Dipartimento di Statistica
"Giuseppe Parenti"

Dipartimento di Statistica "G. Parenti" – Viale Morgagni 59 – 50134 Firenze - www.ds.unifi.it

WORKING PAPER 2001/04

A Nonparametric Bayesian Approach to Detect the Number of Regimes in Markov Switching Models

E. Otranto - G.M. Gallo

Università degli Studi di Firenze

Econometrics

A NONPARAMETRIC BAYESIAN APPROACH TO DETECT THE NUMBER OF REGIMES IN MARKOV SWITCHING MODELS

Edoardo Otranto
Direzione Centrale delle Statistiche Congiunturali
ISTAT, Rome
otranto@istat.it
Giampiero M. Gallo*
Dipartimento di Statistica "G.Parenti"
Università di Firenze
gallog@ds.unifi.it

This Revision October 2001

Key Words: Markov switching models; nuisance parameters; specification testing; exchange rate determination.

JEL Classification: C2, C5, F3

*Corresponding author: Dipartimento di Statistica "G.Parenti" - Viale G.B. Morgagni, 59 - 50134 Firenze. We thank D.Böhning for some useful comments on an earlier version of the paper and to ESEM 99 participants (Santiago de Compostela). The paper greatly benefitted from the suggestions of two anonymous referees and the Associate Editor. Thanks are also due to Jim Hamilton for providing the data used in Engel and Hamilton (1990) and to Charles Engel for the data used in his 1994 article. Financial support from Italian MURST (grant "ex40%-1998") is gratefully acknowledged.

ABSTRACT

The literature on Markov switching models is increasing and producing interesting results both at theoretical and applied levels. Most often the number of regimes, i.e., of data generating processes, is considered known; this strong hypothesis is adopted to somewhat bypass the nuisance parameter problem which affects hypothesis testing for the number of regimes. In this paper we take the view that some results derived from a nonparametric Bayesian approach provide a convenient way to deal with the issue of detecting the number of components in the mixture density, based on the assumption that the parameter distributions are generated by a Dirichlet process. The advantage is that we need no testing (in a classical sense) for the number of regimes, and the approach is not affected by a change point at the beginning or at the end of the sample. A Monte Carlo experiment provides some insights into the performance of the procedure. The potentiality of the approach is illustrated in reference with some well known results on exchange rate modelling.

I Introduction

Some of the recent literature on structural change modeling is related to Markov switching (MS) models applied to economic phenomena (to mention a few, Hamilton, 1989, for the US GNP dynamics; Engel and Hamilton, 1990, (EH) and Engel, 1994, for segmented trends in the US dollar exchange rates;[1] Garcia and Schaller, 1995, for monetary effects on output). These models share the characteristics of assuming known the number of regimes (states), that is, the number of generating data processes which differ from one another just for the value of the parameters. Working with a known number of regimes avoids the problem of selecting it, since it would require hypothesis testing with nuisance parameters identified only under the alternative (Andrews and Ploberger, 1994). This is a common occurrence in many nonlinear models (the threshold parameter in threshold models, the unknown time of a change in regime, the number of regimes in MS models; see Hansen, 1992) and the main consequence is that the regularity conditions required to apply the asymptotic theory are no longer respected. In fact, the likelihood function under the null is non-quadratic and flat with respect to the nuisance parameters at the optimum and the score is identically null when the parameters under the null hypothesis correspond to a maximum, minimum or saddle point for the likelihood function. This latter case is typical for switching models with non-observable regimes (see Hansen, 1992; Andrews and Ploberger, 1994; Garcia, 1995).

There are different approaches to compute the asymptotic distribution of the classical test statistics in this framework; the seminal work is Davies (1977), which Andrews (1993) extended in a dependent data context deriving the expression for the asymptotic distribution of the supremum of a likelihood ratio test statistics relative to a range of nuisance parameters. Hansen (1992) and Garcia (1995) have proposed a similar approach for the specific case of the number of regimes in the context of MS models. However, the first is computationally cumbersome and the second reduces the range of nuisance parameters, possibly excluding from the analysis some important special cases. As an additional difficulty, the asymptotic distribution of both tests changes with the specific model adopted.

As an alternative to testing, one may fix the number of regimes to a safely large number (cf. Franq and Roussignol, 1997).[2] When the number of regimes is overparameterized, for example three estimated regimes instead of two regimes in the data generating process, maximum likelihood estimation provides the same estimated parameters for the second and third regime, so that an overparameterized estimated model does not necessarily perform worse than a correctly specified model.

Yet another alternative, which we investigate in this paper, can be seen as one where

[1] Some results are reported in Engel (1992) and are not present in the 1994 version

[2] We thank an anonymous referee for pointing this out to us.

the number of components of a mixture density has to be determined. As a matter of fact, the density of a MS model is a mixture density with the number of components equal to the number of regimes in the data generating process. In classical terms, the problem is generally bypassed using the likelihood ratio test statistic, obtaining its distribution by simulation techniques; in particular, Feng and McCulloch (1996) provide the mathematical background to justify the use of bootstrap likelihood ratios. Böhning et al. (1994) provide the exact distribution of a likelihood ratio test for mixtures of densities from the one-parametric exponential family. Recently, Böhning et al. (1992) and Böhning et al. (1998) include in their nonparametric maximum likelihood approach a criterion to establish the number of components of a mixture. Briefly put, this last criterion is based on the estimation of a large number of components, down-grouping those that differ by less than a certain threshold.

In a Bayesian context there are different suggestions for choosing the number of components in the mixture. For example, Roeder and Wasserman (1997) propose a Schwarz criterion; Raftery (1996) notes that the natural Bayesian solution would be based on a Bayes factor, but it cannot be used to compare more than two models, so he suggests a Laplace-Metropolis estimator; Richardson and Green (1997) use a fully Bayesian analysis, based on the reversible jump Markov chain Monte Carlo methods, developed in Green (1995), in which the algorithm allows for the change of the dimension of parameter space, changing the number of mixtures from one iteration to the next. Finally, many authors, for example Escobar (1994), West et al. (1994), Escobar and West (1995), adopt a Bayesian nonparametric approach, based on Dirichlet processes.

We think that using a Bayesian approach is more convenient because the probability distribution for the number of components can be easily established leaving the estimated number of regimes to be the mode of that distribution. In this paper, we use the Bayesian nonparametric approach of Escobar and West (1995) to identify the number of regimes in switching models.[3] There are some aspects that make this approach appealing:

1. we do not need tests for the number of regimes k;

2. the analysis does not need strong hypotheses; they are the same used in switching

[3] This issue should be kept separate from the estimation of the switching models in a Bayesian context; for example, Carter and Kohn (1994, 1996) and Shephard (1994) propose Monte Carlo Markov Chain methods to estimate a general model which encompasses the switching model; Albert and Chib (1993) and McCulloch and Tsay (1994) use the Gibbs sampling to estimate the MS model; Hamilton himself (1989) uses a *quasi-Bayesian* approach to bypass singularity problems in estimation. But in this context the number of regimes is considered fixed a priori.

