UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 14-28
Statistics and Econometrics Series (19)
October 2014

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

Particle Learning for Bayesian Non-Parametric Markov Switching Stochastic Volatility Model

Audronė Virbickaitė[*1], Hedibert F. Lopes[2],
M. Concepción Ausín[3], Pedro Galeano[4]

Abstract

This paper designs a Particle Learning (PL) algorithm for estimation of Bayesian non-parametric Stochastic Volatility (SV) models for financial data. The performance of this particle method is then compared with the standard Markov Chain Monte Carlo (MCMC) methods for non-parametric SV models. PL performs as well as MCMC, and at the same time allows for on-line type inference. The posterior distributions are updated as new data is observed, which is prohibitively costly using MCMC. Further, a new non-parametric SV model is proposed that incorporates Markov switching jumps. The proposed model is estimated by using PL and tested on simulated data. Finally, the performance of the two non-parametric SV models, with and without Markov switching, is compared by using real financial time series. The results show that including a Markov switching specification provides higher predictive power in the tails of the distribution.

Keywords: Dirichlet Process Mixture; Markov Switching; MCMC; Particle Learning; Stochastic Volatility; Sequential Monte Carlo.

[*] Corresponding author
[1] Universidad Carlos III de Madrid (UC3M), c/ Madrid 126, Getafe (Madrid), Spain, 28903, audrone.virbickaite@uc3m.es
[2] Insper Institute of Education and Research, Sao Paulo, Brazil, 04546-042, hedibertfl@insper.edu.br
[3] UC3M, Getafe (Madrid), Spain, 28903, concepcion.ausin@uc3m.es
[4] UC3M, Getafe (Madrid), Spain, 28903, pedro.galeano@uc3m.es

Particle Learning for Bayesian Non-Parametric Markov Switching Stochastic Volatility Model

Audronė Virbickaitė [*†] Hedibert F. Lopes [‡] M. Concepción Ausín [§]
Pedro Galeano [¶]

Abstract

This paper designs a Particle Learning (PL) algorithm for estimation of Bayesian non-parametric Stochastic Volatility (SV) models for financial data. The performance of this particle method is then compared with the standard Markov Chain Monte Carlo (MCMC) methods for non-parametric SV models. PL performs as well as MCMC, and at the same time allows for on-line type inference. The posterior distributions are updated as new data is observed, which is prohibitively costly using MCMC. Further, a new non-parametric SV model is proposed that incorporates Markov switching jumps. The proposed model is estimated by using PL and tested on simulated data. Finally, the performance of the two non-parametric SV models, with and without Markov switching, is compared by using real financial time series. The results show that including a Markov switching specification provides higher predictive power in the tails of the distribution.

Keywords: Dirichlet Process Mixture; Markov Switching; MCMC; Particle Learning; Stochastic Volatility; Sequential Monte Carlo.

[*]Corresponding author
[†]Universidad Carlos III de Madrid (UC3M), c/ Madrid 126, Getafe (Madrid), Spain, 28903, audrone.virbickaite@uc3m.es
[‡]Insper Institute of Education and Research, Sao Paulo, Brazil, 04546-042, hedibertfl@insper.edu.br
[§]UC3M, Getafe (Madrid), Spain, 28903, concepcion.ausin@uc3m.es
[¶]UC3M, Getafe (Madrid), Spain, 28903, pedro.galeano@uc3m.es

1 Introduction

Understanding, modeling and predicting volatility of financial time series has been extensively researched for more than 30 years and interest in the subject is far from decreasing. Volatility prediction has a very wide range of applications in finance, for example, in portfolio optimization, risk management, asset allocation, asset pricing, etc. The two most popular approaches to model volatility are based on the Autoregressive Conditional Heteroscedasticity (ARCH) type, first introduced by Engle [1982], and the SV type, first introduced by Taylor [1982], models. There is evidence in the literature that SV models provide more flexibility than Generalized ARCH (GARCH, Bollerslev, 1986) specifications, see e.g. Broto and Ruiz [2004].

The SV model, as introduced by Taylor [1982], allows for time-varying volatility but it is unable to capture the usual heavy-tailed behavior of conditional distribution of the returns, since they are assumed to be Gaussian. One alternative is to abandon parametric assumptions for the returns altogether and consider a semi-parametric SV model, where the distribution of the returns is modeled non-parametrically, at the same time conserving the parametric discrete representation of the SV model.

Bayesian non-parametric approach in SV models is quite a new field of research, with growing popularity due to its flexibility and superior performance, see Jensen and Maheu [2010, 2012] and Delatola and Griffin [2011, 2013]. In these works it is assumed 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) and Bayesian estimation is performed using MCMC methods. The MCMC approach for SV models is the usual methodology since the seminal work by Jacquier et al. [1994], where Bayesian inference for standard SV models was firstly developed. For a survey on Bayesian estimation of time-varying volatility models see Virbickaitė et al. [2013]. However, MCMC methods in general are computationally demanding and inherently non-sequential [Lopes and Polson, 2010]. Alternatively, one can rely on Sequential Monte Carlo (SMC) methods, also known as particle filters, that allow for on-line type inference by updating the posterior distribution as the new data is observed.

Therefore, in this work we use SMC methods for Bayesian non-parametric SV models

which allows us to incorporate new information on-line, i.e. as it arrives. In particular, we make use of the PL approach, which is a particle based method, firstly introduced by Carvalho et al. [2010a]. Differently from other particle methods, it does not suffer from particle degeneracy. It also makes model comparison easy, since at each step we have the predictive likelihood as a by-product. PL methods have been shown to outperform the existing particle filtering alternatives and to be a competitor to MCMC, see Lopes et al. [2011].

In the first part of the paper we design a PL algorithm for a SV model with DPM innovations, referred to as a SV-DPM, similar to that of Delatola and Griffin [2011], and compare the performance of the algorithm to MCMC. We find that PL performs as well as MCMC, but, as commented above, the PL method provides the advantage of easily incorporating the information from the new observation, while MCMC requires to re-run the algorithm again.

In the second part of the paper we augment the SV-DPM model by incorporating Markov switching jumps, as seen in So et al. [1998] and Carvalho and Lopes [2007], resulting into a new MSSV-DPM model. We extend the previously developed PL algorithm to this new model and apply it on simulated data. Then, the performance of the new MSSV-DPM model is compared with the SV-DPM specification using real financial time series and we obtain that the new model provides better predictive power in the tails of the distribution.

The paper is structured as follows. Section 2 presents the linearized SV model with non-parametric errors and compares the estimation output for the SV-DPM model by using PL and MCMC. Then, Section 3 introduces a new MSSV-DPM model and designs a PL algorithm for inference and prediction. Section 4 compares the performance of the two non-parametric models by using real data. Finally, Section 5 concludes.

2 SV-DPM Model

In this section we briefly review a commonly used version of the standard stochastic volatility model with Normal errors. We then drop the normality hypothesis and introduce a novel particle learning scheme to perform sequential Bayesian learning in the class of SV model with Dirichlet Process Mixture models innovations (SV-DPM). We show, via synthetic examples, that our particle filter performs similarly to the standard MCMC scheme, with the advantage of producing online inference and, as a by product, online model comparison/selection statistics.

2.1 Normal errors

The standard SV model looks as follows:

$$y_t = \exp\{h_t/2\}\, v_t,$$
$$h_t = \alpha + \beta h_{t-1} + \tau \eta_t,$$

where we impose $|\beta| < 1$ for the stationarity of the volatilities; v_t and η_t are the error terms, such that $\eta_t \sim \mathcal{N}(0,1)$ and the distribution of the v_t with zero mean and unit variance takes many different forms in the existing literature: from a standard Normal, to heavy-tailed Student-t and others (see Kim et al., 1998, Chib et al., 2002, for example).

Kim et al. [1998] proposed linearlization 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:

$$r_t = h_t + \epsilon_t, \text{ where } \epsilon_t \sim \mathcal{F}, \tag{1}$$
$$h_t = \alpha + \beta h_{t-1} + \tau \eta_t, \text{ where } \eta_t \sim \mathcal{N}(0,1). \tag{2}$$

Observe that the distribution of ϵ_t is a $\log \chi_1^2$ if v_t is Normally distributed. Kim et al. [1998] and Omori et al. [2007] use carefully tuned finite mixtures of Normals to approximate the distribution of $\log \chi_1^2$ and use a data augmentation argument to propose 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 Frühwirth-

Schnatter [1994].

However, the recent literature is abundant in showing that the distribution of v_t has heavier tails than Normal, rendering the above approximations useless. Below we introduce the simple linearlized SV model with non-parametric errors to model the unknown return distribution.

2.2 Non-Normal errors

Delatola and Griffin [2011, 2013], for example, propose to approximate the log-square of the unknown return distribution $\epsilon_t \sim \mathcal{F}$ as an infinite mixture of Normals by relying on DPM models. The simple SV-DPM model presented in this section is of the same spirit as the model in Delatola and Griffin [2011].

Another important issue concerns the moments of the distribution of ϵ_t. Even though the original errors v_t are generated by a process with zero mean and unit variance, the resulting moments of ϵ_t can vary greatly, depending on the distribution of v_t. For example, if $v_t \sim \mathcal{N}(0,1)$, then $E[\epsilon_t] = -1.272$, $V[\epsilon_t] = 4.946$, $S[\epsilon_t] = -1.539$ and $K[\epsilon_t] = 7.015$, where $E[\cdot]$, $V[\cdot]$, $S[\cdot]$ and $K[\cdot]$ denote mean, variance, skewness and kurtosis, respectively. On the other hand, if $v_t \sim \mathcal{ST}(7)$, scaled in such a way that $E[v_t] = 0$ and $V[v_t] = 1$, then $E[\epsilon_t] = -1.428$, $V[\epsilon_t] = 5.218$, $S[\epsilon_t] = -1.404$ and $K[\epsilon_t] = 6.583$. However, Student-t and Normal are not the only possible distributions for the errors. There is an infinite number of possibilities for the distribution of the error term, whose moments are impossible to "map" backwards in order to recover the true error distribution. Actually, the main interest is usually not the distribution of the error term, but filtering and predicting the volatilities of the returns, which are highly sensitive to the choice of the error distribution.

The model specification in (1) and (2) is slightly different from the one in Delatola and Griffin [2011], since we do not sum the constant volatility parameter α into the mixture. We leave this constant separate since in Section 3 we augment the model by considering two different volatility levels, i.e. α_{s_t}, where $s_t \in \{0,1\}$.

