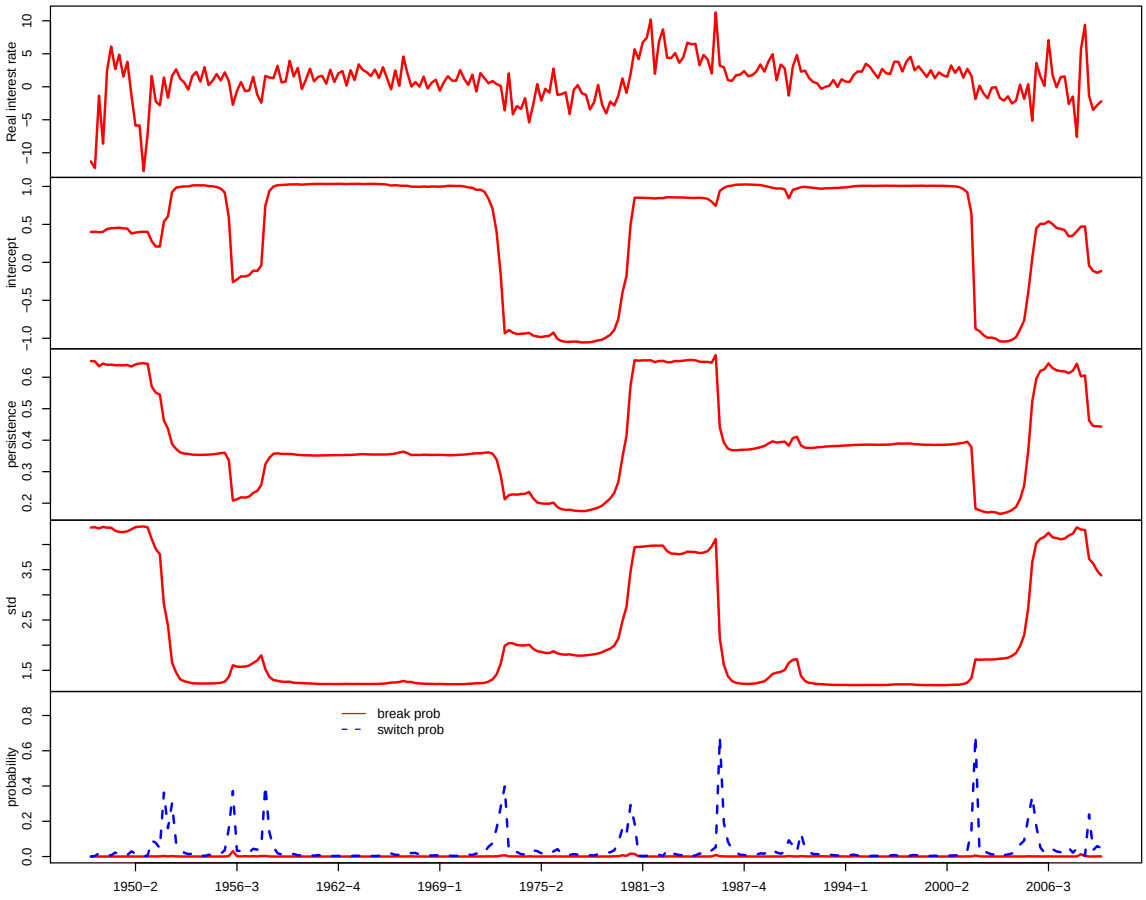

Figure 11: There are 252 observations from 1947Q1 to 2009Q4 for U.S. quarterly real interest rate. The data are estimated by the SDHDP-HMM and each state has Gaussian AR(2) dynamics: $y_t = \phi_{s_t0} + \phi_{s_t1} y_{t-1} + \phi_{s_t2} y_{t-2} + \sigma_{s_t} \varepsilon_t$. The first panel plots the data and the rest plots the posterior mean of different parameters: the second panel plots the intercepts ϕ_{s_t0}, the third panel plots the persistence parameters $\phi_{s_t1} + \phi_{s_t2}$, the fourth panel plots the conditional standard deviations σ_{s_t} and the last panel plots the probabilites of regime switching and structural breaks.

