

Estimation of a Time-varying GQARCH(1,1)-M model

Sofia Anyfantaki and Antonis Demos
Athens University of Economics and Business
E-mail: sanyfantaki@aueb.gr, demos@aueb.gr

Version: April 2009

Time-varying GARCH-M models are commonly used in econometrics and financial economics. Yet the recursive nature of the conditional variance makes exact likelihood analysis of these models computationally infeasible. In this paper we outline the issues and suggest to use a Markov chain Monte Carlo algorithm studied in Fiorentini, Sentana and Sephard (2004) which allows the calculation of a classical estimator via the simulated EM algorithm or a simulated Bayesian solution in only $O(T)$ computational operations, where T is the sample size.

Key Words: dynamic heteroskedasticity, in mean models, time variation, Markov chain Monte Carlo, simulated EM algorithm, Bayesian inference

Subject Classification: [JEL classification] C13;C15;C22

1. INTRODUCTION

Time series data, emerging from diverse fields appear to possess time varying second conditional moments. Furthermore, theoretical results seem to postulate quite oftenly, specific relationships between the second and first conditional moment. For instance, in the stock market context, the first conditional moment of stock market excess returns, given some information set, is a possibly time-varying, linear function of volatility (see eg. Merton, 1980;Glosten et al., 1993). These have led to modifications and extensions of the initial ARCH model of Engle (1982) and it's generalization by Bollerslev (1986), giving rise to a plethora of dynamic heteroscedasticity models. These models have been employed extensively to capture the time variation in the conditional variance of economic series, in general, and of financial time series, in particular (see Bollerslev et al., 1992 for a survey). Although the vast majority of the research in conditional heteroscedasticity is being processed aiming the stylized facts of financial stock returns and of economic time series in general, Arvanitis and Demos (2000), have shown that a family of time varying GARCH-M models can in fact be consistent with the sample characteristics of time series describing the temporal evolution of velocity changes of turbulent fluid and gas molecules. Despite the fact that the latter statistical characteristics match in a considerable degree their financial analogues (for example leptokurtosis, volatility clustering and quasi long range dependence in the squares are common), there are also significant differences in the behaviour of the before mentioned physical systems as opposed to financial markets (examples are the anticorrelation effect and symmetry of velocity changes in contrast to zero autocorrelation and the leverage effect of financial returns) (see Barndorf -Nielsen and Shephard, 2000; as well as Mantegna and Stanley, 1996 and 2000). It was shown that the above mentioned

family of models can even create anticorrelation in the means as far as an AR(1) time varying parameter is introduced.

The analysis of a time-varying GARCH-M model becomes substantially complicated since the log-likelihood of the observed variables can no longer be written in closed form. The main modern way of carrying out likelihood inference in such situations is via a Markov chain Monte Carlo (MCMC) algorithm (see Chib, 2001 for an extensive review). This simulation procedure can be used either to carry out Bayesian inference or to classically estimate the parameters by means of a simulated EM algorithm. Unfortunately, the non-Markovian nature of the GARCH process implies that each time we simulate one error we implicitly change all future conditional variances. As pointed out by Shephard (1996), a regrettable consequence of this path-dependence in volatility is that standard MCMC algorithms will evolve in $O(T^2)$ computational load (see Giakoumatos, Dellaportas and Politis, 1999). Since this cost has to be borne for each parameter value, such procedures are generally infeasible for large financial datasets that we see in practice.

It is clear that from an econometric viewpoint it is important to study how to efficiently estimate models with partially unobserved GARCH processes. In this context, our main contribution is to show how to use the method proposed in Fiorentini, Sentana and Shephard (2004) to achieve MCMC likelihood based estimation of a time-varying GARCH-M model by means of feasible $O(T)$ algorithms. The crucial idea is to transform the GARCH model in a first order Markov model. We prefer to use a GQARCH specification for the conditional variance (Engle, 1990; and Sentana, 1992) since it encompasses all the existing restricted quadratic variance functions (eg. Augmented ARCH model of Bera and Lee, 1990), its properties are very similar to those of GARCH models (eg, stationarity conditions), but avoids some of their criticisms (eg. very easy to generalize to multivariate models). Moreover, many theories in finance involve an explicit trade-off between the risk and the expected returns. For that matter, we use an in mean model which is ideally suited to handling such questions in a time series context where the conditional variance may be time varying. However, a number of studies question the existence of a positive mean/variance ratio directly challenging the mean/variance paradigm. In Glosten, Jagannathan and Runkle (1989) when they explicitly include the nominal risk free rate in the conditioning information set they obtain a negative ARCH-M parameter. For the above, we allow the conditional variance to affect the mean with a possibly time varying coefficient which we assume for simplicity that follows a normal distribution and is independent of the error term. Thus, our model is a Time-Varying GQARCH-M model.

The structure of the paper is as follows. In Section 2 we outline our model. In Section 3 we review both classical and Bayesian likelihood approaches to inference for the time-varying GQARCH-M model. We show that in both cases the key task is to be able to produce simulators from $\{\delta_t\}/\varphi, \{r_t\}, \mathcal{F}_0$ and that the estimation problem arises from the fact that we have two unobserved processes. Section 4 exploits the fact that the method proposed by Fiorentini, Sentana and Shephard (2004) is needed to avoid the huge computational load. An illustrative empirical application for weekly returns from three major stock markets is presented in Section 5. We conclude in Section 6.

2. THE MODEL

The model is a Time-Varying GQARCH(1,1)-M model as described in Chou, Engle and Kane (1992).

$$r_t = \delta_t h_t + \varepsilon_t \tag{1}$$

$$h_t = \omega + \alpha (\varepsilon_{t-1} - \gamma)^2 + \beta h_{t-1} \tag{2}$$

$$\begin{pmatrix} \varepsilon_t \\ \delta_t \end{pmatrix} | \mathcal{F}_{t-1} \sim iidN \left[\begin{pmatrix} 0 \\ \delta_1 \end{pmatrix}, \begin{pmatrix} h_t & 0 \\ 0 & \lambda \end{pmatrix} \right] \tag{3}$$

$$\begin{aligned} \varepsilon_t &= z_t h_t^{1/2} \\ z_t | \mathcal{F}_{t-1} &\sim iidN(0, 1) \end{aligned}$$

where $\{r_t\}_{t=1}^T$ are the observed returns, T is the sample size, $\{\delta_t\}_{t=1}^T$ is an unobserved process independent of $\{\varepsilon_t\}_{t=1}^T$ which follows a normal distribution according to 3 and $\{h_t\}_{t=1}^T$ is the conditional variance which is supposed to follow a GQARCH(1,1). It is obvious that δ_t is the market price of risk (see e.g. Merton, 1980; Glosten et al., 1993). \mathcal{F}_{t-1} is a sequence of natural filtrations generated by the past values of $\{\varepsilon_t\}$ and $\{r_t\}$ at least up to a finite-dimensional vector of unknown parameters φ

3. LIKELIHOOD-INFERENCE: EM AND BAYESIAN APPROACHES

The purpose of the paper is the estimation of a Time-Varying GQARCH-M model. Since our model involves two unobserved components (one from the time-varying in mean parameter which we assume to follow a normal distribution and one from the error term) the estimation method required is an EM and more specifically a simulated EM (SEM) because of the fact that the expectation terms at the E-step cannot be computed. The main way of carrying out likelihood inference in such situations is via a Markov chain Monte Carlo (MCMC) algorithm.

The idea behind the MCMC methods is that in order to sample a given probability distribution, that is referred to as the target distribution, a suitable Markov chain is constructed (using a Metropolis-Hasting (M-H) algorithm or a Gibbs-sampling method) with the property that its limiting, invariant distribution is the target distribution. In most problems, the target distribution is absolutely continuous and as a result the theory of MCMC methods is based on that of Markov chains on continuous state spaces (Mey and Tweedie, 1993). This means that by simulating the Markov chain a large number of times and recording its values a sample of (correlated) draws from the target distribution can be obtained. It should be noted that Markov chain samplers are invariant by construction and therefore the existence of the invariant distribution does not have to be checked in any particular application of MCMC method.

The Metropolis-Hasting algorithm (M-H) is a general MCMC method to produce sample variates from a given multivariate distribution. It is based on a candidate generating density that is used to supply a proposal value that is accepted with probability given as the ratio of the target density times the ratio of the proposal density. There are a number of choices of the proposal density (eg random walk M-H chain, independence M-H chain, tailored M-H chain) and the components may

be revised either in one block or in several blocks. Another MCMC method, which is special case of the multiple block M-H method with acceptance rate always equal to one, is called the Gibbs sampling method and was brought into statistical prominence by Gelfand and Smith (1990). In this algorithm the parameters are grouped into blocks and each block is sampled according to the full conditional distribution denoted as $\pi\left(\varphi_t/\varphi_{/t}\right)$. By Bayes' theorem we have $\pi\left(\varphi_t/\varphi_{/t}\right) \propto \pi\left(\varphi_t\varphi_{/t}\right)$, the joint distribution of all blocks and so full conditional distributions are usually quite simply derived. One cycle of the Gibbs sampling algorithm is completed by simulating $\{\varphi_t\}_{t=1}^p$, where p is the number of blocks, from the full conditional distributions, recursively updating the conditioning variables as one moves through each distribution. Under some general conditions, it is verified that the Markov chain generated by the M-H or the Gibbs sampling algorithm converges to the target density as the number of iterations becomes large.

Within the Bayesian framework MCMC methods have proved very popular and the posterior distribution of the parameters is the target density. Another application of the MCMC is the analysis of hidden Markov models where the approach relies on augmenting the parameter space to include the unobserved states and simulate the target distribution via the conditional distributions. Kim, Shephard and Chib (1998) discuss a MCMC algorithm of the Stochastic Volatility (SV) model which is an example of a state space model in which the state variable h_t (log-volatility) appears non-linearly in the observation equation. The idea is to approximate the model by a conditionally Gaussian state space model with the introduction of multinomial random variables that follow a seven-point discrete distribution.

3.1. Simulated EM algorithm

As mentioned above the estimation problem is that we cannot write down the likelihood function in closed form since we do not observe both ε_t and δ_t . More specifically the conditional log-likelihood function of our model assuming that δ_t were observed would be the following:

$$\begin{aligned}
\ell(r, \delta|\varphi, \mathcal{F}_0) &= \ln p(r|\delta, \varphi, \mathcal{F}_0) + \ln p(\delta|\varphi, \mathcal{F}_0) = & (4) \\
&= -\frac{T}{2} \ln 2\pi - \frac{1}{2} \sum_{t=1}^T \ln h_t - \frac{1}{2} \sum_{t=1}^T \frac{(r_t - \delta_t h_t)^2}{h_t} - \frac{T}{2} \ln 2\pi \\
&\quad - \frac{T}{2} \ln \lambda - \frac{1}{2} \sum_{t=1}^T \frac{(\delta_t - \delta_1)^2}{\lambda} \\
&= -T \ln 2\pi - \frac{1}{2} \sum_{t=1}^T \ln h_t - \frac{1}{2} \sum_{t=1}^T \frac{(\varepsilon_t)^2}{h_t} - \frac{T}{2} \ln \lambda - \frac{1}{2} \sum_{t=1}^T \frac{(\delta_t - \delta_1)^2}{\lambda}
\end{aligned}$$

However, since the δ_t 's are unobserved we rely on an EM algorithm (Dempster, Laird and Rudin, 1977) to obtain estimates as close to the optimum as desired. At each iteration the EM algorithm obtains $\varphi^{(n+1)}$, where φ is the parameter vector, by maximizing the expectation of the log-likelihood conditional on the data and the current parameter values ie $E\left(\ell(\cdot)/r, \varphi^{(n)}, \mathcal{F}_0\right)$ with respect to φ keeping $\varphi^{(n)}$ fixed. The expectation of the complete log-likelihood is:

$$\begin{aligned}
E\left(\ell(\cdot)|r, \varphi^{(n)}, \mathcal{F}_0\right) &= -T \ln 2\pi - \frac{T}{2} \ln \lambda - \frac{1}{2} \sum_{t=1}^T E\left(\ln h_t|r, \varphi^{(n)}, \mathcal{F}_0\right) \\
&\quad - \frac{1}{2} \sum_{t=1}^T E\left(\frac{(\varepsilon_t)^2}{h_t}|r, \varphi^{(n)}, \mathcal{F}_0\right) - \frac{1}{2} \sum_{t=1}^T E\left(\frac{(\delta_t - \delta_1)^2}{\lambda}|r, \varphi^{(n)}, \mathcal{F}_0\right)
\end{aligned} \tag{5}$$

