Forecasting with Non-homogeneous Hidden Markov Models

Loukia Meligkotsidou[1]

Department of Mathematics, University of Athens, Greece

and

Petros Dellaportas

Department of Statistics, Athens University of Economics and Business, Greece

Summary We present a Bayesian forecasting methodology of discrete-time finite state-space hidden Markov models with non-constant transition matrix that depends on a set of exogenous covariates. We describe an MCMC reversible jump algorithm for predictive inference, allowing for model uncertainty regarding the set of covariates that affect the transition matrix. We apply our models to interest rates and we show that our general model formulation improves the predictive ability of standard homogeneous hidden Markov models.

Keywords *Exogenous variables; Markov chain Monte Carlo; Multinomial Regression; Prediction; Reversible jump*

1   Introduction

We deal with discrete-time finite state-space hidden Markov models (HMM) that have been extensively used as standard modelling devices for stochastic processes on which only partial observations are made. The standard form of these models is specified as follows. The data generation mechanism consists of a bivariate process $\{(Y_t, Z_t)\}$ with $\{Z_t\}$ being an unobservable, or hidden, finite Markov chain that governs the distribution of the observable process $\{Y_t\}$.

Modelling economic and financial time series via HMMs has been proposed by Hamilton (1994). There are many applications and generalisations in financial econometrics, see for example Garcia and Perron (1996), Krozlig (1997), Schaller and Van Norden (1997), Kim and Nelson (1999), Franses and van Dijk (2000). MCMC-based advances in computational Bayesian statistics have expanded their wide applicability; see, for example, McCulloch and Tsay (1994), Chib (1996), Billio *et al.* (1999) and Frühwirth-

---

[1]Email for correspondence: meligots@math.uoa.gr

Schnatter (2001). For a general background on HMMs and their applications see the books by MacDonald and Zucchini (1997), Cappé *et al.* (2005) and Frühwirth-Schnatter (2006). For a recent review of classical versus Bayesian inference for HMMs see Ryden (2008).

An important application of HMMs is forecasting. In contrast with the threshold models that are used to just describe the mechanism that produced the data regime-switching, the Markovian structure of HMMs allows for one step ahead forecasts of $\{Z_t\}$ and consequently of $\{Y_t\}$. However, HMMs can be seen as generalisations of threshold models, in the form of the multiple-change point formulation introduced by Chib (1998); for forecasting applications in economic time series see, for example, Koop (2003) and Pesaran *et al.* (2006).

In this paper we are dealing with forecasting using non-homogeneous HMMs. Specifically, we consider models which incorporate covariate information in two ways. First, we assume that Y_t may depend not only on Z_t, but also on a set of exogenous covariates $W_t = (W_{1t}, \ldots, W_{nt})$ that have been observed up to time $t - 1$. Our methodology is immediately applicable to cases in which past values of Y_t can be used instead of the covariates W_t, implying an $AR(p)$ model for $\{Y_t\}$. Secondly, we assume that the hidden process $\{Z_t\}$ is a non-homogeneous Markov chain with time-varying transition matrix $P^{(t)}$, the values of which depend on a further set of covariates $X_t = (X_{1t}, \ldots, X_{kt})$. Thus, our model formulation assumes that some covariates (or lagged values of Y_t) affect directly, in a linear fashion, the density of Y_t, whereas past values of the same or some other covariates affect Y_t through the transition matrix of the hidden process.

Past relevant literature on non-homogeneous HMMs is scarce. Diebold *et al.* (1993) have considered maximum likelihood estimation of the simple two-state Gaussian HMM with time-varying transition matrix. Applications of HMMs with time-varying transitions include Durland and McCardy (1994), Gray (1996), Peria (2002), Masson and Ruge-Murcia (2005), Kim *et al.* (2008), and Banachewicz et al. (2007). Wong and Li (2001) have considered a two-state non-homogeneous Markov switching mixture autoregressive model. All the above papers consider a small number of states (2 or 3 states) and they adopt classical inferential procedures. A Bayesian approach to inference for a two-state non-homogeneous HMM has been proposed by Filardo and Gordon (1998) who adopted a binary regression probit model.

We consider Bayesian inference for a general m-state non-homogeneous HMM, assuming

2

that the elements of the transition matrix $P^{(t)}$ are linked with X_t through a multinomial logistic link. We construct a Markov chain Monte Carlo (MCMC) algorithm and provide detailed implementation guidelines when Y_t, conditional on Z_t, follows an autoregressive process of order p or is linearly regressed on W_t. Although we deal with Gaussian likelihoods, other time series formulations such as discrete-valued time series or spatial autoregressive processes possess no further difficulties, except perhaps the extra care needed for MCMC mixing in some complicated model formulations, since given the unobserved process Z_t the model collapses to well-known MCMC algorithms; see, for example, Spezia (2009). Our proposed models are useful primarily for forecasting rather than being summaries for explaining the data generation mechanism. Therefore, an issue of particular interest incorporated in our MCMC scheme is that we accommodate and exploit model uncertainty within our Bayesian modelling to improve, via model averaging, the predictive ability of the proposed models.

The MCMC algorithm we propose has the following key points. First, conditional on $\{Z_t\}$ the parameters that specify the formulation of the stochastic process $\{Y_t\}$ are, in general, easy to sample from. Second, conditional on $\{Y_t\}$ and on all transition matrices $\{P^{(t)}\}$, a forward-backward filtering algorithm is used to simulate the hidden sequence of states $\{Z_t\}$. Third, the regression coefficients of X_t are obtained via Gibbs steps which utilize the auxiliary variable sampling method of Holmes and Held (2006). Finally, a reversible jump MCMC step is used to simultaneously sample in the product of model and parameter space.

We apply our methodology to a monthly interest rates return series in which it is commonplace that two-state HMMs, or regime-switching models, are appropriate in capturing the stylized facts of the data. We find that non-homogeneous HMMs, using a set of six covariates that affect the transition matrix of the hidden process improve the predictive ability of the model when compared against a standard homogeneous HMM. These results indicate that our proposed modelling strategy may offer important tools in forecasting using HMMs.

The paper proceeds as follows. In section 2 we introduce the discrete-time finite state-space HMM with time varying transition matrix. In section 3 we present Bayesian inference for this model, while in section 4 we consider inference under model uncertainty as well as forecasting using Bayesian model averaging. In section 5 we present results from the application of our approach to US interest rates returns. Finally, section 6

concludes the paper.

