
Sur la stabilité des modèles CHARME

José Gregorio GOMEZ-GARCIA

jose.gomez-garcia@agroparistech.fr

Collaboration avec **J. Fadili** (ENSICAEN) et **C. Chesneau** (UNICAEN)

Processus markoviens, semi-markoviens et leurs applications

5 juin 2023 - IMAG Montpellier, France

Introduction

τ -dépendance

Stabilité

Application et Simulations

Perspectives

Introduction

τ -dépendance

Stabilité

Application et Simulations

Perspectives

Introduction : Motivation

Modèles classiques de séries temporelles sont de la forme :

$$X_t = f(X_{t-1}, \dots, X_{t-p}, \theta^0) + g(X_{t-1}, \dots, X_{t-p}, \lambda^0) \epsilon_t \quad t \in \mathbb{Z}, \quad (1)$$

où les fonctions f et g sont inconnues et $(\epsilon_t)_t$ sont des v.a. centrées i.i.d.

Introduction : Motivation

Modèles classiques de séries temporelles sont de la forme :

$$X_t = f(X_{t-1}, \dots, X_{t-p}, \theta^0) + g(X_{t-1}, \dots, X_{t-p}, \lambda^0) \epsilon_t \quad t \in \mathbb{Z}, \quad (1)$$

où les fonctions f et g sont inconnues et $(\epsilon_t)_t$ sont des v.a. centrées i.i.d.

Remarque

- ▶ Dans la pratique, il n'est pas très réaliste de considérer les mêmes fonctions f et g pour chaque instant de temps t . **Exemple** : les signaux d'EEG.

Introduction : Motivation

Modèles classiques de séries temporelles sont de la forme :

$$X_t = f(X_{t-1}, \dots, X_{t-p}, \theta^0) + g(X_{t-1}, \dots, X_{t-p}, \lambda^0) \epsilon_t \quad t \in \mathbb{Z}, \quad (1)$$

où les fonctions f et g sont inconnues et $(\epsilon_t)_t$ sont des v.a. centrées i.i.d.

Remarque

- ▶ Dans la pratique, il n'est pas très réaliste de considérer les mêmes fonctions f et g pour chaque instant de temps t . *Exemple* : les signaux d'EEG.
- ▶ En particulier, si ces fonctions f et g changent lentement dans le temps, on peut supposer la *stationnarité locale* (Dahlhaus, 2012).

Introduction : Motivation

Modèles classiques de séries temporelles sont de la forme :

$$X_t = f(X_{t-1}, \dots, X_{t-p}, \theta^0) + g(X_{t-1}, \dots, X_{t-p}, \lambda^0) \epsilon_t \quad t \in \mathbb{Z}, \quad (1)$$

où les fonctions f et g sont inconnues et $(\epsilon_t)_t$ sont des v.a. centrées i.i.d.

Remarque

- ▶ Dans la pratique, il n'est pas très réaliste de considérer les mêmes fonctions f et g pour chaque instant de temps t . *Exemple* : les signaux d'EEG.
- ▶ En particulier, si ces fonctions f et g changent lentement dans le temps, on peut supposer la *stationnarité locale* (Dahlhaus, 2012).
 - La théorie est assez développée
 - Méthodes d'estimation sont basées sur des estimateurs de processus stationnaires localement dans le temps.

Ces méthodes d'estimation ne fonctionnent pas bien si la structure de la série temporelle contient des changements plus ou moins brusques

Introduction : Motivation

Modèles classiques de séries temporelles sont de la forme :

$$X_t = f(X_{t-1}, \dots, X_{t-p}, \theta^0) + g(X_{t-1}, \dots, X_{t-p}, \lambda^0) \epsilon_t \quad t \in \mathbb{Z}, \quad (1)$$

où les fonctions f et g sont inconnues et $(\epsilon_t)_t$ sont des v.a. centrées i.i.d.

Remarque

- ▶ Dans la pratique, il n'est pas très réaliste de considérer les mêmes fonctions f et g pour chaque instant de temps t . *Exemple* : les signaux d'EEG.
- ▶ En particulier, si ces fonctions f et g changent lentement dans le temps, on peut supposer la *stationnarité locale* (Dahlhaus, 2012).
 - La théorie est assez développée
 - Méthodes d'estimation basées sur des estimateurs de processus stationnaires localement dans le temps.

Ces méthodes d'estimation ne fonctionnent pas bien si la structure de la série temporelle contient des changements plus ou moins brusques

Nous considérons une classe plus générale de modèles (non)paramétriques (appelé CHARME) qui s'adapte aux situations où ces phases explosives sont aussi incluses.

Introduction : le modèle

$(\mathcal{F}, \|\cdot\|)$ espace de Banach.

Modèle **CHARME**(p) (Conditional Heteroscedastic p -Autoregressive Mixture of Experts) à valeurs dans \mathcal{F} :

$$X_t = \sum_{k=1}^K \xi_t^{(k)} (f_k(X_{t-1}, \dots, X_{t-p}) + g_k(X_{t-1}, \dots, X_{t-p})\epsilon_t) \quad t \in \mathbb{Z}, \quad (2)$$

Introduction : le modèle

$(\mathcal{F}, \|\cdot\|)$ espace de Banach.

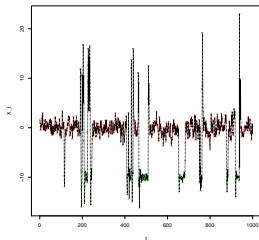
Modèle **CHARME**(p) (Conditional Heteroscedastic p -Autoregressive Mixture of Experts) à valeurs dans \mathcal{F} :

$$X_t = \sum_{k=1}^K \xi_t^{(k)} (f_k(X_{t-1}, \dots, X_{t-p}) + g_k(X_{t-1}, \dots, X_{t-p})\epsilon_t) \quad t \in \mathbb{Z}, \quad (2)$$

où

- $f_k : \mathcal{F}^p \rightarrow \mathcal{F}$ (fonction d'autorégression)
- $g_k : \mathcal{F}^p \rightarrow \mathbb{R}$ (fonction de volatilité)
- $(\epsilon_t)_t$: innovations centrées iid
- $\xi_t^{(k)} := \mathbb{I}_{\{R_t=k\}}$
- $(R_t)_t$: séquence stationnaire dans $\{1, \dots, K\}$, indépendante de $(\epsilon_t)_{t \in \mathbb{Z}}$.

