

## Research Article

# Estimation and Properties of a Time-Varying GQARCH(1,1)-M Model

**Sofia Anyfantaki and Antonis Demos**

*Athens University of Economics and Business, Department of International and European Economic Studies, 10434 Athens, Greece*

Correspondence should be addressed to Antonis Demos, [demos@aueb.gr](mailto:demos@aueb.gr)

Received 16 May 2011; Accepted 14 July 2011

Academic Editor: Mike Tsionas

Copyright © 2011 S. Anyfantaki and A. Demos. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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. This paper outlines the issues and suggests to employ a Markov chain Monte Carlo algorithm 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. Furthermore, the theoretical dynamic properties of a time-varying GQARCH(1,1)-M are derived. We discuss them and apply the suggested Bayesian estimation to three major stock markets.

## 1. Introduction

Time series data, emerging from diverse fields appear to possess time-varying second conditional moments. Furthermore, theoretical results seem to postulate quite often, specific relationships between the second and the first conditional moment. For instance, in the stock market context, the first conditional moment of stock market's excess returns, given some information set, is a possibly time-varying, linear function of volatility (see, e.g., Merton [1], Glosten et al. [2]). These have led to modifications and extensions of the initial ARCH model of Engle [3] and its generalization by Bollerslev [4], 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. [5] for a survey).

Although the vast majority of the research in conditional heteroscedasticity is being processed aiming at the stylized facts of financial stock returns and of economic time series

in general, Arvanitis and Demos [6] 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 (e.g., leptokurtosis, volatility clustering, and quasi long-range dependence in the squares are common), there are also significant differences in the behavior of the before mentioned physical systems as opposed to financial markets (examples are the anticorrelation effect and asymmetry of velocity changes in contrast to zero autocorrelation and the leverage effect of financial returns) (see Barndorff-Nielsen and Shephard [7] as well as Mantegna and Stanley [8, 9]). 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.

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 employ the method proposed in Fiorentini et al. [10] to achieve MCMC likelihood-based estimation of a time-varying GARCH-M model by means of feasible  $O(T)$  algorithms, where  $T$  is the sample size. The crucial idea is to transform the GARCH model in a first-order Markov's model. However, in our model, the error term enters the in-mean equation multiplicatively and not additively as it does in the latent factor models of Fiorentini et al. [10]. Thus, we show that their method applies to more complicated models, as well.

We prefer to employ a GQARCH specification for the conditional variance (Engle [3] and Sentana [11]) since it encompasses all the existing restricted quadratic variance functions (e.g., augmented ARCH model), its properties are very similar to those of GARCH models (e.g., stationarity conditions) but avoids some of their criticisms (e.g., very easy to generalize to multivariate models). Moreover, many theories in finance involve an explicit tradeoff 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 et al. [2] 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 it follows an AR(1) process. Thus, our model is a time-varying GQARCH-M-AR(1) model.

As we shall see in Section 2.1, this model is able to capture the, so-called, stylized facts of excess stock returns. These are (i) the sample mean is positive and much smaller than the standard deviation, that is, high coefficient of variation, (ii) the autocorrelation of excess returns is insignificant with a possible exception of the 1st one, (iii) the distribution of returns is nonnormal mainly due to excess kurtosis and may be asymmetry (negative), (iv) there is strong volatility clustering, that is, significant positive autocorrelation of squared returns even for high lags, and (v) the so-called leverage effect; that is, negative errors increase future volatility more than positive ones of the same size.

The structure of the paper is as follows. In Section 2, we present the model and derive the theoretical properties the GQARCH(1,1)-M-AR(1) model. Next, we review Bayesian and classical likelihood approaches to inference for the time-varying GQARCH-M model. We show that the key task (in both cases) is to be able to produce consistent simulators and that the estimation problem arises from the existence of two unobserved

processes, causing exact likelihood-based estimations computationally infeasible. Hence, we demonstrate that the method proposed by Fiorentini et al. [10] is needed to achieve a first-order Markov's transformation of the model and thus, reducing the computations from  $O(T^2)$  to  $O(T)$ . A comparison of the efficient ( $O(T)$  calculations) and the inefficient ( $O(T^2)$  ones) simulator is also given. An illustrative empirical application on weekly returns from three major stock markets is presented in Section 4, and we conclude in Section 5.

## 2. GQARCH(1,1)-M-AR(1) Model

The definition of our model is as follows.

*Definition 2.1.* The time-varying parameter GQARCH(1,1)-M-AR(1) model:

$$r_t = \delta_t h_t + \varepsilon_t, \quad \varepsilon_t = z_t h_t^{1/2}, \quad (2.1)$$

where

$$\delta_t = (1 - \varphi)\delta + \varphi\delta_{t-1} + \varphi_u u_t, \quad (2.2)$$

$$h_t = \omega + \alpha(\varepsilon_{t-1} - \gamma)^2 + \beta h_{t-1}, \quad (2.3)$$

$z_t \sim \text{i.i.d. } N(0, 1)$ ,  $u_t \sim \text{i.i.d. } N(0, 1)$ ,  $u_t, z_t$  are independent for all  $t$ 's, and where  $\{r_t\}_{t=1}^T$  are the observed excess returns,  $T$  is the sample size,  $\{\delta_t\}_{t=1}^T$  is an unobserved AR(1) process independent (with  $\delta_0 = \delta$ ) of  $\{\varepsilon_t\}_{t=1}^T$ , and  $\{h_t\}_{t=1}^T$  is the conditional variance (with  $h_0$  equal to the unconditional variance and  $\varepsilon_0 = 0$ ) which is supposed to follow a GQARCH(1,1). It is obvious that  $\delta_t$  is the market price of risk (see, e.g., Merton [1] Glosten et al. [2]). Let us call  $\mathcal{F}_{t-1}$  the sequence of natural filtrations generated by the past values of  $\{\varepsilon_t\}$  and  $\{r_t\}$ .

Modelling the theoretical properties of this model has been a quite important issue. Specifically, it would be interesting to investigate whether this model can accommodate the main stylized facts of the financial markets. On other hand, the estimation of the model requires its transformation into a first-order Markov's model to implement the method of Fiorentini et al. [10]. Let us start with the theoretical properties.

### 2.1. Theoretical Properties

Let us consider first the moments of the conditional variance  $h_t$ , needed for the moments of  $r_t$ . The proof of the following lemma is based on raising  $h_t$  to the appropriate power, in (2.3), and taking into account that  $E(z_t^4) = 3$ ,  $E(z_t^6) = 15$  and  $E(z_t^8) = 105$ .

**Lemma 2.2.** *If  $105\alpha^4 + 60\alpha^3\beta + 18\alpha^2\beta^2 + 4\alpha\beta^3 + \beta^4 < 1$ , the first four moments of the conditional variance of (2.3) are given by*

$$\begin{aligned}
E(h_t) &= \frac{\omega + \alpha\gamma^2}{1 - (\alpha + \beta)}, \\
E(h_t^2) &= (\omega + \alpha\gamma^2) \frac{(\omega + \alpha\gamma^2)(1 + \alpha + \beta) + 4\alpha^2\gamma^2}{(1 - 3\alpha^2 - 2\alpha\beta - \beta^2)(1 - \alpha - \beta)}, \\
E(h_t^3) &= \frac{(\omega + \alpha\gamma^2)^3 + 3(\omega + \alpha\gamma^2)[(\omega + \alpha\gamma^2)(\alpha + \beta) + 4\alpha^2\gamma^2]E(h_t)}{1 - \beta^3 - 15\alpha^3 - 9\alpha^2\beta - 3\alpha\beta^2} \\
&\quad + \frac{3[(\omega + \alpha\gamma^2)(3\alpha^2 + \beta^2 + 2\beta\alpha) + 4\alpha^2\gamma^2(3\alpha + \beta)]E(h_t^2)}{1 - \beta^3 - 15\alpha^3 - 9\alpha^2\beta - 3\alpha\beta^2}, \\
E(h_t^4) &= (\omega + \alpha\gamma^2)^2 \frac{(\omega + \alpha\gamma^2)^2 + 4[(\omega + \alpha\gamma^2)(\alpha + \beta) + 6\alpha^2\gamma^2]E(h_t)}{1 - 105\alpha^4 - 60\alpha^3\beta - 18\alpha^2\beta^2 - 4\alpha\beta^3 - \beta^4} \\
&\quad + 6 \frac{(\omega + \alpha\gamma^2)^2(3\alpha^2 + \beta^2 + 2\beta\alpha) + 8[(\omega + \alpha\gamma^2)(3\alpha + \beta) + \alpha^2\gamma^2]\alpha^2\gamma^2}{1 - 105\alpha^4 - 60\alpha^3\beta - 18\alpha^2\beta^2 - 4\alpha\beta^3 - \beta^4} E(h_t^2) \\
&\quad + 4 \frac{(\omega + \alpha\gamma^2)(15\alpha^3 + 9\alpha^2\beta + 3\alpha\beta^2 + \beta^3) + 6\alpha^2\gamma^2(15\alpha^2 + 6\alpha\beta + \beta^2)}{1 - 105\alpha^4 - 60\alpha^3\beta - 18\alpha^2\beta^2 - 4\alpha\beta^3 - \beta^4} E(h_t^3).
\end{aligned} \tag{2.4}$$

Consequently, the moments of  $r_t$  are given in the following theorem taken from Arvanitis and Demos [6].

**Theorem 2.3.** *The first two moments of the model in (2.1), (2.2), and (2.3) are given by*

$$E(r_t) = \delta E(h_t), \quad E(r_t^2) = \left( \delta^2 + \frac{\varphi_u^2}{1 - \varphi^2} \right) E(h_t^2) + E(h_t), \tag{2.5}$$

whereas the skewness and kurtosis coefficients are

$$Sk(r_t) = \frac{S(r_t)}{\text{Var}^{1.5}(r_t)}, \quad \text{kurt}(r_t) = \frac{\kappa}{\text{Var}^2}(r_t), \tag{2.6}$$

where

$$\begin{aligned}
S(r_t) &= \delta \left( \delta^2 + 3 \frac{\varphi_u^2}{1 - \varphi^2} \right) E(h_t^3) + 3\delta E(h_t^2) + 2\delta^3 E^3(h_t) \\
&\quad - 3\delta \left[ \left( \delta^2 + \frac{\varphi_u^2}{1 - \varphi^2} \right) E(h_t^2) + E(h_t) \right] E(h_t),
\end{aligned}$$

$$\begin{aligned}
\kappa = & \left( \delta^4 + 6\delta^2 \frac{\varphi_u^2}{1-\varphi^2} + 3 \frac{\varphi_u^4}{(1-\varphi^2)^2} \right) E(h_t^4) + 3\delta^2 [2 - \delta^2 E(h_t)] E^3(h_t) \\
& + 6 \left( \delta^2 + \frac{\varphi_u^2}{1-\varphi^2} \right) E(h_t^3) - 4\delta^2 \left( \delta^2 + \frac{3\varphi_u^2}{1-\varphi^2} \right) E(h_t^3) E(h_t) \\
& + \left\{ 6\delta^2 \left[ \left( \delta^2 + \frac{\varphi_u^2}{1-\varphi^2} \right) E(h) - 2 \right] E(h) + 3 \right\} E(h_t^2),
\end{aligned} \tag{2.7}$$

and  $E(h_t)$ ,  $E(h_t^2)$ ,  $E(h_t^3)$ , and  $E(h_t^4)$  are given in Lemma 2.2.

In terms of stationarity, the process  $\{r_t\}$  is 4th-order stationary if and only if

$$|\varphi| < 1, \quad 105\alpha^4 + 60\alpha^3\beta + 18\alpha^2\beta^2 + 4\alpha\beta^3 + \beta^4 < 1. \tag{2.8}$$

These conditions are the same as in Arvanitis and Demos [6], indicating that the presence of the asymmetry parameter,  $\gamma$ , does not affect the stationarity conditions (see also Bollerslev [4] and Sentana [11]). Furthermore, the 4th-order stationarity is needed as we would like to measure the autocorrelation of the squared  $r_t$ 's (volatility clustering), as well as the correlation of  $r_t^2$  and  $r_{t-1}$  (leverage effect). The dynamic moments of the conditional variance and those between the conditional variance  $h_t$  and the error  $\varepsilon_t$  are given in the following two lemmas (for a proof see Appendix B).

**Lemma 2.4.** *Under the assumption of Lemma 2.2, one has that*

$$\begin{aligned}
\text{Cov}(h_t, h_{t-k}) &= (\alpha + \beta)^k V(h_t), \\
\text{Cov}(h_t^2, h_{t-k}) &= A^k [E(h_t^3) - E(h_t^2)E(h_t)] + \frac{(\alpha + \beta)^k - A^k}{(\alpha + \beta) - A} B V(h_t), \\
\text{Cov}(h_t, h_{t-k}^2) &= (\alpha + \beta)^k [E(h_{t-1}^3) - E(h_{t-1}^2)E(h_{t-1})], \\
\text{Cov}(h_t^2, h_{t-k}^2) &= A^k V(h_t^2) + B \frac{(\alpha + \beta)^k - A^k}{(\alpha + \beta) - A} [E(h_t^3) - E(h_t^2)E(h_t)],
\end{aligned} \tag{2.9}$$

where  $A = 3\alpha^2 + \beta^2 + 2\alpha\beta$  and  $B = 2[2\alpha^2\gamma^2 + (\omega + \alpha\gamma^2)(\alpha + \beta)]$ .