Bayesian models with the additional hypothesis that the distributions of parameters are generated from a Dirichlet process;

3. the usual tests are not robust to the presence of a structural change-point placed at the beginning or at the end of the time series while this approach does not suffer from this limitation.

Relative to other contributions, ours is empirically motivated and aims at extending some methods developed in the context of mixtures to a dynamic framework which is relevant for econometric applications. In this respect, we want to illustrate the features of the suggested procedure when it supplements the traditional econometric specification of Markov switching models. Rather than working on a novel dataset and models we use well–known studies about exchange rate dynamics (Engel and Hamilton, 1990; Engel, 1994). The structure of the paper is as follows: in Section 2 we establish some notation and we set up the outline of the procedure (the details of which are placed in two appendices). Section 3 contains a Monte Carlo experiment to assess the properties of the approach when the data are generated with a change in regime and when they are not. Finally, in Section 4, we apply the procedure to the data used by EH and by Engel with MS models. Some final remarks follow.

II The Detection of the Number of Regimes

Let us introduce the problem at hand starting from a well–known application of the MS model in the econometric literature suggested by EH: in order to accommodate the persistence around positive and negative trends (especially during the 1980s) for the US dollar exchange rates relative to three major currencies (Deutsche Mark, French Franc and British Pound), EH refer to a MS model with two regimes (with quarterly data; Engel, 1994, works with monthly data in the same context with less clear–cut results about the presence of a segmented trend). Given the good approximation provided by the random walk model to the behavior of the exchange rate, the issue is to characterize whether the exchange rate returns have different means (positive for appreciation, negative for depreciation) and variances for empirically recognizable periods. Therefore, they assume that the return data y_t are generated from two Normal distributions, the first with mean μ_1 and variance σ_1^2 and the second with mean μ_2 and variance σ_2^2, and test the null of a random walk model against the alternative of a MS model, expressed as:

$$y_t = \mu_{s_t} + \varepsilon_t, \quad \varepsilon_t \sim IIN\left(0, \sigma_{s_t}^2\right), \quad t = 1, \ldots, T \tag{1}$$

where the regime is indicated by the discrete random variable $s_t \in \{1, 2\}$ the dynamics of which is regulated by a Markov chain with a transition probability matrix $\mathbf{P} = \{p_{ij}\}$

where p_{ij} is the conditional probability to be in state j at time t given state i at time $t-1$ $(i, j = 1, 2)$. Given the adding–up constraints on the conditional probabilities, the parameters p_{11} and p_{22} are the nuisance parameters present only under the alternative hypothesis when the number of regimes is two.

The aim of EH is to test the null hypothesis of no regimes $H_{0:}\ y_t = \mu + \varepsilon_t$, with $\varepsilon_t \sim IIN\left(0, \sigma^2\right)$, against (1), but this cannot be done directly. To bypass the nuisance parameters problem, they keep the transition probabilities in both the null and the alternative hypotheses by verifying first whether $p_{11} = 1 - p_{22}$, $\mu_1 \neq \mu_2$, $\sigma_1 \neq \sigma_2$ (testing for an independent sequence of non-Markovian regimes), and then verifying whether $\mu_1 = \mu_2$, $\sigma_1 \neq \sigma_2$. In their empirical application, EH favor the MS model as the data generating process. This approach does not test for the number of regimes, but just for whether the data are serially uncorrelated or if they have equal mean.

Without a substantial loss in generality, let us adopt their framework and consider the case of k unknown, allowing it to take an integer value between 1 and T. Let $\theta_t = \left(\mu_t, \sigma_t^2\right)$ be the parameter vector of interest and, in a Bayesian framework, let us assume that it has an unknown distribution G belonging to a class of distributions $\mathcal{F}$. Following a nonparametric Bayesian approach (Ferguson, 1973), we can put a class of priors on $\mathcal{F}$ which should cover every kind of prior for $\mathcal{F}$ and be analitically manageable (Antoniak, 1974). The Dirichlet process, introduced by Ferguson (1973), is one suitable instrument to this end. We recall the main properties of the Dirichlet process in this context in the Appendix A.

Let us then assume:

$$\sigma_t^{-2} \sim \mathcal{G}\left(a/2, b/2\right), \tag{2}$$

$$\mu_t | \sigma_t^2 \sim \mathcal{N}\left(m, \sigma_t^2 \tau\right). \tag{3}$$

where $\mathcal{G}$ is a Gamma distribution and $\mathcal{N}$ is a Normal distribution; a, b, and m are hyperparameters to be chosen. We will also assume that θ_t is generated from an unknown distribution G, that follows a Dirichlet process $\mathcal{D}\left(AG_0\right)$, where A is a hyperparameter which regulates the prior probabilities on the number of regimes k, and G_0 is equal to the bivariate distribution (2)-(3) (cf. the Appendix A for further details). For added generality, one may think of also placing a prior distribution on the precision parameter τ (e.g. an Inverse Gamma, with hyperparameters $w/2$ and $W/2$). In addition, we suppose that for each t $(t = 1, \ldots, T)$ the $(y_t|\theta_t)$ are independent.

We follow Escobar and West (1995) in adopting a technique to estimate the empirical posterior distribution of k, based on the specific assumptions just made. In Appendix B we detail the steps used in the procedure which start from the expression of the distribution for a θ_t conditional on $\mathbf{\Theta}_{[-t]}$ (which includes the other θ_i's, $i = 1, \ldots, t-1, t+1, \ldots, T$) as a sample drawn from G. The posterior distribution $p\left(\theta_t | \mathbf{\Theta}_{[-t]}, \mathbf{Y}_T\right)$ is then computed and used as the basis for a Gibbs sampler to derive the distribution of the distinct values of the parameters as an estimate of the number of regimes k.

We will investigate the matter by first running a Monte Carlo experiment about the performance of the procedure in this context for a suitable choice of hyperparameters and using several data generating processes. We then we will return to the empirical issues about exchange rate behavior, and apply it to the data used by EH and by Engel (1994).

III A Monte Carlo Investigation

In order to get some indications as to the validity of the approach suggested here, we performed a Monte Carlo experiment aimed at assessing the behavior of the mode of the empirical distribution as an estimate of the number of regimes. The idea is to repeatedly generate the data according to a known DGP (either with or without MS properties) and derive the empirical distribution of the mode of number of regimes across replications. To maintain some economic significance to the results of this experiment, at the same time preserving its simplicity, we have chosen to span over number of regimes (1, 2, or 3), presence of an autoregressive component[4] and sample size (T=58, as in EH, and T=116 for a larger sample size).

