COLLEGE ED MATHEMATIQUES
DOCTORAL STIC
PAYS DE LA LOIRE

Le Mans Université

THÈSE DE DOCTORAT DE

LE MANS UNIVERSITÉ

ÉCOLE DOCTORALE N° 641
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : *Mathématiques et leurs Interactions*

Par
Marie BADREAU

Statistique asymptotique pour les séries chronologiques quasi-instables et/ou à mémoire longue

Thèse présentée et soutenue à Le Mans, le 30 mars 2026
Unité de recherche : Laboratoire Manceau de Mathématiques
Thèse N° : 2026LEMA1004

Rapporteurs avant soutenance :

Jean-Marc BARDET	Professeur, Université Paris 1 Panthéon-Sorbonne
Sophie DABO NIANG	Professeur, Université de Lille

Composition du Jury :

Présidente :	Marie KRATZ	Professeur, ESSEC Business School
Examinateurs :	Farida ENIKEEVA	Maitre de conférences, Université de Poitiers
	Anne PHILIPPE	Professeur, Nantes Université
Dir. de thèse :	Alexandre BROUSTE	Professeur, Le Mans Université
Co-dir. de thèse :	Frédéric PROÏA	Professeur, Nantes Université
Co-dir. de thèse :	Youssef ESSTAFA	Maitre de conférences, Le Mans Université

Remerciements

Avant de commencer la présentation de ce travail, je souhaite exprimer ma profonde gratitude à toutes les personnes qui m'ont soutenue, encouragée et accompagnée tout au long de cette thèse, et sans qui ce travail n'aurait pu aboutir.

Je souhaite en premier lieu remercier mon co-directeur de thèse, Frédéric PROÏA, pour l'aide précieuse et le soutien qu'il m'a apportés dans la construction de mon projet doctoral, et tout particulièrement pour m'avoir mis en relation avec l'équipe du Mans. Je le remercie également pour la qualité de son encadrement, qui m'a transmis le goût du travail bien fait, ainsi que pour la richesse de nos échanges, y compris ceux, plus littéraires, qui ont constitué des parenthèses stimulantes et inspirantes au cœur du travail scientifique. Je tiens également à remercier mon directeur de thèse, Alexandre BROUSTE, pour la confiance qu'il m'a accordée en me proposant ce sujet de thèse et en permettant l'obtention du financement nécessaire à sa réalisation. Je le remercie aussi pour sa disponibilité tout au long de ce travail, la pertinence de ses conseils, ainsi que pour son soutien actif dans ma projection professionnelle, et notamment pour m'avoir orientée vers des opportunités prometteuses. Je souhaite adresser enfin mes remerciements les plus sincères à Youssef ESSTAFA, pour son encadrement bienveillant ainsi que pour sa pédagogie et la clarté de ses explications, qui m'ont permis de progresser tant sur le plan scientifique que méthodologique. Je le remercie pour ses encouragements qui m'ont aidée à surmonter les moments de doute et à gagner en autonomie. Je lui suis également reconnaissante pour la perspective de collaborations futures que j'envisage avec enthousiasme.

Je remercie sincèrement l'ensemble de mon équipe encadrante pour leur compréhension, leur patience et leur humanité, en particulier dans les moments plus difficiles que j'ai pu traverser, tant sur le plan professionnel que personnel. Votre confiance et votre bienveillance ont largement contribué à l'aboutissement de cette thèse.

Je remercie sincèrement Jean-Marc BARDET ainsi que Sophie DABO NIANG d'avoir accepté d'être les rapporteurs de cette thèse et pour l'attention qu'ils ont portée à l'évaluation de ce travail. Je leur suis particulièrement reconnaissante pour le temps consacré à la lecture attentive du manuscrit ainsi qu'à la rédaction de leurs rapports.

Je remercie Farida ENIKEEVA, Marie KRATZ et Anne PHILIPPE d'avoir accepté de faire partie de mon jury de thèse et pour l'intérêt porté à ce travail.

Je souhaite également remercier les membres de mon CSI, Marie KRATZ et Gilles STUPFLER, pour leur expertise, leurs conseils avisés et leur soutien tout au long de la réalisation de cette thèse.

Je remercie chaleureusement les laboratoires LMM et LAREMA pour m'avoir accueillie au sein de leurs équipes durant la réalisation de cette thèse. Je leur suis reconnaissante pour l'environnement scientifique stimulant et les conditions de travail qu'ils m'ont offertes. Je remercie ainsi l'ensemble des membres de ces laboratoires, avec un remerciement tout particulier aux doctorants manceaux et angevins pour les moments d'échange, de partage et de convivialité.

Je souhaite également exprimer ma gratitude à l'ensemble de ma famille et amis pour leur soutien constant, leurs encouragements et leur présence à mes côtés pendant toutes ces années. Je souhaite remercier tout spécialement ma maman pour m'avoir aidée et soutenue comme seule une maman peut le faire, ainsi que mon papa pour m'avoir toujours encouragée à devenir docteur.

J'adresse enfin des remerciements tout particuliers à ma famille, à qui je dédie cette thèse. Je ne remercierai jamais assez mon conjoint pour sa patience et son soutien précieux tout au long de mes années d'études ; merci pour tes notes qui ont créé une atmosphère musicale propice à l'écriture de cette thèse. Je souhaite également remercier la plus jolie des clowns, dont l'humour débordant et l'amour sans limite ont largement embelli mes pauses, ainsi que mon petit guerrier, qui m'a rappelé que les batailles les plus dures rendent les victoires encore plus belles.

TABLE DES MATIÈRES

Introduction **9**

Axe 1 : Inférence et test statistique dans la classe des processus autorégressifs quasi-instables . . . 12

Stationnarité et processus autorégressifs . . . 12

Inférence des processus autorégressifs stables . . . 17

Inférence et tests statistiques pour les processus autorégressifs instables . . . 19

Processus autorégressifs quasi-instables . . . 24

Contributions de la thèse . . . 30

Axe 2 : Sélection de modèle dans la classe des processus FARIMA faibles . . . 37

Généralités sur les processus FARIMA . . . 37

Inférence des processus FARIMA . . . 46

Sélection de modèle dans la classe des processus FARIMA . . . 53

Contributions de la thèse . . . 55

1 Consistency and asymptotic normality in a class of nearly unstable processes **59**

1.1 Introduction and motivation . . . 59

1.2 Main results . . . 63

1.2.1 Causal representation and memory . . . 65

1.2.2 OLS estimation . . . 67

1.3 Technical tools . . . 70

1.3.1 Linear Algebra . . . 70

1.3.2 Proof of Propositions 1.2.1 and 1.2.2 . . . 74

1.3.3 Proof of Theorem 1.2.3 . . . 79

1.3.4 Proof of Theorem 1.2.4 . . . 81

1.4 Applications and simulations . . . 84

References . . . 87

2 Testing for the extent of instability in nearly-unstable processes 89
2.1 Introduction and Motivation . . . 90
2.2 A (Quasi-)Unit Root Test . . . 93
2.2.1 The variance parameter c is known . . . 97
2.2.2 The variance parameter c is estimated . . . 99
2.3 Empirical Study . . . 100
2.3.1 On simulated data (c is known) . . . 100
2.3.2 On simulated data (c is estimated) . . . 102
2.3.3 On a real dataset . . . 103
2.4 Conclusion and Perspectives . . . 105
2.5 Technical Developments . . . 108
2.5.1 Known results . . . 108
2.5.2 Technical lemmas . . . 110
2.5.3 Proof of Proposition 2.2.1. . . . 116
2.5.4 Proof of Theorem 2.2.2. . . . 117
2.5.5 Proof of Theorem 2.2.3. . . . 118
2.5.6 The case of a negative unit root . . . 119
References . . . 120

3 Model selection in nonstationary weak FARIMA models 123
3.1 Introduction and Motivation . . . 123
3.2 General framework . . . 126
3.2.1 The Data Generating Process . . . 126
3.2.2 Notation and Model Selection Setup . . . 127
3.2.3 Conditional-Sum-of-Squares Estimator and Its Asymptotic Properties 128
3.3 Bayesian Information Criterion . . . 130
3.3.1 Model Selection Procedure . . . 130
3.3.2 Model Selection Consistency . . . 131
3.4 Simulations and Application to real data . . . 134
3.4.1 Simulations . . . 134
3.4.2 Application to real data . . . 136
3.5 Conclusion . . . 138
References . . . 140
A Proof of Theorem 3.2.4 . . . 143

A.1 Preliminary Lemmas for Proof of Theorem 3.2.4 143
A.2 Proof of Consistency 146
B Proof of Theorem 3.3.2 149
B.1 Preliminary Lemmas for Proof of Theorem 3.3.2 149
B.2 Proof of Theorem 3.3.2 150

Conclusion et perspectives **155**

Bibliographie **157**

INTRODUCTION

Une série chronologique, ou série temporelle, fait généralement référence à une suite d'évènements aléatoires décrits par une quantité numérique et indexés par le temps. Il existe de nombreux champs d'application comme l'économie et la finance (évolution des indices boursiers ou du PIB [1]), la climatologie (suivi des températures, volume des précipitations, etc.), l'hydrologie (observation du débit des fleuves, de leur hauteur, etc.), l'épidémiologie (incidence ou nombre de nouveaux cas dans le suivi d'une épidémie) et bien d'autres domaines. En effet, les séries temporelles peuvent être utilisées dès qu'il est question de modéliser l'évolution des données au cours du temps, qu'il s'agisse d'un enregistrement à intervalles de temps définis (horaire, journalier, mensuel, annuel, etc.) correspondant à du temps discret, ou un enregistrement de haute fréquence, modélisable par du temps continu.

Les premières théories mathématiques dans ce domaine remontent au début du 20^{e} siècle avec notamment Yule (1927) et son étude de l'évolution du nombre de tâches solaires observées à intervalles de temps régulier. Chaque donnée est considérée comme une combinaison linéaire des observations précédentes formalisant de ce fait le concept de processus autorégressif (AR [2]) en temps discret. Cette première modélisation est au cœur de ce travail et sera donc détaillée dans la section suivante. De manière analogue, le processus d'Ornstein et Uhlenbeck (1930) peut être considéré comme l'équivalent en temps continu d'un processus autorégressif d'ordre 1 ou AR(1). Bien qu'il existe un parallèle étroit entre les processus à temps discret et continu, l'ensemble de cette thèse se limitera aux séries chronologiques à temps discret.

Parmi les nombreuses recherches menées dans ce domaine, on peut citer le livre de Doob (1953) qui propose un cadre théorique pour certaines notions émergentes, aujourd'hui fondamentales dans l'étude des séries chronologiques telles qu'on les connaît. Ainsi, les processus stochastiques, terme qui donne son titre à l'ouvrage, y sont définis comme une famille de variables aléatoires $\{X_t, t \in T\}$. Doob va d'ailleurs établir une distinction claire entre les processus à temps continu, lorsque T est un intervalle, et les processus à

1. Produit Intérieur Brut
2. *Autoregressive*, en anglais.

temps discret, lorsque T est une suite infinie. De même, il formalise le concept de stationnarité stricte d'un processus pour lequel la distribution des vecteurs $(X_{t_1+h}, \ldots, X_{t_n+h})$ est identique, indépendamment de la valeur du décalage temporel h. Il distingue également les processus dont la fonction de covariance entre deux variables X_t et X_{t+h} ne dépend que de h, ce qui caractérise la stationnarité au sens large (ou stationnarité au second ordre). À noter que dans ce deuxième cas, des conditions supplémentaires sont requises : la moyenne du processus doit être fixe dans le temps et la variance finie. Par la suite, un processus stationnaire fera naturellement référence à une stationnarité au sens large, l'hypothèse de stationnarité au second ordre étant moins contraignante et donc plus souvent utilisée en pratique. Il existe ainsi une frontière bien définie entre processus stationnaire et non stationnaire, qui explique notamment la différence dans les comportements limites des estimateurs. Les séries dites intégrées, c'est-à-dire dont le processus sous-jacent possède au moins une racine unitaire [3], constituent un cas typique de non-stationnarité. Les tests ADF [4] et KPSS [5], respectivement proposés par Said et Dickey (1984) et Kwiatkowski et al. (1992), sont probablement les plus utilisés en statistique pour évaluer la présence de racine unitaire, et donc la stationnarité d'une série. Toutefois, lorsque le processus étudié se situe à la limite entre stationnarité et non-stationnarité, une certaine incertitude apparaît et les résultats obtenus peuvent sensiblement différer des prévisions théoriques. Partant de ce constat, des travaux récents se sont consacrés à l'étude des propriétés statistiques pour l'estimation de ces processus particuliers. Le premier axe de cette thèse fait partie intégrante de ces nouvelles recherches.

Avec l'émergence des premiers processus autorégressifs, Wold (1938) propose son théorème de décomposition impliquant que tout processus $(X_t)_{t\in\mathbb{Z}}$ stationnaire au second ordre peut être exprimé tel que

$$\forall t \in \mathbb{Z}, \quad X_t = X_t^{(d)} + \sum_{j=0}^{+\infty} \psi_j \epsilon_{t-j}, \tag{1}$$

où $X_t^{(d)}$ est la composante déterministe du processus, la série $(\psi_j)_{j\geqslant 0}$ est telle que $\psi_0 = 1$ et $\sum_j |\psi_j|^2$ est convergente, et $(\epsilon_t)_{t\in\mathbb{Z}}$ est un bruit blanc faible, i.e. une suite de variables aléatoires centrées, de variance commune, et non corrélées (mais pas nécessairement indépendantes), et il s'agit du bruit d'innovation (ϵ_t est décorrélé de X_s pour tout $s < t$).

3. Racine de module 1 dans le polynôme caractéristique associé, voir notamment (3) pour le polynôme AR

4. *Augmented Dickey–Fuller*.

5. Kwiatkowski–Phillips–Schmidt–Shin.

Ainsi, Wold montre que tout processus stationnaire purement non-déterministe admet une représentation en moyenne mobile infinie, ou MA(∞) [6], correspondant à une combinaison linéaire infinie de bruits blancs. Box et Jenkins (1970) reprendront ce modèle pour en proposer une version tronquée à un ordre q fini ($\psi_j = 0$ dès que $j > q$) afin de limiter le nombre de terme du modèle tout en conservant une modélisation adaptée aux dépendances temporelles. Ils développeront également la théorie moderne autour des processus autorégressifs à moyenne mobile ARMA(p, q) [7] en y associant une représentation parcimonieuse d'un AR d'ordre p et d'un MA d'ordre q. Ils suggèrent d'ailleurs une nouvelle approche, aujourd'hui connue sous le nom de *méthode Box-Jenkins*, dont l'objectif est de modéliser l'évolution d'un phénomène au cours du temps. La première étape consiste à sélectionner le modèle, ce qui correspond ici à l'identification des paramètres p et q. Nous reviendrons ultérieurement sur cette problématique fondamentale puisqu'elle est étroitement liée au deuxième axe de recherche de cette thèse. Une fois les ordres p et q fixés, l'estimation des coefficients et la validation du modèle sont nécessaires, avant de pouvoir atteindre l'ultime étape de cette méthode qui est la prédiction de nouvelles données grâce au modèle estimé. Cette nouvelle méthodologie a marqué un tournant dans l'analyse et l'application des séries temporelles à des fins empiriques.

Aujourd'hui largement répandues, les séries chronologiques ont fait l'objet de multiples adaptations depuis leurs premières modélisations mathématiques au début du 20^{e} siècle afin de pouvoir répondre à la complexité des évènements observés et s'ajuster au plus près des données. L'apparition des modèles VAR [8] à la suite des processus autorégressifs (voir notamment Sims (1980) pour leur formalisation théorique) permet par exemple de traiter plusieurs variables simultanément, en supposant que les valeurs futures de l'une dépendent non seulement de son passé mais aussi du passé des autres (très utilisés en macroéconomie pour modéliser les indices boursiers). L'hypothèse de stationnarité, introduite précédemment, est également fréquemment requise dans la modélisation des séries temporelles bien qu'elle ne soit pas toujours adaptée en pratique. En effet, de nombreuses séries empiriques ne vérifient pas cette condition préalable. Pour autant, les accroissements d'une série de données a priori non stationnaire forment parfois une série stationnaire. Cette observation a menée au développement des modèles ARIMA(p, d, q) [9] permettant de passer d'un processus non stationnaire à stationnaire en appliquant un processus ARMA(p, q) à la

6. *Moving Average.*
7. *Autoregressive Moving Average.*
8. *Vector Autoregressive.*
9. *Autoregressive Integrated Moving Average.*

série de données différenciée d fois. Les propriétés de ce modèle ont d'ailleurs été étudiées par Box et Jenkins (1970) à la suite des représentations ARMA, en supposant que d est un entier strictement positif. La nécessité de prendre en compte certains phénomènes de dépendance à long terme, ou mémoire longue, conduira à l'émergence des processus FARIMA [10] pour lesquels l'opérateur de différenciation d peut maintenant prendre n'importe quelle valeur réelle. Granger et Joyeux (1980) proposent une étude de cette différenciation fractionnaire via la classe des modèles FARIMA$(0, d, 0)$ où, dans un premier temps, les processus AR et MA n'interviennent pas. Hosking (1981) généralisera cette approche en considérant l'ensemble des processus FARIMA(p, d, q) où les ordres p et q font référence au processus ARMA sous-jacent. En assouplissant les restrictions sur le paramètre d, il est clair que la classe des processus FARIMA(p, d, q) englobe celle des ARIMA qui contient elle-même les modèles ARMA pour $d = 0$. Il devient alors pertinent de se poser la question du choix des ordres p et q qui ajusteraient ce processus au plus près des données. Le deuxième axe de recherche de cette thèse s'intéresse donc à la sélection de modèle au sein de la classe des processus FARIMA, sous des hypothèses assez larges, qui seront spécifiées par la suite.

Axe 1 : Inférence et test statistique dans la classe des processus autorégressifs quasi-instables

Le premier axe de recherche de cette thèse traite de l'inférence des processus autorégressifs quasi-instables d'ordre p. Après avoir introduit quelques généralités sur la classe des processus autorégressifs et les propriétés asymptotiques de l'estimateur des moindres carrés dans le cas stable et instable, le concept de quasi-instabilité sera précisé avant de terminer par un résumé approfondi des articles de recherche correspondant (voir Chapitre 1 et Chapitre 2).

Stationnarité et processus autorégressifs

Bien que Box et Jenkins (1970) soient les pionniers en matière de théorie des séries chronologiques, on peut également citer d'autres travaux, plus récents, qui ont contribué à l'essor de ces processus comme par exemple Brockwell et Davis (1991), Lütkepohl (2005) ou encore Shumway et Stoffer (2006) pour des applications plus concrètes sur le logiciel R.

10. *Fractionally Autoregressive Integrated Moving Average.*

Processus stationnaires

Avant d'introduire plus amplement le modèle autorégressif, il semble important de revenir sur quelques termes mentionnés précédemment. Ainsi, une famille de variables aléatoires (X_t) indexées par $t \in T$ est appelée *processus stochastique*. Lorsque T est un espace des temps dénombrable (typiquement $\mathbb{Z}$ ou $\mathbb{N}$ si on considère un point de départ bien défini), on parle de processus stochastique à temps discret ou série chronologique discrète [11]. Dans le cadre d'une modélisation, on supposera que l'ensemble des n données observées est la réalisation d'un vecteur aléatoire $(X_1, \ldots, X_n)$ lui-même issu d'un processus stochastique. Ce processus est dit *strictement stationnaire* si pour tout $k \geqslant 1$ et tous $t_1, \ldots, t_k, h \in \mathbb{Z}$, les vecteurs $(X_{t_1}, \ldots, X_{t_k})$ et $(X_{t_1+h}, \ldots, X_{t_k+h})$ ont la même loi jointe. Un processus strictement stationnaire particulièrement utilisé en série temporelle est le *bruit blanc fort*, une suite de variables aléatoires indépendantes et identiquement distribuées (iid). De même, la série $(X_t)_{t\in\mathbb{Z}}$ est dite *stationnaire au second ordre* (ou simplement stationnaire) lorsque

$$\begin{aligned} (i) \quad & \mathbb{E}|X_t|^2 < \infty \quad \forall t \in \mathbb{Z} \\ (ii) \quad & \mathbb{E}[X_t] = m \quad \forall t \in \mathbb{Z} \\ (iii) \quad & \mathrm{Cov}(X_t, X_{t+h}) = \gamma(h) \quad \forall t, h \in \mathbb{Z} \end{aligned}$$

où γ représente la fonction d'autocovariance du processus stationnaire (X_t). De manière complémentaire, on appellera *bruit blanc faible* un processus stationnaire centré [12] tel que $m = 0$ et $\gamma(h) = \sigma^2 \mathbb{1}_{\{h=0\}}$ avec $0 < \sigma^2 < \infty$. On a supposé $T = \mathbb{Z}$ dans les définitions ci-dessus, mais, sous certaines conditions mentionnées plus loin, ces dernières restent valables sur $\mathbb{N}$. De même, les processus considérés ici englobent ceux à valeurs complexes mais, dans le cadre de cette thèse, on se contentera d'étudier les processus à valeurs réelles.

Processus autorégressifs

Les processus AR(p) jouent un rôle fondamental en statistique des séries temporelles tant par la simplicité de leur structure (régression linéaire sur les valeurs passées) que par les multiples généralisations qui en sont dérivées. Soit $p \geqslant 1$, le processus centré $(X_t)_{t\in\mathbb{Z}}$

11. Par opposition, le processus est continu lorsque T est un intervalle de $\mathbb{R}$.
12. Un bruit blanc n'est pas nécessairement centré même si cette hypothèse est souvent ajoutée en pratique. Celle-ci sera admise dans la suite de ce travail.

est dit autorégressif d'ordre p si

$$\forall t \in \mathbb{Z}, \quad X_t = \sum_{i=1}^{p} \theta_i X_{t-i} + \epsilon_t, \tag{2}$$

avec $\theta = (\theta_1, \ldots, \theta_p)^\top$ le vecteur des paramètres à estimer, pour lequel on impose $\theta_p \neq 0$ pour une écriture minimale, et (ϵ_t) un bruit blanc de variance $\sigma^2 < \infty$, que l'on supposera iid dans cette section. En utilisant l'opérateur retard L et le polynôme autorégressif

$$\Theta(z) = 1 - \theta_1 z - \cdots - \theta_p z^p \quad \forall z \in \mathbb{C}, \tag{3}$$

on peut également écrire (2) de la façon suivante

$$\forall t \in \mathbb{Z}, \quad \Theta(L) X_t = \epsilon_t.$$

Le processus AR(p) est *causal* s'il peut être exprimé comme une combinaison linéaire du présent et des valeurs passées d'une autre variable, en l'occurrence du bruit blanc. Plus précisément, cela signifie qu'il existe une suite (ψ_j) telle que $\sum_{j=0}^{\infty} |\psi_j| < \infty$ et

$$\forall t \in \mathbb{Z}, \quad X_t = \epsilon_t + \sum_{j=1}^{+\infty} \psi_j \epsilon_{t-j}. \tag{4}$$

Cette représentation, issue de la décomposition de Wold, peut être obtenue à condition que le polynôme vérifie $\Theta(z) \neq 0$ pour tout $z \in \mathbb{C}$ tel que $|z| \leqslant 1$. Dans ce cas, les coefficients (ψ_j) sont déterminés par la série

$$\frac{1}{\Theta(z)} = \sum_{j \geqslant 0} \psi_j z^j \tag{5}$$

dans son rayon de convergence, et le processus (X_t) est stationnaire. La causalité d'un processus autorégressif implique donc sa stationnarité. D'après le Théorème 3.1.3 de Brockwell et Davis (1991), lorsque l'on relâche les hypothèses de causalité en considérant le cas général où $\Theta(z) \neq 0$ pour tout $z \in \mathbb{C}$ tel que $|z| \neq 1$, on peut noter que ces processus AR(p) restent stationnaires mais l'inversion du polynôme fait intervenir les valeurs futures du bruit blanc. Ce type de processus admet un caractère anticipatif qui impacte leur portée pratique de part leur interprétation. Ainsi, lorsque l'étude porte sur $T = \mathbb{N}$ et non $\mathbb{Z}$, les séries non causales présentent des comportements explosifs, ce qui nécessite une terminologie appropriée.

Prenons l'exemple d'un AR(1) avec $|\theta_1| \neq 1$ tel que

$$\forall t > 0, \quad X_t = \theta_1 X_{t-1} + \epsilon_t, \tag{6}$$

et admettons que l'on observe maintenant la série à partir d'un point de départ défini $t = 0$. Par itération, on obtient alors

$$X_t = \theta_1^t X_0 + \sum_{k=0}^{t-1} \theta_1^k \epsilon_{t-k} \tag{7}$$

et, en supposant que la valeur initiale est décorrélée du bruit, on a

$$\mathbb{E}[X_t] = \theta_1^t \mathbb{E}[X_0] \quad \text{et} \quad \mathrm{Cov}(X_t, X_{t+h}) = \theta_1^{2t} \mathrm{Var}(X_0) + \sigma^2 \theta_1^h \frac{1 - \theta_1^{2t}}{1 - \theta_1^2} \tag{8}$$

pour tout $t > 0$ et $h \geqslant 0$. Il est clair que pour $|\theta_1| > 1$, la variance du processus diverge et ne satisfait donc plus les conditions de stationnarité. On observe par ailleurs que, même sous l'hypothèse de causalité ($|\theta_1| < 1$), les moments d'ordre 1 et 2 du processus ne sont pas invariants dans le temps. Dans ce cas (et uniquement dans ce cas), il est toujours possible de choisir la variable initiale X_0 de façon à rendre le processus stationnaire, en utilisant notamment la décomposition issue de (4). Sous l'hypothèse $|\theta_1| < 1$, l'inversion du polynôme autorégressif de degré 1 fournit naturellement la suite des coefficients (ψ_j) tels que $\psi_j = \theta_1^j$ pour tout $j \geqslant 0$. Cette relation étant valable pour tout t, on peut en particulier poser dans L^2

$$X_0 = \sum_{k \geqslant 0} \theta_1^k \epsilon_{-k}$$

quitte à travailler sur $\mathbb{Z}$, ce qui suggère le choix d'une variable X_0 centrée et de variance $\sigma^2/(1-\theta_1^2)$. Auquel cas, le processus (6) défini sur $\mathbb{N}$ correspond à l'extraction de l'AR(1) centré stationnaire défini sur $\mathbb{Z}$ et de covariance,

$$\mathrm{Cov}(X_t, X_{t+h}) = \frac{\sigma^2 \theta_1^h}{1 - \theta_1^2} \quad \forall\, h \geqslant 0. \tag{9}$$

Sinon, indépendamment de la valeur fixée au départ (sous condition d'existence du moment d'ordre 2), le processus peut être considéré comme *asymptotiquement stationnaire* (voir par exemple Lütkepohl (2005)) dans le sens où le processus atteint un régime stationnaire à partir d'un certain rang, i.e. l'espérance et l'autocovariance du processus établies

dans (8) vérifient

$$\lim_{t\to\infty} \mathbb{E}[X_t] = 0 \quad \text{et, pour tout } h \geqslant 0, \quad \lim_{t\to\infty} \mathrm{Cov}(X_t, X_{t+h}) = \frac{\sigma^2\,\theta_1^h}{1-\theta_1^2}.$$

Évidemment, ce raisonnement n'est plus valable si $|\theta_1| > 1$. Ainsi, la notion de stationnarité sur $\mathbb{N}$ n'est pas toujours un cas particulier de la stationnarité sur $\mathbb{Z}$, c'est pourquoi, dans cet axe de recherche, nous adopterons la terminologie de Duflo (1997).

Soient $\Phi_t = (X_t, \ldots, X_{t-p+1})^\top$ et $E_t = (\epsilon_t, 0, \ldots, 0)^\top$ deux vecteurs de taille p, l'écriture vectorielle du processus autorégressif (2) est

$$\Phi_t = A\Phi_{t-1} + E_t \tag{10}$$

où

$$A = \begin{pmatrix} \theta_1 & \theta_2 & \cdots & \theta_p \\ & I_{p-1} & & 0 \end{pmatrix}$$

est la matrice compagne associée au processus [13]. Soit $\rho(A)$ le rayon spectral de cette matrice, on désignera par $0 < |\lambda_p| \leqslant \cdots \leqslant |\lambda_1| = \rho(A)$ les valeurs propres de A rangées de 1 à p selon la valeur de leur module. En effet, les valeurs propres de A sont non nulles puisque $\det(A) = 0$ impliquerait $\theta_p = 0$, ce qui contredit le postulat d'écriture minimale du modèle. D'après la Définition 2.3.17 de Duflo (1997), le processus est dit *stable* lorsque $\rho(A) < 1$, et *instable* quand $\rho(A) = 1$, autrement dit une ou plusieurs valeurs propres sont situées sur le cercle unité tandis que les autres se trouvent à l'intérieur. Il est qualifié de *purement instable* lorsque l'ensemble des valeurs propres sont sur le cercle unité, et *explosif* quand $|\lambda_k| > 1$ pour au moins un k, et *purement explosif* si $|\lambda_k| > 1$ pour tout $k = 1, \ldots, p$. On vérifie aisément qu'un polynôme autorégressif causal engendre un processus stable et, de même, un processus AR instable présentera un comportement non stationnaire en raison d'une (ou plusieurs) racine unitaire. En effet, on obtient immédiatement (par exemple par une récurrence sur p) que pour tout $\lambda \in \mathbb{C}^*$,

$$\det(A - \lambda I_p) = (-\lambda)^p\,\Theta(\lambda^{-1}).$$

Il en découle naturellement que si λ est valeur propre de A, alors λ^{-1} est racine du polynôme autorégressif Θ et donc $\rho(A) < 1$ est équivalent à l'hypothèse de causalité.

13. Sauf mention contraire, I_k fera référence à la matrice identité d'ordre k dans l'ensemble de ce travail.

Inférence des processus autorégressifs stables

Dans cette section, le processus AR(p) étudié est supposé stable et donc issu d'un polynôme autorégressif causal.

En pratique, on observe une série de n observations $X_1, \ldots, X_n$ que l'on souhaite modéliser à partir d'un processus AR(p) tel que $X_t = \theta^\top \Phi_{t-1} + \epsilon_t$ pour tout $t \geqslant 1$ où l'on suppose que la valeur initiale Φ_0 est dans L^2 et décorrélée du bruit. Un estimateur usuel de θ est l'estimateur des moindres carrés (OLS) qui s'obtient par minimisation de

$$\sum_{t=1}^{n} \epsilon_t^2 = \sum_{t=1}^{n} (X_t - \theta^\top \Phi_{t-1})^2$$

et dont la valeur est directement donnée par la formule

$$\hat{\theta} = S_n^{-1} \sum_{t=1}^{n} \Phi_{t-1} X_t \quad \text{avec} \quad S_n = \sum_{t=1}^{n} \Phi_{t-1} \Phi_{t-1}^\top.$$

Il est toujours possible d'ajouter une légère pénalité $\xi I_p > 0$ à la matrice de covariance empirique S_n pour assurer son inversibilité sans pour autant modifier son comportement asymptotique. La consistance (faible ou forte) de cet estimateur a été largement étudiée dans la littérature. On peut citer notamment le théorème 1 de Lai et Wei (1983) qui fournit, entre autres, la consistance forte de l'OLS dans notre cadre soit

$$\hat{\theta} \xrightarrow{a.s.} \theta.$$

Dans le cas stable, et sous l'hypothèse d'existence du moment d'ordre 2, on a le résultat d'ergodicité (voir Brockwell et Davis (1991))

$$\frac{S_n}{n} \xrightarrow{a.s.} \Gamma_p$$

où la matrice des covariances d'ordre p est donnée par

$$\Gamma_p = \begin{pmatrix} \gamma(0) & \gamma(-1) & \cdots & \gamma(1-p) \\ \gamma(1) & \gamma(0) & & \vdots \\ \vdots & & \ddots & \vdots \\ \gamma(p-1) & \cdots & \cdots & \gamma(0) \end{pmatrix}.$$

Rappelons que $\gamma(.)$ est la fonction d'autocovariance du processus stationnaire (X_t) qui

vérifie $\gamma(h) = \gamma(-h)$, la matrice Γ_p est donc symétrique. Via (10), on peut considérer le vecteur aléatoire Φ_t comme un processus $\mathrm{VAR}_p(1)$ stable, l'ensemble des valeurs propres de A étant à l'intérieur du disque unité. D'après Lütkepohl (2005), la suite $(A^k)_{k\geqslant 0}$ est absolument sommable (voir Section A.9.1) et

$$\lim_{n\to\infty} A^n = 0,$$

ce qui permet d'obtenir (voir Proposition C.9) la décomposition $\mathrm{MA}(\infty)$ suivante

$$\Phi_t = \sum_{k\geqslant 0} A^k E_{t-k}$$

qui mène directement à une autre représentation de la matrice Γ_p donnée par

$$\Gamma_p = \mathbb{E}[\Phi_t \Phi_t^\top] = \sigma^2 \sum_{k\geqslant 0} A^k K_p (A^\top)^k \tag{11}$$

où K_p est une matrice carrée de taille $p \times p$ ayant son premier coefficient égal à 1 et tous les autres nuls [14]. En utilisant l'écriture autorégressive, il est également possible d'établir une relation entre l'autocovariance du processus et ses paramètres, plus connue sous le nom d'*équations de Yule-Walker*,

$$\gamma(0) = \theta^\top \begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(p) \end{pmatrix} + \sigma^2 \quad \text{et} \quad \begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(p) \end{pmatrix} = \Gamma_p \theta.$$

Ainsi, il est possible de déterminer entièrement la matrice de covariance à partir de l'estimation de θ et σ^2 et inversement, la fonction de covariance empirique, fortement consistante pour les processus stationnaires (voir Brockwell et Davis (1991)), définie par

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} X_t X_{t+h} \quad \text{pour tout} \quad 0 \leqslant h < n$$

permet d'obtenir l'estimateur de Yule-Walker de θ et σ^2 qui partage d'ailleurs les mêmes propriétés asymptotiques que l'OLS.

En présence d'un $\mathrm{AR}(p)$ stable, l'estimateur des moindres carrés est asymptotiquement normal et converge à vitesse n comme le montre, par exemple, la proposition 8.10.1 de

14. Voir (1.2.3).

Brockwell et Davis (1991) qui stipule

$$\sqrt{n}\,(\hat{\theta}-\theta) \xrightarrow{d} \mathcal{N}(0,\sigma^2\,\Gamma_p^{-1}).$$

Reprenons l'exemple de l'AR(1) présenté dans (6) en supposant cette fois-ci $|\theta_1|<1$, la variance asymptotique ci-dessus peut alors se calculer aisément à l'aide des équations de Yule-Walker (ou de manière équivalente en utilisant (11) avec $A=(\theta_1)$),

$$\begin{cases} \gamma(0) &= \theta_1\,\gamma(1)+\sigma^2 \\ \gamma(1) &= \gamma(0)\,\theta_1 \end{cases}$$

ce qui donne $\gamma(0)=\sigma^2\,(1-\theta_1^2)^{-1}$ d'où

$$\frac{S_n}{n} \xrightarrow{p} \frac{\sigma^2}{1-\theta_1^2} \quad \text{et} \quad \sqrt{n}\,(\hat{\theta}_1-\theta_1) \xrightarrow{d} \mathcal{N}(0,1-\theta_1^2). \tag{12}$$

Inférence et tests statistiques pour les processus autorégressifs instables

Cette section présente quelques propriétés asymptotiques de l'estimateur des moindres carrés, dans le cadre d'un processus intégré, et évoque le fonctionnement des test de racines unitaires (et de stationnarité) qui en permettent l'utilisation empirique.

Tout d'abord, la consistance forte de l'OLS dans un cadre autorégressif a été démontré par Lai et Wei (1983), et ce indépendamment de la valeur des racines caractéristiques du polynôme AR. En revanche, si celui-ci possède une ou plusieurs racines de module 1, la vitesse de convergence de l'estimateur ainsi que sa distribution asymptotique s'en trouvent modifiées. Chan et Wei (1988) ont même démontré que ces dernières dépendent également de la position des racines sur le cercle unité. En effet, les racines unitaires peuvent être réparties en trois catégories : la racine égale à 1, sans doute la plus étudiée car à l'origine des processus de type *marche aléatoire*, la racine égale à -1 et enfin les racines complexes de module 1 qui interviennent par paires conjuguées dès que les paramètres du modèle sont réels. Cette constatation a servi de point de départ pour le travail de Chan et Wei dans lequel ils supposent que la série chronologique (X_t), définie pour tout $t>0$ par $\phi(L)\,X_t=\epsilon_t$, est issue d'un processus AR instable. Le polynôme autorégressif associé est

alors exprimé de manière à refléter la position des racines sur le disque unité via

$$\phi(z) = (1-z)^{p_+}\,(1+z)^{p_-}\prod_{k=1}^{n_c}\left[1-2\cos(\alpha_k)\,z+z^2\right]^{p_k}\,\Theta(z)$$

où p_+, p_- et $p_1, \ldots, p_{n_c}$, désignent respectivement la multiplicité des racines réelles 1, −1 et des n_c paires de racines complexes conjuguées, $\alpha_k \in]0, \pi[$ pour tout $k = 1, \ldots, n_c$, et $\Theta(z)$ est un polynôme autorégressif causal de degré p. On rappelle que, d'après la terminologie de Duflo (1997), si $p = 0$ le processus est purement instable, sinon, le processus AR est instable d'ordre $\mathbf{P} = p_+ + p_- + 2p_1 + \cdots + 2p_{n_c} + p$. À noter que le bruit blanc considéré par les auteurs est une suite de différences de martingale de variance $\sigma^2 = 1$. On y reviendra plus en détails le moment voulu mais (voir Définition 1.3.5 dans Duflo (1997)) disons simplement qu'une martingale (M_t) est une suite aléatoire intégrable, adaptée à une filtration $\mathcal{F}_t = \sigma(X_s,\, s \leqslant t)$ et qui satisfait

$$\mathbb{E}[M_{t+1}|\mathcal{F}_t] = M_t.$$

Dans un premier temps, Chan et Wei (1988) établissent les propriétés asymptotiques de l'OLS en étudiant séparément chaque catégorie de racines unitaires. Ils introduisent, pour chacune d'elles, des matrices de vitesse adaptées qui garantissent la convergence de la matrice de covariance empirique et permettent d'obtenir la loi asymptotique de l'erreur d'estimation. Ces matrices sont ensuite combinées entre elles pour former une matrice de vitesse diagonale par blocs, nécessaire à l'analyse du cas général. Les auteurs montrent d'ailleurs que, une fois la matrice de covariance empirique renormalisée par cette matrice de vitesse générale, les termes croisés entre composantes stables et instables deviennent asymptotiquement négligeables. Ainsi, soient $\mathbf{\Phi}_t = (X_t, \ldots, X_{t-\mathbf{P}+1})^\top$ et G_n la matrice de vitesse, alors

$$G_n\,Q\,(\sum_{t=1}^{n}\mathbf{\Phi}_t\mathbf{\Phi}_t^\top)\,Q^\top\,G_n^\top \xrightarrow{\;d\;} \Delta$$

avec Q une matrice de taille $\mathbf{P}\times\mathbf{P}$ définie de manière à permettre l'étude séparée des différentes composantes (voir Annexe 1 de Chan et Wei (1988) pour plus de détails) et

$$\Delta = \mathrm{diag}(F_+,\, F_-,\, H_1,\, \ldots,\, H_{n_c},\, \Gamma_p)$$

où F_+, F_-, $H_1, \ldots, H_{n_c}$ sont des fonctionnelles du mouvement brownien et Γ_p est la matrice de variance limite obtenue dans le cas stable (11). De même, soit $\hat{\alpha}$ l'estimateur de α,

le vecteur des paramètres du modèle instable, la distribution asymptotique de l'OLS est telle que

$$(Q^\top G_n^\top)^{-1}(\hat{\alpha} - \alpha) \xrightarrow{d} \mathcal{L}_B$$

où la loi limite $\mathcal{L}_B$ se décompose par bloc de fonctionnelle de type Brownien et un bloc gaussien pour la partie stable.

Reprenons l'exemple d'un AR(1) instable, en considérant $\theta_1 = \pm 1$ et S_n la matrice de covariance empirique, alors $G_n = n^{-1}$, $Q_n = 1$ et

$$\frac{S_n}{n^2} \xrightarrow{d} \sigma^2 \int_0^1 W^2(t)\,\mathrm{d}t \quad \text{et} \quad n(\hat{\theta}_1 - \theta_1) \xrightarrow{d} \frac{\pm \int_0^1 W(t)\,\mathrm{d}W(t)}{\int_0^1 W^2(t)\,\mathrm{d}t} \tag{13}$$

où $(W(t), t \in [0,1])$ est un processus de Wiener standard [15]. Ce résultat a notamment été démontré en amont par White (1958) ou encore Dickey et Fuller (1979), avec $\theta_1 = 1$ et un bruit blanc gaussien de variance 1. Par comparaison avec l'exemple de la section précédente, il est donc clair que non seulement la vitesse de convergence est différente lorsque le processus est instable mais la loi asymptotique de l'estimateur n'est même plus gaussienne.

En pratique, il semble généralement plus judicieux de modéliser une série stationnaire, ce qui implique, dans le cas contraire, un traitement préalable des données. En présence d'une ou plusieurs racines unitaires, on se ramène le plus souvent à un processus stable en différenciant la série autant de fois que nécessaire. On suppose par exemple que les données sont issues du processus (X_t) sous-jacent défini tel que

$$Y_t = (I - L)^d X_t$$

où (Y_t) admet une représentation AR stable (ou plus généralement ARMA) et $d \in \mathbb{N}^\star$ correspond au nombre de différenciation nécessaire pour obtenir une série stationnaire. On dit alors que le processus (X_t) admet une écriture ARI [16] (ou ARIMA lorsque (Y_t) vérifie une équation ARMA). Face à une série empirique, il est donc toujours intéressant de pouvoir évaluer son caractère stationnaire, et en particulier la présence de racine unitaire, afin de garantir une représentation au plus juste des données. Il existe deux tests statistiques assez répandus, ADF et KPSS, qui utilisent tous deux des hypothèses complémentaires.

15. Un processus de Wiener, ou mouvement brownien, est un processus stochastique défini pour tout $t \in \mathbb{R}^+$ tel que $W(0) = 0$ presque sûrement, et dont les incréments sont indépendants et gaussiens, à savoir $W(t) - W(s) \sim \mathcal{N}(0, t-s)$ pour $0 \leqslant s < t$, voir Section 37 de Billingsley (1995).

16. Autoregressive Integrated.

A l'origine du test ADF, Dickey et Fuller (1979) étendent d'abord l'analyse de la distribution asymptotique d'un AR(1) instable en considérant également un modèle non centré ainsi qu'une éventuelle tendance temporelle. Ils établissent alors une statistique de test sous $\mathcal{H}_0$: "$\rho = 1$" pour le modèle

$$X_t = \mu + \beta\, t + \rho\, X_{t-1} + \epsilon_t \quad \text{où} \quad (\epsilon_t) \overset{iid}{\sim} \mathcal{N}(0,1) \tag{14}$$

en considérant successivement le cas centré ($\mu = \beta = 0$), avec constante ($\mu \neq 0$, $\beta = 0$), et avec constante et tendance linéaire ($\mu \neq 0$, $\beta \neq 0$). En cas de rejet de l'hypothèse nulle, l'utilisateur est amené à choisir entre un processus stable ($\rho < 1$) ou explosif ($\rho > 1$), selon la forme de la zone de rejet. Plus tard, Said et Dickey (1981) reprendront cette idée pour en faire une version "augmentée" prenant un compte un modèle plus large tel que

$$(I - \rho L)\, \Theta(L) X_t = \Psi(L)\epsilon_t \tag{15}$$

où $\Theta(.)$ et $\Psi(.)$ sont deux polynômes autorégressifs causaux d'ordre p et q. Sous $\mathcal{H}_0$: "$\rho = 1$", le processus (X_t) suit donc un modèle ARIMA$(p, 1, q)$. Bien que cette généralisation du test présente un intérêt notable pour la caractérisation des séries chronologiques, elle comporte néanmoins un inconvénient majeur : les degrés p et q des polynômes sont supposés connus. Or, en pratique, il semble plus pertinent (voir par exemple le Chapitre 6 de Box et al. (2015)) de tester la présence de racine unitaire avant d'estimer ces valeurs. Dans un second temps, le papier de Said et Dickey (1984) a permis de résoudre ce problème. En effet, en supposant la causalité de l'opérateur $\Psi(.)$, le processus ARMA sous-jacent est qualifié d'*inversible* (voir Théorème 3.1.2 de Brockwell et Davis (1991)) et ils suggèrent alors de tronquer l'écriture AR(∞) à un ordre k_n dépendant du nombre n d'observations. Ils montrent que $n^{1/r} \lesssim k_n \ll n^{1/3}$ pour un $r > 0$ permet de conserver les propriétés théoriques de ce test. Il n'est alors plus nécessaire de connaître p et q pour tester l'hypothèse nulle $\mathcal{H}_0$: "$\rho = 1$". Le test ADF permet donc de rejeter, ou non, la présence de racine unitaire pour un modèle de type ARIMA(p, d, q), ce qui revient à décider si $d = 0$ (ARMA stationnaire) ou $d = 1$, en prenant éventuellement en compte la présence d'une constante voire d'une tendance linéaire associée.

Cette avancée a bien sûr donné lieu à de nouvelles évolutions, on peut citer notamment Phillips (1987a) et Phillips et Perron (1988), ainsi que diverses autres approches reposant toutes sur une hypothèse nulle similaire visant à tester la présence de racine unitaire plutôt que la stationnarité. Il devient vite évident que, pour beaucoup de séries

économiques testées, l'hypothèse nulle n'est pas rejetée. Nelson et Plosser (1982) montrent par exemple que, parmi 14 séries macroéconomiques américaines, une seule apparaît stationnaire d'après différents tests type ADF. Ces résultats mènent à deux interprétations possibles : soit la plupart des séries économiques sont véritablement intégrées, soit les tests ne sont pas assez puissants pour rejeter l'hypothèse nulle de racine unitaire lorsque la stationnarité de la série est moins explicite. C'est sur ce constat que Kwiatkowski et al. (1992) élaborent leur test, en positionnant cette fois-ci la stationnarité au cœur de l'hypothèse nulle et la présence de racine unitaire comme hypothèse alternative. Le modèle étudié,

$$X_t = \xi\, t + r_t + \epsilon_t,$$

représente la somme d'une tendance linéaire $\xi\, t$, d'une marche aléatoire (r_t) définie par

$$r_t = r_{t-1} + u_t$$

où (u_t) est un processus centré, iid et de variance σ_u^2 et d'une erreur stationnaire (ϵ_t) qui peut englober des phénomènes de dépendance. Sous $\mathcal{H}_0$: "$\sigma_u^2 = 0$", le processus (X_t) est bien stationnaire (ou stationnaire en tendance autour d'une constante r_0 quand $\xi \neq 0$) alors que sous $\mathcal{H}_1$: "$\sigma_u^2 > 0$", le processus est non stationnaire mais $\left((I - L)\, X_t\right)$ est stationnaire autour de ξ et se comporte comme un MA(1). En comparant leurs résultats à ceux de Nelson et Plosser (1982), il apparaît que les tests concordent pour cinq des séries étudiées. Trois autres séries semblent potentiellement intégrées, bien que l'hypothèse de stationnarité ne soit pas systématiquement rejetée selon les paramètres retenus. Quant aux autres, les tests de racine unitaire et de stationnarité échouent à rejeter leur hypothèse nulle respective, aboutissant à des conclusions contradictoires. Leybourne et McCabe (1994) reprendront ce modèle avec (ϵ_t) un bruit blanc et les phénomènes de dépendance inclus dans une partie AR(p), ce qui facilite l'analyse mais nécessite la connaissance de l'ordre p. La combinaison de ces deux types de test permet souvent une bonne appréciation de la stationnarité de la série ou la détection d'une éventuelle composante intégrée mais il subsiste tout de même des situations douteuses pour lesquelles ces deux tests ne semblent pas adaptés. Ainsi, à la lumière du mécanisme même de ces tests, il paraît évident qu'une racine unitaire différente de 1, notamment -1, ne sera pas décelée par ces deux méthodes (ceci apparaît clairement dans la Figure 1). Une situation corrigée par l'étude de Proïa (2018) qui reprend le test de Leybourne et McCabe en y incluant la possibilité de multi-intégration et de racine unitaire négative. De plus, DeJong et al. (1992)

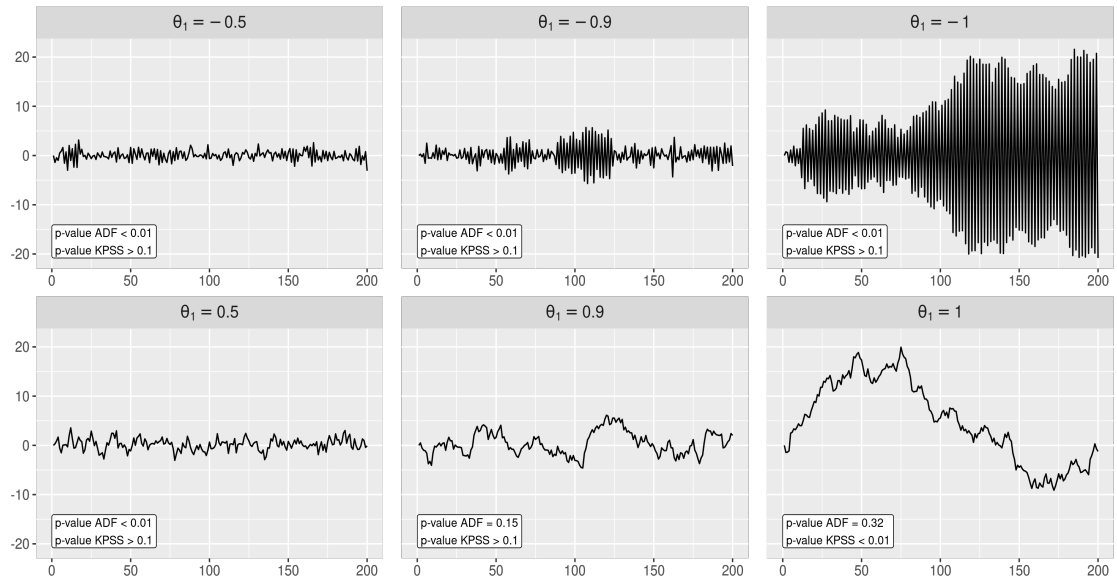


FIGURE 1 – Représentation de $n = 200$ observations du processus $X_t = \theta_1 X_{t-1} + \varepsilon_t$ avec $\theta_1 \in \{\pm 0.5, \pm 0.9, \pm 1\}$ et $\varepsilon_t \sim \mathcal{N}(0, 1)$, et les résultats des tests ADF et KPSS associés.

ont montré que les tests de Dickey-Fuller sont de plus faibles puissances lorsque la racine s'approche de l'unité, tout en restant strictement à l'intérieur du cercle unité. Ils essaient également une approche empirique en réalisant parallèlement le test de racine unitaire $\mathcal{H}_0$: "$\rho = 1$" et le test de stationnarité $\mathcal{H}_0$: "$\rho = 0.85$" sur un modèle de régression similaire à (14). Néanmoins, il demeure difficile de fixer arbitrairement la valeur à retenir pour ce deuxième test. Ce choix met en évidence l'existence d'une zone d'incertitude autour de la frontière entre stationnarité et intégration, ouvrant la voie à l'analyse de processus dits quasi-instables.

Processus autorégressifs quasi-instables

L'analyse empirique des séries temporelles, notamment dans les domaines économique et financier, révèle une frontière plus ou moins floue entre stationnarité et intégration, ce qui est particulièrement visible à la lumière des tests existants. On peut étudier quelques exemples de séries chronologiques issues de processus AR(1) stables et instables (voir Figure 1) parmi lesquels on peut considérer les processus intermédiaires ($|\theta_1| = 0.9$) comme quasi-instables. Du point de vue de la trajectoire en elle-même, on note quelques particularités intéressantes. Ainsi, plus on s'approche de la racine unitaire, plus la variance du

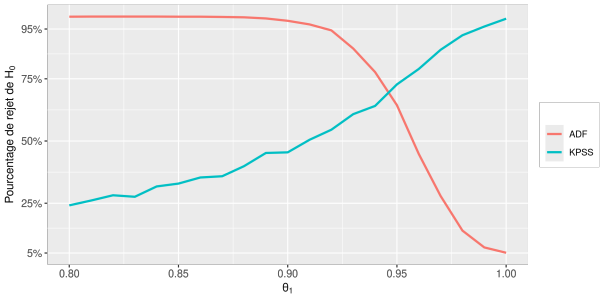


FIGURE 2 – Pourcentage de rejet de $\mathcal{H}_0$ (au seuil de 5%) des tests ADF et KPSS sur des représentations de taille $n = 500$ du processus $X_t = \theta_1 X_{t-1} + \varepsilon_t$ avec $\varepsilon_t \sim \mathcal{N}(0, 1)$.

processus augmente, mais de manière très différente selon le signe de cette racine, avec des oscillations fortes autour de la moyenne pour le cas négatif et le début d'un comportement de marche aléatoire dans le cas positif. De plus, comme évoqué précédemment, les deux tests ne parviennent pas à détecter la racine unitaire négative et s'accordent à dire que la série est stationnaire alors qu'elle est clairement intégrée. De même, on s'aperçoit qu'une racine positive proche du cercle unité, tout en restant strictement < 1, amène des résultats contradictoires sur les tests de stationnarité et racine unitaire. Le test KPSS ne rejette pas la stationnarité tandis que le test ADF échoue à rejeter l'hypothèse nulle de racine unitaire. Bien que les tests soient probants dans les séries restantes, l'incertitude révélée ici n'est pas un cas isolé. La Figure 2 propose les pourcentages de rejet de $\mathcal{H}_0$ des tests ADF et KPSS, effectués sur $n = 500$ observations d'un processus AR(1) dont le paramètre unique varie entre 0.80 et 1 par pas de 0.1. Pour chaque valeur de θ_1, les pourcentages sont calculés sur la base de 2000 représentations. On observe qu'une série AR(1) sur 2 conduit au rejet de l'hypothèse de stationnarité du test KPSS dès que $\theta_1 > 0.90$ tandis que le test ADF échoue à rejeter l'hypothèse nulle de racine unitaire pour près d'une série sur 3 dès que $\theta_1 > 0.95$. D'après Chan et Wei (1987), cette zone d'incertitude, à cheval entre une loi asymptotique normale d'un côté et asymétrique de l'autre, peut impacter la fiabilité de l'estimation, surtout lorsque la taille de l'échantillon est petite. L'étude des processus quasi-instables vient en réponse à cette problématique, dans l'objectif d'établir un pont entre stabilité et instabilité.

