$\mathcal{I}^2\mathcal{SDS}$

The Institute for Integrating Statistics in Decision Sciences

Technical Report TR-2008-9

June 15, 2008

Bayesian Analysis of Multivariate GARCH Models

Refik Soyer

Department of Decision Sciences

The George Washington University, USA

Kadir Tanyeri

Department of Decision Sciences

The George Washington University, USA

BAYESIAN ANALYSIS OF MULTIVARIATE GARCH MODELS

Refik Soyer
Department of Decision Sciences
The George Washington University
Washington, DC, U.S.A.

Kadir Tanyeri
Department of Decision Sciences
The George Washington University
Washington, DC, U.S.A.

Abstract

We present Bayesian analysis of the multivariate autoregressive conditional heteroskedastic (ARCH) and generalized ARCH (GARCH) models as a class of deterministic volatility models. In so doing, we develop a Bayesian inference procedure by extending the Markov chain Monte Carlo method of Muller and Pole (1998) introduced for univariate models. Our approach uses a multivariate Bayesian regression setup in implementation of the Markov chain Monte Carlo.

1. Introduction and Preliminaries

Modeling conditional variances has been interest of many researchers during the last several decades. Many time-series models focus on modeling conditional means, that is, $E(Y_t|Y_{t-1},\dots)$, but they assume that the conditional variance is not a function of the past, implying that by learning from past, we do not learn about the variances. However, empirical evidence suggests that the conditional variances do change in economic and financial time-series. The models that focus on modeling conditional variances are referred to as the conditional heteroskedastic models or stochastic volatility models. The term volatility is used to refer to the conditional variance of a commodity return. In what follows, we first introduce the univariate models and then extend them to multivariate stochastic volatility models. In our development we will focus on commodity returns.

Let r_t denote return on a commodity at time t and we assume that returns have constant means and write

$$r_t = \mu + \xi_t, \tag{1}$$

where ξ_t is a zero mean error term whose probabilistic structure will be discussed in detail. Note that in the above if the mean of the process depends on the past history, then we can replace μ by μ_{t} where $\mu_{\mathrm{t}} = E(r_t|r_{t-1}, \dots)$ is the conditional mean of the commodity returns. For example, if r_t follows a first-order autoregressive process then $\mu_{\mathrm{t}} = C + \phi r_{t-1}$. The common approach in time-series modeling is to assume that the error sequence $\{\xi_t\}$ is an uncorrelated constant variance sequence. In other words, the conditional variance or the volatility of the process is constant, that is, $V(r_t|r_{t-1}, \dots) = \sigma_\xi^2$.

The conditional heteroskedastic variance modeling involves modeling the conditional variance $h_t = V(r_t|r_{t-1}, \dots)$. In other words, it is concerned with the evolution of h_t over time. Note that unlike the time-series r_t the volatility is <u>not</u> observable. Some of these models use an exact function to describe the behavior of h_t and most commonly known models of this type are the autoregressive conditional heteroskedastic (ARCH) and generalized ARCH (GARCH) models. Others use a stochastic equation to describe the behavior of h_t and thus are referred to as stochastic volatility models; see Jacquier et. al (1994) and Harvey et. al (1994). Our focus will be on the ARCH/GARCH models.

The ARCH model is introduced by Engle (1982) and is generalized by Bollerslev (1986) to GARCH models. This class of models assume that the conditional variance is not constant and work with the mean corrected time-series ξ_t. As common to many time-series models it is assumed that ξ_t's are nonautocorrelated but they are not independent. In describing the dependence structure of ξ_t's, one strategy is to model the squared ξ_t's as an AR(q) process using the square of its past values, that is,

$$\xi_{\mathrm{t}}^2 = \alpha_0 + \alpha_1 \xi_{\mathrm{t-1}}^2 + \cdots + \alpha_{\mathrm{q}}\, \xi_{\mathrm{t-q}}^2 + \epsilon_{\mathrm{t}} \tag{2}$$

where ϵ_{t} is a white-noise term. Note that the above implies that ξ_{t}^2 terms are autocorrelated and thus ξ_t's are dependent (even though ξ_t's are unautocorrelated). We can see the presence of such dependence by looking at the autocorrelation function of the ξ_{t}^2 series. Model given by (2) implies a nonlinear dependence structure and thus is not convenient for estimation purposes. Using the similar idea for motivation Engle (1982) suggested the following ARCH(q) model for conditional variance

$$\xi_{\mathrm{t}}^2 = \epsilon_{\mathrm{t}} \sqrt{h_t} \tag{3}$$

$$h_t = \alpha_0 + \alpha_1 \xi_{\mathrm{t-1}}^2 + \cdots + \alpha_{\mathrm{q}} \xi_{\mathrm{t-q}}^2, \tag{4}$$

where ϵ_{t}'s independent and identically distributed random variables with mean 0 and variance 1. There are certain conditions on α_{i}'s that need to be satisfied for $h_t > 0$, that is, $\alpha_{\mathrm{i}} > 0$ for $\mathrm{i} = 0, \ldots, \mathrm{q}$. Note that in the above, since $\xi_t = r_t - \mu_{\mathrm{t}}$, the conditional variance is a function of the past values of r_t series as well. Thus, given the past history of the series, the volatility equation is an exact function for h_t. In the above the mean corrected process ξ_t is stationary.

In the special case $\mathrm{q} = 1$, we have the ARCH(1) model is given by

$$h_t = \alpha_0 + \alpha_1 \xi_{\mathrm{t-1}}^2,$$

where α_0, $\alpha_1 > 0$. We can easily show that

$$E(\xi_t) = E(\xi_t | D_{t-1}) = 0$$

where $D_{t-1} = (r_{t-1}, r_{t-2}, \dots)$ and thus $V(\xi_t) = E(\xi_t^2)$. Due to the stationarity of ξ_t

$$E(\xi_t^2) = \frac{\alpha_0}{1 - \alpha_1} = V(\xi_t). \tag{5}$$

and thus $\alpha_0 > 0$ and $0 < \alpha_1 < 1$. If we need higher order moments to exist then we need to impose other restrictions. It is important to note that in the above

$$E(\xi_t^2 | D_{t-1}) \neq E(\xi_t^2).$$

Similarly, in the q-th order stationary ARCH process (3)-(4), we need $\alpha_i > 0$ for $i = 0, \dots, q$ and

$$E(\xi_t^2) = \frac{\alpha_0}{1 - \alpha_1 - \alpha_2 - \cdots - \alpha_q}. \tag{6}$$

Thus, for the variance to exist we also need $\sum_{i=1}^{q} \alpha_i < 1$.

A generalization of the ARCH model was suggested by Bollerslev (1986) by generalizing the volatility equation to include the past history of h_t's. This resulted in the generalized ARCH(q, p) model

$$\xi_t = \epsilon_t \sqrt{h_t}$$

$$h_t = \alpha_0 + \alpha_1 \xi_{t-1}^2 + \cdots + \alpha_q \xi_{t-q}^2 + \beta_1 h_{t-1} + \cdots + \beta_p h_{t-p}, \tag{7}$$

where ϵ_t's independent and identically distributed random variables with mean 0 and variance 1. As before we have $\alpha_i > 0$ for $i = 0, \dots, q$ and in addition $\beta_i > 0$ for $i = 0, \dots, p$. The GARCH(q, p) model can be motivated by defining $v_t = \xi_t^2 - h_t$, where v_t is a zero-mean white-noise term. By substituting $h_t = \xi_t^2 - v_t$ in (7) we can consider the alternate representation

$$\xi_t^2 = \alpha_0 + \sum_{i=1}^{s} (\alpha_i + \beta_i)\, \xi_{t-i}^2 + v_t - \sum_{j=1}^{p} \beta_j\, v_{t-j} \tag{8}$$

where $s = max(\mathrm{p}, \mathrm{q})$. The above representation gives us an ARMA(s, p) process for ξ_t^2 with $v_t = \xi_\mathrm{t}^2 - E(\xi_\mathrm{t}^2 | D_{t-1})$. Thus, v_t has the usual interpretation of one-step ahead forecast error for ξ_t^2. For example, if $\mathrm{q} = \mathrm{p} = 1$, then we have the GARCH(1, 1) model

$$h_t = \alpha_0 + \alpha_1 \xi_{\mathrm{t}-1}^2 + \beta_1 h_{t-1}$$

implying the form

$$\xi_\mathrm{t}^2 = \alpha_0 + (\alpha_1 + \beta_1)\, \xi_{\mathrm{t}-1}^2 + v_t - \beta_1\, v_{t-1}. \tag{9}$$

The above form gives us ARMA(1,1) for the squared series ξ_t^2.

