Universidad
Carlos III de Madrid
www.uc3m.es

TESIS DOCTORAL

Bayesian Non-Parametrics for Time-Varying Volatility Models

Autor:

Audronė Virbickaitė

Directores:

Concepción Ausín y Pedro Galeano

DEPARTAMENTO DE ESTADÍSTICA

Getafe, diciembre de 2014

Universidad Carlos III de Madrid
www.uc3m.es

TESIS DOCTORAL

BAYESIAN NON-PARAMETRICS FOR TIME-VARYING VOLATILITY MODELS

Autor: *Audronė Virbickaitė*

Directores: Concepción Ausín y Pedro Galeano

Firma del Tribunal Calificador:

Firma

Presidente: Michael P. Wiper

Vocal: Roberto Casarin

Secretario: M. Pilar Muñoz

Calificación:

Getafe, de de

UNIVERSIDAD CARLOS III DE MADRID

DOCTORAL THESIS

Bayesian Non-Parametrics for Time-Varying Volatility Models

Author:
Audronė Virbickaitė

Supervisors:
Dr. Concepción Ausín
Dr. Pedro Galeano

A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy

in the

Business Administration and Quantitative Methods
Statistics Department

December 2014

Acknowledgements

I want to express my deepest appreciation to a multitude of people who have been a big part of my life for the last three years during which this thesis came to life. For professional and personal reasons. For being my mentors and friends. For sharing an office and traveling with me. For giving advice or just listening. For being from all over the world.

First of all, I am very grateful to my thesis advisors Conchi and Pedro. Thank you very much for always finding time for me, for having the patience to teach and let me learn. And most of all for the academic guidance: Conchi for being an excellent Bayesian and Pedro for your methodical and attentive approach; after all, the devil *is* in the detail.

I am especially grateful to Prof. Hedibert Lopes for accepting me as a visiting PhD student at University of Chicago. I feel extremely lucky to have been able to work with and learn from such an extraordinary individual and an excellent researcher. Muito obrigada e espero vê-lo em breve.

Esu labai dėkinga savo šeimai Lietuvoje už tai, kad grįžus visada jaučiuosi namuose. Ačiū mamytei už cepelinus, džiovintus grybus kaskart, kai aplankau, ir už besąlyginę meilę visada. Ačiū tėtei ir Vladai už palaikymą, pilnus šaldytuvus alaus, pirtį ir šaltą protą, kai norisi priimti karštakošį sprendimą. Ačiū Jonytei ir Domukui, Jolitukui ir Andriui ne tik už brolišką/seserišką meilę, bet ir draugystę. Ačiū seneliui Juozui už meilę mokslui ir už buvimą pirmu mano pažintu akademiku. Ačiū Irutėlei už beribę meilę ir dosnumą. Ačiū seneliui Jonui, močiutei Jadvygai.

Lo más importante, muchas gracias a Juanín. No solo por acompañarme todos estos años como mi mejor amigo y familia, pero también por aguantarme en las últimas etapas de la tesis. Gracias por compartir conmigo tu amor a viajar, a historia y a Asturias. También muchas gracias a mi familia política. Gracias a Pilarita, Rodrigo, Chiqui, Mela, Alfredo, R7, Lurditas y Pepe por *adoptarme* y por cuidarme. Por los findes de descanso, playa de Vidiago y sidra sin fin.

Ačiū mano draugams iš Lietuvos ne tik už palaikymą, bet ir už smagiai praleistas poilsio minutes, dienas, savaites. Ačiū Rūtai už nuolatinį *you can do it!* ir už buvimą mano *bff*, Miglutei, Ingai ir Dovilei už be galo smagius susitikimus. Ačiū visai Labanoro kompanijai: Gabrieliui, Agnytei, Tomui, Ramintai, Laimučiui, Gabijai, ir dar kartą Jonytei ir Domukui.

Finalmente, quiero agradecer a todos mis compañeros y amigos de la Universidad Carlos III, el lugar donde durante los últimos cinco años pasé más tiempo que en casa. Muchas gracias a Juanmi por ayudarme en la resolución de dudas, profesionales o personales, y por el fondo musical. Pau y Ana, mis hermanitas mayores, las primeras personas que vi pasar por la tesis y aprendí mucho de ellas. Mil gracias al resto de mis compañeros con quienes los cinco años en Madrid han sido increíbles: Juliana, Diego, Carlo-Charly, Mahmoud, Guille, Huro, Diana, Adolfo, Xiuping. Lo siento mucho si se me olvidó incluir a alguien, sin embargo todos formáis una parte importante de mi vida.

Contents

List of Figures

List of Tables

Skirta mano šeimai ir Juanui.

Chapter 1

Introduction

Understanding, modeling and predicting volatilities of financial time series has been extensively researched for more than 30 years and the 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 and Stochastic Volatility (SV) type models. The seminal paper of Engle (1982) proposed the initial ARCH model while Bollerslev (1986) generalized the purely autoregressive ARCH into an ARMA-type model, called the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model. On the other hand, Taylor (1982, 1986) proposed to model the volatility as an unobserved process, giving the start to SV models. Since then, there has been a very large amount of research on the topic, stretching to various model extensions and generalizations. The supply of models, univariate and multivariate, GARCH and SV, has been growing over the years. Meanwhile, the researchers have been addressing two important topics: looking for the best specification for the errors and selecting the most efficient approach for inference and prediction. This thesis puts emphasis on these two questions as well.

Besides selecting the best model for the volatility, distributional assumptions for the returns are equally important. It is well known, that every prediction, in order to be useful, has to come with a certain precision measurement. In this way the agent can

know the risk she is facing, i.e. uncertainty. Distributional assumptions permit to quantify this uncertainty about the future. Traditionally, the errors have been assumed to be Gaussian, however, it has been widely acknowledged that financial returns display fat tails and are not conditionally Normal. Therefore, it is common to assume a Student-t distribution, see Bollerslev (1987), He & Teräsvirta (1999), Bai et al. (2003) and Jacquier et al. (2004), among many others. However, the assumption of Gaussian or Student-t distributions is rather restrictive. An alternative approach is to use a mixture of distributions, which can approximate arbitrarily any distribution given a sufficient number of mixture components. A mixture of two Normals was used by Bai et al. (2003), Ausín & Galeano (2007) and Giannikis et al. (2008), among others. These authors have shown that the models with the mixture distribution for the errors outperformed the Gaussian ones and do not require additional restrictions on the degrees of freedom parameter as the Student-t one.

As for the inference and prediction, the Bayesian approach is especially well-suited for GARCH and SV models and provides some advantages compared to classical estimation techniques, as outlined by Ardia & Hoogerheide (2010). Firstly, the positivity constraints on the parameters to ensure positive variance, may encumber some optimization procedures. In the Bayesian setting, constraints on the model parameters can be incorporated via priors. Secondly, in most of the cases we are more interested not in the model parameters directly, but in some non-linear functions of them. In the maximum likelihood (ML) setting, it is quite troublesome to perform inference on such quantities, while in the Bayesian setting it is usually straightforward to obtain the posterior distribution of any non-linear function of the model parameters. Furthermore, in the classical approach, models are usually compared by any other means than the likelihood. In the Bayesian setting, marginal likelihoods and Bayes factors allow for consistent comparison of non-nested models while incorporating Occam's razor for parsimony. Also, Bayesian estimation provides reliable results even for finite samples. Finally, Hall & Yao (2003) add that the ML approach presents some limitations when the errors are heavy tailed, also the convergence rate is slow and the estimators may not be asymptotically Gaussian.

Therefore in this thesis we consider different Bayesian non-parametric specifications for the errors for GARCH and SV models. Also, we employ two Bayesian estimation approaches: Markov Chain Monte Carlo (MCMC) and Sequential Monte Carlo (SMC).

The thesis is structured as follows:

Chapter 2 reviews the existing literature on the most relevant Bayesian inference methods for univariate and multivariate GARCH and SV models. The advantages and drawbacks of each procedure are outlined as well as the advantages of the Bayesian approach versus classical procedures. The chapter makes emphasis on Bayesian non-parametrics for time-varying volatility models that avoid imposing arbitrary parametric distributional assumptions. Finally, the chapter presents an alternative Bayesian estimation technique - Sequential Monte Carlo, that allows for an on-line type inference. The major part of the contents of this chapter resulted into a paper by Virbickaitė et al. (2013), which has been accepted in the *Journal of Economic Surveys*.

Chapter 3 considers an asymmetric dynamic conditional correlation (ADCC) model to estimate the time-varying correlations of financial returns where the individual volatilities are driven by GJR-GARCH models. This composite model takes into consideration the asymmetries in individual assets' volatilities, as well as in the correlations. The errors are modeled using a Dirichlet location-scale mixture of multivariate Normals allowing for a flexible return distribution in terms of skewness and kurtosis. This gives rise to a Bayesian non-parametric ADCC (BNP-ADCC) model, as opposed to a symmetric specification, called BNP-DCC. Then these two models are estimated using MCMC and compared by considering a sample of Apple Inc. and NASDAQ Industrial index daily returns. The obtained results reveal that for this particular data set the BNP-ADCC outperforms the BNP-DCC model. Finally, an illustrative asset allocation exercise is presented. The contents of this chapter resulted into a paper by Virbickaitė, Ausín & Galeano (2014), which has been accepted in *Computational Statistics and Data Analysis*.

Chapter 4 designs a Particle Learning (PL) algorithm for estimation of Bayesian non-parametric Stochastic Volatility models for financial data. The performance of this particle method is then compared with the standard 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. The contents of this chapter resulted into a working paper by Virbickaitė, Lopes, Ausín & Galeano (2014).

Finally, Chapter 5 concludes and proposes general future research lines that could be viewed as natural extensions of the ideas presented in the thesis.

Chapter 2

Bayesian Inference for Time-Varying Volatility Models

This chapter reviews the existing Bayesian inference methods for univariate and multivariate GARCH and SV models while having in mind their error specifications. The main emphasis of this chapter is on the recent development of an alternative inference approach for these models using Bayesian non-parametrics. The classical parametric modeling, relying on a finite number of parameters, although so widely used, has some certain drawbacks. Since the number of parameters for any model is fixed, one can encounter underfitting or overfitting, which arises from the misfit between the data available and the parameters needed to estimate. Then, in order to avoid assuming wrong parametric distributions, which may lead to inconsistent estimators, it is better to consider a semi- or non-parametric approach. Bayesian non-parametrics may lead to less constrained models than classical parametric Bayesian statistics and provide an adequate description of the data, especially when the conditional return distribution is far away from Gaussian.

The literature on non-parametric GARCH and SV type models is still very recent, however, the popularity of the topic is rapidly increasing, see Jensen & Maheu (2010, 2013, 2014), Delatola & Griffin (2011, 2013) and Ausín et al. (2014). All of them have considered infinite mixtures of Gaussian distributions with a Dirichlet process (DP) prior over the mixing distribution, which results into DP mixture (DPM) models (see Lo 1984 and Ferguson 1973, among others). This approach proves to be the most popular Bayesian

non-parametric modeling procedure so far. The results over the papers have been consistent: Bayesian non-parametric methods lead to more flexible models and are better in explaining heavy-tailed return distributions, which parametric models cannot fully capture.

The outline of this chapter is as follows. Sections 2.1 and 2.2 shortly introduce univariate and multivariate GARCH and SV models and different inference and prediction methods. Section 2.3 introduces the Bayesian non-parametric modeling approach and reviews the limited literature of this area in time-varying volatility models. Finally, Section 2.5 concludes.

2.1 Univariate and multivariate GARCH

In this section we shortly introduce the most popular univariate and multivariate GARCH specifications. In the description of the models and the review of the inference methods we are not going to enter into the technical details of the Bayesian algorithms and refer to Robert & Casella (2004) for a more detailed description of the mentioned Bayesian techniques.

Univariate GARCH

The general structure of an asset return series modeled by a GARCH-type models can be written as:

$$r_t = \mu_t + a_t = \mu_t + \sqrt{h_t}\epsilon_t, \tag{2.1}$$

where $\mu_t = \mathrm{E}\left[r_t|\mathcal{I}^{t-1}\right]$ is the conditional mean given $\mathcal{I}^{t-1}$, the information up to time $t-1$, a_t is the mean corrected returns of the asset at time t, $h_t = \mathrm{Var}\left[r_t|\mathcal{I}^{t-1}\right]$ is the conditional variance given $\mathcal{I}^{t-1}$ and ϵ_t is the standard white noise shock. There are several ways to model the conditional mean μ_t. The usual assumptions are to consider that the mean is either zero, equal to a constant ($\mu_t = \mu$), or follows an ARMA(p,q) process. However, sometimes the mean is also modeled as a function of the variance, say $g(h_t)$, which leads to the GARCH-in-Mean models. On the other hand, the conditional

variance, h_t, is usually modeled using the GARCH-family models. In the basic GARCH model the conditional volatility of the returns depends on a sum of three parts: a constant variance as the long-run average, a linear combination of the past conditional volatilities and a linear combination of the past mean squared returns. For instance, in the GARCH(1,1) model, the conditional variance at time t is given by

$$h_t = \omega + \alpha a_{t-1}^2 + \beta h_{t-1}, \text{ for } t = 1, \ldots, T. \tag{2.2}$$

There are some restrictions which have to be imposed such as $\omega > 0$, $\alpha, \beta \geq 0$ for positive variance, and $\alpha + \beta < 1$ for covariance stationarity.

Nelson (1991) proposed the exponential GARCH (EGARCH) model that acknowledges the existence of asymmetry in the volatility response to the changes in the returns, sometimes also called "leverage effect", introduced by Black (1976). Negative shocks to the returns have a stronger effect on volatility than positive ones. Other ARCH extensions that try to incorporate the leverage effect are the GJR model by Glosten et al. (1993) and the TGARCH of Zakoian (1994), among many others. As Engle (2004) puts it, "there is now an alphabet soup" of ARCH family models, such as AARCH, APARCH, FIGARCH, STARCH etc, which try to incorporate such return features as fat tails, volatility clustering and volatility asymmetry. Papers by Bollerslev et al. (1992), Bollerslev et al. (1994), Engle (2002*b*), Ishida & Engle (2002) provide extensive reviews of the existing ARCH-type models. Bera & Higgins (1993) review ARCH type models, discuss their extensions, estimation and testing, also numerous applications. Additionally, one can find an explicit review with examples and applications concerning GARCH-family models in Tsay (2010) and Chapter 1 in Teräsvirta (2009).

The most used estimation approach for GARCH-family models is the maximum likelihood method. However, recently there has been a rapid development of Bayesian estimation techniques, which offer some advantages compared to the frequentist approach as already discussed in the beginning of this chapter. In addition, in the empirical finance setting, the frequentist approach presents an uncertainty problem. For instance, optimal allocation is greatly affected by the parameter uncertainty, which has been recognized in a number of papers, see Jorion (1986) and Greyserman et al. (2006), among others. These authors conclude that in the frequentist setting the estimated parameter values are considered to be the true ones, therefore, the optimal portfolio weights tend

to inherit this estimation error. However, instead of solving the optimization problem on the basis of the choice of unique parameter values, the investor can choose the Bayesian approach, because it accounts for parameter uncertainty, as seen in Kang (2011) and Jacquier & Polson (2013), for example. A number of papers in this field have explored different Bayesian procedures for inference and prediction and different approaches to model the fat-tailed errors and/or asymmetric volatility. The recent development of modern Bayesian computational methods, based on Monte Carlo approximations and MCMC methods have facilitated the usage of Bayesian techniques, see e.g. Robert & Casella (2004).

The standard Gibbs sampling procedure does not make the list because it cannot be used due to the recursive nature of the conditional variance: the conditional posterior distributions of the model parameters are not of a simple form. One of the alternatives is the *Griddy-Gibbs* sampler as in Bauwens & Lubrano (1998). They discuss that previously used importance sampling and Metropolis algorithms have certain drawbacks, such as that they require a careful choice of a good approximation of the posterior density. The authors propose a Griddy-Gibbs sampler which explores analytical properties of the posterior density as much as possible. In their paper GARCH model has Student-t errors, which allows for fat tails. The authors choose to use flat (Uniform) priors on parameters (ω, α, β) with whatever region is needed to ensure the positivity of variance, however, the flat prior for the degrees of freedom cannot be used, because then the posterior density is not integrable. Instead, they choose a half-right side of Cauchy. The posteriors of the parameters were found to be skewed, which is a disadvantage for the commonly used Gaussian approximation. On the other hand, Ausín & Galeano (2007) modeled the errors of a GARCH model with a mixture of two Gaussian distributions. The advantage of this approach, compared to that of Student-t errors, is that if the number of the degrees of freedom is very small (less than 5), some moments may not exist. The authors have chosen flat priors for all the parameters, and discovered that there is little sensitivity to the change in the prior distributions (from Uniform to Beta), unlike in Bauwens & Lubrano (1998), where the sensitivity for the prior choice for the degrees of freedom is high. Other articles using a Griddy-Gibbs sampling approach include Bauwens & Lubrano (2002), who have modeled asymmetric volatility with Gaussian innovations and have used Uniform priors for all the parameters, and by Wago (2004), who explored an asymmetric GARCH model with Student-t errors.

Another MCMC algorithm used in estimating GARCH model parameters, is the *Metropolis - Hastings* (MH) method, which samples from a candidate density and then accepts or rejects the draws depending on a certain acceptance probability. Ardia (2006) modeled the errors as Gaussian distributed with zero mean and unit variance while the priors are chosen as Gaussian and a MH algorithm is used to draw samples from the joint posterior distribution. The author has carried out a comparative analysis between ML and Bayesian approaches, finding, as in other papers, that some posterior distributions of the parameters were skewed, thus warning against the abusive use of the Gaussian approximation. Also, Ardia (2006) has performed a sensitivity analysis of the prior means and scale parameters and concluded that the initial priors in this case are vague enough. This approach has been also used by Müller & Pole (1998), Nakatsuma (2000) and Vrontos et al. (2000), among others. A special case of the MH method is the random walk Metropolis-Hastings (RWMH) where the proposal draws are generated by randomly perturbing the current value using a spherically symmetric distribution. A usual choice is to generate candidate values from a Gaussian distribution where the mean is the previous value of the parameter and the variance can be calibrated to achieve the desired acceptance probability. This procedure is repeated at each MCMC iteration. Ausín & Galeano (2007) have also carried out a comparison of estimation approaches, Griddy-Gibbs, RWMH and ML. Apparently, RWMH has difficulties in exploring the tails of the posterior distributions and ML estimates may be rather different for those parameters where posterior distributions are skewed.

In order to select one of the algorithms, one might consider some criteria, such as fast convergence for example. Asai (2006) numerically compares some of these approaches in the context of GARCH models. The Griddy-Gibbs method is capable in handling the shape of the posterior by using shorter MCMC chains comparing with other methods, also, it is flexible regarding parametric specification of the model. However, it can require a lot of computational time. This author also investigates MH, adaptive rejection Metropolis sampling (ARMS), proposed by Gilks et al. (1995), and acceptance-rejection MH algorithms (ARMH), proposed by Tierney (1994). For more in detail about each method in GARCH models see Nakatsuma (2000) and Kim et al. (1998), among others. Using simulated data, Asai (2006) calculated geometric averages of inefficiency factors for each method. Inefficiency factor is just an inverse of Geweke (1992) efficiency factor. According to this, the ARMH algorithm performed the best. Also, computational

time was taken into consideration, where ARMH clearly outperformed MH and ARMS, while Griddy-Gibbs stayed just a bit behind. The author observes that even though the ARMH method showed the best results, the posterior densities for each parameter did not quite explore the tails of the distributions, as desired. In this case Griddy-Gibbs performs better; also, it requires less draws than ARMH. Bauwens & Lubrano (1998) investigate one more convergence criteria, proposed by Yu & Mykland (1998), which is based on cumulative sum (cumsum) statistics. It basically shows that if MCMC is converging, the graph of a certain cumsum statistic against time should approach zero. Their employed Griddy-Gibbs algorithm converged in all four parameters quite fast. Then, the authors explored the advantages and disadvantages of alternative approaches: the importance sampling and the MH algorithm. Considering importance sampling, one of the main disadvantages, as mentioned before, is to find a good approximation of the posterior density (importance function). Also, comparing with Griddy-Gibbs algorithm, the importance sampling requires much more draws to get smooth graphs of the marginal densities. For the MH algorithm, same as in importance sampling, a good approximation needs to be found. Also, compared to Griddy-Gibbs, the MH algorithm did not fully explore the tails of the distribution, unless for a very big number of draws.

Another important aspect of the Bayesian approach, as commented before, is the advantage in model selection compared to classical methods. Miazhynskaia & Dorffner (2006) reviews some Bayesian model selection methods using MCMC for GARCH-type models, which allow for the estimation of either marginal model likelihoods, Bayes factors or posterior model probabilities. These are compared to the classical model selection criteria showing that Bayesian approach clearly considers model complexity in a more unbiased way. Also, Chen et al. (2009) includes a revision of Bayesian selection methods for asymmetric GARCH models, such as the GJR-GARCH and threshold GARCH. They show how using Bayesian approach it is possible to compare complex and non-nested models to choose for example between GARCH and stochastic volatility models, between symmetric or asymmetric GARCH models or to determine the number of regimes in threshold processes, among others.

Markov Switching GARCH (MS-GARCH). One of the most prominent features of the volatilities of financial time series is a very high persistence of the variance process, which in some cases is very close to having a unit root. Some authors argue that the

upward bias in the persistence parameter might occur due to the presence of structural changes in volatility, which simple GARCH models do not account for. Therefore, Hamilton & Susmel (1994) and Cai (1994) independently proposed a Markov Switching ARCH model, which later was generalized by Gray (1996) into MS-GARCH. Differently than simple GARCH model, defined in Equation 2.1 and Equation 2.2, the MS-GARCH model has the following representation:

$$
\begin{aligned}
r_t &= \mu_{s_t} + a_t = \mu_{s_t} + \sqrt{h_t}\epsilon_t, \\
h_t &= \omega_{s_t} + \alpha_{s_t} a_{t-1}^2 + \beta_{s_t} h_{t-1}, \text{ for } t = 1, \ldots, T,
\end{aligned}
$$

where s_t are the regime variables following a J-state first order Markov Process with the following transition probabilities:

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

Kaufmann & Frühwirth-Schnatter (2002) designed an MCMC scheme to generate a sample from the posterior of a MS-ARCH model, which has not been done before, by combining a multi-move sampling of a hidden Markov process with Metropolis – Hastings for parameter estimation. The authors have performed model selection using Bayes factors and model likelihoods to determine the number of states and the number of autoregressive parameters in the volatility process. Bauwens et al. (2010) note that ML estimation of MS-GARCH model is basically impossible, because of the unobservable regimes. Therefore, they propose an MCMC algorithm that evades the problem of path dependence by treating the state variables as additional parameters. The authors carry out an extensive simulation study to evaluate the performance of the algorithm and then apply it to a sample of S&P500 daily returns. Based on the BIC they find that the MS-GARCH model with two regimes fits the data better than the MS-ARCH model. Next, Henneke et al. (2011) generalize the MS-GARCH model by including the ARMA process in the return evolution, resulting into a MS-ARMA–GARCH model. The authors design a MCMC scheme for estimation and compare their model with the one of Hamilton & Susmel (1994) by using the same data set and conclude

that full MS–ARMA–GARCH models outperform models such as of Hamilton & Susmel (1994). Bauwens et al. (2014) design a particle MCMC (PMCMC) method for estimation, called particle Gibbs sampler, which samples state variables jointly, rather than individually, as in Bauwens et al. (2010), and then sample the parameters given the states. The authors compare the performance of the MS-GARCH model with the change point GARCH (CP-GARCH), as in He & Maheu (2010), where the chain is not recurrent, differently than in Markov switching models. Bauwens et al. (2014) introduce an efficient method to compute marginal likelihoods, which was not feasible until then. The authors apply the two models - MS and CP - to several series of financial returns and find that MS-GARCH models with two regimes dominate CP-GARCH models. For some other financial returns, more regimes or breaks are necessary. However, MS-GARCH models are preferable in all cases. Finally, Billio et al. (2014) develop an efficient MCMC estimation approach for MS-GARCH model, which simultaneously generates the states from their joint distribution. The authors design a multiple-try sampling strategy, where a candidate path of the state variable is obtained by applying FFBS algorithm to an auxiliary MS-GARCH model. Billio et al. (2014) use the same data set as in Bauwens et al. (2014) and obtained results that are consistent with the ones in Bauwens et al. (2014).

Multivariate GARCH

Returns and volatilities depend on each other, so multivariate analysis is a more natural and useful approach. The starting point of multivariate volatility models is a univariate GARCH, thus the most simple MGARCH models can be viewed as direct generalizations of their univariate counterparts. Consider a multivariate return series $\{r_t\}_{t=1}^{T}$ of dimension K. Then

$$r_t = \mu_t + a_t = \mu_t + H_t^{1/2}\epsilon_t,$$

where $\mu_t = \mathrm{E}[r_t|\mathcal{I}^{t-1}]$, a_t are mean-corrected returns, ϵ_t is a random vector, such that $\mathrm{E}[\epsilon_t] = 0$ and $\mathrm{Cov}[\epsilon_t] = I_K$ and $H_t^{1/2}$ is a positive definite matrix of dimensions $K \times K$, such that H_t is the conditional covariance matrix of r_t, i.e., $\mathrm{Cov}[r_t|\mathcal{I}^{t-1}] = H_t^{1/2}\mathrm{Cov}[\epsilon_t](H_t^{1/2})' = H_t$. There is a wide range of MGARCH models, where most of them differ in specifying H_t. In the rest of this section we will review the most popular

of them and also the different Bayesian approaches to make inference and prediction. For general reviews on MGARCH models, see Bauwens et al. (2006), Silvennoinen & Teräsvirta (2009) and Tsay (2010) (Chapter 10), among others.

Regarding inference, one can also consider the same arguments provided in the univariate GARCH case above. Maximum likelihood estimation for MGARCH models can be obtained by using numerical optimization algorithms, such as Fisher scoring and Newton-Raphson. Vrontos et al. (2003*b*) have estimated several bivariate ARCH and GARCH models and found that some ML estimates of the parameters were quite different from their Bayesian counterparts. This was due to the non-Normality of the parameters. Thus, the authors suggest careful interpretation of the classical estimation approach. Also, Vrontos et al. (2003*b*) found it difficult to evaluate the classical estimates under the stationarity conditions, and consequently the resulting parameters, evaluated ignoring the stationarity constraints, produced non-stationary estimates. These difficulties can be overcome using the Bayesian approach.

VEC, DVEC and BEKK. The VEC model was proposed by Bollerslev et al. (1988), where every conditional variance and covariance (elements of the H_t matrix) is a function of all lagged conditional variances and covariances, as well as lagged squared mean-corrected returns and cross-products of returns. Using this unrestricted VEC formulation, the number of parameters increases dramatically. For example, if $K = 3$, the number of parameters to estimate will be 78, and if $K = 4$, the number of parameters increases to 210, see Bauwens et al. (2006) for the explicit formula for the number of parameters in VEC models. To overcome this difficulty, Bollerslev et al. (1988) simplified the VEC model by proposing a diagonal VEC model:

$$H_t = \Omega + A \odot (a_{t-1}a'_{t-1}) + B \odot H_{t-1},$$

where $\odot$ indicates the Hadamard product, Ω, A and B are symmetric $K \times K$ matrices. As noted in Bauwens et al. (2006), H_t is positive definite provided that Ω, A, B and the initial matrix H_0 are positive definite. However, these are quite strong restrictions on the parameters. Also, DVEC model does not allow for dynamic dependence between volatility series. In order to avoid such strong restrictions on the parameter matrices, Engle & Kroner (1995) propose the BEKK model, which is just a special case of VEC and,

consequently, less general. It has the attractive property that the conditional covariance matrices are positive definite by construction. The model looks as follows:

$$H_t = \Omega^* \Omega^{*'} + A^* (a_{t-1} a'_{t-1}) A^{*'} + B^* H_{t-1} B^{*'}, \tag{2.3}$$

where Ω^* is a lower triangular matrix and A^* and B^* are $K \times K$ matrices. In the BEKK model it is easy to impose the definite positiveness of the H_t matrix. However, the parameter matrices A^* and B^* do not have direct interpretations since they do not represent directly the size of the impact of the lagged values of volatilities and squared returns.

