

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA
RECHERCHE SCIENTIFIQUE
UNIVERSITÉ ABOU BEKR BELKAID TLEMCEM



Faculté des sciences
Département de Mathématiques
MÉMOIRE DE MASTER

présenté par

Farsi Wissal

Soutenu le : 1 juillet 2026

**Thème : Inférence dans les modèles
autorégressifs à coefficients aléatoires**

Soutenu devant le jury composé de :

Mme. BENSMAN NAWEL	M C A, Université de Tlemcen	Président
Mme. BENYELLES WAFIA	M C B, Université de Tlemcen	Examineur
Mme. BOUKHAR SOUAD	M C B, Université de Tlemcen	Examineur
Mr. ALLAM ABDELAZIZ	M C A, Université de Tlemcen	Encadrent

Année universitaire : 2025 - 2026

Dédicace

Grâce à Allah

Je dédie ce travail :

À mon premier exemple : **mes très chers parents** pour leur soutien inconditionnel, leur amour, leurs sacrifices tout au long de mon cursus scolaire et universitaire et surtout la confiance qu'ils m'ont toujours témoignée. Qu'Allah les protège.

À mes chers frères SIDAHMED et OUSSAMA pour leur aide, leurs encouragements et leur présence dans ma vie.

À mes chères sœurs DOUAA, IMENE et SANAA, mes complices et mes soutiens, merci pour votre affection, votre présence et vos encouragements qui m'ont toujours donné la force d'avancer.

À une personne précieuse, pour son soutien indéfectible et sa confiance en moi.

À mes chers enseignants...

À tous mes chers amis...

À toutes les personnes qui ont contribué à la réalisation de ce travail, de près ou de loin...

Wissal

Remerciements

Tout d'abord, je remercie Allah, le tout puissant de m'avoir accordé la force, le courage et la foi pour accomplir ce modeste travail. J'exprime en premier lieu ma sincère gratitude à mon encadreur Mr. Allam Abdelaziz d'avoir dirigé ce mémoire. Je le remercie pour ses discussions utiles et fructueuses, ainsi que ses conseils qui m'ont été très précieux pour mener à bien ce travail.

Mesdames N. Bensmaine, W. Benyelles et S. Boukhiar, c'est avec une grande considération que je vous exprime mes vifs remerciements pour l'honneur que vous m'avez fait en acceptant de faire partie du jury qui examinera ce modeste travail.

Table des matières

Introduction	8
1 Préliminaires	11
1.1 Séries chronologiques	11
1.1.1 Processus du second ordre et moments associés	11
1.2 Notion de stationnarité	12
1.2.1 Stationnarité du second ordre (ou stationnarité faible)	12
1.2.2 Stationnarité stricte	13
1.2.3 Fonction d'autocovariance (ACV)	14
1.2.4 Fonction d'autocorrélation (ACF)	15
1.2.5 Estimation de l'autocovariance et de l'autocorrélation	15
1.3 Le modèle autorégressif à coefficients aléatoires	17
1.3.1 Définitions et propriétés	17
1.3.2 Solution stationnaire	18
1.4 Estimation dans le modèle RCA(1)	21
1.4.1 Estimation par la méthode des moindres carrés	21
1.4.2 Par la méthode du maximum de vraisemblance	23
1.4.3 Estimateur des moindres carrés pondérés (WLS)	26
2 Vitesse de Convergence vers la loi normale pour une classe d'estimateurs dans un RCA(1)	29
2.1 Position du problème	29
2.2 Lemmes auxiliaires	30
2.3 Normalité asymptotique de l'estimateur $\hat{\beta}_n$	46
3 Simulation	50
3.1 Introduction	51

3.2	Simulation et prévision d'un processus autorégressif à coefficients aléatoires $RCA(1)$ à deux régimes	54
3.2.1	Exemple 1	54
3.2.2	Exemple 2	57
3.2.3	Exemple 3	60
3.3	Prévision des séries chronologiques de lynx à l'aide d'un modèle $RCA(1)$	63
Conclusion		67

Résumé

Ce mémoire est consacré à l'étude de la vitesse de convergence vers la loi normale d'une classe d'estimateurs pour le modèle autorégressif à coefficients aléatoires d'ordre un, noté $RCA(1)$, défini par :

$$X_t = (\beta + B_t)X_{t-1} + \varepsilon_t, \quad (1)$$

où B_t est un coefficient aléatoire. Ce modèle non linéaire s'avère particulièrement pertinent pour la modélisation de séries chronologiques économiques et financières, dont les structures de dépendance sont souvent trop complexes pour être capturées par des modèles linéaires classiques.

L'objectif principal de ce travail est de développer les résultats établis par Vaněček (2006) concernant la vitesse de convergence en loi de l'estimateur de Schick (estimateur des moindres carrés pondérés) vers la distribution normale. L'auteur a démontré que cette vitesse est de l'ordre de $\frac{(\log n)^3}{\sqrt{n}}$.

La performance des estimateurs est évaluée à travers trois études de simulation – mettant en jeu des coefficients aléatoires suivant une loi de Bernoulli généralisée (modèles à deux régimes) ainsi qu'une loi normale – et une application à des données réelles, à savoir la célèbre série historique des captures de lynx de la rivière Mackenzie (1821-1934). Les résultats obtenus montrent que l'estimateur des moindres carrés pondérés (WLS) conduit à des erreurs de prévision légèrement inférieures à celles de l'estimateur des moindres carrés ordinaires (MCO), en particulier lorsque l'hétéroscédasticité est plus prononcée, bien que l'amélioration observée reste modeste.

Enfin, plusieurs perspectives de recherche peuvent être envisagées : l'extension aux modèles RCA d'ordre supérieur ($RCA(p)$), aux processus multivariés, ou encore l'étude de critères de sélection de modèle adaptés à ce type de processus.

Mots-clés : RCA(1), estimation, moindres carrés pondérés (WLS), vitesse de convergence, loi normale, prévision, série des lynx.

Introduction :

Dans ce mémoire, nous étudions la vitesse de convergence, vers la loi normale, d'une certaine classe d'estimateurs pour un modèle autorégressif à coefficients aléatoires (RCA) défini par

$$X_t = (\beta + B_t)X_{t-1} + e_t, \quad (0.1)$$

où $\{B_t, t \in \mathbb{Z}\}$ est une suite de variables aléatoires indépendantes et identiquement distribuées, $\{X_t, t \in \mathbb{Z}\}$ est une suite ergodique et strictement stationnaire de variables aléatoires, et $\{B_t\}$ et $\{e_t\}$ sont mutuellement indépendantes.

L'observation de certaines séries chronologiques issues des domaines économique et financier — telles que les taux de change, les rendements boursiers ou encore les indices financiers — révèle des caractéristiques spécifiques que la modélisation linéaire ne permet pas de capturer pleinement. C'est pourquoi les modèles non linéaires, et en particulier les processus autorégressifs à coefficients aléatoires d'ordre un, notés RCA(1), apparaissent comme une alternative pertinente pour modéliser ce type de données, notamment lorsque le coefficient B_t n'est plus une constante mais une variable aléatoire.

Au cours des trente dernières années, l'intérêt pour le modèle à coefficients aléatoires n'a cessé de croître. Parmi les premiers exemples figure le modèle introduit par Nicholls et Quinn [12]. Ce modèle a suscité un intérêt considérable dans la littérature, principalement en raison de sa flexibilité et de sa maniabilité analytique. Nous renvoyons le lecteur à la monographie de Nicholls et Quinn [13] pour un excellent survol des premiers résultats, ainsi qu'aux références citées dans la contribution de Aue et Horváth [3].

L'estimation du modèle (0.1), et en particulier celle de β , a été largement étudiée. Aue et al. [2] et Berkes et al. [4] utilisent la méthode du quasi-maximum de vraisemblance (QML) pour estimer le coefficient de régression, montrant que la distribution asymptotique de l'estimateur de ϕ est normale

quelle que soit la valeur réelle de ϕ , tant que $\mathbb{E}(B_t^2) > 0$. Enfin, Schick [14] ainsi que Koul et Schick [10] discutent l’approche des moindres carrés pondérés (WLS), montrant qu’elle offre des performances comparables, voire supérieures, à celles de l’approche du maximum de vraisemblance.

Le présent mémoire s’appuie essentiellement sur l’article suivant :

Vaněček, P. (2006). Rate of convergence for a class of RCA estimators. *Kybernetika*, Volume 42, Number 6, pp. 699-709.

Cet article est consacré à l’étude des modèles autorégressifs à coefficients aléatoires dont le processus d’innovation est une différence de martingale. Une classe d’estimateurs du paramètre inconnu, introduite par Schick [14], y est utilisée ; cette classe englobe, entre autres, l’estimateur des moindres carrés et l’estimateur du maximum de vraisemblance. L’auteur examine le comportement asymptotique de ces estimateurs et établit, plus particulièrement, leur vitesse de convergence vers la loi normale.

Le présent mémoire est organisé en quatre chapitres.

Le chapitre préliminaire rappelle quelques notions fondamentales, telles que les fonctions d’autocovariance et d’autocorrélation, ainsi que les concepts de stationnarité faible et stricte. Nous y présentons également la classe des processus autorégressifs à coefficients aléatoires, en mettant l’accent sur les résultats obtenus par différents auteurs concernant la stationnarité de cette classe. Enfin, nous y décrivons quelques méthodes d’estimation des paramètres inconnus.

Le deuxième chapitre est consacré au développement des résultats obtenus par Vaněček [15]. Nous y détaillons les démonstrations des lemmes auxiliaires ainsi que celle du résultat principal relatif à la vitesse de convergence de l’estimateur proposé vers la loi normale.

Le troisième chapitre présente une étude numérique. Nous considérons trois exemples de modèles autorégressifs à coefficients aléatoires. Dans les deux premiers, le coefficient suit une loi de Bernoulli généralisée (modèle à deux régimes) ; dans le troisième, il suit une loi normale. Nous procédons à une simulation du modèle et calculons l’estimateur sur une période d’apprentissage basée sur 80 % des observations. Ensuite, nous calculons les prévisions associées sur la période de test et évaluons les méthodes d’estimation (MCO et WLS) sur la base des erreurs de prévision. Nous traitons également un exemple de données réelles, à savoir les données annuelles du nombre de peaux de lynx récoltées dans la région de la rivière Mackenzie (Canada) par la Compagnie de la Baie d’Hudson entre 1821 et 1934. Nous montrons que l’estimateur WLS, associé à la fonction de poids W optimale, est légèrement

préférable à l'estimateur des moindres carrés ordinaires.

Chapitre 1

Préliminaires

1.1 Séries chronologiques

1.1.1 Processus du second ordre et moments associés

Définition 1.1.1 (Processus du second ordre). *On qualifie de **processus du second ordre** tout processus aléatoire $(X_t)_{t \in T}$, défini sur $\mathbb{R}$, vérifiant pour tout instant $t \in T$:*

$$X_t \in L^2(\Omega, \mathcal{A}, \mathbb{P}).$$

Ceci équivaut à dire que chacune des variables aléatoires X_t admet un moment d'ordre 2 fini.

Pour ce type de processus, on introduit sa moyenne $m(\cdot)$ et sa covariance $\Gamma(\cdot, \cdot)$ de la manière suivante :

$$m(t) = \mathbb{E}[X_t] \quad \text{et} \quad \Gamma(t, s) = \text{Cov}(X_t, X_s),$$

pour tout couple $(t, s) \in T^2$.

La moyenne et la covariance sont ainsi des quantités qui varient dans le temps, reflet de l'évolution dynamique du processus. Concrètement, $m(\cdot)$ est une application de T dans $\mathbb{R}$, tandis que $\Gamma(\cdot, \cdot)$ est une application de T^2 dans $\mathbb{R}$.

Proposition 1.1.1 (Symétrie de la fonction de covariance). *Si $(X_t)_{t \in T}$ est un processus du second ordre, alors sa covariance est symétrique, c'est-à-dire :*

$$\Gamma(t, s) = \Gamma(s, t).$$

Définition 1.1.2 (Série chronologique). *On appelle série chronologique tout processus du second ordre $(X_t)_{t \in T}$ dont l'ensemble des indices temporels T est $\mathbb{N}$ ou $\mathbb{Z}$.*

Comme une série chronologique est un processus du second ordre, on a nécessairement, pour tout $t \in T$,

$$\mathbb{E}[X_t^2] < +\infty.$$

1.2 Notion de stationnarité

La stationnarité traduit l'idée que les caractéristiques du processus restent invariantes au cours du temps. Bien que la loi conjointe du processus soit généralement méconnue, l'analyse est facilitée par l'étude de ses premiers moments (moyenne et covariance).

1.2.1 Stationnarité du second ordre (ou stationnarité faible)

Définition 1.2.1 (Stationnarité faible). *Une série chronologique $(X_t)_{t \in T}$ est qualifiée de stationnaire au second ordre (ou **faiblement stationnaire**) si sa moyenne $m(\cdot)$ et sa covariance $\Gamma(\cdot, \cdot)$ sont inchangées par toute translation dans le temps.*

Formellement, pour tout $t, s \in T$ et tout décalage $k \in \mathbb{Z}$, on doit avoir :

1. $\mathbb{E}[X_t] = m(t) = m$ (la moyenne est indépendante du temps),
2. $\text{Cov}(X_t, X_s) = \Gamma(t, s) = \Gamma(t + k, s + k)$ (la covariance est invariante par translation).

Cette définition est formulée pour $T = \mathbb{Z}$, mais elle se transpose sans difficulté au cas $T = \mathbb{N}$.

On en déduit immédiatement que la variance du processus est elle aussi constante. En prenant $k = -t$ dans la condition de covariance, il vient :

$$\text{Var}(X_t) = \Gamma(t, t) = \Gamma(0, 0).$$

Ainsi, pour une série stationnaire au second ordre, l'espérance et la variance sont fixes, tandis que la covariance entre deux observations ne dépend plus que de l'intervalle qui les sépare.

Dès lors, la covariance, qui était initialement fonction des deux dates t et s , devient uniquement fonction de l'écart $h = t - s$. En choisissant $k = -s$, on obtient en effet :

$$\Gamma(t, s) = \Gamma(t - s, 0) = \Gamma(h, 0) = \gamma(h),$$

où $h = t - s$.

1.2.2 Stationnarité stricte

Définition 1.2.2 (Stationnarité stricte). *Une série chronologique $(X_t)_{t \in T}$ est dite **strictement stationnaire** (ou **fortement stationnaire**) lorsque, pour tout $n \geq 1$, tout n -uplet de dates $(t_1, \dots, t_n) \in T^n$, et tout décalage $k \in \mathbb{Z}$, les vecteurs aléatoires*

$$(X_{t_1}, \dots, X_{t_n}) \quad \text{et} \quad (X_{t_1+k}, \dots, X_{t_n+k})$$

possèdent la même loi.

Il ressort de cette définition que la date à laquelle on observe le processus n'affecte en rien sa distribution. Toute translation temporelle laisse la structure jointe du processus inchangée.

Dans les faits, la vérification de la stationnarité stricte s'avère complexe, puisqu'elle exigerait la connaissance de l'ensemble des lois fini-dimensionnelles. C'est la raison pour laquelle on lui préfère souvent une condition moins exigeante.

Proposition 1.2.1. *Toute série chronologique strictement stationnaire est également stationnaire au second ordre. En revanche, l'implication inverse est fautive dans le cas général.*

Preuve. Si $(X_t)_{t \in T}$ est strictement stationnaire, la définition appliquée avec $n = 1$ implique que toutes les variables X_t partagent la même loi, quelle que soit la date t . Leur espérance et leur variance sont donc bien constantes.

Par ailleurs, en appliquant la définition avec $n = 2$, tous les couples (X_t, X_s) ont la même loi que leurs translatés (X_{t+k}, X_{s+k}) . On en conclut que $\text{Cov}(X_t, X_s) = \text{Cov}(X_{t-s}, X_0)$, ce qui prouve que la covariance ne dépend que de l'écart $t - s$. La série est donc stationnaire au second ordre.

Pour établir la non-réciprocité, considérons le contre-exemple suivant. Soit le processus $(X_t)_{t \in \mathbb{N}}$ défini par

$$X_t = (-1)^t Z,$$

où Z désigne une variable aléatoire asymétrique (autrement dit, les lois de Z et de $-Z$ sont distinctes), avec $\mathbb{E}[Z] = 0$ et $\mathbb{E}[Z^2] = 1$.

Pour tout $t \in \mathbb{N}$, nous obtenons :

$$\mathbb{E}[X_t] = (-1)^t \mathbb{E}[Z] = 0 \quad \text{et} \quad \mathbb{E}[X_t^2] = (-1)^{2t} \mathbb{E}[Z^2] = 1.$$

De plus, pour tous $t, s \in \mathbb{N}$,

$$\text{Cov}(X_t, X_s) = (-1)^t (-1)^s \mathbb{E}[Z^2] = (-1)^{t+s} = (-1)^{t-s}.$$

Par conséquent, l'espérance et la variance demeurent constantes et la covariance est uniquement fonction de l'écart $t - s$. Le processus vérifie donc les conditions de la stationnarité faible.

Néanmoins, puisque X_t alterne entre Z et $-Z$ en fonction de la parité de t , les lois de X_t et de X_{t-1} sont différentes (rappelons que Z est asymétrique). Le processus n'est donc pas strictement stationnaire. $\square$

Remarque 1.2.1. *En ce qui concerne les processus gaussiens, la réciproque est toutefois valable. Un vecteur gaussien étant complètement caractérisé par son espérance et sa matrice de covariance, les deux formes de stationnarité deviennent équivalentes dans ce cadre.*

1.2.3 Fonction d'autocovariance (ACV)

Définition 1.2.3 (Fonction d'autocovariance). *Soit $(X_t)_{t \in T}$ une série chronologique stationnaire. Sa **fonction d'autocovariance**, notée $\gamma(\cdot)$, est définie pour tout décalage $h \in \mathbb{Z}$ par :*

$$\gamma(h) = \text{Cov}(X_h, X_0).$$

Grâce à l'hypothèse de stationnarité, pour tout instant $t \in T$, on a :

$$\text{Cov}(X_{t+h}, X_t) = \gamma(h).$$

Ce résultat confirme que l'autocovariance est déterminée uniquement par l'intervalle h séparant deux observations, indépendamment de la position t dans la série.

