Universitat de les Illes Balears

DEA WP no. 89
Working Paper Series

Bayesian Semi-Parametric Markov Switching Stochastic Volatility Model

Audrone Virbickaite
Universitat de les Illes Balears
E-mail: audrone.virbickaite@uib.es

Hedibert F. Lopes
Insper Institute of Education and Research
E-mail: hedibertfl@insper.edu.br

October 2018

Departament
d'Economia
Aplicada

Bayesian Semi-Parametric Markov Switching Stochastic Volatility Model

Audronė Virbickaitė [*†] Hedibert F. Lopes [‡]

Abstract

This paper proposes a novel Bayesian semi-parametric Stochastic Volatility model with Markov switching regimes for modeling the dynamics of the financial returns. The distribution of the error term of the returns is modeled as an infinite mixture of Normals, meanwhile the intercept of the volatility equation is allowed to switch between two regimes. The proposed model is estimated using a novel sequential Monte Carlo method called Particle Learning that is especially well suited for state-space models. The model is tested on simulated data and, using real financial times series, compared to a model without the Markov switching regimes. The results show that including a Markov switching specification provides higher predictive power for the entire distribution, as well as in the tails of the distribution. Finally, the estimate of the persistence parameter decreases significantly, a finding consistent with previous empirical studies.

Keywords: Bayes Factor; Dirichlet Process Mixture; Particle Learning; Sequential Monte Carlo.

[*]Corresponding author

[†]Department of Applied Economics, Universitat de les Illes Balears (UIB), Cra. de Valldemossa, km 7.5, 07122, Palma de Mallorca, Spain, audrone.virbickaite@uib.es, Phone: 0034 971 17 2791

[‡]Insper Institute of Education and Research, Sao Paulo, Brazil, 04546-042, hedibertfl@insper.edu.br

1 Introduction

Volatility modeling has been of great interest in the last decades, especially after the recent financial crisis when the standard models failed to explain and predict the events that occurred in the financial markets. The two benchmark approaches to model volatility are based on the Autoregressive Conditional Heteroscedasticity (ARCH) type models, proposed by Engle (1982), and the Stochastic Volatility (SV) type models, proposed by Taylor (1982). The models differ in the underlying assumptions of the observability of the volatility: in ARCH-type models the volatility is deterministic and observable, meanwhile in the SV type models the volatility states are latent and stochastic. By allowing for the volatility states to be stochastic, SV models provide more flexibility than the Generalized ARCH (GARCH, Bollerslev, 1986) specifications, see Broto and Ruiz (2004) for example.

The SV model, as introduced by Taylor (1982), assumes the distribution of the error term of the returns to be Normal. Normal distribution was also considered by Taylor (1986, 1994), Jacquier et al. (1994), Kim et al. (1998), just to name a few. However, many empirical studies have shown that the returns exhibit heavy-tailed behavior, see Chib et al. (2002a), Jacquier et al. (2004), Abanto-Valle et al. (2010a), for example. One possibility, instead of Normal distribution, is to employ a distribution that allows for fat tails. The Student-t distribution was used by Harvey et al. (1994), Gallant et al. (1997), Sandmann and Koopman (1998), Chib et al. (2002b), Jacquier et al. (2004), Nakajima and Omori (2009); the Normal-Inverse Gaussian - by Barndorff-Nielsen (1997); the Mixture of Normals - by Mahieu and Schotman (1998); and the Generalized error distribution - by Liesenfeld and Richard (2006), among many others. Another possibility is to abandon parametric assumptions for the distribution of the returns altogether and consider a semi-parametric SV model[1]. In such model the volatility equation maintains the parametric form, meanwhile the distribution of the returns is modeled non-parametrically.

The Bayesian semi-parametric SV models have become rather popular in the last decade, see Jensen and Maheu (2010, 2014) and Delatola and Griffin (2011, 2013) for univariate SV models. Zaharieva et al. (2017) proposed a multivariate extension with

[1] In some papers such models are referred to as non-parametric SV models

the non-parametric errors. In these works the authors assume that the distribution of the returns follows an infinite mixture of Normals via Dirichlet Process Mixture (DPM) models (see Ferguson, 1983 and Lo, 1984, among others). Such infinite mixtures turn out to be a very flexible modeling approach, since they nest other parametric specifications for the error term. The infinite mixture of Normals can approximate other distributions, frequently used in financial time series context, see e.g. Tokdar (2006) and Mencía and Sentana (2009), because of its "universal approximation property" (Titterington et al., 1985).

In the semi-parametric SV models, even when the distribution of the returns is modeled in a flexible non-parametric manner, the volatility equation still maintains its simple AR(1) representation. Such model might have some limitations by not allowing for structural changes in the volatility process. If these changes are not accounted for, the persistence parameter in the volatility equation might be overestimated. Overestimation of the persistence parameter leads to incorrect conclusions about the predictability of the volatility (Vo, 2009). Therefore, in this paper we augment the semi-parametric Stochastic Volatility model, similar to the one in Delatola and Griffin (2011), to include Markov switching regimes in the volatility equation, resulting into a Bayesian semi-parametric Markov switching SV (MSSV-DPM) model. Including shifts in the volatility regimes was first proposed by So et al. (1998). Since then, fully parametric MSSV models have been rather popular in the financial time series context due to its superior performance as compared to the benchmark SV models. Kalimipalli and Susmel (2004) consider two-factor SV model with regime switching and find that the estimated high volatility persistence is reduced when the regimes are incorporated in the model. Shibata and Watanabe (2005) also find that the persistence parameter estimates drop as compared to those of the standard SV models. Moreover, for their data, the MSSV model performs better than the benchmark SV models. Similar findings are also present in Vo (2009) who models oil price movements. Carvalho and Lopes (2007) use an auxiliary particle filter (APF) to sequentially learn about states and parameters of the MSSV model and show the predictive superiority of the MSSV model.

In general, the estimation of SV-type models is rather complex given the unobservable nature of the volatility. The Markov Chain Monte Carlo (MCMC) is the stan-

dard approach in Bayesian context, with the seminal work by Jacquier et al. (1994). For a survey on Bayesian estimation of time-varying volatility models see Virbickaite et al. (2015). Even though MCMC methods are considered to be the gold standard among Bayesian estimation methods, they are computationally costly and inherently non-sequential (Lopes and Polson, 2010). A cost-efficient alternative to MCMC is sequential Monte Carlo (SMC) methods, also known as particle filters, that allow for online type inference by updating the posterior distribution as new observations arrive. By construction, Stochastic Volatility models are state-space models, naturally suggesting the use of particle filters for estimation. Moreover, the model proposed in this paper belongs to a class o models that have the availability of sufficient statistics of the parameters, see Storvik (2002). This permits to track a low-dimensional set of sufficient statistics instead of a high-dimensional vector of parameters. The use of sufficient statistics has been shown to increase the efficiency of the algorithm by reducing the variance of the sampling weights, see Carvalho et al. (2010a). In this paper we make use of the Particle Learning (PL) approach, which is a particle based method, firstly introduced by Carvalho et al. (2010a). For general intoduction to PL and comparison with MCMC see Carvalho et al. (2010a), Lopes and Polson (2010), among others. Warty et al. (2017) propose a sequential estimation algorithm for the SV model with variance-gamma jumps in the returns. The algorithm, a hybrid between the APF and PL, is compared to the MCMC output and used for a real-data application. In a recent paper Virbickaite et al. (2018) have designed a Particle Learning algorithm for a semi-parametric Stochastic Volatility model of Delatola and Griffin (2011). The authors conduct an extensive comparison with the MCMC estimation output and show that both estimation methods present almost identical posterior distributions for model parameters, filtered volatility states and the distribution of the error term. In this paper we construct a PL algoritm similar to the one in Virbickaite et al. (2018) and augment it to include the Markov switching regimes.

The rest of the paper is structured as follows. Section 2 presents the linearized SV model with non-parametric errors and introduces a new MSSV-DPM model. Section 3 designs a PL algorithm for inference and prediction and presents a simulated data example. Section 4 evaluates the performance of the proposed model by using real data. Finally, Section 5 concludes.

2 MSSV-DPM Model

We start this section by reviewing a benchmark Stochastic Volatility model with Normal errors. We then relax the Normality assumption and present the semi-parametric SV model, similar to the one seen in Delatola and Griffin (2011). The innovation distribution is assumed to follow an infinite mixture of Gaussians via Dirichlet Process Mixture models, giving rise to a SV-DPM model. Finally, we augment the semi-parametric SV model with Markov Switching regimes in the volatility equation, resulting into a novel MSSV-DPM model.

Denote y_t as the de-meaned log returns. The standard discrete SV model has the following form:

$$y_t = \exp\{h_t/2\}\, v_t, \tag{1}$$

$$h_t = \alpha + \beta h_{t-1} + \tau \eta_t, \tag{2}$$

where β is the volatility persistence parameter such that $|\beta| < 1$ for the stationarity of the volatilities; v_t and η_t are uncorrelated error terms, such that $\eta_t \sim \mathcal{N}(0,1)$. The distribution of the v_t has zero mean and unit variance and can take many different forms: from the standard Normal, to heavy-tailed Student-t and others (see Kim et al., 1998, Chib et al., 2002b, Mahieu and Schotman, 1998, Liesenfeld and Richard, 2006, for example).

Kim et al. (1998) proposed a linearization of the standard SV model by defining $r_t = \log y_t^2$ and $\epsilon_t = \log v_t^2$, resulting into the following dynamic linear model:

$$r_t = h_t + \epsilon_t, \tag{3}$$

$$h_t = \alpha + \beta h_{t-1} + \tau \eta_t. \tag{4}$$

The distribution of ϵ_t is $\log \chi_1^2$ if v_t in (1) is Normally distributed. Kim et al. (1998) and Omori et al. (2007) use carefully tuned finite mixtures of Normals to approximate the $\log \chi_1^2$ distribution and use a data augmentation argument to design fast MCMC schemes that jointly sample $\{h_1, \ldots, h_T\}$ based on the well-known forward filtering, backward sampling (FFBS) algorithm of Carter and Kohn (1994) and Fruhwirth-Schnatter (1994). However, if v_t is not Normally distributed, then approximations of Kim et al.

(1998) and Omori et al. (2007) are not appropriate anymore.

2.1 DPM errors

As mentioned in the Introduction, it has been shown in multiple empirical studies that the distribution of the returns has heavier tails than permitted by the Normal distribution. Is this is the case, the distribution of ϵ_t in (3) is not $\log \chi_1^2$ anymore. Delatola and Griffin (2011, 2013) propose to approximate the distribution of ϵ_t as an infinite mixture of Normals by relying on DPM models. Dirichlet Process Mixture models, firstly introduced by Lo (1984), have been widely used for modeling time-varying volatilities with univariate and multivariate SV and GARCH-type models, see Jensen and Maheu (2010, 2013, 2014), Delatola and Griffin (2011, 2013), Kalli et al. (2013), Ausín et al. (2014), Virbickaite et al. (2016), Zaharieva et al. (2017).

