An Infinite Hidden Markov Model with Stochastic Volatility

Li, Chenxing and Maheu, John M and Yang, Qiao

25 November 2022

An Infinite Hidden Markov Model with Stochastic Volatility*

Chenxing Li† John M. Maheu‡ Qiao Yang§

November 2022

Abstract

This paper extends the Bayesian semiparametric stochastic volatility (SV-DPM) model of Jensen and Maheu (2010). Instead of using a Dirichlet process mixture (DPM) to model return innovations, we use an infinite hidden Markov model (IHMM). This allows for time variation in the return density beyond that attributed to parametric latent volatility. The new model nests several special cases as well as the SV-DPM. We also discuss posterior and predictive density simulation methods for the model. Applied to equity returns, foreign exchange rates, oil price growth and industrial production growth, the new model improves density forecasts, compared to the SV-DPM, a stochastic volatility with Student-t innovations and other fat-tailed volatility models.

Keywords: stochastic volatility; Markov-switching; MCMC; Bayesian; nonparametric; semiparametric

JEL codes: C58; C14; C32; C11; C34

*We are grateful for the helpful comments from the editor, associate editor, two anonymous referees, seminar participants at the 2018 RECA Bayesian Econometric workshop, 2020 virtual SBIES, Tongji University, Hunan University, 2021 virual CFE, 2021 virtual China Meeting of Econometrics Society. Maheu thanks the SSHRC of Canada for their financial support. Yang thanks the start-up fund of ShanghaiTech and the Young Scientists Fund of NSFC (Project 72103137) for their financial support.

†Center for Economics, Finance and Management Studies, Hunan University. Email:lichenxing@hnu.edu.cn

‡DeGroote School of Business, McMaster University. Email:maheujm@mcmaster.ca

§Corresponding author: School of Entrepreneurship and Management, ShanghaiTech University, China. Email: yangqiao@shanghaitech.edu.cn

1 Introduction

Changing volatility has become ubiquitous in economic time-series data. Besides high frequency asset returns, conditional heteroskedasticity is even found in lower frequency macroeconomic aggregate data (Chan, 2013, 2017; Marcellino et al., 2016; Diebold et al., 2017; Carriero et al., 2019). Generalized autoregressive conditional heteroskedasticity (GARCH, Bollerslev, 1986) and stochastic volatility (SV, Taylor, 1982) are popular modelling approaches used to capture volatility dynamics. However, much less attention has been paid to modelling unknown return innovation distributions.

Flexible modelling of return innovations coupled with parametric volatility models can be found in the work of Jensen and Maheu (2010), Delatola and Griffin (2011, 2013), Kalli et al. (2013) and Liu (2021). Although flexible, these approaches assume that the underlying innovation distribution is constant over time. Volatility changes from the parametric portion of the model, but the underlying return distribution is fixed over time.

This paper explores a SV parametric specification, coupled with an infinite hidden Markov component that governs a mixture of normals. This is a direct extension from Jensen and Maheu (2010) and replaces the Dirichlet process mixture (DPM) with a Markov mixture model. The Markov chain allows for the possibility that the weights on the mixture change over time. In theory this means that the mixture can capture changing conditional skewness, kurtosis as well as changes in the tail dynamics beyond what the SV component can account for.

The infinite hidden Markov model (IHMM) has been fruitfully used in other settings: GARCH modelling (Dufays, 2016), inflation dynamics (Song, 2014; Jochmann, 2015) short-term interest rates (Maheu and Yang, 2016), realized covariance models (Jin and Maheu, 2016; Jin et al., 2019), macroeconomic forecasting (Hou, 2017; Yang, 2019) and model combination (Jin et al., 2022).

The IHMM approximates the unknown conditional return distribution that is nonparametrically similar to the DPM. Unlike the DPM model, the mixture weights in the IHMM are Markovian. The prior on this Markov chain is constructed using two layers of nested Dirichlet processes referred to as a hierarchical Dirichlet process (Teh et al., 2006). The IHMM can be seen as a regime-switching model with an infinite number of states. In each period, the return distribution is approximated by an infinite mixture and the mixture weights depend on the previous state the system is in. In contrast, the DPM approximates the unknown distribution with an infinite mixture, but the weights are constant and independent of the previous states.

Due to the unbounded state space, the IHMM can accommodate both structural breaks and recurrent changes in a unified framework. However, a regime switching model may not be able to capture the strong persistence in the volatility dynamics (Ryden et al., 1998). Our model's SV component captures this and allows the IHMM component to focus on transitory changes in the shape of the unknown distributions.

Our infinite hidden Markov model with stochastic volatility (SV-IHMM), is related to Virbickaitė and Lopes (2019), which has a two-state Markov switching process that affects the conditional mean of log-volatility, while log-squared returns are nonparametrically modelled. The SV-IHMM allows unbounded states for the conditional mean of log-volatility but nonparametrically models return innovations without losing the sign information of returns. Related work that includes discrete parameter changes in volatility modelling include Maheu and McCurdy (2000), Calvet and Fisher (2004), Griffin and Steel (2011) and Bauwens et al. (2014).

Estimation relies on Markov chain Monte Carlo (MCMC) methods. Posterior simulation for the IHMM component comes from Teh et al. (2006) and Maheu and Yang (2016); while the latent stochastic volatility is simulated with the random block sampler of Jensen and Maheu (2010). We apply the model to several different asset classes and compare it with a number of strong benchmark models, including the SV-DPM from Jensen and Maheu (2010) and the SV model with Student-t innovations. While the SV component of the model captures movements that display strong persistence in volatility, the variance component directed from the IHMM portion can be thought of as capturing transitory changes in volatility that could be labelled as jumps. In all applications we find significant evidence of parameter change.

Evaluating forecasts through the predictive likelihood shows that the SV-IHMM is preferred to all other benchmarks. Predictive density plots indicate that the SV-IHMM tends to produce distributions with the fattest tails, when necessary. Comparison of tail forecasts, in the form of value-at-risk and expected shortfall confirm our model's superior performance.

This paper is organized as follows. Section 2 illustrates the specification of the proposed SV-IHMM, along with the sampling algorithm and density forecast computation. Section 3 lists the benchmark models for comparison. Section 4 extensively investigates the model's empirical performance with real world data. Section 5 concludes. An Appendix details the posterior simulation methods used for our model and some benchmark specifications.

2 SV-IHMM

2.1 Model Specification

Our proposed SV-IHMM model includes a parametric SV component and a Bayesian non-parametric portion, following an infinite hidden Markov model (IHMM). The IHMM is constructed from the hierarchical Dirichlet process (HDP) introduced by Teh et al. (2006). Let r_t, denote returns and h_t log-volatility then the hierarchical representation of the SV-IHMM is[1]

$$\Gamma \sim Stick(\eta), \quad \Pi_j \overset{iid}{\sim} Stick2(\alpha, \Gamma), \quad j = 1, \ldots, \infty, \tag{1a}$$

$$s_t | s_{t-1} \sim \Pi_{s_{t-1}}, \tag{1b}$$

$$r_t | s_t, h_t, \theta \sim N(\mu_{s_t}, \omega_{s_t}^2 \exp(h_t)), \tag{1c}$$

$$h_t | h_{t-1} \sim N(\phi h_{t-1}, \sigma_v^2), \tag{1d}$$

$$\theta_j \overset{iid}{\sim} \mathcal{H}, \quad j = 1, \ldots, \infty, \tag{1e}$$

for $t = 1, \ldots, T$. $\theta_{s_t} = \{\mu_{s_t}, \omega_{s_t}\}$ and $\theta = \{\theta_1, \theta_2, ...\}$ is the collection of the state-dependent parameter vectors that are generated from the base measure $\mathcal{H}$. $s_t \in \{1, \ldots, \infty\}$ is the state variable that is governed by the first-order Markov chain of infinite dimension with transition matrix Π. $Stick(\eta)$ and $Stick2(\alpha, \Gamma)$ are stick-breaking representations of the Dirichlet processes (Sethuraman, 1994; Teh et al., 2006). Let $\Gamma = \{\gamma_1 \ldots, \gamma_\infty\}$ then $\Gamma \sim Stick(\eta)$ denotes a discrete distribution with weights generated as

$$\gamma_j = v_j \prod_{l=1}^{j-1} (1 - v_l), \quad v_j \overset{iid}{\sim} Beta(1, \eta), \ j = 1, 2, 3..... \tag{2}$$

Γ serves as a centring distribution. Each row of Π is drawn as $\Pi_j \sim Stick2(\alpha, \Gamma)$. The distribution of Π_j has weights generated as

$$\pi_{ji} = \hat{\pi}_{ji} \prod_{l=1}^{i-1} (1 - \hat{\pi}_{jl}), \quad \hat{\pi}_{ji} \overset{iid}{\sim} Beta\left(\alpha\gamma_i, \alpha\left(1 - \sum_{l=1}^{i} \gamma_l\right)\right), \tag{3}$$

where π_{ji} is an element of Π at the jth row and ith column. π_{ji} represents the probability of moving from parameter θ_j to parameter θ_i.

η and α are concentration parameters that govern the likelihood of new states occurring when the model is applied to a finite dataset. The two DPs in (1a) are linked by sharing

[1]It is possible to allow ϕ and σ_v^2 to be state dependent, however, we omit it here as we found no empirical support for this specification.

the same atom θ. This means that each draw of Π_j has the same support and facilitates the construction of an infinite transition matrix that can be used to govern s_t. $Stick(\eta)$ determines the top-level hierarchy and is shared in the second level. The second layer, $Stick2(\alpha, \Gamma)$, governs each row of the transition matrix and is centred such that $E[\Pi_j] = \Gamma$. The IHMM nests the DPM model of Antoniak (1974) when $\alpha \to \infty$, and each row of the transition matrix converges to the same vector Γ.

The associated stick-breaking representation of the model is

$$p(r_t|\theta, \Pi, s_{t-1}, h_t) = \sum_{k=1}^{\infty} \pi_{s_{t-1}k} N(r_t; \mu_k, \omega_k^2 \exp(h_t)), \tag{4a}$$

$$h_t = \phi h_{t-1} + \sigma_v v_t, \;\; v_t \sim N(0, 1), \tag{4b}$$

where $N(r_t; \mu_k, \omega_k^2 \exp(h_t))$ denotes the normal density function with mean μ_k and variance $\omega_k^2 \exp(h_t)$ evaluated at r_t. $\pi_{s_{t-1}k}$ governs the weight assignments to different normal kernels, where the weights change accordingly over time via the first-order Markov chain. The model in (4) becomes the SV-DPM specification of Jensen and Maheu (2010) if the weights are independent of the previous state, where $\pi_{jk} = \pi_k$ for all j and k.

As in conventional SV models, the conditional mean has the lag term ϕh_{t-1} but the SV-IHMM has a second channel affecting volatility though the Markov chain and the variance component ω_{s_t}. Since the unconditional mean of h_t is zero when $|\phi| < 1$, the parameter $\omega_{s_t}^2$ effectively controls and allows for changes in the log-volatility of the returns. This is seen by rewriting the model as

$$r_t = \mu_{s_t} + \exp(h_t^{'}/2) z_t \tag{5a}$$

$$h_t^{'} - \log \omega_{s_t}^2 = \phi(h_{t-1}^{'} - \log \omega_{s_{t-1}}^2) + \sigma_v v_t, \tag{5b}$$

where $h_t^{'} = h_t + \log \omega_{s_t}^2$. Here the conditional mean of $h_t^{'}$ is $\log \omega_{s_t}^2$ and could capture transitory jumps as well as permanent changes in log-volatility depending on the state process. State changes allow for both the conditional mean and the unconditional mean of $h_t^{'}$ to change over time through $\omega_{s_t}^2$.

Although not modelled parametrically, leverage or asymmetric volatility effects in which price changes result in volatility changes next period can be captured through the nonparametric portion of the model. For instance, a state move that results in a low μ_{s_t} this period and a high $\omega_{s_{t+1}}^2$ next period will capture this. The advantage of modelling this nonparametrically is that a more general relationship can be captured as well as allowing this relationship to change over time.

Finally, one may argue that the extension to allow ϕ and σ_ν^2 to be state dependent could be important. We estimated this for our empirical application and found little evidence of time variation in these parameters. As such, we focus on the simpler specification above.

2.2 Priors and Hierarchical Priors

This subsection defines the priors and hierarchical priors for the SV-IHMM. The priors for the infinite Markov transition matrix Π are formed by $Stick(\eta)$ and $Stick2(\alpha, \Gamma)$, which were discussed in previous section. In order to minimize the impact of the prior, rather than fixing η and α, we follow Fox et al. (2011) and impose the following hyper prior:

$$\eta \sim Gamma(2, 8), \quad \alpha \sim Gamma(2, 8), \quad E(\eta) = E(\alpha) = 0.25. \tag{6}$$

$\mathcal{H}$ is the common base measure of the second layer of the DPs in the model. This prior is specified as $\mu_j \sim N(b_0, B_0)$ and $\omega_j \sim IG(\nu_0, s_0)$. Motivated by Song (2014), a hierarchical prior is used to learn from the data about these prior settings. These are

$$b_0 \sim N(0, 1), \quad B_0 \sim IW(3, I), \quad v_0 \sim Exp(1), \quad s_0 \sim Gamma(5, 1), \tag{7}$$

where I is an identity matrix and $B_0 \sim IW(4, I)$ if the conditional mean is an AR(1) process. When a new state is introduced to the model, the associated draws of a new μ and ω are obtained from the informative priors that were influenced by the data. This can contribute to faster learning about the new states and, thus, improve the forecasts.[2] $\phi \sim N(0, 1)$ is truncated to the stationary region for an AR(1) process and $\sigma_v^2 \sim IG(11, 0.01)$.[3]

2.3 Posterior Sampling

The sampling scheme for the SV-IHMM consists of two parts. First, we sample the state-dependent parameters, transition matrix, latent states and the concentration parameters of the HDP. Second, we sample the log-volatility.