Next, we do not specify a parametric model for the error density, but instead, we assume a Dirichlet Process Mixture prior, firstly introduced by Lo [1984]. DPM models

have been widely used for modeling time-varying volatilities, see Jensen and Maheu [2010, 2012, 2013], Delatola and Griffin [2011, 2013], Kalli et al. [2013], Ausín et al. [2014] and Virbickaitė et al. [2014]. This type of approach is known as time-invariant (independent) DPM.

As seen in Escobar and West [1995], the DPM model has the following density function:

$$f(\epsilon_t; G) = \int k(\epsilon_t; \theta_t) dG(\theta_t),$$

where k is some density kernel with parameters θ_t and the mixing distribution G has a DP prior, denoted here by $G \sim \mathcal{DP}(c, G_0(\theta; \varrho))$. Here the sub-index t in θ_t does not mean time-varying parameters, but refers to the fact that at each time t the observation ϵ_t comes from a different kernel density with some parameters θ_t, following the mixing distribution G. Parameter c is called the concentration parameter and $G_0(\theta; \varrho)$ is called the base distribution. The concentration parameter c can be interpreted as the prior belief about the number of clusters in the mixture. Small values of c assume *a priori* an infinite mixture model with a small number of components with large weights. On the contrary, large values of c assume *a priori* an infinite mixture model with 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. Considering a Gaussian kernel density, $\theta_t \sim \mathcal{N}(\mu_t, \sigma_t^2)$, 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.

Define $\Phi = (\alpha, \beta, \tau^2)$ as the set of parameters associated with the parametric part of the model, $\Omega = \{(\mu, \sigma^2)^{(j)}\}_{j=1}^{\infty}$ as a set of parameters associated with the distribution of the error term, and $\Theta = (\Phi, \Omega)$ as a complete set of all model parameters. Therefore,

the model in (1) and (2) can be rewritten as follows:

$$r_t|h_t, \Theta \sim \frac{c}{c+t-1}\mathcal{N}(r_t; \mu_0 + h_t, \sigma_0^2) + \frac{1}{c+t-1}\sum_{j=1}^{L_{t-1}^\star} n_{t-1,j}\mathcal{N}(r_t; \mu_j + h_t, \sigma_j^2), \quad (3)$$

$$h_t|h_{t-1}, \Theta \sim \mathcal{N}(h_t; \alpha + \beta h_{t-1}, \tau^2), \quad (4)$$

where $n_{t,j}$ is a number of observations assigned to j^{th} component at time t and $L_t^\star$ is a number of non-empty components in the mixture at time t. Given this missing information, the mixture becomes finite, where the maximum number of components theoretically is limited by the number of observations. In practice, data tends to cluster, meaning that some observations come from the same component, therefore $L_t^\star \leq t$.

2.3 MCMC for SV-DPM

The standard Bayesian estimation of SV models, parametric or non-parametric, relies on MCMC methods, which, however, can be costly, because they have to consider a sampler for latent volatilities.

Jensen and Maheu [2010] construct a MCMC scheme for their proposed SV-DPM model, where latent volatilities are sampled via random length block sampler, which helps to reduce correlation between draws. The authors found that the semi-parametric SV model is more robust to non-Normal data and provides better forecasts. In another paper, Jensen and Maheu [2012] consider an asymmetric SV-DPM model. The authors extend their previous semi-parametric sampler to a bivariate setting, where the innovations of the returns and volatilities are modeled jointly via infinite scale mixture of bivariate Normals.

Meanwhile, Delatola and Griffin [2011] use a linearized version of SV model. Conditional on knowing which mixture component the data belongs to, the linearized SV model is just a Normal Dynamic Linear Model (NDLM) and the latent volatilities are updated by FFBS (see the discussion at the end of Section 2.1). The remainder of the model parameters are sampled via an extension of Gibbs sampler, called hybrid Gibbs sampler. In their subsequent paper, Delatola and Griffin [2013] consider an asymmetric SV model. Same as before, they make use of the linearization and update the latent

log-volatilities via FFBS and the other parameters via Metropolis-Hastings. All above MCMC schemes are costly in the context of SV models for at least two reasons: (1) the MCMC sampler has to include a filter for latent volatilities, and (2) the sampler has to be re-run each time a new observation arrives.

2.4 PL for SV-DPM

In this section we present the algorithm to perform PL estimation for a SV model with non-parametric errors. 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 reviews of particle methods in general, see Lopes and Tsay [2011] and Lopes and Carvalho [2013]. The Appendix section at the end of this paper includes a brief description of the main idea behind PL. For a more detailed explanation of PL with illustrations refer to Carvalho et al. [2010a] and Lopes et al. [2011], among others.

The priors for model parameters are chosen to be conditionally conjugate: $h_0 \sim \mathcal{N}(c_0, C_0)$, $\sigma^2 \sim \mathcal{IG}(a_0/2, a_0\sigma_0^2/2)$, $\mu|\sigma^2 \sim \mathcal{N}(m_0, V_0\sigma^2)$, $\tau^2 \sim \mathcal{IG}(b_0/2, b_0\tau_0^2/2)$, $\beta|\tau^2 \sim \mathcal{TN}_{(-1,1)}(m_\beta, V_\beta\tau^2)$, $\alpha \sim \mathcal{N}(m_\alpha, V_\alpha)$. Here $\mathcal{TN}_{(a,b)}$ represents Normal distribution, truncated at a and b. 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_α and V_α are hyper-parameters. Then, a set of sufficient statistics S_t contains all updated hyper-parameters, necessary for the parameter simulation, as well as filtered state variables, which are of two kinds: the latent log-volatilities h_t and the indicator variable k_t, which tells us to which mixture component the error data point belongs to. For $t = 1 \ldots, T$ and for each particle (i) iterate through three steps:

1. **Resampling.**
   Resample old particles (states, sufficient statistics and parameters) with weights

$$w \propto \frac{1}{c+t-1} \sum_{j=0}^{L^\star_{t-1}} n_j f_N(r_t; \alpha + \beta h_{t-1} + \mu_j, \tau^2 + \sigma_j^2),$$

   proportional to the predictive density of the returns ($n_0 = c$). The components of $\Theta = (\alpha, \beta, \tau^2, \mu_1, \ldots, \mu_{L^\star_{t-1}}, \sigma_1^2, \ldots, \sigma^2_{L^\star_{t-1}})$ have been simulated at the end of the

previous period.

2. **Sampling.**

   (a) Sample new log-volatilities h_t from

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

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

   (b) Sample new indicators k_t from $\{1, \ldots, L^{\star}_{t-1} + 1\}$, with weights proportional to

$$\tilde{n}_j f_N(r_t; \alpha + \beta h_{t-1} + \mu_j, \tau^2 + \sigma_j^2), \;\; j = 1, \ldots, L^{\star}_{t-1} + 1,$$

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

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

   (c.2) Sample β from $\mathcal{TN}_{(-1,1)}(\beta; m_{\beta}^{\star}, V_{\beta}^{\star}\tau^2)$, where

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

   (c.3) Sample α from $\mathcal{N}(\alpha; m_{\alpha}^{\star}, V_{\alpha}^{\star})$, where

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

(c.4) Sample $\sigma^2_{k_t}$ from $\mathcal{IG}(\sigma^2_{k_t}; a_0^\star/2, a_0^\star\sigma^{2\star}/2)$, where

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

(c.5) Sample μ_{k_t} from $\mathcal{N}(\mu_{k_t}; m_0^\star, V_0^\star\sigma^2)$, where

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

2.5 Simulation exercise

We compare, based on simulated data, the posterior output for the SV-DPM model, estimated using MCMC and PL. A time series of length $T = 3000$ was simulated directly from the linearized model with $\alpha = 0$, $\beta = 0.98$ and $\tau^2 = 0.10$, where the log-square of the returns ϵ_t comes from the mixture of 7 Normals proposed by Kim et al. [1998] to approximate the $\log\chi^2_1$ distribution. Simulated returns can be recovered as $y_t = \exp\{r_t/2\}$. MCMC results are obtained via `Matlab` code of Delatola and Griffin [2011], which is available on Jim Griffin's website[1]. The MCMC algorithm was run for a total of 100k iterations, with the first 50k discarded as burn-in. The prior on the persistence parameter is $\beta \sim \mathcal{TN}_{(-1,1)}(0.95, 0.1)$ and the prior for the volatility of the volatility is $\tau^2 \sim \mathcal{IG}(8, 0.24)$. The prior for the mixture components is different than in this paper, since Delatola and Griffin [2011] use an alternative specification for the base distribution (see Griffin, 2010). Our PL algorithm, written in `R`, was run for a total of 300k particles. The hyper-parameters in the PL scheme are set as $c_0 = 0$, $C_0 = 0.1$, $m_\alpha = 0$, $V_\alpha = 0.01$, $m_\beta = 0.98$, $V_\beta = 0.1$, $b_0 = 6$, $b_0\tau_0^2 = 1.20$, $a_0 = 6$, $a_0\sigma_0^2 = 19$, $m_0 = -1.27$, $V_0 = 5$. The concentration parameter c in both codes is set to be equal to one. For volatility process and the parameters we report the median particle as an estimate together with 97.5% and 2.5% percentile particles for 95% credible intervals (CIs). For asymmetric distributions instead of quantiles we are using the corresponding HPD (Highest Posterior Density) intervals.

We have split the sample into three data sets of $T = 1000, 2000$ and 3000 observa-

[1] http://www.kent.ac.uk/smsas/personal/jeg28/index.htm

tions. In this way it is possible to see how PL is learning as compared to MCMC. The true advantage of the PL procedure becomes evident at the moment when the new observation arrives. In MCMC setting we need to re-run the entire chain all over again in order to incorporate this new information, meanwhile in PL we just add this new information to the existing output to obtain new updated parameters and states, which is just a matter of seconds. The CPU time for both estimation approaches is presented in Table 1.

Next, we compare the posterior output for both estimation methods graphically in Figures 1, 2 and 3 and Table 2. Figure 1 draws the estimated density at $T = 3000$ for the log of the squared returns for PL and MCMC, compared to the true one. Both estimations seem reasonable and very close to the true data generating density. Table 2 presents the estimated median parameter values with their corresponding 95% CIs or HPDs for the PL and MCMC estimation procedures. Estimation of the persistence parameter β is almost identical among both procedures. The posterior distribution of the volatility parameter τ^2 is always slightly more peaked in PL setting. In fact, as the sample size increases, the width of the HPD intervals for τ^2 for MCMC and PL decreases, and PL always presents around 20% thinner HPD intervals. This might be influenced by the fact the original model specifications are slightly different.

