# Viterbi-Based Estimation for Markov Switching GARCH Model

Robert J. Elliott * John W. Lau † Hong Miao ‡ Tak Kuen Siu §

August 22, 2011

**Abstract**

We outline a two-stage estimation method for a Markov-switching GARCH model modulated by a hidden Markov chain. The first stage involves the estimation of a hidden Markov chain using the Vitberi algorithm given the model parameters. The second stage employs the maximum likelihood method to estimate model parameters given the estimated hidden Markov chain. Applications to financial risk management are discussed via simulated data.

**Keywords**: Volatility; Regime Switching; GARCH; Viterbi algorithm; Reference Probability; Filter; Maximum Likelihood Estimation; Value at Risk

*Corresponding Author: Haskayne School of Business, University of Calgary; Email: relliott@ucalgary.ca; Phone: 001 403 220 5540; Fax: 001 403 210 3327

†Department of Mathematics and Statistics, University of Western Australia; Email: john.lau@uwa.edu.au; Phone:(61-8) 6488 3371

‡Department of Finance and Real Estate, Colorado State University; E-mail: hong.miao@colostate.edu; Phone: (970) 491 2356

§Department of Applied Finance and Actuarial Studies, Faculty of Business and Economics, Macquarie University; Email: ktksiu2005@gmail.com; Phone: (61-2) 9850 8573

# 1 Introduction

The volatility of prices measures the variability of price changes. Volatility is defined as the standard deviation of the return of an asset. Since Engle (1982) in his Nobel Prize winning work introduced the concept of conditional heteroscedasticity and the well known ARCH models, this method of measuring volatility has expanded. Taylor (1986) suggested the GARCH (1,1) model and Bollerslev (1986) independently extended ARCH to GARCH (p, q). Following this work, GARCH models are regarded as the most powerful models for analyzing volatility dynamics.

Regime switching models were introduced to economics in 1972 when researchers recognized that parameters may switch due to structural shifts which divide the sample period into different regimes. Hamilton (1989) introduced Markov switching models to the econometric mainstream. Gray (1996) combined GARCH effects with Markov switching and gave estimates of the parameters by introducing a recombining method and collapsing the conditional variances in each regime into a single variance at each point of time. Haas et al. (2004) solved the path dependence problems by separating the GARCH dynamics from the Markov chain. All of these models employed the maximum likelihood method to estimate parameters, though sometimes the expression for the log maximum likelihood function is difficult to obtain.

Filtering is a commonly used technique, particularly in engineering problems. Using filtering, the estimates of parameters of a model can be continuously updated on the basis of currently available information. The most frequently used filters are the Wonham filter for Markov chain and the Kalman filter. Elliott (1993, 1994) considered a finite-state Markov chain partially observed in Gaussian noises in both the continuous-time and discrete-time frameworks. They used the Expectation Maximization, (EM), algorithm to update parameter estimates. Bhar and Chiarella (1995) used Kalman filtering techniques to estimate the HJM (Heath-Jarrow-Morton) model by reducing it to Markovian dynamics. Chiarella et al. (2001) considered the HJM model and proposed a framework which provides a recursive filtering algorithm. Elliott et al. (2002) applied a robust form of filtering equations for a continuous time hidden Markov model to estimate the volatility of a risky asset. They improved the classical filtering formulae by eliminating stochastic integration. Chiarella et al. (2005) proposed a three-factor volatility specification and analyzed the volatility structure of inter-bank offered rates in three different markets using the extended Kalman filter. Bauwens et. al. (2006) developed univariate regime-switching GARCH (RS-GARCH) models in which the conditional variance switches in time from one GARCH process to another.

In this paper, we outline an estimation method for a Markov-switching GARCH model modulated by a hidden Markov chain using filtering together with the Viterbi algorithm. The proposed estimation method has two stages. At the first stage, we adopt the filtering method and the Viterbi algorithm to estimate the hidden Markov chain. We first derive a recursive equation for the unnormalized filter for the hidden Markov chain using a reference probability and a version of the Bayes' rule. Then the Viterbi algorithm is used to approximate the recursive equation and to estimate the hidden Markov chain given the model parameters. At the second stage, given the estimated hidden Markov chain from the first stage, we adopt the maximum likelihood method to estimate the model parameters. These estimates can then be used to update the inputs at the first stage. Using simulated data, we discuss some applications of the switching GARCH(1,1) model in financial risk management.

This article is organized as follows. In Section 2 we describe the model and state the underlying assumptions. Section 3 gives the recursive filter and discusses the Viterbi algorithm to approximate the recursive equation as well as to estimate the hidden Markov chain. In Section 4, we discuss the maximum likelihood method to estimate the model parameters, given the estimated hidden Markov chain. Applications to financial risk management are discussed in Section 5. The final section summarizes the paper.

