

# On Geometric Ergodicity of CHARME Models

Jürgen Franke\*  
Jean-Pierre Stockis  
Joseph Tadjuidje Kamgaing

*University of Kaiserslautern*

January 10, 2007

## Abstract

In this paper we consider a CHARME Model, a class of generalized mixture of nonlinear nonparametric AR-ARCH time series. We apply the theory of Markov models to derive asymptotic stability of this model. Indeed, the goal is to provide some sets of conditions under which our model is geometric ergodic and therefore satisfies some mixing conditions. This result can be considered as the basis toward an asymptotic theory for our model.

**Keywords:** Nonparametric AR-ARCH; Mixture Models; Markov chain; Geometric Ergodicity

---

\*University of Kaiserslautern, Department of Mathematics, Erwin-Schroedinger-Str., 67663 Kaiserslautern, Germany. E-mail address: franke@mathematik.uni-kl.de

The work was supported by the Deutsche Forschungsgemeinschaft (DFG) as part of the priority research program 1114 *Mathematical Methods of Time Series and Digital Image Analysis*, the center of excellence *Dependable Adaptive Systems and Mathematical Modeling* funded by the state of Rhineland-Palatinate as well as the "Graduiertenkolleg Mathematik und Praxis" and by the Fraunhofer ITWM.

# 1 Introduction

Nonparametric conditional heteroscedastic autoregressive (nonlinear CHARN) models of the form

$$X_t = m(X_{t-1}, \dots, X_{t-p}) + \sigma(X_{t-1}, \dots, X_{t-p})\epsilon_t, \quad (1.1)$$

$m$  and  $\sigma$  unknown functions,  $\epsilon_t$  independent identically distributed (i.i.d.) random variables with mean 0, play an important role in many fields of application, for example in econometrics or finance, see for example Härdle and Tsybakov [2], Franke, Neumann and Stockis [6], Hafner [15]. Theoretical results about stability properties of this processes are available. In particular, the important property of geometric ergodicity is obtained under some conditions.

In practice, it is often not realistic to assume that the observed process has the same trend function  $m$  and the same volatility function  $\sigma$  at each time instant. In this paper we are analyzing the so-called Conditional Heteroscedastic Autoregressive Mixture of Experts, henceforth CHARME, models. Here, a hidden Markov chain  $\{Q_t\}$  with values in a finite set of states  $\{1, 2, \dots, K\}$  drives the dynamics of  $\{X_t\}$  and our model is defined as follows

$$X_t = \sum_{k=1}^K S_{tk} (m_k(X_{t-1}, \dots, X_{t-p}) + \sigma_k(X_{t-1}, \dots, X_{t-p})\epsilon_t) \quad (1.2)$$

with

$$S_{tk} = \begin{cases} 1 & \text{for } Q_t = k \\ 0 & \text{otherwise} \end{cases} \quad (1.3)$$

$m_k, \sigma_k, k = 1, \dots, K$  unknown functions,  $\epsilon_t$  i.i.d. random variables with mean 0.

Notice that for sake of simplicity of notation, we take the same number of components  $p$  in each trend function  $m_k$  and volatility function  $\sigma_k$ . This is done without loss of generality if we take  $p$  large enough.

We call this models CHARME since many authors using a mixture of models, e.g. in engineering are calling them mixture of experts as soon as nonparametric functions estimates, typically neural networks, are considered, see, e.g. Müller et al. [8], Jacob et al. [12], or Jiang and Tanner [10]. CHARME is quite useful for modeling time series data which are piecewise stationary such that their dynamics switch sometimes from one state to another. A typical example is given by stock returns if the market changes from a quiescent to a volatile phase. Tadjuidje [16] gives some applications of such models to financial data in the context of asset management and risk analysis where the state functions  $m_k, \sigma_k, k = 1, \dots, K$ , are estimated by neural networks.

Independently of the type of estimates considered, a crucial condition for developing a theory for estimation and testing in the setting of CHARME is the existence of a stochastic process satisfying (1.2) which is geometric ergodic. In this paper we investigate separately the case  $p = 1$  (section 1) and the case  $p \geq 1$  (section 2) since they differ somewhat with respect to the formulation and proof. In particular, the case  $p = 1$  is interesting on its own. We formulate for both cases two different sets of conditions.