Dans la suite, nous posons $\pi_k = \mathbb{P}(R_0 = k)$



Introduction : le modèle

Modèle (2) peut être étendu au cas $p = \infty$. On l'appelle modèle **CHARME à mémoire infinie** et on le note par **CHARME(∞)**.

Dans ce cadre, les fonctions f_k and g_k seront définies dans les sous-ensemble de $\mathcal{F}^{\mathbb{N}}$:

$$\mathcal{F}^{\infty} := \left\{ (x_k)_{k>0} \in \mathcal{F}^{\mathbb{N}} : x_k = 0 \text{ pour } k > N, \text{ pour tout } N \in \mathbb{N}^* \right\}.$$

Introduction : le modèle

Modèle (2) peut être étendu au cas $p = \infty$. On l'appelle modèle **CHARME à mémoire infinie** et on le note par **CHARME(∞)**.

Dans ce cadre, les fonctions f_k and g_k seront définies dans les sous-ensemble de $\mathcal{F}^{\mathbb{N}}$:

$$\mathcal{F}^{\infty} := \left\{ (x_k)_{k>0} \in \mathcal{F}^{\mathbb{N}} : x_k = 0 \text{ pour } k > N, \text{ pour tout } N \in \mathbb{N}^* \right\}.$$

Objetif

- ▶ Stabilité : stationnarité, ergodicité et τ -dépendance
- ▶ Estimation : consistance et normalité asymptotique

Introduction : le modèle

Modèle (2) peut être étendu au cas $p = \infty$. On l'appelle modèle **CHARME à mémoire infinie** et on le note par **CHARME(∞)**.

Dans ce cadre, les fonctions f_k and g_k seront définies dans les sous-ensemble de $\mathcal{F}^{\mathbb{N}}$:

$$\mathcal{F}^{\infty} := \left\{ (x_k)_{k>0} \in \mathcal{F}^{\mathbb{N}} : x_k = 0 \text{ pour } k > N, \text{ pour tout } N \in \mathbb{N}^* \right\}.$$

Objetif

- ▶ Stabilité : stationnarité, ergodicité et τ -dépendance
- ▶ Estimation : consistance et normalité asymptotique

Exemple : stabilité et estimation des paramètres (poids et biais) des réseaux de neurones d'un modèle CHARME basé sur des réseaux de neurones.

Introduction

τ -dépendance

Stabilité

Application et Simulations

Perspectives

τ - faible dépendance (Dedecker & Prieur, 2004)

- $(\Omega, \mathcal{A}, \mathbb{P})$: espace de probabilité
- \mathcal{M} : σ -sous-tribu de \mathcal{A}
- X : \mathcal{F} -v.a telle que $\|X\|_1 < \infty$.

Le coefficient τ est défini comme suit

$$\begin{aligned} \tau(\mathcal{M}, X) &= \mathbb{E} |\sup \{ |\mathbb{E}[h(X)|\mathcal{M}] - \mathbb{E}[h(X)]| : h \text{ t.q } \text{Lip}(h) \leq 1 \}| \\ &= \mathbb{E} \left| \sup \left\{ \left| \int_{\mathcal{F}} h(x) \mathbb{P}_{X|\mathcal{M}}(dx) - \int_{\mathcal{F}} h(x) \mathbb{P}_X(dx) \right| : h \text{ t.q } \text{Lip}(h) \leq 1 \right\} \right| \end{aligned}$$

τ - faible dépendance (Dedecker & Prieur, 2004)

- $(\Omega, \mathcal{A}, \mathbb{P})$: espace de probabilité
- \mathcal{M} : σ -sous-tribu de \mathcal{A}
- X : \mathcal{F} -v.a telle que $\|X\|_1 < \infty$.

Le coefficient τ est défini comme suit

$$\begin{aligned} \tau(\mathcal{M}, X) &= \mathbb{E} |\sup \{ |\mathbb{E}[h(X)|\mathcal{M}] - \mathbb{E}[h(X)]| : h \text{ t.q } \text{Lip}(h) \leq 1 \}| \\ &= \mathbb{E} \left| \sup \left\{ \left| \int_{\mathcal{F}} h(x) \mathbb{P}_{X|\mathcal{M}}(dx) - \int_{\mathcal{F}} h(x) \mathbb{P}_X(dx) \right| : h \text{ t.q } \text{Lip}(h) \leq 1 \right\} \right| \end{aligned}$$

Pour une série temporelle $(X_t)_{t \in \mathbb{Z}}$, considérons

- la tribu $\mathcal{M}_p = \sigma(X_t, t \leq p)$
- la norme $\|x - y\| = \|x_1 - y_1\| + \dots + \|x_k - y_k\|$ sur \mathcal{F}^k .

La dépendance entre le passé de la série temporelle $(X_t)_{t \in \mathbb{Z}}$ et son future k -tuples peut être évalué avec le coefficient :

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

τ - faible dépendance (Dedecker & Prieur, 2004)

- $(\Omega, \mathcal{A}, \mathbb{P})$: espace de probabilité
- \mathcal{M} : σ -sous-tribu de \mathcal{A}
- X : \mathcal{F} -v.a telle que $\|X\|_1 < \infty$.