Figure 2 presents the posterior distributions for the log-volatilities at three different data points $T = 1000$, 2000 and 3000. The posterior distributions for $T = 2000$ and $T = 3000$ look identical among PL and MCMC. However, in order to obtain these distributions MCMC had to be re-run three times for three "different" data sets, meanwhile PL just incorporated new information sequentially and the posterior distribution of any $p(h_t|r^t)$ is readily available in the estimation output. Finally, Figure 3 draws the PL parameter estimation path with 95% confidence bounds, as compared with the true parameter values. As we can see, the parameter estimations become stable around the 1500^{th} observation. Also, there is no sign of particle degeneracy, which is a problem in other particle filtering methods, see Rios and Lopes [2013] for example. Therefore, PL can be seen as an efficient alternative to MCMC methods. Moreover, once the chain has been run, at the arrival of the new observation the posterior distributions can be updated at a very low computational cost.

In the next section we extend the non-parametric SV model to include Markov switching jumps and design a PL algorithm for inference and prediction.

3 MSSV-DPM Model

The simple SV model has some limitations such as it does not account for structural changes in the volatility process, which we have to take into consideration, otherwise the persistence parameter might be overestimated.

In this paper we consider a two-state Markov switching SV (MSSV) model, based on the model of So et al. [1998], where the log-volatility equation is of the following form:

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

where 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], we have to introduce the following reparametrization for α_{s_t} in order to avoid identification issues:

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

Here $\mathbf{1}\{s_t = 1\}$ is an indicator function that takes values equal to one if the volatility is in the high state ($s_t = 1$) and zero otherwise ($s_t = 0$). We also need to define the transition matrix between the states 0 and 1:

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

There are several papers that consider regime switching SV models in Bayesian context. Kalimipalli and Susmel [2004] have proposed a two-factor SV model with regime switching and estimated it using Gibbs sampler. They find that the high volatility per-

sistence 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. The regime switching SV performs better than the simple SV in sample, however, out of sample, it is only marginally better. Lopes and Carvalho [2007] extend 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 (APF) with the filter of Liu and West [2001] to estimate a SV model with Markov switching regimes. They found that in terms of prediction the Markov switching SV specification outperforms a simple SV model.

Here we extend the SV-DPM model in (3) and (4) to accommodate the above regime-shifting structure:

$$
\begin{aligned}
&r_t|h_t,\Theta \sim \frac{c}{c+t-1}\mathcal{N}(r_t;\mu_0+h_t,\sigma_0^2)+\frac{1}{c+t-1}\sum_{j=1}^{L^\star_{t-1}} n_{t-1,j}\mathcal{N}(r_t;\mu_j+h_t,\sigma_j^2),\\
&h_t|h_{t-1},\lambda_t,\Theta \sim \mathcal{N}(h_t;\gamma_0+\gamma_1\lambda_t+\beta h_{t-1},\tau^2),\\
&\lambda_t|\Theta \sim \mathcal{BER}\left((1-p)^{1-\lambda_{t-1}}q^{\lambda_{t-1}}\right),
\end{aligned}
$$

where $\mathcal{BER}(p)$ denotes a Bernoulli distribution with parameter p and λ_t is a Bernoulli distributed state variable.

3.1 PL for MSSV-DPM

We extend the previous PL algorithm of SV-DPM for MSSV-DPM, by incorporating the estimation of three extra parameters and filtering of one more state variable λ_t. The set of the parameters for the parametric part of the model is $\Phi = (\gamma_0, \gamma_1, \beta, \tau^2, p, q)$. Also, priors for the new parameters are: $\gamma_0 \sim \mathcal{N}(m_{\gamma_0}, V_{\gamma_0})$, $\gamma_1 \sim \mathcal{TN}_{(0,+\infty)}(m_{\gamma_1}, V_{\gamma_1})$, $p \sim \mathcal{B}(\alpha_p, \beta_p)$ and $q \sim \mathcal{B}(\alpha_q, \beta_q)$. For $t = 1 \ldots, T$ and for each particle (i) iterate through three steps:

1. **Resampling**.

   Resample with weights proportional to the predictive density of the returns:

$$w^{(i)} \propto \frac{1}{c+t-1}\sum_{j=1}^{L^\star_{t-1}} n_j f_N(r_t; \gamma_0 + \gamma_1\lambda_{t-1} + \beta h_{t-1} + \mu_j, \tau^2 + \sigma_j^2) + \frac{c}{c+t-1} f_N(r_t; \gamma_0 + \gamma_1\lambda_{t-1} + \beta h_{t-1} + \mu_0, \tau^2 + \sigma_0^2).$$

   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.

2. **Sampling.**

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

$$\lambda_t | \lambda_{t-1}, h_{t-1}, \Theta, r_t \sim \mathcal{BER}\left(\frac{z_2}{z_1 + z_2}\right),$$

   where

$$\begin{aligned} z_1 =& \frac{1}{c+t-1}\sum_{j=1}^{L^\star_{t-1}} n_j f_N(r_t; \gamma_0 + \beta h_{t-1} + \mu_j, \tau^2 + \sigma_j^2) + \\ & \frac{c}{c+t-1} f_N(r_t; \gamma_0 + \beta h_{t-1} + \mu_0, \tau^2 + \sigma_0^2) \times \Pr(\lambda_t = 0 | \lambda_{t-1}, \Theta), \\ z_2 =& \frac{1}{c+t-1}\sum_{j=1}^{L^\star_{t-1}} n_j f_N(r_t; \gamma_0 + \gamma_1\lambda_{t-1} + \beta h_{t-1} + \mu_j, \tau^2 + \sigma_j^2) + \\ & \frac{c}{c+t-1} f_N(r_t; \gamma_0 + \gamma_1\lambda_{t-1} + \beta h_{t-1} + \mu_0, \tau^2 + \sigma_0^2) \times \Pr(\lambda_t = 1 | \lambda_{t-1}, \Theta). \end{aligned}$$

   Then call $\tilde{\alpha} = \tilde{\gamma_0} + \tilde{\gamma_1}\lambda_t$.

   (b) Sample new log-volatilities h_t:

$$p(h_t | \tilde{h}_{t-1}, \tilde{\Theta}, \tilde{n}, \tilde{L}^\star_{t-1}, r_t) = \sum_{j=1}^{L^\star_{t-1}} \frac{n_j}{c+t-1}\mathcal{N}(h_t; m_{hj}, V_{hj}) + \frac{c}{c+t-1}\mathcal{N}(h_t; m_{h0}, V_{h0}),$$

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 $L_{t-1}^\star + 1$ components with the corresponding weights from the previous period.

(c) Sample new indicators k_t from $\{1, \ldots, L_{t-1}^\star + 1\}$, with weights proportional to:

$$\tilde{n}_j f_N(r_t; \alpha + \beta h_{t-1} + \mu_j, \tau^2 + \sigma_j^2), \ j = 1, \ldots, L_{t-1}^\star,$$

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

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

(c.1) Sample γ_0 from $\mathcal{N}(\gamma_0; m_{\gamma_0}^\star, V_{\gamma_0}^\star)$, where

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

(c.2) Sample γ_1 from $\mathcal{TN}_{(0,+\infty)}(\gamma_1; m_{\gamma_1}^\star, V_{\gamma_1}^\star)$, where

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

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

(c.3) Sample p from $\mathcal{B}(p; \alpha_p^\star, \beta_p^\star)$, where

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

(c.4) Sample q from $\mathcal{B}(q; \alpha_q^\star, \beta_q^\star)$, where

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

(c.5) 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.6) Sample β from $\mathcal{TN}_{(-1,1)}(\beta; m_\beta^\star, V_\beta^\star\tau^2)$, where

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

(c.7) Sample σ_{k_t} from $\mathcal{IG}(\sigma_{k_t}^2; a_0^\star/2, a_0^\star\sigma^{2\star}/2)$, 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 μ_{k_t} from $\mathcal{N}(\mu_{k_t}; m_0^\star, V_0^\star\sigma^2)$, where

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

3.2 Simulated Data

In order to test the proposed model we use a simulated data set with the following parameters: $\gamma_0 = -0.06$, $\gamma_1 = 0.08$, $\beta = 0.92$, $\tau^2 = 0.01$, $p = 0.995$, $q = 0.995$. The errors follow a standard Normal distribution $\varepsilon_t \sim \mathcal{N}(0, 1)$. The hyper-parameters are: $m_{\gamma_0} = \gamma_0$, $V_{\gamma_0} = \gamma_0^2$, $m_{\gamma_1} = \gamma_1$, $V_{\gamma_1} = \gamma_1^2$, $\alpha_p = 4$, $\beta_p = 1$, $\alpha_q = 4$, $\beta_q = 1$, $m_\beta = \beta$, $V_\beta = 0.1$, $b_0 = 3$, $b_0\tau_0^2 = 0.01$, $m_0 = -1.2704$, $V_0 = 5$, $a_0 = 5$ and $a_0\sigma_0^2 = 15$. We estimate this data with MSSV-DPM model using PL, number of particles $N = 300$k. The estimation results are presented in the Figures 4, 5 and 6.

Figure 4 top graph draws the simulated returns. The middle graph represents the true realization of the log-volatility (in black) and the mean estimated filtered log-volatility (in grey). The estimation of the latent log-volatility seems reasonable. The bottom graph of the same figure draws the mean probability of being in a state one ($s_t = 1$). 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 iden-

tify the regimes with the overall miss-classification rate equal to 13%. Figure 5 draws the true and estimated density for the log-squared returns, which is $\log \chi_1^2$. Finally, Figure 6 draws the sequential estimation of the model parameters and their 95% HPD intervals. Overall, the obtained estimation results seem quite reasonable and PL is able to correctly identify volatility regimes, filter log-volatilities, estimate the density of the errors and the parameters in an efficient sequential manner.

4 Real Data Application

In this section we present a real data application using return time series for various financial assets, in particular one index - S&P500, one company - Ford - and one commodity - natural gas. The S&P500 prices are from Jan 2nd 1997 till Sept 9th 2014, Ford from Jan 2nd 1997 till Sept 9th 2014 and Henry Hub natural gas spot prices (dollars per million btu) from Jan 5th 1997 till Sept 9th 2014. The summary of descriptive statistics can be seen in Table 3 and Figure 7.

