Universitat de les Illes Balears

DEA WP no. 88
Working Paper Series

Particle Learning for Bayesian Semi-Parametric 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

María Concepción Ausín
Universidad Carlos III de Madrid
E-mail: concepcion.ausin@uc3m.es

Pedro Galeano
Universidad Carlos III de Madrid
E-mail: pedro.galeano@uc3m.es

May 2018

Departament d'Economia Aplicada

Particle Learning for Bayesian Semi-Parametric Stochastic Volatility Model

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

Abstract

This paper designs a Sequential Monte Carlo (SMC) algorithm for estimation of Bayesian semi-parametric Stochastic Volatility model for financial data. In particular, it makes use of one of the most recent particle filters called Particle Learning (PL). SMC methods are especially well suited for state-space models and can be seen as a cost-efficient alternative to MCMC, since they allow for online type inference. The posterior distributions are updated as new data is observed, which is prohibitively costly using MCMC. Also, PL allows for consistent online model comparison using sequential predictive log Bayes factors. A simulated data is used in order to compare the posterior outputs for the PL and MCMC schemes, which are shown to be almost identical. Finally, a short real data application is included.

Keywords: Bayes factor; Dirichlet Process Mixture; MCMC; Sequential Monte Carlo.
JEL classification: C58, C11, C14

[*]Corresponding author
[†]Universitat de les Illes Balears, Departamento de Economia Aplicada, Crta Valldemossa, km 7,5 07122 Palma de Mallorca (Spain), audrone.virbickaite@uib.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 stylized features of financial returns has been extensively researched for more than 30 years and interest in the subject is far from decreasing. Meanwhile mean-variance framework has been of major interest, it is justifiable only for Normally distributed returns. There is overwhelming evidence in the literature that the distribution of the financial returns is far from Normal, in the sense that it exhibits fat tails and occasional asymmetry. Therefore, apart from the mean and variance modeling one also has to consider departures from Normality by allowing for skewness and excess kurtosis via more flexible distributional assumptions for the innovations of the returns.

Modeling conditional mean of the returns is a very challenging task, since they are always very close to zero and exhibit very low levels of autocorrelation. The volatility of the returns, on the other hand, usually exhibits slow decaying autocorrelation function, i.e. high persistence, which can be modeled via auto-regressive process. The two most popular approaches to model volatility are based on the Autoregressive Conditional Heteroscedasticity (ARCH) type models, first introduced by Engle (1982), and the Stochastic volatility (SV) type models, first introduced by Taylor (1982). 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). Therefore, in this work we consider the SV model for the volatilities.

As for the distribution of the error term of the returns, the Normal distribution was considered by Taylor (1986, 1994), Jacquier et al. (1994), Kim et al. (1998), among many others. However, as mentioned above, financial returns depart from Normality since they exhibit fat fails and occasional asymmetry. There has been a multitude of papers considering all kinds of non-Normal distributions. For example, the Student-t distribution was employed by Harvey et al. (1994), Gallant et al. (1997), Sandmann and Koopman (1998), Chib et al. (2002), 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 (2005), among many others.

Another alternative is to abandon parametric assumptions for the distribution of the error term of the returns altogether and consider a semi-parametric SV model, where the distribution of the returns is modeled non-parametrically, and, at the same time, the parametric discrete representation of the SV model is conserved. The Bayesian non-parametric approach for SV models is quite a new field of research, with growing popularity due to its flexibility and superior performance, see Jensen (2004), Jensen and Maheu (2010, 2014) 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). The infinite mixture of Normals can model 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).

The MCMC estimation 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. (2015b). However, MCMC methods in general are computationally demanding for high-frequency data 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 online type inference by updating the posterior distribution as the new data is observed. Stochastic volatility (parametric or semi-parametric) models are state-space models, naturally suggesting SMC scheme. Moreover, model considered in this paper belongs to such a class, that have the availability of sufficient statistics of the parameters. This naturally suggests using a filter that instead of tracking a high-dimensional vector of the parameters tracks a low-dimensional set of sufficient statistics that can be recursively updated. The use of sufficient statistics has been shown to increase the efficiency of the algorithm by reducing the variance of sampling weights, see Carvalho et al. (2010a).

In general, particle filters provide a simulation based approach where a set of particles represent the posterior density. For instance, consider the following state-space

model, where x_t are latent states and Θ are static parameters:

$$r_t|x_t, \Theta \sim p(r_t|x_t, \Theta),$$
$$x_t|x_{t-1}, \Theta \sim p(x_t|x_{t-1}, \Theta),$$

for $t = 1, \ldots, T$, with initial probability density $p(x_0|\Theta)$ and prior $p(\Theta)$. Each particle has an associated weight that is proportional to the predictive $p(r_t|x_t, \Theta)$. The sequential state filtering and parameter learning problem is solved by a sequence of joint posterior distributions $p(x_t, \Theta|r^t)$, where $r^t = (r_1, \ldots, r_t)$. Assume for the time being that Θ is known, which leaves us with a pure filtering problem. Gordon et al. (1993) and Pitt and Shephard (1999) propose bootstrap and auxiliary particle filters, respectively, which are among the most popular ones. However, when Θ is unknown and also needs to be sequentially estimated, the problem becomes more difficult. The approach of directly introducing and resampling Θ breaks down in a few steps, since all the particles collapse into a single point. In order to delay particle degeneracy, Gordon et al. (1993), and later Liu and West (2001), consider artificial evolution for the parameters. On the other hand, Storvik (2002) and Carvalho et al. (2010a) rely on a low-dimensional set of sufficient statistics, instead of the parameters, to be tracked in time. For discussions and illustrations of some of the particle methods or reviews of particle methods in general, see Johansen and Doucet (2008), Kantas et al. (2009), Douc et al. (2009), Lopes and Tsay (2011), Lopes et al. (2011) together with Chopin et al. (2011) for a lively discussion, Lopes and Carvalho (2013) and Rios and Lopes (2013), among many others. Even if particle filters are known to suffer from a fundamental problem called particle degeneracy, i.e., an ever-decreasing set of atoms in the particle approximation of the density of interest (see Section 2.5), the online property of particle filters is definitively an advantage over MCMC.

Among all available Sequential Monte Carlo methods, 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). Surely, alternative particle filters are in order. Nevertheless, comparison of SMC methods in this setting is out of the scope of this paper. One can find extensive empirical results for comparisons of a variety competing filters in Car-

valho et al. (2010a), Lopes and Tsay (2011) and Rios and Lopes (2013) in more general settings. PL incorporates sequential parameter learning, state filtering and smoothing, thus providing an online estimation alternative to MCMC/FFBS methods. For PL comparison with MCMC see Carvalho et al. (2010a), Lopes and Polson (2010), among others. An essential feature of PL is the presence of conditional sufficient statistics for the parameters to be tracked in time. It also makes model comparison easy, since at each step we have the predictive likelihood as a by-product.

The main contribution of the paper is that we design a PL algorithm for a SV model with DPM innovations, referred to as a semi-parametric model (SPM), which is the same as in Delatola and Griffin (2011). We estimate the simulated data via PL and MCMC in order to illustrate that the produced posteriors are almost identical at any given data point. PL method provides the advantage of easily incorporating the information from the new observation, while MCMC requires to re-run the algorithm again. Additionally, PL produces predictive likelihoods for each data point without any additional costs, which allows for sequential model comparison via log predictive Bayes factors. Finally we estimate real data via PL using the SPM and fully parametric model with Normal innovations, referred to as PM (following the nomenclature of Delatola and Griffin 2011), and perform sequential model comparison in order to illustrate the attractiveness of SMC approach.

