A Bayesian algorithm for a Markov Switching GARCH model

Dhiman Das
City University of New York

Abstract

Applications of GARCH methods are now quite widespread in macroeconomic and financial time series. New formulations have been developed in order to address the statistical regularity observed in these time series such as assymetric nature and strong persistence of variances. This paper develops a ARMA-GARCH model with Markov switching conditional variances to simulataneously address the above two conditions. A Bayesian algorithm is developed for the estimation purpose and applied to two datasets.

1 Introduction

Analysis of economic time series started receiving special attention with Box and Jenkins' (1976) initial attempt to characterize the regularity in economic data. Due to its initial success in accounting for important economic time series, Box and Jenkins' methods became extremely popular. The basis for such modelling approach was the Wold representation: any covariance stationary time series can be expressed as a moving average function of present and past innovations. This moving average function can always be approximated by a low order autoregressive process, sometimes with some moving average components. However it did not take too long for econometricians to realize the shortcomings of Box and Jenkins' methods, particularly with the observation that many economic time series show considerable non-linearity, which it did not have the tools to handle. Initially, the focus of most macroeconometric and financial time series modeling centered on the conditional first moments, with any temporal dependencies in the higher order moments treated as nuisance. The realization that economic decisions in the data generating process involve considerable non-linear behavior re-oriented the focus of macroeconomic and financial time series studies.

One of the widely successful approaches in this direction is the Autoregressive Conditional Heteroskedastic (ARCH) model, suggested by Engle(1982), and its various developments and extensions such as the Generalized ARCH(GARCH) model of Bollerslev(1986) and the Exponential GARCH (EGARCH) model of Nelson (1991). The key insight offered by the ARCH model is the distinction between the conditional and the unconditional second order moments. While the unconditional covariance matrix for the variables of interest may be time invariant, the conditional variances and covariances often depend non-trivially on the past states of the world. Understanding the exact nature of this temporal dependence is crucial for many issues in macroeconomics and finance, such as irreversible investments, derivative pricing, the term structure of interest rates, and general dynamic asset pricing relationships.

A large amount of theoretical and empirical research has been done on these models during the last two decades and they have provided an improved description of financial markets' volatility. A usual result of ARCH models is the highly persistent behavior of shocks to conditional variance. This persistence, however, is not consistent with the result of recent papers that analyze the volatility after the stock crash of 1987, as Schwert (1990) and Engle and Mustaffa (1992) argue. On the other hand, some suggest a case for an integrated process. Lamoreux and Lastrapes (1990) argue that the near integrated behavior of the conditional variance might be due to the presence of structural breaks, which are not accounted for by standard ARCH models. In the same article, the authors point out that models with switching parameter values, like the Markov switching model of Hamilton (1989), may provide more appropriate modeling of volatility. Hamilton's Markov Switching model can be viewed as an extension of Goldfeld and Quandt's (1973) model of the important case of structural changes in the parameters of an autoregressive process. In his simple two state processes, Hamilton assumes the existence of an unobserved variable, S_t, which describes the state the process is in. He postulates a Markov Chain for the evolution of the unobserved variable given by a pair of transition probabilities.

Apart from Hamilton's original work on business cycles, many papers use Hamilton's model on stock market returns and other financial time series. Schwert (1989) considers a model in which returns may have either a high or a low variance, switches between these return distributions determined by a two state Markov process. Turner, Startz, and Nelson (1989) consider a Markov switching model in which either the mean, the variance, or both may differ between two regimes. Hamilton and Susmel (1993) propose a model with sudden discrete changes in the process which governs volatility. They found that a Markov switching process provides a better statistical fit to the data than ARCH models without switching. Many economic series show evidences of changes in regime. Even if they are rare, during these events the volatility of the series changes substantially. ARCH models focus on the dynamics of the process itself and fails to account for the switching in the dynamics. It underestimates theconditional variance at the time of the change from a normal volatility state to a high volatility state and over estimates the conditional variance when the economy goes back to normal state.

The reason for the renewed vigor in understanding the nature of the variance of the time series process is that in most cases the variance portrays the risk associated with a financial time series. The recent surge of literature in the field of financial instruments emphasizes the variance process for engineering the risk and return associated with any financial asset. To a great extent the early wave of papers on analyzing financial instruments took a considerably simpler view of the variance structure without recognizing the extent to which the subtleties of the non-linear structures (like GARCH, state dependence, threshold models) might affect the actual outcome of the pricing process of a risky asset.

In the next section we discuss the particular aspects of GARCH modelling that this paper intends to address. We also mention the context of developing a Bayesian estimaiton technique for the same. Section 3 elaborates on the formulation of the model. Section 4 elaborates on the Bayesian procedure. Section 5 concludes with an application of the algorithm on a simulated dataset and a dataset of the Bombay Stock Exchange Index.

2 GARCH, and Bayesian Algorithms

Due to rapid acceptance in various economic application there has been a surge of literature trying to improve on the basic GARCH model to characterize the nature of volatility that is observed. In the following we discuss two main trends which are close to the plan of this essay. The first deals with the assumption, implicit in the GARCH model, that the conditional volatility of the asset is affected symmetrically by positive and negative innovations. For various financnial and macroeconomic time series it is unlikely that positive and negative shocks have the same impact on volatility. In finance theory this asymmetry is sometimes ascribed to a *leverage effect* and sometimes to a *risk premium effect*. By the former theory, as the price of stock falls, its debt-equity ratio rises, increasing the volatility of returns to equity holders. By the latter theory, news of increasing volatility reduces the demand for a stock because of risk aversion. The early attempts in this area were by

Sentana's (1995) quadratic ARCH and Bera and Higgins' (1992) non-linear ARCH models.

The second stream of developments concern the failure to account for all the non linearity that are observed in economic and financial time series, particularly those due to non stationarity. In many high-frequency time-series applications, the conditional variance estimated using a $GARCH(p, q)$ process exhibits a strong persistence. This provides an empirical motivation for the Integrated GARCH model. In these types of models, the autoregressive polynomial of the squared error has a unit root and so the shocks are persistent for future forecasts. Lamoureux and Lastrapes (1990) argue that the persistence in GARCH models might be due to mis-specifications of the variance equation. By introducing dummy variables for deterministic shifts in the unconditional variances, they discover that the duration of the volatility shocks is substantially reduced. A similar point is raised by Diebold (1986), who conjectures that the apparent existence of a unit root as in the IGARCH class of models may be the result of shifts in regimes, which affect the level of the unconditional variances. This has led to a review of the discussion of non-linearity in these type of models. One of the most important research in this direction is introducing Hamilton's Markov switching model to account for the specific type of non-linearity in conditional heteroskedasticity model. The switching process is introduced in various ways by various authors. The simplest way to introduce a switching process to the constant term in the conditional variance equation (Cai(1994)). Hamilton and Susmel(1994) consider introducing the Switching parameter to the coefficients of the conditional variance term while Hansen(1994) considers switching the Student t degrees of freedom parameter where the degree of freedom parameter is allowed to vary over time as a probit type function. Authors like Hamilton and Susmel (1994), Bollen, Gray and Whaley (1996), Susmel (1999), Dueker (1997), and others have found encouraging results in equity price and interest rate data.