**Lemma 2.5.**

$$\begin{aligned}
\text{Cov}(h_t, \varepsilon_{t-k}) &= E(h_t \varepsilon_{t-k}) = -2\alpha\gamma(\alpha + \beta)^{k-1} E(h_t), \\
\text{Cov}(h_t h_{t-k}, \varepsilon_{t-k}) &= E(h_t h_{t-k} \varepsilon_{t-k}) = -2\alpha\gamma(\alpha + \beta)^{k-1} E(h_{t-1}^2), \\
\text{Cov}(h_t, \varepsilon_{t-k}^2) &= (\alpha + \beta)^k V(h_t) + 2\alpha(\alpha + \beta)^{k-1} E(h_t^2),
\end{aligned}$$

$$\begin{aligned}
\text{Cov}(h_t^2, \varepsilon_{t-k}) &= E(h_t^2 \varepsilon_{t-k}) = -4\alpha\gamma(3\alpha + \beta)A^{k-1}E(h_t^2) \\
&\quad - 4\alpha\gamma \left[ (\omega + \alpha\gamma^2) \frac{(\alpha + \beta)^k - A^k}{(\alpha + \beta) - A} + 2\alpha^2\gamma^2 \frac{(\alpha + \beta)^{k-1} - A^{k-1}}{(\alpha + \beta) - A} \right] E(h_t), \\
\text{Cov}(h_t^2, \varepsilon_{t-k}^2) &= A^k \left[ E(h_t^3) - E(h_t^2)E(h_t) \right] + B \frac{(\alpha + \beta)^k - A^k}{(\alpha + \beta) - A} V(h_t) + 4\alpha(3\alpha + \beta)A^{k-1}E(h_t^3) \\
&\quad + \left[ 2\alpha B \frac{(\alpha + \beta)^{k-1} - A^{k-1}}{(\alpha + \beta) - A} + 4(2\alpha^2\gamma^2 + (\omega + \alpha\gamma^2))A^{k-1} \right] E(h_t^2), \\
\text{Cov}(h_t^2 h_{t-k}, \varepsilon_{t-k}) &= E(h_t^2 h_{t-k} \varepsilon_{t-k}) = -4\alpha\gamma A^{k-1} \left[ (3\alpha + \beta)E(h_t^3) + (\omega + \alpha\gamma^2)E(h_t^2) \right] \\
&\quad - 2\alpha\gamma B \frac{(\alpha + \beta)^{k-1} - A^{k-1}}{(\alpha + \beta) - A} E(h_t^2),
\end{aligned} \tag{2.10}$$

where  $A$  and  $B$  are given in Lemma 2.4.

Furthermore, from Arvanitis and Demos [6] we know that the following results hold.

**Theorem 2.6.** *The autocovariance of returns for the model in (2.1)–(2.3) is given by*

$$\gamma_k = \text{Cov}(r_t, r_{t-k}) = \delta^2 \text{Cov}(h_t, h_{t-k}) + \delta E(h_t \varepsilon_{t-k}) + \varphi^k \frac{\varphi_u^2}{1 - \varphi^2} E(h_t h_{t-k}), \tag{2.11}$$

and the covariance of squares' levels and the autocovariance of squares are

$$\begin{aligned}
\text{Cov}(r_t^2, r_{t-k}^2) &= E(\delta_t^2 \delta_{t-k}^2) \text{Cov}(h_t^2, h_{t-k}^2) + \text{Cov}(\delta_t^2, \delta_{t-k}^2) E(h_t^2) E(h_{t-k}^2) \\
&\quad + E(\delta_{t-k}^2) \text{Cov}(h_t, h_{t-k}) + E(\delta_t^2) E(h_t^2 \varepsilon_{t-k}) + E(h_t \varepsilon_{t-k}), \\
\text{Cov}(r_t^2, r_{t-k}^2) &= E^2(\delta_t^2) \text{Cov}(h_t^2, h_{t-k}^2) + \text{Cov}(\delta_t^2, \delta_{t-k}^2) E(h_t^2 h_{t-k}^2) \\
&\quad + E(\delta_t^2) \text{Cov}(h_t^2, \varepsilon_{t-k}^2) + 2E(\delta_t^2 \delta_{t-k}^2) E(h_t^2 h_{t-k} \varepsilon_{t-k}) \\
&\quad + E(\delta_t^2) \text{Cov}(h_t, h_{t-k}^2) + \text{Cov}(h_t, \varepsilon_{t-k}^2) + 2\delta E(h_t h_{t-k} \varepsilon_{t-k}),
\end{aligned} \tag{2.12}$$

where all needed covariances and expectations of the right-hand sides are given in Lemmas 2.4 and 2.5,

$$\begin{aligned}
\text{Cov}(\delta_t^2, \delta_{t-k}^2) &= \text{Cov}(\delta_t, \delta_{t-k}^2) = 2\varphi^k \delta \frac{\varphi_u^2}{1 - \varphi^2}, \\
\text{Cov}(\delta_t^2, \delta_{t-k}^2) &= 4\varphi^k \delta^2 \frac{\varphi_u^2}{1 - \varphi^2} + 2\varphi^{2k} \frac{\varphi_u^4}{(1 - \varphi^2)^2}.
\end{aligned} \tag{2.13}$$

From the above theorems and lemmas it is obvious that our model can accommodate all stylized facts. For example, negative asymmetry is possible; volatility clustering and the leverage effect (negative  $\text{Cov}(h_t, \varepsilon_{t-k})$ ) can be accommodated, and so forth. Furthermore, the model can accommodate negative autocorrelations,  $\gamma_k$ , something that is not possible for the GARCH-M model (see Fiorentini and Sentana [12]). Finally, another interesting issue is the diffusion limit of the time-varying GQARCH-M process. As already presented by Arvanitis [13], the weak convergence of the Time-varying GQARCH(1,1)-M coincides with the general conclusions presented elsewhere in the literature. These are that weak limits of the endogenous volatility models are exogenous (stochastic volatility) continuous-time processes. Moreover, Arvanitis [13] suggests that there is a distributional relation between the GQARCH model and the continuous-time Ornstein-Uhlenbeck models with respect to appropriate nonnegative Levy's processes.

Let us turn our attention to the estimation of our model. We will show that estimating our model is a hard task and the use of well-known methods such as the EM-algorithm cannot handle the problem due to the huge computational load that such methods require.

### 3. Likelihood-Inference: EM and Bayesian Approaches

The purpose of this section is the estimation of the time-varying GQARCH(1,1)-M model. Since our model involves two unobserved components (one from the time-varying in-mean parameter and one from the error term), the estimation method required is an EM and more specifically a simulated EM (SEM), as the expectation terms at the  $E$  step cannot be computed. The main modern way of carrying out likelihood inference in such situations is via a Markov chain Monte Carlo (MCMC) algorithm (see Chib [14] 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.

The idea behind the MCMC methods is that in order to sample a given probability distribution, which 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 the Markov chains on continuous state spaces [15]. 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 the 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 (e.g., 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 [16]. In this algorithm, the parameters are grouped into blocks, and each block is sampled according to the full conditional distribution denoted as



$\pi(\phi_t/\phi_{/t})$ . By Bayes' theorem, we have  $\pi(\phi_t/\phi_{/t}) \propto \pi(\phi_t\phi_{/t})$ , 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  $\{\phi_t\}_{t=1}^p$ , 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 (see [17]). Another application of the MCMC is the analysis of hidden Markov's models where the approach relies on augmenting the parameter space to include the unobserved states and simulate the target distribution via the conditional distributions (this procedure is called data augmentation and was pioneered by Tanner and Wong [18]). Kim et al. [19] discuss an 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.

The analysis of a time-varying GQARCH-M model becomes substantially complicated since the log-likelihood of the observed variables can no longer be written in closed form. In this paper, we focus on both the Bayesian and the classical estimation of the model. 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 [20], a regrettable consequence of this path dependence in volatility is that standard MCMC algorithms will evolve in  $O(T^2)$  computational load (see [21]). 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.

### 3.1. Estimation Problem: Simulated EM Algorithm

As mentioned already, the estimation problem arises because of the fact that we have two unobserved processes. More specifically, we cannot write down the likelihood function in closed form since we do not observe both  $\varepsilon_t$  and  $\delta_t$ . On the other hand, the conditional log-likelihood function of our model assuming that  $\delta_t$  were observed would be the following:

$$\begin{aligned} \ell(\mathbf{r}, \boldsymbol{\delta} \mid \boldsymbol{\phi}, \boldsymbol{\mathcal{T}}_0) &= \ln p(\mathbf{r} \mid \boldsymbol{\delta}, \boldsymbol{\phi}, \boldsymbol{\mathcal{T}}_0) + \ln p(\boldsymbol{\delta} \mid \boldsymbol{\phi}, \boldsymbol{\mathcal{T}}_0) \\ &= -T \ln 2\pi - \frac{1}{2} \sum_{t=1}^T \ln h_t - \frac{1}{2} \sum_{t=1}^T \frac{(\varepsilon_t)^2}{h_t} \\ &\quad - \frac{T}{2} \ln(\varphi_u^2) - \frac{1}{2} \sum_{t=1}^T \frac{(\delta_t - \delta(1 - \varphi) - \varphi\delta_{t-1})^2}{\varphi_u^2}, \end{aligned} \tag{3.1}$$

where  $\mathbf{r} = (r_1, \dots, r_T)'$ ,  $\boldsymbol{\delta} = (\delta_1, \dots, \delta_T)'$ , and  $\mathbf{h} = (h_1, \dots, h_T)'$ .

However, the  $\delta_t$ 's are unobserved, and, thus, to classically estimate the model, we have to rely on an EM algorithm [22] to obtain estimates as close to the optimum as desired. At each iteration, the EM algorithm obtains  $\boldsymbol{\phi}^{(n+1)}$ , where  $\boldsymbol{\phi}$  is the parameter vector, by maximizing



the expectation of the log-likelihood conditional on the data and the current parameter values, that is,  $E(\ell(\cdot) \mid \mathbf{r}, \phi^{(n)}, \mathcal{F}_0)$  with respect to  $\phi$  keeping  $\phi^{(n)}$  fixed.

The  $E$  step, thus, requires the expectation of the complete log-likelihood. For our model, this is given by:

$$\begin{aligned} E(\ell(\cdot) \mid \mathbf{r}, \phi^{(n)}, \mathcal{F}_0) = & -T \ln 2\pi - \frac{T}{2} \ln(\varphi_u^2) - \frac{1}{2} \sum_{t=1}^T E(\ln h_t \mid \mathbf{r}, \phi^{(n)}, \mathcal{F}_0) \\ & - \frac{1}{2} \sum_{t=1}^T E\left(\frac{(\varepsilon_t)^2}{h_t} \mid \mathbf{r}, \phi^{(n)}, \mathcal{F}_0\right) \\ & - \frac{1}{2} \sum_{t=1}^T E\left(\frac{(\delta_t - \delta(1-\varphi) - \varphi\delta_{t-1})^2}{\varphi_u^2} \mid \mathbf{r}, \phi^{(n)}, \mathcal{F}_0\right). \end{aligned} \quad (3.2)$$

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 an SEM or a simulated score. The SEM log-likelihood is:

$$\begin{aligned} \text{SEM}\ell = & -T \ln 2\pi - \frac{T}{2} \ln(\varphi_u^2) - \frac{1}{2} \frac{1}{M} \sum_{i=1}^M \sum_{t=1}^T \ln h_t^{(i)} - \frac{1}{2} \frac{1}{M} \sum_{i=1}^M \sum_{t=1}^T \frac{(\varepsilon_t^{(i)})^2}{h_t^{(i)}} - \frac{T}{2} \frac{(1-\varphi)^2 \delta^2}{\varphi_u^2} \\ & - \frac{1}{2} \frac{1}{\varphi_u^2} \frac{1}{M} \sum_{i=1}^M \sum_{t=1}^T (\delta_t^{(i)})^2 + \frac{(1-\varphi)\delta}{\varphi_u^2} \frac{1}{M} \sum_{i=1}^M \sum_{t=1}^T \delta_t^{(i)} + \frac{\varphi}{\varphi_u^2} \frac{1}{M} \sum_{i=1}^M \sum_{t=1}^T \delta_t^{(i)} \delta_{t-1}^{(i)} \\ & + \frac{(1-\varphi)\varphi\delta}{\varphi_u^2} \frac{1}{M} \sum_{i=1}^M \sum_{t=1}^T \delta_{t-1}^{(i)} - \frac{\varphi^2}{2\varphi_u^2} \frac{1}{M} \sum_{i=1}^M \sum_{t=1}^T (\delta_{t-1}^{(i)})^2. \end{aligned} \quad (3.3)$$

Consequently, we need to obtain the following quantities:  $(1/M) \sum_{i=1}^M \sum_{t=1}^T \ln h_t^{(i)}$ ,  $(1/M) \sum_{i=1}^M \sum_{t=1}^T ((\varepsilon_t^{(i)})^2 / h_t^{(i)})$ ,  $(1/M) \sum_{i=1}^M \sum_{t=1}^T \delta_t^{(i)}$ ,  $(1/M) \sum_{i=1}^M \sum_{t=1}^T \delta_{t-1}^{(i)}$ ,  $(1/M) \sum_{i=1}^M \sum_{t=1}^T \delta_t^{(i)} \delta_{t-1}^{(i)}$  and  $(1/M) \sum_{i=1}^M \sum_{t=1}^T (\delta_t^{(i)})^2$ ,  $(1/M) \sum_{i=1}^M \sum_{t=1}^T \delta_{t-1}^{2(i)}$ , where  $M$  is the number of simulations.

Thus, to classically estimate our model by using an SEM algorithm, the basic problem is to sample from  $\mathbf{h} \mid \phi, \mathbf{r}, \mathcal{F}_0$  where  $\phi$  is the vector of the unknown parameters and also sample from  $\delta \mid \phi, \mathbf{r}, \mathcal{F}_0$ .

In terms of identification, the model is not, up to second moment, identified (see Corollary 1 in Sentana and Fiorentini [23]). The reason is that we can transfer unconditional variance from the error,  $\varepsilon_t$ , to the price of risk,  $\delta_t$ , and vice versa. One possible solution is to fix  $\omega$  such that  $E(h_t)$  is 1 or to set  $\varphi_u$  to a specific value. In fact in an earlier version of the paper, we fixed  $\varphi_u$  to be 1 (see Anyfantaki and Demos [24]). Nevertheless, from a Bayesian viewpoint, the lack of identification is not too much of a problem, as the parameters are identified through their proper priors (see Poirier [25]).

Next, we will exploit the Bayesian estimation of the model, and, since we need to resort to simulations, we will show that the key task is again to simulate from  $\delta \mid \phi, \mathbf{r}, \mathcal{F}_0$ .