1.2.4 Fonction d'autocorrélation (ACF)

Définition 1.2.4 (Fonction d'autocorrélation). Soit $(X_t)_{t \in T}$ une série chronologique faiblement stationnaire. Sa fonction d'autocorrélation, notée $\rho(\cdot)$, est définie pour tout décalage $h \in \mathbb{Z}$ par le rapport :

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}.$$

De même, pour tout $t \in T$, on a :

$$\text{Cov}(X_{t+h}, X_t) = \rho(h).$$

Ainsi, comme pour l'autocovariance, l'autocorrélation ne dépend que du décalage temporel h .

Proposition 1.2.2 (Parité des fonctions d'autocovariance et d'autocorrélation). Pour une série chronologique stationnaire $(X_t)_{t \in T}$, les fonctions $\gamma(\cdot)$ et $\rho(\cdot)$ sont paires. Ainsi, pour tout $h \in \mathbb{Z}$:

$$\gamma(-h) = \gamma(h) \quad \text{et} \quad \rho(-h) = \rho(h).$$

Preuve. En utilisant la propriété de symétrie de la covariance, on montre que pour tout h :

$$\gamma(-h) = \text{Cov}(X_{-h}, X_0) = \text{Cov}(X_0, X_{-h}) = \gamma(h).$$

Le même raisonnement s'applique à la fonction $\rho(\cdot)$, puisque le dénominateur $\gamma(0)$ est une constante positive non nulle. $\square$

1.2.5 Estimation de l'autocovariance et de l'autocorrélation

Soit $\{X_t\}$ un processus stationnaire du second ordre. Sa moyenne est notée μ et sa fonction d'autocovariance est

$$\gamma(h) = \mathbb{E}[(X_{t+h} - \mu)(X_t - \mu)].$$

On suppose que la série des covariances est absolument sommable, c'est-à-dire $\sum_{h \in \mathbb{Z}} |\gamma(h)| < \infty$. Cette hypothèse garantit de bonnes propriétés asymptotiques.

À partir d'un échantillon de n observations, on estime $\gamma(h)$ par la covariance empirique :

$$\hat{\gamma}_n(h) = \begin{cases} \frac{1}{n} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \hat{\mu}_n)(X_t - \hat{\mu}_n), & \text{si } |h| \leq n-1, \\ 0, & \text{sinon,} \end{cases}$$

où $\hat{\mu}_n = \frac{1}{n} \sum_{t=1}^n X_t$ est la moyenne empirique. L'estimateur de l'autocorrélation est alors défini par

$$\hat{\rho}_n(h) = \frac{\hat{\gamma}_n(h)}{\hat{\gamma}_n(0)},$$

à condition que $\hat{\gamma}_n(0) > 0$.

Remarque 1.2.2. *Premièrement, l'estimateur $\hat{\gamma}_n(h)$ est nul pour $|h| \geq n$, car il n'existe pas de paires d'observations (X_t, X_{t+h}) séparées par un aussi grand décalage. Deuxièmement, lorsque $|h|$ est proche de n , le nombre de termes dans la somme est très petit. L'estimateur devient alors instable et peu fiable. En pratique, seuls les décalages $|h|$ petits par rapport à n (notés $|h| \ll n$) fournissent une estimation correcte.*

Pour un échantillon fini, $\hat{\gamma}_n(h)$ est un estimateur biaisé de $\gamma(h)$. Ce biais a deux origines :

- l'utilisation de la moyenne estimée $\hat{\mu}_n$ au lieu de la vraie moyenne μ ;
- le nombre réduit de paires d'observations pour les grands décalages.

Par exemple, pour $h = 0$, on peut montrer que l'espérance de l'estimateur vaut :

$$\mathbb{E}[\hat{\gamma}_n(0)] = \gamma(0) - \frac{1}{n} \sum_{k=-(n-1)}^{n-1} \left(1 - \frac{|k|}{n}\right) \gamma(k).$$

Le facteur $\left(1 - \frac{|k|}{n}\right)$ est appelé fenêtre de Bartlett. Il diminue le poids des covariances $\gamma(k)$ lorsque le retard k augmente.

Cependant, sous l'hypothèse de sommabilité absolue de γ , ce biais disparaît lorsque la taille de l'échantillon augmente. L'estimateur est asymptotiquement sans biais :

$$\lim_{n \rightarrow \infty} \mathbb{E}[\hat{\gamma}_n(h)] = \gamma(h).$$

Plus précisément, ce biais tend vers zéro à la vitesse $O(1/n)$.

1.3 Le modèle autorégressif à coefficients aléatoires

Ce modèle a été introduit et étudié par plusieurs auteurs comme Anděl (1976), Quinn et Nicholls (1983), Brandt (1986), Bougerol et Picard (1992). Le modèle RCA(1) a trouvé une variété d'applications dans différents domaines comme les finances, la biologie, etc. (Nicholls et Quinn, 1982; Tong, 1990).

1.3.1 Définitions et propriétés

Définition 1.3.1 (Modèle RCA(1)). *Un processus stochastique $(X_t, t \in \mathbb{Z})$ est un processus autorégressif à coefficients aléatoires, noté RCA(1), s'il satisfait l'équation*

$$X_t = \varphi_t X_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}, \quad (1.1)$$

où $\{(\varphi_t, \varepsilon_t), t \in \mathbb{Z}\}$ sont des vecteurs aléatoires.

Remarque 1.3.1. *Notons que Si $\sigma_\varphi^2 = \text{Var}(\varphi_t) = 0$, on retrouve le modèle autorégressif d'ordre 1 (AR(1)) de paramètre $\varphi = \mathbb{E}(\varphi_t)$:*

$$X_t = \varphi X_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}.$$

On pose les conditions suivantes :

- (C1) : $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc, c'est à dire une suite de variables aléatoires *i.i.d* centrées et de variance $\sigma_\varepsilon^2 < \infty$.
- (C2) : Les variables aléatoires (φ_t) sont *i.i.d* avec $\mathbb{E}(\varphi_t) = \varphi$.
- (C3) : Les deux suites (φ_t) et (ε_t) sont indépendantes.

Proposition 1.3.1. *Sous les conditions (C1), (C2) et (C3), on a*

$$\gamma(0) = \frac{\sigma^2}{1 - \varphi^2 - \sigma_\varphi^2}, \quad (1.2)$$

et pour $h > 0$

$$\gamma(h) = \varphi^h \gamma(0). \quad (1.3)$$

Preuve. Puisque les variables aléatoires ϕ_t et ε_t sont indépendantes de X_{t-1} alors

$$\begin{aligned}\gamma(0) &= \mathbb{E}(X_t^2) \\ &= \mathbb{E}[(\phi_t X_{t-1} + \varepsilon_t)^2] \\ &= \mathbb{E}(\phi_t^2) \mathbb{E}(X_{t-1}^2) + \mathbb{E}(\varepsilon_t^2) + 2\mathbb{E}(\phi_t) \mathbb{E}(X_{t-1}) \mathbb{E}(\varepsilon_t) \\ &= (\varphi^2 + \sigma_\varphi^2) \gamma(0) + \sigma_\varepsilon^2\end{aligned}$$

D'où la relation (1.2). Soit maintenant $h > 0$, on a

$$\begin{aligned}\mathbb{E}(X_t X_{t-h}) &= \mathbb{E}(\phi_t X_{t-1} X_{t-h}) + \mathbb{E}(\varepsilon_t X_{t-h}) \\ &= \mathbb{E}(\phi_t) \mathbb{E}(X_{t-1} X_{t-h}) \\ &= \varphi \gamma(h-1).\end{aligned}$$

Par récurrence, on obtient la relation (1.3). □

1.3.2 Solution stationnaire

Nombreux auteurs se sont intéressés à l'existence de solutions stationnaires pour les modèles autorégressifs à coefficients aléatoires. Dans le cas univarié, Andél [1] a établi des conditions nécessaires et suffisantes pour l'existence d'une solution stationnaire d'ordre 2 pour le modèle RCA(1). Il a notamment montré que la condition $\varphi^2 + \sigma_\varphi^2 < 1$ est nécessaire et suffisante pour garantir une solution faiblement stationnaire. Par la suite, Nicholls et Quinn [12] ont généralisé ces résultats au cadre multivarié, en dérivant des conditions de stationnarité pour une classe plus large de modèles autorégressifs à coefficients aléatoires. Le théorème suivant, dû à Hwang et Basawa [8], précise ces résultats en donnant, sous la même condition, la représentation explicite de l'unique solution stationnaire ergodique dans L^2 .

Théorème 1.3.1. [12] *Supposons que les conditions (C1), (C2) et (C3) sont satisfaites. Si*

$$\mathbb{E}(\varphi_t^2) < 1, \tag{1.4}$$

c'est-à-dire $\varphi^2 + \sigma_\varphi^2 < 1$, alors le processus 1.1 admet une unique solution stationnaire ergodique dans L^2 , donnée par

$$X_t = \epsilon_t + \sum_{j=1}^{\infty} \left(\prod_{i=0}^{j-1} \varphi_{t-i} \right) \epsilon_{t-j}.$$

Les conditions de stationnarité stricte pour les modèles autorégressifs à coefficients aléatoires reposent sur l'étude de l'exposant de Lyapunov associé au processus. Brandt [6] a considéré l'hypothèse suivante

$$X_n = \varphi_{n-1}X_{n-1} + \epsilon_n, \quad n \in \mathbb{Z}, \quad (1.5)$$

(H1) : La suite $\Psi = \{(\varphi_n, \epsilon_n), n \in \mathbb{Z}\}$ est stationnaire et ergodique.

Contrairement aux approches classiques, aucune condition d'indépendance n'a été requise. Il a établi le résultat fondamental suivant :

Théorème 1.3.2. [6] *Sous la condition (H1) et si*

$$-\infty \leq \mathbb{E}(\log |\varphi_0|) < 0 \quad \text{et} \quad \mathbb{E}(\log^+ |\epsilon_0|) < \infty,$$

ou si $\mathbb{P}(\varphi_0 = 0) > 0$, alors

$$X_n(\Psi) = \sum_{j=0}^{\infty} \left(\prod_{i=n-j}^{n-1} \varphi_i \right) \epsilon_{n-j-1}, \quad n \in \mathbb{Z},$$

est l'unique solution strictement stationnaire de l'équation (1.1). De plus, la série converge absolument presque sûrement.

Par la suite, Bougerol [5] a généralisé ces résultats au cadre multivarié en étudiant le modèle

$$X_n = A_n X_{n-1} + B_n, \quad n \in \mathbb{Z}, \quad (1.6)$$

où A_n, B_n sont des vecteurs aléatoires dans $\mathbb{R}^d$, les A_n sont des matrices $d \times d$ et .

Le théorème suivant, dû à Brandt (1986), est énoncé pour $d = 1$ mais sa preuve est valable pour tout $d \geq 1$.

La norme $\|\cdot\|$ désigne une norme d'opérateur sur l'ensemble des matrices $d \times d$. Considérons a condition :

(H'1) : La suite $\Psi = \{(A_n, B_n), n \in \mathbb{Z}\}$ est stationnaire et ergodique.

Théorème 1.3.3. [5] *Supposons que la l'hypothèse (H'1) est vérifiée. Sous les conditions*

$$\mathbb{E}(\log^+ \|A_0\|) < \infty \quad \text{et} \quad \mathbb{E}(\log^+ \|B_0\|) < \infty,$$

et si l'exposant de Lyapunov supérieur γ , défini par

$$\gamma = \inf_{n \in \mathbb{N}} \frac{1}{n+1} \mathbb{E} \log \|A_0 A_{-1} \cdots A_{-n}\|,$$

est strictement négatif, alors la série

$$X_n = \sum_{k=0}^{\infty} B_{n-k} \left(\prod_{i=0}^{k-1} A_{n-i} \right)$$

converge presque sûrement et définit l'unique solution strictement stationnaire de l'équation `refeq :modelebougerol`.

Bougerol a supposé que

(H2) : La suite $\{(A_n, B_n), n \in \mathbb{Z}\}$ est une suite donnée de variables aléatoires indépendantes et identiquement distribuées, définies sur $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeurs dans $\mathbb{M}(d) \times \mathbb{R}^d$ où $\mathbb{M}(d)$ désigne l'ensemble des matrices réelles $d \times d$.

Et a introduit la notion de solution non anticipative donnée dans la définition suivante :

Définition 1.3.2. Une solution strictement stationnaire non anticipative de l'équation 1.6 est un processus strictement stationnaire $\{X_n, n \in \mathbb{Z}\}$ qui est une solution de 1.6, telle que, pour tout $p \in \mathbb{Z}$, la variable aléatoire X_p est indépendante des variables aléatoires

$$\{(A_n, B_n), n > p\}.$$

Afin de fournir une condition nécessaire et suffisante pour l'existence d'une solution strictement stationnaire, ce qui constitue la réciproque du théorème de Brandt, l'auteur a introduit une condition supplémentaire sur le modèle, appelée "irréductibilité". Cette condition permet d'éviter les cas pathologiques où le processus pourrait être stationnaire sans que l'exposant de Lyapunov soit strictement négatif. La définition suivante formalise cette notion.

Définition 1.3.3. Un sous-espace affine H de $\mathbb{R}^d$ est dit invariant sous le modèle (1.6) si

$$\{A_0 x + B_0 : x \in H\} \subset H \quad \text{presque sûrement.}$$

Le modèle (1.6) est dit **irréductible** si $\mathbb{R}^d$ est le seul sous-espace affine invariant.

Le résultat fondamental de Bougerol est le suivant.

Théorème 1.3.4. [5] *Supposons que la suite $\{(A_n, B_n), n \in \mathbb{Z}\}$ satisfait la condition **(H2)** et que le modèle 1.6 est irréductible. Si*

$$\mathbb{E}[\log^+ \|e_0\|] < \infty \quad \text{et} \quad \mathbb{E}[\log^+ \|\beta + B_0 X\|] < \infty,$$

alors l'équation (1.6) admet une solution stationnaire non anticipative si et seulement si l'exposant de Lyapunov supérieur associé au modèle est strictement négatif :

$$\gamma = \lim_{n \rightarrow \infty} \frac{1}{n+1} \mathbb{E} \log \|(\beta + B_0)(\beta + B_{-1}) \cdots (\beta + B_{-n})\| < 0.$$

Ce théorème est d'une importance capitale car il établit une équivalence entre l'existence d'une solution strictement stationnaire non anticipative et la négativité de l'exposant de Lyapunov, sous l'hypothèse d'irréductibilité. En d'autres termes, la condition $\mathbb{E}[\log \|\beta + B_0\|] < 0$, qui est la version univariée de la condition sur l'exposant de Lyapunov, devient non seulement suffisante (comme l'avait montré Brandt) mais aussi nécessaire.

1.4 Estimation dans le modèle RCA(1)

Dans cette partie, nous présentons quelques études qui traduisent les problèmes d'estimation du paramètre d'un modèle autorégressif à coefficients aléatoires d'ordre un (RCA(1)), faites par de nombreux auteurs cités auparavant.

1.4.1 Estimation par la méthode des moindres carrés

Nicholls et Quinn (1981) considèrent l'échantillon $\{X_0, X_1, \dots, X_k\}$ du modèle défini par l'équation (1.1)(page 14), vérifiant les conditions **(C1)**, **(C2)** et **(C3)** avec

$$\mathbb{E}(\varphi_t) = \varphi, \quad \mathbb{E}(\varepsilon_t) = 0, \quad \text{Var}(\varphi_t) = \sigma_\varphi^2, \quad \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2.$$

L'estimateur des moindres carrés, noté $\hat{\phi}_{MC}$, du paramètre $\phi = \mathbb{E}(\phi_t)$ est obtenu en minimisant l'équation des erreurs suivante :

$$\sum_{t=1}^n \epsilon_t^2 = \sum_{t=1}^n (X_t - \mathbb{E}(X_t \mid \mathcal{F}_{t-1}))^2,$$

où $\mathcal{F}_{t-1} = \sigma(X_s, s \leq t)$ est la tribu engendré par le passé du processus jusqu'à l'ordre $t - 1$. On a

$$\sum_{t=1}^n (X_t - \mathbb{E}(X_t | \mathcal{F}_{t-1}))^2 = \sum_{t=1}^n (X_t - \varphi X_{t-1})^2.$$

En dérivant cette fonction par rapport à φ , on obtient

$$-2 \sum_{t=1}^n (X_t - \varphi X_{t-1}) X_{t-1} = 0,$$

d'où

$$\sum_{t=1}^n X_t X_{t-1} - \varphi \sum_{t=1}^n X_{t-1}^2 = 0,$$

et par conséquent

$$\hat{\varphi}_{MC} = \frac{\sum_{t=1}^n X_t X_{t-1}}{\sum_{t=1}^n X_{t-1}^2}. \quad (1.7)$$

Le théorème ergodique assure que $\hat{\varphi}_{MCO}$ est un estimateur consistant, c'est-à-dire quand $n \rightarrow \infty$, on a presque sûrement

$$\hat{\varphi}_{MCO} \rightarrow \frac{\gamma(1)}{\gamma(0)} = \varphi, \quad ,$$

où

$$\gamma(0) = \frac{\sigma_\epsilon^2}{1 - \phi^2 - \sigma_\varphi^2}, \quad \gamma(k) = \varphi^k \gamma(0), \quad k = 0, 1, 2, \dots$$

Pour déterminer la loi asymptotique de $\hat{\varphi}_{MC}$, les auteurs supposent l'existence du moment d'ordre 4. Ils ont obtenu le résultat suivant :

$$\mathbb{E}(Y_t^4) < \infty. \quad (2)$$

On a alors le théorème suivant :

Théorème 1.4.1. [11] Si $\mathbb{E}(Y_t^4) < \infty$, alors lorsque $n \rightarrow +\infty$ on a

$$\sqrt{n}(\hat{\varphi}_{MC} - \varphi) \xrightarrow{\mathcal{L}} \mathcal{N}(0, v^2),$$

où

$$v^2 = \frac{\sigma_\epsilon^2 \mathbb{E}(X_0^2) + 2 \text{cov}(\varphi, \epsilon) \mathbb{E}(X_0^3) + \sigma_\varphi^2 \mathbb{E}(X_0^4)}{\gamma(0)^2}.$$

Ainsi, la méthode des moindres carrés fournit des estimateurs explicites et facilement calculables, tout en possédant les propriétés de consistance et de normalité asymptotique, ce qui en fait un outil pratique pour l'inférence dans le modèle RCA(1).

1.4.2 Par la méthode du maximum de vraisemblance

Soit (X_t) un processus autorégressif à coefficients aléatoires vérifiant l'équation suivante :

$$X_t = \alpha + \varphi_t X_{t-1} + \varepsilon_t. \quad (1.8)$$

On suppose que les deux suites (φ_t) et (ε_t) vérifient les conditions **(C1)**, **(C2)** et **(C3)**.