For the unconditional variance of ξ_t to exist, we need to impose additional conditions on the parameters in (7). As in the ARCH model, for stationary GARCH process we have $E(\xi_t) = 0$. It can be shown that the unconditional variance is

$$E(\xi_\mathrm{t}^2) = \frac{\alpha_0}{1 - \sum_{\mathrm{i}=1}^{s} (\alpha_\mathrm{i} + \beta_\mathrm{i})} = V(\xi_t), \tag{10}$$

implying that $\sum_{\mathrm{i}=1}^{s} (\alpha_\mathrm{i} + \beta_\mathrm{i}) < 1$. For the special case GARCH(1, 1) this means $(\alpha_1 + \beta_1) < 1$ in (9), that is, the restriction is on the AR coefficient in the ARMA(1,1) representation. More complicated conditions are required for time invariance of higher-order moments.

There are nonstationary versions of the GARCH models that are analogous to the ARIMA models. An integrated GARCH (IGARCH) model is a GARCH model whose characteristic polynomial having a unit root. For example, IGARCH(1,1) model is a GARCH(1,1) model as in (9) where $(\alpha_1 + \beta_1) = 1$ implying that $\alpha_1 = (1 - \beta_1)$. Thus, the model can be written as

$$h_t = \alpha_0 + (1 - \beta_1)\, \xi_{\mathrm{t}-1}^2 + \beta_1 h_{t-1}. \tag{11}$$

In the above the unconditional variance $V(\xi_t)$ does not exist and the process ξ_t is not stationary. Other versions of the GARCH models include GARCH in mean (GARCH-M) and the exponential GARCH model of Nelson (1991).

In the sequel we will first discuss their Bayesian analysis using the Gibbs sampler proposed by Muller and Pole (1998). This is done in Section 2. This will be followed in Section 3 by our consideration of the multivariate ARCH/GARCH models and introduction of their Bayesian analysis by extending the method of Muller and Pole (1998) to the multivariate case. Implementation of our approach will be illustrated with an example in Section 4.

2. Bayesian Analysis of Univariate GARCH Models

Bayesian analysis of the GARCH(q, p) model (7) has been considered by various authors in the literature. In what follows, we will present a slighly modified version of the approach proposed by Muller and Pole (1998) which we will extend to multivariate GARCH models in Section 3. In our development we will focus on the GARCH$(1, 1)$ model, but extension to GARCH(q, p) model is straightforward and will be discussed.

We consider the GARCH(1,1) model

$$r_t = \mu + \xi_t$$
$$h_t = \alpha_0 + \alpha_1 \xi_{t-1}^2 + \beta_1 h_{t-1}, \tag{12}$$

where ξ_t given D_{t-1} is normally distributed with mean 0 and variance h_t, denoted as $\xi_t | D_{t-1} \sim N(0, h_t)$. Given return data $r^{(n)} = (r_1,\, r_2,\, \dots,\, r_n)$ from n periods, the likelihood function under the GARCH model (12) is given by

$$L(\mu,\, \alpha_0,\, \alpha_1,\, \beta_1\,;\; r^{(n)}) \propto (h_t)^{-n/2} exp\big[-\frac{1}{2}\sum_{t=1}^{n}(r_t-\mu)^2/h_t\big]. \tag{13}$$

There is no joint prior $p(\mu,\, \alpha_0,\, \alpha_1,\, \beta_1)$ that provides an analytically tractable posterior analysis with (13). Thus, in what follows, a Gibbs sampler will be presented to generate samples from the joint posterior distribution $p(\mu,\, \alpha_0,\, \alpha_1,\, \beta_1 |\, r^{(n)})$. This requires successive drawings from the *full conditional distributions* of $(\mu,\, \alpha_0,\, \alpha_1,\, \beta_1)$ given $r^{(n)}$; see Gelfand and Smith (1990) for details on the Gibbs sampler.

To obtain the full conditional distribution of μ, that is, $p(\mu|\alpha_0,\, \alpha_1,\, \beta_1, r^{(n)})$, we note that if the prior of μ is normal, say, $\mu \sim N(m_0,\, C_0)$, then posterior analysis follows using the standard Bayesian analysis of the normal model with known variance [see for example, Gelman et al. (2004, pp. 49)]. In our case it is important to note that the variance h_t is not constant and it involves h_{t-1} and $\xi_{\mathrm{t}-1}$ and evaluation of these terms requires some adjustment. Given h_t's are given for all periods, we can obtain the full conditional distribution of μ as $(\mu|\alpha_0,\, \alpha_1,\, \beta_1, r^{(n)}) \sim N(m_1,\, C_1)$ where

$$m_1 = \frac{\sum_{t=1}^{n} r_t/h_t \, + \, m_0/C_0}{\sum_{t=1}^{n} 1/h_t \, + \, 1/C_0}, \tag{14}$$

and

$$C_1 = \Big[\sum_{t=1}^{n} 1/h_t \, + \, 1/C_0\Big]^{-1}. \tag{15}$$

If the initial values $(\mu^0,\, \alpha_0^0,\, \alpha_1^0,\, \beta_1^0)$ and $(h_0$ and $\xi_0)$ are specified then after the $(i-1)th$ iteration of the Gibbs sampler given the values of parameters μ^{i-1} and $\boldsymbol{\gamma}^{i-1} = (\alpha_0^{i-1},\, \alpha_1^{i-1},\, \beta_1^{i-1})$ from the previous iteration we can obtain $(\xi_1^{i-1},\, \xi_2^{i-1},\, \ldots,\, \xi_n^{i-1})$ using

$$\xi_t^{i-1} = r_t - \mu^{i-1} \tag{16}$$

and obtain $(h_1^{i-1},\, h_2^{i-1},\, \ldots,\, h_{n-1}^{i-1})$ via

$$h_t^{i-1} = \alpha_0^{i-1} + \, \alpha_1^{i-1}(\xi_{t-1}^{i-1})^2 + \, \beta_1^{i-1} h_{t-1}^{i-1} \tag{17}$$

for $t = 1, \ldots, n-1$.

The full conditional distribution of $\boldsymbol{\gamma} = (\alpha_0 \;\; \alpha_1 \;\; \beta_1)'$, that is, $p(\boldsymbol{\gamma}|\; \mu,\, r^{(n)})$, can not be obtained analytically. Thus, at each iteration of the Gibbs sampler we can use a Metropolis step to draw from $p(\boldsymbol{\gamma}|\; \mu,\, r^{(n)})$; see for example, Chib and Greenberg (1995) for a review of the Metropolis algorithm. The algorithm uses a sample from a *probing* *distribution* (proposal density) and the selection of this probing distribution plays an important role in the efficiency of the algorithm. At each iteration the candidate sample selected from the probing distribution is accepted or rejected with a probability that depends on the magnitude of the true density $p(\boldsymbol{\gamma}|\ \mu,\, r^{(n)})$ at the sampled value. In selecting the probing distribution we will follow Muller and Pole (1998) who proposed to derive this distribution from an auxiliary regression model. In what follows, we will suppress the dependence of the true and the probing distributions on μ and denote them as $p(\boldsymbol{\gamma}|r^{(n)})$ and $g(\boldsymbol{\gamma}|r^{(n)})$.