### 3.2. Simulation-Based Bayesian Inference

In our problem, the key issue is that the likelihood function of the sample  $p(\mathbf{r} \mid \phi, \mathcal{F}_0)$  is intractable which precludes the direct analysis of the posterior density  $p(\phi \mid \mathbf{r}, \mathcal{F}_0)$ . This problem may be overcome by focusing instead on the posterior density of the model using Bayes' rule:

$$p(\phi, \delta \mid \mathbf{r}) \propto p(\phi, \delta) p(\mathbf{r} \mid \phi, \delta) \propto p(\delta \mid \phi) p(\phi) p(\mathbf{r} \mid \phi, \delta), \quad (3.4)$$

where

$$\phi = (\delta, \varphi, \varphi_u, \alpha, \beta, \gamma, \omega)'. \quad (3.5)$$

Now,

$$p(\delta \mid \phi) = \prod_{t=1}^T p\left(\frac{\delta_t}{\delta}, \varphi, \varphi_u^2\right) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi\varphi_u^2}} \exp\left(-\frac{(\delta_t - \delta(1 - \varphi) - \varphi\delta_{t-1})^2}{2\varphi_u^2}\right). \quad (3.6)$$

On the other hand,

$$p(\mathbf{r} \mid \phi, \delta) = \prod_{t=1}^T p\left(\frac{r_t}{\{r_{t-1}\}}, \delta, \phi\right) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi}h_t} \exp\left(-\frac{\varepsilon_t^2}{2h_t}\right) \quad (3.7)$$

is the full-information likelihood. Once we have the posterior density, we get the parameters' marginal posterior density by integrating the posterior density. MCMC is one way of numerical integration.

The Hammersley-Clifford theorem (see Clifford [26]) says that a joint distribution can be characterized by its complete conditional distribution. Hence, given initial values  $\{\delta_t\}^{(0)}, \phi^{(0)}$ , we draw  $\{\delta_t\}^{(1)}$  from  $p(\{\delta_t\}^{(1)} \mid \mathbf{r}, \phi^{(0)})$  and then  $\phi^{(1)}$  from  $p(\phi^{(1)} \mid \{\delta_t\}^{(1)}, \mathbf{r})$ . Iterating these steps, we finally get  $(\{\delta_t\}^{(i)}, \phi^{(i)})_{i=1}^M$ , and under mild conditions it is shown that the distribution of the sequence converges to the joint posterior distribution  $p(\phi, \delta \mid \mathbf{r})$ .

The above simulation procedure may be carried out by first dividing the parameters into two blocks:

$$\begin{aligned} \phi_1 &= (\delta, \varphi, \varphi_u^2), \\ \phi_2 &= (\alpha, \beta, \gamma, \omega), \end{aligned} \quad (3.8)$$

Then the algorithm is described as follows.

- (1) Initialize  $\phi$ .
- (2) Draw from  $p(\delta_t \mid \delta_{\neq t}, \mathbf{r}, \phi)$ .
- (3) Draw from  $p(\phi \mid \delta, \mathbf{r})$  in the following blocks:
  - (i) draw from  $p(\phi_1 \mid \delta, \mathbf{r})$  using the Gibbs sampling. This is updated in one block;

(ii) draw from  $p(\phi_2 | \mathbf{r})$  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

The task of simulating from an AR model has been already discussed. Here, we will follow the approach of Chib [27], but we do not have any MA terms which makes inference simpler.

Suppose that the prior distribution of  $(\delta, \varphi_u^2, \varphi)$  is given by:

$$p(\delta, \varphi_u^2, \varphi) = p(\delta | \varphi_u^2) p(\varphi_u^2) p(\varphi), \quad (3.9)$$

which means that  $\delta, \varphi_u^2$  is a priory independent of  $\varphi$ .

Also the following holds for the prior distributions of the parameter subvector  $\phi_1$ :

$$\begin{aligned} p(\delta | \varphi_u^2) &\sim N(\delta_{pr}, \varphi_u^2 \sigma_{\delta_{pr}}^2), \\ p(\varphi_u^2) &\sim \text{IG}\left(\frac{v_0}{2}, \frac{d_0}{2}\right), \\ p(\varphi) &\sim N(\varphi_0, \sigma_{\varphi_0}^2) I_\varphi, \end{aligned} \quad (3.10)$$

where  $I_\varphi$  ensures that  $\varphi$  lies outside the unit circle, IG is the inverted gamma distribution, and the hyperparameters  $v_0, d_0, \delta_{pr}, \sigma_{\delta_{pr}}^2, \varphi_0, \sigma_{\varphi_0}^2$  have to be defined.

Now, the joint posterior is proportional to

$$\begin{aligned} p(\delta, \varphi, \varphi_u^2 | \mathbf{r}, \delta) &\propto \prod_{t=1}^T \frac{1}{\sqrt{2\pi\varphi_u^2}} \exp\left\{-\frac{(\delta_t - (1-\varphi)\delta - \varphi\delta_{t-1})^2}{2\varphi_u^2}\right\} \\ &\times N(\delta_{pr}, \varphi_u^2 \sigma_{\delta_{pr}}^2) \times \text{IG}\left(\frac{v_0}{2}, \frac{d_0}{2}\right) \times N(\varphi_0, \sigma_{\varphi_0}^2) I_\varphi. \end{aligned} \quad (3.11)$$

From a Bayesian viewpoint, the right-hand side of the above equation is equal to the “augmented” prior, that is, the prior augmented by the latent  $\delta$  (We would like to thank the associate editor for bringing this to our attention.) We proceed to the generation of these parameters.

#### Generation of $\delta$

First we see how to generate  $\delta$ . Following again Chib [27], we may write

$$\delta_t^* = \delta_t - \varphi\delta_{t-1}, \quad \delta_t^* | \mathcal{F}_{t-1} \sim N((1-\varphi)\delta, \varphi_u^2), \quad (3.12)$$

or, otherwise,

$$\delta_t^* = (1 - \varphi)\delta + v_t, \quad v_t \sim N(0, \varphi_u^2). \quad (3.13)$$

Under the above and using Chib's (1993) notation, we have that the proposal distribution is the following Gaussian distribution (see Chib [27] for a proof).

**Proposition 3.1.** *The proposal distribution of  $\delta$  is*

$$\delta \mid \boldsymbol{\delta}, \phi, \varphi_u^2 \sim N(\tilde{\delta}, \varphi_u^2 \tilde{\sigma}_\delta^2), \quad (3.14)$$

where

$$\begin{aligned} \tilde{\delta} &= \tilde{\sigma}_\delta^2 \left( \frac{\delta_{pr}}{\sigma_{\delta_{pr}}^2} + (1 - \varphi) \sum_{t=1}^T \delta_t^* \right), \\ \tilde{\sigma}_\delta^2 &= \left( \frac{1}{\sigma_{\delta_{pr}}^2} + (1 - \varphi)^2 \right)^{-1}. \end{aligned} \quad (3.15)$$

Hence, the generation of  $\delta$  is completed, and we may turn on the generation of the other parameters.

*Generation of  $\varphi_u^2$*

For the generation of  $\varphi_u^2$  and using [27] notation, we have the following.

**Proposition 3.2.** *The proposal distribution of  $\varphi_u^2$  is*

$$\varphi_u^2 \mid \boldsymbol{\delta}, \phi, \delta \sim \text{IG}\left(\frac{T - v_0}{2}, \frac{d_0 + Q + d}{2}\right), \quad (3.16)$$

where

$$\begin{aligned} Q &= (\delta - \delta_{pr})^2 \sigma_{\delta_{pr}}^{-2}, \\ d &= \sum_{t=2}^T [\delta_t^* - \delta(1 - \varphi)]^2. \end{aligned} \quad (3.17)$$

Finally, we turn on the generation of  $\varphi$ .

### Generation of $\varphi$

For the generation of  $\varphi$ , we follow again Chib [27] and write

$$\delta_t = (1 - \varphi)\delta - \varphi\delta_{t-1} + v_t, \quad v_t \sim N(0, \varphi_u^2). \quad (3.18)$$

We may now state the following proposition (see Chib [27] for a proof).

**Proposition 3.3.** *The proposal distribution of  $\varphi$  is*

$$\varphi^2 \mid \delta, \delta, \varphi_u^2 \sim N(\tilde{\varphi}, \tilde{\sigma}_\varphi^2), \quad (3.19)$$

where

$$\begin{aligned} \tilde{\varphi} &= \tilde{\sigma}_\varphi^2 \left( \sigma_{\varphi_0}^{-2} \varphi_0 + \varphi_u^{-2} \sum_{t=1}^T (\delta_{t-1} - \delta)(\delta_t - \delta) \right), \\ \tilde{\sigma}_\varphi^{-2} &= \sigma_{\varphi_0}^{-2} + \varphi_u^{-2} \sum_{t=1}^T (\delta_{t-1} - \delta)^2. \end{aligned} \quad (3.20)$$

The Gibbs sampling scheme has been completed, and the next step of the algorithm requires the generation of the conditional variance parameters via an M-H algorithm which is now presented.

### 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 et al. [19], Bauwens and Lubrano [28], Nakatsuma [29], Ardia [30] and others.

First, we need to decide on the priors. For the parameters  $\alpha, \beta, \gamma, \omega$ , we use normal densities as priors:

$$\begin{aligned} p(\alpha) &\sim N(\mu_\alpha, \Sigma_\alpha) I_\alpha, \\ p(\beta) &\sim N(\mu_\beta, \sigma_\beta^2) I_\beta \end{aligned} \quad (3.21)$$

and similarly

$$p(\gamma) \sim N(\mu_\gamma, \sigma_\gamma^2), \quad (3.22)$$

where  $\alpha = (\omega, \alpha)'$ ,  $I_\alpha, I_\beta$  are the indicators ensuring the constraints  $\alpha > 0$ ,  $\alpha + \beta < 1$  and  $\beta > 0$ ,  $\alpha + \beta < 1$ , respectively.  $\mu, \sigma^2$  are the hyperparameters.

We form the joint prior by assuming prior independence between  $\alpha, \beta, \gamma$ , and the joint posterior is then obtained by combining the joint prior and likelihood function by Bayes' rule:

$$p(\alpha, \beta, \gamma | \mathbf{r}) \propto \prod_{t=1}^T \frac{1}{\sqrt{2\pi h_t}} \exp\left(-\frac{\varepsilon_t^2}{2h_t}\right) \times N(\mu_\alpha, \Sigma_\alpha) I_\alpha \times N(\mu_\beta, \sigma_\beta^2) I_\beta \times N(\mu_\gamma, \sigma_\gamma^2). \quad (3.23)$$

For the M-H algorithm, we use the following approximated GARCH model as in Nakatsuma [29] which is derived by the well-known property of GARCH models [4]:

$$\varepsilon_t^2 = \omega + \alpha(\varepsilon_{t-1} - \gamma)^2 + \beta\varepsilon_{t-1}^2 + w_t - \beta w_{t-1}, \quad (3.24)$$

where  $w_t = \varepsilon_t^2 - h_t$  with  $w_t \sim N(0, 2h_t^2)$ .

Then the corresponding approximated likelihood is written as

$$p(\varepsilon^2 | \delta, \mathbf{r}, \phi_2) = \prod_{t=1}^T \frac{1}{2h_t\sqrt{\pi}} \exp\left[-\frac{(\varepsilon_t^2 - \omega - \alpha(\varepsilon_{t-1} - \gamma)^2 - \beta\varepsilon_{t-1}^2 + \beta w_{t-1})^2}{4h_t^2}\right], \quad (3.25)$$

and the generation of  $\alpha, \beta, \gamma$  is based on the above likelihood where we update  $\{h_t\}$  each time after the corresponding parameters are updated. The generation of the four variance parameters is given.

#### Generation of $\alpha$

For the generation of  $\alpha$ , we first note that  $w_t$  in (3.32), below, can be written as a linear function of  $\alpha$ :

$$w_t = \bar{\varepsilon}_t^2 - \zeta_t \alpha, \quad (3.26)$$

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

$$\begin{aligned} \bar{\varepsilon}_t^2 &= \tilde{\varepsilon}_t^2 - \beta \tilde{\varepsilon}_{t-1}^2, \\ \tilde{\varepsilon}_t^2 &= \varepsilon_t^2 + \beta \tilde{\varepsilon}_{t-1}^2, \\ \hat{\varepsilon}_t^2 &= (\varepsilon_{t-1} - \gamma)^2 + \beta \hat{\varepsilon}_{t-1}^2, \\ \tilde{l}_t &= 1 + \beta \tilde{l}_{t-1}. \end{aligned} \quad (3.27)$$

Now, let the two following vectors be

$$\begin{aligned} Y_\alpha &= [\bar{\varepsilon}_1^2, \dots, \bar{\varepsilon}_T^2]', \\ X_\alpha &= [\zeta_1', \dots, \zeta_T']'. \end{aligned} \quad (3.28)$$

Then the likelihood function of the approximated model is rewritten as

$$p(\varepsilon^2 \mid \mathbf{r}, \boldsymbol{\delta}, \phi_2) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi(2h_t^2)}} \exp \left[ -\frac{(\bar{\varepsilon}_t^2 - \zeta_t \boldsymbol{\alpha})^2}{2(2h_t^2)} \right]. \quad (3.29)$$

Using this we have the following proposal distribution of  $\boldsymbol{\alpha}$  (see Nakatsuma [29] or Ardia [30] for a proof).

**Proposition 3.4.** *The proposal distribution of  $\boldsymbol{\alpha}$  is*

$$\boldsymbol{\alpha} \mid Y, X, \Sigma, \phi_{2-\boldsymbol{\alpha}} \sim N(\hat{\mu}_{\boldsymbol{\alpha}}, \hat{\Sigma}_{\boldsymbol{\alpha}}) I_{\boldsymbol{\alpha}}, \quad (3.30)$$

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

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

$$\min \left\{ \frac{p(\tilde{\boldsymbol{\alpha}}, \beta, \gamma \mid \boldsymbol{\delta}, \mathbf{r}) q(\boldsymbol{\alpha}^* \mid \tilde{\boldsymbol{\alpha}}, \beta, \gamma, \boldsymbol{\delta}, \mathbf{r})}{p(\boldsymbol{\alpha}^*, \beta, \gamma \mid \boldsymbol{\delta}, \mathbf{r}) q(\tilde{\boldsymbol{\alpha}} \mid \boldsymbol{\alpha}^*, \beta, \gamma, \boldsymbol{\delta}, \mathbf{r})}, 1 \right\}, \quad (3.31)$$

where  $\boldsymbol{\alpha}^*$  is the previous draw.

Similar procedure is used for the generation of  $\beta$  and  $\gamma$ .