L'estimation des paramètres du modèle RCA(1) par la méthode du maximum de vraisemblance est mise en œuvre sous un ensemble d'hypothèses régulières :

C4 : $\mathbb{E}(\varepsilon_t^4) < \infty$ et $\mathbb{E}(\varphi_t^4) < \infty$;

C5 : $\sigma_\varepsilon^2 \geq \delta_1 > 0$ et $\sigma_\varphi^2 \geq \delta_2 > 0$;

C6 : $(\varphi, \sigma_\varphi^2) \in D$, où D est un ensemble fermé contenu dans la région de la stationnarité stricte et de l'ergodicité donnée par l'inégalité (1.4).

Wang [16] a traité la méthode du maximum de vraisemblance pour estimer le paramètre :

$$\Theta = \{(\alpha, \varphi, \sigma_\varphi^2, \sigma^2) : \alpha \in \mathbb{R}, (\varphi, \sigma_\varphi^2) \in D, \sigma^2 \geq \delta_1, \sigma_\varphi^2 \geq \delta_2\}$$

L'hypothèse **C4** sera utilisée pour prouver le théorème central limite pour l'estimateur du maximum de vraisemblance, et les hypothèses **C5** et **C6** sont imposées afin que l'ensemble des paramètres du modèle RCA(1) soit compact.

La construction de la fonction de vraisemblance s'appuie sur l'hypothèse de normalité conditionnelle des erreurs. Dans ce cadre, la loi conditionnelle de l'observation courante X_t sachant la valeur passée X_{t-1} est une loi normale dont l'espérance est donnée par

$$\mathbb{E}(X_t \mid X_{t-1}) = \alpha + \varphi X_{t-1},$$

et dont la variance est

$$\text{Var}(X_t \mid X_{t-1}) = \sigma_\varepsilon^2 + \sigma_\varphi^2 X_{t-1}^2.$$

Cette formulation permet d'écrire explicitement la log-vraisemblance conditionnelle :

$$\begin{aligned}
f_n(X_1, \dots, X_n \mid X_0) &= \prod_{t=1}^n f(X_t \mid X_{t-1}) \\
&= (2\pi)^{-\frac{n}{2}} \prod_{t=1}^n \left\{ (\sigma_\varepsilon^2 + \sigma_\varphi^2 X_{t-1}^2)^{-\frac{1}{2}} \exp \left(-\frac{(x_t - \alpha - \varphi x_{t-1})^2}{2(\sigma_\varepsilon^2 + \sigma_\varphi^2 x_{t-1}^2)} \right) \right\} \\
&= L_n(\alpha, \varphi, \sigma_\varphi^2, \sigma_\varepsilon^2).
\end{aligned}$$

On définit

$$\tilde{\ell}_n(\alpha, \varphi, \sigma_\varphi^2, \sigma_\varepsilon^2) = -\frac{2}{n} \ln L_n(\alpha, \varphi, \sigma_\varphi^2, \sigma_\varepsilon^2) - \ln(2\pi).$$

Il est plus pratique de minimiser $\tilde{\ell}_n(\alpha, \varphi, \sigma_\varphi^2, \sigma_\varepsilon^2)$ que de maximiser $L_n(\alpha, \varphi, \sigma_\varphi^2, \sigma_\varepsilon^2)$. Afin de faciliter la procédure d'optimisation numérique, nous adoptons une reparamétrisation en introduisant le paramètre auxiliaire $\tau = \sigma_\varphi^2 / \sigma_\varepsilon^2$ et on donne la fonction $\tilde{\ell}$ qui dépend de τ :

$$\begin{aligned}
\bar{\ell}_n(\alpha, \varphi, \tau, \sigma_\varepsilon^2) &= \tilde{\ell}_n(\alpha, \varphi, \sigma_\varphi^2, \sigma_\varepsilon^2) \\
&= \ln \sigma_\varepsilon^2 + \frac{1}{n} \sum_{t=1}^n \ln(1 + \tau X_{t-1}^2) + \sigma_\varepsilon^{-2} \frac{1}{n} \sum_{t=1}^n \frac{(X_t - \alpha - \varphi X_{t-1})^2}{1 + \tau X_{t-1}^2}.
\end{aligned}$$

Pour une valeur fixée de τ , l'annulation des dérivées premières de la log-vraisemblance par rapport aux paramètres α et φ conduit à des estimateurs explicites :

$$\hat{\varphi}_n(\tau) = \frac{c_1 c_2 - c_3 c_4}{c_1 c_5 - c_4^2}, \quad \alpha_n(\tau) = \frac{c_3 - \hat{\varphi}_n(\tau) c_4}{c_1},$$

où les coefficients $c_1, \dots, c_5$ sont des sommes pondérées dépendant uniquement de τ et sont définis par

$$\begin{aligned}
c_1 &= \sum_{t=1}^n \frac{1}{1 + \tau X_{t-1}^2}, & c_2 &= \sum_{t=1}^n \frac{X_t X_{t-1}}{1 + \tau X_{t-1}^2}, \\
c_3 &= \sum_{t=1}^n \frac{X_t}{1 + \tau X_{t-1}^2}, & c_4 &= \sum_{t=1}^n \frac{X_{t-1}}{1 + \tau X_{t-1}^2}, \\
c_5 &= \sum_{t=1}^n \frac{X_{t-1}^2}{1 + \tau X_{t-1}^2}, & c_6 &= \sum_{t=1}^n \frac{X_t^2}{1 + \tau X_{t-1}^2}.
\end{aligned}$$

L'estimateur de la variance σ_ε^2 est alors donné par

$$\sigma_n^2(\tau) = \frac{1}{n} \sum_{t=1}^n \frac{(X_t - \alpha_n(\tau) - \hat{\varphi}_n(\tau)X_{t-1})^2}{1 + \tau X_{t-1}^2}.$$

En substituant ces estimateurs profilés dans l'expression de la log-vraisemblance, nous obtenons une fonction qui ne dépend plus que du seul paramètre τ :

$$l_n(\alpha, \varphi, \tau) = \ln \left(\frac{1}{n} \sum_{t=1}^n \frac{(X_t - \alpha - \varphi X_{t-1})^2}{1 + \tau X_{t-1}^2} \right) + \frac{1}{n} \sum_{t=1}^n \ln(1 + \tau X_{t-1}^2).$$

Dès lors, le problème d'optimisation multidimensionnelle initial se réduit à une simple minimisation unidimensionnelle de cette fonction concentrée par rapport à τ . L'estimateur du maximum de vraisemblance $\hat{\tau}_n$ est ainsi obtenu, et il suffit alors de remplacer τ par sa valeur optimale dans les expressions explicites précédentes pour en déduire les estimateurs finaux :

$$\hat{\alpha}_n = \alpha_n(\hat{\tau}_n), \quad \hat{\varphi}_n = \hat{\varphi}_n(\hat{\tau}_n), \quad \hat{\sigma}_n^2 = \sigma_n^2(\hat{\tau}_n), \quad \hat{\sigma}_{\varphi,n}^2 = \hat{\tau}_n \hat{\sigma}_n^2.$$

Cette approche par concentration constitue une simplification majeure qui rend la mise en œuvre numérique de l'estimation du modèle RCA(1) à la fois pratique et efficace. Le résultat suivant assure la consistance forte de l'estimateur du maximum de vraisemblance :

Théorème 1.4.2. [16] Soit $\{X_t\}$ une suite $\mathcal{F}_t$ -mesurable, strictement stationnaire et ergodique qui satisfait l'équation (1.8) avec $\theta = (\alpha, \varphi, \sigma_\varphi^2, \sigma_\varepsilon^2)$ sous les conditions **C1**, **C2**, **C3**, **C5** et **C6**. On définit $\pi = (\alpha, \varphi, \tau)$ où $\tau = \frac{\sigma_\varphi^2}{\sigma_\varepsilon^2}$. Soit Θ^* l'image de Θ par l'application continue $T(\theta) = \pi$, où Θ est l'espace des paramètres du modèle RCA(1).

Soit $\pi_0 = (\alpha_0, \varphi_0, \tau_0)$, où $\tau_0 = \frac{\sigma_{\varphi,0}^2}{\sigma_{\varepsilon,0}^2}$. Alors $\lim_{n \rightarrow \infty} \ell_n^*(\alpha, \varphi, \tau)$ existe presque sûrement pour tout $(\alpha, \varphi, \tau) \in \Theta^*$ et sa limite $\ell^*(\alpha, \varphi, \tau)$ est minimisée uniquement sur Θ^* en π_0 à condition que $\pi_0 \in \text{int}(\Theta^*)$.

Le théorème suivant, dû à Wang (2003), garantit la consistance forte des estimateurs du maximum de vraisemblance dans le modèle RCA(1).

Théorème 1.4.3. [16] Soit $\ell_n(\alpha, \varphi, \tau)$ minimisée sur Θ^* en $\alpha = \hat{\alpha}_n$, $\hat{\varphi}_n(\hat{\tau}_n)$ et $\tau = \hat{\tau}_n$, où Θ^* est défini dans le Théorème 4. Soit $\hat{\pi}_n = (\hat{\alpha}_n, \hat{\varphi}_n, \hat{\tau}_n)'$. Alors $\hat{\pi}_n$ converge presque sûrement vers $\pi_0 = (\alpha_0, \varphi_0, \tau_0)$ si $\pi_0 \in \text{int}(\Theta^*)$. De plus, $\hat{\sigma}_{\varphi,n}^2$ et $\hat{\sigma}_n^2$ convergent presque sûrement vers σ_{varphi}^2 et σ^2 respectivement.

$$\hat{\alpha}_n = \alpha_n(\hat{\tau}_n), \quad \hat{\varphi}_n = \hat{\varphi}_n(\hat{\tau}_n), \quad \hat{\sigma}_n^2 = \sigma_n^2(\hat{\tau}_n), \quad \hat{\sigma}_{\varphi,n}^2 = \hat{\tau}_n \hat{\sigma}_n^2.$$

1.4.3 Estimateur des moindres carrés pondérés (WLS)

Schick [14] a étudié l'estimation du paramètre φ dans le modèle autorégressif à coefficients aléatoires d'ordre 1, RCA(1), défini par l'équation

$$X_t = (\varphi + \varphi_t)X_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z},$$

où $\{\varepsilon_t : t \in \mathbb{Z}\}$ et $\{\varphi_t : t \in \mathbb{Z}\}$ sont deux suites indépendantes de variables aléatoires i.i.d. de moyennes nulles et de variances respectives $\mathbb{E}(\varepsilon_1^2) = \sigma_\varepsilon^2$ et $\mathbb{E}(\varphi_1^2) = \sigma_\varphi^2$ finies et satisfaisant la condition de stationnarité faible

$$\mathbb{E}[(\varphi + \varphi_1)^2] < 1.$$

Soit $\mathcal{D}$ l'ensemble de toutes les fonctions mesurables bornées ϕ qui satisfont $z\phi(z) \geq 0$ pour $z \neq 0$.

L'auteur introduit une classe d'estimateurs de φ , indexée par une famille de fonctions mesurables bornées ϕ , notée $\hat{\varphi}_n(\phi)$ définie par

$$\hat{\varphi}_n(\phi) = \frac{\sum_{j=1}^n \phi(X_{j-1})X_j}{\sum_{j=1}^n \phi(X_{j-1})X_{j-1}}.$$

Posons $w(x) = \sigma_\varepsilon^2 + \sigma_\varphi^2 x^2$.

L'auteur a établi le résultat suivant :

Théorème 1.4.4. [14] *Pour tout $\phi \in \mathcal{D}$,*

$$\sqrt{n}(\hat{\varphi}_n(\phi) - \varphi) \xrightarrow{d} \mathcal{N}(0, V(\phi)),$$

avec

$$V(\phi) = \frac{\mathbb{E}[\phi^2(X_0)w(X_0)]}{(\mathbb{E}[\phi(X_0)X_0])^2}.$$

D'autre part, dans le modèle RCA(1), la variance conditionnelle de ε_t sachant X_{t-1} est donnée par

$$\mathbb{E}(\varepsilon_t^2 \mid X_{t-1}) = \sigma_\varepsilon^2 + \sigma_\varphi^2 X_{t-1}^2.$$

L'auteur s'est ensuite intéressé au problème du choix optimal de la fonction ϕ . La question est de savoir quelle forme de ϕ produit la variance asymptotique la plus petite. Pour ce faire, il a considéré la fonction définie par

$$\psi(x) = \frac{x}{w(x)} = \frac{x}{\sigma_\varepsilon^2 + \sigma_\varphi^2 x^2}$$

et a établi la proposition suivante :

Proposition 1.4.1. *La fonction ψ est un élément de $\mathcal{D}$. De plus*

$$\forall \phi \in \mathcal{D}, \quad V(\phi) \geq V(\psi).$$

Preuve. On a $\psi(X_0)X_0 = \psi^2(X_0)w(X_0)$. Donc

$$V(\psi) = \frac{1}{\mathbb{E}[\psi^2(X_0)w(X_0)]}.$$

D'autre part, l'inégalité de Cauchy-Schwarz donne

$$(\mathbb{E}[\phi(X_0)X_0])^2 = (\mathbb{E}[\phi(X_0)\psi(X_0)w(X_0)])^2 \leq \mathbb{E}[\psi^2(X_0)w(X_0)] \mathbb{E}[\phi^2(X_0)w(X_0)].$$

Il en découle que

$$V(\phi) \geq \frac{1}{\mathbb{E}[\psi^2(X_0)w(X_0)]} = V(\psi),$$

avec égalité si et seulement si $\phi(X_0) = c\psi(X_0)$ presque sûrement pour une certaine constante $c > 0$. $\square$

Définissons, pour tout $r > 0$, les fonctions

$$\phi_r(x) = \frac{x}{r + x^2}, \quad x \in \mathbb{R},$$

et posons $\tau = \sigma_\varphi^2/\sigma_\varepsilon^2$. Alors $\phi_\tau = \sigma_\varepsilon^2\psi$ et $\hat{\varphi}_n(\phi_\tau)$ minimise la variance asymptotique parmi tous les estimateurs $\hat{\varphi}_n(\phi)$ avec $\phi \in \mathcal{D}$. Il en résulte que $\hat{\varphi}_n(\phi_\tau)$ est l'unique estimateur optimal dans cette classe. Il s'avère que l'estimateur du quasi-maximum de vraisemblance (voir section 1.4.2) est asymptotiquement équivalent à $\hat{\varphi}_n(\phi_\tau)$. Cependant, ce dernier nécessite l'existence des moments d'ordre 4 de ε_t et φ_t , et son calcul est difficile.

Le théorème suivant montre qu'il est possible de construire des estimateurs asymptotiquement équivalents à $\hat{\varphi}_n(\phi_\tau)$ si $\tau = \sigma_\varphi^2/\sigma_\varepsilon^2$ peut être estimé de manière convergente :

Théorème 1.4.5. [14] *Supposons que $\hat{\tau}_n$ est un estimateur convergent de $\tau = \sigma_\varphi^2/\sigma_\varepsilon^2$. Alors l'estimateur $\hat{\varphi}_n(\phi_{\hat{\tau}_n})$ vérifie*

$$\sqrt{n}(\hat{\varphi}_n(\phi_{\hat{\tau}_n}) - \varphi) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V(\phi_\tau)).$$

Par conséquent,

$$\hat{\varphi}_n(\phi_{\hat{\tau}_n}) \xrightarrow{p} \varphi.$$

Définition 1.4.1. *L'estimateur des moindres carrés pondérés (WLS) de φ est défini par*

$$\hat{\varphi}_{WLS} = \frac{\sum_{j=1}^n \frac{X_{j-1}X_j}{\sigma_\varepsilon^2 + \sigma_\varphi^2 X_{j-1}^2}}{\sum_{j=1}^n \frac{X_{j-1}^2}{\sigma_\varepsilon^2 + \sigma_\varphi^2 X_{j-1}^2}}.$$

La contribution majeure de Schick est de montrer qu'il est possible de construire un estimateur de type WLS *adaptatif* (ou réalisable) en remplaçant les variances inconnues σ_ε^2 et σ_φ^2 par des estimateurs convergents, sans pour autant imposer l'existence de moments d'ordre 4.

Chapitre 2

Vitesse de Convergence vers la loi normale pour une classe d'estimateurs dans un RCA(1)

Dans ce chapitre, nous développons les résultats présentés dans [15]. L'auteur a considéré un processus autorégressif à coefficients aléatoires du premier ordre RCA(1) solution de l'équation

$$X_t = (\beta + B_t)X_{t-1} + Y_t, \quad (2.1)$$

où β est un paramètre constant inconnu, $\{Y_t; t \in \mathbb{Z}\}$ est un processus d'erreur d'espérance nulle et de variance constante non nulle finie σ^2 , et $\{B_t; t \in \mathbb{Z}\}$ est une suite de variables aléatoires d'espérance nulle et de variance constante non nulle σ_B^2 telle que $\beta^2 + \sigma_B^2 < 1$.

2.1 Position du problème

Soit les conditions :

- (A1) : $\{B_t; t \in \mathbb{Z}\}$ est une suite de variables aléatoires indépendantes et identiquement distribuées, $\{Y_t; t \in \mathbb{Z}\}$ est une suite ergodique et strictement stationnaire de variables aléatoires, $\{B_t\}$ et $\{Y_t\}$ sont mutuellement indépendants.
- (A2) : $\{Y_t; t \in \mathbb{Z}\}$ est une suite de différences de martingale par rapport à la σ -algèbre $\mathcal{F}_t = \sigma(B_s, Y_s; s \leq t)$, c'est-à-dire $Y_t \in \mathcal{F}_t$ et $\mathbb{E}[Y_t | \mathcal{F}_{t-1}] = 0$.

Il est connu que sous l'hypothèse A1, il existe un processus strictement stationnaire et ergodique $\{X_t; t \in \mathbb{Z}\}$ qui satisfait l'équation (2.1) (voir section 1.4). Dans de nombreux travaux, le processus d'erreur est supposé bruit blanc fort c'est-à-dire une suite de variables aléatoires i.i.d. centrées et de variance finie. Dans ce travail, cette condition a été affaiblie en considérant un processus d'erreur sous forme de différence de martingale.

L'objectif principal de ce chapitre consiste à établir la vitesse de convergence des estimateurs $\hat{\beta}_n(\phi)$, introduits dans la section 1.4.3, dans le processus RCA(1) pour une classe suffisamment large de fonctions ϕ .

