

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE
SCIENTIFIQUE

UNIVERSITÉ MOULOUD MAMMARI, TIZI-OUZOU

FACULTÉ DES SCIENCES

DÉPARTEMENT DE MATHÉMATIQUES



Mémoire de Master

Filière : Mathématiques

Spécialité : Probabilités et Statistique

Présenté par :

Amal Kernouf

Sujet :

**SUR L'ESTIMATION DES PARAMÈTRES D'UN
MODÈLE AUTORÉGRESSIF À COEFFICIENTS
ALÉATOIRES AVEC TENDANCE**

Devant le jury d'examen composé de :

Mme. ABDOUCHE Safia	MCB	UMMTO	Présidente
Mme. AIT MOHAMMED Noura	MCB	UMMTO	Promotrice
Mme. MERABET Dalila	MCB	UMMTO	Examinatrice

Année universitaire 2024–2025

Table des matières

Table des matières	4
Introduction générale	7
1 Généralités sur les séries chronologiques	10
1.1 Introduction	10
1.2 Processus stochastiques	10
1.2.1 Une série chronologique	11
1.3 Stationnarité	12
1.3.1 Stationnarité forte	12
1.3.2 Stationnarité faible	13
1.4 Ergodicité	13
1.5 Fonction d'autocovariance	14
1.6 fonction d'autocorrélation	14
1.7 Processus Bruit Blanc (White Noise Process)	14
1.8 Différence de martingale	15
1.8.1 Une différence de martingale homoscédastique	15
1.8.2 Une différence de martingale hétéroscédastique	16
1.9 Opérateurs sur les séries temporelles	16
1.9.1 L'opérateur retard	16
1.9.2 Opérateurs de différenciation	16
1.9.3 Théorème de Wold	17
1.10 Quelques modèles linéaires de séries temporelles	18
1.10.1 Processus autorégressifs AR(p)	18
1.10.2 Processus moyenne mobile MA(q)	18
1.11 Causalité et Inversibilité	19
1.11.1 Causalité	19
1.11.2 Inversibilité	19
1.11.3 Processus autorégressif moyenne mobile (ARMA)	20
2 Étude des modèles autorégressifs à coefficients aléatoires	22
2.1 Introduction	22

2.2	Modèle autorégressif à coefficients aléatoires d'ordre p	23
2.3	Modèle autorégressif à coefficient aléatoire d'ordre un	23
2.4	Stationnarité des modèles à coefficients aléatoires	24
2.4.1	Solution strictement stationnaire	24
2.4.2	Solution faiblement stationnaire	24
2.5	Estimation des paramètres d'un modèle RCA(1)	30
2.5.1	Estimation des paramètres d'un RCA(1) par les moindres carrés . .	30
2.5.2	Consistance et normalité asymptotique de l'estimateur par la méthode des moindres carrés	35
2.5.3	Estimation des paramètres d'un RCA(1) par le maximum de vraisemblance	35
2.6	Estimation des paramètres d'un modèle RCA(1)	42
2.6.1	Estimation des paramètres d'un RCA(1) par les moindres carrés . .	43
2.6.2	Consistance et normalité asymptotique de l'estimateur par la méthode des moindres carrés	47
2.6.3	Estimation des paramètres d'un RCA(1) par le maximum de vraisemblance	47
2.6.4	Consistance et normalité asymptotique de l'estimateur du maximum de vraisemblance	50
3	Modèle de régression linéaire à erreurs RCA(1)	52
3.1	Introduction	52
3.2	Rappel sur les modèles de régression non linéaire	52
3.2.1	Modèle de régression non linéaire	54
3.2.2	Moindres carrés non linéaires	56
3.2.3	Méthodes de calcul de l'estimateur des moindres carrés	57
	Méthode de Gauss-Newton	57
	La méthode de Newton-Raphson	60
	La méthode des scores	61
3.3	Estimation par la méthode des moindres carrés ordinaires dans un modèle de régression linéaire à erreurs RCA(1)	62
3.3.1	Présentation du modèle	62
3.3.2	La procédure d'estimation (Méthode de Gauss-Newton)	63
3.4	Estimation par maximum de vraisemblance	67
4	Application numérique	69
4.1	Étude de simulation	69
4.1.1	Présentation du modèle	69

4.1.2	Application de la méthode des moindres carrés conditionnels pour estimer les paramètres du modèle de régression avec erreurs RCA(1)	71
4.1.3	Application de la méthode du maximum de vraisemblance pour estimer les paramètres du modèle de régression avec erreurs RCA(1)	77
	Conclusion générale	80

Dédicace

À celle dont le cœur est un havre de paix,

À ma mère, lumière de mes jours et force silencieuse de mes nuits.

Ton amour inconditionnel, ta patience infinie et ton regard toujours rassurant ont été les fondations de ma détermination. Ce mémoire t'est dédié, avec tendresse et fierté.

À mon père, pilier de sagesse et de sérénité.

Ta foi en mes capacités, ta discrétion bienveillante et ta droiture m'ont appris l'endurance et l'humilité. Ta présence, même silencieuse, est un moteur invisible dans chacun de mes pas.

À mes frères, Zina et Chabane, compagnons fidèles de cœur et d'esprit.

Votre affection fraternelle, vos encouragements discrets et votre présence rassurante ont allégé mes doutes et porté mes élans.

À toutes les âmes lumineuses qui, d'un mot, d'un geste, ou d'un regard, m'ont insufflé force et espérance,

je vous offre ces pages, empreintes de gratitude et d'hommage.

Amal Kernouf

Remerciements

JE tiens à exprimer, avec une émotion sincère, toute ma gratitude envers DIEU, source de sagesse et de force, qui m'a accompagnée tout au long de ce chemin semé d'efforts et de persévérance.

Ce mémoire n'est pas seulement le fruit d'un travail personnel, mais celui d'un parcours enrichi par le soutien, les conseils et la bienveillance de nombreuses personnes que je souhaite ici remercier chaleureusement.

À *mes parents*, je dédie ce travail avec toute ma reconnaissance. Leur soutien inconditionnel, leur patience silencieuse et leur foi en moi ont été les piliers essentiels de cette réussite. Leur amour est ma plus grande richesse.

Je souhaite témoigner toute ma reconnaissance à ma promotrice, **Madame AIT MOHAMMED Noura**, pour son encadrement exigeant, sa rigueur intellectuelle, et sa présence attentive. Grâce à ses orientations précieuses, j'ai pu progresser avec clarté et assurance.

Je tiens à adresser mes sincères remerciements aux membres de jury, Mme **ABDOUCHE Safia** et Mme **MERABET Dalila**, qui ont accepté d'évaluer ce travail.

Mes sincères remerciements vont également à **Monsieur Fellag**, ainsi qu'à **Mesdames, Atil, Belkacem et Boualem**, dont l'investissement pédagogique, la rigueur et le dévouement ont profondément marqué mon parcours universitaire.

Je remercie de tout cœur mes chères amies et cousines, complices de tous les instants, pour leur tendresse, leur humour, leur écoute, et pour avoir toujours cru en moi, même dans les périodes les plus incertaines.

À tous ceux qui, de près ou de loin, ont contribué à la réussite de ce mémoire, je tiens à dire : MERCI.

Introduction générale

L'étude des séries chronologiques constitue l'un des domaines les plus riches et les plus dynamiques de la statistique appliquée, avec des applications multiples en économétrie, en finance, en climatologie, en biologie, et dans bien d'autres disciplines. Depuis les travaux fondateurs de Yule [29] sur les séries autorégressives et ceux de Slutsky [26] sur la théorie des moyennes mobiles, l'analyse des processus stochastiques temporels n'a cessé d'évoluer, donnant naissance à une panoplie de modèles de plus en plus flexibles et adaptés à la complexité des données observées dans la réalité.

Parmi ces modèles, les modèles classiques tels que le modèle autorégressif d'ordre p ($AR(p)$), le modèle à moyennes mobiles d'ordre q ($MA(q)$) et le modèle autorégressif à moyennes mobiles d'ordre (p, q) ($ARMA(p, q)$), largement développés par Box et Jenkins [7], ont permis de capturer efficacement la dynamique linéaire des séries stationnaires. Dans le cadre bayésien, Ait Mohammed et Guerbyenne [1] ont étudié les modèles de régression à erreurs $AR(p)$ sous contamination. Cependant, ces approches traditionnelles reposent sur l'hypothèse restrictive de constance des paramètres, ce qui peut limiter leur capacité d'adaptation à des phénomènes réels caractérisés par des ruptures de tendance, une volatilité variable ou une structure sous-jacente instable. Face à ces limites, de nouveaux modèles plus souples ont été proposés dans la littérature, à l'instar des modèles à coefficients variables, des modèles de changement de régime (Markov switching) et en particulier, des modèles autorégressifs à coefficients aléatoires (RCA).

Introduits par Nicholls et Quinn [24], les modèles RCA représentent une extension naturelle des modèles AR , dans lesquels les coefficients ne sont plus constants mais suivent

eux-mêmes une dynamique aléatoire. Cette approche permet d'introduire une hétérogénéité temporelle dans la relation autorégressive, rendant ainsi les modèles RCA particulièrement adaptés pour analyser des phénomènes soumis à des chocs exogènes ou à des évolutions structurelles lentes. Ces modèles s'inscrivent également dans une vision plus large de la modélisation stochastique dynamique développée dans les travaux contemporains de Dahlhaus (1997), Stoffer et al. (1990), ou encore Douc, Moulines et Stoffer [14].

Ce mémoire s'intéresse spécifiquement aux modèles de régression linéaire à erreurs autocorrélées suivant un RCA(1). L'objectif principal est d'étudier et de comparer deux méthodes d'estimation des paramètres : la méthode des moindres carrés conditionnels (MCC), simple à implémenter mais parfois inefficace en présence de bruit non gaussien, et la méthode du maximum de vraisemblance (MV), réputée pour sa consistance et son efficacité asymptotique (White [28]; Gouriéroux & Monfort [17]). Cette étude méthodologique s'appuie notamment sur des techniques numériques telles que l'algorithme de Gauss-Newton, fréquemment utilisé pour l'approximation des solutions dans les modèles non linéaires.

La première partie du mémoire est consacrée aux fondements théoriques des séries temporelles : processus stationnaires, fonctions d'autocorrélation, théorème de Wold, causalité et inversibilité, inspirée des références classiques de Brockwell et Davis [9] et Hamilton [18]. Dans une seconde partie, nous présentons en détail la structure et les propriétés des modèles RCA(1), en mettant en évidence leurs conditions de stationnarité et les défis liés à l'estimation des paramètres. La troisième partie propose une étude rigoureuse de l'estimation des paramètres dans un modèle de régression linéaire perturbé par des erreurs RCA(1), via les approches MCC et MV. Enfin, une analyse numérique est menée à travers une étude de simulation.

Les résultats obtenus permettent de comparer la qualité des deux méthodes d'estimation à l'aide de la racine de l'erreur quadratique moyenne (RMSE). Ce travail vise ainsi à contribuer à une meilleure compréhension des modèles à coefficients aléatoires et de leurs propriétés statistiques, tout en mettant en évidence l'importance d'adapter les techniques d'estimation aux structures dynamiques complexes que l'on observe dans les

séries temporelles réelles.

Chapitre 1

Généralités sur les séries chronologiques

1.1 Introduction

Ce premier chapitre est consacré à l'introduction des concepts fondamentaux des séries temporelles, en mettant l'accent sur les propriétés essentielles des processus stochastiques. Nous commençons par la définition formelle d'un processus stochastique, suivie de l'étude des conditions de stationnarité, en distinguant la stationnarité au sens faible et au sens fort. L'analyse des fonctions d'autocovariance et d'autocorrélation sera développée afin de quantifier la dépendance temporelle inhérente aux séries temporelles.

Nous examinerons également la notion de bruit blanc, ainsi que les techniques de différenciation utilisées pour stationnariser une série. Le théorème de Wold sera présenté comme un résultat fondamental garantissant la décomposition des processus stationnaires en une somme de composantes déterministes et aléatoires. Par ailleurs, nous introduirons les principaux modèles de séries temporelles, notamment les modèles AR, MA et ARMA.

1.2 Processus stochastiques

La modélisation d'une série d'observations repose sur la recherche d'un modèle qui approxime au mieux le processus stochastique. Cette approche nécessite de recourir à la théorie des processus stochastiques, ainsi qu'à la compréhension approfondie de leurs propriétés fondamentales. Ces propriétés permettent de décrire de manière précise l'évolution temporelle des phénomènes aléatoires et d'établir les liens entre les différentes réalisations

de ces processus.

Définition 1.2.1. *Un processus stochastique est une famille de variables aléatoires $\{X_t\}_{t \in T}$ définies sur un certain espace probabilisé (Ω, \mathcal{F}, P) et indexées par un certain ensemble T , appelé ensemble de temps.*

- Si $T = \mathbb{Z}$, on parle de processus stochastique à temps discret.
- Si $T = \mathbb{R}_+$, on parle de processus stochastique à temps continu.

1.2.1 Une série chronologique

Définition 1.2.2. *Une série chronologique, également appelée série temporelle ou chronique, est un modèle mathématique qui modélise des phénomènes dépendant du temps. Mathématiquement, on dit que c'est un processus aléatoire noté $(X_t)_{t \in \mathbb{Z}}$ avec X_t une variable aléatoire de densité f . On peut dire aussi que c'est un processus stochastique à temps discret.*

Analyse d'une série chronologique :

Les composantes d'une série temporelle sont généralement les suivantes :

1. Tendances générale (trend).
2. Mouvements saisonniers (seasonality).
3. Composante aléatoire.
4. Éventuellement, on peut ajouter les composantes résiduelles liées à des conjonctures (grèves, confinements, etc.).

On distingue principalement deux types de modèles :

1. **Le modèle additif** : La variable X_t s'écrit comme la somme de trois termes :

$$X_t = m_t + s_t + \epsilon_t, \quad \text{avec} \quad \mathbb{E}(\epsilon_t) = 0.$$

où :

- X_t : série observée,
- m_t : tendance (*trend*),
- s_t : effet saisonnier,
- ϵ_t : composante aléatoire.

2. **Le modèle multiplicatif** : La variable X_t s'écrit, à une erreur près, comme le produit de la tendance et d'une composante de saisonnalité :

$$X_t = m_t \cdot s_t \cdot \epsilon_t, \quad \text{avec} \quad \mathbb{E}(\epsilon_t) = 1.$$

1.3 Stationnarité

Nous commençons par la définition d'un processus **stationnaire au sens strict** (ou *fortement stationnaire*), avant de présenter la notion de **stationnarité au second ordre** (ou *faiblement stationnaire*).

1.3.1 Stationnarité forte

Définition 1.3.1 (Stationnarité au sens strict). *Un processus $\{X_t; t \in \mathbb{Z}\}$ est dit **stationnaire au sens strict** (ou *fortement stationnaire*) si ses lois fini-dimensionnelles sont invariantes par translation temporelle.*

Autrement dit, pour tout $h \in \mathbb{Z}$, pour tout entier $n \in \mathbb{N}^$ et pour tout $(t_1, t_2, \dots, t_n) \in \mathbb{Z}^n$, le vecteur aléatoire*

$$(X_{t_1}, X_{t_2}, \dots, X_{t_n})$$

a la même loi que le vecteur translaté

$$(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_n+h}).$$

*Cette notion de stationnarité forte est difficile à vérifier en pratique. On lui préfère donc la notion de **stationnarité faible**, qui repose sur les moments d'ordre un et deux.*

1.3.2 Stationnarité faible

Définition 1.3.2 (Stationnarité au second ordre). *Un processus $\{X_t; t \in \mathbb{Z}\}$ est dit **stationnaire au second ordre** (ou faiblement stationnaire) s'il vérifie les conditions suivantes :*

1. $\forall t \in \mathbb{Z}, \quad \mathbb{E}[X_t^2] < \infty$ (moment d'ordre deux fini),
2. $\forall t \in \mathbb{Z}, \quad \mathbb{E}[X_t] = m < \infty$ (espérance constante),
3. $\forall (t, h) \in \mathbb{Z}^2, \quad \text{Cov}(X_t, X_{t+h}) = \gamma(h)$ (autocovariance dépendant uniquement du décalage h).

En pratique, on vérifie généralement la constance de l'espérance $\mathbb{E}[X_t] = m$ et de la variance $\text{Var}(X_t) = \sigma^2$ pour tout $t \in \mathbb{Z}$. Ces hypothèses permettent une analyse plus simple des séries temporelles et l'application efficace de nombreux modèles linéaires classiques.

Remarque. *Pour un processus du second ordre, la stationnarité stricte implique la stationnarité faible. Toutefois, la réciproque n'est pas toujours vraie, sauf dans le cas des processus gaussiens.*

1.4 Ergodicité

*On dit qu'une suite stationnaire est **ergodique** si elle satisfait la loi forte des grands nombres.*

Définition 1.4.1 (Processus stationnaire ergodique). *Un processus strictement stationnaire $\{X_t, t \in \mathbb{Z}\}$, à valeurs réelles, est dit **ergodique** si et seulement si, pour tout borélien $B \subset \mathbb{R}^{k+1}$ et tout entier $k \in \mathbb{N}$, on a :*

$$\frac{1}{n} \sum_{t=1}^n \mathbf{1}_B(Y_t, Y_{t+1}, \dots, Y_{t+k}) \longrightarrow \mathbb{P}[(Y_t, Y_{t+1}, \dots, Y_{t+k}) \in B] \quad \text{presque sûrement.}$$

Certaines transformations de suites ergodiques restent ergodiques.

1.5 Fonction d'autocovariance

Définition 1.5.1. Soit $\{X_t, t \in \mathbb{Z}\}$ un processus aléatoire, la **fonction d'autocovariance** mesure la covariance entre deux valeurs du processus séparées d'un intervalle de longueur h , notée $\gamma(h)$:

$$\gamma(h) = \text{Cov}(X_t, X_{t+h}) = E[(X_t - E(X_t))(X_{t+h} - E(X_{t+h}))], \quad \forall t, h \in \mathbb{Z}.$$

1.6 fonction d'autocorrélation

Définition 1.6.1. On appelle **fonction d'autocorrélation**, notée $\rho(h)$, d'un processus $\{X_t; t \in \mathbb{Z}\}$ la fonction définie par :

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}, \quad \forall h \in \mathbb{Z},$$

où $\gamma(h) = \text{Cov}(X_t, X_{t+h})$ est la fonction d'autocovariance du processus.

1.7 Processus Bruit Blanc (White Noise Process)

Définition 1.7.1. "Bruit blanc faible" Un processus $\{X_t; t \in \mathbb{Z}\}$ est appelé bruit blanc faible, si les deux conditions suivantes sont satisfaites :

1. $E(X_t) = 0, \quad \forall t \in \mathbb{Z}$

Ce qui signifie que l'espérance du processus est indépendante du temps et qu'elle est **nulle**.

2. $\gamma(h) = E[X_t X_{t-h}] = \begin{cases} \sigma^2 & \text{si } h = 0, \\ 0 & \text{si } h \neq 0. \end{cases}$

Cela implique bien entendu l'indépendance de la fonction d'autocovariance par rapport au temps. En d'autres termes, cela signifie que les bruits blancs sont des processus stationnaires particuliers sans **mémoire**.

Définition 1.7.2. "Bruit blanc fort" Un processus $\{X_t; t \in \mathbb{Z}\}$ est appelé un **bruit blanc fort** si et seulement si les X_t sont *i.i.d.*, avec :

$$\mathbb{E}(X_t) = 0 \quad \text{et} \quad \text{Var}(X_t) = \sigma^2.$$

On note alors :

$$X_t \sim \mathcal{BB}(0, \sigma^2)$$

1.8 Différence de martingale

Avant de présenter les définitions suivantes, introduisons une notation essentielle :

Soit \mathcal{F}_t la **filtration**, c'est-à-dire l'information disponible jusqu'au temps t inclus.

Plus précisément :

$$\mathcal{F}_t = \sigma(X_t, X_{t-1}, X_{t-2}, \dots)$$

Elle représente l'historique des réalisations passées du processus $\{X_t\}$ jusqu'à l'instant t .

L'espérance conditionnelle $E(X_t | \mathcal{F}_{t-1})$ correspond alors à la meilleure prévision de X_t , en utilisant toute l'information connue jusqu'à $t - 1$.

1.8.1 Une différence de martingale homoscédastique

Définition 1.8.1. Le processus $\{X_t; t \in \mathbb{Z}\}$ est une **différence de martingale homoscédastique** si et seulement si :

$$E(X_t | \mathcal{F}_{t-1}) = 0 \quad \forall t \in \mathbb{Z}$$

$$\text{Var}(X_t) = \sigma^2 \quad \forall t \in \mathbb{Z}$$

(L'espérance conditionnelle est nulle et la variance est constante.)