Le coefficient τ est défini comme suit

$$\begin{aligned} \tau(\mathcal{M}, X) &= \mathbb{E} |\sup \{ |\mathbb{E}[h(X)|\mathcal{M}] - \mathbb{E}[h(X)]| : h \text{ t.q } \text{Lip}(h) \leq 1 \}| \\ &= \mathbb{E} \left| \sup \left\{ \left| \int_{\mathcal{F}} h(x) \mathbb{P}_{X|\mathcal{M}}(dx) - \int_{\mathcal{F}} h(x) \mathbb{P}_X(dx) \right| : h \text{ t.q } \text{Lip}(h) \leq 1 \right\} \right| \end{aligned}$$

Pour une série temporelle $(X_t)_{t \in \mathbb{Z}}$, considérons

- la tribu $\mathcal{M}_p = \sigma(X_t, t \leq p)$
- la norme $\|x - y\| = \|x_1 - y_1\| + \dots + \|x_k - y_k\|$ sur \mathcal{F}^k .

La dépendance entre le passé de la série temporelle $(X_t)_{t \in \mathbb{Z}}$ et son future k -tuples peut être évalué avec le coefficient :

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

On dit que $(X_t)_{t \in \mathbb{Z}}$ est τ -faiblement dépendante si

$$\tau(r) := \tau_\infty(r) = \sup_{k>0} \tau_k(r) \xrightarrow{r \rightarrow \infty} 0$$

τ -faible dépendance : Quelques propriétés

1. Si Y est une v.a. suivant la même loi que X et indépendante de \mathcal{M} , alors

$$\tau(\mathcal{M}, X) \leq \|X - Y\|_1$$

τ -faible dépendance : Quelques propriétés

1. Si Y est une v.a. suivant la même loi que X et indépendante de \mathcal{M} , alors

$$\tau(\mathcal{M}, X) \leq \|X - Y\|_1$$

2. Si $(\Omega, \mathcal{A}, \mathbb{P})$ est assez riche, alors il existe une v.a. X^* suivant la même loi que X et indépendante de \mathcal{M} telle que

$$\|X - X^*\|_1 = \tau(\mathcal{M}, X)$$

Introduction

τ -dépendance

Stabilité

Application et Simulations

Perspectives

Stationnarité du modèle CHARME

Dans la suite, nous considérons les variables du processus de régime $(R_t)_{t \in \mathbb{Z}}$ indépendantes.

Stationnarité du modèle CHARME

Dans la suite, nous considérons les variables du processus de régime $(R_t)_{t \in \mathbb{Z}}$ indépendantes.

Théorème (JGGG et al. 2023)

Pour le modèle $\text{CHARME}(\infty)$, supposons qu'il existe des suites non-négatives $(a_i^{(k)})_{i \geq 1, k \in [K]}$ et $(b_i^{(k)})_{i \geq 1, k \in [K]}$ telles que, pour tout $x, y \in \mathcal{F}^\infty$ et tout $k \in [K]$,

$$\begin{aligned} \|f_k(x) - f_k(y)\| &\leq \sum_{i=1}^{\infty} a_i^{(k)} \|x_i - y_i\| \\ |g_k(x) - g_k(y)| &\leq \sum_{i=1}^{\infty} b_i^{(k)} \|x_i - y_i\|. \end{aligned} \tag{3}$$

Stationnarité du modèle CHARME

Dans la suite, nous considérons les variables du processus de régime $(R_t)_{t \in \mathbb{Z}}$ indépendantes.

Théorème (JGGG et al. 2023)

Pour le modèle $\text{CHARME}(\infty)$, supposons qu'il existe des suites non-négatives $(a_i^{(k)})_{i \geq 1, k \in [K]}$ et $(b_i^{(k)})_{i \geq 1, k \in [K]}$ telles que, pour tout $x, y \in \mathcal{F}^\infty$ et tout $k \in [K]$,

$$\begin{aligned} \|f_k(x) - f_k(y)\| &\leq \sum_{i=1}^{\infty} a_i^{(k)} \|x_i - y_i\| \\ |g_k(x) - g_k(y)| &\leq \sum_{i=1}^{\infty} b_i^{(k)} \|x_i - y_i\|. \end{aligned} \quad (3)$$

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

Stationnarité du modèle CHARME

Dans la suite, nous considérons les variables du processus de régime $(R_t)_{t \in \mathbb{Z}}$ indépendantes.

Théorème (JGGG et al. 2023)

Pour le modèle $\text{CHARME}(\infty)$, supposons qu'il existe des suites non-négatives $(a_i^{(k)})_{i \geq 1, k \in [K]}$ et $(b_i^{(k)})_{i \geq 1, k \in [K]}$ telles que, pour tout $x, y \in \mathcal{F}^\infty$ et tout $k \in [K]$,

$$\begin{aligned} \|f_k(x) - f_k(y)\| &\leq \sum_{i=1}^{\infty} a_i^{(k)} \|x_i - y_i\| \\ |g_k(x) - g_k(y)| &\leq \sum_{i=1}^{\infty} b_i^{(k)} \|x_i - y_i\|. \end{aligned} \quad (3)$$

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

1. si $c := C(1) < 1$, alors il existe une solution **strictement stationnaire et τ -faiblement dépendante** $(X_t)_{t \in \mathbb{Z}}$ du modèle $\text{CHARME}(\infty)$ appartenant à \mathbb{L}^1 , telle que

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

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

Stationnarité du modèle CHARME

Dans la suite, nous considérons les variables du processus de régime $(R_t)_{t \in \mathbb{Z}}$ indépendantes.

