Estimation and Properties of a Time-Varying EGARCH(1,1) in Mean Model

Sofia Anyfantaki[1] and Antonis Demos
Athens University of Economics and Business

Version: July 2012

Time-varying GARCH-M models are commonly employed 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-parameter EGARCH(1,1)-M are derived. We discuss them and apply the suggested Bayesian estimation to three major stock markets.

Key Words: dynamic heteroskedasticity, in mean models, time varying parameter, Markov chain Monte Carlo, simulated EM algorithm, Bayesian inference
Subject Classification: [JEL classification] C13;C15;C63

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 first conditional moment. For instance, in the stock market context, the first conditional moment of stock market excess returns, given some information set, is a possibly time-varying, function of volatility (see e.g. Merton [27]; Glosten, Jagannathan and Runkle [19]). These have led to modifications and extensions of the initial ARCH model of Engle [11] and its generalization by Bollerslev [5], 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. [6] for a survey).

[1]Address of correspondence: Athens University of Economics and Business, Department of International and European Economic Studies, 76 Patission Street, 10434, Athens, Greece. Tel.: +302108203451. Fax: +302108214122 // *E-mail address:* sanyfantaki@aueb.gr (S.Anyfantaki), demos@aueb.gr (A.Demos)

Although the vast majority of the research in conditional heteroscedasticity is being processed aiming the stylized facts of financial stock returns and of economic time series in general, Arvanitis and Demos [3], have shown that a family of time varying GARCH-M models can in fact be consistent with the sample characteristics of time series describing the temporal evolution of velocity changes of turbulent fluid and gas molecules. Despite the fact that the latter statistical characteristics match in a considerable degree their financial analogues (for example leptokurtosis, volatility clustering and quasi long range dependence in the squares are common), there are also significant differences in the behavior of the above 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 Barndorf -Nielsen and Shephard [4]; as well as Mantegna and Stanley [25] and [26]). It was shown that the above mentioned family of models can even create anticorrelation as far as an AR(1) time varying parameter process 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, Sentana and Shephard [13] 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 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. [13]. Thus, we show that their method applies to more complicated models, as well.

We prefer to employ an EGARCH specification (Nelson [30]) for the conditional variance. One common stylized fact in financial economics is the leverage effect, i.e. the fact that negative shocks often increase volatility to a greater extent than positive shocks. Furthermore, the EGARCH model does need positivity constraints, i.e. the conditional variance is positive with probability 1 for all values of the parameter space. This is not the case with most GARCH type models. Finally, the choice of the first-order model here is motivated by the fact that it is the most widely applied exponential GARCH model.

Moreover, many theories in finance involve an explicit trade-off between risk and expected returns. For that matter, we employ 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. [19] when they explicitly include the nominal risk free rate in the conditioning information set they obtain a negative ARCH-M parameter. For the above, we allow the conditional variance to affect the mean with a possibly time varying coefficient which we assume for simplicity that follows an AR(1) process. Thus, our model is a Time-Varying Parameter AR(1) EGARCH-M model (TVPAR(1)-EGARCH(1,1)-M).

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, i.e. high coefficient of variation, ii) the autocorrelations of excess returns are insignificant with a possible exception of the 1^{st} one, iii) the distribution of returns is non-normal mainly due to excess kurtosis and maybe asymmetry (negative), iv) there is strong volatility clustering, i.e. significant positive autocorrelation of squared returns even for high lags, and v) the so called leverage effect, i.e. 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 of our model. Next, we review Bayesian and classical likelihood approaches to inference for the time-varying EGARCH-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. [13] is needed to achieve a first order Markov transformation of the model and thus, reducing the computations from $O\left(T^2\right)$ to $O\left(T\right)$. An illustrative empirical application on weekly returns from three major stock markets is presented in Section 4 and we conclude in Section 5.

2. THE MODEL

The definition of our model is:

Definition 1. The TVPAR(1)-EGARCH(1,1)-M model is:

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

$$\delta_t = (1-\varphi)\,\delta + \varphi\delta_{t-1} + \varphi_u u_t \tag{2}$$

$$\ln h_t = \alpha + \beta \ln h_{t-1} + \gamma z_{t-1} + \theta\,|z_{t-1}| \tag{3}$$

$$z_t \backsim i.i.d.N(0,1), \quad u_t \backsim i.i.d.N(0,1) \quad and\ u_t, z_t\ independent\ for\ all\ t's \tag{4}$$

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 EGARCH(1,1). It is obvious that δ_t is the market price of risk (see e.g. Merton [27]; Glosten at al. [19]). 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 quite an 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 model to implement the method of Fiorentini et al. [13]. Let us start with the theoretical properties.

3. THEORETICAL PROPERTIES

The theoretical properties of a model nesting the one under consideration have been already studied in Demos [9]. Here we will just present his results modified for our case, i.e. here we have that his stochastic volatility, ϕ_η, is 0 and his $\theta_i = \beta^i$.

Provided that $|\beta| < 1$ we can invert the conditional variance equation as:

$$\ln(h_t) = \alpha_0 + \sum_{i=0}^{\infty} \beta^i f\left(z_{t-1-i}\right), \ \ \textit{where } \alpha_0 = \frac{\alpha}{1-\beta} \textit{ and } f\left(z_t\right) = \gamma z_t + \theta \left|z_t\right|.$$

Assuming normality and $|\beta| < 1$ we get the second order and strict stationarity of $\{\varepsilon_t\}$ and we can state the following proposition (for a proof see Demos [9], and He, Terasvirta and Malmsten [22]).

Proposition 1. *For $z_t \backsim i.i.d.N(0,1)$ and $|\beta| < 1$ we have that:*

$$\begin{aligned} Cov[h_t, h_{t-k}] &= \exp\left([2\alpha_0\right) \\ &\quad * \left(\varpi_k^{**} \prod_{i=0}^{k-1} \exp\left(\frac{\beta^{2i}\theta^{*2}}{2}\right) \Phi(\beta^i\theta^*) + \exp\left(\frac{\beta^{2i}\gamma^{*2}}{2}\right) \Phi(\beta^i\gamma^*) - \varpi^2\right), \end{aligned}$$

where $\theta^* = \theta + \gamma$, $\gamma^* = \theta - \gamma$,

$$\varpi = \prod_{i=0}^{\infty} \left(\Phi(\beta^i\theta^*) \exp\left(\frac{\beta^{2i}\theta^{*2}}{2}\right) + \exp\left(\frac{\beta^{2i}\gamma^{*2}}{2}\right) \Phi(\beta^i\gamma^*)\right)$$

and

$$\varpi_k^{**} = \prod_{i=0}^{\infty} \left[\begin{array}{c} \exp\left(\frac{[(1+\beta^k)\beta^i\theta^*]^2}{2}\right) \Phi((1+\beta^k)\beta^i\theta^*) \\ +\exp\left(\frac{[(1+\beta^k)\beta^i\gamma^*]^2}{2}\right) \Phi((1+\beta^k)\beta^i\gamma^*) \end{array} \right]$$

Also,

$$\begin{aligned} E(h_t\varepsilon_{t-k}) &= \beta^{k-1} \exp\left(\frac{3}{2}\alpha_0\right) \varpi_k^* \\ &\quad * \prod_{i=0}^{k-2} \left[\exp\left(\frac{\beta^{2i}\theta^{*2}}{2}\right) \Phi(\beta^i\theta^*) + \exp\left(\frac{\beta^{2i}\gamma^{*2}}{2}\right) \Phi(\beta^i\gamma^*)\right] \\ &\quad * \left[\theta^*\Phi(\beta^{k-1}\theta^*) \exp\left(\frac{\beta^{2k-2}\theta^{*2}}{2}\right) - \gamma^*\Phi(\beta^{k-1}\gamma^*) \exp\left(\frac{\beta^{2k-2}\gamma^{*2}}{2}\right)\right] \end{aligned}$$

where

$$\varpi_k^* = \prod_{i=0}^{\infty} \left[\begin{array}{c} \exp\left(\frac{[(\frac{1}{2}+\beta^k)\beta^i\theta^*]^2}{2}\right)\Phi((\frac{1}{2}+\beta^k)\beta^i\theta^*) \\ +\exp\left(\frac{[(\frac{1}{2}+\beta^k)\beta^i\gamma^*]^2}{2}\right)\Phi((\frac{1}{2}+\beta^k)\beta^i\gamma^*) \end{array} \right]$$

Finally,

$$\begin{aligned} Cov[\varepsilon_t^2, \varepsilon_{t-k}^2] &= Cov[h_t, h_{t-k}] + \beta^{k-1}E\left(h_t h_{t-k}\right) \left[\begin{array}{c} \exp\left(\frac{\beta^{2k-2}(\theta^*)^2}{2}\right)\Phi(\beta^{k-1}\theta^*) \\ +\exp\left(\frac{\beta^{2k-2}(\gamma^*)^2}{2}\right)\Phi(\beta^{k-1}\gamma^*) \end{array} \right]^{-1} \\ &* \left[\begin{array}{c} \beta^{k-1}\theta^{*2}\Phi(\beta^{k-1}\theta^*)\exp\left(\frac{\beta^{2k-2}\theta^{*2}}{2}\right) \\ +\beta^{k-1}\gamma^{*2}\Phi(\beta^{k-1}\gamma^*)\exp\left(\frac{\beta^{2k-2}\gamma^{*2}}{2}\right) \end{array} \right] \end{aligned}$$

Furthermore, we have the following result (for a proof see Demos [9]).

THEOREM 1. *Under the assumptions in equation 4 and for* $|\beta| < 1$ *and* $|\varphi| < 1$*, we have that* $\gamma_k = Cov(r_t, r_{t-k})$ *is a quadratic function in* δ*:*

$$\gamma_k = \delta^2 Cov(h_t, h_{t-k}) + \delta E\left(h_t \varepsilon_{t-k}\right) + \varphi^k \frac{\varphi_u^2}{1-\varphi^2} E\left(h_t h_{t-k}\right)$$

It is obvious that provided that there is no time variation in δ, if $E\left(h_t \varepsilon_{t-k}\right)$ is zero, as in the GARCH-M model of Engle et al. [12], then the k-order autocovariance of the series has the sign of the autocovariance of the conditional variance, irrespective of the value of δ. On the other hand, for the first order EGARCH model, it is clear that the sign of $E\left(h_t \varepsilon_{t-k}\right)$ depends on the relative values of β, θ^*, and γ^*. However, notice that under the assumptions of volatility clustering, leverage and asymmetry effects, i.e. $\beta > 0$, $\theta > 0$ and $\gamma < 0$, we have that $\gamma^* > \theta^*$ and $\gamma^* > 0$. Hence $\theta^* \exp\left(\frac{(\beta^{k-1}\theta^*)^2}{2}\right)\Phi(\beta^{k-1}\theta^*) - \gamma^* \exp\left(\frac{(\beta^{k-1}\gamma^*)^2}{2}\right)\Phi(\beta^{k-1}\gamma^*) < 0$ as the exponential and the cumulative distribution functions are non-decreasing. Consequently, $E\left(h_t \varepsilon_{t-k}\right)$ is negative for any k.