As seen in Escobar and West (1995), the DPM model has the following representation:

$$f(\epsilon_t; G) = \int k(\epsilon_t; \theta_t) dG(\theta_t), \tag{5}$$

where k is some density kernel with parameter vector θ_t and the mixing distribution G has a Dirichlet Process prior, denoted here by $G \sim \mathcal{DP}(c, G_0(\theta; \varrho))$. Each observation ϵ_t comes from a kernel density $k(\cdot)$ with some parameters θ_t, following the mixing distribution G. The parameter c is called the concentration parameter and $G_0(\theta; \varrho)$ is called the base distribution with certain hyperparameters ϱ. The concentration parameter c can be seen as the prior belief about the number of clusters in the mixture. Small values of c assume *a priori* an infinite mixture model with only few components that have large weights. Meanwhile large values of c assume *a priori* an infinite mixture model with many components and all the weights being very small. c is also called a precision parameter and indicates how close G is to the base distribution G_0, where larger c indicates that G is closer to G_0.

Gaussian kernel and conjugate base prior. One of the most popular DPM model variants in the financial time series context assumes a Gaussian kernel for $k(\epsilon_t; \theta_t)$ in (5), i.e. $\epsilon_t \sim \mathcal{N}(\mu_t, \sigma_t^2)$. Then, the conjugate base prior $G_0(\mu, \sigma^2; \varrho)$ is a Normal - Inverse

Gamma prior, denoted here by $G_0 \sim \mathcal{NIG}(\mu, \sigma^2; m_0, V_0, a_0, a_0\sigma_0^2)$, such that $\mu|\sigma^2$ is Normal, $\mathcal{N}(\mu; m_0, V_0\sigma^2)$, and σ^2 is Inverse Gamma, $\mathcal{IG}(\sigma^2; a_0/2, a_0\sigma_0^2/2)$. Here m_0, V_0, a_0 and $a_0\sigma_0^2$ are the hyper-parameters in ϱ.

2.2 Markov switching volatility

As mentioned in the Introduction, the benchmark Stochastic Volatility model has certain limitations. In particular, it does not account for structural changes in the volatility process and if such regime shifts are ignored the persistence parameter is overestimated. In other words, the β parameter in (4) is very close to one and the volatility equation approaches the non-stationary process. In order to incorporate such changes in the regimes, So et al. (1998) introduced the MSSV model, where the log volatility equation is of the following form:

$$h_t = \alpha_{s_t} + \beta h_{t-1} + \tau\eta_t, \ \ \eta_t \sim \mathcal{N}(0,1). \tag{6}$$

Here s_t are the regime variables following a two-state first order Markov process:

$$p_{ij} = \mathrm{P}\left[s_t = j | s_{t-1} = i\right], \ \text{for} \ i, j = 0, 1.$$

As seen in Carvalho and Lopes (2007), it is necessary to introduce the following reparametrization for α_{s_t} in order to avoid identification issues:

$$\alpha_{s_t} = \gamma_0 + \gamma_1 \boldsymbol{I}\left\{s_t = 1\right\}, \ \ \gamma_0 \in \Re \ \text{and} \ \ \gamma_1 > 0.$$

Here $\boldsymbol{I}\{s_t = 1\}$ is an indicator function that takes values equal to one if the volatility is in the high state ($s_t = 1$) and zero in the low state ($s_t = 0$). The transition matrix between the states 0 and 1 is defined as:

$$T = \begin{bmatrix} P(s_t = 0|s_{t-1} = 0) & P(s_t = 1|s_{t-1} = 0) \\ P(s_t = 0|s_{t-1} = 1) & P(s_t = 1|s_{t-1} = 1) \end{bmatrix} = \begin{bmatrix} p & 1-p \\ 1-q & q \end{bmatrix}. \tag{7}$$

There are quite a few papers that consider regime switching SV models in Bayesian

context. Kalimipalli and Susmel (2004) have proposed a two-factor SV model with regime switches and estimated it using Gibbs sampler. They find that the estimate of high volatility persistence is reduced when the regimes are incorporated in the model. Also, the authors compare the new model with other two alternative two-factor models, simple SV and GARCH, and find that SV always outperforms GARCH, both in sample and out of sample. Shibata and Watanabe (2005) design a MCMC scheme to estimate the MSSV model and find that the persistence parameter estimates drop as compared to those of the standard SV models. Moreover, the MSSV model performs better than the benchmark SV models. Lopes and Carvalho (2007) extend the SV model to multivariate case and present a Factor Stochastic Volatility (FSV) model with Markov switching jumps. They construct a novel MCMC scheme for inference and find that the new model can capture market crashes in an instantaneous way, as opposed to the traditional FSV models. Carvalho and Lopes (2007) have constructed a sequential Monte Carlo filter by combining auxiliary particle filter with the filter of Liu and West (2001) to estimate a SV model with Markov switching regimes. They find that in terms of predictions the Markov switching SV specification outperforms a simple SV model. Abanto-Valle et al. (2010b) investigate the relationship between stock return volatility and trading volume by using a MSSV specification, and also find that the persistence parameter drops significantly after introducing the Markov switching jump.

Define $\Phi = \left(\gamma_0, \gamma_1, \beta, \tau^2, p, q\right)$ as a set of parameters associated with the volatility equation, $\Omega = \{(\mu, \sigma^2)^{(j)}\}_{j=1}^{\infty}$ as a set of parameters associated with the distribution of the error term, and $\Theta = (\Phi, \Omega)$ as a complete set of model parameters. Therefore, the complete MSSV-DPM model is a linearlized SV model in (3)-(4) with DPM errors in (5) that accommodates the regime-shifting structure in (6)-(7), and can be written as:

$$r_t|h_t, \Theta \sim \frac{1}{c+t-1} \sum_{j=0}^{L_{t-1}^{\star}} n_{t-1,j} \mathcal{N}(r_t; \mu_j + h_t, \sigma_j^2), \tag{8}$$

$$h_t|h_{t-1}, \lambda_t, \Theta \sim \mathcal{N}(h_t; \gamma_0 + \gamma_1 \lambda_t + \beta h_{t-1}, \tau^2), \tag{9}$$

$$\lambda_t|\Theta \sim \mathcal{BER}\left((1-p)^{1-\lambda_{t-1}} q^{\lambda_{t-1}}\right). \tag{10}$$

Here $n_{t,j}$ is a number of observations assigned to the j^{th} component at time t, $n_0 = c$,

$L_t^\star$ is a number of non-empty components in the mixture at time t, i.e. $L_t^\star$ is not fixed *a priori* and grows if new components are observed. Given this missing information, the mixture becomes finite, and the upper limit for the number of components is the number of observations. In practice, data tends to cluster, meaning that some observations come from the same component, thus $L_t^\star \leq t$. Also, $\mathcal{BER}(\pi)$ denotes a Bernoulli distribution with parameter π and λ_t is a Bernoulli distributed state variable that takes value 1 if the volatility is in the high regime and zero otherwise. The newly proposed MSSV-DPM model contains the SV-DPM model as a special case when there is only one regime, i.e. $p = 0$. It also nests the benchmark SV and MSSV models with Normal innovations when $L_t^\star = 1 \forall t = 1, \ldots, T$.

3 Estimation and simulation study

In this section we present the algorithm to perform PL estimation for the novel MSSV-DPM model. By using simulated data we show that the estimation algorithm is able to precisely estimate the parameters, the density of the squared log returns, filter the latent log volatilities and volatility regimes.

3.1 Particle Learning for the MSSV-DPM model

In this section we modify and augment the PL algorithm presented in Virbickaite et al. (2018) to include the Markov switching specification. PL, as mentioned before, is one of several particle filters that allow to perform sequential state filtering and parameter learning. PL, which was firstly introduced by Carvalho et al. (2010a), allows for sequential filtering, smoothing and parameter learning by including state-sufficient statistics in a set of particles. For a more detailed explanation of PL with illustrations refer to Carvalho et al. (2010a) and Lopes et al. (2011), among others. For comparison between PL and MCMC for the SV-DPM model refer to Virbickaite et al. (2018), and for comparison between APF+PL and MCMC for the SV model with variance-gamma jumps in the returns refer to Warty et al. (2017).

The priors for model parameters and the initial states are chosen to be condition-

ally conjugate: $h_0 \sim \mathcal{N}(c_0, C_0)$, $\sigma^2 \sim \mathcal{IG}(a_0/2, a_0\sigma_0^2/2)$, $\mu|\sigma^2 \sim \mathcal{N}(m_0, V_0\sigma^2)$, $\tau^2 \sim \mathcal{IG}(b_0/2, b_0\tau_0^2/2)$, $\beta|\tau^2 \sim \mathcal{TN}_{(-1,1)}(m_\beta, V_\beta\tau^2)$, $\gamma_0 \sim \mathcal{N}(m_{\gamma_0}, V_{\gamma_0})$, $\gamma_1 \sim \mathcal{TN}_{(0,+\infty)}(m_{\gamma_1}, V_{\gamma_1})$, $p \sim \mathcal{B}(\alpha_p, \beta_p)$ and $q \sim \mathcal{B}(\alpha_q, \beta_q)$. Here $\mathcal{TN}_{(a,b)}$ represents a Normal distribution, truncated at a and b, $\mathcal{B}$ is Beta distribution and c_0, C_0, a_0, $a_0\sigma_0^2$, m_0, V_0, b_0, $b_0\tau_0^2$, m_β, V_β, m_{γ_0}, V_{γ_0}, m_{γ_1}, V_{γ_1}, α_p, β_p, α_q and β_q are the fixed hyper-parameters.

Call S_t a set of sufficient statistics which contains all updated hyper-parameters, necessary for the parameter simulation, as well as the three kinds of filtered state variables: the latent log volatilities h_t, the indicator variable k_t, which tells us to which mixture component the data point belongs to, and λ_t, the volatility regime indicator. The object we call particle at time t contains S_t. All necessary parameters can be easily simulated given the set of sufficient statistics. At each time t we have a collection of N particles that provide approximations to the densities of interest. When this set of N particles passes from one time to another, t to $t+1$, some of the particles disappear (the ones that are not representative with respect to the new data point), and some are repeated more than once to take their place (see the 'Resampling' step below). Then, this resampled set of particles is modified to include the information from the new data point (see the 'Sampling' and 'Propagating' steps below).

In order to initiate the algorithm, initial parameter values are simulated from their corresponding priors. The initial set of sufficient statistics S_0 consists of: $\{h_0^{(i)}\}_{i=1}^N$ which has been simulated from its prior, $\{k_t^{(i)}\}_{i=1}^N$ which at $t = 0$ are all set equal to 1, since when the first observation arrives, it will belong to the first and only component, initial volatility regime $\{\lambda_0\}_{i=1}^N = 0$, and initial hyper-parameters $\{a_0^{(i)}\}_{i=1}^N, \{a_0\sigma_0^{2(i)}\}_{i=1}^N, \ldots$, which at time $t = 0$ are all the same across all particles. Then, for $t = 1 \ldots, T$ and for each particle (i) the algorithm iterates through the following steps. For notation simplicity, we do not include the indicator (i) that refers to a single particle, where $(i) = 1, \ldots, N$.

1. **Resampling**.

   Resample the $(i) = 1, \ldots, N$ particles with weights proportional to the predictive density of the log squared returns $r_t = \log y_t^2$:

$$w^{(i)} \propto \frac{1}{c+t-1} \sum_{j=0}^{L_{t-1}^\star} n_j f_N(r_t; \gamma_0 + \gamma_1\lambda_{t-1} + \beta h_{t-1} + \mu_j, \tau^2 + \sigma_j^2).$$

Here $\Theta = (\gamma_0, \gamma_1, \beta, \tau^2, p, q, \mu_1, \ldots, \mu_{L^\star_{t-1}}, \sigma_1^2, \ldots, \sigma^2_{L^\star_{t-1}})$ have been simulated at the end of the previous period. The resampled particles are denoted by a tilde above the particle, as in $\tilde{\Theta}$.