It is obvious that we cannot compute such quantities. For that matter, we may rely on a simulated EM where the expectation terms are replaced by averages over simulations and so we will have a SEM or a simulated score. The SEM log-likelihood is:

$$\begin{aligned}
SEM\ell &= -T \ln 2\pi - \frac{T}{2} \ln \lambda - \frac{1}{2} \frac{1}{M} \sum_{i=1}^M \sum_{t=1}^T \ln h_t^{(i)} - \frac{1}{2} \sum_{i=1}^M \sum_{t=1}^T \frac{(\varepsilon_t^{(i)})^2}{h_t^{(i)}} \\
&\quad - \frac{T}{2} \frac{\delta_1^2}{\lambda} - \frac{1}{2} \frac{1}{\lambda} \frac{1}{M} \sum_{i=1}^M \sum_{t=1}^T (\delta_t^{(i)})^2 + \frac{\delta_1}{\lambda} \frac{1}{M} \sum_{i=1}^M \sum_{t=1}^T \delta_t^{(i)}
\end{aligned} \tag{6}$$

Hence, we need the following quantities:

$$\sum_{t=1}^T E\left(\ln h_t|r, \varphi^{(n)}, \mathcal{F}_0\right) = \frac{1}{M} \sum_{i=1}^M \sum_{t=1}^T \ln h_t^{(i)} \tag{7}$$

$$\sum_{t=1}^T E\left[\frac{(\varepsilon_t)^2}{h_t}|r, \varphi^{(n)}, \mathcal{F}_0\right] = \frac{1}{M} \sum_{i=1}^M \sum_{t=1}^T \frac{(\varepsilon_t^{(i)})^2}{h_t^{(i)}} \tag{8}$$

$$\sum_{t=1}^T E\left[\frac{\delta_t^{(2)}}{\lambda}|r, \varphi^{(n)}, \mathcal{F}_0\right] = \frac{1}{\lambda} \frac{1}{M} \sum_{i=1}^M \sum_{t=1}^T \delta_t^{(i)(2)} \tag{9}$$

$$\sum_{t=1}^T E\left[\frac{\delta_1 \delta_t}{\lambda}|r, \varphi^{(n)}, \mathcal{F}_0\right] = \frac{\delta_1}{\lambda} \frac{1}{M} \sum_{i=1}^M \sum_{t=1}^T \delta_t^{(i)} \tag{10}$$

where M is the number of simulations.

Thus, the basic problem is to sample from $h|\varphi, r, \mathcal{F}_0$ where φ is the vector of the unknown parameters and also sample from $\delta|\varphi, r, \mathcal{F}_0$ where $r = (r_1, \dots, r_t)$. However, for a given set of parameter values and initial conditions it is generally simpler to simulate $\{\varepsilon_t\}$ for $t = 1, \dots, T$ and then compute $\{\delta_t\}_{t=1}^T$ than to simulate $\{\delta_t\}_{t=1}^T$ directly. For that matter, we concentrate on simulators of ε_t given r and φ . We set the mean and the variance of ε_0 equal to their unconditional values and given that h_t is a sufficient statistic for \mathcal{F}_{t-1} and the unconditional variance is a deterministic function of φ , \mathcal{F}_0 can be eliminated from the information set without any information loss.

Now sampling from:

$$p(\varepsilon|r, \varphi) \propto p(r|\varepsilon, \varphi) p(\varepsilon|\varphi)$$

is feasible by using a M-H algorithm where we update each time only one ε_t leaving all the other unchanged (Shephard, 1996). In particular, let us write the r th

iteration of a Markov chain as ε^r . Then we generate a potential new value of the Markov chain ε^{new} by proposing from some candidate density $g(\varepsilon_t | \varepsilon_{\setminus t}^r, r, \varphi)$ where $\varepsilon_{\setminus t}^r = \{\varepsilon_1^{r+1}, \dots, \varepsilon_{t-1}^{r+1}, \varepsilon_{t+1}^r, \dots, \varepsilon_T^r\}$ which we accept with probability:

$$\min \left[1, \frac{p(\varepsilon_t^{new} | \varepsilon_{\setminus t}^r, r, \varphi) g(\varepsilon_t^{new} | \varepsilon_{\setminus t}^r, r, \varphi)}{p(\varepsilon_t^r | \varepsilon_{\setminus t}^r, r, \varphi) g(\varepsilon_t^r | \varepsilon_{\setminus t}^r, r, \varphi)} \right]$$

If it is accepted then we set $\varepsilon_t^{r+1} = \varepsilon_t^{new}$ and otherwise we keep $\varepsilon_t^{r+1} = \varepsilon_t^r$. Although the proposal is much better since it is only in a single dimension, each time we consider modifying a single error we have to compute:

$$\begin{aligned} \frac{p(\varepsilon_t^{new} | \varepsilon_{\setminus t}^r, r, \varphi)}{p(\varepsilon_t^r | \varepsilon_{\setminus t}^r, r, \varphi)} &= \frac{p(r_t | \varepsilon_t^{new}, h_t^{new,t}, \varphi) p(\varepsilon_t^{new} | h_t^{new,t}, \varphi) p(r_t | h_t^{r,t}, \varphi)}{p(r_t | h_t^{new,t}, \varphi) p(r_t | \varepsilon_t^r, h_t^{r,t}, \varphi) p(\varepsilon_t^{new} | h_t^{r,t}, \varphi)} \\ &* \prod_{s=t+1}^T \frac{p(r_s | \varepsilon_s^r, h_s^{new,t}, \varphi) p(\varepsilon_s^r | h_s^{new,t}, \varphi) p(r_s | h_s^{r,t}, \varphi)}{p(r_s | h_s^{new,t}, \varphi) p(r_s | \varepsilon_s^r, h_s^{r,t}, \varphi) p(\varepsilon_s^r | h_s^{r,t}, \varphi)} \\ &= \frac{p(r_t | \varepsilon_t^{new}, h_t^{new,t}, \varphi) p(\varepsilon_t^{new} | h_t^{new,t}, \varphi)}{p(r_t | \varepsilon_t^r, h_t^{r,t}, \varphi) p(\varepsilon_t^{new} | h_t^{r,t}, \varphi)} \\ &* \prod_{s=t+1}^T \frac{p(r_s | \varepsilon_s^r, h_s^{new,t}, \varphi) p(\varepsilon_s^r | h_s^{new,t}, \varphi)}{p(r_s | \varepsilon_s^r, h_s^{r,t}, \varphi) p(\varepsilon_s^r | h_s^{r,t}, \varphi)} \end{aligned}$$

where for $s = t + 1, \dots, T$

$$h_s^{new,t} = V(\varepsilon_s | \varepsilon_{s-1}^r, \varepsilon_{s-2}^r, \dots, \varepsilon_{t+1}^r, \varepsilon_t^{new}, \varepsilon_{t-1}^{r+1}, \dots, \varepsilon_1^{r+1})$$

$$h_s^{r,t} = V(\varepsilon_s | \varepsilon_{s-1}^r, \varepsilon_{s-2}^r, \dots, \varepsilon_{t+1}^r, \varepsilon_t^r, \varepsilon_{t-1}^{r+1}, \dots, \varepsilon_1^{r+1})$$

while

$$h_t^{new,t} = h_t^{r,t}$$

Nevertheless, each time we revise one ε_t we have also to revise $T - t$ conditional variances because of the recursive nature of the GARCH model which makes $h_s^{new,t}$ depend upon ε_t^{new} for $s = t + 1, \dots, T$. And since $t = 1, \dots, T$ it is obvious that we need to calculate T^2 normal densities and so this algorithm is $O(T^2)$. And this should be done for every φ . To avoid this huge computational load we show how to use the method proposed by Fiorentini, Sentana and Shephard (2004) and so do MCMC with only $O(T)$ calculations.

3.2. Simulation-Based Bayesian Inference

The task of simulating from $\{\delta_t\} | \{r_t\}, \varphi$ also appears in the Bayesian analysis of the model. In particular we recall that in our problem the key issue is that the likelihood function of the sample $p(r | \varphi, \mathcal{F}_0)$ is intractable which precludes the direct analysis of the posterior density $p(\varphi | r, \mathcal{F}_0)$. This problem may be overcome by focusing instead on the posterior density of the model using Bayes' rule:

$$p(\varphi, \delta/r) \propto p(\varphi, \delta) p(r/\varphi, \delta) \propto p(\varphi) p(\delta/\varphi) p(r/\varphi, \delta)$$

where

$$\varphi = (\delta_1, \lambda, \alpha, \beta, \gamma, \omega)' \quad (11)$$

$$\delta = (\delta_1, \delta_2, \dots, \delta_T)'$$

$$r = (r_1, r_2, \dots, r_T)'$$

Assuming independent priors we have:

$$p(\varphi) = p(\delta_1) p(\lambda) p(\alpha) p(\beta) p(\gamma) p(\omega)$$

Also

$$p(\delta/\varphi) = \prod_{t=1}^T p(\delta_t / \{\delta_{t-1}\}, \delta_1, \lambda) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi\lambda}} \exp\left(-\frac{(\delta_t - \delta_1)^2}{2\lambda}\right)$$

and $p(r/\varphi, \delta)$ is the full information likelihood

$$p(r/\varphi, \delta) = \prod_{t=1}^T p(r_t / \{r_{t-1}\}, \delta_1, \delta_2, \dots, \delta_t, \varphi) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi}h_t} \exp\left(-\frac{\varepsilon_t^2}{2h_t}\right)$$

Once we have posterior density function we get marginal posterior density function of parameters by integrating the posterior density function. MCMC is one way of numerical integration. Clifford-Hammersley theorem says that a joint distribution can be characterized by its complete conditional distribution. Given initial values $\{\delta_t\}^{(0)}, \varphi^{(0)}$ we draw $\{\delta_t\}^{(1)}$ from $p(\{\delta_t\}^{(1)} / r, \varphi^{(0)})$ and then $\varphi^{(1)}$ from $p(\varphi^{(1)} / \{\delta_t\}^{(1)}, r)$. Iterating these steps we finally get $(\{\delta_t\}^{(i)}, \varphi^{(i)})_{i=1}^M$ and under mild conditions it is shown that the distribution of the sequence converges to the joint posterior distribution $p(\varphi, \delta/r)$.

The above may be carried out in some blocks. Divide parameters into two categories:

$$\varphi_1 = (\delta_1, \lambda) \quad (12)$$

$$\varphi_2 = (\alpha, \beta, \gamma, \omega) \quad (13)$$

- (1) Initialize φ
- (2) Draw from $p(\delta_t / \delta_{\neq t}, r, \varphi)$
- (3) Draw from $p(\varphi, \delta, r)$ in the following blocks:
 - (i) Draw from $p(\varphi_1 / \delta, r)$ using Gibbs sampling. This is update in one block.
 - (ii) Draw from $p(\varphi_2 / r)$ by M-H. This is updated in a second block.
- (4) Go to (2)

We review the implementation of each step.

3.2.1. Gibbs-sampling

We have that $\delta_t \sim N(\delta_1, \lambda)$. For reasons that will become clear as we proceed, it is convenient to work in terms of τ , the reciprocal of the variance (known as the precision parameter).

Following Poirier(1995) we assume that the joint prior for $\varphi_1 = [\delta_1, \tau]'$ is the normal Gamma distribution denoted by $\varphi_1 \sim NG(\underline{\mu}, \underline{q}, \underline{s}^{-2}, \underline{v})$ with density:

$$\begin{aligned} f_{NG}(\varphi_1/\underline{\mu}, \underline{q}, \underline{s}^{-2}, \underline{v}) &= \phi(\delta_1/\underline{\mu}, \tau^{-1}\underline{q}) \gamma(\tau/\underline{s}^{-2}, \underline{v}) & (14) \\ &\propto \left\{ \tau^{1/2} \exp\left[-\frac{1}{2}\tau\underline{q}^{-1}(\delta_1 - \underline{\mu})^2\right] \right\} \left\{ \tau^{1/2(\underline{v}-2)} \exp\left(-\frac{1}{2}\tau\underline{v}\underline{s}^2\right) \right\} \\ &\propto \tau^{1/2(\underline{v}-1)} \exp\left[-\frac{1}{2}\tau\left\{\underline{v}\underline{s}^2 + \underline{q}^{-1}(\delta_1 - \underline{\mu})^2\right\}\right] \end{aligned}$$

where $\underline{\mu} \in \Re$, \underline{q} , \underline{s}^2 , \underline{v} are known positive constants. The normal density represents the prior conditional density of δ_1 given τ and the gamma density represents the prior marginal density of τ with mean $E(\tau) = \underline{s}^{-2}$ and variance $Var(\tau) = \frac{2\underline{s}^{-4}}{\underline{v}}$. Using change-of-variable techniques it is seen that the distribution of $\lambda \equiv \tau^{-1}$ is the inverted gamma distribution with $E(\lambda) = \underline{v}\underline{s}^2/(\underline{v}-2)$ if $\underline{v} > 2$. The prior marginal distribution of δ_1 is the $t(\underline{\mu}, \underline{s}^2\underline{q}, \underline{v})$ with mean $E(\delta_1) = \underline{\mu}$ if $\underline{v} > 1$ and variance $Var(\delta_1) = \underline{v}\underline{s}^4\underline{q}^2/(\underline{v}-2)$ if $\underline{v} > 2$ (see for example Poirier, 1995).