Théorème (JGGG et al. 2023)

Pour le modèle $\text{CHARME}(\infty)$, supposons qu'il existe des suites non-négatives $(a_i^{(k)})_{i \geq 1, k \in [K]}$ et $(b_i^{(k)})_{i \geq 1, k \in [K]}$ telles que, pour tout $x, y \in \mathcal{F}^\infty$ et tout $k \in [K]$,

$$\begin{aligned} \|f_k(x) - f_k(y)\| &\leq \sum_{i=1}^{\infty} a_i^{(k)} \|x_i - y_i\| \\ |g_k(x) - g_k(y)| &\leq \sum_{i=1}^{\infty} b_i^{(k)} \|x_i - y_i\|. \end{aligned} \quad (3)$$

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

1. si $c := C(1) < 1$, alors il existe une solution **strictement stationnaire et τ -faiblement dépendante** $(X_t)_{t \in \mathbb{Z}}$ du modèle $\text{CHARME}(\infty)$ appartenant à \mathbb{L}^1 , telle que

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

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

2. si en plus $C(m) < 1$ pour certain $m > 1$, alors cette solution appartient à \mathbb{L}^m .

Ergodicité et stationnarité du modèle CHARME

1. Le résultat précédent est également valable dans le cas $p < \infty$.
En effet, il suffit de prendre $a_i^{(k)} = b_i^{(k)} = 0$ pour tout $i > p$ et tout $k \in [K]$ dans les inégalités (3).

Ergodicité et stationnarité du modèle CHARME

1. Le résultat précédent est également valable dans le cas $p < \infty$.
En effet, il suffit de prendre $a_i^{(k)} = b_i^{(k)} = 0$ pour tout $i > p$ et tout $k \in [K]$ dans les inégalités (3).
2. Remarquons que le modèle CHARME(∞) peut être réécrit comme une suite de Markov $X_t = F(X_{t-1}, X_{t-2}, \dots; \tilde{\xi}_t)$, $t \in \mathbb{Z}$, à travers la fonction

$$F(x; (\xi^{(0)}, \dots, \xi^{(K)})) = \sum_{k=1}^K \xi^{(k)} \left(f_k(x) + g_k(x) \xi^{(0)} \right), \quad (4)$$

avec des innovations $\tilde{\xi}_t := (\epsilon_t, \xi_t^{(1)}, \dots, \xi_t^{(K)}) = (\epsilon_t, \xi_t) \in \mathcal{F} \times B_e$, où $B_e := \{e_1, \dots, e_K\}$ est la base canonique de \mathbb{R}^K .

Ergodicité et stationnarité du modèle CHARME

1. Le résultat précédent est également valable dans le cas $p < \infty$.
En effet, il suffit de prendre $a_i^{(k)} = b_i^{(k)} = 0$ pour tout $i > p$ et tout $k \in [K]$ dans les inégalités (3).
2. Remarquons que le modèle CHARME(∞) peut être réécrit comme une suite de Markov $X_t = F(X_{t-1}, X_{t-2}, \dots; \tilde{\xi}_t)$, $t \in \mathbb{Z}$, à travers la fonction

$$F(x; (\xi^{(0)}, \dots, \xi^{(K)})) = \sum_{k=1}^K \xi^{(k)} \left(f_k(x) + g_k(x) \xi^{(0)} \right), \quad (4)$$

avec des innovations $\tilde{\xi}_t := (\epsilon_t, \xi_t^{(1)}, \dots, \xi_t^{(K)}) = (\epsilon_t, \xi_t) \in \mathcal{F} \times B_e$,
où $B_e := \{e_1, \dots, e_K\}$ est la base canonique de \mathbb{R}^K .

- (a) Sous les hypothèses du théorème précédent, la fonction F est continue
(car $f_k(\cdot)$ et $g_k(\cdot)$ sont continues par condition (3))

Ergodicité et stationnarité du modèle CHARME

1. Le résultat précédent est également valable dans le cas $p < \infty$.
En effet, il suffit de prendre $a_i^{(k)} = b_i^{(k)} = 0$ pour tout $i > p$ et tout $k \in [K]$ dans les inégalités (3).
2. Remarquons que le modèle CHARME(∞) peut être réécrit comme une suite de Markov $X_t = F(X_{t-1}, X_{t-2}, \dots; \tilde{\xi}_t)$, $t \in \mathbb{Z}$, à travers la fonction

$$F(x; (\xi^{(0)}, \dots, \xi^{(K)})) = \sum_{k=1}^K \xi^{(k)} (f_k(x) + g_k(x)\xi^{(0)}), \quad (4)$$

avec des innovations $\tilde{\xi}_t := (\epsilon_t, \xi_t^{(1)}, \dots, \xi_t^{(K)}) = (\epsilon_t, \xi_t) \in \mathcal{F} \times B_e$,
où $B_e := \{e_1, \dots, e_K\}$ est la base canonique de \mathbb{R}^K .

- (a) Sous les hypothèses du théorème précédent, la fonction F est continue (car $f_k(\cdot)$ et $g_k(\cdot)$ sont continues par condition (3))
- (b) Il découle alors de (Doukhan et Wintenberger, 2008, Lemma 5.5) et de la complétude de \mathbb{L}^m , qu'il existe une fonction mesurable H telle que le processus CHARME(∞) peut être écrit comme

$$X_t = H(\tilde{\xi}_t, \tilde{\xi}_{t-1}, \dots) \quad (\text{Causal Bernoulli shift (CBS)})$$

Sous ces hypothèses, $(X_t)_{t \in \mathbb{Z}}$ est le seul CBS, solution à (2) avec $p = \infty$.
Donc, la solution $(X_t)_{t \in \mathbb{Z}}$ est automatiquement un processus ergodique.

Ergodicité et stationnarité du modèle CHARME

3. Stockis *et al.* (2010) montre l'ergodicité du modèle CHARME(p) avec $p < \infty$, sous réserve de multiple conditions. En particulier, il faut que $\epsilon_t \underset{iid}{\sim} f$, où f est une densité **continue et positive partout**.