1.8.2 Une différence de martingale hétéroscédastique

Définition 1.8.2. *Le processus $\{X_t; t \in \mathbb{Z}\}$ est une différence de martingale hétéroscédastique si et seulement si :*

$$\begin{aligned} E(X_t | \mathcal{F}_{t-1}) &= 0 \quad \forall t \in \mathbb{Z} \\ \text{Var}(X_t | \mathcal{F}_{t-1}) &= h_t^2 \quad \forall t \in \mathbb{Z} \end{aligned}$$

(L'espérance conditionnelle est nulle et la variance conditionnelle n'est pas constante.)

1.9 Opérateurs sur les séries temporelles

1.9.1 L'opérateur retard

On note indifféremment B (Backward) ou L (lag), l'opérateur retard qui fait passer de X_t à X_{t-1} :

$$BX_t = X_{t-1}.$$

On a également :

$$B^2 X_t = B(BX_t) = BX_{t-1} = X_{t-2}.$$

Par une relation de récurrence, on obtient :

$$B^n X_t = B(B(\dots(BX_t))) = X_{t-n}.$$

1.9.2 Opérateurs de différenciation

Définition 1.9.1. *On définit l'opérateur de différenciation Δ par :*

$$\Delta X_t = X_t - X_{t-1}, \quad \forall t \geq 2.$$

Définition 1.9.2. *On définit l'opérateur linéaire de différenciation d'ordre k par la formule de récurrence :*

$$\Delta^{(k)} X_t = \Delta(\Delta^{(k-1)} X_t).$$

1.9.3 Théorème de Wold

Le théorème de Wold (1938) est le théorème fondamental de l'analyse des séries temporelles stationnaires. Ce résultat constitue, d'une certaine façon, un prérequis aux processus ARMA qui seront abordés à la section suivante.

Définition 1.9.3 (Théorème de Wold). *Tout processus $\{X_t, t \in \mathbb{Z}\}$ faiblement stationnaire peut s'écrire sous la forme :*

$$X_t = \sum_{j=0}^{+\infty} \psi_j \varepsilon_{t-j} + k_t,$$

- ψ_j : paramètres réels du modèle.
- $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc gaussien.
- ε_{t-j} : représentent les "**chocs**" passés (qui ont une variance finie). Cela signifie que la somme des chocs passés correspond à la composante linéaire stochastique de X_t .
- k_t : partie déterministe telle que $\text{Cov}(k_t, \varepsilon_{t-j}) = 0 \quad \forall t \in \mathbb{Z}, \forall j \in \mathbb{N}^*$.

Remarque.

On impose une condition de normalisation :

$$\psi_0 = 1 \quad \text{et} \quad \sum_{j=0}^{\infty} \psi_j^2 < \infty$$

Une autre écriture de ce théorème peut être obtenue comme suit :

$$X_t = \Psi(B)\varepsilon_t + k_t$$

où

$$\Psi(B) = \sum_{j=0}^{\infty} \psi_j B^j = 1 + \psi_1 B + \psi_2 B^2 + \dots$$

et

$$\Psi(B)\varepsilon_t = (1 + \psi_1 B + \psi_2 B^2 + \dots)\varepsilon_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots$$

1.10 Quelques modèles linéaires de séries temporelles

1.10.1 Processus autorégressifs AR(p)

Définition 1.10.1 (Brockwell, P. J. Davis, R. A. (2002) [8]). On dit qu'un processus stationnaire $\{X_t; t \in \mathbb{Z}\}$ admet une représentation autorégressive d'ordre p , notée $AR(p)$, s'il vérifie l'équation suivante :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \cdots + \phi_p X_{t-p} + \varepsilon_t,$$

où

$$\{\varepsilon_t; t \in \mathbb{Z}\} \sim \text{BB}(0, \sigma^2)$$

et ϕ_1, \dots, ϕ_p sont des coefficients réels tels que $\phi_p \neq 0$

On peut aussi écrire :

$$X_t = \sum_{k=1}^p \phi_k X_{t-k} + \varepsilon_t \quad \Leftrightarrow \quad \Phi(B)X_t = \varepsilon_t,$$

avec $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ le polynôme caractéristique de degré p .

Propriété 1. Un processus autorégressif $AR(p)$ est **stationnaire** si et seulement si toutes les racines du polynôme $\Phi(z)$ sont de modules strictement supérieurs à 1.

1.10.2 Processus moyenne mobile MA(q)

Définition 1.10.2 (Brockwell, P. J. Davis, R. A. (2002) [8]). Un processus stationnaire $\{X_t; t \in \mathbb{Z}\}$ admet une représentation moyenne mobile d'ordre q , notée $MA(q)$, s'il vérifie l'équation suivante :

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q},$$

où

$$\{\varepsilon_t; t \in \mathbb{Z}\} \sim \text{BB}(0, \sigma^2)$$

et $\theta_1, \dots, \theta_q$ sont des coefficients réels tels que $\theta_q \neq 0$.

On peut également écrire :

$$X_t = \sum_{k=0}^q \theta_k \varepsilon_{t-k} \Leftrightarrow X_t = \Theta(B) \varepsilon_t,$$

où $\theta_0 = 1$ et $\Theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$ est le polynôme caractéristique de degré q .

Propriété 2. Un processus à moyenne mobile $MA(q)$ est **toujours stationnaire**, quelle que soit la valeur des coefficients $\theta_1, \dots, \theta_q$.

1.11 Causalité et Inversibilité

1.11.1 Causalité

Définition 1.11.1. Processus causal

On dit que $\{X_t; t \in \mathbb{Z}\}$ est **causal** s'il existe une suite $\{a_k\}$ et un processus $\{\varepsilon_t; t \in \mathbb{Z}\}$ telle que :

$$\sum_{j=0}^{\infty} |a_j| < \infty$$

et que

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} \quad \text{où } \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2).$$

Remarque.

Théorème 1.11.1. Le processus $\{X_t, t \in \mathbb{Z}\}$ est causal si et seulement si toutes les racines du polynôme caractéristique $\phi(z)$ sont à l'extérieur du cercle unité, c'est-à-dire :

$$|z| > 1.$$

1.11.2 Inversibilité

Définition 1.11.2. Processus inversible

On dit que $\{X_t; t \in \mathbb{Z}\}$ est **inversible** s'il existe une suite $\{b_j\}$ telle que :

$$\sum_{j=0}^{\infty} |b_j| < \infty$$

et que :

$$\varepsilon_t = \sum_{j=0}^{\infty} b_j X_{t-j}$$

où ε_t est un bruit blanc.

Théorème 1.11.2. *Le processus $\{X_t, t \in \mathbb{Z}\}$ est inversible si et seulement si toutes les racines du polynôme $\Theta(z)$ sont à l'extérieur du cercle unité, c'est-à-dire :*

$$|z| > 1.$$

Remarque.

- Un processus $AR(p)$ est toujours inversible.
- Un processus $MA(q)$ est toujours causal.

1.11.3 Processus autorégressif moyenne mobile (ARMA)

Les modèles ARMA (AutoRegressive Moving Average) combinent une partie autorégressive (AR) et une partie moyenne mobile (MA). Il est possible de vérifier la stationnarité et l'inversibilité du modèle ARMA à l'aide des théorèmes précédemment énoncés.

Définition 1.11.3 (Brockwell, P. J., Davis, R. A. (2002)[8]). *Un modèle **ARMA**(p, q) est défini par l'équation suivante :*

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

Autrement dit,

$$\sum_{k=0}^p \phi_k X_{t-k} = \sum_{k=0}^q \theta_k \varepsilon_{t-k}$$

avec $\phi_0 = 1$ et $\theta_0 = 1$.

Nous pouvons ainsi écrire ce modèle sous forme polynomiale :

$$\Phi(B)X_t = \Theta(B)\varepsilon_t$$

où :

- $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ est le polynôme caractéristique d'ordre p ,
- $\Theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$ est le polynôme caractéristique d'ordre q .

Remarque.

- Si $p = 0$ et $q \neq 0$, alors $X_t \sim \mathbf{MA}(q)$,
- Si $p \neq 0$ et $q = 0$, alors $X_t \sim \mathbf{AR}(p)$,
- Si $p = 0$ et $q = 0$, alors $X_t = \varepsilon_t$, ce qui correspond à un bruit blanc.

Chapitre 2

Étude des modèles autorégressifs à coefficients aléatoires

2.1 Introduction

Au cours des dernières décennies, l'intérêt pour les modèles de séries chronologiques n'a cessé de croître. Parmi eux, les modèles autorégressifs à coefficients aléatoires (RCA), qui généralisent les modèles autorégressifs classiques (AR), ont suscité une attention particulière. Introduits par Anděl [2] et étudiés notamment par Nicholls et Quinn [23], ces modèles permettent de mieux capturer certaines dynamiques aléatoires dans les dépendances temporelles.

Cette section propose une présentation générale du modèle $\text{RCA}(p)$, avant de se focaliser sur le cas particulier $\text{RCA}(1)$, pour lequel nous examinerons les conditions de stationnarité et diverses méthodes d'estimation dans le cadre fréquentiste.

2.2 Modèle autorégressif à coefficients aléatoires d'ordre p

Définition 2.2.1. *Un modèle autorégressif à coefficients aléatoires d'ordre p , noté $RCA(p)$, est représenté sous la forme suivante :*

$$\mu_t = \sum_{i=1}^p (\phi_i + B_i(t)) \mu_{t-i} + \epsilon_t, \quad t \in \mathbb{Z} \quad (2.1)$$

où :

- $\{\epsilon_t, t \in \mathbb{Z}\}$ est une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) de moyenne nulle et de variance $\sigma^2 > 0$.
- $\forall i \in \{1, \dots, p\}$, la suite $\{B_i(t), t \in \mathbb{Z}\}$ est une suite de variables aléatoires i.i.d.
- ϕ_i sont des constantes, pour tout $i = 1, \dots, p$.

Nous nous intéressons à présent au cas particulier du modèle $RCA(1)$, modèle autorégressif à coefficient aléatoire d'ordre un.

2.3 Modèle autorégressif à coefficient aléatoire d'ordre un

Considérons deux suites de variables aléatoires $\{\epsilon_t, t \in \mathbb{Z}\}$ et $\{B(t), t \in \mathbb{Z}\}$ vérifiant les hypothèses suivantes :

- A1 : $\{\epsilon_t\}$ est une suite de variables aléatoires indépendantes, de moyenne nulle et de variance $\sigma_\epsilon^2 > 0$.
- A2 : $\{B_t(t)\}$ est une suite de variables aléatoires indépendantes, de moyenne nulle et de variance $\sigma_B^2 > 0$.
- A3 : Les suites $\{B(t)\}$ et $\{\epsilon_t\}$ sont indépendantes.

Définition 2.3.1. *Un processus $\{X_t, t \in \mathbb{Z}\}$ est dit autorégressif à coefficient aléa-*

toire d'ordre 1, noté $RCA(1)$, s'il vérifie l'équation suivante :

$$\mu_t = (\phi + B(t))\mu_{t-1} + \epsilon_t, \quad t \in \mathbb{Z} \quad (2.2)$$

où $\{\epsilon_t, t \in \mathbb{Z}\}$ et $\{B(t), t \in \mathbb{Z}\}$ satisfont les hypothèses précédemment énoncées, et ϕ est une constante réelle.

Dans le cas particulier où $\phi = 0$, le modèle s'écrit :

$$\mu_t = B(t)\mu_{t-1} + \epsilon_t, \quad t \in \mathbb{Z} \quad (2.3)$$

2.4 Stationnarité des modèles à coefficients aléatoires

2.4.1 Solution strictement stationnaire

Nous commençons par énoncer le résultat de **P. Billingsley** [5] permet de démontrer la stationnarité au sens strict.

Lemme 2.4.1. Soit $\{\mu_t, t \in \mathbb{Z}\}$ un processus stochastique. Considérons un processus stochastique $\{\psi_t, t \in \mathbb{Z}\}$ défini par :

$$\psi_t = \varphi(\dots, \mu_{t-1}, \mu_t, \dots),$$

où φ est une fonction mesurable. Si $\{\mu_t, t \in \mathbb{Z}\}$ est strictement stationnaire et ergodique (en particulier si les μ_t sont indépendants et identiquement distribués), alors $\{\psi_t\}$ est strictement stationnaire et ergodique.

2.4.2 Solution faiblement stationnaire

Théorème 2.4.2. (Hwang et Basawa [19])

Si

$$\mathbb{E}(B(t)^2) < 1,$$

c'est-à-dire :

$$\sigma_B^2 + B^2 < 1,$$

alors le processus (2.3) admet une unique solution

$$\mu_t = \varepsilon_t + \sum_{j=1}^{\infty} \left(\prod_{i=0}^{j-1} B(t-i) \right) \varepsilon_{t-j}. \quad (2.4)$$

Hwang et Basawa [19] démontrent ce théorème en se référant aux travaux de Nicholls et Quinn [23]. Ils notent d'abord que l'expression (2.8) satisfait l'équation (2.3), et que μ_t défini par (2.3) est stationnaire et ergodique.

Démonstration. Partons du modèle suivant :

$$\mu_t = B(t)\mu_{t-1} + \varepsilon_t.$$

En itérant cette relation, on obtient :

$$\begin{aligned} \mu_t &= B(t)(B(t-1)\mu_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= B(t)B(t-1)\mu_{t-2} + B(t)\varepsilon_{t-1} + \varepsilon_t \\ &= B(t)B(t-1)(B(t-2)\mu_{t-3} + \varepsilon_{t-2}) + B(t)\varepsilon_{t-1} + \varepsilon_t \\ &= B(t)B(t-1)B(t-2)\mu_{t-3} + B(t)B(t-1)\varepsilon_{t-2} + B(t)\varepsilon_{t-1} + \varepsilon_t \\ &\vdots \\ &= \left(\prod_{j=0}^{p-1} B(t-j) \right) \mu_{t-p} + \sum_{j=1}^p \left(\prod_{i=0}^{j-1} B(t-i) \right) \varepsilon_{t-j} + \varepsilon_t. \end{aligned}$$

D'après l'inégalité de Tchebychev, lorsque $p \rightarrow \infty$ on a :

$$\begin{aligned} \mathbb{P} \left(\left| \prod_{j=0}^{p-1} B(t-j) \mu_{t-p} \right| > \varepsilon \right) &\leq \frac{\mathbb{E} \left(\left(\prod_{j=0}^{p-1} B(t-j) \mu_{t-p} \right)^2 \right)}{\varepsilon^2} \\ \mathbb{P} \left(\left| \prod_{j=0}^{p-1} B(t-j) \mu_{t-p} \right| > \varepsilon \right) &\leq \frac{\prod_{j=0}^{p-1} \mathbb{E} (B(t-j) \mu_{t-p})^2}{\varepsilon^2} \\ \mathbb{P} \left(\left| \prod_{j=0}^{p-1} B(t-j) \mu_{t-p} \right| > \varepsilon \right) &\leq \frac{\prod_{j=0}^{p-1} \mathbb{E} (B(t-j)^2) \cdot \mathbb{E} (\mu_{t-p}^2)}{\varepsilon^2} \\ \mathbb{P} \left(\left| \prod_{j=0}^{p-1} B(t-j) \mu_{t-p} \right| > \varepsilon \right) &\leq \frac{(\sigma_B^2 + B^2)^p \cdot \mathbb{E} (\mu_{t-p}^2)}{\varepsilon^2} \rightarrow 0 \end{aligned}$$

Etant donné que

$$\mathbb{E} (\mu_{t-p}^2) < \infty$$

et

$$\sigma_B^2 + B^2 < 1$$

Il en résulte que le processus $\{\mu_t, t \in \mathbb{Z}\}$ est stationnaire et ergodique.

Théorème 2.4.3 (Wang et Basawa [19]). *On considère l'équation suivante :*

$$\mu_t = B(t) \mu_{t-1} + \varepsilon_t \tag{2.5}$$

où $\{(B(t), \varepsilon_t), t \in \mathbb{Z}\}$ est une suite de vecteurs aléatoires strictement stationnaire et ergodique telle que :

$$\mathbb{E}[\ln |B(t)|] < \infty \quad \text{et} \quad \mathbb{E}[\ln |\varepsilon_t|] < \infty.$$

Si $\mathbb{E}[\ln |B(t)|] < 0$, alors il existe une solution $\{\mu_t\}$ strictement stationnaire et ergodique pour l'équation (2.9), et telle que pour tout $t \in \mathbb{Z}$, μ_t est mesurable par rapport à $\mathcal{F}_n = \sigma(\{(B(n), \varepsilon_n), n \leq t\})$ la tribu engendrée par $\{(B(n), \varepsilon_n), t \in \mathbb{Z}\}$.

Démonstration. Supposons que

$$\mathbb{E}[\ln |B(t)|] = \alpha < 0$$

Définissons la suite $\{T_j\}$ pour $j \in \mathbb{N}$ par :

$$T_0 = \epsilon_t, \quad T_j = \prod_{i=0}^{j-1} B(n-i) \epsilon_{t-j}, \quad \text{pour } j \geq 1.$$

Soit $S_r = \sum_{j=0}^r T_j$, on a :

$$\ln(|T_j|) = \ln \left(\prod_{i=0}^{j-1} \|B(t-i)\| |\epsilon_j| \right)$$

$$\ln(|T_j|) = \ln |\epsilon_j| + \sum_{i=0}^{j-1} \ln |B(t-i)|$$

En divisant par j , on obtient :

$$\frac{1}{j} \ln |T_j| = \frac{1}{j} \ln |\epsilon_{t-j}| + \frac{1}{j} \sum_{i=0}^{j-1} \ln |B(t-i)|.$$

Par le théorème ergodique, on a :

$$\frac{1}{j} \ln |T_j| \xrightarrow{p.s.} \mathbb{E}[\ln |B(t)|] = \alpha < 0.$$

Donc $|T_j|^{1/j} \xrightarrow{p.s.} e^\alpha < 1$. À présent, posons l'ensemble des réalisations pour lesquelles la convergence est vérifiée. On les note T_j^* .

Comme $\exp(\alpha) < 1$, il existe alors une constante $\delta \in (0, 1)$ telle que :

$$|T_j^*| < \delta^j < 1 \quad \text{pour tout } j \in \mathbb{N}_\delta.$$

On en déduit que la série $\sum_{j=0}^r |T_j^*|$ est convergente, et par conséquent, la somme partielle :

$$S_r = \sum_{j=0}^r T_j$$

converge **presque sûrement** lorsque $r \rightarrow \infty$.

On décide alors de poser :

$$\mu_t^* = \lim_{r \rightarrow \infty} S_r = \epsilon_t + \sum_{j=1}^{\infty} \left(\prod_{i=0}^{j-1} B(t-i) \right) \epsilon_{t-j}.$$

Ainsi, μ_t^* est une solution mesurable à \mathcal{F}_n de l'équation (2.9). Comme $\{B(t), \epsilon_t\}$ est strictement stationnaire et ergodique, on en déduit que $\{\mu_t^*\}$ est également strictement stationnaire et ergodique.

Théorème 2.4.4 (Dazhe Wang[13]). *Soit $\{\mu_t; t \in \mathbb{Z}\}$ un modèle autorégressif à coefficients aléatoires d'ordre 1, défini par l'équation (2.3). Supposons que les hypothèses A1 et A3 (énoncées dans la présentation du modèle RCA(1), section 2.3) soient satisfaites.*

Si les $\{\epsilon_t, B(t)\}$ suivent des lois normales, alors une condition suffisante pour l'existence d'une solution strictement stationnaire et ergodique est :

$$\ln \sigma_B^2 < \zeta + \ln 2 - 2 \int_0^1 \frac{1 - \exp -\lambda(1 - \omega^2)}{1 - \omega^2} d\omega \quad (2.6)$$

où $\zeta \approx 0,57721$ est la constante d'Euler et $\lambda = \frac{\eta_B^2}{2\sigma_B^2}$.

Démonstration. Soit les données suivantes :

- $e_t = \epsilon_t + \alpha$,
- $B(t) = \eta_B + \sigma_B w_t$, avec $w_t \sim \mathcal{N}(0, 1)$,

et $\{(\epsilon_t, B(t)), t \in \mathbb{Z}\}$ est une suite strictement stationnaire et ergodique comme démontré dans le lemme 2.4.1

Il est important de remarquer que :

$$\mathbb{E}(\epsilon_t) < \infty \quad \text{et} \quad \mathbb{E}(\ln |B(t)|) < \infty.$$

Pour pouvoir faire appel au théorème 2.5.2, nous devons montrer avant tout que :

$$\mathbb{E}(\ln |B(t)|) = \mathbb{E}(\ln |\eta_B + \sigma_B w_t|) < 0.$$

On se retrouve face à deux cas : celui où $\eta_B = 0$, et celui où $\eta_B \neq 0$.