Then it is known that the posterior density corresponds to the kernel of a $NG(\bar{\mu}, \bar{q}, \bar{s}^{-2}, \bar{v})$ where:

$$\begin{aligned} \bar{q} &\equiv (\underline{q}^{-1} + T)^{-1} \\ \bar{\mu} &\equiv \bar{q}(\underline{q}^{-1}\underline{\mu} + T\bar{\delta}) \\ \bar{v} &\equiv \underline{v} + T \\ \bar{s}^2 &\equiv \bar{v}^{-1} \left[\underline{v}\underline{s}^2 + \underline{v}\underline{s}^2 + \underline{q}^{-1}\bar{q}T(\bar{\delta} - \underline{\mu})^2 \right] \\ \bar{v} &\equiv T - 1 \\ \bar{s}^2 &\equiv \bar{v}^{-1} \sum_{t=1}^T (\delta_t - \bar{\delta})^2 \end{aligned}$$

Because both the prior and the posterior are proportional to normal-gamma densities, the prior density is the natural conjugate prior when drawing a random sample from normal population with unknown mean and variance. Posterior moments for δ_1, τ can be obtained from prior moments simply by replacing $\underline{\mu}, \underline{q}, \underline{s}^{-2}, \underline{v}$ with $\bar{\mu}, \bar{q}, \bar{s}^{-2}, \bar{v}$.

3.2.2. Metropolis-Hasting

Step(3)(ii) is the task of simulating from the posterior of the parameters of a GQARCH-M process. This has been already addressed by Kim, Shephard and Chib(1998), Bauwens and Lubrano(1998), Nakatsuma(2000), Ardia (2008) and others.

For the parameters $\alpha, \beta, \gamma, \omega$ we use Normal densities as priors:

$$\begin{aligned} p(\boldsymbol{\alpha}) &\sim N(\mu_{\alpha}, \Sigma_{\alpha}) I_{\alpha} \\ p(\beta) &\sim N(\mu_{\beta}, \sigma_{\beta}^2) I_{\beta} \\ p(\gamma) &\sim N(\mu_{\gamma}, \sigma_{\gamma}^2) \end{aligned}$$

where $\boldsymbol{\alpha} = (\omega, \alpha)'$, I_{α}, I_{β} are the indicators ensuring the constraints $\alpha > 0$ and $\beta > 0$ respectively. μ, σ^2 are the hyperparameters.

We form the joint prior by assuming prior independence between α, β, γ ie:

$$p(\boldsymbol{\alpha}, \beta, \gamma) = N(\mu_{\alpha}, \Sigma_{\alpha}) I_{\alpha} \times N(\mu_{\beta}, \sigma_{\beta}^2) I_{\beta} \times N(\mu_{\gamma}, \sigma_{\gamma}^2)$$

The joint posterior is then obtained by combining the joint prior and likelihood function by Bayes' rule:

$$p(\boldsymbol{\alpha}, \beta, \gamma, /Y) \propto \prod_{t=1}^T \frac{1}{\sqrt{2\pi h_t}} \exp\left(-\frac{\varepsilon_t^2}{2h_t}\right) \times N(\mu_{\alpha}, \Sigma_{\alpha}) I_{\alpha} \times N(\mu_{\beta}, \sigma_{\beta}^2) I_{\beta} \times N(\mu_{\gamma}, \sigma_{\gamma}^2) \quad (15)$$

For the M-H algorithm we use the following approximated GARCH model as in Nakatsuma (2000) which is derived by the well known property of GARCH models (Bollerslev, 1986):

$$w_t = \varepsilon_t^2 - \omega - \alpha(\varepsilon_{t-1} - \gamma)^2 + \beta w_{t-1} - \beta \varepsilon_{t-1}^2 \quad (16)$$

Then the corresponding approximated likelihood is written as:

$$p(\varepsilon^2/r, \delta, \varphi_2) = \prod_{t=1}^T \frac{1}{2h_t\sqrt{\pi}} \exp\left[-\frac{w_t^2}{4h_t^2}\right] \quad (17)$$

The generation of α, β, γ is based on the above likelihood where we update $\{h_t\}$ each time after the corresponding parameters are updated

Generation of α We first note that w_t in 12 can be written as a linear function of $\boldsymbol{\alpha}$:

$$w_t = \bar{\varepsilon}_t^2 - \zeta_t \boldsymbol{\alpha} \quad (18)$$

where $\zeta_t = [\tilde{\iota}_t, \hat{\varepsilon}_t^2]$ with

$$\begin{aligned} \bar{\varepsilon}_t^2 &= \tilde{\varepsilon}_t^2 - \beta \tilde{\varepsilon}_{t-1}^2 \\ \tilde{\varepsilon}_t^2 &= \varepsilon_t^2 + \beta \varepsilon_{t-1}^2 \\ \hat{\varepsilon}_t^2 &= (\varepsilon_{t-1} - \gamma)^2 + \beta \hat{\varepsilon}_{t-1}^2 \\ \tilde{\iota}_t &= 1 + \beta \tilde{\iota}_{t-1} \end{aligned}$$

Let

$$\begin{aligned} Y_{\boldsymbol{\alpha}} &= [\bar{\varepsilon}_1^2, \dots, \bar{\varepsilon}_T^2]' \\ X_{\boldsymbol{\alpha}} &= [\zeta_1', \dots, \zeta_T']' \end{aligned}$$

Then the likelihood function of the approximated model is rewritten as:

$$p(\varepsilon^2/r, \delta, \varphi_2) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi(2h_t^2)}} \exp\left[-\frac{(\varepsilon_t^2 - \zeta_t \boldsymbol{\alpha})^2}{2(2h_t^2)}\right] \quad (19)$$

And so using this we have the following proposal distribution of $\boldsymbol{\alpha}$:

$$\boldsymbol{\alpha}/Y, X, \Sigma, \varphi_{2-\boldsymbol{\alpha}} \sim N\left(\widehat{\boldsymbol{\mu}}_{\boldsymbol{\alpha}}, \widehat{\Sigma}_{\boldsymbol{\alpha}}\right) I_{\boldsymbol{\alpha}} \quad (20)$$

where $\widehat{\boldsymbol{\mu}}_{\boldsymbol{\alpha}} = \widehat{\Sigma}_{\boldsymbol{\alpha}} (X'_{\boldsymbol{\alpha}} \Lambda^{-1} Y_{\boldsymbol{\alpha}} + \Sigma_{\boldsymbol{\alpha}}^{-1} \boldsymbol{\mu}_{\boldsymbol{\alpha}})$, $\widehat{\Sigma}_{\boldsymbol{\alpha}} = (X'_{\boldsymbol{\alpha}} \Lambda^{-1} X_{\boldsymbol{\alpha}} + \Sigma_{\boldsymbol{\alpha}}^{-1})^{-1}$ and $\Lambda = \text{diag}(2h_1^2, \dots, 2h_T^2)$. $I_{\boldsymbol{\alpha}}$ imposes the restriction that $\boldsymbol{\alpha} > 0$

Hence a candidate $\tilde{\boldsymbol{\alpha}}$ is sampled from this proposal density and accepted with probability:

$$\min \left\{ \frac{p(\tilde{\boldsymbol{\alpha}}, \beta, \gamma, \delta, r) q(\boldsymbol{\alpha}^*/\tilde{\boldsymbol{\alpha}}, \beta, \gamma, \delta, r)}{p(\boldsymbol{\alpha}^*, \beta, \gamma, \delta, r) q(\tilde{\boldsymbol{\alpha}}/\boldsymbol{\alpha}^*, \beta, \gamma, \delta, r)}, 1 \right\}$$

Generation of β Following Nakatsuma (2000) we linearize w_t by the first-order Taylor expansion

$$w_t(\beta) \approx w_t(\beta^*) + \xi_t(\beta^*)(\beta - \beta^*) \quad (21)$$

where ξ_t is the first-order derivative of $w_t(\beta)$ evaluated at β^* the previous draw of the M-H sampler.

Define as

$$r_t = w_t(\beta^*) + g_t(\beta^*)\beta^*$$

where $g_t(\beta^*) = -\xi_t(\beta^*)$ which is computed by the recursion:

$$g_t = \varepsilon_{t-1}^2 - w_{t-1} + \beta^* g_{t-1}$$

$\xi_t = 0$ for $t \leq 0$ (Ardia 2008). Then:

$$w_t(\beta) \approx r_t - g_t(\beta)\beta$$

Let

$$\begin{aligned} Y_{\beta} &= [r_1, \dots, r_T]' \\ X_{\beta} &= [g_1, \dots, g_T]' \end{aligned}$$

The likelihood function of the approximated model is rewritten as:

$$p(\varepsilon^2/r, \delta, \varphi_2) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi(2h_t^2)}} \exp\left[-\frac{\{w_t(\beta^*) + \xi_t(\beta^*)(\beta - \beta^*)\}^2}{2(2h_t^2)}\right] \quad (22)$$

then we have the following proposal distribution for β

$$\beta/Y, X, \sigma_{\beta}^2, \varphi_{2-\beta} \sim N\left(\widehat{\boldsymbol{\mu}}_{\beta}, \widehat{\sigma}_{\beta}^2\right) I_{\beta} \quad (23)$$

where $\widehat{\boldsymbol{\mu}}_{\beta} = \widehat{\sigma}_{\beta}^2 (X'_{\beta} \Lambda^{-1} Y_{\beta} + \frac{\mu_{\beta}}{\sigma_{\beta}^2})$, $\widehat{\sigma}_{\beta}^2 = (X'_{\beta} \Lambda^{-1} X_{\beta} + \frac{1}{\sigma_{\beta}^2})^{-1}$ and $\Lambda = \text{diag}(2h_1^4, \dots, 2h_T^4)$. I_{β} imposes the restriction that $\beta > 0$

Hence a candidate $\tilde{\beta}$ is sampled from this proposal density and accepted with probability:

$$\min \left\{ \frac{p(\tilde{\beta}, \alpha, \gamma, \delta, \varphi_1/r) q(\beta^*/\tilde{\beta}, \alpha, \gamma, \delta, \varphi_1, r)}{p(\beta^*, \alpha, \gamma, \delta, \varphi_1/r) q(\tilde{\beta}/\beta^*, \alpha, \gamma, \delta, \varphi_1, r)}, 1 \right\}$$

Generation of γ As with β we linearize w_t by a first-order Taylor expansion at a point γ^* the previous draw in the M-H sampler. In this case

$$r_t = w_t(\gamma^*) - g_t(\gamma^*)\gamma^* \quad (24)$$

where $g_t(\gamma^*) = -\xi_t(\gamma^*)$ which is computed by the recursion:

$$g_t = -2\alpha(\varepsilon_{t-1} - \gamma^*) - \beta g_{t-1}$$

and $g_t = 0$ for $t \leq 0$.

