Consistency of Maximum Likelihood Estimators for the Reduced Regime-Switching GARCH Model

Yingfu Xie and Jun Yu

Research Report
Centre of Biostochastics

Swedish University of Agricultural Sciences

Report 2005:2
ISSN 1651-8543

Consistency of Maximum Likelihood Estimators for the Reduced Regime-Switching GARCH Model

Yingfu Xie[1] and Jun Yu

Centre of Biostochastics
Swedish University of Agricultural Sciences
SE-901 83 Umeå, Sweden

Abstract

The regime-switching GARCH model combines the idea of Markov switching and GARCH model, which also extends Hidden Markov models. The statistical inference for this model, however, is rather difficult because the observations depend on the whole regime history. In this paper, we consider a reduced regime-switching GARCH model, that is, the past regimes are integrated out at every step and observations then depend only on the current regimes. We prove the consistency of maximum likelihood estimators for this model. Simulation studies to illustrate consistency, asymptotic normality of the proposed estimators and a model specification problem are also presented.

Keywords: GARCH model, regime-switching, MLE, consistency, asymptotic normality

[1]E-mail address to the correspondence author: yingfu.xie@sekon.slu.se

1 Introduction

A general hidden Markov model (HMM) is a discrete time model for dependent observations $\{Y_t\}_{t\in\mathbb{N}}$, where the dependence is modelled through an unobserved Markov chain $\{X_t\}$ such that the Y_t's are independent given $\{X_t\}$ and the distribution of Y_t depends on X_t only. Hidden Markov models have been popular in the last decade to model weakly dependent sequences in various fields such as speech processing, genetics and biochemistry, financial economics and so on. We refer to the monograph by MacDonald and Zucchini (1997) for comprehensive treatments and references therein.

The statistical inference of HMMs refers back to 1960's (Baum and Petrie, 1966; Petrie, 1969). In the former paper Baum and Petrie obtained the consistency and asymptotic normality of the maximum-likelihood estimator (MLE) when $\{Y_t\}$ take values in a finite set. For the general HMM, Leroux (1992) proved the consistency of the MLE under mild conditions. In a series of papers (Rydén, 1994; Bickel and Ritov, 1996; Bickel et al., 1998), consistent and (locally) asymptotically normal estimators were obtained. Their results were further extended to state space model (i.e. with continuous state space Markov chain) (Jensen and Petersen, 1999) and separable and compact state space, possibly non-stationary HMM (Douc and Matias, 2001).

The autoregressive conditional heteroscedastic (ARCH) model proposed by Engle (1982) (generalized later to GARCH by Bollerslev (1986)) has also been used widely in econometric society for last two decades to capture the time varying variance (see, among others, the review by Bollerslev et al., 1992). Combining the idea of HMM, Hamilton and Susmel (1994) and Cai (1994) proposed the regime-switching ARCH model, i.e., an ARCH model with parameters driven by an unobserved Markov chain with finite states (or regimes in econometric literature). The regime-switching GARCH model has also been considered by Gray (1996), Francq et al. (2001) and Klaassen (2002). Their empirical results show that the regime-switching model outperforms the ordinary (single regime) GARCH model in model interpretation and forecasting. In addition, the strong persistence usually observed in GARCH model (see, Engle et al., 1987, 1990, among others) can be explained by regime-switching.

Concerning the theoretical results of the regime-switching (G)ARCH model, Cai (1994) gave a sufficient and necessary condition for the stationarity of a regime-switching ARCH model. This result was extended to GARCH case by Francq et al. (2001), where they also proved the consistency of the maximum likelihood estimator (MLE) in ARCH case. However, by introducing the GARCH component into the model, a likelihood will depend on the entire regime path due to the recursive structure of GARCH equation (see

Equations (2) and (3) below). This causes an likelihood intractable quickly when the length of observations increases.

In order to overcome the regime path dependence problem while preserving the essential taste of GARCH model, Gray (1996) proposed a reduced regime-switching GARCH(1,1) model in the framework of a generalized regime-switching model. In the reduced model, the past regimes are integrated out at every step and the observations then depend only on the current regime. In this paper we will consider the reduced regime-switching GARCH(p, q) model. The consistency of the MLE of this model is proved.

In our simulation study, we will illustrate the consistency and asymptotic normality of MLE for our model. In another numerical simulation, we are interested in the persistence of model if an ordinary GARCH model is wrongly specified to data generated from regime-switching GARCH models.

Our paper is organized as follows. Section 2 formulates the model and the likelihood. We prove the consistency of MLE in Section 3. The simulation studies are presented in Section 4. The paper ends with some discussions in Section 5.

2 The model and likelihood

The general Regime-switching GARCH(p, q) process $\{Y_t\}_{t\in\mathbb{Z}}$ satisfies

$$Y_t = (h_t)^{1/2}\eta_t,$$
$$h_t = \omega(X_t) + \sum_{i=1}^{q} \alpha_i(X_t)Y_{t-i}^2 + \sum_{j=1}^{p} \beta_j(X_t)h_{t-j}, \tag{1}$$

where $\{\eta_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables with zero mean and unit variance, h_t is the conditional variance of Y_t given information up to time $t-1$, and $\{X_t\}$ is a Markov chain with finite state space $\mathbf{E} = \{1, 2, ..., d\}$. The standard assumptions for the model include:

(M1) $\{\eta_t\}$ and $\{X_t\}$ are independent;

(M2) Parameters $\alpha_i(s), 1 \le i \le q$, and $\beta_j(s), 1 \le j \le p$ are nonnegative and $\omega(s)$ are positive, given $\{X_t = s\}$, $s \in \mathbf{E}$, in order that the conditional variances h_t are almost surely strictly positive;

(M3) The Markov chain $\{X_t\}$ is irreducible and aperiodic, hence stationary.