Ideally, one would like to employ a model which can be compatible with either negative or positive mean autocorrelations, and potentially different from the sign of the autocorrelation of the conditional variance. As an example, consider the volatility clustering observed in financial data. This implies that at least the first order autocorrelation of the returns' conditional variance is positive. However, there are theoretical arguments which support a positive autocorrelation of short horizon stock returns, whereas long horizon ones are negatively autocorrelated (see Poterba and Summers, 1988 [32]).

Let us now turn to higher moments since for the dynamic asymmetry we also need the covariance of squares-levels and levels-squares. Appropriate modifications of Theorem 2 in Demos [9] give the following results.

THEOREM 2. *Under the assumptions of Theorem 1, the covariance of squares-levels and levels-squares is given by:*

$$\begin{aligned} Cov\left(r_t^2, r_{t-k}\right) &= 2\varphi^k \delta \frac{\varphi_u^2}{1-\varphi^2} E\left(h_t^2 h_{t-k}\right) + \delta Cov\left(h_t, h_{t-k}\right) + E\left(h_t h_{t-k}^{1/2}\right) D_{k-1}^{(1)} \left(\Upsilon_{k-1}^{(1)}\right)^{-1} \\ &+ \left(\frac{\varphi_u^2}{1-\varphi^2} + \delta^2\right) \left[\delta Cov\left(h_t^2, h_{t-k}\right) + E\left(h_t^2 h_{t-k}^{1/2}\right) D_{k-1}^{(2)} \left(\Upsilon_{k-1}^{(2)}\right)^{-1}\right] \end{aligned}$$

and

$$\begin{aligned} Cov(r_t, r_{t-k}^2) &= \left(\frac{\varphi_u^2}{1-\varphi^2} + \delta^2\right) \delta Cov\left(h_t, h_{t-k}^2\right) \\ &+ \delta \left[E(h_t h_{t-k}) F_{k-1}^{(1)} \left(\Upsilon_{k-1}^{(1)}\right)^{-1} + Cov\left(h_t, h_{t-k}\right)\right] \\ &+ 2\left(\varphi^k \frac{\varphi_u^2}{1-\varphi^2} + \delta^2\right) E\left(h_t h_{t-k}^{3/2}\right) D_{k-1}^{(1)} \left(\Upsilon_{k-1}^{(1)}\right)^{-1}, \end{aligned}$$

respectively, where $D_{k-1}^{(s)}$ and $\Upsilon_{k-1}^{(s)}$ are provided in the following Proposition.

Setting $\phi_\eta = 0$ and $\theta_i = \beta^i$ in Proposition 1 and 2 of Demos [9] we get.

PROPOSITION 2.

$$E\left(h_t^s h_{t-k}^d\right) = \exp\left[(s+d)\alpha_0\right] \varpi_k^{(s,d)} \Lambda_{(k-1)}^{(s)}, \; E\left(h_t^s h_{t-k}^d z_{t-k}\right) = E\left(h_t^s h_{t-k}^d\right) D_{k-1}^{(s)} \left(\Upsilon_{k-1}^{(s)}\right)^{-1},$$

$$E\left(h_t^s h_{t-k}^d z_{t-k}^2\right) = E\left(h_t^s h_{t-k}^d\right) \left[1 + F_{k-1}^{(s)} \left(\Upsilon_{k-1}^{(s)}\right)^{-1}\right]$$

where

$$\Upsilon_{k-1}^{(s)} = \Phi\left(A_{0,k-1}^{(0,s)}\right) \exp\left(\Gamma_{0,k-1}^{(0,s)}\right) + \exp\left(\Delta_{0,k-1}^{(0,s)}\right) \Phi\left(-B_{0,k-1}^{(0,s)}\right),$$

$$D_{k-1}^{(s)} = A_{0,k-1}^{(0,s)} \exp\left(\Gamma_{0,k-1}^{(0,s)}\right) \Phi\left(A_{0,k-1}^{(0,s)}\right) + B_{0,k-1}^{(0,s)} \exp\left(\Delta_{0,k-1}^{(0,s)}\right) \Phi\left(-B_{0,k-1}^{(0,s)}\right),$$

$$F_{k-1}^{(s)} = \left(A_{0,k-1}^{(0,s)}\right)^2 \exp\left(\Gamma_{0,k-1}^{(0,s)}\right) \Phi\left(A_{0,k-1}^{(0,s)}\right) + \left(B_{0,k-1}^{(0,s)}\right)^2 \exp\left(\Delta_{0,k-1}^{(0,s)}\right) \Phi\left(-B_{0,k-1}^{(0,s)}\right),$$

$$A_{k,i}^{(m,j)} = \left(m\beta^{i+k} + j\beta^i\right)\theta^*, \quad B_{k,i}^{(m,j)} = -\left(m\beta^{i+k} + j\beta^{i+k}\right)\gamma^*,$$

$$\Gamma_{k,i}^{(m,j)} = \frac{\left[\left(m\beta^{i+k} + j\beta^i\right)\theta^*\right]^2}{2}, \quad \Delta_{k,i}^{(m,j)} = \frac{\left[\left(m\beta^{i+k} + j\beta^i\right)\gamma^*\right]^2}{2},$$

$$\varpi_k^{(m,j)} = \prod_{i=0}^{\infty}\left[\Phi\left(A_{k,i}^{(m,j)}\right) \exp\left(\Gamma_{k,i}^{(m,j)}\right) + \exp\left(\Delta_{k,i}^{(m,j)}\right) \Phi\left(-B_{k,i}^{(m,j)}\right)\right]$$

and

$$\Lambda_{(k-1)}^{(s)} = \prod_{i=0}^{k-1} \left[\Phi\left(A_{0,i}^{(0,s)}\right) \exp\left(\Gamma_{0,i}^{(0,s)}\right) + \exp\left(\Delta_{0,i}^{(0,s)}\right) \Phi\left(-B_{0,i}^{(0,s)}\right) \right]$$

with $\Lambda_{(-1)}^{(s)} = 1$.

Notice that for a time-invariant EGARCH model, i.e. when $\delta_t = 0$ for all t's, we have that $Cov(r_t, r_{t-k}^2) = 0$ whereas $Cov\left(r_t^2, r_{t-k}\right) = E\left(h_t h_{t-k}^{1/2}\right) D_{k-1}^{(1)} \left(\Upsilon_{k-1}^{(1)}\right)^{-1} \neq 0$ unless $\theta = 0$. On the other hand, if $\theta = 0$ but δ_t is time varying, $Cov(r_t, r_{t-k}^2)$ and $Cov\left(r_t^2, r_{t-k}\right)$ can be nonzero and thus, we do not need the asymmetric EGARCH effect to get dynamic asymmetry, even under the assumption of symmetric distribution of the errors (normality). Furthermore, for large values of δ, then $Corr\left(r_t, r_{t-k}^2\right)$ and $Corr\left(r_t^2, r_{t-k}\right)$ behave like $Corr\left(h_t, h_{t-k}^2\right)$ and $Corr\left(h_t^2, h_{t-k}\right)$, respectively.

Moreover, we also have the covariance of squares which is stated in the theorem below (for a proof see Demos [9]).

Theorem 3. *Under the assumptions of Theorem 1, the autocovariance of squares is:*

$$\begin{aligned} Cov\left(r_t^2, r_{t-k}^2\right) &= \frac{2\varphi^k \varphi_u^2}{1-\varphi^2}\left[2\delta^2 + \frac{\varphi^k \varphi_u^2}{1-\varphi^2}\right] E\left(h_t^2 h_{t-k}^2\right) + E^2\left(\delta_t^2\right) Cov\left(h_t^2, h_{t-k}^2\right) \\ &+ E\left(\delta_t^2\right)\left\{Cov\left(h_t, h_{t-k}^2\right) + Cov\left(h_t^2, h_{t-k}\right) + E\left(h_t^2 h_{t-k}\right) F_{k-1}^{(2)} \left(\Upsilon_{k-1}^{(2)}\right)^{-1}\right\} \\ &+ \left[E\left(\delta_t^2\right) + \frac{2\varphi^k \varphi_u^2}{1-\varphi^2}\right] \delta E\left(h_t^2 h_{t-k}^{3/2}\right) D_{k-1}^{(2)} \left(\Upsilon_{k-1}^{(2)}\right)^{-1} + Cov\left(h_t, h_{t-k}\right) \\ &+ \left[2\delta E\left(h_t h_{t-k}^{3/2}\right) D_{k-1}^{(1)} + E\left(h_t h_{t-k}\right) F_{k-1}^{(1)}\right] \left(\Upsilon_{k-1}^{(1)}\right)^{-1} \end{aligned}$$

Again for large values of δ^2, $Corr\left(r_t^2, r_{t-k}^2\right)$ behaves like $Corr\left(h_t^2, h_{t-k}^2\right)$.

Finally, the skewness and kurtosis are now stated (for a proof see Demos [9]).

Theorem 4. *Under the assumptions of Theorem 1, the third and fourth central moments of* $\{r_t\}$ *are :*

$$E\left[r_t - E\left(r_t\right)\right]^3 = \delta\left(\delta^2 + \frac{3\varphi_u^2}{1-\varphi^2}\right)\left[E\left(h_t^3\right) - E\left(h_t^2\right) E\left(h_t\right)\right] + \delta Var\left(h_t\right)\left[3 - 2\delta^2 E\left(h_t\right)\right]$$

and

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

First, notice that if $\delta = 0$, $E\left[r_t - E\left(r_t\right)\right]^3 = 0$. As mentioned in Demos [9] the absolute value of the skewness of $\{r_t\}$, $|sk_r| = \frac{\left|E[r_t - E(r_t)]^3\right|}{[Var(r_t)]^{3/2}}$, is a decreasing function of φ_u^2 and φ^2, whereas for $\delta \in \left(-\sqrt{\frac{E(h_t)}{2E(h_t^2)}}, \sqrt{\frac{E(h_t)}{2E(h_t^2)}}\right)$, sk_r is an increasing function in δ. Although unconditional asymmetry is not observed in stock market data, this is not the case for exchange rates (see Gallant et al., 1997 [14]) and turbulence datasets (see Barndorff-Nielsen [4]). Furthermore, for $\delta = 0$ the kurtosis of r_t, $\kappa_r = \frac{E[r_t - E(r_t)]^4}{Var^2(r_t)}$, is an increasing function in φ_u^2 and φ^2. However notice that in this case sk_r is zero.

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.