Conditional on the log-volatility, the sampling algorithm for the state-dependent parameters is similar to that of the IHMM. We use the beam sampler from Van Gael et al. (2008). This randomly generates the auxiliary variables (slices) that stochastically truncate the infinitely dimensional transition matrix Π into a finite size so that the forward-filtering backward-sampling (FFBS) an be applied (Chib, 1996).

[2]Maheu and Yang (2016) documents significant improvements in the density forecast accuracy.

[3]We apply a very informative prior to separately identify the SV and IHMM components. A prior of $\sigma_v^2 \sim IG(5, 0.25)$ provides similar forecast results.

We define an auxiliary variable $u_t > 0$ (slice) that is generated by a uniform density as follows:

$$p(u_t|s_t, s_{t-1}, \Pi,) = \frac{\mathbb{1}(u_t < \pi_{s_{t-1},s_t})}{\pi_{s_{t-1},s_t}} \quad t = 1, \dots, T, \tag{8}$$

where $\mathbb{1}(\cdot)$ denotes the indicator function. Augmenting the model with u_t gives us the following target density:

$$p(r_t, u_t|\theta, \Pi, s_{t-1}, h_t) = \sum_{k=1}^{\infty} \mathbb{1}(u_t > \pi_{s_{t-1}k}) N(r_t; \mu_k, \omega_k^2 \exp(h_t)). \tag{9a}$$

Integrating out the slice yields (4a), but given u_t there are now a finite number of non-zero terms $\mathbb{1}(u_t > \pi_{s_{t-1}k})$ that we need to account for. This is easily found by defining K to satisfy $\max_{i\in\{1,\dots,K\}}\{1 - \sum_{j=1}^{K} \pi_{i,j}\} < \min_{t\in\{1,\dots,T\}}\{u_t\}$. Then $j = 1, \dots, K$ cover all non-zero terms $\mathbb{1}(u_t > \pi_{s_{t-1}k})$.

Now, sampling the states and the state-dependent parameters is done on a finite Markov switching model. In each iteration of the posterior sample, K will change.

The FFBS within the Beam sampler is applied in the following way:

The prediction step for $k = 1, \dots, K$ calculates as

$$p(s_t = k|u_{1:T}, \Pi, r_{1:t-1}) \propto \sum_{j=1}^{K} \mathbb{1}(u_t < \pi_{j,k}) p(s_{t-1} = j|u_{1:T}, \Pi, r_{1:t-1}, h_t). \tag{10}$$

The update step for $k = 1, \dots, K$ calculates as

$$p(s_t = k|u_{1:T}, \Pi, r_{1:t}) \propto p(s_t = k|u_{1:T}, \Pi, r_{1:t-1}) p(r_t|r_{1:t-1}, \mu_k, \omega_k, h_t). \tag{11}$$

After $s_{1:T}$ are sampled, we update K by excluding the states for which there are no observation assignment. The slices are drawn from the uniform distribution.

To sample h_t, a random length block-move Metropolis-Hastings (MH) sampler of Jensen and Maheu (2010) is used. The block size of this sampler is randomly drawn from a Poisson distribution with preset hyperparameter λ_h, and the expected block size is $\lambda_h + 1$. Once h_t is sampled, θ and σ_v can be easily sampled via conjugacy. $c_{1:K}$ represents the oracle counts that help us sample α and η. All of the posterior steps are summarized in the following:

$$
\begin{array}{lll}
p(u_{1:T}|s_{1:T}, \Pi) & p(s_{1:T}|\Pi, u_{1:T}, r_{1:T}, h_{1:T}, \theta) & p(c_{1:K}|s_{1:T}, \Gamma, \alpha) \\
p(\Gamma|s_{1:T}, \eta, \alpha, c_{1:K}) & p(\Pi|s_{1:T}, \Gamma, \alpha, c_{1:K}) & p(\mu_{1:K}, \omega_{1:K}|r_{1:T}, s_{1:T}) \\
p(\alpha, \eta|s_{1:T}, c_{1:K}) & p(h_{1:T}|r_{1:T}, \theta) & p(\phi, \sigma_v^2|h_{1:T}) \\
p(b_0, B_0, v_0, s_0|\mu_{1:K}, \omega_{1:K}) & &
\end{array}
$$

Appendix A.2 describes the details of each sampling step. Let $\Theta = \{s_{1:T}, u_{1:T}, \Pi, \alpha, \eta, c_{1:K}, \mu_{1:K}, \omega_{1:K}, \phi, \sigma_v, h$ Sampling each of the conditional posterior distributions provides one iteration of the sampler and MCMC theory ensures these draws converge to a sample from the desired posterior density, $p(\Theta|r_{1:T})$. After dropping the burn-in draws, the sample average of $g(\Theta^{(i)})$ provides a simulation consistent estimate of the posterior moment, $E[g(\Theta)|r_{1:T}]$, for some function of interest $g(\cdot)$. For example, given N MCMC draws,

$$E(\mu_{s_t}|r_{1:T}) \approx \frac{1}{N}\sum_{i=1}^{N} \mu^{(i)}_{s_t^{(i)}}, \text{ for } t = 1, \ldots, T, \tag{13}$$

is the posterior mean estimate of μ_{s_t} at each point in time.

2.4 Out-of-Sample Forecasts

This subsection describes the simulation details to compute forecasts. The predictive distribution of the returns integrates out all of the parameter uncertainty and has the following generic form:

$$p(r_{t+1}|r_{1:t}) = \int p(r_{t+1}|\Theta, r_{1:t})p(\Theta|r_{1:t})d\Theta, \tag{14}$$

where $p(r_{t+1}|\Theta, r_{1:t})$ is the density of r_{t+1}, given the parameter set Θ and the past returns. $p(\Theta|r_{1:t})$ is the posterior density of Θ, given the data. Any feature of the predictive density, such as the predictive mean, can be obtained through simulation methods.

A central component in a Bayesian model comparison is the predictive likelihood. This is obtained for a model by evaluating the predictive density at the realized data point r_{t+1}. The predictive likelihood measures the accuracy of the density forecasts, with larger values being better.

To compute the log-predictive likelihood (LPL) for the SV-IHMM, we do the following: Given the posterior draws from each iteration of the MCMC sampler $\{\Theta^{(i)}\}_{i=1}^{N}$, we draw $s_{t+1}^{(i)} \in \{1, \ldots, K^{(i)} + 1\}$, where $K^{(i)}$ is the total number of active states:

1. Simulate the state variable $s_{t+1}^{(i)}$ through $\Pi^{(i)}_{s_t^{(i)}}$, conditional on $s_t^{(i)}$.

2. If $s_{t+1}^{(i)} \leq K^{(i)}$, then r_{t+1} is assigned to an existing state, with state-dependent parameter $\theta_{s_{t+1}^{(i)}} = (\mu_{s_{t+1}^{(i)}}^{(i)}, \omega_{s_{t+1}^{(i)}}^{(i)})$. Otherwise, r_{t+1} is assigned to a new state, $s_{t+1}^{(i)} = K^{(i)} + 1$, where $(\mu_{s_{t+1}^{(i)}}^{(i)}, \omega_{s_{t+1}^{(i)}}^{(i)})$ is drawn from the hierarchical prior, $\mu_{s_{t+1}^{(i)}} \sim N(b_0^{(i)}, B_0^{(i)})$ and $\omega_{s_{t+1}^{(i)}}^2 \sim IG(\nu_0^{(i)}, s_0^{(i)})$.

The predictive likelihood estimate at $t+1$ is computed over all MCMC draws:

$$p(r_{t+1}|r_{1:t}) \approx \frac{1}{N} \sum_{i=1}^{N} p(r_{t+1}|\mu_{s_{t+1}^{(i)}}^{(i)}, \omega_{s_{t+1}^{(i)}}^{(i)2} \exp(h_{t+1}^{(i)})), \tag{15}$$

where $p(r_{t+1}|\mu_{s_{t+1}^{(i)}}^{(i)}, \omega_{s_{t+1}^{(i)}}^{(i)2} \exp(h_{t+1}^{(i)}))$ denotes the normal density evaluated at r_{t+1} with mean $\mu_{s_{t+1}^{(i)}}^{(i)}$ and variance $\omega_{s_{t+1}^{(i)}}^{(i)2} \exp(h_{t+1}^{(i)})$. $h_{t+1}^{(i)}$ is obtained by simulating forward a value from the existing MCMC draw $h_{t+1}^{(i)} \sim N(\phi^{(i)} h_t^{(i)}, \sigma_v^{(i)2})$.

Equation (15) measures the predictive likelihood of forecast accuracy at period $t+1$. The forecast performance over the entire out-of-sample period, $t_0, \ldots, t_1$ $t_0 \leq t_1$, is determined by computing the joint predictive likelihood of model $\mathcal{M}_A$ in the following way:

$$LPL_A = \log p(r_{t_0:t_1}|r_{1:t_0}, \mathcal{M}_A) = \sum_{t=t_0}^{t_1} \log p(r_t|r_{1:t-1}, \mathcal{M}_A) \tag{16}$$

Two models, $\mathcal{M}_A$ and $\mathcal{M}_B$, can be compared with a log-predictive Bayes factor (BF) defined as $BF_{AB} = LPL_A - LPL_B$. Positive values favour $\mathcal{M}_A$. Values above 5 are regarded as strong evidence for $\mathcal{M}_A$.

The root mean squared forecast error (RMSFE) for $\mathcal{M}_A$ is computed in a similar way:

$$\text{RMSFE} = \sqrt{\frac{\sum_{t=t_0}^{t_1} (r_t - E(r_t \mid r_{1:t-1}, \mathcal{M}_A))^2}{t_1 - t_0 + 1}}, \tag{17}$$

where $E(r_t|r_{1:t-1}, \mathcal{M}_A)$ is the predictive mean for r_t given data $r_{1:t-1}$. For each out-of-sample period, we re-estimate the model to compute the predictive quantities.

To further evaluate the model forecasts we compute the value-at-risk for quantile q along with the expected shortfall by simulating from the predictive density. To compare models we report the scoring rule of Taylor (2019). Let VaR_{t+1}^q and ES_{t+1}^q denote the value-at-risk and expected shortfall for a model using information $r_{1:t}$ at percentile q. We simulate from the predictive distribution by adding a third step above that simulates $r_{t+1}^{(i)} \sim N(\mu_{s_{t+1}^{(i)}}^{(i)}, \omega_{s_{t+1}^{(i)}}^{(i)2} \exp(h_{t+1}^{(i)}))$. From these draws we numerically estimate the VaR_{t+1}^q and ES_{t+1}^q

accordingly. The scoring function is

$$L(r_{r+1}, VaR^q_{t+1}, ES^q_{t+1}) = -\ln\left(\frac{1-q}{ES^q_{t+1}}\right) - \frac{(r_{t+1} - VaR^q_{t+1})\left[q - \mathbb{1}(r_{t+1} < VaR^q_{t+1})\right]}{qES^q_{T+1}} + \frac{r_{t+1}}{ES^q_{t+1}}.$$

The average score, TS(q), is measured over entire out-of-sample period in the following way,

$$TS(q) = \frac{\sum_{t=t_0}^{t_1} L(r_{r+1}, VaR^q_{t+1}, ES^q_{t+1})}{t_1 - t_0 + 1}, \tag{18}$$

with models producing smaller values being preferred.

3 Benchmark Models

We consider the following benchmark models for comparison. The GARCH-N is defined as

$$r_t = \mu + \sigma_t \epsilon_t, \quad \epsilon_t \sim N(0,1), \quad \sigma_t^2 = \beta_0 + \beta_1(r_{t-1} - \mu)^2 + \beta_2 \sigma^2_{t-1}. \tag{19}$$

The GARCH-t replaces the normal distribution with a Student-t distribution:

$$r_r = \mu + \sigma_t u_t, \quad u_t \sim t(\nu), \quad \sigma_t^2 = \beta_0 + \beta_1(r_{t-1} - \mu)^2 + \beta_2 \sigma^2_{t-1}, \tag{20}$$

where $t(\nu)$ denotes a Student-t distribution with mean 0, scale parameter 1 and degree of freedom ν.