Cas 1 : $\eta_B = 0$ Dans ce cas, on a :

$$\mathbb{E}(\ln |B(t)|) = \mathbb{E}(\ln |\sigma_B w_t|).$$

En utilisant les règles de base des logarithmes, on obtient :

$$\mathbb{E}(\ln |B(t)|) = \mathbb{E}(\ln |\sigma_B|) + \mathbb{E}(\ln |w_t|).$$

Comme w_t est une variable aléatoire gaussienne standard, son espérance logarithmique existe. De plus, σ_B est une constante réelle strictement positive, donc :

$$\mathbb{E}(\ln |B(t)|) = \ln |\sigma_B| + \mathbb{E}(\ln |w_t|),$$

$$\mathbb{E}(\ln |B(t)|) = \frac{1}{2} \ln |\sigma_B|^2 + \mathbb{E}(\ln |w_t|),$$

$$\mathbb{E}(\ln |B(t)|) = \frac{1}{2} \ln |\sigma_B|^2 - \frac{1}{2}(\zeta + \ln 2),$$

Par conséquent $\mathbb{E}(\ln |B(t)|) < 0$.

Cas 2 : $\eta_B \neq 0$ On a alors :

$$\mathbb{E}(\ln |B(t)|) = \mathbb{E}(\ln |\eta_B + \sigma_B w_t|)$$

$$\mathbb{E}(\ln |B(t)|) = \mathbb{E} \left(\ln \left(|\eta_B| \cdot \left| 1 + \frac{\sigma_B}{\eta_B} w_t \right| \right) \right)$$

$$\mathbb{E}(\ln |B(t)|) = \mathbb{E}(\ln |\eta_B|) + \mathbb{E} \left(\ln \left| 1 + \frac{\sigma_B}{\eta_B} w_t \right| \right)$$

$$\mathbb{E}(\ln |B(t)|) = \ln |\eta_B| + \mathbb{E} \left(\ln \left| 1 + \frac{\sigma_B}{\eta_B} w_t \right| \right)$$

$$\mathbb{E}(\ln |B(t)|) = \frac{1}{2} \ln(|\eta_B|^2) + \mathbb{E}(\ln |1 + Z|)$$

où

$$Z \sim \mathcal{N}\left(0, \frac{\sigma_B^2}{\eta_B^2}\right)$$

En utilisant l'évaluation de $\mathbb{E}(\ln |1 + Z|)$, dans [23] on obtient :

$$\mathbb{E}(\ln |B(t)|) < 0 \iff \ln \sigma_B^2 < \zeta + \ln 2 - 2 \int_0^1 \frac{1 - \exp -\lambda(1 - \omega^2)}{1 - \omega^2} d\omega,$$

2.5 Estimation des paramètres d'un modèle RCA(1)

L'estimation des paramètres d'un modèle autorégressif à coefficient aléatoire, en particulier le modèle RCA(1), est étudiée à travers les méthodes des moindres carrés et du maximum de vraisemblance, en mettant l'accent sur les propriétés statistiques des estimateurs. Nous considérons le modèle :

$$\mu_t = (\phi + B(t))\mu_{t-1} + \varepsilon_t \tag{2.7}$$

où :

- $\{B(t), t \in \mathbb{Z}\}$ est une suite de variables aléatoires indépendantes et identiquement distribuées de moyenne nulle et de variance σ_B^2 ,
- $\{\varepsilon_t, t \in \mathbb{Z}\}$ est une suite de variables aléatoires indépendantes et identiquement distribuées de moyenne nulle et de variance σ_ε^2 .

2.5.1 Estimation des paramètres d'un RCA(1) par les moindres carrés

Soit $\mathcal{F}_t = \sigma(\mu_s, s \leq t)$ la tribu engendrée par les observations du processus jusqu'à l'instant t . On a :

$$\begin{aligned}\mathbb{E}(\mu_t | \mathcal{F}_{t-1}) &= \mathbb{E}(B(t) + \phi)\mu_{t-1} + \varepsilon_t | \mathcal{F}_{t-1}) \\ &= \mathbb{E}(B(t) + \phi)\mu_{t-1} | \mathcal{F}_{t-1}) + \mathbb{E}(\varepsilon_t | \mathcal{F}_{t-1})\end{aligned}$$

Comme $B(t)$ et ε_t sont indépendants de \mathcal{F}_{t-1} et μ_{t-1} est \mathcal{F}_{t-1} -mesurable, on a :

$$\begin{aligned}\mathbb{E}(\mu_{t-1} | \mathcal{F}_{t-1}) &= \mu_{t-1}, \\ \mathbb{E}(B(t) | \mathcal{F}_{t-1}) &= 0, \\ \mathbb{E}(\varepsilon_t | \mathcal{F}_{t-1}) &= 0.\end{aligned}$$

Donc :

$$\mathbb{E}(\mu_t | \mathcal{F}_{t-1}) = \phi\mu_{t-1}$$

On peut réécrire le modèle comme :

$$\mu_t = \phi\mu_{t-1} + B(t)\mu_{t-1} + \varepsilon_t$$

Posons $w_t = B(t)\mu_{t-1} + \varepsilon_t$,

soit $w_t = \mu_t - \phi\mu_{t-1}$, on pourra écrire :

$$\begin{aligned}\mathbb{E}(w_t | \mathcal{F}_{t-1}) &= \mathbb{E}(\mu_t - \phi\mu_{t-1} | \mathcal{F}_{t-1}) \\ &= \mathbb{E}(\mu_t | \mathcal{F}_{t-1}) - \mathbb{E}(\phi\mu_{t-1} | \mathcal{F}_{t-1}) \\ &= \mathbb{E}(w_t | \mathcal{F}_{t-1}) = 0\end{aligned}$$

Ainsi de la même manière nous pouvons écrire que :

$$\begin{aligned}\mathbb{E}(w_t^2 | \mathcal{F}_{t-1}) &= \mathbb{E}((\mu_t - \phi\mu_{t-1})^2 | \mathcal{F}_{t-1}) \\ &= \mathbb{E}(\mu_t^2 - 2\phi\mu_t\mu_{t-1} + \phi^2\mu_{t-1}^2 | \mathcal{F}_{t-1}) \\ &= \mathbb{E}(\mu_t^2 | \mathcal{F}_{t-1}) - 2\phi\mathbb{E}(\mu_t\mu_{t-1} | \mathcal{F}_{t-1}) + \phi^2\mathbb{E}(\mu_{t-1}^2 | \mathcal{F}_{t-1})\end{aligned}$$

Puisque :

$$\begin{aligned}\mathbb{E}(\mu_t^2 | \mathcal{F}_{t-1}) &= \mathbb{E}((B(t) + \phi)^2\mu_{t-1}^2 | \mathcal{F}_{t-1}) + \mathbb{E}(\varepsilon_t^2 | \mathcal{F}_{t-1}) \\ &\quad - 2\mathbb{E}(\varepsilon_t(B(t) + \phi)\mu_{t-1} | \mathcal{F}_{t-1})\end{aligned}$$

Comme on a :

$$\begin{aligned}\mathbb{E}(\mu_t^2 | \mathcal{F}_{t-1}) &= \mu_{t-1}^2\phi^2 + \mu_{t-1}^2\mathbb{E}(B(t)^2 | \mathcal{F}_{t-1}) + \mathbb{E}(\varepsilon_t^2 | \mathcal{F}_{t-1}) \\ &= \mu_{t-1}^2(\sigma_B^2 + \phi^2) + \sigma_\varepsilon^2\end{aligned}$$

Ainsi on peut conclure que :

$$\begin{aligned}\mathbb{E}(w_t^2 | \mathcal{F}_{t-1}) &= \mu_{t-1}^2(\sigma_B^2 + \phi^2) + \sigma_\varepsilon^2 + \phi^2\mu_{t-1}^2 - 2\phi^2\mu_{t-1}^2 \\ &= \sigma_\varepsilon^2 + \sigma_B^2\mu_{t-1}^2\end{aligned}$$

En considérant un échantillon $(\mu_0, \mu_1, \dots, \mu_n)$, nous estimons $\hat{\phi}$ en minimisant $\sum_{t=1}^n w_t^2$ par rapport à ϕ :

$$\frac{\partial}{\partial \phi} \sum_{t=1}^n w_t^2 = \frac{\partial}{\partial \phi} \sum_{t=1}^n (\mu_t - \phi\mu_{t-1})^2$$

Nous développons le terme de droite afin de pouvoir dériver par rapport à ϕ :

$$\frac{\partial}{\partial \phi} \sum_{t=1}^n (\mu_t^2 - 2\phi\mu_{t-1}\mu_t + \phi^2\mu_{t-1}^2)$$

En dérivant :

$$\frac{\partial}{\partial \phi} \sum_{t=1}^n w_t^2 = \sum_{t=1}^n (-2\mu_{t-1}\mu_t + 2\phi\mu_{t-1}^2)$$

Pour estimer $\hat{\phi}$, nous annulons la dérivée première. On remarque que :

$$\frac{\partial^2}{\partial \phi^2} \sum_{t=1}^n w_t^2 = 2 \sum_{t=1}^n \mu_{t-1}^2 > 0$$

La condition du minimum est donc bien vérifiée, d'où :

$$\sum_{t=1}^n (-2\mu_{t-1}X_t + 2\hat{\phi}_1\mu_{t-1}^2) = 0$$

Ainsi, on obtient :

$$\hat{\phi} = \frac{\sum_{t=1}^n \mu_t\mu_{t-1}}{\sum_{t=1}^n \mu_{t-1}^2}$$

La seconde étape de l'estimation par les moindres carrés consiste à former le résidu \hat{w}_t à partir de l'estimateur $\hat{\phi}$, afin de pouvoir ensuite estimer σ_ε^2 et σ_B^2 :

$$\hat{w}_t = \mu_t - \hat{\phi}_1\mu_{t-1}$$

On note

$$\ell(\sigma_\varepsilon^2, \sigma_B^2) = \sum_{t=1}^n (\hat{w}_t^2 - (\sigma_\varepsilon^2 + \sigma_B^2\mu_{t-1}^2))^2$$

En dérivant ℓ par rapport à σ_B^2 :

$$\frac{\partial \ell}{\partial \sigma_B^2} = -2 \sum_{t=1}^n \mu_{t-1}^2 (\hat{w}_t^2 - \sigma_\varepsilon^2 - \sigma_B^2\mu_{t-1}^2)$$

On voit de la même manière que la seconde dérivée est strictement positive :

$$\frac{\partial^2 \ell(\sigma_\varepsilon^2, \sigma_B^2)}{\partial^2 \sigma_B^2} = \frac{\partial}{\partial \sigma_B^2} \left(-2 \sum_{t=1}^n \mu_{t-1}^2 (\hat{w}_t^2 - \sigma_\varepsilon^2 - \sigma_B^2 \mu_{t-1}^2) \right)$$

$$\frac{\partial^2 \ell(\sigma_\varepsilon^2, \sigma_B^2)}{\partial^2 \sigma_B^2} = 2 \sum_{t=1}^n \mu_{t-1}^4 > 0$$

Il y a donc lieu de résoudre :

$$\frac{\partial \ell(\sigma_\varepsilon^2, \sigma_B^2)}{\partial \sigma_B^2} = 0$$

$$-2 \sum_{t=1}^n \mu_{t-1}^2 (\hat{w}_t^2 - \sigma_\varepsilon^2 - \hat{\sigma}_B^2 \mu_{t-1}^2) = 0$$

On peut en déduire que :

$$\hat{\sigma}_B^2 = \frac{\sum_{t=1}^n \mu_{t-1}^2 \hat{w}_t^2 - \hat{\sigma}_\varepsilon^2 \sum_{t=1}^n \mu_{t-1}^2}{\sum_{t=1}^n \mu_{t-1}^4}$$

À présent, nous estimons $\hat{\sigma}_\varepsilon^2$ de la même manière :

$$\frac{\partial \ell(\sigma_\varepsilon^2, \sigma_B^2)}{\partial \sigma_\varepsilon^2} = \frac{\partial}{\partial \sigma_\varepsilon^2} \sum_{t=1}^n (\hat{w}_t^2 - \sigma_\varepsilon^2 - \sigma_B^2 \mu_{t-1}^2)^2$$

$$\frac{\partial \ell(\sigma_\varepsilon^2, \sigma_B^2)}{\partial \sigma_\varepsilon^2} = -2 \sum_{t=1}^n (\hat{w}_t^2 - \sigma_\varepsilon^2 - \sigma_B^2 \mu_{t-1}^2) = 0$$

$$n \hat{\sigma}_\varepsilon^2 = \sum_{t=1}^n (\hat{w}_t^2 - \sigma_B^2 \mu_{t-1}^2)$$

Ainsi, nous pouvons conclure que :

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n} \sum_{t=1}^n (\hat{w}_t^2 - \hat{\sigma}_B^2 \mu_{t-1}^2)$$

La méthode est itérative, car l'estimation de chaque paramètre repose sur celle des autres, impliquant un processus de mise à jour successive jusqu'à convergence.

2.5.2 Consistance et normalité asymptotique de l'estimateur par la méthode des moindres carrés

D'après Nicholls et Quinn [23], pour déterminer la loi asymptotique de $\hat{\phi}$ la première étape consiste à établir que l'estimateur $\hat{\phi}$ est consistant, c'est-à-dire que :

$$\hat{\phi} \xrightarrow[n \rightarrow \infty]{p.s.} \frac{\gamma(1)}{\gamma(0)}$$

où :

$$\gamma(0) = \frac{\sigma_\varepsilon^2}{1 - \phi^2 - \sigma_B^2}, \quad \text{et} \quad \gamma(h) = \phi^h \gamma(0), \quad \forall h \in \mathbb{N}.$$

Pour déterminer la loi asymptotique de $\hat{\eta}_B$, certaines conditions doivent être remplies :

- (ε_t) et $(B(t))$ sont des suites de variables aléatoires indépendantes et identiquement distribuées ;
- $\mathbb{E}(\mu_t^4) < \infty$.

Ces conditions impliquent que le processus (μ_t) est markovien. Pour montrer que $\mathbb{E}(\mu_t^4) < \infty$, il est nécessaire de vérifier que $\mathbb{E}(B(t)^4) < 1$.

Théorème 2.5.1 (Nicholls et Quinn[23]). *Sous les conditions établies précédemment, nous avons :*

$$\sqrt{n}(\hat{\phi} - \phi) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, v^2),$$

avec :

$$v^2 = \frac{\sigma_\varepsilon^2 \mathbb{E}(\mu_0^2) + 2\sigma_\varepsilon^2 \sigma_B \mathbb{E}(\mu_0^3) + \sigma_B^2 \mathbb{E}(\mu_0^4)}{\gamma(0)^2}.$$

2.5.3 Estimation des paramètres d'un RCA(1) par le maximum de vraisemblance

Pour estimer les paramètres d'un modèle autoregressif à coefficient aléatoire d'ordre 1 (RCA(1)) par la méthode du maximum de vraisemblance, on considère le processus $\{\mu_t, t \in \mathbb{Z}\}$ strictement ergodique et stationnaire, \mathcal{F}_t -mesurable, avec :

$$\mu_t = (\phi + B(t))\mu_{t-1} + \varepsilon_t$$

sous les conditions suivantes :

- $(\phi, \sigma_B^2) \in \mathcal{D}$, où \mathcal{D} est un ensemble fermé inclus dans la région de stationnarité stricte et d'ergodicité, donnée par l'inégalité :

$$\sigma_B^2 < \zeta + \ln 2 - 2 \int_0^1 \frac{1 - \exp(-\theta(1 - \omega^2))}{1 - \omega^2} d\omega$$

- $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ et $\{B(t)\}_{t \in \mathbb{Z}}$ sont deux suites de variables aléatoires indépendantes, avec $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ et $B(t) \sim \mathcal{N}(0, \sigma_B^2)$.

Les propriétés de l'espérance conditionnelle donnent :

$$\mathbb{E}(\mu_t | \mu_{t-1}) = \mathbb{E}((B(t) + \phi)\mu_{t-1} | \mu_{t-1}) + \mathbb{E}(\varepsilon_t | \mu_{t-1}) = \phi\mu_{t-1}$$

L'application de ces propriétés à la variance nous permet d'écrire :

$$\text{Var}(\mu_t | \mu_{t-1}) = \text{Var}((B(t) + \phi)\mu_{t-1} | \mu_{t-1}) + \text{Var}(\varepsilon_t | \mu_{t-1}) = \sigma_B^2 \mu_{t-1}^2 + \sigma_\varepsilon^2$$

En supposant la normalité de $B(t)$ et ε_t , la fonction de vraisemblance conditionnelle sachant μ_0 s'écrit :

$$f_n(\mu_1, \dots, \mu_n | \mu_0) = \prod_{t=1}^n f(\mu_t | \mu_{t-1}) = \left(\frac{1}{2\pi}\right)^{n/2} \prod_{t=1}^n \left(\frac{1}{\sigma_\varepsilon^2 + \sigma_B^2 \mu_{t-1}^2}\right)^{1/2} \exp\left(-\frac{(\mu_t - \phi\mu_{t-1})^2}{2(\sigma_\varepsilon^2 + \sigma_B^2 \mu_{t-1}^2)}\right)$$

On note :

$$L(\sigma_\varepsilon^2, \phi, \sigma_B^2) = f_n(\mu_1, \dots, \mu_n | \mu_0)$$

et on travaille avec la fonction logarithmique de vraisemblance notée \bar{l} :

$$\bar{l}(\sigma_\varepsilon^2, \phi, \sigma_B^2) = \frac{1}{n} \sum_{t=1}^n \ln(\sigma_\varepsilon^2 + \sigma_B^2 \mu_{t-1}^2) + \frac{1}{n} \sum_{t=1}^n \frac{(\mu_t - \phi\mu_{t-1})^2}{\sigma_\varepsilon^2 + \sigma_B^2 \mu_{t-1}^2}$$

Pour simplifier, posons $\tau = \frac{\sigma_B^2}{\sigma_\varepsilon^2}$, alors :

$$\bar{l}(\sigma_\varepsilon^2, \phi, \sigma_B^2) = g(\phi, \tau, \sigma_\varepsilon^2)$$

où :

$$g(\phi, \tau, \sigma_\varepsilon^2) = \ln(\sigma_\varepsilon^2) + \frac{1}{n} \sum_{t=1}^n \ln(1 + \tau \mu_{t-1}^2) + \frac{1}{n \sigma_\varepsilon^2} \sum_{t=1}^n \frac{(\mu_t - \phi \mu_{t-1})^2}{1 + \tau \mu_{t-1}^2}$$

En dérivant par rapport à ϕ :

$$\frac{\partial g}{\partial \phi} = -\frac{2}{n \sigma_\varepsilon^2} \sum_{t=1}^n \frac{\mu_{t-1}(\mu_t - \phi \mu_{t-1})}{1 + \tau \mu_{t-1}^2}$$

En résolvant $\frac{\partial g}{\partial \phi} = 0$, on obtient l'estimateur :

$$\hat{\phi} = \left[\frac{1}{n} \sum_{t=1}^n \frac{\mu_t \mu_{t-1}}{1 + \tau \mu_{t-1}^2} \right] \left[\frac{1}{n} \sum_{t=1}^n \frac{\mu_{t-1}^2}{1 + \tau \mu_{t-1}^2} \right]^{-1}$$

De la même manière, en dérivant g par rapport à σ_ε^2 :

$$\frac{\partial g}{\partial \sigma_\varepsilon^2} = \frac{1}{\sigma_\varepsilon^2} - \frac{1}{n(\sigma_\varepsilon^2)^2} \sum_{t=1}^n \frac{(\mu_t - \phi \mu_{t-1})^2}{1 + \tau \mu_{t-1}^2}$$

ce qui mène à :

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n} \sum_{t=1}^n \frac{(\mu_t - \phi \mu_{t-1})^2}{1 + \tau \mu_{t-1}^2}$$

Enfin, on considère la fonction minimisée :

$\bar{g}(\phi, \tau, \sigma$ stationnaire ergodique dans $\mathbb{L}^\#$, donnée par :

$$\mu_t = \varepsilon_t + \sum_{j=1}^{\infty} \left(\prod_{i=0}^{j-1} B(t-i) \right) \varepsilon_{t-j}. \quad (2.8)$$

Hwang et Basawa [19] démontrent ce théorème en se référant aux travaux de Nicholls et Quinn [23]. Ils notent d'abord que l'expression (2.8) satisfait l'équation (2.3), et que μ_t défini par (2.3) est stationnaire et ergodique.

