JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **29**: 825–842 (2014)
Published online 26 June 2013 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/jae.2337

MODELLING REGIME SWITCHING AND STRUCTURAL BREAKS WITH AN INFINITE HIDDEN MARKOV MODEL

YONG SONG*
University of Technology, Sydney, Australia; and Rimini Centre for Economic Analysis, Italy

SUMMARY

This paper proposes an infinite hidden Markov model to integrate the regime switching and structural break dynamics in a unified 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 flexible approach allows for regime persistence and estimates the number of states automatically. An application to US real interest rates compares the new model to existing parametric alternatives. Copyright © 2013 John Wiley & Sons, Ltd.

Received 5 August 2011; Revised 26 March 2013;

Supporting information may be found in the online version of this article.

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 a loss in the estimation precision because the past states cannot recur. An infinite hidden Markov model (iHMM) is proposed to accommodate both types of models and provide much richer dynamics. The estimation and forecasting are performed in a Bayesian framework. The model is applied to US real interest rates and I find that both the regime switching and the structural break dynamics are important for the data.

Regime switching models were first applied by Hamilton (1989). It is an important methodology to model nonlinear dynamics and widely applied to economic data including business cycles (Hamilton, 1989), bull and bear markets (Maheu *et al*., 2012), 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 and small (it is usually two and at most four). In the rest of the paper, a regime switching model is assumed to have both features. In practice, 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 parameters, and could incur some loss in estimation precision. In the current literature, structural break models such as Chib (1998); Wang and Zivot (2000); Pesaran *et al*. (2006) and Maheu and

* Correspondence to: Yong Song, Economics Discipline Group, University of Technology, Level 2, Block D, 1–59 Quay Street, Haymarket, Sydney, Australia. E-mail: Yong.Song@uts.edu.au

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. Garcia and Perron (1996) used a three-regime Markov switching model for US real interest rates, while Wang and Zivot (2000) applied a model with structural breaks in the mean and the volatility. Did the real interest rates in 1981 have distinct dynamics or revert to a historical state with the same dynamics? Existing econometric models have difficulty answering such questions.

This paper provides a solution by proposing an infinite hidden Markov 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 to capture the dynamic instability. Differently from the Bayesian model averaging methodology, this model combines different dynamics in the estimation.

The proposed model builds on and extends Fox *et al*. (2011). They used Dirichlet processes as the prior on the transition probabilities of an infinite hidden Markov model. The key innovation in their work is introducing a *sticky* parameter that favours state persistence and avoids the saturation of states. Jochmann (2010) applies their model to detect the number of regimes in US inflation. Their model is denoted by FSJW in the rest of the paper.[1]

The first contribution of this paper is introducing a second hierarchical structure in addition to FSJW to allow learning and sharing of information for the parameters of the conditional data density in each state.

The second contribution is to apply the model to US real interest rates, which is a widely analysed series in the existing literature. Specifically, the iHMM is compared to regime switching models, structural break models and linear AR models with and without rolling windows. The model comparison shows that the iHMM provides the best out-of-sample density forecast and both regime switching and structural break dynamics are important for US real interest rates.

The rest of the paper is organized as follows. Section 2 introduces the Dirichlet process and its stick-breaking representation to make this paper self-contained. Section 3 outlines the infinite hidden Markov model and discusses its structure and implications. Section 4 sketches the posterior sampling algorithm and the forecasting method. Section 5 studies the dynamics of US real interest rates and performs a prior sensitivity check. Section 6 concludes.

2. DIRICHLET PROCESS

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: a shape parameter G_0, which is a probability measure over a sample space Ω, and a scalar concentration parameter $\alpha_0 > 0$. Ferguson (1973) has shown that the random distribution F drawn from a Dirichlet process is almost surely discrete, although the shape parameter G_0 can be continuous. So F can be written as $F = (\Theta, p)$, where $\Theta = (\theta_1, \theta_2, \ldots)'$, $p = (p_1, p_2, \ldots)'$ with $p_i > 0$ and $\sum_{i=1}^{\infty} p_i = 1$. Each distinct value is represented by θ_i and its corresponding probability is p_i, for $i = 1, 2, \ldots$. Sethuraman (1994) defined the stick-breaking representation of the Dirichlet process as follows:

[1] For an infinite hidden Markov model without regime persistence, see Teh *et al*. (2006). For other nonparametric modelling methods using the Dirichlet processes but without the hidden Markov representation, see Bassetti *et al*. (2011), Griffin and Steel (2011) and Griffin (2011).

DOI: 10.1002/jae

$$V_i \overset{\text{i.i.d.}}{\sim} \mathbf{B}(1, \alpha_0), \quad p_i = V_i \prod_{j=1}^{i-1} \left(1 - V_j\right) \tag{1}$$

$$\theta_i \overset{\text{i.i.d.}}{\sim} G_0 \tag{2}$$

The notation **B** represents the beta distribution. This representation shows that p and Θ are independent. The process (1), which generates p, is called the *stick-breaking process* and is denoted by $\mathbf{SBP}(\alpha_0)$ in the rest of the paper.

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) as follows:

$$p \sim \mathbf{SBP}(\alpha_0), \quad \theta_i \overset{\text{i.i.d.}}{\sim} G_0 \quad \text{for } i = 1, 2, \ldots, \quad g(y) = \sum_{i=1}^{\infty} p_i f\left(y|\theta_i\right). \tag{3}$$

The probability density function $g(y)$ is an infinite mixture of the probability density functions $f(y|\theta_i)$'s. If $f(y|\theta_i)$ is the normal distribution density function and θ_i represents the mean and the variance y is distributed as an infinite mixture of normal distributions. Hence continuous random variables can be modelled nonparametrically by the DPM model. Otranto and Gallo (2002) applied this approach to detect the number of regimes. The DPM model, however, is usually applied to cross-sectional data because it ignores regime dependence over time.

3. INFINITE HIDDEN MARKOV MODEL

First, I introduce some notation. Let y_t represent the observed variable of interest and s_t the regime indicator, which is a discrete unobserved random variable on the natural numbers. An infinite-dimensional vector $\pi_i = (\pi_{i1}, \pi_{i2}, \ldots)$ represents the distribution of s_t when $s_{t-1} = i$. An infinite-dimensional matrix $P = \left(\pi_1', \pi_2', \ldots\right)'$ is the collection of all π_is. Each π_i is generated based on an infinite discrete distribution $\pi_0 = (\pi_{01}, \pi_{02}, \ldots)$, whose prior is a stick-breaking process with parameter γ. The rest of the parameters for generating π_is are the concentration parameter c and the *sticky coefficient* ρ, which will be discussed in detail later. Let θ_i represent the parameter which characterizes regime i and define Θ as the collection of all θ_is. The parameter θ_i is assumed to follow a hierarchical distribution $G_0(\lambda)$. A vector $Y_{1,t} = (y_1, \ldots, y_t)'$ represents the data up to time t. The symbol δ_i denotes the measure of a point mass of 1 at i. The model is as follows:

$$y_t|s_t, \Theta, Y_{1,t-1} \sim f\left(y_t|\theta_{s_t}, Y_{1,t-1}\right) \tag{4}$$

$$s_t|s_{t-1} = i, P \sim \pi_i \tag{5}$$

$$\pi_i|\pi_0, c, \rho \sim \mathbf{DP}(c, (1-\rho)\pi_0 + \rho\delta_i), \quad c > 0, \rho \in (0, 1) \tag{6}$$

$$\pi_0|\gamma \sim \mathbf{SBP}(\gamma), \quad \gamma > 0 \tag{7}$$

$$\theta_i|\lambda \sim G_0(\lambda) \tag{8}$$

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

where $i = 1, 2, \ldots$ and $\mathcal{G}$ is the prior for λ. Note that the hyperparameters γ, c and ρ are fixed in this model. We can draw inference for these parameters by assuming priors on them, which will be discussed in Section 5.1.