Note that after the $(i-1)th$ iteration of the Gibbs sampler given the values of $\boldsymbol{\xi}^{i-1} = (\xi_1^{i-1}, \xi_2^{i-1}, \dots, \xi_n^{i-1})$ and $\boldsymbol{h}^{i-1} = (h_1^{i-1}, h_2^{i-1}, \dots, h_n^{i-1})$ obtained via (16) and (17), we can consider the regression model

$$\xi_{\mathrm{t}}^2 = \alpha_0 + \alpha_1 \xi_{\mathrm{t}-1}^2 + \beta_1 h_{t-1} + w_t, \tag{18}$$

which is motivated by the conditional expectation of ξ_t^2 given D_{t-1}, where the auxiliary error term $w_t|D_{t-1} \sim N(0, \sigma_w^2)$. Thus, the auxiliary regression model can be written as

$$\boldsymbol{U} = \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{w} \tag{19}$$

where $= (w_1\, w_2 \cdots\cdots w_n)'$,

$$\boldsymbol{U} = \begin{bmatrix} (\xi_1^{i-1})^2 \\ \vdots \\ (\xi_n^{i-1})^2 \end{bmatrix} \text{and } \boldsymbol{Z} = \begin{bmatrix} 1 & (\xi_0^{i-1})^2 & h_0^{i-1} \\ \vdots & \vdots & \vdots \\ 1 & (\xi_{n-1}^{i-1})^2 & h_{n-1}^{i-1} \end{bmatrix}.$$

The probing distribution at each iteration can be derived by obtaining the posterior distribution of $\boldsymbol{\gamma}$ given $(\boldsymbol{U}, \boldsymbol{Z})$ in (18). In so doing, we can use an improper joint prior for $(\boldsymbol{\gamma},\, \sigma_w^2)$ as being proportional to $1/\sigma_w$ and obtain the probing distribution as a multivariate normal density given by

$$(\boldsymbol{\gamma}|\boldsymbol{U}, \boldsymbol{Z},\, \sigma_w^2) \sim N(\widehat{\boldsymbol{\gamma}},\, \widehat{\boldsymbol{V}}\sigma_w^2), \tag{20}$$

where

$$\widehat{\boldsymbol{\gamma}} = \ (\boldsymbol{Z}'\ \boldsymbol{Z})^{-1}\,\boldsymbol{Z}'\boldsymbol{U}, \tag{21}$$

and $\widehat{\boldsymbol{V}} = \ (\boldsymbol{Z}'\ \boldsymbol{Z})^{-1}$. It can be also shown that the distribution of σ_w^2 is an inverted gamma distribution with parameters $(n-3)/2$ and $\widehat{\sigma}_w^2/2$ where

$$\widehat{\sigma}_w^2 = \frac{1}{n-3}\,(\boldsymbol{U}-\boldsymbol{Z}\widehat{\boldsymbol{\gamma}})'(\boldsymbol{U}-\boldsymbol{Z}\widehat{\boldsymbol{\gamma}}). \tag{22}$$

Details of the above development can be found in Gelman et. al. (2004, pp. 356). Thus, at the ith iteration of the Gibbs sampler the probing distribution $g(\boldsymbol{\gamma}|r^{(n)})$ is given by (20). We note that in drawing a *candidate sample* from (20) we can either first draw from the inverted gamma distribution of σ_w^2 or use (22) as an estimate of the σ_w^2.

At the ith iteration we draw a candidate, say $\boldsymbol{\gamma}^{\boldsymbol{c}}$ from the probing distribution (20) and then the new value $\boldsymbol{\gamma}^i$ is set to the candidate value, that is, $\boldsymbol{\gamma}^i = \boldsymbol{\gamma}^{\boldsymbol{c}}$ with probability

$$a(\boldsymbol{\gamma}^{i-1},\boldsymbol{\gamma}^{\boldsymbol{c}}) = min\Big\{1,\ \frac{\widehat{p}(\boldsymbol{\gamma}^{\boldsymbol{c}})\ g(\boldsymbol{\gamma}^{\boldsymbol{i-1}})}{g(\boldsymbol{\gamma}^{\boldsymbol{c}})\ \widehat{p}(\boldsymbol{\gamma}^{\boldsymbol{i-1}})}\Big\} \tag{23}$$

where

$$\widehat{p}(\boldsymbol{\gamma}) = (h_t)^{-1/2}exp\Big[-\frac{1}{2}\sum_{t=1}^{n}(r_t-\mu)^2/h_t\Big]p(\boldsymbol{\gamma}) \tag{24}$$

and $p(\boldsymbol{\gamma})$ is the joint prior density. Note that the Metropolis algorithm requires that the full conditional $p(\boldsymbol{\gamma}|\mu,\, r^{(n)})$ is only specified to a normalizing constant as given by (24). The probability $a(\boldsymbol{\gamma}^{i-1},\boldsymbol{\gamma}^{\boldsymbol{c}})$ implies that if the ratio of the distributions in (23) is large then the probability of acceptance is high. At each iteration, we generate a uniform $(0,\ 1)$ random variable, say u, and if $u \leq\ a(\boldsymbol{\gamma}^{i-1},\boldsymbol{\gamma}^{\boldsymbol{c}})$, then the candidate is accepted, that is, $\boldsymbol{\gamma}^i = \boldsymbol{\gamma}^{\boldsymbol{c}}$, otherwise we set $\boldsymbol{\gamma}^i = \boldsymbol{\gamma}^{\boldsymbol{i-1}}$. Note that the candidate $\boldsymbol{\gamma}^{\boldsymbol{c}}$ is considered for acceptance only if $\alpha_0 > 0$ and $0 < (\alpha_1+\beta_1) < 1$.

Once $\boldsymbol{\gamma}^i$ is generated, we update h_t's based on the new parameters, that is, via

$$\widehat{h}_t^i = \alpha_0^i + \alpha_1^i(\xi_{t-1}^{i-1})^2 + \beta_1^i \widehat{h}_{t-1}^i. \tag{25}$$

Note that the updating given by (25) is based on previous error estimates, that is, based on ξ_t^{i-1}'s. Thus, (25) is different than the updating given by (17) which is based on current error terms. Once h_t's are obtained via (25) μ^i is drawn from the normal density whose mean and variance are given by (14) and (15). Once μ^i is drawn then a new set of error terms and h_t's are obtained via (16) and (17) and the above process is repeated for iteration $(i+1)$. Continuing with these successive draws samples are obtained from the posterior distribution $p(\mu,\, \alpha_0,\, \alpha_1,\, \beta_1 |\, r^{(n)})$.

The above algorithm can be easily generalized for the GARCH(q, p) model where simply the dimension of the parameter vector $\boldsymbol{\gamma}$ increases to $(\mathrm{q}+\mathrm{p}+1)$. Similarly, in other cases, where the observation model $r_t = \mu + \xi_t$ may include covariates, the algorithm can be modified so that the mean and variance terms (14) and (15) represent posterior mean and variance of regression parameters in the updating.

3. Multivariate ARCH/GARCH Models

A multivariate extension of the univariate ARCH and generalized ARCH models of Engle (1982) and Bollerslev (1986), is introduced by Bollerslev et al. (1988). We consider the multivariate version of the observation model (1) as

$$\boldsymbol{r}_t = \boldsymbol{\mu} + \boldsymbol{\xi}_t, \tag{26}$$

where $\boldsymbol{r}_t$ is the K dimensional return vector, $\boldsymbol{\mu}$ is the K dimensional mean return, $\boldsymbol{\xi}_t \big| \mathbf{D}_{t-1} \sim \boldsymbol{N}(\mathbf{0},\, \boldsymbol{H}_t)$ and $\boldsymbol{H}_t$ is $K \times K$ variance-covariance matrix. Models for $\boldsymbol{H}_t$ are referred to as multivariate volatility models for the return series $\boldsymbol{r}_t$.

Different modeling strategies have been suggested in the literature to describe the evolution of $\boldsymbol{H}_t$. In what follows we will consider the setup given by Bollerslev et al.