Démonstration. Partons du modèle suivant :

$$\mu_t = B(t)\mu_{t-1} + \varepsilon_t.$$

En itérant cette relation, on obtient :

$$\begin{aligned} \mu_t &= B(t)(B(t-1)\mu_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= B(t)B(t-1)\mu_{t-2} + B(t)\varepsilon_{t-1} + \varepsilon_t \\ &= B(t)B(t-1)(B(t-2)\mu_{t-3} + \varepsilon_{t-2}) + B(t)\varepsilon_{t-1} + \varepsilon_t \\ &= B(t)B(t-1)B(t-2)\mu_{t-3} + B(t)B(t-1)\varepsilon_{t-2} + B(t)\varepsilon_{t-1} + \varepsilon_t \\ &\vdots \\ &= \left(\prod_{j=0}^{p-1} B(t-j) \right) \mu_{t-p} + \sum_{j=1}^p \left(\prod_{i=0}^{j-1} B(t-i) \right) \varepsilon_{t-j} + \varepsilon_t. \end{aligned}$$

D'après l'inégalité de Tchebychev, lorsque $p \rightarrow \infty$ on a :

$$\begin{aligned} \mathbb{P} \left(\left| \prod_{j=0}^{p-1} B(t-j) \mu_{t-p} \right| > \varepsilon \right) &\leq \frac{\mathbb{E} \left(\left(\prod_{j=0}^{p-1} B(t-j) \mu_{t-p} \right)^2 \right)}{\varepsilon^2} \\ \mathbb{P} \left(\left| \prod_{j=0}^{p-1} B(t-j) \mu_{t-p} \right| > \varepsilon \right) &\leq \frac{\prod_{j=0}^{p-1} \mathbb{E}(B(t-j) \mu_{t-p})^2}{\varepsilon^2} \\ \mathbb{P} \left(\left| \prod_{j=0}^{p-1} B(t-j) \mu_{t-p} \right| > \varepsilon \right) &\leq \frac{\prod_{j=0}^{p-1} \mathbb{E}(B(t-j)^2) \cdot \mathbb{E}(\mu_{t-p}^2)}{\varepsilon^2} \\ \mathbb{P} \left(\left| \prod_{j=0}^{p-1} B(t-j) \mu_{t-p} \right| > \varepsilon \right) &\leq \frac{(\sigma_B^2 + B^2)^p \cdot \mathbb{E}(\mu_{t-p}^2)}{\varepsilon^2} \longrightarrow 0 \end{aligned}$$

Etant donné que

$$\mathbb{E}(\mu_{t-p}^2) < \infty$$

et

$$\sigma_B^2 + B^2 < 1$$

Il en résulte que le processus $\{\mu_t, t \in \mathbb{Z}\}$ est stationnaire et ergodique.

Théorème 2.5.2 (Wang et Basawa [19]). *On considère l'équation suivante :*

$$\mu_t = B(t)\mu_{t-1} + \varepsilon_t \quad (2.9)$$

où $\{(B(t), \varepsilon_t), t \in \mathbb{Z}\}$ est une suite de vecteurs aléatoires strictement stationnaire et ergodique telle que :

$$\mathbb{E}[\ln |B(t)|] < \infty \quad \text{et} \quad \mathbb{E}[\ln |\varepsilon_t|] < \infty.$$

Si $\mathbb{E}[\ln |B(t)|] < 0$, alors il existe une solution $\{\mu_t\}$ strictement stationnaire et ergodique pour l'équation (2.9), et telle que pour tout $t \in \mathbb{Z}$, μ_t est mesurable par rapport à $\mathcal{F}_n = \sigma(\{(B(n), \varepsilon_n), n \leq t\})$ la tribu engendrée par $\{(B(n), \varepsilon_n), t \in \mathbb{Z}\}$.

Démonstration. Supposons que

$$\mathbb{E}[\ln |B(t)|] = \alpha < 0$$

Définissons la suite $\{T_j\}$ pour $j \in \mathbb{N}$ par :

$$T_0 = \varepsilon_t, \quad T_j = \prod_{i=0}^{j-1} B(n-i) \varepsilon_{t-j}, \quad \text{pour } j \geq 1.$$

Soit $S_r = \sum_{j=0}^r T_j$, on a :

$$\ln(|T_j|) = \ln \left(\prod_{i=0}^{j-1} |B(t-i)| |\varepsilon_j| \right)$$

$$\ln(|T_j|) = \ln |\varepsilon_j| + \sum_{i=0}^{j-1} \ln |B(t-i)|$$

En divisant par j , on obtient :

$$\frac{1}{j} \ln |T_j| = \frac{1}{j} \ln |\varepsilon_{t-j}| + \frac{1}{j} \sum_{i=0}^{j-1} \ln |B(t-i)|.$$

Par le théorème ergodique, on a :

$$\frac{1}{j} \ln |T_j| \xrightarrow{p.s.} \mathbb{E}[\ln |B(t)|] = \alpha < 0.$$

Donc $|T_j|^{1/j} \xrightarrow{p.s.} e^\alpha < 1$. À présent, posons l'ensemble des réalisations pour lesquelles la convergence est vérifiée. On les note T_j^* .

Comme $\exp(\alpha) < 1$, il existe alors une constante $\delta \in (0, 1)$ telle que :

$$|T_j^*| < \delta^j < 1 \quad \text{pour tout } j \in \mathbb{N}_\delta.$$

On en déduit que la série $\sum_{j=0}^r |T_j^*|$ est convergente, et par conséquent, la somme partielle :

$$S_r = \sum_{j=0}^r T_j$$

converge **presque sûrement** lorsque $r \rightarrow \infty$.

On décide alors de poser :

$$\mu_t^* = \lim_{r \rightarrow \infty} S_r = \epsilon_t + \sum_{j=1}^{\infty} \left(\prod_{i=0}^{j-1} B(t-i) \right) \epsilon_{t-j}.$$

Ainsi, μ_t^* est une solution mesurable à \mathcal{F}_n de l'équation (2.9). Comme $\{B(t), \epsilon_t\}$ est strictement stationnaire et ergodique, on en déduit que $\{\mu_t^*\}$ est également strictement stationnaire et ergodique.

Théorème 2.5.3 (Dazhe Wang[13]). *Soit $\{\mu_t; t \in \mathbb{Z}\}$ un modèle autorégressif à coefficients aléatoires d'ordre 1, défini par l'équation (2.3). Supposons que les hypothèses A1 et A3 (énoncées dans la présentation du modèle RCA(1), section 2.3) soient satisfaites.*

Si les $\{\epsilon_t, B(t)\}$ suivent des lois normales, alors une condition suffisante pour l'existence d'une solution strictement stationnaire et ergodique est :

$$\ln \sigma_B^2 < \zeta + \ln 2 - 2 \int_0^1 \frac{1 - \exp -\lambda(1 - \omega^2)}{1 - \omega^2} d\omega \quad (2.10)$$

où $\zeta \approx 0,57721$ est la constante d'Euler et $\lambda = \frac{\eta_B^2}{2\sigma_B^2}$.

Démonstration. Soit les données suivantes :

$$- e_t = \epsilon_t + \alpha,$$

$$— B(t) = \eta_B + \sigma_B w_t, \text{ avec } w_t \sim \mathcal{N}(0, 1),$$

et $\{(\varepsilon_t, B(t)), t \in \mathbb{Z}\}$ est une suite strictement stationnaire et ergodique comme démontré dans le lemme 2.4.1

Il est important de remarquer que :

$$\mathbb{E}(\varepsilon_t) < \infty \quad \text{et} \quad \mathbb{E}(\ln |B(t)|) < \infty.$$

Pour pouvoir faire appel au théorème 2.5.2, nous devons montrer avant tout que :

$$\mathbb{E}(\ln |B(t)|) = \mathbb{E}(\ln |\eta_B + \sigma_B w_t|) < 0.$$

On se retrouve face à deux cas : celui où $\eta_B = 0$, et celui où $\eta_B \neq 0$.

Cas 1 : $\eta_B = 0$ Dans ce cas, on a :

$$\mathbb{E}(\ln |B(t)|) = \mathbb{E}(\ln |\sigma_B w_t|).$$

En utilisant les règles de base des logarithmes, on obtient :

$$\mathbb{E}(\ln |B(t)|) = \mathbb{E}(\ln |\sigma_B|) + \mathbb{E}(\ln |w_t|).$$

Comme w_t est une variable aléatoire gaussienne standard, son espérance logarithmique existe. De plus, σ_B est une constante réelle strictement positive, donc :

$$\mathbb{E}(\ln |B(t)|) = \ln |\sigma_B| + \mathbb{E}(\ln |w_t|),$$

$$\mathbb{E}(\ln |B(t)|) = \frac{1}{2} \ln |\sigma_B|^2 + \mathbb{E}(\ln |w_t|),$$

$$\mathbb{E}(\ln |B(t)|) = \frac{1}{2} \ln |\sigma_B|^2 - \frac{1}{2}(\zeta + \ln 2),$$

Par conséquent $\mathbb{E}(\ln |B(t)|) < 0$.

Cas 2 : $\eta_B \neq 0$ On a alors :

$$\mathbb{E}(\ln |B(t)|) = \mathbb{E}(\ln |\eta_B + \sigma_B w_t|)$$

$$\mathbb{E}(\ln |B(t)|) = \mathbb{E} \left(\ln \left(|\eta_B| \cdot \left| 1 + \frac{\sigma_B}{\eta_B} w_t \right| \right) \right)$$

$$\mathbb{E}(\ln |B(t)|) = \mathbb{E}(\ln |\eta_B|) + \mathbb{E} \left(\ln \left| 1 + \frac{\sigma_B}{\eta_B} w_t \right| \right)$$

$$\mathbb{E}(\ln |B(t)|) = \ln |\eta_B| + \mathbb{E} \left(\ln \left| 1 + \frac{\sigma_B}{\eta_B} w_t \right| \right)$$

$$\mathbb{E}(\ln |B(t)|) = \frac{1}{2} \ln(|\eta_B|^2) + \mathbb{E}(\ln |1 + Z|)$$

où

$$Z \sim \mathcal{N} \left(0, \frac{\sigma_B^2}{\eta_B^2} \right)$$

En utilisant l'évaluation de $\mathbb{E}(\ln |1 + Z|)$, dans [23] on obtient :

$$\mathbb{E}(\ln |B(t)|) < 0 \iff \ln \sigma_B^2 < \zeta + \ln 2 - 2 \int_0^1 \frac{1 - \exp -\lambda(1 - \omega^2)}{1 - \omega^2} d\omega,$$

2.6 Estimation des paramètres d'un modèle RCA(1)

L'estimation des paramètres d'un modèle autorégressif à coefficient aléatoire, en particulier le modèle RCA(1), est étudiée à travers les méthodes des moindres carrés et du maximum de vraisemblance, en mettant l'accent sur les propriétés statistiques des estimateurs. Nous considérons le modèle :

$$\mu_t = (\phi + B(t))\mu_{t-1} + \varepsilon_t \tag{2.11}$$

où :

- $\{B(t), t \in \mathbb{Z}\}$ est une suite de variables aléatoires indépendantes et identiquement distribuées de moyenne nulle et de variance σ_B^2 ,
- $\{\varepsilon_t, t \in \mathbb{Z}\}$ est une suite de variables aléatoires indépendantes et identiquement distribuées de moyenne nulle et de variance σ_ε^2 .

2.6.1 Estimation des paramètres d'un RCA(1) par les moindres carrés

Soit $\mathcal{F}_t = \sigma(\mu_s, s \leq t)$ la tribu engendrée par les observations du processus jusqu'à l'instant t . On a :

$$\begin{aligned} \mathbb{E}(\mu_t | \mathcal{F}_{t-1}) &= \mathbb{E}(B(t) + \phi)\mu_{t-1} + \varepsilon_t | \mathcal{F}_{t-1}) \\ &= \mathbb{E}(B(t) + \phi)\mu_{t-1} | \mathcal{F}_{t-1}) + \mathbb{E}(\varepsilon_t | \mathcal{F}_{t-1}) \end{aligned}$$

Comme $B(t)$ et ε_t sont indépendants de \mathcal{F}_{t-1} et μ_{t-1} est \mathcal{F}_{t-1} -mesurable, on a :

$$\begin{aligned} \mathbb{E}(\mu_{t-1} | \mathcal{F}_{t-1}) &= \mu_{t-1}, \\ \mathbb{E}(B(t) | \mathcal{F}_{t-1}) &= 0, \\ \mathbb{E}(\varepsilon_t | \mathcal{F}_{t-1}) &= 0. \end{aligned}$$

Donc :

$$\mathbb{E}(\mu_t | \mathcal{F}_{t-1}) = \phi\mu_{t-1}$$

On peut réécrire le modèle comme :

$$\mu_t = \phi\mu_{t-1} + B(t)\mu_{t-1} + \varepsilon_t$$

Posons $w_t = B(t)\mu_{t-1} + \varepsilon_t$,

soit $w_t = \mu_t - \phi\mu_{t-1}$, on pourra écrire :

$$\begin{aligned}\mathbb{E}(w_t \mid \mathcal{F}_{t-1}) &= \mathbb{E}(\mu_t - \phi\mu_{t-1} \mid \mathcal{F}_{t-1}) \\ &= \mathbb{E}(\mu_t \mid \mathcal{F}_{t-1}) - \mathbb{E}(\phi\mu_{t-1} \mid \mathcal{F}_{t-1}) \\ &= \mathbb{E}(w_t \mid \mathcal{F}_{t-1}) = 0\end{aligned}$$

Ainsi de la même manière nous pouvons écrire que :

$$\begin{aligned}\mathbb{E}(w_t^2 \mid \mathcal{F}_{t-1}) &= \mathbb{E}((\mu_t - \phi\mu_{t-1})^2 \mid \mathcal{F}_{t-1}) \\ &= \mathbb{E}(\mu_t^2 - 2\phi\mu_t\mu_{t-1} + \phi^2\mu_{t-1}^2 \mid \mathcal{F}_{t-1}) \\ &= \mathbb{E}(\mu_t^2 \mid \mathcal{F}_{t-1}) - 2\phi\mathbb{E}(\mu_t\mu_{t-1} \mid \mathcal{F}_{t-1}) + \phi^2\mathbb{E}(\mu_{t-1}^2 \mid \mathcal{F}_{t-1})\end{aligned}$$

Puisque :

$$\begin{aligned}\mathbb{E}(\mu_t^2 \mid \mathcal{F}_{t-1}) &= \mathbb{E}((B(t) + \phi)^2\mu_{t-1}^2 \mid \mathcal{F}_{t-1}) + \mathbb{E}(\varepsilon_t^2 \mid \mathcal{F}_{t-1}) \\ &\quad - 2\mathbb{E}(\varepsilon_t(B(t) + \phi)\mu_{t-1} \mid \mathcal{F}_{t-1})\end{aligned}$$

Comme on a :

$$\begin{aligned}\mathbb{E}(\mu_t^2 \mid \mathcal{F}_{t-1}) &= \mu_{t-1}^2\phi^2 + \mu_{t-1}^2\mathbb{E}(B(t)^2 \mid \mathcal{F}_{t-1}) + \mathbb{E}(\varepsilon_t^2 \mid \mathcal{F}_{t-1}) \\ &= \mu_{t-1}^2(\sigma_B^2 + \phi^2) + \sigma_\varepsilon^2\end{aligned}$$

Ainsi on peut conclure que :

$$\begin{aligned}\mathbb{E}(w_t^2 \mid \mathcal{F}_{t-1}) &= \mu_{t-1}^2(\sigma_B^2 + \phi^2) + \sigma_\varepsilon^2 + \phi^2\mu_{t-1}^2 - 2\phi^2\mu_{t-1}^2 \\ &= \sigma_\varepsilon^2 + \sigma_B^2\mu_{t-1}^2\end{aligned}$$

En considérant un échantillon $(\mu_0, \mu_1, \dots, \mu_n)$, nous estimons $\hat{\phi}$ en minimisant $\sum_{t=1}^n w_t^2$

par rapport à ϕ :

$$\frac{\partial}{\partial \phi} \sum_{t=1}^n w_t^2 = \frac{\partial}{\partial \phi} \sum_{t=1}^n (\mu_t - \phi \mu_{t-1})^2$$

Nous développons le terme de droite afin de pouvoir dériver par rapport à ϕ :

$$\frac{\partial}{\partial \phi} \sum_{t=1}^n (\mu_t^2 - 2\phi \mu_{t-1} \mu_t + \phi^2 \mu_{t-1}^2)$$

En dérivant :

$$\frac{\partial}{\partial \phi} \sum_{t=1}^n w_t^2 = \sum_{t=1}^n (-2\mu_{t-1} \mu_t + 2\phi \mu_{t-1}^2)$$

Pour estimer $\hat{\phi}$, nous annulons la dérivée première. On remarque que :

$$\frac{\partial^2}{\partial \phi^2} \sum_{t=1}^n w_t^2 = 2 \sum_{t=1}^n \mu_{t-1}^2 > 0$$

La condition du minimum est donc bien vérifiée, d'où :

$$\sum_{t=1}^n (-2\mu_{t-1} X_t + 2\hat{\phi}_1 \mu_{t-1}^2) = 0$$

Ainsi, on obtient :

$$\hat{\phi} = \frac{\sum_{t=1}^n \mu_t \mu_{t-1}}{\sum_{t=1}^n \mu_{t-1}^2}$$

La seconde étape de l'estimation par les moindres carrés consiste à former le résidu \hat{w}_t à partir de l'estimateur $\hat{\phi}$, afin de pouvoir ensuite estimer σ_ε^2 et σ_B^2 :

$$\hat{w}_t = \mu_t - \hat{\phi}_1 \mu_{t-1}$$

On note

$$\ell(\sigma_\varepsilon^2, \sigma_B^2) = \sum_{t=1}^n (\hat{w}_t^2 - (\sigma_\varepsilon^2 + \sigma_B^2 \mu_{t-1}^2))^2$$

En dérivant ℓ par rapport à σ_B^2 :

$$\frac{\partial \ell}{\partial \sigma_B^2} = -2 \sum_{t=1}^n \mu_{t-1}^2 \left(\hat{w}_t^2 - \sigma_\varepsilon^2 - \sigma_B^2 \mu_{t-1}^2 \right)$$

On voit de la même manière que la seconde dérivée est strictement positive :

$$\frac{\partial^2 \ell(\sigma_\varepsilon^2, \sigma_B^2)}{\partial^2 \sigma_B^2} = \frac{\partial}{\partial \sigma_B^2} \left(-2 \sum_{t=1}^n \mu_{t-1}^2 \left(\hat{w}_t^2 - \sigma_\varepsilon^2 - \sigma_B^2 \mu_{t-1}^2 \right) \right)$$

$$\frac{\partial^2 \ell(\sigma_\varepsilon^2, \sigma_B^2)}{\partial^2 \sigma_B^2} = 2 \sum_{t=1}^n \mu_{t-1}^4 > 0$$

Il y a donc lieu de résoudre :

$$\frac{\partial \ell(\sigma_\varepsilon^2, \sigma_B^2)}{\partial \sigma_B^2} = 0$$

$$-2 \sum_{t=1}^n \mu_{t-1}^2 \left(\hat{w}_t^2 - \sigma_\varepsilon^2 - \hat{\sigma}_B^2 \mu_{t-1}^2 \right) = 0$$

On peut en déduire que :

$$\hat{\sigma}_B^2 = \frac{\sum_{t=1}^n \mu_{t-1}^2 \hat{w}_t^2 - \hat{\sigma}_\varepsilon^2 \sum_{t=1}^n \mu_{t-1}^2}{\sum_{t=1}^n \mu_{t-1}^4}$$

À présent, nous estimons $\hat{\sigma}_\varepsilon^2$ de la même manière :

$$\frac{\partial \ell(\sigma_\varepsilon^2, \sigma_B^2)}{\partial \sigma_\varepsilon^2} = \frac{\partial}{\partial \sigma_\varepsilon^2} \sum_{t=1}^n \left(\hat{w}_t^2 - \sigma_\varepsilon^2 - \sigma_B^2 \mu_{t-1}^2 \right)^2$$

$$\frac{\partial \ell(\sigma_\varepsilon^2, \sigma_B^2)}{\partial \sigma_\varepsilon^2} = -2 \sum_{t=1}^n \left(\hat{w}_t^2 - \sigma_\varepsilon^2 - \sigma_B^2 \mu_{t-1}^2 \right) = 0$$

$$n \hat{\sigma}_\varepsilon^2 = \sum_{t=1}^n \left(\hat{w}_t^2 - \sigma_B^2 \mu_{t-1}^2 \right)$$

Ainsi, nous pouvons conclure que :

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n} \sum_{t=1}^n \left(\hat{w}_t^2 - \hat{\sigma}_B^2 \mu_{t-1}^2 \right)$$

La méthode est itérative, car l'estimation de chaque paramètre repose sur celle des autres, impliquant un processus de mise à jour successive jusqu'à convergence.

2.6.2 Consistance et normalité asymptotique de l'estimateur par la méthode des moindres carrés

D'après Nicholls et Quinn [23], pour déterminer la loi asymptotique de $\hat{\phi}$ la première étape consiste à établir que l'estimateur $\hat{\phi}$ est consistant, c'est-à-dire que :

$$\hat{\phi} \xrightarrow[n \rightarrow \infty]{p.s.} \frac{\gamma(1)}{\gamma(0)}$$

où :

$$\gamma(0) = \frac{\sigma_\varepsilon^2}{1 - \phi^2 - \sigma_B^2}, \quad \text{et} \quad \gamma(h) = \phi^h \gamma(0), \quad \forall h \in \mathbb{N}.$$

Pour déterminer la loi asymptotique de $\hat{\eta}_B$, certaines conditions doivent être remplies :

- (ε_t) et $(B(t))$ sont des suites de variables aléatoires indépendantes et identiquement distribuées ;
- $\mathbb{E}(\mu_t^4) < \infty$.

Ces conditions impliquent que le processus (μ_t) est markovien. Pour montrer que $\mathbb{E}(\mu_t^4) < \infty$, il est nécessaire de vérifier que $\mathbb{E}(B(t)^4) < 1$.

