

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Volatility Time Series Models</b>	<b>3</b>
2.1	Introduction . . . . .	3
2.2	Volatility Clustering . . . . .	3
2.3	Classifying Models of Changing Volatility . . . . .	4
2.4	Univariate Models . . . . .	6
2.4.1	ARCH-type Models . . . . .	6
2.4.2	Stochastic Volatility Models . . . . .	9
2.5	Multivariate Models . . . . .	12
2.5.1	ARCH/GARCH Unconstrained Models . . . . .	12
2.5.2	ARCH/GARCH Constrained Models . . . . .	13
2.5.3	ARCH/GARCH Models With Constant Conditional Correlation .	14
2.5.4	Factor ARCH/GARCH Models . . . . .	15
2.5.5	The Multivariate Stochastic Volatility Model . . . . .	15
2.5.6	The Multivariate Factor Stochastic Volatility Model . . . . .	16
<b>3</b>	<b>The Bayes Framework and MCMC</b>	<b>17</b>
3.1	Bayes's Theorem . . . . .	17
3.2	The Computational Difficulty of Bayesian Analysis . . . . .	18

3.3	Sampling Based Methods . . . . .	20
3.3.1	The Gibbs Sampler . . . . .	22
3.3.2	The Metropolis-Hastings Algorithm . . . . .	24
3.3.3	Example . . . . .	27
3.3.4	The Auxiliary Variable Sampler . . . . .	28
<b>4</b>	<b>A New Convergence Diagnostic Based on Subsampling</b>	<b>31</b>
4.1	Introduction . . . . .	31
4.2	CODA tests . . . . .	33
4.3	An MCMC Convergence Diagnostic using Subsampling . . . . .	34
4.3.1	Introduction . . . . .	34
4.3.2	The Subsampling Methodology . . . . .	35
4.3.3	The MCMC Subsampling Diagnostic . . . . .	38
4.3.4	An Alternative ‘Hybrid’ Diagnostic . . . . .	41
4.3.5	Implementation . . . . .	43
4.3.6	Examples and simulations . . . . .	46
<b>5</b>	<b>The Auxiliary Variable Sampler</b>	<b>55</b>
5.1	Introduction . . . . .	55
5.2	The Auxiliary Variable Sampler . . . . .	56
5.3	Variations of the Auxiliary Variable Sampler . . . . .	57
5.3.1	Introduction . . . . .	57
5.3.2	The Simple Auxiliary Variable Sampler . . . . .	58
5.3.3	The Multiple Auxiliary Variable Sampler . . . . .	58
5.3.4	The Product Auxiliary Variable Sampler . . . . .	59
5.4	Properties of the Auxiliary Variable sampler . . . . .	60
5.5	Comparison of the Auxiliary Variable Sampler with the Metropolis-Hastings scheme . . . . .	64

5.6	Auxiliary Variable Sampler in Bayesian Modeling . . . . .	65
5.7	Main Drawbacks of Auxiliary Variable Sampler . . . . .	67
5.8	The Metropolis scheme within the Auxiliary Variable Sampler . . . . .	69
5.9	Neal's Approach . . . . .	70
5.9.1	The Stepping-out Procedure . . . . .	71
5.9.2	The Doubling Procedure . . . . .	71
5.9.3	Discussion . . . . .	72
5.10	The Multivariate Auxiliary Variable Sampler . . . . .	74
5.11	Other Aspects of the Auxiliary Variable Sampler . . . . .	76
5.11.1	Partial Decoupling Method . . . . .	76
5.11.2	The Overrelaxed Auxiliary Variable Sampler . . . . .	76
5.11.3	The Adaptive Multivariate Auxiliary Variable Sampler . . . . .	77
<b>6</b>	<b>A Bayesian Approach to Univariate Stochastic Volatility Models</b>	<b>79</b>
6.1	Introduction . . . . .	79
6.2	Existing MCMC Algorithms for the Stochastic Volatility Model . . . . .	80
6.2.1	Singe-Move Algorithms . . . . .	80
6.2.2	Block-Move Algorithms . . . . .	86
<b>7</b>	<b>Auxiliary Variable Sampler and Univariate Time-Varying Volatility Models</b>	<b>91</b>
7.1	Introduction . . . . .	91
7.2	Stochastic volatility model . . . . .	92
7.2.1	Introduction . . . . .	92
7.2.2	Bayesian Approach and the Auxiliary Variable Sampler . . . . .	93
7.2.3	Some Applications . . . . .	97
7.3	Unobserved ARCH model . . . . .	107
7.3.1	Introduction . . . . .	107
7.3.2	Bayesian Approach and the Auxiliary Variable Sampler . . . . .	109

7.3.3	Usage of the Auxiliary Variable Sampling . . . . .	112
7.3.4	Some Applications . . . . .	113
7.4	ARCH model . . . . .	123
7.4.1	Introduction . . . . .	123
7.4.2	Bayesian Approach and the Auxiliary Variable Sampler . . . . .	123
7.4.3	Some Applications . . . . .	126
7.5	GARCH models . . . . .	133
7.5.1	Introduction . . . . .	133
7.5.2	Bayesian Approach and the Auxiliary Variable Sampler . . . . .	133
7.5.3	Some Applications . . . . .	135
<b>8</b>	<b>Comparison of the Univariate Models</b>	<b>145</b>
8.1	Introduction . . . . .	145
8.2	The Comparison Method . . . . .	145
8.3	The Results . . . . .	147
<b>9</b>	<b>Existing Multivariate Parameter-Driven Models</b>	<b>149</b>
9.1	Introduction . . . . .	149
9.2	Multivariate Stochastic Volatility Models . . . . .	150
9.3	Factor Stochastic Volatility Models . . . . .	151
9.4	Latent Factor ARCH Models . . . . .	153
<b>10</b>	<b>Auxiliary Variable Sampler and Multivariate Time-Varying Volatility Models</b>	<b>155</b>
10.1	Multivariate Stochastic Volatility Model . . . . .	155
10.1.1	Introduction . . . . .	155
10.1.2	Bayesian Approach and the Auxiliary Variable Sampler . . . . .	156
10.1.3	Application . . . . .	162
10.2	Multivariate Unobserved ARCH Model . . . . .	170

10.2.1	Introduction . . . . .	170
10.2.2	Bayesian Approach and the Auxiliary Variable Sampler . . . . .	171
10.2.3	Application . . . . .	177
10.3	Latent Factor ARCH Model . . . . .	184
10.3.1	Introduction . . . . .	184
10.3.2	Bayesian Approach and the Auxiliary Variable Sampler . . . . .	185
10.3.3	Application . . . . .	188
10.4	Latent Factor GARCH Model . . . . .	192
10.4.1	Introduction . . . . .	192
10.4.2	Bayesian Approach and the Auxiliary Variable Sampler . . . . .	193
10.4.3	Application . . . . .	195
10.5	Comparison . . . . .	199
10.6	Technical Details . . . . .	201
10.6.1	Sampling from truncated Inverse Gamma . . . . .	201
10.6.2	Construction of (1-a)100% credible interval for the mean using sub- sampling . . . . .	201
<b>11</b>	<b>Discussion</b>	<b>203</b>

# List of Figures

2-1	Exchange rates of USD and DEM versus GRD . . . . .	4
4-1	Coefficient of determination. Solid line: threshold $d=0.999$ , dashed line: subsampling method, dotted line: alternative method. (a) Example 1, (b) Example 2, first chain, (c) Example2, second chain, (d) Example 3, problem 1, (e) Example 3, problem 2. . . . .	49
7-1	Exchange Rates Series . . . . .	98
7-2	Autocorrelation function plots for the parameters of the SV model . . . .	99
7-3	MCMC output for parameters of SV for USD . . . . .	101
7-4	Posterior Histograms for the parameters of SV for USD . . . . .	102
7-5	Subsampling diagnostic plot for USD . . . . .	103
7-6	MCMC output for parameters of SV for JPY . . . . .	104
7-7	Posterior Histograms for the parameters of sv for JPY . . . . .	105
7-8	Subsampling diagnostic plot for JPY . . . . .	106
7-9	Autocorrelation function plots for the parameters of the unobserved ARCH model . . . . .	115
7-10	MCMC output for parameters of unobserved ARCH for USD . . . . .	117
7-11	Posterior Histograms for the parameters of unobserved ARCH for USD .	118
7-12	Subsampling diagnostic plot for USD . . . . .	119
7-13	MCMC output for parameters of unobserved ARCH for JPY . . . . .	120
7-14	MCMC output for parameters of unobserved ARCH for JPY . . . . .	121

7-15	Subsampling diagnostic plot for JPY . . . . .	122
7-16	Autocorrelation function plots for the parameters of the ARCH model . .	127
7-17	MCMC output for parameters of ARCH(1) for USD . . . . .	128
7-18	Posterior Histograms for the parameters of ARCH(1) for USD . . . . .	129
7-19	Subsampling diagnostic plot for the USD . . . . .	130
7-20	MCMC output for parameters of ARCH(1) for JPY . . . . .	131
7-21	Posterior Histograms for the parameters of ARCH(1) for JPY . . . . .	131
7-22	Subsampling diagnostic plot for the JPY . . . . .	132
7-23	Autocorrelation function plots for the parameters of the GARCH model .	137
7-24	MCMC output for the parameters of the GARCH(1,1) for USD . . . . .	139
7-25	Posterior Histograms for the parameters of the GARCH(1,1) for USD . .	140
7-26	Subsampling diagnostic plot for the GARCH(1,1) for USD . . . . .	141
7-27	MCMC output for the parameters of the GARCH(1,1) for JPY . . . . .	142
7-28	Posterior Histograms for the parameters of the GARCH(1,1) for JPY . .	143
7-29	Subsampling diagnostic plot for the GARCH(1,1) for JPY . . . . .	144
10-1	Exchange Rates Series . . . . .	163
10-2	Subsampling diagnostic plot for Multivariate SV model . . . . .	165
10-3	Autocorrelation function plots for the hyperparameters of the multivariate SV model . . . . .	166
10-4	MCMC output for parameters of multivariate SV model . . . . .	167
10-5	Posterior Histograms for the hyperparameters of the multivariate SV model	168
10-6	Posterior Histograms for the variance covariance matrix $\Lambda$ of the multi- variate SV model . . . . .	169
10-7	MCMC output for parameters of multivariate unobserved ARCH model .	178
10-8	Subsampling diagnostic plot for multivariate unobserved ARCH model .	179
10-9	Autocorrelation function plots for the hyperparameters of the multivariate unobserved ARCH model . . . . .	180

10-10	Posterior Histograms for the hyperparameters of the multivariate unobserved ARCH model . . . . .	182
10-11	Posterior Histograms for the variance covariance matrix $\Sigma$ of the multivariate unobserved ARCH model . . . . .	183
10-12	Subsampling diagnostic plot for latent factor ARCH model . . . . .	190
10-13	MCMC output for parameters of latent factor ARCH model . . . . .	191
10-14	Subsampling diagnostic plot for latent factor GARCH model . . . . .	197
10-15	MCMC output for parameters of latent factor GARCH model . . . . .	198

# List of Tables

4.1	<i>Accuracy of the posterior mean of the model parameters . . . . .</i>	47
4.2	<i>Posterior model probabilities and their accuracies . . . . .</i>	52
4.3	<i>Number of iterations for specific accuracy . . . . .</i>	53
7.1	Autocorrelation function results for the parameters of the SV model . . .	98
7.2	Posterior Statistics for the papameters of the SV model for USD . . . . .	100
7.3	Posterior Statistics for the papameters of the SV model for JPY . . . . .	103
7.4	Autocorrelation function results for the parameters of the unobserved ARCH model . . . . .	114
7.5	Posterior Statistics for the papameters of the unobserved ARCH model for USD . . . . .	116
7.6	Posterior Statistics for the papameters of the unobserved ARCH model for JPY . . . . .	119
7.7	Autocorrelation function results for the parameters of the ARCH model .	126
7.8	Posterior Statistics for the papameters of the ARCH(1) model for USD .	128
7.9	Posterior Statistics for the papameters of the ARCH(1) model for JPY .	130
7.10	Autocorrelation function results for the parameters of the GARCH model	136
7.11	Posterior Statistics for the papameters of the GARCH(1) model for USD	138
7.12	Posterior Statistics for the papameters of the GARCH(1) model for JPY	141
8.1	Model comparison . . . . .	148
10.1	Summary statistics for the hyperparameters of the multivariate SV model	164

10.2	Summary statistics for the hyperparameters of the Unobserved ARCH model	181
10.3	Estimates of the covariance matrix. Diagonal elements: variances. Upper diagonal elements: covariances. Lower diagonal elements: correlations . .	181
10.4	Posterior means and 95% credible intervals for the parameters of the latent factor ARCH model . . . . .	189
10.5	Posterior means and 95% credible intervals for the parameters of the latent factor GARCH model. . . . .	196
10.6	Model comparison . . . . .	200

# Chapter 1

## Introduction

Time-varying volatility models have drawn lately a lot of attention due to their ability to capture the behavior of financial data such as exchange rates and stock prices. These models are capable of describing the phenomenon of volatility clustering that is present in financial data. The estimation of time-varying volatilities and covariances is crucial for areas such as asset pricing, risk management and portfolio analysis.

In this Thesis the univariate and multivariate models are analyzed. Special effort is given to models that are incorporated to one of the two main categories of the time-varying volatility models, the stochastic volatility models (parameter-driven models). Furthermore, new Markov Chain Monte Carlo algorithms are proposed for certain univariate and multivariate models that are easy to apply and fast to converge.

This Thesis is organized in eleven chapters. The first chapter contains the introduction of this Thesis. In the second chapter a more inclusive presentation of the time-varying volatility models can be found. There, the phenomenon of volatility clustering along with the models that have been developed to explain this phenomenon is presented. In the third chapter the Bayesian framework and the Markov Chain Monte Carlo (MCMC) methods are presented. In chapter four the problem of the convergence of the MCMC algorithms is presented and a new diagnostic that can be used for the detection of the convergence of the MCMC algorithms is proposed. This convergence diagnostic is based

on the subsampling methodology and on the construction of  $(1 - a)$  100% confidence regions for the mean and for the  $t$ -percentile. In the fifth chapter the Auxiliary Variable sampler is described and all the versions of this algorithm and its theoretical properties are presented.

The sixth chapter presents the existing MCMC algorithms for the case of stochastic volatility models and in the seventh chapter new MCMC algorithms for the cases of stochastic volatility model, the unobserved ARCH model, the ARCH model and the GARCH model are proposed. These new MCMC algorithms are using the Auxiliary Variable sampling methodology and are consisted by only Gibbs steps. Furthermore, real data are used for the application of these algorithms and in the eighth chapter the aforementioned models are compared.

In the ninth chapter a summary of the existing algorithms for the multivariate stochastic volatility is given. Additionally, in the tenth chapter new MCMC algorithms are proposed for a version of the multivariate stochastic volatility model, the multivariate unobserved ARCH, the latent factor ARCH model and the latent factor GARCH model. Furthermore, these models are compared with criteria based on their predictive distribution.

Finally, chapter eleven contains a summary of this thesis and some points for future research.

# Chapter 2

## Volatility Time Series Models

### 2.1 Introduction

Up to 1980 the financial time series modeling have been centered on the conditional first moments, with any temporal dependencies in the higher order moments treated as a nuisance. The last two decades, however, researchers turn their interest in developing new techniques that allow the modeling of time-varying variances and covariances. This shifting of interest was the result of the increasing importance that the risk plays in modern economics. The phenomenon of volatility clustering that is apparent in most financial time series, such as asset returns and exchange rates can be captured by modeling the conditional second moments.

### 2.2 Volatility Clustering

The phenomenon of volatility clustering is immediately turned up when we plot an asset returns series or an exchange rates series through time. As Mandelbrot (1963) wrote,

*...large changes tend to be followed by large changes, of either sign, and small changes tends to be followed by small changes....*

This statement is confirmed in Figure 2-1, where the compounded return of the US

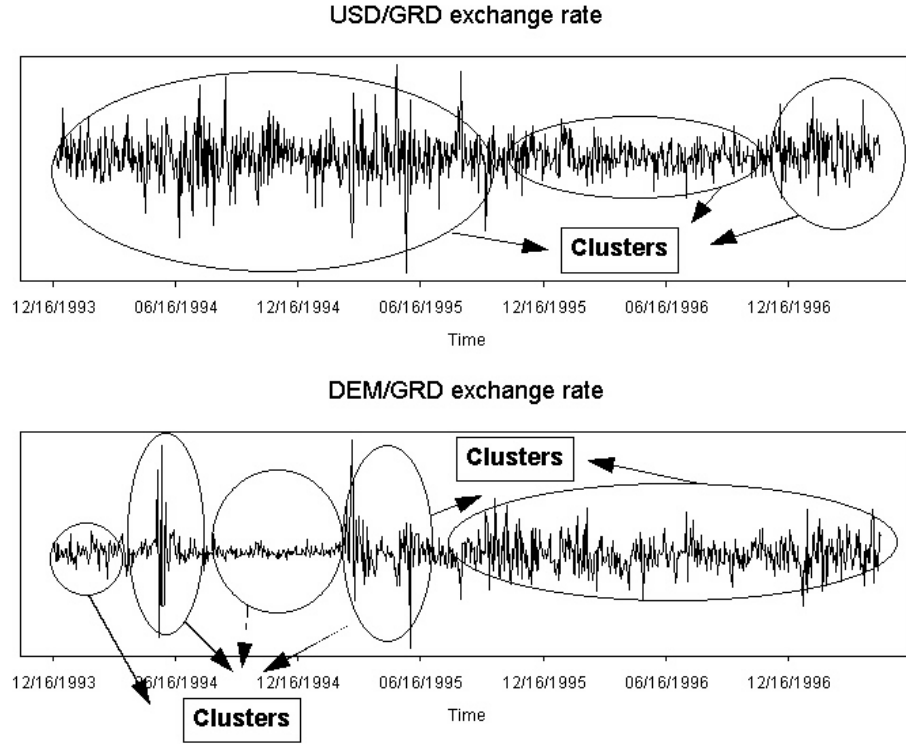


Figure 2-1: Exchange rates of USD and DEM versus GRD

dollar (USD) and the Germany Marc (DEM) with respect to the Greek Drachma (GRD) is plotted through time (where the compounded return on the series  $y_t$  is defined as  $y_t = 1000 \log(x_t/x_{t-1})$ ). In this plot it is obvious that the series appear major fluctuation with time periods of large disturbances followed by periods with small disturbances and vice-versa. Researchers have tried to analyze this behavior using models that allow the variance of the series to change through time.

## 2.3 Classifying Models of Changing Volatility

There are numerous models that can capture the phenomenon of volatility clustering, modeling the conditional variance and covariance of the underline time series (see Shep-

hard, 1996; Bollerslev, Chou and Kroner, 1992; Gouriéroux, 1997). Cox (1981) divided them into two major classes, **observation driven** and **parameter-driven** models. In order to discuss these two classes we will assume

$$y_t | \Psi_t \sim N(0, \sigma_t^2),$$

where  $y_t$  is a realization of the stochastic process at time  $t$ ,  $N(\cdot, \cdot)$  is the Normal density and  $\Psi_t$  contains all the past information up to time  $t$ , i.e.  $\Psi_t = \{y_t, y_{t-1}, \dots\}$ . The **observation-driven** models allow  $\Psi_t$  to be a function of lagged values of  $y_t$ . The most typical examples are the Autoregressive Conditional Heteroskedasticity (ARCH) model (Engle, 1982) and the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model (Bollerslev, 1986). ARCH model put the conditional variance of  $y_t$  as a linear function of the squares of the past observation

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

Similarly in GARCH model, the conditional variance depends on the lag observations plus the lag conditional variances

$$\sigma_t^2 = a_0 + a_1 y_{t-1}^2 + \dots + a_p y_{t-p}^2 + b_1 \sigma_{t-1}^2 + \dots + b_q \sigma_{t-q}^2.$$

On the other hand, **parameter-driven** (or state-space) models put  $\Psi_t$  as a function of some unobserved or "latent" component. The main representative of this class of models is the stochastic volatility model (Taylor, 1982). The general form of the stochastic volatility model

$$\begin{aligned} y_t | h_t &\sim N(0, \exp(h_t)), \\ h_t | a, d, h_{t-1}, \sigma_\eta^2 &\sim N(a + d \cdot h_{t-1}, \sigma_\eta^2). \end{aligned}$$

Here  $h_1, \dots, h_t$  (log-volatilities) are unobserved, something which classifies this model as the parameter-driven models.

Reviews of the literature on this topic are given by Shephard (1996), Bollerslev, Chou and Kroner (1992), Bera and Higgins (1993), Bollerslev, Engle and Nelson (1994), Diebold and Lopez (1995), Gouriéroux (1997), Hafner (1998), and a collection of ARCH papers is in Engle (1995).

## 2.4 Univariate Models

### 2.4.1 ARCH-type Models

The simplest observation-driven time-varying volatility model is the ARCH model, which has been proposed by the Engle (1982). The ARCH model of order  $p$  can be written as

$$y_t | \mathbf{a}, \mathbf{y}_{t-1} \sim N(0, \sigma_t^2), \quad (2.1)$$

$$\sigma_t^2 = a_0 + \sum_{i=1}^p a_i y_{t-i}^2, \quad (2.2)$$

where  $y_t$  is the time series at time  $t$ ,  $\mathbf{a} = (a_0, \dots, a_p)$  and  $\mathbf{y}_{t-1} = (y_{t-1}, y_{t-2}, \dots)$ . To prevent the nonnegativity of the conditional variance (2.2) the parameters  $a_0, a_1, \dots, a_p$  are restricted to be nonnegative. Following Theorem 2 of Engle (1982), the ARCH( $p$ ) process (2.1) is covariance stationary if and only if all the roots of the associated characteristic equation lie outside the unit circle, therefore  $\sum_{i=1}^p a_i < 1$ . Under these restrictions the

unconditional variance exists and it is equal to  $Var(y_t) = E(y_t^2) = a_0 / \left(1 - \sum_{i=1}^p a_i\right)$ .

The moment conditions of the ARCH(1) model are established by Engle (1982). By the assumed normality in (2.1) all the odds moments are zero. The second moment  $E(y_t^2)$  exists only if  $a_1 < 1$  and it is equal to  $Var(y_t) = E(y_t^2) = a_0 / (1 - a_1^2)$ , the  $E(y_t^4)$  exists if  $3a_1^2 < 1$  and it is equal

$$E(y_t^4) = \frac{3a_0^2(1 - a_1^2)}{(1 - a_1^2)^2(1 - 3a_1^2)}.$$

The kurtosis of the unconditional distribution of  $y_t$  is

$$Kurtosis(y_t) = \frac{E(y_t^4)}{E(y_t^2)^2} - 3 = \frac{3(1 - a_1^2)}{(1 - 3a_1^2)} - 3.$$

The kurtosis for  $a_1 > 0$  is greater than zero therefore the unconditional distribution of  $y_t$  is leptokurtic. The properties and the estimation of the parameters of the ARCH model are discussed by Engle (1982), White (1982) and Gouriéroux, Monfort and Tragnon (1984) among several others. Bayesian inference procedures for ARCH models have been developed by Geweke (1989a, b), Polasek and Muller (1995) and Vrontos (1997).

The GARCH model (Bollerslev, 1986; Taylor, 1986), is a generalization of the ARCH model. In detail Bollerslev included in the equation (2.2) of the conditional variance autoregressive terms. The GARCH( $p, q$ ) model has the following formula:

$$y_t | \mathbf{a}, \mathbf{b}, \mathbf{y}_{t-1} \sim N(0, \sigma_t), \quad (2.3)$$

$$\sigma_t^2 = a_0 + \sum_{i=1}^p a_i y_{t-i}^2 + \sum_{j=1}^q b_j \sigma_{t-j}^2, \quad (2.4)$$

where  $y_t$  is the time series at time  $t$ ,  $\mathbf{a} = (a_0, \dots, a_p)$ ,  $\mathbf{b} = (b_1, \dots, b_q)$ , and  $\mathbf{y}_{t-1} = (y_{t-1}, y_{t-2}, \dots)$ . Sufficient, but not necessary conditions such that  $\sigma_t^2 > 0$  for all  $t$  are  $a_0 > 0$ ,  $a_i \geq 0$ ,  $i = 1, \dots, p$  and  $b_i \geq 0$ ,  $i = 1, \dots, q$ . From Theorem 1 of Bollerslev (1986),  $y_t$  is covariance stationary - which is equivalent to the existence of the unconditional variance - if and only if  $\sum_{i=1}^p a_i + \sum_{j=1}^q b_j < 1$ . In special case where  $\sum_{i=1}^p a_i + \sum_{j=1}^q b_j = 1$  the model is called Integrated GARCH (IGARCH). Under the above restrictions the unconditional variance of  $y_t$  under the GARCH( $p, q$ ) model is

$$Var(y_t) = E(y_t^2) = a_0 / \left( 1 - \sum_{i=1}^p a_i - \sum_{j=1}^q b_j \right).$$

Under the assumption of normality in (2.3) all the odds moments are zero. The kurtosis

of the unconditional distribution of  $y_t$  under the GARCH(1,1) model is

$$Kurt(y_t) = \frac{E(y_t^4)}{E(y_t^2)^2} - 3 = \frac{3 + 6a_1^2}{(1 - 3a_1^2 - 2a_1b_1 - b_1^2)} - 3.$$

Note that the forth moment of the GARCH(1,1) exists if and only if  $3a_1^2 + 2a_1b_1 + b_1^2 < 1$ . If this condition is true the unconditional density is leptokurtic because the kurtosis is greater than zero.

Regarding the estimation of the parameters of the GARCH model see Bollerslev (1986, 1987), Baillie and Bollerslev (1989), Fiorentini, Calzolari and Panattoni (1996). A Bayesian analysis of GARCH models was proposed by Bauwens and Lubrano (1998), Muller and Pole (1999), Bos, Mahieu and van Dijk (1999), Vrontos, Dellaportas and Politis (2000) among several others.

The generalized exponential GARCH model (Nelson 1991)

$$\begin{aligned} y_t | \mathbf{a}, \theta, \gamma &\sim N(0, \sigma_t), \\ \log(\sigma_t^2) &= \omega_t + \sum_{i=1}^{\infty} a_i g(\varepsilon_{t-i}), \\ g(\varepsilon_t) &= \theta \varepsilon_t + \gamma (|\varepsilon_t| - E|\varepsilon_t|), \end{aligned} \tag{2.5}$$

where  $y_t$  is the time series at time  $t$ ,  $\omega_t$  and  $\mathbf{a} = (a_1, a_2, \dots)$  and  $\varepsilon_t$  is a sequence of independent and identically distributed (i.i.d.) random variables with mean zero and variance one. For more information regarding the exponential ARCH model see Engle and Ng (1993), Taylor (1994), Poon and Taylor (1992), Day and Lewis (1992), Kuwahara and Marsh (1992) and Vrontos, Dellaportas and Politis (2000).

Apart from the above models, many other versions of the ARCH/GARCH model have been proposed in the literature such as Weak GARCH, Fractional Integrated GARCH, Log GARCH, nonlinear ARCH, asymmetric GARCH, threshold GARCH, ARCH-M etc. For more information regarding the extensions of the ARCH model see Shephard (1996), Taylor (1986), Geweke (1988), Pantula (1986), Higgins and Bera (1992), Glosten, Ja-

gannathan and Runkle (1993), Davidian and Carroll (1987), Lee and Hansen (1994), Gouriéroux and Monfort (1992), He and Terasvirta (1999), Engle, Lilien and Robins (1987).

## 2.4.2 Stochastic Volatility Models

The basic alternative to ARCH class of models is the stochastic volatility model which is included to the class of parameter-driven models. In this class of models the conditional variance  $\sigma_t^2$  depends, not on past observation such as ARCH and GARCH processes, but on some unobserved component or latent structure.

The form of the univariate Stochastic Volatility Model (Taylor, 1982, 1986; Tauchen and Pitts, 1983) is

$$y_t|h_t \sim N(0, h_t), \quad (2.6)$$

$$h_t|a, d, \sigma_h^2 \sim LN(a + d \log(h_{t-1}), \sigma_h^2), \quad (2.7)$$

where  $y_t$  is the time series at time  $t$ ,  $a$ ,  $d$  and  $\sigma_h^2$  are the hyperparameters and  $LN(\cdot, \cdot)$  is the LogNormal density. 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 persistence of the volatility (standard deviation) and  $\sigma_h^2$  is the volatility of the latent parameters  $\mathbf{h} = (h_1, h_2, \dots, h_t, \dots)$ .

To ensure that the series of the log-volatilities to be covariance stationary, the restriction  $|d| < 1$  is imposed. In case that  $d = 1$  then the series of  $h_t$  is random walk.

An alternative parameterization of the stochastic volatility model that has also attracted some attention is given by

$$y_t|h_t \sim N(0, \exp(h_t)), \quad (2.8)$$

$$h_t|a, d, \sigma_h^2 \sim N(a + dh_{t-1}, \sigma_h^2). \quad (2.9)$$

Note that, in (2.9) the series of log-volatilities  $h_t$  follow a standard Gaussian autoregressive process (AR(1)).

The unconditional mean and the variance of the latent parameter  $h_t$  are given by

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

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

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

Moreover, the coefficient of kurtosis is

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

From the above formulas, it is obvious that the distribution of the time series  $y_t$  is leptokurtic and symmetric. The dynamic properties of the stochastic volatility model appear most clearly if we square  $y_t$  and take it logs. Then

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

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). The theoretical properties of the absolute values of  $y_t$  are given in Ghysels, Harvey and Renault (1999).

As far as the estimation of the parameters of the stochastic volatility model is concerned, Duffie and Singleton (1993), Melino and Turnbull (1990), Andersen (1996), An-

dersen and Sorensen (1996) and Andersen *et al.* (1999) used the Generalized Method of Moments (GMM) (see, Hamilton, 1994 and Rothenberg, 1973). Quasi-Likelihood method was used by Harvey, Ruiz and Shephard (1994). Bayesian methods are followed by Shephard (1993, 1996), Pitt and Shephard (1999b), Gilks and Wild (1992), Wild and Gilks (1993), Jacquier, Polson and Rossi (1994), Kim, Shephard and Chib (1998) and Chib, Nardari and Shephard (2002).

Apart from ARCH and stochastic volatility models, an additional class of models has been proposed by Harvey *et al.* (1992). These models have a different intellectual perspective but they imply particular forms of conditional heteroskedasticity. The time series is composed of several sources of errors where all or some of them have an ARCH form. Since the error components cannot be separately observed given the past information, these models can be classified as stochastic volatility models (Shephard 1996). In general, we can say that this class is a “connection” between ARCH-type and stochastic volatility models. A distinguished member of this class is the unobserved ARCH model, presented by Shephard (1996). The ARCH components in this model are observed with errors. The form of this model can be written using the following hierarchical structure of conditional densities:

$$\begin{aligned} y_t | f_t, \sigma^2 &\sim \mathbf{N}(f_t, \sigma^2), \\ f_t | f_{t-1}, \alpha, b, f_0 &\sim \mathbf{N}(0, h_t), \\ h_t &= \alpha + b \cdot f_{t-1}^2, \end{aligned} \tag{2.10}$$

where  $y_1, \dots, y_T$  is a realization of the process,  $f_t$  is the unobserved ARCH component at time  $t$ ,  $f_0$  is the initial state or the “history” of the unobserved components and  $\mathbf{N}(\cdot, \cdot)$  is the Normal distribution. For the nonnegativity of the variance  $h_t$  of the factor the parameters  $a$  and  $b$  are restricted to be positive. The additional restriction  $0 < b \leq 1$  is placed so that the ARCH component of the model to be covariance stationary (Engle 1982). Note that the unobserved component  $f_t$  is not measurable with respect to the

available information at time  $t$ , something which characterizes this class of models. The unconditional and conditional variances of  $y_t$  are given by  $Var(y_t) = \sigma^2 + a/(1 - b)$  and  $Var(y_t|y_{t-1}, a, b) = \sigma^2 + h_t$ . Therefore, the stochastic process  $y_t$  can be considered to have an underline variance on which it is added the variability which is caused by the effect of volatility clustering.

## 2.5 Multivariate Models

### 2.5.1 ARCH/GARCH Unconstrained Models

Kraft and Engle (1982) introduce the basic multivariate GARCH( $p, q$ ) model

$$\mathbf{y}_t | \cdot \sim MN_k(0, \Sigma_t) , \quad (2.11)$$

where

$$\Sigma_t = \begin{bmatrix} \sigma_{11t}^2 & \sigma_{12t}^2 & \cdots & \sigma_{1kt}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1t}^2 & \sigma_{k2t}^2 & \cdots & \sigma_{kkt}^2 \end{bmatrix} ,$$

$$vech(\Sigma_t) = vech(C) + \sum_{i=1}^p A_i vech(\boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}_{t-1}^*) + \sum_{j=1}^q B_j vech(\varepsilon_{t-j}) .$$

Here  $\mathbf{y}_t = (y_{1t}, \dots, y_{kt})'$  is the realization of the  $k$ -dimensional stochastic process of time  $t$ ,  $MN_k(\cdot, \cdot)$  is the  $k$ -variate Normal density,  $vec(\cdot)$  denotes the column stacking operator of the lower portion of a symmetric matrix and  $C, A_i$  ( $i = 1, \dots, p$ ) and  $B_j$  ( $j = 1, \dots, q$ ) are the matrices of the parameters. The above model is heavily parameterized. There is  $\{k(k+1)/2\} + (p+q)\{k(k+1)/2\}^2$  unknown parameters to be estimated ( $k = 4$  delivers 212 parameters). Moreover it is extremely difficult to state the conditions to ensure that  $\Sigma_t$  is always positive definite. For these reasons, many scientists search for plausible

constraints on the parameters of this model.

### 2.5.2 ARCH/GARCH Constrained Models

As noted in the previous section, the direct extension of the univariate ARCH/GARCH models to the multivariate case leads to a very large number of unknown parameters to be estimated. An attempt to reduce the number of parameters of the multivariate ARCH/GARCH model produced the diagonal ARCH/GARCH model (Bollerslev, Engle and Wooldridge 1988). This model sets the matrix  $A_i$  and  $B_i$  of the model (2.11) to be diagonal. In detail, the diagonal GARCH( $p, q$ ) is written as

$$y_t | \cdot \sim MN_k(0, \Sigma_t),$$

where

$$\Sigma_t = \begin{bmatrix} \sigma_{11t}^2 & \sigma_{12t}^2 & \cdots & \sigma_{1kt}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1t}^2 & \sigma_{k2t}^2 & \cdots & \sigma_{kkt}^2 \end{bmatrix},$$

and

$$vech(\Sigma_t) = C + \sum_{i=1}^p \mathbf{a}_i vech(\boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}_{t-1}') + \sum_{j=1}^q \beta_j vech(\boldsymbol{\varepsilon}_{t-j}),$$

where  $\mathbf{a}_i$  ( $i = 1, \dots, p$ ) and  $\beta_j$  ( $j = 1, \dots, q$ ) is  $n(n+1)/2$  vectors of parameters. More simply each element of the conditional covariance matrix  $\varepsilon_t$  can be written as

$$\sigma_{klt}^2 = C_{kl} + \sum_{i=1}^p a_{kli} \varepsilon_{k,t-1} \varepsilon_{l,t-1} + \sum_{j=1}^q \beta_{klj} \sigma_{kl,t-j}^2.$$

Note that, and for this model the constraints that is needed such as  $\Sigma_t$  to be always positive definite is not an easy task.

### 2.5.3 ARCH/GARCH Models With Constant Conditional Correlation

A popular constrained ARCH/GARCH model is the constant conditional correlation model of Bollerslev (1990). Here the correlation among series is constant  $\sigma_{ijt}^2 = \rho_{ij}\sigma_{iit}\sigma_{jjt}$  over time. In detail, the GARCH( $p, q$ ) with constant conditional correlation is written as

$$\begin{aligned}
 y_t | \cdot &\sim MN_K \left( 0, \Sigma_t^{1/2} R \Sigma_t^{1/2} \right), \\
 \Sigma_t &= \begin{bmatrix} \sigma_{1t}^2 & 0 & \cdots & 0 & 0 \\ 0 & \sigma_{2t}^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_{k-1t}^2 & 0 \\ 0 & 0 & \cdots & 0 & \sigma_{kt}^2 \end{bmatrix}, \\
 R &= \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1(k-1)} & \rho_{1k} \\ \rho_{21} & 1 & \cdots & \rho_{2(k-1)} & \rho_{2k} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{(k-1)1} & \rho_{(k-1)2} & \cdots & 1 & \rho_{(k-1)k} \\ \rho_{k1} & \rho_{k2} & \cdots & \rho_{k(k-1)} & 1 \end{bmatrix}, \\
 \sigma_{kt}^2 &= C_k + \sum_{i=1}^p a_{ki} \varepsilon_{k,t-1}^2 + \sum_{j=1}^q b_{kj} \sigma_{k,t-j}^2,
 \end{aligned}$$

The restriction of constant correlation  $\text{corr}(y_{it}, y_{jt} | Y_{t-1}) = \rho_{ij}$  has been found to be empirically reasonable (Baillie and Bollerslev, 1990).

Regarding the applications of the above multivariate model see Giovannini and Jorion (1989), Baillie and Bollerslev (1990), Schwert and Seguin (1990), Kroner and Claessens (1991), Kroner and Sultan (1991), Ng (1991), Bekaert and Hodrick (1993), Turtle, Buse and Korkie (1994) and Jeantheau (1998) among others.

### 2.5.4 Factor ARCH/GARCH Models

The latent factor ARCH model is proposed by Diebold and Nerlove (1989). Here each component of the stochastic process is expressed as a linear combination of a few underlying processes called factors, and of a noise. In detail, the one-factor model is written as

$$\begin{aligned} y_t | \boldsymbol{\lambda}, f_t, \Sigma &\sim MN_k(\boldsymbol{\lambda}, f_t, \Sigma), \\ f_t | f_{t-1}, a, b &\sim N(0, a + bf_{t-1}^2), \end{aligned}$$

where  $\mathbf{y}_t$  is a  $k$ -variate realization of the stochastic process,  $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_k)$  is the vector of loadings,  $a, b$  are the hyperparameters of the ARCH components and  $f_t$  is the factor component. Here  $f_t$ 's are unobserved (latent) quantities and they must be estimated from the data.

King, Sentana and Wadhawani (1994) modeled the common factor to follow a GARCH process. Overall, different factor models have been proposed in the literature, and have been analyzed by many researches (see, for example, Diebold and Nerlove (1989), Laux and Ng (1993), Engle, Ng and Rothschild (1990), King, Sentana and Wadhawani (1994), Demos and Sentana (1996), Fiorentini *et al.* (2004) among several others),

### 2.5.5 The Multivariate Stochastic Volatility Model

Based on the approach of Bollerslev (1990) constant correlation ARCH model, Harvey, Ruiz and Shephard (1994) proposed the analog in stochastic volatility model. The form

of this multivariate model is given by

$$\begin{aligned}
\mathbf{y}_t | H_t, R &\sim MN_K \left( 0, H_t^{1/2} R H_t^{1/2} \right), \\
H_t &= \text{diag} (h_{1t}, \dots, h_{kt}), \\
R &= \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1(k-1)} & \rho_{1k} \\ \rho_{21} & 1 & \cdots & \rho_{2(k-1)} & \rho_{2k} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{(k-1)1} & \rho_{(k-1)2} & \cdots & 1 & \rho_{(k-1)k} \\ \rho_{k1} & \rho_{k2} & \cdots & \rho_{k(k-1)} & 1 \end{bmatrix}, \\
h_{it} | a, d, h_{i,t-1}, \sigma_i &\sim LN \left( a_i + d \log (h_{i,t-1}), \sigma_i^2 \right),
\end{aligned}$$

where  $R$  is a constant correlation matrix,  $h_t$  is the *log*-volatility at time  $t$  and  $a_i, d_i, \sigma_i^2$  for  $i = 1, 2, \dots, k$  are the unknown hyperparameters of the model.

### 2.5.6 The Multivariate Factor Stochastic Volatility Model

The idea of Diebold and Nerlove (1989) for the factor ARCH model can be also applied to the stochastic volatility models. The simplest one-factor stochastic volatility model puts

$$\begin{aligned}
y_t | \boldsymbol{\lambda}, f_t, \Sigma &\sim MN_K (\boldsymbol{\lambda} f_t, \Sigma), \\
f_t | h_t &\sim N (0, \exp (h_t)), \\
h_t | h_{t-1}, d, \sigma_n^2 &\sim N (d h_{t-1}, \sigma_n^2),
\end{aligned}$$

where  $\boldsymbol{\lambda}$  is the  $k$ -variate vector of loadings,  $f_t$  is the latent (common) factor,  $h_t$  is the log-volatility of the factor,  $a, d, \sigma_n^2$  are the parameters of the log-volatilities and  $\Sigma$  is a diagonal covariance matrix. Moreover the elements of  $\Sigma$  can be driven by independent stochastic volatility models. Models including in the general framework of factor stochastic volatility have been studied by Jacquier, Polson and Rossi (1999) and Aguilar and West (2000).

# Chapter 3

## The Bayes Framework and MCMC

### 3.1 Bayes's Theorem

The framework that is chosen to work with is the Bayesian analysis (Jeffreys 1939, Bernardo and Smith 1994, O'Hagan 1994). This approach of statistics is totally based on Bayes theorem (Bayes, 1763):

**Theorem 1** *Bayes Theorem(continuous random variables)*

*Let  $\theta$  and  $y$  continuous random variables, then*

$$f(\theta|y) = \frac{f(y|\theta) \pi(\theta)}{\int_{\theta} f(y|\theta) \pi(\theta) d\theta}.$$

The application of the above theorem is straightforward in applied Statistics. Let  $\mathbf{y}$  the set of the data which we have at hand. The data are assumed to come from a density  $f(\cdot|\boldsymbol{\theta})$  where  $\boldsymbol{\theta}$  is the vector of parameters of the assumed distribution. The goal is to investigate the distribution of parameters  $\boldsymbol{\theta}$  given the  $\mathbf{y}$ , i.e. given the data at hand. Suppose that  $\pi(\boldsymbol{\theta})$  is the a priori information about  $\boldsymbol{\theta}$  then the density of  $\boldsymbol{\theta}$  given the

data (a posteriori density) is extracted via Bayes theorem

$$f(\boldsymbol{\theta}|\mathbf{y}) = \frac{f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}} \quad (3.1)$$

*or*

$$f(\boldsymbol{\theta}|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}),$$

where  $\propto$  means equal up to a constant. For more information about the Bayesian Statistics see for example, Bernardo and Smith (1994), Jeffreys (1939), Pole, West and Harrison (1994).

In many cases, there is 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 non informative. Commonly, non-informative priors based on Jeffreys priors for location and scale are used. For example: if  $\theta \in \mathbf{R}$  then  $[\theta] \propto 1$  and if  $\theta \in \mathbf{R}^+$  then  $[\theta] \propto \frac{1}{\theta}$ . Another approach is to use flat distributions and specifically  $N(0, \sigma^2)$ , where  $\sigma^2$  is considered a large number.

The challenging part in Bayesian analysis is to calculate the integral in the denominator of (3.1). In most cases, this cannot be done analytically and other techniques must be used. The last two decades Markov Chain Monte Carlo (MCMC) techniques are used to overcome this problem.

## 3.2 The Computational Difficulty of Bayesian Analysis

The Bayesian approach presents some difficulties regarding the computational part. To be more specific, the case of the stochastic volatility model is examined.

Let the univariate stochastic volatility model (Jacquier, Polson and Rossi, 1994)

$$\begin{aligned}
y_t|h_t &\sim N(0, h_t) \\
h_t|a, d, \sigma^2, h_0 &\sim LN(a + d \log h_{t-1}, \sigma^2)
\end{aligned}$$

where  $y_t$  is the stochastic process at time  $t$ ,  $a, d, \sigma^2, h_0$  are the hyperparameters of the model and  $h_t$  is the volatilities.

Suppose that we have at hand  $T$  realizations of the  $y_t$ , therefore  $\mathbf{y} = (y_1, \dots, y_T)'$ . Our aim is to investigate the distribution of the hyperparameters and the volatilities  $\mathbf{h} = (h_0, h_1, \dots, h_T)$  of the stochastic volatility model. For this reason, we apply the Bayes theorem as follows:

$$f(a, d, \sigma^2, \mathbf{h}|\mathbf{y}) = \frac{f(\mathbf{y}|a, d, \sigma^2, \mathbf{h}) \pi(a, d, \sigma^2, h_0)}{\int_{\Theta} f(\mathbf{y}|a, d, \sigma^2, \mathbf{h}) \pi(a, d, \sigma^2, h_0) d\Theta} \quad (3.2)$$

where  $\Theta = (a, d, \sigma^2, h_0, \mathbf{h})$  is the parameter space. The formula (3.2) can be written as

$$\begin{aligned}
f(a, d, \sigma^2, \mathbf{h}|\mathbf{y}) &= c(\Theta) f(\mathbf{y}|a, d, \sigma^2, \mathbf{h}) \pi(a, d, \sigma^2, h_0) \\
&\propto f(\mathbf{y}|a, d, \sigma^2, \mathbf{h}) \pi(a, d, \sigma^2, h_0)
\end{aligned}$$

where

$$c(\Theta) = \left\{ \int_{\Theta} f(\mathbf{y}|a, d, \sigma^2, \mathbf{h}) \pi(a, d, \sigma^2, h_0) d\Theta \right\}^{-1}.$$

Note that  $c(\Theta)$  is the normalizing constant of the joint posterior distribution of parameters of interest. To calculate the constant  $c(\Theta)$  requires a  $(T + 4)$  dimensional (numeric) integration (4 hyperparameters and T volatilities).

The marginal posterior distribution for a particular parameter of the model, say  $d$ , is given from the solution of the integral bellow

$$f(d|\mathbf{y}) = c(\Theta) \int \int \dots \int f(\mathbf{y}|a, d, \sigma^2, \mathbf{h}) \pi(a, d, \sigma^2, h_0) da d\sigma^2 dh_0 \dots dh_T,$$

which means that apart from the  $(T + 4)$  dimensional integration for calculating  $c(\Theta)$  we need a  $(T + 3)$ -dimensional integration to marginalize out the rest hyperparameters and volatilities.

Moreover if we want to find out the posterior mean for the parameter  $a$

$$E(a|\mathbf{y}) = \int_a a f(a|\mathbf{y}) da ,$$

one more integration is required. Additionally, more high order integrations are needed in order to calculate other characteristics of the posterior distribution of a specific parameter, say  $Var(a|\mathbf{y})$ ,  $Kurt(a|\mathbf{y})$  or  $Cov(a|\mathbf{y})$ .

From the above it becomes obvious that Bayesian analysis requires high dimensional integration; an approach that many times is not possible to be followed as for example in the case of stochastic volatility model.

Tierney and Kadane (1984) introduce an asymptotic Laplace approximation to calculate the expectation of a function of a parameter of interest, which still involves high dimensional functional maximization and the calculation of the Hessian matrix for each specific approximation. Furthermore, Achcar and Smith (1990) and Hills and Smith (1992) show that these approximations can be very sensitive to the parameterization of the parameters. Other approach to overcome the computational difficulty of the Bayes Analysis is via sampling methods (see Stewart, 1983 and 1987; Steward and Davis, 1986; Geweke, 1988 and 1989b).

### 3.3 Sampling Based Methods

Smith (1991) observed that there is a duality between a distribution and a sample which is generated from this distribution. In detail, if there is a sample that is generated from a specific distribution its distribution can be approximately reconstructed along with its features. This could be a fair solution to the computational problem of the Bayesian Statistics.

Suppose that we have a random sample of size  $n$ ,  $\theta_1^*, \dots, \theta_k^*, \dots, \theta_n^*$  from the joint posterior distribution of the parameters of the stochastic volatility model, where  $\theta_k^* = (a_k^*, d_k^*, \sigma_k^{2*}, \mathbf{h}_k^*)$ . Most of the features that characterize the distribution of interest can be easily calculated from this random sample. More specifically, the posterior mean of the parameter  $a$  is approximately

$$E(a|\mathbf{y}) \cong \sum_{i=1}^n \frac{a_i^*}{n}.$$

The posterior covariance between  $a$  and  $d$  is approximately

$$Cov(a, d|\mathbf{y}) \cong \sum_{i=1}^n \frac{(a_i^* - E(a|\mathbf{y}))(d_i^* - E(d|\mathbf{y}))}{n}.$$

Moreover, the marginal distribution of any parameter can be estimated via non-parametric density estimation, say Kernel estimator

$$f(d|\mathbf{y}) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{a - a_i^*}{h}\right),$$

where  $h$  is the window width, and  $K(\cdot)$  is the Kernel function which satisfies  $\int K(x)dx = 1$ . Moreover, a simple estimation of the posterior density could be given by the histogram of the random sample of the parameter  $a$ .

Therefore, by employing a random sample, most of the features of the posterior density can be approximately estimated. From the above, emerges a question that has to be answered: “Is it possible to sample from the joint posterior density?”, especially if this posterior density is known up to a constant.

In case where it is possible to sample from the posterior density of the parameters of interest, the computational difficulty of Bayesian analysis is eliminated. Thus, the issue is whether it is possible to sample from a density. The answer to this issue can be given by Markov chain Monte Carlo techniques, such as Gibbs sampler and Metropolis-Hastings algorithm. An explicit description of these simulation strategies is given in Smith and

Roberts (1993), Gilks *et al.* (1995), Dellaportas and Roberts (2003).

### 3.3.1 The Gibbs Sampler

The Gibbs sampler is a Markovian technique that enables the researcher to sample from multidimensional density functions.

Suppose that we want to take a sample from a posterior density  $f(\boldsymbol{\theta}|\mathbf{y})$  which is known up to a constant. In this case,  $\boldsymbol{\theta}$  denotes the vector of  $k$  parameters of interest and  $\mathbf{y}$  denotes the data at hand. Let  $f(\theta_i|\boldsymbol{\theta}_{/i}, \mathbf{y})$  denotes the full conditional distribution of parameter  $\theta_i$  and  $\boldsymbol{\theta}_{/i}$  is the vector of the parameters apart from  $\theta_i$ , i.e.  $\boldsymbol{\theta}_{/i} = (\theta_1, \theta_2, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_k)$ .

The basic idea behind Gibbs sampler is relying on the fact that the full conditional distributions  $f(\theta_i|\boldsymbol{\theta}_{/i}, \mathbf{y})$ , for  $i = 1, \dots, k$  uniquely determine the joint posterior density as originally observed by Besag (1974). Based on this, Gelfand and Smith (1990) formulated the Gibbs sampler. Specifically, given an arbitrary set of starting values for the parameters of the model under consideration,  $\boldsymbol{\theta}^0 = (\theta_1^0, \theta_2^0, \dots, \theta_k^0)$  the following algorithm can be implemented

*First iteration*

draw  $\theta_1^1$  from  $f(\theta_1|\mathbf{y}, \theta_2^0, \theta_3^0, \dots, \theta_k^0)$   
draw  $\theta_2^1$  from  $f(\theta_2|\mathbf{y}, \theta_1^1, \theta_3^0, \dots, \theta_k^0)$   
 $\vdots$   
draw  $\theta_k^1$  from  $f(\theta_k|\mathbf{y}, \theta_1^1, \theta_2^1, \dots, \theta_{k-1}^1)$

*Second iteration*

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

*i-th iteration*

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

Having the above iterative algorithm repeated for  $t$  times, the transition probability of going from  $\boldsymbol{\theta}^t = (\theta_1^t, \theta_2^t, \dots, \theta_k^t)$  to  $\boldsymbol{\theta}^{t+1} = (\theta_1^{t+1}, \theta_2^{t+1}, \dots, \theta_k^{t+1})$  is given by

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

As Geman and Geman (1984) proved that under mild conditions

$$\boldsymbol{\theta}^t = (\theta_1^t, \theta_2^t, \dots, \theta_k^t) \xrightarrow{L} \boldsymbol{\theta}$$

as  $t$  goes to infinity. Note that, Gibbs sampler is a Markovian chain since the vector  $\boldsymbol{\theta}^t = (\theta_1^t, \theta_2^t, \dots, \theta_k^t)$  depends only on the position of the previous vector  $\boldsymbol{\theta}^{t-1} = (\theta_1^{t-1}, \theta_2^{t-1}, \dots, \theta_k^{t-1})$ .

The simplicity of implementation of the aforementioned algorithm, has made this approach very popular among Bayesians, which explains the existence of many examples

found on related literature that are based on this method. For more information about the Gibbs Sampler and its applications see Besag and Green (1993), Casela and George (1992), Dellaportas and Smith (1993), Gelfand and Smith (1990), Gelfand *et al.* (1990), Geman and Geman (1984), Roberts and Sahu (1997), Gilks *et al.* (1995), Spiegelhalter *et al.* (1995a, 1995b).

A critical point to this sampling scheme, is the requirement of having all the full conditional densities tractable to sampling, something that is not very common in practice. In cases where a full conditional density cannot be sampled, then the Metropolis-Hastings algorithm can offer some solutions.

### 3.3.2 The Metropolis-Hastings Algorithm

In many applications, there are no specific algorithms to sample from the conditional densities resulting from the Gibbs sampler. Still, there are cases where a sample from all the parameters or a block of parameters must be drawn at once, something that might be proven difficult to be directly performed. In such cases an iterative sampling scheme, Metropolis-Hastings algorithm (Metropolis *et al.*, 1953 & Hastings, 1970), can be used as an alternative for sampling from the distribution of interest.

Suppose that a sample must be retrieved from a density  $f(\boldsymbol{\theta}|\mathbf{y})$  which is known up to a constant. Assume that in order to construct the Markov chain, a candidate-generating density is  $q(\boldsymbol{\theta}^{old}, \boldsymbol{\theta}^{new})$  where  $\int q(\boldsymbol{\theta}^{old}, \boldsymbol{\theta}^{new}) d\boldsymbol{\theta}^{new} = 1$  (Metropolis *et al.*, 1953 & Hastings, 1970). This density generates values  $\boldsymbol{\theta}^{new}$  when a process is at the point  $\boldsymbol{\theta}^{old}$ . Of course, this candidate density must follow some conditions, such as the reversibility condition for all  $q(\boldsymbol{\theta}^{old}, \boldsymbol{\theta}^{new})$ . In most cases, this condition is not satisfied. Therefore, for some pairs  $(\boldsymbol{\theta}^{old}, \boldsymbol{\theta}^{new})$ , the following might occur:

$$f(\boldsymbol{\theta}^{old}|\mathbf{y}) q(\boldsymbol{\theta}^{old}, \boldsymbol{\theta}^{new}) > f(\boldsymbol{\theta}^{new}|\mathbf{y}) q(\boldsymbol{\theta}^{old}, \boldsymbol{\theta}^{new}).$$

In these cases, the process moves from  $\boldsymbol{\theta}^{old}$  to  $\boldsymbol{\theta}^{new}$  too often, while the opposite

(from  $\boldsymbol{\theta}^{new}$  to  $\boldsymbol{\theta}^{old}$ ) happens very rarely. To repair this, a probability  $a(\boldsymbol{\theta}^{old}, \boldsymbol{\theta}^{new}) < 1$  is introduced, which is called *probability of move*, that corrects the condition of reversibility and reduces the number of moves from  $\boldsymbol{\theta}^{old}$  to  $\boldsymbol{\theta}^{new}$ .

Therefore, if  $p(\boldsymbol{\theta}^{old}, \boldsymbol{\theta}^{new})$  is the transition matrix of the chain (when the states are discrete) or the transition kernel (when the states are continuous),

$$\begin{aligned} f(\boldsymbol{\theta}^{old}|\mathbf{y}) p(\boldsymbol{\theta}^{old}, \boldsymbol{\theta}^{new}) &= f(\boldsymbol{\theta}^{new}|\mathbf{y}) p(\boldsymbol{\theta}^{new}, \boldsymbol{\theta}^{old}), \\ f(\boldsymbol{\theta}^{old}|\mathbf{y}) q(\boldsymbol{\theta}^{old}, \boldsymbol{\theta}^{new}) \cdot a(\boldsymbol{\theta}^{old}, \boldsymbol{\theta}^{new}) &> f(\boldsymbol{\theta}^{new}|\mathbf{y}) q(\boldsymbol{\theta}^{new}, \boldsymbol{\theta}^{old}) \cdot a(\boldsymbol{\theta}^{new}, \boldsymbol{\theta}^{old}). \end{aligned}$$

As it has been already stated, the movement from  $\boldsymbol{\theta}^{new}$  to  $\boldsymbol{\theta}^{old}$  is not occurring often enough, therefore  $a(\boldsymbol{\theta}^{new}, \boldsymbol{\theta}^{old})$  must be defined to be as large as possible. Still, having  $a(\boldsymbol{\theta}^{new}, \boldsymbol{\theta}^{old})$  being a probability, it must be set as  $a(\boldsymbol{\theta}^{new}, \boldsymbol{\theta}^{old}) = 1$ ; then

$$f(\boldsymbol{\theta}^{old}|\mathbf{y}) q(\boldsymbol{\theta}^{old}, \boldsymbol{\theta}^{new}) \cdot a(\boldsymbol{\theta}^{old}, \boldsymbol{\theta}^{new}) = f(\boldsymbol{\theta}^{new}|\mathbf{y}) q(\boldsymbol{\theta}^{new}, \boldsymbol{\theta}^{old}),$$

and the probability of move is produced by the form

$$a(\boldsymbol{\theta}^{old}, \boldsymbol{\theta}^{new}) = \min \left[ \frac{f(\boldsymbol{\theta}^{new}|\mathbf{y}) q(\boldsymbol{\theta}^{new}, \boldsymbol{\theta}^{old})}{f(\boldsymbol{\theta}^{old}|\mathbf{y}) q(\boldsymbol{\theta}^{old}, \boldsymbol{\theta}^{new})}, 1 \right].$$

if  $f(\boldsymbol{\theta}^{old}|\mathbf{y}) q(\boldsymbol{\theta}^{old}, \boldsymbol{\theta}^{new}) > 0$  or 1 otherwise.

Therefore, the Metropolis-Hasting Algorithm takes the following form:

- Step 1: Generate  $\mathbf{z}$  from  $q(\boldsymbol{\theta}^{old}, \cdot)$  and  $u$  from  $\mathbf{U}(0, 1)$ , where  $\mathbf{U}(\cdot, \cdot)$  denotes the uniform distribution.
- Step 2: Let  $\boldsymbol{\theta}^{new} = \mathbf{z}$  if  $u \leq a(\boldsymbol{\theta}^{old}, \mathbf{z})$ ; otherwise let  $\boldsymbol{\theta}^{new} = \boldsymbol{\theta}^{old}$ .
- Repeat the steps 1 and 2  $n$  times in order to take a sample of size  $n$ .

An introduction to the Metropolis-Hastings is contained in Chib and Greenberg (1995a).

A crucial detail in application of Metropolis-Hastings algorithm is the choice of the candidate distribution. There are many options of candidate (proposal) densities and those will be summarized below:

$$\alpha) a(\boldsymbol{\theta}^{old}, \boldsymbol{\theta}^{new}) = a(\boldsymbol{\theta}^{new}, \boldsymbol{\theta}^{old})$$

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

$$\beta) a(\boldsymbol{\theta}^{old}, \boldsymbol{\theta}^{new}) = a(\boldsymbol{\theta}^{new})$$

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.

$\gamma$ ) Other choices

There is a wide set of possible candidate densities. For more information see, Chib and Greenberg (1995b), Tierney (1994) and Hastings (1970). Furthermore, other updating schemes of the Metropolis-Hastings algorithm are discussed in Marinari and Parisi (1992), Geyer and Thompson (1995), Mira (1999), and Tierney and Mira (1999).

Note that, in the use Metropolis-Hastings algorithm if the candidate distribution is a symmetric<sup>1</sup> distribution such as Normal then the probability of move is

$$a(\boldsymbol{\theta}^{old}, \boldsymbol{\theta}^{new}) = \min \left[ \frac{f(\boldsymbol{\theta}^{new}|\mathbf{y})}{f(\boldsymbol{\theta}^{old}|\mathbf{y})}, 1 \right].$$

The Metropolis-Hastings algorithm can be applied as itself or with combination with the Gibbs sampler. In detail, if in a Gibbs scheme, the full conditional densities of a posterior density are not of known forms then the Metropolis-Hastings must be applied in order to sample from them. This last remark is very crucial since this combination of the two

---

<sup>1</sup>Symmetric with respect to its arguments,  $q(\boldsymbol{\theta}^{new}, \boldsymbol{\theta}^{old}) = q(\boldsymbol{\theta}^{old}, \boldsymbol{\theta}^{new})$ .

algorithms constructs a powerful tool in order to exploit the properties of the posterior distribution (Chib and Greenberg, 1995a).

On the other hand, the Metropolis-Hastings algorithm suffers from a serious drawback when compared to the Gibbs sampler, which is the necessity of tuning to the spread of the proposal density (Chib and Greenberg 1995a), in order to succeed efficient sampling.

### 3.3.3 Example

For the case of the stochastic volatility model the MCMC algorithm that can be used in order to sample from the posterior density of the parameters of interest is the following (Jacquier, Polson and Rossi, 1994; Giakoumatos 1997). Using non-informative priors for the hyperparameter of the model the joint posterior density up to a constant is

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

where  $\boldsymbol{\theta}$  is the vector of the hyperparameters  $(a, d, \sigma_\eta^2)'$  and  $\mathbf{h} = (h_0, h_1, \dots, h_T)$  is the vector of log-volatilities.

The full conditional densities are of the form

- $a | \cdot \sim \mathbf{Normal} \left( \frac{\sum h_t - d \cdot \sum h_{t-1}}{T}, \frac{\sigma_\eta^2}{T} \right)$ , where  $a | \cdot$  means the parameter  $a$  given the rest of the parameters and the data.
- $d | \cdot \sim \mathbf{Normal} \left( \frac{\sum h_t \cdot h_{t-1} - a \cdot \sum h_{t-1}}{\sum h_{t-1}^2}, \frac{\sigma_\eta^2}{\sum h_{t-1}^2} \right)$ .
- $\sigma_\eta^2 | \cdot \sim \mathbf{IG} \left( \frac{T}{2}, \frac{1}{2} \cdot \sum_{t=1}^T (h_t - a - d \cdot h_{t-1})^2 \right)$ , where  $\mathbf{IG}(\cdot, \cdot)$  is the inverse gamma distribution.

- $h_0|\cdot \sim \mathbf{Normal}\left(\frac{h_1-a}{d}, \frac{\sigma_\eta^2}{d^2}\right)$ .
- $f(h_t|\cdot) \propto \exp[-0.5(h_t + y_t^2 \exp(-h_t))] \exp\left[-\frac{1}{2s^2}(h_t - m_t)^2\right]$ , for  $t = 1, \dots, T$ . Here  $m_t = \frac{[a \cdot (1-d) + d \cdot (h_{t+1} + h_{t-1})]}{(1+d^2)}$  and  $s^2 = \frac{\sigma_\eta^2}{(1+d)}$ .

The above full conditional densities of the log-volatilities, do not have trivial forms. They are consisted of a Normal term and a term that is very unusual in common practice.