2 The Model

Consider a random process $\{Y_t\}$ and suppose that we wish to use covariates X_t and W_t for forecasting purposes. Let y_t, x_{it}, w_{jt} be the corresponding realisations at time t for the process $\{Y_t\}$ and the i-th covariate in X_t and j-th covariate in W_t, $i = 1, \ldots, k$, $j = 1, \ldots, n$. We model the dynamics of $\{Y_t\}$ through a hidden Markov model (HMM) with time-varying transition matrix. Specifically, we assume that there exists a hidden underlying process $\{Z_t\}$ which is a non-homogeneous discrete-time Markov chain on a finite state-space $S = \{1, \ldots, m\}$ with transition probability matrix at time t denoted by $P^{(t)} = [p_{ij}^{(t)}]$. The latent variables $Z_1, \ldots, Z_T$ represent the hidden states of some underlying mechanism that has generated the observed data $y_1, \ldots, y_T$. For $Z_t = s$, $s \in S$, we assume that the distribution of Y_t is given by

$$Y_t | Z_t = s \sim N(\mu_s + w_t b_s, \sigma_s^2), \tag{1}$$

where $w_t = (w_{1t}, \ldots, w_{kt})$, b_s is a $n \times 1$ vector of coefficients and σ_s^2 is the state specific variance of the Gaussian process.

The time-varying transition matrix $P^{(t)}$ of the hidden process is modelled as a function of the observed covariates, $x_{1t}, \ldots, x_{kt}$, at time t. Time-varying transition probabilities incorporate covariate information to describe the transition dynamics of the underlying hidden process. Thus, we allow the covariates to affect the observed process $\{Y_t\}$ in a non-linear fashion. We use the multinomial function to link X_t and the entries of $P^{(t)}$. Specifically, the ij-th entry of $P^{(t)}$ is assumed to depend on $x_t = (1, x_{1t}, \ldots, x_{kt})$ via

$$p_{ij}^{(t)} = \frac{\exp(x_t \beta_{ij})}{\sum_{\ell=1}^{m} \exp(x_t \beta_{i\ell})}, \text{ for } i, j = 1, \ldots, m, \tag{2}$$

where β_{ij} is a $(k+1) \times 1$ vector of regression coefficients. For identifiability reasons, in each row of the transition matrix we restrict one of the β_{ij}'s to be equal to a $(k+1) \times 1$ vector of zeros. For example, for $m = 2$ the above modelling approach reduces to adopting a logistic link function to model two transition probabilities. A similar non-homogeneous HMM has been formulated by Banachewicz et al. (2007) for modelling portfolio defaults.

4

3   Bayesian Inference

We consider Bayesian inference on the parameters of our non-homogeneous HMM by constructing a Markov chain Monte Carlo algorithm which updates, in turn, the multinomial regression coefficients $\beta = \{\beta_{ij}; i, j = 1, \ldots, m\}$ in (2), the state specific model parameters $\theta = \{\mu_s, b_s, \sigma_s^2; s = 1, \ldots, m\}$ in (1), and the latent variables $Z^T = (Z_1, \ldots, Z_T)$. In detail, this algorithm is described as follows.

Let $Y^t = (Y_1, \ldots, Y_t)$ be the history of the observed process, $Z^t = (Z_1, \ldots, Z_t)$ the sequence of states up to time t, and let $f_s(.)$ denote the normal probability density function of $Y_t \mid Z_t = s$, $s \in S$, as given in (1). The joint likelihood function of the observed data, y^T, and the unobserved sequence of states, z^T, is given by

$$\pi(y^T, z^T \mid \theta, \beta) = \pi(y^T \mid z^T, \theta, \beta)\pi(z^T \mid \theta, \beta) = f_{z_1}(y_1)p_{z_1,z_2}^{(1)} f_{z_2}(y_2) \ldots p_{z_{T-1},z_T}^{(T-1)} f_{z_T}(y_T)$$

where the entries of the transition matrix are linked to parameters β via (2). We assume that the chain is equally likely to start off at any state, which amounts to assuming a uniform distribution on the initial state.

If a prior distribution $\pi(\theta, \beta) = \pi(\theta)\pi(\beta)$ is specified for the model parameters, then inference on all the unknown quantities in the model is based on their joint posterior distribution

$$\pi(\theta, \beta, z^T \mid y^T) \propto \pi(\theta, \beta)\pi(y^T, z^T \mid \theta, \beta).$$

The prior specification we consider is the following. For the multinomial regression coefficients β_{ij} we assume independent multivariate normal prior distributions $N(\xi, v)$, restricting in each row of the transition matrix one of the β_{ij}'s to be a vector of zeros. For the state specific parameters μ_s, b_s, σ_s^2 we use conjugate prior distributions $\sigma_s^2 \sim IG(p, q)$ and $(\mu_s, b_s) \mid \sigma_s^2 \sim N(\tilde{\xi}, \sigma_s^2\Omega)$, where IG denotes the inverted Gamma distribution.

The MCMC sampling scheme is constructed by successively updating the latent variables z^T given the current value of the model parameters θ, β by using the Forward-backward algorithm outlined in Subsection 3.1; the multinomial regression coefficients β given the sequence of states z^T by adopting the auxiliary variable method of Holmes and Held (2006) discussed in Subsection 3.2 (we note here that the data-augmented Metropolis-Hastings sampling scheme of Scott (2009) could be used instead); and the state specific parameters θ conditional on z^T by using well-known conjugate Gibbs sampling steps the details of which are omitted; see, for example, O'Hagan and Forster, 2004).

3.1 Simulating the Hidden States

The Forward-backward algorithm (Baum et al., 1970, Scott, 2002) consists of a set of filtering recursions that can be used for calculating the likelihood and simulating realisations of the hidden underlying process of a partially observed Markovian model given the values of the model parameters. It can be constructed either to implement the EM algorithm by maximizing the likelihood function, as for example in Banachewicz et al. (2007), or within the context of an MCMC scheme for Bayesian inference as it is applied here, see Carter and Kohn (1993), Frühwirth-Schnatter (1994, 2006).

The forward recursion for the evaluation of the likelihood is constructed as follows. For time t and state s define the forward variables $a_t(s)$ as the joint probability of the data up to time t and the state of the hidden process at time t given the values of the parameters θ, β:

$$a_t(s) = \Pr(y^t, Z_t = s \mid \theta, \beta) = \Pr(Z_t = s) \Pr(y^t \mid Z_t = s, \theta, \beta),$$

with $a_1(s) \propto f_s(y_1)$. The forward variables can be calculated recursively via the relationship

$$a_t(s) = \sum_{j=1}^{m} \Pr(y^t, Z_{t-1} = j, Z_t = s \mid \theta, \beta) = \left[\sum_{j=1}^{m} a_{t-1}(j) p_{js}^{(t-1)} \right] f_s(y_t).$$

The final step of the recursion allows the calculation of $a_T(s) = \Pr(y^T, Z_T = s \mid \theta, \beta)$, $s \in S$, and, therefore, the likelihood $L(y^T \mid \theta, \beta)$ can be obtained as

$$L(y^T \mid \theta, \beta) = \sum_{s=1}^{m} a_T(s).$$

The calculation of the likelihood function via the forward recursion requires $O(m^2 T)$ steps, instead of the $O(m^T)$ steps needed for direct evaluation of the likelihood. This reduces the computational complexity of the problem substantially.