Bayesian estimation methods for time series processes are by now quite popular. In various macroeconomic and financial time series it has found major success. A detailed discussion of such methods can be found in Tsurumi(00) and refernces there in. In most time series formulations, alternative methods of Maximum Likelihood gives rise to complex likelihood functions either with too many parameters or a computationally intractable form, or both. Besides available maximization algorithms may reach optimality slowly or not at all. On the other hand Bayesian estimation techinques are based on Markov Chain Monte Carlo (MCMC) principle. The main idea behind the Monte Carlo principle is that anything we want to know about a random variable x can be learned by sampling many times from $f(x)$, the probability density function of x. First discussed by Metropolis and Ulam(49), these techniques have gained wide acceptance in recent years with availibility of advanced computational power. The idea behind the Markov Chain principle is the fact that high dimensional joint densities are completely characterized by lower dimensional conditional densities. To learn about $f(x, y)$ the idea is to learn about x conditional on what we know about y, $g(x|y)$ and then to learn about y conditional on what we know about x, $g(y|x)$. Iterate these two steps till we have enough information. This in fact is the Bayesian idea about the Bayesian estimation.

Any model that can be estimated by maximum likelihood method can be estimated using Bayesian methods. But Bayesian simulation lets us work with models and data set thought to be difficult otherwise. Particularly there are major successes in areas like hierarchical models, data set with missing values, item response models, heavy tailed distributions, mixture models, models with dynamics in the latent variable, among others.

3 The Model

In this section, we consider an extension of the ARMA-GARCH model in which the conditional variance term has a regime switching term, S_t which follows a Markov Switching process. A number of papers in the areas of financial macroeconomics discuss why regime switching could be possible. Cecchetti, Lam and Mark(1990) consider a Lucas asset-pricing model in which the economy's endowment switches between high economic growth and low economic growth. They show that such switching in fundamentals accounts for a number of features of stock market returns, such as leptokurtosis and mean reversion. Blanchard and Watson's (1982) model of stochastic bubbles also points in that direction. In each period, a bubble may either survive or collapse; in such a world, returns could be drawn from one of two distributions - surviving bubbles or collapsing bubbles.

There are a number of papers which attempted extending the ideas of Hamilton's Markov Switching model to examine the variance structure of stock market returns.The most important aspect of having a regime switching process in influencing the dynamics is the fact that a movement in either direction (up or down) at any point of time might have different implication depending on the regime the process is in. Thus, a certain increase in volatility in a collapsing bubble will have completely different implication, had it been a surviving bubble. The GARCH models are successful in allowing a change of variance over time, but seem to overlook the special characteristic of such changing variance, depending on regimes, that is so typical of stock market returns and other financial time series. A markov regime switching model provides a more flexible structure than the basic GARCH model. Schwert(1989), Turner, Startz and Nelson (1989) and Hamilton and Susmel (1993), all use certain variations of introducing a Markov Switching element in the mean or the variance structure of the returns.

This paper attempts to develop a Bayesian estimation procedure for such a process. Besides there is one way in which this model differs from the general class of regime switching GARCH models. In most models, the variance structure switched regimes according to the state of volatility. In this model the variance structure switch regimes based on the state of the process and not the variance. This is to highlight the essential differences that variance might have in determining the dynamics of the process, during phases of upward and downward trend in the time series.

Another important aspect of the Bayesian techniques that is used in this model is the fact that it is based on a hybrid model involving both MetropolisHastings and Gibbs technique. The main reason which prompted the introduction of the Gibbs sampling is the estimation of the state variables. The state variables are as long as the data series. At every iteration, under a Metropolis Hastings algorithm it will require that we generate the whole vector of state variables one at a time. The multi-move Gibbs sampling on the other hand hastens the process.

The following describes the basic structure of the model:

$$\begin{aligned} y_t &= \gamma x_t + u_t \\ u_t &= \sum_{j=1}^{p} \phi_j u_{t-j} + \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j}, \qquad \epsilon_t | I_{t-1} \sim N(0, \sigma_t^2) \\ \sigma_t^2 &= \mu_0 + \mu_1 S_t + \sum_{j=1}^{r} \alpha_j \epsilon_{t-j}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2 \end{aligned} \tag{1}$$

where y_t is the dependent variable; x_t is the independent variable;γ is the regression coefficient; σ_t^2 is the conditional variance ofϵ; S_t is the state dummy variable taking integer values in [0,1]; α and β are the coefficient of the GARCH process.

We will closely follow Nakatsuma(2000) who constructs the posterior density functions of the model:

$$p(\delta|Y,X) = \frac{l(Y|X,\delta)p(\delta)}{\int l(Y|X,\delta)p(\delta)d\delta} \tag{2}$$

where δ is the set of all the parameters of the model, $l(Y|X,\delta)$ is the likelihood function, and $p(\delta)$ is the prior.

Likelihood function

Likelihood function for the above model:

$$l(Y|X,\delta) \;=\; \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2}} exp\Big[-\frac{\hat{\epsilon}_t^2}{2\sigma_t^2}\Big] \tag{3}$$

where,

$$\hat{\epsilon}_t = \begin{cases} \epsilon_0 & t=0 \\ y_t - x_t\gamma - \sum_{j=1}^{p}\phi_j(y_{t-j} - x_{t-j}\gamma) - \sum_{j=1}^{q}\theta_j\hat{\epsilon}_{t-j} & t=1,\dots,n \end{cases} \tag{4}$$

and we assume $y_0 = \epsilon_0$, $y_t = 0$ for $t < 0$ and $x_t = 0$ for $t \leq 0$. Pre-sample error ϵ_0 is treated as a parameter.

Prior

We use the following prior:

$$\begin{aligned} p(\epsilon_0,\gamma,\phi,\theta,\mu 0,\mu 1,\alpha,\beta) \;=\; & N(\mu_{\epsilon_0},\Sigma_{\epsilon_0}) \times N(\mu_\gamma,\Sigma_\gamma) \\ & \times N(\mu_\phi,\Sigma_\phi) \times N(\mu_\theta,\Sigma_\theta) \\ & \times N(\mu_\varphi,\Sigma_\alpha) \times N(\mu_\beta,\Sigma_\beta) \end{aligned} \tag{5}$$

where $\varphi = \{\mu_0, \mu_1, \alpha\}$.

4 MCMC procedure

To apply the Monte Carlo method, we need to generate samples $\{\delta^1, \dots \delta^m\}$ from the posterior distribution. Since we cannot generate them directly, we use the MH algorithm. In the MH algorithm, we generate a value$\hat{\delta}$ from the proposal distribution $g(\delta)$ and accept the proposal value with probability:

$$\lambda(\delta,\hat{\delta}) \;=\; min\Big\{1, \frac{p(\hat{\delta}|Y,X)/g(\hat{\delta})}{p(\delta|Y,X)/g(\delta)}\Big\} \tag{6}$$

To construct a MCMC procedure for the model, Nakatsuma divides the parameters into two groups. Let $\delta_1 = (\epsilon_0,\gamma,\phi,\theta)$ be the first group and $\delta_2 = (\mu_0,\mu_1,\alpha,\beta)$ be the second group.For each group of parameters he uses different proposal distributions, which are discussed below.