Denote the stationary distribution as $\pi(s) := P(X_1 = s), 1 \le s \le d$ and transition probabilities as $p(k, l) := P(X_t = l|X_{t-1} = k)$. Note that $\pi(s) > 0, s \in \mathbf{E}$ under assumptions.

In this model, sequence $\{Y_t\}$ is not conditionally independent even given the Markov chain $\{X_t\}$, which is the key difference from general HMMs. The Markov chain $\{X_t\}$ is unobservable and its transition probabilities and stationary distribution are unknown. Our aim is to draw statistical inference only based on the observed $\{Y_t\}$.

Now, suppose that we are given a realization $\{y_1, ..., y_n\}$, while the noise η_t and the Markov chain X_t are unobservable. The parameters to be estimated from y_t are usually chosen to be

$$\boldsymbol{\theta} := \{p(k,l), \omega(s), \alpha_i(s), \beta_j(s), k \neq l, 1 \leq k, l, s \leq d, 1 \leq i \leq q, 1 \leq j \leq p\},$$

which contains the transition probabilities of the Markov chain and parameters of the GARCH equation (1). The stationary distribution $\pi(s), 1 \leq s \leq d$ are not included since asymptotically the stationary distribution will not affect the estimation (see equation (11) below). We assume that p, q and d are known and will not discuss order selection in this paper. Readers who are interested in this issue are referred to Rydén (1995) for hidden Markov models and Francq et al. (2001) for regime-switching ARCH models.

In this paper, we consider the MLE for regime-switching GARCH models and its asymptotic properties. Assume that η_t is standard Gaussian. From (1), by summing up the probability density over all possible paths of the Markov chain, we get the likelihood function (x_t denotes the value of X_t)

$$p_{\boldsymbol{\theta}}(y_1, ..., y_n) = \sum_{(x_1,...,x_n)\in\mathbf{E}^n} \pi(x_1) \left\{ \prod_{t=2}^{n} p(x_{t-1}, x_t) \right\} \left\{ \prod_{t=1}^{n} f_{x_1,...,x_t}(y_1, ..., y_t) \right\}, \tag{2}$$

where

$$f_{x_1,...,x_t}(y_1, ..., y_t) = \frac{1}{(2\pi h_{x_1,...,x_t}(y_1, ..., y_{t-1}))^{1/2}} \exp\{-\frac{y_t^2}{2h_{x_1,...,x_t}(y_1, ..., y_{t-1})}\},$$

and the conditional variance process follows

$$\begin{aligned} h_{x_1,...,x_t}(y_1, ..., y_{t-1}) &= \omega(x_t) + \sum_{i=1}^{q} \alpha_i(x_t) y_{t-i}^2 \\ &\quad + \sum_{j=1}^{p} \beta_j(x_t) h_{x_1,...,x_{t-j}}(y_1, ..., y_{t-j-1}), \end{aligned} \tag{3}$$

starting with

$$h_{x_1} = \omega(x_1), \quad h_{x_1,x_2}(y_1) = \omega(x_2) + \alpha_1(x_2) y_1^2 + \beta_1(x_2) h_{x_1}$$

and continuing recursively.

From (2), it follows, which was also pointed out by Cai (1994), Hamilton and Susmel (1994) and Francq et al. (2001) among others, that the likelihood becomes intractable very quickly as n increases, mainly because of the recursive structure in (3)— $h_{x_1,...,x_t}$ depends on the whole regime path through $h_{x_1,...,x_{t-j}}$ and further on. Because the number of possible regime paths grows exponentially with t, this leads to an enormous number of addends in (2).

To avoid the path dependence problem, we propose a reduced regime-switching GARCH(p,q) model, inspired from Gray (1996). We replace equation (1) by

$$Y_t = (h_t)^{1/2}\eta_t,$$

$$h_t = \omega(X_t) + \sum_{i=1}^{q} \alpha_i(X_t) Y_{t-i}^2 + E_{\tilde{X}_{t-1}} \left[\sum_{j=1}^{p} \beta_j(X_t) h_{t-j} \right], \qquad (4)$$

where the expectation is across the regime path $\tilde{X}_{t-1} := \{X_{t-1}, X_{t-2}, ...\}$, conditional on available information. (M1), (M2) and (M3) are assumed for this model. Note that actually we only need to integrate out the single regime X_{t-1} at time point t since recursively h_{t-1} is already independent of $\tilde{X}_{t-2}$. As Gray (1996) illustrated, for $p = q = 1$, since Y_t (Δr_t in his context) was essentially a mixture of distributions with respect to different regimes (with time-varying mixing parameters), it was natural to consider to take expectation of individual conditional variances over regimes. This averaged variance was used as the lagged conditional variance in constructing conditional variance of next time period, and the path dependence problem was overcome while the essential nature of GARCH process was reserved. Here we extend this idea to GARCH(p,q) model and use p previous averaged conditional variance in the GARCH equation (4).

Now, the likelihood for this model can be written as

$$p_{\boldsymbol{\theta}}(y_1, ..., y_n) = \sum_{(x_1,...,x_n)\in \mathbf{E}^n} \pi(x_1) \left\{ \prod_{t=2}^{n} p(x_{t-1}, x_t) \right\} \left\{ \prod_{t=1}^{n} f_{x_t}(y_1, ..., y_t) \right\}, \quad (5)$$

where $f_{x_t}(.)$ only depends on the current regime x_t and the conditional variance entering $f_{x_t}(.)$ is (4) instead of (3). Similar to the case of general HMM (MacDonald and Zucchini, 1997, pp.78-79), we can write (5) as a product of matrices due to this simplified structure.