## 2 A Markov Switching GARCH Model

In this section, we present a Markov-switching GARCH model for describing the volatility of a risky asset. To model uncertainty, we consider a complete probability space $(\Omega, \mathcal{F}, P)$, where $P$ is the historical probability measure. Suppose the price process of an individual risky asset follows:

$$S_t = S_{t-1} \exp\left(\mu - \frac{\sigma_t^2}{2} + \sigma_t v_t\right)$$

where $v = \{v_t, t = 1, 2, ...\}$ is a sequence of i.i.d. $N(0,1)$ random variables defined on $(\Omega, \mathcal{F}, P)$. Then its return at time $t$ is

$$Y_t = \log S_t - \log S_{t-1} = \mu - \frac{\sigma_t^2}{2} + \sigma_t v_t, \tag{2.1}$$

where $S_t$ and $S_{t-1}$ are the prices of the individual asset at times $t$ and $t-1$.

Given $\sigma_t$, the return of the asset is a Gaussian function of mean $\mu$ and volatility $\sigma^2$. Volatility forecasts are important for hedging risky assets and pricing options. Although the conventional GARCH models are applied in forecasting volatility, they provide forecasts which are too high following above-normal periods. In our model we allow the volatility to be influenced by the "state of the world". Its dynamics can experience discrete jumps in the parameters. That is, the volatility of the asset follows different dynamics in different states of the world, or different conditions of the market. However, the states of the world are not observable directly. They are hidden in the observed return process. We represent the states of the world by a Markov chain with a transition matrix $A$. We estimate the "state of the world" by observing the returns of the risky asset. Let $X$ be the state of the world. We suppose that $X$ and $v$ are independent under $P$. Following the canonical representation of a Markov process introduced in Elliott (1993), without loss of generality, we can take the state space of $X$ to be the set $S = \{e_1, e_2, ..., e_n\}$, where $e_i$ is a column vector in $\mathbb{R}^n$ with unity in the $i$th position and zero elsewhere.

In our model, we assume that the world has two states, say, a "good" state and a "bad" one. Consequently, the state space of $X$ is taken to be the set of unit vectors

$S = \{e_1, e_2\}$, where $e_1 = (1, 0)$ and $e_2 = (0, 1)$. It is straightforward to generalize our model to any number of states.

Suppose $p_{ji} = P(X_t = e_j | X_{t-1} = e_i)$, and write $A = (p_{ji}), 1 \leq i, j \leq 2$, for the transition matrix of the chain $X$. Then, as in Elliott et al. (2008), the dynamics of the chain can be written as:

$$X_t = AX_{t-1} + M_t. \tag{2.2}$$

where $\{M_t, t = 1, 2, \cdots\}$ is a martingale increment process under $P$.

We suppose that the volatility of the return has dynamics given by a Markov-switching GARCH (1, 1) model as follows:

$$\sigma_t^2 = \alpha(t) + \beta(t) E[\sigma_{t-1}^2 | \mathcal{F}_{t-1}^Y] + \theta(t) E[\sigma_{t-1}^2 | \mathcal{F}_{t-1}^Y] v_{t-1}^2. \tag{2.3}$$

Here the coefficients $\alpha(t)$, $\beta(t)$ and $\theta(t)$ are given by:

$$\begin{aligned}
\alpha(t) &= \langle \alpha, X_{t-1} \rangle, \\
\beta(t) &= \langle \beta, X_{t-1} \rangle, \\
\theta(t) &= \langle \theta, X_{t-1} \rangle,
\end{aligned}$$

where

$$\alpha = \begin{pmatrix} \alpha^1 \\ \alpha^2 \end{pmatrix}, \beta = \begin{pmatrix} \beta^1 \\ \beta^2 \end{pmatrix}, \theta = \begin{pmatrix} \theta^1 \\ \theta^2 \end{pmatrix}.$$

Write $h_t = \sigma_t^2$, so that $\sigma_t = \sqrt{h_t}$. Introducing the state process $X$, we have the following dynamics:

$$\begin{aligned}
h_t &= \alpha(t) + \beta(t) E[h_{t-1} | \mathcal{F}_{t-1}^Y] + \theta(t) E[h_{t-1} | \mathcal{F}_{t-1}^Y] v_{t-1}^2 \\
&= \langle \alpha, X_{t-1} \rangle + \langle \beta, X_{t-1} \rangle E[h_{t-1} | \mathcal{F}_{t-1}^Y] + \langle \theta, X_{t-1} \rangle E[h_{t-1} | \mathcal{F}_{t-1}^Y] v_{t-1}^2, \qquad (2.4) \\
Y_t &= \mu - \frac{h_t}{2} + \sqrt{h_t} v_t. \qquad (2.5)
\end{aligned}$$

Consequently the volatility dynamics change between two different regimes. The shifts are determined by the state process $X$.

