Bayesian analysis of predictive Non-Homogeneous hidden Markov models using Pólya-Gamma data augmentation[*]

Constandina Koki[† 1], Loukia Meligkotsidou[‡2], and Ioannis Vrontos[§1]

[1]*Department of Statistics, University of Economics and Bussiness*
[2]*Department of Mathematics, National and Kapodistrian University of Athens*

February 9, 2018

Abstract

We consider Non-Homogeneous Hidden Markov Models (NHHMMs) for forecasting univariate time series. We introduce two state NHHMMs where the time series are modeled via different predictive regression models for each state. Also, the time-varying transition probabilities depend on exogenous variables through a logistic function. In a hidden Markov setting, inference for logistic regression coefficients becomes complicated and in some cases impossible due to convergence issues. To address this problem, we use a new latent variable scheme, that utilizes the Pólya-Gamma class of distributions, introduced by Polson et al. (2013). Given an available set of predictors, we allow for model uncertainty regarding the predictors that affect the series both linearly – in the mean – and non-linearly – in the transition matrix. Predictor selection and inference on the model parameters are based on a MCMC scheme with reversible jump steps. Single-step and multiple-steps-ahead predictions are obtained based on the most probable model, median probability model or a Bayesian Model Averaging (BMA) approach. Simulation experiments, as well as an empirical study on real financial data, illustrate the performance of our algorithm in various setups, in terms of mixing properties, model selection and predictive ability.

Keywords: Non Homogeneous Hidden Markov Models; Model selection; Bayesian model averaging; Forecasting; Pólya-Gamma Data Augmentation, Continuous Ranked Probability Score

1 Introduction

This paper follows and extends the work of Meligkotsidou and Dellaportas (2011) by constructing a **broad class** of discrete-time, finite state-space, Non-Homogeneous Hidden Markov Models (NHHMMs) that can be used for modeling and predicting univariate time-series. We introduce a two-states NHHMM where the time series are modeled via different predictive regression models for each state, whereas the transition probabilities are modeled via logistic regressions. Given

1

an available set of predictors we allow for model uncertainty, regarding the predictors that affect the series both linearly, that is directly in the mean regressions, and non-linearly, that is in the transition matrix.

Discrete-time finite state-space Homogeneous Hidden Markov models (HHMMs) have been extensively studied and widely used to model stochastic processes consisting of an observed process and a latent (hidden) sequence of states which is assumed to affect the observation sequence, see for example Cappé et al. (2005) and Billio et al. (1999). We refer to Scott (2002) and Rydén (2008) for a review of a Bayesian approach of HHMMs. Bayesian inference, using Markov chain Monte Carlo (MCMC) techniques, has enhanced the applicability of HHMMs and has led to the construction of more complex model specifications including NHHMMs. Initially Diebold et al. (1994) studied the two-state Gaussian NHMMs where the time varying transition probabilities were modeled via logistic functions and their approach was based on the Expectation-Maximization algorithm (EM). Filardo and Gordon (1998) adopted a Bayesian perspective to overcome technical and calculation issues of classical approaches. Since then, various Bayesian methods have been proposed in the literature. For example Spezia (2006) modeled the time-varying transition probabilities via a logistic function depending on exogenous variables and performed model selection based on the Bayes factor. In the same spirit, Meligkotsidou and Dellaportas (2011) considered a m-stage NHMM and assumed that the elements of the transition matrix are linked through exogenous variables with a multinomial logistic link whereas the observed process conditional on the unobserved process follows an autoregressive process of order p.

This paper considers inference and variable selection for predictive NHHMMs which exploit a set of available predictors by allowing them to affect the transition probabilities and/or the state-specific predictive regressions. We perform Bayesian inference for the proposed models based on MCMC. Our MCMC scheme aims at overcoming difficulties and solving convergence issues arising with existing MCMC algorithms, thus offering an improved inferential procedure for estimation and variable selection in NHHMMs. We exploit the missing data representation of hidden Markov models and construct an MCMC algorithm based on data augmentation, consisting of several steps. First, the latent sequence of states is sampled via the Scaled Forward-Backward algorithm (Scott, 2002), which is a modification of the Forward-Backward algorithm of Baum et al. (1970) who used it to implement the classical EM algorithm. Then we simulate the parameters of the mean predictive regression model for each stage, via Gibbs sampling step and following Meligkotsidou and Dellaportas (2011) we use a logistic regression representation of the transition probabilities. Using the data augmentation scheme of Holmes and Held (2006) as Meligkotsidou and Dellaportas (2011), may result in serious convergence issues, especially in cases that there exists model uncertainty. Frühwirth-Schnatter and Frühwirth (2010) give a detailed comparison between various methods for dealing with binary and multinomial logit data and argue about the efficiency of Holmes and Held (2006) data augmentation scheme. Exploring various options, to deal with this problem we decided to use the Pólya-Gamma data augmentation scheme of Polson et al. (2013) which has a significantly improved performance. The recent work of Holsclaw et al. (2017) confirms that using Pólya-Gamma data augmentation to parametrize the transition probabilities, the algorithm mixes well and gives adequate estimates of the parameters. Finally, we account for model uncertainty by using the well-known model selection method of Green (1995), the reversible jump algorithm. We perform a couple of reversible jump steps within our algorithm since we allow for different covariates affecting the mean equation and the transition probabilities.

Regarding stochastic variable selection in the transition matrix of the NHHMM we are the first to perform a reversible jump algorithm using the Pólya-Gamma data augmentation representation for the logistic regression coefficients. Recently, Holsclaw et al. (2017) proposed a

model similar to ours for modeling multivariate meteorological time series data. In that paper, the hidden states are modeled via multinomial logistic regression affected by exogenous variables. The authors use the BIC criterion for choosing the best model among some prefixed models.

Apart from developing a stable algorithm for inferring NHHMMs in the presence of model uncertainty, another main goal of our work is to apply our model to forecasting real financial data. Financial data sets are characterized by large volatilities due to high market uncertainties. We argue that the logarithmic scoring rule for evaluating different models and their predictive accuracy is not appropriate, since the data have multimodal and fat-tailed distributions. We propose the use of Continuous Ranked Probability Score as a better alternative for assessing the quality of forecasts as well as for validating the model performance.

The paper proceeds as follows. In Section 2 we briefly describe the proposed model and in Section 3 we present analytically the Bayesian inference for this model, specifically for the model with a fixed number of predictors (covariates), as well as for the model with unknown number of predictors. Then, in Section 4, we present the forecasting criteria we used to asses the predictive ability of our method. To check our methodology we performed extensive experiments with simulated data. We provide in Section 5 some indicative case studies. We illustrate our results regarding the variable selection and forecasting evaluation and we make comparisons with other simpler models such as the Homogeneous Hidden Markov Model and the regression model with autoregressive terms. Next, we apply our methodology in monthly realized volatility data set (Section 6). Finally a summary of our work and possible extensions are discussed in **??**.

2 The Non-Homogeneous Hidden Markov Model

In this section, we present the proposed Non-Homogeneous Hidden Markov Model (NHHMM) for modeling univariate time series. Consider an observed random process $\{Y_t\}$ and a hidden underlying process $\{Z_t\}$ which is a two-state non-homogeneous discrete-time Markov chain that determines the states of the observed process. Let y_t and z_t be the realizations of the observed random process $\{Y_t\}$ and of the hidden process $\{Zt\}$, respectively. We assume that at time t, y_t depends on the current state z_t and not on the previous states. Consider also a set of available predictors $\{X_t\}$ with realization $x_t = (1, x_{1t}, \ldots, x_{K-1t})$ at time t. A subset of the predictors $X_t^{(1)} \subseteq \{X_t\}$ is used in the regression model for the observed process and a subset $X_t^{(2)} \subseteq \{X_t\}$ is used to model/describe the dynamics of the time-varying transition probabilities. Thus, we allow the covariates to affect the observed process $\{Y_t\}$ in a non-linear fashion.

In the case of a univariate random process $\{Y_t\}$, the NHHMM can be written in the form

$$Y_t = g(Z_t) + \epsilon_t$$

where $g(Z_t) = X_{t-1}^{(1)} B_{Z_t}$, $X_{t-1}^{(1)} = (1, x_{1t-1}^{(1)}, \ldots, x_{K-1t-1}^{(1)})$, $B_{Z_t} = (b_{0Z_t}, b_{1z_t}, \ldots b_{K-1Z_t})'$ and $\epsilon_t \sim \mathcal{N}(0, \sigma_{Z_t}^2)$. We use $\mathcal{N}(\mu, \sigma^2)$ to denote the normal distribution with mean μ and variance σ^2. In a less formal way, if s represents the hidden states, the observed series given the unobserved process has the form

$$Y_t \mid Z_t = s \sim \mathcal{N}(X_{t-1}^{(1)} B_s, \sigma_s^2) \ , s = 1, 2$$

The dynamics of the unobserved process $\{Z_t\}$ can be described by the time-varying transition probabilities, which depend on the predictors $X_t^{(2)}$ and are given by the following relationship

$$p_{ij}^{(t)} = \frac{\exp(x_t^{(2)} \beta_{ij})}{\sum_{l=1}^m \exp(x_t^{(2)} \beta_{il})}$$

3

where $\beta_{ij} = (\beta_{0,ij}, \beta_{1,ij} \ldots, \beta_{n-1,ij})'$ is the vector of the logistic regression coefficients to be estimated and $x_t^{(2)} = (1, x_{1t}^{(2)}, \ldots, x_{n-1t}^{(2)})$ is the set of covariates that affect the transition probabilities. Note that for identifiability reasons we adopt the convention of setting, for each row of the transition matrix, one of the β_{ij}'s to be a vector of zeros. The unknown quantities of the NHHMM are $\{\theta_s = (B_s, \sigma_s^2), \beta_{ss}, s = 1, 2\}$, that is the parameters in the mean predictive regression equation and the parameters in the logistic regression equation for the transition probabilities of the unobserved process $\{Z_t\}$, $t = 1, ..., T$. Our model can easily be generalized into an m-state NHHMM using almost the same methodology as the one that we describe.

3 Bayesian Inference and Computational Strategy

This section presents the Bayesian approach to inference for the Non-Homogeneous Hidden Markov model. The key steps in our proposed framework are as follow. First, for a given Hidden Markov model with time-varying probabilities, we construct a Markov chain which has as a stationary distribution the posterior distribution of the model parameters. Simulation of this Markov chain provides, after some burn-in period and adequately many iterations, samples from the posterior distribution of interest; see, for details, Besag et al. (1995). Second, for a given set of competing models each including a different set of predictors in the mean regression and/or in the transition probabilities equation, we base our inference about the models on their posterior probabilities. Thus, we avoid the usual approach which considers the models separately and chooses the best model via significance tests or via model selection criteria (AIC, BIC, DIC, WBIC).

3.1 Inference for fixed sets of predictors

Below we provide detailed guidelines on how to estimate the parameters of a given NHHMM, i.e. for fixed sets of predictors $X^{(1)}$ and $X^{(2)}$. The proposed approach to Bayesian inference on the parameters of the NHHMM is based on constructing a Markov chain Monte Carlo algorithm which updates, in turn, the mean regression parameters, the logistic regression coefficients, and the latent variables Z. 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 $Y_t \mid Z_t = s$, $s \in S$ and $\pi_1(z_1)$ the initial distribution of Z_1. The joint likelihood function of the observed data, y^T, and the unobserved sequence of states, z^T, is given by

$$\pi(y^T, z^T \mid X, \theta, \beta) = \pi(y^T \mid z^T, X, \theta, \beta)\pi(z^T \mid X, \theta, \beta) = \pi_1(z_1)f_{z_1}(y_1) \prod_{t=2}^{T} \prod_{i=1}^{s} \prod_{j=1}^{s} p_{ij}^{(t-1)} f_j(y_t)$$

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

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

For the parameters in the mean predictive regression equation, we use conjugate prior distributions, i.e.

$$\sigma_s^2 \sim \mathcal{IG}(p, q), \ B_s \mid, \sigma_s^2 \sim \mathcal{N}(L_0, \sigma_s^2 V_0), \ s = 1, 2,$$

where $\mathcal{IG}$ denotes the Inverted-Gamma distribution. To make inference about the logistic regression coefficients we use the auxiliary variables method of Polson et al. (2013) as described in Subsection 3.1.1. Given an auxiliary variable ω, a conjugate prior for the logistic regression coefficients β_{ss}, $s = 1, 2$ is multivariate normal distribution $\mathcal{N}(m_\omega, V_\omega)$.

The joint likelihood of the observed data y^T and the hidden states z^T is

$$\mathcal{L}(\theta, \beta) = \pi\left(y^T, z^T \mid X, \theta, \beta\right) = \pi_1(z_1) f_{z_1}(y_1) \prod_{t=2}^{T} \prod_{i=1}^{m} \prod_{j=1}^{m} p_{ij}^{(t-1)} f_j(y_t)$$

$$= \prod_{i=1}^{m} \prod_{j=1}^{m} \left[\prod_{t:z_t=j} p_{ij}^{(t-1)} \right] \left(\frac{1}{2\pi\sigma_j^2} \right)^{n_j/2}$$

$$\times \exp\left\{ -\frac{1}{2\sigma_j^2} (Y_j - X_j^{(1)'} B_j)'(Y_j - X_j^{(1)'} B_j) \right\}$$

$$= \prod_{i=1}^{m} \prod_{j=1}^{m} \left[\prod_{t:z_t=j} \frac{\exp(x_{t-1}^{(2)} \beta_{ij})}{\sum_{l=1}^{m} \exp(x_{t-1}^{(2)} \beta_{il})} \right] \left(\frac{1}{2\pi\sigma_j^2} \right)^{n_j/2}$$

$$\times \exp\left\{ -\frac{1}{2\sigma_j^2} (Y_j - X_j^{(1)'} B_j)'(Y_j - X_j^{(1)'} B_j) \right\}$$

We use the notation n_s, $s = 1, 2$ for the number of times the chain was in state s, that is $n_s = \sum_{t=1}^{T} I(Z_t = s)$, the $n_s \times T$ with I the indicator function.

The MCMC sampling scheme is constructed with recursive updates of (i) the latent variables z^T given the current value of the model parameters by using the scaled Forward-backward algorithm (Scott (2002)) (ii) the logistic regression coefficients by adopting the auxiliary variables method of Polson et al. (2013) given the sequence of states z^T, and (iii) the mean regression coefficients conditional on z^T by using the Gibbs sampling algorithm.