In order to simulate a realisation $z_1, \ldots, z_T$ from the joint distribution of the hidden state variables $Z_1, \ldots, Z_T$ a second -backward- recursion is constructed. The state at time T is first simulated from

$$\Pr(Z_T = s \mid y^T) = \frac{\Pr(y^T, Z_T = s)}{L(y^T)} = \frac{a_T(s)}{\sum_{j=1}^{m} a_T(j)}, \quad s = 1, \ldots, m.$$

Then, for $t = T - 1, \ldots, 1$, the state at time t given the state at time $t + 1$ is simulated

from

$$\begin{aligned}
\Pr(Z_t = s \mid y^T, z_{t+1}) &= \frac{\Pr(Z_t = s \mid y^t)\Pr(z_t \mid Z_{t+1} = s)}{\sum_{j=1}^m \Pr(Z_t = j \mid y^t)\Pr(z_t \mid Z_{t+1} = j)} \\
&= \frac{a_t(s)p_{sz_{t+1}}^{(t)}}{\sum_{j=1}^m a_t(j)p_{jz_{t+1}}^{(t)}}, \quad s = 1, \ldots, m,
\end{aligned}$$

where the forward variables $a_t(s)$ have been calculated during the forward step of the algorithm. Note that had we assumed that the chain is not slow mixing we could have gained some efficiency by performing several forward recursions.

3.2   Simulating the Multinomial Regression Coefficients

We will first consider the simple case $m = 2$. Given the latent data z^T, we choose to model the two diagonal elements of the transition matrix by linking them to the set of covariates with a logistic link function g. First define

$$\tilde{Z}_t^s = \mathbb{1}[Z_{t+1} = Z_t = s]$$

with $\mathbb{1}[.]$ denoting the indicator function, and then model $\tilde{Z}_t^s$, $s = 1, 2$, as

$$\begin{aligned}
\Pr(\tilde{Z}_t^s = 1) &= \Pr(u_t > 0) \\
u_t &= x_t\beta_{ss} + e_t \\
e_t &\sim logistic(0, 1)
\end{aligned}$$

The logistic regression coefficients are updated using auxiliary variable sampling as follows. For given s, the above model has the auxiliary variable representation

$$\begin{aligned}
\Pr(\tilde{Z}_t^s = 1) &= \Pr(u_t > 0) \\
u_t &= x_t\beta_{ss} + e_t \\
e_t &\sim N(0, \lambda_t) \\
\lambda_t &= (2\psi_t)^2 \\
\psi_t &\sim KS
\end{aligned}$$

where the KS denotes the Kolmogorov-Smirnov distribution. In this model specification, u_t follows a scale mixture of normal densities with a marginal logistic distribution (Andrews and Mallows, 1974). This representation allows for sampling the logistic regression coefficients via iterative Gibbs steps from the conditional posterior distributions of β_{ss}, u_t and λ_t.

Let X_s denote the design matrix consisting of the covariate values at times t in which $z_t = s$ and let $\tilde{Z}_s$ and λ_s be the corresponding vectors of $\tilde{Z}_t^s$ and λ_t. Using a multivariate normal $N(\xi, v)$ prior for β_{ss}, the full conditional posterior distribution of β_{ss} given $\tilde{Z}_s$ and λ_s is $N(\Xi_s, V_s)$, where $V_s = (v^{-1} + X_s' B_s^{-1} X_s)^{-1}$, $\Xi_s = V_s(v^{-1}\xi + X_s' B_s^{-1} \tilde{Z}_s)$ and $B_s = diag(\lambda_s)$. The posterior densities of the u_t's, after marginalising out the λ_t's, are independent truncated logistic distributions, while the latent variables λ_t are simulated from their full conditional posterior distributions via Gibbs steps based on rejection sampling; see Holmes and Held (2006).

The above approach can be generalised to the multinomial regression case. For the $m > 2$ state HMM the likelihood for the regression coefficients used to model the s-th row of the transition matrix corresponds to a multinomial distribution with m classes. The likelihood for β_{sj}, conditional on all other coefficients of the s-th row, is proportional to

$$\prod_{z_t = s} \left[\frac{\exp(x_t \beta_{sj} - C_{sj}^{(t)})}{1 + \exp(x_t \beta_{sj} - C_{sj}^{(t)})} \right]^{\mathbb{1}[z_{t+1}=j]} \left[\frac{1}{1 + \exp(x_t \beta_{sj} - C_{sj}^{(t)})} \right]^{\mathbb{1}[z_{t+1} \neq j]}, \quad (3)$$

where $C_{ij}^{(t)} = \log \sum_{s \neq j} \exp(x_t \beta_{sj})$. As Holmes and Held (2006) mention, this conditional likelihood has the form of a logistic regression on class indicator $\mathbb{1}[z_{t+1} = j]$ and therefore a generalisation of the above logistic sampling scheme can be used in this setting.

3.3 Label Switching

An important problem in Bayesian analysis of HMM is the non identifiability of the hidden states. This problem occurs when exchangeable priors are used for the state specific parameters, which is common practice if there is no prior information about the hidden states. In these cases, the posterior distribution is invariant to permutations of the state labels and, hence, the marginal posterior distributions of the state specific parameters are identical for all states. Therefore, direct inferences about the state specific parameters are not available from the MCMC output.

Various approaches to dealing with the label switching problem in finite mixture models have been proposed in the literature; see Jasra *et al.* (2003) for a recent review. Recently, Spezia (2008) considered the label switching problem in HMMs with unknown number of states. In the present paper we focus our attention on predicting future observations of economic series that can be modelled using some HMM and on examining whether the predictive ability of the HMM is improved if the transition matrix is

assumed to vary in time. Therefore, we consider applications for which the label switching problem does not matter, since prediction is based on parameter functions that are invariant of label switching rather than on the parameters directly. Fortunately, in the application in this paper we did not face the label switching problem and therefore were able to produce inferences on the model parameters. In general, we propose using some technique of post-processing the MCMC output in order to be able to infer the state specific parameters.

4   Variable Selection

We extend our methodology to accommodate model uncertainty, regarding the set of covariates to be included in the model, through reversible jump MCMC. Since forecasting is our primary focus, it is natural not to focus on the choice of single subset of variables, but use a model averaging approach which provides composite predictions. For evidence that accounting for model uncertainty improves predictive performance see, for example, Hoeting *et al.*(1999).