When there is only one regime in the model, the Markov-switching GARCH (1, 1) model becomes the standard GARCH (1, 1) model:

$$h_t = \alpha + \beta h_{t-1} + \theta h_{t-1} v_{t-1}^2, \tag{2.6}$$

$$Y_t = \mu - \frac{h_t}{2} + \sqrt{h_t} v_t. \tag{2.7}$$

Define the following filtrations:

$$\begin{aligned}
\mathcal{F}_t^X &= \sigma \left\{ X_0, X_1, ..., X_t \right\}, \\
\mathcal{F}_t^Y &= \sigma \left\{ Y_1, Y_2, ..., Y_t \right\}, \\
\mathcal{F}_t^{X,Y} &= \sigma \left\{ X_0, X_1, ..., X_t, Y_1, Y_2, ..., Y_t \right\}.
\end{aligned}$$

Then for the Markov-switching GARCH(1, 1) model,

$$\begin{aligned}
E\left[ Y_t | \mathcal{F}_{t-1}^{X,Y} \right] &= E\left[ \mu - \frac{\sigma_t^2}{2} + \sigma_t v_t | \mathcal{F}_{t-1}^{X,Y} \right] \\
&= \mu - \frac{\sigma_t^2}{2} = \mu - \frac{1}{2} h_t.
\end{aligned}$$

Since $v_t \sim N(0,1)$,

$$\begin{aligned}
Var\left[ Y_t | \mathcal{F}_{t-1}^{X,Y} \right] &= E\left[ \left( Y_t - E\left[ Y_t | \mathcal{F}_{t-1}^{X,Y} \right] \right)^2 | \mathcal{F}_{t-1}^{X,Y} \right] \\
&= E\left[ (\sigma_t v_t)^2 | \mathcal{F}_{t-1}^{X,Y} \right] = \sigma_t^2 = h_t.
\end{aligned}$$

Note that $\sigma_t$ (and $h_t$) is $\mathcal{F}_{t-1}^{X,Y}$-measurable. Consequently, to estimate the Markov-switching GARCH (1, 1) model, we must estimate the hidden Markov chain $X$ and the unknown model parameters $\mu$, $\alpha_i$, $\beta_i$ and $\theta_i$, $i = 1, 2$. We shall propose a two-stage estimation method, where we estimate the hidden Markov chain given the model parameters at the first stage and estimate the unknown parameters given the estimated hidden Markov chain at the second stage.

## 3 Filtering: The First-Stage of Estimation

In this section, we discuss the first stage of the estimation method, where the filtering approach is used to estimate the hidden Markov chain $X$ given the model parameters $\mu$, $\alpha_i$, $\beta_i$ and $\theta_i$, $i = 1, 2$. In particular, given the returns data $Y_1$, $Y_2$, $\cdots$, $Y_t$ up to and including time $t$, we wish to estimate the state $X_t$ of the hidden Markov chain $X$. To derive the filter for $X_t$ given $\mathcal{F}_t^Y$, we use the change of measure technique and start with a "reference" probability $\bar{P}$. We suppose that under the probability measure $\bar{P}$, $\{Y_1, Y_2, \cdots, Y_t\}$ is a sequence of independent and identically distributed, (i.i.d.), random variables each of which is $N(0, 1)$, the standard Normal distribution. Write

$$\phi(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} ,$$

for the density function of $N(0, 1)$.

Define, for each $k = 1, 2, \cdots$,

$$\lambda_k := \frac{\phi((h_k)^{-\frac{1}{2}}(Y_k - \mu + \frac{1}{2}h_k))}{\sqrt{h_k}\phi(Y_k)} ,$$

with $\lambda_0 = 1$.

Consider an $\mathcal{F}^{X,Y}$-adapted process $\{\Lambda_t, t = 0, 1, \cdots\}$ defined by putting:

$$\Lambda_t := \prod_{k=1}^{t} \lambda_k , \quad \Lambda_0 = 1 .$$

We now define a measure $P$ by setting

$$\left.\frac{dP}{d\bar{P}}\right|_{\mathcal{F}_t^{X,Y}} := \Lambda_t .$$

Then as shown in the following lemma, $P$ is the "real world" probability.