3.1.1 Simulation of the logistic regression coefficients

In a two-state NHHMM, as Meligkotsidou and Dellaportas (2011) observe, we can model the two diagonal elements of probability transition matrix by linking them to the set of covariates using a logistic link. However, the algorithm adopting the auxiliary representation of Holmes and Held (2006) in the method of Meligkotsidou and Dellaportas (2011) does not converge in some cases, especially if there is model uncertainty in transition probabilities. Many data-augmentation or Metropolis-Hastings algorithms are proposed to model the logistic regression model, see for example O'Brien and Dunson (2004); Fussl et al. (2013); Polson et al. (2013). We follow Polson et al. (2013) since as shown in their work, using Pólya-Gamma data augmentation gives superior results, it terms of convergence and mixing, among all the competing models.

Given the unobserved (latent) data $z^T = (z_1, \ldots, z_T)$ we define $\tilde{Z}_t^s = I[Z_{t+1} = Z_t = s]$. In words $\tilde{Z}_t^s$ is the number of times where the chain was at the same stage for two consecutive time periods. Then,

$$p\left(\tilde{Z}_t^s\right) = p_{ss}^t = \frac{\exp\left(x_t^{(2)} \beta_{ss}\right)}{1 + \exp\left(x_t^{(2)} \beta_{ss}\right)} \Leftrightarrow logit(p_{ss}^t) = x_t^{(2)} \beta_{ss}, \ s = 1, 2.$$

Polson et al. (2013) proved that binomial likelihoods (thus Bernoulli likelihoods in our simpler case) parametrized by log odds can be represented as mixtures of Gaussian distributions with respect to Pólya-Gamma distribution. The main result of Polson et al. (2013) is that letting $p(\omega)$ be the density of a latent variable ω with $\omega \sim \mathcal{PG}(b, 0)$ for $b > 0$, the following integral identity

holds for all $a \in \mathbb{R}$:

$$\frac{\exp\left(\psi\right)^a}{\left(1+\exp\left(\psi\right)\right)^b} = 2^{-b} \exp\left(k\psi\right) \int_0^\infty \exp\left(-\omega\psi^2/2\right) p\left(\omega\right) d\omega,$$

where $k = a - b/2$. Furthermore, the conditional distribution of $\omega \mid \psi$ is also Pólya-Gamma, $\mathcal{PG}(b, \psi)$. Using the previous result and setting $\Omega_s = diag\{\omega_{1,s}, \ldots, \omega_{N_s,s}\}$ as a set of latent variables the likelihood for each state $s = 1, 2$ is

$$\mathcal{L}\left(\beta, \omega\right) = \prod_{t=1}^{N_s} \left\{\frac{\exp\left(x_t^{(2)}\beta_{ss}\right)}{1+\exp\left(x_t^{(2)}\beta_{ss}\right)}\right\}^{\tilde{z}_t} \left\{\frac{1}{1+\exp\left(x_t^{(2)}\beta_{ss}\right)}\right\}^{1-\tilde{z}_t}$$

$$= \prod_{t=1}^{N_s} \frac{\exp\left(x_t^{(2)}\beta_{ss}\right)^{\tilde{z}_t}}{1+\exp\left(x_t^{(2)}\beta_{ss}\right)}$$

$$\propto \prod_{t=1}^{N_s} \exp\left(k_t x_t^{(2)}\beta_{ss}\right) \int_0^\infty \exp\left\{-\omega_t \left(x_t^{(2)}\beta_{ss}\right)^2/2\right\} p(\omega_t) d\omega_t$$

Conditioning on Ω, one can derive the proportion

$$\pi\left(\beta \mid z^t, \omega\right) \propto \prod_{t=1}^{N_s} \exp\left\{k_t x_t^{(2)}\beta_{ss} - \frac{\omega_t}{2}\left(x_t^{(2)}\beta_{ss}\right)^2\right\} \pi\left(\beta_{ss}\right)$$

$$\propto \pi\left(\beta\right) \prod_{t=1}^{N_s} \exp\left\{-\frac{\omega_t}{2}\left(\left(x_t^{(2)}\beta_{ss}\right)^2 - \frac{2k_t x_t^{(2)}\beta_{ss}}{\omega_t}\right)\right\}$$

$$\propto \pi\left(\beta\right) \prod_{t=1}^{N_s} \exp\left\{-\frac{\omega_t}{2}\left(\left(x_t^{(2)}\beta_{ss}\right)^2 - \frac{2k_t x_t^{(2)}\beta_{ss}}{\omega_t} + \frac{k_t^2}{\omega_t^2}\right)\right\}.$$

Assuming as prior distributions $\omega \sim \mathcal{PG}(b, 0)$ and $\beta \sim \mathcal{N}(m_{0\omega}, V_{0\omega})$, simulation from the posterior distribution can be done iteratively in two steps:

$$\omega_{t,s} \mid \tilde{z}_t \sim PG\left(1, x_t^{(2)}\beta_{ss}\right), \ t = 1 : N_s, \ s = 1, 2,$$

$$\beta_{ss} \mid \tilde{Z}, \Omega_s \sim \mathcal{N}(m_{\omega_s}, V_{\omega_s}),$$

$$V_{\omega_s} = \left(X^{(2)\prime}\Omega_s X^{(2)} + m_{0\omega}^{-1}\right)^{-1} \text{ and } m_{\omega_s} = V_{\omega_s}\left(X^{(2)\prime}k + V_{0\omega}^{-1}m_{0\omega}\right),$$

where $\mathcal{PG}$ denotes the Pólya-Gamma distribution and $k = (\tilde{z}_1 - 1/2, \ldots, \tilde{z}_{N_s} - 1/2)$.

3.1.2 Simulation of the mean equation parameters using the Gibbs algorithm

We used conditionally conjugate prior distributions on the parameters of the mean predictive distribution, i.e.

$$\sigma_s^2 \sim \mathcal{IG}\left(p, q\right) \text{ and } B_s \mid \sigma_s^2 \sim \mathcal{N}\left(L_{0s}, \sigma_s^2 V_{0s}\right), \ s = 1, 2.$$

After some straightforward algebra we derive the conditional and the marginal posterior distribution for the state specific parameters σ_s and B_s,

$$\sigma_s^2 \mid y^T, z^T \sim \mathcal{IG}\left(p + \frac{n}{2} + \frac{n_s}{2}, q + \frac{1}{2}\left(Y_s - X_s^{(1)} B_s\right)'\left(Y_s - X_s^{(1)} B_s\right) + \frac{1}{2}\left(B_s - L_{0s}\right)' V_{0s}^{-1}\left(B_s - L_{0s}\right)\right)$$

$$\sim \mathcal{IG}\left(p + \frac{n_s}{2}, q + \frac{1}{2}\left(L_{0s}' V_{0s}^{-1} L_{0s} + Y_s' Y_s - L_s' V_s^{-1} L_s\right)\right),$$

$$B_s \mid \sigma_s^2, z^T, y^T \sim \mathcal{N}\left(L_s, \sigma_s^2 V_s\right),$$

with $V_s = \left(V_{0s}^{-1} + X_s^{(1)'} X_s^{(1)}\right)^{-1}$, $L_s = V_s\left(V_{0s}^{-1} L_{0s} + X_s^{(1)'} Y_s\right)$.

3.2 Inference under model uncertainty

Here, we consider the full model comparison problem where the uncertainty about which predictors should be included in the mean regression model is taken into account together with the uncertainty about the predictors that should be included in the transition probability equation. The proposed model is flexible, since we do not decide a priori which covariates affect the observed or the unobserved process. Instead, we have a common pool of covariates $\{X\}$ and within the MCMC algorithm we gauge which covariates are included in subset $\{X^{(1)}\}$, affecting the mean predictive equation of the observed process, and which covariates are included in subset $\{X^{(2)}\}$, affecting the time-varying transition probabilities. It is worth mentioning that within the framework of the proposed model, a set of $p-1$ covariates implies that there are $2^{p-1} \times 2^{p-1} = 2^{2(p-1)}$ possible models, hence the model selection problem becomes complicated.

Different approaches have been used in the literature to cope with the model selection problem. The use of information criteria, such as Akaike's Information Criterion (AIC, Akaike et al. (1973)), Bayesian Information Criterion (BIC) of Schwarz (1978) or Deviance Information Criterion (DIC, Spiegelhalter et al. (2002)) or Widely applicable Bayesian Information Criterion (WBIC, Watanabe (2013)), is another approach to variable selection.

We propose a probabilistic approach to inference, which is based on the calculation of the posterior distribution of different NHHMMs or equivalently on computing the posterior probabilities of different hidden Markov models. Posterior probabilities can be used either for selecting the most probable model (i.e. making inference using the model with the highest posterior probability), or for Bayesian model averaging (i.e. producing inferences averaged over different NHHMM). Barbieri and Berger (2004) argue that the optimal predictive model is not necessarily the model with highest posterior probability but the median probability model, which is defined as the model consisting of those covariates which have overall posterior probability of being included in the model (inclusion probability) greater or equal to 0.5. Our method allows us to calculate the posterior probability of the model as well as the probabilities of inclusion. To this end, we develop a Reversible Jump Markov Chain Monte Carlo (RJMCMC) algorithm (Green (1995), Green and Hastie (2009)) which explores the model space by jumping between different hidden Markov models. A more applied tutorial based on the lines of Green (1995), can be found Waagepetersen and Sorensen (2001). A study for comparing variable selection methods is well presented in O'Hara and Sillanpää (2009) whilst Dellaportas et al. (2002) study the variable selection methods in the context of model choice.

As Green and Hastie (2009) noticed, reversible jump MCMC is in fact a Metropolis-Hastings algorithm, formulated to allow sampling from a distribution on a union of spaces of differing dimension and to permit state-dependent choice of move type. Suppose that a prior $\pi(k)$ is specified over k models $(M_1, M_2, \ldots, M_k)$ in a countable set $\mathcal{K}$ and for each k we are given a

prior distribution $\pi(\tilde{\theta}_k \mid k)$ along with a likelihood $\mathcal{L}(y \mid k, \tilde{\theta}_k)$ for data y. The joint prior for $\tilde{\theta}_k$ and k is $\pi(k, \tilde{\theta}_k) = \pi(\tilde{\theta}_k \mid k)p(k)$ and obviously the joint posterior distribution is

$$\pi(k, \tilde{\theta}_k \mid y) = \frac{\pi(k, \tilde{\theta}_k)\mathcal{L}(y \mid k, \tilde{\theta}_k)}{\sum_{k' \in \mathcal{K}} \int \pi(k', \tilde{\theta}_{k'})\mathcal{L}(y \mid k', \tilde{\theta}_{k'})d\tilde{\theta}'_{k'}}$$

The standard formulation of the Metropolis-Hastings algorithm relies on the construction of a time-reversible Markov chain, that satisfies the detailed balance condition. This condition means that the probability that the state of a chain is in a general set A and moves to a general set B is the same with the probability that the state is in the set B and moves to set A. Note that this is a simple way to ensure that the limiting distribution of the chain is the desired target distribution. Define $\tilde{x} = (k, \tilde{\theta}_k) \in \bigcup_{k \in \mathcal{K}} (\{k\} \times \mathcal{X}_k)$ as the state and state space of chain. At the current state $\tilde{x}$ we generate r random numbers u from a known joint density g. The new state of chain $\tilde{x}'$ is constructed by some suitable deterministic function $h : \mathcal{R}^n \times \mathcal{R}^r \to \mathcal{R}^{n'} \times \mathcal{R}^{r'}$ such that $(\tilde{x}', u') = h(\tilde{x}, u)$. The inverse move is made by generating u' from a suitable known density g' and then using the inverse function $h' : \mathcal{R}^{n'} \times \mathcal{R}^{r'} \to \mathcal{R}^n \times \mathcal{R}^r$ of h to move from $\tilde{x}'$ to $\tilde{x}$. In practice, the construction of proposal moves between different models is achieved via the concept of 'dimension matching'. Specifically, if the dimensions of $\tilde{x}, \tilde{x}', u$ and u' are n, n', r and r' respectively, then $n + r = n' + r'$. If the move from $\tilde{x}$ to $\tilde{x}'$ is accepted with probability $\alpha(\tilde{x}, \tilde{x}')$ then, the reverse move is accepted with probability $\alpha(\tilde{x}', \tilde{x})$. Under this formulation the detailed balance condition is

$$\int_{(\tilde{x}, \tilde{x}') \in A \times B} \pi(\tilde{x})g(u)\alpha(\tilde{x}, \tilde{x}') \, dx du = \int_{(\tilde{x}, \tilde{x}') \in A \times B} \pi(\tilde{x}')g'(u')\alpha(\tilde{x}', \tilde{x}) \, dx' du'.$$

If the transformation h from (x, u) to (x', u') and its inverse h' are differentiable then the equality holds if

$$\pi(\tilde{x}) g(u)\alpha(\tilde{x}, \tilde{x}') = \pi(\tilde{x}')g'(u')\alpha(\tilde{x}', \tilde{x}) \left| \frac{\partial(\tilde{x}', u')}{\partial(\tilde{x}, u)} \right|,$$

where $J = \left| \frac{\partial(\tilde{x}', u')}{\partial(\tilde{x}, u)} \right|$ is the Jacobian of the transformation. Thus, the acceptance probability α, for moving is

$$\alpha(\tilde{x}, \tilde{x}') = \min \left\{ 1, \frac{\pi(\tilde{x}') g'(u')}{\pi(\tilde{x}) g(u)} \left| \frac{\partial(\tilde{x}', u')}{\partial(\tilde{x}, u)} \right| \right\}.$$

Holsclaw et al. (2017) recently introduced NHHMMs using the Pólya-gamma latent data method. As a model selection strategy, they use BIC values to choose the best model among a prespecified set of models. We apply a reversible jump algorithm to account model uncertainty. Due to the fact that in the proposed model there is uncertainty both in the mean predictive equation, as well as in the transition probability equation, we will have to perform two reversible jump steps within our MCMC algorithm. Let us now assume that in a reversible jump step we propose a move type m from $\tilde{x} = (k, \theta_k)$ to $\tilde{x}^* = (k^*, \theta_{k^*}^*)$. Let $j_m(\tilde{x})$ denote the probability that move m is attempted at state $\tilde{x}$ and $j_{m^*}(\tilde{x}^*)$ the probability of the reverse move attempted at state $\tilde{x}^*$. We accept the proposed move with probability $\alpha_m(\tilde{x}, \tilde{x}^*) = \min\{1, A_m(\tilde{x}, \tilde{x}^*)\}$ where

$$A_m(\tilde{x}, \tilde{x}^*) = \frac{\mathcal{L}(y^T \mid \tilde{x}^*) p(\theta^* \mid k^*) p(k^*) j_{m^*}(\tilde{x}^*) g'_m(u^* \mid \tilde{x}^*, k)}{\mathcal{L}(y^T \mid \tilde{x}) p(\theta \mid k) p(k) j_m(\tilde{x}) g_m(u \mid \tilde{x}, k^*)} \left| \frac{\partial(\theta_{k^*}^*, u^*)}{\partial(\theta_k, u)} \right|.$$