Important to notice, that the proposed efficient SMC scheme for this type of models does not come without a cost. Apart from the limitations of particle filters in general, which are outlined in Section 2.6, there is an important shortcoming of PL algorithm for the specific class of models considered in this paper. In particular, in order to design a fully-adapted PL algorithm, the returns have to be transformed by applying a log-square transformation. This transformation masks possible skewness of the distribution of the returns.[1] As acknowledged in Delatola and Griffin (2011), this is a strong assumption, however, they refer to the work of Jensen and Maheu (2010) and argue that the authors found little evidence of skewed returns and showed that a symmetric model exhibits better out of sample performance as compared to an asymmetric one.

The paper is structured as follows. Section 2 presents the linearized SV model with

[1] We would like to thank the Referee for pointing this out.

non-parametric errors and designs a PL algorithm for this model. It also includes a discussion on the limitations of the particle methods in general. Then, Section 3 presents simulated data exercise and comparison with the MCMC estimation output. Section 4 compares the performance of the parametric and semiparametric models 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 semi-parametric SV models. The innovation distribution is assumed to follow an infinite mixture of Gaussians via Dirichlet Process Mixture models, giving rise to the SPM. We show the differences in the computational aspects between PL and MCMC. Meanwhile MCMC is a gold standard in this type of models, PL has 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:

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

$$h_t^* = \alpha + \beta(h_{t-1}^* - \alpha) + \tau\eta_t, \quad t = 1, \ldots, T, \tag{2}$$

where $|\beta| < 1$ 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 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, Mahieu and Schotman, 1998, Liesenfeld and Richard, 2005, for example).

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

$$r_t = h_t^* + \epsilon_t^*, \text{ where } \epsilon_t^* \sim \mathcal{F}, \tag{3}$$

$$h_t^* = \alpha + \beta(h_{t-1}^* - \alpha) + \tau\eta_t, \text{ where } \eta_t \sim \mathcal{N}(0,1). \tag{4}$$

Observe that the distribution $\mathcal{F}$ 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 $\log \chi_1^2$ distribution 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). Moreover, c_O is an offset parameter that is needed in order to avoid the logarithm to be undefined in case zero returns. Delatola and Griffin (2011) have tried several different values for c_O and presented their real data application with $c_O = 10^{-4}$, meanwhile Jensen (2004) has used the value of $c_O = 0.0005$. Therefore, in this paper we fix $c_O = 0.0003$ for all simulated and real data applications.

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

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. Nonetheless, Delatola and Griffin (2011) propose an approximate procedure in

order to recover the underlying true distribution.

2.2 Non-Normal errors

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 (2004), Jensen and Maheu (2010, 2013, 2014), Delatola and Griffin (2011, 2013), Kalli et al. (2013), Ausín et al. (2014) and Virbickaitė et al. (2015a). This type of approach is known as time-invariant (independent) DPM.

Delatola and Griffin (2011, 2013), for example, propose to approximate the log-square of the unknown return distribution $\mathcal{F}$ as an infinite mixture of Normals by relying on DPM models. The SPM presented in this section is of the same spirit as the model in Delatola and Griffin (2011). As noted by the authors, since the mean of the disturbance ϵ_t^* is not fixed and is not known, there might arise some identification issues. Therefore, the mean of the volatility process in (4) can be subsumed into ϵ_t^*, leading to the following reparametrized model:

$$r_t = h_t + \epsilon_t, \text{ where } \epsilon_t \sim \mathcal{F}, \tag{5}$$

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

such that $h_t = h_t^* - \alpha$ and $\epsilon_t = \epsilon_t^* + \alpha$. Here the log volatility process has the unconditional mean equal to zero.

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 $G \sim \mathcal{DP}(c, G_0(\theta; \varrho))$. Each observation ϵ_t comes from a different kernel density 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 that depends on certain hyperparameters ϱ. 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. A rather standard approach is to consider a Gaussian kernel density, $\epsilon_t \sim \mathcal{N}(\mu_t, \sigma_t^2)$, and follow the procedure outlined in Escobar and West (1995) and put a prior on the mixing mean and the variance. Alternatively, we rely on an approach proposed by Griffin (2010) and Delatola and Griffin (2011):

$$\begin{aligned}
\epsilon_t|\mu_t &\sim \mathcal{N}(\mu_t, \alpha\sigma^2), \quad t = 1, \ldots, T, \\
\mu_t &\sim G, \\
G &\sim \mathcal{DP}(c, G_0), \\
G_0 &\sim \mathcal{N}(\mu_0, (1-\alpha)\sigma^2).
\end{aligned}$$

Here μ_0 is the overall location parameter and mixing is done over μ_j, meanwhile σ^2 is the overall scale and is constant. Moreover, the uncertainty associated with μ_j can be integrated out and the prior predictive for ϵ_t is just a single Normal $\mathcal{N}(\mu_0, \sigma^2)$. Parameter α is a smoothness parameter and is fixed to 0.05 throughout the paper. Delatola and Griffin (2011) have also considered a different value of $\alpha = 0.01$; alternatively, α can also be estimated with the rest of the model parameters, see Griffin (2010) for details. The concentration parameter c is set to be equal to one, as seen in Carvalho et al. (2010b), however, it can be estimated together with the rest of model parameters. One can specify some informative priors for μ_0 and σ^2, however, following Delatola and Griffin (2011), we allow for completely uninformative priors.

Define $\Phi = (\beta, \tau^2)$ as the set of parameters associated with the parametric part of the model, $\Omega = \{(\mu_0, \mu_1, \ldots, \sigma^2)\}$ 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,

using a Polya urn representation of DPM, the model in (5) and (6) can be rewritten as follows:

$$r_t|h_t, \Theta \sim \frac{1}{c+t-1} \sum_{j=0}^{L^\star_{t-1}} n_{t-1,j} \mathcal{N}(r_t; \mu_j + h_t, \alpha\sigma_j^2), \tag{7}$$

$$h_t|h_{t-1}, \Theta \sim \mathcal{N}(h_t; \beta h_{t-1}, \tau^2), \tag{8}$$

where $n_{t,j}$ is a number of observations assigned to j^{th} component at time t, $n_0 = c$, $\sigma_j^2 = \sigma^2 \, \forall j > 0$, $\sigma_0^2 = \sigma^2/\alpha$ and $L_t^\star$ 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, 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 $L_t^\star \ll t$.

2.3 MCMC for SPM

The standard Bayesian estimation of SV models, parametric or semi-parametric, relies on MCMC methods, which, however, can be costly, because, additionally to the parameter estimation, they have to consider a sampler for latent volatilities. One notable exception is a work by Jensen (2004), who proposes a highly efficient MCMC sampler for a long memory semiparametric SV model by making use of the SV model's wavelet representation and near-independence of the wavelet coefficients.

Jensen and Maheu (2010) construct an 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 (2014) 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 the 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 high-frequency data for at least three reasons: (1) the MCMC sampler has to include a filter for latent volatilities, (2) the sampler has to be re-run each time a new observation arrives, and (3) sequential consistent model comparison is nearly impossible due to computational burden.