(1988) where the multivariate GARCH(q, p) model is defined by the volatility equation given by

$$vech\left(\boldsymbol{H}_t\right) = \boldsymbol{A}_0 + \sum_{i=1}^{q} \boldsymbol{A}_i\, vech\left(\boldsymbol{\xi}_{t-i}\boldsymbol{\xi}_{t-i}'\right) + \sum_{j=1}^{p} \boldsymbol{B}_j\, vech\left(\boldsymbol{H}_{t-j}\right), \tag{27}$$

where $vech(\cdot)$ denotes the column stacking operator of the lower portion of a symmetric matrix, $\boldsymbol{A}_0$ is a $s = K(K+1)/2$ vector, $\boldsymbol{A}_i$'s and $\boldsymbol{B}_j$ are s $\times$ s matrices. We note that certain conditions need to be satisfied in (27) for $\boldsymbol{H}_t$ to be a positive definite matrix. Similar to the univariate case a multivariate GARCH(q, p) model can be motivated as a multivariate ARMA(q, p) for $vech\left(\boldsymbol{\xi}_t\boldsymbol{\xi}_t'\right)$. As a result the stationarity conditions for multivariate ARMA(q, p) processes need to be satisfied in (27). These conditions are equivalent to conditions on the roots of the matrix polynomial $\boldsymbol{A}(L) = (\boldsymbol{I}_s - \boldsymbol{A}_1 L - \cdots - \boldsymbol{A}_q L^q)$, where L is the lag operator and is $\boldsymbol{I}_s$ the $s \times s$ identity matrix. The requirement is that the roots of the determinant $|\boldsymbol{A}(L)|$ be outside the unit circle; see for example, Tsay (2002), pp. 322.

The multivariate ARCH models are obtained from (27) by setting $p = 0$. Similar to the univariate ARCH models, multivariate ARCH(q) models imply that $vech\left(\boldsymbol{\xi}_t\boldsymbol{\xi}_t'\right)$ follows a vector autoregressive process of order q. In other words, we can motivate an ARCH(q) process via

$$vech\left(\boldsymbol{\xi}_t\boldsymbol{\xi}_t'\right) = \boldsymbol{A}_0 + \sum_{i=1}^{q} \boldsymbol{A}_i vech\left(\boldsymbol{\xi}_{t-i}\boldsymbol{\xi}_{t-i}'\right) + \boldsymbol{\omega}_t, \tag{28}$$

where $\boldsymbol{\omega}_t = vech\left(\boldsymbol{H}_t\right) - vech\left(\boldsymbol{\xi}_t\boldsymbol{\xi}_t'\right)$ is a zero-mean white-noise process.

We note that the model (27) is a highly parameterized representation. For example, the first order ARCH model is given by

$$vech\left(\boldsymbol{H}_t\right) = \boldsymbol{A}_0 + \boldsymbol{A}_1\, vech\left(\boldsymbol{\xi}_{t-i}\boldsymbol{\xi}_{t-i}'\right), \tag{29}$$

where for the two-dimensional case, $K = 2$, $\boldsymbol{A}_1$ is a (3×3) matrix and the model can be written as

$$
vech(\boldsymbol{H}_t) = \begin{pmatrix} H_{11t} \\ H_{12t} \\ H_{22t} \end{pmatrix} = \begin{bmatrix} \alpha_{01} \\ \alpha_{02} \\ \alpha_{03} \end{bmatrix} + \begin{bmatrix} \alpha_{11} & \alpha_{21} & \alpha_{31} \\ \alpha_{12} & \alpha_{22} & \alpha_{32} \\ \alpha_{13} & \alpha_{23} & \alpha_{33} \end{bmatrix} \begin{bmatrix} \xi_{1,t-1}^2 \\ \xi_{1,t-1} \cdot \xi_{2,t-1} \\ \xi_{2,t-1}^2 \end{bmatrix}. \tag{30}
$$

The stationarity of $vech\left(\boldsymbol{\xi}_t \boldsymbol{\xi}_t'\right)$ in (29) requires that the root of the determinant $|(\boldsymbol{I}_s - \boldsymbol{A}_1 L)|$ be outside the unit circle or equivalently all the eigenvalues of $\boldsymbol{A}_1$ to be less than 1.

A natural simplification is obtained by assuming a diagonal structure for the $\boldsymbol{A}_1$ matrix in (30) as

$$
\boldsymbol{A}_1 = \begin{bmatrix} \alpha_{11} & 0 & 0 \\ 0 & \alpha_{22} & 0 \\ 0 & 0 & \alpha_{33} \end{bmatrix}
$$

that is, $\alpha_{\text{ij}} = 0$ if i $\neq$ j, As pointed out by Bollerslev et al. (1988), this simply implies that each covariance depends only on its past values, that is,

$$
\begin{pmatrix} H_{11t} \\ H_{12t} \\ H_{22t} \end{pmatrix} = \begin{bmatrix} \alpha_{01} + \alpha_{11}\, \xi_{1,t-1}^2 \\ \alpha_{02} + \alpha_{22}\, \xi_{1,t-1} \cdot \xi_{2,t-1} \\ \alpha_{03} + \alpha_{33}\, \xi_{2,t-1}^2 \end{bmatrix}. \tag{31}
$$

Similar simplifications can be obtained for multivariate GARCH models which are motivated by $vech\left(\boldsymbol{\xi}_t \boldsymbol{\xi}_t'\right)$being a vector ARMA process.

3.1 Bayesian Analysis of Multivariate ARCH/GARCH Models

In what follows we will introduce a Bayesian approach for the analysis of multivariate ARCH/GARCH models. Our approach is an extension of the Markov chain Monte Carlo method of Muller and Pole (1998), presented in section 2, to multivariate models. For illustrative purposes we will present our approach using the first order ARCH model of (29). Extension to the general GARCH$(q,\, p)$ models is straightforward and the necessary details will be summarized at the end of the section.

We consider the multivariate GARCH(1, 1) model

$$\boldsymbol{r}_t = \boldsymbol{\mu} + \boldsymbol{\xi}_t,$$
$$vech\Big(\boldsymbol{H}_t\Big) = \boldsymbol{A}_0 + \boldsymbol{A}_1\, vech\Big(\boldsymbol{\xi}_{t-i}\boldsymbol{\xi}_{t-i}'\Big) + \boldsymbol{B}_1\, vech\Big(\boldsymbol{H}_{t-1}\Big), \tag{32}$$

where $\boldsymbol{\xi}_t\big|\mathbf{D}_{t-1} \sim \boldsymbol{N}(\mathbf{0},\ \ \boldsymbol{H}_t)$, $\boldsymbol{\xi}_t = (\xi_{1t}\ \xi_{2t} \cdots\cdots \xi_{Kt})'$, $\boldsymbol{\mu} = (\mu_1\ \mu_2 \cdots\cdots \mu_K)'$ and $vech\big(\boldsymbol{H}_t\big)$ is the extension of (30) with dimension $s = K(K+1)/2$. Note that (32) is a multivariate extension of (12). For example, for the bivariate case, that is, $K = 2$, (32) reduces to

$$\begin{pmatrix} H_{11t} \\ H_{12t} \\ H_{22t} \end{pmatrix} = \begin{bmatrix} \alpha_{01} \\ \alpha_{02} \\ \alpha_{03} \end{bmatrix} + \begin{bmatrix} \alpha_{11} & \alpha_{21} & \alpha_{31} \\ \alpha_{12} & \alpha_{22} & \alpha_{32} \\ \alpha_{13} & \alpha_{23} & \alpha_{33} \end{bmatrix} \begin{bmatrix} \xi_{1,t-1}^2 \\ \xi_{1,t-1}\xi_{2,t-1} \\ \xi_{2,t-1}^2 \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{21} & \beta_{31} \\ \beta_{12} & \beta_{22} & \beta_{32} \\ \beta_{13} & \beta_{23} & \beta_{33} \end{bmatrix} \begin{pmatrix} H_{11,t-1} \\ H_{12,t-1} \\ H_{22,t-1} \end{pmatrix}$$