Next, we estimate the data with two non-parametric models, SV-DPM and MSSV-DPM. The hyper-parameters for the priors are as follows: $c_0 = 0$, $C_0 = 0.1$, $m_\alpha = 0$, $V_\alpha = 0.001$, $m_\beta = 0.95$, $V_\beta = 0.1$, $b_0 = 8$, $b_0\tau_0^2 = 0.24$, $a_0 = 6$, $a_0\sigma_0^2 = 18$, $m_0 = -1.26$, $V_0 = 5$ for SV-DPM and $m_{\gamma_0} = -0.10$, $V_{\gamma_0} = 0.01$, $m_{\gamma_1} = 0.20$, $V_{\gamma_1} = 0.04$, $\alpha_p = 7$, $\beta_p = 1$, $\alpha_q = 7$, $\beta_q = 1$, $m_\beta = 0.95$, $V_\beta = 0.01$, $b_0 = 8$, $b_0\tau_0^2 = 0.456$, $m_0 = -1.26$, $V_0 = 5$, $a_0 = 6$, $a_0\sigma_0^2 = 18$ for MSSV-DPM. The codes were run for 500k particles each.

To compare the performance of the models, we use the average log-predictive score (LPS) and average log-predictive tail score (LPTS_α), which restrics attention to the events in the upper $100\alpha\%$ of the empirical distribution of the squared returns, as seen in Delatola and Griffin [2011]. 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:

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

where z_α is the upper 100α percentile of the empirical distribution of r_t. As Delatola and Griffin [2011] point out, the LPTS_α is not considered a proper scoring rule, however, it can be very useful for understanding how the model performs in the tails.

The log-predictive densities are very easy to obtain in SMC setting, since they are a by-product of the estimation procedure and, for each $t = 1, \ldots, T$, are calculated as

$$\log p(r_t|r^{t-1}) = \frac{1}{N} \sum_{i=1}^{N} p(r_t|(\Theta, h_t, k_t)^{(i)}). \tag{5}$$

Differently than in Delatola and Griffin [2011], there is no need to fix a certain $\hat{\Theta}$ for the calculation of the LPS and LPTS_α, since we can account for parameter and state uncertainty by using the approximation in (5).

Next, we present the estimation results for the S&P500 data set. Figures 8 and 9 present estimated predictive densities, filtered volatilities and volatility states and Table 4 presents the estimated parameters. Figure 8 shows the estimated densities for the error term as compared to the frequently used mixture of 7 Normals, as an approximation of $\log \chi_1^2$. SV-DPM and MSSV-DPM models estimates are very similar to each other and different from the 7N approximation. This shows that the assumption of Normality is very restrictive and in most cases incorrect. As we can see in Figure 9, the filtered volatility for both models is very similar (second and third graphs). Additionally, the MSSV-DPM model is able to identify some different volatility regimes, especially in the second half of the data series when the algorithm had time to learn (bottom graph). As for the parameter estimation in Table 4, the volatility persistence parameter tends to be larger for the SV-DPM model, as expected, see So et al. [1998] and Kalimipalli and Susmel [2004], among others.

Table 5 presents the LPS and LPTS_α for the S&P500 data. Same as in the paper by Delatola and Griffin [2011], where the authors compare parametric and non-parametric

SV models, the LPS are very similar thus making the models virtually indistinguishable. However, once we concentrate on the tails, the MSSV-DPM model tends to perform better, especially if we consider the very extreme events (the 99th percentile).

Similar results can be seen in the estimation of the other two data sets, see Tables 6, 7, 8 and 9 and Figures 10 and 11. For Ford and Natural gas data the SV-DPM model estimates larger persistence parameter, same as in the S&P500 data set. Also, the LPS for both models are very similar, but the differences appear when we consider only the tails of the distribution.

To conclude, it seems that the SV-DPM and MSSV-DPM models tend to perform similarly, if we consider the entire predictive distribution of the returns. However, the identification of different volatility regimes becomes important if we consider the tails of the distributions, where the MSSV-DPM model performs better. This is of major interest to the investors, who are usually more interested in the tails than the entire distribution.

5 Discussion

This paper designs a more efficient estimation procedure, based on SMC schemes, for a non-parametric SV-DPM model. We compare the performance of PL with the standard Bayesian estimation methods - MCMC. PL performs as well as MCMC, however, at a much lower computational cost whenever the new observation arrives. PL provides on-line type inference, which enables us to see the evolution of parameter learning and also provides the predictive likelihoods at each data point as a by-product. Next, the existing SV-DPM model is augmented with Markov switching jumps to capture different volatility regimes. We test the new model on simulated data and find that it is able to identify different volatility regimes. Finally, we present a real data application using three financial time series of the returns for one index - S&P500, one company - Ford, and one commodity - Natural gas. We find that the new MSSV-DPM model performs as well as the SV-DPM model if we consider the entire predictive distribution of the returns. However, the MSSV-DPM model outperforms the SV-DPM model if we consider only the tails of the distribution, especially, very rare events (the 99th percentile).

Appendix: A brief review of particle learning

Define S_t as an essential state vector to be tracked in time. S_t is sufficient for the computation of $p(r_{t+1}|S_t)$, $p(S_{t+1}|S_t, r_{t+1})$ and $p(\Theta|S_{t+1})$. Usually it contains the filtered states and the hyper-parameters for the distributions of the model parameters Θ. PL, differently than other particle methods, relies on a resample-propagate scheme, that can be understood by rewriting the Bayes theorem:

$$
\begin{aligned}
&p(S_t|r^{t+1}) \propto p(r_{t+1}|S_t)p(S_t|r^t) : \\
&\text{Resample } \; p(S_t|r^t) \; \text{ with weights } \; p(r_{t+1}|S_t), \\
&p(S_{t+1}|r^{t+1}) = \int p(S_{t+1}|S_t, r_{t+1}) dP(S_t|r^{t+1}) : \\
&\text{Propagate } \; S_{t+1} \; \text{ via some propagation rules.}
\end{aligned}
$$

Here $r^{t+1} = (r_1, \ldots, r_{t+1})$. At $t = 0$ initial values for parameters and states are simulated from their prior distributions: Φ_0 of dimension $K \times N$ (N is the number of particles and K is the number of model parameters), Ω_0 of dimension $2 \times N$ (at time $t = 1$ there is only one mixture component, having only two parameters) and h_0 of dimension $1 \times N$. Also, an essential state vector S_0 is constructed, containing all the hyper-parameters for the parameters of the model and mixture components, the volatility states and other information about the mixture. This vector is of dimension $Z_t \times N$, where Z_t changes in time depending on the number of the mixture components. Then, PL iterates through three steps, for each particle (i), for $i = 1, \ldots, N$:

1. **Resample** the particles with weights proportional to the posterior predictive density $w^{(i)} \propto p(r_{t+1}|S_t^{(i)})$ to obtain resampled particles $\tilde{S}_t^{(i)}$. In other words, we obtain a new essential state vector $\tilde{S}_t$ by sampling from the existing essential state vector S_t with weights that give more importance to the particles that produce higher likelihood with respect to the new data point.

2. **Propagate** the particles $S_{t+1}^{(i)} \sim p(S_{t+1}|\tilde{S}_t^{(i)}, r_{t+1})$. In this step we update all the elements of the essential state vector through some propagation rules.

3. **Learn** about the parameters on-line or off-line by approximating $p(\Theta|r^{t+1})$ as fol-

lows:

$$p(\Theta|r^{t+1}) \approx 1/N \sum_{i=1}^{N} p(\Theta|S_{t+1}).$$

In this step, once the elements of the essential state vector have been propagated, we use those updated hyper-parameters to sample from the posterior distributions of the parameters, obtaining new samples for the parameters Θ. In some cases it is possible to integrate out the parameter uncertainty in resample step. Then, the predictive density depends only on the essential state vector $p(r_{t+1}|S_t^{(i)})$. However, in many other cases it is not possible to integrate out the parameter uncertainty analytically. Then, in order to calculate the predictive density in the resample step, we use the sampled parameters, obtained from the hyper-parameters in the essential state vector: $p(r_{t+1}|\Theta_t^{(i)}, S_t^{(i)})$.

Carvalho et al. [2010b] presented a detailed explanation of PL methods for general mixtures, including DPM models. As before, $n_{t,j}$ is a number of observations assigned to the j^{th} mixture component at time t and k_t is an allocation variable that indicates which mixture component the observation belongs to. We can augment the essential state vector S_t by including $n_{t,j}$ and k_t. Then density estimation by using a infinite location-scale mixture of Normals via PL can be carried out by iterating through the following two steps, for each particle i:

1. **Resample** with weights proportional to the predictive density $w^{(i)} \propto p(r_{t+1}|S_t^{(i)})$ to obtain resampled particles $\tilde{S}_t^{(i)}$;

2. **Propagate** allocation variable $k_{t+1}^{(i)} \sim p(k_{t+1}|\tilde{S}_t^{(i)}, y_{t+1})$, and the rest of the sufficient statistics $S_{t+1}^{(i)} = p(S_{t+1}|\tilde{S}_t^{(i)}, k_{t+1}, y_{t+1})$, including $n_{t+1}^{(i)}$.

The third step, parameter learning, can be performed off-line since the parameter uncertainty, as mentioned before, can be integrated out. In various simulation studies, presented in the papers above, the authors show that PL outperforms other particle filtering approaches, and is a cost-efficient alternative to MCMC methods.

Acknowledgments

Virbickaitė, A. and Ausín, C.M. are grateful for the financial support from MEC grant ECO2011-25706. Galeano, P. acknowledges financial support from MEC grant ECO2012-38442.

References

M. Concepción Ausín, Pedro Galeano, and Pulak Ghosh. 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, January 2014. ISSN 03772217. doi: 10.1016/j.ejor.2013.07.008.

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

Carmen Broto and Esther Ruiz. Estimation Methods for Stochastic Volatility Models: a Survey. *Journal of Economic Surveys*, 18(5):613–650, 2004. doi: 10.1111/j.1467-6419.2004.00232.x.

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

Carlos M. Carvalho and Hedibert F. Lopes. Simulation-Based Sequential Analysis of Markov Switching Stochastic Volatility Models. *Computational Statistics & Data Analysis*, 51(9):4526–4542, May 2007. ISSN 01679473. doi: 10.1016/j.csda.2006.07.019.

Carlos M. Carvalho, Michael S. Johannes, Hedibert F. Lopes, and Nicholas G. Polson. Particle Learning and Smoothing. *Statistical Science*, 25(1):88–106, February 2010a. ISSN 0883-4237. doi: 10.1214/10-STS325.