**Lemma 3.1.** *Under $P$, $\{v_t, t = 0, 1, 2, \cdots\}$ is a sequence of i.i.d. $N(0, 1)$ random vari-*

*ables, where*

$$\nu_t = \frac{Y_t - \mu + \frac{1}{2}h_t}{\sqrt{h_t}}.$$

**Proof.** Let $f$ be a real-valued, measurable test function on $\Re$. Then by a version of the Bayes' rule and the $\mathcal{F}_{t-1}^{X,Y}$-measurability of $\Lambda_{t-1}$,

$$\begin{aligned} E[f(v_t)|\mathcal{F}_{t-1}^{X,Y}] &= \frac{\bar{E}[\Lambda_t f(v_t)|\mathcal{F}_{t-1}^{X,Y}]}{\bar{E}[\Lambda_t|\mathcal{F}_{t-1}^{X,Y}]} \\ &= \frac{\bar{E}[\lambda_t f(v_t)|\mathcal{F}_{t-1}^{X,Y}]}{\bar{E}[\lambda_t|\mathcal{F}_{t-1}^{X,Y}]} \ . \end{aligned}$$

Note that

$$\begin{aligned} E[\lambda_t|\mathcal{F}_{t-1}^{X,Y}] &= E\left[\frac{\phi((h_t)^{-\frac{1}{2}}(Y_t - \mu + \frac{1}{2}h_t))}{\sqrt{h_t}\phi(Y_t)}|\mathcal{F}_{t-1}^{X,Y}\right] \\ &= \int_{-\infty}^{\infty} \frac{\phi((h_t)^{-\frac{1}{2}}(y - \mu + \frac{1}{2}h_t))}{\sqrt{h_t}\phi(y)}\phi(y)dy \\ &= \int_{-\infty}^{\infty} \frac{\phi((h_t)^{-\frac{1}{2}}(y - \mu + \frac{1}{2}h_t))}{\sqrt{h_t}}dy \\ &= \int_{-\infty}^{\infty} \phi(z)dz = 1 \ . \end{aligned}$$

The second last equality follows by letting $z := (h_t)^{-\frac{1}{2}}(y - \mu + \frac{1}{2}h_t)$, for each $t = 1, 2, \cdots$. Now

$$\begin{aligned} \bar{E}[\Lambda_t f(v_t)|\mathcal{F}_{t-1}^{X,Y}] &= \int_{-\infty}^{\infty} \frac{\phi((h_t)^{-\frac{1}{2}}(y - \mu + \frac{1}{2}h_t))}{\sqrt{h_t}\phi(y)} f\left(\frac{y - \mu + \frac{1}{2}h_t}{\sqrt{h_t}}\right)\phi(y)dy \\ &= \int_{-\infty}^{\infty} \frac{\phi((h_t)^{-\frac{1}{2}}(y - \mu + \frac{1}{2}h_t))}{\sqrt{h_t}} f\left(\frac{y - \mu + \frac{1}{2}h_t}{\sqrt{h_t}}\right)dy \\ &= \int_{-\infty}^{\infty} f(z)\phi(z)dz \ . \end{aligned}$$

Consequently,

$$E[f(v_t)|\mathcal{F}_{t-1}^{X,Y}] = \int_{-\infty}^{\infty} f(z)\phi(z)dz \ .$$

This does not depend on $\mathcal{F}_{t-1}^{X,Y}$, so the result follows. $\square$

We wish to estimate $X_t$ given $\mathcal{F}_t^Y$ under the "real world" probability $P$. That is, we evaluate $E[X_t|\mathcal{F}_t^Y]$ which is an optimal estimate of $X_t$ given $\mathcal{F}_t^Y$ in the mean-square sense. Again by a version of the Bayes' rule,

$$E[X_t|\mathcal{F}_t^Y] = \frac{\bar{E}[\Lambda_t X_t|\mathcal{F}_t^Y]}{\bar{E}[\Lambda_t|\mathcal{F}_t^Y]} \ . \tag{3.8}$$

Write, for each $t = 1, 2, \cdots$,

$$q_t := \bar{E}[\Lambda_t X_t|\mathcal{F}_t^Y] \in \Re^2 \ .$$

Instead of evaluating $E[X_t|\mathcal{F}_t^Y]$ directly, it is more convenient to evaluate $q_t$. Then from (3.8),

$$E[X_t|\mathcal{F}_t^Y] = \frac{q_t}{\langle q_t, \mathbf{1}\rangle} \ .$$

Here $\mathbf{1} := (1,1)' \in \Re^2$ and $q_0 = E[X_0]$, which is the initial distribution of $X_0$.

The following theorem gives a recursion for $q_t$, $t = 1, 2, \cdots$.