### Generation of $\beta$

Following Nakatsuma [29], we linearize  $w_t$  by the first-order Taylor expansion

$$w_t(\beta) \approx w_t(\beta^*) + \xi_t(\beta^*)(\beta - \beta^*), \quad (3.32)$$

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

Define as

$$r_t = w_t(\beta^*) + g_t(\beta^*)\beta^*, \quad (3.33)$$

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

$$g_t = \varepsilon_{t-1}^2 - w_{t-1} + \beta^* g_{t-1}, \quad (3.34)$$

$\xi_t = 0$  for  $t \leq 0$  [30]. Then,

$$w_t(\beta) \approx r_t - g_t(\beta)\beta. \quad (3.35)$$



Let the following two vectors be

$$\begin{aligned} Y_\beta &= [r_1, \dots, r_T]', \\ X_\beta &= [g_1, \dots, g_T]'. \end{aligned} \quad (3.36)$$

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

$$p(\varepsilon^2 \mid \delta, \mathbf{r}, \phi_2) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi(2h_t^2)}} \exp \left[ -\frac{\{w_t(\beta^*) + \xi_t(\beta^*)(\beta - \beta^*)\}^2}{2(2h_t^2)} \right]. \quad (3.37)$$

We have the following proposal distribution for  $\beta$  (for a proof see Nakatsuma [29] or Ardia [30]).

**Proposition 3.5.** *The proposal distribution for  $\beta$  is*

$$\beta \mid Y, X, \sigma_\beta^2, \phi_{2-\beta} \sim N(\hat{\mu}_\beta, \hat{\sigma}_\beta^2) I_\beta, \quad (3.38)$$

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

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

$$\min \left\{ \frac{p(\tilde{\beta}, \alpha, \gamma \mid \delta, \mathbf{r}) q(\beta^* \mid \tilde{\beta}, \alpha, \gamma, \delta, \mathbf{r})}{p(\beta^*, \alpha, \gamma \mid \delta, \mathbf{r}) q(\tilde{\beta} \mid \beta^*, \alpha, \gamma, \delta, \mathbf{r})}, 1 \right\} \quad (3.39)$$

Finally, we explain the generation of  $\gamma$ .

#### Generation of $\gamma$

As with  $\beta$ , we linearize  $w_t$  by a first-order Taylor, expansion at a point  $\gamma^*$  the previous draw in the M-H sampler. In this case,

$$r_t = w_t(\gamma^*) - g_t(\gamma^*)\gamma^*, \quad (3.40)$$

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

$$g_t = -2\alpha(\varepsilon_{t-1} - \gamma^*) + \beta g_{t-1}, \quad (3.41)$$

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

Then

$$w_t(\gamma) \approx r_t - g_t \gamma. \quad (3.42)$$

Let again

$$\begin{aligned} Y_Y &= [r_1, \dots, r_T]', \\ X_Y &= [g_1, \dots, g_T]' \end{aligned} \quad (3.43)$$

and the likelihood function of the approximated model is rewritten as

$$p(\varepsilon^2 \mid \delta, \mathbf{r}, \phi_2) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi(2h_t^2)}} \exp \left[ -\frac{\{w_t(\gamma^*) - g_t(\gamma^*)\gamma^*\}^2}{2(2h_t^2)} \right]. \quad (3.44)$$

Thus, we have the following proposal distribution for  $\gamma$  (for proof see Nakatsuma [29] and Ardia [30]).

**Proposition 3.6.** *The proposal distribution of  $\gamma$  is*

$$\gamma \mid Y, X, \sigma_Y^2, \phi_{2-Y} \sim N(\hat{\mu}_Y, \hat{\sigma}_Y^2), \quad (3.45)$$

where  $\hat{\mu}_Y = \hat{\sigma}_Y^2(X_Y' \Lambda^{-1} Y_Y + (\mu_Y / \sigma_Y^2))$ ,  $\hat{\sigma}_Y^2 = (X_Y' \Lambda^{-1} X_Y + (1 / \sigma_Y^2))^{-1}$ , and  $\Lambda = \text{diag}(2h_1^4, \dots, 2h_T^4)$ . A candidate  $\tilde{\gamma}$  is sampled from this proposal density and accepted with probability:

$$\min \left\{ \frac{p(\tilde{\gamma}, \alpha, \beta \mid \delta, \mathbf{r}) q(\gamma^* \mid \tilde{\gamma}, \alpha, \beta, \delta, \mathbf{r})}{p(\gamma^*, \alpha, \beta \mid \delta, \mathbf{r}) q(\tilde{\gamma} \mid \gamma^*, \alpha, \beta, \delta, \mathbf{r})}, 1 \right\}. \quad (3.46)$$

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

All that remains, therefore, is step (2). Thus, from the above, it is seen that the main task is again as with the classical estimation of the model, to simulate from  $\delta \mid \phi, \mathbf{r}, \mathcal{F}_0$ .

### 3.2.3. MCMC Simulation of $\varepsilon \mid \phi, \mathbf{r}, \mathcal{F}_0$

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  $\varepsilon_t$  given  $\mathbf{r}$  and  $\phi$ . We set the mean and the variance of  $\varepsilon_0$  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  $\phi$ ,  $\mathcal{F}_0$  can be eliminated from the information set without any information loss.

Now sampling from  $p(\varepsilon \mid \mathbf{r}, \phi) \propto p(\mathbf{r} \mid \varepsilon, \phi)p(\varepsilon \mid \phi)$  is feasible by using an M-H algorithm where we update each time only one  $\varepsilon_t$  leaving all the other unchanged [20]. In particular, let us write the  $n$ th iteration of a Markov chain as  $\varepsilon_t^n$ . Then we generate a potential new value of the Markov chain  $\varepsilon_t^{\text{new}}$  by proposing from some candidate density  $g(\varepsilon_t \mid \varepsilon_{\setminus t}^n, \mathbf{r}, \phi)$  where  $\varepsilon_{\setminus t}^n = \{\varepsilon_1^{n+1}, \dots, \varepsilon_{t-1}^{n+1}, \varepsilon_{t+1}^n, \dots, \varepsilon_T^n\}$  which we accept with probability

$$\min \left[ 1, \frac{p(\varepsilon_t^{\text{new}} \mid \varepsilon_{\setminus t}^n, \mathbf{r}, \phi)g(\varepsilon_t^{\text{new}} \mid \varepsilon_{\setminus t}^n, \mathbf{r}, \phi)}{p(\varepsilon_t^n \mid \varepsilon_{\setminus t}^n, \mathbf{r}, \phi)g(\varepsilon_t^n \mid \varepsilon_{\setminus t}^n, \mathbf{r}, \phi)} \right]. \quad (3.47)$$

If it is accepted then, we set  $\varepsilon_t^{n+1} = \varepsilon_t^{\text{new}}$  and otherwise we keep  $\varepsilon_t^{n+1} = \varepsilon_t^n$ . 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:

$$\begin{aligned} \frac{p(\varepsilon_t^{\text{new}} \mid \varepsilon_{\setminus t}^n, \mathbf{r}, \phi)}{p(\varepsilon_t^n \mid \varepsilon_{\setminus t}^n, \mathbf{r}, \phi)} &= \frac{p(r_t \mid \varepsilon_t^{\text{new}}, h_t^{\text{new}, t}, \phi)p(\varepsilon_t^{\text{new}} \mid h_t^{\text{new}, t}, \phi)p(r_t \mid h_t^{n, t}, \phi)}{p(r_t \mid h_t^{\text{new}, t}, \phi)p(r_t \mid \varepsilon_t^n, h_t^{n, t}, \phi)p(\varepsilon_t^{\text{new}} \mid h_t^{n, t}, \phi)} \\ &\quad * \prod_{s=t+1}^T \frac{p(r_s \mid \varepsilon_s^r, h_s^{\text{new}, t}, \phi)p(\varepsilon_s^r \mid h_s^{\text{new}, t}, \phi)p(r_s \mid h_s^{n, t}, \phi)}{p(r_s \mid h_s^{\text{new}, t}, \phi)p(r_s \mid \varepsilon_s^n, h_s^{n, t}, \phi)p(\varepsilon_s^n \mid h_s^{n, t}, \phi)} \\ &= \frac{p(r_t \mid \varepsilon_t^{\text{new}}, h_t^{\text{new}, t}, \phi)p(\varepsilon_t^{\text{new}} \mid h_t^{\text{new}, t}, \phi)}{p(r_t \mid \varepsilon_t^n, h_t^{n, t}, \phi)p(\varepsilon_t^{\text{new}} \mid h_t^{n, t}, \phi)} \\ &\quad * \prod_{s=t+1}^T \frac{p(r_s \mid \varepsilon_s^n, h_s^{\text{new}, t}, \phi)p(\varepsilon_s^n \mid h_s^{\text{new}, t}, \phi)}{p(r_s \mid \varepsilon_s^n, h_s^{n, t}, \phi)p(\varepsilon_s^n \mid h_s^{n, t}, \phi)}, \end{aligned} \quad (3.48)$$

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

$$\begin{aligned} h_s^{\text{new}, t} &= V(\varepsilon_s \mid \varepsilon_{s-1}^n, \varepsilon_{s-2}^n, \dots, \varepsilon_{t+1}^n, \varepsilon_t^{\text{new}}, \varepsilon_{t-1}^{n+1}, \dots, \varepsilon_1^{n+1}), \\ h_s^{n, t} &= V(\varepsilon_s \mid \varepsilon_{s-1}^n, \varepsilon_{s-2}^n, \dots, \varepsilon_{t+1}^n, \varepsilon_t^n, \varepsilon_{t-1}^{n+1}, \dots, \varepsilon_1^{n+1}) \end{aligned} \quad (3.49)$$

while

$$h_t^{\text{new}, t} = h_t^{n, t}. \quad (3.50)$$

Nevertheless, each time we revise one  $\varepsilon_t$ , we have also to revise  $T - t$  conditional variances because of the recursive nature of the GARCH model which makes  $h_s^{\text{new}, t}$  depend upon  $\varepsilon_t^{\text{new}}$  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  $\phi$ . To

avoid this huge computational load, we show how to use the method proposed by Fiorentini et al. [10] and so do MCMC with only  $O(T)$  calculations. The method is described in the following subsection.

### 3.3. Estimation Method Proposed: Classical and Bayesian Estimation

The method proposed by Fiorentini et al. [10] is to transform the GARCH model into a first-order Markov's 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  $\{h_{t+1}, s_t\} \mid \mathbf{r}, \phi \in \mathcal{F}_t$  where

$$s_t = \text{sign}(\varepsilon_t - \gamma), \quad (3.51)$$

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  $\phi$ , then we know the value of

$$(\varepsilon_t - \gamma)^2 = \frac{h_{t+1} - \omega - \beta h_t}{\alpha} \quad \forall t \geq 1. \quad (3.52)$$

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  $\{\delta_t\} \mid \mathbf{r}, \phi, \{h_{t+1}\}$ .

Now we have to sample from

$$p(\{s_t, h_{t+1}\} \mid \mathbf{r}, \phi) \propto \prod_{t=1}^T p(s_t \mid h_{t+1}, h_t, \phi) p(h_{t+1} \mid h_t, \phi) p(r_t \mid s_t, h_t, h_{t+1}, \phi), \quad (3.53)$$

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

$$\varepsilon_t = \gamma \pm d_t, \quad (3.54)$$

where

$$d_t = \sqrt{\frac{h_{t+1} - \omega - \beta h_t}{\alpha}}. \quad (3.55)$$

From the above, it is seen that we should first simulate  $\{h_{t+1}\} \mid \mathbf{r}, \phi$  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  $\varepsilon_t$  and then simulate  $\{s_t\} \mid \{h_{t+1}\}, \mathbf{r}, \phi$ . The second step is a Gibbs sampling scheme whose acceptance rate is always one and also conditional on  $\{h_{t+1}\}, \mathbf{r}, \phi$  the elements of  $\{s_t\}$  are independent which further simplifies the calculations. We prefer to review first the Gibbs sampling scheme and then the simulation of the conditional variance.

### 3.3.1. Simulations of $\{s_t\} \mid \{h_{t+1}\}, \mathbf{r}, \phi$

First, we see how to sample from  $\{s_t\} \mid \{h_{t+1}\}, \mathbf{r}, \phi$ . To obtain the required conditionally Bernoulli distribution, we establish first some notation. We have the following (see Appendix A):

$$c_t = \frac{1}{\sqrt{v_{t|r_t, h_t}}} \left[ \varphi\left(\frac{\gamma + d_t - \varepsilon_{t|r_t, h_t}}{\sqrt{v_{t|r_t, h_t}}}\right) + \varphi\left(\frac{\gamma - d_t - \varepsilon_{t|r_t, h_t}}{\sqrt{v_{t|r_t, h_t}}}\right) \right], \quad (3.56)$$

$$\varepsilon_{t|r_t, h_t} = E(\varepsilon_t \mid r_t, h_t) = \frac{(1 - \varphi^2)(r_t - \delta h_t)}{\varphi_u^2 h_t + 1 - \varphi^2}, \quad v_{t|r_t, h_t} = \text{Var}(\varepsilon_t \mid r_t, h_t) = \frac{\varphi_u^2 h_t^2}{\varphi_u^2 h_t + 1 - \varphi^2}.$$

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  $h_{t+1}, h_t, r_t, \phi$ , where  $d_t$  is given by (3.70), which is given by

$$\begin{aligned} p(s_t = 1 \mid \{h_{t+1}\}, \mathbf{r}, \phi) &= p(\varepsilon_t = \gamma + d_t \mid h_{t+1}, h_t, r_t, \phi) \\ &= \frac{1}{c_t \sqrt{v_{t|r_t, h_t}}} \varphi\left(\frac{\gamma + d_t - \varepsilon_{t/r_t, h_t}}{\sqrt{v_{t|r_t, h_t}}}\right). \end{aligned} \quad (3.57)$$

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

$$\varphi\left(\frac{\gamma + d_t - \varepsilon_{t/r_t, h_t}}{\sqrt{v_{t|r_t, h_t}}}\right) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\gamma + d_t - \varepsilon_{t/r_t, h_t}}{\sqrt{v_{t|r_t, h_t}}}\right)^2\right\} \quad (3.58)$$

and so we may simulate  $\{s_t\} \mid \{h_{t+1}\}, \mathbf{r}, \phi$  using a Gibbs sampling scheme. Specifically, since conditional on  $\{h_{t+1}\}, \mathbf{r}, \phi$  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  $\{s_t\} \mid \{h_{t+1}\}, \mathbf{r}, \phi$  may be described as below.