Carlos M. Carvalho, Hedibert F. Lopes, Nicholas G. Polson, and Matt A. Taddy. Particle Learning for General Mixtures. *Bayesian Analysis*, 5(4):709–740, December 2010b. ISSN 1936-0975. doi: 10.1214/10-BA525.

Siddhartha Chib, Federico Nardari, and Neil Shephard. Markov Chain Monte Carlo Methods for Stochastic Volatility Models. *Journal of Econometrics*, 108(2):281–316, June 2002. ISSN 03044076. doi: 10.1016/S0304-4076(01)00137-3.

Eleni-Ioanna Delatola and Jim E. Griffin. Bayesian Nonparametric Modelling of the Return Distribution with Stochastic Volatility. *Bayesian Analysis*, 6(4):901–926, 2011. ISSN 1931-6690. doi: 10.1214/11-BA632.

Eleni-Ioanna Delatola and Jim E. Griffin. A Bayesian Semiparametric Model for Volatility with a Leverage Effect. *Computational Statistics & Data Analysis*, 60:97–110, 2013. doi: 10.1016/j.csda.2012.10.023.

Robert F. Engle. Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4):987–1008, 1982. doi: 10.2307/1912773.

Michael D Escobar and Mike West. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90(430):577–588, 1995.

Thomas S Ferguson. Bayesian Density Estimation by Mixtures of Normal Distributions. In H. Rizvi and J. Rustagi, editors, *Recent Advances in Statistics*, volume 24, pages 287–302. Academic Press, New York, 1983.

S. Frühwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15:183–802, 1994.

J. E. Griffin. Default priors for density estimation with mixture models. *Bayesian Analysis*, 5(1):45–64, March 2010. ISSN 1936-0975. doi: 10.1214/10-BA502.

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

Mark J. Jensen and John M. Maheu. Bayesian Semiparametric Stochastic Volatility Modeling. *Journal of Econometrics*, 157(2):306–316, 2010. ISSN 0304-4076. doi: 10.1016/j.jeconom.2010.01.014.

Mark J. Jensen and John M. Maheu. Estimating a Semiparametric Asymmetric Stochastic Volatility Model with a Dirichlet Process Mixture. *Working Paper, University of Toronto*, 453, 2012.

Mark J. Jensen and John M. Maheu. Bayesian Semiparametric Multivariate GARCH Modeling. *Journal of Econometrics*, 176(1):3–17, September 2013. ISSN 03044076. doi: 10.1016/j.jeconom.2013.03.009.

Madhu Kalimipalli and Raul Susmel. Regime-Switching Stochastic Volatility and Short-Term Interest Rates. *Journal of Empirical Finance*, 11(3):309–329, June 2004. ISSN 09275398. doi: 10.1016/j.jempfin.2003.10.001.

Maria Kalli, Stephen G. Walker, and Paul Damien. Modeling the Conditional Distribution of Daily Stock Index Returns: An Alternative Bayesian Semiparametric Model. *Journal of Business & Economic Statistics*, 31(4):371–383, October 2013. ISSN 0735-0015. doi: 10.1080/07350015.2013.794142.

Sangjoon Kim, Neil Shephard, and Siddhartha Chib. Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *Review of Economic Studies*, 65:361–393, 1998.

Jane Liu and Mike West. Combined Parameter and State Estimation in Simulation-Based Filtering. In Arnaud Doucet, Nando Freitas, and Neil Gordon, editors, *Sequential Monte Carlo Methods in Practice*, pages 197–223. Springer New York, 2001. ISBN 978-1-4419-2887-0. doi: 10.1007/978-1-4757-3437-9_10.

Albert Y. Lo. On a Class of Bayesian Nonparametric Estimates: I. Density Estimates. *The Annals of Statistics*, 12(1):351–357, 1984. doi: 10.1214/aos/1176346412.

Hedibert F. Lopes and Carlos M. Carvalho. Online Bayesian Learning in Dynamic Models: an Illustrative Introduction to Particle Methods. In Paul Damien, Petros Dellaportas, Nicholas G. Polson, and David A. Stephens, editors, *Bayesian Theory and Applications*, chapter 11, pages 203–228. Oxford University Press, Oxford, 1 edition, 2013. ISBN 978-0-19-969560-7.

Hedibert F. Lopes and Ruey S. Tsay. Particle Filters and Bayesian Inference in Financial Econometrics. *Journal of Forecasting*, 30(July):168–209, 2011. doi: 10.1002/for.1195.

Hedibert F. Lopes, Carlos M. Carvalho, Michael S. Johannes, and Nicholas G. Polson. Particle Learning for Sequential Bayesian Computation. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West, editors, *Bayesian Statistics 9*, volume 9, pages 317–360. Oxford University Press, Oxford, 2011.

Hedibert Freitas Lopes and Carlos Marinho Carvalho. Factor Stochastic Volatility with Time Varying Loadings and Markov Switching Regimes. *Journal of Statistical Planning and Inference*, 137(10):3082–3091, October 2007. ISSN 03783758. doi: 10.1016/j.jspi.2006.06.047.

Hedibert Freitas Lopes and Nicholas G Polson. Extracting S&P500 and NASDAQ volatility: The Credit Crisis of 2007-2008. In Tony O'Hagan and Mike West, editors, *The Oxford Handbook of Applied Bayesian Analysis*, chapter 13, pages 319–342. Oxford University Press, New York, 2010. ISBN The Oxford Handbook of Applied Bayesian Analysis.

Yasuhiro Omori, Siddhartha Chib, Neil Shephard, and Jouchi Nakajima. Stochastic Volatility with Leverage: Fast and Efficient Likelihood Inference. *Journal of Econometrics*, 140(2):425–449, October 2007. ISSN 03044076. doi: 10.1016/j.jeconom.2006.07.008.

Maria Paula Rios and Hedibert F. Lopes. The Extended Liu and West Filter: Parameter Learning in Markov Switching Stochastic Volatility Models. In Yong Zeng and Shu Wu, editors, *State-Space Models: Statistics and Econometrics for Finance*, chapter 2, pages 23–61. Springer New York, 2013. ISBN 978-1-4614-7788-4. doi: 10.1007/978-1-4614-7789-1_2.

Mike EC P So, Kin Lam, and Wai Keung Li. A stochastic volatility model with markov switching. *Journal of Business & Economic Statistics*, 16(2):244–253, 1998.

Stephen J. Taylor. Financial Returns Modeled by the Product of two Stochastic Processes - a Study of Daily Sugar Prices 1961 - 79. In O. D. Anderson, editor, *Time Series Analysis: Theory and Practice 1*. North-Holland, 1982.

Audronė Virbickaitė, M. Concepción Ausín, and Pedro Galeano. Bayesian Inference Methods for Univariate and Multivariate Garch Models: a Survey. *Journal of Economic Surveys*, In Press, September 2013. doi: 10.1111/joes.12046.

Audronė Virbickaitė, M. Concepción Ausín, and Pedro Galeano. A Bayesian Non-Parametric Approach to Asymmetric Dynamic Conditional Correlation Model with Application to Portfolio Selection. *Revised and submitted to CSDA*, arXiv:1301.5129v2, 2014.

Tables and Figures

Table 1: CPU time in seconds for MCMC and PL.

T	MCMC (50k+50k)	PL (300k particles)
1000	23356	
2000	51796	
3000	80401	56999

Table 2: Parameter estimates and their corresponding 95% CIs for PL and MCMC for $T = 1000, 2000, 3000$.

		MCMC		PL	
	T	Mean	95%CI	Mean	95%CI
	1000	0.9616	(0.9368, 0.9826)	0.9671	(0.9464, 0.9841)
$\beta = 0.98$	2000	0.9701	(0.9552, 0.9833)	0.9753	(0.9627, 0.9854)
	3000	0.9809	(0.9721, 0.9889)	0.9843	(0.9768, 0.9906)
	1000	0.1303	(0.0856, 0.1887)	0.1060	(0.0705, 0.1457)
$\tau^2 = 0.10$	2000	0.1140	(0.0805, 0.1538)	0.0837	(0.0590, 0.1147)
	3000	0.1021	(0.0774, 0.1295)	0.0727	(0.0610, 0.1010)

Table 3: Descriptive statistics for S&P500, Ford and 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 4: 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	0.0144	(0.0098, 0.0190)	-	-
β	0.9792	(0.9747, 0.9837)	0.9474	(0.9383, 0.9550)
τ^2	0.0187	(0.0172, 0.0202)	0.0255	(0.0239, 0.0276)
γ_0	-	-	0.0052	(-0.0010, 0.0131)
γ_1	-	-	0.1279	(0.1069, 0.1497)
p	-	-	0.9943	(0.9898, 0.9973)
q	-	-	0.9585	(0.9352, 0.9782)

Table 5: LPS and LPTS_α for SV-DPM and MSSV-DPM for S&P500 data.

	SV-DPM	MSSV-DPM	difference
LPS	2.1907	2.2029	-0.0122
$\text{LPTS}_{0.10}$	2.6444	2.6610	-0.0166
$\text{LPTS}_{0.05}$	2.9369	2.9282	0.0087
$\text{LPTS}_{0.01}$	3.6168	3.5068	0.1100

Table 6: 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	0.0198	(-0.0238, 0.0264)	-	-
β	0.9738	(0.9678, 0.9791)	0.9389	(0.9287, 0.9481)
τ^2	0.0274	(0.0171, 0.0366)	0.0474	(0.0442, 0.0509)
γ_0	-	-	0.0013	(-0.0088, 0.0089)
γ_1	-	-	0.0875	(0.0731, 0.1087)
p	-	-	0.9944	(0.9909, 0.9974)
q	-	-	0.9854	(0.9755, 0.9929)

Table 7: LPS and LPTS_α for SV-DPM and MSSV-DPM for Ford data.

	SV-DPM	MSSV-DPM	difference
LPS	2.0718	2.0851	-0.0133
$\text{LPTS}_{0.10}$	2.7639	2.7687	-0.0048
$\text{LPTS}_{0.05}$	3.1086	3.0956	0.0130
$\text{LPTS}_{0.01}$	4.1864	4.1007	0.0857

Table 8: 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	-0.0430	(-0.0481, -0.0342)	-	-
β	0.9823	(0.9755, 0.9845)	0.9458	(0.9414, 0.9502)
τ^2	0.0278	(0.0253, 0.0441)	0.0374	(0.0340, 0.0404)
γ_0	-	-	-0.1103	(-0.1173, -0.1030)
γ_1	-	-	0.3961	(0.3574, 0.4314)
p	-	-	0.9682	(0.9596, 0.9761)
q	-	-	0.6458	(0.5762, 0.7125)

Table 9: LPS and LPTS_α for SV-DPM and MSSV-DPM for Gas data.