Présentons maintenant plus en détails les articles ayant servi de socle à notre réflexion. En premier lieu, le travail de Chan et Wei (1987) dans lequel ils proposent un processus

AR(1) écrit sous la forme triangulaire suivante

$$X_{n,t} = \rho_n X_{n,t-1} + \epsilon_t \tag{16}$$

où (ϵ_t) est par exemple une suite iid de variables centrées et de variance finie [17] et $\rho_n = 1 - c/v_n$, avec $c \in \mathbb{R}$ une constante fixe et $v_n = n$. Le paramètre AR évolue donc vers la frontière unité lorsque n grandit. On parle de processus stable mais quasi-instable pour $c > 0$ (processus stable à n fixé mais instable en limite) ou explosif quasi-instable dans le cas contraire. On notera que pour $c = 0$, le processus est purement instable. L'écriture du processus quasi-instable ci-dessus admet une double indexation qui assez fréquente lorsqu'on utilise un modèle dont l'un des paramètres varie au cours du temps. Le processus est alors traité comme un tableau triangulaire,

$$\begin{array}{c|c|cccccc} n=1 & X_{1,0} & X_{1,1} & & & & & \\ n=2 & X_{2,0} & X_{2,1} & X_{2,2} & & & & \\ n=3 & X_{3,0} & X_{3,1} & X_{3,2} & X_{3,3} & & & \\ \vdots & \vdots & \vdots & & & \ddots & & \\ n & X_{n,0} & X_{n,1} & \dots & & & X_{n,n} \end{array}$$

qui prend en compte la réactualisation du modèle pour chaque n. En supposant $X_{n,0} = 0$ pour tout n, les auteurs montrent que l'erreur d'estimation auto-normalisée converge vers une distribution limite non Gaussienne qui dépend entièrement de c. Pour $c = 0$, on retrouve par exemple la loi asymptotique correspondant à un AR(1) instable de racine unitaire 1. D'après leur Théorème 2, cette loi devient asymptotiquement gaussienne dès que $|c| \to \infty$, ce qui a également été démontré dans Phillips (1987a) à partir d'un modèle légèrement différent. Ce résultat est attendu dans le sens où, pour une constante $c > 0$ suffisamment grande, le modèle devient stable et l'erreur d'estimation admet alors une distribution limite normale. On notera que ce théorème est aussi vrai dans le cas $c \to -\infty$ qui correspond au cas explosif quasi-instable, ce cadre ne sera toutefois pas étudié durant cette thèse. Bien que le modèle proposé dans ce travail offre une nouvelle perspective sur les processus AR quasi-instables, la suite $v_n = n$ ne semble pas adaptée pour établir un pont entre stable et instable puisque la vitesse de convergence, trop rapide, amène directement à la distribution asymptotique typique du cadre instable. Giraitis et Phillips (2006) reprendront le modèle (16) et développeront une théorie limite normale uniforme par rap-

17. Les auteurs admettent même des conditions moins strictes sur le bruit.

port au coefficient autorégressif ρ_n sur lequel ils imposeront la condition $n(1-\rho_n) \to \infty$. Cette théorie sera complétée par Phillips et Magdalinos (2007) dans un papier qui a servi de référence pour le travail présenté dans la section suivante. Les auteurs considèrent ainsi $\rho_n = 1 - c/v_n$ où $(v_n)_{n\in\mathbb{N}}$ est une suite divergente vers $+\infty$ telle que $v_n = o(n)$. Cette dernière hypothèse va dans le sens de Giraitis et Phillips (2006) et permet de ralentir la convergence vers la racine unitaire. Les auteurs proposent notamment $v_n = n^\alpha$ $(0 < \alpha < 1)$, un choix qui relie le cas stable, quand $\alpha \to 0$ (ou explosif selon le signe de c), et le cas instable, quand $\alpha \to 1$. Comme indiqué précédemment, nous nous limiterons à présenter les résultats obtenus pour le processus AR(1) stable quasi-instable, soit $c > 0$, même si les auteurs se sont aussi intéressés au cas $c < 0$. L'analyse des propriétés asymptotiques de l'erreur d'estimation

$$\hat{\rho}_n - \rho_n = S_{n,n-1}^{-1} \sum_{t=1}^{n} X_{n,t-1}\, \epsilon_t \tag{17}$$

s'articule alors en deux étapes distinctes. On rappelle que, dans ce cadre,

$$\hat{\rho}_n = S_{n,n-1}^{-1} \sum_{t=1}^{n} X_{n,t-1}\, X_{n,t} \quad \text{où} \quad S_{n,n-1} = \sum_{t=1}^{n} X_{n,t-1}^2.$$

Dans un premier temps, la loi des grands nombres leur permet d'établir la consistance de la variance empirique soit

$$(1-\rho_n)\, \frac{S_{n,n-1}}{n} \xrightarrow{\ p\ } \frac{\sigma^2}{2} \tag{18}$$

où $\sigma^2 < \infty$ est la variance du bruit (ϵ_t), supposé iid, centré, et tel que $\mathbb{E}|\epsilon_1|^{2+\delta} < \infty$ pour un $\delta > 0$, cette dernière contrainte étant surtout nécessaire pour ce qui suit. Les conditions initiales sont également supposées indépendantes du bruit et telles que $X_{n,0} = o(v_n^{1/2})$ pour tout $n \in \mathbb{N}$. Dans un deuxième temps, ils démontrent la normalité asymptotique du numérateur dans (17) en appliquant le théorème central limite pour les tableaux de différences de martingale de Pollard (1984), voir Section 1 du Chapitre VIII. Soient $m_{n,1}, \ldots, m_{n,n}$, des variables aléatoires intégrables adaptées à la filtration formée par les σ-algèbres $\mathcal{F}_{n,0} \subseteq \cdots \subseteq \mathcal{F}_{n,n}$, on dit que $(m_{n,t})_{t\geqslant 1}$ est une suite de différences de martingale si elle vérifie, pour tout t,

$$\mathbb{E}[m_{n,t} \,|\, \mathcal{F}_{n,t-1}] = 0.$$

Le théorème central limite évoqué ci-dessus implique que si

$$(i)\ \sum_{t=1}^{n} \mathbb{E}[m_{n,t}^2 \,|\, \mathcal{F}_{n,t-1}] \ \xrightarrow{\ p\ }\ \sigma^2,$$
$$(ii)\ \forall\, \eta > 0,\ \sum_{t=1}^{n} \mathbb{E}[m_{n,t}^2\, \mathbb{1}_{\{|m_{n,t}|>\eta\}} \,|\, \mathcal{F}_{n,t-1}] \ \xrightarrow{\ p\ }\ 0 \text{ (condition de Lindeberg)},$$

alors

$$\sum_{t=1}^{n} m_{n,t} \ \xrightarrow{\ d\ }\ \mathcal{N}(0, \sigma^2).$$

En posant maintenant

$$m_{n,t} = \left(\frac{1-\rho_n}{n}\right)^{1/2} X_{n,t-1}\, \epsilon_t$$

pour tout $n \geqslant 1$ et $1 \leqslant t \leqslant n$, il est clair que $(m_{n,t})$ est une suite de différences de martingale adaptée à la filtration $\mathcal{F}_{n,t} = \sigma(X_{n,0}, \epsilon_1, \ldots, \epsilon_t)$ puisque $X_{n,t-1}$ est $\mathcal{F}_{n,t-1}$-mesurable et (ϵ_t) est centré. De même, on obtient facilement que

$$\sum_{t=1}^{n} \mathbb{E}[m_{n,t}^2 \,|\, \mathcal{F}_{n,t-1}] = \sigma^2(1-\rho_n)\, \frac{S_{n,n-1}}{n} \ \xrightarrow{\ p\ }\ \frac{\sigma^4}{2},$$

ce qui, ajouté à la condition de Lindeberg démontrée en annexe de Phillips et Magdalinos (2007), conduit à

$$\frac{1}{\sqrt{n\, v_n}} \sum_{t=1}^{n} X_{n,t-1}\, \epsilon_t \ \xrightarrow{\ d\ }\ \mathcal{N}(0, \frac{\sigma^4}{2c}), \tag{19}$$

en se rappelant que $1 - \rho_n = c/v_n$. Le lemme de Slutsky associé à (18) et (19) fournit alors la normalité asymptotique recherchée, à savoir

$$\sqrt{n\, v_n}\,(\hat{\rho}_n - \rho_n) \ \xrightarrow{\ d\ }\ \mathcal{N}(0, 2c).$$

Avec une hypothèse supplémentaire de symétrie sur la distribution du bruit, les auteurs retrouvent le même résultat pour $\rho_n = -1 + c/v_n$, soit une racine unitaire négative à la limite. Les résultats ainsi obtenus permettent bien d'établir un pont entre la vitesse de convergence dans le cas stable et celle du cas instable. Le phénomène est d'autant plus prononcé lorsque $v_n = n^\alpha$, avec $0 < \alpha < 1$, dans la mesure où la vitesse de convergence est alors donnée par $n^{1/2+\alpha/2}$. En laissant $\alpha \to 0$, le processus devient stable et on retrouve une vitesse en $\sqrt{n}$, tandis que du côté instable, $\alpha \to 1$ et l'erreur d'estimation converge à vitesse n. On remarque cependant quelques discontinuités autour de ces deux limites. Posons par exemple $\alpha = 0$ et étudions la limite du processus AR(1) stable dont le paramètre est donc

tel que $\theta_1 = 1 - c$ (pour $0 < c < 2$). D'après (12), on a la normalité asymptotique de l'OLS dans le cas stable donnée par

$$\sqrt{n}\,(\hat{\theta}_1 - \theta_1) \xrightarrow{d} \mathcal{N}(0, 2c - c^2).$$

Lorsque $\alpha \to 0$, la variance asymptotique est donc surestimée dans la convergence obtenue par Phillips et Magdalinos (2007). De manière encore plus flagrante, mais sans surprise, on observe une rupture entre cette même convergence, lorsque $\alpha \to 1$, et la loi asymptotique obtenue dans le cadre d'un processus instable qui n'est plus du tout gaussienne.

A noter que, dans ce cadre quasi-instable, des principes de déviations modérées ont été établies pour $p = 1$ par Miao et al. (2015) et pour p quelconque par Proïa (2020). Pour ce qui est de la distribution asymptotique de l'OLS, les modèles présentés ci-dessus ont tous été étudié en considérant $p = 1$. Dans cette thèse, nous nous proposons donc d'étudier le cas général en reprenant notamment les techniques de preuves liées aux martingales. On peut néanmoins mentionner le travail de Park (2003) qui propose une étude assez générale des modèles de séries temporelles avec une racine proche de l'unité, incluant les processus AR(p) générés par

$$X_{n,t} = \beta_n X_{n,t-1} + V_{n,t}$$

où $\beta_n = 1 - \frac{m}{n}$, tel que $m \to \infty$ et $m/n \to 0$, et $(V_{n,t})$ est un processus AR($p-1$) stable. Ce modèle, analogue à celui utilisé par Phillips et Magdalinos (2007), conduit à des résultats similaires et l'auteur suggère même une procédure de test dans ce cadre quasi-instable. Cependant, l'une des hypothèses semble assez restrictive puisqu'elle impose que tous les coefficients convergent à la même vitesse. De plus, la convergence vers une racine unitaire négative n'est pas envisagée. L'article présenté en Chapitre 1 généralise donc ces résultats à un cadre plus large. L'approche de Park (2003), et en particulier la décomposition ci-dessus, a toutefois servi de base au développement d'un test statistique pour l'évaluation de la quasi-instabilité, développé dans le Chapitre 2.

Il existe aujourd'hui un intérêt croissant pour les processus quasi-instables dans la littérature, abordés selon des approches méthodologiques assez variées comme en témoignent les références suivantes. Ainsi, Phillips et Lee (2015) reprendront l'idée de Phillips et Magdalinos (2007) pour l'appliquer à des processus VAR(1) bivariés possédant des racines mixtes proches du cercle unité. Dans une autre approche, Buchmann et Chan (2013) travaillent sur un AR(p) quasi-instable où la perturbation est introduite dans une nouvelle version de la décomposition de Jordan de la matrice compagne. On peut également

citer Horváth et Trapani (2016) qui montrent que, pour un modèle autorégressif à coefficients aléatoires (RCA [18]), l'estimateur des moindres carrés pondérés est asymptotiquement normal peu importe la valeur moyenne de la racine, et concluent donc sur l'absence de problème de racine unitaire pour les modèles de type RCA. Trapani (2021) proposera également un test de stationnarité stricte pour ce type de modèle.

Contributions de la thèse

L'objectif principal de ce premier axe de recherche est donc d'étendre l'étude des propriétés asymptotiques de l'OLS pour des processus autorégressifs quasi-instables d'ordre p quelconque. A partir de ce travail, un test statistique a été développé pour permettre d'évaluer, d'une certaine manière, le degré d'instabilité d'une série. Il permet notamment, pour un certain type de modèle, de mesurer la proximité de la plus grande valeur propre (en module) au cercle unité.

Le résultats obtenus dans ce cadre ont été établi en collaboration avec Frédéric Proïa et ont donné lieu à deux articles présentés en Chapitre 1, Badreau et Proïa (2023), publié dans *Stastical Inference for Stochastic Processes*, et en Chapitre 2, Badreau et Proïa (2025), publié dans *Journal of Time Series Analysis*.

Résumé du Chapitre 1

On considère dans ce travail un processus autorégressif d'ordre $p \geqslant 1$ généré par la forme triangulaire suivante

$$\forall\, n \geqslant 1,\ \forall\, 1 \leqslant k \leqslant n, \qquad X_{n,k} = \sum_{i=1}^{p} \theta_{n,i}\, X_{n,k-i} + \varepsilon_k \tag{20}$$

et dont l'écriture vectorielle associée est donnée par

$$\Phi_{n,k} = A_n\, \Phi_{n,k-1} + E_k \tag{21}$$

où $\Phi_{n,k} = (X_{n,k}, \ldots, X_{n,k-p+1})^\intercal$ et $E_k = (\varepsilon_k, 0, \ldots, 0)^\intercal$ sont des vecteurs de taille p avec (ε_k) une suite iid centrée et de variance $\sigma^2 > 0$ et A_n est la matrice compagne associée au

18. *Random Coefficient Autoregressive*

processus, de taille $p \times p$, telle que

$$A_n = \begin{pmatrix} \theta_{n,1} & \theta_{n,2} & \cdots & \theta_{n,p} \\ & I_{p-1} & & 0 \end{pmatrix}. \tag{22}$$

On suppose que la valeur initiale $\Phi_{n,0}$ admet un moment d'ordre 2 fini et est indépendant du processus (E_k). Avant de présenter les principaux résultats de ce papier, introduisons tout d'abord les trois hypothèses majeures nécessaires à leur établissement.

(H_1) *Convergence de la matrice compagne.* La matrice A_n converge vers une matrice A de taille $p \times p$ dont les valeurs propres sont toutes distinctes et triées par ordre décroissant des modules (et par ordre lexicographique en cas d'égalité), c'est-à-dire $0 < |\lambda_p| \leqslant \ldots \leqslant |\lambda_2| \leqslant |\lambda_1| = \rho(A) = 1$ où $\rho(A)$ est le rayon spectral de A. De plus, l'élément en haut à droite de A est non-nul (i.e. $\theta_p \neq 0$).

(H_2) *Nombre de racines unitaires.* On admet soit (H_{21}) ou (H_{22}).

(H_{21}) La matrice A possède une seule racine unitaire ($\lambda_1 = \pm 1$ et $|\lambda_2| < 1$ si $p \geqslant 2$).

(H_{22}) La matrice A possède deux racines unitaires ($\lambda_1 = 1$, $\lambda_2 = -1$ et $|\lambda_3| < 1$ si $p \geqslant 3$).

(H_3) *Rayon spectral de la matrice compagne.* Le rayon spectral de A_n est donné par

$$\rho_n \triangleq \rho(A_n) = 1 - \frac{c}{v_n}$$

pour un $c > 0$ et une suite (v_n) telle que $1 \ll v_n \ll n$.

La première hypothèse simplifie les raisonnements à venir puisqu'elle assure, comme dans le travail de Proïa (2020), que la matrice A_n est diagonalisable à partir d'un certain rang. En particulier, il existe $n_0 \geqslant 1$ tel que, pour tout $n > n_0$, A_n est diagonalisable sous la forme $A_n = P_n D_n P_n^{-1}$ où $D_n = \mathrm{diag}(\lambda_{n,1}, \ldots, \lambda_{n,p})$ contient les valeurs propres distinctes et ordonnées $1 > |\lambda_{n,1}| \geqslant \ldots \geqslant |\lambda_{n,p}| > 0$ de A_n qui convergent vers les valeurs propres de A données dans (H_1). De plus, la matrice de passage P_n peut être exprimée comme la matrice de Vandermonde suivante,

$$P_n = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \frac{1}{\lambda_{n,1}} & \frac{1}{\lambda_{n,2}} & \cdots & \frac{1}{\lambda_{n,p}} \\ \vdots & \vdots & & \vdots \\ \frac{1}{\lambda_{n,1}^{p-1}} & \frac{1}{\lambda_{n,2}^{p-1}} & \cdots & \frac{1}{\lambda_{n,p}^{p-1}} \end{pmatrix}$$

où $P_n \to P$ et $P_n^{-1} \to P^{-1}$. La connaissance d'une partie de cette dernière matrice se

révélera essentielle dans l'établissement de la matrice de précision liée à la normalité asymptotique de l'OLS. La dernière hypothèse est, quant à elle, directement liée au cadre quasi-instable établi dans Phillips et Magdalinos (2007). En effet, d'après la Définition 2.3.17 de Duflo (1997), le processus ainsi défini est stable à n fixé mais instable lorsque $n \to \infty$ car alors $\rho(A_n) \to 1$.

On rappelle que l'estimateur OLS de $\theta_n = (\theta_{n,1}, \ldots, \theta_{n,p})^\top$ est donné par

$$\widehat{\theta}_n = S_{n,n-1}^{-1} \sum_{k=1}^{n} \Phi_{n,k-1} \, X_{n,k}$$

où la matrice de covariance empirique est définie par

$$S_n = S_{n,n} = \sum_{k=0}^{n} \Phi_{n,k} \, \Phi_{n,k}^\top.$$

Les propriétés asymptotiques de cette matrice sont données dans la proposition ci-dessous.

Proposition (Consistance de la covariance empirique)**.** *On suppose que (H_1), (H_2) et (H_3) sont vérifiées, et on admet que $\mathbb{E}[\varepsilon_1^2] = \sigma^2 < +\infty$. Alors*

$$(1-\rho_n)\,\frac{S_n}{n} \xrightarrow{\;p\;} \Gamma \quad \textit{et} \quad (1-\rho_n)\left\|\frac{S_n}{n} - \Gamma_n\right\| \xrightarrow{\;p\;} 0$$

où $\Gamma_n = \sigma^2 \sum_{\ell \geqslant 0} A_n^\ell \, K_p \, (A_n^\top)^\ell$ est la covariance du processus stationnaire à n fixé.

La consistance de la covariance empirique est donc établie sous condition de renormalisation par la vitesse $1-\rho_n$ et on peut montrer, par calcul direct, que la matrice limite Γ est égale à

$$\pi_{11}^2 \, \frac{\sigma^2}{2} \begin{pmatrix} 1 & \lambda_1 & \cdots & \lambda_1^{p-1} \\ \lambda_1 & 1 & \cdots & \lambda_1^{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{p-1} & \lambda_1^{p-2} & \cdots & 1 \end{pmatrix}$$

sous (H_{21}) avec $\lambda_1 = \pm 1$ et

$$\frac{\sigma^2}{2} \begin{pmatrix} \pi_{11}^2 + \pi_{21}^2 & \pi_{11}^2 - \pi_{21}^2 & \cdots & \pi_{11}^2 + (-1)^{p-1} \, \pi_{21}^2 \\ \pi_{11}^2 - \pi_{21}^2 & \pi_{11}^2 + \pi_{21}^2 & \cdots & \pi_{11}^2 + (-1)^{p-2} \, \pi_{21}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{11}^2 + (-1)^{p-1} \, \pi_{21}^2 & \pi_{11}^2 + (-1)^{p-2} \, \pi_{21}^2 & \cdots & \pi_{11}^2 + \pi_{21}^2 \end{pmatrix}$$

sous (H_{22}), avec π_{11} et π_{21} les deux premiers éléments de la première colonne de P^{-1} définie plus haut. En notant que $\pi_{11} = 1$ quand $p = 1$, on retrouve immédiatement la convergence établie par Phillips et Magdalinos (2007) dans (18). De même, en posant $v_n = n^\alpha$, on peut établir un parallèle avec les propriétés établies dans le cas stable (12) et instable (13) même si la connexion n'est pas parfaite. En revanche, dès que $p > 1$, la matrice Γ n'est clairement plus inversible sous (H_{21}), et de même sous (H_{22}) dès que $p > 2$. Comme dans Chan et Wei (1988), la vitesse de convergence doit être adaptée à chaque valeur propre afin d'obtenir une matrice limite non singulière. On introduit alors la matrice de vitesse $V_n = \mathrm{diag}(1 - \rho_n, 1, \ldots, 1)$, de taille $p \times p$, et on montre que l'erreur d'estimation satisfait

$$\widehat{\theta}_n - \theta_n = (P_n^\top)^{-1} V_n^{1/2} \left(V_n^{1/2} P_n^{-1} \frac{S_{n,n-1}}{n} (P_n^{-1})^\top V_n^{1/2} \right)^{-1} \left(V_n^{1/2} P_n^{-1} \frac{Z_n}{n} e_p \right)$$

où $Z_n = \sum_{k=1}^n \Phi_{n,k-1} E_k^\top$ est un tableau de martingales, ce que l'on exploitera dans le théorème suivant.

Théorème (Consistance et Normalité asymptotique de l'OLS)**.** *On suppose que (H_1), (H_{21}) et (H_3) sont vérifiées, et on admet que $\mathbb{E}[\varepsilon_1^2] = \sigma^2 < +\infty$. Alors, on a la consistance de l'estimateur OLS*

$$\left\| \widehat{\theta}_n - \theta_n \right\| \xrightarrow{p} 0.$$

On suppose de plus que $\mathbb{E}[|\varepsilon_1|^{2+\nu}] = \eta_\nu < +\infty$ pour un $\nu > 0$ et que les valeurs propres de A sont réelles. Alors, on a la normalité asymptotique

$$\sqrt{n}\, V_n^{-1/2} P_n^\top \left(\widehat{\theta}_n - \theta_n \right) \xrightarrow{d} \mathcal{N}_p(0, H_0^{-1})$$

où H_0 est une matrice définie positive de taille $p \times p$, diagonale par bloc, et dont le premier bloc est égal à $\pi_{11}^2/2$.

A noter que la matrice H_0, composée de deux blocs inversibles, est donnée explicitement dans le Chapitre 1, voir (1.2.2). Ce théorème est également valable sous (H_{22}) à condition d'adapter la vitesse de convergence telle que $V_n = \mathrm{diag}(1-\lambda_{n,1}, 1+\lambda_{n,2}, 1, \ldots, 1)$ et alors H_0 est constituée d'un premier bloc 2×2 égal à $\mathrm{diag}(\pi_{11}^2/2, \pi_{21}^2/2)$ et d'un second bloc précisé dans la Remarque 1.3.2. De même que pour la proposition ci-dessus, et sous des hypothèses légèrement moins restrictives (du côté de la racine unitaire négative), on retrouve le résultat de normalité asymptotique obtenu par Phillips et Magdalinos (2007). En effet, pour $p = 1$, on a $V_n = 1 - \rho_n = c/v_n$, $P_n = 1$ et $H_0 = 1/2$.

Ainsi, les propriétés asymptotiques de l'OLS ont été généralisées dans le cadre d'un AR(p) quasi-instable pour p quelconque. Néanmoins, la matrice de passage P_n, nécessaire à l'établissement de la convergence, constitue un obstacle technique non négligeable. De part la consistance de l'OLS, il est possible d'obtenir des estimateurs consistant des valeurs propres de A_n, constituant les seuls éléments de P_n. En revanche, il paraît plus compliqué de construire un test statistique pour évaluer l'instabilité dans ces conditions. Cette observation a constitué le point de départ de notre second article.

Résumé du Chapitre 2

Dans ce chapitre, le processus triangulaire $(X_{n,k})$ d'ordre p généré par (20), est exprimé, à l'instar de Park (2003), sous la forme hiérarchique suivante

$$\forall\, n \geqslant 1,\ \forall\, 1 \leqslant k \leqslant n, \quad \left\{ \begin{array}{rcl} X_{n,k} & = & \lambda_{n,1}\, X_{n,k-1} + V_{n,k} \\ V_{n,k} & = & \beta_{n,1}\, V_{n,k-1} + \ldots + \beta_{n,p-1}\, V_{n,k-p+1} + \varepsilon_k \end{array} \right.$$

où $\lambda_{n,1}$ est la valeur propre de plus grand module de la matrice compagne A_n, voir (21) et (22), qui contiendra toute la partie "quasi-instable" du processus telle que définie par les hypothèses ci-dessous. En effet, on peut montrer que le processus $(V_{n,k})$ est un processus AR($p-1$) dont la matrice compagne $\bar{A}_n$ converge vers une matrice $\bar{A}$ de rayon spectral strictement plus petit que 1 en module. Le processus $(V_{n,k})$ est donc stable à n fixé et en limite, et sa matrice de covariance (normalisée) est donnée par

$$\Sigma_{p-1} = \sum_{k \geqslant 0} \bar{A}^k\, K_{p-1}\, (\bar{A}^T)^k. \tag{23}$$

En posant $\vartheta_n = (\lambda_{n,1},\, \beta_{n,1},\, \ldots,\, \beta_{n,p-1})^\top$ et $\Psi_{n,k} = (X_{n,k},\, V_{n,k},\, \ldots,\, V_{n,k-p+2})^\top$, le processus $(X_{n,k})$ peut également s'écrire

$$\forall\, n \geqslant 1,\ \forall\, 1 \leqslant k \leqslant n, \quad X_{n,k} = \vartheta_n^\top\, \Psi_{n,k-1} + \varepsilon_k$$

où la valeur initiale $\Psi_{n,0}$ admet un moment fini d'ordre 2. Comme pour l'estimateur OLS, la somme des carrés $\sum_k \|\varepsilon_k\|^2$ dans l'équation précédente est minimisée par

$$\widehat{\vartheta}_n = T_{n,n-1}^{-1} \sum_{k=1}^{n} \Psi_{n,k-1}\, X_{n,k} \quad \text{avec} \quad T_{n,n} = \sum_{k=0}^{n} \Psi_{n,k}\, \Psi_{n,k}^\top. \tag{24}$$

On retrouve les hypothèses techniques suivantes, assez similaires aux travaux précédents.

(A1) *Convergence de la matrice compagne.* La matrice A_n converge vers une matrice A de taille $p \times p$ dont les valeurs propres sont toutes distinctes et ordonnées telles que $0 < |\lambda_p| \leqslant \ldots \leqslant |\lambda_2| < |\lambda_1| = \rho(A) = 1$. On suppose de plus $\theta_p \neq 0$.

(A2) *Une seule racine unitaire.* La matrice A possède une seule racine unitaire, $\lambda_1 = \pm 1$, mais $|\lambda_2| < 1$ si $p \geqslant 2$.

(A3) *Rayon spectral de la matrice compagne.* Le rayon spectral de A_n est donné par

$$\forall\, n \geqslant 1, \quad \rho_n(\alpha) = 1 - \frac{c}{n^\alpha}$$

pour un $c > 0$ et $0 < \alpha < 1$.

L'hypothèse sur le rayon spectral de la matrice A_n permet d'aborder la question du choix de α, en particulier pour évaluer la proximité de la racine à la frontière unité. L'idée est donc d'établir un test statistique de type $\mathcal{H}_0$: "$\alpha = \alpha_0$" contre $\mathcal{H}_1$: "$\alpha > \alpha_0$" pour une certaine valeur test α_0. On notera également l'importance du paramètre α, dont dépend la vitesse de convergence, vis-à-vis du coefficient c correspondant plutôt à un paramètre de variance. Cependant, l'un et l'autre ne peuvent être estimés séparément sans contraintes supplémentaires sur le modèle qui est non identifiable (car sur-paramétré), c'est pourquoi on supposera c connu dans un premier temps avant de proposer une procédure empirique d'estimation jointe de ces deux valeurs.

La valeur α joue ici un rôle primordial dans la formulation de la représentation hiérarchique du modèle et en particulier dans l'expression des variables $V_{n,k}$. Le vecteur $\widehat{\vartheta}_n$ introduit précédemment dans (24) est lui-même défini comme une fonction de α et les propriétés asymptotiques associées sont donc naturellement dépendantes de la valeur qui lui est associée. On soulignera cette dépendance par la notation $\widehat{\vartheta}_n(\alpha)$ lorsque cela sera nécessaire. Le théorème suivant repose sur des techniques de preuves semblables à celles de l'analyse qui précède et nécessite la matrice de vitesse suivante

$$C_n(\alpha) = \operatorname{diag}\left(\frac{1}{1-\rho_n(\alpha)}, 1, \ldots, 1\right).$$

Théorème. *On suppose que (A1), (A2) et (A3) sont vérifiées et que* $\mathbb{E}[|\epsilon_1|^{2+\nu}] = \eta_\nu < +\infty$ *pour un* $\nu > 0$. *Alors, si* α *est la "vraie" valeur,*

$$\sqrt{n}\, C_n^{\frac{1}{2}}(\alpha)\left(\widehat{\vartheta}_n(\alpha) - \vartheta_n(\alpha)\right) \xrightarrow{d} \mathcal{N}\left(0,\, \Upsilon_p^{-1}\right) \tag{25}$$

avec la covariance limite $\Upsilon_p = \operatorname{diag}\left(\pi_{11}^2/2,\, \Sigma_{p-1}\right)$ *où*

$$\pi_{11} = \frac{\lambda_1^{p-1}}{\prod_{j=2}^{p}\left(1-\lambda_j\right)}$$

et Σ_{p-1} *est définie dans* (23)*. De plus, pour tout* $\alpha/2 < \alpha_0 < \alpha$*,*

$$\left\| \sqrt{n}\, C_n^{\frac{1}{2}}(\alpha_0)\left(\widehat{\vartheta}_n(\alpha_0) - \vartheta_n(\alpha_0)\right)\right\| \xrightarrow{\ p\ } +\infty. \tag{26}$$

Considérons maintenant la statistique

$$\forall\, n > 1, \qquad \widehat{\alpha}_n(\alpha) = \frac{\ln c - \ln\left(1 - \lambda_1 \widehat{v}_n(\alpha)\right)}{\ln n}$$

définie à condition que $\lambda_1 \widehat{v}_n(\alpha) < 1$ avec $\widehat{v}_n(\alpha)$ la première composante de $\widehat{\vartheta}_n(\alpha)$, autrement dit un estimateur de $\lambda_{n,1}$. On cherche donc à tester une valeur $\alpha_0 > \alpha/2$ via la statistique de test

$$Z_n^2(\alpha_0) = \frac{c\,\widehat{\pi}_n^2}{2}\,(\ln n)^2\, n^{1-\alpha_0}\left(\widehat{\alpha}_n(\alpha_0) - \alpha_0\right)^2$$

où $\widehat{\pi}_n$ est un estimateur consistant de π_{11}. Le théorème précédent permet finalement d'établir le comportement asymptotique de $Z_n^2(\alpha_0)$ selon les hypothèses étudiées.

Théorème. *On suppose que (A1), (A2) et (A3) sont vérifiées et que* $\mathbb{E}\left[|\epsilon_1|^{2+\nu}\right] = \eta_\nu < +\infty$ *pour un* $\nu > 0$*. Alors, sous l'hypothèse nulle* $\mathcal{H}_0$: "$\alpha = \alpha_0$"*, la statistique de test satisfait*

$$Z_n^2(\alpha_0) \xrightarrow{\ d\ } \chi_1^2$$

où χ_1^2 *suit une loi du khi-deux à un degré de liberté. En supposant de plus* $\alpha < 2\alpha_0$*, alors, sous l'hypothèse alternative* $\mathcal{H}_1$: "$\alpha > \alpha_0$"*,*

$$Z_n^2(\alpha_0) \xrightarrow{\ p\ } +\infty. \tag{27}$$

Plus concrètement, il suffit de choisir $\alpha_0 \geqslant 1/2$ et de rejeter $\mathcal{H}_0$ en faveur de $\mathcal{H}_1$ dès que $Z_n^2(\alpha_0) > z_{1-\zeta}$ où $z_{1-\zeta}$ est le quantile d'ordre $1-\zeta$ d'une loi du khi-deux. Ainsi, pour évaluer cette notion de quasi-instabilité en supposant c connu, un choix naturel pour α serait

$$\widehat{\alpha}_c = \inf\left\{\alpha_0 \in \mathcal{A} : \mathcal{H}_0 \text{ n'est pas rejetée}\right\}$$

où $\mathcal{A} \subset [1/2, 1[$ est une grille de valeur à tester (par convention, $\inf \varnothing = +\infty$ et on conclura alors que la série est intégrée si $\mathcal{H}_0$ est toujours rejeté et $\alpha \geqslant 1/2$). Si on souhaite maintenant estimer c conjointement à α, un protocole de sélection est proposé, reposant sur un principe de validation croisée. Bien que ce protocole soit empirique, les simulations suggèrent une amélioration des résultats par compensation du biais d'estimation de α découlant naturellement de la méthode proposée ci-dessus.

Le test statistique établi permet ainsi de mesurer la proximité d'une racine à la frontière unité, en l'occurrence ± 1. Ce test découle des propriétés asymptotiques de l'OLS obtenu dans un cadre général d'AR(p) quasi-instable. Étendre ces résultats à des modèles plus larges (ARMA ou même FARIMA) est l'une des perspectives de ces recherches.

Axe 2 : Sélection de modèle dans la classe des processus FARIMA faibles

Le deuxième axe de recherche de cette thèse porte sur la sélection des ordres AR et MA dans la classe des modèles FARIMA faibles. Après avoir introduit de manière générale les processus FARIMA, nous aborderons les questions d'inférence de ces processus et la sélection de modèle dans les sous-classes des processus FARIMA forts et semi-forts avant de terminer par un résumé détaillé du dernier chapitre de cette thèse.

Généralités sur les processus FARIMA

Le modèle autorégressif étudié jusqu'ici est l'un des premiers processus de série chronologique a avoir été développé d'un point de vue théorique. Il présente l'avantage d'être simple, en terme de construction, tout en permettant la modélisation de structure de dépendance linéaire et de corrélations entre des observations proches dans le temps. En effet, il est bien connu, voir par exemple Brockwell et Davis (1991), que la fonction d'autocovariance d'un processus AR stationnaire décroît de manière géométrique. Cependant, on observe dans de nombreuses séries empiriques a priori stationnaires, et en particulier en économie et en hydrologie, des phénomènes de dépendance entre des évènements distants dans le temps, qui se traduisent par une décroissance lente de la fonction d'autocorrélation. On parle alors de processus à mémoire longue, pour lesquels les modèles FARIMA font figure de référence, par opposition aux processus à mémoire courte, comme les modèles AR, MA ou encore ARMA.

Classification des types de mémoire

Ce phénomène est décrit pour la première fois dans les travaux de Hurst (1951), qui étudie les propriétés hydrologiques du Nil afin d'évaluer la taille optimale d'un réservoir. Il constate alors que les mesures ne concordent pas avec les prévisions des modèles statistiques standards et introduit une nouvelle statistique, dite statistique R/S, afin de quantifier ce phénomène de dépendance à long terme. Soient y_k le volume total d'eau écoulé sur une année k et u_n la moyenne des n premières années, on définit S_k les sommes cumulées telles que

$$\forall 1 \leqslant k \leqslant n, \quad S_k = \sum_{j=1}^{k} (y_j - u_n).$$

La statistique de Hurst est alors donnée par

$$R/S = \left(\max_{1 \leqslant k \leqslant n} (S_k) - \min_{1 \leqslant k \leqslant n} (S_k) \right) / \sigma$$

avec σ l'écart-type de la série (y_k). Hurst (1951) observe notamment que ce ratio augmente avec la taille de la série et montre empiriquement que R/S est équivalent, à une constante près, à n^H où H est devenu le *coefficient de Hurst*, utilisé de nos jours comme un indicateur de mémoire longue. D'après Hurst (1951), une série gaussienne supposée indépendante implique $H = 0.5$ alors que l'analyse de nombreuses séries hydrologiques suggère plutôt une valeur autour de 0.7. Les travaux de Mandelbrot et Wallis (1968), (1969a) et (1969b), contribueront à établir les fondements théoriques de ce phénomène en démontrant que $H = 0.5$ pour tout processus à mémoire courte, typiquement un bruit blanc gaussien, tandis qu'un processus dont l'exposant de Hurst est différent de 0.5 présentera une structure de dépendance particulière. Ainsi, $H > 0.5$ indique une persistance de l'autocorrélation tandis que $H < 0.5$ correspond à un processus dont les valeurs ont tendance à se compenser, induisant un retour à la moyenne (on parlera plus tard de processus antipersistant). En réponse à cette découverte, Mandelbrot et Van Ness (1968) développent le mouvement brownien fractionnaire, l'un des premiers processus stochastiques permettant de formaliser théoriquement la mémoire longue par l'intermédiaire du paramètre H caractérisant la structure de dépendance du processus. Plus tard, Granger et Joyeux (1980) puis Hosking (1981) proposeront une nouvelle classe de modèles capables de modéliser la mémoire longue : les processus FARIMA. Avant d'aborder plus en détails la nature de ce dernier modèle, il semble important de préciser les différents types de mémoire mentionnés ci-dessus.

La classification proposée ci-dessous a été introduite par Beran et al. (2013). On considère la densité spectrale d'un processus stationnaire $(X_t)_{t\in\mathbb{Z}}$ donnée par

$$f_X(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h)\, e^{-ih\omega} \quad \forall \omega \in [-\pi, \pi]$$

où $\gamma(.)$ est la fonction d'autocovariance. De manière équivalente, on a

$$\sum_{h=-\infty}^{\infty} \gamma(h) = 2\pi\, f_X(0).$$

La dépendance linéaire d'un processus peut être définie selon les catégories suivantes :

- (X_t) est un processus à *mémoire longue* si

$$\lim_{|\omega|\to 0} f_X(\omega) = \infty \quad \text{ou} \quad \sum_{h=-\infty}^{\infty} \gamma(h) = \infty.$$

- (X_t) est un processus à *mémoire courte* si

$$0 < \lim_{|\omega|\to 0} f_X(\omega) < \infty \quad \text{ou} \quad 0 < \sum_{h=-\infty}^{\infty} \gamma(h) < \infty.$$

- (X_t) est un processus *antipersistant* si

$$\lim_{|\omega|\to 0} f_X(\omega) = 0 \quad \text{ou} \quad \sum_{h=-\infty}^{\infty} \gamma(h) = 0.$$

Par exemple, soit $|\theta_1| < 1$, on rappelle que la fonction d'autocovariance d'un processus AR(1) stationnaire est donnée[19] par

$$\forall h \geqslant 0, \quad \gamma(h) = \frac{\sigma^2}{1-\theta_1^2}\, \theta_1^h$$

où $\sigma^2 < \infty$ est la variance du bruit blanc considéré. On a de plus $\gamma(h) = \gamma(-h)$, par

19. Voir (9).

stationnarité, ce qui implique

$$\begin{aligned}\sum_{h=-\infty}^{\infty} \gamma(h) &= 2 \sum_{h \geqslant 0} \gamma(h) - \gamma(0) \\ &= \frac{2\,\sigma^2}{(1-\theta_1^2)\,(1-\theta_1)} - \frac{\sigma^2}{1-\theta_1^2} \\ &= \frac{\sigma^2}{(1-\theta_1)^2}.\end{aligned}$$

Sous l'hypothèse de stationnarité, la série des covariance converge donc clairement vers une quantité finie et non nulle. D'après la classification de Beran et al. (2013), le processus AR(1) est donc à mémoire courte. Cette dépendance à courte portée se traduit notamment par une décroissance rapide de la fonction d'autocorrélation empirique. Ce résultat est notamment visible sur la Figure 3, où l'on retrouve un processus AR(1) de paramètre $\theta_1 = 0.6$ quand $d_0 = 0$ (graphique central).

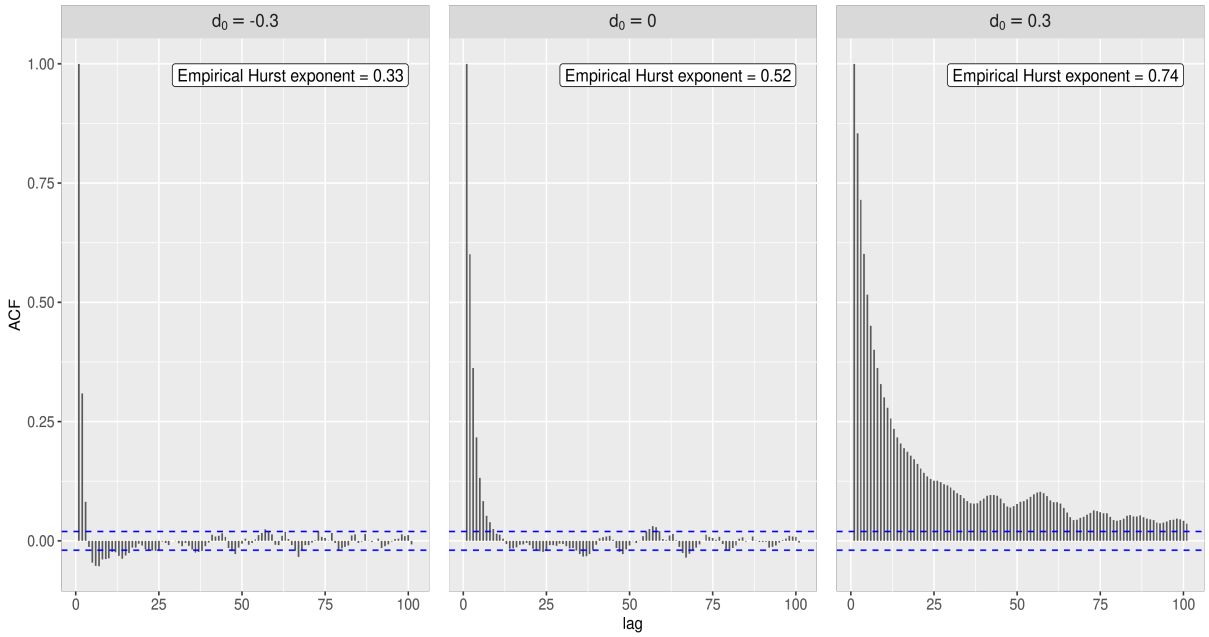


FIGURE 3 – Fonction d'autocorrélation empirique (ACF) du processus FARIMA$(1, d_0, 0)$ avec $a_{01} = 0.6$, $d_0 \in \{-0.3, 0, 0.3\}$ et $\epsilon_t \sim \mathcal{N}(0, 1)$, et les intervalles de confiance à 95% associés. Les trajectoires des processus sous-jacents sont de taille $n = 10000$, et l'exposant de Hurst empirique est estimé pour chaque réalisation.

Une définition plus formelle de cette classification est également proposée lorsque la densité spectrale d'un processus stationnaire peut s'écrire

$$f_X(\omega) = H_X(\omega)\,|\omega|^{-2d} \qquad \forall \omega \in [-\pi,\, \pi]$$

où $H_X(\omega)$ est une fonction positive et symétrique à variation lente à l'origine (voir notamment la Définition 1.1 de Beran et al. (2013) pour les définitions standard de Karamata et Zygmund). Le processus est alors dit

- à mémoire longue si $0 < d < 1/2$,
- à mémoire courte si $d = 0$ et $0 < \lim_{|\omega|\to 0} H_X(\omega) < \infty$,
- antipersistant si $-1/2 < d < 0$.

On parle de dépendance *intermédiaire* quand $d = 0$ mais $\lim_{|\omega|\to 0} H_X(\omega) = \infty$ car alors la série des autocovariances n'est pas sommable, ce qui s'apparente à une structure de dépendance à long terme.

Le processus stationnaire défini pour tout $t \in \mathbb{Z}$ par

$$(1 - aL)\,(1 - L)^d\, X_t = \epsilon_t$$

permet par exemple d'illustrer la différence entre ces trois principaux types de mémoire. On pose $|a| < 1$ et $d \in]-1/2,\, 1/2[$ pour garantir les hypothèses de causalité et inversibilité du processus (voir section suivante) et on suppose que (ϵ_t) est un bruit blanc de variance $\sigma^2 < \infty$. D'après le Théorème 13.2.2 de Brockwell et Davis (1991), il existe, dans ce cas, une unique solution stationnaire à l'équation ci-dessus donnée par

$$X_t = \sum_{k\in\mathbb{N}} \psi_k\, \epsilon_{t-k}$$

où $\psi(z) = \sum_{k\in\mathbb{Z}} \psi_k\, z^k = (1 - z)^{-d}\,(1 - az)^{-1}$. Soit $\omega \in [-\pi,\, \pi]$, on a alors la densité spectrale du processus telle que

$$\begin{aligned} 2\pi\, f_X(\omega) &= \sum_{h\in\mathbb{Z}} \mathrm{Cov}(X_t,\, X_{t+h})\, e^{-ih\omega} \\ &= \sum_{h\in\mathbb{Z}} \sum_{k\in\mathbb{N}} \sum_{\ell\in\mathbb{N}} \psi_k\, \psi_\ell\, \mathrm{Cov}(\epsilon_{t-k},\, \epsilon_{t+h-\ell})\, e^{-ih\omega} \\ &= \sigma^2 \sum_{k\in\mathbb{N}} \sum_{\ell\in\mathbb{N}} \psi_k\, \psi_\ell\, e^{ik\omega} e^{-i\ell\omega} \end{aligned}$$

par décorrélation du bruit. De plus, puisque $z\bar{z} = |z|^2$ pour tout nombre complexe z,

$$\begin{aligned} 2\pi\, f_X(\omega) &= \sigma^2 \,|\sum_{k\in\mathbb{N}} \psi_k\, e^{ik\omega}|^2 \\ &= \sigma^2\, |1 - e^{i\omega}|^{-2d}\, |1 - a\, e^{i\omega}|^{-2} \\ &= \sigma^2\, |2 - e^{i\omega} - e^{-i\omega}|^{-d}\, |1 - a\, e^{i\omega}|^{-2} \\ &= \sigma^2\, |2\,(1 - \cos(\omega))|^{-d}\, |1 - a\, e^{i\omega}|^{-2} \\ &= \sigma^2\, |2\sin(\frac{\omega}{2})|^{-2d}\, |1 - a\, e^{i\omega}|^{-2}. \end{aligned}$$

En notant que $\sin(x) \sim x$ quand $x \to 0$, on obtient que

$$f_X(\omega) \underset{|\omega|\to 0}{\sim} H_X(\omega)\, |\omega|^{-2d}$$

où $H_X(\omega) = (\sigma^2/2\pi)\,|1-a|^{-2}$ est une fonction constante strictement positive. En reprenant la définition ci-dessus, on voit immédiatement que la nature de la mémoire du processus (X_t) dépend uniquement du paramètre de différenciation d. On observe également le comportement de la fonction d'autocovariance empirique d'un tel processus pour différentes valeurs de $d = d_0$ dans la Figure 3. Il est clair que lorsque $d_0 = 0.3$, le processus présente une mémoire longue caractérisée par la lente décroissance de ses autocorrélations, contrairement aux deux autres cas. Bien qu'il soit difficile de distinguer visuellement le processus antipersistant du processus à mémoire courte, le coefficient de Hurst semble néanmoins capable de les discriminer efficacement. Le modèle introduit ici, qui n'est autre qu'un FARIMA$(1, d_0, 0)$, permet ainsi de représenter aussi bien la mémoire longue que la mémoire courte ou l'antipersistance. Cet exemple souligne la flexibilité d'un tel modèle dont nous allons maintenant présenter la formulation générale.

Présentation générale du modèle

Les processus FARIMA apparaissent dans les années 1980, en réponse au phénomène de Hurst et à la mise en évidence de ces dynamiques de dépendance à long terme. Leur formulation théorique, dérivée des processus ARMA et ARIMA, en font des outils facilement applicables, permettant de modéliser aussi bien la mémoire longue que les structures de dépendance plus classiques. Avant d'introduire cette extension, il semble donc opportun de rappeler les fondements des modèles ARMA et ARIMA. Dans la suite, le processus (ϵ_t) considéré est un bruit blanc dont les caractéristiques seront précisées si nécessaire.

Un modèle ARIMA(p_0, d_0, q_0) est en fait une combinaison d'un processus AR(p_0) et d'un processus MA(q_0), auquel on a ajouté un opérateur de différenciation $d_0 \in \mathbb{N}$. L'écriture d'un tel processus est donnée par

$$\forall t \in \mathbb{Z}, \quad a_0(L)\,(1-L)^{d_0}\,X_t = b_0(L)\,\epsilon_t \tag{28}$$

où on rappelle que L est l'opérateur retard, et où les opérateurs autorégressif (AR) et moyenne mobile (MA) sont respectivement définis tels que

$$a_0(z) = 1 - a_{0,1}z - \cdots - a_{0,p_0}\,z^{p_0} \quad \text{et} \quad b_0(z) = 1 - b_{0,1}z - \cdots - b_{0,q_0}\,z^{q_0}. \tag{29}$$

On suppose également que les polynômes $a_0(.)$ et $b_0(.)$ n'ont aucun facteur en commun pour garantir l'irréductibilité du modèle. On notera que pour $d_0 = 0$, l'opérateur de différenciation disparaît et le processus devient un ARMA(p_0, q_0). D'après les définitions 3.1.3 et 3.1.4 de Brockwell et Davis (1991), un processus ARMA est causal s'il existe une suite de constante (ψ_j) telle que la série des ψ_j est absolument sommable et

$$X_t = \sum_{j \geqslant 0} \psi_j\,\epsilon_{t-j} \quad \forall t \in \mathbb{Z}$$

et il est dit inversible s'il existe une suite de constante (π_i) telle que $\sum_{i \geqslant 0} |\pi_i| < \infty$ et

$$\sum_{i \geqslant 0} \pi_i\,X_{t-i} = \epsilon_t \quad \forall t \in \mathbb{Z}.$$

On a évoqué précédemment une condition nécessaire et suffisante pour obtenir la causalité d'un processus, à savoir $a_0(z) \neq 0$ pour tout $z \in \mathbb{C}$ tel que $|z| \leqslant 1$. De manière complémentaire, le processus ARMA est inversible si et seulement si le polynôme $b_0(z)$ a toutes ses racines en dehors du cercle unité. On suppose généralement la causalité et l'inversibilité dans (28). Ainsi, puisque la causalité implique la stationnarité, le processus ARIMA ci-dessus est stationnaire pour $d_0 = 0$ et devient non stationnaire dès que $d_0 \geqslant 1$ permettant de modéliser l'existence d'une (ou plusieurs) racine unitaire.

Dans l'idée de conserver une structure proche du modèle ARIMA, Granger et Joyeux (1980) introduisent la notion d'opérateur de différenciation fractionnaire $(1-L)^{d_0}$, avec $d_0 \in \mathbb{R}$, et posent les bases d'un premier processus FARIMA$(0, d_0, 0)$ tel que

$$\forall t \in \mathbb{Z}, \quad (1-L)^{d_0}\,X_t = \epsilon_t.$$

Par la formule du binôme de Newton généralisé,

$$(1-L)^{d_0} = \sum_{k\geqslant 0} \alpha_k(d_0)\, L^k \quad \text{avec} \quad \alpha_k(d_0) = \frac{\Gamma(k-d_0)}{\Gamma(k+1)\,\Gamma(-d_0)} \tag{30}$$

où $\Gamma(.)$ est la fonction Gamma. Par l'approximation de Stirling, on a $\Gamma(k) \sim \sqrt{2\pi}\, k^{k-\frac{1}{2}}\, e^{-k}$ pour k assez grand ce qui implique $\alpha_k(d_0) \sim k^{-d_0-1}/\Gamma(-d_0)$ et

$$\mathrm{Var}(X_t) = \sum_{k\geqslant 0} \alpha_k^2(-d_0)\,\mathrm{Var}(\epsilon_{t-k}) \;\sim\; \frac{\sigma^2}{\Gamma(d_0)^2} \sum_{k\geqslant 0} k^{2(d_0-1)}$$

où σ^2 est la variance (finie) du bruit blanc. La série de Riemann diverge dès que $d_0 \geqslant 1/2$ et, dans ce cas, la variance du processus (X_t) est infinie. On retrouve un raisonnement similaire dans le travail de Granger et Joyeux (1980) et dans celui de Hosking (1981) qui ira encore plus loin en démontrant qu'un tel processus est stationnaire pour $d_0 < 1/2$ et inversible quand $d_0 > -1/2$. Afin d'ajouter davantage de flexibilité à ce modèle, Hosking (1981) généralise l'étude de ce nouveau processus à des ordres AR et MA quelconques en supposant $d_0 \in \mathbb{R}$ dans l'écriture ARIMA. D'après Hosking (1981) ou encore Brockwell et Davis (1991), l'équation (28) admet alors une solution (X_t)

- stationnaire (et unique) si $d_0 < 1/2$ et $a_0(z) \neq 0$ pour tout $|z| \neq 1$,
- causale si $d_0 < 1/2$ et $a_0(z) \neq 0$ pour tout $|z| \leqslant 1$,
- inversible si $d_0 > -1/2$ et $b_0(z) \neq 0$ pour tout $|z| \leqslant 1$.

De la même façon que dans la section précédente, on peut montrer que la densité spectrale d'un processus FARIMA(p_0, d_0, q_0) causal et inversible est telle que

$$f_X(\omega) \underset{|\omega|\to 0}{\sim} \frac{\sigma^2}{2\pi} \frac{|b_0(1)|^2}{|a_0(1)|^2} |\omega|^{-2d_0}.$$

Il apparaît clairement que la dynamique de dépendance locale, issue de la partie ARMA du modèle, n'influence pas la structure de dépendance globale du processus, qui est finalement entièrement déterminée par le paramètre de différenciation fractionnaire d_0. L'expression de l'autocovariance du processus, que l'on retrouve dans Palma (2007), illustre également ce résultat. En tant qu'extension des processus ARIMA, cette nouvelle classe de modèle offre un cadre de travail flexible capable de représenter un large éventail de séries chronologiques. La section suivante présente différentes hypothèses sur le bruit élargissant encore le champ d'application des processus FARIMA.

Hypothèses sur le bruit

Les processus FARIMA occupent ainsi une place importante dans la modélisation des dépendances linéaires puisqu'ils peuvent représenter aussi bien des corrélations à décroissance rapide que des phénomènes de mémoire longue associés à une décroissance lente des corrélations. Néanmoins, les hypothèses fortes généralement imposées sur le bruit (le plus souvent iid ou gaussien) ne permettent pas d'intégrer d'éventuelles composantes non linéaires susceptibles d'affecter la série.

Typiquement, on suppose que le terme d'erreur est un bruit blanc fort, autrement dit une suite de variables aléatoires iid, dont la définition exclut naturellement toute forme de dépendance. On parlera dans ce cas de *FARIMA forts*. Lorsque l'hypothèse d'indépendance est relâchée, on autorise l'apparition de dépendance non linéaire dans le modèle ce qui rend l'analyse statistique beaucoup plus délicate. En règle générale, on suppose alors une structure de dépendance explicite, telle que le modèle GARCH [20] qui constitue un exemple classique en finance. Introduit par Bollerslev (1986), ce modèle généralise le cadre ARCH [21] proposé par Engle (1982) dans lequel la variance conditionnelle évolue selon une dynamique autorégressive sur le carré des innovations passées. D'après Francq et Zakoïan (2019), un processus (ϵ_t) est appelé GARCH(p, q) si ses deux premiers moments conditionnels existent et sont tels que, pour tout $t \in \mathbb{Z}$, $\mathbb{E}[\epsilon_t \,|\, \sigma(\epsilon_s,\, s \leqslant t)] = 0$ et sa variance conditionnelle $\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$ pour des constantes ω, $\alpha_1, \ldots, \alpha_q$ et $\beta_1, \ldots, \beta_p$. Cette définition illustre clairement qu'un processus GARCH est une suite de différences de martingale dont la variance conditionnelle dépend du passé. Les processus FARIMA dont les innovations représentent une suite de différences de martingale seront ainsi appelés *FARIMA semi-forts* tandis que ceux dont le bruit blanc est seulement supposé décorrélé seront désignés comme *FARIMA faibles*. Ces trois sous-classes de modèles s'imbriquent de la manière suivante :

$$\{\text{FARIMA forts}\} \subset \{\text{FARIMA semi-forts}\} \subset \{\text{FARIMA faibles}\}.$$

En effet, un bruit blanc fort est naturellement décorrélé et, puisque ce processus est centré, il constitue à la fois un bruit blanc faible et une suite de différences de martingale. De plus, soit (ϵ_t) une telle suite adaptée à la filtration $\mathcal{F}_t = \sigma(\epsilon_s,\, s \leqslant t)$, alors pour tout $t \in \mathbb{Z}$ et tout $s < t$,

$$\mathbb{E}[\epsilon_t\, \epsilon_s] = \mathbb{E}[\mathbb{E}[\epsilon_t\, \epsilon_s \,|\, \mathcal{F}_{t-1}]] = \mathbb{E}[\epsilon_s\, \mathbb{E}[\epsilon_t \,|\, \mathcal{F}_{t-1}]] = 0,$$

20. *Generalized Autoregressive Conditional Heteroskedasticity.*
21. *Autoregressive Conditional Heteroskedasticity.*

ce qui implique qu'une suite de différences de martingale est aussi un bruit blanc faible. Nous verrons plus tard que cette structure de dépendance particulière se rapproche néanmoins d'un bruit blanc fort du point de vue des propriétés asymptotiques des FARIMA associés (et notamment de la matrice de variance), d'où le terme "semi-fort".

L'utilisation de modèles FARIMA faibles permettraient donc non seulement de modéliser des phénomènes de dépendance linéaire à court et long terme mais également d'éventuelles dépendances non linéaires largement présentes dans les séries temporelles empiriques. Ainsi, Francq et Zakoïan (1998), (2005) et Francq et al. (2005) ont montré qu'il existe de nombreuses représentations ARMA de processus non linéaires comme les modèles à changements de régime, les processus bilinéaires ou encore les représentations causales de processus ARMA forts non causaux. Par exemple, dans ce dernier cas, il suffit de prendre un processus AR(1) fort, $X_t = \theta X_{t-1} + \eta_t$, avec $|\theta| > 1$, puis de poser $\epsilon_t = X_t - (1/\theta)\, X_{t-1}$. Alors (X_t) admet une représentation AR(1) causale dont le processus des innovations (ϵ_t) est un bruit blanc faible qui est ni indépendant, ni une différence de martingale. De même, l'hypothèse d'innovations iid exclut, par construction, la capacité de générer une dynamique de variance conditionnelle. Il apparaît donc essentiel d'utiliser un modèle capable de détecter ces dépendances non linéaires, afin d'éviter des erreurs d'estimation trop significatives. Ceci est d'autant plus vrai que les propriétés asymptotiques du modèle associé dépendent fortement des hypothèses faites sur le bruit, comme nous le verrons à la suite.