The SV parametric versions, including the SV-N, are defined as

$$r_t = \mu + \exp(h_t/2)\, \epsilon_t, \quad \epsilon \sim N(0,1), \quad h_t = \xi + \phi h_{t-1} + \sigma_v v_t. \tag{21}$$

Similarly, SV-t has the following Student-t return innovations:

$$r_t = \mu + \exp(h_t/2)\, u_t, \quad u_t \sim t(\nu), \quad h_t = \xi + \phi h_{t-1} + \sigma_v v_t. \tag{22}$$

The SV-IHMM nests several models of interest that we can compare our model to. The first is an IHMM without the SV component. If $\sigma_v = 0$, and $h_t = 0, \forall t$ in the SV-IHMM then we have the following IHMM:

$$\Gamma \sim Stick(\eta), \quad \Pi_j \overset{iid}{\sim} Stick2(\alpha, \Gamma), \quad j = 1, \ldots, \infty, \tag{23a}$$

$$s_t | s_{t-1} \sim \Pi_{s_{t-1}}, \tag{23b}$$

$$r_t | s_t, h_t, \theta \sim N(\mu_{s_t}, \omega^2_{s_t}), \tag{23c}$$

$$\theta_j \overset{iid}{\sim} \mathcal{H}, \quad j = 1, \ldots, \infty, \tag{23d}$$

As mentioned above, the infinite hidden Markov chain nests the DPM as a special case and, therefore, the SV-IHMM nests the SV-DPM of Jensen and Maheu (2010). The SV-DPM model is obtained by replacing the first two lines in (1a)–(1b) with

$$\Gamma \sim Stick(\eta), \tag{24a}$$

$$s_t \sim \Gamma, \quad t = 1, \ldots, T. \tag{24b}$$

Finally, since the SV-IHMM nests the SV-DPM, it also nests the SV-t under certain parameter restrictions and prior assumptions.

The model by Amado and Terasvirta (2013), is a multiplicative time-varying GJR-GARCH that decomposes volatility to a GJR-GARCH specification and a multiplicative time-varying component. According to Amado and Terasvirta (2013), the TV-GJR-GARCH model is written as,

$$r_t = \mu + e_t, \quad e_t = \sigma_t^2 g_t \epsilon_t, \quad \epsilon_t \sim N(0, 1), \tag{25a}$$

$$\sigma_t^2 = \beta_0 + \beta_1 e_{t-1}^2 + \beta_2 \sigma_{t-1}^2 + \beta_3 e_{t-1}^2 \mathrm{I}(e_{t-1} < 0), \tag{25b}$$

$$g_t = g_t(t/T, \gamma, c_{1:K}) = \sum_{l=1}^{r} \delta_l G_l(t/T, \gamma, c_{l,1:K}), \tag{25c}$$

$$G_l(t/T, \gamma, c_{l,1:K}) = \left(1 + \exp\{-\gamma \prod_{k=1}^{K} (t/T - c_{lk})\}\right)^{-1}. \tag{25d}$$

The $g_t(\cdot)$ is a time-varying deterministic function with $\delta_l > 0$, $\gamma > 0$ and $c_1 \leq c_2 \leq \cdots \leq c_K$. The choice of $r = 1$ and $K = 2$ are preselected and suggested by Amado and Terasvirta (2013) as the optimal choice.[4] The extended model labelled TV-GJR-GARCH-t replaces the normal innovations with Student-t innovations in the return equation. Posterior simulation steps follow the GARCH model and details are discussed on the Appendix A.4.

The multifractal volatility of Calvet and Fisher (2004) decomposes the volatility into sev-

[4]We also tried $r = 1$ and $K = 1$ ranked as the second best in Amado and Terasvirta (2013). The alternative choice does not display better forecast performance.

eral latent multiplicative components, each multinomially distributed. Their model, labelled as MMV-K, is.

$$
\begin{aligned}
r_t &= \mu + \omega e_t, \quad e_t \sim N(0, \sigma_t^2) && \text{(26a)} \\
\sigma_t^2 &= M_{1t} \cdot M_{2t} \cdots M_{Kt} && \text{(26b)} \\
M_{kt} &\sim \begin{cases} \alpha & \text{with probability} \quad \frac{1}{2}\gamma_k \\ 2-\alpha & \text{with probability} \quad \frac{1}{2}\gamma_k \\ M_{kt-1} & \text{with probability} \quad 1-\gamma_k \end{cases} && \text{(26c)} \\
\gamma_k &= 1 - (1-\gamma_1)^{b^{k-1}}, && \text{(26d)}
\end{aligned}
$$

with $\gamma_k \in (0,1)$ and $b \in (1, \infty)$.[5] The σ_t^2 is a joint multiplication of K multipliers and each multiplier is either α or $2-\alpha$. The number of multipliers K is preset and denoted as MMV-K in the paper. Given the K and two choices for each multiple (α and $2-\alpha$), there is a total of $d = 2^K$ combinations. As suggested by Calvet and Fisher (2004), the model can be written as a special case of Markov-switching model and as such estimation follows Chib (1996) to sample the latent states. The rest of the parameters (γ_1, b, α) are sampled conditioned on the latent state sequences via a single-move random-walk Metropolis-Hastings step.[6]

The prior and the hierarchical prior of the IHMM are the same as that of the SV-IHMM. For the SV-DPM, we keep the same priors, hyper-priors and hierarchical priors as in SV-IHMM. The key difference is that there is only one concentration parameter, $\eta \sim Gamma(2, 8)$, in the SV-DPM. Let μ, β_0, β_1, β_2 follow an independent $N(0, 1)$ in GARCH-N and GARCH-t. Similarly, μ, ξ, ρ follow an independent $N(0, 1)$ and $\sigma_v^2 \sim IG(11, 0.01)$ in both the SV-N and SV-t. The prior for ν in the Student-t is uniform: $\nu \sim U[2, 50]$ which applies to all model using student-t distribution. For the TV-GJR-GARCH model, μ, β_0, β_1, β_2 and β_3 follow independent $N(0, 1)$. The other parameters, δ_1, γ, $c_{1:2}$ follow truncated $N(0, 1)$ with the restrictions such that $\delta_1 > 0$, $\gamma > 0$, $c_1 \le c_2 \le \cdots \le c_K$. For the MMV-K model, $\gamma_1 \sim Beta(2, 2)$, $\alpha \sim U[0, 1]$, $\omega^{-2} \sim Gamma(5, 1)$. There is no need to sample $\gamma_{2:K}$ as they are deterministic conditioned on γ_1 and b.

---

[5] γ_1 is sampled and $\gamma_{2:K}$ are deterministic given γ_1 and b.

[6] Calvet and Fisher (2004) indicate the MMV-K is ultimately a Markov switching model with state dimension of 2^K. The computation of Markov transition probability is referred to Appendix A.5.

4 Empirical Results

4.1 Data

Four time series datasets are studied using the SV-IHMM and the benchmark models. These datasets cover three assets from equity, commodity, and foreign exchange markets and a macroeconomic indicator. We select Apple Inc. (AAPL) as a large cap equity and use its common stock returns at daily frequencies, dated from December 15th, 1980 to December 31, 2020, and obtain a sample size of 10,099, which we retrieved from the CRSP.[7] For the foreign exchange rate, we study the daily exchange rates of the Canada-US dollar for the period January 5th, 1971, to December 31, 2020 (12,057 observations), which we obtained from the FRED.[8] West Texas Intermediate (WTI) crude oil spot free on board (FOB) prices are selected for our commodity prices and run from January 2, 1986 to December 31, 2020. There are 8,819 daily observations and these are downloaded from the U.S. Energy Information Administration. The U.S. industrial production index is downloaded from FRED and is a monthly measure of real output. There are 1,222 observations, dating from March, 1919 to December, 2020. All of the time series are transformed into rates of change by taking the log difference and scaling it by 100. The data series are labelled AAPL, USD/CAD, Crude Oil, and IP Growth, respectively. Table 1 illustrates some descriptive statistics of the data. AAPL and Crude Oil have greater volatility and skewness than USD/CAD and IP Growth.

4.2 Posterior Analysis

Table 2 summarizes the posterior parameter estimates of only the most competitive models: the SV-IHMM, SV-DPM, SV-t and GARCH-t across the four datasets. The posterior means and 0.95 density interval estimates are reported. The burn-in MCMC draws are 20,000 and another $N = 20,000$ draws are used for posterior inference. In the case of IP Growth, we include an AR(1) term in the parametric models with a fixed coefficient in the conditional mean and denote it as ρ in the table.[9] In the nonparametric models (SV-IHMM, SV-DPM and IHMM), ρ is also state dependent along with the intercept.

First, introducing a second dynamic structure on the volatility through $\omega^2_{s_t}$ does not weaken the volatility persistence of h_t. For instance, ϕ is in the range of 0.993 – 0.999 for all of the models. Second, in the nonparametric components, we find that the SV-IHMM model uses more active states than the SV-DPM in applications of AAPL and USD/CAD, whereas

[7] Center for Research in Security Prices.

[8] Federal Reserve Economic Data, U.S. Federal Reserve Bank of St. Louis.

[9] See Maheu et al. (2020) for the importance of a lag of IP growth for forecasting.

it is about the same in applications of Crude Oil and IP Growth, as shown in Table 2.[10]

The estimates for the SV-t and the GARCH-t are typical, with a small degree of freedom in the t-distribution and strong persistence measures of ϕ and $\beta_1 + \beta_2$ in volatility. The exception is for the SV-t applied to IP Growth, where the degree of freedom is larger other applications. In this case, the fat-tails are generated though the log-volatility, which has a much larger σ_v^2 than the other data.

Figure 1 shows the posterior mean of the variance components for the SV-IHMM model that is applied to AAPL for the period 2012 to 2020. As discussed earlier, the h_t process by construction captures the smooth changes in the volatility. Deviations from this are controlled by ω_{s_t}, which captures short-term changes in volatility and, as the bottom plot shows, is much more transitory in nature than h_t. This allows for a volatility shock with little to no persistence, where abrupt breaks are captured by ω_{s_t} and we avoid the problem that is common to standard GARCH and SV models, in which the effects of large volatility shocks last too long (Mikosch and Stărică, 2004; Stărică and Granger, 2005).

Figures 2 and 3 display a heatmap for the state used in USD/CAD and IP Growth application. The heatmap is a $T \times T$ matrix that displays $p(s_i = s_j | r_{1:T})$ at entry (i, j) with colours. A colour closer to red (yellow) indicates a probability closer to one (zero). Excluding the diagonal which is all red, off diagonal elements that are closer to red indicate that those dates, (i, j), are more likely to share the same state and hence parameters. The USD/CAD heatmap indicates frequent recurrent states over the whole sample. Similarly, for IP growth we see clear evidence of past states from 1920s and 1960s used to capture the Great Moderation in the early 1980s.

Figures 4 and 5 display the posterior mean of selected state dependent parameters (e.g. $E[\mu_{s_t} | r_{1:T}]$) for two applications. A 0.95 density intervals is included along with colours to indicate the most likely parameter in use at each point in time. $\omega_{s_t}^2$ captures the transitory changes such as the recent COVID-19 shock, where $\omega_{s_t}^2$ displays a spike in early 2020 as shown in Figures 4 and 5. On the contrary, the persistent volatility component, $\exp(h_t)$, hardly moves at all.

Estimates of conditional skewness and conditional kurtosis at each point in the sample from the posterior predictive density are displayed in Figure 6 for the SV-IHMM. There is considerable variation due to the mixture component weights changing over time as well as stochastic volatility.

[10] However, no concrete conclusion can be drawn from the K estimates as it is not a consistent estimator of the number of components. (Miller and Harrison, 2013)

4.3 Density, Point and Tail Forecasts

We perform recursive one-period-ahead out-of-sample forecasts by using each of the models. Three measurements are computed. We report the log-predictive likelihood (LPL), which evaluates the predictive accuracy of the entire predictive distribution. The second measure is the root-mean squared forecast error (RMSFE) of the predictive mean. The final is the scoring rule by Taylor (2019) that evaluates the forecast accuracy of value-at-risk and expected shortfall.

Table 3 shows the LPL, the log-Bayes factor in favour of the SV-IHMM against the benchmarks, and the RMSFE among all of the models for the four datasets. Table 4 compares the performance of tail forecasts of SV-IHMM and the benchmarks. We consider a large out-of-sample period with which to compare the models. In this paper, the training sample is set to approximately to the first five-years of data, and the out-of-sample period starts from the closest start of a calendar year. For example, the out-of-sample size is 8,823 observations for AAPL, 10,856 observations for USD/CAD, 7,543 observations for Crude Oil, and 1,164 observations for IP Growth. To compute the forecasts, each model is re-estimated in each out-of-sample period.

There are several points worth mentioning. First, the SV-IHMM provides the best forecast results, compared to all of the benchmark models, in terms of density forecasts. This model has a positive log-Bayes factor against all competitors. The SV-DPM specification is the second-best model and is always superior to or marginally better than the SV-t. Third, the GARCH-t is quite competitive and is better than the SV-t for IP Growth. The fat tails of the SV-t are always preferred to the SV-N, except for the case of IP Growth where the two models predict equally well. The SV-N does produce fat tails in the predictive density but the generally small degree of freedom parameter estimates in the SV-t (see Table 2) model indicate that this is insufficient. The SV-DPM and SV-IHMM capture the non-Gaussian fat tails through a discrete mixture of distributions.