2. **Sampling.**

   (a) Sample new states of the log volatilities λ_t:

$$\lambda_t | \tilde{\lambda}_{t-1}, \tilde{h}_{t-1}, \tilde{\Theta}, r_t \sim \mathcal{BER}\left(\frac{z_2}{z_1 + z_2}\right),$$

   where

$$\begin{aligned} z_1 =& \left[\frac{1}{c+t-1}\sum_{j=1}^{\tilde{L}^\star_{t-1}} \tilde{n}_j f_N(r_t; \tilde{\gamma}_0 + \tilde{\beta}\tilde{h}_{t-1} + \tilde{\mu}_j, \tilde{\tau}^2 + \tilde{\sigma}_j^2) + \right. \\ & \left. \frac{c}{c+t-1} f_N(r_t; \tilde{\gamma}_0 + \tilde{\beta}\tilde{h}_{t-1} + \mu_0, \tilde{\tau}^2 + \sigma_0^2)\right] \times \Pr(\lambda_t = 0 | \tilde{\lambda}_{t-1}, \tilde{\Theta}), \\ z_2 =& \left[\frac{1}{c+t-1}\sum_{j=1}^{\tilde{L}^\star_{t-1}} \tilde{n}_j f_N(r_t; \tilde{\gamma}_0 + \tilde{\gamma}_1\tilde{\lambda}_{t-1} + \tilde{\beta}\tilde{h}_{t-1} + \tilde{\mu}_j, \tilde{\tau}^2 + \tilde{\sigma}_j^2) + \right. \\ & \left. \frac{c}{c+t-1} f_N(r_t; \tilde{\gamma}_0 + \tilde{\gamma}_1\tilde{\lambda}_{t-1} + \tilde{\beta}\tilde{h}_{t-1} + \mu_0, \tilde{\tau}^2 + \sigma_0^2)\right] \times \Pr(\lambda_t = 1 | \tilde{\lambda}_{t-1}, \tilde{\Theta}). \end{aligned} \tag{11}$$

   Call $\tilde{\alpha} = \tilde{\gamma}_0 + \tilde{\gamma}_1\lambda_t$.

   (b) Sample new log volatilities h_t:

$$h_t | \tilde{h}_{t-1}, \tilde{\Theta}, \tilde{L}^\star_{t-1}, \lambda_t, r_t \sim \sum_{j=0}^{\tilde{L}^\star_{t-1}} \frac{\tilde{n}_j}{c+t-1} \mathcal{N}(h_t; m_{hj}, V_{hj}),$$

   where

$$m_{hj} = \frac{\tilde{\tau}^2(r_t - \tilde{\mu}_j) + \tilde{\sigma}_j^2(\tilde{\alpha} + \tilde{\beta}\tilde{h}_{t-1})}{\tilde{\tau}^2 + \tilde{\sigma}_j^2} \text{ and } V_{hj} = \frac{\tilde{\sigma}_j^2\tilde{\tau}^2}{\tilde{\sigma}_j^2 + \tilde{\tau}^2}.$$

   For each particle we sample h_t from a mixture of $L^\star_{t-1} + 1$ components with the corresponding weights from the previous period.

   (c) Sample new indicators k_t from $\{1, \ldots, L^\star_{t-1} + 1\}$, with weights proportional

to:

$$\tilde{n}_j f_N(r_t; \tilde{\alpha} + \tilde{\beta}\tilde{h}_{t-1} + \tilde{\mu}_j, \tilde{\tau}^2 + \tilde{\sigma}_j^2),\ j = 1, \ldots, L^\star_{t-1},$$

where $\tilde{n}_{L^\star_{t-1}+1} = c$ and $\sigma^2_{L^\star_{t-1}+1} = \sigma_0^2$. If $k_t \leq L^\star_{t-1}$, $n_{k_t} = \tilde{n}_{k_t} + 1$ and $L^\star_t = L^\star_{t-1}$, otherwise, $L^\star_t = L^\star_{t-1} + 1$ and $n_{k_t} = 1$.

3. **Propagating sufficient statistics and learning Θ.**

(c.1) Sample γ_0 from $\mathcal{N}(\gamma_0; m^\star_{\gamma_0}, V^\star_{\gamma_0})$, where

$$m^\star_{\gamma_0} = \frac{\tilde{m}_{\gamma_0}\tilde{\tau}^2 + \tilde{V}_{\gamma_0}(h_t - (\tilde{\gamma}_1\lambda_t + \tilde{\beta}\tilde{h}_{t-1}))}{\tilde{\tau}^2 + \tilde{V}_{\gamma_0}} \text{ and } V^\star_{\gamma_0} = \frac{\tilde{\tau}^2\tilde{V}_{\gamma_0}}{\tilde{\tau}^2 + \tilde{V}_{\gamma_0}}.$$

(c.2) Sample γ_1 from $\mathcal{TN}_{(0,+\infty)}(\gamma_1; m^\star_{\gamma_1}, V^\star_{\gamma_1})$, where

$$m^\star_{\gamma_1} = \frac{\tilde{m}_{\gamma_1}\tilde{\tau}^2 + \tilde{V}_{\gamma_1}\lambda_t(h_t - (\gamma_0 + \tilde{\beta}\tilde{h}_{t-1}))}{\tilde{V}_{\gamma_1}\lambda_t + \tilde{\tau}^2} \text{ and } V^\star_{\gamma_1} = \frac{\tilde{\tau}^2\tilde{V}_{\gamma_1}}{\tilde{\tau}^2 + \lambda_t\tilde{V}_{\gamma_1}}.$$

Call $\alpha = \gamma_0 + \gamma_1\lambda_t$.

(c.3) Sample τ^2 from $\mathcal{IG}(\tau^2; b_0^\star/2, b_0^\star\tau_0^{2\star}/2)$, where

$$b_0^\star = \tilde{b}_0 + 1 \text{ and } b_0^\star\tau_0^{2\star} = \tilde{b}_0\tilde{\tau}_0^2 + \frac{(\tilde{m}_\beta\tilde{h}_{t-1} - (h_t - \alpha))^2}{1 + \tilde{V}_\beta\tilde{h}_{t-1}^2}.$$

(c.4) Sample β from $\mathcal{TN}_{(-1,1)}(\beta; m^\star_\beta, V^\star_\beta\tau^2)$, where

$$m^\star_\beta = \frac{\tilde{m}_\beta + \tilde{V}_\beta\tilde{h}_{t-1}(h_t - \alpha)}{1 + \tilde{V}_\beta\tilde{h}_{t-1}^2} \text{ and } V^\star_\beta = \frac{\tilde{V}_\beta}{1 + \tilde{V}_\beta\tilde{h}_{t-1}^2}.$$

(c.5) Sample p from $\mathcal{B}(p; \alpha^\star_p, \beta^\star_p)$, where

$$\alpha^\star_p = \alpha_p + 1 \text{ if } \lambda_t = 0|\lambda_{t-1} = 0 \text{ and } \beta^\star_p = \beta_p + 1 \text{ if } \lambda_t = 1|\lambda_{t-1} = 0.$$

(c.6) Sample q from $\mathcal{B}(q; \alpha^\star_q, \beta^\star_q)$, where

$$\alpha^\star_q = \alpha_q + 1 \text{ if } \lambda_t = 1|\lambda_{t-1} = 1 \text{ and } \beta^\star_q = \beta_q + 1 \text{ if } \lambda_t = 0|\lambda_{t-1} = 1.$$

(c.7) Sample $\sigma_{k_t=j}$ only for that component j where the data point at time t is assigned to, i.e. $k_t = j$, from $\mathcal{IG}(\sigma^2_{k_t=j}; a_0^\star/2, a_0^\star\sigma^{2\star}/2)$, where

$$a_0^\star = \tilde{a}_0 + 1 \text{ and } a_0^\star\sigma_0^{2\star} = \tilde{a}_0\tilde{\sigma}_0^2 + \frac{(r_t - h_t - \tilde{m}_0)^2}{1+\tilde{V}_0}.$$

(c.8) Sample $\mu_{k_t=j}$ only for that component j where the data point at time t is assigned to, i.e. $k_t = j$, from $\mathcal{N}(\mu_{k_t=j}; m_0^\star, V_0^\star\sigma^2_{k_t=j})$, where

$$m_0^\star = \frac{\tilde{m}_0 + \tilde{V}_0(r_t - h_t)}{1+\tilde{V}_0} \text{ and } V_0^\star = \frac{\tilde{V}_0}{1+\tilde{V}_0}.$$

Parts of the derivations of the equations are available in the Appendix at the end of the manuscript, meanwhile the rest are available in Virbickaite et al. (2018).

3.2 Simulation study

In order to asses the estimation accuracy of the PL algorithm for the proposed model we use a simulated data set of $T = 3000$ observations with the following parameters: $\gamma_0 = -0.06$, $\gamma_1 = 0.20$, $\beta = 0.92$, $\tau^2 = 0.05$, $p = 0.996$, $q = 0.996$. The error term for the returns follows a standard Normal distribution $v_t \sim \mathcal{N}(0,1)$, thus the true DGP for the linearlized model is $\epsilon_t \sim \log\chi_1^2$. The hyper-parameters are $c_0 = 0$, $C_0 = 0.1$, $m_\beta = 0.95$, $V_\beta = 0.1$, $b_0 = 4$, $b_0\tau_0^2 = 0.2$, $a_0 = 5$, $a_0\sigma_0^2 = 15$, $m_0 = -1.27$ (this specific value is chosen because the mean of the $\log\chi_1^2$ distribution is equal to -1.27), $V_0 = 0.1$, $m_{\gamma_0} = 0$, $V_{\gamma_0} = 1$, $m_{\gamma_1} = 0$, $V_{\gamma_1} = 0.1$, $\alpha_p = 3$, $\beta_p = 0.1$, $\alpha_q = 3$ and $\beta_q = 0.1$. The concentration parameter c is set to be equal to 1, as in Carvalho et al. (2010b). Using the simulated data, we fit the MSSV-DPM model using PL, number of particles $N = 300$k. All codes were written in R.

Figure 1 top graph draws the simulated log returns y_t. The middle graph represents the true realization of the log volatility (in black) and the mean estimated filtered log volatility (in grey). The 95% credible intervals (CIs) for the estimated filtered log volatility almost always capture the true realization of the log volatility (not reported). The bottom graph of the same figure draws the mean probability of being in a state

one ($s_t = 1$), compared to the true state. As seen from the figure, PL takes some time to learn, since at first it is not able to distinguish the regimes well. However, around observation 1000 the algorithm is able to correctly identify the regimes with the overall regime miss-classification rate equal to 13%. Figure 2 draws the sequential estimation of the model parameters and their 95% CIs. The parameter estimates seem reasonable and very close to their true values. It takes some time to learn the true value of the parameter q because there is no information available about it before the regime switch, which happens around time $t = 700$. The estimation of the parameter τ^2 is the least precise due to the fact that this parameter represents the volatility of the volatility, and is notoriously difficult to estimate. Finally, Figure 3 draws the estimated non-parametric density of the log squared error term, which is almost identical to the true DGP, which is $\log \chi^2_1$ distributed.