Specifically, the series of y_t is generated from

$$(y_t - \mu_{s_t}) = \phi(y_{t-1} - \mu_{s_{t-1}}) + \varepsilon_t, \quad \varepsilon_t \sim IIN\left(0, \sigma^2_{s_t}\right), \quad t = 1, \dots T, \ s_t = \{1, \dots, k\}$$

according to the following set of DGPs:

1. DGP1_a: No AR ($\phi = 0$), one regime ($\mu_{s_t} = \mu = 0$, $\sigma^2_{s_t} = \sigma^2 = 9.991$, $\forall\, t$), T=58;

2. DGP1_b: AR(1) ($\phi = 0.7$), one regime ($\mu_{s_t} = \mu = 0$, $\sigma^2_{s_t} = \sigma^2 = 9.991$, $\forall\, t$), T=58;

3. DGP1_c: same as DGP1_a, but T=116;

4. DGP1_d: same as DGP1_b, but T=116;

5. DGP2_a: No AR ($\phi = 0$), two regimes (in regime 1, $\mu_1 = 3.256$, and $\sigma^2_1 = 9.991$; in regime 2, $\mu_2 = -2.712$, and $\sigma^2_2 = 36.921$, $p_{11} = 0.822$, $p_{22} = 0.908$, T=58. Note that for this DGP we have chosen the parameter values taken from the French Franc/US dollar estimation in EH.

[4]In the presence of autoregression, the conditional independence of $(y_t|\boldsymbol{\theta}_t)$ is lost. The rationale for working with a DGP with dependent observations is to check the robustness of our procedure relative to the violation of this assumption.

6. DGP2_b: AR(1) ($\phi = 0.7$), two regimes (in regime 1, $\mu_1 = 3.256$, and $\sigma_1^2 = 9.991$; in regime 2, $\mu_2 = -2.712$, and $\sigma_2^2 = 36.921$, $p_{11} = 0.822$, $p_{22} = 0.908$, T=58.

7. DGP2_c: same as DGP2_a, but T=116;

8. DGP2_d: same as DGP2_b, but T=116;

9. DGP3_a: No AR ($\phi = 0$), three regimes (in regime 1, $\mu_1 = -2$, and $\sigma_1^2 = 36$; in regime 2, $\mu_2 = 0$, and $\sigma_2^2 = 9$; in regime 3, $\mu_3 = 2$, and $\sigma_3^2 = 36$; the unconstrained transition probabilities are $p_{11} = 0.6$, $p_{12} = 0.2$, $p_{21} = 0.1$, $p_{22} = 0.7$, $p_{31} = .2$, $p_{32} = 0.1$; T=58.

10. DGP3_b: AR(1) ($\phi = 0.7$), three regimes (in regime 1, $\mu_1 = -2$, and $\sigma_1^2 = 36$; in regime 2, $\mu_2 = 0$, and $\sigma_2^2 = 9$; in regime 3, $\mu_3 = 2$, and $\sigma_3^2 = 36$; the unconstrained transition probabilities are $p_{11} = 0.6$, $p_{12} = 0.2$, $p_{21} = 0.1$, $p_{22} = 0.7$, $p_{31} = .2$, $p_{32} = 0.1$; T=58.

11. DGP3_c: same as DGP3_a, but T=116;

12. DGP3_d: same as DGP3_b, but T=116;

To apply the suggested Bayesian procedure, we need to choose the hyperparameters to be used in the successive Gibbs sampling involved (cf. Appendix B to recall the specifics used here). For practical purposes, the choice of prior probabilities on the number of regimes must be inspired by the kind of problem at hand: we want to show the details of our choice as an example of the kind of reasoning that could be repeated in other situations. Since the final results strongly depend on the choice of the priors, we could trick the cards and intervene in order to favor a specific number of regimes. More realistically, from a preliminary analysis of the data, investigation may result in some beliefs about the possible number of regimes k: for example, in the case studied by EH, the uncertainty is as to whether there is one regime or two. Accordingly, for a given sample size T, an impartial position could be to choose the hyperparameter A so as to assign maximum prior probabilities on $k = 1$ and $k = 2$. This strategy reflects a reasonable stance of uncertainty about the true number of regimes, while maintaining a connection with the type of data we work with. In Table 1, we report a few examples of prior distributions for k, based on (8) in the Appendix A, adopted to implement our procedure on the DGPs used in the simulations. The values should be seen as reflecting uncertainty between one or two regimes in the first two rows, and two or three in the other two rows.

Table 1: Prior Probability Distribution of k for Several Choices of A and T

		Number of Regimes							
A	**T**	1	2	3	4	5	6	7	8
0.19	116	0.374	0.377	0.181	0.054	0.012	0.002		
0.22	58	0.374	0.381	0.179	0.053	0.011	0.002		
0.40	116	0.133	0.283	0.284	0.180	0.082	0.028	0.008	0.002
0.48	58	0.126	0.281	0.288	0.184	0.083	0.028	0.008	0.002

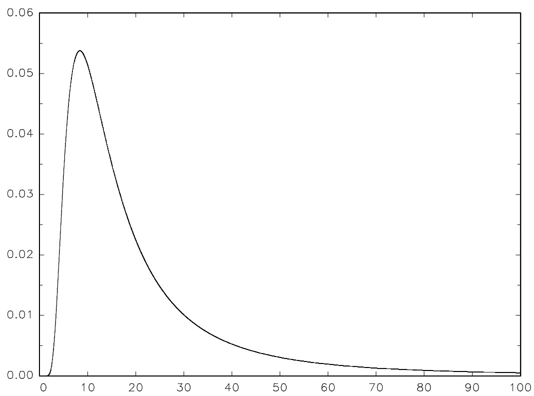

Figure 1: Prior distribution for σ_t^2.

As far as the mean distribution is concerned and in view of EH's assumptions, we can note that the observations with increasing trend should have positive mean and the observations with decreasing trend should have negative mean. To be realistic and not induce unrealistically high values (in reference to quarterly returns on exchange rate), we can safely choose a Normal with zero mean, i.e. $m = 0$ as a prior distribution for μ_t.

As far as σ_t^2 is concerned, an impartial choice of Inverse Gamma prior would favor one that assigns large probabilities to a large range of values. Recall that the Inverse Gamma random variable X with parameters $a/2$ $(a \geq 4)$ and $b/2$ has mean $b/(a-2)$ and variance $2b^2/((a-2)^2(a-4))$. If $a = 4$ then we obtain a distribution with infinite variance.