Some insight into model performance is seen in Figure 7, which plots the cumulative log-BF between the SV-IHMM and other top performing benchmark models at each point in time. If the curve is sloping upward (downward), this indicates the SV-IHMM does better (worse) in accounting for the associated realized data at time t.

Overall, each of the plots shows gradually increasing log-Bayes factors in support of the SV-IHMM. None of the results of the final log-Bayes factors are driven by a few influential outliers and, instead, come from consistent gains over the out-of-sample period. The SV-IHMM can take some time to show improvements over the SV-DPM in the case of Crude Oil and IP Growth. This is likely due to needing more data to learn about the more complex transition matrix here.

Although the evidence for the SV-IHMM is very strong over the SV-DPM, we acknowledge that the out-of-sample period is very large. This means that it takes a significant amount of data to uncover the gains the SV-IHMM has over the SV-DPM. The key difference in these models is the Markov chain structure governing the states in the SV-IHMM. Finally, the differences in RMSFE are very minor across models.

Some differences in the models can be seen in Figure 8, which shows log-predictive densities for various dates. Generally, when necessary, the SV-IHMM can produce thicker tails than the SV-DPM model.

While the RMSFE and log-predictive likelihood assess the center and the whole predictive distribution, the Taylor scoring rule focuses on accuracy in the lower tail of the distribution. The average score according to (18) is reported in Table 4 and shows the SV-IHMM performs the best for a q of 5%. At the 1% level the SV-t and GARCH-N models are better for AAPL and IP growth, respectively, while the SV-IHMM is competitive.[11] This scoring rule requires $ES^q_{T+1} < 0$ which is violated for the MMV-K models applied to IP growth.

4.4 Robustness

The hierarchical prior in the SV-IHMM automatically provides some robustness to prior settings, but the priors on the precision parameters η and α are informative. This is standard and necessary as it imposes some weak structure on density estimation. Broadly speaking, these parameters control the number of active states in the model and, as such, govern parsimony. To explore their impact on the results, we report the posterior estimates for the full sample and we recompute the out-of-sample forecasts for a loose prior for $\eta \sim Gamma(5,5)$ and $\alpha \sim Gamma(5,5)$ and a tight prior for $\eta \sim Gamma(0.5,8)$ and $\alpha \sim Gamma(0.5,8)$.

Table 5 compares the results of the two different prior settings. The posterior estimates of the SV component are very similar over all prior settings, but more states are used on average for the loose prior, as expected. The loose prior tends to reduce the LPL in the USD/CAD application while it improves in IP Growth. The tighter prior does not show significant changes in the LPL with respect to the benchmark prior. For AAPL and Crude Oil, the alternative priors have a small impact on the LPL and the RMSFE.

The Appendix A.1 includes additional results for the top models using looser priors for σ^2_ν. These result in the same ranking of models. Except for the IP Growth, the tighter prior results in larger LPL values.

[11] Similar results are documented while using the alternative scoring rule of Patton et al. (2019).

5 Conclusion

This paper proposes a new Bayesian semiparametric stochastic volatility model with Markovian mixtures. The model nests the SV-DPM model proposed by Jensen and Maheu (2010) but allows the unknown innovation distribution to change over time. The empirical results show that this change is important. In general, the SV-IHMM consistently outperforms all of the benchmark models in terms of out-of-sample density forecasts and tail forecasts. The results for the SV-IHMM are robust to different prior settings.

References

Amado, C. and Terasvirta, T. (2013). Modelling volatility by variance decomposition. *Journal of Econometrics*, 175(1):142–153.

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics*, 2(6):1152–1174.

Bauwens, L., Dufays, A., and Rombouts, J. V. (2014). Marginal likelihood computation for Markov switching and change-point GARCH models. *Journal of Econometrics*, 178(3):508–522.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Calvet, L. E. and Fisher, A. J. (2004). How to forecast long-run volatility: Regime switching and the estimation of multifractal processes. *Journal of Financial Econometrics*, 2(1):49–83.

Carriero, A., Clark, T. E., and Marcellino, M. (2019). Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors. *Journal of Econometrics*, 212(1):137–154.

Chan, J. (2013). Moving average stochastic volatility models with application to inflation forecast. *Journal of Econometrics*, 176(2):162–172.

Chan, J. C. (2017). The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling. *Journal of Business & Economic Statistics*, 35(1):17–28.

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

Delatola, E.-I. and Griffin, J. E. (2011). Bayesian nonparametric modelling of the return distribution with stochastic volatility. *Bayesian Analysis*, 6(4):901–926.

Delatola, E.-I. and Griffin, J. E. (2013). A Bayesian semiparametric model for volatility with a leverage effect. *Computational Statistics & Data Analysis*, 60:97–110.

Diebold, F. X., Schorfheide, F., and Shin, M. (2017). Real-time forecast evaluation of DSGE models with stochastic volatility. *Journal of Econometrics*, 201(2):322–332.

Dufays, A. (2016). Infinite-state Markov-switching for dynamic volatility. *Journal of Financial Econometrics*, 14(2):418–460.

Fox, E. B., Sudderth, E. B., Jordan, M. I., and Willsky, A. S. (2011). A sticky HDP-HMM with application to speaker diarization. *The Annals of Applied Statistics*, 5(2A):1020–1056.

Griffin, J. E. and Steel, M. F. (2011). Stick-breaking autoregressive processes. *Journal of Econometrics*, 162(2):383–396.

Hou, C. (2017). Infinite hidden Markov switching VARs with application to macroeconomic forecast. *International Journal of Forecasting*, 33(4):1025–1043.

Jensen, M. J. and Maheu, J. M. (2010). Bayesian semiparametric stochastic volatility modeling. *Journal of Econometrics*, 157(2):306–316.

Jin, X. and Maheu, J. M. (2016). Bayesian semiparametric modeling of realized covariance matrices. *Journal of Econometrics*, 192(1):19–39.

Jin, X., Maheu, J. M., and Yang, Q. (2019). Bayesian parametric and semiparametric factor models for large realized covariance matrices. *Journal of Applied Econometrics*, 34(5):641–660.

Jin, X., Maheu, J. M., and Yang, Q. (2022). Infinite Markov pooling of predictive distributions. *Journal of Econometrics*, 228(2):302–321.

Jochmann, M. (2015). Modeling U.S. inflation dynamics: A Bayesian nonparametric approach. *Econometric Reviews*, 34(5):537–558.

Kalli, M., Walker, S. G., and Damien, P. (2013). Modeling the conditional distribution of daily stock index returns: An alternative Bayesian semiparametric model. *Journal of Business & Economic Statistics*, 31(4):371–383.

Liu, J. (2021). A Bayesian semiparametric realized stochastic volatility model. *Journal of Risk and Financial Management*, 14(12):617.

Maheu, J. M. and McCurdy, T. H. (2000). Volatility dynamics under duration-dependent mixing. *Journal of Empirical Finance*, 7(3):345–372.

Maheu, J. M., Song, Y., and Yang, Q. (2020). Oil price shocks and economic growth: the volatility link. *International Journal of Forecasting*, 36(2):570–587.

Maheu, J. M. and Yang, Q. (2016). An infinite hidden Markov model for short-term interest rates. *Journal of Empirical Finance*, 38:202–220.

Marcellino, M., Porqueddu, M., and Venditti, F. (2016). Short-term GDP forecasting with a mixed-frequency dynamic factor model with stochastic volatility. *Journal of Business & Economic Statistics*, 34(1):118–127.

Mikosch, T. and Stărică, C. (2004). Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *The Review of Economics and Statistics*, 86(1):378–390.

Miller, J. W. and Harrison, M. T. (2013). A simple example of Dirichlet process mixture inconsistency for the number of components.

Patton, A. J., Ziegel, J. F., and Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics*, 211(2):388–413.

Ryden, T., Terasvirta, T., and Asbrink, S. (1998). Stylized facts of daily return series and the hidden Markov model. *Journal of Applied Econometrics*, 13(3):217–244.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, 4(2):639–650.

Song, Y. (2014). Modelling regime switching and structural breaks with an infinite hidden Markov model. *Journal of Applied Econometrics*, 29(5):825–842.

Stărică, C. and Granger, C. (2005). Nonstationarities in stock returns. *The Review of Economics and Statistics*, 87(3):503–522.

Taylor, J. W. (2019). Forecasting value-at-risk and expected shortfall using a semiparametric approach based on the asymmetric Laplace distribution. *Journal of Business & Economic Statistics*, 37(1):121–133.

Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes - a study of the daily sugar prices 1961-75. *Time Series Analysis: Theory and Practice*, 1:203–226.

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

Van Gael, J., Saatci, Y., Teh, Y. W., and Ghahramani, Z. (2008). Beam sampling for the infinite hidden Markov model. In *Proceedings of the 25th International Conference on Machine Learning*, pages 1088–1095. ACM.

Virbickaitė, A. and Lopes, H. F. (2019). Bayesian semiparametric Markov switching stochastic volatility model. *Applied Stochastic Models in Business and Industry*, 35(4):978–997.

Yang, Q. (2019). Stock returns and real growth: A Bayesian nonparametric approach. *Journal of Empirical Finance*, 53:53–69.

Table 1: Descriptive Statistics

Returns	Mean	Median	StDev	Skewness	Ex.Kurtosis	Min	Max
AAPL	0.0711	0.0000	2.9081	-1.7501	46.5407	-73.1248	28.6890
USD/CAD	-0.0006	0.0000	0.4087	0.1098	10.1554	-3.8070	5.0716
Crude Oil	-0.0117	-0.0213	2.5514	1.8373	69.8919	-41.2023	64.3699
IP Growth	0.2493	0.2800	1.9409	-0.0607	12.8184	-14.6100	15.3219

Table 2: Posterior Summary of Parameters

Panel A: AAPL

	SV-IHMM		SV-DPM		SV-t			GARCH-t	
	Mean	0.95 DI	Mean	0.95 DI	Mean	0.95 DI		Mean	0.95 DI
μ					0.1188	(0.0816, 0.1559)		0.1417	(0.1239, 0.1563)
ξ					0.0122	(0.0067, 0.0188)	β_0	0.0260	(0.0139, 0.0404)
ϕ	0.9993	(0.9985, 0.9999)	0.9928	(0.9888, 0.9962)	0.9909	(0.9864, 0.9947)	β_2	0.9332	(0.9217, 0.9461)
σ_ν^2	0.0011	(0.0007, 0.0017)	0.0098	(0.0058, 0.0147)	0.0122	(0.0078, 0.0182)	β_1	0.0394	(0.0324, 0.0452)
ν					6.1802	(5.5081, 6.9721)		5.2532	(4.9630, 5.4631)
α	1.2308	(0.7548, 1.8503)							
η	0.9454	(0.4342, 1.6495)	0.4132	(0.1180, 0.8610)					
K	10.3052	(8.0000,13.0000)	6.1295	(3.0000,10.0000)					

Panel B: USD/CAD

	SV-IHMM		SV-DPM		SV-t			GARCH-t	
	Mean	0.95 DI	Mean	0.95 DI	Mean	0.95 DI		Mean	0.95 DI
μ					-0.0001	(-0.0038, 0.0035)		-0.0005	(-0.0043, 0.0031)
ξ					-0.0127	(-0.0189,-0.0069)	β_0	0.0001	(0.0000, 0.0002)
ϕ	0.9993	(0.9987, 0.9998)	0.9962	(0.9943, 0.9979)	0.9951	(0.9929, 0.9971)	β_2	0.9268	(0.9173, 0.9351)
σ_ν^2	0.0024	(0.0015, 0.0033)	0.0116	(0.0091, 0.0148)	0.0132	(0.0100, 0.0174)	β_1	0.0542	(0.0473, 0.0620)
ν					10.0817	(8.2735,12.5834)		6.3324	(5.6626, 7.0463)
α	0.6543	(0.3615, 1.0422)							
η	1.0455	(0.4979, 1.8023)	0.3647	(0.1060, 0.7685)					
K	10.9187	(9.0000,14.0000)	5.3633	(3.0000, 9.0000)					

Panel C: Crude Oil

	SV-IHMM		SV-DPM		SV-t			GARCH-t	
	Mean	0.95 DI	Mean	0.95 DI	Mean	0.95 DI		Mean	0.95 DI
μ					-0.0594	(-0.0965,-0.0227)		-0.0754	(-0.0995,-0.0484)
ξ					0.0148	(0.0091, 0.0214)	β_0	0.0488	(0.0337, 0.0657)
ϕ	0.9933	(0.9901, 0.9961)	0.9893	(0.9851, 0.9931)	0.9875	(0.9826, 0.9916)	β_2	0.9068	(0.8920, 0.9197)
σ_ν^2	0.0098	(0.0066, 0.0137)	0.0164	(0.0120, 0.0214)	0.0182	(0.0139, 0.0237)	β_1	0.0525	(0.0448, 0.0622)
ν					8.9162	(7.4014,10.9299)		5.1749	(4.8852, 5.7158)
α	1.5493	(0.8461, 2.4802)							
η	0.5526	(0.1694, 1.2078)	0.4203	(0.1155, 0.9213)					
K	5.6379	(4.0000,10.0000)	6.1468	(3.0000,12.0000)					

Panel D: IP Growth

	SV-IHMM		SV-DPM		SV-t			GARCH-t	
	Mean	0.95 DI	Mean	0.95 DI	Mean	0.95 DI		Mean	0.95 DI
μ					0.1411	(0.0956, 0.1869)		0.1631	(0.1183, 0.2040)
ρ					0.4158	(0.3576, 0.4734)		0.3973	(0.3466, 0.4574)
ξ					-0.0098	(-0.0335, 0.0118)	β_0	0.0521	(0.0249, 0.0825)
ϕ	0.9966	(0.9941, 0.9990)	0.9970	(0.9948, 0.9992)	0.9622	(0.9354, 0.9866)	β_2	0.5993	(0.4912, 0.7464)
σ_ν^2	0.0017	(0.0006, 0.0046)	0.0020	(0.0008, 0.0039)	0.1411	(0.0518, 0.2265)	β_1	0.2400	(0.1466, 0.3233)
ν					24.6920	(6.1601,48.5491)		4.5288	(3.5544, 5.5427)
α	1.2772	(0.7094, 2.1243)							
η	0.6872	(0.2446, 1.3464)	0.5785	(0.2191, 1.1131)					
K	6.9515	(5.0000,11.0000)	7.6072	(5.0000,11.0000)					

Note 1: ρ denotes the parameter of the additional AR(1) term for each model.
Note 2: μ, ρ and ξ are state-dependent parameters for SV-IHMM and SV-DPM.
Note 3: β_0, β_1 and β_2 are the GARCH parameters from (20).