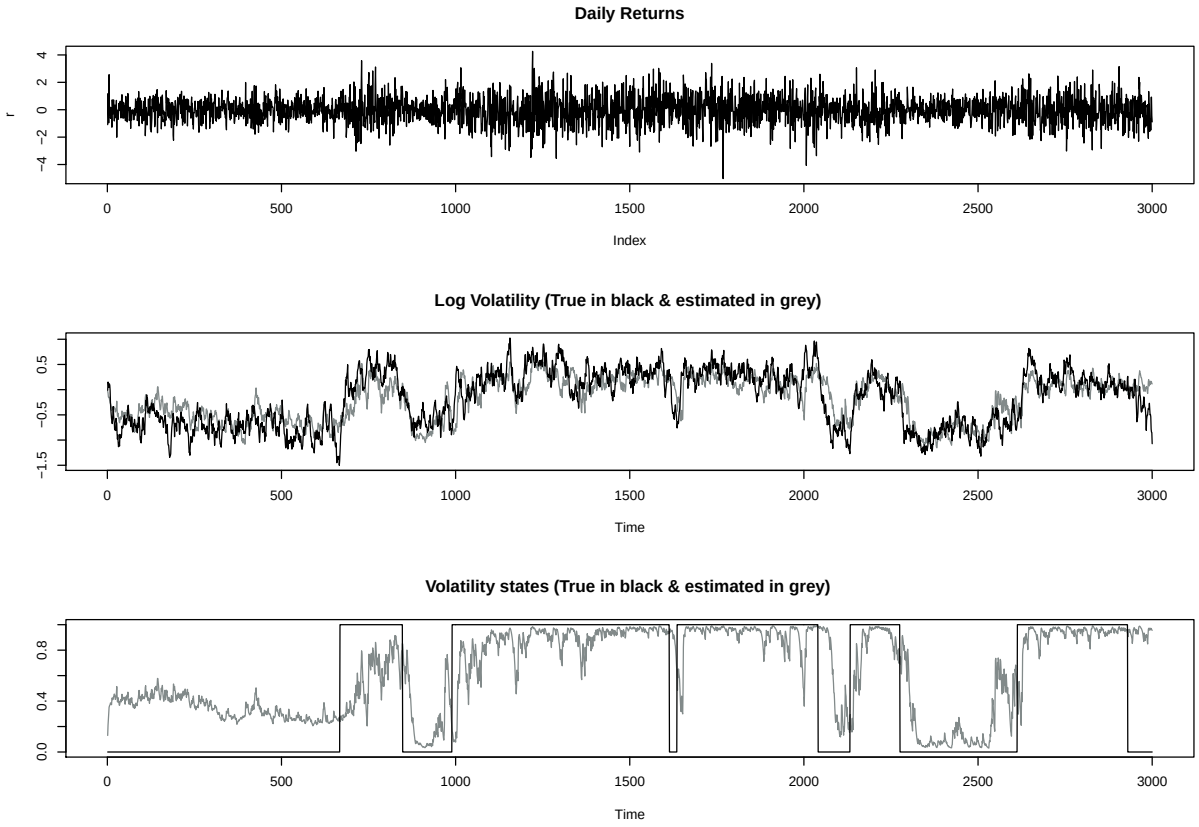
Overall, the obtained estimation results seem quite reasonable and PL is able to correctly identify the hidden volatility regimes, filter log volatilities, estimate the density of the errors and the parameters in an efficient sequential manner.

However, one has to be aware of the shortcomings of the PL procedure. Virbickaite et al. (2018) summarize the main weaknesses of particle filters, with the main disadvantage being an ever-decreasing set of atoms in the particle approximation of the density of interest. As noted by Chopin et al. (2011), increasing the number of observations will lead to degenerating paths, unless the number of particles is being increased simultaneously. Therefore, the use of PL or any particle-based filter in general is advantageous only if one is interested in fast one-step-ahead predictions. However, once the number of observations has increased one should consider restarting the filter at a later time t with a smaller number of observations, or anticipate the large sample size and employ more particles (which would slow down the estimation, since the sizes of matrices that need to be carried from one time to another increase dramatically).

4 Real Data Application

In this section we present a real data application using log returns of three financial assets: S&P500 index, Ford company and a commodity - natural gas. The S&P500 and

Figure 1: Simulated data: daily log returns (top), true and estimated log volatilities (middle), true and estimated volatility regimes (bottom).

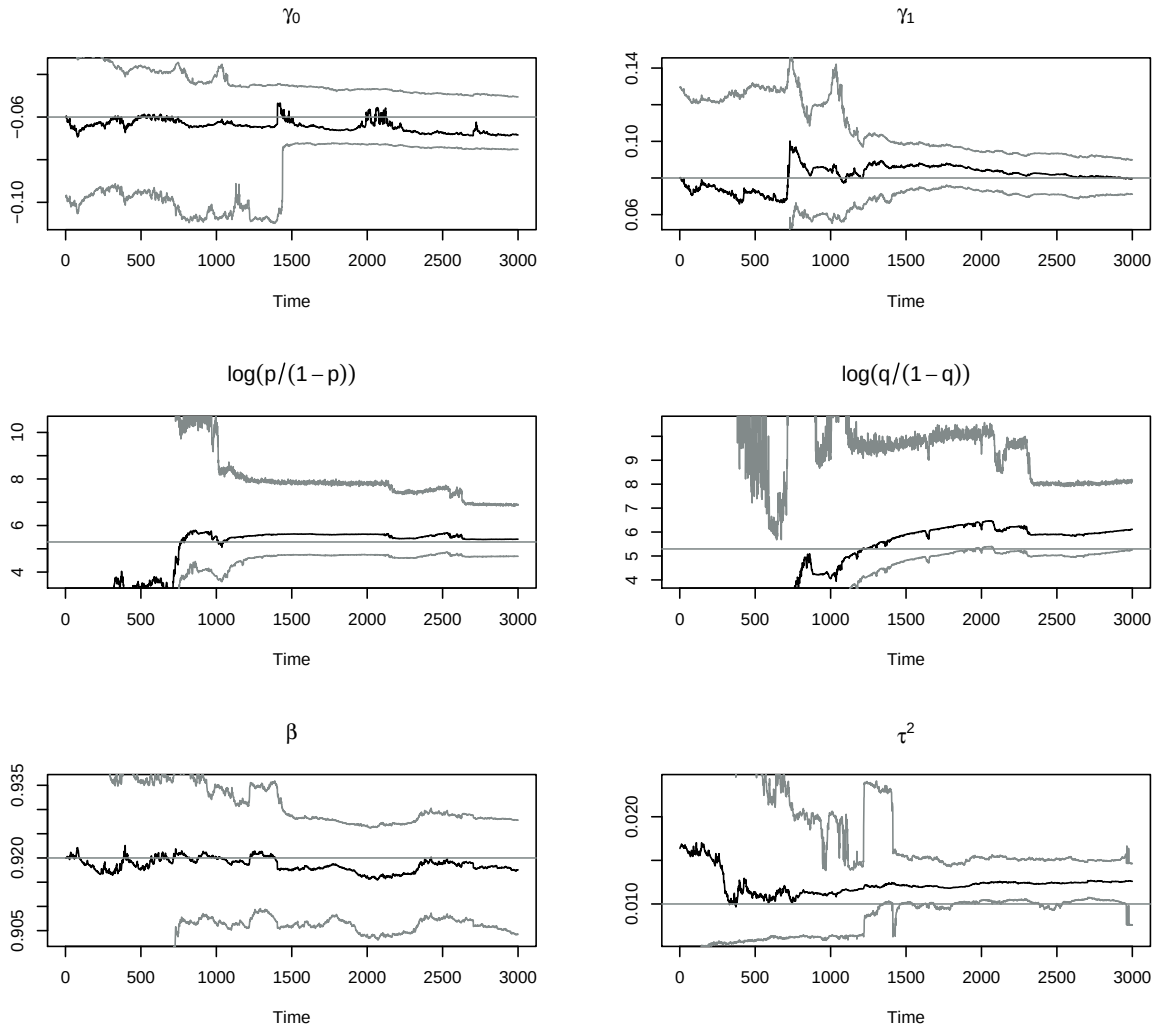


Ford prices are from Jan 2nd 1997 till Sept 9th 2014 and Henry Hub natural gas spot prices (dollars per million btu) are from Jan 5th 1997 till Sept 9th 2014. All data is obtained from the Datastream database. The descriptive statistics are in Table 1 and the descriptive graphs are in Figure 4.

[Table 1 about here.]

Next, using the de-meaned data we fit two semi-parametric models, SV-DPM and the newly introduced MSSV-DPM. The hyper-parameters for the priors are the same as in the simulation study in Section 3.2. Note that the SV-DPM model is a restricted version of the MSSV-DPM model where the probability of staying in the same regime is set equal to one, $p = 1$, therefore, neither q, nor γ_1 nor λ_t are possible to estimate. Also,

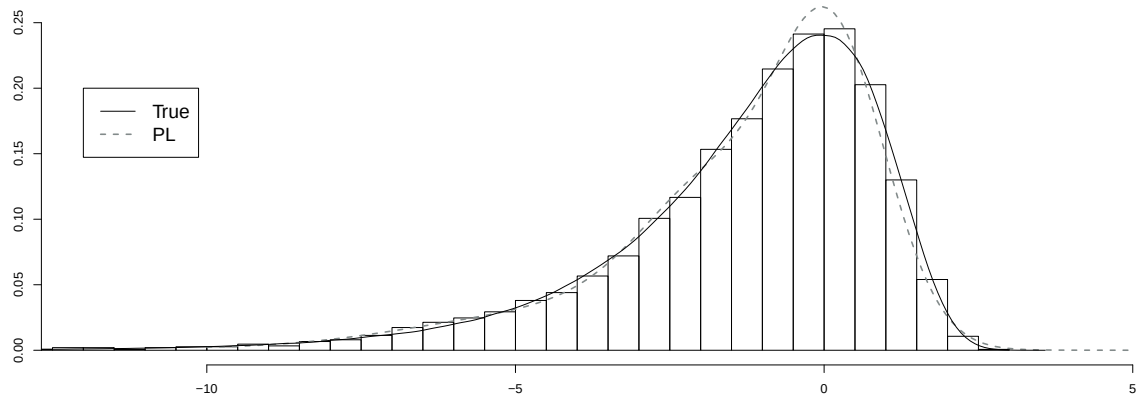
Figure 2: Simulated data: sequential estimation of the model parameters, their 95% credible intervals and their true values.



as mentioned before, both models nest a benchmark SV model with Normal errors as a special case. The number of particles is set to $N = 500$k since the number of observations is larger than in the simulation study (4000+).