Choosing the other hyperparameter b to be equal to 50, we obtain the distribution in Figure 1, which has the desired properties. Finally, we have to choose the prior distribution

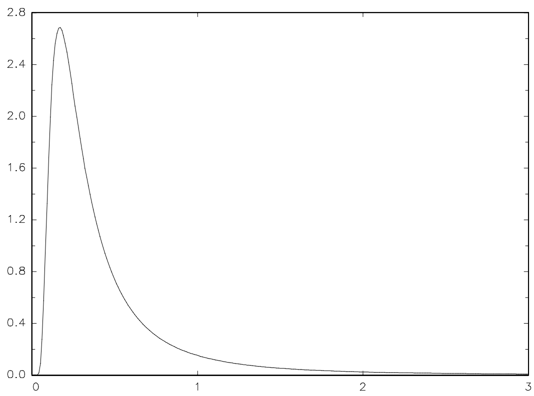

Figure 2: Prior distribution for τ.

for τ, which determines the variance of μ_t. If τ is 1, $E(\sigma_t^2)$= 25, hence smaller values of τ are preferable with a good distribution for τ assigning a large probability to the values within the interval between 0 and 1, to limit the effect of σ_t^2, but at the same time it must have infinite variance not to steer results in a specific direction and hence guarantee a certain degree of uncertainty. The distribution that we use is an Inverse Gamma with hyperparameters $w = 4$ and $W = 1$ (an Inverse Chi-Square with 4 degree of freedom), with the graph displayed in Figure 2. When applying the procedure for the Monte Carlo experiment, the only hyperparameter which we vary across DGPs is A, with values to reflect uncertainty between k=1 and k=2 for DGP1 and DGP2 and between k=2 and k=3 for DGP3 (cf. Table 1).

Now we are in a position to apply the suggested Gibbs sampling procedure to determine the number of regimes. For the choice of the number of iterations for burn-in to convergence, we have applied the Gelman and Rubin (1992) methodology, using the CODA package of Best et al. (1995). Across esperiments we usually obtain convergence in distribution after 2000 iterations, choosing an observation every 10, to eliminate autocorrelation in the sample. At any rate, for our application, to be sure that the Gibbs sampler has converged, we select the observations after 5000 iterations, saving draws every other 10 iterations, to obtain a sample of 500 observations on $\mathbf{\Theta}_T$. The empirical posterior distribution $p\left(k|\mathbf{Y}_T\right)$ is evaluated on 1000 series generated from the DGPs with T=58 and 500 series generated from the DGPs of length T=116.[5]

[5]Even on a Pentium III with 1Ghz processor, the time required to generate the results is very high, given the high number of times the Gibbs sampler is run, so that experiments with a larger sample size would

Table 2: Empirical Distributions of k for various DGPs

DGPs	Number of Regimes			
	1	2	3	4
1a: No AR, k=1, T=58, A=0.22	0.870	0.130		
1b: AR(1), k=1, T=58, A=0.22	0.587	0.412	0.001	
1c: No AR, k=1, T=116, A=0.19	0.882	0.118		
1d: AR(1), k=1, T=116, A=0.19	0.608	0.382	0.010	
2a: No AR, k=2, T=58, A=0.22	0.350	0.610	0.040	
2b: AR(1), k=2, T=58, A=0.22	0.285	0.677	0.038	
2c: No AR, k=2, T=116, A=0.19	0.246	0.640	0.114	
2d: AR(1), k=1, T=116, A=0.19	0.190	0.684	0.126	
3a: No AR, k=3, T=58, A=0.48		0.113	0.794	0.093
3b: AR(1), k=3, T=58, A=0.48		0.107	0.763	0.130
3c: No AR, k=3, T=116, A=0.40		0.086	0.818	0.096
3d: AR(1), k=1, T=116, A=0.40		0.072	0.784	0.144

The results of the experiment provide the evidence presented in Table 2 which contains the relative frequencies of the times that the mode of the posterior distribution in each replication points to the number of regimes in the column. For ease of reference we recall the main characteristics of each DGP by row. The general features of the experiments point to a general satisfactory performance of the procedure, since the correct number of regimes is always detected with the highest frequency. The presence of autocorrelation in the DGP seems to induce a tendency of the procedure to increase the relative frequency with which a larger k is detected.[6] When no MS is present, therefore, this may signal more than one regime, while in the presence of two regimes the problem is less relavant. When the sample size increases, the performance of the procedure improves slightly.

become very demanding.

[6]We thank the Associate Editor for pointing out this possibility to us.

Table 3: Posterior distributions for dollar exchange rates series.

$Pr(k) \backslash k$	1	2	3	4	5	6
DEM	0.273	**0.408**	0.238	0.062	0.014	0.005
FRF	0.355	**0.369**	0.190	0.069	0.015	0.002
GBP	**0.412**	0.402	0.132	0.042	0.009	0.003

IV Long Swings Revisited

We have run the same procedure with the same hyperparameters on the exchange rate returns for the three currencies used by EH and then for the seven currencies used in Engel (1994).

IV.1 Quarterly Data

The quarterly data are referred to the US dollar relative to the Deutsche Mark (DEM), the French Franc (FRF) and the British Pound (GBP); the empirical posterior distributions are reported in Table 3 (the boldface numbers correspond to the distribution mode by row). There is a strong evidence of two regimes for the series DEM. For the series FRF the evidence is less strong, since the mode is in $k = 2$ with a high value for $k = 1$ as well. For the GBP, the evidence is in favor of $k = 1$ even if, again, the value for $k = 2$ is not much smaller.

We will maintain, therefore, that the results provided by EH are a good representation for the DEM and FRF series. For the case of the GBP, we will investigate whether the true model is the linear model (random walk) or an homoskedastic Markov switching model. This time, though, we can follow a classical approach by estimating the MS model:[7] $y_t = \mu_{s_t} + \varepsilon_t$, with $\varepsilon_t \sim IIN\left(0, \sigma^2\right)$.[8] In this case, the likelihood ratio test can be used to verify this

[7]The log-likelihood function of a Markov switching model presents numerous local maxima, so that the final estimation depends on the starting values. To choose the starting values, we have selected various grids for unknown parameters, starting from the combination with highest log-likelihood. The grids are: $\mu_1 \in [0, 5]$ and $\mu_2 \in [-5, 0]$ both with step-length 1; $\sigma^2 \in [8, 28]$ with step-length 2; $p_{11} \in [0.5, 0.95]$ and $p_{22} \in [0.5, 0.95]$ both with step-length 0.05, for a total of 39,600 combinations.

[8]We obtain the following values for the parameters (standard errors are in parentheses):

$$\widehat{\mu}_1 = \underset{(0.823)}{2.661}, \quad \widehat{\mu}_2 = \underset{(1.398)}{-3.838}, \quad \widehat{\sigma}^2 = 18.583, \quad \hat{p}_{11} = \underset{(0.093)}{0.935}, \quad \hat{p}_{22} = \underset{(0.112)}{0.902}.$$

model against the EH model, because both of them contain the transition probabilities; the likelihood ratio statistics is equal to 0.238 (to be compared with critical values from a χ^2 with one degree of freedom), not allowing the rejection of the null hypothesis of no difference between the variances across regimes.

