Identifying Speculative Bubbles with an Infinite Hidden Markov Model *

Shu-Ping Shi
The Australian National University

Yong Song
University of Toronto

October 5, 2011

Abstract

This paper proposes an infinite hidden Markov model (iHMM) to detect, date stamp and estimate speculative bubbles. Three features make this new approach attractive to practitioners. First, the iHMM is capable of capturing the nonlinear dynamics of different types of bubble behaviors as it allows an infinite number of regimes. Second, the implementation of this procedure is straightforward as the detection, dating and estimation of bubbles are done simultaneously in a coherent Bayesian framework. Third, the iHMM by assuming hierarchical structures is parsimonious and superior in out-of-sample forecast. Two empirical applications, one to Argentina money base, exchange rate and consumer price from January 1983 to November 1989, the other to U.S. oil price from April 1983 to December 2010, are presented. We find prominent results, which have not been discovered by the existing finite hidden Markov model. Model comparison shows that the iHMM is strongly supported by the predictive likelihood.

Keywords: speculative bubbles, data generating process, infinite hidden Markov model

JEL classification: C11, C15

*Shuping Shi, School of Finance, Actuarial Studies and Applied Statistics, The Australian National University; Email: shuping.shi@anu.edu.au. Yong Song, Department of Economics, University of Toronto; Email: tommy.song@utoronto.ca.

1 Introduction

Bubbles, which are recognized as germs of economic and financial instability, have drawn considerable attention over the last several decades. Nevertheless, a general agreement on specific data generating processes of bubbles has not yet been reached. Evans (1991), for example, recommends a periodically collapsing explosive process for bubbles. The explosive behavior of bubbles prevails throughout the sample period, however faces a probability of collapsing (to a non-zero value) when it exceeds a certain threshold. If the bubble survives, it expands at a rate faster than the previous stage.[1] In contrast, Phillips et al. (2011b, PWY hereafter) propose a locally explosive bubble process in which the explosive behavior is a temporary phenomenon. Namely, asset prices transit from a unit root regime to a mildly explosive regime when bubbles originate and slide back to the level before origination (with a small perturbation) upon collapsing.[2] Although the aforementioned processes have been frequently utilized as the data generating processes when comparing different econometric tests or dating algorithms for bubbles[3], they have not been justified by any data series containing bubble episodes.

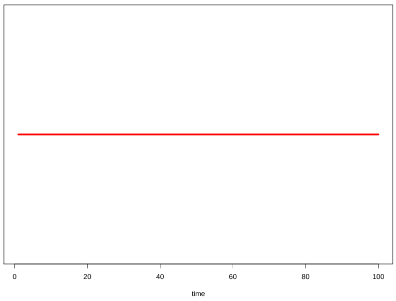
This paper applies an infinite hidden Markov model (iHMM) to reconcile existing data generating processes within a unified and coherent Bayesian framework. This new approach is attractive to practitioners from three perspectives.

First of all, the iHMM is generic to the bubble problem, since current literature assumes that the data dynamics in the presence of bubbles are associated with regime changes. Heuristically, Figure 1b illustrates the periodically collapsing bubble process of Evans (1991) while Figure 1c shows the locally explosive bubble behavior of PWY. One common feature is that they both assume a fixed number of regimes a priori, hence the model uncertainty problem is ignored. One contribution of the iHMM is to provide much more flexibility by allowing an unknown number of regimes, which is treated as a parameter and estimated endogenously. This model

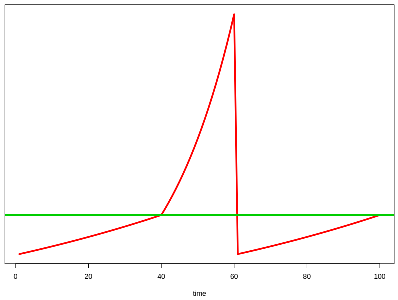
[1] Charemza and Deadman (1995) propose a stochastic explosive bubble process based on the periodically collapsing process of Evans (1991).

[2] Shi et al. (2011) modify the locally explosive process by replacing the one-period bubble collapsing process with a stationary mean-reverting process.

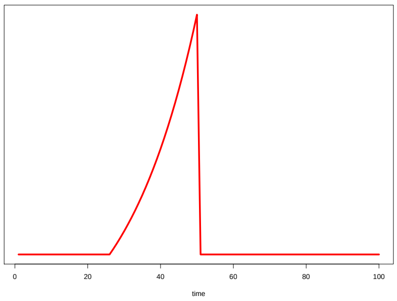
[3] See Evans (1991) and Phillips et al. (2011b), among others.



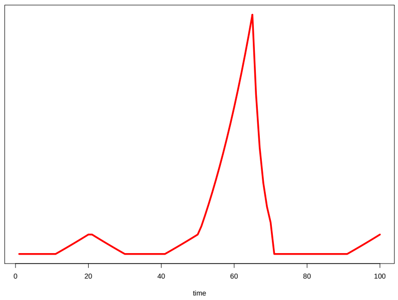
(a) The non-bubble scenario



(b) The periodically collapsing explosive behavior of Evans (1991)



(c) The locally explosive behavior of PWY



(d) The hybrid bubble behavior

Figure 1: Illustration of different data generating processes of bubbles

incorporates linear dynamic (Figure 1a), existing nonlinear dynamics (Figure 1a-1b), and some much richer dynamic with multiple and heterogeneous bubbles (Figure 1d). [4]

The second contribution is that we propose an easy and coherent dating algorithm for bubbles by estimating the iHMM in the Bayesian framework. One of the prevailing approaches for date stamping bubbles[5] is the Markov-switching Augmented Dickey-Fuller (MSADF) test proposed by Hall et al. (1999, HPS hereafter). The MSADF test requires to assume or test for state dimension before estimating the model. However, to the best of our knowledge, the performance of testing procedures for the state dimension of a Markov-switching model which

[4]Phillips et al. (2011a) argues that multiples bubbles is a inherent feature of a long-span economic or financial price series.

[5]Another prevailing approach is the sup type unit root test of Phillips et al. (2011b) and Phillips et al. (2011a).

involves nonstationary (especially explosive) behavior has not yet been investigated. A subjective or an inaccurate selection of the state dimension may cause significant bias in parameter estimation and regime classification. Moreover, the boostrapping procedure embedded in the MSADF test is computationally burdensome as Psaradakis et al. (2001) pointed out and the asymptotic correctness of such bootstrapping procedure has not yet been established and is far from obvious. In contrast to existing frequentists' approaches, the Bayesian methodology allows us to draw inference with a small sample size. The number of regimes and other model parameters are estimated simultaneously using Markov Chain Monte Carlo methods. The dating algorithm is then built on the posterior distributions of the iHMM's parameters. The implementation of this algorithm is much less computational demanding compared with HPS.