## 2 First conditions for geometric ergodicity of CHARME processes

We focus on our CHARME model (1.2) and make the following assumptions

**A. 1** *The process  $\{Q_t\}$  with values on  $\{1, \dots, K\}$  is a first order strictly stationary Markov chain which is irreducible and aperiodic with probability distribution  $(\pi_1, \dots, \pi_K)$  and transition probability matrix  $A = (a_{ij})_{1 \leq i, j \leq K}$ .*

Obviously,  $\{S_t = (S_{t1}, \dots, S_{tK})'\}$  inherits the properties of  $\{Q_t\}$ .

**A. 2** Let  $\mathcal{G}_{t-1} = \sigma\{X_r, r \leq t-1\}$  be the  $\sigma$ -algebra generated by  $\{X_r, r \leq t-1\}$  and  $G_{t-1}$  any event in  $\mathcal{G}_{t-1}$ . Then

$$P(Q_t = j | Q_{t-1} = i, G_{t-1}) = P(Q_t = j | Q_{t-1} = i), \quad \forall i, j$$

This assumption means that the hidden process  $Q_t$  is independent of the past observations given its own past, i.e.  $Q_{t-1}$ .

**A. 3** Given  $(Q_{t-1}, X_{t-1}, X_{t-2}, \dots)$ ,  $Q_t$  is uncorrelated with the innovation  $\epsilon_t$ .

**A. 4**  $\epsilon_t$  is independent of  $X_{t-1}, X_{t-2}, \dots$ .

**A. 5** The functions  $m_k$  and  $\sigma_k$  are bounded on compact sets for all  $k$ , there exists a  $\delta$  such that  $\sigma_k(u) \geq \delta > 0$ , for all  $k, u$ .

**A. 6** The i.i.d. random variables  $\epsilon_t$  have a density  $f$  which is continuous and positive everywhere.

These assumptions are reasonable conditions for hidden Markov chain models, see e.g. Francq and Roussignol [11] or Francq, Roussignol and Zakoian [9].

Now, we restrict ourselves for the rest of this section to the case  $p = 1$ , i.e.  $m_k, \sigma_k$  are functions on the real line. We first assume

**A. 7** The i.i.d. random variables  $\epsilon_t$  have mean 0 and variance  $\sigma^2 = 1$

**A. 8**

$$\max_{l \in \{1, \dots, K\}} \limsup_{|x| \rightarrow \infty} \frac{\sum_k a_{lk} (m_k^2(x) + \sigma_k^2(x))}{x^2} < 1$$

A.8 is the generalization of the well-known sufficient condition for geometric ergodicity in the case of model (1.1). Now we need a Markov chain representing the transformed mixture process: under assumptions A.1 to A.4 it is easily seen that if we define, as previously,  $S_t = (S_{t1}, \dots, S_{tK})'$ , then,

$$\zeta_t = (S_t, X_t)'$$

is a Markov chain.

**Theorem 1** Under A.1 to A.8,  $\{\zeta_t\}$  is geometrically ergodic.

**Proof:** We are going to prove that the conditions of Theorem 15.0.1, (iii) of Meyn and Tweedie [7], pp 354 – 355, are satisfied.

- $\{\zeta_t\}$  is  $\varphi$ -irreducible if we take  $\varphi$  as the product of the stationary probability distribution measure on  $\{1, \dots, K\}$  and the Lebesgue measure on  $\mathbb{R}$

This can be proven as follows:

Let  $A = A_1 \times A_2$  be such that  $\varphi(A) > 0$ . Then  $A_1$  contains at least one integer between 1 and  $K$  and it is enough to prove that there exists  $t$  such that

$$P\left(\left(\begin{matrix} S_{t+1} \\ X_{t+1} \end{matrix}\right) \in \{e\} \times A_2 \mid S_1 = s_l, X_1 = x\right) > 0$$

with  $e$  a unit vector with the the  $k$ th component equal 1 and  $s_l$  a unit vector with the  $l$ th component equal 1. By definition,

$$\begin{aligned}
& P\left(\begin{pmatrix} S_2 \\ X_2 \end{pmatrix} \in \{e\} \times A_2 \mid S_1 = s_l, X_1 = x\right) \\
&= P(Q_2 = k, X_2 \in A_2 \mid S_1 = s_l, X_1 = x) \\
&= P(X_2 \in A_2 \mid Q_2 = k, S_1 = s_l, X_1 = x) P(Q_2 = k \mid Q_1 = l, X_1 = x) \\
&= a_{lk} P(m_k(x) + \sigma_k(x)\epsilon_2 \in A_2) \\
&= a_{lk} \int_{A_2} \frac{1}{\sigma_k(x)} f\left(\frac{u - m_k(x)}{\sigma_k(x)}\right) du \\
&= a_{lk} b_k(x) \quad \text{with } b_k(x) > 0
\end{aligned}$$

Further,

$$\begin{aligned}
& P\left(\begin{pmatrix} S_3 \\ X_3 \end{pmatrix} \in \{e\} \times A_2 \mid S_1 = s_l, X_1 = x\right) \\
&= \sum_{j=1}^K a_{lj} a_{jk} \int_{A_2} \int_{\mathbb{R}} \frac{1}{\sigma_k(y)} f\left(\frac{u - m_k(y)}{\sigma_k(y)}\right) \frac{1}{\sigma_j(x)} f\left(\frac{y - m_j(x)}{\sigma_j(x)}\right) dy du \\
&= \sum_{j=1}^K a_{lj} a_{jk} b_{jk}(x) \quad \text{with } b_{jk}(x) > 0
\end{aligned}$$

and doing so iteratively, we obtain

$$\begin{aligned}
& P\left(\begin{pmatrix} S_{t+1} \\ X_{t+1} \end{pmatrix} \in \{e\} \times A_2 \mid S_1 = s_l, X_1 = x\right) \\
&= \sum_{j, \dots, j_{t-1}}^K a_{lj_1} \cdots a_{j_{t-1}k} b_{j, \dots, j_{t-1}}(x)
\end{aligned}$$

which is strictly greater than 0 for some  $t$  because of the irreducibility of  $\{Q_t\}$  and the fact that  $b_{j, \dots, j_{t-1}}(x) > 0$ .

- Analogously it can easily be seen that  $\{\zeta_t\}$  is aperiodic.
- In the drift criterion of Theorem 15.0.1, (iii) mentioned previously appears the notion of a petite set. In our case, it can be shown that each compact set is indeed a small set and thus a petite set, see for example, Bhattacharya and Lee [3] and Lee and Shin [5].
- So, to apply the drift criterion, we need to find a function  $g(\zeta) > 1$ ,  $\beta > 0$  and  $M > 0$  such that

$$\frac{\mathbb{E}(g(\zeta_t) \mid \zeta_{t-1} = \binom{s_l}{x}) - g\left(\binom{s_l}{x}\right)}{g\left(\binom{s_l}{x}\right)} \leq -\beta \quad \text{for } \|\zeta_{t-1}\| > M$$

Let

$$g(\zeta_t) = 1 + X_t^2.$$

Then,

$$\begin{aligned}
\frac{\mathbb{E}(g(\zeta_t) \mid \zeta_{t-1} = \binom{s_l}{x}) - g\left(\binom{s_l}{x}\right)}{g\left(\binom{s_l}{x}\right)} &= \frac{\sum_k (m_k^2(x) + \sigma_k^2(x)) \mathbb{E}(S_{tk} \mid S_{t-1} = s_l) - x^2}{1 + x^2} \\
&\leq \frac{\sum_k (m_k^2(x) + \sigma_k^2(x)) a_{lk}}{x^2} - 1
\end{aligned}$$

and the conclusion is obtained by A.8. ■

However, in financial time series which are very often heavy-tailed, the existence of  $\sigma^2 = \text{var}(\epsilon_t)$  is not necessarily guaranteed. Therefore, instead of A.7 and A.8 we assume