Given $\boldsymbol{r}^{(n)} = (\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_n)$ the return data from n periods, the likelihood function of $\boldsymbol{\mu}$, $\boldsymbol{A}_0$, $\boldsymbol{A}_1$ and $\boldsymbol{B}_1$ is given by

$$L(\boldsymbol{\mu}, \boldsymbol{A}_0, \boldsymbol{A}_1, \boldsymbol{B}_1;\ \boldsymbol{r}^{(n)}) \propto |\boldsymbol{H}_t|^{-1/2} exp\Big[-\frac{1}{2}\sum_{t=1}^{n}(\boldsymbol{r}_t - \boldsymbol{\mu})'\boldsymbol{H}_t^{-1}(\boldsymbol{r}_t - \boldsymbol{\mu})\Big], \tag{33}$$

where $\boldsymbol{H}_t$ is defined via (32) and $\boldsymbol{A}_0$ is an $s-$ dimensional vector and $\boldsymbol{A}_1$ and $\boldsymbol{B}_1$ are $s \times s$ matrices. The Bayesian analysis involves specification of the joint prior $p(\boldsymbol{\mu}, \boldsymbol{A}_0, \boldsymbol{A}_1, \boldsymbol{B}_1)$. As in the univariate case, there is no joint prior that provides an analytically tractable posterior analysis when combined with the likelihood (33). Thus, we will develop a Gibbs sampler algorithm to generate samples from the posterior distribution $p(\boldsymbol{\mu}, \boldsymbol{A}_0, \boldsymbol{A}_1, \boldsymbol{B}_1 | \boldsymbol{r}^{(n)})$ by generating successive drawings from the full conditional distributions of $\boldsymbol{\mu}$, $\boldsymbol{A}_0, \boldsymbol{A}_1$ and $\boldsymbol{B}_1$ given $\boldsymbol{r}^{(n)}$.

For our development of the Gibbs sampler, in (32) we assume that the mean return vector $\boldsymbol{\mu}$ has a normal prior, say, $\boldsymbol{\mu} \sim N(\boldsymbol{m}, \boldsymbol{C})$ where $\boldsymbol{m}$ is a specified $K \times \mathbf{1}$ vector and $\boldsymbol{C}$ is a specified $K \times K$ matrix. The full conditional posterior distribution of $\boldsymbol{\mu}$ is given by

$$p(\boldsymbol{\mu}\,|\,\boldsymbol{A}_0, \boldsymbol{A}_1, \boldsymbol{B}_1, \boldsymbol{r}^{(n)}) \propto L(\boldsymbol{\mu}\,; \boldsymbol{A}_0, \boldsymbol{A}_1, \boldsymbol{B}_1, \boldsymbol{r}^{(n)})\, p(\boldsymbol{\mu})$$

$$\propto exp\Big[-\frac{1}{2}\Big(\sum_{t=1}^{n}(\boldsymbol{r}_t-\boldsymbol{\mu})'\boldsymbol{H}_t^{-1}(\boldsymbol{r}_t-\boldsymbol{\mu})+(\boldsymbol{\mu}-\boldsymbol{m})'\boldsymbol{C}^{-1}(\boldsymbol{\mu}-\boldsymbol{m})\Big)\Big]. \quad (34)$$

The above can be written as proportional to

$$\propto exp\Big[-\frac{1}{2}\Big(\boldsymbol{\mu}'\,(\sum_{t=1}^{n}\boldsymbol{H}_t^{-1}+\boldsymbol{C}^{-1})\boldsymbol{\mu}-2\boldsymbol{\mu}'\,(\sum_{t=1}^{n}(\boldsymbol{H}_t^{-1}\boldsymbol{r}_t+\boldsymbol{C}^{-1}\boldsymbol{m})\Big)\Big],$$

implying that

$$p(\boldsymbol{\mu}\,|\,\boldsymbol{A}_0,\boldsymbol{A}_1,\boldsymbol{B}_1,\boldsymbol{r}^{(n)})\propto exp\Big[-\frac{1}{2}\Big((\boldsymbol{\mu}-\boldsymbol{m}^*)'(\boldsymbol{C}^*)^{-1}(\boldsymbol{\mu}-\boldsymbol{m}^*)\Big) \quad (35)$$

where

$$\boldsymbol{m}^*=\Big(\sum_{t=1}^{n}\boldsymbol{H}_t^{-1}+\boldsymbol{C}^{-1}\Big)^{-1}\Big(\sum_{t=1}^{n}\boldsymbol{H}_t^{-1}\boldsymbol{r}_t+\boldsymbol{C}^{-1}\boldsymbol{m}\Big) \quad (36)$$

and

$$(\boldsymbol{C}^*)^{-1}=\Big(\sum_{t=1}^{n}\boldsymbol{H}_t^{-1}+\boldsymbol{C}^{-1}\Big)^{-1}. \quad (37)$$

Thus, the full conditional posterior distribution of $\boldsymbol{\mu}$ is obtained as a multivariate normal as $(\boldsymbol{\mu}\,|\,\boldsymbol{A}_0,\boldsymbol{A}_1,\boldsymbol{B}_1,\boldsymbol{r}^{(n)})\sim N(\boldsymbol{m}^*,\boldsymbol{C}^*)$. Note that the posterior mean and precision updates given by (36) and (37) are multivariate versions of (14) and (15).

The full conditional distributions of $\boldsymbol{A}_0,\boldsymbol{A}_1$ and $\boldsymbol{B}_1$, that is, $p(\boldsymbol{A}_0|\,\boldsymbol{\mu},\boldsymbol{A}_1,\,\boldsymbol{B}_1,\,\boldsymbol{r}^{(n)})$, $p(\boldsymbol{A}_1|\,\boldsymbol{\mu},\boldsymbol{A}_0,\boldsymbol{B}_1,\boldsymbol{r}^{(n)})$, $p(\boldsymbol{B}_1|\,\boldsymbol{\mu},\boldsymbol{A}_0,\boldsymbol{A}_1,\boldsymbol{r}^{(n)})$ and are not available as known distributional forms. Thus, we will present an extension of the approach presented in Section 2 for the univariate ARCH/GARCH models. Similar to the development, we will use a Metropolis step at each iteration of the Gibbs sampler to draw from the full conditional distribution of $(\boldsymbol{A}_0,\boldsymbol{A}_1,\boldsymbol{B}_1)$. In so doing, our probing distribution for $(\boldsymbol{A}_0,\,\boldsymbol{A}_1,\boldsymbol{B}_1)$ is derived from an auxiliary multivariate linear regression model and as a result we will use results from Bayesian multivariate analysis; see for example Press (1989, pp. 131-137). In the sequel, we will review the Bayesian multivariate regression analysis and adopt some of the results to our problem.