2.4 PL for the SPM

In this section we present the algorithm to perform PL estimation for a SV model with non-parametric errors. PL, as mentioned before, is one of several particle filters that consider 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. The Appendix I 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: $h_0 \sim \mathcal{N}(c_0, C_0)$, $\tau^2 \sim \mathcal{IG}(b_0/2, b_0\tau_0^2/2)$ and $\beta \sim \mathcal{TN}_{(-1,1)}(m_\beta, V_\beta)$. Here $\mathcal{TN}_{(a,b)}$ represents Normal distribution, truncated at a and b, while c_0, C_0, b_0, $b_0\tau_0^2$, m_β, and V_β 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. The object we call particle at time t thus will contain S_t and corresponding parameters, simulated from the hyper-parameters in S_t. At each time t we have a collection of N particles. When this set of

N particles passes from t to $t+1$, some of the particles disappear, some are repeated (sampling with replacement, corresponds to the Resampling step defined below) and then modified (Sampling and Propagating steps).

In order to initiate the algorithm, we need to have the initial set of sufficient statistics S_0 and initial parameter values. The set S_0 consists of: initial $\{h_0^{(i)}\}_{i=1}^N$, that has been simulated from its prior, initial overall location $\{\mu_0^{(i)}\}_{i=1}^N$, which is set to -1.272 for all particles, $\{\sigma^{2(i)}\}_{i=1}^N$, which is set to 4.946. These specific values correspond to the fist two moments of the log χ^2 distribution, which would correspond to Normally distributed returns. The rest of the initial hyper-parameters $\{b_0^{(i)}\}_{i=1}^N, \{b_0\tau_0^{2(i)}\}_{i=1}^N, \ldots$ are all the same across all particles at $t=0$.

For $t = 1 \ldots, T$ and for each particle (i) the algorithm iterates through three steps (the derivations of the posterior distributions are rather straightforward and very similar to the ones available in Griffin 2010 and Delatola and Griffin 2011):

1. **Resampling.**
   Resample the particles from the previous period $t-1$ with weights

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

   that are proportional to the predictive density of the returns ($n_0 = c$, $\sigma_j^2 = \sigma^2 \, \forall j > 0$, $\sigma_0^2 = \sigma^2/\alpha$). The components of $\Theta = (\beta, \tau^2, \mu_1, \ldots, \mu_{L_{t-1}^\star}, \mu_0, \sigma^2)$ 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 log volatilities h_t from

$$h_t | \tilde{h}_{t-1}, \tilde{\Theta}, r_t \sim \mathcal{N}(h_t; m_h, V_h),$$

   where, $V_h = A\tilde{\sigma}^2$, $m_h = A(r_t - \tilde{\mu}_0) + (1-A)\tilde{\beta}\tilde{h}_{t-1}$, and $A = \tilde{\tau}^2/(\tilde{\tau}^2 + \tilde{\sigma}^2)$.

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

to

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

where $\tilde{n}_{L_{t-1}^\star+1} = c$, $\tilde{\sigma}_j^2 = \tilde{\sigma}^2 \, \forall j \leq L_{t-1}^\star$ and $\sigma_{L_{t-1}^\star+1}^2 = \sigma^2/\alpha$. If $k_t \leq L_{t-1}^\star$, $n_{k_t} = \tilde{n}_{k_t} + 1$ and $L_t^\star = L_{t-1}^\star$, otherwise, $L_t^\star = L_{t-1}^\star + 1$ and $n_{k_t} = 1$.

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

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

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

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

(c.3) Sample μ_{k_t} from $\mathcal{N}(\mu_{k_t}; m, V\tilde{\sigma}^2)$, where

$$m = V\left(\frac{\tilde{\mu}_0}{(1-\alpha)} + \frac{s_{k_t}}{\alpha}\right) \quad \text{and} \quad V = \frac{\alpha(1-\alpha)}{\alpha + (1-\alpha)\cdot t'},$$

such that $s_{k_t} = \tilde{s}_{j=k_t} + (r_t - h_t)$.

(c.4) Sample μ_0 from $\mathcal{N}(\alpha; m, V)$, where

$$m = \sum_{j=1}^{L_t^\star} \mu_j \quad \text{and} \quad V = \frac{\tilde{\sigma}^2(1-\alpha)}{L_t^\star}.$$

(c.5) Sample σ^2 from $\mathcal{IG}(\sigma^2; a, b)$, where

$$a = \frac{t + L_t^\star}{2} \quad \text{and} \quad b = \frac{l_{k_t}}{2\alpha} + \frac{\sum_{j=1}^{L_t^\star}(\mu_j - \mu_0)^2}{2(1-\alpha)},$$

such that $l_{k_t} = \tilde{l}_{j=k_t}(r_t - h_t - \mu_{k_t})^2$.

2.5 Limitations of particle filters

Particle filters, PL included, are known to suffer from a problem called particle degeneracy: 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. This has to be monitored carefully for the chosen filter and can be seen as a trade-off between the sequential nature of the algorithm and stability of MCMC for very large samples. Therefore, the *a priori* consideration of the sample size of interest directly influences the choice of number of particles in order not to reach the stage where particles start to degenerate.

Although the development of particle filters is not that new, it is a very active field of research. The ever going quest to avoid or at least postpone particle degeneracy has lead to Gordon et al. (1993) and Liu and West (2001) introducing artificial evolution in the parameters. Another strategy is to use *resample – propagate* strategy rather than *propagate – resample*, as seen in Carvalho et al. (2010a), Lopes and Tsay (2011). Finally, the use of sufficient statistics produces lower MC error than other filters (given the same number of particles), which in turn implies that filters, making use of sufficient statistics – such as PL or Storvik (2002), can reach the same accuracy with a smaller number of particles as other filters. This leaves more room for increase in a number of particles to accommodate desired time-horizon before the particles start vanishing.

Finally, if the interest is not online type inference, MCMC is still a gold standard in the area. Recently other approaches, such as Particle MCMC, that combine MCMC and particle filters, have been emerging, see Andrieu et al. (2010) and Pitt et al. (2012), among others.

3 Simulation exercise and comparison with MCMC

We perform a simulation exercise based on synthetic data to illustrate computational aspects of MCMC and PL approaches. A data set of length $T = 500$ is simulated from the model in (1)-(2) with $\alpha = 0$, $\beta = 0.97$, $\tau^2 = 0.0225$, where v_t is distributed as a

standard Normal. We estimate the SPM using the simulated data with PL and MCMC schemes. The priors for the unknown parameters are the same for MCMC and PL and are given by

$$\beta \sim \mathcal{N}(0.95, 0.1) \quad \text{and} \quad \tau^2 \sim \mathcal{IG}(10/2, 0.1/2).$$

Also, initial values for μ_0 and σ^2 are set the same for both algorithms to match the first two moments of the $\log \chi^2$ distribution. PL is run for 100k particles, meanwhile the MCMC is run for 100000 iterations, keeping every 10th. MCMC results are obtained via `Matlab` code of Delatola and Griffin (2011), which is available on Jim Griffin's website[2]. We have modified the code accordingly, to exactly match our model specification. In particular, the concentration parameter is set to be $c = 1$, the probability of zero returns is always set to be equal to zero and we do not switch between two alternative reparametrizations, as described in Delatola and Griffin (2011). Also, the draws for parameter β are obtained via Gibbs rather than MH step, as in the original code.