4.1   Reversible jump MCMC

Suppose that there are K available covariates that can be included in the multinomial regression model (2). For simplicity of exposition we assume that the same subset of covariates are used to model all the entries of the transition probability matrix. This results in 2^K possible models. Denote by M one of these models and β_M the corresponding vector of multinomial regression coefficients. We use a non-informative discrete uniform prior density $\pi(M) = 2^{-K}$ on the model space. The reversible jump MCMC strategy, introduced by Green (1995), can be used for generating from the joint posterior density $\pi\left(M, \beta_M, \theta \mid y^T\right)$, based on the standard Metropolis Hastings approach. During reversible jump MCMC sampling, the constructed Markov chain moves within and between models so that the limiting proportion of visits to a given model is the required marginal posterior density $\pi\left(M \mid y^T\right)$.

The algorithm works as follows. Suppose that the current state of the Markov chain is (M, β_M), where β_M has dimension $d(\beta_M)$, and a move is proposed to a new model $M^{'}$ with probability $j\left(M, M^{'}\right)$ and corresponding parameter vector $\beta_{M'}$. Then, a vector u is generated from a specified proposal density $q\left(u \mid \beta_M, M, M^{'}\right)$ and we set $\left(\beta_{M'}, u^{'}\right) =$

$g_{M,M'}(\beta_M, u)$ for a specified invertible function $g_{M,M'}$ such that $g_{M',M} = g_{M,M'}^{-1}$. Note that $d(\beta_M) + d(u) = d(\beta_{M'}) + d(u')$. Green (1995) showed that if the new move is accepted as the next realization of the Markov chain with probability $a = \min\{1, A\}$ where

$$A = \frac{L\left(y^T \mid M', \beta_{M'}, \theta\right) \pi\left(\beta_{M'} \mid M'\right) \pi\left(M'\right) j\left(M', M\right) q\left(u' \mid \beta_{M'}, M', M\right)}{L\left(y^T \mid M, \beta_M, \theta\right) \pi\left(\beta_M \mid M\right) \pi\left(M\right) j\left(M, M'\right) q\left(u \mid \beta_M, M, M'\right)} |J|, \tag{4}$$

with $J = \partial\left(\beta_{M'}, u'\right) / \partial(\beta_M, u)$ denoting the Jacobian of the transformation, then the chain satisfies the condition of detailed balance and has the required limiting distribution $\pi(M, \beta_M, \theta \mid y^T)$.

We implement the above algorithm as follows. First, we choose to jump only to 'neighboring' models which are obtained by removing or adding a randomly chosen covariate and setting the corresponding jumping probabilities $j\left(M, M'\right)$ to be equal for every possible proposed model M'. Thus, in every iteration, we add a fourth step in the MCMC algorithm of Section 3, at which we choose (when this is possible), to add or to remove a covariate with equal probability and then we randomly choose which covariate will be added or removed from the set of available covariates. For example, if there are currently $k < K - 1$ covariates in the model, we propose to add a covariate with probability $1/2$ and select a covariate to add from those not included in the model, each with probability $1/(K - k)$. The probability of proposing the reverse move is $1/2(k + 1)$. Hence, in this case, $j\left(M', M\right) / j\left(M, M'\right) = (K - k)/(k + 1)$. When we add or remove a covariate, the proposal distribution $q(u \mid \beta_M, M, M')$ is just the full conditional density of the multinomial regression coefficients in M', $\beta_{M'}$. That is the multinomial regression coefficients are jointly updated with the model, while all other parameters remain unchanged. Therefore, $\beta_{M'} = u$, and $g_{M,M'}(\beta_M, \beta_{M'}) = (\beta_{M'}, \beta_M)$, so the Jacobian term in (4) is one. As an example, in the case of a two-state HMM, $q(u \mid \beta_M, M, M') = q(u \mid M')$ is a product of two multivariate normal distributions, $N(\Xi_{M',s}, V_{M',s})$, $s = 1, 2$, where the parameters $\Xi_{M',s}, V_{M',s}$ are defined as in Section 3.2 and depend on the proposed model M'. The above choices do not affect the results but may be crucial for the convergence rate of the Markov chain. In our case the performance was immediately excellent, with probability of acceptance at around .75, so we did not experiment with further choices.

We note that our reversible jump algorithm is implemented by choosing to jump only to neighboring models using proposal distributions such that the Jacobian term is equal

to one. Therefore, our reversible jump step reduces to a standard Metropolis-Hastings step similar to that of a proposed binary indicator update as in Holmes and Held (2006). However, a more complex model specification for the transition matrix can be easily incorporated within this algorithm by allowing for different subsets of covariates to affect the transition probabilities. This can be achieved by simply modifying the proposal distributions for moving among different models.

4.2 Forecasting with Bayesian Model Averaging

Since forecasting of y_{t+1} is of primary interest, and having been able to calculate the posterior probabilities of each model via the reversible jump MCMC algorithm, it seems natural to account for this model uncertainty in our predictive inferences. Rather than choosing a single "best" model and then make inferences as if the selected model was the true model, we can use a model averaging approach which provides composite predictions. In the real data applications of the next section we focus our attention at examining whether covariate information can be useful in modelling the transition matrix of a HMM, so different models correspond to all subsets of covariates X_t.

Under a discrete-time finite state-space HMM one-step-ahead predictions are straightforward. Given the model specification, M say, the predictive distribution of the next observation y_{T+1}, is given by

$$f(y_{T+1}|y^T, M) = \int f\left(y_{T+1}|y^T, z^T, M, \beta_M, \theta\right) \pi(\beta_M)\pi(\theta)d\beta_M d\theta \tag{5}$$

where

$$f\left(y_{T+1}|y^T, z^T, M, \beta_M, \theta\right) = \sum_{s=1}^{m} \Pr(Z_{T+1} = s \mid Z_T = z_T)f_s(y_{T+1}).$$

The multidimentional integration required in (5) can not be performed analytically, but it can instead be evaluated numerically by exploiting the samples from the posterior distribution obtained from the reversible jump MCMC algorithm. Assume that R_M values $z^{T(r)}, \beta_M^{(r)}, \theta^{(r)}, \ r = 1, \ldots, R_M$ are available from model M at the end of the MCMC run; then, an approximation of $f(y_{T+1}|y^T, M)$ is given by

$$\hat{f}(y_{T+1}|y^T, M) = R_M^{-1} \sum_{r=1}^{R_M} f\left(y_{T+1}|y^T, z^{T(r)}, M, \beta_M^{(r)}, \theta^{(r)}\right) \tag{6}$$

If model uncertainty is taken into account, a Bayesian model averaging estimate of

$f(y_{T+1} \mid y^T)$ is computed by

$$f(y^{T+1}|y^T) = \sum_M f(y^{T+1}|y^T, M) \Pr(M|y^T). \tag{7}$$

In practice, a sample from $f(y_{T+1} \mid y^T)$ can be obtained as follows. At the rth iteration of the algorithm, after all the unknown quantities in the model have been simulated, a value $z_{T+1}^{(r)}$ can be simulated first from the discrete distribution with probabilities $\Pr(Z_{T+1} = s \mid Z_T = z_T^{(r)})$, $s = 1, \ldots, m$, and then, conditional on the value $z_{T+1}^{(r)}$, $y_{T+1}^{(r)}$ can be drawn from $f_{z_{T+1}^{(r)}}(y_{T+1})$. Then, inferences on the future observation Y_{T+1} can be based on the sample $y_{T+1}^{(1)}, \ldots, y_{T+1}^{(R)}$.

