

# Maximum Likelihood Estimators for Markov Switching Autoregressive Processes with ARCH Component

Jürgen Franke\*      Joseph Tadjuidje Kamgaing †

September 28, 2009

## Abstract

We consider a mixture of AR-ARCH models where the switching between the basic states of the observed time series is controlled by a hidden Markov chain. Under simple conditions, we prove consistency and asymptotic normality of the maximum likelihood parameter estimates combining general results on asymptotics of Douc et al (2004) and of geometric ergodicity of Franke et al (2007).

**Keywords:** AR-ARCH; mixture models; Markov switching; geometric ergodicity; consistency; asymptotic normality.

---

\*University of Kaiserslautern, Department of Mathematics, D-67653 Kaiserslautern, Germany.

Email: franke@mathematik.uni-kl.de

†Email: tadjuidj@mathematik.uni-kl.de

# 1 Introduction

Autoregressive (AR) processes and autoregressive conditionally heteroscedastic (ARCH) processes are well established and very popular models. While AR processes can be used for forecasting, ARCH processes are, e.g., used in the risk estimation. The combination of both models, the so-called AR-ARCH process, combines both the forecasting and the risk estimation capability. AR-ARCH models exhibit some other nice properties like their tail behavior compare, e.g., Borkovec and Klüppelberg (2001) or Cline (2006) for the special case of AR(1)-ARCH(1). Another variant of this model class is introduced in Lange et al. (2006). They consider an AR model with ARCH residuals for which the geometric ergodicity is investigated and the asymptotic behavior the Quasi maximum likelihood estimator is derived.

Recently, AR-ARCH models have repeatedly served as building blocks for switching regimes processes which allow for more flexibility in modeling data which only show locally a homogeneous behaviour. Wong and Li (2000, 2001) used them in an independent mixture of AR-ARCH model. Lanne and Saikkonen (2003) considered a threshold like mixture of AR-ARCH model, and Lee (2006) developed stability results for a mixture of AR-ARCH model where the switching is controlled in a threshold like manner by past observations of the time series and by a hidden white noise sequence. Franke et al. (2007) provided stability results for mixtures of nonlinear and nonparametric AR-ARCH models with Markovian switching regimes.

In this paper, we consider a hidden Markov mixture of autoregressive processes with ARCH component combining the models proposed by Wong and Li (2000, 2001), but with a general Markov switching regime. We apply general results given for such processes by Franke et al. (2007) to derive simple conditions for the geometric ergodicity of the time series. To illustrate the usefulness of that property, we apply the asymptotic theory of Douc et al. (2004) to prove consistency and asymptotic normality of the maximum likelihood parameter estimates. For sake of simplicity, we restrict our discussion to processes of order 1 in the autoregressive and ARCH component, but the arguments can be straightforwardly extended to hidden Markov mixtures of AR(p)-ARCH(q) models. This paper also serves as a role model how to derive the asymptotics for other, not only linear, parametric Markov switching autoregressive processes with ARCH component by combining the results of Franke et al. (2007) with those of Douc et al. (2004).

In the next two chapters, we first introduce the model under consideration and the main results. Technical lemmas and proofs are postponed to the appendix.

## 2 The Model and the Parameter Estimates

Let  $\{Q_t\}$  be a hidden stationary Markov chain with a finite number  $K$  of states which controls the data generating mechanism of the observed time series  $\{X_t\}$ . Let  $A = (a_{ij})_{i,j=1,\dots,K}$  denote the corresponding transition matrix and  $\pi = (\pi_1, \dots, \pi_K)'$  the stationary distribution of the chain. To simplify notation, we consider a hidden Markov mixture of AR-ARCH processes of first order only, i.e.

$$X_t = \sum_{k=1}^K S_{tk} (m_k(X_{t-1}, \theta) + \sigma_k(X_{t-1}, \theta) \varepsilon_t) \quad (2.1)$$

where the current state is indicated by

$$S_{tk} = \begin{cases} 1 & \text{if } Q_t = k \\ 0 & \text{otherwise} \end{cases} \quad (2.2)$$

The trend and volatility functions are of the standard parametric autoregressive and ARCH form

$$m_k(x, \theta) = \alpha_k x \quad \text{and} \quad \sigma_k^2(x, \theta) = \omega_k + \beta_k x^2, \quad k = 1, \dots, K,$$

and the innovations  $\varepsilon_t$  are i.i.d.  $\mathcal{N}(0, 1)$  variables.  $\theta$  combines all the free parameters of the model, i.e.  $\alpha_k \in \mathbb{R}, \beta_k \geq 0, \omega_k > 0, k = 1, \dots, K$ , as well as  $a_{ij} \geq 0, i = 1, \dots, K, j = 1, \dots, K - 1$ , into a vector of dimension  $K(K + 2)$ . The latter have to satisfy  $\sum_j a_{ij} = 1, i = 1, \dots, K$ , additionally. In the following,  $\Theta \subset \mathbb{R}^{K(K+2)}$  denotes the parameter set, and we sometimes write  $A = A_\theta$  to stress the dependence of the transition matrix on the parameters.

Below, we shall give conditions on  $\theta$  which guarantee the existence of a stationary solution to equation (2.1) as well as its geometric ergodicity. Given those conditions are satisfied, we consider the observed process to be sampled from a stationarity and geometrically ergodic mechanism  $\{(Q_t, X_t)\}$ , and we assume the starting values  $(Q_0, X_0)$  to be generated accordingly to the corresponding stationary distribution. Then, the combined process  $\{(Q_t, X_t)\}_{t=0}^\infty$  is a stationary Markov process defined on the product space  $\{1, \dots, K\} \times \mathbb{R}$ .

