https://doi.org/10.26219/heal.aueb.5968

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

STOCHASTIC VOLATILITY MODELS
A BAYESIAN APPROACH

By

Stefanos G. Giakoumatos

A THESIS
Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfillment of the requirements for
the degree of Master of Science in Statistics

Athens, Greece
1997

ISBN : 960-7929-03-9

https://doi.org/10.26219/heal.aueb.5968

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

BAYESIAN STOCHASTIC VOLATILITY MODELS

By

Stefanos G. Giakoumatos

A THESIS
Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfillment of the requirements for
the degree of Master of Science in Statistics

Athens, Greece
May 1997

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΜΠΕΫΖΙΑΝΑ ΜΟΝΤΕΛΑ ΣΤΟΧΑΣΤΙΚΗΣ ΔΙΑΚΥΜΑΝΣΗΣ

Στέφανος Γ. Γιακουμάτος

ΔΙΑΤΡΙΒΗ

Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος Ειδίκευσης στη Στατιστική

Αθήνα
Μάης 1997

https://doi.org/10.26219/heal.aueb.5968

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

A Thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Science

BAYESIAN STOCHASTIC VOLATILITY MODELS

Stefanos G. Giakoumatos

Supervisor :
Dr. P. Dellaportas

External examiner :
Professor Peter Mueller
Institute of Statistics and Decision Sciences,
Duke University, USA

Approved by the Graduate Committee

Professor E. Xekalaki
Director of the Graduate Program
September 1997

DEDICATION

To ever memorable **Ioannis Koklas**,

my grandfather.

https://doi.org/10.26219/heal.aueb.5968

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Petros Dellaportas for his endless support and care; he taught me so much in so little time. Moreover, I would like to express my thanks to Dr. Kostas Christopoulos because he provided many data from the Athens Exchange Stock Market.

My mother and my sister know that they have a special part of my heart and I will never stop loving and thanking them for their love and endless support.

https://doi.org/10.26219/heal.aueb.5968

VITA

I was born in Athens in 22 of July in 1972. I graduated from the 3[d] Highschool of Koridalos in 1990, and at the same year I succeeded in the Athens University of Economics. In 1994, I took my degree in Statistics (9.2/10, excellent).

I should also mention that during all my academic years I was proficient enough so as to gain scholarships from the State Scholarships Foundation and also from the George Xalkiopoulou Foundation. I have also attended courses of the MSc in Statistics at the same University.

https://doi.org/10.26219/heal.aueb.5968

ABSTRACT

Stefanos G. Giakoumatos

BAYESIAN STOCHASTIC VOLATILITY MODELS

February 1997

The phenomenon of changing variance and covariance is often encountered in financial time series. As a result, the last ten years researchers move their interests from the homoscedastic time series models to conditional heteroscedastic time series models.

In general, the models of changing variance and covariance are called Volatility models. The main representatives of this class of models are the **Autoregressive Conditional Heteroscedasticity** (ARCH) **models** (Engle, 1982), the **Generalized Autoregressive Conditional Heteroscedasticity** (GARCH) **models** (Bollerslev, 1986), and the **Stochastic Volatility models** (SV) (Taylor, 1986).

The SV model is a very promising alternative of the ARCH and GARCH models and has been the focus of considerable attention in the recent years. From the classical view of the Statistics the SV models have been investigated by Taylor (1986) and Vetzal (1992). On the other hand, from the Bayesian approach little work has been done. A recent paper of Jacquier et al (1994) was the first step to this direction.

In this dissertation we focus our interest on the Stochastic Volatility models using the Bayesian framework. We develop an MCMC algorithm that converges to the joint posterior distribution of the parameters of the SV model. The MCMC algorithm that has been proposed by Jacquier et al (1994) has been studied and we have evidence that it is not very efficient.

https://doi.org/10.26219/heal.aueb.5968

In our proposed MCMC algorithm, we use some techniques so that the algorithm achieves better convergence characteristics. Our experience show that the random-scan MCMC algorithm converges faster than the general MCMC algorithm. In addition to that a reparameterisation is used which gives better performance than the random-scan MCMC algorithm.

Finally, we illustrate our methodology by modeling the weekly rate of return of the General Index of the Athens Stock Exchange Market.

https://doi.org/10.26219/heal.aueb.5968

ΠΕΡΙΛΗΨΗ

Στέφανος Γ. Γιακουμάτος

ΜΠΕΫΖΙΑΝΑ ΜΟΝΤΕΛΑ ΣΤΟΧΑΣΤΙΚΗΣ ΔΙΑΚΥΜΑΝΣΗΣ

Φεβρουάριος 1997

Στις χρονολογικές σειρές των οικονομικών δεδομένων - και κυρίως σε χρηματιστηριακές σειρές - έχει παρατηρηθεί ότι η διακύμανση και η συνδιακύμανση δεν είναι σταθερές αλλά μεταβάλλονται κατά την διάρκεια του χρόνου. Σαν αποτέλεσμα αυτού του γεγονότος, οι ερευνητές έχουν μετακινήσει τα ενδιαφέροντα τους τα τελευταία χρόνια από τα ομοσκεδαστικά μοντέλα χρονοσειρών στα ετεροσκεδαστικά μοντέλα.

Τα μοντέλα αυτά, που επιτρέπουν στην διακύμανση της σειράς να μεταβάλλεται, ονομάζονται **Volatility models**. Οι κύριοι αντιπρόσωποι αυτής της κατηγορίας μοντέλων είναι τα **Αυτοπαλίνδρομα Εξαρτώμενης Ετεροσκεδαστικότητας (ΑΕΕ) μοντέλα**[1] (Engle, 1982), τα **Γενικευμένα Αυτοπαλίνδρομα Εξαρτώμενης Ετεροσκεδαστικότητας (ΓΑΕΕ) μοντέλα**[2] (Bollerslev, 1986) και τα **Στοχαστικής Διακύμανσης (ΣΔ) μοντέλα**[3] (Taylor, 1986).

---

[1] Autoregressive Conditional Heteroscedasticity (ARCH) models
[2] Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models
[3] Stochastic Volatility (SV) models

Τα ΣΔ μοντέλα είναι μία πολλά υποσχόμενη εναλλακτική πρόταση ως προς τα ΑΕΕ και ΓΑΕΕ μοντέλα και έχουν έλξει το ενδιαφέρον των ερευνητών τα τελευταία χρόνια. Η εκτίμηση των παραμέτρων του ΣΔ μοντέλου μέσω της κλασσικής Στατιστικής έχει ερευνηθεί από τους Taylor (1986) και Vetzal (1992). Από την άποψη της Bayesian Στατιστικής λίγη έρευνα έχει γίνει. Στην διεθνής βιβλιογραφία ένα μόνο paper από τους Jacquier et al (1994) υπάρχει και αυτό ήταν το πρώτο βήμα προς αυτή την κατεύθυνση.

Σε αυτή την διατριβή επικεντρώνουμε το ενδιαφέρον μας στα ΣΔ μοντέλα χρησιμοποιώντας την Bayesian ανάλυση για να κάνουμε συμπερασματολογία σχετικά με τις παραμέτρους του μοντέλου. Προτείνουμε ένα MCMC[4] αλγόριθμο που συγκλίνει στην εκ των υστέρων κατανομή των παραμέτρων του ΣΔ μοντέλου. Θεωρούμε πως αυτός ο αλγόριθμος είναι πιο αποτελεσματικός από τον αντίστοιχο αλγόριθμο των Jacquier et al (1994). Επιπλέον χρησιμοποιούμε διάφορες τεχνικές έτσι ώστε να επιταχυνθεί η σύγκλιση.

Τα πειράματά μας δείχνουν πως ο random-scan MCMC αλγόριθμος συγκλίνει γρηγορότερα από τον απλό MCMC αλγόριθμο. Επιπλέον η αλλαγή παραμέτρων του ΣΔ μοντέλου δίνει ακόμα καλύτερα αποτελέσματα.

Τέλος, εφαρμόσαμε το ΣΔ μοντέλο στον γενικό δείκτη του Χρηματιστηρίου Αξιών Αθηνών.

[4] Markov Chain Monte Carlo (MCMC)

https://doi.org/10.26219/heal.aueb.5968

LIST OF TABLES

https://doi.org/10.26219/heal.aueb.5968

LIST OF FIGURES

https://doi.org/10.26219/heal.aueb.5968

Contents

https://doi.org/10.26219/heal.aueb.5968

https://doi.org/10.26219/heal.aueb.5968

Chapter 1

Introduction

Time varying volatility (variance) is a main characteristic of many financial time series. Engle (1982) modelled this phenomenon introducing a new class of models which called Autoregressive Conditional Heteroscedasticity models (ARCH) and Bollerslev (1986) extended this class of models with Generalized Autoregressive Conditional Heteroscedasticity models (GARCH).

The basic alternative to these models is the Stochastic Volatility model (SV) which allows both the conditional mean and variance to be driven by separate stochastic processes. The stochastic conditional variance of the SV is the main advantage of this model with respect to ARCH & GARCH class of models.

The general form of the Stochastic volatility model is

$$y_t = \varepsilon_t \exp(h_t/2)$$

$$h_t = a + dh_{t-1} + \eta_t$$

$$\eta_t \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_\eta^2\right)$$

$$\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, 1)$$

where one interpretation of the latent variable h_t is that it represents the random and uneven flow of a new information, which is very difficult to model directly into financial markets.

The main disadvantage of the SV model is that, unlike the ARCH and GARCH models, has a very large number of parameters which are very difficult to be estimated.

Many methods have been proposed in order to solve the problem of estimating the parameters of the SV model. Some of these use the Maximum Likelihood theory (Taylor, 1986), some the Generalized Methods of Moments (Chesney and Scott, 1989), and some others the Quasi-Maximum likelihood (Harvey, Ruiz and Shephard, 1994). On the other hand, there is little work which adopts the Bayesian framework for the estimation of the parameters of the SV model.

In this dissertation we focus our analysis in the Bayesian approach of the SV model. It must be noted that it is more logical to view the SV model with a hierarchical structure, therefore the Bayesian framework is the most natural choice.

In the literature, only one paper has appeared (Jacquier et al, 1994) which uses this approach. In this paper an MCMC algorithm is introduced which converges to the distribution of the parameter of the SV model. We study this algorithm and we provide some evidence that its performance is not very efficient. Moreover following the steps of Jacquier et al, we introduce a somewhat similar MCMC algorithm that works more efficient. In this algorithm we use the well known MCMC techniques such that the Gibbs sampler and the Metropolis-Hasting algorithm.

This algorithm is more efficient than Jacquier et al algorithm but the burn-in sample that is needed to converge to the equilibrium distribution is somehow big. This problem appears due to high autocorrelation of the Markov chain and the high correlation of the parameters of the SV model.

To handle this problem we use some techniques so that the algorithm converges to the desired posterior distribution more quickly. Firstly, we use a non-sequential MCMC algorithm but the outcomes are the same as the sequential MCMC algorithm. Secondly, we use the random scan MCMC algorithm in which we update the latent parameters (log-volatilities) not in the usual sequential order, but randomly. This technique reduces the necessary burn-in sample. Finally, we try to reparameterise the latent parameters **h** so that each of them is independent from the others. Our experiments provided evidence that this transformation decrease the time to convergence of the Markov chain.

The final conclusions are, that the reparameterisation of the latent parameters **h** (log-volatilities) seems better than random scan MCMC and the latter seems better than sequential MCMC.

All these algorithms are applied to the weekly rate of return of the General Index of Athens Stock Exchange Market. It must be noted that such analysis is used for first time for the Greek General Index.

https://doi.org/10.26219/heal.aueb.5968

Chapter 2

The Basic aspects of Bayesian Statistics

2.1 Introduction

The Bayesian Statistics is a dynamic branch of the Statistics science which is based on the Bayes's Theorem (Bayes, 1763). In the next sections we will give the main points of this theory. For more information about the Bayesian Statistics see Bernardo and Smith (1995), Magdalinos (1992), Robert (1994).

2.2 Bayes's Theorem

Theorem 1 *Bayes (discrete r.v.)*

For a given probability space $(\Omega, \mathcal{A}, P[\cdot])$, if $B_1, B_2, ..., B_n$ is a collection of mutually disjoint events in $\mathcal{A}$ satisfying $\Omega = \bigcup_{j=1}^{n} B_j$ and $P[B_j] > 0$ for

$j = 1, ..., n$, then for every $A \in \mathcal{A}$ for which $P[A] > 0$

$$P[B_k|A] = \frac{P[A|B_k]P[B_k]}{\sum_{j=1}^{n} P[A|B_j]P[B_j]}$$

Theorem 2 *Bayes (continuous r.v.)*

Let x and y continuous random variables, then

$$f(x|y) = \frac{f(y|x)f(x)}{\int_x f(y|x)f(x)dx}$$

2.3 Interpretation

Let **D** the set of the data which we have at hand. The data comes from a distribution $f(\cdot|\boldsymbol{\theta})$ where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is the parameter vector. Now we want to investigate the distribution of $\boldsymbol{\theta}|\mathbf{D}$. Let $\pi(\boldsymbol{\theta})$ the **a priori** information or subjective opinion that we have about $\boldsymbol{\theta}$. According to the Bayes theorem we have

$$f(\boldsymbol{\theta}|\mathbf{D}) = \frac{f(\mathbf{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int \cdots \int f(\mathbf{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}$$

or

https://doi.org/10.26219/heal.aueb.5968

$$f(\boldsymbol{\theta}|\mathbf{D}) \quad \propto \quad f(\mathbf{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$$

where $\propto$ means equal up to a constant.

From a different side of view, $f(\mathbf{D}|\boldsymbol{\theta})$ can be thought as the likelihood of the data so $L(\boldsymbol{\theta}|\mathbf{D}) = f(\mathbf{D}|\boldsymbol{\theta})$

$$f(\boldsymbol{\theta}|\mathbf{D}) \quad \propto \quad L(\boldsymbol{\theta}|\mathbf{D})\pi(\boldsymbol{\theta}).$$

The final result $f(\boldsymbol{\theta}|\mathbf{D})$ is the a posteriori probability or distribution of $\boldsymbol{\theta}$ conditioned on $\mathbf{D}$.

This magnitude is the challenging part of all the analysis and based on this, we can take the right decision.

2.4 Marginal & Conditional Distributions

Let $\boldsymbol{\theta} = (\theta_1, \theta_2) \in \mathbf{R}^2$. The marginal distribution of θ_1 is

$$f(\theta_1|\mathbf{D}) \quad = \quad \int f(\boldsymbol{\theta}|\mathbf{D})d\theta_2$$

and the conditional distribution of θ_1 is $f(\theta_1|\theta_2, \mathbf{D})$.

2.5 Predictions

Many times, we want to know the behavior of an observation x_0 which has not turned up. Let $f(\boldsymbol{\theta}|\mathrm{D})$ to be the a posteriori distribution of $\boldsymbol{\theta}$, then the posterior distribution of $\boldsymbol{\theta}$ and x_0 given the data, is the following

$$f(x_0, \boldsymbol{\theta}|\mathrm{D}) = f(x_0|\boldsymbol{\theta}, \mathrm{D}) f(\boldsymbol{\theta}|\mathrm{D})$$

and the predictive distribution is

$$f(x_0|\mathrm{D}) = \int \cdots \int f(x_0, \boldsymbol{\theta}|\mathrm{D}) d\boldsymbol{\theta}.$$

2.6 Informative and non-informative priors

Many times we have no prior information about the parameters of interest. In such cases, it is common to use as a priori distributions for these parameters, the distributions which are called non informative. A commonly used non-informative prior based on Jeffreys prior for location and scale is: if $\theta \in \mathbf{R}$ then $\pi(\theta) \propto 1$ and if $\theta \in \mathbf{R}^+$ then $\pi(\theta) \propto \frac{1}{\theta}$.