DOI: 10.1002/jae

The *first* hierarchical structure, which governs the transition probabilities, comprises equations (6) and (7). The distribution π_0 is the hierarchical distribution drawn from a stick-breaking process and has the natural numbers as its support. Each infinite-dimensional vector π_i is drawn from a Dirichlet process with the concentration parameter c and the shape parameter $(1-\rho)\pi_0+\rho\delta_i$. There are two points worth noting for clarity. First, because the shape parameter $(1-\rho)\pi_0+\rho\delta_i$ has support only on the natural numbers and each number is associated with a non-zero probability, the random distribution π_i can only take values on the natural numbers and each value has positive probability. We can sum up the probabilities for the same integer value so that an element π_{ij} represents the probability of s_t taking the integer value j given that $s_{t-1}=i$. Then, the infinite dimensional matrix P is simply the transition matrix of a hidden Markov model. Second, the larger ρ, the larger is the expected probability of π_i at integer i. This implies that s_t is more likely to be equal to s_{t-1}. Hence ρ captures state persistence and is referred as the sticky coefficient. Conditional on π_0 and ρ, the mean of the transition matrix $E(P|\pi_0,\rho)=(1-\rho)\left(\pi_0^{'},\pi_0^{'},\ldots\right)^{'}+\rho\mathbf{I}_\infty$ is a convex combination of two infinite-dimensional matrices. The sticky coefficient ρ increases the self-transition probabilities by adding weights to the infinite-dimensional identity matrix $\mathbf{I}_\infty$. The concentration parameter c controls how close P is to $E(P|\pi_0,\rho)$.

A common practice of setting the prior for the transition matrix of a finite Markov switching model assumes each row of the transition matrix is drawn from a Dirichlet distribution independently. In the extension to the infinite dimension, each row π_i should be drawn from a stick-breaking process. However, Teh *et al*. (2006) argued that this prior may have an overparametrization problem without a hierarchical structure similar to equations (6) and (7), because it precludes each π_i from sharing information with each other. In terms of parsimony, the iHMM only needs one stick-breaking process for the hierarchical distribution π_0, instead of assuming an infinite number of stick-breaking processes for the transition matrix P.

The *second* hierarchical structure, which governs the parameters of the conditional data density, includes equations (8) and (9). A non-hierarchical prior for θ_i would only include equation (8) with a fixed value of λ. In the case of a structural break, a new conditional data density parameter will be drawn from such pre-fixed distribution $G_0(\lambda)$, which may harm the out-of-sample forecasting. In this paper, I assume a prior $\mathcal{G}$ on λ to draw inference from its posterior distribution. This methodology provides a way of learning λ from past values of θ_i to improve estimation and forecasting.[2]

In comparison to the iHMM, FSJW does not have equation (9). The stick-breaking representation of the Dirichlet process is not fully exploited by FSJW, since it has only one hierarchical structure, which is on the transition probabilities. In fact, the stick-breaking representation (1)–(2) decomposes the random probability measure F drawn from a Dirichlet process into two independent parts: the probabilities are generated from a stick-breaking process and the parameter values are drawn from the shape parameter. The iHMM takes fuller advantage of this structure than FSJW by modelling two parallel hierarchical structures.

4. ESTIMATION AND FORECAST

I assume the conditional density of y_t in Equation (4) follows a Gaussian AR(q) process:

$$y_t|s_t,\Theta,Y_{1,t-1}\sim\mathbf{N}\left(\phi_{s_t,0}+\phi_{s_t,1}y_{t-1}+\ldots+\phi_{s_t,q}y_{t-q},\sigma_{s_t}^2\right) \tag{10}$$

The conditional data density parameter is $\theta_i=\left(\phi_i^{'},\sigma_i\right)^{'}$ with $\phi_i=(\phi_{i0},\phi_{i1},\ldots,\phi_{iq})'$ for $i=1,2,\ldots$.

[2] Pesaran *et al*. (2006) emphasized the importance of modelling the hierarchical distribution in the presence of structural breaks.

DOI: 10.1002/jae

4.1. Priors

The hierarchical distribution $G_0(\lambda)$ in (8) is assumed as the standard conjugate normal-gamma distribution in the Bayesian literature (Geweke, 2005):

$$\sigma_i^{-2} \sim \mathbf{G}\left(\frac{\chi}{2}, \frac{v}{2}\right), \quad \phi_i|\sigma_i \sim \mathbf{N}\left(\phi, \sigma_i^2 H^{-1}\right), \qquad \text{for } i = 1, 2, \ldots \tag{11}$$

By definition, the collection of the hierarchical parameters is $\lambda = (\phi, H, \chi, v)$. The $(q+1) \times 1$ vector ϕ is the mean of the regression coefficients ϕ_i. The $(q+1) \times (q+1)$ positive definite matrix H is proportional to the precision matrix of ϕ_i. The scalars $\frac{\chi}{2} > 0$ and $\frac{v}{2} > 0$ are the scale parameter and the degree of freedom of the gamma distribution, respectively.

The prior on the collection of the hierarchical parameters λ in equation (9) is set as

$$H \sim \mathbf{W}(A_0, a_0); \quad \phi|H \sim \mathbf{N}\left(m_0, \tau_0 H^{-1}\right); \quad \chi \sim \mathbf{G}\left(\frac{d_0}{2}, \frac{c_0}{2}\right); \quad v \sim \mathbf{Exp}(\rho_v) \tag{12}$$

The matrix H is drawn from a Wishart distribution with parameter A_0, which is a $(q+1) \times (q+1)$ positive definite matrix, and the degree of freedom $a_0 > q$. The vector m_0 is $(q+1) \times 1$ and represents the mean of ϕ. The scalar $\tau_0 > 0$ controls for the dispersion of φ conditional on H. The parameter χ's prior is a gamma distribution with the scale parameter of $\frac{d_0}{2}$ and the degree of freedom of $\frac{c_0}{2}$. The parameter v has an exponential distribution with mean ρ_v.

4.2. Beam Sampler

Fox *et al*. (2011) truncate the number of regimes in order to simulate the model parameters from their posterior distribution. This paper follows Jochmann (2010) by applying the beam sampler of Van Gael *et al*. (2008) to avoid such approximation.

The idea is to introduce a latent variable $u_t > 0$ in equation ((5)) such that the joint density of (u_t, s_t) conditional on $s_{t-1} = i$ and P is given by $1\left(u_t < \pi_{i,s_t}\right)$. Clearly, marginalizing out u_t returns π_{i,s_t} and gives the original model back. The conditional density $f(u_t|s_t)$ is simply $\frac{1\left(u_t < \pi_{i,s_t}\right)}{\pi_{i,s_t}}$, so u_t is conditionally uniform distributed. The conditional probability $f(s_t|u_t)$ is $\frac{1\left(u_t < \pi_{i,s_t}\right)}{\sum_{j=1}^{\infty} 1\left(u_t < \pi_{ij}\right)}$. It is easy to see that $\{j : u_t < \pi_{ij}\}$ is a finite set for all $u_t > 0$. Hence the variable u_t turns an infinite dimension problem into a finite one. The details are shown in the web Appendix (supporting information).

4.3. Estimation

The posterior sampling is based on Markov chain Monte Carlo (MCMC) methods and shown in the web Appendix. For a brief summary, the parameter space is partitioned into four parts: S, (Θ, P, π_0), (ϕ, H, χ) and v.[3] Each part is sampled conditional on the others and the data.

After initializing the parameter values, the algorithm is applied iteratively many times to obtain a large sample of the model parameters. A first block of the sample is discarded as the burn-in sample. The rest of the sample, $\left\{S^{(i)}, \Theta^{(i)}, P^{(i)}\pi_0^{(i)}, \phi^{(i)}, H^{(i)}, \chi^{(i)}, v^{(i)}\right\}_{i=1}^{N}$, is used for inferences as

[3] The auxiliary variables are discussed in detail in the web Appendix.

DOI: 10.1002/jae

if they were drawn from the posterior distribution. Simulation-consistent posterior statistics are computed as sample averages. For example, the posterior mean of ϕ is computed as $\frac{1}{N}\sum_{i=1}^{N}\phi^{(i)}$.