We adopt the proposal of Meligkotsidou and Dellaportas (2011) to implement the above algorithm. In each step, we choose to add or remove one covariate with probability 0.5 and then we randomly choose which covariate we will add/remove. We propose a new value for the mean

equation coefficients B^* or for the regression equation coefficients β^* from the full conditional posterior density, conditionally on the other coefficients, thus the Jacobian of transformation will be equal to unity. To be more specific, if we want to update the covariates in the mean equation, the proposal distribution g' is just the product of the two conditional posterior distributions for $s = 1, 2$ derived in Subsection 3.1.2 given σ and if we want to update the covariates that affect the transition matrix, the proposal distribution is the product of conditional normal distributions derived in Subsection 3.1.1 given Ω_s for $s = 1, 2$. With some straightforward matrix algebra, the acceptance probability for the mean equation is $\alpha_B = \min\{1, A_B\}$ and the acceptance probability for the transition matrix is $\alpha_\beta = \min\{1, A_\beta\}$ where

$$A_B = \frac{j_{m^*}(k^*)}{j_m(k)} \prod_{s=1}^{2} \frac{|V_s^*|^{1/2} |V_{0s}|^{1/2}}{|V_s|^{1/2} |V_{0s}^*|^{1/2}}$$
$$\times \exp\left\{ -\frac{1}{2\sigma_s^2} \left(L_{0s}^{*'} V_{0s}^{*-1} L_{0s}^* - L_s^{*'} V_s^{*-1} L_s^* - L_{0s} V_{0s}^{-1} L_{0s} + L_s' V_s^{-1} L_s \right) \right\},$$

and

$$A_\beta = \frac{j_{m^*}(k^*)}{j_m(k)} \prod_{s=1}^{2} \frac{|V_{\omega s}^*|^{1/2} |V_{0\omega s}|^{1/2}}{|V_{\omega s}|^{1/2} |V_{0\omega s}^*|^{1/2}}$$
$$\times \exp\left\{ -\frac{1}{2\sigma_s^2} \left(L_{0\omega s}^{*'} V_{0\omega s}^{*-1} L_{0\omega s}^* - L_\omega^{*'} V_{\omega s}^{*-1} L_{\omega s}^* - L_{0\omega s} V_{0\omega s}^{-1} L_{0\omega s} + L_{\omega s}' V_{\omega s}^{-1} L_{\omega s} \right) \right\}.$$

3.3   MCMC Sampling Scheme

In the next lines, we summarize the predictive MCMC algorithm that we have constructed for joint inference on model specification and model parameters.

1. Start with initial values of $\beta, \theta = [B, \sigma]$.

2. Calculate the probabilities of time-varying transition matrix for $t = 1 : T$.

3. Run a Scaled Forward-Backward (Scott (2002)) algorithm to simulate the hidden states given the parameters of the model.

4. Simulate the parameters of the regression model for the mean via a Gibbs sampler method.

5. Simulate the coefficients β using the Pólya-gamma representation by Polson et al. (2013).

6. Do a double reversible jump algorithm to define which covariates will affect the transition matrix and which covariates will affect the mean regression model.

7. Make one-step-ahead predictions conditional on the simulated unknown quantities.

Repeat steps 3-6 until convergence and then repeat steps 3-7.

4   Bayesian Forecasting and Scoring rules

4.1   One-step-ahead predictions

The proposed modeling and inferential approach is used for forecasting. The posterior predictive density can not be found in closed form, but it can instead be evaluated numerically. Given

model M, the predictive distribution of y_{T+1} is

$$f_p\left(y_{T+1} \mid y^T\right) = \int f\left(y_{T+1} \mid y^T, z^T, M, \beta_M, \theta_M\right) \pi\left(\beta_M, \theta_M \mid y^T\right) d\beta_M d\theta_M,$$

where

$$f\left(y_{T+1} \mid y^T, z^T, \beta_M, \theta_M\right) = \sum_{s=1}^{2} P\left(Z_{T+1} = s \mid Z_T = z_T\right) f_s\left(y_{T+1}\right).$$

In practice we follow an iterative procedure within the MCMC algorithm to draw a sample from the posterior predictive distribution. At the r-th iteration of our algorithm, the algorithm chooses model M. Also, the hidden states and the unknown parameters $\beta_M^{(r)}, \theta_M^{(r)}$ are simulated as described before. Let $z_T^{(r)}$ be the hidden state is at time T. To make an one-step-ahead prediction (i.e. simulate y_{T+1}), we first simulate the hidden state for time $T+1$, from the discrete distribution based on the transition probabilities $P\left(Z_{T+1}^{(r)} = s \mid Z_T = z_T^{(r)}\right)$, $s = 1, 2$ and then conditional on the hidden state we draw a value y_{T+1}^r from

$$Y_{T+1}^{(r)} \mid Z_{T+1}^{(r)} = s \sim \mathcal{N}\left(X_T^{(1)} B_s, \sigma_s^2\right), \ s = 1, 2.$$

Given Y_{T+l}, $l = 1, \ldots, L$, the hidden state Z_{T+l} and the covariates X_{T+l-1}, we update the transition matrix P^{T+l-1}, simulate Z_{T+l+1} and simulate the prediction y_{T+l+1} from the predicting distribution, in sequence.

4.2   Forecasting criteria

One way to evaluate a model is through the accuracy of its forecasts (Geweke and Whiteman (2006)). As Gelman et al. (2014) observe, predictive accuracy is valued not only for its own sake but also for comparing different models. Advances in numerical integration via MCMC algorithms made probabilistic forecasts possible, which are in most cases, preferable. Besides, having the posterior predictive distribution, one can obtain a point effective forecasts, using suitable scoring functions (Gneiting (2011)).

Scoring rules provide summary measures for the evaluation of probabilistic forecasts, by assigning a numerical score based on the forecast and on the event or value that materializes. We refer to Gneiting and Raftery (2007) for a review on the theory and properties of scoring rules. If the forecaster quotes the predictive distribution P and the event x materializes, then his reward is $S(P, x)$. If Q is the forecaster judgment then $S(P, Q)$ is the expected value of his reward under Q. A scoring rule is said to be proper if $S(Q, Q) \geq S(P, Q)$ and if the equality $S(Q, Q) = S(P, Q)$ holds if and only if $Q = P$, then the rule is strictly proper. Intuitively, proper scoring rules encourage a forecaster to report the truth about his judgment distribution and they are strictly proper is the issued forecasts are the same with the forecaster's judgment.

There are various proper scoring rules for continuous variables, such as the probabilistic score, the Linear Score (LinS), the Quadratic Score (QS), the Logarithmic Score (LogS) and the Continuous Ranked Probability Score (CRPS), among others (Machete (2013); Gneiting and Raftery (2007)). Probabilistic and Linear score are often used, nevertheless they are improper and can have misleading results. A widely used and quite powerful criterion is the Logarithmic Score. It is based on the posterior predictive density evaluated at the observed value and there is a strong connection with the classical Kullback-Leibler divergence, which is one of the reasons for being studied quite intensively (see for example Gelman et al. (2014), GschlÃűÃ§l and Czado (2007)). The Logarithmic Score is defined as follows: Let T be the total sample size and N the

number of out-of-sample forecasts. For $i = T - N, T - N + 1, \ldots, T$ let y_i be the real observed values of the forecasts, $\hat{y}_i$ the estimated forecasts and $\tilde{\theta}$ the unknown quantities. Finally, let D be the number of draws of the simulation. Using the notation $f_p(\hat{y}_i)$ for the posterior predictive density of the new data, f for the distribution of the true model and $\pi(y \mid \tilde{\theta})$ for the likelihood of the data, the predictive fit for the i-th observation would be,

$$\log\left(f_p\left(\hat{y}_i\right)\right) = \log \int \pi\left(\hat{y}_i \mid \tilde{\theta}\right) \pi\left(\tilde{\theta}\right) d\tilde{\theta}.$$

Having in mind though that the new data are unknown Gelman et al. (2014) defined the expected out-of-sample log predictive density for the i-th observation as

$$E_f\left(\log\left(f_p\left(\hat{y}_i\right)\right)\right) = \int \left(\log\left(f_p\left(\hat{y}_i\right)\right)\right) f\left(\hat{y}_i\right) d\hat{y}_i.$$

To compute the predictive density for N estimated forecasts in practice, since $\tilde{\theta}$ is unknown, we shall calculate the

$$LogS = \sum_{i=T-N}^{T} \log\left(\frac{1}{D}\sum_{j=1}^{D} \pi\left(\hat{y}_i \mid \tilde{\theta}^j\right)\right).$$

Although the Logarithmic Score is a strictly proper rule, it lacks robustness as it involves harsh penalty for low probability events and thus is sensitive to extreme cases (Boero et al. (2011)). Besides, comparing the entropies of the forecasts, Machete (2013) showed that LS prefers the forecast density that is less informative. As Gneiting and Raftery (2007) notice, measures which are not sensitive to distance give no credit for assigning high probabilities to values near but not identical to the one materializing. Sensitivity to distance seems desirable when predictive distributions tend to be multimodal, which is the case for our model. To deal with this, one could calculate the Continuous Ranked Probability Score which is based on the cumulative predictive distribution. Recently, Boero et al. (2011) made a comparison between LS, CRPS and QS and they argued that when density forecasts are collected in histogram format, then the ranked probability score has advantages over the other scoring rules. Note than when the forecast is deterministic, CRPS is reduced to the Mean Absolute Forecast Error $\left(\frac{1}{N}\sum_{i=T-N}^{T}|y_i - \hat{y}_i|\right)$ and therefore it is easily interpretable and provides a direct way for comparing various deterministic and probabilistic forecasts.

The posterior predictive cumulative density function is defined as $F_p(x) = \int_{-\infty}^{x} f_p\left(\hat{y} \mid \mathbf{y}\right) d\hat{y}$, where $\hat{y}$ is the forecast of interest and $\mathbf{y}$ is the history of the predictive quantity. Using the same notation as in Logarithmic score, the CRPS for the real observed value y_i is defined as,

$$CRPS(F_{p,i}, y_i) = -\int_{-\infty}^{\infty} \left(F_{p,i}\left(\hat{y}_i\right) - I\left(\hat{y}_i \geq y_i\right)\right)^2 d\hat{y}_i = -\int_{-\infty}^{\infty} \left(F_{p,i}\left(\hat{y}_i\right) - F_{y_i}\left(\hat{y}_i\right)\right)^2 d\hat{y}_i,$$

where $I\left(x \geq y\right)$ denotes a step function along the real line that attains the value 1 if $x \geq y$ and the value 0 otherwise, $F_{y_i} = H(\hat{y}_i - y_i)$ is the cumulative distribution of the real value y_i and H is the known Heaviside function (Hersbach (2000))

$$H(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ 1 & \text{otherwise.} \end{cases}$$

As we described above, at each iteration of the MCMC algorithm, we obtain a sample of length L from the predictive distributions of $y_{T+1}, \ldots, y_{T+L}$. Hence we can evaluate both the posterior

predictive probability and cumulative distribution function. Instead,there is an easier way of evaluating CRPS using the identity of SzÃĺkely and Rizzo (2005).

$$CRPS(F, y_i) = \frac{1}{2} E_F \left| Y - Y' \right| - E_F \left| Y - y_i \right|,$$

were Y, Y' are independent copies of a random variable with distribution function F (see also GschlÃűÃ§l and Czado (2007)).

Finally, for the sake of completeness we estimate the mean, mode and median of the predictive distribution and compare them using two well known point forecasting criteria, the Mean Square Forecast Error, $MSFE = \frac{1}{N} \sum_{i=T-N}^{T} (y_i - \hat{y}_i)^2$ and the Mean Absolute Forecast Error, $MAFE = \frac{1}{N} \sum_{i=T-N}^{T} |y_i - \hat{y}_i|$. Based on the MCMC sample, the computed point forecasting criteria is just the average score of MSFE or MAFE for all the sample values.

5   Simulation Study

We have tested our method using a number of simulated data sets. We conducted a series of simulation experiments to validate the superiority of our method. We tested quite extensively our algorithms, accounting for model uncertainty or not, using different sample sizes and assigning various values to the parameters. In all our experiments our model outperforms the competing models in forecasting the observed process y^T. We compared our model with a Homogeneous Hidden Markov model and a linear model with autoregression terms (qAR). Although considered standard, for the sake of completeness, we give the definitions of HHM and qAR models. In the Homogeneous Hidden Markov model, covariates affect the mean equation but the transition probability matrix is constant. That is,

$$Y_t \mid Z_t = s \sim \mathcal{N}(X_{t-1}^{(1)} B_s, \sigma_s^2), \ s = 1, 2, \ t = 1, \dots, T,$$

$$P\left(z_t = j \mid z_{t-1} = i\right) = p_{ij}, \ i, j = 1, 2 \ \forall \ t = 1, \dots T.$$

In the linear regression with autocovariates, besides covariates X_{t-1}, Y_{t-1} lagged values of Y_t are assumed to affect the mean equation. Specifically,

$$Y_t = \left[X_{t-1} Y_{t-1}\right] B + \epsilon_t \ , \ t = 1, \dots, T \text{ and } \epsilon_t \sim \mathcal{N}(0, \sigma^2).$$

The data were generated either from a Homogeneous Hidden Markov model (HHMM) or from a NHHMM model with independent normal distributed covariates. We found that the mean equation parameters converge rapidly whereas the logistic regression coefficients needed some burn in period to converge. The hidden chain Z^T was well estimated. For each iteration we kept a replication of the hidden chain (thus the number of the hidden-chain replications is the same with the MCMC sample size) and compared it with the real simulated hidden chain, using a 1-0 loss function. Also the reversible jump algorithm behaved really well in all of our simulations. Even in the case where the data were homogeneous, no covariates were selected in the logistic regression equation, thus the transition probability matrix remained constant through time. Furthermore, in order to test the predictive ability of our model we kept N out-of-sample observations and we calculated the CRPS for every observation. We confirm that our forecasts were more accurate than the forecasts of the competing models. A summary of our comparisons are presented and discussed in the following subsections.