On the other hand, if we have some information from previous surveys about the parameters or we want to include our subjective opinion in the

https://doi.org/10.26219/heal.aueb.5968

survey scheme, then we define the prior distributions of the parameters of interest according to this (subjective) information. These distributions are called informative distributions, e.g. $\theta \sim \mathbf{N}(\mu, \sigma^2)$.

2.7 Example

Let $\mathbf{X} = (X_1, X_2, ..., X_n)^T$ comes from the Normal distribution $\mathbf{N}(\mu, \sigma^2)$ where the two parameters are unknown.

The likelihood of the data is

$$f(\mathbf{x}|\mu, \sigma^2) \quad \propto \quad (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{n}(x_i - \mu)^2\right\}.$$

If the a priori distributions of these parameters are considered as non informative distributions, then the full prior distribution takes the form

$$\pi(\mu, \sigma^2) \quad \propto \quad 1/\sigma^2.$$

Finally the a posteriori distribution is

$$f(\mu, \sigma^2|\mathbf{x}) \quad \propto \quad (\sigma^2)^{-\frac{n}{2}+1} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum(x_i - \bar{x})^2 + n(\mu - \bar{x})^2\right)\right\}.$$

The marginal distributions are

$$
\begin{aligned}
f(\mu|\mathbf{x}) &\propto \left(\sum(x_i-\bar{x})^2+n(\mu-\bar{x})^2\right)^{-n/2} \\
&\equiv t\left(n-1,\bar{x},\frac{\sum(x_i-\bar{x})^2}{n(n-1)}\right).
\end{aligned}
$$

$$
\begin{aligned}
f(\sigma^2|\mathbf{x}) &\propto (\sigma^2)^{-n/2}\exp\left(-\frac{\sum(x_i-\bar{x})^2}{2\sigma^2}\right) \\
&\equiv \mathbf{IG}\left(\frac{n}{2},\frac{\sum(x_i-\bar{x})^2}{2}\right).
\end{aligned}
$$

where **IG** denotes the inverse Gamma distribution. The conditional distributions are

$$
\begin{aligned}
f(\mu|\sigma^2,\mathbf{x}) &\propto \exp\left\{-\frac{n}{2\sigma^2}(\mu-\bar{x})\right\} \\
&\equiv \mathbf{N}\left(\bar{x},\frac{\sigma^2}{n}\right)
\end{aligned}
$$

$$
\begin{aligned}
f(\sigma^2|\mu,\mathbf{x}) &\propto (\sigma^2)^{-n/2}\exp\left(-\frac{\sum(x_i-\bar{x})^2}{2\sigma^2}\right) \\
&\equiv \mathbf{IG}\left(\frac{n}{2},\frac{\sum(x_i-\bar{x})^2}{2}\right).
\end{aligned}
$$

https://doi.org/10.26219/heal.aueb.5968

Chapter 3

Markov Chain Monte Carlo (MCMC)

3.1 Introduction

From the form of Bayes's theorem it is obvious that technical difficulties arise in the calculation of the posterior and marginal posterior densities which are needed for the Bayesian inference. See, for example, the integrals in sections 2.2, 2.3, 2.4, 2.5. These integrals, many times, are intractable so analytic calculations are not possible. In order to solve this problem -main problem for the Bayesian inference- it has been proposed the use of numerical integration or analytic approximation techniques. The most popular technique is **Markov Chain Monte Carlo** .

https://doi.org/10.26219/heal.aueb.5968

3.2 Monte Carlo Methods

The key of this idea is very simple. Suppose that we want an approximation of the integral

$$\mathbf{I} = \int_{\Theta} g(\theta) f(\theta) d\theta$$

where $f(\theta)$ is proportional to a distribution from which we can easily take a sample. A straightforward way to do this is to generate random variables $\theta^1, \theta^2, ..., \theta^n$ from $f(\theta)$ and then according to the law of large numbers

$$\lim_{n \to +\infty} \left(\frac{1}{n} \sum_{i=1}^{n} g(\theta^i) f(\theta^i) \right) = \mathbf{I}.$$

See, for example, Hammersley and Handscomb (1964).

3.3 Markov Chain Monte Carlo

The Markov chain Monte Carlo method is a more general Monte Carlo method which approximates the generation of random variables from a posterior density, when this density cannot be directly simulated (Gilks et al, 1996 & Tierney, 1994) .

The idea is the following. Suppose that we want to sample from a posterior distribution $f(\boldsymbol{\theta}|\mathrm{D})$ where $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbf{R}^k$ denotes the unknown parameters and D denotes the data, but we cannot do this directly. On the other hand, we can construct a Markov chain with state space $\boldsymbol{\Theta}$, which is straightforward to sample from and whose equilibrium distribution is $f(\boldsymbol{\theta}|\mathrm{D})$. If then we run this chain for a long time, we would take a sample from its equilibrium distribution therefore we would take a sample from $f(\boldsymbol{\theta}|\mathrm{D})$.

For example if we want to integrate

$$\int_{\Theta} g(\boldsymbol{\theta}) f(\boldsymbol{\theta}|\mathrm{D}) d\boldsymbol{\theta}$$

where the $f(\boldsymbol{\theta}|\mathrm{D})$ is not one of the known form distributions and we cannot simulate from it. So we construct a Markov chain with equilibrium distribution the $f(\boldsymbol{\theta}|\mathrm{D})$. If now the $\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, ..., \boldsymbol{\theta}^n, ...$ is a realization from this chain, then as $n \rightarrow +\infty$

$$\boldsymbol{\theta}^n \rightarrow \boldsymbol{\theta} \sim f(\boldsymbol{\theta}|\mathrm{D}),$$

in distribution and

$$\frac{1}{n} \sum_{i=1}^{n} g(\boldsymbol{\theta}^i) f(\boldsymbol{\theta}^i|\mathrm{D})$$

is the approximate solution of the above integral.

Of course, successively $\boldsymbol{\theta}^n$ will be highly correlated, so that, if the first of these asymptotic results are to be exploited to mimic a random sample from $f(\boldsymbol{\theta}|\mathrm{D})$, suitable spacings will be required between realizations, or parallel independent runs of the chain might be considered.

In the following sections we will describe the most popular forms of Markov chain scheme, the *Gibbs sampler* and the *Metropolis sampler*.

3.4 The Gibbs sampler

Suppose that we want to sample from an a posteriori density $f(\boldsymbol{\theta}|\mathbf{D})$ which is, probably, known up to a constant, and let $\boldsymbol{\theta}$ a vector of dimension k. Also, let

$$\begin{aligned} &f(\theta_1|\boldsymbol{\theta}_{-1}, \mathbf{D}) \\ &f(\theta_2|\boldsymbol{\theta}_{-2}, \mathbf{D}) \\ &\vdots \\ &f(\theta_i|\boldsymbol{\theta}_{-i}, \mathbf{D}) \\ &\vdots \\ &f(\theta_k|\boldsymbol{\theta}_{-k}, \mathbf{D}) \end{aligned}$$

be the so called **full conditional distributions** for the individual components given the data and specified values of all the other components of $\boldsymbol{\theta}$. In the above, $\boldsymbol{\theta}_{-i}$ means a vector, that includes all the components, apart from the θ_i, i.e. $\boldsymbol{\theta}_{-i} = (\theta_1, ..., \theta_{i-1}, \theta_{i+1}, ..., \theta_k)$.

Suppose then, that given an arbitrary set of starting values, $\theta_1^0, ..., \theta_k^0$, for the unknown quantities, we implement the following iterative algorithm

First iteration

$$\begin{aligned} &\text{draw } \theta_1^1 \text{ from } f(\theta_1|\mathbf{D}, \theta_2^0, \theta_3^0, ..., \theta_k^0) \\ &\text{draw } \theta_2^1 \text{ from } f(\theta_2|\mathbf{D}, \theta_1^1, \theta_3^0, ..., \theta_k^0) \\ &\vdots \\ &\text{draw } \theta_k^1 \text{ from } f(\theta_k|\mathbf{D}, \theta_1^1, \theta_2^1, ..., \theta_{k-1}^1) \end{aligned}$$

Second iteration

$$\text{draw } \theta_1^2 \text{ from } f(\theta_1|\mathbf{D}, \theta_2^1, \theta_3^1, ..., \theta_k^1)$$
$$\text{draw } \theta_2^2 \text{ from } f(\theta_2|\mathbf{D}, \theta_1^2, \theta_3^1, ..., \theta_k^1)$$
$$\vdots$$
$$\text{draw } \theta_k^2 \text{ from } f(\theta_k|\mathbf{D}, \theta_1^2, \theta_2^2, ..., \theta_{k-1}^2)$$

i-th iteration

$$\text{draw } \theta_1^i \text{ from } f(\theta_1|\mathbf{D}, \theta_2^{i-1}, \theta_3^{i-1}, ..., \theta_k^{i-1})$$
$$\text{draw } \theta_2^i \text{ from } f(\theta_2|\mathbf{D}, \theta_1^i, \theta_3^{i-1}, ..., \theta_k^{i-1})$$
$$\vdots$$
$$\text{draw } \theta_k^i \text{ from } f(\theta_k|\mathbf{D}, \theta_1^i, \theta_2^i, ..., \theta_{k-1}^i).$$

Having repeated the above iterative algorithm t times, the transition probability of going from $\boldsymbol{\theta}^t = \left(\theta_1^t, ..., \theta_k^t\right)$ to $\boldsymbol{\theta}^{t+1} = \left(\theta_1^{t+1}, ..., \theta_k^{t+1}\right)$ is

$$\pi(\boldsymbol{\theta}^t, \boldsymbol{\theta}^{t+1}) = \prod_{i=1}^{k} f(\theta_i^{t+1}|\theta_j^t, j > i, \theta_j^{t+1}, j < i, D).$$

Then as $t \to +\infty$, $\boldsymbol{\theta}^t = \left(\theta_1^t, ..., \theta_k^t\right) \sim f(\boldsymbol{\theta}|\mathbf{D})$. Moreover, θ_i^t tends in distribution to a random quantity that is drawn from the marginal distribution $f(\theta_i|\mathbf{D})$.

This scheme is very pleasing to the statisticians because of the easiness of its implementation and many examples are available in the literature. For more information about the Gibbs Sampler and its applications see Casela and George (1992), Dellaportas (1993), Gelfand and Smith (1990), Gelfand

https://doi.org/10.26219/heal.aueb.5968

et al (1990), Geman and Geman (1984), Roberts and Sahu (1996), Tai-Ming Lee (1992).

3.5 The Metropolis-Hastings sampler

3.5.1 General points

As we have, already said, in Monte Carlo Markov Chains we want to construct a Markov chain in which its equilibrium distribution is the posterior distribution. Assume that a candidate-generating density in order to construct the Markov chain is the $\mathbf{q(x,y)}$ where $\int q(x,y)dy = 1$ (Metropolis et al, 1953 & Hastings, 1970). This density can be interpreted as saying that when a process is at the point x, the density generates values y from $q(x,y)$. Of course this candidate must follow some conditions, such as the reversibility condition for all (x,y) . Many times -we can say all the times-this condition is not satisfied. So for some pairs (x,y) we might find that

$$\pi(x)q(x,y) > \pi(y)q(y,x).$$

In these cases, the process moves from x to y too often and from y to x too rarely. Thus in order to repair this mistake we introduce a probability $a(x,y) < 1$, which is called probability of move, that corrects the condition of reversibility and reduces the number of moves from x to y.

Therefore if $p(x,y)$ is the transition matrix of the chain (when the states are discrete) or the transition kernel (when the states are continuous)

$$\pi(x)p(x,y) = \pi(y)p(y,x)$$

$$\pi(x)q(x,y)a(x,y) = \pi(y)q(y,x)a(y,x).$$

As we have said, the movement from y to x is not made often enough, so we must define $a(y,x)$ to be as large as possible and since it is a probability we set $a(y,x) = 1$; therefore

$$\pi(x)q(x,y)a(x,y) = \pi(y)q(y,x)$$

and we conclude that the probability of move is produced by the form

$$a(x,y) = \min\left[\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\right]$$

if $\pi(x)q(x,y) > 0$ or 1 otherwise.

Therefore, the **Metropolis-Hasting Algorithm** is:

- Step 1: Generate y from $q(x^t, .)$ and U from $\mathcal{U}(0,1)$.
- Step 2: Let $x^{t+1} = y$ if $U \leq a(x^t, y)$; otherwise let $x^{t+1} = x^t$.
- Repeat the steps 1 and 2 n times in order to take a sample of size n.

The full algorithm is described in Chib and Greenberg (1995).

3.5.2 Choices of $q(x, y)$

There are many options of candidate (proposal) densities and in summary they will be mentioned in this subsection

α) $q(x, y) = q_1(y - x)$

This choice is called the random walk chain because the candidate value y is drawn according to the process $y = x + z$ where the increment random variable $z \sim q_1$. Possible choices for q_1 include the multivariate normal or the multivariate-t distribution.

β) $q(x, y) = q_2(y)$

This choice is called independent chain . Possible choices is the multivariate normal distribution or some other densities and it is required to specify the location of the generating density in addition to the spread.

γ) Other choices

There is a wide set of possible candidate densities. For more information see, Chib and Greenberg (1993), Tierney (1994) and Hastings (1970).

3.6 Remarks

First of all, in order to apply the two algorithms described above, it is needed the equilibrium density to be known up to a constant, something very useful for the Bayesian analysis. Secondly, in the Metropolis-Hastings algorithm if the candidate distribution is a symmetric[1] distribution such as Normal then the probability of move is

[1]Symmetric with respect to it's arguments, $q(x, y) = q(y, x)$

$$a(x, y) = \min\left[\frac{\pi(y)}{\pi(x)}, 1\right].$$

Finally, the two algorithms can be combined, so in the Gibbs sampler if the full conditional densities are not one of the known forms then we can use the Metropolis-Hastings in order to sample from them. This last remark is very important because this combination of the two algorithms construct a powerful tool in order to exploit the properties of the posterior distribution (Chib and Greenberg, 1993).

3.7 Convergence issue

The main problem in the MCMC algorithms is the specification of the convergence to the equilibrium distribution. The theory says that when the number of iterations is large then the theory assures convergence of ergodic averages to the desired state space averages. Of course this number is not known and many authors have proposed many criteria in order to detect the convergence (Brooks and Roberts, 1995, Brooks et al, 1996, Raftery and Lewis, 1992).

Here in order to detect the convergence we use the CODA software (Best & Cowles, 1995) which contains the most popular convergence tests. These tests are represented in the next table.For more details see Best and Cowles (1995).

https://doi.org/10.26219/heal.aueb.5968

1	Geweke
2	Gelman & Rubin
3	Raftery & Lewis
4	Heidelberger & Welch

$\square$

3.8 Examples

In this section we present two examples of the application of the Gibbs sampler. These two examples present the application of the Gibbs sampler in the theory of linear regression model and in the theory of autoregressive time series model (AR(1)).

3.8.1 Linear regression

Let $\mathbf{y} = (y_1, ..., y_T)^T$ be a $T \times 1$ vector of observations and $y_t = a + \beta x_t + u_t$, where $\mathbf{u} = (u_1, ..., u_T)^T \sim \mathrm{N}(0, \mathrm{I}\sigma^2)$ and $\mathbf{x} = (x_1, ..., x_T)^T$ be a $T \times 1$ vector of constants. We are interested for the posterior distribution of the three parameters α, β, σ^2 .