One problem for the posterior inference from mixture models is label switching (Celeux *et al.*, 2000; Frühwirth-Schnatter, 2001). Namely, if we switch the labels of any two regimes and their corresponding parameter values, the likelihood value does not change. Hence the regime indicator s_t and other label-dependent statistics are not identified. For instance, the probability for the observation at time t being in regime 1, $P(s_t = 1|Y_{1,T})$, is label dependent and meaningless. Because, if regime 1 and regime 2's labels are switched, the new regime indicator $s_t^{'} = 1$ if $s_t = 2$ and $s_t^{'} = 2$ if $s_t = 1$. Hence the value of $P\left(s_t^{'} = 1 \middle| Y_{1,T}\right)$ is actually $P(s_t = 2|Y_{1,T})$. This example is simplified. In practice, the label switching could happen back and forth without being noticed in an MCMC, which makes the inference on the label dependent statistics misleading.

Meanwhile, Geweke (2007) has shown that the inference on the label-invariant statistics is not affected. For the same example, we are able to draw inference on whether the data at time t_0 and t_1 are in the same regime as $P\left(s_{t_0} = s_{t_1} | Y_{1,T}\right)$. This is because $P\left(s_{t_0}^{'} = s_{t_1}^{'} \middle| Y_{1,T}\right) = P\left(s_{t_0} = s_{t_1} \middle| Y_{1,T}\right)$ in the presence of label switching.

In this paper, I compute the predictive likelihood $p(y_{t+1}|Y_{1,t})$, the posterior mean of φ_{s_t} and σ_{s_t} and the posterior probability of clustering $P(s_t = s_\tau|Y_{1,T})$. All these statistics are label invariant. Among these values, the predictive likelihood is the most important and will be explained next in detail.

4.4. Model Comparison

I define $\Psi^{(i)} = \left\{S^{(i)}, \Theta^{(i)}, P^{(i)}\pi_0^{(i)}, \phi^{(i)}, H^{(i)}, \chi^{(i)}, v^{(i)}\right\}$ as one draw of the parameters from the posterior distribution conditional on data $Y_{1,t}$. The number of active regimes $L^{(i)}$, where an active regime means a regime having at least one observation, can be inferred from the regime allocation $S^{(i)}$. After relabelling the active regimes to $1, 2, \ldots, L^{(i)}$, the conditional predictive density at time $t+1$ is $p\left(\widetilde{y}_{t+1} \middle| \Psi^{(i)}, Y_{1,t}\right) = \sum_{j=1}^{L^{(i)}} \pi_{s_t^{(i)}, j}^{(i)} f\left(\widetilde{y}_{t+1} \middle| \theta_j^{(i)}, Y_{1,t}\right) + \pi_{s_t^{(i)}, \widetilde{L}^{(i)}}^{(i)} f\left(\widetilde{y}_{t+1} \middle| \theta_{\widetilde{L}^{(i)}}^{(i)}, Y_{1,t}\right)$, where $\widetilde{y}_{t+1}$ is the random variable of interest at time $t+1$. The residual probability $\pi_{s_t^{(i)}, \widetilde{L}^{(i)}}^{(i)} = 1 - \sum_{j=1}^{L^{(i)}} \pi_{s_t^{(i)}, j}^{(i)}$ represents the structural break probability. If a structural break occurs, the new parameter for the conditional data density is $\theta_{\widetilde{L}^{(i)}}^{(i)}$. After integrating out these parameters, we can have the predictive density as $\hat{p}\left(\widetilde{y}_{t+1} \middle| Y_{1,t}\right) = \frac{1}{N}\sum_{i=1}^{N} p\left(\widetilde{y}_{t+1} \middle| \Psi^{(i)}, Y_{1,t}\right)$. The predictive likelihood $\hat{p}\left(y_{t+1} \middle| Y_{1,t}\right)$ is obtained if we replace $\widetilde{y}_{t+1}$ by the observation y_{t+1}. The marginal likelihood is computed as the product of the predictive likelihoods from $t = 1$ to T as $\hat{p}\left(Y_{1,T}\right) = \prod_{t=1}^{T} \hat{p}\left(y_t | Y_{1,t-1}\right)$. If we use a training sample from $t = 1$ to $t_0 - 1$, the predictive likelihood from $t = t_0$ to T is $\hat{p}\left(Y_{t_0,T} \middle| Y_{1,t_0-1}\right) = \prod_{t=t_0}^{T} \hat{p}\left(y_t | Y_{1,t-1}\right)$.

I compare the iHMM to alternative models by using both the marginal and the predictive likelihood. Kass and Raftery (1995) compared model M_i to M_j by the log Bayes factor: $log\left(\mathrm{BF}_{ij}\right) = log\left(\frac{p\left(Y_{1,T}|M_i\right)}{p\left(Y_{1,T}|M_j\right)}\right)$. They suggested interpreting the evidence for M_i versus M_j as: not worth more than a bare mention if $0 \le log(\mathrm{BF}_{ij})$ 1; positive if $1 \le log(\mathrm{BF}_{ij}) < 3$; strong if $3 \le log(\mathrm{BF}_{ij}) < 5$; very strong if $log(\mathrm{BF}_{ij}) \ge 5$. Geweke and Amisano (2010) argued that the predictive likelihood comparison is robust to the prior elicitation.

DOI: 10.1002/jae

The interpretation of the log predictive Bayes factor $log\left(\mathrm{BF}_{ij}\middle|Y_{1,t_0-1}\right) = log\left(\frac{p\left(Y_{t_0,T}\middle|Y_{1,t_0-1},M_i\right)}{p\left(Y_{t_0,T}\middle|Y_{1,t_0-1},M_j\right)}\right)$ is the same as the log Bayes factor if the initial sample Y_{1,t_0-1} is regarded as a training sample.

Similarly, the predictive mean is obtained as $\hat{E}\left(\widetilde{y}_{t+1}\middle|Y_{1,t}\right) = \frac{1}{N}\sum_{i=1}^{N} E\left(\widetilde{y}_{t+1}\middle|\Psi^{(i)}, Y_{1,t}\right)$. A traditional measure for model selection is the root mean squared forecast error (RMSFE), which is computed as $\sqrt{\frac{\sum_{t=t0}^{T}\left(y_t - \hat{E}\left(\widetilde{y}_t\middle|Y_{1,t-1}\right)\right)^2}{T-t_0+1}}$.

To evaluate the long-run forecasting ability, I compute the long-run predictive likelihood $p(y_{t+h}|Y_{1,t})$, where $h > 1$ is the forecasting horizon. In detail, for each draw of $s_t^{(i)}$ from the posterior sample conditional on $Y_{1,t}$, the regime indicators $\widetilde{s}_{t+1}^{(i)}, \ldots, \widetilde{s}_{t+h}^{(i)}$ are simulated according to the transition matrix $P^{(i)}$. If a structural break occurs (the regime indicator exceeds the number of active regimes $L^{(i)}$), the state space is expanded and a new set of parameters is drawn for the new regime. Then, we can simulate $\widetilde{y}_{t+1}^{(i)}, \ldots, \widetilde{y}_{t+h-1}^{(i)}$ by following equation (10) and use them as the observed data to evaluate the conditional long-run predictive likelihood $p\left(y_{t+h}\middle|\Psi^{(i)}, Y_{1,t}, \widetilde{y}_{t+1}^{(i)}, \ldots, \widetilde{y}_{t+h-1}^{(i)}\right) = f\left(y_{t+h}\middle|\theta_{\widetilde{s}_{t+h}}^{(i)}, Y_{1,t}, \widetilde{y}_{t+1}^{(i)}, \ldots, \widetilde{y}_{t+h-1}^{(i)}\right)$. The long-run predictive likelihood is obtained by integrating out the estimation uncertainties as $\hat{p}\left(y_{t+h}\middle|Y_{1,t}\right) = \frac{1}{N}\sum_{i=1}^{N} p\left(y_{t+h}\middle|\Psi^{(i)}, Y_{1,t}, \widetilde{y}_{t+1}^{(i)}, \ldots, \widetilde{y}_{t+h-1}^{(i)}\right)$.