Lastly, out approach is less *subjective* than the iHMM of Teh et al. (2006) and Fox et al. (2011) by using two parallel hierarchical structures for the model parameters. Geweke and Jiang (2011) emphasize the importance of the prior elicitation for regime change models. One prominent approach to dealing this problem is using hierarchical structures as Pesaran et al. (2006) among many. Simply speaking, we estimate the prior for the parameters which characterize each regime instead of assuming it as fixed. This methodology produces results robust to the prior choice from an empirical point of view. It is also very convenient from the computational perspective, since regime switching may be practically infeasible with some wild prior. The hierarchical structure will shrink it to a reasonable one, hence facilitates the mixing of the Markov chain.

The first application of the iHMM is to the money base, exchange rate and consumer price in Argentina from January 1983 to November 1989 as in HPS. It is designed to investigate if there exist any new discoveries after we extend the finite hidden Markov model to the infinite dimension. The two-regime Markov switching model of HPS (MS2 thereafter) is estimated in the Bayesian framework as a benchmark. On one hand, some similarities exist between the iHMM and the MS2. On the other hand, we find new prominent features implied by the iHMM.

First, the iHMM and MS2 have the same results for the money base, which are resemble to the locally explosive behavior of PWY. Second, the iHMM implies that the exchange rate's dynamic is similar to to the periodically collapsing process of Evans (1991), while the MS2 finds no sign of bubble collapse. Lastly, the iHMM finds evidence of bubble existence in the consumer price throughout the whole sample period, whereas the MS2 suggests that the explosive behaviors only appear for two short periods starting in June 1985 and July 1989, respectively.

We use the predictive likelihood as the criterion for model comparison. It is based on prediction and acts as the *Ockham's razor* by automatically punishing overparametrization. The results show that the two-regime specification is as good as the iHMM for the money base. However, for the exchange rate and the consumer price, the predictive likelihood strongly supports the iHMM against the MS2. Hence, the results found by the iHMM is more credible from the statistical point of view. We find the explosive money growth in June 1985 did not trigger significant change of dynamics for the exchange rate and consumer price. On the other hand, the explosive growth of the money base in July 1989 is associated with both the exchange rate and the consumer price switching to explosive dynamic regimes.

The second application is to U.S. oil price from April 1983 to December 2010. The oil inventory is used as a proxy to the market fundamental. According to the predictive likelihoods, the iHMM fits the oil price better than the MS2. The iHMM suggests that there exist mild bubbles in the oil price during most of the sample period with four major bubble collapsing periods following the 1985 oil price war, the first Persian Gulf War, the 1998 Asian financial crisis and the subprime mortgage crisis. On the other hand, no explosive dynamic is discovered in the oil inventory data.

The rest of the paper is organized as follows. Section 2 introduces the infinite hidden Markov model. The estimation procedure, along with the dating algorithm of bubbles and the model comparison method, are described in Section 3. The application to the Argentina hyperinflation period and the U.S oil price are in Section 4 and Section 5. Section 6 concludes the paper.

2 Infinite Hidden Markov Model

The infinite hidden Markov model is expressed as

$$y_t \mid s_t = i, \Theta, Y_{1,t-1} \sim f(y_t \mid \theta_i, Y_{1,t-1}), \tag{1}$$

$$\Pr(s_t = i \mid s_{t-1} = j, S_{1,t-2}, P, Y_{1,t-1}) = \Pr(s_t = i \mid s_{t-1} = j, P) = \pi_{ji}, \tag{2}$$

where $i, j = 1, 2, \cdots$. y_t is the data at time t and $Y_{1,t-1} = (y_1, \cdots, y_{t-1})$. s_t is the regime indicator at time t and $S_{1,t-2} = (s_1, \cdots, s_{t-2})$. $\Theta = (\theta_1, \theta_2, \cdots)$ is the collection of parameter θ_i's, which characterize each regime. P is an infinite dimensional transition matrix with π_{ji} on its jth row and ith column.[6] (1) shows that the conditional density of y_t depends on s_t and the past information $Y_{1,t-1}$. (2) implies that the dynamic of s_t only depends on s_{t-1}.

A finite hidden Markov model (HMM) with K regimes, for instance the two-regime Markov-switching model of HPS ($K = 2$), is nested in the iHMM by assuming $\sum_{i=1}^{K} \pi_{ji} = 1$ for $j = 1, \cdots, K$, and the initial regime $s_1 \in \{1, \cdots, K\}$. In addition, the periodically collapsing process of Evans (1991) and the locally explosive process of Phillips et al. (2011b) can be captured by a three-regime HMM.

In order to identify exuberant dynamics, this paper assumes (1) to have a form as of the augmented Dickey-Fuller (ADF) test:

$$f(\Delta y_t \mid \theta_{s_t}, Y_{1,t-1}) \sim \mathbf{N}(\phi_{s_t,0} + \beta_{s_t} y_{t-1} + \phi_{s_t,1} \Delta y_{t-1} + \cdots + \Delta \phi_{s_t,q} y_{t-q},\ \sigma^2_{s_t}), \tag{3}$$

where q is the lag order and $\theta_{s_t} = (\phi'_{s_t}, \sigma_{s_t})$ by construction with $\phi_{s_t} = (\phi_{s_t,0}, \beta_{s_t}, \phi_{s_t,1}, \cdots, \phi_{s_t,q})$. Notice that we model Δy_t instead of y_t. The existence of explosive behaviors is determined by the coefficient of y_{t-1}, namely β_{s_t}. A random walk process implies $\beta_{s_t} = 0$ in an ADF test. In this paper, a positive β_{s_t} shows that y_t is explosive at time t.

The Bayesian approach is applied in estimation to deal with infinite dimensionality. Two parallel hierarchical priors, one governing Θ and the other governing P, are introduced as

[6] By definition, $\pi_{ji} \geq 0$ for all $j, i = 1, 2, \cdots$ and $\sum_{i=1}^{\infty} \pi_{ji} = 1$ for all $j = 1, 2, \cdots$.

follows.