The likelihood of the data is

$$f(\mathbf{y}|\alpha, \beta, \sigma^2, \mathbf{x}) \propto (\sigma^2)^{-\frac{T}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t - \alpha - \beta x_t)^2\right\}.$$

https://doi.org/10.26219/heal.aueb.5968

Non informative case

The non-informative a priori distribution of the three parameters is

$$f(\alpha, \beta, \sigma^2) \;\propto\; \sigma^{-2},$$

the full posterior distribution follows

$$f(\alpha, \beta, \sigma^2|\mathbf{y}, \mathbf{x}) \;\propto\; (\sigma^2)^{-(\frac{T}{2}+1)} \cdot \exp\left\{-\frac{1}{2\sigma^2} \cdot \sum_{t=1}^{T}(y_t - \alpha - \beta x_t)^2\right\}.$$

The full condition distributions take the forms

$$\begin{aligned} f(\alpha|\beta, \sigma^2, \mathbf{y}, \mathbf{x}) \;&\propto\; \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t - \alpha - \beta x_t)^2\right\} \\ &\propto\; \exp\left\{-\frac{T}{2\sigma^2}\left(\alpha - \frac{\sum y_t - \beta \sum x_t}{T}\right)^2\right\} \\ &\equiv\; \mathrm{N}\left(\frac{\sum y_t - \beta \sum x_t}{T}, \frac{\sigma^2}{T}\right) \end{aligned}$$

$$f(\beta|\alpha, \sigma^2, \mathbf{y}, \mathbf{x}) \;\propto\; \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t - \alpha - \beta x_t)^2\right\}$$

https://doi.org/10.26219/heal.aueb.5968

$$\begin{aligned} &\propto \exp\left\{-\frac{\sum x_t^2}{2\sigma^2}\left(\beta - \frac{\sum y_t x_t - \alpha \sum x_t}{\sum x_t^2}\right)^2\right\} \\ &\equiv \mathbf{N}\left(\frac{\sum y_t x_t - \alpha \sum x_t}{\sum x_t^2}, \frac{\sigma^2}{\sum x_t^2}\right) \end{aligned}$$

$$\begin{aligned} f(\sigma^2|\alpha,\beta,\mathbf{y},\mathbf{x}) &\propto (\sigma^2)^{-(\frac{T}{2}+1)} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t - \alpha - \beta x_t)^2\right\} \\ &\equiv \mathbf{IG}\left(\frac{T}{2}, \frac{1}{2}\sum_{t=1}^{T}(y_t - \alpha - \beta x_t)^2\right), \end{aligned}$$

where **IG** denotes the inverse Gamma distribution.

https://doi.org/10.26219/heal.aueb.5968

Informative case

The informative prior distributions of the three parameters are

$$\alpha \sim \mathbf{N}\left(\mu_\alpha, \sigma_\alpha^2\right), \; f(\alpha) \propto \sigma_\alpha^{-2} \exp\left\{-\frac{1}{2\sigma_\alpha^2}(\alpha-\mu_\alpha)^2\right\}$$

$$\beta \sim \mathbf{N}\left(\mu_\beta, \sigma_\beta^2\right), \; f(\beta) \propto \sigma_\beta^{-2} \exp\left\{-\frac{1}{2\sigma_\beta^2}\left(\beta-\mu_\beta\right)^2\right\}$$

$$\sigma^2 \sim \mathbf{IG}\left(\nu_0, \frac{S_0}{2}\right), \; f(\sigma^2) \propto \sigma^{2^{-(\nu_0+1)}} \exp\left\{-\frac{S_0^2}{2\sigma^2}\right\}.$$

The posterior distribution follows

$$\begin{aligned} f(\alpha,\beta,\sigma^2|\mathbf{y},\mathbf{x}) \;\propto\; & (\sigma^2)^{-(\frac{T}{2}+\nu_0+2)} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t-\alpha-\beta x_t)^2\right\}\cdot \\ & \exp\left\{-\frac{1}{2\sigma_\alpha^2}(\alpha-\mu_\alpha)^2\right\} * \exp\left\{-\frac{1}{2\sigma_\beta^2}\left(\beta-\mu_\beta\right)^2\right\}\cdot \\ & \exp\left\{-\frac{S^2}{2\sigma^2}\right\}. \end{aligned}$$

https://doi.org/10.26219/heal.aueb.5968

The full condition distributions are

$$f(\alpha|\beta,\sigma^2,\mathbf{y},\mathbf{x}) \propto \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2}\sum_{t=1}^{T}(y_t-\alpha-\beta x_t)^2+\frac{1}{\sigma_\alpha^2}(\alpha-\mu_\alpha)^2\right)\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left(\frac{T}{\sigma^2}+\frac{1}{\sigma_\alpha^2}\right)\left(\alpha-\frac{\frac{\sum y_t-\beta\sum x_t}{\sigma^2}+\frac{\mu_\alpha}{\sigma_\alpha^2}}{\left(\frac{T}{\sigma^2}+\frac{1}{\sigma_\alpha^2}\right)}\right)^2\right\}$$

$$\equiv \mathbf{N}\left(\frac{\frac{\sum y_t-\beta\sum x_t}{\sigma^2}+\frac{\mu_\alpha}{\sigma_\alpha^2}}{\left(\frac{T}{\sigma^2}+\frac{1}{\sigma_\alpha^2}\right)},\left(\frac{T}{\sigma^2}+\frac{1}{\sigma_\alpha^2}\right)^{-1}\right)$$

$$f(\beta|\alpha,\sigma^2,\mathbf{y},\mathbf{x}) \propto \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2}\sum_{t=1}^{T}(y_t-\alpha-\beta x_t)^2+\frac{1}{\sigma_\beta^2}\left(\beta-\mu_\beta\right)^2\right)\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left(\frac{\sum x_t^2}{\sigma^2}+\frac{1}{\sigma_\beta^2}\right)\left(\beta-\frac{\frac{\sum x_t y_t-\alpha\sum x_t}{\sigma^2}+\frac{\mu_\beta}{\sigma_\beta^2}}{\left(\frac{\sum x_t^2}{\sigma^2}+\frac{1}{\sigma_\beta^2}\right)}\right)^2\right\}$$

$$\equiv \mathbf{N}\left(\frac{\frac{\sum x_t y_t-\alpha\sum x_t}{\sigma^2}+\frac{\mu_\beta}{\sigma_\beta^2}}{\left(\frac{\sum x_t^2}{\sigma^2}+\frac{1}{\sigma_\beta^2}\right)},\left(\frac{\sum x_t^2}{\sigma^2}+\frac{1}{\sigma_\beta^2}\right)^{-1}\right)$$

$$f(\sigma^2|\alpha,\beta,\mathbf{y},\mathbf{x}) \propto \left(\sigma^2\right)^{-\left(\frac{T}{2}+\nu_0+2\right)}\cdot$$

$$\exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{t=1}^{T}(y_t-\alpha-\beta x_t)^2+S_0^2\right)\right\}.$$

https://doi.org/10.26219/heal.aueb.5968

Let

$$S = \frac{1}{2}\left(\sum_{t=1}^{T}(y_t - \alpha - \beta x_t)^2 + S_0^2\right),$$

then

$$\sigma^2 \sim \mathbf{IG}\left((\frac{T}{2} + \nu_0 + 1), S\right),$$

where **IG** denotes the inverse Gamma distribution.

https://doi.org/10.26219/heal.aueb.5968

3.8.2 Autoregressive model

Let y be a $(T+1) \times 1$ vector of observations and $y_t = \mu + \varphi y_{t-1} + u_t$ (Hamilton, 1994 & Marriott et al, 1994) , where $\mathbf{u} = (u_1, ...u_T)^T \sim \mathbf{N}(\mathbf{0}, \mathbf{I}\sigma^2)$. In this case we consider y_0 as known constant. We are interested in the distribution of the three parameters μ, φ, σ^2 .

The likelihood of the data given y_0 is

$$f(\mathbf{y}|\mu, \varphi, \sigma^2, y_0) \propto (\sigma^2)^{-\frac{T}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t - \mu - \varphi y_{t-1})^2\right\}$$

Non informative case

The non-informative a priori distribution of the three parameters is

$$f(\mu, \varphi, \sigma^2) \propto \sigma^{-2}$$

and the full posterior is

$$f(\mu, \varphi, \sigma^2|\mathbf{y}, y_0) \propto (\sigma^2)^{-(\frac{T}{2}+1)} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t - \mu - \varphi y_{t-1})^2\right\}.$$

The full condition distributions take the forms

https://doi.org/10.26219/heal.aueb.5968

$$\begin{aligned} f(\mu|\varphi,\sigma^2,\mathbf{y},y_0) &\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t-\mu-\varphi y_{t-1})^2\right\} \\ &\propto \exp\left\{-\frac{T}{2\sigma^2}\left(\mu-\frac{\sum y_t-\varphi\sum y_{t-1}}{T}\right)^2\right\} \\ &\equiv \mathbf{N}\left(\frac{\sum y_t-\varphi\sum y_{t-1}}{T},\frac{\sigma^2}{T}\right) \end{aligned}$$

$$\begin{aligned} f(\varphi|\mu,\sigma^2,\mathbf{y},y_0) &\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t-\mu-\varphi y_{t-1})^2\right\} \\ &\propto \exp\left\{-\frac{\sum y_{t-1}^2}{2\sigma^2}\left(\varphi-\frac{\sum y_t y_{t-1}-\mu\sum y_{t-1}}{\sum y_{t-1}^2}\right)^2\right\} \\ &\equiv \mathbf{N}\left(\frac{\sum y_t y_{t-1}-\mu\sum y_{t-1}}{\sum y_{t-1}^2},\frac{\sigma^2}{\sum y_{t-1}^2}\right) \end{aligned}$$

$$\begin{aligned} f(\sigma^2|\mu,\varphi,\mathbf{y},y_0) &\propto (\sigma^2)^{-(\frac{T}{2}+1)}\exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t-\mu-\varphi y_{t-1})^2\right\} \\ &\equiv \mathbf{IG}\left(\frac{T}{2},\frac{1}{2}\sum_{t=1}^{T}(y_t-\mu-\varphi y_{t-1})^2\right), \end{aligned}$$

where **IG** denotes the inverse Gamma distribution.

https://doi.org/10.26219/heal.aueb.5968

Informative case

The informative prior distributions of the three parameters are

$$\mu \sim \mathbf{N}\left(\mu_{\mu}, \sigma_{\mu}^{2}\right), \quad f(\mu) \propto \sigma_{\mu}^{-2} \exp\left\{-\frac{1}{2\sigma_{\mu}^{2}}\left(\mu-\mu_{\mu}\right)^{2}\right\}$$

$$\varphi \sim \mathbf{N}\left(\mu_{\varphi}, \sigma_{\varphi}^{2}\right), \quad f(\varphi) \propto \sigma_{\varphi}^{-2} \exp\left\{-\frac{1}{2\sigma_{\varphi}^{2}}\left(\varphi-\mu_{\varphi}\right)^{2}\right\}$$

$$\sigma^{2} \sim \mathbf{IG}\left(\nu_{0}, \frac{S_{0}}{2}\right), \quad f(\sigma^{2}) \propto \sigma^{2^{-(\nu_{0}+1)}} \exp\left\{-\frac{S_{0}^{2}}{2\sigma^{2}}\right\}.$$

The posterior distribution follows

$$\begin{aligned} f(\mu, \varphi, \sigma^{2}|\mathbf{y}, y_{0}) \propto\; & (\sigma^{2})^{-(\frac{T}{2}+\nu_{0}+2)} \exp\left\{-\frac{1}{2\sigma^{2}}\sum_{t=1}^{T}(y_{t}-\mu-\varphi y_{t-1})^{2}\right\}\cdot \\ & \exp\left\{-\frac{1}{2\sigma_{\mu}^{2}}\left(\mu-\mu_{\mu}\right)^{2}\right\}\cdot \exp\left\{-\frac{1}{2\sigma_{\varphi}^{2}}\left(\varphi-\mu_{\varphi}\right)^{2}\right\}\cdot \\ & \exp\left\{-\frac{S^{2}}{2\sigma^{2}}\right\}. \end{aligned}$$

The full condition distributions are

https://doi.org/10.26219/heal.aueb.5968

$$
\begin{aligned}
f(\mu|\varphi,\sigma^2,\mathbf{y},y_0) &\propto \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2}\sum_{t=1}^{T}(y_t-\mu-\varphi y_0)^2+\frac{1}{\sigma_\mu^2}\left(\mu-\mu_\mu\right)^2\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\frac{T}{\sigma^2}+\frac{1}{\sigma_\mu^2}\right)\left(\mu-\frac{\frac{\sum y_t-\varphi\sum y_{t-1}}{\sigma^2}+\frac{\mu_\mu}{\sigma_\mu^2}}{\left(\frac{T}{\sigma^2}+\frac{1}{\sigma_\mu^2}\right)}\right)^2\right\} \\
&\equiv \mathbf{N}\left(\frac{\frac{\sum y_t-\varphi\sum y_{t-1}}{\sigma^2}+\frac{\mu_\mu}{\sigma_\mu^2}}{\left(\frac{T}{\sigma^2}+\frac{1}{\sigma_\mu^2}\right)},\left(\frac{T}{\sigma^2}+\frac{1}{\sigma_\mu^2}\right)^{-1}\right)
\end{aligned}
$$

$$
\begin{aligned}
f(\varphi|\mu,\sigma^2,\mathbf{y},y_0) &\propto \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2}\sum_{t=1}^{T}(y_t-\mu-\varphi y_{t-1})^2+\frac{1}{\sigma_\varphi^2}\left(\varphi-\mu_\varphi\right)^2\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\frac{\sum y_{t-1}^2}{\sigma^2}+\frac{1}{\sigma_\varphi^2}\right)\left(\varphi-\frac{\frac{\sum y_{t-1}y_t-\mu\sum y_{t-1}}{\sigma^2}+\frac{\mu_\varphi}{\sigma_\varphi^2}}{\left(\frac{\sum y_{t-1}^2}{\sigma^2}+\frac{1}{\sigma_\varphi^2}\right)}\right)^2\right\} \\
&\equiv \mathbf{N}\left(\frac{\frac{\sum y_{t-1}y_t-\mu\sum y_{t-1}}{\sigma^2}+\frac{\mu_\varphi}{\sigma_\varphi^2}}{\left(\frac{\sum y_{t-1}^2}{\sigma^2}+\frac{1}{\sigma_\varphi^2}\right)},\left(\frac{\sum y_{t-1}^2}{\sigma^2}+\frac{1}{\sigma_\varphi^2}\right)^{-1}\right)
\end{aligned}
$$

$$
\begin{aligned}
f(\sigma^2|\mu,\varphi,\mathbf{y},y_0) &\propto (\sigma^2)^{-(\frac{T}{2}+\nu_0+2)} \\
&\cdot\exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{t=1}^{T}(y_t-\mu-\varphi y_{t-1})^2+S_0^2\right)\right\}.
\end{aligned}
$$

https://doi.org/10.26219/heal.aueb.5968

Let

$$S = \left(\sum_{t=1}^{T}(y_t - \mu - \varphi y_{t-1})^2 + S_0^2\right),$$

then

$$\sigma^2 \sim \mathbf{IG}\left((\frac{T}{2} + \nu_0 + 1), S\right).$$

where **IG** denotes the inverse Gamma distribution.

https://doi.org/10.26219/heal.aueb.5968

Chapter 4

Volatility Time Series models

4.1 Introduction

Uncertainty is central to much of modern finance theory. According to most asset pricing theories the risk premium is determined by the covariance between the future return on the asset and one or more benchmark portfolios; e.g. the market portfolio or the growth rate in consumption. In option pricing the uncertainty associated with the future price of the underlying asset is the most important determinant in the pricing function. The construction of hedge portfolios is another example where the conditional future variances and covariances among the different assets involved play an important role (Bollerslev, Chou and Kroner, 1992).

It has been recognized that, the uncertainty of prices, as measured by the variances and covariances, is changing through time. In order to model such changes in the variance a category of models has been proposed, such as Autoregressive Conditional Heteroskedasticity model (ARCH), Generalized Autoregressive Conditional Heteroskedasticity model (GARCH) and Stochastic Volatility model (SV). These models have been studied via the classical statistical approach.