5.1 The fixed model

Below, we present two experiments with simulated data using the fixed NHHMM. For all our experiments we used non-informative priors for the unknown parameters $\sigma_s, B_s, \beta_{ss}, \quad s = 1, 2$, that is $\sigma_s^2 \sim \mathcal{IG}(0.1, 0.1)$, $B_s \mid \sigma_s^2 \sim \mathcal{N}(0, 100\sigma_s^2)$ and finally $\beta_{ss} \sim \mathcal{N}(0, 80 \times I)$. Inferences are based on an MCMC sample of 15000 iterations after 10000 burn in period. In the first experiment we used a sample of $T = 1000$ observations. From a common pool of covariates $X = \{1, X_1, X_2, X_3, X_4, X_5, X_6, X_7\}$ we chose $X^{(1)} = \{1, X_1, X_2, X_3, X_7\}$ to affect the mean equation and $X^{(2)} = \{1, X_1, X_3, X_4, X_6, X_6\}$ to affect the transition matrix. The mean equation parameters were $B_1 = [5, 1, 3, 2, 4]'$, $\sigma^2 = 0.8$ and $B_2 = [3, -4, 3, 1, 2]'$, $\sigma^2 = 1.3$ whereas the logistic regression coefficients where $\beta_1 = [-2, 3, 2, 4, 1, -2]'$ and $\beta_2 = [2, 4, 1, 2, -1, 3]'$ for states $s = 1, 2$ respectively. Secondly, we run another experiment with sample size $T = 1500$. We used 3 covariates $X = \{1, X_1, X_2, X_3\}$ affecting the mean equation and $X = \{1, X_1, X_2, X_4\}$ the transition matrix. The mean equation parameters were $B_1 = [2, -0.3, 2, 2]'$, $\sigma^2 = 0.1.5$ and $B_2 = [1, 3, 4, 3]'$, $\sigma^2 = 0.8$ whereas the logistic regression coefficients where $\beta_1 = [1.5, 1, 2, 3]'$ and $\beta_2 = [3, -2.5, 4, 1]'$ for states $s = 1, 2$ respectively. For both experiments, we kept 10 out-of-sample observations, we computed a sequence of one-step-ahead forecasts of the real observed process. For completenes we present in Section 7 various plots that confirm the estimation performance and convergence of our methodology, parameter estimates and summary statistics of these experiment studies (Tables 7 and 9) and a comparison of the three competing models in terms of forecasting (Tables 8 and 10).
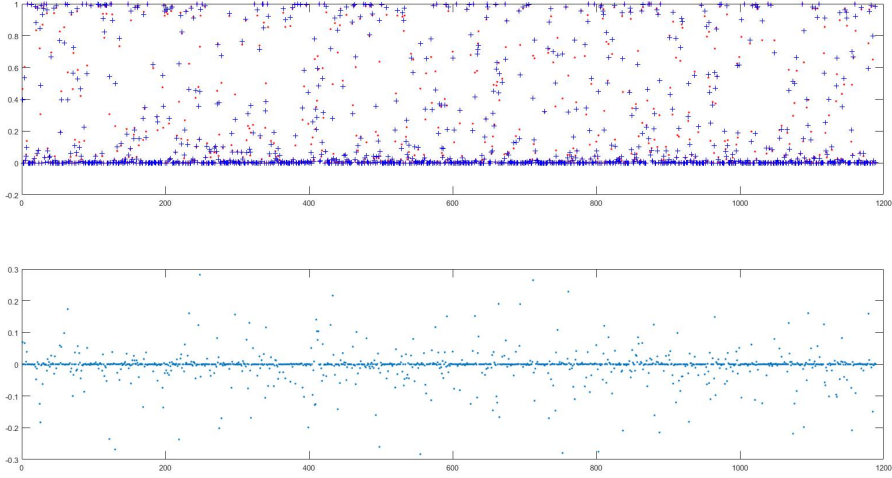

Figure 1: Experiment 1: The upper panel shows a comparison between the transition probabilities p_{ss}. True probability values are marked with red dots whereas the posterior mean estimated transition probabilities are marked with blue crosses. Lower panel: The difference between real and mean estimated probabilities.
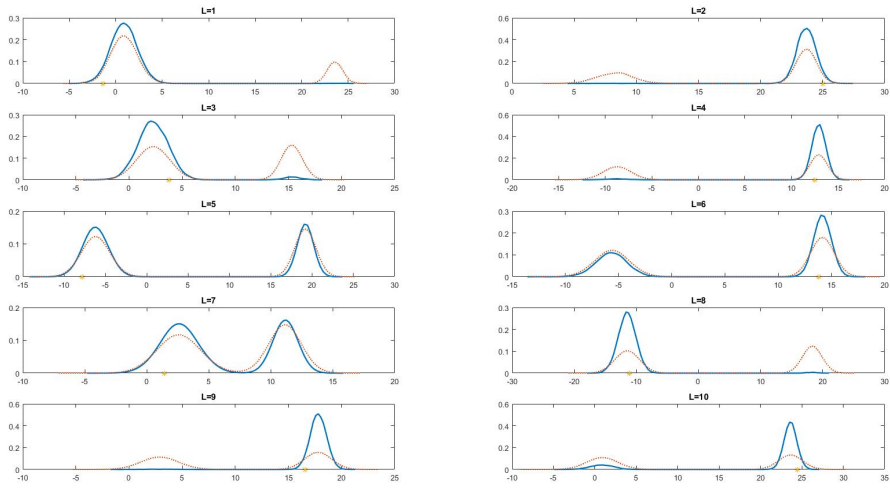
Figure 2: Forecasts of experiment 1. Empirical continuous approximation of the posterior predictive distribution (based on a normal kernel function) using the NHHMM model (blue continuous line) and the HHMM (red dotted line). True out-of-sample values are marked with yellow asterisk.
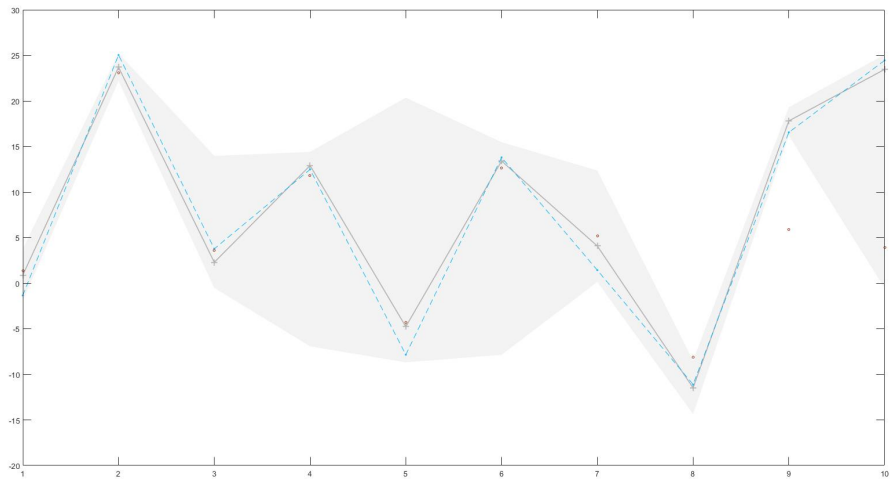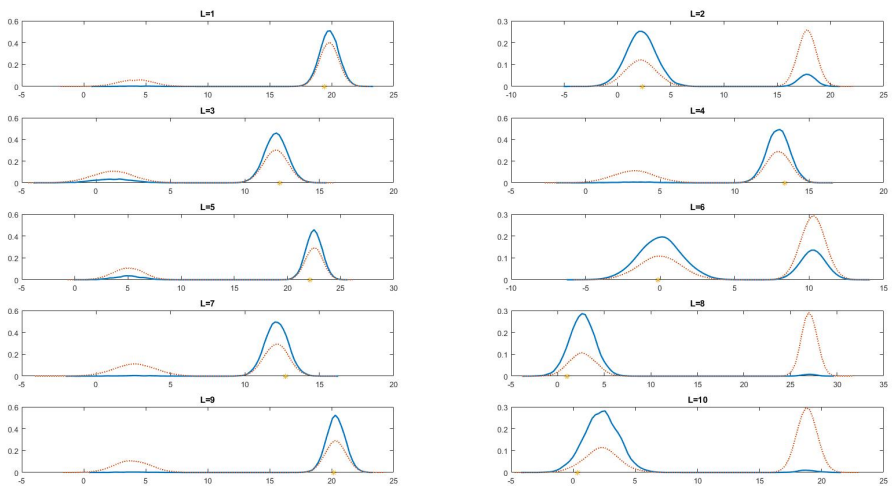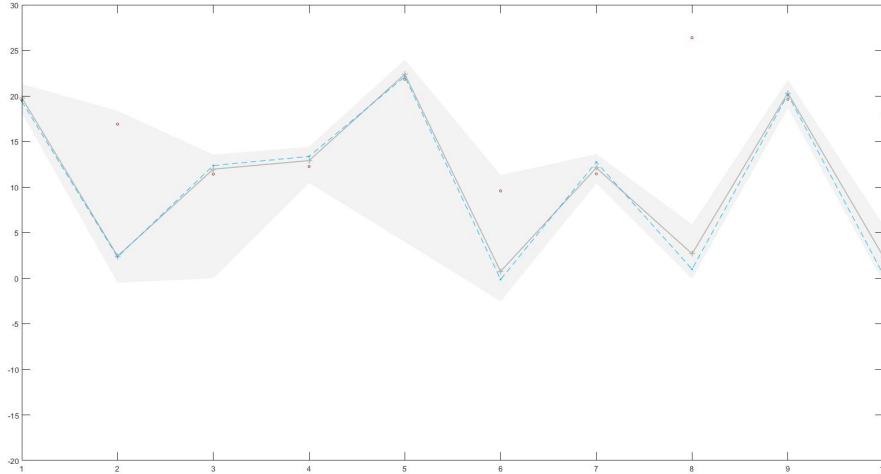

Figure 3: Experiment 1: This figure presents the quantiles of the predictive distribution of the studied model.. Blue dashed line presents the real observed out-of-sample values of interest. Gray bold line represents the median of the predictive distribution. The gray area around the median is defined by the 2.5% and 9.5% quantiles. Red circles represent the median values of the predictive distribution using the HHMM model.
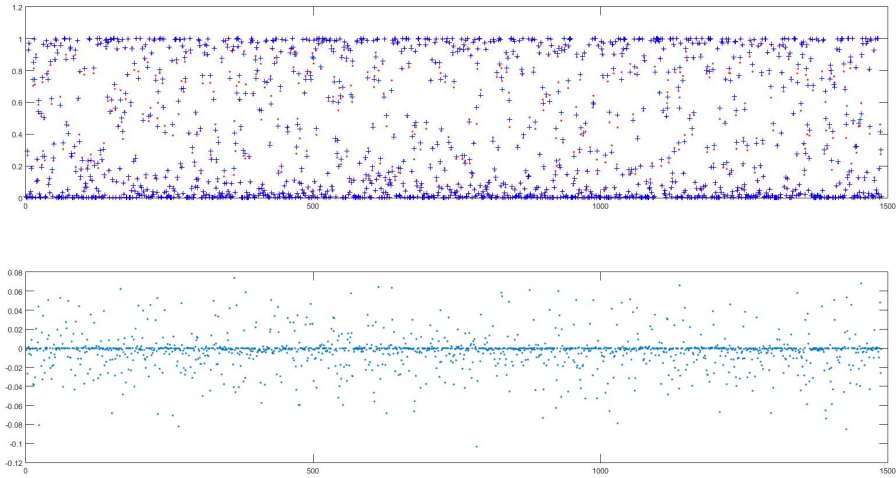
Figure 4: Experiment 2: The upper panel shows a comparison between the transition probabilities p_{11}^t and p_{22}^t. True probability values are marked with red dots whereas the posterior mean estimated transition probabilities are marked with blue crosses. Lower panel: The difference between real and mean estimated probabilities.

Figure 5: Forecasts of experiment 2. Empirical continuous approximation of the posterior predictive distribution (based on a normal kernel function) using the NHHMM model (blue continuous line) and the HHMM (red dotted line). True out-of-sample values are marked with yellow asterisk.

Figure 6: This figure presents the quantiles of the predictive distribution of the studied model for the second experiment. Blue dashed line presents the real observed out-of-sample values of interest. Gray bold line represents the median of the predictive distribution. The gray area around the median is defined by the 2.5% and 9.5% quantiles. Red circles represent the median values of the predictive distribution using the HHMM model.

5.2   Model selection

In this section we present two case studies (experiments 3 and 4). We compared outputs of three models (NHHMM, HHMM, qAR) with model uncertainty. We performed stochastic variable selection with reversible jump for all the competing models. As in Subsection 5.1 we used the same non informative priors and burn-in period. For the first experiment we simulated data of size $T = 1200$. We used the same two covariates affecting the mean equation and transition matrix via the logistic regression. To check the variable selection ability of our method we included three more covariates in the common pool of predictors. Next, we present a second, more complicated, experiment. We used 3 covariates affecting the mean equation and 5 covariates affecting the transition probability matrix. In Tables 1 and 3 we show that the most probable model and the Median probability model is the true one for the two experiments. Additionally, as in the fixed model case studies we kept 10 out-of-sample observations for forecasting evaluation exercise and we confirmed that our model performed better than the other models in terms of CRPS, MAFE and MSFE (see Tables 2 and 4). Finally, for each MCMC iteration we kept a replicated chain of the hidden process and we compared it with the true simulated chain. Using the 0-1 Loss function we computed the mean mis-estimated states in each chain. On average from the chain with 1200 hidden states of the first experiment, our model failed to recognize 22 states per iteration and on average from the chain of 1500 hidden states of the second experiment we failed to recognize 8 states. We note that we did not encounter any label switching problems.

In the following pages we present the mean estimated probabilities (for being at the same stage two consecutive time periods, that is p_{11} or p_{22}) of the transition matrix and the differences with the true simulated probabilities (Figure 7) for the first case study and in Figure 11 for the second case study. A realization of the observed process and hidden process as well as one realization of the simulated hidden process for the two case studies presented in Figures 8 and 12 respectively.