We always assume that the evolution of the hidden Markov chain does not directly depend on the observed time series, which follows from

**A. 2.1.**  $\{\varepsilon_t\}_{t=0}^\infty$  is independent of  $\{Q_t\}_{t=0}^\infty$ .

Then, we have e.g. for  $t > 0, k = 1, \dots, K$ ,

$$\mathbb{P}(Q_t = k | Q_s, X_s, s = 0, \dots, t-1) = \mathbb{P}(Q_t = k | Q_s, s = 0, \dots, t-1) = \mathbb{P}(Q_t = k | Q_{t-1})$$

by the Markov property.

To define the parameter estimates of interest, we first have to introduce some notation. Let  $g_\theta(\cdot | x, k)$  denote the conditional density of  $X_t$  given  $X_{t-1} = x, Q_t = k$ , which under model (2.1) is the Gaussian density with mean  $m_k(x, \theta)$  and variance  $\sigma_k^2(x, \theta)$ .

### 3 Asymptotic Properties of the Parameter Estimate

Given a sequence  $\{y_t\}_{t \in \mathbb{Z}}$  of deterministic or random real numbers and  $m, n \in \mathbb{Z}$ ,  $m \leq n$ , we set  $y_m^n = \{y_m, \dots, y_n\}$ . For  $q_0 \in \{1, \dots, K\}$ , we then get conditional likelihood function as the conditional density of  $X_1^n$  given  $X_0$  and  $Q_0 = q_0$

$$p_\theta(X_1^n | X_0, Q_0 = q_0) = \sum_{q_n=1}^K \dots \sum_{q_1=1}^K \prod_{t=1}^n a_{q_{t-1}, q_t} g_\theta(X_t | X_{t-1}, q_t) \quad (2.3)$$

and the conditional log-likelihood function given  $X_0$  and  $Q_0 = q_0$

$$\begin{aligned} l_n(\theta, q_0) &= \log p_\theta(X_1^n | X_0, Q_0 = q_0) \\ &= \sum_{t=1}^n \log p_\theta(X_t | X_0^{t-1}, Q_0 = q_0). \end{aligned} \quad (2.4)$$

with

$$\begin{aligned} p_\theta(X_t | X_0^{t-1}, Q_0 = q_0) \\ &= \sum_{q_{t-1}=1}^K \sum_{q_t=1}^K g_\theta(X_t | X_{t-1}, q_t) a_{q_{t-1}, q_t} \mathbb{P}(Q_{t-1} = q_{t-1} | X_0^{t-1}, Q_0 = q_0). \end{aligned}$$

Similarly, let us introduce the conditional log-likelihood function given only  $X_0$

$$l_n(\theta) = \sum_{t=1}^n \log \bar{p}_\theta(X_t | X_0^{t-1}) \quad (2.5)$$

with

$$\bar{p}_\theta(X_t | X_0^{t-1}) = \sum_{q_{t-1}=1}^K \sum_{q_t=1}^K g_\theta(X_t | X_{t-1}, q_t) a_{q_{t-1}, q_t} \mathbb{P}(Q_{t-1} = q_{t-1} | X_0^{t-1}). \quad (2.6)$$

The maximum likelihood estimates  $\hat{\theta}_n = \arg \max_{\theta \in \Theta} l_n(\theta)$  is hard to evaluate compared to

$$\hat{\theta}_{n, q_0} = \arg \max_{\theta \in \Theta} l_n(\theta, q_0) \quad (2.7)$$

where the function to be maximized has a simple explicit form given by (2.3) and (2.4). However,  $Q_0$  is not known. Nevertheless, we may use  $\hat{\theta}_{n, q_0}$  with an arbitrary initial value  $q_0 \in \{1, \dots, K\}$  which asymptotically will make no difference by Proposition 3.2 below.

## 3 Asymptotic Properties of the Parameter Estimate

In this section, we state our main results. Under rather weak conditions, we may conclude that the assumptions of Douc et al. (2004) are fulfilled for our model and,

### 3 Asymptotic Properties of the Parameter Estimate

therefore, make use of their results to derive the asymptotic of the parameter estimates. We assume that the data  $X_0, \dots, X_n$  are generated by a hidden Markov mixture of AR-ARCH processes as in (2.1) with parameter  $\theta^*$ . Moreover, we assume that assumption A.2.1 holds.

**A. 3.1.** The parameter set  $\Theta$  is a compact subset of  $\mathbb{R}^{K(K+2)}$ , and  $\theta^*$  is an interior point of  $\Theta$ .

That assumption on the parameter set  $\Theta$  as well as on the true parameter value  $\theta^*$  are quite standard in the literature and will be considered here without any further justification.

A major condition to apply the results of Douc et al. (2004) is stationarity, irreducibility and geometric ergodicity of the Markov process  $\{(Q_t, X_t)\}$ . Using the results of Franke et al. (2007), this follows from the subsequent assumptions.

**A. 3.2.**  $\mathbb{E}|X_t| < \infty$

**A. 3.3.**  $0 < a_- \leq \inf_{\theta \in \Theta} \inf_{i,j} a_{ij} \leq \sup_{\theta \in \Theta} \sup_{i,j} a_{ij} = a_+ < 1$ ,

**A. 3.4.**

$$\sum_{k=1}^K a_{lk} (\alpha_k^2 + \beta_k) < 1 \quad \text{for all } l = 1, \dots, K.$$

Mark that A.3.4 allows for mixture components with even explosive behaviour ( $|\alpha_k| > 1$ ) if the corresponding state is visited rarely, i.e.  $a_{lk}$  is for all  $l$  small enough.

**Proposition 3.1.** *Let  $\{X_t\}$  be generated from model (2.1), and let A.2.1 and A.3.1 to A.3.4 hold. It follows*

1.  $\{Q_t\}$  is a strictly stationary, irreducible and aperiodic Markov chain with finite state space  $\{1, \dots, K\}$ .
2.  $\{(Q_t, X_t)\}$  is geometrically ergodic.