The above procedure can be generalised for forecasting when further observations y_{T+2}, y_{T+3}, ... arrive, without the need to re-run the MCMC algorithm. Assume that for some integer $\ell > 1$, at time $T + \ell - 1$ the data $y^{T+\ell-1}$ and $x^{T+\ell-1}$ have been observed. Then the simulation of the $y_{T+\ell}$ future observation is performed as follows. First we draw a value $z_{T+\ell}^{(r)}$ from a discrete distribution with probabilities $\Pr(Z_{T+\ell} = s \mid Z_{T+\ell-1} = z_{T+\ell-1}^{(r)})f_{z_{T+\ell-1}^{(r)}}(y_{T+\ell-1})$, $s = 1, \ldots, m$ and then we draw $y_{T+\ell}^{(r)}$ from $f_{z_{T+\ell}^{(r)}}(y_{T+\ell})$.

A further issue that is necessary to complete our methodological contribution is the evaluation of the predictive ability of our proposed models by comparing it to that of the standard homogeneous HMM. In the applications section that follows we conduct an out-of-sample exercise based on the logarithmic scoring rule; see, for example, Gneiting and Raftery (2007). Under each model, we calculate the conditional predictive ordinate so models which assign large predictive probability to the values that actually occur return a large log score. For example, for a model M the predictive log score is calculated as $\sum_{\ell=1}^L \log \hat{f}(y_{T+\ell}|y^{T+\ell-1}, M)$, where $\hat{f}(y_{T+\ell})$ is calculated in the same manner as in (6). When inferences are based on Bayesian model averaging, the predictive log score is given by $\sum_{\ell=1}^L \log \hat{f}(y_{T+\ell}|y^{T+\ell-1})$, where

$$\hat{f}(y_{T+\ell}|y^{T+\ell-1}) = R^{-1} \sum_{r=1}^R f\left(y_{T+\ell}|y^{T+\ell-1}, z^{T+\ell-1(r)}, M^{(r)}, \beta_{M^{(r)}}^{(r)}, \theta^{(r)}\right)$$

is the estimate of the predictive density based on a sample of R MCMC draws.

We note that our reversible jump algorithm does not account for model uncertainty regarding the number of states. For homogeneous HMMs such algorithms have been proposed in the literature, see Robert *et. al* (2000) and Cappe *et. al* (2003). Unfortunately these approaches contain moves between models that are rather complicated and extensions in our setup is extremely difficult. Nevertheless, since the focus of our

approach is prediction rather than parameter inference, we suggest choosing m based on the predictive ability of the respective models. On the other hand, model choice regarding the set of covariates which affect directly the mean of the Gaussian process can be easily incorporated within our algorithm. This would simply involve adding steps that move between models allowing for different covariates or AR effects in the mean of the analysed series.

5   Application to US Interest Rates

We illustrate our proposed modelling and inferential approach using the monthly US 3-month treasury bill rates from January 1962 until December 1999. For this data we use a two-state non-homogeneous HMM with six economic variables as covariates. We make predictions both conditional on a given covariate set and following a BMA approach. The plot of the analysed series is shown in Figure 1. These data has been previously analysed by Pfann *et al.* (1996) and Dellaportas *et al.* (2007).

It is commonplace that such time series are best modelled by incorporating some sort of HMM, since regime switching structures are well suited to capture non-linearities in interest rates. Ang and Bekaert (2002) have shown that regime-switching models correspond well with business cycles in the US, and have better forecasting ability than single regime models. Dellaportas *et al.* (2007) have identified different regimes in the US 3-month treasury bill rates, also considered here, corresponding to clusters of data points with high or low volatility. Furthermore, they have advocated adopting a nonlinear model by using six economic variables to explain the episodes of mean-reverting tendencies that appear in interest rate levels. The six series that were used as covariates are the US annual inflation rate (INF), a trade-weighted measure of the US dollar against other major currencies (CUR), the US producer price index (PPI),the national association of purchasing management index (NAPM), the US consumer price index (CPI) and the 10 year US treasury yield (LTR). All the series of explanatory variables were differenced before the analysis apart from LTR and the data has been retrieved from DATASTREAM.

In our study we use the above variables to model the transition matrix of a non-homogeneous HMM. The non-homogeneous HMM we use to analyse this data is of

the form

$$\Delta y_t = \mu_{z_t} + b_{z_t} y_{t-1} + \sigma_{z_t} \epsilon_t, \quad \epsilon_t \sim N(0,1),$$

where $\Delta y_t = y_t - y_{t-1}$ and Z_t follows a 2-state Markov chain with transition matrix $P^{(t)}$ which is modelled as described in Section 2 using the six covariates mentioned above.

We assign non-informative prior specifications to β_{ss}, μ_s, b_s, $s = 1, 2$, in the form of a product of univariate $N(0, 10\sigma_s^2)$ normal densities, and to σ_s^2, $s = 1, 2$, in the form of an inverted-Gamma $IG(c, d)$ density with $c = d = 0.01$. Finally, for the reversible jump MCMC, each model is assigned equal prior probability. For our comparisons, we also used the respective homogeneous HMM, for which the constant transition probabilities p_{ss}, $s = 1, 2$ are assigned independent uniform densities in $(0, 1)$. Inferences for all models are based on MCMC samples consisting of 20000 draws with a burn-in period of 5000 iterations.

The most probable model, with posterior probability of 0.165, is the model containing all variables. The posterior model probabilities of the other models range between 0.001 for the model including just CUR and 0.096 for the model including all covariates except for CPI.

Our out-of-sample evaluation exercise is designed as follows. We use the data from January 1962 to December 1997 for inference (estimation period) and we consider the period from January 1998 to December 1999 as our two-year out-of-sample evaluation period. The data in the evaluation period are predicted under each model using the model parameters obtained from the estimation data within an MCMC sampling scheme.

We consider two non-homogeneous HMM for forecasting; the saturated model with all six covariates and the Bayesian model averaging approach. The BMA approach with predictive log score of 8.66 outperforms both the standard homogeneous HMM which obtained predictive log score of 4.35 and the non-homogeneous saturated HMM which obtained a predictive log score of 8.39.