2.1 Prior of Θ

We assume θ_i to have the regular normal gamma distribution $\mathbf{NG}(\phi, H, \frac{\chi}{2}, \frac{\nu}{2})$ (see Geweke (2009)), which is conjugate to linear models. In detail,

$$\sigma_i^{-2} \sim \mathbf{G}\left(\frac{\chi}{2}, \frac{\nu}{2}\right) \text{ and } \quad \phi_i \mid \sigma_i \sim \mathbf{N}\left(\phi, \sigma_i^2 H^{-1}\right). \tag{4}$$

The inverse of the variance σ_i^{-2} is drawn from a gamma distribution with degree of freedom $\frac{\nu}{2}$ and multiplier $\frac{\chi}{2}$. Conditional on σ_i, the vector of regression coefficients, ϕ_i, has a multivariate normal distribution with mean ϕ and covariance matrix $\sigma_i^2 H^{-1}$.[7]

Define $\lambda = (\phi, H, \chi, \nu)$ as the collection of the parameters in the normal gamma distribution. A common practice for a finite hidden Markov model is to assume λ as constant. For the iHMM, however, the number of regimes can grow with the sample size, which allows us to learn λ by using the information across regimes. Hence, we estimate it by giving it the prior as follows:

$$H \sim \mathbf{W}(A_0, a_0), \tag{5}$$

$$\phi \mid H \sim \mathbf{N}(m_0, \tau_0 H^{-1}), \tag{6}$$

$$\chi \sim \mathbf{G}(\frac{d_0}{2}, \frac{c_0}{2}), \tag{7}$$

$$\nu \sim \mathbf{Exp}(\rho_\nu). \tag{8}$$

H is a positive definite matrix and drawn from a Wishart distribution.[8] Conditional on H, ϕ has a multivariate normal distribution with mean m_0 and covariance matrix $\tau_0 H^{-1}$. The prior of χ is a gamma distribution with multiplier $d_0/2$ and degree of freedom of $c_0/2$. The prior of ν is an exponential distribution with parameter ρ_ν.

[7] χ and ν are positive scalars. The mean and variance of σ_i^{-2} are $\frac{\nu}{\chi}$ and $\frac{2\nu}{\chi^2}$, respectively. The mean of σ_i^2 is $\frac{\chi}{(\nu-2)}$. ϕ is a $(q+2)\times 1$ vector and H is a $(q+2)\times(q+2)$ positive definite matrix.

[8] The Wishart distribution is the extension of the gamma distribution to the multivariate setting. Parameters A_0 is a $(q+2)\times(q+2)$ positive definite matrix and a_0 is a positive scalar. A sample from Wishart distribution $W(A_0, a_0)$ is a positive definite matrix. The prior mean of H is $A_0 a_0$. The variance of H_{ij}, the ith row and the jth column element of H, is $a_0(A_{ij}^2 + A_{ii}A_{ij})$, where A_{ij} is the ith row and the jth column element of A_0.

This methodology is less subjective than Teh et al. (2006) and Fox et al. (2011) because the hierarchical structure is robust to prior elicitation. It also facilitates the mixing of the Markov chain by shrinking λ to a reasonable region so that a new regime can be easily born if a structural change is implied by the data.

2.2 Prior of P

The infinite-dimensional transition matrix P is comprised of an infinite number of infinite-dimensional row vector π_j's, where $j = 1, 2, \cdots$. Each $\pi_j = (\pi_{j1}, \pi_{j2}, \cdots)$ represents a probability measure on the natural numbers, namely, $\mathbf{Pr}(A \mid \pi_j) = \sum_{i=1}^{\infty} \pi_{ji} \mathbf{1}_{i \in A}$. By definition, we should have $\pi_{ji} \geq 0$ for each j and i and $\sum_{i=1}^{\infty} \pi_{ji} = 1$ for each j. The prior of P is set as

$$\pi_0 \sim \mathbf{SBP}(\gamma), \tag{9}$$

$$\pi_j \mid \pi_0 \sim \mathbf{DP}\left(c,\ (1-\rho)\pi_0 + \rho\delta_j\right). \tag{10}$$

π_0 is a random probability measure on the natural numbers and drawn from a stick breaking process (SBP).[9] It serves as a hierarchical parameter of all π_j's. Conditional on π_0, each π_j is drawn from a Dirichlet process (DP) with concentration parameter c and shape parameter $(1-\rho)\pi_0 + \rho\delta_j$.[10] The aforementioned constraints on π_j's are automatically satisfied by this prior.

From (10), the shape parameter, $(1-\rho)\pi_0 + \rho\delta_j$, is an infinite discrete distribution and represents the mean of π_j by the definition of DP. It is a convex combination of the hierarchical distribution π_0 and a degenerate distribution at integer j, δ_j[11], with $\rho \in [0, 1]$. The hierarchical

[9] The stick breaking process generates a probability measure over natural numbers. Each number is associated with a non-zero probability. For a probability measure $p \equiv (p_1, p_2, \cdots) \sim \mathbf{SBP}(\gamma)$, where γ is a positive scalar which controls the concentration of a random probability measure, p_i is the probability associated with integer i with $i = 1, 2, \cdots$. Appendix A.2 provides detailed explanation of this process.

[10] A Dirichlet process is a distribution of discrete distributions. It has two parameters: the shape parameter and the concentration parameter. The shape parameter is a probability measure and controls the centre of the random samples, which is analogous to the mean of a distribution. The concentration parameter is a positive scalar and controls the tightness of a random draw, which is analogous to the inverse of the variance of a distribution. Appendix A.1 provides a detailed discussion of this process.

[11] δ_j is a probability measure with $\delta_j(A) = \begin{cases} 1 & \text{if } j \in A \\ 0 & \text{o.w.} \end{cases}$.

distribution π_0 creates a *common* shape for each π_j and δ_j reflects the prior belief of regime persistence. By construction, conditional on π_0 and ρ, the mean of the transition matrix P is a convex combination of two infinite-dimensional matrices, expressed by

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

The above conditional mean of P shows that the self-transition probability is larger as ρ goes closer to 1. In the rest of the paper, ρ is referred to as the sticky coefficient. It is introduced to the iHMM for two reasons. First, empirical evidence shows that regime persistence is a salient feature of many macroeconomic and finance variables. The sticky coefficient explicitly embeds this feature into the prior. Second, a finite hidden Markov model usually has a small number of regimes, which guarantees that each regime can have a reasonable amount of data. The infinite hidden Markov model, however, may assign each data to one distinct regime. This phenomenon is called state saturation, which is obviously not interesting and harmful to forecasting. The sticky coefficient shrinks the over-dispersed regime allocation towards a coherent one and hence avoids the state saturation problem.

In summary, the iHMM is comprised of (1) and (2), in which (1) takes the form of (3) for bubble detection and estimation. (4)-(8) comprise the hierarchical prior for Θ, and (9)-(10) comprise the hierarchical prior for P.