4.1 Proposal for δ_1

The proposal distribution for the first group δ_1 is based on the original model:

$$y_t = x_t\gamma + \sum_{j=1}^{p}\phi_j(y_{t-j} - x_{t-j}\gamma) + \epsilon_t + \sum_{j=1}^{q}\theta_j\epsilon_{t-j} \qquad \epsilon_t \sim N(0,\sigma_t^2) \tag{7}$$

under the assumption that the conditional variances $\{\sigma_t^2\}_{t=1}^n$ are fixed and known. Using the above equation Nakatsuma generates δ_1 from their proposal distributions by the MCMC procedure by Chib and Greenberg(1994) with some modifications.

4.1.1 ϵ_0 : Pre-sample error

Given the pre-sample error ϵ_0, the model can be rewritten as:

$$y_t = x_t\gamma + \sum_{j=1}^{p}\phi_j(y_{t-j} - x_{t-j}\gamma) + \epsilon_t + \sum_{j=1}^{q}\theta_j\epsilon_{t-j} + (\phi_t + \theta_t)\epsilon_0 \tag{8}$$

This follows closely Chib and Greenberg (1994) estimation of ARMA(p,q) error process. The likelihood function for the model is rewritten as

$$f(Y|X,\Sigma,\delta_1,\delta_2) = \prod_{t=1}^{n}\frac{1}{\sqrt{2\pi\sigma_t^2}}exp\Big[-\frac{(y_t^\dagger - x_t^\dagger\epsilon_0)^2}{2\sigma_t^2}\Big] \tag{9}$$

Where

$$y_t^\dagger = y_t - x_t\gamma - \sum_{j=1}^{p}\phi_j(y_{t-j} - x_{t-j}\gamma) - \sum_{j=1}^{q}\theta_j y_{t-j}^\dagger \tag{10}$$

$$x_t^\dagger = (\phi_t + \theta_t) - \sum_{j=1}^{q}\theta_j x_{t-j}^\dagger \tag{11}$$

$$\epsilon_t = y_t^\dagger - x_t^\dagger\epsilon_0 \tag{12}$$

We have the following proposal distribution of ϵ_0:

$$\epsilon_0|Y,X,\Sigma, \sim N(\hat{\mu}_{\epsilon_0}, \hat{\sigma}_{\epsilon_0}) \tag{13}$$

$$\text{where, } \hat{\mu}_{\epsilon_0} = \hat{\Sigma}_{\epsilon_0}(\sum_{t=1}^{n} x_{\epsilon_t}^2/\sigma_t^2 + \mu_{\epsilon_0}/\sigma_{\epsilon_0}^2) \tag{14}$$

$$\hat{\Sigma}_{\epsilon_0} = (\sum_{t=1}^{n} x_{\epsilon_t}y_{\epsilon_t}/\sigma_t^2 + \sigma_{\epsilon_0}^{-2})^{-1} \tag{15}$$

4.1.2 γ : Regression coefficient

The likelihood function of the model is rewritten as:

$$f(Y|X,\Sigma,\delta_1,\delta_2) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2}} exp\Big[-\frac{(y_t^* - x_t^*\gamma)^2}{2\sigma_t^2}\Big] \tag{16}$$

where,

$$y_t^* = y_t - \sum_{j=1}^{p} \phi_j y_{t-j} - \sum_{j=1}^{p} \theta_j y_{t-j}^* \tag{17}$$

$$x_t^* = x_t - \sum_{j=1}^{p} \phi_j x_{t-j} - \sum_{j=1}^{p} \theta_j x_{t-j}^* \tag{18}$$

Denoting $Y_\gamma = [y_1^*, \ldots, y_n^*]'$ and $X_\gamma = [x_1^*, \ldots, x_n^*]'$, we use the following proposal distribution of γ:

$$\begin{aligned} \gamma|Y,X,\Sigma,\delta_{-\gamma} &\sim N(\hat{\mu}_\gamma, \hat{\Sigma}_\gamma) \\ \text{where, } \hat{\mu}_\gamma &= \hat{\Sigma}_\gamma(X_\gamma'\Sigma^{-1}Y_\gamma + \Sigma_\gamma^{-1}\mu_\gamma) \\ \hat{\Sigma}_\gamma &= (X_\gamma'\Sigma^{-1}X_\gamma + \Sigma_\gamma^{-1})^{-1} \end{aligned} \tag{19}$$

4.1.3 ϕ : AR coefficient

The likelihood function of the model can also be written as:

$$f(Y|X,\Sigma,\delta_1,\delta_2) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2}} exp\Big[-\frac{(\tilde{y}_t - \tilde{x}_t\phi)^2}{2\sigma_t^2}\Big] \tag{20}$$

where,

$$\tilde{y}_t = y_t - x_t\gamma - \sum_{j=1}^{p} \theta_j \tilde{y}_{t-j} \tag{21}$$

$$\tilde{x}_t = [\tilde{y}_{t-1}, \ldots, \tilde{y}_{t-p}] \tag{22}$$

Denoting $Y_\phi = [\tilde{y}_1, \ldots, \tilde{y}_n]'$ and$X_\phi = [\tilde{x}_1', \ldots, \tilde{x}_n']'$, we use the following proposal distribution of ϕ:

$$\begin{aligned} \phi|Y,X,\Sigma,\delta_{-\phi} &\sim N(\hat{\mu}_\phi, \hat{\Sigma}_\phi) \\ \text{where, } \hat{\mu}_\phi &= \hat{\Sigma}_\phi(X_\phi'\Sigma^{-1}Y_\phi + \Sigma_\phi^{-1}\mu_\phi) \\ \hat{\Sigma}_\phi &= (X_\phi'\Sigma^{-1}X_\phi + \Sigma_\phi^{-1})^{-1} \end{aligned} \tag{23}$$

4.1.4 θ : MA coefficient

One important innovation in Nakatsuma's algorithm is the estimation of the moving average term which in this case is a non-linear function of the error. Chib and Greenberg (1994) proposed in their estimation of a simple moving average term to linearize ϵ_t by the first-order Taylor series expansion

$$\epsilon_t(\theta) \approx \epsilon_t(\theta^*) + \psi_t(\theta - \theta^*) \tag{24}$$

where $\epsilon_t(\theta^*) = y_t^*(\theta^*) - x_t^*(\theta^*)$ and$\psi = [\psi_{1t}, \ldots, \psi_{qt}]$ is the first order derivative of$\epsilon_t(\theta)$ evaluated at θ^* given by the following recursion:

$$\psi_{it} = -\epsilon_{t-i}(\theta^*) - \sum_{j=1}^{q} \theta_j^* \psi_{it-j} \qquad (i = 1, \ldots, q) \tag{25}$$

