



HAL
open science

Deep neural network-based CHARME models with infinite memory

José G. Gómez-García, Jalal Fadili, Christophe Chesneau

► **To cite this version:**

José G. Gómez-García, Jalal Fadili, Christophe Chesneau. Deep neural network-based CHARME models with infinite memory. Data Science Summer School (DS3), Jun 2019, Paris - Saclay, France. hal-02518028

HAL Id: hal-02518028

<https://hal.science/hal-02518028>

Submitted on 24 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Deep neural network-based CHARME models with infinite memory

José G. Gómez-García⁽¹⁾, Jalal Fadili⁽²⁾ and Christophe Chesneau⁽¹⁾

(1) Lab. of Mathematics Nicolas Oresme (LMNO), Université de Caen Normandie
(2) Ecole Nationale Supérieure d'Ingénieurs de Caen (ENSICAEN)



Abstract

We consider a model called CHARME (Conditional Heteroscedastic Autoregressive Mixture of Experts), a class of generalized mixture of nonlinear nonparametric AR-ARCH times series. Under certain Lipschitz-type conditions on the autoregressive and volatility functions, we prove that this model is τ -weakly dependent in the sense of Dedecker & Prieur (2004) [1], and therefore, ergodic and stationary. This result forms the theoretical basis for deriving an asymptotic theory of the underlying nonparametric estimation. As application, for the case of a single expert, we use the universal approximation property of the neural networks in order to develop an estimation theory for the autoregressive function by deep neural networks, where the consistency of the estimator of neurons and bias are guaranteed.

The model

Let $(E, \|\cdot\|)$ be a Banach space. The conditional heteroscedastic p -autoregressive mixture of experts (CHARME(p)) model, with values in E , is defined by:

$$X_t = \sum_{k=1}^K \xi_t^{(k)} [f_k(X_{t-1}, \dots, X_{t-p}) + g_k(X_{t-1}, \dots, X_{t-p}) \epsilon_t] \quad t \in \mathbb{Z}, \quad (1)$$

where

- ▶ $f_k : (E^p, \mathcal{E}^{\otimes p}) \rightarrow (E, \mathcal{E})$ and $g_k : (E^p, \mathcal{E}^{\otimes p}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, with $k \in [K] := \{1, 2, \dots, K\}$, are arbitrary unknown functions,
- ▶ $(\epsilon_t)_t$ are E -valued independent identically distributed (iid) zero-mean innovations, and
- ▶ $\xi_t^{(k)} = \mathbb{I}_{\{Q_t=k\}}$, where $(Q_t)_t$ is an iid sequence with values in the finite set of states $[K]$, which is independent of the innovations $(\epsilon_t)_t$.

In particular, if $p = \infty$, we call this model CHARME with infinite memory (CHARME(∞)).

Weak dependence

Let $(E, \|\cdot\|)$ be a Banach space and let $h : E \rightarrow \mathbb{R}$. We define $\|h\|_\infty = \sup_{x \in E} |h(x)|$ and

$$\text{Lip}(h) = \sup_{x \neq y} \frac{|h(x) - h(y)|}{\|x - y\|}.$$

Moreover, we denote by $\Lambda_1(E) := \{h : E \rightarrow \mathbb{R} : \text{Lip}(h) \leq 1\}$.

The appropriate notion of weak dependence for the CHARME model was introduced in [1]. It is based on the concept of the coefficient τ defined below.

Def. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, \mathcal{M} a σ -sub-algebra of \mathcal{A} and X a random variable with values in E such that $\|X\|_1 < \infty$. The coefficient τ is defined as

$$\tau(\mathcal{M}, X) = \left\| \sup \left\{ \left| \int h(x) \mathbb{P}_{X|\mathcal{M}}(dx) - \int h(x) \mathbb{P}_X(dx) \right| : h \in \Lambda_1(E) \right\} \right\|_1.$$

Using the definition of this τ coefficient with the σ -algebra

$$\mathcal{M}_p = \sigma(X_t, t \leq p)$$

and the norm $\|x - y\| = \|x_1 - y_1\| + \dots + \|x_k - y_k\|$ on E^k , we can assess the dependence between the past of the sequence $(X_t)_{t \in \mathbb{Z}}$ and its future k -tuples through the coefficients

$$\tau_k(r) = \max_{1 \leq l \leq k} \frac{1}{l} \sup \{ \tau(\mathcal{M}_p, (X_{j_1}, \dots, X_{j_l})) \text{ with } p+r \leq j_1 < \dots < j_l \}.$$

Finally, denoting $\tau(r) := \tau_\infty(r) = \sup_{k>0} \tau_k(r)$, the time series $(X_t)_{t \in \mathbb{Z}}$ is called τ -weakly dependent if its coefficients $\tau(r)$ tend to 0 as r tends to infinity.

Deep neural networks (DNN)

Def. Let $d, L \in \mathbb{N}$. A deep neural network (architecture) θ with input dimension d and L layers is a sequence of matrix-vector tuples

$$\theta = \left((A^{(1)}, b^{(1)}), (A^{(2)}, b^{(2)}), \dots, (A^{(L)}, b^{(L)}) \right),$$

where $A^{(l)}$ is a $N_l \times N_{l-1}$ matrix and $b^{(l)} \in \mathbb{R}^{N_l}$, with $N_0 = d$ and $N_1, \dots, N_L \in \mathbb{N}$, the number of neurons for each layer.