**Theorem 3.1.** *Let* $h_t^i = \alpha_i + \beta_i E[h_{t-1}|\mathcal{F}_{t-1}^Y] + \theta_i E[h_{t-1}|\mathcal{F}_{t-1}^Y] v_{t-1}^2$*, for each* $t = 1, 2, \cdots$ *and* $i = 1, 2$*. Write*

$$B(Y_t) := \begin{pmatrix} \frac{\phi((h_t^1)^{-\frac{1}{2}}(Y_t - \mu + \frac{1}{2}h_t^1))}{\sqrt{h_t^1}\phi(Y_t)} & 0 \\ 0 & \frac{\phi((h_t^2)^{-\frac{1}{2}}(Y_t - \mu + \frac{1}{2}h_t^2))}{\sqrt{h_t^2}\phi(Y_t)} \end{pmatrix} .$$

*Then* $q$ *satisfies the recursion:*

$$q_t\,(z) = AB(Y_t)q_{t-1} \ , \quad t = 1, 2, \cdots$$

**Proof.** For each $t = 1, 2, \cdots$,

$$
\begin{aligned}
q_t &= E[\Lambda_t X_t | \mathcal{F}_t^Y] \\
&= E[\Lambda_{t-1}\lambda_t(AX_{t-1} + M_t)|\mathcal{F}_t^Y] \\
&= E\left[\Lambda_{t-1}AX_{t-1}\frac{\phi((h_t)^{-\frac{1}{2}}(Y_t - \mu + \frac{1}{2}h_t))}{\sqrt{h_t}\phi(Y_t)}|\mathcal{F}_t^Y\right] \\
&= \sum_{i=1}^{2} E\left[\Lambda_{t-1}Ae_i\frac{\phi((h_t^i)^{-\frac{1}{2}}(Y_t - \mu + \frac{1}{2}h_t^i))}{\sqrt{h_t^i}\phi(Y_t)}\langle X_{t-1}, e_i\rangle |\mathcal{F}_t^Y\right] \\
&= \sum_{i=1}^{2} A\langle q_{t-1}, e_i\rangle \frac{\phi((h_t^i)^{-\frac{1}{2}}(Y_t - \mu + \frac{1}{2}h_t^i))}{\sqrt{h_t^i}\phi(Y_t)} e_i \\
&= AB(Y_t)q_{t-1} \ .
\end{aligned}
$$

□

We now describe the Viterbi algorithm to estimate the hidden Markov chain $X$. The main idea of the Viterbi algorithm is that the expected values represented by the summations in the recursion in Theorem 3.1. are replaced by the corresponding maximum likelihoods. In other words, the sums are replaced by the maxima.

Let $[B(Y_t)]_{ii}$ be the $(i, i)$-element of the matrix $B(Y_t)$, for each $k = 1, 2, \cdots$ and $i = 1, 2$. Then using the Viterbi algorithm, a new unnormalized estimate $q_t^* = (q_t^*(1), q_t^*(2))' \in \Re^2$ is given by the following recursion:

$$
q_t^*(i) = \max\{p_{i1}[B(Y_t)]_{11}q_{t-1}^*(1), p_{i2}[B(Y_t)]_{22}q_{t-1}^*(2)\} \ , \quad i = 1, 2 \ .
$$

It is obvious that $q_t^*(i) > 0$, for each $i = 1, 2$.

The Viterbi probabilities are then defined as:

$$
\rho_t(i) := \frac{q_t^*(i)}{q_t^*(1) + q_t^*(2)} \ , \quad i = 1, 2 \ ,
$$

so $\rho_t(1) + \rho_t(2) = 1$.

Since $q_t^*$ is an approximation to $q_t$, $\rho_i(i)$ is an estimate of the conditional probability that $X_t = e_i$ given $\mathcal{F}_{t-1}^Y$, for each $i = 1, 2$ and $t = 1, 2, \cdots$.

We can then estimate the state $X_t$ of the hidden Markov chain $X$ at time $t$ by maximizing the approximated likelihood function by the Viterbi algorithm. That is, the estimate $\widehat{X}_t$ is $e_k$ if

$$k = \arg\max\{\rho_t(1), \rho_t(2)\} \ .$$

# 4 Maximum Likelihood Estimation: The Second-Stage of Estimation

In this section, we estimate the model parameters using the maximum likelihood estimation. Suppose we have estimated the hidden Markov chain $X$ using the method described in the last section. Given the estimated hidden Markov chain $\widehat{X}$, we wish to estimate the transition matrix $A = (p_{ji})$, the mean return $\mu$ and the GARCH parameters $\alpha_i$, $\beta_i$ and $\theta_i$, for $i = 1, 2$, where

1. $p_{ji} \geq 0$ and $\sum_{j=1}^{2} p_{ji} = 1$;
2. $\alpha_i, \beta_i, \theta_i \in \Re^+$;
3. for each state $\mathbf{e}_i$, the peristence parameter $\alpha_i + \beta_i < 1$.