Osiewalski & Pipien (2004) present a paper that compares the performance of various bivariate ARCH and GARCH models, such as VEC, BEKK, etc., estimated using Bayesian techniques. As the authors observe, they are the first to perform model comparison using Bayes factors and posterior odds in the MGARCH setting. The algorithm used for parameter estimation and inference is Metropolis-Hastings, and to check for convergence they rely on cumsum statistics, introduced by Yu & Mykland (1998), and used by Bauwens & Lubrano (1998) in the univariate GARCH setting. Using the real data and assuming Student-t distribution for the mean-corrected returns, the authors found that BEKK models performed best, leaving VEC not so far behind. To sum up, the authors choose t-BEKK model as clearly better than the t-VEC, because it is relatively simple and has less parameters to estimate.

On the other hand, Hudson & Gerlach (2008) developed a prior distribution for a VEC specification that directly satisfies both necessary and sufficient conditions for positive definiteness and covariance stationarity, while remaining diffuse and non-informative over the allowable parameter space. These authors employed MCMC methods, including Metropolis-Hastings, to help enforce the conditions in this prior.

More recently, Burda & Maheu (2013) use the BEKK-GARCH model to show the usefulness of a new posterior sampler called the Adaptive Hamiltonian Monte Carlo (AHMC). Hamiltonian Monte Carlo (HMC) is a procedure to sample from complex distributions. The AHMC is an alternative inferential method based on HMC that is both fast and locally adaptive. The AHMC appears to work very well when the dimensions of the parameter space are very high. Model selection based on marginal likelihood is used

to show that full BEKK models are preferred to restricted diagonal specifications. Additionally, Burda (2013) suggests an approach called Constrained Hamiltonian Monte Carlo (CHMC) in order to deal with high-dimensional BEKK models with targeting, which allows for a parameter dimension reduction without compromising the model fit, unlike the diagonal BEKK. Model comparison of the full BEKK and the BEKK with targeting is performed indicating that the latter dominates the former in terms of marginal likelihood.

Factor-GARCH. Factor-GARCH was first proposed by Engle et al. (1990) to reduce the dimension of the multivariate model of interest using an accurate approximation of the multivariate volatility. The definition of the Factor-GARCH model, proposed by Lin (1992), says that BEKK model in Equation 2.3 is a Factor-GARCH, if A^* and B^* have rank one and the same left and right eigenvalues: $A^* = \alpha w\lambda'$, $B^* = \beta w\lambda'$, where α and β are scalars and w and λ are eigenvectors. Several variants of the factor model have been proposed. One of them is the full-factor multivariate GARCH by Vrontos et al. (2003*a*):

$$
\begin{aligned}
r_t &= \mu + a_t, \\
a_t &= WX_t, \\
X_t|\mathcal{I}^{t-1} &\sim \mathcal{N}_K(0, \Sigma_t),
\end{aligned}
$$

where μ is a $K \times 1$ vector of constants, which is time invariant, W is a $K \times K$ parameter matrix, X_t is a $K \times 1$ vector of factors and $\Sigma_t = \text{diag}(\sigma_{1t}^2, \ldots, \sigma_{Kt}^2)$ is a $K \times K$ diagonal variance matrix such that $\sigma_{it}^2 = c_i + b_i x_{i,t-1}^2 + g_i \sigma_{i,t-1}^2$, where σ_{it}^2 is the conditional variance of the ith factor at time t such that $c_i > 0$, $b_i \geq 0$, $g_i \geq 0$. Then, the factors in the X_t vector are GARCH(1,1) processes and the vector a_t is a linear combination of such factors. It can be easily shown that H_t is always positive definite by construction. However, the structure of H_t depends on the order of the time series in r_t. Vrontos et al. (2003*a*) have considered the problem of finding the best ordering under the proposed model. Furthermore, Vrontos et al. (2003*a*) investigate a full-factor MGARCH model using the ML and Bayesian approaches. The authors compute maximum likelihood estimates using Fisher scoring algorithm. As for the Bayesian analysis, the authors have adopted a Metropolis-Hastings algorithm, and found that the algorithm is very

time consuming, especially in high-dimensional data. To speed-up the convergence, Vrontos et al. (2003*a*) have proposed reparametrization of positive parameters and also a blocking sampling scheme, where the parameter vector is divided into three blocks: mean, variance and the matrix of constants W. As mentioned before, the ordering of the univariate time series in full-factor models is important, thus to select "the best" model one has to consider $K!$ possibilities for a multivariate dataset of dimension K. Instead of choosing one model and making inference (as if the selected model was the true one), the authors employ a Bayesian approach by calculating the posterior probabilities for all competing models and model averaging to provide "combined" predictions. The main contribution of this paper is that the authors were able to carry out an extensive Bayesian analysis of a full-factor MGARCH model considering not only parameter uncertainty, but model uncertainty as well.

As already discussed above, a very common stylized feature of financial time series is the asymmetric volatility. Dellaportas & Vrontos (2007) have proposed a new class of tree structured MGARCH models that explore the asymmetric volatility effect. Same as the paper by Vrontos et al. (2003*a*), the authors consider not only parameter-related uncertainty, but also uncertainty corresponding to model selection. Thus in this case the Bayesian approach becomes particularly useful because an alternative method based on maximizing the pseudo-likelihood is only able to work after selecting a single model. The authors develop an MCMC stochastic search algorithm that generates candidate tree structures and their posterior probabilities. The proposed algorithm converged fast. Such modeling and inference approach leads to more reliable and more informative results concerning model-selection and individual parameter inference.

There are more models that are nested in BEKK, such as the Orthogonal GARCH for example, see Alexander & Chibumba (1997) and Van der Weide (2002), among others. All of them fall into the class of direct generalizations of univariate GARCH or linear combinations of univariate GARCH models. Another class of models are the nonlinear combinations of univariate GARCH models, such as constant conditional correlation (CCC), dynamic condition correlation (DCC), general dynamic covariance (GDC) and Copula-GARCH models. A very recent alternative approach that also considers Bayesian estimation can be found in Jin & Maheu (2013) who proposes a new dynamic

component models of returns and realized covariance (RCOV) matrices based on time-varying Wishart distributions. In particular, Bayesian estimation and model comparison is conducted with the existing range of multivariate GARCH models and RCOV models.

CCC. The CCC model, proposed by Bollerslev (1990) and the simplest in its class, is based on the decomposition of the conditional covariance matrix into conditional standard deviations and correlations. Then, the conditional covariance matrix H_t looks as follows:

$$H_t = D_t R D_t,$$

where D_t is diagonal matrix with the K conditional standard deviations and R is a time-invariant conditional correlation matrix such that $R = (\rho_{ij})$ and $\rho_{ij} = 1, \forall i = j$. The CCC approach can be applied to a wide range of univariate GARCH family models, such as exponential GARCH or GJR-GARCH, for example.

Vrontos et al. (2003*b*) have estimated some real data using a variety of bivariate ARCH and GARCH models in order to select the best model specification and to compare the Bayesian parameter estimates to those of the ML. The authors have considered three ARCH and three GARCH models, all of them with constant conditional correlations (CCC). They have used a Metropolis-Hastings algorithm, which allows to simulate from the joint posterior distribution of the parameters. For model comparison and selection, Vrontos et al. (2003*b*) have obtained predictive distributions and assessed comparative validity of the analyzed models, according to which the CCC model with diagonal covariance matrix performed the best considering one-step-ahead predictions.

DCC. A natural extension of the simple CCC model are the dynamic conditional correlation (DCC) models, firstly proposed by Tse & Tsui (2002) and Engle (2002*a*). The DCC approach is more realistic, because the dependence between returns is likely to be time-varying.

The models proposed by Tse & Tsui (2002) and Engle (2002*a*) consider that the conditional covariance matrix H_t looks as

$$H_t = D_t R_t D_t,$$

where R_t is now a time-varying correlation matrix at time t. The models differ in the specification of R_t. In the paper by Tse & Tsui (2002), the conditional correlation matrix is

$$R_t = (1 - \theta_1 - \theta_2)R + \theta_1 R_{t-1} + \theta_2 \Psi_{t-1},$$

where θ_1 and θ_2 are non-negative scalar parameters, such that $\theta_1 + \theta_2 < 1$, R is a positive definite matrix such that $\rho_{ii} = 1$ and Ψ_{t-1} is a $K \times K$ sample correlation matrix of the past m standardized mean-corrected returns $u_t = D_t^{-1} a_t$.

On the other hand, in the paper by Engle (2002*a*), the specification of R_t is

$$R_t = (I \odot Q_t)^{-1/2} Q_t (I \odot Q_t)^{-1/2},$$

where

$$Q_t = (1 - \alpha - \beta)\bar{Q} + \alpha(u_{t-1}u'_{t-1}) + \beta Q_{t-1}.$$

$u_{i,t} = a_{i,t}/\sqrt{h_{ii,t}}$ is a mean-corrected standardized returns, α and β are non-negative scalar parameters, such that $\alpha + \beta < 1$ and $\bar{Q}$ is unconditional covariance matrix of u_t. As noted in Bauwens et al. (2006), the model by Engle (2002*a*) does not formulate the conditional correlation as a weighted sum of past correlations, unlike in the DCC model by Tse & Tsui (2002), seen above. The drawback of both these models is that θ_1, θ_2, α and β are scalar parameters, so all conditional correlations have the same dynamics. However, as Tsay (2010) notes it, the models are parsimonious.

Moreover, as financial returns display not only asymmetric volatility, but also excess kurtosis, previous research, as in univariate case, has mostly considered using a multivariate Student-t distribution for the errors. However, as already discussed above, this approach has several limitations. Galeano & Ausín (2010) propose a DCC model, where the standardized innovations follow a mixture of Gaussian distributions. This allows to capture long tails without being limited by the degrees of freedom constraint,

which is necessary to impose in the Student-t distribution so that the higher moments could exist. The authors estimate the proposed model using the classical MLE and Bayesian approaches. In order to estimate the parameters of the dynamics of individual assets and dynamic correlations, and the parameters of the Gaussian mixture, Galeano & Ausín (2010) have relied on RWMH algorithm. BIC criteria was used for selecting the number of mixture components, which performed well in simulated data. Using real data, the authors provide an application to calculating the Value at Risk (VaR) and solving a portfolio selection problem. MLE and Bayesian approaches have performed similarly in point estimation, however, the Bayesian approach, besides giving just point estimates, allows the derivation of predictive distributions for the portfolio VaR.

An extension of the DCC model of Engle (2002*a*) is the Asymmetric DCC also proposed by Engle (2002*a*), which incorporates an asymmetric correlation effect. It means that correlations between asset returns might be higher after a negative return than after a positive one of the same size. Cappiello et al. (2006) generalizes the ADCC model into the AGDCC model, where the parameters of the correlation equation are vectors, and not scalars. This allows for asset-specific correlation dynamics. In the AGDCC model, the Q_t matrix in the DCC model is replaced with:

$$Q_t = S(1 - \bar{\kappa}^2 - \bar{\lambda}^2 - \bar{\delta}^2/2) + \kappa\kappa' \odot u'_{t-1}u_{t-1} + \lambda\lambda' \odot Q_{t-1} + \delta\delta' \odot \eta'_{t-1}\eta_{t-1},$$

where $u_t = D_t^{-1}a_t$ are mean corrected standardized returns, $\eta_t = u_t \odot I(u_t < 0)$ selects just negative returns, “diag" stands for either taking just the diagonal elements from the matrix, or making a diagonal matrix from a vector, S is a sample correlation matrix of u_t, κ, λ and δ are $K \times 1$ vectors, $\bar{\kappa} = K^{-1}\sum_{i=1}^{K}\kappa_i$, $\bar{\lambda} = K^{-1}\sum_{i=1}^{K}\lambda_i$ and $\bar{\delta} = K^{-1}\sum_{i=1}^{K}\delta_i$. To ensure positivity and stationarity of Q_t, it is necessary to impose $\kappa_i, \lambda_i, \delta_i > 0$ and $\kappa_i^2 + \lambda_i^2 + \delta_i^2/2 < 1$, $\forall i = 1, \ldots, K$. The ADCC by Engle (2002*a*) is just a special case where $\kappa_1 = \ldots = \kappa_K$, $\lambda_1 = \ldots = \lambda_K$ and $\delta_1 = \ldots = \delta_K$.

Copula-GARCH. The use of copulas is an alternative approach to study dependencies between individual returns and their volatilities. The main convenience of using copulas is that individual marginal densities of the returns can be defined separately from their dependence structure. Then, each marginal time series can be modeled using univariate specification and the dependence between the returns can be modeled

by selecting an appropriate copula function. A K-dimensional copula $C(u_1, \ldots, u_K)$, is a multivariate distribution function in the unit hypercube $[0,1]^K$, with Uniform $[0,1]$ marginal distributions. Under certain conditions, the Sklar Theorem (Sklar 1959) affirms that, every joint distribution $F(x_1, \ldots, x_K)$, whose marginals are given by $F_1(x_1), \ldots, F_K(x_K)$, can be written as

$$F(x_1, \ldots, x_K) = C(F_1(x_1), \ldots, F_K(x_K)),$$

where C is a copula function of F, which is unique if the marginal distributions are continuous.

The most popular approach for modeling time-varying volatilities through copulas is called the Copula-GARCH model, where univariate GARCH models are specified for each marginal series and the dependence structure between them is described using a copula function. A very useful feature of copulas, as noted by Patton (2009), is that the marginal distributions of each random variable do not need to be similar to each other. This is very important in modeling return time series, because each of them might be following different distributions. The choice of copulas can vary from a simple Gaussian copula to more flexible ones, such as Clayton, Gumbel, mixed Gaussian, etc. In the existing literature different parametric and non-parametric specifications can be used for the marginals and copula function C. Also, the copula function can be assumed to be constant or time-varying, as seen in Ausín & Lopes (2010), among others.

The estimation for Copula-GARCH models can be performed in a variety of ways. Maximum likelihood is the obvious choice for fully parametric models. Estimation is generally based on a multi-stage method, where firstly the parameters of the marginal univariate distributions are estimated and then used to condition in estimating the parameters of the copula. Another approach is non- or semi-parametric estimation of the univariate marginal distributions followed by a parametric estimation of the copula parameters. As Patton (2006) has showed, the two-stage maximum likelihood approach lead to consistent, but not efficient estimators.

An alternative is to employ a Bayesian approach, as done by Ausín & Lopes (2010). The

authors have developed a one-step Bayesian procedure, where all parameters are estimated at the same time using the entire likelihood function and provided the methodology for obtaining optimal portfolio, calculating VaR and CVaR. Ausín & Lopes (2010) have used a Gibbs sampler to sample from a joint posterior, where each parameter is updated using a RWMH. In order to reduce computational cost, the model and copula parameters are updated not one-by-one, but rather by blocks, that consist of highly correlated vectors of model parameters.

Arakelian & Dellaportas (2012) have also used Bayesian inference for Copula-GARCH models. These authors have proposed a methodology for modeling dynamic dependence structure by allowing copula functions or copula parameters to change across time. The idea is to use a threshold approach so these changes, that are assumed to be unknown, do not evolve in time but occur in distinct points. These authors have also employed a RWMH for parameter estimation together with a Laplace approximation. The adoption of an MCMC algorithm allows the choice of different copula functions and/or different parameter values between two time thresholds. Bayesian model averaging is considered for predicting dependence measures such as the Kendall's correlation. They conclude that the new model performs well and offers a good insight into the time-varying dependencies between the financial returns.

Hofmann & Czado (2010) developed Bayesian inference of a multivariate GARCH model where the dependence is introduced by a D-vine copula on the innovations. A D-vine copula is a special case of vine copulas which are very flexible to construct multivariate copulas because it allows to model dependency between pairs of margins individually. Inference is carried out using a two-step MCMC method closely related with the usual two-step maximum likelihood procedure for estimating Copula-GARCH models. The authors then focus on estimating VaR of a portfolio that shows asymmetric dependencies between some pairs of assets and symmetric dependency between others.

An alternative approach to the previous parametric GARCH specifications is the use of Bayesian non-parametric methods, that allow to model the errors as an infinite mixture of Normals, as seen in Ausín et al. (2014) and Jensen & Maheu (2013). The Bayesian non-parametric approach for time-varying volatility models will be discussed in detail in Section 2.3.

To sum up, considering the amount of articles published quite recently regarding the topic of estimating univariate and multivariate GARCH models using MCMC methods indicates still growing interest in the area. Although numerous GARCH-family models have been investigated using different MCMC algorithms, there are still a lot of areas that need further research and development.

2.2 Univariate and multivariate SV

SV models are a closely related to GARCH and are also used to model time-varying volatility. SV models express the logarithm of volatility as dependent on the past volatilities and an error term, thus making volatility not deterministic anymore.

Univariate SV

The basic autoregressive SV(1) (ARSV) model for regularly spaced data looks as follows:

$$a_t = \exp\{h_t/2\}\epsilon_t, \tag{2.4}$$

$$h_t = \omega + \alpha h_{t-1} + \sigma_\eta \eta_t, \;\; t = 1, \ldots, T, \tag{2.5}$$

where a_t is the mean corrected return of the asset at time t, ϵ_t and η_t are uncorrelated standard white noise shocks, $\log h_t$ is log-volatility, which is a stationary process, provided that the absolute value of α, which is also called the persistence parameter, is $\alpha < |1|$ and σ_η is the standard deviation of the shock to $\log h_t$.

There has been a plentiful amount of research on this model and extensions. For a review on the properties of SV models see Taylor (1994) and Shephard (1996), for example. Instead of considering Gaussian errors, some authors investigate heavy-tailed distributions, or correlated errors to include the asymmetry effect, see e.g. Harvey & Shephard (1996) and Jacquier et al. (2004), among others. Shephard & Andersen (2009) overview the origins of the SV models and Broto & Ruiz (2004) discuss in detail the estimation methods for the SV models.

There has been some discussion concerning the comparison of GARCH and SV models. A number of papers have provided empirical evidence of better fit of SV rather than GARCH models, however, as commented in the next section, SV models are harder to estimate, which is a serious drawback concerning a choice of a model, see e.g. Kim et al. (1998), Ghysels et al. (1996), Shephard (1996) and Taylor (1994).

The main estimation methods for SV models include Method of Moments (MM), Generalized MM (GMM), Efficient Method of Moments (EMM) and Quasi-Maximum Likelihood approach, among others. The use of Sequential Monte Carlo (SMC) and MCMC algorithms for inference and prediction in the context of SV models is more recent: Geweke (1994) was the first to apply importance sampling algorithm to SV models. Algorithms developed by Jacquier et al. (1994), Kim et al. (1998) and Shephard (1993) have been the basis for numerous subsequent papers, see Broto & Ruiz (2004) for a detailed review.

The paper by Jacquier et al. (1994) is one of the first articles to propose a new Bayesian approach for inference and prediction for SV models, which allows to conduct finite sample inference and calculate predictive distributions (as opposed to previously dominating estimation procedures, where one had to rely on asymptotic approximations for inference and the uncertainty of forecasted variability was not accounted for at all). The joint posterior of interest is given by the Bayes theorem:

$$\pi(h, \theta|y) \propto p(y|h) \times p(h|\theta) \times p(\theta),$$

where θ is the vector of parameters $\theta = (\omega, \alpha, \sigma_\eta)$, $h = (h_1, \ldots, h_T)$ is the vector of volatilities and $y = (y_1, \ldots, y_T)$ is the vector of returns, as seen in the basic ARSV(1) formulation, Equation 2.4 and Equation 2.5. The authors use a cyclic independence Metropolis chain. Instead of sampling directly from $p(h|\theta, y)$, they sample indirectly from $p(h_t|h_{-t}, \theta, y)$, where h_{-t} is the rest of the vector h, except for h_t. The empirical analysis using real data sets revealed that almost all marginal posterior distributions were skewed, which is a strong evidence against the usage of Gaussian approximations. The authors also found out that the Method of Moments produced very different estimates. Thus, they investigated the sampling properties of Bayes, MM and QML-Kalman filtering estimators and concluded that Bayesian approach outperforms the other two.

Another very important paper is by Kim et al. (1998), which has been later cited in many subsequent papers, see Omori et al. (2007), Chib et al. (2002), and their developed algorithm, referred as KSC (named after Kim, Shephard and Chib), has been extended later in a numerous ways. The authors first use a simple Gibbs sampling algorithm, which proves to be inefficient and converge very slowly. Then, they develop another method, that samples unobserved volatilities using an offset mixture of seven Normal distributions to accurately approximate the exact likelihood, followed by an importance re-weighting procedure. Using simulated data, the authors conclude that their proposed method is significantly more efficient than previously suggested methods for estimating stochastic volatility, as the one proposed by Jacquier et al. (1994).

The first work to model correlated errors in order to include the asymmetric volatility effect using likelihood-based inference is developed by Jacquier et al. (2004). The authors make use of Bayes Factors to provide justification for using fat-tailed asymmetric SV model. The basic ARSV(1) model given in Equation 2.4 and Equation 2.5 is extended by defining ρ as the correlation between errors (ϵ_t, η_t), and assuming that the marginal distribution of ϵ_t is Student-t to incorporate fat tails. The priors proposed by the authors are Normal-Gamma for the $\theta = (\omega, \alpha, \sigma_\eta)$, an Inverse-Gamma for the correlation ρ and a discrete Uniform prior for the degrees of freedom ν of a Student-t distribution. The authors combine rejection and Metropolis-Hastings algorithms, finding strong evidence for fat tails and asymmetry effect. Finally, they investigate the sampling properties of their proposed MCMC algorithm and the convergence of the parameters, concluding that the algorithm is reliable and fast.

A paper by Omori et al. (2007) extends Kim et al. (1998) approach by approximating the joint distribution of the outcome and volatility innovations by ten-component mixture of bivariate gaussian distributions, followed by a re-weighting procedure. In this manner, the authors are able to extend the previous model into the SV model with leverage, and also include the heavy-tailed feature of the returns. They show that the new algorithm performs as well as the one developed by Kim et al. (1998), and is applicable under wider conditions.

Markov Switching SV (MSSV). The motivation of introducing Markov switching jumps in the volatility process of the SV models is the same as in the case of GARCH

models, discussed above. MSSV model, first introduced by So et al. (1998), is of the following form:

$$a_t = \exp\{h_t/2\}\epsilon_t,$$
$$h_t = \alpha_{s_t} + \beta h_{t-1} + \tau\eta_t, \ \ \eta_t \sim \mathcal{N}(0,1),$$

where s_t is a state variable, defined the same as in MS-GARCH above. Differently than in MS-GARCH models, here only parameter α displays different regimes. The authors propose a MCMC and the data-augmentation methodology, where they use Gibbs sampler to generate samples from the joint posterior distribution of the unknown parameters and the latent variables. Kalimipalli & Susmel (2004) have proposed a two-factor SV model with regime switching and estimated it using Gibbs sampler. They find that the high volatility persistence is reduced when the regimes are incorporated in the model. Also, the authors compare the new model with other two alternative two-factor models, simple SV and GARCH, and find that SV always outperforms GARCH, both in sample and out of sample. The regime switching SV performs better than the simple SV in sample, however, out of sample, it is only marginally better. Lopes & 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 & Lopes (2007) have constructed a sequential Monte Carlo filter by combining auxiliary particle filter (APF) with the filter of Liu & 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.

Multivariate SV

As for the multivariate case, the basic setting for a MSV model, proposed by Harvey et al. (1994), is given by

$$a_t = H_t^{1/2}\epsilon_t \tag{2.6}$$

$$H_t^{1/2} = \text{diag}\left(\exp\{h_{i,t}/2\}\right),\ i = 1,\ldots,K,\ t = 1,\ldots,T \tag{2.7}$$

$$h_{t+1} = \omega + \alpha \odot h_t + \eta_t \tag{2.8}$$

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} | h_t \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{\epsilon\epsilon} & 0 \\ 0 & \Sigma_{\eta\eta} \end{pmatrix}\right], \tag{2.9}$$

where $a_t = (a_{1,t},\ldots,a_{K,t})$ is a vector of mean-corrected K assets returns at time t, $h_t = (h_{1,t},\ldots,h_{K,t})$ is a vector of unobserved log-volatilities, ω and α are $K \times 1$ parameters vectors, $\Sigma_{\eta\eta} = \sigma_{\eta,ij}$ is a positive definite covariance matrix and $\Sigma_{\epsilon\epsilon} = \rho_{ij}$ is the correlation matrix such that $\rho_{ij} = 1$, $\forall i = j$ and $V[y_t|h_t] = H_t^{1/2}\Sigma_{\epsilon\epsilon}H_t^{1/2}$. There has been other proposals of similar basic MSV models by So et al. (1997), Daníelsson (1998) and Smith & Pitts (2006), among others. For an extensive review of Multivariate SV models, see e.g. Asai et al. (2006) and Chib et al. (2009).

Same as in GARCH models, the SV model can be augmented in order to incorporate the asymmetric volatility, or the so-called leverage effect. This can be done by letting the errors to be correlated. Daníelsson (1998) and Chan et al. (2006) consider the correlation between ϵ_t and η_{t-1}, which means that there is not time lag between the shock to the return and the volatility shock. As Yu (2005) discussed it in the univariate setting, this is not a correct way to introduce the leverage effect, because such correlations do not have clear interpretations and the volatility needs time to react to the shock. Thus, the correlation between ϵ_t and η_t makes much more sense. Asai & McAleer (2006) introduced a MSV model with leverage effect where a covariance matrix, L, between ϵ_t and η_t is defined such that $L = \text{diag}\{\lambda_1\sigma_{v,11},\ldots,\lambda_N\sigma_{v,KK}\}$, where the parameter λ_i is expected to be negative.

Factor models. There are two kinds of factor models, as discussed by Asai et al. (2006). The first one is called Additive Factor Models, proposed by Harvey et al. (1994) and extended by Jacquier et al. (1995). The basic idea of this model is that the mean-corrected returns are decomposed into two parts: the first one has a smaller number of factors to capture the information relevant to all the assets, and the second component is idiosyncratic noise,

$$a_t = Df_t + e_t \tag{2.10}$$

$$f_{it} = \exp\{h_{it}/2\}\,\epsilon_{it} \tag{2.11}$$

$$h_{i,t+1} = \omega_i + \alpha_i h_{it} + \eta_{it}, \tag{2.12}$$

where f_t is a $B \times 1$ vector of factors, such that $(B < K)$, where K is a number of assets, D is a $K \times B$ matrix of factors loadings, $e_t \sim N(0, \text{diag}\{\sigma_1^2, \ldots, \sigma_K^2\})$, $\epsilon_{it} \sim N(0,1)$ and $\eta_{it} \sim N(0, \sigma_\eta^2)$. The covariance matrix of a_t is always positive definite.

On the other hand, the multiplicative factor model, also called stochastic discount factor model, was proposed by Quintana & West (1987), where the returns are decomposed into two multiplicative components. However, unlike the additive factor model, the correlations are time-invariant, which is quite a strong restriction.

Time-varying correlation models. To allow for time-varying correlations, firstly the assumption of constant correlation in Equation 2.9 has to be relaxed, such that $\Sigma_{\epsilon\epsilon,t} = \rho_{ij,t}$. Asai & McAleer (2009) proposed a MSV model based on Wishart distribution. For more details, see Asai et al. (2006) and the original paper by Asai & McAleer (2009).

To sum up, there exists a wide variety of MSV models, including alternative specifications, such as based on the matrix exponential transformation, Cholesky decomposition, etc. Concerning the choice of a MSV model, see Yu & Meyer (2006), who have estimated and compared nine MSV models and found strong evidence in favor of asymmetric models, allowing for time-varying correlations.

As seen above, the use of MCMC methods in SV context and beyond is quite recent and developing fast. One can be referred to a survey by Chib (2001) on MCMC methods in a general context. Even though earlier MCMC estimation of SV models was very computationally demanding, nowadays it can be easily implemented using basic software, such as BUGS (Bayesian analysis using Gibbs sampling), as demonstrated by Meyer & Yu (2000).

All the previously introduced methods rely on parametric assumptions for the distribution of the errors. However, imposing a certain distribution can be rather restrictive.

Bayesian non-parametric methods become especially useful, since they do not impose any specific distribution on the standardized returns.

2.3 Dirichlet Process Mixture

Bayesian non-parametrics is an alternative approach to the classical parametric Bayesian statistics, where one usually gives some prior for the parameters of interest, whose distribution is unknown, and then observes the data and calculates the posterior. The priors come from the family of parametric distributions. Bayesian non-parametrics uses a prior over distributions with the support being the space of all distributions. Then, it can be viewed as a distribution over distributions.