3 Estimation, Dating Algorithm and Model Comparison

3.1 Estimation

The posterior sampling is based on a Markov Chain Monte Carlo (MCMC) method. Fox et al. (2011) show that the block sampler which approximates the iHMM with truncation is more efficient than the individual sampler.[12] We approximate the iHMM by truncation with a finite but large number of regimes. If the number of regimes is large enough, the finite approximation is equivalent to the iHMM in practice.[13]

Suppose L is the maximal number of regimes in the approximation, the model is as follows:

$$\pi_0 \sim \mathbf{Dir}\left(\frac{\gamma}{L}, \cdots, \frac{\gamma}{L}\right), \tag{11}$$

$$\pi_j \mid \pi_0 \sim \mathbf{Dir}\left((1-\rho)c\pi_{01}, ..., (1-\rho)c\pi_{0i} + \rho c, \cdots, (1-\rho)c\pi_{0L}\right), \tag{12}$$

$$s_t \mid s_{t-1} = j \sim \pi_j, \tag{13}$$

$$(\phi, H, \chi, \nu) \sim \mathcal{G}, \tag{14}$$

$$\theta_i \sim \mathbf{NG}(\phi, H, \frac{\chi}{2}, \frac{\nu}{2}), \tag{15}$$

$$\Delta y_t \mid Y_{1,t-1} \sim \mathbf{N}(\phi_{s_t,0} + \beta_{s_t} y_{t-1} + \phi_{s_t,1}\Delta y_{t-1} + \cdots + \Delta\phi_{s_t,q} y_{t-q},\ \sigma^2_{s_t}), \tag{16}$$

where $j = 1, 2, \cdots, L$ and **Dir** represents the Dirichlet distribution. Notice that the only approximation is (11), which is from Ishwaran and Zarepour (2000). They show that $\mathbf{Dir}\left(\frac{\gamma}{L}, \cdots, \frac{\gamma}{L}\right)$ converges to $\mathbf{SBP}(\gamma)$ as $L \to \infty$. (12) is not an approximation, because a DP is equivalent to a Dirichlet distribution if its shape parameter only has support on a finite set. (11) and (12) comprise the prior of P. The dynamic of the regime indicator (13) is the same as (2). (14) and (15) comprise the prior of Θ, which is the same as (4) - (8) in the iHMM. The conditional data density (16) is the same as (1). Different L's are tried to investigate the robustness of the block sampler in applications.

To sample from the posterior distribution, the MCMC method partitions the parameter space into four parts: (S, I), (Θ, P, π_0), (ϕ, H, χ) and ν, where S and I are the collection of regime indicators s_t's and binary auxiliary variables I_t's respectively.[14] Each part is randomly sampled conditional on the other parts and the data $Y = (y_1, \cdots, y_T)$. The sampling algorithms are as follows (see Appendix B for more details):

[12]Consistency proof of the approximation can be found in Ishwaran and Zarepour (2000), Ishwaran and Zarepour (2002). Ishwaran and James (2001) compare the individual sampler with the block sampler and find that the later one is more efficient in terms of mixing.

[13]For example, the approximation is exact if the number of regimes equals to the number of data T.

[14]I_t is an auxiliary variable to sample π_0. Details are in the appendix B.

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

    (a) Sample $S \mid \Theta, P, Y$ by the forward filtering and backward sampling method of Chib (1996).

    (b) Sample $I \mid S$ by a Polya Urn scheme.

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

    (a) Sample $\pi_0 \mid I$ from a Dirichlet distribution.

    (b) Sample $P \mid \pi_0, S$ from Dirichlet distributions.

    (c) Sample $\Theta \mid S, Y$ by the regular linear model results.

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 $\chi \mid \nu, S, \Theta$ from a gamma distribution.

4. Sample $\nu \mid \chi, S, \Theta$ by a Metropolis-Hastings algorithm.

After initiating the parameter values, the algorithm is applied iteratively to obtain a large number of samples of the model parameters. We discard the first block of the samples to remove dependence on the initial values. The rest of the samples, $\{S^{(i)}, \Theta^{(i)}, P^{(i)}, \pi_0^{(i)}, \phi^{(i)}, H^{(i)}, \chi^{(i)}, \nu^{(i)}\}_{i=1}^{N}$, are used for inferences as if they were drawn from their posterior distributions. Simulation consistent posterior statistics are computed as sample averages. For example, the posterior mean of ϕ is calculated by $\frac{1}{N}\sum_{i=1}^{N}\phi^{(i)}$. In order to avoid the label-switching problem in mixture models[15], we use label-invariant statistics suggested by Geweke (2007) so that the posterior sampling algorithm can be implemented without modification.

[15]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 results in the same likelihood.

3.2 Dating Algorithm of Bubbles

The explosive behavior of bubbles depends on the estimates of (16). The autoregressive coefficient of y_{t-1} (i.e. β_{s_t}) should be positive in the presence of explosive bubbles and negative when the data is locally stationary mean-reverting.

The Bayesian approach produces the whole posterior distribution of β_{s_t} instead of a point estimate, so one can make decisions based on his/her specific loss function. This paper considers two intuitive posterior statistics, which are derived from two simple loss functions, to help identify bubbles. One is the posterior probability $P\left(\beta_{s_t} > 0 \mid Y\right)$ and the other is the posterior mean $\mathbb{E}\left(\beta_{s_t} \mid Y\right)$.

For the first statistic, we claim that a bubble exists at time t if the posterior probability of $\beta_{s_t} > 0$ is above 0.5 and there is no bubble otherwise. More concisely,

$$\text{bubble exists in period } t \quad \text{if } P\left(\beta_{s_t} > 0 \mid Y\right) > 0.5 \ \cdot$$

It is easy to see that this criterion is derived from an absolute value loss function. The cutoff value can be different from 0.5 if the loss function is asymmetric.

The second statistic is based on a quadratic loss function. We claim that a bubble exists if the posterior mean of β_{s_t} is above zero and there is no bubble otherwise. More concisely,

$$\text{bubble exists in period } t \quad \text{if } \mathbb{E}\left(\beta_{s_t} \mid Y\right) > 0 \ \cdot$$

This statistic also shows the magnitude of explosiveness. The higher the value is, the faster the bubble expanses.