For illustrative purposes we also estimate a fully parametric model, where the error term is assumed to be Normally distributed. The $\log \chi^2$ distribution is approximated via carefully tuned mixture of Normals, as seen in Kim et al. (1998). Such approximation allows us to implement the fully adapted filter and allows us to illustrate one of the advantages of the PL algorithm: sequential predictive model performance. In this case we know the underlying DGP, therefore, the sequential predictive Bayes factors should prefer the fully parametric model purely due to much smaller parameter space.

We report estimation results at 5 points of the sample, in particular, at observations $t = 100, 200, 300, 400, 500$. For PL, the algorithm has to be run only once, meanwhile for MCMC it had to be run 5 times. We present the PL results for 4 independent runs in order to get some idea about the Monte Carlo error (the codes were run on a standard desktop computer with four cores, this way all four runs could be carried out in parallel). The smaller the number of particles, the more variability is observed across runs, see Carvalho et al. (2010a) for example.

Figure 1 plots the posterior distributions for the model parameters associated with

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

the non-parametric part - μ_0 and σ^2, and the parameters, governing the volatility process - β and τ^2 at time $T = 100$. The four grey lines correspond to the four independent PL runs, meanwhile the dotted black line draws the MCMC produced posterior distributions. As seen, at time $T = 100$ all posterior distributions are nearly identical. Similar plots can be drawn for each time point t. In order to save space, for the rest of time points instead of drawing all posterior distributions, we plot the PL median, 2.5 and 97.5 percentile paths and the corresponding MCMC medians and 95% credible intervals, see Figure 2. As seen from the plots, the posterior distributions seem very similar for all data points. Instead of the medians and credible bounds for the MCMC only at specific time points, one could also draw the exact paths for all ts, however, this would mean that MCMC algorithm would have to be re-run 500 times.

[Figure 1 about here.]

[Figure 2 about here.]

Next, Figure 3 draws the posterior median, 2.5 and 97.5 percentile paths for PL and corresponding MCMC medians with 95% credible intervals at the selected time cuts for the *filtered* log volatilities. Although for the MCMC we have the entire path of volatilities available, it is important to distinguish that these are smoothed paths, therefore, are not comparable with the only filtered PL paths. If one wishes to obtained smoothed paths in PL setting, it is possible to perform the backwards smoothing afterwards the algorithm has been run, see Carvalho et al. (2010a) for details on smoothing. As seen, the filtered median log volatilities and 95% credible intervals are almost identical for both algorithms.

As mentioned in the Introduction, the predictive distribution of the returns (or their log square transformation) is of major interest. Figure 4 draws posterior predictive distributions for each of the time cuts for MCMC and PL. As seen from the plot, there is very little MC variability among the PL runs and the posterior predictives are identical to those produced by the MCMC. The figure presents such posteriors only for five selected time cuts, however, for PL there are 500 such posterior predictive distributions

readily available. On the other hand, as mentioned before, the MCMC has to be re-run each time a new observation arrives, resulting into prohitively large computational burden if one wants to produce online type inference.

[Figure 3 about here.]

[Figure 4 about here.]

Model comparison. To compare the performance of the models, we use the sequential predictive log Bayes factor (BF). As pointed out in Koop (2003), Bayes factors permit consistent model comparison even for non-nested models. Also, it contains rewards for model fit, accounts for coherency between the prior and the information arising from the data, as well as rewards parsimony. As seen in Kass and Raftery (1995), Bayes factor between two competing models is defined as

$$BF_{12} = \frac{p(D|\mathcal{M}_1)}{p(D|\mathcal{M}_2)},$$

where $p(D|\mathcal{M}_r)$ is the marginal likelihood for data D given a model $\mathcal{M}_r$. Then the log predictive Bayes factor at time $t-1$ for data point r_t is defined as

$$\log BF_{12,t} = \sum_{k=1}^{t} \log p(r_k|r^{k-1}, \mathcal{M}_1) - \sum_{k=1}^{t} \log p(r_k|r^{k-1}, \mathcal{M}_2).$$

The posterior predictive $p(r_t|r^{t-1}, \mathcal{M}_r)$ for model $\mathcal{M}_r$ is obtained as follows:

$$p(r_t|r^{t-1}, \mathcal{M}_r) = \int p(r_t|r^{t-1}, \mathcal{M}_r, \Theta_r)\pi(\Theta_r|r^{t-1}, \mathcal{M}_r)d\Theta_r,$$

where Θ_r is a set of parameters associated with model $\mathcal{M}_r$. The integral above is not always analytically tractable and can be either approximated by using the MCMC output, or is readily available as a by-product in PL scheme. In particular, for each $t = 1, \ldots, T$, the log predictive densities are calculated as

$$\log p(r_t|r^{t-1}) = \frac{1}{N}\sum_{i=1}^{N} \log p(r_t|(\Theta, h_t, k_t)^{(i)}). \tag{9}$$

Finally, Figure 5 illustrates the attractiveness of PL: availability of sequential log predictive likelihoods and Bayes factors, which allow for fast and consistent model comparison. The top panel draws the simulated zero mean return process with Normal errors meanwhile the bottom panel draws the sequential predictive log Bayes factors (across four independent runs for SPM and PM). Since the true data generating process is Normal, as expected, the Bayes factors are negative, showing strong support for the PM. Even though SPM includes PM as a special case, it has much more parameters to estimate, therefore, Bayes factors are negative since they reward parsimony.

[Figure 5 about here.]

This simulation study demonstrates that the posterior distributions for the parameters, filtered volatilities and posterior predictive distribution for the one step ahead squared log returns are identical for both estimation schemes. Moreover, PL allows for sequential consistent model comparison, which is prohibitively costly in MCMC setting.

4 Real Data Application

In this section we present a real data application using return time series for two financial assets, which are the same as in Delatola and Griffin (2011). In particular, we consider the Microsoft company and the SP500 index. The daily prices from Jan/01/2007 till Oct/31/2016 for both assets are obtained from Datastream. The summary of the descriptive statistics for the de-meaned log returns (in %) can be seen in Table 1. In order to closer illustrate the ability of the SPM to capture different distributions of the squared log returns, we split the data into two disjoint periods: a volatile one that includes the financial crisis (Jan/01/2007 - Nov/01/2010) and a calm one (Jan/01/2013 - Oct/31/2016), both containing 1000 observations. Figure 6 draws the daily prices (panels (a) and (b)), the log returns in (%) for the entire period, where the two sub-periods of interest are in black (panels (c) and (d)) and the densities for the squared log returns for the two different sub-periods (panels (e) and (f)). Obviously, the densities for the two sub-periods are very different, expecting calm period returns to be closer to Normal

distribution. The SPM can capture such different shapes via the infinite mixture of Normals, meanwhile the purely parametric model will be fitting the exact same distribution in all four cases.

[Table 1 about here.]

[Figure 6 about here.]