4. LIKELIHOOD-INFERENCE: EM AND BAYESIAN APPROACHES

The purpose of this section is the estimation of a TVPAR(1)-EGARCH(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) because of the fact that 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 [8] 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, that is referred to as the target distribution, a suitable Markov chain is constructed (using a Metropolis-Hasting (M-H) algorithm or a Gibbs-sampling method) with the property that its limiting, invariant distribution is the target distribution. In most problems, the target distribution is absolutely continuous and as a result the theory of MCMC methods is based on that of Markov chains on continuous state spaces (Meyn and Tweedie [29]). This means that by simulating the

Markov chain a large number of times and recording its values a sample of (correlated) draws from the target distribution can be obtained. It should be noted that Markov chain samplers are invariant by construction and therefore the existence of the invariant distribution does not have to be checked in any particular application of MCMC method.

The Metropolis-Hasting algorithm (M-H) is a general MCMC method to produce sample variates from a given multivariate distribution. It is based on a candidate generating density that is used to supply a proposal value that is accepted with probability given as the ratio of the target density times the ratio of the proposal density. There are a number of choices of the proposal density (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 [15]. In this algorithm the parameters are grouped into blocks and each block is sampled according to the full conditional distribution denoted as $\pi\left(\phi_t/\phi_{/t}\right)$. By Bayes' theorem we have $\pi\left(\phi_t/\phi_{/t}\right) \propto \pi\left(\phi_t\phi_{/t}\right)$, the joint distribution of all blocks and so full conditional distributions are usually quite simply derived. One cycle of the Gibbs sampling algorithm is completed by simulating $\{\phi_t\}_{t=1}^{p}$, 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 Tierney [38]). Another application of the MCMC is the analysis of hidden Markov models where the approach relies on augmenting the parameter space to include the unobserved states and simulate the target distribution via the conditional distributions (this procedure is called data augmentation and was pioneered by Tanner and Wong [37]). Kim, Shephard and Chib [24] discuss a MCMC algorithm of the Stochastic Volatility (SV) model which is an example of a state space model in which the state variable h_t (log-volatility) appears non-linearly in the observation equation. The idea is to approximate the model by a conditionally Gaussian state space model with the introduction of multinomial random variables that follow a seven-point discrete distribution.

The analysis of a time-varying EGARCH-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 [35], a regrettable consequence of this path-dependence in volatility is that standard MCMC algorithms will evolve in $O\left(T^2\right)$ computational load (see Giakoumatos, Dellaportas and Politis [18]). 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.

4.1. Estimation problem: Simulated EM algorithm

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

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

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

However, the $\delta_t's$ are unobserved and thus, to classically estimate the model, we have to rely on an EM algorithm (Dempster, Laird and Rudin [10]) to obtain estimates as close to the optimum as desired. At each iteration the EM algorithm obtains $\phi^{(n+1)}$,where ϕ is the parameter vector, by maximizing the expectation of the log-likelihood conditional on the data and the current parameter values i.e. $E\left(\ell\left(.\right)|\mathbf{r},\phi^{(n)},\mathcal{F}_0\right)$ with respect to ϕ keeping $\phi^{(n)}$ fixed.

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

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

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

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

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

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

In terms of identification, the model is not, up to second moment, identified (see Corollary 1 in Sentana and Fiorentini [34]). The reason is that we can transfer unconditional variance from the error, ε_t, to the price of risk, δ_t, and vice versa. One possible solution is to fix α such that $E\left(h_t\right)$ is 1, or to set φ_u to a specific value. In fact in an earlier version of the paper we fixed φ_u to be 1 (see Anyfantaki and Demos [1]). 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 [31]).

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 $\boldsymbol{\delta}|\phi,\mathbf{r},\mathcal{F}_0$.

4.2. Simulation based Bayesian inference

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

$$p(\phi, \boldsymbol{\delta}|\mathbf{r}) \propto p(\phi, \boldsymbol{\delta})\, p(\mathbf{r}|\phi, \boldsymbol{\delta}) \propto p(\phi)\, p(\boldsymbol{\delta}|\phi)\, p(\mathbf{r}|\phi, \boldsymbol{\delta})$$

where

$$\phi = \left(\delta, \varphi, \varphi_u^2, \alpha, \beta, \gamma, \theta\right)'$$

Now,

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

On the other hand,

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

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 [20] 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 $p\left(\{\delta_t\}^{(1)} |\mathbf{r}, \phi^{(0)}\right)$ and then $\phi^{(1)}$ from $p\left(\phi^{(1)}| \{\delta_t\}^{(1)}, \mathbf{r}\right)$. Iterating these steps we finally get $\left(\{\delta_t\}^{(i)}, \phi^{(i)}\right)_{i=1}^{M}$ and under mild conditions it is shown that the distribution of the sequence converges to the joint posterior distribution $p(\phi, \boldsymbol{\delta}|\mathbf{r})$.

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

$$\phi_1 = \left(\delta, \varphi, \varphi_u^2\right)$$

$$\phi_2 = (\alpha, \beta, \gamma, \theta)$$

Then the algorithm is described as follows.

(1) Initialize ϕ

(2) Draw from $p(\delta_t|\delta_{\neq t}, \mathbf{r}, \phi)$

(3) Draw from $p(\phi|\boldsymbol{\delta}, \mathbf{r})$ in the following blocks:

(i)Draw from $p(\phi_1|\boldsymbol{\delta}, \mathbf{r})$ using Gibbs sampling. This is update 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.

4.2.1. Gibbs-Sampling

The task of simulating from an AR model has been already discussed. Here, we will follow the approach of Chib [8] but we do not have any MA terms which makes inference simpler.

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

$$p\left(\delta, \varphi_u^2, \varphi\right) = p\left(\delta|\varphi_u^2\right) p\left(\varphi_u^2\right) p\left(\varphi\right)$$

which means that δ, φ_u^2 is a priory independent of φ.

Also the following holds for the prior distributions of the parameter subvector ϕ_1:

$$p\left(\delta|\varphi_u^2\right) \sim N\left(\delta_{pr}, \varphi_u^2\sigma_{\delta_{pr}}^2\right),$$

$$p\left(\varphi_u^2\right) \sim IG\left(\frac{v_0}{2}, \frac{d_0}{2}\right)$$

and

$$p\left(\varphi\right) \sim N\left(\varphi_0, \sigma_{\varphi_0}^2\right) I_\varphi$$

where I_φ ensures that φ 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

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

From a Bayesian viewpoint the right hand side of the above equation is equal to the "augmented" prior, i.e. the prior augmented by the latent $\boldsymbol{\delta}$. We proceed to the generation of these parameters..

First we see how to generate δ. Following again Chib [7] we may write:

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

or otherwise,

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

Under the above and using Chib's [7] notation we have that the proposal distribution is the following Gaussian distribution (see Chib [7] for a proof).

Proposition 3. *The proposal distribution of δ is:*

$$\delta|\boldsymbol{\delta}, \phi, \varphi_u^2 \sim N\left(\widetilde{\delta}, \varphi_u^2\widetilde{\sigma}_\delta^2\right)$$

where

$$\widetilde{\delta} = \widetilde{\sigma}_{\delta}^2 \left(\frac{\delta_{pr}}{\sigma_{\delta_{pr}}^2} + (1-\varphi) \sum_{t=1}^{T} \delta_t^* \right)$$

and

$$\widetilde{\sigma}_{\delta}^2 = \left(\frac{1}{\sigma_{\delta_{pr}}^2} + (1-\varphi)^2 \right)^{-1}$$

Hence, the generation of δ is completed and we may turn on the generation of the other parameters.

Generation of φ_u^2 For the generation of φ_u^2 and using Chib's [7] notation we have that:

PROPOSITION 4. *The proposal distribution of* φ_u^2 *is:*

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

where

$$Q = (\delta - \delta_{pr})^2 \, \sigma_{\delta_{pr}}^{-2}, \quad \textit{and} \quad d = \sum_{t=2}^{T} \left[\delta_t^* - \delta (1-\varphi) \right]^2.$$

Finally, we turn on the generation of φ.

Generation of φ For the generation of φ we follow again Chib [7] and write:

$$\delta_t = (1-\varphi)\,\delta - \varphi \delta_{t-1} + v_t, \qquad v_t \sim N\left(0, \varphi_u^2\right)$$

We may now state the following Proposition (see Chib [7] for a proof).

PROPOSITION 5. *The proposal distribution of* φ *is:*

$$\varphi^2 | \boldsymbol{\delta}, \delta, \varphi_u^2 \sim N\left(\widetilde{\varphi}, \widetilde{\sigma}_{\varphi}^2 \right)$$

where

$$\widetilde{\varphi} = \widetilde{\sigma}_{\varphi}^2 \left(\sigma_{\varphi_0}^{-2} \varphi_0 + \varphi_u^{-2} \sum_{t=1}^{T} (\delta_{t-1} - \delta)(\delta_t - \delta) \right)$$

and

$$\widetilde{\sigma}_{\varphi}^{-2} = \sigma_{\varphi_0}^{-2} + \varphi_u^{-2} \sum_{t=1}^{T} (\delta_{t-1} - \delta)^2$$

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

4.2.2. Metropolis-Hasting

Step(3)(ii) is the task of simulating from the posterior of the parameters of a EGARCH-M process. This has been already addressed by Vrontos et al. [39]. We could use a simple M-H algorithm where each parameter is updated one at a time (single-component update) by using independent Metropolis steps where a random walk chain is adopted with an increment normal density. However, as Vrontos et al. [39] mention the efficiency of the algorithm may improved if we simultaneously sample a chosen subvector of the parameter vector using multivariate Metropolis steps (simultaneously update). According to their method one should first estimate the sample covariance matrix $\mathbf{\Sigma}$ of the correlated elements of the subvector from an initial explanatory run (e.g. random walk chain) and then update using a multivariate normal proposal density $N\left(\boldsymbol{\mu}^r, c\mathbf{\Sigma}\right)$ where $\boldsymbol{\mu}^r$ denotes the vector at the rth iteration and c is a constant to tune the acceptance rate. Also it would be better to transform the subvector to take values on $(-\infty, +\infty)$.

To illustrate we suppose that first we adopt the random-walk MH as our initial explanatory run. Thus, suppose that we have the general parameter subvector θ. Then the random walk MH requires the following two steps:

Step 1: At iteration j, generate a point θ^* from the random walk kernel, $\theta^* = \theta^{[j-1]} + \varepsilon, \varepsilon \sim N(0, \Omega)$ where $\theta^{[j-1]}$ is the $(j-1)$th MCMC iterate of θ .

Step 2: Accept θ^* as $\theta^{[j]}$ with probability $\pi = min\left(1, f(\theta^*)/f\left(\theta^{[j-1]}\right)\right)$

We select Ω to be a diagonal matrix, whose elements are tuned by monitoring the MH acceptance rate to lie between 25% and 50%, as in Gelfand et al. [16]. The function f is the conditional posterior density.