Define vector $\mathbf{1} = (1, ..., 1)^T \in \mathbb{R}^d$, $\mathbf{p} = (\pi(1) f_1(y_1), ..., \pi(d) f_d(y_1))^T \in \mathbb{R}^d$ and matrix

$M_{\boldsymbol{\theta}}(y_1, ..., y_t)$

$$= \begin{pmatrix} p(1,1)f_1(y_1,...,y_t) & p(2,1)f_1(y_1,...,y_t) & ... & p(d,1)f_1(y_1,...,y_t) \\ p(1,2)f_2(y_1,...,y_t) & p(2,2)f_2(y_1,...,y_t) & ... & p(d,2)f_2(y_1,...,y_t) \\ ... & ... & ... & \\ p(1,d)f_d(y_1,...,y_t) & p(2,d)f_d(y_1,...,y_t) & ... & p(d,d)f_d(y_1,...,y_t) \end{pmatrix},$$

then the likelihood can be written as

$$p_{\boldsymbol{\theta}}(y_1, ..., y_n) = \mathbf{1}^T \left\{ \prod_{t=2}^{n} M_{\boldsymbol{\theta}}(y_1, ..., y_t) \right\} \mathbf{p}, \tag{6}$$

which can be computed numerically, by using, e.g., the so-called forward-backward algorithm.

To calculate $f_s(y_1, ..., y_t), s = 1, ..., d$, write the conditional variance as $h_t(s)$ given $X_t = s$, since from (4) it depends on the regimes only through X_t. Then

$$f_s(y_1, ..., y_t) = (2\pi h_t(s))^{-1/2} \exp\{-y_t^2/(2h_t(s))\},$$

and

$$h_t(s) = \omega(s) + \sum_{j=1}^{q} \alpha_j(s) y_{t-j}^2 + \sum_{j=1}^{p} \beta_j(s) h'_{t-j},$$

where h'_{t-j} denotes the conditional variances that have been taken expectation over all regime path and are path-independent. Recall that we only need to integrate out one regime at each step, for example, X_{t-1} for h'_t due to the recursive structure. Let $\lambda_{st} = P(X_t = s|I_{t-1})$. Then we have $h'_t = \sum_{s=1}^{d} \lambda_{st} h_t(s)$. Making use of Bayesian rule, it is not difficult to obtain (cf. Gray (1996))

$$\lambda_{st} = \sum_{j=1}^{d} p(j, s) \frac{f_j(y_1, ..., y_{t-1})\lambda_{j,t-1}}{\sum_{k=1}^{d} f_k(y_1, ..., y_{t-1})\lambda_{k,t-1}},$$

which is of the recursive form too.

3 Consistency of the MLEs

In this section, we will give conditions and the main theorem for the consistency of MLEs for the reduced regime-switching GARCH model. Our method is benefited from Francq et al. (2001).

First, we assume that

(A1) $\{Y_t\}_{t\in\mathbb{Z}}$ in model (4) is strictly stationary and ergodic. In addition, the unconditional variance of Y_t is finite.

Following Brandt (1986) and Bougerol and Picard (1992), Francq et al. (2001, Theorem 2) gave a sufficient and necessary condition for the existence of a second-order stationarity solution of (1). However, their result does not hold for our model (4). It is of interest to find a similar condition for the reduced regime-switching GARCH model and we leave it as a separate work.

The parameter space Θ is defined as the subset of $\mathbb{R}^{d^2+pd+qd}$ of the parameters satisfying (M2), (M3) and (A1). Assume also that the true parameter value $\boldsymbol{\theta}_0$ belongs to Θ.

Because the indices of the states of the Markov chain can be permuted without changing the law of the model, the parameters are not strictly identifiable up to permutation. Hence, it is necessary to introduce a condition for identifiability. Define $p_{\boldsymbol{\theta}}(Y_t|Y_{t-1}, Y_{t-2}, ...)$ as the conditional density of Y_t given all previous observations, $\{Y_{t-1}, Y_{t-2}, ...\}$, and $p_{\boldsymbol{\theta}}(Y_t|Y_{t-1}, ..., Y_1)$ given $\{Y_{t-1}, Y_{t-2}, ..., Y_1\}$. When $t = 1$, $p_{\boldsymbol{\theta}}(Y_t|Y_{t-1}, ..., Y_1)$ becomes $p_{\boldsymbol{\theta}}(Y_1)$, the unconditional density of Y_1. Let $g_{\boldsymbol{\theta}}(Y_t|Y_{t-1}, Y_{t-2}, ...)$ and $g_{\boldsymbol{\theta}}(Y_t|Y_{t-1}, ..., Y_1)$ be the corresponding logarithms. We also define the conditional likelihood function based on all observations from infinite past, $p_{\boldsymbol{\theta}}(y_1, ..., y_n|y_0, y_{-1}, ...)$ similar to (5).

(A2) *Identifiability Condition*: For any $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2} \in \Theta$ and all $Y_t, Y_{t-1}, ...$, if $p_{\boldsymbol{\theta_1}}(Y_t|Y_{t-1}, Y_{t-2}, ...) = p_{\boldsymbol{\theta}_2}(Y_t|\ Y_{t-1}, Y_{t-2}, ...)$, $P_{\boldsymbol{\theta}_0} - a.s.$, then $\boldsymbol{\theta_1} = \boldsymbol{\theta}_2$.

After describing all assumptions, we are now in a position to state our main theorem.

Theorem 1 *For the reduced regime-switching GARCH model (4), assume (A1) and (A2). Suppose that Θ^* is a compact subset of Θ and $\boldsymbol{\theta}_0 \in \Theta^*$; $(\hat{\boldsymbol{\theta}}_n)$ is an MLE sequence, satisfying almost surely*

$$p_{\hat{\boldsymbol{\theta}}_n}(Y_1, ..., Y_n) = \sup_{\boldsymbol{\theta}\in\Theta^*} p_{\boldsymbol{\theta}}(Y_1, ..., Y_n) \quad \forall\ n.$$

Then $(\hat{\boldsymbol{\theta}}_n)$ tends almost surely to $\boldsymbol{\theta}_0$ as $n \to \infty$.