Table 3: Out-of-Sample Forecast Performance: Density and Point Forecast

	AAPL			USD/CAD		
	LPL	log BF	RMSFE	LPL	log BF	RMSFE
SV-IHMM	**-19846.89**	—	2.8373	**-3581.02**	—	**0.4283**
SV-DPM	-19893.40	46.50	**2.8341**	-3616.36	35.34	0.4284
SV-t	-19892.58	45.68	2.8345	-3629.96	48.93	0.4284
GARCH-t	-19953.12	106.23	2.8347	-3636.74	55.72	0.4284
IHMM	-19934.70	87.81	2.8382	-3716.15	135.13	0.4297
SV-N	-20037.88	190.99	2.8350	-3688.19	107.17	0.4284
GARCH-N	-20542.51	695.62	2.8343	-3914.81	333.79	0.4284
TV-GJR-GARCH	-20830.94	984.01	2.8340	-4063.41	482.38	0.4284
TV-GJR-GARCH-t	-20578.67	731.78	2.8349	-3966.95	385.93	0.4284
MMV-2	-20507.10	660.21	2.8350	-5002.40	1421.38	0.4284
MMV-3	-20240.00	393.11	2.8348	-4392.00	810.98	0.4284
MMV-4	-20187.90	341.01	2.8347	-4867.90	1286.88	0.4284
MMV-5	-20211.60	364.71	2.8348	-4644.30	1063.28	0.4284
MMV-6	-20285.40	438.51	2.8348	-4818.10	1238.08	0.4284
	Crude Oil			IP Growth		
	LPL	log BF	RMSFE	LPL	log BF	RMSFE
SV-IHMM	**-16189.88**	—	**2.6687**	**-1622.73**	—	1.6058
SV-DPM	-16213.17	23.29	2.6688	-1641.43	18.69	1.5835
SV-t	-16221.76	31.87	2.6690	-1661.11	38.37	1.5823
GARCH-t	-16226.84	36.95	2.6689	-1649.76	27.02	1.5808
IHMM	-16231.72	41.83	2.6706	-1635.47	12.74	1.5903
SV-N	-17019.77	829.88	**2.6687**	-1662.67	39.94	1.5805
GARCH-N	-16492.99	303.10	2.6688	-1791.79	169.06	1.5837
TV-GJR-GARCH	-16572.35	382.42	2.6689	-1721.94	99.17	**1.5753**
TV-GJR-GARCH-t	-17004.03	814.15	2.6692	-1661.55	38.82	1.5817
MMV-2	-16648.60	458.72	2.6694	-1772.30	149.57	1.5923
MMV-3	-16525.70	335.82	2.6698	-1727.50	104.77	1.5905
MMV-4	-16432.60	242.72	2.6693	-1720.30	97.57	1.5910
MMV-5	-16418.70	228.82	2.6693	-1715.40	92.67	1.5906
MMV-6	-16459.80	269.92	2.6693	-1705.30	82.57	1.5875

Note 1: The number of out-of-sample observations for AAPL, USD/CAD, Crude Oil and IP Growth are 8823, 10856, 7543 and 1164, respectively.
Note 2: The log Bayes factors are the difference between the log-predictive likelihoods of the SV-IHMM model and each corresponding model. Bold entries are for the largest LPL and the smallest RMSFE.

Table 4: Out-of-Sample Forecast Performance: Tail Forecasts

	AAPL		USD/CAD		Crude Oil		IP Growth	
	1%	5%	1%	5%	1%	5%	1%	5%
SV-IHMM	3.219	**2.726**	**1.157**	**0.801**	**2.907**	**2.542**	2.513	**1.803**
SV-DPM	3.248	2.742	1.192	0.814	2.936	2.553	2.561	1.850
SV-t	**3.205**	2.735	1.172	0.808	2.938	2.557	2.581	1.846
GARCH-t	3.213	2.754	1.165	0.807	2.920	2.552	2.474	1.825
IHMM	3.223	2.739	1.223	0.833	2.923	2.557	2.506	1.830
SV-N	3.239	2.763	1.184	0.820	3.566	2.737	2.568	1.844
GARCH-N	3.255	2.758	1.262	0.814	**2.907**	2.555	**2.473**	1.807
TV-GJR-GARCH	3.305	2.823	1.241	0.832	2.925	2.567	2.871	1.908
TV-GJR-GARCH-t	4.842	3.712	1.241	0.834	5.121	3.865	—	—
MMV-2	3.375	2.887	1.404	1.032	3.191	2.698	2.843	—
MMV-3	3.291	2.820	1.309	0.925	3.080	2.661	2.713	—
MMV-4	3.283	2.805	1.358	1.027	3.008	2.619	2.681	—
MMV-5	3.276	2.810	1.384	0.996	3.000	2.614	2.669	—
MMV-6	3.298	2.833	1.340	1.016	3.025	2.630	2.634	—

Note 1: The number of out-of-sample observations for AAPL, USD/CAD, Crude Oil and IP Growth are 8823, 10856, 7543 and 1164, respectively.
Note 2: The entries are computed according to the scoring rule by Taylor (2019), which considers the value-at-risk and the expected shortfall jointly. Bold entries are for the smallest value in a column.
Note 3: For IP Growth at 5% level, the MMV-K and TV-GJR-GARCH-t models generate positive VaR and ES in IP Growth, which violates the strictly negative constraint of the scoring rule.

Table 5: Robustness: Posterior Estimates and Forecast Performance

AAPL

	ϕ	σ_v^2	α	η	K	LPL	RMSFE
Loose	0.9989	0.0014	3.3395	1.8419	13.1340	-19843.27	2.8373
Benchmark	0.9993	0.0011	1.2308	0.9454	10.3052	-19846.89	2.8373
Tight	0.9990	0.0011	0.7743	0.8449	10.4248	-19843.70	2.8379

USD/CAD

	ϕ	σ_v^2	α	η	K	LPL	RMSFE
Loose	0.9994	0.0015	1.2870	2.3436	16.3146	-3591.81	0.4283
Benchmark	0.9993	0.0024	0.6543	1.0455	10.9187	-3581.02	0.4283
Tight	0.9991	0.0023	0.6986	0.7403	9.0708	-3577.30	0.4283

Crude Oil

	ϕ	σ_v^2	α	η	K	LPL	RMSFE
Loose	0.9932	0.0099	3.6505	1.1750	6.9398	-16190.03	2.6695
Benchmark	0.9933	0.0098	1.5493	0.5526	5.6379	-16189.88	2.6687
Tight	0.9938	0.0081	1.5841	0.3766	5.0579	-16192.12	2.6691

IP Growth

	ϕ	σ_v^2	α	η	K	LPL	RMSFE
Loose	0.9964	0.0014	2.1213	1.5240	8.8890	-1616.03	1.6060
Benchmark	0.9966	0.0017	1.2772	0.6872	6.9515	-1622.73	1.6058
Tight	0.9964	0.0014	1.0268	0.4840	6.1176	-1624.92	1.6075

Note 1: This table reports posterior mean estimates for ϕ, σ_v^2, α, η and K, in addition to out-of-sample LPL and RMSFE using the same out-of-sample period as before.
Note 2: The loose prior represents $Gamma(5, 5)$; the benchmark prior represents $Gamma(2, 8)$; and the tight prior represents $Gamma(0.5, 8)$.

Figure 1: AAPL Application: Posterior Mean of Variance Components

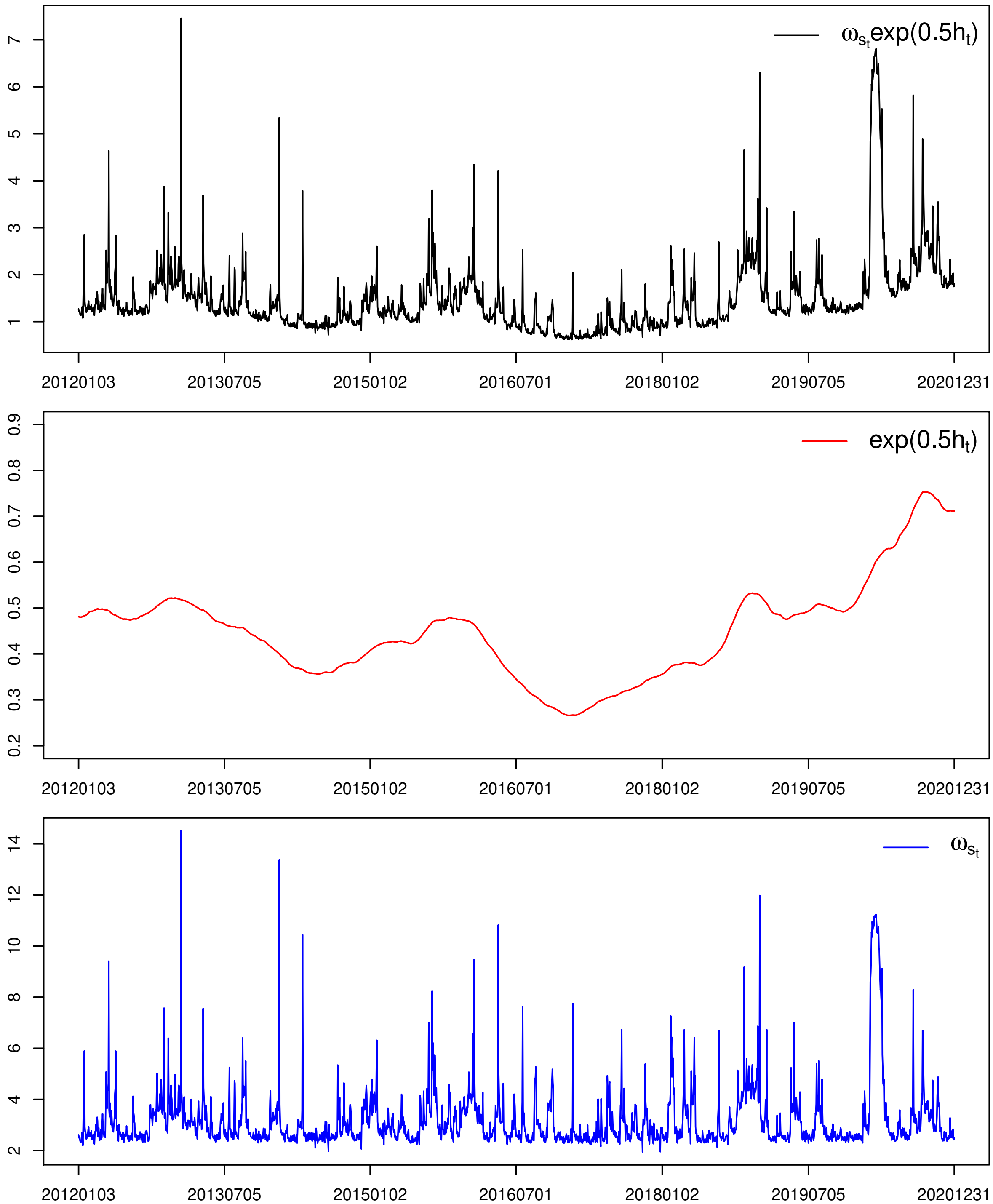
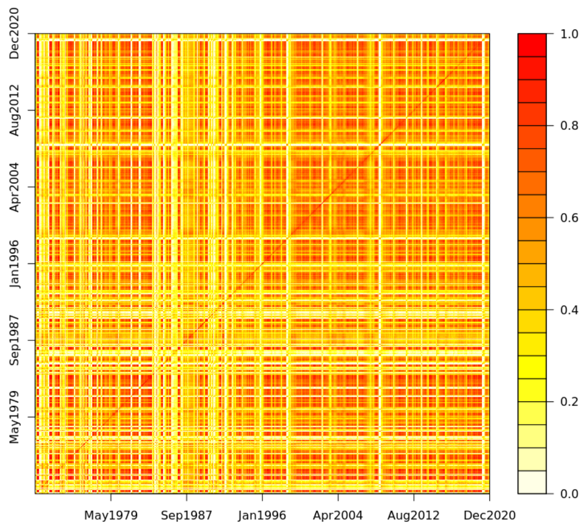
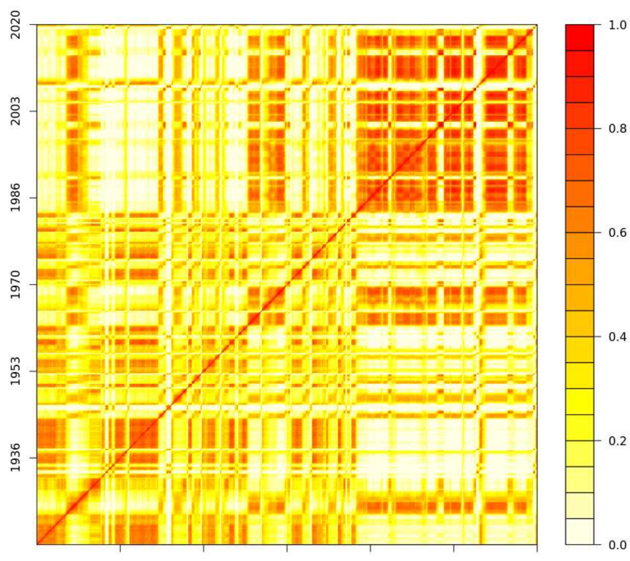


