A Bayesian MCMC Algorithm for Markov Switching GARCH models

Dhiman Das
City University of New York
B.Hark Yoo
Rutgers University

May 2004

Abstract

Markov switching GARCH models have been developed in order to address the statistical regularity observed in financial time series such as strong persistence of conditional variance. However, Maximum Likelihood Estimation faces a implementation problem since the conditional variance depends on all the past history of state. This paper shows that this problem can be handled easily in Bayesian inference. A new Markov Chain Monte Carlo algorithm is introduced and proves to work well in a numerical example.

1 Introduction

Since the Autoregressive Conditional Heteroskedastic (ARCH) model was suggested by Engle(1982), a large amount of theoretical and empirical research has been done during the last two decades and they have provided an improved description of financial markets' volatility. 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.

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 GARCH 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 the conditional variance at the time of the change from a normal volatility state to a high volatility state and overestimates the conditional variance when the economy goes back to normal state.

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.

When we estimate Markov switching GARCH models by Maximum Likelihood Estimation, we confront an implementation problem. Because of the structure of GARCH, the conditional variance depends on all the past history of the state variable. This means that if we have K-state and T-sample size, we need to consider K^T cases to get likelihood function. It is practically impossible to implement. Hamilton and Susmel(1994) and Cai(1994) use Markov switching ARCH models to avoid this problem. Gray(1996) and Dueker (1997) estimate Markov switching GARCH models by approximating likelihood function which depends on only a few of the state variables.

In this paper, we show that the problem can be easily dealt with in Bayesian context. In Bayesian inference, we treat the state variables as random variables and construct the likelihood function assuming we know the states. This structure makes construction of the likelihood function easy. We construct posterior distribution using priors and the likelihood function and integrate the posterior density function with respect to parameters and state variables. For the integration, we devise a new Markov Chain Monte Carlo (MCMC) algorithm and show the result of a numerical example of the algorithm.

In Section 2, we addresse Markov switching GARCH models. In Section 3, we show the general idea about Bayesian inference. In Section 4, our MCMC algorithm is explained in detail. A numerical example is presented in Section 5. Section 6 concludes.

2 The Model

Markov switching GARCH(r,s) model is

$$\begin{aligned}
y_t &= x_t'\gamma + \epsilon_t, \qquad \epsilon_t = \sigma_t\omega_t, \qquad \omega_t \sim N(0,1) \\
\sigma_t^2 &= \mu_0 + \mu_1 S_t + \sum_{j=1}^{r}\alpha_j\epsilon_{t-j}^2 + \sum_{j=1}^{s}\beta_j\sigma_{t-j}^2
\end{aligned}$$

where y_t is dependent variable; x_t is vector of independent variables; γ is regression coefficient; α and β are coefficients of the GARCH process; S_t is state variable taking 0 or 1. The state variable, S_t, evolves according to a two state, first order Markov Switching process with the following hidden transition probabilities:

$$Pr[S_t = 0|S_{t-1} = 0] = p_{00}, \quad Pr[S_t = 1|S_{t-1} = 1] = p_{11}.$$

3 Bayesian Inference

The goal of Bayesian inference is to derive the distributions of the parameters and the state variables conditional on the data. First, we construct posterior distribution via the Bayes' rule.

The posterior density of our model is

$$\begin{aligned}
p(\Theta, S|Y) &\propto p(\Theta, S)\, p(Y|\Theta, S) \\
&\propto p(\Theta)\, p(S|\Theta)\, p(Y|\Theta, S) \qquad (1)
\end{aligned}$$

where $\Theta = (\gamma, \alpha, \beta, \mu_0, \mu_1, p_{00}, p_{11})$, $S = (S_1, ..., S_T)$, and $Y = (y_1, ..., y_T)$. The Bayes' rule is applied in the first line. The second line comes from the definition of conditional probability.

$p(\Theta)$ is the prior for the parameters. Under the assumption of independence, the prior density is chosen as

$$\begin{aligned}
p(\Theta) &= p(\gamma)p(\mu_0)p(\mu_1)p(\alpha)p(\beta)p(p_{00})p(p_{11}) \\
&= \mathrm{N}(\mu_\gamma, \Sigma_\gamma) \times \mathrm{N}(\mu_{\mu_0}, \Sigma_{\mu_0}) \times \mathrm{N}(\mu_{\mu_1}, \Sigma_{\mu_1}) \\
&\quad \times \mathrm{N}(\mu_\alpha, \Sigma_\alpha) \times \mathrm{N}(\mu_\beta, \Sigma_\beta) \\
&\quad \times Beta(u_{00}, u_{01}) \times Beta(u_{11}, u_{10})
\end{aligned}$$

where $\mathrm{N}(\cdot)$ is the normal density function, and $Beta(\cdot)$ is the beta density function.