For the estimation of the transition probabilities $p_{ji}$, $i, j = 1, 2$, we resort to the simple counting method. Define, for each $i, j = 1, 2$,

$$\widehat{N}_t^{ji} := \sum_{k=1}^{t} \left\langle \widehat{X}_k, e_j \right\rangle \left\langle \widehat{X}_{k-1}, e_i \right\rangle \ ,$$

which represents an estimate of the number of transitions of the estimated hidden Markov chain $\widehat{X}$ from state $e_i$ to state $e_j$ up to and including time $t$ given the observed returns. Note that $\widehat{N}_t^{ji}$ is $\mathcal{F}_t^Y$-measurable.

Similarly, we define the estimated occupation time of the estimated hidden Markov chain $\widehat{X}$ in state $e_i$ up to time $t$ as:

$$\widehat{O}_t^i := \sum_{k=0}^{t} \langle X_k, e_i \rangle \ .$$

Note also that $\widehat{O}_t^i$ is $\mathcal{F}_t^Y$-measurable.

For each $i, j = 1, 2$, an estimate of the transition probability $p_{ji}$ given $\mathcal{F}_t^Y$ is:

$$\widehat{p}_{ji}(t) := \frac{\widehat{N}_t^{ji}}{\widehat{O}_t^i} \ .$$

We now estimate the mean return $\mu$ and the GARCH parameters $\alpha_i$, $\beta_i$ and $\theta_i$, for $i = 1, 2$ using maximum likelihood estimation. Write $\eta := (\mu, \alpha_1, \alpha_2, \beta_1, \beta_2, \theta_1, \theta_2)$, for the vector of unknown parameters. Then it is easy to show that the log likelihood function of $\eta$ given $\mathcal{F}_t^Y$ and the estimated hidden Markov chain $\widehat{X}$ is:

$$\begin{aligned} & l(\eta | \mathcal{F}_t^Y, \widehat{X}_1, \widehat{X}_2, \cdots, \widehat{X}_t) \\ & := \sum_{k=1}^{t} \left[ \sum_{i=1}^{2} \left( -\frac{1}{2} \ln(2\pi h_k(\alpha_i, \beta_i, \theta_i)) - \frac{(Y_t - \mu)^2}{2h_k(\alpha_i, \beta_i, \theta_i)} \right) \langle X_{k-1}, e_i \rangle \right] , \end{aligned}$$

where

$$h_k(\alpha_i, \beta_i, \theta_i) := \alpha_i + \beta_i E[h_{k-1} | \mathcal{F}_{k-1}^Y] + \theta_i E[h_{k-1} | \mathcal{F}_{k-1}^Y] v_{k-1}^2 \ .$$

We estimate $\eta$ given $\mathcal{F}_t^Y$, $\widehat{X}_1$, $\widehat{X}_2$, $\cdots$, $\widehat{X}_t$ as follows:

$$\widehat{\eta}_t := \arg\max_{\eta} l(\eta | \mathcal{F}_t^Y, \widehat{X}_1, \widehat{X}_2, \cdots, \widehat{X}_t) \ .$$

After we determine $\widehat{\eta}_t$, we use it as an input in the first stage of estimation in Section 3. We repeat the two stages iteratively until convergence is achieved.

Here we outline the main idea and some theoretical results of the two-stage estimation scheme based on filtering. For practical implementation of the proposed method, there are some remaining issues to be discussed. For example, what are the conditions for the convergence of the two-stage iterative scheme? What is the rate of convergence of the scheme? What are the statistical properties of the estimators of the scheme, such as asymptotic properties? These are open issues and represent potential topics for further econometric and statistical research. Intuitively, one may discuss the convergence and statistical properties of the estimators for the two-stage scheme by looking at these issues at each stage of the scheme. The underlying principle for this strategy is divide-and-rule; this breaks down a complex procedure into more simpler sub-procedures and analyses the properties of each. At the first-stage of the estimation, we adopt the filtering approach together with the Viterbi algorithm to estimate the states of the hidden Markov chain. There is previous work which discuss the convergence and statistical properties of filtering hidden Markov models, or related models, and the Viterbi algorithm. For example, the convergence of filtered estimates for hidden Markov models is discussed in Elliott and Moore (1997). Dufour, Elliott and Tsoi (1995) provided an asymptotic study for filtering linear systems with jump parameters. The convergence of the Vitberi algorithm was discussed in Elliott, Aggoun and Moore (2008). These works provide some insights into developing convergence results and statistical properties for the estimators in the first stage of the scheme. The second-stage of the estimation scheme involves the the maximum likelihood estimation of the model parameters in regime-switching GARCH models with observable regimes given by the estimated states of the hidden Markov chain. There is some published work on asymptotic properties and convergence of the maximum likelihood estimates of, (regime-switching), GARCH models. Examples include Francq and Zakoian (2004), Bauwens, Preminger and Rombouts (2006) and Winker and Maringer (2009). These references provide some indication how to develop convergence results and statistical properties for the estimators in the second stage of the scheme.