Figure 2: Heat Map for USD/CAD



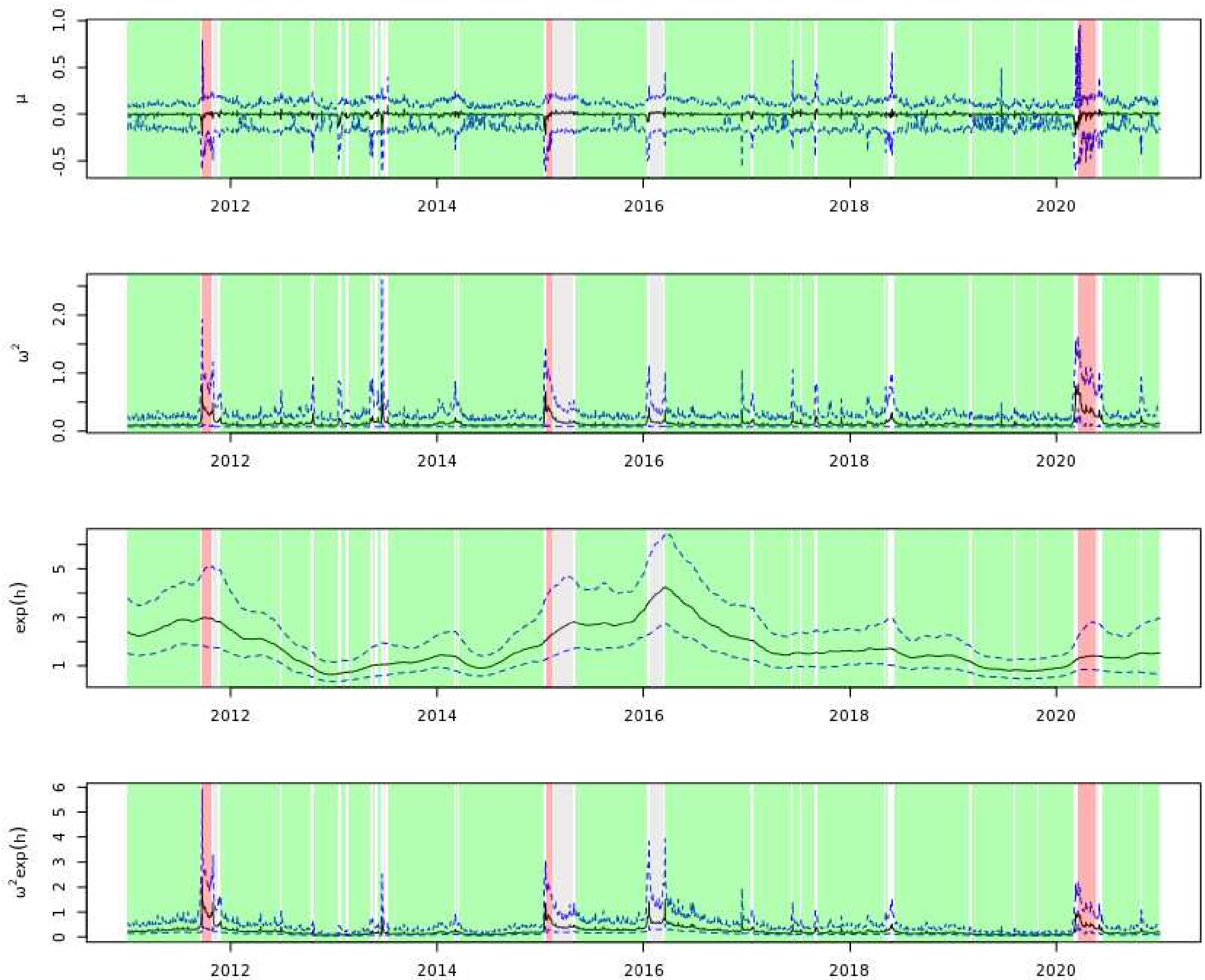
Note 1: The redder the colour, the higher the probability that two periods sharing the same state.

Figure 3: Heat Map for IP Growth



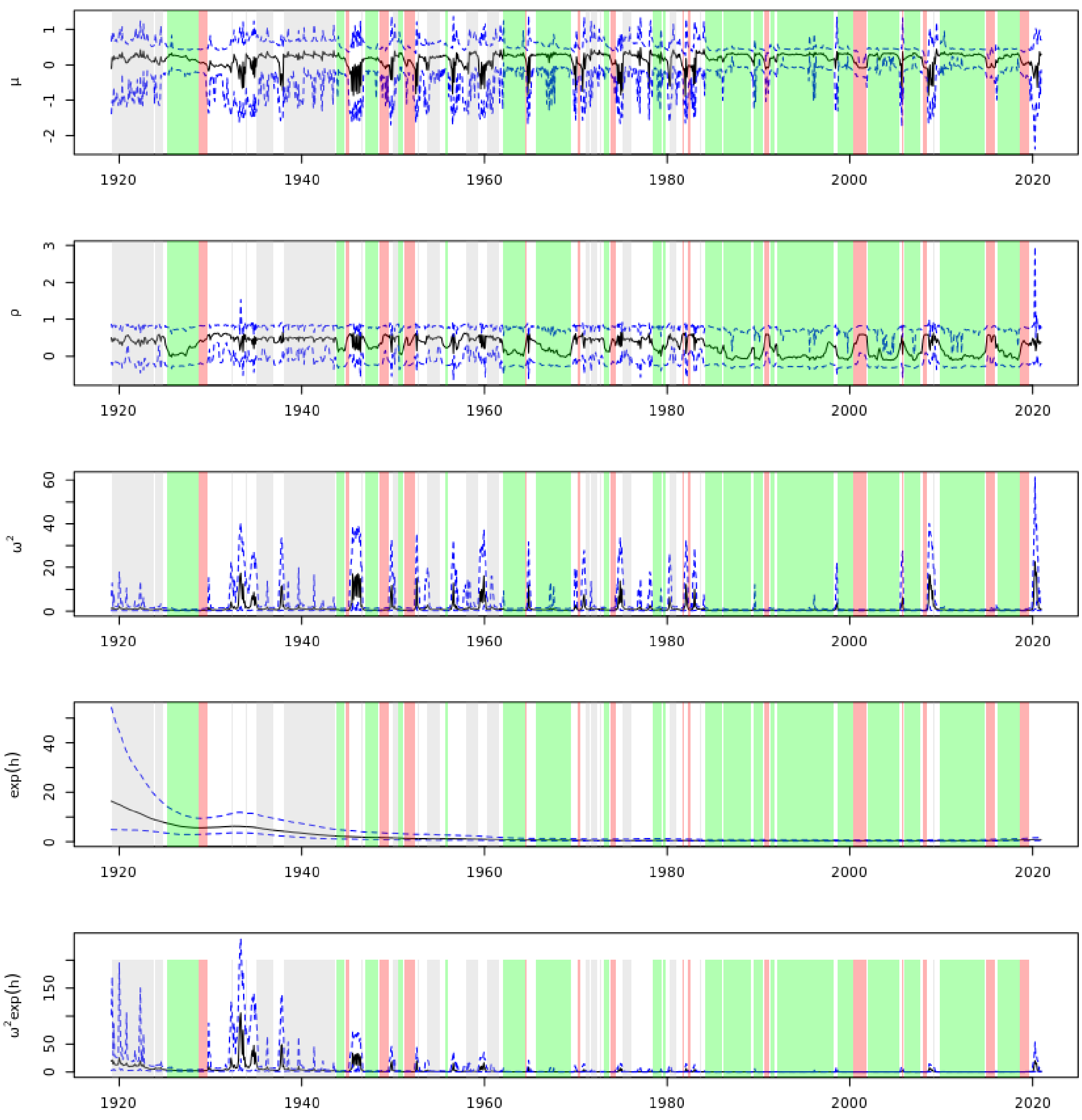
Note 1: The redder the colour, the higher the probability that two periods sharing the same state.

Figure 4: State-dependent parameters over time for USD/CAD



Note 1: The black solid line shows the posterior average of the state-dependent parameter and the blue dotted line shows the corresponding 0.95 DI.
Note 2: The shaded area indicates the most probable state of the period by different colours.

Figure 5: State-dependent parameters over time for IP Growth



Note 1: The black solid line shows the posterior average of the state-dependent parameter and the blue dotted line shows the corresponding 0.95 DI.
Note 2: The shaded area indicates the most probable state of the period by different colours.