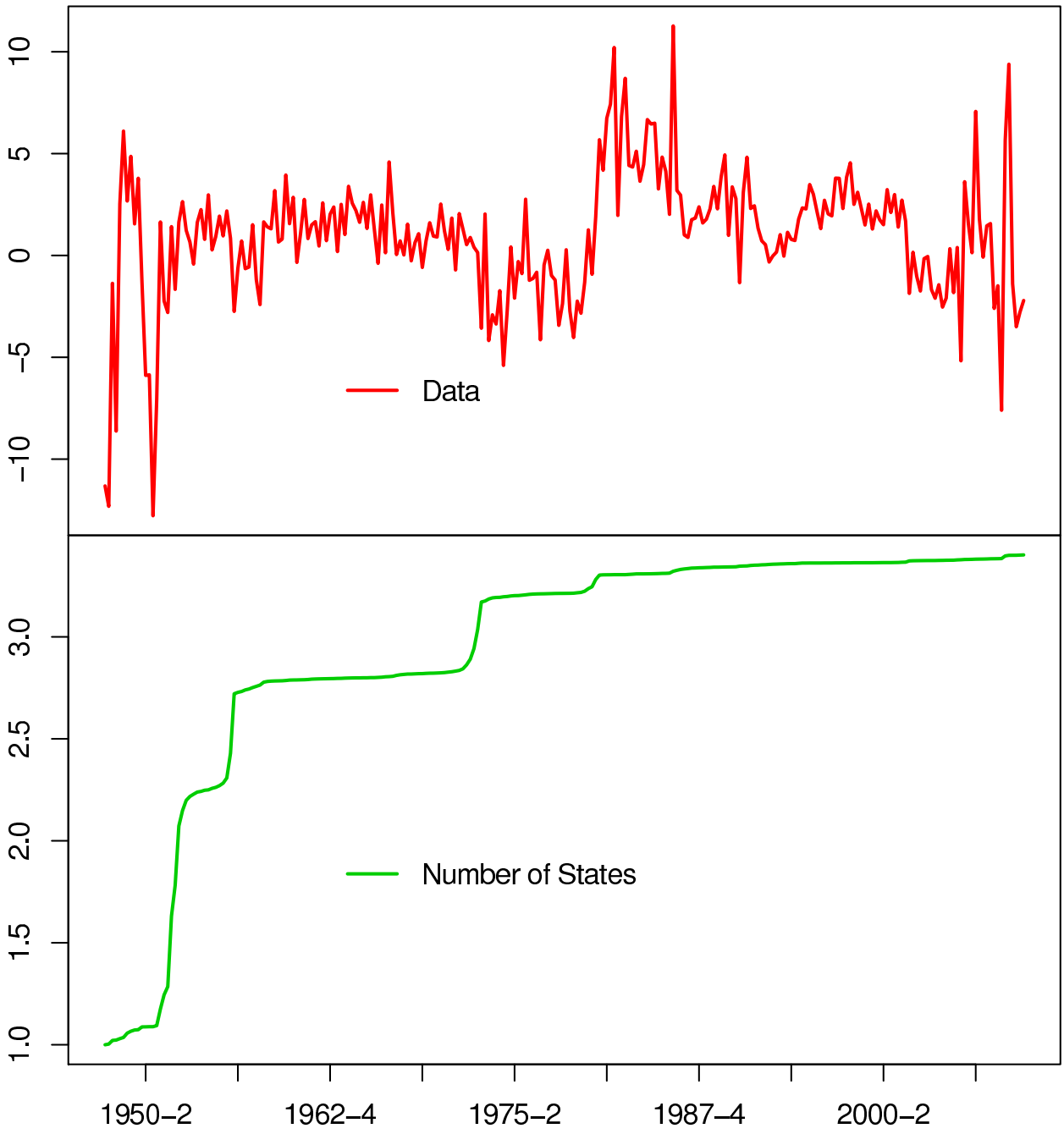

Figure 12: There are 252 observations from 1947Q1 to 2009Q4 for U.S. quarterly real interest rate. The data are estimated by the SDHDP-HMM and each state has Gaussian AR(2) dynamics: $y_t = \phi_{s_t0} + \phi_{s_t1} y_{t-1} + \phi_{s_t2} y_{t-2} + \sigma_{s_t} \varepsilon_t$. The top panel plots the data and the bottom panel plots the posterior mean of the cumulative number of active states (active state means it has been visited at least once).