# 5 Applications to Financial Risk Management: Simulated Data

In this section we discuss some applications of the switching GARCH (1, 1) model considered here to financial risk management using simulated data. In particular, we apply the switching GARCH (1, 1) model for evaluating Value at Risk, (VaR), and compare the VaR evaluated from the switching GARCH (1, 1) model with that from the GARCH (1, 1) model.

For the simulation study, we adopt the following values of the model parameters in the switching GARCH (1, 1) model:

$$\mu_1 = 0.11; \quad \mu_2 = -0.005; \quad \alpha_1 = 0.0073; \quad \alpha_2 = 0.11; \quad \beta_1 = 0.91; \quad \beta_2 = 1.45;$$
$$\theta_1 = 0.012; \quad \theta_2 = 0.43; \quad p_{11} = 0.9066; \quad p_{22} = 0.0917 \ ,$$

and in the GARCH(1,1) model:

$$\mu = 0.082; \quad \alpha = 0.022; \quad \beta = 0.85; \quad \theta = 0.15.$$

These parameters are consistent those used in Bauwens et al. (2006).

We first simulate the hidden Markov chain $\mathbf{X}$ over a time period with $T = 10,000$ using transition probabilities $p_{11}$ and $p_{22}$ as well as an initial probability $P(\mathbf{X}_0 = \mathbf{e}_1) = 0.5$. The simulated sequence of the hidden Markov chain is treated as if it were the "true" underlying sequence of the hidden Markov chain $\mathbf{X}$. Then given the simulated hidden Markov chain, we simulate the return data process $Y$ and the conditional variance process $h$ over a time period with $T = 10,000$. These simulated sequences of return data and conditional variances are also treated as if they were the "true" returns data and conditional variance process.

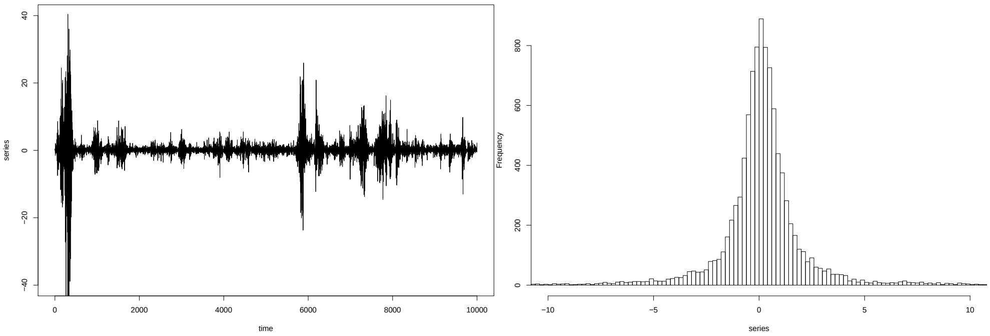


*Figure 1. trace plot [left] and histogram [right] of a series generated from the GARCH(1,1) Markov switching model*

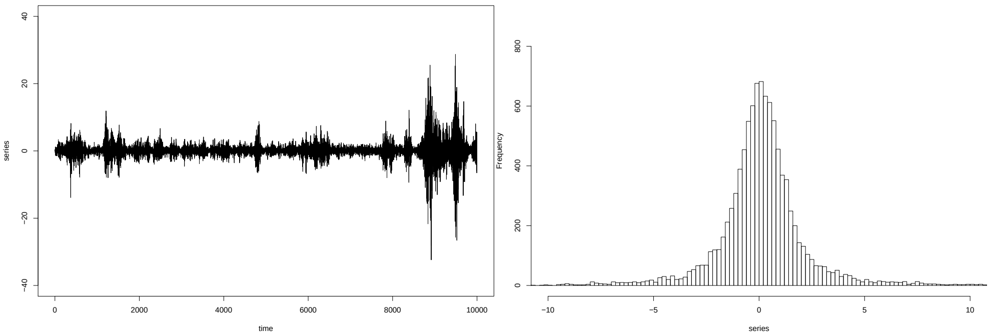


*Figure 2. trace plot [left] and histogram [right] of a series generated from the GARCH(1,1) model*