Lastly for each experiment, the predictive distribution using a kernel density function is plotted, for both NHHM and HHM models, are presented in Figures 10 and 14 and the quantiles of the predictive distributions in Figures 9 and 13.
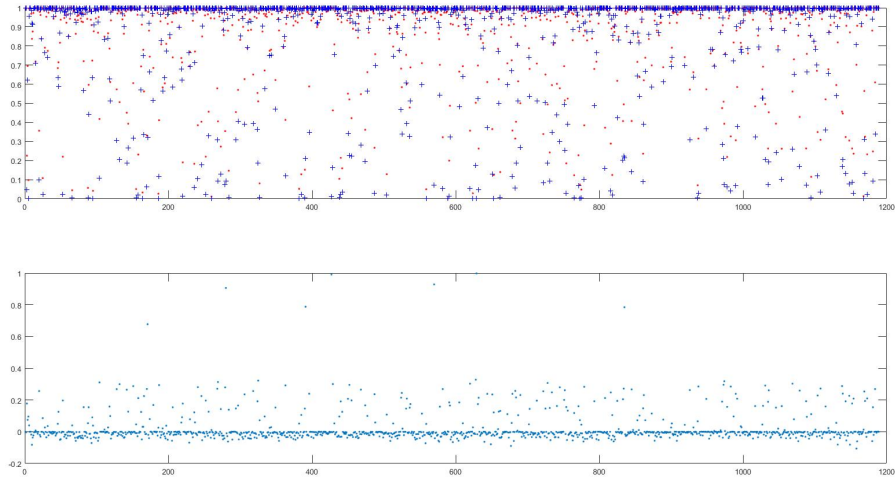

Figure 7: Experiment 3: True simulated transition probabilities vs mean estimated transition probabilities. Upper panel: True probabilities are marked with blue crosses and mean estimated probabilities are estimated with red dots. Lower panel: The difference between mean estimated probabilities for one estimated realized hidden process and the true simulated probabilities.
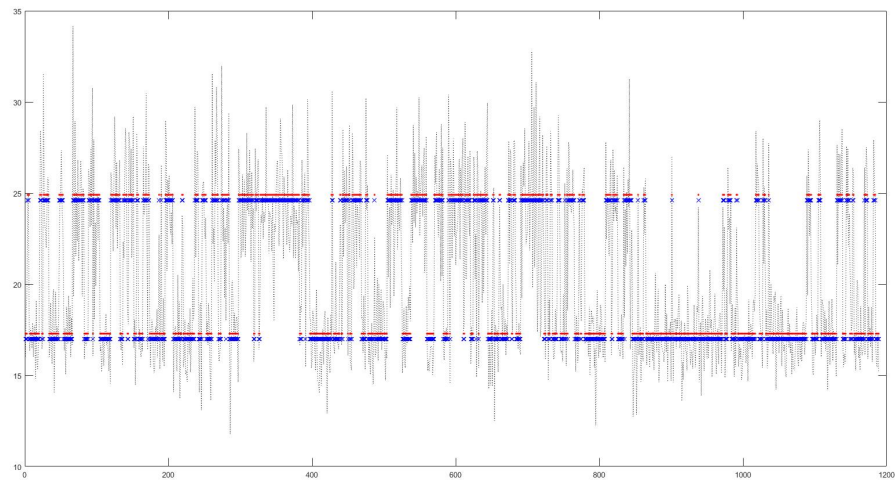

Figure 8: Observed process (black dotted line) and hidden process. True hidden states are marked with blue x and a realized simulated states are marked with red dots.

17

True model	
Included Covariates, Mean	X_1, X_2
Included Covariates, Transition Matrix	X_1, X_2
Total Covariates	X_1, X_2, X_3, X_4, X_5
Highest Posterior Probability model	
Included Covariates, Mean	X_1, X_2
Included Covariates, Transition Matrix	X_1, X_2
Posterior probability	0.9958
Median Probability model	
Included Covariates, Mean	X_1, X_2
Included Covariates, Transition Matrix	X_1, X_2

Table 1: Experiment 3: Highest posterior probability model and median probability model.
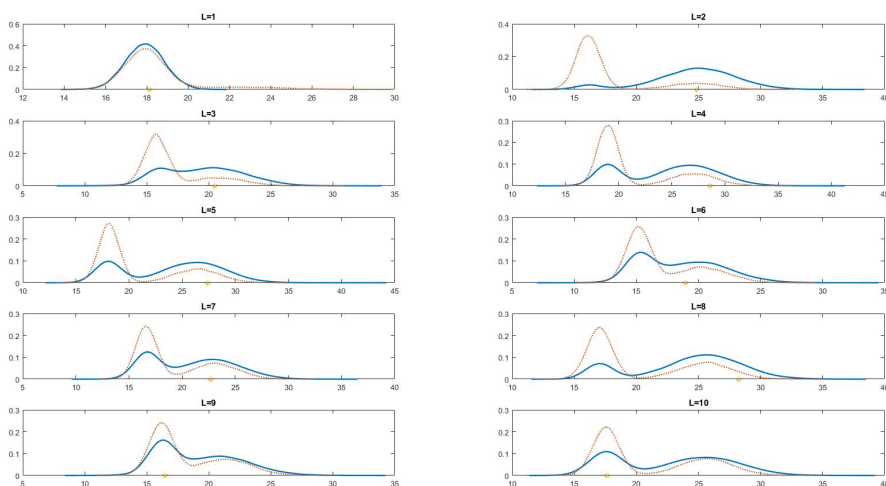

Figure 9: Experiment 3: Empirical posterior predictive distribution for the 10 out-of-sample forecasts using the NHHMM (blue line) and the (red dotted line). Yellow asterisk represents the true simulated values.
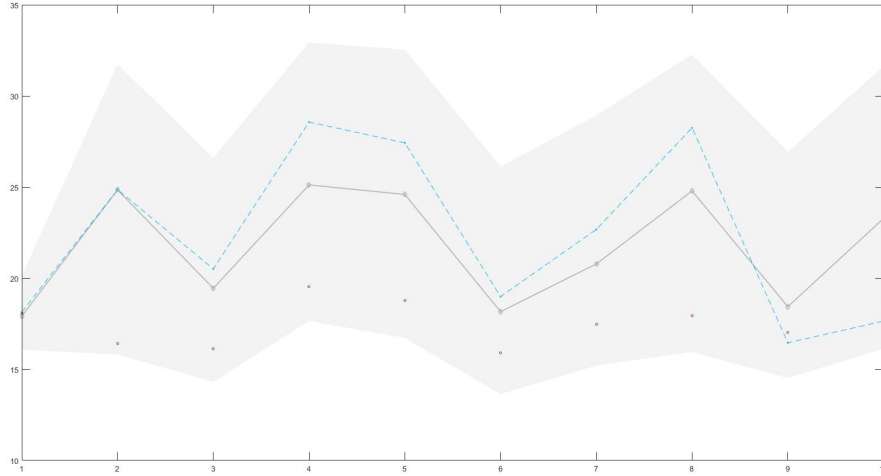
Figure 10: Experiment 3: Quantiles of the predictive distribution. Ten one-step-look-ahead observed values (as taken from the sample) are marked with dashed blue line (and grey dots). Gray bold line represents the median of the predictive distribution. The gray area around the median is defined by the 2.5% and 9.5% quantiles. Red circles represent the median values of the predictive distribution using the HHMM model.

Probabilistic Forecasting Criterion				
	y	NHHMM	HHMM	qAR
	y_1	-0.2428	-0.2735	-6.0555
	y_2	-0.7669	-5.5975	-6.2297
	y_3	-0.9958	-2.6073	-6.0018
	y_4	-2.3435	-5.1726	-6.6054
$CRPS(y_i)$	y_5	-2.0090	-4.6137	-6.4204
	y_6	-0.9844	-1.15774	-5.8633
	y_7	-1.4488	-2.4993	-6.0581
	y_8	-2.4239	-5.5700	-6.6718
	y_9	-1.0679	-0.7877	-6.1227
	y_{10}	-2.6806	-1.3794	-5.6358
$E(CRPS)$		-1.5163	-2.9957	-6.1728
$MAFE$		3.5105	4.8270	20.5394
$MSFE$		22.2409	37.0315	665.1931

Table 2: Experiment 3: Predictive ability of the three competing models.
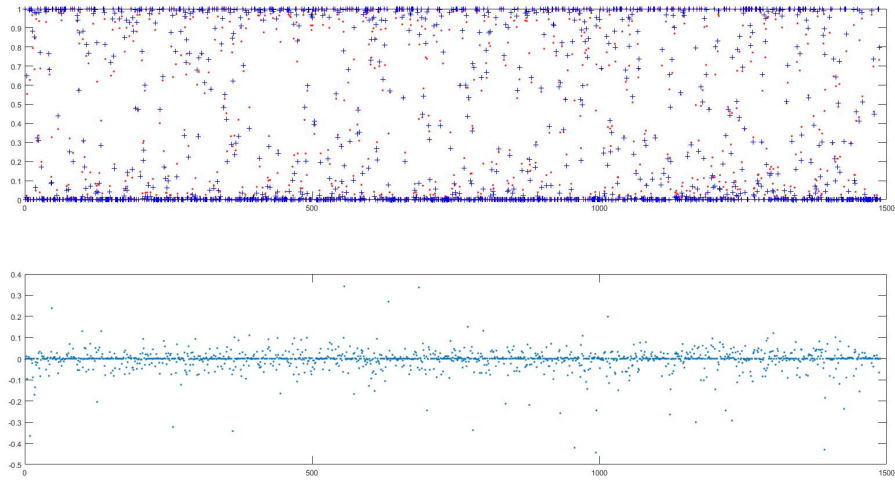
19

Figure 11: Experiment 4: True simulated transition probabilities vs mean estimated transition probabilities . Upper panel: True probabilities are marked with blue crosses and mean estimated probabilities are estimated with red dots. Lower panel: The difference between mean estimated probabilities for one estimated realized hidden process and the true simulated probabilities.
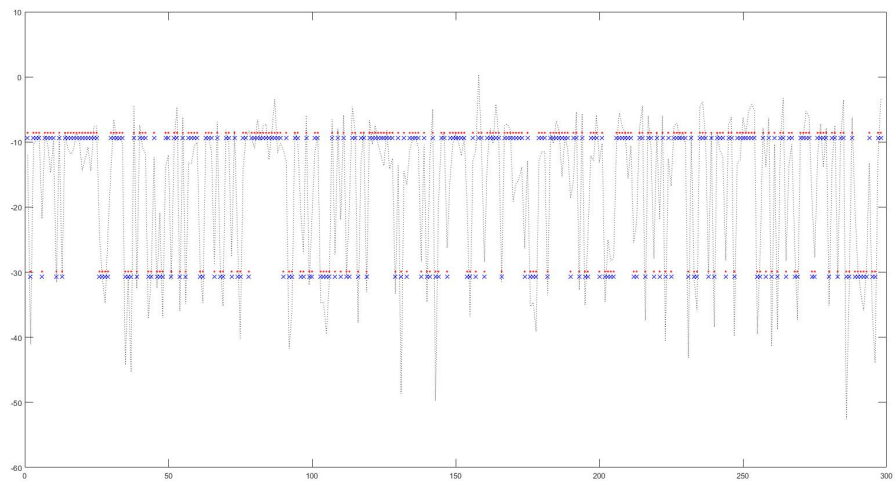

Figure 12: Experiment 4: Thinned observed process (black dotted line) and hidden process. True hidden states are marked with blue x and a realized simulated states are marked with red dots.

True model	
Included Covariates, Mean	X_1, X_2, X_3
Included Covariates, Transition Matrix	X_1, X_2, X_5, X_6, X_7
Total Covariates	$X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8$
Highest Posterior Probability model	
Included Covariates, Mean	X_1, X_2, X_3
Included Covariates, Transition Matrix	X_1, X_2, X_3, X_4, X_5
Posterior probability	0.9775
Median Probability model	
Included Covariates, Mean	X_1, X_2, X_3
Included Covariates, Transition Matrix	X_1, X_2, X_5, X_6, X_7

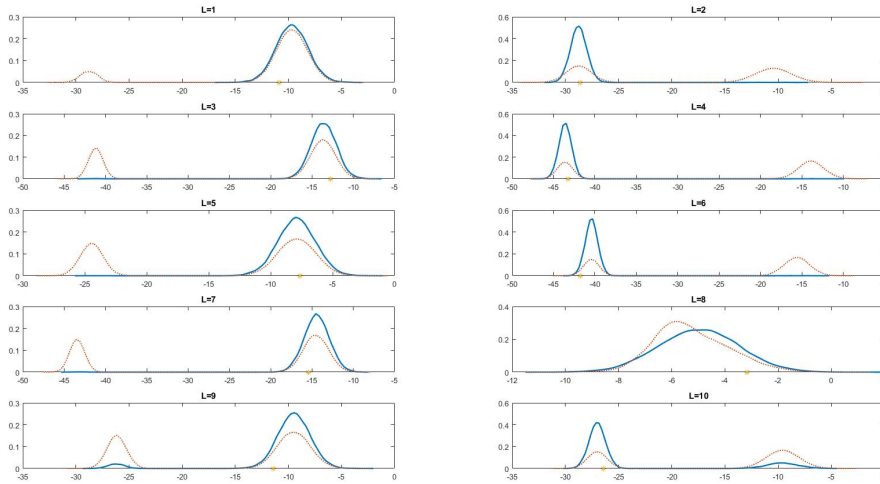Table 3: Experiment 4: Highest posterior probability model and median probability model.