Inférence des processus FARIMA

La méthode Box-Jenkins, formalisée par Box et Jenkins (1970), est un protocole initialement développé pour la modélisation ARMA des séries temporelles mais qui s'étend naturellement à toute extension de ce processus. Cette approche se fait en trois étapes distinctes qui sont, pour un modèle FARIMA :

1. Identification du modèle par sélection des ordres p et q des opérateurs AR et MA,
2. Estimation des paramètres $a_1, \ldots, a_p, b_1, \ldots, b_q$ et d,
3. Validation du modèle par l'analyse des résidus obtenus.

La suite de cette section se consacre à la présentation des étapes 2 et 3 de cette méthodologie, en considérant $p = p_0$ et $q = q_0$ connus et en mettant l'accent sur la théorie développée sous les hypothèses les plus générales concernant le bruit. L'étape d'identification sera développée dans les sections suivantes.

L'estimation des processus FARIMA a fait l'objet de nombreuses études, même si la prise en compte d'une dynamique non linéaire générale apparaît plus délicate à formaliser. Les processus à mémoire longue sont généralement modélisés à partir de l'estimateur du maximum de vraisemblance (MLE) dans le domaine fréquentiel. Ainsi, Fox et Taqqu (1986) proposent une approximation du MLE adaptée de l'approche de Whittle (1953) et montrent la consistance et la normalité asymptotique de cet estimateur pour une série chronologique gaussienne de densité spectrale $f_\theta(\omega) \underset{|\omega|\to 0}{\sim} H_\theta(\omega)\,|\omega|^{-\alpha(\theta_0)}$ où $H_\theta(.)$ varie lentement à l'origine et $0 < \alpha(\theta_0) < 1$. Cette définition inclut, par exemple, les bruits gaussiens fractionnaires ou encore les processus FARIMA à mémoire longue, ie $0 < d_0 < 1/2$, et dont le bruit blanc est supposé gaussien. Dahlhaus (1989) étendra ce travail au MLE exact de θ_0 et établira de plus, sous les mêmes hypothèses et toujours dans le domaine fréquentiel, l'efficacité asymptotique de ces deux estimateurs via la matrice de variance limite donnée par $\Gamma(\theta_0)^{-1}$ où

$$\Gamma(\theta_0) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial\theta} \log f_{\theta_0}(\omega)\, \frac{\partial}{\partial\theta^\top} \log f_{\theta_0}(\omega)\, \mathrm{d}\omega.$$

Dans une étude centrée maintenant sur le processus FARIMA associé à un bruit fort gaussien, Beran (1995) démontre de nouveau la consistance et la normalité asymptotique du MLE mais pour tout $d_0 > -1/2$, couvrant ainsi le cas non stationnaire ainsi que les situations d'antipersistance et de mémoire courte. Parmi les recherches effectuées dans le cadre des FARIMA semi-forts, Hualde et Robinson (2011) montrent que l'estimateur obtenu par minimisation de la somme des carrés des erreurs (CSS) [22] est consistant et asymptotiquement normal pour tout $d_0 \in \mathbb{R}$, tel que

$$\sqrt{n}\,(\widehat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0,\, A(\theta_0)^{-1})$$

où la matrice $A(\theta_0)$ est donnée de manière explicite dans la section 2.2 de Hualde et Robinson (2011). À notre connaissance, cet article figure parmi les rares contributions fournissant une matrice de covariance explicite dans un cadre temporel, contrairement à l'approche fréquentielle privilégiée dans les travaux précédents. En considérant une structure de dépendance non linéaire plus générale, Shao (2010) établit, dans le domaine fréquentiel, la consistance et la normalité asymptotique de l'estimateur de Whittle pour $-1/2 < d_0 < 1/2$ et ce résultat est étendu au domaine non-stationnaire en adaptant l'esti-

22. *Conditional Sum of Squares*

mateur. La matrice de variance limite obtenue dans les deux cas s'écrit sous la forme sandwich $W(\theta_0)^{-1} V(\theta_0) W(\theta_0)^{-1}$ et coïncide exactement avec $\Gamma(\theta_0)$ définie ci-dessus lorsque le processus des innovations réponds à des contraintes iid ou gaussiennes (voir Remarque 2.4 de Shao (2010)). Ce travail constitue une première approche pour l'estimation des processus FARIMA faibles. Cependant, lorsque les hypothèses sur le bruit sont relâchées, l'estimation de leur matrice $V(\theta_0)$ semble difficilement réalisable puisqu'elle dépend des cumulants joints d'ordre 4 du processus des innovations et il n'existe aucun estimateur explicite dans la littérature. Shao (2012) propose une approche visant à construire des intervalles de confiance pour pallier à cette difficulté mais uniquement pour les processus FARIMA à mémoire courte (voir Hypothèse 1). Plus récemment, Ben Hariz et al. (2025) montrent qu'il est possible d'estimer cette matrice limite à l'aide d'un estimateur one-step partageant les mêmes propriétés asymptotiques que l'estimateur de Whittle.

Abordons maintenant plus en détail l'un des travaux fondateurs relatifs à l'inférence des processus FARIMA faibles, développé dans Boubacar Maïnassara et al. (2021). Soit (X_t) centré et stationnaire au second ordre défini par

$$\forall t \in \mathbb{Z}, \quad a_0(L)\,(1-L)^{d_0}\,X_t = b_0(L)\,\epsilon_t \tag{31}$$

où le paramètre de longue mémoire est tel que $-1/2 < d_0 < 1/2$ et les opérateurs AR et MA sont donnés par (29). Le processus (X_t) admet une représentation FARIMA(p_0, d_0, q_0) faible si la suite des innovations (ϵ_t) est un bruit blanc faible tel que

(H$_0$) $\mathbb{E}[\epsilon_t] = 0$, $\mathrm{Var}(\epsilon_t) = \sigma^2 > 0$, et $\mathrm{Cov}(\epsilon_t, \epsilon_s) = 0$ pour tout $t \neq s$.

On pose $\theta_0 = (a_{0,1}, \ldots, a_{0,p_0}, b_{0,1}, \ldots, b_{0,q_0}\ d_0)^\intercal$ le vecteur des paramètres vérifiant (31), le théorème 13.2.2 de Brockwell et Davis (1991) garantit l'existence et l'unicité de cette solution qui est stationnaire, causale et inversible sous l'hypothèse

(H$_1$) Les polynômes $a_0(.)$ et $b_0(.)$ n'ont aucun facteur en commun et admettent toutes leurs racines en dehors du cercle unité.

L'espace des paramètres est alors défini tel que $\Theta_\delta := \Theta_\delta^* \times [d_1\, d_2]$ avec

$$\Theta_\delta^* := \{\theta \in \mathbb{R}^{p_0+q_0} \mid \text{les racines } z^* \text{ des polynômes } a_\theta(z) = 1 - \textstyle\sum_{i=1}^{p_0} \theta_i\, z^i \text{ et } b_\theta(z) = 1 - \sum_{j=1}^{q_0} \theta_{p_0+j}\, z^j \text{ vérifient } |z^*| \geqslant 1+\delta\}$$

et $[d_1\, d_2] \subset]-1/2, 1/2[$ avec $d_1 - d_0 > -1/2$ (nous verrons dans le Chapitre 3 que l'on peut se passer de cette dernière condition). Pour tout $\theta \in \Theta_\delta$, considérons maintenant le

processus stationnaire $(\epsilon_t(\theta))$ solution de

$$\forall t \in \mathbb{Z}, \quad \epsilon_t(\theta) = \sum_{k\geqslant 0} \alpha_k(d)\, X_{t-k} - \sum_{i=1}^{p_0} \theta_i \sum_{k\geqslant 0} \alpha_k(d)\, X_{t-i-k} + \sum_{j=1}^{q_0} \theta_{p_0+j}\, \epsilon_{t-j}(\theta)$$

où la suite des coefficients $(\alpha_k(d))$ est explicitée dans (30) et, pour tout $t \in \mathbb{Z}$, on a, presque sûrement (p.s.), $\epsilon_t(\theta_0) = \epsilon_t$. Connaissant $X_1, \ldots, X_n,$ on peut approximer ce processus par la série des erreurs tronquées, définie récursivement par

$$\forall 1 \leqslant t \leqslant n, \quad \tilde{\epsilon}_t(\theta) = \sum_{k=0}^{t-1} \alpha_k(d)\, X_{t-k} - \sum_{i=1}^{p_0} \theta_i \sum_{k=0}^{t-i-1} \alpha_k(d)\, X_{t-i-k} + \sum_{j=1}^{q_0} \theta_{p_0+j}\, \tilde{\epsilon}_{t-j}(\theta) \tag{32}$$

avec $\tilde{\epsilon}_t(\theta) = X_t = 0$ dès que $t \leqslant 0$. L'estimateur des moindres carrés (LSE) de θ_0 est alors donné par

$$\widehat{\theta}_n = \underset{\theta\in\Theta_\delta}{\operatorname{argmin}}\, \tilde{\sigma}_n^2(\theta) \quad \text{où} \quad \tilde{\sigma}_n^2(\theta) := \frac{1}{n}\sum_{t=1}^{n} \tilde{\epsilon}_t^2(\theta). \tag{33}$$

Boubacar Maïnassara et al. (2021) établissent alors la consistance de l'estimateur des moindres carrés sous (H_0), (H_1) et

(H_2) Le processus (ϵ_t) est strictement stationnaire et ergodique.

Des hypothèses supplémentaires sont néanmoins requises pour l'obtention de la normalité asymptotique du LSE.

(H_3) θ_0 se situe strictement à l'intérieur de l'espace des paramètres Θ_δ.

Dans le cadre de modèles non linéaires, où la structure de dépendance est généralement inconnue, l'étude des propriétés asymptotiques de l'estimateur requiert des hypothèses de sommabilité des cumulants joints du processus des innovations voire le recours à des coefficients de mélange fort afin de contrôler la dépendance non linéaire. Ces hypothèses sont classiques et apparaissent notamment dans Francq et Zakoïan (1998) et Boubacar Maïnassara et Francq (2011) pour l'étude des ARMA faibles univariés et multivariés (respectivement) ainsi que dans Shao (2010) pour les FARIMA faibles. Les coefficients de mélange forts associés au bruit $\epsilon := (\epsilon_t)$ sont définis, pour tout $h \geqslant 0$, par

$$\alpha_\epsilon(h) = \sup_{A\in\mathcal{F}_{-\infty}^{t},\, B\in\mathcal{F}_{t+h}^{+\infty}} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)|$$

où $\mathcal{F}_{-\infty}^{t} = \sigma(\epsilon_u,\, u \leqslant t)$ et $\mathcal{F}_{t+h}^{+\infty} = \sigma(\epsilon_u,\, u \geqslant t+h)$. La dépendance non linéaire générée par le processus (ϵ_t) est alors contrôlée par l'hypothèse

(H_4) Il existe un entier $\tau \geqslant 2$ tel que, pour $0 < \nu \leqslant 1$, $\mathbb{E}[|\epsilon_t|^{\tau+\nu}] < \infty$ et, pour $k = 1, \ldots, \tau$, $\sum_{h\geqslant 0} (h+1)^{k-2} \{\alpha_\epsilon(h)\}^{\frac{\nu}{k+\nu}} < \infty$.

D'après la proposition 2.2 de Doukhan et Léon (1989), (H_4) implique la sommabilité des cumulants joints telle que, pour cette même valeur de τ,

$$\sum_{t_1,\ldots,t_{\tau-1}\in\mathbb{Z}} |\mathrm{cum}(\epsilon_0, \epsilon_{t_1}, \ldots, \epsilon_{t_{\tau-1}})| < \infty.$$

On rappelle ci-dessous l'expression des premiers cumulants d'ordre joints d'un processus centré, à partir des moments de ce processus :

$$\begin{aligned}
\mathbb{E}[\epsilon_t] &= \mathrm{cum}(\epsilon_t) = 0 \\
\mathbb{E}[\epsilon_{t_1}\,\epsilon_{t_2}] &= \mathrm{cum}(\epsilon_{t_1}, \epsilon_{t_2}) \\
\mathbb{E}[\epsilon_{t_1}\,\epsilon_{t_2}\,\epsilon_{t_3}] &= \mathrm{cum}(\epsilon_{t_1}, \epsilon_{t_2}, \epsilon_{t_3}) \\
\mathbb{E}[\epsilon_{t_1}\,\epsilon_{t_2}\,\epsilon_{t_3}\,\epsilon_{t_4}] &= \mathrm{cum}(\epsilon_{t_1}, \epsilon_{t_2}, \epsilon_{t_3}, \epsilon_{t_4}) + \mathrm{cum}(\epsilon_{t_1}, \epsilon_{t_2})\,\mathrm{cum}(\epsilon_{t_3}, \epsilon_{t_4}) \\
&\quad + \mathrm{cum}(\epsilon_{t_1}, \epsilon_{t_3})\,\mathrm{cum}(\epsilon_{t_2}, \epsilon_{t_4}) + \mathrm{cum}(\epsilon_{t_1}, \epsilon_{t_4})\,\mathrm{cum}(\epsilon_{t_2}, \epsilon_{t_3}).
\end{aligned}$$

Sous les hypothèses (H_0)-(H_3) et (H_4) avec $\tau = 4$, on a alors

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}\Big(0,\, J(\theta_0)^{-1}\, I(\theta_0)\, J(\theta_0)^{-1}\Big) \tag{34}$$

avec les matrices limites

$$J(\theta_0) = 2\,\mathbb{E}\left[\frac{\partial}{\partial\theta}\epsilon_t(\theta_0)\,\frac{\partial}{\partial\theta^\intercal}\epsilon_t(\theta_0)\right] \text{ p.s.}$$

et

$$I(\theta_0) = \sum_{h\in\mathbb{Z}} \mathrm{Cov}(H_t(\theta_0),\, H_{t-h}(\theta_0)) \quad \text{où} \quad H_t(\theta) = 2\,\epsilon_t(\theta)\frac{\partial}{\partial\theta}\epsilon_t(\theta).$$

On retrouve une matrice de variance de type sandwich, similaire à celle présentée dans l'étude de Shao (2010), et l'on observe que $J(\theta_0)$ présente une structure proche de celle de leur matrice $W(\theta_0)$ formulée par densité spectrale, ou celle de $\Gamma(\theta_0)$ obtenue par Dahlhaus (1989), Fox et Taqqu (1986). Dans le domaine temporel, on peut comparer la matrice limite $A(\theta_0)$ obtenue par Hualde et Robinson (2011) avec la matrice $J(\theta_0)$ calculée de manière explicite dans la section 2.3 de l'article Ben Hariz et al. (2023), on obtient exactement $A(\theta_0) = (2\sigma^2)^{-1} J(\theta_0)$. Il apparaît que, dans les cas forts et semi-forts, la matrice sandwich est simplement réduite à $2\sigma^2 J(\theta_0)^{-1}$. En effet, soit (ϵ_t) une suite de

différence de martingale adaptée à la filtration $\mathcal{F}_t = \sigma(\epsilon_s,\, s \leqslant t)$ et de variance σ^2, alors

$$\begin{aligned}
I(\theta_0) &= \sum_{h\in\mathbb{Z}} \operatorname{Cov}\left(2\,\epsilon_t\left(\frac{\partial}{\partial\theta}\epsilon_t\right),\, 2\,\epsilon_{t-h}\left(\frac{\partial}{\partial\theta}\epsilon_{t-h}\right)\right) \\
&= 4\sum_{h\leqslant -1} \mathbb{E}\left[\epsilon_t\left(\frac{\partial}{\partial\theta}\epsilon_t\right)\epsilon_{t-h}\left(\frac{\partial}{\partial\theta^\top}\epsilon_{t-h}\right)\right] + 4\,\mathbb{E}\left[\epsilon_t^2\left(\frac{\partial}{\partial\theta}\epsilon_t\right)\left(\frac{\partial}{\partial\theta^\top}\epsilon_t\right)\right] \\
&\quad + 4\sum_{h\geqslant 1} \mathbb{E}\left[\epsilon_t\left(\frac{\partial}{\partial\theta}\epsilon_t\right)\epsilon_{t-h}\left(\frac{\partial}{\partial\theta^\top}\epsilon_{t-h}\right)\right] \\
&= 4\sum_{h\leqslant -1} \mathbb{E}\left[\mathbb{E}\left[\epsilon_t\left(\frac{\partial}{\partial\theta}\epsilon_t\right)\epsilon_{t-h}\left(\frac{\partial}{\partial\theta^\top}\epsilon_{t-h}\right)\Big|\,\mathcal{F}_{t-h-1}\right]\right] + 4\,\mathbb{E}\left[\mathbb{E}\left[\epsilon_t^2\left(\frac{\partial}{\partial\theta}\epsilon_t\right)\left(\frac{\partial}{\partial\theta^\top}\epsilon_t\right)\Big|\,\mathcal{F}_{t-1}\right]\right] \\
&\quad + 4\sum_{h\geqslant 1} \mathbb{E}\left[\mathbb{E}\left[\epsilon_t\left(\frac{\partial}{\partial\theta}\epsilon_t\right)\epsilon_{t-h}\left(\frac{\partial}{\partial\theta^\top}\epsilon_{t-h}\right)\Big|\,\mathcal{F}_{t-1}\right]\right].
\end{aligned}$$

En remarquant que

$$\forall t \quad \left(\frac{\partial}{\partial\theta}\epsilon_t\right) \subset \mathcal{F}_{t-1},$$

on obtient

$$\begin{aligned}
I(\theta_0) &= 4\sum_{h\leqslant -1} \mathbb{E}\left[\epsilon_t\left(\frac{\partial}{\partial\theta}\epsilon_t\right)\mathbb{E}\left[\epsilon_{t-h}\Big|\,\mathcal{F}_{t-h-1}\right]\left(\frac{\partial}{\partial\theta^\top}\epsilon_{t-h}\right)\right] + 4\,\mathbb{E}\left[\mathbb{E}\left[\epsilon_t^2\Big|\,\mathcal{F}_{t-1}\right]\left(\frac{\partial}{\partial\theta}\epsilon_t\right)\left(\frac{\partial}{\partial\theta^\top}\epsilon_t\right)\right] \\
&\quad + 4\sum_{h\geqslant 1} \mathbb{E}\left[\mathbb{E}\left[\epsilon_t\Big|\,\mathcal{F}_{t-1}\right]\left(\frac{\partial}{\partial\theta}\epsilon_t\right)\epsilon_{t-h}\left(\frac{\partial}{\partial\theta^\top}\epsilon_{t-h}\right)\right] \\
&= 2\sigma^2 J(\theta_0).
\end{aligned}$$

Par définition, un bruit blanc iid peut être représenté comme une différence de martingale, ce résultat reste donc valable dans le cadre des FARIMA forts. Il se démontre d'ailleurs aisément en exploitant l'indépendance du bruit dans un raisonnement analogue.

Pour compléter l'inférence des processus faibles, Boubacar Maïnassara et al. (2021) proposent un estimateur fortement consistant de la matrice $J(\theta_0)$, en remplaçant le bruit par la série des erreurs tronquées évaluées en $\widehat{\theta}_n$. La matrice $I(\theta_0)$ peut elle être interprétée comme la densité spectrale, à un facteur près, du processus $(H_t(\theta_0))$, et être estimée de manière consistante à partir d'un modèle autorégressif tronqué à un ordre dépendant de n. Cette approche a été initialement développée par Berk (1974) pour des processus linéaires à bruit indépendant, puis étendue notamment par Boubacar Maïnassara et Francq (2011) et Boubacar Maïnassara et al. (2012) aux modèles multivariés où les innovations ne sont plus supposées indépendantes. Une estimation non paramétrique, mais toujours consistante, de la matrice de variance est également proposée par Boubacar Maïnassara et

Esstafa (2025).

La dernière étape de la méthode Box-Jenkins consiste à valider, ou non, le modèle en testant les résidus obtenus, c'est-à-dire en testant la série des $(\tilde{\epsilon}_t(\widehat{\theta}_n))$ définie par (32). Les tests de portmanteau, développés par Box et Pierce (1970) et Ljung et Box (1978) dans une version modifiée, figurent probablement parmi les méthodes les plus utilisées pour vérifier l'absence d'autocorrélation résiduelle après estimation d'un modèle ARMA. Les statistiques associées à ces tests sont données respectivement par

$$Q_\ell^{BP} = n \sum_{h=1}^{\ell} \hat{\rho}^2(h) \quad \text{et} \quad Q_\ell^{LB} = n(n+2) \sum_{h=1}^{\ell} \frac{\hat{\rho}^2(h)}{n-h}$$

où $\hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0)$ est l'autocorrélation empirique calculée à partir de l'autocovariance de la série des résidus

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=h+1}^{n} \tilde{\epsilon}_t(\widehat{\theta}_n)\, \tilde{\epsilon}_{t-h}(\widehat{\theta}_n) \quad \text{pour tout} \quad 0 \leqslant h < n.$$

Li et McLeod (1986) ont montré que, lorsque le modèle FARIMA(p, d, q) est correctement spécifié, ces statistiques convergent vers une loi du χ^2 à $\ell - (p+q+1)$ degrés de liberté. Toutefois, ces résultats reposent sur l'hypothèse stricte d'un processus des innovations iid et peuvent devenir invalides lorsque le bruit contient des dépendances non linéaires (tout en restant décorrélé). Ling et Li (1997) et plus tard Shao (2011) [23] généraliseront les propriétés asymptotiques du test de Box-Pierce au cadre des processus FARIMA semi-forts puis Boubacar Maïnassara et al. (2023) étendront la procédure de validation des modèles ARMA faibles, établie par Francq et al. (2005), aux processus FARIMA faibles. Ils démontrent ainsi que, en relâchant les hypothèses sur le bruit, les statistiques de Box-Pierce et Ljung-Box convergent vers une loi correspondant à une combinaison de χ^2 dont la distribution peut être approchée à l'aide de l'algorithme de Imhof (1961). Cette méthode nécessite cependant l'estimation d'une matrice de covariance asymptotique qui peut s'avérer délicate. Une méthode alternative est donc proposée, permettant de construire un nouveau test statistique basé sur une approche auto-normalisée utilisée par Boubacar Maïnassara et Saussereau (2018).

Si les procédures d'estimation et de validation des processus FARIMA sont désormais bien établies dans un cadre général, il n'en va pas de même pour l'identification du modèle,

23. L'auteur propose également un test de validation pour une sous-classe restreinte des processus FARIMA faibles.

qui constitue pourtant une étape cruciale de la modélisation. La section suivante propose un aperçu des principales approches de sélection pour les modèles FARIMA.

Sélection de modèle dans la classe des processus FARIMA

Dans la section précédente, les ordres autorégressif p_0 et moyenne mobile q_0 sont supposés connus. La sélection de ces ordres devient le problème central pour l'identification des modèles FARIMA puisqu'une mauvaise spécification peut mener à une dégradation significative des performances d'estimation et de prévision. D'un côté, le sous-ajustement des paramètres conduit à un modèle trop simple, incapable de traduire pleinement la complexité des données, ce qui entraîne généralement une dépendance résiduelle et des estimateurs inconsistants. D'un autre côté, le sur-ajustement contrevient au principe de parcimonie en introduisant des paramètres superflus, pouvant provoquer une augmentation de la variance et une inférence instable.

L'approche de Box et Jenkins (1970) pour la sélection des ordres AR et MA repose initialement sur une identification graphique, fondée sur l'analyse des fonctions empiriques d'autocorrélation et d'autocorrélation partielle. Cette méthode devient cependant délicate dès que le modèle dépasse le cadre du processus ARMA et nécessite le recours à d'autres procédures de sélection. La minimisation d'un critère d'information, tel que l'AIC (*Akaike Information Criterion*) ou le BIC (*Bayesian Information Criterion*), constitue alors une alternative classique pour déterminer les ordres AR et MA optimaux. Notons κ_m la dimension d'un modèle candidat m et $L_n(m)$ le maximum de vraisemblance obtenu dans ce modèle, le célèbre critère développé par Akaike (1974) est donné par

$$\mathrm{AIC}_m = -2 \log L_n(m) + 2\kappa_m.$$

La minimisation de ce critère doit permettre d'obtenir le modèle qui engendre la plus petite distance de Kullback–Leibler par rapport au modèle générateur des données. Un peu plus tard, Schwarz (1978) propose le critère BIC, défini par

$$\mathrm{BIC}_m = -2 \log L_n(m) + \kappa_m \log n,$$

reposant sur une approximation bayésienne visant à identifier le modèle le plus probable compte tenu des données observées. À la différence de l'AIC, le BIC applique une pénalité plus importante sur le nombre de paramètres, favorisant ainsi les modèles moins

complexes et plus parcimonieux. Hannan (1980) montre par exemple que les estimateurs $\hat{p}_n$ et $\hat{q}_n$ des ordres d'un processus ARMA, obtenus par minimisation du BIC, sont fortement consistants, ce qui n'est pas le cas pour l'AIC qui a tendance à sélectionner des modèles sur-ajustés (voir Shibata (1976)). De même, Bardet et al. (2020) (2023) suggèrent des conditions générales suffisantes sur le terme de pénalité pour garantir la consistance d'un critère, dans le sens où la probabilité de sélectionner le bon modèle tend vers 1, au sein d'une large classe de modèles linéaires causaux à mémoire courte comprenant, entre autres, les modèles ARMA. Ils établissent notamment la consistance du BIC et remarquent que les critères à pénalité constante, typiquement l'AIC, ne satisfont pas ces conditions. Dans le même cadre de travail, Kamila (2022) développe un critère fortement consistant (sélection asymptotiquement presque sûre du vrai modèle) dérivé du HQC ou critère de Hannan et Quinn. D'autres critères ont également été analysés pour la sélection de modèles à mémoire courte parmi lesquels on peut citer Hannan et Rissanen (1982) pour les ARMA semi-forts ou encore Sin et Yu (2019) qui suggèrent une combinaison AIC-BIC pour la sélection d'un modèle autorégressif d'ordre infini, éventuellement intégré, adapté à la représentation d'un processus ARIMA. De plus, en supposant uniquement la décorrélation des innovations, Boubacar Maïnassara (2011) propose un critère AIC modifié, adapté à la sélection des modèles ARMA multivariés faibles.

Si les propriétés de ces critères ont été largement étudiées dans le cadre des séries temporelles à dépendance à court terme, la littérature est moins abondantes en ce qui concerne la mémoire longue et reste très limitée lorsqu'elle est associée à des hypothèses faibles sur le bruit. La consistance du BIC a été démontrée par Beran et al. (1998) dans le cadre restreint des processus FARIMA forts sans partie moyenne mobile, c'est-à-dire en supposant $q_0 = 0$. Plus tard, Baillie et al. (2014) présenteront, dans un cadre général incluant les processus FARIMA semi-forts, une procédure de sélection des ordres de la partie mémoire courte, après estimation semi-paramétrique du paramètre de mémoire longue. La critère d'information utilisé met en compétition deux modèles, et les auteurs montrent que la probabilité de choisir celui qui minimise la divergence de Kullback–Leibler converge vers 1. Toujours dans un cadre général, Kengne (2022) fournit des conditions asymptotiques pour la consistance faible et forte de la sélection pour une large classe de série temporelle incluant potentiellement des dépendances à long terme, et dont les innovations forment une suite de différences de martingale. La procédure de sélection repose sur la minimisation d'une fonction de contraste générale, capable d'englober un large éventail de critères. Finalement, Huang et al. (2022) démontrent la consistance du

BIC pour l'identification des processus FARIMA forts et semi-forts, constituant ainsi une extension notable du travail de Beran et al. (1998).

Contributions de la thèse

A notre connaissance, l'identification de modèle dans le cadre des processus FARIMA faibles n'a pas encore été développée dans la littérature. L'objectif de ce second axe de recherche est donc d'étendre les travaux de Huang et al. (2022) en démontrant notamment la consistance du BIC pour la sélection des modèles FARIMA dont le bruit est uniquement supposé décorrélé. Il s'agit de l'article présenté en Chapitre 3 qui a fait l'objet d'une soumission pour publication début 2026. Ce travail, résumé dans la section suivante, a été effectué en collaboration avec Alexandre Brouste, Youssef Esstafa et Tetsuya Takabatake.

Résumé du Chapitre 3

Soit la série $(X_t)_{t\in\mathbb{Z}}$ générée par le modèle FARIMA(p_0, d_0, q_0) suivant

$$\begin{aligned} X_t &= (1-L)^{-d_0}\left\{Y_t\,\mathbb{1}_{\{t\geqslant 1\}}\right\}, && t\in\mathbb{Z},\\ a_{\xi_0}(L)\,Y_t &= b_{\xi_0}(L)\,\epsilon_t, && t\in\mathbb{Z}, \end{aligned}$$

où les polynômes autorégressif et moyenne mobile sont définis par $a_{\xi_0}(z) = 1 - \sum_{i=1}^{p_0} a_{0,i} z^i$ et $b_{\xi_0}(z) = 1 - \sum_{j=1}^{q_0} b_{0,j} z^j$. Le vecteur des paramètres est alors donné par $\theta_0 = (\xi_0^\top, d_0)^\top$ avec $\xi_0 = (a_{0,1}, \ldots, a_{0,p_0}, b_{0,1}, \ldots, b_{0,q_0}, d_0)^\top \in \mathbb{R}^{p_0+q_0}$ et $d_0 \in \mathbb{R}$. On suppose également que les polynômes AR et MA vérifient les hypothèses classiques, à savoir (H_1) assurant l'inversibilité et l'identifiabilité du modèle, et que leurs coefficients de plus haut degré sont non-nuls. Enfin, le processus d'innovation (ϵ_t) est un bruit blanc faible, supposé ergodique, tel que $\mathbb{E}[\epsilon_t] = 0$, $\mathrm{Var}(\epsilon_t) = \sigma^2 > 0$, et $\mathrm{Cov}(\epsilon_t, \epsilon_s) = 0$ pour tout $t \neq s$. Le processus ainsi défini est qualifié de FARIMA faible et englobe les processus FARIMA fort et semi-fort tels que présenté plus haut. On impose de plus l'hypothèse (H_4) pour contrôler la vitesse de décroissance des coefficients de mélange forts associés au bruit.

Dans le cadre de la sélection de modèle considérée ici, les paramètres d'intérêt p_0 et q_0 sont supposés inconnus et estimés à partir à partir d'une série de n observations. On cherche ainsi à identifier le vrai modèle $m_0 = (p_0, q_0)$ parmi un ensemble de modèles candidats $\mathcal{M} = \{m = (p, q)\,|\,0 \leqslant p \leqslant P, 0 \leqslant q \leqslant Q\}$ où P et Q sont des entiers fixés. Pour chaque modèle candidat m, on pose $\theta_m = (a_1, \ldots, a_p, b_1, \ldots, b_p, d)^\top \in \Theta_m$. L'espace des

paramètres est défini par $\Theta_m = \Pi_m \times \mathcal{D}$, avec

$$\Pi_m := \{\xi_m = (a_1, \ldots, a_p, b_1, \ldots, b_q)^\top \,|\, (a_1, \ldots, a_p, 0, \ldots, 0, b_1, \ldots, b_q, 0, \ldots, 0)^\top \in \Pi_M\}$$

où $\Pi_M \subset \mathbb{R}^{P+Q}$ est un ensemble compact qui contient tous les vecteurs de paramètres $\xi_M = (a_1, \ldots, a_P, b_1, \ldots, b_Q)^\top$ correspondant à des polynômes AR et MA causaux, et l'ensemble $\mathcal{D} = [d_L, d_U]$, tel que $-\infty < d_L < d_U < \infty$, est l'espace compact dans lequel évolue le paramètre de mémoire d. La stationnarité et l'inversibilité de chaque modèle candidat est donc assurée par les conditions de causalité imposées sur l'espace des paramètres commun Π_M. On suppose de plus que le vrai paramètre θ_0 se situe à l'intérieur de l'ensemble $\Theta_{m_0} = \Pi_{m_0} \times \mathcal{D}$.

Pour chaque modèle $m \in \mathcal{M}$, le processus des erreurs est défini tel que

$$\epsilon_t(\theta_m) := \sum_{i=0}^{t-1} \phi_i(\theta_m)\, X_{t-i} \quad \text{for } t = 1, \ldots, n,$$

où la suite de coefficients $(\phi_i(\theta_m))_{i\geqslant 0}$ provient du développement en série entière de la fonction $z \mapsto b_{\xi_m}^{-1}(z) a_{\xi_m}(z)(1-z)^d$. L'estimateur des moindres carrés est alors donné par

$$\widehat{\theta}_{m,n} \in \operatorname*{argmin}_{\theta_m \in \Theta_m} \tilde{\sigma}_n^2(\theta_m) \quad \text{où} \quad \tilde{\sigma}_n^2(\theta_m) := \frac{1}{n}\sum_{t=1}^{n} \epsilon_t^2(\theta_m)$$

et on définit les matrices suivantes :

$$I_m(\theta_m) := \lim_{n\to\infty} \operatorname{Var}\left[\sqrt{n}\, \frac{\partial}{\partial \theta_m} \tilde{\sigma}_n^2(\theta_m)\right] \tag{35}$$

et

$$J_m(\theta_m) := \lim_{n\to\infty} \frac{\partial^2}{\partial \theta_m \partial \theta_m^\top} \tilde{\sigma}_n^2(\theta_m) \quad \text{a.s.} \tag{36}$$

Le premier résultat notable de cette étude est l'extension du théorème (34) concernant la normalité asymptotique pour l'estimateur des moindres carrés $\widehat{\theta}_{m_0,n}$, établi sous le vrai modèle m_0, et étendu au cas où $d_0 \in \mathbb{R}$. Cette généralisation permet notamment d'intégrer les processus non-stationnaires. Elle est de plus obtenue sans imposer de restriction supplémentaire sur l'espace des paramètres et s'appuie, en particulier, sur les travaux de Hualde et Robinson (2011).

On considère ensuite la sélection des ordres p et q pour les modèles FARIMA faibles,

basée sur la fonction de vraisemblance pseudo-gaussienne donnée par

$$\tilde{L}_n(\theta_m) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\,\tilde{\sigma}_n^2(\theta_m)}} \exp\left\{-\frac{\epsilon_t^2(\theta_m)}{2\,\tilde{\sigma}_n^2(\theta_m)}\right\}$$

ou encore

$$\log \tilde{L}_n(\theta_m) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\tilde{\sigma}_n^2(\theta_m) - \frac{1}{2\,\tilde{\sigma}_n^2(\theta_m)}\sum_{t=1}^{n}\epsilon_t^2(\theta_m).$$

En ne gardant que ce qui dépend du modèle $m = (p, q)$, à un facteur multiplicatif près, et en ajoutant une pénalisation dépendant de la dimension du modèle $\kappa_m = p + q + 1$, on obtient le critère BIC suivant,

$$\mathrm{BIC}_{m,n} := n\log\tilde{\sigma}_n^2(\widehat{\theta}_{m,n}) + \kappa_m \log n,$$

ainsi que son variant le Quasi-BIC (QBIC) défini par

$$\mathrm{QBIC}_{m,n} := n\log\tilde{\sigma}_n^2(\widehat{\theta}_{m,n}) + \mathbb{1}_{A^c_{m,,n}}\log\det\left[n\,\partial^2_{\theta_m}\log\tilde{\sigma}_n^2(\widehat{\theta}_{m,n})\right] + \mathbb{1}_{A_{m,n}}\kappa_m\log n.$$

où $A_{m,n} := \{\det[\partial^2_{\theta_m}\log\tilde{\sigma}_n^2(\widehat{\theta}_{m,n})] \leqslant 0\}$. On pose

$$\widehat{m}_n \in \operatorname*{argmin}_{m\in\mathcal{M}} \mathrm{QBIC}_{m,n}$$

le modèle optimal sélectionné à partir des données. On admet par ailleurs l'identifiabilité des modèles candidats en supposant que la fonction $\theta_m \mapsto \sigma^2(\theta_m)$, limite de $\tilde{\sigma}_n^2(\theta_m)$, admet un minimum unique sur Θ_m. Sous les hypothèses additionnelles d'ergodicité et (H_4) avec $\tau = 4$, on a alors la consistance du critère QBIC (et par conséquent du BIC) au sens où

$$\lim_{n\to\infty} \mathbb{P}_{\theta_0}\left(\widehat{m}_n = m_0\right) = 1.$$

Dans l'ensemble, les simulations effectuées confirment le comportement attendu des critères proposés, avec des performances comparables entre les deux critères (plus de 90% de sélection du vrai modèle). Ce travail établit donc la consistance des critères BIC et QBIC pour une classe très générale de modèles FARIMA, intégrant à la fois la dépendance à long terme et/ou la non-stationnarité via le paramètre de mémoire d_0, ainsi qu'une structure de dépendance non linéaire très générale pour le bruit. Il s'inscrit dans le prolongement des travaux de Beran et al. (1998) et Huang et al. (2022), tout en élargissant leur cadre. Il convient cependant de souligner que Huang et al. (2022) ne font aucune hypo-

thèse d'identifiabilité sur les modèles candidats, ce qui marque une avancée notable dans l'étude de la sélection de modèles pour les processus à mémoire longue. S'affranchir de cette hypothèse d'identifiabilité dans le contexte de processus FARIMA faibles constitue de ce fait une perspective de recherche intéressante.

En terme de sélection de modèles, la minimisation d'un critère oppose fréquemment la consistance du BIC à l'efficacité prédictive de l'AIC. Dans le cadre de cette thèse, un critère AIC adapté aux processus FARIMA faibles a été construit, même si certains développements théoriques restent à finaliser. L'étude de la convergence des moments de l'erreur d'estimation dans le cadre des processus FARIMA faibles, ou, à défaut, l'existence de moments finis d'ordre strictement supérieur à 2, permettrait d'achever ce travail et représente une piste de recherche future prometteuse. On notera que, contrairement au critère BIC et QBIC, le critère AIC prend en compte la structure de dépendance du bruit dans son terme de pénalité ce qui laisse envisager une meilleure adaptation aux données lorsque le bruit présente de fortes dépendances non linéaires. Comme évoqué au début du paragraphe, les performances de l'AIC en terme de sélection restent limités. Néanmoins, ces résultats peuvent être largement optimisés grâce à une adaptation du terme de pénalité qui se rapproche du BIC. Cette nouvelle approche, combinant AIC et BIC, assure une sélection consistante des modèles FARIMA faibles par la minimisation du critère suivant :

$$n \log \tilde{\sigma}_n^2(\widehat{\theta}_{m,n}) + \frac{\log n}{2\,\tilde{\sigma}_n^2(\widehat{\theta}_{m,n})} \operatorname{tr}\{I_m(\widehat{\theta}_{m,n})\, J_m^{-1}(\widehat{\theta}_{m,n})\}$$

où les matrices $I_m(\theta_m)$ et $J_m(\theta_m)$ sont respectivement données par (35) et (36). Puisque $I(\theta_0) = 2\sigma^2 J(\theta_0)$ lorsque le processus des innovations est iid ou forme une suite de différences de martingale, on retrouve exactement le critère BIC dans le cadre fort et semi-fort, tout en conservant une pénalisation capable de s'adapter à une structure de dépendance générale sur le bruit.

Chapter 1

Consistency and asymptotic normality in a class of nearly unstable processes

This chapter is based on the following publication: Badreau and Proïa (2023)

Abstract: This paper deals with inference in a class of stable but nearly-unstable processes. Autoregressive processes are considered, in which the bridge between stability and instability is expressed by a time-varying companion matrix A_n with spectral radius $\rho(A_n) < 1$ satisfying $\rho(A_n) \to 1$. This framework is particularly suitable to understand unit root issues by focusing on the inner boundary of the unit circle. Consistency is established for the empirical covariance and the OLS estimation together with asymptotic normality under appropriate hypotheses when A, the limit of A_n, has a real spectrum, and a particular case is deduced when A also contains complex eigenvalues. The asymptotic process is integrated with either one unit root (located at 1 or -1), or even two unit roots located at 1 and -1. Finally, a set of simulations illustrate the asymptotic behavior of the OLS. The results are essentially proved by L^2 computations and the limit theory of triangular arrays of martingales.

Keywords and phrases: Nearly unstable autoregressive process; OLS estimation; Asymptotic behavior; Unit root; Martingales.

Mathematics Subject Classification: 62M10, 62F12, 60G52, 60G10, 60G42.

1.1 Introduction and motivation

This paper is dedicated to the boundary between stationarity and integration in time series, which has long since proven to be crucial for practitioners, *e.g.* in econometrics or finance. First of all, we refer the reader to Brockwell and Davis (1991) for a large overview

of linear time series. In the context of autoregressive processes (AR), it is well-known that the least squares (OLS) estimator is strongly consistent wherever its characteristic roots lie, see Lai and Wei (1983), but with very different convergence rates and limit distributions whether the process is stationary or integrated, or even explosive. According to the terminology of Duflo (1997), a stationary process will be described as *stable* and an integrated process as *unstable* in what follows. In the first case, the OLS estimator is $\sqrt{n}$-consistent with a Gaussian asymptotic behavior whereas in the other case, it is (at least) n-consistent with an asymmetrical asymptotic behavior that can be written as functionals of a standard Wiener process $(W(t),\, t \in [0,1])$. Precisely, for example in the univariate case, in the stable (left-hand side) and unstable (right-hand side, $\theta = \pm 1$) cases, either

$$\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}(0, 1-\theta^2) \quad \text{or} \quad n\left(\widehat{\theta}_n - \theta\right) \xrightarrow{d} \frac{\pm \int_0^1 W(t)\,\mathrm{d}W(t)}{\int_0^1 W^2(t)\,\mathrm{d}t}$$

and these results are extended to the general case (see *e.g.* the substantial work of Chan and Wei (1988)). This discontinuity has motivated numerous studies devoted to intermediate AR models aimed at bridging these two situations. Most of these studies involve the AR(1) process with time-varying coefficients, which is an easy-to-use relevant case to understand the underlying dynamics. Random coefficients have been suggested to bypass the difficulty: as it is explained in Horváth and Trapani (2016), thanks to a specific self-normalized WLS estimator of θ, "there is no *unit root problem* in case of the RCA model". Indeed, the authors get the asymptotic normality irrespective of the average value of the coefficient, only the rate is affected. In a different context, the so-called *volatility induced stationarity* detailed *e.g.* in Ling (2004) and Nielsen and Rahbek (2014), enable to accommodate integration and stationarity in a double autoregression through stochastic volatility despite a unit root in the conditional mean of the process, see also Trapani (2021). We refer the reader to all the references contained in those papers. An AR(1) with non-random but time-varying coefficients may also be written in the triangular form [1]

$$\forall\, n \geqslant 1,\ \forall\, 1 \leqslant k \leqslant n, \qquad X_{n,k} = \theta_n\, X_{n,k-1} + \varepsilon_k$$

1. Double indexing is customary to such representations: $X_{n,k}$ is the k-th observation of a time series of size n (apart from the initial value). The triangular form of the process is $\{X_{1,0}, X_{1,1}\}, \{X_{2,0}, X_{2,1}, X_{2,2}\}, \ldots, \{X_{n,0}, \ldots, X_{n,n}\}$.

where (ε_k) is a sequence of zero-mean perturbations (usually independent and identically distributed (i.i.d.) random variables, or differences of martingale) and $X_{n,0}$ is an arbitrary initial value. In that case, the OLS estimator is obviously given by

$$\widehat{\theta}_n = \frac{\sum_{k=1}^n X_{n,k-1} X_{n,k}}{\sum_{k=1}^n X_{n,k-1}^2}. \tag{1.1.1}$$

To focus on the inner neighborhood of the unit root, the idea is to consider a sequence of coefficients that satisfies $|\theta_n| < 1$ for all $n \geqslant 1$ but $|\theta_n| \to 1$, so that a new model corresponds to each new observation, always stable but increasingly close to instability. Chan and Wei (1987) show that, letting $\theta_n = 1 - c/n$ and under appropriate assumptions, the OLS estimator (1.1.1) is n-consistent whether $c = 0$ (which corresponds to the standard integrated AR(1) process) or $c > 0$. Once self-normalized, it is shown in Thm. 1 that the estimation error has a non-Gaussian limit distribution $\mathcal{L}(c)$ which is asymptotically $\mathcal{N}(0,1)$ as $c \to +\infty$ (see also Thm. 2 of Phillips (1987b)). In this context, it is natural to conjecture that any rate faster than n shall lead to the same conclusion. But slowing down the convergence of θ_n to 1 in order to make the aforementioned bridge, Giraitis and Phillips (2006) and later Phillips and Magdalinos (2007) suggest to fix $\theta_n = 1 - c/v_n$ for $1 \ll v_n \ll n$. Among other results, they establish that

$$\sqrt{n\,v_n}\left(\widehat{\theta}_n - \theta_n\right) \xrightarrow{d} \mathcal{N}(0,\, 2c)$$

as soon as $c > 0$, that is in the stable but nearly unstable case, and that the result also holds, under suitable assumptions, around the negative unit root with $\theta_n = -1 + c/v_n$. In the special case where $v_n = n^\alpha$ $(0 < \alpha < 1)$, the rate is $n^{(1+\alpha)/2}$ and the authors note that the interval $\sqrt{n} \ll n^{(1+\alpha)/2} \ll n$ is covered but they emphasize that boundaries do not match. The asymptotic variance is overestimated for $\alpha \to 0^+$ ($2c$ instead of $2c - c^2$ when $\alpha = 0$) whereas the Gaussian limit distribution for $\alpha \to 1^-$ is no longer Gaussian for $\alpha = 1$. Note that the authors also consider the explosive case $|\theta_n| > 1$, whereas we will only focus here on the stable alternative to the unstable process. In the same vein, Phillips and Lee (2015) develop a limit theory for nonstationary vector autoregressions with mixed roots in the vicinity of unity involving persistent and explosive components.

Now for any $p \geqslant 1$, consider the process generated according to the triangular form

$$\forall\, n \geqslant 1,\ \forall\, 1 \leqslant k \leqslant n, \qquad X_{n,k} = \sum_{i=1}^{p} \theta_{n,i}\, X_{n,k-i} + \varepsilon_k \tag{1.1.2}$$

where (ε_k) is a sequence of zero-mean i.i.d. random variables with variance $\sigma^2 > 0$. In an equivalent way, the vector expression of this process is

$$\Phi_{n,k} = A_n \, \Phi_{n,k-1} + E_k \tag{1.1.3}$$

where $E_k = (\varepsilon_k, 0, \ldots, 0)^\top$ is a p-vectorial noise, $\Phi_{n,k} = (X_{n,k}, \ldots, X_{n,k-p+1})^\top$ and

$$A_n = \begin{pmatrix} \theta_{n,1} & \theta_{n,2} & \cdots & \theta_{n,p} \\ & I_{p-1} & & 0 \end{pmatrix}$$

is the associated $p \times p$ companion matrix. The initial value $\Phi_{n,0}$ is supposed to have a finite moment of order 2 and to be independent of (E_k). The OLS estimator of $\theta_n = (\theta_{n,1}, \ldots, \theta_{n,p})^\top$ is then given by

$$\widehat{\theta}_n = S_{n,n-1}^{-1} \sum_{k=1}^{n} \Phi_{n,k-1} \, X_{n,k} \quad \text{where} \quad S_{n,n-1} = \sum_{k=0}^{n-1} \Phi_{n,k} \, \Phi_{n,k}^\top. \tag{1.1.4}$$

Note that we may add a small $\epsilon > 0$ to $S_{n,n-1}$ to avoid a useless invertibility assumption, without disturbing the asymptotic behavior. [2] The p-dimensional process (1.1.3) is stable when $\rho(A_n) < 1$, that is, when the largest modulus of its eigenvalues is less than 1, see Def. 2.3.17 of Duflo (1997). Since the eigenvalues of A_n are the inverses of the roots of the complex polynomial $\Theta_n(z) = 1 - \theta_{n,1} \, z - \ldots - \theta_{n,p} \, z^p$, this is equivalent to say that $\Theta_n(z) \neq 0$ for all $|z| \leqslant 1$. On the contrary, it is unstable when $\rho(A_n) = 1$. Along the same lines, we will thus consider that $\rho(A_n) < 1$ for all $n \geqslant 1$ but $\rho(A_n) \to 1$, which corresponds to a stable but nearly unstable AR(p) process with time-varying coefficients. In this context and under suitable assumptions, Proïa (2020) has established some moderate deviation principles for the empirical covariance and the OLS estimator that hold for any $p \geqslant 1$ (extending those of Miao et al. (2015) valid for $p = 1$). In particular, the sequence

$$\left(\frac{\sqrt{n}}{b_n \, (1 - \rho(A_n))^{1/2}} \left(\widehat{\theta}_n - \theta_n \right) \right)_{n \geqslant 1}$$

satisfies a large deviation principle with a speed (b_n^2) and a rate function depending on the renormalized limit covariance of the process. Let us also mention the weak unit roots of

2. To be rigorous, one should write $\widehat{\theta}_{n,n}$ instead of $\widehat{\theta}_n$ to emphasize that the OLS is a function of $X_{n,0}, \ldots, X_{n,n}$. Similarly, S_n will be used for $S_{n,n}$ (and T_n for $T_{n,n}$, etc.) to lighten the notation when no confusion can arise.

Park (2003) containing more applications than ours, including faster rates of convergence for the nearly unit root and non-linear models, but in a more restricted setting (we will come back to this in due time), or the recent work of Buchmann and Chan (2013) who introduce a perturbation in the Jordan canonical form of the AR(p) model (see Thm. 2.1) and get a set of convergences in a context close to ours (although more general). However, by directly dealing with the spectral radius of the companion matrix, the approach of this paper seems easier to interpret and we will see that *in fine* different kind of results are obtained *via* different technical tools. Now, to complete these deviations and to generalize the results of Phillips and Magdalinos (2007), we aim at proving the consistency and asymptotic normality of the OLS estimator (1.1.4). The strategy remains the same, but the calculation steps will prove to be much trickier. In the second section, the assumptions and main results are provided together with comments and examples. The third section is dedicated to the technical proofs whereas the fourth section is the empirical part of the paper, containing simulations. A quick conclusion with considerations about further improvements ends the paper.

1.2 Main results

First, let us start by describing three technical hypotheses that will be needed to achieve our goals. In particular, the first one is a matter of simplification of the reasonings since A_n turns out to be diagonalizable for a sufficiently large n, and that specificity will prove to be very useful. The second hypothesis is related to the number of unit roots in the asymptotic process (either ± 1, or both 1 and -1). The third assumption characterizes our stable but nearly-unstable setting. The complex eigenvalues are sorted according to their modulus (in descending order), with ties broken by lexicographic order (also in descending order).

(H_1) *Convergence of the companion matrix.* There exists a $p \times p$ matrix A such that

$$\lim_{n \to +\infty} A_n = A$$

with distinct eigenvalues $0 < |\lambda_p| \leqslant \ldots \leqslant |\lambda_2| \leqslant |\lambda_1| = \rho(A) = 1$, and the top-right element of A is non-zero ($\theta_p \neq 0$).

(H_2) *Number of unit roots.* Either (H_{21}) or (H_{22}) is true.

(H_{21}) There is exactly one unit root in A ($\lambda_1 = \pm 1$ but $|\lambda_2| < 1$ if $p \geqslant 2$).

(H_{22}) There is exactly two unit roots in A ($\lambda_1 = 1$, $\lambda_2 = -1$ but $|\lambda_3| < 1$ if $p \geqslant 3$).

(H_3) *Spectral radius of the companion matrix.* The spectral radius of A_n is given by

$$\rho(A_n) = 1 - \frac{c}{v_n}$$

for some $c > 0$ and $1 \ll v_n \ll n$.

Most of our results will be stated under (H_{21}) but we will explain at the end of the section that, in fact, they are still valid under (H_{22}) with only slight adjustments of the rates and limit behaviors. For readability purposes, the calculations will not be developed in that case but left to the reader since they follow exactly the same lines. A decisive argument in the technical part of the paper rests on the diagonalization of A_n, on which we will give more details in due course. But to summarize, we will explain that there exists $n_0 \geqslant 1$ such that, for all $n > n_0$, $A_n = P_n D_n P_n^{-1}$ where $D_n = \mathrm{diag}(\lambda_{n,1}, \ldots, \lambda_{n,p})$ contains the ordered distinct eigenvalues $1 > |\lambda_{n,1}| \geqslant \ldots \geqslant |\lambda_{n,p}| > 0$ of A_n which converge to those of A (see (H_1) above), and the basis of eigenvectors of A_n can be written in the standardized form

$$P_n = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \frac{1}{\lambda_{n,1}} & \frac{1}{\lambda_{n,2}} & \cdots & \frac{1}{\lambda_{n,p}} \\ \vdots & \vdots & & \vdots \\ \frac{1}{\lambda_{n,1}^{p-1}} & \frac{1}{\lambda_{n,2}^{p-1}} & \cdots & \frac{1}{\lambda_{n,p}^{p-1}} \end{pmatrix}. \tag{1.2.1}$$

In addition, $P_n \to P$ and $P_n^{-1} \to P^{-1}$ whose entries of the first column are denoted by $\pi_{11}, \ldots, \pi_{p1}$. The $p \times p$ symmetric matrix H and its standardized version H_0, that will play the role of the precision matrix in the asymptotic normality of the OLS estimate, are then defined as

$$H = \sigma^2 \begin{pmatrix} \frac{\pi_{11}^2}{2} & 0 & \cdots & 0 \\ 0 & \frac{\pi_{21}^2}{1-\lambda_2^2} & \cdots & \frac{\pi_{21}\pi_{p1}}{1-\lambda_2\lambda_p} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \frac{\pi_{p1}\pi_{21}}{1-\lambda_p\lambda_2} & \cdots & \frac{\pi_{p1}^2}{1-\lambda_p^2} \end{pmatrix} \triangleq \sigma^2 H_0. \tag{1.2.2}$$

From now on, to lighten the expressions we will rather use $\rho_n \triangleq \rho(A_n)$. Also, for any $d \geqslant 1$, the usual $d \times d$ matrices

$$I_d = \begin{pmatrix} 1 & 0 & \cdots \\ 0 & \ddots & \\ \vdots & & 1 \end{pmatrix} \quad \text{and} \quad K_d = \begin{pmatrix} 1 & 0 & \cdots \\ 0 & 0 & \\ \vdots & & \ddots \end{pmatrix} \tag{1.2.3}$$

will be frequently encountered, both in statements and proofs. Similarly, the first vector of the canonical basis of $\mathbb{R}^d$ will be denoted by $e_d = (1, 0, \ldots, 0)^\intercal$. Finally, $f_n \asymp g_n$ will have the meaning that both $f_n = O(g_n)$ and $g_n = O(f_n)$.

1.2.1 Causal representation and memory

Under (H_3), there exists a causal representation of $(\Phi_{n,k})$ given by

$$\forall\, n \geqslant 1,\ \forall\, 1 \leqslant k \leqslant n, \quad \Phi_{n,k} = \sum_{\ell \geqslant 0} A_n^\ell\, E_{k-\ell}$$

that directly leads to the autocovariance function

$$\forall\, h \geqslant 0, \quad \Gamma_n(h) = \sigma^2 A_n^h \sum_{\ell \geqslant 0} A_n^\ell\, K_p\, (A_n^\intercal)^\ell \quad \text{and} \quad \Gamma_n(-h) = \Gamma_n^\intercal(h) \tag{1.2.4}$$

provided that the noise has a finite moment of order 2. In particular, we have the following result related to the memory of the process.

Proposition 1.2.1. *Assume that (H_1), (H_2) and (H_3) hold, and that $\mathbb{E}[\varepsilon_1^2] = \sigma^2 < +\infty$. Then,*

$$\forall\, n > n_0, \quad \left\| \sum_{h \in \mathbb{Z}} \Gamma_n(h) \right\| < +\infty.$$

Besides, as $n \to +\infty$,

$$\sum_{h \in \mathbb{Z}} \Gamma_n(h) \asymp \begin{cases} (1-\rho_n)^{-2} & \text{under } (\mathrm{H}_{21}) \text{ with } \lambda_1 = 1 \text{ or } (\mathrm{H}_{22}), \\ 1 & \text{under } (\mathrm{H}_{21}) \text{ with } \lambda_1 = -1. \end{cases}$$

Proof. See Section 1.3.2. □

In other terms, using the terminology of Sec. 1.3.1 of Beran et al. (2013), $(\Phi_{n,k})$ has a *short memory* at fixed n whereas, as n tends to infinity, it turns to a *long memory* process when $\lambda_1 = 1$ but keeps a short memory when $\lambda_1 = -1$. This is especially clear in the univariate case for which

$$\sum_{h \in \mathbb{Z}} \gamma_n(h) = \frac{\sigma^2}{(1-\theta_n)^2}$$

is an increasing function of θ_n, minimal for $\theta_n \to -1^+$ and diverging to infinity for $\theta_n \to 1^-$. We are now ready to state the following result dedicated to the asymptotic behavior of the covariance matrix of the process.