5. APPLICATION TO US REAL INTEREST RATES

The dynamic stability was tested by Fama (1975); Rose (1988) and Walsh (1987). 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 US real interest rates of Huizinga and Mishkin (1986) from 1961Q1 to 1986Q3. The real interest rate dynamics in each state are characterized by a Gaussian AR(2) process. Wang and Zivot (2000) used the same data to investigate structural breaks and found three breaks at 1970Q3, 1980Q2 and 1985Q4.

This paper constructs US quarterly real interest rates in the same way as Huizinga and Mishkin (1986) and extends their dataset to a total of 252 observations from 1947Q1 to 2009Q4. The data are plotted in the top panel of Figure 1.

5.1. Models and Priors

For the iHMM, I assume that each regime has an AR(q) representation and label the model as iHMM-AR(q). It uses the following prior: $\gamma = 1, c = 10, a_0 = q + 10, A_0 = \frac{1}{a_0}\mathbf{I}_{q+1}, m_0 = \underset{\bar{}q+1}{0}, \tau_0 = 1, d_0 = 1, c_0 = 5$ and $\rho_v = c_0 + 2$. I estimate the model for $\rho = 0.5$, 0.9 or 0.95 and $q = 1$, 2, 3 or 4.[4] The values of A_0 and a_0 imply that the prior mean of H is an identity matrix. The zero vector m_0 reflects that the prior mean of ϕ is zero. The values of d_0, c_0 and ρ_v imply that a typical draw of σ_i^{-2} is centred around 1. This prior is informative and its range is shown in Table I if each regime has AR(1) dynamics.

An extension of the iHMM, which is labelled as iHMM-hyper-AR(q) for $q = 1, 2, 3$ and 4, estimates the hyperparameters γ, c and ρ. I assume $\gamma \sim \mathbf{G}(1,1)$ and $c \sim \mathbf{G}(1,10)$. The prior means of γ and c are

[4] Section 7 of the web Appendix reports robustness checks on γ and c.

DOI: 10.1002/jae

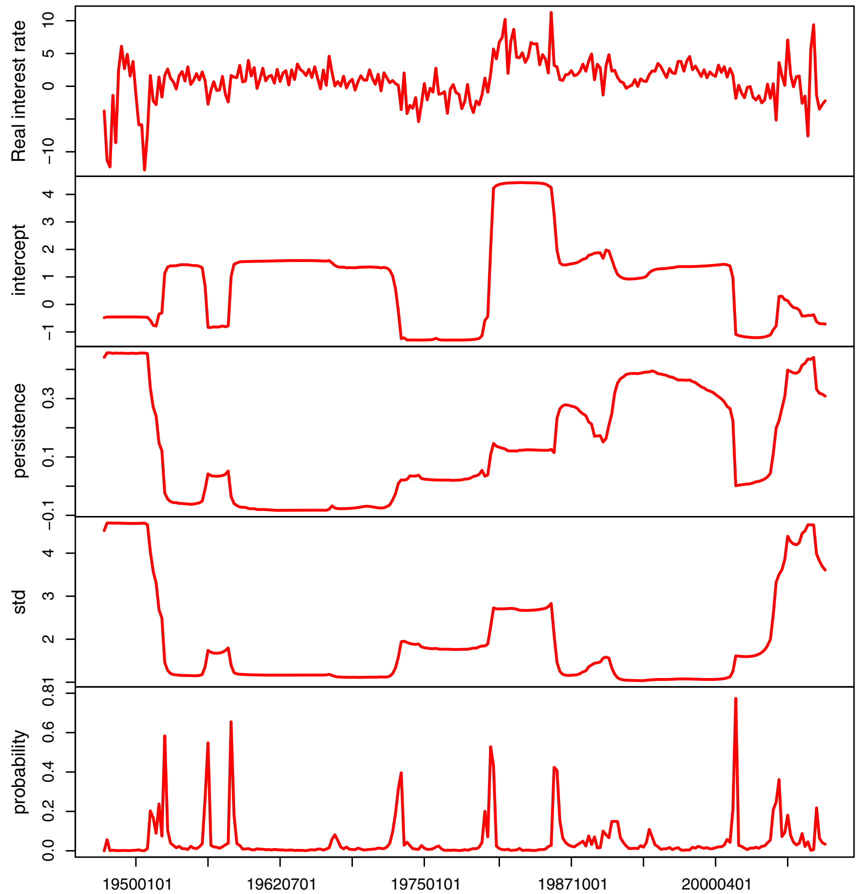

Figure 1. There are 252 observations from 1947Q1 to 2009Q4 for US quarterly real interest rates. The data are estimated by iHMM-AR(1) with $\rho = 0.95$. The first panel plots the data; the second panel plots the posterior mean of the intercepts; the third panel plots the posterior mean of the persistence parameters; the fourth panel plots the posterior mean of the standard deviations; and the last panel plots the posterior probabilities of regime changing

Table I. Prior and posterior summary

	Prior			Posterior		
	Mean	SD	90% DI	Mean	SD	90% DI
χ	5.00	2.24	(1.97, 9.15)	6.26	2.72	(2.38, 13.0)
v	7.00	7.00	(0.36, 21.0)	3.03	0.91	(2.03, 5.33)
φ_0	0.00	1.41	(−1.90, 1.90)	0.62	0.67	(−0.73, 1.93)
φ_1	0.00	1.41	(−1.90, 1.90)	0.08	0.49	(−0.89, 1.10)
H_{00}	1.00	0.58	(0.42, 1.79)	0.98	0.36	(0.42, 1.80)
H_{11}	1.00	0.58	(0.42, 1.79)	1.52	0.55	(0.67, 2.76)
H_{01}	0.00	0.41	(−0.49, 0.49)	−0.02	0.30	(−0.60, 0.55)
# of regimes: L	1.75	0.89	(1.00, 3.00)	6.33	1.50	(4.00, 9.00)

Note: This table shows the prior and posterior summary of the parameters of the iHMM-AR(1) with $\rho = 0.95$. There are 252 observations from 1947Q1 to 2009Q4 for US quarterly real interest rates.

equal to the fixed values used in the iHMM. For ρ, I tried $\rho \sim \mathbf{B}(9.5,0.5)$ as an informative prior and $\rho \sim \mathbf{B}(1,1)$ as an uninformative prior. The rest of the priors are the same as that of the iHMM-AR(q).[5]

The finite Markov switching and the structural break models are estimated for model comparison. For a K-regime Markov switching model with AR(q) dynamics in each regime, which is denoted by MS(K)-AR(q), I assume that the transition probability is given by

$$P(s_t = j | s_{t-1} = i) = P_{ij} \tag{13}$$

$$(P_{i1}, \ldots, , P_{iK}) \sim \mathbf{Dir}(1, \ldots, 1) \tag{14}$$

for $i, j = 1, \ldots, K$. The MS(K) model is comprised of equations (4)–(5), (13)–(14) and (8)–(9). Hence the difference between an MS(K) model and the iHMM is how the transition probabilities are constructed. The model is estimated for $K = 2, 3, 4$ and $q = 1, 2, 3, 4$.

For an MS(K) model with a larger number of regimes ($K > 4$), it is equivalent to an iHMM without the sticky coefficient from the empirical view. Because some regimes become inactive for a large K, the structural breaks are implicitly modelled as in Fox *et al.*'s (2011) block sampling scheme. Hence a large K is able to show whether it is useful to include structural breaks in the regime switching dynamics. I call a regime switching model *large* MS(K)-AR(q) if $K \geq 5$. The model is estimated for $K = 5, \ldots, 10$ and $q = 1, \ldots, 4$. The rest of the priors are the same as that of the iHMM.

For the structural break dynamics, Pesaran *et al.*'s (2006) model with some modification is estimated. For a finite K-regime structural break model with AR(q) dynamics in each regime, which is denoted by SB(K)-AR(q), the transition matrix is constructed as

$$P(s_t = i | s_{t-1} = i) = \begin{cases} p & \text{if } \ i < K(15) \\ 1 & \quad \text{if } \ i = K \end{cases} \tag{15}$$

$$P(s_t = i + 1 | s_{t-1} = i) = 1 - p \quad \text{if } \ i < K \tag{16}$$

$$p \sim \mathbf{B}(9, 1) \tag{17}$$

where $i = 1, \ldots, K$. The SB(K) model is comprised of equations (4)–(5), (15)–(17) and (8)–(9). Unlike Pesaran *et al.* (2006), I assume a conjugate prior for θ_i and a homogeneous structural break probability *1-p*. The rest of the priors are the same as that of the iHMM. The SB(K)-AR(q) model is estimated for $K = 2, \ldots, 10$ and $q = 1, \ldots, 4$. In the h-step-ahead out-of-sample forecast, I allow a maximum of h out-of-sample breaks.