To compare the performance of the models, we use the average log predictive score (LPS) and average log predictive tail score (LPTS_α). LPTS_α considers the predictive performance only in the upper $100\alpha\%$ of the empirical distribution of the squared log returns. LPTS_α was also employed by Delatola and Griffin (2011). As the authors point

Figure 3: Simulated data: MSSV-DPM estimated non-parametric density of the log squared error term compared to the true DGP of $\log \chi_1^2$.



out, the LPTS_α is not a proper scoring rule, however, it can be very helpful in understanding how the model performs in the tails. The LPS is defined as follows:

$$\text{LPS} = -\frac{1}{T}\sum_{t=1}^{T} \log p(r_t|r^{t-1}),$$

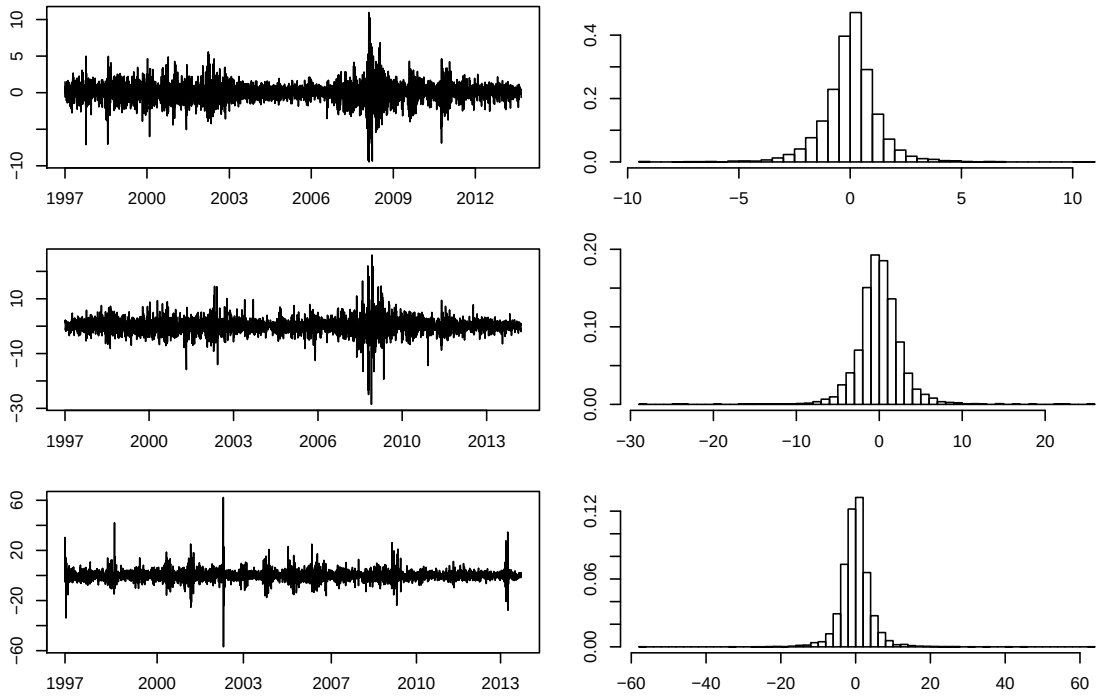
and LPTS_α is defined as:

$$\text{LPTS}_\alpha = -\frac{1}{\sum_{t=1}^{T} \boldsymbol{I}\{r_t > z_\alpha\}} \sum_{t=1}^{T} \boldsymbol{I}\{r_t > z_\alpha\} \log p(r_t|r^{t-1}),$$

where z_α is the upper 100α percentile of the empirical distribution of r_t. Note that smaller values of the LPS and LPTS_α correspond to a better model.

The log predictive densities are straightforward to obtain with the PL algorithm, since they are a by-product of the estimation procedure. For each $t = 1, \ldots, T$ the log

Figure 4: Daily log returns (in %) and histograms for the S&P500 (top), Ford (middle) and natural gas (bottom) data.



predictive densities are calculated as:

$$\log p(r_t|r^{t-1}) = \frac{1}{N}\sum_{i=1}^{N} \log p(r_t|(\Theta, h_t, k_t, \lambda_t)^{(i)}). \tag{12}$$

Differently than in MCMC setting, there is no need to fix a certain $\hat{\Theta}$ for the calculation of the LPS and LPTS$_\alpha$, and we can account for the parameter and state uncertainty by using the approximation in (12). Accounting for parameter and state uncertainty in MCMC setting at each time t without fixing certain $\hat{\Theta}$ would be prohibitively costly. The LPS and LPTS$_\alpha$ report the average predictive performance for the entire distribution and the tails of the distribution, respectively. Comparing these scores is not straightforward, since there is no established decision rule to decide whether the difference in the scores between the two models is statistically significant.

Therefore, we also report the cumulative LPS, which can be seen as a log predictive Bayes Factor (BF). Bayes factor allows for consistent model comparison even for non-nested models, it contains rewards for model fit, accounts for coherency between the prior and the information arising from the data, as well as rewards parsimony, see Koop (2003). BF between model $\mathcal{M}_1$ and model $\mathcal{M}_2$ is defined as $BF_{12} = p(D|\mathcal{M}_1)/p(D|\mathcal{M}_2)$, see Kass and Raftery (1995). Here $p(D|\mathcal{M})$ is the marginal likelihood for data D given a certain model $\mathcal{M}$. For a predictive BF, this marginal likelihood is nothing else but the log predictive density in (12). Then, the difference between such cumulative log predictive densities is a log predictive Bayes Factor (LPBF). Kass and Raftery (1995) also provide a scale for the strength of preference of one model against another, and if the $2 \times \text{LPBF} > 10$, the evidence in favor of one model against another is very strong.

Next, we present the estimation results for the S&P500 index. Figure 5 draws the estimated densities for the error term for the SV-DPM and the MSSV-DPM models as compared to the frequently used mixture of 7 Normals of Kim et al. (1998), as an approximation for $\log \chi_1^2$. SV-DPM and MSSV-DPM models estimates are very similar to each other and different from the 7-mixture approximation. This shows that the assumption of Normality is restrictive and, for this data set, would be inappropriate. As seen in Figure 6, the filtered log volatilities and volatilities for both models are very similar (second and third graphs). The differences arise during the financial crisis period, where the MSSV-DPM model estimates higher volatility. This is one of the main advantages of the Markov Switching specification, since it allows for high volatility via the shift in the regime and not by artificial inflation of the volatility persistence parameter β. The filtered volatility regimes in the MSSV-DPM model are in the bottom part of the Figure 6. The volatility of the S&P500 index is in the high regime during the years 1999-2001 and 2008-2010 (financial crisis). Table 2 presents the estimated parameter means and 95% credible intervals. The volatility persistence parameter is significantly larger for the SV-DPM model, i.e. the estimated 95% CIs do not overlap. This result is in line with the findings present in other papers, see So et al. (1998), Kalimipalli and Susmel (2004), Shibata and Watanabe (2005), Vo (2009), among others.

[Table 2 about here.]

Figure 5: Estimated densities for the log squared error term for the SV-DPM and MSSV-DPM models, as compared to the approximation of 7 Normals.

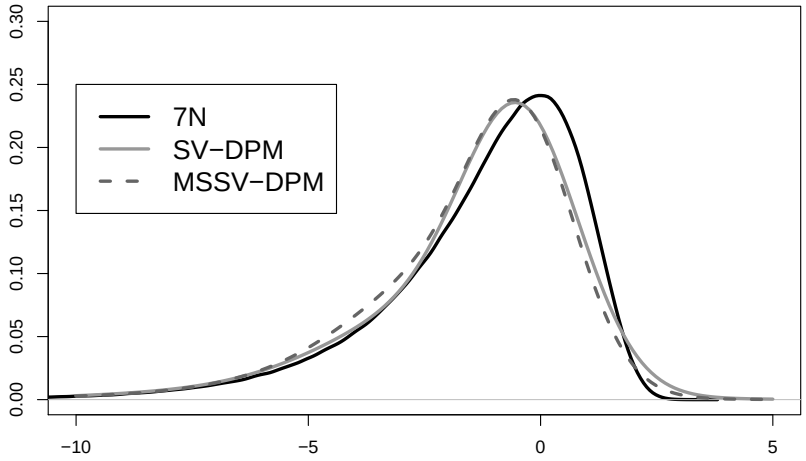
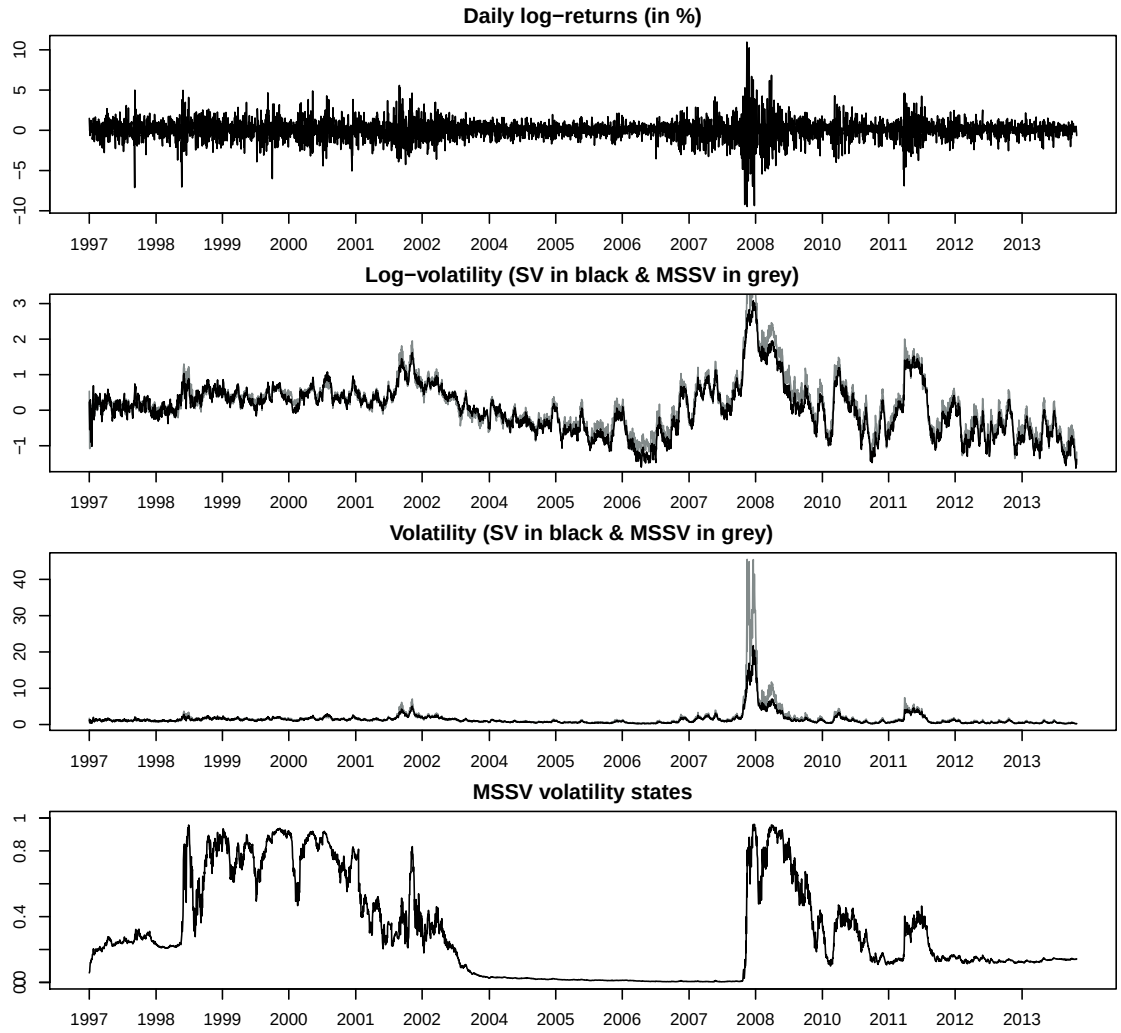


Table 3 presents the LPS, LPTS_α and LPBF for the S&P500 data. As mentioned before, the LPS measures the average predictive model performance for the entire distribution of the log squared returns and LPTS_α only for the tails. Even thought the averages are virtually indistinguishable, the cumulative differences - LPBFs - are significantly different. In particular, for the S&P500 data, the SV-DPM model performs better for the entire distribution, but the results change if we consider only the tails of the predictive distribution, where the newly introduced MSSV-DPM model provides significantly better performance.