In this dissertation we will study the **Stochastic Volatility Model** via the Bayesian approach.

4.2 Classifying models of changing volatility

All these models can be classified in two categories. These two categories are: **Observation-driven models** and **Parameter-driven model**.

In order to discuss these two categories we will assume

$$y_t|_{z_t} \sim \mathcal{N}(0, \sigma_t^2)$$

where the z_t contains all the past information up to time t.

4.2.1 Parameter-driven model

Parameter-driven or **state space models** allow z_t to be a function of some unobserved or "latent" component. The main representative of this class of models is the Stochastic Volatility model (SV). The general form of SV model is the follow

$$y_t|_{z_t} \sim \mathcal{N}(0, \exp(h_t))$$

$$h_t = a + d \cdot h_{t-1} + \eta_t$$

$$\eta_t \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_\eta^2\right).$$

Here the latent parameters h_t/s are unobserved but they can be estimated using the observations.

4.2.2 Observation-driven model

In this class of models we put z_t as a function of lagged values of y_t. The simplest example is the Autoregressive Conditional Heteroscedasticity (ARCH) model.

$$y_t|_{z_t} \sim \mathcal{N}(0, \sigma_t^2)$$

$$\sigma_t^2 = a_0 + a_1 y_{t-1}^2 + \ldots + a_p y_{t-p}^2.$$

https://doi.org/10.26219/heal.aueb.5968

4.3 ARCH, GARCH and others observation-driven models

The simplest ARCH(1) model puts:

$$y_t = \varepsilon_t \sigma_t$$

$$\sigma_t^2 = a_0 + a_1 y_{t-1}^2$$

$$\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0,1)$$

and it had been proposed by the Engle (1982). The GARCH model is an extension of the ARCH model and it had been proposed by the Bollerslev (1986). The simple GARCH(1,1) model puts

$$y_t = \varepsilon_t \sigma_t$$

$$\sigma_t^2 = a_0 + a_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

$$\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0,1)$$

and in order to be the time series covariance stationary, the parameters must follow the condition

$$a_1 + \beta_1 < 1.$$

https://doi.org/10.26219/heal.aueb.5968

In cases where

$$a_1 + \beta_1 = 1$$

the model is called Integrated GARCH (IGARCH). Other types of these models are: Weak GARCH, Unobserved ARCH, Log GARCH, Exponential GARCH. etc. For more information see Shephard (1994).

4.4 Stochastic Volatility Model

The basic alternative to ARCH class of models is the **Stochastic Volatility model** which belongs to the parameter-driven models. In this class of models the variance conditional σ_t^2 depends, not on past observation such as ARCH and GARCH, but on some unobserved component or latent structure.

The general form of Stochastic Volatility Model is

$$\begin{aligned} y_t &= \varepsilon_t \exp(h_t/2) \\ h_t &= a + dh_{t-1} + \eta_t \\ &\eta_t \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_\eta^2\right) \\ &\varepsilon_t \overset{iid}{\sim} \mathcal{N}\left(0, 1\right). \end{aligned}$$

This model has been proposed by Taylor (1986). We can interpret the latent variable h_t as a random and uneven flow of new information (which is difficult to model directly) into financial markets. The parameter d can be interpreted as the persistent of the volatility. The σ_η^2 is the volatility of the latent parameters $\mathbf{h}$ (log-volatilities).

4.4.1 Basic properties

In the above formula of the SV model the ε_t and η_t are assumed to be independent of one another Gaussian white noise. As η_t is gaussian then h_t is a standard Gaussian autoregressive model (AR(1)). In order the series of log-volatilities to be covariance stationary we must impose that $|d| < 1$. If $d = 1$ then the series is random walk.

The mean and the variance of the latent parameter h_t are

$$\begin{aligned} \mu_h &= E(h_t) = \frac{a}{1-d} \\ \sigma_h^2 &= \mathcal{V}ar(h_t) = \frac{\sigma_\eta^2}{1-d^2}. \end{aligned}$$

The odds moments of the time series are zero and the even moments can be found from the formula

$$\begin{aligned} E(y_t^r) &= E\left[(\varepsilon_t)^r\right] E\left[\exp(\frac{r}{2}h_t)\right] = \\ &= \frac{r!}{(2)^{r/2}(r/2)!}\exp\left(\frac{r}{2}\mu_h + r^2\sigma_h^2/8\right). \end{aligned}$$

https://doi.org/10.26219/heal.aueb.5968

Moreover, the coefficient of kyrtosis can be found if we apply the next formula

$$\frac{E\left(y_t^4\right)}{\left(\sigma_{y^2}^2\right)^2} = 3\exp\left(\sigma_h^2\right) > 3.$$

From the above formulas, it is obvious that the distribution of the time series y_t is leptokyrtic and symmetric. The dynamic properties of the SV model appear most clearly if we square y_t and take it in the log-scale. Then

$$\log\left(y_t^2\right) = h_t + \log\left(\varepsilon_t^2\right).$$

The $\log(y_t^2)$ is the sum of an AR(1) component and a white noise, so its autocorrelation function (ACF) is equivalent to the ACF of an ARMA(1,1) .

4.4.2 Estimation of the parameters of the SV model

The main disadvantage of the SV models is that, unlike with the ARCH models, it is not immediately clear how to evaluate the likelihood. Like most non-Gaussian parameter-driven models, there many ways to perform estimation.

The main methods that are used is:

1) Generalized method-of-moments (GMM)

2) Quasi-likelihood

3) MCMC

The last years there are a huge interest about the MCMC estimation and in this dissertation we focus our interest in this philosophy of estimating the parameters of the SV model.

https://doi.org/10.26219/heal.aueb.5968

Chapter 5

Bayesian approach of the SV model

5.1 Introduction

The Stochastic Volatility model views the observed data $\mathbf{y}$, as a vector generated from a probability model $p(\mathbf{y}|\mathbf{h})$, where $\mathbf{h}$ is the vector of the latent parameters (log-volatilities). Each data point y_t has conditional variance $\exp(h_t)$, which is dependent on time. The latent parameters $\mathbf{h}$ are unobserved and they are produced from a probability mechanism $p(\mathbf{h}|\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (a, d, \sigma_h^2)$. The likelihood of the data is a mixture over the probabilistic space of the log-volatilities, therefore

$$p(\mathbf{y}|\boldsymbol{\theta}) = \int p(\mathbf{y}|\mathbf{h})p(\mathbf{h}|\boldsymbol{\theta})\mathrm{d}\mathbf{h}.$$

https://doi.org/10.26219/heal.aueb.5968

The main problem of the above formula is that a T-dimensional integration is needed. In order to handle this problem we use the MCMC algorithm via a Bayesian framework.

5.2 Hierarchical structure of SV.

As it is shown in the previous section, we can view the SV model via a hierarchical structure of the conditional distributions. This hierarchy can be specified by the sequence of three conditional distributions, $p(\mathbf{y}|\mathbf{h})$, $p(\mathbf{h}|\boldsymbol{\theta})$ and the prior distribution $p(\boldsymbol{\theta})$.

In the Bayesian approach the subjective information is included in the prior distribution of the parameters, $\boldsymbol{\theta} = (a, d, \sigma_h^2)$. This prior distribution of the hyperparameters a, d and σ_h^2 could be non-informative or informative.

The posterior density of the latent parameters and the hyperparameters can be found by applying the Bayes's theorem

$$p(\boldsymbol{\theta}, \mathbf{h}) \propto p(\mathbf{y}|\mathbf{h})p(\mathbf{h}|\boldsymbol{\theta})p(\boldsymbol{\theta}).$$

5.2.1 Non-informative case

In this case, the prior distributions of the model's parameters are

$$\begin{aligned} p(a) &= p(d) \propto 1 \\ p(\sigma_\eta^2) &\propto \frac{1}{\sigma_\eta^2}, \end{aligned}$$

so the full prior distribution is

$$p(\boldsymbol{\theta}) = \mathbf{p}\left(a, d, \sigma_\eta^2\right) \propto \frac{1}{\sigma_\eta^2}.$$

The distribution of interest takes the form

$$p(\mathbf{y}|\mathbf{h}) \propto \exp\left[-\frac{\sum_{t=1}^{T} h_t}{2}\right] \cdot \exp\left[-\frac{1}{2} \cdot \sum_{t=1}^{T} y_t^2 \cdot \exp(-h_t)\right]$$

$$p(\mathbf{h}|\boldsymbol{\theta}) \propto \frac{1}{\sigma_\eta^T} \cdot \exp\left[-\frac{1}{2 \cdot \sigma_h^2} \cdot \sum_{t=1}^{T} (h_t - a - d \cdot h_{t-1})^2\right]$$

$$p(\boldsymbol{\theta}) = \mathbf{p}\left(a, d, \sigma_\eta^2\right) \propto \frac{1}{\sigma_\eta^2}$$

and applying the Bayes's theorem we take the posterior distribution

$$p(\mathbf{h}, \boldsymbol{\theta}|\mathbf{y}) \propto \exp\left[-\frac{\sum_{t=1}^{T} h_t}{2}\right] \cdot \exp\left[-\frac{1}{2} \cdot \sum_{t=1}^{T} y_t^2 \cdot \exp(-h_t)\right] \cdot$$
$$\frac{1}{\left(\sigma_\eta^2\right)^{\frac{T}{2}+1}} \cdot \exp\left[-\frac{1}{2 \cdot \sigma_h^2} \cdot \sum_{t=1}^{T} (h_t - a - d \cdot h_{t-1})^2\right].$$

https://doi.org/10.26219/heal.aueb.5968

5.2.2 Informative case

In this case, the prior distributions of the model's parameters are

$$p(a) \propto \exp\left[-\frac{1}{2\cdot s_a^2}\cdot(a-\mu_a)^2\right] \equiv \mathbf{Normal}\left(\mu_a, s_a^2\right)$$

$$p(d) \propto \exp\left[-\frac{1}{2\cdot s_d^2}\cdot(d-\mu_d)^2\right] \equiv \mathbf{Normal}\left(\mu_d, s_d^2\right)$$

$$p(\sigma_\eta^2) \propto \frac{1}{\sigma_\eta^{2v_0}}\cdot\exp\left[-\frac{S_0^2}{2\cdot\sigma_\eta^2}\right] \equiv \mathbf{IG}\left(v_0-1,\frac{S_0^2}{2}\right).$$

The full prior distribution is

$$p(\boldsymbol{\theta}) = \mathbf{p}\left(a,d,\sigma_\eta^2\right) = \frac{1}{\sigma_\eta^{2v_0}}\cdot\exp\left[-\frac{1}{2}\left(\frac{1}{s_a^2}\cdot(a-\mu_a)^2+\frac{1}{s_d^2}\cdot(d-\mu_d)^2+\frac{S_0^2}{\sigma_\eta^2}\right)\right].$$

The distribution of interest takes the form

$$p(\mathbf{y}|\mathbf{h}) \propto \exp\left[-\frac{\sum_{t=1}^{T}h_t}{2}\right]\cdot\exp\left[-\frac{1}{2}\cdot\sum_{t=1}^{T}y_t^2\cdot\exp(-h_t)\right]$$

$$p(\mathbf{h}|\boldsymbol{\theta}) \propto \frac{1}{\sigma_\eta^T}\cdot\exp\left[-\frac{1}{2\cdot\sigma_h^2}\cdot\sum_{t=1}^{T}(h_t-a-d\cdot h_{t-1})^2\right]$$

$$p(\boldsymbol{\theta}) = \mathbf{p}\left(a,d,\sigma_\eta^2\right) = \frac{1}{\sigma_\eta^{2v_0}}\cdot\exp\left[-\frac{1}{2}\left(\frac{1}{s_a^2}\cdot(a-\mu_a)^2+\frac{1}{s_d^2}\cdot(d-\mu_d)^2+\frac{S_0^2}{\sigma_\eta^2}\right)\right]$$

so applying the Bayes's theorem we take the posterior distribution

$$
\begin{aligned}
p(\mathbf{h},\boldsymbol{\theta}|\mathbf{y}) \quad \propto \quad & \exp\left[-\frac{\sum_{t=1}^{T} h_t}{2}\right] \cdot \exp\left[-\frac{1}{2} \cdot \sum_{t=1}^{T} y_t^2 \cdot \exp\left(-h_t\right)\right] \cdot \\
& \frac{1}{\left(\sigma_\eta^2\right)^{\frac{T}{2}+v_0+1}} \cdot \exp\left[-\frac{1}{2 \cdot \sigma_h^2} \cdot \sum_{t=1}^{T} \left(h_t - a - d \cdot h_{t-1}\right)^2\right] \cdot \\
& \exp\left[-\frac{1}{2}\left(\frac{1}{s_a^2} \cdot (a-\mu_a)^2 + \frac{1}{s_d^2} \cdot (d-\mu_d)^2 + \frac{S_0^2}{\sigma_\eta^2}\right)\right].
\end{aligned}
$$

5.3 The MCMC algorithm

5.3.1 The conditional distributions

In order to apply the *Gibbs sampler* we need the full conditional distributions. In the next sections we represent the full conditional distribution for each parameter of the SV model.

Conditional distribution of parameter a

Non-Informative case

$$
\begin{aligned}
p(a|\cdot) \quad \propto \quad & \exp\left[-\frac{1}{2 \cdot \sigma_\eta^2} \cdot \sum_{t=1}^{T} \left(h_t - a - d \cdot h_{t-1}\right)^2\right] \\
= \quad & \exp\left[-\frac{T}{2 \cdot \sigma_\eta^2} \cdot \left(a - \frac{\sum h_t - d \cdot \sum h_{t-1}}{T}\right)^2\right]
\end{aligned}
$$

https://doi.org/10.26219/heal.aueb.5968

$$\equiv \text{Normal}\left(\frac{\sum h_t - d \cdot \sum h_{t-1}}{T}, \frac{\sigma_\eta^2}{T}\right).$$

Informative case

$$p(a|\cdot) \propto \exp\left[-\frac{1}{2 \cdot \sigma_\eta^2} \cdot \sum_{t=1}^{T} (h_t - a - d \cdot h_{t-1})^2 - \frac{1}{2 \cdot s_a^2} \cdot (a - \mu_a)^2\right]$$

$$= \exp\left[-\frac{T \cdot s_a^2 + \sigma_\eta^2}{2 \cdot \sigma_\eta^2 \cdot s_a^2} \cdot \left(a - \frac{\frac{\sum h_t - d \cdot \sum h_{t-1}}{\sigma_\eta^2} + \frac{\mu_a}{s_a^2}}{\frac{T \cdot s_a^2 + \sigma_\eta^2}{\sigma_\eta^2 \cdot s_a^2}}\right)^2\right]$$

$$\equiv \text{Normal}\left(\frac{s_a^2 \cdot (\sum h_t - d \cdot \sum h_{t-1}) + \sigma_\eta^2 \cdot \mu_a}{s_a^2 - \sigma_\eta^2}, \frac{T \cdot s_a^2 + \sigma_\eta^2}{\sigma_\eta^2 \cdot s_a^2}\right).$$

Conditional distribution of parameter d

Non-Informative case

$$p(d|\cdot) \propto \exp\left[-\frac{1}{2 \cdot \sigma_\eta^2} \cdot \sum_{t=1}^{T} (h_t - a - d \cdot h_{t-1})^2\right]$$

$$= \exp\left[-\frac{\sum h_{t-1}^2}{2 \cdot \sigma_\eta^2} \cdot \left(d - \frac{\sum h_t \cdot h_{t-1} - a \cdot \sum h_{t-1}}{\sum h_{t-1}^2}\right)^2\right]$$

$$\equiv \text{Normal}\left(\frac{\sum h_t \cdot h_{t-1} - a \cdot \sum h_{t-1}}{\sum h_{t-1}^2}, \frac{\sigma_\eta^2}{\sum h_{t-1}^2}\right).$$

https://doi.org/10.26219/heal.aueb.5968

Informative case

$$\begin{aligned}
p(d|\cdot) &\propto \exp\left[-\frac{1}{2\cdot\sigma_\eta^2}\cdot\sum_{t=1}^{T}(h_t-a-d\cdot h_{t-1})^2-\frac{1}{2\cdot s_d^2}\cdot(d-\mu_d)^2\right]\\
&= \exp\left[-\frac{s_d^2\cdot\sum h_{t-1}^2+\sigma_\eta^2}{2\cdot\sigma_\eta^2\cdot s_d^2}\cdot\left(d-\frac{\frac{\sum h_t\cdot h_{t-1}-a\cdot\sum h_{t-1}}{\sigma_\eta^2}+\frac{\mu_d}{s_d^2}}{\frac{s_d^2\cdot\sum h_{t-1}^2+\sigma_\eta^2}{\sigma_\eta^2\cdot s_d^2}}\right)^2\right]\\
&\equiv \textbf{Normal}\left(\frac{s_d^2\cdot(\sum h_t\cdot h_{t-1}-a\cdot\sum h_{t-1})+\sigma_\eta^2\cdot\mu_d}{s_d^2\cdot\sum h_{t-1}^2+\sigma_\eta^2},\frac{s_d^2\cdot\sum h_{t-1}^2+\sigma_\eta^2}{\sigma_\eta^2\cdot s_d^2}\right).
\end{aligned}$$

Conditional distribution of parameter σ_η^2

Non-Informative case

$$\begin{aligned}
p(\sigma_\eta^2|\cdot) &\propto \frac{1}{\left(\sigma_\eta^2\right)^{\frac{T}{2}+1}}\cdot\exp\left[-\frac{1}{2\cdot\sigma_\eta^2}\cdot\sum_{t=1}^{T}(h_t-a-d\cdot h_{t-1})^2\right]\\
&\equiv \textbf{IG}\left(\frac{T}{2},\frac{1}{2}\cdot\sum_{t=1}^{T}(h_t-a-d\cdot h_{t-1})^2\right)
\end{aligned}$$

where IG is the inverse gamma distribution.