Nous étendons les résultats à une classe générale d'estimateurs pour le modèle RCA(1). En premier lieu, nous devons définir la classe de toutes les fonctions admissibles $\phi(x)$. Posons $h(x) = x\phi(x)$ et supposons que

(A3) : $h(x) > 0$ pour $x \neq 0$ et h vérifie la condition de Lipschitz

$$|h(X_s) - h(X_t)| \leq c_h |X_s - X_t| \quad \text{p.s.}$$

pour le processus $\{X_t, t \in \mathbb{Z}\}$, pour tous $s, t \in \mathbb{Z}$, et pour une certaine constante $c_h > 0$.

2.2 Lemmes auxiliaires

La démonstration du résultat principal nécessite quelques lemmes préliminaires. Le lemme suivant Précision de l'approximation normale

Lemme 2.2.1. *Soit X et Y des variables aléatoires réelles définies sur $(\Omega, \mathcal{A}, \mathbb{P})$.*

Alors, pour tout $\varepsilon > 0$,

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left(\frac{X}{Y} \leq x \right) - \Phi(x) \right| \leq \sup_{y \in \mathbb{R}} |\mathbb{P}(X \leq y) - \Phi(y)| + \mathbb{P}(|Y - 1| > \varepsilon) + \varepsilon,$$

où $\Phi(x)$ est la fonction de répartition de la loi normale standard.

Preuve. L'inégalité est triviale pour $\varepsilon \geq 1$. On suppose donc que $\varepsilon \in]0, 1[$.

Notons

$$D = \sup_{y \in \mathbb{R}} |\mathbb{P}(X \leq y) - \Phi(y)|.$$

Fixons $x \in \mathbb{R}$ et $\varepsilon > 0$. On décompose :

$$\mathbb{P} \left(\frac{X}{Y} \leq x \right) = \mathbb{P}(X \leq xY, |Y - 1| \leq \varepsilon) + \mathbb{P}(X \leq xY, |Y - 1| > \varepsilon).$$

Pour $|Y - 1| \leq \varepsilon$, on a

$$|xY - x| = |x| |Y - 1| \leq |x| \varepsilon,$$

donc

$$xY \leq x + |x| \varepsilon.$$

Ainsi,

$$\{X \leq xY, |Y - 1| \leq \varepsilon\} \subseteq \{X \leq x + |x| \varepsilon\}.$$

Par conséquent,

$$\mathbb{P}\left(\frac{X}{Y} \leq x\right) \leq \mathbb{P}(X \leq x + |x| \varepsilon, |Y - 1| \leq \varepsilon) + \mathbb{P}(|Y - 1| > \varepsilon). \quad (2.2)$$

D'autre part, lorsque $|Y - 1| \leq \varepsilon$, on a $1 - \varepsilon \leq Y \leq 1 + \varepsilon$. Nous distinguons deux cas :

— Si $x \geq 0$,

$$xY \geq x(1 - \varepsilon) = x - x\varepsilon = x - |x| \varepsilon.$$

— Si $x < 0$, posons $x = -a$ avec $a > 0$. Alors

$$xY = -aY \geq -a(1 + \varepsilon) = -a - a\varepsilon = x - |x| \varepsilon.$$

ceci assure que, pour tout nombre réel x , on ait $xY \geq x - |x| \varepsilon$. Ainsi

$$\{X \leq x - |x| \varepsilon, |Y - 1| \leq \varepsilon\} \subseteq \{X \leq xY, |Y - 1| \leq \varepsilon\}.$$

Or,

$$\begin{aligned} \mathbb{P}(X \leq x - |x| \varepsilon, |Y - 1| \leq \varepsilon) &= \mathbb{P}(X \leq x - |x| \varepsilon) - \mathbb{P}(X \leq x - |x| \varepsilon, |Y - 1| > \varepsilon) \\ &\geq \mathbb{P}(X \leq x - |x| \varepsilon) - \mathbb{P}(|Y - 1| > \varepsilon). \end{aligned}$$

On obtient donc

$$\begin{aligned} \mathbb{P}\left(\frac{X}{Y} \leq x\right) &= \mathbb{P}(X \leq xY, |Y - 1| \leq \varepsilon) + \underbrace{\mathbb{P}(X \leq xY, |Y - 1| > \varepsilon)}_{\geq 0} \\ &\geq \mathbb{P}(X \leq xY, |Y - 1| \leq \varepsilon) \\ &\geq \mathbb{P}(X \leq x - |x| \varepsilon) - \mathbb{P}(|Y - 1| > \varepsilon). \end{aligned} \quad (2.3)$$

En utilisant les inégalités 2.2 et 2.3 on tire que

$$\begin{aligned}\mathbb{P}\left(\frac{X}{Y} \leq x\right) - \Phi(x) &\leq [\mathbb{P}(X \leq x + |x|\varepsilon) - \Phi(x + |x|\varepsilon)] \\ &\quad + [\Phi(x + |x|\varepsilon) - \Phi(x)] + \mathbb{P}(|Y - 1| > \varepsilon), \\ \mathbb{P}\left(\frac{X}{Y} \leq x\right) - \Phi(x) &\geq [\mathbb{P}(X \leq x - |x|\varepsilon) - \Phi(x - |x|\varepsilon)] \\ &\quad + [\Phi(x - |x|\varepsilon) - \Phi(x)] - \mathbb{P}(|Y - 1| > \varepsilon).\end{aligned}$$

En prenant la valeur absolue, on obtient

$$\left|\mathbb{P}\left(\frac{X}{Y} \leq x\right) - \Phi(x)\right| \leq D + \max\left\{|\Phi(x + |x|\varepsilon) - \Phi(x)|, |\Phi(x - |x|\varepsilon) - \Phi(x)|\right\} + \mathbb{P}(|Y - 1| > \varepsilon).$$

Montrons à présent que pour tout $x \in \mathbb{R}$ et $\varepsilon > 0$,

$$|\Phi(x + |x|\varepsilon) - \Phi(x)| \leq \varepsilon.$$

Supposons dans un premier temps que $x \geq 0$.

Par le théorème des accroissements finis, il existe $\xi \in [x, x(1 + \varepsilon)]$ tel que

$$\Phi(x(1 + \varepsilon)) - \Phi(x) = (x(1 + \varepsilon) - x) f(\xi) = \varepsilon x f(\xi).$$

où $f = \phi'$ est la densité de la loi normale centrée réduite

Il nous faut majorer $\varepsilon x f(\xi)$. Comme $x \leq \xi$, alors

$$\varepsilon x f(\xi) \leq \varepsilon \xi f(\xi).$$

Étudions la fonction $k(t) = t f(t)$ pour $t \geq 0$. Sa dérivée est

$$k'(t) = f(t) + t f'(t) = f(t) - t^2 f(t) = (1 - t^2) f(t).$$

Le signe de $k'(t)$ est celui de $1 - t^2$: positif pour $0 < t < 1$, nul en $t = 1$, négatif pour $t > 1$. Donc $k(t)$ atteint son maximum en $t = 1$:

$$\max_{t \geq 0} k(t) = k(1) = 1 \cdot f(1) = \frac{1}{\sqrt{2\pi}} e^{-1/2} = \frac{1}{\sqrt{2\pi e}} \leq 1.$$

Ainsi, pour tout $\xi \geq 0$, $\xi f(\xi) \leq 1$. Par conséquent,

$$\varepsilon x f(\xi) \leq \varepsilon \cdot \max_{t \geq 0} t f(t) \leq \varepsilon.$$

Pour $x \geq 0$, on a donc

$$\Phi(x(1 + \varepsilon)) - \Phi(x) \leq \varepsilon.$$

Comme la différence est positive (car Φ est croissante), il en résulte

$$|\Phi(x(1 + \varepsilon)) - \Phi(x)| \leq \varepsilon.$$

Considérons maintenant le deuxième cas $x < 0$. Posons $x = -a$ avec $a > 0$. Alors $|x| = a$. On veut majorer

$$|\Phi(x - |x|\varepsilon) - \Phi(x)| = |\Phi(-a - a\varepsilon) - \Phi(-a)|.$$

Par symétrie de la loi normale, $\Phi(-u) = 1 - \Phi(u)$ pour tout réel u . Donc

$$\Phi(-a - a\varepsilon) = 1 - \Phi(a + a\varepsilon) = 1 - \Phi(a(1 + \varepsilon)), \quad \Phi(-a) = 1 - \Phi(a).$$

La différence devient

$$|(1 - \Phi(a(1 + \varepsilon))) - (1 - \Phi(a))| = |\Phi(a) - \Phi(a(1 + \varepsilon))| = \Phi(a(1 + \varepsilon)) - \Phi(a),$$

car $a(1 + \varepsilon) > a$ et Φ est croissante. On est donc ramené au cas $x \geq 0$ avec $x = a$. On a déjà montré que cette différence est majorée par ε . D'où

$$|\Phi(x - |x|\varepsilon) - \Phi(x)| \leq \varepsilon.$$

On conclue que, pour tout $x \in \mathbb{R}$ et tout $\varepsilon > 0$,

$$|\Phi(x + |x|\varepsilon) - \Phi(x)| \leq \varepsilon \quad \text{et} \quad |\Phi(x - |x|\varepsilon) - \Phi(x)| \leq \varepsilon.$$

□

Lemme 2.2.2 (Berry-Esseen). *Soit $\{M_t, t \in \mathbb{N}\}$ une suite de différences de martingale de variance constante non nulle σ^2 .*

Supposons que $\mathbb{P}(|M_t| \leq c, \forall t \in \mathbb{N}) = 1$ pour une constante $c > 0$. Alors il existe $d > 0$ tel que pour chaque $n \in \mathbb{N}$

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left(\frac{1}{\sqrt{\sigma^2}} \cdot \frac{1}{\sqrt{n}} \sum_{t=1}^n M_t \leq x \right) - \Phi(x) \right| \leq d \cdot \frac{(\ln n)^3}{\sqrt{n}}.$$

Il s'avère également qu'une inégalité exponentielle de type Hoeffding est requise pour une fonction du processus RCA(1) $\{X_t\}$. Celle-ci est donnée par :

$$\mathbb{P} \left(\left| \frac{1}{n} \sum_{t=0}^{n-1} (h(X_t) - \mathbb{E}h(X_t)) \right| > \epsilon \right) \leq c e^{-dn\epsilon^2}, \quad (2.4)$$

où h est une fonction mesurable et $c, d > 0$. Cette inégalité est classique dans le cas où $h(x) = x$, que la suite $\{X_t\}$ soit indépendante ou qu'elle constitue une suite de différences de martingale.

Certains concepts et lemmes nécessaires à la démonstration du résultat principal, sont introduits ci-après (on renvoie à [7] pour une présentation détaillée).

Définition 2.2.1. Soit $\{V_t, t \in \mathbb{Z}\}$ une suite de vecteurs aléatoires. Soit $\mathcal{F}_{-\infty}^t = \sigma(V_s, s \leq t)$, et $\mathcal{F}_{t+m}^{+\infty} = \sigma(V_s, s \geq t+m)$ pour chaque $t, m \in \mathbb{Z}$. Le coefficient du mélange fort est

$$\alpha_m = \sup_{t \in \mathbb{Z}} \left(\sup_{G \in \mathcal{F}_{-\infty}^t, H \in \mathcal{F}_{t+m}^{+\infty}} |\mathbb{P}(G \cap H) - \mathbb{P}(G)\mathbb{P}(H)| \right).$$

La suite $\{V_t\}$ est dite α -mélangeante (ou fortement mélangeante) si

$$\lim_{m \rightarrow \infty} \alpha_m = 0.$$

Une suite est dite α -mélangeante de taux $-a_0$ si $\alpha_m = O(m^{-a})$ pour un certain $a > a_0$.

Le résultat suivant, dû à Ibragimov [9] permet de majorer la norme de l'écart entre une variable et son espérance conditionnelle par rapport au passé en fonction du coefficient de mélange fort.

Théorème 2.2.1. [7]

Soit $\{V_t, t \in \mathbb{Z}\}$ une suite de vecteurs aléatoires α -mélangeante. Alors, pour tout $r \geq p \geq 1$ on a l'inégalité suivante :

$$\|\mathbb{E}(X_{t+m} \mid \mathcal{F}_{-\infty}^t) - \mathbb{E}(X_{t+m})\|_p \leq 2(2^{1/p} + 1) \alpha_m^{1/p-1/r} \|X_{t+m}\|_r.$$

Afin de démontrer le théorème principal, nous devons faire une autre hypothèse sur la structure mutuelle entre les coefficients aléatoires et le processus d'erreur.

(A5) : $\{(B_t, Y_t), t \in \mathbb{Z}\}$ est α -mélangeant de taux $-a$ pour un certain $a > 2$.

il est commode d'introduire la notion de *mixingale*, qui généralise à la fois les suites de différences de martingale et les processus mélangeants.

Définition 2.2.2. Soit $(X_t)_{t \in \mathbb{Z}}$ une suite de variables aléatoires intégrables. Soit $(\mathcal{F}_t)_{t \in \mathbb{Z}}$ une filtration.

La suite $\{X_t, \mathcal{F}_t\}$ est appelée un L_p -mixingale si, pour $p \geq 1$, il existe des suites de constantes non négatives $\{c_t, t \in \mathbb{Z}\}$ et $\{\zeta_m, m \in \mathbb{N}_0\}$ telles que $\zeta_m \rightarrow 0$ quand $m \rightarrow +\infty$ et

$$\|\mathbb{E}[X_t | \mathcal{F}_{t-m}]\|_p \leq c_t \zeta_m, \quad \|X_t - \mathbb{E}[X_t | \mathcal{F}_{t+m}]\|_p \leq c_t \zeta_{m+1}$$

sont vérifiées pour tous $t \in \mathbb{Z}$ et $m \in \mathbb{N}_0$.

Le L_p -mixingale est dit de taille $-a_0 > 0$ si $\zeta_m = O(m^{-a})$ pour un certain $a > a_0$.

On introduit le concept de dépendance de proche époque (Near Epoch Dependence), qui mesure la capacité d'une série à être approximée par une fonction d'un voisinage fini d'un processus sous-jacent.

Définition 2.2.3. Soit $\{X_t, t \in \mathbb{Z}\}$ une suite de variables aléatoires intégrables. Soit $\{V_t, t \in \mathbb{Z}\}$ une suite, éventuellement vectorielle, de variables aléatoire.

Soit la filtration $\{\mathcal{F}_{t-m}^{t+m}, m \in \mathbb{N}_0\}$ définie pour tout $t \in \mathbb{Z}$ et $m \in \mathbb{N}$ par $\mathcal{F}_{t-m}^{t+m} = \sigma(V_{t-m}, \dots, V_{t+m})$.

La suite $\{X_t\}$ est dite dépendante de proche époque en norme L_p (L_p -NED) sur $\{V_t\}$ si, pour $p > 0$, il existe des suites de constantes non négatives $\{d_t, t \in \mathbb{Z}\}$ et $\{\nu_m, m \in \mathbb{N}_0\}$ telles que $\nu_m \rightarrow 0$ quand $m \rightarrow +\infty$, et pour tous $t \in \mathbb{Z}$ et $m \in \mathbb{N}_0$ on ait

$$\|X_t - \mathbb{E}[X_t | \mathcal{F}_{t-m}^{t+m}]\|_p \leq d_t \nu_m$$

Le L_p -NED est de taille $-a_0 > 0$ si $\nu_m = O(m^{-a})$ pour un certain $a > a_0$.

L'idée derrière l'utilisation des concepts que l'on vient de définir pour démontrer l'inégalité exponentielle (2.4) est brièvement la suivante : une fonction lipschitzienne préserve la propriété L_p -NED, un processus L_p -NED est un L_p -mixingale sous certaines hypothèses, et la somme d'éléments d'un L_p -mixingale peut être divisée en somme d'éléments de différences de martingale et d'un résidu.

Lemme 2.2.3. Soit $\{X_t, t \in \mathbb{Z}\}$ une suite L_2 -NED (Near-Epoch Dependent) de taille $-a$ sur $\{V_t\}$ avec constantes $\{d_t\}$. Soit h une fonction vérifiant l'hypothèse **(A3)**. Alors $\{h(X_t), t \in \mathbb{Z}\}$ est aussi L_2 -NED de taille $-a$ sur $\{V_t\}$, avec constantes $\{c_h \cdot d_t\}$.

Preuve. Fixons $t \in \mathbb{Z}$ et $m \geq 0$. Pour alléger l'écriture, notons

$$\mathcal{F} = \mathcal{F}_{t-m}^{t+m}, \quad \tilde{X}_t = \mathbb{E}[X_t \mid \mathcal{F}].$$

La fonction h est continue donc borélienne et $\tilde{X}_t$ est $\mathcal{F}$ -mesurable. Il en découle que $h(\tilde{X}_t)$ est $\mathcal{F}$ -mesurable.

Par définition de l'espérance conditionnelle, celle-ci minimise la distance L_2 à $h(X_t)$ parmi toutes les variables $\mathcal{F}$ -mesurables. Donc

$$\|h(X_t) - \mathbb{E}[h(X_t) \mid \mathcal{F}]\|_2 \leq \|h(X_t) - h(\tilde{X}_t)\|_2.$$

En utilisant l'hypothèse A3, on peut écrire

$$|h(X_t) - h(\tilde{X}_t)| \leq c_h |X_t - \tilde{X}_t|.$$

En prenant la norme L_2 des deux côtés, on obtient

$$\|h(X_t) - h(\tilde{X}_t)\|_2 \leq c_h \|X_t - \tilde{X}_t\|_2.$$

La suite $\{X_t\}$ étant NED, on a directement

$$\|X_t - \tilde{X}_t\|_2 = \|X_t - \mathbb{E}[X_t \mid \mathcal{F}_{t-m}^{t+m}]\|_2 \leq d_t \nu_m.$$

On conclue que

$$\|h(X_t) - \mathbb{E}[h(X_t) \mid \mathcal{F}_{t-m}^{t+m}]\|_2 \leq c_h d_t \nu_m.$$

Ainsi, $\{h(X_t)\}$ est L_2 -NED sur $\{V_t\}$ avec les mêmes coefficients ν_m (donc la même taille $-a$) et des constantes $c_h d_t$. $\square$

Lemme 2.2.4. Soit $\{X_t, t \in \mathbb{Z}\}$ une suite centrée et L_r -bornée, avec $r > 1$. Soit $\{V_t, t \in \mathbb{Z}\}$ un processus α -mélangeant de taille $-a$. Si $\{X_t\}$ est L_p -NED de taille $-b$ sur $\{V_t\}$, avec $1 \leq p < r$ et des constantes $\{d_t\}$, alors $\{X_t, \mathcal{F}_t\}$ est un L_p -mixingale de taille $-\min(b, a(1/p - 1/r))$ avec des constantes $\{c_t\}$ telles que $c_t = O(\max(\|X_t\|_r, d_t))$.