*Proof.* 1. follows immediately from A3.3 and the properties of Markov chains with finite state space. 2. follows from Theorem 1 of Franke et al. (2007), where we only have to check the assumptions A.1 to A.8 of that theorem. In particular, A.1 follows from A.3.3, A.2, A.3 from A.2.1 and A.8 from A.3.4. A.4, A.6 and A.7 are implied by the our choice of i.i.d. standard normal innovations  $\varepsilon_t$ , and A.5 follows from the special form of  $m_k, \sigma_k^2$  and from  $\omega_1, \dots, \omega_K > 0$ .  $\square$

The following result implies that in estimating the model parameter by maximizing the log likelihood, it makes no difference if assume  $Q_0 = q_0$  to be given. The proof is postponed to the appendix.

**Proposition 3.2.** *Consider A.3.1 to A.3.3 hold. It follows*

$$\sup_{\theta} \sup_{1 \leq q_0 \leq K} \left| \frac{1}{n} l_n(\theta, q_0) - l(\theta) \right| \longrightarrow 0 \quad \text{a.s. as } n \rightarrow \infty$$

## 4 Technical Appendix

For proving consistency, we need a standard identifiability condition for the true parameter vector which is essentially a condition on the parameter set  $\Theta$ .

**A. 3.5.** For all  $n \geq 1$ ,  $\theta^*$  is the unique solution in  $\Theta$  of

$$\mathbb{E} \left( \log \frac{p_{\theta^*}(X_1^n | X_0)}{p_{\theta}(X_1^n | X_0)} \right) = 0$$

**Theorem 3.1.** *Let A.2.1 and A.3.1 to A.3.5 hold. Then, for all  $q_0 = 1, \dots, K$ ,*

$$\lim_{n \rightarrow \infty} \hat{\theta}_{n, q_0} = \theta^* \text{ a.s.}$$

where

$$\hat{\theta}_{n, q_0} = \arg \max_{\theta \in \Theta} l_n(\theta, q_0).$$

To formulate the asymptotic normality of the parameter estimate, we have to introduce the notation

$$I(\theta) = -\mathbb{E}_{\theta} \nabla_{\theta}^2 \log \bar{p}_{\theta}(X_t | X_{-\infty}^{t-1}),$$

which does not depend on  $t$  for stationary processes.  $I(\theta^*)$  is the Fisher information in our model, and we can estimate it consistently as described in the following theorem.

**Theorem 3.2.** *Consider A.3.1 to A.3.5 and assume, additionally, that  $\mathbb{E}(\varepsilon_t^4) < \infty$ , and that  $I(\theta^*)$  is positive definite. Then, for all  $q_0$*

$$\frac{1}{n} \nabla_{\theta}^2 l_n(\hat{\theta}_{n, q_0}, q_0) \longrightarrow I(\theta^*) \text{ a.s.}$$

and

$$\sqrt{n}(\hat{\theta}_{n, q_0} - \theta^*) \longrightarrow \mathcal{N}(0, (I(\theta^*))^{-1})$$

**Remark 3.1.** From our proof, it is clear that we need a 4<sup>th</sup> moment assumption, which is quite natural for proving asymptotic normality if parameters of variance like  $\omega_k, \beta_k$  are involved. For allowing even more heavy-tailed distributions and, therefore, further weakening the moment condition, one could think of modifying the volatility function, e.g. defining

$$\sigma_k^2(\theta, x) = \omega_k + \beta_k |x|^{\gamma} \text{ for some } 0 \leq \gamma < 2$$

will allow for weaker moment conditions.

## 4 Technical Appendix

Throughout the whole appendix, we assume that  $\{Q_t, X_t\}$  is a stationary process generated from model (2.1). We first start with some technical lemmas which are

## 4 Technical Appendix

needed for proving consistency and asymptotic normality. Under model (2.1), the conditional density of  $X_t$  given  $X_{t-1} = \bar{x}$  is

$$\begin{aligned} g_\theta(x|\bar{x}) &= \sum_{k=1}^K \pi_k g_\theta(x|\bar{x}, k) = \sum_{k=1}^K \frac{\pi_k}{\sigma_k(\bar{x}, \theta)} \varphi\left(\frac{x - m_k(\bar{x}, \theta)}{\sigma_k(\bar{x}, \theta)}\right), \\ &= \sum_{k=1}^K \frac{\pi_k}{\sqrt{\omega_k + \beta_k \bar{x}^2}} \varphi\left(\frac{x - \alpha_k \bar{x}}{\sqrt{\omega_k + \beta_k \bar{x}^2}}\right), \end{aligned} \quad (4.1)$$

where  $\varphi$  denotes the standard normal density.

**Lemma 4.1.** *Consider A.3.1 to A.3.3 hold. Then,*

1. For all  $x, \bar{x} \in \mathbb{R}$ ,

$$\inf_{\theta \in \Theta} g_\theta(x|\bar{x}) > 0, \quad \sup_{\theta \in \Theta} g_\theta(x|\bar{x}) < \infty \quad (4.2)$$

2.

$$b_+ = \sup_{\theta} \sup_{x, \bar{x}, k} g_\theta(x | \bar{x}, k) < \infty \quad (4.3)$$

and

$$\mathbb{E} \left| \log \inf_{\theta} g_\theta(X_1 | X_0) \right| < \infty. \quad (4.4)$$