One of the most popular Bayesian non-parametric modeling approach is based on Dirichlet processes (DP) and mixtures of Dirichlet processes (DPM), see Ferguson (1973), Antoniak (1974) and Lo (1984) among others. Suppose that we have a sequence of exchangeably distributed random variables $X_1, X_2, \ldots$ from an unknown distribution F, where the support for X_i is Ω. In order to perform Bayesian inference, we need to define the prior for F. This can be done by considering partitions of Ω, such as $\Omega = C_1 \cup C_2 \cup \ldots \cup C_m$, and defining priors over all possible partitions. We say that F has a Dirichlet process prior, denoted as

$$F \sim \mathcal{DP}(\alpha, F_0),$$

if the set of associated probabilities given F for any partition follows a Dirichlet distribution,

$$\{F(C_1), \ldots, F(C_m)\} \sim \text{Dirichlet } (\alpha F_0(C_1), \ldots, \alpha F_0(C_m)),$$

where $\alpha > 0$ is a precision parameter that represents our prior certainty of how concentrated the distribution is around F_0, which is a known base distribution on Ω. The Dirichlet process is a conjugate prior. Thus, given n independent and identically distributed samples from F, the posterior distribution of F is also a Dirichlet process such that

$$F \sim \mathcal{DP}\left(\alpha + n, \frac{\alpha F_0 + nF_n}{\alpha + n}\right),$$

where F_n is the empirical distribution function of $X_1, \ldots, X_n$.

There are two main ways for generating a sample from the marginal distribution of X, where $X|F \sim F$ and $F \sim \mathcal{DP}(\alpha, F_0)$: the Polya urn and stick breaking procedures. On the one hand, the Polya urn scheme can be illustrated in terms of a urn with α black balls; when a non-black ball is drawn, it is placed back in the urn together with another ball of the same color. If the drawn ball is black, a new color is generated from F_0 and a ball of this new color is added to the urn together with the black ball we drew. This process gives a discrete marginal distribution for X since there is always a probability that a previously seen value is repeated. Regarding inference algorithms, the marginal methods, such as those proposed by Escobar & West (1995), MacEachern (1994) and Neal (2000), are based on integrating out the infinite dimensional part of the model.

On the other hand, the stick-breaking procedure is based on the representation of the random distribution F as a countably infinite mixture:

$$F = \sum_{m=1}^{\infty} \omega_m \delta_{X_m},$$

where δ_X is a Dirac measure, $X_m \sim F_0$ and the weights are such that $\omega_1 = \beta_1$, $\omega_m = \beta_m \prod_{i=1}^{m-1}(1-\beta_i)$, for $m = 1, \ldots,$ where $\beta_m \sim \text{Beta}(1, \alpha)$. This implies that the weights $\omega \to \text{Dirichlet}(\alpha/K, \ldots, \alpha/K)$ as $K \to \infty$. This class of algorithms, called conditional methods, leave the infinite part in the model and sample a finite number of variables. These include procedures by Ishwaran & James (2001), Walker (2007), Papaspiliopoulos & Roberts (2008), Papaspiliopoulos (2008) and Kalli et al. (2011).

The discreteness of the Dirichlet process is clearly a disadvantage in practice. A solution was proposed by Antoniak (1974) by using DPM models where a DP prior is imposed over the distribution of the model parameters, θ, as follows:

$$\begin{aligned} X_i|\theta_i &\sim F(X|\theta_i), \\ \theta_i|G &\sim G(\theta), \\ G|\alpha, G_0 &\sim \mathcal{DP}(\alpha, G_0). \end{aligned}$$

Observe that G is a random distribution drawn from the DP and because it is discrete, multiple θ_i's can take the same value simultaneously, making it a mixture model. In fact, using the stick-breaking representation, the hierarchical model above can be seen as an infinite mixture of distributions:

$$f(X|\theta,\omega) = \sum_{m=1}^{\infty} \omega_m k(X|\theta_m),$$

where the weights are obtained as before: $\omega_1 = \beta_1$, $\omega_m = \beta_m \prod_{i=1}^{m-1}(1-\beta_i)$, for $m = 1, \ldots$, and where $\beta_m \sim \text{Beta}\,(1,\alpha)$ and $\theta_m \sim G_0$ and k is some density kernel with parameters θ. For more on DPM see the discussion in Section 3.1.1.

Bayesian non-parametric models in econometrics. The use of Bayesian non - parametric methods in econometric modeling is relatively recent with increasing number of papers, starting after the seminal work of Hirano (2002). The author proposes a semi-parametric random effects autoregressive model for dynamic panel data, where the error term is modeled using DPM models. Rodriguez & ter Horst (2008) develop statistical methods to estimate and predict time-varying densities, which can be viewed as an extension of the DPM models to a sequence of distributions that evolve in discrete time. The authors present an illustration that estimates the distribution of travel reimbursement claims, noting, that the proposed methodology can be adapted to a wide range of econometric models in areas such as insurance, time-varying volatilities, risk management, actuarial science, epidemiology, climatology etc. Taddy & Kottas (2009) have presented a general framework for a semi - parametric hidden Markov switching regression, where the model formulation involves a finite mixture of conditionally independent DPMs, with a Markov chain for the mixing distribution. The proposed methodology has been illustrated to a problem from fisheries research, that investigates stock - recruitment data under shifts in the ecosystem state. Taddy (2010) proposes a dependent stick-breaking mixture model with marginal DP priors, that allows a non - parametric density to be time-varying. The introduced methodology is applied to tracking weekly maps of violent crime events in Cincinnati, and can be adapted to a wide variety of alternative settings in a straightforward manner. Bassetti et al. (2014) propose two classes of prior distributions, called beta-product dependent

Pitman-Yor (DPY) processes, which extend the independent Dirichlet priors. The authors present an application to a multi-country macroeconomic data set of industrial production indexes growth rate series for the EU and US and find that the results favor DPY against the independent Dirichlet prior. Jochmann (2014) have proposed a Bayesian non-parametric approach, in particular sticky infinite hidden Markov model, to explain U.S. inflation dynamics. This approach is capable of incorporating unknown number of breakpoints in the time series process. Burda et al. (2014) have analyzed unemployment duration data from the US Department of Labor by considering a new specification for the competing risk model, where unobserved individual heterogeneity is modeled via DPM.

The above papers illustrate the flexibility of the DPM approach in econometric modeling, since it can be applied in various settings and in diverse areas of research. Next, we briefly review the latest developments of Bayesian non-parametrics in time-varying volatility modeling.

2.3.1 Volatility modeling using DPM

As mentioned above, modeling time-varying volatilities with non-parametric errors is quite a recent topic of research with increasing popularity due to its flexibility. Ausín et al. (2014) propose models for univariate GARCH, Jensen & Maheu (2013) for MGARCH, Jensen & Maheu (2010, 2014) and Delatola & Griffin (2011, 2013) for univariate SV.

Non-parametric GARCH-type models. Ausín et al. (2014) have applied semi-parametric Bayesian techniques to estimate univariate GARCH-type models. These authors have used the class of scale mixtures of Gaussian distributions, that allow for the variances to change over components, with a Dirichlet process prior on the mixing distribution to model innovations of the GARCH process. The resulting class of models is called DPM-GARCH type models. In order to perform Bayesian inference on the new model, the authors employ a stick-breaking sampling scheme and make use of the ideas proposed in Walker (2007), Papaspiliopoulos & Roberts (2008) and Papaspiliopoulos (2008). The new scale mixture model was compared to a simpler mixture of two Gaussians, Student-t and the usual Gaussian models. The estimation

results in all three cases were quite similar, however, the scale mixture model is able to capture skewness as well as kurtosis and, based on the approximated Log Marginal Likelihood (LML) and DIC, provided the best performance in simulated and real data. Finally, Ausín et al. (2014) have applied the resulting model to perform one-step-ahead predictions for volatilities and VaR. In general, the non-parametric approach leads to wider Bayesian credible intervals and can better describe long tails.

Jensen & Maheu (2013) propose a Bayesian non-parametric modeling approach for the innovations in MGARCH models. They use a MGARCH specification, proposed by Ding & Engle (2001), which is a different representation of a well known DVEC model, introduced above. The innovations are modeled as an infinite mixture of multivariate Normals with a DP prior. The authors have employed Polya urn and stick-breaking schemes and, using two data sets, compared the three model specifications: parametric MGARCH with Student-t innovations (MGARCH-t), GARCH-DPM-Λ that allows for different covariances (scale mixture) and MGARCH-DPM, allowing for different means and covariances of each component (location-scale mixture). In general, both semi-parametric models produced wider density intervals. However, in MGARCH-t model a single degree of freedom parameter determines the tail thickness in all directions of the density, meanwhile the non-parametric models are able to capture various deviations from Normality by using a certain number of components. These results are consistent with the findings of Ausín et al. (2014) in univariate setting. As for predictions, both semi-parametric models performed equally good and outperformed the parametric MGARCH-t specification.

Non-parametric SV-type models. Jensen & Maheu (2010) construct an MCMC sampler for their proposed SV-DPM model, where latent volatilities are sampled via random length block sample, 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 & 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 & 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 forward filtering backward sampling (FFBS). The rest of the model parameters are sampled via an extension of Gibbs sampler, called hybrid Gibbs sampler. In their subsequent paper, Delatola & Griffin (2013) consider and asymmetric SV model. Same as before, they make use of the linearization and update the latent log-volatilities by FFBS. The rest of the parameters are sampled via Metropolis-Hastings. All above MCMC schemes are costly in the context of SV models where we consider daily return data.

To sum up, the findings in the above papers are consistent: the Bayesian semi-parametric approach leads to more flexible models and is better in explaining heavy-tailed return distributions, which parametric models cannot fully capture. This provides a more adequate measure of uncertainty. However, the question if non-parametric errors can incorporate model-specific asymmetries or Markov switching jumps in volatilities is still open.

2.4 Sequential Monte Carlo

As seen from the previous sections, MCMC has been a dominant Bayesian estimation approach for time-varying volatility models. Meanwhile MCMC estimation of GARCH-type models most of the time does not present problems of elevated computational cost and auto-correlated draws, the estimation of SV-type models is rather more complicated. SV models, as compared to GARCH, include an extra level of complexity by considering non-deterministic volatility evolution. In other words, SV is a state-space model, where apart from parameter estimation one also has to consider filtering of the latent states. MCMC estimation can become computationally costly, since inside the chain some volatility filter has to be considered as well. This presents unbearable computational burden whenever a new observation arrives and the chain has to be re-run all over again.

An alternative approach to MCMC is to rely on Sequential Monte Carlo techniques, or particle filters (PF), for state filtering and parameter learning. For thorough reviews of particle methods in general, see Lopes & Tsay (2011) and Lopes & Carvalho (2013). Basically, there are a lot of filers in the literature that are able to recover the unobserved

states, however, sequential parameter learning is a particularly difficult problem. Three of the most popular approaches for sequential state filtering and parameter learning include filters by Liu & West (2001), Storvik (2002) and a Particle Learning filter by Carvalho, Johannes, Lopes & Polson (2010).

Using PF for SV-type models is quite a new field of research and recent works include Carvalho & Lopes (2007), Lopes & Polson (2010) and Rios & Lopes (2013), among others. The main concern of using PF is particle degeneracy, which means that after certain time the variability among particles decreases and they collapse into one point. The most recent PF method, called Particle Learning (PL), avoids this problem by relying on resample-propagate type algorithm and by using a set of sufficient statistics to be tracked in time, which is sufficient for learning about the parameters. Next, we present in short the main idea behind PL.

Consider a general state space model for the returns defined by the observation and evolution equations:

$$
\begin{aligned}
r_t &\sim p(r_t|h_t, \Theta), \\
h_t &\sim p(h_t|h_{t-1}, \Theta).
\end{aligned}
$$

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

$$
\begin{aligned}
&p(S_t|r^{t+1}) \propto p(r_{t+1}|S_t)p(S_t|r^t) : \\
&\text{Resample } \; p(S_t|r^t) \; \text{ with weights } \; p(r_{t+1}|S_t),
\end{aligned}
$$

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

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: Θ_0 of dimension $K \times N$ (N is the number of particles and

K is the number of model parameters) and h_0 of dimension $1 \times N$ (in case of a single state variable). Also, an essential state vector S_0 is constructed, containing all the hyper-parameters for the parameters of the model and the volatility states. Then, PL iterates through three steps, for each particle (i), for $i = 1, \ldots, N$:

1. **Resampling**.

   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. **Propagation.**

   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. The hyper-parameters of the distributions of the static parameters Θ can be updated via deterministic rules. The latent states are propagated by sampling from $p(h_{t+1}|\tilde{S}_t, r_{t+1})$.

3. **Learning.**

   Learn about the parameters on-line 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, Lopes, Polson & Taddy (2010) presented a detailed explanation of PL methods for general mixtures, including DPM models. Call $n_{t,j}$ 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. **Resampling.**
   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. **Propagation**.

   (a) Propagate allocation variable $k_{t+1}^{(i)} \sim p(k_{t+1}|\tilde{S}_t^{(i)}, y_{t+1})$.

   (b) Propagate 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 $n_{t+1}^{(i)}$, via some deterministic rules.

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 aforementioned papers, the authors show that PL outperforms other particle filtering approaches, and is a cost-efficient alternative to MCMC methods. For a more detailed explanation of PL with illustrations refer to Carvalho, Johannes, Lopes & Polson (2010) and Lopes et al. (2011), among others.

2.5 Conclusions

In this chapter we reviewed univariate and multivariate GARCH and SV models and inference methods, putting emphasis on the Bayesian approach. We have surveyed the existing literature that concerns various Bayesian inference methods for these models, outlining the advantages of the Bayesian approach versus the classical procedures. We have also discussed in more detail recent Bayesian non-parametric methods for GARCH and SV models, which avoid imposing arbitrary parametric distributional assumptions. Additionally, we have reviewed an alternative Bayesian estimation approach - SMC, in particular, PL. The mentioned SMC technique can be seen as true

competitor to MCMC, since it produces a very similar posterior output at a much lower computational cost and allows for on-line estimation.

The main contributions of this thesis are included in the following chapters. Chapter 3 explores an asymmetric multivariate GARCH model with non-parametric errors. The inference is carried out using MCMC techniques. We show that non-parametric errors are not able to account for leverage effect, since the asymmetric non-parametric MGARCH specification outperforms its symmetric counterpart. We also present an illustrative portfolio allocation exercise. Chapter 4 employs the aforementioned PL to estimate a non-parametric SV model. We show that PL and MCMC produce very similar outputs for this model. Then the SV-DPM is augmented to incorporate Markov switching jumps, resulting into a MSSV-DPM model, which is tested on synthetic data and validated on real returns.

Chapter 3

A Bayesian Approach to the ADCC Model with Application to Portfolio Selection

This chapter presents a non-parametric multivariate asymmetric GARCH model for multiple financial returns and the MCMC scheme for inference and prediction. As mentioned before, ARCH-family models, first introduced by Engle (1982) and then generalized by Bollerslev (1986), are without doubt the most analyzed and used in practice to explain time-varying volatilities, see Bollerslev et al. (1992), Bollerslev et al. (1994), Engle (2002*b*), Teräsvirta (2009) and Tsay (2010).

Empirical evidence shows that returns and volatilities exhibit three types of asymmetries. The first two asymmetries are present in the dynamics of volatilities and correlations: these respond to changes in returns in an asymmetric manner that depends on the sign of the return. The third type of asymmetry is present in the unconditional distribution of the returns and is modeled via distributional assumptions for the error term.

Asymmetry in the volatility response to the changes in the returns, sometimes also called “leverage effect", was first introduced by Black (1976). It means that negative shocks to the returns have a stronger effect on volatility than positive ones of the same magnitude. When dealing with multiple returns, one must also take into consideration the mutual dependence between them, see Bauwens et al. (2006), Silvennoinen &

Teräsvirta (2009) and Tsay (2010). In particular, conditional correlation models, firstly proposed by Engle (2002*b*), Tse & Tsui (2002) and Christodoulakis & Satchell (2002), play an important role because there is evidence that conditional correlations between returns are time dependent. More recently, Cappiello et al. (2006) have proposed Asymmetric Dynamic Conditional Correlation (ADCC) model for time-varying correlations. Cappiello et al. (2006) argue that correlations between asset returns might be higher after a negative return than after a positive one of the same size. These two types of asymmetries govern the deterministic evolution of volatilities and correlations.

On the other hand, the third type of asymmetry - the unconditional one - can be modeled via the distribution of the returns. Many of the GARCH models, univariate or multivatriate, rely on Gaussianity assumption for the error term. However, the traditional premises of Normal distribution may be rather restrictive because the empirical unconditional distribution of returns, as mentioned before, is usually slightly skewed (asymmetric) and fat-tailed, see Rossi & Spazzini (2010), for example. Alternative parametric choices, such as the Student-t density, see Fiorentini et al. (2003), the skew-Student-t distribution, see Bauwens & Laurent (2005), or finite mixtures of Gaussian distributions, see e.g. Ausín & Galeano (2007) and Galeano & Ausín (2010), have been proposed in the literature and they usually improve the fit of GARCH models. However, all of them require the assumption of a certain parametric model. An alternative is to abandon the parametric setting altogether and consider a Dirichlet Process Mixture (DPM) model of Gaussian distributions, firstly introduced by Lo (1984). This is a very flexible model which can be viewed as a location-scale mixture of Gaussian distributions and is capable of modeling the Gaussian, Student-t, logistic, double exponential, Cauchy and generalized hyperbolic distributions, among others, see e.g. Tokdar (2006) and Mencía & Sentana (2009). The use of Bayesian non-parametrics in econometric modeling is relatively recent with rapidly increasing popularity for its superior performance. For a short review of existing works and possible econometric applications refer to the last paragraph of Section 2.3.

Therefore, in this chapter we consider an ADCC model for time-varying correlations with GJR-GARCH for individual volatilities (Glosten et al. 1993) and a DPM model for the return innovations, resulting into Bayesian non-parametric ADCC model (BNP-ADCC). We follow closely the works of Kalli et al. (2013) and Ausín et al. (2014), who have applied the DPM models for univariate GJR-GARCH and Jensen & Maheu (2013),

who have used DPM models for the multivariate symmetric DVEC (Ding & Engle 2001). Non-parametric time-varying volatility models have been of great interest in the recent literature, in both, GARCH and Stochastic Volatility setting, see Jensen & Maheu (2010, 2013, 2014), Delatola & Griffin (2011, 2013), Kalli et al. (2013), Ausín et al. (2014). For a survey on Bayesian inference methods for univariate and multivariate GARCH models see Chapter 2.

Jensen & Maheu (2013) have established the superiority of non-parametric errors in MGARCH models as compared to most commonly used parametric distributions, such as Gaussian and Student-t. In this chapter, we carry out an extensive comparison between our proposed BNP-ADCC model and its fully symmetric version, BNP-DCC, both with non-parametric errors.

The improved fit of the model to multiple financial time series can be applied to risk management problems, such as, portfolio optimization, for example. In this chapter, we propose a Bayesian method which provides the posterior distributions of the one-step-ahead optimal portfolio weights, which are more informative than simple point estimates. The Bayesian approach also helps to deal with parameter uncertainty in portfolio decision problems, see e.g. Jorion (1986), Greyserman et al. (2006), Avramov & Zhou (2010) and Kang (2011), among others. This is in contrast with the usual maximum likelihood estimation approach, which assumes a "certainty equivalence" viewpoint, where the sample estimates are treated as the true values, which is not in general correct and has been criticized in a number of papers, see Brown (1978) and Jorion (1986), among others.

Therefore, the main contribution of this chapter is the proposal of a Bayesian non-parametric method for explaining the asymmetric dynamics of the assets' returns via a BNP-ADCC model. We carry out a comparison with a fully symmetric BNP-DCC model to examine if by considering the asymmetric volatility and correlation response we can improve the prediction accuracy. Also, we present an application of Bayesian non-parametric techniques in portfolio decision problems and explore the differences in uncertainty between the two models. This chapter extends the work by Ausín et al. (2014) to the multivariate framework and the recent work by Jensen & Maheu (2013) to the asymmetric setting. Also, differently from the work of Jensen & Maheu (2013),

we always assume a conjugate prior specification and we use a different sampling approach.

The outline of the chapter is as follows: Section 3.1 describes the model, inference and prediction from a Bayesian perspective. Section 3.2 introduces the time-varying portfolio optimization problem. Section 3.3 presents a short simulation study. Section 3.4 illustrates the proposed approach using a real data example, compares the models and solves a portfolio allocation problem. Section 3.5 concludes.

3.1 Model, inference and prediction

This section describes the asymmetric dynamic conditional correlation GJR-GARCH model used for volatilities and correlations and the DPM specification for the error term, resulting in the BNP-ADCC model. Then we provide a detailed explanation of the implementation of Bayesian non-parametric inference and the methodology of obtaining predictive densities of returns and volatilities.

3.1.1 The Bayesian non-parametric ADCC model

Financial returns usually exhibit asymmetries in individual volatilities and in conditional correlations. Therefore, on the one hand, we choose the GJR-GARCH model proposed by Glosten et al. (1993) for individual returns to incorporate asymmetric volatility effects, while, on the other hand, we use the ADCC model proposed by Cappiello et al. (2006) to model joint volatilities. Then, we assume that the vector of K asset returns is given by:

$$r_t = H_t^{1/2} \epsilon_t, \tag{3.1}$$

for $t = 1, 2, \ldots$, where H_t is a symmetric $K \times K$ scale matrix and ϵ_t are a sequence of *iid* random variables with an unknown K-dimensional distribution $\mathcal{F}_K$. As usual in all DCC models, the matrix H_t can be decomposed as follows:

$$H_t = D_t R_t D_t, \tag{3.2}$$

where D_t is a diagonal matrix that contains the square root of the elements of the main diagonal of H_t, denoted by $d_{ii,t}$, for $i = 1, 2, \ldots, K$, and R_t is a time-varying correlation matrix. The $d_{ii,t}$'s are assumed to follow GJR-GARCH (1,1) models given by:

$$d_{ii,t}^2 = \omega_i + (\alpha_i + \phi_i I_{i,t-1}) r_{i,t-1}^2 + \beta_i d_{ii,t-1}^2, \tag{3.3}$$

with parameters $\omega_i, \alpha_i, \phi_i, \beta_i > 0$ and $\alpha_i + \beta_i + \phi_i/2 < 1$, for $i = 1, 2, \ldots, K$ and where $I_{i,t-1}$ is an indicator function such that $I_{i,t-1} = 1$ if $r_{i,t-1} < 0$ and $I_{i,t-1} = 0$ otherwise. On the other hand, to introduce R_t, we need to define

$$\varepsilon_t = D_t^{-1} r_t, \text{ and } \eta_t = \varepsilon_t \odot I(\varepsilon_t < 0), \tag{3.4}$$

where $\odot$ denotes Hadamard matrix product operator and $I(\varepsilon_t < 0)$ is a vector with ith component equal to 1 if $\varepsilon_{i,t} < 0$, and 0 otherwise. The Hadamard product operator helps to create a new vector of only negative residuals by multiplying entry-wise the original residual vector ε_t by an indicator function. Then, R_t is given by:

$$R_t = Q_t^{\star -1} Q_t Q_t^{\star -1}, \tag{3.5}$$

where Q_t is the $K \times K$ matrix given by:

$$Q_t = S(1 - \kappa - \lambda - \delta/2) + \kappa \times \varepsilon_{t-1}\varepsilon_{t-1}' + \lambda \times Q_{t-1} + \delta \times \eta_{t-1}\eta_{t-1}', \tag{3.6}$$

and $Q_t^\star$ is a diagonal matrix with the square root of the ith diagonal element of Q_t on its ith diagonal position. As pointed out by Cappiello et al. (2006), $Q_t^\star$ is a matrix that guarantees R_t is a correlation matrix, as long as Q_t is positive definite, S being the sample correlation matrix of ε_t. In Equation 3.6 we impose that $\kappa, \lambda, \delta > 0$ and $\kappa + \lambda + \delta/2 < 1$ to ensure the positivity and stationarity of Q_t. Finally, the vector $\Phi = (\omega, \alpha, \beta, \phi, \kappa, \lambda, \delta)$ summarizes the set of parameters describing the matrices H_t, for $t = 1, 2, \ldots$.

Popular parametric choices for the unknown distribution of $\epsilon_t \sim \mathcal{F}_K$ include Student-t, skewed Student-t, finite mixtures of Gaussian distributions. However, as pointed out by Jensen & Maheu (2013), these models still remain parametric. Next, we present a

flexible DPM specification for the errors and some of the most important special cases arising from this model. Using the stick-breaking representation by Sethuraman (1994), a DPM of Gaussian distributions can be expressed as a location-scale Gaussian mixture model with infinitely many components and therefore, it can be easily defined as an extension of a parametric mixture model. The base distribution of the DP, usually denoted by G_0, corresponds to the prior distribution of the component parameters in the infinite mixture.

The concentration parameter, denoted by 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. Additionally, c can be seen as a precision parameter that indicates how close G is to the base distribution G_0, where larger c indicates that G is closer to G_0. So, if the base measure is not chosen well, it will impact on $\mathcal{F}_K$. As pointed out by Kalli et al. (2013), c controls the decay of weights, which is exponential. This might be a disadvantage of DPM models in the case that more mixture components are needed. An alternative could be to consider more general stick-breaking processes, for more details refer to Kalli et al. (2013).

Therefore, the resulting density function of ϵ_t can be written as:

$$f\left(\epsilon_t|w,\mu,\Lambda\right) = \sum_{j=1}^{\infty} w_j \mathcal{N}_K\left(\epsilon_t|\mu_j,\Lambda_j^{-1}\right), \tag{3.7}$$

where $\mathcal{N}_K$ denotes a K-variate Normal density. Let us denote by $\Omega = \left\{w_j,\mu_j,\Lambda_j\right\}_{j=1}^{\infty}$ the infinite-dimensional parameter vector describing the mixture distribution for the innovations. Here w_j represent the component weights, μ_j are the component means and Λ_j are the precision matrices, for $j = 1,2,\ldots$. Using the stick-breaking representation, the weights of the infinite mixture components are reparameterized as follows: $w_1 = v_1$, $w_j = (1-v_1)\cdots(1-v_{j-1})v_j$, where we assume a Beta prior distribution for v_j, $v_j \sim \mathcal{B}(1,c)$, for $j = 1,2,\ldots$. Clearly, there will be some sensitivity to the choice of the concentration parameter c. Therefore, we further assume a Gamma hyper-prior distribution for c, $c \sim \mathcal{G}(a_0,b_0)$, as seen in Escobar & West (1995). Finally, as a base distribution, we assume a conjugate Normal-Wishart prior for (μ_j,Λ_j), $(\mu_j,\Lambda_j) \sim \mathcal{NW}(m_0,s_0,W_0,d_0)$, where

$$\mu_j|\Lambda_j \sim \mathcal{N}_K\left(m_0, (s_0\Lambda_j)^{-1}\right),$$
$$\Lambda_j \sim \mathcal{W}(W_0, d_0),$$

for $j = 1, 2, \ldots$, such that $\mathrm{E}\left[\Lambda_j\right] = d_0 \times W_0^{-1}$ and $\mathrm{E}\left[\Lambda_j^{-1}\right] = (d_0 - (K+1)/2)^{-1} \times W_0$.

In summary, the complete set of model parameters is denoted by $\Theta = (\Phi, \Omega)$. Given the information available up to time $t-1$, denoted by r^{t-1}, the conditional density of the returns can be written as follows:

$$f(r_t|\Theta, r^{t-1}) = \sum_{j=1}^{\infty} w_j \mathcal{N}_K\left(r_t|H_t^{1/2}\mu_j, H_t^{1/2}\Lambda_j^{-1}(H_t^{1/2})'\right), \tag{3.8}$$

with conditional mean given by:

$$\mu_t^\star = E\left[r_t|\Theta, r^{t-1}\right] = H_t^{1/2}\sum_{j=1}^{\infty} w_j\mu_j, \tag{3.9}$$

and conditional covariance matrix:

$$H_t^\star = \mathrm{Cov}\left[r_t|\Theta, r^{t-1}\right] = H_t^{1/2}\mathrm{Cov}\left[\epsilon_t|\Omega\right](H_t^{1/2})', \tag{3.10}$$

where

$$\mathrm{Cov}\left[\epsilon_t|\Omega\right] = \sum_{j=1}^{\infty} w_j\left(\Lambda_j^{-1} + \mu_j(\mu_j)'\right) - \left(\sum_{j=1}^{\infty} w_j\mu_j\right)\left(\sum_{j=1}^{\infty} w_j\mu_j\right)'.$$

It is important to notice that this full unrestricted model induces GARCH-in-Mean effects, since the conditional mean of the returns is not restricted to be zero. Moreover, the DPM model for ϵ_t does not assume an identity covariance matrix. As noted in Jensen & Maheu (2013), imposing moment restrictions in DPM models is still an open question. However, the prior information considered centers ϵ_t around an identity covariance matrix.