Preuve. On pose $\mathcal{F}_t = \sigma(V_s, s \leq t)$, $\mathcal{F}_t^{t+k} = \sigma(V_t, \dots, V_{t+k})$, et $\mathcal{G}_{t,k} = \mathcal{F}_{t-k}^{t+k}$. Soit $m \geq 1$ et $k = \lfloor m/2 \rfloor$. Puisque $t - m \leq t - k$, alors $\mathcal{F}_{t-m} \subset \mathcal{F}_{t-k}^{t+k}$. Par l'inégalité triangulaire

$$\begin{aligned} \|\mathbb{E}(X_t \mid \mathcal{F}_{t-m})\|_p &\leq \|\mathbb{E}(X_t - \mathbb{E}(X_t \mid \mathcal{F}_{t-k}^{t+k}) \mid \mathcal{F}_{t-m})\|_p \\ &\quad + \|\mathbb{E}(\mathbb{E}(X_t \mid \mathcal{F}_{t-k}^{t+k}) \mid \mathcal{F}_{t-m})\|_p. \end{aligned} \quad (2.5)$$

En utilisant l'inégalité de Jensen et le fait que la suite $\{X_t\}$ est L_p -NED sur $\{V_t\}$, on obtient

$$\|\mathbb{E}(X_t - \mathbb{E}(X_t \mid \mathcal{F}_{t-k}^{t+k}) \mid \mathcal{F}_{t-m})\|_p \leq \|X_t - \mathbb{E}(X_t \mid \mathcal{F}_{t-k}^{t+k})\|_p \leq d_t \nu_k,$$

où ν_k est la suite de taille $-b$ donnée par la propriété NED. Comme $k \sim m/2$, $\nu_k = O(m^{-b})$.

Pour le deuxième terme dans 2.5, la variable aléatoire $Z = \mathbb{E}(X_t \mid \mathcal{F}_{t-k}^{t+k})$ est une fonction mesurable dépendant d'un nombre fini de retards de $V_{t-k}, \dots, V_{t+k}$, et possède donc, pour k fini, la même propriété de mélange que la suite $\{V_t\}$ avec la même taille. D'après le théorème 2.2.1, pour tout $p < r$ on a

$$\|\mathbb{E}(Z \mid \mathcal{F}_{t-m})\|_p \leq 6 \alpha_{m-k}^{(p-1)/p} \|Z\|_p.$$

Puisque $m - k \geq k$ et α_n est décroissante, $\alpha_{m-k} \leq \alpha_k$. De plus, $p < r$ donc

$$\|Z\|_p \leq \|X_t\|_p \leq \|X_t\|_r.$$

Ainsi

$$\|\mathbb{E}(\mathbb{E}(X_t \mid \mathcal{F}_{t-k}^{t+k}) \mid \mathcal{F}_{t-m})\|_p \leq 6 \|X_t\|_r \alpha_k^{(p-1)/p}.$$

Par suite

$$\|\mathbb{E}(X_t \mid \mathcal{F}_{t-m})\|_p \leq d_t \nu_k + 6 \|X_t\|_r \alpha_k^{(p-1)/p}.$$

On a $\nu_k = O(m^{-b})$ et $\alpha_k^{(p-1)/p} = O(m^{-a(1-1/p)})$. La condition $p < r$ implique $1 - 1/p \geq 1/p - 1/r$ et comme $\alpha_k \leq 1$, alors

$$\alpha_k^{(p-1)/p} \leq \alpha_k^{1/p-1/r}.$$

Or

$$\alpha_k^{1/p-1/r} = O(k^{-a(1/p-1/r)}) = O(m^{-a(1/p-1/r)}).$$

D'où

$$\|\mathbb{E}(X_t \mid \mathcal{F}_{t-m})\|_p \leq (d_t + 6 \|X_t\|_r) \zeta_m.$$

où $\zeta_m = m^{-\min(b, a(1/p-1/r))}$. D'autre part $k \leq m$ entraîne que $\mathcal{G}_{t,k} \subset \mathcal{F}_{t+m}$. Par conséquent

$$\begin{aligned} \|X_t - \mathbb{E}(X_t \mid \mathcal{F}_{t+m})\|_p &\leq 2 \|X_t - \mathbb{E}(X_t \mid \mathcal{F}_{t-k}^{t+k})\|_p \\ &\leq 2d_t\nu_m \\ &\leq 2d_t\zeta_m \end{aligned}$$

On conclue que $\{X_t, \mathcal{F}_t\}_{t \in \mathbb{Z}}$ est une L_p -mixingale de taille $-\min(b, a(1/p - 1/r))$. $\square$

Lemme 2.2.5. *Soit $\{X_t, \mathcal{F}_t\}_{t \in \mathbb{Z}}$ un L_1 -mixingale stationnaire d'ordre -1 . Il existe une décomposition*

$$X_t = R_t - R_{t+1} + W_t, \quad (2.6)$$

où $\mathbb{E}|R_t| < +\infty$ et $\{W_t, \mathcal{F}_t\}$ est une suite de martingale différence stationnaire.

L'équation 2.6 implique immédiatement que

$$\sum_{t=0}^{n-1} X_t = R_0 - R_n + \sum_{t=0}^{n-1} W_t.$$

Preuve. Pour chaque $t \in \mathbb{Z}$ et $m \geq 0$, on pose

$$R_{m,t} = \sum_{s=0}^m [\mathbb{E}(X_{t+s} \mid \mathcal{F}_{t-1}) - X_{t-s-1} + \mathbb{E}(X_{t-s-1} \mid \mathcal{F}_{t-1})], \quad (2.7)$$

et

$$W_{m,t} = \sum_{s=-m_0}^m [\mathbb{E}(X_{t+s} \mid \mathcal{F}_t) - \mathbb{E}(X_{t+s} \mid \mathcal{F}_{t-1})] + \mathbb{E}(X_{t+m+1} \mid \mathcal{F}_t) + X_{t-m-1} - \mathbb{E}(X_{t-m-1} \mid \mathcal{F}_{t-1}). \quad (2.8)$$

On va d'abord étudier, pour t fixé, la convergence des suites $(W_{m,t})$ et $(R_{m,t})$ quand $m \rightarrow +\infty$.

Par stationnarité, la loi du vecteur aléatoire $(X_{t+s}, X_{t-1}, X_{t-2}, \dots)$ est la même que celle $(X_t, X_{t-s-1}, X_{t-s-2}, \dots)$. Ceci donne

$$\mathbb{E}(X_{t+s} \mid \mathcal{F}_{t-1}) = \mathbb{E}(X_t \mid \mathcal{F}_{t-s-1}),$$

de même

$$\mathbb{E} |X_{t-s-1} - \mathbb{E}(X_{t-s-1} | \mathcal{F}_{t-1})| = \mathbb{E} |X_t - \mathbb{E}(X_t | \mathcal{F}_{t+s})|.$$

Par l'inégalité triangulaire, et comme (X_t) est une mixingale de taille -1 ,

$$\begin{aligned} \mathbb{E} |R_t| &\leq \sum_{s=0}^{\infty} \mathbb{E} |\mathbb{E}(X_t | \mathcal{F}_{t-s-1})| + \sum_{s=0}^{\infty} \mathbb{E} |X_t - \mathbb{E}(X_t | \mathcal{F}_{t+s})| \\ &\leq 2 \sum_{s=0}^{\infty} \zeta_s < \infty. \end{aligned}$$

la mixingale (X_t) est dans L_1 . De l'écriture $W_t = X_t - R_t + R_{t+1}$, on tire que

$$\mathbb{E} |W_t| \leq \mathbb{E} |X_t| + 2\mathbb{E} |R_t| < \infty.$$

Montrons que (W_t) est une différence de martingale. Puisque $\mathcal{F}_{t-1} \subset \mathcal{F}_t$, alors

$$\begin{aligned} \mathbb{E}(W_{m,t} | \mathcal{F}_{t-1}) &= \sum_{s=-m}^m \{ \mathbb{E} [\mathbb{E}(X_{t+s} | \mathcal{F}_t) | \mathcal{F}_{t-1}] - \mathbb{E} [\mathbb{E}(X_{t+s} | \mathcal{F}_{t-1}) | \mathcal{F}_{t-1}] \} \\ &\quad + \mathbb{E} [\mathbb{E}(X_{t+m+1} | \mathcal{F}_t) | \mathcal{F}_{t-1}] + \mathbb{E}(X_{t-m-1} | \mathcal{F}_{t-1}) \\ &\quad - \mathbb{E} [\mathbb{E}(X_{t-m-1} | \mathcal{F}_{t-1}) | \mathcal{F}_{t-1}] \\ &= \mathbb{E}(X_{t+m+1} | \mathcal{F}_{t-1}) \quad p.s. \end{aligned}$$

Mais

$$\mathbb{E} |\mathbb{E}(X_{t+m+1} | \mathcal{F}_{t-1})| = \|\mathbb{E}(X_{t+m+1} | \mathcal{F}_{t-1})\|_1 \leq c_t \zeta_{m+1} \xrightarrow{m \rightarrow \infty} 0,$$

soit $\mathbb{E}(W_{m,t} | \mathcal{F}_{t-1}) \xrightarrow[m \rightarrow \infty]{L^1} 0$ et donc

$$\mathbb{E}(W_{m,t} | \mathcal{F}_{t-1}) \xrightarrow[m \rightarrow \infty]{\mathbb{P}} 0.$$

Il en découle que toute sous-suite $\{m_k, k \in \mathbb{N}\}$ contient une sous-sous-suite $\{m_{k(j)}, j \in \mathbb{N}\}$ telle que

$$\left| \mathbb{E} \left(W_{m_{k(j)}, t} | \mathcal{F}_{t-1} \right) \right| \xrightarrow[j \rightarrow \infty]{p.s.} 0$$

Puisque $W_{m_{k(j)},t} \rightarrow W_t$ pour toute sous-suite de ce type, on peut conclure que

$$\mathbb{E}(W_t \mid \mathcal{F}_{t-1}) = 0 \quad \text{p.s.}$$

Ainsi $\{W_t, \mathcal{F}_t\}$ est une différence de martingale.

Vérifions à présent que pour tout $m \in \mathbb{N}$, on a

$$W_{m,t} = X_t - R_{m,t} + R_{m,t+1} \quad (2.9)$$

Commençons par expliciter $R_{m,t+1}$. En utilisant la définition $W_{m,t}$ dans (2.8) avec t remplacé par $t+1$, on a

$$\begin{aligned} R_{m,t+1} &= \sum_{s=0}^m [\mathbb{E}(X_{t+1+s} \mid \mathcal{F}_t) - X_{t-s} + \mathbb{E}(X_{t-s} \mid \mathcal{F}_t)] \\ &= \sum_{s=1}^{m+1} \mathbb{E}(X_{t+s} \mid \mathcal{F}_t) - \sum_{s=0}^m X_{t-s} + \sum_{s=0}^m \mathbb{E}(X_{t-s} \mid \mathcal{F}_t) \\ &= \sum_{s=1}^{m+1} \mathbb{E}(X_{t+s} \mid \mathcal{F}_t) - \sum_{s=0}^m X_{t-s} + \sum_{s=-m}^0 \mathbb{E}(X_{t+s} \mid \mathcal{F}_t) \\ &= \sum_{s=-m}^{m+1} \mathbb{E}(X_{t+s} \mid \mathcal{F}_t) - \sum_{s=0}^m X_{t-s} \end{aligned}$$

De la même façon, on peut écrire

$$= \sum_{s=-m-1}^m \mathbb{E}(X_{t+s} \mid \mathcal{F}_{t-1}) - \sum_{s=1}^{m+1} X_{t-s}$$

Par suite

$$\begin{aligned} X_t - R_{m,t} + R_{m,t+1} &= X_{t-m-1} - \sum_{s=-m-1}^m \mathbb{E}(X_{t+s} \mid \mathcal{F}_{t-1}) + \sum_{s=-m}^{m+1} \mathbb{E}(X_{t+s} \mid \mathcal{F}_t) \\ &= X_{t-m-1} - \mathbb{E}(X_{t-m-1} \mid \mathcal{F}_{t-1}) \\ &\quad + \sum_{s=-m}^m [\mathbb{E}(X_{t+s} \mid \mathcal{F}_t) - \mathbb{E}(X_{t+s} \mid \mathcal{F}_{t-1})] + \mathbb{E}(X_{t+m+1} \mid \mathcal{F}_t) \\ &= W_{m,t} \end{aligned}$$

Ainsi la relation 2.9 est satisfaite. En passant à la limite lorsque $m \rightarrow \infty$, on obtient

$$W_t = X_t - R_t + R_{t+1}. \quad (2.10)$$

□

Corollaire 2.2.1 (bornes uniformes). *Si, de plus la suite (X_t) dans le lemme 2.2.5 est une L^1 -mixingale uniformément bornée de taille -1 avec une suite de constantes c_t dans la définition 2.2.2 est bornée, alors les deux suites R_t et W_t sont uniformément bornées.*

Preuve. On a de la preuve du lemme 2.2.5

$$R_{m,t} = \sum_{s=0}^m \left(\mathbb{E}[X_{t+s} \mid \mathcal{F}_{t-1}] - X_{t-s-1} + \mathbb{E}[X_{t-s-1} \mid \mathcal{F}_{t-1}] \right)$$

et

$$R_t = \lim_{m \rightarrow \infty} R_{m,t}.$$

Donc, en utilisant l'inégalité triangulaire

$$|R_{m,t}| \leq \sum_{s=0}^m \left(|\mathbb{E}[X_{t+s} \mid \mathcal{F}_{t-1}]| + |X_{t-s-1} - \mathbb{E}[X_{t-s-1} \mid \mathcal{F}_{t-1}]| \right).$$

Comme (X_t) est une L^1 -mixingale, on a :

$$|\mathbb{E}[X_{t+s} \mid \mathcal{F}_{t-1}]| \leq c_{t+s} \zeta_{s+1},$$

et

$$|X_{t-s-1} - \mathbb{E}[X_{t-s-1} \mid \mathcal{F}_{t-1}]| \leq c_{t-s-1} \zeta_{s+1}.$$

Par conséquent

$$|R_{m,t}| \leq \sum_{s=0}^m (c_{t+s} + c_{t-s-1}) \zeta_{s+1}.$$

La suite (c_t) étant borée par hypothèse, il existe donc une constante $c > 0$ telle que $c_t \leq c$ pour tout t . Il s'en suit que

$$|R_{m,t}| \leq 2c \sum_{s=0}^m \zeta_{s+1}.$$

Comme $\zeta_s = O(s^{-a})$ avec $a > 1$, alors la série $\sum \zeta_s$ converge et la suite $(R_{m,t})$ est uniformément bornée c'est à dire il existe une constante $d > 0$, telle que

$$\forall t \in \mathbb{Z}, \forall m \in \mathbb{N}, |R_{m,t}| \leq d.$$

Or $R_t = \lim_{m \rightarrow \infty} R_{m,t}$. Donc $(R_{m,t})$ est uniformément bornée.
En suite,

$$\forall t \in \mathbb{Z}, |R_t| \leq d.$$

Comme (X_t) est uniformément bornée par hypothèse, alors (W_t) est aussi uniformément bornée $\square$

Considérons les hypothèses supplémentaires suivantes sur le processus RCA(1) $\{X_t\}$ pour utiliser les techniques précédentes.

(A5) : $\mathbb{P}(|X_0| \leq c_X) = 1$ pour une certaine constante $c_X > 0$.

(A6) : $\mathbb{P}(|B_0| \leq c_B) = 1$ pour une certaine constante $c_B > 0$.

Puisque le processus RCA(1) $\{X_t\}$ est strictement stationnaire, la bornitude de X_0 selon A5 assure la bornitude uniforme de toute la suite $\{X_t\}$, c'est-à-dire qu'il existe une constante $c_X > 0$ telle que $\mathbb{P}(|X_t| \leq c_X, \forall t \in \mathbb{Z}) = 1$.

Lemme 2.2.6. *Sous les hypothèses A5 et A6, les processus $\{B_t\}$ et $\{Y_t\}$ de la définition 1.3.1 sont uniformément bornés.*

Preuve. En utilisant l'équation 2.1, on tire que

$$|X_t| = |(\beta + B_t)X_{t-1} + Y_t|.$$

Par suite

$$|X_t| \geq |B_t X_{t-1} + Y_t| - |\beta X_{t-1}|.$$

Ainsi,

$$|B_t X_{t-1} + Y_t| \leq |X_t| + |\beta X_{t-1}|.$$

Comme la suite (X_t) est uniformément bornée, alors il existe $c_X > 0$ tel que :

$$|X_t| \leq c_X, \quad \forall t \in \mathbb{Z}.$$

On en déduit :

$$|B_t X_{t-1} + Y_t| \leq c_X + |\beta|c_X = (1 + |\beta|)c_X.$$

Ainsi, la suite $(B_t X_{t-1} + Y_t)$ est uniformément bornée.

D'autre part

$$|B_t X_{t-1} + Y_t| \geq |Y_t| - |B_t X_{t-1}|.$$

Donc :

$$|Y_t| \leq |B_t X_{t-1} + Y_t| + |B_t X_{t-1}|.$$

Or,

$$|B_t X_{t-1}| \leq |B_t| \cdot |X_{t-1}| \leq c_B \cdot c_X.$$

Finalement,

$$|Y_t| \leq (1 + |\beta|)c_X + c_B c_X.$$

Donc, il existe une constante $c > 0$ telle que :

$$|Y_t| \leq c, \quad \forall t \in \mathbb{Z}.$$

Ainsi les deux suites (B_t) et (Y_t) sont uniformément bornées. $\square$

Lemme 2.2.7. Soit $\{X_t, t \in \mathbb{Z}\}$ un processus RCA(1) vérifiant les conditions A4, A5 et A6.

Soit $\phi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue telle que $h(x) = x\phi(x)$ satisfasse l'hypothèse A3. On pose

$$Z_t = h(X_t) - \mathbb{E}[h(X_t)].$$

Alors il existe une décomposition de $\{Z_t\}$ donnée par le Lemme 2.2.5 avec $\mathcal{F}_t = \sigma(B_s, Y_s; s \leq t)$, telles que les deux suites $\{R_t\}$ et $\{W_t\}$ soient uniformément bornées.