First, we will give some lemmas. For a general strictly stationary and ergodic sequence $\{Z_t\}_{t\in\mathbb{Z}}$, suppose that the density of Z_t depends on parameter $\boldsymbol{\psi}$ in some Euclidean space Ψ and let the true parameter be $\boldsymbol{\psi_0}$.

Lemma 1 *For the sequence* $\{Z_t\}_{t\in\mathbb{Z}}$*, the likelihood of* $(Z_1, ..., Z_n)$ *is asymptotically equivalent to the likelihood conditioning observations from infinite past with probability one, i.e., for all* $\boldsymbol{\psi} \in \Psi$*, with probability one we have*

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} \log p_{\boldsymbol{\psi}}(Z_1, ..., Z_n) &= \lim_{n\to\infty} \frac{1}{n} \log p_{\boldsymbol{\psi}}(Z_1, ..., Z_n|Z_0, Z_{-1}, ...) \\ &= E_{\boldsymbol{\psi}_0} g_{\boldsymbol{\psi}}(Z_t|Z_{t-1}, ...), \end{aligned} \tag{7}$$

provided that

$$E_{\boldsymbol{\psi_0}}|g_{\boldsymbol{\psi}}(Z_t|Z_{t-1}, Z_{t-2}, ...)| < \infty, \quad \forall \boldsymbol{\psi} \in \Psi, \tag{8}$$

where $g_{\boldsymbol{\psi}}(Z_t|Z_{t-1}, Z_{t-2}, ...)$ *is similarly defined as* $g_{\boldsymbol{\theta}}(Y_t|Y_{t-1}, Y_{t-2}, ...)$ *for* $\{Y_t\}$.

Proof. First, note that

$$\log p_{\boldsymbol{\psi}}(Z_1, ..., Z_n|Z_0, Z_{-1}, ...) = \sum_{t=1}^{n} g_{\boldsymbol{\psi}}(Z_t|Z_{t-1}, Z_{t-2}, ...) \tag{9}$$

and

$$\log p_{\boldsymbol{\psi}}(Z_1, ..., Z_n) = \sum_{t=1}^{n} g_{\boldsymbol{\psi}}(Z_t|Z_{t-1}, ..., Z_1).$$

Write

$$\begin{aligned}&\frac{1}{n}\sum_{t=1}^{n} g_{\boldsymbol{\psi}}(Z_t|Z_{t-1}, ..., Z_1) \\ &= \frac{1}{n}\sum_{t=1}^{n} g_{\boldsymbol{\psi}}(Z_t|Z_{t-1}, Z_{t-2}, ...) \\ &\quad + \frac{1}{n}\sum_{t=1}^{n} \{g_{\boldsymbol{\psi}}(Z_t|Z_{t-1}, ..., Z_1) - g_{\boldsymbol{\psi}}(Z_t|Z_{t-1}, Z_{t-2}, ...)\}. \end{aligned} \tag{10}$$

Analogous to Karlin and Taylor (1975, pp.502), define

$$\phi_N(z_0, z_{-1}, ...) = \sup_{l\geq N} |g_{\boldsymbol{\psi}}(z_0|z_{-1}, ..., z_{-l}) - g_{\boldsymbol{\psi}}(z_0|z_{-1}, z_{-2}, ...)|,$$

and

$$Z_t^N = \phi_N(Z_t, Z_{t-1}, ...).$$

Under assumptions on $\{Z_t\}$, $\{Z_t^N\}$ is stationary, ergodic, and $E[|Z_t^N|] < \infty$. We have

$$\begin{aligned}
&\limsup_{n\to\infty}\left|\frac{1}{n}\sum_{t=1}^{n}\{g_{\boldsymbol{\psi}}(Z_t|Z_{t-1},...,Z_1) - g_{\boldsymbol{\psi}}(Z_t|Z_{t-1},Z_{t-2},...)\}\right| \\
\leq\ & \limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\left|g_{\boldsymbol{\psi}}(Z_t|Z_{t-1},...,Z_1) - g_{\boldsymbol{\psi}}(Z_t|Z_{t-1},Z_{t-2},...)\right| \\
\leq\ & \limsup_{n\to\infty}\frac{1}{n}\sum_{t=N+1}^{n} Z_t^N = E[Z_1^N].
\end{aligned}$$

But as $N \to \infty$, $Z_1^N \to 0$, and the interchange of limit and expectation can be justified to give $\lim_{N\to\infty} E[Z_1^N] = 0$. So the second term on the right-hand side of (10) goes to zero as $n \to \infty$. And the second equality in (7) follows by applying ergodic theorem, which completes the proof. □

We will next compare the likelihood $p_{\boldsymbol{\psi}}(Z_1,...,Z_n)$ with the one evaluated at the true parameter $\boldsymbol{\psi}_0$, $p_{\boldsymbol{\psi}_0}(Z_1,...,Z_n)$. Define

$$O_n(\boldsymbol{\psi}) = \frac{1}{n}\log\frac{p_{\boldsymbol{\psi}}(Z_1,...,Z_n)}{p_{\boldsymbol{\psi}_0}(Z_1,...,Z_n)},$$

and we have the following lemma.

Lemma 2 *For the sequence $\{Z_t\}$ and parameter space Ψ, in addition to (8), assume also that the identifiability condition (A2) holds. Then for all $\boldsymbol{\psi} \in \Psi$, with probability one,*

$$\lim_{n\to\infty} O_n(\boldsymbol{\psi}) \leq 0$$

and the limit is almost surely equal to zero if and only if $\boldsymbol{\psi} = \boldsymbol{\psi}_0$.