**A. 9** *The i.i.d. random variables  $\epsilon_t$  are such that  $\mathbb{E}(|\epsilon_t|^\alpha) < \infty$  for some  $0 < \alpha \leq 1$*

**A. 10**

$$\max_{l \in \{1, \dots, K\}} \limsup_{|x| \rightarrow \infty} \frac{\sum_k a_{lk} (|m_k(x)|^\alpha + \sigma_k^\alpha(x) \mathbb{E}|\epsilon_t|^\alpha)}{|x|^\alpha} < 1$$

with  $\alpha$  as in A.9

**Theorem 2** *Under Assumptions A.1 to A.6, A.9 and A.10,  $\{\zeta_t\}$  is geometrically ergodic.*

**Proof:** The only part of this proof which is not similar to the proof of Lemma 1 is the drift criterion. Here we consider

$$g(\zeta) = 1 + |X_t|^\alpha.$$

Then,

$$\begin{aligned} & \frac{\mathbb{E}(g(\zeta_t) | \zeta_{t-1} = \binom{s_l}{x}) - g(\binom{s_l}{x})}{g(\binom{s_l}{x})} \\ & \leq \frac{\sum_k (|m_k(x)|^\alpha + \sigma_k^\alpha(x) \mathbb{E}|\epsilon_t|^\alpha) \mathbb{E}(|S_{tk}|^\alpha | S_{t-1} = s_l) - |x|^\alpha}{1 + |x|^\alpha} \\ & \leq \frac{\sum_k a_{lk} (|m_k(x)|^\alpha + \sigma_k^\alpha(x) \mathbb{E}|\epsilon_t|^\alpha)}{|x|^\alpha} - 1 \end{aligned}$$

and we conclude the proof by using A.10. ■

### 3 Geometric ergodicity for higher order CHARME processes

We now follow a slightly different route to geometric ergodicity of CHARME processes. We first state an auxiliary result that we are going to use for proving a condition for geometric ergodicity.

**Lemma 1** *Let  $\phi, \psi$  be random variables with values in  $\mathbb{R}^d$ ,  $C \subset \mathbb{R}^d$  a measurable set,  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  measurable and bounded on  $C$  satisfying  $g \geq 1$ . If there exist constants  $0 < r < 1$ ,  $B > 0$  such that*

$$\begin{aligned} \mathbb{E}(g(\phi) | \psi = x) &< rg(x), \text{ if } x \notin C \\ \mathbb{E}(g(\phi) | \psi = x) &< B, \text{ if } x \in C \end{aligned}$$

then, there exist  $\beta > 0$ ,  $b < \infty$  such that

$$\mathbb{E}(g(\phi) | \psi = x) - g(x) < -\beta g(x) + b \mathbb{I}_C(x).$$

**Proof:** The result follows if we choose  $\beta = 1 - r$  and  $b = B + \beta \sup_{x \in C} g(x)$  ■

In particular in our case where  $\{\phi_t\}$  is a Markov chain, it is enough to prove the existence of a petite set  $C$ , a function  $g \geq 1$  and constants  $0 < r < 1$  and  $B > 0$  such that

$$\begin{aligned}\mathbb{E}(g(\phi_t) | \phi_{t-1} = x) &< rg(x), x \notin C \\ \mathbb{E}(g(\phi_t) | \phi_{t-1} = x) &< B, x \in C.\end{aligned}$$

To achieve the desired results in this section, we use some new sets of assumptions.

**A. 11** *There exist  $\alpha_k, d_k \in \mathbb{R}^p$  with  $d_{ki} \geq 0, i = 1, \dots, p$  such that as  $\|u\| \rightarrow \infty$ ,*

$$m_k(u) = \sum_{i=1}^p \alpha_{ki} u_i + o(\|u\|)$$

and

$$\sigma_k^2(u) = \sum_{i=1}^p d_{ki} u_i^2 + o(\|u\|^2)$$

**A. 12**

$$\max_{l \in \{1, \dots, K\}} \sum_{i=1}^p \sum_{k=1}^K a_{lk} \left( d_{ki} + |\alpha_{ki}| \sum_{j=1}^p |\alpha_{kj}| \right) < 1$$

Remark that assumption A.11, similar to Assumption 3.2, (c) in Masry and Tjostheim [1], does not imply that the system has to be parametric or linear in any sense. It just bounds the trend functions asymptotically for  $\|u\| \rightarrow \infty$  by linear functions and the squared volatility functions by some quadratic functions in order to avoid an explosion of the system for large values of  $\|u\|$ .