Preuve. Premièrement, nous allons montrer que $\{X_t\}$ est L2-NED sur $\{B_t, Y_t\}$ de taille arbitraire.

Notons $F_{t-m}^{t+m} = \sigma(B_s, Y_s; s = t-m, \dots, t+m)$ et vérifions la définition 2.2.3. La relation (2.1) et la mesurabilité de B_s et Y_s par rapport à F_{t-m}^{t+m} pour $s = t, t-1, \dots, t-m$ donnent :

$$\begin{aligned} \|X_t - \mathbb{E}[X_t | F_{t-m}^{t+m}]\|_2 &= \|(\beta + B_t)X_{t-1} + Y_t - \mathbb{E}[(\beta + B_t)X_{t-1} + Y_t | F_{t-m}^{t+m}]\|_2 \\ &= \|(\beta + B_t)X_{t-1} - (\beta + B_t)\mathbb{E}[X_{t-1} | F_{t-m}^{t+m}]\|_2 \\ &= \|(\beta + B_t)(X_{t-1} - \mathbb{E}[X_{t-1} | F_{t-m}^{t+m}])\|_2 \\ &= \dots \\ &= \left\| \left(\prod_{i=0}^m (\beta + B_{t-i}) \right) \cdot (X_{t-m-1} - \mathbb{E}[X_{t-m-1} | F_{t-m}^{t+m}]) \right\|_2. \end{aligned}$$

En utilisant l'inégalité triangulaire et la borne uniforme $|X_t| \leq c_X$ (constante positive), on obtient :

$$\begin{aligned} \|X_t - \mathbb{E}[X_t \mid F_{t-m}^{t+m}]\|_2 &\leq \left\| \left(\prod_{i=0}^m (\beta + B_{t-i}) \right) X_{t-m-1} \right\|_2 + \left\| \left(\prod_{i=0}^m (\beta + B_{t-i}) \right) \mathbb{E}[X_{t-m-1} \mid F_{t-m}^{t+m}] \right\|_2 \\ &\leq \left\| \prod_{i=0}^m (\beta + B_{t-i}) \right\|_2 c_X + \left\| \prod_{i=0}^m (\beta + B_{t-i}) \right\|_2 c_X. \end{aligned}$$

Calculons la norme du produit :

$$\left\| \prod_{i=0}^m (\beta + B_{t-i}) \right\|_2 = \left(\mathbb{E} \prod_{i=0}^m (\beta + B_{t-i})^2 \right)^{\frac{1}{2}} = (\mathbb{E}(\beta + B_t)^2)^{\frac{m+1}{2}} = (\beta^2 + \sigma_B^2)^{\frac{m+1}{2}}.$$

Ainsi,

$$\|X_t - \mathbb{E}[X_t \mid F_{t-m}^{t+m}]\|_2 \leq 2 c_X (\beta^2 + \sigma_B^2)^{\frac{m+1}{2}}.$$

On suppose $\beta^2 + \sigma_B^2 < 1$; alors $(\beta^2 + \sigma_B^2)^{(m+1)/2}$ est $O(m^{-a})$ pour tout $a > 0$. Par conséquent, $\{X_t\}$ est L2-NED avec des constantes $d_t = 2c_X$ dans la définition 2.2.3

Maintenant, la fonction h étant de Lipschitz avec une constante c_h , le lemme 2.2.5 assure que $(h(X_t))$ est L2-NED sur $\{B_t, Y_t\}$ de taille arbitraire avec des constantes $d_t = 2cc_h$. Comme $Z_t = h(X_t) + C^{ste}$ alors même (Z_t) est L2-NED sur $\{B_t, Y_t\}$ de taille arbitraire avec des constantes $d_t = 2cc_h$.

$\{X_t\}$ est bornée et h est continue, donc la suite $\{Z_t\}$ est L_r -bornée pour tout $r > 1$. La suite (B_t, Y_t) est par hypothèse α -mélangeante, en appliquant le lemme 2.2.4, on obtient que $\{Z_t\}$ est une $L2$ -mixingale de taille $-\min(b, a(1/2 - 1/r))$. Observons que $\min(b, a(1/2 - 1/r)) = a/2 - a/r > 1$ car a/r peut être rendu arbitrairement petit en augmentant r . Les constantes c_t dans la définition 3 sont égales à $O(\max(\|X_t\|_r, 2cc_h))$; elles sont bornées en raison de la bornitude de $\{X_t\}$. Ainsi, les deux suites $\{R_t\}$ et $\{W_t\}$ dans la décomposition sont uniformément bornées d'après le corollaire du lemme 2.2.5. $\square$

Lemme 2.2.8. *Soit $\{X_t\}$ un processus RCA(1) vérifiant les hypothèses A4 à A6. Soit $\phi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue et mesurable telle que $h(x) =$*

$x\phi(x)$ satisfasse l'hypothèse A3 (condition de Lipschitz). Alors il existe des constantes $c > 0$ et $d > 0$ telles que, pour tout $\varepsilon > 0$ et tout $n \in \mathbb{N}$,

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{t=0}^{n-1} (h(X_t) - \mathbb{E}[h(X_t)])\right| > \varepsilon\right) \leq c e^{-dn\varepsilon^2}.$$

Preuve. Posons

$$Z_t = h(X_t) - \mathbb{E}[h(X_t)], \quad \forall t \in \mathbb{Z}.$$

D'après le lemme 2.2.5, la suite (Z_t) admet la décomposition suivante :

$$\sum_{t=0}^{n-1} Z_t = R_0 - R_n + \sum_{t=0}^{n-1} W_t,$$

où (W_t) est une suite de différences de martingale uniformément bornée, et (R_t) est une suite uniformément bornée.

On a

$$\begin{aligned} \mathbb{P}\left(\left|\frac{1}{n} \sum_{t=0}^{n-1} Z_t\right| > \varepsilon\right) &= P\left(\left|\frac{1}{n} (R_0 - R_n + \sum_{t=0}^{n-1} W_t)\right| > \varepsilon\right) \\ &\leq \mathbb{P}\left(\frac{1}{n} |R_0 - R_n| + \left|\frac{1}{n} \sum_{t=0}^{n-1} W_t\right| > \varepsilon\right) \\ &\leq \mathbb{P}\left(|R_0 - R_n| > \frac{n\varepsilon}{2}\right) + \mathbb{P}\left(\left|\frac{1}{n} \sum_{t=0}^{n-1} W_t\right| > \frac{\varepsilon}{2}\right). \end{aligned}$$

(W_t) étant une suite de différences de martingale uniformément bornée, en appliquant inégalité de Hoeffding, on obtient

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{t=0}^{n-1} W_t\right| > \frac{\varepsilon}{2}\right) \leq c_1 \exp(-d_1 n \varepsilon^2),$$

où $c_1, d_1 > 0$. D'autre part

$$\mathbb{P}(|R_0 - R_n| > n\varepsilon/2) \leq \mathbb{P}(|R_0| + |R_n| > n\varepsilon/2).$$

Ensuite, par inégalité de Markov

$$\mathbb{P}(|R_0| + |R_n| > n\varepsilon/2) \leq \mathbb{E}(e^{\varepsilon(|R_0| + |R_n|)}) e^{-n\varepsilon^2/2}.$$

Puisque (R_t) est uniformément bornée, alors l'espérance ci-dessus est finie. Ainsi

$$\mathbb{P}(|R_0| + |R_n| > n\varepsilon/2) \leq c_2 \exp(-d_2 n \varepsilon^2),$$

avec $c_2, d_2 > 0$. En combinant les deux estimations, on obtient

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{t=0}^{n-1} Z_t\right| > \varepsilon\right) \leq c \exp(-dn \varepsilon^2),$$

où $c = \max(c_1, c_2)$ et $d = \min(d_1, d_2)$.

□

2.3 Normalité asymptotique de l'estimateur $\widehat{\beta}_n$

Dans cette section, nous allons démontrer le résultat suivant :

Théorème 2.3.1. *Soit $(X_t)_{t \in \mathbb{Z}}$ un processus RCA(1) satisfaisant les hypothèses A4 à A6. Soit $\phi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction mesurable continue telle que $h(x) = x\phi(x)$ vérifie l'hypothèse A3.*

Alors, il existe une constante $c > 0$ telle que, pour tout $n \in \mathbb{N}$,

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left(\frac{\sqrt{n}(\widehat{\beta}_n(\phi) - \beta)}{\sqrt{V(\phi)}} \leq x\right) - \Phi(x) \right| \leq c \frac{(\ln n)^3}{\sqrt{n}},$$

où $\widehat{\beta}_n(\phi)$ est l'estimateur défini par (2) et $V(\phi)$ est la variance asymptotique définie par (3).

Preuve. On introduit les quantités suivantes :

$$f_n = \frac{1}{\sqrt{n}} \sum_{t=1}^n \varphi(X_{t-1})(X_t - \beta X_{t-1}), \quad g_n = \frac{1}{n} \sum_{t=1}^n h(X_{t-1}).$$

Alors :

$$\sqrt{n}(\widehat{\beta}_n(\phi) - \beta) = \frac{f_n}{g_n}.$$

Posons :

$$U(\phi) = \mathbb{E}(\phi^2(X_1) w(X_1)), \quad V(\phi) = \frac{U(\phi)}{(\mathbb{E}[h(X_1)])^2}.$$

On obtient :

$$\frac{\sqrt{n}(\widehat{\beta}_n(\phi) - \beta)}{\sqrt{V(\phi)}} = \frac{\mathbb{E}[h(X_1)]}{\sqrt{U(\phi)}} \cdot \frac{f_n}{g_n}.$$

Pour tout $\varepsilon > 0$, le Lemme 2.2.1 donne :

$$\begin{aligned} & \sup_x \left| \mathbb{P} \left(\frac{\mathbb{E}[h(X_1)]}{\sqrt{U(\phi)}} \frac{f_n}{g_n} \leq x \right) - \Phi(x) \right| \\ & \leq \sup_y \left| \mathbb{P} \left(\frac{f_n}{\sqrt{U(\phi)}} \leq y \right) - \Phi(y) \right| + \mathbb{P} \left(\left| \frac{g_n}{\mathbb{E}[h(X_1)]} - 1 \right| > \varepsilon \right) + \varepsilon. \end{aligned}$$

La suite

$$U_t = \phi(X_{t-1})(X_t - \beta X_{t-1})$$

est une suite de différences de martingale, bornée. Donc, par le lemme de Berry–Esseen pour martingales :

$$\sup_y \left| \mathbb{P} \left(\frac{f_n}{\sqrt{U(\phi)}} \leq y \right) - \Phi(y) \right| \leq d \frac{(\ln n)^3}{\sqrt{n}}.$$

D'autre part,

$$\begin{aligned} \mathbb{P} \left(\left| \frac{g_n}{\mathbb{E}[h(X_1)]} - 1 \right| > \varepsilon \right) &= \mathbb{P} \left(\left| g_n - \mathbb{E}[h(X_1)] \right| > \varepsilon \mathbb{E}[h(X_1)] \right) \\ &= \mathbb{P} \left(\frac{1}{n} \sum_{t=1}^n (h(X_{t-1}) - \mathbb{E}h(X_{t-1})) > \varepsilon \mathbb{E}[h(X_1)] \right). \end{aligned}$$

D'après le Lemme 2.2.8, il existe $p, q > 0$ tels que :

$$\mathbb{P} \left(\frac{1}{n} \sum_{t=1}^n (h(X_{t-1}) - \mathbb{E}h(X_{t-1})) > \varepsilon \mathbb{E}[h(X_1)] \right) \leq p e^{-qn\varepsilon^2}.$$

On obtient :

$$\sup_x \left| \mathbb{P} \left(\frac{\mathbb{E}[h(X_1)]}{\sqrt{U(\phi)}} \frac{f_n}{g_n} \leq x \right) - \Phi(x) \right| \leq d \frac{(\ln n)^3}{\sqrt{n}} + p e^{-qn\varepsilon^2} + \varepsilon.$$

On choisit :

$$\varepsilon = \frac{(\ln n)^3}{\sqrt{n}}.$$

Alors $p e^{-qn\varepsilon^2} = p e^{-q(\ln n)^6}$ est négligeable et ε est du même ordre que le premier terme.

Par conséquent,

$$\sup_x \mathbb{P} \left(\frac{\sqrt{n}(\hat{\beta}_n(\varphi) - \beta)}{\sqrt{V(\phi)}} \leq x \right) - \Phi(x) \leq c \frac{(\ln n)^3}{\sqrt{n}}.$$

On a montré que

$$\frac{\sqrt{n}(\hat{\beta}_n(\phi) - \beta)}{\sqrt{V(\phi)}}$$

converge vers la loi normale standard avec une vitesse de convergence de l'ordre

$$\frac{(\ln n)^3}{\sqrt{n}}.$$

Remarque Ce théorème couvre à la fois l'estimateur des moindres carrés (choix $\phi(x) = x$) et l'estimateur de variance asymptotique minimale dans le modèle où $\sigma_B^2 = \sigma_Y^2$ (choix $\phi(x) = \frac{x}{1+x^2}$). En effet

1. Cas $\phi(\mathbf{x}) = \mathbf{x}$

Alors $h(x) = x\phi(x) = x^2$. Pour tout s, t ,

$$|h(X_s) - h(X_t)| = |X_s^2 - X_t^2| = |X_s + X_t| |X_s - X_t| \leq (|X_s| + |X_t|) |X_s - X_t| \leq 2c_X |X_s - X_t|$$

presque sûrement, car $|X_s|, |X_t| \leq c_X$. Donc h vérifie la condition de Lipschitz A3 avec $c_h = 2c_X$. De plus $h(x) > 0$ pour $x \neq 0$. Les conditions sont remplies.

2. Cas $\phi(\mathbf{x}) = \frac{\mathbf{x}}{1 + \mathbf{x}^2}$ (lorsque $\sigma_B^2 = \sigma_Y^2$)

Alors $h(x) = x\phi(x) = \frac{x^2}{1 + x^2}$. h est de classe C^1 sur $\mathbb{R}$ et sa dérivée est

$$h'(x) = \frac{2x}{(1 + x^2)^2}.$$

Sur l'intervalle $[-c_X, c_X]$ (qui contient presque sûrement toutes les réalisations de X_t), h' est continue, donc bornée. Posons

$$c_h = \max_{|x| \leq c_X} |h'(x)| < \infty.$$

Par le théorème des accroissements finis, pour tout s, t , il existe ξ entre X_s et X_t tel que

$$|h(X_s) - h(X_t)| = |h'(\xi)| |X_s - X_t| \leq c_h |X_s - X_t| \quad \text{p.s.}$$

Ainsi A3 est satisfaite. On a bien $h(x) > 0$ pour $x \neq 0$.

Conclusion

Dans les deux cas, les fonctions considérées vérifient toutes les hypothèses du théorème. Par conséquent, le résultat s'applique :

- à l'estimateur des moindres carrés ($\phi(x) = x$),
- ainsi qu'à l'estimateur optimal ($\phi(x) = \frac{x}{1+x^2}$).

Chapitre 3

Simulation

Ce chapitre propose une étude par simulation de modèles autorégressifs d'ordre 1 à coefficients aléatoires ($RCA(1)$). Nous considérons différentes distributions pour le coefficient aléatoire B_t (loi de Bernoulli et loi gaussienne). L'objectif est d'évaluer l'estimateur $\hat{\beta}_n(\phi)$ de $\beta = \mathbb{E}[a_t]$ introduit par Schick [14], ainsi que les prédicteurs qui lui sont associés, pour diverses fonctions de pondération ϕ . À partir de séries simulées, nous analysons l'impact du choix de Φ sur la qualité des prédictions hors échantillon, à l'aide de plusieurs critères d'erreur (notamment la MSE, la RMSE, et la MAE).

Nous réaliserons les simulations avec la version 4.2.2 du logiciel R, disponible au téléchargement à l'adresse suivante :

<https://cran.r-project.org/bin/windows/base/old/4.2.2>.

Notons que R est un langage de programmation libre et un environnement dédié aux statistiques ainsi qu'à la science des données. Il permet la manipulation de données, les calculs, la création de graphiques et la réalisation de diverses analyses statistiques, telles que la régression linéaire, les tests d'hypothèses, l'analyse de la variance, etc.

On va considérer un processus autorégressif à coefficients aléatoires $RCA(1)$ satisfaisant la relation

$$X_{t+1} = (\beta + B_t)X_t + Y_t, \quad t \in \mathbb{Z},$$

3.1 Introduction

On considère le processus RCA(1) défini par

$$X_t = (\beta + B_t) X_{t-1} + Y_t, \quad t \geq 1,$$

où β est un nombre réel et $(B_t)_{t \geq 1}$ est une suite de variables aléatoires centrées indépendantes et identiquement distribuées et de variance $\mathbb{E}[B_t^2] = \sigma_\beta^2$ et (Y_t) est un bruit blanc fort, c'est à dire une suite de variables aléatoires *i.i.d* centrées et de variance $\mathbb{E}[Y_0^2] = \sigma^2 < \infty$.

On pose $a_t = \beta + B_t$ et on suppose que la condition de stationnarité stricte (voir section 1.3.2) est satisfaite.

$$\mathbb{E}[\log |a|] = -\gamma < 0 \quad (\gamma > 0),$$

est satisfaite. La solution strictement stationnaire s'écrit, pour une réalisation donnée des bruits passés,

$$X_t^{\text{stat}} = \sum_{j=0}^{\infty} P_t^j Y_{t-j},$$

avec $P_t^j = \prod_{k=1}^j a_{t-i+1}$. On simule une trajectoire X_t^{sim} à l'instant t à partir d'une condition initiale déterministe $X_0^{\text{sim}} = x_0$ en utilisant, par récurrence, les relations

$$X_1^{\text{sim}} = a_1 X_0^{\text{sim}} + Y_1,$$

et pour $t \geq 1$,

$$\begin{aligned} X_t^{\text{sim}} &= a_t X_{t-1}^{\text{sim}} + Y_t \\ &= a_t (a_{t-1} X_{t-2}^{\text{sim}} + Y_{t-1}) + Y_t \\ &= a_t a_{t-1} (X_{t-2}^{\text{sim}} + a_t Y_{t-1}) + Y_t \\ &= a_t a_{t-1} X_{t-2}^{\text{sim}} + a_t Y_{t-1} + Y_t \\ &\vdots \\ &= \sum_{j=0}^{t-1} P_t^j Y_{t-j} + P_t^0 x_0. \end{aligned}$$

Remarquons que, pour tout entiers $t \geq 1$ et tout entiers j, k avec $k \leq j$ on a

$$P_t^j = P_t^k P_k^j.$$

En notant X_0^{stat} la valeur stationnaire à l'instant 0, on obtient pour tout $t \geq 1$

$$X_t^{\text{stat}} = \sum_{j=0}^{t-1} P_t^j Y_{t-j} + P_t^0 X_0^{\text{stat}}.$$

Par suite

$$X_t^{\text{stat}} - X_t^{\text{sim}} = P_t^0 (X_0^{\text{stat}} - x_0).$$

Ainsi, l'écart entre la trajectoire simulée et la trajectoire stationnaire est exactement le produit aléatoire P_t^0 multiplié par une constante aléatoire $(X_0^{\text{stat}} - x_0)$.