3.3 Model Comparison

We use the predictive likelihood for model comparison as suggested by Geweke and Amisano (2010). Conditional on an initial data set $Y_{1,t}$, the predictive likelihood of $Y_{t+1,T} = (y_{t+1}, \cdots, y_T)$

by model M_i is calculated as

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

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

The one-period predictive likelihood of a model M_i is calculated by

$$\hat{p}(y_t \mid Y_{1,t-1}, M_i) = \frac{1}{N}\sum_{i=1}^{N} f(y_t \mid \Upsilon^{(i)}, Y_{1,t-1}, M_i), \tag{17}$$

where $\Upsilon^{(i)}$ is one sample of the parameters from the posterior distribution conditional on the past data $Y_{1,t-1}$. For the iHMM, the one-period predictive likelihood is

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

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

The differenced log predictive likelihood, namely $\log(BF_{ij} \mid Y_{1,t}) = \log(Y_{t+1,T} \mid Y_{1,t}, M_i) - \log(Y_{t+1,T} \mid Y_{1,t}, M_j)$, is used for model comparison. Kass and Raftery (1995) suggest comparing the difference between the log marginal likelihoods. Geweke and Amisano (2010) show that the interpretation for the predictive likelihood is the same as the marginal likelihood if we regard the initial data set $Y_{1,t}$ as a training sample. Therefore, we use the following criteria from Kass and Raftery (1995) as evidence of model M_i against M_j.

$$\begin{array}{cc}
\text{Not worth more than a bare mention,} & \text{if } 0 \leq \log(BF_{ij}) < 1; \\
\text{Positive,} & \text{if } 1 \leq \log(BF_{ij}) < 3; \\
\text{Strong,} & \text{if } 3 \leq \log(BF_{ij}) < 5; \\
\text{Very strong,} & \text{if } \log(BF_{ij}) \geq 5.
\end{array}$$

4 Empirical Application: Argentina Hyperinflation

In this section, we apply the iHMM approach to the money base, exchange rate and consumer price in Argentina from January 1983 to November 1989. The money base is used as a proxy for market fundamental and the exchange rate data series is to capture fundamentally determined bubble-like behavior. The purpose is to investigate whether there is evidence of bubble behaviors in the consumer price.

These three data series are also examined in HPS and Shi (2010). Both HPS and Shi (2010) conduct a two-regime Markov-switching ADF (MSADF) test (with different specifications in the error variance) on these three data series and conclude no evidence of bubbles in the consumer price. The two-regime Markov-switching (MS2) models of HPS and Shi (2010) are both estimated by MLE.

As a benchmark, we estimate a MS2 model using the Bayesian approach, which is[16]

$$\Delta y_t \mid s_t = j, Y_{1,t-1} \sim \mathbf{N}(\phi_{j0} + \beta_j y_{t-1} + \phi_{j1}\Delta y_{t-1} + \cdots + \phi_{j4}\Delta y_{t-4}, \sigma_j^2) \tag{18}$$

$$\Pr(s_t = j \mid s_{t-1} = j) = p_{jj} \tag{19}$$

with $j = 1, 2$. The prior of self-transition probabilities p_{11} and p_{22} is $\mathbf{Beta}(9, 1)$ and the prior of (ϕ_j, σ_j) is a normal-gamma distribution, namely $\sigma_j^{-2} \sim \mathbf{G}(1, 1)$ and $\phi_j \mid \sigma_j \sim \mathbf{N}(\mathbf{0}, \sigma^2\mathbf{I})$. The prior mean and variance of precision σ_j^{-2} are unity. The infinite hidden Markov model, (11)-(16) is estimated by setting $L = 5$ and with priors $\phi, H \sim \mathbf{NW}(0, 1, 0.2\mathbf{I}, 5)$, $\chi \sim \mathbf{G}(1, 1)$ and $\nu \sim \mathbf{Exp}(1)$.[17] We set this prior in order to make the prior parameters of the MS2 model equal to the mean of the hierarchical prior of the iHMM.

Figure 2 illustrates the posterior probabilities of $\beta_{s_t} > 0$ (i.e $P(\beta_{s_t} > 0|Y)$) for the logarithmic money base, exchange rate and consumer price. From the MS2 model (dotted line), we can see that the posterior probability exceeds the 0.5 in June 1985 and July 1989 for all three data series, which suggests the existence of explosive behaviors. Meanwhile, since the spikes appear simultaneous in these two periods, the explosive behavior of market fundamentals

[16]The lag order is the same as that in HPS and Shi (2010).
[17]Larger Ls produce the same results in bubble detection.

(money growth) is consistent with the explosive dynamics of the exchange rate and consumer price. However, we also find bubbles in the exchange rate emerge in April 1987, October 1987 and September 1988, which have some locally explosive dynamics that can not be explained by the money base. This is further supported by the posterior mean of β_{s_t} (i.e. $\mathbb{E}(\beta_{s_t}|Y)$) displayed in Figure 3.

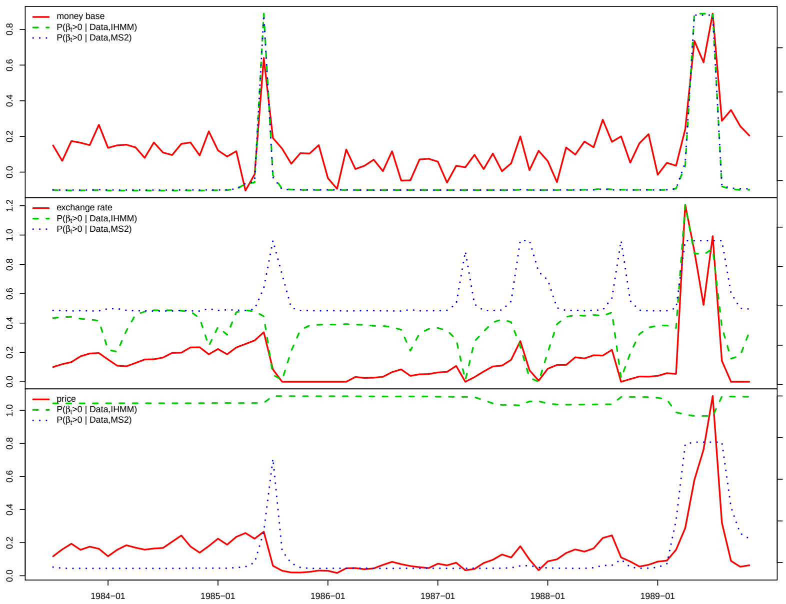


Figure 2: The growth rate and the posterior probabilities of $\beta_{s_t} > 0$ for the money base, exchange rate and consumer price

The posterior probability and posterior mean of the money base from the iHMM (dashed line) are identical to those from the MS2 model. Therefore, it is reasonable to have a two-regime specification for the money base. On the other hand, the iHMM shows distinct probability and mean patterns for the exchange rate and consumer price.