- (1) Specify an initial value  $s^{(0)} = (s_1^{(0)}, \dots, s_T^{(0)})$ .
- (2) Repeat for  $k = 1, \dots, M$ .
  - (a) Repeat for  $t = 0, \dots, T - 1$ .
    - (i) Draw  $s^{(k)} = 1$  with probability,

$$\frac{1}{c_t \sqrt{v_{t|r_t, h_t}}} \varphi\left(\frac{\gamma + d_t - \varepsilon_{t/r_t, h_t}}{\sqrt{v_{t|r_t, h_t}}}\right), \quad (3.59)$$

and  $s^{(k)} = -1$  with probability,

$$1 - \frac{1}{c_t \sqrt{v_{t|r_t, h_t}}} \varphi\left(\frac{\gamma + d_t - \varepsilon_{t/r_t, h_t}}{\sqrt{v_{t|r_t, h_t}}}\right). \quad (3.60)$$

- (3) Return the values  $\{s^{(1)}, \dots, s^{(M)}\}$ .

### 3.3.2. Simulations of $\{h_{t+1}\} \mid \mathbf{r}, \phi$ (Single Move Samplers)

On the other hand, the first step involves simulating from  $\{h_{t+1}\} \mid \mathbf{r}, \phi$ . To avoid large dependence in the chain, we use an M-H algorithm where we simulate one  $h_{t+1}$  at a time leaving the others unchanged [20, 31]. So if  $(h_{t+1})^n$  is the current value of the  $n$ th iteration of a Markov chain, then we draw a candidate value of the Markov chain  $h_{t+1}^{\text{new}}$  by proposing it from a candidate density (proposal density)  $g(h_{t+1} \mid (h)_{/t+1}^n, \mathbf{r}, \phi)$  where  $(h)_{/t+1}^n = \{h_1^{n+1}, h_2^{n+1}, \dots, h_t^{n+1}, h_{t+2}^n, \dots, h_{T+1}^n\}$ . We set  $(h_{t+1})^{n+1} = (h_{t+1})^{\text{new}}$  with acceptance probability

$$\min \left[ 1, \frac{p(h_{t+1}^{\text{new}} \mid (h)_{/t+1}^n, \mathbf{r}, \phi) g(h_{t+1}^n \mid (h)_{/t+1}^n, \mathbf{r}, \phi)}{p(h_{t+1}^n \mid (h)_{/t+1}^n, \mathbf{r}, \phi) g(h_{t+1}^{\text{new}} \mid (h)_{/t+1}^n, \mathbf{r}, \phi)} \right], \quad (3.61)$$

where we have used the fact that

$$p(\mathbf{h} \mid \mathbf{r}, \phi) = p((h)_{/t} \mid \mathbf{r}, \phi) p(h_t \mid (h)_{/t}, \mathbf{r}, \phi). \quad (3.62)$$

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

$$p(h_{t+1} \mid (h)_{/t+1}, \mathbf{r}, \phi) \propto p(h_{t+2} \mid h_{t+1}, \phi) p(h_{t+1} \mid h_t, \phi) p(r_{t+1} \mid h_{t+2}, h_{t+1}, \phi) p(r_t \mid h_{t+1}, h_t, \phi). \quad (3.63)$$

Now, since the following should hold:

$$h_{t+1} \geq \omega + \beta h_t \quad (3.64)$$

and similarly

$$h_{t+1} \leq \beta^{-1}(h_{t+2} - \omega), \quad (3.65)$$

we have the support of the conditional distribution of  $h_{t+1}$  given that  $h_t$  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; that is:

$$h_{t+1} \in [\omega + \beta h_t, \beta^{-1}(h_{t+2} - \omega)]. \quad (3.66)$$

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

$$g(h_{t+1} \mid (h)_{/t+1}, \mathbf{r}, \phi) = p(h_{t+1} \mid h_t, \phi) \quad (3.67)$$

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 et al. [10] we can achieve a substantially higher acceptance rate if we propose from

$$g(h_{t+1} | (h)_{/t+1}, \mathbf{r}, \phi) = p(h_{t+1} | r_t, h_t, \phi). \quad (3.68)$$

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

$$\varepsilon_t | r_t, h_t, \phi \sim N\left(\frac{(1 - \varphi^2)(r_t - \delta h_t)}{\varphi_u^2 h_t + 1 - \varphi^2}, \frac{\varphi_u^2 h_t^2}{\varphi_u^2 h_t + 1 - \varphi^2}\right) \quad (3.69)$$

doubly truncated so that it remains within the following bounds:

$$\varepsilon_t^{\text{new}} \in [\gamma - l_t, \gamma + l_t], \quad (3.70)$$

where

$$l_t = \sqrt{\frac{h_{t+2} - \omega - \beta\omega - \beta^2 h_t}{\beta\alpha}}, \quad (3.71)$$

using an inverse transform method and then compute

$$h_{t+1}^{\text{new}} = \omega + \alpha(\varepsilon_t^{\text{new}} - \gamma)^2 + \beta h_t, \quad (3.72)$$

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

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

$$u \sim U(0, 1) \quad (3.73)$$

and then computes the following:

$$\begin{aligned} \bar{u} = (1 - u)\Phi\left(\frac{\gamma - l_t - ((1 - \varphi^2)(r_t - \delta h_t)/(\varphi_u^2 h_t + 1 - \varphi^2))}{\sqrt{\varphi_u^2 h_t^2 / (\varphi_u^2 h_t + 1 - \varphi^2)}}\right) \\ + u\Phi\left(\frac{\gamma + l_t - ((1 - \varphi^2)(r_t - \delta h_t)/(\varphi_u^2 h_t + 1 - \varphi^2))}{\sqrt{\varphi_u^2 h_t^2 / (\varphi_u^2 h_t + 1 - \varphi^2)}}\right). \end{aligned} \quad (3.74)$$



A draw is then given by

$$\varepsilon_t^{\text{new}} = \Phi^{-1}(\bar{u}). \quad (3.75)$$

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

$$\varepsilon_t^{\text{new}} \mid r_t, h_t, \phi \sim N\left(\frac{(1-\varphi^2)(r_t - \delta h_t)}{\varphi_u^2 h_t + 1 - \varphi^2}, \frac{\varphi_u^2 h_t^2}{\varphi_u^2 h_t + 1 - \varphi^2}\right) \quad (3.76)$$

and accept the draw if  $\gamma - l_t \leq \varepsilon_t^{\text{new}} \leq \gamma + l_t$ , 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  $\varepsilon_t^{\text{new}}$  will be given according to the definition of a truncated normal distribution:

$$\begin{aligned} p(\varepsilon_t^{\text{new}} \mid |\varepsilon_t^{\text{new}} - \gamma| \leq l_t, r_t, h_t, \phi) &= \frac{1}{\sqrt{v_{t/r_t, h_t}}} \varphi\left(\frac{\varepsilon_t^{\text{new}} - \varepsilon_{t/r_t, h_t}}{\sqrt{v_{t/r_t, h_t}}}\right) \\ &\times \left[ \Phi\left(\frac{\gamma + l_t - \varepsilon_{t/r_t, h_t}}{\sqrt{v_{t/r_t, h_t}}}\right) - \Phi\left(\frac{\gamma - l_t - \varepsilon_{t/r_t, h_t}}{\sqrt{v_{t/r_t, h_t}}}\right) \right]^{-1}, \end{aligned} \quad (3.77)$$

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

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

$$\begin{aligned} p(h_{t+1}^{\text{new}} \mid h_{t+1}^{\text{new}} \in [\omega + \beta h_t, \beta^{-1}(h_{t+2} - \omega)], r_t, h_t, \phi) \\ = \frac{c_t^{\text{new}}}{|2\alpha d_t^{\text{new}}|} \times \left[ \Phi\left(\frac{\gamma + l_t - \varepsilon_{t/r_t, h_t}}{\sqrt{v_{t/r_t, h_t}}}\right) - \Phi\left(\frac{\gamma - l_t - \varepsilon_{t/r_t, h_t}}{\sqrt{v_{t/r_t, h_t}}}\right) \right]^{-1}. \end{aligned} \quad (3.78)$$

Using Bayes theorem we have that the acceptance probability will be

$$\min\left(1, \frac{p(h_{t+2} \mid h_{t+1}^{\text{new}}, r_{t+1}, \phi) p(r_{t+1} \mid h_{t+1}^{\text{new}}, \phi)}{p(r_{t+1} \mid h_{t+1}^{\text{new}}, \phi) p(h_{t+2} \mid h_{t+1}^{\text{new}}, r_{t+1}, \phi)}\right). \quad (3.79)$$

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

$$\min\left(1, \frac{p(r_{t+1} \mid h_{t+1}^{\text{new}}) \frac{c_{t+1}^{\text{new}}}{c_{t+1}^{\text{new}}} \frac{d_{t+1}^{\text{new}}}{d_{t+1}^{\text{new}}}}{p(r_{t+1} \mid h_{t+1}^{\text{new}}) \frac{c_{t+1}^{\text{new}}}{c_{t+1}^{\text{new}}} \frac{d_{t+1}^{\text{new}}}{d_{t+1}^{\text{new}}}}\right), \quad (3.80)$$

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

$$r_{t+1} | h_{t+1} \sim N\left(\delta h_{t+1}, \left(\frac{\varphi_u^2}{1 - \varphi^2} h_{t+1} + 1\right) h_{t+1}\right). \quad (3.81)$$

Hence,

$$p(r_{t+1} | h_{t+1}^n) = \frac{1}{\sqrt{2\pi((\varphi_u^2/(1 - \varphi^2))h_{t+1}^n + 1)h_{t+1}^n}} \exp\left(-\frac{(r_{t+1} - \delta h_{t+1}^n)^2}{2((\varphi_u^2/(1 - \varphi^2))h_{t+1}^n + 1)h_{t+1}^n}\right), \quad (3.82)$$

and the acceptance probability becomes

$$\min\left[1, \left(\frac{h_{t+1}^n}{h_{t+1}^{\text{new}}}\right)^{3/2} \frac{\sqrt{h_{t+2}^n - \omega - \beta h_{t+1}^n} \kappa(h_{t+1}^{\text{new}})}{\sqrt{h_{t+2}^n - \omega - \beta h_{t+1}^{\text{new}}} \kappa(h_{t+1}^n)}\right], \quad (3.83)$$

where

$$\begin{aligned} \kappa(h_{t+1}^i) = & \exp\left[-\frac{(r_{t+1} - \delta h_{t+1}^i)^2}{2\left(\frac{\varphi_u^2}{1 - \varphi^2} h_{t+1}^i + 1\right) h_{t+1}^i} - \frac{\frac{\varphi_u^2}{1 - \varphi^2} h_{t+1}^i + 1}{2\frac{\varphi_u^2}{1 - \varphi^2} (h_{t+1}^i)^2} \left(\left(\gamma - \frac{r_{t+1} - \delta h_{t+1}^i}{\frac{\varphi_u^2}{1 - \varphi^2} h_{t+1}^i + 1}\right)^2\right)\right] \\ & \times \frac{\left[1 + \exp\left[2\frac{\frac{\varphi_u^2}{1 - \varphi^2} h_{t+1}^i + 1}{\frac{\varphi_u^2}{1 - \varphi^2} (h_{t+1}^i)^2} \left(\gamma - \frac{r_{t+1} - \delta h_{t+1}^i}{\frac{\varphi_u^2}{1 - \varphi^2} h_{t+1}^i + 1}\right) \sqrt{\frac{h_{t+2}^n - \omega - \beta h_{t+1}^i}{\alpha}}\right]\right]}{\left[\exp\left[\frac{\frac{\varphi_u^2}{1 - \varphi^2} h_{t+1}^i + 1}{\frac{\varphi_u^2}{1 - \varphi^2} (h_{t+1}^i)^2} \left(\gamma - \frac{r_{t+1} - \delta h_{t+1}^i}{\frac{\varphi_u^2}{1 - \varphi^2} h_{t+1}^i + 1}\right) \sqrt{\frac{h_{t+2}^n - \omega - \beta h_{t+1}^i}{\alpha}}\right]\right]}. \end{aligned} \quad (3.84)$$

Overall the MCMC of  $\{h_{t+1}\} | \mathbf{r}, \phi$  includes the following steps.

- (1) Specify an initial value  $\{h^{(0)}\}$ .
- (2) Repeat for  $n = 1, \dots, M$ .
  - (a) Repeat for  $t = 0, \dots, T - 1$ .

(i) Use an inverse transform method to simulate

$$\varepsilon_t^{\text{new}} \mid r_t, h_t, \phi \sim N\left(\frac{(1 - \varphi^2)(r_t - \delta h_t)}{\varphi_u^2 h_t + 1 - \varphi^2}, \frac{\varphi_u^2 h_t^2}{\varphi_u^2 h_t + 1 - \varphi^2}\right) \quad (3.85)$$

doubly truncated.

(ii) Calculate

$$h_{t+1}^{\text{new}} = \omega + \alpha(\varepsilon_t^c - \gamma)^2 + \beta h_t. \quad (3.86)$$

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

$$(h_{t+1})^{\text{new}} \sim p(h_{t+1}^{\text{new}} \mid r_t, h_t, \phi) \quad (3.87)$$

appropriately truncated.

(iii) Calculate

$$\alpha_r = \min\left[1, \frac{p(r_{t+1} \mid h_{t+1}^{\text{new}}) \frac{c_{t+1}^{\text{new}}}{c_{t+1}^n} \frac{d_{t+1}^n}{d_{t+1}^{\text{new}}}}{p(r_{t+1} \mid h_{t+1}^n)}\right]. \quad (3.88)$$

(iv) Set

$$(h_{t+1})^{n+1} = \begin{cases} (h_{t+1})^{\text{new}} & \text{if } \text{Unif}(0, 1) \leq \alpha_r \\ (h_{t+1})^n & \text{otherwise} \end{cases} \quad (3.89)$$

*Remark 3.7.* Every time 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  $h_{t+1}^{\text{new}}$ , then  $\varepsilon_t^{\text{new}}$  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  $\varepsilon_T$ . The final step involves computing

$$\delta_{t+1}^{(n)} = \frac{r_{t+1} - \varepsilon_{t+1}^{(n)}}{h_{t+1}^{(n)}}, \quad t = 0, \dots, T - 1, \quad n = 1, \dots, M. \quad (3.90)$$

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 (2) 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 an 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.