Now, using these iterations the sample covariance matrix $\mathbf{\Sigma}$ is found. We then update simultaneously the elements of the subvector using a multivariate normal proposal density $N\left(\boldsymbol{\mu}^r, c\mathbf{\Sigma}\right)$.

Step 1: At iteration i, generate a point θ^{**} from the multivariate normal proposal density $N\left(\boldsymbol{\mu}^{i-1}, c\mathbf{\Sigma}\right)$ where $\boldsymbol{\mu}^{i-1}$ denotes the vector at the $(i-1)$th iteration and c is a constant to tune the acceptance rate.

Step 2: Accept θ^{**} as $\theta^{[i]}$ with probability $\pi = min\left(1, \frac{f(\theta^{**})}{f\left(\theta^{[i-1]}\right)} \frac{g\left(\theta^{[i-1]}\right)}{g((\theta^{**})}\right)$

Where $g\left(\theta\right) \propto \exp\left(-\frac{1}{2}\left(\theta - \boldsymbol{\mu}\right)^\top \left(c\mathbf{\Sigma}\right)^{-1}\left(\theta - \boldsymbol{\mu}\right)\right)$ is the Gaussian proposal density.

The advantage of using $\mathbf{\Sigma}$ in the MH algorithm is that the posterior correlations among the elements of θ can be accounted for, increasing the efficiency of the Markov chain and thus speeding up convergence.

For the parameters of the EGARCH model we use $U\left(-1, 1\right)$ prior for β and normal priors for the other parameters α, γ, δ as $N\left(0, 10\right)$. These priors are practically noninformative.

The algorithm described above is a special case of a MCMC algorithm, which converges as it iterates, to draws from the required density $p(\phi, \boldsymbol{\delta}|\mathbf{r})$. 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 ϕ is simply estimated by the sample mean of the simulated ϕ values. These estimated values can be made arbitrarily accurate by increasing the simulation sample size. However, it should be remembered that sample variates from a MCMC algorithm are a high dimensional (correlated) sample from the target density and sometimes the serial correlation can be quite high for badly behaved 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 $\boldsymbol{\delta}|\phi, \mathbf{r}, \mathcal{F}_0$.

4.2.3. MCMC Simulation of $\boldsymbol{\varepsilon}|\phi, \mathbf{r}, \mathcal{F}_0$

For a given set of parameter values and initial conditions it is generally simpler to simulate $\{\varepsilon_t\}$ for $t = 1, ..., T$ and then compute $\{\delta_t\}_{t=1}^{T}$ than to simulate $\{\delta_t\}_{t=1}^{T}$ directly. For that matter, we concentrate on simulators of ε_t given $\mathbf{r}$ and ϕ. We set the mean and the variance of ε_0 equal to their unconditional values and given that h_t is a sufficient statistic for $\mathcal{F}_{t-1}$and the unconditional variance is a deterministic function of ϕ, $\mathcal{F}_0$ can be eliminated from the information set without any information loss.

Now sampling from $p(\varepsilon|\mathbf{r}, \phi) \propto p(\mathbf{r}|\varepsilon, \phi)\, p(\varepsilon|\phi)$ is feasible by using a M-H algorithm where we update each time only one ε_t leaving all the other unchanged (Shephard, [35]). In particular, let us write the n^{th} iteration of a Markov chain as ε^n. Then we generate a potential new value of the Markov chain ε^{new} by proposing from some candidate density $g\left(\varepsilon_t|\varepsilon^n_{\backslash t}, \mathbf{r}, \phi\right)$ where $\varepsilon^n_{\backslash t} = \left\{\varepsilon_1^{n+1}, ..., \varepsilon_{t-1}^{n+1}, \varepsilon_{t+1}^{n}, ..., \varepsilon_T^n\right\}$ which we accept with probability

$$\min\left[1, \frac{p\left(\varepsilon_t^{new}|\varepsilon^n_{\backslash t}, \mathbf{r}, \phi\right) g\left(\varepsilon_t^{new}|\varepsilon^n_{\backslash t}, \mathbf{r}, \phi\right)}{p\left(\varepsilon_t^{n}|\varepsilon^n_{\backslash t}, \mathbf{r}, \phi\right) g\left(\varepsilon_t^{n}|\varepsilon^n_{\backslash t}, \mathbf{r}, \phi\right)}\right]$$

If it is accepted then we set $\varepsilon_t^{n+1} = \varepsilon_t^{new}$ 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:

$$\frac{p\left(\varepsilon_t^{new}|\varepsilon_{\backslash t}^{n},\mathbf{r},\phi\right)}{p\left(\varepsilon_t^{n}|\varepsilon_{\backslash t}^{n},\mathbf{r},\phi\right)} = \frac{p\left(r_t|\varepsilon_t^{new},h_t^{new,t},\phi\right)p\left(\varepsilon_t^{new}|h_t^{new,t},\phi\right)p\left(r_t|h_t^{n,t},\phi\right)}{p\left(r_t|h_t^{new,t},\phi\right)p\left(r_t|\varepsilon_t^{n},h_t^{n,t},\phi\right)p\left(\varepsilon_t^{new}|h_t^{n,t},\phi\right)}$$

$$* \prod_{s=t+1}^{T}\frac{p\left(r_s|\varepsilon_s^{r},h_s^{new,t},\phi\right)p\left(\varepsilon_s^{r}|h_s^{new,t},\phi\right)p\left(r_s|h_s^{n,t},\phi\right)}{p\left(r_s|h_s^{new,t},\phi\right)p\left(r_s|\varepsilon_s^{n},h_s^{n,t},\phi\right)p\left(\varepsilon_s^{n}|h_s^{n,t},\phi\right)}$$

$$= \frac{p\left(r_t|\varepsilon_t^{new},h_t^{new,t},\phi\right)p\left(\varepsilon_t^{new}|h_t^{new,t},\phi\right)}{p\left(r_t|\varepsilon_t^{n},h_t^{n,t},\phi\right)p\left(\varepsilon_t^{new}|h_t^{n,t},\phi\right)}$$

$$* \prod_{s=t+1}^{T}\frac{p\left(r_s|\varepsilon_s^{n},h_s^{new,t},\phi\right)p\left(\varepsilon_s^{n}|h_s^{new,t},\phi\right)}{p\left(r_s|\varepsilon_s^{n},h_s^{n,t},\phi\right)p\left(\varepsilon_s^{n}|h_s^{n,t},\phi\right)}$$

where for $s = t+1,..,T$

$$h_s^{new,t} = V\left(\varepsilon_s|\varepsilon_{s-1}^{n},\varepsilon_{s-2}^{n},...,\varepsilon_{t+1}^{n},\varepsilon_t^{new},\varepsilon_{t-1}^{n+1},..,\varepsilon_1^{n+1}\right)$$

$$h_s^{n,t} = V\left(\varepsilon_s|\varepsilon_{s-1}^{n},\varepsilon_{s-2}^{n},..,\varepsilon_{t+1}^{n},\varepsilon_t^{n},\varepsilon_{t-1}^{n+1},..,\varepsilon_1^{n+1}\right)$$

while

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

Nevertheless, each time we revise one ε_t we have also to revise $T-t$ conditional variances because of the recursive nature of the GARCH model which makes $h_s^{new,t}$ depend upon ε_t^{new} for $s = t+1,..,T$. And since $t = 1,..,T$ it is obvious that we need to calculate T^2 normal densities and so this algorithm is $O\left(T^2\right)$. And this should be done for every ϕ. To avoid this huge computational load we show how to use the method proposed by Fiorentini et al. [13] and so do MCMC with only $O\left(T\right)$ calculations. The method is described in the following subsection.

4.3. Estimation method proposed: Classical and Bayesian estimation

The method proposed by Fiorentini et al. [13] is to transform the GARCH model into a first order Markov model and so do MCMC with only $O\left(T\right)$ calculations.

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

$$s_t = sign\left(z_t\right)$$

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

$$z_t = \frac{\ln h_{t+1} - \alpha - \beta \ln h_t}{\gamma \pm \theta} \quad \forall t \geq 1$$

where we have $\gamma+\theta$ for positive z_t and $\gamma-\theta$ for negative z_t.

Hence the additional knowledge of the signs of z_t would reveal the entire path of $\{z_t\}$ 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 $\{\delta_t\}|\mathbf{r},\varphi,\{h_{t+1}\}$.

Now we have to sample from:

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

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

$$z_t=\frac{\ln h_{t+1}-\alpha-\beta\ln h_t}{\gamma\pm\theta}$$

where again we have $\gamma+\theta$ for positive z_t and $\gamma-\theta$ for negative z_t.

Although, we alter the volatility process when we flip from $s_t=-1$ to $s_t=1$ (implying that the signs cause the volatility process) and also we alter z_t, as it will become obvious below we do not have to simulate the signs but only for $t=T$. Hence, we may ignore this and simulate first $\{h_{t+1}\}|\mathbf{r},\phi$ and then simulate $\{s_t\}|\{h_{t+1}\},\mathbf{r},\phi$. 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.

4.3.1. Simulations of $\{s_t\}|\{h_{t+1}\},r,\phi$

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

$$c_t=\frac{1}{\sqrt{v_{t|r_t,h_t}}}\left[\varphi\left(\frac{\frac{\ln h_{t+1}-\alpha-\beta\ln h_t}{\gamma+\delta}-z_{t|r_r,h_t}}{\sqrt{v_{t|r_t,h_t}}}\right)+\varphi\left(\frac{\frac{\ln h_{t+1}-\alpha-\beta\ln h_t}{\gamma-\delta}-z_{t|r_r,h_t}}{\sqrt{v_{t|r_t,h_t}}}\right)\right]$$

where

$$z_{t|r_t,h_t}=E(z_t|r_t,h_t)=\frac{(1-\varphi^2)(r_t-\delta h_t)}{h_t^{1/2}(\varphi_u^2h_t+1-\varphi^2)},\qquad v_{t|r_t,h_t}=Var(z_t|r_t,h_t)=\frac{\varphi_u^2h_t}{\varphi_u^2h_t+1-\varphi^2}$$

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 $z_t=\frac{\ln h_{t+1}-\alpha-\beta\ln h_t}{\gamma+\theta}$

conditional on h_{t+1}, h_t, r_t, ϕ, which is given by:

$$\begin{aligned} p\left(s_t = 1 | \left\{h_{t+1}\right\}, r, \varphi\right) &= p\left(z_t = \frac{\ln h_{t+1} - \alpha - \beta \ln h_t}{\gamma + \theta} | h_{t+1}, h_t, r_t, \varphi\right) \\ &= \frac{1}{c_t \sqrt{v_{t|r_t,h_t}}} \varphi\left(\frac{\frac{\ln h_{t+1} - \alpha - \beta \ln h_t}{\gamma + \theta} - z_{t|r_r,h_t}}{\sqrt{v_{t|r_t,h_t}}}\right) \end{aligned}$$