In order to sample from these full conditional densities, the Metropolis-Hasting algorithm can be applied. Giakoumatos (1997) used the dependent Metropolis-Hastings algorithm with Normal proposal density. Thus, the proposed sampling point for the  $i$ -th value of a specific log-volatility, say  $h_t$  is sampled from

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

and the probability of acceptance is

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

Chib and Greenberg (1994) proposed that the constant  $c$  must be chosen in such a way that the acceptance probability will be about 50%. Still, there is a major scientific debate concerning the proper value of the acceptance probability. In this algorithm Giakoumatos (1997) proposed the constant  $c$  to be equal to  $\frac{1}{(1+d^2)}$  which gives probability of acceptance approximately 65%.

### 3.3.4 The Auxiliary Variable Sampler

The basic idea of the auxiliary variable sampler can be traced back to the papers of Swendsen and Wang (1987), Edwards and Sokal (1988), Besag and Green (1993), Damien and Walker (1996), Damien *et al.* (1999), Higdon (1998) and Neal (2003). All the above authors pointed out that if the parameter space of a posterior density is properly

increased by including extra latent variables, this could make the resulting posterior density more tractable by sampling methods. In particular, Damien *et al.* (1999) gave specific guidelines to cope with a series of popular statistical models. As an initial approach to this method the following examples display the usefulness and the easiness of the auxiliary variable sampler.

**Example 2** Suppose that we want to take sample from  $f(\theta) \propto \pi(\theta) \cdot \prod_{i=1}^N l_i(\theta)$ , and assume that this is not possible by employing standard methods. Applying the auxiliary variable sampler the set of latent variables  $\mathbf{u} = (u_1, \dots, u_N)$  can be introduced, where  $u_i \in (0, \infty)$  such that the joint density of  $\theta$  and  $\mathbf{u}$  to become  $f(\theta, \mathbf{u}) \propto \pi(\theta) \prod_i \mathbf{I}(u_i \leq l_i(\theta))$ , where  $\mathbf{I}(\cdot)$  is the indicator function. Note that, the marginal density of  $\theta$  is  $f(\theta)$ . Now it is easy to sample from  $f(\theta)$  using the Gibbs sampler. First we sample  $\mathbf{u}$  from their full conditional densities which are uniforms i.e.  $u_i \sim U(0, l_i(\theta))$  and then we sample from the full conditional density of  $\theta$  which is  $\pi(\theta) I_A$ ;  $A = \{\theta : l_i(\theta) \geq u_i, i = 1 \dots N\}$ .

**Example 3** Suppose that we want to take sample from  $f(\theta) \equiv \mathbf{N}(m, \tau)$ , where  $\tau$  denotes the accuracy. In this case the latent variables  $u \in (0, \infty)$  can be introduced resulting the following joint density of  $\theta$  and  $u$

$$f(\theta, u) \propto \sqrt{u} \exp(-0.5ut) \mathbf{I}(u \geq (\theta - m)^2). \quad (3.3)$$

Note that if the latent variable  $u$  is integrated out from (3.3), then the marginal density of  $\theta$  becomes the density of interest. Thus, a sample from  $f(\theta, u)$  can be taken using the following Gibbs steps:

1.  $\theta|u \sim \mathbf{U}(m - \sqrt{u}, m + \sqrt{u})$ .
2.  $u|\theta \sim \mathbf{Gamma}(0.5, 0.5\tau) \mathbf{I}(u \geq (\theta - m)^2)$ .

Moreover if the full conditional density of  $u|\theta$  is marginalized with respect to  $\theta$ , a sample from  $f(\theta, u)$  can be drawn using the following steps:

1.  $\theta|u \sim \mathbf{U}(m - \sqrt{u}, m - \sqrt{u})$ .

2.  $u|\theta \sim \mathbf{Gamma}(\frac{3}{2}, \frac{\tau}{2})$ .

A detailed presentation of the auxiliary variable sampler will be provided on chapter 5.

# Chapter 4

## A New Convergence Diagnostic Based on Subsampling

### 4.1 Introduction

Due to the computational difficulty of Bayesian Statistical analysis, statisticians nowadays routinely use Markov Chain Monte Carlo (MCMC) methods to simulate from complex, non standard posterior distributions; see, for example Tierney (1994). As noted in previous sections, the Gibbs sampler (Geman and Geman, 1984; Gelfand and Smith, 1990). the Metropolis-Hastings algorithm (Metropolis, Rosenbluth, Rosenbluth, Teller and Teller, 1953; Hastings, 1970) and the auxiliary variable sampler (Damien *et al.*, 1999; and Neal, 2003) are the most popular methods for the analysis of complex statistical models.

The main problem that turns up, using such methods, is to gauge when convergence is achieved; that is, to assess at what point the Markov chain “gets in” the target distribution, and to figure out how many points will have to be taken from this distribution in order to estimate, with the desired accuracy, the parameters of interest. The first part of the problem is that of determining the length of the required “burn-in” period or the

point in which the Markov chain has “forgotten” its starting point; observations within an initial transient phase are discarded, in order to reduce the bias caused by the effect of starting values. The second part is that of determining the number of sample points which “adequately” represent the posterior distribution and on which inference will be based. We will be using the expression “the Markov chain gets in the target distribution” meaning that the Markov chain has effectively converged to its asymptotic distribution.

In order to solve the problem of convergence, a number of different diagnostics have been proposed in the literature. From a theoretical point of view, there is an attempt to predetermine the number of iterations that will ensure convergence; see, for example, Polson (1996), Rosenthal (1993), Roberts and Sahu (1997) and Schervish and Carlin (1992). Due to the difficulty of this approach in practice, almost all of the applied works are based on the output produced by running Markov Chain Monte Carlo algorithms.

Some of the proposed methods try to assess convergence, while others attempt to “measure” the performance of any particular sampler. The choice of diagnostic depends on the problem at hand. The theoretical background, the assumptions, the assessment of convergence of the joint or the marginal density of an MCMC output, the number of chains that are needed, the range of the samplers to which the diagnostic is applied, the computational expense and the interpretability are some of the criteria for the choice of the diagnostic. These convergence diagnostics are described in the review papers of Cowles and Carlin (1996) Robert and Mengersen (1999) and Brooks and Roberts (1999).

In the next sections some of the most popular convergence tests (included in CODA software (Best and Cowles, 1995)) are presented and a new convergence test is proposed (Giakoumatos, Vrontos, Delaportas and Politis, 1999) based on the subsampling methodology.

## 4.2 CODA tests

Most of the MCMC users based their decision that their Markov chains converge to the distribution of interest to the results of the convergence diagnostics that included in CODA software (Best and Cowles, 1995). These tests are

- 1 Geweke test
- 2 Gelman and Rubin test
- 3 Raftery and Lewis test
- 4 Heidelberger and Welch test

Geweke (1992) proposed a test that regards the output of the MCMC as a time series and uses methods from spectral analysis to estimate the asymptotic variance of the MCMC sampler. Let  $n$  the number of iterations of the MCMC sampler,  $q(\cdot)$  the function that we want to estimate from the MCMC output. Assume that the spectral density  $S_q(\omega)$  exists and has not discontinuities at the frequency 0. If these assumptions are satisfied then for the estimation

$$\bar{q}_n = \frac{\sum_{i=1}^n q(\theta^{(i)})}{n}$$

of  $E(q(\theta))$  the asymptotic variance is  $\frac{S_q(0)}{n}$ , where  $\theta^{(i)}$  for  $i = 1, \dots, n$  is the output of the sampler. Using this estimation of the variance we can test if the difference of mean  $\bar{q}(\theta)_n^A$ , based on the first  $n_A$  iterations and  $\bar{q}(\theta)_n^B$ , based on the last  $n_B$  iterations is statistically significant. Geweke suggests using  $n_A = 0.1n$  and  $n_B = 0.5n$ .

The test of Gelman and Rubin (1992) that was included in the CODA software is applied in the output of two or more independent chains beginning from different starting points. Assume that  $m$  independent chains are implemented for  $2n$  iterations. Based on the last  $n$  iterations Gelman and Rubin method estimate the target distribution of each scalar of interest as a conservative *Student - t* distribution.. Convergence is assumed if

the "shrink factor" for each scalar is near to 1. The "shrink factor" is estimated as

$$\sqrt{\widehat{R}} = \sqrt{\left(\frac{n-1}{n} + \frac{m+1}{mn} \frac{B}{W}\right) \frac{df}{df-2}},$$

where  $B$  is the variance between means from the  $m$  chains,  $W$  is the average of the  $m$  within-chain variances and  $df$  is the degrees of freedom of the approximating  $t$  density.

Raftery and Lewis's (1992) test is based on two-state Markov chain theory. A minimum number of iterations, say  $n_{\min}$ , is needed to obtain the desired precision of estimation if the samples are independent. For any specific quantile  $q$  of the posterior density, and desired accuracy the program reports the total number of iterations that algorithms should be run and how many of the beginning iterations should be discarded. Moreover, the test indicates  $k$  where only every  $k^{th}$  iterations should be kept in order to have independent sample.

Heidelberger and Welch (1983) proposed a procedure for generating a confidence interval of prespecified width for the mean of the parameters of interest. Let  $n$  the number of iterations that MCMC has been run. Let  $n_1 = 0.1n$  an initial number of iterations. Using the rest iterations a spectral density estimation of  $S(0)$  is computed. Then a Schrunben's stationarity test on the entire run is performed. If the null hypothesis is rejected the first  $n_1$  iterations are discarded and the test is repeated until the null hypothesis does not rejected or the half of the  $n$  iterations have been discarded.

## 4.3 An MCMC Convergence Diagnostic using Sub-sampling

### 4.3.1 Introduction

In this section, a new diagnostic will be proposed (Giakoumatos, Vrontos, Dellaportas and Politis, 1999) which assesses the convergence of both marginal and joint posterior

densities. It can be applied to any MCMC sampler and uses the output from a single chain to gauge convergence. Generalization for multiple chains is straightforward, and this diagnostic can be very useful when there is suspicion of very influential starting points; see Gelman (1996). This test can be used to detect MCMC convergence in great generality and does not need to be combined with other methods in order to be effectively implemented. The subsampling method for statistical inference as developed by Politis and Romano (1994) and Politis, Romano and Wolf (1997) is used. The diagnostic is based on obtaining (via subsampling)  $(1 - \alpha)100\%$  confidence regions for the posterior mean and for the 90th percentile of the first marginal distribution of the Markov chain at hand; the assessment of convergence can be done through looking at a set of graphs. An alternative diagnostic is also proposed that uses the asymptotic normal distribution together with a subsampling estimate of the asymptotic variance-covariance matrix. A recent approach, which also looks with different perspective at the use of confidence intervals as a means of convergence diagnosis, is given by Brooks and Gelman (1998). The authors generalized the method of Gelman and Rubin (1992) and they used the (squared) ratio of the lengths of the empirical estimated confidence intervals for the parameters of interest as an alternative interpretation of the  $\hat{R}$  diagnostic (Gelman and Rubin, 1992), which is defined as the ratio of the between and within variances of the MCMC sequences. This alternative calculation of  $\hat{R}$  is simpler than the original ratio of variances and is free from the assumption of normality.

### 4.3.2 The Subsampling Methodology

In this section the basic ideas of subsampling methodology for time series (see Politis and Romano, 1994 and Politis, Romano and Wolf, 1997) are reviewed. Let  $(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots, \boldsymbol{\theta}^{(N)})$  be an observed stretch of a multivariate time series  $\{\boldsymbol{\theta}^{(s)}, s = 1, 2, \dots\}$ ; each  $\boldsymbol{\theta}^{(s)}$  is assumed to be a  $(p \times 1)$  vector. The time series is also assumed to be strong mixing, and asymptotically stationary.

The assumption of strong mixing is an assumption of “asymptotic independence”:

for any positive integers  $i$  and  $n$ , the two sets of random variables  $(\boldsymbol{\theta}^{(i)}, \boldsymbol{\theta}^{(i+1)}, \dots, \boldsymbol{\theta}^{(i+n)})$  and  $(\boldsymbol{\theta}^{(i+n+k)}, \boldsymbol{\theta}^{(i+n+k+1)}, \dots, \boldsymbol{\theta}^{(i+2n+k)})$  should be approximately independent if  $k$  is large enough; see Rosenblatt (1956), Doukhan (1994) or Politis, Romano and Wolf (1997) for a precise definition. Similarly, the notion of asymptotic stationarity means that the sequence  $(\boldsymbol{\theta}^{(k)}, \boldsymbol{\theta}^{(k+1)}, \dots)$  is approximately stationary (in the strict sense – see Brockwell and Davis, 1991) if  $k$  is large enough.

The basic idea of subsampling is to approximate the sampling distribution of a statistic based on the values of the same statistic recomputed over smaller subsets of the data that retain the dependence structure of the observations. If we are willing to consider subsets of size  $b(< N)$ , where  $b$  is a positive integer that in general may depend on  $N$ , then we are led to consider the  $B = N - b + 1$  “blocks” of consecutive observations of the type  $(\boldsymbol{\theta}^{(i)}, \boldsymbol{\theta}^{(i+1)}, \dots, \boldsymbol{\theta}^{(i+b-1)})$ , for  $i = 1, \dots, B$ .

Let  $T_N$  be a statistic of interest that is a function of the data sequence  $(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots, \boldsymbol{\theta}^{(N)})$ . The statistic  $T_N$  is generally a vector (say  $q$ -dimensional), and is employed as an estimator of an unknown parameter  $\theta$ . In general,  $\theta$  can be a parameter of the whole (infinite-dimensional) joint distribution of  $\{\boldsymbol{\theta}^{(s)}, s = 1, 2, \dots\}$ ; however, in the MCMC case considered in the next sections,  $\theta$  will almost always be a parameter of the invariant distribution of the Markov chain, i.e., the “asymptotic” first marginal of the sequence  $\{\boldsymbol{\theta}^{(s)}, s = 1, 2, \dots\}$ .

It is assumed that  $T_N$  is consistent for  $\theta$  as the sample size  $N \rightarrow \infty$ . Specifically, it is assumed that the statistic  $T_N$ , suitably centered and normalized, possesses a non-degenerate large-sample distribution. To be more precise, let  $\{\tau_n, n = 1, 2, \dots\}$  be an increasing sequence that diverges to  $\infty$  as  $n \rightarrow \infty$ , let  $\|\cdot\|$  denote a norm on space  $\mathbf{R}^q$ , and let  $J_N(x) = \text{Prob}[\tau_n \|T_N - \theta\| \leq x]$ . The assumption required is that there exists some nondegenerate continuous distribution function  $J(\cdot)$  such that

$$J_N(x) \rightarrow J(x) \tag{4.1}$$

for all  $x$  as  $N \rightarrow \infty$ .

Let  $T_{i,b}$  be the statistic of interest computed from block  $(\boldsymbol{\theta}^{(i)}, \boldsymbol{\theta}^{(i+1)}, \dots, \boldsymbol{\theta}^{(i+b-1)})$ , with all other data temporarily ignored, and construct an “empirical” distribution of the “subsample values”  $\{T_{i,b}, i = 1, \dots, B\}$  by

$$L_N(x) = \frac{1}{B} \cdot \sum_{i=1}^B 1\{\tau_b \|T_{i,b} - T_N\| \leq x\}. \quad (4.2)$$

If the series  $\{\boldsymbol{\theta}^{(s)}, s = 1, 2, \dots\}$  is strong mixing and asymptotically stationary, and if the statistic possesses a nondegenerate large-sample distribution (i.e., equation (4.1) holds), then  $L_N(\cdot)$  is a consistent estimator of the (generally unknown) limit distribution  $J(\cdot)$ , provided  $b$  is chosen in a way that:  $b \rightarrow \infty$  as  $N \rightarrow \infty$ , but  $b/N \rightarrow 0$  and  $\tau_b/\tau_N \rightarrow 0$ ; see Politis, Romano and Wolf (1997). Perhaps more important is that consistent estimation of the quantiles of  $J(\cdot)$  can be achieved by looking at the quantiles of  $L_N(\cdot)$ ; in other words, for any  $t \in (0, 1)$ ,

$$L_N^{-1}(t) \rightarrow J^{-1}(t) \quad (4.3)$$

in probability as  $N \rightarrow \infty$ , where  $L_N^{-1}(t) \equiv \inf\{x : L_N(x) \geq t\}$  and  $J^{-1}(t) = \inf\{x : J(x) \geq t\}$  are the  $t$  quantiles of  $L_N(\cdot)$  and  $J(\cdot)$  respectively.

Using the quantiles of  $L_N(\cdot)$  we can now construct confidence regions for  $\theta$  with a prescribed coverage level (to be attained in large samples). Relation (4.3) implies that the set  $\{\theta : \tau_N \|T_N - \theta\| \leq L_N^{-1}(1 - \alpha)\}$  is a confidence region for  $\theta$  with asymptotic coverage probability equal to the nominal  $1 - \alpha$ .

Note that our choice of norm  $\|\cdot\|$  will dictate the shape of those confidence regions. If  $\|\cdot\|$  is the Euclidean norm, then the confidence regions for  $\theta$  will be spheres centered at  $T_N$ . Notably, the choice of sup-norm for  $\|\cdot\|$  (i.e., the  $l_\infty$  norm which is nothing other than the maximum absolute coordinatewise deviation) results in confidence regions for  $\theta$  that have the shape of hypercubes with edges that are perpendicular to the axis; the

sup-norm will be our choice in what follows as confidence regions that are hypercubes have the useful alternative interpretation as confidence intervals for the coordinates of  $\theta$  with *simultaneous* coverage equal to the coverage level of the whole confidence region.

**Remark 2.1.** If the variances of different coordinates of the multivariate statistic  $T_N$  are of different orders of magnitude it may be inefficient to construct a hypercube for a confidence region for  $\mu$ ; rather, a “hyper-parallelepiped” should be constructed instead. To achieve this, the subsampling methodology must be applied to a “studentized” version of our statistic  $T_N$  as discussed in Politis and Romano (1994). In other words, a new “studentized” multivariate statistic  $\tilde{T}_N$  is defined with the property that the coordinates of  $\tilde{T}_N$  have all approximately equal variances; for example, we can define  $\tilde{T}_N^{(k)} = T_N^{(k)} / \hat{S}_{T_N^{(k)}}^2$ , where  $\hat{S}_{T_N^{(k)}}^2$  is a consistent estimate of the variance of  $T_N^{(k)}$ , and thus ensure that all coordinates of  $\tilde{T}_N$  have approximately variance equal to one. Notably,  $\hat{S}_{T_N^{(k)}}^2$  may even be a subsampling estimate of variance so that an “iterated” subsampling takes effect. However, for practical purposes, even a rough preliminary variance estimate can be used in this type of studentization with good ensuing results.

### 4.3.3 The MCMC Subsampling Diagnostic

In this section, the proposed method for assessing convergence of the MCMC output is presented. Let  $(\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(N)})$  be the multivariate output of an MCMC simulation; as before, each  $\theta^{(s)}$  is assumed to be a  $(p \times 1)$  vector with  $k$ th coordinate denoted by  $\theta_k^{(s)}$ .

Notably, both assumptions required for subsampling to work (i.e., strong mixing and asymptotic stationarity) hold true in the MCMC case that interests us where the sequence  $\{\theta^{(s)}, s = 1, 2, \dots\}$  is a Markov chain that possesses a unique invariant (i.e., stationary) distribution but the starting value  $\theta^{(1)}$  may follow a different distribution; see e.g. Meyn and Tweedie (1993).

Let  $\bar{\theta}_N = N^{-1} \sum_{i=1}^N \theta^{(i)}$  denote the sample mean of the observed sequence  $(\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(N)})$ ;  $\bar{\theta}_N$  is of course a  $(p \times 1)$  vector as well, and its  $k$ th coordinate will

be denoted by  $\bar{\boldsymbol{\theta}}_{N,k}$ . For  $t \in (0, 1)$ , let  $\mathbf{q}_N^t$  be a  $(p \times 1)$  vector with  $k$ th coordinate denoted by  $q_{N,k}^t$ , where  $q_{N,k}^t$  is the empirical  $t$ -quantile of the  $k$ th coordinate data sequence  $(\theta_k^{(1)}, \theta_k^{(2)}, \dots, \theta_k^{(N)})$ ; in other words, if  $(\theta_k^{*(1)} \leq \theta_k^{*(2)} \leq \dots \leq \theta_k^{*(N)})$  are the order statistics of the  $k$ th coordinate data sequence  $(\theta_k^{(1)}, \theta_k^{(2)}, \dots, \theta_k^{(N)})$ , then  $q_{N,k}^t = \theta_k^{*(\lfloor tN+1 \rfloor)}$ , where  $\lfloor \cdot \rfloor$  is the integer part.

Using the subsampling methodology and choosing for our statistic  $\mathbf{T}_N$  either  $\bar{\boldsymbol{\theta}}_N$  or  $\mathbf{q}_N^t$  (with some choice of  $t$  that is of interest, e.g.  $t = 0.90$ ), we can construct confidence regions for the mean and the  $t$  quantile of the “asymptotic” first marginal, i.e., the unique invariant distribution of the Markov chain  $\{\boldsymbol{\theta}^{(s)}, s = 1, 2, \dots\}$ . Note that in either case ( $\bar{\boldsymbol{\theta}}_N$  or  $\mathbf{q}_N^t$ ) we have in general that  $\tau_N = N^{1/2}$ , i.e.,  $\sqrt{N}$ -convergent statistics. Thus, we can choose the block size  $b$  proportional to  $N^\gamma$  (for some constant  $\gamma \in (0, 1)$ ), and thus ensure that the conditions for (4.3) –on which the construction of confidence regions is based– are fulfilled.

As mentioned before, the sup-norm is recommended for use, i.e.,  $\|\mathbf{T}_N\| = \sup_{k=1, \dots, q} |T_{N,k}|$ , where  $T_{N,k}$  is the  $k$ th coordinate of vector  $\mathbf{T}_N$ ; therefore, the confidence regions are hypercubes in  $\mathbf{R}^q$ . Note also that in both cases ( $\bar{\boldsymbol{\theta}}_N$  and  $\mathbf{q}_N^t$ ), the “observation” dimension  $p$  coincides with the “parameter” dimension  $q$ .

Since the confidence regions are hypercubes in  $\mathbf{R}^q$ , we can define the “range” of such a confidence region as the  $q$ th root of its volume. The proposed diagnostic can now be viewed as a consequence of the following fact:

- The “range” of a  $(1-\alpha)100\%$  confidence region for either the mean or the  $t$  quantile of the “asymptotic” first marginal is (asymptotically) proportional to  $1/\sqrt{N}$ .

Based on the above fact, our diagnostic can be formulated as follows:

- (A) **Estimation of “burn-in” time.** As the simulation is running and  $N$  increases, construct  $(1-\alpha)100\%$  confidence regions for the  $t$  quantile based on different (increasing) values of  $N$ ; we used  $\alpha = 0.05$  and  $t = 0.90$  but other choices are possible as well. Now plot the “range” of the confidence region versus  $1/\sqrt{N}$ . We would

estimate the “burn-in” time to be  $N^*$  if the plot of “range” versus  $1/\sqrt{N}$  is approximately linear for  $N > N^*$ . Linearity can be checked by visual inspection of the plot, but we also recommend to use a plot of the coefficient of determination of a weighted linear regression between the dependent variable “range” and  $1/\sqrt{N}$ . Having estimated the “burn-in” time to be  $N^*$ , observations  $(\theta_k^{(1)}, \theta_k^{(2)}, \dots, \theta_k^{(N^*)})$  are discarded from the simulation as they could introduce undesired bias.

(B) **Figuring out when to stop the simulation.** Again as the simulation is running and  $N$  increases, construct  $(1 - \alpha)100\%$  confidence regions for the mean  $\mu$  of the “asymptotic” first marginal of the Markov chain  $\{\boldsymbol{\theta}^{(s)}\}$ . Since the main objective of the simulation is to estimate  $\mu$  by Monte Carlo, and since the error in this estimation can be quantified by the range of a  $(1 - \alpha)100\%$  confidence region for  $\mu$ , we would then propose to stop the MCMC simulation when the range of this  $(1 - \alpha)100\%$  confidence region (with  $\alpha = 0.05$ , say) is appropriately small, smaller than some prespecified absolute or relative measure of accuracy; for example, we could stop when the range becomes smaller than 0.001 (say), thus obeying an absolute measure, or when it is smaller than  $0.001 \|\bar{\boldsymbol{\theta}}_N\|$  which is then a relative measure (relative to  $\|\mu\|$  which is estimated by  $\|\bar{\boldsymbol{\theta}}_N\|$ ).

The proposed diagnostic is simple and easy to use, and has the significant advantage of being valid in *asymptotically* stationary settings such as the MCMC case of interest; see e.g. Yue and Chan (1996). Notably, other diagnostics (e.g. Geweke, 1992) are shown to be valid only under the assumption of stationarity, thus neglecting the fact that the MCMC output is *not* exactly stationary.

**Remark 3.1.** The reason why the  $t$  quantile (with a large  $t$ , say  $t = 0.90$ ) is considered in part A of the diagnostic as opposed to a similar graphical plot of the confidence range of other statistics (e.g. the sample mean) is based on the idea that a stabilization of estimates of the invariant distribution of the Markov chain (especially in the tails) is a reliable indicator of the target distribution having been achieved.

**Remark 3.2.** Other statistics are also possible even in part B of the diagnostic. For example, in a simulated annealing setting where the posterior mode is the objective, our statistic  $T_N$  could have as  $k$ th coordinate the sample mode of the  $k$ th coordinate data sequence  $(\theta_k^{(1)}, \theta_k^{(2)}, \dots, \theta_k^{(N)})$ ; for concreteness however, we will focus in the sequel on the sample mean and  $t$  quantile of an MCMC output.

Given the burn-in sample of the MCMC chain and the sample size needed to estimate the parameters of interest with the desired accuracy, we can estimate the variance covariance matrix of these parameters using the whole sample (excluding the burn-in) with a subsampling estimate. For example, we could use the blocked sample variance as introduced by Politis and Romano (1993) and its formula is presented in the next section. This treatment could be beneficial as indicated by MacEachern and Berliner (1994).

#### 4.3.4 An Alternative ‘Hybrid’ Diagnostic

In cases where the asymptotic distribution of our statistic  $T_N$  is known to be normal, then a ‘hybrid’ method that combines subsampling with the information regarding asymptotic normality may be used. Note that, the sample mean and sample 0.90 quantile, which are chosen as statistics of interest, are both asymptotically normal under standard regularity conditions; see e.g. Brockwell and Davis (1991). To elaborate, let  $(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots, \boldsymbol{\theta}^{(N)})$  be the multivariate output of an MCMC simulation; as before, each  $\boldsymbol{\theta}^{(s)}$  is assumed to be a  $(p \times 1)$  vector with  $k$ th coordinate denoted by  $\theta_k^{(s)}$ . The assumptions that are required for subsampling (i.e., strong mixing and asymptotic stationarity) still are assumed to hold together with the additional assumption that  $\sqrt{N}(\mathbf{T}_N - \boldsymbol{\theta}) \xrightarrow{L} N(0, \boldsymbol{\Sigma}_\infty)$ , for some nonnegative definite matrix  $\boldsymbol{\Sigma}_\infty$ .

The problem that now turns up is the estimation of the unknown asymptotic variance-covariance matrix  $\boldsymbol{\Sigma}_\infty$ . Subsampling can be used for this purpose as well, under some additional regularity conditions. Carlstein (1986) proposed the subsampling estimators of variance based on nonoverlapping blocks from a stationary sequence, while Künsch (1989) addressed the case of overlapping blocks. Recently, Fukuchi (1997) demonstrated

the asymptotic consistency of the subsampling estimator of variance in the case of asymptotically stationary time series considered in this paper.

Thus, as an estimator of  $\Sigma_\infty$  we may use the blocked sample variance matrix  $\hat{\mathbf{V}}_{b/N}$ , where  $b$  is the block size as in the previous section. The blocked sampled variance is given from the following formula (cf. Politis and Romano, 1993):

$$\hat{\mathbf{V}}_{b/N} = \begin{bmatrix} \hat{\sigma}_{b/N}^{11} & \hat{\sigma}_{b/N}^{12} & \cdots & \hat{\sigma}_{b/N}^{1p} \\ \hat{\sigma}_{b/N}^{21} & \hat{\sigma}_{b/N}^{22} & \cdots & \hat{\sigma}_{b/N}^{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{b/N}^{p1} & \hat{\sigma}_{b/N}^{p2} & \cdots & \hat{\sigma}_{b/N}^{pp} \end{bmatrix}$$

where

$$\hat{\sigma}_{b/N}^{ij} = \frac{b}{N-b+1} \sum_{m=1}^{N-b+1} [T_{m,b,i} - T_{N,i}] [T_{m,b,j} - T_{N,j}], \quad i, j = 1, 2, \dots, p \quad (4.4)$$

Since  $\hat{\mathbf{V}}_{b/N}$  is an asymptotically consistent estimator of  $\Sigma_\infty$  it follows that we can approximate the probability law of

$\left\| \sqrt{N}(\mathbf{T}_N - \boldsymbol{\theta}) \right\| = \sup_{k=1, \dots, q} \left| \sqrt{N}(T_{N,k} - \theta_k) \right|$ , by the probability law of  $\sup_{k=1, \dots, q} |Z_k|$ , where the multivariate random variable  $\mathbf{Z}$  has the  $N(0, \hat{\mathbf{V}}_{b/N})$  distribution. Although the latter is difficult to evaluate analytically, it is nevertheless very easy to approximate by Monte Carlo. For this purpose, we let  $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_M$  be i.i.d. random variables having distribution  $N(0, \hat{\mathbf{V}}_{b/N})$ , where  $M$  is large. Now we have

$$Prob\left[ \sup_{k=1, \dots, q} \left| \sqrt{N}(T_{N,k} - \theta_k) \right| \leq x \right] \simeq M^{-1} \sum_{m=1}^M 1\left\{ \sup_{k=1, \dots, q} |Z_{m,k}| \leq x \right\} \simeq \tilde{L}_N(x), \quad (4.5)$$

where we define

$$\tilde{L}_N(x) \equiv \lim_{M \rightarrow \infty} M^{-1} \sum_{m=1}^M 1\left\{ \sup_{k=1, \dots, q} |Z_{m,k}| \leq x \right\};$$

note that the approximations " $\simeq$ " in equation (4.5) occur with high probability for

large  $M$  and  $N$ , i.e. they are justified as convergences in probability. The bootstrap ‘rule-of-thumb’ is to take  $M$  of the order of a few thousands, and in our examples we chose  $M = 10000$ .

The alternative ‘hybrid’ subsampling diagnostic proceeds the same way (parts A and B) as the first subsampling diagnostic, the only difference being that now our confidence regions for the mean and  $t$ -quantile are based on the quantiles of the ‘hybrid’ distribution  $\tilde{L}_N(\cdot)$ , and not on the quantiles of the subsampling distribution  $L_N(\cdot)$ . In cases where the assumption of asymptotic normality is valid then the alternative ‘hybrid’ method is expected to work better than the subsampling diagnostic of the previous section as it uses this additional information (on the asymptotic normality). On the other hand, if the assumption of normality does not hold then the subsampling diagnostic of the previous section must be used.

### 4.3.5 Implementation

#### Description of the simulations

Using the MCMC algorithm, we recursively create the Markov sequence  $(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots, \boldsymbol{\theta}^{(N)})$ , where each  $\boldsymbol{\theta}^{(i)}$  is a  $(p \times 1)$  vector;  $p$  denotes the number of parameters and  $N$  is the number of iterates (the total sample size). As before the quantity of interest is a function of the data sequence  $(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots, \boldsymbol{\theta}^{(N)})$  which we denote by the  $(q \times 1)$  vector  $\mathbf{T}_N$ ; in our case we will consider  $\bar{\boldsymbol{\theta}}_N$  or  $\mathbf{q}_N^t$  for our  $\mathbf{T}_N$  statistic, and therefore  $q = p$ .

To fix ideas, let  $N_j = jN/100$ , for  $j = 1, \dots, 100$ . The simulation algorithm is now precisely described; for  $j = 1, \dots, 100$  do the following:

1) As discussed in the previous section, identify the  $B_j = N_j - b_j + 1$  subsamples  $(\boldsymbol{\theta}^{(i)}, \boldsymbol{\theta}^{(i+1)}, \dots, \boldsymbol{\theta}^{(i+b_j-1)})$ , for  $i = 1, \dots, B_j$ ; note that the subsample size  $b_j$  depends on  $N_j$ . We used the simple choice  $b_j = \sqrt{N_j}$ , although other choices are possible as well; see Hall, Horowitz and Jing (1996), Politis, Romano and Wolf (1997).

2) From the sequence  $(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots, \boldsymbol{\theta}^{(N_j)})$ , calculate the quantity of interest  $\mathbf{T}_{N_j}$  (sample mean or  $t$  quantile with  $t = 0.90$ ).

3) From each subsample  $(\boldsymbol{\theta}^{(i)}, \boldsymbol{\theta}^{(i+1)}, \dots, \boldsymbol{\theta}^{(i+b_j-1)})$ , recalculate the quantity of interest  $\mathbf{T}_{i,N_j}$ , where  $i = 1, 2, \dots, B_j$ .

4) Let  $T_{i,N_j,k}$  and  $T_{N_j,k}$  denote the  $k$ th coordinates of the vectors  $\mathbf{T}_{i,N_j}$  and  $\mathbf{T}_{N_j}$  respectively. For  $i = 1, 2, \dots, B_j$ , compute the “maximum deviation” (i.e. sup-norm) from block  $i$  as

$$d_{i,j} = \max_{k=1,\dots,q} |T_{i,N_j,k} - T_{N_j,k}|,$$

and

$$D_{i,j} = \sqrt{b_j} \cdot d_{i,j}$$

5) In order to find the estimated quantile  $L_{N_j}^{-1}(1 - \alpha)$  (with  $\alpha = 0.05$ ), we sort  $D_{i,j}$ ,  $i = 1, 2, \dots, B_j$  in an ascending order to come up with the order statistics  $D_{(1,j)} \leq D_{(2,j)} \leq \dots \leq D_{(B_j,j)}$ , and then let

$$L_{N_j}^{-1}(1 - \alpha) = D_{(\lfloor (1-\alpha) \cdot B_j + 1 \rfloor, j)}$$

where  $\lfloor \cdot \rfloor$  is the integer part.

6) As discussed before, the confidence region at “time”  $N_j$  is a hypercube in  $q$  dimensions that is centered at the value  $\mathbf{T}_{N_j}$  and has “sides” that are perpendicular to the coordinate axes. The length of each “edge” of the hypercube is identical to the “range”  $R_j$  of the hypercube region which is given by

$$R_j = \frac{2L_{N_j}^{-1}(1 - \alpha)}{\sqrt{N_j}}.$$

A different way of describing this situation is to say that the  $q$  confidence intervals of the type  $\left(T_{N_j,k} \pm \frac{L_{N_j}^{-1}(1-\alpha)}{\sqrt{N_j}}\right)$ , for  $k = 1, \dots, q$ , have *simultaneous* coverage  $1 - \alpha$  for the  $q$  respective coordinate parameters, provided of course the sample size  $N_j$  is large.

7) Finally plot  $R_j$  versus  $1/\sqrt{N_j}$  and take appropriate action based on the plot

(quantile case), or just the magnitude of the range for the largest  $N_j$  considered (sample mean case).

The implementation of the alternative ‘hybrid’ diagnostic is exactly like the implementation of the subsampling diagnostic above, the only difference being that the quantile  $\tilde{L}_{N_j}^{-1}(1 - \alpha)$  is put instead of  $L_{N_j}^{-1}(1 - \alpha)$  in all occurrences of the latter.

### Estimation of the ‘burn-in’ using the coefficient of determination

Although linearity can be visually assessed by inspecting the graph, a more objective/automatic method might be desirable. One way to automatically check the linearity between the range of the confidence region and the  $1/\sqrt{N_j}$ , in order to estimate the burn-in, is the coefficient of determination ( $R^2$ ) between these two variables. In detail, the assumed linear relation is given by the linear model

$$\mathbf{y}_{j:j+c-1} = \beta \mathbf{x}_{j:j+c-1} + \boldsymbol{\varepsilon}$$

where  $\mathbf{y}_{j:j+c-1}$ ,  $\mathbf{x}_{j:j+c-1}$  denote the vectors  $(y_j, \dots, y_{j+c-1})$ ,  $(x_j, \dots, x_{j+c-1})$  respectively and  $\boldsymbol{\varepsilon}$  represents a mean-zero error term. Each element of the  $\mathbf{y}_{j:j+c-1}$  is the estimated range of the confidence region of the 0.90 quantile (or of the mean) using  $N_j$  iterations and each element of  $\mathbf{x}_{j:j+c-1}$  is the corresponding  $1/\sqrt{N_j}$ . The quantity  $c$  is chosen by the practitioner and represents the window of our regression, that is, the number of elements  $(y_j, x_j)$  in the vectors  $\mathbf{y}_{j:j+c-1}$ ,  $\mathbf{x}_{j:j+c-1}$  that is used to calculate the coefficient of determination. For each pair of vectors  $(\mathbf{y}_{j:j+c-1}, \mathbf{x}_{j:j+c-1})$  the  $R^2$  is calculated, using the weighted least squares method with weights  $w_j = \sqrt{b_j}/N_j$ . Weighted linear regression is used because the dependent variable is not homoscedastic and its standard deviation is proportional to the ratio  $\sqrt{b_j}/N_j$ ; see, for example, Politis and Romano (1993) where the variance of the subsampling estimate of variance is calculated. In the examples  $c = 20$  is chosen and therefore there are  $(100 - c + 1) = 81$  different  $\mathbf{y}_{j:j+c-1}$ ,  $\mathbf{x}_{j:j+c-1}$  vectors; for each of them we calculate the  $R_k^2$ , for  $k = 1, 2, \dots, 100 - c + 1$ .

If  $R_k^2 > d$  for all  $k > k^*$  and some prespecified threshold  $d$  we discard the first  $N_{k^*-1}$  iterations as burn-in. A typical value of  $d$  we use is 0.999.

### 4.3.6 Examples and simulations

To illustrate how the proposed diagnostic performs in practice, the following three examples are presented. The first example concerns a trivariate normal distribution with high correlation, the second deals with a bimodal mixture of trivariate normals and the third refers to the stochastic search variable selection (SSVS) MCMC output introduced by George and McCulloch (1993).

#### Example 1

This example is taken from the MCMC diagnostics review paper of Cowles and Carlin (1996). It is assumed that we deal with a three parameter joint posterior density which is a zero-mean trivariate normal with correlations 0.90, 0.90 and 0.98 and covariance matrix

$$\begin{pmatrix} 1.0 & 4.5 & 9.0 \\ 4.5 & 25.0 & 49.0 \\ 9.0 & 49.0 & 100.0 \end{pmatrix}.$$

The Gibbs sampler is initialized with  $\theta_2 = 10$ ,  $\theta_3 = -10$ . The algorithm then proceeds by sampling from the full conditional densities, which generally are given by:

$$\theta_i \mid \boldsymbol{\theta}_j \sim N\left(\mu_i + \boldsymbol{\Sigma}_{ij} \cdot \boldsymbol{\Sigma}_{jj}^{-1} \cdot (\boldsymbol{\theta}_j - \boldsymbol{\mu}_j), \boldsymbol{\Sigma}_{ii} - \boldsymbol{\Sigma}_{ij} \cdot \boldsymbol{\Sigma}_{jj}^{-1} \cdot \boldsymbol{\Sigma}_{ji}\right), \quad i = 1, 2, 3$$

where,  $\theta_i$  is the  $i$ th element of the vector of parameters,  $\boldsymbol{\theta}_j$  is the vector of all the other parameters except the  $i$ th,  $\mu_i$  is the  $i$ th element of the mean vector,  $\boldsymbol{\mu}_j$  is the vector of all the other elements except the  $i$ th, and  $\boldsymbol{\Sigma}_{ij}$ ,  $\boldsymbol{\Sigma}_{jj}$ ,  $\boldsymbol{\Sigma}_{ii}$ ,  $\boldsymbol{\Sigma}_{ji}$  are the corresponding partitions of the variance-covariance matrix.

	Iterations				
	6000	12000	18000	24000	30000
$\theta_1$	0.2783	0.2251	0.1824	0.1624	0.1481
$\theta_2$	1.4894	1.2206	0.9867	0.8746	0.8023
$\theta_3$	2.9927	2.4593	1.9854	1.7529	1.6083

Table 4.1: *Accuracy of the posterior mean of the model parameters*

To illustrate the diagnostic test,  $N = 30000$  values are generated from the above iterative Gibbs sampling scheme. A studentization of the values of the parameters has been made using the simple variance estimator. The range of the confidence interval for the 0.90 quantile is estimated for  $N_j = \frac{jN}{100}$ ,  $j = 1, \dots, 100$ , samples. The weighted linear regression of the range versus  $(N_j)^{-1/2}$  is our burn-in indicator. Adopting  $d = 0.999$  as a threshold for  $R^2$ , we discard as burn-in the first 4800 iterations because after that  $R^2 > d$ ; see Figure 4-1, where the values of  $R^2$  are presented across iterations. The results of the alternative ‘hybrid’ method are also illustrated in Figure 4-1 (a) and are similar with the results of our subsampling diagnostic.

The second part of the problem is to estimate how many points are needed to estimate the parameters of interest with the desired accuracy. This can be done by using the range of the confidence interval for the mean. The accuracy of the parameters is the difference of the upper and lower limit of the confidence interval of the mean. Table 4.1 gives the accuracy of the posterior mean of these parameters for some iterations.

A diagnostic which is similar in spirit with this is the one suggested by Raftery and Lewis (1992). To compare the two methodologies, we run the Raftery and Lewis diagnostic requiring the precision achieved by the proposed diagnostic for the 0.90 quantile after 30000 iterations. The input values were  $q = 0.90$ , the 0.90 quantile,  $r = \pm 0.085$ , the precision of the 0.90 quantile and  $s = 0.95$ , the probability of estimating the 0.90 quantile within  $\pm 0.085$  for the parameter  $\theta_1$ . The precision of the 0.90 quantile for parameters  $\theta_2$  and  $\theta_3$  was  $\pm 0.458$  and  $\pm 0.924$  respectively. The required burn-in and sample size

results were 104 and 1105 iterations. It is evident that subsampling diagnostic is much more conservative than the one by Raftery and Lewis; this is in part due to the high choice (0.999) as a linearity threshold in our  $R^2$  criterion. Nevertheless, in this example the Raftery and Lewis diagnostic seems to underestimate the sample size required to get the claimed accuracy. To elaborate, after 1105 iterations and excluding the burn-in of 104 iterations the estimated 0.90 quantiles turn out to be 1.65, 9.02 and 17.78 for the parameters  $\theta_1, \theta_2$  and  $\theta_3$  respectively, whereas after 30000 iterations the 0.90 quantiles are estimated to be 1.31, 6.69, and 13.34 respectively, within the desired accuracy of the true 0.90 quantiles 1.28, 6.40, and 12.81. Note that in both the Raftery and Lewis diagnostic and subsampling diagnostic, convergence detection is related to precision required. By demanding smaller accuracy in the posterior quantity of interest, one needs more iterations for convergence. Cowles and Carlin (1996) also suggest another reason for the Raftery and Lewis method to be less conservative: more iterations are required for estimating quantiles near the median than extreme quantiles due to a formula based on binomial variance.

We ran the subsampling and the alternative method on a Sun Ultra-2, Sparcstation. The computational time for the subsampling diagnostic we propose is around 3180 seconds, while the alternative hybrid method needs 3306 seconds. For both methods the above times were used to calculate confidence intervals for both the mean and the 0.90 quantile of the joint density. The Raftery and Lewis diagnostic required only a fraction of this time (7 seconds) being one of the cheapest diagnostic available; see Brooks and Roberts (1999).

## Example 2

This example is also taken from Cowles and Carlin (1996) and refers to a bimodal target density consisting of a mixture of two trivariate normals with equal probability. These

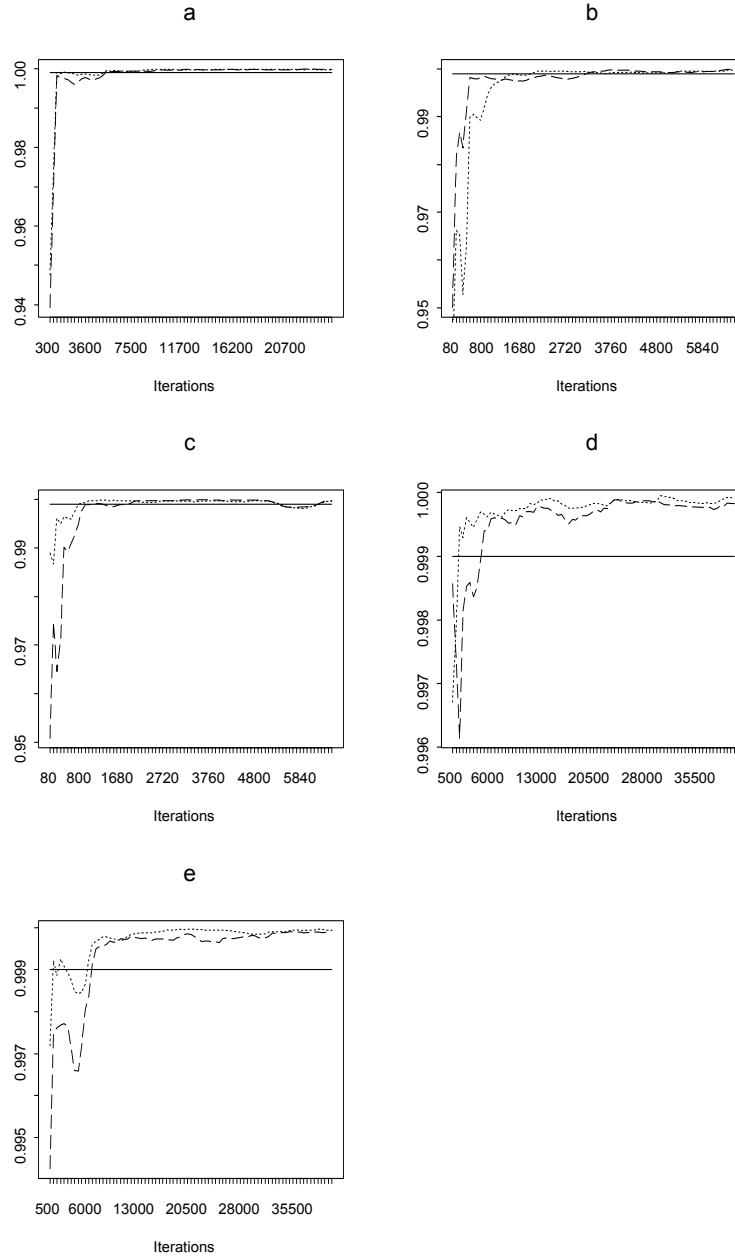


Figure 4-1: Coefficient of determination. Solid line: threshold  $d=0.999$ , dashed line: subsampling method, dotted line: alternative method. (a) Example 1, (b) Example 2, first chain, (c) Example 2, second chain, (d) Example 3, problem 1, (e) Example 3, problem 2.

two normals have a common covariance matrix

$$\begin{pmatrix} 1 & 1.3 & 1.5 \\ 1.3 & 2 & 2 \\ 1.5 & 2 & 4 \end{pmatrix},$$

which produces correlations 0.919, 0.75, 0.707, and mean vectors  $\begin{pmatrix} 0 & 0 & 0 \end{pmatrix}$  and  $\begin{pmatrix} -6 & -8.49 & -12 \end{pmatrix}$ . We generate an MCMC output using a random walk Metropolis Hastings algorithm consisting of 8000 values.

Two independent chains with starting values  $(\theta_1 = 5, \theta_2 = 15, \theta_3 = 10)$  and  $(\theta_1 = -15, \theta_2 = -20, \theta_3 = -25)$  were used to illustrate our diagnostic. The Metropolis proposal density was chosen so that the two chains remain in the area of one mixture component and therefore the chain does not visit the whole parameter space. Figures 4-1 (b) and 4-1 (c) depict the resulting  $R^2$  for both our suggested diagnostics for the two chains respectively. Although clearly the Markov chain has not converged to its stationary distribution, in Figure 4-1 (b) we receive the wrong signal of “getting in” the target distribution, after 3040 iterations using the subsampling and after 1840 iterations using the alternative method. On the other hand, in Figure 4-1 (c) we correctly detect that more iterations are needed because  $R^2$  is not continuously higher than the threshold 0.999. The confidence intervals for the mean of the parameters, for the first and the second chain respectively. Note that the estimated accuracy of the mean of the parameters  $\theta_1, \theta_2, \theta_3$  is 0.3188, 0.4422, 0.6571, respectively for the first chain and 0.3387, 0.4802, 0.5613, for the second chain after 8000 iterations. Also, the estimated accuracy of the 0.90 quantile of the parameters  $\theta_1, \theta_2, \theta_3$  is 0.3873, 0.5161 and 0.7270 respectively for the first chain, and 0.3935, 0.5106 and 0.7073 for the second chain after 8000 iterations, which clearly suggests that only a crude estimate is available.

### Example 3

There are cases in which the posterior summary of interest may be only the posterior mean with a corresponding confidence interval. These cases are particularly suited to our methodology. For example, take the usual model choice or variable selection approaches dealt with the MCMC algorithm; see for example Green (1995), George and McCulloch (1993). In these models the MCMC output contains a variable, say  $\gamma$ , which expresses the probability of a model or the probability that a variable is included in the model. This variable is a string of 0 or 1, and in stationarity, the mean and confidence intervals of  $\gamma$  are the desired posterior summaries of interest.

A major problem in the linear model theory is the choice of the appropriate set of regressors which explain satisfactorily the variability of the dependent variable. Recently George and McCulloch (1993) developed the Stochastic Search variable Selection (SSVS) which enables the calculation of the posterior probability of inclusion of a regressor.

To illustrate our methodology, we use the example 4.1 of George and McCulloch (1993). There are five regressors  $\mathbf{X}_1, \dots, \mathbf{X}_5 \stackrel{iid}{\sim} \mathbf{N}(0, 1)$  of size  $n = 60$  which are used in two variable selection problems. In Problem 1, the dependent variable is generated according to the model

$$\mathbf{Y} = \mathbf{X}_4 + 1.2\mathbf{X}_5 + \boldsymbol{\varepsilon},$$

where  $\boldsymbol{\varepsilon} \sim \mathbf{N}_{60}(0, \sigma^2 \mathbf{I})$  with  $\sigma = 2.5$ . Problem 2 is identical to Problem 1, apart from the regressor  $\mathbf{X}_3$  which is replaced by  $\mathbf{X}_3^* = \mathbf{X}_4 + 0.15\mathbf{Z}$  where  $\mathbf{Z} \sim \mathbf{N}(0, 1)$ , yielding  $\text{corr}(\mathbf{X}_3, \mathbf{X}_5) = 0.99$ .

For each of the potential models of the above problems we can construct, using the SSVS method, a variable which takes values 1 or 0, depending on whether the particular model is chosen or not in the current iteration. To obtain a sample from these posterior model probabilities we construct an MCMC chain that converges to the posterior distribution of interest. For more details of the above methodology and the example, see George and McCulloch (1993).

Problem 1			Problem 2		
Model	Prob	Accuracy	Model	Prob	Accuracy
$\{\mathbf{X}_4, \mathbf{X}_5\}$	0.23	0.010311	$\{\mathbf{X}_4, \mathbf{X}_5\}$	0.10	0.010450
$\{\mathbf{X}_5\}$	0.16	0.008893	$\{\mathbf{X}_3, \mathbf{X}_4\}$	0.09	0.009580
$\{\mathbf{X}_2, \mathbf{X}_4, \mathbf{X}_5\}$	0.09	0.006071	$\{\mathbf{X}_4\}$	0.08	0.007388
$\{\mathbf{X}_2, \mathbf{X}_5\}$	0.06	0.005508	$\{\mathbf{X}_5\}$	0.07	0.007678
$\{\mathbf{X}_4\}$	0.06	0.005530	$\{\mathbf{X}_3\}$	0.06	0.006669
			$\{\mathbf{X}_3, \mathbf{X}_4, \mathbf{X}_5\}$	0.05	0.009486

Table 4.2: *Posterior model probabilities and their accuracies*

We focus our analysis on the models that have more than 0.05 posterior probability. Using an MCMC chain of 50000 iterations and choosing as threshold value  $d = 0.999$ , we estimate the burn-in period. This comes out to be 4500 and 6000 iterations for Problems 1 and 2 respectively (see Figures 4-1(*d*) and 4-1(*e*)). The resulting model probabilities and corresponding accuracies after 50000 iterations are presented in Table 4.2. The variables in the curly brackets are contained in the resulting models.

Assume that, in a hypothetical scenario, one needs to obtain posterior model probabilities with a required accuracy of 2% of their estimated value. The results of our diagnostic can guide the MCMC to run for a number of iterations such that the above requirement is satisfied. In detail, discarding the burn-in period for each model, there is a strong linear relation between the range of the confidence region and the  $1/\sqrt{N_j}$  as indicated in previous sections. Therefore, a weighted regression between the range, as dependent variable, and  $1/\sqrt{N_j}$ , could give an estimate for the required iterations that are needed to estimate the posterior model probabilities with the desired accuracy. Table 4.3 contains the estimated iterations that an MCMC needs in order to estimate the model probabilities for Problems 1 and 2 with accuracy of 2% of their estimated value. Note that if our desired accuracy was produced in less than 50000 iterations, this could be a significant drawback of our diagnostic: we produced unnecessary iterations to estimate that only a portion of them is needed!

Problem 1		Problem 2	
Model	Iterations	Model	Iterations
$\{\mathbf{X}_4, \mathbf{X}_5\}$	252286	$\{\mathbf{X}_4, \mathbf{X}_5\}$	1150652
$\{\mathbf{X}_5\}$	345571	$\{\mathbf{X}_3, \mathbf{X}_4\}$	1125107
$\{\mathbf{X}_2, \mathbf{X}_4, \mathbf{X}_5\}$	521015	$\{\mathbf{X}_4\}$	883304
$\{\mathbf{X}_2, \mathbf{X}_5\}$	817300	$\{\mathbf{X}_5\}$	1107036
$\{\mathbf{X}_4\}$	855504	$\{\mathbf{X}_3\}$	1359638
		$\{\mathbf{X}_3, \mathbf{X}_4, \mathbf{X}_5\}$	3941288

Table 4.3: *Number of iterations for specific accuracy*



# Chapter 5

## The Auxiliary Variable Sampler

### 5.1 Introduction

During the last decade, statisticians have been focused on the Markov Chain Monte Carlo techniques such as Gibbs (Gelfand and Smith,1990) and Metropolis - Hastings algorithm (Metropolis *et al.*, 1953; Hastings, 1970), in order to sample from the complex multivariate posterior distributions appeared in many statistical problems (Gilks et al.,1995).

Both these algorithms may present difficulties in certain problems. Firstly, the implementation of the Gibbs sampler may lead to the need of sampling from non-standard conditional densities. In the last case, methods to sample from these non-standard densities must be developed, something which is not always possible or efficient. Alternatively, Metropolis-Hastings algorithm must be utilized in order to update the parameters from these non-standard densities. However, the Metropolis-Hastings algorithm presupposes the choice of an appropriate “proposal” density that will lead to efficient sampling and this can be succeeded only with a lot of tuning on the proposal density, for the specific application.

*These major drawbacks of the two popular MCMC algorithms have restrained the*

*development of computer programs that will automatically simulate Markov Chains for Bayesian modelling.*

Lately, a new MCMC algorithm has been proposed, which overcomes the aforementioned problems in many statistical applications. This algorithm is based on the idea of introducing one or more auxiliary variables to the Markov chain Monte Carlo scheme, in such a way that enables the development of a Markov chain that is easier to be simulated. This algorithm is called auxiliary variable algorithm or slice sampler.

The idea of using auxiliary variables to a MCMC scheme was firstly introduced by Swendsen and Wang (1987) for Ising model. Edwards and Sokal (1988), Besag and Green (1993), Higdon (1998), Damien *et al.* (1999) and Neal (2003) demonstrate the usefulness of this technique in a variety of statistical applications.

The theoretical properties of the auxiliary variable sampler have been studied by Fishman (1999), Mira and Tierney (1998) and Roberts and Rosenthal (1999).

## 5.2 The Auxiliary Variable Sampler

As already mentioned, the auxiliary variable sampler technique is based on the use of latent variables in order to develop easy to simulate Markov chains. The application of the auxiliary variable sampler is based on the following Theorem of Damien *et al.* (1999, theorem 1, page 332).

**Theorem 4** *Suppose that we wish to generate random variates from a density  $f$  given by*

$$f(x) \propto \pi(x) \prod_{i=1}^n l_i(x),$$

*where  $\pi$  is a density of known form and  $l_i$  are non-negative invertible functions (not necessarily densities), i.e. if  $l_i(x) > u$  is possible to obtain the set  $A_u^c = \{x : l_i(x) > u\}$ . Then a Gibbs sampler for generating random variates from  $f$  exists, where all except one of the full conditions are uniform densities and the remaining full conditional is a*

truncated version of  $\pi$ .

**Proof:** see Damien et al. (1999).

Based on the above Theorem, Markov chains can be constructed by introducing latent variables in the distribution of interest in such a way that the MCMC algorithm which is comprised entirely on Gibbs steps.

Assume that a sample from the density  $p(\theta|\mathbf{y}) \propto \pi(\theta) l(\mathbf{y}|\theta)$  is required. However, this is not feasible to be accomplished by applying standard methodology. Instead of using Metropolis-Hastings algorithm, a latent variable, say  $u$ , defined on the interval  $(0, \infty)$  is introduced in such a way that the joint density of  $\theta$  and  $u$  will be given by

$$p(\theta, u|\mathbf{y}) \propto \pi(\theta) \mathbf{I}\{u < l(\mathbf{y}|\theta)\}. \quad (5.1)$$

In (5.1) the resulting full conditional densities for  $u$  and  $\theta$  are of standard form and a MCMC scheme can be constructed based upon those full conditional densities. In detail, the MCMC algorithm for this case is given below:

1. Give initial value to  $\theta$ ,  $\theta^0$
2. Sample  $u^1|\theta^0 \sim Uniform(0, l(\mathbf{y}|\theta^0))$
3. Sample  $\theta^1|u^1 \sim \pi(\cdot)$  truncated on the interval  $S = \{\theta : u^1 < l(\mathbf{y}|\theta)\}$ .
4. Repeat from step 2.

## 5.3 Variations of the Auxiliary Variable Sampler

### 5.3.1 Introduction

A wide variety of alternative versions referring to the auxiliary variable sampler have been proposed in related literature. A standard classification scheme for these versions

of auxiliary variable sampler has been proposed by Mira and Tierney (1998) and will be presented in this section. Note that, in the Bayesian framework the distribution of interest is known up to a constant and it is extracted via the Bayes theorem, i.e.

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto \pi(\boldsymbol{\theta}) l(\mathbf{y}|\boldsymbol{\theta}). \quad (5.2)$$

### 5.3.2 The Simple Auxiliary Variable Sampler

This version of the auxiliary variable sampler is fully presented in Damien *et al.* (1999), and it is based on the Theorem (4). One positive latent variable, say  $u$  is introduced in (5.2) such as:

$$p(\boldsymbol{\theta}, u|\mathbf{y}) \propto \pi(\boldsymbol{\theta}) \mathbf{I}\{u < l(\mathbf{y}|\boldsymbol{\theta})\}.$$

Then the MCMC algorithm takes the following form:

1. Give initial value to  $\boldsymbol{\theta}$ ,  $\boldsymbol{\theta}^0$
2. Sample  $u^1|\boldsymbol{\theta}^0 \sim \text{Uniform}(0, l(\mathbf{y}|\boldsymbol{\theta}^0))$
3. Sample  $\boldsymbol{\theta}^1|u^1 \sim \pi(\cdot)$  truncated on the interval  $S = \{\boldsymbol{\theta} : u^1 < l(\mathbf{y}|\boldsymbol{\theta})\}$ .
4. Repeat from step 2.

Note that, the full conditional density of the latent variable  $u$  is uniform and the full conditional density of  $\boldsymbol{\theta}$  is a truncated version of the prior  $\pi(\cdot)$ .

### 5.3.3 The Multiple Auxiliary Variable Sampler

The variation of the auxiliary variable sampler that is presented in this section has been introduced by Edward and Sokal (1988). Suppose that the full posterior density can be

written as:

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{y}) &\propto \pi(\boldsymbol{\theta}) l(\mathbf{y}|\boldsymbol{\theta}) \\ &\propto \pi(\boldsymbol{\theta}) \prod_{i=1}^k g_i(\mathbf{y}|\boldsymbol{\theta}), \end{aligned}$$

then a collection of  $k$  positive latent variables can be introduced,  $\mathbf{u} = (u_1, \dots, u_k)$ , such as the joint density of  $\boldsymbol{\theta}$  and  $\mathbf{u}$  is given by:

$$p(\boldsymbol{\theta}, \mathbf{u}|\mathbf{y}) \propto \pi(\boldsymbol{\theta}) \prod_{i=1}^k \mathbf{I}\{u_i < g_i(\mathbf{y}|\boldsymbol{\theta})\}.$$

Then the MCMC algorithm takes the form:

1. Give initial value to  $\boldsymbol{\theta}$ ,  $\boldsymbol{\theta}^0$
2. Sample  $u_i^1 | \boldsymbol{\theta}^0 \sim \text{Uniform}(0, g_i(\mathbf{y}|\boldsymbol{\theta}^0))$ , for  $i = 1, \dots, k$ .
3. Sample  $\boldsymbol{\theta}^1 | u_k^1, \dots, u_1^1 \sim \pi(\cdot)$  truncated on the interval  
 $S = \{\boldsymbol{\theta} : u_i^1 < g_i(\mathbf{y}|\boldsymbol{\theta}), \text{ for } i = 1, \dots, k\}.$
4. Repeat from step 2.

### 5.3.4 The Product Auxiliary Variable Sampler

As in the previous case of the multiple auxiliary variable sampler, suppose that the full posterior density can be written as:

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{y}) &\propto \pi(\boldsymbol{\theta}) l(\mathbf{y}|\boldsymbol{\theta}) \\ &\propto \pi(\boldsymbol{\theta}) \prod_{i=1}^k g_i(\mathbf{y}|\boldsymbol{\theta}). \end{aligned}$$

Instead of introducing a collection of auxiliary positive variables, only one random variable  $u$  is introduced such as:

$$p(\boldsymbol{\theta}, \mathbf{u}|\mathbf{y}) \propto \pi(\boldsymbol{\theta}) \mathbf{I}\left\{u < \prod_{i=1}^k g_i(\mathbf{y}|\boldsymbol{\theta})\right\}.$$

Then, the MCMC algorithm takes the following form:

1. Give initial value to  $\boldsymbol{\theta}$ ,  $\boldsymbol{\theta}^0$
2. Sample  $u^1|\boldsymbol{\theta}^0 \sim \text{Uniform}\left(0, \prod_{i=1}^k g_i(\mathbf{y}|\boldsymbol{\theta}^0)\right)$ .
3. Sample  $\boldsymbol{\theta}^1|u_1^1 \sim \pi(\cdot)$  truncated on the interval  $S = \left\{\boldsymbol{\theta} : u^1 < \prod_{i=1}^k g_i(\mathbf{y}|\boldsymbol{\theta})\right\}$ .
4. Repeat from step 2.

Note that, the above algorithm is exactly the same with the algorithm for the simple auxiliary variable sampler if we replace  $\prod_{i=1}^k g_i(\mathbf{y}|\boldsymbol{\theta})$  with  $l(\mathbf{y}|\boldsymbol{\theta})$ .