### 3.4. A Comparison of the Simulators

In order to compare the performance of the inefficient and the efficient MCMC sampler introduced in the previous subsection, we have generated realizations of size  $T = 240$  for the simple GQARCH(1,1)-M-AR(1) model with parameters  $\delta = 0.1$ ,  $\varphi = 0.85$ ,  $\alpha = 0.084$ ,  $\beta = 0.688$ ,  $\gamma = 0.314$  (which are centered around typical values that we tend to see in the empirical literature). We first examine the increase in the variance of the sample mean of  $\varepsilon_t$  across 500,000 simulations due to the autocorrelation in the drawings relative to an ideal but infeasible independent sampler.

We do so by recording the inefficient ratios for the observations  $t = 80$  and  $t = 160$  using standard spectral density techniques. In addition, we record the mean acceptance probabilities over all observations and the average CPU time needed to simulate one complete drawing. The behavior of the two simulators is summarized in Table 1 and is very much as one would expect. The computationally inefficient sampler shows high serial correlation for both  $t = 80$  and  $t = 160$  and a low acceptance rate for each individual  $t$ . Moreover, it is extremely time consuming to compute even though our sample size is fairly small. In fact, when we increase  $T$  from 240 to 2400, and 24000 the average CPU time increases by a factor of 100 and 10000, respectively, as opposed to 10 and 100 for the other one (the efficient), which makes it impossible to implement in most cases of practical interest. On the other hand, the single-move efficient sampler produces results much faster, with a reasonably high acceptance rate but more autocorrelation in the drawings for  $t = 160$ .

## 4. Empirical Application: Bayesian Estimation of Weekly Excess Returns from Three Major Stock Markets: Dow-Jones, FTSE, and Nikkei

In this section we apply the procedures described above to weekly excess returns from three major stock markets: Dow-Jones, FTSE, and Nikkei for the period of the last week of 1979:8 to the second to the last week of 2008:5 (1,500 observations). To guarantee  $0 \leq \beta \leq 1 - \alpha \leq 1$  and to ensure that  $\omega > 0$  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  $\omega > 0$ .

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 is to employ informative priors that would be in accordance with the “received wisdom.” In particular, for all data sets, we set the prior mean for  $\beta$  equal to 0.7, and for  $\alpha$ ,  $\omega$ , and  $\gamma$  we decided to set their prior means equal to 0.15, 0.4, and 0.0, respectively. We had also to decide on the prior mean of  $\delta$ . We set its prior mean equal to 0.05, for all markets. These prior means imply an annual excess return of around 4%, which is a typical value for annualized stock excess returns. Finally, we set the prior mean of  $\varphi$  equal to 0.75, of  $\varphi_u^2$  equal to 0.01, and the hyperparameters  $\nu_0$  and  $d_0$  equal to 1550 and 3, respectively, for all three datasets, something which is consistent with the “common wisdom” of high autocorrelation of the price of risk. We employ a rather vague prior and set its prior variance equal to 10,000 for all datasets.

We ran a chain for 200,000 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 Dow-Jones, FTSE, and Nikkei are reported in Table 2. Inefficiency

**Table 1:** Comparison between the efficient and inefficient MCMC simulator.

	MAP	IR		Time (CPU effort)		
		$t = 80$	$t = 160$	$T = 240$	$T = 2400$	$T = 24000$
Inefficient	0.16338	45.1748	47.3602	0.031250	4.046875	382.05532
Single-move	0.60585	8.74229	110.245	0.001000	0.015625	0.1083750

Note: MAP denotes mean acceptance probability over the whole sample, while IR refers to inefficiency ratio of the MCMC drawings at observations 80 and 160. Time refers to the total CPU time (in seconds) taken to simulate one complete drawing.

**Table 2:** Bayesian inference results.

Dow-Jones	PM	PSD	$\phi_{0.5}$	$\phi_{\min}$	$\phi_{\max}$	IF
$\delta$	0.052	0.041	0.054	-0.048	0.103	3.001
$\varphi$	0.812	0.082	0.854	-0.992	0.999	2.115
$\varphi_u^2$	0.010	0.034	0.013	0.002	0.018	1.509
$\omega$	0.405	0.071	0.461	0.004	0.816	2.367
$\alpha$	0.152	0.040	0.291	0.001	0.873	1.048
$\beta$	0.651	0.168	0.629	0.003	0.984	2.994
$\gamma$	0.392	0.112	0.393	-0.681	0.418	5.108
FTSE	PM	PSD	$\phi_{0.5}$	$\phi_{\min}$	$\phi_{\max}$	IF
$\delta$	0.059	0.036	0.059	-0.051	0.395	3.111
$\varphi$	0.811	0.096	0.839	-0.809	0.999	2.154
$\varphi_u^2$	0.009	0.029	0.012	0.005	0.017	1.995
$\omega$	0.205	0.087	0.398	0.003	0.995	1.457
$\alpha$	0.140	0.055	0.187	0.001	0.931	3.458
$\beta$	0.682	0.153	0.701	0.001	0.988	2.721
$\gamma$	0.374	0.102	0.381	-0.615	0.401	1.254
Nikkei	PM	PSD	$\phi_{0.5}$	$\phi_{\min}$	$\phi_{\max}$	IF
$\delta$	0.068	0.051	0.068	-0.064	0.211	2.998
$\varphi$	0.809	0.090	0.837	-0.831	0.999	1.211
$\varphi_u^2$	0.010	0.031	0.010	0.004	0.019	2.001
$\omega$	0.195	0.079	0.228	0.004	0.501	2.789
$\alpha$	0.149	0.052	0.197	0.001	0.893	3.974
$\beta$	0.634	0.119	0.645	0.006	0.989	1.988
$\gamma$	0.408	0.123	0.409	-0.587	0.487	4.007

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

factors are calculated using a Parzen window equal to  $0.1T$  (where, recall,  $T$  is the number of observations) and indicate that the M-H sampling algorithm has converged and well behaved (This is also justified by the ACFs of the draws. However, they are not presented for space considerations and are available upon request.) With the exception of the constants  $\delta$  and the  $\varphi_u^2$ 's, there is low uncertainty with the estimation of the parameters. The estimated persistence,  $\alpha + \beta$ , for all three markets is close to 0.8 with the highest being the one of

FTSE (0.822), indicating that the half life of a shock is around 3.5. The estimated asymmetry parameters are round 0.4 with “*t*-statistics” higher than 3.2, indicating that the leverage effect is important in all three markets. In a nut shell, all estimated parameters have plausible values, which are in accordance with previous results in the literature.

We have also performed a sensitivity analysis to our choice of priors. In particular, we have halved and doubled the dispersion of the prior distributions around their respective means. Figures 1, 2, and 3 show the kernel density estimates for all parameters for all datasets for the posterior distributions for the three cases: when the variances are 10,000 (baseline posterior), when the variances are halved (small variance posterior), and when the variances are doubled (large variance posterior). We used a canonical Epanechnikov kernel, and the optimal bandwidth was determined automatically by the data. The results which are reported in Figures 1, 2, and 3 indicate that the choice of priors does not unduly influence our conclusions.

Finally, treating the estimated posterior means as the “true” parameters, we can employ the formulae of Section 2.1 and compare the moments implied by the estimates and the sample ones. One fact is immediately obvious. All order autocorrelations of excess returns implied by the estimates are positive but small, with the 1st one being around 0.04, which is in accordance with the (ii) stylized fact (see Section 1). However, for all the three markets, the sample skewness coefficients are negative, ranging from  $-0.89$  (FTSE) to  $-0.12$  (Nikkei), whereas the implied ones are all positive, ranging from 0.036 (FTSE) to 0.042 (Dow-Jones). Nevertheless, the model is matching all the other stylized facts satisfactorily, that is, the estimated parameter values accommodate high coefficient of variation, leptokurtosis as well the volatility clustering and leverage effect.

## 5. Conclusions

In this paper, 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. Furthermore, we derive the theoretical properties of this model, as far as moments and dynamic moments are concerned.

In order to investigate the practical performance of the proposed procedure, we estimate within a Bayesian context our time-varying GQARCH(1,1)-M-AR(1) model for weekly excess stock returns from the Dow-Jones, Nikkei, and FTSE index. With the exception of the returns’ skewness, the suggested likelihood-based estimation method and our model is producing satisfactory results, as far as a comparison between sample and theoretical moments is concerned.

Although we have developed the method within the context of an AR(1) 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’s switching process. A Bernoulli’s distributed price of risk would allow a negative third moment by appropriately choosing the two values of

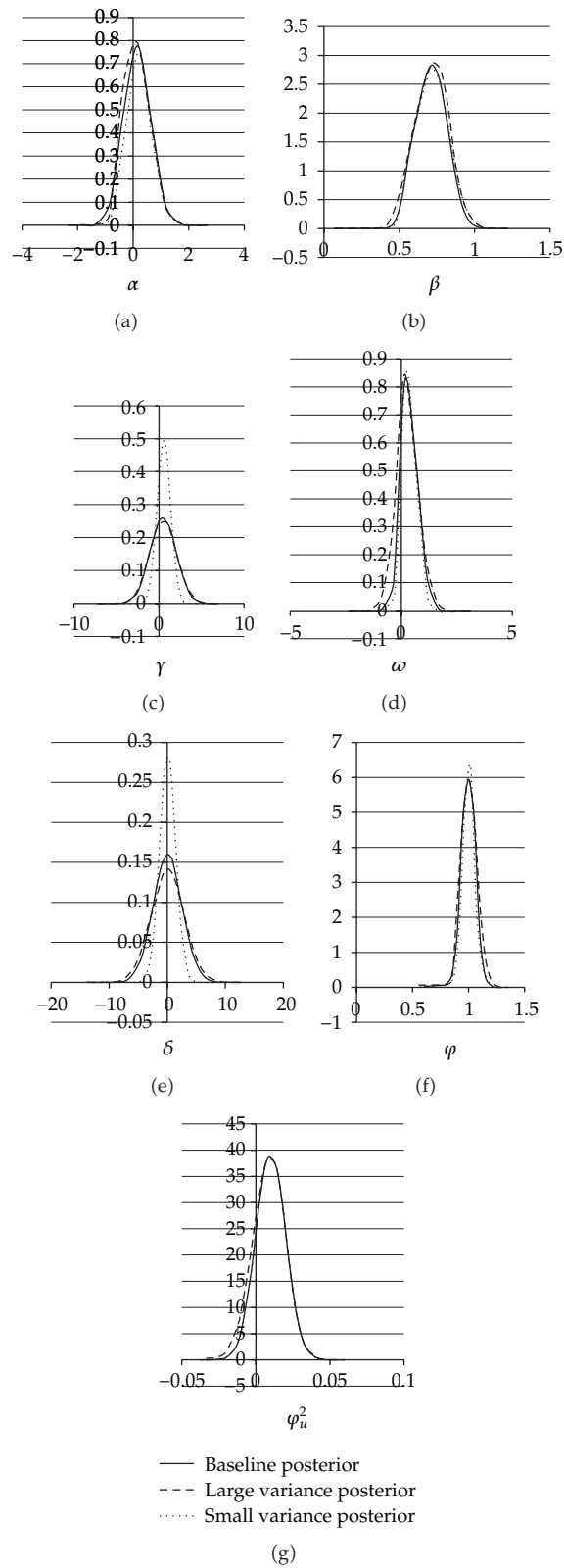


Figure 1: Dow-Jones: posterior density estimates and sensitivity analysis.



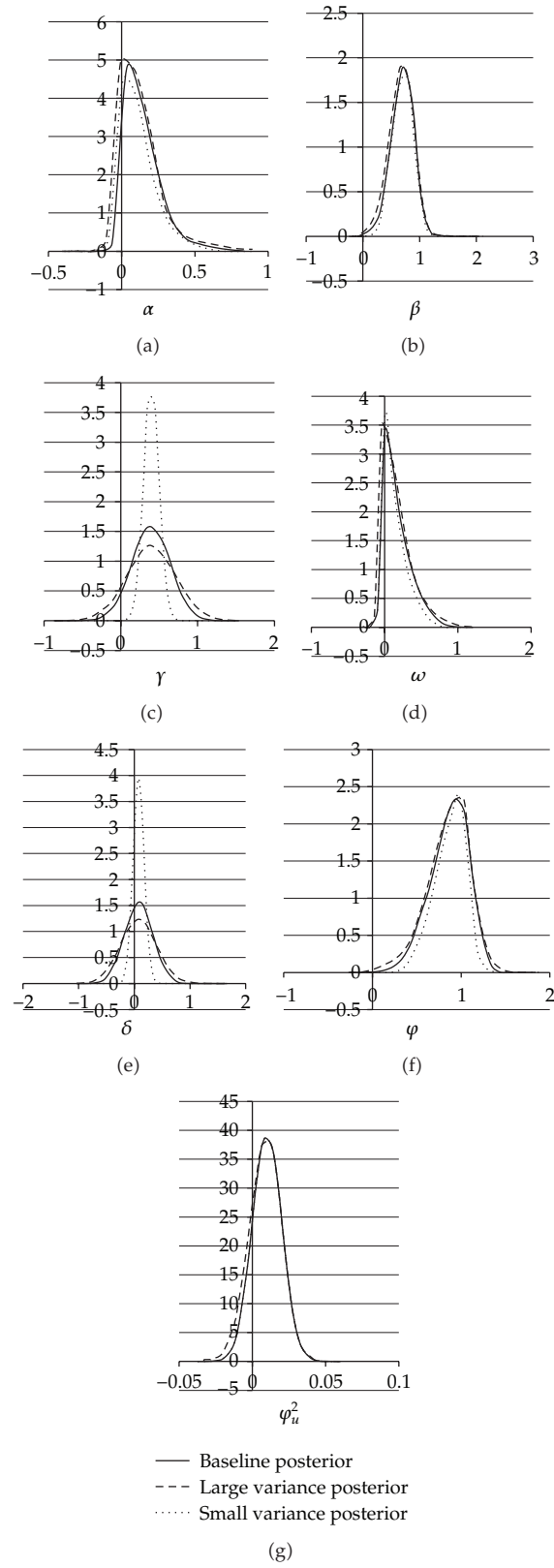
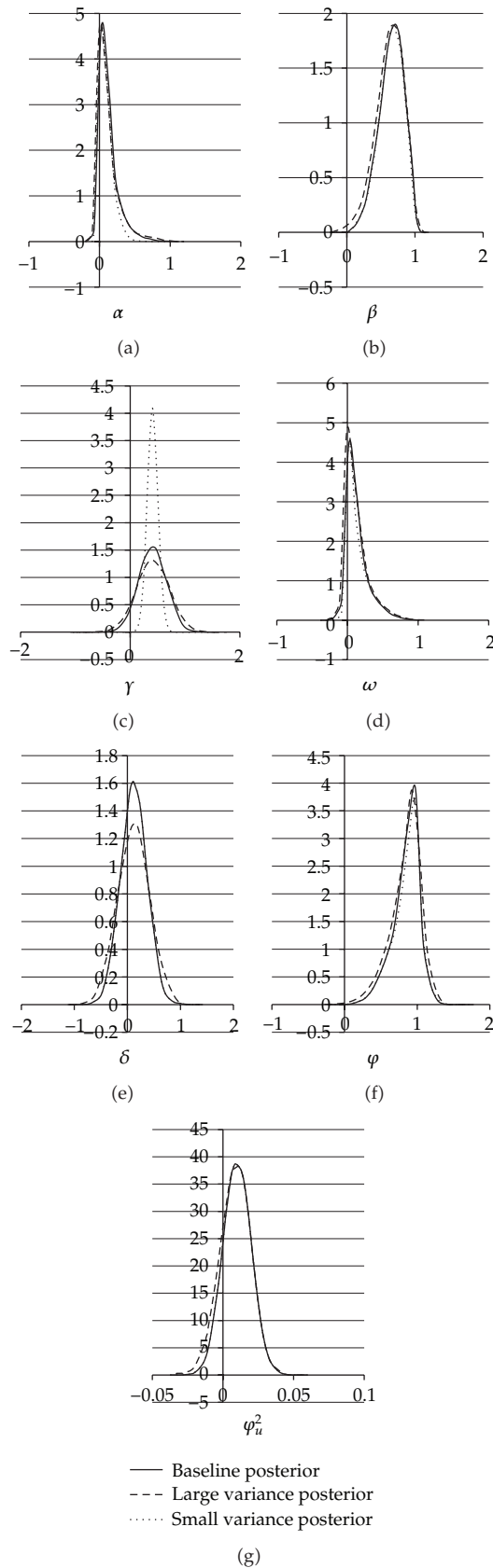


Figure 2: FTSE: posterior density estimates and sensitivity analysis.