For the exchange rate, the posterior probability of $\beta_{s_t} > 0$ implied by the iHMM in Figure 2 is almost a mirror image of that by the MS2 model except in 1989. Any plunge of $P(\beta_{s_t} > 0 \mid Y)$ of the iHMM is associated with a decrease of the exchange rate, which is consistent with Evans's (1991) bubble collapse regime. Ironically, for the MS2 model, 4 out of 5 spikes in Figure 2 correspond to decreasing of the exchange rate, which is counter intuitive. This is caused by the

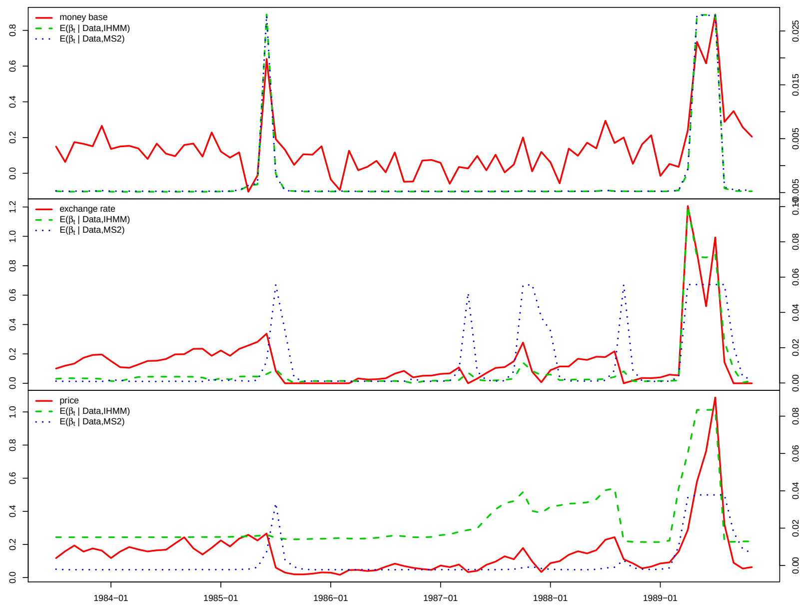


Figure 3: The growth rate and posterior mean of β_{s_t} for the money base, exchange rate and consumer price

limited number of regimes in the MS2 model. The iHMM in Figure 2 and 3 shows that there are approximately 3 regimes: one is associated with the bubble emerging at July 1989; one is related to the exchange rate decreasing periods starting at June 1985, April 1987, October 1987 and September 1988; and one for the rest of the sample. The MS2 model clearly combines the first two regimes into one and produces confusing results.

For the consumer price, the iHMM implies that bubbles exist throughout the whole sample period in the bottom panel of Figure 2. The bottom panel of Figure 3 shows the dynamics of the consumer price is consistent with Evans's (1991) assumption with time-varying degrees of explosiveness. The expansion rate is relatively higher in the one and a half years spanning from June 1987 to August 1988, and a considerable increase occurred over the period from March 1989 to July 1989. The posterior mean of β_{s_t} at the peak of this episode is 0.08. The rate then dropped rapidly so that by August 1989 it was 0.01 (same as the rate before the increase occurred). On the other hand, the MS2 is unable to capture the dynamics because of assuming a limited number of regimes. Although it identifies two bubble periods staring at

June 1985 and July 1989, we can visually find that they are falsely grouped together since the first of these two episodes is associated with price decrease while the last one corresponds to price increase from Figure 3.

In summary, according to the iHMM, the dynamic pattern of the money base appears like the locally explosive explosive of PWY, where the explosive bubble behaviors are transitory. More specifically, a data series switches between a non-explosive regime and an explosive regime. In contrast, the dynamics of the exchange rate and the consumer price are similar to the periodically collapsing behavior of Evans (1991), where bubbles exist throughout the sample period. The iHMM also implies that the switching of the money supply's dynamic to an explosive regime at July 1989 was closely related to the change of dynamics for the exchange rate and the consumer price (switching to a regime with a faster expansion rate). On the other hand, the explosive behavior of the money base in June 1985 did not cause regime switching in either the exchange rate or the consumer price.

Table 1: Log predictive likelihoods

	MS2	iHMM
Money Base	34.1	34.4
Exchange Rate	34.9	43.5
Consumer Price	54.4	67.0

The last 50 observations (out of 82) are used to calculate the predictive likelihood.

For formal model comparison, Table 1 reports the log predictive likelihoods. For the money base, the predictive likelihoods of the MS2 model and the iHMM are very close, which are consistent with the conclusion from Figure 2 and 3. The log predictive likelihoods of the exchange rate and the consumer price from the iHMM are much higher than those from the MS2 model. The differences in the log predictive likelihoods for the exchange rate and the consumer price are $43.5 - 34.9 = 8.6$ and $67.0 - 54.4 = 12.6$, respectively. Therefore, we strongly reject the MS2 model for the exchange rate and the consumer price by Kass and

Raftery's (1995) criterion.

5 Empirical Application: U.S Oil Price

The second application investigates the existence of speculative bubble behaviors in the U.S oil price. Several papers have studied the evidence for bubbles in the oil price (among others, Phillips and Yu (2011), Sornette et al. (2009) and Shi and Arora (2011)). Although the sample periods and methodologies used in these papers are different, most studies have found evidence of bubble existence.

Our data is sampled from April 1983 to December 2010. The price of the nearest-month West Texas Intermediate futures contract, obtained from DataStream International, is used as a proxy for the spot oil price. We deflate the oil price by U.S. Consumer Price Index obtained from the U.S Bureau of Labour Statistics. The market fundamental is proxied by the oil inventory, which is the ending stocks excluding Strategic Petroleum Reserve of crude oil and petroleum products (thousand barrels) and comes from the U.S. Energy Information Administration.

We apply both the MS2 model and the iHMM to the logarithmic real oil price and the logarithmic oil inventory.[18] The priors are assumed to be the same as in Section 4. The posterior probabilities of $\beta_{s_t} > 0$ and the posterior mean of β_{s_t} are presented in Figure 4 and 5, respectively.

For the oil inventory, both models suggest that there is no evidence of explosive behavior. Namely, the posterior probabilities (posterior means) of the inventory from both models are below 0.5 (zero). In fact, the posterior mean of β_{s_t} remains the same throughout the sample period.[19] This supports that regime switching is not a feature of the oil inventory dynamic in the sample period.

For the oil price , the patterns of the posterior probability $P\left(\beta_{s_t} > 0|Y\right)$ and the posterior

[18]For the oil inventory, we have estimated three versions of the data: the raw data, the data that is scaled by 10^{-5} and the log of the raw data. All results are the same for bubble detection and regime identification. So we only report the results based on the log of the raw data.