To verify the latter model against the linear, we can resort to a Bayesian procedure, along the lines suggested by West et al. (1994) to take into account homoskedasticity. Let A equal 0.22, let $\mathcal{G}(4/2, 50/2)$ be the σ^{-2} prior distribution and let $N(0, 10)$ be the prior distribution for μ_t (thus allowing high draw probability for a large range of values). The empirical posterior distribution for the number of regimes is:

k :	1	2	3	4	5	6
$\Pr(k)$:	0.352	**0.388**	0.193	0.053	0.009	0.005

that has mode $k = 2$. Hence for the exchange rate British Pound/US Dollar, the homoskedastic MS model receives empirical support.

We can note that the posterior probability of states 1 and 2 are very similar in various experiments and we acknowledge that this may be caused by the small number of observations. In such cases the final result would not be robust with respect to the choice of the prior, but this is a common problem to other Bayesian methods, especially in the presence of a small sample size.

IV.2 Monthly Data

Let us follow the extension of the analysis, as suggested by Engel (1992), and consider the performance of the procedure when applied to monthly observations on seven exchange rates (Canadian Dollar - CAD, French Franc – FRF , Italian Lira – ITL, Japanese Yen - JPY, Swiss Franc - CHF, British Pound - GBP, and Deutsche Mark – DEM), using the same MS model. We have chosen to use three priors: one which assigns a high probability to $k = 1$, the second which corresponds to a prior with modes in $k = 1$ and $k = 2$, and the third which places a mode at $k = 2$ with a non-zero probability at $k = 1$ and $k = 3$. The A values chosen and the corresponding priors[9] are summarized in the Table 4.

The asymptotic theory for ML estimation of MS models is by no means standard. The consistency of the ML estimator and the consistency and asymptotic normality of the pseudo-ML estimator are ensured under given conditions as shown by Francq and Roussignol (1997).

[9]We have had numerical problems calculating the prior with (8) in the Appendix, using T=214 as in Engel (1992). Following the result by Escobar and West (1995) that the prior is not sensible to changes in T, when the sample size is sufficiently high, we have calculated the prior fixing $T = 170$.

Table 4: Priors for different values of A, with T=170.

	k						
A	1	2	3	4	5	6	7
0.01	0.945	0.054	0.001				
0.18	0.367	0.377	0.184	0.057	0.013	0.002	
0.22	0.295	0.371	0.221	0.084	0.023	0.005	0.001

By applying our procedure we obtain the posterior distributions for the regimes reported in Table 5.

We can note the strong evidence of two regimes for the Canadian Dollar, Japanese Yen, and Swiss Franc, one regime for the French Franc and three regimes for the British Pound; the Italian Lira seems to be particularly sensible to the prior, since the mode is 1 for $A = 0.01$ and $A = 0.18$ while it is 2 for the third prior. The Deutsche Mark is a puzzling case in which the mode changes without any seemingly clear pattern with respect to the changes in the prior. The case of the British Pound is interesting since it provides evidence in favor of three states, which is in agreement with the results in Engel (1994).

Repeating the experiment for the case with equal variance for FRF, ITL and DEM, using the same priors as in the quarterly case, except for the A parameter, we obtain the results in Table 6.

The evidence in favor of one state is now rather clear. These results are coherent with the classical tests. In fact, the likelihood ratio tests show the rejection of the null of MS model with just a changing mean, versus the MS model with both mean and variance switching. Engel (1994) notes that the null hypothesis of monthly heteroskedastic MS model with unchanging means is accepted against the MS model with mean and variance switching; this, along with our results, confirms the impression that the segmented trends get hidden by the size of monthly returns with time-varying volatility.

V Concluding Remarks

In this paper, we have proposed a new method to detect the number of regimes for MS models, building on some results of nonparametric Bayesian statistics. The idea is to consider the number of regimes as a discrete random variable with a prior distribution; the

Table 5: Posterior distributions for different values of A.

		k						
		1	2	3	4	5	6	7
A=0.01	CAD	0.211	**0.773**	0.016				
	FRF	**0.994**	0.006					
	ITL	**0.737**	0.244	0.018	0.001			
	JPY		**0.934**	0.064	0.002			
	CHF		**0.972**	0.027	0.001			
	GBP		0.020	**0.945**	0.035			
	DEM	**0.884**	0.110	0.006				
A=0.18	CAD	0.069	**0.522**	0.300	0.099	0.010		
	FRF	**0.870**	0.121	0.009				
	ITL	**0.435**	0.383	0.145	0.033	0.002	0.002	
	JPY	0.115	**0.523**	0.271	0.076	0.014	0.001	
	CHF		**0.405**	0.389	0.154	0.042	0.008	0.002
	GBP	0.129	0.233	**0.384**	0.181	0.059	0.013	0.001
	DEM	0.023	**0.409**	0.377	0.154	0.029	0.008	
A=0.22	CAD	0.237	**0.391**	0.241	0.106	0.020	0.004	0.001
	FRF	**0.841**	0.148	0.011				
	ITL		**0.401**	0.370	0.167	0.049	0.012	0.001
	JPY	0.029	**0.529**	0.309	0.114	0.015	0.003	0.001
	CHF	0.005	**0.396**	0.369	0.162	0.057	0.010	0.001
	GBP	0.026	0.314	**0.373**	0.189	0.075	0.021	0.002
	DEM	**0.553**	0.339	0.090	0.016	0.002		

Table 6: Posterior distributions for different values of A: selected exchange rates.

		k					
		1	2	3	4	5	6
	FRF	**0.956**	0.044				
A=0.01	ITL	**0.968**	0.031	0.001			
	DEM	**0.960**	0.040				
	FRF	**0.436**	0.364	0.148	0.039	0.012	0.001
A=0.18	ITL	**0.488**	0.349	0.124	0.036	0.003	
	DEM	**0.513**	0.350	0.111	0.019	0.007	
	FRF	0.397	**0.400**	0.161	0.034	0.008	
A=0.22	ITL	**0.448**	0.362	0.146	0.038	0.003	0.003
	DEM	**0.503**	0.337	0.122	0.028	0.010	

mode of empirical posterior distribution will provide evidence about the number of regimes for the model. In practical applications, our procedure can be seen as a preliminary step to determine the number of regimes before estimating an MS model.

As shown in the applications, this procedure can be used also after the estimation of a given MS model with a fixed number of regimes, using the information provided by the parameter estimates to express the priors; in this case the present procedure would serve as a test.