Proof. From Lemma 1 and Jensen's inequality,

$$\begin{aligned}
\lim_{n\to\infty} O_n(\boldsymbol{\psi}) &= E_{\boldsymbol{\psi}_0}\log\frac{p_{\boldsymbol{\psi}}(Z_t|Z_{t-1},Z_{t-2},...)}{p_{\boldsymbol{\psi}_0}(Z_t|Z_{t-1},Z_{t-2},...)} \\
&\leq \log E_{\boldsymbol{\psi}_0}\frac{p_{\boldsymbol{\psi}}(Z_t|Z_{t-1},Z_{t-2},...)}{p_{\boldsymbol{\psi}_0}(Z_t|Z_{t-1},Z_{t-2},...)} = 0.
\end{aligned}$$

By the identifiability condition, this limit equals to zero if and only if $\boldsymbol{\psi} = \boldsymbol{\psi}_0$. □

Now consider the reduced regime-switching GARCH model (4). (8) is satisfied for our model under (A1) and standard Gaussian η_t since the conditional distribution in (8) is now just a mixture of normal. Hence Lemma 1 and Lemma 2 hold for $\{Y_t\}$ and Θ. By similarly defining

$$O_n(\boldsymbol{\theta}) = \frac{1}{n} \log \frac{p_{\boldsymbol{\theta}}(Y_1, ..., Y_n)}{p_{\boldsymbol{\theta}_0}(Y_1, ..., Y_n)},$$

we have

Lemma 3 *For the reduced regime-switching GARCH model (4), assume (A1) and (A2). Then, for any $\boldsymbol{\theta}_1 \in \Theta$, $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_0$, there exists a neighborhood $V(\boldsymbol{\theta}_1)$ of $\boldsymbol{\theta}_1$ such that*

$$\limsup_{n\to\infty} \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_1)} O_n(\boldsymbol{\theta}) < 0 \quad \text{a.s.}$$

Proof. First, define the matrix norm $\| \, . \, \|$ as the sum of all element of the matrix. Making use of the product of matrices (6), we have

$$\begin{aligned} \min_j \pi(j) f_j(Y_1) \left\| \prod_{t=2}^{n} M_{\boldsymbol{\theta}}(Y_1, ..., Y_t) \right\| &\leq p_{\boldsymbol{\theta}}(Y_1, ..., Y_n) \\ = \left\| \left\{ \prod_{t=2}^{n} M_{\boldsymbol{\theta}}(Y_1, ..., Y_t) \right\} \mathbf{p} \right\| &\leq \max_j \pi(j) f_j(Y_1) \left\| \prod_{t=2}^{n} M_{\boldsymbol{\theta}}(Y_1, ..., Y_t) \right\|. \end{aligned}$$

Hence, it is straightforward to obtain

$$\lim_{n\to\infty} \frac{1}{n} \log p_{\boldsymbol{\theta}}(Y_1, ..., Y_n) = \lim_{n\to\infty} \frac{1}{n} \log \left\| \prod_{t=2}^{n} M_{\boldsymbol{\theta}}(Y_1, ..., Y_t) \right\|. \tag{11}$$

Next, let $V_r(\boldsymbol{\theta}_1)$ be the open sphere with center $\boldsymbol{\theta}_1$ and radius $1/r$ and define

$$S_{2,n}^r = \sup_{\boldsymbol{\theta}\in V_r(\boldsymbol{\theta}_1)} \left\| \prod_{t=2}^{n} M_{\boldsymbol{\theta}}(Y_1, ..., Y_t) \right\|.$$

Because this matrix norm is multiplicative, we have, for $\boldsymbol{\theta} \in V_r(\boldsymbol{\theta}_1)$

$$\sup_{\boldsymbol{\theta}} \left\| \prod_{t=2}^{n+k} M_{\boldsymbol{\theta}}(Y_1, ..., Y_t) \right\| \leq \sup_{\boldsymbol{\theta}} \left\| \prod_{t=2}^{n} M_{\boldsymbol{\theta}}(Y_1, ..., Y_t) \right\| \cdot \sup_{\boldsymbol{\theta}} \left\| \prod_{t=n+1}^{n+k} M_{\boldsymbol{\theta}}(Y_1, ..., Y_t) \right\|,$$

that is

$$\log S_{2,n+k}^r(Y_1, ..., Y_{n+k}) \leq \log S_{2,n}^r(Y_1, ..., Y_n) + \log S_{n+1,n+k}^r(Y_1, ..., Y_{n+k})$$

for any positive integers $n(\geq 2)$, k and r. From the Kingman's ergodic theorem (1973) for the subadditive processes we can readily get

$$\lim_{n\to\infty}\frac{1}{n}\log S^r_{2,n}(Y_1,...,Y_n)=\lambda_r(\boldsymbol{\theta}_1):=\inf_{n>1}\frac{1}{n}E_{\boldsymbol{\theta}_0}\log S^r_{2,n}(Y_1,...,Y_n),\ P_{\boldsymbol{\theta}_0}-a.s.$$

Recall that almost surely

$$\begin{aligned}\lambda(\boldsymbol{\theta}_0) &:= \lim_{n\to\infty}\frac{1}{n}\log\left\|\prod_{t=2}^{n}M_{\boldsymbol{\theta}_0}(Y_1,...,Y_t)\right\| \\ &= \inf_{n>1}\frac{1}{n}E_{\boldsymbol{\theta}_0}\log\left\|\prod_{t=2}^{n}M_{\boldsymbol{\theta}_0}(Y_1,...,Y_t)\right\|.\end{aligned}$$

Thus, from Lemma 2 and (11), there exists $\varepsilon>0$ and $n_\varepsilon\in\mathbb{N}$ such that

$$\frac{1}{n_\varepsilon}E_{\boldsymbol{\theta}_0}\log\left\|\prod_{t=2}^{n_\varepsilon}M_{\boldsymbol{\theta}_1}(Y_1,...,Y_t)\right\|<\lambda(\boldsymbol{\theta}_0)-\varepsilon.$$