Proposition 1.2.2. *Assume that* (H_1)*,* (H_{21}) *and* (H_3) *hold, and that* $\mathbb{E}[\varepsilon_1^2] = \sigma^2 < +\infty$*. Then, the empirical covariance matrix* $S_{n,n} = S_n$ *given in* (1.1.4) *satisfies*

$$(1-\rho_n)\,\frac{S_n}{n} \xrightarrow{\ p\ } \frac{\sigma^2}{2}\,\mathrm{vec}^{-1}\big((A^* \otimes A^*)\,e_{p^2}\big) \triangleq \Gamma \quad \textit{with} \quad A^* = P\,K_p\,P^{-1}$$

where $\mathrm{vec}^{-1} : \mathbb{R}^{p^2} \to \mathbb{R}^{p\times p}$ *is the vectorization inverse operator and the invertible matrix* P *is the limit of* P_n *in* (1.2.1)*. In addition,*

$$(1-\rho_n)\left\|\frac{S_n}{n} - \Gamma_n\right\| \xrightarrow{\ p\ } 0 \quad \textit{where} \quad \Gamma_n = \Gamma_n(0) = \sigma^2 \sum_{\ell \geqslant 0} A_n^\ell\,K_p\,(A_n^\top)^\ell$$

is the covariance of the stationary process (at fixed n*).*

Proof. See Section 1.3.2. □

Remark 1.2.1. For the empirical covariance of the process to be consistent, it must be renormalized by $1-\rho_n$. In the stable case (stationary and ergodic), it is well-known that S_n converges at rate n whereas in the unstable case (integrated with one unit root, either positive or negative), S_n converges at rate n^2. Let $p = 1$ to simplify. Then, under suitable hypotheses, in the stable (left-hand side) and unstable (right-hand side) cases, either

$$\frac{S_n}{n} \xrightarrow{\ a.s.\ } \gamma_0 \quad \text{or} \quad \frac{S_n}{n^2} \xrightarrow{\ d\ } \sigma^2 \int_0^1 W^2(t)\,\mathrm{d}t$$

where $\gamma_0 > 0$ is the variance of the stationary process and $(W(t),\, t \in [0,1])$ is a standard Wiener process. Proposition 1.2.2 establishes the convergence of $S_{n,n}$ at rate $n/(1-\rho_n) \propto n^{1+\alpha}$ for $v_n = n^\alpha$ and $0 < \alpha < 1$. One can see that our model bridges the stable and unstable cases in terms of empirical covariance even if the limit behaviors as $\alpha \to 0^+$ and $\alpha \to 1^-$ do not make a connection with $\alpha = 0$ and $\alpha = 1$.

Even if it is instructive to control the behavior of the covariance matrix in a stationary process, that result shall *not* help to prove the following theorems, dedicated to the OLS estimator, since the limit matrix is not invertible as soon as $p > 1$. Indeed a direct calculation shows that $(A^* \otimes A^*)\,e_{p^2}$ is a column vector of size p^2 with entries $\pm\pi_{11}^2$ and different rates of convergence must be used to achieve our objectives.

1.2.2 OLS estimation

In view of the above, consider the $p \times p$ diagonal matrix defined as

$$V_n = \mathrm{diag}(1 - \rho_n, 1, \ldots, 1) \tag{1.2.5}$$

together with the $p^2 \times p^2$ matrix given, for $n > n_0$, by

$$W_n = (V_n^{1/2} P_n^{-1}) \otimes (V_n^{1/2} P_n^{-1}) \tag{1.2.6}$$

where P_n is the (invertible) basis of eigenvectors of A_n.

Theorem 1.2.3. *Assume that (H_1), (H_{21}) and (H_3) hold, and that $\mathbb{E}[\varepsilon_1^2] = \sigma^2 < +\infty$. Then we have the consistency*

$$\left\| \widehat{\theta}_n - \theta_n \right\| \xrightarrow{p} 0$$

where $\widehat{\theta}_n$ is the OLS estimator (1.1.4) *in the nearly unstable $AR(p)$ process* (1.1.2).

Proof. See Section 1.3.3. □

For the asymptotic normality, we need a slightly stronger hypothesis on the noise (ε_k). Let us assume from now on that there exists a moment of order strictly greater than 2.

Theorem 1.2.4. *Assume that (H_1), (H_{21}) and (H_3) hold, and that $\mathbb{E}[|\varepsilon_1|^{2+\nu}] = \eta_\nu < +\infty$ for some $\nu > 0$. Then, if the eigenvalues of A are real, we have the asymptotic normality*

$$\sqrt{n}\, V_n^{-1/2} P_n^T \left(\widehat{\theta}_n - \theta_n \right) \xrightarrow{d} \mathcal{N}_p(0, H_0^{-1})$$

where $\widehat{\theta}_n$ is the OLS estimator (1.1.4) *in the nearly unstable $AR(p)$ process* (1.1.2)*, the matrix of rates V_n is given in* (1.2.5)*, P_n is given in* (1.2.1)*, and the standardized positive definite precision matrix H_0 is given in* (1.2.2)*. If some eigenvalues of A are complex, we have the asymptotic normality*

$$\sqrt{n\, v_n} \left\langle L_n, \widehat{\theta}_n - \theta_n \right\rangle \xrightarrow{d} \mathcal{N}(0, h_0^2)$$

where

$$L_n = \left(1, \frac{1}{\lambda_1 \rho_n}, \ldots, \frac{1}{(\lambda_1 \rho_n)^{p-1}} \right)^\top \quad \text{and} \quad h_0^2 = \frac{2c}{\pi_{11}^2} > 0.$$

Proof. See Section 1.3.4. □

Corollary 1.2.5. *Under the same assumptions as in Theorem 1.2.4 with $p = 1$, we have the asymptotic normality*

$$\sqrt{n\,v_n}\left(\widehat{\theta}_n - \theta_n\right) \xrightarrow{d} \mathcal{N}(0,\, 2c).$$

Proof. Just take $V_n = 1 - \rho_n = c/v_n$, $P_n = 1$ and $H_0 = 1/2$ in Theorem 1.2.4. □

One can note that this is precisely the statement of Thm. 3.2(c) of Phillips and Magdalinos (2007) when the unit root is positive ($\lambda_1 = 1$). Nevertheless, our result also holds for a negative unit root ($\lambda_1 = -1$) whereas in the reference just mentioned, an extra symmetry argument is needed on the distribution of (ε_k). We may therefore say that Theorem 1.2.4 reinforces and extends the result to the general near-stationary setting.

Remark 1.2.2. On the basis of Theorem 1.2.3, the triangle inequality directly implies that, under the same hypotheses,

$$\widehat{\theta}_n \xrightarrow{p} \theta. \tag{1.2.7}$$

But it is important to note that Theorem 1.2.4 cannot lead to the corresponding asymptotic normality. To understand this, consider for example the univariate setting of Corollary 1.2.5 where $\theta_n - \theta \propto 1/v_n$. Then,

$$\sqrt{n\,v_n}\left|\theta_n - \theta\right| \propto \sqrt{\frac{n}{v_n}} \longrightarrow +\infty.$$

However, this is not surprising. Indeed, the limit AR process generated by θ is unstable. In that case, the OLS estimate is still consistent but not asymptotically normal.

In terms of estimation, the mixing induced by P_n^T is somewhat troublesome because it relies on the true and unknown eigenvalues of A_n. Since the eigenvalues of a square matrix depend continuously on its entries, see *e.g.* Thm. 2.4.9.2 of Horn and Johnson (2012), the consistency of $\widehat{\theta}_n$ would justify replacing A_n by $\widehat{A}_n$ before calculating its spectrum, so as to get consistent estimates of all quantities appearing in Theorem 1.2.4 and build hypotheses tests. For simulations (Section 1.4), the following corollary will be used to illustrate the latter result.

Corollary 1.2.6. *Assume that the assumptions of Theorem 1.2.4 hold. Then, for $\lambda_1 = \pm 1$, we have the asymptotic distribution*

$$\frac{\pi_{11}^2\, n\, v_n}{2\,c}\left[\sum_{i=1}^{p} \frac{1}{(\lambda_1 \rho_n)^{i-1}}\left(\widehat{\theta}_{n,\,i} - \theta_{n,\,i}\right)\right]^2 \xrightarrow{d} \chi_1^2.$$

Proof. This is in fact the first marginal convergence of the asymptotic normality combined with the continuous mapping theorem, (1.2.2) and the fact that the first column of P_n is $(1, 1/\lambda_{n,1}, \ldots, 1/\lambda_{n,1}^{p-1})$. □

Before switching to the proofs of the results, some examples are provided. The goal is to get a clear insight into the structure of the fundamental matrix H_0 in simple cases.

Examples. Let us consider the examples $p \in \{2, 3, 4\}$ with $\lambda_1 = \pm 1$ (see Corollary 1.2.5 for $p = 1$). In fact, H_0 has a simple form deduced from the expressions

$$\forall\, 1 \leqslant i \leqslant p, \qquad f_i = \prod_{j \neq i} (\lambda_j - \lambda_i).$$

— For $p = 2$ and $\mathrm{sp}(A) = \{\pm 1, \lambda_2\}$ with $0 < |\lambda_2| < 1$,

$$H_0 = \begin{pmatrix} \frac{\lambda_1^2}{2 f_1^2} & 0 \\ 0 & \frac{\lambda_2^2}{f_2^2 (1-\lambda_2^2)} \end{pmatrix}.$$

— For $p = 3$ and $\mathrm{sp}(A) = \{\pm 1, \lambda_2, \lambda_3\}$ with $0 < |\lambda_3| \leqslant |\lambda_2| < 1$ and $\lambda_2 \neq \lambda_3$,

$$H_0 = \begin{pmatrix} \frac{\lambda_1^4}{2 f_1^2} & 0 & 0 \\ 0 & \frac{\lambda_2^4}{f_2^2 (1-\lambda_2^2)} & \frac{\lambda_2^2 \lambda_3^2}{f_2 f_3 (1-\lambda_2 \lambda_3)} \\ 0 & \frac{\lambda_3^2 \lambda_2^2}{f_3 f_2 (1-\lambda_3 \lambda_2)} & \frac{\lambda_3^4}{f_3^2 (1-\lambda_3^2)} \end{pmatrix}.$$

— For $p = 4$ and $\mathrm{sp}(A) = \{\pm 1, \lambda_2, \lambda_3, \lambda_4\}$ with $0 < |\lambda_4| \leqslant |\lambda_3| \leqslant |\lambda_2| < 1$ and $\lambda_2 \neq \lambda_3 \neq \lambda_4$,

$$H_0 = \begin{pmatrix} \frac{\lambda_1^6}{2 f_1^2} & 0 & 0 & 0 \\ 0 & \frac{\lambda_2^6}{f_2^2 (1-\lambda_2^2)} & \frac{\lambda_2^3 \lambda_3^3}{f_2 f_3 (1-\lambda_2 \lambda_3)} & \frac{\lambda_2^3 \lambda_4^3}{f_2 f_4 (1-\lambda_2 \lambda_4)} \\ 0 & \frac{\lambda_3^3 \lambda_2^3}{f_3 f_2 (1-\lambda_3 \lambda_2)} & \frac{\lambda_3^6}{f_3^2 (1-\lambda_3^2)} & \frac{\lambda_3^3 \lambda_4^3}{f_3 f_4 (1-\lambda_3 \lambda_4)} \\ 0 & \frac{\lambda_4^3 \lambda_2^3}{f_4 f_2 (1-\lambda_4 \lambda_2)} & \frac{\lambda_4^3 \lambda_3^3}{f_4 f_3 (1-\lambda_4 \lambda_3)} & \frac{\lambda_4^6}{f_4^2 (1-\lambda_4^2)} \end{pmatrix}.$$

As mentioned at the beginning of the section, Proposition 1.2.2 and Theorems 1.2.3 and 1.2.4 are true under (H_{22}), *i.e.* when A contains the eigenvalues 1 and -1 ordered this way (according to lexicographic descending order). That could correspond to a situation in which $\lambda_{n,1} = 1 - c/v_n$ and $\lambda_{n,2} = -1 + d/w_n$ with $c, d > 0$, $1 \ll v_n, w_n \ll n$ and $c/v_n \leqslant d/w_n$ (otherwise switching $\lambda_{n,1}$ and $\lambda_{n,2}$ in P_n and D_n). More details will be given in due course (Remarks 1.3.1–1.3.4). Let us now prove our results.

1.3 Technical tools

In this section, C_{st} denotes a generic positive constant that is not necessarily identical from one line to another and we use the conventions $\sum_{\varnothing} = 0$ and $\Pi_{\varnothing} = 1$. Let us also define a fundamental matrix for the reasoning to come,

$$B_n = I_{p^2} - A_n \otimes A_n. \tag{1.3.1}$$

In all the sequel, $\|\cdot\|$ will refer to the Frobenius norm $\|\cdot\|_F = \|\mathrm{vec}(\cdot)\|_2$ induced by the inner product $\langle\!\langle \cdot, \cdot \rangle\!\rangle = \langle\!\langle \cdot, \cdot \rangle\!\rangle_F = \langle \mathrm{vec}(\cdot), \mathrm{vec}(\cdot) \rangle$. The distinction is not made between matrices and vectors for which we simply have $\|\cdot\|_F = \|\cdot\|_2$ and $\langle\!\langle \cdot, \cdot \rangle\!\rangle_F = \langle \cdot, \cdot \rangle$.

1.3.1 Linear Algebra

This section gathers all the linear algebra tools that we shall need in the proofs of our results.

Lemma 1.3.1. *Assume that (H_1) holds. Then, there exists $n_0 \geqslant 1$ such that, for all $n > n_0$, A_n is diagonalizable in the form $A_n = P_n D_n P_n^{-1}$ with $D_n = \mathrm{diag}(\lambda_{n,1}, \ldots, \lambda_{n,p})$ containing ordered distinct eigenvalues $1 > |\lambda_{n,1}| \geqslant \ldots \geqslant |\lambda_{n,p}| > 0$. In addition, $\|P_n\| \leqslant C_{st}$ and $\|P_n^{-1}\| \leqslant C_{st}$.*

Proof. See Lem. 3.1 in Proïa (2020). □

Lemma 1.3.2. *Assume that (H_1) holds. Then in the context of Lemma 1.3.1, for all $n > n_0$, the first row of P_n^{-1} is real.*

Proof. This is obvious for $p = 1$, so let $p > 1$. Suppose that, for $j \geqslant 2$, $\lambda_{n,j} \in \mathrm{sp}(A_n)$ is a complex eigenvalue so that, since A_n is real, we must also have $\bar{\lambda}_{n,j} \in \mathrm{sp}(A_n)$. Taking $i = 1$ in formula (3.4) of Proïa (2020), the first row of P_n^{-1} is written

$$\left(\frac{T_{n,1,k}\, \lambda_{n,1}^{p-1} \prod_{j=2}^{p} \lambda_{n,j}}{\prod_{j=2}^{p} (\lambda_{n,j} - \lambda_{n,1})} \right)_{1 \leqslant k \leqslant p}$$

where $T_{n,1,k}$ can be retrieved from the relation

$$\prod_{j=2}^{p} \left(X - \frac{1}{\lambda_{n,j}} \right) = T_{n,1,1} + T_{n,1,2}\, X + \ldots + T_{n,1,p}\, X^{p-1}$$

for real values of X. Combining such terms together with the fact that, for any $x \in \mathbb{R}$ and for any $z \in \mathbb{C}$,

$$(x-z)(x-\bar{z}) = x^2 + |z|^2 - 2\,\mathrm{Re}(z)x \in \mathbb{R} \quad \text{and} \quad \lambda_{n,1} = \pm\rho_n \in \mathbb{R},$$

each entry of the first row of P_n^{-1} is real. □

Lemma 1.3.3. *Assume that (H_1) holds. Then, each element of the first column of P^{-1} is non-zero.*

Proof. We remind that the entries of the first column of P^{-1} are denoted by π_{k1}. Then, taking $j = 1$ in the same formula as in the previous proof,

$$\pi_{k1} = \frac{(-\lambda_k)^{p-1}}{\prod_{\ell \neq k}(\lambda_\ell - \lambda_k)}. \tag{1.3.2}$$

Under (H_1), all eigenvalues are distinct and non-zero, which concludes the proof. □

Lemma 1.3.4. *Assume that (H_3) holds. Then, for all $n \geqslant 1$, the matrix B_n given in* (1.3.1) *is invertible.*

Proof. Indeed, since $\rho(A_n \otimes A_n) = \rho_n^2 < 1$ under (H_3), the conclusion follows *e.g.* from Cor. 5.6.16 of Horn and Johnson (2012). □

Lemma 1.3.5. *Assume that (H_1), (H_{21}) and (H_3) hold. Then,*

$$\lim_{n \to +\infty} (1-\rho_n)\, B_n^{-1} = \frac{1}{2}\,(A^* \otimes A^*)$$

where A^* is defined in Proposition 1.2.2.

Proof. Once again the result is obvious for $p = 1$, so let $p > 1$. Because $\rho(A_n \otimes A_n) < 1$ (see the proof of Lemma 1.3.4),

$$B_n^{-1} = \sum_{k=0}^{+\infty} A_n^k \otimes A_n^k.$$

For $n > n_0$, it follows from simple linear algebra that

$$B_n^{-1} = (P_n \otimes P_n) \left[\sum_{k=0}^{+\infty} D_n^k \otimes D_n^k \right] (P_n^{-1} \otimes P_n^{-1}). \tag{1.3.3}$$

From Lem. 3.1 of Proïa (2020), we know that $P_n \to P$ where the limit matrix P is invertible under (H_1), as a Vandermonde matrix with distinct entries. Thus we also have

$P_n^{-1} \to P^{-1}$. It remains to observe that $D_n^k \otimes D_n^k$ is a $p^2 \times p^2$ diagonal matrix with top-left element $(\pm\rho_n)^{2k}$ and all other elements given either by $(\pm\rho_n)^k \lambda_{n,i}^k$, depending on whether $\lambda_1 = 1$ or $\lambda_1 = -1$, or by $\lambda_{n,i}^k \lambda_{n,j}^k$ for $2 \leqslant i, j \leqslant p$. Under (H_3),

$$\lim_{n \to +\infty} (1 - \rho_n) \sum_{k=0}^{+\infty} (\pm\rho_n)^{2k} = \frac{1}{2} \tag{1.3.4}$$

and, for $2 \leqslant i, j \leqslant p$,

$$\lim_{n \to +\infty} \sum_{k=0}^{+\infty} (\pm\rho_n)^k \lambda_{n,i}^k = \frac{1}{1 \mp \lambda_i} \quad \text{and} \quad \lim_{n \to +\infty} \sum_{k=0}^{+\infty} \lambda_{n,i}^k \lambda_{n,j}^k = \frac{1}{1 - \lambda_i \lambda_j}. \tag{1.3.5}$$

Under (H_{21}), we can conclude that

$$\lim_{n \to +\infty} (1 - \rho_n) B_n^{-1} = \frac{1}{2} (P \otimes P) K_{p^2} (P^{-1} \otimes P^{-1}) \tag{1.3.6}$$

where K_{p^2} is given in (1.2.3). In a more 'elegant' way,

$$\frac{1}{2} (P \otimes P) K_{p^2} (P^{-1} \otimes P^{-1}) = \frac{1}{2} (P K_p P^{-1}) \otimes (P K_p P^{-1})$$

using the fact that $K_{p^2} = K_p \otimes K_p$, which concludes the proof. □

Remark 1.3.1. If (H_{22}) is true instead of (H_{21}), the lemma still holds and the reasoning is similar but K_{p^2} in (1.3.6) must be replaced by the matrix of size $p^2 \times p^2$ having the diagonal block structure

$$K_{p^2}^* = \begin{pmatrix} K_p & & \\ & \bar{K}_p & \\ & & 0 \end{pmatrix} \quad \text{with} \quad \bar{K}_p = \mathrm{diag}(0, 1, 0, \ldots, 0).$$

Lemma 1.3.6. *Assume that (H_1) and (H_{21}) hold, and that $p > 1$. Then, the $(p-1)\times(p-1)$ bottom-right block of H_0, which is the symmetric matrix given by*

$$\Lambda = \begin{pmatrix} \frac{\pi_{21}^2}{1-\lambda_2^2} & \cdots & \frac{\pi_{21}\pi_{p1}}{1-\lambda_2\lambda_p} \\ \vdots & \ddots & \vdots \\ \frac{\pi_{p1}\pi_{21}}{1-\lambda_p\lambda_2} & \cdots & \frac{\pi_{p1}^2}{1-\lambda_p^2} \end{pmatrix}, \tag{1.3.7}$$

is invertible.

Proof. First under (H_1), Lemma 1.3.3 implies $\pi_{k1} \neq 0$ for all k, so the result is obvious for $p = 2$. Now let $p > 2$ and denote by Λ_k the $(k-1)\times(k-1)$ top-left submatrix of Λ (so that $\Lambda_p = \Lambda$) and note that, although it is not a Cauchy matrix, it is closely related to it (see Sec. 0.9.12 of Horn and Johnson (2012)). Let us adopt the usual reasoning for calculating the determinant of such matrices. Multiplying the $(j-1)$-th column by $(1-\lambda_j\lambda_k)/(\pi_{j1}\pi_{k1})$ for $2 \leqslant j \leqslant k$, and subtracting the last column from each other column, it follows that

$$d_k = \prod_{j=2}^{k} \frac{\pi_{j1}\pi_{k1}}{(1-\lambda_j\lambda_k)} \times \begin{vmatrix} \frac{b_2(\lambda_2-\lambda_k)^2}{(1-\lambda_2^2)(1-\lambda_2\lambda_k)} & \frac{b_2(\lambda_2-\lambda_k)(\lambda_3-\lambda_k)}{(1-\lambda_2\lambda_3)(1-\lambda_2\lambda_k)} & \cdots & \frac{b_2(\lambda_2-\lambda_k)(\lambda_{k-1}-\lambda_k)}{(1-\lambda_2\lambda_{k-1})(1-\lambda_2\lambda_k)} & * \\ \frac{b_3(\lambda_3-\lambda_k)(\lambda_2-\lambda_k)}{(1-\lambda_3\lambda_2)(1-\lambda_3\lambda_k)} & \frac{b_3(\lambda_3-\lambda_k)^2}{(1-\lambda_3^2)(1-\lambda_3\lambda_k)} & \cdots & \frac{b_3(\lambda_3-\lambda_k)(\lambda_{k-1}-\lambda_k)}{(1-\lambda_3\lambda_{k-1})(1-\lambda_3\lambda_k)} & * \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{b_{k-1}(\lambda_{k-1}-\lambda_k)(\lambda_2-\lambda_k)}{(1-\lambda_{k-1}\lambda_2)(1-\lambda_{k-1}\lambda_k)} & \frac{b_{k-1}(\lambda_{k-1}-\lambda_k)(\lambda_3-\lambda_k)}{(1-\lambda_{k-1}\lambda_3)(1-\lambda_{k-1}\lambda_k)} & \cdots & \frac{b_{k-1}(\lambda_{k-1}-\lambda_k)^2}{(1-\lambda_{k-1}^2)(1-\lambda_{k-1}\lambda_k)} & * \\ 0 & 0 & \cdots & 0 & 1 \end{vmatrix}$$

where $d_k = \det(\Lambda_k)$, $b_i = (\pi_{i1}\pi_{j1})/(\pi_{j1}\pi_{k1}) = \pi_{i1}/\pi_{k1}$ (j stands for the column), and $*$ symbolizes useless entries for d_k. Thus, factorizing everything that can be factorized yields, for every $3 \leqslant k \leqslant p$,

$$d_k = \frac{\pi_{k1}^2}{1-\lambda_k^2}\left(\prod_{j=2}^{k-1}\frac{(\lambda_j-\lambda_k)^2}{(1-\lambda_j\lambda_k)^2}\right)d_{k-1} \quad \text{with} \quad d_2 = \frac{\pi_{21}^2}{1-\lambda_2^2}.$$

The solution of the recurrence is given by

$$\forall\, 2 \leqslant k \leqslant p, \quad d_k = \left(\prod_{j=2}^{k}\frac{\pi_{j1}^2}{1-\lambda_j^2}\right)\left(\prod_{j=2}^{k}\prod_{i<j}\frac{(\lambda_i-\lambda_j)^2}{(1-\lambda_i\lambda_j)^2}\right). \tag{1.3.8}$$

If (H_1) holds, then all eigenvalues are distinct and since π_{j1} is non-zero for all j, d_k is non-zero for all k. Taking $k = p$, the proof is now complete. If A has only real eigenvalues, we can even show that, under (H_1), $d_k > 0$ for all k which implies the positive definiteness of Λ using Sylvester's criterion, see *e.g.* Thm. 7.2.5 of Horn and Johnson (2012). □

Remark 1.3.2. Under (H_{22}) and for $p > 2$, we only have to consider the $(p-2)\times(p-2)$ bottom-right submatrix of Λ, which is obviously invertible (and positive definite when A has only real eigenvalues), based on the above.

1.3.2 Proof of Propositions 1.2.1 and 1.2.2

Let us start by noting that, under Lemma 1.3.1 and (H_3), for a fixed $n > n_0$,

$$\begin{aligned}\|\Gamma_n(0)\| &\leqslant C_{st}\sum_{\ell\geqslant 0}\|A_n^\ell K_p (A_n^\top)^\ell\| \\ &\leqslant C_{st}\sum_{\ell\geqslant 0}\|D_n^\ell\|^2 \leqslant \frac{C_{st}}{1-\rho_n^2} < +\infty\end{aligned}$$

where $\Gamma_n(\cdot)$ is the autocovariance function given in (1.2.4). Thus, by the same reasoning,

$$\begin{aligned}\left\|\sum_{h\in\mathbb{Z}}\Gamma_n(h)\right\| \leqslant \sum_{h\in\mathbb{Z}}\|\Gamma_n(h)\| &\leqslant C_{st}\|\Gamma_n(0)\|\sum_{h\in\mathbb{Z}}\|A_n^{|h|}\| \\ &\leqslant \frac{C_{st}}{1-\rho_n^2}\sum_{h\in\mathbb{Z}}\|D_n^{|h|}\| \leqslant \frac{C_{st}}{(1-\rho_n)^2} < +\infty.\end{aligned}$$

Moreover, as we will see later in (1.3.22), there exists a positive semi-definite matrix Γ such that $(1-\rho_n)\Gamma_n(0) \to \Gamma$. Hence, since $\Gamma_n(-h) = \Gamma_n^\top(h)$,

$$(1-\rho_n)^2\sum_{h\in\mathbb{Z}}\Gamma_n(h) \longrightarrow A^*\Gamma + (A^*\Gamma)^\top \triangleq \Gamma^+ \tag{1.3.9}$$

under (H_{21}) with $\lambda_1 = 1$ or (H_{22}), where $A^* = P K_p P^{-1}$ and K_p is given in (1.2.3). The top-left element of Γ^+ is $\sigma^2\pi_{11}^2 \neq 0$ (take *e.g.* Lemma 1.3.3 with $k = 1$). Under (H_{21}) with $\lambda_1 = -1$, by the same calculations, it turns out that

$$\sum_{h\in\mathbb{Z}}\Gamma_n(h) \longrightarrow \Gamma^- \tag{1.3.10}$$

where Γ^- is also a non-zero matrix. The result is established by combining (1.3.9) and (1.3.10). □

As for the proof of Proposition 1.2.2, it will result from a chain of intermediate lemmas. See footnote 2 on page 62 for notation.

Lemma 1.3.7 (Variance decomposition)**.** *Assume that (H_3) holds. Then, for all $n \geqslant 1$,*

$$\begin{aligned}\mathrm{vec}\left(\frac{S_{n,n-1}}{n}\right) &= B_n^{-1}\,\mathrm{vec}\left(\frac{T_n}{n}\right) + B_n^{-1}(I_p\otimes A_n)\,\mathrm{vec}\left(\frac{Z_n}{n}\right) \\ &\quad + B_n^{-1}(A_n\otimes I_p)\,\mathrm{vec}\left(\frac{Z_n^\top}{n}\right) + B_n^{-1}\,\mathrm{vec}\left(\frac{L_n}{n}\right)\end{aligned} \tag{1.3.11}$$

where $S_{n,n-1}$ *is given in* (1.1.4), B_n *is given in* (1.3.1),

$$Z_n = \sum_{k=1}^{n} \Phi_{n,k-1}\, E_k^\top, \qquad L_n = \sum_{k=1}^{n} E_k\, E_k^\top \tag{1.3.12}$$

and $T_n = \Phi_{n,0}\, \Phi_{n,0}^\top - \Phi_{n,n}\, \Phi_{n,n}^\top$.

Proof. By direct calculation, we first obtain that, for all $n \geqslant 1$,

$$\Phi_{n,k}\, \Phi_{n,k}^\top = A_n\, \Phi_{n,k-1}\, \Phi_{n,k-1}^\top\, A_n^\top + A_n\, \Phi_{n,k-1}\, E_k^\top + E_k\, \Phi_{n,k-1}^\top\, A_n^\top + E_k\, E_k^\top.$$

Then, summing over k, it is not hard to see that

$$S_{n,n-1} = T_n + A_n\, S_{n,n-1}\, A_n^\top + A_n\, Z_n + Z_n^\top A_n^\top + L_n.$$

This is a generalized Sylvester equation w.r.t. $S_{n,n-1}$ having a unique solution since $\rho_n < 1$, see *e.g.* Lem. 2.1(2) of Jiang and Wei (2003). We easily deduce that

$$B_n\, \mathrm{vec}(S_{n,n-1}) = \mathrm{vec}(T_n) + (I_p \otimes A_n)\, \mathrm{vec}(Z_n) + (A_n \otimes I_p)\, \mathrm{vec}(Z_n^\top) + \mathrm{vec}(L_n)$$

which, *via* Lemma 1.3.4, gives the result. □

Lemma 1.3.8 (Isolated terms)**.** *Assume that* (H_1) *and* (H_3) *hold. Then,*

$$\frac{T_n}{n} \xrightarrow{\ p\ } 0$$

where T_n *is given in* (1.3.12).

Proof. First, since $\Phi_{n,0}$ has a finite moment of order 2, Markov's inequality directly gives

$$\frac{\|\Phi_{n,0}\, \Phi_{n,0}^\top\|}{n} \xrightarrow{\ p\ } 0. \tag{1.3.13}$$

Then, by the triangle inequality and exploiting Lemma 1.3.1, for all $n > n_0$,

$$\begin{aligned} \|\Phi_{n,n}\, \Phi_{n,n}^\top\| &\leqslant 2\, \|A_n^n\|^2\, \|\Phi_{n,0}\|^2 + 2 \left\| \sum_{k=0}^{n-1} A_n^k\, E_{n-k} \right\|^2 \\ &\leqslant C_{st}\, \rho_n^{2n}\, \|\Phi_{n,0}\|^2 + 2 \left\| \sum_{k=0}^{n-1} A_n^k\, E_{n-k} \right\|^2 \triangleq T_{1,n} + T_{2,n}. \end{aligned}$$

Note that, according to (H_3),

$$\begin{aligned}\rho_n^{2n} &= \left(1-\frac{c}{v_n}\right)^{2n} \\ &= \exp\left(\frac{-2c\,n}{v_n}+o\left(\frac{n}{v_n}\right)\right) \longrightarrow 0.\end{aligned}$$

We can conclude that $\mathbb{E}[|T_{1,n}|]\to 0$, and thus

$$\frac{T_{1,n}}{n} \xrightarrow{p} 0. \tag{1.3.14}$$

Moreover,

$$\begin{aligned}\left\|\sum_{k=0}^{n-1} A_n^k\,E_{n-k}\right\|^2 &= \sum_{k=0}^{n-1}\sum_{\ell=0}^{n-1}\langle\!\langle P_n\,D_n^k\,P_n^{-1}\,E_{n-k},\,P_n\,D_n^\ell\,P_n^{-1}\,E_{n-\ell}\rangle\!\rangle \\ &= \sum_{k=0}^{n-1}\sum_{\ell=0}^{n-1}\mathrm{tr}\big(P_n\,D_n^\ell\,P_n^{-1}\,E_{n-\ell}\,E_{n-k}^\top\,(P_n^{-1})^\top\,D_n^k\,P_n^\top\big).\end{aligned}$$

Consequently,

$$\begin{aligned}\mathbb{E}[|T_{2,n}|] &= 2\,\sigma^2\sum_{k=0}^{n-1}\mathrm{tr}\big(P_n\,D_n^k\,P_n^{-1}\,K_p\,(P_n^{-1})^\top\,D_n^k\,P_n^\top\big) \\ &= C_{st}\sum_{k=0}^{n-1}\|P_n\,D_n^k\,P_n^{-1}\,K_p\|^2\end{aligned}$$

where K_p is given in (1.2.3), and obviously satisfies $K_p = K_p^\top$ and $K_p^2 = K_p$. Hence, using the same arguments as before,

$$\begin{aligned}\mathbb{E}[|T_{2,n}|] &\leqslant C_{st}\sum_{k=0}^{n-1}\rho_n^{2k} \\ &= C_{st}\,v_n\,\frac{1-\exp\left(\frac{-2c\,n}{v_n}+o\left(\frac{n}{v_n}\right)\right)}{2c-\frac{c^2}{v_n}} = O(v_n).\end{aligned} \tag{1.3.15}$$

Since $v_n = o(n)$, that leads to

$$\frac{T_{2,n}}{n} \xrightarrow{p} 0. \tag{1.3.16}$$

The combination of (1.3.13), (1.3.14) and (1.3.16) concludes the proof. □

Lemma 1.3.9 (Martingale terms)**.** *Assume that (H_1) and (H_3) hold. Then,*

$$\frac{Z_n}{n} \xrightarrow{p} 0$$

where Z_n is given in (1.3.12)*.*

Proof. By the triangle inequality,

$$\begin{aligned} \|Z_n\| &\leqslant \|\Phi_{n,0}\, E_1^\top\| + \left\|\sum_{k=2}^{n} A_n^{k-1}\, \Phi_{n,0}\, E_k^\top\right\| + \left\|\sum_{k=2}^{n}\sum_{\ell=0}^{k-2} A_n^\ell\, E_{k-1-\ell}\, E_k^\top\right\| \\ &\triangleq T_{0,n} + T_{1,n} + T_{2,n}. \end{aligned}$$

Clearly, by independence and existence of moments of order 2,

$$\frac{T_{0,n}}{n} \xrightarrow{p} 0. \tag{1.3.17}$$

Then, a direct calculation shows that

$$\begin{aligned} T_{1,n}^2 &= \sum_{k=2}^{n}\sum_{\ell=2}^{n} \langle\!\langle A_n^{k-1}\, \Phi_{n,0}\, E_k^\top,\, A_n^{\ell-1}\, \Phi_{n,0}\, E_\ell^\top \rangle\!\rangle \\ &= \sum_{k=2}^{n}\sum_{\ell=2}^{n} \mathrm{tr}\big(\Phi_{n,0}^\top\, (A_n^\top)^{\ell-1}\, A_n^{k-1}\, \Phi_{n,0}\, E_k^\top E_\ell\big) \end{aligned}$$

exploiting the cyclic property of the trace. Hence, for any $n > n_0$,

$$\begin{aligned} \mathbb{E}[T_{1,n}^2] &= \sum_{k=2}^{n}\sum_{\ell=2}^{n} \mathrm{tr}\big(\mathbb{E}\big[\Phi_{n,0}^\top\, (A_n^\top)^{\ell-1}\, A_n^{k-1}\, \Phi_{n,0}\big]\, \mathbb{E}[E_k^\top E_\ell]\big) \\ &= \sigma^2 \sum_{k=2}^{n} \mathbb{E}\big[\mathrm{tr}\big(\Phi_{n,0}^\top\, (A_n^{k-1})^\top\, A_n^{k-1}\, \Phi_{n,0}\big)\big] \\ &= \sigma^2 \sum_{k=2}^{n} \mathbb{E}\big[\|A_n^{k-1}\, \Phi_{n,0}\|^2\big] \\ &\leqslant C_{st} \sum_{k=2}^{n} \rho_n^{2(k-1)}\, \mathbb{E}[\|\Phi_{n,0}\|^2] \;=\; O(v_n) \end{aligned}$$

using (1.3.15), Lemma 1.3.1 and the second-order moments of $\Phi_{n,0}$. Since $v_n = o(n)$,

$$\frac{T_{1,n}^2}{n} \xrightarrow{p} 0 \quad \text{so that} \quad \frac{T_{1,n}}{n} \xrightarrow{p} 0. \tag{1.3.18}$$

To handle the last term of the decomposition, let, for any $k \geqslant 1$,

$$U_k = \sum_{j=1}^{k} A_n^{k-j} E_j$$

and note that

$$\forall\, k \neq \ell, \quad \mathbb{E}\big[U_{\ell-1}^\top U_{k-1} E_k^\top E_\ell\big] = 0. \tag{1.3.19}$$

Now,

$$\begin{aligned} T_{2,n}^2 &= \left\| \sum_{k=2}^{n} U_{k-1} E_k^\top \right\|^2 \\ &= \sum_{k=2}^{n}\sum_{\ell=2}^{n} \langle\!\langle U_{k-1} E_k^\top, U_{\ell-1} E_\ell^\top \rangle\!\rangle = \sum_{k=2}^{n}\sum_{\ell=2}^{n} \mathrm{tr}\big(U_{\ell-1}^\top U_{k-1} E_k^\top E_\ell\big) \end{aligned}$$

exploiting again the cyclic property of the trace. Thanks to relations (1.3.19), it follows that, for $n > n_0$,

$$\begin{aligned} \mathbb{E}[T_{2,n}^2] &= \sigma^2 \sum_{k=2}^{n} \mathrm{tr}\big(\mathbb{E}\big[U_{k-1}^\top U_{k-1}\big]\big) \\ &= \sigma^2 \sum_{k=2}^{n}\sum_{j=1}^{k-1}\sum_{i=1}^{k-1} \mathrm{tr}\big((A_n^\top)^{k-1-i} A_n^{k-1-j}\, \mathbb{E}[E_j E_i^\top]\big) \\ &= \sigma^4 \sum_{k=2}^{n}\sum_{j=1}^{k-1} \mathrm{tr}\big((A_n^\top)^{k-1-j} A_n^{k-1-j} K_p\big) = \sigma^4 \sum_{k=2}^{n}\sum_{\ell=0}^{k-2} \|A_n^\ell K_p\|^2 = O(nv_n) \end{aligned}$$

where K_p is given in (1.2.3), and where the rate v_n is obtained *via* the same lines as (1.3.15). Consequently,

$$\frac{T_{2,n}^2}{n^2} \xrightarrow{\ p\ } 0 \quad \text{so that} \quad \frac{T_{2,n}}{n} \xrightarrow{\ p\ } 0. \tag{1.3.20}$$

The combination of (1.3.17), (1.3.18) and (1.3.20) concludes the proof. □

Let us now return to the proof of Proposition 1.2.2. By the law of large numbers, the last term of (1.3.11) is such that

$$\frac{L_n}{n} \xrightarrow{\ p\ } \sigma^2 K_p \tag{1.3.21}$$

where K_p is given in (1.2.3). By combining Lemmas 1.3.5, 1.3.7, 1.3.8 and 1.3.9, con-

vergence (1.3.21) and hypotheses (H_1), (H_{21}) and (H_3), the first result is proved. It is important to note that $A^* = P\,K_p\,P^{-1}$ is a real matrix (see Lemma 1.3.2), so that the limit of the renormalized empirical covariance is obviously real itself. For the second result, we have

$$\begin{aligned}(1-\rho_n)\,\mathrm{vec}(\Gamma_n) &= \sigma^2\,(1-\rho_n)\left[\sum_{k=0}^{+\infty} A_n^k \otimes A_n^k\right]\mathrm{vec}(K_p)\\ &= \sigma^2\left((1-\rho_n)\,B_n^{-1}\right)e_{p^2}\\ &\longrightarrow \frac{\sigma^2}{2}\,(A^*\otimes A^*)e_{p^2} \;=\; \mathrm{vec}(\Gamma)\end{aligned} \tag{1.3.22}$$

from Lemma 1.3.5, where Γ is defined in Proposition 1.2.2 and noting that $\mathrm{vec}(K_p) = e_{p^2}$. Finally,

$$(1-\rho_n)\left\|\frac{S_n}{n}-\Gamma_n\right\| \leqslant \left\|(1-\rho_n)\,\frac{S_n}{n}-\Gamma\right\| + \left\|(1-\rho_n)\,\Gamma_n-\Gamma\right\| \xrightarrow{\;p\;} 0$$

exploiting the first part of the proposition and (1.3.22). The proof is now complete. □

Remark 1.3.3. If (H_{22}) is true instead of (H_{21}), the results of the proposition still hold provided that the limit is adjusted (see Remark 1.3.1).

1.3.3 Proof of Theorem 1.2.3

Note that for all $n > n_0$, under (H_1) and like in (1.3.3),

$$W_n\,B_n^{-1} = (V_n^{1/2}\otimes V_n^{1/2})\left[\sum_{k=0}^{+\infty} D_n^k\otimes D_n^k\right](P_n^{-1}\otimes P_n^{-1})$$

where W_n comes from (1.2.6). The matrix of rates is structured in p blocks as follows,

$$V_n^{1/2}\otimes V_n^{1/2} = \mathrm{diag}\big(\sqrt{1-\rho_n}\,V_n^{1/2},\,V_n^{1/2},\,\ldots,\,V_n^{1/2}\big).$$

Thus, taking over the reasoning of (1.3.4) and (1.3.5), we can show that

$$\lim_{n\to+\infty} W_n\,B_n^{-1} = \mathrm{diag}\left(\frac{K_p}{2},\,\Delta_2,\,\ldots,\,\Delta_p\right)(P^{-1}\otimes P^{-1}) \;\triangleq\; H^* \tag{1.3.23}$$

using (H_{21}) and (H_3), where K_p is given in (1.2.3) and

$$\forall\, 2\leqslant i\leqslant p,\qquad \Delta_i = \mathrm{diag}\left(0,\,\frac{1}{1-\lambda_i\,\lambda_2},\,\ldots,\,\frac{1}{1-\lambda_i\,\lambda_p}\right).$$

That leads, together with Lemmas 1.3.7, 1.3.8, 1.3.9 and formula (1.3.21), to the convergence

$$W_n \operatorname{vec}\left(\frac{S_{n,n-1}}{n}\right) \xrightarrow{p} \sigma^2 H^* e_{p^2}.$$

Since

$$W_n \operatorname{vec}\left(\frac{S_{n,n-1}}{n}\right) = \operatorname{vec}\left(V_n^{1/2} P_n^{-1} \frac{S_{n,n-1}}{n} (P_n^{-1})^\top V_n^{1/2}\right),$$

we obtain that

$$V_n^{1/2} P_n^{-1} \frac{S_{n,n-1}}{n} (P_n^{-1})^\top V_n^{1/2} \xrightarrow{p} \operatorname{vec}^{-1}(\sigma^2 H^* e_{p^2}) \tag{1.3.24}$$

where $\operatorname{vec}^{-1} : \mathbb{C}^{p^2} \to \mathbb{C}^{p\times p}$ is the vectorization inverse operator. A straightforward calculation shows that $\operatorname{vec}^{-1}(\sigma^2 H^* e_{p^2}) = H$ where H is given in (1.2.2), and this limit is invertible. Indeed, on the one hand,

$$\pi_{11} = \frac{(-\lambda_1)^{p-1}}{\prod_{j=2}^{p}(\lambda_j - \lambda_1)}$$

from formula (1.3.2) with $k = 1$, and we know from Lemma 1.3.2 that π_{11} is real and non-zero (positive). On the other hand, Lemma 1.3.6 implies that Λ is invertible so that $\det(H)$ is obviously non-zero. In addition, the estimation error satisfies

$$\begin{aligned}\widehat{\theta}_n - \theta_n &= (P_n^\top)^{-1} V_n^{1/2} \left(V_n^{1/2} P_n^{-1} \frac{S_{n,n-1}}{n} (P_n^{-1})^\top V_n^{1/2}\right)^{-1} \\ &\quad \times \left(V_n^{1/2} P_n^{-1} \frac{Z_n}{n} e_p\right)\end{aligned} \tag{1.3.25}$$

where in this expression, Z_n comes from (1.3.12). By the reasoning above and Lemmas 1.3.1 and 1.3.9,

$$\left\|\widehat{\theta}_n - \theta_n\right\| \leqslant C_{st} \left\|\left(V_n^{1/2} P_n^{-1} \frac{S_{n,n-1}}{n} (P_n^{-1})^\top V_n^{1/2}\right)^{-1}\right\| \left\|\frac{Z_n}{n}\right\| \xrightarrow{p} 0.$$

That establishes the consistency of the OLS estimator. □

1.3.4 Proof of Theorem 1.2.4

Suppose first that the eigenvalues of A are real, and consider the filtration

$$\forall\, n \geqslant 1,\ \forall\, 1 \leqslant k \leqslant n, \quad \mathcal{F}_{n,k} = \sigma(\Phi_{n,0},\, \varepsilon_1,\, \ldots,\, \varepsilon_k).$$

For all $a \in \mathbb{R}^p \backslash \{0\}$, let also

$$m_{n,k}^{(a)} = a^\top V_n^{1/2} P_n^{-1} \Phi_{n,k-1}\, \varepsilon_k \quad \text{and} \quad M_n^{(a)} = \sum_{k=1}^{n} m_{n,k}^{(a)} \tag{1.3.26}$$

the way it appears in the right-hand side of (1.3.25). The sequence $(m_{n,k}^{(a)})$ is clearly a scalar martingale difference array w.r.t. $\mathcal{F}_{n,k}$ at fixed n and for $1 \leqslant k \leqslant n$. The predictable quadratic variation of $M_n^{(a)}$ is

$$\begin{aligned} \langle M^{(a)} \rangle_n &= \sum_{k=1}^{n} \mathbb{E}[(m_{n,k}^{(a)})^2 \,|\, \mathcal{F}_{n,k-1}] \\ &= \sigma^2\, a^\top V_n^{1/2} P_n^{-1} \left[\sum_{k=1}^{n} \Phi_{n,k-1}\, \Phi_{n,k-1}^\top \right] (P_n^{-1})^\top V_n^{1/2}\, a \end{aligned}$$

since (ε_k) is a white noise and $\Phi_{n,k-1}$ is obviously $\mathcal{F}_{n,k-1}$-measurable. Together with (1.3.24) and the definition of $S_{n,n-1}$ in (1.1.4), that leads to the convergence

$$\frac{\langle M^{(a)} \rangle_n}{n} \xrightarrow{\ p\ } \sigma^2\, a^\top H a > 0 \tag{1.3.27}$$

where H is the covariance matrix given in (1.2.2), positive definite as shown in the proof of Lemma 1.3.6 when A has only real eigenvalues. To apply the central limit theorem for arrays of martingales, see *e.g.* Thm. 1 of Sec. 8 in Pollard (1984), it remains to show that the Lindeberg's condition is satisfied, in other words that

$$\forall\, \epsilon > 0, \quad \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\Big[(m_{n,k}^{(a)})^2\, \mathbb{1}_{\{|m_{n,k}^{(a)}| > \epsilon\sqrt{n}\}} \,\Big|\, \mathcal{F}_{n,k-1} \Big] \xrightarrow{\ p\ } 0. \tag{1.3.28}$$

To prove (1.3.28), we can first see that, for any $1 \leqslant k \leqslant n$,

$$\begin{aligned} \mathbb{E}\Big[(m_{n,k}^{(a)})^2\, \mathbb{1}_{\{|m_{n,k}^{(a)}| > \epsilon\sqrt{n}\}} \,\Big|\, \mathcal{F}_{n,k-1} \Big] &= T_{n,k-1}^{(a)}\, \xi_{n,k} \\ &\leqslant T_{n,k-1}^{(a)} \sup_{1 \leqslant k \leqslant n} \xi_{n,k} \end{aligned}$$

where $T_{n,k-1}^{(a)} \triangleq a^\top V_n^{1/2} P_n^{-1} \Phi_{n,k-1} \Phi_{n,k-1}^\top (P_n^{-1})^\top V_n^{1/2} a > 0$ and with, by Hölder's and Markov's inequalities,

$$\begin{aligned}\xi_{n,k} &\triangleq \mathbb{E}\big[\varepsilon_k^2 \mathbb{1}_{\{|m_{n,k}^{(a)}|>\epsilon\sqrt{n}\}} \,|\, \mathcal{F}_{n,k-1}\big] \\ &\leqslant \eta_\nu^{2/(2+\nu)} \mathbb{P}\big((m_{n,k}^{(a)})^2 > \epsilon^2 n \,|\, \mathcal{F}_{n,k-1}\big)^{\nu/(2+\nu)} \leqslant C_{st}\left(\frac{T_{n,k-1}^{(a)}}{n}\right)^{\nu/(2+\nu)}\end{aligned}$$

in which η_ν is the moment of order $2+\nu$ of (ε_k), as defined in the statement of the theorem. But choosing $k = n$, (1.3.27) directly entails

$$\xi_{n,n} \xrightarrow{p} 0 \quad \text{so that} \quad \xi_{n,n}^\sharp \triangleq \sup_{1\leqslant k\leqslant n} \xi_{n,k} \xrightarrow{p} 0 \tag{1.3.29}$$

according to Lem. 1.3.20 of Duflo (1997) adapted to convergence in probability. Finally,

$$\frac{1}{n}\sum_{k=1}^n \mathbb{E}\big[(m_{n,k}^{(a)})^2 \mathbb{1}_{\{|m_{n,k}^{(a)}|>\epsilon\sqrt{n}\}} \,|\, \mathcal{F}_{n,k-1}\big] \leqslant \left(\frac{1}{n}\sum_{k=1}^n T_{n,k-1}^{(a)}\right)\xi_{n,n}^\sharp$$

and, *via* (1.3.24) and (1.3.29), Lindeberg's condition (1.3.28) is satisfied. Together with (1.3.27), this is sufficient to establish that

$$\forall\, a \in \mathbb{R}^p\backslash\{0\}, \quad \frac{M_n^{(a)}}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \sigma^2 a^\top H a).$$

Because a is arbitrary, we can now make use of the Cramér-Wold device to get the p-vectorial convergence

$$\frac{M_n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}_p(0, \sigma^2 H) \tag{1.3.30}$$

with

$$M_n = V_n^{1/2} P_n^{-1} \sum_{k=1}^n \Phi_{n,k-1}\, \varepsilon_k.$$

Coming back to (1.3.24) and (1.3.25), and applying Slutsky's lemma, the proof is complete for the real case. If the spectrum of A contains complex eigenvalues, we shall exploit the fact that the first row of P_n^{-1} (say, $\ell_n^\top$) still contains real entries, as it is established in Lemma 1.3.2. The scalar array of $\mathcal{F}_{n,k}$-martingale to use is now

$$\forall\, n \geqslant 1, \quad M_n = \sum_{k=1}^n m_{n,k}$$

with

$$\forall\, 1 \leqslant k \leqslant n, \quad m_{n,k} = \sqrt{1-\rho_n}\,\ell_n^\top\,\Phi_{n,k-1}\,\varepsilon_k.$$

Similarly, the predictable quadratic variation of M_n is

$$\langle M\rangle_n = \sigma^2\,(1-\rho_n)\,\ell_n^\top\left[\sum_{k=1}^{n}\Phi_{n,k-1}\,\Phi_{n,k-1}^\top\right]\ell_n.$$

From (1.3.24),

$$\frac{\langle M\rangle_n}{n} \xrightarrow{\ p\ } \sigma^4\,\frac{\pi_{11}^2}{2} > 0 \quad \text{so that} \quad \frac{M_n}{\sqrt{n}} \xrightarrow{\ d\ } \mathcal{N}\left(0,\,\sigma^4\,\frac{\pi_{11}^2}{2}\right) \tag{1.3.31}$$

by the same reasoning, the real and positive limit being the top-left element of H (times σ^2). Since L_n as defined in the statement of the theorem is the first column of P_n, it follows that $L_n^\top(P_n^\top)^{-1} = e_p^\top$. Thus, by (1.3.25) and noting that $v_n\,(1-\rho_n) = c$ under (H_3),

$$\begin{aligned}
\sqrt{n\,v_n}\left\langle L_n,\,\widehat{\theta}_n - \theta_n\right\rangle &= \sqrt{n\,v_n\,(1-\rho_n)}\,e_p^\top\left(V_n^{1/2}\,P_n^{-1}\,\frac{S_{n,n-1}}{n}\,(P_n^{-1})^\top\,V_n^{1/2}\right)^{-1} \\
&\qquad \times\left(V_n^{1/2}\,P_n^{-1}\frac{Z_n}{n}\right)e_p \\
&= \sqrt{c}\,e_p^\top\left[\left(V_n^{1/2}\,P_n^{-1}\,\frac{S_{n,n-1}}{n}\,(P_n^{-1})^\top\,V_n^{1/2}\right)^{-1} - H^{-1}\right] \\
&\qquad \times\left(V_n^{1/2}\,P_n^{-1}\frac{Z_n}{\sqrt{n}}\right)e_p \\
&\qquad + \frac{2\sqrt{c}\,M_n}{\sigma^2\,\pi_{11}^2\,\sqrt{n}} \triangleq T_{1,n} + T_{2,n} \xrightarrow{\ d\ } \mathcal{N}\left(0,\,\frac{2c}{\pi_{11}^2}\right)
\end{aligned}$$

exploiting (1.3.31), the block diagonal structure of H which directly gives $e_p^\top H^{-1} = 2\,e_p^\top/(\sigma^2\,\pi_{11}^2)$, Slutsky's lemma and the fact that $\|T_{1,n}\| = o_p(\|T_{2,n}\|)$. The proof is now complete. □

Remark 1.3.4. If we replace (H_{21}) by (H_{22}) in Theorems 1.2.3 and 1.2.4, the results still hold with slight adjustments. In particular, we have now to consider

$$V_n = \mathrm{diag}(1-\lambda_{n,1},\,1+\lambda_{n,2},\,1,\,\ldots,\,1)$$

as rates, and this implies different limits in (1.3.23) and (1.3.24). The main consequence

is that the top-left 1×1 block of H_0 is changed into a 2×2 block given by

$$\begin{pmatrix} \frac{\pi_{11}^2}{2} & 0 \\ 0 & \frac{\pi_{21}^2}{2} \end{pmatrix}.$$

This new submatrix is obviously real and positive definite (see Lemma 1.3.3 with $k = 2$), and all the subsequent reasoning is perfectly similar.

1.4 Applications and simulations

To illustrate the asymptotic normality of Theorem 1.2.4, we will use the marginal convergence stated in Corollary 1.2.6, *i.e.*

$$Z_n^2 \triangleq \frac{\pi_{11}^2 \, n \, v_n}{2\,c} \left[\sum_{i=1}^{p} \frac{1}{(\lambda_1 \rho_n)^{\,i-1}} \left(\widehat{\theta}_{n,i} - \theta_{n,i} \right) \right]^2 \xrightarrow{\;d\;} \chi_1^2 \tag{1.4.1}$$

including positive and negative unit root situations ($\lambda_1 = \pm 1$). According to (H_3), we set $\rho_n = 1 - c/v_n$ with $c = 1$ and $v_n = n^\alpha$ for some $0 < \alpha < 1$. For each simulation, $\lambda_{n,2}, \ldots, \lambda_{n,p}$ are randomly chosen out of $[-|\lambda_{n,1}| + \epsilon, |\lambda_{n,1}| - \epsilon]$ for $\epsilon = 0.1$, then the process is generated with $n = 5000$ and Z_n^2 is computed. The experiment is repeated 3000 times. On Figure 1.1, the simulations are conducted with $p \in \{2, 3, 4\}$ for $\alpha = 1/2$, $\lambda_1 \in \{-1, 1\}$ and $\sigma^2 = 1$. The empirical distributions of Z_n^2 are superimposed with the 'true' χ_1^2 density (in red), and the frequencies of observed values greater than 6.5 are indicated. On Figure 1.2, the simulations are conducted with $p = 3$, $\lambda_1 \in \{-1, 1\}$, $\alpha \in \{1/5, 1/4, 1/3, 1/2\}$ and $\sigma^2 = 1$. The empirical cumulative distributions of Z_n^2 are superimposed with the 'true' χ_1^2 cumulative distribution (in red). Figure 1.3 displays the same experiments with $\alpha \in \{1/2, 2/3, 3/4, 4/5\}$. Now let us comment our observations. One can see that there are a few more outliers than expected: $\approx 2.5\%$ of observed values greater than 6.5 whereas $\mathbb{P}(\chi_1^2 > 6.5) \approx 1.1\%$. This is due to the fact that, as we have noticed, the convergence may be slow (see the perspectives in the conclusion) and the Gaussian distribution tails remain somewhat overloaded. It is particularly clear on Figure 1.2 when α is close to 0 and, unsurprisingly, this phenomenon is widely attenuated when we choose larger values of n. The best results are obtained for $\alpha = 1/2$ and we observe a drift which occurs faster when α tends to 0 (Figure 1.2) than when α tends to 1 (Figure 1.3). This is probably a consequence of the well-known fact that the rates of convergence are faster in an unstable model that in a stable one, as

we recalled in the introduction. On the whole, we may roughly say that for the moderate values of α (between 1/3 and 2/3, say), the simulations are convincing and clearly fit with the theoretical behavior.