On étudie le comportement asymptotique de $P_t^0 = \prod_{i=1}^t a_i$, pour cela, on prend les logarithmes,

$$\log |P_t^0| = \sum_{i=1}^t \log |a_i|.$$

La loi forte des grands nombres (LGN) donne

$$\frac{1}{t} \log |P_t^0| \xrightarrow[t \rightarrow \infty]{p.s.} \mathbb{E}[\log |a|] = -\gamma.$$

Donc, pour presque toute réalisation, pour tout $\delta \in (0, \gamma)$, il existe $T(\omega)$ tel que pour tout $t \geq T(\omega)$,

$$\log |P_t^0| \leq -(\gamma - \delta)t$$

en particulier pour $\delta = \frac{\gamma}{2}$, on obtient

$$|P_t^0| \leq r^t,$$

avec $r = e^{-\frac{\gamma}{2}} < 1$. Ainsi, la différence $X_t^{\text{sim}} - X_t^{\text{stat}}$ tend vers 0 en probabilité (et même en moyenne). Par conséquent, pour t suffisamment grand, la loi de X_t^{sim} est très proche de la loi stationnaire π (celle de X_t^{stat} .) En pratique, On fixe un seuil de précision $\varepsilon > 0$. On veut t tel que

$$e^{-\gamma t} \leq \varepsilon.$$

Prenons $\varepsilon = e^{-6} \approx 0,0025$ (un choix conventionnel). on résout

$$e^{-\gamma t} \leq e^{-6} \iff t \geq \frac{6}{\gamma}.$$

Ce raisonnement conduit à la proposition suivante, qui définit la période de chauffe utilisée dans nos simulations.

Proposition 3.1.1. *[Choix du burn-in n_{start}] On pose*

$$n_{start} = 1 + \left\lceil \frac{6}{\gamma} \right\rceil.$$

Cette valeur garantit que, pour $t = n_{start}$, l'influence de la condition initiale est réduite à un facteur e^{-6} en espérance, et donc que la loi de X_t^{sim} est à une distance en variation totale inférieure à e^{-6} de la loi stationnaire π .

Remarque 3.1.1. *Dans la fonction de simulation, lorsque la durée du burn-in n_{start} n'est pas spécifiée et que $\mathbb{E}[\log |a|] < 0$, elle est déterminée automatiquement par la proposition 3.1.1. Si $\mathbb{E}[\log |a|] \geq 0$, le processus n'est pas strictement stationnaire ; on fixe alors $n_{start} = 100$ et un avertissement est émis.*

Avant d'illustrer les résultats théoriques par des exemples concrets, il convient de rappeler quelques indicateurs usuelles permettant d'évaluer la qualité des prévisions. Soient y_t la valeur observée et $\hat{y}_t$ la prévision associée, pour $t = 1, \dots, n$. Les erreurs de prévision les plus couramment utilisées sont les suivantes :

— L'erreur absolue moyenne (**MAE**) :

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|.$$

— L'erreur quadratique moyenne (**MSE**) :

$$\text{MSE} = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2.$$

— La racine carrée de l'erreur quadratique moyenne (**RMSE**) :

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}.$$

Ces indicateurs permettent de quantifier la précision des modèles de prévision et facilitent la comparaison entre différentes méthodes. Dans les exemples qui suivent, nous utiliserons ces critères pour évaluer la performance des procédures proposées.

3.2 Simulation et prévision d'un processus autorégressif à coefficients aléatoires $RCA(1)$ à deux régimes

Dans cette section, nous traitons le cas où la variable aléatoire B_t suit une loi de Bernoulli généralisée donnée par :

$$\mathbb{P}(B_t = a_1) = p_1, \quad \mathbb{P}(B_t = a_2) = p_2,$$

avec a_1, a_2, p_1 et p_2 sont des nombres réels qui seront choisis à travers les exemples avec la condition $p_1 + p_2 = 1$.

Dans ce cas, le processus $RCA(1)$ est dit à double régimes (double swtching). Il est clair que

$$\sigma_\beta^2 = \mathbb{E}(B_t^2) = a_1^2 p_1 + a_2^2 p_2.$$

La condition de stationnarité faible exige que

$$\beta^2 + a_1^2 p_1 + a_2^2 p_2 < 1.$$

3.2.1 Exemple 1

Dans un premier exemple, on prend $\beta = \frac{1}{2}$, et on suppose que la variable B_t prend les valeurs $-0,4$ et $0,4$ chacune avec probabilité $\frac{1}{2}$ (c'est-à-dire $\mathbb{P}(B_t = -0,4) = \frac{1}{2}$, $\mathbb{P}(B_t = 0,4) = \frac{1}{2}$). Ainsi, la variable $a_t = \beta + B_t$ prend deux valeurs : l'une faiblement positive $(0,1)$ et l'autre fortement positive $(0,9)$. Dans ce cas, le modèle est strictement stationnaire.

La Figure 3.1 présente le diagnostic de la série simulée issue d'un processus $RCA(1)$ à deux régimes :

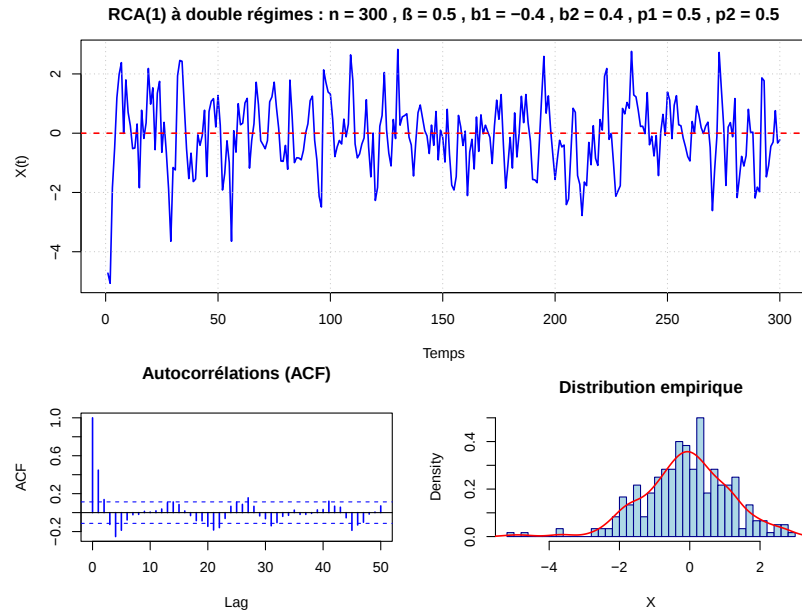


FIGURE 3.1 – Processus RCA(1) à double régimes

La Figure 3.2 compare les prédictions à un pas issues de l'estimateur des moindres carrés ordinaires (MCO) et de l'estimateur pondéré optimal ($\phi(x) = x/(\sigma^2 + \sigma_B^2 x^2)$) sur l'échantillon test.

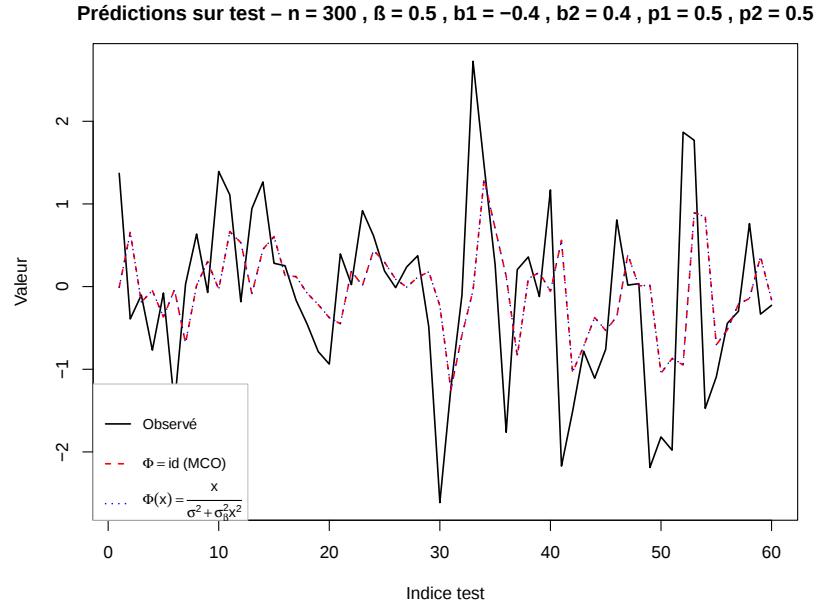


FIGURE 3.2 – Prédiction sur test

Le tableau 3.1 présente les estimateurs une comparaison des erreurs de prédictions à un pas sur l'échantillon test pour les deux méthodes d'estimation.

TABLE 3.1 – Comparaison des erreurs de prédiction (processus Bernoulli)

Méthode	$\hat{\beta}_n$	MSE	RMSE	MAE
MCO ($\Phi=id$)	0.45014	1.529650	1.236790	1.036558
$\Phi(x) = \frac{x}{\sigma^2 + \sigma_\beta^2 x^2}$	0.48082	1.509803	1.228740	1.023596

Dans le premier exemple, et comme le montre la tableau 3.1, l'estimateur des moindres carrés pondérés (WLS) s'avère plus proche de la vraie valeur du paramètre β que l'estimateur MCO, et conduit à des erreurs quadratiques légèrement inférieures. Cet exemple présente une hétéroscédasticité modérée, avec une variance conditionnelle

$$Var(X_t | X_{t-1}) = \sigma^2 + \sigma_B^2 X_{t-1}^2 = 1 + 0,16X_{t-1}^2.$$

On peut dire aussi que l'estimateur WLS n'améliore que très marginalement les performances prédictives par rapport au MCO. Ceci s'explique par la

faible amplitude des variations de B_t ($\pm 0,4$), qui limite l'effet de la pondération.

3.2.2 Exemple 2

Dans ce second exemple, on fixe $\beta = 0,32$ et l'on suppose que la variable aléatoire B_t prend les valeurs $-1,12$ et $0,58$ avec probabilités respectives $0,3$ et $0,7$. Ainsi, le coefficient autorégressif $a_t = \beta + B_t$ peut valoir

$$\mathbb{P}(a_t = -0,8) = 0,3, \quad \mathbb{P}(a_t = 0,9) = 0,7$$

Le processus des coefficients (a_t) alterne donc entre un régime négatif $(-0,80)$ et un régime fortement positif $(0,90)$. On a

$$\mathbb{E}[B_t^2] = 0,3 \times (-1,12)^2 + 0,7 \times (0,58)^2 = 0,6566.$$

La condition de stationnarité est satisfaite puisque

$$\mathbb{E}[a_t^2] = \beta^2 + \mathbb{E}[B_t^2] = 0,759 < 1.$$

La Figure 3.3 présente le diagnostic de la série simulée issue d'un processus RCA(1) à deux régimes.

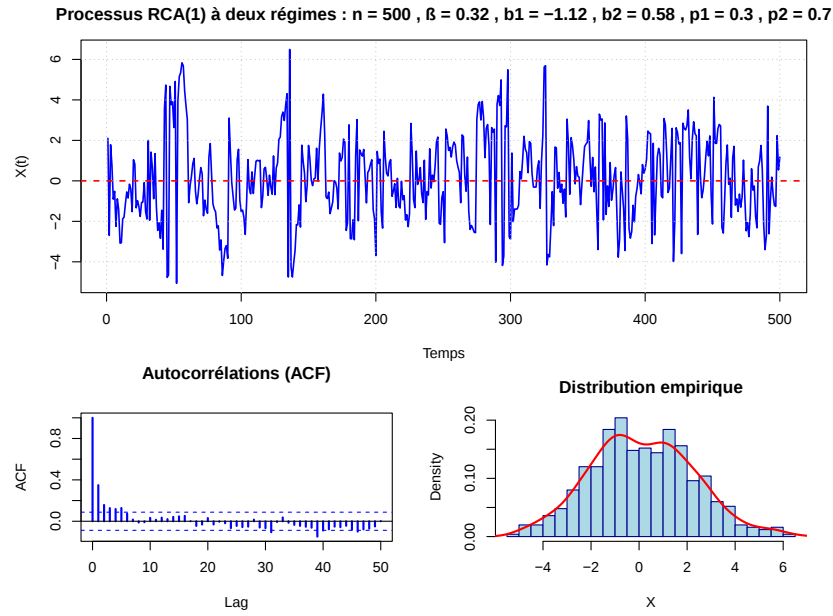


FIGURE 3.3 – Processus RCA(1) à double régimes

La Figure 3.4 compare les prédictions à un pas issues de l'estimateur des moindres carrés ordinaires (MCO) et de l'estimateur pondéré optimal ($\phi(x) = x/(\sigma^2 + \sigma_B^2 x^2)$) sur l'échantillon test.

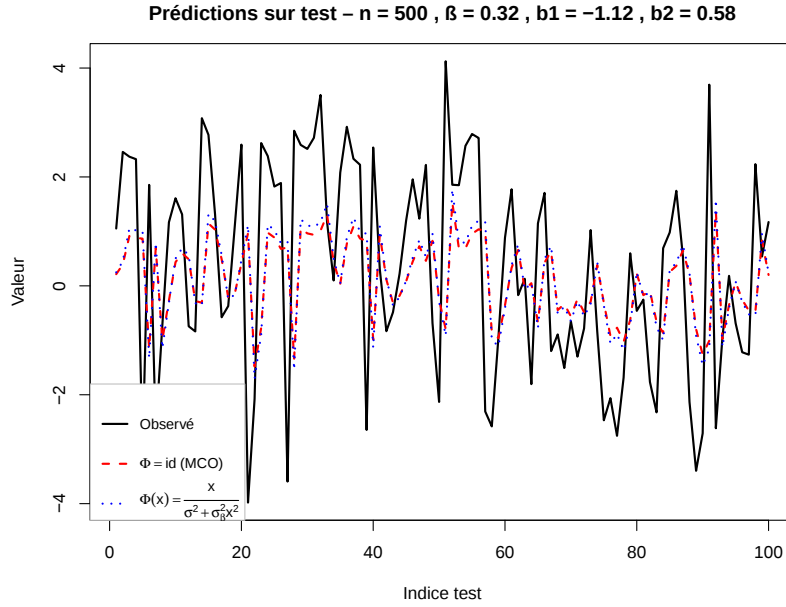


FIGURE 3.4 – Prédiction sur test

Le tableau suivant présente une comparaison des estimateurs et des erreurs de prédiction à un pas sur l'échantillon test pour les deux méthodes d'estimation.

1.955674 1.398454 1.085165

TABLE 3.2 – Comparaison des erreurs de prédiction (processus Bernoulli, exemple 2)

Méthode	$\hat{\beta}_n$	MSE	RMSE	MAE
MCO ($\Phi=\text{id}$)	0.26792	1.997028	1.413162	1.082957
$\Phi(x) = \frac{x}{\sigma^2 + \sigma_\beta^2 x^2}$	0.32559	1.955674	1.398454	1.085165

Le tableau 3.2 montre que l'estimateur MCO s'éloigne sensiblement de la vraie valeur, tandis que l'estimateur pondéré est plus proche. L'estimateur pondéré améliorerait très légèrement les erreurs quadratiques (MSE, RMSE, MAE). Ceci peut être justifié par le fait que la variance conditionnelle dans l'exemple 2

$$\sigma^2 + \sigma_B^2 X_{t-1}^2 = 1 + 0,6118 X_{t-1}^2,$$

soit une hétéroscédasticité nettement plus marquée que dans l'exemple 1.

La Figure 3.4 présente la série simulée du processus RCA(1) à deux régimes

avec les paramètres $n = 500$, $\beta = 0,32$, $b_1 = -1,12$, $b_2 = 0,58$, $p_1 = 0,3$, $p_2 = 0,7$. Le tracé de la série (en haut à gauche) montre des fluctuations d'amplitude variable, avec des pics positifs et négatifs marqués, traduisant l'alternance des deux régimes ($a_t = -0,80$ et $a_t = 0,90$). L'autocorrélogramme (ACF) présente une décroissance lente, caractéristique d'un processus autorégressif d'ordre 1. L'histogramme de la distribution empirique est unimodal et légèrement asymétrique, reflétant la nature hétéroscédastique du modèle. La Figure 3.4 compare les prédictions à un pas issues de l'estimateur MCO (rouge pointillé) et de l'estimateur pondéré optimal $\phi(x) = x/(\sigma^2 + \sigma_B^2 x^2)$ (bleu tirets) sur l'échantillon test. Les deux séries de prédictions sont très proches, avec une légère sous-estimation des pics par l'estimateur MCO. Les métriques du tableau 3.2 confirment cette proximité : la MSE du MCO (5,218) est très voisine de celle du pondéré (5,213). L'avantage de l'estimateur pondéré reste modeste, mais cohérent avec la présence d'hétéroscédasticité. La difficulté à distinguer visuellement les courbes s'explique par la faible proportion d'observations où X_{t-1} est suffisamment grand pour que la pondération ait un effet notable.

3.2.3 Exemple 3

Dans cet exemple, on suppose que le coefficient aléatoire B_t suit une loi normale centrée d'écart-type 0,6 : $B_t \sim \mathcal{N}(0, 0,6^2)$. On fixe $\beta = 0,5$ et $\sigma_\varepsilon = 1$. Ainsi, le coefficient autorégressif $a_t = \beta + B_t$ est gaussien avec $\mathbb{E}[a_t] = 0,5$ et $\text{Var}(a_t) = 0,36$. La condition de stationnarité au second ordre exige

$$\mathbb{E}[a_t^2] = 0,5^2 + 0,36 = 0,61 < 1,$$

ce qui est vérifié. La variance conditionnelle de X_t sachant X_{t-1} s'écrit

$$\text{Var}(X_t \mid X_{t-1}) = \sigma_\varepsilon^2 + \text{Var}(a_t)X_{t-1}^2 = 1 + 0,36X_{t-1}^2.$$

La Figure 3.5 présente le diagnostic de la série simulée issue de ce processus RCA(1) à coefficient gaussien. La Figure 3.5 donne une vue d'ensemble graphique de la série simulée pour un processus RCA(1) avec coefficient gaussien.