[Table 3 about here.]

Similar results can be seen in the estimation of the other two data sets, see Tables 4, 5, 6 and 7 and Figures 7 and 8. For Ford data the estimated mean of the persistence parameter drops from 0.9514 to 0.908 and the 95% CIs do not overlap. For the natural gas data the SV-DPM model estimates the persistence parameter to be 0.8889 as compared to the MSSV-DPM model estimate of 0.8137, and the 95% CIs do not overlap either. The average LPS and LPTS_α for both models are very similar, however, the LPBFs favor the MSSV-DPM model in the entire distribution and in the tails for both data sets, see Tables

Figure 6: Filtered volatilities and volatility states for S&P500 data for SV-DPM and MSSV-DPM models.



5 and 7. Figures 7 and 8 draw the estimated volatilities and volatility states for Ford and natural gas data. The Ford returns exhibit three periods of volatility increases. The increase in the years 1998-1999 might be due to a series of lawsuits against the company, which later were dismissed. The volatility increase in 2002-2003 might be due to the early 2000s economic recession in the US, meanwhile the volatility increase in the 2008-2009 is clearly due to the global financial crisis. Natural gas data exhibits a very clear periodical pattern of shifts in the volatility regimes due to industry-specific events.

Interestingly, during the global financial crisis the natural gas volatility does not exhibit a long-lasting shift in the regime.

Important to mention, that Figures 6-8 present only filtered, but not smoothed volatilities and volatility state estimates. In general, the purpose of the analysis plays essential role. If one is interested in understanding the historical behavior of the series and the effects of, say, economic factors on the changes in volatility, it is important to consider the information from the entire sequence, as is done by MCMC. In SMC setting this can be achieved by performing the backwards smoothing procedure, where PL provides a procedure to perform direct backwards smoothing, see Carvalho et al. (2010a). If one is interested in the prediction of the volatility in the next period $t + 1$, backwards smoothing does not apply. As noted in Lopes et al. (2011), in most models estimated using PL smoothing can effectively be performed after the estimation.

[Table 4 about here.]

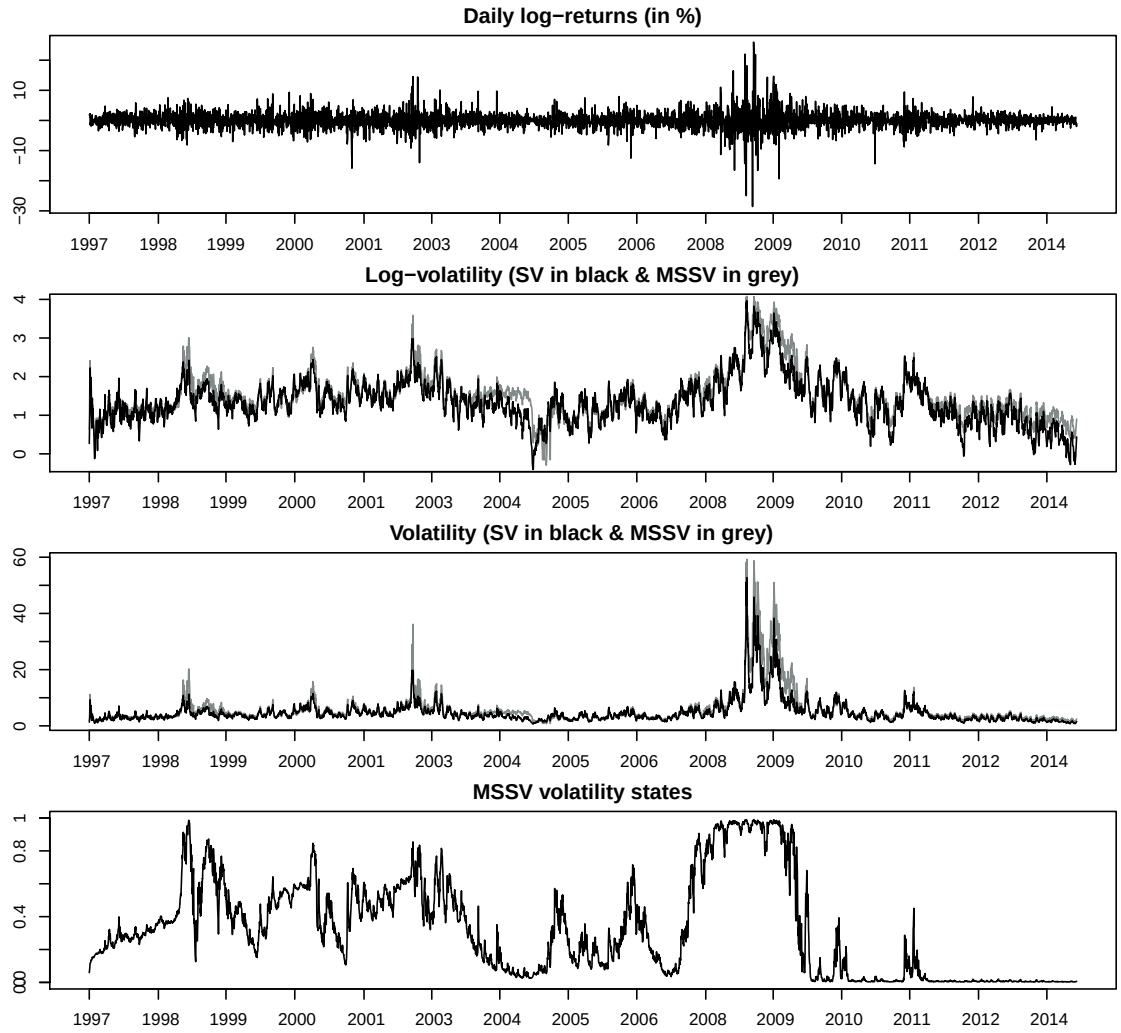
[Table 5 about here.]

[Table 6 about here.]

[Table 7 about here.]

To conclude, in majority of the cases the newly proposed MSSV-DPM model outperforms the SV-DPM model in terms of one-step-ahead prediction for the entire distribution (except for the S&P500 data) and for the tails (for all three data sets). Moreover, including the regime shifts in the mean of the volatility equation reduces the value of the estimated persistence parameter. As noted in Vo (2009), ignoring the shifts in the regimes gives the impression that the volatility is highly persistent, therefore, highly predictable, which is not the case. The half-life of a volatility shock, defined as 'the time it takes for a shock to decay half of its initial value' (Vo, 2009,) for the S&P500 data drops from 61 days to 24 days, for Ford data - from 14 to 7 days, and for natural gas data - from 6 days to 3 days.

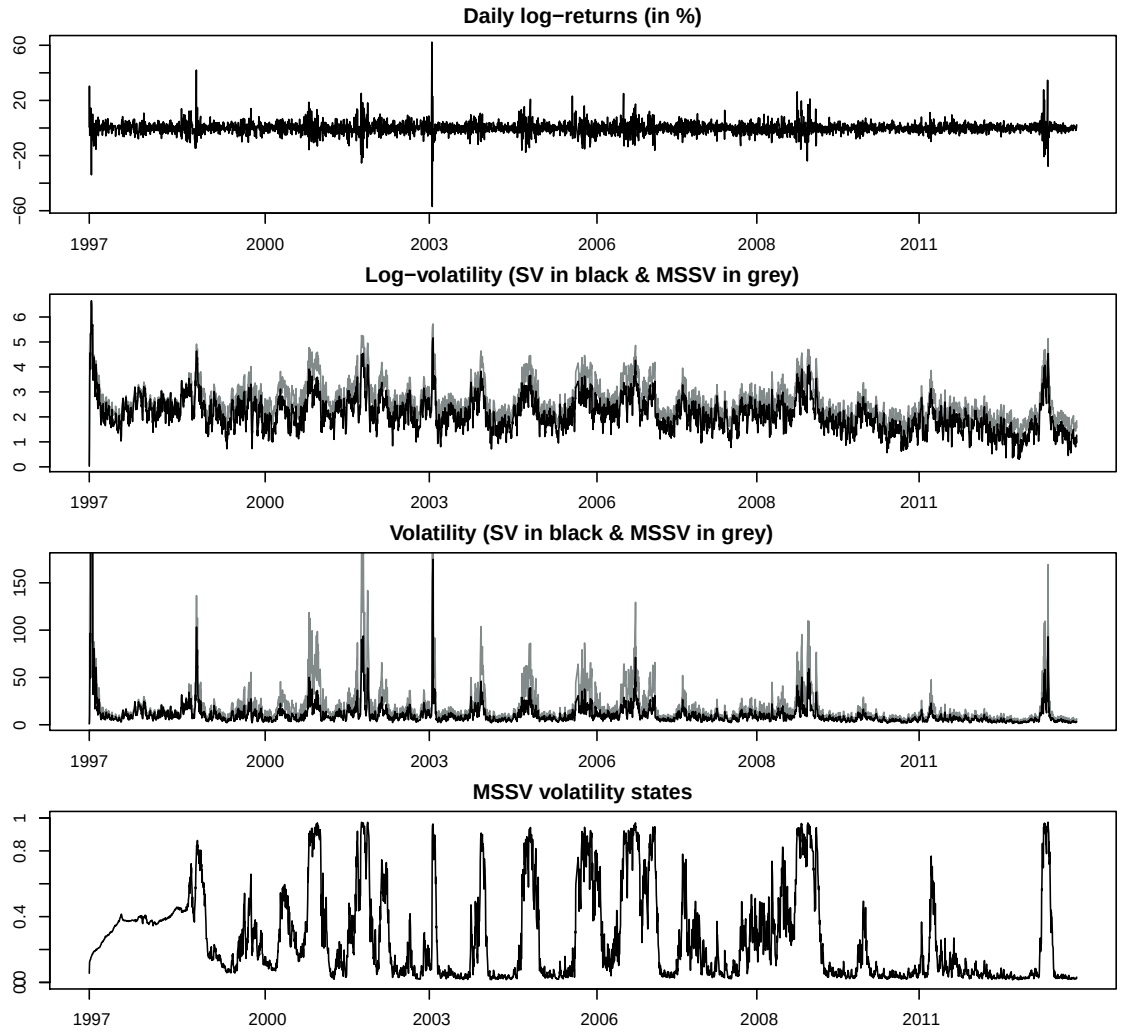
Figure 7: Filtered volatilities and volatility states for Ford data for SV-DPM and MSSV-DPM models.



5 Discussion

This paper augments the existing SV-DPM model with Markov switching jumps to capture different volatility regimes, resulting into a MSSV-DPM model. We test the newly proposed model on simulated data and find that the Particle Learning estimation procedure it is able to identify different volatility regimes. We present a real data application using three financial time series of the returns for one index - S&P500, one company -

Figure 8: Filtered volatilities and volatility states for natural gas data for SV-DPM and MSSV-DPM models.