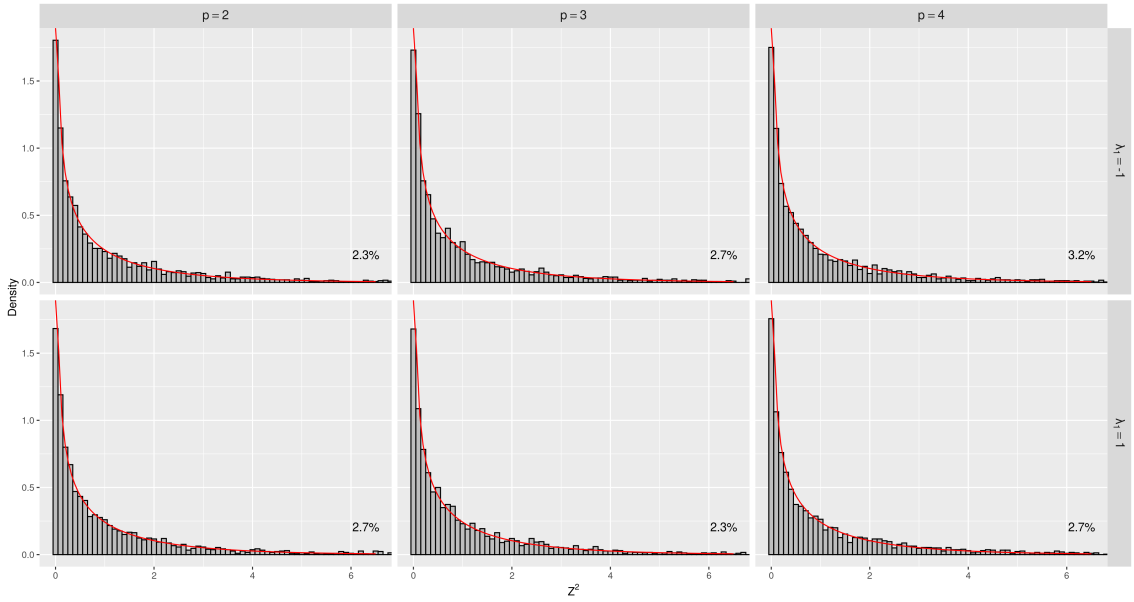


Figure 1.1 – Empirical distributions of Z_n^2 for $p = 2$ (left), $p = 3$ (middle) and $p = 4$ (right), with $\lambda_1 = -1$ (top) and $\lambda_1 = 1$ (bottom), on the basis of 3000 experiments of size $n = 5000$ with $\alpha = 1/2$. The red curve is the χ_1^2 distribution. The frequencies of observed values greater than 6.5 are indicated, to be compared with $\mathbb{P}(\chi_1^2 > 6.5) \approx 1.1\%$.

To sum up, aligned with the work of Phillips and Magdalinos (2007), this study completes the moderate deviations of Proïa (2020) by providing a sharp analysis of the asymptotic behavior of the OLS between the rates $\sqrt{n}$ and n corresponding to the stable and unstable AR(p) processes, respectively, when the unit root in A is either positive or negative (and even when both unit roots are present). The same conclusion prevails concerning the lack of continuity at the boundaries: $\alpha \to 0^+$ and $\alpha = 0$ do not match and the same is true for $\alpha \to 1^-$ and $\alpha = 1$. However, by focusing on the inner neighborhood of the unit root, nearly-unstable time-varying AR processes are instructive, especially considering the extension of the asymptotic normality when ρ_n does not converge 'too fast' to 1. There are still many improvements to be made, the main one being whether a unit root test can

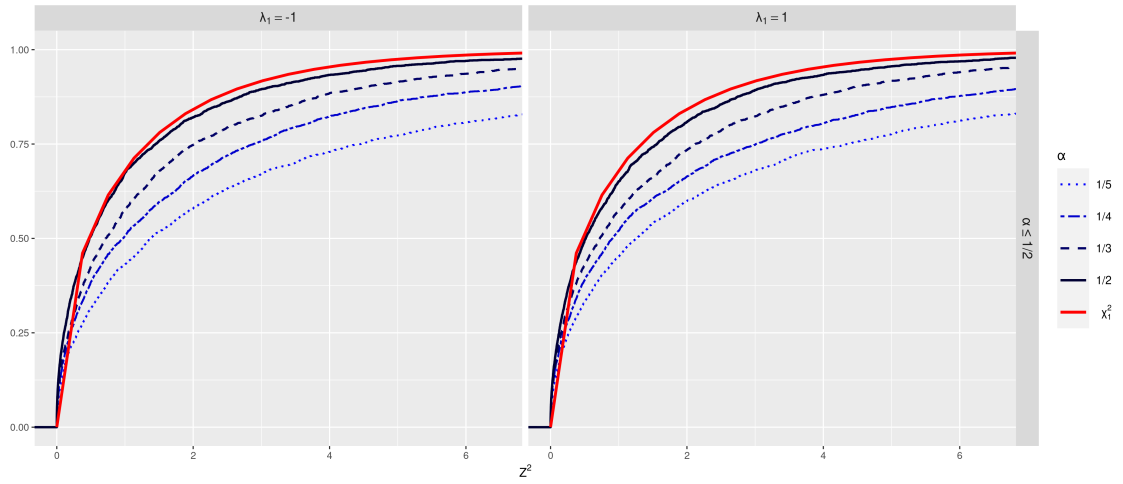


Figure 1.2 – Empirical cumulative distributions of Z_n^2 for $p = 3$ with $\lambda_1 = -1$ (left), $\lambda_1 = 1$ (right) and $\alpha \in \{1/5, 1/4, 1/3, 1/2\}$, on the basis of 3000 experiments of size $n = 5000$. The red curve is the χ_1^2 cumulative distribution.

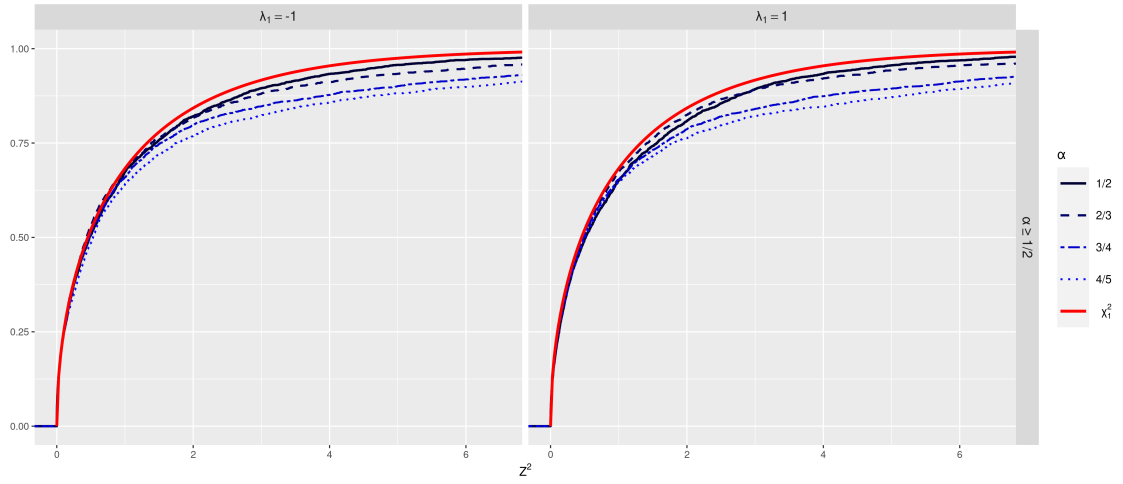


Figure 1.3 – Empirical cumulative distributions of Z_n^2 for $p = 3$ with $\lambda_1 = -1$ (left), $\lambda_1 = 1$ (right), $\alpha \in \{1/2, 2/3, 3/4, 4/5\}$, on the basis of 3000 experiments of size $n = 5000$. The red curve is the χ_1^2 cumulative distribution.

be derived. For the processes which admit a decomposition of the form

$$\forall\, n \geqslant 1,\ \forall\, 1 \leqslant k \leqslant n, \quad \left\{ \begin{array}{rcl} X_{n,k} & = & \beta_n\, X_{n,k-1} + W_k \\ (W_k) & \sim & \text{stable AR}(p-1), \end{array} \right.$$

an identification with (1.1.2) is possible and shows that $(X_{n,k})$ is in fact a time-varying AR(p) process such that if β_n converges to β, then all its coefficients converge *at the same rate*. In this subclass of processes, Park (2003) suggests a theoretical procedure to test for $\mathcal{H}_0$: "$\beta = 1$" *vs* $\mathcal{H}_1$: "$\beta < 1$" provided that $|\beta_n - \beta| \asymp 1/v_n$. The authors are pretty convinced that this could be extended to all the nearly unstable processes covered by (1.1.2) and that such a trail is likely to outperform the usual Dickey-Fuller tests for unit root. This is a work in progress. In addition, the simulations highlight an asymmetry in the convergence rates: we observed, on many more experiments than those presented here, that the asymptotic distribution deteriorates less quickly when α moves away from 1/2 towards 1 than towards 0 (see again Figures 1.2 and 1.3). It should therefore be interesting to investigate more deeply the convergence rate of the OLS. Finally, it would be instructive but very challenging to look at the behavior of the OLS on the *outer* neighborhood of the unit root, that is, for explosive but nearly unstable AR(p) processes, as it is already done for $p = 1$.

Acknowledgements. This research benefited from the support of the ANR project 'Efficient inference for large and high-frequency data' (ANR-21-CE40-0021). The authors sincerely thank the anonymous reviewer and the associate editor for their comments and references which have clearly contributed to the improvement of the paper.

References

Beran, J., Feng, Y., Ghosh, S., and Kulik, R. (2013), *Long-Memory Processes*, Probabilistic Properties and Statistical Methods, Springer, Heidelberg.

Brockwell, P. J. and Davis, R. A. (1991), *Time Series: Theory and Methods*, 2nd ed., Springer-Verlag: New York.

Buchmann, B. and Chan, N. H. (2013), « Unified asymptotic theory for nearly unstable AR(p) processes », *in*: *Stochastic Processes and their Applications* 123, pp. 952–985.

Chan, N. H. and Wei, C. Z. (1987), « Asymptotic inference for nearly nonstationary AR(1) processes », *in*: *Annals of Statistics* 15, pp. 1050–1063.

Chan, N. H. and Wei, C. Z. (1988), « Limiting distributions of least squares estimates of unstable autoregressive processes », *in*: *Ann. Stat.* 16, pp. 367–401.

Duflo, M. (1997), *Random Iterative Models*, Applications of Mathematics, Vol. 34, Springer-Verlag: Berlin.

Giraitis, L. and Phillips, P. C. B. (2006), « Uniform limit theory for stationary autoregression », *in*: *Journal of Time Series Analysis* 27, pp. 51–60.

Horn, R. A. and Johnson, C. R. (2012), *Matrix Analysis (Second Edition)*, Cambridge University Press: Cambridge, New-York.

Horváth, L. and Trapani, L. (2016), « Statistical inference in a random coefficient panel model », *in*: *Journal of Econometrics* 193, pp. 54–75.

Jiang, T. and Wei, M. (2003), « On solutions of the matrix equations X - A X B = C and X - A $\bar{X}$ B = C », *in*: *Linear Algebra Appl.* 367, pp. 225–233.

Lai, T. L. and Wei, C. Z. (1983), « Asymptotic Properties of General Autoregressive Models and Strong Consistency of Least-Squares Estimates of Their Parameters », *in*: *Journal of Multivariate Analysis* 13, pp. 1–23.

Ling, S. (2004), « Estimation and testing stationarity for double-autoregressive models », *in*: *J. R. Stat. Soc. B.* 66, pp. 63–78.

Miao, Y., Wang, Y., and Yang, G. (2015), « Moderate deviation principles for empirical covariance in the neighbourhood of the unit root », *in*: *Scandinavian Journal of Statistics* 42, pp. 234–255.

Nielsen, H. B. and Rahbek, A. (2014), « Unit root vector autoregression with volatility induced stationarity », *in*: *J. Empir. Financ.* 29, pp. 144–167.

Park, J. W. (2003), *Weak unit roots*, Department of Economics, Rice University.

Phillips, P. C. B. (1987b), « Towards a unified asymptotic theory for autoregression », *in*: *Biometrika* 74, pp. 535–547.

Phillips, P. C. B. and Lee, J. H. (2015), « Limit theory for VARs with mixed roots near unity », *in*: *Econometric Review* 34, pp. 1034–1055.

Phillips, P. C. B. and Magdalinos, T. (2007), « Limit theory for moderate deviations from a unit root », *in*: *Journal of Econometrics* 136, pp. 115–130.

Pollard, D. (1984), *Convergence of Stochastic Processes*, Springer: Berlin.

Proïa, F. (2020), « Moderate deviations in a class of stable but nearly unstable processes », *in*: *Journal of Statistical Planning and Inference* 208, pp. 66–81.

Trapani, L. (2021), « Testing for strict stationarity in a Random Coefficient AutoRegression », *in*: *Econometric Review* 40, pp. 220–256.

Chapter 2

Testing for the extent of instability in nearly-unstable processes

This chapter is based on the following publication: Badreau and Proïa (2025)

Abstract: This paper deals with unit root issues in time series analysis. It has been known for a long time that unit root tests may be flawed when a series although stationary has a root close to unity. That motivated recent papers dedicated to autoregressive processes where the bridge between stability and instability is expressed by means of time-varying coefficients. The process we consider has a companion matrix A_n with spectral radius $\rho(A_n) < 1$ satisfying $\rho(A_n) \to 1$, a situation described as 'nearly-unstable'. The question we investigate is: given an observed path supposed to come from a nearly-unstable process, is it possible to test for the 'extent of instability', *i.e.* to test how close we are to the unit root? In this regard, we develop a strategy to evaluate α and to test for $\mathcal{H}_0 : \text{“}\alpha = \alpha_0\text{”}$ against $\mathcal{H}_1 : \text{“}\alpha > \alpha_0\text{”}$ when $\rho(A_n)$ lies in an inner $O(n^{-\alpha})$-neighborhood of the unity, for some $0 < \alpha < 1$. Empirical evidence is given about the advantages of the flexibility induced by such a procedure compared to the common unit root tests. We also build a symmetric procedure for the usually left out situation where the dominant root lies around -1.

Keywords and phrases: Nearly unstable autoregressive process; Unit root test; Martingales.

Mathematics Subject Classification: 62M10, 62F12, 62F03, 60G52, 60G42.

2.1 Introduction and Motivation

This paper is devoted to the well-known unit root issue, still at the heart of many concerns in finance, econometrics, and more generally in time series analysis, at the origin, when undetected, of spurious regressions through unreliable estimates. In order to investigate the stochastic non-stationarity of an observed autoregressive path, two strategies are frequently encountered. The first one consists in estimating the coefficient behind the (supposed) unit root, and the second one directly looks for a hidden random walk in the disturbance. In a very particular case, Dickey and Fuller (1979) first tackled the unit root issue by considering an AR(1) model with a linear trend and concluded that, when there is a unit root in the generating process, the least-squares estimator of the autoregressive parameter has an asymmetric distribution that can be written asymptotically as a functional of a family of detrended Wiener processes, actually depending on the trend. Such a result had been anticipated in the Gaussian case by White (1958) in the absence of trend and, among numerous other studies, Chan and Wei (1988) improved the assumptions on the noise by considering a sequence of martingale differences (with sufficient moments). The behavior of the associated t-statistic has been tabulated by Dickey and Fuller (1979) for testing purposes, see also MacKinnon (1991). The *ADF test* (Augmented Dickey-Fuller) is probably the most commonly used nowadays to evaluate the presence of a unit root in a causal ARMA(p, q) process. It was formalized by Said and Dickey (1981) when p and q are known, and by Said and Dickey (1984) under a truncated approximation of the AR(∞) expression, therefore working for any p and q. At the basis of their study is the following differentiated form of an AR(k) process with parameter θ,

$$\forall\, n \geqslant 1, \quad \Delta X_n = (\theta_0 - 1) X_{n-1} + \sum_{i=1}^{k-1} \delta_i \, \Delta X_{n-i} + \varepsilon_n \tag{2.1.1}$$

where (ε_n) is a white noise sequence with finite variance, $\delta_i = -(\theta_{i+1} + \cdots + \theta_k)$ for $1 \leqslant i \leqslant k-1$ and $\theta_0 = \theta_1 + \cdots + \theta_k$, so that (ΔX_n) is stationary under the null hypothesis of unit root $\mathcal{H}_0$: "$\theta_0 = 1$" (actually trend-stationary if a trend is added in the model). That was an inspiration for us, and we will find a similar reasoning in this paper. The asymptotic behavior of the estimate is still the one of the seminal work of Dickey and Fuller (1979) as long as $k = O(n^{1/3})$ but the power of the test is impacted and, as it is explained by Schwert (1989) or by Ng and Perron (1995), some distortions may occur for badly truncated processes. Let us also mention the non-parametric approach suggested

by Phillips (1987a) and deepened by Phillips and Perron (1988) in which all correlation phenomenon are postponed to the noise of the model, which is now supposed to be strongly mixing. Overall there is an abundant literature on unit roots tests, we have just summarized here some important topics. The interested reader will find in-depth studies for example in Bhargava (1986), Perron (1988), Schmidt and Phillips (1992), etc. From a practical point of view, Nelson and Plosser (1982) had already highlighted the presence of unit roots in a number of macroeconomic series *via* the Dickey-Fuller strategy. Although not directly related to this paper, we found it useful to mention the complementary approach consisting in testing for trend-stationarity instead of integration. The noise is now assumed to include a random walk whose variance is zero under the null hypothesis and among such procedures, the most famous seem to be the *KPSS test* of Kwiatkowski et al. (1992) and the quite similar *LMC test* of Leybourne and McCabe (1994). Here again there is an ample literature but too far from our scope to dwell on it, let us instead return to unit roots.

In the context of autoregressive processes, the substantial works of Lai and Wei (1983) and Chan and Wei (1988) helped to establish that the least-squares estimator is strongly consistent wherever the characteristic roots lie, but with very different convergence rates and limit distributions whether the process is stationary (stable), integrated (unstable), or even explosive. However, DeJong et al. (1992) observed that the unit root tests may be empirically less powerful than their stable counterparts when, although very close to it, the spectral radius remains less than 1, and some econometricians are not fully convinced by the compatibility between the theoretical impacts of this discontinuity – from $n^{1/2}$-normality to n–asymmetry while the spectral radius continuously moves towards the unit root – and real-word data. These are strong arguments in favor of attempts to bridge stability and instability that motivated numerous studies devoted to intermediate models, including ours. Most of them involve a first-order process with time-varying coefficients $(X_{n,k})$, generated by the triangular form

$$\forall\, n \geqslant 1,\ \forall\, 1 \leqslant k \leqslant n, \qquad X_{n,k} = \theta_n\, X_{n,k-1} + \varepsilon_k. \tag{2.1.2}$$

This is an easy-to-use relevant case to understand the underlying dynamics since to focus on the inner neighborhood of the unit root, we may only consider a sequence of coefficients satisfying $|\theta_n| < 1$ for all $n \geqslant 1$ but $|\theta_n| \to 1$, so that the model is always stable but increasingly close to instability. Chan and Wei (1987), Phillips (1987b), Giraitis

and Phillips (2006) and later Phillips and Magdalinos (2007) were interested in such a model and, among other conclusions, it was established that the least-squares estimator is asymptotically normal around θ_n with rate $(nv_n)^{1/2}$ when $\theta_n = 1 - c/v_n$, for any $1 \ll v_n \ll n$ and $c > 0$. The junction is not perfect – there is still a discontinuity in the asymptotic variance when $v_n = 1$ and in the asymptotic distribution when $v_n = n$ – but these are promising results. Indeed, by setting $v_n = n^\alpha$ for some $0 < \alpha < 1$, the gap between stability and instability is filled with rates of the form $n^{1/2} \ll n^{(1+\alpha)/2} \ll n$. The studies of Meer et al. (1999), Park (2003), Buchmann and Chan (2013), or the more recent ones of Proïa (2020) and Badreau and Proïa (2023) aimed at finding a kind of generalization to nearly-unstable AR(p) processes. In the latter, the process $(\Phi_{n,k})$ is assumed to follow the $\mathrm{VAR}_p(1)$ triangular form

$$\forall\, n \geqslant 1,\ \forall\, 1 \leqslant k \leqslant n, \qquad \Phi_{n,k} = A_n\, \Phi_{n,k-1} + E_k \tag{2.1.3}$$

where (E_k) is a p-vectorial noise and

$$A_n = \begin{pmatrix} \theta_{n,1} & \theta_{n,2} & \cdots & \theta_{n,p} \\ \ddots & & & 0 \\ & I_{p-1} & & \vdots \\ & & \ddots & 0 \end{pmatrix} \tag{2.1.4}$$

is the associated $p \times p$ companion matrix with spectral radius satisfying $\rho(A_n) < 1$ for all $n \geqslant 1$ and $\rho(A_n) \to 1$. Provided suitable assumptions of moments, the least-squares estimator is still shown to be consistent and asymptotically normal around θ_n with mixed rates (see Thm. 3 and 4), and it is deduced as a corollary that, when $\rho(A_n) = 1 - c/v_n$,

$$\sqrt{n\, v_n}\,\big\langle L_n,\, \widehat{\theta}_n - \theta_n \big\rangle \xrightarrow{\ d\ } \mathcal{N}\big(0,\, h_0^2\big) \quad \text{with} \quad L_n^\top = \big(\lambda_1^{i-1} \rho(A_n)^{1-i}\big)_{1 \leqslant i \leqslant p} \tag{2.1.5}$$

where $\lambda_1 = \pm 1$ (depending on the sign of the unit root in the unstable limit process) and the asymptotic variance $h_0^2 > 0$ is given as a function of c and the eigenvalues of A, the limit matrix of A_n. However, although this result and the ones that may be found in the references above are very useful for understanding the behavior of the process and the estimator in the inner neighborhood of the unit circle, they definitely lack practical application. Beyond the scope of this study, the concept of near-instability has also been defined for first-order INAR processes (AR with integer values), see *e.g.* Ispány et al. (2003). In

that case, the mean value of the counting sequence is assumed to converge to one, that is to the unstable setting of an INAR(1) process, and some kind of bridges are built between the stable and unstable rates. An application to nearly-critical Galton-Watson processes can be found in the same reference or more deeply in Ispány et al. (2005). Let us also mention the study of Barreto-Souza and Chan (2024) related to nearly-unstable INARCH(1) processes, still dealing with counting time series but in a heteroscedastic context. The reader is referred to the large bibliography on the topic given in this very recent article.

The question we wish to investigate in this paper is the following: given an observed path supposed to come from a nearly-unstable process (*e.g.* doubtful ADF/KPSS diagnoses), is it possible to test for the 'extent of instability', *i.e.* to admit that the process is stable but to test how close we are to the unit root? Setting $\rho(A_n) = 1 - c/v_n$ with $v_n = n^\alpha$, the question is clearly related to the choice of α and may reduce to a test of $\mathcal{H}_0 : "\alpha = \alpha_0"$ against $\mathcal{H}_1 : "\alpha > \alpha_0"$ for some test value α_0. In addition, evaluating α is likely to provide adjustments in the confidence intervals associated with the autoregressive parameters. This is an innovative approach of the unit root issue (to the best of our knowledge), especially from a practical point of view. In Section 2.2, we give some specific features of the model and the hypotheses before developing the test procedure just mentioned. We will see that, as a corollary of our reasonings, we obtain a version of the asymptotic normality for a corrected statistic closely related to the least-squares estimation that does not need a mixing of the rates matrix, actually deepening Badreau and Proïa (2023):Thm. 4. An empirical study is provided in Section 2.3, first on simulated data and then on real time series – the famous dataset coming from Nelson and Plosser (1982) – in order to assess the performances of our procedure and justify its utility. After a short conclusion in Section 2.4, we finally prove our results in Section 2.5.

2.2 A (Quasi-)Unit Root Test

Let us consider the triangular model (2.1.3), where

$$\forall\, n \geqslant 1,\ \forall\, 1 \leqslant k \leqslant n, \quad \Phi_{n,k}^\top = \begin{pmatrix} X_{n,k} & \cdots & X_{n,k-p+1} \end{pmatrix} \quad \text{and} \quad E_k^\top = \begin{pmatrix} \varepsilon_k & 0 & \cdots & 0 \end{pmatrix}.$$

The process (ε_k) is assumed to be a white noise with finite variance $\sigma^2 > 0$, and the arbitrary initial value $\Phi_{n,0}$ is square-integrable and independent of the noise. The companion matrix A_n of the process is given in (2.1.4) and has spectrum $\{\lambda_{n,i}\}_{1 \leqslant i \leqslant p}$ where $\lambda_{n,1}$

is the one with largest modulus (the complex eigenvalues are sorted according to their modulus, in descending order, with ties broken by lexicographic order, also in descending order). The following technical hypotheses are intended to reflect our understanding of 'near instability'.

(A1) *Convergence of the companion matrix.* There exists a $p \times p$ matrix A such that

$$\lim_{n \to +\infty} A_n = A$$

with distinct eigenvalues $0 < |\lambda_p| \leqslant \ldots \leqslant |\lambda_2| \leqslant |\lambda_1| = \rho(A) = 1$, and the top-right element of A is non-zero ($\theta_p \neq 0$).

(A2) *Number of unit roots.* There is exactly one unit root in A ($\lambda_1 = \pm 1$ but $|\lambda_2| < 1$).

(A3) *Spectral radius of the companion matrix.* The spectral radius of A_n is given by

$$\forall\, n \geqslant 1, \quad \rho(A_n) = 1 - \frac{c}{n^\alpha}$$

for some $c > 0$ and $0 < \alpha < 1$.

For any $d \geqslant 1$, the usual $d \times d$ matrices

$$I_d = \begin{pmatrix} 1 & 0 & \cdots \\ 0 & \ddots & \\ \vdots & & 1 \end{pmatrix} \quad \text{and} \quad K_d = \begin{pmatrix} 1 & 0 & \cdots \\ 0 & 0 & \\ \vdots & & \ddots \end{pmatrix} \tag{2.2.1}$$

will be frequently encountered and the first vector of the canonical basis of $\mathbb{R}^d$ will be denoted by $e_d^\top = (1, 0, \ldots, 0)$. For readability purposes, we prove the results as if the limit process had a positive unit root (that is, $\lambda_1 = 1$), so that $\lambda_{n,1}$ is equal (to be rigorous, this is true only for n sufficiently large) to the spectral radius $\rho_n = \rho(A_n)$ of the companion matrix. Nevertheless they remain valid for $\lambda_1 = -1$, at the cost of some adjustments that will be clarified alongside (see Section 2.5.6). Technically, we should assume in addition that $0 < c < n^\alpha$ to ensure that for all n, $0 < \rho_n < 1$. Since our results are asymptotic w.r.t. n, we will allow ourselves a slight abuse of notation and simply set $c > 0$. The parameters c and α have a very different role to play: c is a *variance parameter* and is clearly less important than α, which is a *rate parameter*. In fact, the over-parametrization issued from the ratio c/n^α makes it impossible to dissociate c from α without additional constraints, which is the reason why we first assume in Section 2.2.1 that c is known to focus on α. However, a procedure will also be detailed in Section 2.2.2 for a joint evaluation of c and

α, according to a certain empirical protocol. First, like Park (2003), it is convenient to express the process under the hierarchical form

$$\forall\, n \geqslant 1,\ \forall\, 1 \leqslant k \leqslant n, \quad \left\{ \begin{array}{rcl} X_{n,k} & = & \lambda_{n,1}\, X_{n,k-1} + V_{n,k} \\ V_{n,k} & = & \beta_{n,1}\, V_{n,k-1} + \ldots + \beta_{n,p-1}\, V_{n,k-p+1} + \varepsilon_k \end{array} \right. \tag{2.2.2}$$

where we merely set $V_{n,k} = \varepsilon_k$ if $p = 1$. As it can be easily seen by factoring the autoregressive polynomial of $(X_{n,k})$ by $(1 - \lambda_{n,1} z)$, the companion matrix $\bar{A}_n$ associated with $(V_{n,k})$ has spectrum $\{\lambda_{n,i}\}_{2 \leqslant i \leqslant p}$. Under (A1) and (A2), it is therefore a stable AR$(p-1)$ process with time-varying coefficients $\beta_{n,i} \to \beta_i$, such that $\bar{A}_n \to \bar{A}$ and $\rho(\bar{A}_n) \to \rho(\bar{A}) = |\lambda_2| < 1$. So the limit process is itself stable and its (normalized) covariance matrix is given by

$$\Sigma_{p-1} = \sum_{k \geqslant 0} \bar{A}^k\, K_{p-1}\, (\bar{A}^\top)^k. \tag{2.2.3}$$

But that also leads to another representation of the process, closely related to (2.1.1) mentioned in the introduction,

$$\forall\, n \geqslant 1,\ \forall\, 1 \leqslant k \leqslant n, \quad X_{n,k} = \lambda_{n,1}\, X_{n,k-1} + \sum_{i=1}^{p-1} \beta_{n,i}\, (X_{n,k-i} - \lambda_{n,1}\, X_{n,k-i-1}) + \varepsilon_k \tag{2.2.4}$$

from which we will benefit in our following results. In consequence, it is straightforward to see that $\theta_n^\top = (\theta_{n,1},\, \ldots,\, \theta_{n,p})$ and $\beta_n^\top = (\beta_{n,1},\, \ldots,\, \beta_{n,p-1})$ are such that

$$\forall n \geqslant 1, \quad \theta_n = J_n\, \beta_n + \lambda_{n,1}\, e_p \quad \text{where} \quad J_n = \begin{pmatrix} 1 & & \\ -\lambda_{n,1} & \ddots & \\ & \ddots & 1 \\ & & -\lambda_{n,1} \end{pmatrix} \quad (= 0 \text{ if } p = 1) \tag{2.2.5}$$

is a rectangular matrix of size $p \times (p-1)$. To emphasize that the latent variables $V_{n,k}$ depend on α, from now on we will use $V_{n,k}(\alpha)$ instead and we will do the same for all confusing notation. Accordingly, let

$$\forall\, n \geqslant 1,\ \forall\, 1 \leqslant k \leqslant n, \quad \Psi_{n,k}^\top(\alpha) = \begin{pmatrix} X_{n,k} & V_{n,k}(\alpha) & \cdots & V_{n,k-p+2}(\alpha) \end{pmatrix}$$

where the arbitrary initial value $\Psi_{n,0}(\alpha)$ is square-integrable, so that from (2.2.4),

$$\forall\, n \geqslant 1,\ \forall\, 1 \leqslant k \leqslant n, \quad X_{n,k} = \vartheta_n^\top(\alpha)\, \Psi_{n,k-1}(\alpha) + \varepsilon_k \tag{2.2.6}$$

where $\vartheta_n^\top(\alpha) = (\lambda_{n,1}, \beta_{n,1}, \ldots, \beta_{n,p-1})$. Let also

$$S_{n,n} = \sum_{k=0}^{n} \Phi_{n,k}\, \Phi_{n,k}^\top \quad \text{and} \quad T_{n,n}(\alpha) = \sum_{k=0}^{n} \Psi_{n,k}(\alpha)\, \Psi_{n,k}^\top(\alpha) \tag{2.2.7}$$

together with the rate matrix of size $p \times p$, that shall help to control the convergence of $S_{n,n}$ and $T_{n,n}(\alpha)$, given by

$$C_n(\alpha) = \operatorname{diag}\left(\frac{1}{1-\rho_n(\alpha)}, 1, \ldots, 1\right). \tag{2.2.8}$$

The sum of squares $\sum_k \|E_k\|^2$ in (2.1.3) is obviously minimized by

$$\widehat{\theta}_n = S_{n,n-1}^{-1} \sum_{k=1}^{n} \Phi_{n,k-1}\, X_{n,k} \tag{2.2.9}$$

whereas in (2.2.6), it is minimized (as a function of α), by

$$\widehat{\vartheta}_n(\alpha) = T_{n,n-1}^{-1}(\alpha) \sum_{k=1}^{n} \Psi_{n,k-1}(\alpha)\, X_{n,k}. \tag{2.2.10}$$

If $p = 1$, then $\widehat{\vartheta}_n(\alpha) = \widehat{\vartheta}_n$ coincides with $\widehat{\theta}_n$. But before coming to the study of the rate parameter α, it should be noted that as an interesting consequence of our study, it is now possible to build a consistent and asymptotically normal estimator of θ_n. On the basis of (2.2.5), let

$$\widetilde{\theta}_n = \widehat{J}_n\, \widehat{\beta}_n + \widehat{v}_n\, e_p \tag{2.2.11}$$

where $\widehat{J}_n$ is J_n with $\lambda_{n,1}$ that can be estimated by $\widehat{v}_n = e_p^\top \widehat{\vartheta}_n$, the first component of $\widehat{\vartheta}_n$, and $\widehat{\beta}_n$ is $\widehat{\vartheta}_n$ deprived of its first component (we drop the dependency on α in the notation of this paragraph, for readability). Then, we have the following result.

Proposition 2.2.1. *Assume that (A1), (A2) and (A3) hold, and that* $\mathbb{E}\big[|\varepsilon_1|^{2+\nu}\big] = \eta_\nu < +\infty$ *for some* $\nu > 0$. *Then,*

$$\big\|\widetilde{\theta}_n - \theta_n\big\| \xrightarrow{\;p\;} 0 \quad \textit{and} \quad \sqrt{n}\,\big(\widetilde{\theta}_n - \theta_n\big) \xrightarrow{\;d\;} \mathcal{N}\big(0,\, J\,\Sigma_{p-1}^{-1} J^\top\big) \tag{2.2.12}$$

where Σ_{p-1} *is given in* (2.2.3) *and*

$$J = \begin{pmatrix} 1 & & \\ -\lambda_1 & \ddots & \\ & \ddots & 1 \\ & & -\lambda_1 \end{pmatrix} \quad (= 0 \text{ if } p = 1).$$

Proof. See Section 2.5.3. □

The asymptotic normality is degenerate in the sense that $J\,\Sigma_{p-1}^{-1}J^\top$ is not invertible (as a $p \times p$ matrix of rank at most $p-1$). This may be illustrated by the relation $v_p^\top J = 0$ where $v_p^\top = (\lambda_1^{i-1})_{1 \leqslant i \leqslant p}$ so that (2.2.12) becomes

$$\sqrt{n}\, v_p^\top\big(\widetilde{\theta}_n - \theta_n\big) \xrightarrow{\ p\ } 0$$

and a parallel can be made with (2.1.5). It is not surprising for $p = 1$ since it is well-known that in that case, $|\widetilde{\theta}_n - \theta_n| = O_p(n^{-(1+\alpha)/2})$, see *e.g.* Phillips and Magdalinos (2007).

Remark 2.2.1. Strictly speaking, $\widetilde{\theta}_n$ – as well as $\widehat{\theta}_n$ – is a consistent estimation of θ, the limit of θ_n, for obviously $\|\widetilde{\theta}_n - \theta\| \leqslant \|\widetilde{\theta}_n - \theta_n\| + \|\theta_n - \theta\|$. But it can only be asymptotically normal under the stronger hypothesis $\|\theta_n - \theta\| = o_p(n^{-1/2})$, which requires in particular that $\alpha > 1/2$. In this case, an immediate consequence of (2.2.12) is

$$\sqrt{n}\,\big(\widetilde{\theta}_n - \theta\big) \xrightarrow{\ d\ } \mathcal{N}\big(0,\, J\,\Sigma_{p-1}^{-1}J^\top\big). \tag{2.2.13}$$

2.2.1 The variance parameter c is known

After this contextualization, we are now ready to define a statistic that will be useful to achieve our objectives (recall that $\lambda_1 = \pm 1$ means that the limit process has a unit root located at ± 1). Suppose first that the variance parameter c is known and consider, provided that $\lambda_1 \widehat{v}_n(\alpha) < 1$, the statistic

$$\forall\, n > 1, \qquad \widehat{\alpha}_n(\alpha) = \frac{\ln c - \ln\big(1 - \lambda_1 \widehat{v}_n(\alpha)\big)}{\ln n} \tag{2.2.14}$$

where $\widehat{v}_n(\alpha) = e_p^\top\, \widehat{\vartheta}_n(\alpha)$ is the first component of $\widehat{\vartheta}_n(\alpha)$. Despite appearances, $\widehat{\alpha}_n(\alpha)$ is *not* an estimator of α because it actually depends on α. It is a statistic whose behavior, as a function of α, will help to formalize the test.

Theorem 2.2.2. *Assume that (A1), (A2) and (A3) hold, and that* $\mathbb{E}\big[|\varepsilon_1|^{2+\nu}\big] = \eta_\nu < +\infty$ *for some* $\nu > 0$*. Then, for the 'true' values* $0 < \alpha < 1$ *and* $c > 0$*,*

$$(\ln n)\sqrt{n^{1-\alpha}}\big(\widehat{\alpha}_n(\alpha) - \alpha\big) \xrightarrow{d} \mathcal{N}\big(0,\, 2\,c^{-1}\pi_{11}^{-2}\big) \tag{2.2.15}$$

where

$$\pi_{11} = \frac{\lambda_1^{p-1}}{\prod_{j=2}^{p}(1-\lambda_j)} \quad (= 1 \text{ if } p = 1). \tag{2.2.16}$$

Proof. See Section 2.5.4. □

Now choose some $\alpha/2 < \alpha_0 < 1$ (in particular, any $1/2 \leqslant \alpha_0 < 1$ is suitable), and consider the test statistic

$$Z_n^2(\alpha_0) = \frac{c\,\widehat{\pi}_n^2}{2}\,(\ln n)^2\, n^{1-\alpha_0}\big(\widehat{\alpha}_n(\alpha_0) - \alpha_0\big)^2 \tag{2.2.17}$$

where

$$\widehat{\pi}_n = \frac{\lambda_1^{p-1}}{\prod_{i=2}^{p}(1-\widehat{\lambda}_{n,i})} \quad (= 1 \text{ if } p = 1)$$

and $\widehat{\lambda}_{n,i}$ is the i-th eigenvalue of the estimated companion matrix $\widehat{A}_n$ built by plugging $\widehat{\theta}_n$ into A_n, ordered as indicated in the hypotheses. Whenever $\widehat{\alpha}_n(\alpha_0)$ is not defined for the reason mentioned above, we set $Z_n^2(\alpha_0) = +\infty$ by convention (so that $\mathcal{H}_0$ below is rejected). The following result is devoted to the asymptotic behavior of $Z_n^2(\alpha_0)$ under the null $\mathcal{H}_0 : \text{“}\alpha = \alpha_0\text{”}$ as well as the alternative $\mathcal{H}_1 : \text{“}\alpha > \alpha_0\text{”}$.

Theorem 2.2.3. *Assume that (A1), (A2) and (A3) hold, and that* $\mathbb{E}\big[|\varepsilon_1|^{2+\nu}\big] = \eta_\nu < +\infty$ *for some* $\nu > 0$*. Then, under* $\mathcal{H}_0 : \text{“}\alpha = \alpha_0\text{”}$*, the test statistic* (2.2.17) *satisfies*

$$Z_n^2(\alpha_0) \xrightarrow{d} \chi_1^2 \tag{2.2.18}$$

where χ_1^2 *has a chi-square distribution with one degree of freedom. If we additionally assume that* $\alpha < 2\alpha_0$*, then under* $\mathcal{H}_1 : \text{“}\alpha > \alpha_0\text{”}$*,*

$$Z_n^2(\alpha_0) \xrightarrow{p} +\infty. \tag{2.2.19}$$

Proof. See Section 2.5.5. □

From now on, we replace the assumption $\alpha < 2\alpha_0$ by the stronger one $\alpha_0 \geqslant 1/2$ (this will be justified in the empirical section). As a corollary, for some $\alpha_0 \geqslant 1/2$, the rejection of $\mathcal{H}_0 : \text{“}\alpha = \alpha_0\text{”}$ in favor of $\mathcal{H}_1 : \text{“}\alpha > \alpha_0\text{”}$ will happen as soon as $Z_n^2(\alpha_0) > z_{1-\epsilon}$ where

$z_{1-\epsilon}$ is the quantile of order $1-\epsilon$ of the χ^2_1 distribution. All the empirical study of Section 2.3 is based on this rule (but note that instead of (2.2.17), we could as well use the first component of $\widehat{\vartheta}_n(\alpha_0)$ in (2.2.10) and reach the same conclusions, through the limiting distribution of Lemma 2.5.4). To evaluate the extent of instability, a natural choice may be, given c,

$$\widehat{\alpha}_c = \inf\{\alpha_0 \in \mathcal{A} : \mathcal{H}_0 \text{ is not rejected}\} \tag{2.2.20}$$

where $\mathcal{A} \subset [1/2, 1[$ is a grid of test values (by convention $\inf \varnothing = +\infty$ so that our conclusion is that the series is integrated if $\mathcal{H}_0$ is always rejected and $\alpha \geqslant 1/2$). It remains to generalize the procedure to the less common situation where the limit process has a negative unit root, that is $\lambda_1 = -1$. Actually, Theorems 2.2.2 and 2.2.3 are still valid: λ_1 already appears in the first component of $\vartheta_n(\alpha)$ in (2.2.6), *i.e.* $\lambda_{n,1} = \lambda_1\rho_n$, and in $\widehat{\alpha}_n(\alpha)$ given in (2.2.14). The reader is referred to Section 2.5.6 for more technical details.

2.2.2 The variance parameter c is estimated

As mentioned in the beginning of the section, the over-parametrization makes it impossible to dissociate c from α without additional constraints. We suggest instead a cross-validation with the following protocol:

— For each $c_0 \in \mathcal{C}$, compute $\widehat{\alpha}_{c_0}$ based on the first $(n - n_{val})$ values of the series.

— Choose $(\widehat{c}_0, \widehat{\alpha}_{\hat{c}_0})$ as the pair that optimizes some criterion stemming from the one-step predictions of the n_{val} remaining values.

Clearly, the value of α selected by (2.2.20) has to be biased downwards (due to overlapping confidence intervals for increasing values of α_0). Thus, we must also expect $\widehat{c}_0$ to be biased downwards because of the ratio c/n^α. Paradoxically, despite unsatisfactory individual evaluations of c and α, the compensated result may prove interesting, as the empirical section suggests (see Figure 2.6). In terms of computational costs, proceeding like this is a great faster than testing all $(c, \alpha) \in \mathcal{C} \times \mathcal{A}$ since the cross-validation criterion is only computed once for each c. Overall, the procedure contains $O(1+|\mathcal{C}||\mathcal{A}|+|\mathcal{C}|n_{val})$ steps, where in this context a 'step' requires at worst the inversion of a $p \times p$ matrix ($p \ll n$). Hence, this is not so large considering that n_{val} is usually a small percentage of n. Especially as from a practical point of view, it does not seem necessary for $\mathcal{C}$ to have a very fine mesh (unlike $\mathcal{A}$).

2.3 Empirical Study

This section is dedicated to the empirical performances of the test procedure described in Theorem 2.2.3 together with the selection rule (2.2.20), and also to the cross-validation process presented in Section 2.2.2.

2.3.1 On simulated data (c is known)

Without loss of generality, let us consider $c = 1$ to produce the data, and set $\rho_n = 1 - n^{-\alpha} = \pm\lambda_{n,1}$ with $1/2 \leqslant \alpha < 1$. For each simulation, $\lambda_{n,2}, \ldots, \lambda_{n,p}$ are randomly chosen in the interval $[-\rho_n + 0.1, \rho_n - 0.1]$, and a process of size n is generated with a standard Gaussian noise. The test of $\mathcal{H}_0 : "\alpha = \alpha_0"$ against $\mathcal{H}_1 : "\alpha > \alpha_0"$ is done *via* Theorem 2.2.3 for $\alpha_0 \in \mathcal{A} = \{1/2 + k/50,\, k = 0, \ldots, 24\}$. The rejection frequencies are evaluated on the basis of 5000 replications for each α_0. The distribution of $Z_n^2(\alpha_0)$ and the empirical power of the test with 5% type I risk are represented on Figure 2.1 ($\alpha = 2/3$), Figure 2.2 ($\alpha = 3/4$) and Figure 2.3 ($\alpha = 4/5$), for $\lambda_1 = 1$, $n = 1000$ and $p \in \{1, 2, 3, 4\}$. To get an idea of the impact of smaller samples and a negative unit root, similar experiments are conducted on Figure 2.4 (with $\lambda_1 = 1$, $\alpha = 4/5$, $p = 1$, $n = 250$ on the top, and $\lambda_1 = -1$, $\alpha = 4/5$, $p = 1$, $n = 1000$ on the bottom). The light-gray items (at the right of the vertical line) correspond to $\alpha_0 > \alpha$, situations beyond the scope of our study that we decided anyway to incorporate in the simulations for informational purposes. Finally, the distribution of $\widehat{\alpha}_{c=1}$ over 5000 replications is outlined for $\alpha \in \{2/3, 3/4, 4/5\}$ with $n = 250$ and $n = 1000$, on Figure 2.5. For the latter simulations, we have adopted the rule (2.2.20) with 5% type I risk.

From the simulations of Figures 2.1–2.3, we observe that p has a negligible influence on the asymptotic behaviors when $\lambda_1 = 1$. Under $\mathcal{H}_0$, the rejection rate is $\approx 5\%$ as expected and the power of the test is quite satisfying under $\mathcal{H}_1$ ($\geqslant 75\%$ for $\alpha_0 \leqslant \alpha - 0.2$, $\approx 100\%$ for $\alpha_0 \leqslant \alpha - 0.3$). The decrease in the rejection rates is rather similar for α_0 growing from $1/2$ to α, whether $\alpha = 2/3$, $\alpha = 3/4$ or $\alpha = 4/5$. However, we can see that the trickier alternatives where $\alpha_0 \leqslant \alpha - 0.1$ are underestimated ($\approx 25\%$): this is due to the fine mesh of $\mathcal{A}$, it is difficult to discern substantial differences in the statistic between values of α_0 that increases 0.02 by 0.02 and overall, the numerical results are encouraging.

In terms of sample sizes, the results are obviously impacted when we set $n = 250$ as can be seen on the upper part of Figure 2.4 with $p = 1$ and $\alpha = 4/5$ ($\geqslant 50\%$ for $\alpha_0 \leqslant \alpha - 0.2$), but the global behavior follows the same lines. The real time series of next subsection are

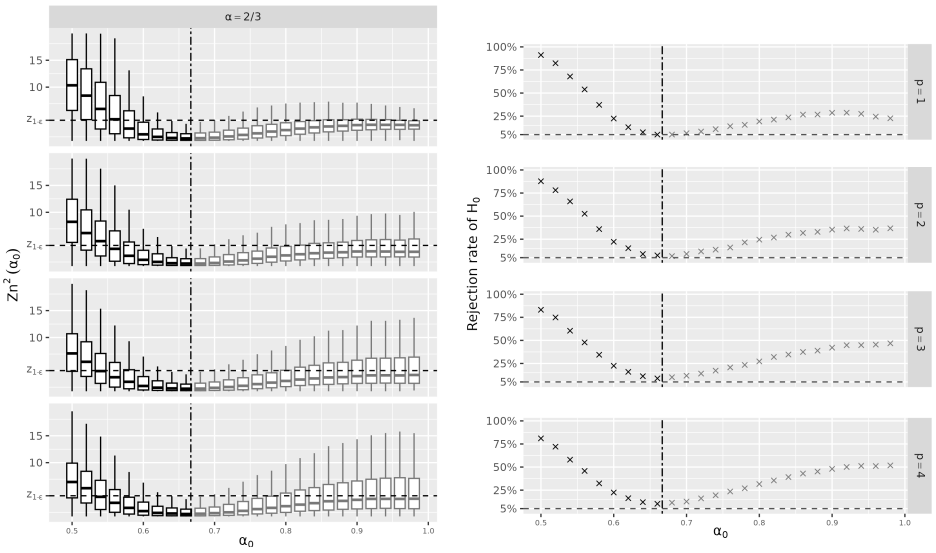


Figure 2.1 – Empirical distribution of $Z_n^2(\alpha_0)$ for $n = 1000$, $\lambda_1 = 1$, $\alpha = 2/3$, $\alpha_0 \in \mathcal{A}$ and $p \in \{1, 2, 3, 4\}$ on the left. The corresponding rejection frequencies of $\mathcal{H}_0 : \text{“}\alpha = \alpha_0\text{”}$ against $\mathcal{H}_1 : \text{“}\alpha > \alpha_0\text{”}$ (5% level) are represented on the right.

even smaller ($n = 80$ to $n = 129$), that will mainly result in larger confidence intervals. Under the hypothesis that the limit process has a negative unit root, Figure 2.4 (on the bottom) also highlights the same efficiency but we must precise here that, unlike $\lambda_1 = 1$, more experiments than those presented here showed a notable degradation with p: the decrease is faster and the statistic falls below $z_{1-\epsilon}$ too quickly. However, despite the lack of power observed for $p \geqslant 2$, this remains an innovative approach since we recall that the ADF, KPSS, and all usual tests for integration or trend-stationarity are unable to detect instability coming from negative unit roots.

As expected, the selected values of α on Figure 2.5 are underestimated, on average: for $n = 1000$, we can roughly say that $\widehat{\alpha}_{c=1}$ is close to $(\alpha{-}0.1)$, yet greater than $(\alpha{-}0.2)$ with an overwhelming probability. It is a direct consequence of the construction rule (2.2.20) and overlapping confidence intervals (in particular, together with $\inf \mathcal{A} = 1/2$, that explains the initial peak on smaller samples). So, the spectral radius must be underestimated as well. Fortunately the next section will show that, by evaluating c together with α instead of arbitrarily setting $c = 1$, the estimation is clearly improved. Of course a fully unbiased strategy may be an interesting trail for future studies (even if the procedure is asymptotically unbiased), see the conclusion for some comments on the topic.

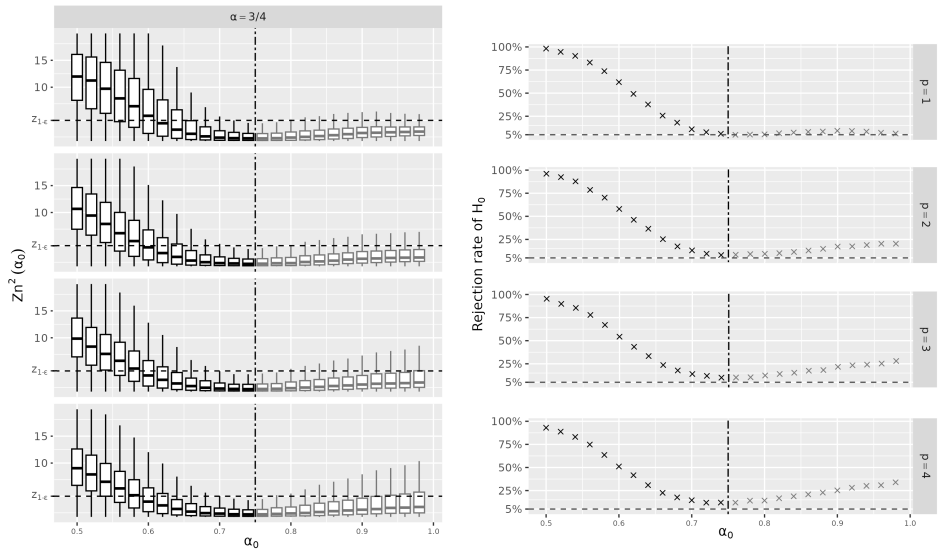


Figure 2.2 – Empirical distribution of $Z_n^2(\alpha_0)$ for $n = 1000$, $\lambda_1 = 1$, $\alpha = 3/4$, $\alpha_0 \in \mathcal{A}$ and $p \in \{1, 2, 3, 4\}$ on the left. The corresponding rejection frequencies of $\mathcal{H}_0 : \text{“}\alpha = \alpha_0\text{”}$ against $\mathcal{H}_1 : \text{“}\alpha > \alpha_0\text{”}$ (5% level) are represented on the right.

2.3.2 On simulated data (c is estimated)

The estimation of c according to the cross-validation described in Section 2.2.2 actually proves beneficial. Indeed, let us consider three groups of experiments generated as in the previous section, with $n = 1000$, $\lambda_1 = 1$, $p = 2$ and (α, c) uniformly chosen out of $\mathcal{A} \times \mathcal{C} = \{1/2 + k/50,\ k = 0, \dots, 24\} \times \{1/2 + k/5,\ k = 0, \dots, 22\}$. Figure 2.6 shows the scatterplot of $\widehat{c}/n^{\widehat{\alpha}}$ as a function of c/n^{α} together with a linear fit (with 95% confidence interval) for each of them. In the first one, we basically set $c = 1$ and estimate α using (2.2.20). This corresponds to the decreasing linear fit on the plot and reveals a significant deterioration, hence the inadequacy of making $\widehat{c} = 1$ independent of the data: the bias of $\widehat{\alpha}_{\hat{c}=1}$ may strongly affect the results. The second group of experiments consists of giving $\widehat{c} = c$ the true value before computing $\widehat{\alpha}_{\hat{c}=c}$ by the same strategy. The linear fit is clearly better, but the parallelism with the bisector of the coordinate axes unveils a systematic bias. This is a predictable consequence of overlapping confidence intervals of (2.2.20) already mentioned in the previous section: $\widehat{\alpha}_{\hat{c}}$ is underestimated so that $c/n^{\widehat{\alpha}_{\hat{c}}}$ overestimates the true value and the points are generally above the identity function. Finally, in the third group of experiments we estimate the pair (c, α) by the cross-validation of Section 2.2.2 with $c_0 \in \mathcal{C}$ defined above, $n_{val} = n/10$ and the MSE criterion. One can see that except for the small values of c/n^{α} (either c is small or α is large) corresponding to a spectral radius

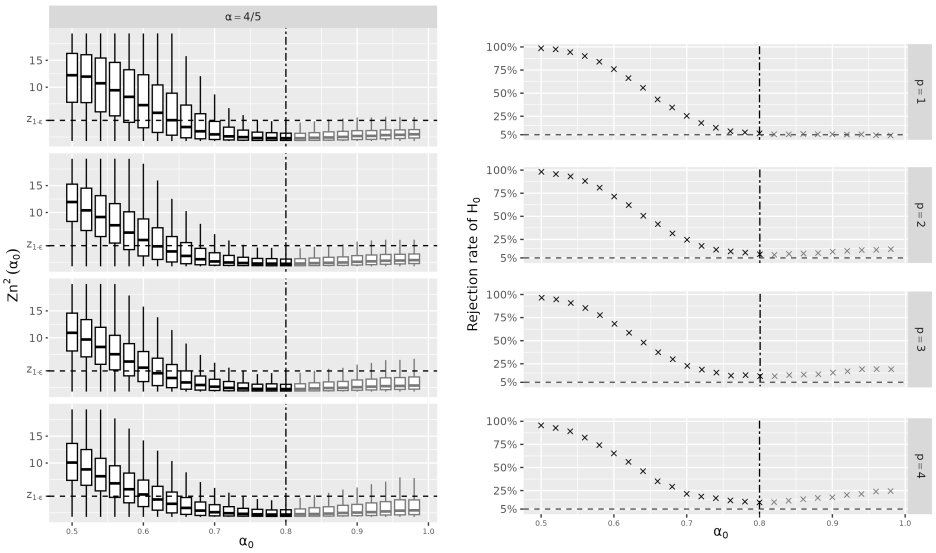


Figure 2.3 – Empirical distribution of $Z_n^2(\alpha_0)$ for $n = 1000$, $\lambda_1 = 1$, $\alpha = 4/5$, $\alpha_0 \in \mathcal{A}$ and $p \in \{1, 2, 3, 4\}$ on the left. The corresponding rejection frequencies of $\mathcal{H}_0 : "\alpha = \alpha_0"$ against $\mathcal{H}_1 : "\alpha > \alpha_0"$ (5% level) are represented on the right.

close to one and where all methods perform quite similarly, only the last one is able to counterbalance the bias of $\widehat{\alpha}_{\hat{c}}$ and results in a more reliable estimation. Incidentally, the t-tests for equality of slopes in the groups confirm the visual intuition through a clear rejection. The latter approach is now going to be used on a real dataset.