Ergodicité et stationnarité du modèle CHARME

3. Stockis *et al.* (2010) montre l'ergodicité du modèle CHARME(p) avec $p < \infty$, sous réserve de multiple conditions. En particulier, il faut que $\epsilon_t \underset{iid}{\sim} f$, où f est une densité **continue et positive partout**.

★ Nous n'avons pas besoin de cette restriction ici et notre approche peut être appliqué à des processus à espace d'états discrets.

Ergodicité et stationnarité du modèle CHARME

3. Stockis *et al.* (2010) montre l'ergodicité du modèle CHARME(p) avec $p < \infty$, sous réserve de multiple conditions. En particulier, il faut que $\epsilon_t \underset{iid}{\sim} f$, où f est une densité **continue et positive partout**.

★ Nous n'avons pas besoin de cette restriction ici et notre approche peut être appliqué à des processus à espace d'états discrets.

En outre, les conditions de Stockis *et al.* étaient données pour obtenir des certaines conditions de "mixing" du processus CHARME(p) afin d'en déduire la stabilité du modèle.

Ergodicité et stationnarité du modèle CHARME

3. Stockis *et al.* (2010) montre l'ergodicité du modèle CHARME(p) avec $p < \infty$, sous réserve de multiple conditions. En particulier, il faut que $\epsilon_t \underset{iid}{\sim} f$, où f est une densité **continue et positive partout**.

★ Nous n'avons pas besoin de cette restriction ici et notre approche peut être appliqué à des processus à espace d'états discrets.

En outre, les conditions de Stockis *et al.* étaient données pour obtenir des certaines conditions de "mixing" du processus CHARME(p) afin d'en déduire la stabilité du modèle.

Cependant, remarquons que si nous prenons un modèle simple comme l'AR(1) :

$$X_t = \frac{1}{2} (X_{t-1} + \epsilon_t), \quad t \in \mathbb{Z} \quad (5)$$

avec $(\epsilon_t)_{t \in \mathbb{Z}}$ i.i.d. tel que $\mathbb{P}(\epsilon_0 = 0) = \mathbb{P}(\epsilon_0 = 1) = 1/2$, nous pouvons observer que les hypothèses de Stockis *et al.* ne sont pas satisfaites. En fait, ce modèle n'est pas mélangeant (Andrews, 1984).

Ergodicité et stationnarité du modèle CHARME

3. Stockis *et al.* (2010) montre l'ergodicité du modèle CHARME(p) avec $p < \infty$, sous réserve de multiple conditions. En particulier, il faut que $\epsilon_t \underset{iid}{\sim} f$, où f est une densité **continue et positive partout**.

★ Nous n'avons pas besoin de cette restriction ici et notre approche peut être appliqué à des processus à espace d'états discrets.

En outre, les conditions de Stockis *et al.* étaient données pour obtenir des certaines conditions de "mixing" du processus CHARME(p) afin d'en déduire la stabilité du modèle.

Cependant, remarquons que si nous prenons un modèle simple comme l'AR(1) :

$$X_t = \frac{1}{2} (X_{t-1} + \epsilon_t), \quad t \in \mathbb{Z} \quad (5)$$

avec $(\epsilon_t)_{t \in \mathbb{Z}}$ i.i.d. tel que $\mathbb{P}(\epsilon_0 = 0) = \mathbb{P}(\epsilon_0 = 1) = 1/2$, nous pouvons observer que les hypothèses de Stockis *et al.* ne sont pas satisfaites. En fait, ce modèle n'est pas mélangeant (Andrews, 1984).

- ▶ Les principales restrictions des processus mélangeants (mixing) sont les conditions de régularité demandés pour le processus $(\epsilon_t)_{t \in \mathbb{Z}}$

Ergodicité et stationnarité du modèle CHARME

3. Stockis *et al.* (2010) montre l'ergodicité du modèle CHARME(p) avec $p < \infty$, sous réserve de multiple conditions. En particulier, il faut que $\epsilon_t \underset{iid}{\sim} f$, où f est une densité **continue et positive partout**.

★ Nous n'avons pas besoin de cette restriction ici et notre approche peut être appliqué à des processus à espace d'états discrets.

En outre, les conditions de Stockis *et al.* étaient données pour obtenir des certaines conditions de "mixing" du processus CHARME(p) afin d'en déduire la stabilité du modèle.

Cependant, remarquons que si nous prenons un modèle simple comme l'AR(1) :

$$X_t = \frac{1}{2} (X_{t-1} + \epsilon_t), \quad t \in \mathbb{Z} \quad (5)$$

avec $(\epsilon_t)_{t \in \mathbb{Z}}$ i.i.d. tel que $\mathbb{P}(\epsilon_0 = 0) = \mathbb{P}(\epsilon_0 = 1) = 1/2$, nous pouvons observer que les hypothèses de Stockis *et al.* ne sont pas satisfaites. En fait, ce modèle n'est pas mélangeant (Andrews, 1984).

- ▶ Les principales restrictions des processus mélangeants (mixing) sont les conditions de régularité demandés pour le processus $(\epsilon_t)_{t \in \mathbb{Z}}$
- ▶ Ces conditions de régularité ne sont pas nécessaires dans notre approche. Par exemple, le processus (5) est τ -faiblement dépendant, avec

$$\tau(r) \leq 2^{-r} \sqrt{1/6} \quad (\text{Dedecker et Prieur, 2004}).$$

De plus,

$$C(m) = 3/4 < 1, \quad \text{pour tout } m \geq 1.$$

Introduction

τ -dépendance

Stabilité

Application et Simulations

Perspectives

Estimation de paramètres du modèle

On suppose des modèles paramétriques pour les fonctions

$$f_k(\cdot) = f_k(\cdot, \theta_k^0) \quad \text{et} \quad g_k(\cdot) = g_k(\cdot, \lambda_k^0).$$