The second term in Equation 1 is $p(S|\Theta)$ [1]. Note that $p(S|\Theta) = p(S|p_{00}, p_{11})$ since the dynamics of S is dependent of p_{00}, p_{11}.

$$\begin{aligned} p(S|\Theta) &= p(S|p_{00}, p_{11}) \\ &= \prod_{t=1}^{T} p(S_{t+1}|S_t, p_{00}, p_{11}) \\ &= p_{00}{}^{\eta_{00}}(1-p_{00})^{\eta_{01}} p_{11}{}^{\eta_{11}}(1-p_{11})^{\eta_{10}} \end{aligned}$$

where η_{ij} refers to the number of the transitions from state i to j. The second line is due to the Markov property of S.

The last term, $p(Y|\Theta, S)$, is the likelihood function [2] given Θ and S.

$$\begin{aligned} p(Y|\Theta, S) &= \prod_{t=1}^{T} p(y_t|Y_{t-1}, S_t, ..., S_1, \Theta) \\ &= \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left[-\frac{\epsilon_t^2}{2\sigma_t^2}\right]. \end{aligned}$$

We can construct this likelihood function easily because we hypothetically consider S known as well as Θ. This is one of Bayesian features which is different from Classical inference. The classical likelihood is contructed with only parameters, Θ, being treated known.

Once we have posterior density function, we get marginal posterior density functions of parameters and state variables by integrating the posterior density function. Markov Chain Monte Carlo(MCMC) is one way of numerical intergration. MCMC algorithms are based on the Clifford-Hammersley theorem. The theorem says that a joint distribution can be characterized by its complete conditional distributions. In our context, the posterior distribution, $p(\Theta, S|Y)$, is characterized by the complete conditional distributions, $p(\Theta|S, Y)$ and $p(S|\Theta, Y)$.

Given the initial values, $\Theta^{(0)}$ and $S^{(0)}$, we draw $\Theta^{(1)}$ from $p(\Theta|S^{(0)}, Y)$ and then $S^{(1)}$ from $p(S|\Theta^{(1)}, Y)$. Iterating this steps, we finally get $\{S^{(g)}, \Theta^{(g)}\}_{g=1}^{G}$. Under some mild conditions, it is shown that the distribution of the sequence converges to the joint posterior distribution, $p(\Theta, S|Y)$.

---

[1] It is sometimes called likelihood function of S. One can consider it the prior for S.

[2] Some call it full information likelihood.

When a complete conditional distribution is known, such as Normal distribution or Beta distribution, we use Gibbs sampler to draw the random variable. When it is not, we may use the Metropolis-Hasting(MH) algorithm.

In the MH algorithm we generate a value $\hat{\Theta}$ from its proposal distribution $g(.)$ and accept the proposal value, i.e $\Theta^{(g+1)} = \hat{\Theta}$, with probability:

$$\lambda(\Theta^{(g)}, \hat{\Theta}) = \min\left\{ \frac{p(\hat{\Theta}|Y,S)}{p(\Theta^{(g)}|Y,S)} \Big/ \frac{g(\hat{\Theta})}{g(\Theta^{(g)})}, 1 \right\}.$$

Theoretically, we can use almost any distribution for the proposal distribution. In practice, however, we need to choose proposal distribution very carefully to ensure fast convergence of MCMC samples.

One version of MH algorithm is random walk MH algorithm. We generate $\hat{\Theta}$ from the random walk model of $\hat{\Theta} = \Theta^{(g)} + \eta_t$ where η_t has zero mean. The variance of η_t should be tuned carefully. The acceptance probability of random walk MH algorithm is

$$\lambda(\Theta^{(g)}, \hat{\Theta}) = \min\left\{ \frac{p(\hat{\Theta}|Y,S)}{p(\Theta^{(g)}|Y,S)}, 1 \right\}.$$

In this paper, we use random walk MH algorithm whenever we need to use MH algorithm.