Once more we need a Markov chain; under assumptions A.1 to A.4 as for the case  $p = 1$ , it is easily seen that, if we define as previously,  $S_t = (S_{t1}, \dots, S_{tK})'$ , then

$$\zeta_t = (S_t, X_t, \dots, X_{t-p+1})'$$

is a Markov chain.

**Theorem 3** *Under A.1 to A.7, A.11 and A.12,  $\{\zeta_t\}$  is geometrically ergodic*

**Proof:** As in the case  $p = 1$  we are going to prove that the conditions of Theorem 15.0.1, (iii) of Meyn and Tweedie [7] pp 354 – 355 are satisfied. Irreducibility, aperiodicity and the fact that each compact set is a petite set are proved in a very similar way as in the case  $p = 1$ , and therefore the proofs are omitted here. Hence, it remains to prove that the conditions of Theorem 1 hold. To achieve our goal, we need to consider a function  $g$  that we define as

$$g(\zeta_t) = 1 + X_t^2 + b_{p-1} X_{t-1}^2 + \dots + b_1 X_{t-p+1}^2$$

where  $b_{p-1}, \dots, b_1 > 0$  are not yet determined and will be suitably chosen later.

Let  $x = (x_t, \dots, x_{t-p+1})'$  be a vector of real numbers and  $s_l$  a  $K$ -dimensional unit vector with the  $l$ -th component equal 1 and consider

$$\begin{aligned}\mathbb{E}(g(\zeta_{t+1}) | (X_t, \dots, X_{t-p+1})' = x, S_t = s_l) \\ = 1 + \sum_{k=1}^K a_{lk} (m_k^2(x) + \sigma_k^2(x)) + b_{p-1} x_t^2 + \dots + b_1 x_{t-p+1}^2\end{aligned}\tag{3.1}$$

Now, let us focus on

$$\begin{aligned}
& \sum_{k=1}^K a_{lk}(m_k^2(x) + \sigma_k^2(x)) \\
&= \sum_{k=1}^K a_{lk} \left( \sum_{i=1}^p \alpha_{ki} x_{t-i+1} \right)^2 + \sum_{k=1}^K a_{lk} \left( \sum_{i=1}^p d_{ki} x_{t-i+1}^2 \right) + o(\|x\|^2) \\
&= \sum_{k=1}^K a_{lk} \left( \sum_{i=1}^p \alpha_{ki}^2 x_{t-i+1}^2 \right) + \sum_{k=1}^K a_{lk} \left( \sum_{i=1}^p d_{ki} x_{t-i+1}^2 \right) \\
&\quad + 2 \sum_{k=1}^K \sum_{i=1}^{p-1} \sum_{j=i+1}^p a_{lk} (\alpha_{ki} x_{t-i+1} \alpha_{kj} x_{t-j+1}) + o(\|x\|^2),
\end{aligned}$$

and considering the fact that  $2ab \leq 2|ab| \leq a^2 + b^2, \forall a, b \in \mathbb{R}$  it follows

$$\begin{aligned}
& \sum_{k=1}^K a_{lk}(m_k^2(x) + \sigma_k^2(x)) \\
&= \sum_{k=1}^K a_{lk} \left( \sum_{i=1}^p \alpha_{ki}^2 x_{t-i+1}^2 \right) + \sum_{k=1}^K a_{lk} \left( \sum_{i=1}^p d_{ki} x_{t-i+1}^2 \right) \\
&\quad + \sum_{k=1}^K \sum_{i=1}^{p-1} \sum_{j=i+1}^p a_{lk} |\alpha_{ki}| |\alpha_{kj}| (x_{t-i+1}^2 + x_{t-j+1}^2) + o(\|x\|^2) \\
&\leq \sum_{i=1}^p \sum_{k=1}^K a_{lk} \left( \alpha_{ki}^2 + d_{ki} + |\alpha_{ki}| \sum_{j=1, j \neq i}^p |\alpha_{kj}| \right) x_{t-i+1}^2 + o(\|x\|^2).
\end{aligned}$$