*Proof.* 1. Using  $\delta_\sigma^2 \equiv \min\{\omega_1, \dots, \omega_K\} > 0$  and  $\sigma_k(u) \geq \delta_\sigma$  for all  $k, u$ , we get for all  $x, \bar{x} \in \mathbb{R}$ ,

$$g_\theta(x|\bar{x}) \leq \frac{1}{\delta_\sigma}$$

since  $\varphi(u) \leq 1$  for all  $u$ . On the other hand, by compactness of  $\Theta$ , we can choose  $M > 0$  such that  $\alpha_k^2, \omega_k, \beta_k \leq M^2, k = 1, \dots, K$ . Then,

$$(x - \alpha_k \bar{x})^2 = (|x + \alpha_k \bar{x}|)^2 \leq (|x| + M|\bar{x}|)^2,$$

and

$$0 < \delta_\sigma \leq \sigma_k(\bar{x}, \theta) = \sqrt{\omega_k + \beta_k \bar{x}^2} \leq M(1 + |\bar{x}|),$$

we get for all  $\theta \in \Theta$ , as  $\varphi(u)$  is decreasing in  $|u|$ ,

$$\begin{aligned} g_\theta(x|\bar{x}) &\geq \max_{k=1, \dots, K} \frac{\pi_k}{\sigma_k(\bar{x}, \theta)} \varphi\left(\frac{x - \alpha_k \bar{x}}{\sigma_k(\bar{x}, \theta)}\right) \\ &\geq \max_{k=1, \dots, K} \frac{\pi_k}{M(1 + |\bar{x}|)} \varphi\left(\frac{x - \alpha_k \bar{x}}{\sigma_k(\bar{x}, \theta)}\right) \\ &\geq \frac{1}{K} \frac{1}{M(1 + |\bar{x}|)} \varphi\left(\frac{|x| + M|\bar{x}|}{\delta_\sigma}\right) > 0, \end{aligned} \quad (4.5)$$

## 4 Technical Appendix

as  $\max_{k=1,\dots,K} \pi_k \geq \frac{1}{K}$  due to  $\pi_1 + \dots + \pi_K = 1$ .

2. By definition and moving along the same line of arguments as in 1., we see that  $b_+$  is trivially dominated by a positive constant. Hence the first part of our assertion holds. For the second part, we get from (4.5):

$$\frac{1}{\delta_\sigma} \geq \inf_{\theta} g_{\theta}(X_1|X_0) \geq \frac{1}{KM} \frac{1}{1+|X_0|} \varphi\left(\frac{|X_1| + M|X_0|}{\delta_\sigma}\right)$$

Henceforth,

$$\begin{aligned} \left| \log \inf_{\theta} g_{\theta}(X_1|X_0) \right| &\leq |\log \delta_\sigma| + \log \sqrt{2\pi} + \left| \frac{1}{KM} \right| \\ &\quad + \log(1 + |X_0|) + \frac{(|X_1| + M|X_0|)^2}{2\delta_\sigma^2} \end{aligned}$$

Using the moment assumptions and observing that  $\mathbb{E} \log(1 + |X_0|) \leq \mathbb{E}|X_0| < \infty$ , the assertion follows.  $\square$

### Proof of Proposition 3.2 and Theorem 3.1:

Under our conditions, the proof of Theorem 1 of Franke et al. (2007) implies that  $\{(Q_t, X_t)\}$  is not only geometrically ergodic, but also irreducible and aperiodic, and every compact set is a petite set. Choosing only the  $\theta$  for which the drift condition is fulfilled, the transition kernel of the combined Markov process  $\{(Q_t, X_t)\}$  is positive Harris recurrent. Therefore, assumption (A2) of Douc et al. (2004) is satisfied. Then, Proposition 3.2 follows from going through the proof of Proposition 2 of Douc et al. (2004), where we only have to check, if their other assumptions are satisfied too. Our assumptions A.3.3 represents (A1) of Douc et al., (A3) is implied by our Lemma 4.1 below, and (A4) is immediate from the representation (4.1) of  $g_{\theta}(x|\bar{x})$ .

Marking that any stationary process  $\{Z_t\}_{t \geq 0}$  can be extended to a two-sided process  $\{Z_t\}_{-\infty < t < \infty}$ , see e.g. Theorem 4.8 of Krengel (1985), our Theorem 3.1 follows from Theorem 1 of Douc et al. (2004) once we have checked their conditions. (A1)-(A4) have been discussed already in the previous paragraph. The identifiability condition (A5) follows immediately from our Assumption 3.5. Finally, the required geometric ergodicity follows from Proposition 3.1.

For sake of reference, we give some details on the line of arguments used by Douc et al. (2004) using our notation. The consistency proof for the conditional likelihood, i.e. the proof of Proposition 3.2, follows the classical scheme in the literature, which consists of proving the existence of a deterministic function  $l(\theta)$  such that asymptotically  $\frac{1}{n} l_n(\theta, q_0) \rightarrow l(\theta)$  a.s. uniformly w.r.t  $\theta \in \Theta$ . This argument requires a uniform law of large numbers, and  $\theta^*$ , the desired optimum of  $l(\theta)$ , has to be identifiable in a sense to be specified later. We also need to emphasize that  $l(\theta)$  should not depend on the initial state value  $Q_0 = q_0$ . To check those requirements Douc et al. (2004) first show that under some suitable conditions and for any  $q_0 \in \{1, \dots, K\}$

$$\sup_{\theta \in \Theta} |l_n(\theta, q_0) - l_n(\theta)| \leq \frac{1}{(1 - \rho)^2}, \quad \text{a.s. for some } 0 \leq \rho < 1.$$



## 4 Technical Appendix

Therefore asymptotically,

$$\frac{1}{n}(l_n(\theta, q_0) - l_n(\theta)) \longrightarrow 0 \text{ a.s and in } L^1 \quad \text{uniformly in } \theta \in \Theta$$

In the next step, they prove that

$$\begin{aligned} \frac{1}{n}l_n(\theta) &= \frac{1}{n} \sum_{t=1}^n \log \bar{p}_\theta(X_t|X_0^{t-1}) \\ &\approx \frac{1}{n} \sum_{t=1}^n \log \bar{p}_\theta(X_t|X_{-\infty}^{t-1}), \end{aligned}$$

compare (2.5), where  $\frac{1}{n} \sum_{t=1}^n \log \bar{p}_\theta(X_t|X_{-\infty}^{t-1})$  is a sample mean of observations from a two-sided stationary ergodic sequence of random variables in  $L^1$ .