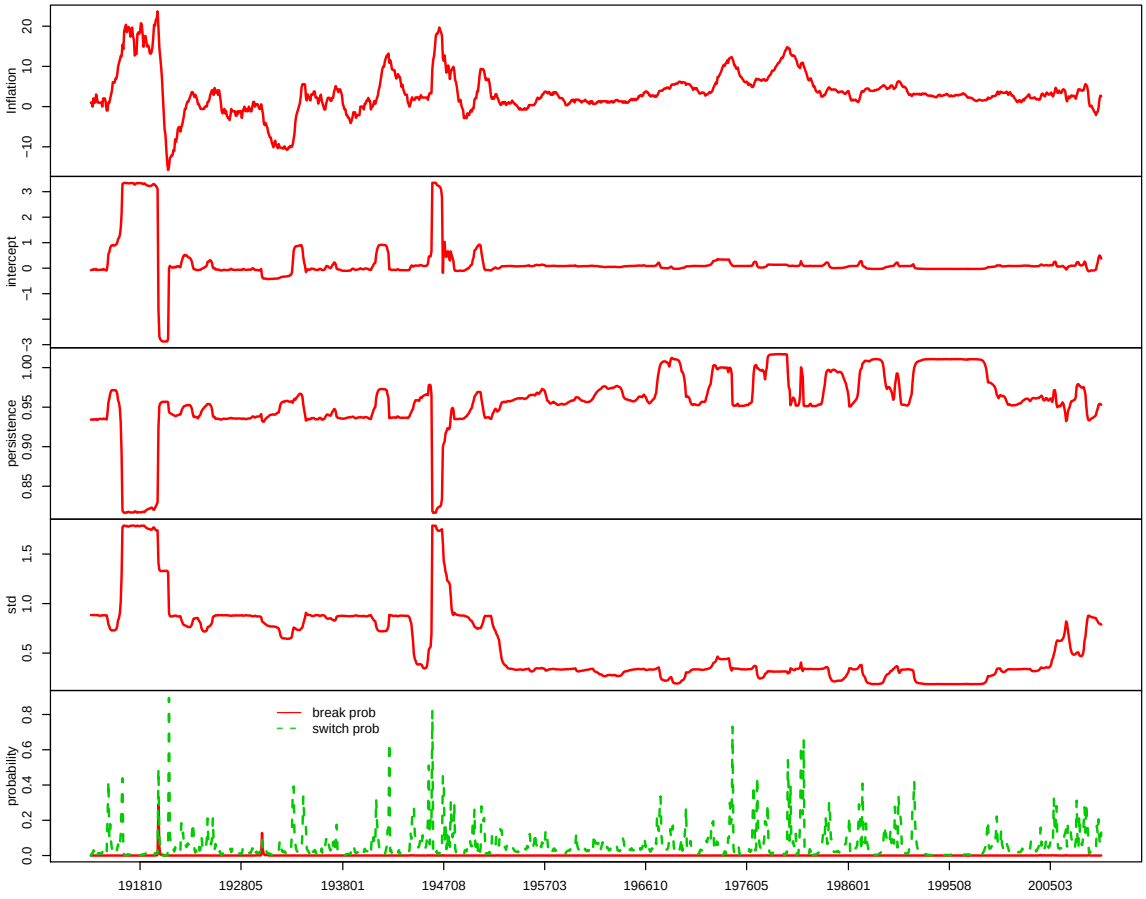

Figure 13: There are 1153 observations from Feb 1914 to Jan 2010 for U.S. monthly inflation rate. The data are estimated by the SDHDP-HMM and each state has Gaussian AR(1) dynamics: $y_t = \phi_{s_t 0} + \phi_{s_t 1} y_{t-1} + \sigma_{s_t} \varepsilon_t$. The first panel plots the data and the rest plots the posterior mean of different parameters: the second panel plots the intercepts $\phi_{s_t 0}$, the third panel plots the persistence parameters $\phi_{s_t 1}$, the fourth panel plots the conditional standard deviations σ_{s_t} and the last panel plots the probabilites of regime switching and structural breaks.