Back to the original problem, we have

$$\begin{aligned}
& \mathbb{E}(g(\zeta_{t+1}) | (X_t, \dots, X_{t-p+1})' = x, S_t = s_l) \\
&\leq 1 + \sum_{i=1}^{p-1} \sum_{k=1}^K a_{lk} \left( d_{ki} + |\alpha_{ki}| \sum_{j=1}^p |\alpha_{kj}| + b_{p-i} \right) x_{t-i+1}^2 \\
&\quad + \sum_{k=1}^K a_{lk} \left( d_{kp} + |\alpha_{kp}| \sum_{j=1}^p |\alpha_{kj}| \right) x_{t-p+1}^2 + o(\|x\|^2)
\end{aligned}$$

which we can rewrite in the following way

$$\begin{aligned}
& \mathbb{E}(g(\zeta_{t+1}) | (X_t, \dots, X_{t-p+1})' = x, S_t = s_l) \\
&\leq 1 + \sum_{k=1}^K a_{lk} \left( d_{kp} + |\alpha_{kp}| \sum_{j=1}^p |\alpha_{kj}| \right) \frac{b_1}{b_1} x_{t-p+1}^2 \\
&\quad + \sum_{i=2}^{p-1} \sum_{k=1}^K a_{lk} \left( d_{ki} + |\alpha_{ki}| \sum_{j=1}^p |\alpha_{kj}| + b_{p-i} \right) \frac{b_{p-i+1}}{b_{p-i+1}} x_{t-i+1}^2 \\
&\quad + \sum_{k=1}^K a_{lk} \left( d_{k1} + |\alpha_{k1}| \sum_{j=1}^p |\alpha_{kj}| + b_{p-1} \right) x_t^2 + o(\|x\|^2).
\end{aligned}$$

Hence,

$$\mathbb{E}(g(\zeta_{t+1}) | (X_t, \dots, X_{t-p+1})' = x, S_t = s_l) < g(\zeta_t).$$

If for example by assumption A.12 we choose  $b_{p-1}, \dots, b_1$  such that

$$\begin{aligned} \sum_{k=1}^K a_{lk} \left( d_{k1} + |\alpha_{k1}| \sum_{j=1}^p |\alpha_{kj}| + b_{p-1} \right) &< 1, \\ \sum_{k=1}^K a_{lk} \left( d_{ki} + |\alpha_{ki}| \sum_{j=1}^p |\alpha_{kj}| + b_{p-i} \right) &< 1 \quad \text{for } i = 2, \dots, p-1 \\ \text{and } \frac{\sum_{k=1}^K a_{lk} \left( d_{kp} + |\alpha_{kp}| \sum_{j=1}^p |\alpha_{kj}| \right)}{b_1} &< 1, \end{aligned}$$

we obtain

$$\begin{aligned} \mathbb{E}(g(\zeta_{t+1}) | (X_t, \dots, X_{t-p+1})' = x, S_t = s_l) \\ \leq 1 + \gamma_1 x_t^2 + \gamma_2 b_{p-1} x_{t-1}^2 + \dots + \gamma_p b_1 x_{t-p+1}^2 \end{aligned}$$

with  $\gamma_1, \gamma_2, \dots, \gamma_p < 1$  which leads us to the conclusion.  $\blacksquare$

Lee and Shin [5] in their Theorem 2.1 require, instead of A.12, a condition which can be rewritten in our notation as

$$\sum_{i=1}^p \max_{l \in \{1, \dots, K\}} \sum_{k=1}^K a_{lk} \left( \sum_{j=1}^p |\alpha_{ki}| |\alpha_{kj}| + d_{ki} \right) < 1.$$