If the initial values $(\boldsymbol{\mu}^{\mathbf{0}},\ \boldsymbol{A}_0^{\mathbf{0}}, \boldsymbol{A}_1^{\mathbf{0}}, \boldsymbol{B}_1^{\mathbf{0}})$ and $(\boldsymbol{H}_0,\ \boldsymbol{\xi}_0)$ are specified then after the $(i-1)th$ iteration of the Gibbs sampler, given the values of parameter vectors and matrix $(\boldsymbol{\mu}^{\boldsymbol{i-1}},\ \boldsymbol{A}_0^{\boldsymbol{i-1}}, \boldsymbol{A}_1^{\boldsymbol{i-1}}, \boldsymbol{B}_1^{\boldsymbol{i-1}})$ from the previous iteration we can obtain the error vectors $(\boldsymbol{\xi}_1^{i-1},\ \boldsymbol{\xi}_2^{i-1},\ \ldots,\ \boldsymbol{\xi}_n^{i-1})$ using

$$\boldsymbol{\xi}_t^{i-1} = \boldsymbol{r}_t - \boldsymbol{\mu}^{i-1}. \tag{38}$$

Also, we can obtain $vech\left(\boldsymbol{H}_1^{\boldsymbol{i-1}}\right), \ldots, vech\left(\boldsymbol{H}_{n-1}^{\boldsymbol{i-1}}\right)$ via the volatilility equation (32). Once these values are available we consider the auxiliary multivariate regression model

$$vech\left(\boldsymbol{\xi}_t\boldsymbol{\xi}_t^{'}\right) = \boldsymbol{A}_0 + \boldsymbol{A}_1 vech\left(\boldsymbol{\xi}_{t-i}\boldsymbol{\xi}_{t-i}^{'}\right) + \boldsymbol{B}_1\, vech\left(\boldsymbol{H}_{t-1}\right) + \boldsymbol{\omega}_t, \tag{39}$$

where the auxiliary error vector $\boldsymbol{\omega}_t\big|\mathbf{D}_{t-1} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_w)$ with $s \times\ s$ covariance matrix $\boldsymbol{\Sigma}_w$. Note that (39) is motivated by the vector ARMA$(1,1)$ process representation of the multivariate GARCH$(1,1)$ model. Given n, $(s \times 1)$ dimensional vectors $(\boldsymbol{\xi}_1^{i-1},\ \boldsymbol{\xi}_2^{i-1},\ \ldots,$ $\boldsymbol{\xi}_n^{i-1})$ and the (39) we can represent this as a multivariate linear model at the $(i-1)th$ iteration where the dependent variable matrix is given by $\boldsymbol{U} = \left[vech\left(\boldsymbol{\xi}_1\boldsymbol{\xi}_1^{'}\right)^{'}\right.$ $\left.vech\left(\boldsymbol{\xi}_2\boldsymbol{\xi}_2^{'}\right)^{'} \cdots\cdots vech\left(\boldsymbol{\xi}_n\boldsymbol{\xi}_n^{'}\right)^{'}\right]^{'}$. Note that $\boldsymbol{U}$ is a $(n \times s)$ matrix where each row is the s dimensional row vector $vech\left(\boldsymbol{\xi}_t\boldsymbol{\xi}_t^{'}\right)^{'}$ based on $(i-1)th$ iteration values. For the case where $K = 2$, we have $s = 3$ and $\boldsymbol{U}$ is given by

$$\boldsymbol{U} = \begin{bmatrix} \xi_{11}^2 & \xi_{11}\,\xi_{21} & \xi_{21}^2 \\ \vdots & \vdots & \vdots \\ \xi_{1n}^2 & \xi_{1n}\,\xi_{2n} & \xi_{2n}^2 \end{bmatrix}. \tag{40}$$

Note that for the $K > 2$ dimension case, $t-th$ row of the $\boldsymbol{U}$ matrix will be given by

$$(\xi_{1t}^2\ \ \xi_{1t}\,\xi_{2t} \cdots \xi_{1t}\,\xi_{Kt}\ \ \xi_{2t}^2\ \ \xi_{2t}\,\xi_{3t}\ \cdots\ \xi_{2t}\,\xi_{Kt} \cdots\cdots \xi_{K-1,t}\,\xi_{Kt}\ \ \xi_{Kt}^2).$$

We will define the coefficient matrix of the auxiliary multivariate regression model by $\boldsymbol{\Gamma} = [\boldsymbol{A}_0^{\prime}\ \boldsymbol{A}_1^{\prime}\ \boldsymbol{B}_1^{\prime}]^{\prime}$ which is a $(2s+1) \times s$ matrix. For the bivariate case where $s = 3,\ \boldsymbol{\Gamma}$ is given by

$$
\boldsymbol{\Gamma} = \begin{bmatrix} \alpha_{01} & \alpha_{02} & \alpha_{03} \\ \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \\ \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \end{bmatrix} \tag{41}
$$

Note that for the ARCH(q) models with $q > 1$, we define $\boldsymbol{\Gamma} = [\boldsymbol{A}_0' \ \boldsymbol{A}_1' \ \cdots \ \boldsymbol{A}_q' \ \boldsymbol{B}_1' \ \cdots \ \boldsymbol{B}_p']'$.

The design matrix of the multivariate linear regression will be given by $n \times (2s+1)$ matrix $\boldsymbol{Z}$ with row t is given by

$$
(1 \ \xi_{1,t-1}^2 \cdots \xi_{1,t-1}\xi_{K,t-1} \ \ \xi_{2,t-1}^2 \cdots \xi_{2,t-1}\xi_{K,t-1} \cdots\cdots \xi_{K,t-1}^2 \ H_{11,t-1} \ \cdots H_{1K,t-1} \cdots H_{KK,t-1}).
$$

For example, for the two-dimensional case where $s = 3$ we have $(n \times 7)$ matrix

$$
\boldsymbol{Z} = \begin{bmatrix} 1 & \xi_{10}^2 & \xi_{10}\,\xi_{20} & \xi_{20}^2 & H_{110} & H_{120} & H_{220} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \xi_{1,n-1} & \xi_{1,n-1}\,\xi_{2,n-1} & \xi_{2,n-1} & H_{11,n-1} & H_{12,n-1} & H_{22,n-1} \end{bmatrix}. \tag{42}
$$

We also define the $(n \times s)$ error matrix as

$$
\boldsymbol{\Omega} = \begin{bmatrix} \omega_{11} & \omega_{21} & \omega_{s1} \\ \vdots & \vdots & \vdots \\ \omega_{1n} & \omega_{2n} & \omega_{sn} \end{bmatrix}. \tag{43}
$$

Thus, at the $(i-1)th$ iteration of the Gibbs sampler, the auxiliary multivariate linear model can be written as

$$
\boldsymbol{U} = \boldsymbol{Z}\boldsymbol{\Gamma} + \boldsymbol{\Omega}. \tag{44}
$$

Given $\boldsymbol{U}$ and $\boldsymbol{Z}$, the likelihood function of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}_w$, where $\boldsymbol{\Sigma}_w$ is the variance-covariance matrix of the auxiliary error vector $\boldsymbol{\omega}_t$, is given by

$$
L(\boldsymbol{\Gamma}, \boldsymbol{\Sigma}_w; \boldsymbol{U}, \boldsymbol{Z}) \propto |\boldsymbol{\Sigma}_w|^{-n/2}\, exp\Big[-tr(\boldsymbol{\Sigma}_w^{-1})(\boldsymbol{U} - \boldsymbol{Z}\boldsymbol{\Gamma})'(\boldsymbol{U} - \boldsymbol{Z}\boldsymbol{\Gamma})/2 \Big], \tag{45}
$$

where $tr(\mathbf{\Sigma}_w^{-1})$ denotes the trace of the matrix. The probing distribution at iteration i can be derived by obtaining the posterior distribution of the random matrix $\mathbf{\Gamma}$ given $\boldsymbol{U}$ and $\boldsymbol{Z}$. In so doing, we can use an improper prior for $\mathbf{\Gamma}$ and $\mathbf{\Sigma}_w$ as

$$p(\mathbf{\Gamma}, \mathbf{\Sigma}_w) \propto |\mathbf{\Sigma}_w|^{-(s+1)/2}. \tag{46}$$

It can be shown that the least squares estimator of matrix $\mathbf{\Gamma}$ is given by

$$\widehat{\mathbf{\Gamma}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{U} \tag{47}$$

and using the orthogonality property of the least squares estimators [see for example, Press (1989), pp. 134], the joint posterior distribution $p(\mathbf{\Gamma}, \mathbf{\Sigma}_w | \boldsymbol{U}, \boldsymbol{Z})$ can be written as

$$\propto |\mathbf{\Sigma}_w|^{-(n+s+1)/2} exp\Big[-tr(\mathbf{\Sigma}_w^{-1})\big[\boldsymbol{E} + (\mathbf{\Gamma} - \widehat{\mathbf{\Gamma}})'\boldsymbol{Z}'\boldsymbol{Z}(\mathbf{\Gamma} - \widehat{\mathbf{\Gamma}})\big]/2\Big], \tag{48}$$

where $\boldsymbol{E} = (\boldsymbol{U} - \boldsymbol{Z}\widehat{\mathbf{\Gamma}})'(\boldsymbol{U} - \boldsymbol{Z}\widehat{\mathbf{\Gamma}})$.