- ▶ Consistance (sans besoin de la différentiabilité de la fonction de coût)
- ▶ TLC (hypothesis classiques)

Estimation de paramètres du modèle

On suppose des modèles paramétriques pour les fonctions

$$f_k(\cdot) = f_k(\cdot, \theta_k^0) \quad \text{et} \quad g_k(\cdot) = g_k(\cdot, \lambda_k^0).$$

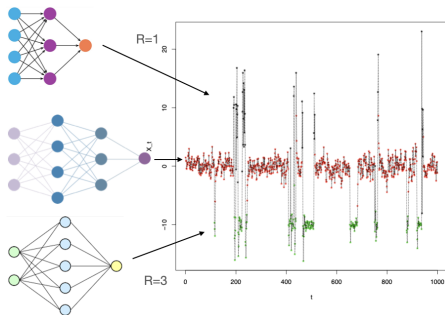
- ▶ Consistance (sans besoin de la différentiabilité de la fonction de coût)
- ▶ TLC (hypothesis classiques)

GG, Fadili & Chesneau, 2023 (Statistical Papers)

CHARME basé-sur-RN

Ici, nous travaillerons sous le contexte suivant :

- ▶ $\mathcal{F} = \mathbb{R}^d$
- ▶ le modèle CHARME(p), avec $p < \infty$, est tel que chaque fonction f_k et g_k est exactement modélisée par un réseau de neurones (RN) : CHARME basé-sur-RN.



CHARME basé-sur-RN

Plus précisément, étant donné une fonction d'activation φ , pour chaque $k \in [K]$, les fonctions f_k et g_k sont les réseaux de neurones paramétrisés par

$$\theta_k = \left((W_k^{(1)}, b_k^{(1)}), \dots, (W_k^{(L_k)}, b_k^{(L_k)}) \right) \text{ et } \lambda_k = \left((\bar{W}_k^{(1)}, \bar{b}_k^{(1)}), \dots, (\bar{W}_k^{(\bar{L}_k)}, \bar{b}_k^{(\bar{L}_k)}) \right),$$

respectivement.

Ici, $N_{k,0} = \bar{N}_{k,0} = d \cdot p$ (dim d'entrée de f_k et g_k), $N_{k,L} = d$ (dim de sortie f_k) et $\bar{N}_{k,L} = 1$ (dim de sortie g_k).

CHARME basé-sur-RN : ergodicité et stationnarité

Soient $x^\top = (x_1, \dots, x_p) \in \mathbb{R}^{dp}$ et $y^\top = (y_1, \dots, y_p) \in \mathbb{R}^{dp}$.

Découpons la matrice $W_k^{(1)}$ en p blocs-colonnes $W_{k,i}^{(1)} \in \mathbb{R}^{N_{k,1} \times d}$ telle que

$$W_k^{(1)} = \left(W_{k,1}^{(1)} \quad W_{k,2}^{(1)} \quad \dots \quad W_{k,p}^{(1)} \right).$$

Alors,

$$\begin{aligned} \|f_k(x, \theta_k) - f_k(y, \theta_k)\| &\leq (\text{Lip}(\varphi))^{L_k-1} \prod_{l=2}^{L_k} \|W_k^{(l)}\| \|W_k^{(1)}(x - y)\| \\ &= (\text{Lip}(\varphi))^{L_k-1} \prod_{l=2}^{L_k} \|W_k^{(l)}\| \left\| \sum_{i=1}^p W_{k,i}^{(1)}(x_i - y_i) \right\| \\ &\leq (\text{Lip}(\varphi))^{L_k-1} \prod_{l=2}^{L_k} \|W_k^{(l)}\| \sum_{i=1}^p \|W_{k,i}^{(1)}\| \|x_i - y_i\|, \end{aligned}$$

où $\|W_k^{(l)}\|$ est la norme spectrale de la matrice correspondante.

CHARME basé-sur-RN : ergodicité et stationnarité

De manière similaire, nous obtenons que

$$|g_k(x, \lambda_k) - g_k(y, \lambda_k)| \leq (\text{Lip}(\varphi))^{\bar{L}_k - 1} \prod_{l=2}^{\bar{L}_k} \|\bar{W}_k^{(l)}\| \sum_{i=1}^p \|\bar{W}_{k,i}^{(1)}\| \|x_i - y_i\|.$$

En s'identifiant avec (3), nous reconnaissons

$$A_k = (\text{Lip}(\varphi))^{L_k - 1} \prod_{l=2}^{L_k} \|W_k^{(l)}\| \sum_{i=1}^p \|W_{k,i}^{(1)}\|$$

$$B_k = (\text{Lip}(\varphi))^{\bar{L}_k - 1} \prod_{l=2}^{\bar{L}_k} \|\bar{W}_k^{(l)}\| \sum_{i=1}^p \|\bar{W}_{k,i}^{(1)}\|.$$

CHARME basé-sur-RN : ergodicité et stationnarité

De manière similaire, nous obtenons que

$$|g_k(x, \lambda_k) - g_k(y, \lambda_k)| \leq (\text{Lip}(\varphi))^{\bar{L}_k - 1} \prod_{l=2}^{\bar{L}_k} \|\bar{W}_k^{(l)}\| \sum_{i=1}^p \|\bar{W}_{k,i}^{(1)}\| \|x_i - y_i\|.$$

En s'identifiant avec (3), nous reconnaissons

$$A_k = (\text{Lip}(\varphi))^{L_k - 1} \prod_{l=2}^{L_k} \|W_k^{(l)}\| \sum_{i=1}^p \|W_{k,i}^{(1)}\|$$

$$B_k = (\text{Lip}(\varphi))^{\bar{L}_k - 1} \prod_{l=2}^{\bar{L}_k} \|\bar{W}_k^{(l)}\| \sum_{i=1}^p \|\bar{W}_{k,i}^{(1)}\|.$$

Donc, si $C(m) = 2^{m-1} \sum_{k=1}^K \pi_k (A_k^m + B_k^m \|\epsilon_0\|_m^m) < 1$ pour certain $m \geq 1$, il existe une solution strictement stationnaire et ergodique du modèle CHARME(p) basé-sur-RN.