Ford, and one commodity - natural gas. We find that the MSSV-DPM model performs significantly better than the SV-DPM model if we consider the entire predictive distribution of the returns for Ford and natural gas data, but not for the S&P500 data. If we consider the tails of the distributions, the MSSV-DPM model significantly outperforms the SV-DPM model for the three data sets for all percentiles (1, 5 and 10) of the tail. Finally, the volatility persistence parameter estimates drop significantly after including the Markov Switching specification, a finding in line with the results in multiple previous

studies. Overestimation of the volatility persistence leads to incorrect understanding of the predictability of the volatility.

Appendix: PL for MSSV-DPM

1. **Resampling.** Resample old particles (parameters and the set of sufficient statistics, including the three state variables, h_t, k_t and λ_t) with weights proportional to the predictive density of the returns, that can be obtained as below, where $p(r_t|h_t, \Theta)$ and $p(h_t|h_{t-1}, \Theta)$ are as in (8) and (9):

$$\begin{aligned}
p(r_t|h_{t-1}, \Theta) =& \int p(r_t|h_t, \Theta) p(h_t|h_{t-1}, \Theta) dh_t \\
=& \frac{1}{c+t-1} \sum_{j=0}^{L^\star_{t-1}} n_j \int f_{\mathcal{N}}(r_t; h_t + \mu_j, \sigma_j^2) f_{\mathcal{N}}(h_t; \gamma_0 + \gamma_1 \lambda_t + \beta h_{t-1}, \tau^2) dh_t \\
=& \dots \int \frac{\exp\{-(r_t - (h_t + \mu_j))^2/(2\sigma_j^2)\}}{\sqrt{2\pi\sigma_j^2}} \frac{\exp\{-(h_t - (\gamma_0 + \gamma_1\lambda_t + \beta h_{t-1}))^2/(2\tau^2)\}}{\sqrt{2\pi\tau^2}} dh_t \\
=& \frac{1}{c+t-1} \sum_{j=0}^{L^\star_{t-1}} n_j f_N(r_t; \gamma_0 + \gamma_1\lambda_t + \beta h_{t-1} + \mu_j, \tau^2 + \sigma_j^2),
\end{aligned}$$

   where $n_0 = c$ and $(\gamma_0, \gamma_1, \beta, \tau^2, p, q, \mu_1, \dots, \mu_{L^\star_{t-1}}, \sigma_1^2, \dots, \sigma^2_{L^\star_{t-1}})$ have been simulated at the end of the previous period.

2. **Sampling.** In this step we propagate the latent states h_t, the latent volatility states λ_t and the indicator variables k_t, that indicate to which mixture component the observation belongs to. Note that the tilde above the parameter indicates that the particle has been resampled in the first step.

   (a) The volatility state variable λ_t is propagated according to the following:

$$\begin{aligned}
p(\lambda_t|\tilde{\lambda}_{t-1}, \tilde{h}_{t-1}, \tilde{\Theta}, r_t) &\propto p(r_t|\tilde{\lambda}_{t-1}, \tilde{h}_{t-1}, \tilde{\Theta}) p(\lambda_t|\tilde{\lambda}_{t-1}) \\
\lambda_t|\tilde{\lambda}_{t-1}, \tilde{h}_{t-1}, \tilde{\Theta}, r_t &\sim \mathcal{BER}\left(\frac{z_2}{z_1+z_2}\right),
\end{aligned}$$

   where z_1 and z_2 as in (11) and $\tilde{\alpha} = \tilde{\gamma}_0 + \tilde{\gamma}_1\lambda_t$.

(b) For sampling the h_t make use of $p(h_t|\tilde{h}_{t-1}, r_t, \lambda_t, \tilde{\Theta}) \propto p(r_t|h_t, \tilde{\Theta})p(h_t|\tilde{h}_{t-1}, \lambda_t, \tilde{\Theta})$, where $p(r_t|h_t, \Theta)$ and $p(h_t|h_{t-1}, \lambda_t, \Theta)$ are as in (8) and (9):

$$p(h_t|\tilde{h}_{t-1}, \tilde{\Theta}, \tilde{n}, \tilde{L}^{\star}_{t-1}, \lambda_t, r_t) \propto \sum_{j=0}^{\tilde{L}^{\star}_{t-1}} \frac{\tilde{n}_j}{c+t-1} f_{\mathcal{N}}(r_t; h_t + \tilde{\mu}_j, \tilde{\sigma}_j^2) f_{\mathcal{N}}(h_t; \tilde{\alpha} + \tilde{\beta}\tilde{h}_{t-1}, \tilde{\tau}^2)$$

$$h_t|\tilde{h}_{t-1}, \tilde{\Theta}, \tilde{n}, \tilde{L}^{\star}_{t-1}, \lambda_t, r_t \sim \sum_{j=0}^{L^{\star}_{t-1}} \frac{\tilde{n}_j}{c+t-1} \mathcal{N}(h_t; m_{hj}, V_{hj}),$$

where, $V_{hj} = A_j\tilde{\sigma}_j^2$, $m_{hj} = A_j(r_t - \tilde{\mu}_j) + (1 - A_j)(\tilde{\alpha} + \tilde{\beta}\tilde{h}_{t-1})$, $A_j = \tilde{\tau}^2/(\tilde{\tau}^2 + \tilde{\sigma}_j^2)$ and $\tilde{\alpha} = \tilde{\gamma}_0 + \tilde{\gamma}_1\lambda_t$.

(c) For sampling new indicators k_t, make use of $p(k_t = j|r_t, \tilde{h}_{t-1}, \tilde{\Theta}) \propto p(r_t|k_t = j, \tilde{h}_{t-1}, \tilde{\Theta})p(k_t = j|\tilde{h}_{t-1}, \tilde{\Theta})$, where

$$p(r_t|k_t = j, \tilde{h}_{t-1}, \tilde{\Theta}) = \int p(r_t|h_t, \tilde{\Theta})p(h_t|\tilde{h}_{t-1}, \tilde{\Theta})dh_t$$

and

$$p(k_t = j|\tilde{h}_{t-1}, \tilde{\Theta}) \propto \frac{\tilde{n}_j}{c+t-1},$$

therefore,

$$p(k_t = j|r_t, \tilde{h}_{t-1}, \tilde{\Theta}) \propto \tilde{n}_j f_N(r_t; \tilde{\alpha} + \tilde{\beta}\tilde{h}_{t-1} + \tilde{\mu}_j, \tilde{\tau}^2 + \tilde{\sigma}_j^2), \;\; j = 1, \ldots, L^{\star}_{t-1} + 1,$$

where $\tilde{n}_{L^{\star}_{t-1}+1} = c$ and $\sigma^2_{L^{\star}_{t-1}+1} = \sigma_0^2$.

3. **Propagating sufficient statistics and learning Θ.**

(c.3) & (c.4) Sampling τ^2 and β:

$$\begin{aligned}
p(\beta, \tau^2|h_t) &\propto p(h_t|\beta, \tau^2)p(\beta, \tau^2) \\
&\propto f_{\mathcal{N}}(h_t; \tilde{\alpha} + \beta\tilde{h}_{t-1}, \tau^2) f_{\mathcal{TN}_{(-1,1)}}(\beta; \tilde{m}_\beta, \tilde{V}_\beta\tau^2) f_{\mathcal{IG}}(\tau^2; \tilde{b}_0/2, \tilde{b}_0\tilde{\tau}_0^2/2) \\
\tau^2 &\sim \mathcal{IG}(\tau^2; \frac{\tilde{b}_0 + 1}{2}, \frac{\tilde{b}_0\tilde{\tau}_0^2 + (\tilde{m}_\beta\tilde{h}_{t-1} - (h_t - \tilde{\alpha})^2)/(1 + \tilde{V}_\beta\tilde{h}_{t-1}^2)}{2}) \\
\beta &\sim \mathcal{TN}_{(-1,1)}(\beta; \frac{\tilde{m}_\beta + \tilde{V}_\beta\tilde{h}_{t-1}(h_t - \tilde{\alpha})}{1 + \tilde{V}_\beta\tilde{h}_{t-1}^2}, \frac{\tilde{V}_\beta\tau^2}{1 + \tilde{V}_\beta\tilde{h}_{t-1}^2})
\end{aligned}$$

Sufficient statistics updates and sampling for the rest of the parameters is analogous to (c.3) & (c.4).

References

Abanto-Valle, C. A., Bandyopadhyay, D., Lachos, V. H., and Enriquez, I. (2010a). "Robust Bayesian analysis of heavy-tailed stochastic volatility models using scale mixtures of normal distributions." *Computational Statistics and Data Analysis*, 54(12): 2883–2898.

Abanto-Valle, C. A., Migon, H. S., and Lopes, H. F. (2010b). "Bayesian modeling of financial returns: A relationship between volatility nad trading volume." *Appl .Stochastic Models Bus. Ind.*, 26(2): 172–193.

Ausín, M. C., Galeano, P., and Ghosh, P. (2014). "A semiparametric Bayesian approach to the analysis of financial time series with applications to value at risk estimation." *European Journal of Operational Research*, 232(2): 350–358.

Barndorff-Nielsen, O. E. (1997). "Normal Inverse Gaussian Distributions and Stochastic Volatility Modelling." *Scandinavian Journal of Statistics*, 24(1): 1–13.

Bollerslev, T. (1986). "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics*, 31(3): 307 – 327.

Broto, C. and Ruiz, E. (2004). "Estimation Methods for Stochastic Volatility Models: a Survey." *Journal of Economic Surveys*, 18(5): 613–650.

Carter, A. C. K. and Kohn, R. (1994). "On Gibbs Sampling for State Space Models." *Biometrika*, 81(3): 541–553.

Carvalho, C. M., Johannes, M. S., Lopes, H. F., and Polson, N. G. (2010a). "Particle Learning and Smoothing." *Statistical Science*, 25(1): 88–106.

Carvalho, C. M. and Lopes, H. F. (2007). "Simulation-Based Sequential Analysis of Markov Switching Stochastic Volatility Models." *Computational Statistics & Data Analysis*, 51(9): 4526–4542.

Carvalho, C. M., Lopes, H. F., Polson, N. G., and Taddy, M. A. (2010b). "Particle Learning for General Mixtures." *Bayesian Analysis*, 5(4): 709–740.

Chib, S., Nardari, F., and Shephard, N. (2002a). "Markov chain Monte Carlo methods for stochastic volatility models." *Journal of Econometrics*, 108(2): 281–316.

— (2002b). "Markov Chain Monte Carlo Methods for Stochastic Volatility Models." *Journal of Econometrics*, 108(2): 281–316.

Chopin, N., Iacobucci, A., Marin, J.-M., Mengersen, K. L., Robert, C. P., Ryder, R., and Schäfer, C. (2011). "On Particle Learning." *comments on "Particle learning for sequential Bayesian computation" by Lopes, Carvalho, Johannes, and Polson. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, Smith A. F. M., and M. West, editors*.

Delatola, E.-I. and Griffin, J. E. (2011). "Bayesian Nonparametric Modelling of the Return Distribution with Stochastic Volatility." *Bayesian Analysis*, 6(4): 901–926.