Informative case

$$p(\sigma_\eta^2|\cdot) \propto \frac{1}{\left(\sigma_\eta^2\right)^{\frac{T}{2}+v_0+1}}\cdot\exp\left[-\frac{1}{2\cdot\sigma_\eta^2}\cdot\sum_{t=1}^{T}(h_t-a-d\cdot h_{t-1})^2-\frac{S_0^2}{2\cdot\sigma_\eta^2}\right]$$

$$\equiv \text{IG}\left(\frac{T}{2}+v_0, \frac{1}{2}\left(\sum_{t=1}^{T}(h_t-a-d\cdot h_{t-1})^2+S_0^2\right)\right)$$

where IG is the inverse gamma distribution.

Conditional distribution of latent parameters h

$$p(\mathbf{h}|\cdot) \propto \exp\left[-\frac{\sum_{t=1}^{T}h_t}{2}\right]\cdot\exp\left[-\frac{1}{2}\cdot\sum_{t=1}^{T}y_t^2\cdot\exp(-h_t)\right]\cdot$$
$$\exp\left[-\frac{1}{2\cdot\sigma_\eta^2}\cdot\sum_{t=1}^{T}(h_t-a-d\cdot h_{t-1})^2\right].$$

$$\begin{aligned} p(h_0|h_1,\cdot) &\propto \exp\left[-\frac{1}{2\cdot\sigma_\eta^2}\cdot(h_1-a-d\cdot h_0)^2\right] \\ &= \exp\left[-\frac{d^2}{2\cdot\sigma_\eta^2}\cdot\left(h_0-\frac{h_1-a}{d}\right)^2\right] \\ &\equiv \text{Normal}\left(\frac{h_1-a}{d},\frac{\sigma_\eta^2}{d^2}\right). \end{aligned}$$

5.3.2 The algorithm

The algorithm of the Gibbs sampler has been described in a previous chapter and this algorithm is applied here but some difficulties appear.

The full conditional distribution of the latent parameters is not one of the known forms so we must use Metropolis sampler in order to have sample from it. Moreover this density has T-dimension, so it is well understood that, in order to sample from it requires huge computational effort. On the other hand, this distribution can be seen as the product of the distribution of each latent parameter

$$p(\mathbf{h}|\cdot) = \prod_{t=1}^{T} p(h_t|h_{t-1},\cdot).$$

The univariate conditional densities have the form

$$\begin{aligned} p(h_t|h_{t-1},h_{t+1},\cdot) &\propto \exp\left[-\frac{1}{2}\cdot\left(h_t+y_t^2\cdot\exp\left(-h_t\right)\right)\right]\cdot \\ &\quad \exp\left[-\frac{1}{2\cdot\sigma_\eta^2}\cdot\left((h_t-a-d\cdot h_{t-1})^2+(h_{t+1}-a-d\cdot h_t)^2\right)\right] \\ &\propto \exp\left[-\frac{1}{2}\cdot\left(h_t+y_t^2\cdot\exp\left(-h_t\right)\right)\right]\cdot \\ &\quad \exp\left[-\frac{1}{2\cdot s^2}\cdot(h_t-m_t)^2\right], \end{aligned}$$

where

$$m_t = \frac{[a\cdot(1-d)+d\cdot(h_{t+1}+h_{t-1})]}{(1+d^2)}$$

and

$$s^2 = \frac{\sigma_\eta^2}{(1+d)}.$$

https://doi.org/10.26219/heal.aueb.5968

The above density has not a trivial form. It is consisted of a Normal term and another part that is very unusual.

In order to sample from it, we use the Metropolis-Hasting algorithm. In detail, we use the dependent Metropolis-Hastings algorithm and as proposal we use a Normal density. So the $i-th$ value of the h_t is sampled from

$$h_t^{(i)} \sim \textbf{Normal}\left(h_t^{(i-1)}, c \cdot \sigma_\eta^2\right)$$

and the probability of acceptance is

$$a\left(h_t^{(i-1)} \to h_t^{(i)}\right) = \frac{\exp\left[-\frac{1}{2} \cdot \left(h_t^{(i)} + y_t^2 \cdot \exp\left(-h_t^{(i)}\right)\right)\right]}{\exp\left[-\frac{1}{2} \cdot \left(h_t^{(i-1)} + y_t^2 \cdot \exp\left(-h_t^{(i-1)}\right)\right)\right]} \cdot \frac{\exp\left[-\frac{1}{2 \cdot s^2} \cdot \left(h_t^{(i)} - m_t\right)^2\right]}{\exp\left[-\frac{1}{2 \cdot s^2} \cdot \left(h_t^{(i-1)} - m_t\right)^2\right]}.$$

Chib and Greenberg (1994) tell us that the constant c must chosen such that the acceptance probability is 50% or more, but there is a debate in scientific community about this issue. For this situation we choose the constant c equal to $\frac{1}{(1+d^2)}$ which gives very good acceptance probability, approximately 65%.

Finally, the algorithm takes the form:

1) Give initial values to: $\mathbf{h}^{(0)}, a^{(0)}, d^{(0)}, \sigma_h^{2(0)}$.

2) Sample $a^{(1)} \sim p(a|\cdot)$.

3) Sample $d^{(1)} \sim p(d|\cdot)$.

4) Sample $\sigma_\eta^{2(1)} \sim p(\sigma_\eta^2|\cdot)$.

5) Sample $h_0^{(1)} \sim p(h_0|\cdot)$.

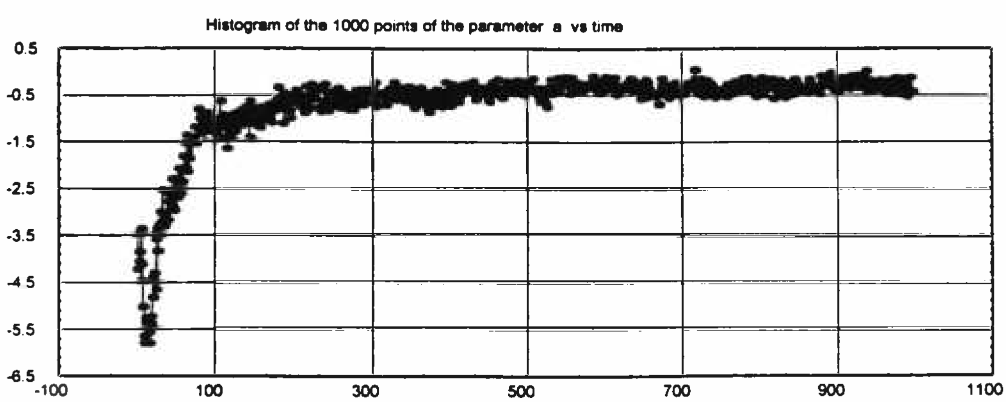
6) Sample $h_t^{(1)} \sim p(h_t|h_{t+1}, h_{t-1}, \cdot)$ for t=1,...,T.

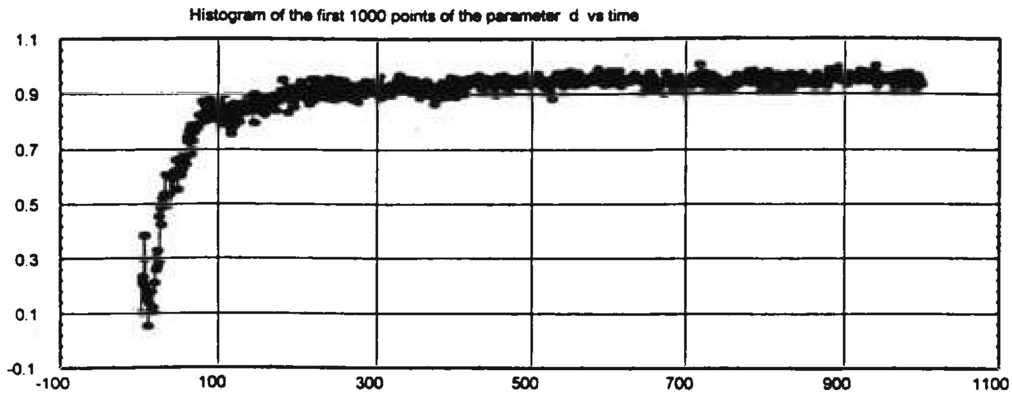
7) Repeat steps **2 - 6** n times with $n \to +\infty$.

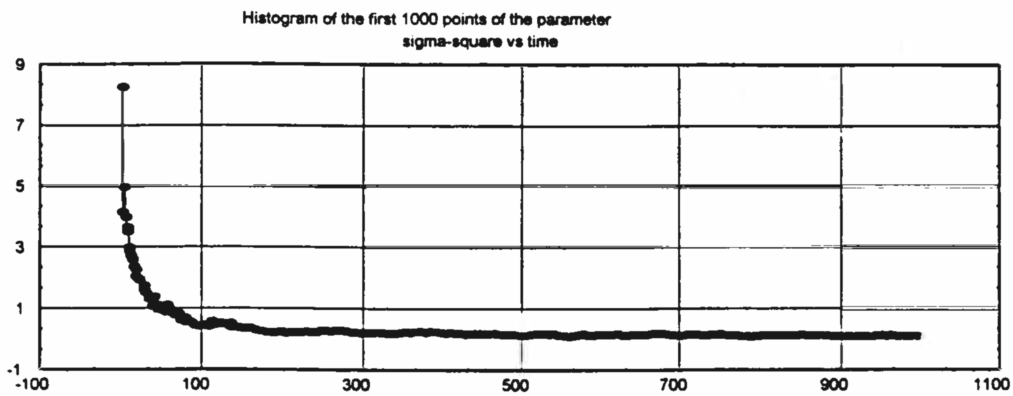
The sample we finally collect comes from the posterior distribution of interest.

5.3.3 Convergence issues

The main problem of the SV models is that the MCMC sampler converges slowly to the equilibrium distribution. This slowness is caused because of the high autocorrelation of the parameters and the correlation between the parameters of the SV models. The following figures show the first 1000 points of the MCMC algorithm.







In order to handle this problem we follow three different strategies. The first is based on a reparameterisation of the latent parameters, in the second we use the random scan MCMC and finally the non-sequential MCMC is used.

Reparameteterisation

In order to speed up the convergence of the MCMC we use a reparameterisation of the latent parameters . A proper reparameterisation is usually a

https://doi.org/10.26219/heal.aueb.5968

good way to speed up the MCMC algorithm (Phillips and Smith, 1993 & Dellaportas, 1995). The general sketch of this reparameterisation follows.

For $h_0 = 0$ we have

$$\begin{aligned}
w_1 &= h_1 - a \\
w_2 &= h_2 - a - d \cdot h_1 \\
w_3 &= h_3 - a - d \cdot h_2 \\
&\vdots \\
w_T &= h_T - a - d \cdot h_{T-1}
\end{aligned}$$

so we have that

$$h_k = a \cdot \frac{d^k - 1}{d - 1} + \sum_{i=1}^{k} d^{k-i} \cdot w_i.$$

https://doi.org/10.26219/heal.aueb.5968

$$\frac{\partial(h_1,...,h_T)}{\partial(w_1,...,w_T)} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ d & 1 & 0 & 0 & \cdots & 0 \\ d^2 & d & 1 & 0 & \cdots & 0 \\ d^3 & d^2 & d & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ d^{T-1} & d^{T-2} & d^{T-3} & d^{T-4} & \cdots & 1 \end{bmatrix}$$

and

$$J = \left|\det\left(\frac{\partial(h_1,...,h_T)}{\partial(w_1,...,w_T)}\right)\right| = 1$$

The full posterior distribution after the reparameterisation is

$$\begin{aligned} p_{\mathbf{w}}(a,d,\sigma_\eta^2,\mathbf{w}|\mathbf{y}) &= p_{\mathbf{h}}(a,d,\sigma_\eta^2,\mathbf{w}|\mathbf{y})\cdot J \\ &\propto \frac{1}{\left(\sigma_\eta^2\right)^{\frac{T}{2}+1}}\cdot\exp\left[-\frac{1}{2}\cdot\frac{\sum_{t=1}^{T}w_t^2}{\sigma_h^2}\right]\cdot \\ &\quad \exp\left[-\frac{1}{2}\cdot\sum_{t=1}^{T}\left(a\cdot\frac{d^t-1}{d-1}+\sum_{i=1}^{t}d^{t-i}\cdot w_i\right)\right]\cdot \\ &\quad \exp\left[-\frac{1}{2}\cdot\sum_{t=1}^{T}y_t^2\cdot\exp\left(-a\cdot\frac{d^t-1}{d-1}-\sum_{i=1}^{t}d^{t-i}\cdot w_i\right)\right] \end{aligned}$$

https://doi.org/10.26219/heal.aueb.5968

$$\propto \frac{1}{\left(\sigma_\eta^2\right)^{\frac{T}{2}+1}} \cdot \exp\left[-\frac{1}{2} \cdot \frac{\sum_{t=1}^{T} w_t^2}{\sigma_\eta^2}\right] \cdot$$

$$\exp\left[-\frac{1}{2} \cdot \left(\frac{a}{d-1} \cdot \left(\frac{d^t - d}{d-1} - T\right) + \sum_{t=1}^{T} w_t \cdot \frac{d^{T-t+1} - 1}{d-1}\right)\right] \cdot$$

$$\exp\left[-\frac{1}{2} \cdot \sum_{t=1}^{T} y_t^2 \cdot \exp\left(-a \cdot \frac{d^t - 1}{d-1} - \sum_{i=1}^{t} d^{t-i} \cdot w_i\right)\right].$$

The above reparameterisation gives enough evidence that it speeds up the convergence and that reduces the autocorrelation between the parameters of interest.