**Figure 3:** Nikkei: posterior density estimates and sensitivity analysis.

the in-mean process. However, this would make all computations much more complicated. In an earlier version of the paper, we assumed that the market price of risk follows a normal distribution, and we applied both the classical and the Bayesian procedure to three stock markets (where we decided to set the posterior means as initial values for the simulated EM algorithm). The results suggested that the Bayesian and the classical procedures are quite in agreement (see Anyfantaki and Demos [24]).

Finally, it is known that (e.g., [32, pages 84 and 85]) 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

$$E\left(\frac{\partial \ln p(\boldsymbol{\delta} \mid \mathbf{r}, \phi, \boldsymbol{\mathcal{T}}_0)}{\partial \phi} \mid \mathbf{r}, \phi^{(n)}, \boldsymbol{\mathcal{T}}_0\right) = 0, \quad (5.1)$$

so it is clear that the score can be obtained as the expected value given  $\mathbf{r}, \phi, \boldsymbol{\mathcal{T}}_0$  of the sum of the unobservable scores corresponding to  $\ln p(\mathbf{r} \mid \boldsymbol{\delta}, \phi, \boldsymbol{\mathcal{T}}_0)$  and  $\ln p(\boldsymbol{\delta} \mid \phi, \boldsymbol{\mathcal{T}}_0)$ . This could be very useful for the classical estimation procedure, not presented here, as even though our algorithm is an  $O(T)$  one, it is still rather slow. We leave these issues for further research.

## Appendices

### A. Proof of (3.69) and (3.82)

*Proof of Equation (3.69).* This is easily derived using the fact that

$$r_t = \delta_t h_t + \varepsilon_t, \quad (A.1)$$

where

$$r_t \mid h_t \sim N\left(\delta h_t, \left(\frac{\varphi_u^2}{1 - \varphi^2} h_t + 1\right) h_t\right), \quad (A.2)$$

and, consequently,

$$\begin{pmatrix} \varepsilon_t \\ r_t \end{pmatrix} \mid h_t \sim N\left(\begin{pmatrix} 0 \\ \delta h_t \end{pmatrix}, \begin{pmatrix} h_t & h_t \\ h_t & \left(\frac{\varphi_u^2}{1 - \varphi^2} h_t + 1\right) h_t \end{pmatrix}\right), \quad (A.3)$$

and, thus; from the definition of the bivariate normal,

$$\begin{aligned} E(\varepsilon_t \mid r_t, h_t) &= \frac{(1 - \varphi^2)(r_t - \delta h_t)}{\varphi_u^2 h_t + 1 - \varphi^2}, \\ \text{Var}(\varepsilon_t \mid r_t, h_t) &= \frac{\varphi_u^2 h_t^2}{\varphi_u^2 h_t + 1 - \varphi^2}. \end{aligned} \quad (\text{A.4})$$

Consequently,

$$\varepsilon_t \mid r_t, h_t, \phi \sim N\left(\frac{(1 - \varphi^2)(r_t - \delta h_t)}{\varphi_u^2 h_t + 1 - \varphi^2}, \frac{\varphi_u^2 h_t^2}{\varphi_u^2 h_t + 1 - \varphi^2}\right). \quad (\text{A.5})$$

□

*Proof of Equation (3.82).* We have that

$$p(r_{t+1} \mid h_{t+1}^r) = \frac{1}{\sqrt{2\pi((\varphi_u^2/(1 - \varphi^2))h_{t+1}^r + 1)h_{t+1}^r}} \exp\left(-\frac{(r_{t+1} - \delta h_{t+1}^r)^2}{2((\varphi_u^2/(1 - \varphi^2))h_{t+1}^r + 1)h_{t+1}^r}\right), \quad (\text{A.6})$$

and, thus,

$$\begin{aligned} \frac{p(r_{t+1} \mid h_{t+1}^{\text{new}})}{p(r_{t+1} \mid h_{t+1}^r)} &= \frac{\sqrt{((\varphi_u^2/(1 - \varphi^2))h_{t+1}^r + 1)h_{t+1}^r}}{\sqrt{((\varphi_u^2/(1 - \varphi^2))h_{t+1}^{\text{new}} + 1)h_{t+1}^{\text{new}}}} \\ &\times \exp\left(\frac{(r_{t+1} - \delta h_{t+1}^r)^2}{2((\varphi_u^2/(1 - \varphi^2))h_{t+1}^r + 1)h_{t+1}^r} - \frac{(r_{t+1} - \delta h_{t+1}^{\text{new}})^2}{2((\varphi_u^2/(1 - \varphi^2))h_{t+1}^{\text{new}} + 1)h_{t+1}^{\text{new}}}\right). \end{aligned} \quad (\text{A.7})$$

Also,

$$d_{t+1}^r = \sqrt{\frac{h_{t+2}^r - \omega - \beta h_{t+1}^r}{\alpha}}, \quad (\text{A.8})$$

and so

$$\frac{d_{t+1}^r}{d_{t+1}^{\text{new}}} = \frac{\sqrt{h_{t+2}^r - \omega - \beta h_{t+1}^r}}{\sqrt{h_{t+2}^r - \omega - \beta h_{t+1}^{\text{new}}}}. \quad (\text{A.9})$$

Moreover,

$$\frac{c_{t+1}^{\text{new}}}{c_{t+1}^r} = \frac{h_{t+1}^r \sqrt{\varphi_u^2 h_{t+1}^{\text{new}} + 1 - \varphi^2} \left[ \begin{aligned} & \exp \left( -\frac{(\varphi_u^2 h_{t+1}^{\text{new}} + 1 - \varphi^2)}{2\varphi_u^2 h_{t+1}^{\text{new}2}} \left( \frac{\gamma + \sqrt{\frac{h_{t+2}^r - \omega - \beta h_{t+1}^{\text{new}}}{\alpha}}}{(1 - \varphi^2)(r_{t+1} - \delta h_{t+1}^{\text{new}})} \right)^2 \right) \\ & + \exp \left( -\frac{(\varphi_u^2 h_{t+1}^{\text{new}} + 1 - \varphi^2)}{2\varphi_u^2 h_{t+1}^{\text{new}2}} \left( \frac{\gamma - \sqrt{\frac{h_{t+2}^r - \omega - \beta h_{t+1}^{\text{new}}}{\alpha}}}{(1 - \varphi^2)(r_{t+1} - \delta h_{t+1}^{\text{new}})} \right)^2 \right) \end{aligned} \right]}{h_{t+1}^{\text{new}} \sqrt{\varphi_u^2 h_{t+1}^r + 1 - \varphi^2} \left[ \begin{aligned} & \exp \left( -\frac{(\varphi_u^2 h_{t+1}^r + 1 - \varphi^2)}{2\varphi_u^2 h_{t+1}^{r2}} \left( \frac{\gamma + \sqrt{\frac{h_{t+2}^r - \omega - \beta h_{t+1}^r}{\alpha}}}{(1 - \varphi^2)(r_{t+1} - \delta h_{t+1}^r)} \right)^2 \right) \\ & + \exp \left( -\frac{(\varphi_u^2 h_{t+1}^r + 1 - \varphi^2)}{2\varphi_u^2 h_{t+1}^{r2}} \left( \frac{\gamma - \sqrt{\frac{h_{t+2}^r - \omega - \beta h_{t+1}^r}{\alpha}}}{(1 - \varphi^2)(r_{t+1} - \delta h_{t+1}^r)} \right)^2 \right) \end{aligned} \right]}. \quad (\text{A.10})$$

And the result comes straightforward.  $\square$

## B. Dynamic Moments of the Conditional Variance

We have

$$\begin{aligned} \text{Cov}(h_t, h_{t-k}) &= \alpha \text{Cov}(\varepsilon_{t-1}^2, h_{t-k}) - 2\alpha\gamma \text{Cov}(\varepsilon_{t-1}, h_{t-k}) + \beta \text{Cov}(h_{t-1}, h_{t-k}) \\ &= (\alpha + \beta) \text{Cov}(h_{t-1}, h_{t-k}) = \cdots = (\alpha + \beta)^k V(h_t), \\ E(h_t \varepsilon_{t-k}) &= \alpha E(\varepsilon_{t-1}^2 \varepsilon_{t-k}) - 2\alpha\gamma E(\varepsilon_{t-1} \varepsilon_{t-k}) + \beta E(h_{t-1} \varepsilon_{t-k}) \\ &= (\alpha + \beta) E(h_{t-1} \varepsilon_{t-k}) = \cdots = (\alpha + \beta)^{k-1} E(h_{t-k+1} \varepsilon_{t-k}) \\ &= -2\alpha\gamma (\alpha + \beta)^{k-1} E(h_{t-1}), \end{aligned} \quad (\text{B.1})$$

since

$$\begin{aligned} E(h_t \varepsilon_{t-1}) &= E(\alpha \varepsilon_{t-1}^3 - 2\alpha\gamma \varepsilon_{t-1}^2 + \beta h_{t-1} \varepsilon_{t-1}) \\ &= -2\alpha\gamma E(\varepsilon_{t-1}^2) = -2\alpha\gamma E(h_{t-1}). \end{aligned} \quad (\text{B.2})$$

Also,

$$\begin{aligned}
\text{Cov}(h_t^2, h_{t-k}) &= A\text{Cov}(h_{t-1}^2, h_{t-k}) + B\text{Cov}(h_{t-1}, h_{t-k}) \\
&= A^2\text{Cov}(h_{t-2}^2, h_{t-k}) + ABC\text{Cov}(h_{t-2}, h_{t-k}) + B\text{Cov}(h_{t-1}, h_{t-k}) \\
&= \dots = A^{k-1}\text{Cov}(h_{t-k+1}^2, h_{t-k}) + A^{k-2}B\text{Cov}(h_{t-k+1}, h_{t-k}) \\
&\quad + \dots + ABC\text{Cov}(h_{t-2}, h_{t-k}) + B\text{Cov}(h_{t-1}, h_{t-k}) \\
&= A^{k-1}\text{Cov}(h_{t-k+1}^2, h_{t-k}) + (\alpha + \beta)B \frac{(\alpha + \beta)^{k-1} - A^{k-1}}{(\alpha + \beta) - A} V(h_t) \\
&= A^k \left[ E(h_{t-1}^3) - E(h_{t-1}^2)E(h_{t-1}) \right] + \frac{(\alpha + \beta)^k - A^k}{(\alpha + \beta) - A} BV(h_t),
\end{aligned} \tag{B.3}$$

as

$$\text{Cov}(h_t^2, h_{t-1}) = A \left[ E(h_{t-1}^3) - E(h_{t-1}^2)E(h_{t-1}) \right] + BV(h_{t-1}), \tag{B.4}$$

where  $A = 3\alpha^2 + \beta^2 + 2\alpha\beta$  and  $B = 2[2\alpha^2\gamma^2 + (\omega + \alpha\gamma^2)(\alpha + \beta)]$ .