A linear AR(q) model, for $q = 1, \ldots, 4$, is applied as a benchmark:

$$y_t | Y_{t-1}, \phi, \sigma \sim \mathbf{N}\left(\phi_0 + \phi_1 y_{t-1} + \ldots + \phi_q y_{t-q}, \sigma^2\right) \tag{18}$$

The prior of (ϕ,σ) is set as $\sigma^{-2} \sim \mathbf{G}(2.5,3.5)$ and $\phi | \sigma \sim \mathbf{N}(0, \sigma^2 I)$. These prior parameter values are equal to the means of the hierarchical priors of the iHMM.

I also estimate a linear AR(q) model, for $q = 1, \ldots, 4$, with rolling windows in order to check whether it can account for the nonlinear dynamics of the data. The prior is the same as that of the linear AR(q) model. The rolling windows are 3, 5, 10 and 20 years.

[5] Section 7 of the web Appendix shows the posterior summary of γ, c and ρ from the iHMM-hyper-AR(1).

DOI: 10.1002/jae

The last candidate is the Bayesian model averaging (BMA) approach. The first BMA, denoted by BMA-MS-AR(q), includes the MS(K)-AR(q) models for $K = 1, \ldots, 4$. The MS(1)-AR(q) model is equivalent to the linear AR(q) model (19) plus the hierarchical structure (8) and (9). The second is denoted by BMA-SB-AR(q) and includes 10 structural break models from SB(1)-AR(q) to SB(10)-AR(q), among which the SB(1)-AR(q) is equivalent to the MS(1)-AR(q). The last one is denoted by BMA-MS+SB-AR(q) and includes all of the above Markov switching and structural break models. I estimate these models for $q = 1, 2, 3$ and 4.

5.2. Marginal Likelihood

Table II shows the log marginal likelihoods. Each column represents the dynamics in each regime. The first panel shows the iHMM and the iHMM-hyper models; the second panel reports the MS(K) and the SB(K) models; the BMA approaches and the large MS(K) models are in the third panel; and the linear and the rolling AR(q) models are in the last panel. For the SB(K) and the large MS(K) models, I only report the largest value for each q and the corresponding value of K. Figure 2 plots all the log marginal likelihoods for the MS(K), the large MS(K) and the SB(K) models.

The best model is the iHMM-AR(1) with $\rho = 0.95$. It dominates the MS(K) and the SB(K) models strongly. The log Bayes factors of the best iHMM against the best MS(K) and the best SB(K) model are 11.1 and 18.3, respectively. Hence a regime switching or structural break model alone is not enough to describe the data dynamics. Figure 2 plots the marginal likelihood of the best model as a horizontal line. The second and the third best model are the iHMM-AR(1) with $\rho = 0.9$ and the iHMM-hyper-AR(1) with $\rho \sim \mathbf{B}(9.5,0.5)$, which support the high regime persistence. The iHMM-hyper with $\rho \sim \mathbf{B}(1,1)$ is robust to the number of lags q and superior to the MS(K) and the SB(K) models as well.

The BMA approaches are not able to improve the log marginal likelihood. A large MS(K) model, however, can improve the density forecast. Since adding more regimes to the MS(K) model is able to accommodate structural breaks, Table II implies that the structural break dynamics are important

Table II. Log marginal likelihood

Model	AR(1)	AR(2)	AR(3)	AR(4)
iHMM $\rho = 0.5$	−551.8	−554.8	−547.6	−555.0
iHMM $\rho = 0.9$	−539.0	−550.7	−553.0	−551.6
iHMM $\rho = 0.95$	−538.6	−554.2	−551.4	−554.6
iHMM-hyper $\rho \sim \mathbf{B}(1,1)$	−542.5	−545.1	−544.5	−547.1
iHMM-hyper $\rho \sim \mathbf{B}(9.5,0.5)$	−539.2	−554.1	−550.7	−551.6
MS(2)	−556.4	−559.2	−554.9	−555.0
MS(3)	−555.0	−553.9	−554.5	−558.2
MS(4)	−557.4	−549.7	−552.7	−555.7
SB(K)	−556.9 SB(8)	−572.2 SB(5)	−573.4 SB(5)	−570.8 SB(6)
BMA-MS	−556.0	−551.1	−553.8	−556.0
BMA-SB	−560.6	−560.8	−566.2	−573.4
BMA-MS+SB	−557.2	−552.2	−555.0	−557.1
Large MS(K)	−559.4 MS(5)	−542.7 MS(10)	−542.6 MS(8)	−546.8 MS(9)
AR model	−591.0	−592.6	−600.0	−586.9
Rolling AR (3 years)	−558.0	−573.9	−599.5	−595.0
Rolling AR (5 years)	−557.3	−564.7	−584.4	−575.4
Rolling AR (10 years)	−585.9	−583.4	−593.6	−579.8
Rolling AR (20 years)	−614.2	−611.4	−620.8	−602.2

Note: There are 252 observations from 1947Q1 to 2009Q4 for US quarterly real interest rates. The marginal likelihoods are computed by using the last 248 observations for all models. This is because the AR(4) dynamics require 4 lags from the data.

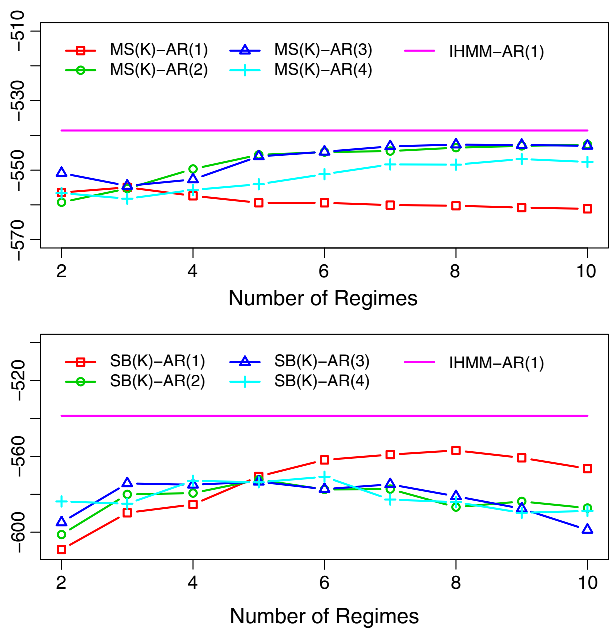

Figure 2. The top panel plots the marginal likelihoods of the MS(K) models; the bottom panel plots the marginal likelihoods of the SB(K) models

for the data. Meanwhile, because the SB(K) model or the BMA-SB approach does not perform as well as the best large MS(K) model, we can say that the regime switching dynamics are also crucial for the data. Lastly, the best iHMM still strongly dominates the large MS(K) models. The log Bayes factor of the best iHMM against the best large MS(K) model is 4.0. Hence, explicitly modelling the regime persistence can improve the marginal likelihood.

The linear and rolling AR(q) models perform the worst. The best of these models is the 5-year rolling AR(2) model, but it is still strongly dominated by the iHMM, the iHMM-hyper, the MS(K) and the SB(K) models.

5.3. Predictive Likelihood

Table III shows the log predictive likelihoods of the last 200 observations, which account for 80% of the sample size. For the large MS(K) models, the value from the best model is reported in the table for each q. Figure 3 plots all the log predictive likelihoods of the MS(K) and the large MS (K) models.

For the SB(K) models, Pesaran *et al.*'s (2006) approach implicitly assumes that the number of regimes depends on the sample size. Hence, for a fixed K, an SB(K) model estimated on different subsamples has different interpretations. Following their paper, I report the log predictive likelihoods of the BMA-SB-AR(q) models.