Figure 6: Posterior Estimates of Conditional Skewness and Kurtosis

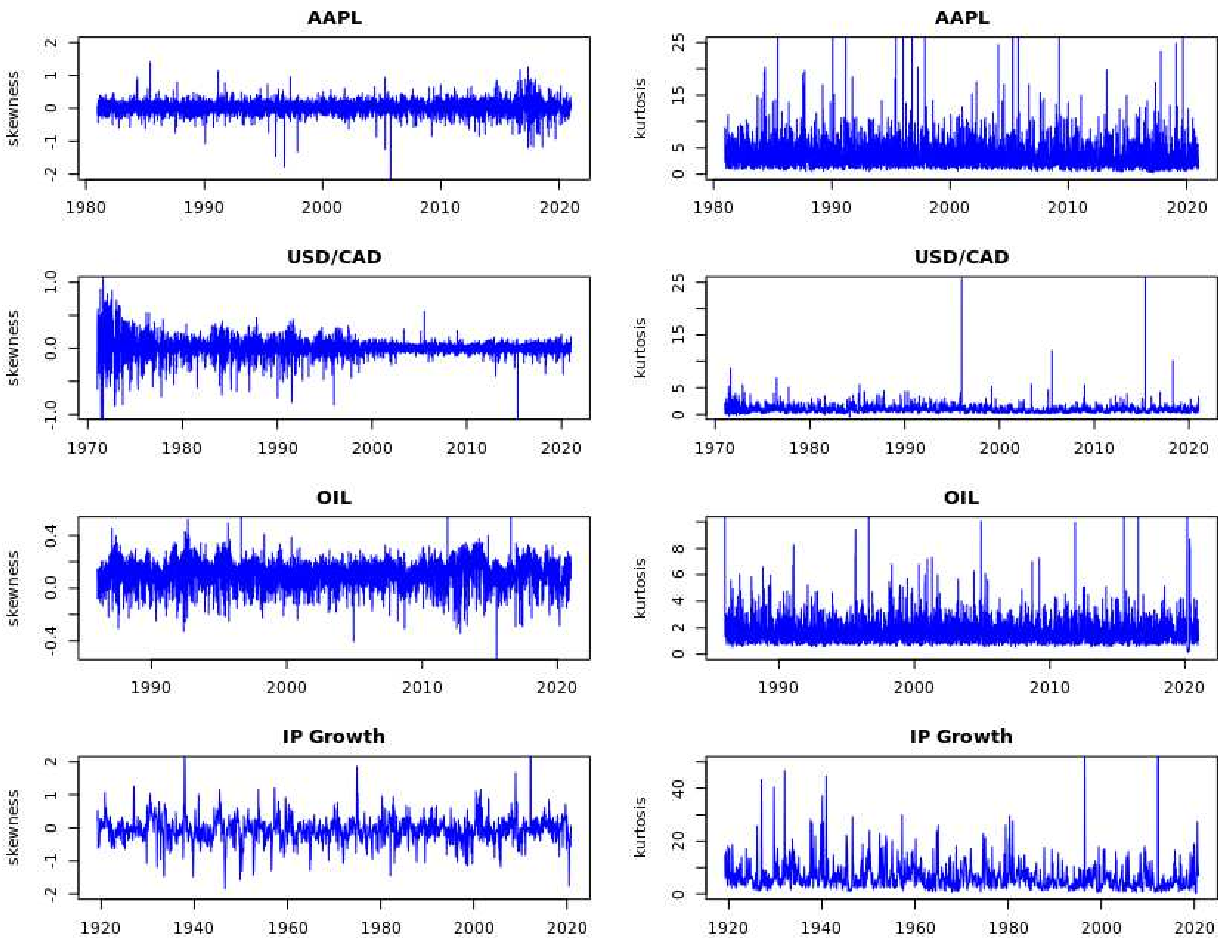


Figure 7: Cumulative Log-Bayes Factor for SV-IHMM

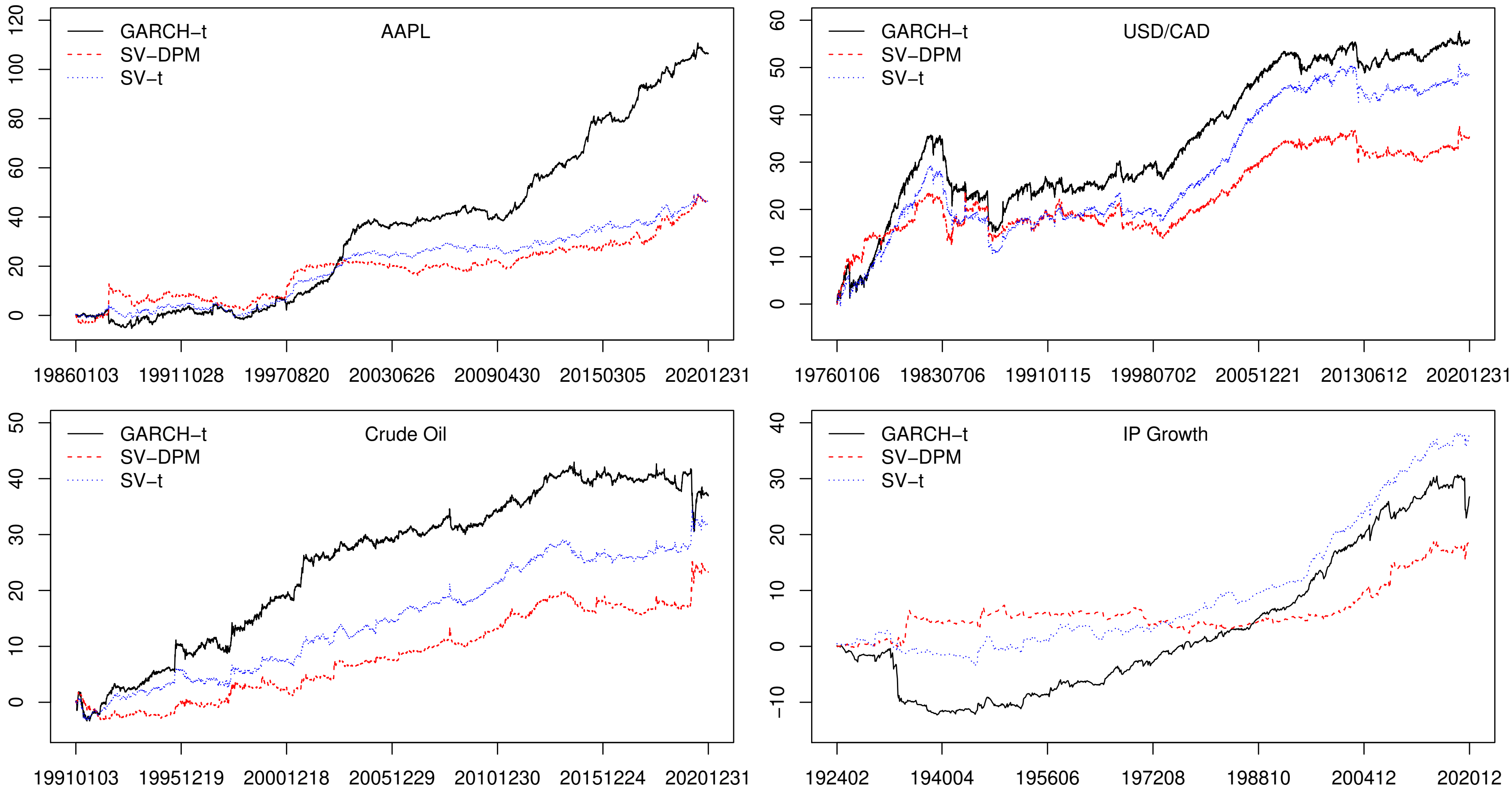
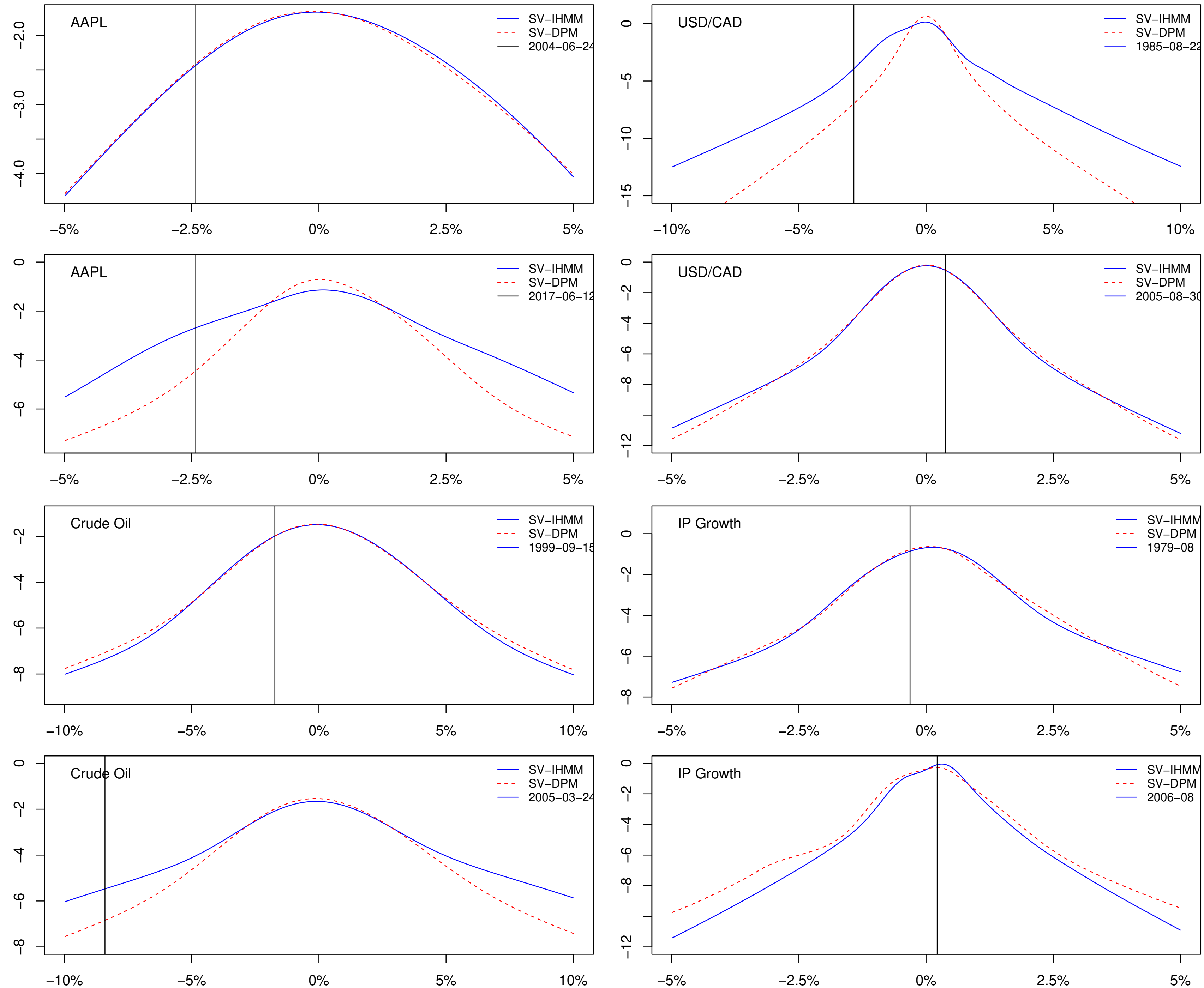


Figure 8: Log-Predictive Densities at Selected Dates



A Appendix

A.1 Appendix: Robustness Test for Different σ_ν^2 Priors

Table 6: Log Predictive Likelihoods for SV-IHMM with Different σ_ν^2 Priors

AAPL	IG(11,0.01)	IG(5,0.25)	IG(2.5,0.5)
SV-IHMM	**-19846.89**	**-19868.08**	**-19871.51**
SV-DPM	-19893.40	-19891.52	-19902.91
SV-t	-19892.58	-19900.91	-19905.87
SV-N	-20037.88	-20038.83	-20042.03
FX			
SV-IHMM	**-3581.02**	**-3611.51**	**-3637.44**
SV-DPM	-3616.36	-3630.36	-3654.82
SV-t	-3629.96	-3638.41	-3644.20
SV-N	-3688.19	-3691.60	-3694.79
OIL			
SV-IHMM	**-16189.88**	**-16196.07**	**-16203.09**
SV-DPM	-16213.17	-16212.94	-16220.30
SV-t	-16221.76	-16224.20	-16227.19
SV-N	-17019.77	-17038.88	-16963.99
IP Growth			
SV-IHMM	**-1622.73**	**-1608.83**	**-1615.09**
SV-DPM	-1641.43	-1631.08	-1633.45
SV-t	-1661.11	-1652.57	-1651.97
SV-N	-1662.67	-1652.25	-1651.96

Note 1: The number of out-of-sample observations for AAPL, USD/CAD, Crude Oil and IP Gr owth are 8823, 10856, 7543 and 1164, respectively.
Note 2: For SV-IHMM and SV-DPM, the SV dynamics follow $h_t = \phi h_{t-1} + \sigma_\nu \nu_t,\ \nu_t \sim N(0,1)$.
Note 3: For SV-t and SV-N, the SV dynamics follow $h_t = \xi + \phi h_{t-1} + \sigma_\nu \nu_t,\ \nu_t \sim N(0,1)$.

A.2 Appendix: Posterior Sampling Steps for SV-IHMM

1. We sample $u_{1:T}|\Gamma, \Pi$: The auxiliary slice variable $U = \{u_t\}_{t=1}^T$ is drawn from $u_1 \sim U(0, \gamma_{s_1})$ and $u_t \sim U\left(0, \pi_{s_{t-1}s_t}\right)$.

2. We update K. Similar to DPM model, if K does not meet the following condition

$$\min \{u_t\}_{t=1}^T > \max \{\pi_{jR}\}_{j=1}^K \tag{27}$$

then K needs to be increased by 1 ($K' = K + 1$), and all of the parameters need to be drawn from the base measure. In addition, since a new "major" state is introduced, Γ and Π also need to be updated accordingly:

(a) $\Theta_{K'} \sim H$;

(b) We draw $v \sim Beta\,(1, \eta)$, then we update $\Gamma = (\gamma_1, \ldots, \gamma_K, \gamma_{K'}, \gamma_R)'$, where $\gamma_{K'} = v\gamma_R$ and $\gamma_R = (1 - v)\,\gamma_R$;

(c) We draw $v_j \sim Beta\,(\alpha\gamma_{K'}, \alpha\gamma_R)$, then we update $\Pi_j = (\pi_{j1}, \ldots, \pi_{jK}, \pi_{jK'}, \pi_{jR})$ for $j = 1, \ldots, K$, where $\pi_{jK'} = v\pi_{jR}$ and $\pi_{jR} = (1 - v)\,\pi_{jR}$;

(d) We draw the K'th row of Π, $\Pi_{K'}$, by $\Pi_{K'} \sim Dir\,(\alpha\gamma_1, \ldots, \alpha\gamma_K, \alpha\gamma_{K'}, \alpha\gamma_R)$.

The above steps are repeated until inequality (27) holds.