## 5.4 Properties of the Auxiliary Variable sampler

Mira and Tierney (1998), Roberts and Rosenthal (1999) and Fishman (1999) have studied the theoretical properties of the auxiliary variable algorithm and came up with a series of promising results about the ergodicity of the sampler and the quantitative bounds on the total variation distance from stationarity after a given number of iterations. Initially, the Markov chain that is constructed from the application of this sampler is irreducible and aperiodic. These properties have as result that the Markov Chain converges to the distribution of interest as  $n \rightarrow \infty$ , independently from the starting points. Moreover, all the variations of the auxiliary variable sampler are uniformly ergodic under some mild regularity conditions (Mira and Tierney, 1998). The Theorem below, introduced by Mira

and Tierney (1998) presents these regularity conditions for the uniform ergodicity and the rate of convergence  $r_c$  for all versions of auxiliary variable sampler

**Theorem 5** *The main variations of the auxiliary variable sampler are uniform ergodic. In detail,*

1. *The simple auxiliary variable sampler,  $f(\theta, u|\mathbf{y}) \propto \pi(\theta) \mathbf{I}\{u \leq l(\mathbf{y}|\theta)\}$ , is uniform ergodic if  $l(\cdot)$  is a bounded function, and the rate of convergence to stationarity in total variation distance is such that*

$$r \leq \left\{ 1 - h \left[ \sup_{\theta \in \Theta} l(\theta) \right]^{-1} \right\},$$

where

$$h = \int_{\Theta} \pi(\theta) l(\mathbf{y}|\theta) d\theta,$$

2. *The product auxiliary variable sampler  $f(\theta, u|\mathbf{y}) \propto \pi(\theta) \prod_i \mathbf{I}\{u_i \leq l_i(\mathbf{y}|\theta)\}$ , is uniform ergodic if  $\prod_i l_i(\mathbf{y}|\theta)$  is bounded and the rate of convergence to stationarity in total variation distance is such that*

$$r \leq \left\{ 1 - h \left[ \sup_{\theta \in \Theta} \prod_i l_i(\theta) \right]^{-1} \right\},$$

where  $h = \int_{\Theta} \pi(\theta) \prod_i l_i(\mathbf{y}|\theta) d\theta$ ,

3. *The multiple auxiliary variable sampler  $f(\theta, u|\mathbf{y}) \propto \pi(\theta) \mathbf{I}\left\{u \leq \prod_i l_i(\mathbf{y}|\theta)\right\}$ , is uniform ergodic if  $l_i(\mathbf{y}|\theta)$  are bounded functions and the rate of convergence to stationarity in total variation distance is such that*

$$r \leq \left\{ 1 - h \left[ \prod_i \sup_{\theta \in \Theta} l_i(\theta) \right]^{-1} \right\},$$

where

$$h = \int_{\Theta} \pi(\theta) \prod_i l_i(\mathbf{y}|\theta) d\theta.$$

**Proof:** see Mira and Tierney (1998).

In addition to that, Roberts and Rosenthal (1999) proved that all the variations of the auxiliary variable sampler are geometrically ergodic. The following Theorems of Roberts and Rosenthal (1999) established this property for the versions of auxiliary variable sampler

**Theorem 6** Consider the simple auxiliary variable sampler,  $f^*(\theta, u|\mathbf{y}) \propto \mathbf{I}\{u \leq f(\theta|\mathbf{y})\}$ , on a bounded density  $f(\theta|\mathbf{y})$ . Let  $L(u) = \{\theta : f(\theta|\mathbf{y}) \geq u\}$  and  $Q(u)$  for  $m(L(u))$ , where  $m$  is Lebesgue measure. Suppose that the function  $Q(u)$  is differentiable, and that there is a constant  $a > 1$  such that  $Q'(u)u^{1+1/a}$  is non-increasing, at least on an open set containing 0. Then the Markov chain for the simple auxiliary sampler is geometrically ergodic.

**Proof:** see Roberts and Rosenthal (1999).

The above Theorem is referred to the case where  $f(\theta|\mathbf{y})$  is bounded. On the other hand, if  $f(\theta|\mathbf{y})$  is unbounded the geometrically ergodicity is still hold under some conditions that presented below

**Theorem 7** Consider the simple auxiliary variable sampler,  $f^*(\theta, u|\mathbf{y}) \propto \mathbf{I}\{u \leq f(\theta|\mathbf{y})\}$ , on a unbounded density  $f(\theta|\mathbf{y})$  with infinite support. Suppose that the function  $Q(u)$  is differentiable, and that there is a constant  $a > 1$  such that  $Q'(u)u^{1+1/a}$  is non-increasing for  $u$  in an open set containing 0, and furthermore that  $(Q^{-1})'(w)w^{1+1/a}$  is non-increasing for  $w$  in an open set containing 0. Then the Markov chain for the simple auxiliary sampler is geometrically ergodic.

**Proof:** see Roberts and Rosenthal (1999).

**Theorem 8** Consider the product auxiliary variable sampler,  $f(\boldsymbol{\theta}, \mathbf{u}|\mathbf{y}) \propto \prod_i^k \mathbf{I}\{u_i \leq f_i(\boldsymbol{\theta}|\mathbf{y})\}$ . Suppose that, for each  $i$   $f_i(\boldsymbol{\theta}|\mathbf{y})$  is bounded. Set  $Q_1(u) = m(L(u; f_1))$  and suppose that the function  $Q_1$  is differentiable with  $Q_1'(u) u^{1+1/a}$  non-increasing, at least in some open set containing 0. Suppose that, for all  $\varepsilon > 0$ , the set  $\{\mathbf{z} : f_1(\mathbf{z}) \geq \varepsilon\}$  is compact, and for each  $1 \leq i \leq k$  the function  $f_i$  is bounded away from 0 on compact intervals. Finally suppose that,  $f_i(\theta_1) \leq f_i(\theta_2)$  holds, then the Markov chain for the product auxiliary sampler is geometrically ergodic.

**Proof:** see Roberts and Rosenthal (1999).

In addition to that, Roberts and Rosenthal (1999) proved that geometrical ergodicity still holds and for the case of the product auxiliary sampler where  $f_i$  are not all decreasing in the same direction.

The most profound result of the work of Roberts and Rosenthal (1999) is the fact that they have managed to provide rigorous quantitative bounds on the total variation distance from stationarity after a given number of iterations for the case of simple auxiliary sampler:

**Theorem 9** Consider the simple auxiliary variable sampler,  $f^*(\theta, u|\mathbf{y}) \propto \mathbf{I}\{u \leq f(\theta|\mathbf{y})\}$ , on a bounded density  $f(\theta|\mathbf{y})$  such as the function  $Q(u)$  is differentiable and  $uQ'(u)$  is non-increasing. Assume  $E_{p^* \wedge \delta_u}(V) \leq 3$ , that is the expected value of  $V$  under the stochastic minorant (with respect to the ordering  $\leq$ ) of the stationary distribution  $f(\theta|\mathbf{y})$  (normalizing) and the point mass  $\delta_u(\cdot)$ . Then the simple auxiliary variable sampler for  $f$  satisfies

$$\|P^n(\theta, \cdot) - f(\cdot)\| \leq 0.054865 (0.985015)^n (n - 15.7043), \quad n \geq 23,$$

where  $P^n(\theta, \cdot)$  is the transition kernel of the Markov chain.

**Proof:** see Roberts and Rosenthal (1999).

For example, if we run a Markov chain, which is constructed from the simple auxiliary variable sampler for  $n = 530$ , we obtain  $\|P^{530}(\theta, \cdot) - f(\cdot)\| \leq 0.0095$ . Therefore, the total

variation distance to stationarity is less than 1%.

## 5.5 Comparison of the Auxiliary Variable Sampler with the Metropolis-Hastings scheme

A relatively small number of publications is concerned with the direct comparison of the auxiliary variable sampler with the other MCMC schemes. A first approach on this issue was conducted by Damien et al (1999). In this paper, authors compared the auxiliary variable sampler with the independent Metropolis-Hastings algorithm and concluded that auxiliary variable sampler is more efficient. Based on their approach, let  $f(\theta|\mathbf{y}) \propto \pi(\theta) l(\mathbf{y}|\theta)$ , the distribution from which a sample must be drawn. Consider a specific version of the independent Metropolis-Hastings where  $\pi(\cdot)$  is the proposal density. Let  $\theta^{(t)}$  the current point of the MCMC chain. In this case, the Metropolis-Hastings step in order to update the parameter  $\theta$  takes the form:

1. sample  $\theta^{pr}$  from  $\pi(\cdot)$ ,
2. sample  $u$  from  $Uniform(0, 1)$ , if  $u \leq \left( l(\theta^{pr}) / l(\theta^{(t)}) \right)$  accept the new value, i.e.  $\theta^{(t+1)} = \theta^{pr}$ , otherwise  $\theta^{(t+1)} = \theta^{(t)}$ .

Suppose now that the order under which the steps are taking place on the above algorithm is changed. As a consequence of this change, it is  $u$  that must be sampled firstly from  $Uniform(0, 1)$ , while the second step will be to sample  $\theta$  from  $\pi(\cdot) \mathbf{I}(\theta : l(\theta) \geq l(\theta))$ . In detail

1. sample  $u$  from  $Uniform(0, 1)$
2. sample  $\theta$  from  $\pi(\cdot) \mathbf{I}(\theta : l(\theta) \geq l(\theta))$

The second step of the above algorithm is an auxiliary variable step and the Markov chain always proceeds with respect to the Metropolis-Hastings step of the initial algo-

rithm, where the motion is succeeded with probability  $\left(l(\theta^{pr})/l(\theta^{(t)})\right)$ , otherwise the new sampling point is the previous one.

In addition to that, Mira and Tierney (1998) proved that the simple and the product auxiliary variable samplers dominate off the diagonal the independence Metropolis-Hastings algorithm that use as proposal  $\pi(\cdot)$ , where  $f(\theta|\mathbf{y}) \propto \pi(\theta)l(\mathbf{y}|\theta)$ . Moreover, they proved that the transition kernels of the simple and the product auxiliary variable samplers have smaller second largest eigenvalues than these corresponding to the independence Metropolis-Hastings kernel. This result means that the auxiliary variable samplers converge faster to the distribution of interest than the independent Metropolis-Hastings algorithm.

Another result of Mira and Tierney (1998) is that a simple or product auxiliary variable sampler can be set up to have smaller asymptotic variance of the sample paths averages than the specific Metropolis-Hastings algorithm. For example, suppose that a sample is needed to be taken from  $f(\theta|\mathbf{y})$ . Let  $q(\theta)$  the proposal density of the independent Metropolis-Hastings algorithm. Then, by setting  $l(\mathbf{y}|\theta) = f(\theta|\mathbf{y})/q(\theta)$  the distribution of interest can be factorized as  $f(\theta|\mathbf{y}) = q(\theta)l(\mathbf{y}|\theta)$ . The resulting simple auxiliary variable sampler  $f(\theta, u|\mathbf{y}) \propto q(\theta) \mathbf{I}\{u \leq l(\mathbf{y}|\theta)\}$ , has smaller asymptotic variance of the sample paths averages.

## 5.6 Auxiliary Variable Sampler in Bayesian Modeling

Damien *et al.* (1999) provided various examples of using auxiliary variable sampler for non-conjugate and hierarchical models. For example, for the class of non-conjugate models, the Poisson *log*-Normal model is presented. In this case,  $\tau \sim \text{Poisson}(\exp(\lambda))$  is observed. If Normal prior with mean 0 and variance 1 ( $N(0, 1)$ ) is assumed for  $\lambda$ , then

the posterior distribution of  $\lambda$  takes the form:

$$f(\lambda|\tau) \propto \exp\{\tau\lambda - \exp(\lambda)\} \exp\{-0.5\lambda^2\}.$$

A positive latent variable  $u$  can be included such as:

$$f(\lambda, u|\tau) \propto \exp\{-u\} \exp\{-0.5(\lambda^2 - 2\tau\lambda)\} \mathbf{I}\{u > \exp(\lambda)\}.$$

Then the MCMC algorithm for sampling from the posterior density of  $\lambda$  becomes:

1. Give initial value to  $\lambda$ ,  $\lambda^0$
2. Sample  $u^1|\lambda^0 \sim \text{exponential}(1) \mathbf{I}(u > \exp(\lambda^0))$ .
3. Sample  $\lambda^1|u^1 \sim \text{Normal}(\tau, 1) \mathbf{I}(\lambda < \log(u^1))$ .
4. Repeat from step 2.

In the context of hierarchical models Damien *et al.* (1999) proposed algorithms for both generalized linear mixed models and for non-linear mixed models.

In the case of non-linear mixed models we have the following model:

$$y_{ij}|\theta_i, \sigma^2 \sim N(g(\theta_i, x_{ij}), \sigma^2)$$

where  $y_{ij}$  represents the observations,  $i = 1, \dots, n$ ,  $j = 1, \dots, n_i$ ,  $N = \sum_i^n n_i$ ,  $\theta_i$  is the random effect,  $x_{ij}$  are the explanatory variables and  $g(\theta_i, x_{ij})$  is a known non-linear mean response function. For this model the posterior density for  $\theta_i$  takes the form:

$$f(\theta_i, \dots, \theta_n|\mathbf{Y}) \propto \prod_{i=1}^n \left\{ \prod_{j=1}^{n_i} \exp\left\{-\frac{1}{2\sigma^2} (y_{ij} - g(\theta_i, x_{ij}))^2\right\} \right\} \cdot \pi(\theta_i)$$

where  $\pi(\cdot)$  is the prior for  $\theta_i$ .

Due to the non-standard form of the conditionals densities of  $\theta_i$  the model can be rewritten considering a positive latent variable  $u_{ij}$  for each  $y_{ij}$ . Therefore, the model takes the following form:

$$y_{ij}/u_{ij}, \theta_i \sim \text{Uniform} \left( g_{ij}(\theta_i, x_{ij}) - \sqrt{u_{ij}}, g_{ij}(\theta_i, x_{ij}) + \sqrt{u_{ij}} \right),$$

and

$$u_{ij} \sim G \left( \frac{3}{2}, \frac{1}{2\sigma^2} \right).$$

So, the posterior (density) is given by:

$$f(\theta_i, \dots, \theta_n | \mathbf{Y}) \propto \exp \left\{ -0.5 \sum_i \sum_j u_{ij} \right\} \cdot \prod_i \prod_j \{ \mathbf{I}(u_{ij} > (y_{ij} - g(\theta_i, x_{ij}))^2) \} \pi(\theta_i).$$

Based on the above posterior density the full conditionals take the form:

$$u_{ij} | \cdot \sim \text{exponential}(0.5) \mathbf{I}(u_{ij} > (y_{ij} - g(\theta_i, x_{ij}))^2),$$

for  $i = 1, \dots, n$  and  $j = 1, \dots, n_i$

$$\theta_i | \cdot \sim \pi(\cdot) \prod_j^{n_i} \mathbf{I}(u_{ij} > (y_{ij} - g(\theta_i, x_{ij}))^2),$$

for  $i = 1, \dots, n$ .

## 5.7 Main Drawbacks of Auxiliary Variable Sampler

The auxiliary variable sampler is a new MCMC scheme with a lot of useful properties. One of the most important ones, is that enables the construction of easy-to-implement and fast-to-converge Markov chains. On the other hand, this technique has some negative points.

Suppose, that a random sample is needed from the posterior density  $f(\theta|\mathbf{y}) \propto \pi(\theta) \cdot l(\mathbf{y}|\theta)$ . The auxiliary variable sampler can be utilized by applying Theorem 1 of Damien *et al.* (1999) i.e.  $f(\theta, u|\mathbf{y}) \propto \pi(\theta) \cdot \mathbf{I}(u < l(\mathbf{y}|\theta))$ .

In Theorem 1 of Damien *et al.* (1999),  $l(\theta)$  is assumed to be invertible, i.e. if  $l(\mathbf{y}|\theta) > u$  then it is possible to obtain, explicitly, the set  $S = \{\theta : l(\mathbf{y}|\theta) > u\}$ . This specific condition is satisfied in many statistical models, but there are others where all solutions of  $l(\theta) = u$  cannot be tracked analytically. In order to overcome this problem robust numeric methods must be applied but there are cases where the usage of these numerical methods is not feasible.

Another drawback of the auxiliary variable sampler is that does not allow the updating of all parameters or a block of them simultaneously. Suppose that  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$  with posterior density  $f(\boldsymbol{\theta}|\mathbf{y}) \propto \pi(\boldsymbol{\theta}) l(\mathbf{y}|\boldsymbol{\theta})$ . If any version of auxiliary variable sampler is applied then the full conditional density for  $\boldsymbol{\theta}$  is a multivariate density with a set of truncations. Sampling from truncated multivariate densities is not feasible, apart from some special cases, see Robert (1995). In order to overcome this problem, the simultaneous updating of all the parameters  $\boldsymbol{\theta}$  is not applied but instead, each parameter  $\theta_i$  is updated given the rest parameters using the univariate full conditionals  $f(\theta_1|\cdot), f(\theta_2|\cdot), \dots, f(\theta_k|\cdot)$ , which are univariate truncated densities and therefore it is more easy to sample from them.

On the other hand, it is commonly accepted that sampling all or a block of parameters simultaneously leads to faster convergence rates, '*The larger the block that are updated simultaneously the faster the convergence*' (Amit and Grenader, 1991). There are some counterexamples for this statement (Roberts and Sahu, 1997), but it is a common practice to update block of parameters, instead of updating parameters one by one. This approach is used in order to speed up convergence and reduce the correlation between the parameters.

## 5.8 The Metropolis scheme within the Auxiliary Variable Sampler

Suppose that a sample from  $f(\theta|\mathbf{y})$  is needed. For this reason, the simple auxiliary variable sampler scheme is applied by including one positive latent variable  $u$ , such as:

$$f(\theta, u|\mathbf{y}) \propto \mathbf{I}\{u < f(\theta|\mathbf{y})\}.$$

The corresponding MCMC algorithm is:

1. Give initial value to  $\theta$ ,  $\theta^0$ .
2. Sample  $u|\theta^0, \mathbf{y} \sim \text{Uniform}(0, f(\theta|\mathbf{y}))$ .
3. Sample  $\theta^1|u, \mathbf{y} \sim \text{Uniform}(0, \{\theta : f(\theta|\mathbf{y}) \geq u\})$ .
4. Repeat from step 2.

Suppose now, that  $f(\theta|\mathbf{y})$  is not invertible with respect to  $\theta$ . In such case, numerical methods or a Metropolis-Hastings step could be used as Higdon (1998) proposed. Following Higdon (1998) the MCMC algorithm takes the form:

1. Give initial value to  $\theta$ ,  $\theta^0$ .
2. Sample  $u|\theta^0, \mathbf{y} \sim \text{Uniform}(0, f(\theta|\mathbf{y}))$ .
3. Sample  $\theta^{new}$  from a symmetric proposal density  $g(\cdot|\theta)$  and keep  $\theta^{new}$  as a new point  $\theta^1$  if  $f(\theta^{new}|\mathbf{y}) \geq u$  or else the new point is the previous one, i.e.  $\theta^1 = \theta^0$ .

It should be mentioned that, the above algorithm inherits the disadvantage of the Metropolis-Hastings sampler. Therefore, a lot of tuning for the proposal  $g(\cdot|\theta)$  is needed in order to have efficient sampling and the most of the advantages of the auxiliary variable sampler are vanished.

## 5.9 Neal's Approach

Neal's (2003) extensive and innovative study on the auxiliary variable sampler had as an outcome the suggestion of a number of procedures that efficiently enabled researchers to overcome all the aforementioned drawbacks of this method. His research focused on the simple auxiliary sampler and incorporated automated procedures for solving the problem of non-invertibility as defined in section 5.7 and the problem of updating blocks of parameters.

Following Neal (2003), let  $\pi(\theta)$  the density from which a sample must be taken. Adopting the scheme of simple auxiliary variable sampler, one positive latent variable  $u$  is introduced such as:

$$f(\theta, u) \propto \mathbf{I}(u < f(\theta)).$$

The resulting MCMC algorithm takes the form:

1. Give initial value to  $\theta$ ,  $\theta^0$ .
2. Sample  $u|\theta^0 \sim \text{uniform}(0, f(\theta^0))$ . Then a set (slice)  $S = \{\theta : u \leq f(\theta^0)\}$  is defined.
3. Find an interval  $I = (L, R)$  around  $\theta^0$  that contains at least a big part of set  $S$ .
4. Sample  $\theta^1$  from Uniform density defined on  $S \cap I$ .
5. Repeat from step 2.

In case that  $f(\theta)$  is invertible with respect to  $\theta$  then  $L = \inf(S)$  and  $R = \sup(S)$  can be set. The last result is equivalent to the simple auxiliary variable sampler defined by Damien *et al.* (1999).

If  $f(\theta)$  is not invertible, then Neal (2003) suggested to pick randomly an initial interval of size  $w$  that contains  $\theta^0$  and expanding this interval in order to contain a big part of  $S$ . For this expansion, Neal (2003) proposed two procedures, the 'stepping out' procedure and the 'doubling' procedure.

### 5.9.1 The Stepping-out Procedure

Let  $f(\theta|u)$  be the conditional density from which a sample is required. Let  $\theta^k$  denote the current point,  $u$  the latent variable that defines the slice  $S = \{\theta : u \leq f(\theta^k)\}$ ,  $w$  is the interval width,  $m$  an integer that sets the maximum size of the interval equal to  $mw$  and  $I = (L, R)$  the interval that has to be defined, where  $L$  is the lower bound and  $R$  is the upper bound respectively. Then

1. Sample  $z, v \sim \text{Uniform}(0, 1)$ .
2. Set  $L = \theta^k - wz$  and  $R = L + w$ .
3. Set  $K_1 = \text{Floor}(mv)$  and  $K_2 = (m - 1) - K_1$ .
4. Repeat while  $K_1 > 0$  and  $u < f(L|u) : L = L - w$  and  $K_1 = K_1 - 1$ .
5. Repeat while  $K_2 > 0$  and  $u < f(R|u) : R = R + w$  and  $K_2 = K_2 - 1$ .

The initial interval of size  $w$  is posed randomly around  $\theta^0$  and this is essential for the correctness of the method. In the application of the above procedure a rough estimate for  $w$ , the size of the interval, is needed. The size of the interval can be allowed to grow to any size or to a specific size  $mw$ . Of course, the size of the interval can be set always to be  $w$ , but this approach might proven to be inefficient in case that the interval is too small. In case that  $\theta$  is bounded, then the interval can be set to be equal with the region that is defined by  $\theta$ . However, this may be inefficient if the  $S \cap I$  is much smaller than this region.

### 5.9.2 The Doubling Procedure

Let  $f(\theta|u)$  the conditional density from which a sample should be drawn. Let  $\theta^k$  the current point,  $u$  the latent variable that defines the slice  $S = \{\theta : u \leq f(\theta^k)\}$ ,  $w$  is the interval width,  $p$  is an integer that set the maximum size of the interval equal to  $2^p w$  and

$I=(L, R)$  is the interval that has to be defined, where  $L$  is the lower bound and  $R$  is the upper bound respectively. Then

1. Sample  $z \sim Uniform(0, 1)$ .
2. Set  $L = \theta^k - wz$ ,  $R = L + w$ , and  $K = p$ .
3. Repeat while  $K > 0$  and  $\{u < f(L|u) \text{ or } u < f(R|u)\}$  :
  - Sample  $v \sim Uniform(0, 1)$ .
  - If  $v < 0.5$  then  $L = L - (R - L)$  else  $R = R + (R - L)$ .
  - $K = K - 1$ .

As previously described, an initial interval of size  $w$  is assigned randomly around  $\theta^0$  and a rough estimation for  $w$  is needed. This procedure expands the initial interval faster than the stepping out procedure and this could be more efficient if the initial interval of size  $w$  is too small.

### 5.9.3 Discussion

Neal (2003) proved that both procedures in the context of simple auxiliary variable sampler leave the density of interest invariant and the resulting Markov chain ergodic. In case of doubling procedure one test must be done, in order the above properties to be valid. In detail, let  $f(\theta|u)$  the conditional density from which the sample must be extracted. Let  $\theta^k, \theta^{k+1}$  denote the current and the possible next point respectively,  $u$  the latent variable that defines the slice  $S = \{\theta : u \leq f(\theta^k)\}$ ,  $w$  the interval width and let  $I=(L, R)$  be the interval found by the 'doubling' procedure. Then

1. Set  $\hat{L} = L$ ,  $\hat{R} = R$ , and  $B = false$ .
2. Repeat while  $\hat{R} - \hat{L} > 1.1w$  :
  - $M = (\hat{R} + \hat{L})/2$ .

- If  $\{\theta^k < M \text{ and } \theta^{k+1} \geq M\}$  or  $\{\theta^k \geq M \text{ and } \theta^{k+1} < M\}$ , then  $B = true$ .
- If  $\theta^{k+1} < M$  then  $\hat{R} = M$ ; else  $\hat{L} = M$ .
- If  $B = true$  and  $u \geq f(\hat{L}|u)$  and  $u \geq f(\hat{R}|u)$ , then the  $\theta^{k+1}$  is not acceptable.

3.  $\theta^{k+1}$  is acceptable if not rejected in the loop above

This test ensures that the new value of  $\theta$  is sampled in such a way that the above properties are hold. In cases that the conditional distribution is unimodal the above test can be omitted.

By using the 'stepping out' procedure or the 'doubling' procedure an interval  $\mathbf{I}$  that contains a big part of  $S\{\theta : u < f(\theta)\}$  can be found. Then sample points must be repeatedly drawn from a Uniform density defined on  $\mathbf{I}$  till the moment that the first of those sample points is drawn from  $S$ . This sampling procedure might proven to be inefficient in case that  $I$  is much larger than  $S$ . A more efficient method is to sample uniformly from an interval that is initially equal to  $I$  and shrink the interval each time that a new point is sampled out of  $S$ , until a  $\theta$  from  $S$  is sampled. A procedure that shrinks the initial interval  $\mathbf{I}$  has been proposed by Neal (2003) and is presented below.

Let  $f(\theta|u)$  be the conditional density from which a sample must be drawn. Let  $\theta^k$  denote the current point,  $u$  the latent variable that defines the slice  $S = \{\theta : u \leq f(\theta^k)\}$ ,  $w$  the interval width and let  $I = (L, R)$  be this interval. Then,

1. Set  $\hat{L} = L$ ,  $\hat{R} = R$ , and  $B = false$ .
2. Repeat
  - Sample  $z \sim Uniform(0, 1)$
  - Set  $\theta^* = \hat{L} + z(\hat{R} - \hat{L})$ .
  - If  $u < f(\theta^*|u)$  and  $Accept(\theta^*)$  then exit loop.
  - If  $\theta^* < \theta^k$  then  $\hat{L} = \theta^*$  else  $\hat{R} = \theta^*$ .

3.  $\theta^{k+1} = \theta^*$ .

Accept( $\theta^*$ ) means that in the case of doubling procedure the value is accepted from the test above. Up to now experience indicates that the aforementioned shrinkage procedure could be very useful in case that  $\theta$  is bounded. In such cases the initial interval could be set to be the interval that is defined by  $\theta$  and sample from this interval uniformly and shrink it by using the above procedure each time that the new point is not in  $S$ .

## 5.10 The Multivariate Auxiliary Variable Sampler

In many cases it is more preferable to update blocks of parameters instead of updating the parameters one by one. Suppose that  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$  is a vector of  $k$  parameters with posterior density  $f(\boldsymbol{\theta}|\mathbf{y})$  known up to a constant. Our aim in this case is to take sample from  $f(\cdot|\mathbf{y})$ . One way is to use one of the described variations of auxiliary variable sampler. In detail, one Markov Chain could be constructed in a way that will update the elements of  $\boldsymbol{\theta}$  one by one. On the other hand, the following algorithm proposed by Neal (2003) could be used by updating the whole vector  $\boldsymbol{\theta}$  at once.

Specifically, Neal (2003) proposed to include in the density of interest one random variable  $u$  so that

$$f(\boldsymbol{\theta}, u|\mathbf{y}) \propto \mathbf{I}(u < f(\boldsymbol{\theta}|\mathbf{y})).$$

Then the MCMC algorithm takes the form:

1. Give initial values to  $\boldsymbol{\theta}$  such as  $\boldsymbol{\theta}^0 = (\theta_1^0, \dots, \theta_k^0)$ .
2. Sample  $u$  from  $Uniform(0, f(\boldsymbol{\theta}^0|\mathbf{y}))$ . The sampled  $u$  defines the set  $S\{\boldsymbol{\theta} : u < f(\boldsymbol{\theta}|\mathbf{y})\}$  from which we must sample  $\boldsymbol{\theta}$ .
3. Pose randomly a hyperrectangle  $H = (L_1, R_1) \times \dots \times (L_k, R_k)$  around  $\boldsymbol{\theta}^0$ , which preferably contains a big part of  $S$ .
4. Sample uniformly  $\boldsymbol{\theta}$  from  $S \cap H$ .

5. Repeat from step 2.

Ideally,  $H$  to be the smallest hyperrectangle containing  $S$ , but this is unlikely to be feasible. Neal (2003) proposed to use a hyperrectangle around  $\theta^0$ , where the width of each edge  $w_i = R_i - L_i$  is decided by our experience. Furthermore, sample points are repeatedly taken uniformly from this hyperrectangle up to find a point that belongs to  $S$ .

Another approach is to sample uniformly from  $H$  and shrink  $H$  every time that the sampled point does not belong to  $S$ . All the dimensions of the hyperrectangle can be shrunk when the point that it was sampled does not belong to  $S$  or only the axis where the following quantity is maximized can be shrank :

$$Q_i = (R_i - L_i) |G_i|,$$

where  $G$  is the gradient of  $\log f(\theta|\mathbf{y})$  and  $\theta$  is the sampled point.

The above procedure may be time consuming if  $H$  is much greater than  $S$ . Doubling and stepping out procedures, that described previously in the context of one-dimension auxiliary variable sampler, may be appropriate extended to fit in multivariate auxiliary variable sampler.

Multivariate auxiliary variable sampling schemes are in their early stages of development. Therefore, it becomes evident that apart from the above described scheme and a framework for adaptive multivariate auxiliary variable sampling (Neal 2003) which will be presented in a following section, a minimal number of effort has been made in this area.

## 5.11 Other Aspects of the Auxiliary Variable Sampler

### 5.11.1 Partial Decoupling Method

The method of partial decoupling was first proposed by Higdon (1993). In the context of the auxiliary variable sampler the method is introduced by Higdon (1998), and has been proved to be useful for sampling from posterior distributions resulting from binary imaging problems. In detail, the posterior is of the form:

$$f(\theta|\mathbf{y}) \propto \pi(\theta) \prod_k b_k(\theta).$$

Higdon (1998) introduced  $k$  latent parameters,  $\mathbf{u} = (u_1, \dots, u_k)$ , and transforms the density such as:

$$f(\theta, \mathbf{u}|\mathbf{y}) \propto \pi(\theta) \prod_k b_k(\theta)^{1-\delta_k} \cdot \mathbf{I}(u_k \leq b_k(\theta)^{\delta_k}).$$

The marginal density with respect to  $\theta$  is

$$f(\theta|\mathbf{y}) \propto \pi(\theta) \prod_k b_k(\theta).$$

The choice of  $\delta$  is either given by the user or by a decreasing function of data ranging from 1 to 0.

### 5.11.2 The Overrelaxed Auxiliary Variable Sampler

Overrelaxed methods were introduced by Adler (1981), and discussed by Barone and Frigessi (1990) and Neal (2003). Here, instead of sampling a new point from its full conditionals, a point is chosen that is on the opposite side of the mode of the full conditional from the previous point.

In detail, if the full conditional is the Normal density then the new value is chosen as

(Adler 1981):

$$\theta^{new} = \mu + \alpha (\theta^{previous} - \mu) + \sigma (1 - \alpha^2)^{\frac{1}{2}} \varepsilon,$$

where  $\mu, \sigma$  are the conditional mean and standard deviation of  $\theta$ ,  $\varepsilon \sim N(0, 1)$  and  $\alpha > -1$ . This method takes into account the dependencies between variables and moves faster from one part of the distribution to another.

For the case of unimodal distributions Neal (2003) introduced an overrelaxed auxiliary variable scheme. More specifically, let  $f(\theta|\mathbf{y})$  the posterior density from which we want to sample. Introducing one latent variable  $u$  such as  $f(\theta, u|\mathbf{y})$  then the Markov chain Monte Carlo algorithm is given by:

1. Give initial value  $\theta^0$ .
2. Sample  $u$  from  $Uniform(0, f(\theta^0|\mathbf{y}))$ . Then  $S = \{\theta : u < f(\theta|\mathbf{y})\}$  is defined.
3. Find an interval  $I = (L, R)$  that contains the biggest part from  $S$ . In case that  $I$  is bigger than  $S$  then estimate the endpoints of  $S$  using bisection.
4. Set the new value for  $\theta$  as  $\theta^1 = L + R - \theta^0$ .
5. Repeat from step 2.

Note, that the new point must be rejected if it does not belong to  $S$ . Apart from that, the application of bisection in order to locate the endpoints of  $S$  might be proven to be time consuming.

### 5.11.3 The Adaptive Multivariate Auxiliary Variable Sampler

The adaptive multivariate auxiliary variable sampler has been introduced by Neal (2003) and can be viewed as a generalization of the multivariate method that has been described in section 5.10. According to this method, the points that do not belong to  $S$  can be used to guide the selection of the new point. Neal (2003) uses 'crumbs' that lead the selection of the new point inside  $S$ .

In detail, let  $f(\boldsymbol{\theta}|\mathbf{y})$  be the density of interest and  $f(\boldsymbol{\theta}, u|\mathbf{y}) \propto \mathbf{I}(u < f(\boldsymbol{\theta}|\mathbf{y}))$ . The Markov chain Monte Carlo algorithm takes the form:

1. Give initial value to  $\boldsymbol{\theta}$ ,  $\boldsymbol{\theta}^0$ .
2. Sample  $c_1$  from  $g_1(c; \boldsymbol{\theta}^0, u)$ .
3. Sample  $\boldsymbol{\theta}^*$  from  $h_1(\boldsymbol{\theta}^*; u, c_1)$ , where

$$h_1(\boldsymbol{\theta}^*; u, c_1) = \frac{g_1(c_1; \boldsymbol{\theta}^*, u)}{\int g_1(c_1; \boldsymbol{\theta}^*, u) d\boldsymbol{\theta}^*}$$

4. If  $\boldsymbol{\theta}^*$  in  $S$  then  $\boldsymbol{\theta}^1 = \boldsymbol{\theta}^{1*}$  and go to step 9, else
5. Sample  $c_2$  from  $g_2(c_1; \boldsymbol{\theta}^0, u, c_1, \boldsymbol{\theta}^{*1})$
6. Sample  $\boldsymbol{\theta}^{*2}$  from  $h_2(\boldsymbol{\theta}; u, c_1, \boldsymbol{\theta}^{*1})$  where

$$h_2(\boldsymbol{\theta}; u, c_1, \boldsymbol{\theta}^{*1}) = \frac{g_1(\cdot) g_2(\cdot)}{\int g_1(\cdot) g_2(\cdot) d\boldsymbol{\theta}^*}$$

7. If  $\boldsymbol{\theta}^{*2}$  in  $S$  then  $\boldsymbol{\theta}' = \boldsymbol{\theta}^{*2}$  and go to 9, else ...
8. Repeat from step 2.

The distributions  $g_1(\cdot), g_2(\cdot), g_3(\cdot), \dots$ , may depend on  $\theta$  and  $u$ . A simple choice for  $g_i(\cdot)$  could be the Normal density with mean  $\boldsymbol{\theta}^0$  and variance-covariance matrix  $\Sigma = \sigma^2 \mathbf{I}_{matrix}$ , where  $\mathbf{I}_{matrix}$  is the identity matrix. Then, the distribution  $h_i$  is  $\mathbf{N}(\bar{c}, (\sigma^2/i) \mathbf{I}_{matrix})$ , where

$$\bar{c} = \frac{c_1 + c_2 + \dots + c_i}{i}.$$

As a result, the larger the number of points that are rejected the narrower the density from which sample is drawn it becomes. Note that, above procedure leaves the density of interest invariant (Neal, 2003).

# Chapter 6

## A Bayesian Approach to Univariate Stochastic Volatility Models

### 6.1 Introduction

Univariate stochastic volatility models have been under extensive investigation by many authors during the last decades. From the view of classical statistics, Taylor (1986) used a likelihood-based approach to estimate the parameters of the stochastic volatility model, Melino and Turnbull (1990) and Andersen *et al.* (1999) used the method of moments, while Nelson (1988), Harvey, Ruiz and Shephard (1994) Ruiz (1994) and Kim, Shephard and Chib (1998) were based on quasi-maximum likelihood estimator. From the Bayesian statistics point of view, a number of MCMC algorithms have been proposed that converge to the posterior density of the parameters of stochastic volatility model. The existing MCMC algorithms can be separated in two main categories. The single move MCMC algorithms that update parameters one by one, and the block move MCMC algorithms that update all or a block of parameters simultaneously. This chapter will be devoted on the Bayesian perspective presenting at the same time the existing algorithms used in the case of univariate stochastic volatility model.

## 6.2 Existing MCMC Algorithms for the Stochastic Volatility Model

### 6.2.1 Singe-Move Algorithms

Many single move MCMC algorithms, for sampling from the posterior density of the stochastic volatility model, have been proposed in the literature. The main difference of those algorithms can be traced on the way that are updating the volatilities  $h_t$ . Note, that under conjugate priors the rest parameters of the stochastic volatility model have full posterior densities of standard form.

Initially, Jacquier, Polson and Rossi (1994) proposed a MCMC single move algorithm for the case of stochastic volatility model based on the following form of the model

$$\begin{aligned} y_t | h_t &\sim N(0, h_t), \\ h_t | a, b, \sigma_h^2 &\sim LN(a + b \log h_{t-1}, \sigma_h^2), \end{aligned}$$

where the volatilities follow LogNormal distribution  $LN(\cdot, \cdot)$ . Based on standard conjugate prior  $\pi(\mathbf{b}, \sigma_h^2) = \pi(\mathbf{b} | \sigma^2) \pi(\sigma_h^2)$ , where  $\mathbf{b} = (a, b)$ ,  $\mathbf{b} | \sigma^2 \sim MN(\mathbf{b}^*, \sigma^2 A^{-1})$  and  $\sigma^2 \sim IG(n_0, s_0^2)$ , Jacquier, Polson and Rossi (1994) construct the following MCMC algorithm

1. Give initial values to the parameters  $\mathbf{b}^0, \sigma_h^{2^0}, h_1^0, \dots, h_T^0$ .
2. Sample  $\mathbf{b}^1 | \sigma_h^{2^0}, h_1^0, \dots, h_T^0 \sim MN(\cdot, \cdot)$ .
3. Sample  $\sigma_h^{2^1} | \mathbf{b}^1, h_1^0, \dots, h_T^0$  from  $\sim IG(\cdot, \cdot)$ .
4. Sample  $h_t^1 \sim f(h_t | \mathbf{h}_{/t}, \mathbf{b}^1, \sigma_h^{2^1})$ , where  $\mathbf{h}_{/t} = (h_1, \dots, h_{t-1}, h_{t+1}, \dots, h_T)$ , for  $t = 1, \dots, T$ .
5. Repeat from step 2.

The above algorithm is consisted of Gibbs steps apart from step (4) where the full conditional density is not of standard form. More specifically, the full conditional density for the volatilities is

$$\begin{aligned}
f(h_t | \mathbf{h}_{/t}, \mathbf{b}, \sigma_h^2) &\propto f(h_t | h_{t-1}, h_{t+1}, \mathbf{b}, \sigma_h^2) \\
&\propto f(y_t | h_t) f(h_t | h_{t-1}) f(h_{t+1} | h_t, ) \\
&\propto h_t^{-0.5} \exp \left\{ -0.5 y_t^2 / h_t \right\} h_t^{-1} \exp \left\{ -\frac{(\ln h_t - m_t)^2}{2s^2} \right\},
\end{aligned} \tag{6.1}$$

where

$$\begin{aligned}
m_t &= \frac{a(1-b) + b(\ln h_{t-1} + \ln h_{t+1})}{1+b^2}, \\
s^2 &= \frac{\sigma_h^2}{1+b^2}.
\end{aligned}$$

As a solution, Jacquier, Polson and Rossi (1994) proposed a Metropolis-Hastings step to update the volatilities  $h'_t s$ . As a proposal density they have chosen an Inverse Gamma density. Jacquier, Polson and Rossi (1994) noticed that the full conditional (6.1) is consisted of two terms, an Inverse Gamma term and a LogNormal term. They approximated the LogNormal term by matching the first and second moments of a LogNormal distribution to the corresponding moments of an Inverse Gamma distribution. The result was that the remaining Inverse Gamma term and the new Inverse Gamma term (which is the result of the approximation of the LogNormal term) can be combined to an Inverse Gamma density with parameters

$$\begin{aligned}
n_1 &= \frac{1 - 2 \exp(s^2)}{(1 - \exp(s^2)) + 0.5}, \\
s_1 &= (n_1 - 1) \exp(m_t + 0.5s^2) + 0.5y_t^2.
\end{aligned}$$

The proposed MCMC algorithm of Jacquier, Polson and Rossi (1994) using

as proposal  $IG(n_1, s_1)$  for updating the volatilities is given bellow

1. Give initial values to the parameters  $\mathbf{b}^0, \sigma_h^{2^0}, h_1^0, \dots, h_T^0$
2. Sample  $\mathbf{b}^1 | \sigma_h^{2^0}, h_1^0, \dots, h_T^0 \sim MN(\cdot, \cdot),$
3. Sample  $\sigma_h^{2^1} | \mathbf{b}^1, h_1^0, \dots, h_T^0 \sim IG(\cdot, \cdot),$
4. Sample  $h_t^* \sim IG(n_1, s_1)$  and  $u \sim Uniform(0, 1)$ . If  $u < q(h_t^0 \rightarrow h^*)$  then  $h_t^1 = h^*$ ; else  $h_t^1 = h_t^0$ , for  $t = 1, \dots, T$
5. Repeat from step 2.

Jacquier, Polson and Rossi (1994) stated that the acceptance probability in step 4 of their algorithm is between 70% and 80%. Giakoumatos (1997) proposed a new candidate density for the Metropolis-Hastings step in order to update volatilities  $h'_t$ s. In detail, he proposed the use of a dependent Metropolis-Hastings step with proposal density the Normal distribution. Step 4 of the late algorithm become as following:

- Sample  $h_t^* \sim N\left(h_t^0, \frac{\sigma_h^2}{(1+b^2)}\right)$  and  $u \sim Uniform(0, 1)$ . If  $u < q(h_t^0 \rightarrow h^*)$ , then  $h_t^1 = h^*$ ; else  $h_t^1 = h_t^0$ , for  $t = 1, \dots, T$ .

Moreover, Jacquier, Polson and Rossi (1999)- based on the above algorithm - proposed Metropolis within Gibbs algorithms for other extensions of the univariate stochastic volatility model.

Pitt (1997) argued that the MCMC algorithm of Jacquier, Polson and Rossi is not so efficient in the way that updates volatilities and for this reason proposed a different way of updating. Pitt (1997) concentrated on the following form of the stochastic volatility model

$$\begin{aligned}
y_t|\cdot &\sim N\left(0, z^2 \exp(h_t)\right), \\
h_t|\cdot &\sim N\left(b * h_{t-1}, \sigma_h^2\right), \\
h_1|\cdot &\sim N\left(0, \frac{\sigma_h^2}{1-b^2}\right).
\end{aligned}$$

In order to update the log-volatilities, Pitt (1997) used the following Theorem that has been proposed in the context of non-Gaussian State Space models. Note that the stochastic volatility model is a member of the class of non-Gaussian state space models.

Let the non-Gaussian state space model

$$\begin{aligned}
y_t|\cdot &\sim f(y_t|c_t + Z_th_t), \\
h_{t+1}|\cdot &\sim N(d_t + T_th_t, A), \\
h_1|Y_0 &\sim N(h_{1|0}, P_{1|0}).
\end{aligned}$$

Pitt (1997) and Carlin, Polson and Stoffer (1992) observed that the full conditional density for  $h_t$  can be written as

$$f(h_t|\mathbf{y}, \mathbf{h}_{/t}) \propto f(y_t|h_t)f(h_t|h_{t-1}, h_{t+1}),$$

where  $\mathbf{h}_{/t} = (h_1, \dots, h_{t-1}, h_{t+1}, \dots, h_T)$ . The conditional prior density  $f(h_t|h_{t-1}, h_{t+1})$  dominates the likelihood  $f(y_t|h_t)$ . Therefore, Pitt (1997) focused on the approximation of this density contrary to Jacquier, Polson and Rossi (1994), who focused on  $f(y_t|h_t)$ , in their algorithm for the stochastic volatility model and developed a rather inefficient MCMC.

The conditional prior density  $f(h_t|h_{t-1}, h_{t+1})$  of the above non-Gaussian state space model follows  $N(\mu_t, S_t)$ , where  $\mu_t$  is a linear combination of  $h_{t-1}$  and  $h_{t+1}$ . Let the

log-likelihood  $l(h_t) = \log f(y_t|h_t)$  and

$$\begin{aligned} l'(h_t) &= \frac{\partial l(h_t)}{\partial h_t}, \\ l''(h_t) &= \frac{\partial^2 l(h_t)}{\partial h_t \partial h_t'}. \end{aligned}$$

The following theorem of Pitt (1997) can be used to find an approximate density for  $f(h_t|\mathbf{y}, \mathbf{h}_{/t})$

**Theorem 10** *Suppose  $S_t$  is non-singular and that  $l''(h_t)$  is negative semi-defined for all values of  $h_t$ . Then, the following two results hold.*

1. *We can sample from  $h_t|h_{t-1}, h_{t+1}, y_t$  by making suggestions  $h_t$  from*

$$N\left(\mu_t + S_t l'(m_t), S_t\right), \quad (6.2)$$

*which are accepted with probability*

$$\exp\left\{l(h_t) - l(m_t) - l'(m_t)^T (h_t - m_t)\right\}, \quad (6.3)$$

*whatever the value of  $m_t$ .*

2. *The probability of rejecting the suggestion made in (6.2) is minimized by selection of  $m_t$  as the mode of  $f(h_t|h_{t-1}, h_{t+1}, y_t)$ .*

The above Theorem can be used in the case of the stochastic volatility model in order to approximate the non-standard density

$$\begin{aligned} f(h_t|\cdot) &\propto f(y_t|h_t) f(h_t|h_{t+1}, h_{t-1}) \\ &\quad \frac{\exp(-0.5h_t)}{z} \exp\left(-\frac{1}{2z^2}y_t^2 \exp(-h_t)\right) \\ &\quad \exp\left\{-\frac{1}{2\sigma_h^2}[(h_t - bh_{t-1})^2 + (h_{t+1} - bh_t)^2]\right\}. \end{aligned}$$

The prior conditional density is

$$\begin{aligned} f(h_t|h_{t+1}, h_{t-1}) &\propto \exp \left\{ -\frac{1}{2\sigma_h^2} [(h_t - bh_{t-1})^2 + (h_{t+1} - bh_t)^2] \right\} \\ &\propto \exp \left( -\frac{1+b^2}{2\sigma_h^2} (h_t - \mu_t)^2 \right), \end{aligned}$$

where  $\mu_t = b(h_{t+1} + h_{t-1}) / (1 + b^2)$ . The log-likelihood takes the form

$$l(h_t) = \log f(y_t|h_t) \propto -0.5h_t - \frac{y_t^2}{2z^2} \exp(-h_t).$$

Applying the results of the previous theorem we obtain

$$\begin{aligned} l'(h_t) &= \frac{y_t^2}{2z^2} \exp(-h_t) - 0.5, \\ l''(h_t) &= -\frac{y_t^2}{2z^2} \exp(-h_t) < 0. \end{aligned}$$

Therefore,  $h_t|\cdot$  can be drawn by sampling from  $N\left(\mu_t^*, \frac{\sigma_h^2}{1+b^2}\right)$ , where

$$\mu_t^* = \mu_t + 0.5 \frac{\sigma_h^2}{1+b^2} \left\{ \frac{y_t^2}{2z^2} \exp(-\mu_t) - 1 \right\}.$$

The probability of acceptance is approximately

$$1 + \frac{y_t^2}{4z^2} \exp(-\mu_t) \left\{ \frac{\sigma_h^2}{1+b^2} + (\mu_t - \mu_t^*)^2 \right\}.$$

Pitt (1997) states that the probability of acceptance is over 99% for most of the financial datasets.

Moreover Pitt (1997) proposes a pseudo-dominating Metropolis-Hastings step (Tierney, 1994) in order to update the log-volatilities  $h'_t$ s. The proposal density for the

Metropolis-Hastings step is  $N(m_t, \sigma_t^{*2})$ , where

$$m_t = \frac{\sigma_t^{*2}}{\sigma_t^2} \mu_t + \frac{\sigma_t^{*2}}{2} \left\{ \frac{y_t^2}{z^2} \exp(-\mu_t) (1 + \mu_t) - 1 \right\},$$

and

$$\sigma_t^{*2} = \frac{1 + b^2}{\sigma_h^2} + \frac{y_t^2}{2z^2} \exp(-\mu_t).$$

In this case we sample  $h_t^*$  from  $N(m_t, \sigma_t^{*2})$  and we accept it as a new value from  $f(h_t|\cdot)$  with probability

$$\min \left[ \frac{f(h_t^*|\cdot) \min \{f(h_t|\cdot), g(h_t)\}}{f(h_t|\cdot) \min \{f(h_t^*|\cdot), g(h_t^*)\}}, 1 \right],$$

where  $g(z) \equiv N(m_t, \sigma_t^{*2})$ .

Apart from the way that volatilities are being updated, Pitt and Shephard (1999c) proved that the analytical convergence rate for the volatilities is

$$\frac{4b^2}{\left\{ \frac{1+b^2+\sigma_h^2}{\text{Var}(\log \varepsilon_t^2)} \right\}^2}.$$

### 6.2.2 Block-Move Algorithms

Apart from the single move algorithms, some authors are proposing algorithms that are updating a block or all the parameters of the stochastic volatility model simultaneously. Regarding the benefits of updating a block of parameters in the Bayesian analysis see Smith and Roberts, (1993) and Liu, Wong and Kong (1994) among others. In the context of time-series models, Fruhwirth-Schnatter (1994) and de Jong and Shephard (1995) have suggested a number of block move algorithms. In this context, an interest approach was introduced by Pitt (1997). A fixed number - knots - of log-volatilities are randomly chosen to remain fixed for one sweep of the MCMC algorithm. Let  $k$  knots are

selected  $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \dots, \kappa_k)$  corresponding to log-volatilities  $\mathbf{h}_/ = (h_{\kappa_1}, h_{\kappa_2}, \dots, h_{\kappa_k})$ . The idea is to update simultaneously all the log-volatilities between two of the selected knots. Pitt (1997) employed a Taylor expansion on the logarithm of the conditional density of log-volatilities between two knots

$$\log f(\mathbf{h}_{t,k} | h_{t-1}, h_{t+k+1}, y_t, \dots, y_{t+k}),$$

where  $\mathbf{h}_{t,k} = (h_t, h_{t+1}, \dots, h_{t+k})$ . The Taylor expansion took place around some preliminary estimates, say  $\hat{\mathbf{h}}_{t,k} = (\hat{h}_t, \hat{h}_{t+1}, \dots, \hat{h}_{t+k})$  of  $\mathbf{h}_{t,k}$ . This expansion leads us to a multivariate Normal density that can be used as proposal for sampling  $\mathbf{h}_{t,k}$ . As preliminary estimates  $\hat{\mathbf{h}}_{t,k}$  can be used the mode of the  $f(\mathbf{h}_{t,k} | h_{t-1}, h_{t+k+1}, y_t, \dots, y_{t+k})$ . The desired mode can be calculated using the Newton-Raphson method.

As a consequence of the above method, a block of  $h'_t$ s can be updated by using Metropolis-Hastings step where the proposal is the multivariate Normal density resulting from the Taylor expansion.

Pitt (1997) offered results by using the above block sampling for log-volatilities - for some optimal number of knots - where the MCMC algorithm converges faster than single-move algorithms.

Similarly, Taylor expansion can be applied to the full posterior density of the parameters of the stochastic volatility model and to lead us to a multivariate Normal density. This multivariate Normal density is used by Pitt (1997) as a proposal to a Metropolis-Hastings step that updates all the parameters and the log-volatilities of the stochastic volatility model at once.

Kim, Shephard and Chib (1998) proposed another way to sample the log-volatilities on a simultaneous manner. Their offset mixture method based on the feature of the stochastic volatility model that can be transformed into a linear model if the logarithm

of the squares of  $y_t$  observations is taken. In detail

$$\begin{aligned} y_t &= \exp\left(\frac{h_t}{2}\right) \varepsilon_t \Rightarrow \\ \log y_t^2 &= h_t + \log \varepsilon_t^2, \end{aligned}$$

where  $Var(\log y_t^2) = 4.93$ .

Kim, Shephard and Chib (1998) approximated the term  $\log \varepsilon_t^2$  using a mixture of 7 Normal densities with component probabilities  $\pi_1, \pi_2, \dots, \pi_k$ , means  $\mu_i - 1.2704$ , for  $i = 1, \dots, 7$  and variances  $\sigma_1^2, \sigma_2^2, \dots, \sigma_7^2$ . As a result, the transformed stochastic volatility model takes the form

$$\begin{aligned} y_t^* | h_t, \omega_t &= i, \mu_i, \sigma_i^2 \sim N(h_t + \mu_i - 1.2704, \sigma_i^2), \\ h_t | a, b, h_{t-1}, \sigma_h^2 &\sim N(a + bh_{t-1}, \sigma_h^2), \\ p(\omega_t = i) &= \pi_i, \end{aligned}$$

where  $y_t^* = \log(y_t^2 + c)$  and  $c$  is an offset that is set by the authors  $c = 0.001$ . The constants  $\{\pi_i, \mu_i, \sigma_i^2; i = 1, \dots, 7\}$  are selected to closely approximate the density  $\log \varepsilon_t^2$  via a non-linear least squares program. The following table contains the values of the constants that Kim, Shephard and Chib (1998) proposed.

$\omega$	$p(\omega)$	$\mu_i$	$\sigma_i$
1	0.00730	-10.12999	5.79596
2	0.10556	-3.97281	2.61369
3	0.00002	-8.56686	5.17950
4	0.04395	2.77786	0.16735
5	0.34001	0.61942	0.64009
6	0.24566	1.79518	0.34023
7	0.25750	-1.08819	1.26261

Based on the above mixture of Normal densities, the log-volatilities can be sampled

simultaneously from a multivariate Normal density. A new step is added in the algorithm to sample the component of the normal mixture, which is straightforward. Finally, a reweighting must be done on the MCMC sample so as to follow the exact posterior density (see Kim, Shephard and Chib, 1998).



# Chapter 7

## Auxiliary Variable Sampler and Univariate Time-Varying Volatility Models

### 7.1 Introduction

Many financial time series, such as stock returns and exchange rates, can be successfully modeled by assuming that the error variance fluctuates over time. Thus, time-varying volatility models can capture a usual phenomenon, common in financial time series, the “volatility clustering”. The familiar modeling approaches are the Autoregressive Conditional Heteroskedasticity (ARCH) models and their variants (Bollerslev, Engle and Nelson 1994). An alternative to those models is given by the Stochastic Volatility models. As a reference for recent descriptions of the stochastic volatility models see: Shephard (1996) and Ghysels, Harvey and Renault (1996). Another model that draw attention was the Unobserved ARCH model (Shephard, 1996), which belongs to a class of models, introduced by Harvey, Ruiz and Sentana (1992). The unobserved ARCH model can be classified as parameter-driven model, therefore, it is a stochastic volatility model.

In this chapter, Bayesian inference is adopted for the aforementioned time-varying

volatility models. In addition, new easy to implement MCMC algorithms are proposed. To achieve that, the Auxiliary Variable sampler is employed (Swendsen and Wang 1987) and sets of latent variables are included in the full posterior density of each model such as all the full conditional densities of the parameters of interest to be of known forms. Still, in certain cases (unobserved ARCH model and stochastic volatility model), the parameter space is transformed so as the resulting full conditional posterior densities are simplified. The proposed MCMC algorithms are consisting of *only* Gibbs steps. This emanated from the need to develop algorithms which are straightforward to use (unlike Metropolis-Hastings, Gibbs sampling requires no tuning) and have better convergence behavior than existing MCMC samplers.

## 7.2 Stochastic volatility model

### 7.2.1 Introduction

In this section, full Bayesian analysis of the stochastic volatility model (Taylor 1982, Shephard 1996) is presented. The form of the stochastic volatility model is given by the following hierarchical structure

$$\begin{aligned} \mathbf{y}_t | \cdot &\sim N(0, ah_t), \\ h_t | \cdot &\sim LN(\beta \log h_{t-1}, \sigma^2). \end{aligned} \tag{7.1}$$

where  $y_t$  is the realization of the stochastic process at time  $t$ ,  $h_t$ ,  $t = 1, \dots, T$  are the volatilities,  $a, \beta$  and  $\sigma^2$  are the hyperparameters of the model and  $LN(\cdot, \cdot)$  is the Log-Normal distribution. Restriction,  $0 < \beta < 1$  is imposed such as the series of volatilities to be covariance stationary. This form of the model leads to a MCMC algorithm where some of the full conditionals are not of standard form (see Jacquier *et al.*, 1994; Giakoumatos 1997; Pitt 1997).

In the following section, the model (7.1) is written by the inclusion  $T$  latent variables

$\mathbf{u} = (u_1, \dots, u_T)$  and a MCMC algorithm which is based only on Gibbs steps is proposed.

### 7.2.2 Bayesian Approach and the Auxiliary Variable Sampler

The model (7.1) can be rewritten using the following Theorem of the decomposition of the LogNormal density.

**Theorem 11** *Suppose that  $h_t|u_t$  follows  $f(\cdot)$ , where*

$$f(h_t|u_t) = \frac{1}{2h_t\sqrt{u}} I_{A_t},$$

where

$$A_t = \{h_t : \exp(\beta \log h_{t-1} - \sqrt{u}) \leq h_t \leq \exp(\beta \log h_{t-1} + \sqrt{u})\}$$

and  $u$  follows  $G\left(\frac{3}{2}, \frac{1}{2\sigma^2}\right)$ . Then the marginal density of  $h_t$  is the LogNormal density with parameters  $\beta \log h_{t-1}$  and  $\sigma^2$ .

**Proof.** Just integrate out the latent variable  $u_t$ . ■

Utilizing the above Theorem, the stochastic volatility model of (7.1) can be written as

$$\begin{aligned} y_t|\cdot &\sim N(0, ah_t), \\ f(h_t|u_t, \beta, h_{t-1}, \sigma^2) &= \frac{1}{2h_t\sqrt{u}} I_{A_t}, \text{ for } t = 1, \dots, T, \\ u_t|\cdot &\sim G\left(\frac{3}{2}, \frac{1}{2\sigma^2}\right), \text{ for } t = 1, \dots, T, \end{aligned} \tag{7.2}$$

where

$$A_t = \{\exp(\beta \log h_{t-1} - \sqrt{u_t}) \leq h_t \leq \exp(\beta \log h_{t-1} + \sqrt{u_t})\}. \tag{7.3}$$

Using non-informative priors for the parameters  $a, \beta, \sigma^2$ , i.e.  $\pi(a, \beta, \sigma^2) \propto (a\sigma^2)^{-1}$  and LogNormal prior for the  $h_0$ , i.e.  $\pi(h_0) \propto (h_0)^{-1} \exp\left\{-0.5\left(\frac{\ln h_0}{v}\right)^2\right\}$ , the joint posterior distribution -known up to a constant- for the parameters of interest takes the following form:

$$\begin{aligned}
f(\beta, \sigma^2, a, h_0, \mathbf{h}, \mathbf{u} | \mathbf{y}) &\propto \left( a^{(T/2)+1} \sigma^{2\frac{3T}{2}+1} h_0 \prod_{t=1}^T \left( h_t^{3/2} \right) \right)^{-1} \\
&\exp \left\{ -\frac{1}{2} \sum_{t=1}^T \frac{y_t^2}{a h_t} \right\} \exp \left\{ -0.5 \left( \frac{\ln h_o}{v} \right)^2 \right\} \\
&\exp \left\{ -\frac{1}{2} \sum_{t=1}^T \frac{u_t}{2\sigma^2} \right\} \prod_{t=1}^T I_{A_t}.
\end{aligned} \tag{7.4}$$

where  $I_{A_i}$  is defined in (7.3).

The full conditional densities for the parameters of interest, based on the posterior (7.4) is given below

- $\sigma^2 | \cdot \sim IG \left( \frac{3T}{2}, 0.5 \sum_{t=1}^T u_t \right).$
- $a | \cdot \sim IG \left( \frac{T}{2}, 0.5 \sum_{t=1}^T \frac{y_t^2}{h_t} \right).$
- $u_t | \cdot \sim \exp \left( \frac{1}{2\sigma^2} \right) I_{A_u}$ , where  $\exp(\cdot)$  is the exponential density,

$$A_u = \{u_t : u_t \geq (\log h_t - \beta \log h_{t-1})^2\},$$

for  $t = 1, \dots, T$ .

- Marginalizing out (Chib and Carlin 1999) the latent parameters  $\mathbf{u}$  the full conditional of  $\beta$  is  $N(\mu, s^2) I(0, 1)$ , where

$$\begin{aligned}
\mu &= \frac{\sum_{t=1}^T \log h_{t-1} \log h_t}{\sum_{t=1}^T \log h_{t-1}^2} \\
s^2 &= \frac{\sigma^2}{\sum_{t=1}^T \log h_{t-1}^2}.
\end{aligned}$$

- Marginalizing out (Chib and Carlin 1999) the latent parameter  $u_1$  the full conditional of  $h_0$  is  $LN(\mu, s^2)$ , where

$$\begin{aligned}\mu &= \frac{\beta \ln(h_1) v^2}{\sigma^2 + \beta v^2} \\ s^2 &= \frac{\sigma^2 v^2}{\sigma^2 + \beta v^2}.\end{aligned}$$

- The full conditional density of the volatilities  $\mathbf{h}$  is given by

$$f(\mathbf{h}|\cdot) \propto \frac{1}{\prod_{t=1}^T h_t^{3/2}} \exp \left\{ -\frac{1}{2a} \sum_{t=1}^T \frac{y_t^2}{h_t} \right\} \prod_{t=1}^T I_{A_t}. \quad (7.5)$$

In order to handle the problem that the full conditional density of volatilities  $\mathbf{h}$  is not of standard form, some *non-linear* transformations of the volatilities are applied. Firstly, note that

**Theorem 12** *If in the posterior density  $f(\mathbf{h}|\cdot)$  defined in (7.5) we perform the following transformations  $k_t = 1/\sqrt{h_t}$ ; for  $t = 0, \dots, T$ , the posterior density takes the form*

$$[\mathbf{k}|\cdot] \propto \exp \left\{ -\frac{1}{2a} \sum_{t=1}^T y_t^2 k_t^2 \right\} \prod_{t=1}^T I_{A_t^*}, \quad (7.6)$$

where

$$A_t^* = \left\{ \exp \left( 0.5 \left( \beta \log k_{t-1}^2 - \sqrt{u_t} \right) \right) \leq k_t \leq \exp \left( 0.5 \left( \beta \log k_{t-1}^2 + \sqrt{u_t} \right) \right) \right\}, \quad (7.7)$$

and  $\mathbf{k} = (k_1, \dots, k_T)$ .