4 MCMC Implementation

For MCMC implementation, we divide the parameters, Θ, into three categories:

$$\begin{aligned} \Theta_1 &= (p_{00}, p_{11}) \\ \Theta_2 &= (\gamma) \\ \Theta_3 &= (\mu_0, \mu_1, \alpha, \beta) \end{aligned}$$

The MCMC algorithm is summarized as followed:

- Draw $S_t, (t = 1, ..., T)$ from $p(S_t|S_{\neq t}, Y, \Theta)$ by the Single Move procedure
- Draw Θ_1 from $p(\Theta_1|S) \sim Beta$

- Draw Θ_2 from $p(\Theta_2|S, Y, \Theta_3)$ by MH
- Draw Θ_3 from $p(\Theta_3|S, Y, \Theta_2)$ by MH

For the MH algorithm for Θ_3, I use the following model, as in Nakatsuma(2000),

$$\epsilon_t^2 = \mu_0 + \mu_1 S_t + \sum_{j=1}^{l}(\alpha_j + \beta_j)\epsilon_{t-j}^2 + w_t - \sum_{j=1}^{s}\beta_j w_{t-j}, \quad w_t \sim \mathrm{N}(0, 2\sigma_t^4), \quad (2)$$

where $w_t = \epsilon_t^2 - \sigma_t^2$, $l = \max\{r, s\}$, $\alpha_j = 0$ for $j > r$,$\beta_j = 0$ for $j > s$ and $\epsilon_t^2 = 0$ and $w_t = 0$ for $t \leq 0$. Then, the corresponding likelihood function is

$$p(\epsilon^2|Y, S, \Theta_2, \Theta_3) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi(2\sigma_t^4)}} \exp\left[-\frac{w_t^2}{2(2\sigma_t^4)}\right], \quad (3)$$

where $\epsilon^2 = [\epsilon_1^2, \cdots, \epsilon_T^2]'$.

The details will be explained in the following sections

4.1 Generating S

Following the single move scheme suggested by Carlin, Polson, and Stoffer (1992) and Yoo(2004), the conditional posterior of S_t is derived from Equation 1 as follows:

$$\begin{aligned} p(S_t|Y, S_{\neq t}, \Theta) &\propto p(S|\Theta)p(Y|S, \Theta) \\ &\propto p(S_{t+1}|S_t, \Theta_1)p(S_t|S_{t-1}, \Theta_1) \\ &\quad p(y_t|Y_{t-1}, S_t, ..., S_1, \Theta)...p(y_T|Y_{T-1}, S_T, ..., S_1, \Theta) \\ &\propto p(S_{t+1}|S_t, \Theta_1)p(S_t|S_{t-1}, \Theta_1) \\ &\quad \prod_{s=t}^{T}(\sigma_s^2)^{-1/2} \exp\left[-\frac{1}{2}\sum_{s=t}^{T}\frac{\epsilon_s^2}{\sigma_s^2}\right]. \qquad (t = 1, ..., T) \quad (4) \end{aligned}$$

We calculate $Pr(S_t = 0|Y, S_{\neq t}, \Theta)$ and generate a random number from uniform distribution between 0 and 1. If the random number is less than the probability we set $S_t = 0$, otherwise 1.

The third factor shows how to handle the problem that σ_t^2 depends on not only S_t but also $S_{t-1}, ..., S_1$ in Bayesian context.

4.2 Generating p_{00} and p_{11}

The conditional distribution of p_{00} is given by

$$\begin{aligned}
p(p_{00}|Y, S, \Theta - p_{00}) &\propto p(p_{00})p(S|p_{00}, p_{11}) \\
&\propto {p_{00}}^{u_{00}-1}(1-p_{00})^{u_{01}-1}{p_{00}}^{\eta_{00}}(1-p_{00})^{\eta_{01}} \\
&\propto {p_{00}}^{\eta_{00}+u_{00}-1}(1-p_{00})^{\eta_{01}+u_{01}-1}
\end{aligned}$$

This is the Beta density function. Therefore, we generate p_{00} by Gibbs sampler from the following Beta distributions:

$$p_{00}|S \sim Beta(u_{00} + \eta_{00}, u_{01} + \eta_{01})$$

We generate p_{11} in a similar way.

$$p_{11}|S \sim Beta(u_{11} + \eta_{11}, u_{10} + \eta_{10}).$$

4.3 Generating γ: Regression Coefficients

The conditional distribution of γ is given by

$$\begin{aligned}
p(\gamma|Y, S, \Theta - \gamma) &\propto p(\gamma)p(Y|S, \Theta) \\
&\propto \mathrm{N}(\mu_\gamma, \Sigma_\gamma) \times \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left[-\frac{(y_t - x_t'\gamma)^2}{2\sigma_t^2}\right]
\end{aligned}$$

This is not any standard or known distribution because σ_t^2 is also a function of γ. Therefore, we use MH algorithm in this case.