[19]Notice that the posterior mean of β_{s_t} from the iHMM is smaller than that from the MS2 model. This is because the iHMM uses hierarchical structures but the two-regime MS model does not.

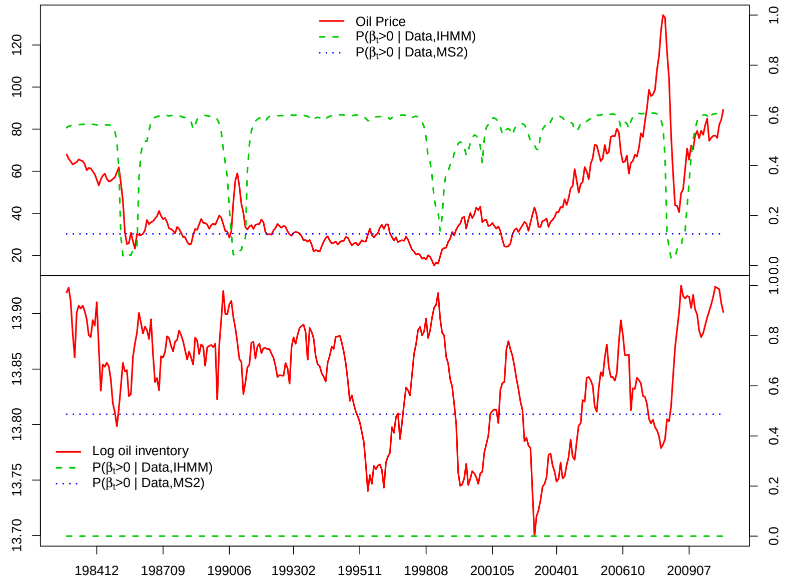


Figure 4: Posterior probabilities of $\beta_{s_t} > 0$ for the logarithmic oil price and inventory

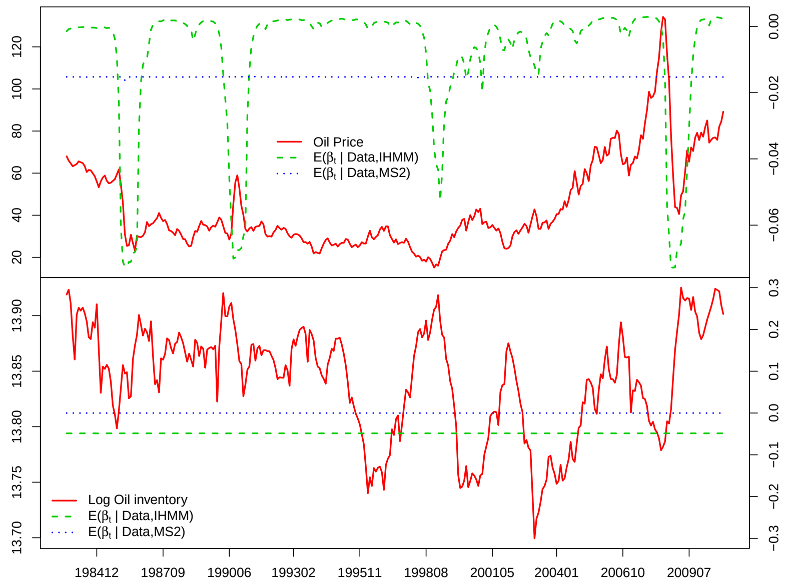


Figure 5: Posterior mean of β_{s_t} for the logarithmic oil price and inventory

mean $\mathbb{E}(\beta_{s_t}|Y)$ from the MS2 model are different from the iHMM. For the MS2 model, the posterior probabilities of $\beta_{s_t} > 0$ are always smaller than 0.5 and the posterior means of β_{s_t} are negative. Therefore, no evidence of bubble exists based on the MS2 model. However, the iHMM implies that the oil price dynamic is comprised of mild explosive behavior and bubble collapse phases, which is a prominent feature and different from Evans (1991) and PWY. For most of the sample period, the posterior mean of β_{s_t} is slightly positive. There are significant falls (below zero) in periods following the 1985 oil price war (1985M11-1986M09), the first Persian Gulf War (1990M04-1991M03), the 1998 Asian financial crisis (1998M12-1999M08) and the subprime mortgage crisis (2008M07-2009M07).

The formal model comparison is in Table 2. It reports the log predictive likelihoods for the oil price and the oil inventory. We can see that the iHMM outperforms the two-regime MS model for both data series. The differences of the log predictive likelihoods are 72 and 488 for the log price and the log inventory respectively. The results strongly support the iHMM against the two-regime switching model.

Table 2: Log predictive likelihoods

	2-regime MS model	iHMM
Oil price	249	321
Oil inventory	-2397	-1909

The last 300 observations (out of 333) are use to calculate the predictive likelihood.

6 Conclusions

This paper proposes an infinite hidden Markov model to integrate the detection, date-stamping and estimation of bubble behaviors in a coherent Bayesian framework. It reconciles the existing data generating processes of speculative bubbles and is able to reveal the dynamic patterns of real data series. Two parallel hierarchical structures provide a parsimonious methodology for prior elicitation and improve out-of-sample forecasts.

The iHMM is applied to Argentina money base, exchange rate and consumer price from January 1983 to November 1989 and U.S. oil price and oil inventory from April 1983 to December 2010. The predictive likelihoods strongly support the iHMM against a finite Markov switching model.

The dynamic of the Argentina money base is similar to the locally explosive behavior of PWY, where the explosive behavior is a transitory phenomenon. The Argentina exchange rate and consumer price, on the other hand, resemble the periodically collapsing explosive behavior of Evans (1991), where the explosive behavior prevails throughout the sample period and the rate of explosiveness is time-varying. Furthermore, we discover that the expansion of the money supply in July 1989 is closely related to the simultaneous regime changes of the exchange rate and the consumer price (to a regime with a faster expansion rate).

For the U.S. oil price, we find that mild explosive behavior exists for most of the time except 4 major bubble collapsing periods following the 1985 oil price war, the first Persian Gulf War, the 1998 Asian financial crisis and the subprime mortgage crisis. This feature is different from existing data generating processes as Evans (1991) and PWY. On the other hand, regime change is not a feature of U.S. oil inventory data.

References

J. Berger. The case for objective bayesian analysis. *Bayesian Analysis*, 1(3):385–402, 2006.

Wojciech W. Charemza and Derek F. Deadman. Speculative bubbles with stochastic explosive roots: The failure of unit root testing. *Journal of Empirical Finance*, 2:153–163, 1995.