To show this result, they introduce

$$\Delta_{t,m,q}(\theta) = \log p_\theta(X_t|X_{-m}^{t-1}, Q_{-m} = q)$$

and

$$\begin{aligned} \Delta_{t,m}(\theta) &= \log p_\theta(X_t|X_{-m}^{t-1}) \\ &= \log \sum_{q=1}^K p_\theta(X_t|X_{-m}^{t-1}, Q_{-m} = q) \mathbb{P}(Q_{-m} = q|X_{-m}^{t-1}), \end{aligned}$$

in particular  $l_n(\theta) = \sum_{t=1}^n \Delta_{t,0}(\theta)$ .

They then show that  $\{\Delta_{t,m}(\theta)\}_{m \geq 0}$  and  $\{\Delta_{t,m,q}(\theta)\}_{m \geq 0}$  are uniform Cauchy sequences w.r.t.  $\theta \in \Theta$  almost surely and derive that they converge uniformly w.r.t.  $\theta$  almost surely. Furthermore, they also show that  $\{\Delta_{t,m,q}(\theta)\}_{m \geq 0}$  is uniformly bounded in  $L^1$ , and for all  $m$ ,

$$\sup_{\theta} \sup_q |\Delta_{t,m,q}(\theta) - \Delta_{t,m}(\theta)| \leq \frac{\rho^{k+m-1}}{1-\rho}.$$

From the above properties, they derive,

$$\sum_{t=1}^n \sup_{\theta} |\Delta_{t,0}(\theta) - \Delta_{t,\infty}(\theta)| \leq \frac{2}{(1-\rho)^2} \text{ a.s.}$$

such that, in particular,

$$\frac{1}{n}l_n(\theta) \longrightarrow l(\theta) = \mathbb{E} \Delta_{0,\infty}(\theta) \text{ a.s.}$$

Hence, for all  $q_0$  and  $\theta \in \Theta$ ,

$$\frac{1}{n}l_n(\theta, q_0) \longrightarrow l(\theta) \text{ a.s.}$$

## 4 Technical Appendix

Finally, using the regularity implied by their assumptions, they establish the uniform law of large numbers formulated as Proposition 3.2 above:

$$\sup_{\theta} \sup_{q_0} \left| \frac{1}{n} l_n(\theta, q_0) - l(\theta) \right| \longrightarrow 0 \text{ as } n \rightarrow \infty$$

which together with their identifiability condition provides the consistency of the MLE.

In the next lemma,  $\Theta^* \subset \Theta$  denotes an open neighbourhood of  $\theta^*$  contained in  $\Theta$  which exists by A.3.1. To stress the dependence on the model parameters, we sometimes write  $A_{\theta}, a_{kl}(\theta)$  for the transition matrix of  $\{Q_t\}$  and its elements.

**Lemma 4.2.** *Consider A.3.1 to A.3.3 hold. It follows*

- (a) *for all  $k, l \in \{1, \dots, K\}$  and  $x, \bar{x} \in \mathbb{R}$ , the function  $\theta \mapsto a_{kl}(\theta)$  and  $\theta \mapsto g_{\theta}(x|\bar{x}, k)$  are twice continuously differentiable on  $\Theta^*$ .*
- (b)  *$\sup_{\theta \in \Theta^*} \sup_{k,l} \|\nabla_{\theta} \log a_{kl}(\theta)\| < \infty$  and  $\sup_{\theta \in \Theta^*} \sup_{k,l} \|\nabla_{\theta}^2 \log a_{kl}(\theta)\| < \infty$*
- (c)  *$\mathbb{E}\{\sup_{\theta \in \Theta^*} \sup_k \|\nabla_{\theta} \log g_{\theta}(X_1|X_0, k)\|\} < \infty$  and  $\mathbb{E}\{\sup_{\theta \in \Theta^*} \sup_k \|\nabla_{\theta}^2 \log g_{\theta}(X_1|X_0, k)\|\} < \infty$  for all  $k = 1, \dots, K$ .*

*Proof.* The first part of (a) and (b) follow immediately from the fact that the transition probabilities  $a_{kl}(\theta)$  are parameters themselves or, for  $l = K$ , linear functions of the parameters. For the other assertions, let us recall that  $g_{\theta}(\cdot|\bar{x}, k)$  is a Gaussian density, and in particular

$$\begin{aligned} G_k(\theta) &= \log g_{\theta}(X_t|X_{t-1}, k) \\ &= -\frac{1}{2} \log(\omega_k + \beta_k X_{t-1}^2) - \log \sqrt{2\pi} - \frac{1}{2} \frac{(X_t - \alpha_k X_{t-1})^2}{\omega_k + \beta_k X_{t-1}^2}. \end{aligned} \quad (4.6)$$

The required differentiability of  $g_{\theta}(\cdot|\bar{x}, k)$  follows immediately, recalling that  $\alpha_k, \omega_k, \beta_k$  are bounded and bounded away from 0 on  $\Theta^*$ . To prove (c), it is enough to investigate