The dominated convergence theorem shows that, for r large enough,

$$\lambda_r(\boldsymbol{\theta}_1)\leq\frac{1}{n_\varepsilon}E_{\boldsymbol{\theta}_0}\log S^r_{2,n_\varepsilon}(Y_1,...,Y_{n_\varepsilon})<\lambda(\boldsymbol{\theta}_0)-\frac{\varepsilon}{2}.$$

□

Proof of Theorem 1. Suppose that $\hat{\boldsymbol{\theta}}_n$ didn't tend to $\boldsymbol{\theta}_0$ with probability one as $n\to\infty$, i.e., for arbitrarily large integer M, there exist a $\delta>0$ and at least one $n^*, n^*\geq M$ such that $|\hat{\boldsymbol{\theta}}_{n^*}-\boldsymbol{\theta}_0|>\delta$ with positive probability. By Lemma 3, it follows that $p_{\hat{\boldsymbol{\theta}}_{n^*}}(Y_1,...,Y_{n^*})$ is strictly less than $p_{\boldsymbol{\theta_0}}(Y_1,...,Y_{n^*})$ with positive probability. However, with probability one, we have

$$p_{\hat{\boldsymbol{\theta}}_n}(Y_1,...,Y_n)=\sup_{\boldsymbol{\theta}\in\Theta^*}p_{\boldsymbol{\theta}}(Y_1,...,Y_n)\geq p_{\boldsymbol{\theta_0}}(Y_1,...,Y_n)$$

for all n. The contradiction gives our result. □

4 Simulation studies

In our simulation studies we will illustrate the consistency and asymptotic normality of the MLE of model (4), and present a model specification problem. The most commonly used two-regime switching model is considered. During the estimation, in addition to the standard constraints $0\leq p(k,l)\leq 1$, $\sum_{l=2}^{d}p(k,l)\leq 1$, $\alpha_i(s)\geq 0$, $\beta_j(s)\geq 0, 1\leq i\leq q, 1\leq j\leq p, 1\leq$

$k, l, s \leq d$ and $\sum_{i=1}^{q} \alpha_i(s) + \sum_{j=1}^{p} \beta_j(s) \leq 1, 1 \leq s \leq d$, we also assume $\omega(s) \geq 0.001, 1 \leq s \leq d$ to avoid the underflow of the numerical algorithm, and $\omega(1) \leq \omega(2) \leq ... \leq \omega(d)$ for identifiability. Several randomly chosen starting point set are used to pursue the global maxima.

4.1 Consistency and asymptotic normality

In the first experiment, 100 independent series from model (4), each with length 10000, are generated. The true model has two regimes with parameters : $\omega(1) = 1, \alpha_1(1) = 0.4$, $\beta_1(1) = 0.2$, and $\omega(2) = 20, \alpha_1(2) = 0.2$, $\beta_1(2) = 0.4$. The transition probabilities between these two regimes are set to be $p(1, 2) = p(2, 1) = 0.1$. These parameters are chosen such that they are fairly away from the boundary of parameter space.

Table 1: The mean, bias and standard deviation of MLE for model (4)

	$\hat{\omega}(1)$	$\hat{\alpha}_1(1)$	$\hat{\beta}_1(1)$	$\hat{\omega}(2)$	$\hat{\alpha}_1(2)$	$\hat{\beta}_1(2)$	$\hat{p}(1,2)$	$\hat{p}(2,1)$
mean	0.959	0.406	0.205	20.329	0.192	0.4004	0.0997	0.102
bias	0.041	-0.006	0.005	-0.329	0.008	-0.0004	0.0003	-0.002
std.	0.259	0.044	0.03	1.808	0.030	0.057	0.009	0.014

Table 1 summarizes the mean, bias and standard deviation of the MLE for the reduced regime-switching GARCH model. The biases are all well close to zero, which illustrates the result for consistency.

Next we want to shed a few light on the asymptotic normality issue. Bickel et al. (1998) proved the asymptotic normality of MLE for general HMM model. Their results, however, cannot be readily extended to regime-switching GARCH model. We will first consider the case where $p = 0$ in model (4), i.e., no GARCH effects exist. It is worth noting that by taking away the GARCH effect, our model is same as the one Francq et al. (2001) considered. Douc et al. (2004) showed the normality of MLE for the regime-switching ARCH model in a more general autoregressive model, making use of the uniform exponential forgetting of the initial distribution for the hidden Markov chain conditional on the observations (see Douc et al. (2004) for details).

We obtain MLEs for 100 independent series, each with length 10000, from model (4) with the same parameter values as Francq et al. (2001), i.e., $p = 0$, $\omega(1) = 1$, $\omega(2) = 20$, $\alpha_1(1) = 0.4$, $\alpha_1(2) = 0.2$ and $p(1, 2) = p(2, 1) = 0.1$. The Quantile-Quantile (QQ) plots together with 0-1 line are presented in Figure 1. The P-values of ks.gof (Kolmogorov-Smirnov goodness-of-fit) test with null hypothesis normal are 0.804, 1.097, 0.15, 0.445, 1.049, and 0.64 for estimators of $\omega(1)$, $\alpha_1(1)$, $\omega(2)$, $\alpha_1(2)$, $p(1,2)$ and $p(2,1)$ respectively.(The P-value is calculated using the approximation of Dallal and Wilkinson (1986), which is most accurate for P-value less than 0.1. So these P-values are all set to 0.5 in statistical software S-plus®, with which we carry out our analysis.)