We are aware of the fact that, strictly speaking, the use of the number of regimes first as a random variable (in the specification step) and then as a constant (in the estimation) is not correct in a Bayesian context. It would be more correct to consider various MS models each with a different number of states, weighted with the corresponding prior probability assigned to each model, and then choose the model with the highest posterior probability: but this procedure would be cumbersome. Invoking the principle of ecumenism in statistics put forth by Box (1983), we mix classical and Bayesian approaches (not an uncommon practice in simulation-based econometrics) achieving a simplification in the computational efforts.

In this paper we have made a specific reference to the MS models popularized by Hamilton, but the approach is valid for every switching model with unknown number of regimes. The choice of the Hamilton model for the examples here is instructive because it is very popular in econometrics and it is a leading case for the nuisance parameters problem in testing. The empirical exercise performed here on exchange rate long swings shows that the procedure is helpful in refining model specification.

References

[1] Abramowitz M. and Stegun I. A. (1972): *Handbook of Mathematical Functions*. New York: Dover Publications.

[2] Albert J. and Chib S. (1993): Bayes Inference Via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance Shifts. *Journal of Business & Economic Statistics* 11, 1:15.

[3] Andrews D. W. K. (1993): Tests for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica* 61, 821:856.

[4] Andrews D. W. K. and Ploberger W. (1994): Optimal Tests when a Nuisance Parameter Is Present only under the Alternative. *Econometrica* 62, 1383:1414.

[5] Antoniak C. E. (1974): Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems. *The Annals of Statistics* 2, 1152:1174.

[6] Best N. G., Cowles M. K. and Vines K. (1995): CODA: Convergence Diagnosis and Output Analysis Software for Gibbs Sampling Output, Version 3.0. *Technical Report*, University of Cambridge, MRC Biostatistics Unit.

[7] Böhning D., Dietz E., Schaub R., Schlattmann P. and Lindsay B. G. (1994): The Distribution of the Likelihood Ratio for Mixtures of Densities from the One-Parametric Exponential Family. *Annals of the Institute of Statistical Mathematics* 46, 373:388.

[8] Böhning D., Dietz E. and Schlattmann P. (1998): Recent Developments in Computer-Assisted Analysis of Mixtures. *Biometrics* 54, 367:377.

[9] Böhning D., Schlattmann P. and Lindsay B. G. (1992): Computer-Assisted Analysis of Mixtures (C.A.MAN): Statistical Algorithms. *Biometrics* 48, 283:303.

[10] Box G. E. P. (1983): An Apology for Ecumenism in Statistics. In *Scientific Inference, Data Analysis, and Robustness* (Box G. E. P., Leonard T. e Wu C.-F. eds.), 51:83. Academic Press.

[11] Carter C. K. and Kohn R. (1994): On Gibbs Sampling for State Space Models. *Biometrika* 81, 541:553.

[12] Carter C. K. and Kohn R. (1996): Markov Chain Monte Carlo in Conditionally Gaussian State Space Models. *Biometrika* 83, 589:601.

[13] Davies R. B. (1977): Hypotheses Testing when a Nuisance Parameter Is Present only under the Alternative, *Biometrika* 64, 247:254.

[14] Engel C. (1992): Can the Markov Switching Model Forecast Exchange Rates?, *NBER* WP 4210.

[15] Engel C. (1994): Can the Markov Switching Model Forecast Exchange Rates?, *Journal of International Economics,* 36, 151:165.

[16] Engel C. and Hamilton J. D. (1990): Long Swings in the Dollar: Are They in the Data and Do Markets Know It?, *American Economic Review* 80, 689:713.

[17] Escobar M. D. (1994): Estimating Normal Means with a Dirichlet Process Prior. *Journal of the American Statistical Association* 89, 268:277.

[18] Escobar M. D. and West M. (1995): Bayesian Density Estimation and Inference Using Mixtures. *Journal of the American Statistical Association* 90, 577:588.

[19] Feng Z. D. and McCulloch C. E. (1996): Using Bootstrap Likelihood Ratios in Finite Mixture Models. *Journal of the Royal Statistical Society, Series B* 58, 609:617.

[20] Ferguson T. S. (1973): A Bayesian Analysis of Some Nonparametric Problems. *The Annals of Statistics* 1, 209:230.

[21] Francq C. and Roussignol M. (1997): On White Noises Driven by Hidden Markov Chains. *Journal of Time series Analysis* 18, 553:578.

[22] Garcia R. (1995): Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models. *International Economic Review, 39, 763-88*.

[23] Garcia R. and Schaller H. (1995): Are the Effects of Monetary Policy Symmetric? *Discussion Paper* 9510. Université de Montréal, Département des sciences économiques.

[24] Gelman A. and Rubin D. B. (1992): Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science* 7, 457:511.

[25] Green P. J. (1995): Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika* 82, 711:732.

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

[27] Hansen B. E. (1992): The Likelihood Ratio Test under Nonstandard Conditions: Testing the Markov Switching Model of GNP, *Journal of Applied Econometrics* 7, S61:S82.

[28] McCulloch R. E. and Tsay R. S. (1994): Statistical Analysis of Economic Time Series Via Markov Switching Models. *Journal of Time Series Analysis* 15, 523:539.

[29] Raftery A. E. (1996): Hypothesis Testing and Model Selection, in *Markov Chain Monte Carlo in Practice* (eds. W. R. Gilks, S. Richardson and D. J. Spiegelhalter), ch. 10, 163:188. London: Chapman and Hall.

[30] Richardson S. and Green P. J. (1997): On Bayesian Analysis of Mixtures with an Unknown Number of Components. *Journal of the Royal Statistical Society, Series B* 4, 731:792.

[31] Roeder K. and Wasserman L. (1997): Practical Bayesian Density Estimation Using Mixtures of Normals. *Journal of the American Statistical Association* 92, 894:902.

[32] Shephard N. (1994): Partial Non-Gaussian State-Space. *Biometrika* 1, 115:131.

[33] West M. (1992): Modelling with Mixtures. *Bayesian Statistics 4* (Bernardo J. M., Berger a. P. e Smith A. F. M. eds.), 503:524.

[34] West M., Müller P. and Escobar M. D. (1994): Hierarchical Priors and Mixture Models, with Application in Regression and Density Estimation *Aspects of Uncertainty* (Freeman P. R. e Smith A. F. M. eds.): 363:386, J. Wiley & Sons Ltd.

Appendix A: Dirichlet Processes

In a Bayesian context, let us consider θ to be a parameter vector with unknown distribution G and let $\mathcal{F}$ be a class of distributions containing G; in a nonparametric Bayesian approach we can put a class of priors, say $\mathcal{P}$, on $\mathcal{F}$. Antoniak (1974) indicates some properties for $\mathcal{P}$: they consist, essentially, in covering every kind of prior for $\mathcal{F}$ and in being analitically manageable. The Dirichlet process, introduced by Ferguson (1973), is one suitable instrument to this end.