Figure 13: Experiment 4: Empirical posterior predictive distribution for the 10 out-of-sample forecasts using the NHHMM (blue line) and the HHMM (red dotted line). Yellow asterisk represents the true simulated values.
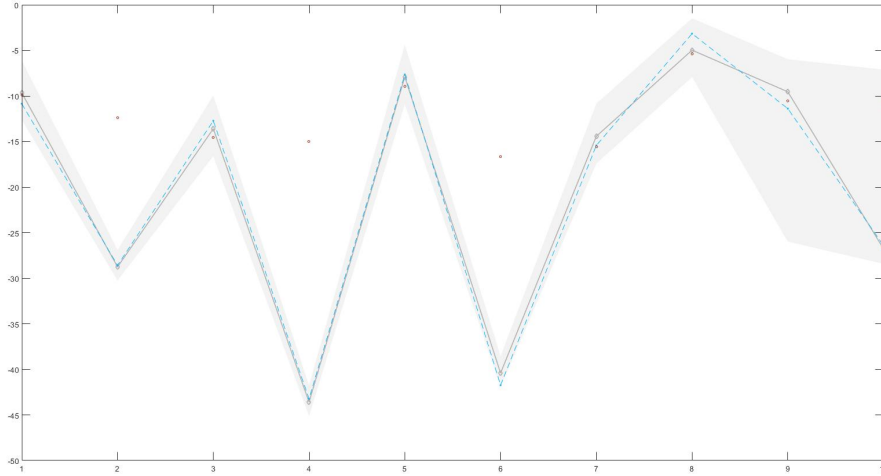
Figure 14: Quantiles of the predictive distribution. Ten one-step-look-ahead observed values are marked with dashed blue line. Gray bold line represents the median of the predictive distribution. The gray area around the median is defined by the 2.5% and 9.5% quantiles. Red circles represent the median values of the predictive distribution using the HHMM model.

Probabilistic Forecasting Criterion				
	y	NHHMM	HHMM	qAR
	y_1	-0.7132	-0.7163	-26.5950
	y_2	-0.1909	-5.5165	-26.2714
	y_3	-0.5462	-3.6384	-26.5136
	y_4	-0.2409	-11.7427	-26.4477
$CRPS(y_i)$	y_5	-0.3732	-2.3383	-26.3171
	y_6	-0.9319	-11.1786	-27.6324
	y_7	-0.6246	-3.6083	-26.1485
	y_8	-1.1246	-1.4325	-26.5387
	y_9	-1.1691	-2.2991	-26.6335
	y_{10}	-0.8270	-6.8545	-25.7379
$E(\bar{CRPS})$		-0.6742	-2.9957	-26.4565
$MAFE$		1.6932	9.6738	88.5193
$MSFE$		9.7859	211.4431	1.2×10^4

Table 4: Experiment 4: Predictive ability of the three competing models.

6 Empirical Application: Realized volatility data

Financial volatility has been extensively studied in the literature, due to the crucial role in various financial fields, such as pricing, risk managment, investment, asset allocation among others. Many models have been proposed, exploring different methodologies. A review on the realized volatility literature can be found in ÎIJichael McAleer and Medeiros (2008). There are

though only few studies in predicting realized stock volatility using various financial predictors (Mittnik et al. (2015); Meligkotsidou et al. (2015); Christiansen et al. (2012); Paye (2012)). We use the described model and methodology to asses the predictive ability of 13 financial variables in realized volatility data.

6.1   The data

We applied our method to Realized stock market Volatility data, as described in detail by Christiansen et al. (2012). Specifically, we use the 'long' sample of U.S. equity market of the S&P500. The realized volatility is the squared root of the realized variance for asset class i in month t us the sum of squared intra-period (daily) returns

$$RV_{i,t} = \sqrt{\sum_{\tau=1}^{M_t} r_{i,t,\tau}^2}, \ t = 1 \dots, T.$$

Where $r_{i,t,\tau}$ is the rth daily continuously compounded return of month t for asset i and M_t denotes the trading days during month t. Thus $\sum_{\tau=1}^{M_t} r_{i,t,\tau}^2$ is the realized variance for asset class i in month t. The distribution of the realized daily variances are highly non-normal and skewed to the right, but the logarithms of the realized variances are approximately normal and thus they have better behavior (see Andersen et al. (2003)). Hence, in the following analysis we study the natural logarithm of the realized volatility series,

$$RV_{i,t} = \ln \sqrt{\sum_{\tau=1}^{M_t} r_{i,t,\tau}^2}, \ t = 1 \dots, T.$$

The data are in a monthly basis, starting from December 1926 until October 2002. The out-of-sample forecast evaluation period is two years, i.e. 24 observations from October of 2000 until October 2002. In both simulations, for realized volatility and for log-realized volatility, we had a burn-in period of 20000 and we generated 20000 MCMC samples. We used non-informative priors for the unknown parameters $\sigma_s, B_s, \beta_{ss}, \quad s = 1, 2$, that is $\sigma_s^2 \sim \mathcal{IG}(0.1, 0.1)$, $B_s \mid \sigma_s^2 \sim \mathcal{N}(0, 1000\sigma_s^2)$ and finally $\beta_{ss} \sim \mathcal{N}(0, 100 \times I)$. Following Christiansen et al. (2012) and Meligkotsidou et al. (2015) we take into account 13 macroeconomic and financial predictive covariates. Particularly, from a list of equity market variables and risk factors we consider dividend price ratio (DP) and earnings price ratio (EP) (Welch and Goyal (2008)), lagged equity market returns (MKT), in order to capture the leverage effect, that is the asymmetric response of volatility to positive and negative returns (Nelson (1991)). We also use the risk factors of Fama and French (1993), that is, the size factor (SMB), value factor (HML), a short-term reversal factor (STR). From the set of interest rates, spreads and bond market factors we included the treasure bill rate (TBL), i.e. the interest rate on a three-month Treasure bill, the long-term return (LTR) on long-term government bonds, the term spread (TMS), i.e. the difference between the log-term yield and Treasure bill rate, the relative T-bill rate (RTB) as the difference between, T-bill rate and its 12-month moving average and the relative bond rate (RBR), as the difference between LTR and its 12 month moving average (Welch and Goyal (2008)). To proxy for weighted credit risk we also used the default spread (DEF) defined as the yield spread between BAA and AAA rated bonds. Lastly we consider the macroeconomic variable, inflation rate (INF), the monthly growth rate of CPI. The strong contemporaneous relation between the volatility and the business conditions implies that lagged volatility plays important role in forecasting (see Paye (2012)). We run a series of experiments for this data. Specifically we performed our analysis using the

predictors described and then we repeated the analysis using the predictors plus autoregressive terms (AR) of lag 1, 2 and lag 3 to investigate if there is additional predictive content of the macroeconomic and financial variables that go beyond the information contained in time-series history of volatility.

6.2 Results

Our in-sample analysis revealed some interesting results. Based on the posterior probabilities of inclusion we see that if we do not include any AR terms in the pool of the predictors, then the median NHHMM model has 9 predictors that affect the series linearly or not. When we added the AR(1) term, the included predictors in the median probability model immediately reduced in 6. Adding more AR terms (of lag 2 and lag 3), the median probability model remained almost the same as in the case of the model with one autoregressive term. Furthermore, in our out-of-sample analysis, we did not encounter any significant improvement in the forecasting ability of the models with AR(2) and AR(3) terms. We note that this result confirms the analysis of Christiansen et al. (2012), who also used only one AR term of lag 1 in their analysis. Thus, from hereafter even though based on the CRPS the model with the best performance was the one with AR(1) term we present both results of the model with no autoregressive terms (NHHMM$_0$) and the model with one autoregressive term (NHHMM$_1$), for the sake of completeness. Also, we compare our results with two benchmark models, the Homogeneous model with one autoregressive term (HHMM) and the quasi-AR model (qAR), as described in Section 5.

Figure 15 shows a plot of the realized volatility data (blue line), the probability of staying at the same stage (e.g. if at time t we are on state 1 then the red dot at time t shows the value of probability p_{11}) and the shaded bars represent the time period that the chain was in state 2, based on the smoothed probabilities of being above 0.5 for NHHMM$_1$. We observe that both states are highly persistent, i.e. probabilities of staying at the same state are high. Also, we present a thinned in-sample realization of the observed process inferred by our algorithm along with the real data in Figure 17.

6.2.1 Model Selection

The double reversible jump algorithm of our methodology, did not assign high probability to a specific model. Instead, we summarize in Table 6 (Section 7) the posterior probabilities of inclusion for each predictor, both for the mean equation and for the transition matrix for the NHHMM$_0$ and NHHMM$_1$. With our methodology we are not only able to identify which covariates affect the realized volatility series but to comment on how the covariates affect the series (linearly or not). Based on the median model of NHHMM$_0$ we found that predicting realized volatility series can be improved using several predictors. Earnings price ratio (EP), lagged equity market returns (MKT), value factor HML, relative T-bill rate (RTB), long-term goverment bonds (LTR), relative T-bill rate (RBR) affect the series linearly, term spread (TMS), difference between LTR and TBL (RBR),inflation (INF) affect the series non linearly and term spread (TMS), default spread (DEF) affect the series in a linear and non-linear fashion. The median proabability NHHMM with AR(1) term NHHMM$_1$ has notably less predictors affecting the RV-series. Specifically we found that MKT affect the series linearly, TBL,RTB,LTR,TMS affect the series non-linearly and DEF both linearly and non linearly. Our analysis agrees with the analysis of Christiansen et al. (2012) who find that MKT, TBL,LTR, DEF improve volatility forecasts, but we also find significant two more predictors (TMS, RTB), whereas we do not include the variables STR and LTR that Christiansen et al. (2012) found also significant.

6.2.2 Forecasting

As far as forecasting, we report in Table 5 the values of the forecasting criteria that we used for all the competing models. We confirm that $NHHMM_1$ performs better than all the other models, since it has higher scores in Mean Ranked Probability Score (E(CPRS)) and also in MAFE and MSFE. Here we would like to remark that eventhough the $NHHMM_0$ performs worse than all the other models, it can predict the moves of the series better.
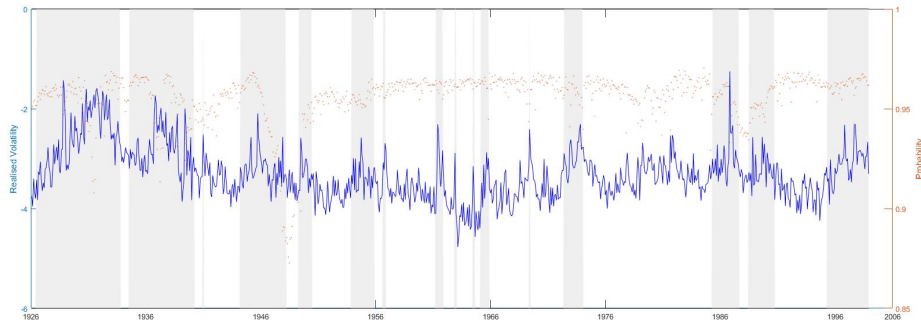

Figure 15: Time series of the monthly realized volatility of the Standard & Poor ($S\&P$) 500 index (in logarithmic scale) for the period 1926-2002 ,using the $NHHMM_0$ (top figure) and $NHHMM_1$ (lower figure). Red dots are the posterior mean probabilities of staying at the same stage. We calculated the smoothed probability of staying at state 1. If the probability is above 0.5 we marked this event with gray-shaded bar.
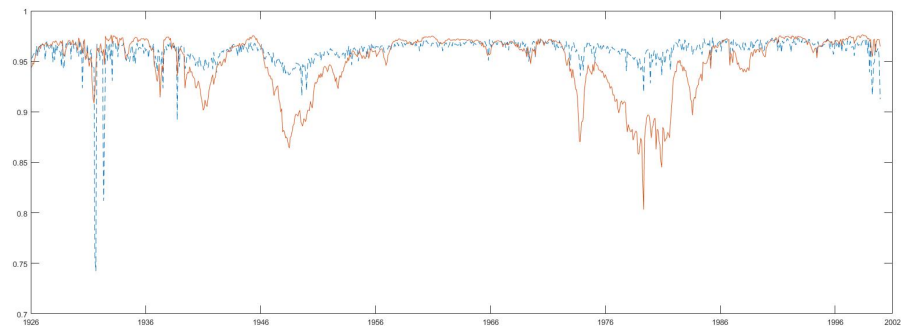

Figure 16: This figure shows the probabilities of staying at state 1 (blue solid line) and the probabilities of staying at state 2 (red dashed line).
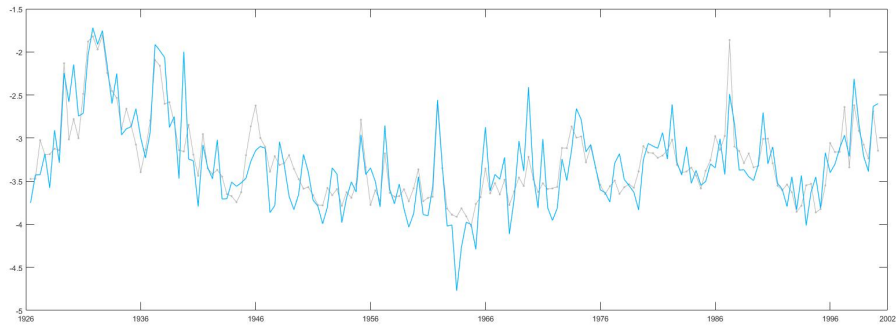
Figure 17: This figure presents a thinned realization of the observed process (1:5 observations) as calculated using our NHHM model (gray solid line). For comparison we also present the thinned time series of the studied variable.

Probabilistic Forecasting Criterion				
	$NHHMM_0$	$NHHMM_1$	HHMM	qAR
$E(CRPS)$	-0.2388	-0.1474^*	-0.1745	-0.1822
$MAFE$	0.4873	0.2398^*	0.3465	0.3035
$MSFE$	0.3737	0.0914^*	0.1821	0.1652

Table 5: Out-of-Sample results: Predictive ability of the competing models ($NHHMM_0$,$NHHMM_1$, HHMM, qAR(3)) for the log-realized volatility data. For each of the competing models the table presents the mean Continuous Ranked Probability Score E(CRPS). Also Mean Absolute Forecasting Error and Mean Squared Forecasting Error are reported. Better performance by means of forecasting criteria is marked with asterisks.
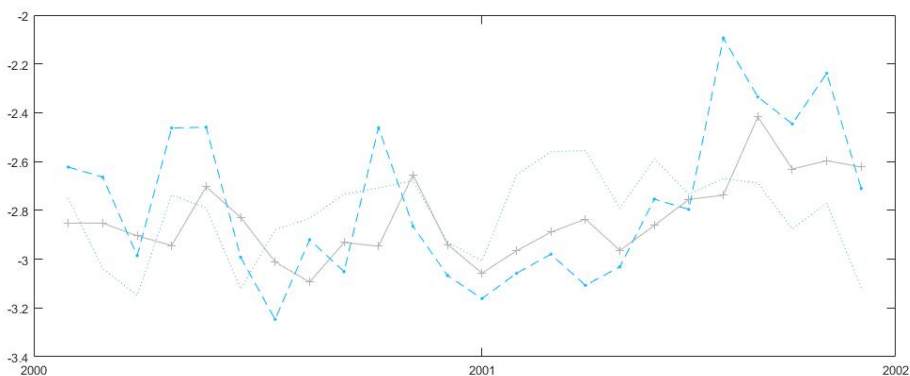

Figure 18: This figure shows the out-of-sample forecasting ability of the $NHHMM_0$ (dotted line) and $NHHMM_1$ (gray solid line). The real values of the realized volatility series is marked with blue dashed line.