	SV-DPM	MSSV-DPM	difference
LPS	2.1431	2.1485	-0.0054
$\text{LPTS}_{0.10}$	2.7865	2.8091	-0.0226
$\text{LPTS}_{0.05}$	3.1608	3.1560	0.0049
$\text{LPTS}_{0.01}$	4.5336	4.2702	0.2634

Figure 1: Density of a mixture of 7 Normals and the density of the simulated data compared to the predictive density for $\epsilon_t = \log \varepsilon_t^2$, estimated by PL and MCMC for $T = 3000$.

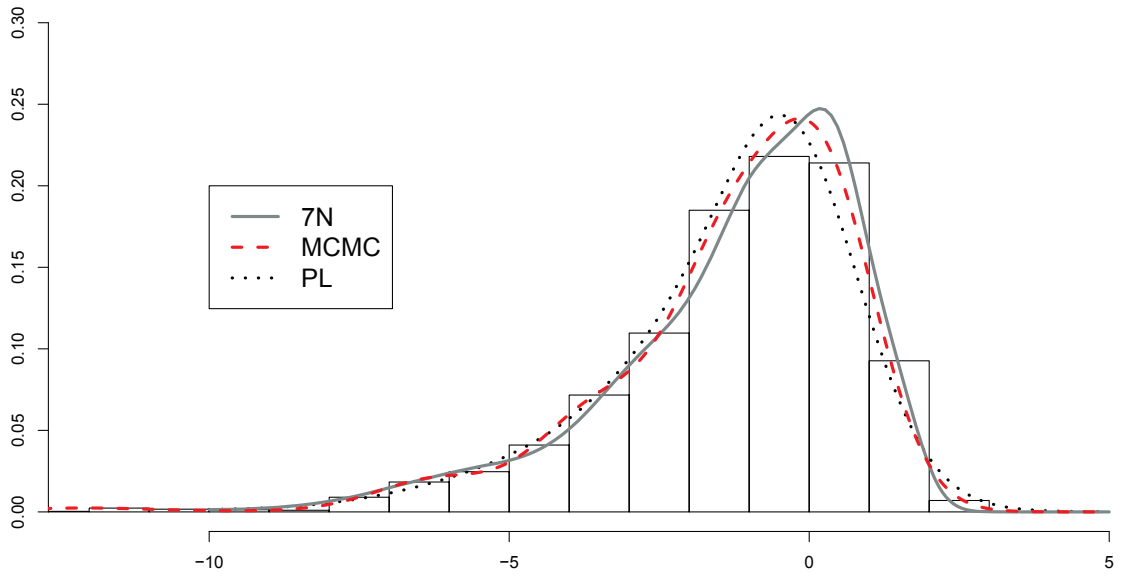

Figure 2: Posterior distributions of the log-volatilities for MCMC and PL for $T = 1000, 2000$ and 3000.

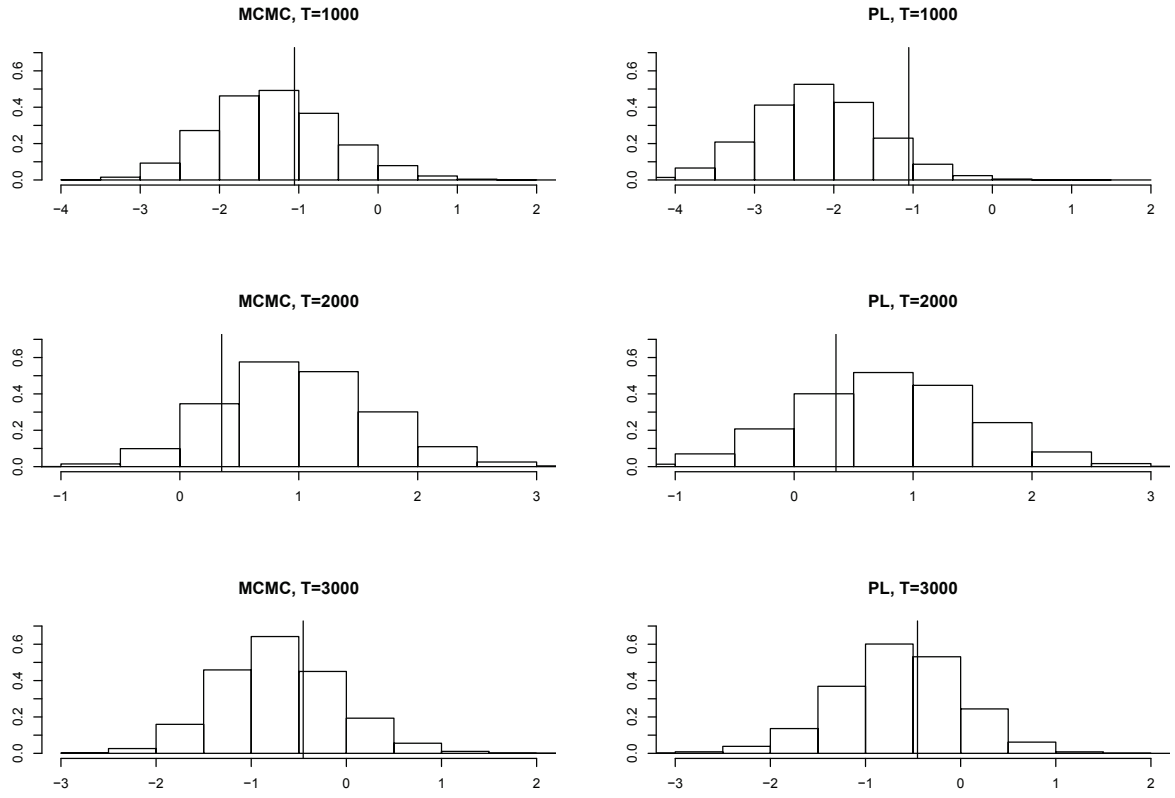

Figure 3: PL parameter estimates with 95% CI for one run of 300k particles, compared to the true parameter values.

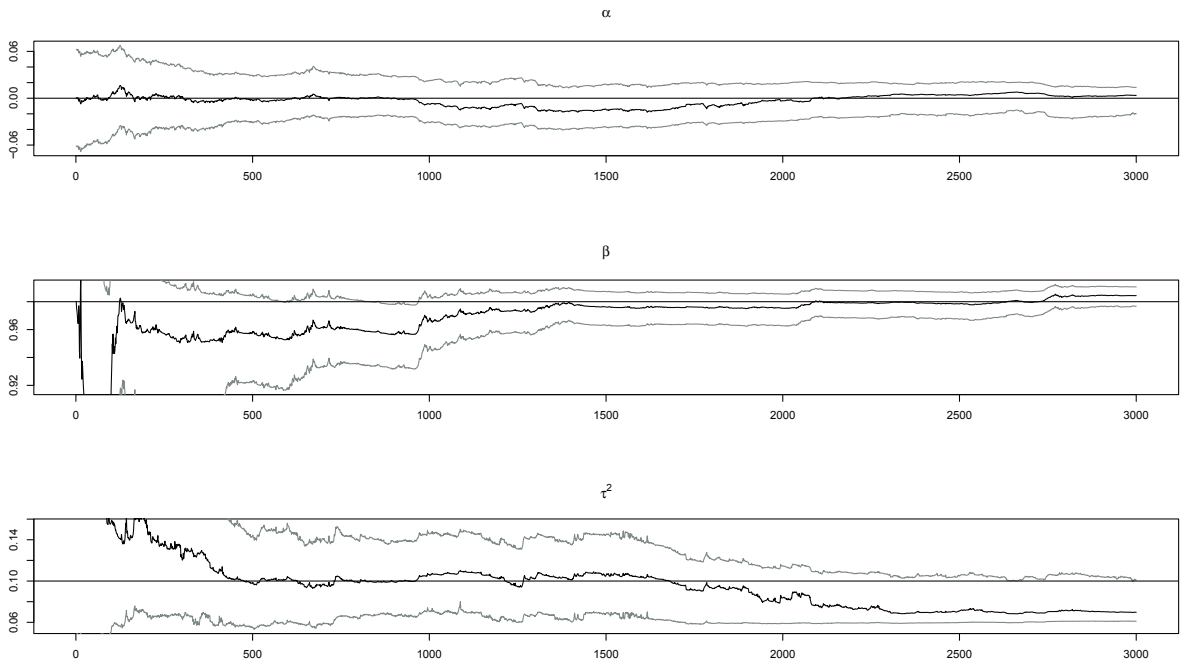

Figure 4: Simulated data: daily returns (top graph), true and estimated log-volatilities (middle graph) and true and estimated regimes (bottom graph).

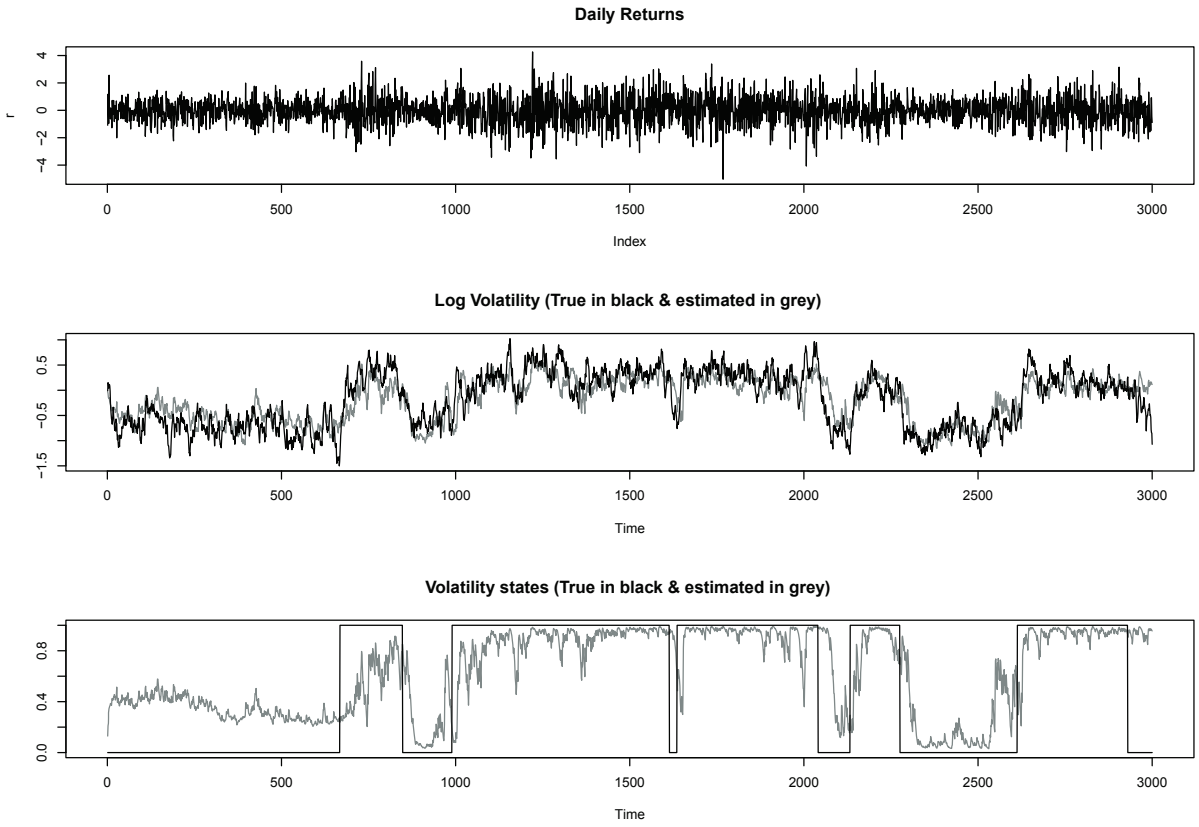

Figure 5: Simulated data: true and estimated density for log-squared return distribution.