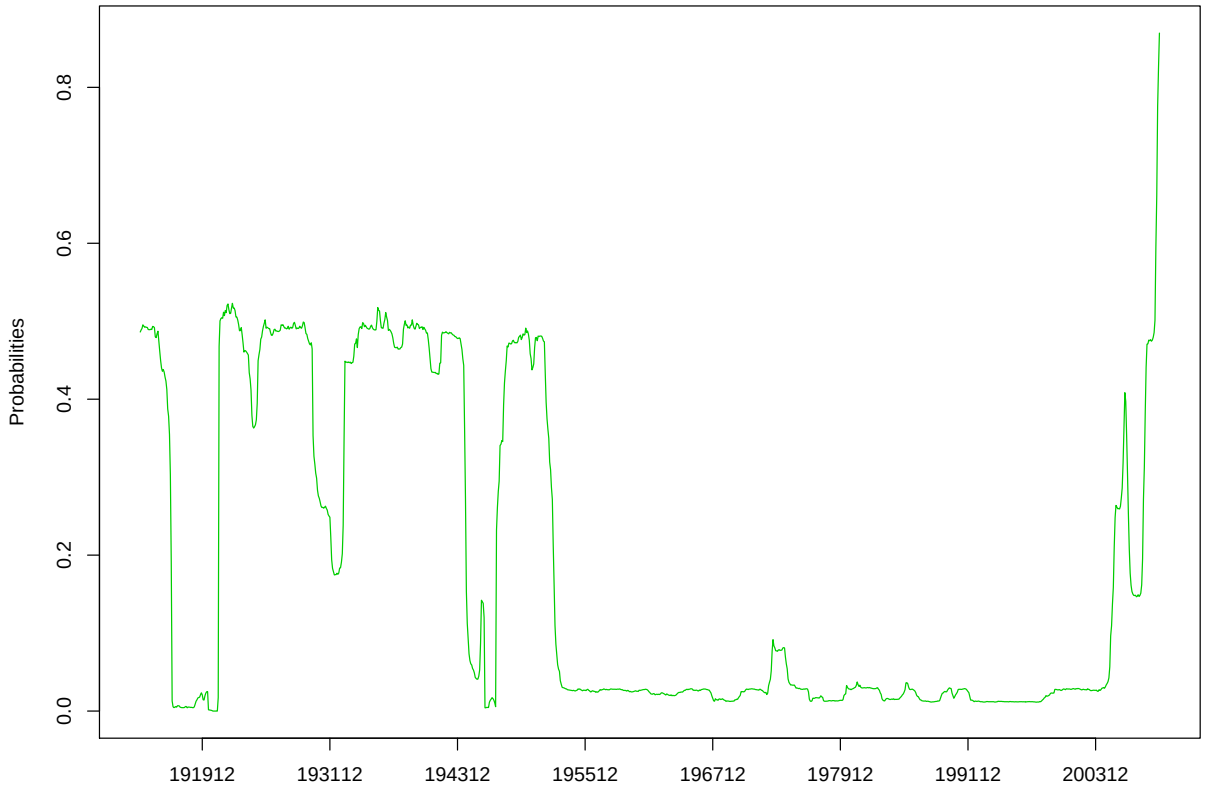

Figure 14: There are 1153 observations from Feb 1914 to Jan 2010 for U.S. monthly inflation rate. The data are estimated by the SDHDP-HMM and each state has Gaussian AR(1) dynamics: $y_t = \phi_{s_t 0} + \phi_{s_t 1} y_{t-1} + \sigma_{s_t} \varepsilon_t$. This figure plots the smoothed probabilities of past states of U.S. inflation to be the same as Jan 2010, or $p(z_\tau = z_{201001} \mid Y)$.