	NHHM_0		NHHM_1		HHM	AR
Posterior probabilities of inclusion						
Covariates	Mean Equation	Transition Matrix	Mean Equation	Transition Matrix	Mean Equation	Mean Equation
DP	0.0016	0.0452	0.0051	0.0592	0.0099	0.0083
EP	1.0000*	0.0629	0.4709	0.2565	0.9991*	0.0182
MKT	1.0000*	0.3372	0.9867	0.3162	1.0000*	0.3297
SMB	0.0182	0.4228	0.0153	0.4452	0.1444	0.0367
HML	0.6249*	0.4304	0.0156	0.4397	0.6519*	0.0297
STR	0.0013	0.4322	0.0908	0.4438	0.0059	0.4087
TBL	0.0173	0.3896	0.0124	0.5300*	0.0581	0.0466
RTB	0.9733*	0.4663	0.0786	0.5145*	0.8265*	0.1485
LTR	1.0000*	0.4799	0.0494	0.5014*	0.9431*	0.1467
RBR	1.0000*	0.4731	0.0589	0.4860	0.9448*	0.1119
TMS	0.9241*	0.5082*	0.0517*	0.5018*	0.6313	0.0961
DEF	1.0000*	0.5102*	1.0000*	0.5025*	1.0000*	1.0000*
INF	0.4130	0.5074*	0.3018	0.4964	0.5444*	0.1669

Table 6: Posterior probabilities of inclusion for the competing models. Predictors with inclusion probability above 0.5 (median probability model) are marked with asterisks.

7 Appendix

7.1 Simulation study plots

In this section we present tables and figures of the case-studies described in Section 5.

States	True Values	Posterior Mean	Std	2.5%	Median	97,5 %
$s = 1$	$b_0 = 3$	3.5566	0.3282	2.9187	3.5566	4.2020
	$b_1 = -4$	-4.0533	0.0481	4.1483	-4.0533	-3.9592
	$b_2 = 3$	2.9584	0.0491	2.8619	2.9584	-3.0544
	$b_3 = 1$	1.0855	0.0978	0.8929	1.0855	1.2748
	$b_4 = 2$	1.8942	0.0954	1.7083	1.8942	2.0809
	$\sigma^2 = 1.3$	1.4228	0.0829	1.2709	1.4202	1.5947
	$\beta_0 = 2$	3.1800	1.8851	-0.4722	3.1685	6.9587
	$\beta_1 = 4$	3.7009	0.4236	2.9081	3.6835	4.3753
	$\beta_2 = 1$	0.7587	0.3742	0.0439	0.7527	1.5111
	$\beta_3 = 2$	1.1884	0.2659	1.3518	1.8290	2.3900
	$\beta_4 = -1$	-1.1579	0.2074	-1.5772	-1.1528	-0.7683
	$\beta_5 = 3$	3.2933	0.3852	2.5767	3.2777	4.0901
$s = 2$	$b_0 = 5$	4.8797	0.2422	-3.5162	4.8797	5.3552
	$b_1 = 1$	1.0220	0.0382	0.9467	1.0220	1.0971
	$b_2 = 3$	2.9519	0.0366	2.8802	2.9519	3.0234
	$b_3 = 2$	1.9104	0.0688	1.7739	1.9104	2.0448
	$b_4 = 4$	3.9987	0.0744	3.8555	3.9987	4.1465
	$\sigma^2 = 0.8$	0.7598	0.0446	0.6766	0.07589	0.8515
	$\beta_0 = -2$	0.8237	2.2179	-3.5296	0.8211	5.2151
	$\beta_1 = 3$	3.0154	0.3732	2.3360	2.9994	3.7981
	$\beta_2 = 2$	2.0436	0.4490	1.1941	2.0327	2.9592
	$\beta_3 = 4$	4.2320	0.4990	3.3177	4.2092	5.2799
	$\beta_4 = 1$	0.6664	0.1944	0.2906	0.6627	1.0621
	$\beta_5 = -2$	-2.4036	0.3300	-3.0887	-2.3927	-1.7888

Table 7: The fixed model: Summary statistics for the explanatory variables of experiment 1.

Out-of-sample		Forecasting Criteria		
y		NHHMM	HHMM	qAR
$y_1 = -1.3404$		-1.5227	-2.7418	-22.8871
$y_2 = 25.0334$		-0.9262	-2.9887	-22.9191
$y_3 = 3.7751$		-0.9208	-2.2213	-21.9447
$y_4 = 12.4843$	$CRPS(y_i)$	-0.3128	-4.4827	-25.5160
$y_5 = -7.8430$		-5.5556	-6.8169	-23.7165
$y_6 = 13.8031$		-3.2179	-4.7337	-23.2957
$y_7 = 1.4307$		-2.4628	-2.9928	-22.7315
$y_8 = -11.1139$		-0.3740	-7.4199	-22.5383
$y_9 = 16.5539$		-0.8481	-3.6279	-22.6431
$y_1 0 = 24.4482$		-1.2314	-6.3935	-22.7568
	$E(CRPS)$	-1.7372	-4.4419	-22.7949
	$MAFE$	3.9173	9.3282	77.2418
	$MSFE$	62.0849	187.9507	9.4×10^4

Table 8: Comparing the forecasting ability of the three competing models of experiment 1.
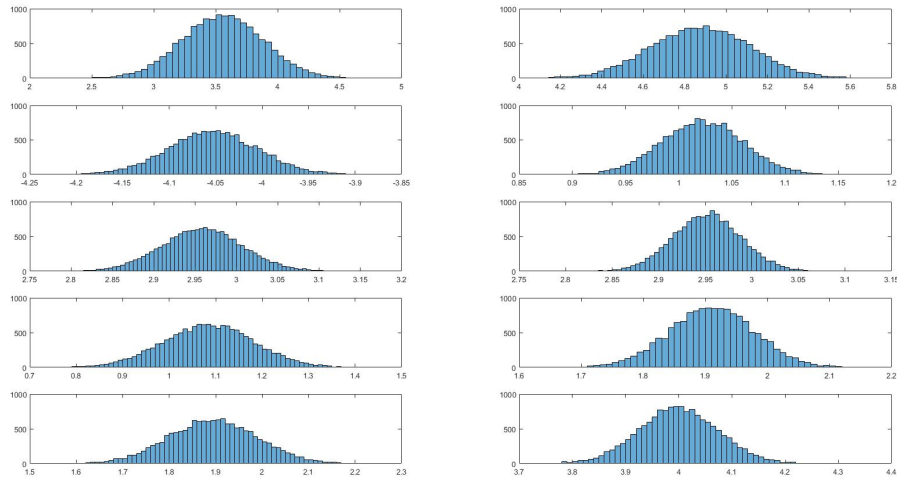

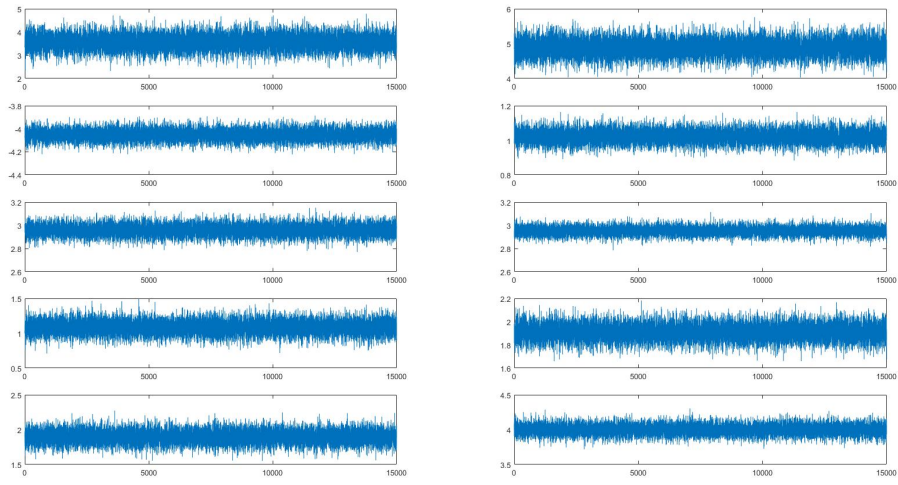Figure 19: Experiment 1: Histograms of the mean equation parameters

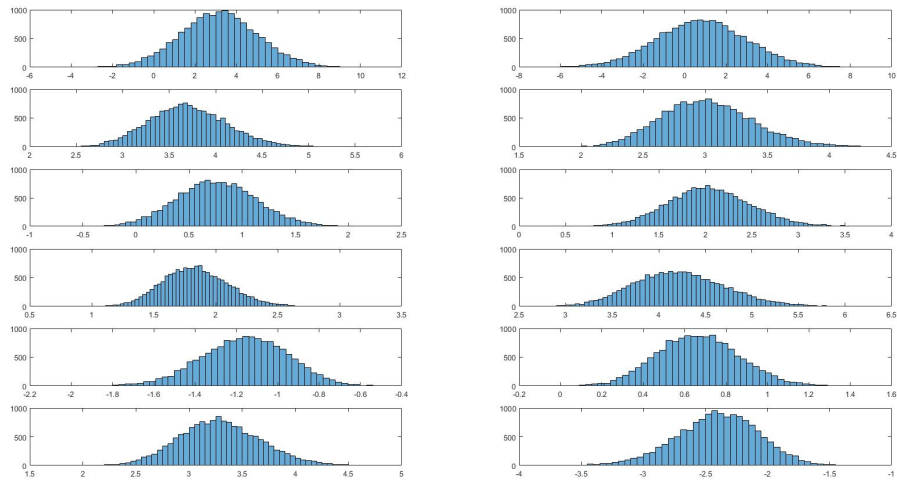Figure 20: Experiment 1: Convergence plots of the mean equation parameters

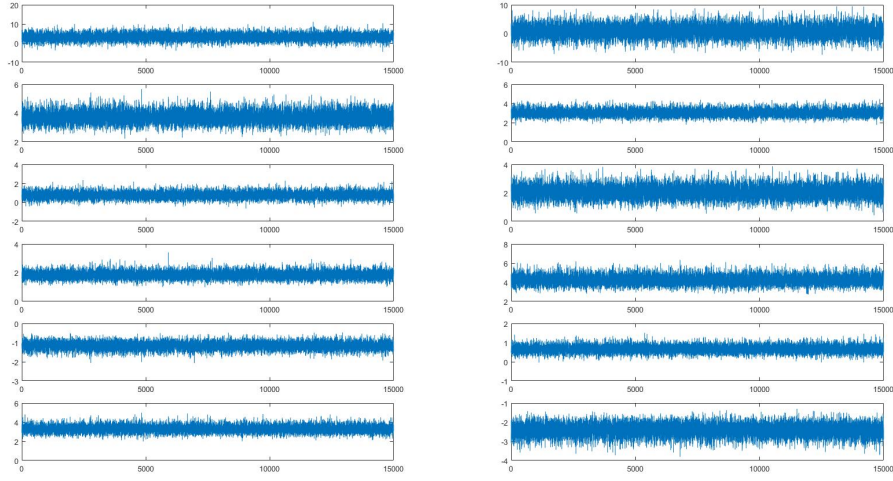Figure 21: Experiment 1: Histograms of the logistic regression coefficients

Figure 22: Experiment 1: Convergence plots of the logistic regression coefficients

States	True Values	Posterior Mean	Std	2.5%	Median	97,5 %
$s = 1$	$b_0 = 2$	1.7134	0.3147	1.0985	1.7147	2.3336
	$b_1 = -0.3$	-0.2569	0.0538	-0.3632	-0.2572	-0.1536
	$b_2 = 2$	2.0721	0.0501	1.9730	2.0729	2.1696
	$b_3 = 2$	2.0488	0.0975	1.8604	2.0481	2.2406
	$\sigma^2 = 1.5$	1.0245	0.0822	1.2312	1.375	1.5559
	$\beta_0 = 1.5$	1.3815	1.1805	0.04722	2.2870	4.6590
	$\beta_1 = 1$	1.0364	0.1902	0.6770	1.0314	1.4203
	$\beta_2 = 2$	2.0923	0.2343	1.6540	2.0840	2.5682
	$\beta_3 = 3$	3.2644	0.3325	2.6567	3.2515	3.9610
$s = 2$	$b_0 = 1$	1.1955	0.1878	0.8263	1.1935	1.5650
	$b_1 = 3$	2.9608	0.0298	2.9018	2.9608	3.0192
	$b_2 = 4$	3.9925	0.0302	3.9333	3.9927	4.0514
	$b_3 = 3$	3.0312	0.0595	2.9142	3.0315	3.1471
	$\sigma^2 = 0.8$	0.7635	0.0359	0.6968	0.7626	0.8369
	$\beta_0 = 3$	3.5477	0.9022	1.8020	3.5363	5.3628
	$\beta_1 = -2.5$	-2.5752	0.2103	-3.0022	-2.5683	-2.1785
	$\beta_2 = 4$	3.9736	0.2917	3.4352	3.9674	4.5709
	$\beta_3 = 1$	1.0602	0.1405	0.7397	1.0573	1.3453

Table 9: The fixed model. Summary statistics for the explanatory variables of the experiment 2 using the NHMM model.