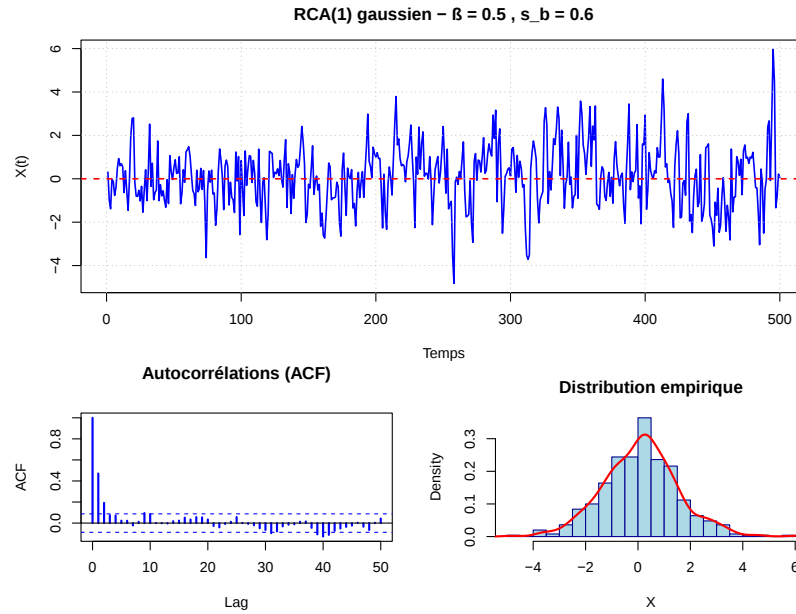


FIGURE 3.5 – Processus RCA(1) à coefficient gaussien

La figure permet d’observer une série temporelle dont l’amplitude varie au cours du temps, avec des creux et des pics prononcés, traduisant une variance conditionnelle non constante. L’autocorrélogramme (ACF) affiche une décroissance progressive, caractéristique d’une dépendance linéaire marquée. Enfin, l’histogramme des données simulées est unimodal et présente une très légère asymétrie, soulignant le caractère aléatoire du coefficient autorégressif.

L’analyse des prévisions à un pas est illustrée dans la Figure 3.6, où l’on compare les séries prédites par le MCO et par l’estimateur pondéré sur l’échantillon test.

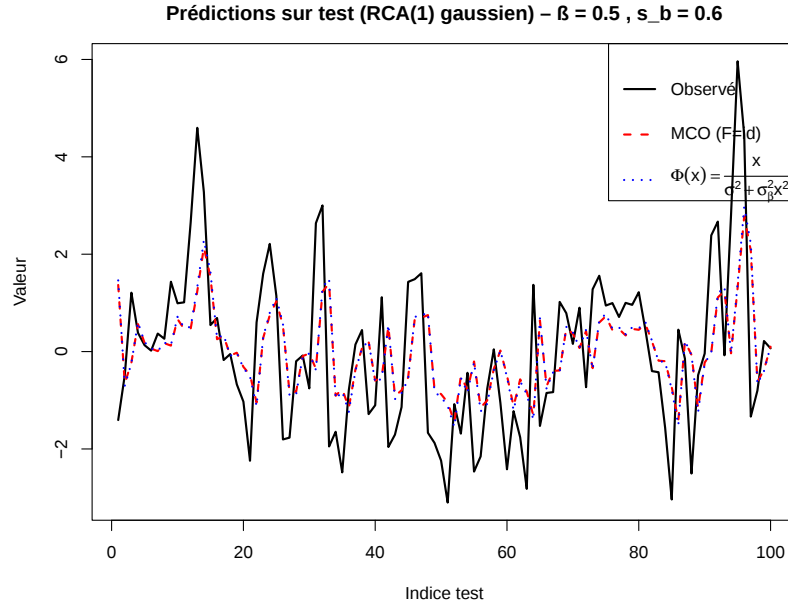


FIGURE 3.6 – Prédiction sur test

Le tableau 3.3 compare les performances des deux estimateurs (MCO et pondéré) sur les données simulées d'un processus RCA(1) à coefficient gaussien.

TABLE 3.3 – Comparaison des erreurs de prédiction (processus RCA(1) gaussien)

Méthode	$\hat{\beta}_n$	MSE	RMSE	MAE
MCO ($\Phi=id$)	0.46604	2.185452	1.478327	1.137018
$\Phi(x) = \frac{x}{\sigma^2 + \sigma_\beta^2 x^2}$	0.49824	2.183504	1.477668	1.131752

Les courbes de prédiction (Figure 3.6) se superposent en grande partie, mais l'estimateur pondéré (bleu tirets) réagit légèrement mieux aux fortes variations de la série test. Le tableau 3.3 confirme cet avantage : l'estimateur pondéré donne une estimation $\hat{\phi} = 0,49824$ plus proche de la vraie valeur $\beta = 0,5$ que le MCO (0,46604). Les erreurs quadratiques (MSE, RMSE) et absolues (MAE) sont très légèrement inférieures pour l'estimateur pondéré

3.3 Prédiction des séries chronologiques de lynx à l'aide d'un modèle RCA(1)

Les données annuelles du nombre de peaux de lynx récoltées dans la région de la rivière Mackenzie (Canada) par la Compagnie de la Baie d'Hudson entre 1821 et 1934 constituent une série temporelle de référence en écologie comme en statistique. Connues sous le nom de *Canadian Lynx data*, elles sont directement accessibles dans R via la commande `data(lynx)` et servent fréquemment pour l'étude des oscillations proie-prédateur. Ces données sont réputées pour leur dynamique cyclique (période d'environ 10 ans) et leur non-linéarité potentielle.

Dans ce mémoire, nous utilisons ces données pour prévoir les fluctuations des populations de lynx à l'aide d'un modèle autorégressif à coefficients aléatoires d'ordre 1, noté RCA(1). Contrairement à un modèle AR(1) standard, le RCA(1) intègre une composante stochastique dans son coefficient autorégressif, ce qui lui permet de saisir à la fois des variations temporelles de la structure de dépendance et d'éventuels effets non linéaires. Nous estimons le modèle RCA(1) sur les données historiques, évaluons ses capacités prédictives et comparons ses prévisions à celles obtenues avec des modèles autorégressifs classiques.

La figure ci-dessous montre l'évolution du nombre annuel de peaux de lynx recensées par la Compagnie de la Baie d'Hudson entre 1821 et 1934. :

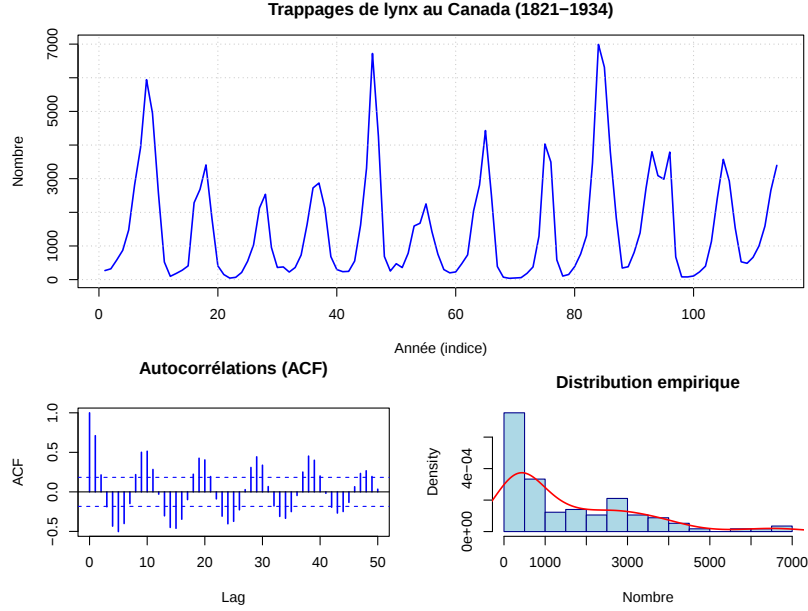


FIGURE 3.7 – Diagnostic de la série lynx : évolution, autocorrélations et distribution

On observe une dynamique cyclique caractéristique, avec des pics rapprochés tous les 8 à 11 ans – un phénomène typique des interactions proie-prédateur (lynx – lièvre d’Amérique).

L’autocorrélogramme (ACF) de la série lynx présente une décroissance lente et régulière, caractéristique d’une forte dépendance linéaire. Il montre que les valeurs d’une année dépendent fortement de celles des années précédentes, en particulier des premières années de retard. Cela confirme le caractère cyclique de la série : une année avec beaucoup de lynx a tendance à être suivie d’une autre année encore riche, ce qui traduit une certaine persistance.

L’histogramme de la distribution empirique montre une forme unimodal mais asymétrique, avec une queue à droite allongée traduisant la présence d’années exceptionnellement fortes.

Pour estimer le paramètre $\tau = \sigma_B^2 / \sigma_\varepsilon^2$, nous utilisons la méthode du maximum de vraisemblance (voir section 1.4.2). La minimisation de la fonction

$$l_n(\beta, \tau) = \ln \left(\frac{1}{n} \sum_{t=1}^n \frac{(X_t - \beta X_{t-1})^2}{1 + \tau X_{t-1}^2} \right) + \frac{1}{n} \sum_{t=1}^n \ln(1 + \tau X_{t-1}^2),$$

par rapport à τ (sur un intervalle approprié) fournit l'estimateur $\hat{\tau}_n$.

Dans notre étude numérique, cette optimisation est effectuée à l'aide de la fonction `optimize()` de R, qui implémente une recherche unidimensionnelle robuste. L'estimateur $\hat{\tau}_n$ est ensuite utilisé pour construire la fonction de pondération optimale

$$\phi_*(x) = \frac{x}{\sigma_\varepsilon^2 + \tau x^2}.$$

Cette approche permet d'obtenir un estimateur des moindres carrés pondérés (WLS) asymptotiquement équivalent à l'estimateur du maximum de vraisemblance, tout en étant plus simple à calculer.

Le graphique suivant compare les valeurs observées aux prédictions obtenues par les deux modèles l'estimateur des moindres carrés ordinaires (MCO) et de l'estimateur pondéré optimal ($\phi(x) = x/(\sigma^2 + \sigma_B^2 x^2)$) sur l'échantillon test :

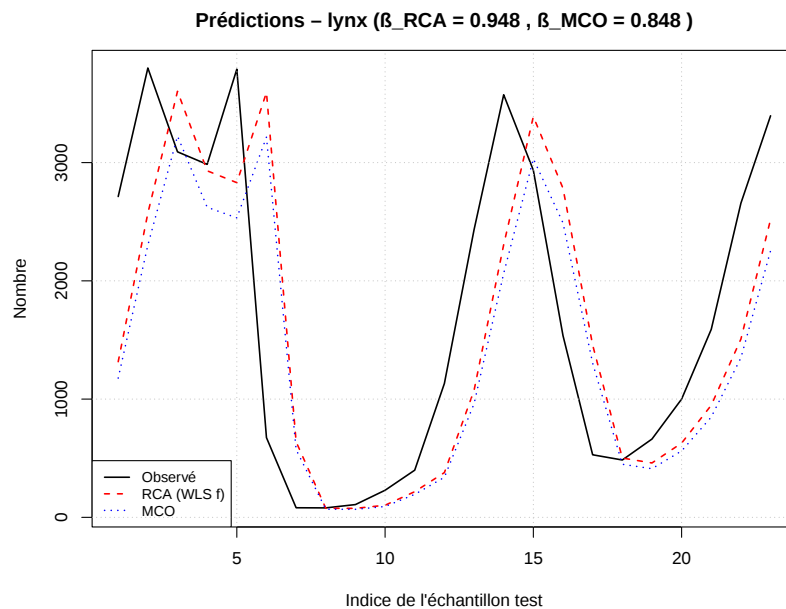


FIGURE 3.8 – Prévisions des captures de lynx sur l'échantillon test par les méthodes MCO et WLSS

Le tableau ci-dessous résume les performances prédictives des deux modèles à l'aide des trois critères d'erreur classiques :

TABLE 3.4 – Comparaison des erreurs de prédiction – Série lynx (trappages annuels)

Méthode	$\hat{\beta}_n$	MSE	RMSE	MAE
MCO (AR(1) sans constante)	0.84845	1022223.993889	1011.050935	770.860965
RCA(1) (WLS $\varphi(x) = x/(\sigma^2 + \sigma_B^2 x^2)$)	0.94820	993632.434469	996.811133	749.531010

Le graphique 3.8 montre que le modèle RCA(1) estimé par WLS ($\phi = 0,948$) suit plus fidèlement les pics d'abondance de lynx que le modèle AR(1) par MCO ($\phi = 0,848$), en particulier pour les valeurs extrêmes. Cette meilleure adéquation visuelle est confirmée par le tableau des erreurs 3.4 : le RCA(1) réduit le MSE d'environ 2,8 % par rapport au MCO, avec un RMSE et un MAE également plus faibles. Globalement, le modèle RCA(1) s'avère plus performant pour prévoir les fluctuations cycliques de cette série.

Conclusion

Ce mémoire a été consacré à l'étude de l'estimation et de la prévision dans le cadre des modèles autorégressifs à coefficients aléatoires d'ordre 1, notés RCA(1). Trois approches d'estimation ont été présentées et comparées : la méthode des moindres carrés ordinaires (MCO), la méthode du maximum de vraisemblance (MV), ainsi que la méthode des moindres carrés pondérés (WLS), avec un intérêt particulier pour la classe d'estimateurs introduite par Schick [14].

L'objectif principal de ce travail était de développer les résultats établis par Vaněček [15] concernant la vitesse de convergence en loi de l'estimateur de Schick vers la loi normale. L'auteur a montré que cette vitesse est de l'ordre de

$$\frac{(\log n)^3}{\sqrt{n}},$$

ce qui constitue un résultat important pour l'inférence asymptotique dans les modèles RCA(1). Ce résultat améliore les vitesses de convergence classiques obtenues dans le cadre des modèles autorégressifs linéaires.

La performance des estimateurs et des prédicteurs associés a été illustrée à travers quatre exemples : trois reposant sur des données simulées (coefficients aléatoires suivant une loi de Bernoulli généralisée — modèle à deux régimes — et suivant une loi normale) et un dernier basé sur des données réelles, à savoir la célèbre série des captures de lynx de la région de la rivière Mackenzie.

Les résultats numériques obtenus montrent que l'estimateur des moindres carrés pondérés (WLS) conduit à des erreurs de prévision (MSE, RMSE, MAE) légèrement inférieures à celles de l'estimateur des moindres carrés ordinaires, en particulier lorsque l'hétéroscédasticité est plus marquée. Cette observation est cohérente avec la théorie, puisque l'estimateur WLS tire parti de la structure de la variance conditionnelle pour améliorer l'efficacité des estimations. Toutefois, dans les exemples considérés, l'amélioration reste mo-

deste, ce qui s'explique par la faible proportion d'observations où la pondération a un effet notable.

En perspective, plusieurs axes de recherche peuvent être envisagés. Il serait intéressant d'étendre ces résultats à des modèles RCA d'ordre supérieur ($\text{RCA}(p)$) ou à des modèles multivariés. Une autre piste consisterait à examiner le comportement des estimateurs sous des hypothèses de dépendance plus faibles, ou encore à étudier des procédures de sélection de modèle adaptées à ce type de processus. Enfin, l'application à d'autres séries économiques ou financières pourrait constituer un prolongement naturel de ce travail.

Bibliographie

- [1] Anděl, J. (1976). *Autoregressive series with random parameters*. Mathematische Operationsforschung und Statistik, **7**(5), 735–741.
- [2] Aue, A., Horváth, L. and Steinebach, J. (2006). *Estimation in random coefficient autoregressive models*. Journal of Time Series Analysis, **27**(1), 61–76.
- [3] Aue, A. and Horváth, L. (2011). Quasi-likelihood estimation in stationary and nonstationary autoregressive models with random coefficients. *Statistica Sinica*, **21**, 973–999.
- [4] Berkes, I., Horváth, L. and Schick, A. (2009). Quasi-likelihood estimation in random coefficient autoregressive models. *Journal of Statistical Planning and Inference*, **139**(6), 2041–2054.
- [5] Bougerol, P. and Picard, N. (1992). *Strict stationarity of generalized autoregressive processes*. Annals of Probability, **20**, 1714–1730.
- [6] Brandt, A. (1986). *The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients*. Advances in Applied Probability, **18**, 211–220.
- [7] Davidson, J. (2002). *Stochastic Limit Theory*. Advanced Texts in Econometrics. Oxford University Press, New York. Reprinted edition.
- [8] Hwang, S. Y. and Basawa, I. V. (1993). *Asymptotic optimal inference for a class of nonlinear time series models*. Stochastic Processes and their Applications, **46**, 91–113.
- [9] Ibragimov, I. A. (1962). Some limit theorems for stationary processes. *Theory of Probability and Its Applications*, **7**, 349–382.
- [10] Koul, H. L. and Schick, A. (1996). Adaptive estimation in a random coefficient autoregressive model. *The Annals of Statistics*, **24**(3), 1025–1052.

- [11] Nicholls, D. F. and Quinn, B. G. (1981). *The estimation of random coefficient autoregressive models. II.* Journal of Time Series Analysis, **2**, 185–203.
- [12] Nicholls, D. F. and Quinn, B. G. (1982). *Random Coefficient Autoregressive Models : An Introduction.* Springer-Verlag, New York.
- [13] Nicholls, D. F. and Quinn, B. G. (2012). *Random Coefficient Autoregressive Models : An Introduction.* Volume 11. Springer Science & Business Media.
- [14] Schick, A. (1996). $\sqrt{n}$ -consistent estimation in a random coefficient autoregressive model. Australian Journal of Statistics, **38**(2), 155–160.
- [15] Vaněček, P. (2006). Rate of convergence for a class of RCA estimators. *Kybernetika*, **42**(6), 699–709.
- [16] Wang, D. (2003). *Frequentist and Bayesian analysis of random coefficient autoregressive models.* Thèse de doctorat, University of California, Berkeley.