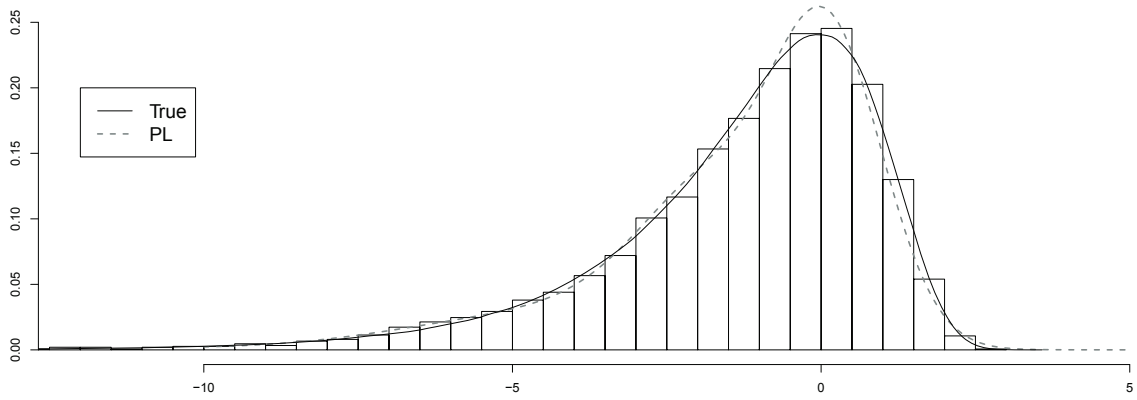

Figure 6: Simulated data: true and estimated parameters with 95% HPD intervals.

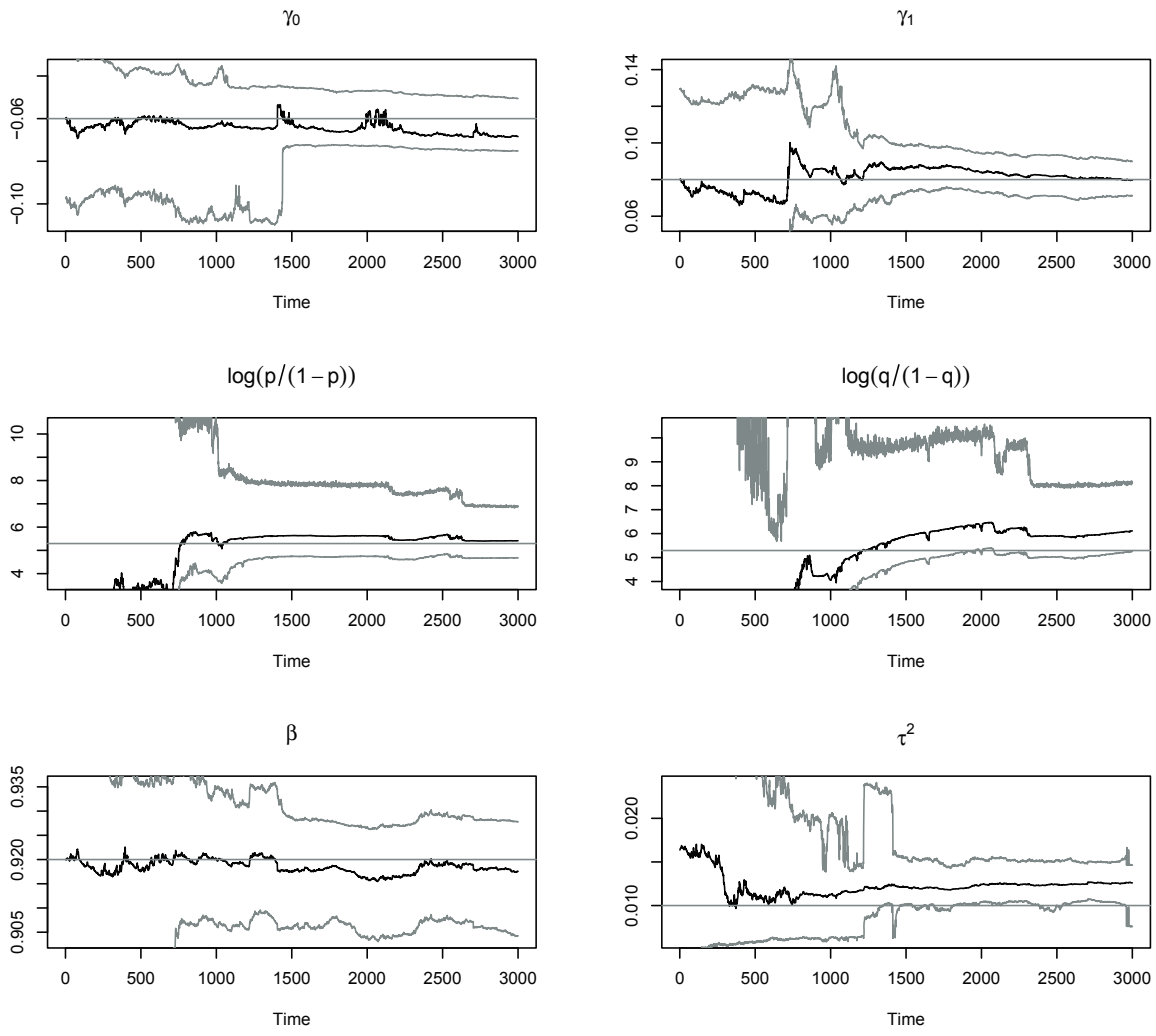

Figure 7: Daily log-returns (in %) and corresponding histograms for S&P500, Ford and Natural gas data.

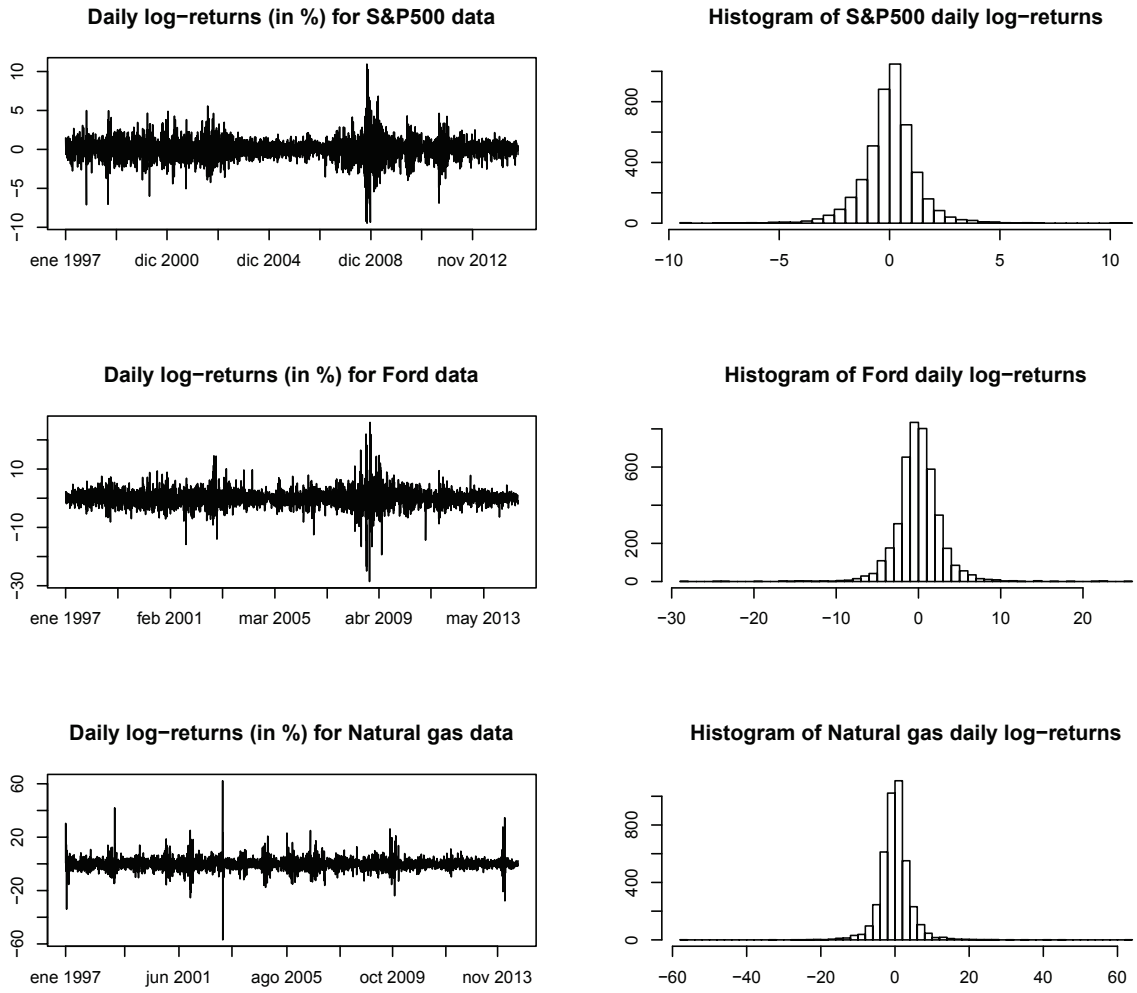

Figure 8: Estimated densities for the log-squared error term for SV-DPM and MSSV-DPM models.