**Proof.** Note that the Jacobian of the above transformations is  $|J| \propto \prod_{t=1}^T k_t^{-3}$ . The remaining calculations are straightforward. ■

Applying the previous Theorem the full conditional for the vector  $\mathbf{k} = (k_1, \dots, k_T)$  is truncated  $T$ - variate Normal density,  $MN_T(\boldsymbol{\mu}, \mathbf{S}) \prod_{t=1}^T I_{A_t^*}$ , where  $\boldsymbol{\mu} = (0, \dots, 0)'$ ,

$$\mathbf{S} = \begin{bmatrix} \frac{a}{y_1^2} & 0 & \cdots & 0 \\ 0 & \frac{a}{y_2^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{a}{y_T^2} \end{bmatrix},$$

and  $A_t^*$  is give in (7.7).

Because of the difficulty to sample from the above truncated multivariate Normal distribution we sample from the univariate density of each  $k_t$ ,  $t = 1, \dots, T$ . In detail

- $k_1 | \cdot \sim N\left(0, \frac{a}{y_1^2}\right) I_{(L,U)}$ , where

$$L = \begin{cases} \max \left\{ 0, \exp \left( -0.5 \left( \beta \log h_0 + \sqrt{u_1} \right) \right), \exp \left( \frac{1}{2\beta} \left( \log k_2^2 - \sqrt{u_2} \right) \right) \right\}, & \text{if } \beta > 0 \\ \max \left\{ 0, \exp \left( -0.5 \left( \beta \log h_0 + \sqrt{u_1} \right) \right), \exp \left( \frac{1}{2\beta} \left( \log k_2^2 + \sqrt{u_2} \right) \right) \right\}, & \text{if } \beta < 0 \end{cases},$$

$$U = \begin{cases} \min \left\{ \exp \left( -0.5 \left( \beta \log h_0 - \sqrt{u_1} \right) \right), \exp \left( \frac{1}{2\beta} \left( \log k_2^2 + \sqrt{u_2} \right) \right) \right\}, & \text{if } \beta > 0 \\ \min \left\{ \exp \left( -0.5 \left( \beta \log h_0 - \sqrt{u_1} \right) \right), \exp \left( \frac{1}{2\beta} \left( \log k_2^2 - \sqrt{u_2} \right) \right) \right\}, & \text{if } \beta < 0 \end{cases}.$$

- $k_t | \cdot \sim N\left(0, \frac{a}{y_t^2}\right) I_{(L,U)}$ , for  $t = 2, \dots, T - 1$ , where

$$L = \begin{cases} \max \left\{ 0, \exp \left( 0.5 \left( \beta \log k_{t-1}^2 - \sqrt{u_t} \right) \right), \exp \left( \frac{1}{2\beta} \left( \log k_{t+1}^2 - \sqrt{u_{t+1}} \right) \right) \right\}, & \text{if } \beta > 0 \\ \max \left\{ 0, \exp \left( -0.5 \left( \beta \log k_{t-1}^2 + \sqrt{u_t} \right) \right), \exp \left( \frac{1}{2\beta} \left( \log k_{t+1}^2 + \sqrt{u_{t+1}} \right) \right) \right\}, & \text{if } \beta < 0 \end{cases},$$

$$U = \begin{cases} \min \left\{ 0, \exp \left( -0.5 \left( \beta \log k_{t-1}^2 - \sqrt{u_t} \right) \right), \exp \left( \frac{1}{2\beta} \left( \log k_{t+1}^2 + \sqrt{u_{t+1}} \right) \right) \right\}, & \text{if } \beta > 0 \\ \min \left\{ 0, \exp \left( -0.5 \left( \beta \log k_{t-1}^2 - \sqrt{u_t} \right) \right), \exp \left( \frac{1}{2\beta} \left( \log k_{t+1}^2 - \sqrt{u_{t+1}} \right) \right) \right\}, & \text{if } \beta < 0 \end{cases}.$$

- $k_T|\cdot \sim N\left(0, \frac{a}{y_T^2}\right) I_{(L,U)}$ , for  $t = 2, \dots, T - 1$ , where

$$L = \max \left\{ 0, \exp \left( 0.5 \left( \beta \log k_{t-1}^2 - \sqrt{u_t} \right) \right) \right\}$$

$$U = \min \left\{ 0, \exp \left( -0.5 \left( \beta \log k_{t-1}^2 - \sqrt{u_t} \right) \right) \right\}.$$

In the end of each sweep of the algorithm we re-transform  $k_t$  to  $h_t$  using the reverse transformation  $h_t = 1/k_t^2$ ; for  $t = 0, \dots, T$ ,

### 7.2.3 Some Applications

The above algorithm for the estimation of the stochastic volatility model parameters' is applied to two series of data sets. Both of them are consisted of 844 daily exchange rates multiplied by 10000. In detail, the daily exchange rate of the US dollar (USD) and the Japanese Yen (JPY) with respect to the Greek Drachma (GRD) are used (Figure 7-1). The first 100000 iterations of the algorithm were dropped as burn-in and 1 sample point is kept every 500 iterations so as the final samples, that are consisting of 1000 values each, to be approximately independent and identically distributed samples from the marginal densities of the parameters of interest.

The final posterior samples were checked for convergence to the limiting distribution by the criteria of Geweke(1992), Raftery and Lewis (1992) and Heidelberger and Welch (1983). Apart from this result, the subsampling diagnostic (see: Section 4.3 and Giakoumatos *et al.*, 1999) was used in order the convergence to be checked. This criterion was applied to the initial 100000 iterations of each chain by setting:  $a = 0.05$ ,  $t = 0.90$ ,  $d = 0.999$ . Note that, the subsampling diagnostic is considered by its authors (Giakoumatos, Vrontos, Dellaportas and Politis, 1999) as a 'very conservative' one.

Table 7.1 presents the number of lagged values where the autocorrelation dies out for each series of data. Graphically, these results are presented in Figure 7-2

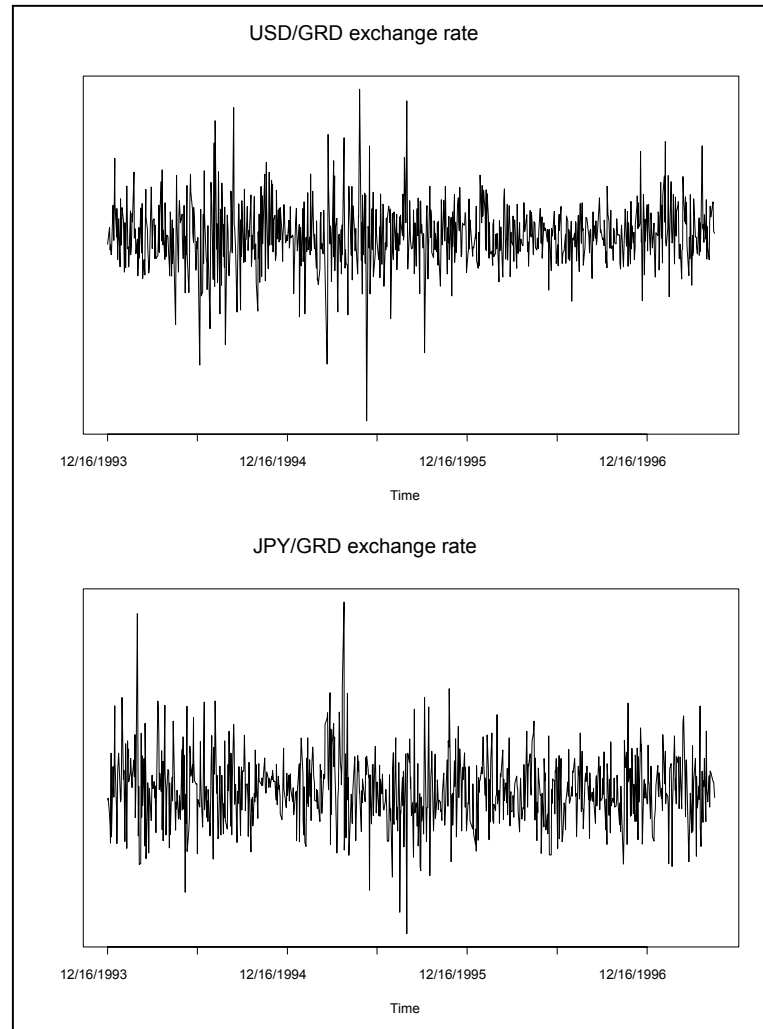


Figure 7-1: Exchange Rates Series

	USD	JPY
$a$	34	18
$b$	195	85
$\sigma^2$	134	170

Table 7.1: Autocorrelation function results for the parameters of the SV model

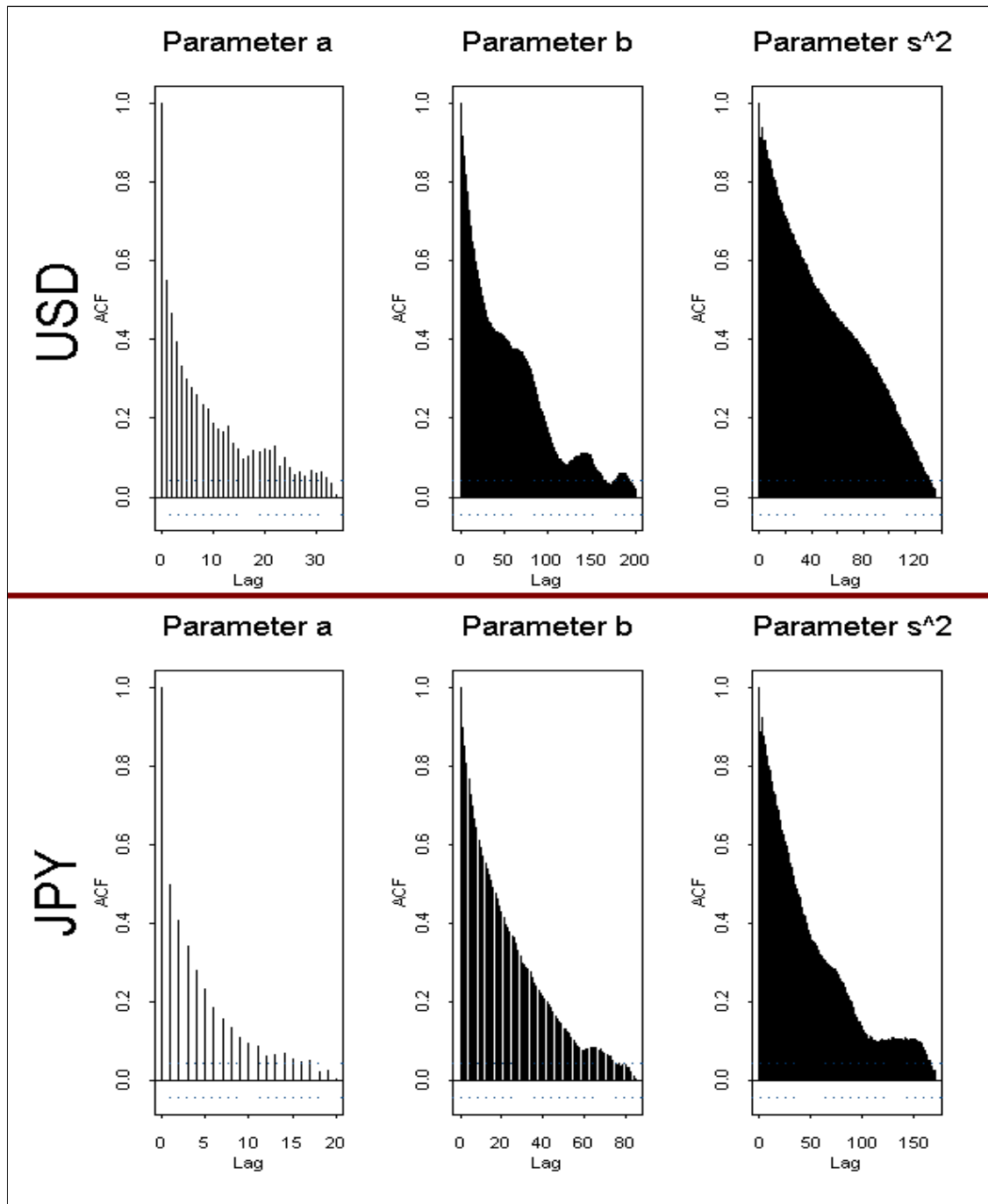


Figure 7-2: Autocorrelation function plots for the parameters of the SV model

Parameters	Mean	Variance
$a$	19.82866	2.23853
$\beta$	0.4693494	0.01269
$\sigma^2$	0.4659143	0.00896

Table 7.2: Posterior Statistics for the parameters of the SV model for USD

## Results for USD

In the case of stochastic volatility model as this applied to USD, Figure 7-3 presents the 1000 iterations of the MCMC algorithm for each parameter of interest. Based on this i.i.d. sample, the posterior mean and variance of the parameters  $a$ ,  $\beta$  and  $\sigma^2$  are estimated (see Table 7.2).

These posterior summary statistics indicate that the USD series is 'weakly' volatility persistent ( $\beta = 0.4693494$ ). The posterior histograms of the parameters  $a$ ,  $\beta$  and  $\sigma^2$  are presented in Figure 7-4.

As far as the convergence of the Markov chain is concerned, the subsampling diagnostic points out that the proposed algorithm needs approximately 25000 iterations to get in the target distribution. The results of the diagnostic are presented in Figure 7-5.

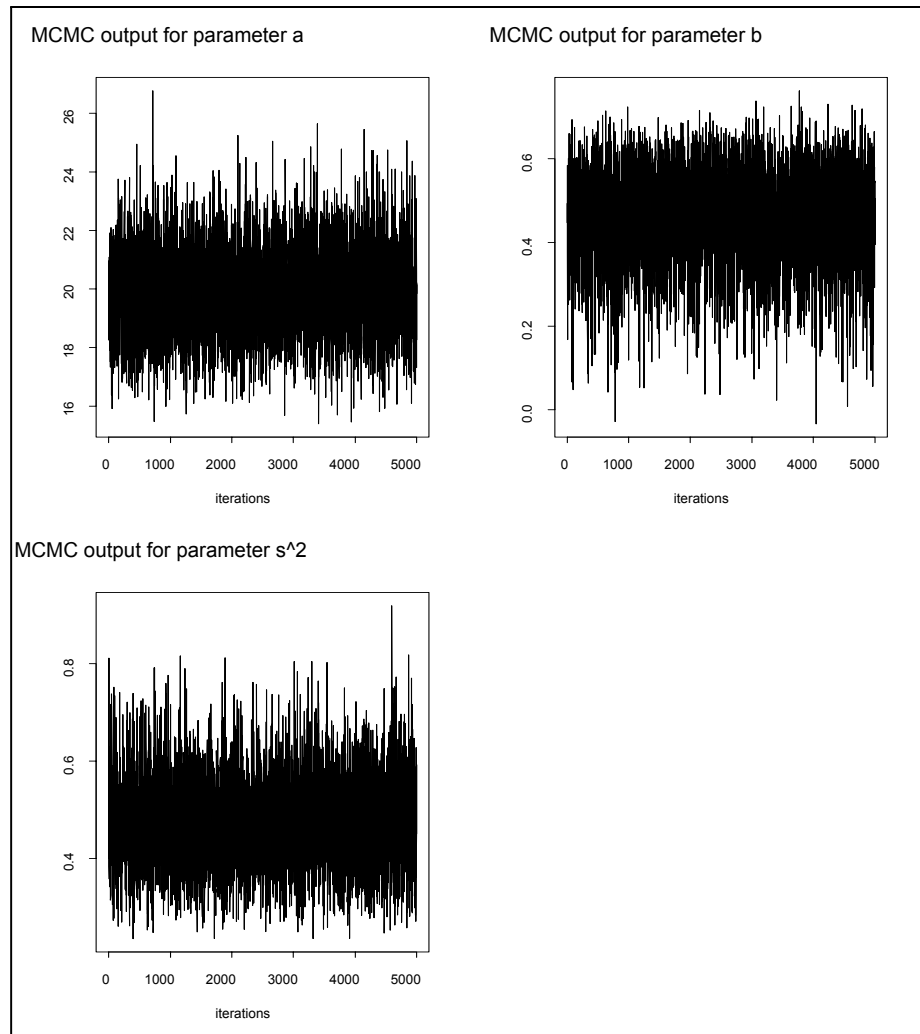


Figure 7-3: MCMC output for parameters of SV for USD

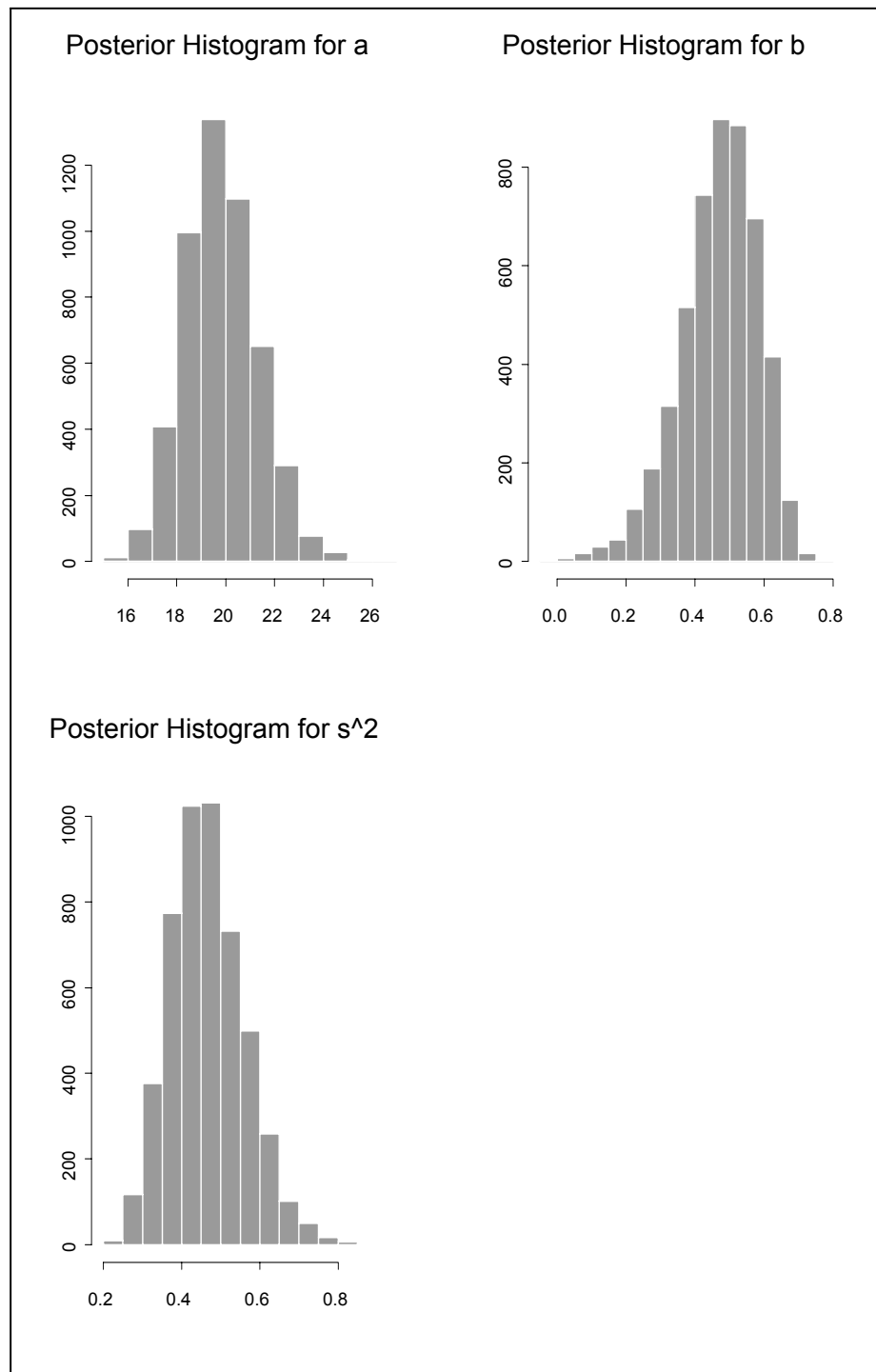


Figure 7-4: Posterior Histograms for the parameters of SV for USD

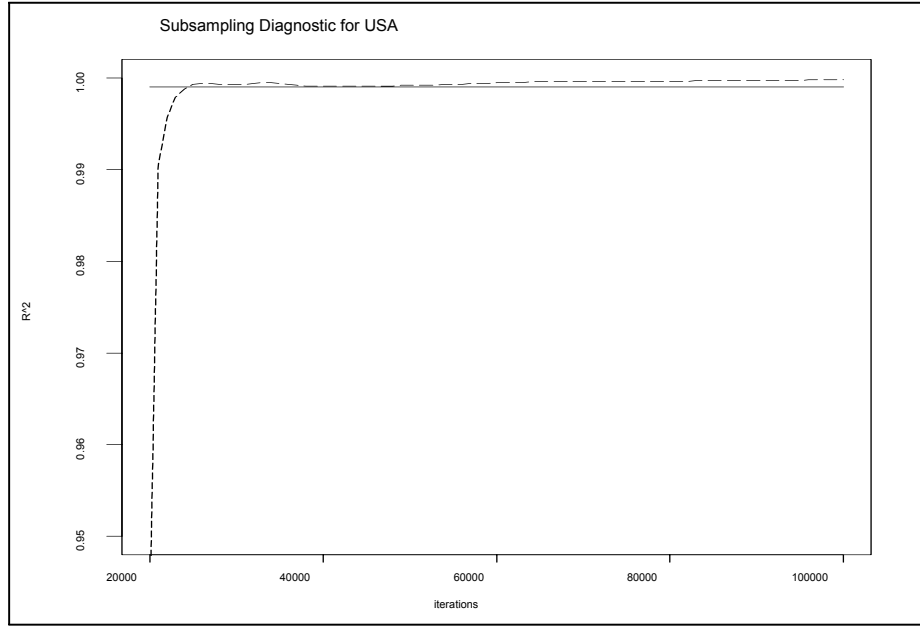


Figure 7-5: Subsampling diagnostic plot for USD

Parameters	Mean	Variance
$a$	21.80848	2.336953
$\beta$	0.430045	0.009154
$\sigma^2$	0.403155	0.006299

Table 7.3: Posterior Statistics for the parameters of the SV model for JPY

## Results for JPY

Figure 7-6 presents the 1000 iterations of the MCMC algorithms for each parameter of interest when the stochastic volatility model applied to the JPY. Based on this i.i.d. sample, the posterior mean and variance of the parameters are estimated. The Table 4.3 presents these estimates.

Based on these posterior summary statistics, it can be concluded that the JPY series is 'weakly' volatility persistent ( $\beta = 0.430045$ ). The posterior histograms of the parameters  $a$ ,  $\beta$  and  $\sigma^2$  are presented in Figure 7-7.

Considering the convergence of the Markov chain, the subsampling diagnostic points

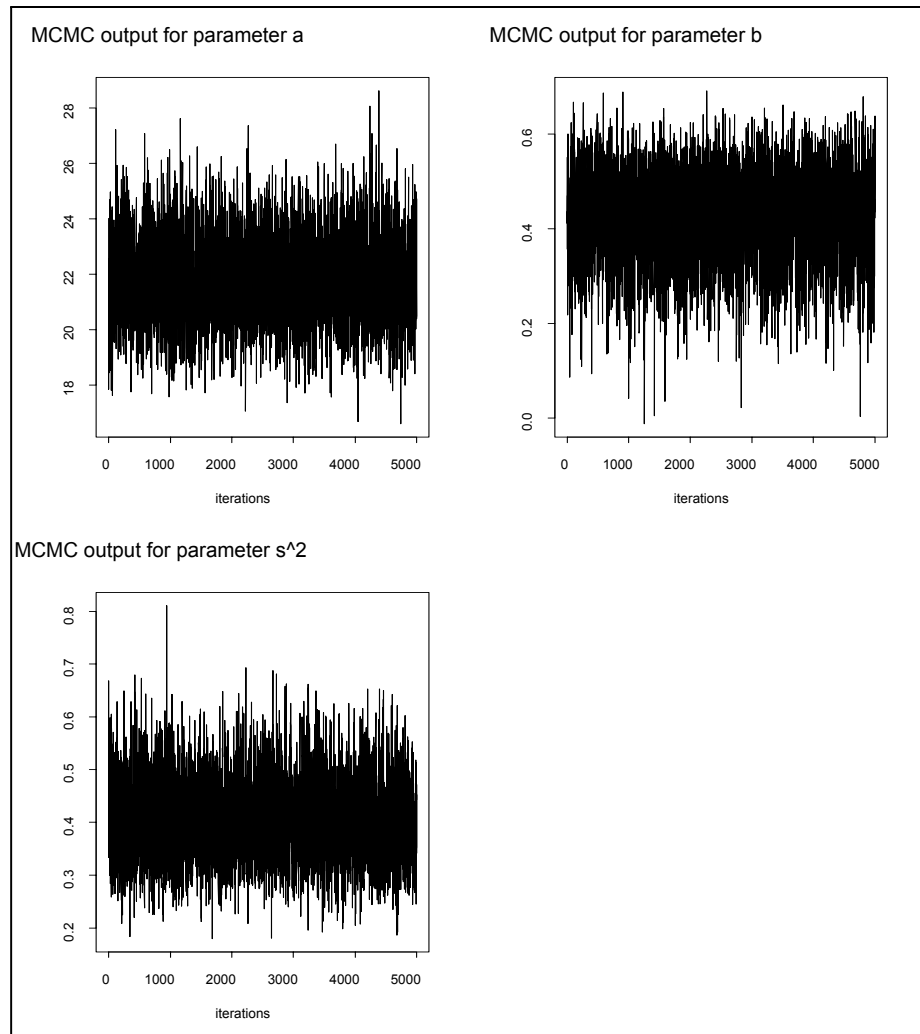


Figure 7-6: MCMC output for parameters of SV for JPY

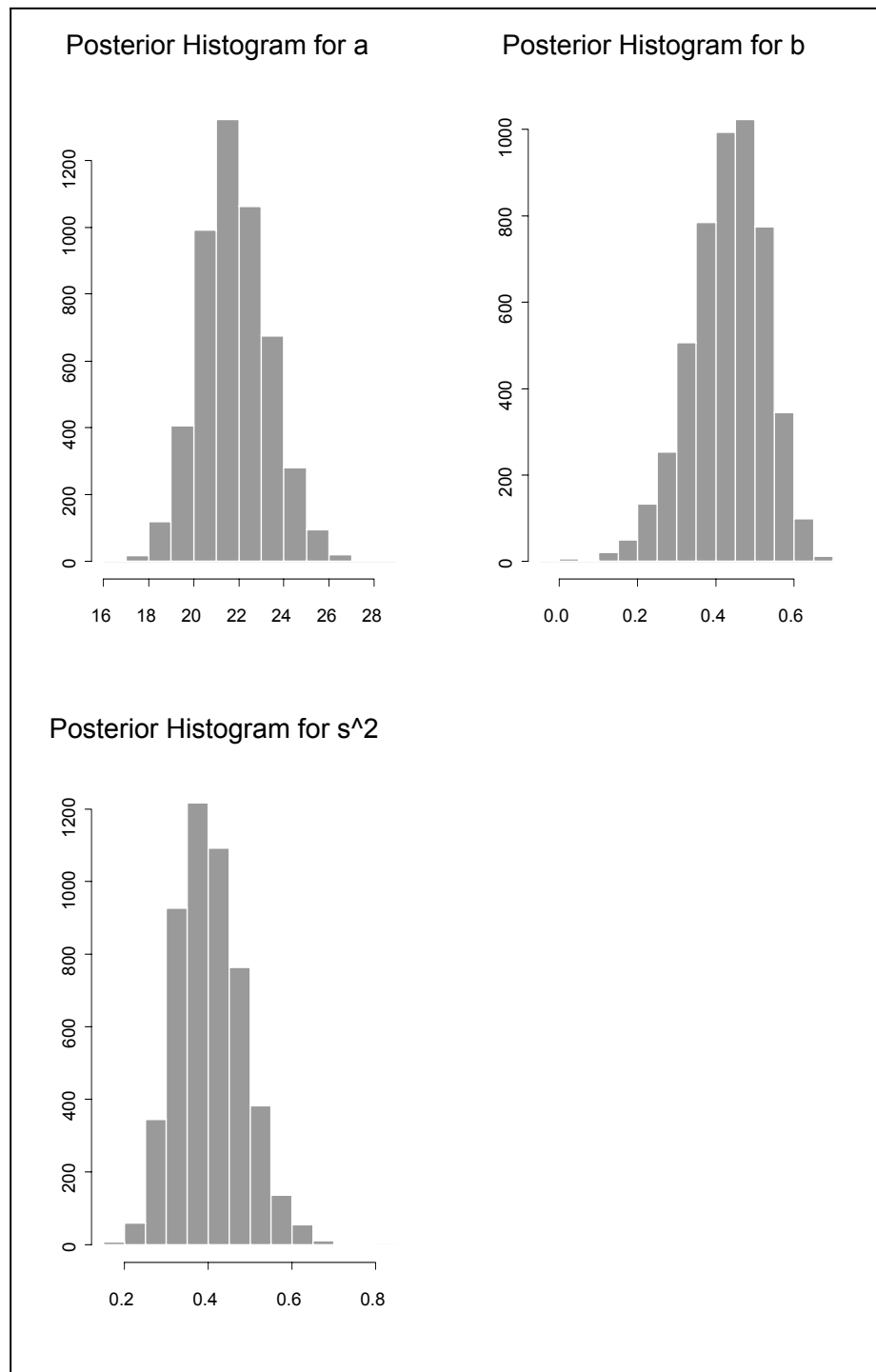


Figure 7-7: Posterior Histograms for the parameters of  $sv$  for JPY

out that the proposed algorithm needs approximately 32000 iterations to get in the target distribution. The results of the diagnostic are presented in Figure 7-8.

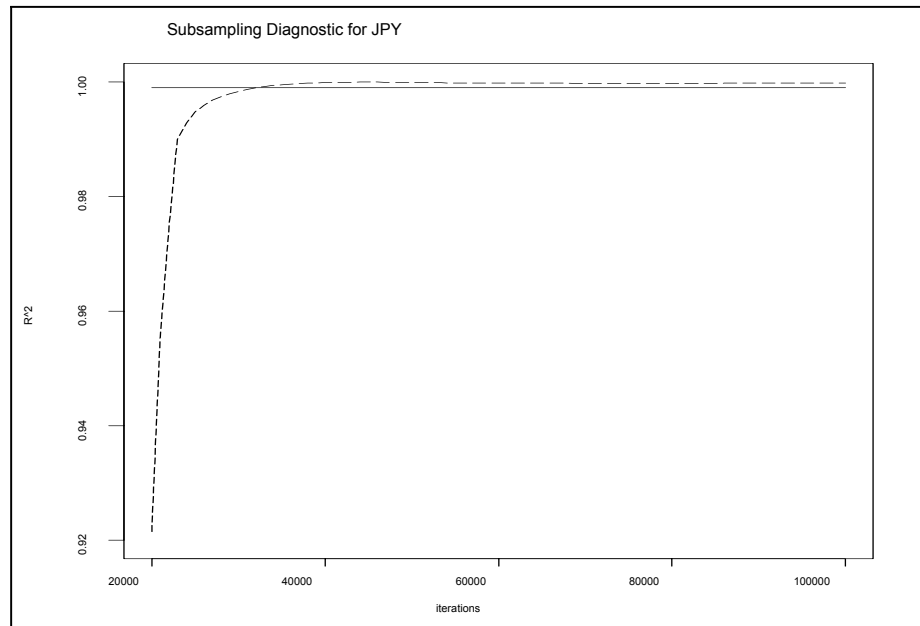


Figure 7-8: Subsampling diagnostic plot for JPY

## 7.3 Unobserved ARCH model

### 7.3.1 Introduction

The Unobserved ARCH model has been presented by Shephard (1996). The ARCH components in this model are observed with disturbances. The form of this model can be written using the following hierarchical structure of conditional densities:

$$\begin{aligned}y_t|f_t, \sigma^2 &\sim \mathbf{N}(f_t, \sigma^2), \\f_t|f_{t-1}, \alpha, b, f_0 &\sim \mathbf{N}(0, h_t), \\h_t &= \alpha + b \cdot f_{t-1}^2.\end{aligned}\tag{7.8}$$

where  $y_1, \dots, y_T$  is a realization of the process,  $f_t$  is the unobserved ARCH component at time  $t$ ,  $f_0$  is the initial state or the “history” of the unobserved components and  $\mathbf{N}(\cdot, \cdot)$  is the Normal distribution. To obtain  $h_t > 0$ , the parameters  $a$  and  $b$  are restricted to be positive. The additional restriction  $0 < b \leq 1$  is placed so that the ARCH component of the model to be covariance stationary (Engle 1982). Note that the unobserved component  $f_t$  is not measurable with respect to the available information at time  $t$ , something which characterizes this class of models. The unconditional and conditional variances of  $y_t$  are given by  $Var(y_t) = \sigma^2 + a/(1 - b)$  and  $Var(y_t|y_{t-1}, a, b) = \sigma^2 + h_t$ . Therefore, the stochastic process  $y_t$  can be considered to have an underline variance on which it is added the variability which is caused by the effect of volatility clustering.

We now proceed by investigating the behavior of the squares of the returns of the unobserved ARCH model, recalling that

**Theorem 13** *Let  $y_t$  follow an unobserved ARCH( $p$ ) model of the form*

$$\begin{aligned} y_t &= f_t + \sigma e_t, \\ f_t &= u_t \sqrt{h_t}, \\ h_t &= \alpha + \sum_{i=1}^p b_i \cdot f_{t-i}^2, \end{aligned}$$

where  $e_t, u_t \sim \mathbf{N}(0, 1)$ . Then,  $y_t^2$  follows a non-Gaussian ARMA( $p, p$ ) process.

**Proof.**

$$y_t^2 = y_t^2 - \sigma^2 + \sigma^2 = f_t^2 + k_t + \sigma^2, \quad (7.9)$$

where  $k_t = \sigma^2 e_t^2 + 2f_t \sigma e_t - \sigma^2$ . The ARCH component in (7.9) follows an AR( $p$ ) process since

$$f_t^2 = f_t^2 + h_t - h_t = a + \sum_{i=1}^p b_i \cdot f_{t-i}^2 + z_t,$$

where  $z_t = f_t^2 - h_t$ . Therefore,  $z_t + a = b(B) f_t^2$ , where  $b(\cdot)$  is the  $p^{\text{th}}$  degree polynomial ( $b(\xi) = 1 - b_1 \xi - \dots - b_p \xi^p$ ) and  $B$  is the backward shift operator. Then, (7.9) obtains

$$\begin{aligned} b(B) y_t^2 &= b(B) f_t^2 + b(B) k_t + \left(1 - \sum_{i=1}^p b_i\right) \sigma^2 \\ &= \left(1 - \sum_{i=1}^p b_i\right) \sigma^2 + a + w_t \end{aligned}$$

where  $w_t = b(B) k_t + z_t$ , and  $y_t^2$  is an ARMA( $p, p$ ) process since  $w_t$  is an MA( $p$ ) process as a sum of an MA( $p$ ) process and a white noise. ■

### 7.3.2 Bayesian Approach and the Auxiliary Variable Sampler

The posterior density of the parameters of the model (7.8) can be extracted via Bayes theorem, by

$$\pi(a, b, \sigma^2, f_0, \mathbf{f}|\mathbf{y}) \propto \prod_{t=1}^T (\pi(y_t|f_t, \sigma^2) \pi(f_t|f_{t-1}, a, b)) \pi(a, b, \sigma^2, f_0). \quad (7.10)$$

The first two terms in the above product are derived from the hierarchical structure in (7.8) and the last term,  $\pi(a, b, \sigma^2, f_0)$ , is the joint prior density of  $a$ ,  $b$ ,  $\sigma^2$  and  $f_0$ . These parameters are assumed a priori independent and we choose improper priors for the  $a$ ,  $b$ ,  $\sigma^2$  and a vague Normal density  $\mathbf{N}(0, v)$  for  $f_0$ , so that the joint prior density takes the form  $\pi(a, b, \sigma^2, f_0) \propto (a \cdot \sigma^2)^{-1} \exp\{-0.5f_0^2/v\}$ . Using the above joint prior density, the joint posterior density (7.10) takes the form

$$\begin{aligned} \pi(a, b, \sigma^2, f_0, \mathbf{f}|\mathbf{y}) \propto & \frac{1}{\prod_{t=1}^T \sqrt{a + bf_{t-1}^2}} \exp \left\{ -\frac{1}{2} \sum_{t=1}^T \left( \frac{f_t^2}{a + bf_{t-1}^2} \right) \right\} \\ & \frac{1}{a\sigma^2^{\frac{T+2}{2}}} \exp \left\{ -\frac{1}{2} \left( \frac{1}{\sigma^2} \sum_{t=1}^T (y_t - f_t)^2 + \frac{f_0^2}{v} \right) \right\}. \end{aligned} \quad (7.11)$$

The above posterior (7.11) is heavily parameterized and the full posterior conditional densities (i.e. the full posterior conditional density means, the posterior density of one parameter condition on all the remaining parameters) are not of standard forms. Therefore, the construction of the MCMC algorithm is not at all simple. In order to handle this problem, some *non-linear* transformations of the parameter space are adopted. Firstly, note that

**Theorem 14** *If in the posterior density  $\pi(a, b, \sigma^2, f_0, \mathbf{f}|\mathbf{y})$  defined in (7.11) we perform the following transformations  $g = \sqrt{a/b}$  and  $w_t = \sqrt{b/af_t}$ ; for  $t = 0, \dots, T$ , the posterior*

density takes the form

$$\pi(g, b, \sigma^2, \mathbf{w} | \mathbf{y}) \propto \frac{1}{\prod_{t=1}^T \sqrt{1 + w_{t-1}^2}} \exp \left\{ -\frac{1}{2b} \sum_{t=1}^T \frac{w_t^2}{(1 + w_{t-1}^2)} \right\} \quad (7.12)$$

$$\frac{1}{\sigma^{2\frac{T+2}{2}} b^{\frac{T}{2}}} \exp \left\{ -\frac{1}{2} \left( \frac{1}{\sigma^2} \sum_{t=1}^T (y_t - gw_t)^2 + \frac{(gw_0)^2}{v} \right) \right\},$$

where  $\mathbf{w} = (w_0, \dots, w_T)$ .

**Proof.** Note that the Jacobian of the above transformations is  $|J| = 2bg^{T+2}$ . The remaining calculations are straightforward. ■

By using Theorem 14, the resulting posterior density (7.12) has full conditional densities of a rather convenient form. In particular,

- $\sigma^2 | \cdot \sim \mathbf{IG} \left( \frac{T}{2}, \frac{1}{2} \sum_{t=1}^T (y_t - gw_t)^2 \right)$ , where  $\mathbf{IG}(a, b)$  denotes the Inverse Gamma density with mean  $b/(a - 1)$ ; the notation  $|\cdot$  implies conditioning on all the remaining parameters.
- $[b | \cdot] \sim \mathbf{IG} \left( \frac{T-2}{2}, \frac{1}{2} \sum_{t=1}^T \frac{w_t^2}{1+w_{t-1}^2} \right) \mathbf{I}(b \leq 1)$ , where  $\mathbf{I}(\cdot)$  is the indicator function.
- $g | \cdot \sim \mathbf{N}(m, s) \mathbf{I}(g \geq 0)$ , where  $m = \left( v \sum_{t=1}^T w_t y_t \right) / \left( \sigma^2 w_0^2 + v \sum_{t=1}^T w_t^2 \right)$  and  $s = \left( \sigma^2 v \right) / \left( \sigma^2 w_0^2 + v \sum_{t=1}^T w_t^2 \right)$ .
- $\pi(w_0 | \cdot) \propto \mathbf{ND} \left( 0, \frac{v}{g^2} \right) \frac{1}{\sqrt{1+w_0^2}} \exp \left\{ -\frac{1}{2b} \frac{w_1^2}{(1+w_0^2)} \right\}$ .
- $\pi(w_t | \cdot) \propto \mathbf{ND} \left( m_t, s_t^2 \right) \frac{1}{\sqrt{1+w_t^2}} \exp \left\{ -\frac{1}{2b} \frac{w_{t+1}^2}{(1+w_t^2)} \right\}$ , for  $t = 1, \dots, T-1$ .
- $w_T | \cdot \sim \mathbf{N}(m_T, s_T^2)$ ,

where  $\mathbf{ND}(\cdot, \cdot)$  denotes the probability density function of the Normal distribution and  $m_t$  and  $s_t^2$  are given by

$$m_t = \frac{y_t g b (1 + w_{t-1}^2)}{g b (1 + w_{t-1}^2) + \sigma^2}, \quad s_t^2 = \frac{\sigma^2 b (1 + w_{t-1}^2)}{g^2 b (1 + w_{t-1}^2) + \sigma^2}. \quad (7.13)$$

Again, the full conditional densities of  $w_t$ , for  $t = 0, \dots, T-1$ , are not of known forms. One way to deal with that, is to use Metropolis-Hastings steps (Hastings 1970; Metropolis, Rosenbluth, Rosenbluth, Teller and Teller 1953) which allow us to sample from non-standard densities. Giakoumatos, Dellaportas and Politis (2004a) tried a random walk Metropolis-Hastings step with a Normal proposal density with variance given by  $s_t^2$ . For a series of data sets that analyzed in this Thesis, the probability of acceptance is approximately 0.5, a value which has been considered satisfactory by Chib and Greenberg (1995a).

However, note that the full conditional densities of  $w_t$ ,  $t = 0, \dots, T-1$ , can be written as

$$\pi(w_t | \cdot) \propto \mathbf{N}(m_t, s_t^2) \Psi(w_t), \quad t = 0, \dots, T-1$$

where  $\Psi(w_t)$  is a function of  $w_t$ . In that case, Giakoumatos, Dellaportas and Politis (2004a) followed Chib and Greenberg (1994) (see also Chib and Greenberg 1995b) and sample from these full conditional densities by a Metropolis-Hastings step using as proposal density  $\mathbf{N}(m_t, s_t^2)$ . In this case the probability of acceptance reduces to  $\min\{1, \Psi(w'_t) / \Psi(w_t)\}$ , where  $w'_t$  is the proposal value.

Another way to sample from  $w_t$ ,  $t = 0, \dots, T-1$ , is to use auxiliary variable sampling techniques (Swendsen and Wang 1987; Edwards and Sokal 1988; Besag and Green 1993; Higdon 1998; Damien, Wakefield and Walker 1999; Neal 2003). The way that this can be achieved becomes evident in the next section.

### 7.3.3 Usage of the Auxiliary Variable Sampling

In this section the auxiliary variable sampler is utilized to construct an MCMC algorithm from which a sample can be easily drawn from the posterior of the Unobserved ARCH model. The parameter space is expanded, introducing  $2T$  auxiliary variables in such a way that the resulting MCMC algorithm to consist of only Gibbs steps. To construct the proposed algorithm the following Theorem is used.

**Theorem 15** *If we include  $2T$  positive latent variables  $\mathbf{u} = (u_1, \dots, u_T)$  and  $\mathbf{k} = (k_1, \dots, k_T)$  in the posterior density (7.12) such that the resulting joint density is given by*

$$\begin{aligned} \pi(g, b, \sigma^2, \mathbf{w}, \mathbf{u}, \mathbf{k} | \mathbf{y}) \propto & \frac{1}{\sigma^{2\frac{T+2}{2}} b^{\frac{T}{2}}} \left( \prod_{t=1}^T \mathbf{I} \left( u_t \leq \frac{1}{\sqrt{1 + w_{t-1}^2}} \right) \right) \\ & \left( \prod_{t=1}^T \mathbf{I} \left( k_t \leq \exp \left( -\frac{1}{2b} \frac{w_t^2}{(1 + w_{t-1}^2)} \right) \right) \right) \\ & \exp \left\{ -\frac{1}{2} \left( \frac{1}{\sigma^2} \sum_{t=1}^T (y_t - gw_t)^2 + \frac{(gw_0)^2}{v} \right) \right\}, \end{aligned}$$

then, the marginal density  $\pi(g, b, \sigma^2, \mathbf{w} | \mathbf{y})$  is given by (7.12).

The above Theorem guarantees that a MCMC algorithm which obtains samples from  $\pi(\mathbf{u}, \mathbf{k}, g, b, \mathbf{w}, \sigma^2 | \mathbf{y})$  obtains also samples from  $\pi(g, b, \mathbf{w}, \sigma^2 | \mathbf{y})$ . To utilize Theorem 15, we need to further elaborate on the resulting full conditional densities of  $\pi(g, b, \sigma^2, \mathbf{w}, \mathbf{u}, \mathbf{k} | \mathbf{y})$ . In fact, it is readily evident that it is more convenient to use some forms of conditional densities appropriately marginalised over some parameters (Chib and Carlin 1999). In particular, to sample from  $\pi(g, b, \sigma^2, \mathbf{w}, \mathbf{u}, \mathbf{k} | \mathbf{y})$  we use the full conditional densities of  $g$  and  $\sigma^2$ , which are presented in section 7.3.2, because they are independent of  $\mathbf{u}$  and  $\mathbf{k}$ . For the remaining of the parameters, the following updating steps are used:

- Instead of sampling from  $\pi(b|\cdot) \equiv \pi(b|\mathbf{k}, \mathbf{w})$  sampling procedure is taking place from

- $$\pi(b|\mathbf{w}) \equiv \mathbf{IG}\left(\frac{T-2}{2}, \frac{1}{2} \sum_{t=1}^T \frac{w_t^2}{1+w_{t-1}^2}\right) \mathbf{I}(b \leq 1).$$
- $\pi(u_t|\cdot) \equiv \mathbf{U}\left(0, \frac{1}{\sqrt{1+w_{t-1}^2}}\right)$ , for all  $t = 1, \dots, T$ .
  - $\pi(k_t|\cdot) \equiv \mathbf{U}\left(0, \exp\left\{-\frac{w_t^2}{2b(1+w_{t-1}^2)}\right\}\right)$ , for all  $t = 1, \dots, T$ .
  - $\pi(w_0|\cdot) \equiv \mathbf{N}(0, v) \mathbf{I}\left(u_1 \leq \frac{1}{\sqrt{1+w_0^2}}\right) \cdot \mathbf{I}\left(k_1 \leq \exp\left\{-\frac{w_1^2}{2b(1+w_0^2)}\right\}\right)$ .
  - Instead of sampling from  $\pi(w_t|\cdot) \equiv \pi(w_t|u_{t+1}, k_t, k_{t+1}, g, b, w_{t-1}, w_{t+1}, \sigma^2, \mathbf{y})$ ; for  $t = 1, \dots, T-1$ , sampling is taking place from
 
$$\pi(w_t|u_{t+1}, k_{t+1}, g, b, w_{t-1}, w_{t+1}, \sigma^2, \mathbf{y}) \equiv \mathbf{N}(m_t, s_t^2) \mathbf{I}\left(u_{t+1} \leq \frac{1}{\sqrt{1+w_t^2}}\right) \mathbf{I}\left(k_{t+1} \leq \exp\left\{-\frac{w_{t+1}^2}{2b(1+w_t^2)}\right\}\right),$$
 where  $m_t$  and  $s_t^2$  are defined in (7.13).
  - Instead of sampling from  $\pi(w_T|\cdot) \equiv \pi(w_T|k_T, g, b, w_{T-1}, \sigma^2, \mathbf{y})$ , sampling is taking place from
 
$$\pi(w_T|g, b, w_{T-1}, \sigma^2, \mathbf{y}) \equiv \mathbf{N}(m_T, s_T^2),$$
 where  $m_T$  and  $s_T^2$  are defined in (7.13).

In order to sample from the above truncated Normal density, rejection sampling (Gelfand, Smith and Lee 1992) is used. On the other hand, the method of Robert (1995) could be used. For the full conditional density of  $b$ , which is truncated Inverse Gamma density, the *AV* sampler is used, introducing a latent variable; see section 10.6.

### 7.3.4 Some Applications

The aforementioned algorithm for the estimation of the parameters of the unobserved ARCH model is applied to two series of data sets. Both of them are consisted of 844 daily exchange rates multiplied by 10000. In detail, the daily exchange rate of the US dollar (USD) and the Japanese Yen (JPY) with respect to the Greek Drachma (GRD) are used (Figure 7-1, page 98). The first 50000 iterations of the algorithm were dropped as burn-in and 1 sample point is kept every 500 iterations such as the final samples, that are

	USD	JPY
$a$	190	270
$b$	140	290
$\sigma^2$	210	120

Table 7.4: Autocorrelation function results for the parameters of the unobserved ARCH model

consisting of 5000 values, to be approximately, independent and identically distributed samples from the marginal densities of the parameters of interest.

The final posterior samples were checked for convergence to the limiting distribution by the criteria of Geweke (1992), Raftery and Lewis (1992) and Heidelberger and Welch (1983). Apart from this result, the subsampling diagnostic (see: section 4.3 and Giakoumatos *et al.*, 1999) was used in order the convergence to be checked. This criterion was applied to the initial 50000 iterations of each chain by setting:  $a = 0.05, t = 0.90, d = 0.999$ . Note that, the subsampling diagnostic is considered by its authors (Giakoumatos *et al.*, 1999) ‘very conservative’.

Table 7.4 presents the number of lagged values where the autocorrelation dies out for each series of data. Graphically, these results are presented in Figure 7-9.

Finally, note that Giakoumatos, Dellaportas and Politis (2004a) compared the auxiliary variable algorithm with respect to the Metropolis-Hastings algorithm and the Chib & Greenberg algorithm for the unobserved ARCH model and they found enough evidences that the auxiliary variable MCMC algorithm converges faster than the other two algorithms.

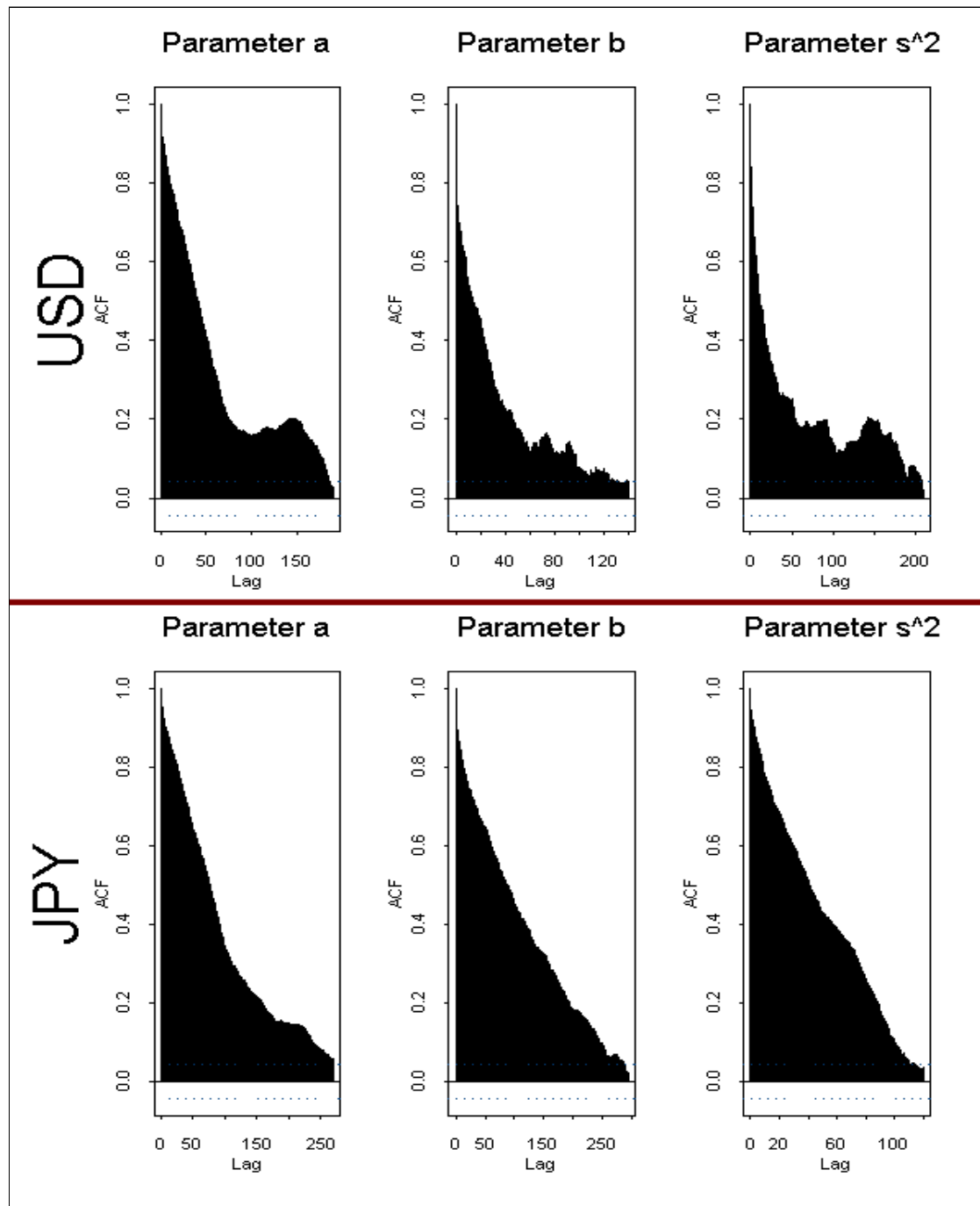


Figure 7-9: Autocorrelation function plots for the parameters of the unobserved ARCH model

Parameters	Mean	Variance
$a$	2.666473	1.041731
$b$	0.9055209	0.006319529
$\sigma^2$	14.61337	3.401016

Table 7.5: Posterior Statistics for the parameters of the unobserved ARCH model for USD

### Results for USD

In the case of unobserved ARCH model as this applied to USD, Figure 7-10 presents the 5000 iterations of the MCMC algorithm for each parameter of interest. Based on this i.i.d. sample, the posterior mean and variance of the parameters of the model are estimated. The following Table presents these estimates.

These posterior summary statistics indicate that the USD series is 'strongly' volatility persistent ( $b = 0.9055209$ ). The posterior histograms of the parameters  $a$ ,  $b$  and  $\sigma^2$  are presented in Figure 7-11.

For the convergence of the Markov chain, the subsampling diagnostic points out that the proposed algorithm needs approximately 27500 iterations to get in the target distribution. The results of the diagnostic are presented in Figure 7-12.

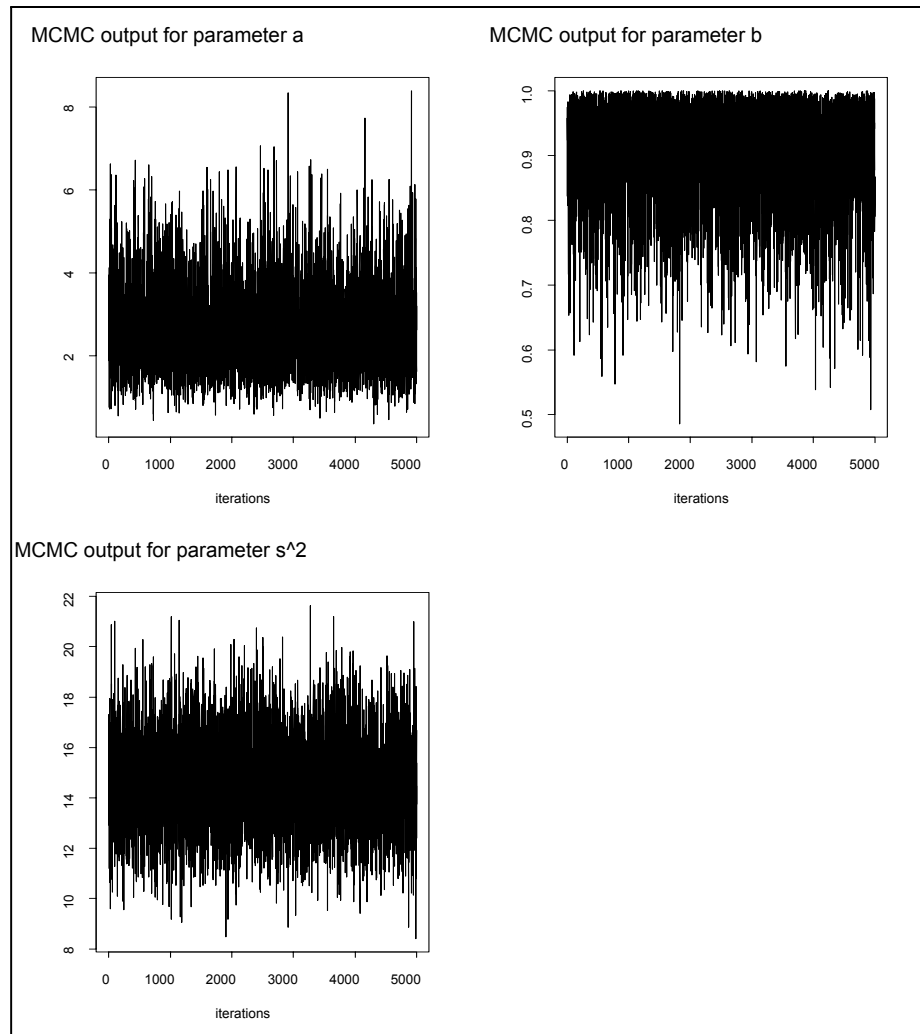


Figure 7-10: MCMC output for parameters of unobserved ARCH for USD

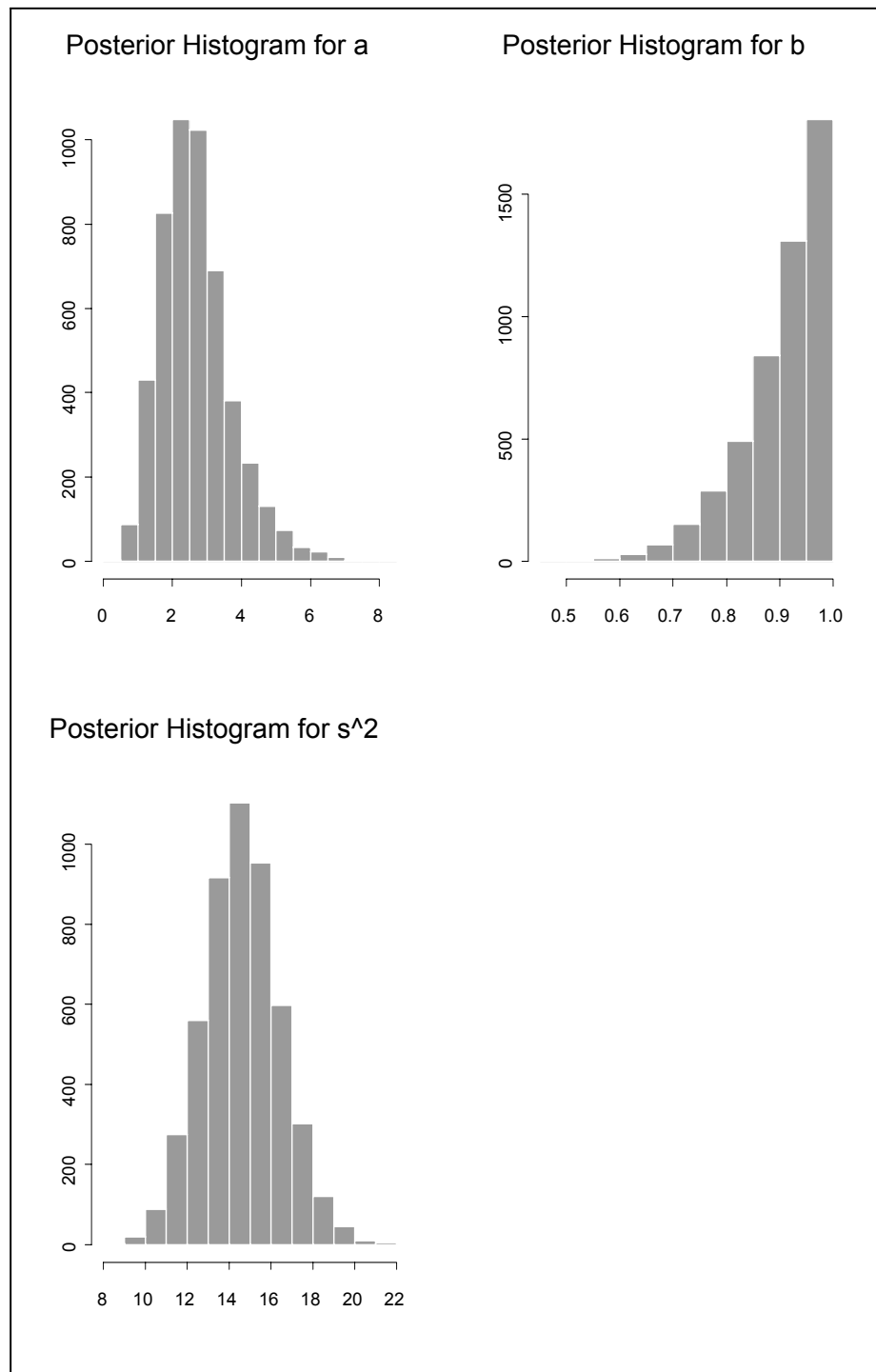


Figure 7-11: Posterior Histograms for the parameters of unobserved ARCH for USD

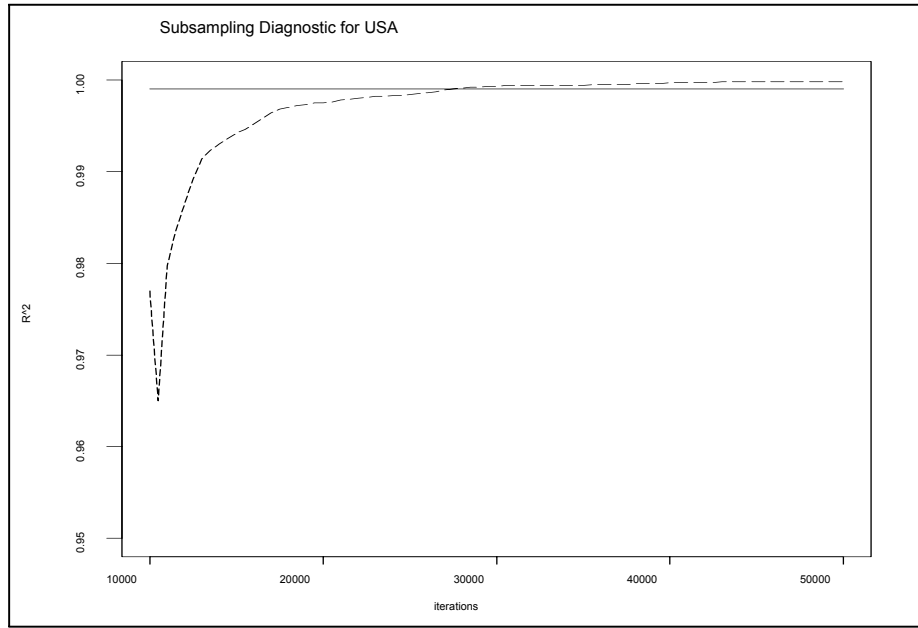


Figure 7-12: Subsampling diagnostic plot for USD

Parameters	Mean	Variance
$a$	4.76157	8.224736
$\beta$	0.7497178	0.02615015
$\sigma^2$	14.37208	10.78908

Table 7.6: Posterior Statistics for the parameters of the unobserved ARCH model for JPY

## Results for JPY

Figure 7-13 presents the 1000 iterations from the MCMC for each parameter of interest when the unobserved ARCH model applied to the JPY. Using this i.i.d. sample, the posterior mean and variance of the parameters of the model are estimated (see Table 7.6).