The choice of θ^* is crucial to obtain a suitable approximation. Chib and Greenberg (1994) chose the non-linear least squares estimate of θ given the other parameters of θ^*

$$\theta^* = argmin \sum_{t=1}^{n} \{\epsilon_t(\theta)\}^2 \tag{26}$$

However, the error term in the ARMA-GARCH model is heteroskedastic while Chib and Greenberg applied their approximation to the homoskedastic error ARMA model.Thus instead of the above, Nakatsuma proposes the following weighted non-linear least square estimate of θ

$$\theta^* = argmin \sum_{t=1}^{n} \{\epsilon_t(\theta)\}^2 / \sigma_t^2 \tag{27}$$

to approximate the likelihood function of the model. The approximated likelihood function of the model can also be written as:

$$f(Y|X, \Sigma, \delta_1, \delta_2) \;=\; \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2}} exp\Big[- \frac{\{\epsilon_t(\theta^*) + \psi_t(\theta - \theta^*)^2\}^2}{2\sigma_t^2} \Big] \tag{28}$$

Denoting $Y_\theta = [\psi_1\theta^* - \epsilon_1(\theta^*), \ldots, \psi_n\theta^* - \epsilon_n(\theta^*)]'$ and $X_\theta = [\psi_1', \ldots, \psi_n']'$, using the approximated distribution we use the following proposal distribution of θ:

$$\begin{aligned} \theta|Y, X, \Sigma, \delta_{-\theta} &\sim N(\hat{\mu}_\theta, \hat{\Sigma}_\theta) \\ \text{where, } \hat{\mu}_\theta &= \hat{\Sigma}_\theta(X_\theta' \Sigma^{-1} Y_\theta + \Sigma_\theta^{-1} \mu_\theta) \\ \hat{\Sigma}_\theta &= (X_\theta' \Sigma^{-1} X_\theta + \Sigma_\theta^{-1})^{-1} \end{aligned} \tag{29}$$

4.2 Proposal for δ_2

The proposal distribution for the second group δ_2 are based on an approximated GARCH model:

$$\begin{aligned} \epsilon_t^2 &= \mu_0 + \mu_1 S_t + \sum_{j=1}^{l} (\alpha_j + \beta_j)\epsilon_{t-j}^2 + w_t - \sum_{j=1}^{s} \beta_j w_{t-j} + \alpha_t w_0 \\ w_t &\sim N(0, 2\sigma_t^4) \end{aligned} \tag{30}$$

Given $\{\epsilon_t^2\}_{t=1}^n$ and conditional variances $\{\sigma_t^2\}_{t=1}^n$ it is possible to generate δ_2 from their proposal distributions by the MCMC procedure similar to Chib and Greenberg's method. However, before obtaining the ARCH and the GARCH coefficients, the next important step would be to obtain the state variables in the variance expression.

The Markov Switching Process

At this point, we need to discuss the Bayesian estimation procedure for the Markov Switching states. Let S_t be a stochastic variable taking the values 0 or 1. Furthermore, we assume S_t evolves according to a Markov Chain, independent of past observations of y_t i.e.,

$$P[S_t = j | S_{t-1} = i, S_{t-2} = i', ..., y_{t-1}] = P[S_t = j | S_{t-1} = i] = p_{ij} \tag{31}$$

The p_{ij}'s will be elements of a (2×2) transition matrix of states P which collects probability of going to state j at time t given being at state i at time $t-1$.

$$P = \begin{bmatrix} p_{00} & p_{10} \\ p_{01} & p_{11} \end{bmatrix} \tag{32}$$

The columns should sum to unity. So that we can write P as

$$P = \begin{bmatrix} q & 1-p \\ 1-q & p \end{bmatrix} \tag{33}$$

Assuming we have complete information about the vector of parameters $\{\mu_0, \mu_1\}$ and the transition probabilities p_{ij}, we collect them into a setθ. If we have a sequence of observation on y_t for $t \in \{0, T\}$ we now want to calculate the likelihood of these observations as a function of the parameter vector θ and the sequence of the state probabilities $P[S_t = 0 | y_t; \theta, P(S_{t-1} = 1)]$ for

$t = 0, \ldots T$. For this we first denote $P(S_t = 0|y_t, \ldots, y_0), P(S_t = 1|y_t, \ldots, y_0)$ by $P(S_t|y_t)$ and let $P(S_{t+1}, S_t|y_t)$ denote the (4×1) vector

$$\begin{bmatrix} P(S_{t+1} = 0, S_t = 0|y_t, \ldots, y_0) \\ P(S_{t+1} = 1, S_t = 0|y_t, \ldots, y_0) \\ P(S_{t+1} = 0, S_t = 1|y_t, \ldots, y_0) \\ P(S_{t+1} = 1, S_t = 1|y_t, \ldots, y_0) \end{bmatrix} \tag{34}$$

For this purpose, we construct a filter that takes as *input*

$$P(S_{t-1}|y_{t-1}) \tag{35}$$

and produces as *output*

$$P(S_t|y_t) \text{ and } f(y_t|y_{t-1}, \ldots, y_0) \tag{36}$$

We construct this filter in five steps. The technique follows closely Hamilton's(1994) description.

4.2.1 Step 1 : Calculate $P(S_t, S_{t-1}|y_{t-1})$

This is done by using the rule $P(A \cap B) = P(B|A)P(A)$. So in this case

$$P(S_t = i, S_{t-1} = j|y_{t-1}) = P(S_t = i|S_{t-1} = j, y_{t-1})P(S_{t-1} = j|y_{t-1}) \tag{37}$$

Element by element multiplication of $P(S_t, S_{t-1}|y_{t-1}) = P. * P(S_{t-1}|y_{t-1})$ produces a (2×2) matrix.

$$\begin{bmatrix} P(S_t = 0, S_{t-1} = 0) & P(S_t = 1, S_{t-1} = 0) \\ P(S_t = 0, S_{t-1} = 1) & P(S_t = 1, S_{t-1} = 1) \end{bmatrix} \tag{38}$$

4.2.2 Step 2: Joint density

We now calculate the joint density $f(y_t, S_t, S_{t-1})$ for the four possible combination of S_t and S_{t-1}

$$\begin{aligned} f(y_t, S_t = i, S_{t-1} = j|y_{t-1}) &= f(y_t|S_t = i, S_{t-1} = j, y_{t-1}) \\ &\quad \times P(S_t = i, S_{t-1} = j|y_{t-1}) \end{aligned} \tag{39}$$

To obtain that, we create a (2×2) matrix of the expected values ofy_t forthe state sequences.

$$M_t = \begin{bmatrix} \mu_0 & \mu_0 + \mu_1 \\ \mu_0 & \mu_0 + \mu_1 \end{bmatrix} \tag{40}$$

Then

$$\frac{1}{\sqrt{2\pi\sigma^2}} exp[-\frac{1}{2\sigma^2}(M_t - y_t)^2] \tag{41}$$

produces the (2×2) matrix of $f(y_t|S_t, S_{t-1}, y_{t-1})$

4.2.3 Step 3 : Calculating density $f(y_t|y_{t-1})$

This is obtained be summing all elements of $f(y_t, S_t = i, S_{t-1} = j|y_{t-1})$

4.2.4 Step 4 : Calculate $P(S_t, S_{t-1}|y_t)$

For this we use the Bayes law

$$P(S_t = i, S_{t-1} = j|y_t) = \frac{f(y_t, S_t = i, S_{t-1} = j|y_{t-1})}{f(y_t|y_{t-1})} \quad (42)$$

This is obtained by dividing the matrix from Step 2 by the scalar from Step 3.