Théorème 2.6.1 (Nicholls et Quinn[23]). *Sous les conditions établies précédemment, nous avons :*

$$\sqrt{n}(\hat{\phi} - \phi) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, v^2),$$

avec :

$$v^2 = \frac{\sigma_\varepsilon^2 \mathbb{E}(\mu_0^2) + 2\sigma_\varepsilon^2 \sigma_B \mathbb{E}(\mu_0^3) + \sigma_B^2 \mathbb{E}(\mu_0^4)}{\gamma(0)^2}.$$

2.6.3 Estimation des paramètres d'un RCA(1) par le maximum de vraisemblance

Pour estimer les paramètres d'un modèle autoregressif à coefficient aléatoire d'ordre 1 (RCA(1)) par la méthode du maximum de vraisemblance, on considère le processus $\{\mu_t, t \in \mathbb{Z}\}$ strictement ergodique et stationnaire, \mathcal{F}_t -mesurable, avec :

$$\mu_t = (\phi + B(t))\mu_{t-1} + \varepsilon_t$$

sous les conditions suivantes :

- $(\phi, \sigma_B^2) \in \mathcal{D}$, où \mathcal{D} est un ensemble fermé inclus dans la région de stationnarité stricte et d'ergodicité, donnée par l'inégalité :

$$\sigma_B^2 < \zeta + \ln 2 - 2 \int_0^1 \frac{1 - \exp(-\theta(1 - \omega^2))}{1 - \omega^2} d\omega$$

- $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ et $\{B(t)\}_{t \in \mathbb{Z}}$ sont deux suites de variables aléatoires indépendantes, avec $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ et $B(t) \sim \mathcal{N}(0, \sigma_B^2)$.

Les propriétés de l'espérance conditionnelle donnent :

$$\mathbb{E}(\mu_t | \mu_{t-1}) = \mathbb{E}((B(t) + \phi)\mu_{t-1} | \mu_{t-1}) + \mathbb{E}(\varepsilon_t | \mu_{t-1}) = \phi\mu_{t-1}$$

L'application de ces propriétés à la variance nous permet d'écrire :

$$\text{Var}(\mu_t | \mu_{t-1}) = \text{Var}((B(t) + \phi)\mu_{t-1} | \mu_{t-1}) + \text{Var}(\varepsilon_t | \mu_{t-1}) = \sigma_B^2 \mu_{t-1}^2 + \sigma_\varepsilon^2$$

En supposant la normalité de $B(t)$ et ε_t , la fonction de vraisemblance conditionnelle sachant μ_0 s'écrit :

$$f_n(\mu_1, \dots, \mu_n | \mu_0) = \prod_{t=1}^n f(\mu_t | \mu_{t-1}) = \left(\frac{1}{2\pi}\right)^{n/2} \prod_{t=1}^n \left(\frac{1}{\sigma_\varepsilon^2 + \sigma_B^2 \mu_{t-1}^2}\right)^{1/2} \exp\left(-\frac{(\mu_t - \phi\mu_{t-1})^2}{2(\sigma_\varepsilon^2 + \sigma_B^2 \mu_{t-1}^2)}\right)$$

On note :

$$L(\sigma_\varepsilon^2, \phi, \sigma_B^2) = f_n(\mu_1, \dots, \mu_n | \mu_0)$$

et on travaille avec la fonction logarithmique de vraisemblance notée \bar{l} :

$$\bar{l}(\sigma_\varepsilon^2, \phi, \sigma_B^2) = \frac{1}{n} \sum_{t=1}^n \ln(\sigma_\varepsilon^2 + \sigma_B^2 \mu_{t-1}^2) + \frac{1}{n} \sum_{t=1}^n \frac{(\mu_t - \phi\mu_{t-1})^2}{\sigma_\varepsilon^2 + \sigma_B^2 \mu_{t-1}^2}$$

Pour simplifier, posons $\tau = \frac{\sigma_B^2}{\sigma_\varepsilon^2}$, alors :

$$\bar{l}(\sigma_\varepsilon^2, \phi, \sigma_B^2) = g(\phi, \tau, \sigma_\varepsilon^2)$$

où :

$$g(\phi, \tau, \sigma_\varepsilon^2) = \ln(\sigma_\varepsilon^2) + \frac{1}{n} \sum_{t=1}^n \ln(1 + \tau \mu_{t-1}^2) + \frac{1}{n \sigma_\varepsilon^2} \sum_{t=1}^n \frac{(\mu_t - \phi \mu_{t-1})^2}{1 + \tau \mu_{t-1}^2}$$

En dérivant par rapport à ϕ :

$$\frac{\partial g}{\partial \phi} = -\frac{2}{n \sigma_\varepsilon^2} \sum_{t=1}^n \frac{\mu_{t-1} (\mu_t - \phi \mu_{t-1})}{1 + \tau \mu_{t-1}^2}$$

En résolvant $\frac{\partial g}{\partial \phi} = 0$, on obtient l'estimateur :

$$\hat{\phi} = \left[\frac{1}{n} \sum_{t=1}^n \frac{\mu_t \mu_{t-1}}{1 + \tau \mu_{t-1}^2} \right] \left[\frac{1}{n} \sum_{t=1}^n \frac{\mu_{t-1}^2}{1 + \tau \mu_{t-1}^2} \right]^{-1}$$

De la même manière, en dérivant g par rapport à σ_ε^2 :

$$\frac{\partial g}{\partial \sigma_\varepsilon^2} = \frac{1}{\sigma_\varepsilon^2} - \frac{1}{n (\sigma_\varepsilon^2)^2} \sum_{t=1}^n \frac{(\mu_t - \phi \mu_{t-1})^2}{1 + \tau \mu_{t-1}^2}$$

ce qui mène à :

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n} \sum_{t=1}^n \frac{(\mu_t - \phi \mu_{t-1})^2}{1 + \tau \mu_{t-1}^2}$$

Enfin, on considère la fonction minimisée :

$$\bar{g}(\phi, \tau, \sigma_\varepsilon^2) = \ln \left(\frac{1}{n} \sum_{t=1}^n \frac{(\mu_t - \phi \mu_{t-1})^2}{1 + \tau \mu_{t-1}^2} \right) + \frac{1}{n} \sum_{t=1}^n \ln(1 + \tau \mu_{t-1}^2)$$

L'estimation finale de τ se fait par minimisation de $\bar{g}(\hat{\phi}, \tau, \sigma_\varepsilon^2)$, et on en déduit :

$$\hat{\sigma}_B^2 = \hat{\sigma}_\varepsilon^2 \cdot \hat{\tau}$$

2.6.4 Consistance et normalité asymptotique de l'estimateur du maximum de vraisemblance

Soit $\{\mu_t, t \in \mathbb{Z}\}$ une solution de (2.2) telle que $\mathbb{E}(|\phi + B(1)|) < \infty$, alors l'estimateur du maximum de vraisemblance, noté $\hat{\phi}_{mv}$, est faiblement consistant pour ϕ :

$$\hat{\phi}_{mv} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \phi$$

À présent, nous présentons un autre résultat de **Aue et Horvath** [4] qui établit la normalité asymptotique de $\hat{\phi}_{mv}$. Nous commençons par la première composante, $\hat{\phi}_{mv,1}$.

Théorème 2.6.2 (Aue et Horvath [4]). *Soit μ_t une solution de (??) telle que $\mathbb{E}(\log^+ |\phi + B(1)|) < \infty$, alors :*

$$\sqrt{n}(\hat{\phi}_{mv,1} - \phi) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \theta^2)$$

où :

$$\theta^2 = \begin{cases} \sigma_B^2 & \text{si } \mathbb{E}(\log |\phi + B(1)|) \geq 0 \\ \frac{4}{\alpha^2(2, 1)} [\sigma_\beta^2 \alpha(4, 2) + \sigma_\varepsilon^2 \alpha(2, 2)] & \text{sinon} \end{cases}$$

avec :

$$\alpha(k, N) = \mathbb{E} \left(\frac{\mu_1^k}{(\sigma_B^2 \mu_1^2 + \sigma_\varepsilon^2)^N} \right), \quad k = 0, 1, 2, \dots, 2N, \quad N \in \mathbb{N}$$

À partir du théorème (2.6.2), il est important de noter que la normalité asymptotique de $\hat{\phi}_{mv,1} - \phi$ est valable sans restriction sur la stationnarité du modèle (2.1).

On remarque également que la variance asymptotique θ^2 diffère selon que le modèle soit stationnaire ou non.

Ainsi, **Aue et Horvath** [4] ont proposé un estimateur de θ^2 , noté $\hat{\theta}^2$, qui fonctionne dans les deux cas (stationnaire et non stationnaire) :

$$\hat{\alpha}(k, N) = \frac{1}{n} \sum_{t=1}^n \frac{\mu_t^k}{(\hat{\phi}_{mv,2} \mu_t^2 + \hat{\phi}_{mv,3})^N}$$

Si toutes les conditions du théorème sont vérifiées, alors on peut conclure :

$$\sqrt{n} \left(\frac{\hat{\phi}_{\text{mv},1} - \phi}{\hat{\theta}} \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

Chapitre 3

Modèle de régression linéaire à erreurs

RCA(1)

3.1 Introduction

Dans ce chapitre, nous nous intéressons à l'étude d'un modèle de régression linéaire dans lequel le terme d'erreur suit un modèle autorégressif à coefficient aléatoire RCA(1). Ce type de structure introduit une hétéroscédasticité conditionnelle aléatoire dans les résidus, ce qui rend l'analyse statistique plus complexe que dans le cadre classique.

Ce chapitre est divisé en deux parties. Dans la première partie, nous présentons quelques généralités sur les modèles de régression non linéaires, en mettant l'accent sur leurs propriétés, leurs applications et les difficultés inhérentes à leur estimation. Nous y introduisons également les notations et les hypothèses utilisées tout au long du chapitre.

Dans la seconde partie, nous abordons la problématique de l'estimation des paramètres du modèle à l'aide de la méthode des moindres carrés non linéaires. Nous décrivons les étapes de l'algorithme itératif utilisé pour obtenir les estimateurs des coefficients du modèle.

3.2 Rappel sur les modèles de régression non linéaire

L'objet de la régression est de modéliser et d'étudier la relation entre une variable dite réponse et une ou plusieurs autres variables dites explicatives. Les méthodes linéaires

peuvent ne pas donner des résultats satisfaisants et, il faudra donc faire appel à la théorie des modèles de régression non linéaire. La plupart des résultats, en régression paramétrique non linéaire sont de nature asymptotique. D'autre part, une des difficultés majeures des problèmes de régression non linéaires provient de la nécessité de résoudre des problèmes d'optimisation et les algorithmes ne sont pas toujours convergents.

Exemple 1 :

La fonction de production de type Cobb-Douglas est donnée par

$$Q = AK^\beta L^\gamma \tag{3.1}$$

où Q est la quantité produite, K est le capital, L le travail et A, β et γ les paramètres. D'un point de vue de la théorie économique, (3.1) a un nombre de propriétés intéressantes et elle est choisie comme base dans des travaux appliqués. En particulier, β et γ sont les élasticités de l'output, par rapport au capital et au travail respectivement.

Supposons que l'on veuille estimer les paramètres A, β et γ sur la base d'un ensemble d'observations sur Q, K et L . Une formulation stochastique de la relation exacte (3.1) peut être établie. Par exemple :

$$Q_t = AK_t^\beta L_t^\gamma \exp(\epsilon_t), \quad t = 1, \dots, T \tag{3.2}$$

où ϵ_t est la perturbation aléatoire. Ce modèle est intrinsèquement linéaire. En effet, si on applique le logarithme naturel aux deux membres de (3.2), on obtient :

$$\log Q_t = \log A + \beta \log K_t + \gamma \log L_t + \epsilon_t, \quad t = 1, \dots, T \tag{3.3}$$

on note ! $\tau = \log A$. Le modèle n'est pas linéaire en les variables, mais il est linéaire en les paramètres et s'écrit donc :

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \epsilon_t, \quad t = 1, \dots, T \tag{3.4}$$

où $\beta_1 = \tau, x_{1t} = 1, \beta_2 = \beta, x_{2t} = \log K_t, \beta_3 = \gamma, x_{3t} = \log L_t$. L'estimation des paramètres inconnus peut être menée grâce aux techniques connues de régression linéaire.

Donc, l'hypothèse de linéarité peut être moins restrictive qu'il n'apparaît au premier regard.

Mais supposons qu'une information a priori nous impose une perturbation aléatoire additive dans (3.1), i.e.,

$$Q_t = AK_t^\beta L_t^\gamma + \varepsilon_t, \quad t = 1, \dots, T \quad (3.5)$$

On ne peut, en aucune manière, transformer (3.5) en un modèle linéaire et dans ce cas, l'estimation des paramètres peut être menée en utilisant des techniques de régression non linéaires.

Exemple 2

La fonction de production à élasticité de substitution constante a une forme plus générale que (3.1). Elle est spécifiée par la relation fonctionnelle

$$Q = \exp(\tau) \left\{ (1 - \delta)L^{-\gamma} + \delta K^{-\beta} \right\}^{-\frac{\tau}{\gamma}} \quad (3.6)$$

où τ, β, γ et δ sont des paramètres. Ce modèle à estimer est non linéaire aussi bien lorsque la perturbation aléatoire est additive ou multiplicative.

On remarque, cependant, qu'une perturbation multiplicative est plus simple à prendre en charge. En effet, si on prend le logarithme des deux membres et sur la base d'un ensemble d'observations, on obtient :

$$\log Q_t = \tau - \frac{\beta}{\gamma} \log \left\{ (1 - \delta)L_t^{-\gamma} + \delta K_t^{-\beta} \right\} + \varepsilon_t, \quad t = 1, \dots, T \quad (3.7)$$

Le modèle est encore non linéaire par rapport à tous les paramètres, mais pour γ et δ fixés, il est linéaire par rapport à τ et β .

3.2.1 Modèle de régression non linéaire

La formulation générale d'un modèle de régression non linéaire est comme suit :

$$y_t = h(\underline{x}_t, \underline{\beta}) + \varepsilon_t, \quad t = 1, \dots, T \quad (3.8)$$

où \underline{x}_t est le vecteur à k composantes (ci-après, vecteur $k \times 1$) des variables exogènes et $\underline{\beta}$ est un vecteur $n \times 1$ des paramètres du modèle et h une fonction de \underline{x}_t et de $\underline{\beta}$.

Les hypothèses sont les suivantes :

1. Le paramètre inconnu $\beta \in \Theta^\circ \subset \mathbb{R}^k, k \geq 1$, où Θ° est l'intérieur de l'ensemble Θ supposé non vide, h est deux fois continûment différentiable et $\mathbb{E}(y_t | \underline{x}_t) = h(\underline{x}_t; \underline{\beta}), t = 1, \dots, T$.
2. Les paramètres du modèle sont identifiables, i.e., si $\underline{\beta}_1 \neq \underline{\beta}_2$ alors $h(\underline{x}_t, \underline{\beta}_1) \neq h(\underline{x}_t, \underline{\beta}_2), \forall \underline{x}_t, t = 1, \dots, T$. La fonction h est une fonction injective du paramètre $\underline{\beta}, \forall \underline{x}_t, t = 1, \dots, T$ (Cette hypothèse correspond à la condition de plein

rang pour la matrice du plan d'expérience) $X = \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \vdots \\ \underline{x}_T \end{bmatrix}$.

3. Le modèle (3.8) et l'hypothèse 1 impliquent que les perturbations $\varepsilon_t, t = 1, \dots, T$ sont telles que $\mathbb{E}(\varepsilon_t | h(\underline{x}_t; \underline{\beta})) = 0, t = 1, \dots, T$. La t -ème perturbation est non corrélée avec la moyenne conditionnelle.
4. **Homoscédasticité** On suppose l'homoscédasticité conditionnelle :

$$\mathbb{E}(\varepsilon_t^2 | h(\underline{x}_t; \underline{\beta})) = \sigma^2 \quad t = 1, \dots, T \text{ (constante finie)} \quad (3.9)$$

5. **Absence d'autocorrélation**

$$\mathbb{E}(\varepsilon_t \varepsilon_s | h(\underline{x}_t; \underline{\beta}), h(\underline{x}_s; \underline{\beta})) = 0 \quad s = 1, \dots, T \quad (3.10)$$

Ainsi, le modèle (3.8) s'écrit de façon plus compacte de la manière suivante :

$$\underline{Y} = \eta(\underline{\beta}) + \underline{\varepsilon} \quad (3.11)$$

où $\underline{Y} = [y_1, y_2, \dots, y_T]', \eta(\underline{\beta}) = [h(\underline{x}_1; \underline{\beta}), \dots, h(\underline{x}_T; \underline{\beta})]'$ et $\underline{\varepsilon} = [\varepsilon_1, \dots, \varepsilon_T]'$.

3.2.2 Moindres carrés non linéaires

Pour un modèle de régression non linéaire, l'ensemble des valeurs possibles pour la moyenne de \underline{Y} , est la surface $\mathcal{M} = \{\eta(\underline{\beta}), \underline{\beta} \in \Theta\} \subset \mathbb{R}^T$ des observations.

Remarque. Dans les modèles linéaires, cette surface est l'espace vectoriel $\mathcal{M}(X)$ engendré par les colonnes de la matrice X du plan d'expérience. L'estimateur est unique et s'obtient en projetant \underline{Y} sur $\mathcal{M}(X)$.

Dans les modèles non linéaires, cette surface est une variété différentiable de l'espace \mathbb{R}^T . Le critère des moindres carrés consiste à choisir, pour estimer la moyenne de \underline{Y} (et donc aussi le paramètre), un point de la surface \mathcal{M} le plus proche possible, au sens de la norme euclidienne, du point \underline{Y} .

Définition 3.2.1. Un estimateur, au sens des moindres carrés de $\underline{\beta}$ lorsqu'il existe, est un vecteur aléatoire dont la réalisation $\hat{\underline{\beta}}$ minimise la fonction :

$$S(\underline{\beta}) = \|\underline{Y} - \eta(\underline{\beta})\|^2 = (\underline{Y} - \eta(\underline{\beta}))'(\underline{Y} - \eta(\underline{\beta})) \quad (3.12)$$

Il s'agit de minimiser la somme des résidus au carré.

Remarque. Cet estimateur existe si l'espace des paramètres Θ est fermé et borné.

En développant $S(\underline{\beta})$, on obtient :

$$S(\underline{\beta}) = \underline{Y}'\underline{Y} - 2\underline{Y}'\eta(\underline{\beta}) + \eta(\underline{\beta})'\eta(\underline{\beta})$$

Les équations normales sont données par :

$$\frac{\partial S(\underline{\beta})}{\partial \underline{\beta}} = 0$$

C'est-à-dire :

$$\left(\underline{Y} - \eta(\underline{\beta})\right)' \left(\frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}}\right)_{\underline{\beta}=\hat{\underline{\beta}}} = 0_{\mathbb{R}^k} \quad (3.13)$$

où $\frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}}$ est une matrice de taille $T \times k$.

Remarque. Pour un modèle de régression linéaire, $\frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}} = X$ est indépendante de $\underline{\beta}$. En revanche, pour un modèle non linéaire, $\frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}}$ dépend de $\underline{\beta}$. Ainsi, les équations normales (3.13) ne possèdent pas de solution analytique explicite comme dans le cas linéaire. Il est alors nécessaire de recourir à des méthodes numériques pour les résoudre.

3.2.3 Méthodes de calcul de l'estimateur des moindres carrés

Méthode de Gauss-Newton

La méthode de Gauss-Newton est une procédure itérative permettant de résoudre les équations normales dans le cas non linéaire. Elle repose uniquement sur la dérivée première de la fonction $\eta(\underline{\beta})$.

Soit $\hat{\underline{\beta}}_0$ une valeur initiale. On effectue un développement de Taylor de $\eta(\underline{\beta})$ au voisinage de $\hat{\underline{\beta}}_0$:

$$\eta(\underline{\beta}) \approx \eta(\hat{\underline{\beta}}_0) + \left. \frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}} \right|_{\underline{\beta}=\hat{\underline{\beta}}_0} (\underline{\beta} - \hat{\underline{\beta}}_0) \quad (3.14)$$

En remplaçant cette approximation dans l'équation du modèle (3.11) on obtient :

$$\underline{Y} = \eta(\hat{\underline{\beta}}_0) + \left. \frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}} \right|_{\hat{\underline{\beta}}_0} (\underline{\beta} - \hat{\underline{\beta}}_0) + \varepsilon^0 \quad (3.15)$$

Ce qui donne :

$$\underline{Y} - \eta(\hat{\underline{\beta}}_0) + \left. \frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}} \right|_{\hat{\underline{\beta}}_0} \hat{\underline{\beta}}_0 = \left. \frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}} \right|_{\hat{\underline{\beta}}_0} \underline{\beta} + \varepsilon^0 \quad (3.16)$$

ou encore :

$$\underline{Y}^0 = \left. \frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}} \right|_{\hat{\underline{\beta}}_0} \underline{\beta} + \varepsilon^0 \quad (3.17)$$

avec $\underline{Y}^0 = \underline{Y} - \eta(\hat{\underline{\beta}}_0) + \left. \frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}} \right|_{\hat{\underline{\beta}}_0} \hat{\underline{\beta}}_0$, qui est une fonction des données et non du paramètre inconnu et ε^0 contient à la fois la vraie perturbation $\underline{\varepsilon}$ et l'erreur de l'approximation de Taylor à l'ordre 1. Il en résulte une égalité donnée par (3.18) au lieu d'une approximation.