The conclusion is basically the same as that of the log marginal likelihood comparison. The best three models are still the iHMM-AR(1) with $\rho = 0.95$, the iHMM-AR(1) with $\rho = 0.9$ and the iHMM-hyper-AR(1) with $\rho \sim \mathbf{B}(9.5,0.5)$. Figure 3 plots the predictive likelihood of the best model as a horizontal line. For the iHMM-hyper with $\rho \sim \mathbf{B}(1,1)$, it is not only still robust to the choice of q, but also dominates the rest of the alternative models.

DOI: 10.1002/jae

Table III. Log predictive likelihood

Model	AR(1)	AR(2)	AR(3)	AR(4)
iHMM $\rho = 0.5$	−428.2	−429.9	−424.1	−429.8
iHMM $\rho = 0.9$	−418.9	−421.6	−421.5	−425.7
iHMM $\rho = 0.95$	−418.8	−423.3	−421.0	−426.7
iHMM-hyper $\rho \sim \mathbf{B}(1,1)$	−419.9	−422.6	−422.1	−426.1
iHMM-hyper $\rho \sim \mathbf{B}(9.5,0.5)$	−419.0	−427.0	−422.2	−427.0
MS(2)	−433.4	−431.5	−423.8	−429.0
MS(3)	−432.3	−432.2	−429.3	−435.4
MS(4)	−435.7	−428.7	−431.1	−434.5
BMA-MS	−433.0	−429.1	−430.9	−433.5
BMA-SB	−435.6	−432.3	−438.1	−444.7
BMA-MS+SB	−433.7	−429.1	−430.9	−433.5
Large MS(K)	−438.2 MS(5)	−423.7 MS(10)	−422.9 MS(9)	−426.5 MS(9)
AR model	−462.4	−456.5	−450.1	−455.8
Rolling AR (3 years)	−436.2	−441.5	−453.0	−466.1
Rolling AR (5 years)	−434.2	−433.2	−437.3	−443.6
Rolling AR (10 years)	−458.4	−448.3	−444.2	−449.2
Rolling AR (20 years)	−485.6	−475.4	−470.9	−471.1

Note: There are 252 observations from 1947Q1 to 2009Q4 for US quarterly real interest rates. The last 200 observations are used for computing the predictive likelihoods.

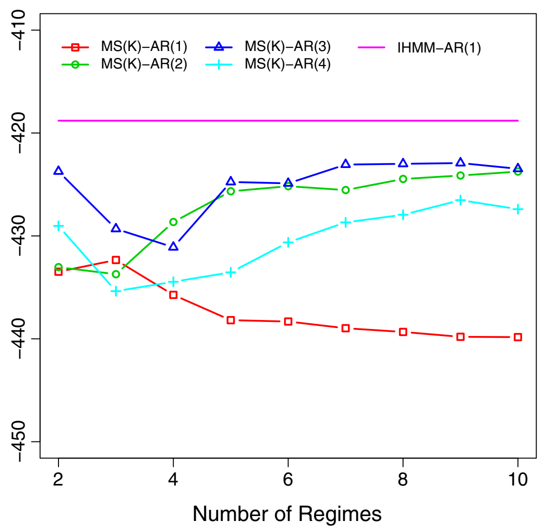

Figure 3. This figure plots the predictive likelihoods of the MS(K) models

A large MS(K) model is still able to improve the density forecast over the MS(K) model or the BMA approach, but to a less extent. The best large MS(K) model is still strongly dominated by the best iHMM, so explicit modelling the regime persistence can improve the density forecast.

The linear and rolling AR(q) models perform the worst. Among these models, the best is still the 5-year rolling AR(2) model.

To further investigate how the iHMM performs in comparison to the regime switching and the structural break dynamics over time, the top panel of Figure 4 plots the cumulative log Bayes factors of the best iHMM against the best BMA-MS model in Table II. The bottom panel of

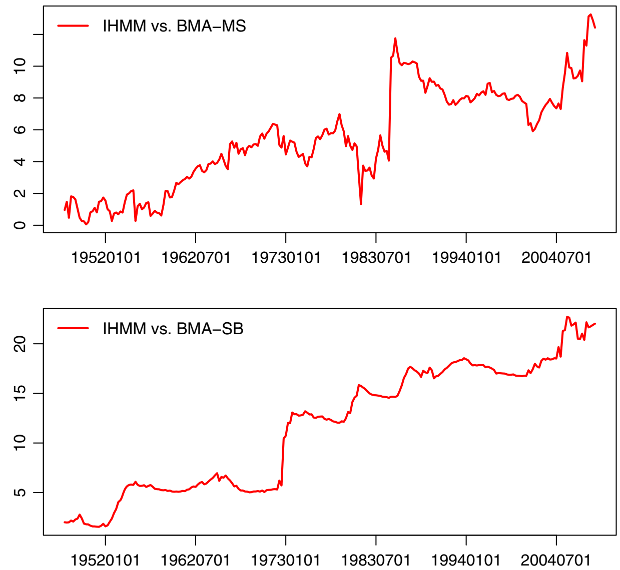

Figure 4. The top panel plots the cumulative log Bayes factors of iHMM-AR(1) with $\rho = 0.95$ against BMA-MS-AR(2). The bottom panel plots the cumulative log Bayes factors of iHMM-AR(1) with $\rho = 0.95$ against BMA-SB-AR(1)

Figure 4 plots the cumulative log Bayes factors of the best iHMM against the best BMA-SB model in Table II.

From the top panel, an upward trend shows that the iHMM is gaining more evidence against the BMA-MS as more data are observed. Before 1972, the cumulative log Bayes factor increases steadily. After 1972, its value is more volatile; however, the upward trend of this series is not affected.

In the bottom panel, the iHMM performs almost uniformly better than the BMA-SB approach. An upward jump of the cumulative log Bayes factor is usually associated with a regime change to an existing regime. For example, it is increased by 4.7 in 1972Q4 and the probability of switching to an existing regime is 0.31, while the probability of entering a new regime is 0.04.[6]

5.4. Point Forecast and Long-Run Forecast

Table IV shows the RMSFE and the log long-run predictive likelihoods. I report the models with the largest predictive likelihoods from each model category in Table III. The same training sample as in the predictive likelihood comparison is used. A number in bold font represents the optimal value in its column.

For the RMSFE, the linear AR(3) model is the best. The MS(2)-AR(3) and the iHMM also perform well. The large MS(K) model and the iHMM-hyper perform worse than the best models. The BMA approaches are the worst for point forecasting.

The reason for the iHMM does not win in the RMSFE comparison could be attributed to the fact that it focuses on the overall density forecast instead of the point forecast.[7] This is analogous to forecasting at different horizons. A model that is good at short-run forecasting is not necessarily good at long-run forecasting. Next, I investigate the long-run forecasting ability of the iHMM.

[6] The algorithm to identify the regime switching and structural break points is in the web Appendix.
[7] Pesaran *et al*. (2011) discussed mean forecasting in the presence of structural instability.

DOI: 10.1002/jae

Table IV. RMSFE and long-run density forecast

Model	RMSFE	2Qs	3Qs	1Y	2Ys	3Ys	5Ys	10Ys
iHMM-AR(1) $\rho = 0.95$	2.30	**−427.7**	**−439.3**	**−453.8**	−487.8	−498.7	−489.8	−469.7
iHMM-hyper-AR(1) $\rho \sim \mathbf{B}(9.5,0.5)$	2.48	−431.9	−451.4	−461.2	−486.3	−491.7	−485.2	−457.1
MS(2)-AR(3)	2.29	−437.0	−449.0	−461.5	−480.1	−483.3	−470.2	−433.5
MS(9)-AR(3)	2.32	−430.0	−443.8	−454.9	**−474.6**	**−479.9**	−470.7	−435.2
BMA-MS-AR(2)	2.82	−431.8	−449.4	−460.0	−477.4	−480.8	**−469.3**	**−424.8**
BMA-SB-AR(2)	3.29	−436.2	−454.0	−469.8	−514.0	−531.5	−532.9	−551.2
BMA-MS+SB-AR(2)	2.90	−431.7	−449.3	−460.1	−477.9	−481.0	−469.8	−426.3
AR(3)	**2.25**	−468.7	−480.0	−490.3	−533.6	−564.7	−615.2	−729.4
Rolling AR(2) (5 years)	2.42	−437.5	−453.0	−471.2	−513.1	−526.4	−543.4	−565.0

Note: There are 252 observations from 1947Q1 to 2009Q4 for US quarterly real interest rates. The same training sample is used as that in in Table III. For the h-period-ahead forecast, the number of observations used is $200 - h + 1$.