4.2.5 Step 5 : Output : $P(S_t|y_t)$

This is obtained be summing $P(S_t, S_{t-1}|y_t)$ over all S_{t-1}. By repeating this from $t = 1$ to T we get a sequence of state probabilities and a sequence of likelihoods $f(y_t|y_{t-1}, \ldots, y_0)$.

4.2.6 Smoothed Probabilities

Above we have calculated the probabilities $P(S_t|y_t)$. However we may be interested to find out $P(S_t|y_T)$ which is supposed to give better inference about the states. To obtain the "smoothed" state probabilities Hamilton uses the following lemma

$$P(S_t = j|S_{t+1} = i, y_T) = P(S_t = j|S_{t+1} = i, y_t) \quad (43)$$

which follows from $P(S_{t+1} = j|S_t = i, y_T) = P(S_{t+1} = j|S_t = i, y_t)$ which follows form the assumption of Markov Chain. He also uses another lemma

$$\begin{aligned} P(S_t = j|S_{t+1} = i, y_t) &= \frac{P(S_t = j, S_{t+1} = i|y_t)}{P(S_{t+1} = i|y_t)} \\ &= \frac{P(S_t = j|y_t)P(S_{t+1} = i|S_t = j)}{P(S_{t+1} = i|y_t)} \\ &= \frac{P(S_t = j|y_t)p_{ji}}{P(S_{t+1} = i|y_t)} \end{aligned} \quad (44)$$

We start from last-but-one period and use the first lemma:

$$P(S_{T-1} = j|S_T = i, y_T) = P(S_{T-1} = j|S_T = i, y_{T-1})P(S_T = i|y_T) \quad (45)$$

Then we use the second lemma

$$P(S_{T-1}=j,S_T=i|y_T) \quad = \quad \frac{P(S_{T-1}=j|y_{T-1})p_{ji}}{P(S_T=i|y_{T-1})}P(S_T=i|y_T) \quad (46)$$

The right hand side of the above equation was calculated in Step 5 above. Then to get the smoothed probability for $T-1$ we just sum over all the S_Ts.

$$P(S_{t-1}=j|y_T)=\sum_i P(S_{T-1}=j,S_T=i|y_T) \quad (47)$$

Doing this for all states gives a vector of smoothed state probabilities for$T-1$. The same can be repeated for $T-2$ to period 0.The probability distribution of $S_t|y_t$ along with the transition probabilities are used to draw the state in period t, given the state in earlier period, using a binomial distribution.The multi-move sampling technique requires the whole sequence of state variable is generated at the same time to be used in the generation of paremeters in δ_2.

4.2.7 Transition Probabilities

For generating transition probabilities, p and q, conditional on $\tilde{S_T}$ we assume independent beta distributions for the priors of p and q, we have:
Prior

$$\begin{aligned} p &\sim Beta(u_{11},u_{10}) \\ q &\sim Beta(u_{00},u_{01}) \end{aligned}$$

where,

$$g(p,q)\propto p^{u_{11}-1}(1-p)^{u_{10}-1}q^{u_{00}-1}(1-q)^{u_{01}-1} \quad (48)$$

where $u_{ij};i,j=0,1$, are the hyper-parameters of the priors.The likelihood function for p and q is given by:

$$f(p,q|\tilde{S_T})=p^{n_{11}}(1-p)^{n_{10}}q^{n_{00}}(1-q)^{n_{01}} \quad (49)$$

where n_{ij} refers to the number of transitions from state i to state j, which can be easily counted for given $\tilde{S_T}$. Combining the prior distribution and the likelihood function, we get the following posterior distribution:

$$\begin{aligned} p(p,q|\tilde{S_T}) &= g(p,q)f(p,q|\tilde{S_T}) \\ &\propto p^{u_{11}-1}(1-p)^{u_{10}-1}q^{u_{00}-1}(1-q)^{u_{01}-1}p^{n_{11}}(1-p)^{n_{10}}q^{n_{00}}(1-q)^{n_{01}} \\ &= p^{u_{11}+n_{11}-1}(1-p)^{u_{10}+n_{10}-1}q^{u_{00}+n_{00}-1}(1-q)^{u_{01}+n_{01}-1} \end{aligned}$$

4.3 α : ARCH coefficient

The likelihood function of the approximated GARCH model is rewritten as

$$f(\epsilon_t^2|Y,X,\Sigma,\delta_1,\delta_2) = \prod_{t=1}^{n}\frac{1}{\sqrt{2\pi(2\sigma_t^4)}}exp\Big[-\frac{(\bar{\epsilon}_t^2-\zeta_t\alpha)^2}{2(2\sigma_t^4)}\Big] \tag{50}$$

where $\epsilon^2 = [\epsilon_1^2,\dots,\epsilon_n^2]'$, $\bar{\epsilon}_t^2 = \tilde{\epsilon}_t^2 - \sum_{j=1}^{s}\beta_j\tilde{\epsilon}_{t-j}^2$ and$\zeta_t = [\tilde{\iota}_t, \tilde{\epsilon}_{t-1}^2,\dots,\tilde{\epsilon}_{t-r}]$. Let $Y_\alpha = [\bar{\epsilon}_1^2,\dots,\bar{\epsilon}_n^2]'$ and$X_\alpha = [\zeta_1',\dots,\zeta_n']'$. We have the following proposaldistribution of α:

$$\begin{aligned}\alpha|Y,X,\Sigma,\delta_{-\alpha} &\sim N(\hat{\mu}_\alpha,\hat{\Sigma}_\alpha) \\ \text{where, } \hat{\mu}_\alpha &= \hat{\Sigma}_\alpha(X_\alpha'\Lambda^{-1}Y_\alpha+\Sigma_\alpha^{-1}\mu_\alpha) \\ \hat{\Sigma}_\alpha &= (X_\alpha'\Lambda^{-1}X_\alpha+\Sigma_\alpha^{-1})^{-1} \\ \Lambda &= diag\{2\sigma_1^4,\dots,2\sigma_n^4\}\end{aligned} \tag{51}$$