For  $p > 1$  this condition is obviously stronger than A.12 since

$$\max_{l \in \{1, \dots, K\}} \sum_{i=1}^p \sum_{k=1}^K a_{lk} (d_{ki} + |\alpha_{ki}| \sum_{j=1}^p |\alpha_{kj}|) \leq \sum_{i=1}^p \max_{l \in \{1, \dots, K\}} \sum_{k=1}^K a_{lk} \left( \sum_{j=1}^p |\alpha_{ki}| |\alpha_{kj}| + d_{ki} \right).$$

As in the case  $p = 1$ , we may replace assumption A.7 by A.9 and instead of A.11 we now require

**A. 13** *There exist  $\alpha_k, d_k \in \mathbb{R}^p$  with  $d_{ki} \geq 0$ ,  $i = 1, \dots, p$  such that as  $\|u\| \rightarrow \infty$ ,*

$$|m_k(u)|^\alpha = \sum_{i=1}^p |\alpha_{ki}| |u_i|^\alpha + o(\|u\|^\alpha)$$

and

$$\sigma_k^\alpha(u) = \sum_{i=1}^p d_{ki} |u_i|^\alpha + o(\|u\|^\alpha)$$

**A. 14**

$$\max_{l \in \{1, \dots, K\}} \sum_{i=1}^p \left( \sum_{k=1}^K a_{lk} (\mathbb{E}|\epsilon_t|^\alpha d_{ki} + |\alpha_{ki}|) \right) < 1$$

**Theorem 4** *Under assumptions A.1 to A.6, A.9, A.13 and A.14,  $\{\zeta_t\}$  is geometrically ergodic.*

**Proof:** We start the proof similar as in the proof of Theorem 3. However, we now consider

$$g(\zeta_t) = 1 + |X_t|^\alpha + b_{p-1} |X_{t-1}|^\alpha + \dots + b_1 |X_{t-p+1}|^\alpha$$



with  $b_{p-1}, \dots, b_1 > 0$  to be suitably chosen as previously. Let us consider  $x$  and  $s_l$  as defined in the proof of the previous theorem, we can now focus on

$$\begin{aligned}
& \mathbb{E}(g(\zeta_{t+1}) | (X_t, \dots, X_{t-p+1})' = x, S_t = s_l) \\
& \leq 1 + \sum_{k=1}^K a_{lk} (|m_k(x)|^\alpha + \sigma_k^\alpha(x) \mathbb{E}|\epsilon_t|^\alpha + b_{p-1}|x_t|^\alpha + \dots + b_1|x_{t-p+2}|^\alpha) \\
& \leq 1 + \sum_{i=1}^{p-1} \left( \sum_{k=1}^K a_{lk} (\mathbb{E}|\epsilon_t|^\alpha d_{ki} + \alpha_{ki}) + b_{p-i} \right) |x_{t-i+1}|^\alpha \\
& \quad + \sum_{k=1}^K a_{lk} (\mathbb{E}|\epsilon_t|^\alpha d_{kp} + \alpha_{kp}) |x_{t-p+1}|^\alpha + o(\|u\|^\alpha)
\end{aligned}$$

and the conclusion follows in the same way as in Theorem 3. ■

## 4 Concluding remarks

We have considered Conditional Heteroscedastic Autoregressive Mixture of Experts (CHARME) models, a form of hidden Markov model with nonlinear autoregressive-ARCH components, which can be useful in, e.g. financial econometrics. We have proven geometric ergodicity of  $\zeta_t = (S_t, X_t, \dots, X_{t-p+1})'$  under several sets of conditions. If the process  $\{X_t\}$  is also strictly stationary it is well known that this implies that  $\{X_t\}$  is absolutely regular with geometric decreasing rate, which gives a very useful condition for deriving limit theorems like the central limit theorem.

Geometric ergodicity of  $\zeta_t$  can clearly be obtained even if some of the underlying dynamics taken on their own are not geometric ergodic or even stationary, provided the probability to go from a stable dynamic to a non stable dynamic is low enough and the probability to move from a non stable dynamic to a stable dynamic is large enough.