Next, we estimate the data using the SPM and PM specifications. The hyper- parameters for the priors are the same as in the simulation study, the offset parameter value is set to $c_O = 0.0003$. The codes were run for 500k particles each. Figures 7 and 8 present the estimation results for the Microsoft data set. The figures draw sequential predictive Bayes factors as compared to the PM specification and the estimated predictive densities at time $T + 1$ for the two sub-periods. The PM density corresponds to the mixture of 7 Normals, as an approximation of $\log \chi^2_1$. Only by looking at the plots, it is obvious that SPM estimates different densities than the one provided by the fully parametric model. The sequential predictive log Bayes factors confirm the non-Normally distributed returns, i.e. SPM is strongly preferred to PM for both sub-periods. BFs are much larger for the volatile data as compared to the calm period, indicating that calm period returns are closer to Normal.

[Figure 7 about here.]

[Figure 8 about here.]

Figures 9 and 10 present estimation results for the the two sub-periods of the SP500 data set. Same as for the Microsoft data, the SPM is strongly preferred to PM for both sub-periods. Also, the shapes of the predictive distributions for the log squared returns differ dramatically from the ones produced by Normally distributed errors.

[Figure 9 about here.]

[Figure 10 about here.]

To conclude, there is strong evidence that SPM outperforms PM for the selected data sets, confirming the finding present in previous empirical studies. Consistent sequential model comparison is possible via the use of the proposed PL algorithm for semi-parametric SV models.

5 Discussion

This paper designs a sequential estimation procedure, based on PL, for a semi-parametric SV model. PL is comparable to MCMC and allows for sequential inference, which is important in high-frequency data context. SMC also produces the picture of the evolution of parameter learning and provides the predictive likelihoods at each data point as a by-product. The availability of predictive likelihoods at each time point enables to perform fast online model comparison using sequential predictive log Bayes factors. Finally, we present a real data application using two financial time series of the returns for one index - SP500 and one company - Microsoft. As already confirmed in prior empirical semi-parametric SV studies, non-parametric errors provide a better model fit for both, volatile and calm periods.

As noted in the introduction, we use PL to perform sequential Monte Carlo for non-parametric SV models. Nevertheless, other particle filter alternatives are in order. Comparison of these methodologies for the particular models considered in this paper is of interest and we believe it deserves its own space.

Appendix I: 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 online or off-line by approximating $p(\Theta|r^{t+1})$ as follows:

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

   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.

Acknowledgements. We are very grateful to the Associate Editor and Referee for insightful comments and suggestions that helped to improve the manuscript.

References

Andrieu, C., Doucet, A., and Holenstein, R. (2010). "Particle Markov chain Monte Carlo methods." *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 72(3): 269–342.

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, C. K. and Kohn, R. (1994). "On Gibbs sampling for state space models." *Biometrika*, 81: 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., 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. (2002). "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, *Bayesian Statistics 9*, pages 317–360. Oxford University Press."

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.

Douc, R., Moulines, E., and Olsson, J. (2009). "Optimality of the auxiliary particle filter." *Probability and Mathematical Statistics*, 29(1): 1–28.

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.

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

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

Gordon, N., Salmond, D., and Smith, A. (1993). "Novel appraoch to nonlinear/non-Gaussian Bayesian state estimation." *IEE Proceedings F (Radar and Signal Processing)*, 140(2): 107–113.

Griffin, J. E. (2010). "Default priors for density estimation with mixture models." *Bayesian Analysis*, 5(1): 45–64.

Harvey, A., Ruiz, E., and Shephard, N. (1994). "Multivariate Stochastic Variance Models." *The Review of Economic Studies*, 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. (2004). "Semiparametric Bayesian Inference of Long-Memory Stochastic Volatility Models." *Journal of Time Series Analysis*, 25(6).

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.

Johansen, A. M. and Doucet, A. (2008). "A note on auxiliary particle filters." *Statistics & Probability Letters*, 78(12): 1498–1504.

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.

Kantas, N., Doucet, a., Singh, S. S., and Maciejowski, J. M. (2009). "An Overview of Sequential Monte Carlo Methods for Parameter Estimation in General State-Space Models." *Engineering*, 44(Ml): 774–785.

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.

Liesenfeld, R. and Richard, J.-F. (2005). "Classical and Bayesian Analysis of Univariate and Multivariate Stochastic Volatility Models."

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. (2013). "Online Bayesian Learning in Dynamic Models: an Illustrative Introduction to Particle Methods." In Damien, P., Dellaportas, P., Polson, N. G., and Stephens, D. A. (eds.), *Bayesian Theory and Applications*, chapter 11, 203–228. Oxford: Oxford University Press, 1 edition.

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.

Lopes, H. F. and Tsay, R. S. (2011). "Particle Filters and Bayesian Inference in Financial Econometrics." *Journal of Forecasting*, 30(July): 168–209.

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.

Pitt, M. K. and Shephard, N. (1999). "Filtering via Simulation: Auxiliary Particle Filters." *Journal of the American Statistical Association*, 94(446): 590–599.

Pitt, M. K., Silva, R. D. S., Giordani, P., and Kohn, R. (2012). "On some properties of Markov chain Monte Carlo simulation methods based on the particle filter." *Journal of Econometrics*, 171(2): 134–151.

Rios, M. P. and Lopes, H. F. (2013). "The Extended Liu and West Filter: Parameter Learning in Markov Switching Stochastic Volatility Models." In Zeng, Y. and Wu, S. (eds.), *State-Space Models: Statistics and Econometrics for Finance*, chapter 2, 23–61. Springer New York.

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

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 1961 - 79." In Anderson, O. D. (ed.), *Time Series Analysis: Theory and Practice 1*. 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.

Virbickaitė, A., Ausín, M. C., and Galeano, P. (2015a). "A Bayesian Non-Parametric Approach to Asymmetric Dynamic Conditional Correlation Model with Application to Portfolio Selection." *Computational Statistics and Data Analysis*, Forthcoming.

— (2015b). "Bayesian Inference Methods for Univariate and Multivariate Garch Models: a Survey." *Journal of Economic Surveys*, 29(1): 76–96.

Figure 1: Posterior distributions for the parameters for MCMC (black dotted line) and four runs of PL (grey lines) at time $T = 100$.