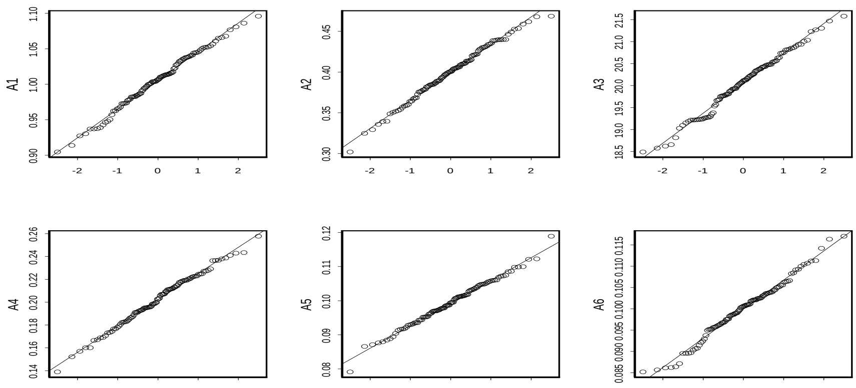

Figure 1: The QQ-plot of MLE for regime-switching ARCH model with 0-1 line: A1–A6 represent estimators of $\omega(1)$, $\alpha_1(1)$, $\omega(2)$, $\alpha_1(2)$, $p(1,2)$ and $p(2,1)$, respectively.

For the asymptotic normality of the reduced regime-switching GARCH model, we adopt the same settings as the first experiment for consistency. The QQ-plots with 0-1 line are shown in Figure 2, where the P-values of ks.gof test are 0.458, 0.587, 0.161, 0.609, 0.593, 0.128, 0.839 and 1.018 respectively.

4.2 A model specification problem

From existing literature on regime-switching model, we are seemingly impressed that if we specify an ordinary GARCH model to sequences generated from regime-switching GARCH models, the obtained estimates are typically close to non-stationary region, i.e., the sum of GARCH parameters is larger than or equal to unity ($\sum_{j=1}^{q}\alpha_j + \sum_{k=1}^{p}\beta_k \simeq 1$). For instance, Francq et al. (2001, pp.198) stated that '..., for a series with at least two GARCH regimes, the standard GARCH parameter estimates are generally explosive (..., and) close to the nonstationary region.' However, we observe here that this statement depends crucially on the scale of transition probabilities between regimes.

We first adopt the parameter values used by Francq et al. (2001): $\omega(1) = 1$, $\alpha_1(1) = 0.6$, $\omega(2) = 100$, $\alpha_1(2) = 0$, $p(1,2) = p(2,1) = 0.01$. Then we change the pair of transition probabilities $(p(1,2), p(2,1))$ to (0.1, 0.1), (0.2, 0.2) and (0.5, 0.5). From each parameter set, we generate 100 independent

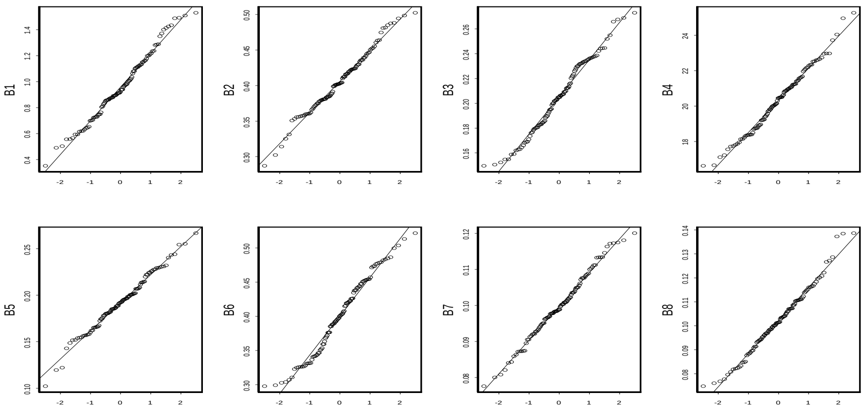

Figure 2: The QQ-plot of MLE for regime-switching GARCH model with 0-1 line: B1–B8 represent estimators of $\omega(1)$, $\omega(2)$, $\alpha_1(1)$, $\beta_1(1)$, $\alpha_1(2)$, $\beta_1(2)$, $p(1,2)$ and $p(2,1)$, respectively.

Table 2: Estimated parameters of standard GARCH(1,1) model fitted to two-regime ARCH(1) models with different transition probabilities, where ω, α and β are model parameters of standard GARCH(1,1) model.

$(p(1,2), p(2,1))$	(0.01, 0.01)	(0.1, 0.1)	(0.2, 0.2)	(0.5,0.5)
$\hat{\omega}$	0.9907	8.9981	24.294	46.9367
$\hat{\alpha}+\hat{\beta}$	0.9991	0.9458	0.6717	0.3505

series of length 1000 and apply to them the standard GARCH(1,1) model. The estimate $\hat{\omega}$ and sum of estimated GARCH parameters are summarized in Table 2.

We can see clearly from Table 2 that when $p(1,2)$ and $p(2,1)$ are small, it is true that the estimates are close to non-stationary region. In fact, there are 97 estimate with $\hat{\alpha} + \hat{\beta}$ equal exactly to unity. If we relax the restriction that $\alpha + \beta \leq 1$, the sum will be generally large than unity. However, when the transition probabilities increase, they move more and more away from the non-stationary region, to, maybe not so surprising, seemingly the average of these two regimes. Note that this result still holds with other settings of parameters (other than transition probabilities), higher order GARCH models or longer sequence. Table 3 gives us similar result when the data are from two-regime switching GARCH model. A possible explanation is that when the transition probabilities are large, the average effect of these regimes dominates

Table 3: Estimated parameters of standard GARCH(1,1) model fitted to two-regime GARCH(1,1) models with different transition probabilities

$(p(1,2), p(2,1))$	(0.01, 0.01)	(0.1, 0.1)	(0.2, 0.2)	(0.5, 0.5)
$\hat{\omega}$	0.6012	4.412	7.227	11.0346
$\hat{\alpha}+\hat{\beta}$	0.9803	0.7452	0.5703	0.3709

the sequence. Then it will be more stationary and the parameters tend to mean of parameters of all regimes. However, while regimes seldom switch to each other, there is significant structure change in the sequence and the sequence is more likely to be treated as non-stationary.