On appelle (3.18) le modèle linéarisé de régression.

On estime $\underline{\beta}$ par la méthode des moindres carrés linéaires. On obtient

$$\hat{\underline{\beta}}_1 = \left[\left(\frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}} \Big|_{\hat{\underline{\beta}}_0} \right)^T \left(\frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}} \Big|_{\hat{\underline{\beta}}_0} \right) \right]^{-1} \left(\frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}} \Big|_{\hat{\underline{\beta}}_0} \right)^T \underline{Y}^0 \quad (3.18)$$

C'est l'estimateur de la première étape de l'itération. \underline{Y}^0 est la nouvelle variable dépendante et $\frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}} \Big|_{\hat{\underline{\beta}}_0}$ le vecteur des régresseurs. Nous retournons maintenant en (3.16) où $\hat{\underline{\beta}}_0$ est remplacé par $\hat{\underline{\beta}}_1$, puis en (3.15) et enfin (3.18) qui sera remplacée par

$$\underline{Y}^1 = \frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}} \Big|_{\hat{\underline{\beta}}_1} \underline{\beta} + \varepsilon^1 \quad (3.19)$$

L'estimateur de la deuxième étape sera noté $\hat{\underline{\beta}}_2$.

L'équation (3.18) s'écrit encore

$$\hat{\underline{\beta}}_2 = \left[\left(\frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}} \Big|_{\hat{\underline{\beta}}_1} \right)^T \left(\frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}} \Big|_{\hat{\underline{\beta}}_1} \right) \right]^{-1} \left(\frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}} \Big|_{\hat{\underline{\beta}}_1} \right)^T (\underline{Y} - \eta(\hat{\underline{\beta}}_1) + \frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}} \Big|_{\hat{\underline{\beta}}_1} \hat{\underline{\beta}}_1) \quad (3.20)$$

$$\hat{\underline{\beta}}_2 = \hat{\underline{\beta}}_1 + \left[\left(\frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}} \Big|_{\hat{\underline{\beta}}_1} \right)^T \left(\frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}} \Big|_{\hat{\underline{\beta}}_1} \right) \right]^{-1} \left(\frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}} \Big|_{\hat{\underline{\beta}}_1} \right)^T (\underline{Y} - \eta(\hat{\underline{\beta}}_1)) \quad (3.21)$$

De façon générale, l'algorithme itératif est le suivant

$$\hat{\underline{\beta}}_{i+1} = \hat{\underline{\beta}}_i + \left[\left(\frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}} \Big|_{\hat{\underline{\beta}}_i} \right)^T \left(\frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}} \Big|_{\hat{\underline{\beta}}_i} \right) \right]^{-1} \left(\frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}} \Big|_{\hat{\underline{\beta}}_i} \right)^T (\underline{Y} - \eta(\hat{\underline{\beta}}_i)) \quad (3.22)$$

Exemple

Supposons que $T = 2$, $k = 1$, et que le modèle soit :

$$y_t = h(\underline{x}_t, \underline{\beta}) + \varepsilon_t,$$

avec $\underline{x}_1 = 1$, $\underline{x}_2 = 2$, et $h(\underline{x}_t, \underline{\beta}) = e^{\underline{\beta} \cdot \underline{x}_t}$.

Données observées : $y_1 = 1$, $y_2 = 8$

On a :

$$\eta(\underline{\beta}) = \begin{pmatrix} e^{\underline{\beta}} \\ e^{2\underline{\beta}} \end{pmatrix} \quad \text{et} \quad Y = \begin{pmatrix} 1 \\ 8 \end{pmatrix}$$

La fonction des moindres carrés s'écrit :

$$S(\underline{\beta}) = (1 - e^{\underline{\beta}})^2 + (8 - e^{2\underline{\beta}})^2$$

Annulons la dérivée :

$$\frac{dS(\underline{\beta})}{d\underline{\beta}} = -2(1 - e^{\underline{\beta}})e^{\underline{\beta}} - 4(8 - e^{2\underline{\beta}})e^{2\underline{\beta}} = 0$$

Une solution numérique donne :

$$\hat{\underline{\beta}} \approx \log(2,7714) \approx 1,01934$$

Développement de Taylor au premier ordre

Autour de $\underline{\beta}_0 = 0$, on a :

$$\eta(\underline{\beta}) \approx \eta(\underline{\beta}_0) + \left. \frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}} \right|_{\underline{\beta}_0} (\underline{\beta} - \underline{\beta}_0)$$

$$\eta(\underline{\beta}_0) = \begin{pmatrix} e^0 \\ e^0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \left. \frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}} \right|_{\underline{\beta}_0} = \begin{pmatrix} 1 \cdot e^0 \\ 2 \cdot e^0 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Donc :

$$\eta(\underline{\beta}) \approx \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \underline{\beta} \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Première itération de Gauss-Newton

$$\text{Jacobienne } J = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \text{résidu } r = Y - \eta(\underline{\beta}_0) = \begin{pmatrix} 0 \\ 7 \end{pmatrix}$$

$$J^T J = 1^2 + 2^2 = 5, \quad J^T r = 1 \cdot 0 + 2 \cdot 7 = 14$$

$$\hat{\underline{\beta}}_1 = \underline{\beta}_0 + (J^T J)^{-1} J^T r = 0 + \frac{14}{5} = 2,8$$

Développement quadratique de $S(\underline{\beta})$

On écrit une approximation quadratique de $S(\underline{\beta})$ autour de $\hat{\underline{\beta}}_1$:

$$S(\underline{\beta}) \approx \left\| Y - \eta(\hat{\underline{\beta}}_1) - \left. \frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}} \right|_{\hat{\underline{\beta}}_1} (\underline{\beta} - \hat{\underline{\beta}}_1) \right\|^2 = a + b^T \underline{\beta} + \frac{1}{2} \underline{\beta}^T H \underline{\beta}$$

où :

- $r = Y - \eta(\hat{\underline{\beta}}_1)$
- $J = \left. \frac{\partial \eta(\underline{\beta})}{\partial \underline{\beta}} \right|_{\hat{\underline{\beta}}_1}$
- $a = r^T r$: terme constant
- $b = -2J^T r$: gradient
- $H = 2J^T J$: hessienne approchée

La solution optimale est alors donnée par :

$$\frac{\partial S(\underline{\beta})}{\partial \underline{\beta}} = b + H \hat{\underline{\beta}} = 0 \quad \Rightarrow \quad \hat{\underline{\beta}} = -H^{-1} b$$

Si H est définie positive, alors ce point est un minimum local unique.

La méthode de Newton-Raphson

La méthode de Gauss-Newton repose sur une approximation linéaire à l'ordre 1 de la fonction de régression $\eta(\underline{\beta})$. La méthode de Newton-Raphson est une méthode de descente qui repose sur l'approximation de Taylor à l'ordre 2 du gradient $\frac{\partial S(\underline{\beta})}{\partial \underline{\beta}}$. Cela revient à développer la fonction de régression à l'ordre 2.

On pose donc :

$$\frac{\partial S(\underline{\beta})}{\partial \underline{\beta}} \approx \frac{\partial S(\hat{\underline{\beta}}_i)}{\partial \underline{\beta}} + \frac{\partial^2 S(\hat{\underline{\beta}}_i)}{\partial \underline{\beta} \partial \underline{\beta}'} (\underline{\beta} - \hat{\underline{\beta}}_i)$$

Si $\hat{\underline{\beta}}$ est la solution du problème de minimisation, alors

$$\frac{\partial S(\hat{\underline{\beta}})}{\partial \underline{\beta}} = 0_k$$

et on a approximativement :

$$\frac{\partial S(\hat{\underline{\beta}}_i)}{\partial \underline{\beta}} \approx \frac{\partial^2 S(\hat{\underline{\beta}}_i)}{\partial \underline{\beta} \partial \underline{\beta}'} (\underline{\beta} - \hat{\underline{\beta}}_i)$$

On en déduit l'algorithme itératif :

$$\hat{\underline{\beta}}_{i+1} = \hat{\underline{\beta}}_i - \left(\frac{\partial^2 S(\hat{\underline{\beta}}_i)}{\partial \underline{\beta} \partial \underline{\beta}'} \right)^{-1} \frac{\partial S(\hat{\underline{\beta}}_i)}{\partial \underline{\beta}}$$

Remarque. $\frac{\partial^2 S(\hat{\underline{\beta}}_i)}{\partial \underline{\beta} \partial \underline{\beta}'}$ est appelée la **matrice hessienne**, notée $\text{Hess}(S(\hat{\underline{\beta}}_i))$.

La méthode des scores

Cette méthode consiste à remplacer, dans l'algorithme de Newton-Raphson, $\text{Hess}(S(\hat{\underline{\beta}}_t))$ par son espérance mathématique.

Remarque. *Comme tout algorithme itératif, ces algorithmes sont sensibles à l'initialisation. Il n'y a pas de règle pour choisir cette initialisation. Un mauvais choix de l'initialisation peut conduire à la divergence de l'algorithme. Lorsque l'algorithme converge, le minimum obtenu peut être local et différents choix de l'initialisation peuvent conduire à des minima différents. Lorsque la convexité stricte de $S(\underline{\beta})$ n'est pas prouvée, on applique l'algorithme avec plusieurs initialisations différentes et on choisit pour minimum le point le plus souvent atteint.*

3.3 Estimation par la méthode des moindres carrés ordinaires dans un modèle de régression linéaire à erreurs RCA(1)

Nicholls et Quinn[24] ont développé une procédure d'estimation fondée sur la méthode des moindres carrés pour les processus autorégressifs à coefficients aléatoires univariés, laquelle, sous certaines conditions, donne des estimateurs fortement consistants pour les paramètres.

3.3.1 Présentation du modèle

Nous considérons le modèle linéaire simple où les erreurs μ_i suivent un RCA(1) :

$$y_t = \beta x_t + \mu_t, \quad t = 1, \dots, n \quad (3.23)$$

où $\mu_t = (\phi + B(t))\mu_{t-1} + \varepsilon_t$

avec $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ et $B(t) \sim \mathcal{N}(0, \sigma_B^2)$.

On suppose que $B(t)$ est indépendant de ε_t pour tout $t = 1, \dots, n$ et les processus $\{\varepsilon_t\}$ et $\{B(t)\}$ sont i.i.d.

Le modèle (3.23) peut se réécrire de la manière suivante :

$$y_t = \beta x_t + (\phi + B(t))(y_{t-1} - \beta x_{t-1}) + \varepsilon_t, \quad t = 1, \dots, n$$

ou bien

$$y_t = \beta x_t + \phi(y_{t-1} - \beta x_{t-1}) + B(t)(y_{t-1} - \beta x_{t-1}) + \varepsilon_t, \quad t = 1, \dots, n$$

Posons $\Gamma(t) = B(t)(y_{t-1} - \beta x_{t-1}) + \varepsilon_t$, on obtient :

$$y_t = \beta x_t + \phi(y_{t-1} - \beta x_{t-1}) + \Gamma(t), \quad t = 1, \dots, n$$

Soit \mathcal{F}_t la σ -algèbre générée par $\{\varepsilon_t, B(t)\}, \{\varepsilon_{t-1}, B(t-1), \dots\}$

Nous avons :

$$\begin{aligned}\mathbb{E}[\Gamma_t | \mathcal{F}_{t-1}] &= \mathbb{E}[B(t)(y_{t-1} - \beta x_{t-1}) + \varepsilon_t | \mathcal{F}_{t-1}] \\ &= \mathbb{E}[B(t)(y_{t-1} - \beta x_{t-1}) | \mathcal{F}_{t-1}] + \mathbb{E}[\varepsilon_t | \mathcal{F}_{t-1}] \\ &= \mathbb{E}[B(t) | \mathcal{F}_{t-1}](y_{t-1} - \beta x_{t-1}) + \mathbb{E}[\varepsilon_t | \mathcal{F}_{t-1}] \\ &= 0\end{aligned}$$

C'est donc une différence de martingales, puisque $\{B(t)\}$ et $\{\varepsilon_t\}$ sont indépendants de $\{(\varepsilon_{t-1}, B(t-1)), (\varepsilon_{t-2}, B(t-2)), \dots\}$ et $(y_{t-1} - \beta x_{t-1})$ est une fonction mesurable.

Calcul de la variance conditionnelle :

$$\begin{aligned}\mathbb{E}[\Gamma_t^2 | \mathcal{F}_{t-1}] &= \mathbb{E}[(B(t)(y_{t-1} - \beta x_{t-1}) + \varepsilon_t)^2 | \mathcal{F}_{t-1}] \\ &= \mathbb{E}[\varepsilon_t^2 + (B(t)(y_{t-1} - \beta x_{t-1}))'(B(t)(y_{t-1} - \beta x_{t-1})) + 2\varepsilon_t B(t)(y_{t-1} - \beta x_{t-1}) | \mathcal{F}_{t-1}] \\ &= \mathbb{E}[\varepsilon_t^2 | \mathcal{F}_{t-1}] + \mathbb{E}[(y_{t-1} - \beta x_{t-1})' B(t)' B(t) (y_{t-1} - \beta x_{t-1}) | \mathcal{F}_{t-1}] \\ &\quad + 2 \underbrace{\mathbb{E}[\varepsilon_t | \mathcal{F}_{t-1}]}_{=0} \underbrace{\mathbb{E}[B(t) | \mathcal{F}_{t-1}]}_{=0} (y_{t-1} - \beta x_{t-1}) \\ &= \sigma_\varepsilon^2 + \mathbb{E}[B(t)' B(t) | \mathcal{F}_{t-1}] (y_{t-1} - \beta x_{t-1})^2 + 0 \\ &= \sigma_\varepsilon^2 + \sigma_B^2 (y_{t-1} - \beta x_{t-1})^2\end{aligned}$$

3.3.2 La procédure d'estimation (Méthode de Gauss-Newton)

La procédure d'estimation est une généralisation d'une procédure en deux étapes, proposée par Rosenberg [25], pour estimer les paramètres d'un modèle de régression à coefficients aléatoires.

La première étape consiste à estimer les paramètres β et ϕ du modèle (3.23), en minimisant :

$$\begin{aligned} S(\beta, \phi) &= \sum_{t=1}^n [y_t - \mathbb{E}(y_t | \mathcal{F}_{t-1})]^2 \\ &= \sum_{t=1}^n [y_t - \beta x_t - \phi(y_{t-1} - \beta x_{t-1})]^2 \end{aligned}$$

On pose $\theta = (\beta, \phi)'$, on obtient la forme matricielle du modèle (3.23) :

$$\underline{y} = f(\underline{\theta}) + \underline{\Gamma} \quad (3.24)$$

où $f(\underline{\theta})$ est défini par :

$$f(\underline{\theta}) = \beta x_t + \phi(y_{t-1} - \beta x_{t-1}), \quad t = 1, \dots, n$$

Soit $\hat{\underline{\theta}}^{(0)}$ une valeur initiale de $\hat{\underline{\theta}}$. On donne un développement de Taylor à l'ordre 1 de $f(\underline{\theta})$ au voisinage de $\hat{\underline{\theta}}^{(0)}$, soit :

$$f(\underline{\theta}) \approx f(\hat{\underline{\theta}}^{(0)}) + \left. \frac{\partial f(\underline{\theta})}{\partial \underline{\theta}} \right|_{\underline{\theta}=\hat{\underline{\theta}}^{(0)}} (\underline{\theta} - \hat{\underline{\theta}}^{(0)}) \quad (3.25)$$

où

$$f(\hat{\underline{\theta}}^{(0)}) = \hat{\beta}^{(0)} x_t + \hat{\phi}^{(0)}(y_{t-1} - \hat{\beta}^{(0)} x_{t-1}), \quad t = 1, \dots, n$$

En remplaçant (3.25) dans (3.24), on obtient :

$$\underline{y} = f(\hat{\underline{\theta}}^{(0)}) + \left. \frac{\partial f(\underline{\theta})}{\partial \underline{\theta}} \right|_{\underline{\theta}=\hat{\underline{\theta}}^{(0)}} (\underline{\theta} - \hat{\underline{\theta}}^{(0)}) + \underline{\Gamma}_{(0)} \quad (3.26)$$

Le modèle (3.26) s'écrit :

$$\underline{y} - f(\hat{\underline{\theta}}^{(0)}) + \left. \frac{\partial f(\underline{\theta})}{\partial \underline{\theta}} \right|_{\hat{\underline{\theta}}^{(0)}} \hat{\underline{\theta}}^{(0)} = \left. \frac{\partial f(\underline{\theta})}{\partial \underline{\theta}} \right|_{\hat{\underline{\theta}}^{(0)}} \underline{\theta} + \underline{\Gamma}_{(0)}$$

ou encore :

$$\underline{y}_{(0)} = \left. \frac{\partial f(\underline{\theta})}{\partial \underline{\theta}} \right|_{\underline{\theta}=\hat{\underline{\theta}}^{(0)}} \underline{\theta} + \underline{\Gamma}_{(0)} \quad (3.27)$$

avec :

$$\underline{y}_{(0)} = \underline{y} - f(\hat{\underline{\theta}}^{(0)}) + \left. \frac{\partial f(\underline{\theta})}{\partial \underline{\theta}} \right|_{\underline{\theta}=\hat{\underline{\theta}}^{(0)}} \hat{\underline{\theta}}^{(0)}$$

qui est une fonction des données et non un paramètre inconnu. $\underline{\Gamma}_{(0)}$ contient à la fois la vraie perturbation $\underline{\Gamma}$ et l'erreur de l'approximation de Taylor à l'ordre 1. On appelle (3.27) le modèle linéarisé de régression.

On estime $\hat{\underline{\theta}}_{(1)}$ par la méthode des moindres carrés non linéaires. On obtient :

$$\hat{\underline{\theta}}_{(1)} = \left[\left(\left. \frac{\partial f(\underline{\theta})}{\partial \underline{\theta}} \right|_{\hat{\underline{\theta}}^{(0)}} \right)^\top \left(\left. \frac{\partial f(\underline{\theta})}{\partial \underline{\theta}} \right|_{\hat{\underline{\theta}}^{(0)}} \right) \right]^{-1} \left(\left. \frac{\partial f(\underline{\theta})}{\partial \underline{\theta}} \right|_{\hat{\underline{\theta}}^{(0)}} \right)^\top \underline{y}_{(0)} \quad (3.28)$$

Il s'agit de l'estimateur de la première étape de l'itération. Ici, $\underline{y}_{(0)}$ est la nouvelle variable dépendante, et $\left. \frac{\partial f(\underline{\theta})}{\partial \underline{\theta}} \right|_{\hat{\underline{\theta}}^{(0)}}$ représente le vecteur des régresseurs.