If θ is a deep neural network architecture as above and if $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ is an arbitrary function, then we define the **deep neural network (DNN)** associated to θ with activation function φ as the map $f_{\theta, \varphi} : \mathbb{R}^d \rightarrow \mathbb{R}^{N_L}$ such that

$$f_{\theta, \varphi}(x) = x_L,$$

where x_L results from the following scheme:

$$\begin{aligned} x_0 &:= x, \\ x_l &:= \varphi(A_l x_{l-1} + b_l), \quad \text{for } l = 1, \dots, L-1, \\ x_L &:= A_L x_{L-1} + b_L, \end{aligned}$$

where φ acts componentwise, i.e., for $y = (y_1, \dots, y_N) \in \mathbb{R}^N$, $\varphi(y) = (\varphi(y_1), \dots, \varphi(y_N))$.

Theorem (Stationarity of CHARME models)

Let $E^\infty := \{(x_k)_{k>0} \in E^\mathbb{N} : x_k = \mathbf{0} \text{ for } k > N, \text{ for some } N \in \mathbb{N}^*\}$ endowed with its product σ -algebra $\mathcal{E}^{\otimes \mathbb{N}}$.

Consider the CHARME(∞) model and denote $\pi_k = \mathbb{P}(Q_0 = k)$, with $k = 1, \dots, K$. Assume that there exist non-negative real sequences $(a_i^{(k)})_{i \geq 1}$ and $(b_i^{(k)})_{i \geq 1}$, for $k = 1, 2, \dots, K$, such that for any $x, y \in E^\infty$,

$$\begin{aligned} \|f_k(x) - f_k(y)\| &\leq \sum_{i=1}^{\infty} a_i^{(k)} \|x_i - y_i\|, \\ |g_k(x) - g_k(y)| &\leq \sum_{i=1}^{\infty} b_i^{(k)} \|x_i - y_i\|, \quad k = 1, \dots, K. \end{aligned} \quad (2)$$

Denote $a(m) = 2^{m-1} \sum_{k=1}^K \pi_k (A_k^m + B_k^m \|\epsilon_0\|_m^m)$, where $A_k = \sum_{i=1}^{\infty} a_i^{(k)}$ and $B_k = \sum_{i=1}^{\infty} b_i^{(k)}$. Then,

1. if $a(\mathbf{1}) < 1$, there exists a τ -weakly dependent strictly stationary solution $(X_t)_{t \in \mathbb{Z}}$ of (1, with $p = \infty$) which belongs to \mathbb{L}^1 , and such that

$$\tau(r) \leq 2 \frac{\mu_1}{1-a} \inf_{1 \leq s \leq r} \left(a^{r/s} + \frac{1}{1-a} \sum_{i=s+1}^{\infty} a_i \right) \xrightarrow{r \rightarrow \infty} 0, \quad (3)$$

where $\mu_1 = \sum_{k=1}^K \pi_k (\|f_k(\mathbf{0})\| + |g_k(\mathbf{0})| \|\epsilon_0\|_1)$ and $a_i = \sum_{k=1}^K \pi_k (a_i^{(k)} + b_i^{(k)} \|\epsilon_0\|_1)$.

2. if moreover $a(m) < 1$ for some $m \geq 1$, the stat. solution belong to \mathbb{L}^m .

Application-Example

Suppose that $(X_t)_t$ is a time series such that

$$X_t = f_{\theta, \varphi}(X_{t-1}, \dots, X_{t-p}) + \epsilon_t, \quad (4)$$

where $f_{\theta, \varphi} : \mathbb{R}^p \rightarrow \mathbb{R}$ is a DNN with parameter

$$\theta = \left((A^{(1)}, b^{(1)}), (A^{(2)}, b^{(2)}), \dots, (A^{(L)}, b^{(L)}) \right) \in \prod_{l=1}^L (\mathbb{M}_{N_l \times N_{l-1}}(\mathbb{R}) \times \mathbb{M}_{N_l \times 1}(\mathbb{R}))$$

and Lipschitz activation function φ . Then, if $\|\epsilon_0\|_1 < \infty$ and

$$\tilde{a} = (\text{Lip}(\varphi))^{L-1} \sum_{(j_0, \dots, j_L) \in \prod_{l=0}^L [M]} \prod_{l=1}^L |a_{j_l j_{l-1}}^{(l)}| < 1, \quad (5)$$

the time series $(X_t)_t$ is stationary and belongs to \mathbb{L}^1 .

Some conclusions

- ▶ Under the assumptions of Theorem, the stationary solution of the CHARME(∞) model can be represented as a causal Bernoulli shift (see [2]). Moreover, this solution is the unique causal Bernoulli shift solution of the model. Therefore, the solution is automatically an ergodic process.
- ▶ Under the assumptions of Theorem and additional classical conditions, the least squares estimator for the DNN-based CHARME(p) model (1) is consistent.

Future work

- ▶ To provide a central limit theorem for the least squares estimator of the DNN-based CHARME(p) model (1) under flexible conditions.

References

- [1] J. Dedecker and C. Prieur. Coupling for τ -dependent sequences and applications. *Journal of Theoretical Probability*, 17(4):861–885, October 2004.
- [2] P. Doukhan and O. Wintenberger. Weakly dependent chains with infinite memory. *Stochastic Processes and their Applications*, 118:1997–2013, 2008.
- [3] JP. Stockis, J. Franke, and J. Tadjuidje Kamgaing. On geometric ergodicity of CHARME models. *Journal of Time Series Analysis*, 31(3):141–152, 2010.
- [4] DWK. Andrews. Non strong mixing autoregressive processes. *Journal of Applied Probability*, 21:930–934, 1984.

Acknowledgments

This work has been developed with the help of the Normandy Region.

Contact Information

- ▶ Web: <https://gomezgarcia.users.lmno.cnrs.fr/>
- ▶ Email: jose3g@gmail.com; jose-gregorio.gomez-garcia@unicaen.fr