Then:

$$w_t(\gamma) \approx r_t - g_t\gamma$$

Let

$$\begin{aligned} Y_\gamma &= [r_1, \dots, r_T]' \\ X_\gamma &= [g_1, \dots, g_T]' \end{aligned}$$

The likelihood function of the approximated model is rewritted as:

$$p(\varepsilon^2/r, \delta, \varphi_2) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi}(2h_t^2)} \exp \left[-\frac{\{w_t(\gamma^*) - g_t(\gamma^*)\gamma^*\}^2}{2(2h_t^2)} \right] \quad (25)$$

then we have the following proposal distribution for γ

$$\gamma/Y, X, \sigma_\gamma^2, \varphi_{2-\gamma} \sim N(\hat{\mu}_\gamma, \hat{\sigma}_\gamma^2) \quad (26)$$

where $\hat{\mu}_\gamma = \hat{\sigma}_\gamma^2 \left(X_\gamma' \Lambda^{-1} Y_\gamma + \frac{\mu_\gamma}{\sigma_\gamma^2} \right)$, $\hat{\sigma}_\gamma^2 = \left(X_\gamma' \Lambda^{-1} X_\gamma + \frac{1}{\sigma_\gamma^2} \right)^{-1}$ and $\Lambda = \text{diag}(2h_1^4, \dots, 2h_T^4)$

Hence a candidate $\tilde{\gamma}$ is sampled from this proposal density and accepted with probability:

$$\min \left\{ \frac{p(\tilde{\gamma}, \alpha, \beta, \delta, \Theta_1/Y) q(\gamma^*/\tilde{\gamma}, \alpha, \beta, \delta, \varphi_1, r)}{p(\gamma^*, \alpha, \beta, \delta, \varphi_1/Y) q(\tilde{\gamma}/\gamma^*, \alpha, \beta, \delta, \varphi_1, r)}, 1 \right\}$$

The algorithm described above is a special case of a MCMC algorithm, which converges as it iterates, to draws from the required density $p(\varphi, \delta/r)$. Posterior moments and marginal densities can be estimated (simulation consistently) by averaging the releveant function of interest over the sample variates. The posterior mean of φ is simply estimated by the sample mean of the simulated φ values. These estimated can be made arbitrarily accurate by increasing the simulation sample size. However, it should be remembered that sample variates from a MCMC algorithm are a high dimensional (correlated) sample from the target density and sometimes the serial correlation can be quite high for badly behaved algoritihms.

All that remains therefore is Step (ii). Thus, from the above it is seen that whether we follow a classical or a Bayesian approach to estimation, the main task is to simulate from $\delta/\varphi, r, \mathcal{F}_0$.

4. MCMC SIMULATION OF $\delta/\varphi, R, \mathcal{F}_0$

The method proposed by Fiorentini, Sentana and Shephard (2004) is to transform the GARCH model into a first order Markov model and so do MCMC with only $O(T)$ calculations.

Following their transformation we augment the state vector with the variables h_{t+1} and then sample the joint Markov process $\{h_{t+1}, s_t\} | r, \varphi \in \mathcal{F}_t$ where

$$s_t = \text{sign}(\varepsilon_t - \gamma) \quad (27)$$

so that $s_t = \pm 1$ with probability one. The mapping is one-to-one and has no singularities. More specifically if we know $\{h_{t+1}\}$ and φ then we know the value of

$$(\varepsilon_t - \gamma)^2 = \frac{h_{t+1} - \omega - \beta h_t}{\alpha} \quad \forall t \geq 1$$

Hence the additional knowledge of the signs of $(\varepsilon_t - \gamma)$ would reveal the entire path of $\{\varepsilon_t\}$ so long as h_0 (which equals the unconditional value in our case) is known and thus we may now reveal also the unobserved random variable $\{\delta_t\} / r, \varphi, \{h_{t+1}\}$.

Now we have to sample from:

$$p(\{s_t, h_{t+1}\} / r, \varphi) \propto \prod_{t=1}^T p(s_t / h_{t+1}, h_t, \varphi) p(h_{t+1} / h_t, \varphi) p(r_t / s_t, h_t, h_{t+1}, \varphi) \quad (28)$$

where the second and the third term come from the model and the first comes from the fact that $\varepsilon_t / \mathcal{F}_{t-1} \sim N(0, h_t)$ but $\varepsilon_t / \{h_{t+1}\}, \mathcal{F}_{t-1}$ takes values:

$$\begin{aligned} h_{t+1} &= \omega + \alpha(\varepsilon_t - \gamma)^2 + \beta h_t \Rightarrow \\ (\varepsilon_t - \gamma)^2 &= \frac{h_{t+1} - \omega - \beta h_t}{\alpha} \Rightarrow \\ \varepsilon_t &= \gamma \pm \sqrt{\frac{h_{t+1} - \omega - \beta h_t}{\alpha}} \end{aligned}$$

$$\varepsilon_t = \gamma \pm d_t$$

where

$$d_t = \sqrt{\frac{h_{t+1} - \omega - \beta h_t}{\alpha}} \quad (29)$$

From the above it is seen that we should first simulate $\{h_{t+1}\} / r, \varphi$ since we do not alter the volatility process when we flip from $s_t = -1$ to $s_t = 1$ (implying that the signs do not cause the volatility process) but we do alter ε_t and then simulate $\{s_t\} / \{h_{t+1}\}, r, \varphi$. The second step is a Gibbs sampling scheme whose acceptance rate is always one and also conditional on $\{h_{t+1}\}, r, \varphi$ the elements of $\{s_t\}$ are independent which further simplifies the calculations.

4.1. Simulations of $\{s_t\} / \{h_{t+1}\}, r, \varphi$

First, we see how to sample from $\{s_t\} / \{h_{t+1}\}, r, \varphi$. To obtain the required conditionally Bernoulli distribution we establish first some notation. We have the following:

$$c_t = \frac{1}{\sqrt{v_{t|r_t, h_t}}} \left[\varphi \left(\frac{\gamma + d_t - \varepsilon_{t|r_t, h_t}}{\sqrt{v_{t|r_t, h_t}}} \right) + \varphi \left(\frac{\gamma - d_t - \varepsilon_{t|r_t, h_t}}{\sqrt{v_{t|r_t, h_t}}} \right) \right] \quad (30)$$

where

$$\varepsilon_{t|r_t, h_t} = E(\varepsilon_t | r_t, h_t) = \frac{r_t - \delta_1 h_t}{\lambda h_t + 1}, \quad v_{t|r_t, h_t} = \text{Var}(\varepsilon_t | r_t, h_t) = \frac{\lambda h_t^2}{(\lambda h_t + 1)} \quad (31)$$

Using the above notation we see that the probability of drawing $s_t = 1$ conditional on $\{h_{t+1}\}$ is equal to the probability of drawing $\varepsilon_t = \gamma + d_t$ conditional on $h_{t+1}, h_t, r_t, \varphi$, where d_t is given by 29, which is given by:

$$\begin{aligned} p(s_t = s_1 / \{h_{t+1}\}, r, \varphi) &= p(\varepsilon_t = \gamma + d_t / h_{t+1}, h_t, r_t, \varphi) = \\ &= \frac{\varphi \left(\frac{\gamma + d_t - \varepsilon_{t/r_t, h_t}}{\sqrt{v_{t/r_t, h_t}}} \right)}{\varphi \left(\frac{\gamma + d_t - \varepsilon_{t/r_t, h_t}}{\sqrt{v_{t/r_t, h_t}}} \right) + \varphi \left(\frac{\gamma - d_t - \varepsilon_{t/r_t, h_t}}{\sqrt{v_{t/r_t, h_t}}} \right)} = \\ &= \frac{1}{c_t \sqrt{v_{t/r_t, h_t}}} \varphi \left(\frac{\gamma + d_t - \varepsilon_{t/r_t, h_t}}{\sqrt{v_{t/r_t, h_t}}} \right) \end{aligned}$$

Similarly for the probability of drawing $s_t = -1$. Both these quantities are easy to compute eg.

$$\varphi \left(\frac{\gamma + d_t - \varepsilon_{t/r_t, h_t}}{\sqrt{v_{t/r_t, h_t}}} \right) = \frac{1}{\sqrt{2\pi}} \exp \left\{ - \left(\frac{\gamma + d_t - \varepsilon_{t/r_t, h_t}}{\sqrt{v_{t/r_t, h_t}}} \right)^2 \right\}$$

and so we may simulate $\{s_t\} / \{h_{t+1}\}, r, \varphi$ using a Gibbs sampling scheme. More specifically, since conditional on $\{h_{t+1}\}, r, \varphi$ the elements of $\{s_t\}$ are independent we actually draw from the marginal distribution and the acceptance rate for this algorithm is always one.

The Gibbs sampling algorithm for drawing $\{s_t\} / \{h_{t+1}\}, r, \varphi$ may be described as below:

- (1) Specify an initial value $s^{(0)} = (s_1^{(0)}, \dots, s_T^{(0)})$
- (2) Repeat for $r=1, \dots, M$
 - (a) Repeat for $t=0, \dots, T-1$
 - (i) Draw $s^{(r)} = 1$ with probability $\frac{1}{c_t \sqrt{v_{t/r_t, h_t}}} \varphi \left(\frac{\gamma + d_t - \varepsilon_{t/r_t, h_t}}{\sqrt{v_{t/r_t, h_t}}} \right)$ and $s^{(r)} = -1$ with probability $1 - \frac{1}{c_t \sqrt{v_{t/r_t, h_t}}} \varphi \left(\frac{\gamma + d_t - \varepsilon_{t/r_t, h_t}}{\sqrt{v_{t/r_t, h_t}}} \right)$
- (3) Return the values $\{s^{(1)}, \dots, s^{(M)}\}$

4.2. Simulations of $\{h_{t+1}\} / r, \varphi$ (single move samplers)

On the other hand, the first step involves simulating from $\{h_{t+1}\} / r, \varphi$. To avoid large dependence in the chain we use a M-H algorithm where we simulate one h_{t+1} at a time leaving the others unchanged (Shephard (1996) and Wei

(2002)). So if $(h_{t+1})^r$ is the current value of the r th iteration of a Markov chain then we draw a candidate value of the Markov chain h_{t+1}^{new} by proposing it from a candidate density (proposal density) $g\left(h_{t+1}/(h)_{/t+1}^r, r, \varphi\right)$ where $(h)_{/t+1}^r = \{h_1^{r+1}, h_2^{r+1}, \dots, h_t^{r+1}, h_{t+2}^r, \dots, h_{T+1}^r\}$. We set $(h_{t+1})^{r+1} = (h_{t+1})^{new}$ with acceptance probability

$$\min \left[1, \frac{p\left(h_{t+1}^{new}/(h)_{/t+1}^r, r, \varphi\right) g\left(h_{t+1}^r/(h)_{/t+1}^r, r, \varphi\right)}{p\left(h_{t+1}^r/(h)_{/t+1}^r, r, \varphi\right) g\left(h_{t+1}^{new}/(h)_{/t+1}^r, r, \varphi\right)} \right]$$

where we have used the fact that:

$$p(h/r, \varphi) = p\left((h)_{/t}/r, \varphi\right) p\left(h_t/(h)_{/t}, r, \varphi\right)$$

However, we may simplify further the acceptance rate. More specifically, we have that:

$$p\left(h_{t+1}/(h)_{/t+1}, r, \varphi\right) \propto p(h_{t+2}/h_{t+1}, \varphi) p(h_{t+1}/h_t, \varphi) \\ p(r_{t+1}/h_{t+2}, h_{t+1}, \varphi) p(r_t/h_{t+1}, h_t, \varphi)$$