5 Discussions

The regime-switching GARCH model can be considered as an extension of HMM with dependent conditional distribution of $\{Y_t\}$ given the regimes $\{X_t\}$. One problem with this extension is that the observations at any time point will then depend on the whole regime path due to the recursive structure of GARCH equation. It causes an intractable likelihood for the full model and the statistical inference is difficult. Francq et al. (2001) obtained the consistency of the MLE and some empirical results only in case that $p = 0$, i.e., taking away the GARCH part of the model. Cai (1994) and Hamilton and Susmel (1994) also only used regime-switching ARCH models. Gray (1996) proposed a reduced regime-switching GARCH(1,1) model, which integrated out the previous regime path at every step. It is hence tractable for estimation while keeps the parsimony style and preserves the essential nature of GARCH model. We generalized their model to GARCH(p, q) in this paper and proved the consistency of the MLE for this model. Lemma 1 itself is also of interests.

Another problem with the extension is that observations are then not independent given the regimes. Many results for HMM, e.g., the asymptotic normality of MLE, don't hold for the regime-switching GARCH model. The framework of Bickel et al. (1998) for asymptotic normality of HMM is not easy to extend to include this model. Douc et al. (2004) obtained the asymptotic normality for a class of autoregressive models with Markov regime, which includes HMM and regime-switching ARCH model as special cases. However, they instantly use the Markov property of $\{X_t, Y_t, ..., Y_{t-r+1}\}$ (assume an r-th order ARCH model), which doesn't hold for GARCH models. Frow our numerical experiments, we conjecture in this paper that the MLE of the reduced regime-switching GARCH model is asymptotically normal distributed, but we need to find other approach to theoretically prove it, which will be a challenging but rewarding task in the future.

In a model specification problem, we observed that wrongly specifying an ordinary GARCH model to series with (at least) two GARCH regimes will result in persistent, explosive estimates, if the transition probabilities between these regimes are quite small. However, regime-switching GARCH models (with larger transition probabilities) can also produce less persistent sequences. This implies that a GARCH model whose parameter is far from the non-stationary region could also come from a regime-switching model, which we didn't realize before and could lead to interesting applications.

Acknowledgements

The authors are grateful to Prof. Bo Ranneby for the helpful discussions and advices.

References

Baum, L.E. and Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Statist.* **37**, 1554-1563.

Bickel, P.J. and Ritov, Y. (1996). Inference in hidden Markov models I: local asymptotic normality in the stationary case. *Bernoulli* **2**, 199-228.

Bickel, P.J., Ritov, Y. and Rydén, T. (1998). Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. *Ann. Statist.* **26**, 1614-1635.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31**, 307-327.

Bollerslev, T., Chou, R.Y. and Kroner, K.F. (1992). ARCH modelling in finance. *J. Econometrics* **52**, 5-59.

Bougerol, P. and Picard, N. (1992). Strict stationary of generalized autoregressive processes. *Ann. Probab.* **20**, 1714-1729.

Brandt, A. (1986). The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Adv. Applied Probab.* **18**, 221-254.

Cai, J. (1994). A Markov model of switching-regime ARCH. *J. Busi. Econ. Statist.* **12**, 309-316.

Dallal, G.E. and Wilkinson, L. (1986). An analytic approximation to the distribution of Lilliefors's test statistic for normality. *The Amer. Statistician* **40**, 294-296.

Douc, R. and Matias, C. (2001). Asymptotics of the maximum likelihood estimator for general hidden Markov models. *Bernoulli* **7**, 381-420.

Douc, R., É. Moulines and Rydén, T. (2004). Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. *Ann. Statist.* **32**, 2254–2304.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of U.K. inflation. *Econometrica* **50**, 987-1008.

Engle, R.F., Lilien, D.M. and Robbins, R.P. (1987). Estimating time varying risk premia in the term structure: The ARCH-M model. *Econometrica* **55**, 391-407.

Engle, R.F., Ng, V. and Rothschild, M. (1990). Asset pricing with a factor ARCH covariance structure: Empirical estimates for treasury bills. *J. Econometrics* **45**, 213-238.

Francq, C., Roussignol, M. and Zakoïan, J. (2001). Conditional heteroskedasticity driven by hidden Markov chains. *J. Time Series Analy.* **22**, 197-220.

Gray, S.F. (1996). Modelling the conditional distribution of interest rates as a regime-switching process. *J. Financial Econ.* **42**, 27-62.

Hamilton, J.D. and Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *J. Econometrics* **64**, 307-333.

Jensen, J.L. and Petersen, N.V. (1999). Asymptotic normality of the maximum-likelihood estimator in state space models. *Ann. Statist.* **27**, 514-535.

Karlin, S. and Taylor, H.M. (1975). *A First Course in Stochastic Processes.* New York: Academic Press.

Kingman, J.F.C. (1973). Subadditive ergodic theory. *Ann. Probab.* **1**, 883-909.

Klaassen, F. (2002). Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Econ.* **27**, 363-394.

Leroux, B.G. (1992). Maximum-likelihood estimation for hidden Markov models. *Stochastic Proc. Appl.* **40**, 127-143.

MacDonald, I.L. and Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-valued Time Series.* London: Chapman and Hall.

Petrie, T. (1969). Probabilistic functions of finite state Markov chains. *Ann. Math. Statist.* **40**, 97-115.

Rydén, T. (1994). Consistent and asymptotically normal parameter estimates for hidden Markov models. *Ann. Statist.* **22**, 1884-1895.

Rydén, T. (1995). Estimating the order of hidden Markov models. Statist. **26**, 345-354.