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

$$\varphi\left(\frac{\frac{\ln h_{t+1} - \alpha - \beta \ln h_t}{\gamma + \theta} - z_{t|r_r,h_t}}{\sqrt{v_{t|r_t,h_t}}}\right) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\left(\frac{\frac{\ln h_{t+1} - \alpha - \beta \ln h_t}{\gamma + \theta} - z_{t|r_r,h_t}}{\sqrt{v_{t|r_t,h_t}}}\right)^2\right\}$$

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

1) Specify an initial value $s^{(0)} = \left(s_1^{(0)}, ..., s_T^{(0)}\right)$

2) Repeat for $n = 1, ..., M$

(a) Repeat for $t = 0, .., T - 1$

(i) Draw $s^{(n)} = 1$ with probability

$$\frac{1}{c_t \sqrt{v_{t|r_t,h_t}}} \varphi\left(\frac{\frac{\ln h_{t+1} - \alpha - \beta \ln h_t}{\gamma + \theta} - z_{t|r_r,h_t}}{\sqrt{v_{t|r_t,h_t}}}\right)$$

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

$$1 - \frac{1}{c_t \sqrt{v_{t|r_t,h_t}}} \varphi\left(\frac{\frac{\ln h_{t+1} - \alpha - \beta \ln h_t}{\gamma + \theta} - z_{t|r_r,h_t}}{\sqrt{v_{t|r_t,h_t}}}\right)$$

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

4.3.2. Simulations of $\{h_{t+1}\} / \mathbf{r}, \phi$ (single move samplers)

On the other hand, the first step involves simulating from $\{h_{t+1}\} | \mathbf{r}, \phi$. To avoid large dependence in the chain we use a M-H algorithm where we simulate one h_{t+1} at a time leaving the others unchanged (Shephard [35] and Wei [40]). So if $\left(h_{t+1}\right)^n$is the current value of the nth iteration of a Markov chain then we draw a candidate value of the Markov chain h_{t+1}^{new} by proposing it from a candidate density (proposal density)

$g\left(h_{t+1}|\,(h)^{n}_{/t+1}\,,\mathbf{r},\phi\right)$ where $(h)^{n}_{/t+1}=\left\{h^{n+1}_{1},h^{n+1}_{2},..,h^{n+1}_{t},h^{n}_{t+2},..,h^{n}_{T+1}\right\}$.
We set $(h_{t+1})^{n+1}=(h_{t+1})^{new}$ with acceptance probability

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

where we have used the fact that

$$p\left(\mathbf{h}|\mathbf{r},\phi\right)=p\left((h)_{/t}\,|\mathbf{r},\phi\right)p\left(h_{t}|\,(h)_{/t}\,,\mathbf{r},\phi\right)$$

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

$$p\left(h_{t+1}|\,(h)_{/t+1}\,,\mathbf{r},\phi\right)\propto p\left(h_{t+2}|h_{t+1},\phi\right)p\left(h_{t+1}|h_{t},\phi\right)p\left(r_{t+1}|h_{t+2},h_{t+1},\phi\right)p\left(r_{t}|h_{t+1},h_{t},\phi\right)$$

Now, the following should hold:

$$\ln h_{t+1}\geq\alpha+\beta\ln h_{t}$$

assuming that $\gamma+\theta\geq 0$ and $\gamma-\theta\leq 0$. This makes sense from an economics view point. The squared volatility as a function of z_{t-1} should be nondecreasing on the positive real line (i.e. $\gamma+\delta\geq 0$) and nonincreasing on the negative real line (i.e. $\delta-\gamma\geq 0$). Altogether $\gamma z+\theta\,|z|\geq 0$ for all $z\in R$ or $\theta\geq|\gamma|$.

Similarly

$$\ln h_{t+1}\leq\beta^{-1}\left(\ln h_{t+2}-\alpha\right)$$

and thus, we have that the support of the conditional distribution of h_{t+1} given h_t is bounded from below and the same applies to the distribution of h_{t+2} given h_{t+1}. This means that the range of values of $\ln h_{t+1}$ compatible with $\ln h_t$ and $\ln h_{t+2}$ in the EGARCH case is bounded from above and below i.e.:

$$\ln h_{t+1}\epsilon\left[\alpha+\beta\ln h_{t},\beta^{-1}\left(\ln h_{t+2}-\alpha\right)\right]$$

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

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

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. [13] we can achieve a substantially higher acceptance rate if we propose from

$$g\left(h_{t+1}|\,(h)_{/t+1}\,,\mathbf{r},\phi\right)=p\left(h_{t+1}|r_{t},h_{t},\phi\right)$$

A numerically efficient way to simulate h_{t+1} from $p(h_{t+1}|r_t, h_t, \phi)$ is to sample

$$z_t|r_t, h_t, \phi \sim N\left(\frac{\left(1-\varphi^2\right)(r_t - \delta h_t)}{h_t^{1/2}\left(\varphi_u^2 h_t + 1 - \varphi^2\right)}, \frac{\varphi_u^2 h_t}{\varphi_u^2 h_t + 1 - \varphi^2}\right) \tag{5}$$

so that the following upper bound is satisfied:

$$\gamma z_t^{new} + \theta \left|z_t^{new}\right| \leq \frac{\ln h_{t+2} - \alpha - \beta\alpha - \beta^2 \ln h_t}{\beta} \tag{6}$$

using an accept-reject method and then compute

$$\ln h_{t+1}^{new} = \alpha + \beta \ln h_t + \gamma z_t^{new} + \theta \left|z_t^{new}\right|$$

which in turn guarantees that $\ln h_{t+1}^{new}$ lies within the acceptance bounds.

For the accept-reject method we draw

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

and accept the draw if $\gamma z_t^{new} + \theta \left|z_t^{new}\right| \leq \frac{\ln h_{t+2} - \alpha - \beta\alpha - \beta^2 \ln h_t}{\beta}$ and otherwise we repeat the drawing (this method is inefficient if the truncation lies in the tails of the distribution).

The conditional density of z_t^{new} will be given according to the definition of a truncated normal distribution and by using the change of variable formula we have that the density of h_{t+1}^{new} will be:

$$\begin{aligned}
& p\left(h_{t+1}^{new} | \ln h_{t+1}^{new} \in \left[\alpha + \beta \ln h_t, \beta^{-1}\left(\ln h_{t+2} - \alpha\right)\right], r_t, h_t, \varphi\right) \\
= \quad & \frac{c_t^{new}}{(\gamma \pm \theta)\, h_{t+1}^{new}} \left[\Phi\left(\frac{\frac{\ln h_{t+2} - \alpha - \beta\alpha - \beta^2 \ln h_t}{\beta(\gamma \pm \theta)} - z_{t|r_r,h_t}}{\sqrt{v_{t/r_t,h_t}}}\right)\right]^{-1}
\end{aligned}$$

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

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

$$\min\left(1, \frac{p\left(h_{t+2}|h_{t+1}^{new}, r_{t+1}, \phi\right) p\left(r_{t+1}|h_{t+1}^{new}, \phi\right)}{p\left(r_{t+1}|h_{t+1}^{n}, \phi\right) p\left(h_{t+2}|h_{t+1}^{n}, r_{t+1}, \phi\right)}\right)$$

Now, since the degree of truncation is same for old and new the acceptance probability will be:

$$\min\left(1, \frac{p\left(r_{t+1}|h_{t+1}^{new}\right)}{p\left(r_{t+1}|h_{t+1}^{n}\right)} \frac{c_{t+1}^{new} h_{t+1}^{n}}{c_{t+1}^{n} h_{t+1}^{new}}\right)$$

where $p\left(r_{t+1}|h_{t+1}\right)$ is a mixture of two univariate normal densities and hence:

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

So the acceptance probability becomes:

$$\min\left[1,\left(\frac{h_{t+1}^{2n}}{h_{t+1}^{2new}}\right)\frac{\kappa\left(h_{t+1}^{new}\right)}{\kappa\left(h_{t+1}^{n}\right)}\right] \tag{7}$$

where

$$\begin{aligned}\kappa\left(h_{t+1}^{i}\right) &= \left[\begin{array}{c}\exp\left\{-\frac{1}{2}\frac{\left(\varphi_u^2h_{t+1}^{i}+1-\varphi^2\right)}{\varphi_u^2h_{t+1}^{i}}\left(\frac{\ln h_{t+2}-\alpha-\beta\ln h_{t+1}^{i}}{\gamma+\theta}-\frac{\left(1-\varphi^2\right)\left(r_{t+1}-\delta h_{t+1}^{i}\right)}{h_{t+1}^{1/2i}\left(\varphi_u^2h_{t+1}^{i}+1-\varphi^2\right)}\right)^2\right\}\\ +\exp\left\{-\frac{1}{2}\frac{\left(\varphi_u^2h_{t+1}^{i}+1-\varphi^2\right)}{\varphi_u^2h_{t+1}^{i}}\left(\frac{\ln h_{t+2}-\alpha-\beta\ln h_{t+1}^{i}}{\gamma-\theta}-\frac{\left(1-\varphi^2\right)\left(r_{t+1}-\delta h_{t+1}^{i}\right)}{h_{t+1}^{1/2i}\left(\varphi_u^2h_{t+1}^{i}+1-\varphi^2\right)}\right)^2\right\}\end{array}\right]\\ &\times\exp\left(-\frac{\left(r_{t+1}-\delta_1h_{t+1}^{i}\right)^2}{2\left(\frac{\varphi_u^2}{1-\varphi^2}h_{t+1}^{i}+1\right)h_{t+1}^{i}}\right)\end{aligned}$$

Overall the MCMC of $\{h_{t+1}\}\,|\mathbf{r},\varphi$ includes the following steps:
(1) Specify an initial value $\left\{h^{(0)}\right\}$
(2) Repeat for $n=1,...,M$
(a) Repeat for $t=0,..,T-1$
(i) Use an accept-reject method to simulate

$$z_t^{new}|r_t,h_t,\varphi\sim N\left(\frac{\left(1-\varphi^2\right)\left(r_t-\delta h_t\right)}{h_t^{1/2}\left(\varphi_u^2h_t+1-\varphi^2\right)},\frac{\varphi_u^2h_t}{\varphi_u^2h_t+1-\varphi^2}\right)$$

truncated from above

$$\gamma z_t^{new}+\delta\left|z_t^{new}\right|\leq\frac{\ln h_{t+2}-\alpha-\beta\alpha-\beta^2\ln h_t}{\beta}$$

(ii) Calculate

$$\ln h_{t+1}^{new}=\alpha+\beta\ln h_t+\gamma z_t^{new}+\theta\left|z_t^{new}\right|$$

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

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

appropriately truncated so that:

$$\ln h_{t+1} \epsilon \left[\alpha + \beta \ln h_t, \beta^{-1} \left(\ln h_{t+2} - \alpha\right)\right]$$

(iii) Calculate

$$\alpha_n = \min \left[1, \left(\frac{h_{t+1}^n}{h_{t+1}^{new}}\right) \frac{\kappa\left(h_{t+1}^{new}\right)}{\kappa\left(h_{t+1}^n\right)}\right]$$

(iii) Set

$$\left(h_{t+1}\right)^{n+1} = \begin{Bmatrix} \left(h_{t+1}\right)^{new} & if & Unif\,(0,1) \leq \alpha_r \\ \left(h_{t+1}\right)^n & & otherwise \end{Bmatrix}$$

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

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

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

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

5. EMPIRICAL APPLICATION: BAYESIAN ESTIMATION OF WEEKLY EXCESS RETURNS FROM THREE MAJOR STOCK MARKETS: DOW-JONES, FTSE AND NIKKEI

In this section we investigate the practical performance of the procedures described above. To do this, we use weekly excess returns from three major stock markets: Dow-Jones, FTSE and Nikkei for the period 1979:8 to 2008:5 (1500 observations). To guarantee $|\varphi| \geq 1$ and to ensure that $0 \leq \beta \leq 1$ and $\theta \geq |\gamma|$ 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 $\theta \geq |\gamma|$, $\beta > 0, \beta < 1$ and $|\varphi| \geq 1$.