CHARME basé-sur-RN : ergodicité et stationnarité

De manière similaire, nous obtenons que

$$|g_k(x, \lambda_k) - g_k(y, \lambda_k)| \leq (\text{Lip}(\varphi))^{\bar{L}_k - 1} \prod_{l=2}^{\bar{L}_k} \|\bar{W}_k^{(l)}\| \sum_{i=1}^p \|\bar{W}_{k,i}^{(1)}\| \|x_i - y_i\|.$$

En s'identifiant avec (3), nous reconnaissons

$$A_k = (\text{Lip}(\varphi))^{L_k - 1} \prod_{l=2}^{L_k} \|W_k^{(l)}\| \sum_{i=1}^p \|W_{k,i}^{(1)}\|$$

$$B_k = (\text{Lip}(\varphi))^{\bar{L}_k - 1} \prod_{l=2}^{\bar{L}_k} \|\bar{W}_k^{(l)}\| \sum_{i=1}^p \|\bar{W}_{k,i}^{(1)}\|.$$

Donc, si $C(m) = 2^{m-1} \sum_{k=1}^K \pi_k (A_k^m + B_k^m \|\epsilon_0\|_m^m) < 1$ pour certain $m \geq 1$, il existe une solution strictement stationnaire et ergodique du modèle CHARME(p) basé-sur-RN.

Remarque

Cette condition est suffisant pour garantir l'apprentissage du modèle.

Expérience 1 : apprentissage de RN pour des données générées avec un CHARME basé-sur-RN

Modèle : CHARME(30)-basé-sur-RN avec $K = 3$, $g_k = 1$ et $f_k = f_k(\cdot, \theta_k^0)$ pour $k = 1, 2, 3$ sont des RN tels que

$$\#neu(\theta_1^0) = (N_{1,0}, \dots, N_{1,5}) = (30, 50, 60, 40, 20, 1),$$

$$\#neu(\theta_2^0) = (N_{2,0}, \dots, N_{2,3}) = (30, 20, 5, 1) \text{ et}$$

$$\#neu(\theta_3^0) = (N_{3,0}, \dots, N_{3,3}) = (30, 25, 30, 1), \text{ et fonction d'activation ReLU.}$$

Expérience 1 : apprentissage de RN pour des données générées avec un CHARME basé-sur-RN

Modèle : CHARME(30)-basé-sur-RN avec $K = 3$, $g_k = 1$ et $f_k = f_k(\cdot, \theta_k^0)$ pour $k = 1, 2, 3$ sont des RN tels que

$$\#neu(\theta_1^0) = (N_{1,0}, \dots, N_{1,5}) = (30, 50, 60, 40, 20, 1),$$

$$\#neu(\theta_2^0) = (N_{2,0}, \dots, N_{2,3}) = (30, 20, 5, 1) \text{ et}$$

$$\#neu(\theta_3^0) = (N_{3,0}, \dots, N_{3,3}) = (30, 25, 30, 1), \text{ et fonction d'activation ReLU.}$$

- Nous avons pris $(\pi_1, \pi_2, \pi_3) = (0.1, 0.4, 0.5)$ et $(b_1^{(5)}, b_2^{(3)}, b_3^{(3)}) = (1, 0, -1)$, et donné des poids arbitraires des matrices $W_k^{(l)}$ tels que $C(1) < 1$. En effet, nous avons choisi les poids tels que $C(1) = 0.9882$. Ici $\epsilon \sim \mathcal{N}(0, 1)$ et $n = 10^5$.

Expérience 1 : apprentissage de RN pour des données générées avec un CHARME basé-sur-RN

Modèle : CHARME(30)-basé-sur-RN avec $K = 3$, $g_k = 1$ et $f_k = f_k(\cdot, \theta_k^0)$ pour $k = 1, 2, 3$ sont des RN tels que

$\#neu(\theta_1^0) = (N_{1,0}, \dots, N_{1,5}) = (30, 50, 60, 40, 20, 1)$,

$\#neu(\theta_2^0) = (N_{2,0}, \dots, N_{2,3}) = (30, 20, 5, 1)$ et

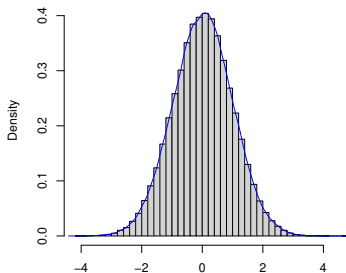
$\#neu(\theta_3^0) = (N_{3,0}, \dots, N_{3,3}) = (30, 25, 30, 1)$, et fonction d'activation ReLU.

- Nous avons pris $(\pi_1, \pi_2, \pi_3) = (0.1, 0.4, 0.5)$ et $(b_1^{(5)}, b_2^{(3)}, b_3^{(3)}) = (1, 0, -1)$, et donné des poids arbitraires des matrices $W_k^{(l)}$ tels que $C(1) < 1$. En effet, nous avons choisi les poids tels que $C(1) = 0.9882$. Ici $\epsilon \sim \mathcal{N}(0, 1)$ et $n = 10^5$.