2.3.3 On a real dataset

The extended dataset `NelPlo` is available as open-source `https://www.rdocumentation.org/packages/tseries/versions/0.9-9/topics/NelPlo` and contains 14 macroeconomic series ending in 1988. According to Nelson and Plosser (1982), those data are known to be either integrated (in the sense of the ADF test) or trend-stationary with spectral radius close to 1, so let us approach them from the nearly-unstable point of view, with positive unit root ($\lambda_1 = 1$). It is important to note that the data of `NelPlo` are longer than the original ones so that our conclusions may be slightly different. The statistical protocol itself is different: p is chosen by looking at the partial autocorrelations of the differentiated series and no trend is estimated. Then, the pair $(\widehat{c}, \widehat{\alpha}_{\hat{c}})$ is selected through the cross-validation described in Section 2.2.2, with $n_{val} = n/10$ and the MSE criterion. An (approximate) 95%-confidence interval of $\widehat{\alpha}_{\hat{c}}$ is also derived from Theorem 2.2.2 with $c = \widehat{c}$. Because n is usually small, in any case much smaller than in the simulation study,

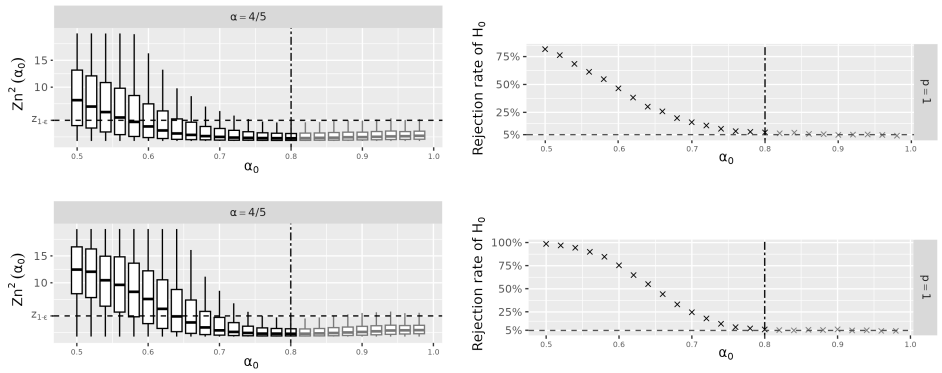


Figure 2.4 – Empirical distribution of $Z_n^2(\alpha_0)$ for $n = 250$, $\lambda_1 = 1$, $\alpha = 4/5$, $\alpha_0 \in \mathcal{A}$ and $p = 1$ on the top left. The corresponding rejection frequencies of $\mathcal{H}_0 : \text{“}\alpha = \alpha_0\text{”}$ against $\mathcal{H}_1 : \text{“}\alpha > \alpha_0\text{”}$ (5% level) are represented on the right. Same experiments, but with $n = 1000$, $\lambda_1 = -1$, $\alpha = 4/5$, $\alpha_0 \in \mathcal{A}$ and $p = 1$ on the bottom.

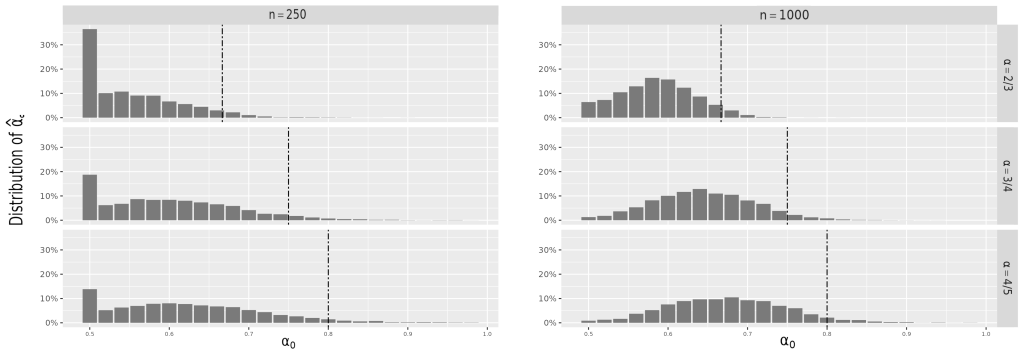


Figure 2.5 – Empirical distribution of $\widehat{\alpha}_c$ (5% level) with known value $c = 1$, for $n = 250$ (left) and $n = 1000$ (right), $\lambda_1 = 1$, $\alpha \in \{2/3, 3/4, 4/5\}$, $\alpha_0 \in \mathcal{A}$ and $p = 1$.

we have restrained the domain of c and only considered a search grid $\mathcal{C} \subset [0.8, 1.2]$, while as before $\mathcal{A} \subset [0.5, 1[$. See Table 2.1 below for the numerical outcomes.

Although the statistical context is different, our results are rather consistent with Table 5 of Nelson and Plosser (1982). According to these authors, all series except ‘Unemployment rate’ are integrated. In fact it seems that among them, only ‘Interest rate’ has a unit root whereas the nearly-unstable approach may be suitable for the others. It should be noted that the only stationary series of the authors (radius ≈ 0.71) corresponds to one of the lowest spectral radius given by our procedure. Similarly, the most integrated series of the authors is ‘Interest rate’ (radius ≈ 1.03), that is the only one being considered as integrated in our study. There are obviously points of divergence due to different protocols, to the fact that we did not try to detrend the series beforehand, and so on, but it

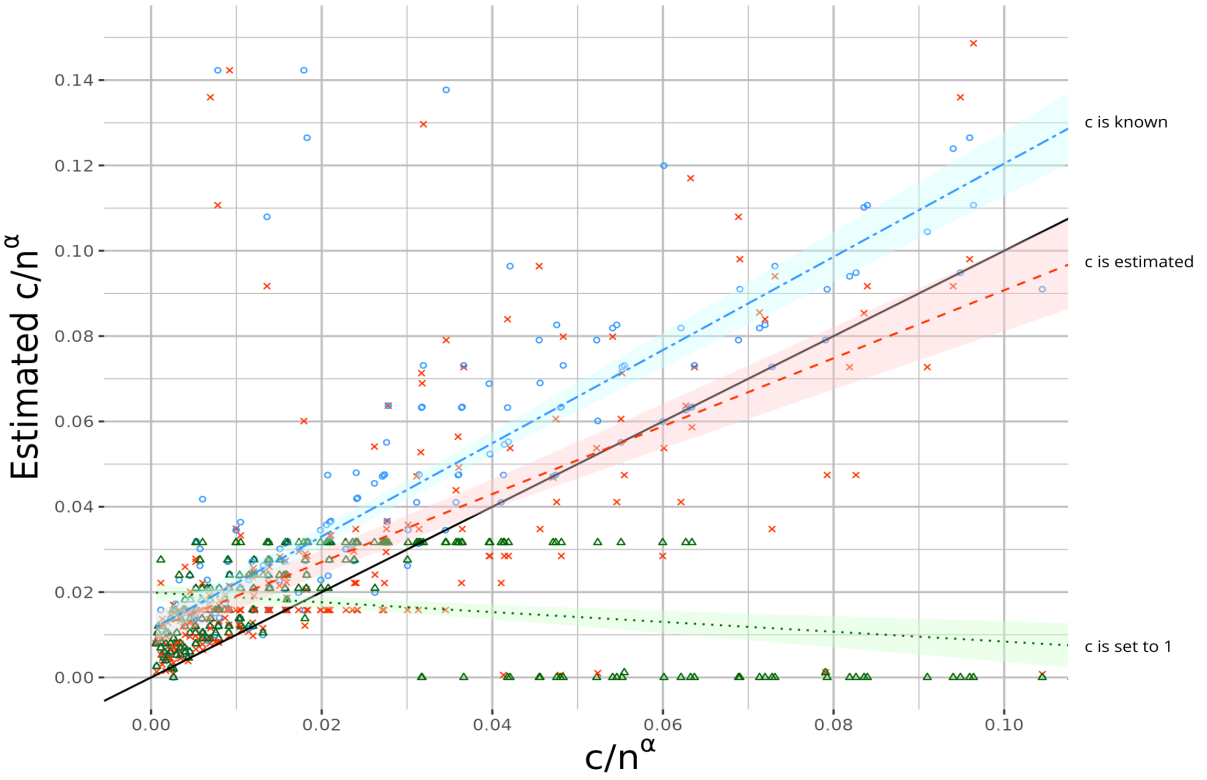


Figure 2.6 – Scatterplot of $\widehat{c}/n^{\widehat{\alpha}}$ as a function of c/n^{α} in three different cases: c is set to 1 (triangles), c is known (circles) and c is estimated by cross-validation (crosses). Experiments are conducted according to the given protocol to simulate the data and estimate α. A linear fit (with 95% confidence interval) shows the mean evolution in each group and the solid line is the identity function

.

is interesting to note that we share the same general ideas on these data. For illustrative purposes, Figures 2.7-2.10 contain some curves from the dataset and associated forecasts.

2.4 Conclusion and Perspectives

Our motivation was to bring a fresh look to the unit root issues in time series analysis, through the concept of near-instability. On a dubious series in terms of stationarity, the 'extent of instability' may now be subject to a testing procedure, and many opportunities of improvements should follow from this innovative approach. For example, it would be interesting to unveil the limit behavior of the test statistic with no constraints on α.

Series	n	p	$\widehat{c}$	$\widehat{\alpha}_{\hat{c}}$	$\widehat{\rho}_n = 1 - \widehat{c}/n^{\widehat{\alpha}_{\hat{c}}}$
Velocity	120	1	0.80	$\in [0.50, 0.66]$	$\in [0.93, 0.97]$
Industrial production	129	6	0.80	$\in [0.50, 0.72]$	$\in [0.93, 0.98]$
Nominal GNP	80	2	0.80	$\in [0.50, 0.57]$	$\in [0.91, 0.93]$
Consumer prices	129	4	1.08	$\in [0.73, 0.98]$	$\in [0.97, 0.99]$
Employment	99	5	0.80	$\in [0.50, 0.59]$	$\in [0.92, 0.95]$
Interest rate	89	4	0	$+\infty$	1
Wages	89	2	0.80	$\in [0.50, 0.56]$	$\in [0.92, 0.94]$
GNP deflator	100	6	1.20	$\in [0.50, 0.53]$	$\in [0.88, 0.90]$
Money stock	100	3	0.80	$\in [0.50, 0.53]$	$\in [0.92, 0.93]$
Real GNP	80	4	1.10	$\in [0.50, 0.74]$	$\in [0.88, 0.96]$
Common stock prices	118	6	0.82	$\in [0.50, 0.98]$	$\in [0.92, 0.99]$
Real per capita GNP	80	2	0.80	$\in [0.50, 0.62]$	$\in [0.91, 0.95]$
Real wages	89	2	1.04	$\in [0.50, 0.79]$	$\in [0.89, 0.97]$
Unemployment rate	99	3	1.08	$\in [0.50, 0.65]$	$\in [0.89, 0.95]$

Table 2.1 – `NelPlo` dataset from the nearly-unstable point of view. For each series, $(\widehat{c}, \widehat{\alpha}_{\hat{c}})$ coming from the cross-validation of Section 2.2.2 is given together with an (approximate) 95% confidence interval for $\widehat{\alpha}_{\hat{c}}$. The estimated spectral radius is also precised.

The light-gray boxplots of Figures 2.1–2.3 suggest that $\mathcal{H}_1$: "$\alpha < \alpha_0$" does not lead to a divergence but to some kind of stabilization away from zero (in distribution ?). But a spectral radius close to unity together with $\alpha < 1/2$ must correspond to high values of n, questioning the practical interest (a series with moderate size is more likely stable in that context). On the other hand, when $\alpha \geqslant 1/2$, a bilateral test would undoubtedly refine the selection of $\widehat{\alpha}_c$ and contribute to eliminate the observed bias, so this is an important issue. Also, the choice of c presented here is based on empirical foundations, which makes it difficult to obtain theoretical grounds related to consistency or normality. The authors are pretty convinced that a joint estimation of (c, α) through a penalized Gaussian likelihood may be a solution to improve the estimation, bypass over-parametrization through well-chosen constraints and overall, get guarantees. Authorizing $p = p_n$ to grow to infinity could lead to the formalization of nearly-unstable ARMA(p, q) processes under an hypothesis of invertibility at fixed n and some restrictions on p_n – like the strategy of ADF. The results of this paper shall then probably be generalized without much difficulty to that wider class of models. Especially as, on a broader scale, any model with a critical parameter could be responsive to this kind of approach: the features of the reproduction law for counting and Galton-Watson processes (see Ispány et al. (2003), Ispány et al. (2005)), the mean

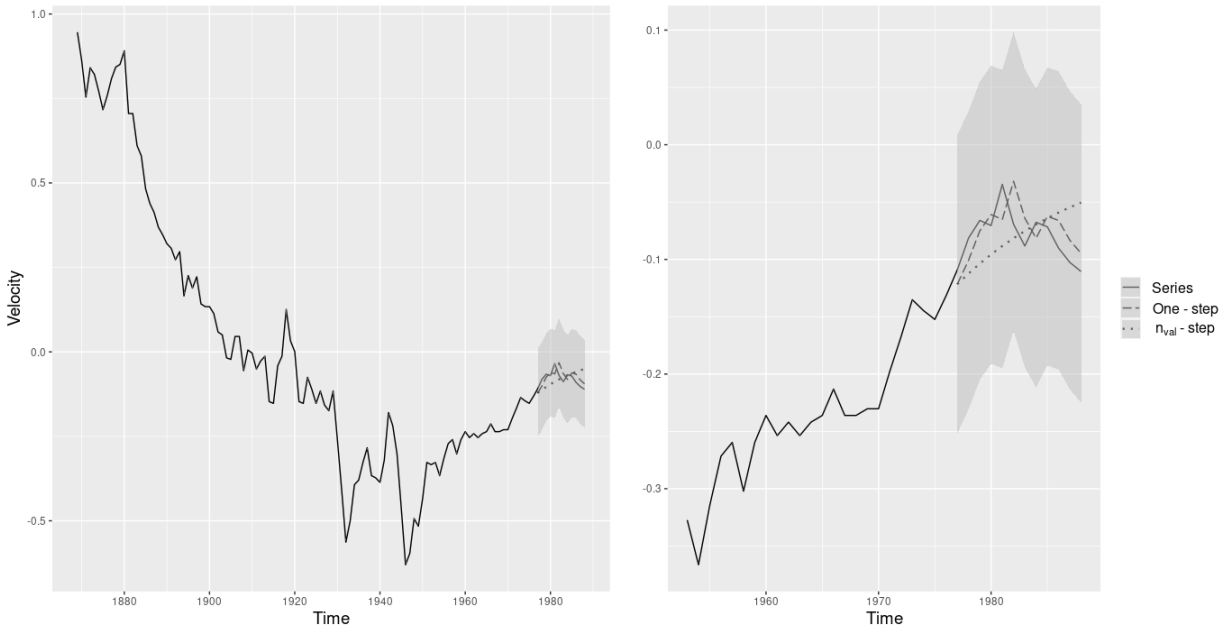


Figure 2.7 – Series 'Velocity' and associated predictions for the last $n_{val} = 12$ values (one-step and n_{val}-step forecasts). The Gaussian 95% confidence interval for one-step predictions is added. The whole series is represented on the left whereas on the right a zoom focuses on the last decades.

value of the random coefficients in RCAR processes, the fractional order of integration in FARIMA processes, the Hurst exponent of long-term memory time series, and so on. For p-dimensional processes, the first step should be to express the stability/instability criterion through the spectral radius of a companion matrix. For example in Barczy et al. (2011), an INAR(p) process with $\rho(A) < 1$ is stable whereas it is unstable for $\rho(A) = 1$ (see A in (2.2), Prop. 2.6 and the definition that follows). Hence, through probably arduous demonstrations, it seems that our approach might be adaptable: let $\rho(A_n) < 1$ with $\rho(A_n) \to 1$, and consider near-unstability in such a count time series to generalize from first-order to p-order. Finally, a natural question arising here concerns explosive nearly-unstable processes, that is when the spectral radius tends to one from the *outer* neighborhood of the unit circle. Although less interesting from a practical point of view, this seems to be a challenging trail for future studies.

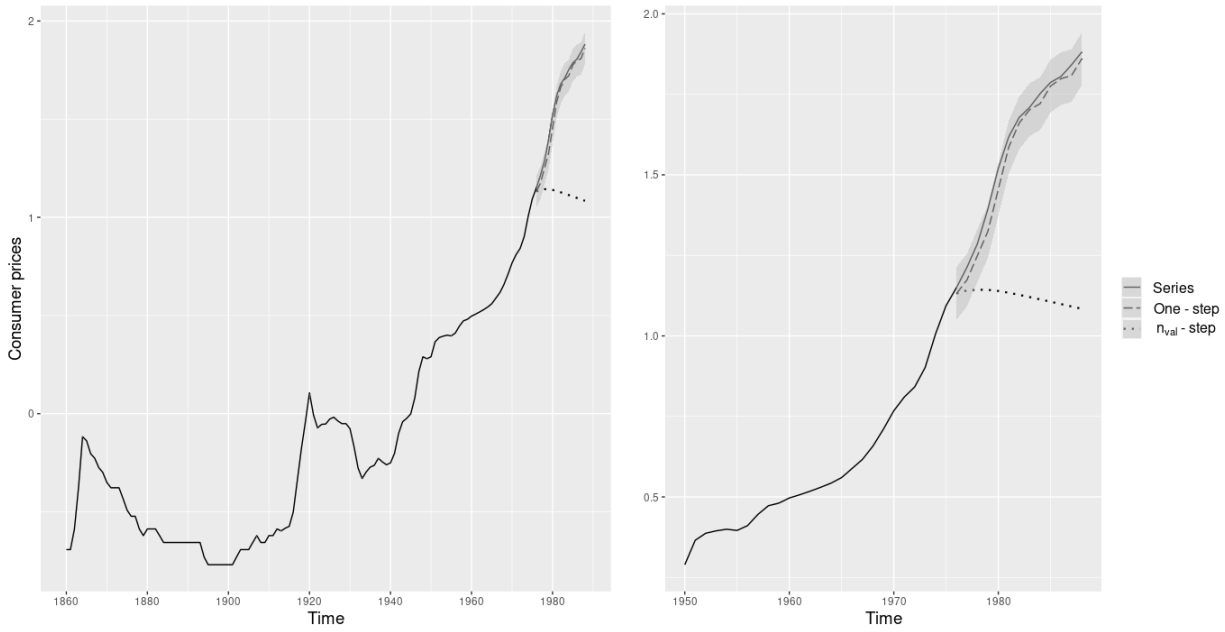


Figure 2.8 – Same graphs as in Figure 2.7 with the series 'Consumer prices' and $n_{val} = 13$ values.

2.5 Technical Developments

To deal with martingale terms, we need to consider the triangular array of filtrations

$$\forall\, n \geqslant 1,\ \forall\, 1 \leqslant k \leqslant n, \qquad \mathcal{F}_{n,k} = \sigma(\Phi_{n,0},\, \varepsilon_1,\, \ldots,\, \varepsilon_k) \tag{2.5.1}$$

obviously satisfying $\mathcal{F}_{n,0} \subseteq \mathcal{F}_{n,1} \subseteq \ldots \subseteq \mathcal{F}_{n,n}$ for each $n \geqslant 1$.

2.5.1 Known results

The two lemmas given in this section may be found with far more details in Badreau and Proïa (2023).

Lemma 2.5.1. *Assume that (A1) holds. Then, there exists $n_0 \geqslant 1$ such that, for all $n > n_0$, A_n is diagonalizable in the form $A_n = P_n\, D_n\, P_n^{-1}$ with $D_n = \mathrm{diag}(\lambda_{n,1},\, \ldots,\, \lambda_{n,p})$ containing ordered distinct eigenvalues $1 > |\lambda_{n,1}| \geqslant \ldots \geqslant |\lambda_{n,p}| > 0$. In addition, $\|P_n\|$ and $\|P_n^{-1}\|$ are bounded and there exists an invertible matrix P such that $P_n \to P$ and $P_n^{-1} \to P^{-1}$.*

Note that because its spectrum is a subset of the one of A_n, the companion matrix $\bar{A}_n$ associated with $(V_{n,k})$ is also diagonalizable for all $n > n_0$ and the lemma still holds

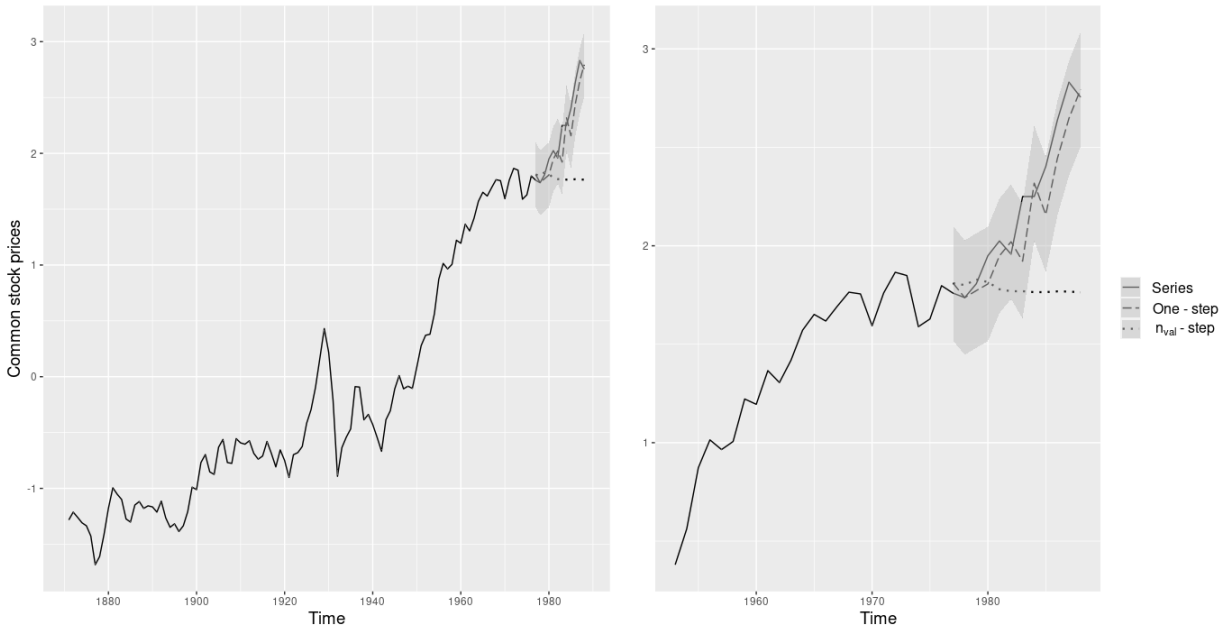


Figure 2.9 – Same graphs as in Figure 2.7 with the series 'Common stock prices' and $n_{val} = 12$ values.

when replacing A_n by $\bar{A}_n$ and obviously p by $p-1$. This will be a key argument to justify (2.5.3) in the next section.

Lemma 2.5.2. *Assume that (A1), (A2) and (A3) hold and that* $\mathbb{E}[\varepsilon_1^2] = \sigma^2 < +\infty$. *Then, as* n *tends to infinity,*

$$(1-\rho_n(\alpha))\,\frac{S_{n,n}}{n} \xrightarrow{\ p\ } \Gamma_p \triangleq \sigma^2 \ell \left(\lambda_1^{i+j}\right)_{1\leqslant i,j\leqslant p} \qquad \textit{with} \qquad \ell = \frac{\pi_{11}^2}{2} > 0,$$

where $S_{n,n}$ *is given in* (2.2.7) *and* π_{i1} $(i = 1,\dots,p)$ *are the (non-zero) entries of the first column of* P^{-1}. *Moreover,*

$$C_n^{-\frac{1}{2}}(\alpha)\,P_n^{-1}\,\frac{S_{n,n}}{n}\,(P_n^{\top})^{-1}C_n^{-\frac{1}{2}}(\alpha) \xrightarrow{\ p\ } \sigma^2 \begin{pmatrix} \frac{\pi_{11}^2}{2} & 0 & \cdots & 0 \\ 0 & \frac{\pi_{21}^2}{1-\lambda_2^2} & \cdots & \frac{\pi_{21}\pi_{p1}}{1-\lambda_2\lambda_p} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \frac{\pi_{p1}\pi_{21}}{1-\lambda_p\lambda_2} & \cdots & \frac{\pi_{p1}^2}{1-\lambda_p^2} \end{pmatrix} \triangleq H_p$$

where $C_n(\alpha)$ *is defined in* (2.2.8). *In addition,* H_p *is invertible (and positive definite if* A *has only real eigenvalues).*

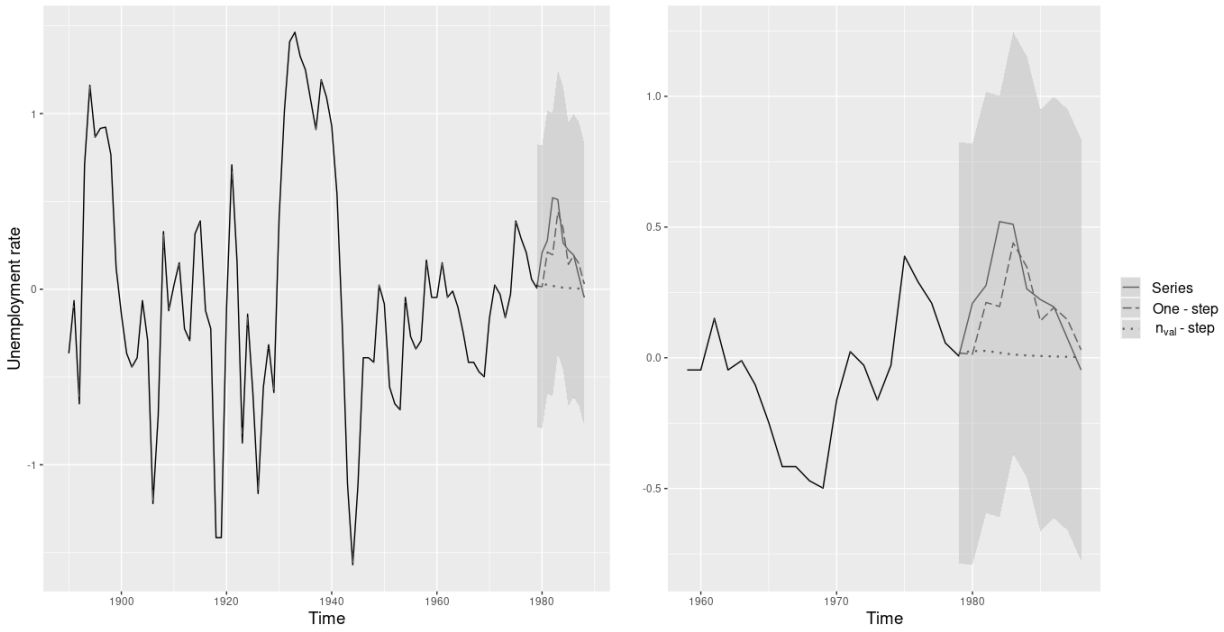


Figure 2.10 – Same graphs as in Figure 2.7 with the series 'Unemployment rate' and $n_{val} = 10$ values.

2.5.2 Technical lemmas

Lemma 2.5.3. *Assume that (A1), (A2) and (A3) hold and that* $\mathbb{E}[\varepsilon_1^2] = \sigma^2 < +\infty$. *Then, as* n *tends to infinity, for any* $\alpha/2 < \alpha_0 \leqslant \alpha$,

$$C_n^{-\frac{1}{2}}(\alpha)\,\frac{T_{n,n}(\alpha_0)}{n}\,C_n^{-\frac{1}{2}}(\alpha) \;\xrightarrow{p}\; \Upsilon_p \triangleq \sigma^2\,\mathrm{diag}\,(\ell,\,\Sigma_{p-1})$$

where $T_{n,n}(\alpha)$ *is defined in* (2.2.7), $C_n(\alpha)$ *in* (2.2.8), *and where* ℓ *and* Σ_{p-1} *are defined in Lemma 2.5.2 and* (2.2.3), *respectively. In addition,* Υ_p *is positive definite.*

Proof. Note that $S_{n,n}$ and $T_{n,n}(\alpha)$ share the same top-left element (which actually does not depend on α), so we directly get

$$\frac{1-\rho_n(\alpha)}{n}\sum_{k=0}^{n} X_{n,k}^2 = \frac{c}{n^{1+\alpha}}\sum_{k=0}^{n} X_{n,k}^2 \;\xrightarrow{p}\; \sigma^2\ell. \tag{2.5.2}$$

Moreover, coming back to the reasoning of Badreau and Proïa (2023), especially Lem. 13–15 and Prop. 2, and adapting it to the present case (see the remark after Lemma 2.5.1),

$$\frac{1}{n}\sum_{k=0}^{n}\Lambda_{n,k}(\alpha)\,\Lambda_{n,k}^{\top}(\alpha)\ \xrightarrow{\ p\ }\ \sigma^2\,\Sigma_{p-1}\quad\text{where}\quad\Lambda_{n,k}^{\top}(\alpha)=\Big(V_{n,k}(\alpha)\quad\cdots\quad V_{n,k-p+2}(\alpha)\Big). \tag{2.5.3}$$

Indeed, the equivalence

$$\frac{1}{n}\sum_{k=0}^{n}\Lambda_{n,k}(\alpha)\,\Lambda_{n,k}^{\top}(\alpha)\ \overset{p}{\sim}\ \left(\sum_{k\geqslant 0}\bar{A}_n^k\,K_{p-1}\,(\bar{A}_n^{\top})^k\right)\frac{1}{n}\sum_{k=1}^{n}\varepsilon_k^2,$$

the law of large numbers together with the fact that $\rho(\bar{A}_n)<1$ and $\rho(\bar{A})<1$ lead to (2.5.3). It remains to treat the (symmetrical) non-diagonal blocks. First,

$$\begin{aligned}\sum_{k=0}^{n}\Lambda_{n,k}^{(p)}(\alpha)\,X_{n,k}&=\Lambda_{n,0}^{(p)}(\alpha)X_{n,0}+\left[\sum_{k=1}^{n}\left(\Phi_{n,k}-\rho_n(\alpha)\,\Phi_{n,k-1}\right)\Phi_{n,k}^{\top}\right]e_p\\&=\left[S_{n,n}\,(I_p-\rho_n(\alpha)A_n^{\top})-\rho_n(\alpha)M_n+r_n\right]e_p\end{aligned} \tag{2.5.4}$$

where $\Lambda_{n,k}^{(p)}$ is the p-dimensional version of $\Lambda_{n,k}$ (with last element $V_{n,k-p+1}(\alpha)$),

$$M_n=\sum_{k=1}^{n}\Phi_{n,k-1}\,E_k^{\top}$$

and r_n is made of isolated terms. It is already proven (Lem. 14–15) that $\|M_n\|$ and $\|r_n\|$ are $o_p(n)$. As an intermediate result, we can see that for $n>n_0$,

$$\begin{aligned}C_n^{\frac{1}{2}}(\alpha)\,P_n^{\top}\,(I_p-\rho_n(\alpha)A_n^{\top})&=C_n^{\frac{1}{2}}(\alpha)\,(I_p-\rho_n(\alpha)D_n)\,P_n^{\top}\\&\xrightarrow{\ p\ }\operatorname{diag}\left(0,(1-\lambda_i)_{2\leqslant i\leqslant p}\right)P^{\top}\triangleq\Delta_p\,P^{\top}.\end{aligned} \tag{2.5.5}$$

Hence,

$$\begin{aligned}\frac{\sqrt{1-\rho_n(\alpha)}}{n}\,S_{n,n}\,(I_p-\rho_n(\alpha)A_n^{\top})&=P_n\,\frac{C_n^{\frac{1}{2}}(\alpha)\sqrt{c}}{n^{\frac{\alpha}{2}}}\left[C_n^{-\frac{1}{2}}(\alpha)\,P_n^{-1}\,\frac{S_{n,n}}{n}\,(P_n^{\top})^{-1}C_n^{-\frac{1}{2}}(\alpha)\right]\\&\qquad\times\left[C_n^{\frac{1}{2}}(\alpha)\,P_n^{\top}\,(I_p-\rho_n(\alpha)A_n^{\top})\right]\\&\xrightarrow{\ p\ }P\,K_p\,H_p\,\Delta_p\,P^{\top}\end{aligned} \tag{2.5.6}$$

where the block diagonal matrix H_p is given in Lemma 2.5.2 and Δ is defined above. But K_p, H_p and Δ_p have particular structures consisting of zeros in key locations and implying

the following pattern,

$$P\,K_p\,H_p\,\Delta_p = \begin{pmatrix} * & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ * & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} 0 & \cdots & \cdots & 0 \\ \vdots & * & \cdots & * \\ \vdots & \vdots & & \vdots \\ 0 & * & \cdots & * \end{pmatrix} = 0.$$

The combination of (2.5.2), (2.5.3), (2.5.6) and the remarks that follow lead to the convergence

$$C_n^{-\frac{1}{2}}(\alpha)\,\frac{T_{n,n}(\alpha)}{n}\,C_n^{-\frac{1}{2}}(\alpha) \;\xrightarrow{\;p\;}\; \Upsilon_p \tag{2.5.7}$$

that is, when α_0 takes the 'true' value α. We now wish to generalize the result to $T_{n,n}(\alpha_0)$ for any $\alpha/2 < \alpha_0 \leqslant \alpha$. Regarding this, note that

$$\Psi_{n,k}(\alpha_0) = \Psi_{n,k}(\alpha) + c\,(n^{-\alpha_0} - n^{-\alpha})\,\dot{\Phi}_{n,k} \tag{2.5.8}$$

where

$$\forall\, n \geqslant 1,\ \forall\, 1 \leqslant k \leqslant n, \qquad \dot{\Phi}_{n,k}^{\top} = \begin{pmatrix} 0 & X_{n,k-1} & \cdots & X_{n,k-p+1} \end{pmatrix}.$$

Thus, we have the decomposition

$$T_{n,n}(\alpha_0) = T_{n,n}(\alpha) + R_n \tag{2.5.9}$$

with

$$\begin{aligned} R_n \;&=\; c\,(n^{-\alpha_0} - n^{-\alpha}) \left[\sum_{k=0}^{n} \Psi_{n,k}(\alpha)\,\dot{\Phi}_{n,k}^{\top} + \sum_{k=0}^{n} \dot{\Phi}_{n,k}\Psi_{n,k}^{\top}(\alpha) \right] \\ &\quad + c^2\,(n^{-\alpha_0} - n^{-\alpha})^2 \sum_{k=0}^{n} \dot{\Phi}_{n,k}\,\dot{\Phi}_{n,k}^{\top} \;\triangleq\; R_{n,1} + R_{n,1}^{\top} + R_{n,2}. \end{aligned}$$

The direct computation shows that $\Psi_{n,k}(\alpha)\,\dot{\Phi}_{n,k}^{\top}$ contains a column of zeros, a line made of $X_{n,k}\,X_{n,k-i}$ $(1 \leqslant i \leqslant p-1)$, and a $(p-1)\times(p-1)$ block of terms of the form $V_{n,k-j}(\alpha)\,X_{n,k-i}$ $(0 \leqslant j \leqslant p-2,\ 1 \leqslant i \leqslant p-1)$. For $2 \leqslant i \leqslant p-1$, according to the first expression in (2.2.2),

$$\begin{aligned} \rho_n^{2i}(\alpha)\,X_{n,k-i} &= \rho_n^{2i-1}(\alpha)\,(X_{n,k-i+1} - V_{n,k-i+1}(\alpha)) \\ &= \rho_n^{i+1}(\alpha)\,X_{n,k-1} - \sum_{j=1}^{i-1} \rho_n^{2i-j}(\alpha)\,V_{n,k-i+j}(\alpha) \end{aligned}$$

which entails that, for any $0 \leqslant u \leqslant p-2$,

$$\rho_n^{2i}(\alpha) \sum_{k=1}^{n} V_{n,k-u}(\alpha)\, X_{n,k-i} = \rho_n^{i+1}(\alpha) \sum_{k=1}^{n} V_{n,k-u}(\alpha)\, X_{n,k-1} - \sum_{j=1}^{i-1} \rho_n^{2i-j}(\alpha) \sum_{k=1}^{n} V_{n,k-u}(\alpha)\, V_{n,k-i+j}(\alpha). \quad (2.5.10)$$

The first part of this equation is $o_p(n^{1+\alpha/2})$ and the second part is a finite sum of $O_p(n)$, see for instance (2.5.7). Thus, *via* Lemma 2.5.2,

$$\left\| C_n^{-\frac{1}{2}}(\alpha)\, \frac{R_{n,1}}{n}\, C_n^{-\frac{1}{2}}(\alpha) \right\| = O_p\left(\frac{n^{-\alpha_0} - n^{-\alpha}}{n^{1+\frac{\alpha}{2}}}\, n^{1+\alpha} \right) = o_p(1) \quad (2.5.11)$$

as soon as $\alpha_0 > \alpha/2$. In the same way, $R_{n,2}$ is closely related to the covariance matrix $S_{n,n}$ but with zeros on the first line and column. So,

$$\left\| C_n^{-\frac{1}{2}}(\alpha)\, \frac{R_{n,2}}{n}\, C_n^{-\frac{1}{2}}(\alpha) \right\| = O_p\left(\frac{(n^{-\alpha_0} - n^{-\alpha})^2}{n}\, n^{1+\alpha} \right) = o_p(1) \quad (2.5.12)$$

provided that $\alpha_0 > \alpha/2$. Using (2.5.7), (2.5.9), (2.5.11) and (2.5.12), the convergence is now established. Since $\ell > 0$ (Lemma 2.5.2), the positive-definiteness of Υ_p relies on the invertibility of Σ_{p-1}, which can be found *e.g.* in Brockwell and Davis (1991) in the context of a stable AR$(p-1)$ process with $\sigma^2 > 0$. □

Lemma 2.5.4. *Assume that (A1), (A2) and (A3) hold and that* $\mathbb{E}\big[|\varepsilon_1|^{2+\nu}\big] = \eta_\nu < +\infty$. *Then, as n tends to infinity, for the 'true' value of* α,

$$\sqrt{n}\, C_n^{\frac{1}{2}}(\alpha) \left(\widehat{\vartheta}_n(\alpha) - \vartheta_n(\alpha) \right) \xrightarrow{d} \mathcal{N}\big(0,\, \sigma^2\, \Upsilon_p^{-1}\big) \quad (2.5.13)$$

where the limit covariance Υ_p *is given in Lemma 2.5.3. Moreover, for any* $\alpha/2 < \alpha_0 < \alpha$,

$$\left\| \sqrt{n}\, C_n^{\frac{1}{2}}(\alpha_0) \left(\widehat{\vartheta}_n(\alpha_0) - \vartheta_n(\alpha_0) \right) \right\| \xrightarrow{p} +\infty. \quad (2.5.14)$$

Proof. Developing $X_{n,k} = \vartheta_n^{\top}(\alpha)\Psi_{n,k-1}(\alpha) + \varepsilon_k$ in (2.2.10), the estimation error satisfies

$$\sqrt{n}\, C_n^{\frac{1}{2}}(\alpha) \left(\widehat{\vartheta}_n(\alpha) - \vartheta_n(\alpha) \right) = \sqrt{n}\, C_n^{\frac{1}{2}}(\alpha)\, T_{n,n-1}^{-1}(\alpha) \sum_{k=1}^{n} \Psi_{n,k-1}(\alpha)\, \varepsilon_k. \quad (2.5.15)$$

The sequence $(\Psi_{n,k-1}(\alpha)\, \varepsilon_k)_{n,k}$ is a martingale-difference array w.r.t. $\mathcal{F}_{n,k}$ defined in

(2.5.1) so, for all $u \in \mathbb{R}^p \backslash \{0\}$,

$$M_{n,n}^{(u)} = u^\top \frac{C_n^{-\frac{1}{2}}(\alpha)}{\sqrt{n}} \sum_{k=1}^{n} \Psi_{n,k-1}(\alpha)\, \varepsilon_k \tag{2.5.16}$$

is a martingale array having predictable quadratic variation

$$\begin{aligned} \langle M^{(u)} \rangle_{n,n} &= \sigma^2\, u^\top C_n^{-\frac{1}{2}}(\alpha)\, \frac{T_{n,n-1}(\alpha)}{n}\, C_n^{-\frac{1}{2}}(\alpha)\, u \\ &\xrightarrow{p} \sigma^2\, u^\top \Upsilon_p\, u > 0 \end{aligned} \tag{2.5.17}$$

by the convergence in Lemma 2.5.3 (with $\alpha_0 = \alpha$) and the positive definiteness of Υ_p. It is now necessary to establish the Lindeberg's condition in order to apply the central limit theorem for arrays of martingales (see *e.g.* Thm. 1 of Sec. 8 of Pollard (1984)), namely that

$$\forall\, \epsilon > 0, \quad \sum_{k=1}^{n} \mathbb{E}\left[(\Delta M_{n,k}^{(u)})^2\, \mathbb{1}_{\{|\Delta M_{n,k}^{(u)}| > \epsilon\}} \,|\, \mathcal{F}_{n,k-1} \right] \xrightarrow{p} 0 \tag{2.5.18}$$

where $\Delta M_{n,k}^{(u)} = M_{n,k}^{(u)} - M_{n,k-1}^{(u)}$. But for any $1 \leqslant k \leqslant n$,

$$\mathbb{E}\big[(\Delta M_{n,k}^{(u)})^2\, \mathbb{1}_{\{|\Delta M_{n,k}^{(u)}| > \epsilon\}} \,|\, \mathcal{F}_{n,k-1} \big] = \varphi_{n,k-1}^{(u)}(\alpha)\, \frac{\xi_{n,k}}{n} \leqslant \varphi_{n,k-1}^{(u)}(\alpha)\, \frac{\xi_n^*}{n} \tag{2.5.19}$$

where $\varphi_{n,k-1}^{(u)}(\alpha) = u^\top C_n^{-1/2}(\alpha)\, \Psi_{n,k-1}(\alpha)\, \Psi_{n,k-1}^\top(\alpha)\, C_n^{-1/2}(\alpha)\, u$,

$$\xi_{n,k} = \mathbb{E}\big[\varepsilon_k^2\, \mathbb{1}_{\{|\Delta M_{n,k}^{(u)}| > \epsilon\}} \,|\, \mathcal{F}_{n,k-1} \big] \quad \text{and} \quad \xi_n^* = \sup_{1 \leqslant k \leqslant n} \xi_{n,k}.$$

Applying Hölder's and Markov's inequalities,

$$\begin{aligned} \forall\, 1 \leqslant k \leqslant n, \quad \xi_{n,k} &\leqslant \eta_\nu^{\frac{2}{2+\nu}}\, \mathbb{P}\big((\Delta M_{n,k}^{(u)})^2 > \epsilon^2 \,|\, \mathcal{F}_{n,k-1} \big)^{\frac{\nu}{2+\nu}} \\ &\leqslant \eta_\nu^{\frac{2}{2+\nu}} \left(\frac{\sigma^2\, \varphi_{n,k-1}^{(u)}(\alpha)}{\epsilon^2\, n} \right)^{\frac{\nu}{2+\nu}} \xrightarrow{p} 0 \end{aligned}$$

from Lemma 2.5.3 (with $\alpha_0 = \alpha$), where η_ν is the moment of order $2 + \nu$ of the white noise (ε_k). Thus $\xi_n^* = o_p(1)$, see *e.g.* Lem. 1.3.20 of Duflo (1997). Together with (2.5.19) and again Lemma 2.5.3, (2.5.18) follows. Consequently, the martingale defined in (2.5.16) satisfies

$$M_{n,n}^{(u)} \xrightarrow{d} \mathcal{N}(0,\, \sigma^2\, u^\top \Upsilon_p\, u)$$

and because that holds for any $u \neq 0$, by the Cramér-Wold device,

$$\frac{C_n^{-\frac{1}{2}}(\alpha)}{\sqrt{n}} \sum_{k=1}^{n} \Psi_{n,k-1}(\alpha)\,\varepsilon_k \xrightarrow{d} \mathcal{N}\big(0,\, \sigma^2\,\Upsilon_p\big). \tag{2.5.20}$$

Combined with (2.5.17), Slutsky's lemma and (2.5.15),

$$\sqrt{n}\,C_n^{\frac{1}{2}}(\alpha)\left(\widehat{\vartheta}_n(\alpha) - \vartheta_n(\alpha)\right) \xrightarrow{d} \mathcal{N}\big(0,\, \sigma^2\,\Upsilon_p^{-1}\big).$$

That concludes the proof of the first part of the lemma. For the other part, note that using (2.2.6), (2.5.8) and (2.2.10), we have the decomposition

$$\begin{aligned}\widehat{\vartheta}_n(\alpha_0) - \vartheta_n(\alpha_0) &= \left[\vartheta_n(\alpha) - \vartheta_n(\alpha_0)\right] + T_{n,n-1}^{-1}(\alpha_0) \sum_{k=1}^{n} \Psi_{n,k-1}(\alpha_0)\,\varepsilon_k \\ &\quad + \; c\,(n^{-\alpha} - n^{-\alpha_0})\,T_{n,n-1}^{-1}(\alpha_0) \sum_{k=1}^{n} \Psi_{n,k-1}(\alpha_0)\,\dot{\Phi}_{n,k-1}^{\top}\,\vartheta_n(\alpha) \\ &\triangleq \tilde{R}_{n,1} + \tilde{R}_{n,2} + \tilde{R}_{n,3}.\end{aligned} \tag{2.5.21}$$

The reasoning we have just developed also applies to $\tilde{R}_{n,2}$ *via* the same lines and tools, and as a result we obtain the convergence

$$\sqrt{n}\,C_n^{\frac{1}{2}}(\alpha)\,\tilde{R}_{n,2} \xrightarrow{d} \mathcal{N}\big(0,\, \sigma^2\,\Upsilon_p^{-1}\big). \tag{2.5.22}$$

That leads to the first intermediate step,

$$\left|e_p^{\top}\sqrt{n}\,C_n^{\frac{1}{2}}(\alpha_0)\,\tilde{R}_{n,2}\right| \;=\; \left|\sqrt{c\,n^{\alpha_0-\alpha}}\,e_p^{\top}\,\sqrt{n}\,C_n^{\frac{1}{2}}(\alpha)\,\tilde{R}_{n,2}\right| \;=\; o_p(1) \tag{2.5.23}$$

as soon as $\alpha_0 < \alpha$. We can consider the rest of the decomposition similarly, that is

$$\left|e_p^{\top}\sqrt{n}\,C_n^{\frac{1}{2}}(\alpha_0)\left(\tilde{R}_{n,1} + \tilde{R}_{n,3}\right)\right| = \left|\sqrt{c\,n^{1-\alpha_0}}\,(1 - n^{\alpha_0-\alpha})\right| |1 - u_n| \tag{2.5.24}$$

where the first factor is clearly divergent when $\alpha_0 < \alpha$ and

$$u_n \;\triangleq\; e_p^{\top}\,T_{n,n-1}^{-1}(\alpha_0)\left[\sum_{k=1}^{n} \Psi_{n,k-1}(\alpha_0)\,\dot{\Phi}_{n,k-1}^{\top}\right]\vartheta_n(\alpha).$$

As in the proof of Lemma 2.5.3, it is easy to see that $\Psi_{n,k-1}(\alpha_0)\,\dot{\Phi}_{n,k-1}^{\top}$ is a $p \times p$ matrix made of a first column of zeros, terms of the form $X_{n,k-1}\,X_{n,k-j}$ on the first line and

$V_{n,k-i}(\alpha_0)\,X_{n,k-j}$ in the bottom-right block ($1 \leqslant i \leqslant p-1$, $2 \leqslant j \leqslant p$), broken down as

$$\begin{aligned}\rho_n^{2j}(\alpha)\sum_{k=1}^{n} V_{n,k-i}(\alpha_0)\,X_{n,k-j} &= \rho_n^{j+1}(\alpha)\sum_{k=1}^{n} V_{n,k-i}(\alpha_0)\,X_{n,k-1} \\ &\quad - \sum_{s=1}^{j-1}\rho_n^{2j-s}(\alpha)\sum_{k=1}^{n} V_{n,k-i}(\alpha_0)\,V_{n,k-j+s}(\alpha)\end{aligned}$$

as it is done in (2.5.10). As soon as $\alpha/2 < \alpha_0 < \alpha$, Lemma 2.5.3 entails that the first part of this equation is $o_p(n^{1+\alpha/2})$ and, using Cauchy–Schwarz inequality,

$$\left|\sum_{k=1}^{n} V_{n,k-i}(\alpha_0)\,V_{n,k-j+l}(\alpha)\right| \leqslant \sqrt{\sum_{k=1}^{n} V_{n,k-i}^2(\alpha_0)}\sqrt{\sum_{k=1}^{n} V_{n,k-j+l}^2(\alpha)}$$

where both sums are $O_p(n)$. Finally, it is straightforward to obtain the simplification

$$u_n \xrightarrow{\;p\;} e_p^\top\,\Upsilon_p^{-1}\,\sigma^2\,\ell\begin{pmatrix}0 & 1 & \cdots & 1\\ \vdots & 0 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & 0\end{pmatrix}\begin{pmatrix}1\\ \vdots\\ \beta_i\\ \vdots\end{pmatrix} = \sum_{i=1}^{p-1}\beta_i \triangleq s_\beta \tag{2.5.25}$$

where ℓ and Υ are defined in Lemmas 2.5.2 and 2.5.3, respectively. Note that $1 - s_\beta$ is obviously non-zero since by stability, the polynomial $\Theta_\beta(z) = 1 - \beta_1 z - \ldots - \beta_{p-1}z^{p-1}$ has no zero on the unit circle. Going back to (2.5.23) and (2.5.24), we then have, for any $\alpha/2 < \alpha_0 < \alpha$,

$$\left|e_p^\top\sqrt{n}\,C_n^{\frac{1}{2}}(\alpha_0)\left(\widehat{\vartheta}_n(\alpha_0) - \vartheta_n(\alpha_0)\right)\right| \xrightarrow{\;p\;} +\infty \tag{2.5.26}$$

which concludes the proof. □

2.5.3 Proof of Proposition 2.2.1.

First, note that

$$\begin{aligned}\sqrt{n}\left(\widetilde{\theta}_n - \theta_n\right) &= \sqrt{n}\left(\widehat{J}_n - J_n\right)\left(\widehat{\beta}_n - \beta_n\right) + \sqrt{n}\left(\widehat{J}_n - J_n\right)\beta_n \\ &\quad + \sqrt{n}\,J_n\left(\widehat{\beta}_n - \beta_n\right) + \sqrt{n}\left(\widehat{v}_n - \lambda_{n,1}\right)e_p.\end{aligned}$$

In fact, the proof rests entirely on Lemma 2.5.4. The first line of the asymptotic normality gives $|\widehat{v}_n - \lambda_{n,1}| = O_p(n^{-(1+\alpha)/2}) = o_p(n^{-1/2})$ and obviously $\|\widehat{J}_n - J_n\| = o_p(n^{-1/2})$. Now from

the remaining part of (2.5.13),

$$\sqrt{n}\left(\widehat{\beta}_n - \beta_n\right) \xrightarrow{d} \mathcal{N}\left(0, \Sigma_{p-1}^{-1}\right)$$

where the (invertible) covariance Σ_{p-1} is given in (2.2.3). So $\|\widehat{J}_n - J_n\| \, \|\widehat{\beta}_n - \beta_n\| = o_p(n^{-1/2})$ and the result follows from (A1), that is from $J_n \to J$. □

2.5.4 Proof of Theorem 2.2.2.

Taking the first line of the convergence in Lemma 2.5.4,

$$\sqrt{n^{1+\alpha}}\left(\widehat{v}_n(\alpha) - \rho_n(\alpha)\right) \xrightarrow{d} \mathcal{N}\left(0, c\,\ell^{-1}\right) \tag{2.5.27}$$

where we recall that $\widehat{v}_n(\alpha)$ is the first component of $\widehat{\vartheta}_n(\alpha)$, $\rho_n(\alpha) = 1 - c\,n^{-\alpha}$ and $\ell > 0$ is given in Lemma 2.5.2. Now define the sequence of $\mathcal{C}^1$-mappings from $]0,1[$ to $\mathbb{R}$ given by

$$\forall\, n > 1, \quad g_n(x) = \frac{\ln c - \ln(1-x)}{\ln n}$$

so that $\alpha = g_n(\rho_n(\alpha))$ and $\widehat{\alpha}_n(\alpha) = g_n(\widehat{v}_n(\alpha))$. Using Taylor's expansion, there is a sequence (ϑ_n^*) such that

$$\widehat{\alpha}_n(\alpha) = \alpha + \left(\widehat{v}_n(\alpha) - \rho_n(\alpha)\right) g_n'(\vartheta_n^*)$$

with, for all $n > 1$, $(\rho_n(\alpha) \wedge \widehat{v}_n(\alpha)) < \vartheta_n^* < (\rho_n(\alpha) \vee \widehat{v}_n(\alpha))$. So,

$$(\ln n)\sqrt{n^{1-\alpha}}\,\left(\widehat{\alpha}_n(\alpha) - \alpha\right) = \sqrt{n^{1+\alpha}}\left(\widehat{v}_n(\alpha) - \rho_n(\alpha)\right)\frac{n^{-\alpha}}{1-\vartheta_n^*}. \tag{2.5.28}$$

We also note that

$$n^\alpha\left(1 - \rho_n(\alpha) \vee \widehat{v}_n(\alpha)\right) \;<\; n^\alpha\left(1 - \vartheta_n^*\right) \;<\; n^\alpha\left(1 - \rho_n(\alpha) \wedge \widehat{v}_n(\alpha)\right)$$

and

$$\begin{aligned} n^\alpha\left(1 - \widehat{v}_n(\alpha)\right) &= n^\alpha\left(1 - \rho_n(\alpha)\right) + n^\alpha\left(\rho_n(\alpha) - \widehat{v}_n(\alpha)\right) \\ &= c + O_p\left(\sqrt{n^{\alpha-1}}\right) \xrightarrow{p} c \end{aligned}$$

from Lemma 2.5.4 and since $\alpha < 1$. That clearly implies that

$$n^{\alpha}\left(1-\vartheta_n^*\right) \xrightarrow{p} c. \tag{2.5.29}$$

The result follows from (2.5.27), (2.5.28) and (2.5.29) together with Slutsky's lemma. □

2.5.5 Proof of Theorem 2.2.3.

First, a direct computation of the entries of P^{-1} shows that

$$\pi_{11} = \frac{1}{\prod_{j=2}^{p}(1-\lambda_j)} \quad (= 1 \text{ if } p = 1),$$

see (2.2.16) with $\lambda_1 = 1$. We know by Rem. 2.2 in Badreau and Proïa (2023) that the OLS of θ_n defined in (2.2.9) is consistent for θ, so obviously

$$\widehat{A}_n \xrightarrow{p} A$$

also holds. Since the eigenvalues of a square matrix are continuous functions of its entries, see e.g. Thm. 2.4.9.2 of Horn and Johnson (2012), it follows by the continuous mapping theorem that

$$\widehat{\pi}_n \xrightarrow{p} \pi_{11}.$$

Then, taking for α_0 the 'true' value α and using Slutsky's lemma, (2.2.18) is a straightforward consequence of Theorem 2.2.2. Now suppose that $\alpha/2 < \alpha_0 < \alpha$. Coming back to (2.5.28) with α_0, through the reasoning of the previous proof, there is a sequence (ϑ_n^{**}) such that

$$(\ln n)\sqrt{n^{1-\alpha_0}}\,\left(\widehat{\alpha}_n(\alpha_0)-\alpha_0\right) = \sqrt{n^{1+\alpha_0}}\left(\widehat{v}_n(\alpha_0)-\rho_n(\alpha_0)\right)\frac{n^{-\alpha_0}}{1-\vartheta_n^{**}} \tag{2.5.30}$$

where $(\rho_n(\alpha_0)\wedge\widehat{v}_n(\alpha_0)) < \vartheta_n^{**} < (\rho_n(\alpha_0)\vee\widehat{v}_n(\alpha_0))$. Using again the decomposition (2.5.21), we first obtain with (2.5.22),

$$\left|n^{\alpha_0}\, e_p^{\top}\tilde{R}_{n,2}\right| = O_p\Big(\sqrt{n^{2\alpha_0-1-\alpha}}\Big) \xrightarrow{p} 0$$

since $2\alpha_0 < \alpha + 1$. Then,

$$n^{\alpha_0}\, e_p^{\top}\left(\tilde{R}_{n,1}+\tilde{R}_{n,3}\right) = c\left(1-n^{\alpha_0-\alpha}\right)(1-u_n) \xrightarrow{p} c\left(1-s_\beta\right)$$

where u_n has been introduced in (2.5.24) and its limit behavior established in (2.5.25). Hence,

$$n^{\alpha_0}\left(1-\widehat{v}_n(\alpha_0)\right) = c + n^{\alpha_0}\left(\rho_n(\alpha_0)-\widehat{v}_n(\alpha_0)\right) \xrightarrow{p} c\, s_\beta.$$

This is sufficient to get that, for any $\alpha/2 < \alpha_0 < \alpha$,

$$\liminf_{n\to+\infty} \frac{n^{-\alpha_0}}{|1-\vartheta_n^{**}|} \geqslant \frac{1}{c\,(1\vee|s_\beta|)} > 0. \tag{2.5.31}$$

Finally, (2.5.26) entails

$$\left|\sqrt{n^{1+\alpha_0}}\left(\widehat{v}_n(\alpha_0)-\rho_n(\alpha_0)\right)\right| \xrightarrow{p} +\infty. \tag{2.5.32}$$

The combination of (2.5.30), (2.5.31) and (2.5.32) leads to (2.2.19), which concludes the proof. □

2.5.6 The case of a negative unit root

Our results have all been established for $\lambda_1 = 1$ but they remain true in a context of a negative unit root, at the cost of some slight adjustments. In what follows, it is thus implied that $\lambda_1 = -1$ so that we may now consider, $\lambda_{n,1} = \lambda_1\,\rho_n(\alpha) = -\rho_n(\alpha)$. The immediate consequence is that, in the definition of the latent variables $V_{n,k}$ in (2.2.2), $V_{n,k}(\alpha) = X_{n,k} + \rho_n(\alpha)\,X_{n,k-1}$. So, in the proof of Lemma 2.5.3, $\rho_n(\alpha)$ must be replaced by $-\rho_n(\alpha)$ in (2.5.4) and (2.5.10), and in particular that leads to

$$\begin{array}{rcl} C_n^{\frac{1}{2}}(\alpha)\,P_n^\top\,(I_p+\rho_n(\alpha)A_n^\top) & = & C_n^{\frac{1}{2}}(\alpha)\,(I_p+\rho_n(\alpha)D_n)\,P_n^\top \\ & \xrightarrow{p} & \operatorname{diag}\left(0,(1+\lambda_i)_{2\leqslant i\leqslant p}\right)P^\top \triangleq \Delta_p\,P^\top \end{array}$$

in place of (2.5.5). We may also note that c merely becomes $-c$ in (2.5.8), (2.5.9) and (2.5.21). The sequence (u_n) defined in (2.5.24) now satisfies

$$u_n \xrightarrow{p} e_p^\top\,\Upsilon_p^{-1}\,\sigma^2\,\ell \begin{pmatrix} 0 & -1 & \cdots & (-1)^{p-1} \\ \vdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} -1 \\ \vdots \\ \beta_i \\ \vdots \end{pmatrix} = \sum_{i=1}^{p-1}(-1)^i\beta_i \triangleq s_\beta$$

where $1 - s_\beta = \Theta_\beta(-1) \neq 0$. It is interesting to note that, for $\lambda_1 = \pm 1$, we have in general terms

$$1 - u_n \xrightarrow{p} \Theta_\beta(\lambda_1)$$

where $\Theta_\beta(z) = 1 - \beta_1 z - \ldots - \beta_{p-1} z^{p-1}$ has no zero on the unit circle, by stability: the sign of the asymptotic behavior of Z_n under $\mathcal{H}_1$ thus depends on the sign of $\Theta_\beta(\lambda_1) \neq 0$. Finally, by considering the $\mathcal{C}^1$-mappings from $]-1, 0[$ to $\mathbb{R}$ given by

$$\forall\, n > 1, \quad g_n(x) = \frac{\ln c - \ln(1+x)}{\ln n},$$

one can see that $\alpha = g_n(-\rho_n(\alpha))$ and $\widehat{\alpha}_n(\alpha) = g_n(\widehat{v}_n(\alpha))$ as defined in (2.2.14). Theorems 2.2.2 and 2.2.3 easily follow.