Other strategies

The second strategy was the random scan MCMC. The main difference from the usual MCMC algorithm that we present in the section 5.3.2, is that the latent parameters are updated not in a sequential order but randomly. This strategy has an advantage to the convergence with respect to the usual MCMC algorithm but it is not as good as the reparameterisation strategy.

Another approach of the MCMC is the non-sequential MCMC, which does not update the latent parameters in a sequential order but with a step each time. For example, for step 5 we update h_1, h_6, $h_{11}, \ldots$.

The outcomes of this strategy show that the non-sequential MCMC does not affect on the convergence of the parameters of the SV model.

Chapter 6

Implementation

6.1 Introduction

In this study we try to investigate the weekly return of the General Index of the Athens Stock Exchange Market (composite Index) over the period 1986 - 1996. The dataset consists of 491 observations of weekly return.

We define the weekly rate of return of the General index as follow

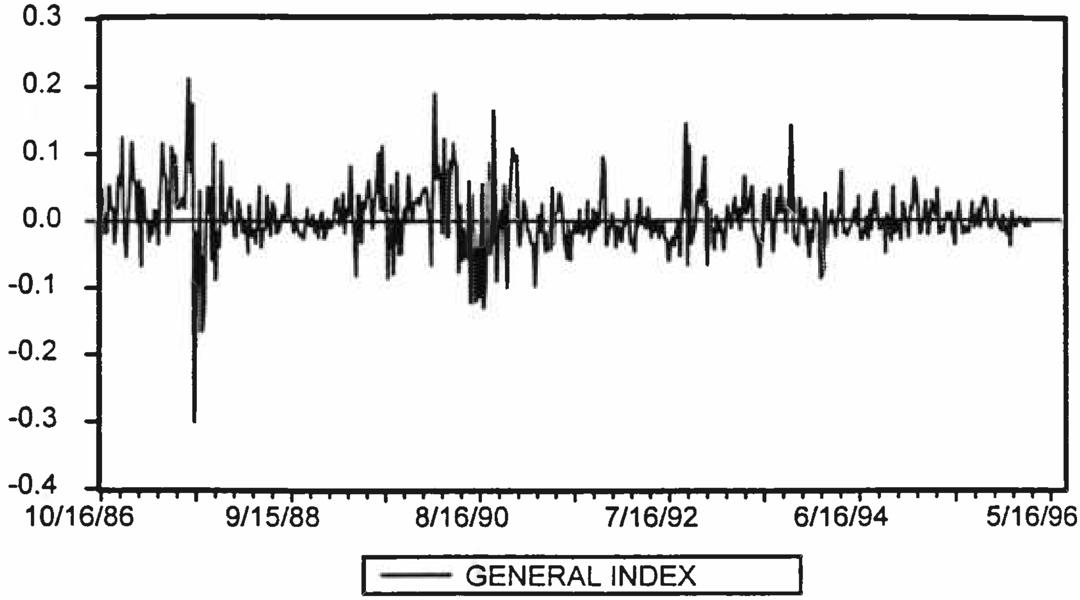
$$y_t = \ln\left(\frac{G_t}{G_{t-1}}\right)$$

where the G_t is the value of the General Index at the Thursday of a week t. We try to model y_t with the Stochastic Volatility model.

https://doi.org/10.26219/heal.aueb.5968

6.2 Data

The present study investigates the weekly rate of return of the General Index of the Athens Stock Exchange over the period 1986-1996. The data are represented in the next figure



From the above figure it is obvious that the data have volatility clustering. In consequence to that, this time series cannot be modeled by the usual homoscedastic models such ARMA or ARIMA and ARCH or SV models must be used.

https://doi.org/10.26219/heal.aueb.5968

6.3 Methodology

Because of the volatility clustering we use the Stochastic Volatility model in order to analyze the dataset. The methodology we adopt is the Markov Chain Monte Carlo (MCMC) sampling approach. Particularly, we use Metropolis-Hasting within Gibbs. In order to apply this methodology, we calculate the posterior distribution for the model and then the full conditional posterior distributions for each parameter of the model.

We use the criteria that included in the CODA software (Best & Cowles, 1995) in order to examine the convergence and especially the criterion of Raftery & Lewis (1992) in order to infer about the speed of the convergence between the different MCMC samplers that we use.

6.4 Results

6.4.1 Diagnosing the Convergence

We fit the Stochastic Volatility model to the above data using the algorithm of the MCMC . As a burn-in we take 150,000 points in order to converge to the distribution of interest and then we take a sample of 15,000 points with lag 100. We take such a big lag so that to eliminate the high autocorrelation that turns up.

Using the CODA software we find that all the criteria of convergence are satisfied. In the next paragraphs we provide the outcomes of the main criteria that we use.

Geweke test

The Geweke test (Geweke, 1992) divides the chain of each parameter in two parts which contain the first 10% and the last 50% of the iterates. If the whole chain is stationary then the means of the iterates early and late in the chain should be similar. The sample mean and the asymptotic variance -using spectral density estimation- in each part are calculated. The convergence diagnostic Z is the difference between these two means divided by the asymptotic standard error of their difference. As the chain's length $\rightarrow \infty$, the sampling distribution of $Z \rightarrow N(0,1)$, if the chain has converged (Best and Cowles, 1995) . Therefore, the results from the following table provide no evidence against convergence for each parameter of the SV model.

	Geweke
variable	**Z**
a	-1,33
d	-1,17
σ_η^2	0,738

https://doi.org/10.26219/heal.aueb.5968

Heidelberger & Welch

Applying this test (Heidelberger & Welch, 1983) we take the following outcomes

variable	Test	# to keep	# to discard	C-vonM
a	passed	15000	0	0,336
d	passed	15000	0	0,280
σ_η^2	passed	15000	0	0,169

This test based on Brownian Bridge theory and uses the Cramer-von-Mises statistic to test the null hypothesis that the sampled values for each parameter comes from a stationary process (Best and Cowles, 1995). The above table contains the outcomes of this test. The last column contains the result of the Cramer-von-Mises statistic for each parameter, the second column informs us if our chain passes this test, the third column contains the number of the iterations which can be considered that come from the distribution of interest (here all the iterations) and the fourth column gives us the number of the initial iterations that must discard. Therefore based on this test, we can conclude that our chain has converged to the posterior distribution.

Moreover, a halfwidth test is considered which examines the precision of the posterior means. The outcomes are included in the following table. Here the precision is considered $\epsilon = 0.1$, and our chain passes this test.

https://doi.org/10.26219/heal.aueb.5968

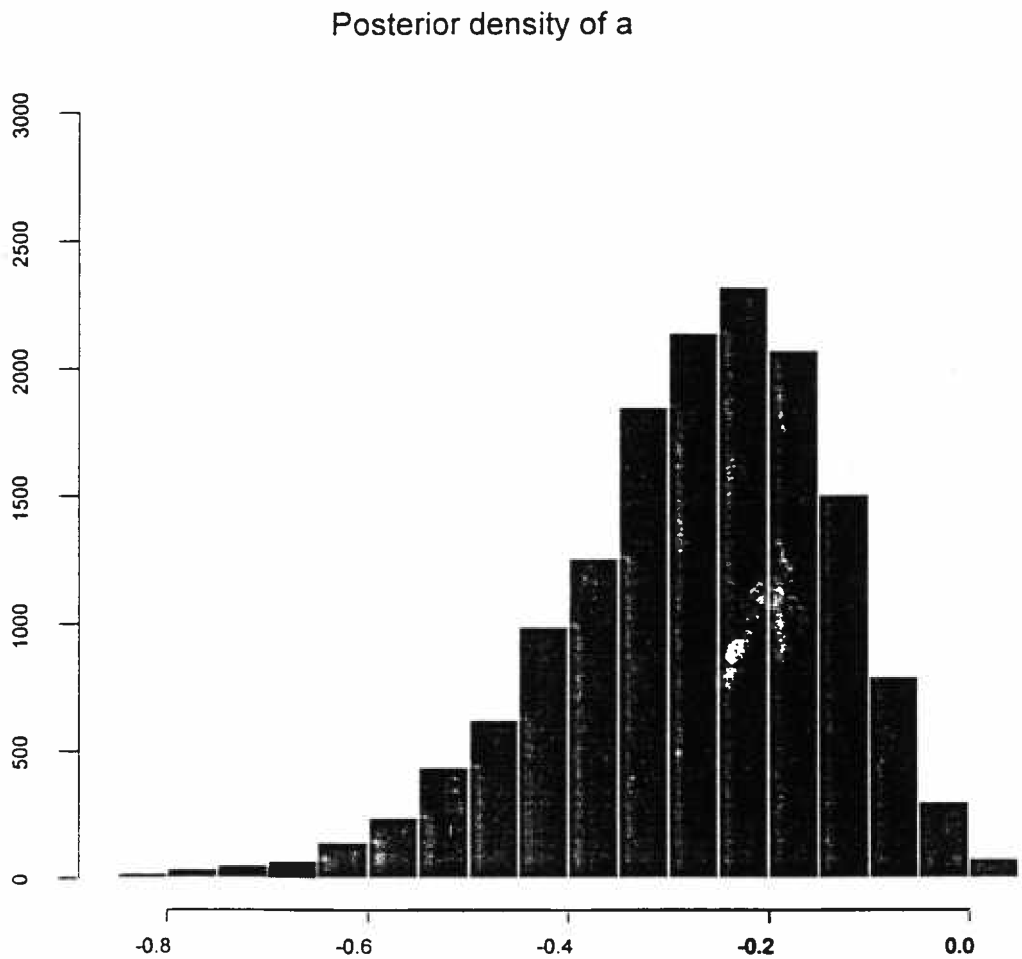
variable	Halfwidth test	Mean	Halfwidth
a	passed	-0,276	0,0024
d	passed	0,960	0,0004
σ_η^2	passed	0,110	0,0008

From the above results we can assume that the chain has converged to the equilibrium distribution so the sample that we have taken can be used for inferences about the parameters of the SV model.

The following histograms are the results of the analysis. From the histogram of d, we can see that the Greek Stock Market appears high persistent volatility and the conditional variance (σ_η^2) of the latent parameters is small. Moreover the contour graph of a and d shows the high correlation between these two parameters.

https://doi.org/10.26219/heal.aueb.5968

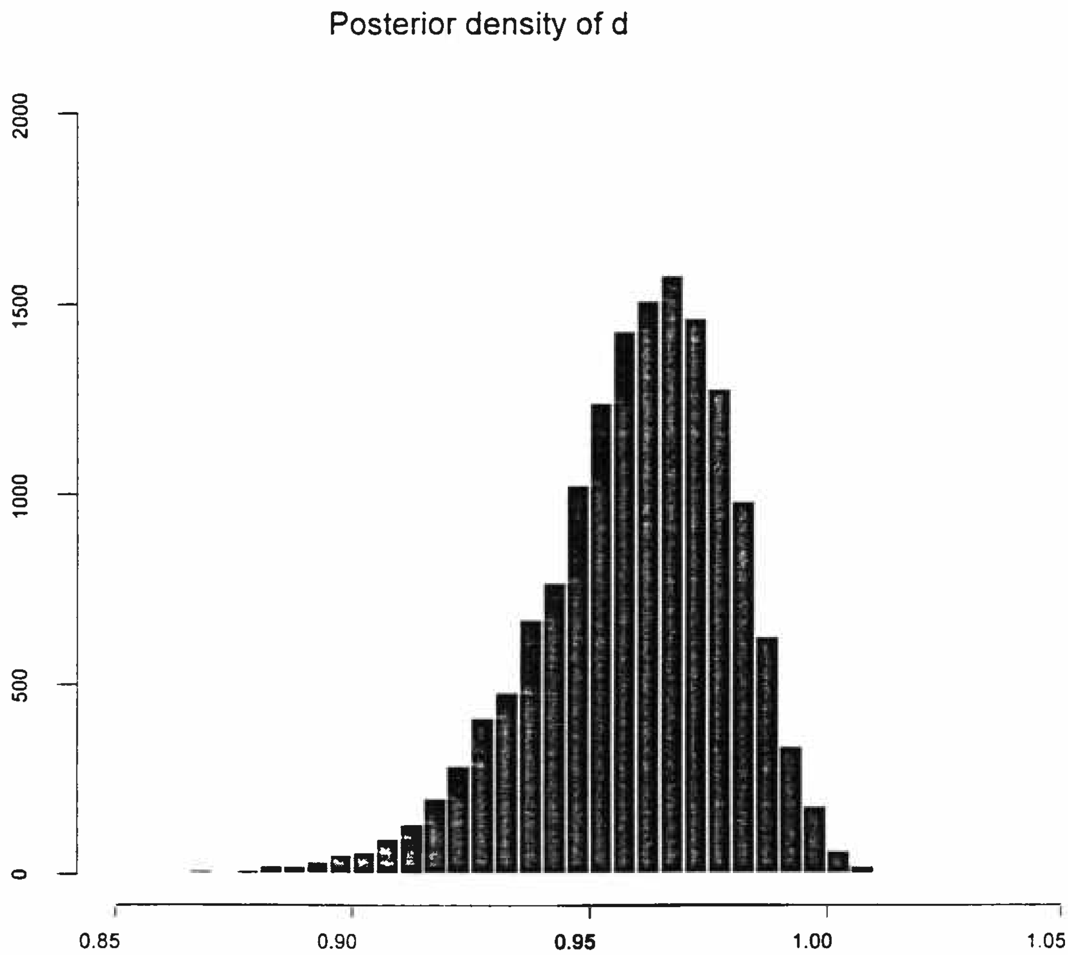
6.4.2 Histogram and summary statistics of parameter a



Mean	Median	Q_1	Q_3
-0,275681	-0,258661	-0,355484	-0,176068
Variance	**Stand. Dev.**	**Skewness**	**Kyrtosis**
0,01991	0,141104	-0,78195	1,149566