- Estimation par la méthode des moindres carrés.
- Même configurations des RN que ceux du modèle générateur des données.
- Tournage de 10^3 itérations de l'algorithme SGD avec "learning rate" = 0.001 ("decay rate" = 0.5).
- Soit $\hat{\theta}_n^* = (\hat{\theta}_{n,1}^*, \hat{\theta}_{n,2}^*, \hat{\theta}_{n,3}^*)$ le paramètre obtenu de la dernière itération et soit $\hat{\epsilon}_t = X_t - \hat{X}_t$, où

$$\hat{X}_t = \sum_{k=1}^K \xi_t^{(k)} f_k(X_{t-1}, \dots, X_{t-p}, \hat{\theta}_{n,k}^*).$$

Histogram of residuals



Expérience 2 : apprentissage de RN pour des données générées avec un CHARME

Modèle : CHARME(5) défini comme suit

$$X_t = \epsilon_t + \begin{cases} X_{t-1} + 3, & \text{if } R_t = 1 \\ \sqrt{0.2X_{t-1}^2 + 0.1X_{t-2}^2 + 0.25X_{t-3}^2 + 0.2X_{t-4}^2 + 0.05X_{t-5}^2} - 3, & \text{if } R_t = 2 \\ 0.05X_{t-1} + 0.2X_{t-2} + 0.15X_{t-3} + 0.03X_{t-4} + 0.01X_{t-5} + 0.1, & \text{if } R_t = 3 \end{cases}$$

avec $(\pi_1, \pi_2, \pi_3) = (0.15, 0.35, 0.5)$. Remarquons que la série $X_t^{(1)} = X_{t-1} + 3 + \epsilon_t$ n'est pas stationnaire malgré le fait que le processus entier le soit (car $C(1) < 1$). Comme avant, nous avons pris $\epsilon_t \sim \mathcal{N}(0, 1)$ et $n = 10^5$.

Expérience 2 : apprentissage de RN pour des données générées avec un CHARME

Modèle : CHARME(5) défini comme suit

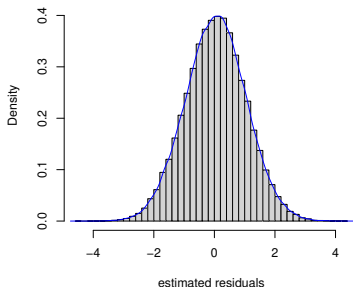
$$X_t = \epsilon_t + \begin{cases} X_{t-1} + 3, & \text{if } R_t = 1 \\ \sqrt{0.2X_{t-1}^2 + 0.1X_{t-2}^2 + 0.25X_{t-3}^2 + 0.2X_{t-4}^2 + 0.05X_{t-5}^2} - 3, & \text{if } R_t = 2 \\ 0.05X_{t-1} + 0.2X_{t-2} + 0.15X_{t-3} + 0.03X_{t-4} + 0.01X_{t-5} + 0.1, & \text{if } R_t = 3 \end{cases}$$

avec $(\pi_1, \pi_2, \pi_3) = (0.15, 0.35, 0.5)$. Remarquons que la série $X_t^{(1)} = X_{t-1} + 3 + \epsilon_t$ n'est pas stationnaire malgré le fait que le processus entier le soit (car $C(1) < 1$). Comme avant, nous avons pris $\epsilon_t \sim \mathcal{N}(0, 1)$ et $n = 10^5$.

- (i) Estimation par la méthode des MC avec trois RN $f_k(\cdot, \theta_k)$, $k = 1, 2, 3$, tels que
 - $\#neu(\theta_1) = (5, 300, 400, 200, 1)$,
 - $\#neu(\theta_2) = (5, 500, 600, 400, 1)$,
 - $\#neu(\theta_3) = (5, 300, 400, 200, 1)$,
 tous avec la fonction d'activation ReLU.
- (ii) Tournage de $2 \cdot 10^3$ itérations de l'algorithme SGD avec "learning rate" = 0.01 ("decay rate" = 10^{-6}).
- (iii) Soit $\hat{\theta}_n^* = (\hat{\theta}_{n,1}^*, \hat{\theta}_{n,2}^*, \hat{\theta}_{n,3}^*)$ le paramètre obtenu de la dernière itération et soit $\hat{\epsilon}_t = X_t - \hat{X}_t$, où

$$\hat{X}_t = \sum_{k=1}^K \xi_t^{(k)} f_k(X_{t-1}, \dots, X_{t-p}, \hat{\theta}_{n,k}^*).$$

Histogram of residuals



Introduction

τ -dépendance

Stabilité

Application et Simulations

Perspectives

Perspectives théoriques

- ▶ Solutions stationnaires dans L^α ($0 < \alpha < 1$)
- ▶ Stationnarité et Ergodicité du modèle CHARME sous des conditions plus faibles pour le processus de changement de régime $(R_t)_{t \in \mathbb{Z}}$
- ▶ Critère de selection du nombre de régimes K .

Références (principales)

1. Andrews, D. (1984) *Non strong mixing autoregressive processes*. J. Appl. Prob. 21 (1984): 930 – 934.
2. Dahlhaus, R. (2000) *A likelihood approximation for locally stationary processes*. Annals of Statistics, 28:1762 - 1794.
3. Dedecker, J. & Prieur, C. *Coupling for τ -dependent sequences and applications*. Journal of Theor. Prob. 17 – 4 (2004) : 861 – 885.
4. Doukhan, P. et Wintenberger, O. (2008) *Weakly dependent chains with infinite memory*. Stoc. Proc. and their App. 118:1997-2013.
5. **Gómez García, JG., Fadili, J. et Chesneau, C.** (2023) *Learning CHARME models with neural networks*. Statistical Papers.
6. Rockafellar, R.T. et Wets, R.J.B. (1998) *Variational Analysis* Springer.
7. Stockis, J-P., Franke, J. and Tadjuidje Kamgaing, J. (2010) *On geometric ergodicity of charme models*. J. of Time Series Analysis, 31: 141-152.
8. Tadjuidje-Kamgaing, J. (2005) *Competing neural networks as model for nonstationary financial time series*. PhD thesis, University of Kaiserslautern.
9. Yarotsky, D. (2017) *Error bounds for approximations with deep relu networks*. Neural Networks, 94 :103 - 114.

Merci de votre attention