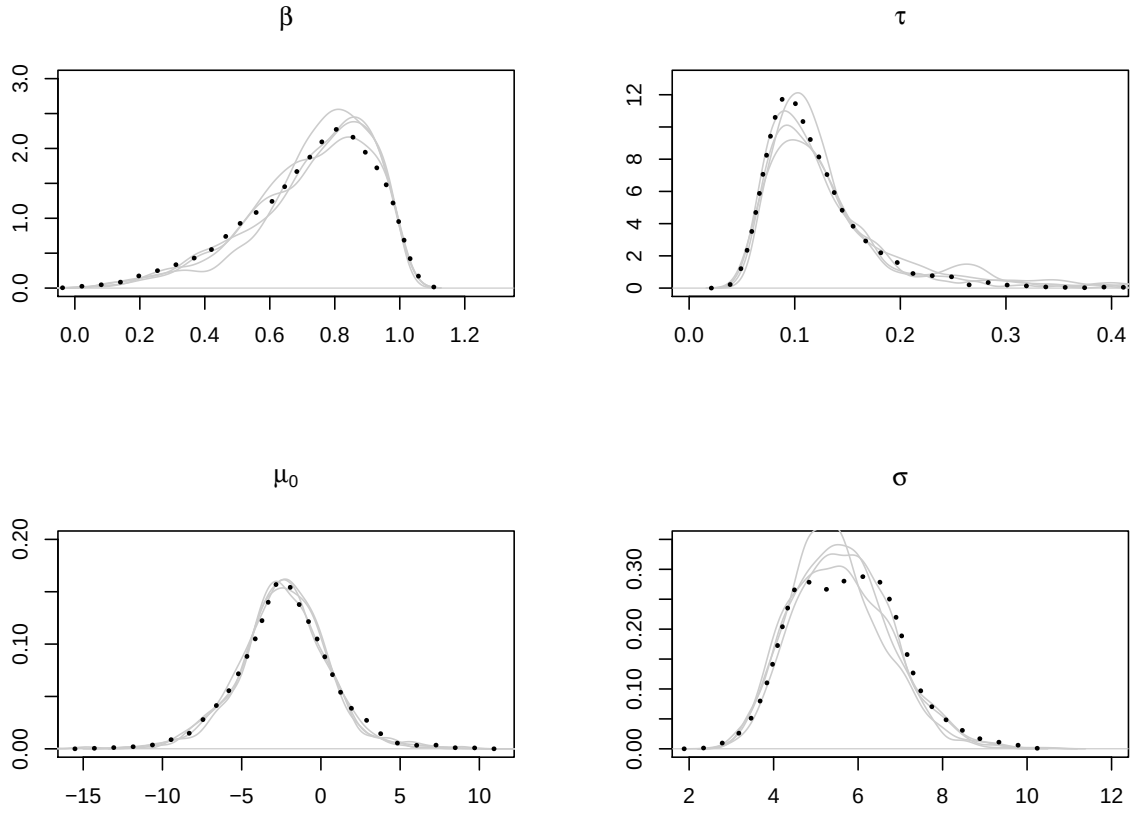

Figure 2: Posterior median, 2.5 and 97.5 percentile paths for PL and corresponding MCMC medians with 95% credible intervals at $T = \{100, 200, 300, 400, 500\}$ for the model parameters.

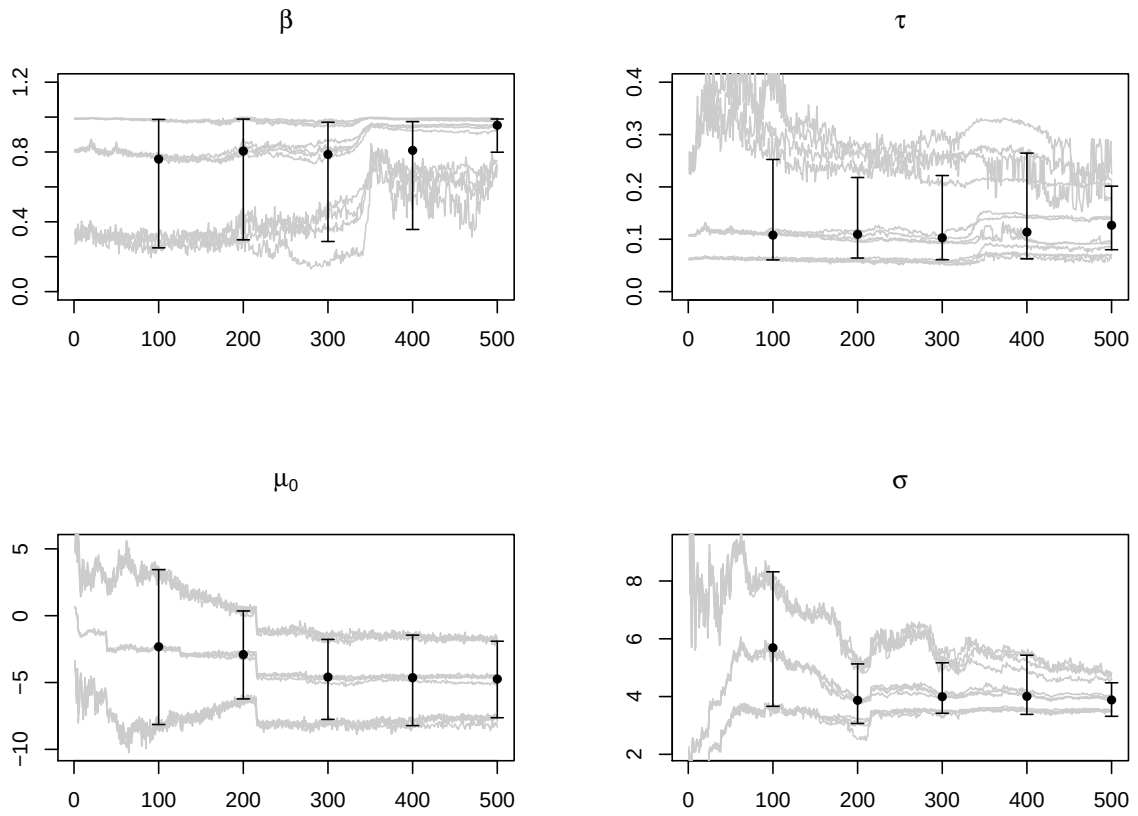

Figure 3: Posterior median, 2.5 and 97.5 percentile paths for PL and corresponding MCMC medians with 95% credible intervals at $T = \{100, 200, 300, 400, 500\}$ for the filtered log volatility process.

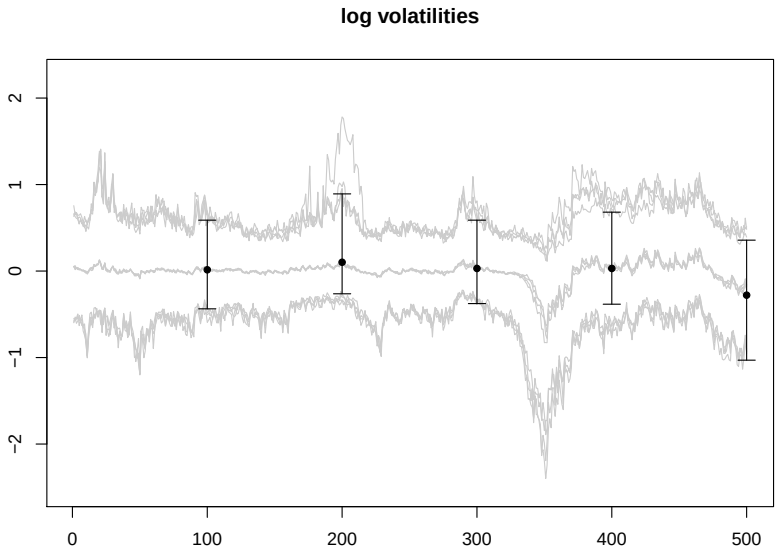

Figure 4: Posterior predictive distribution of the squared log returns at $T = \{100, 200, 300, 400, 500\}$ for PL (grey lines) and MCMC (black dotted line).

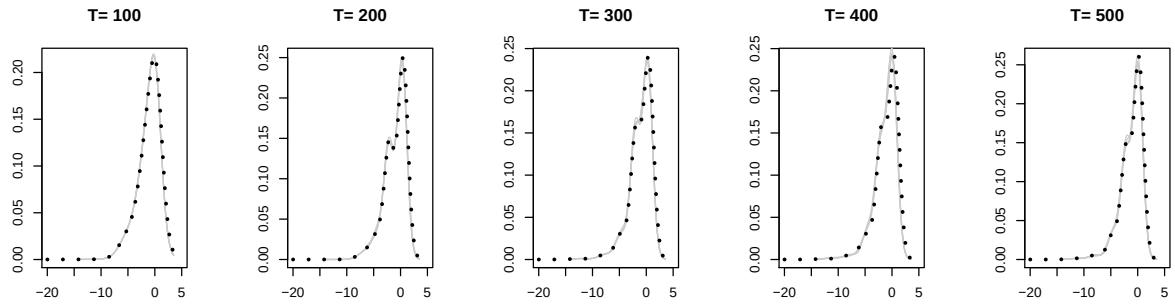

Figure 5: Simulated data (top panel) and sequential predictive log Bayes factor for SPM vs PM (bottom panel).