These posterior summery statistics indicate that the JPY series is volatility persistent ( $b = 0.7497178$ ). The posterior histograms of the parameters  $a$ ,  $\beta$  and  $\sigma^2$  are presented in Figure 7-14.

For the convergence of the Markov chain, the subsampling diagnostic points out

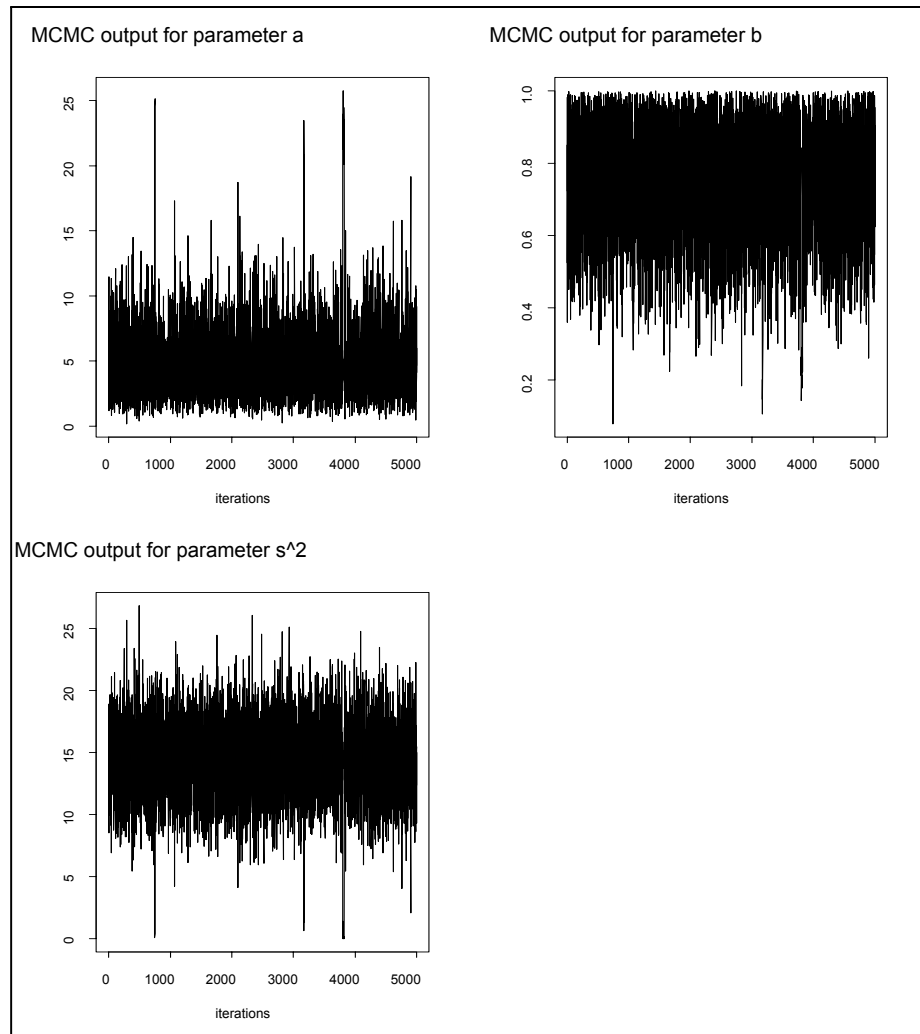


Figure 7-13: MCMC output for parameters of unobserved ARCH for JPY

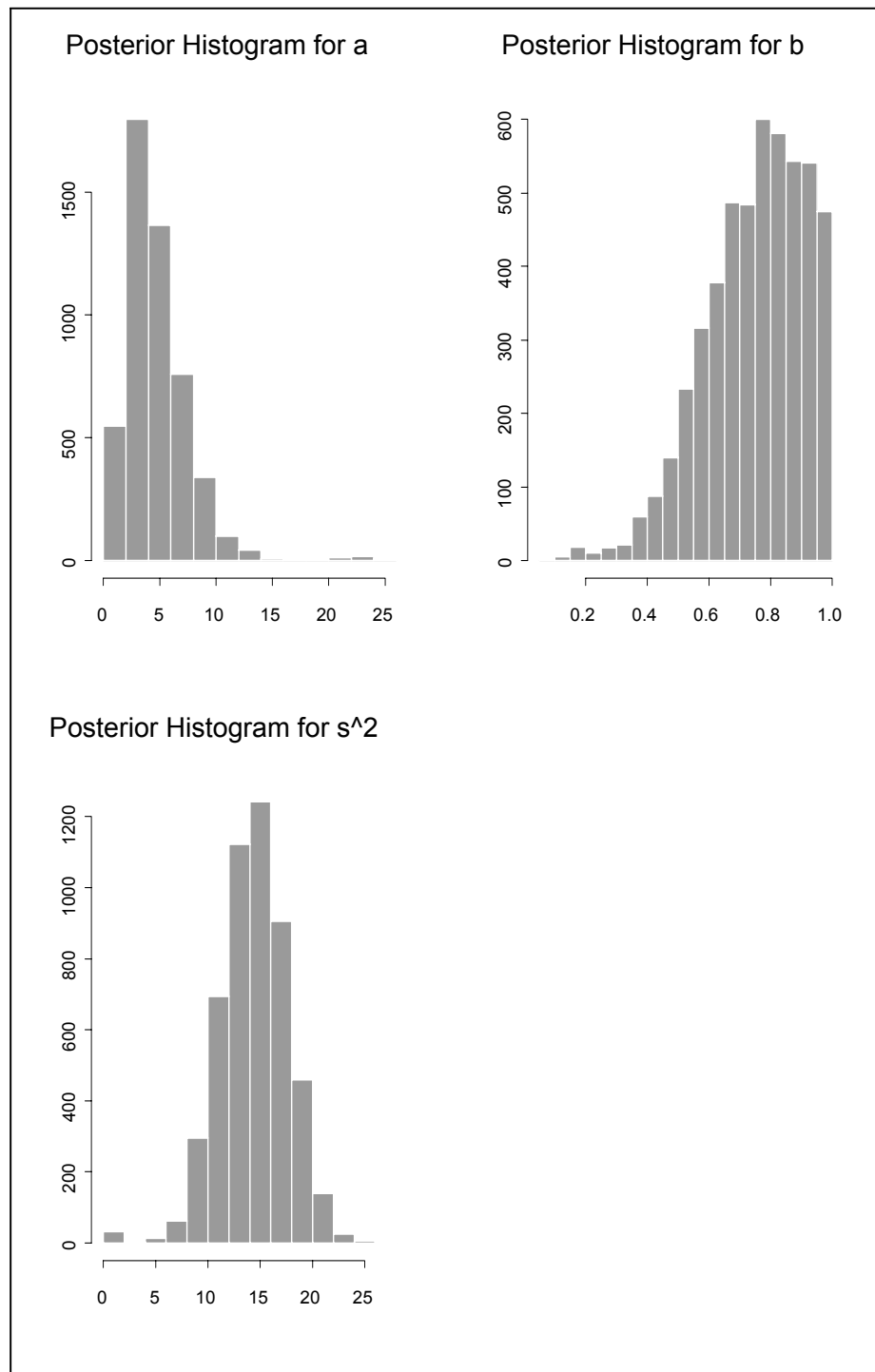


Figure 7-14: MCMC output for parameters of unobserved ARCH for JPY

that the proposed algorithm needs approximately 26500 iterations to get in the target distribution. The results of the diagnostic are presented in Figure 7-15.

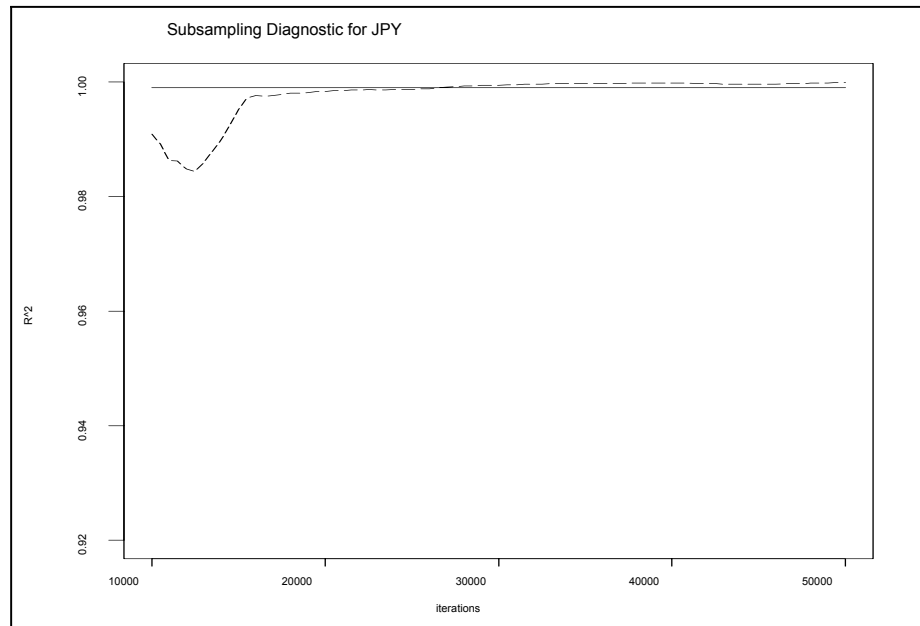


Figure 7-15: Subsampling diagnostic plot for JPY

## 7.4 ARCH model

### 7.4.1 Introduction

In this section an auxiliary variable MCMC algorithm is proposed in order to take sample from the posterior density of the parameters of ARCH(1) model (Engle, 1982). The ARCH(1) model can be written as

$$\begin{aligned} y_t | a_0, a_1, \mathbf{y}_{t-1} &\sim N(0, \sigma_t^2), \\ \sigma_t^2 &= a_0 + a_1 y_{t-1}^2, \end{aligned} \quad (7.14)$$

where  $y_t$  is the time series at time  $t$ ,  $a_0, a_1$  are the parameters of interest and  $\mathbf{y}_{t-1}$  is the vector of information up to time  $t-1$ , i.e.  $\mathbf{y}_t = (y_1, y_2, \dots, y_{t-1})$ . Restrictions  $0 < a_1 < 1$  and  $a_0 > 0$  are imposed to prevent the nonnegativity of the conditional variance and ensure that the ARCH(1) process is covariance stationary (Engle, 1982).

In the next section an easy to construct and fast to converge MCMC algorithm is proposed for the aforementioned model. This MCMC algorithm is based only on Gibbs steps.

Monte Carlo methods for the analysis of the ARCH models have been also proposed by Geweke (1989a, 1989b), Polasek and Muller (1995) and Vrontos (2001).

### 7.4.2 Bayesian Approach and the Auxiliary Variable Sampler

Using non-informative priors for the parameters of interest,  $\pi(a_0, a_1) \propto a_0^{-1}$ , the joint posterior distribution -known up to a constant- for the parameters  $a_0$  and  $a_1$  of model (7.14) takes the following form:

$$f(a_0, a_1 | \mathbf{y}_T) \propto \left( a_0 \prod_{i=1}^T \sqrt{a_0 + a_1 y_{i-1}^2} \right)^{-1} \exp \left\{ -0.5 \sum_{i=1}^T \frac{y_i^2}{a_0 + a_1 y_{i-1}^2} \right\}. \quad (7.15)$$

It is obvious from the above posterior density that the full conditional densities for

$a_0$  and  $a_1$  are not of standard forms, therefore Gibbs algorithm cannot be applied here. Metropolis-Hastings algorithm can be used (Metropolis *et al.*, 1953; Hastings, 1970) but as noted in previous sections Metropolis-Hastings requires a lot of tuning for the proposal density. On the other hand, the product auxiliary variable sampler can be utilized in such a way that constructs an easy to apply algorithm with only Gibbs steps. In the following Theorem  $2T + 1$  auxiliary variables are introduced in the posterior density (7.15) such as the resulting full conditional densities for  $a_0$  and  $a_1$  to be Uniforms.

**Theorem 16** *Let the posterior density given by (7.15). If we include  $2T + 1$  positive latent variables  $w$ ,  $\mathbf{u} = (u_1, \dots, u_T)$  and  $\mathbf{k} = (k_1, \dots, k_T)$  in (7.15) such as the resulting joint density for  $a_0, a_1, w, \mathbf{u}$  and  $\mathbf{k}$  to be given by*

$$f(a_0, a_1, w, \mathbf{u}, \mathbf{k} | \mathbf{y}) \propto \exp \left\{ -0.5 \sum_{i=1}^T u_t \right\} \mathbf{I} \left( w < \frac{1}{a_0} \right) \quad (7.16)$$

$$\prod_{i=1}^T \left\{ \mathbf{I} \left( u_t > \frac{y_t^2}{a_0 + a_1 y_{t-1}^2} \right) \right\}$$

$$\prod_{i=1}^T \left\{ \mathbf{I} \left( k_t < 1 / \sqrt{a_0 + a_1 y_{t-1}^2} \right) \right\}.$$

then, the marginal density  $f(a_0, a_1 | \mathbf{y})$  is given by (7.15).

**Proof:** *It is enough to integrate out all the latent parameters from the resulting*

posterior density (7.16)

$$\begin{aligned}
f(a_0, a_1 | \mathbf{y}) &= \int_w \int_{u_1} \cdots \int_{u_T} \int_{k_1} \cdots \int_{k_T} f(a_0, a_1, w, \mathbf{u}, \mathbf{k} | \mathbf{y}) dw du_1 \cdots du_T dk_1 \cdots dk_T \\
&= \int_0^{1/a_0} dw \int_0^{1/\sqrt{a_0+a_1 y_0^2}} dk_1 \cdots \int_0^{1/\sqrt{a_0+a_1 y_0^2}} dk_T \\
&\quad \int_{\frac{y_1^2}{a_0+a_1 y_0^2}}^{\infty} \exp\{-0.5u_1\} du_1 \cdots \int_{\frac{y_T^2}{a_0+a_1 y_{T-1}^2}}^{\infty} \exp\{-0.5u_T\} du_T \\
&= \left( a_0 \prod_{i=1}^T \sqrt{a_0 + a_1 y_{i-1}^2} \right)^{-1} \exp \left\{ -0.5 \sum_{i=1}^T \frac{y_i^2}{a_0 + a_1 y_{i-1}^2} \right\}. \blacksquare
\end{aligned}$$

Using the posterior density (7.16) an MCMC algorithm which is consisted only of Gibbs steps can be easily constructed. In detail, the proposed algorithm is consisted of the following steps

1. Give initial values  $a_0^0, a_1^0$ .
2. Sample  $u_t | a_0^0, a_1^0, \mathbf{y} \sim \text{exp onential}(0.5) \mathbf{I} \left( u_t > \frac{y_t^2}{a_0^0 + a_1^0 y_{t-1}^2} \right)$ , for all  $t = 1, \dots, T$ .
3. Sample  $k_t | a_0^0, a_1^0, \mathbf{y} \sim \mathbf{U} \left( 0, 1/\sqrt{a_0^0 + a_1^0 y_{t-1}^2} \right)$ , for all  $t = 1, \dots, T$ .
4. Sample  $w | a_0^0 \sim \mathbf{U} \left( 0, \frac{1}{a_0^0} \right)$ .
5. Sample  $a_0^1 | a_1^0, \mathbf{y}, \mathbf{u}, \mathbf{k}, w \sim \mathbf{U}(l, m)$ ,  
where  $l = \max \left\{ \max \left\{ \frac{y_t^2}{u_t} - a_1^0 y_{t-1}^2 \right\}, \{0\} \right\}$  and  
 $m = \min \left\{ \min \left\{ \frac{1}{k_t^2} - a_1^0 y_{t-1}^2 \right\}, \{1/w\}, \{1 - a_1^0\} \right\}$ .
6. Sample  $a_1^1 | a_0^1, \mathbf{y}, \mathbf{u}, \mathbf{k}, w \sim \mathbf{U}(l, m)$ , where  $l = \max \left\{ \max \left\{ \left( \frac{y_t^2}{u_t} - a_0^1 \right) / y_{t-1}^2 \right\}, \{0\} \right\}$   
and  $m = \min \left\{ \min \left\{ \left( \frac{1}{k_t^2} - a_0^1 \right) / y_{t-1}^2 \right\}, \{1 - a_0^1\} \right\}$ .
7. Repeat from step 2.

	USD	JPY
$a_0$	490	600
$a_1$	450	360

Table 7.7: Autocorrelation function results for the parameters of the ARCH model

Note that the above full conditional densities are Uniforms for  $a_0, a_1, \mathbf{k}, \mathbf{w}$  and truncated exponential for  $\mathbf{k}$  therefore it is easy to sample from them using standard procedures (reference for uniform) for Uniforms and following Damien *et al.* (1999) for truncated exponential. Moreover the above algorithm can be easily extended for higher order ARCH models.

### 7.4.3 Some Applications

The algorithm for ARCH(1) model that was described in the previous section is applied to two series of data sets. Both of them are consisted of 844 daily exchange rates (multiplied by 10000). These concern the daily exchange rate of the US dollar (USD) and the Japanese Yen (JPY) with respect to the Greek Drachma (GRD) (Figure 7-1, page 98).

The initial 50000 iterations from each chain are dropped as burn-in and 1 sample point every 600 iterations is kept so as the final samples, that are consisting of 5000 iterations, to be approximately independent and identically distributed samples from the marginal densities of the parameters of interest of each exchange rate. The final samples of 5000 iterations were checked for convergence to the limiting distribution by the criteria of Geweke (1992), Raftery and Lewis (1992) and Heidelberger and Welch (1983). Apart from this result, the subsampling diagnostic was applied in order to check the convergence. This criterion was applied to the initial 50000 iterations of each chain ( $a = 0.05, t = 0.90, d = 0.999$ ). It has to be mentioned that, the subsampling diagnostic is considered by its authors (Giakoumatos, Vrontos, Dellaportas and Politis, 1999) ‘very conservative’.

Table 7.7 presents the number of lagged values where the autocorrelation dies out for each series of data. Graphically, these results are presented in Figure 7-16.

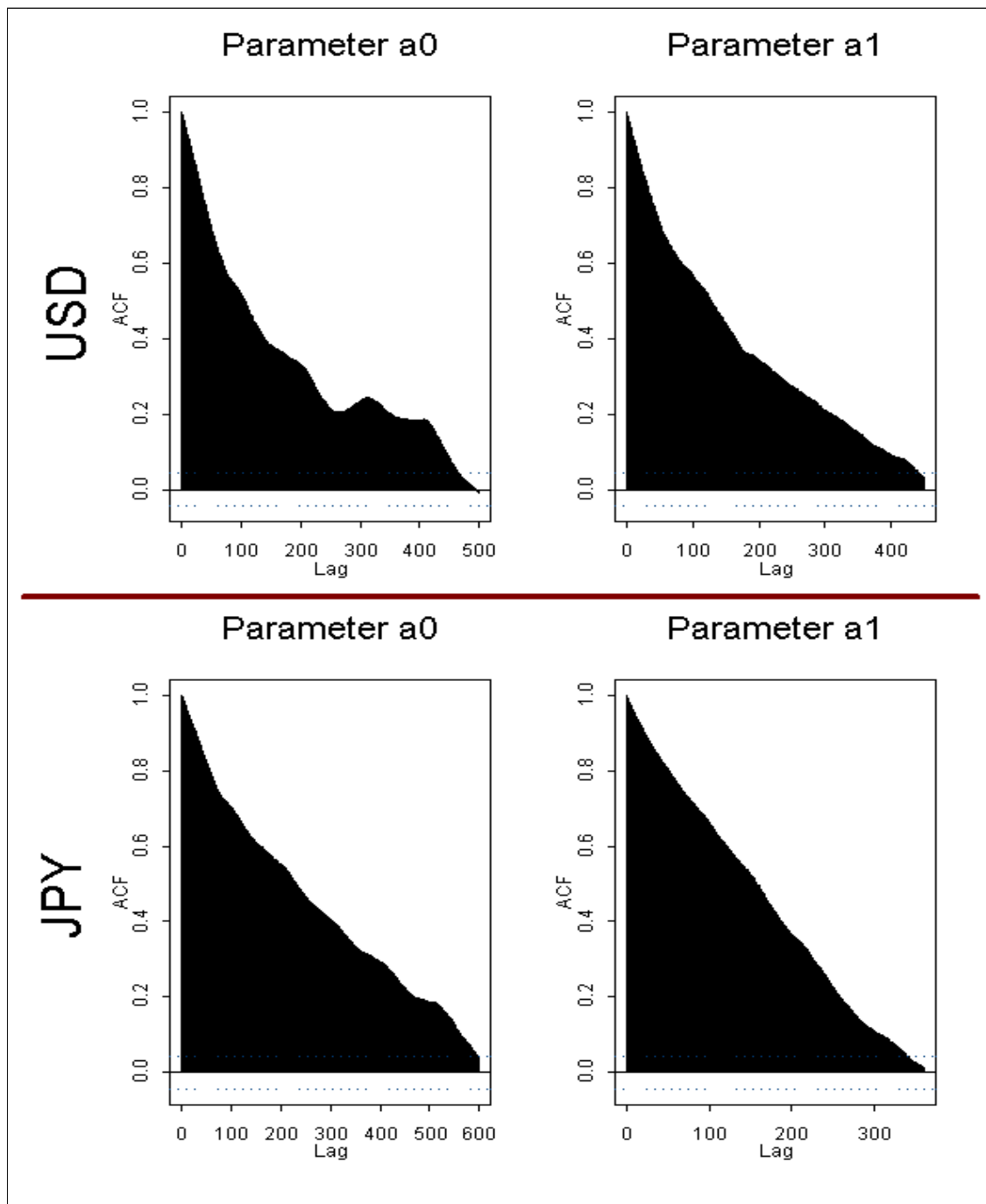


Figure 7-16: Autocorrelation function plots for the parameters of the ARCH model

Parameters	Mean	Variance
$a_0$	23.33826	1.927203
$a_1$	0.15366	0.002395

Table 7.8: Posterior Statistics for the parameters of the ARCH(1) model for USD

## Results for USD

Figure 7-17 presents the 5000 iterations of the MCMC algorithm for each parameter of interest for the case of ARCH(1) model applied to the USD. Based on this i.i.d. sample, the posterior mean and variance of the parameters of the model are estimated. The Table 7.8 presents these estimates.

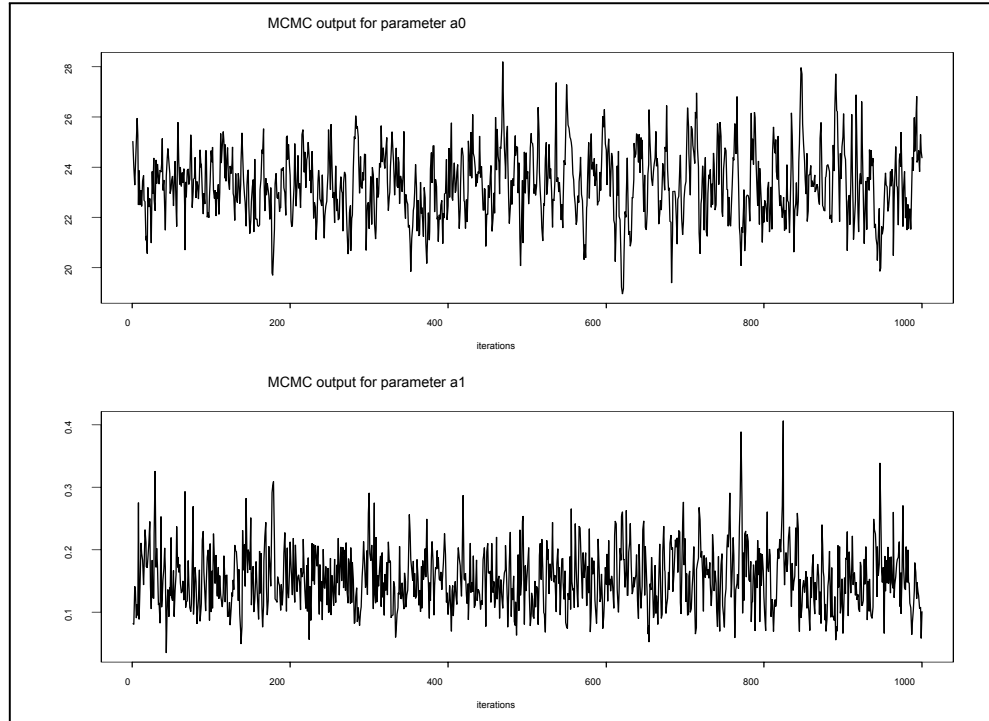


Figure 7-17: MCMC output for parameters of ARCH(1) for USD

These posterior summary statistics indicate that the USD series is “weakly” volatility persistent. The posterior histograms of the parameters  $a_0$  and  $a_1$  are presented in Figure 7-18.

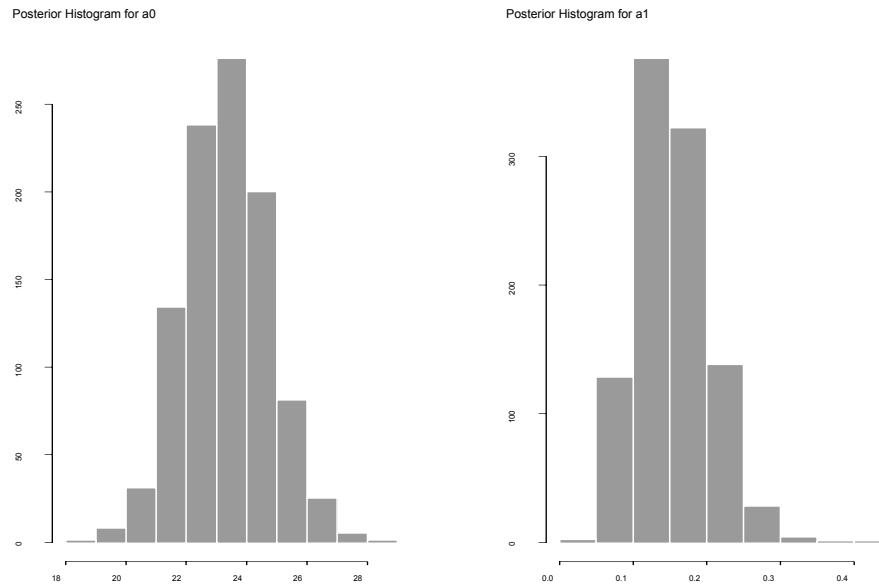


Figure 7-18: Posterior Histograms for the parameters of ARCH(1) for USD

For the convergence of the Markov chain, the subsampling diagnostic points out that the proposed algorithm needs approximately 29000 iterations to get in the target distribution. The results of the diagnostic are presented in Figure 7-19.

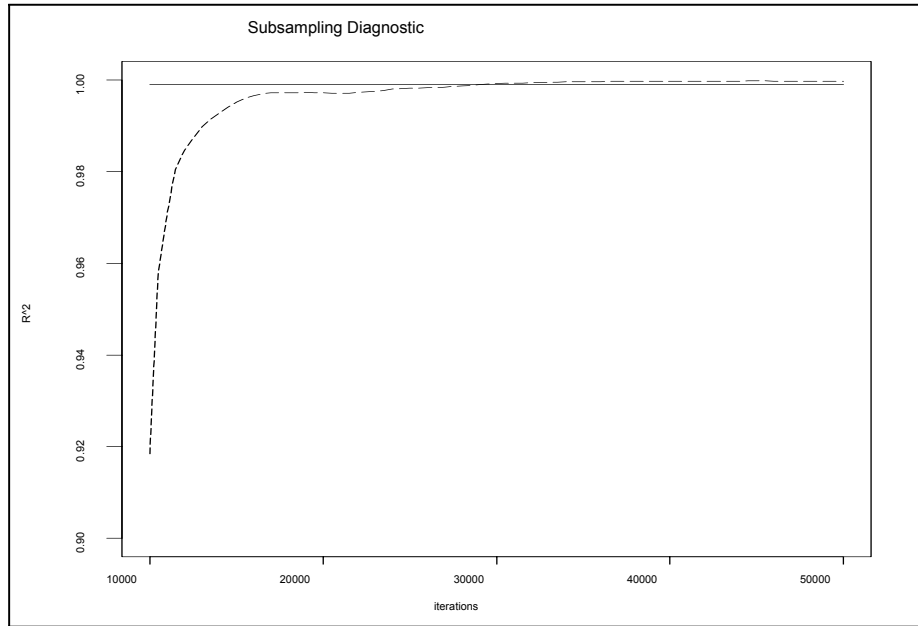


Figure 7-19: Subsampling diagnostic plot for the USD

Parameters	Mean	Variance
$a_0$	23.26966	1.859893
$a_1$	0.16612	0.002317

Table 7.9: Posterior Statistics for the parameters of the ARCH(1) model for JPY

## Results for JPY

Figure 7-20 presents the 5000 iterations of the MCMC algorithm for each parameter of interest for the case of ARCH(1) model applied to the JPY. Based on this i.i.d. sample, the posterior mean and variance of the parameters are estimated. The Table 7.9 presents these estimates.

Based on these posterior summary statistics, it can be concluded that the USD series is “weakly” volatility persistent. The posterior histograms of the parameters  $a_0$  and  $a_1$  are presented in Figure 7-21.

For the convergence of the Markov chain, the subsampling diagnostic points out that the proposed algorithm needs approximately 30000 iterations to get in the target

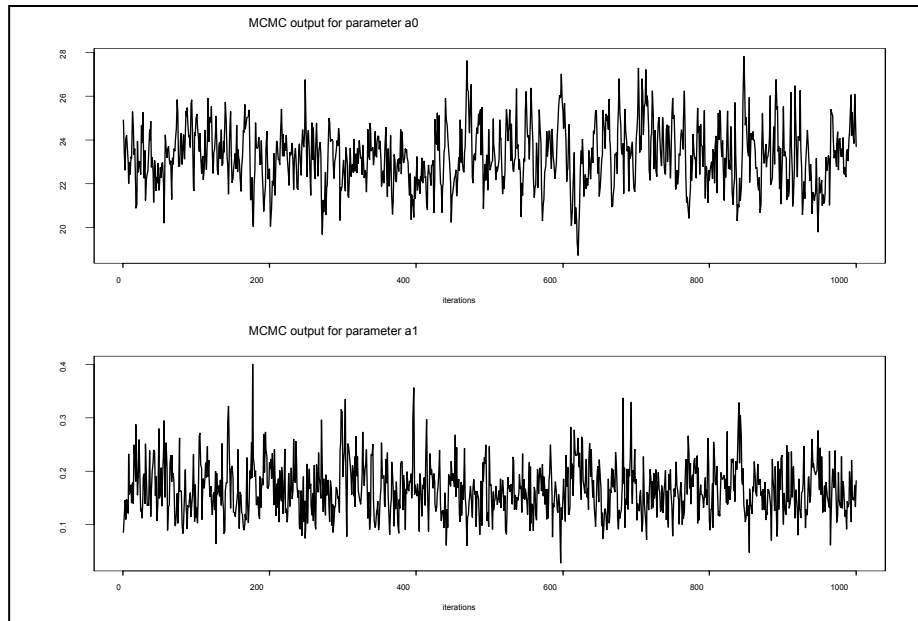


Figure 7-20: MCMC output for parameters of ARCH(1) for JPY

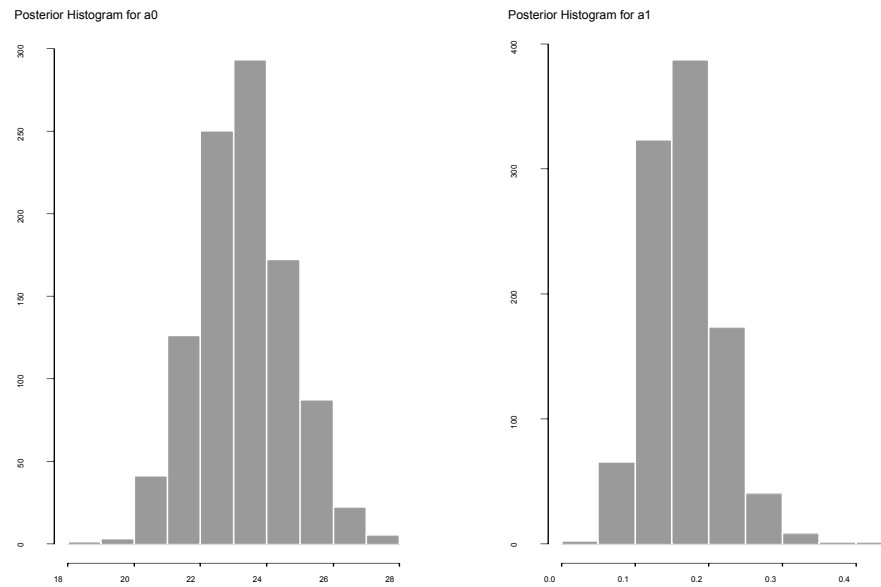


Figure 7-21: Posterior Histograms for the parameters of ARCH(1) for JPY

distribution. The results of the diagnostic are presented in Figure 7-22.

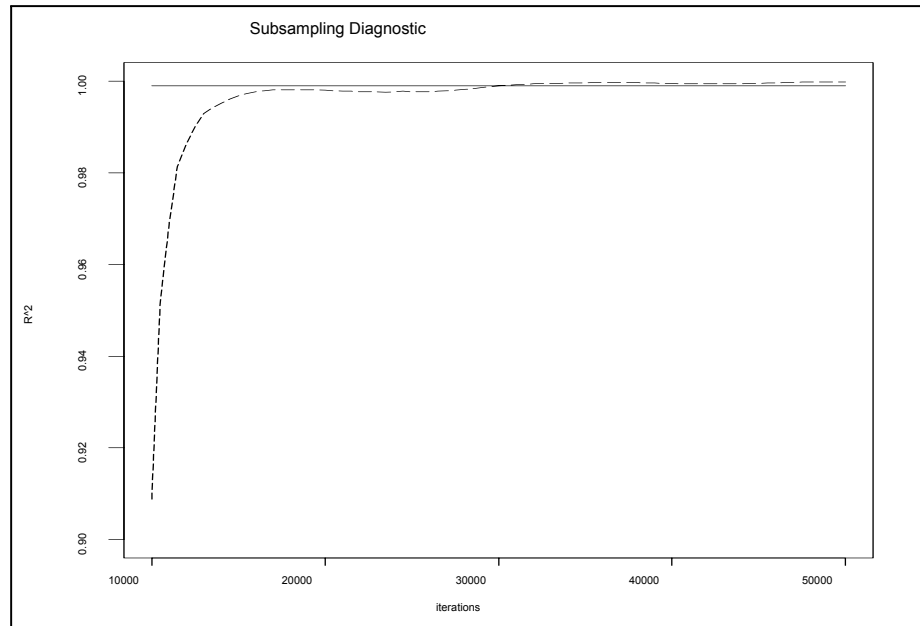


Figure 7-22: Subsampling diagnostic plot for the JPY

## 7.5 GARCH models

### 7.5.1 Introduction

In this section an auxiliary variable algorithm is proposed in order to take sample from the posterior density of the parameters of GARCH(1, 1) model (Bollerslev, 1986). The GARCH(1, 1) model can be written as

$$\begin{aligned} y_t | a_0, a_1, b_1, \mathbf{y}_{t-1} &\sim N(0, \sigma_t^2), \\ \sigma_t^2 &= a_0 + a_1 y_{t-1}^2 + b_1 \sigma_{t-1}^2, \end{aligned} \quad (7.17)$$

where  $y_t$  is the time series at time  $t$ ,  $a_0, a_1, b_1$  are the parameters of interest and  $\mathbf{y}_{t-1}$  is the vector of information up to time  $t-1$ , i.e.  $\mathbf{y}_t = (y_1, y_2, \dots, y_{t-1})$ . Sufficient, but not necessary conditions such that  $\sigma_t^2 > 0$  for all  $t$  are  $a_0 > 0$ ,  $a_1 \geq 0$ , and  $b_1 \geq 0$ . An additional restriction is imposed,  $a_1 + b_1 < 1$ , to ensure that  $y_t$  is covariance stationary (Bollerslev (1986), Theorem 1).

A Bayesian analysis of GARCH model using other MCMC algorithms has been also proposed by Bauwens and Lubrano (1998), Muler and Pole (1999), Bos, Mahieu and van Dijk (1999), Vrontos, Dellaportas and Politis (2000).

### 7.5.2 Bayesian Approach and the Auxiliary Variable Sampler

Using non-informative priors for the parameters of interest,  $\pi(a_0, a_1, b_1) \propto a_0^{-1}$ , the joint posterior distribution -known up to a constant- for  $a_0, a_1$  and  $b_1$  takes the following form:

$$\begin{aligned} f(a_0, a_1, b_1 | \mathbf{y}) &\propto \left( a_0 \prod_{i=1}^T \sqrt{a_0 + a_1 y_{i-1}^2 + b_1 \sigma_{i-1}^2} \right)^{-1} \\ &\exp \left\{ -0.5 \sum_{i=1}^T \frac{y_i^2}{a_0 + a_1 y_{i-1}^2 + b_1 \sigma_{i-1}^2} \right\}. \end{aligned} \quad (7.18)$$

The full conditional densities for  $a_0, a_1$  and  $b_1$  are not of standard forms, therefore

Gibbs algorithm cannot be applied here. In order to handle this problem the simple auxiliary variable sampler is adopted such as the resulting MCMC algorithm to be consisted of only Gibbs steps. In the following Theorem one auxiliary variable is introduced in the posterior density (7.18) such as the resulting full conditional densities for  $a_0, a_1$  and  $b_1$  to be of standard forms.

**Theorem 17** *Let the posterior density given by (7.18). If we include one positive latent variables  $u$ , in (7.18) such as the resulting joint density for  $a_0, a_1, b_1$  and  $u$  to be given by*

$$f(a_0, a_1, b_1, u | \mathbf{y}) \propto \mathbf{I}(u < f(a_0, a_1, b_1 | \mathbf{y})) \quad (7.19)$$

$$\mathbf{I} \left( u < \left( \prod_{t=1}^T \sqrt{\sigma_t^2} \right)^{-1} \exp \left\{ -0.5 \sum_{i=1}^T \frac{y_t^2}{\sigma_t^2} \right\} \right).$$

then, the marginal density  $f(a_0, a_1, b_1 | \mathbf{y})$  is given by (7.18).

**Proof:** *It is enough to integrate out  $w$  from the resulting posterior density (7.19)*

$$f(a_0, a_1, b_1 | \mathbf{y}) = \int_u f(a_0, a_1, b_1, u, | \mathbf{y}) du = \int_0^{f(a_0, a_1, b_1 | \mathbf{y})} du$$

$$= \left( a_0 \prod_{i=1}^T \sqrt{a_0 + a_1 y_{t-1}^2 + b_1 \sigma_{t-1}^2} \right)^{-1}$$

$$\exp \left\{ -0.5 \sum_{i=1}^T \frac{y_t^2}{a_0 + a_1 y_{t-1}^2 + b_1 \sigma_{t-1}^2} \right\} \cdot \blacksquare$$

Using the posterior density (7.19) an MCMC algorithm which is consisted only of Gibbs steps can be easily constructed. In detail, the proposed algorithm is consisted of the following steps

1. Give initial values  $a_0^0, a_1^0, b_1^0$ .
2. Sample  $u^1 | a_0^0, a_1^0, b_1^0, \mathbf{y} \sim \mathbf{U}(0, f(a_0^0, a_1^0, b_1^0 | \mathbf{y}))$ .
3. Sample  $a_0^1$  from  $\mathbf{U}(\cdot, \cdot)$  such as  $\{u^1 < f(a_0^1, a_1^0, b_1^0 | \mathbf{y})\}$ .

4. Sample  $a_1^1$  from  $\mathbf{U}(\cdot, \cdot)$  such as  $\{u^1 < f(a_0^1, a_1^1, b_1^0 | \mathbf{y})\}$ .
5. Sample  $b_1^1$  from  $\mathbf{U}(\cdot, \cdot)$  such as  $\{u^1 < f(a_0^1, a_1^1, b_1^1 | \mathbf{y})\}$ .
6. Repeat from step 2.

In the above algorithm the  $f(a_0, a_1, b_1 | \mathbf{y})$  is not invertible with respect to its parameters  $a_0, a_1$  and  $b_1$ . For this reason, Neal's approach (2003) is adopted. An interval  $I = (L, R)$  is taken around the parameter of interest that contains at least a big part of set  $S = \{u < f(a_0, a_1, b_1 | \mathbf{y})\}$ . More specifically, for parameter  $a_0$  an initial interval  $I = (L, R)$  of size  $w$  is randomly picked that contains  $a_0^0$  and this interval is expanded using the 'stepping out' procedure (Neal, 2003) to contain a big part of  $S$ . Then,  $a_0$  is sampled uniformly from the interval  $I$  and the interval is narrowed - using the shrink procedure of Neal (2003) - each time that a new point is sampled out of  $S$ , until an  $a_0$  from  $S$  is sampled. For the parameters  $a_1$  and  $b_1$  the initial intervals are set equal to  $I_{a_1} = (0, 1 - b_1)$  and  $I_{b_1} = (0, 1 - a_1)$  respectively. Then,  $a_1$  and  $b_1$  are uniformly sampled from  $I_{a_1}$  and  $I_{b_1}$  respectively and the intervals are shrank- using the shrink procedure- each time that a new point is sampled out of  $S$ .

### 7.5.3 Some Applications

The above algorithm for the estimation of the parameters of the GARCH(1) model is applied to two series of data sets. Both of them are consisted of 844 daily exchange rates multiplied by 1000. In detail, the daily exchange rate of the US dollar (USD) and the Japanese Yen (JPY) with respect to the Greek Drachma (GRD) are used (Figure 7-1, page 98). The initial 50000 iterations of the algorithm were dropped as burn-in and 1 sample point out of 100 iterations is kept such that the final samples, which consist of 5000 values, to be approximately, independent and identically distributed samples from the marginal densities of the parameters of interest.

The final posterior samples were checked for convergence to the limiting distribution by the criteria of Geweke (1992), Raftery and Lewis (1992) and Heidelberger and Welch

	USD	JPY
$a_0$	40	160
$a_1$	31	30
$b_1$	39	160

Table 7.10: Autocorrelation function results for the parameters of the GARCH model

(1983). Apart from this result, the subsampling diagnostic (see section 4.3 and Giakoumatos *et al.*, 1999) was applied to the initial 50000 iterations of each chain in order to check the convergence ( $a = 0.05, t = 0.90, d = 0.999$ ). Note that, the subsampling diagnostic is considered by its authors (Giakoumatos, Vrontos, Dellaportas and Politis, 1999) ‘very conservative’.

Table 7.10 presents the number of lagged values where the autocorrelation dies out for each series of data. Graphically, these results are presented in Figure 7-23.

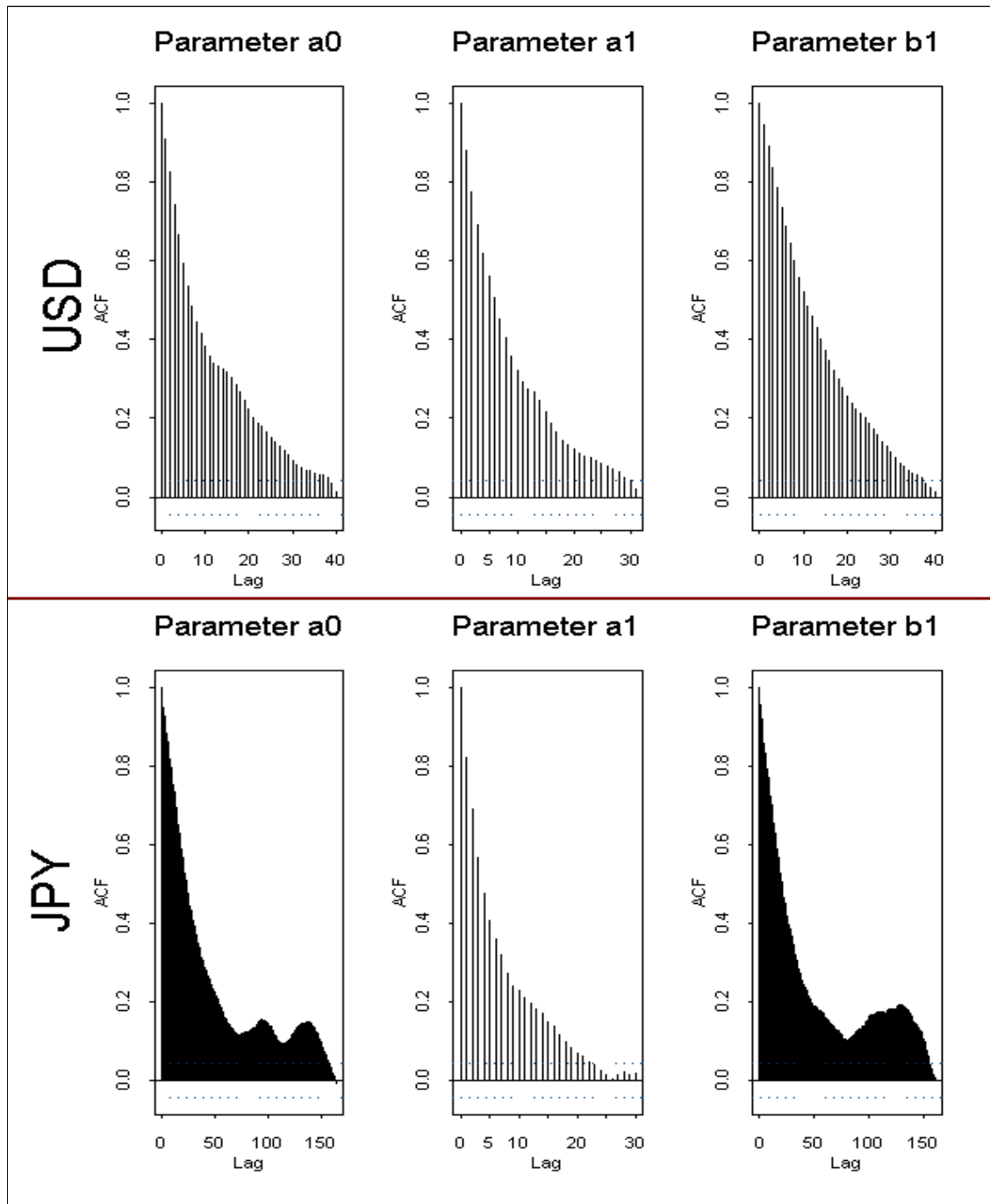


Figure 7-23: Autocorrelation function plots for the parameters of the GARCH model

Parameters	Mean	Variance
$a_0$	0.940828	0.0871954
$a_1$	0.065460	0.0002280
$b_1$	0.902724	0.0003988

Table 7.11: Posterior Statistics for the parameters of the GARCH(1) model for USD

## Results for USD

Figure 7-24 presents the 5000 iterations of the MCMC algorithm for each parameter of interest for the case of GARCH(1,1) model applied to the USD. Based on this i.i.d. sample, the posterior mean and variance of the parameters of the models are estimated. Table 7.11 presents these estimates.

These posterior summary statistics indicate that the USD series is 'heavily' volatility persistent ( $a_1 + b_1 = 0.968174$ ). The posterior histograms of the parameters  $a_0$ ,  $a_1$  and  $b_1$  are presented in Figure 7-25.

For the convergence of the Markov chain, the subsampling diagnostic points out that the proposed algorithm needs approximately 27000 iterations to get in the target distribution. The results of the diagnostic are presented in Figure 7-26.

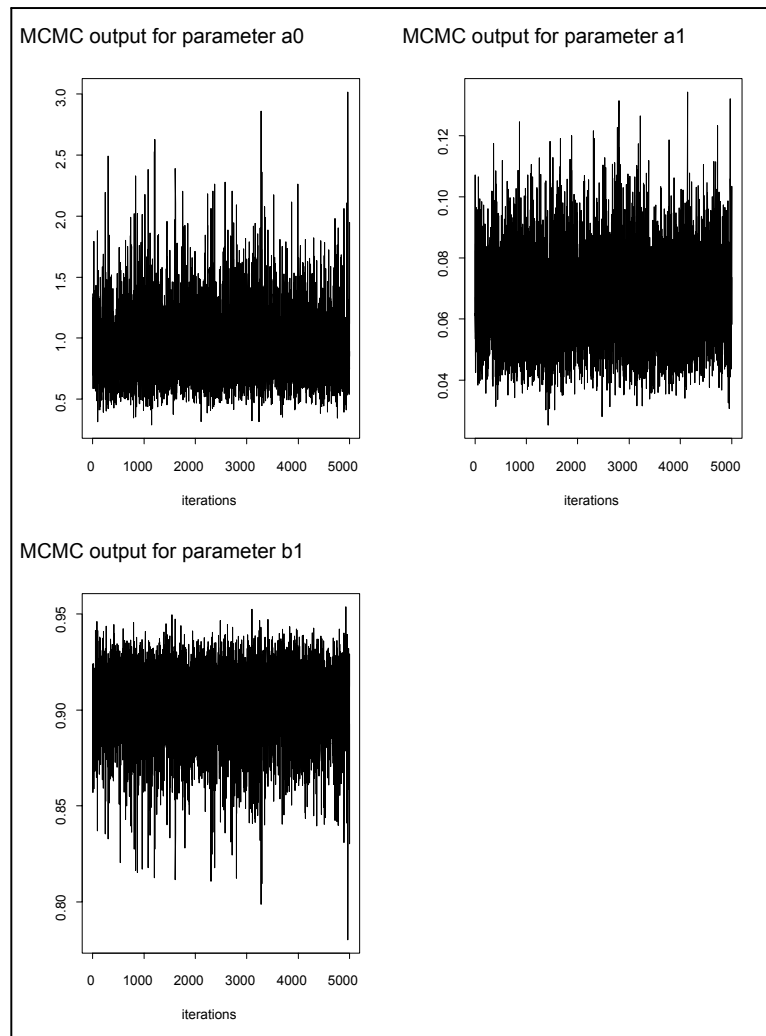


Figure 7-24: MCMC output for the parameters of the GARCH(1.1) for USD

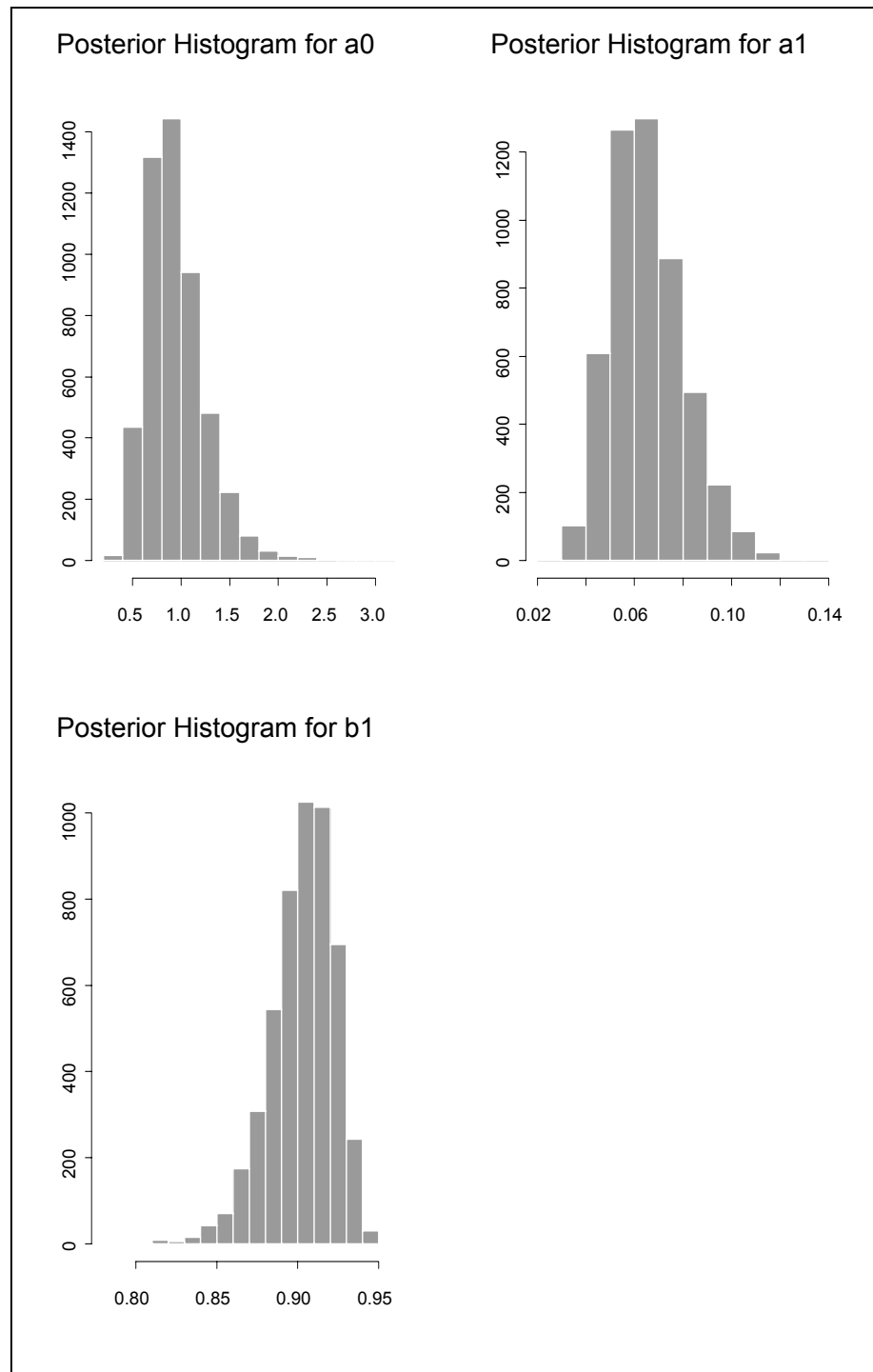


Figure 7-25: Posterior Histograms for the parameters of the GARCH(1,1) for USD

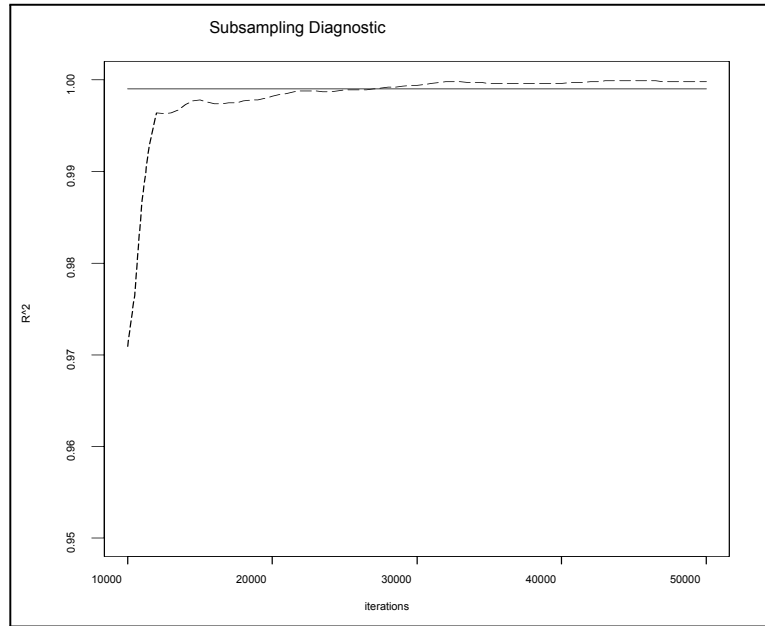


Figure 7-26: Subsampling diagnostic plot for the GARCH(1,1) for USD

Parameters	Mean	Variance
$a_0$	2.934831	1.4824450
$a_1$	0.123073	0.0009159
$b_1$	0.779819	0.0030641

Table 7.12: Posterior Statistics for the parameters of the GARCH(1) model for JPY

## Results for JPY

Regarding the case where GARCH(1,1) model is applied to JPY series, Figure 7-27 presents the 5000 iterations of the MCMC algorithm for each parameter of interest. Based on this i.i.d. sample, the posterior mean and variance of the parameters are estimated. The Table 7.12 presents these estimates.

These posterior summary statistics indicate that the JPY series is volatility persistent ( $a_1 + b_1 = 0.902892$ ). The posterior histograms of the parameters  $a_0$ ,  $a_1$  and  $b_1$  are presented in Figure 7-28.

For the convergence of the Markov chain, the subsampling diagnostic points out

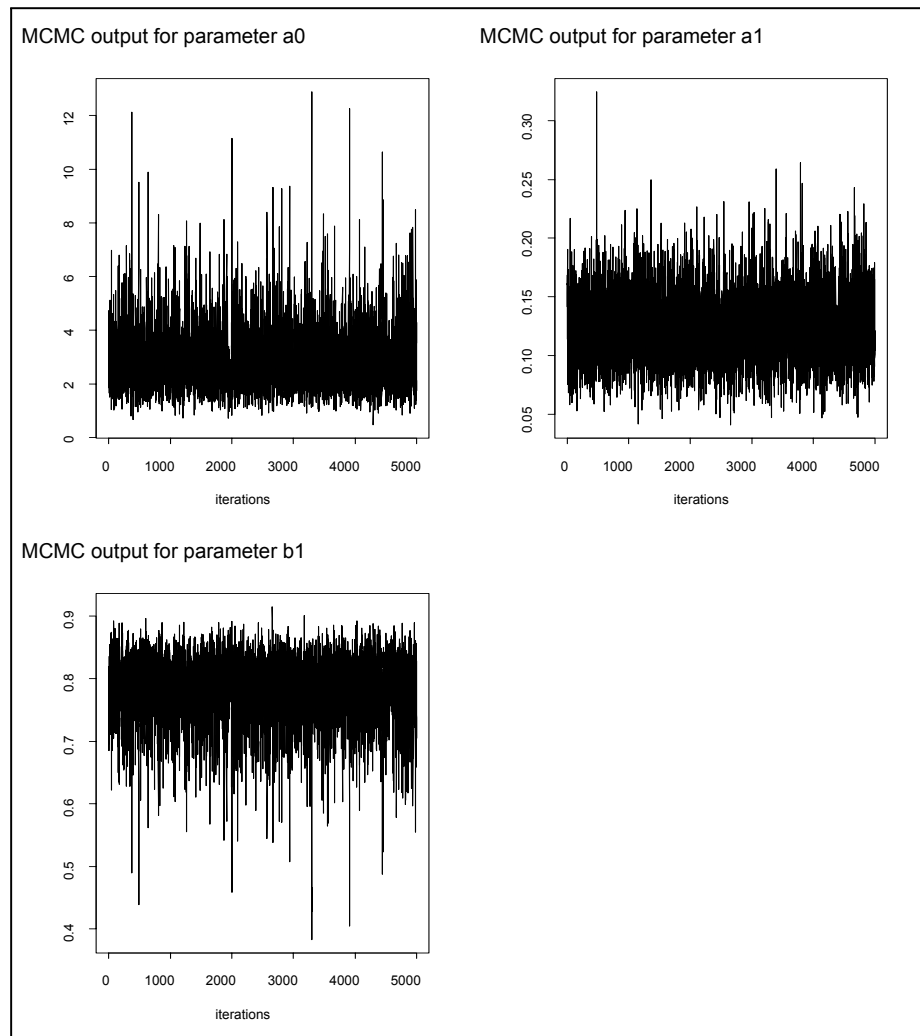


Figure 7-27: MCMC output for the parameters of the GARCH(1.1) for JPY

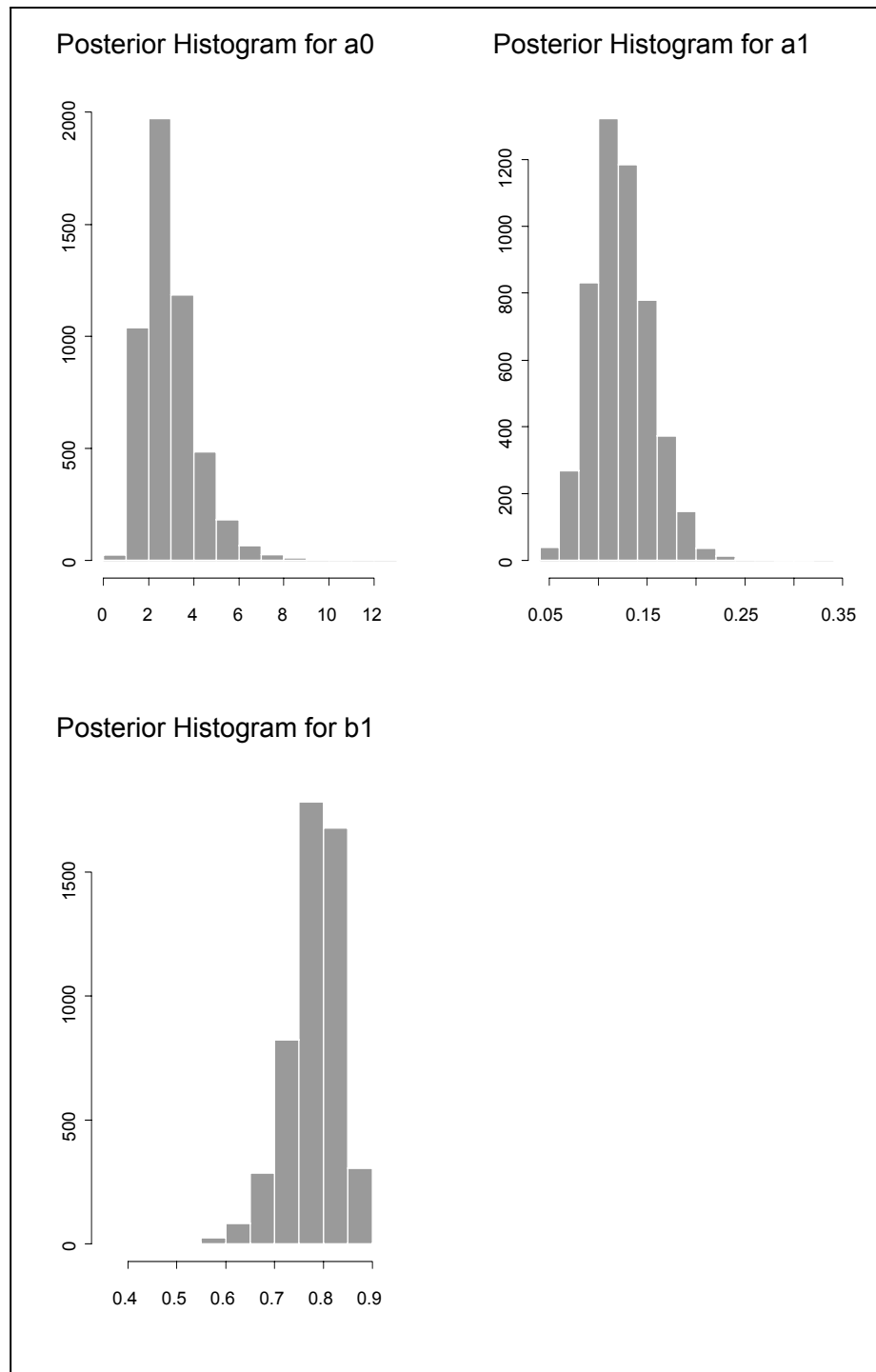


Figure 7-28: Posterior Histograms for the parameters of the GARCH(1,1) for JPY

that the proposed algorithm needs approximately 19000 iterations to get in the target distribution. The results of the diagnostic are presented in Figure 7-29.