From (48) we can obtain the posterior distribution of $\mathbf{\Gamma}$ given $\mathbf{\Sigma}_w$ as

$$p(\mathbf{\Gamma}| \mathbf{\Sigma}_w, \boldsymbol{U}, \boldsymbol{Z}) \propto exp\Big[-tr(\mathbf{\Sigma}_w^{-1})\big[(\mathbf{\Gamma} - \widehat{\mathbf{\Gamma}})'\boldsymbol{Z}'\boldsymbol{Z}(\mathbf{\Gamma} - \widehat{\mathbf{\Gamma}})\big]/2\Big]. \tag{49}$$

The distribution given by (49) is known as a *matrix normal distribution* with mean matrix $\widehat{\mathbf{\Gamma}}$, $(s \times s)$ left variance matrix $\mathbf{\Sigma}_w$ and $(2s+1) \times (2s+1)$ right variance matrix $(\boldsymbol{Z}'\boldsymbol{Z})^{-1}$. Note that is $\mathbf{\Gamma}$ a $(2s+1) \times s$ random matrix and the variance-covariance matrix of $\mathbf{\Gamma}$ is given by the $[(2s+1)s \times (2s+1)s]$ matrix $\mathbf{\Sigma}_w \otimes (\boldsymbol{Z}'\boldsymbol{Z})^{-1}$ where $\otimes$ is the Kroenecker product. Thus, $(\mathbf{\Sigma}_w)_{ii}\,(\boldsymbol{Z}'\boldsymbol{Z})^{-1}$ defines the covariance matrix for the ith row of $\mathbf{\Gamma}$ whereas $[(\boldsymbol{Z}'\boldsymbol{Z})^{-1}]_{jj}\,\mathbf{\Sigma}_w$ defines the covariance matrix for the jth column. The above implies that all the elements of the random matrix $\mathbf{\Gamma}$ have univariate, multivariate or matrix normal distributions [see Dawid (1981) for a review of matrix normal distribution].

Using (48) and (49), it can be shown that the posterior distribution of $\mathbf{\Sigma}_w$ is an *inverse-Wishart distribution* with degrees of freedom $(n-2s-1)$ and $(n \times s)$ scale matrix

$$\boldsymbol{E} = (\boldsymbol{U} - \boldsymbol{Z}\widehat{\boldsymbol{\Gamma}})'(\boldsymbol{U} - \boldsymbol{Z}\widehat{\boldsymbol{\Gamma}}); \tag{50}$$

see Press (1982), pp. 136, for details. Thus, at the $ithe$ iteration of the Gibbs sampler the probing distribution of the full conditional of $\boldsymbol{\Gamma}$ is given by the matrix normal distribution (49) where a value for $\boldsymbol{\Sigma}_w$ can be either drawn from the inverse-Wishart distribution or (50) can be used to estimate $\boldsymbol{\Sigma}_w$, after divided by the degrees of freedom $(n-2s-1)$.

As before, we suppress the dependence of the true and the probing distributions on $\boldsymbol{\mu}$ and data and denote them as $p(\boldsymbol{\Gamma}|r^{(n)})$ and $g(\boldsymbol{\Gamma}|r^{(n)})$. Then at the ith iteration we draw a candidate, say $\boldsymbol{\Gamma^c}$ from the probing distribution (49) and then the new value $\boldsymbol{\Gamma}^i$ is set to the candidate value, that is, $\boldsymbol{\Gamma}^i = \boldsymbol{\Gamma^c}$ with probability

$$a(\boldsymbol{\Gamma}^{i-1}, \boldsymbol{\Gamma^c}) = min\left\{1,\ \frac{\widehat{p}(\boldsymbol{\Gamma^c})\ g(\boldsymbol{\Gamma^{i-1}})}{g(\boldsymbol{\Gamma^c})\ \widehat{p}(\boldsymbol{\Gamma^{i-1}})}\right\} \tag{51}$$

where

$$\widehat{p}(\boldsymbol{\Gamma}) = |\boldsymbol{\Sigma}_w|^{-n/2}\, exp\Big[-tr(\boldsymbol{\Sigma}_w^{-1})(\boldsymbol{U} - \boldsymbol{Z\Gamma})'(\boldsymbol{U} - \boldsymbol{Z\Gamma})/2\Big]p(\boldsymbol{\Gamma}) \tag{52}$$

and $p(\boldsymbol{\Gamma})$ is the prior density for the matrix of coefficients. Note that the probability $a(\boldsymbol{\Gamma}^{i-1}, \boldsymbol{\Gamma^c})$ implies that if the ratio of the distributions in (51) is large then the probability of acceptance is high. At each iteration, we generate a uniform $(0,\ 1)$ random variable, say u, and if $u \leq\ a(\boldsymbol{\Gamma}^{i-1}, \boldsymbol{\Gamma^c})$, then the candidate is accepted, that is, $\boldsymbol{\Gamma}^i = \boldsymbol{\Gamma^c}$**,** otherwise we set $\boldsymbol{\Gamma}^i = \boldsymbol{\Gamma^{i-1}}$. Note that the candidate $\boldsymbol{\Gamma^c}$ is considered for acceptance only if conditions for positive definiteness are satisfied in (32). More specifically these imply the positive definiteness of the $\boldsymbol{A}_1$ and $\boldsymbol{B}_1$ matrices. Thus, at each iteration the algorithm checks whether the resulting matrices have positive eigen values.

Once $\boldsymbol{\Gamma}^i$ is generated, we update $vech\Big(\boldsymbol{H}_t^{\boldsymbol{i-1}}\Big)$'s via the volatility equation (32) based on the new parameters, that is, via

$$vech\big(\widehat{\boldsymbol{H}}_t^{\boldsymbol{i-1}}\big) = \boldsymbol{A}_0^{\boldsymbol{i-1}} +\ \boldsymbol{A}_1^{\boldsymbol{i-1}}\, vech\Big(\boldsymbol{\xi}_{t-i}\boldsymbol{\xi}_{t-i}'\Big)^{i-1} +\ \boldsymbol{B}_1^{\boldsymbol{i}}\, vech\Big(\widehat{\boldsymbol{H}}_{t-1}^{\boldsymbol{i}}\Big). \tag{53}$$

Note that, as before, the updating given by (53) is based on previous error estimates, that is, based on $\boldsymbol{\xi}_{t-i}^{\boldsymbol{i-1}}$'s. Thus, (53) is different than the earlier updating after iteration $(i-1)$, which is based on current error terms. Once $vech\left(\boldsymbol{H}_t^{\boldsymbol{i-1}}\right)$'s are obtained via (53) $\boldsymbol{\mu}^i$ is drawn from the multivariate normal density given by (35). Continuing with these successive draws samples are obtained from the posterior distribution $p(\boldsymbol{\mu},\, \boldsymbol{A}_0, \boldsymbol{A}_1, \boldsymbol{B}_1|$ $\boldsymbol{r}^{(n)})$.

4. A Numerical Illustration

We first consider a multivariate ARCH(1) model with diagonal structure given by (30) in our illustration. We assume that the components of $\boldsymbol{A}_0$ vector are independent gamma distributed random quantities, that is, $\boldsymbol{\alpha}_{0,i} \sim Gamma(0.75,\, 1),\, i = 1, \ldots, 3.$ Similarly, the diagonal matrix $\boldsymbol{A}_1$ has independent gamma components, $\boldsymbol{\alpha}_{1,i,i} \sim Gamma(0.75,\, 1),\, i = 1, \ldots, 3.$ The mean vector $\boldsymbol{\mu}$ is assumed to have a normal prior with mean vector $\boldsymbol{m}_0 = \begin{bmatrix} 0.030 \\ 0.024 \end{bmatrix}$ and variance covariance matrix $\boldsymbol{C}_0 = \begin{bmatrix} 0.010 & 0.006 \\ 0.006 & 0.010 \end{bmatrix}$. Furthermore, we assume that apriori $\boldsymbol{A}_0$, $\boldsymbol{A}_1$ and $\boldsymbol{\mu}$ independent of each other and specify $\boldsymbol{\xi}_0 = \begin{bmatrix} 0.0 \\ 0.0 \end{bmatrix}$.