Now, since the following should hold:

$$h_{t+1} = \omega + \alpha(\varepsilon_t - \gamma)^2 + \beta h_t \Rightarrow \\ h_{t+1} \geq \omega + \beta h_t$$

and similarly

$$h_{t+2} \geq \omega + \beta h_{t+1} \Rightarrow \\ h_{t+1} \leq \beta^{-1}(h_{t+2} - \omega)$$

we have that the support of the conditional distribution of h_{t+1} given h_t is bounded from below by $\omega + \beta h_t$ and the same applies to the distribution of h_{t+2} given h_{t+1} (lower limit corresponds to $d_t = 0$ and the upper limit to $d_{t+1} = 0$). This means that the range of values of h_{t+1} compatible with h_t and h_{t+2} in the GQARCH case is bounded from above and below ie:

$$h_{t+1} \in [\omega + \beta h_t, \beta^{-1}(h_{t+2} - \omega)] \quad (32)$$

From the above we understand that it makes sense to make the proposal to obey the support of the density and so it is seen that we can simplify the acceptance rate by setting:

$$g\left(h_{t+1}/(h)_{/t+1}, r, \varphi\right) = p(h_{t+1}/h_t, \varphi)$$

appropriately truncated from above (since the truncation from below will automatically be satisfied). But the above proposal density ignores the information contained in r_{t+1} and so according to Fiorentini, Sentana and Shephard (2004) we can achieve a substantially higher acceptance rate if we propose from:

$$g\left(h_{t+1}/(h)_{/t+1}, r, \varphi\right) = p(h_{t+1}/r_t, h_t, \varphi)$$

An numerically efficient way to simulate h_{t+1} from $p(h_{t+1}/r_t, h_t, \varphi)$ is to sample an underlying Gaussian random variable doubly truncated by using an inverse transform method. More specifically, we may draw

$$\varepsilon_t/r_t, h_t, \varphi \sim N\left(\frac{r_t - \delta_1 h_t}{\lambda h_t + 1}, \frac{\lambda h_t^2}{(\lambda h_t + 1)}\right) \quad (33)$$

doubly truncated so that it remains within the following bounds:

$$\begin{aligned} (\varepsilon_t - \gamma)^2 &= \frac{h_{t+1} - \omega - \beta h_t}{\alpha} = \\ &= \frac{\beta^{-1}(h_{t+2} - \omega) - \omega - \beta h_t}{\alpha} = \\ &= \frac{h_{t+2} - \omega - \beta\omega - \beta^2 h_t}{\beta\alpha} \Rightarrow \\ \varepsilon_t^{new} &\in \left[\gamma - \sqrt{\frac{h_{t+2} - \omega - \beta\omega - \beta^2 h_t}{\beta\alpha}}, \gamma + \sqrt{\frac{h_{t+2} - \omega - \beta\omega - \beta^2 h_t}{\beta\alpha}} \right] \\ \varepsilon_t^{new} &\in [\gamma - l_t, \gamma + l_t] \end{aligned}$$

where

$$l_t = \sqrt{\frac{h_{t+2} - \omega - \beta\omega - \beta^2 h_t}{\beta\alpha}} \quad (34)$$

using an inverse transform method and then compute

$$h_{t+1}^{new} = \omega + \alpha (\varepsilon_t^{new} - \gamma)^2 + \beta h_t$$

which in turn implies a real value for $d_{t+1}^{new} = \sqrt{\frac{h_{t+2} - \omega - \beta h_{t+1}^{new}}{\alpha}}$ and so guarantees that h_{t+1}^{new} lies within the acceptance bounds.

The inverse transform method to draw the doubly truncated Gaussian random variable first draws a uniform random number

$$u \sim U(0, 1)$$

and then computes the following:

$$\bar{u} = (1 - u) \Phi\left(\frac{\gamma - l_t - \frac{r_t - \delta_1 h_t}{\lambda h_t + 1}}{\sqrt{\frac{(\lambda h_t + 1)^2 - h_t^2}{h_t(\lambda h_t + 1)}}}\right) + u \Phi\left(\frac{\gamma + l_t - \frac{r_t - \delta_1 h_t}{\lambda h_t + 1}}{\sqrt{\frac{(\lambda h_t + 1)^2 - h_t^2}{h_t(\lambda h_t + 1)}}}\right)$$

A draw is then given by:

$$\varepsilon_t^{new} = \Phi^{-1}(\bar{u})$$

However, if the bounds are close to each other (the degree of truncation is small) the extra computations involved make this method unnecessarily slow and so we prefer to use the accept-reject method where we draw $\varepsilon_t^{new}/r_t, h_t, \varphi \sim N\left(\frac{r_t - \delta_1 h_t}{\lambda h_t + 1}, \frac{\lambda h_t^2}{(\lambda h_t + 1)}\right)$ and accept the draw if $\gamma - l_t \leq \varepsilon_t^{new} \leq \gamma + l_t$ and otherwise we repeat the drawing (this method is inefficient if the truncation lies in the

tails of the distribution). It may be worth assessing the degree of truncation first, and depending on its tightness, choose one simulation method or the other.

The conditional density of ε_t^{new} will be given according to the definition of a truncated normal distribution:

$$\begin{aligned} & p\left(\varepsilon_t^{new} / \left| \varepsilon_t^{new} - \gamma \right| \leq l_t, r_t, h_t, \varphi\right) \\ &= \frac{1}{\sqrt{v_t/r_t, h_t}} \varphi\left(\frac{\varepsilon_t^{new} - \varepsilon_t/r_r, h_t}{\sqrt{v_t/r_t, h_t}}\right) \times \\ & \quad \times \left[\Phi\left(\frac{\gamma + l_t - \varepsilon_t/r_r, h_t}{\sqrt{v_t/r_t, h_t}}\right) - \Phi\left(\frac{\gamma - l_t - \varepsilon_t/r_r, h_t}{\sqrt{v_t/r_t, h_t}}\right) \right]^{-1} \end{aligned}$$

where $\Phi(\cdot)$ is the cdf of the standard normal.

By using the change of variable formula we have that the density of h_{t+1}^{new} will be:

$$\begin{aligned} & p\left(h_{t+1}^{new}/h_{t+1}^{new} \in [\omega + \beta h_t, \beta^{-1}(h_{t+2} - \omega)], r_t, h_t, \varphi\right) \\ &= \frac{c_t^{new}}{|2\alpha d_t^{new}|} \times \left[\Phi\left(\frac{\gamma + l_t - \varepsilon_t/r_r, h_t}{\sqrt{v_t/r_t, h_t}}\right) - \Phi\left(\frac{\gamma - l_t - \varepsilon_t/r_r, h_t}{\sqrt{v_t/r_t, h_t}}\right) \right]^{-1} \end{aligned}$$

Using Bayes theorem we have that the acceptance probability will be:

$$\min\left(1, \frac{p(h_{t+2}/h_{t+1}^{new}, r_{t+1}, \varphi) p(r_{t+1}/h_{t+1}^{new}, \varphi)}{p(r_{t+1}/h_{t+1}^r, \varphi) p(h_{t+2}/h_{t+1}^r, r_{t+1}, \varphi)}\right)$$

Since the degree of truncation is same for old and new the acceptance probability will be:

$$\min\left(1, \frac{p(r_{t+1}/h_{t+1}^{new}) c_{t+1}^{new} d_{t+1}^{new}}{p(r_{t+1}/h_{t+1}^r) c_{t+1}^r d_{t+1}^{new}}\right) \quad (35)$$

where $p(r_{t+1}/h_{t+1})$ is a mixture of two univariate normal densities so:

$$r_{t+1}/h_{t+1} \sim N(\delta_1 h_{t+1}, h_{t+1}(h_{t+1}\lambda + 1))$$

Hence:

$$p(r_{t+1}|h_{t+1}^r) = \frac{1}{\sqrt{2\pi(\lambda h_{t+1}^r + 1)h_{t+1}^r}} \exp\left(-\frac{(r_{t+1} - \delta_1 h_{t+1}^r)^2}{2(\lambda h_{t+1}^r + 1)h_{t+1}^r}\right)$$

and the acceptance probability becomes:

$$\min\left[1, \left(\frac{h_{t+1}^r}{h_{t+1}^{new}}\right)^{\frac{3}{2}} \frac{\sqrt{h_{t+2}^r - \omega - \beta h_{t+1}^r} \kappa(h_{t+1}^{new})}{\sqrt{h_{t+2}^r - \omega - \beta h_{t+1}^{new}} \kappa(h_{t+1}^r)}\right] \quad (36)$$

where

$$\kappa(h_{t+1}^i) = \exp \left[-\frac{(r_{t+1} - \delta_1 h_{t+1}^i)^2}{2(\lambda h_{t+1}^i + 1) h_{t+1}^i} - \frac{\lambda h_{t+1}^i + 1}{2\lambda (h_{t+1}^i)^2} \left(\left(\gamma - \frac{r_{t+1} - \delta_1 h_{t+1}^i}{\lambda h_{t+1}^i + 1} \right)^2 + \frac{h_{t+2}^r - \omega - \beta h_{t+1}^i}{\alpha} \right) \right] \quad (37)$$

$$\times \left[\frac{1 + \exp \left[2 \frac{\lambda h_{t+1}^i + 1}{\lambda (h_{t+1}^i)^2} \left(\gamma - \frac{r_{t+1} - \delta_1 h_{t+1}^i}{\lambda h_{t+1}^i + 1} \right) \sqrt{\frac{h_{t+2}^r - \omega - \beta h_{t+1}^i}{\alpha}} \right]}{\exp \left[\frac{\lambda h_{t+1}^i + 1}{\lambda (h_{t+1}^i)^2} \left(\gamma - \frac{r_{t+1} - \delta_1 h_{t+1}^i}{\lambda h_{t+1}^i + 1} \right) \sqrt{\frac{h_{t+2}^r - \omega - \beta h_{t+1}^i}{\alpha}} \right]} \right]$$

Overall the MCMC of $\{h_{t+1}\}/r, \varphi$ includes the following steps:

- (1) Specify an initial value $\{h^{(0)}\}$
- (2) Repeat for $r=1, \dots, M$
 - (a) Repeat for $t=0, \dots, T-1$
 - (i) Use an inverse transform method to simulate

$$\varepsilon_t^{new}/r_t, h_t, \varphi \sim N \left(\frac{r_t - \delta_1 h_t}{\lambda h_t + 1}, \frac{\lambda h_t^2}{(\lambda h_t + 1)} \right)$$

doubly truncated

$$(ii) \text{ Calculate } h_{t+1}^{new'} = \omega + \alpha (\varepsilon_t^c - \gamma)^2 + \beta h_t$$

Steps (2)(a)(i) and (2)(a)(ii) are equivalent to draw

$$(h_{t+1})^{new} \sim p \left(h_{t+1}^{new}/r_t, h_t, \varphi \right)$$

appropriately truncated

(iii) Calculate

$$\alpha_r = \min \left[1, \frac{p(r_{t+1}/h_{t+1}^{new}) c_{t+1}^{new} d_{t+1}^{new}}{p(r_{t+1}/h_{t+1}^r) c_{t+1}^r d_{t+1}^{new}} \right]$$

(iii) Set

$$(h_{t+1})^{r+1} = \begin{cases} (h_{t+1})^{new} & \text{if } Unif(0, 1) \leq \alpha_r \\ (h_{t+1})^r & \text{otherwise} \end{cases}$$

Everytime we change h_{t+1} we calculate only one normal density since the transformation is Markovian and since $t = 0, \dots, T-1$ we need $O(T)$ calculations.

Notice that if we retain h_{t+1}^{new} , then ε_t^{new} is retained and we will not need to simulate s_t at a later stage. In fact we only need to simulate s_t at $t = T$ since we need to know ε_T . The final step involves computing:

$$\delta_{t+1}^{(i)} = \left(r_{t+1} - \varepsilon_{t+1}^{(i)} \right) / h_{t+1}^{(i)}, \quad t = 0, \dots, T-1 \quad \text{and} \quad i = 1, \dots, M$$

Using all the above simulated values we may now take average of simulations and compute the quantities needed for the SEM algorithm. As for the Bayesian inference, having completed Step (ii) we may now proceed to the Gibbs-sampling and M-H steps to obtain draws from the required posterior density. Thus, the first order Markov transformation of the model made feasible a MCMC algorithm which allows the calculation of a classical estimator via the simulated EM algorithm and a simulation-based Bayesian inference in $O(T)$ computational operations.

5. EMPIRICAL APPLICATION

In this section we investigate the practical performance of the procedures described above. To do this, we use weekly excess returns from three major stock markets: Dow-Jones, FTSE and Nikkei for the period 1979:8 to 2008:5 (1500 observations). We take Monday as the first day after Friday

Before applying either the classical or Bayesian estimation procedures, we should talk a little bit about the identification issues of our model. More specifically, to avoid the identification problem we set the unconditional variance of ε_t equal to one. This implies that the constant term is set to $\omega = 1 - \alpha - \beta - \alpha\gamma^2$ and so we have to estimate now five parameters instead of six. This implies that $E(h_t) = 1$ and $\omega \geq 0$ as long as $\alpha + \beta < 1$ and γ is not too large. In particular, we use the reparameterization $\alpha + \beta = \sin^2(\psi_1^*)$ and $\beta/(\alpha + \beta) = \psi_2 = \sin^2(\psi_2^*)$ to guarantee $0 \leq \beta \leq 1 - \alpha \leq 1$. In addition, we set $\gamma = [(1 - \alpha - \beta)/\alpha]^{1/2} \sin(\psi_3^*)$ to ensure $\omega \geq 0$ and $E(h_t) = 1$. However, the performance of the Gibbs sampler can be very sensitive to the normalization choice and so we also used some accept-reject method for the Bayesian inference. This means that when drawing from the posterior (as well as from the prior) we had to ensure that $\alpha, \beta > 0, \alpha + \beta < 1$ and $\omega \geq 0$.