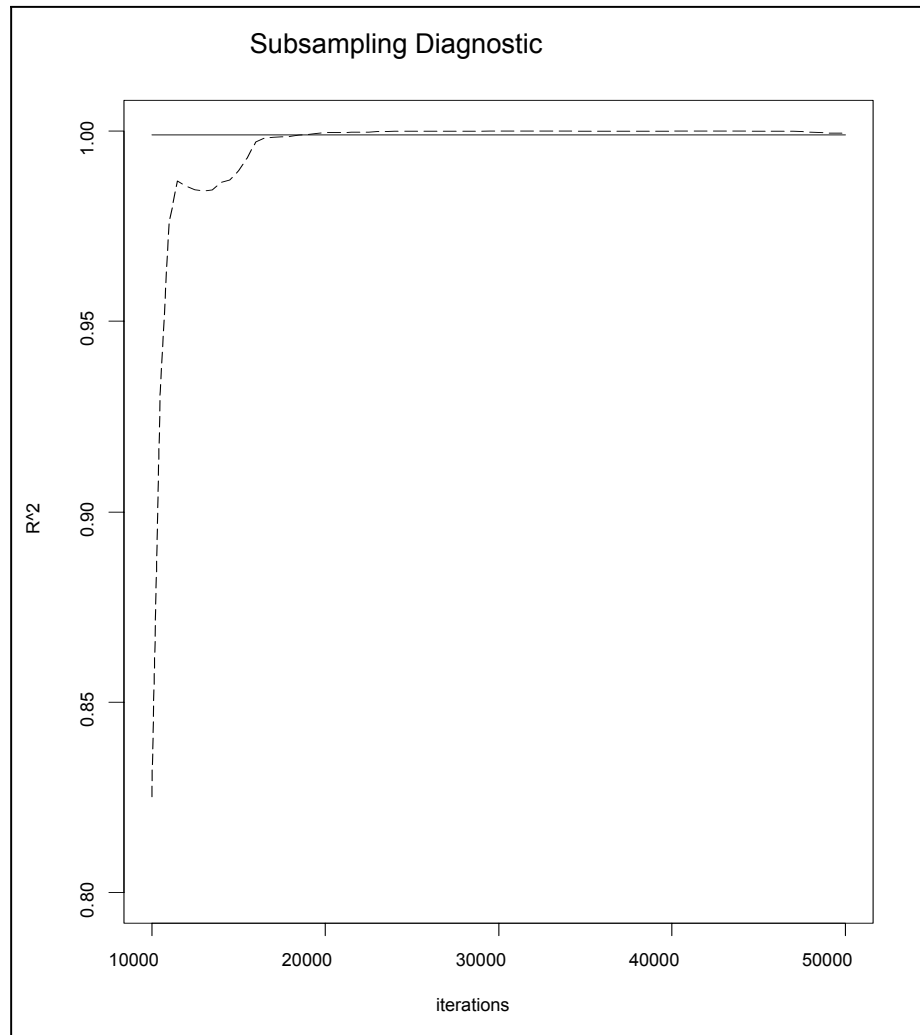


Figure 7-29: Subsampling diagnostic plot for the GARCH(1,1) for JPY

# Chapter 8

## Comparison of the Univariate Models

### 8.1 Introduction

In this chapter the univariate models that have been presented and analyzed in the previous chapter are compared. The model comparison method which is chosen is based on the predictive distributions of time-varying volatilities. This choice is sensible due to the predictive nature of the applicability of these models and due to the fact that volatilities are of primary interest. As a criterion for the model comparison one of those suggested by Gelfand, Dey and Chang (1992), has been chosen. Similar diagnostic measures were presented by Pitt and Shephard (1999a, 1999b) in order to compare stochastic volatility models.

### 8.2 The Comparison Method

The MCMC techniques for the time-varying volatility models that were described on the previous chapter produce sample points from the joint posterior  $\pi\left(\boldsymbol{\theta}^{(t)} \mid \Phi_t\right)$ , where  $\Phi_t = (y_1, \dots, y_t)'$  and  $\boldsymbol{\theta}_s^{(t)}$ ,  $s = 1, \dots, B$ , denotes the MCMC output for the model parameters

based on samples of size  $B$ . Hence,  $\boldsymbol{\theta}_s^{(t)}$  can be used to carry out computations required for model comparison. The estimate of the predictive density is computed using the outcome of the MCMC algorithm:

$$\hat{\pi}(y_t | \Phi_{t-1}) = B^{-1} \sum_{s=1}^B \pi(y_t | \boldsymbol{\theta}_s^{(t-1)}, \Phi_{t-1}).$$

Using the above equation and  $\boldsymbol{\theta}_s^{(T-R)}, \boldsymbol{\theta}_s^{(T-R+1)}, \dots, \boldsymbol{\theta}_s^{(T-1)}$   $R$  one-step-ahead predictive densities  $\hat{\pi}(y_{T-R+1} | \Phi_{T-R}), \hat{\pi}(y_{T-R+2} | \Phi_{T-R+1}), \dots, \hat{\pi}(y_T | \Phi_{T-1})$  are constructed.

Let  $Z_t$  be the random variable representing a future point. Then, model  $M_1$  ( $M_2$ ) is chosen according to whether  $D > 0$  ( $D < 0$ ), where

$$D = \log \left( \frac{\prod_{t=T-R+1}^T \hat{\pi}(Z_t = y_t | \Phi_{t-1}; M_1)}{\prod_{t=T-R+1}^T \hat{\pi}(Z_t = y_t | \Phi_{t-1}; M_2)} \right),$$

and  $\mathbf{y}_t$  is the realization of the stochastic process at time  $t$ . Note that  $\exp(D)$  is called pseudo-Bayes factor (Gelfand, Dey and Chang 1992).

The above method is straightforward when applied to the case of ARCH and GARCH model. Unfortunately, for the case of the unobserved ARCH and stochastic volatility model the estimation of the predictive distributions for the last  $R$  data points is time consuming. For this reason Pitt and Shephard (1999a, 1999b) methodology is followed and  $\pi(Z_t = y_t | \Phi_{t-1}; M)$  is estimated via filtering methods using the Bayesian mean  $\bar{\boldsymbol{\theta}}^{(T-R)}$  for the parameters of the model of interest. In particular, an estimate of  $\pi(Z_t = y_t | \Phi_{t-1}, \bar{\boldsymbol{\theta}}^{(T-R)}; M)$  is given as follows.

- Firstly, samples of size  $B$  is obtained (in the application it has be chosen  $B$  to be equal to 10000) of the unobserved component of the unobserved ARCH model and the unobserved volatilities of the stochastic volatility model, for  $t = T-R+1, \dots, T$ , using  $\mathbf{f}_t^i \sim \pi(\mathbf{f}_t | \Phi_{t-1}, \bar{\boldsymbol{\theta}}^{(T-R)}; M)$  and  $\mathbf{h}_t^i \sim \pi(\mathbf{h}_t | \Phi_{t-1}, \bar{\boldsymbol{\theta}}^{(T-R)}; M)$ , respectively, for

$i = 1, \dots, B$ . These densities are easily derived for both models.

Then an estimate is given by

$$\hat{\pi} \left( Z_t = y_t | \Phi_{t-1}, \bar{\boldsymbol{\theta}}^{(T-R)}; M \right) = \frac{1}{K} \sum_{i=1}^K \hat{\pi} \left( Z_t = y_t | \mathbf{f}_t^i, \bar{\boldsymbol{\theta}}^{(T-R)}; M \right),$$

which is the result of the Monte Carlo integration of

$$\pi \left( \mathbf{y}_t | \Phi_{t-1}, \bar{\boldsymbol{\theta}}^{(T-R)}; M \right) = \int \pi \left( \mathbf{y}_t | \mathbf{f}_t, \bar{\boldsymbol{\theta}}^{(T-R)}; M \right) \pi \left( \mathbf{f}_t | \Phi_{t-1}, \bar{\boldsymbol{\theta}}^{(T-R)}; M \right) d\mathbf{f}_t.$$

This technique presupposes that the densities  $\pi \left( \mathbf{y}_t | \mathbf{f}_t, \bar{\boldsymbol{\theta}}^{(T-R)}; M \right)$  can be evaluated and simulated. This is valid for the unobserved ARCH model and stochastic volatility model ; see equation (7.8) and equation (7.1) respectively. For more details about this technique and the filtering method see Pitt and Shephard (1999a).

## 8.3 The Results

The MCMC algorithms that were presented in the previous chapter have been applied using  $T = 844$  daily exchange rates. In detail, the daily exchange rate of the US dollar (USD) and the Japanese Yen (JPY) with respect to the Greek Drachma (GRD) are used (Figure 7-1, page 98). In order to compare these models, Bayesian analysis is used by estimating the predictive distributions  $\pi(\mathbf{y}_t | \Phi_{t-1})$  under each model. In this implementation the first 820 real data points are used, and  $R = 24$  one-step-ahead predictive distributions  $\pi(y_t | \Phi_{t-1})$  are estimated for  $t = 821, \dots, 844$  data points. For each one of those 24 time periods the predictive densities  $\pi(y_t | \Phi_{t-1})$  are estimated at the real data point  $y_t$ , denoted by  $\hat{\pi}(y_t | \Phi_{t-1})$ , under each model, as described in section 8.2. As a criterion, the

$$D = \sum_{t=T-R+1}^T \log(\hat{\pi}(Z_t = y_t | \Phi_{t-1}; M_1)) - \sum_{t=T-R+1}^T \log(\hat{\pi}(Z_t = y_t | \Phi_{t-1}; M_2)),$$

<b>Model</b>	<b>USD</b>	<b>JPY</b>
ARCH(1)	-69.6882327	-70.47012514
GARCH(1,1)	-69.65541316	-70.41296153
Unobserved ARCH	-68.05946666	-70.13310226
Stochastic volatility	-67.33219424	-69.54337267

Table 8.1: Model comparison

is used and the model  $M_1$  ( $M_2$ ) is chosen in case  $D > 0$  ( $D < 0$ ). The quantity

$$\sum_{t=T-R+1}^T \log (\hat{\pi}(Z_t = y_t | \Phi_{t-1}; M))$$

is estimated for the models under consideration and the results are presented in Table 8.1 for the ARCH, GARCH, unobserved ARCH and stochastic volatility models respectively.

According to these results, for the two exchange rates the stochastic volatility model seems to be more preferable for the one-ahead predictions in comparison to the Unobserved ARCH model which is preferable than GARCH(1,1) model and ARCH(1) model. Analogous results have been found by Kim, Shephard and Chib (1998) for the comparison of the univariate GARCH and stochastic volatility models.

# Chapter 9

## Existing Multivariate Parameter-Driven Models

### 9.1 Introduction

Lately, there is a large scientific interest in the study of multivariate time-varying volatility models (Chib *et al.*, 1999 and 2002). These models can capture not only the changing volatility of the financial time series but also the correlation structure existing between two or more time series. Regarding the multivariate observation-driven models (ARCH/GARCH), a variety of different models has been proposed (Kraft and Engle, 1983; Bollerslev *et al.*, 1988; Bollerslev 1990; Vrontos 2001).

This chapter focuses on the multivariate parameter-driven models. This specific class of time-varying volatility models contains not only the versions of the multivariate stochastic volatility models but also includes versions of the latent factor stochastic volatility model (Pitt and Shephard, 1999a; Aguilar, Huerta, Prado and West 1999), the multivariate unobserved ARCH model and the latent factor ARCH/GARCH models (King, Sentana and Wadhwani, 1994; Diebold and Nerlove, 1989). In the following sections the most important of these multivariate parameter-driven models will be presented along with the methods that have been proposed for the estimation of their parameters.

## 9.2 Multivariate Stochastic Volatility Models

The most known multivariate stochastic volatility model has been proposed by Harvey, Ruiz and Shephard (1994). In detail, the authors proposed a straightforward multivariate version of the univariate stochastic volatility model.

This model can be written as

$$\mathbf{y}_t | \cdot \sim MN_k(\mathbf{0}, \mathbf{H}),$$

where

$$\mathbf{H}_t = \begin{bmatrix} \sigma_1^2 \exp(h_{1,t}) & 0 & \cdots & 0 \\ 0 & \sigma_2^2 \exp(h_{2,t}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_k^2 \exp(h_{k,t}) \end{bmatrix},$$

and

$$\mathbf{h}_t | \cdot \sim MN_k(\mathbf{a} + \mathbf{B}\mathbf{h}_{t-1}, \mathbf{\Sigma}_h),$$

where  $\mathbf{y}_t$  is the  $k$ - variate vector of the realization of the time series at time  $t$ ,  $\mathbf{h}_t = (h_1, h_2, \dots, h_k)'$  is the vector of volatilities at time  $t$ ,  $\mathbf{a} = (a_1, a_2, \dots, a_k)'$ ,  $\mathbf{\Sigma}_h$  is a full covariance matrix of the volatilities and

$$\mathbf{B} = \begin{bmatrix} b_1 & 0 & \cdots & 0 \\ 0 & b_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b_k \end{bmatrix}.$$

Harvey, Ruiz and Shephard (1994) proposed to approximate this model by using the process  $\log(y_{it}^2)$  in the context of Gaussian state space model. This linearization of the stochastic volatility model introduces non-Gaussian errors in the measurement equation (Pitt, 1997). The authors treat these non-gaussian errors as Gaussian because the

Kalman Filter can be used. Unfortunately, the resulting estimates seem to be inefficient due to the previous assumption. Another attempt was made by Mahieu and Schotman (1998) who estimated the parameters of the above model using stochastic EM algorithm.

A serious disadvantage of multivariate stochastic volatility model is that it has a very large number of parameters. For this reason, many scientists directed their research to latent factor structure models.

### 9.3 Factor Stochastic Volatility Models

Apart from the straightforward extensions of the stochastic volatility model to the multivariate case, a very popular approach in modeling covariance matrices in financial literature is the latent factor structure.

Pitt and Shephard (1999a) and Pitt (1997) proposed a version of the latent factor model that incorporates stochastic volatility processes. This model is presented in the following hierarchical formula of conditional densities

$$\begin{aligned}\mathbf{y}_t | \cdot &\sim MN_n(\mathbf{B}\mathbf{f}_t, \mathbf{\Sigma}_w), \\ \mathbf{f}_t | \cdot &\sim MN_k(\mathbf{0}, D(\mathbf{h}_t) \mathbf{I} D(\mathbf{h}_t)), \\ \mathbf{h}_t | \cdot &\sim MN_k(\mathbf{L}\mathbf{h}_{t-1}, \mathbf{H}\mathbf{I}\mathbf{H}),\end{aligned}$$

where  $\mathbf{y}_t$  is a  $n$  – *variate* vector of stochastic process at time  $t$ ,  $\mathbf{B}$  is the  $n \times k$  matrix of factor weights,  $\mathbf{f}_t$  is the  $k \times 1$  common factors, the  $n \times n$  matrix  $\mathbf{\Sigma}_w$  and the  $k \times k$  matrix  $\mathbf{H}$  are full covariance matrices, and  $D(\mathbf{h}_t)$  is a diagonal  $k \times k$  matrix with elements  $D_{ii} = \exp(h_t/2)$ . The  $k \times 1$  vector  $\mathbf{h}_t$  are the volatilities that follow a *VAR* process and the  $k \times k$  matrix  $\mathbf{L}$  is the hyperparameters of the stochastic volatility part. Pitt (1997) considered the matrix  $\mathbf{L}$  to be diagonal with all of its diagonal elements less than 1 to ensure stationarity for the process of the volatilities  $\mathbf{h}_t$ . Moreover the diagonal elements of  $\mathbf{B}$  is restricted to be 1 for identifiability reasons. Pitt (1997) and Pitt and Shephard

(1999a) proposed to sample from the joint posterior density of the aforementioned model using a MCMC algorithm. In detail, they used a block sampling scheme based on Pitt's (1997) MCMC algorithm for the case of univariate stochastic volatility model.

Aguilar and West (2000) proposed a similar version for the factor stochastic volatility model. This model can be written as

$$\begin{aligned}\mathbf{y}_t | \cdot &\sim MN_n(\mathbf{m} + \mathbf{B}\mathbf{f}_t, \mathbf{\Sigma}_w), \\ \mathbf{f}_t | \cdot &\sim MN_k(\mathbf{0}, \mathbf{H}_t), \\ \mathbf{h}_t^* | \cdot &\sim MN_k(\mathbf{m}_{h^*} + \mathbf{L}\mathbf{h}_{t-1}^*, \mathbf{U}),\end{aligned}$$

where  $\mathbf{y}_t$  is a  $n$  - *variate* vector of stochastic process at time  $t$ ,  $\mathbf{B}$  is a  $n \times k$  matrix of factor weights,  $\mathbf{f}_t$  is the  $k \times 1$  common factors, the  $n \times n$  matrix  $\mathbf{\Sigma}_w$  is a full covariance matrix and the diagonal  $k \times k$  matrix  $\mathbf{H}$  contains the volatilities  $h_{t,i}$ . The vector of *log*- volatilities  $\mathbf{h}_t^* = (\log(h_{t,1}), \log(h_{t,2}), \dots, \log(h_{t,k}))$  follow a *VAR* process and the  $k \times k$  matrix  $\mathbf{L}$  is the hyperparameters of the stochastic volatility part. For identification reasons, Aguilar and West (2000) set the matrix  $\mathbf{B}$  to

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ b_{21} & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ b_{k,1} & b_{k,2} & b_{k,3} & \cdots & 1 \\ b_{k+1,1} & b_{k+1,2} & b_{k+1,3} & \cdots & b_{k+1,k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{n,1} & b_{n,2} & b_{n,3} & \cdots & b_{n,k} \end{bmatrix}$$

It becomes evident that the order of the univariate time series  $y_{t,i}$  defines the estimation of the elements of matrix  $\mathbf{B}$ . The first time-series is the first factor plus a noise term and so forth.

For the estimation of parameters of this model, Aguilar and West (2000) adopted a

Metropolis-Hastings MCMC algorithm to take sample from the posterior density.

A more simplified factor stochastic volatility model has been proposed by Jacquier, Polson and Rossi (1999). This model can be presented in the following three stage hierarchical formula of conditional densities

$$\begin{aligned} \mathbf{y}_t | \cdot &\sim MN_n(\mathbf{m} + \mathbf{B}\mathbf{F}_t + \boldsymbol{\Sigma}), \\ \mathbf{f}_t | \cdot &\sim MN_k(\mathbf{0}, \mathbf{diag}(h_1, \dots, h_k)), \\ \log(h_{it}) | \cdot &\sim N(a_i + b_i \log(h_{i,t-1}), v_i^2), \end{aligned}$$

where  $\mathbf{y}_t$  is a  $n$  - variate vector of stochastic process at time  $t$ ,  $\mathbf{B}$  is a  $n \times k$  matrix of factor loadings,  $\mathbf{F}_t$  is the  $k \times 1$  common factors, the  $n \times n$  matrix  $\boldsymbol{\Sigma}$  is a full covariance matrix. The unobserved factors  $\mathbf{f}_t = (f_{1,t}, \dots, f_{k,t})$  follow independent Normal distributions with variances,  $h_1, \dots, h_k$ . Each of the  $k$  log-volatilities,  $(\log(h_{t,1}), \log(h_{t,2}), \dots, \log(h_{t,k}))$  follows an  $AR(1)$  process - as the univariate stochastic volatility model - with hyperparameters  $\theta = (\dots, a_i, b_i, v_i^2, \dots)$ . The aforementioned restrictions of Aguilar and West (2000) can be set in such a manner that the model to become identifiable. Another approach, is to set the variance of its factor  $f_{i,t}$ ,  $i = 1, \dots, k$  equal to one, i.e. to set  $a_i = -0.5v_i^2 / (1 + b_i)$ .

Jacquier *et al.* (1999) proposed MCMC algorithms for the case where the factors are observed and for the case of unobserved factors. For both cases they proposed a Metropolis within Gibbs algorithm in order to take sample from the posterior density of this model. This algorithm is similar with the one proposed by the aforementioned authors for the case of multivariate stochastic volatility model.

## 9.4 Latent Factor ARCH Models

Diebold and Nerlove (1989) proposed a latent factor model where the unobserved factors come from a ARCH(12) process. In detail, this model is presented in the following

hierarchical formula of conditional densities

$$\begin{aligned} \mathbf{y}_t | \cdot &\sim MN_n(\boldsymbol{\lambda} f_t, \boldsymbol{\Sigma}_w), \\ f_t | \cdot &\sim N(0, \sigma_t^2), \\ \sigma_t^2 &= a_0 + \frac{1 - a_0}{78} \sum_{i=1}^{12} (13 - i) f_{t-i}^2, \end{aligned}$$

where  $\mathbf{y}_t$  is a  $n - variate$  vector of stochastic process at time  $t$ ,  $\boldsymbol{\lambda}$  is the  $n \times 1$  vector of factor weights,  $f_t$  is the common factor and  $\boldsymbol{\Sigma}_w$  is a  $n \times n$  diagonal covariance matrix. The authors used the Kalman filter to estimate the parameters of the model.

In a following step King, Sentana and Wadhwani (1994) extended this model such as the unobserved factors to follow a GARCH process.

# Chapter 10

## Auxiliary Variable Sampler and Multivariate Time-Varying Volatility Models

### 10.1 Multivariate Stochastic Volatility Model

#### 10.1.1 Introduction

This section focuses on the Bayesian analysis of the multivariate stochastic volatility model.

More specifically, the form of the proposed multivariate stochastic volatility model is given by the following hierarchical structure of conditional densities

$$\mathbf{y}_t | \cdot \sim MN_k \left( \mathbf{0}, \mathbf{H}_t^{1/2} \mathbf{\Lambda} \mathbf{H}_t^{1/2} \right), \quad (10.1)$$

where

$$\mathbf{H}_t = \begin{bmatrix} h_{1,t} & 0 & \cdots & 0 \\ 0 & h_{2,t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_{k,t} \end{bmatrix},$$

and

$$h_{i,t}|\cdot \sim LN\left(\beta_i \log h_{i,t-1}, \sigma_i^2\right).$$

where  $\mathbf{y}_t$  is a  $k$ -variate vector of realizations of the stochastic process at time  $t$ ,  $\mathbf{H}_t$  is  $k \times k$  diagonal matrix that contains the volatilities,  $\mathbf{\Lambda}$  is a positive defined covariance matrix,  $LN(\cdot, \cdot)$  is the LogNormal density and  $\beta_1, \beta_2, \dots, \beta_k, \sigma_1^2, \sigma_2^2, \dots, \sigma_k^2$  are the hyperparameters of the model. Restrictions,  $0 < \beta_i < 1$ , for  $i = 1, \dots, k$  are imposed such as the series of volatilities to be covariance stationary. This form of the model leads to a MCMC algorithm where some of the full conditional densities are not of standard form.

In the next section the model (10.1) is rewritten by the introduction of  $kT$  latent variables such as the constructing MCMC algorithm to be entirely based on Gibbs steps.

### 10.1.2 Bayesian Approach and the Auxiliary Variable Sampler

The model (10.1) can be rewritten applying the following Theorem of the decomposition of the LogNormal density.

**Theorem 18** Suppose that  $x|u$  follows  $f(\cdot)$ , where

$$\begin{aligned} f(x|u) &= \frac{1}{2x\sqrt{u}} I_A, \\ A &= \{x : \exp(\mu - \sqrt{u}) \leq x \leq \exp(\mu + \sqrt{u})\} \end{aligned}$$

and  $u$  follows  $G\left(\frac{3}{2}, \frac{1}{2\sigma^2}\right)$ . Then the marginal density of  $x$  is the LogNormal density with parameters  $\mu$  and  $\sigma^2$ .

**Proof:**

$$\begin{aligned}
f(x) &= \int_u f(x|u) f(u) du \\
&= \frac{1}{2x\Gamma(3/2)(2\sigma^2)^{3/2}} \int_u \exp\left(-\frac{u}{2\sigma^2}\right) du \\
&= \frac{1}{x\sqrt{\pi}(2\sigma^2)^{3/2}} \int_{(\log x - \mu)^2}^{\infty} \exp\left(-\frac{u}{2\sigma^2}\right) du \\
&= \frac{1}{x\sqrt{2\pi}\sigma^2} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right) \\
&\equiv LN(\mu, \sigma^2). \blacksquare
\end{aligned}$$

By utilizing the above Theorem, the stochastic volatility model of (10.1) can be rewritten with the following hierarchical structure

$$\begin{aligned}
\mathbf{y}_t | \cdot &\sim MN_k(\mathbf{0}, \mathbf{H}_t^{1/2} \mathbf{\Lambda} \mathbf{H}_t^{1/2}), \\
f(h_{i,t} | u_{i,t}, \beta_i, h_{i,t-1}, \sigma_i^2) &= \frac{1}{2h_{i,t}\sqrt{p}} I_{A_{i,t}}, \text{ for } i = 1, \dots, k, \\
u_{i,t} | \cdot &\sim G\left(\frac{3}{2}, \frac{1}{2\sigma_i^2}\right), \text{ for } i = 1, \dots, k,
\end{aligned} \tag{10.2}$$

where

$$A_{i,t} = \left\{ \exp(\beta_i \log h_{i,t-1} - \sqrt{u_{i,t}}) \leq h_{i,t} \leq \exp(\beta_i \log h_{i,t-1} + \sqrt{u_{i,t}}) \right\}. \tag{10.3}$$

Based on (10.2), the joint posterior distribution -known up to a constant- for the parameters of the multivariate stochastic volatility model takes the following form:

Let  $\boldsymbol{\theta} = (\beta_1, \beta_2, \dots, \beta_k, \sigma_1^2, \sigma_2^2, \dots, \sigma_k^2, \boldsymbol{\Lambda}, \mathbf{H}, \mathbf{U})$  is the parameter vector, where

$$\mathbf{H} = \begin{bmatrix} h_{1,T} & h_{2,T} & \cdots & h_{k,T} \\ h_{1,T-1} & h_{2,T-1} & \cdots & h_{k,T-1} \\ \vdots & \vdots & \ddots & \vdots \\ h_{1,1} & h_{2,1} & \cdots & h_{k,1} \end{bmatrix},$$

then

$$\begin{aligned} f(\boldsymbol{\theta}|\mathbf{Y}) &\propto \left( |\boldsymbol{\Lambda}|^{T/2} \prod_{i=1}^k \left( \sigma_i^{2\frac{3T}{2}} \prod_{t=1}^T \left( h_{i,t}^{3/2} \right) \right) \right)^{-1} \\ &\exp \left\{ -\frac{1}{2} \sum_{t=1}^T \left( \mathbf{y}_t' \mathbf{H}_t^{-1/2} \boldsymbol{\Lambda}^{-1} \mathbf{H}_t^{-1/2} \mathbf{y}_t \right) \right\} \\ &\exp \left\{ -\frac{1}{2} \sum_{i=1}^k \sum_{t=1}^T \frac{u_{i,t}}{2\sigma_i^2} \right\} \prod_{i=1}^k \prod_{t=1}^T I_{A_{i,t}} \\ &\pi(\beta_1, \beta_2, \dots, \beta_k, \sigma_1^2, \sigma_2^2, \dots, \sigma_k^2, \boldsymbol{\Lambda}, \mathbf{h}_0), \end{aligned} \quad (10.4)$$

where  $I_{A_{i,t}}$  is defined in (10.3) and  $\pi(\cdot)$  is the joint a priori distribution of the parameters of interest.

The full conditional densities for the parameters of stochastic volatility model, based on the posterior (10.4) are given below

- $\sigma_i^2 | \cdot \sim IG \left( \frac{3T+1}{2}, 0.5 \sum_{t=1}^T u_{i,t} + 0.5 \right)$ , for  $i = 1, \dots, k$ .
- $u_{i,t} | \cdot \sim \exp onential \left( \frac{1}{2\sigma_i^2} \right) I_A$ , where

$$A = \{u_{i,t} : u_{i,t} \geq (\log h_{i,t} - \beta_i \log h_{i,t-1})^2\},$$

for  $i = 1, \dots, T$  and  $t = 1, \dots, T$ .

- $\boldsymbol{\Lambda}^{-1} | \cdot \sim \mathbf{W} \left( \left( \mathbf{K} + \sum_{t=1}^T \left( \mathbf{y}_t \mathbf{H}_t^{-1/2} \mathbf{H}_t^{-1/2} \mathbf{y}_t' \right) \right)^{-1}, T + 2 \right)$ .

- Marginalizing out the  $u_{i,t}$  (Chib and Carlin 1999) the full conditional for  $\beta_i$  is  $N(\mu_i, s_i^2)$ , where

$$\mu_i = \frac{\sum_{t=1}^T \log h_{i,t-1} \log h_{i,t}}{\sum_{t=1}^T \log h_{i,t-1}^2}$$

$$s_i^2 = \frac{\sigma_i^2}{\sum_{t=1}^T \log h_{i,t-1}^2},$$

- The full conditional density for the matrix of volatilities  $\mathbf{H} = (\mathbf{h}_T, \mathbf{h}_{T-1}, \dots)'$  is not of standard form

$$f(\mathbf{H}|\cdot) \propto \left( \prod_{t=1}^T \prod_{i=1}^k h_{i,t}^{3/2} \right) \exp \left\{ -0.5 \sum_{t=1}^T \mathbf{y}_t' \mathbf{H}_t^{-1/2} \mathbf{\Lambda}^{-1} \mathbf{H}_t^{-1/2} \mathbf{y}_t \right\} \quad (10.5)$$

$$\prod_{t=1}^T \prod_{i=1}^k I_{A_{i,t}}.$$

In order to handle the problem that the full conditional density of volatilities which is not of standard form, certain *non-linear* transformations of the volatilities are adopted. Firstly, note that

**Theorem 19** *Let the density given by (10.5). If we apply the following transformation*

$$w_{i,t} = h_{i,t}^{-0.5}, \text{ for } i = 1, \dots, k \text{ and } t = 1, \dots, T$$

*then the distribution of*

$$\mathbf{W} = \begin{bmatrix} w_{1,T} & w_{2,T} & \cdots & w_{k,T} \\ w_{1,T-1} & w_{2,T-1} & \cdots & w_{k,T-1} \\ \vdots & \vdots & \ddots & \vdots \\ w_{1,1} & w_{2,1} & \cdots & w_{k,1} \end{bmatrix},$$

is given by the following formula

$$f(\mathbf{W}|\cdot) \propto \exp \left\{ -0.5 \sum_{t=1}^T \mathbf{y}_t' \mathbf{G}_t^{1/2} \mathbf{\Lambda}_t^{-1} \mathbf{G}_t^{1/2} \mathbf{y}_t \right\} \prod_{t=1}^T \prod_{i=1}^k I_{A_{i,t}^*}, \quad (10.6)$$

where

$$\mathbf{G}_t^{1/2} = \text{diag}(w_{1,t}, w_{2,t}, \dots, w_{k,t}),$$

$$\begin{aligned} A_{i,t}^* &= \left\{ \exp(0.5(\beta_i \log w_{i,t-1}^2 - \sqrt{u_{i,t}})) \leq w_{i,t} \leq \exp(0.5(\beta_i \log w_{i,t-1}^2 + \sqrt{u_{i,t}})) \right\} \\ \text{for } t &= 2, \dots, T, \text{ and} \\ A_{i,1}^* &= \left\{ \exp(-0.5(\beta_i \log h_{i,0} + \sqrt{u_{i,1}})) \leq w_{i,1} \leq \exp(-0.5(\beta_i \log h_{i,0} - \sqrt{u_{i,1}})) \right\}, \\ \text{for } t &= 1. \end{aligned} \quad (10.7)$$

**Proof.** Note that the Jacobian of the above transformations is  $|J| \propto \prod_{t=1}^T \prod_{i=1}^p w_t^{-3}$ . The remaining calculations are straightforward. ■

The following remark turns up straightforwardly from the above theorem:

**Remark 1** Following (10.6).

1. The marginal density of  $\mathbf{w}_t = (w_{1,t}, w_{2,t}, \dots, w_{k,t})$  is a truncated multivariate Normal density with mean vector equals to  $\mathbf{0}$  and variance covariance matrix  $\mathbf{S}_t$  equals to

$$\mathbf{S}_t = (\mathbf{y}_t^* \mathbf{\Lambda}^{-1} \mathbf{y}_t^*)^{-1}, \quad (10.8)$$

where

$$\mathbf{y}_t^* = \text{diag}(y_{1,t}, y_{2,t}, \dots, y_{k,t}).$$

The truncation is given by  $\prod_{i=1}^k I_{A_{i,t}^*}$ , where  $A_{i,t}^*$  is given by (10.7).

2. The conditional density of  $w_{i,t}$  given  $\mathbf{w}_{t/-i} = (w_{1,t}, \dots, w_{i-1,t}, w_{i+1,t}, \dots, w_{k,t})$  is Normal with mean and variance given by the properties of the Multivariate distribution (Johnson and Wichern, 1999), truncated on  $\{A_{i,t}^* \cap A_{i,t+1}^*\}$ .

Due to the difficulty of sampling from the truncated multivariate Normal distribution of the Remark (1), sampling procedure is taking place from the univariate density of each  $w_{i,t}$ , (see 10.7) for  $t = 1, \dots, T$  and  $i = 1, \dots, k$ .

In detail

- $w_{i,1}|\cdot \sim N(\cdot, \cdot) I_{(L,U)}$ , where

$$L = \begin{cases} \max \left\{ \begin{array}{l} 0, \exp(-0.5(\beta_i \log h_{i,0} + \sqrt{u_1})) \\ \exp\left(\frac{1}{2\beta_i}(\log w_{i,2}^2 - \sqrt{u_{i,2}})\right) \end{array} \right\}, & \text{if } \beta > 0 \\ \max \left\{ \begin{array}{l} 0, \exp(-0.5(\beta_i \log h_{i,0} + \sqrt{u_1})) \\ \exp\left(\frac{1}{2\beta_i}(\log w_{i,2}^2 + \sqrt{u_{i,2}})\right) \end{array} \right\}, & \text{if } \beta < 0 \end{cases},$$

$$U = \begin{cases} \min \left\{ \begin{array}{l} \exp(-0.5(\beta_i \log h_{i,0} - \sqrt{u_1})) \\ \exp\left(\frac{1}{2\beta_i}(\log w_{i,2}^2 + \sqrt{u_{i,2}})\right) \end{array} \right\}, & \text{if } \beta > 0 \\ \min \left\{ \begin{array}{l} \exp(-0.5(\beta_i \log h_{i,0} - \sqrt{u_1})) \\ \exp\left(\frac{1}{2\beta_i}(\log w_{i,2}^2 - \sqrt{u_{i,2}})\right) \end{array} \right\}, & \text{if } \beta < 0 \end{cases}.$$

- $w_{i,t}|\cdot \sim N(\cdot, \cdot) I_{(L,U)}$ , for  $t = 2, \dots, T-1$ , where

$$L = \begin{cases} \max \left\{ \begin{array}{l} 0, \exp(0.5(\beta_i \log w_{i,t-1}^2 - \sqrt{u_{i,t}})) \\ \exp\left(\frac{1}{2\beta_i}(\log w_{i,t+1}^2 - \sqrt{u_{i,t+1}})\right) \end{array} \right\}, & \text{if } \beta_i > 0 \\ \max \left\{ \begin{array}{l} 0, \exp(0.5(\beta_i \log w_{i,t-1}^2 - \sqrt{u_{i,t}})) \\ \exp\left(\frac{1}{2\beta_i}(\log w_{i,t+1}^2 + \sqrt{u_{i,t+1}})\right) \end{array} \right\}, & \text{if } \beta_i < 0 \end{cases},$$

$$U = \begin{cases} \min \left\{ \begin{array}{l} 0, \exp \left( 0.5 \left( \beta_i \log w_{i,t-1}^2 + \sqrt{u_{i,t}} \right) \right), \\ \exp \left( \frac{1}{2\beta_i} \left( \log w_{i,t+1}^2 + \sqrt{u_{i,t+1}} \right) \right) \end{array} \right\}, & \text{if } \beta_i > 0 \\ \min \left\{ \begin{array}{l} 0, \exp \left( 0.5 \left( \beta_i \log w_{i,t-1}^2 + \sqrt{u_{i,t}} \right) \right), \\ \exp \left( \frac{1}{2\beta_i} \left( \log w_{i,t+1}^2 - \sqrt{u_{i,t+1}} \right) \right) \end{array} \right\}, & \text{if } \beta_i < 0 \end{cases}.$$

- $w_{i,T}|\cdot \sim N(\cdot, \cdot) I_{(L,U)}$ , for  $t = 2, \dots, T-1$ , where

$$L = \max \{0, \exp (0.5 (\beta_i \log w_{i,T-1}^2 - \sqrt{u_{i,T}}))\}$$

$$U = \min \{0, \exp (-0.5 (\beta \log w_{i,T-1}^2 - \sqrt{u_{i,T}}))\}.$$

In the end of each sweep of the algorithm we re-transform  $w_{i,t}$  to  $h_{i,t}$  by using the reverse transformation  $h_{i,t} = 1/w_{i,t}^2$ ; for  $t = 0, \dots, T$ , and  $i = 1, \dots, k$

### 10.1.3 Application

In this section an application is presented using real data. In detail, the daily exchange rate of the US dollar (USD) and the Japanese Yen (JPY) with respect to the Greek Drachma (GRD) are used. To elaborate, let  $c_t$  be the exchange rate of a currency with respect to the drachma on day  $t$ ; then (one component of the multivariate) data series is given by  $y_t = \log \left( \frac{c_t}{c_{t-1}} \right) \cdot 1000$ , that represents the daily relative (percentage) change of the exchange rate since

$$\log \left( \frac{c_t}{c_{t-1}} \right) \simeq \frac{c_t}{c_{t-1}} - 1 = \frac{c_t - c_{t-1}}{c_{t-1}}, \text{ for } \frac{c_t}{c_{t-1}} \simeq 1.$$

The data set is consisted of 845 multivariate observations concerning the period (16/12/93 – 2/5/97) and illustrated in Figure 10-1.

Applying the algorithm that has been described in the previous section, a sample from the posterior density of the model is taken and used to make inference about the parameters. In detail, the first 50000 sample points of the MCMC algorithm are discarded

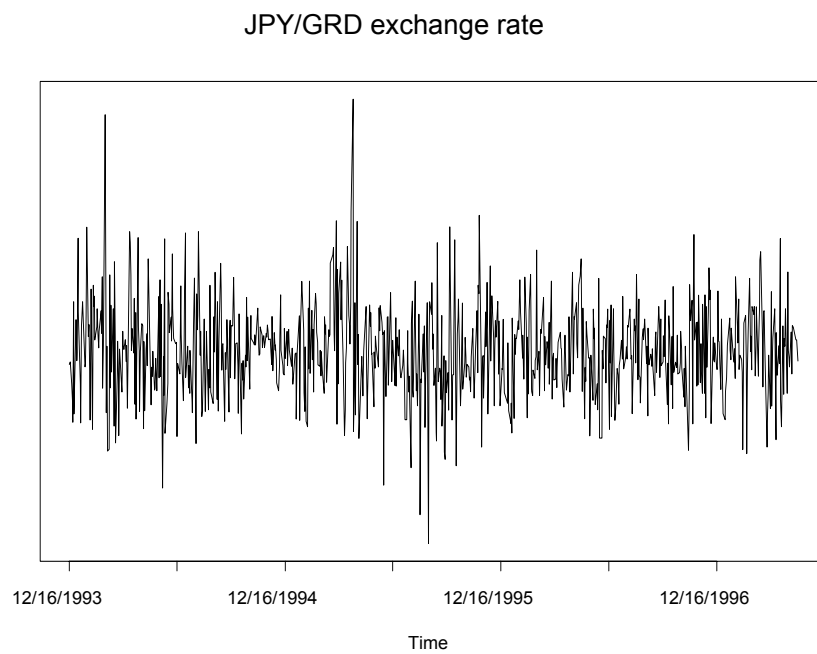
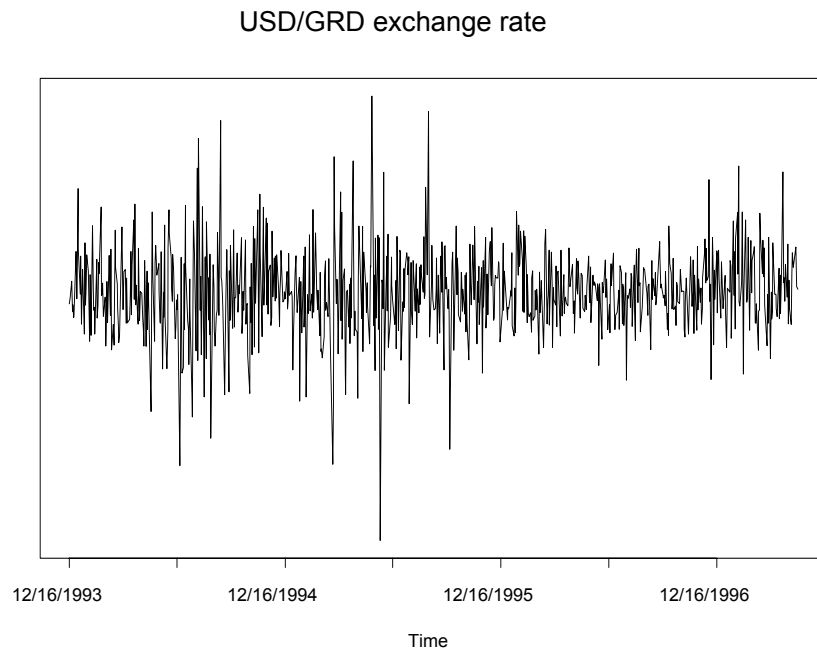


Figure 10-1: Exchange Rates Series

Parameter	Mean	Variance
USD $b$	0.9536141	0.001116
USD $\sigma^2$	0.0454296	0.001083
JPY $b$	0.9246696	0.001479
JPY $\sigma^2$	0.0763938	0.001639
$\Lambda(1, 1)$	21.54266	27.94165
$\Lambda(1, 2)$	4.757719	1.129672
$\Lambda(2, 2)$	21.5663	12.81458
$Corr(USD, JPY)$	0.2228579	0.001273

Table 10.1: Summary statistics for the hyperparameters of the multivariate SV model

as burn-in. After the drop of the burn-in, one sample point is kept out of 2000 iterations in order the final sample of 5000 iterations to be an approximately independent sample from the posterior density of the parameters of the multivariate stochastic volatility model.

The final posterior sample was checked for convergence to the limiting distribution. All the chains have been converged based on the criteria of Geweke (1992), Raftery and Lewis (1992) and Heidelberger and Welch (1983). Apart from this result the subsampling diagnostic (see section 4.3 and Giakoumatos *et al.*, 1999) is applied in order to check the convergence. This criterion is applied to the initial 50000 iterations of the Markov chain by setting:  $a = 0.05$ ,  $t = 0.99$ ,  $d = 0.999$ . The subsampling diagnostic points out that the proposed algorithm needs approximately 33000 iterations to get in the target distribution. The results of this diagnostic are presented in Figure 10-2. Moreover, Figure 10-2 presents the autocorrelation plots for the hyperparameters from the MCMC series

Figure 10-4 presents the 5000 iterations from the MCMC for each parameter of interest for the stochastic volatility model. Based on this i.i.d. sample the posterior mean and variance of the parameters of the model are estimated.

The summary statistics of these marginal densities are presented in the Table 10.1.

From the above, it becomes obvious that both exchange rates (USD and JPY) present high volatility persistence ( $E(b_{USD})=0.95$  and  $E(b_{JPY})=0.92$ ) and the correlation between these two exchange rates is significant ( $Corr(USA, JPY)=0.2228579$ ). The pos-

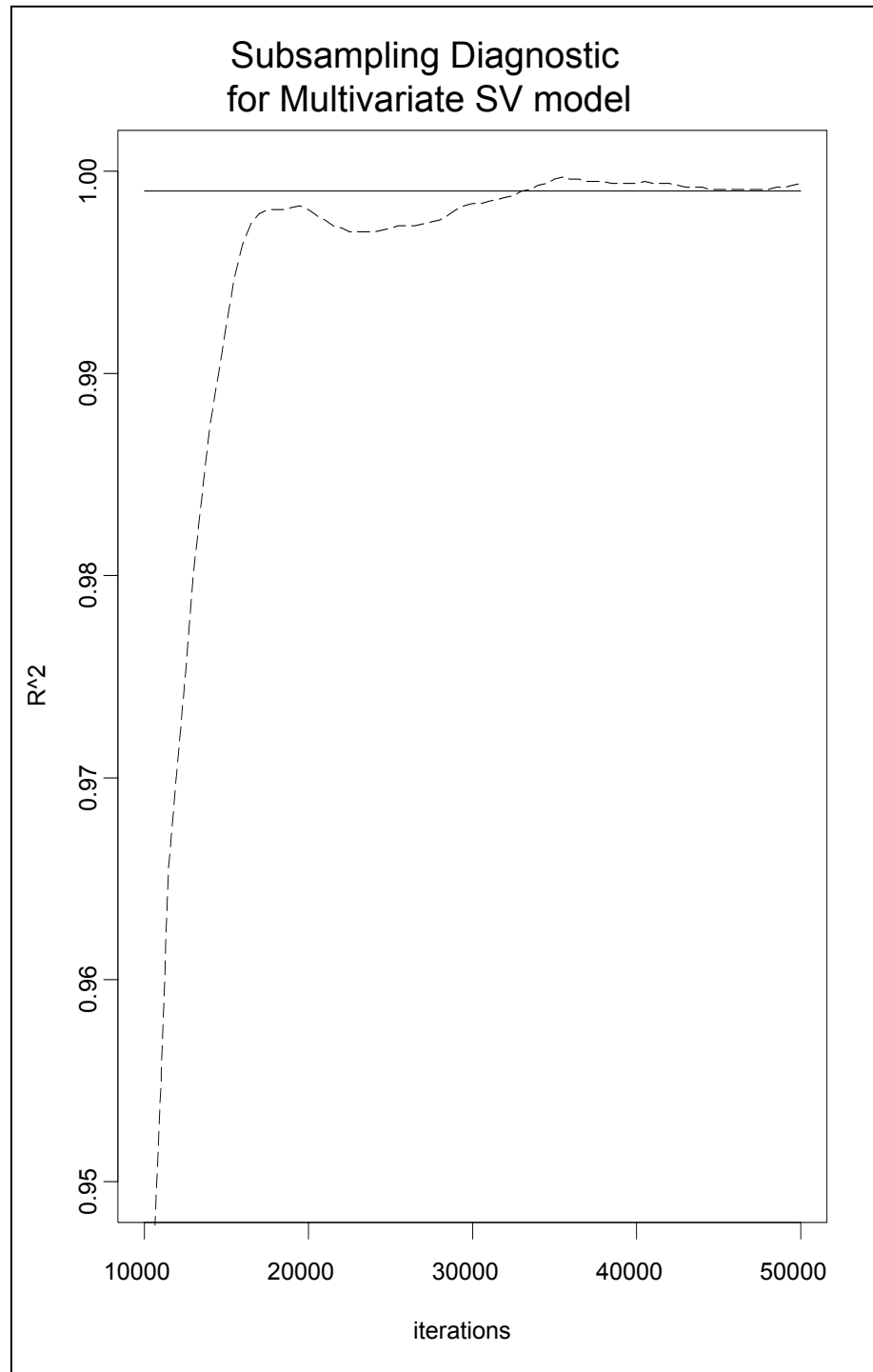


Figure 10-2: Subsampling diagnostic plot for Multivariate SV model

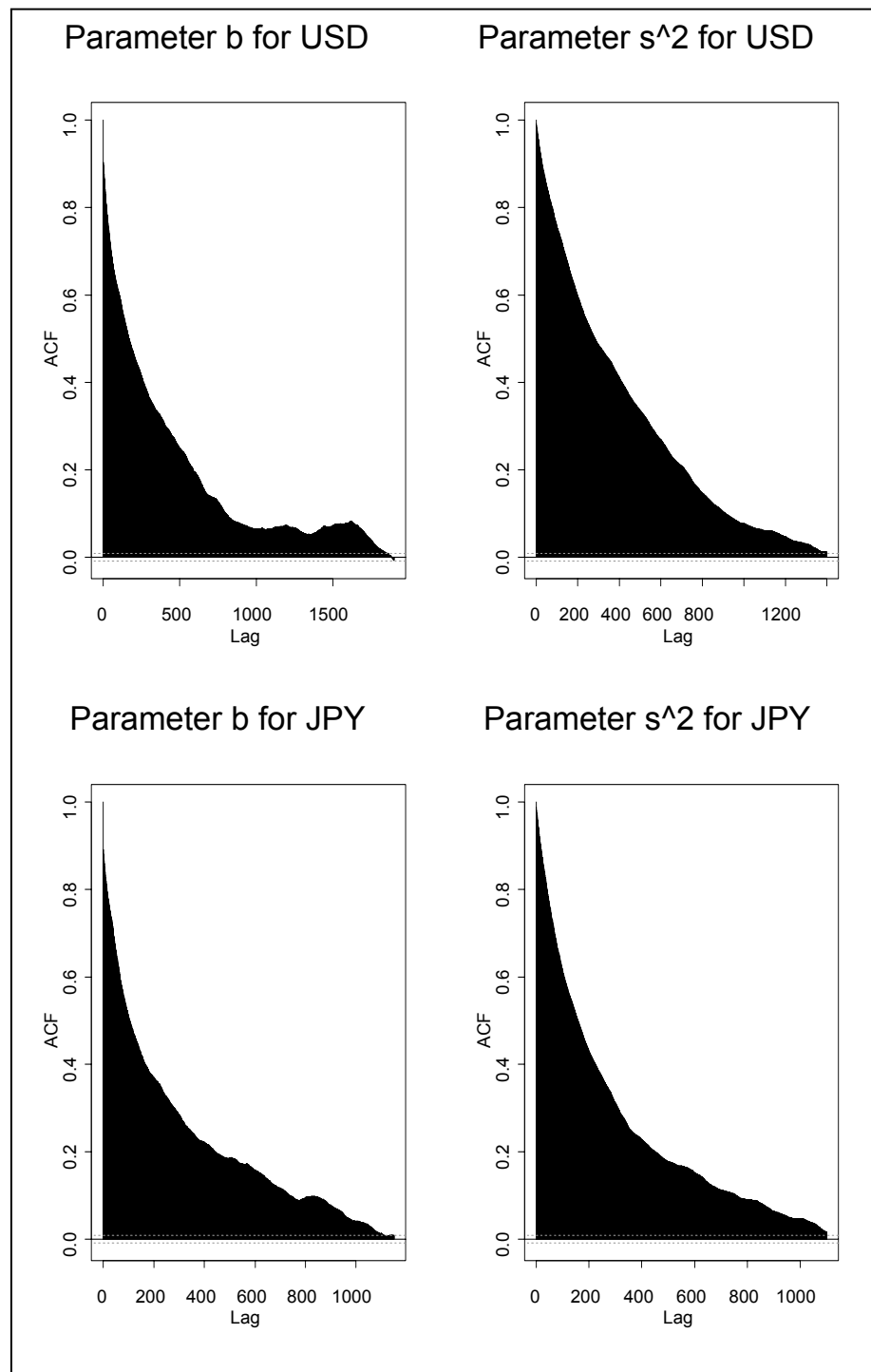


Figure 10-3: Autocorrelation function plots for the hyperparameters of the multivariate SV model

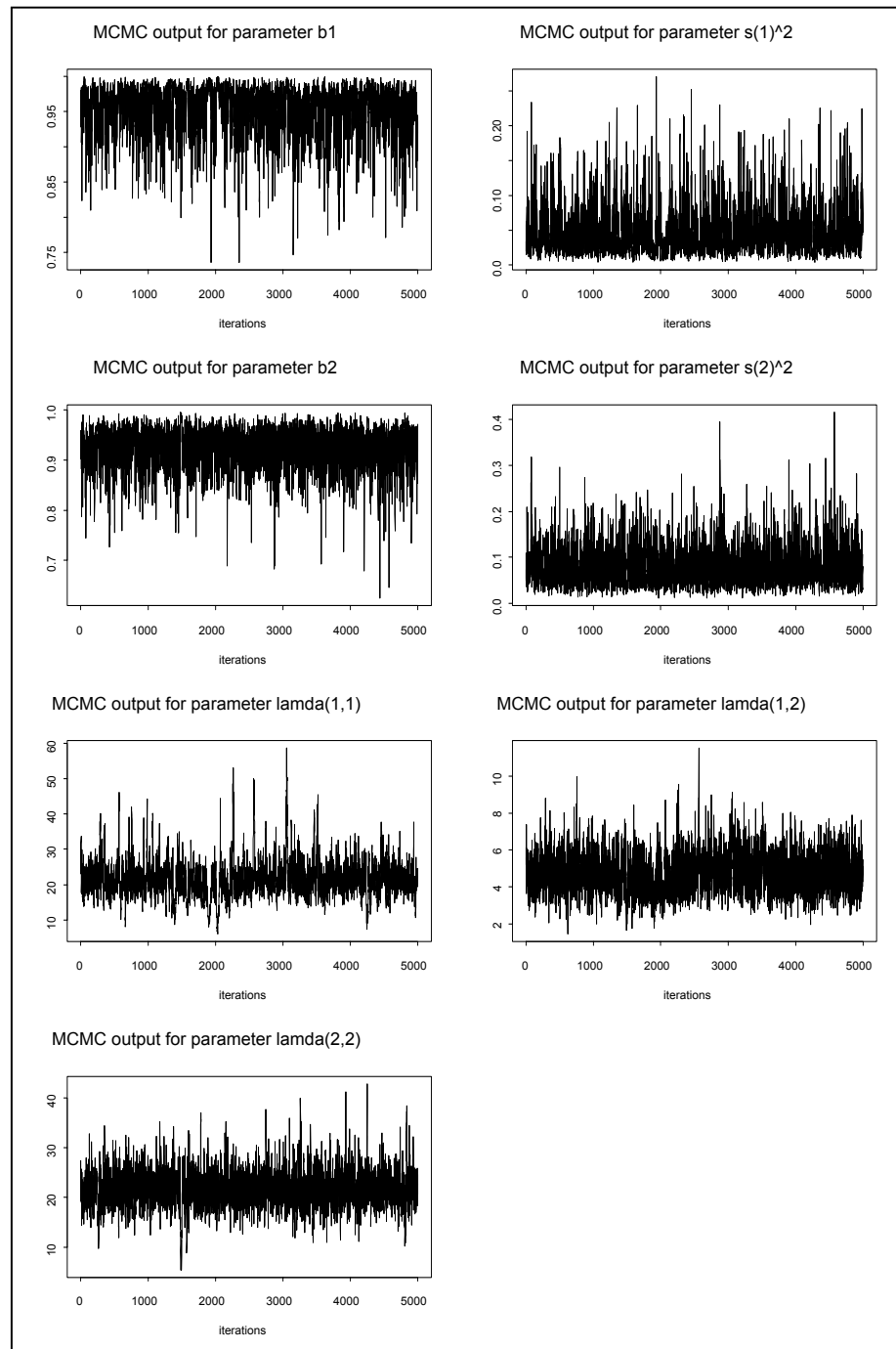


Figure 10-4: MCMC output for parameters of multivariate SV model

terior histograms of the hyperparameters of the multivariate stochastic volatility model are presented in Figure 10-5 and for the parameters of  $\Lambda$  in Figure 10-6.

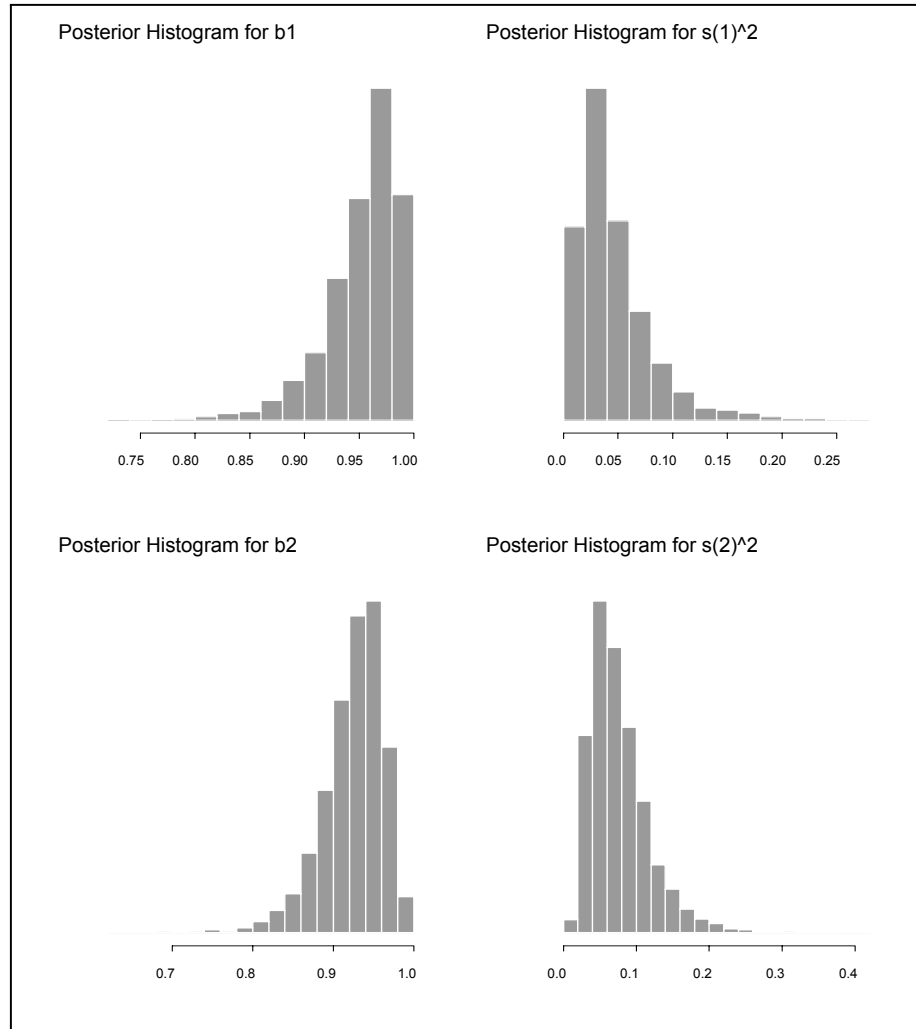


Figure 10-5: Posterior Histograms for the hyperparameters of the multivariate SV model

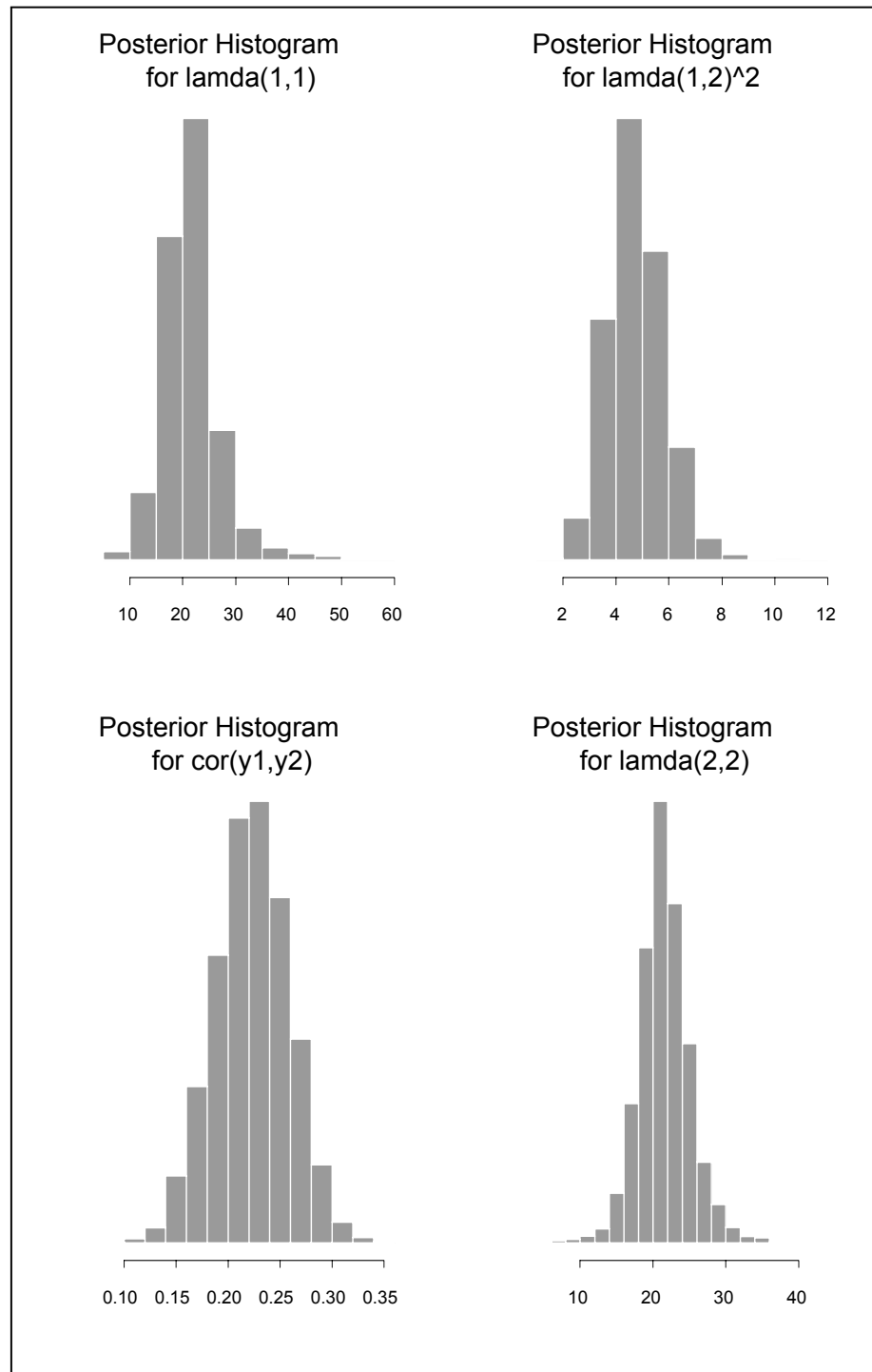


Figure 10-6: Posterior Histograms for the variance covariance matrix  $\Lambda$  of the multivariate SV model

## 10.2 Multivariate Unobserved ARCH Model

### 10.2.1 Introduction

A straightforward extension of the univariate unobserved ARCH model to multivariate case, is the model that presented bellow

$$\begin{aligned}
 \mathbf{y}_t | \mathbf{f}_t, \boldsymbol{\Sigma} &\sim \mathbf{N}_p(\mathbf{f}_t, \boldsymbol{\Sigma}), \\
 \mathbf{f}_t | \mathbf{f}_{t-1}, \boldsymbol{\theta}, \mathbf{f}_0 &\sim \mathbf{N}_p(\mathbf{0}, \boldsymbol{\Delta}_t) \\
 \boldsymbol{\Delta}_t &= \text{diag}(\sigma_{1t}^2, \dots, \sigma_{pt}^2), \sigma_{it}^2 = a_i + b_i \cdot f_{i,t-1}^2, \\
 \text{for } t &= 1, \dots, T;
 \end{aligned} \tag{10.9}$$

where  $\mathbf{y}_t$  is the  $p$ -variate realization of the stochastic process,  $\boldsymbol{\theta} = (a_1, \dots, a_p, b_1, \dots, b_p)$  is the vector of hyperparameters,  $\mathbf{f}_0$  is a  $p$ -variate vector which denotes the "history" of the unobserved ARCH components,  $\mathbf{f}_t$  is the time varying  $p$ -dimension unobserved component,  $\boldsymbol{\Sigma}$  and  $\boldsymbol{\Delta}_t$  are  $(p \times p)$  covariance matrices of the stochastic process and the unobserved component at time  $t$  respectively, and  $\mathbf{N}_p(\cdot, \cdot)$  denotes the  $p$ -variate Normal distribution. Restrictions  $a_i > 0$  and  $b_i > 0$ , for  $i = 1, \dots, p$ , are imposed to (10.9) so that the elements of  $\boldsymbol{\Delta}_t$  are always positive. The additional restrictions  $b_i \leq 1$  for  $i = 1, \dots, p$ , are placed so that each of the  $p$  ARCH components of the model to be covariance stationary (Engle, 1982). The unconditional and conditional variance of the stochastic process  $\mathbf{y}_t$  are given by

$$\begin{aligned}
 \text{Var}(\mathbf{y}_t) &= \boldsymbol{\Sigma} + \mathbf{H}, \\
 \text{Var}(\mathbf{y}_t | \mathbf{y}_{t-1}, \boldsymbol{\theta}) &= \boldsymbol{\Sigma} + \boldsymbol{\Delta}_t,
 \end{aligned}$$

where  $\mathbf{H} = \text{diag} \left( \frac{a_1}{1-b_1}, \dots, \frac{a_p}{1-b_p} \right)$ ; thus, the unconditional and conditional variance can be decomposed into two parts respectively. The first part is the variance-covariance matrix  $\Sigma$ , which is independent of time, and can be considered as the underlying variance structure of the stochastic process  $\mathbf{y}_t$ . The second part is the matrix  $\mathbf{H}$  for the unconditional variance and the matrix  $\Delta_t$  for the conditional variance respectively, and expresses the variability which is caused by the phenomenon of volatility clustering.