## 4 Technical Appendix

the first and second order partial derivatives

$$\begin{aligned}
\frac{\partial G_k(\theta)}{\partial \alpha_k} &= \frac{X_{t-1}(X_t - \alpha_k X_{t-1})}{\omega_k + \beta_k X_{t-1}^2} \\
\frac{\partial G_k(\theta)}{\partial \omega_k} &= -\frac{1}{2(\omega_k + \beta_k X_{t-1}^2)} + \frac{1}{2} \frac{(X_t - \alpha_k X_{t-1})^2}{(\omega_k + \beta_k X_{t-1}^2)^2}, \\
\frac{\partial G_k(\theta)}{\partial \beta_k} &= -\frac{X_{t-1}^2}{2(\omega_k + \beta_k X_{t-1}^2)} + \frac{1}{2} \frac{X_{t-1}^2 (X_t - \alpha_k X_{t-1})^2}{(\omega_k + \beta_k X_{t-1}^2)^2}, \\
\frac{\partial^2 G_k(\theta)}{\partial \omega_k \partial \alpha_k} &= -\frac{X_{t-1}(X_t - \alpha_k X_{t-1})}{(\omega_k + \beta_k X_{t-1}^2)^2}, \\
\frac{\partial^2 G_k(\theta)}{\partial \beta_k \partial \alpha_k} &= \frac{-X_{t-1}^3 (X_t - \alpha_k X_{t-1})}{(\omega_k + \beta_k X_{t-1}^2)^2}, \\
\frac{\partial^2 G_k(\theta)}{\partial \omega_k \partial \beta_k} &= \frac{X_{t-1}^2}{2(\omega_k + \beta_k X_{t-1}^2)^2} - \frac{X_{t-1}^2 (X_t - \alpha_k X_{t-1})^2}{(\omega_k + \beta_k X_{t-1}^2)^3} \\
\frac{\partial^2 G_k(\theta)}{\partial \alpha_k^2} &= \frac{-X_{t-1}^2}{\omega_k + \beta_k X_{t-1}^2} \\
\frac{\partial^2 G_k(\theta)}{\partial \omega_k^2} &= \frac{1}{2(\omega_k + \beta_k X_{t-1}^2)^2} - \frac{(X_t - \alpha_k X_{t-1})^2}{(\omega_k + \beta_k X_{t-1}^2)^3} \\
\frac{\partial^2 G_k(\theta)}{\partial \beta_k^2} &= \frac{X_{t-1}^4}{2(\omega_k + \beta_k X_{t-1}^2)^2} - \frac{X_{t-1}^4 (X_t - \alpha_k X_{t-1})^2}{(\omega_k + \beta_k X_{t-1}^2)^3}
\end{aligned}$$

Under model (2.1) there are constants  $\tilde{\alpha}, \tilde{\beta} > 0$  such that for all and for  $\theta \in \Theta^*$ ,  $|X_t| \leq \tilde{\alpha}|X_{t-1}| + \tilde{\beta}|X_{t-1}||\varepsilon_t|$  and additionally we have  $X_{t-1}, \varepsilon_t$  are independent. (c) follows from observing that  $\varepsilon_t$  has finite variance and that, as  $\omega_1, \dots, \omega_K$  are uniformly bounded away from 0 in  $\Theta^*$ , functions of the form

$$\frac{z^\mu}{(\omega_k + \beta_k z^2)^\nu} \leq c_{\mu\nu}, \tag{4.7}$$

are uniformly bounded by a suitable constant  $c_{\mu\nu}$  in  $0 \leq z < \infty, k = 1, \dots, K$  for all  $\mu \leq 2\nu$ .  $\square$

**Remark 4.1.** Since  $|X_t| \leq \tilde{\alpha}|X_{t-1}| + \tilde{\beta}|X_{t-1}||\varepsilon_t|$  it follows that

$$|X_t - \alpha_k X_{t-1}| \leq |X_{t-1}|(|\alpha_k| + \tilde{\alpha} + \tilde{\beta}|\varepsilon_t|)$$

and

$$(X_t - \alpha_k X_{t-1})^2 \leq 2X_{t-1}^2(\{|\alpha_k| + \tilde{\alpha}\}^2 + \tilde{\beta}^2 \varepsilon_t^2). \tag{4.8}$$

Therefore, using equation (4.7) we can easily observe that the highest moment condition required is  $\mathbb{E}\varepsilon_t^2 < \infty$ . Using the same argument, we also observe from equation (4.9) that we only need  $\mathbb{E}\varepsilon_t^4 < \infty$  to conclude the moment assumption for the asymptotic normality proof. However, as one can observe from the proof of Lemma 4.1, we still need to assume  $\mathbb{E}|X_0| < \infty$ , which is granted as the geometric ergodicity proof implies even  $\mathbb{E}X_t^2 < \infty$ .

## 4 Technical Appendix

**Lemma 4.3.** *If A.3.1 to A.3.3 and, additionally,  $\mathbb{E}(X_t^4) < \infty$  hold then*

1. *There exists a function  $f_0 : \mathbb{R}^2 \mapsto \mathbb{R}^+$  satisfying  $\mathbb{E}f_0(X_t, X_{t-1}) < \infty$ , such that*

$$\sup_{\theta \in \Theta^*} g_\theta(x|\bar{x}, k) \leq f_0(x, \bar{x}) \quad \text{for all } x, \bar{x} \in \mathbb{R}$$

2. *There exist functions  $f_1, f_2 : \mathbb{R}^2 \mapsto \mathbb{R}^+$  satisfying  $\mathbb{E}f_i(X_t, X_{t-1}) < \infty, i = 1, 2$ , such that*

$$\|\nabla_\theta g_\theta(x|\bar{x}, k)\| \leq f_1(x, \bar{x}) \quad \text{and} \quad \|\nabla_\theta^2 g_\theta(x|\bar{x}, k)\| \leq f_2(x, \bar{x}) \quad \text{for all } x, \bar{x} \in \mathbb{R}$$

*Proof.* We use the notation

$$g_k(\theta) = g_\theta(x | \bar{x}, k) = \exp(G_k(\theta))$$

1. follows immediately with a constant  $f_0$  from  $0 < g_k(\theta) \leq \frac{1}{\delta_\sigma}$  which we have shown above.