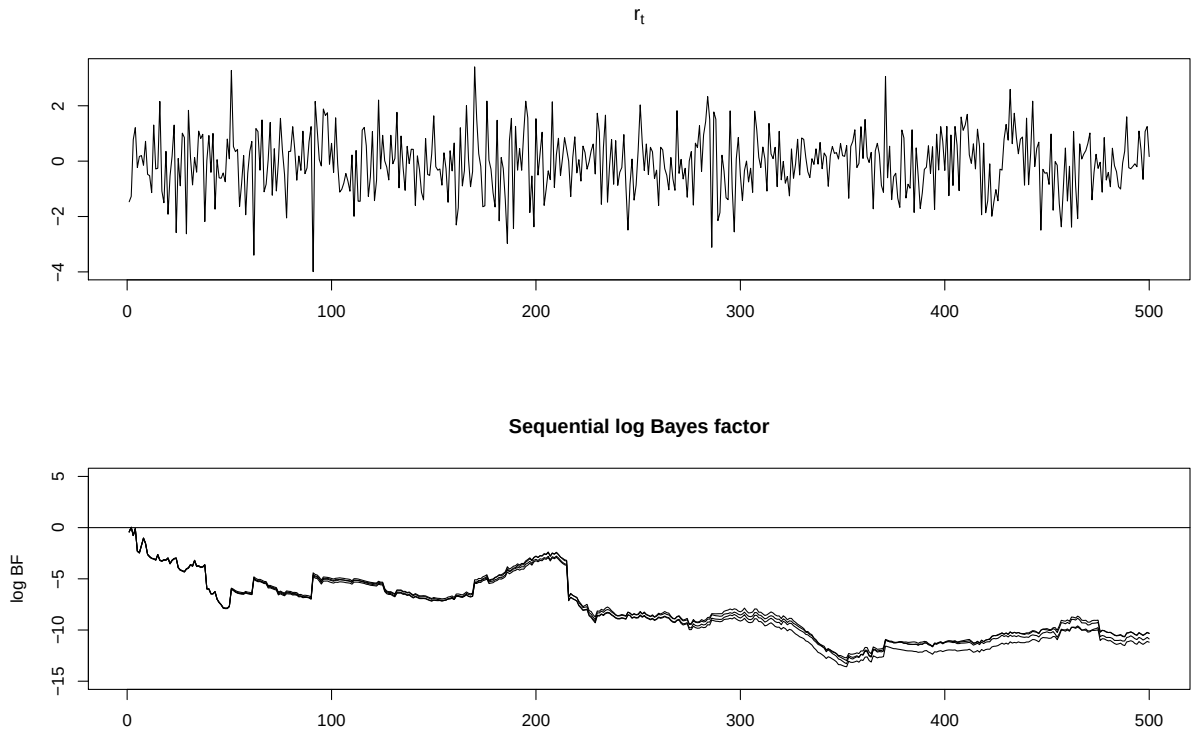

Figure 6: Daily prices, log-returns (in %) and densities fort the log-squared returns for two sub-periods for Microsoft & SP500 data.

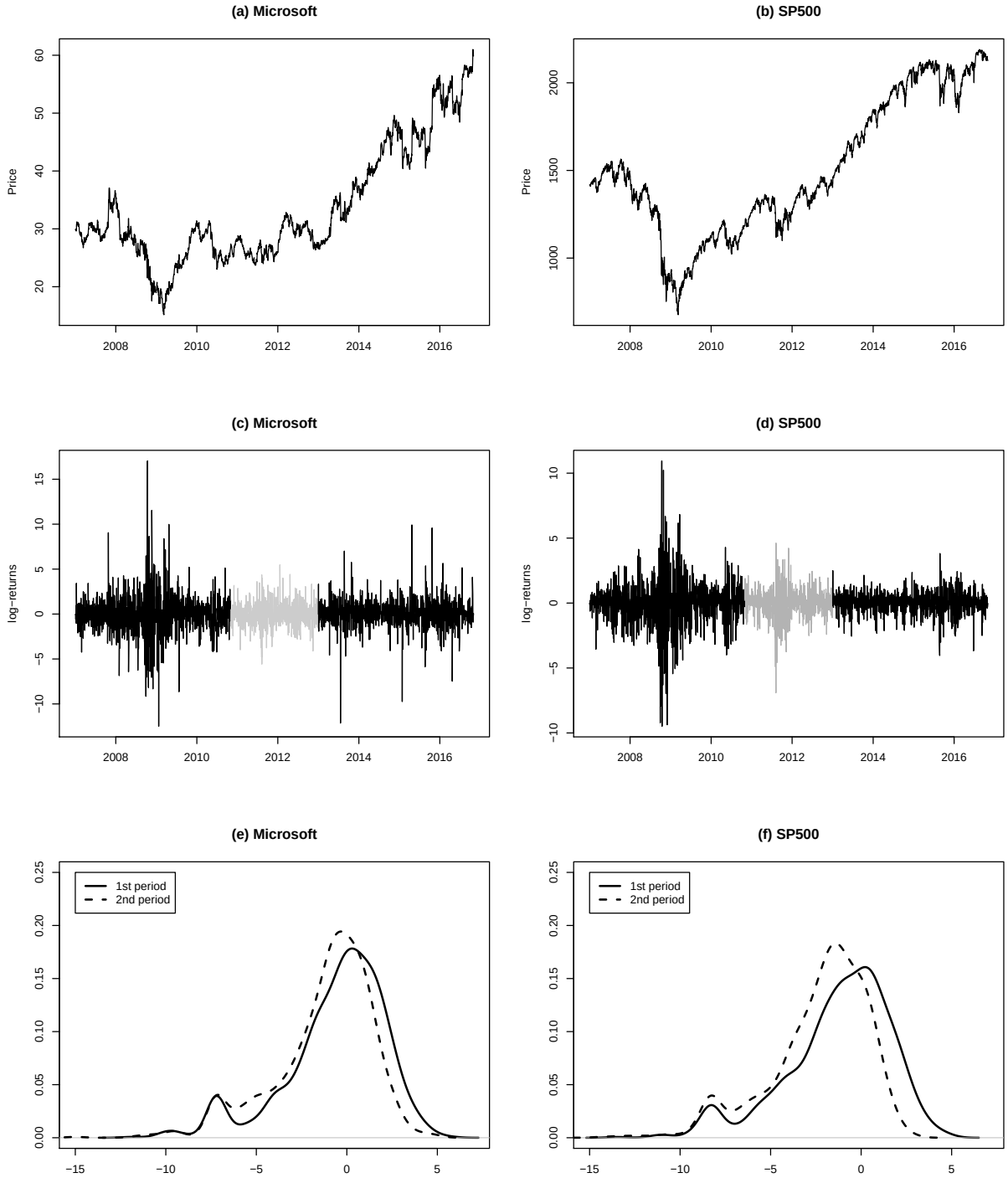

Figure 7: Sequential log predictive Bayes factors and estimated densities for the log-squared error term for SPM, as compared to the PM for Microsoft data for the fist period.