3. The forward filter for $s_{1:T}|r_{1:T}, u_{1:T}, \Gamma, \Pi, \Theta, h_{1:T}$. Iterating the following steps forward from 1 to T, we have the following:

(a) The prediction step for initial state s_1 is as follows:

$$p(s_1 = k|u_1, \Gamma) \propto \mathbb{1}\,(u_1 < \gamma_k)\,, \qquad k = 1, \ldots, K \tag{28}$$

for the following states $s_{2:T}$:

$$p(s_t = k|r_{1:t-1}, u_{1:t}, \Pi, \Theta, h_{1:t-1}) \propto \sum_{j=1}^{K} \mathbb{1}\,(u_t < \pi_{jk})\, p\,(s_{t-1} = j|r_{1:t-1}, u_{1:t-1}, \Pi, \Theta, h_{1:t-1}) \tag{29}$$

(b) We update the step for $s_{1:T}$:

$$p\,(s_t = k|r_{1:t}, u_{1:t}, \Pi, \Theta, h_{1:t}) \propto p\,(r_t|r_{t-1}, \theta_k, h_t)\, p\,(s_t = k|r_{1:t-1}, u_{1:t}, \Pi, \Theta, h_{1:t-1}) \tag{30}$$

4. The backward sampler for $s_{1:T}|r_{1:T}, u_{1:T}, \Pi, \Theta, h_{1:T}$. We sample states $s_{1:T}$ using the previously filtered values backward from T to 1:

(a) for the terminal state s_T, we sample directly from $p\,(s_T|r_{1:T}, u_{1:T}, \Pi, \Theta, h_{1:T})$

(b) for the rest states, we sample from the following,

$$p\,(s_t = k|s_{t+1} = j, r_{1:t}, u_{1:t+1}, \Pi, \Theta, h_{1:T}) \propto \mathbb{1}\,(u_{t+1} < \pi_{kj})\, p\,(s_t = k|r_{1:t}, u_{1:t}, \Pi, \Theta, h_{1:T}) \tag{31}$$

5. Sample $c_{1:K}|s_{1:T},\Gamma,\alpha$. Following the sampling approach of Fox et al. (2011), we perform the following:

   (a) We count the number of each transition type, n_{jk}, number of times state j switches to state k.

   (b) We simulate an auxiliary trail variable $x_i \sim Bernoulli\left(\frac{\alpha\gamma_k}{i-1+\alpha\gamma_k}\right)$, for $i=1,\ldots,n_{jk}$. If the trial is successful, then an "oracle" step is involved at the ith step toward n_{jk} and we increase the corresponding "oracle" counts, o_{jk}, by one.

   (c) $c_k = \sum_{j=1}^{K} o_{jk}$.

6. Sample η: Following Fox et al. (2011) and Maheu and Yang (2016), we assume a Gamma prior $\eta \sim Gamma\,(a_1, b_1)$, and let $c = \sum_{j=1}^{K} c_j$,

   (a) $\nu \sim Bernoulli\left(\frac{c}{c+\eta}\right)$

   (b) $\lambda \sim Beta\,(\eta+1, c)$

   (c) $\eta \sim Gamma\,(a_1 + K - \nu, b_1 - \log\lambda)$

7. Sample α: Following Fox et al. (2011), we assume a Gamma prior $\alpha \sim Gamma\,(a_2, b_2)$ and let $n_j = \sum_{k=1}^{K} n_{jk}$,

   (a) $\nu_j \sim Bernoulli\left(\frac{n_j}{n_j+\alpha}\right)$

   (b) $\lambda_j \sim Beta\,(\alpha+1, n_j)$

   (c) $\alpha \sim Gamma\left(a_2 + c - \sum_{j=1}^{K}\nu_j, b_2 - \sum_{j=1}^{K}\log(\lambda_j)\right)$

8. Sample $\Gamma|c_{1:K},\eta$: Given the "oracle" counts $c_{1:K}$ and the property of Dirichlet process, the conjugate posterior is

$$\Gamma|c_{1:K},\eta \sim Dir\,(c_1,\ldots,c_K,\eta) \tag{32}$$

9. Sample $\Pi|n_{1:K,1:K},\Gamma,\alpha$: Similarly, the conjugate posterior of Π_j is

$$\Pi_j|n_{j,1:K},\Gamma,\alpha \sim Dir\,(\alpha\gamma_1 + n_{j1},\ldots,\alpha\gamma_K + n_{jK},\alpha\gamma_R) \tag{33}$$

10. Sample $\Theta|r_{1:T},s_{1:T},h_{1:T}$. We define $Y_k \equiv \left(e^{-\frac{1}{2}h_t} r_t|s_t=k\right)_{t=2}^{T}$, and $X_k \equiv \left(e^{-\frac{1}{2}h_t}|s_t=k\right)_{t=2}^{T}$. The linear model is now

$$Y_k = X_k\mu_k + \omega_k\epsilon_k, \quad \epsilon_k \sim N\,(0,I) \tag{34}$$

The posteriors are

$$p(\mu_k|Y_k,\omega_k) \sim \prod_{t:s_t=k} p(r_t|\mu_k,\omega_k)\, p(\mu_k) \tag{35}$$

$$\sim N(M_\mu, V_\mu) \tag{36}$$

where

$$M_\mu = V_\mu \left(\omega_k^{-1} X_k' Y_k + B_0^{-1} b_0\right) \tag{37}$$

$$V_\mu = \left(\omega_k^{-1} X_k' X_k + B_0^{-1}\right)^{-1} \tag{38}$$

and

$$p(\omega_k|Y_{,}\mu_k) \propto \prod_{t:s_t=k} p(r_t|\mu_k,\omega_k)\, p(\omega_k) \tag{39}$$

$$\sim IG(\bar{v}, \bar{s}) \tag{40}$$

where

$$\bar{v} = \frac{T_k}{2} + v_0 = \frac{1}{2}\sum_{t=1}^{T} \mathbb{1}(s_t = k) + v_0 \tag{41}$$

$$\bar{s} = \frac{1}{2}(Y_k - X_k\mu_k)'(Y_k - X_k\mu_k) + s_0 \tag{42}$$

11. Sample hierarchical priors.

 (a) Sample $b_0|\mu_{1:K}, B_0, h_0, H_0 \sim N(\mu_b, \Sigma_b)$, where

$$\mu_b = \Sigma_b \left(B_0^{-1} \sum_{k=1}^{K} \mu_k + H_0^{-1} h_0\right) \tag{43}$$

$$\Sigma_b = \left(K B_0^{-1} + H_0^{-1}\right)^{-1} \tag{44}$$

 (b) Sample $B_0|\mu_{1:K}, b_0, a_0, A_0 \sim IW(\Omega_B, \omega_b)$, where

$$\omega_b = K + a_0 \tag{45}$$

$$\Omega_B = \sum_{k=1}^{K} (\mu_k - b_0)(\mu_k - b_0)' + A_0 \tag{46}$$

(c) Sample $s_0|\sigma^2_{1:K}, v_0, c_0, d_0 \sim Gamma\left(c_s, d_s\right)$, where

$$c_s = Kv_0 + c_0 \tag{47}$$

$$d_s = \sum_{k=1}^{K} \sigma_k^{-2} + d_0 \tag{48}$$

(d) Sample $v_0|\sigma^2_{1:K}, s_0, g_0$. There IS no easily applicable conjugate prior for v_0, so a Metropolis-Hastings step needs to be applied. We implement a Gamma proposal, following Maheu and Yang (2016):

$$v_0'|v_0 \sim Gamma\left(\tau, \frac{\tau}{v_0}\right) \tag{49}$$

and the acceptance rate is

$$\min\left\{1, \frac{p\left(v_0'|\sigma^2_{1:K}, s_0, g_0\right)/q\left(v_0'|v_0\right)}{p\left(v_0|\sigma^2_{1:K}, s_0, g_0\right)/q\left(v_0|v_0'\right)}\right\} \tag{50}$$

12. $\theta_h|h_{1:T}$: Equation (1d) is simply a linear regression model. Assuming conjugate prior $\beta \sim N\left(b_h, B_h\right)$, the posterior is

$$\delta|\sigma_v, h_{1:T} \sim N\left(M, V\right) \tag{51}$$

$$M = V\left(\sigma_v^{-2}\sum_{t=1}^{T-1} h_t h_{t+1} + b_h B_h^{-1}\right) \tag{52}$$

$$V = \left(\sigma_v^{-2}\sum_{t=1}^{T-1} h_t^2 + B_h^{-1}\right)^{-1} \tag{53}$$

Based on the above linear regression model with conjugate prior $\sigma_v^2 \sim IG\left(v_h, s_h\right)$, the posterior is

$$\sigma_v^2|\delta, h_{1:T} \sim IG\left(\frac{T}{2} + v_h, \frac{\sum_{t=1}^{T-1}\left(h_{t+1} - \delta h_t\right)^2}{2} + s_h\right) \tag{54}$$

13. Sample $h_t|h_{-t}, r_{1:T}, \Theta, s_{1:T}$: We use the block Metropolis-Hastings (MH) sampler as in Jensen and Maheu (2010) with random block size $k = Poisson\left(\lambda_h\right) + 1$. The proposal density is derived by approximating the autoregressive coefficient to 1. This approximation provides an analytic inversion of the covariance matrix. We draw $h'_{(t,\tau)}$

from the following proposal density

$$g\left(h_{(t,\tau)}|\cdots\right)=N\left(h_{(t,\tau)};M_h-0.5V_h\left(\iota-\tilde{y}\right),V_h\right)\tag{55}$$

where

$$\tilde{y}_i=\frac{\left(r_i-\mu_{s_i}\right)^2}{\omega_{s_i}}\exp\left(-M_{h,i}\right)\tag{56}$$

$$M_{h,i}=\frac{\left(k+1-i\right)h_{t-1}+ih_{\tau+1}}{k+1},\ \ i=1,2,\ldots,k\tag{57}$$

$$V_{h,ij}=\sigma_v^2\frac{\min\left(i,j\right)\left(1+k\right)-ij}{k+1}\tag{58}$$

$$V_{h,ij}^{-1}=\begin{cases}2\sigma_v^2 & i=j\\ -\sigma_v^2 & j=i\pm 1\\ 0 & \text{otherwise}\end{cases}\tag{59}$$

We accept $h'_{(t,\tau)}$ with probability

$$\min\left\{1,\frac{p\left(h'_{(t,\tau)}|r_{1:T},h_{-(t,\tau)},\Theta,s_{1:T}\right)/g\left(h'_{(t,\tau)}|h_{-(t,\tau)}\right)}{p\left(h_{(t,\tau)}|r_{1:T},h_{-(t,\tau)},\Theta,s_{1:T}\right)/g\left(h_{(t,\tau)}|h_{-(t,\tau)}\right)}\right\}\tag{60}$$

A.3 Appendix: Posterior Sampling for GARCH-N and GARCH-t

Let $\Theta=\{\mu,\beta_0,\beta_1,\beta_2,\nu\}$ where ν is irrelevant for GARCH-N. We apply a random-walk MH (RWMH) algorithm to sample the whole Θ vector jointly. A single-move RWMH is used to compute the proposal covariance and then a block-move RWMH for better sampling efficiency. A N(0,1) prior is employed for μ, β_0, β_1 and β_2 with restrictions of $\beta_0>0$, $\beta_1>0$, $\beta_2>0$ and $\beta_1+\beta_2<1$. The prior for ν is $U(2,50)$.

A.4 Appendix: Posterior Sampling for TV-GJR-GARCH

We sample $\Theta=\{\mu,\beta_0,\beta_1,\beta_2,\beta_3,\delta_1,\gamma,c_{1:2}\}$ sampled via single-move RWMH with random walk, then a block-move RWMH with random is applied to improve the efficiency. g_t imposes a deterministic function of time t in addition to the GARCH persistence. The priors for μ, β_0, β_1, β_2, β_3 are independent $N(0,1)$ with the same restrictions in GARCH-N. The priors for δ_1, γ, $c_{1:2}$ are truncated $N(0,1)$ with restrictions of $\delta_1>0$, $\gamma>0$ and $c_1\leq c_2\leq\cdots\leq c_K$.

A.5 Appendix: Posterior Sampling for MMV-K

According to Calvet and Fisher (2004), the MMV-K model can be written as a restricted version of Markov-switching model. Given K and two choices for each multiplier (α and $2-\alpha$), there are total of $d = 2^K$ combinations. If we let each combination be a particular state, then each state corresponds to a sequence of K multipliers with each multiplier being either α or $2-\alpha$. Alternatively, let $m_i \in \{m_1, \ldots, m_d\}$ represent one sequence combination where $m_i \in \mathbb{R}_+^K$ for $i = 1, \ldots, d$. By definitions, there is a substantial number of combinations such that $m_i \neq m_j$ for $i \neq j$ but $\prod_{k=1}^K m_i^k = \prod_{k=1}^K m_j^k$, where m_i^k is the kth element in the combination vector m_i for state i. Apparently, σ_t will be the same for these combinations. In short, we may have a large number of unique combination sequences but most of them result in the same σ_t.

Given that $\gamma_{2:K}$ is a deterministic function of b and γ_1, the corresponding Markov transition probability becomes the following,

$$\pi_{ij} = \prod_{k=1}^{K} \left\{ (1-\gamma_k)\mathbb{1}(m_i^k = m_j^k) + 0.5\gamma_k \right\}, \quad i, j \in (1, \ldots, d)$$

where π_{ij} represent the probability of moving from state i to j. With the Markov transition probability matrix and corresponding state variable $\sigma_t \in \{\prod_{k=1}^K m_1^k, \ldots, \prod_{k=1}^K m_d^k\}$, we could generate a large Markov-switching model with dimension of $d = 2^K$. The latent state variable can be sampled via the Forward-filtering Backward-sampling (FFBS) by Chib (1996). Conditioned on the sampled latent states, γ_1 and b are jointly sampled via RWMH. This can be computationally expensive as a new path of the state variable $\sigma_{1:t}$ need to be sampled during the MH whenever new γ_1 and b are generated from the proposal distribution. Conditioned on $\sigma_{1:t}$, γ_1 and b, parameters μ and ω are sampled via conjugate Gibbs.