The difference between predictive log scores of two models, respectively denoted by $\widehat{pls}_1$ and $\widehat{pls}_2$, can be interpreted on a 'per month' basis as an improvement in the predictive performance by a factor of $\exp\left\{(\widehat{pls}_1 - \widehat{pls}_2)/24\right\}$. For example, the difference in log score between the Bayesian model averaging approach and the homogeneous HMM represents an improvement in the predictive performance by a factor of

$\exp\{(8.66 - 4.35)/24\} = 1.20$ or by about 20%.

We now restrict our attention to the most probable non-homogeneous HMM, the saturated model, examining how the different covariates considered affect the transition probabilities. As mentioned earlier, this involves solving the label switching problem. In this application, though, our algorithm did not present any label switching. In terms of MCMC performance, this fact is clearly an indication of poor mixing. However, although we realize the possible dangers of such a problem (see, for example, Marin *et al.*, 2005), we believe that we have a clear situation in which the modes are well separated by an area with effectively zero probability and that the single mode is being explored thoroughly. A close inspection of the MCMC output revealed that the two states of the hidden Markov chain correspond to periods of high and low volatility, respectively. Figure 2 shows the plot of the analysed series of first differences of US interest rates together with the hidden Markov chain inferred by our MCMC algorithm. It is evident from this plot that one of the underlying states corresponds to periods of high volatility (state 1) and the other to periods of low volatility (state 2). To investigate the sources of the non-homogeneity of this hidden Markov chain, we examined how the probabilities p_{11} and p_{22}, vary with the six different covariates.

In Figure 3 we show plots of p_{11} and p_{22} calculated for a grid of values for each covariate ranging from the minimum to the maximum value observed in the sample. In each case the remaining 5 covariates were set equal to their mean values. It seems that, in general p_{22} is higher than p_{11}. Both probabilities decrease as the value of US annual inflation rate (INF) increases, while they both increase as the value of the trade-weighted measure of the US dollar against other major currencies (CUR) increases. The remaining four covariates affect p_{11} and p_{22} in opposite ways. For example, p_{11} decreases with national association of purchasing management index (NAPM), while p_{22} increases.

Finally, in Figure 4 we show plots of $p_{11}^{(t)}$ and $p_{22}^{(t)}$ as they vary with time. It is evident that $p_{22}^{(t)}$ is higher than $p_{11}^{(t)}$, implying that the chain is more probable to remain in the low volatility than in the high volatility state, while it is more probable to move from the high volatility state to the low volatility one than vice-versa since $1 - p_{11}^{(t)} > 1 - p_{22}^{(t)}$. Moreover, as expected, $p_{11}^{(t)}$ is lower in periods of low volatility and gets higher in periods of high volatility. Similar results are obtained for $p_{22}^{(t)}$.

6 Conclusions

In this paper we have considered Bayesian predictive inference for discrete-time finite state-space HMMs with transition matrix that varies in time depending on a set of covariates. We have constructed a reversible jump MCMC algorithm for inference on the model parameters which includes drawing the hidden sequence of states using the Forward-backward algorithm, simulating the multinomial regression coefficients using auxiliary variable methods drawing the state specific parameters using standard Gibbs steps, and sampling from the product space of models and parameters by using reversible jump MCMC.

The proposed approach to modelling and inference has been used for the prediction of future observations of economic series. The advantage of our model is that it allows the transition probabilities of the HMM to vary in time by exploiting covariate information. Our model assumes that the covariates affect some economic variable of interest in a non-linear fashion, while other exogenous covariates or lagged values of the economic variable can be used to model the variable of interest directly. This can be a reasonable assumption in the analysis of many economic series that has been entirely ignored in the forecasting literature. Our empirical results indicate that for the dataset we analysed, our proposed model strategy has better predictive ability than a standard homogeneous HMMs.

References

Andrews, D. and Mallows, C. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B*, **36**, 99-102.

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

Banachewicz, K., Lucas, A. and Vaart, A. (2007). Modeling Portfolio Defaults Using Hidden Markov Models with Covariates. *Econometrics Journal*, **10**, 1-18.

Baum, L. E., Petrie, T., Soules, G. and Weiss, N. (1970). A maximisation technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, **41**, 164-171.

Billio, M., Monfort, A., and Robert, C.P. (1999). Bayesian estimation of switching ARMA models. *Journal of Econometrics*, **93**, 229-55.

Cappé, O., Moulines, E. and Ryden, T. (2005). *Inference in Hidden Markov Models.* Springer.

Cappé, O., Robert, C. P. and Ryden, T. (2003). Reversible jump, birth-and-death and more general continuous time Markov chain Monte Carlo samplers. *Journal of the Royal Statistical Society, Series B*, **65**, 679-700.

Carter, C.K. and Kohn, R. (1994) On Gibbs sampling for state space models. *Biometrika*, **81**, 541-553.

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

Chib, S. (1998). Estimation and Comparison of Multiple Change Point Models, *Journal of Econometrics*, **86**, 221241.

Dellaportas, P., Denison, D. G. T and Holmes, C. C. (2007). Flexible Threshold Models for Modelling Interest Rate Volatility. *Econometric Reviews*, **26**, 419 - 437.

Diebold, F. X., Lee, J. and Weinbach, G. C. (1994). Regime switching with time-varying transition probabilities, In C. P. Hargreaves (Ed.), *Non-stationary Time Series Analysis and Cointegration*, pp 283-302. Oxford: Oxford University Press.

Durland, J. M. and McCurdy (1994). Duration-dependent transitions in a Markov model of U.S. GNP growth. *Journal of Business and Economic Statistics*, **12**, 279-288.

Filardo, A. J. and Gordon, S. F. (1998). Business cycle durations. *Journal of Econometrics*, **85**, 99-123.

Franses, P.H. and van Dijk, D. (2000). *Nonlinear Time Series Models in Empirical Finance.* Cambridge: Cambridge University Press.

Frühwirth-Schnatter, S. (1994). Data Augmentation and Dynamic Linear Models. *Journal of Time Series Analysis*, **15**, 183-202.

Frühwirth-Schnatter, S. (2001). Markov Chain Monte Carlo Estimation of Classical and Dynamic Switching and Mixture Models. *Journal of the American Statistical Association*, **96**, 194-209.

Frühwirth-Schnatter (2006). *Finite Mixture and Markov Switching Models*, Springer, New York.

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

Gneiting, T. and Raftery, A. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, **102**, 359-378.

Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, **42**, 27-62.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, **82**, 711-732.

Hamilton J. D. (1994). *Time Series Analysis.* Princeton: Princeton University Press.