Using real securities data from companies Amoco and GM, we employ the Gibbs sampler described in Section 3.1 to sample from $\boldsymbol{A}_0$, $\boldsymbol{A}_1$ and $\boldsymbol{\mu}$. The Gibbs sampler was run for 10,000 iterations. In this case the full conditional distribution of $\boldsymbol{A}_0$ and $\boldsymbol{A}_1$ can not be obtained analytically. Thus, at each iteration of the Gibbs sampler we use a Metropolis step to draw from $\boldsymbol{A}_0$ and $\boldsymbol{A}_1$.

Figure 1 is autocorrelation function graph of one of the accepted $\boldsymbol{\alpha}_{11}$ series, showing the autocorrelation is rapidly decreasing. Figure 2 shows trace plot of the same $\boldsymbol{\alpha}_{11}$, showing there is no trends involved. Figure 3 shows histogram with the density line of the same $\boldsymbol{\alpha}_{11}$.

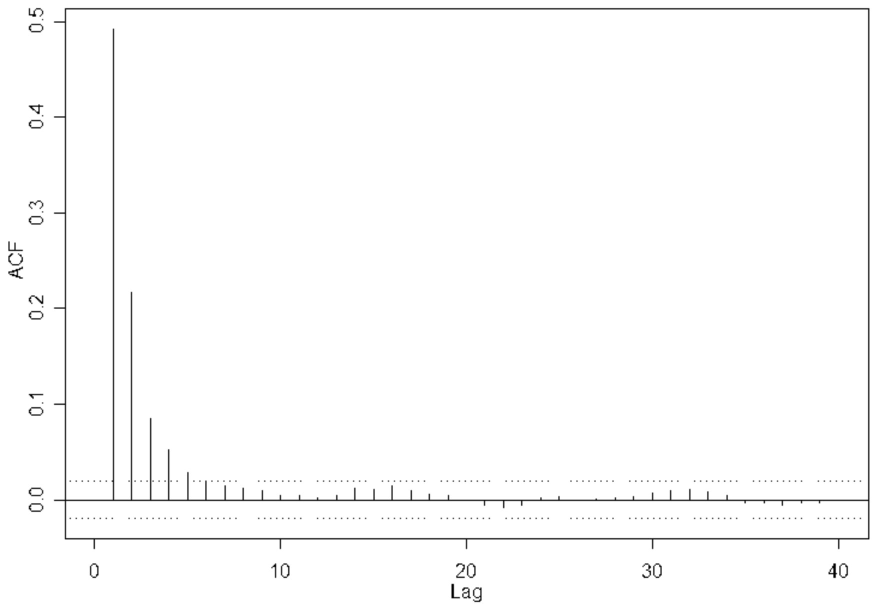

Figure 1. Autocorrelation function of posterir samples of $\boldsymbol{\alpha}_{11}$.

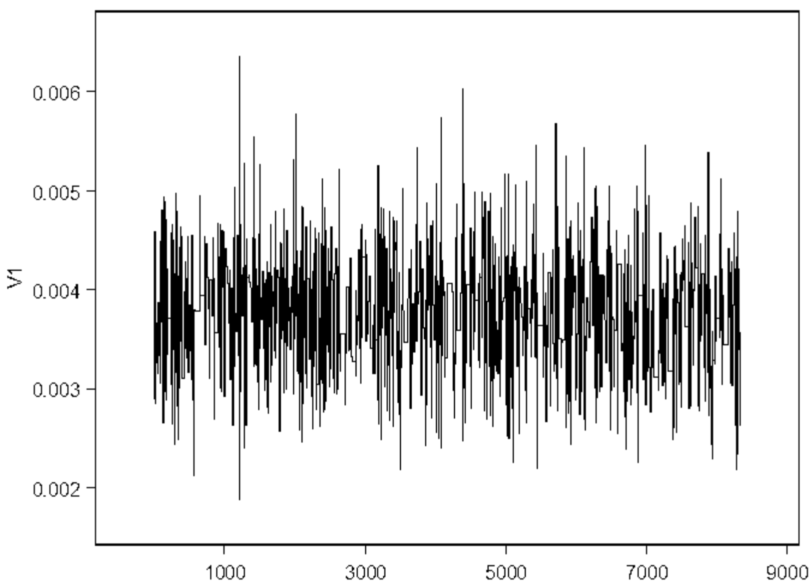

Figure 2. Trace plot of $\boldsymbol{\alpha}_{11}$.

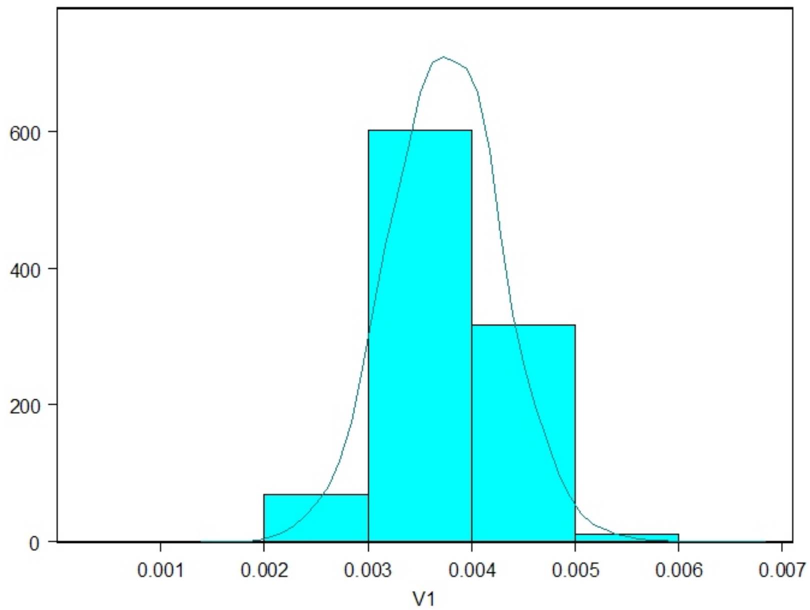

Figure 3. Posterior histogram and density plot for $\boldsymbol{\alpha}_{11}$.

REFERENCES

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

Bollerslev, T., Engle, R. F. and Wooldridge, J. M. (1988). A Capital Asset Pricing Model with Time-varying Covariances. *Journal of Political Economy*, 96, 116-131.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm, *The American Statistician*, 49, 327-335.

Dawid, A. P. (1981). Some matrix-variate distribution theory: notational considerations and a Bayesian application, *Biometrika*, 68, 265-274.

Engle, R. F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, 50, 987-1007.

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.

Gelman, A., Carlin, J. B., Stern, H.S., and Rubin, D. B. (2004). *Bayesian Data Analysis*, Second Edition, Chapman and Hall.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996). Model *Markov Chain Monte Carlo in Practice*, Chapman & Hall.

Harvey, A., Ruiz, E. and Shephard, N. (1994). Multivariate Stochastic Variance Models, *Review of Economic Studies*, 61, 247-264.

Jacquier, E., Polson, N. G. and Rossi, P. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics*, 12, 371-388.

Muller, P. and Pole, A.(1998). Monte Carlo Posterior Integration in GARCH Models. *Sankhya, Series B*, 60, 127-144.

Nelson, D. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59, 347-370.

Press, S. J. (1989). *Bayesian Statistics: Principles, Models, and Applications*. Wiley.

Tsay, R. S. (2002). *Analysis of Financial Time Series*. Wiley.