The last seven columns of Table IV show the log long-run predictive likelihoods at different horizons. I choose the horizon $h = 2, 3, 4, 8, 12, 20, 40$, which represent 2 and 3 quarters, and 1, 2, 3, 5 and 10 years. Since the same training sample as in Table III is used, there are $200 - h + 1$ observations for the h-period-ahead forecast.

For the 2-quarter, the 3-quarter and the 1-year-ahead density forecast, the iHMM performs the best. For the 2-year and the 3-year-ahead density forecast, the large MS(K) model performs better. Finally, for the the 5-year and 10-year-ahead density forecast, the BMA-MS performs the best.

The dominance of the large MS(K) and MS(K) models in the longer horizon may be because the data are likely to return to the existing regimes in the long run. Fixing the number of regimes is able to prevent the data from having too many structural breaks. This could also explain why the BMA-SB works so poorly.

The linear and rolling AR(q) models do not work well at any forecasting horizon. This shows that a nonlinear model is valuable for the long-run density forecast as well.

5.5. Prior Sensitivity

This subsection investigates the sensitivity of the marginal and predictive likelihood to the prior assumption in Section 5.1. I focus on the hierarchical parameters (H, φ, χ, v) in the iHMM-AR(1) with $\rho = 0.95$. One loose and one tight prior are proposed for each of H, φ, χ and v while holding the others' priors unchanged. Finally, a loose and a tight prior for all these parameters are tried.

For H, a more dispersed/shrinking distribution can be obtained by decreasing/increasing a_0 while preserving its prior mean equal to the identity matrix. I use $a_0 = q + 5$ as a loose prior and $a_0 = q + 20$ as a tight prior for H. For φ, I use $\tau_0 = 10$ and $\tau_0 = 0.1$ to represent a loose and a tight prior. For χ, I set $c_0 = 0.5$ and $d_0 = 0.1$ as the loose prior, which preserves the prior mean of χ. A tight prior for χ has $c_0 = 50$ and $d_0 = 10$. For v, because its prior mean and variance both depend on ρ_v, it is impossible to make the prior distribution more dispersed or shrinking without changing the mean. I set $\rho_v = 10$ and $\rho_v = 3$ as a loose and a tight prior; their 90% density intervals are (0.51, 30.0) and (0.15, 9.0), respectively. Lastly, for the whole set of the aforementioned hierarchical parameters, I set the loose/tight prior as the combination of their loose/tight priors. Namely, the loose prior has $a_0 = q + 5, \tau_0 = 10, c_0 = 0.5, d_0 = 0.1, \rho_v = 10$ and the tight prior has $a_0 = q + 20, \tau_0 = 0.1, c_0 = 50, d_0 = 10, \rho_v = 3$.

Table V shows the log Bayes factors and the log predictive Bayes factors of the alternative prior settings against the original one in Section 5.1. The log predictive Bayes factors are computed from the last 200 observations. We can see that the data do not favour any prior setting strongly, since all the absolute values of the log Bayes factors and the log predictive Bayes factors are less than 3.

Table V. Prior sensitivity

	Log BF	Log PBF
Loose H	1.16	1.01
Tight H	−0.16	−0.22
Loose ϕ	−0.80	0.31
Tight ϕ	0.17	0.13
Loose χ	−1.27	−0.79
Tight χ	−0.06	0.24
Loose v	0.55	1.47
Tight v	−0.26	−0.24
Loose (H,ϕ,χ,v)	−2.18	−0.64
Tight (H,ϕ,χ,v)	1.97	1.37

Note: The original model is the iHMM-AR(1) with $\rho = 0.95$. There are 252 observations from 1947Q1 to 2009Q4 for US quarterly real interest rates. The last 200 observations are used to calculate the predictive likelihoods.

5.6. Posterior Analysis

The best model, iHMM-AR(1) with $\rho = 0.95$, is estimated on the full sample. The posterior summary statistics are reported in Table I. It shows that the data are informative for the degree of freedom v and the number of active regimes L. The posterior mean of v is less than half of its prior mean, which is not included in the posterior 90% density interval. For L, its prior mean is only 1.75, but the posterior mean is 6.33. This number is larger than what is usually used in the existing literature.

Figure 1 plots the posterior mean of the intercept $\varphi_{s_t,0}$, the persistence $\varphi_{s_t,1}$, the standard deviation σ_{s_t} and the regime change probabilities for $t = 1, \ldots, T$. The posterior mean of the persistence parameter in the third panel shows a decrease in 1952Q2. After that, the persistence increases step by step until it suddenly drops to almost 0 in early 2000. It quickly bounces back to the highest level after 2005Q2. The largest posterior mean of the persistence parameter is less than 0.5, so the autocorrelation in each regime is not high. The standard deviations are largest at the beginning and the end of the sample. Between early 1970s and late 1980s, the volatilities are lower than that in the aforementioned two periods but higher than that in the rest of the sample period.

The clustering of the regimes can be shown through a $T \times T$ matrix, in which the value of the tth row and τth column is the probability of time t and τ being in the same regime, $p(s_t = s_\tau | Y)$. I use a temperature plot similar to Geweke and Jiang (2011) to visually show this matrix in Figure 5. A large (small) value is represented by a dark (light) colour. The figure is symmetric against the 45° line, on which the values equal to 1 by construction.

The dark squares on the 45° line imply that the US real interest rate has experienced many regimes. Among these regimes, some are unique (structural break) and some are recurrent (regime switching). The most visible recurrent regimes are associated with the beginning (1947Q2–1952Q2) and the end of the sample period (2005Q2–2009Q4). They are represented by the black squares at the four corners of the figure. A typical non-recurrent regime is associated with the period from 1981Q1 to 1985Q1 in the middle of the figure. If we search horizontally or vertically from this square, the colours for the other points are white, which means no other periods are in the same regime as this block. There are also some grey areas such as the intersection of 1958Q2–1972Q3 and 1986Q2–2001Q3. These two periods share some similarities but also have some uncertainties of being grouped as one regime.

Garcia and Perron (1996) found switching points at the beginning of 1973 and the middle of 1981. From the iHMM, the probability of a regime change in 1973Q1 is 0.41, which is consistent with their finding. The probabilities of regime changes for 1980Q4 and 1981Q1 are 0.56 and 0.21, respectively. So the iHMM identifies the second turning point earlier than Garcia and Perron (1996). Meanwhile, Huizinga and Mishkin (1986) identified October 1979 and October 1982 as the turning points.

DOI: 10.1002/jae

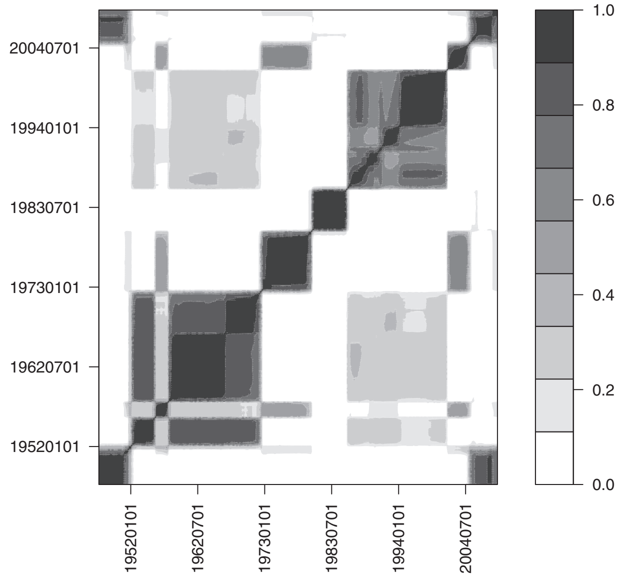

Figure 5. This figure plots the clustering of regimes implied by iHMM-AR(1) with $\rho = 0.95$

The probabilities of regime changes in 1979Q3 and Q4 are 0.01 and 0.03, respectively, while in 1982Q3 and 1982Q4 they are less than 0.001. Thus the iHMM supports Garcia and Perron (1996) against Huizinga and Mishkin (1986).