Hoeting,J.A., Madigan,D., Raftery, A.E., and Volinsky, C.T. (1999). Bayesian Model Averaging: A Tutorial, *Statistical Science*, **14**, 382-401.

Holmes, C. C. and Held, L. (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis*, **1**, 145-168.

Jasra, A., Holmes, C. C. and Stephens, D. A. (2005). Markov chain Monte Carlo and the label switching problem in Bayesian mixture modelling. *Statistical Science*, **20**, 50-67.

Kim, C.J. and Nelson, C.R. (1999). *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications.* Cambridge: The MIT Press.

Kim, C., Piger, J. and Startz, R. (2008). Estimation of Markov regime-switching regression models with endogenous switching. *Journal of Econometrics*, **143**, 263-273.

Koop, G. (2003). *Bayesian Econometrics*, New York: John Wiley.

Krozlig, H-M. (1997). *Markov-Switching Vector Autoregression Modelling, Statistical Inference and Applications to Business Cycle Analysis.* Berlin: Springer.

MacDonald, I. L. and Zucchini, W. (1997). *Hidden Markov and other models for discrete-valued time series.* Chapman and Hall, New York.

Marin, J. M., Mengersen, K. and Robert, C. (2005). Bayesian modelling and inference on mixtures of distributions. In Rao C. R. and Dey D., editors, *Handbook of Statistics: Volume 25*, Springer-Verlag, New York.

Masson, P. and Ruge-Murcia, F. J. (2005). Explaining the transition between exchange rate regimes. *Scandinavian Journal of Economics*, **107**, 261-278.

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.

O'Hagan, A. and Forster, J. J. (2004). *Kendall's Advanced Theory of Statistics, volume 2B: Bayesian Inference*, Arnold, London.

Peria, M. S. M. (2002). A regime-switching approach to the study of speculative attacks: A focus on the EMS crisis. *Empirical Economics*, **27**, 299-334.

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

Pfann G. A., Schotman P.C. and Tschernig R. (1996) Nonlinear interest rate dynamics and implications for the term structure. *Journal of Econometrics*, **74**, 1549-176.

Robert, C. P., Ryden, T. and Titterington D. M. (2000). Bayesian inference in hidden Markov models through the reversible jump Markov chain Monte Carlo method. *Journal of the Royal Statistical Society, Series B*, **62**, 57-75.

Ryden, T. (2008). EM versus Markov chain Monte Carlo for estimation of hidden Markov models: A computational perspective. *Bayesian Analysis*, **3**, 659-688.

Schaller, H. and Van Norden, S. (1997). Regime switching in stock market returns. *Applied Financial Economics*, **7**, 177-191.

Scott, S. L. (2002). Bayesian methods for hidden Markov models: recursive computing in the 21st century. *Journal of the American Statistical Association*, **97**, 337-351.

Scott, S. L. (2009). Data augmentation and the Bayesian analysis of multinomial logit models. *Statistical papers*, (*forthcoming*).

Spezia, L. (2009). Reversible jump and the label switching problem in hidden Markov models. *Journal of Statistical Planning and Inference*, **139**, 2305-2315 .

Wong, C. S. and Li, W. K. (2001). On a logistic mixture autoregressive model. *Biometrika*, **88**, 833-846.
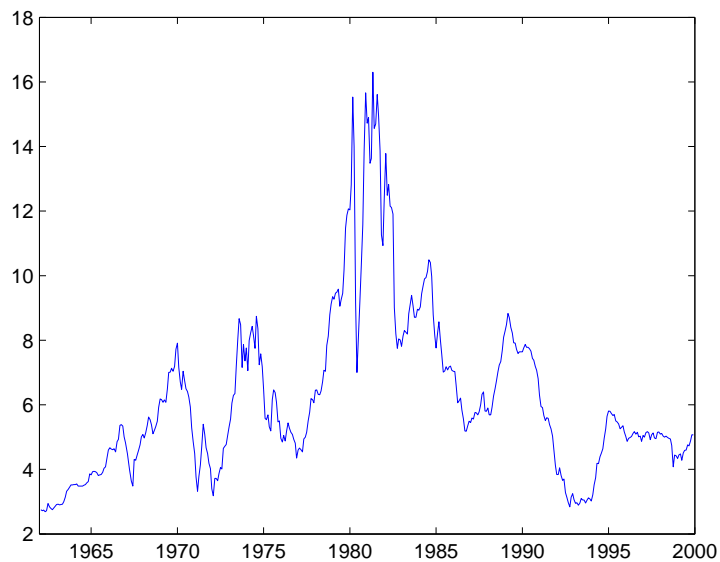
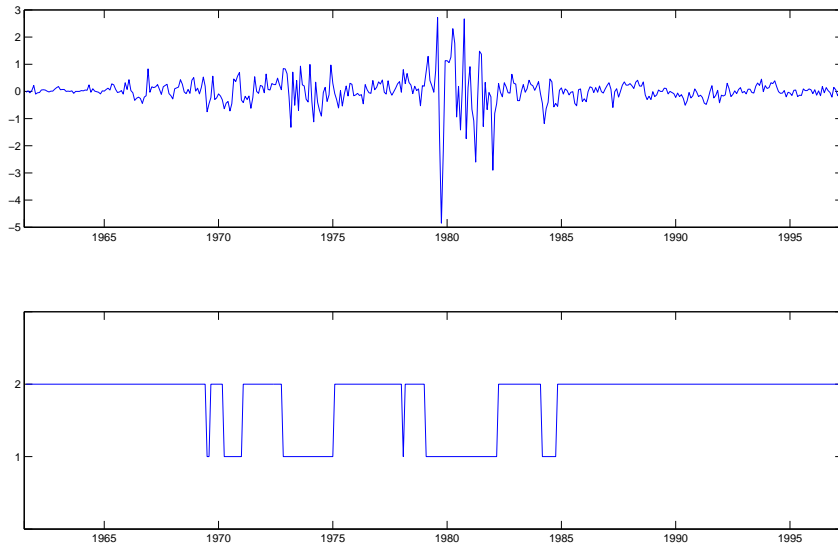Figure 1: The monthly US 3-month treasury bill rates for the period 1962 to 1999.