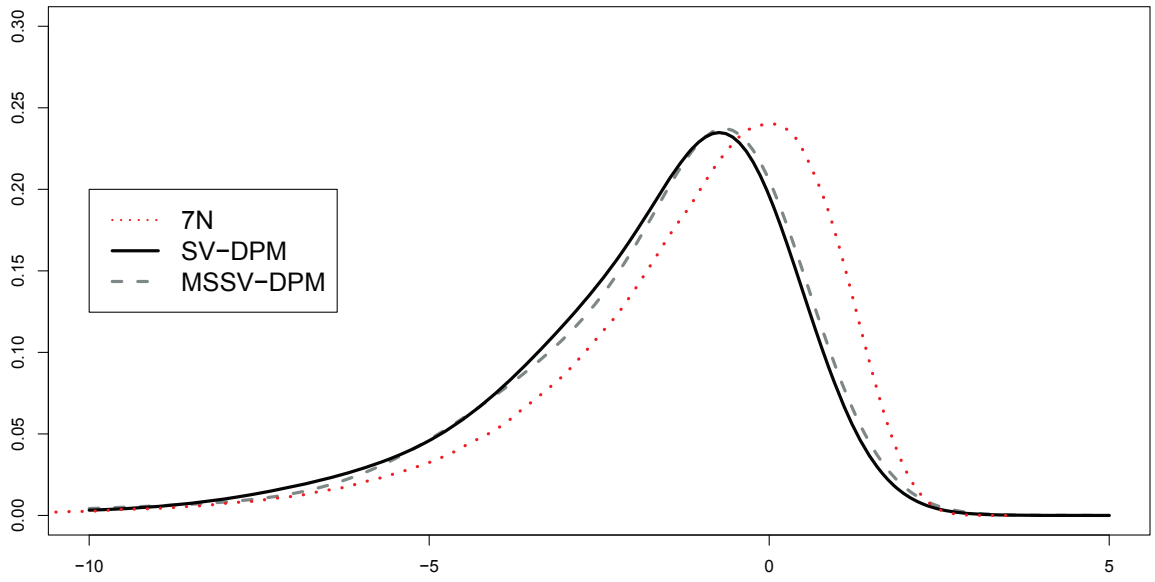

Figure 9: Filtered volatilities and volatility states for S&P500 data for SV-DPM and MSSV-DPM models.

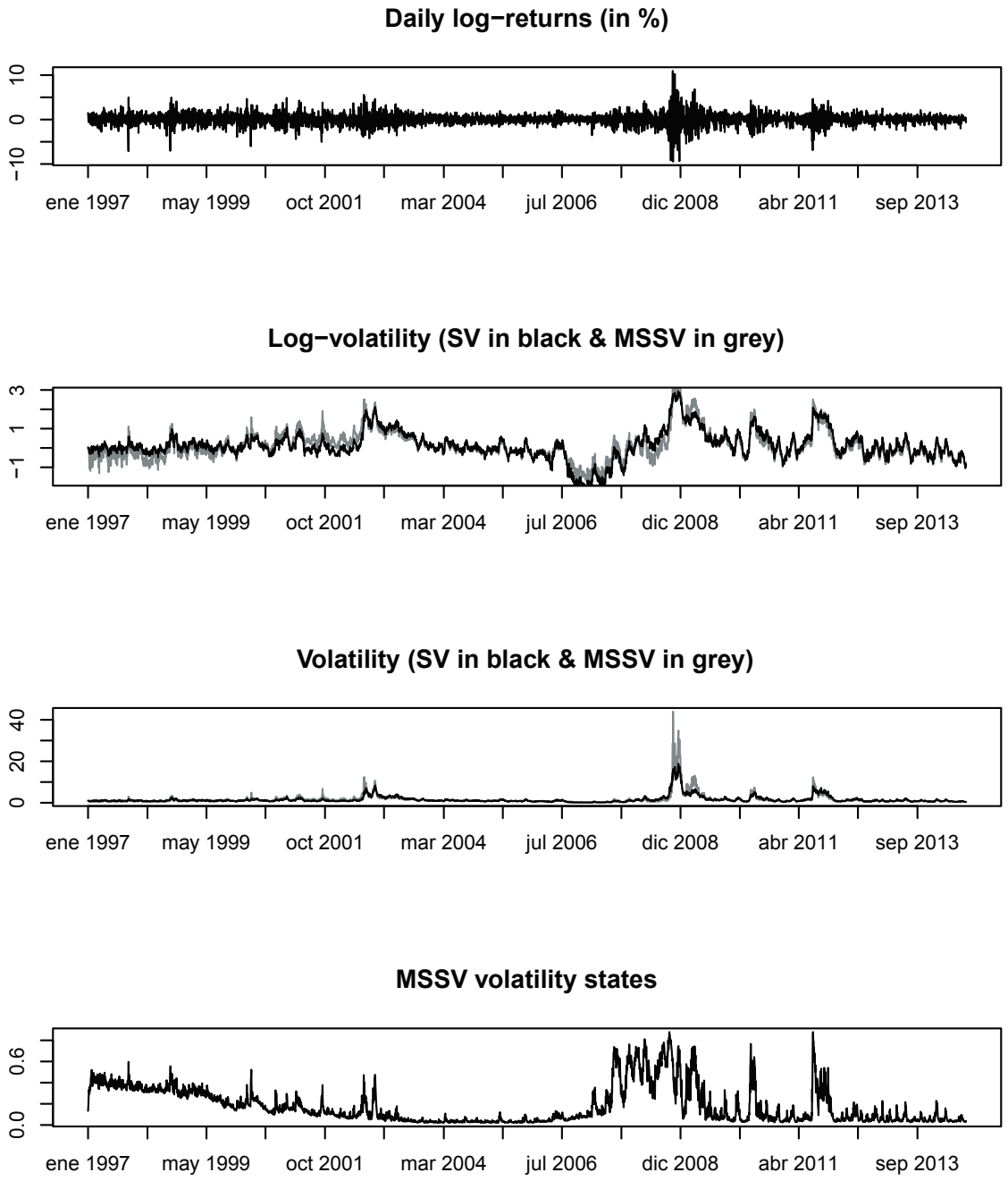

Figure 10: Filtered volatilities and volatility states for Ford data for SV-DPM and MSSV-DPM models.

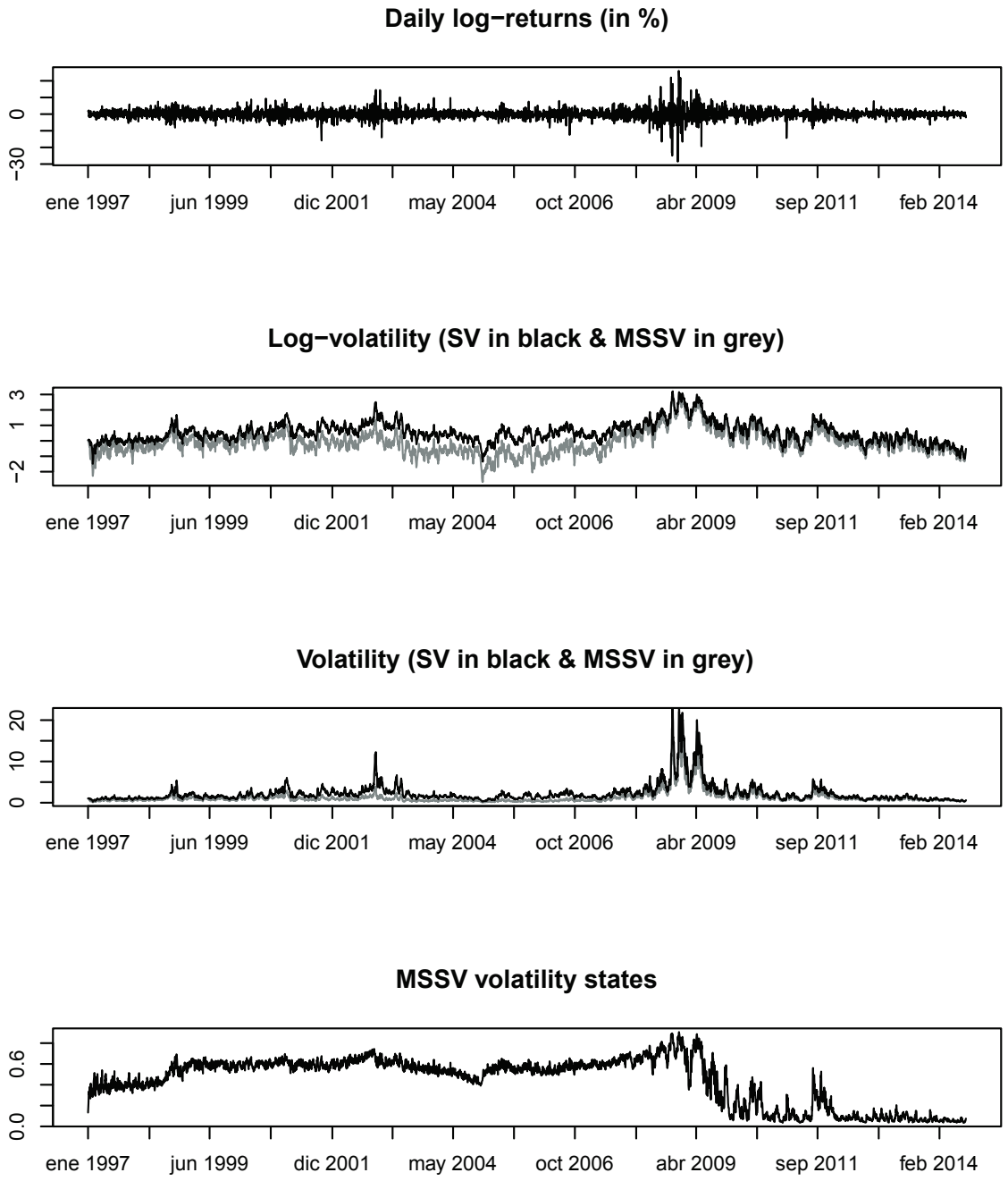

Figure 11: Filtered volatilities and volatility states for Gas data for SV-DPM and MSSV-DPM models.