On retourne maintenant à l'équation (3.25), où $\hat{\underline{\theta}}^{(0)}$ est remplacé par $\hat{\underline{\theta}}^{(1)}$, puis à (3.26), et enfin à (3.27), qui devient :

$$\underline{y}_{(1)} = \left. \frac{\partial f(\underline{\theta})}{\partial \underline{\theta}} \right|_{\hat{\underline{\theta}}^{(1)}} \hat{\underline{\theta}} + \underline{\Gamma}_{(1)} \quad (3.29)$$

L'estimateur de la deuxième itération est noté $\hat{\underline{\theta}}_{(2)}$, donné par la formule :

$$\begin{aligned} \hat{\underline{\theta}}_{(2)} &= \left[\left(\frac{\partial f(\underline{\theta})}{\partial \underline{\theta}} \Big|_{\hat{\underline{\theta}}_{(1)}} \right)^\top \left(\frac{\partial f(\underline{\theta})}{\partial \underline{\theta}} \Big|_{\hat{\underline{\theta}}_{(1)}} \right) \right]^{-1} \left(\frac{\partial f(\underline{\theta})}{\partial \underline{\theta}} \Big|_{\hat{\underline{\theta}}_{(1)}} \right)^\top \\ &\quad \times \left(\underline{y} - \underline{f}(\hat{\underline{\theta}}_{(1)}) + \frac{\partial f(\underline{\theta})}{\partial \underline{\theta}} \Big|_{\hat{\underline{\theta}}_{(1)}} \hat{\underline{\theta}}_{(1)} \right) \end{aligned} \quad (3.30)$$

Ce qui se simplifie :

$$\hat{\underline{\theta}}_{(2)} = \hat{\underline{\theta}}_{(1)} + \left[\left(\frac{\partial f(\hat{\underline{\theta}})}{\partial \hat{\underline{\theta}}} \Big|_{\hat{\underline{\theta}}_{(1)}} \right)^\top \left(\frac{\partial f(\hat{\underline{\theta}})}{\partial \hat{\underline{\theta}}} \Big|_{\hat{\underline{\theta}}_{(1)}} \right) \right]^{-1} \left(\frac{\partial f(\hat{\underline{\theta}})}{\partial \hat{\underline{\theta}}} \Big|_{\hat{\underline{\theta}}_{(1)}} \right)^\top (y - f(\hat{\underline{\theta}}_{(1)})) \quad (3.31)$$

De manière générale, l'algorithme itératif de Gauss-Newton s'écrit comme suit :

$$\hat{\underline{\theta}}_{(i+1)} = \hat{\underline{\theta}}_{(i)} + \left[\left(\frac{\partial f(\hat{\underline{\theta}})}{\partial \hat{\underline{\theta}}} \Big|_{\hat{\underline{\theta}}_{(i)}} \right)^\top \left(\frac{\partial f(\hat{\underline{\theta}})}{\partial \hat{\underline{\theta}}} \Big|_{\hat{\underline{\theta}}_{(i)}} \right) \right]^{-1} \left(\frac{\partial f(\hat{\underline{\theta}})}{\partial \hat{\underline{\theta}}} \Big|_{\hat{\underline{\theta}}_{(i)}} \right)^\top (y - f(\hat{\underline{\theta}}_{(i)})) \quad (3.32)$$

Les dérivées partielles premières sont données par :

$$\frac{\partial f(\hat{\underline{\theta}})}{\partial \beta} = x_t - \hat{\phi}x_{t-1}, \quad t = 1, \dots, \quad \frac{\partial f(\hat{\underline{\theta}})}{\partial \phi} = y_{t-1} - \hat{\beta}x_{t-1}, \quad t = 1, \dots, n$$

La deuxième étape consiste à poser :

$$R(t) = \hat{\Gamma}^2(t) - \sigma_\varepsilon^2 - \sigma_B^2(y_{t-1} - \hat{\beta}x_{t-1})^2$$

avec $\hat{\Gamma}(t) = y_t - \hat{\beta}x_t - \hat{\phi}(y_{t-1} - \hat{\beta}x_{t-1})$ pour $t = 1, \dots, n$ (résidus).

Les estimateurs de σ_ε^2 et σ_B^2 sont obtenus en minimisant :

$$\sum_{t=1}^n R(t)^2 \quad \text{par rapport à } \sigma_\varepsilon^2 \text{ et } \sigma_B^2, \quad \text{c'est-à-dire en régressant } \hat{\Gamma}^2(t) \text{ sur } 1 \text{ et } (y_{t-1} - \hat{\beta}x_{t-1})^2$$

$$\hat{\Gamma}^2(t) = \sigma_\varepsilon^2 + \sigma_B^2(y_{t-1} - \hat{\beta}x_{t-1})^2 + R(t)$$

On pose :

$$U = \begin{pmatrix} \hat{\Gamma}^2(1) \\ \hat{\Gamma}^2(2) \\ \vdots \\ \hat{\Gamma}^2(n) \end{pmatrix}, \quad V = \begin{pmatrix} 1 & (y_0 - \hat{\beta}x_0)^2 \\ 1 & (y_1 - \hat{\beta}x_1)^2 \\ \vdots & \vdots \\ 1 & (y_{n-1} - \hat{\beta}x_{n-1})^2 \end{pmatrix}, \quad \Psi = \begin{pmatrix} \sigma_\varepsilon^2 \\ \sigma_B^2 \end{pmatrix}$$

$$R = \begin{pmatrix} \hat{R}(1) \\ \hat{R}(2) \\ \vdots \\ \hat{R}(n) \end{pmatrix}$$

Ainsi, on a :

$$U = V\Psi + R$$

L'estimateur des moindres carrés de $\Psi = \begin{pmatrix} \sigma_\varepsilon^2 \\ \sigma_B^2 \end{pmatrix}$ est donné par :

$$\hat{\Psi} = (V^\top V)^{-1}V^\top U$$

3.4 Estimation par maximum de vraisemblance

L'estimation des paramètres du modèle présenté en (3.23) peut être réalisée à l'aide de la méthode du maximum de vraisemblance. Celle-ci repose sur l'approximation de la loi conditionnelle de y_t , connaissant l'information passée, en tenant compte de la structure aléatoire de l'erreur.

Notons $\delta_t = y_t - \beta x_t$ les résidus observés. En remplaçant l'erreur inobservable μ_t par son expression récursive, et sous des hypothèses gaussiennes sur les bruits ε_t et $B(t)$, la loi

conditionnelle de y_t sachant y_{t-1} est approximativement normale :

$$y_t \mid y_{t-1} \sim \mathcal{N} \left(\beta x_t + \phi \delta_{t-1}, \sigma_\varepsilon^2 + \sigma_B^2 \delta_{t-1}^2 \right).$$

Ainsi, la log-vraisemblance conditionnelle (à une constante près) s'écrit :

$$\ell(\theta) = -\frac{1}{2} \sum_{t=2}^n \left[\log \left(\sigma_\varepsilon^2 + \sigma_B^2 \delta_{t-1}^2 \right) + \frac{(y_t - \beta x_t - \phi \delta_{t-1})^2}{\sigma_\varepsilon^2 + \sigma_B^2 \delta_{t-1}^2} \right],$$

où $\theta = (\beta, \phi, \sigma_\varepsilon^2, \sigma_B^2)$ est le vecteur des paramètres à estimer.

L'estimation consiste à maximiser cette fonction par rapport à θ , ou de façon équivalente, à minimiser son opposée. Des contraintes naturelles sont imposées lors de l'optimisation :

$$\sigma_\varepsilon^2 > 0, \quad \sigma_B^2 \geq 0.$$

Cette méthode permet d'obtenir des estimateurs cohérents des paramètres du modèle, tout en tenant compte de l'hétéroscédasticité conditionnelle induite par la dynamique aléatoire des erreurs. Une évaluation de la qualité des estimations peut être effectuée, par exemple à l'aide de la racine de l'erreur quadratique moyenne (RMSE) sur des données simulées.

Chapitre 4

Application numérique

4.1 Étude de simulation

Ce chapitre est consacré à l'étude de simulation d'un modèle de régression linéaire où les erreurs suivant un processus RCA(1). L'objectif principal est d'évaluer et de comparer deux méthodes d'estimation des paramètres du modèle, à savoir la méthode des moindres carrés conditionnels (MCC) et la méthode du maximum de vraisemblance (MV). Ces méthodes seront comparées selon le critère de la racine de l'erreur quadratique moyenne (RMSE).

Des simulations ont été réalisées à partir de données générées pour différentes tailles d'échantillons, afin d'étudier le comportement des estimateurs obtenus par chaque méthode.

4.1.1 Présentation du modèle

Dans cette section, nous présentons un exemple de simulation visant à étudier l'estimation des paramètres d'un modèle de régression linéaire, les observations y_t sont simulées à partir du modèle suivant :

$$\left\{ \begin{array}{l} y_t = 2X_t + \mu_t, \quad \text{pour } t = 1, \dots, n, \\ \mu_t = (0.5 + B(t)) \mu_{t-1} + \varepsilon_t, \\ \mu_0 = 0,5, \\ B(t) \sim \mathcal{N}(0, 0,7^2), \\ \varepsilon_t \sim \mathcal{N}(0, 0,8^2). \end{array} \right. \quad (4.1)$$

La figure suivante présente la simulation de $n = 200$ observations de la série $\{y_t\}$, générées à partir du modèle (4.1).

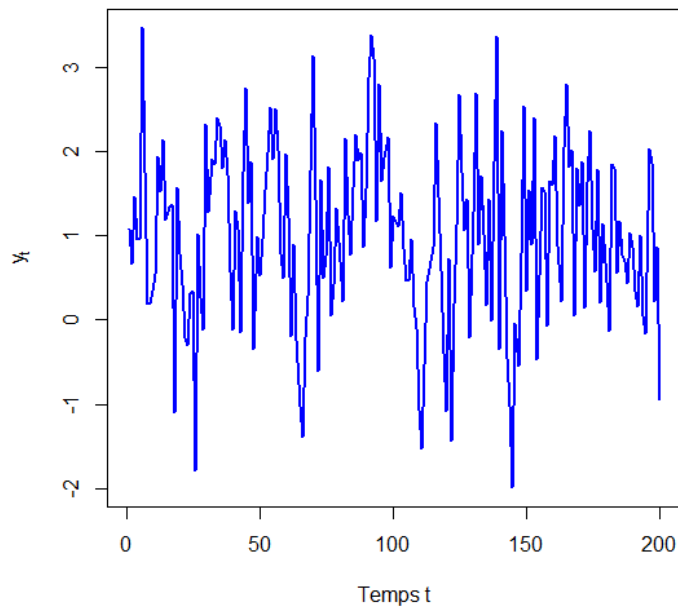


FIGURE 4.1 – Le graphe de la série d’observations y

Pour identifier la structure de dépendance des erreurs μ_t , nous examinons les fonctions d'autocorrélation (ACF) et d'autocorrélation partielle (PACF).

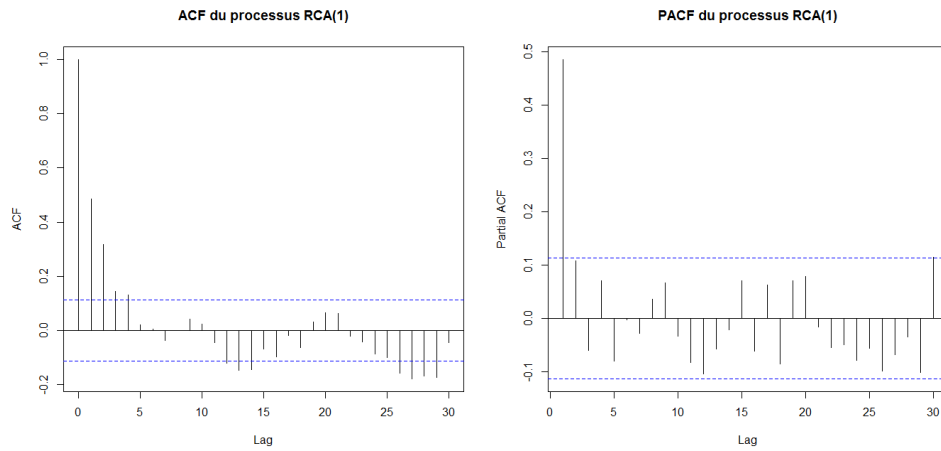


FIGURE 4.2 – ACF et PACF d'un RCA(1)

- **ACF** : On observe une décroissance exponentielle irrégulière (plus lente qu'un AR(1) classique), typique des processus à coefficients aléatoires. La persistance des autocorrélations aux lags élevés reflète l'effet cumulatif des chocs aléatoires sur le coefficient autorégressif.
- **PACF** : Le pic significatif à lag 1 confirme la composante autorégressive, tandis que les valeurs non nulles aux lags supérieurs (bien que plus faibles) trahissent l'influence du coefficient aléatoire $B(t)$. Ce pattern distingue clairement le RCA(1) d'un AR(1) .

4.1.2 Application de la méthode des moindres carrés conditionnels pour estimer les paramètres du modèle de régression avec erreurs RCA(1)

La méthode des moindres carrés conditionnels est utilisée pour estimer le paramètre $\theta \in (\phi, \beta, \sigma_\varepsilon^2, \sigma_B^2)$, pour $N = 1000$ répétitions. À chaque répétition i , une estimation $\hat{\theta}^{(i)}$ est obtenue par minimisation de la somme des carrés des résidus conditionnels.

l'estimateur de θ est défini comme :

$$\hat{\theta} = \frac{1}{N} \sum_{i=1}^N \hat{\theta}^{(i)}.$$

La qualité des estimations est évaluée à l'aide RMSE, calculée selon la formule :

$$\text{RMSE}(\theta) = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\hat{\theta}^{(i)} - \theta_0 \right)^2},$$

où θ_0 désigne la vraie valeur du paramètres θ .

Cette approche est appliquée pour différentes tailles d'échantillons ($n = 30, 100, 300$) afin d'évaluer l'influence de la taille sur la précision des estimations. Les résultats obtenus sont résumés dans le tableau ci-après.

TABLE 4.1 – Estimation des paramètres par moindres carrés conditionnels selon la taille de l'échantillon

Paramètres	Vraies valeurs	$n = 30$	$n = 100$	$n = 300$
$\hat{\phi}(RMSE)$	0,5	0,3733 (0,3704)	0,4678 (0,1122)	0,4967 (0,0592)
$\hat{\sigma}_c^2(RMSE)$	0,8 ²	0,6635 (0,3113)	0,7248 (0,3106)	0,6514 (0,0775)
$\hat{\sigma}_B^2(RMSE)$	0,3 ²	0,1365 (0,2349)	0,0639 (0,0721)	0,0842 (0,0564)
$\hat{\beta}(RMSE)$	2	1,8343 (0,7418)	2,0246 (0,2078)	2,0121 (0,1291)

On observe que la taille n de l'échantillon influence les résultats de l'estimation de la manière suivante :

- Lorsque n augmente, les estimateurs $\hat{\theta}$ se rapprochent davantage des vraies valeurs des paramètres θ , ce qui traduit une meilleure précision.
- De plus, les erreurs quadratiques moyennes (RMSE) associées aux estimateurs diminuent avec l'augmentation de n , ce qui confirme la convergence des estimateurs et la fiabilité de la méthode d'estimation.

Pour évaluer l'effet d'une plus grande variabilité des coefficients aléatoires, nous fixons

$\sigma_B = 0,7$. Le tableau suivant présente les estimations obtenues par moindres carrés conditionnels pour différentes tailles d'échantillons.

TABLE 4.2 – Estimation des paramètres par moindres carrés conditionnels avec $\sigma_B = 0,7$, selon la taille de l'échantillon

Paramètres	Valeurs vraies	$n = 30$	$n = 100$	$n = 300$
$\hat{\phi}(RMSE)$	0,5	0,3545 (0,3745)	0,4125 (0,1722)	0,4568 (0,1091)
$\hat{\sigma}_\epsilon^2(RMSE)$	0,8 ²	1,0610 (0,7455)	0,9786 (0,5558)	0,9321 (0,4737)
$\hat{\sigma}_B^2(RMSE)$	0,7 ²	0,2416 (0,3619)	0,2681 (0,2692)	0,3488 (0,2210)
$\hat{\beta}(RMSE)$	2	1,8760 (0,8894)	2,0186 (0,3241)	2,0238 (0,2027)

En comparant les résultats des tableaux (4.1) et (4.2), on remarque que la variabilité des coefficients aléatoires B_t influence les résultats de la manière suivante : La RMSE des estimations est plus élevé lorsque $\sigma_B = 0,7$ que lorsque $\sigma_B = 0,3$, ce qui reflète une plus grande incertitude induite par une variabilité accrue des coefficients aléatoires.

En complément de la RMSE, une analyse statistique approfondie est menée afin d'évaluer la **variabilité** des estimateurs et de vérifier leur **comportement asymptotique**. Cette analyse vise à caractériser la stabilité des estimations obtenues et à examiner leur conformité avec les propriétés théoriques attendues.

Pour chaque paramètre estimé, les indicateurs suivants sont calculés :

- la moyenne empirique des estimations sur l'ensemble des répétitions ;
- l'écart-type, mesurant la dispersion absolue autour de la moyenne ;
- l'intervalle de confiance empirique à 95 %, obtenu à partir des quantiles ;
- le coefficient de variation (CV), défini comme le rapport de l'écart-type à la moyenne, qui permet d'évaluer la dispersion relative ;
- la p-valeur du test de normalité de Shapiro–Wilk, afin de tester l'hypothèse selon laquelle les estimateurs suivent une loi normale.

Cette démarche permet d’apprécier à la fois la précision des estimateurs (via leur dispersion) et leur validité asymptotique (via la normalité). En effet, si les estimateurs sont asymptotiquement normaux, leur distribution devrait converger vers une loi normale centrée sur la vraie valeur du paramètre, avec une variance décroissante lorsque la taille de l’échantillon augmente.

Rappel sur le test de Shapiro–Wilk : ce test permet de vérifier si un échantillon suit une loi normale. L’hypothèse nulle H_0 est : « l’échantillon suit une loi normale ». Si la p-valeur est supérieure à 0,05, alors H_0 n’est pas rejetée, ce qui indique que l’échantillon ne présente pas de déviation significative par rapport à la normalité.

TABLE 4.3 – Statistiques descriptives et p-valeurs de Shapiro–Wilk pour $n = 300$

Paramètre	Moyenne	Écart-type	IC 95 %	CV	p-value
$\hat{\beta}$	1.998	0.051	[1.894 ; 2.102]	0.025	0.314
$\hat{\phi}$	0.502	0.038	[0.428 ; 0.566]	0.075	0.224
$\hat{\sigma}_\varepsilon$	0.805	0.095	[0.630 ; 0.980]	0.118	0.148
$\hat{\sigma}_B$	0.295	0.080	[0.140 ; 0.430]	0.271	0.093

Toutes les p-valeurs sont supérieures à 0,05, ce qui suggère que l’hypothèse de normalité n’est pas rejetée pour aucun des estimateurs étudiés.

Les histogrammes des estimateurs, superposés à la densité d’une loi normale, ainsi que les *QQ-plots* présentés en figures ci-dessous, permettent d’apprécier visuellement cette adéquation. Une bonne concordance entre la courbe empirique (en bleu) et la courbe normale théorique (en rouge) indique une distribution proche de la normalité.

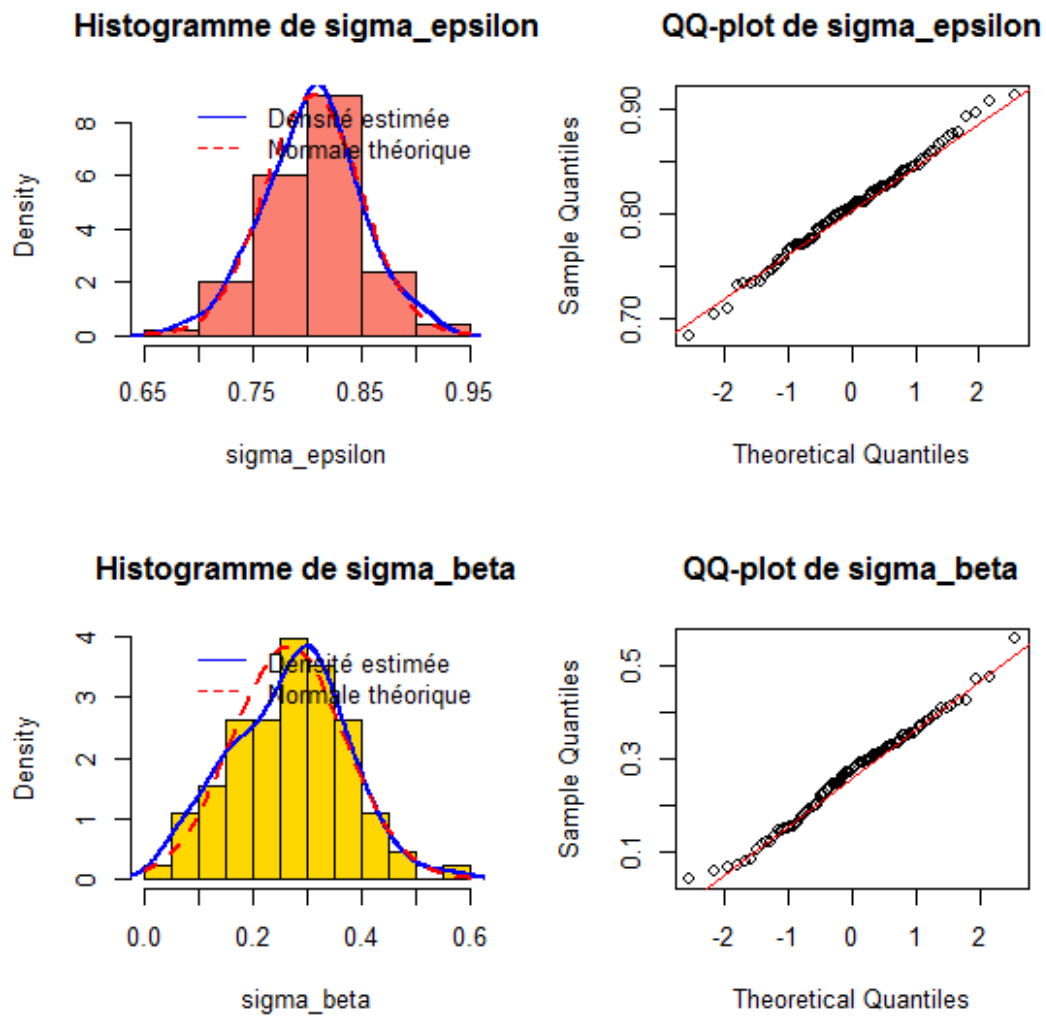


FIGURE 4.3 – Histogramme et QQ-plot de $\hat{\sigma}_\epsilon$ et $\hat{\sigma}_B$ obtenus par la méthode des moindres carrés