In order to implement our proposed Bayesian approach, we first have to specify the hyperparameters that characterize the prior distributions of the parameters. In this respect, our aim was to use informative priors that would be in accordance with the "received wisdom". In particular, for all data sets we set the prior mean for a equal to -0.06 and for θ, γ and β equal to $0.25, -0.09$ and 0.78 respectively. These prior means imply an annual excess return of around 3%, which is a typical value for annualized stock excess returns. In order to diminish the impact of the prior on the joint posterior, we use rather vague priors by setting the prior variance of the skedastic function's parameters a ,θ, γ and β to $10,000$ for all datasets. Moreover, for prior mean of δ we used 0.03 for all markets and again we set its prior variance equal to $10,000$. Finally, we set the prior mean of φ equal to 0.80 for all three datasets, of φ_u^2 equal to 0.01 and the hyperparameters ν_0 and d_o 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. In any case, we performed a sensitivity analysis with respect to the variance hyperparameters and confirmed that our initial choice is vague enough and does not introduce significant information in our estimation.

We run a chain for $200,000$ 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 1**. Inefficiency 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.[2] The parameter γ which measures the sign effect is as expected negative for all datasets while the parameter θ is positive since it measures the size effect of the shocks on the volatility. This means that volatility reacts asymmetrically to the bad and good news. 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 1,2,3** 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 1,2,3** indicate that the choice of priors does not unduly influence our conclusions. In particular, the negativity of the parameter γ and the positivity of the price of

[2] This is also justified by the ACFs of the draws. However, they are not presented for space considerations and are available upon request.

risk coefficient δ seem to be robust.

Finally, treating the posterior means as the "true" parameters and employing the results in Section 3, one can compare the theoretical moments and correlations with the sample equivalents. This comparison is presented in **Table 2**. In the second and third columns the sample and theoretical moments and correlations are presented for the Dow Jones, whereas the analogous results for the FTSE and Nikkei are presented in columns four to seven. It is obvious that the theoretical standard deviations are uniformly, over the three markets, smaller than the sample ones and the same is true for the kurtosises. On the other hand the model is delivering skewnesses that have opposite sign than the sample counterparts. The same is true for the autocorrelations, i.e. the theoretical ones are only positive. However, the theoretical and the sample dynamic asymmetries, what is commonly called leverage effect, are very close, and the same is true for the theoretical and sample volatility clustering measures, i.e. $\rho\left(r_t^2, r_{t-k}^2\right)$. In short, treating the posterior means as the "true" parameters, the model delivers positive autocorrelations, negative leverage effects, volatility clustering, and satisfactory approximations to the sample means. However, it underestimates the variances and the kurtosises and overestimates the skewnesses.

6. EXTENSIONS AND CONCLUSIONS

In this paper, we derive exact likelihood based estimators for our time varying EGARCH(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 is concerned.

In order to investigate the practical performance of the proposed procedure, we estimate within a Bayesian context our TVPAR(1)-EGARCH(1,1)-M model for weekly excess stock returns from the Dow-Jones, Nikkei and FTSE index. We leave for further research the empirical application of the classical estimation procedure.

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 switching process. A Bernoulli distributed price of risk would allow a negative third moment by appropriately choosing the two values of the in-mean process. However, 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 classical procedures are quite in agreement (see Anyfantaki and Demos [1]).

Furthermore, one can extend the proposed method to other conditionally heteroskedastic models. Let us consider the model of Hentschel [23], i.e. let us modify 3 so that now it reads

$$\begin{aligned} \frac{h_t^\lambda - 1}{\lambda} &= \alpha + \beta \frac{h_{t-1}^\lambda - 1}{\lambda} + \gamma h_{t-1}^\lambda f^\nu\left(z_{t-1}\right), \quad \textit{where} \\ f(z_t) &= \left|z_t - b\right| - c\left(z_t - b\right). \end{aligned} \tag{8}$$

Notice that this model nests most popular symmetric and asymmetric GARCH models, e.g. for $\lambda = b = 0$, $\nu = 1$ and c free is the EGARCH of Nelson (1991) [30], $\lambda = \nu = 2$, and $b = c = 0$ is the GARCH of Bollerslev [5], $\lambda = \nu = 2$, $b = 0$, and c free is the GARCH of Glosten et al. [19], etc. (see the paper of Hentschel [23] for detailed discussion on this).[3]

Again, we can augment the state vector with the variables h_{t+1} and then sample the joint Markov process $\{h_{t+1}, s_t\}|\mathbf{r}, \phi \in \mathcal{F}_t$, where now $s_t = sign\left(z_t - b\right)$ so that $s_t = \pm 1$ with probability one, as before. The mapping is one-to-one and has no singularities, provide that $c \neq \pm 1$. More specifically if we know $\{h_{t+1}\}$ and ϕ then we know the value of

$$z_t = b - \left(\frac{\frac{h_{t+1}^\lambda - 1}{\lambda} - \alpha - \beta \frac{h_t^\lambda - 1}{\lambda}}{\gamma h_t^\lambda} \right)^{1/\nu} \frac{1}{1 \mp c} \quad \forall t \geq 1$$

where we have $1 - c$ for $z_t \geq b$ and $1 + c$ for negative $z_t < b$. Again, the additional knowledge of the signs of $(z_t - b)$ would reveal the entire path of $\{z_t\}$ so long as h_0 is known and thus the unobserved random variable $\{\delta_t\}|\mathbf{r}, \varphi, \{h_{t+1}\}$ is also revealed. Of course, the formulae in sections 4.3.1 and 4.3.2 are modified accordingly.

Finally, it is known that (e.g. Tanner [36], pp. 84-85) 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

[3] Notice that the GQARCH model of Sentana [33] is not nested in this specification. However, the GQARCH model is analized in Fiorentini et al. [13] and in Anyfantaki and Demos [2].

allowing the computation of the score. In particular, it is easy to see that:

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

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

ACKNOWLEDGMENTS

We would like to thank S. Arvanitis, G. Fiorentinni 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 PhD Conference in Economics 2009, in Memory of Vassilis Patsatzis (Athens, May 2009), the 3rd International Conference (CFE) in Computational and Financial Econometrics (Limassol, October 2009), the 4th International Conference (CFE) in Computational and Financial Econometrics (London, December 2010), the 5th International Conference (CFE) in Computational and Financial Econometrics (London, December 2011), the 4th Ioannina Meeting (IMAEF) in Applied Economics and Finance (June 2012), the 2012 European Meeting of the Econometric Society (Malaga, August 2012) and the seminar participants at Athens University of Economics and Business for useful discussions. Finally, we acknowledge Landis Conrad's help for the collection of the data.

APPENDIX A: *PROOF OF (5) AND (7)*

Proof. of 5
This is easily derived using the fact that:

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

where

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

and consequently,

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

and thus from the definition of the bivariate normal:

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

$$Var(\varepsilon_t|r_t, h_t) = \frac{\varphi_u^2 h_t^2}{\varphi_u^2 h_t + 1 - \varphi^2}$$

Consequently,

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

∎

Proof. of 7
We have that

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

and thus,

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

Also:

$$c_t = \frac{1}{\sqrt{v_{t|r_t,h_t}}}\left[\varphi\left(\frac{\frac{\ln h_{t+1} - \alpha - \beta\ln h_t}{\gamma+\delta} - z_{t|r_r,h_t}}{\sqrt{v_{t|r_t,h_t}}}\right) + \varphi\left(\frac{\frac{\ln h_{t+1} - \alpha - \beta\ln h_t}{\gamma-\delta} - z_{t|r_r,h_t}}{\sqrt{v_{t|r_t,h_t}}}\right)\right]$$

where

$$z_{t|r_t,h_t} = E\left(z_t|r_t,h_t\right) = \frac{\left(1-\varphi^2\right)\left(r_t-\delta h_t\right)}{h_t^{1/2}\left(\varphi_u^2 h_t + 1 - \varphi^2\right)}, \qquad v_{t|r_t,h_t} = Var\left(z_t|r_t,h_t\right) = \frac{\varphi_u^2 h_t}{\varphi_u^2 h_t + 1 - \varphi^2}$$

and so,

$$\frac{c_{t+1}^{new}}{c_{t+1}^{n}} = \frac{h_{t+1}^{1/2n}\sqrt{\varphi_u^2 h_{t+1}^{new} + 1 - \varphi^2}\left[\begin{array}{c} \exp\left(-\frac{1}{2}\frac{\left(\varphi_u^2 h_{t+1}^{new}+1-\varphi^2\right)}{\varphi_u^2 h_{t+1}^{new}}\left(\begin{array}{c}\frac{\ln h_{t+2}-\alpha-\beta\ln h_{t+1}^{new}}{\gamma+\delta} \\ -\frac{\left(1-\varphi^2\right)\left(r_{t+1}-\delta h_{t+1}^{new}\right)}{h_{t+1}^{1/2new}\left(\varphi_u^2 h_{t+1}^{new}+1-\varphi^2\right)}\end{array}\right)^2\right) \\ +\exp\left(-\frac{1}{2}\frac{\left(\varphi_u^2 h_{t+1}^{new}+1-\varphi^2\right)}{\varphi_u^2 h_{t+1}^{new}}\left(\begin{array}{c}\frac{\ln h_{t+2}-\alpha-\beta\ln h_{t+1}^{new}}{\gamma-\delta} \\ -\frac{\left(1-\varphi^2\right)\left(r_{t+1}-\delta h_{t+1}^{new}\right)}{h_{t+1}^{1/2new}\left(\varphi_u^2 h_{t+1}^{new}+1-\varphi^2\right)}\end{array}\right)^2\right)\end{array}\right]}{h_{t+1}^{1/2new}\sqrt{\varphi_u^2 h_{t+1}^{n} + 1 - \varphi^2}\left[\begin{array}{c} \exp\left(-\frac{1}{2}\frac{\left(\varphi_u^2 h_{t+1}^{r}+1-\varphi^2\right)}{\varphi_u^2 h_{t+1}^{r}}\left(\begin{array}{c}\frac{\ln h_{t+2}-\alpha-\beta\ln h_{t+1}^{n}}{\gamma+\delta} \\ -\frac{\left(1-\varphi^2\right)\left(r_{t+1}-\delta h_{t+1}^{n}\right)}{h_{t+1}^{1/2n}\left(\varphi_u^2 h_{t+1}^{n}+1-\varphi^2\right)}\end{array}\right)^2\right) \\ +\exp\left(-\frac{1}{2}\frac{\left(\varphi_u^2 h_{t+1}^{n}+1-\varphi^2\right)}{\varphi_u^2 h_{t+1}^{n}}\left(\begin{array}{c}\frac{\ln h_{t+2}-\alpha-\beta\ln h_{t+1}^{n}}{\gamma-\delta} \\ -\frac{\left(1-\varphi^2\right)\left(r_{t+1}-\delta h_{t+1}^{n}\right)}{h_{t+1}^{1/2n}\left(\varphi_u^2 h_{t+1}^{n}+1-\varphi^2\right)}\end{array}\right)^2\right)\end{array}\right]}$$