The different sets of conditions are based on the existence of some moments for  $\epsilon_t$  and on the behavior of  $m_k(x)$  and  $\sigma_k(x)$  as  $\|x\| \rightarrow \infty$ ; some sets of conditions relaxing the moment assumptions for the  $\epsilon_t$  are at the expenses of stronger conditions for the  $m_k$  and  $\sigma_k$  at the tails. Notice that our conditions for the  $m_k$  and  $\sigma_k$  are easier to verify than some kind of sub linearity conditions or Lyapounov exponents conditions.

It would be quite straightforward, at least in the case  $p = 1$ , to relax the assumption of an everywhere positive density for  $\epsilon_t$  in favor of a positive density on a big enough compact in a similar manner as in Franke, Neumann and Stockis [6].

Subsequent work based on our results can lead to sufficient conditions for the existence of moments for  $X_t$  or for the existence of limit theorems for  $X_t$ .

This work provides stability properties which will be useful when dealing with the problem of estimating the  $m_k$  and  $\sigma_k$ ,  $k = 1, \dots, K$  in presence of real data. This practical aspect is the subject of forthcoming publication.

# References

- [1] MASRY, E. AND TJOSTHEIM, D. (1995). Nonparametric estimation and identification of nonlinear ARCH time series: Strong convergence and asymptotic normality. *Econometric theory* **95**, 258–289
- [2] HÄRDLE, W. AND TSYBAKOV, A. (1997). Local polynomial estimation of volatility function. *Journal of Econometrics* **81**, 221–245
- [3] BHATTACHARYA, R. AND LEE, C. (1995). On geometric ergodicity of nonlinear autoregressive models. *Statistics & Probability Letters* **22**, 439–440
- [4] BHATTACHARYA, R. AND LEE, C. (1999). Erratum: On geometric ergodicity of nonlinear autoregressive models. *Statistics & Probability Letters* **41**, 311–315
- [5] LEE, O. AND SHIN, D.W. (2004). On geometric ergodicity of an AR-ARCH type process with Markov Switching. *Journal of Korean Mathematical Society* **41**, 309–318
- [6] FRANKE, J., NEUMANN, M.H. AND STOCKIS, J.-P. (2004). Bootstrapping nonparametric estimators of the volatility function. *Journal of Econometrics* **118**, 189–218
- [7] MEYN, S.P. AND TWEEDIE, R.L. (1993). *Markov Chain and Stochastic Stability*, Springer-Verlag, London.
- [8] MÜLLER, K.R., KOHLMORGEN, J., PAWELZIK, K. AND NON-MEMBERS (1995). Analysis of switching Dynamics with competing neural networks. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, E78-A* **10**, 1306–1315
- [9] FRANCO, C., ROUSSIGNOL, M. AND ZAKOIAN, J.-M. (2001). Conditional Heteroskedasticity Driven by Hidden Markov Chains. *Journal of Time Series Analysis* **10**, 197–220
- [10] JIANG, W. AND TANNER, M.A. (1999). On the Identifiability of Mixtures-of-Experts. *Neural Networks* **12**, 197–220
- [11] FRANCO, C. AND ROUSSIGNOL, M. (1997). On white noises driven by Hidden Markov Chains. *Journal of Time Series Analysis* **18**, 553–578
- [12] JACOB, R.A., JORDAN, M.I., NOWLAN, S.J. AND HINTON, G.E. (1991). Adaptive mixtures of Local experts. *Neural Computation* **3**, 79–87
- [13] JORDAN, M.I. AND JACOB, R.A. (1994). Hierarchical mixtures of experts and the EM Algorithm. *Neural Computation* **6**, 181–214
- [14] CARVALHO, A.X. AND TANNER, M.A. (2005). Mixtures of experts of autoregressive time series: asymptotic normality and model specification. *IEEE Trans Neural Computation* **16**, 39–56
- [15] HAFNER, C. (1998). *Nonlinear Time Series Analysis with Applications to Foreign Exchange Rate Volatility*, Physica Verlag, Heidelberg.
- [16] TADJUIDJE KAMGAING, J. (2005). Competing Neural Networks as Model for Non stationary Financial Time Series. Doctoral Thesis, University of Kaiserslautern.