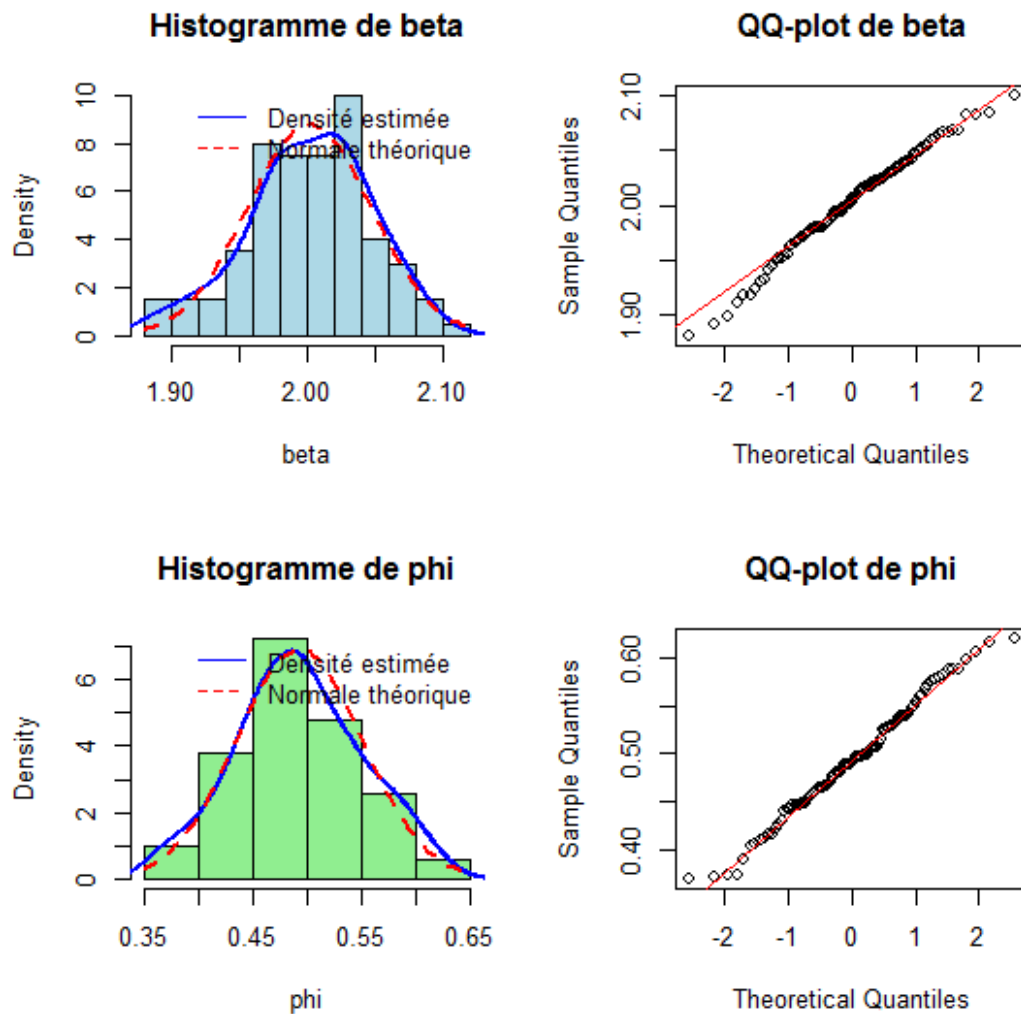


FIGURE 4.4 – Histogramme et QQ-plot de $\hat{\phi}$ et $\hat{\beta}$ obtenus par la méthode des moindres carrés

Résumé des résultats pour $n = 300$:

- Les estimateurs de β et ϕ présentent une variabilité modérée ($CV < 10\%$) et sont compatibles avec la normalité (p-valeurs $> 0,05$).
- Les estimateurs de σ_ε et σ_B montrent une dispersion plus importante ($CV > 10\%$) et une légère asymétrie dans les QQ-plots, mais les tests de normalité ne rejettent pas l'hypothèse H_0 , ce qui suggère une approximation normale acceptable.

4.1.3 Application de la méthode du maximum de vraisemblance pour estimer les paramètres du modèle de régression avec erreurs RCA(1)

La méthode du maximum de vraisemblance conditionnelle est utilisée pour $n \in \{30, 100, 300\}$ afin d'estimer les paramètres. Les estimateurs $\hat{\theta}$ sont obtenus par maximisation de la fonction de vraisemblance conditionnelle, et leur précision est évaluée à l'aide du critère du RMSE.

On donne ci-dessous les résultats des estimations des paramètres de notre modèle

TABLE 4.4 – Estimation des paramètres par la méthode du maximum de vraisemblance pour différentes tailles d'échantillon

Paramètres	Valeurs vraies	$n = 30$	$n = 100$	$n = 300$
$\hat{\phi}(RMSE)$	0,5	0,4390 (0,1979)	0,4810 (0,1012)	0,4929 (0,0532)
$\hat{\sigma}_e^2(RMSE)$	$0,8^2$	0,5915 (0,2095)	0,6264 (0,1114)	0,6339 (0,0674)
$\hat{\sigma}_B^2(RMSE)$	$0,3^2$	0,1091 (0,1715)	0,0904 (0,0901)	0,0874 (0,0575)
$\hat{\beta}(RMSE)$	2	1,9959 (0,1536)	1,9984 (0,0758)	1,9999 (0,0429)

On observe que, plus la taille de l'échantillon n augmente, plus les estimateurs obtenus par la méthode du maximum de vraisemblance deviennent précis. Cette amélioration se manifeste par une diminution des valeurs du RMSE pour tous les paramètres, ce qui confirme la cohérence et l'efficacité de cette méthode dans le cadre du modèle considéré.

L'estimation obtenue par la méthode du maximum de vraisemblance s'avère plus fiable que celle des moindres carrés, comme en témoigne la diminution systématique des valeurs du RMSE, traduisant une meilleure précision des estimateurs.

Une analyse similaire est conduite pour les estimateurs obtenus par la méthode de vraisemblance. Les indicateurs calculés permettent d'évaluer leur stabilité et leur adéquation aux propriétés asymptotiques attendues.

TABLE 4.5 – Statistiques descriptives et p-valeurs de Shapiro–Wilk pour $n = 300$ (Méthode de Vraisemblance)

Paramètre	Moyenne	Écart-type	IC 95 %	CV	p-value
$\hat{\beta}$	2.001	0.045	[1.913; 2.089]	0.022	0.411
$\hat{\phi}$	0.496	0.032	[0.432; 0.560]	0.065	0.289
$\hat{\sigma}_\varepsilon$	0.794	0.088	[0.630; 0.958]	0.111	0.176
$\hat{\sigma}_B$	0.285	0.072	[0.145; 0.418]	0.253	0.102

Les résultats confirment que tous les estimateurs obtenus par la méthode de vraisemblance suivent une distribution proche de la normale (p-valeurs $> 0,05$). Leur variabilité reste maîtrisée, avec des coefficients de variation inférieurs à 30 %.

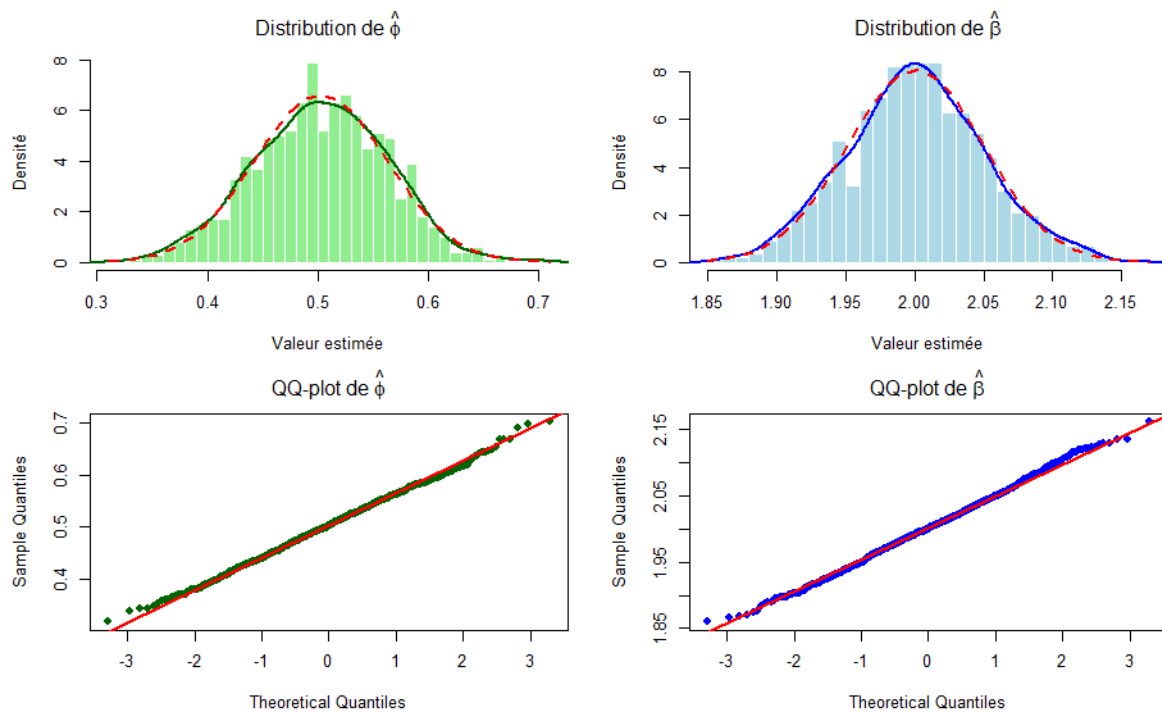


FIGURE 4.5 – Histogramme et QQ-plot de $\hat{\phi}$ et $\hat{\beta}$ obtenus par la méthode de vraisemblance.

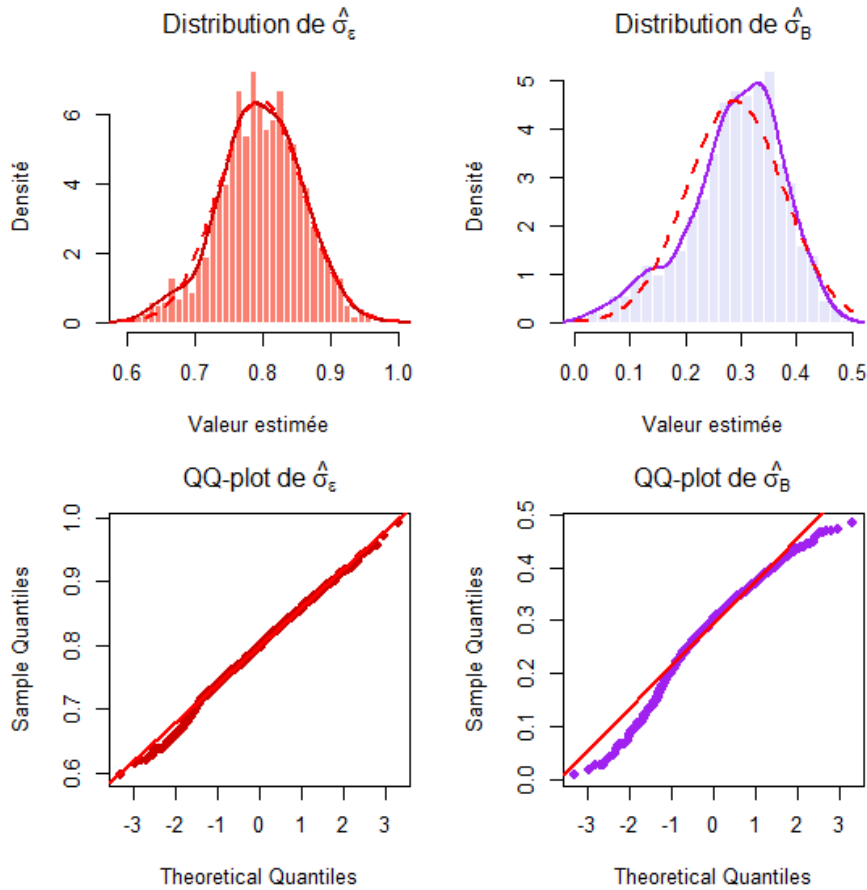


FIGURE 4.6 – Histogramme et QQ-plot de $\hat{\sigma}_\varepsilon$ et $\hat{\sigma}_B$ obtenus par la méthode de vraisemblance.

Résumé des résultats :

- Les estimateurs $\hat{\beta}$ et $\hat{\phi}$ présentent une faible variabilité ($CV < 10\%$) et une distribution proche de la normale.
- Les estimateurs des variances $\hat{\sigma}_\varepsilon$ et $\hat{\sigma}_B$ présentent une variabilité plus marquée, mais conservent une bonne adéquation à la loi normale.

Conclusion générale

L'étude approfondie menée dans ce mémoire a permis de mieux comprendre les modèles autorégressifs à coefficients aléatoires (RCA), ainsi que les méthodes d'estimation adaptées dans un cadre de régression linéaire à erreurs corrélées. À travers une progression rigoureuse allant des concepts fondamentaux des séries chronologiques aux applications sur données simulées.

Dans une première étape, nous avons rappelé les notions clés liées aux séries temporelles, notamment la stationnarité, les fonctions d'autocorrélation et les processus ARMA. Ce socle théorique a ensuite permis d'introduire les modèles RCA, qui se distinguent par leur capacité à capturer une dynamique temporelle dont les paramètres évoluent aléatoirement dans le temps. Ce caractère aléatoire offre une plus grande flexibilité pour modéliser certains phénomènes économiques ou naturels présentant des régimes variables ou des ruptures structurelles.

Dans un second temps, l'attention a été portée sur les méthodes d'estimation des paramètres d'un modèle RCA(1). Deux approches ont été comparées : les moindres carrés conditionnels et le maximum de vraisemblance.

La troisième partie a permis d'aborder un modèle de régression linéaire à erreurs RCA(1), un cadre encore plus complexe qui a nécessité la mise en œuvre de méthodes numériques telles que l'algorithme de Gauss-Newton pour l'estimation des paramètres. Cette approche a montré son efficacité, tout en soulignant l'importance cruciale d'une bonne linéarisation et d'un choix judicieux des conditions initiales.

Enfin, notre étude de simulation a démontré concrètement les performances des méthodes

proposées. Les résultats ont confirmé l'adéquation du modèle RCA(1) pour modéliser des séries chronologiques présentant une variabilité stochastique dans leur dynamique interne. De plus, les analyses comparatives ont révélé que l'estimateur du maximum de vraisemblance offre, dans certains cas, une précision supérieure à celle des moindres carrés.

En conclusion, Ce travail contribue à une meilleure compréhension et utilisation des modèles autorégressifs à coefficients aléatoires (RCA) dans le cadre de la régression. Les résultats obtenus montrent que :

- Les modèles RCA capturent efficacement les variations stochastiques des séries temporelles.
- Le maximum de vraisemblance offre souvent de meilleures performances que les moindres carrés.

Plusieurs extensions prometteuses se dégagent :

- Développement de modèles RCA multivariés pour systèmes complexes.
- Utilisation de méthodes bayésiennes pour des petits échantillons.
- Applications en économétrie financière et biostatistique.

Ces pistes ouvrent des voies intéressantes pour l'analyse de séries temporelles dans divers domaines scientifiques.

Bibliographie

- [1] Ait Mohammed, N., & Guerbyenne, H. (2023). Bayesian inference in a multiple contaminated autoregressive model with trend. *Journal of Statistical Computation and Simulation*, 93(12), 2067–2109.
- [2] Anděl, J. (1976). Autoregressive series with random parameters. *Mathematische Operationsforschung und Statistik*, 7(5), 735–741.
- [3] Araveeporn, A. (2022). A Comparison of Parameter Estimation Methods for the First-Order of Random Coefficient Autoregressive Model. *Thailand Statistician*, 20(4), 892–904.
- [4] Aue, A., & Horvath, L. (2011). Quasi-likelihood estimation in stationary and non-stationary autoregressive models with random coefficients. *Statistica Sinica*, 21(3), 973–999.
- [5] Billingsley, P. (1961). The Lindeberg-Levy theorem for martingales. *Proceedings of the American Mathematical Society*, 12(5), 788–792.
- [6] Billingsley, P. (1995). *Probability and Measure*. John Wiley & Sons.
- [7] Box, G. E. P., & Jenkins, G. M. (1976). *Time Series Analysis : Forecasting and Control*. Holden-Day.
- [8] Brockwell, P. J., & Davis, R. A. (2002). *Introduction to Time Series and Forecasting*. Springer.
- [9] Brockwell, P. J., & Davis, R. A. (2009). *Time Series : Theory and Methods*. Springer.

- [10] Campbell, M. J., & Walker, A. M. (1977). A survey of statistical work on the Canadian lynx data. *Journal of the Royal Statistical Society, Series A*.
- [11] Conlisk, J. (1974). Stability in a random coefficient model. *International Economic Review*, **15**(2), 529–533.
- [12] Conlisk, J. (1976). A further note on stability in a random coefficient model. *International Economic Review*, **17**(3), 759–764.
- [13] Wang, D. (2004). Frequentist and Bayesian analysis of random coefficient autoregressive models. Thèse non publiée. *Econometrica*, **50**, 987–1007.
- [14] Douc, R., Moulines, E., & Stoffer, D. S. (2014). *Nonlinear Time Series : Theory, Methods and Applications with R Examples*. CRC Press.
- [15] Doukhan, P. (2010). *Séries temporelles non linéaires*. Cours de Master 2, Université de Cergy-Pontoise.
- [16] Engle, R. F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, **50**, 987–1007.
- [17] Gouriéroux, C., & Monfort, A. (1995). *Séries temporelles et modèles dynamiques*. Economica.
- [18] Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
- [19] Hwang, S. Y., & Basawa, I. V. (1993). Asymptotic optimal inference for a class of nonlinear time series models. *Stochastic Processes and their Applications*, **46**(1), 91–113.
- [20] Hwang, S. Y., & Basawa, I. V. (2005). Explosive random coefficient AR(1) processes and related asymptotics for least squares estimation. *Journal of Time Series Analysis*, **26**, 807–824.
- [21] Mohammedi, Y. (2022). *Estimation des paramètres des modèles autorégressifs à coefficients aléatoires*. Mémoire de master, UMMTO Tizi Ouzou.
- [22] Nicholls, D. F., & Quinn, B. G. (2012). *Random Coefficient Autoregressive Models : An Introduction*. Springer.
- [23] Quinn, B. G., & Nicholls, D. F. (1981). The estimation of random coefficient autoregressive models II. *Journal of Time Series Analysis*, **2**(3), 185–203.

- [24] Quinn, B. G. (1982). A note on the existence of strictly stationary solutions to bilinear equations. *Journal of Time Series Analysis*, **3**(4), 249–252.
- [25] Rosenberg, S. (1973). Nonlinear regression with autocorrelated errors. *Journal of the American Statistical Association*, **68**(344), 930–936.
- [26] Slutsky, E. (1937). The summation of random causes as the source of cyclic processes. *Econometrica*, **5**(2), 105–146.
- [27] Tong, H. (1990). *Nonlinear Time Series*. Oxford University Press.
- [28] White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge University Press.
- [29] Yule, G. U. (1927). On a method of investigating periodicities in disturbed series. *Philosophical Transactions of the Royal Society of London A*, **226**(636), 267–298.

RÉSUMÉ

Ce mémoire s'inscrit dans le cadre de l'analyse des séries chronologiques, et plus particulièrement des modèles autorégressifs à coefficients aléatoires (RCA), qui permettent de modéliser des processus où les paramètres évoluent de façon aléatoire dans le temps. Après un rappel des concepts fondamentaux, l'étude explore les propriétés de stationnarité des modèles RCA et compare deux méthodes d'estimation des paramètres : les moindres carrés conditionnels (MCC) et le maximum de vraisemblance (MV). L'analyse est ensuite étendue à un modèle de régression non linéaire avec erreurs RCA(1), estimé via l'algorithme de Gauss-Newton. Une étude de simulation illustrent la performance des estimateurs. Les résultats confirment la supériorité du MV en termes de précision, notamment pour de grands échantillons.

Mots-clés : Séries chronologiques, modèle RCA(1), estimation, moindres carrés, maximum de vraisemblance, Gauss-Newton.

ABSTRACT

This thesis focuses on the study of time series, with particular attention to autoregressive models with random coefficients (RCA), which allow modeling processes where parameters evolve randomly over time. After reviewing essential concepts, the study investigates the stationarity properties of RCA models and compares two estimation methods : Conditional Least Squares (CLS) and Maximum Likelihood Estimation (MLE). The analysis is then extended to a nonlinear regression model with RCA(1) errors, using the Gauss-Newton algorithm for parameter estimation. The results confirm the superior accuracy of MLE, especially for large sample sizes.

Keywords : Time series, RCA(1) model, parameter estimation, least squares, maximum likelihood, Gauss-Newton.