Out-of-sample		Forecasting Criteria		
y		NHHMM	HHMM	qAR
$y_1 = 19.4094$		-0.2674	-0.8738	-27.3658
$y_2 = 2.6234$		-0.6003	-4.9020	-26.5102
$y_3 = 12.3667$		-0.4427	-1.7485	-28.4347
$y_4 = 13.3616$	$CRPS(y_i)$	-0.2889	-1.7146	-26.5481
$y_5 = 22.12014$		-0.4673	-2.5203	-27.2373
$y_6 = -0.1348$		-1.2171	-4.1459	-26.6438
$y_7 = 12.7453$		-0.3824	-1.7933	-26.6438
$y_8 = 0.9820$		-1.0587	-10.5293	-26.6438
$y_9 = 20.1527$		-0.1921	-2.6289	-26.6438
$y_{10} = 0.3442$		-1.2822	-8.0522	-26.6438
	$E(CRPS)$	-0.6199	-3.8929	-27.1233
	$MAFE$	1.8977	7.56560	91.8549
	$MSFE$	15.0261	123.1742	1.3255×10^4

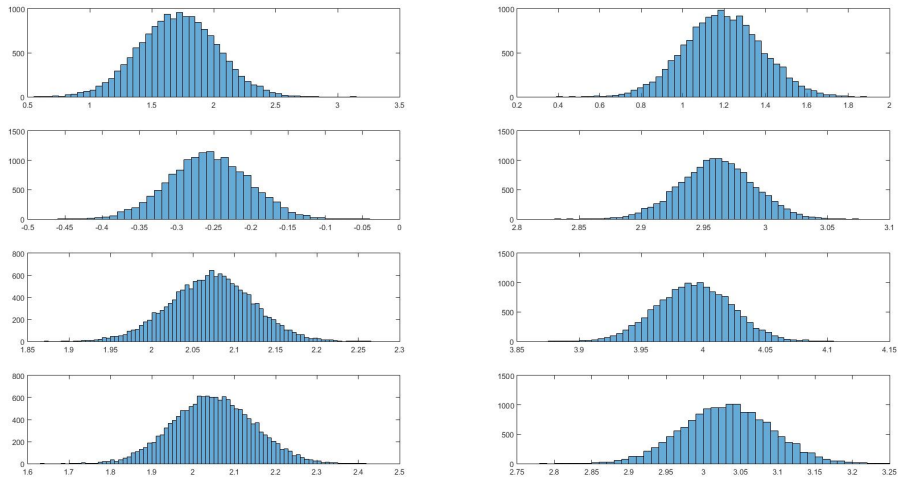Table 10: Comparing the forecasting ability of the three competing models of experiment 2

Figure 23: Experiment 2: Histograms of the mean equation parameters
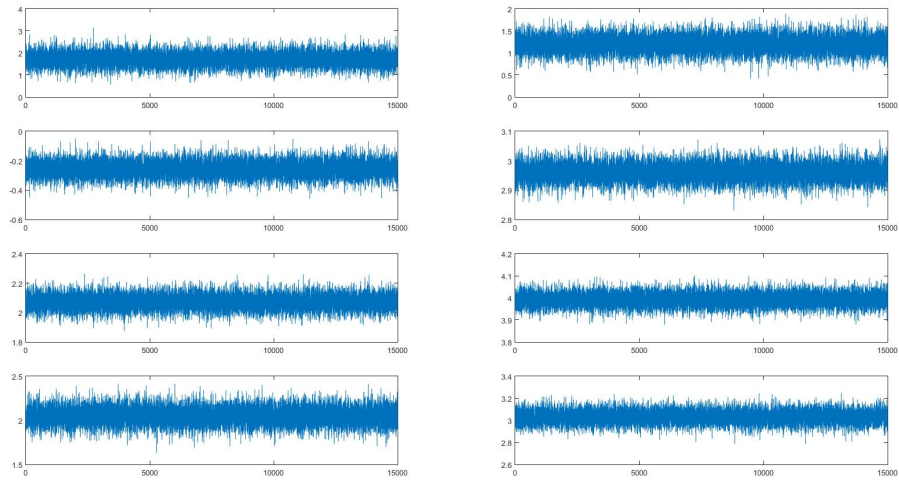
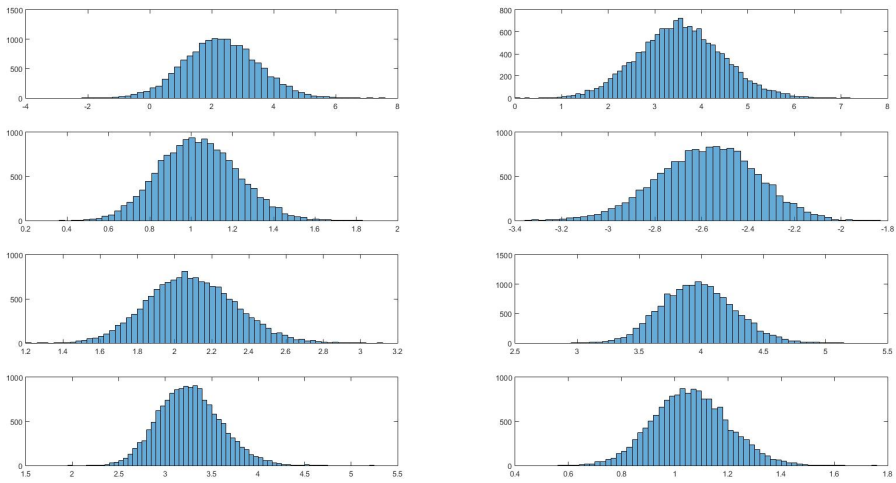Figure 24: Experiment 2:Convergence plots of the mean equation parameters

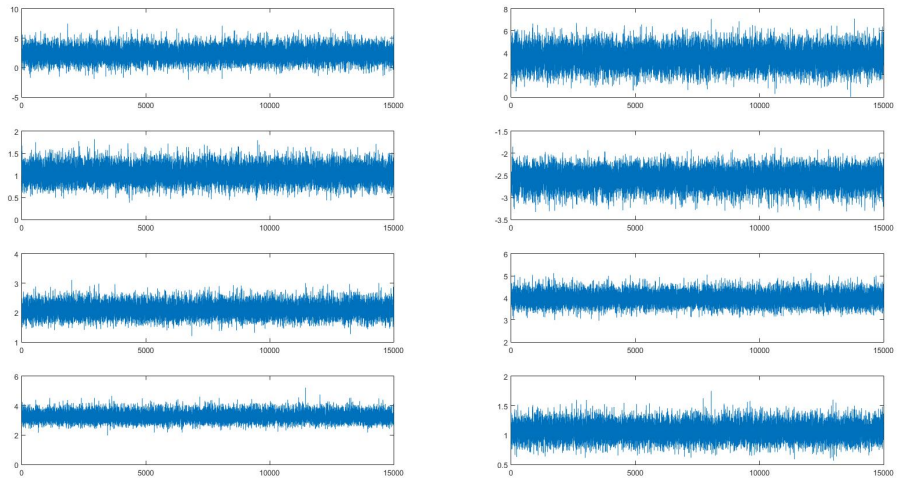Figure 25: Experiment 2: Histograms of the logistic regression coefficients

33

Figure 26: Experiment 2: Convergence plots of the logistic regression coefficients

References

H. Akaike, B. Petrov, and F. Csaki. Information theory and an extension of the maximum likelihood principle. *Proceedings of the Second International Symposium on Information Theory*, pages 267–281, 1973.

T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–62, 2003. doi:10.1111/1468-0262.00418.

M. M. Barbieri and J. O. Berger. Optimal predictive model selection. *The Annals of Statistics*, 32(3):870–897, 2004. doi:10.1214/009053604000000238.

L. E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The Annals of Mathematical Statistics*, 41(1):164–171, 1970. doi:10.1214/aoms/1177697196.

J. Besag, P. Green, D. Higdon, and K. Mengersen. Bayesian computation and stochastic systems. *Statistical Science*, 10(1):3–41, 1995. doi:10.2307/2246224.

M. Billio, A. Monfort, and C. Robert. Bayesian estimation of switching arma models. *Journal of Econometrics*, 93(2):229 – 255, 1999. ISSN 0304-4076. doi:10.1016/S0304-4076(99)00010-X.

G. Boero, J. Smith, and K. F. Wallis. Scoring rules and survey density forecasts. *International Journal of Forecasting*, 27(2):379 – 393, 2011. doi:10.1016/j.ijforecast.2010.04.003.

O. Cappé, E. Moulines, and T. Ryden. *Inference in Hidden Markov Models (Springer Series in Statistics)*. Springer-Verlag New York, Inc., 2005. ISBN 0387402640.

C. Christiansen, M. Schmeling, and A. Schrimpf. A comprehensive look at financial volatility prediction by economic variables. *Journal of Applied Econometrics*, 27(6):956–977, 2012. doi:10.1002/jae.2298.

P. Dellaportas, J. Forster, and I. Ntzoufras. On bayesian model and variable selection using mcmc. *Statistics and Computing*, 12(1):27–36, 2002. doi:10.1023/A:1013164120801.

F. X. Diebold, J.-H. Lee, and G. C. Weinbach. *Regime switching with time-varying transition probabilities. In: Hargreaves, C.P. (ed.) Nonstationary Time Series Analysis and Cointegration.* Oxford: Oxford University Press, 1994.

E. F. Fama and K. R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3 – 56, 1993. doi:10.1016/0304-405X(93)90023-5.

A. J. Filardo and S. F. Gordon. Business cycle durations. *Journal of Econometrics*, 85(1):99 – 123, 1998. doi:10.1016/S0304-4076(97)00096-1.

S. Frühwirth-Schnatter and R. Frühwirth. *Data Augmentation and MCMC for Binary and Multinomial Logit Models.* Physica-Verlag HD, 2010. doi:10.1007/978-3-7908-2413-1_7.

A. Fussl, S. Frühwirth-Schnatter, and R. Frühwirth. Efficient mcmc for binomial logit models. *ACM Trans. Model. Comput. Simul.*, 23(1):3:1–3:21, 2013. doi:10.1145/2414416.2414419.

A. Gelman, J. Hwang, and A. Vehtari. Understanding predictive information criteria for bayesian models. *Statistics and Computing*, 24(6):997–1016, 2014. doi:10.1007/s11222-013-9416-2.

J. Geweke and C. Whiteman. Chapter 1 bayesian forecasting. In G. Elliott, C. Granger, and A. Timmermann, editors, *Handbook of Economic Forecasting*, volume 1, pages 3 – 80. Elsevier, 2006. doi:10.1016/S1574-0706(05)01001-3.

T. Gneiting. Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762, 2011. doi:10.1198/jasa.2011.r10138.

T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. doi:10.1198/016214506000001437.

P. J. Green. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732, 1995. doi:10.1093/biomet/82.4.711.

P. J. Green and D. I. Hastie. Reversible jump mcmc. *Genetics*, 155(3):1391–1403, 2009.

S. GschlÃűÃ§l and C. Czado. Spatial modelling of claim frequency and claim size in non-life insurance. *Scandinavian Actuarial Journal*, 2007(3):202–225, 2007. doi:10.1080/03461230701414764.

H. Hersbach. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15(5):559–570, 2000. doi:10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.

C. C. Holmes and L. Held. Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis*, 1(1):145–168, 2006. doi:10.1214/06-BA105.

T. Holsclaw, A. M. Greene, A. W. Robertson, and P. Smyth. Bayesian nonhomogeneous markov models via pólya-gamma data augmentation with applications to rainfall modeling. *The Annals of Applied Statistics*, 11(1):393–426, 03 2017. doi:10.1214/16-AOAS1009.

ÎŁIJichael McAleer and M. C. Medeiros. Realized volatility: A review. *Econometric Reviews*, 27 (1-3):10–45, 2008. doi:10.1080/07474930701853509.

R. L. Machete. Contrasting probabilistic scoring rules. *Journal of Statistical Planning and Inference*, 143(10):1781 – 1790, 2013. doi:10.1016/j.jspi.2013.05.012.

L. Meligkotsidou and P. Dellaportas. Forecasting with non-homogeneous hidden markov models. *Statistics and Computing*, 21(3):439–449, 2011. doi:10.1007/s11222-010-9180-5.

L. Meligkotsidou, E. Panopoulou, I. D. Vrontos, and S. D. Vrontos. Quantile forecast combinations in realised volatility prediction. *SSRN Electronic Journal*, 2015. doi:10.2139/ssrn.2673529.

S. Mittnik, N. Robinzonov, and M. Spindler. Stock market volatility: Identifying major drivers and the nature of their impact. *Journal of Banking* & *Finance*, 58:1 – 14, 2015. ISSN 0378-4266. doi:10.1016/j.jbankfin.2015.04.003.

D. B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370, 1991. ISSN 00129682, 14680262. doi:10.2307/2938260.

S. M. O'Brien and D. B. Dunson. Bayesian multivariate logistic regression. *Biometrics*, 60(3): 739–746, 2004. doi:10.1111/j.0006-341X.2004.00224.x.

R. O'Hara and M. J. Sillanpää. A review of bayesian variable selection methods: what, how and which. *Bayesian Analysis*, 4(1):85–117, 2009. doi:10.1214/09-BA403.

B. S. Paye. âĂŸdÃŋjÃă volâĂŹ: Predictive regressions for aggregate stock market volatility using macroeconomic variables. *Journal of Financial Economics*, 106(3):527 – 546, 2012. doi:10.1016/j.jfineco.2012.06.005.

N. G. Polson, J. G. Scott, and J. Windle. Bayesian inference for logistic models using polyagamma latent variables. *Journal of the American Statistical Association*, 108(504):1339–1349, 2013. doi:10.1080/01621459.2013.829001.

T. Rydén. Em versus markov chain monte carlo for estimation of hidden markov models: a computational perspective. *Bayesian Analysis*, 3(4):659–688, 2008. doi:10.1214/08-BA326.

G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 03 1978. doi:10.1214/aos/1176344136.

S. L. Scott. Bayesian methods for hidden markov models. *Journal of the American Statistical Association*, 97(457):337–351, 2002. doi:10.1198/016214502753479464.

L. Spezia. Bayesian analysis of non-homogeneous hidden markov models. *Journal of Statistical Computation and Simulation*, 76(8):713–725, 2006. doi:10.1080/10629360500108798.

D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, 2002. doi:10.1111/1467-9868.00353.

G. J. SzÃŋkely and M. L. Rizzo. A new test for multivariate normality. *Journal of Multivariate Analysis*, 93(1):58 – 80, 2005. doi:10.1016/j.jmva.2003.12.002.

R. Waagepetersen and D. Sorensen. A tutorial on reversible jump mcmc with a view toward applications in qtl-mapping. *International Statistical Review*, 69(1):49–61, 2001. doi:10.1111/j.1751-5823.2001.tb00479.x.

S. Watanabe. A widely applicable bayesian information criterion. *J. Mach. Learn. Res.*, 14(1): 867–897, 2013. ISSN 1532-4435. URL http://dl.acm.org/citation.cfm?id=2567709.2502609.

I. Welch and A. Goyal. A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4):1455–1508, 2008. doi:10.1093/rfs/hhm014.