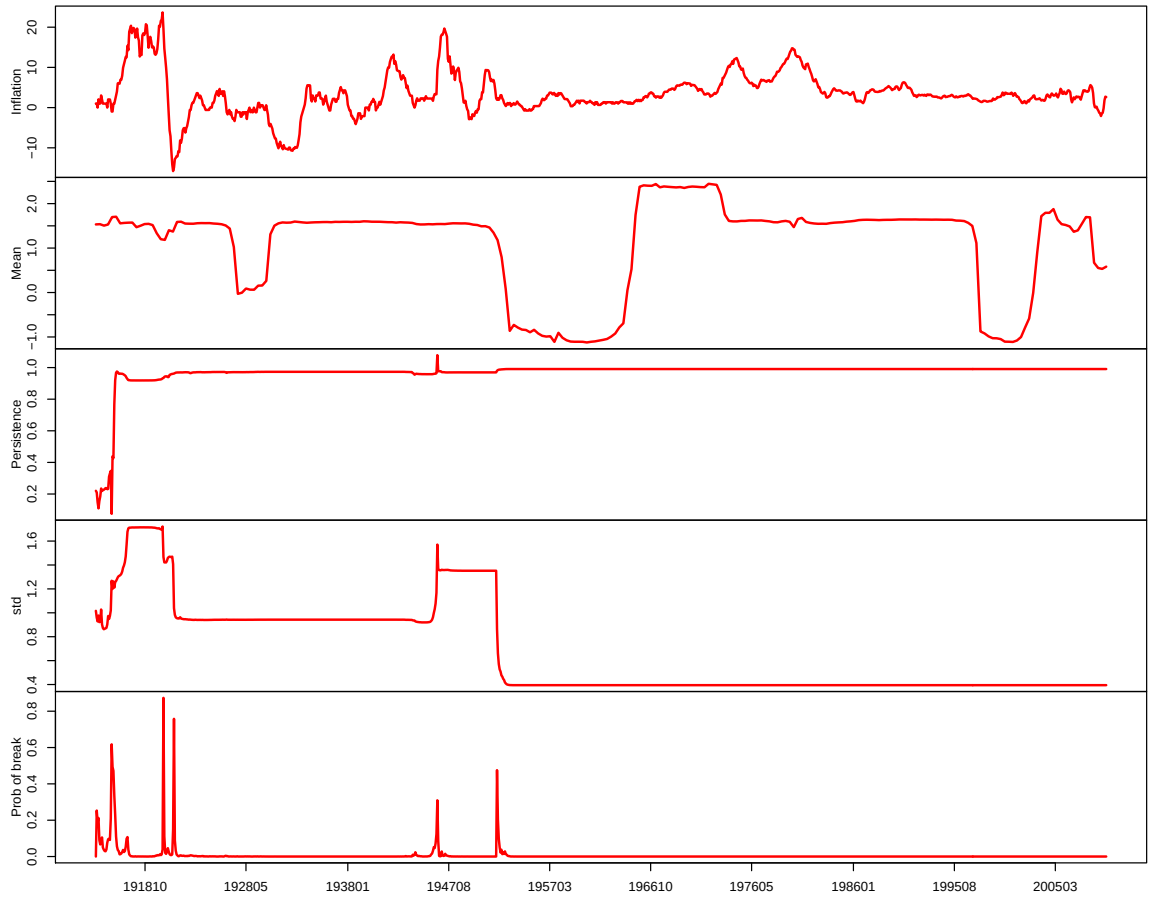

Figure 15: There are 1153 observations from Feb 1914 to Jan 2010 for U.S. monthly inflation rate . The data are estimated by the structural break model of Chib(1998) and each state has Gaussian AR(1) dynamics: $y_t = \phi_{s_t 0} + \phi_{s_t 1} y_{t-1} + \sigma_{s_t} \varepsilon_t$. The first panel plots the data and the rest plots the posterior mean of different parameters: the second panel plots the intercepts $\phi_{s_t 0}$, the third panel plots the persistence parameters $\phi_{s_t 1}$, the fourth panel plots the conditional standard deviations σ_{s_t} and the last panel plots the probabilites of structural breaks.