As an attempt to locate potential regime changing points, I define a time as a candidate turning point if the regime change probability is greater than 0.4. There are seven points in total: 1952Q3, 1956Q2,

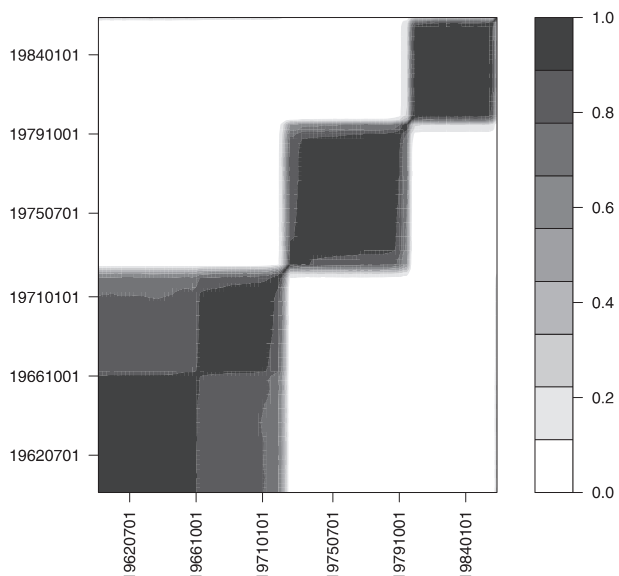

Figure 6. This figure is the truncation of Figure 5. It represents the same sample from 1961Q1 to 1986Q3 used in Wang and Zivot (2000)

DOI: 10.1002/jae

1958Q2, 1973Q1, 1980Q4, 1986Q2 and 2002Q1. 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. The iHMM finds that 1980Q4 and 1986Q2 are close to their finding. However, the iHMM does not identify late 1970 as a turning point.

Another interesting result is shown in Figure 6, which is the truncation of Figure 5 by focusing on the sample period used by Wang and Zivot (2000). The dark blocks are only on the 45° line, which is consistent with a structural break model. Hence Wang and Zivot's (2000) model is a reasonable approach for this sample period.

6. CONCLUSION

This paper proposes to apply an infinite hidden Markov model (iHMM) to integrate current Markov switching and structural break models in a unified Bayesian framework. 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.

The application to US real interest rates shows that the iHMM is robust to model uncertainty and provides superior out-of-sample forecasts to the existing Markov switching and structural break models. The second hierarchical structure is robust to the prior elicitation and able to learn extra information from the data. From both the density forecasts and the posterior analysis, US real interest rates are characterized by both the regime switching and the structural break dynamics.

ACKNOWLEDGEMENTS

I am sincerely thankful to my supervisor, Professor John M. Maheu, for guidance and many helpful comments. I am also grateful to Herman K. van Dijk(coeditor), two anonymous referees, an anonymous commenter, Martin Burda, John Geweke, Christian Gourieroux, Robert Kohn, Thomas McCurdy, James Morley, Rodney Strachan and seminar participants at the Australian National University, Bank of Canada, CEA conference, University of Technology Sydney, University of Melbourne, University of New South Wales, University of Toronto and Wilfrid Laurier University. A previous version is titled *Modelling Regime Switching and Structural Breaks with an Infinite Dimension Markov Switching Model*.

REFERENCES

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

Bassetti F, Casarin R, Leisen F. 2011. Beta-product Poisson–Dirichlet processes. *Arxiv* preprint arXiv:1109.4777.

Celeux G, Hurn M, Robert CP. 2000. Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association* **95**: 957–970.

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

Escobar MD, West M. 1995. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90**: 577–588.

Evans M, Wachtel P. 1993. Inflation regimes and the sources of inflation uncertainty. *Journal of Money, Credit, and Banking* **25**: 475–511.

Fama E. 1975. Short-term interest rates as predictors of inflation. *American Economic Review* **65**: 269–282.

Ferguson TS. 1973. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* **1**: 209–230.

Fox E, Sudderth E, Jordan M, Willsky A. 2011. A sticky HDP-HMM with application to speaker diarization. *Annals of Applied Statistics* **5**: 1020–1056.

Frühwirth-Schnatter S. 2001. Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association* **96**: 194–209.

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

DOI: 10.1002/jae

Geweke J. 2005. Contemporary Bayesian Econometrics and Statistics, Vol. 537. Wiley-Interscience: Hoboken, NJ.
Geweke J. 2007. Interpretation and inference in mixture models: simple MCMC works. *Computational Statistics and Data Analysis* **51**: 3529–3550.
Geweke J, Amisano G. 2010. Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting* **26**: 216–230.
Geweke J, Jiang Y. 2011. Inference and prediction in a multiple structural break model. *Journal of Econometrics* **163**: 172–185.
Griffin JE. 2011. Inference in infinite superpositions of non-Gaussian Ornstein–Uhlenbeck processes using Bayesian nonparametric methods. *Journal of Financial Econometrics* **9**: 519–549.
Griffin J, Steel M. 2011. Stick-breaking autoregressive processes. *Journal of Econometrics* **162**: 383–396.
Hamilton J. 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* **57**: 357–384.
Huizinga J, Mishkin F. 1986. Monetary policy regime shifts and the unusual behavior of real interest rates. *Carnegie-Rochester Conference Series on Public Policy* **24**: 231–274.
Jochmann M. 2010. Modeling us inflation dynamics: a Bayesian nonparametric approach. Working paper, Rimini Centre for Economic Analysis.
Kass RE, Raftery AE. 1995. Bayes factors. *Journal of the American Statistical Association* **90**: 773–795.
Koop G, Potter S. 2007. Estimation and forecasting in models with multiple breaks. *Review of Economic Studies* **74**: 763–789.
Maheu J, Gordon S. 2008. Learning, forecasting and structural breaks. *Journal of Applied Econometrics* **23**: 553–584.
Maheu J, McCurdy T, Song Y. 2012. Components of bull and bear markets: bull corrections and bear rallies. *Journal of Business and Economic Statistics* **30**: 391–403.
Otranto E, Gallo G. 2002. A nonparametric Bayesian approach to detect the number of regimes in Markov switching models. *Econometric Reviews* **21**: 477–496.
Pesaran M, Pettenuzzo D, Timmermann A. 2006. Forecasting time series subject to multiple structural breaks. *Review of Economic Studies* **73**: 1057–1084.
Pesaran M, Pick A, Pranovich M. 2011. Optimal forecasts in the presence of structural breaks. Cambridge Working Papers in Economics.
Rose A. 1988. Is the real interest rate stable? *Journal of Finance* **43**: 1095–1112.
Sethuraman J. 1994. A constructive definition of Dirichlet priors. *Statistica Sinica* **4**: 639–650.
Teh Y, Jordan M, Beal M, Blei D. 2006. Hierarchical Dirichlet processes. *Journal of the American Statistical Association* **101**: 1566–1581.
Van Gael J, Saatci Y, Teh Y, Ghahramani Z. 2008. Beam sampling for the infinite hidden Markov model. In Proceedings of the 25th International Conference on Machine Learning. ACM: New York; 1088–1095.
Walsh C. 1987. Three questions concerning nominal and real interest rates. *Economic Review* **5–19**.
Wang J, Zivot E. 2000. A Bayesian time series model of multiple structural changes in level, trend, and variance. *Journal of Business and Economic Statistics* **18**: 374–386.
West M, Müller P, Escobar M. 1994. Hierarchical priors and mixture models, with application in regression and density estimation. In Aspects of uncertainty: A Tribute to DV Lindley. Wiley: New York; 363–386.

DOI: 10.1002/jae