4.4 β : GARCH coefficient

The likelihood function of the approximated GARCH model is rewritten as

$$f(\epsilon^2|Y,X,\Sigma,\delta_1,\delta_2) = \prod_{t=1}^{n}\frac{1}{\sqrt{2\pi(2\sigma_t^4)}}exp\Big[-\frac{\{w_t(\beta^*)-\xi_t(\beta-\beta^*)\}^2}{2(2\sigma_t^4)}\Big] \tag{52}$$

where $\xi_t = [\xi_{1t},\xi_{2t}]$ is the first-order derivative of $\sigma_t^2(\beta)$ evaluated at β^*. Let $Y_\beta = [w_1(\beta^*)+\xi_1\beta^*,\dots,w_n(\beta^*)+\xi_n\beta^*]'$and $X_\beta = [\xi_1',\dots,\xi_n']'$. Using the approximated likelihood function, we have the following proposal distribution of β:

$$\begin{aligned}\beta|Y,X,\Sigma,\delta_\beta &\sim N(\hat{\mu}_\beta,\hat{\Sigma}_\beta) \\ \text{where, } \hat{\mu}_\beta &= \hat{\Sigma}_\beta(X_\beta'\Lambda^{-1}Y_\beta+\Sigma_\beta^{-1}\mu_\beta) \\ \hat{\Sigma}_\beta &= (X_\beta'\Lambda^{-1}X_\beta+\Sigma_\beta^{-1})^{-1} \\ \Lambda &= diag\{2\sigma_1^4,\dots,2\sigma_n^4\}\end{aligned} \tag{53}$$

5 Results and Conclusions

We apply the algorithms for a ARMA(1,1) Markov Switching GARCH model on simulated data and also on a market data set. The market data that is being considered is one of the most important market index in Indian Stock market, the Bombay Stock Exchange (BSE) SENSEX, which is composed of a basket of 30 Indian stocks. The above procedure was implemented on a simulated Switching GARCH data and the results are as follows:

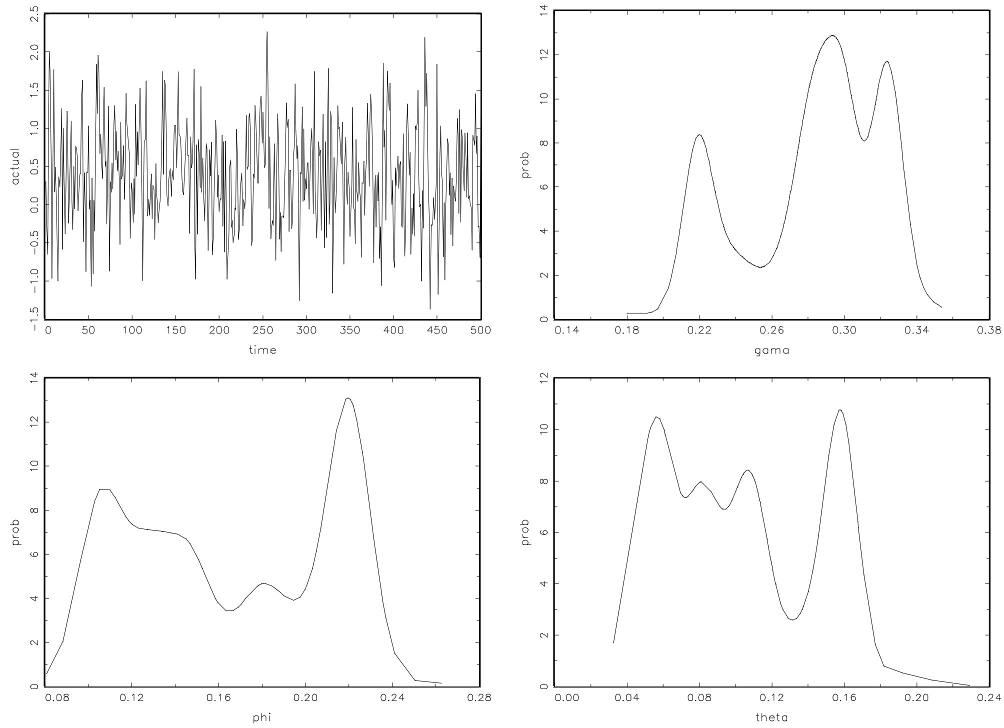

Figure 1: Kernel Density for parameters; simulated data

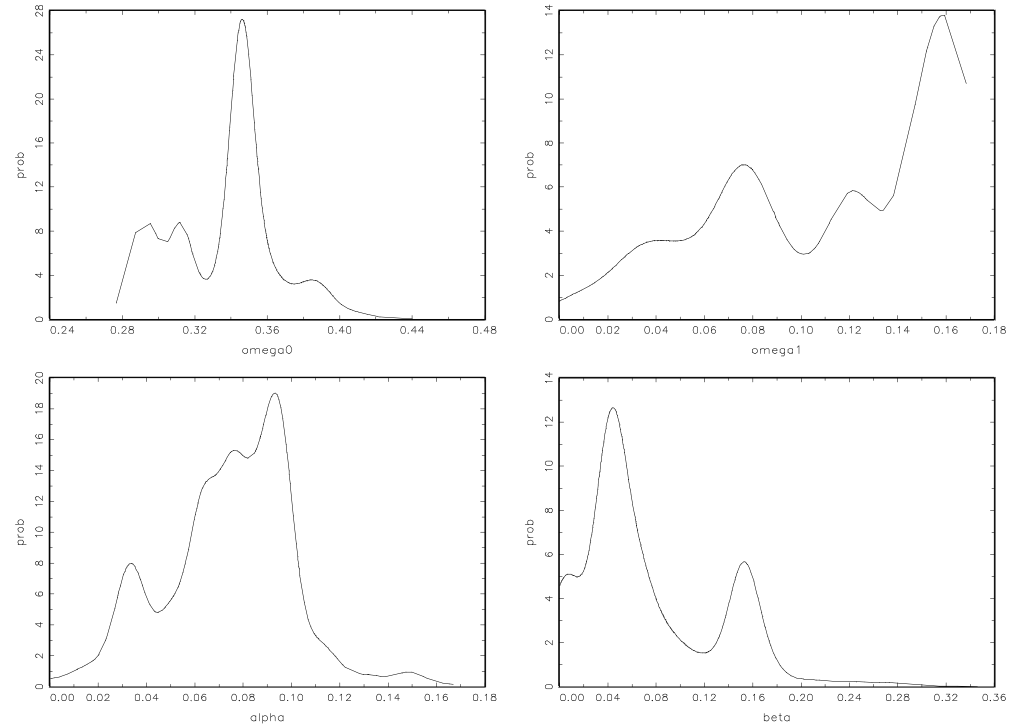

Figure 2: Kernel Density for parameters; simulated data *contd*.