Ferguson considers a non-null, finite, non-negative and finitely additive measure α on the space $(\mathbf{\Theta}, \mathcal{A})$, where $\mathbf{\Theta}$ is a set and $\mathcal{A}$ a sigma-algebra of $\mathbf{\Theta}$. In addition, let P be a random probability measure, that is a probability satisfying the following conditions:

1) $P(B) \in [0,1]$ for every $B \in \mathcal{A}$;

2) $P(\mathbf{\Theta}) = 1$ a. s.;

3) if $B_1^{'}, B_2^{'}, ..., B_n^{'}$ is a measurable partition of $\mathbf{\Theta}$, and $B_1 = \bigcup_{i=1}^{n_1} B_i^{'}$, $B_2 = \bigcup_{i=n_1+1}^{n_2} B_i^{'}$, ... $B_s = \bigcup_{i=n_{s-1}+1}^{n} B_i^{'}$, then the joint distribution of $[P(B_1), ..., P(B_s)]$ is identical to the distribution of:

$$\left[\sum_{i=1}^{n_1} P\left(B_i^{'}\right), \sum_{i=n_1+1}^{n_2} P\left(B_i^{'}\right), ... \sum_{i=n_{s-1}+1}^{n} P\left(B_i^{'}\right)\right].$$

We say that P is a Dirichlet process on $(\boldsymbol{\Theta}, \mathcal{A})$ with parameter α, if, for every strictly positive integer h and for every measurable partition $(B_1, B_2, ..., B_h)$ of $\boldsymbol{\Theta}$, $[P(B_1), ..., P(B_h)]$ has a Dirichlet distribution with parameter α measurable on the partition, that is

$[\alpha(B_1), \alpha(B_2), ...\alpha(B_h)]$, or:

$$[P(B_1), ..., P(B_h)] \sim \mathcal{D}[\alpha(B_1), \alpha(B_2), ...\alpha(B_h)]$$

or simply $[P(B_1), ..., P(B_h)] \sim \mathcal{D}(\alpha)$.

Ferguson verifies that this definition satisfies the Kolmogorov criteria, existing a probability $\mathcal{P}$ on the space of all functions from $\mathcal{A}$ into $[0, 1]$ the $\sigma-$field generated by the field of cylinder sets; $\mathcal{P}$ yields these distributions.

Three properties of this process are useful in the context of the determination of the number of regimes:

Property 1:

$$E[P(B_i)] = \alpha(B_i) / \alpha(\boldsymbol{\Theta}), \tag{4}$$

where $\alpha(\boldsymbol{\Theta}) = \sum_{i=1}^{h} \alpha(B_i)$.

Property 2:

$(\theta_1, ...\theta_n)$ is a sample of size n from P if, for any $k = 1, 2, \ldots$ and measurable sets $C_1, ..., C_n$, $B_1, ..., B_k$:

$$\mathcal{P}[(\theta_1 \in C_1), ..., (\theta_n \in C_n) | P(B_1), ..., P(B_k), P(C_1), ..., P(C_n)] = \prod_{i=1}^{n} P(C_i), \tag{5}$$

where $(C_1, ..., C_n) \in \mathcal{A}$.

Property 3:

if $[P(B_1), ..., P(B_h)] \sim \mathcal{D}[\alpha(B_1), \alpha(B_2), ...\alpha(B_h)]$ and $(\theta_1, ...\theta_n)$ is a sample from P, then:

$$[P(B_1), ..., P(B_h) | \theta_1, ...\theta_n] \sim \mathcal{D}\left[\alpha(B_1) + \sum_{i=1}^{n} \delta_{\theta_i}(B_1), ..., \alpha(B_h) + \sum_{i=1}^{n} \delta_{\theta_i}(B_h)\right], \tag{6}$$

where $\delta_{\theta_i}(B_j)$ is 1 if $\theta_i \in B_j$, 0 otherwise.

Let us suppose that G follows a Dirichlet process $\mathcal{D}(\alpha)$. Let G_0 be the prior expectation value of G and $A = \alpha(\boldsymbol{\Theta})$, so, for equation (4), $G_0(\cdot) = \alpha(\cdot)/A$ and $\alpha = AG_0$. In other terms, we can write $G \sim \mathcal{D}(AG_0)$, where A indicates the concentration of the prior distribution for G around G_0; a high A indicates that there is a strong prior probability that the true G is G_0. The previous properties can be used to establish the posterior expected value of G.

Let us draw a sample $(\theta_1, ...\theta_n)$ from G (in the sense of Property 2); if we use a quadratic loss function, the estimation of the distribution of $\left(\theta|\theta_1, ...\theta_n\right)$ is the expected value of $[G(\theta)|\theta_1, ...\theta_n]$, and, using properties 1 and 3,

$$\left[\hat{G}(\theta)|\theta_1, ...\theta_n\right] = \frac{\alpha(\theta) + \sum_{i=1}^{n} \delta_{\theta_i}(\theta)}{\alpha(\mathbf{\Theta}) + n} = \frac{A}{A+n} G_0 + \frac{1}{A+n} \sum_{i=1}^{n} \delta_{\theta_i}(\theta). \tag{7}$$

Note that (7) is simple to interpret as a parameterization of the distribution of θ; in addition, G_0 can assume all the standard distributions used in a Bayesian context.

There are some interesting results in Antoniak (1974) which can be exploited for our purposes, namely the expression for the probability of obtaining k different components in the last term of (7), that is:

$$\Pr(k_n) = {}_n a_k A^k / A^{(n)}, \tag{8}$$

where ${}_n a_k$ is the first type Stirling number in absolute value (tabulated in Abramowitz and Stegun, 1972, p. 833) and $A^{(n)} = A(A+1)...(A+n-1)$. In addition, the expected value of k depends only on A and n; in fact:

$$E(k_n) = \sum_{i=1}^{n} A/(A+i-1) \approx A\left[\log\left(\frac{n+A}{A}\right)\right]. \tag{9}$$

When the size of parameter vector θ increases, we can no longer express the estimation process of G and θ analitically. The Gibbs sampling method is a good way to bypass this problem (West, 1992; Escobar, 1994; Escobar and West, 1995).