In order to implement our proposed Bayesian approach, we first have to specify the hyperparameters that characterize the prior distributions of the parameters. In this respect, our aim was to use informative priors that would be in accordance with the “received wisdom”. In particular, we set the prior mean for δ_1 equal to 0.01 which implies average excess returns (%) 0.52 on a yearly basis. Similarly, we choose for β for all stock markets prior mean of 0.88 which is centered around typical values estimated usually with weekly data. For a we decided to set the prior mean equal to 0.05 for the DowJones, equal to 0.01 for the Nikkei and 0.1 for the FTSE which are also centered around typical values. In order to diminish the impact of the prior on the joint posterior, we use rather vague priors by setting the prior variance of the skedastic function’s parameters a and β to 1000. We do not take any ex ante view on the sign of asymmetry effect by imposing a normal prior with zero mean and variance 300 for all stock markets. For the variance of the in-mean variable we use a standard inverted gamma prior with hyperparameters set to $\underline{v} = 80$ and $\underline{s}^2 = 0.5$. which imply that $E(\lambda) = 2.05$. Finally, we have to decide on \underline{q} . One attractive possibility is to set its inverse to \underline{T} , where \underline{T} is a scalar. Such a choice for \underline{q} is compatible with the usual interpretation that a conjugate prior can be viewed as the posterior distribution that could have been obtained from a fictitious sample from the same population (see e.g. Poirier, 1995). Thus, \underline{T} could be interpreted as the number of observations, or more precisely, the degrees of freedom in the fictitious sample. For that reason, we also set \underline{T} to 80 and this implies that $\underline{q} = 0.0125$. Thus, the prior standard deviation of δ_1 is 0.025. In any case, we performed a sensitivity analysis with respect to the variance hyperparameters and confirmed that our initial choice is vague enough and does not introduce significant information in our estimation.

We run a chain for 100,000 simulations for the three datasets and decided to use every tenth point, instead of all points, in the sample path to avoid strong serial correlation. The posterior statistics for the DowJones, Nikkei and FTSE are reported in Table 1.

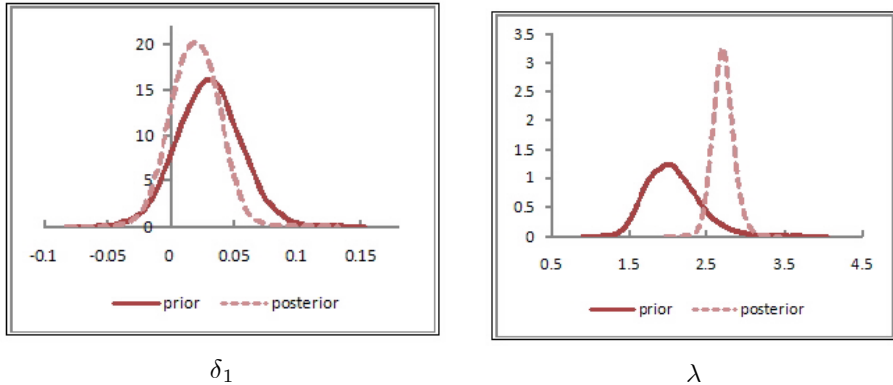
Table 1: Bayesian inference results

Dow Jones	PM	PSD	$\phi_{0.5}$	ϕ_{\min}	ϕ_{\max}	IF
δ_1	0.019	0.019	0.019	-0.049	0.150	5.031
λ	2.709	0.132	2.704	2.099	3.301	4.146
α	0.057	0.008	0.059	0.01	0.065	623.738
β	0.721	0.035	0.720	0.582	0.909	61.977
γ	1.377	0.352	1.350	-2.229	3.713	114.101
FTSE	PM	PSD	$\phi_{0.5}$	ϕ_{\min}	ϕ_{\max}	IF
δ_1	0.003	0.020	0.003	-0.075	0.086	19.691
λ	2.472	0.120	2.467	2.040	3.265	14.842
α	0.071	0.024	0.081	0.011	0.089	625.5
β	0.752	0.046	0.753	0.624	0.891	251.0
γ	1.324	0.943	0.869	0.031	3.720	357.8
Nikkei	PM	PSD	$\phi_{0.5}$	ϕ_{\min}	ϕ_{\max}	IF
δ_1	0.060	0.019	0.060	-0.009	0.122	11.334
λ	3.113	0.137	3.109	2.643	3.804	60.2
α	0.012	0.002	0.012	0.007	0.016	829.441
β	0.811	0.014	0.810	0.789	0.856	745.525
γ	3.883	0.329	3.800	2.436	4.424	562.4

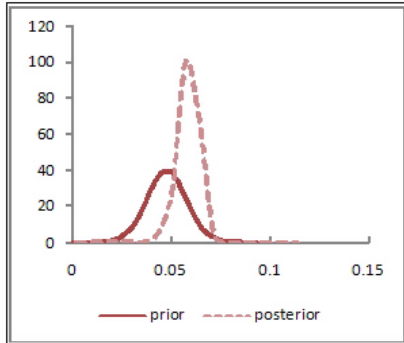
Note:PM denotes posterior mean, PSD posterior standard deviation, $\phi_{0.5}$ posterior median, ϕ_{\min} posterior minimum, ϕ_{\max} posterior maximum and IF inefficiency factor.¹

In all cases, there is a large positive assymetry effect and the price of risk coefficient is also positive. Inefficiency factors where calculated using a Parzen window equal to $0.1M$. Figures 1,2,3 show the kernel density estimates for all parameters for all datasets both for the posterior and the prior distribution. We used a canonical Epanechnikov kernel and the optimal bandwintdh was determined automatically by the data.

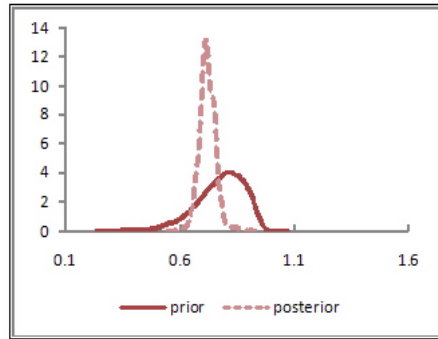
Figure 1:DowJones: Posterior and prior density estimates



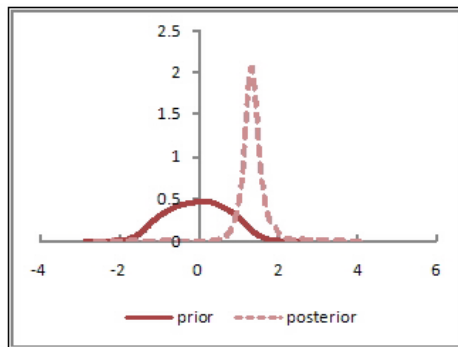
¹That is, the ratio of the variance of the sample mean of the drawings to the variance of the drawings divided by the number of replications.



α

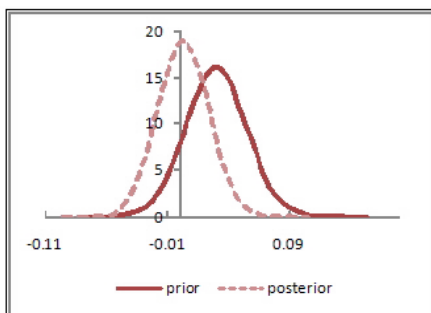


β

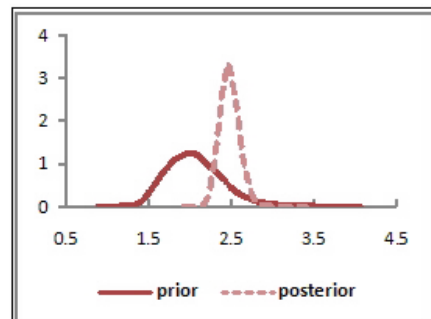


γ

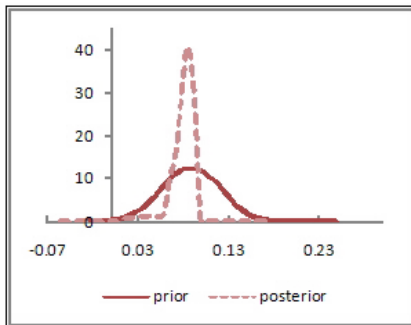
Figure 2:FTSE: Posterior and prior density estimates



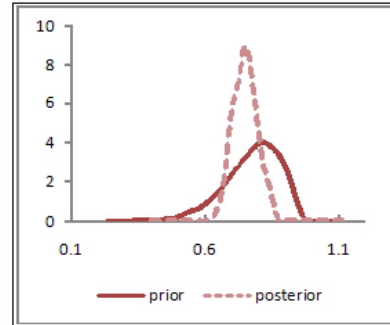
δ_1



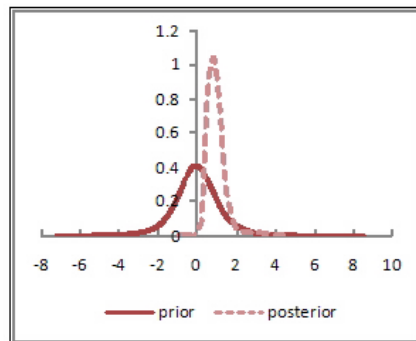
λ



α

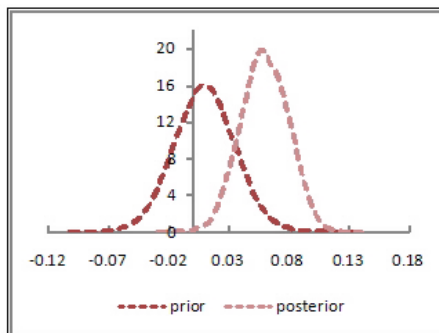


β

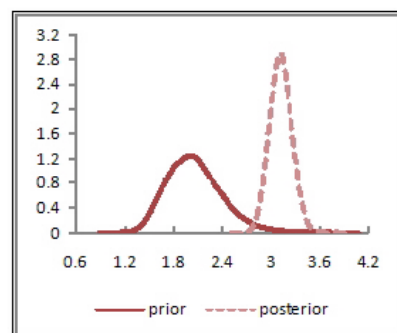


γ

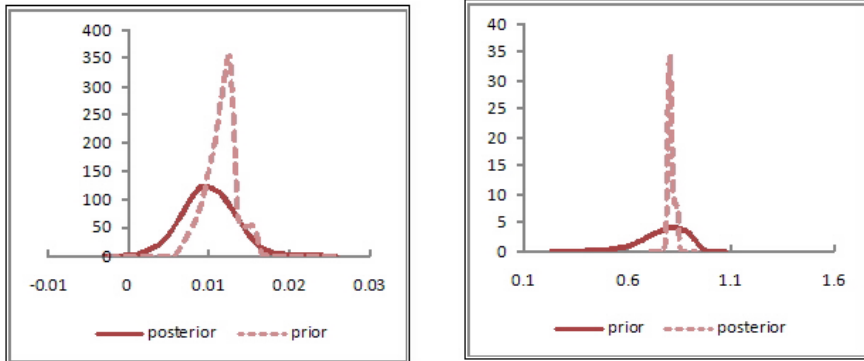
Figure 3: Nikkei: Posterior and prior density estimates



δ_1

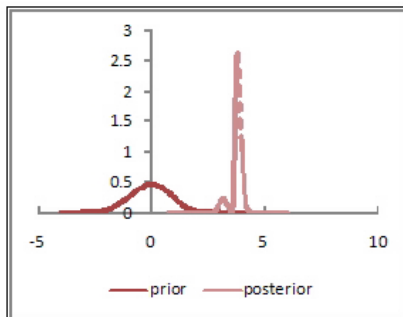


λ



α

β



γ

For the simulated EM algorithm, we decided to set all initial parameter values to their bayesian posterior means shown above. Hence, for example for the DowJones we set $\delta_1 = 0.019$, $\lambda = 2.709$, the asymmetry parameter γ to 1.377 and the GARCH parameters α, β to 0.057 and 0.721 respectively. Importantly, the numerical maximization with respect to the underlying parameters $\psi_1^*, \psi_2^*, \psi_3^*, \delta_1$ and λ was performed every single iteration. Note that in order to maximize at each EM iteration, we must maintain the latest simulated values constant and we must also maintain the underlying random drawings constant across EM iterations in order to allow the algorithm to converge (Nielsen, 2000).