3.1.2 MCMC algorithm

The following section describes a Markov Chain Monte Carlo (MCMC) algorithm to sample from the posterior distribution of the parameters of the BNP-ADCC model introduced in the previous section. The algorithm is based on the slice-sampler by Kalli et al. (2011), which is an improved version of the algorithm by Walker (2007). This algorithm belongs to a general class of conditional methods, that sample a sufficient and finite number of variables to deal with the infiniteness problem, see Ishwaran & James (2001), Walker (2007), Papaspiliopoulos & Roberts (2008), Papaspiliopoulos (2008) and Kalli et al. (2011).

Following Kalli et al. (2011), in order to avoid the problem of sampling an infinite number of values at each MCMC step, we introduce a latent variable u_t, such that the joint density of (ϵ_t, u_t) given Ω is:

$$f(\epsilon_t, u_t|\Omega) = \sum_{j=1}^{\infty} \underline{1}(u_t < w_j)\mathcal{N}_K(\epsilon_t|\mu_j, \Lambda_j^{-1}). \tag{3.11}$$

Let $A_w(u_t) = \{j : w_j > u_t\}$ be a set of size N_{u_t}, which is finite for all $u_t > 0$. Then the joint density of (ϵ_t, u_t) in Equation 3.11 can be equivalently written as $f(\epsilon_t, u_t|\Omega) = \sum_{j\in A_w(u_t)} \mathcal{N}_K(\epsilon_t|\mu_j, \Lambda_j^{-1})$. Integrating over u_t gives us the density of infinite mixture of distributions Equation 3.7. Finally, given u_t, the number of mixture components is finite.

In order to simplify the likelihood, we also need to introduce a further indicator latent variable z_t, which indicates the mixture component that ϵ_t comes from:

$$f(\epsilon_t, z_t = j, u_t|\Omega) = \mathcal{N}_K(\epsilon_t|\mu, \Lambda^{-1})\underline{1}(j \in A_w(u_t)).$$

Then, the log-likelihood of Θ, given the latent variables u_t and z_t looks as follows:

$$\log L(\Theta|\{r_t, u_t, z_t\}_{t=1}^T) = -\frac{1}{2}\sum_{t=1}^{T}\left(K\log(2\pi) + \log|H_{t,z_t}^{\star}| + (r_t - \mu_{t,z_t}^{\star})H_{t,z_t}^{\star -1}(r_t - \mu_{t,z_t}^{\star})'\right). \tag{3.12}$$

Here $\mu_{t,z_t}^{\star}$ and $H_{t,z_t}^{\star}$ are the conditional mean vector and conditional covariance matrix given z_t, i.e., $\mu_{t,z_t}^{\star} = H_t^{1/2}\mu_{z_t}$ and $H_{t,z_t}^{\star} = H_t^{1/2}\Lambda_{z_t}^{-1}H_t^{1/2}$, respectively. Using these latent

variables, we now construct the following MCMC algorithm that is described step by step.

1. **Sampling c**

   Firstly, given z_t, for $t = 1, 2, \ldots, T$, the conditional posterior distribution of the concentration parameter c is independent of the rest of the parameters, as seen in Escobar & West (1995). So, we first sample an auxiliary variable $\xi \sim \mathcal{B}(c+1, T)$ and then c from a Gamma mixture:

   $$\pi_\xi \mathcal{G}(a_0 + z^\star, b_0 - \log(\xi)) + (1 - \pi_\xi)\mathcal{G}(a_0 + z^\star - 1, b_0 - \log(\xi)),$$

   where $z^\star = \max(z_1, \ldots, z_T)$ and $\pi_\xi = (a_0 + z^\star - 1)/(a_0 + z^\star - 1 + T(b_0 - \log(\xi)))$.

2. **Sampling v**

   In the second step, we sample from the conditional posterior of v_j for $j = 1, 2, \ldots, z^\star$, which is given by:

   $$v_j | \{z_t\}_{t=1}^T \sim \mathcal{B}(n_j + 1, T - \sum_{l=1}^{j} n_l + c),$$

   where n_j is the number of observations in the jth component and $\sum_{l=1}^{j} n_l$ gives the cumulative sum of the groups. Also, $w_1 = v_1$, $w_j = (1 - v_1) \cdots (1 - v_{j-1}) v_j$, for $j = 2, \ldots, z^\star$. Then, we sample the uniform latent variables $u_t \sim \mathcal{U}(0, w_{z_t})$, for $t = 1, 2, \ldots, T$. Following Kalli et al. (2011), we need to find the smallest $j^\star$ such that $\sum_{j=1}^{j^\star} w_j > (1 - u^\star)$, where $u^\star = \min(u_1, \ldots, u_T)$. Then, if $z^\star < j^\star$, we need to sample v_j from the prior and sample w_j accordingly, for $j = z^\star + 1, \ldots, j^\star$.

3. **Sampling μ and Λ**

   As for the mixture parameters, we sample them from the conditional posterior Normal-Wishart distribution $(\mu_j, \Lambda_j) \sim \mathcal{NW}(m_j, s_j, W_j, d_j)$, for $j = 1, 2, \ldots, j^\star$, where:

$$
\begin{aligned}
m_j &= \frac{s_0 m_0 + n_j \bar{\epsilon}_j}{s_0 + n_j}, \qquad s_j = s_0 + n_j, \\
W_j &= W_0^{-1} + S_j + \frac{s_0 n_j}{s_0 + n_j}(m_0 - \bar{\epsilon}_j)(m_0 - \bar{\epsilon}_j)', \\
S_j &= \frac{1}{n_j} \sum_{t:z_t=j}^{T} (\epsilon_t - \bar{\epsilon}_j)(\epsilon_t - \bar{\epsilon}_j)', \qquad \bar{\epsilon}_j = \frac{1}{n_j} \sum_{t:z_t=j}^{T} \epsilon_t, \\
d_j &= d_0 + n_j.
\end{aligned}
$$

Note that this approach is different from the one described in Jensen & Maheu (2013) since they assume independent prior distributions for μ_j and Λ_j and include some Gibbs steps to sample from the conditional posterior.

4. **Sampling z**

   In this step we assign to which component each observation belongs to by sampling the latent indicator variable z_t from its conditional posterior distribution:

$$
\Pr(z_t = j|...) \propto \underline{1}\left(j \in A_w\left(u_t\right)\right) \mathcal{N}_K(\epsilon_t|\mu_j, \Lambda_j^{-1}).
$$

5. **Sampling parameters $\boldsymbol{\Phi}$**

   The rest of the steps of the algorithm concern updating the parameters of the BNP-ADCC model. For that, we use the Random Walk Metropolis Hasting (RWMH), following a similar procedure as in Jensen & Maheu (2013). For the set of parameters Φ, we generate a candidate value $\tilde{\Phi}$ from a D-variate Normal distribution with mean equal to the previous value of the parameter, where $D = 4K + 3$ is the number of parameters in Φ, as follows:

$$
\tilde{\Phi} \sim \begin{cases} \mathcal{N}_D\left(\Phi, V\right) & \text{w.p.} \quad p \\ \mathcal{N}_D\left(\Phi, 100V\right) & \text{w.p.} \quad 1-p \end{cases}
$$

   The probability of accepting a proposed value $\tilde{\Phi}$, given the current value Φ, is $\alpha(\Phi, \tilde{\Phi}) = \min\left\{1, \prod_{t=1}^{T} l(r_t|\Phi) / \prod_{t=1}^{T} \tilde{l}(r_t|\tilde{\Phi})\right\}$, where the likelihood used is as in Equation 3.12, see e.g. Robert & Casella (2004). The covariance matrix V is obtained by running some initial MCMC iterations and then adjusting the sample covariance matrix by some factor in order to achieve the desired acceptance probability. In this chapter the acceptance probabilities are adjusted to be between 20% and 50%, while we fix $p = 0.9$.

In the simulation and real data application studies in the following sections we assume uninformative uniform priors restricted to the stationary region for Φ and setting $m_0 = 0_2$, $s_0 = 0.1$, $d_0 = 5$, $W_0 = I_2/5$, $a_0 = 4$ and $b_0 = 4$. The choice of hyper-parameters is such that the mixture components would initially center at zero and have unit variances. The MCMC algorithm is run for 10k burn-in plus 40k iterations for the simulation study and 50k + 50k for the real data application, in order to ensure convergence.

3.1.3 Prediction

In this section, we are mainly interested in estimating the one-step-ahead predictive density of the returns:

$$f(r_{T+1}|r^T) = \int f(r_{T+1}|\Theta, r^T) f(\Theta|r^T) d\Theta, \tag{3.13}$$

where $f(r_{T+1}|\Theta, r^T)$ is specified in Equation 3.8. Although this integral is not analytically tractable, it can be approximated using the MCMC output:

$$f(r_{T+1}|r^T) \simeq \frac{1}{M} \sum_{m=1}^{M} f(r_{T+1}|\Theta^{(m)}, r^T), \tag{3.14}$$

where M is the length of the MCMC chain and $\Theta^{(m)}$ is the infinite set of parameters at the m-th iteration. However, in practice, at each iteration, there is a finite number of weights $w_j^{(m)}$, means $\mu_j^{(m)}$ and precision matrices $\Lambda_j^{(m)}$, for $j = 1, 2, \ldots, j^{\star(m)}$, where $j^{\star(m)}$ is a number of components to sample at the step 2 of MCMC sampler at the m-th iteration. Then, as seen in Jensen & Maheu (2013), we repeat for $r = 1, 2, \ldots, R$ at each MCMC iteration, where R a is number of components to sample and is fixed a priori:

i. Sample a random variable $a \sim \mathcal{U}(0, 1)$.

ii. Take such $w_r^{(m)}$ for which $\sum_{j=1}^{r-1} w_{j-1}^{(m)} < a < \sum_{j=1}^{r} w_j^{(m)}$ and the corresponding $(\mu_r, \Lambda_r)^{(m)}$.

iii. If $\sum_{j=1}^{j^{\star(m)}} w_j^{(m)} < a$, sample $(\mu_r, \Lambda_r)^{(m)}$ from the Normal-Wishart prior.

Then, approximate the one-step-ahead density in Equation 3.14 by

$$f(r_{T+1}|\Theta^{(m)}, r^T) \simeq \frac{1}{R}\sum_{r=1}^{R}\mathcal{N}_K\left(r_{T+1}|\mu_r^{(m)}H_{T+1}^{(m)1/2}, H_{T+1}^{(m)1/2}\left(\Lambda_r^{(m)}\right)^{-1}(H_{T+1}^{(m)1/2})'\right), \tag{3.15}$$

where $(\mu_r, \Lambda_r)^{(m)}$ are the R pairs of means and precision matrices simulated for $r = 1, 2, \ldots, R$, and $H_{T+1}^{(m)}$ is the value of the H_{T+1} matrix at the m-th MCMC iteration.

This simulation procedure also delivers predictions for many other important measures. For example, the posterior expected value of the adjusted one-step-ahead mean and volatility matrix, introduced in Equation 3.9 and Equation 3.10, can be approximated by:

$$\mathrm{E}\left[\mu_{T+1}^{\star} \mid r^T\right] \simeq \frac{1}{M}\sum_{m=1}^{M}\mu_{T+1}^{\star(m)}, \tag{3.16}$$

and

$$\mathrm{E}\left[H_{T+1}^{\star} \mid r^T\right] \simeq \frac{1}{M}\sum_{m=1}^{M}H_{T+1}^{\star(m)}, \tag{3.17}$$

respectively, where,

$$\mu_{T+1}^{\star(m)} = H_{T+1}^{(m)1/2}\left(\frac{1}{R}\sum_{r=1}^{R}\mu_r^{(m)}\right), \tag{3.18}$$

and

$$H_{T+1}^{\star(m)} = H_{T+1}^{(m)1/2}\left(\frac{1}{R}\sum_{r=1}^{R}\left(\left(\Lambda_r^{(m)}\right)^{-1} + \mu_r^{(m)}(\mu_r^{(m)})'\right) - \left(\frac{1}{R}\sum_{r=1}^{R}\mu_r^{(m)}\right)\left(\frac{1}{R}\sum_{r=1}^{R}\mu_r^{(m)}\right)'\right) \times \left(H_{T+1}^{(m)1/2}\right)'. \tag{3.19}$$

In order to obtain the posterior distributions of the adjusted means and volatilities, one should fix a certain R. Since the number of components in the data is not known a priori, one might choose R depending on the number of clusters in the data. However, this implies that there is no upper limit for R, which might result into sampling a very large number of components at each step and increasing computational cost. Instead,

we propose to use $j^\star$ components with their corresponding weights w at each step, such that equations Equation 3.18 and Equation 3.19 become the following:

$$\mu_{T+1}^{\star(m)} = H_{T+1}^{(m)1/2} \left(\sum_{j=1}^{j^{\star(m)}} w_j^{(m)} \mu_j^{(m)} \right), \tag{3.20}$$

and

$$\begin{aligned} H_{T+1}^{\star(m)} &= H_{T+1}^{(m)1/2} \times \\ &\left(\sum_{j=1}^{j^{\star(m)}} w_j^{(m)} \left(\left(\Lambda_j^{(m)} \right)^{-1} + \mu_j^{(m)} (\mu_j^{(m)})' \right) - \left(\sum_{j=1}^{j^{\star(m)}} w_j^{(m)} \mu_j^{(m)} \right) \left(\sum_{j=1}^{j^{\star(m)}} w_j^{(m)} \mu_j^{(m)} \right)' \right) \\ &\times \left(H_{T+1}^{(m)1/2} \right)'. \end{aligned} \tag{3.21}$$

Similarly, we can approximate the posterior median and credible intervals using the quantiles of the posterior samples $\left\{ \mu_{T+1}^{\star(m)} \right\}_{m=1}^{M}$ and $\left\{ H_{T+1}^{\star(m)} \right\}_{m=1}^{M}$.

3.2 Portfolio decisions

Optimal asset allocation is greatly affected by the parameter uncertainty, see Jorion (1986) and Greyserman et al. (2006), among others. In the frequentist setting, the estimated parameter values are considered to be the true ones, therefore, the optimal portfolio weights tend to inherit this estimation error. Instead of solving the optimization problem on the basis of the choice of unique parameter values, the investor can choose the Bayesian approach, because it accounts for parameter uncertainty, as seen in Kang (2011) and Jacquier & Polson (2013), for example.

The main objective of diversification is to reduce investor's exposure to risk. See Markowitz (1952) and Merton (1972) for some classical portfolio optimization references. Nowadays, there is a wide variety of portfolio optimization objectives, such as maximizing agent's utility or minimizing expected shortfall, among many others. In this chapter we consider one of the most frequently used objectives, where the investor minimizes the portfolio variance. The Global Minimum Variance (GMV) portfolio can be found at the very peak of the efficient frontier. Given the time series of returns

$r_1, \ldots, r_T$, the standard approach is to consider the unconditional covariance matrix of the returns, $\Sigma = \text{Cov}[r_t]$, and solve the following optimization problem:

$$p^\star = \arg \min_{p: p'1_K = 1} \text{Var}[r_t^P],$$

where p is the weight vector, 1_K is a K-vector of ones and $r_t^P = p'r_t$ is the portfolio return at time t. The problem has the solution:

$$p^\star = \frac{\Sigma^{-1}1_K}{1_K'\Sigma^{-1}1_K},$$

that is independent of the time point T. Note that if we choose to impose the short sale constraint, i.e., $p_i \geq 0$, $\forall i = 1, 2, \ldots, K$, the problem cannot be solved analytically anymore and it requires numerical optimization techniques.

However, recent results suggest that the use of the time-varying covariance matrix to determine portfolio weights leads to better performing portfolios than the use of a constant covariance matrix. For instance, Giamouridis & Vrontos (2007) find that portfolios, constructed under a dynamic approach, have lower average risk and higher out-of-sample risk-adjusted realized return, see also Yilmaz (2011). Cecchetti et al. (1988) was the first to suggest the use of MGARCH models in optimal allocation context. Since then, there has been a number of papers investigating the differences in estimation and evaluating their performance using various approaches, from simple OLS, to bivariate vector autorregression (VAR), to GARCH. In particular, Kroner & Sultan (1993), Rossi & Zucca (2002) and Yang & Allen (2004), among others, have shown that GARCH-type models lead to the overall portfolio risk reduction.

Consequentely, to solve the portfolio allocation problem in our case, instead of Σ, we use the adjusted one-step-ahead conditional covariance matrix for the assets returns $H_{T+1}^\star$, defined in Equation 3.10, which varies continuously on the basis of available information up to time T, r^T. Therefore, we are able to obtain optimal portfolio weights for time point $T+1$ as follows:

$$p_{T+1}^\star = \frac{H_{T+1}^{\star -1}1_K}{1_K'H_{T+1}^{\star -1}1_K}. \tag{3.22}$$

Using the MCMC output, we can obtain samples from the entire posterior distribution of optimal portfolio weights for $T+1$, $f(p^{\star}_{T+1}|r^T)$. This approach relies on solving the allocation problem at every MCMC iteration and approximating the posterior mean of the optimal portfolio weights as follows:

$$\mathrm{E}[p^{\star}_{T+1}|r^T] = \int p^{\star}_{T+1} f(\Theta|r^T) d\Theta \approx \frac{1}{M}\sum_{m=1}^{M} p^{\star(m)}_{T+1},$$

where $\left\{p^{\star(m)}_{T+1}\right\}_{m=1}^{M}$ is a posterior sample of optimal portfolio weights obtained from Equation 3.22 for each value of one-step-ahead conditional covariance matrix of the returns $\left\{H^{\star(m)}_{T+1}\right\}_{m=1}^{M}$ in the MCMC sample. In other words, since we have assembled M one-step-ahead volatility matrices, we can solve the portfolio allocation problem M times. As in the previous section, we can similarly approximate the posterior median and credible intervals of $p^{\star}_{T+1}$ by using the quantiles of the sample of optimal portfolio weights. In this manner, we are able to draw samples from the posterior distribution of the optimal portfolio variance, $\sigma^{2\star}_{T+1}$, and optimal portfolio gain, $g^{\star}_{T+1}$ as follows:

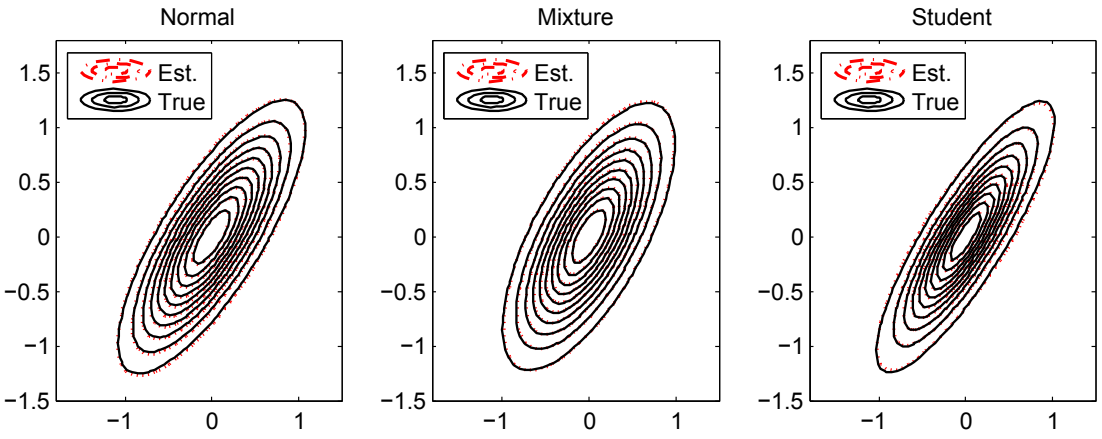
$$\left\{\left(\sigma^{2\star}_{T+1}\right)^{(m)}\right\}_{m=1}^{M} \sim p\left(\sigma^{2\star}_{T+1}|r^T\right),$$
$$\left\{\left(g^{\star}_{T+1}\right)^{(m)}\right\}_{m=1}^{M} \sim p\left(g^{\star}_{T+1}|r^T\right),$$

where $\{(\sigma^{2\star}_{T+1})^{(m)}\}_{m=1}^{M} = \{(p^{\star'}_{T+1}H^{\star}_{T+1}p^{\star}_{T+1})^{(m)}\}_{m=1}^{M}$ and $\{(g^{\star}_{T+1})^{(m)}\}_{m=1}^{M} = \{(p^{\star}_{T+1}\mu^{\star'}_{T+1})^{(m)}\}_{m=1}^{M}$.

3.3 Simulation study

The goal of this simulation study is to show the flexibility and adaptability of the DPM specification for the innovations for the BNP-ADCC model introduced in Section 3.1. In particular, we demonstrate that the DPM error specification can adapt to some of the most popular parametric distributions used in financial return data. For this, we consider three bivariate time series of 3000 observations simulated from a BNP-ADCC model with the following innovation distributions: (a) Gaussian $\mathcal{N}(0, I_2)$; (b) Student-t $\mathcal{T}(I_2, \nu = 8)$; (c) Mixture of two bivariate Normals $0.9\mathcal{N}(0, \sigma^2_1 = 0.8, \sigma_{12} = 0.0849, \sigma^2_2 = 0.9) + 0.1\mathcal{N}(0, \sigma^2_1 = 2.8, \sigma_{12} = -0.7637, \sigma^2_2 = 1.9)$. Note that, in the third case, we have

FIGURE 3.1: Contour plots of the true and estimated one-step-ahead predictive densities, $f(r_{T+1} \mid r^T)$, for the three simulated data sets.



chosen larger variances for the second mixture component to allow for the presence of extreme returns but preserving an identity covariance matrix. Then, we estimate all three data sets using the proposed BNP-ADCC model.

Figure 3.1 presents the contour plots that compare the true one-step-ahead predictive densities of returns, given the model parameters, with the estimated ones, obtained from Equation 3.14 by setting $R = 3$. As we can see, the estimated contours of the one-step ahead return densities are very close to the true ones. Note that these contours can be seen as a summary of the estimation results for all 11 model parameters $\Phi = (\omega, \alpha, \beta, \phi, \kappa, \lambda, \delta)$ and the distribution for the error term. Therefore, it seems that the proposed infinite mixture model is a very flexible tool that is able to adjust to rather different return specifications. This is of primary interest because in practice one never knows which is the true error distribution.

Table 3.1 presents the estimated posterior mean, median and 95% credible intervals for the number of clusters, $z^\star$, for the three generated datasets. For the Gaussian dataset, the proposed DPM model estimates very few non-empty components, 1.23 on average, where there is always a clear dominant weight. For the Student-t dataset, the proposed DPM model estimates a large number of clusters, around 19.66, with similar small weights. This is expected since, as commented in Jensen & Maheu (2013), the Student-t distribution can be viewed as a limiting case of a DPM model when the concentration parameter goes to infinity and, consequently, the number of clusters increases indefinitely. Finally, for the two-component mixture data, the DPM model can

identify very well the two underlying clusters with posterior mean around 2.68. Next, Table 3.1 shows the estimation results for the concentration parameter, c, and its transformed value $A = c/(1+c)$, where $0 < A < 1$, that has been used by Jensen & Maheu (2013) to provide an intuition of the probability of having infinite clusters in the mixture. However, note that, different to Jensen & Maheu (2013), we have previously defined a Gamma prior on c instead of a Uniform prior on A. Observe that the obtained results are coherent, since the posterior mean of A for the Gaussian case is the smallest ($A = 0.2509$), while for the Student-t case is the largest ($A = 0.6892$). Finally, for the two-component mixture dataset, the posterior mean of A is between the corresponding values of the Gaussian and Student-t, that can be seen as a compromise between the two extreme cases. The rest of the Table 3.1 contains the true and estimated parameter values and the corresponding 95% Bayesian credible intervals. As can be seen from the table, the estimation of parameters is quite good.

Finally, we have estimated the generated Normal and Student-t data sets assuming Gaussian and Student-t distributions, respectively. We used the RWMH with 10k burn-in plus 40k iterations. This way we were able to obtain a sample of one-step-ahead covariance matrices $\{H_{t+1}^{(m)}\}_{m=1}^{M}$, estimated using the true return distributions. Figure 3.2 compares the densities for one-step-ahead covariances $\{H_{t+1}^{\star(m)}\}_{m=1}^{M}$ assuming a DPM for (a) and (b) with the true data generating model, Gaussian and Student-t, respectively. As we can see, the mean estimates, the width and shape of the posterior distributions are very similar for DPM and the ones obtained using the true return distribution. A few different data sets were used to test the performance of the DPM model. It was always able to “recover" the underlying distribution of the error term, and, in turn, estimate well the posterior densities of the elements of the one-step-ahead covariance matrix. Therefore, we can conclude that DPM model can adjust to different frequently used distributions for the return data without making any restrictive distributional assumptions.

3.4 Real data and results

In this section, we illustrate the performance of the proposed methodology using a real dataset and solve a portfolio allocation problem as described in Section 3.2.

TABLE 3.1: Posterior means, medians and 95% credible intervals for the number of non-empty clusters $z^\star$, concentration parameter c, quantity $A = c/(1+c)$, and model parameters for the three simulated data sets.

		Gaussian		Student-t		2 comp. mixture	
	True value	Mean Median	95% CI	Mean Median	95% CI	Mean Median	95% CI
$z^\star$		1.2330 1.0000	(1.0000, 3.0000)	19.6612 19.0000	(9.0000, 33.0000)	2.6765 2.0000	(2.0000, 5.0000)
c		0.3578 0.3252	(0.0934, 0.8072)	2.5037 2.3308	(0.9327, 5.0059)	0.4863 0.4462	(0.1512, 1.0408)
A		0.2509 0.2454	(0.0855, 0.4467)	0.6892 0.6998	(0.4826, 0.8335)	0.3122 0.3085	(0.1314, 0.5100)
ω_1	0.01	0.0075 0.0077	(0.0045, 0.0104)	0.0110 0.0110	(0.0072, 0.0154)	0.0201 0.0201	(0.0104, 0.0315)
ω_2	0.01	0.0138 0.0142	(0.0071, 0.0204)	0.0109 0.0103	(0.0065, 0.0182)	0.0130 0.0118	(0.0071, 0.0218)
α_1	0.1	0.0595 0.0582	(0.0383, 0.0858)	0.0970 0.0953	(0.0740, 0.1277)	0.1473 0.1353	(0.0808, 0.2570)
α_2	0.08	0.0699 0.0716	(0.0402, 0.0930)	0.0710 0.0693	(0.0500, 0.0998)	0.0962 0.0944	(0.0512, 0.1408)
β_1	0.85	0.8625 0.8629	(0.8388, 0.8816)	0.8244 0.8262	(0.7920, 0.8481)	0.8426 0.8441	(0.8076, 0.8710)
β_2	0.88	0.8695 0.8686	(0.8478, 0.8946)	0.8727 0.8734	(0.8539, 0.8888)	0.8939 0.8952	(0.8589, 0.9130)
ϕ_1	0.025	0.0299 0.0285	(0.0085, 0.0544)	0.0395 0.0392	(0.0049, 0.0749)	0.0463 0.0475	(0.0034, 0.0800)
ϕ_2	0.025	0.0385 0.0373	(0.0137, 0.0705)	0.0229 0.0204	(0.0031, 0.0554)	0.0438 0.0443	(0.0147, 0.0729)
κ	0.05	0.0504 0.0495	(0.0366, 0.0681)	0.0501 0.0503	(0.0347, 0.0648)	0.0615 0.0588	(0.0407, 0.0878)
λ	0.9	0.8908 0.8903	(0.8670, 0.9136)	0.8946 0.8957	(0.8786, 0.9075)	0.8898 0.8884	(0.8674, 0.9167)
δ	0.025	0.0221 0.0216	(0.0015, 0.0475)	0.0312 0.0311	(0.0098, 0.0536)	0.0265 0.0264	(0.0011, 0.0551)

3.4.1 Estimation

We consider the daily price data of Apple Inc. company (P_t^A) and NASDAQ Industrial index (P_t^N) from January 1, 2000 till May 7, 2012, obtained from Yahoo Finance. Then, daily prices are transformed into daily logarithmic returns (in %), resulting in $T = 3098$ observations. Table 3.2 provides the basic descriptive statistics and Figure 3.3 illustrates the dynamics of the log-returns.

As expected, Apple Inc. has higher overall variance because of the higher mean return. Both returns do not exhibit any evidence of autoregressive behavior. Apple Inc. returns

FIGURE 3.2: Densities of the elements of the one-step-ahead covariance matrices for Normal and Student-t data estimated using (a) DPM and Normal and (b) DPM and Student-t errors.