Let $Y_\gamma = [y_1^*, \cdots, y_T^*]'$ and $X_\gamma = [x_1^{*\prime}, \cdots, x_T^{*\prime}]'$. Then we have the following proposal distribution for γ:

$$\gamma|Y, S, \Theta - \gamma \sim \mathrm{N}(\hat{\mu}_\gamma, \hat{\Sigma}_\gamma),$$

where $\hat{\mu}_\gamma = \hat{\Sigma}_\gamma(X_\gamma'\Sigma^{-1}Y_\gamma + \Sigma_\gamma^{-1}\mu_\gamma)$, $\hat{\Sigma}_\gamma = (X_\gamma'\Sigma^{-1}X_\gamma + \Sigma_\gamma^{-1})^{-1}$, and $\Sigma = \mathrm{diag}\{\sigma_1^2, ..., \sigma_T^2\}$.

4.4 Generating (μ_0, α)

The conditional distribution of (μ_0, α) is

$$\begin{aligned} p(\mu_0, \alpha | Y, S, \Theta - (\mu_0, \alpha)) &\propto p(\mu_0, \alpha) p(Y|S, \Theta) \\ &\propto \mathrm{N}(\mu_\alpha, \Sigma_\alpha) \times \mathrm{N}(\mu_{\mu_0}, \Sigma_{\mu_0}) \\ & \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left[-\frac{\epsilon_t^2}{2\sigma_t^2}\right] \end{aligned}$$

The likelihood function is a very complicated function of (μ_0, α). To use MH algorithm and get a good proposal density, we use Equation 2 and choose the equations in which $S_t = 0$.

$$\epsilon_t^2 = \mu_0 + \sum_{j=1}^{l} (\alpha_j + \beta_j)\epsilon_{t-j}^2 + w_t - \sum_{j=1}^{s} \beta_j w_{t-j}, \quad w_t \sim \mathrm{N}(0, 2\sigma_t^4), \tag{5}$$

where $t \in I_0$ and I_0 is an index set of t such that $S_t = 0$.

Calculate $\tilde{\iota}_t$, $\tilde{\epsilon}_t^2$, and ζ_t by the following transformation:

$$\begin{aligned} \tilde{\iota}_t &= 1 + [\tilde{\iota}_{t-1}, \cdots, \tilde{\iota}_{t-s}]'\beta \\ \tilde{\epsilon}_t^2 &= \epsilon_t^2 + [\tilde{\epsilon}_{t-1}^2, \cdots, \tilde{\epsilon}_{t-s}^2]'\beta \\ \zeta_t &= [\tilde{\iota}_t, \tilde{\epsilon}_{t-1}^2, \cdots, \tilde{\epsilon}_{t-r}^2] \end{aligned}$$

where $\tilde{\iota}_t = 0$ and $\tilde{\epsilon}_t^2 = 0$ for $t \leq 0$. It turns out that $w_t = \epsilon_t^2 - \zeta_t'\alpha$. The likelihood function (Eq 3) can be rewritten as

$$f(\epsilon^2 | Y, S, \Theta_2, \Theta_3) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi(2\sigma_t^4)}} \exp\left[-\frac{(\epsilon_t^2 - \zeta_t'\alpha)^2}{2(2\sigma_t^4)}\right].$$

Let $Y_\alpha = [\epsilon_1^2, \cdots, \epsilon_T^2]'$ and $X_\alpha = [\zeta_1', \cdots, \zeta_T']'$. Then we have the following proposal distribution for (μ_0, α):

$$\mu_0, \alpha | Y, S, \Theta_{-(\mu_0, \alpha)} \sim \mathrm{N}(\hat{\mu}_\alpha, \hat{\Sigma}_\alpha)$$

where $\hat{\mu}_\alpha = \hat{\Sigma}_\alpha (X_\alpha' \Lambda^{-1} Y_\alpha + \Sigma_\alpha^{-1}\mu_\alpha)$, $\hat{\Sigma}_\alpha = (X_\alpha' \Lambda^{-1} X_\alpha + \Sigma_\alpha^{-1})^{-1}$, and $\Lambda = \mathrm{diag}\{2\sigma_1^4, \cdots, 2\sigma_n^4\}$.

4.5 Generating μ_1

Generating μ_1 is similar to generating (μ_0, α). In this case, however, we consider the case of $S_t = 1$. We use the following model:

$$\epsilon_t^2 = \mu_0 + \mu_1 + \sum_{j=1}^{l} (\alpha_j + \beta_j)\epsilon_{t-j}^2 + w_t - \sum_{j=1}^{s} \beta_j w_{t-j}, \quad w_t \sim \mathrm{N}(0, 2\sigma_t^4), \tag{6}$$

where $t \in I_1$ and I_1 is an index set of t such that $S_t = 1$. We draw $(\mu_0+\mu_1, \alpha)$ as in previous procesure and get μ_1 by subtracting μ_0 from $\mu_0 + \mu_1$.