Table 2 shows the parameter estimates using the simulated EM algorithm described above for the three datasets. If we use the posterior means of the parameters reported in Tables 1 as point estimates, our results suggest that the Bayesian and classical procedures are quite in agreement.

Table 2: SEM estimate results

	δ_1	λ	α	β	γ
DowJones	0.021	2.708	0.058	0.722	1.366
Ftse	0.009	2.425	0.070	0.751	1.330
Nikkei	0.052	3.124	0.013	0.814	3.616

Figure 4 shows also the density estimate of $\left\{\delta_{426}^{(i)}\right\}_{i=1}^{900,000}$ and $\left\{h_{426}^{(i)}\right\}_{i=1}^{900,000}$ for the DowJones. These are the simulated values obtained for the in mean process and the conditional variance for the 426th observation which corresponds to the third week of October 1987. As it is widely known this is just the week beginning with Black Monday, which refers to Monday, October 19, 1987, when stock markets around the world crashed, shedding a huge value in a very short time. On that day, the Dow Jones Industrial Average (DJIA) dropped by 508 points to 1739 (22.6%) and this decline was the largest one-day percentage decline in stock market history. On Friday, October 16, the DJIA closed at 2246.74 on record volume. As expected this dramatic event is illustrated in our Figures as well as in Table 3 where the posterior statistics are presented. The price of risk, as given from the posterior mean, is negative with a quite high posterior standard deviation. Moreover, the conditional variance is high indicating the presence of a large negative innovation. The October 1987 crash was unusual in many ways. It was characterized by the largest one-day drop in the history of stock markets indices since 1985, followed by a dramatic jump in stock volatility. Both these characteristics are captured by our estimation results.

Figure 4: DowJones: Density estimate of simulated δ and h in 3rd week of Oct.1987

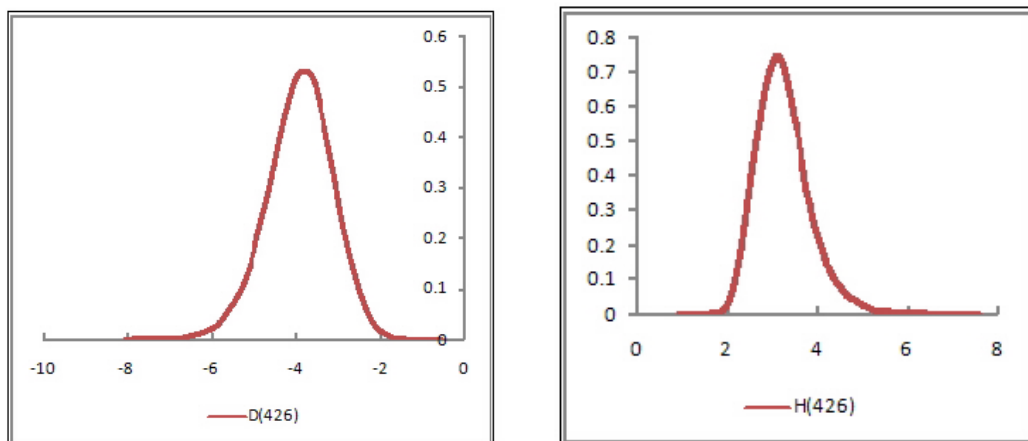


Table 3:DowJones:Posterior statistics for simulated δ and h in 3rd week of Oct.1987

DowJones	PM	PSD	PMD	Min	Max	IF
δ	-3.907	0.772	-3.869	-8.483	-1.341	7.723
h	3.229	0.577	3.168	1.500	6.956	9.879

Note:PM denotes posterior mean, PSD posterior standard deviation, PMD posterior median, Min minimum value, Max maximum value and IF inefficiency factor

Finally, Figures 5 and 6 show the average δ_t and h_t over the 900,000 simulations for illustrative reasons. This means that we have taken the average over simulations

for every t where $t = 1, \dots, 1500$. The largest average value for h_t corresponds to $t = 476$ and one of the most negative average value for δ_t is again for the same date.

Figure 5:DowJones: Average over 900,000 simulations of δ_t

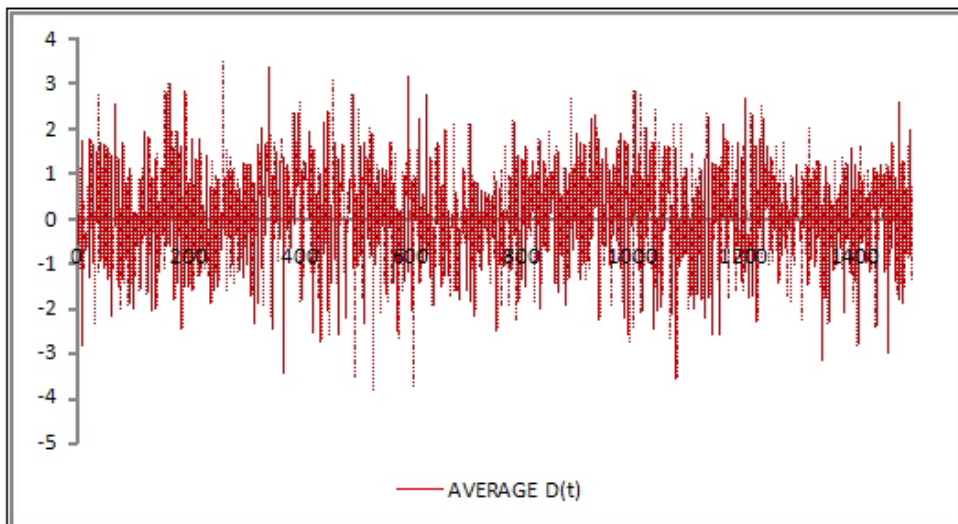
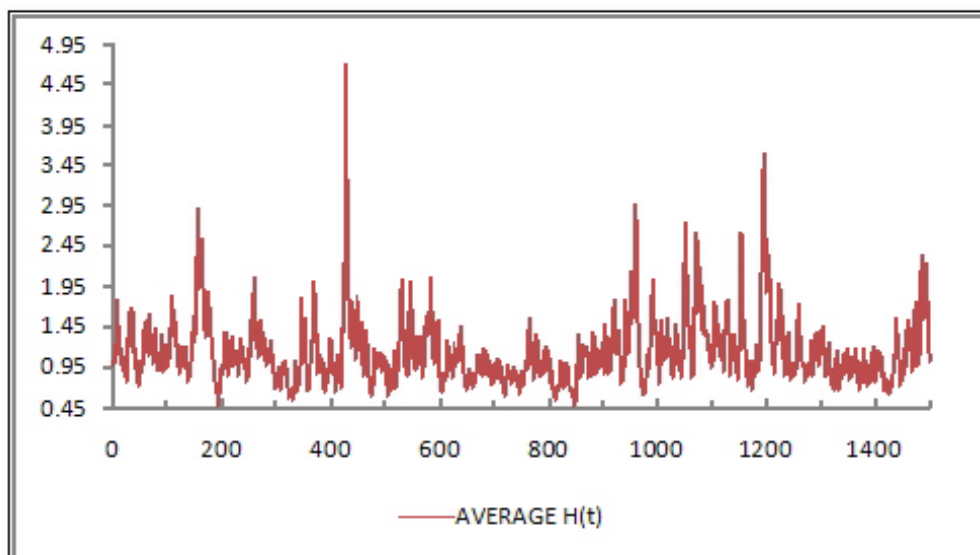


Figure 6:DowJones: Average over 900,000 simulations of h_t



6. CONCLUSIONS

We derive exact likelihood based estimators for our time-varying GQARCH(1,1)-M model. Since in general the expression for the likelihood function is unknown, we resort to simulation methods. In this context, we show that MCMC likelihood-based estimation of such a model can in fact be handled by means of feasible $O(T)$ algorithms. Our samplers involve two main steps. First we augment the state vector to achieve a first-order Markovian process in an analogous manner to the way in which GARCH models are simulated in practice. Then, we discuss how to simulate first the conditional variance and then the sign given these simulated series so that the unobserved in mean process is revealed as a residual term. We also develop simulation-based Bayesian inference procedures by combining within a Gibbs sampler the MCMC simulators.

In order to investigate the practical performance of the proposed procedure, we estimate both classically and within a Bayesian context, a GQARCH(1,1)-M model for weekly excess stock returns from the Dow-Jones, Nikkei and FTSE index.

Although we have developed the method within the context of normally distributed time-varying price of risk it applies much more widely. For example, we could assume that the market price of risk is a Bernoulli process or a Markov switching process. A Bernoulli distributed price of risk would allow a negative third moment by appropriately choosing the two values of the in-mean process. However, we preferred to assume normality for computational simplicity.

Finally, it is known that (eg. Tanner, 1996, pp. 84-85) that the EM algorithm slows down significantly in the neighborhood of the optimum. As a result, after some initial EM iterations it is tempting to switch to a derivative based optimization routine, which is more likely to quickly converge to the maximum. EM type arguments can be used to facilitate this switch by allowing the computation of the score. In particular, it is easy to see that:

$$E \left(\frac{\partial \ln p(\delta h | r, \varphi, \mathcal{F}_0)}{\partial \varphi} | r, \varphi^{(n)}, \mathcal{F}_0 \right) = 0$$

so it is clear that the score can be obtained as the expected value given $r, \varphi, \mathcal{F}_0$ of the sum of the unobservable scores corresponding to $\ln p(r | \delta h, \varphi, \mathcal{F}_0)$ and $\ln p(\delta h | \varphi, \mathcal{F}_0)$.

APPENDIX A: PROOFS

Proof of 4

$$E(\delta_t h_t | \mathcal{F}_{t-1}) = \delta_t h_t$$

$$V(\delta_t h_t | \mathcal{F}_{t-1}) = \lambda h_t^2$$

and

$$E(r_t | \delta_t h_t, \mathcal{F}_{t-1}) = \delta_t h_t$$

$$V(r_t | \delta_t h_t, \mathcal{F}_{t-1}) = h_t$$

Proof of 33

This comes from the fact that:

$$r_t = \delta_t h_t + \varepsilon_t$$

where

$$r_t | h_t \sim N(\delta_1 h_t, (\lambda h_t + 1) h_t)$$

and consequently,

$$\begin{pmatrix} \varepsilon_t \\ r_t \end{pmatrix} | h_t \sim N \left(\begin{pmatrix} 0 \\ \delta_1 h_t \end{pmatrix}, \begin{pmatrix} h_t & h_t \\ h_t & (\lambda h_t + 1) h_t \end{pmatrix} \right)$$

and thus from the definition of the bivariate normal:

$$E(\varepsilon_t | r_t, h_t) = 0 + h_t ((\lambda h_t + 1) h_t)^{-1} (r_t - \delta_1 h_t) = \frac{r_t - \delta_1 h_t}{\lambda h_t + 1}$$

$$\begin{aligned} \text{Var}(\varepsilon_t | r_t, h_t) &= h_t - h_t ((\lambda h_t + 1) h_t)^{-1} h_t = h_t - \frac{h_t}{\lambda h_t + 1} = \\ &= \frac{\lambda h_t^2 + h_t - h_t}{\lambda h_t + 1} = \frac{\lambda h_t^2}{(\lambda h_t + 1)} \end{aligned}$$

$$\varepsilon_t | r_t, h_t, \varphi \sim N \left(\frac{r_t - \delta_1 h_t}{\lambda h_t + 1}, \frac{\lambda h_t^2}{(\lambda h_t + 1)} \right)$$

Proof of 34

This may be shown by:

$$\begin{aligned} h_{t+1} &\leq \frac{h_{t+2} - \omega}{\beta} \Rightarrow \\ \omega + \alpha(\varepsilon_t - \gamma)^2 + \beta h_t &\leq \frac{h_{t+2} - \omega}{\beta} \Rightarrow \\ \alpha(\varepsilon_t - \gamma)^2 &\leq \frac{h_{t+2} - \omega}{\beta} - \omega - \beta h_t \Rightarrow \\ (\varepsilon_t - \gamma)^2 &\leq \frac{h_{t+2} - \omega - \omega\beta - \beta^2 h_t}{\alpha\beta} \Rightarrow \\ |(\varepsilon_t - \gamma)| &\leq \sqrt{\frac{h_{t+2} - \omega - \omega\beta - \beta^2 h_t}{\alpha\beta}} \Rightarrow \\ \sqrt{\frac{h_{t+2} - \omega - \omega\beta - \beta^2 h_t}{\alpha\beta}} &\leq (\varepsilon_t - \gamma) \leq \sqrt{\frac{h_{t+2} - \omega - \omega\beta - \beta^2 h_t}{\alpha\beta}} \Rightarrow \\ \gamma - \sqrt{\frac{h_{t+2} - \omega - \omega\beta - \beta^2 h_t}{\alpha\beta}} &\leq \varepsilon_t \leq \gamma + \sqrt{\frac{h_{t+2} - \omega - \omega\beta - \beta^2 h_t}{\alpha\beta}} \Rightarrow \\ \gamma - l_t &\leq \varepsilon_t \leq \gamma + l_t \end{aligned}$$