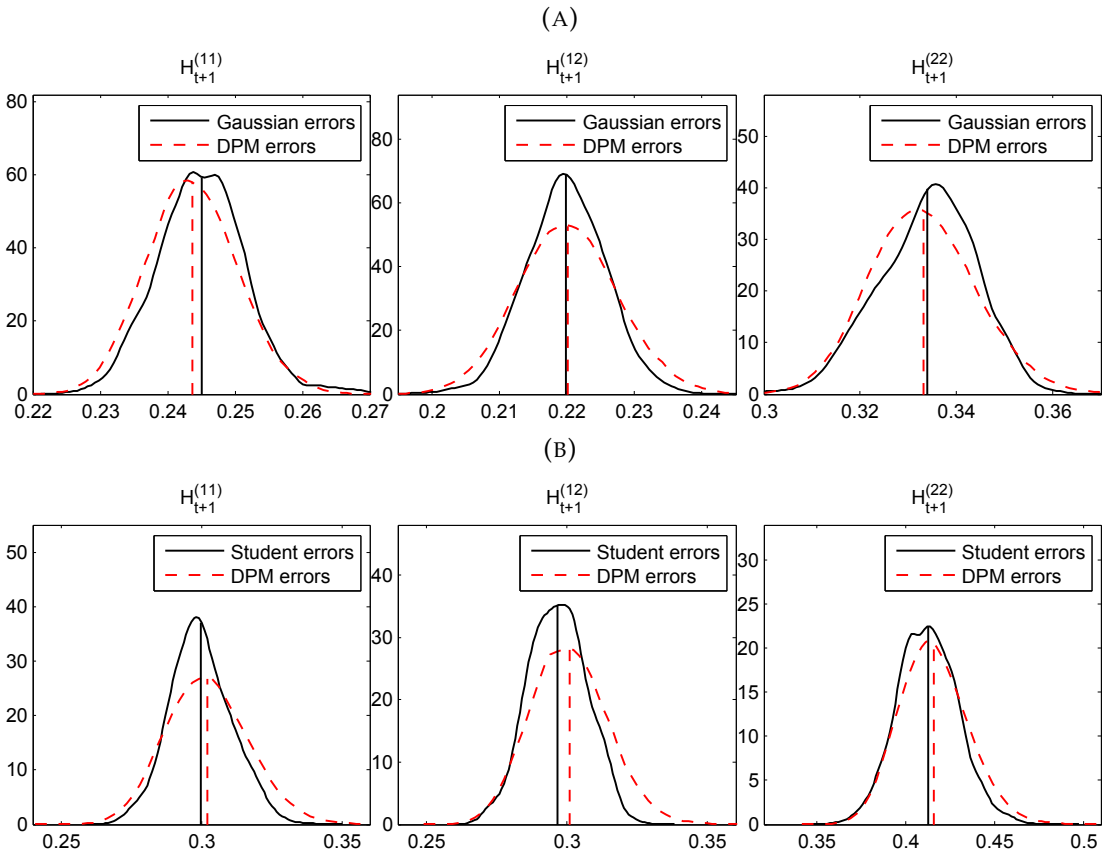


FIGURE 3.3: Log-returns (in %) and histograms of Apple Inc. and NASDAQ Ind. index.

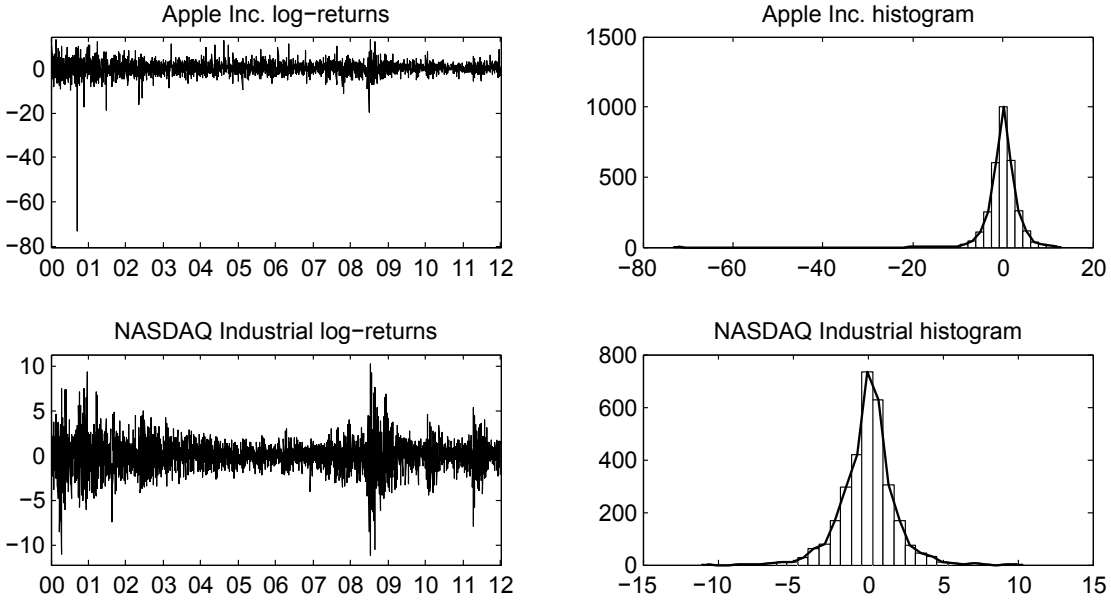


TABLE 3.2: Descriptive statistics of the Apple Inc. and NASDAQ Ind. index return series.

	$100 \times \ln\left(\frac{P_t^A}{P_{t-1}^A}\right)$	$100 \times \ln\left(\frac{P_t^N}{P_{t-1}^N}\right)$
Mean	0.0973	0.0020
Median	0.1007	0.0766
Variance	9.7482	3.1537
Skewness	−4.2492	−0.1487
Kurtosis	102.0411	7.1513
Correlation	0.5376	

contain one atypical data point, corresponding to September 29, 2000. The very low return is due to an announcement the day before about lower than expected sales.

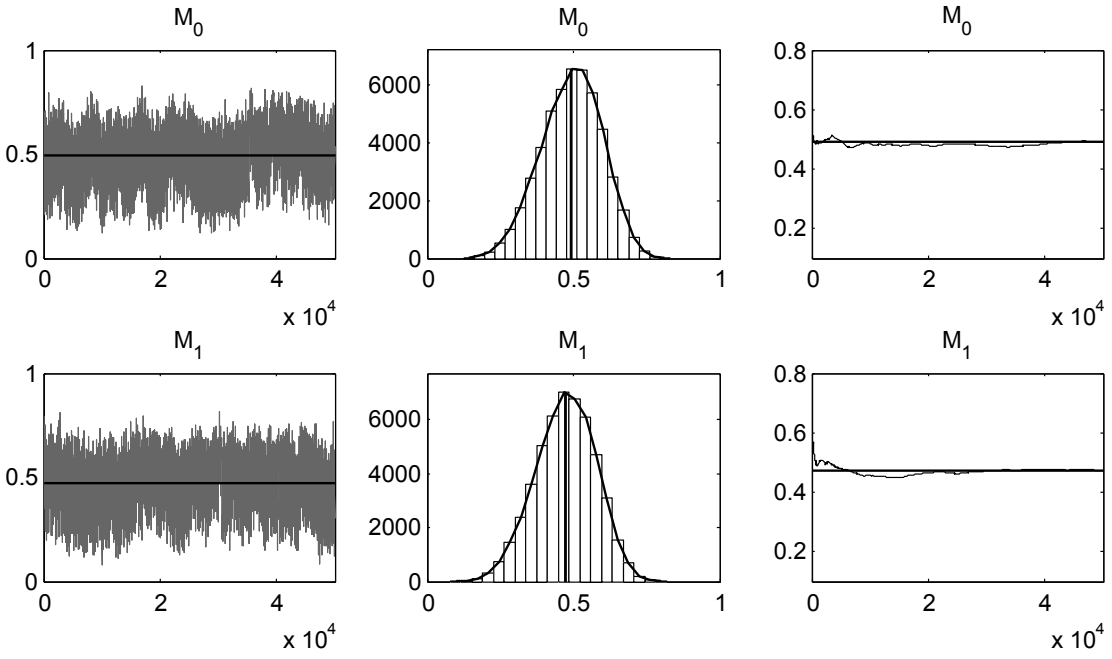
Next, the return series was estimated assuming fully symmetric BNP-DCC (M_0) and asymmetric BNP-ADCC (M_1) models, both with DPM errors. Table 3.3 reports the parameter estimation results. As we can see from the table, the constant volatility parameter for the first series is twice as big for the asymmetric model. On the other hand, the volatility persistence parameter for the first series is bigger for the symmetric model, as well as the correlation persistence. Without the asymmetry parameters (ϕ_1, ϕ_2, δ), these persistence parameters incorporate some part of the persistence that otherwise would be captured by the asymmetry parameters. As expected, the average number of non-empty clusters for the mixture distribution $z^\star$ is greater for the symmetric (M_0) model, since it tries to compensate the absence of the asymmetry parameters by including extra mixture components. The same conclusion is reflected in the parameter A, which is greater for the symmetric model meaning that the probability of observing an extra mixture component is greater for the BNP-DCC model. Figure 3.4 presents the traces, histograms and running mean plots of the parameter A. The autocorrelation is smaller in the BNP-ADCC model.

Next, Figure 3.5 compares the predictive densities of the one-step-ahead returns r_{T+1}. The densities are rather similar, but they present important differences in the tail behavior. These differences can be better understood from Figure 3.6, where the marginal log-predictive densities are presented. The log-predictive for Apple Inc. data is almost identical for both models, whereas the log-predictive for the NASDAQ Ind. data has fatter tails for the asymmetric model M_1. Therefore, it seems that although the DPM

TABLE 3.3: Estimation results for Apple Inc. (1) and NASDAQ Ind. (2) returns assuming fully symmetric (M_0) and asymmetric (M_1) multivariate GARCH models.

	BNP-DCC (M_0)		BNP-ADCC (M_1)	
	Mean	95% CI	Mean	95% CI
ω_1	0.0653	(0.0228, 0.1305)	0.1344	(0.0738, 0.2097)
ω_2	0.0072	(0.0034, 0.0131)	0.0095	(0.0050, 0.0217)
α_1	0.0527	(0.0312, 0.0763)	0.0608	(0.0412, 0.0790)
α_2	0.0269	(0.0164, 0.0411)	0.0060	(0.0007, 0.0128)
β_1	0.9306	(0.9027, 0.9532)	0.8950	(0.8565, 0.9188)
β_2	0.9198	(0.9049, 0.9375)	0.9235	(0.9079, 0.9378)
ϕ_1			0.0500	(0.0216, 0.0765)
ϕ_2			0.0393	(0.0250, 0.0700)
κ	0.0211	(0.0064, 0.0361)	0.0213	(0.0068, 0.0392)
λ	0.9080	(0.7995, 0.9854)	0.8494	(0.7855, 0.9185)
δ			0.0285	(0.0036, 0.0557)
$z^\star$	8.1838	(4.0000, 14.0000)	7.3411	(4.0000, 12.0000)
A	0.4933	(0.2806, 0.6839)	0.4726	(0.2650, 0.6596)

FIGURE 3.4: Traces, histograms and running mean plots of $A = c/(1+c)$ for fully symmetric BNP-DCC (M_0) and asymmetric BNP-ADCC (M_1) models.



in the asymmetric model M_1 uses fewer mixture components, it can better capture the heavy tails in one-step-ahead predictive densities of the returns.

Next, similar to Jensen & Maheu (2013), we compare the two estimated models using predictive likelihoods based on a small set of out-of-sample observations

FIGURE 3.5: Contours of the predictive densities for r_{T+1} for fully symmetric BNP-DCC (M_0) and asymmetric BNP-ADCC (M_1) models.

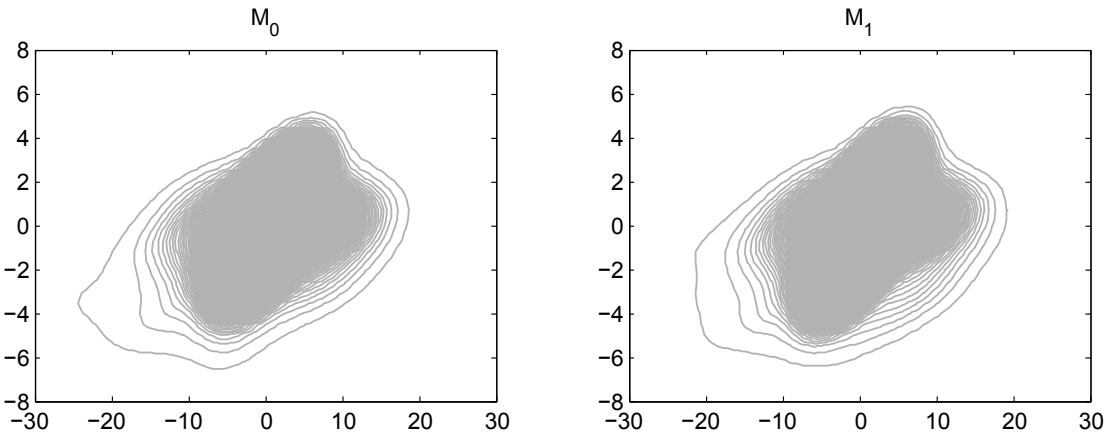
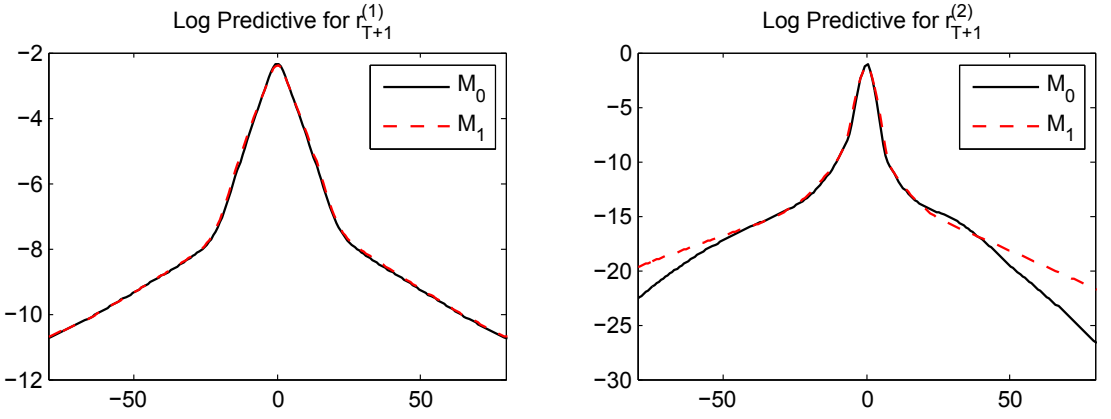


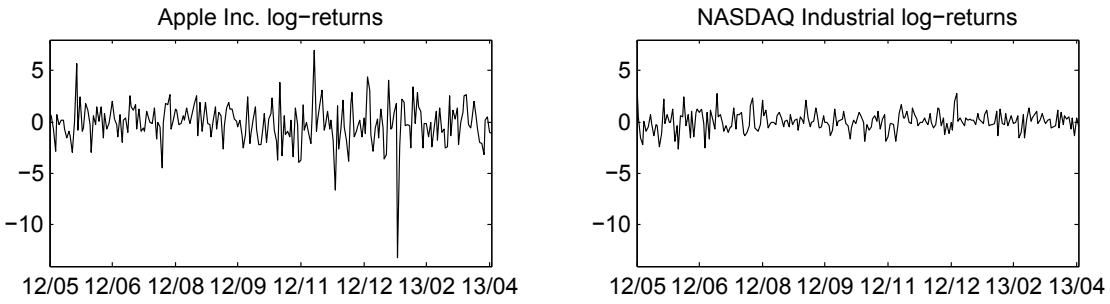
FIGURE 3.6: Marginal log-predictive densities for the one-step-ahead Apple Inc. ($r^{(1)}_{T+1}$) and NASDAQ Ind. ($r^{(2)}_{T+1}$) return data, for fully symmetric BNP-DCC (M_0) and asymmetric BNP-ADCC (M_1) models.



$\{T+1,\ldots,T+\tau\}$, where $\tau = 233$, the end date of the series is April 5, 2013, see Figure 3.7. However, differently from Jensen & Maheu (2013), we do not re-estimate the model whenever a new observation arrives to avoid an increase in the computational cost, but we use the already estimated model parameters up to time T. This results into approximated predictive likelihoods rather than pure predictive likelihoods as considered in Jensen & Maheu (2014). Since predictive likelihoods are very sensitive to the behavior of the last return used on the model's estimation, the obtained results can be generalized only to some extent, and a full valid model comparison still needs to be carried out.

The ratio of predictive likelihoods is called Bayes factor. Since in this chapter we are

FIGURE 3.7: Log-returns (in %) of Apple Inc. and NASDAQ Ind. index for $t = 3099, \ldots, 3331$.



considering approximated predictive likelihoods, we will refer as pseudo Bayes factor to the resulting ratios. As seen in Kass & Raftery (1995), the predictive likelihood for the i^{th} data point, given the k^{th} model, can be obtained as

$$p(r_{T+i}|r^{T+i-1}, M_k) = \int p(r_{T+i}|r^{T+i-1}, M_k, \Theta_k)\pi(\Theta_k|r^{T+i-1}, M_k)d\Theta_k,$$

where Θ_k is a set of parameters associated with the k^{th} model. Since this integral is not analytically tractable, we can approximate it using the MCMC output:

$$p(r_{T+i}|r^{T+i-1}, M_k) = \frac{1}{M}\sum_{m=1}^{M} p(r_{T+i}|r^{T+i-1}, M_k, \Theta_k^{(m)}), \text{ for } i = 1, 2, \ldots, \tau,$$

and then calculate the sum of the logarithms over the entire out-of-sample period:

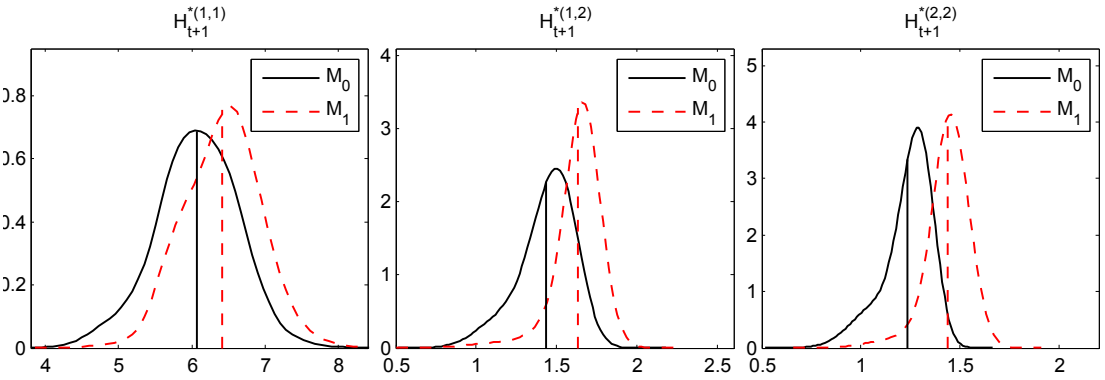
$$\log p(r_{T+1}, \ldots, r_{T+\tau}|M_k) = \sum_{i=1}^{\tau} \log p(r_{T+i}|r^{T+i-1}, M_k).$$

Table 3.4 presents the cumulative approximated log-predictive likelihood for the two models using $\tau = 233$ out-of-sample observations. The log of the pseudo Bayes factor can be obtained as a difference of the likelihoods:

$$\log BF_{10} = \log p(r_{T+1}, \ldots, r_{T+\tau}|M_1) - \log p(r_{T+1}, \ldots, r_{T+\tau}|M_0),$$

where BF_{ij} represents a pseudo Bayes Factor comparing models i and j. As seen in Kass & Raftery (1995), if $2\log BF_{10} \geq 10$, then M_1 is strongly preferred to M_0, since the

FIGURE 3.8: Posterior distributions of one-step-ahead volatilities for fully symmetric BNP-DCC (M_0) and asymmetric BNP-ADCC (M_1) models.



former can predict the data better. In our case, $2\log BF_{10} = 9.9876$, which favors the fully asymmetric BNP-ADCC model.

TABLE 3.4: Cumulative log-predictive likelihoods for fully symmetric BNP-DCC (M_0) and asymmetric BNP-ADCC (M_1) models.

Model	$\log p(r_{T+1}, \ldots, r_{T+\tau} \mid r^T)$
M_0	-751.5094
M_1	-746.5156
# of out-of-sample obs.	$\tau = 233$

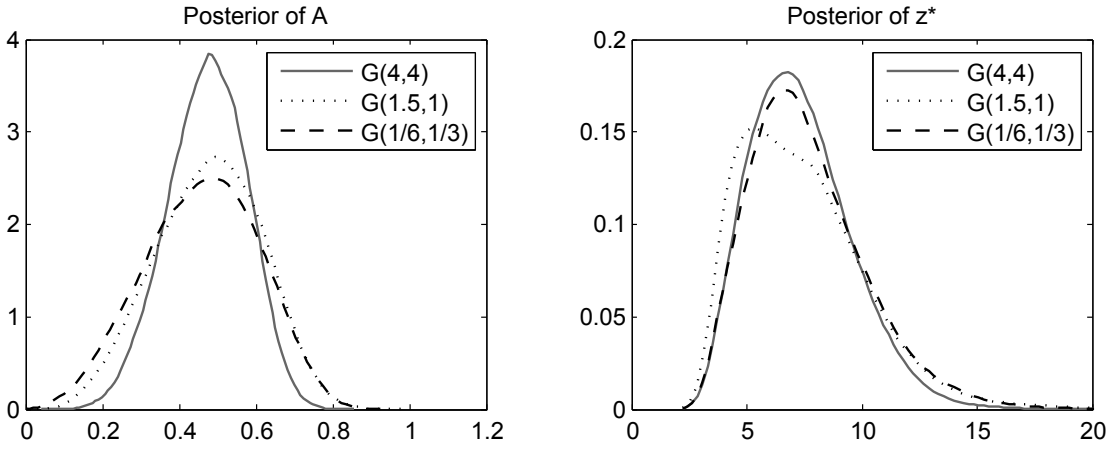
Figure 3.8 draws the posterior densities of the volatilities and Table 3.5 presents the posterior means, medians and confidence intervals for the elements of the one-step-ahead volatility matrix. These are obtained using Equation 3.17 and the explanations given in Section 3.1.3. Even though the posterior densities are of similar shape, the CI width in all cases is smaller for the asymmetric model M_1. It seems that by incorporating the asymmetric assumption in the model we may obtain more precision in the predicted volatilities and, consequently, in optimal portfolio weights and variances.

Finally, we have performed a sensitivity analysis on the posterior distribution of $A = c/(1+c)$ and number of non-empty clusters $z^\star$ by changing the hyper-parameters of the concentration parameter c. In the simulation study and real data application we have assumed $c \sim \mathcal{G}(4,4)$. In the sensitivity analysis we have tried two more combinations: $c \sim \mathcal{G}(1.5,1)$ and $c \sim \mathcal{G}(1/6,1/3)$, such that *a priori* c would center at 0.5 and 1.5, respectively. Table 3.6, Figure 3.9 and Figure 3.10 present the estimation results, which

TABLE 3.5: Posterior means, medians and confidence intervals for the elements of the one-step-ahead volatility matrix for fully symmetric BNP-DCC (M_0) and asymmetric BNP-ADCC (M_1) models.

	BNP-DCC (M_0)		BNP-ADCC (M_1)	
	Mean	95% HPD	Mean	95% HPD
	Median		Median	
$H_{T+1}^{\star(1,1)}$	6.0656	(4.8345, 7.1482)	6.4182	(5.3852, 7.4298)
	6.0780		6.4377	
$H_{T+1}^{\star(1,2)}$	1.4350	(1.0478, 1.7431)	1.6296	(1.3340, 1.8923)
	1.4614		1.6504	
$H_{T+1}^{\star(2,2)}$	1.2346	(0.9647, 1.4132)	1.4369	(1.2345, 1.6322)
	1.2636		1.4469	

FIGURE 3.9: Posterior distributions for $A = c/(1+c)$ and a number of non-empty clusters $z^\star$ for different hyper-parameters for $c \sim \mathcal{G}(a_0, b_0)$ for BNP-ADCC model for Apple-NASDAQ data.

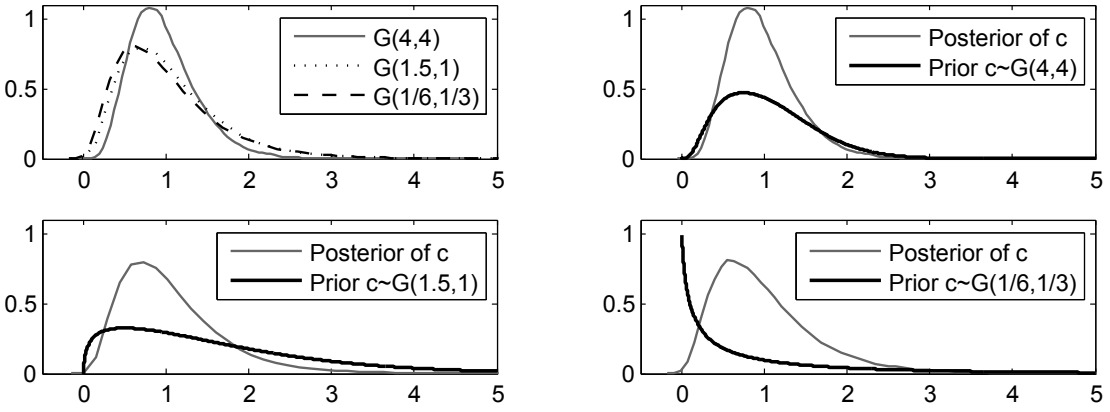


seem to be rather robust to the changes in hyper-parameters for the prior of c as long as the priors are not very informative.

TABLE 3.6: Posterior means and 95% credible intervals for $A = c/(1+c)$ and a number of non-empty clusters $z^\star$ for different hyper-parameters for $c \sim \mathcal{G}(a_0, b_0)$ for BNP-ADCC model for Apple-NASDAQ data.

	$c \sim \mathcal{G}(4,4)$	$c \sim \mathcal{G}(1.5,1)$	$c \sim \mathcal{G}(1/6,1/3)$
Mean of A	0.4726	0.4766	0.4588
95% CI of A	(0.2650, 0.6596)	(0.2011, 0.7305)	(0.1648, 0.7296)
Mean of $z^\star$	7.3411	7.4071	7.6391
95% CI of $z^\star$	(4, 12)	(4, 14)	(4, 14)

FIGURE 3.10: Prior and posterior distributions for c for BNP-ADCC model for Apple-NASDAQ data.



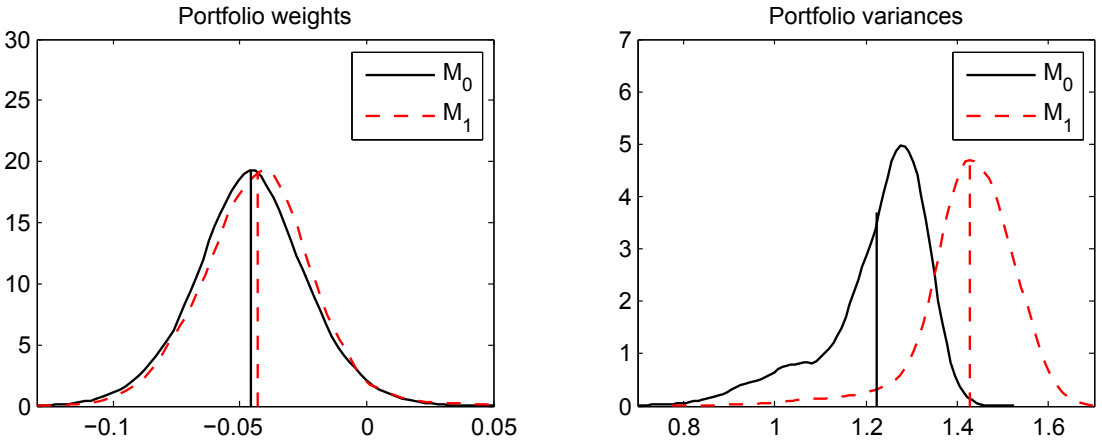
3.4.2 Portfolio allocation

Here we are interested in estimating the GMV optimal portfolio of the two real assets, without the short-sale constraint, using the procedure described in Section 3.2. Firstly, we will make predictions on the optimal one-step-ahead portfolio and then, we will consider all the 233 out-of-sample future observations, adjusting the optimal portfolio weights at each time period. Throughout the portfolio allocation exercise, we report only the first portfolio weight $p_{T+1}^{\star(1)}$, as the other can be obtained as $p_{T+1}^{\star(2)} = 1 - p_{T+1}^{\star(1)}$. The estimation results for the $T+1$ period are presented in Table 3.7 and the posterior densities for optimal portfolio weights and variances can be seen in Figure 3.11. The point estimates for the optimal portfolio weights for $T+1$ period are very similar, however, the asymmetric model estimates greater portfolio variance for $T+1$ since the estimated one-step-ahead volatilites for the BNP-ADCC model are greater. However, note that, as before, the CI width is smaller for the M_1 model.