4.6 Generating β

Following Nakatsuma(2000), we linearize w_t by the first-order Taylor expansion.

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

where $w_t(\beta^*) = \epsilon_t^2 - \mu_0 - \mu_1 S_t - \sum_{j=1}^{l}(\alpha_j + \beta_j^*)\epsilon_{t-j}^2 + \sum_{j=1}^{s} \beta_j^* w_{t-j}$, $\xi_t = [\xi_{1t}, \cdots, \xi_{qt}]$ is the first-order derivative of $w_t(\beta)$ evaluated at β^* given by the following recursion:

$$\xi_{it} = -\epsilon_{t-i}^2 + w_{t-i}(\beta^*) + \sum_{j=1}^{s} \beta_j^* \xi_{i,t-j}, \quad (i = 1, \cdots, s),$$

where $\xi_{it} = 0$ for $t \leq 0$. β^* is the non-linear least squares estimate of β,

$$\beta^* = \arg\min_{\beta} \sum_{t=1}^{T} \{w_t(\beta)\}^2/(2\sigma_t^4).$$

Then the likelihood function (Eq 3) can be rewritten as

$$f(\epsilon^2|Y, S, \Theta_2, \Theta_3) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi(2\sigma_t^4)}} \exp\left[-\frac{\{w_t(\beta^*) - \xi_t(\beta - \beta^*)\}^2}{2(2\sigma_t^4)}\right].$$

Let $Y_\beta = [w_1(\beta^*)+\xi_1\beta^*, \cdots, w_T(\beta^*)+\xi_T\beta^*]'$ and $X_\beta = [\xi_1', \cdots, \xi_T']'$. Then, we have the following proposal distribution of β:

$$\beta|Y, S, \Theta_{-\beta} \sim \mathrm{N}(\hat{\mu}_\beta, \hat{\Sigma}_\beta)$$

where $\hat{\mu}_\beta = \hat{\Sigma}_\beta(X_\beta'\Lambda^{-1}Y_\beta + \Sigma_\beta^{-1}\mu_\beta)$, $\hat{\Sigma}_\beta = (X_\beta'\Lambda^{-1}X_\beta + \Sigma_\beta^{-1})^{-1}$ and $\Lambda = \mathrm{diag}\{2\sigma_1^4, \cdots, 2\sigma_n^4\}$.

Table 1: Case: Markov switching GARCH(1,1)

	true	posterior statistics			acceptance
	value	mean	s.d.	$\hat{\rho}^a$	rate
γ	0	-0.002	0.010	0.86	0.17
μ_0	0.01	0.009	0.002	0.96	0.13
μ_1	0.05	0.150	0.099	0.88	0.14
α	0.2	0.195	0.060	0.93	0.13
β	0.5	0.553	0.079	0.98	0.19
p_{00}	0.98	0.990	0.006	0.38	
p_{11}	0.95	0.881	0.077	0.32	

Notes: (a) the first-order autocorrelation in a sample path.

5 Numerical Example

In this section, we simulate a Markov switching GARCH model and estimate the model by our MCMC algorithm.

The model is Markov switching GARCH(1,1).

$$\begin{aligned} y_t &= \gamma + \epsilon_t, \qquad \epsilon_t = \sigma_t \omega_t, \qquad \omega_t \sim N(0,1) \\ \sigma_t^2 &= \mu_0 + \mu_1 S_t + \alpha \epsilon_t^2 + \beta \sigma_t^2 \end{aligned}$$

The sample size is $T = 500$. We generate 6,000 iterations for MCMC and discard the first 1,000. Every third draw is selected to construct posterior densities. Table 1 shows that our algorithm works reasonably.

Conclusion

When we estimate Markov switching GARCH models by MLE, it is not easy to construct likelihood function because the conditional variance depends on all the history of state variable. We showed that one can handle this problem easily in Bayesian inference. A numerical illustration shows that our MCMC algorithm works well.

Our model can be extended to include ARMA structure and state dependent mean. our MCMC algorithm need to be improved especially in drawing the state varible. Kaufman and Fruhwirth-Schnatter (2002) estimates Markov switching ARCH models in Bayesian context because, as they say, the mulimove procesure of state variable does not work in GARCH. On the other hand, we show that we can use the single move procesure in Markov switching GARCH.

Literature starts to incoporate conditional variance dynamics in derivative pricing.Previously the 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.

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.

[53] B.Hark Yoo. A Bayesian Analysis of Markov Switching Models with ARMA-GARCH Errors. *Working Paper, Rutgers University*, 2004.