And the result comes straightforward.

REFERENCES

[1] Anyfantaki, S. and A. Demos (2010), Estimation of Time-Varying GARCH-M Models, *Working Paper*

[2] Anyfantaki, S. and A. Demos (2011), Estimation and Properties of a Time-Varying GQARCH(1,1)-M Model, *Journal of Probability and Statistics*, Article ID 718647, 39 pages.

[3] Arvanitis, S. and A. Demos (2004), Time Dependence and Moments of a Family of Time-Varying Parameter GARCH in Mean Models, *Journal of Time Series Analysis*, Vol.25, No1,1-25.

[4] Barndorff-Nielsen, O.E. and N. Shephard (2000), Modelling by Levy Processes for Financial Econometrics, *D.P. No 2000-W3, Nuffield College,* Oxford University.

[5] Bollerslev, T. (1986), Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*,31,307-27.

[6] Bollerslev, T., R.Y. Chou and K.F. Kroner (1992), ARCH Modelling in Finance: A Review of the Theory and Empirical Evidence, *Journal of Econometrics* ,52,5-59.

[7] Chib, S. (1993), Bayes Regression with Autoregressive Errors, *Journal of Econometrics,* 58, 275-294.

[8] Chib, S. (2001), Markov Chain Monte Carlo Methods: Computation and Inference, in *Handbook of Econometrics*, Vol.5, ed. by J.J.Heckman and E.Leamer. Amsterdam:North-Holland,3569-3649.

[9] Demos, A. (2002), Moments and Dynamic Structure of a Time-varying Parameter Stochastic Volatility in Mean Model, *The Econometrics Journal*, 5, 345-357.

[10] Demster, A.P., N. Laird, and D.B. Rubin(1977), Maximum likelihood from Incomplete Data via the EM Algortihm, J*ournal of the Royal Statistical Society*, Series B, 39,1-38.

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

[12] Engle, R.F., D.M. Lilien and R.P. Robins (1987), Estimating Time-Varying Risk-Premia in the Term Structure: The ARCH-Models, *Econometrica*,55, 391-407.

[13] Fiorentini, G., E. Sentana and N. Shephard (2004), Likelihood-Based Estimation of Latent Generalized ARCH Structures, *Econometrica*,Vol.72, No.5,1481-1517.

[14] Gallant, A.R., D. Hsieh and G. Tauchen (1997), Estimation of Stochastic Volatility Models with Diagnostics, *Journal of Econometrics,* 81, 159-192.

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

[16] Gelfand, A .E., A.F.M. Smith and T.M. L ee (1992), BayesianA nalysis of Constrained Parameter and Truncated Data Problems Using Gibbs Sampling, *Journal of the American Statistical Association*, 873, 523-532.

[17] Geweke, J. (1989), Bayesian Inference in Econometric Models Using Monte Carlo Integration", *Econometrica*, 57, 1317-1339.

[18] Giakoumatos, S.G., P. Dellaportas and D.M. Politis (2005), Bayesian Analysis of the Unobserved ARCH Model, *Statistics and Computing*, 15, 103-111.

[19] Glosten, L.R., R. Jaganathan and D.E. Runkle (1993), On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, *Journal of Finance*,48, 1101-801.

[20] Hammersley, J. M. and P. Clifford (1971), Markov Random Fields in Statistics, unpublished, Oxford University.

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

[22] He C., T. Terasvirta and H. Malmsten (2002), Moment Structure of a Family of a Family of First-Order Exponential GARCH Models, *Econometric Theory*, 18, 868–885.

[23] Hentschel, L. (1995), All in the family. Nesting symmetric and asymmetric GARCH models, *Journal of Financial Economics*,39,71-104

[24] Kim, S., N. Shephard and S. Chib (1998), Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models, *Review of Economic Studies*, 65, 361-393.

[25] Mantegna, R.N. and H.E. Stanley (1996), Turbulence and Financial Markets, *Nature*, 383,587-8.

[26] Mantegna, R.N. and H.E. Stanley (2000), *An Introduction to Econophysics. Correlations and Complexity in Finance*, Cambridge, UK: Campridge University Press.

[27] Merton, R.C. (1980a), Stationarity and Persistence in the GARCH (1,1) Model, *Econometric Theory* ,6, 318-34.

[28] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and E. Teller (1953), Equations of State Calculations by Fast Computing Machines, *Journal of Chemical Physics*, 2, 1087-1091.

[29] Meyn, S. P. and R.L. Tweedie (1993), *Markov Chains and Stochastic Stability,* Springer, New York.

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

[31] Poirier, D. J. (1998), Revising beliefs in nonidentified models , *Econometric Theory,* 14, 483–509.

[32] Poterba, J.M. and L.H. Summers (1988), Mean Reversion in Stock Prices: Evidence and Implications, *Journal of Financial Economics* , 22, 27-59.

[33] Sentana, E. (1995), Quadratic ARCH Models, *The Review of Economic Studies*, 62, 639-661.

[34] Sentana, E. and G. Fiorentini (2001), Identication, estimation and testing of conditionally heteroskedastic factor models, *Journal of Econometrics* ,102, 143-164.

[35] Shephard, N. (1996), Statistical Aspects of ARCH and Stochastic Volatility, in *Time Series Models in Econometrics, Finance and Other Fields*, ed. by D.R. Cox, D.V. Hinkley, and O.E. Barndorff-Nielsen. London:Chapman and Hall. 1-67.

[36] Tanner, M.A. (1996), *Tools for Statisical Inference:Methods for Exploration of Posterior Distributions and Likelihood Functions* (Third ed.). New-York:Springer-Verlag.

[37] Tanner, M.A., and W.H. Wong (1987), The calculation of posterior distributions by data augmentation, *Journal of the American Statistical Association*, 82, 528-540.

[38] Tierney, L. (1994), Markov chains for exploring posterior distribution, *The Annals of Statistics*, 22.4, 1701-1728.

[39] Vrontos, I.D., Dellaportas, P. and D.N. Politis (2000), Full Bayesian Inference for GARCH and EGARCH Models, *Journal of Business and Economic Statistics*, 18,187-198.

[40] Wei, S.X. (2002), A Censored-GARCH Model of Asset Returns with Price Limits, *Journal of Empirical Finance*, 9,197-223.

TABLE 1
Bayesian inference results

Dow Jones	PM	PSD	$\phi_{0.5}$	$\phi_{\min}$	$\phi_{\max}$	IF
δ	0.030	0.018	0.030	0.025	0.039	2.289
φ	0.822	0.068	0.803	0.513	0.999	1.199
φ_u^2	0.010	0.008	0.007	0.003	0.012	2.956
θ	0.287	0.076	0.265	-0.191	0.354	1.565
α	-0.047	0.022	-0.043	-0.107	0.041	2.841
β	0.792	0.033	0.793	0.701	0.872	2.001
γ	-0.099	0.009	-0.098	-0.111	0.015	1.942
FTSE	PM	PSD	$\phi_{0.5}$	$\phi_{\min}$	$\phi_{\max}$	IF
δ	0.034	0.014	0.033	0.016	0.039	1.611
φ	0.812	0.059	0.802	0.508	0.999	3.786
φ_u^2	0.010	0.009	0.010	0.008	0.011	1.979
θ	0.226	0.078	0.205	-0.127	0.348	1.998
α	-0.152	0.045	-0.152	-0.451	0.181	1.546
β	0.649	0.055	0.650	0.549	0.898	2.222
γ	-0.099	0.005	-0.098	-0.107	0.018	1.902
Nikkei	PM	PSD	$\phi_{0.5}$	$\phi_{\min}$	$\phi_{\max}$	IF
δ	0.041	0.010	0.041	0.037	0.046	1.710
φ	0.650	0.061	0.637	0.442	0.999	3.159
φ_u^2	0.009	0.007	0.010	0.008	0.015	1.980
θ	0.288	0.080	0.275	-0.187	0.360	2.001
α	-0.253	0.022	-0.253	-0.464	0.103	2.996
β	0.789	0.019	0.789	0.762	0.880	1.999
γ	-0.099	0.008	-0.099	-0.110	0.010	1.997

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.