Proof of 36

We have that

$$p(r_{t+1} | h_{t+1}^r) = \frac{1}{\sqrt{2\pi (\lambda h_{t+1}^r + 1) h_{t+1}^r}} \exp \left(-\frac{(r_{t+1} - \delta_1 h_{t+1}^r)^2}{2 (\lambda h_{t+1}^r + 1) h_{t+1}^r} \right)$$

and

$$\frac{p(r_{t+1}|h_{t+1}^{new})}{p(r_{t+1}|h_{t+1}^r)} = \frac{\sqrt{(\lambda h_{t+1}^r + 1) h_{t+1}^r}}{\sqrt{(\lambda h_{t+1}^{new} + 1) h_{t+1}^{new}}} \exp\left(\frac{(r_{t+1} - \delta_1 h_{t+1}^r)^2}{2(\lambda h_{t+1}^r + 1) h_{t+1}^r} - \frac{(r_{t+1} - \delta_1 h_{t+1}^{new})^2}{2(\lambda h_{t+1}^{new} + 1) h_{t+1}^{new}}\right)$$

Also:

$$d_{t+1}^r = \sqrt{\frac{h_{t+2}^r - \omega - \beta h_{t+1}^r}{\alpha}}, \quad d_{t+1}^{new} = \frac{\sqrt{h_{t+2}^r - \omega - \beta h_{t+1}^r}}{\sqrt{h_{t+2}^r - \omega - \beta h_{t+1}^{new}}}$$

and

$$c_{t+1}^r = \frac{\sqrt{\lambda h_{t+1}^r + 1}}{\sqrt{2\pi\lambda h_{t+1}^r}} \left[\begin{array}{l} \exp\left(-\frac{1}{2} \frac{(\lambda h_{t+1}^r + 1)}{\lambda (h_{t+1}^r)^2} \left(\gamma + \sqrt{\frac{h_{t+2}^r - \omega - \beta h_{t+1}^r}{\alpha}} - \frac{r_{t+1} - \delta_1 h_{t+1}^r}{\lambda h_{t+1}^r + 1}\right)^2\right) \\ + \exp\left(-\frac{1}{2} \frac{(\lambda h_{t+1}^r + 1)}{\lambda (h_{t+1}^r)^2} \left(\gamma - \sqrt{\frac{h_{t+2}^r - \omega - \beta h_{t+1}^r}{\alpha}} - \frac{r_{t+1} - \delta_1 h_{t+1}^r}{\lambda h_{t+1}^r + 1}\right)^2\right) \end{array} \right]$$

$$\frac{c_{t+1}^{new}}{c_{t+1}^r} = \frac{h_{t+1}^r \sqrt{\lambda h_{t+1}^{new} + 1} \left[\begin{array}{l} \exp\left(-\frac{1}{2} \frac{(\lambda h_{t+1}^{new} + 1)}{\lambda (h_{t+1}^{new})^2} \left(\gamma + \sqrt{\frac{h_{t+2}^r - \omega - \beta h_{t+1}^{new}}{\alpha}} - \frac{r_{t+1} - \delta_1 h_{t+1}^{new}}{\lambda h_{t+1}^{new} + 1}\right)^2\right) + \\ + \exp\left(-\frac{1}{2} \frac{(\lambda h_{t+1}^{new} + 1)}{\lambda (h_{t+1}^{new})^2} \left(\gamma - \sqrt{\frac{h_{t+2}^r - \omega - \beta h_{t+1}^{new}}{\alpha}} - \frac{r_{t+1} - \delta_1 h_{t+1}^{new}}{\lambda h_{t+1}^{new} + 1}\right)^2\right) \end{array} \right]}{h_{t+1}^{new} \sqrt{\lambda h_{t+1}^r + 1} \left[\begin{array}{l} \exp\left(-\frac{1}{2} \frac{(\lambda h_{t+1}^r + 1)}{\lambda (h_{t+1}^r)^2} \left(\gamma + \sqrt{\frac{h_{t+2}^r - \omega - \beta h_{t+1}^r}{\alpha}} - \frac{r_{t+1} - \delta_1 h_{t+1}^r}{\lambda h_{t+1}^r + 1}\right)^2\right) + \\ + \exp\left(-\frac{1}{2} \frac{(\lambda h_{t+1}^r + 1)}{\lambda (h_{t+1}^r)^2} \left(\gamma - \sqrt{\frac{h_{t+2}^r - \omega - \beta h_{t+1}^r}{\alpha}} - \frac{r_{t+1} - \delta_1 h_{t+1}^r}{\lambda h_{t+1}^r + 1}\right)^2\right) \end{array} \right]}$$

So the result comes straightforward.

REFERENCES

- [1] Akaike, H. (1974): "Markovian Representation of Stochastic Processes and its Application to the Analysis of Autoregressive Moving Average Processes", *Annals of the Institute of Statistical Mathematics*, 26, 363-387.
- [2] Albert, J.H., and S.Chib (1993): "Bayesian Inference via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance Shifts", *Journal of Business and Economic Statistics*, 11, 1-15.
- [3] Ardia, D. (2008): "Bayesian Estimation of a Markov-Switching Threshold Asymmetric GARCH Model with Student-t Innovations", *Econometrics Journal*, 00, 1-22.
- [4] Arvanitis, S. and A. Demos (2004): "Time Dependence and Moments of a Family of Time-Varying Parameter GARCH in Mean Models", *Journal of Time Series Analysis*, Vol.25, No1, 1-25.

- [5] Barndorff-Nielsen, O.E. and N. Shephard (2000): "Modelling by Levy Processes for Financial Econometrics", *D.P. No 2000-W3, Nuffield College, Oxford University*.
- [6] Bauwens, L., and M. Lubrano (1998): "Bayesian Inference on Garch Models Using the Gibbs Sampler", *Econometrics Journal*,1, C23-C46.
- [7] Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*,31,307-27.
- [8] Bollerslev, T., Chou, R.Y. and K.F. Kroner (1992): "ARCH Modelling in Finance: A Review of the Theory and Empirical Evidence", *Journal of Econometrics* ,52,5-59.
- [9] Bollerslev, T., R.F. Engle and D.B. Nelson (1994): "ARCH Models", in *The Handbook of Econometrics*, Vol.4, ed. by R.F. Engle and D.McFadden. Amsterdam:North-Holland, 2959-3038.
- [10] Box, G.E.P. and G.M. Jenkins(1996): *Time Series Analysis Forecasting and Control*, San Francisco:Holden-Day.
- [11] Campbell, J.Y. (1991): "A Variance Decomposition for Stock Returns", *Economic Journal*,101, 157-179.
- [12] Campbell, J.Y., Lo, A. W. and A.C. MacKinlay (1997): *The Econometrics of Financial Markets*, Princeton University Press.
- [13] Chib, S. (2001): "Markov Chain Monte Carlo Methods: Computation and Inference", in *Handbook of Econometrics*, Vol.5, ed. by J.J.Heckman and E.Leamer. Amsterdam:North-Holland,3569-3649.
- [14] Chib, S., and E. Greenberg (1994): "Bayes Inference for Regression Models with ARMA(p,q) Errors", *Journal of Econometrics*, 64, 183-206.
- [15] Chou, R., Engle, R.F. and A. Kane (1992): "Measuring Risk Aversion from Excess Returns on a Stock Index", *Journal of Econometrics*,52, 201-24.
- [16] Demos, A. and E.Sentana (1998): "An EM Algorithm for Conditionally Heteroskedastic Factor Models", *Journal of Business and Economic Statistics*,16, 357-361.
- [17] Demster, A.P., N.Laird, and D.B. Rubin(1977): "Maximum likelihood from Incomplete Data via the EM Algorithm", *Journal of the Royal Statistical Society*, Series B, 39,1-38.
- [18] Dueker, M. (1999): "Conditional Heteroskedasticity in Qualitative Response Models of Time Series: A Gibbs-Sampling Approach to the Bank Prime Rate", *Journal of Business and Economic Statistics*, 17, 466-472.
- [19] Engle, R.F. (1982): "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation", *Econometrica*,50, 987-1007.
- [20] Engle, R.F., Lilien, D.M. and R.P. Robins (1987): "Estimating Time-Varying Risk-Premia in the Term Structure: The ARCH-Models", *Econometrica*,55, 391-407.

- [21] Fiorentini, G., Sentana, E. and N.Shephard (2004): "Likelihood-Based Estimation of Latent Generalized ARCH Structures", *Econometrica*, Vol.72, No.5,1481-1517.
- [22] Gelfand, A.E. and A.F.M. Smith (1990): "Sampling-Based Approaches to Calculating Marginal Densities", *Journal of the American Statistical Association*,85,398-409.
- [23] Geweke, J. (1989): "Bayesian Inference in Econometric Models Using Monte Carlo Integration", *Econometrica*, 57, 1317-1339.
- [24] Giakoumatos, S.G., P. Dellaportas and D.M. Politis (1999): "Bayesian Analysis of the Unobserved ARCH Model", *Department of Statistics*, Athens University of Economics and Business.
- [25] Glosten, L.R., Jaganathan, R. and D.E. Runkle (1993): "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks", *Journal of Finance*,48, 1101-801.
- [26] Hastings, W.K. (1970):"Monte Carlo Sampling Methods Using Markov Chains and their Applications", *Biometrika*,57,97-109
- [27] Higgins, M.L, and Bera, A.K. (1992): "A Class of Nonlinear ARCH Models", *International Economic Review* ,33, 137-58.
- [28] Kim, S. and C.R. Nelson (1999): *State-Space Models with Regime Switching. Classical and Gibbs-Sampling Approaches with Applications*. Cambridge:MIT.
- [29] Kim, S., N. Shephard and S. Chib (1998): "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models", *Review of Economic Studies*, 65, 361-393.
- [30] Louis, T.A. (1982):"Finding Observed Information Using the EM Algorithm", *Journal of the Royal Statistical Society*, Series B 44, 98-103.
- [31] Mantegna, R.N. and H.E. Stanley (1996): "Turbulence and Financial Markets", *Nature*, 383,587-8.
- [32] Mantegna, R.N. and H.E. Stanley (2000): *An Introduction to Econophysics. Correlations and Complexity in Finance*, Cambridge, UK: Cambridge University Press.
- [33] Merton, R.C. (1980): "Stationarity and Persistence in the GARCH (1,1) Model", *Econometric Theory* ,6, 318-34.
- [34] Nakatsuma, T. (2000): "Bayesian Analysis of the ARMA-GARCH Models: A Markov Chain Sampling Approach", *Journal of Econometrics*, 95, 57-69.
- [35] Nielsen, S.F. (2000): "On Simulated EM Algorithms", *Journal of Econometrics* ,96,267-292.
- [36] Poirier, D.J. (1995): *Intermediate Statistics and Econometrics*, Cambridge:MIT.
- [37] Ruud, P. (1991): "Extensions of Estimation Methods Using the EM Algorithm", *Journal of Econometrics* ,49, 305-341.

- [38] Sentana, E. (1995): "Quadratic ARCH Models", *Review of Economic Studies*, 62, 639-661.
- [39] Shephard, N. (1996): "Statistical Aspects of ARCH and Stochastic Volatility", in *Time Series Models in Econometrics, Finance and Other Fields*, ed. by D.R. Cox, D.V. Hinkley, and O.E. Barndorff-Nielsen. London:Chapman and Hall. 1-67.
- [40] Spivak, M. (1965): *Calculus on Manifolds. A Modern Approach to Classical Theorems of Advanced Calculus*. New York:Benjamin.
- [41] Tanner, M.A. (1996): *Tools for Statistical Inference:Methods for Exploration of Posterior Distributions and Likelihood Functions* (Third ed.). New-York:Springer-Verlag.
- [42] Wei, S.X. (2002): "A Censored-GARCH Model of Asset Returns with Price Limits", *Journal of Empirical Finance*, 9,197-223.