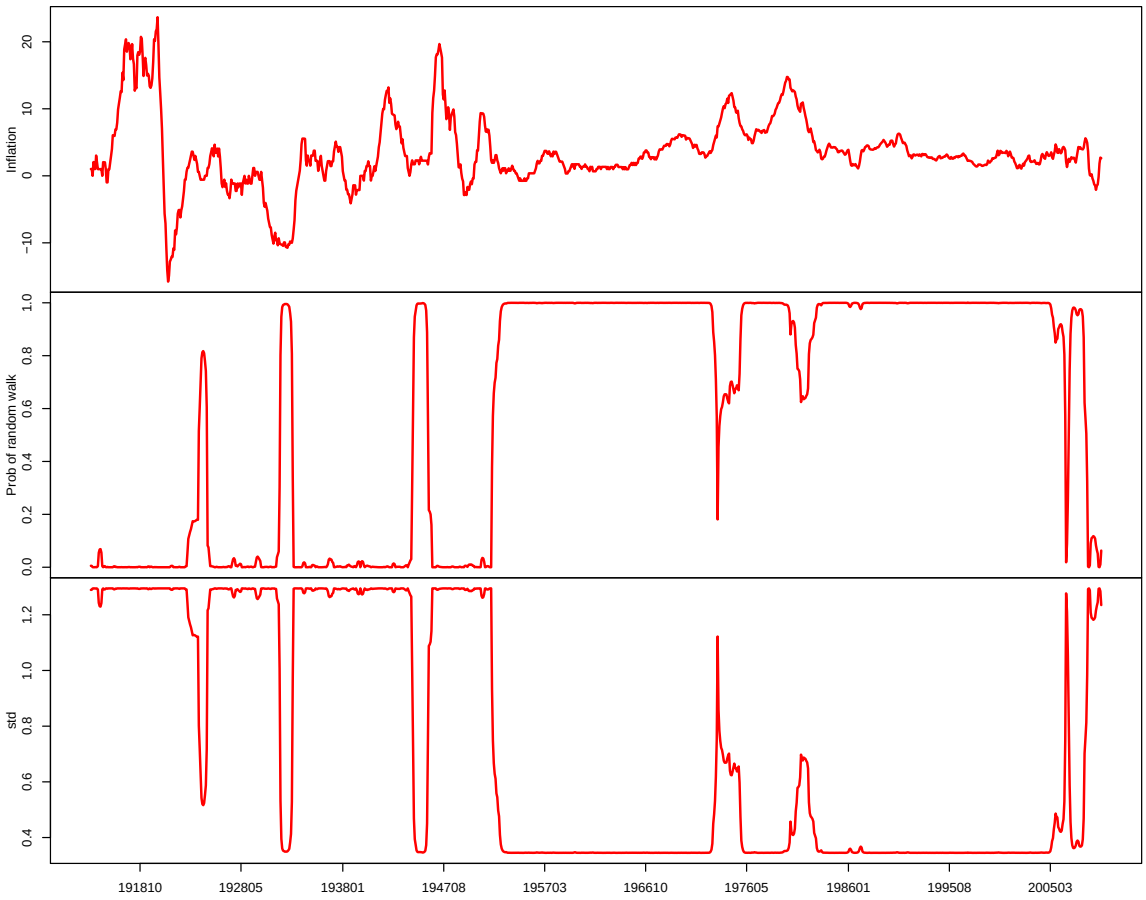

Figure 16: There are 1153 observations from Feb 1914 to Jan 2010 for U.S. monthly inflation rate. The data are estimated by the 2-state Markov switching model of Evans and Wachtel (1993). The first panel plots the data and the rest plots the posterior mean of different parameters: the second panel plots the probabilities of in the random walk state and the last panel plots the conditional standard deviations.