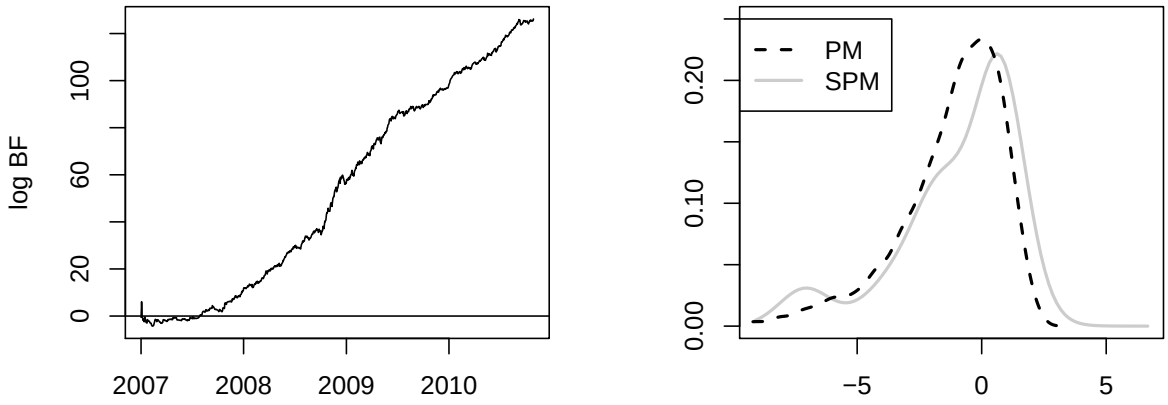

Figure 8: Sequential log predictive Bayes factors and estimated densities for the log-squared error term for SPM, as compared to the PM for Microsoft data for the second period.

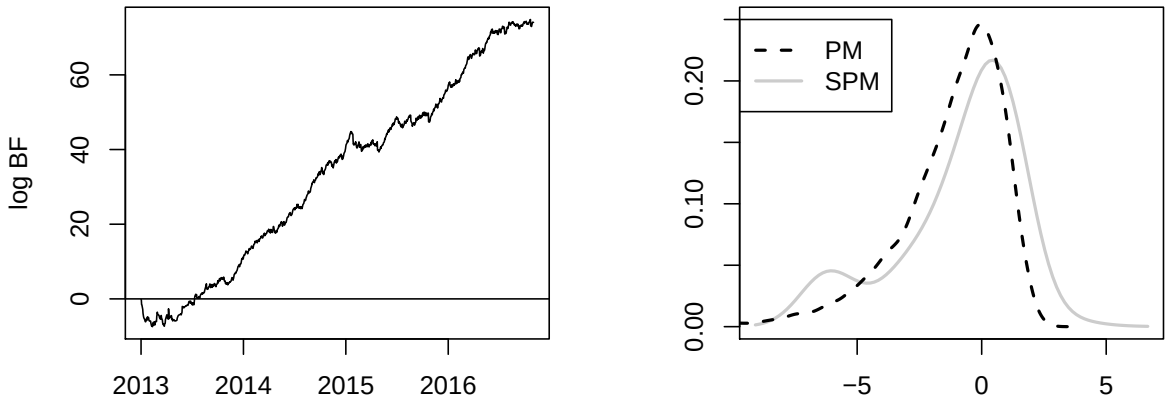

Figure 9: Sequential log predictive Bayes factors and estimated densities for the log-squared error term for SPM, as compared to the PM for SP500 data for the first period.

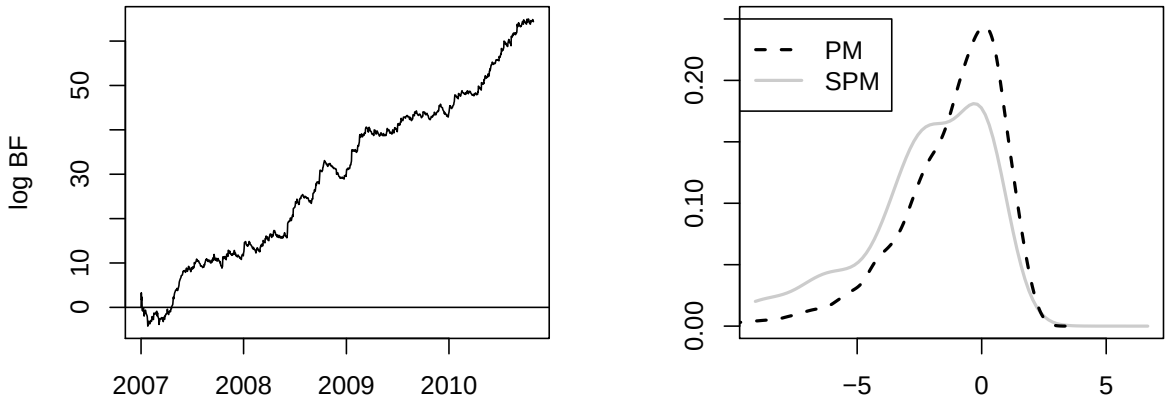

Figure 10: Sequential log predictive Bayes factors and estimated densities for the log-squared error term for SPM, as compared to the PM for SP500 data for the second period.

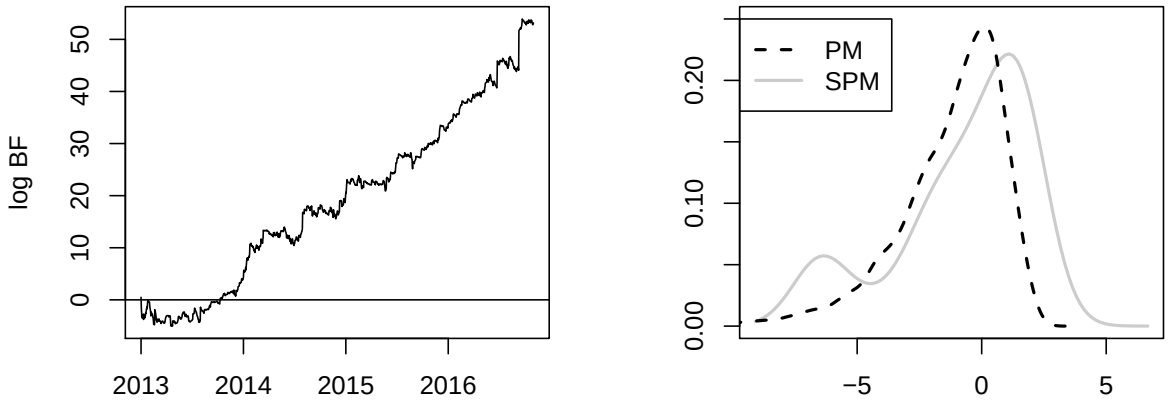

Table 1: Descriptive statistics for Microsoft and SP500 data.

	Microsoft	SP500
Mean	0.0000	0.0000
Median	-0.0271	0.0129
St.dev.	1.7581	1.3051
Skewness	0.1926	-0.3273
Kurtosis	12.7410	13.2423