For showing 2., let  $\gamma_k, \rho_k$  represent an arbitrary selection from  $\alpha_k, \beta_k, \omega_k$  with repetitions allowed. We have

$$\frac{\partial g_k(\theta)}{\partial \gamma_k} = \frac{\partial G_k(\theta)}{\partial \gamma_k} g_k(\theta), \quad \frac{\partial^2 g_k(\theta)}{\partial \gamma_k \partial \rho_k} = \left( \frac{\partial^2 G_k(\theta)}{\partial \gamma_k \partial \rho_k} + \frac{\partial G_k(\theta)}{\partial \gamma_k} \frac{\partial G_k(\theta)}{\partial \rho_k} \right) g_k(\theta), \quad (4.9)$$

where  $G_k$  and its partial derivatives are given in the proof of Lemma 4.2. Using those relations and, again, (4.7), we get with suitable constants  $c_1, c_2$  such that all partial derivatives of  $g_k(\theta)$  of first and second order with respect to  $\alpha_k, \beta_k, \omega_k$  are bounded by

$$(c_2 x^4 + c_1) g_\theta(x|\bar{x}, k) \leq \frac{1}{\delta_\sigma} (c_2 x^4 + c_1)$$

for all  $x, \bar{x} \in \mathbb{R}$  and all  $k = 1, \dots, K$ . Setting  $f_1, f_2$  equal to the right-hand side of this inequality, the assertion follows from the moment condition on  $X_t$ .  $\square$

### Proof of Theorem 3.2:

*Proof.* The assertion follows from Theorems 3 and 4 of Douc et al. (2004). We have already discussed in the proof of Theorem 3.1 that the majority of their assumptions are fulfilled. The remaining assumptions (A6)-(A8) follow from Lemma 4.2 and 4.3.  $\square$

Again, for sake of reference, we give some more details of the asymptotic normality proof due to Douc et al. (2004). As usual, they rely on

1. a (local) uniform law of large number for the observed Fisher information  $\frac{1}{n} \nabla_\theta^2 l_n(\theta, q_0)$  for  $\theta$  in the vicinity of  $\theta^*$ .
2. A CLT for the Fisher score function  $\frac{1}{\sqrt{n}} \nabla_\theta l_n(\theta^*, q_0)$ .

## 4 Technical Appendix

For the CLT, as in Douc et al. (2004), we define

$$\begin{aligned} \frac{1}{\sqrt{n}} \nabla_{\theta} l_n(\theta^*, q_0) &= \frac{1}{\sqrt{n}} \sum_{t=1}^n \nabla_{\theta} \log p_{\theta^*}(X_t | X_0^{t-1}, Q_0 = q_0) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^n \Delta'_{t,0,q_0}(\theta^*), \end{aligned} \quad (4.10)$$

where, using the notation of the proof of Theorem 3.1,

$$\begin{aligned} \Delta'_{t,0,q_0}(\theta) = \nabla_{\theta} \Delta_{t,0,q_0}(\theta) &= \mathbb{E}_{\theta} \left[ \sum_{i=1}^t \phi(\theta, Z_{i-1}^i) | X_0^t, Q_0 = q_0 \right] \\ &\quad - \mathbb{E}_{\theta} \left[ \sum_{i=1}^{t-1} \phi(\theta, Z_{i-1}^i) | X_0^{t-1}, Q_0 = q_0 \right] \end{aligned} \quad (4.11)$$

with

$$\begin{aligned} \phi(\theta, Z_{i-1}^i) &= \phi(\theta, Z_{i-1}, Z_i) \\ &= \phi(\theta, (Q_{i-1}, X_{i-1}), (Q_i, X_i)) \\ &= \nabla_{\theta} \log(a_{Q_{i-1}, Q_i} g_{\theta}(X_i | X_{i-1}, Q_i)). \end{aligned} \quad (4.12)$$

Similarly,

$$\Delta'_{t,m}(\theta) = \mathbb{E}_{\theta} \left[ \sum_{i=1}^t \phi(\theta, Z_{i-1}^i) | X_{-m}^t \right] - \mathbb{E}_{\theta} \left[ \sum_{i=1}^{t-1} \phi(\theta, Z_{i-1}^i) | X_{-m}^{t-1} \right] \quad (4.13)$$

Under suitable assumptions, we can show  $\{\Delta'_{t,\infty}\}_{t=-\infty}^{\infty}$  is an  $\mathcal{F}_t = \sigma(X_s, s \leq t)$  adapted, stationary, ergodic and square integrable martingale sequence for which we can apply a CLT, Durrett 1996, p 418.

In other words, for

$$I(\theta^*) = \mathbb{E}_{\theta^*} [\Delta'_{0,\infty}(\theta^*) \Delta'_{0,\infty}(\theta^*)^T]$$

we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \Delta'_{t,\infty}(\theta^*) \longrightarrow \mathcal{N}(0, I(\theta^*))$$

Additionally,

$$\lim_{n \rightarrow \infty} \mathbb{E} \left\| \frac{1}{\sqrt{n}} \sum_{t=1}^n (\Delta'_{t,0}(\theta^*) - \Delta'_{t,\infty}(\theta^*)) \right\|^2 = 0$$

and

$$\lim_{n \rightarrow \infty} \mathbb{E} \left\| \frac{1}{\sqrt{n}} \sum_{t=1}^n (\Delta'_{t,0,q}(\theta^*) - \Delta'_{t,0}(\theta^*)) \right\|^2 = 0.$$

#### 4 Technical Appendix

This proves that  $\Delta'_{t,0,q}$  can be approximated in  $L^2$  by a stationary martingale increment sequence, for which a CLT for stationary martingale has been applied. Therefore,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \Delta'_{t,0,q}(\theta^*) \longrightarrow \mathcal{N}(0, I(\theta^*)).$$