Acknowledgements. The authors warmly thank the Associate Editor and the two anonymous Reviewers for their suggestions and constructive comments which helped to improve this work. This research benefited from the support of the ANR project 'Efficient inference for large and high-frequency data' (ANR-21-CE40-0021).

References

Badreau, M. and Proïa, F. (2023), « Consistency and asymptotic normality in a class of nearly unstable processes », *in*: *Statistical Inference for Stochastic Processes*, DOI: 10.1007/s11203-023-09290-2.

Barczy, M., Ispány, M., and Pap, G. (2011), « Asymptotic behavior of unstable INAR(p) processes », *in*: *Stochastic Process. Appl.* 121, pp. 563–608.

Barreto-Souza, W. and Chan, N. H. (2024), « Nearly unstable integer-valued ARCH process and unit root testing », *in*: *Scand. J. Stat.* 51, pp. 402–424.

Bhargava, A. (1986), « On the theory of testing for unit roots in observed time series », *in*: *Rev. Econom. Stud.* 53, pp. 369–384.

Brockwell, P. J. and Davis, R. A. (1991), *Time Series: Theory and Methods*, 2nd ed., Springer-Verlag: New York.

Buchmann, B. and Chan, N. H. (2013), « Unified asymptotic theory for nearly unstable AR(p) processes », *in*: *Stochastic Processes and their Applications* 123, pp. 952–985.

Chan, N. H. and Wei, C. Z. (1987), « Asymptotic inference for nearly nonstationary AR(1) processes », *in*: *Annals of Statistics* 15, pp. 1050–1063.

Chan, N. H. and Wei, C. Z. (1988), « Limiting distributions of least squares estimates of unstable autoregressive processes », *in*: *Ann. Stat.* 16, pp. 367–401.

DeJong, D. N., Nankervis, J. C., Savin, N. E., and Whiteman, C. H. (1992), « Integration Versus Trend Stationarity in Time Series », *in*: *Econometrica* 60, pp. 423–433.

Dickey, D. A. and Fuller, W. A. (1979), « Distribution of the estimators for autoregressive time series with a unit root », *in*: *Journal of the American Statistical Association* 74(*366*), pp. 427–431.

Duflo, M. (1997), *Random Iterative Models*, Applications of Mathematics, Vol. 34, Springer-Verlag: Berlin.

Giraitis, L. and Phillips, P. C. B. (2006), « Uniform limit theory for stationary autoregression », *in*: *Journal of Time Series Analysis* 27, pp. 51–60.

Horn, R. A. and Johnson, C. R. (2012), *Matrix Analysis (Second Edition)*, Cambridge University Press: Cambridge, New-York.

Horváth, L. and Trapani, L. (2016), « Statistical inference in a random coefficient panel model », *in*: *Journal of Econometrics* 193, pp. 54–75.

Ispány, M., Pap, G., and Van Zuijlen, M. C. A. (2003), « Asymptotic inference for nearly unstable INAR(1) models », *in*: *J. Appl. Probab.* 40, pp. 750–765.

Ispány, M., Pap, G., and Van Zuijlen, M. C. A. (2005), « Fluctuation limit of branching processes with immigration and estimation of the means », *in*: *Adv. Appl. Probab.* 37, pp. 523–538.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. (1992), « Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We that Economic Time Series Have a Unit Root? », *in*: *Journal of Econometrics* 54, pp. 159–178.

Lai, T. L. and Wei, C. Z. (1983), « Asymptotic Properties of General Autoregressive Models and Strong Consistency of Least-Squares Estimates of Their Parameters », *in*: *Journal of Multivariate Analysis* 13, pp. 1–23.

Leybourne, S. J. and McCabe, B. P. M. (1994), « A consistent test for a unit root », *in*: *J. Bus. Econ. Stat.* 12, pp. 157–166.

MacKinnon, J. G. (1991), *Critical values for cointegration tests*, Long-Run Economic Relationships, ed. by R. F. Engle and C. W. Granger, Oxford University Press: Oxford.

Meer, T. Van der, Pap, G., and Van Zuijlen, M. C. A. (1999), « Asymptotic inference for nearly unstable AR(p) processes », *in*: *Econom. Theory* 15, pp. 184–217.

Nelson, C. R. and Plosser, C. I. (1982), « Trends and random walks in macroeconomic time series: Some evidence and implications », *in*: *J. Monet. Econ.* 10, pp. 139–162.

Ng, S. and Perron, P. (1995), « Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag », *in*: *J. Amer. Statist. Assoc.* 90, pp. 268–281.

Park, J. W. (2003), *Weak unit roots*, Department of Economics, Rice University.

Perron, P. (1988), « Trends and random walks in macroeconomic time series: Further evidence from a new approach », *in*: *J. Econom. Dynam. Control* 12, pp. 297–332.

Phillips, P. C. B. (1987a), « Time Series Regression with a Unit Root », *in*: *Econometrica* 55, pp. 277–301.

Phillips, P. C. B. (1987b), « Towards a unified asymptotic theory for autoregression », *in*: *Biometrika* 74, pp. 535–547.

Phillips, P. C. B. and Lee, J. H. (2015), « Limit theory for VARs with mixed roots near unity », *in*: *Econometric Review* 34, pp. 1034–1055.

Phillips, P. C. B. and Magdalinos, T. (2007), « Limit theory for moderate deviations from a unit root », *in*: *Journal of Econometrics* 136, pp. 115–130.

Phillips, P. C. B. and Perron, P. (1988), « Testing for a Unit Root in Time Series Regression », *in*: *Biometrika* 75, pp. 335–346.

Pollard, D. (1984), *Convergence of Stochastic Processes*, Springer: Berlin.

Proïa, F. (2020), « Moderate deviations in a class of stable but nearly unstable processes », *in*: *Journal of Statistical Planning and Inference* 208, pp. 66–81.

Said, E. S. and Dickey, D. A. (1981), « Testing ARIMA(p,1,q) versus ARMA(p+1,q) », *in*: *Proc. Bus. Econ. Statist. Sect., Am. Statist. Assoc.*, pp. 318–322.

Said, E. S. and Dickey, D. A. (1984), « Testing for Unit Roots in Autoregressive Moving Average Models of Unknown Order », *in*: *Biometrika* 71(*3*), pp. 599–607.

Schmidt, P. and Phillips, P. (1992), « LM test for a unit root in the presence of deterministic trends », *in*: *Oxford Bull. Econ. Stat.* 54, pp. 257–287.

Schwert, G. (1989), « Tests for unit roots: a Monte Carlo investigation », *in*: *J. Bus. Econom. Statist.* 7, pp. 147–160.

White, J. S. (1958), « The limiting distribution of the serial correlation coefficient in the explosive case », *in*: *Ann. Math. Statist.* 29, pp. 1188–1197.

Chapter 3

Model selection in nonstationary weak FARIMA models

Abstract: We investigate order selection for Fractional AutoRegressive Integrated Moving Average (FARIMA) models driven by uncorrelated innovations that may still exhibit temporal dependence through an α-mixing structure. This framework goes beyond the classical setting of independent or martingale difference innovations and covers a broader class of weak FARIMA models, including situations in which nonlinear dependence may be present despite the absence of serial correlation. Our contribution is primarily theoretical. Rather than introducing a new penalization term, we establish that standard BIC-type model selection procedures remain valid in this substantially more general long-memory framework. More precisely, we study the asymptotic behavior of the Bayesian Information Criterion (BIC) and of a quasi-Bayesian Information Criterion (QBIC) built from quasi-maximum likelihood estimation, and we prove the consistency of both criteria for selecting the autoregressive and moving average orders under mild regularity assumptions. These results extend existing consistency theorems available for FARIMA models with independent or martingale difference innovations and, to the best of our knowledge, provide the first theoretical justification for BIC-type order selection in long-memory models driven by uncorrelated yet dependent noise.

Keywords and phrases: FARIMA models; Nonlinear processes; Order Selection; BIC; Consistency.

Mathematics Subject Classification: 62M10, 62F07.

3.1 Introduction and Motivation

Long-memory processes arise in many fields where dependence persists over long horizons, including hydrology, economics, finance, and environmental sciences. Their defining feature is a hyperbolic decay of autocovariances, which cannot be captured by standard

short-memory models. Fractional AutoRegressive Integrated Moving Average (FARIMA) models, introduced independently by Granger and Joyeux (1980) and Hosking (1981), provide a flexible and widely used parametric framework for representing such long-range dependence through a fractional differencing parameter.

Most of the asymptotic theory for FARIMA models has been developed under strong or semi-strong assumptions on the innovation process, typically requiring independence or a martingale difference structure; see, among others, Fox and Taqqu (1986), Dahlhaus (1989), Beran (1995), Hualde and Robinson (2011), and Huang et al. (2022). Although mathematically convenient, these assumptions exclude many practically relevant situations in which the innovations are uncorrelated but still exhibit nonlinear or higher-order dependence. This phenomenon naturally appears, for example, in time series with conditional heteroskedasticity or in nonlinear transformations of linear processes; see Francq and Zakoïan (1998), Francq et al. (2005), and Francq and Zakoïan (2005). Such situations motivate the study of weak FARIMA models, recently considered in Boubacar Maïnassara et al. (2021); Boubacar Maïnassara et al. (2023); Ben Hariz et al. (2023), where the innovations are only required to be uncorrelated and may otherwise display substantial temporal dependence.

The present paper focuses on order selection for such weak FARIMA processes. More precisely, we consider the problem of selecting the autoregressive and moving average orders when the innovations are uncorrelated but not necessarily independent. This question is important both methodologically and theoretically. Methodologically, order misspecification may lead to unstable inference, loss of efficiency, or residual dependence in fitted models. Theoretically, once one departs from the classical independent or martingale difference framework, it is no longer clear whether standard information criteria retain their usual consistency properties. In the weak FARIMA setting, this issue is especially delicate because long-range dependence in the signal interacts with nonlinear dependence in the innovations.

Among model selection tools, penalized likelihood criteria play a central role. In particular, the Bayesian Information Criterion (BIC) is well known for its consistency in many parametric settings, whereas the Akaike Information Criterion is not designed for consistent order recovery. For short-memory models, consistency results for BIC-type criteria are now available in a wide variety of frameworks, including AR, ARMA, VARMA, and more general affine causal processes; see Hannan and Rissanen (1982), Boubacar Maïnassara (2011), Sin and Yu (2019), Bardet et al. (2020), Kamila (2022), Kengne (2022),

and Bardet et al. (2023). In the long-memory case, however, the available theory remains much more limited. Existing results essentially concern strong or semi-strong FARIMA models. For instance, Beran et al. (1998) proved the consistency of BIC for fractional autoregressive models without a moving average component, while Huang et al. (2022) obtained consistency results for strong and semi-strong FARIMA models. In a related direction, Baillie et al. (2014) proposed a modified information criterion for selecting the short-memory component of FARIMA models with martingale difference noise.

Our contribution is primarily theoretical. We do not introduce a new penalty term, nor do we claim that BIC-type procedures systematically outperform alternative criteria in finite samples. Instead, the purpose of this paper is to show that standard BIC-type order selection procedures remain asymptotically valid in a substantially broader long-memory framework than previously covered in the literature. More specifically, we establish the consistency of the BIC for selecting the autoregressive and moving average orders of weak FARIMA models under mild regularity assumptions, and we also prove the consistency of a quasi-Bayesian Information Criterion (QBIC) based on quasi-maximum likelihood estimation. The interest of these results lies precisely in their scope: the innovation process is only assumed to be uncorrelated and to satisfy an α-mixing condition, allowing for general forms of temporal dependence beyond independence and martingale difference assumptions.

To the best of our knowledge, this provides the first rigorous justification of BIC-type order selection for FARIMA models driven by uncorrelated yet dependent innovations. In this sense, the paper extends to the long-memory setting a line of research that is by now well developed for weak short-memory models, and helps bridge the gap between the weak ARMA literature and the theory of fractional processes. The α-mixing assumption is used here as a tractable and widely adopted dependence condition ensuring the asymptotic validity of the selection procedure over a broad class of weak innovations, rather than as a feature tailored to a particular empirical dataset.

The rest of the paper is organized as follows. Section 1 introduces the general framework, notation, and assumptions, and establishes the asymptotic normality of the conditional sum-of-squares estimator in a general setting. Section 2 is devoted to the study of BIC-type criteria. Section 3 reports simulation results and an empirical illustration.

3.2 General framework

3.2.1 The Data Generating Process

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which the innovation process $(\epsilon_t)_{t\in\mathbb{Z}}$ is defined. Fix $p_0, q_0 \in \mathbb{N}$ and $d_0 \in \mathbb{R}$. Consider a centered, second-order stationary process $(X_t)_{t\in\mathbb{Z}}$ admitting a FARIMA(p_0, d_0, q_0) representation

$$\begin{aligned} X_t &= (1-L)^{-d_0}\{Y_t \mathbb{1}_{\{t\geqslant 1\}}\}, && t \in \mathbb{Z}, \\ a_{\xi_0}(L)\, Y_t &= b_{\xi_0}(L)\, \epsilon_t, && t \in \mathbb{Z}, \end{aligned} \tag{3.2.1}$$

where L denotes the back-shift operator and $\mathbb{1}_A$ is the indicator function that is equal to 1 on the measurable set A and 0 otherwise. The autoregressive and moving-average polynomials are defined by $a_{\xi_0}(z) = 1 - \sum_{i=1}^{p_0} a_{0,i} z^i$ and $b_{\xi_0}(z) = 1 - \sum_{j=1}^{q_0} b_{0,j} z^j$, with parameter vector $\xi_0 = (a_{0,1}, \ldots, a_{0,p_0}, b_{0,1}, \ldots, b_{0,q_0})^\top \in \mathbb{R}^{p_0+q_0}$.

Assumption 3.2.1. *We assume that a_{ξ_0} and b_{ξ_0} have no common zero and that all their roots lie outside the unit disk.*

Notice that Assumption 3.2.1 ensures the invertibility of the associated operators and identifiability of the parameter $\theta_0 = (\xi_0^\top, d_0)^\top$. In the model selection problem considered in this paper, the orders p_0 and q_0 are unknown and must be inferred from the data.

The parameter d_0 governs the long-memory behavior of the process through the fractional differencing operator $(1-L)^{-d_0}$, whose coefficients $\alpha_k(d_0)$, defined via the generalized binomial expansion $(1-L)^{-d_0} = \sum_{k\geq 0} \alpha_k(d_0) L^k$ with $\alpha_k(d) = \Gamma(k+d)\big/(\Gamma(k+1)\Gamma(d))$, satisfy the asymptotic relation $\alpha_k(d_0) \sim k^{d_0-1}/\Gamma(d_0)$ as $k \to \infty$, see Beran et al. (2013). This slow decay characterizes the long-range dependence of the process.

The innovation process $(\epsilon_t)_{t\in\mathbb{Z}}$ is assumed to be a weak white noise in the sense that

$$\mathbb{E}[\epsilon_t] = 0, \mathrm{Var}[\epsilon_t] = \sigma^2 > 0, \text{ and } \mathrm{Cov}[\epsilon_t, \epsilon_s] = 0 \text{ for all } t \neq s.$$

Under this assumption, Model (3.2.1) is referred to as a weak FARIMA model. This terminology contrasts with the strong FARIMA case, where the innovations are independent and identically distributed, and with the semi-strong FARIMA case, where $(\epsilon_t)_{t\in\mathbb{Z}}$ forms a martingale difference sequence. Accordingly,

$$\{\text{strong FARIMA}\} \subset \{\text{semi-strong FARIMA}\} \subset \{\text{weak FARIMA}\}.$$

As in Boubacar Maïnassara et al. (2021), we further assume that the innovation process satisfies the following condition:

Assumption 3.2.2. *The stochastic process* $(\epsilon_t)_{t\in\mathbb{Z}}$ *is strictly stationary and ergodic.*

3.2.2 Notation and Model Selection Setup

Let $\mathcal{D} = [d_L, d_U]$ be a compact parameter space for the memory parameter, with $-\infty < d_L < d_U < \infty$, and assume that the true memory parameter d_0 lies in the interior of $\mathcal{D}$. Let P and Q be fixed nonnegative integers representing prescribed upper bounds for the autoregressive and moving average orders, respectively.

We consider the problem of selecting the true order pair $m_0 = (p_0, q_0)$ from the finite collection of candidate models

$$\mathcal{M} := \{(p,q) \in \mathbb{N}^2 : \ 0 \le p \le P, \ 0 \le q \le Q\}.$$

Each element $m = (p,q) \in \mathcal{M}$ is referred to as a candidate model. For a given candidate model $m = (p,q)$, the associated FARIMA model is characterized by autoregressive and moving average polynomials a_{ξ_m} and b_{ξ_m} of respective orders p and q, with coefficient vector $\xi_m = (a_1, \ldots, a_p, b_1, \ldots, b_q)^\top$. Note that, for the sake of notational simplicity, we omit the dependence on the model whenever it is clear from the context.

To embed all candidate models in a common parameter space, we introduce a compact set $\Pi_M \subset \mathbb{R}^{P+Q}$, for the largest model $M = (P,Q) \in \mathcal{M}$, whose generic element is denoted by $\xi_M = (a_1, \ldots, a_P, b_1, \ldots, b_Q)^\top$. We assume that Π_M is chosen such that, for all $\xi_M \in \Pi_M$, the corresponding autoregressive and moving average polynomials have no zeros in the closed unit disk. This condition ensures stationarity and invertibility uniformly over the collection of candidate models.

For each candidate model $m = (p,q) \in \mathcal{M}$, the associated parameter space is defined as $\Theta_m = \Pi_m \times \mathcal{D}$, where Π_m denotes the set of admissible AR and MA coefficients of orders p and q. More precisely, Π_m is obtained by embedding the coefficient vector $\xi_m = (a_1, \ldots, a_p, b_1, \ldots, b_q)^\top$ into the common parameter space $\Pi_M \subset \mathbb{R}^{P+Q}$ by padding with zeros whenever $p < P$ or $q < Q$, *i.e.*

$$\Pi_m := \{\xi_m = (a_1, \ldots, a_p, b_1, \ldots, b_q)^\top | (a_1, \ldots, a_p, 0, \ldots, 0, b_1, \ldots, b_q, 0, \ldots, 0)^\top \in \Pi_M\}.$$

By convention, the autoregressive (respectively, moving-average) component is absent

when $p = 0$ (respectively, $q = 0$). Since Π_M is compact, it follows that Θ_m is a compact subset of $\mathbb{R}^{\kappa_m}$, where $\kappa_m = p+q+1$ denotes the dimension of the parameter vector, including the memory parameter. In particular, $\Pi_{(0,0)}$ reduces to the trivial set corresponding to the pure fractional noise model.

The data-generating process is assumed to arise from a unique true model $m_0 = (p_0, q_0)$ with parameter

$$\theta_0 = (a_{0,1}, \ldots, a_{0,p_0}, b_{0,1}, \ldots, b_{0,q_0}, d_0)^\top,$$

satisfying $a_{0,p_0} \neq 0$ and $b_{0,q_0} \neq 0$. Furthermore, we assume that θ_0 belongs to the interior of $\Theta_{m_0} = \Pi_{m_0} \times \mathcal{D}$.

3.2.3 Conditional-Sum-of-Squares Estimator and Its Asymptotic Properties

For each candidate model $m = (p, q) \in \mathcal{M}$ and each parameter $\theta_m = (\xi_m^\top, d)^\top \in \Theta_m$, the error process is recursively defined by

$$\epsilon_t(\theta_m) := \sum_{i=0}^{t-1} \phi_i(\theta_m)\, X_{t-i} \quad \text{for } t = 1, \ldots, n, \tag{3.2.2}$$

where the sequence of coefficients $(\phi_i(\theta_m))_{i=0}^{\infty}$ is the sequence of coefficients of the power series of the function $z \mapsto b_{\xi_m}^{-1}(z) a_{\xi_m}(z)(1-z)^d$. The Conditional-Sum-of-Squares (CSS) estimator for the model $m = (p, q)$ is then given by

$$\widehat{\theta}_{m,n} \in \underset{\theta_m \in \Theta_m}{\arg\min}\, \tilde{\sigma}_n^2(\theta_m), \quad \text{where} \quad \tilde{\sigma}_n^2(\theta_m) := \frac{1}{n} \sum_{t=1}^{n} \epsilon_t^2(\theta_m).$$

In the remainder of this section, we briefly recall existing results on the consistency and asymptotic normality of CSS estimators for weak FARIMA models when the true model m_0 is known, and we present their extension.

The consistency and asymptotic normality of the CSS estimator sequence $(\widehat{\theta}_{m_0,n})_{n\geq 1}$ for correctly specified stationary weak FARIMA models with $d_0 \in (-1/2, 1/2)$ were established in Boubacar Maïnassara et al. (2021). These results were obtained under the restriction that the parameter space $\Theta_{m_0} = \Pi_{m_0} \times \mathcal{D}$ satisfies $\mathcal{D} \subset (-1/2, 1/2)$ and $d - d_0 > -1/2$.

The contribution of the present work is to extend these results beyond the aforementioned restrictions. In particular, we allow for memory parameters d_0 lying in the nonstationary region $[1/2, \infty)$ as well as in the entire stationary region $(-\infty, 1/2)$, and we

relax the constraint on the parameter space by removing the condition $d - d_0 > -1/2$.

Before stating the generalized results, we introduce assumptions on a moment condition on the innovation process and the decay speed of the α-mixing coefficients of the innovation process $(\epsilon_t)_{t\in\mathbb{Z}}$ in (3.2.1) to validate the consistency and asymptotic normality of the sequence of CSS estimators. Remark that these assumptions are also required to validate the consistency of the model selection information criteria discussed in the next section. The sequence of α-mixing coefficients of the innovation process $(\epsilon_t)_{t\in\mathbb{Z}}$, denoted by $\{\alpha_\epsilon(h)\}_{h=0}^{\infty}$, is defined by

$$\alpha_\epsilon(h) := \sup_{A\in\mathcal{F}_{-\infty}^{t}, B\in\mathcal{F}_{t+h}^{\infty}} \left|\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)\right|,$$

where $\mathcal{F}_{-\infty}^{t} := \sigma(\epsilon_u; u \leqslant t)$ and $\mathcal{F}_{t+h}^{\infty} := \sigma(\epsilon_u; u \geqslant t+h)$. Then we impose an integrability condition on the moments of the innovation noise process and a geometric decay condition on the sequence $\{\alpha_\epsilon(h)\}_{h=0}^{\infty}$ stated below.

Assumption 3.2.3. *There exists a positive integer τ such that, for some $\nu \in (0,1]$, $\mathbb{E}[|\epsilon_t|^{\tau+\nu}] < \infty$ and $\sum_{h=0}^{\infty}(h+1)^{k-2}\{\alpha_\epsilon(h)\}^{\frac{\nu}{k+\nu}} < \infty$ hold for all $k \in \{1,\ldots,\tau\}$.*

Notice that Assumption 3.2.3 with $\tau \geqslant 2$ implies a useful condition on the absolute summability of the τ-th order joint cumulants of the innovation process $(\epsilon_t)_{t\in\mathbb{Z}}$ (see Doukhan and Léon (1989) for details):

- For the same $\tau \geqslant 2$ in Assumption 3.2.3, we have

$$\sum_{t_1,\ldots,t_{\tau-1}\in\mathbb{Z}} |\mathrm{cum}(\epsilon_0, \epsilon_{t_1}, \ldots, \epsilon_{t_{\tau-1}})| < \infty. \tag{3.2.3}$$

Moreover, for $\theta_m \in \Theta_m$, we introduce two matrices

$$I_m(\theta_m) := \lim_{n\to\infty} \mathrm{Var}\left[\sqrt{n}\,\frac{\partial}{\partial\theta_m}\tilde{\sigma}_n^2(\theta_m)\right] \tag{3.2.4}$$

and

$$J_m(\theta_m) := \lim_{n\to\infty} \frac{\partial^2}{\partial\theta_m\partial\theta_m^\top}\,\tilde{\sigma}_n^2(\theta_m) \quad \text{a.s.} \tag{3.2.5}$$

For notational simplicity, we abbreviate the dependence of the true model m_0 in the following, *i.e.* set $\widehat{\theta}_n := \widehat{\theta}_{m_0,n}$, $I(\theta_0) := I_{m_0}(\theta_0)$ and $J(\theta_0) := J_{m_0}(\theta_0)$.

Now we can state a main result in this section, whose proof is left for Section A.

Theorem 3.2.4. *Under Assumptions* 3.2.1*,* 3.2.2 *and* 3.2.3 *with* $\tau = 4$*, we have*

$$\mathcal{L}\left\{\sqrt{n}(\widehat{\theta}_n - \theta_0)\middle|\mathbb{P}_{\theta_0}\right\} \underset{n\to\infty}{\to} \mathcal{N}\left(0, J(\theta_0)^{-1} I(\theta_0) J(\theta_0)^{-1}\right).$$

Remark 3.2.1. In both strong, semi-strong and weak FARIMA models, the matrix $J(\theta_0)$ has the same analytical form (see Theorem 1 in Beran (1995) and Theorem 2.2 in Hualde and Robinson (2011)). However, the matrix $I(\theta_0)$, which captures the dependence structure of the innovations, is generally more intricate in the weak case. In particular, when the noise is i.i.d. or martingale difference, we have the simplification $I(\theta_0) = 2\sigma^2 J(\theta_0)$.

3.3 Bayesian Information Criterion

Recall that $\mathcal{M} = \{(p,q) \in \mathbb{N}^2 : \ 0 \le p \le P, \ 0 \le q \le Q\}$ denotes the set of candidate models and $m_0 = (p_0, q_0)$ denotes the true model. Let $\mathcal{M}_-$ and $\mathcal{M}_+$ be the following sets:

$$\begin{aligned}
\mathcal{M}_- :=& \{m = (p,q) \in \mathcal{M} : p < p_0 \text{ or } q < q_0\},\\
\mathcal{M}_+ :=& \mathcal{M} \setminus (\mathcal{M}_- \cup \{m_0\})\\
=& \{m = (p,q) \in \mathcal{M} : p \geqslant p_0 \text{ and } q > q_0\} \cup \{m = (p,q) \in \mathcal{M} : p > p_0 \text{ and } q \geqslant q_0\}.
\end{aligned}$$

Note that $\mathcal{M}_+$ contains overfitted candidate models and $\mathcal{M}_-$ refers to the set of underfitting and wrong candidate models in the sense that the model can be overfitted for one component (AR or MA) and underfitted for the other. Then, since $\kappa_m = p + q + 1$, it is clear that $\kappa_m > \kappa_{m_0}$ for any $m \in \mathcal{M}_+$ whereas no general statement can be made about κ_m for $m \in \mathcal{M}_-$.

3.3.1 Model Selection Procedure

Now we introduce information criteria for model selection to determine the orders of weak FARIMA models based on the pseudo-Gaussian likelihood function by the Bayesian Information Criterion (BIC) and its variant Quasi-BIC (QBIC) defined as follows:

$$\mathrm{BIC}_{m,n} := n \log \tilde{\sigma}_n^2(\widehat{\theta}_{m,n}) + \kappa_m \log n, \tag{3.3.1}$$

$$\mathrm{QBIC}^{\dagger}_{m,n} := n \log \tilde{\sigma}_n^2(\widehat{\theta}_{m,n}) + \log\det\left[n\, \partial^2_{\theta_m} \log \tilde{\sigma}_n^2(\widehat{\theta}_{m,n})\right]. \tag{3.3.2}$$

Unlike the penalty term of the BIC in (3.3.1), the penalty component of the QBIC in (3.3.2) may be ill-defined under finite-sample manipulations. This issue, which arises from the non–positive definiteness of the matrix $\partial^2_{\theta_m} \log \tilde{\sigma}^2_n(\widehat{\theta}_{m,n})$, can be circumvented by modifying the correction term in (3.3.2) as follows:

$$\mathrm{QBIC}_{m,n} := n \log \tilde{\sigma}^2_n(\widehat{\theta}_{m,n}) + \mathbb{1}_{A^c_{m,n}} \log\det\left[n\, \partial^2_{\theta_m} \log \tilde{\sigma}^2_n(\widehat{\theta}_{m,n})\right] + \mathbb{1}_{A_{m,n}}\, \kappa_m \log n, \quad (3.3.3)$$

where the set $A_{m,n}$ is defined by $A_{m,n} := \{\det[\partial^2_{\theta_m} \log \tilde{\sigma}^2_n(\widehat{\theta}_{m,n})] \leqslant 0\}$.

Then an optimal model $\widehat{m}_n$ is selected by

$$\widehat{m}_n \in \arg\min_{m \in \mathcal{M}} \mathrm{QBIC}_{m,n}.$$

3.3.2 Model Selection Consistency

Recall that $(\phi_i(\theta_m))_{i=0}^{\infty}$ is the sequence of coefficients of the power series of the function $z \mapsto b^{-1}_{\xi_m}(z) a_{\xi_m}(z)(1-z)^d$.

Let us introduce a restricted parameter space $\mathcal{D}_4(\iota)$ by

$$\mathcal{D}_4(\iota) := \{d \in \mathcal{D} : d - d_0 > -1/2 + \iota\} \quad \text{with } \iota \in (0, 1/2)$$

and set $\Theta_{m,4}(\iota) := \Pi_m \times \mathcal{D}_4(\iota)$. Moreover, the error process $(\epsilon_t(\theta_m))_{t \geqslant 1}$ can be written as

$$\epsilon_t(\theta_m) := \sum_{i=0}^{t-1} \phi_i(\theta_m) X_{t-i} = \sum_{i=0}^{t-1} \lambda_i(\theta_m) \epsilon_{t-i}$$

where the sequence of coefficients $(\lambda_i(\theta_m))_{i=0}^{\infty}$ is defined by the Cauchy products

$$\lambda_i(\theta_m) := \sum_{j=0}^{i} \phi_j(\theta_m) \psi_{i-j}(\theta_0),$$

with $(\psi_i(\theta_0))_{i=0}^{\infty}$ the coefficients of the power series $(1-z)^{-d_0} a^{-1}_{\xi_0}(z) b_{\xi_0}(z)$. Consider the non-truncated process $(\zeta_t(\theta_m))_{t \in \mathbb{Z}}$ defined by

$$\zeta_t(\theta_m) := \sum_{i=0}^{\infty} \lambda_i(\theta_m)\, \epsilon_{t-i}. \quad (3.3.4)$$

Note that, for each $\theta_m \in \Theta_{m,4}(\iota)$, the sequence $(\lambda_i(\theta_m))_{i=0}^{\infty}$ is square summable from (31) in Boubacar Maïnassara et al. (2021) so that the random process $(\zeta_t(\theta_m))_{t \in \mathbb{Z}}$ is well-defined

and second-order stationary. Its variance $\sigma^2(\theta_m) \coloneqq \mathrm{Var}_{\theta_0}[\zeta_1(\theta_m)]$ is then given by

$$\sigma^2(\theta_m) = \sum_{i,j=0}^{\infty} \lambda_i(\theta_m)\lambda_j(\theta_m)\mathbb{E}[\epsilon_{t-i}\,\epsilon_{t-j}] = \sigma^2 \sum_{i=0}^{\infty} \lambda_i^2(\theta_m) \tag{3.3.5}$$

since $(\epsilon_t)_{t\in\mathbb{Z}}$ is a white noise with mean 0 and variance σ^2.

For the true value $\theta_0 = (\xi_0^\top, d_0)^\top$ and each overfitting model $m = (p,q) \in \mathcal{M}_+$, *i.e.* either of the conditions 1) $p_0 \leqslant p$ and $q_0 < q$ or 2) $p_0 < p$ and $q_0 \leqslant q$ holds, let $\xi_{m,0}^\dagger$ denote the zero–padded vector extension of $\xi_0 = (a_{0,1}, \ldots, a_{0,p_0}, b_{0,1}, \ldots, b_{0,q_0})^\top$ into $\mathbb{R}^{p+q}$, that is

$$\xi_{m,0}^\dagger \coloneqq (a_{0,1}, \ldots, a_{0,p_0}, \underbrace{0, \ldots, 0}_{(p-p_0)\text{-times}}, b_{0,1}, \ldots, b_{0,q_0}, \underbrace{0, \ldots, 0}_{(q-q_0)\text{-times}})^\top \in \mathbb{R}^{p+q}. \tag{3.3.6}$$

We also write $\theta_{m,0}^\dagger \coloneqq ((\xi_{m,0}^\dagger)^\top, d_0)^\top$, which is an interior point of the parameter space Θ_m of the overfitting model m from the definition of Θ_m.

The function $\sigma^2(\theta_m)$ in (3.3.5) can be rewritten as

$$\sigma^2(\theta_m) = \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} \frac{S_{\theta_0}(\omega)}{S_{\theta_m}(\omega)}\,\mathrm{d}\omega, \text{ where } S_\theta(\omega) \coloneqq \frac{1}{2\pi}|1-e^{-i\omega}|^{-2d}\left|\frac{b_\xi(e^{-i\omega})}{a_\xi(e^{-i\omega})}\right|^2, \theta = (\xi^\top, d)^\top \in \Theta.$$

For any $m \in \mathcal{M}$ and $\theta_m \in \Theta_m$, we have $(2\pi)^{-1}\int_{-\pi}^{\pi} \log(S_{\theta_0}(\omega)/S_{\theta_m}(\omega))\,\mathrm{d}\omega = 0$ so that Jensen's inequality yields

$$\sigma^2(\theta_m) = \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} \frac{S_{\theta_0}(\omega)}{S_{\theta_m}(\omega)}\,\mathrm{d}\omega \geqslant \sigma^2, \quad \forall \theta_m \in \Theta_m, \quad \forall m \in \mathcal{M},$$

and the equality holds if and only if $S_{\theta_m} = S_{\theta_0}$ almost everywhere. This condition is satisfied when $\theta_m = (\xi_m^\top, d)^\top \in \Theta_m$ fulfills $d = d_0$ and

$$\frac{b_{\xi_m}(z)}{a_{\xi_m}(z)} = \frac{b_{\xi_0}(z)}{a_{\xi_0}(z)} \quad \text{for all } z \in \mathbb{C} \text{ with } |z| = 1. \tag{3.3.7}$$

Remark that the minimizers of the function $\sigma^2(\theta_m)$ over Θ_m might not be unique in general so that we impose the following assumption.

Assumption 3.3.1. *For each candidate model $m \in \mathcal{M}$, the function $\theta_m \mapsto \sigma^2(\theta_m)$ has a unique minimizer in the parameter space Θ_m, denoted by $\theta_{m,0}$. Moreover, we also assume that $\theta_{m,0}$ is an interior point of Θ_m.*

Obviously, the equality $\theta_{m_0,0} = \theta_0$ holds since $\theta_0 = (\xi_0^\top, d_0)^\top$ satisfies (3.3.7) so that θ_0 is a

minimizer of the function $\sigma^2(\theta_{m_0})$ on Θ_{m_0}. We make several comments on the uniqueness of minimizers given in Assumption 3.3.1 for the true model as well as overfitted/underfitted or wrong models separately.

- **True Model $\mathbf{m_0}$**: As mentioned earlier, the uniqueness of the minimizer $\theta_{m_0,0}$ is equivalent to the uniqueness of parameters satisfying (3.3.7), and this holds if and only if the rational function $b_{\xi_0}(z)/a_{\xi_0}(z)$ is irreducible when the candidate model is the true model. Hence, under Assumption 3.2.1, θ_0 is the unique parameter satisfying (3.3.7) and a unique minimizer of the function $\sigma^2(\theta_{m_0})$ on Θ_{m_0} satisfying $\sigma^2(\theta_0) = \sigma^2$.
- **Overfitted Models $\mathbf{m} \in \mathcal{M}_+$**: First notice that $\theta^\dagger_{m,0}$ always satisfies the equalities (3.3.7) and $\sigma^2(\theta^\dagger_{m,0}) = \sigma^2(\theta_0) = \sigma^2$ since the definition of $\theta^\dagger_{m,0}$ in (3.3.6) yields

$$a_{\xi^\dagger_{m,0}} = a_{\xi_0} \quad \text{and} \quad b_{\xi^\dagger_{m,0}} = b_{\xi_0}.$$

  Therefore, there exists at least one parameter $\theta_m \in \Theta_m$ satisfying (3.3.7) and, together with Assumption 3.3.1, we conclude that $\theta^\dagger_{m,0}$ is a unique minimizer of $\sigma^2(\theta_m)$ on Θ_m.
- **Underfitted or Wrong Models $\mathbf{m} \in \mathcal{M}_-$**: Recall that for the true model $m_0 = (p_0, q_0)$, we assumed that $a_{0,p_0} \neq 0$ and $b_{0,q_0} \neq 0$ hold. Moreover, as mentioned in the above comment for the true model, the rational function $b_{\xi_0}(z)/a_{\xi_0}(z)$ is irreducible under Assumption 3.3.1 so that, for models $m \in \mathcal{M}_-$, with either $p < p_0$ or $q < q_0$, there is no parameter θ_m in Θ_m satisfying (3.3.7), which implies $\sigma^2(\theta_{m,0}) > \sigma^2(\theta_0)$.

The consistency of the information criteria is established in the following sense.

Theorem 3.3.2. *Under Assumptions* 3.2.2, 3.2.3 *with* $\tau = 4$ *and* 3.3.1, *we have*

$$\lim_{n\to\infty} \mathbb{P}_{\theta_0}[\widehat{m}_n = m_0] = 1.$$

Remark 3.3.1. We give a comparison between the proof of Theorem 3.3.2 and the one of Theorem 5.1 of Eguchi and Masuda (2018). Even though their Theorem 5.1 provides a unified framework to verify the model selection consistency of the QBIC and BIC, their result is not directly applicable to prove Theorem 3.3.2. First, note that Assumptions 3.2 and 3.3 of Eguchi and Masuda (2018) are actually not required to establish the model selection consistency of the BIC/QBIC stated above. Moreover, the uniform convergences of the estimation functions on the whole parameter spaces stated in (5.1) of Eguchi and Masuda (2018), which is required as an assumption in their theorem, is not satisfied for

fractional time series like FARIMA processes in general. Thus we give a self-contained proof of the model selection consistency of the BIC/QBIC in Section B.

3.4 Simulations and Application to real data

3.4.1 Simulations

This simulation study is intended as an illustration of the theoretical results rather than as a comprehensive finite-sample comparison of competing model selection criteria. Its main purpose is to assess whether the consistency established above translates into satisfactory empirical behavior in weak FARIMA settings, and to evaluate how nonlinear dependence in the innovations affects order recovery in finite samples.

We assess the performance of the QBIC criterion for weak FARIMA model selection through simulations comparable to those conducted in Huang et al. (2022). Specifically, we generated $N = 1{,}000$ trajectories of lengths $n = 250$ and $n = 500$ from each of the following FARIMA processes:

(S1) FARIMA$(1, d_0, 0)$ with $a_{0,1} = -0.7$,

(S2) FARIMA$(1, d_0, 1)$ with $a_{0,1} = 0.8$ and $b_{0,1} = -0.5$,

(S3) FARIMA$(2, d_0, 2)$ with $a_{0,1} = 1.8$, $a_{0,2} = -0.9$, $b_{0,1} = 1.42$, and $b_{0,2} = -0.73$.

Here, $d_0 \in \{-0.5, -0.25, 0, 0.25, 0.5\}$. The innovation process $(\epsilon_t)_{t\in\mathbb{Z}}$ is defined by

$$\epsilon_t = \eta_t^2 \eta_{t-1}, \tag{3.4.1}$$

where $(\eta_t)_{t\in\mathbb{Z}}$ is a sequence of independent and identically distributed standard Gaussian random variables. This weak noise specification is chosen because it generates innovations that are uncorrelated while being neither independent nor martingale differences, thereby providing a simple setting in which the weak FARIMA framework genuinely differs from the classical strong and semi-strong noise settings. For Setup (S2), additional comparisons are performed under strong noise, that is, $\epsilon_t = \eta_t$, for $d_0 \in \{-0.5, 0, 0.5, 1, 1.5\}$, which also allows us to cover the nonstationary region.

In Huang et al. (2022), the authors introduced a refinement of the model selection procedure that remains consistent while substantially reducing the computational cost. We adopt the same algorithm here. For clarity, we denote the criterion by $\mathrm{QBIC}_{m,n}(p, q)$,

where p and q denote the AR and MA orders of model m, respectively. Let $P = Q = 5$. The procedure then consists of

— Computing

$$\widehat{R}_n = \arg\min_{0\leqslant r\leqslant 5} \mathrm{QBIC}_{m,n}(r,r).$$

— Estimating the AR order

$$\hat{p}_n = \arg\min_{0\leqslant p\leqslant \widehat{R}_n} \mathrm{QBIC}_{m,n}(p,\widehat{R}_n).$$

— Estimating the MA order

$$\hat{q}_n = \arg\min_{0\leqslant q\leqslant \widehat{R}_n} \mathrm{QBIC}_{m,n}(\widehat{R}_n,q).$$

The results are reported in Tables 3.1 and 3.2. The following indicators are used to summarize model selection performance:

$$\begin{aligned}
\text{(W) Wrong model selection} &= \mathbb{1}_{\{\hat{p}_n<p_0 \text{ or } \hat{q}_n<q_0\}},\\
\text{(T) True model selection} &= \mathbb{1}_{\{\hat{p}_n=p_0 \text{ and } \hat{q}_n=q_0\}},\\
\text{(O) Overfitted model selection} &= \mathbb{1}_{\{\hat{p}_n\geqslant p_0,\,\hat{q}_n\geqslant q_0\}} - \mathbb{1}_{\{\hat{p}_n=p_0 \text{ and } \hat{q}_n=q_0\}}.
\end{aligned}$$

Table 3.1 – Model selection results over 1,000 replications of FARIMA Processes (S1)–(S3), for different values of the memory parameter d_0 and sample size n, under the weak white noise process defined in (3.4.1).

d_0		−0.5		−0.25		0		0.25		0.5	
n		250	500	250	500	250	500	250	500	250	500
	W	4	4	1	0	5	1	3	0	3	0
(S1)	T	890	929	901	934	923	939	928	943	932	930
	O	106	67	98	66	72	60	69	57	65	70
	W	110	21	129	15	114	13	116	18	131	20
(S2)	T	882	971	866	976	878	980	877	977	867	977
	O	8	8	5	9	8	7	7	5	2	3
	W	6	0	5	2	8	3	6	0	2	3
(S3)	T	958	992	954	968	949	973	959	977	950	972
	O	36	8	41	30	43	24	35	23	48	25

The results reported in Tables 3.1 and 3.2 are in line with the theoretical consistency

results. First, the true model is selected with high frequency across all configurations, and the selection accuracy systematically improves with the sample size. In particular, for $n = 500$, the proportion of correct selections is very high in all experiments.

Second, the comparison between strong and weak noise in Table 3.2 shows that nonlinear dependence in the innovations makes finite-sample identification more challenging. This is expected, since weak dependence creates a more difficult environment for recovering the correct short-memory orders. Nevertheless, the deterioration remains moderate, and the improvement observed when passing from $n = 250$ to $n = 500$ is substantial, which is fully consistent with the asymptotic theory.

Overall, the simulation study should therefore be read as supporting the main message of the paper: BIC-type order selection remains practically informative beyond the classical independent-innovation framework, even though weak dependence may induce a non-negligible finite-sample cost.

Although no extensive comparison with AIC is reported here, it is worth emphasizing that the objective of the present paper is consistent order identification rather than predictive optimization. From this perspective, BIC-type criteria are the natural benchmark.

Table 3.2 – Model selection over 1,000 replications of the process in (S2) with memory parameter d_0 and sample size n induced by strong or weak white noise.

d_0		−0.5		0		0.5		1		1.5	
n		250	500	250	500	250	500	250	500	250	500
Strong noise	W	76	7	77	4	105	11	110	9	112	14
	T	918	990	916	994	894	986	889	990	885	984
	O	6	3	7	2	1	3	1	1	3	2
Weak noise	W	110	21	114	13	131	20	130	21	128	30
	T	882	971	878	980	867	977	869	979	870	970
	O	8	8	8	7	2	3	1	0	2	0

3.4.2 Application to real data

We now provide an empirical illustration of the weak FARIMA methodology using the mean daily discharge of the Seine River (see Figure 3.1). The purpose of this application is not to validate the dependence assumptions for this specific dataset, but rather to illustrate that the weak FARIMA framework may yield a coherent modelling and diagnostic strategy when standard strong-noise tools appear too restrictive.

The dataset covers the period from 1 January 2000 to 31 December 2020 on a daily basis, yielding a sample size of 7,671. Since the seminal work of Hurst (1951), it has been widely recognized that hydrological time series may exhibit long-memory behavior, thereby motivating the use of FARIMA models for such data. However, the original series is nonstationary (KPSS test p-value < 0.01) and displays clear seasonality with a period of approximately 365/366 days, that is, one year. A preprocessing step is therefore required before parameter estimation.

We consider the transformed series (X_t) obtained by applying a differencing operator of order 365 to the original data. The resulting series still exhibits long-memory features, as can be seen from the empirical autocorrelation function displayed in Figure 3.2, and from the empirical Hurst exponent, which is approximately 0.80. In addition, the differenced series appears to be stationary according to both the ADF and KPSS tests. We may therefore apply our model selection procedure to (X_t).

The model selected by QBIC is a FARIMA$(1, d, 1)$ process of the form

$$(1 - L)^d(X_t - aX_{t-1}) = \epsilon_t - b\epsilon_{t-1}.$$

The CSS estimator of the parameter $\theta = (a, b, d)^\top$ is

$$\hat{\theta}_n = (0.8760, -0.1436, 0.3586)^\top,$$

with associated standard deviations

$$(0.0265, 0.0350, 0.0566)^\top,$$

computed from Theorem 3.2.4. The estimated long-memory parameter is positive, which is consistent with the persistence suggested by the data.

To assess the adequacy of the fitted model, we applied Portmanteau tests to the residuals up to lag 30. The standard Box–Pierce and Ljung–Box tests, designed for strong FARIMA models, produced very small p-values at all lags, leading to systematic rejection of the fitted model. By contrast, the modified Portmanteau test proposed by Boubacar Maïnassara et al. (2023) for weak FARIMA models did not indicate significant residual autocorrelation. This contrast is informative: it suggests that, for this series, a modelling strategy based exclusively on strong-noise diagnostics may be misleading, whereas the weak FARIMA framework yields a coherent specification-and-diagnostics pipeline.

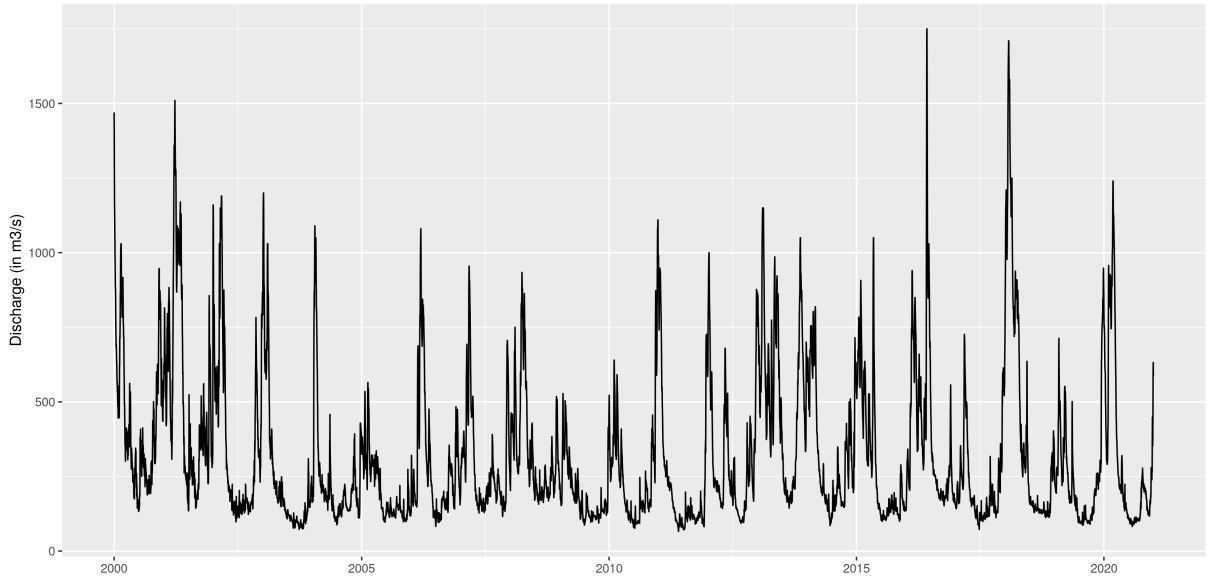


Figure 3.1 – Mean daily discharge of the Seine River.

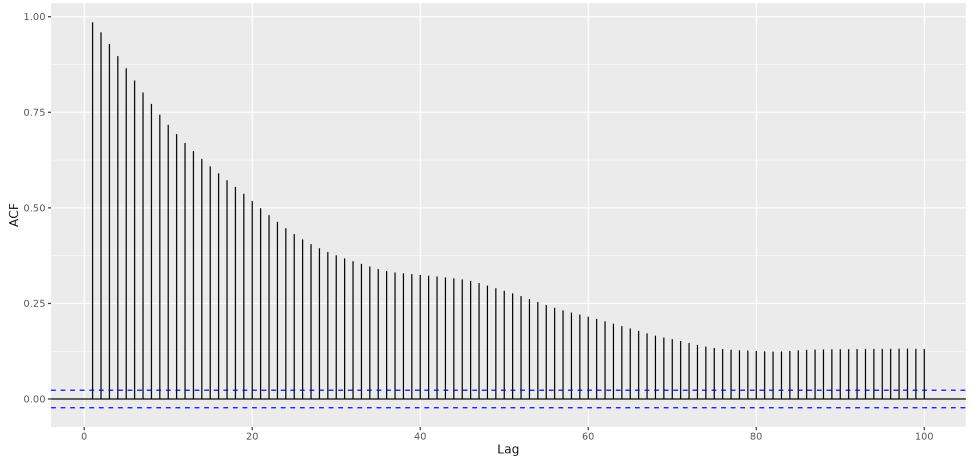


Figure 3.2 – Empirical autocorrelation function of (X_t).

3.5 Conclusion

This paper provides a theoretical contribution to model selection for FARIMA processes in a substantially broader setting than that usually considered in the literature. More precisely, we establish the consistency of BIC-type criteria for selecting the short-memory component of FARIMA models when the innovations are only assumed to be uncorrelated and may exhibit general forms of temporal dependence. Our framework also accommodates a wide range of values of the memory parameter, including nonstationary regimes.

The main contribution of the paper is therefore not the introduction of a new penalization scheme, but the justification of standard BIC-type procedures in weak FARIMA models, where independence or martingale difference assumptions are no longer imposed on the innovations. In this sense, our results extend previous consistency results such as those of Beran et al. (1998) and Huang et al. (2022), and help bridge the gap between the literature on weak short-memory models and the theory of long-memory processes.

The numerical experiments support this theoretical message. They show that the proposed criteria retain a high selection accuracy in finite samples, that weak dependence in the innovations makes order recovery more challenging, and that performance improves markedly with the sample size, in agreement with the asymptotic theory. The empirical illustration further suggests that the weak FARIMA framework can provide a coherent modelling strategy in situations where standard strong-noise diagnostics may be too restrictive.

Several extensions deserve further investigation. In particular, our analysis relies on identifiability assumptions, whereas Huang et al. (2022) obtain important results without imposing such conditions. Extending the present theory to non-identifiable weak FARIMA settings, especially in the presence of general nonlinear dependence in the innovations, would be a natural and challenging direction for future research. A broader finite-sample comparison with alternative criteria, including AIC-type procedures, would also usefully complement the asymptotic perspective developed in this paper.

Acknowledgements

This research benefited from the support of the ANR project ‘Efficient inference for large and high-frequency data’ (ANR-21-CE40-0021). The third co-author acknowledges

support from the 'Etoiles Montantes en Pays de la Loire' project, funded by the Pays de la Loire Region.

References

Baillie, R. T., Kapetanios, G., and Papailias, F. (2014), « Modified information criteria and selection of long memory time series models », *in*: *Computational Statistics & Data Analysis* 76, pp. 116–131.

Bardet, J. M., Kamila, K., and Kengne, W. (2020), « Consistent model selection criteria and goodness-of-fit test for common time series models », *in*: *Electronic Journal of Statistics* 14(*1*), pp. 2009–2052.

Bardet, J. M., Kare, K., and Kengne, W. (2023), « Efficient and consistent model selection procedures for time series », *in*: *Bernoulli* 29(*4*), pp. 2652–2690.

Ben Hariz, S., Brouste, A., Esstafa, Y., and Soltane, M. (2023), « Fast calibration of weak FARIMA models », *in*: *ESAIM: Probability and Statistics* 27, pp. 156–173.

Beran, J. (1995), « Maximum likelihood estimation of the differencing parameter for invertible short- and long-memory ARIMA models », *in*: *Journal of the Royal Statistical Society, Series B* 57(*4*), pp. 659–672.

Beran, J., Bhansali, R. J., and Ocker, D. (1998), « On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes », *in*: *Biometrika* 85(*4*), pp. 921–934.

Beran, J., Feng, Y., Ghosh, S., and Kulik, R. (2013), *Long-Memory Processes*, Probabilistic Properties and Statistical Methods, Springer, Heidelberg.

Billingsley, P. (1968), *Convergence of Probability Measures*, John Wiley: New York.

Boubacar Maïnassara, Y. (2011), « Selection of weak VARMA models by modified Akaike's information criteria », *in*: *Journal of Time Series Analysis* 33(*1*), pp. 121–130.