From the simulated time series of returns, it appears that the effect of volatility clustering is more significant in the GARCH (1, 1) model than in the switching GARCH (1, 1) model. By looking at the return distributions, it seems that the return distribution generated from the switching GARCH (1, 1) model has a heavier tail than that generated from the GARCH (1, 1) model. The additional amount of tail risk in the switching GARCH (1, 1) model may be attributed to the presence of the regime-switching effect.

| | 5% VAR | 1% VAR |
|---|---|---|
| MS-GARCH(1,1) | 3.0815 | 9.1205 |
| GARCH(1,1) | 3.0675 | 7.6644 |

*Table 1. Value at Risk for the two models*

From Table 1, we can see that the switching GARCH (1, 1) model gives a more prudent, or conservative, estimate for the VaR at each of the two probability levels than the GARCH (1, 1) model.

# 6 Conclusion

A two-stage method for estimating a switching GARCH model based on filtering theory was given. In the first stage, the hidden Markov chain was estimated using a filtering method and the Viterbi algorithm, while, in the second stage, the maximum likelihood method was used to estimate the model parameters given the estimated hidden Markov chain at the first stage. Applications of the switching GARCH (1, 1) model in financial risk management were discussed using simulated data.

**Acknowledgement:** The authors would like to thank the referee for helpful comments. Robert J. Elliott and Tak Kuen Siu wish to acknowledge the Discovery Grant from the Australian Research Council (ARC), (Project No.: DP1096243). Robert J. Elliott would also like to thank the ARC for support through grant DP0877639.

# References

[1] Bauwens, L., Preminger, A., Rombouts, J. V. K. (2006), Regime Switching GARCH Models. Working Paper. July, 2006.

[2] Bhar, R., and Chiarella, C. (1997). Estimation of the Heath-Jarrow-Morton Models by Use of Kalman Filtering Techniques. In *Computational Approaches to Economic Problems*, ed. H. Amman, B. Rustem and A. Whinston, Kluwer, The Netherlands, pp. 113-126.

[3] Chiarella, C., Pasquali, S., and Runggaldier, W. J. (2001). On Filtering in Markovian Term Structure Models: An Approximation Approach. *Advances in Applied Probability*, 33, 794 - 809.

[4] Chiarella, C., Hung, H., and To, T. (2005). The Volatility Structure of the Fixed Income Market under the HJM Framework: A Nonlinear Filtering Approach. Research Paper Series, 151, Quantitative Finance Research Centre, Universiy of Technology, Sydney, January 2005.

[5] Dufour, F., Elliott, R.J. and Tsoi, A. (1995). Asymptotic study of estimation in filtering for linear systems with jump parameters. IEEE Conference on Decision and Control. New Orleans, December 1995 IEEE Press, Piscataway, NJ, 3349-3353.

[6] Elliott, R. J., Hunter W. C., and Jamieson. B. M. (1998). Drift and volatility estimation in discrete time. *Journal of Economic Dynamics & Control*, 22, 209-218.

[7] Elliott, R. J. (1993). New finite dimentional filters and smoothers for noisly observed Markov chains. *IEEE Transactions on Information Theory*, 39, 265-271.

[8] Elliott, R. J. (1994). Exact adaptive filters for Markov chains observed in Gaussian noise. *Automatica,* 30, 1399-1408.

[9] Elliott, R.J. and Moore, J.B. (1997). Almost sure parameter estimation and convergence rates for Hidden Markov Models. *Systems and Control Letters*, 32, 203-207.

[10] Elliott, R. J., Aggoun, L. and Moore, J.B. (2008). *Hidden Markov Models: Estimation and Control*, 2nd edition, Springer-Verlag.

[11] Engle, R. R. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica,* 50, 987-1008.

[12] Francq, C. and Zakoian, J.-M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli*, 10(4), 605-637.

[13] Gray, S. (1996). Modeling the conditional distribution of interest rates as a regime switching process. *Journal of Financial Economics*, 42, 27-62.

[14] Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57, 354-384.

[15] Haas, M., Mittnik, S., and Paolella, M.S. (2004). A new approach to Markov switching GARCH models. *Journal of Financial Econometrics*, 2, 493-530.

[16] Pardoux, E., and Savona, C. (1988). Piecewise linear filtering. In *Stochastic Differential Systems, Stochastic Control Theory and Applications*, ed. W. Fleming and P.L. Lions, IMA Vol.10, Springer, pp. 429-437.

[17] Winker, P., and Maringer, D. (2009). The convergence of estimators based on heuristics: theory and application to a GARCH model. *Computational Statistics*, 24, 533-550.