#### Uniform Law of large Number for the observed Fisher Information

Let us define

$$\begin{aligned} \nabla_{\theta}^2 \log \bar{p}_{\theta}(X_1^n | X_0, Q_0 = q_0) &= \mathbb{E}_{\theta} \left[ \sum_{i=1}^t \psi(\theta, Z_{i-1}^i) | X_0^t, Q_0 = q_0 \right] \\ &\quad + \text{var}_{\theta} \left[ \sum_{i=1}^t \phi(\theta, Z_{i-1}^i) | X_0^t, Q_0 = q_0 \right] \end{aligned} \quad (4.14)$$

where

$$\begin{aligned} \psi(\theta, Z_{i-1}^i) &= \psi(\theta, Z_{i-1}, Z_i) \\ &= \psi(\theta, (Q_{i-1}, X_{i-1}), (Q_i, X_i)) \\ &= \nabla_{\theta}^2 \log(a_{Q_{i-1}, Q_i} g_{\theta}(X_i | X_{i-1}, Q_i)). \end{aligned} \quad (4.15)$$

As previously,

$$\begin{aligned} \mathbb{E}_{\theta} \left[ \sum_{i=1}^t \psi(\theta, Z_{i-1}^i) | X_0^t, Q_0 = q_0 \right] &= \sum_{t=1}^n \left( \mathbb{E}_{\theta} \left[ \sum_{i=1}^t \psi(\theta, Z_{i-1}^i) | X_0^t, Q_0 = q_0 \right] \right. \\ &\quad \left. - \mathbb{E}_{\theta} \left[ \sum_{i=1}^{t-1} \psi(\theta, Z_{i-1}^i) | X_0^{t-1}, Q_0 = q_0 \right] \right) \end{aligned}$$

and

$$\begin{aligned} \text{var}_{\theta} \left[ \sum_{i=1}^t \phi(\theta, Z_{i-1}^i) | X_0^t, Q_0 = q_0 \right] &= \sum_{t=1}^n \left( \text{var}_{\theta} \left[ \sum_{i=1}^t \phi(\theta, Z_{i-1}^i) | X_0^t, Q_0 = q_0 \right] \right. \\ &\quad \left. - \text{var}_{\theta} \left[ \sum_{i=1}^{t-1} \phi(\theta, Z_{i-1}^i) | X_0^{t-1}, Q_0 = q_0 \right] \right) \end{aligned}$$

Let further define, for  $t \geq 1, m \geq 0$

$$\Delta'_{t,m}(\theta) = \mathbb{E}_{\theta} \left[ \sum_{i=-m+1}^t \psi(\theta, Z_{i-1}^i) | X_{-m}^t \right] - \mathbb{E}_{\theta} \left[ \sum_{i=-m+1}^{t-1} \psi(\theta, Z_{i-1}^i) | X_{-m}^{t-1} \right]$$

and

$$\Gamma_{t,m}(\theta) = \text{var}_{\theta} \left[ \sum_{i=-m+1}^t \phi(\theta, Z_{i-1}^i) | X_{-m}^t \right] - \text{var}_{\theta} \left[ \sum_{i=-m+1}^{t-1} \phi(\theta, Z_{i-1}^i) | X_{-m}^{t-1} \right]$$

$\Delta'_{t,m}$  and  $\Gamma_{t,m}$  converge to  $\Delta'_{t,\infty}$  and  $\Gamma_{t,\infty}$  in  $L^1$  as  $m \rightarrow \infty$ . Furthermore,  $\{\Delta'_{t,\infty}\}$  and  $\{\Gamma_{t,\infty}\}$  are stationary and ergodic. Therefore, the observed Fisher information matrix converges to  $-\mathbb{E}_{\theta^*}[\Delta'_{0,\infty}(\theta^*) + \Gamma_{0,\infty}(\theta^*)]$ .

## References

- BORKOVEC, M. AND KLÜPPELBERG, C. (2001) The tail of the stationary distribution of an autoregressive process with ARCH(1) errors. *The Annals of Applied Probability*, **11**, 4, 1220-1241
- CLINE, D.B.H. (2006) Regular variation of order 1 nonlinear AR-ARCH models *Stochastic Process and their applications*, **117**, 840-861
- DOUC, R., MOULINES, E. AND RYDEN, T. (2004) Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. *Ann. Statist.*, **32**, 5, 2254-2304
- DURRETT, R. (1996). *Stochastic Calculus: A Practical Introduction*. CRC Press, Boca Raton.
- FRANKE, J., STOCKIS, J-P. AND TADJUIDJE-KAMGAING, J. (2007) On Geometric Ergodicity of CHARME Models. <http://kluedo.ub.uni-kl.de/volltexte/2007/2057/>
- KRENGEL, U. (1985). *Ergodic Theorems*. De Gruyter, Berlin.
- LANGE, T., RAHBK, A. AND JENSEN, S.T. (2006) Estimation and asymptotic inference in the first order AR-ARCH model. *Technical report, Institute for Mathematical Sciences, University of Copenhagen*.
- LANNE, M. AND SAIKKONEN, P. (2003) Modeling the US short-term interest rate by mixture autoregressive processes. *J. of Financial Econometrics*, **1**, 1, 96-125.
- LEE, O. (2006) Stationarity and  $\beta$ -mixing property of a mixture AR-ARCH models. *Bull. Korean Math. Soc.*, **43**, 813-820.
- WONG, C.S. AND LI, W.K. (2000) On a mixture autoregressive model. *J. Royal Statist. Soc. B*, **62**, 95-115.
- WONG, C.S. AND LI, W.K. (2001) On a mixture autoregressive conditional heteroscedastic model. *J. American Statist. Assoc.*, **96**, 982-995.