Appendix B: The Gibbs Sampler in this Context

As mentioned in the main text, we can consider θ_t as the θ in (7) and the other θ_i ($i = 1, \ldots, t-1, t+1, \ldots, T$) as the sample drawn from G. We can then use (7) to express the conditional distribution $p\left(\theta_t|\mathbf{\Theta}_{[-t]}\right)$, where $\mathbf{\Theta}_{[-t]}$ denotes $\left\{\theta_1, ...\theta_{t-1}, \theta_{t+1} ..., \theta_T\right\}$:

$$p\left(\theta_t|\mathbf{\Theta}_{[-t]}\right) = \frac{A}{A+T-1} G_0(\theta_t) + \frac{1}{A+T-1} \sum_{j=1, j\neq t}^{T} \delta_{\theta_j}(\theta_t). \tag{10}$$

There is a probability $\frac{A}{A+T-1}$ that θ_t is different from the other terms in $\mathbf{\Theta}_T$ and a probability $\frac{1}{A+T-1}$ that θ_t is equal to the j-th term in the matrix $\mathbf{\Theta}_T \equiv (\theta_1\theta_2 \ldots \theta_T)$. Of course, if there are n_j terms of the sample equal to θ_j, the probability that $\theta_t = \theta_j$ would be $\frac{n_j}{A+T-1}$.

Based on Bayesian statistical properties, the posterior distribution of $\left(\theta_t|\mathbf{\Theta}_{[-t]}\right)$ is:

$$p\left(\theta_t|\mathbf{\Theta}_{[-t]},\mathbf{Y}_T\right)=q_0G_t\left(\theta_t\right)+\sum_{j=1,j\neq t}^{T}q_j\delta_{\theta_j}\left(\theta_t\right), \tag{11}$$

where $\mathbf{Y}_T$ is $(y_1,y_2,\ldots y_T)$, and $G_t\left(\theta_t\right)$ is the bivariate Normal-Inverse Gamma, with components:

$$\sigma_t^{-2}\sim\mathcal{G}\left[(a+1)/2,\beta_t/2\right], \tag{12}$$

$$\mu_t|\sigma_t^2\sim N\left(x_t,X\sigma_t^2\right). \tag{13}$$

The new parameters are expressed by (see Escobar and West, 1995):

$$q_0\propto A\frac{\Gamma\left[(1+a)/2\right]}{\Gamma\left(a/2\right)a^{1/2}}\left\{1+\left(y_t-m\right)^2/\left[(1+\tau)b\right]\right\}^{-(1+a)/2}\left[(1+\tau)b/a\right]^{-1/2},$$

$$q_j\propto\exp\left[-\left(y_t-\mu_j\right)^2/\left(2\sigma_j^2\right)\right]\left(2\sigma_j^2\right)^{-1/2},\qquad j=1,\ldots,T$$

with $q_0+...+q_{t-1}+q_{t+1}+...+q_T=1$;

$$\beta_t=b+\left(y_t-m\right)^2/\left(1+\tau\right),$$

$$x_t=\left(m+\tau y_t\right)/\left(1+\tau\right),$$

$$X=\tau/\left(1+\tau\right).$$

If we can specify the values of a, b, m and τ from available prior information, the only parameter to specify is the precision of the Dirichlet process, A. Note that we know the prior probability for k, expressed by (8) and its expected value, expressed by (9), which depends on the sample size (in this case the length of the series, T) and on A. Therefore, if we have some expectation about the number of regimes, we can choose the A that fits this expectation.

Now, we can use the Gibbs sampler to infer the number of regimes; the steps are:

1) choose a starting value for $\mathbf{\Theta}_T$, call it $\mathbf{\Theta}_T^0$ drawing every θ_t from $G_t\left(\theta_t^0\right)$;

2) sample θ_1 from $\left(\theta_1|\mathbf{\Theta}_{[-1]},\mathbf{Y}_T\right)$, θ_2 from $\left(\theta_2|\mathbf{\Theta}_{[-2]},\mathbf{Y}_T\right)$, ..., θ_T from $\left(\theta_T|\mathbf{\Theta}_{[-T]},\mathbf{Y}_T\right)$, obtaining a new $\mathbf{\Theta}_T$. Note that the last θ_t sampled is inserted immediately in $\mathbf{\Theta}_{[-(t+1)]}$ for the subsequent draw;

3) iterate step 2 until convergence;[10] we thereby obtain the first element $\mathbf{\Theta}_T(1)$ for the first replication;

[10] The convergence is in distribution, so that every θ_i is considered sampled from the posterior distribution $\left(\mathbf{\Theta}_T|\mathbf{Y}_T\right)$.

4) repeat step 2 N times, obtaining $\mathbf{\Theta}_T(2)$, ..., $\mathbf{\Theta}_T(N)$;

5) enumerate the k distinct values in $\mathbf{\Theta}_T(i)$ and construct the empirical posterior distribution $p(k|\mathbf{Y}_T)$;

6) the mode of $p(k|\mathbf{Y}_T)$ can be taken as an estimate of the number of regimes for the switching model.

The analysis exposed here is the simplest, because the hyperparameters are considered known. Surely it is convenient to establish a prior for τ, which is a crucial task in the determination of the form of the prior distribution for $\mu_t|\sigma_t^2$. The prior used by Escobar and West (1995) is (w and W are constants):

$$\tau^{-1} \sim \mathcal{G}\left(\frac{w}{2}, \frac{W}{2}\right), \tag{14}$$

that has a posterior distribution that does not depend on y_t; in the Gibbs sampling τ^{-1} is to be sampled as:

$$\left(\tau^{-1}|\mathbf{\Theta}_T\right) \sim \mathcal{G}\left(\frac{w+k}{2}, \frac{W+K}{2}\right), \tag{15}$$

where k is the number of different components of $\mathbf{\Theta}_T$ and $K = \sum_{i=1}^{k} \frac{(\mu_i - m)^2}{\sigma_i^2}$. The steps in the Gibbs sampler do not change, but we have to add the drawing of τ (from (14) for the starting value and from (15) for the other drawings), before the steps 1-4.

The simpler case, with unchanged variance, is illustrated in West et al. (1994). The principal difference relative to our case is that the distribution of σ and μ_t are independent; in fact:

$$\sigma^{-2} \sim \mathcal{G}(a/2, b/2),$$

$$\mu_t \sim N\left(m, v^2\right).$$

The posterior distributions corresponding to (12) and (13), are:

$$\sigma^{-2} \sim \mathcal{G}\left[(a+T)/2, \beta/2\right],$$

$$\mu_t \sim N(x_t, X),$$

where:

$$\beta = b + \sum_{t=1}^{T} (y_t - \mu_t)^2,$$

$$x_t = \left(\sigma^2 m + v^2 y_t\right) / \left(\sigma^2 + v^2\right),$$

$$X = v^2\sigma^2 / \left(v^2 + \sigma^2\right),$$

and the drawing probabilities:

$$q_0 \propto A \exp\left[-\left(y_t - m\right)^2 / 2\left(v^2 + \sigma^2\right)\right] \left[2\left(v^2 + \sigma^2\right)\right]^{-1/2},$$

$$q_j \propto \exp\left[-\left(y_t - \mu_j\right)^2 / \left(2\sigma^2\right)\right] \left(2\sigma^2\right)^{-1/2}, \qquad j = 1, \ldots, T.$$