Furthermore,

$$\begin{aligned}
E(h_t^2 \varepsilon_{t-k}) &= AE(h_{t-1}^2 \varepsilon_{t-k}) + BE(h_{t-1} \varepsilon_{t-k}) \\
&= A^2 E(h_{t-2}^2 \varepsilon_{t-k}) + ABE(h_{t-2} \varepsilon_{t-k}) + BE(h_{t-1} \varepsilon_{t-k}) \\
&= \dots = A^{k-1} E(h_{t-k+1}^2 \varepsilon_{t-k}) - 2\alpha\gamma B \left[ A^{k-2} + \dots + A(\alpha + \beta)^{k-3} + (\alpha + \beta)^{k-2} \right] E(h_{t-1}) \\
&= -4\alpha\gamma(3\alpha + \beta)A^{k-1} E(h_{t-1}^2) \\
&\quad - 4\alpha\gamma \left[ (\omega + \alpha\gamma^2) \frac{(\alpha + \beta)^k - A^k}{(\alpha + \beta) - A} + 2\alpha^2\gamma^2 \frac{(\alpha + \beta)^{k-1} - A^{k-1}}{(\alpha + \beta) - A} \right] E(h_{t-1}),
\end{aligned} \tag{B.5}$$

as

$$E(h_t^2 \varepsilon_{t-1}) = -4\alpha\gamma(3\alpha + \beta)E(h_{t-1}^2) - 4(\omega + \alpha\gamma^2)\alpha\gamma E(h_{t-1}). \tag{B.6}$$

Also we have that

$$\begin{aligned}
\text{Cov}(h_t, h_{t-k}^2) &= (\alpha + \beta)\text{Cov}(h_{t-1}, h_{t-k}^2) \\
&= (\alpha + \beta)^k \left[ E(h_{t-1}^3) - E(h_{t-1}^2)E(h_{t-1}) \right],
\end{aligned} \tag{B.7}$$

as

$$\text{Cov}(h_t, h_{t-1}^2) = (\alpha + \beta) \left[ E(h_{t-1}^3) - E(h_{t-1}^2)E(h_{t-1}) \right]. \quad (\text{B.8})$$

Now,

$$\begin{aligned} \text{Cov}(h_t^2, h_{t-k}^2) &= A\text{Cov}(h_{t-1}^2, h_{t-k}^2) + B\text{Cov}(h_{t-1}, h_{t-k}^2) \\ &= A^2\text{Cov}(h_{t-2}^2, h_{t-k}^2) + AB\text{Cov}(h_{t-2}, h_{t-k}^2) + B\text{Cov}(h_{t-1}, h_{t-k}^2) \\ &= \dots = A^{k-1}\text{Cov}(h_{t-k+1}^2, h_{t-k}^2) + A^{k-1}B\text{Cov}(h_{t-k+1}, h_{t-k}^2) \\ &\quad + \dots + AB\text{Cov}(h_{t-2}, h_{t-k}^2) + B\text{Cov}(h_{t-1}, h_{t-k}^2) \\ &= A^k V(h_{t-1}^2) + B \frac{(\alpha + \beta)^k - A^k}{(\alpha + \beta) - A} \left[ E(h_{t-1}^3) - E(h_{t-1}^2)E(h_{t-1}) \right], \end{aligned} \quad (\text{B.9})$$

as

$$\text{Cov}(h_t^2, h_{t-1}^2) = AV(h_{t-1}^2) + B \left[ E(h_{t-1}^3) - E(h_{t-1}^2)E(h_{t-1}) \right]. \quad (\text{B.10})$$

Further,

$$\begin{aligned} \text{Cov}(h_t, \varepsilon_{t-k}^2) &= (\alpha + \beta)\text{Cov}(h_{t-1}, \varepsilon_{t-k}^2) \\ &= (\alpha + \beta)^{k-1}\text{Cov}(h_{t-k+1}, \varepsilon_{t-k}^2) \\ &= (\alpha + \beta)^k V(h_{t-1}) + 2\alpha(\alpha + \beta)^{k-1} E(h_t^2), \end{aligned} \quad (\text{B.11})$$

as

$$\begin{aligned} \text{Cov}(h_t, \varepsilon_{t-1}^2) &= \alpha \left[ 2E(h_t^2) + V(h_{t-1}) \right] + \beta V(h_{t-1}) \\ &= (\alpha + \beta) V(h_{t-1}) + 2\alpha E(h_t^2). \end{aligned} \quad (\text{B.12})$$

Further

$$\begin{aligned} \text{Cov}(h_t^2, \varepsilon_{t-k}^2) &= A\text{Cov}(h_{t-1}^2, \varepsilon_{t-k}^2) + B\text{Cov}(h_{t-1}, \varepsilon_{t-k}^2) \\ &= A\text{Cov}(h_{t-2}^2, \varepsilon_{t-k}^2) + AB\text{Cov}(h_{t-2}, \varepsilon_{t-k}^2) + B\text{Cov}(h_{t-1}, \varepsilon_{t-k}^2) \\ &= \dots = A^{k-1}\text{Cov}(h_{t-k+1}^2, \varepsilon_{t-k}^2) + A^{k-2}B\text{Cov}(h_{t-k+1}, \varepsilon_{t-k}^2) \\ &\quad + \dots + AB\text{Cov}(h_{t-2}, \varepsilon_{t-k}^2) + B\text{Cov}(h_{t-1}, \varepsilon_{t-k}^2) \end{aligned}$$



$$\begin{aligned}
&= A^{k-1} \text{Cov}\left(h_{t-k+1}^2, \varepsilon_{t-k}^2\right) + B(\alpha + \beta) \frac{(\alpha + \beta)^{k-1} - A^{k-1}}{(\alpha + \beta) - A} V(h_{t-1}) \\
&\quad + 2\alpha B \frac{(\alpha + \beta)^{k-1} - A^{k-1}}{(\alpha + \beta) - A} E\left(h_t^2\right) \\
&= A^k \left[ E\left(h_{t-1}^3\right) - E\left(h_{t-1}^2\right) E(h_{t-1}) \right] + B \frac{(\alpha + \beta)^k - A^k}{(\alpha + \beta) - A} V(h_{t-1}) \\
&\quad + 4\alpha(3\alpha + \beta) A^{k-1} E\left(h_{t-1}^3\right) \\
&\quad + \left[ 2\alpha B \frac{(\alpha + \beta)^{k-1} - A^{k-1}}{(\alpha + \beta) - A} + 4\left(2\alpha^2 \gamma^2 + (\omega + \alpha \gamma^2)\right) A^{k-1} \right] E\left(h_t^2\right),
\end{aligned} \tag{B.13}$$

as

$$\begin{aligned}
\text{Cov}\left(h_t^2, \varepsilon_{t-1}^2\right) &= A \left[ E\left(h_{t-1}^3\right) - E\left(h_{t-1}^2\right) E(h_{t-1}) \right] + 4\alpha(3\alpha + \beta) E\left(h_{t-1}^3\right) \\
&\quad + 4\left(2\alpha^2 \gamma^2 + (\omega + \alpha \gamma^2)\right) E\left(h_t^2\right) + BV(h_t).
\end{aligned} \tag{B.14}$$

Further,

$$\begin{aligned}
E(h_t h_{t-k} \varepsilon_{t-k}) &= (\alpha + \beta) E(h_{t-1} h_{t-k} \varepsilon_{t-k}) = \cdots = (\alpha + \beta)^{k-1} E(h_{t-k+1} h_{t-k} \varepsilon_{t-k}) \\
&= -2\alpha \gamma (\alpha + \beta)^{k-1} E\left(h_{t-1}^2\right),
\end{aligned} \tag{B.15}$$

as

$$E(h_t h_{t-1} \varepsilon_{t-1}) = -2\alpha \gamma E\left(h_{t-1}^2\right). \tag{B.16}$$

Finally,

$$\begin{aligned}
E\left(h_t^2 h_{t-k} \varepsilon_{t-k}\right) &= AE\left(h_{t-1}^2 h_{t-k} \varepsilon_{t-k}\right) + BE(h_{t-1} h_{t-k} \varepsilon_{t-k}) \\
&= \cdots = A^{k-1} E\left(h_{t-k+1}^2 h_{t-k} \varepsilon_{t-k}\right) + A^{k-2} BE(h_{t-k+1} h_{t-k} \varepsilon_{t-k}) \\
&\quad + \cdots + ABE(h_{t-2} h_{t-k} \varepsilon_{t-k}) + BE(h_{t-1} h_{t-k} \varepsilon_{t-k}) \\
&= -4\alpha \gamma A^{k-1} \left[ (3\alpha + \beta) E\left(h_{t-1}^3\right) + (\omega + \alpha \gamma^2) E\left(h_{t-1}^2\right) \right] \\
&\quad - 2\alpha \gamma B \frac{(\alpha + \beta)^{k-1} - A^{k-1}}{(\alpha + \beta) - A} E\left(h_{t-1}^2\right),
\end{aligned} \tag{B.17}$$

as

$$E\left(h_t^2 h_{t-1} \varepsilon_{t-1}\right) = -4\alpha\gamma(3\alpha + \beta)E\left(h_{t-1}^3\right) - 4\left(\omega + \alpha\gamma^2\right)\alpha\gamma E\left(h_{t-1}^2\right). \quad (\text{B.18})$$

## Acknowledgments

The authors would like to thank S. Arvanitis, G. Fiorentini, and E. Sentana for their valuable comments. We wish also to thank the participants of the 7th Conference on Research on Economic Theory and Econometrics (Naxos, July 2008), the 14th Spring Meeting of Young Economists (Istanbul, April 2009), the Doctoral Meeting of Montpellier (D.M.M.) in Economics, Management and Finance (Montpellier, May 2009), the 2nd Ph.D. Conference in Economics 2009, in Memory of Vassilis Patsatzis (Athens, May 2009), the 3rd International Conference in Computational and Financial Econometrics (Limassol, October 2009), the 4th International Conference in Computational and Financial Econometrics (London, December 2010) and the seminar participants at the Athens University of Economics and Business for useful discussions. They acknowledge Landis Conrad's help for the collection of the data. Finally, they greatly benefited by the comments of the Associate Editor and an anonymous referee.

## References

- [1] R. C. Merton, "Stationarity and persistence in the GARCH (1,1) model," *Econometric Theory*, vol. 6, no. 3, pp. 318–334, 1980.
- [2] L. R. Glosten, R. Jaganathan, and D. E. Runkle, "On the relation between the expected value and the volatility of the nominal excess return on stocks," *Journal of Finance*, vol. 48, no. 5, pp. 1101–1801, 1993.
- [3] R. F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, vol. 50, no. 4, pp. 987–1007, 1982.
- [4] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, vol. 31, no. 3, pp. 307–327, 1986.
- [5] T. Bollerslev, R. Y. Chou, and K. F. Kroner, "ARCH modeling in finance: a review of the theory and empirical evidence," *Journal of Econometrics*, vol. 52, no. 1-2, pp. 5–59, 1992.
- [6] S. Arvanitis and A. Demos, "Time dependence and moments of a family of time-varying parameter GARCH in mean models," *Journal of Time Series Analysis*, vol. 25, no. 1, pp. 1–25, 2004.
- [7] O. E. Barndorff-Nielsen and N. Shephard, *Modelling by Levy Processes for Financial Econometrics*, Nuffield College Economics Working Papers, Oxford University, London, UK, 2000, D.P. No 2000-W3.
- [8] R. N. Mantegna and H. E. Stanley, "Turbulence and financial markets," *Nature*, vol. 383, no. 6601, pp. 587–588, 1996.
- [9] R. N. Mantegna and H. E. Stanley, *An Introduction to Econophysics Correlations and Complexity in Finance*, Cambridge University Press, Cambridge, UK, 2000.
- [10] G. Fiorentini, E. Sentana, and N. Shephard, "Likelihood-based estimation of latent generalized ARCH structures," *Econometrica*, vol. 72, no. 5, pp. 1481–1517, 2004.
- [11] E. Sentana, "Quadratic ARCH models," *The Review of Economic Studies*, vol. 62, no. 4, pp. 639–661, 1995.
- [12] G. Fiorentini and E. Sentana, "Conditional means of time series processes and time series processes for conditional means," *International Economic Review*, vol. 39, no. 4, pp. 1101–1118, 1998.
- [13] S. Arvanitis, "The diffusion limit of a TVP-GQARCH-M(1,1) model," *Econometric Theory*, vol. 20, no. 1, pp. 161–175, 2004.
- [14] S. Chib, "Markov chain monte carlo methods: computation and Inference," in *Handbook of Econometrics*, J. J. Heckman and E. Leamer, Eds., vol. 5, pp. 3569–3649, North-Holland, Amsterdam, The Netherlands, 2001.
- [15] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*, Springer, New York, NY, USA, 1993.

- [16] A. E. Gelfand and A. F. M. Smith, "Sampling-based approaches to calculating marginal densities," *Journal of the American Statistical Association*, vol. 85, no. 410, pp. 398–409, 1990.
- [17] L. Tierney, "Markov chains for exploring posterior distributions," *The Annals of Statistics*, vol. 22, no. 4, pp. 1701–1728, 1994.
- [18] M. A. Tanner and W. H. Wong, "The calculation of posterior distributions by data augmentation," *Journal of the American Statistical Association*, vol. 82, no. 398, pp. 528–540, 1987.
- [19] S. Kim, N. Shephard, and S. Chib, "Stochastic volatility: likelihood inference and comparison with ARCH models," *Review of Economic Studies*, vol. 65, no. 3, pp. 361–393, 1998.
- [20] N. Shephard, "Statistical aspects of ARCH and stochastic volatility," in *Time Series Models in Econometrics, Finance and Other Fields*, D. R. Cox, D. V. Hinkley, and O. E. B. Nielsen, Eds., pp. 1–67, Chapman & Hall, London, UK, 1996.
- [21] S. G. Giakoumatos, P. Dellaportas, and D. M. Politis, *Bayesian Analysis of the Unobserved ARCH Model*, vol. 15, Department of Statistics, Athens University of Economics and Business, Athens, Greece, 1999.
- [22] A. P. Dempster, N. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society B*, vol. 39, no. 1, pp. 1–38, 1977.
- [23] E. Sentana and G. Fiorentini, "Identification, estimation and testing of conditionally heteroskedastic factor models," *Journal of Econometrics*, vol. 102, no. 2, pp. 143–164, 2001.
- [24] S. Anyfantaki and A. Demos, "Estimation of a time-varying GQARCH(1,1)-M model," Working paper, 2009, <http://www.addegem-asso.fr/docs/PapersDMM2009/2.pdf>.
- [25] D. J. Poirier, "Revising beliefs in nonidentified models," *Econometric Theory*, vol. 14, no. 4, pp. 483–509, 1998.
- [26] P. Clifford, "Markov random fields in statistics," in *Disorder in Physical Systems: A Volume in Honor of John M. Hammersley*, G. Grimmett and D. Welsh, Eds., pp. 19–32, Oxford University Press, 1990.
- [27] S. Chib, "Bayes regression with autoregressive errors," *Journal of Econometrics*, vol. 58, no. 3, pp. 275–294, 1993.
- [28] L. Bauwens and M. Lubrano, "Bayesian inference on Garch models using the Gibbs sampler," *Econometrics Journal*, vol. 1, no. 1, pp. C23–C46, 1998.
- [29] T. Nakatsuma, "Bayesian analysis of ARMA-GARCH models: a Markov chain sampling approach," *Journal of Econometrics*, vol. 95, no. 1, pp. 57–69, 2000.
- [30] D. Ardia, "Bayesian estimation of a Markov-switching threshold asymmetric GARCH model with student-t innovations," *The Econometrics Journal*, vol. 12, no. 1, pp. 105–126, 2009.
- [31] S. X. Wei, "A censored-GARCH model of asset returns with price limits," *Journal of Empirical Finance*, vol. 9, no. 2, pp. 197–223, 2002.
- [32] M. A. Tanner, *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Function*, Springer, New York, NY, USA, 3rd edition, 1996.