TABLE 3.7: Posterior mean, median and 95% credible intervals for the optimal one-step-ahead portfolio weight and variance.

	BNP-DCC (M_0)		BNP-ADCC (M_1)	
	Mean Median	95% CI	Mean Median	95% CI
$p_{T+1}^{\star}$	-0.0457 -0.0457	(-0.0907, -0.0011)	-0.0428 -0.0426	(-0.0869, 0.0017)
$\sigma_{T+1}^{2\star}$	1.2234 1.2522	(1.1570, 1.5924)	1.4266 1.4352	(1.5175, 1.7722)

FIGURE 3.11: Posterior distributions of one-step-ahead optimal portfolio weights $p^{\star}_{T+1}$ and overall portfolio variances $\sigma^{2\star}_{T+1}$ for fully symmetric BNP-DCC (M_0) and asymmetric BNP-ADCC (M_1) models .



Next, we estimate the optimal portfolio weights for the entire out-of-sample period of 233 observations. Figure 3.12 and Figure 3.13 present the dynamics of the estimated portfolio weights and variances with corresponding CIs for each of the models. It shows that along time the mean portfolio weights are rather similar across the two models. Since the asymmetric model can predict the data better, the estimated portfolio weights for the M_1 model would be more precise. NASDAQ Ind. exhibits more volatile behavior in the first half of the data set, as compared to the second half, meanwhile Apple Inc. returns behave in a reverse manner - the first half is less volatile than the second, see Figure 3.7. This is is reflected in the optimal portfolio allocation: till around 12/10 Apple Inc. has a positive portfolio weight, see Figure 3.12, and from 12/10 the optimal decision is to short-sell Apple Inc. shares. Also, looking at the model parameters, greater differences in marginal volatilities between models M_0 and M_1 are observed for negative returns. Therefore, greater differences in portfolio weights and variances (Figure 3.12 and Figure 3.13) between models M_0 and M_1 should be observed when the returns are negative. Finally, the BNP-ADCC model has thinner CIs (8 to 9 percent on average), as seen in Figure 3.14, making the point estimation more precise.

To sum up, these portfolio allocation exercises helped to illustrate the consequences for financial decisions of assuming different models for return series. The DPM model permits the investor to perform inference and prediction about the returns and their volatilities without imposing arbitrary restrictions on the data generating process. Additionally, the use of asymmetric model results into more precise point estimates and

FIGURE 3.12: A sequence of portfolio weights and their corresponding 95% CIs for $\tau = 1, \ldots, 233$ for fully symmetric BNP-DCC (M_0) and asymmetric BNP-ADCC (M_1) models.

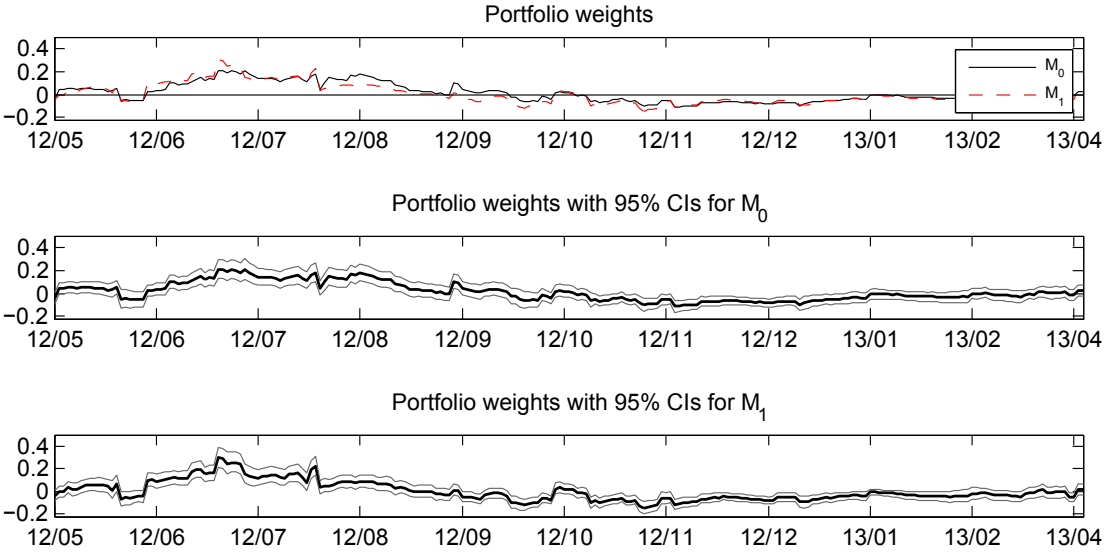


FIGURE 3.13: A sequence of portfolio variances and their corresponding 95% CIs for $\tau = 1, \ldots, 233$ for fully symmetric BNP-DCC (M_0) and asymmetric BNP-ADCC (M_1) models.

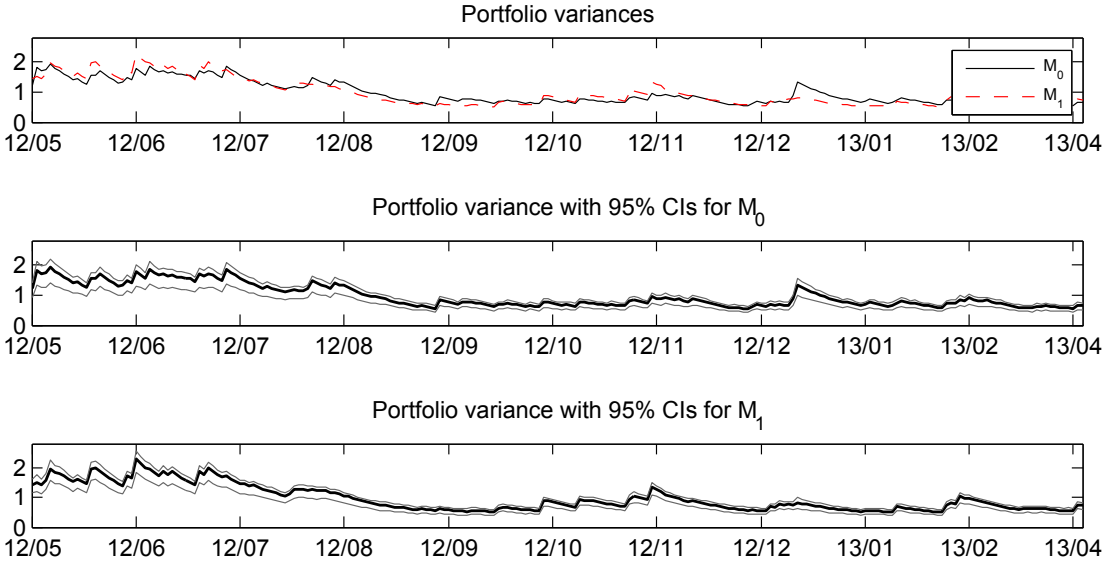
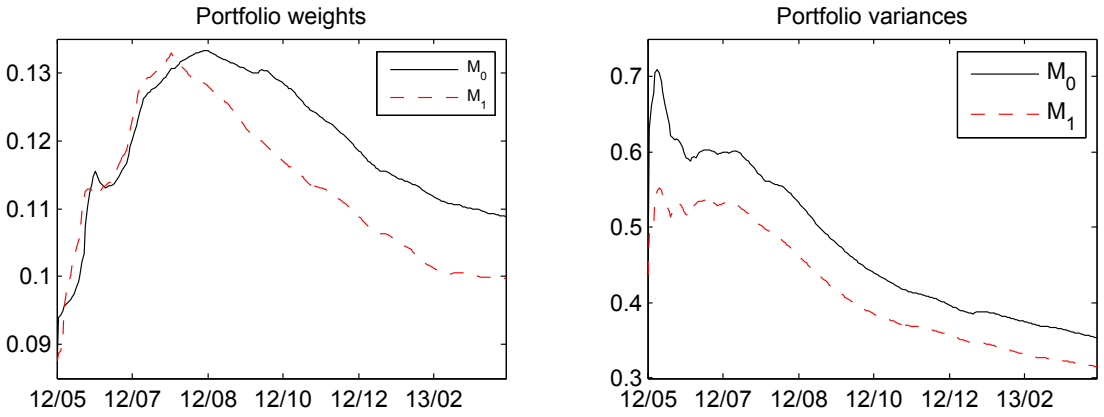


FIGURE 3.14: Mean cumsum of the of 95% CI width for the optimal portfolio weights and variances for $\tau = 1, \ldots, 233$ for fully symmetric BNP-DCC (M_0) and asymmetric BNP-ADCC (M_1) models.



better predictive performance. In the portfolio allocation context, adjusting portfolio weights at each period might lead to high transaction costs, thus the investor will adjust her portfolio only if the expected utility after the adjustment minus the transaction costs is greater than the expected utility without the adjustment.

On the other hand, an essential issue in choosing more complicated models versus the simple ones is the ability to handle numerous assets, especially in financial applications, such as portfolio optimization. The DPM model is very flexible in this sense, since the general specification described before contains numerous simplified models. For example, it clearly contains the single Gaussian as a special case when the first mixture weight is equal to one. Also, it is possible to impose a symmetric distribution for the innovations by simply assuming that the mixture means are all equal and, in particular, it could be reasonable to impose $\mu_j = 0$, for $j = 1, 2, \ldots$. If we further assume that the precision matrices are all diagonal, $\Lambda_j = \text{diag}\left(\lambda_{j1}, \ldots, \lambda_{jK}\right)$, this will lead to uncorrelated innovations. Finally, we could in addition assume that the diagonal elements of each precision matrix are all equal by considering $\Lambda_j = \lambda_j I_K$. In this chapter we have used the full version of the DPM model to illustrate the flexibility of it. The adaptation of the model to these particular cases in order to simplify the problem of many assets is straightforward from the theoretical point of view, but it might present a heavy computational burden and some additional issues, such as convergence to a stable posterior.

3.5 Conclusions

In this chapter we have proposed a Bayesian non-parametric approach for modeling the distribution of multiple returns. We have used a BNP-ADCC model to explain the individual volatilities and the time-varying correlations, taking into consideration the asymmetries in individual assets' volatilities, as well as in the correlations. The errors are modeled using a location-scale mixture of infinite Gaussian distributions that has been shown to allow for a great flexibility in the return distribution in terms of skewness and kurtosis. An MCMC method has been presented for model estimation and prediction. For that, DPM prior has been given to the infinite mixture of multivariate Gaussian distribution. We have presented a short simulation study that illustrates the differences arising from different assumptions for the errors and shows the adaptability of the DPM model. The simulation results suggest that the proposed approach appears to be able to fit adequately several frequently used distributions. Finally, we have presented a real data application that compares the proposed BNP-ADCC with a fully symmetric BNP-DCC model by using return series of Apple Inc. and NASDAQ Industrial index. Model comparison via approximated log-predictive likelihood favors the asymmetric BNP-ADCC, which also produces thinner credible intervals for one-step-ahead volatilities. Additionally, we have employed the proposed approach to solve a portfolio allocation problem. The explained methodology and obtained results are not limited to this specific risk management problem and could be expanded into various other topics in applied finance and risk management.

Chapter 4

A Bayesian Non-Parametric Approach to a MSSV Model with Particle Learning

The previously reviewed Stochastic Volatility 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 & Maheu (2010, 2014) and Delatola & 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. However, MCMC methods in general are computationally demanding and inherently non-sequential (Lopes & 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 chapter 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, Johannes, Lopes & Polson (2010). 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 chapter we design a PL algorithm for a SV model with DPM innovations, referred to as a SV-DPM, similar to that of Delatola & 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 chapter we augment the SV-DPM model by incorporating Markov switching jumps, as seen in So et al. (1998) and Carvalho & 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 chapter is structured as follows. Section 4.1 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 4.2 introduces a new MSSV-DPM model and designs a PL algorithm for inference and prediction. Section 4.3 compares the performance of the two non-parametric models by using real data. Finally, Section 4.4 concludes.

4.1 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.

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:

$$r_t = h_t + \epsilon_t, \text{ where } \epsilon_t \sim \mathcal{F}, \tag{4.1}$$
$$h_t = \alpha + \beta h_{t-1} + \tau \eta_t, \text{ where } \eta_t \sim \mathcal{N}(0,1). \tag{4.2}$$

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 & 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.

Non-Normal errors

Delatola & 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 & 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 Equation 4.1 and Equation 4.2 is slightly different from the one in Delatola & Griffin (2011), since we do not sum the constant volatility parameter α into the mixture. We leave this constant separate since in Section 4.2 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 & Maheu

(2010, 2013, 2014), Delatola & Griffin (2011, 2013), Kalli et al. (2013), and Ausín et al. (2014). This type of approach is known as time-invariant (independent) DPM.

As seen in Escobar & 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, $\epsilon_t \sim \mathcal{N}(\mu_t, \sigma_t^2)$, 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 Equation 4.1 and Equation 4.2 can be rewritten as follows:

$$r_t|h_t, \Theta \sim \frac{c}{c+t-1}\mathcal{N}(r_t; \mu_0 + h_t, \sigma_0^2) + \frac{1}{c+t-1}\sum_{j=1}^{L_{t-1}^{\star}} n_{t-1,j}\mathcal{N}(r_t; \mu_j + h_t, \sigma_j^2), \tag{4.3}$$

$$h_t|h_{t-1}, \Theta \sim \mathcal{N}(h_t; \alpha + \beta h_{t-1}, \tau^2), \tag{4.4}$$

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 $L_t^\star \leq t$.

4.1.1 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 & 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 & 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 & 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 4.1). The remainder of the model parameters are sampled via an extension of Gibbs sampler, called hybrid Gibbs sampler. In their subsequent paper, Delatola & 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.

4.1.2 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, Johannes, Lopes & Polson (2010), 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 & Tsay (2011) and Lopes & Carvalho (2013). For a more detailed explanation of PL with illustrations refer to Carvalho, Johannes, Lopes & Polson (2010) 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

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

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

2. **Sampling.**

   (a) Sample new log-volatilities h_t from

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

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 $\{1, \ldots, L_{t-1}^\star + 1\}$, with weights proportional to

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

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 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

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

4.1.3 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 $\log \chi_1^2$ distribution. Simulated returns can be recovered as $y_t = \exp\{r_t/2\}$. MCMC results are obtained via `Matlab` code of Delatola & 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 ours, since Delatola & 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 observations. 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 4.1.

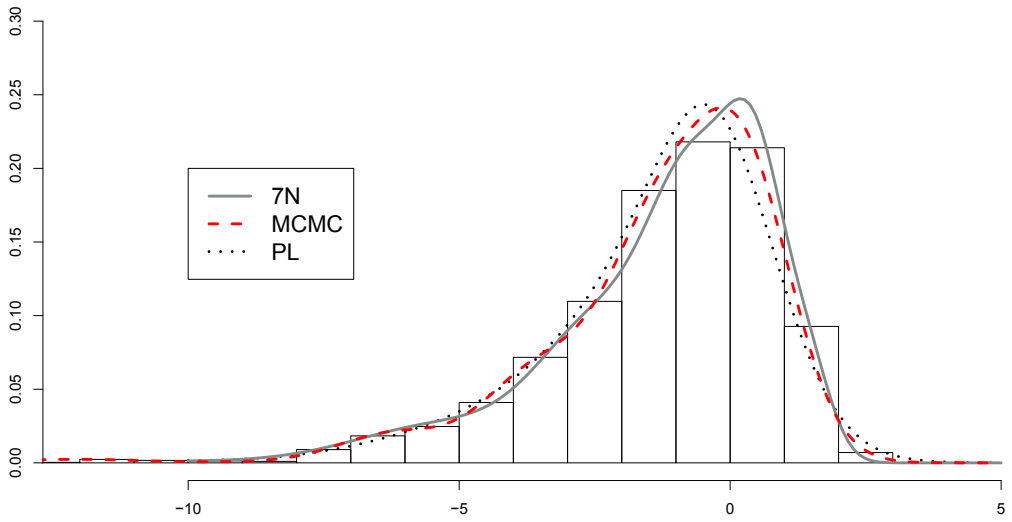
TABLE 4.1: CPU time in seconds for MCMC and PL.

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

Next, we compare the posterior output for both estimation methods graphically in Figure 4.1, Figure 4.2 and Figure 4.3 and Table 4.2. Figure 4.1 draws the estimated density at $T = 3000$ for the log of the squared returns for PL and MCMC, compared to the true

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

FIGURE 4.1: 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 $T = 3000$.



one. Both estimations seem reasonable and very close to the true data generating density. Table 4.2 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.

TABLE 4.2: 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
$\beta = 0.98$	1000	0.9616	(0.9368, 0.9826)	0.9671	(0.9464, 0.9841)
	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)
$\tau^2 = 0.10$	1000	0.1303	(0.0856, 0.1887)	0.1060	(0.0705, 0.1457)
	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)

Figure 4.2 presents the posterior distributions for the log-volatilities at three different

FIGURE 4.2: Posterior distributions of the log-volatilities for MCMC and PL for $T = 1000, 2000$ and 3000.

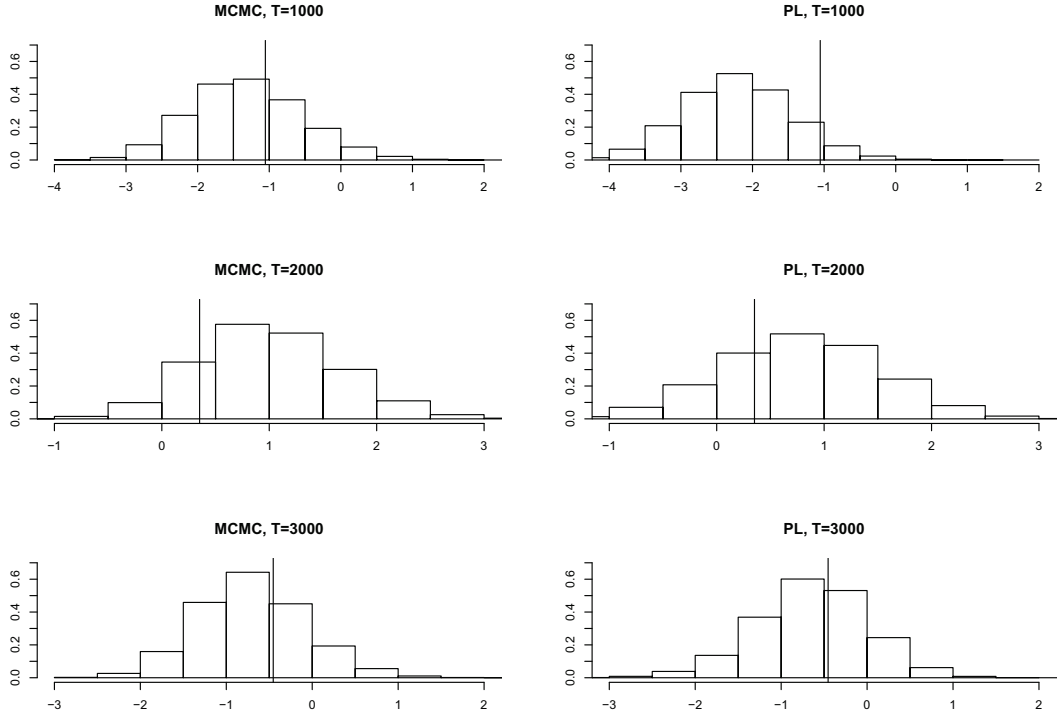
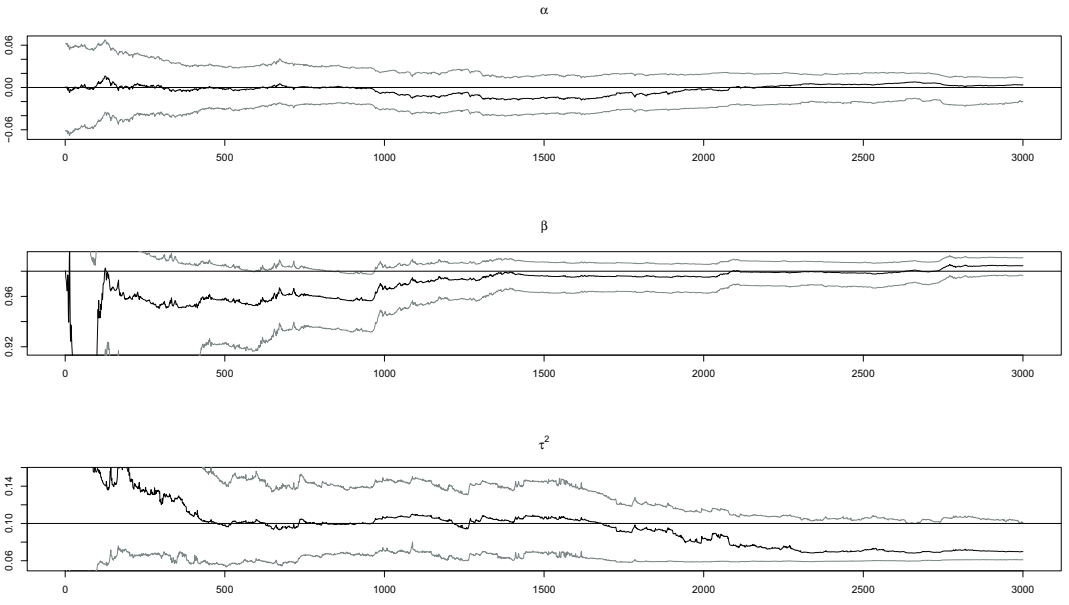


FIGURE 4.3: PL parameter estimates with 95% CI for one run of 300k particles, compared to the true parameter values.



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 4.3 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 & 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.

4.2 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. For a brief review of Markov switching GARCH and SV models refer to Section 2.1 and Section 2.2.

Consider a J-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:

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

As seen in Carvalho & Lopes (2007), we have to introduce the following reparametrization for α_{s_t} in order to avoid identification issues:

$$\alpha_{s_t} = \gamma_1 + \sum_{j=2}^{J} \gamma_j \mathbf{1}\{s_t \geq j\}, \quad \gamma_1 \in \Re \text{ and } \gamma_j > 0 \text{ for } j > 1.$$

Here $\mathbf{1}\{s_t \geq j\}$ is an indicator function that takes values equal to one if $s_t \geq j$ and zero otherwise ($s_t < j$).

In this section we consider a two-state MSSV model, as seen in Carvalho & Lopes (2007), since it is a natural starting point. However, the *a priori* selection of number of states might not fully capture the underlying data structure, since more regimes might be necessary, especially after the 2007-2009 financial crisis. Therefore, a natural extension could be to compare models with different number of volatility states and select the most appropriate one, as has been done by Bauwens et al. (2014) in MS-GARCH setting, among many others. Therefore, in the case of only two regimes the s_t variables follow a two-state first order Markov Process:

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

And the α_{s_t} reparametrization is as follows:

$$\alpha_{s_t} = \gamma_0 + \gamma_1 \mathbf{1}\{s_t = 1\}, \quad \gamma_0 \in \Re \text{ and } \gamma_1 > 0,$$

where the indicator function $\mathbf{1}\{s_t = 1\}$ 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 & Susmel (2004) have proposed a two-factor SV model with regime switching and estimated it using Gibbs sampler. They find that the high volatility persistence is reduced when the regimes are incorporated in the model. Also, the authors compare the new model with other two alternative two-factor models, simple SV and GARCH, and find that SV always outperforms GARCH, both in sample and out of

sample. The regime switching SV performs better than the simple SV in sample, however, out of sample, it is only marginally better. Lopes & 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 & Lopes (2007) have constructed a sequential Monte Carlo filter by combining auxiliary particle filter (APF) with the filter of Liu & 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 Equation 4.3 and Equation 4.4 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.

4.2.1 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

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

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

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

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

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

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

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

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

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

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

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

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

(c.3) Sample p from $\mathcal{B}(p; \alpha^\star_p, \beta^\star_p)$, where

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

(c.4) Sample q from $\mathcal{B}(q; \alpha^\star_q, \beta^\star_q)$, where

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

(c.5) Sample τ^2 from $\mathcal{IG}(\tau^2; b^\star_0/2, b^\star_0\tau^{2\star}_0/2)$, where

$$b^\star_0 = \tilde{b}_0 + 1 \text{ and } b^\star_0\tau^{2\star}_0 = \tilde{b_0}\tilde{\tau}^2_0 + \frac{(\tilde{m}_\beta\tilde{h}_{t-1} - (h_t - \alpha))^2}{1 + \tilde{V}_\beta\tilde{h}^2_{t-1}}.$$

(c.6) Sample β from $\mathcal{TN}_{(-1,1)}(\beta; m^\star_\beta, V^\star_\beta\tau^2)$, where

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

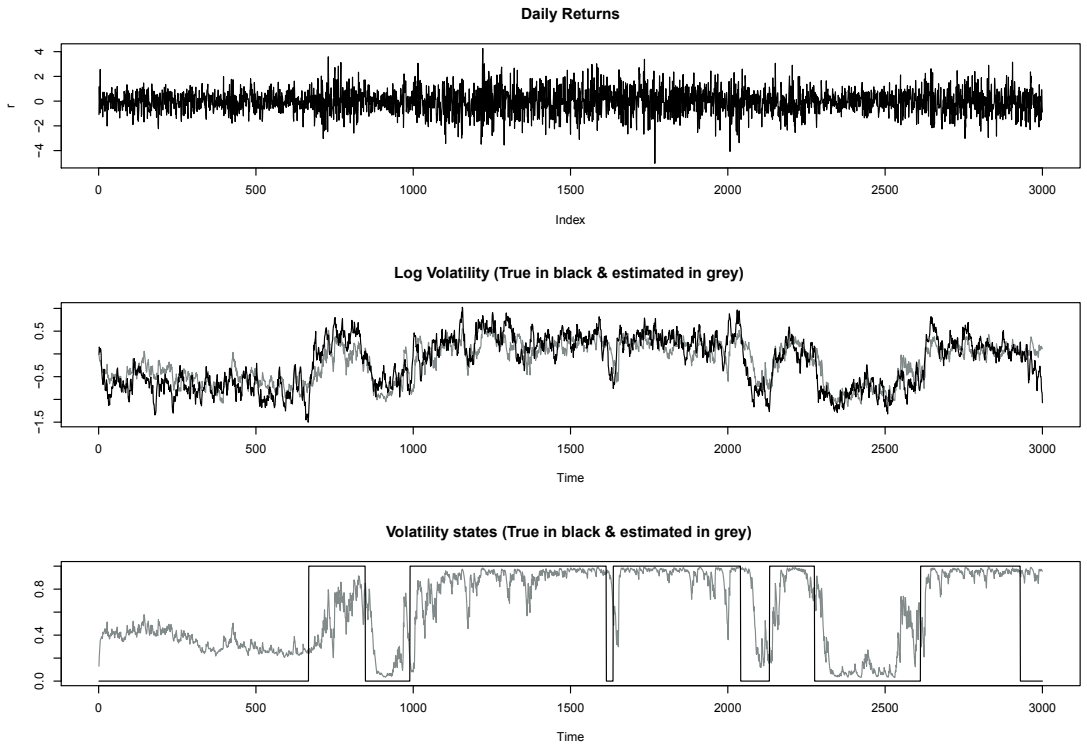
(c.7) Sample σ_{k_t} from $\mathcal{IG}(\sigma^2_{k_t}; a^\star_0/2, a^\star_0\sigma^{2\star}/2)$, where

$$a^\star_0 = \tilde{a}_0 + 1 \text{ and } a^\star_0\sigma^{2\star}_0 = \tilde{a}_0\tilde{\sigma}^2_0 + \frac{(r_t - h_t - \tilde{m}_0)^2}{1 + \tilde{V}_0}.$$

(c.8) Sample μ_{k_t} from $\mathcal{N}(\mu_{k_t}; m^\star_0, V^\star_0\sigma^2)$, where

$$m^\star_0 = \frac{\tilde{m}_0 + \tilde{V}_0(r_t - h_t)}{1 + \tilde{V}_0} \text{ and } V^\star_0 = \frac{\tilde{V}_0}{1 + \tilde{V}_0}.$$

FIGURE 4.4: Simulated data: daily returns (top graph), true and estimated log-volatilities (middle graph) and true and estimated regimes (bottom graph).