Parameters	**Original values**	**Estimated values**	**Standard deviation**
γ	0.3000	0.2820	0.0383
ϕ	0.1000	0.1657	0.0462
θ	0.1000	0.1032	0.0419
ω_0	0.3000	0.3360	0.0289
ω_1	0.1000	0.1141	0.0481
α	0.1000	0.0742	0.0267
β	0.1000	0.0725	0.0567

The model seems to have satisfactory performance with respect to the original values.The same algorithm is then applied on the BSE sensex data and the results are asfollows:

Parameters	**Estimated values**	**Standard deviation**
γ	-0.1081	0.0696
ϕ	0.5337	0.0439
θ	-0.4286	0.0563
ω_0	1.0398	0.1841
ω_1	0.5640	0.1793
α	0.1516	0.0210
β	0.6118	0.0750

The statistical significance of the switching parameter makes the point regarding the applicability of a Markov switching GARCH model on the market data.

Conclusion

Time series processes study systems, whose dynamics are strongly dependent on itself or past values of related variables. Considerable information can however be also gained from the second moments. The GARCH model advanced the understanding of time series process by incorporating the second moments. A Markov Switching GARCH process accomodates further the aspect of time series dynamics being affected by external unpredictable processes. Incorporating a Markov Switching process for the conditional variance of the underlying asset prices will only enhance predictive power. Literature on derivative pricing generally does not consider these aspects of the time dynamics, which are so common in observed data. Previously the

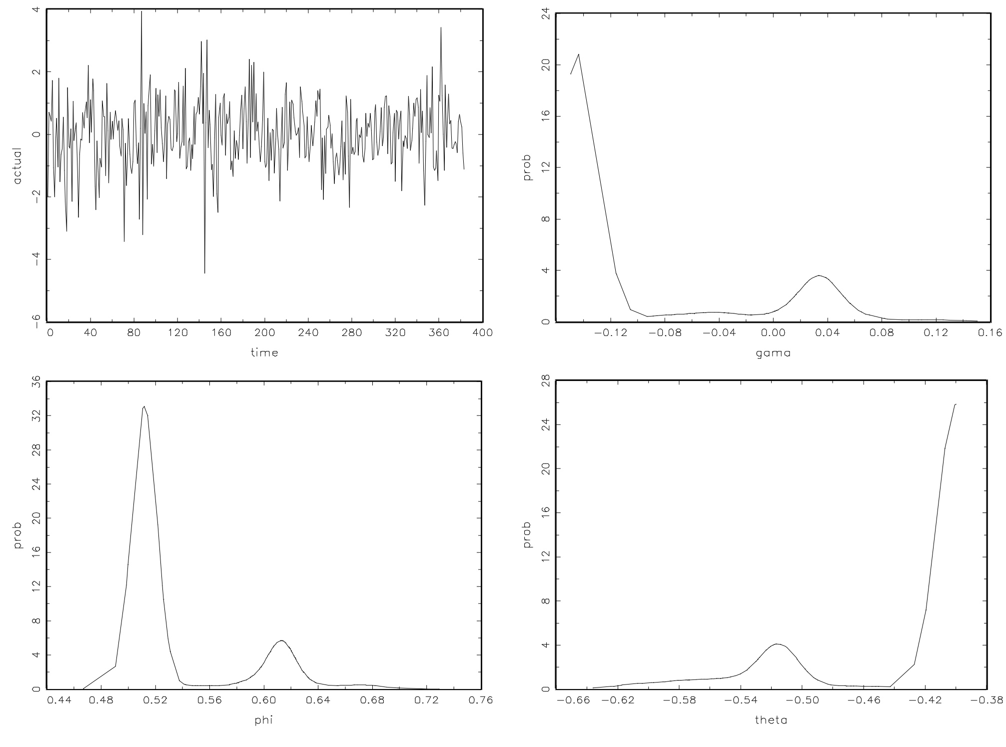

Figure 3: Kernel Density for parameters; BSE Sensex data

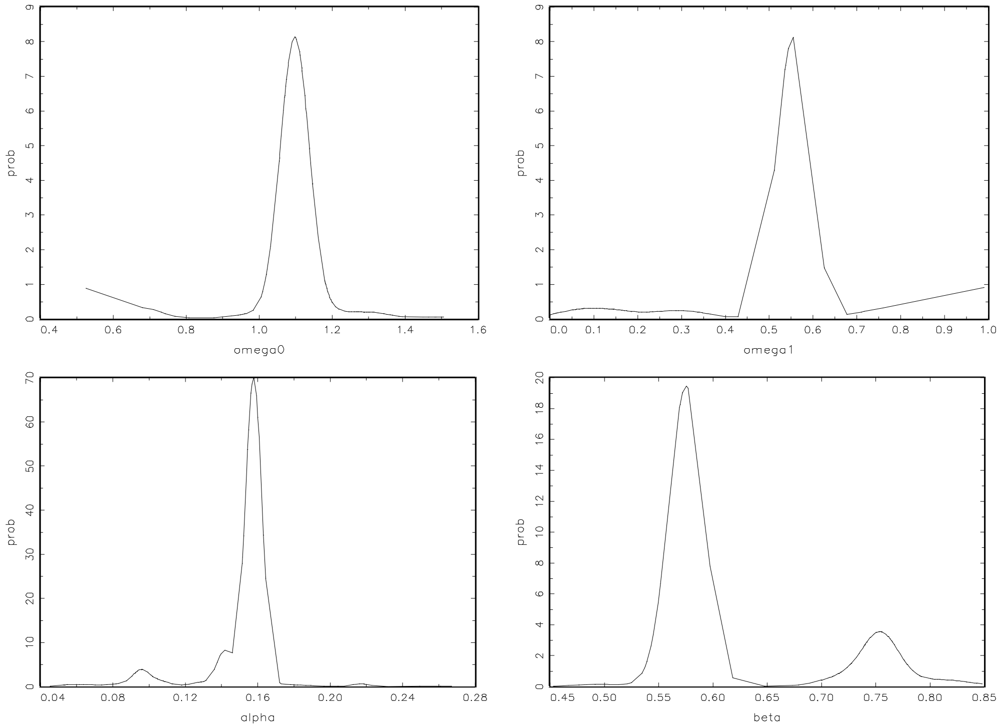

Figure 4: Kernel Density for parameters; BSE Sensex data *contd*.

variances were plugged into the Black Scholes option pricing formula (Engle, Hong, Kane and Noh(1993)). Recently Duan (89) developed a GARCH option-pricing model, where he introduced the concept of local risk neutral valuation with respect to an underlying GARCH asset pricing process. Using local risk neutral valuation, the Markov Switching GARCH can also be applied in derivative pricing. Further research in this area requires the clarification of certain aspects of the Bayesian estimation technique. We need to look into the convergence of the MCMC draws using plots and test statistic such as the fluctuation test discussed in Goldman, Valieva and Tsurumi (2003). We should also look at identification and the applicability of the filter that generates the state probabilities for a more precise algorithm for estimating such processes as discussed in Kaufman and Fruhwirth-Schnatter (2002). In addition, the performance of this technique in comparison to other specifications of asymmetric GARCH model should be investigated using Bayes factor and predictive densities.

References

[1] J. Albert and S. Chib. A Markov model of unconditional variance in ARCH. *Journal of Econometrics*, 58, 1993.