FIG. 1 Dow-Jones: Posterior density estimates and sensitivity analysis

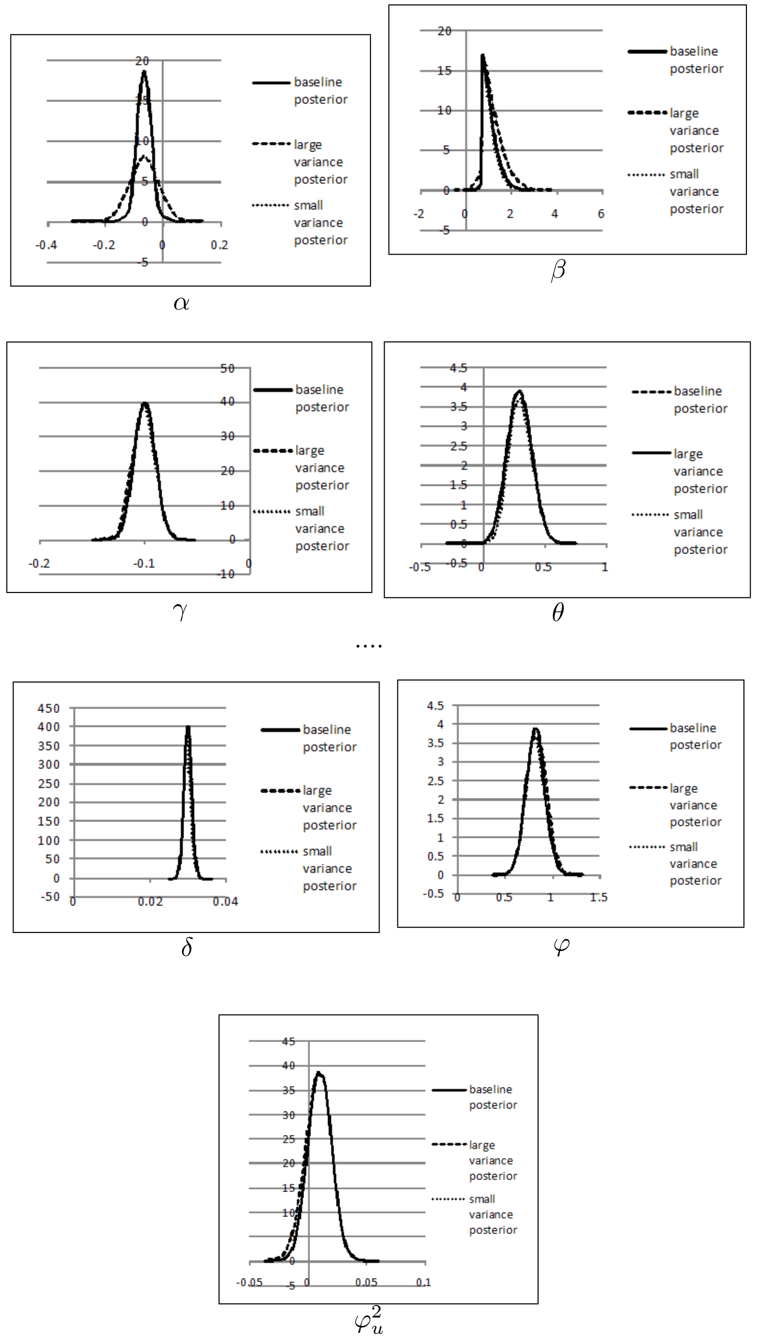

FIG. 2 FTSE: Posterior density estimates and sensitivity analysis

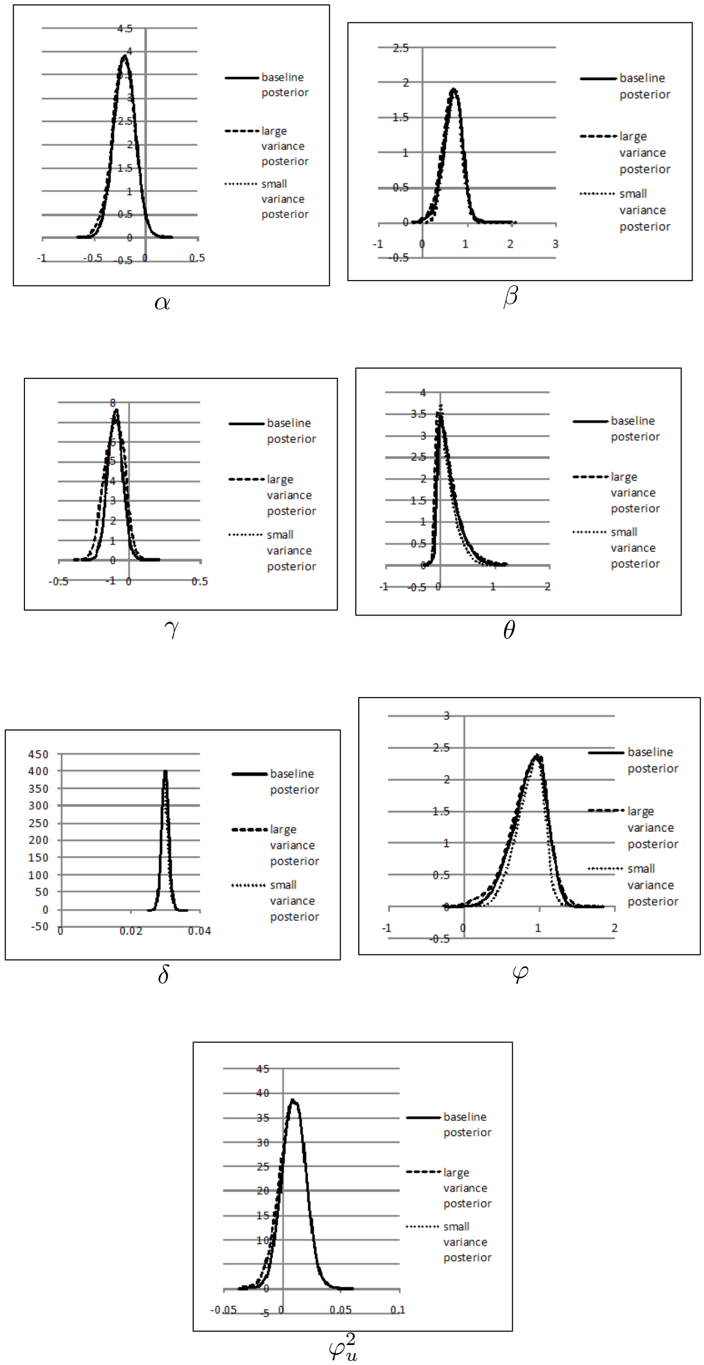

FIG. 3 Nikkei: Posterior density estimates and sensitivity analysis

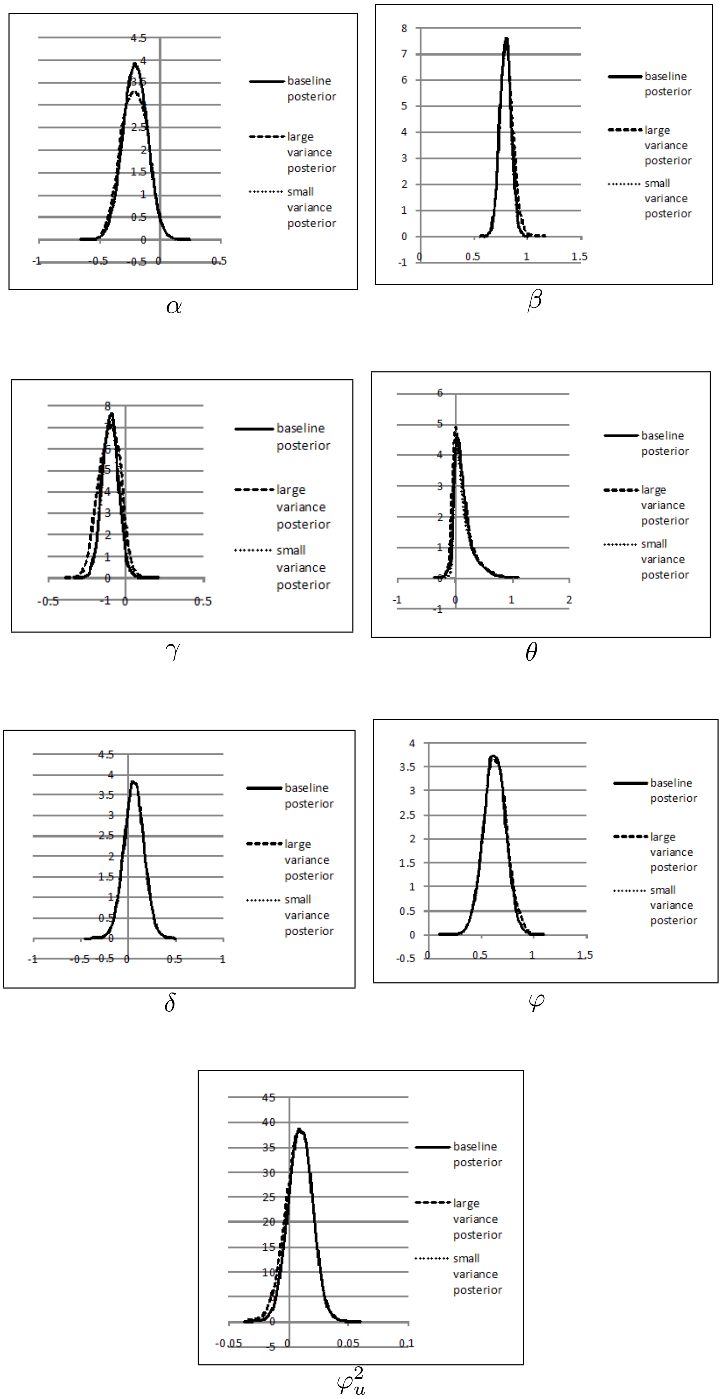

TABLE 2
Sample and Theoretical Moments Comparison

	D-J-Sam.	D-J-The.	FTSE-Sam.	FTSE-The.	Nik.-Sam.	Nik.-The.
Mean	0.0929	0.0740	0.0422	0.0360	0.0380	0.0389
St. Dev.	2.1762	1.6373	2.1231	1.0784	2.5674	0.9822
Skew.	-0.4187	0.0211	-0.8913	0.0067	-0.1221	0.0174
Kurt.	6.2763	3.5278	11.4281	3.2121	5.1262	3.4525
$\rho\left(r_t, r_{t-1}\right)$	-0.0631	0.0586	0.0330	0.0227	-0.0221	0.0056
$\rho\left(r_t, r_{t-2}\right)$	0.0523	0.0475	0.0835	0.0189	0.0447	0.0030
$\rho\left(r_t^2, r_{t-1}\right)$	-0.1262	-0.0800	-0.0932	-0.0814	-0.1227	-0.0808
$\rho\left(r_t^2, r_{t-2}\right)$	-0.0954	-0.0589	-0.1053	-0.0484	-0.0661	-0.0599
$\rho\left(r_t^2, r_{t-3}\right)$	-0.0753	-0.0440	-0.0315	-0.0296	-0.1244	-0.0450
$\rho\left(r_t^2, r_{t-1}^2\right)$	0.2313	0.1741	0.0796	0.1189	0.1766	0.1679
$\rho\left(r_t^2, r_{t-2}^2\right)$	0.0940	0.1289	0.1321	0.0715	0.1050	0.1250
$\rho\left(r_t^2, r_{t-3}^2\right)$	0.0599	0.0971	0.0276	0.0442	0.1817	0.0944
$\rho\left(r_t^2, r_{t-4}^2\right)$	0.0396	0.0740	0.0283	0.0278	0.1183	0.0720

Note:D-J-Sam. stands for the sample moments of Dow Jones and D-J-The. stands for the theoretical moments employing the forulae in section 3 and treating the parameter postirior means as the "true" parameters. Analogously for FTSE and Nikkei.