4.2.2 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 Figure 4.4, Figure 4.5 and Figure 4.6.

Figure 4.4 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 identify the regimes with the overall miss-classification rate equal to 13%.

FIGURE 4.5: Simulated data: true and estimated density for log-squared return distribution.

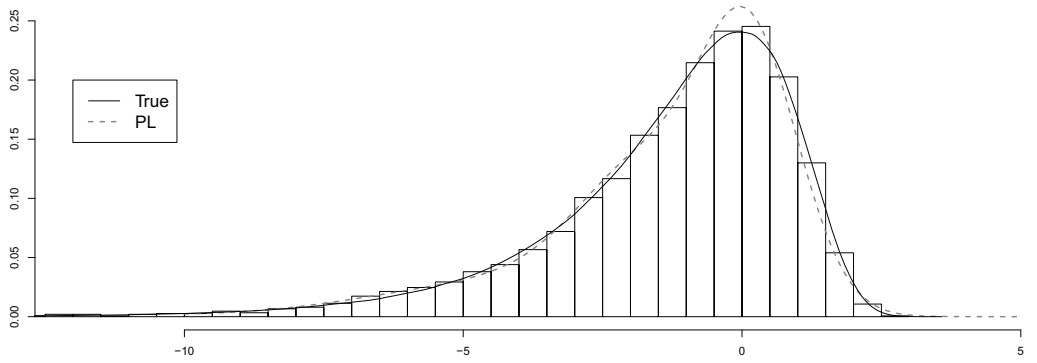


FIGURE 4.6: Simulated data: true and estimated parameters with 95% HPD intervals.

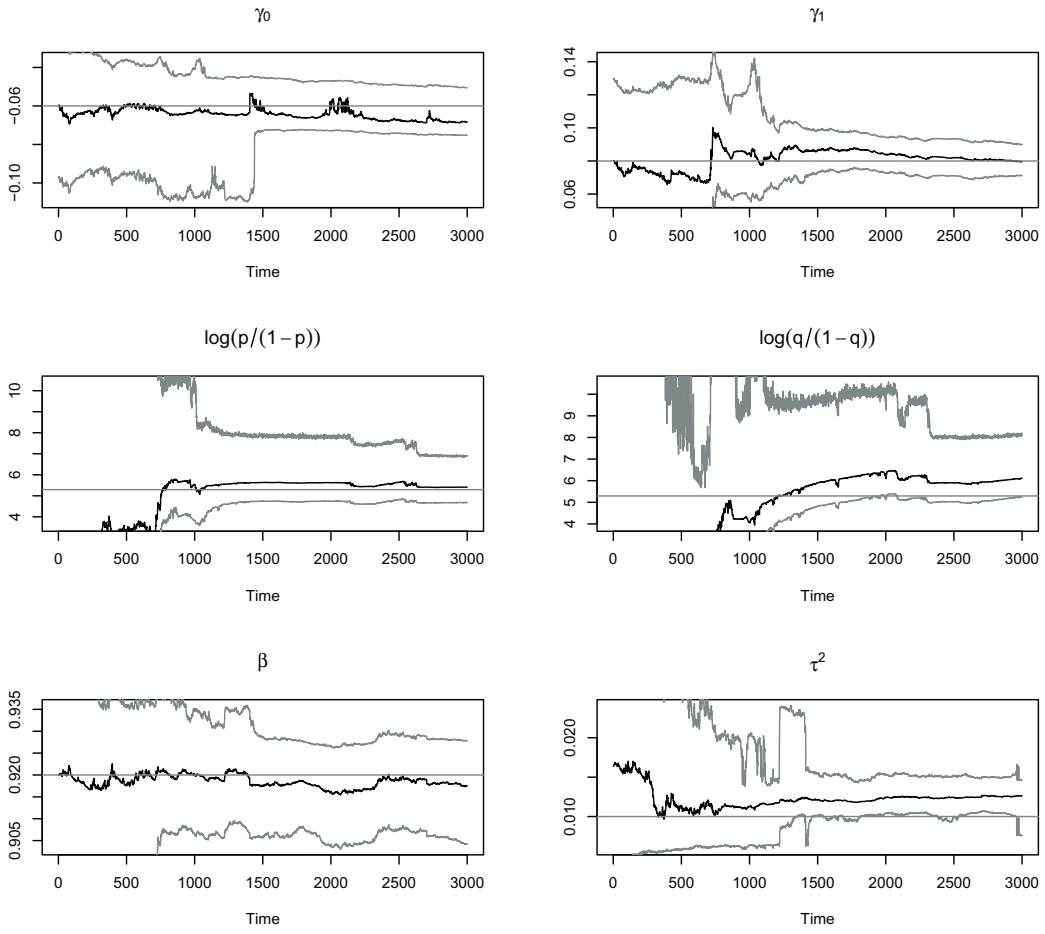


FIGURE 4.7: Simulated data II: daily returns (top graph), true and estimated volatilities (middle graph) and true and estimated regimes (bottom graph).

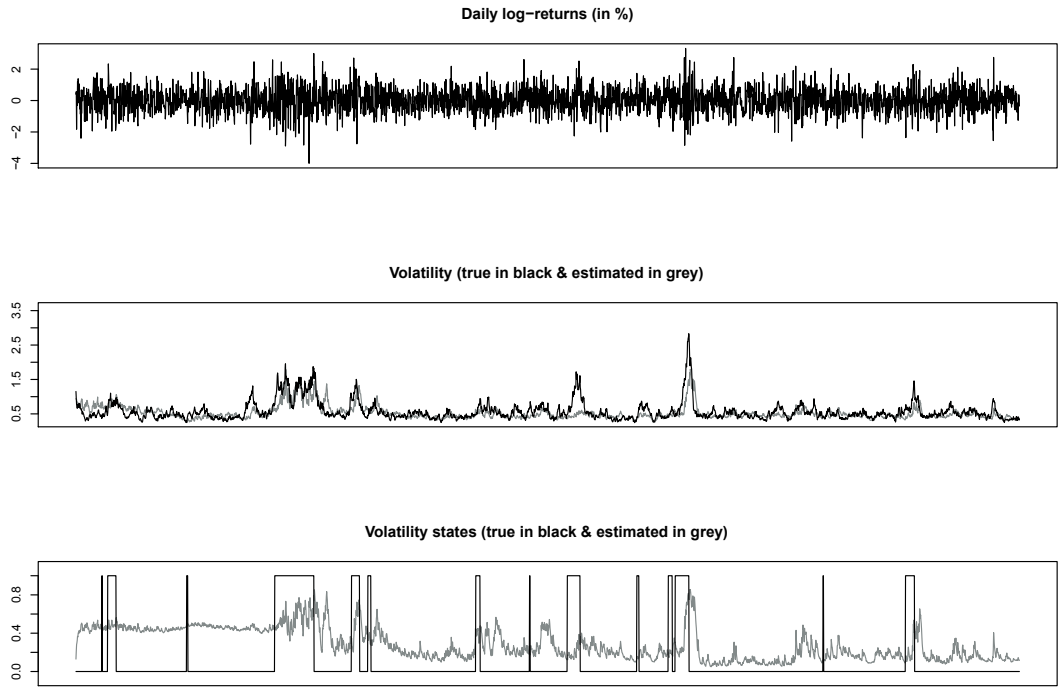
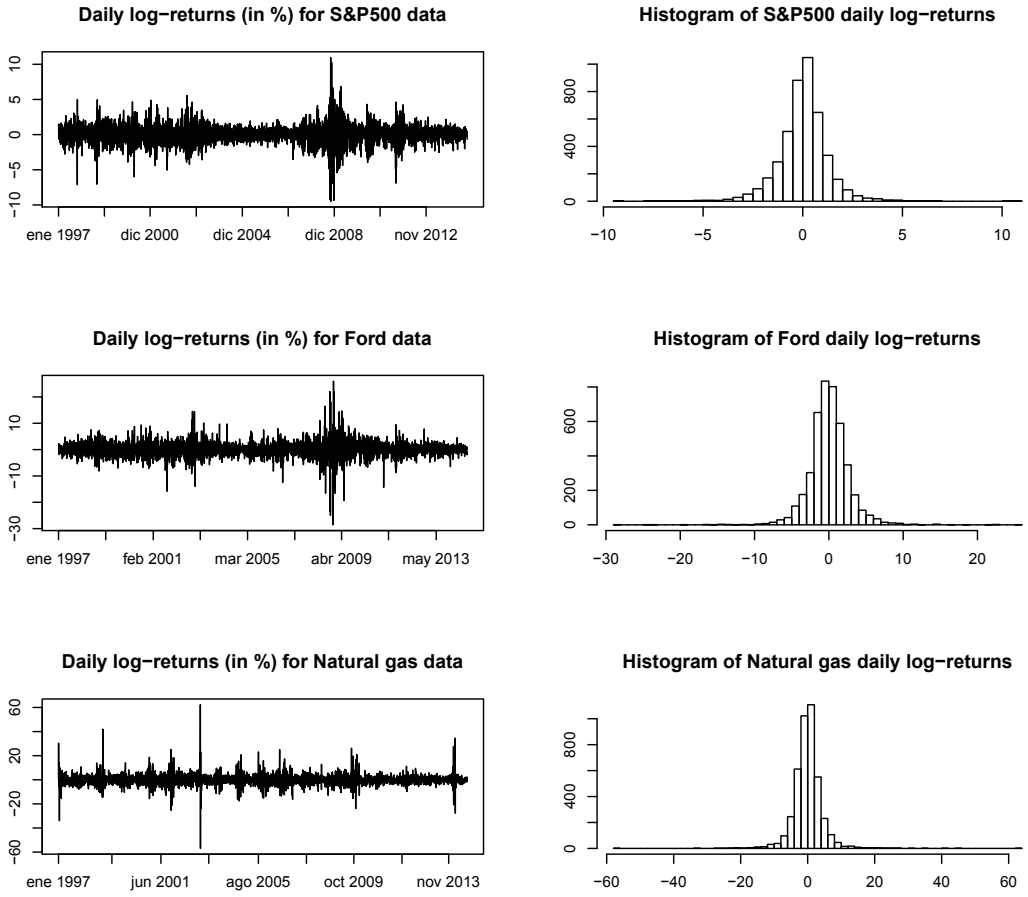


Figure 4.5 draws the true and estimated density for the log-squared returns, which is $\log \chi_1^2$. Finally, Figure 4.6 draws the sequential estimation of the model parameters and their 95% HPD intervals. Overall, the obtained estimation results seem quite reasonable and PL is able to correctly identify volatility regimes, filter log-volatilities, estimate the density of the errors and the parameters in an efficient sequential manner. The classification would become more troublesome once one of the regimes becomes less persistent, and it would even worsen if both of the regimes become less persistent. This would mean very frequent changes in volatility states that the model would not have time to capture properly. See, for example, estimation results of another simulation study in Figure 4.7, where the high volatility state regime is less persistent $q = 0.950$.

4.3 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

FIGURE 4.8: Daily log-returns (in %) and corresponding histograms for S&P500, Ford and Natural gas data.



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 4.3 and Figure 4.8.

TABLE 4.3: 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

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 & 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 & 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.

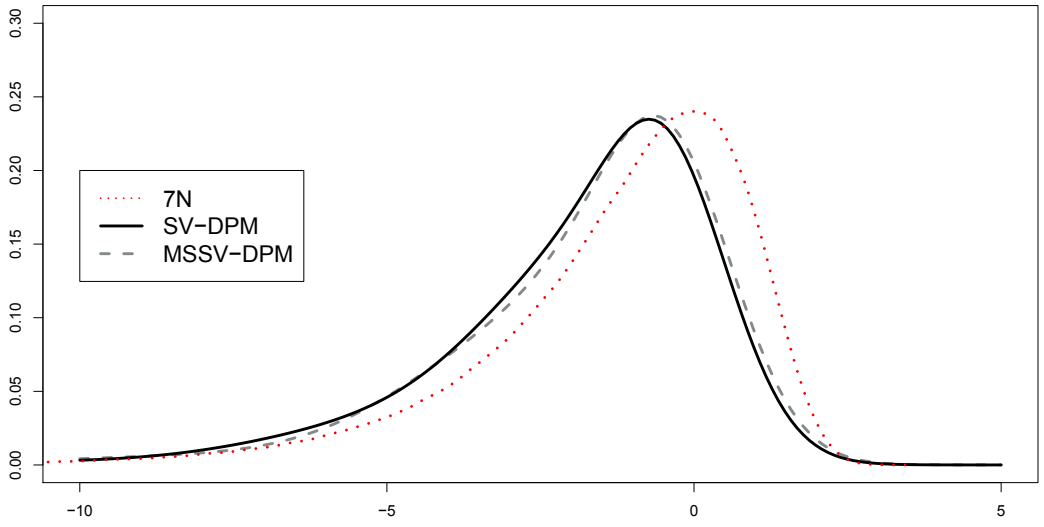
The log-predictive densities are very easy to obtain in SMC setting, since they are a by-product of the estimation procedure and, for each $t = 1, \ldots, T$, are calculated as

$$\log p(r_t|r^{t-1}) = \frac{1}{N}\sum_{i=1}^{N}\log p(r_t|(\Theta, h_t, k_t)^{(i)}). \tag{4.5}$$

Differently than in Delatola & 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 Equation 4.5.

Next, we present the estimation results for the S&P500 data set. Figure 4.9 and Figure 4.10 present estimated predictive densities, filtered volatilities and volatility states and Table 4.4 presents the estimated parameters. Figure 4.9 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

FIGURE 4.9: Estimated densities for the log-squared error term for SV-DPM and MSSV-DPM models.



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 4.10, 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 4.4, the volatility persistence parameter tends to be larger for the SV-DPM model, as expected, see So et al. (1998) and Kalimipalli & Susmel (2004), among others.

TABLE 4.4: 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)

FIGURE 4.10: Filtered volatilities and volatility states for S&P500 data for SV-DPM and MSSV-DPM models.

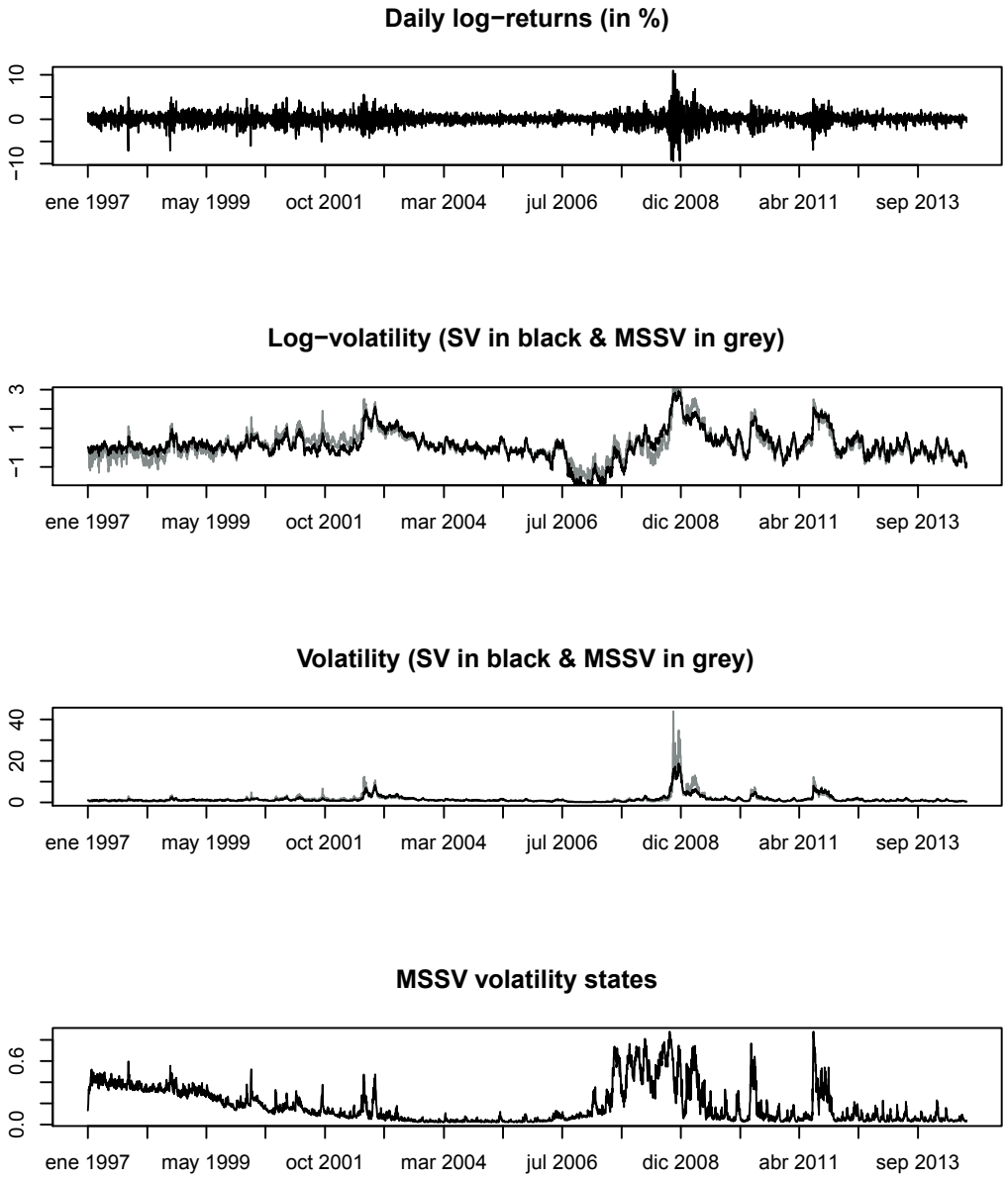


TABLE 4.5: LPS and $LPTS_\alpha$ for SV-DPM and MSSV-DPM for S&P500 data.

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

Table 4.5 presents the LPS and $LPTS_\alpha$ for the S&P500 data. Same as in the paper by Delatola & 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 Table 4.6, Table 4.7, Table 4.8 and Table 4.9 and Figure 4.11 and Figure 4.12. 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.

TABLE 4.6: 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)

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 not only in portfolio allocation setting, but also in risk measurement and management problems, where the agents are usually more interested in the tails than the entire distribution.

TABLE 4.7: LPS and $LPTS_\alpha$ for SV-DPM and MSSV-DPM for Ford data.

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

FIGURE 4.11: Filtered volatilities and volatility states for Ford data for SV-DPM and MSSV-DPM models.

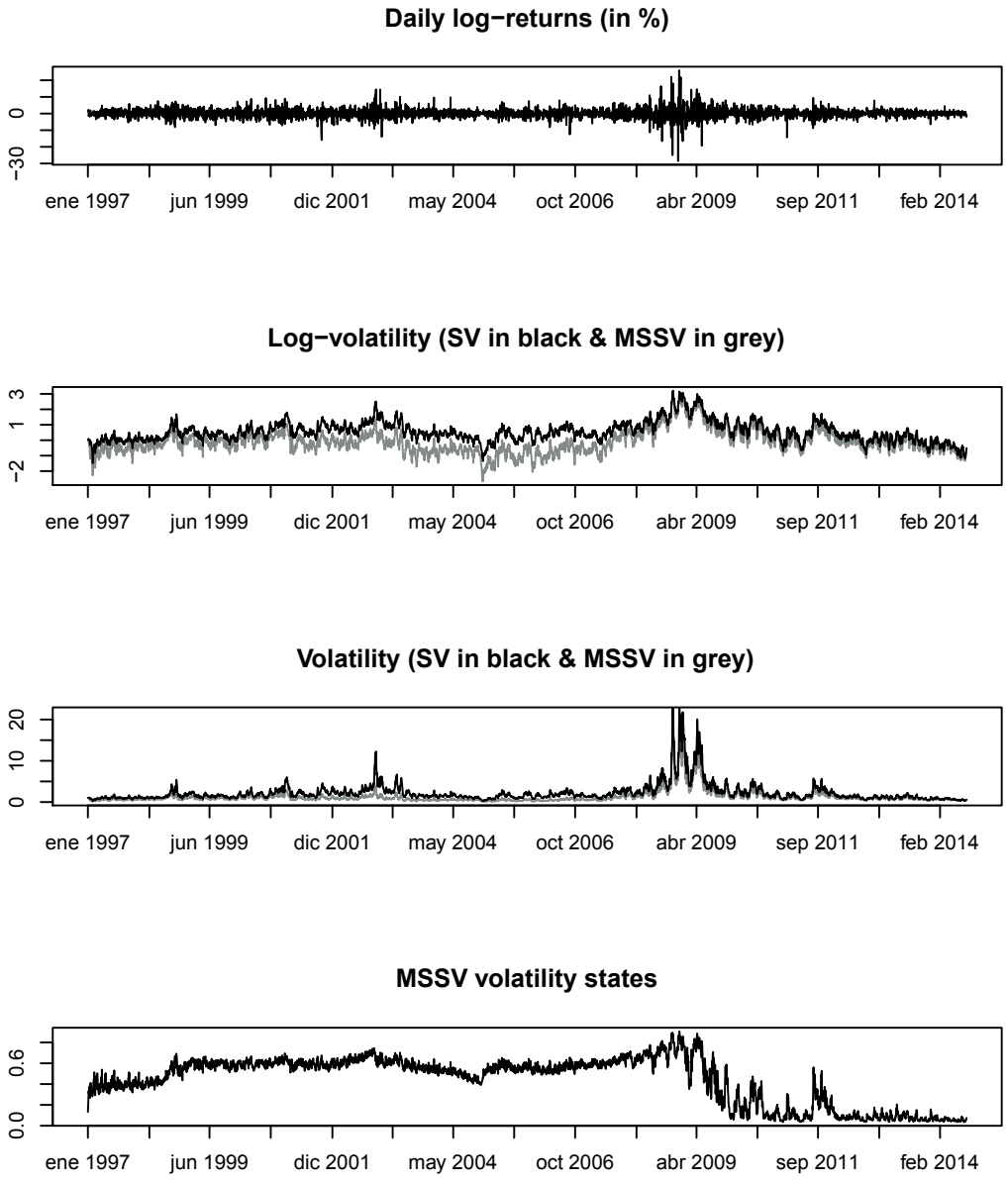


FIGURE 4.12: Filtered volatilities and volatility states for Gas data for SV-DPM and MSSV-DPM models.

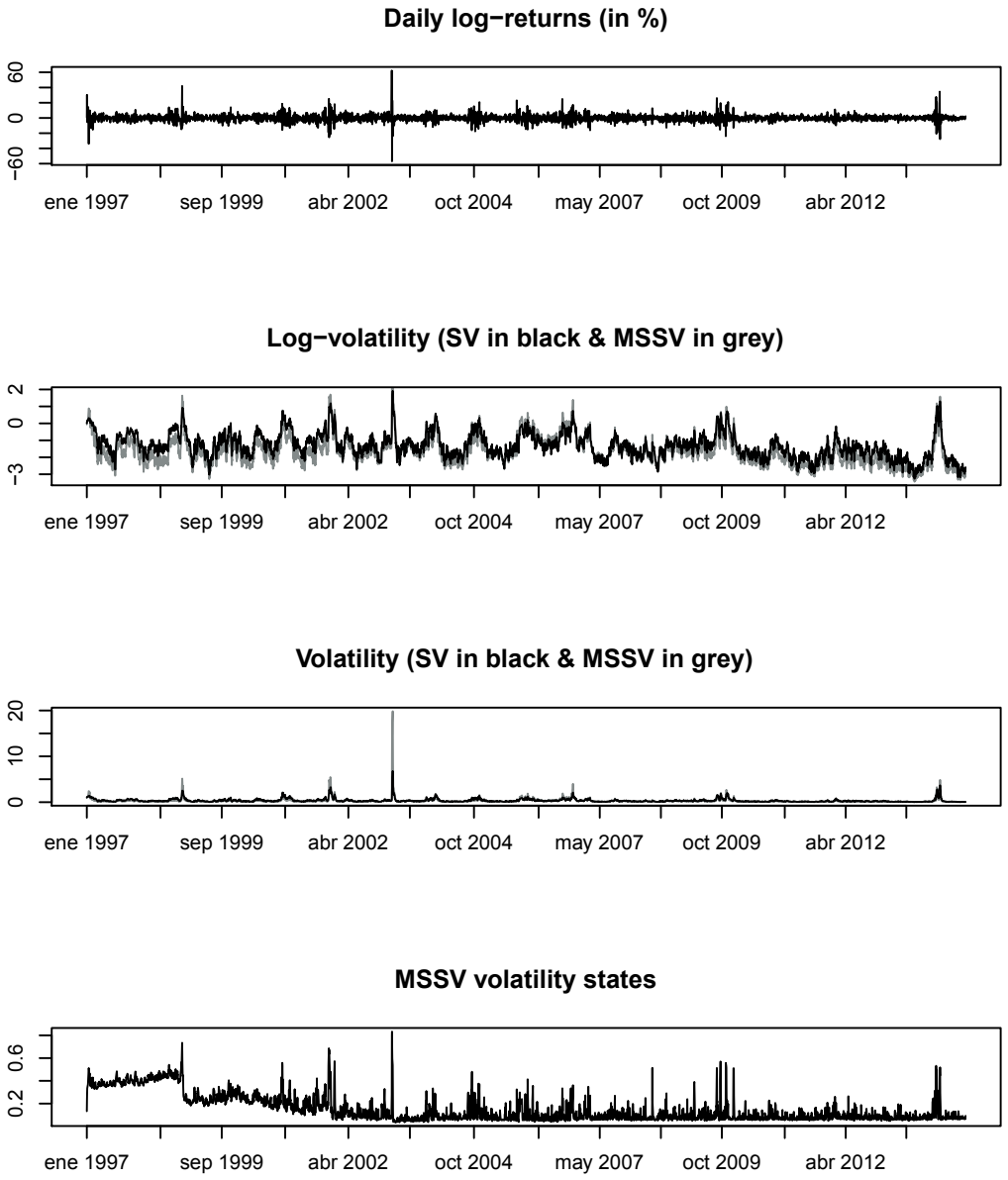


TABLE 4.8: Parameter estimation for SV-DPM and MSSV-DPM models for Gas data at time T.

	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 4.9: 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

4.4 Discussion

This chapter 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 good 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). This result leads to a straightforward extension for future research,

which is comparison of competing parametric and non-parametric SV-type models in terms of VaR and CVaR estimation having in mind the computational cost and increase in the parameter space for more complicated models.

Chapter 5

Conclusions and Extensions

5.1 Conclusions

In this thesis we have put emphasis on two recent developments in time-varying volatility literature: the non-parametric error specification and efficient sequential estimation. We have expanded on time-varying volatility models in two directions. Firstly, we have showed that even though non-parametric errors are very flexible as compared to the parametric specifications, they are still not able to account for the asymmetries in time-varying volatilities and correlations. By using real data we have showed that a Bayesian non-parametric ADCC model has higher predictive power than its symmetric counterpart. Secondly, we have presented the most recent SMC estimation technique - PL. We have designed an efficient PL scheme for the estimation of non-parametric SV model and compared the output with MCMC. PL performs as well as MCMC, however, presents a true advantage in terms of computational costs when a new data point arrives, since it is not necessary to run the entire chain again. We have generalized the SV-DPM model by introducing Markov switching jumps and showed that the new model has higher predictive power in the tails of the distribution. Therefore, even though it has been established that non-parametric error specification outperforms the parametric one, it still cannot account for the asymmetries or jumps in the volatilities.

5.2 Extensions

The first extension concerns Chapter 3. There are many ways to improve model comparison, both, via predictive likelihoods and via VaR/CVaR estimation. The improvement, as already discussed in the paper, would be is not to approximate the log - predictive likelihoods, but to re-estimate the model as new data arrives and obtained the exact log-predictive likelihoods. Also, model performance in the tails is of special interest, since it is relevant in many fields, such as portfolio allocation problems, risk measurement and management. Therefore, one could compare the models in terms of the estimation of VaR and/or CVaR.