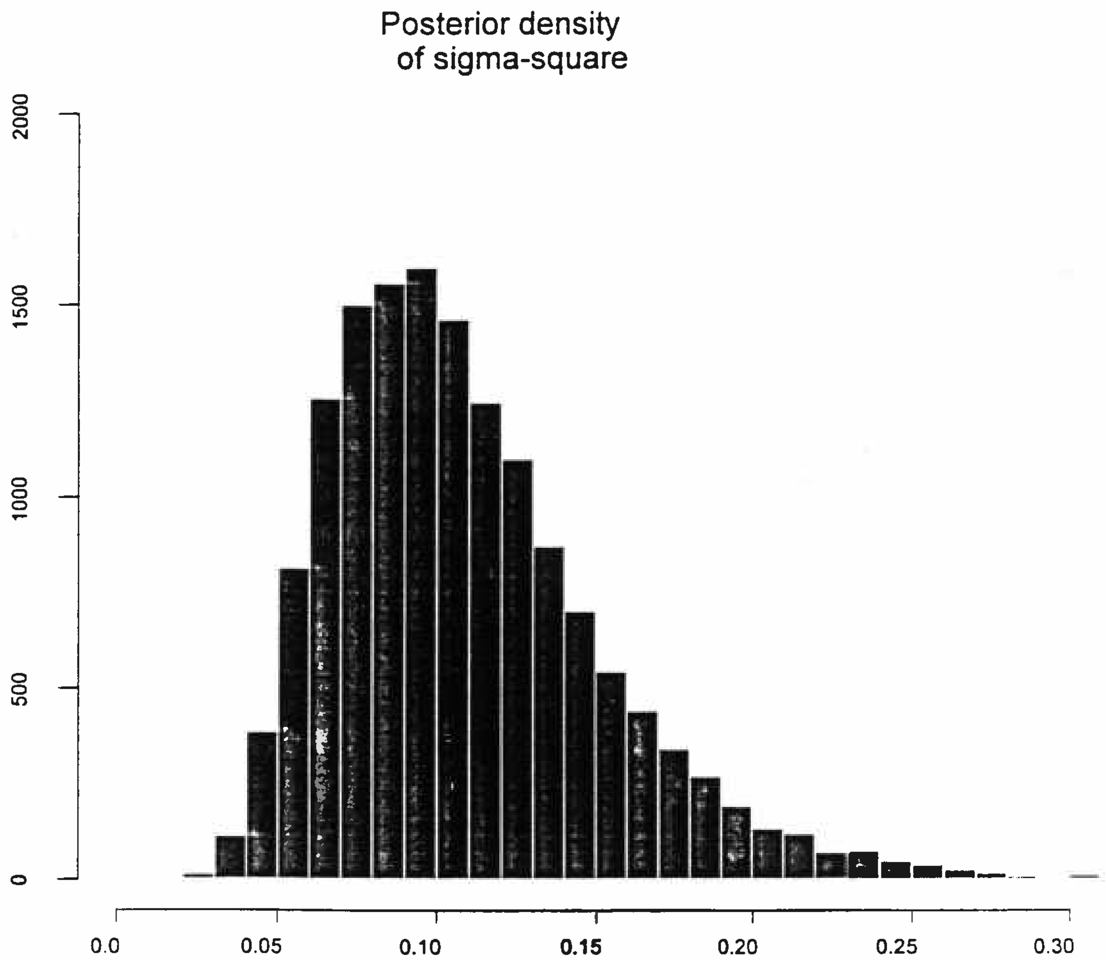
https://doi.org/10.26219/heal.aueb.5968

6.4.3 Histogram and summary statistics of parameter d



Mean	Median	Q_1	Q_3
0,959821	0,962220	0,947869	0,974557
Variance	**Stand. Dev.**	**Skewness**	**Kyrtosis**
0,000435	0,020865	-0,772202	1,10809

6.4.4 Histogram of and summary statistics parameter σ_η^2



Mean	Median	Q_1	Q_3
0,109865	0,102206	0,078397	0,132544
Variance	**Stand. Dev.**	**Skewness**	**Kyrtosis**
0,001907	0,043675	1,153098	2,180811

https://doi.org/10.26219/heal.aueb.5968

6.4.5 Posterior distribution of the mean and variance of the h_t

The mean and variance of the latent parameter h_t are produced from the following forms

$$\mu_h = E(h_t) = \frac{a}{1-d}$$

$$\sigma_h^2 = \mathcal{V}ar(h_t) = \frac{\sigma_\eta^2}{1-d^2}$$

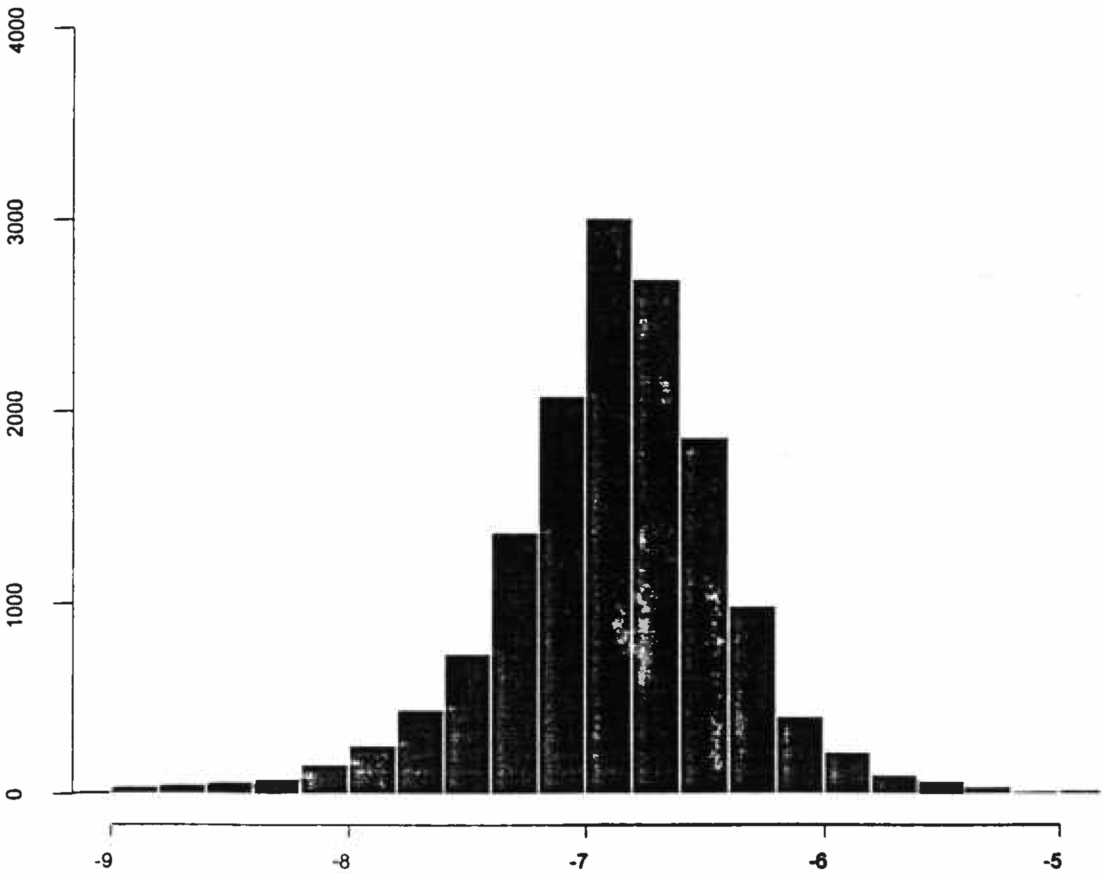
Using our sample from the MCMC algorithm we can obtain the posterior distributions of these characteristics

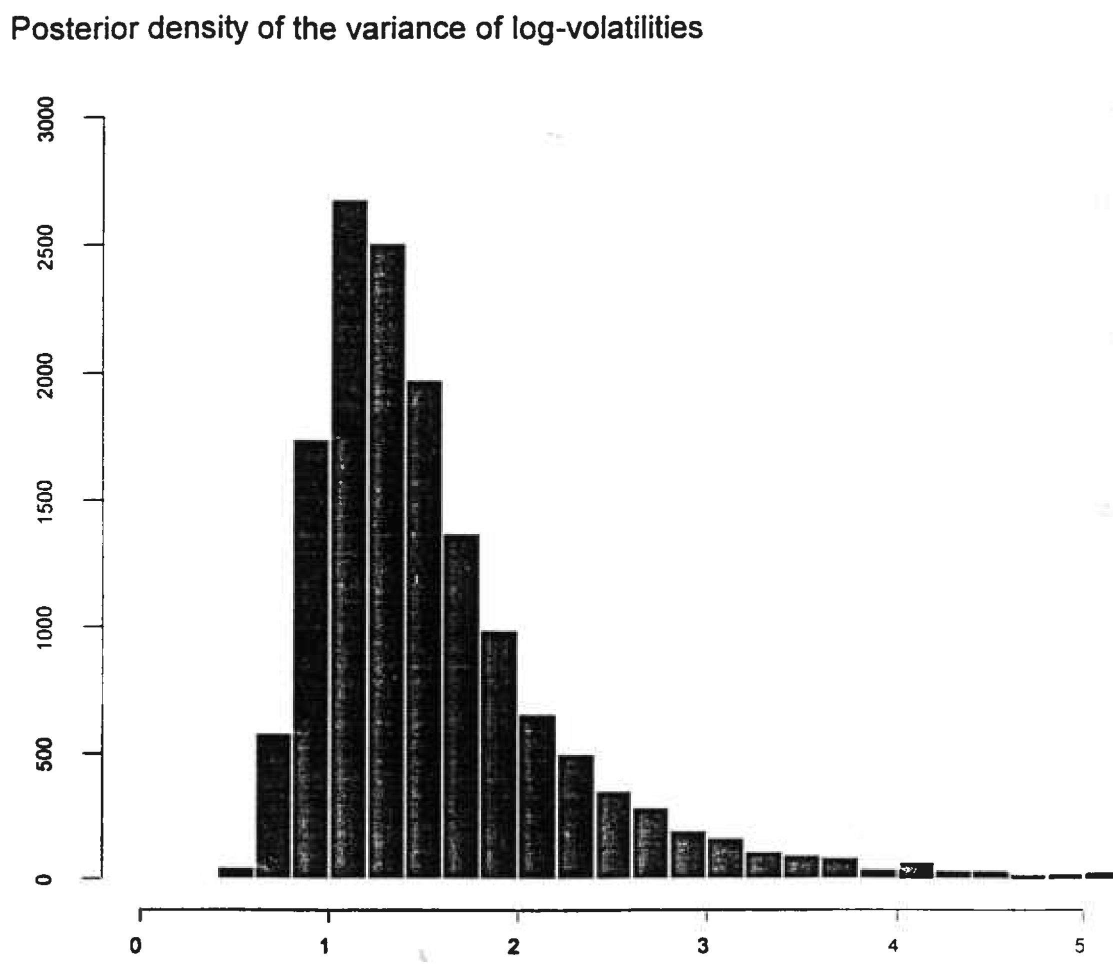
The summary statistics of these distribution are

	μ_h	σ_h^2
mean	-6,868	1,389
median	-7,161	1,097

The histograms of these distributions follow

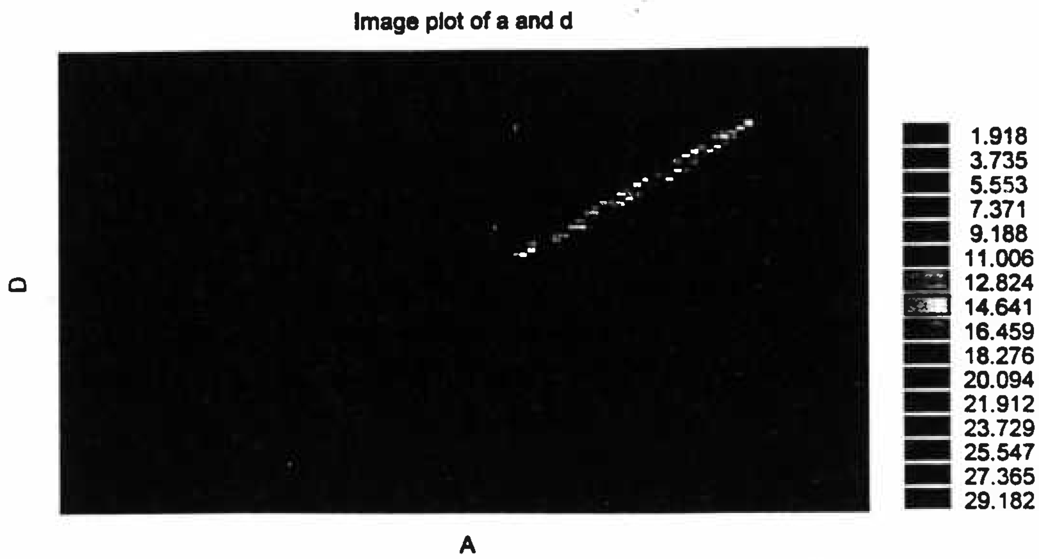
Posterior density of the mean of log-volatilities




Posterior density of the variance of log-volatilities
0
500
1000
1500
2000
2500
3000
0
1
2
3
4
5

https://doi.org/10.26219/heal.aueb.5968

6.4.6 Image plot of a and d

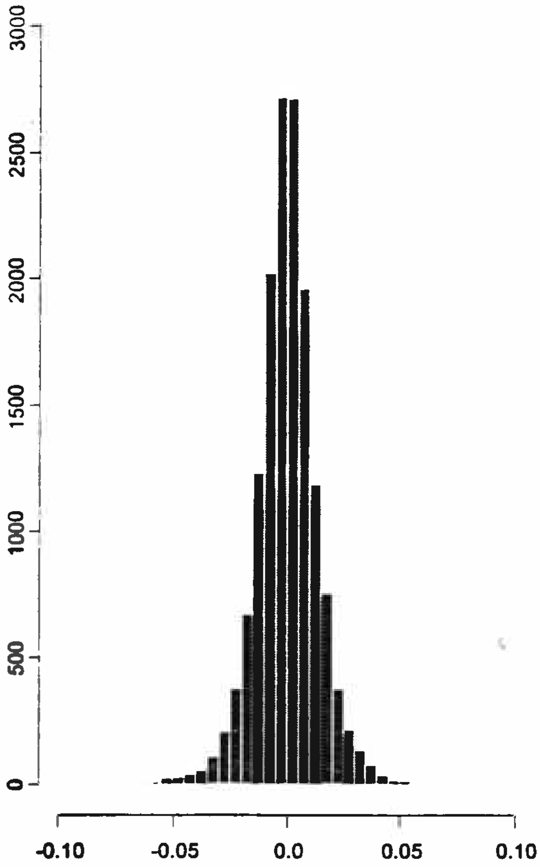


6.5 Predictions

Using the samples that we have already produced from the equilibrium distribution we can calculate the predictive density of the $y^{T+1} = y^{491}$ and the predictive distribution of the latent parameter h^{491} and the variance of this prediction $Var\left(y^{491}\right) = \exp\left(h^{491}\right)$. The histograms of the future values y^{491}, y^{492} the variance of these predictions $Var\left(y^{491}\right)$, $Var\left(y^{492}\right)$ and the latent parameters h^{491}, h^{492} follow.