Figure 2: First differences of the monthly US 3-month treasury bill rates for the estimation period 1962 to 1997 (left axis) and sequence of states of the underlying Markov process (right axis).
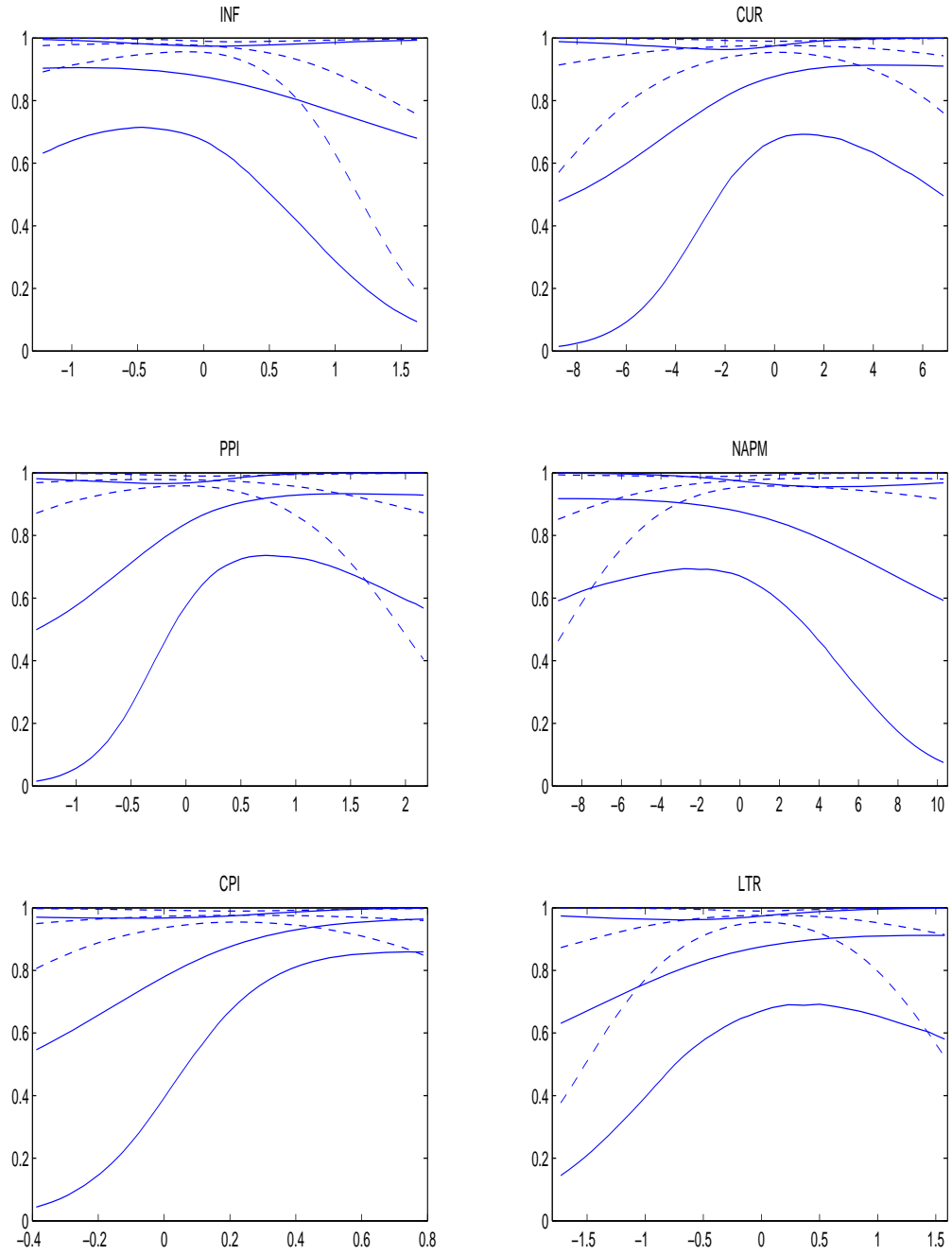
Figure 3: Means and 90% credibility intervals of how p_{11} (solid lines) and p_{22} (dashed lines) vary with different covariates.
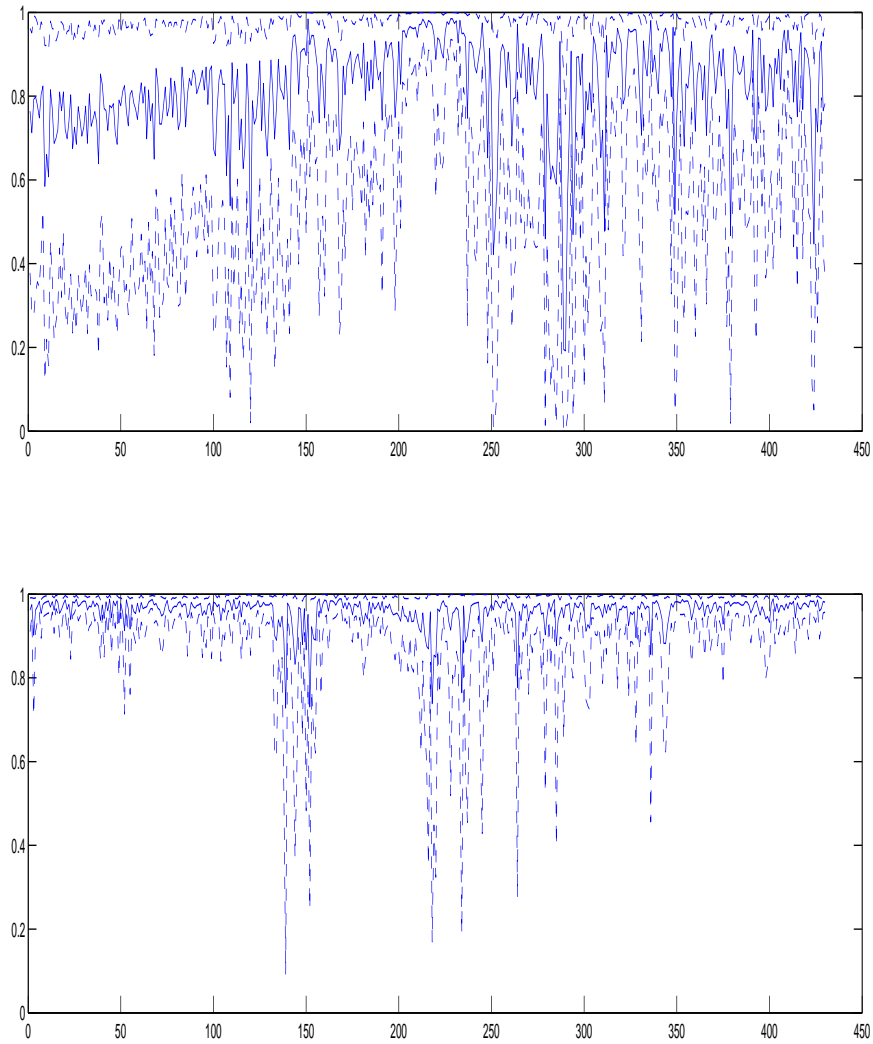
Figure 4: Plots of how p_{11} (top figure) and p_{22} (bottom figure) vary in time. Also shown are 90% credibility intervals, obtained from the MCMC output.