The superiority of non-parametric time-varying volatility models over their parametric counterparts has already been established in the previous literature, see Jensen & Maheu (2010, 2013), Kalli et al. (2013), Ausín et al. (2014), Delatola & Griffin (2010). However, as we have seen in the previous chapters, there are some model specifications that the non-parametric errors cannot capture. Those include the asymmetric volatility effect or Markov switching jumps in the volatility process. Therefore, the most natural extension is to consider a non-parametric SV model with leverage effect. In fact, there has been two papers concerning this topic, those are Delatola & Griffin (2013) and Jensen & Maheu (2014). However, in both works the authors use MCMC schemes for inference and prediction, which, as already shown before, are prohibitively costly when new observation arrives. Thus the first extension is to estimate the asymmetric SV-DPM model using sequential Monte Carlo techniques, in particular, Particle Learning.

As already seen in Chapter 4, PL can be seen as a true competitor to MCMC methods and in many cases provides more for less. Therefore, the use of an algorithm that combines auxiliary particle filters (APF) and sufficient statistics (SS) can be extended to more complex models, such as time-varying dependencies between two series, that in turn could be individual asset returns. Thus the second extension concerns sequential estimation of a time-varying copula in order to capture complex dependencies between two series in an efficient manner. Almeida & Czado (2012) have proposed an efficient Bayesian estimation for stochastic time-varying copulas via MCMC. However, differently from the mentioned paper, SMC scheme for the estimation of the parameters

could be developed, since the time-varying copula parameters can be seen as a latent state of a general state-space model.

Bibliography

Alexander, C. O. & Chibumba, A. M. (1997), 'Multivariate Orthogonal Factor GARCH', *Mimeo, University of Sussex* .

Almeida, C. & Czado, C. (2012), 'Efficient Bayesian Inference for Stochastic Time-Varying Copula Models', *Computational Statistics & Data Analysis* **56**(6), 1511–1527.

Antoniak, C. E. (1974), 'Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems', *Annals of Statistics* **2**(6), 1152–1174.

Arakelian, V. & Dellaportas, P. (2012), 'Contagion Determination via Copula and Volatility Threshold Models', *Quantitative Finance* **12**(2), 295–310.

Ardia, D. (2006), 'Bayesian Estimation of the GARCH (1,1) Model with Normal Innovations', *Student* **5**(September), 1–13.

Ardia, D. & Hoogerheide, L. F. (2010), Efficient Bayesian estimation and combination of GARCH-type models, *in* K. Bocker, ed., 'Rethinking Risk Measurement and Reporting: Examples and Applications from Finance: Vol II', RiskBooks, London, chapter 1, pp. 1–22.

Asai, M. (2006), 'Comparison of MCMC Methods for Estimating GARCH Models', *Journal of the Japan Statistical Society* **36**(2), 199–212.

Asai, M. & McAleer, M. (2006), 'Asymmetric Multivariate Stochastic Volatility', *Econometric Reviews* **25**(2-3), 453–473.

Asai, M. & McAleer, M. (2009), 'The Structure of Dynamic Correlations in Multivariate Stochastic Volatility Models', *Journal of Econometrics* **150**(2), 182–192.

Asai, M., McAleer, M. & Yu, J. (2006), 'Multivariate Stochastic Volatility: A Review', *Econometric Reviews* **25**(2-3), 145–175.

Ausín, M. C. & Galeano, P. (2007), 'Bayesian Estimation of the Gaussian Mixture GARCH Model', *Computational Statistics & Data Analysis* **51**, 2636–2652.

Ausín, M. C., Galeano, P. & 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.

Ausín, M. C. & Lopes, H. F. (2010), 'Time-Varying Joint Distribution Through Copulas', *Computational Statistics & Data Analysis* **54**, 2383–2399.

Avramov, D. & Zhou, G. (2010), 'Bayesian Portfolio Analysis', *Annual Review of Financial Economics* **2**(1), 25–47.

Bai, X., Russell, J. R. & Tiao, G. C. (2003), 'Kurtosis of GARCH and Stochastic Volatility Models with Non-Normal Innovations', *Journal of Econometrics* **114**(2), 349–360.

Bassetti, F., Casarin, R. & Leisen, F. (2014), 'Beta-product dependent Pitman–Yor processes for Bayesian inference', *Journal of Econometrics* **180**(1), 49–72.

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

Bauwens, L. & Laurent, S. (2005), 'A New Class of Multivariate Skew Densities, with Application to Generalized Autoregressive Conditional Heteroscedasticity Models', *Journal of Business & Economic Statistics* **23**(3), 346–354.

Bauwens, L., Laurent, S. & Rombouts, J. V. K. (2006), 'Multivariate GARCH Models: a Survey', *Journal of Applied Econometrics* **21**(1), 79–109.

Bauwens, L. & Lubrano, M. (1998), 'Bayesian Inference on GARCH Models Using the Gibbs Sampler', *Econometrics Journal* **1**, 23–46.

Bauwens, L. & Lubrano, M. (2002), 'Bayesian Option Pricing Using Asymmetric GARCH Models', *Journal of Empirical Finance* **9**, 321–342.

Bauwens, L., Preminger, A. & Rombouts, J. V. K. (2010), 'Theory and inference for a Markov switching GARCH model', *The Econometrics Journal* **13**(2), 218–244.

Bera, A. K. & Higgins, M. L. (1993), 'ARCH Models: Properties, Estimation and Testing', *Journal of Economic Surveys* **7**(4), 305–362.

Billio, M., Casarin, R. & Osuntuyi, A. (2014), 'Efficient Gibbs sampling for Markov switching GARCH models', *Computational Statistics & Data Analysis* .

Black, F. (1976), 'Studies of Stock Market Volatility Changes', *Proceedings of the American Statistical Association; Business and Economic Statistics Section* pp. 177–181.

Bollerslev, T. (1986), 'Generalized Autoregressive Conditional Heteroskedasticity', *Journal of Econometrics* **31**(3), 307 – 327.

Bollerslev, T. (1987), 'A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return', *The Review of Economics and Statistics* **69**(3), 542–547.

Bollerslev, T. (1990), 'Modelling the Coherence in Short-Run Nominal Exchange Rates: a Multivariate Generalized ARCH Model', *The Review of Economics and Statistics* **72**(3), 498–505.

Bollerslev, T., Chou, R. Y. & Kroner, K. F. (1992), 'ARCH Modeling in Finance: a Review of the Theory and Empirical Evidence', *Journal of Econometrics* **52**(1-2), 5–59.

Bollerslev, T., Engle, R. F. & Nelson, D. B. (1994), ARCH Models, *in* R. F. Engle & D. McFadden, eds, 'Handbook of Econometrics, Vol. 4', number November, Elsevier, Amsterdam, pp. 2959–3038.

Bollerslev, T., Engle, R. F. & Wooldridge, J. M. (1988), 'A Capital Asset Pricing Model with Time-Varying Covariances', *Journal of Political Economy* **96**(1), 116–131.

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

Brown, S. J. (1978), 'The portfolio choice problem: comparison of certainty equivalence and optimal Bayes portfolios', *Communications in Statistics-Simulation and Computation* **4**(7), 321–334.

Burda, M. (2013), 'Constrained Hamiltonian Monte Carlo in BEKK GARCH with Targeting', *Working Paper, University of Toronto* .

Burda, M., Harding, M. & Hausman, J. (2014), 'A Bayesian semiparametric competing risk model with unobserved heterogeneity', *Journal of Applied Econometrics* **Forthcomin**.

Burda, M. & Maheu, J. M. (2013), 'Bayesian Adaptively Updated Hamiltonian Monte Carlo with an Application to High-Dimensional BEKK GARCH Models', *Studies in Nonlinear Dynamics & Econometrics* **17**(4), 345–372.

Cai, J. (1994), 'A Markov Model of Switching-Regime ARCH', *Journal of Business & Economic Statistics* **12**(3), 309–316.

Cappiello, L., Engle, R. F. & Sheppard, K. (2006), 'Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns', *Journal of Financial Econometrics* **4**(4), 537–572.

Carter, C. K. & Kohn, R. (1994), 'On Gibbs sampling for state space models', *Biometrika* **81**, 541–553.

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

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

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

Cecchetti, S. G., Cumby, R. E. & Figlewski, S. (1988), 'Estimation of the Optimal Futures Hedge', *The Review of Economics and Statistics* **70**(4), 623–630.

Chan, D., Kohn, R. & Kirby, C. (2006), 'Multivariate Stochastic Volatility Models with Correlated Errors', *Econometric Reviews* **25**(2-3), 245–274.

Chen, C. W. S., Gerlach, R. & So, M. K. P. (2009), 'Bayesian Model Selection for Heteroskedastic Models', *Advances in Econometrics* **23**, 567–594.

Chib, S. (2001), Markov Chain Monte Carlo Methods: Computation and Inference, *in* J. Heckman & E. Leamer, eds, 'Handbook of Econometrics, Vol. 5', Vol. 5, Elsevier Science B.V., chapter 57, pp. 3570–3649.

Chib, S., Nardari, F. & Shephard, N. (2002), 'Markov Chain Monte Carlo Methods for Stochastic Volatility Models', *Journal of Econometrics* **108**(2), 281–316.

Chib, S., Omori, Y. & Asai, M. (2009), Multivariate Stochastic Volatility, *in* T. Mikosch, J.-P. Kreiß, R. A. Davis & T. G. Andersen, eds, 'Time', Springer Berlin Heidelberg, Berlin, Heidelberg, chapter 16, pp. 365–400.

Christodoulakis, G. a. & Satchell, S. E. (2002), 'Correlated ARCH (CorrARCH): Modelling the Time-varying Conditional Correlation Between Financial Asset Returns', *European Journal of Operational Research* **139**(2), 351–370.

Daníelsson, J. (1998), 'Multivariate Stochastic Volatility Models: Estimation and a Comparison with VGARCH Models', *Journal of Empirical Finance* **5**(2), 155–173.

Delatola, E. & Griffin, J. E. (2010), 'Bayesian Nonparametric Modelling of the Return Distribution with Stochastic Volatility', *Working Paper* pp. 1–24.

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

Delatola, E.-I. & Griffin, J. E. (2013), 'A Bayesian Semiparametric Model for Volatility with a Leverage Effect', *Computational Statistics & Data Analysis* **60**, 97–110.

Dellaportas, P. & Vrontos, I. D. (2007), 'Modelling Volatility Asymmetries: a Bayesian Analysis of a Class of Tree Structured Multivariate GARCH Models', *The Econometrics Journal* **10**(3), 503–520.

Ding, Z. & Engle, R. F. (2001), 'Large Scale Conditional Covariance Matrix Modeling , Estimation and Testing', *Academia Economic Papers* **29**(2), 157–184.

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

Engle, R. F. (2002*a*), 'Dynamic Conditional Correlation', *Journal of Business & Economic Statistics* **20**(3), 339–350.

Engle, R. F. (2002*b*), 'New Frontiers for ARCH Models', *Journal of Applied Econometrics* **17**, 425–446.

Engle, R. F. (2004), 'Risk and Volatility : Econometric Models and Financial Practice', *The American Economic Review* **94**(3), 405–420.

Engle, R. F. & Kroner, K. F. (1995), 'Multivariate Simultaneous Generalized ARCH', *Econometric Theory* **11**(1), 122–150.

Engle, R. F., Ng, V. K. & Rothschild, M. (1990), 'Asset Pricing With a Factor-ARCH Covariance Structure', *Journal of Econometrics* **45**, 213–237.

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

Ferguson, T. S. (1973), 'A Bayesian Analysis of Some Nonparametric Problems', *Annals of Statistics* **1**(2), 209–230.

Ferguson, T. S. (1983), Bayesian Density Estimation by Mixtures of Normal Distributions, *in* H. Rizvi & J. Rustagi, eds, 'Recent Advances in Statistics', Vol. 24, Academic Press, New York, pp. 287–302.

Fiorentini, G., Sentana, E. & Calzolari, G. (2003), 'Maximum likelihood estimation and inference in multivariate conditionally heteroscedastic dynamic regression models with student t innovations', *Journal of Business and Economic Statistics* **21**, 532–546.

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

Galeano, P. & Ausín, M. C. (2010), 'The Gaussian Mixture Dynamic Conditional Correlation Model: Parameter Estimation, Value at Risk Calculation, and Portfolio Selection', *Journal of Business and Economic Statistics* **28**(4), 559–571.

Geweke, J. (1992), 'Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments', *Bayesian Statistics* **4**(3), 169–193.

Geweke, J. (1994), 'Bayesian Comparison of Econometric Models', *Econometrica* **57**, 1317–1340.

Ghysels, E., Harvey, A. & Renault, E. (1996), Stochastic Volatility, *in* G. Maddala & C. Rao, eds, 'Handbook of Statistics 14: Statistical Methods in Finance', North-Holland, Amsterdam, pp. 119–191.

Giamouridis, D. & Vrontos, I. D. (2007), 'Hedge Fund Portfolio Construction: a Comparison of Static and Dynamic Approaches', *Journal of Banking & Finance* **31**, 199–217.

Giannikis, D., Vrontos, I. D. & Dellaportas, P. (2008), 'Modelling Nonlinearities and Heavy Tails via Threshold Normal Mixture GARCH Models', *Computational Statistics & Data Analysis* **52**, 1549–1571.

Gilks, W. R., Best, N. G. & Tan, K. K. C. (1995), 'Adaptive Rejection Metropolis Sampling within Gibbs Sampling', *Journal of the Royal Statistical Society Series C* **44**(4), 455–472.

Glosten, L. R., Jagannathan, R. & Runkle, D. E. (1993), 'On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks', *The Journal of Finance* **48**(5), 1779–1801.

Gray, S. F. (1996), 'Modeling the conditional distribution of interest rates as a regime-switching process', *Journal of Financial Economics* **42**(1), 27–62.

Greyserman, A., Jones, D. H. & Strawderman, W. E. (2006), 'Portfolio selection using hierarchical Bayesian analysis and MCMC methods', *Journal of Banking & Finance* **30**(2), 669–678.

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

Hall, P. & Yao, Q. (2003), 'Inference in ARCH and GARCH Models with Heavy-Tailed Errors', *Econometrica* **71**(1), 285–317.

Hamilton, J. D. & Susmel, R. (1994), 'Autoregressive conditional heteroskedasticity and changes in regime', *Journal of Econometrics* **64**(1-2), 307–333.

Harvey, A. C. & Shephard, N. (1996), 'Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns', *Journal of Business & Economic Statistics* **14**(4), 429–434.

Harvey, A., Ruiz, E. & Shephard, N. (1994), 'Multivariate Stochastic Variance Models', *The Review of Economic Studies* **61**(2), 247–264.

He, C. & Teräsvirta, T. (1999), 'Properties of Moments of a Family of GARCH Processes', *Journal of Econometrics* **92**, 173–192.

He, Z. & Maheu, J. M. (2010), 'Real time detection of structural breaks in GARCH models', *Computational Statistics & Data Analysis* **54**(11), 2628–2640.

Henneke, J. S., Rachev, S. T., Fabozzi, F. J. & Nikolov, M. (2011), 'MCMC-based estimation of Markov Switching ARMA–GARCH models', *Applied Economics* **43**(3), 259–271.

Hirano, K. (2002), 'Semiparametric Bayesian Inference in Autoregressive Panel Data Models', *Econometrica* **70**(2), 781–799.

Hofmann, M. & Czado, C. (2010), 'Assessing the VaR of a portfolio using D-vine copula based multivariate GARCH models', *Working Paper, Technische Universität München Zentrum Mathematik.* .

Hudson, B. G. & Gerlach, R. H. (2008), 'A Bayesian approach to relaxing parameter restrictions in multivariate GARCH models', *Test* **17**, 606–627.

Ishida, I. & Engle, R. F. (2002), 'Modeling Variance of Variance: the Square-Root, the Affine, and the CEV GARCH Models', *Working Paper, New York University* pp. 1–47.

Ishwaran, H. & James, L. F. (2001), 'Gibbs Sampling Methods for Stick-Breaking Priors', *Journal of the American Statistical Association* **96**(453), 161–173.

Jacquier, E. & Polson, N. G. (2013), Asset Allocation in Finance: A Bayesian Perspective, *in* P. Demian, P. Dellaportas, N. G. Polson & D. A. Stephen, eds, 'Bayesian Theory and Applications', 1 edn, Oxford University Press, Oxford, chapter 25, pp. 501–516.

Jacquier, E., Polson, N. G. & Rossi, P. (1995), 'Models and Priors for Multivariate Stochastic Volatility', *CIRANO Working Papers* **95s-18**.

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

Jacquier, E., Polson, N. G. & Rossi, P. E. (2004), 'Bayesian Analysis of Stochastic Volatility Models with Fat-Tails and Correlated Errors', *Journal of Econometrics* **122**(1), 185–212.

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

Jensen, M. J. & Maheu, J. M. (2013), 'Bayesian Semiparametric Multivariate GARCH Modeling', *Journal of Econometrics* **176**(1), 3–17.

Jensen, M. J. & Maheu, J. M. (2014), 'Estimating a semiparametric asymmetric stochastic volatility model with a Dirichlet process mixture', *Journal of Econometrics* **178**, 523–538.

Jin, X. & Maheu, J. M. (2013), 'Modeling Realized Covariances and Returns', *Journal of Financial Econometrics* **11**, 335–369.

Jochmann, M. (2014), 'Modeling U.S. Inflation Dynamics: A Bayesian Nonparametric Approach', *Econometric Reviews* **34**(5), 537–558.

Jorion, P. (1986), 'Bayes-Stein Estimation for Portfolio Analysis', *The Journal of Financial and Quantitative Analysis* **21**(3), 279–292.

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

Kalli, M., Griffin, J. E. & Walker, S. G. (2011), 'Slice sampling mixture models', *Statistics and Computing* **21**(1), 93–105.

Kalli, M., Walker, S. G. & 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.

Kang, L. (2011), 'Asset allocation in a Bayesian Copula-GARCH Framework: An Application to the "Passive Funds Versus Active Funds" Problem', *Journal of Asset Management* **12**(1), 45–66.

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

Kaufmann, S. & Frühwirth-Schnatter, S. (2002), 'Bayesian analysis of switching ARCH models', *Journal of Time Series Analysis* **23**(4), 425–458.

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

Kroner, K. F. & Sultan, J. (1993), 'Time-Varying Distributions and Dynamic Hedging with Foreign Currency Futures', *The Journal of Financial and Quantitative Analysis* **28**(4), 535–551.

Lin, W.-L. (1992), 'Alternative Estimators for Factor GARCH Models - a Monte Carlo Comparison', *Journal of Applied Econometrics* **7**(3), 259–279.

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

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

Lopes, H. F. & Carvalho, C. M. (2013), Online Bayesian Learning in Dynamic Models: an Illustrative Introduction to Particle Methods, *in* P. Damien, P. Dellaportas, N. G. Polson & D. A. Stephens, eds, 'Bayesian Theory and Applications', 1 edn, Oxford University Press, Oxford, chapter 11, pp. 203–228.

Lopes, H. F., Carvalho, C. M., Johannes, M. S. & Polson, N. G. (2011), 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 & M. West, eds, 'Bayesian Statistics 9', Vol. 9, Oxford University Press, Oxford, pp. 317–360.

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

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

MacEachern, S. N. (1994), 'Estimating Normal Means with a Conjugate Style Dirichlet Process Prior', *Communications in Statistics - Simulation and Computation* **23**(3), 727–741.

Markowitz, H. M. (1952), 'Portfolio selection', *Journal of Finance* **7**, 77–91.

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

Merton, R. C. (1972), 'An Analytic Derivation of the Efficient Portfolio Frontier', *The Journal of Financial and Quantitative Analysis* **7**(4), 1851–1872.

Meyer, R. & Yu, J. (2000), 'BUGS for a Bayesian Analysis of Stochastic Volatility Models', *Econometrics Journal* **3**(2), 198–215.

Miazhynskaia, T. & Dorffner, G. (2006), 'A Comparison of Bayesian Model Selection Based on MCMC with Application to GARCH-type Models', *Statistical Papers* **47**(4), 525–549.

Müller, P. & Pole, A. (1998), 'Monte Carlo Posterior Integration in GARCH Models', *Sankhya, Series B* **60**(1), 127–14.

Nakatsuma, T. (2000), 'Bayesian Analysis of ARMA- GARCH Models: a Markov Chain Sampling Approach', *Journal of Econometrics* **95**, 57–69.

Neal, R. M. (2000), 'Markov Chain Sampling Methods for Dirichlet Process Mixture Models', *Journal of Computational and Graphical Statistics* **9**(2), 249–265.

Nelson, D. B. (1991), 'Conditional Heteroskedasticity in Asset Returns: a New Approach', *Econometrica* **59**(2), 347–370.

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

Osiewalski, J. & Pipien, M. (2004), 'Bayesian Comparison of Bivariate ARCH - Type Models for the Main Exchange Rates in Poland', *Journal of Econometrics* **123**(2), 371–391.

Papaspiliopoulos, O. (2008), 'A Note on Posterior Sampling from Dirichlet Mixture Models', *Working Paper, University of Warwick. Centre for Research in Statistical Methodology, Coventry* pp. 1–8.

Papaspiliopoulos, O. & Roberts, G. O. (2008), 'Retrospective Markov Chain Monte Carlo Methods for Dirichlet Process Hierarchical Models', *Biometrika* **95**(1), 169–186.

Patton, A. J. (2006), 'Modelling Asymmetric Exchange Rate Dependence', *International Economic Review* **47**(2), 527–556.

Patton, A. J. (2009), Copula - Based Models for Financial Time Series, *in* T. G. Andersen, T. Mikosch, J.-P. Kreiß& R. A. Davis, eds, 'Handbook of Financial Time Series', Springer Berlin Heidelberg, Berlin, Heidelberg, chapter 36, pp. 767–785.

Quintana, J. M. & West, M. (1987), 'An Analysis of International Exchange Rates Using Multivariate DLM's', *The Statistician* **36**, 275.

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

Robert, C. P. & Casella, G. (2004), *Monte Carlo Statistical Methods*, 2nd edn, Springer Texts in Statistics.

Rodriguez, A. & ter Horst, E. (2008), 'Bayesian dynamic density estimation', *Bayesian Analysis* **3**(2), 339–365.

Rossi, E. & Spazzini, F. (2010), 'Model and Distribution Uncertainty in Multivariate GARCH Estimation: A Monte Carlo Analysis', *Computational Statistics & Data Analysis* **54**(11), 2786–2800.

Rossi, E. & Zucca, C. (2002), 'Hedging Interest Rate Risk with Multivariate GARCH', *Applied Financial Economics* **12**(4), 241–251.

Sethuraman, J. (1994), 'A Constructive Definition of Dirichlet Priors', *Statistica Sinica* **4**, 639–650.

Shephard, N. (1993), 'Fitting Nonlinear Time-Series Models with Applications to Stochastic Variance Models', *Journal of Applied Econometrics* **8**, S135–S152.

Shephard, N. (1996), Statistical Aspects of ARCH and Stochastic Volatility, *in* D. R. Cox, D. V. Hinkley & O. E. Barndorff-Nielsen, eds, 'Time Series Models', number 1, Springer US, pp. 1–67.

Shephard, N. & Andersen, T. G. (2009), Stochastic Volatility: Origins and Overview, *in* T. Mikosch, J.-P. Kreiß, R. A. Davis & T. G. Andersen, eds, 'Handbook of Financial Time Series', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 233–254.

Silvennoinen, A. & Teräsvirta, T. (2009), Multivariate GARCH Models, *in* T. G. Andersen, R. A. Davis, J.-P. Kreiß& T. Mikosch, eds, 'Handbook of Financial Time Series', Springer, Berlin, Heidelberg, pp. 201–229.

Sklar, A. (1959), 'Fonctions de répartition á n Dimensions et Leur Marges', *Publ. Inst. Statist. Univ. Paris* **8**, 229–231.

Smith, M. & Pitts, A. (2006), 'Foreign Exchange Intervention by the Bank of Japan: Bayesian Analysis Using a Bivariate Stochastic Volatility Model', *Econometric Reviews* **25**(2-3), 425–451.

So, M. E. P., Lam, K. & Li, W. K. (1998), 'A stochastic volatility model with markov switching', *Journal of Business & Economic Statistics* **16**(2), 244–253.

So, M. K. P., Li, W. K. & Lam, K. (1997), 'Multivariate Modelling of the Autoregressive Random Variance Process', *Journal of Time Series Analysis* **18**(4), 429–446.

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.

Taddy, M. A. (2010), 'Autoregressive Mixture Models for Dynamic Spatial Poisson Processes: Application to Tracking Intensity of Violent Crime', *Journal of the American Statistical Association* **105**(492), 1403–1417.

Taddy, M. A. & Kottas, A. (2009), 'Markov Switching Dirichlet Process Mixture Regression', *Bayesian Analysis* **4**(4), 793–816.

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

Taylor, S. J. (1986), *Modelling Financial Time Series*, 1 edn, John Willey & Sons, New York.

Taylor, S. J. (1994), 'Modeling Stochastic Volatility: a Review and Comparative Study', *Mathematical Finance* **4**(2), 183–204.

Teräsvirta, T. (2009), An Introduction to Univariate GARCH Models, *in* T. G. Andersen, T. Mikosch, J.-P. Kreiß& R. A. Davis, eds, 'Handbook of Financial Time Series', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 17–42.

Tierney, L. (1994), 'Markov Chains for Exploring Posterior Distributions', *Annals of Statistics* **22**(4), 1701–1728.

Tokdar, S. T. (2006), 'Posterior Consistency of Dirichlet Location-scale Mixture of Normals in Density Estimation and Regression', *Sankhyā: Indian Statistical Institute* **67**(4), 90–110.

Tsay, R. S. (2010), *Analysis of Financial Time Series*, 3 edn, John Willey & Sons, Hoboken.

Tse, Y. K. & Tsui, A. K. C. (2002), 'A Multivariate Generalized Autoregressive Conditional Heteroscedasticity Model With Time-Varying Correlations', *Journal of Business & Economic Statistics* **20**(3), 351–362.

Van der Weide, R. (2002), 'GO-GARCH: a Multivariate Generalized Orthogonal GARCH Model', *Journal of Applied Econometrics* **17**, 549–564.

Virbickaitė, A., Ausín, M. C. & Galeano, P. (2013), 'Bayesian Inference Methods for Univariate and Multivariate Garch Models: a Survey', *Journal of Economic Surveys* **Forthcoming**.

Virbickaitė, A., Ausín, M. C. & Galeano, P. (2014), 'A Bayesian Non-Parametric Approach to Asymmetric Dynamic Conditional Correlation Model with Application to Portfolio Selection', *Computational Statistics & Data Analysis* **Forthcoming**.

Virbickaitė, A., Lopes, H. F., Ausín, M. C. & Galeano, P. (2014), 'Particle Leaning for Bayesian Non-Parametric Markov Switching Stochastic Volatility Model', *UC3M, Working Paper* **14-28**(Statistics and Econometrics Series (19)).

Vrontos, I. D., Dellaportas, P. & Politis, D. N. (2000), 'Full Bayesian Inference for GARCH and EGARCH Models', *Journal of Business and Economic Statistics* **18**(2), 187–198.

Vrontos, I. D., Dellaportas, P. & Politis, D. N. (2003*a*), 'A Full-Factor Multivariate GARCH Model', *Econometrics Journal* **6**, 312–334.

Vrontos, I. D., Dellaportas, P. & Politis, D. N. (2003*b*), 'Inference for Some Multivariate ARCH and GARCH Models', *Journal of Forecasting* **22**(6-7), 427–446.

Wago, H. (2004), 'Bayesian Estimation of Smooth Transition GARCH Model Using Gibbs Sampling', *Mathematics and Computers in Simulation* **64**, 63–78.

Walker, S. G. (2007), 'Sampling the Dirichlet Mixture Model with Slices', *Communications in Statistics - Simulation and Computation* **36**(1), 45–54.

Yang, W. & Allen, D. E. (2004), 'Multivariate GARCH Hedge Ratios and Hedging Effectiveness in Australian Futures Markets', *Accounting and Finance* **45**, 301–321.

Yilmaz, T. (2011), 'Improving Portfolio Optimization by DCC and DECO GARCH: Evidence from Istanbul Stock Exchange', *International Journal of Advanced Economics and Business Management* **1**(2), 81–92.

Yu, B. & Mykland, P. (1998), 'Looking at Markov Samplers Through Cusum Path Plots: a Simple Diagnostic Idea', *Statistics and Computing* **8**, 275–286.

Yu, J. (2005), 'On leverage in a Stochastic Volatility Model', *Journal of Econometrics* **127**(2), 165–178.

Yu, J. & Meyer, R. (2006), 'Multivariate Stochastic Volatility Models: Bayesian Estimation and Model Comparison', *Econometric Reviews* **25**(2-3), 361–384.

Zakoian, J.-M. (1994), 'Threshold Heteroskedastic Models', *Journal of Economic Dynamics and Control* **18**(5), 931–955.