— (2013). "A Bayesian Semiparametric Model for Volatility with a Leverage Effect." *Computational Statistics & Data Analysis*, 60: 97–110.

Engle, R. F. (1982). "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica*, 50(4): 987–1008.

Escobar, M. D. and West, M. (1995). "Bayesian Density Estimation and Inference Using Mixtures." *Journal of the American Statistical Association*, 90(430): 577–588.

Ferguson, T. S. (1983). "Bayesian Density Estimation by Mixtures of Normal Distributions." In Rizvi, H. and Rustagi, J. (eds.), *Recent Advances in Statistics*, volume 24, 287–302. New York: Academic Press.

Fruhwirth-Schnatter, S. (1994). "Data Augmentation and Dynamic Linear Models." *Journal of Time Series Analysis*, 15(2): 183–202.

Gallant, A., Hsieh, D., and Tauchen, G. (1997). "Estimation of Stochastic Volatility Models with Diagnostics." *Journal of Econometrics*, 81: 159–192.

Harvey, A. C., Ruiz, E., and Shephard, N. (1994). "Multivariate Stochastic Variance Models." *The Review of Economic Studies1*, 61(2): 247–264.

Jacquier, E., Polson, N. G., and Rossi, P. E. (1994). "Bayesian Analysis of Stochastic Volatility Models." *Journal of Business & Economic Statistics*, 12(4): 371–389.

— (2004). "Bayesian Analysis of Stochastic Volatility Models with Fat-Tails and Correlated Errors." *Journal of Econometrics*, 122(1): 185–212.

Jensen, M. J. and Maheu, J. M. (2010). "Bayesian Semiparametric Stochastic Volatility Modeling." *Journal of Econometrics*, 157(2): 306–316.

— (2013). "Bayesian Semiparametric Multivariate GARCH Modeling." *Journal of Econometrics*, 176(1): 3–17.

— (2014). "Estimating a semiparametric asymmetric stochastic volatility model with a Dirichlet process mixture." *Journal of Econometrics*, 178: 523–538.

Kalimipalli, M. and Susmel, R. (2004). "Regime-Switching Stochastic Volatility and Short-Term Interest Rates." *Journal of Empirical Finance*, 11(3): 309–329.

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.

Kass, R. E. and Raftery, A. E. (1995). "Bayes Factors." *Journal of the American Statistical Association*, 90(430): 773–795.

Kim, S., Shephard, N., and Chib, S. (1998). "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models." *Review of Economic Studies*, 65: 361–393.

Koop, G. (2003). *Bayesian Econometrics*. Wiley-Interscience.

Liesenfeld, R. and Richard, J. (2006). "Classical and Bayesian Analysis of Univariate and Multivariate Stochastic Volatility Models." *Econometric Reviews*, 25(2-3): 335–360.

Liu, J. and West, M. (2001). "Combined Parameter and State Estimation in Simulation-Based Filtering." In Doucet, A., Freitas, N., and Gordon, N. (eds.), *Sequential Monte Carlo Methods in Practice*, 197–223. Springer New York.

Lo, A. Y. (1984). "On a Class of Bayesian Nonparametric Estimates: I. Density Estimates." *The Annals of Statistics*, 12(1): 351–357.

Lopes, H. F. and Carvalho, C. M. (2007). "Factor Stochastic Volatility with Time Varying Loadings and Markov Switching Regimes." *Journal of Statistical Planning and Inference*, 137(10): 3082–3091.

Lopes, H. F., Carvalho, C. M., Johannes, M. S., and Polson, N. G. (2011). "Particle Learning for Sequential Bayesian Computation." In Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M. (eds.), *Bayesian Statistics 9*, volume 9, 317–360. Oxford: Oxford University Press.

Lopes, H. F. and Polson, N. G. (2010). "Extracting S&P500 and NASDAQ volatility: The Credit Crisis of 2007-2008." In O'Hagan, T. and West, M. (eds.), *The Oxford Handbook of Applied Bayesian Analysis*, chapter 13, 319–342. New York: Oxford University Press.

Mahieu, R. J. and Schotman, P. C. (1998). "An empirical application of stochastic volatility models." *Journal of Applied Econometrics*, 13(4): 333–360.

Mencía, J. and Sentana, E. (2009). "Multivariate location scale mixtures of normals and mean variance skewness portfolio allocation." *Journal of Econometrics*, 153(2): 105–121.

Nakajima, J. and Omori, Y. (2009). "Leverage, heavy-tails and correlated jumps in stochastic volatility models." *Computational Statistics and Data Analysis*, 53(6): 2335–2353.

Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007). "Stochastic Volatility with Leverage: Fast and Efficient Likelihood Inference." *Journal of Econometrics*, 140(2): 425–449.

Sandmann, G. and Koopman, S. J. (1998). "Estimation of stochastic volatility models via Monte Carlo maximum likelihood." *Journal of Econometrics*, 87(2): 271–301.

Shibata, M. and Watanabe, T. (2005). "Bayesian analysis of a Markov switching stochastic volatility model." *Journal of the Japan Statistical Society*, 35(2): 205–219.

So, M. K. P., Lam, K., and Li, W. K. (1998). "A Stochastic Volatility Model with Markov Switching." *Journal of Business & Economic Statistics*, 16(2): 244–253.

Storvik, G. (2002). "Particle filters for state-space models with the presence of unknown static parameters." *IEEE Transactions on Signal Processing*, 50(2): 281–289.

Taylor, S. J. (1982). "Financial Returns Modeled by the Product of two Stochastic Processes - a Study of Daily Sugar Prices." In Anderson, O. D. (ed.), *Time Series Analysis: Theory and Practice 1*, 203–226. Amsterdam: North-Holland.

— (1986). *Modelling Financial Time Series*. New York: John Willey & Sons, 1 edition.

— (1994). "Modeling Stochastic Volatility: a Review and Comparative Study." *Mathematical Finance*, 4(2): 183–204.

Titterington, D., Smith, A., and Makov, U. (1985). *Statistical Analysis of Finite Mixture Distributions*. Chichester: Willey.

Tokdar, S. T. (2006). "Posterior Consistency of Dirichlet Location-scale Mixture of Normals in Density Estimation and Regression." *Sankhya: Indian Statistical Institute*, 67(4): 90–110.

Virbickaite, A., Ausín, M. C., and Galeano, P. (2015). "Bayesian Inference Methods for Univariate and Multivariate Garch Models: a Survey." *Journal of Economic Surveys*, 29(1): 76–96.

— (2016). "A Bayesian non-parametric approach to asymmetric dynamic conditional correlation model with application to portfolio selection." *Computational Statistics and Data Analysis*, 100: 814–829.

Virbickaite, A., Lopes, H. F., Ausín, M. C., and Galeano, P. (2018). "Particle Learning for Bayesian Semi-Parametric Stochastic Volatility Model." *Econometric Reviews*, In press.

Vo, M. T. (2009). "Regime-switching stochastic volatility: Evidence from the crude oil market." *Energy Economics*, 31(5): 779–788.

Warty, S. P., Lopes, H. F., and Polson, N. G. (2017). "Sequential Bayesian learning for stochastic volatility with variance-gamma jumps in returns." *Appl .Stochastic Models Bus. Ind.*, Early View: 1–20.

Zaharieva, M. D., Trede, M., and Wilfling, B. (2017). "Bayesian semiparametric multivariate stochastic volatility with an application to international volatility co-movements." *CQE Working Papers*.

Table 1: Descriptive statistics for S&P500, Ford and natural gas data.

	S&P500	Ford	Gas
Mean	0.0223	0.0182	0.0104
Median	0.0690	-0.0778	0.0668
St.dev.	1.2752	2.8026	4.4554
Skewness	-0.2237	-0.0220	0.7370
Kurtosis	10.4789	15.8981	28.3024
T	4447	4329	4193

Table 2: Parameter estimation for SV-DPM and MSSV-DPM models for S&P500 data at time T.

	SV-DPM		MSSV-DPM	
	Mean	95% CI	Mean	95% CI
α	0.0019	(-0.002, 0.0061)	-	-
β	0.9887	(0.984, 0.9932)	0.9712	(0.9592, 0.9822)
τ^2	0.016	(0.0144, 0.0173)	0.0319	(0.0293, 0.0351)
γ_0	-	-	0.0074	(6e-04, 0.0136)
γ_1	-	-	0.0737	(6e-04, 0.5593)
p	-	-	0.9998	(0.9987, 1)
q	-	-	0.9971	(0.8617, 1)

Table 3: LPS and LPTS for SV-DPM and MSSV-DPM for S&P500 data ($T = 4447$).

	SV-DPM	MSSV-DPM	difference	$2\times$LPBF
LPS	2.1956	2.1991	-0.0036	-32.0184
$\text{LPTS}_{0.10}$	2.5953	2.5527	0.0426	37.8884
$\text{LPTS}_{0.05}$	2.8400	2.7826	0.0574	25.5258
$\text{LPTS}_{0.01}$	3.3949	3.2398	0.1550	13.7857

Table 4: Parameter estimation for SV-DPM and MSSV-DPM models for Ford data at time T.

	SV-DPM		MSSV-DPM	
	Mean	95% CI	Mean	95% CI
α	0.0707	(0.0641, 0.0774)	-	-
β	0.9514	(0.9461, 0.9563)	0.908	(0.9041, 0.9119)
τ^2	0.0432	(0.0339, 0.0487)	0.0395	(0.0374, 0.0417)
γ_0	-	-	0.1287	(0.1223, 0.1351)
γ_1	-	-	0.139	(0.113, 0.167)
p	-	-	0.9994	(0.9981, 0.9999)
q	-	-	0.9949	(0.9847, 0.9992)

Table 5: LPS and LPTS for SV-DPM and MSSV-DPM for Ford data ($T = 4329$).

	SV-DPM	MSSV-DPM	difference	$2\times$ LPBF
LPS	2.0783	2.0756	0.0027	23.3766
$\text{LPTS}_{0.10}$	2.8134	2.7260	0.0873	75.5843
$\text{LPTS}_{0.05}$	3.1713	3.0334	0.1379	59.6969
$\text{LPTS}_{0.01}$	4.2408	3.9699	0.2709	23.4545

Table 6: Parameter estimation for SV-DPM and MSSV-DPM models for Gas data at time T.

	SV-DPM		MSSV-DPM	
	Mean	95% CI	Mean	95% CI
α	0.2262	(0.2125, 0.2408)	-	-
β	0.8889	(0.8812, 0.8961)	0.8137	(0.807, 0.8206)
τ^2	0.1219	(0.1129, 0.1458)	0.1338	(0.1245, 0.1492)
γ_0	-	-	0.4161	(0.397, 0.4368)
γ_1	-	-	0.2758	(0.2255, 0.3205)
p	-	-	0.9951	(0.9907, 0.9976)
q	-	-	0.9802	(0.9655, 0.9904)

Table 7: LPS and LPTS for SV-DPM and MSSV-DPM for Gas data ($T = 4193$).

	SV-DPM	MSSV-DPM	difference	$2\times$LPBF
LPS	2.1592	2.1529	0.0063	52.8318
$\text{LPTS}_{0.10}$	2.8845	2.7875	0.0970	81.3442
$\text{LPTS}_{0.05}$	3.2812	3.1095	0.1717	71.9938
$\text{LPTS}_{0.01}$	4.5151	4.2300	0.2851	23.9085