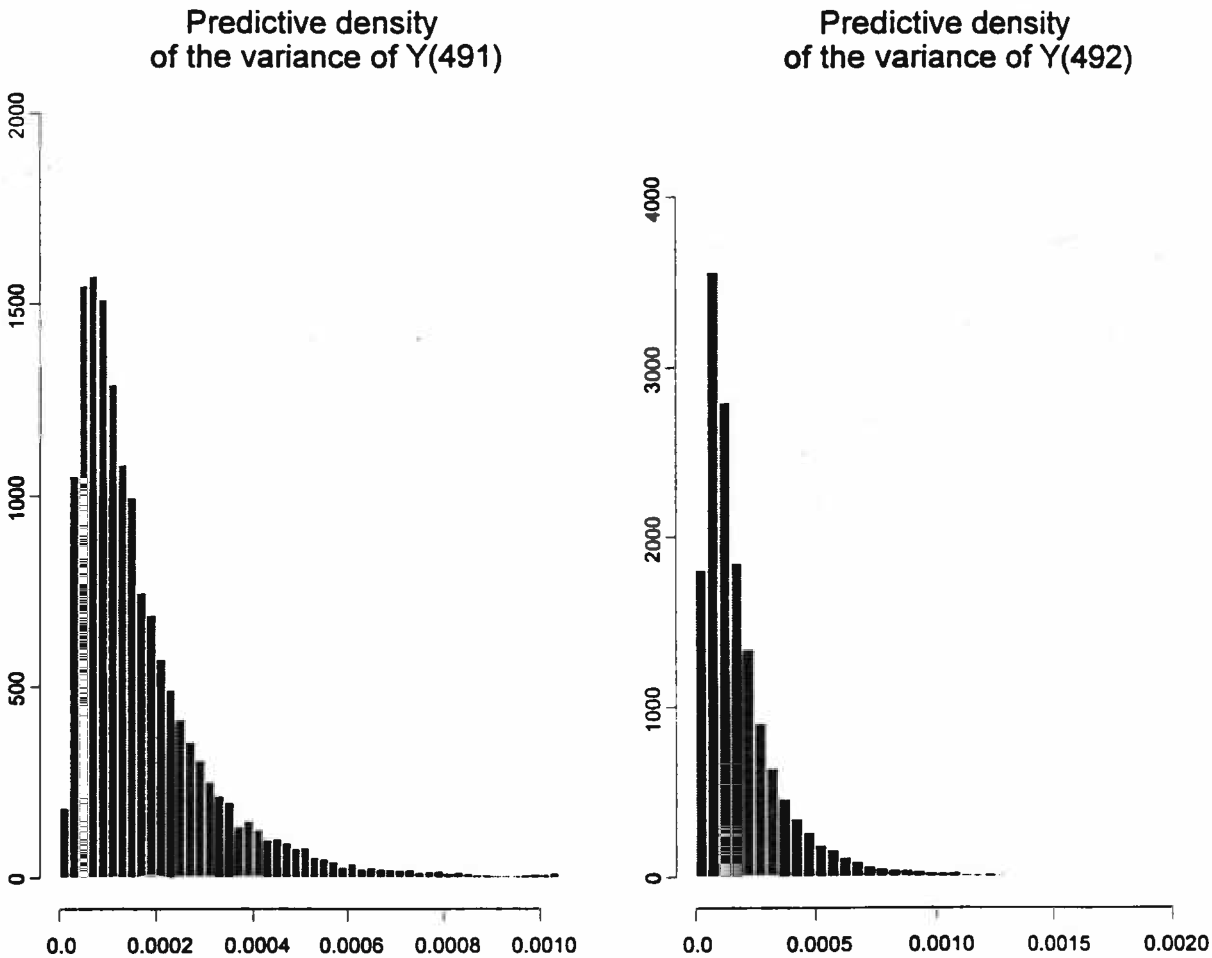
https://doi.org/10.26219/heal.aueb.5968

6.5.1 Histograms of y^{491}, y^{492}

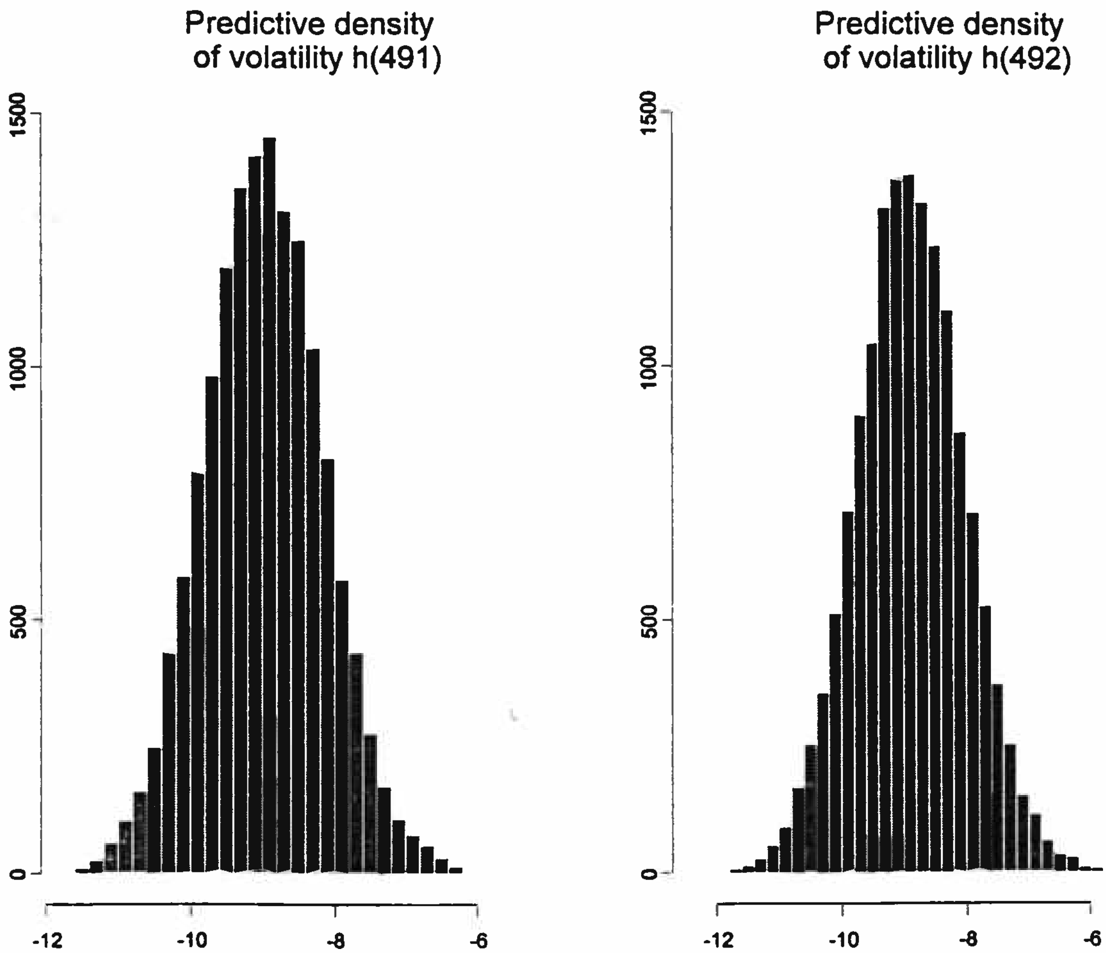


https://doi.org/10.26219/heal.aueb.5968

6.5.2 Histograms of $Var\left(y^{491}\right)$, $Var\left(y^{492}\right)$



6.5.3 Histograms of h^{491}, h^{492}



6.6 Testing results the other strategies

In order to test which of the proposed strategies is better with respect to the speed of the convergence, we run all the algorithms and we take samples of 10,000 points and we use the **Raftery & Lewis** (1992) criterion that is included in CODA software. This test examines the convergence of the chain and provides us the total number of iterations that we must run our chain in order to have sample from the posterior distribution with a specific accuracy for a specific percentile of the distribution. Here, we choose the 2.5*th* percentile and the desired accuracy is chosen ±0.005. The results of this test follow.

The sequential MCMC

variable	Thin (k)	Burn-in	Total	Lower Bound	factor (I)
a	8	312	**206936**	3746	55,2
d	4	140	**139056**	3746	37,1
σ^2_η	7	91	**114247**	3746	30,5

so using this algorithm we must take 206,936 points in order to make inference about the parameters

Random scan MCMC

variable	Thin (k)	Burn-in	Total	Lower Bound	factor (I)
a	10	150	**136000**	3746	36,3
d	8	152	**132456**	3746	35,4
σ^2_η	7	175	**182616**	3746	48,7

so with this algorithm we must take 182,616 points

Non-sequential MCMC

variable	Thin (k)	Burn-in	Total	Lower Bound	factor (I)
a	8	312	**206936**	3746	55,2
d	4	140	**139056**	3746	37,1
σ_η^2	7	91	**114247**	3746	30,5

so using this algorithm we must take 206,936 points in order to make inference about the parameters.

MCMC with transformation on the latent parameters

variable	Thin (k)	Burn-in	Total	Lower Bound	factor (I)
a	10	140	**123860**	3746	33,1
d	10	110	**115930**	3746	30,9
σ_η^2	5	50	**56895**	3746	15,2

so this algorithm forces us a run the MCMC 124,000 times, therefore seems better than the other three. In addition to that, this algorithm reduces the autocorrelation and therefore we propose it as the more appropriate MCMC strategy for the Stochastic Volatility model.

https://doi.org/10.26219/heal.aueb.5968

Chapter 7

Future Research

The class of Stochastic Volatility models is a very promising area for research if we have in mind that it is only few years since Taylor (1986) introduced this class and only recently Jacquier et al (1994) investigated this model via the Bayesian framework.

Many extension of the stochastic model can be studied in univariate case. Moreover the **Multivariate Stochastic Volatility** models (Jacquier et al, 1995) is a totally new and hopeful area of research.

A very natural extensions of the SV could be the above

$$y_t|_{h_t} \sim \mathcal{N}\left(0, \exp(h_t)\right) \tag{7.1}$$

$$h_t = a + d \cdot h_{t-1} + \eta_t - \beta \cdot \eta_{t-1}$$

$$\eta_t \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_\eta^2\right)$$

and

https://doi.org/10.26219/heal.aueb.5968

$$y_t|_{h_t} \sim \mathcal{N}\left(0, \exp(h_t)\right)$$

$$h_t = a + d \cdot h_{t-1} + g \cdot \ln\left(y_{t-1}^2\right) + \eta_t \tag{7.2}$$

$$\eta_t \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_\eta^2\right).$$

The first model is a SV model but in this form the variance equation has an ARMA(1,1) pattern than the usual AR(1). The second model is like the GARCH(1,1) and allows the previous value of the time series to affect the conditional variance.

Another topic for research, is the model selection. In this case, an algorithm can be constructed - following, probably the Reversible Jump (Green 1995) - in order to select a time series model from a set of models, where this set can be contained ARCH, GARCH and SV model and the above extension of SV.

https://doi.org/10.26219/heal.aueb.5968

Bibliography

[1] **Bayes, T. (1763)**, "An essay towards solving a problem in doctrine of chances", **Phil. Trans. Roy. Soc.**, 53, 370-418.

[2] **Best, N. and M. K. Cowles (1995)**, "CODA: Convergence Diagnosis and Output Analysis Software for Gibbs sampling output, Version 0.30", MRC Biostatistics Unit, Institute of Public Health, Cambridge.

[3] **Bernardo, J. M. and A. F. Smith (1995)**, "Bayesian Theory", John Wiley and Sons.

[4] **Bollerslev, T. (1986)**, "Generalized Autoregressive Conditional Heteroskedasticity", **Journal of Econometrics**, 31, 307-327.

[5] **Bollerslev, T., R. Y. Chou and K.F. Kroner (1992)**, "ARCH modeling in finance", **Journal of Econometrics**, 52, 5-59.

[6] **Brooks, S. P. and G. O. Roberts (1995)**, "Diagnosing Convergence of Markov Chain Monte Carlo Algorithms", **Statistics and Computing**, in Publication..

[7] **Brooks, S. P., P. Dellaportas and G. O. Roberts (1996)**, "An Approach to Diagnosing Total Variation Convergence of MCMC algo-

rithm", **Journal of Computational and Graphical Statistics**, in Publication..

[8] **Casella, C. and E. I. George (1992)**, "Explain the Gibbs Sampler", **The American Statistician**, 46, 167-174.

[9] **Chesney, M. and L. O. Scott (1989)**, "Pricing European options: a comparison of the modified Black-Scholes model and a random variance model", **Journal of Financial and Qualitative Analysis**, 24, 267-284.

[10] **Chib, S. and E. Greenberg (1993)**, "Markov Chain Monte Carlo Simulation Methods in Econometrics", **Econometric Theory**, in press.

[11] **Chib, S. and E. Greenberg (1995)**, "Understanding the Metropolis-Hastings Algorithm", **Journal of the American Statistical Association**, 19, 327-335.

[12] **Dellaportas, P. (1993)**, "Simulation and Stochastic Models", Lecture notes, Athens University of Economics, Department of Statistics.

[13] **Dellaportas, P. (1995)**, "Random variate Transformations in the Gibbs Sampler: Issues of efficiency and Convergence", **Statistics and computing**, 5, 133-140.

[14] **Engle, R. F. (1982)**, "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation", **Econometrica**, 50, 987-1008.

https://doi.org/10.26219/heal.aueb.5968

[15] **Gelfand, A. E. and A. F. M. Smith (1990)**, "Sampling-based approaches to calculating Marginal Densities", **Journal of the American Statistical Association**, 85, 398-409

[16] **Gelfand, A. E., S. E. Hills, A. Racine-Poon and A. F. M. Smith (1990)**, "Models Using Gibbs Sampling", **Journal of American Statistical Association**, 85, 972-985.

[17] **Gelman, A. and D. B. Rubin (1992)**, "inference from iterative Simulation Using Multiple Sequences", **Statistical science**, 7, 457-511.

[18] **Geman, S. and D. Geman (1984)**, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images", **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 6, 721-741.

[19] **Geweke, J. (1989)**, "Bayesian inference in econometric model using Monte Carlo integration", **Econometrica**, 57, 1317-1340.

[20] **Geweke, J. (1992)**, "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments", In Bayesian Statistics 4 (eds J. M. Bernardo et al), Oxford University Press, 169-193.

[21] **Gilks W. R., S. Richarson and D. J. Spiegelhalter (1996)**, "-Markov Chain Monte Carlo in Practice", Chapman and Hall.

[22] **Green, P. J. (1995)**, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination", **Biometrika**, 82, 711-732.

[23] **Hammersley, J. M. and D. C. Handscomb (1964)**, "Monte Carlo Methods", Chapman and Hall, UK.

[24] **Hamilton, J. D. (1994)**, "Time Series Analysis", Princenton University Press.

[25] **Harvey, A. C., E. Ruiz and N. Shephard (1994)**, "Multivariate stochastic variance models", **Review of Economic Studies**, 61,247-264.

[26] **Hastings, W. K. (1970)**, "Monte Carlo Sampling Methods Using Markov Chains and Their Applications", **Biometrika**, 57, 97-109.

[27] **Heidelberger, P. and P. D. Welch (1983)**, "Simulation Run Length Control in the Presence of an Initial Transient", **Operations Research**, 31, 1109-1144.

[28] **Magdalinos, M. (1992)**, "Introduction to Bayesian Statistics", Lecture Notes, Athens University of Economics, Department of Statistics.

[29] **Marriott, J., N. Ravishanker, A. Gelfand and J. Pai (1994)**, "Bayesian Analysis of ARMA processes: Complete Sampling based Inference under Full Likelihoods", Bayesian Statistics and Econometrics: Essays in honor of Arnold Zellner. Eds D. Berry, K. Chaloner and J. Geweke. To appear.

[30] **Metropolis, N., Rosenbluth, A. W., Reller, A. H. and E. Teller (1953)**, "Equations of State Calculations by Fast Computing Machines", **Journal of Chemical Physics**, 21, 1087-1092.

[31] **Lee Tai-Ming (1992)**, "Comparison of sampling based approaches for Bayesian computations", Ph. D. Dissertation, University of Connecticut.

[32] **Nelson, D. B. (1991)**, "Conditional Heteroskedasticity in Asset Returns: A New Approach", **Econometrica**, 59, 347-370.

[33] **Jacquier, E:, N. G. Polson and E. Rossi (1994)**, "Bayesian Analysis of Stochastic Volatility Models", **Journal of Business & Economic Statistics**, 12, 371-417.

[34] **Jacquier, E:, N. G. Polson and E. Rossi (1995)**, "Models and Priors for Multivariate Stochastic Volatility", Working Paper, University of Chicago.

[35] **Phillips, D. B. and A. F. M. Smith (1993)**, "Orthogonal random-direction sampling in Markov Chain Monte Carlo", personal communication.

[36] **Raftery, A. and S. Lewis (1992)**, "How Many Iterations in the Gibbs Sampler?", Bayesian statistics 4, Oxford University Press.

[37] **Robert, C. P. (1994)**, "The Bayesian Choice: a decision-theoretic motivation", Springer-Verlag New York.

[38] **Roberts, G. O. and S. K. Sahu (1996)**, "Updating Schemes, Correlation Structure, Blocking and parameterisation for the Gibbs Sampler", in Publication..

[39] **Shephard, N. (1996)**, "Time Series Models in econometrics, finance and other fields", Chapman and Hall, 1-55.

[40] **Taylor, S. J. (1986)**, "Modelling Financial Time Series", John Wiley, Chrichester.

[41] **Tierney, L. (1994)**, "Markov Chains for Exploring Posterior Distributions" (with discussion), **Annals of Statistics**, 22, 1701-1762.

[42] **Vetzal, K. (1992)**, "Stochastic Short Rate Volatility and the Pricing of Bonds and Bond options", Ph. D., University of Toronto.

https://doi.org/10.26219/heal.aueb.5968