[2] R.T. Baillie and T. Bollerslev. Prediction in Dynamic Models with Time-Dependent Conditional Variances. *Journal of Econometrics*, 52, 1992.

[3] R. Barnett, G. Kohn and S. Sheather. Bayesian estimation of an autoregressive model using Markov Chain Monte Carlo. *Journal of Econometrics*, 74, 1996.

[4] L. Bauwens and M. Lubrano. Bayesian inference on GARCH models using Gibbs sampler. *Econometrics Journal*, 1, 1998.

[5] A.K. Bera and H.L. Higgins. A Survey of ARCH Models: Properties, Estimation and Testing. *Journal of Economic Surveys*, 4, 1993.

[6] O.J. Blanchard and M.W. Watson. *Bubbles, rational expectations and financial markets.* Lexington Books, 1982.

[7] S. F. Bollen, N. P. B. Gray and Whaley R. E. Regime switching in foreign exchange rates: Evidence from currency option prices. *Journal of Econometrics*, 94, 2000.

[8] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 1986.

[9] G. Box and G. Jenkins. *Time Series Analysis: Forecasting and Control.* Holden-Day, 1976.

[10] J. Cai. Bayes Regression and Autocorrelated Errors: A Gibbs Sample Approach. *Journal of Business and Economic Statistics*, 12, 1994.

[11] B. P. Carlin and A.E. Gelfand. Approaches for Empirical Bayes Confidence Intervals. *Journal of the American Statistical Association*, 85, 1990.

[12] N. G. Carlin, B.P. Polson and D.S. Stoffer. A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modeling. *Journal of the American Statistical Association*, 87, 1992.

[13] C. Carter and R. Kohn. On Gibbs Sampling for State Space Models. *Biometrika*, (81), 1994.

[14] P. Cecchetti, S. G. Lam and M. Nelson. Mean Reversion in Equilibrium Asset Prices. *American Economic Review*, 80, 1990.

[15] S. Chib and E. Greenberg. Understanding the Metropolis Hastings Algorithm. *The American Statistician*, 49, 1995.

[16] F. X. Diebold. Modeling the Persistence of Conditional Variances: A Comment. *Econometric Reviews*, 5, 1986.

[17] J. C. Duan. The GARCH option Pricing model. *Mathematical Finance*, 1995.

[18] M. Dueker. Markov Switching in GARCH Processes and Mean- Reverting Stock Market Volatility. *Journal of Business and Economic Statistics*, 15, 1997.

[19] R. Engle and C. Mustafa. Implied ARCH Models from Option Prices. *Journal of Econometrics*, 52, 1992.

[20] R. F. Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50, 1982.

[21] E. F. Fama. The Behaviour of Stock Market Prices. *Journal of Business*, 38, 1965.

[22] J. Geweke. Bayesian Inference in Econometric Models Using Monte Carlo Integration. *Econometrica*, 57, 1989.

[23] R. Glosten, L.R. Jagannathan and D.E. Runkle. On the Relation between Expected Value and the Volatility of the Nominal Excess Return on Stocks. *The Journal of Finance*, 48, 1993.

[24] S. Goldfeld and R. Quandt. A Markov Model for Switching Regression. *Journal of Econometrics*, 1, 1973.

[25] E. Goldman, E. Valieva and H. Tsurumi. Convergence Tests for Markov Chain Monte Carlo Draws. Technical report, Rutgers University, 2003.

[26] J Hamilton. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57, 1989.

[27] J. Hamilton. *Time Series Analysis.* Princeton University Press, 1994.

[28] J. Hamilton and R. Susmel. Autoregressive Conditional Heteroskedasticity and Changes in Regime. *Journal of Econometrics*, 64, 1994.

[29] B. E. Hansen. Autoregressive Conditional Density Estimation. *International Economic Review*, 35, 1994.

[30] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, (57), 1970.

[31] S. L. Heston and S. Nandi. A Closed-Form GARCH Option Pricing Model. Working Papers, 1997. Federal Reserve Bank of Atlanta.

[32] S. Kaufmann and S. Fruhwirth-Schnatter. Bayesian Analysis of Switching ARCH Models. *Journal of Time Series Analysis*, 23, 2002.

[33] C. Kim and C. Nelson. *State Space Models with Regime Switching: Classical and Gibbs Sampling Approaches with Aplications.* MIT Press, Cambridge, 1999.

[34] F. Kleibergen and H.K. van Dijk. Nonstationarity in GARCH Models: A Bayesian Analysis. *Journal of Applied Econometrics*, 8, 1993.

[35] C. Lamoreux and W. Lastrapes. Persistence in Variance, Structural Change and the GARCH Model. *Journal of Business and Economic Statistics*, 5, 1990.

[36] B.B. Mandelbrot. New methods in statistical economics. *Journal of Political Economy*, 71, 1963.

[37] M.N. Metropolis N.; Rosenbluth, A.W.; Rosenbluth and Teller A.H. Equation of State Calculations by Fast Computer Machines. *Journal of Chemical Physics*, 21, 1953.

[38] P. Muller and A. Pole. Monte Carlo Posterior Integration in GARCH models. *Sankhya*, 6, 1998.

[39] T. Nakatsuma. A Markov-Chain Sampling Algorithm for GARCH Models. *Studies in Nonlinear Dynamics and Econometrics*, 3, 1998.

[40] T. Nakatsuma. Bayesian analyisis of ARMA-GARCH models: A Markov chain sampling approach. *Journal of Econometrics*, 95, 2000.

[41] D. B. Nelson. Conditional heteroskedasticity in asset pricing: a new approach. *Econometrica*, 59, 1991.

[42] R. Rabemananjara and J. M. Zakoian. Threshold ARCH Models and Asymmetries in Volatility. *Journal of Applied Econometrics*, 1, 1993.

[43] C. Ritter and M. Tanner. Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler. *Journal of American Statistical Association*, 87(419), 1992.

[44] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, 1999.

[45] G. W. Schwert. Why Does Stock Market Volatility Change Over Time? *Journal of Finance*, 44, 1989.

[46] G. W. Schwert. Stock volatility and the crash of '87. *Review of Finanical Studies*, 3, 1990.

[47] E. Sentana. Quadratic ARCH Models. *Review of Economic Studies*, 62, 1995.

[48] R. Susmel. Switching Volatility in International Equity Markets. *Journal of Econometrics*, 1992.

[49] H. Tsurumi. Bayesian Statistical Computations of Nonlinear Financial Time Series Models: A Survey with Illustrations. *Asia-Pacific Financial Markets*, 7, 2000.

[50] R. Turner, C. Startz and C. Nelson. A Markov model of heteroscedasticity, risk, and learning in the stock market. *Journal of Financial Economics*, 25, 1989.

[51] S. vanNorden and H. Schaller. Regime Switching in Stock Market Returns. Technical report, Bank of Canada, 1993. Working Papers.

[52] G. Wei and M. Tanner. A Monte Carlo implementation of the EM algorithm and the poor mans data augmentation algorithms. *Journal of the American Statistical Association*, 85, 1990.