It is readily shown that each element of  $\mathbf{y}_t$  follows a univariate unobserved ARCH process defined in (7.8), but (10.9) achieves the modelling of the covariance structure of  $\mathbf{y}_t$  through  $\Sigma$ . The parameter vector of (10.9) contains the hyperparameters  $\theta$ , the unobserved components  $\mathbf{F} = (\mathbf{f}_1, \dots, \mathbf{f}_T)$ , the ‘history’  $\mathbf{f}_0$  of the unobserved components and the covariance matrix  $\Sigma$ .

## 10.2.2 Bayesian Approach and the Auxiliary Variable Sampler

In this section, the auxiliary variable algorithm for sampling from the posterior density of the multivariate unobserved ARCH model is analyzed.

Denoting all data by the  $T \times p$  matrix  $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_T)$ , the joint posterior density of model (7.8) can be calculated, via Bayes theorem, by

$$f(\mathbf{F}, \theta, \Sigma, \mathbf{f}_0 | \mathbf{Y}) \propto \prod_{t=1}^T (f(\mathbf{y}_t | \mathbf{f}_t, \Sigma) f(\mathbf{f}_t | \mathbf{f}_{t-1}, \theta, \mathbf{f}_0)) \pi(\theta, \Sigma, \mathbf{f}_0). \quad (10.10)$$

The first two terms in the above product are derived from the hierarchical structure in (10.9). The last term,  $\pi(\theta, \Sigma, \mathbf{f}_0)$ , is the joint prior density of  $\theta$ ,  $\Sigma$  and  $\mathbf{f}_0$ . These vectors are assumed a priori independent with prior densities given by  $\pi(\theta) \propto \left( \prod_{i=1}^P a_i \right)^{-1}$ ,  $\mathbf{f}_0 \sim \mathbf{N}_p(\mathbf{0}, \mathbf{V})$  for a diagonal covariance matrix  $\mathbf{V}$ , and  $\Sigma \sim \mathbf{IW}(k, \mathbf{A})$ , where  $\mathbf{IW}(\cdot, \cdot)$  denotes the inverse Wishart distribution with mean  $E(\Sigma) = \frac{1}{k-2p-2} \mathbf{A}$ , where  $\mathbf{A}$  is a  $p \times p$

matrix. Using these prior densities, the posterior (10.10) takes the form

$$\begin{aligned}
f(\mathbf{F}, \boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{f}_0 | \mathbf{Y}) \propto & \frac{1}{\prod_{i=1}^p \left( a_i \prod_{t=1}^T \left( \sqrt{a_i + b_i f_{i,t-1}^2} \right) \right)} \cdot \frac{1}{|\boldsymbol{\Sigma}|^{\frac{T+k-p-1}{2}}} \\
& \cdot \exp \left\{ -\frac{1}{2} \left( \sum_{t=1}^T \left( \sum_{i=1}^p \left( \frac{f_{i,t}^2}{a_i + b_i f_{i,t-1}^2} \right) \right) + \text{tr}(\mathbf{A} \boldsymbol{\Sigma}^{-1}) \right) \right\} \\
& \cdot \exp \left\{ -\frac{1}{2} \sum_{t=1}^T \left( (\mathbf{y}_t - \mathbf{f}_t)' \boldsymbol{\Sigma}^{-1} (\mathbf{y}_t - \mathbf{f}_t) \right) \right\} \\
& \cdot \exp \left\{ -\frac{1}{2} \mathbf{f}_0' \mathbf{V}^{-1} \mathbf{f}_0 \right\}.
\end{aligned} \tag{10.11}$$

From the above, it is clear that (10.11) is heavily parameterized and that construction of a MCMC sampling strategy is not at all simple. In detail, all full posterior conditional densities are not of known form, so Gibbs sampler steps are problematic. Furthermore, it is known that Metropolis-Hastings requires a high degree of sophistication (Shephard and Pitt, 1997; Kim *et al*, 1998).

The following theorem resolves this problem.

**Theorem 20** *If in the posterior density  $f(\mathbf{F}, \boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{f}_0 | \mathbf{Y})$  defined in (10.11) we perform*

*(a) the transformations*

$$w_{i,t} = \sqrt{\frac{b_i}{a_i}} \cdot f_{i,t}, \quad g_i = \sqrt{\frac{a_i}{b_i}}, \quad \gamma_i = b_i,$$

*for  $t = 0, \dots, T$  and  $i = 1, \dots, p$ ,*

*(b) the inclusions of  $2T$  positive latent variables  $\mathbf{u} = (u_1, \dots, u_T)$  and  $\mathbf{k} = (k_1, \dots, k_T)$*

such that the resulting joint density is given by

$$\begin{aligned}
f(\mathbf{u}, \mathbf{k}, \mathbf{g}, \boldsymbol{\gamma}, \mathbf{W}, \boldsymbol{\Sigma} | \mathbf{Y}) &\propto \left( \prod_{t=1}^T \mathbf{I} \left( u_t \leq \frac{1}{\prod_{i=1}^p \sqrt{1 + w_{i,t-1}^2}} \right) \right) \\
&\cdot \left( \prod_{t=1}^T \mathbf{I} \left( k_t \leq \exp \left\{ -\frac{1}{2} \sum_{i=1}^p \frac{w_{i,t}^2}{\gamma_i [1 + w_{i,t-1}^2]} \right\} \right) \right) \\
&\cdot \exp \left\{ -\frac{1}{2} \mathbf{z}_0' \mathbf{V}^{-1} \mathbf{z}_0 \right\} \frac{1}{|\boldsymbol{\Sigma}|^{\frac{T+k-p-1}{2}} \prod_{i=1}^p \gamma_i^{\frac{T}{2}}} \\
&\cdot \exp \left\{ -\frac{1}{2} \left( \text{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \sum_{t=1}^T (\mathbf{y}_t - \mathbf{z}_t)' \boldsymbol{\Sigma}^{-1} (\mathbf{y}_t - \mathbf{z}_t) \right) \right\},
\end{aligned}$$

where  $\mathbf{I}(\cdot)$  is the indicator function,  $\mathbf{g} = (g_1, \dots, g_p)'$ ,  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_p)'$ ,  $\mathbf{z}_t = (g_1 \cdot w_{1,t}, \dots, g_p \cdot w_{p,t})'$  for  $t = 0, \dots, T$  and  $\mathbf{W} = (\mathbf{w}_0, \dots, \mathbf{w}_T)'$ , then the marginal density  $f(\mathbf{F}, \boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{f}_0 | \mathbf{Y})$  is given by (10.11).

*Proof:* For the first step of the theorem note that the Jacobian takes the form

$$|\mathbf{J}| = \prod_{i=1}^P [g_i^{T+2} \cdot \gamma_i]$$

and

$$\begin{aligned}
f_{\mathbf{W}, \mathbf{g}, \boldsymbol{\gamma}}(\mathbf{W}, \boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{w}_0 | \mathbf{Y}) &= f_{\mathbf{a}, \mathbf{b}, \mathbf{F}}(\mathbf{W}, \boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{w}_0 | \mathbf{Y}) \cdot |\mathbf{J}| \\
&= \frac{1}{\prod_{i=1}^p \sqrt{1 + w_{i,t-1}^2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^p \frac{w_{i,t}^2}{\gamma_i [1 + w_{i,t-1}^2]} \right\} \\
&\cdot \exp \left\{ -\frac{1}{2} \mathbf{z}_0' \mathbf{V}^{-1} \mathbf{z}_0 \right\} \frac{1}{|\boldsymbol{\Sigma}|^{\frac{T+k-p-1}{2}} \prod_{i=1}^p \gamma_i^{\frac{T}{2}}} \\
&\cdot \exp \left\{ -\frac{1}{2} \left( \text{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \sum_{t=1}^T (\mathbf{y}_t - \mathbf{z}_t)' \boldsymbol{\Sigma}^{-1} (\mathbf{y}_t - \mathbf{z}_t) \right) \right\}.
\end{aligned}$$

For the second step, it is enough to integrate all the latent parameters of the resulting

posterior density

$$\begin{aligned}
f(\mathbf{W}, \boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{w}_0 | \mathbf{Y}) &= \int_0^{+\infty} \cdots \int_0^{+\infty} f(\mathbf{W}, \boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{w}_0, \mathbf{u}, \mathbf{k} | \mathbf{Y}) du_1 \cdots du_T dk_1 \cdots dk_T \\
&= \exp \left\{ -\frac{1}{2} \mathbf{z}_0' \mathbf{V}^{-1} \mathbf{z}_0 \right\} \cdot \frac{1}{|\boldsymbol{\Sigma}|^{\frac{T+k-p-1}{2}} \cdot \prod_{i=1}^p \gamma_i^{\frac{T}{2}}} \\
&\quad \cdot \exp \left\{ -\frac{1}{2} \left( \text{tr}(\mathbf{A} \cdot \boldsymbol{\Sigma}) + \sum_{t=1}^T (\mathbf{y}_t - \mathbf{z}_t)' \cdot \boldsymbol{\Sigma}^{-1} \cdot (\mathbf{y}_t - \mathbf{z}_t) \right) \right\} \\
&\quad \cdot \int_0^{\frac{1}{\prod_{i=1}^p \sqrt{1+w_{1,t}^2}}} du_1 \cdots \int_0^{\frac{1}{\prod_{i=1}^p \sqrt{1+w_{p,t}^2}}} du_T \\
&\quad \cdot \int_0^{\exp \left\{ \sum_{i=1}^p \frac{w_{1,t+1}^2}{\gamma_1 \cdot [1+w_{1,t}^2]} \right\}} dk_1 \cdots \int_0^{\exp \left\{ \sum_{i=1}^p \frac{w_{p,t+1}^2}{\gamma_p \cdot [1+w_{p,t}^2]} \right\}} dk_T \\
&= \exp \left\{ -\frac{1}{2} \left( \text{tr}(\mathbf{A} \cdot \boldsymbol{\Sigma}) + \sum_{t=1}^T (\mathbf{y}_t - \mathbf{z}_t)' \cdot \boldsymbol{\Sigma}^{-1} \cdot (\mathbf{y}_t - \mathbf{z}_t) \right) \right\} \\
&\quad \cdot \exp \left\{ -\frac{1}{2} \mathbf{z}_0' \mathbf{V}^{-1} \mathbf{z}_0 \right\} \cdot \frac{1}{|\boldsymbol{\Sigma}|^{\frac{T+k-p-1}{2}} \cdot \prod_{i=1}^p \gamma_i^{\frac{T}{2}}} \\
&\quad \cdot \left[ \frac{1}{\prod_{i=1}^p \sqrt{1+w_{i,t}^2}} \right] \cdot \exp \left\{ \sum_{i=1}^p \frac{w_{i,t+1}^2}{\gamma_i \cdot [1+w_{i,t}^2]} \right\} \blacksquare
\end{aligned}$$

The above Theorem guarantees that a MCMC algorithm which obtains samples from  $f(\mathbf{u}, \mathbf{k}, \mathbf{g}, \boldsymbol{\gamma}, \mathbf{w}_t, \boldsymbol{\Sigma} | \mathbf{Y})$  also obtains samples from the marginal density (with respect to  $\mathbf{u}, \mathbf{k}$ )  $f(\mathbf{g}, \boldsymbol{\gamma}, \mathbf{w}_t, \boldsymbol{\Sigma} | \mathbf{Y})$  and using the reverse transformations of (a) samples from  $f(\mathbf{F}, \boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{f}_0 | \mathbf{Y})$  are obtained. In addition to that, using the strategy suggested by Chib and Carlin (1999) some of the full conditional densities are marginalized over some parameters so that the final full conditional distributions to have elegant forms. The following remarks explain in detail these marginalizations.

**Remark 2** If the full conditional density  $f(\gamma_i | \mathbf{u}, \mathbf{k}, \mathbf{g}, \boldsymbol{\gamma}_{-i}, \mathbf{W}, \boldsymbol{\Sigma}, \mathbf{Y}) \equiv$

$f(\gamma_i | \mathbf{k}, \boldsymbol{\gamma}_{-i}, \mathbf{W})$ , for  $i = 1, \dots, p$ , where  $\boldsymbol{\gamma}_{-i}$  denotes the vector that contains all the

elements of  $\gamma$  apart from the element  $\gamma_i$ , is marginalized over the  $T$  latent parameters  $\mathbf{k} = (k_1, \dots, k_T)$ , then

$$f(\gamma_i | \gamma_{-i}, \mathbf{W}) \equiv f(\gamma_i | \mathbf{W}) \equiv \mathbf{IG} \left( \frac{T-2}{2}, \frac{1}{2} \sum_{t=1}^T \frac{w_{i,t}^2}{1+w_{i,t-1}^2} \right) \mathbf{I}(\gamma_i \leq 1),$$

where  $\mathbf{IG}(a, b)$  denotes the Inverse Gamma density with mean  $\frac{b^2}{(a-1)^2(a-2)}$ .

**Remark 3** If the full conditional density is given by  $f(\mathbf{w}_t | \mathbf{u}, \mathbf{k}, \mathbf{g}, \gamma, \mathbf{W}_{-i}, \Sigma, \mathbf{Y}) \equiv f(\mathbf{w}_t | u_{t+1}, k_t, k_{t+1}, \mathbf{g}, \gamma, \mathbf{w}_{t-1}, \mathbf{w}_{t+1}, \Sigma, \mathbf{Y})$  for  $t = 1, \dots, T-1$ , is marginalized over the parameter  $k_t$  then,

$$\begin{aligned} & f(\mathbf{w}_t | u_{t+1}, k_{t+1}, \mathbf{g}, \gamma, \mathbf{w}_{t-1}, \mathbf{w}_{t+1}, \Sigma, \mathbf{Y}) \\ \equiv & \mathbf{N}_p(\mathbf{m}_t, \mathbf{S}_t) \mathbf{I} \left( u_{t+1} \leq \frac{1}{\prod_{i=1}^p \sqrt{1+w_{i,t}^2}} \right) \mathbf{I} \left( k_{t+1} \leq \exp \left\{ -\frac{1}{2} \sum_{i=1}^p \frac{w_{i,t+1}^2}{\gamma_i (1+w_{i,t}^2)} \right\} \right), \end{aligned}$$

where

$$\begin{aligned} \mathbf{S}_t^{-1} &= \mathbf{G}' \cdot \Sigma^{-1} \cdot \mathbf{G} + \mathbf{\Gamma}_t, \\ \mathbf{m}_t &= \mathbf{S}_t \cdot \mathbf{G}' \cdot \Sigma^{-1} \cdot \mathbf{y}_t, \end{aligned}$$

where  $\mathbf{G} = \text{diag}(g_1, \dots, g_p)$  and  $\mathbf{\Gamma}_t = \text{diag}((\gamma_1 (1+w_{1,t-1}^2)), \dots, (\gamma_p (1+w_{p,t-1}^2)))$  are diagonal  $p \times p$  matrices.

**Remark 4** If the full conditional density is given by  $f(\mathbf{w}_T | \mathbf{u}, \mathbf{k}, \mathbf{g}, \gamma, \mathbf{W}_{-i}, \Sigma, \mathbf{Y}) \equiv f(\mathbf{w}_T | k_T, \mathbf{g}, \gamma, \mathbf{w}_{T-1}, \mathbf{w}_T, \Sigma, \mathbf{Y})$  is marginalized over the parameter  $k_T$  then,

$$f(\mathbf{w}_T | \mathbf{g}, \gamma, \mathbf{w}_{t-1}, \mathbf{w}_t, \Sigma, \mathbf{Y}) \equiv \mathbf{N}_p(\mathbf{m}_T, \mathbf{S}_T),$$

where  $\mathbf{S}_T^{-1} = (\mathbf{G}' \Sigma^{-1} \mathbf{G}) + \mathbf{\Gamma}_T$ ,  $\mathbf{m}_T = \mathbf{S}_T \mathbf{G}' \Sigma^{-1} \mathbf{y}_T$ ,  $\mathbf{G} = \text{diag}(g_1, \dots, g_p)$ ,

$$\mathbf{\Gamma}_T = \text{diag}((\gamma_1 (1+w_{1,T-1}^2)), \dots, (\gamma_p (1+w_{p,T-1}^2))).$$

Thus, by using the previous Remarks, the MCMC algorithm is readily constructed by using the following Gibbs steps; the notation  $|\cdot$  implies conditioning on all remaining parameters.

- $\Sigma|\cdot \sim \mathbf{IW}_p \left( T + k, \left( \mathbf{A} + \sum_{t=1}^T \left( (\mathbf{y}_t - \mathbf{z}_t) (\mathbf{y}_t - \mathbf{z}_t)' \right) \right) \right),$   
where  $\mathbf{z}_t = (g_1 \cdot w_{1,t}, \dots, g_{p,t} \cdot w_{p,t})'$ .
- $\gamma_i|\cdot \sim \mathbf{IG} \left( \frac{T-2}{2}, \frac{1}{2} \sum_{t=1}^T \frac{w_{i,t}^2}{1+w_{i,t-1}^2} \right) \mathbf{I}(\gamma_i \leq 1),$  for  $i = 1, \dots, p.$
- $\mathbf{g}|\cdot \sim \mathbf{N}_p(\mathbf{m}, \mathbf{S}) \prod_{i=1}^p \mathbf{I}(g_i \geq 0),$  where  $\mathbf{S}^{-1} = \mathbf{X}_0 \mathbf{V}^{-1} \mathbf{X}_0 + \sum_{t=1}^T (\mathbf{X}_t \Sigma^{-1} \mathbf{X}_t),$   $\mathbf{m} = \mathbf{S} \sum_{t=1}^T (\mathbf{X}_t \Sigma^{-1} \mathbf{y}_t)$  and  $\mathbf{X}_t = \text{diag}(w_{1,t}, \dots, w_{p,t}).$
- $u_t|\cdot \sim \mathbf{U} \left( 0, \frac{1}{\prod_{i=1}^p \sqrt{1+w_{i,t-1}^2}} \right),$  for all  $t = 1, \dots, T.$
- $k_t|\cdot \sim \mathbf{U} \left( 0, \exp \left\{ \sum_{i=1}^p \frac{w_{i,t}^2}{\gamma_i [1+w_{i,t-1}^2]} \right\} \right),$  for all  $t = 1, \dots, T.$
- $\mathbf{w}_0|\cdot \sim \mathbf{N}_p(\mathbf{0}, \mathbf{V}) \mathbf{I} \left( u_1 \leq \frac{1}{\prod_{i=1}^p \sqrt{1+w_{i,0}^2}} \right) \cdot \mathbf{I} \left( k_1 \leq \exp \left\{ -\frac{1}{2} \cdot \sum_{i=1}^p \frac{w_{i,1}^2}{\gamma_i (1+w_{i,0}^2)} \right\} \right)$
- $\mathbf{w}_t|\cdot \sim \mathbf{N}(\mathbf{m}_t, \mathbf{S}_t) \mathbf{I} \left( u_t \leq \frac{1}{\prod_{i=1}^p \sqrt{1+w_{i,t-1}^2}} \right) \cdot \mathbf{I} \left( k_t \leq \exp \left\{ -\frac{1}{2} \sum_{i=1}^p \frac{w_{i,t}^2}{\gamma_i (1+w_{i,t-1}^2)} \right\} \right),$  where  $\mathbf{S}_T^{-1}$  and  $\mathbf{m}_T$  are defined in *Remark 3*, for all  $t = 1, \dots, T-1.$
- $\mathbf{w}_T \sim \mathbf{N}_p(\mathbf{m}_T, \mathbf{S}_T),$  where  $\mathbf{S}_T^{-1}$  and  $\mathbf{m}_T$  are defined in *Remark 4*.

It has been mentioned that, in order to sample from the above truncated Normal densities the method of Robert (1995) is followed, and from the truncated Inverse Gamma densities we suggest using an additional latent variable as described in section 10.6.

### 10.2.3 Application

In this section an application of the multivariate unobserved ARCH model is presented using real data. In detail, the daily exchange rate of the US dollar (USD) and the Japanese Yen (JPY) with respect to the Greek Drachma (GRD) are used. The data set is consisted of 845 multivariate observations concerning the period (16/12/93 – 2/5/97) and illustrated in Figure 10-1 (pp 163). Using the algorithm that was described in the previous section, a sample from the posterior density of the models of interest can be generated and by using this sample inferences about the parameters can be made.

In detail, the MCMC algorithm was applied for 10000 iterations keeping the non-diagonal elements of the variance-covariance matrix  $\Sigma$  equal to 0; then the full algorithm was applied for 550000 iterations. This enabled the algorithm to quickly “locate” the high posterior regions of the posterior density. The first 50000 iterations were discarded as a burn-in and, for the remaining 500000 iterations one sample point is kept every 100 iterations in order to have an approximately independent sample from the posterior density of the parameters of the multivariate unobserved ARCH model. Figure 10-4 presents the final sample from the MCMC algorithm for each parameter of the multivariate unobserved ARCH model. In order to assess the convergence of the MCMC algorithm to the posterior distribution, the subsampling diagnostic (see section 4.3 and Giakoumatos *et al.*, 1999) is used to the initial 50000 iterations of the full algorithm. For the application of this criterion the following quantities were set:  $a = 0.05$ ,  $t = 0.99$ ,  $d = 0.999$ . The subsampling diagnostic pointed out that the proposed algorithm needs approximately 29000 iterations to get in the target distribution. The results of the diagnostic are presented in Figure 10-8. Moreover, Figure 10-9 presents the autocorrelation plots for the hyperparameters for the MCMC series

The summary statistics of the posterior marginal densities are presented in the Table 10.2.

From the above, it is obvious that both ARCH components, for USD and JPY, present high volatility persistence (  $E(b_{USD}) = 0.9077$  and  $E(b_{JPY}) = 0.8319$  ).

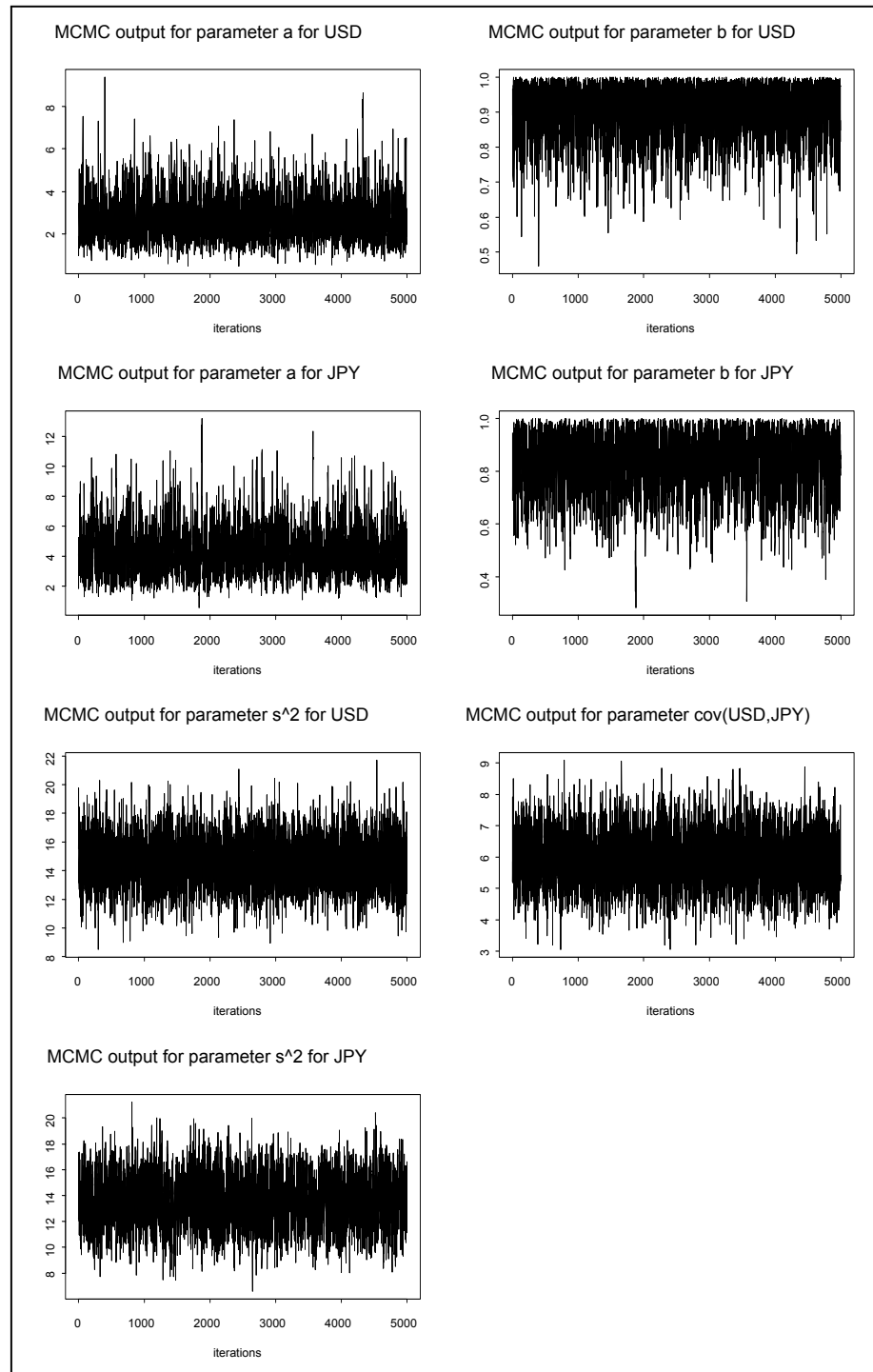


Figure 10-7: MCMC output for parameters of multivariate unobserved ARCH model

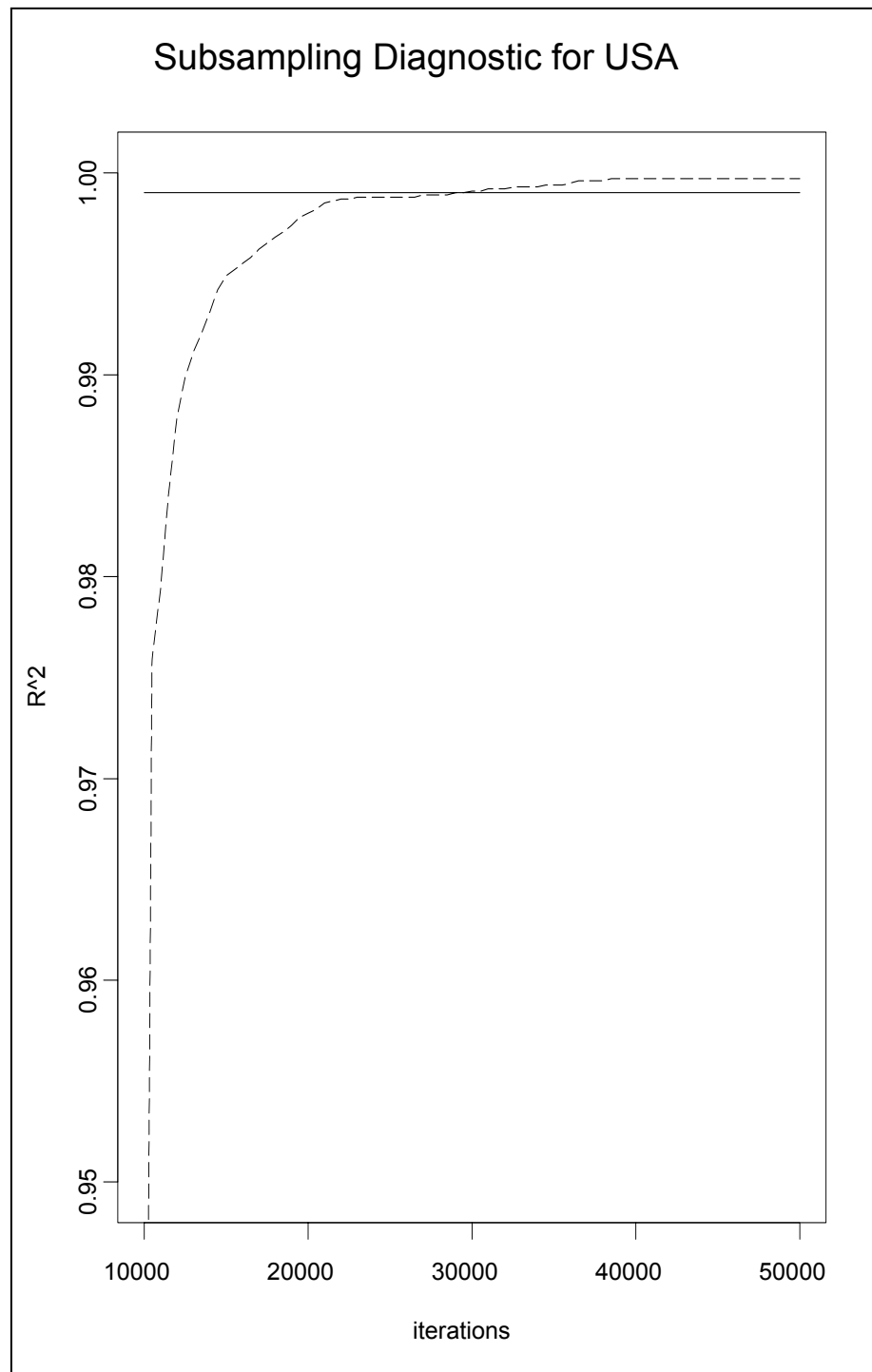


Figure 10-8: Subsampling diagnostic plot for multivariate unobserved ARCH model

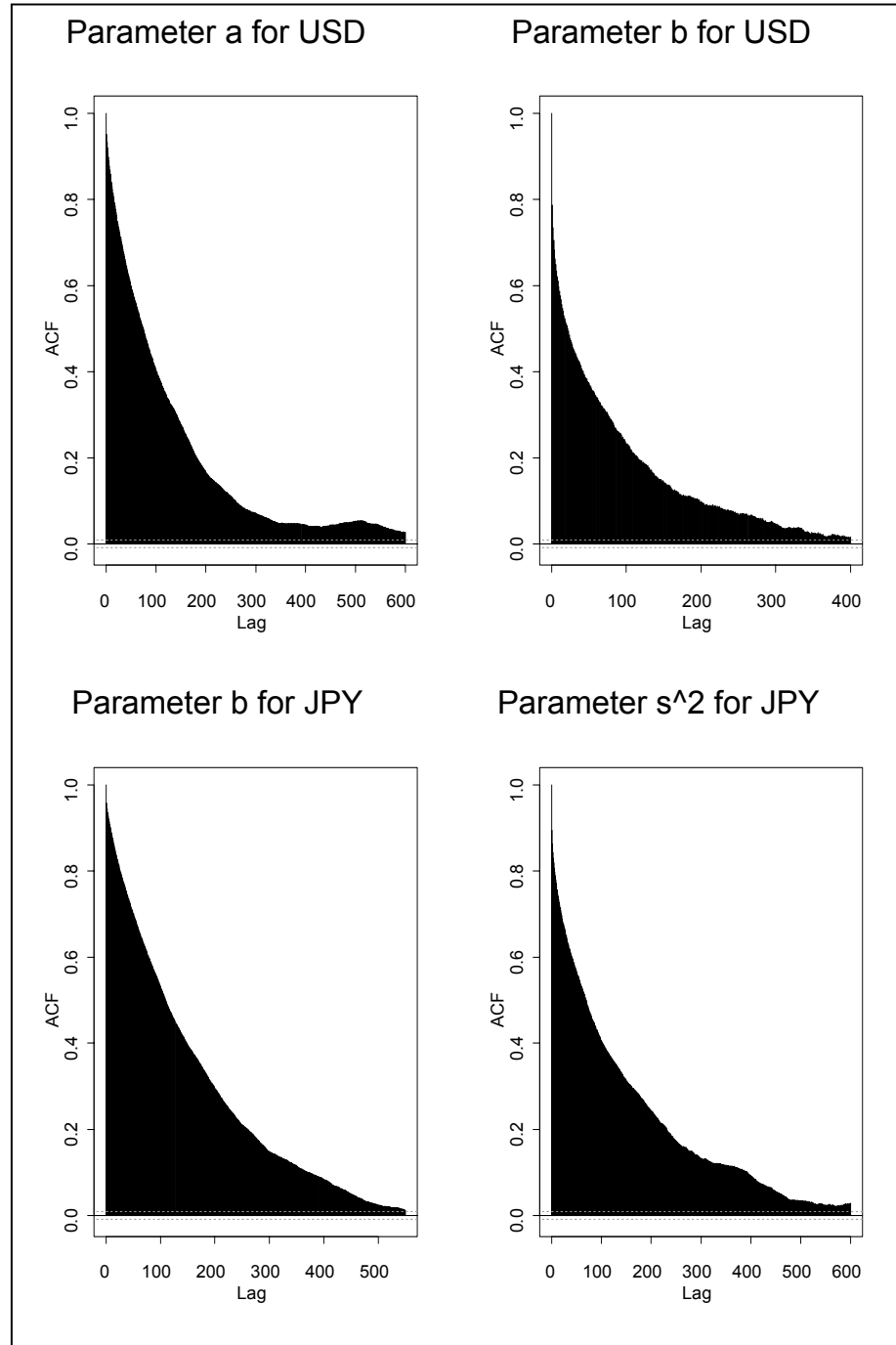


Figure 10-9: Autocorrelation function plots for the hyperparameters of the multivariate unobserved ARCH model

Parameter	Mean	Variance
USD $a$	2.717691	1.012218
USD $b$	0.9077305	0.006078998
JPY $a$	4.326326	2.508016
JPY $b$	0.8318802	0.01364302

Table 10.2: Summary statistics for the hyperparameters of the Unobserved ARCH model

	USD	JPY
USD	14.58924 (3.11593)	5.934796 (0.77334)
JPY	0.427934 (0.004285)	13.45407 (4.063)

Table 10.3: Estimates of the covariance matrix. Diagonal elements: variances. Upper diagonal elements: covariances. Lower diagonal elements: correlations

Table 10.3 contains the posterior mean and standard deviation estimates of the components of the underlying variance-covariance matrix  $\Sigma$ . It can be seen that there is positive correlation (0.43) between the USD and the JPY.

The posterior histograms of the hyperparameters of the multivariate unobserved ARCH model are presented in Figure 10-10 and for the parameters of  $\Sigma$  in Figure 10-11.

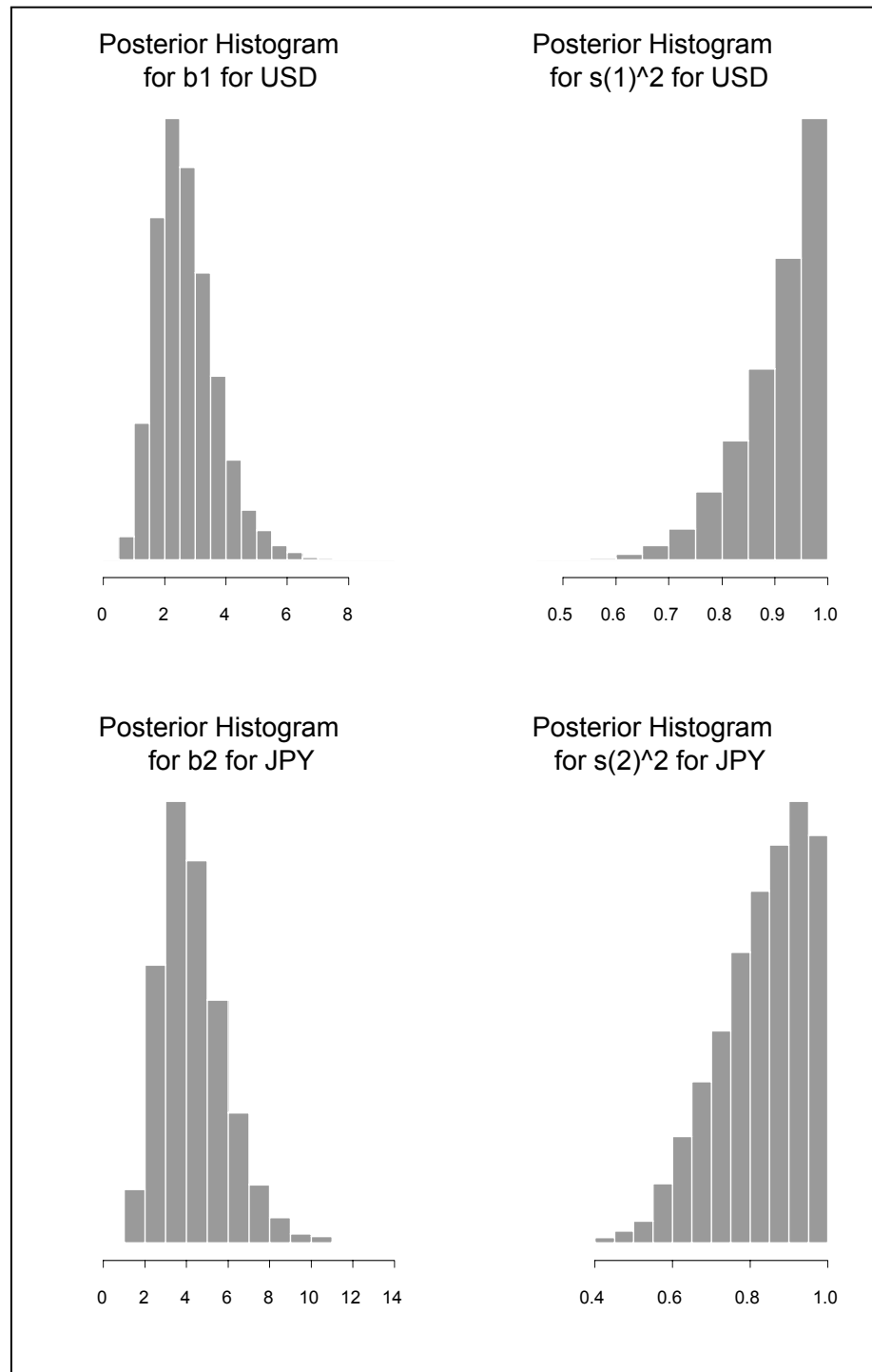


Figure 10-10: Posterior Histograms for the hyperparameters of the multivariate unobserved ARCH model

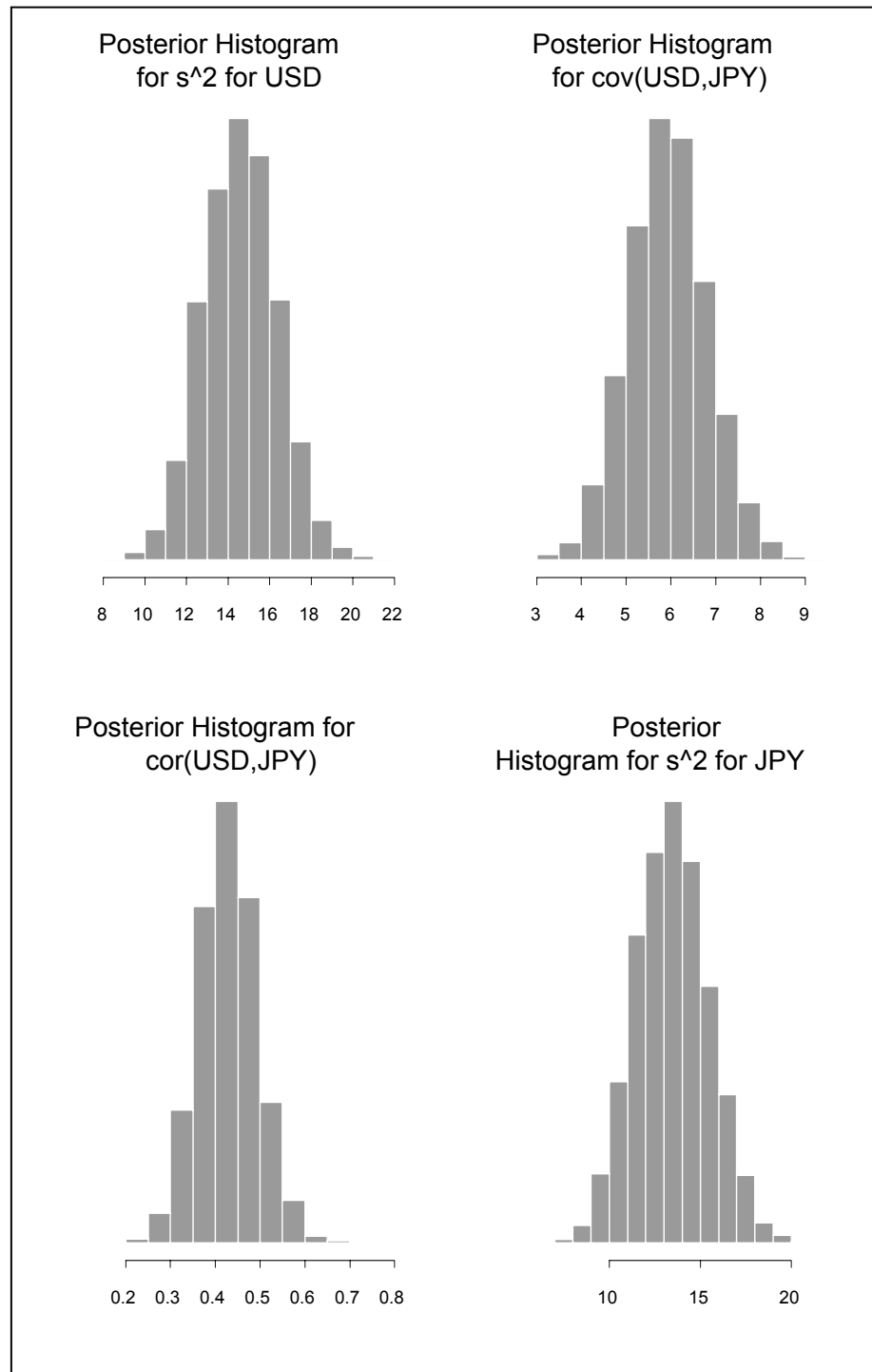


Figure 10-11: Posterior Histograms for the variance covariance matrix  $\Sigma$  of the multi-variate unobserved ARCH model

## 10.3 Latent Factor ARCH Model

### 10.3.1 Introduction

A more parsimonious representation of the multivariate unobserved ARCH model (10.9) is given by the latent factor ARCH model introduced by Diebold and Nerlove (1989) and King *et al.* (1994). Adopting again the hierarchical conditional densities, the latent factor ARCH model can be defined as following:

$$\begin{aligned} \mathbf{y}_t | \boldsymbol{\lambda}, f_t, \boldsymbol{\Sigma} &\sim \mathbf{N}_p(\boldsymbol{\lambda} f_t, \boldsymbol{\Sigma}), \\ f_t | a, f_{t-1} &\sim \mathbf{N}(0, \Delta_t), \\ \Delta_t &= 1 - a + a f_{t-1}^2, \quad t = 1, \dots, T; \end{aligned} \tag{10.12}$$

where  $\mathbf{y}_t$  is a  $p$ -variate realization of the stochastic process at time  $t$ ,  $f_t$  is the ‘unique’ common unobserved ARCH factor at time  $t$ ,  $\boldsymbol{\lambda}$  is the  $(p \times 1)$ -variate vector of parameters reflecting the sensitivity of the common factor  $f_t$ ,

$\boldsymbol{\Sigma} = \text{diag}(\gamma_1, \dots, \gamma_p)$  is a diagonal  $(p \times p)$  covariance matrix and  $a$  is the hyperparameter of the ARCH component that takes values in  $(0, 1)$ . The model (10.12) satisfies sufficient conditions which ensure parameter identifiability; see Diebold and Nerlove (1989). The unconditional variance-covariance matrix of the process  $\mathbf{y}_t$  is

$$\text{Var}(\mathbf{y}_t) = \boldsymbol{\lambda} \cdot \boldsymbol{\lambda}' + \boldsymbol{\Sigma}.$$

The parameter vector of (10.12) contains the hyperparameter  $a$ , the vector of loadings  $\boldsymbol{\lambda}$ , the vector of the unobserved component  $\mathbf{f} = (f_1, \dots, f_T)$ , the ‘history’  $f_0$  of the unique unobserved component and the vector  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_p)$  of the covariance matrix  $\boldsymbol{\Sigma}$ .

### 10.3.2 Bayesian Approach and the Auxiliary Variable Sampler

In this section a full Bayesian analysis for the latent factor ARCH model is presented and an auxiliary variable algorithm is proposed in order to sample from the parameters of interest of the latent factor ARCH model.

Denoting all data by the  $T \times p$  matrix  $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_T)$ , the joint posterior density of model (10.12) can be written as

$$f(\boldsymbol{\lambda}, \mathbf{f}, a, \boldsymbol{\gamma}, f_0 | \mathbf{Y}) \propto \prod_{t=1}^T (f(\mathbf{y}_t | \boldsymbol{\lambda}, f_t, \boldsymbol{\gamma}) f(f_t | f_{t-1}, a)) \pi(\boldsymbol{\lambda}, \boldsymbol{\gamma}, a, f_0)$$

where  $\pi(\boldsymbol{\lambda}, a, \boldsymbol{\gamma}, f_0)$  denotes the joint prior density of the parameters of the models. Assuming a priori independence for the parameters,  $\pi(\boldsymbol{\lambda}) \propto \text{constant}$ ,  $\pi(\boldsymbol{\gamma}) \propto \left( \prod_{i=1}^p \gamma_i \right)^{-1}$ , for  $i = 1, \dots, p$ ,  $a \sim \mathbf{U}(0, 1)$ , where  $\mathbf{U}(\cdot, \cdot)$  denotes the uniform distribution and  $f_0 \sim \mathbf{N}(0, s)$  are chosen. This results to a joint prior density of the form

$$\pi(\boldsymbol{\lambda}, \boldsymbol{\gamma}, a, f_0) \propto \left( \prod_{i=1}^p \gamma_i \right)^{-1} \exp \left\{ -\frac{1}{2} \cdot \frac{f_0^2}{s} \right\}.$$

Then, the joint posterior density of the parameters of the latent factor ARCH model can be written as

$$\begin{aligned} f(\boldsymbol{\lambda}, \mathbf{f}, a, \boldsymbol{\gamma}, f_0 | \mathbf{Y}) &\propto \left( \prod_{i=1}^p \gamma_i^{\frac{T}{2}+1} \right)^{-1} \left( \prod_{t=1}^T \sqrt{1 - a + a f_{t-1}^2} \right)^{-1} \exp \left\{ -\frac{1}{2} \frac{f_0^2}{s} \right\} \\ &\cdot \exp \left\{ -\frac{1}{2} \sum_{t=1}^T \left( \frac{f_t^2}{1 - a + a f_{t-1}^2} + \sum_{i=1}^p \frac{(y_{i,t} - \lambda_i f_t)^2}{\gamma_i} \right) \right\}. \end{aligned} \quad (10.13)$$

The above joint posterior density does not have full conditional densities of known form. In order to overcome this problem, the following theorem is utilized.

**Theorem 21** *If in the posterior density  $\pi(\boldsymbol{\lambda}, \mathbf{f}, a, \boldsymbol{\gamma}, f_0 | \mathbf{Y})$  in (10.13) we include*

*$\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_T)'$  and  $\mathbf{z} = (z_1, \dots, z_T)'$  positive random variates such that the joint density*

$f(\boldsymbol{\lambda}, \mathbf{f}, a, \boldsymbol{\gamma}, f_0, \boldsymbol{\zeta}, \mathbf{z} | \mathbf{Y})$  takes the form

$$\begin{aligned} f(\boldsymbol{\lambda}, \mathbf{f}, a, \boldsymbol{\gamma}, f_0, \boldsymbol{\zeta}, \mathbf{z} | \mathbf{Y}) &\propto \left( \prod_{i=1}^p \gamma_i^{\frac{T}{2}+1} \right)^{-1} \exp \left\{ -\frac{1}{2} \left( \frac{f_0^2}{s} + \sum_{t=1}^T \sum_{i=1}^p \frac{(y_{i,t} - \lambda_i f_t)^2}{\gamma_i} \right) \right\} \\ &\cdot \left\{ \prod_{t=1}^T \mathbf{I} \left( z_t \leq \exp \left\{ -\frac{1}{2} \frac{f_t^2}{1-a+a f_{t-1}^2} \right\} \right) \right\} \\ &\cdot \left\{ \prod_{t=1}^T \mathbf{I} \left( \zeta_t \leq \frac{1}{\sqrt{1-a+a f_{t-1}^2}} \right) \right\}, \end{aligned}$$

then the marginal density of  $f(\boldsymbol{\lambda}, \mathbf{f}, a, \boldsymbol{\gamma}, f_0 | \mathbf{Y})$  is given by (10.13).

*Proof:* Let  $\boldsymbol{\theta} = (\boldsymbol{\lambda}, \mathbf{f}, a, \boldsymbol{\gamma}, f_0)$ , then

$$\begin{aligned} f(\boldsymbol{\theta} | \mathbf{Y}) &= \int \cdots \int [\boldsymbol{\theta}, \mathbf{k}, \mathbf{z} | \mathbf{Y}] d\zeta_1 \cdots d\zeta_T dz_1 \cdots dz_T \\ &= \frac{1}{\left( \prod_{i=1}^p \gamma_i^{\frac{T}{2}+1} \right)} \cdot \exp \left\{ -\frac{1}{2} \cdot \left( \frac{f_0^2}{s} + \sum_{t=1}^T \sum_{i=1}^p \frac{(y_{i,t} - \lambda_i \cdot f_t)^2}{\gamma_i} \right) \right\} \\ &\cdot \int_0^{\exp \left\{ -\frac{1}{2} \cdot \frac{f_1^2}{1-a+a \cdot f_0^2} \right\}} dz_1 \cdots \int_0^{\exp \left\{ -\frac{1}{2} \cdot \frac{f_T^2}{1-a+a \cdot f_{T-1}^2} \right\}} dz_T \\ &\cdot \int_0^{\frac{1}{\sqrt{1-a+a \cdot f_0^2}}} d\zeta_1 \cdots \int_0^{\frac{1}{\sqrt{1-a+a \cdot f_{T-1}^2}}} d\zeta_T \\ &= \frac{1}{\left( \prod_{i=1}^p \gamma_i^{\frac{T}{2}+1} \right)} \cdot \frac{1}{\left( \prod_{t=1}^T \sqrt{1-a+a \cdot f_{t-1}^2} \right)} \cdot \exp \left\{ -\frac{1}{2} \cdot \frac{f_0^2}{s} \right\} \\ &\exp \left\{ -\frac{1}{2} \cdot \sum_{t=1}^T \left( \frac{f_t^2}{1-a+a \cdot f_{t-1}^2} + \sum_{i=1}^p \frac{(y_{i,t} - \lambda_i \cdot f_t)^2}{\gamma_i} \right) \right\} \blacksquare \end{aligned}$$

The above theorem guarantees that a MCMC algorithm which obtains samples from  $f(\boldsymbol{\lambda}, \mathbf{f}, a, \boldsymbol{\gamma}, f_0, \boldsymbol{\zeta}, \mathbf{z} | \mathbf{Y})$  also obtains samples from  $f(\boldsymbol{\lambda}, \mathbf{f}, a, \boldsymbol{\gamma}, f_0 | \mathbf{Y})$ . Moreover, the construction of this MCMC algorithm is straightforward using the following Gibbs steps:

- $\gamma_i | \cdot \sim \mathbf{IG} \left( \frac{T}{2}, \frac{\sum_{t=1}^T (y_{i,t} - \lambda_i \cdot f_t)^2}{2} \right)$ , for all  $i = 1, \dots, p$ .

- $\lambda_i | \cdot \sim \mathbf{N} \left( \frac{\sum_{t=1}^T f_t \cdot y_{i,t}}{\sum_{t=1}^T f_t^2}, \frac{\gamma_i}{\sum_{t=1}^T f_t^2} \right)$ , for all  $i = 1, \dots, p$ .

- $a | \cdot \sim \mathbf{U}(L, U)$ , where

$$L = \max \left\{ \max_t \{ \phi_t : |f_{t-1}| \geq 1 \}, \max_t \{ \rho_t : |f_{t-1}| < 1 \}, 0 \right\},$$

$$U = \min \left\{ \min_t \{ \phi_t : |f_{t-1}| \leq 1 \}, \min_t \{ \rho_t : |f_{t-1}| > 1 \}, 1 \right\}$$

and

$$\phi_t = \frac{\left( -0.5 \frac{f_t^2}{\ln(z_t)} - 1 \right)}{(f_{t-1}^2 - 1)}, \quad \rho_t = \frac{\left( \frac{1}{\zeta_t^2} - 1 \right)}{(f_{t-1}^2 - 1)}.$$

- $\zeta_t | \cdot \sim \mathbf{U} \left( 0, \frac{1}{\sqrt{1-a+a \cdot f_{t-1}^2}} \right)$ , for all  $t = 1, \dots, T$ .

- $z_t | \cdot \sim \mathbf{U} \left( 0, \exp \left\{ -\frac{1}{2} \cdot \frac{f_t^2}{1-a+a \cdot f_{t-1}^2} \right\} \right)$ , for all  $t = 1, \dots, T$ .

- $f_0 | \cdot \sim N(0, s) \mathbf{I}(\Upsilon)$ , where  $\Upsilon = \{L \cup U\}$  and

$$L = \left( -\sqrt{\frac{1}{a\zeta_1^2} - \frac{1}{a} + 1}, -\sqrt{-\frac{f_1^2}{2a \ln(z_1)} - \frac{1}{a} + 1} \right),$$

$$U = \left( \sqrt{-\frac{f_1^2}{2a \ln(z_1)} - \frac{1}{a} + 1}, \sqrt{\frac{1}{a\zeta_1^2} - \frac{1}{a} + 1} \right)$$

- $f_t | \cdot \sim N(\mu_t, \sigma_t^2) \mathbf{I}(\Upsilon)$ , where

$$\mu_t = \frac{\sum_{i=1}^P \lambda_i \gamma_i y_{i,t}}{\sum_{i=1}^P \lambda_i^2 \gamma_i}, \quad \sigma_t^2 = \frac{\prod_{i=1}^P \gamma_i}{\sum_{i=1}^P \lambda_i^2 \gamma_i} \quad (10.14)$$

$$\Upsilon = \left\{ \left( L, -\sqrt{-\frac{f_{t+1}^2}{2a \ln(z_{t+1})} - \frac{1}{a} + 1} \right) \cup \left( \sqrt{-\frac{f_{t+1}^2}{2a \ln(z_{t+1})} - \frac{1}{a} + 1}, U \right) \right\},$$

and

$$L = \max \left( -\sqrt{\frac{1}{a \cdot \zeta_{t+1}^2} - \frac{1}{a}} + 1, -\sqrt{-2(1 - a + af_{t-1}^2) \ln(z_t)} \right),$$

$$U = \min \left( \sqrt{\frac{1}{a \zeta_{t+1}^2} - \frac{1}{a}} + 1, \sqrt{-2(1 - a + af_{t-1}^2) \ln(z_t)} \right),$$

for  $t = 1, \dots, T - 1$ .

- $f_T | \cdot \sim N(\mu_T, \sigma_T^2) \mathbf{I}(L, U)$ , where  $\mu_T$  and  $\sigma_T^2$  are calculated as in formula (10.14) and the bounds  $L, U$  are given by

$$L = \left( -\sqrt{-2(1 - a + af_{T-1}^2) \ln(z_T)} \right),$$

$$U = \left( \sqrt{-2(1 - a + a \cdot f_{T-1}^2) \ln(z_T)} \right).$$

### 10.3.3 Application

In this section an application of the latent factor ARCH model is presented using real data. In detail, the daily exchange rate of the US dollar (USD) and the Japanese Yen (JPY) with respect to the Greek Drachma (GRD) are used. The data set is consisted of 845 multivariate observations concerning the period (16/12/93 – 2/5/97) and illustrated in Figure 10-1 (pp 163). Using the algorithm that has been described in the previous section, a sample from the posterior density of the latent factor ARCH model is generated and by using this sample, inferences about the parameters can be derived.

In detail the MCMC algorithm was applied for 350000 iterations. The initial 150000 iterations were discarded as burn-in. In order to ensure that the Markov chain produced by the algorithm gets in the target distributions the subsampling diagnostic was used. This test pointed out that the proposed algorithm requires approximately 81000 iterations

		95% Credible Interval	
Parameter	Posterior Mean	Lower Limit	Upper Limit
$\lambda_{\text{USD}}$	-16.43321606	-28.40609990	51.26371137
$\lambda_{\text{JPY}}$	-2.27515464	-3.91430508	6.97497419
$\alpha$	0.92577876	0.85203707	0.99999999
$\gamma_{\text{USD}}$	13.55768267	12.92785260	14.43662310
$\gamma_{\text{JPY}}$	27.07353838	27.34464947	27.62347193

Table 10.4: Posterior means and 95% credible intervals for the parameters of the latent factor ARCH model

to get in the target distribution. The results of the diagnostic are presented in Figure 10-12.

The final sample was selected, by keeping 1 sample point every 20 iterations from the remaining 200000 iterations . The final sample was consisted by 10000 points and presented in Figure 10-13.

In addition, the final posterior sample was checked for convergence to the limiting distribution by applying the test of Geweke (1992), Raftery and Lewis (1992) and Heidelberger and Welch (1983).

However, it has been observed the existence of an extremely high autocorrelation of the parameter chains as it becomes obvious from Figure 10-13. For this reason, and in order to save computer disk and memory, subsampling techniques were applied (see Politis 1998; Politis *et al.*, 1999) for the calculation of the 95% credible intervals for the ergodic means of the parameters of the latent factor ARCH model. The construction of the credible intervals is presented in detail in section 10.6.

The ergodic means and the 95% credible intervals of the parameters of the latent factor ARCH model are included in Table 10.4.

Based upon the posterior ergodic mean for the parameter  $a$  (0.926) it becomes obvious that the ARCH component is highly volatility persistent.