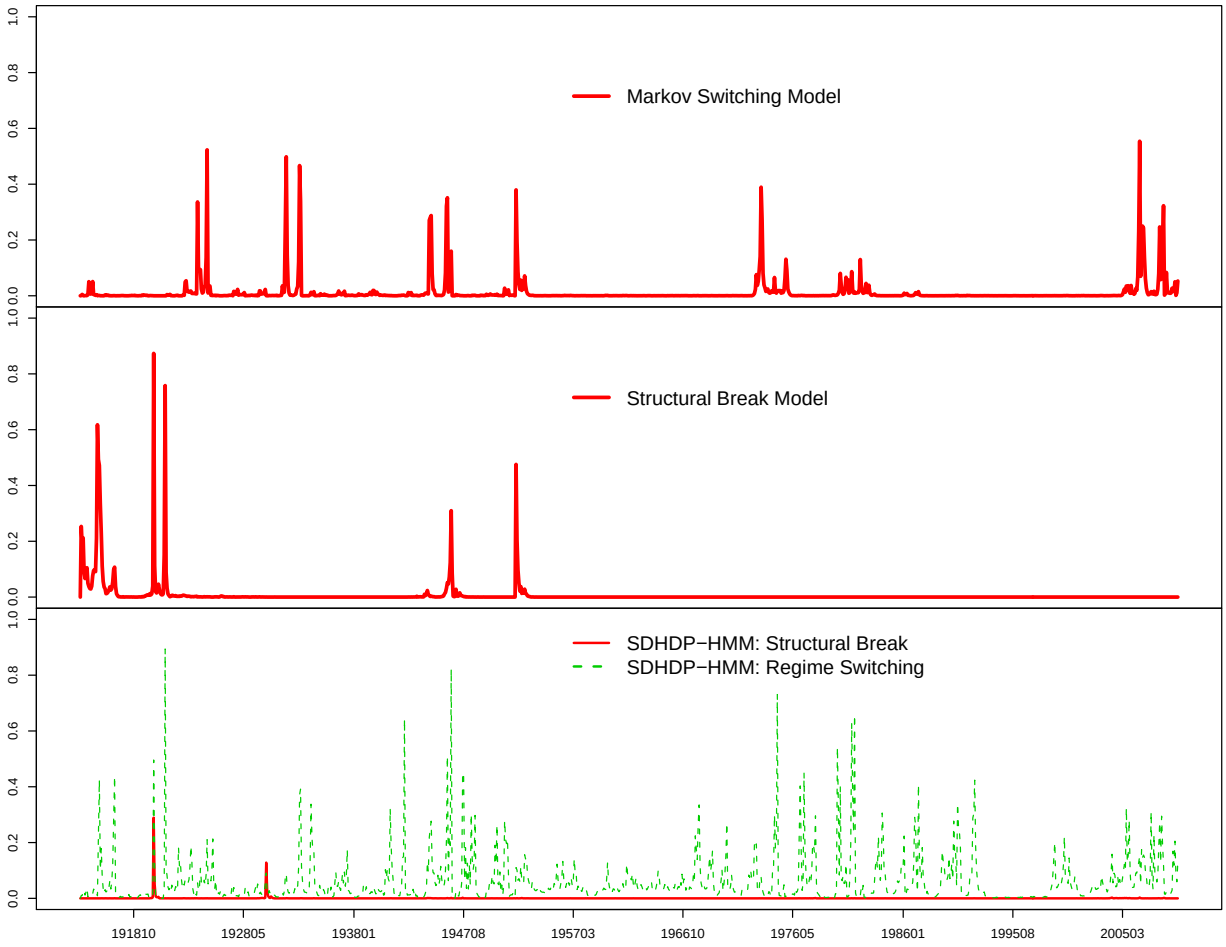

Figure 17: There are 1153 observations from Feb 1914 to Jan 2010 for U.S. monthly inflation rate. The first panel plots the posterior probabilities of regime switching by the 2-state Markov switching model of Evans and Wachtel (1993); the second panel plots the posterior probabilities of structural breaks by the structural break model of Chib (1998) and the last panel plots the posterior probabilities of regime switching and structural breaks by the SDHDP-HMM.