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

George W. Evans. Pitfalls in testing for explosive bubbles in asset prices. *The American Economic Review*, 81(4):922–930, Sep. 1991.

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

E.B. Fox, E.B. Sudderth, M.I. Jordan, and A.S. Willsky. A Sticky HDP-HMM with Application to Speaker Diarization. *Annals of Applied Statistics*, 5(2A):1020–1056, 2011.

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

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

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

J. Geweke and Y. Jiang. Inference and prediction in a multiple structural break model. *Journal of Econometrics*, 2011.

Stephen G. Hall, Zacharias Psaradakis, and Martin Sola. Detecting periodically collapsing bubbles: A markov-switching unit root test. *Journal of Applied Econometrics*, 14:143–154, 1999.

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

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

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

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

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

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

Peter C. B. Phillips and Jun Yu. Dating the timeline of financial bubbles during the subprime crisis. *Quantitative Economics, forthcoming*, 2011.

Peter C. B. Phillips, Shu-Ping Shi, and Jun Yu. Testing for multiple bubbles. *Singapore Management University, Center for Financial Econometrics Working Paper 2010-03*, 2011a.

Peter C. B. Phillips, Yangru Wu, and Jun Yu. Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values? *International Economic Review*, 52:201–226, 2011b.

Zacharias Psaradakis, Martin Sola, and Fabio Spagnolo. A simple procedure for detecting periodically collapsing rationa bubbles. *Economics Letters*, 24(72):317–323, 2001.

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

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

Shu-Ping Shi. Bubbles or volatility: A markov-switching adf test with regime-varying error variance. *Australian National University, School of Economics Working Paper 2010-524*, 2010.

Shu-Ping Shi and Vipin Arora. An application of models of speculative behaviour to oil prices. *Australian National University, CAMA Working paper 2011-11*, 2011.

Shu-Ping Shi, Peter C. B. Phillips, and Jun Yu. Specification sensitivities in right-tailed unit root testing for financial bubbles. *Hong Kong Institute for Monetary Research, Working Paper 17*, 2011.

D. Sornette, R. Woodard, and W. Zhou. The 2006-2008 oil bubble and beyond. *Physica A*, 388:1571–1576, 2009.

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

A Dirichlet Process and Stick Breaking Process

A.1 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-dimensional vector with $x_i \in (0,1)$ and $\sum\limits_{i=1}^{K} x_i = 1$. The probability density function is

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

A special case is the **Beta distribution**, where $K = 2$. $\blacksquare$

Define $\alpha_0 = \sum\limits_{i=1}^{K} \alpha_i$ and X_i as the ith element of the random vector X from a Dirichlet distribution $\mathbf{Dir}(\alpha)$. The random variable X_i 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 multi-nominal distribution in the following sense: if

$$X \sim \mathbf{Dir}(\alpha),$$

$$\beta = (n_1, \ldots, n_K) \mid X \sim \mathbf{Mult}(X),$$

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 finite dimensions to infinite dimensions. 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 interpretation 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 form 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 center of a random distribution F drawn from a Dirichlet process $\mathbf{DP}(\alpha_0, G_0)$. 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 $P(a_i \in A \mid G_0) = \mathbb{E}(P(a_i \in A \mid F) \mid G_0) = \mathbb{E}(F(A) \mid G_0) = G_0(A)$. 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 multi-nominal distribution. By the conjugacy of Dirichlet distribution to the multi-nominal 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 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. Ferguson (1973) showed that a random distribution drawn from a Dirichlet process is almost sure discrete, although the shape parameter G_0 can be continuous.

A.2 Stick breaking process

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 θ_i's and their corresponding probabilities p_i's, 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

$$V_i \overset{iid}{\sim} \mathbf{Beta}(1, \alpha_0) \tag{20}$$

$$p_i = V_i \prod_{j=1}^{i-1} (1 - V_j) \tag{21}$$

$$\theta_i \overset{iid}{\sim} G_0 \tag{22}$$

where $i = 1, 2, \cdots$. In this representation, p and θ are generated independently. The process generating p, (20) and (21), is called the stick breaking process and denoted by $p \sim \mathbf{SBP}(\alpha_0)$. 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.

B Block sampler

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

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

I is introduced to facilitate the π_0 sampling. From (11) and (12), the filtered distribution of π_j conditional on $S_{1,t}$ and π_0 is a Dirichlet distribution:

$$\pi_j \mid S_{1,t}, \pi_0 \sim \mathbf{Dir}\left(c(1-\rho)\pi_{01} + n_{j1}^{(t)}, \cdots, c(1-\rho)\pi_{0j} + c\rho + n_{jj}^{(t)}, \cdots, c(1-\rho)\pi_{0L} + n_{jL}^{(t)}\right)$$

where $n_{ji}^{(t)}$ is the number of $\{\tau \mid s_\tau = i, s_{\tau-1} = j, \tau \le t\}$. Integrating out π_j, the conditional distribution of s_{t+1} given $S_{1,t}$ and π_0 is:

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

Construct a variable I_t with a Bernoulli distribution

$$p(I_{t+1} \mid s_t = j, S_{1,t}, \pi_0) \propto \begin{cases} c\rho + \sum\limits_{j=1}^{L} n_{ji}^{(t)} & \text{if } I_{t+1} = 0 \\ c(1-\rho) & \text{if } I_{t+1} = 1 \end{cases}$$

and the conditional distribution:

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

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

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

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

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

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

This approach of sampling π_0 is simpler than Fox et al. (2011).

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

$$\pi_j \mid \pi_0, S \sim \mathbf{Dir}(c(1-\rho)\pi_{01} + n_{j1}, \cdots, c(1-\rho)\pi_{0j} + c\rho + n_{jj}, \cdots, c(1-\rho)\pi_{0L} + n_{jL})$$

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

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

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

By conjugacy, the posterior is:

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

with

$$\begin{aligned}
\overline{\phi}_i &= \overline{H}_i^{-1}(H\phi + X_i'Y_i) \\
\overline{H}_i &= H + X_i'X_i \\
\overline{\chi}_i &= \chi + Y_i'Y_i + \phi' H \phi - \overline{\phi}' \overline{H\phi} \\
\overline{\nu}_i &= \nu + n_i
\end{aligned}$$

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

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

The conditional posterior is:

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

where K is the number of active regimes, with which at least one data point is associated. ϕ_i and σ_i are the parameters of regime i.

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

The conditional posterior of χ is:

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

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

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

The conditional posterior of ν has no regular density form:

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

The Metropolis-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).