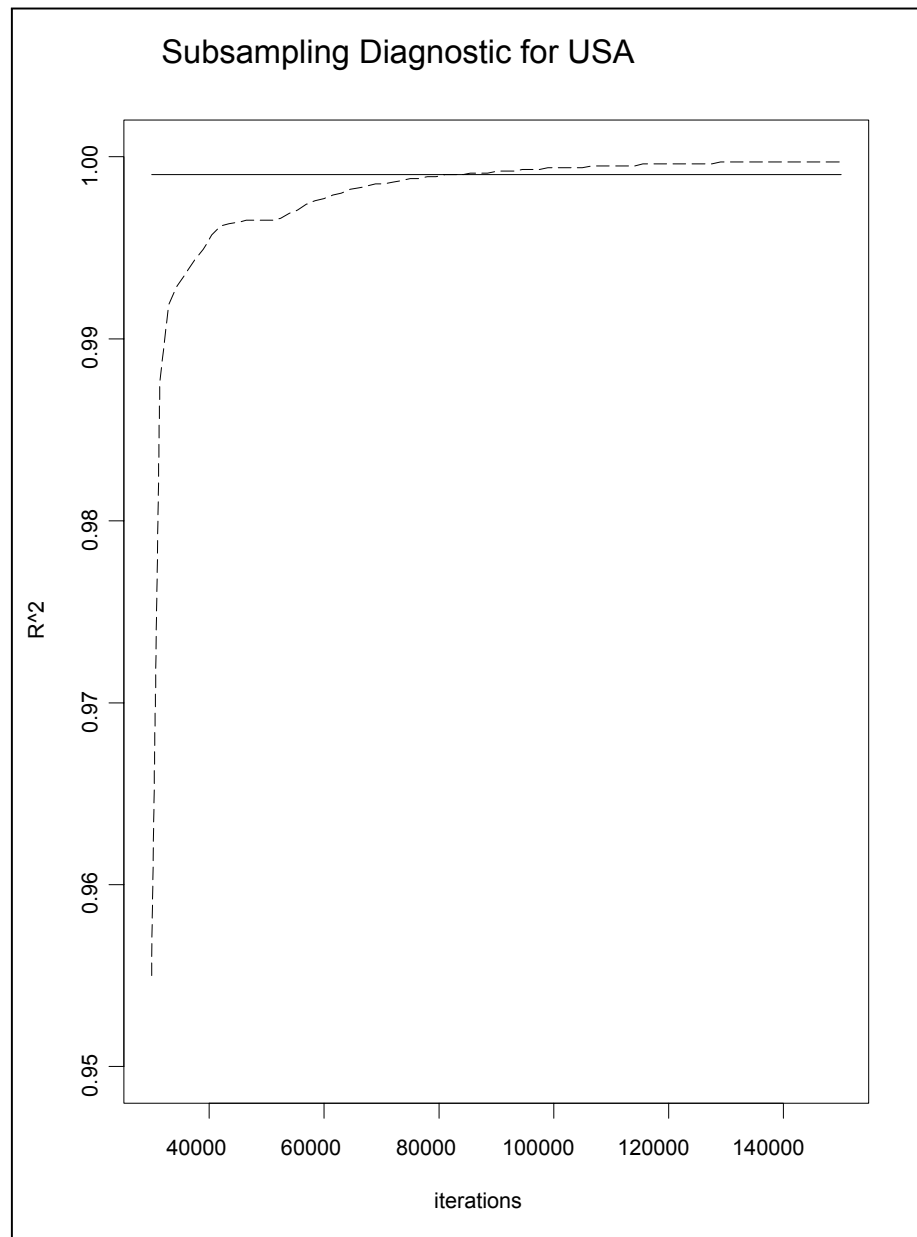


Figure 10-12: Subsampling diagnostic plot for latent factor ARCH model

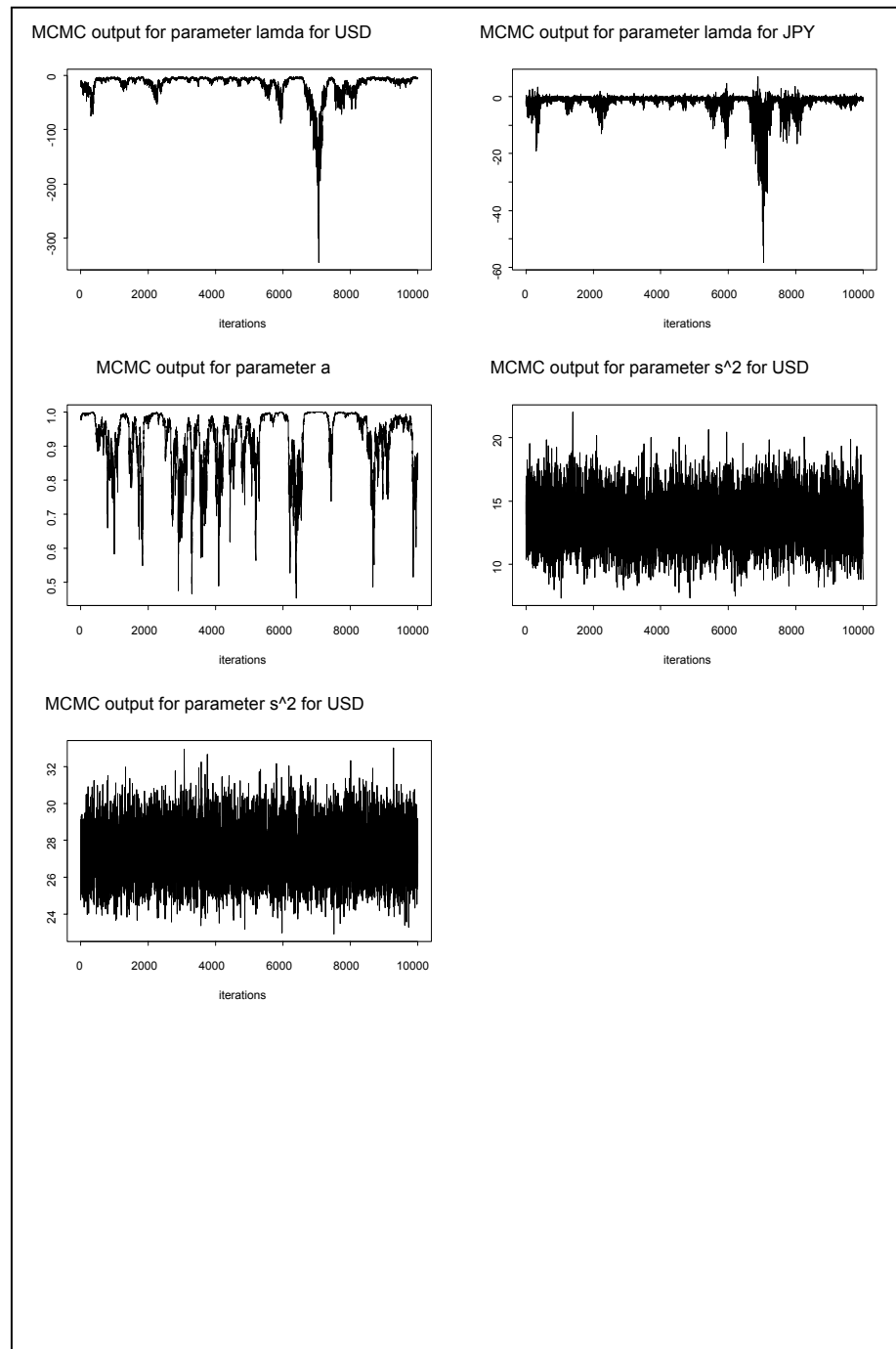


Figure 10-13: MCMC output for parameters of latent factor ARCH model

## 10.4 Latent Factor GARCH Model

### 10.4.1 Introduction

A straightforward extension of the latent factor ARCH model is the latent factor GARCH model. In this case, the unobserved component follows a GARCH process. In detail, the latent factor GARCH model can be written with the following hierarchical structure of conditional densities.

$$\begin{aligned} \mathbf{y}_t | \boldsymbol{\lambda}, f_t, \boldsymbol{\Sigma} &\sim \mathbf{N}_p(\boldsymbol{\lambda} f_t, \boldsymbol{\Sigma}), \\ f_t | a, b, f_{t-1} &\sim \mathbf{N}(0, \sigma_t^2), \\ \Delta_t &= 1 - a - b + a f_{t-1}^2 + b \sigma_{t-1}^2, \quad t = 1, \dots, T; \end{aligned} \tag{10.15}$$

where  $\mathbf{y}_t$  is a  $p$ -variate realization of the stochastic process at time  $t$ ,  $f_t$  is the ‘unique’ common unobserved GARCH factor at time  $t$ ,  $\boldsymbol{\lambda}$  is the  $(p \times 1)$ -variate vector of parameters reflecting the sensitivity of the common factor  $f_t$ ,  $\boldsymbol{\Sigma} = \text{diag}(\gamma_1, \dots, \gamma_p)$  is a diagonal  $(p \times p)$  covariance matrix and  $a, b$  are the hyperparameter of the GARCH component that take values in  $(0, 1)$  with restriction  $a + b < 1$ . The model (10.15) satisfies the sufficient conditions which ensure the parameter identifiability; see Diebold and Nerlove (1989). The unconditional variance-covariance matrix of the process  $\mathbf{y}_t$  is

$$\text{Var}(\mathbf{y}_t) = \boldsymbol{\lambda} \cdot \boldsymbol{\lambda}' + \boldsymbol{\Sigma}.$$

The parameter vector of (10.15) contains the hyperparameter  $a$ , the vector of loadings  $\boldsymbol{\lambda}$ , the vector of the unobserved component  $\mathbf{f} = (f_1, \dots, f_T)$ , the ‘history’  $f_0$  of the unique unobserved component and the vector  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_p)$  of the covariance matrix  $\boldsymbol{\Sigma}$ .

## 10.4.2 Bayesian Approach and the Auxiliary Variable Sampler

In this section a full Bayesian analysis of the latent factor GARCH model is presented and an MCMC algorithm is proposed in order to take a sample from the posterior density of model's parameters.

Denoting all data by the  $T \times p$  matrix  $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_T)$ , the joint posterior density of model (10.15) can be written as

$$f(\boldsymbol{\lambda}, \mathbf{f}, a, b, \boldsymbol{\gamma}, f_0 | \mathbf{Y}) \propto \prod_{t=1}^T (f(\mathbf{y}_t | \boldsymbol{\lambda}, f_t, \boldsymbol{\gamma}) f(f_t | f_{t-1}, a, b)) \pi(\boldsymbol{\lambda}, \boldsymbol{\gamma}, a, b, f_0) \quad (10.16)$$

where  $\pi(\boldsymbol{\lambda}, a, b, \boldsymbol{\gamma}, f_0)$  denotes the joint prior density of the parameters of the model. Assuming a priori independence for the parameters,  $\pi(\boldsymbol{\lambda}) \propto \text{constant}$ ,  $\pi(\boldsymbol{\gamma}) \propto \left(\prod_{i=1}^p \gamma_i\right)^{-1}$ , for  $i = 1, \dots, p$ ,  $a, b \sim \mathbf{U}(0, 1)$ , where  $\mathbf{U}(\cdot, \cdot)$  denotes the uniform distribution and  $f_0 \sim \mathbf{N}(0, s)$  are chosen. This results to a joint prior density of the form

$$f(\boldsymbol{\lambda}, \boldsymbol{\gamma}, a, b, f_0) \propto \left(\prod_{i=1}^p \gamma_i\right)^{-1} \exp\left\{-\frac{1}{2} \cdot \frac{f_0^2}{s}\right\}.$$

Then, the joint posterior density of the parameters of the latent factor GARCH model can be written as

$$\begin{aligned} f(\boldsymbol{\lambda}, \mathbf{f}, a, b, \boldsymbol{\gamma}, f_0 | \mathbf{Y}) &= g(\boldsymbol{\lambda}, \mathbf{f}, a, b, \boldsymbol{\gamma}, f_0) \\ &\propto \left(\prod_{i=1}^p \gamma_i^{\frac{T}{2}+1}\right)^{-1} \left(\prod_{t=1}^T \sqrt{1 - a - b + a f_{t-1}^2 + b \sigma_t^2}\right)^{-1} \exp\left\{-\frac{1}{2} \frac{f_0^2}{s}\right\} \\ &\quad \exp\left\{-\frac{1}{2} \sum_{t=1}^T \left(\frac{f_t^2}{1 - a - b + a f_{t-1}^2 + b \sigma_t^2} + \sum_{i=1}^p \frac{(y_{i,t} - \lambda_i f_t)^2}{\gamma_i}\right)\right\}. \end{aligned} \quad (10.17)$$

The resulting full conditional densities of the above joint posterior density (10.17) are not of standard form, so the auxiliary variable sampler is adopted in order to take sample from the posterior density. The following Theorem reveals the usage of the auxiliary

sampler in this case.

**Theorem 22** *If in the posterior density  $f(\boldsymbol{\lambda}, \mathbf{f}, a, b, \boldsymbol{\gamma}, f_0 | \mathbf{Y})$  in (10.17) we include a positive random variate  $u$  such that the joint density  $f(\boldsymbol{\lambda}, \mathbf{f}, a, b, \boldsymbol{\gamma}, f_0, u | \mathbf{Y})$  takes the form*

$$f(\boldsymbol{\lambda}, \mathbf{f}, a, b, \boldsymbol{\gamma}, f_0, u | \mathbf{Y}) \propto I(u \leq g(\boldsymbol{\lambda}, \mathbf{f}, a, b, \boldsymbol{\gamma}, f_0)), \quad (10.18)$$

where  $I(\cdot)$  is the indicator function, then the marginal density of (10.18) marginalize out the auxiliary parameter  $u$  is given by (10.13).

**Proof:**

$$\begin{aligned} f(x) &= \int_u I(u \leq g(\boldsymbol{\lambda}, \mathbf{f}, a, b, \boldsymbol{\gamma}, f_0)) du \\ &= \int_0^{g(\boldsymbol{\lambda}, \mathbf{f}, a, b, \boldsymbol{\gamma}, f_0)} 1 du \\ &= g(\boldsymbol{\lambda}, \mathbf{f}, a, b, \boldsymbol{\gamma}, f_0). \blacksquare \end{aligned}$$

Based on the above Theorem, the proposed MCMC algorithm is given by the following full conditional densities:

- $u | \cdot \sim \mathbf{U}(\cdot, \cdot)$
- $\gamma_i | \cdot \sim \mathbf{IG}(\cdot, \cdot)$ , for all  $i = 1, \dots, p$ , given that  $u$  is marginalized out.
- $\lambda_i | \cdot \sim \mathbf{N}(\cdot, \cdot)$ , for all  $i = 1, \dots, p$ , given that  $u$  is marginalized out.
- $f_T | \cdot \sim \mathbf{N}(\cdot, \cdot)$ , given that  $u$  is marginalized out.
- Sample  $a$  from  $\mathbf{U}(\cdot, \cdot)$  such as  $\{u < g(\boldsymbol{\lambda}, \mathbf{f}, a, b, \boldsymbol{\gamma}, f_0)\}$ .
- Sample  $b$  from  $\mathbf{U}(\cdot, \cdot)$  such as  $\{u < g(\boldsymbol{\lambda}, \mathbf{f}, a, b, \boldsymbol{\gamma}, f_0)\}$ .
- Sample  $f_t$ ,  $t = 1, \dots, T - 1$ , from  $\mathbf{U}(\cdot, \cdot)$  such as  $\{u < g(\boldsymbol{\lambda}, \mathbf{f}, a, b, \boldsymbol{\gamma}, f_0)\}$ .

In the above algorithm the  $g(\boldsymbol{\lambda}, \mathbf{f}, a, b, \boldsymbol{\gamma}, f_0)$  is not invertible with respect to its parameters  $a$ ,  $b$  and  $f_t$ ,  $t = 1, \dots, T - 1$ . For this reason Neal's approach (2003) is adopted in order to find out an interval  $I = (L, R)$  around the parameter of interest that contains at least a big part of set  $S = \{u < g(\boldsymbol{\lambda}, \mathbf{f}, a, b, \boldsymbol{\gamma}, f_0)\}$ . In detail, for parameter  $f_t$ ,  $t = 1, \dots, T - 1$  an initial interval  $I = (L, R)$  of size  $w$  is randomly picked such as to contain the current value of  $f_t$ . This interval is expanded using the 'stepping out' procedure (Neal, 2003) such as to contain a big part of  $S$ . Then  $f_t$  is uniformly sampled from the interval  $I$  and each time that a new point is sampled out of  $S$  the interval is shrunk - using Neal's (2003) shrink procedure - , until an  $f_t$  from  $S$  is sampled. For the parameters  $a$  and  $b$  the initial intervals are set equal to  $I_a = (0, 1 - b)$  and  $I_b = (0, 1 - a)$  respectively. Then,  $a$  and  $b$  are uniformly sampled from  $I_a$  and  $I_b$  respectively and each time that a new point is sampled out of  $S$  the intervals are narrowed - using the shrink procedure (Neal, 2003). However, the correlation between these parameters is very high and for this reason we choose to update the parameters  $a$  and  $b$  simultaneously by a dependent Metropolis-Hasting step. In detail, as candidate density for the Metropolis-Hastings step is chosen to be the bivariate Normal density. The mean vector of this density is the previous values for  $a$  and  $b$  and the variance covariance matrix is chosen to incorporate the high correlation between  $a$  and  $b$  and the acceptance rate to be approximately 50%.

### 10.4.3 Application

In this section an application of the latent factor GARCH model is presented using real data. In detail, the daily exchange rate of the US dollar (USD) and the Japanese Yen (JPY) with respect to the Greek Drachma (GRD) is used. The data set is consisted of 845 multivariate observations concerning the period (16/12/93 – 2/5/97) and illustrated in Figure 10-1 (pp 163). Using the algorithm that were described in the previous section, a sample was generated from the posterior density of the model and used for inference purposes about the parameters.

In order to generate a sample from the posterior density of the parameters of the

		95% Credible Interval	
Parameter	Posterior Mean	Lower Limit	Upper Limit
$\lambda_{\text{USD}}$	6.99887723	8.85351217	10.80469852
$\gamma_{\text{USD}}$	12.37400806	0.95026350	14.23300884
$\lambda_{\text{JPY}}$	1.29230672	0.95026350	1.65374465
$\gamma_{\text{JPY}}$	27.30040048	27.04663690	27.54712884
$\alpha$	0.71165224	0.71020596	0.71284911
$b$	0.26516888	0.26400163	0.26653133

Table 10.5: Posterior means and 95% credible intervals for the parameters of the latent factor GARCH model.

latent factor GARCH model, the proposed MCMC algorithm was applied for 550000 iterations. The initial 50000 sample points were discarded as burn-in. In order to ensure that the Markov chain that produced by the algorithm gets in the target distributions, the subsampling diagnostic (see section 4.3 and Giakoumatos *et al.*, 1999) was applied. This test pointed out that the proposed algorithm requires approximately 27000 iterations to get in the target distribution. The results of the diagnostic are presented in Figure 10-14.

The final sample was selected, by choosing 1 sample point every 20 iterations from the remaining 500000 iterations (in order minimize the requirements of CPU and storage memory capacity). The final sample is consisted by 25000 points and presented in Figure 10-15.

In addition, the final sample was also checked for convergence to the limiting distribution by the criteria of Geweke (1992), Raftery and Lewis (1992) and Heidelberger and Welch (1983).

However, an extremely high autocorrelation of the parameter chains was observed (as this becomes obvious from Figure 10-15). For this reason, subsampling techniques (see Politis 1998, Politis *et al.*, 1999) are used to calculate the 95% credible intervals for the ergodic means of the parameters of the latent factor GARCH model. The construction of the credible intervals is presented analytically in section 10.6.

The ergodic means and the 95% credible intervals of the parameters of the latent factor GARCH model are included in Table 10.5.

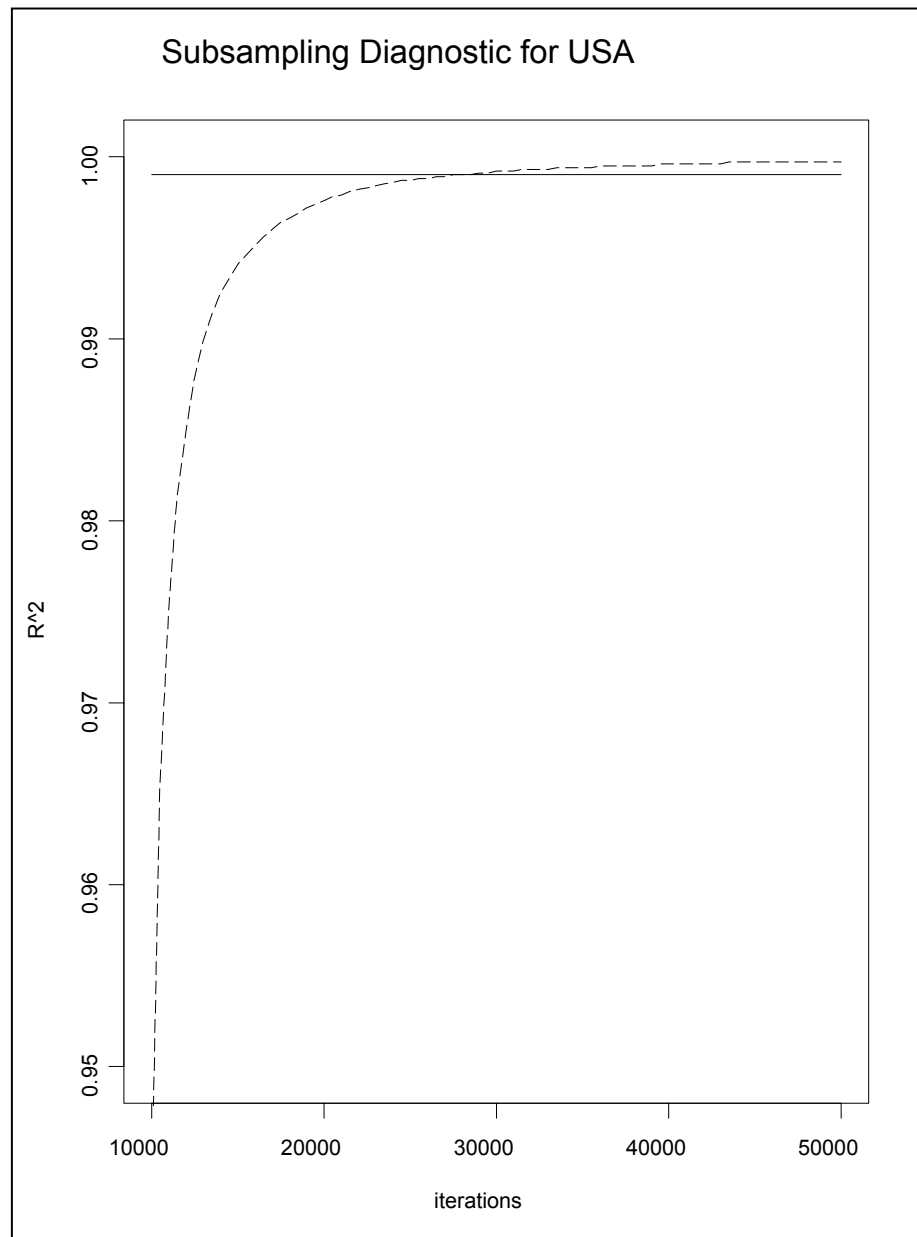


Figure 10-14: Subsampling diagnostic plot for latent factor GARCH model

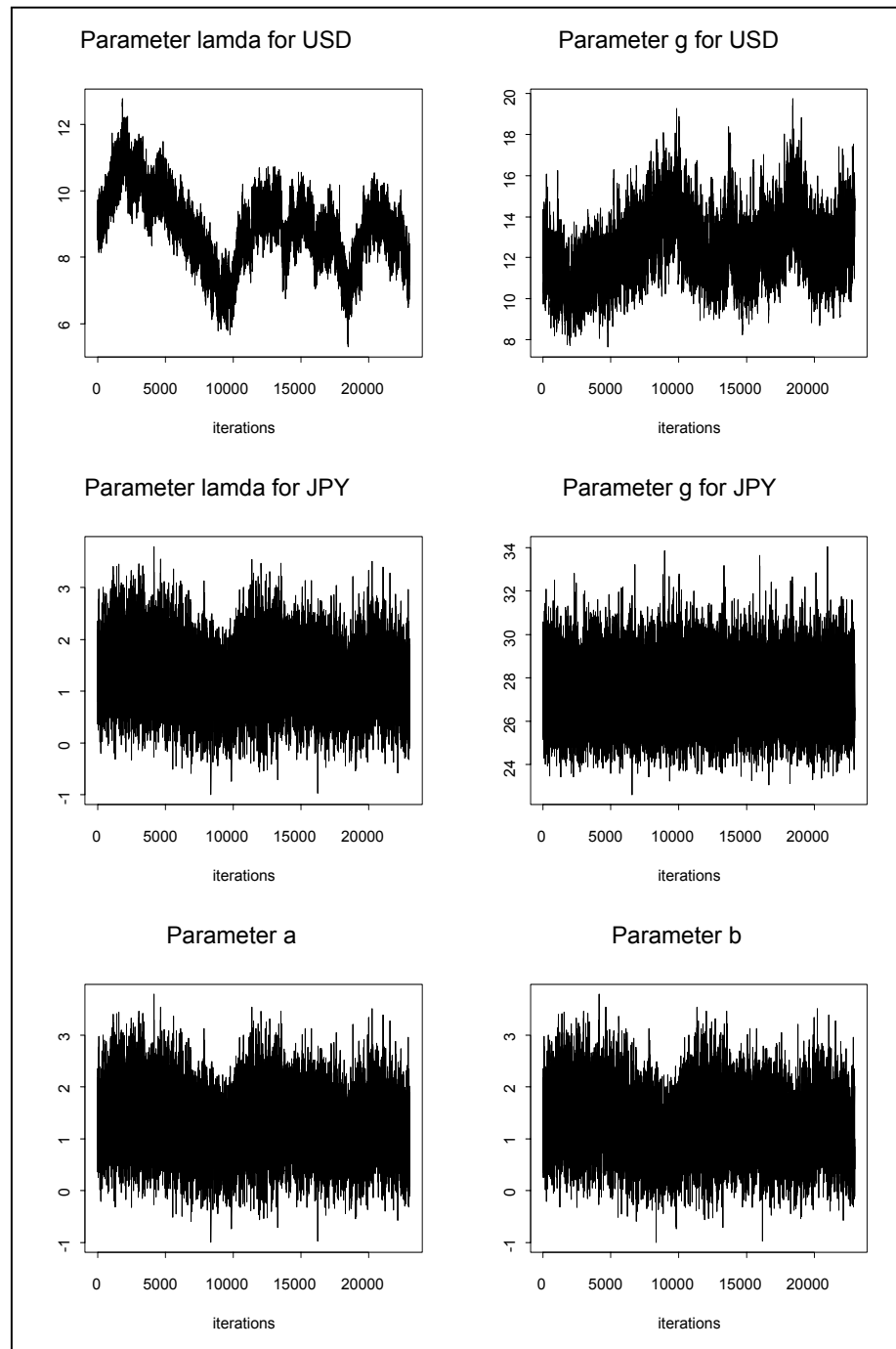


Figure 10-15: MCMC output for parameters of latent factor GARCH model

From the posterior ergodic means for the parameters  $a$  (0.71) and  $b$  (0.26) it is obvious that the GARCH component is highly volatility persistent ( $E(a) + E(b) = 0.97$ ).

## 10.5 Comparison

The same idea of comparing models that has been presented on section 8.2 for the case of univariate time-varying volatility models, has been also applied here for the comparison of the presented multivariate models. In order to choose between two models  $M_1$  and  $M_2$ , one of the procedures suggested by Gelfand *et al.* (1992) is followed. In detail, let  $Z_t$  be the random variable representing future points. Then, the model  $M_1$  ( $M_2$ ) is chosen according to whether  $D > 0$  ( $D < 0$ ), where

$$D = \log \left( \frac{\prod_{t=1}^{T-1} \pi(Z_{t+1} = \mathbf{y}_{t+1} | \Psi_t, \bar{\boldsymbol{\theta}}; M_1)}{\prod_{t=1}^{T-1} \pi(Z_{t+1} = \mathbf{y}_{t+1} | \Psi_t, \bar{\boldsymbol{\theta}}; M_2)} \right),$$

$\mathbf{y}_{t+1}$  is the realization of the stochastic process at time  $t+1$ ,  $\Psi_t$  is the available information up to time  $t$  and  $\bar{\boldsymbol{\theta}}$  is the vector of the posterior means of the parameters of the specific model. Note that  $\exp(D)$  is called pseudo-Bayes factor (Gelfand *et al.*, 1992).

Following Pitt and Shephard (1999a, 1999b),  $\pi(Z_{t+1} = \mathbf{y}_{t+1} | \Psi_t, \bar{\boldsymbol{\theta}}; M_k)$  is estimated via filtering methods. In detail, an estimate of  $\pi(Z_{t+1} = \mathbf{y}_{t+1} | \Psi_t, \bar{\boldsymbol{\theta}}; M_k)$  is given as follows. Initially, samples of size  $K$  (we use  $K = 10000$ ) are obtained of the unobserved component (or components) of the model  $M_k$ , for  $t = 1, \dots, T$ , using  $\mathbf{F}_{t+1}^i \sim \pi(\mathbf{F}_{t+1} | \Psi_t, \bar{\boldsymbol{\theta}}; M_k)$ , for  $i = 1, \dots, K$ . The above density is easily derived for both models (7.8) and (10.12). Then an estimate is given by

$$\hat{\pi}(Z_{t+1} = \mathbf{y}_{t+1} | \Psi_t, \bar{\boldsymbol{\theta}}; M_k) = \frac{1}{K} \sum_{i=1}^K \hat{\pi}(Z_{t+1} = \mathbf{y}_{t+1} | \mathbf{F}_{t+1}^i, \bar{\boldsymbol{\theta}}; M_k),$$

Model	
Latent factor ARCH	-4899.79
Latent factor GARCH	-4818.11
Unobserved ARCH	-5263.43
Stochastic volatility	-4910.13

Table 10.6: Model comparison

which is the result of the Monte Carlo integration of

$$\pi(\mathbf{y}_{t+1}|\Psi_t, \bar{\boldsymbol{\theta}}; M_k) = \int \pi(\mathbf{y}_{t+1}|\mathbf{F}_{t+1}, \bar{\boldsymbol{\theta}}; M_k) \pi(\mathbf{F}_{t+1}|\Psi_t, \bar{\boldsymbol{\theta}}; M_k) d\mathbf{F}_{t+1}.$$

This technique presupposes that the density  $\pi(\mathbf{y}_t|\mathbf{F}_t, \bar{\boldsymbol{\theta}}; M_k)$  can be evaluated and simulated, something which is valid for parameter-driven models that are analyzed in this chapter.

In the Table 10.6 the results for the quantity  $\log \left( \prod_{t=1}^{T-1} \hat{\pi}(Z_{t+1} = \mathbf{y}_{t+1}|\Psi_t, \bar{\boldsymbol{\theta}}; M_1) \right)$  are presented and used for comparison purposes.

Note, that similar comparisons have been conducted by Vrontos, Giakoumatos, Dellaportas and Politis (2001) between the bivariate models. In detail, Vrontos, Giakoumatos, Dellaportas and Politis (2001) compared the unobserved ARCH model, the GARCH model with constant conditional correlations (Bollerslev 1990) and the ARCH model recently introduced by Jeantheau (1998). According to Vrontos, Giakoumatos, Dellaportas and Politis (2001) the unobserved ARCH model seems to be preferable for the one ahead predictions for the analyzed dataset.

In addition, Giakoumatos, Dellaportas and Politis (2004b), compared the multivariate unobserved ARCH, the latent Factor ARCH and the latent factor GARCH model with a dataset of four exchange rates and the results indicated that the latent factor GARCH is preferable than the alternative models.

## 10.6 Technical Details

### 10.6.1 Sampling from truncated Inverse Gamma

Suppose, that a sample point  $x$  is required from the density  $f(x) \propto \frac{1}{x^{a+1}} \cdot \exp\left(-\frac{b}{x}\right) \cdot \mathbf{I}(x < d)$ . Instead of sampling  $x$ ,  $y$  can be sampled from  $f(y) \propto y^{a-1} \cdot \exp(-by) \cdot \mathbf{I}(y > \frac{1}{d})$ , and we set  $x = \frac{1}{y}$ . In order to sample from the last density, one positive latent variable  $m$  is introduced such that  $f(y, m) \propto \exp(-by) \cdot \mathbf{I}(m < y^{a-1}) \cdot \mathbf{I}(y > \frac{1}{d})$ . Then using the Gibbs sampler, sampling points are drawn from the full conditional densities which are of known form:

- $y|m \equiv \exp(-by) \mathbf{I}(y > k)$  where  $k = \max_x \left( \sqrt[a-1]{m}, \frac{1}{d} \right)$ . In order to sample from truncated exponential see Damien *et al* (1999).
- $m|y \sim \mathbf{U}(0, y^{a-1})$

### 10.6.2 Construction of (1-a)100% credible interval for the mean using subsampling

Let  $(X_1, X_2, \dots, X_T)$  be the observed time series. This time series is assumed to be strong mixing, and asymptotically stationary. The basic idea of subsampling is to approximate the sampling distribution of a statistic ( in our case the sample mean  $\bar{x}$ ) based on the values of the same statistic recomputed over smaller subsets of the data that retain the dependence structure of the observations. Subsets of size  $b = \left\lfloor \sqrt{T} \right\rfloor$  are considered, where  $\lfloor \cdot \rfloor$  is the integer part of a number, so that we are led to consider the  $B = N - b + 1$  “blocks” of consecutive observations of the type  $(X_i, X_{i+1}, \dots, X_{i+b-1})$ , for  $i = 1, \dots, B$ .

Let  $\bar{x}^i$  the sample mean of the subset  $(X_i, X_{i+1}, \dots, X_{i+b-1})$  and  $\bar{x}$  the overall mean, then the subsampling estimate of the sampling distribution function  $Dist_{\bar{x}-E(X)}(x) \equiv$

$P(\bar{x} - E(X) \leq x)$  is given by

$$L(x) \equiv \frac{1}{B} \sum_{i=1}^B \mathbf{I} \left( \bar{x}^i \leq x \frac{\sqrt{T}}{\sqrt{b}} + \bar{x} \right)$$

Using the above sampling distribution function we can calculate a valid  $(1 - \alpha)$  100% credible interval for the mean which is given by

$$(\bar{x} - q(1 - a/2), \bar{x} - q(a/2))$$

where  $q(t)$  is the  $t$ -quantile of the subsampling distribution  $L(x)$ , i.e.,  $q(t) = \inf \{x : L(x) \geq t\}$ . Note that, for the parameter  $a$  of the model (10.12) which is constrained to  $(0, 1)$ , a logit transformation has been applied and then it is constructed the credible interval defined in  $(-\infty, +\infty)$  and obtain the required interval bounds by retransforming to the parameter  $a$ . For more details subsampling, see Politis (1998), Politis, Romano and Wolf (1999).

# Chapter 11

## Discussion

This Thesis presents new MCMC algorithms for certain univariate and multivariate time-varying volatility models. Easy to implement and fast to converge MCMC algorithms are presented for the ARCH, GARCH, stochastic volatility, unobserved ARCH, multivariate stochastic volatility, multivariate unobserved ARCH, latent factor ARCH and latent factor GARCH models. All the proposed MCMC algorithms utilize the power of the auxiliary variable sampler and are consisted only by Gibbs steps.

Overall, the auxiliary variable sampler is a very powerful technique for sampling from posterior distribution of univariate and multivariate models. However, in some cases it produces high autocorrelation in the MCMC. In this topic additional research is required. A recent approach has been suggested by Neal (2003), and on these lines further investigation might be worthwhile.

Furthermore, this Thesis exploits the predictive ability of the models in order to construct model comparison tests. We feel that Bayesian model determination methods such as reversible jump (Green, 1995) or marginal likelihood approximations may not well suited for financial time series data where forecasting is of primary importance.

Finally, another promising topic for future research is the application of the MCMC methods in parameter-driven models defined on a continuous time pattern (Hull and White, 1987; Barndorff-Nielsen and Shephard, 2001; Roberts and Stramer, 2001; Eraker,

Johannes and Polson, 2003; Roberts, Papaspiliopoulos and Dellaportas, 2004).

# Bibliography

- [1] **Achcar, J. A. and Smith, A . F. M. (1990).** *Aspects of reparametrization in approximate Bayesian inference, in Bayesian and likelihood methods in statistics and econometrics* (editors: G.Geisser,J.S. Hodges, S.J. Press and A. Zellner). Elsevier science Publishers, 439-452.
- [2] **Adler, S.L. (1981).** Over-relaxation methods for the Monte Carlo Evaluation of the partition function for multiquadratic actions. *Phisical Review*, D, 23, 2901-2904.
- [3] **Aguilar, O., G. Huerta, R. Prado and M. West (1999).** Bayesian Inference on Latent Structure in Time Series. *Bayesian Statistics 6*, pp. 1–16, (editors: J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith). Oxford University Press.
- [4] **Aguilar, O., G. and M. West (2000).** Bayesian Dynamic factor Models and Portofolio Allocation. *Journal of Business and Economic Statistics*, 18, 338-357.
- [5] **Amit, Y., and Grenander, U. (1991).** Comparing sweep strategies for stochastic relaxation. *Journal of Multivariate Analysis*, 37, 197-222.
- [6] **Andersen, T. (1996).** Return volatility and trading volume: an information flow interpretation of stochastic volatility. *Journal of Finance*, 51, 169-204.
- [7] **Andersen, T. G., B. E. Sorensen (1996).** GMM estimation of a stochastic volatility model. a Monte Carlo study. *Journal of Business and Economic statistics*, 14, 328-352.

- [8] **Andersen, T. G., H. J. Chung and B. E. Sorensen (1999).** Efficient method of moments estimation of a stochastic volatility model: a Monte Carlo study. *Journal of Econometrics*, 91, 61-87.
- [9] **Baillie, R.T. and Bollerslev, T. (1989).** The message in daily exchange rates. A conditional variance tale. *Journal of business and Economic Statistics*, 7, 295-305.
- [10] **Baillie, R.T. and Bollerslev, T. (1990).** A multivariate generalized ARCH approach to modeling risk premia in forward foreign exchange rate markets. *Journal of International Money and Finance*, 9, 309-324.
- [11] **Barndorff-Nielsen, O.E. and N. Shephard (2001).** Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics (with discussion). *Journal of the Royal Statistical Society, Series B*, 63, 167-241.
- [12] **Barone, P. and A. Frigessi, (1990).** Improving Stochastic Relaxation for Gaussian Random Fields. *Probability in the Engineering and Informational Sciences*, vol. 4, 369-381.
- [13] **Bauwens, L. and Lubrano, M. (1998).** Bayesian inference on GARCH models using the Gibbs sampler. *CORE discussion paper*, Universite Catholique de Louvain.
- [14] **Bayes, Rev. T., (1763).** An Essay Toward Solving a Problem in the Doctrine of Chances. *Philos. Trans. R. Soc. London*, 53, pp. 370-418; (1958) reprinted in *Biometrika*, 45, pp. 293-315.
- [15] **Bekaert, G. and Hodrick, R.J. (1993).** On Biases in the Measurement of Foreign Exchange Risk premiums. *Journal of International Money and Finance*, 12, 115-138.
- [16] **Bera, A.K. and Higgins, M.L (1993).** ARCH models: properties, estimation and testing. *Journal of Economics Survey*, 7, 305-366.

- [17] **Bernardo, J. M. and A. F. M. Smith (1994).** *Bayesian Theory*. John Wiley & Sons.
- [18] **Besag, J. (1974).** Spatial interaction and the statistical analysis of lattice systems (with discussions). *Journal of the Royal Statistical Society, Series B*, 36, 192–236.
- [19] **Besag J. E. and P. J. Green (1993).** Spatial statistics and Bayesian computation (with discussion). *Journal of Royal Statistical Society, B* 48, 259-302.
- [20] **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.
- [21] **Bollerslev T. (1986).** Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 51, 307-327.
- [22] **Bollerslev, T. (1987).** A conditional heteroskedastic time series model for speculative prices and rates of returns. *Review of economics and Statistics*, 69, 542-547.
- [23] **Bollerslev T. (1990).** Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *The review of Economics and Statistics*, 43, 498-505.
- [24] **Bollerslev T, Engle R, Wooldridge J.M. (1988).** A Capital Asset Pricing Model with Time-Varying Covariances. *Journal of Political Economy*, 96, 116-131.
- [25] **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, 5-59.
- [26] **Bollerslev T., R. F. Engle and D. B. Nelson (1994).** ARCH Models. In *Handbook Of Econometrics*, Volume IV, 2961-3038. Eds: R. F. Engle and D. L. McFadden.

- [27] **Bos, C.S., Mahieu, R.J. and van Dijk, H.K. (1999).** Daily Exchange Rate behavior and Hedging of Currency Risk. Econometric Institute Report Ei-9936/A, Erasmus University Rotterdam.
- [28] **Brockwell, P.J. and Davis, R.A. (1991).** *Time Series: Theory and Models*, 2nd edition, Springer-Verlag, New York.
- [29] **Brooks, S.P. and A. Gelman (1998).** General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, 7, 434-455..
- [30] **Brooks, S. P. and G. O. Roberts (1999).** Assessing Convergence of Markov Chain Monte Carlo Algorithms. working paper. *Statistics and Computing*, 8, 319-335.
- [31] **Carlstein, E. (1986).** The use of subseries values for estimating the variance of a general statistic from a stationary time series. *Annals of Statistics*, 14, 1171–1179.
- [32] **Casella, C. and E. I. George (1992),** Explain the Gibbs Sampler. *The American Statistician*, 46, 167-174.
- [33] **Chib, S., and Greenberg, E. (1994).** Bayes inference in regression models with ARMA(p,q) errors. *Journal of Econometrics*, 64, 183-206.
- [34] **Chib S. and E. Greenberg (1995a).** Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, 19, 327-335.
- [35] **Chib, S., and Greenberg, E. (1995b).** Hierarchical analysis of SUR models with extensions to correlated errors and time-varying parameter models. *Journal of Econometrics*, 68, 339-360.
- [36] **Chib S. and B. P. Carlin (1999).** On MCMC Sampling in Hierarchical longitudinal Models. *Statistics and Computing*, 9, 17-26.

- [37] **Chib, S., Nardari F. and N. Shephard, (2002).** Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics*, 108, 281–316.
- [38] **Cox, D.R. (1981).** Statistical Analysis of Time Series: Some Recent Developments. *Scandinavian Journal of Statistics*, 8, 93-115.
- [39] **Cowles, M.K. and B.P. Carlin (1996).** Markov Chain Monte Carlo Convergence Diagnostics: A comparative Review. *Journal of the American Statistical Association*, 91, 883-904.
- [40] **Damien P. and S. Walker (1996).** Sampling Probability Densities via Uniform Random Variables and a Gibbs sample. Technical report, Imperial College.
- [41] **Damien, P., Wakefield, J., and Walker, S. (1999).** Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables. *Journal of Royal Society*, series B, 61, 331 - 344.
- [42] **Davidian, M. and Carroll, R.J. (1987).** Variances function estimation. *Journal of the American Statistical Association*, 82, 1079-1091.
- [43] **Day, T.E. and Lewis, C.M. (1992).** Stock market volatility and the information content of stock index options. *Journal of Econometrics*, 52, 267-287.
- [44] **De Jong, P. and N. Shephard (1995).** The simulation smoother for time series models. *Biometrika*, 82, 339-350.
- [45] **Dellaportas, P. (1995).** Random variate Transformations in the Gibbs Sampler: Issues of efficiency and Convergence. *Statistics and Computing*, 5, 133-140.
- [46] **Dellaportas, P. (1996).** Computational strategies for the implementation of the Bayesian paradigm. Hermis 1996: 3rd Hellenic European Conference on Mathematics and Informatics.

- [47] **Dellaportas, P. and Roberts, G.O. (2003).** An introduction to MCMC. In *Spatial Statistics and Computational methods*, (eds: J. Muller), 1-42, Springer-Verlag, NY.
- [48] **Dellaportas, P. and Smith, A.F.M. (1993).** Bayesian Inference for Generalized Linear and Proportional hazards Models via Gibbs Sampling. *Applied Statistics*, 42, 443-459.
- [49] **Demos A. and E. Sentana (1996).** An EM Algorithm for Conditional Heteroskedastic Factor Models. *Journal of Business and Economics Statistics*, 16, 357-361.
- [50] **Diebold F. X. and M. Nerlove (1989).** The Dynamics of Exchange Rate Volatility: A Multivariate Latent Factor Arch Model. *Journal of Applied Econometrics*, 4, 1-21.
- [51] **Diebold, F.X. and Lopez, J.A. (1995).** *ARCH models. Macroeconomics: Developments, Tensions and Prospects* (K. Hoover, ed).
- [52] **Doukhan, P. (1994).** *Mixing*. Lecture Notes in Statistics No 85. Springer-Verlag: New York.
- [53] **Duffie, D. and K.J. Singleton (1993).** Simulated moments estimation of Markov models of asset process. *Econometrica*, 61, 929-952.
- [54] **Edwards R. G. and A. D. Sokal (1988).** Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo Algorithm. *Phys. Rev. Lett.*, 38, 2009-2012.
- [55] **Engle R. F. (1982).** Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation. *Econometrica*, 50, 987-1008.
- [56] **Engle, R.F. (1995).** ARCH selected readings. New York: Oxford University Press.

- [57] **Engle, R.F. and Ng, V.K. (1993).** Measuring and Testing the Impact of News on Volatility. *The Journal of Finance*, 48, 1749-1778.
- [58] **Engle, R.F., Ng VK, Rothschild M. (1990).** Asset Pricing with a Factor-ARCH Covariance Structure: Empirical Estimates for Treasury Bills. *Journal of Econometrics*. 45, 213-237.
- [59] **Engle, R.F., Lilien, D.M. and Robins, R.P. (1987).** Estimating time varying risk premia in the term structure: The ARCH-M model. *Econometrica*, 55, 391-407.
- [60] **Eraker, B., M. Johannes and N.G. Polson (2003).** The impact of jumps in returns and volatility. *Journal of Finance*, 53, 1269-1300.
- [61] **Fiorentini, G., Calzolari, G. and Panattoni, L. (1996).** Analytic Derivatives and the computation of GARCH Estimates. *Journal of Applied Econometrics*, 11, 399-417.
- [62] **Fiorentini, G., Sentana, E. and N. Shephard (2004).** Likelihood-based estimation of latent generalised ARCH structures. *Econometrica*, forthcoming.
- [63] **Fishman, G. S. (1999).** An analysis of Swendsen-Wang and related sampling methods. *Journal of Royal Statistical Society*, B, 61, 623-641.
- [64] **Fruhwirth-Schnatter, S. (1994).** Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15, 183-202.
- [65] **Fukuchi, J. (1997).** Subsampling and model selection in time series analysis. Technical report 97-2, Statistical Research Group, Hiroshima University. Japan.
- [66] **Gelfand, A.E., and Smith, A.F.M. (1990).** Sampling - Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, 85, 398-409.

- [67] **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.
- [68] **Gelfand, E. A., Smith, A.F.M., and Lee, T. (1992).** Bayesian Analysis of Constrained parameter and Truncated data Problems using Gibbs Sampling. *Journal of the American Statistical Association*, 87, 523-532.
- [69] **Gelfand, A.E., Dey and H.Chang (1992).** Model determination using predictive distributions with implementation via sampling based methods. *Bayesian Statistics 4* (Bernardo *et al.*, editors), 147-167.
- [70] **Gelman, A. (1996).** Inference and Monitoring Convergence, in *Markov Chain Monte Carlo in Practice*. eds: W.R. Gilks, S. Richardson and D.J. Spiegelhalter, page 131-143, Chapman and Hall.
- [71] **Gelman, A. and D. Rubin (1992).** Inference from Iterative Simulation using Multiple Sequences. *Statistical Science* 7, 457-511.
- [72] **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.
- [73] **George, E. I. and R. E. McCulloch (1993).** Variable Selection Via Gibbs Sampling. *Journal of the American Statistical Association*, 88, 881-889.
- [74] **Geweke, J. (1988).** Antithetic acceleration of Monte Carlo integration in Bayesian inference. *Journal of Econometrics*, 38, 73-89.
- [75] **Geweke, J.F. (1989a).** Exact Predictive Densities in Linear Models with ARCH Disturbances. *Journal of Econometrics*, 40, 63-86.
- [76] **Geweke, J.F. (1989b).** Bayesian Inference in Econometric Models Using Monte Carlo Integration. *Econometrica*, 57, 1317-1339.

- [77] **Geweke J. (1992).** Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In *Bayesian Statistics 4*, Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds). Oxford University Press, 169-193.
- [78] **Geyer, C. J. and Thompson (1995).** Annealing Markov Chain Monte Carlo with Applications to Ancestral Inference. *Journal of the American Statistical Association*, 90, 909-920.
- [79] **Ghysels, E., Harvey, A. C. and Renault, E. (1996).** Stochastic Volatility. In Rao, C. R., Maddala, G. S. (Eds), *Statistical Method in Finance*, pp 119-191, North-Holland, Amsterdam.
- [80] **Giakoumatos, S.G. (1997).** *Stochastic Volatility Models: A Bayesian Approach*. M.Sc. Thesis. Department of Statistics, Athens University Of Economics and Business.
- [81] **Giakoumatos, S. G., Vrontos, I. D., Dellaportas, P., and Politis, D. N. (1999).** An MCMC Convergence Diagnostic using Subsampling. *Journal of Computational and Graphical Statistics*, volume 8, number 3, 431-451.
- [82] **Giakoumatos S. G., Dellaportas, P. and D. N. Politis (2004a).** Bayesian Analysis of the Unobserved ARCH model. Submitted.
- [83] **Giakoumatos S. G., Dellaportas, P. and D. N. Politis (2004b).** Bayesian analysis of some Multivariate time varying volatility models. Submitted.
- [84] **Gilks W.R., and P. Wild (1992).** Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, 41, 337-348.
- [85] **Gilks W.R., Richardson S. and D. J. Spiegelhalter (1995).** *Markov Chain Monte Carlo in Practice*. CRC Press, ISBN: 0412055511, (December 1, 1995)
- [86] **Giovannini, A. and Jorion, P. (1989).** The time variation of risk and return in the foreign exchange and stock markets. *The Journal of Finance*, 44, 307-325.

- [87] **Glosten, L.R., Jagannathan, R. and Runkle, D.E. (1993).** On the Relation between the Expected Value and the Volatility of the Nominal Excess returns on Stocks. *The Journal of Finance*, 48, 1779-1801.
- [88] **Gourieroux C. (1997).** *ARCH Models and Financial Applications*. Springer-Verlag New York. ISBN 0-387-94876-7.
- [89] **Gourieroux C., and A. Monfort (1992).** Qualitative threshold ARCH models. **Journal of Econometrics**, 52, 159-199.
- [90] **Gourieroux C., A. Monfort and A. Tragnon (1984).** Pseudo Maximum Likelihood Methods: Theory. *Econometrica*, 52, 681-700.
- [91] **Green, P. J. (1995).** Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82, 711-732.
- [92] **Hafner, C.M. (1998).** *Nonlinear Time Series Analysis with Application to foreign Exchange rate Volatility*. Heidelberg: Physica-Verlag.
- [93] **Hall, P., Horowitz, J.L. and Jing, B. (1996).** On blocking rules for the bootstrap with dependent data. *Biometrika*, 50, 561-574.
- [94] **Hamilton, J. (1994).** *Time Series Analysis*. Princeton University Press, Princeton.
- [95] **Harvey A., E. Ruiz and E. Sentana (1992).** Unobserved component time series models with ARCH disturbances. *Journal of Econometrics*, 52, 129-157.
- [96] **Harvey A., E. Ruiz and N. Shephard (1994).** Multivariate Stochastic variance Models. *Review of Economic Studies*, 61, 247-264.
- [97] **Hastings, W. K. (1970).** Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57, 97-109.

- [98] **He, C. and Terasvirta, T. (1999).** Properties of moments of a family of GARCH processes. *Journal of Econometrics*, 92, 173-192.
- [99] **Heidelberger, P. and P. D. Welch (1983).** Simulation Run Length Control in the Presence of an Initial Transient. *Operations Research*, 31, 1109-1144.
- [100] **Higgins, M.L. and Bera, A.K. (1992).** A Class of Nonlinear ARCH models. *International Economic Review*, 33, 137-158.
- [101] **Higdon, D.M. (1993).** Comment on *Spatial statistics and Bayesian computation* by J. Besag and P. Green, *Journal of the Royal Statistical Society, Series B*, 55-78.
- [102] **Higdon, D. M. (1998).** Auxiliary variable methods for Markov chain Monte Carlo with applications. *Journal of the American Statistical Association*, 93, 585-595.
- [103] **Hills, S.E and A.F.M Smith (1992).** Parametrization issues in Bayesian inference. *Bayesian Statistics 4* (Proceedings of the Third Valencia International Meeting on Bayesian Statistics).
- [104] **Hull, J. and A. White (1987).** The pricing of options on assets with stochastic volatilities. *Journal of Finance*, 42, 281-300.
- [105] **Jacquier, E., N. G. Polson and E. Rossi (1994).** Bayesian Analysis of Stochastic Volatility Models. *Journal of Business & Economic Statistics*, 12, 371-417.
- [106] **Jacquier, E., Polson N.G. and Rossi, P. (1999).** Models and priors for multivariate stochastic volatility. Discussion Paper, Graduate School of Business, University of Chicago.
- [107] **Jeantheau, T. (1998).** Strong Consistency of estimators for multivariate ARCH models. *Econometric Theory*, 14, 70-86.
- [108] **Jeffreys, H. (1939).** *Theory of Probability*. Oxford University Press, NY.

- [109] **Johnson, R.A. and D.W. Wichern (1999).** *Applied Multivariate Statistical Analysis*. Prentice Hall, New Jersey.
- [110] **Kim, S, Shephard, N., and Chib S. (1998).** Stochastic Volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies*, 65, 361-393.
- [111] **King M., E. Sentana and S. Wadhwani (1994).** Volatility and Links between National Stock Markets. *Econometrica*, 62, 901-933.
- [112] **Kraft, D. F. and R. F. Engle (1982).** Autoregressive Conditional Heteroscedasticity in Multiple Time Series. Discussion paper 82-23, Department of Economics, University of California, San Diego, USA.
- [113] **Kroner, K.F. and Claessens, S. (1991).** Optimal dynamic hedging portfolios and currency composition of external debt. *Journal of International Money and Finance*, 10, 131-148.
- [114] **Kroner, K.F. and Sultan, J. (1991).** Exchange rate volatility and time varying hedge ratios. *Pacific-Basin Capital Markets Research* 11.
- [115] **Künsch, H.R. (1989).** The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17, 1217–1241.
- [116] **Kuwahara, H. and Marsh, T.A. (1992).** The Pricing of Japanese Equity Warrants. *Management Science*, 38, 1610-1641.
- [117] **Laux, P.A. and Ng, L.K. (1993).** The sources of GARCH: empirical evidence from an intraday returns model incorporating systematic and unique risks. *Journal of International Money and Finance*, 12, 543-560.
- [118] **Lee, S.W. and Hansen, B.E. (1994).** Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory*, 10, 29-52.

- [119] **Liu, J., W.H. Wong and A. Kong (1994).** Covariance structure of the Gibbs sampler with application to the comparison of estimators and augmentation schemes. *Biometrika*, 81, 27-40.
- [120] **MacEachern, S.N. and L.M. Berliner (1994).** Subsampling the Gibbs Sampler. *The American Statistician*, 48, 188-190.
- [121] **Mahieu, R., Schotman, P. (1998).** An empirical application of stochastic volatility models. *Journal of Applied Econometrics*, 13 (4), 333-360.
- [122] **Mandelbrot, B. (1963).** The variation of certain speculative prices. *Journal of Business*, 36, 349-419.
- [123] **Marinari, E. and Parisi, G. (1992).** Simulated Tempering: a new Monte Carlo Scheme. *Europhysics Letter*, 19, 451-458.
- [124] **Melino, A. and S.M. Turnbull (1990).** Pricing foreign currency options with stochastic volatility. *Journal of Econometrics*, 45, 239-265.
- [125] **Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller A. H., and E. Teller (1953).** Equations of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, 21, 1087-1092.
- [126] **Meyn, S. P. and Tweedie, R. L. (1993).** *Markov Chains and Stochastic Stability*. London: Springer-Verlag.
- [127] **Mira, A. and Tierney, L. (1998).** On the use of the auxiliary variables in markov chain Monte carlo sampling. *Scandinavian Journal of Statistics*, To appear
- [128] **Mira, A. (1999).** On Metropolis-Hastings algorithm with delayed rejection. Technical Report, Universita dell'Insubria, Varese, Italy.
- [129] **Muller, P. and Pole, A. (1999).** Monte Carlo Posterior Integration in GARCH models. *Sankhya*, 62.

- [130] **Neal, R. M. (2003).** Slice sampling (with discussion). *Annals of Statistics*, vol. 31, pp. 705-767.
- [131] **Nelson, D.B. (1988).** Time Series Behavior of Stock Market Volatility and Returns. Ph.D. Dissertation, MIT.
- [132] **Nelson, D.B. (1991).** Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 59, 347-370.
- [133] **Ng L. (1991).** Tests of the CAPM with Time-Varying Covariances: A Multivariate GARCH Approach. *Journal of Finance*, 46, 1507-1521.
- [134] **O'Hagan, A (1994).** *Kendall's Advanced Theory of Statistics: Bayesian Inference*. University Press, Cambridge.
- [135] **Pantula, S.G. (1986).** Modelling the persistence of conditional variances; a comment. *Econometric Review*, 5, 71-74.
- [136] **Pitt, M.K., (1997).** Bayesian methods for non-Gaussian state space models. D. Phil. (supervised by Neil Shephard), University of Oxford.
- [137] **Pitt M.K., Shephard N. (1999a).** Time varying covariances: a factor stochastic volatility approach (with discussion). In *Bayesian Statistics 6*, Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds). Oxford University Press, 547-570.
- [138] **Pitt M.K., Shephard N. (1999b).** Filtering via simulation:auxiliary particle filter. *The Journal of the American Statistical Association*, 94, 590-599.
- [139] **Pitt M.K., Shephard N. (1999c).** Analytic convergence rates and parameterisation issues for the Gibbs sampler applied to state space models. *Journal of Time Series Analysis*, 20, 63-85.
- [140] **Polasek, W. and Muller, P. (1995).** Gibbs Sampling for ARCH models in Finance. Proceedings of MODA 4 (C.P. Kitsos and W.G. Muller, eds) Heidelberg, Physica Verlag.

- [141] **Pole, A., West, M. and J. Harrison. (1994).** *Applied Bayesian forecasting and time-series analysis*. Chapman & Hall, New York.
- [142] **Politis, D.N. (1998).** Computer-Intensive Methods in Statistical Analysis. *IEEE Signal Proc. Magazine*, Vol. 15, No. 1, pp. 39-55.
- [143] **Politis, D.N. and Romano, J.P. (1993).** On the sample variance of linear statistics derived from mixing sequences. *Stochastic Processes and their Applications*, 45, 155-167.
- [144] **Politis, D.N. and Romano, J.P. (1994).** Large sample confidence regions based on subsamples under minimal assumptions. *Annals of Statistics*, 22, 2031–2050.
- [145] **Politis, D.N., Romano, J.P., and Wolf, M. (1997).** Subsampling for heteroskedastic time series. *Journal of Econometrics*, 81, 281-318.
- [146] **Politis, D.N., J. P. Romano and M. Wolf (1999).** *Subsampling*. Springer Verlag, New York.
- [147] **Polson, N. G. (1996).** Convergence of the MCMC algorithms, in *Bayesian Statistics 5*, Bernardo, Berger, Dawid, Smith (eds.), 297-322, Oxford University Press.
- [148] **Poon, S. and Taylor, S.J. (1992).** Stock returns and volatility: an empirical study of the UK stock market. *Journal of Banking and Finance*, 16, 37-59.
- [149] **Raftery, A. and S. Lewis (1992).** How Many Iterations in the Gibbs Sampler?. In *Bayesian Statistics 4*, Oxford University Press.
- [150] **Robert C. P. (1995).** Simulation of truncated normal variables. *Statistics and Computing*, 5, 121-125.
- [151] **Robert, C.P. and K.L. Mengersen (1999).** MCMC convergence Diagnostics: A “Reviewww”. In *Bayesian Statistics 6* (eds: Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds)). Oxford University Press.

- [152] **Roberts G. O. and J. S. Rosenthal (1999).** Convergence of slice sampler Markov chains. *Journal of Royal Statistical Society*, B, 61, 643-660.
- [153] **Roberts, G.O. and S. K. Sahu (1997).** Updating Schemes, Correlation Structure, Blocking and Parametrization for the Gibbs Sampler. *Journal of Royal Statistical Society*, B, 291-317.
- [154] **Roberts, G.O. and O. Stramer (2001).** On inference for nonlinear diffusion models using the Hastings-Metropolis algorithms. *Biometrika*, 88, Issue 3, 603-621.
- [155] **Roberts, G.O., Papaspiliopoulos, O. and P. Dellaportas (2004).** Bayesian inference for non-Gaussian Ornstein-Uhlenbeck stochastic volatility processes. *Journal of the Royal Statistical Society*, Series B, 66, 369-393.
- [156] **Rosenblatt, M. (1956).** A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci*, 42, 43-47.
- [157] **Rosenthal, J. S. (1993).** Rates of Convergence for Data Augmentation on the Finite Sample Space. *Annals of Applied Probability*, 3, 819-839.
- [158] **Rothenberg, T.J. (1973).** *Efficient Estimation with a priori Information*. Yale University Press, New Haven.
- [159] **Ruiz, E. (1994).** Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics*, 63, 289-306.
- [160] **Schervish, M. J. and B. P. Carlin (1992).** On the Convergence of Successive Substitution Sampling. *Journal of Computational and Graphical Statistics*, 1, 111-127.
- [161] **Schwert, G.W. and Sequin, P.J. (1990).** Heteroskedasticity in Stock Returns. *Journal of Finance*, 45, 1129-1155.

- [162] **Shephard, N. (1993).** Filtering non-linear time series models with application to stochastic variances models. *Journal of Applied Econometrics*, 8, S135-S152.
- [163] **Shephard N. (1996).** Statistical aspects of ARCH and Stochastic Volatility. In *Time Series Models In Econometrics, finance and other fields*. Eds: Cox D. R., D. V. Hinkley and O. E. Barndorff-Nielsen.
- [164] **Shephard, N., and M. K. Pitt (1997).** Likelihood analysis of non-Gaussian measurement time series, *Biometrika*, 84, 653-667.
- [165] **Smith A.F.M. (1991).** Bayesian computational methods. *Phil. Trans. Roy. Soc. London A*, 337, 369-386.
- [166] **Smith A.F.M. and A.E. Gelfand (1992).** Bayesian Statistics without tears: A sampling - resampling perspective. *The American Statistician*, 46, 84-88.
- [167] **Smith A.F.M. and G.O. Roberts (1993).** Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society*, B, 55, 3-24.
- [168] **Spiegelhalter, D.J., Thomas, A., Best, N. and W.R. Gilks (1995a).** BUGS: Bayesian Inference Using Gibbs Sampling, Version 0.50. Technical Report, Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge University.
- [169] **Spiegelhalter, D.J., Thomas, A., Best, N. and W.R. Gilks (1995b).** BUGS: Examples, Version 0.50. Technical Report, Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge University.
- [170] **Stewart, L. (1983).** Bayesian analysis using Monte Carlo integration, a powerful methodology for handling some difficult problems. *The Statistician*, 32, 195-200.

- [171] **Stewart, L. (1987).** Hierarchical bayesian analysis using Monte Carlo Integration: computing posterior distributions when there are many models. *The Statistician*, 36, 211-219.
- [172] **Stewart, L. and Davis, W.W. (1986).** Bayesian posterior distributions over sets of possible models with inferences computed by Monte carlo integration. **The Statistician**, 35, 175-182.
- [173] **Swendsen R. H. and J. S. Wang (1987).** Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.*, 58, 86-88.
- [174] **Tauchen, G. and M. Pitts (1983).** The price variability–volume relationship on speculative markets. *Econometrica* 51, 485–505.
- [175] **Taylor S. J. (1982).** Financial returns modelled by the product of two stochastic processes - a study of the daily sugar prices 1961-75. In *Time Series Analysis Theory and Practice*, Anderson OD (eds). North-Holland: Amsterdam.
- [176] **Taylor, S. J. (1986).** Modelling Financial Time Series. John Wiley, Chrichester.
- [177] **Taylor, S.J. (1994).** Modeling stochastic volatility: A review and comparative study. *Mathematical Finance*, 4, 183-204.
- [178] **Tierney, L. (1994).** Markov chain for exploring posterior distributions (with discussion). *The Annals of Statistics*, 22, 1701-1762.
- [179] **Tierney, L. and Mira, A. (1999).** Some adaptive Monte Carlo methods for Bayesian inference. *Statistics in Medicine*, 18, 2507-2515.
- [180] **Turtle, H., Buse, A. and Korkie, B. (1994).** Test of Conditional Asset Pricing with Time-Varying Momments and Risk Prices. *Journal of Financial and Quantitative Analysis*, 29, 15-29.

- [181] **Vrontos, I.D. (1997).** *Bayesian Autoregressive Conditional Heteroscedasticity Models*. MSc Thesis. Department of Statistics. Athens University of Economics and Business.
- [182] **Vrontos, I. D. (2001).** *MCMC Applications in Time-Varying Volatility Models*. Ph.D. Thesis. Athens University of Economics and Business.
- [183] **Vrontos, I.D., Dellaportas, P. and Politis, D.N. (2000).** Full Bayesian Inference for GARCH and EGARCH Models. *Journal of Business and Economic Statistics*, 18, 187-198.
- [184] **Vrontos, I.D., Giakoumatos, S.G., Dellaportas, P., and Politis D.N. (2001).** An application of three bivariate time varying volatility models. *Applied stochastic models in business and industry*, 17, 121-133
- [185] **Wild, P. and Gilks, W.R. (1993).** S 287: Adaptive rejection sampling from log-concave density functions. *Applied Statistics*, 42, 701-709.
- [186] **White, H. (1982).** maximum likelihood estimation of misspecified models. *Econometrica*, 50, 1-25.
- [187] **Yue, H., and Chan, K.S. (1996).** Asymptotic efficiency of the sample mean in Markov Chain Monte Carlo schemes. *Journal of the Royal Statistical Society, B*, 58, 525-539.