Boubacar Maïnassara, Y., Esstafa, Y., and Saussereau, B. (2021), « Estimating FARIMA models with uncorrelated but non-independent error terms », *in*: *Statistical Inference for Stochastic Processes* 24, pp. 549–608.

Boubacar Maïnassara, Y., Esstafa, Y., and Saussereau, B. (2023), « Diagnostic checking in FARIMA models with uncorrelated but non-independent error terms », *in*: *Electronic Journal of Statistics* 17(*1*), pp. 1160–1239.

Dahlhaus, R. (1989), « Efficient parameter estimation for self-similar processes », *in*: *Annals of Statistics* 17(*4*), pp. 1749–1766.

Doukhan, P. and Léon, J. R. (1989), « Cumulants for stationary mixing random sequences and applications to empirical spectral density », *in*: *Probability and Mathematical Statistics* 10(*1*), pp. 11–26.

Eguchi, S. and Masuda, H. (2018), « Schwarz type model comparison for LAQ models », *in*: *Bernoulli* 24(*3*), pp. 2278–2327.

Fox, R. and Taqqu, M. S. (1986), « Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series », *in*: *Annals of Statistics* 14(*3*), pp. 517–532.

Francq, C., Roy, R., and Zakoïan, J. M. (2005), « Diagnostic checking in ARMA models with uncorrelated errors », *in*: *Journal of Time Series Analysis* 26(*2*), pp. 251–272.

Francq, C. and Zakoïan, J. M. (1998), « Estimating linear representations of nonlinear processes », *in*: *Journal of Time Series Analysis* 19(*3*), pp. 253–273.

Francq, C. and Zakoïan, J. M. (2005), « Recent results for linear time series models with non independent innovations », *in*: *Statistical Modeling and Analysis for Complex Data Problems*, vol. 1, GERAD 25th Anniversary Series, Springer: New York, pp. 241–265.

Granger, C. W. J. and Joyeux, R. (1980), « An Introduction to Long-Memory Time Series Models and Fractional Differencing », *in*: *Journal of Time Series Analysis* 1, pp. 15–29.

Hallin, M., Taniguchi, M., Serroukh, A., and Choy, K. (1999), « Local asymptotic normality for regression models with long-memory disturbance », *in*: *Annals of Statistics* 27(*6*), pp. 2054–2080.

Hannan, E. J. and Rissanen, J. (1982), « Recursive estimation of mixed autoregressive-moving average order », *in*: *Biometrika* 69(*1*), pp. 81–94.

Hosking, J. R. M. (1981), « Fractional Differencing », *in*: *Biometrika* 68, pp. 165–176.

Hosoya, Y. (2005), « Fractional invariance principle », *in*: *Journal of Time Series Analysis* 26(*3*), pp. 463–486.

Hualde, J. and Robinson, P. M. (2011), « Gaussian pseudo-maximum likelihood estimation of fractional time series models », *in*: *Annals of Statistics* 39(*6*), pp. 3152–3181.

Huang, H.-H., Chan, N. H., Chen, K., and Ing, C.-K. (2022), « Consistent order selection for ARFIMA processes », *in*: *Annals of Statistics* 50(*3*), pp. 1297–1319.

Hurst, H. E. (1951), « Long-Term Storage Capacity of Reservoirs », *in*: *Transactions of the American Society of Civil Engineers* 116, pp. 770–799.

Kamila, K. (2022), « General Hannan and Quinn criterion for common time series », *in*: *Statistics* 56(*1*), pp. 222–241.

Kengne, W. (2022), « On consistency for time series model selection », *in*: *Statistical Inference for Stochastic Processes* 26(*2*), pp. 437–458.

McLeod, A. I. (1999), « Necessary and Sufficient Condition for Nonsingular Fisher Information Matrix in ARMA and Fractional ARIMA Models », *in*: *The American Statistician* 53(*1*), pp. 71–72.

Sin, C.-Y. and Yu, S.-H. (2019), « Order selection for possibly infinite-order non-stationary time series », *in*: *AStA Advances in Statistical Analysis* 103(*2*), pp. 187–216.

A Proof of Theorem 3.2.4

A.1 Preliminary Lemmas for Proof of Theorem 3.2.4

Lemma A.1. *Let $(\epsilon_t)_{t\in\mathbb{Z}}$ be a white noise process such that $\epsilon_t \in \mathbb{L}^{2+\nu}$ for some $\nu > 0$, and suppose that the process is α-mixing with mixing coefficients $\alpha_\epsilon(h)$ satisfying*

$$\alpha_\epsilon(h) \leqslant C_1 h^{-\gamma}, \quad \text{for some } \gamma > \frac{(2+c)(2+\nu)(2+\nu-\eta)}{\nu(2+\nu)-\eta(1+\nu)},$$

where $c > 0$, $\eta \in (0,\nu)$, and $C_1 > 0$ are constants. Then, there exists a constant $C_2 > 0$ such that for all $s > t$,

$$\operatorname{Var}\left(\mathbb{E}[\epsilon_s \mid \mathcal{F}_{-\infty}^t]\right) \leqslant C_2 (s-t)^{-2-c}.$$

Proof of Lemma A.1. Let $s > t$ and define the conditional expectation

$$Z := \mathbb{E}[\epsilon_s \mid \mathcal{F}_{-\infty}^t],$$

which is the orthogonal projection of ϵ_s onto the sigma-algebra $\mathcal{F}_{-\infty}^t$. Since $\mathbb{E}[\epsilon_s] = 0$, we have

$$\operatorname{Var}(Z) = \mathbb{E}[Z^2] = \operatorname{Cov}(Z, \epsilon_s).$$

To bound this covariance, we apply Davydov's inequality for α-mixing sequences: Let $X \in \mathbb{L}^p$ and $Y \in \mathbb{L}^q$ with $1 < p, q < \infty$ and $1/p + 1/q < 1$, then there exists a constant $K_0 > 0$ such that

$$|\operatorname{Cov}(X,Y)| \leqslant K_0 \|X\|_p \|Y\|_q \left[\alpha\{\sigma(X), \sigma(Y)\}\right]^{1-\frac{1}{p}-\frac{1}{q}}.$$

Fix $\nu > 0$, choose a small $\eta \in (0,\nu)$, then define

$$p := 2 + \nu - \eta, \quad q := 2 + \nu.$$

Since $\eta \in (0,\nu)$, we ensure that

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{2+\nu-\eta} + \frac{1}{2+\nu} < 1,$$

so the inequality is applicable.

Because $Z = \mathbb{E}[\epsilon_s \mid \mathcal{F}_{-\infty}^t]$ and $\epsilon_s \in \mathbb{L}^{2+\nu}$, we can use Jensen's inequality to obtain

$$\|Z\|_p \leqslant \|\epsilon_s\|_p \leqslant \|\epsilon_0\|_{2+\nu}, \quad \|\epsilon_s\|_q = \|\epsilon_0\|_{2+\nu}.$$

We now apply Davydov's inequality with the above values

$$\operatorname{Var}(Z) = |\operatorname{Cov}(Z, \epsilon_s)| \leqslant C\|\epsilon_0\|_p\|\epsilon_0\|_q\alpha_\epsilon(s-t)^{1-\frac{1}{p}-\frac{1}{q}}.$$

Therefore, there exists a constant $K_1 > 0$ such that

$$\operatorname{Var}(\mathbb{E}[\epsilon_s \mid \mathcal{F}_{-\infty}^t]) \leqslant K_1\alpha_\epsilon(s-t)^\theta,$$

where $\theta := 1 - 1/p - 1/q > 0$. Now, using the assumption $\alpha_\epsilon(h) \leqslant Ch^{-\gamma}$ for some $\gamma > \frac{2+c}{\theta}$, we get

$$\operatorname{Var}(\mathbb{E}[\epsilon_s \mid \mathcal{F}_{-\infty}^t]) \leqslant K_2(s-t)^{-\gamma\theta},$$

for a constant $K_2 > 0$. The condition on γ implies that

$$\gamma\theta > 2 + c,$$

and thus

$$\operatorname{Var}(\mathbb{E}[\epsilon_s \mid \mathcal{F}_{-\infty}^t]) \leqslant K_3(s-t)^{-2-c},$$

for some constant $K_3 > 0$, which concludes the proof. □

Lemma A.2. *Let $(\epsilon_t)_{t\in\mathbb{Z}}$ be a white noise process such that $\epsilon_t \in \mathbb{L}^{4+\nu}$ for some $\nu > 0$, and suppose that $(\epsilon_t)_{t\in\mathbb{Z}}$ is α-mixing with mixing coefficients satisfying*

$$\alpha_\epsilon(h) \leqslant C_1h^{-\gamma},$$

for some $\gamma > (1+2c)/\theta$, where $c > 0$,

$$\theta := \frac{\delta}{2+\delta}$$

for some fixed $\delta \in (0, \nu/2]$ and $C_1 > 0$. Then there exists a constant $C_2 > 0$ such that for all $\ell > t$, $m > t$,

$$\mathbb{E}\left[|\mathbb{E}[\epsilon_\ell\epsilon_m \mid \mathcal{F}_{-\infty}^t] - \mathbb{E}[\epsilon_\ell\epsilon_m]|\right] \leqslant C_2 \min(\ell - t, m - t)^{-1/2-c}.$$

Proof. Let $\ell > t$, $m > t$, and define

$$Z := \mathbb{E}[\epsilon_\ell\epsilon_m \mid \mathcal{F}_{-\infty}^t] - \mathbb{E}[\epsilon_\ell\epsilon_m].$$

By the Cauchy–Schwarz inequality,

$$\mathbb{E}[|Z|] \leqslant \big(\mathbb{E}[Z^2]\big)^{1/2} = \Big(\mathrm{Var}\big(\mathbb{E}[\epsilon_\ell \epsilon_m \mid \mathcal{F}_{-\infty}^t]\big)\Big)^{1/2}.$$

Let

$$X := \mathbb{E}[\epsilon_\ell \epsilon_m \mid \mathcal{F}_{-\infty}^t], \qquad Y := \epsilon_\ell \epsilon_m.$$

Since $X = \mathbb{E}[Y \mid \mathcal{F}_{-\infty}^t]$ is $\mathcal{F}_{-\infty}^t$-measurable, we have

$$\mathbb{E}[XY] = \mathbb{E}[X\,\mathbb{E}[Y \mid \mathcal{F}_{-\infty}^t]] = \mathbb{E}[X^2],$$

and $\mathbb{E}[X] = \mathbb{E}[Y]$. Hence,

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[Y]^2 = \mathrm{Cov}(X, Y).$$

We now apply Davydov's inequality. Fix $\delta \in (0, \nu/2]$ and set

$$p = q = 2 + \delta.$$

Then $1/p + 1/q < 1$ and

$$\theta := 1 - \frac{1}{p} - \frac{1}{q} = \frac{\delta}{2+\delta} > 0.$$

Since X is $\mathcal{F}_{-\infty}^t$-measurable and Y is measurable with respect to $\sigma(\epsilon_\ell, \epsilon_m) \subset \mathcal{F}_{t+h}^\infty$, where

$$h := \min(\ell - t, m - t),$$

Davydov's inequality yields

$$|\mathrm{Cov}(X, Y)| \leqslant K_0 \|X\|_p \|Y\|_q\, \alpha_\epsilon(h)^\theta.$$

By Jensen's inequality,

$$\|X\|_p = \|\mathbb{E}[Y \mid \mathcal{F}_{-\infty}^t]\|_p \leqslant \|Y\|_p.$$

Moreover, by Hölder's inequality,

$$\|Y\|_p = \|\epsilon_\ell \epsilon_m\|_p \leqslant \|\epsilon_\ell\|_{2p} \|\epsilon_m\|_{2p}.$$

Since $2p = 4 + 2\delta \leq 4 + \nu$, the assumption $\epsilon_t \in \mathbb{L}^{4+\nu}$ ensures that these norms are finite and uniformly bounded in t. The same holds for $\|Y\|_q$. Hence,

$$\|X\|_p \|Y\|_q \leqslant K_1,$$

for some constant $K_1 > 0$, and therefore

$$\mathrm{Var}(X) \leqslant K_2 \alpha_\epsilon(h)^\theta.$$

If $\alpha_\epsilon(h) \leq C_1 h^{-\gamma}$ with $\gamma > (1 + 2c)/\theta$, then

$$\mathrm{Var}(X) \leq K_3 h^{-1-2c}.$$

Taking square roots yields

$$\mathbb{E}[|Z|] \leq K_4 h^{-1/2-c} = K_4 \min(\ell - t, m - t)^{-1/2-c},$$

which completes the proof. □

A.2 Proof of Consistency

We only provide a sketch of the proof of consistency, since the asymptotic normality stated in Theorem 3.2.4 can be established in the same manner as Theorem 2 in Boubacar Maïnassara et al. (2021). First notice that we can assume $d_L < d_0 - \frac{1}{2}$ without loss of generality, since our proof trivially covers the case $d_L \geqslant d_0 - \frac{1}{2}$. We start from a standard argument. For $\delta > 0$, define $B_\delta(\theta_0) := \{\theta \in \Theta_{m_0} : \|\theta - \theta_0\| < \delta\}$. For small enough $\delta > 0$,

$$\mathbb{P}\Big[\|\widehat{\theta}_n - \theta_0\| \geqslant \delta\Big] \leqslant \mathbb{P}\Big[\inf_{\theta \in B_\delta(\theta_0)^c} \widetilde{\sigma}_n^2(\theta) \leqslant \widetilde{\sigma}_n^2(\theta_0)\Big] = \mathbb{P}\Big[\inf_{\theta \in B_\delta(\theta_0)^c} S_n(\theta) \leqslant 0\Big], \tag{A.1}$$

where $S_n(\theta) := \widetilde{\sigma}_n^2(\theta) - \widetilde{\sigma}_n^2(\theta_0)$ and $B_\delta(\theta_0)^c$ denotes the complement of $B_\delta(\theta_0)$. The remainder of the proof reflects the fact that $\widetilde{\sigma}_n^2(\theta)$, and thus $S_n(\theta)$, converges in probability to a well-behaved function when $d > d_0 - \frac{1}{2}$, and diverges when $d \leqslant d_0 - \frac{1}{2}$, while the need to establish uniform convergence, especially in a neighborhood of $d = d_0 - \frac{1}{2}$, requires additional special treatment. Consequently, for sufficiently small $\iota > 0$ such that $\iota < d_0 - \frac{1}{2} - d_L$,

we define the non-intersecting sets

$$\mathcal{D}_1(\iota) \coloneqq \{d \in \mathcal{D} : d_L \leqslant d \leqslant d_0 - \frac{1}{2} - \iota\}, \quad \mathcal{D}_2(\iota) \coloneqq \{d \in \mathcal{D} : d_0 - \frac{1}{2} - \iota < d < d_0 - \frac{1}{2}\},$$
$$\mathcal{D}_3(\iota) \coloneqq \{d \in \mathcal{D} : d_0 - \frac{1}{2} \leqslant d \leqslant d_0 - \frac{1}{2} + \iota\}, \quad \mathcal{D}_4(\iota) \coloneqq \{d \in \mathcal{D} : d_0 - \frac{1}{2} + \iota < d \leqslant d_U\}.$$

Correspondingly, define $\Theta_i(\iota) \coloneqq \Pi_{m_0} \times \mathcal{D}_i(\iota)$, for $i \in \{1,2,3,4\}$, so that $\Theta_{m_0} = \bigcup_{i=1}^{4} \Theta_i(\iota)$. Thus, from (A.1) it remains to prove

$$\lim_{n\to\infty} \mathbb{P}\Big[\inf_{\theta \in B_\delta(\theta_0)^c \cap \Theta_i(\iota)} S_n(\theta) \leqslant 0\Big] = 0, \quad \forall i \in \{1,2,3,4\}. \tag{A.2}$$

We prove (A.2) only in the case $i = 1$, since the cases $i = 2, 3$ can be established in the same way as (2.7) in Hualde and Robinson (2011) and the case $i = 4$ is implied by the following lemma, where $\sigma^2(\theta)$ is defined in 3.3.5.

Lemma A.3. *Under the assumptions in Theorem* 3.2.4*, the following uniform convergence holds for each* $\iota \in (0, 1/2)$*:*

$$\sup_{\theta \in \Theta_4(\iota)} |\tilde{\sigma}_n^2(\theta) - \sigma^2(\theta)| = o_{\mathbb{P}_{\theta_0}}(1) \quad \textit{as } n \to \infty.$$

Notice that Lemma A.3 is proved in Theorem 1 of Boubacar Maïnassara et al. (2021) only in the case $d_U < 1/2$. However, the same arguments remain valid when $d_U \geqslant 1/2$, including the non-stationary region, by following a similar line of reasoning to the proof of Theorem 2.1 of Hualde and Robinson (2011). Therefore, we omit the technical details.

Moreover, using similar arguments to those employed in the proof of (2.7) in Hualde and Robinson (2011), we can see that (A.2) with $i = 1$ follows once condition (2.41) is verified, that is, there exists some constant $c > 0$ such that

$$\lim_{n\to\infty} \mathbb{P}\Big[\inf_{d \in \mathcal{D}_1} h_n^2(d) > c\Big] = 1, \tag{A.3}$$

where

$$h_n(d) \coloneqq \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} b_{j,n}(d_0 - d + 1)\, \epsilon_{n-j}, \quad b_{j,n}(d) \coloneqq \frac{\alpha_j(d)}{n^{d-1}},$$

and this is sufficient because the verifications of (2.39) and (2.40) in Hualde and Robinson (2011) can be carried out in the same way.

Verification of Condition (A.3)

We first introduce notation. Let $C_{\mathcal{D}_1}$ be the set of continuous functions on the (non-empty) closed interval $\mathcal{D}_1$ and $d_{\infty,\mathcal{D}_1}$ be the uniform metric on $\mathcal{D}_1$, *i.e.* $d_{\infty,\mathcal{D}_1}(f,g) := \max_{d\in\mathcal{D}_1}|f(d)-g(d)|$ for $f,g\in C_{\mathcal{D}_1}$. Notice that $(C_{\mathcal{D}_1}, d_{\infty,\mathcal{D}_1})$ is a complete and separable metric space. Let $\mathcal{B}(C_{\mathcal{D}_1})$ be the Borel σ-algebra generated by the topology induced from the metric $d_{\infty,\mathcal{D}_1}$.

Write $h_n = (h_n(d))_{d\in\mathcal{D}_1}$ for $n\in\mathbb{N}$ and then $(h_n)_{n=1}^{\infty}$ is a sequence of continuous stochastic processes with the index set $\mathcal{D}_1$. Following the same way as the proof of (2.41) instead of (2.46) in Hualde and Robinson (2011), it suffices to prove that

$$h_n \Rightarrow \left(\int_0^1 \frac{\sqrt{\sigma^2}(1-t)^{d_0-d}}{\Gamma(d_0-d+1)}\,\mathrm{d}B_t\right)_{d\in\mathcal{D}_1} \quad \text{as } n\to\infty, \tag{A.4}$$

where $B = (B_t)_{t\in[0,1]}$ is a one-dimensional standard Brownian motion and $\Rightarrow$ denotes the weak convergence in the measurable space $(C_{\mathcal{D}_1}, \mathcal{B}(C_{\mathcal{D}_1}))$. In order to prove (A.4), it suffices to verify the following two conditions: 1) convergence of the finite-dimensional distributions in (A.4) and 2) tightness of $(h_n)_{n=1}^{\infty}$. We summarize their outlines of proofs below.

1) Convergence of the finite-dimensional distributions: Similarly to the proof of (2.7) for $i=1$, the convergence of the finite-dimensional distributions follows from Theorem 1 of Hosoya (2005) once we have verified the conditions A(i), A(ii) and A(iii) in Hosoya (2005). Notice that the condition (3.2.3), that holds under Assumption 3.2.3, implies that the fourth-order cumulant spectral density function of ϵ is bounded. Then, Lemma 1 of Hosoya (2005) implies that A(iii) is verified for $\gamma_3 = 1/2$. Moreover, Lemmas A.1 and A.2 validate the conditions A(i) and A(ii) under our current setup. Therefore we conclude the convergence of the finite-dimensional distributions in (A.4).

2) Tightness: By Theorem 12.3 of Billingsley (1968), if it holds the following two conditions: a) for each $d\in\mathcal{D}_1$, $(h_n(d))_{n=1}^{\infty}$ is a tight sequence, and b) there exists some positive constant K, independent of n, such that for all $d_1, d_2\in\mathcal{D}_1$, we have

$$\mathbb{E}\big[\big(h_n(d_1)-h_n(d_2)\big)^2\big] \leqslant K(d_1-d_2)^2, \tag{A.5}$$

then the sequence of stochastic processes $(h_n)_{n=1}^{\infty}$ is tight as well. However, the same

calculations as in Hualde and Robinson (2011) yield the above conditions a) and b) even when the innovation noise ϵ is general white noise and not necessarily the martingale difference they assumed. Therefore we conclude the tightness of $(h_n)_{n=1}^{\infty}$ and the weak convergence in (A.4).

B Proof of Theorem 3.3.2

B.1 Preliminary Lemmas for Proof of Theorem 3.3.2

Recall that for each candidate model $m \in \mathcal{M}$, $\theta_{m,0}$ denotes the unique global minimizer of the function $\theta_m \mapsto \sigma^2(\theta_m)$ on Θ_m defined in (3.3.5), which is a limit of the function $\theta_m \mapsto \tilde{\sigma}_n^2(\theta_m)$ as $n \to \infty$.

Lemma B.1. *Under the assumptions in Theorem* 3.3.2*, for all candidate models $m \in \mathcal{M}$, we have the consistency of the sequence of estimators $(\widehat{\theta}_{m,n})_{n=1}^{\infty}$ to $\theta_{m,0}$, i.e. $\widehat{\theta}_{m,n} = \theta_{m,0} + o_{\mathbb{P}_{\theta_0}}(1)$ as $n \to \infty$, and the convergence*

$$\tilde{\sigma}_n^2(\widehat{\theta}_{m,n}) = \sigma^2(\theta_{m,0}) + o_{\mathbb{P}_{\theta_0}}(1) \quad \text{as } n \to \infty. \tag{B.1}$$

Proof. In the same way as the proof of the consistency in Theorem 3.2.4, we get the uniform convergence

$$\sup_{\theta_m \in \Theta_{m,4}(\iota)} |\tilde{\sigma}_n^2(\theta_m) - \sigma^2(\theta_m)| = o_{\mathbb{P}_{\theta_0}}(1) \quad \text{as } n \to \infty. \tag{B.2}$$

for each $\iota \in (0, 1/2)$, where

$$\Theta_{m,4}(\iota) := \{\theta_m = (a_1, \ldots, a_p, b_1, \ldots, b_q, d)^\top \in \Theta_m : d > d_0 - 1/2 + \iota\},$$

and then we conclude the consistency of $(\widehat{\theta}_{m,n})_{n=1}^{\infty}$ to $\theta_{m,0}$ since $\theta_{m,0}$ is a unique global minimizer of the limit function $\theta_m \mapsto \sigma^2(\theta_m)$ on Θ_m. Moreover, combining with the uniform convergence (B.2) and the consistency of $(\widehat{\theta}_{m,n})_{n=1}^{\infty}$, we also conclude the convergence (B.1). □

Lemma B.2. *Under the assumptions in Theorem* 3.3.2*, for any $m \in \mathcal{M}_+$, we have*

$$n\left(\tilde{\sigma}_n^2(\widehat{\theta}_n) - \tilde{\sigma}_n^2(\widehat{\theta}_{m,n})\right) = O_{\mathbb{P}_{\theta_0}}(1) \quad \text{as } n \to \infty.$$

Proof. Recall that we write $\widehat{\theta}_n = \widehat{\theta}_{m_0,n}$ and that, under Assumption 3.3.1, the equality $\tilde{\sigma}_n^2(\theta_0) = \tilde{\sigma}_n^2(\theta_{m,0})$ holds for all $m \in \mathcal{M}_+ \cup \{m_0\}$, see the remark right after the assumption. Then,

$$\tilde{\sigma}_n^2(\widehat{\theta}_n) - \tilde{\sigma}_n^2(\widehat{\theta}_{m,n}) = -\Big(\tilde{\sigma}_n^2(\theta_0) - \tilde{\sigma}_n^2(\widehat{\theta}_n)\Big) + \Big(\tilde{\sigma}_n^2(\theta_{m,0}) - \tilde{\sigma}_n^2(\widehat{\theta}_{m,n})\Big) \tag{B.3}$$

and it suffices to prove that, for any $m \in \mathcal{M}_+ \cup \{m_0\}$,

$$\tilde{\sigma}_n^2(\theta_{m,0}) - \tilde{\sigma}_n^2(\widehat{\theta}_{m,n}) = O_{\mathbb{P}_{\theta_0}}(n^{-1}) \quad \text{as } n \to \infty.$$

First, in the same way as the proof of the asymptotic normality of $(\widehat{\theta}_n)_{n=1}^{\infty}$ in Section A and using the consistency of $(\widehat{\theta}_{m,n})_{n=1}^{\infty}$ in Lemma B.1, we obtain the convergence

$$\sup_{\theta_m \in \Theta_{m,4}(\iota)} \left|\partial_{\theta_m}^2 \tilde{\sigma}_n^2(\theta_m) - \partial_{\theta_m}^2 \tilde{\sigma}_{m,0}^2(\theta_{m,0})\right| = o_{\mathbb{P}_{\theta_0}}(1) \quad \text{as } n \to \infty, \text{ for any } \iota \in (0, 1/2),$$

together with (B.6). Besides, the asymptotic normality of $(\widehat{\theta}_{m,n})_{n=1}^{\infty}$ in the overfitted case, that is $m \in \mathcal{M}_+$, follows by arguments similar to those used for the true model and the fact that the Fisher information matrix $J_m(\theta_{m,0}^{\dagger})$ in (3.2.5) is invertible under Assumption 3.3.1, see McLeod (1999). Hence, we have $\sqrt{n}(\widehat{\theta}_{m,n} - \theta_{m,0}) = O_{\mathbb{P}_{\theta_0}}(1)$ as $n \to \infty$ for any $m \in \mathcal{M}_+ \cup \{m_0\}$. Then the Taylor theorem yields that for any $m \in \mathcal{M}_+ \cup \{m_0\}$,

$$\begin{aligned} n\Big(\tilde{\sigma}_n^2(\theta_{m,0}) - \tilde{\sigma}_n^2(\widehat{\theta}_{m,n})\Big) =& \sqrt{n}(\theta_{m,0} - \widehat{\theta}_{m,n})^{\top} \sqrt{n}\, \partial_{\theta_m} \tilde{\sigma}_n^2(\widehat{\theta}_{m,n}) + \frac{1}{2}\sqrt{n}(\widehat{\theta}_{m,n} - \theta_{m,0})^{\top} \\ & \times \Big[\int_0^1 \partial_{\theta_m}^2 \tilde{\sigma}_n^2\big(\widehat{\theta}_{m,n} + u(\theta_{m,0} - \widehat{\theta}_{m,n})\big)\, \mathrm{d}u\Big] \sqrt{n}(\widehat{\theta}_{m,n} - \theta_{m,0}) \\ =& O_{\mathbb{P}_{\theta_0}}(1). \end{aligned}$$

This completes the proof. □

B.2 Proof of Theorem 3.3.2

First recall that $m_0 = (p_0, q_0)$ denotes the true model and, for each $m = (p, q) \in \mathcal{M}$, $\kappa_m = p+q+1$ is the dimension of the model m and we write $A_{m,n} := \{\det[\partial_{\theta_m}^2 \log \tilde{\sigma}_n^2(\widehat{\theta}_{m,n})] \leqslant 0\}$. Using the notation

$$\widehat{V}_{m,n} := \log\det\left[n\, \partial_{\theta_m}^2 \log \tilde{\sigma}_n^2(\widehat{\theta}_{m,n})\right] = \kappa_m \log n + \log\det\left[\partial_{\theta_m}^2 \log \tilde{\sigma}_n^2(\widehat{\theta}_{m,n})\right], \tag{B.4}$$

and the definition of $\mathrm{QBIC}_{m,n}$ in (3.3.3) we have

$$\begin{aligned}\Delta_{m,m_0} &:= \mathrm{QBIC}_{m,n} - \mathrm{QBIC}_{m_0,n} \\ &= n \log \frac{\tilde{\sigma}_n^2(\widehat{\theta}_{m,n})}{\tilde{\sigma}_n^2(\widehat{\theta}_n)} + \left(\mathbb{1}_{A_{m,n}^c} \widehat{V}_{m,n} - \mathbb{1}_{A_{m_0,n}^c} \widehat{V}_{m_0,n}\right) + \left(\mathbb{1}_{A_{m,n}} \kappa_m - \mathbb{1}_{A_{m_0,n}} \kappa_{m_0}\right) \log n\end{aligned}$$

for any $m \in \mathcal{M} \setminus \{m_0\}$. Then, the following equality holds:

$$\mathbb{P}_{\theta_0}[\widehat{m}_n = m_0] = 1 - \mathbb{P}_{\theta_0}[\widehat{m}_n \neq m_0] = 1 - \{\mathbb{P}_{\theta_0}[\widehat{m}_n \in \mathcal{M}_+] + \mathbb{P}_{\theta_0}[\widehat{m}_n \in \mathcal{M}_-]\}.$$

Thereby, it suffices to show that a) $\mathbb{P}_{\theta_0}[\widehat{m}_n \in \mathcal{M}_+] \underset{n\to\infty}{\to} 0$ and b) $\mathbb{P}_{\theta_0}[\widehat{m}_n \in \mathcal{M}_-] \underset{n\to\infty}{\to} 0$. In the rest of this section, we prove these convergences separately using

$$\mathbb{P}_{\theta_0}[\widehat{m}_n \in \mathcal{M}_+] = \mathbb{P}_{\theta_0}\Big[\min_{m\in\mathcal{M}_+} \Delta_{m,m_0} \leqslant 0\Big] \leqslant \sum_{m\in\mathcal{M}_+} \mathbb{P}_{\theta_0}\big[\Delta_{m,m_0} \leqslant 0\big] \tag{B.5}$$

where the same result holds for $\mathbb{P}_{\theta_0}[\widehat{m}_n \in \mathcal{M}_-]$.

a) Proof of $\lim_{n\to\infty} \mathbb{P}_{\theta_0}[\widehat{m}_n \in \mathcal{M}_+] = 0$

First, using analogous arguments to the proofs of Lemmas 16 and 17 in Boubacar Maïnassara et al. (2021) and the consistency of $(\widehat{\theta}_{m,n})_{n=1}^{\infty}$ to $\theta_{m,0}$ established in Lemma B.1, we obtain that

$$\frac{\partial^2}{\partial\theta_i\,\partial\theta_j} \log \tilde{\sigma}_n^2(\widehat{\theta}_{m,n}) \underset{n\to\infty}{\to} \frac{1}{\sigma^2} J_m(\theta_{m,0}^{\dagger})(i,j) < \infty \quad \forall 1 \leqslant i,j \leqslant \kappa_m \tag{B.6}$$

for any $m \in \mathcal{M}_+ \cup m_0$, where obviously $J_{m_0}(\theta_{m_0,0}^{\dagger}) = J(\theta_0)$. Moreover, the Taylor theorem and Lemmas B.1 and B.2 yield

$$R_n := n \log \frac{\tilde{\sigma}_n^2(\widehat{\theta}_n)}{\tilde{\sigma}_n^2(\widehat{\theta}_{m,n})} = \int_0^1 \frac{n\left(\tilde{\sigma}_n^2(\widehat{\theta}_n) - \tilde{\sigma}_n^2(\widehat{\theta}_{m,n})\right)}{\tilde{\sigma}_n^2(\widehat{\theta}_{m,n}) + u\left(\tilde{\sigma}_n^2(\widehat{\theta}_n) - \tilde{\sigma}_n^2(\widehat{\theta}_{m,n})\right)}\,\mathrm{d}u = O_{\mathbb{P}_{\theta_0}}(1). \tag{B.7}$$

Then, for any $m \in \mathcal{M}_+$ and $n \geqslant 2$, one has

$$\begin{aligned}\mathbb{P}_{\theta_0}\big[\Delta_{m,m_0} \leqslant 0\big] &= \mathbb{P}_{\theta_0}\big[\{\Delta_{m,m_0} \leqslant 0\} \cap A_{m,n}\big] + \mathbb{P}_{\theta_0}\big[\{\Delta_{m,m_0} \leqslant 0\} \cap A_{m,n}^c\big] \\ &\leqslant \mathbb{P}_{\theta_0}\big[R_n + \mathbb{1}_{A_{m_0,n}^c} \widehat{V}_{m_0,n} \geqslant (\kappa_m - \mathbb{1}_{A_{m_0,n}} \kappa_{m_0}) \log n\big] \\ &\quad + \mathbb{P}_{\theta_0}\big[R_n + \mathbb{1}_{A_{m_0,n}^c} \widehat{V}_{m_0,n} - \widehat{V}_{m,n} \geqslant (-\mathbb{1}_{A_{m_0,n}} \kappa_{m_0}) \log n\big]\end{aligned}$$

$$= \mathbb{P}_{\theta_0}\Big[\frac{R_n}{\log n} \geqslant \kappa_m - \kappa_{m_0}\Big] + \mathbb{P}_{\theta_0}\Big[\frac{R_{n,1}}{\log n} \geqslant \kappa_m - \kappa_{m_0}\Big]$$
$$+ \mathbb{P}_{\theta_0}\Big[\frac{R_{n,2}}{\log n} \geqslant \kappa_m - \kappa_{m_0}\Big] + \mathbb{P}_{\theta_0}\Big[\frac{R_{n,3}}{\log n} \geqslant \kappa_m - \kappa_{m_0}\Big]$$

where (B.6) and (B.7) imply

$$R_{n,1} := R_n + \log\det\Big[\partial^2_\theta \log\tilde\sigma^2_n(\widehat\theta_n)\Big] = O_{\mathbb{P}_{\theta_0}}(1),$$
$$R_{n,2} := R_n - \log\det\Big[\partial^2_{\theta_m} \log\tilde\sigma^2_n(\widehat\theta_{m,n})\Big] = O_{\mathbb{P}_{\theta_0}}(1) \quad \text{and}$$
$$R_{n,3} := R_n + \log\det\Big[\partial^2_\theta \log\tilde\sigma^2_n(\widehat\theta_n)\Big] - \log\det\Big[\partial^2_{\theta_m} \log\tilde\sigma^2_n(\widehat\theta_{m,n})\Big] = O_{\mathbb{P}_{\theta_0}}(1).$$

The conclusion follows from the fact that $\kappa_m > \kappa_{m_0}$ and (B.5).

b) Proof of $\lim_{n\to\infty} \mathbb{P}_{\theta_0}[\widehat m_n \in \mathcal{M}_-] = 0$

First recall that, under Assumption 3.3.1, we have $\sigma^2(\theta_{m,0}) > \sigma^2(\theta_0) = \sigma^2$ for any $m \in \mathcal{M}_-$. Thus, since the logarithmic function is strictly increasing, Lemma B.1 and the Continuous Mapping Theorem yield

$$\log\tilde\sigma^2_{m,n}(\widehat\theta_{m,n}) - \log\tilde\sigma^2_n(\widehat\theta_n) = o_{\mathbb{P}_{\theta_0}}(1) + \log\frac{\sigma^2(\theta_{m,0})}{\sigma^2} \quad \text{where} \quad \log\frac{\sigma^2(\theta_{m,0})}{\sigma^2} > 0. \tag{B.8}$$

Then, as in the previous section a) with Lemma 17 in Boubacar Maïnassara et al. (2021) and Lemma B.1, we establish that, for any $m \in \mathcal{M}_-$,

$$\frac{\partial^2}{\partial\theta_m\,\partial\theta_m^\top} \log\tilde\sigma^2_n(\widehat\theta_{m,n}) \underset{n\to\infty}{\to} \frac{1}{\sigma^2(\theta_{m,0})} J_m(\theta_{m,0}) + G_m(\theta_{m,0}) + H_m(\theta_{m,0})\, H_m(\theta_{m,0})^\top$$

where $J_m(\theta_m)$ is given in (3.2.5),

$$G_m(\theta_{m,0}) = \frac{2}{\sigma^2(\theta_{m,0})}\,\mathbb{E}_{\theta_m}\left[\epsilon_t(\theta_{m,0})\,\frac{\partial^2}{\partial\theta_m\,\partial\theta_m^\top}\epsilon_t(\theta_{m,0})\right]$$

and

$$H_m(\theta_{m,0}) = \frac{2}{\sigma^2(\theta_{m,0})}\,\mathbb{E}_{\theta_m}\left[\epsilon_t(\theta_{m,0})\,\frac{\partial}{\partial\theta_m}\epsilon_t(\theta_{m,0})\right].$$

As in (32) and (33) of Boubacar Maïnassara et al. (2021), for any $\theta_m = (\xi_m^\top, d)^\top \in \Theta_m$ and $1 \leqslant i, j \leqslant \kappa_m$, there exist sequences $(\dot{\lambda}_{k,i}(\theta_m))_{k=0}^{\infty}$ and $(\ddot{\lambda}_{k,i,j}(\theta_m))_{k=0}^{\infty}$ directly constructed from the derivatives of $\zeta_t(\theta_m)$, such that

$$\frac{\partial}{\partial \theta_{m,i}} \epsilon_t(\theta_m) = \sum_{k=0}^{t-1} \dot{\lambda}_{k,i}(\theta_m)\, \epsilon_{t-k} \quad \text{and} \quad \frac{\partial^2}{\partial \theta_{m,i}\, \partial \theta_{m,j}^\top} \epsilon_t(\theta_m) = \sum_{k=0}^{t-1} \ddot{\lambda}_{k,i,j}(\theta_m)\, \epsilon_{t-k}.$$

For any small enough $\eta > 0$, those coefficients are of order $O_{\mathbb{P}_{\theta_0}}(k^{-1-(d-d_0)+\eta})$, see for instance Hallin et al. (1999). This also implies that, for any $\iota \in (0, 1/2)$,

$$\sup_{\theta_m \in \Theta_{m,4}(\iota)} \frac{\partial^2}{\partial \theta_i\, \partial \theta_j} \log \tilde{\sigma}_n^2(\theta_m) = O_{\mathbb{P}_{\theta_0}}(1) \quad \text{as } n \to \infty, \tag{B.9}$$

for any $1 \leqslant i, j \leqslant \kappa_m$ and $m \in \mathcal{M}_-$. Finally, we have

$$\mathbb{P}_{\theta_0}\big[\Delta_{m,m_0} \leqslant 0\big] = \mathbb{P}_{\theta_0}\left[\frac{\mathbb{V}(m_0, m)}{n} + \left(\mathbb{1}_{A_{m_0,n}}\, \kappa_{m_0} - \mathbb{1}_{A_{m,n}}\, \kappa_m\right) \frac{\log n}{n} \geqslant \log \frac{\tilde{\sigma}_n^2(\widehat{\theta}_{m,n})}{\tilde{\sigma}_n^2(\widehat{\theta}_n)}\right]$$

where

$$\mathbb{V}(m_0, m) := \left(\mathbb{1}_{A_{m_0,n}^c}\, \widehat{V}_{m_0,n} - \mathbb{1}_{A_{m,n}^c}\, \widehat{V}_{m,n}\right) = O_{\mathbb{P}_{\theta_0}}(1)$$

from (B.4), (B.6) and (B.9). The conclusion follows from (B.8).

Conclusion et perspectives

Cette thèse s'inscrit dans le domaine de la statistique asymptotique pour les séries chronologiques en y abordant deux problématiques distinctes.

Elle s'intéresse tout d'abord à l'inférence des processus autorégressifs stables dont les paramètres dépendent de la taille de l'échantillon, conduisant à un régime instable à la limite. Dans ce cadre, la consistance et la normalité asymptotique de l'estimateur des moindres carrés ont été établies. Sur la base de ces résultats, un test statistique a été développé, offrant une perspective nouvelle sur la problématique de racine unitaire, complémentaire aux tests classiques de stationnarité. Ces développements théoriques obtenus dans un cadre autorégressif, pourraient être étendus à des processus plus généraux, tels que les modèles ARMA, FARIMA ou à toute autre classe de modèles présentant une rupture nette du comportement asymptotique des estimateurs. Une autre perspective d'extension serait de lever la première hypothèse impliquant la diagonalisation de la matrice compagne en s'appuyant sur une décomposition en blocs de Jordan. Si cette généralisation reste essentiellement technique dans le cas des valeurs propres évoluant strictement dans le domaine stable, la présence de plusieurs racines unitaires à la limite conduit à une situation plus délicate, susceptible d'affecter les vitesses de convergence comme l'illustrent les travaux de Chan and Wei (1988). Enfin, un prolongement naturel de ces recherches réside dans l'analyse du cadre explosif de ces processus autorégressifs à coefficients variant dans le temps. Plusieurs situations peuvent être envisagées selon la position des valeurs propres : toutes peuvent être à l'extérieur du cercle unité et la plus petite converge vers celui-ci, ou à l'inverse, toutes peuvent se trouver à l'intérieur du cercle sauf la plus grande qui est à l'extérieur mais tend vers 1 en module, sans oublier les situations mixtes où certaines valeurs propres sont à l'intérieur, d'autres à l'extérieur, et l'une d'elles évolue vers la frontière unité depuis l'intérieur ou l'extérieur du cercle. Cette étude est actuellement en cours avec Frédéric Proïa et déjà bien avancée.

Le second axe de ce travail est consacré à la sélection de modèle par minimisation d'un critère, au sein d'une classe générale de processus FARIMA. La consistance de deux critères de sélection y est démontrée, garantissant l'identification du modèle dans une classe de processus autorisant la stationnarité ou la non-stationnarité, la présence de

dépendance à court terme et/ou la mémoire longue, ainsi qu'une structure de dépendance non linéaire du bruit. Il resterait toutefois à s'affranchir de l'hypothèse d'identifiabilité des modèles candidats afin d'étendre ces résultats à des cadres applicatifs très généraux. Plus largement, ce travail s'inscrit dans l'étude de l'inférence pour les processus FARIMA faibles, domaine dans lequel la littérature demeure encore relativement restreinte. Ainsi, établir les conditions d'existence de moments finis pour l'erreur d'estimation pourrait non seulement enrichir ce cadre de recherche mais également permettre de caractériser les propriétés prédictives associées à ce modèle. En revenant sur la problématique de sélection de modèle, une perspective secondaire, déjà largement développée et directement liée aux conditions d'existence des moments finis, consiste à proposer un nouveau critère pour l'identification des processus FARIMA faibles, combinant la consistance du BIC avec une pénalité type AIC adaptée à la structure de dépendance du bruit. La construction de ce critère et la preuve de sa consistance doit faire l'objet d'une publication en collaboration avec Alexandre Brouste et Youssef Esstafa. Enfin, un package R est en cours de développement. Il regroupera les procédures de simulation, sélection, estimation et validation des processus FARIMA faibles implémentées durant cette thèse.

BIBLIOGRAPHIE

Akaike, H. (1974), « A new look at the statistical model identification », *in*: *IEEE Transactions on Automatic Control* AC-19, pp. 716–723.

Badreau, M. and Proïa, F. (2023), « Consistency and asymptotic normality in a class of nearly unstable processes », *in*: *Statistical Inference for Stochastic Processes*, DOI: 10.1007/s11203-023-09290-2.

Badreau, M. and Proïa, F. (2025), « Testing for the extent of instability in nearly unstable processes », *in*: *Journal of Time Series Analysis*, DOI: 10.1111/jtsa.12751.

Baillie, R. T., Kapetanios, G., and Papailias, F. (2014), « Modified information criteria and selection of long memory time series models », *in*: *Computational Statistics & Data Analysis* 76, pp. 116–131.

Barczy, M., Ispány, M., and Pap, G. (2011), « Asymptotic behavior of unstable INAR(p) processes », *in*: *Stochastic Process. Appl.* 121, pp. 563–608.

Bardet, J. M., Kamila, K., and Kengne, W. (2020), « Consistent model selection criteria and goodness-of-fit test for common time series models », *in*: *Electronic Journal of Statistics* 14(*1*), pp. 2009–2052.

Bardet, J. M., Kare, K., and Kengne, W. (2023), « Efficient and consistent model selection procedures for time series », *in*: *Bernoulli* 29(*4*), pp. 2652–2690.

Barreto-Souza, W. and Chan, N. H. (2024), « Nearly unstable integer-valued ARCH process and unit root testing », *in*: *Scand. J. Stat.* 51, pp. 402–424.

Ben Hariz, S., Brouste, A., Esstafa, Y., and Soltane, M. (2023), « Fast calibration of weak FARIMA models », *in*: *ESAIM: Probability and Statistics* 27, pp. 156–173.

Ben Hariz, S., Brouste, A., Esstafa, Y., and Soltane, M. (2025), « Full Whittle inference for weak FARIMA models », *in*: preprint, URL: https://hal.science/hal-04702353v1.

Beran, J. (1995), « Maximum likelihood estimation of the differencing parameter for invertible short- and long-memory ARIMA models », *in*: *Journal of the Royal Statistical Society, Series B* 57(*4*), pp. 659–672.

Beran, J., Bhansali, R. J., and Ocker, D. (1998), « On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes », *in*: *Biometrika* 85(*4*), pp. 921–934.

Beran, J., Feng, Y., Ghosh, S., and Kulik, R. (2013), *Long-Memory Processes*, Probabilistic Properties and Statistical Methods, Springer, Heidelberg.

Berk, K. N. (1974), « Consistent autoregressive spectral estimates », *in*: *The Annals of Statistics* 2, pp. 489–502.

Bhargava, A. (1986), « On the theory of testing for unit roots in observed time series », *in*: *Rev. Econom. Stud.* 53, pp. 369–384.

Billingsley, P. (1968), *Convergence of Probability Measures*, John Wiley: New York.

Billingsley, P. (1995), *Probability and Measure*, 3rd ed., Wiley Series in Probability and Statistics, Wiley: New York.

Bollerslev, T. (1986), « Generalized Autoregressive Conditional Heteroskedasticity », *in*: *Journal of Econometrics* 31(*3*), pp. 307–327.

Boubacar Maïnassara, Y. (2011), « Selection of weak VARMA models by modified Akaike's information criteria », *in*: *Journal of Time Series Analysis* 33(*1*), pp. 121–130.

Boubacar Maïnassara, Y., Carbon, M., and Francq, C. (2012), « Computing and estimating information matrices of weak ARMA models », *in*: *Computational Statistics & Data Analysis* 56(*2*), pp. 345–358.

Boubacar Maïnassara, Y. and Esstafa, Y. (2025), « Estimation of the asymptotic variance matrix of the least squares estimator of weak FARIMA models », *in*: *Statistica Neerlandica* 79(*3*).

Boubacar Maïnassara, Y., Esstafa, Y., and Saussereau, B. (2021), « Estimating FARIMA models with uncorrelated but non-independent error terms », *in*: *Statistical Inference for Stochastic Processes* 24, pp. 549–608.

Boubacar Maïnassara, Y., Esstafa, Y., and Saussereau, B. (2023), « Diagnostic checking in FARIMA models with uncorrelated but non-independent error terms », *in*: *Electronic Journal of Statistics* 17(*1*), pp. 1160–1239.

Boubacar Maïnassara, Y. and Francq, C. (2011), « Estimating Structural VARMA Models with Uncorrelated but Non-Independent Error Terms », *in*: *Journal of Multivariate Analysis* 102(*3*), pp. 496–505.

Boubacar Maïnassara, Y. and Saussereau, B. (2018), « Diagnostic Checking in Multivariate ARMA Models With Dependent Errors Using Normalized Residual Autocorrelations », *in*: *Journal of the American Statistical Association* 113(*524*), pp. 1813–1827.

Box, G. E. P. and Jenkins, G. M. (1970), *Time Series Analysis: Forecasting and Control*, Holden-Day: San Francisco.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015), *Time Series Analysis: Forecasting and Control*, 5th ed., Wiley.

Box, G. E. P. and Pierce, D. A. (1970), « Distribution of residual autocorrelations in autoregressive-integrated moving average time series models », *in*: *Journal of the American Statistical Association* 65, pp. 1509–1526.

Brockwell, P. J. and Davis, R. A. (1991), *Time Series: Theory and Methods*, 2nd ed., Springer-Verlag: New York.

Buchmann, B. and Chan, N. H. (2013), « Unified asymptotic theory for nearly unstable AR(p) processes », *in*: *Stochastic Processes and their Applications* 123, pp. 952–985.

Chan, N. H. and Wei, C. Z. (1987), « Asymptotic inference for nearly nonstationary AR(1) processes », *in*: *Annals of Statistics* 15, pp. 1050–1063.

Chan, N. H. and Wei, C. Z. (1988), « Limiting distributions of least squares estimates of unstable autoregressive processes », *in*: *Ann. Stat.* 16, pp. 367–401.

Dahlhaus, R. (1989), « Efficient parameter estimation for self-similar processes », *in*: *Annals of Statistics* 17(*4*), pp. 1749–1766.

DeJong, D. N., Nankervis, J. C., Savin, N. E., and Whiteman, C. H. (1992), « Integration Versus Trend Stationarity in Time Series », *in*: *Econometrica* 60, pp. 423–433.

Dickey, D. A. and Fuller, W. A. (1979), « Distribution of the estimators for autoregressive time series with a unit root », *in*: *Journal of the American Statistical Association* 74(*366*), pp. 427–431.

Doob, J. L. (1953), *Stochastic Processes*, Wiley Publications in Statistics, John Wiley & Sons: New York, pp. vii+654.

Doukhan, P. and Léon, J. R. (1989), « Cumulants for stationary mixing random sequences and applications to empirical spectral density », *in*: *Probability and Mathematical Statistics* 10(*1*), pp. 11–26.

Duflo, M. (1997), *Random Iterative Models*, Applications of Mathematics, Vol. 34, Springer-Verlag: Berlin.

Eguchi, S. and Masuda, H. (2018), « Schwarz type model comparison for LAQ models », *in*: *Bernoulli* 24(*3*), pp. 2278–2327.

Engle, R. F. (1982), « Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation », *in*: *Econometrica* 50, pp. 987–1007.

Fox, R. and Taqqu, M. S. (1986), « Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series », *in*: *Annals of Statistics* 14(*3*), pp. 517–532.

Francq, C., Roy, R., and Zakoïan, J. M. (2005), « Diagnostic checking in ARMA models with uncorrelated errors », *in*: *Journal of Time Series Analysis* 26(*2*), pp. 251–272.

Francq, C. and Zakoïan, J. M. (1998), « Estimating linear representations of nonlinear processes », *in*: *Journal of Time Series Analysis* 19(*3*), pp. 253–273.

Francq, C. and Zakoïan, J. M. (2005), « Recent results for linear time series models with non independent innovations », *in*: *Statistical Modeling and Analysis for Complex Data Problems*, vol. 1, GERAD 25th Anniversary Series, Springer: New York, pp. 241–265.

Francq, C. and Zakoïan, J. M. (2019), *GARCH Models: Structure, Statistical Inference and Financial Applications*, Second, Wiley: Chichester.

Giraitis, L. and Phillips, P. C. B. (2006), « Uniform limit theory for stationary autoregression », *in*: *Journal of Time Series Analysis* 27, pp. 51–60.

Granger, C. W. J. and Joyeux, R. (1980), « An Introduction to Long-Memory Time Series Models and Fractional Differencing », *in*: *Journal of Time Series Analysis* 1, pp. 15–29.

Hallin, M., Taniguchi, M., Serroukh, A., and Choy, K. (1999), « Local asymptotic normality for regression models with long-memory disturbance », *in*: *Annals of Statistics* 27(*6*), pp. 2054–2080.

Hannan, E. J. (1980), « The estimation of the order of an ARMA process », *in*: *Annals of Statistics* 8(*5*), pp. 1071–1081.

Hannan, E. J. and Rissanen, J. (1982), « Recursive estimation of mixed autoregressive-moving average order », *in*: *Biometrika* 69(*1*), pp. 81–94.

Horn, R. A. and Johnson, C. R. (2012), *Matrix Analysis (Second Edition)*, Cambridge University Press: Cambridge, New-York.

Horváth, L. and Trapani, L. (2016), « Statistical inference in a random coefficient panel model », *in*: *Journal of Econometrics* 193, pp. 54–75.

Hosking, J. R. M. (1981), « Fractional Differencing », *in*: *Biometrika* 68, pp. 165–176.

Hosoya, Y. (2005), « Fractional invariance principle », *in*: *Journal of Time Series Analysis* 26(*3*), pp. 463–486.

Hualde, J. and Robinson, P. M. (2011), « Gaussian pseudo-maximum likelihood estimation of fractional time series models », *in*: *Annals of Statistics* 39(*6*), pp. 3152–3181.

Huang, H.-H., Chan, N. H., Chen, K., and Ing, C.-K. (2022), « Consistent order selection for ARFIMA processes », *in*: *Annals of Statistics* 50(*3*), pp. 1297–1319.

Hurst, H. E. (1951), « Long-Term Storage Capacity of Reservoirs », *in*: *Transactions of the American Society of Civil Engineers* 116, pp. 770–799.

Imhof, J. P. (1961), « Computing the distribution of quadratic forms in normal variables », *in*: *Biometrika* 48, pp. 419–426.

Ispány, M., Pap, G., and Van Zuijlen, M. C. A. (2003), « Asymptotic inference for nearly unstable INAR(1) models », *in*: *J. Appl. Probab.* 40, pp. 750–765.

Ispány, M., Pap, G., and Van Zuijlen, M. C. A. (2005), « Fluctuation limit of branching processes with immigration and estimation of the means », *in*: *Adv. Appl. Probab.* 37, pp. 523–538.

Jiang, T. and Wei, M. (2003), « On solutions of the matrix equations X - A X B = C and X - A $\bar{X}$ B = C », *in*: *Linear Algebra Appl.* 367, pp. 225–233.

Kamila, K. (2022), « General Hannan and Quinn criterion for common time series », *in*: *Statistics* 56(*1*), pp. 222–241.

Kengne, W. (2022), « On consistency for time series model selection », *in*: *Statistical Inference for Stochastic Processes* 26(*2*), pp. 437–458.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. (1992), « Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We that Economic Time Series Have a Unit Root? », *in*: *Journal of Econometrics* 54, pp. 159–178.

Lai, T. L. and Wei, C. Z. (1983), « Asymptotic Properties of General Autoregressive Models and Strong Consistency of Least-Squares Estimates of Their Parameters », *in*: *Journal of Multivariate Analysis* 13, pp. 1–23.

Leybourne, S. J. and McCabe, B. P. M. (1994), « A consistent test for a unit root », *in*: *J. Bus. Econ. Stat.* 12, pp. 157–166.

Li, W. K. and McLeod, A. I. (1986), « Fractional time series modelling », *in*: *Biometrika* 73(*1*), pp. 217–221.

Ling, S. (2004), « Estimation and testing stationarity for double-autoregressive models », *in*: *J. R. Stat. Soc. B.* 66, pp. 63–78.

Ling, S. and Li, W. K. (1997), « On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity », *in*: *Journal of the American Statistical Association* 92(*439*), pp. 1184–1194.

Ljung, G. M. and Box, G. E. P. (1978), « On a measure of lack of fit in time series models », *in*: *Biometrika* 65(*2*), pp. 297–303.

Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Springer: Berlin and Heidelberg.

MacKinnon, J. G. (1991), *Critical values for cointegration tests*, Long-Run Economic Relationships, ed. by R. F. Engle and C. W. Granger, Oxford University Press: Oxford.

Mandelbrot, B. and Van Ness, J. W. (1968), « Fractional Brownian Motions, Fractional Noises and Applications », *in*: *SIAM Review* 10(*4*), pp. 422–437.

Mandelbrot, B. and Wallis, J. R. (1968), « Noah, Joseph, and Operational Hydrology », *in*: *Water Resources Research* 4(*5*), pp. 909–918.

Mandelbrot, B. and Wallis, J. R. (1969a), « Robustness of the Rescaled Range R/S in the Measurement of Noncyclic Long-Run Statistical Dependence », *in*: *Water Resources Research* 5(*5*), pp. 967–988.

Mandelbrot, B. and Wallis, J. R. (1969b), « Some Long-Run Properties of Geophysical Records », *in*: *Water Resources Research* 5(*2*), pp. 321–340.

McLeod, A. I. (1999), « Necessary and Sufficient Condition for Nonsingular Fisher Information Matrix in ARMA and Fractional ARIMA Models », *in*: *The American Statistician* 53(*1*), pp. 71–72.

Meer, T. Van der, Pap, G., and Van Zuijlen, M. C. A. (1999), « Asymptotic inference for nearly unstable AR(p) processes », *in*: *Econom. Theory* 15, pp. 184–217.

Miao, Y., Wang, Y., and Yang, G. (2015), « Moderate deviation principles for empirical covariance in the neighbourhood of the unit root », *in*: *Scandinavian Journal of Statistics* 42, pp. 234–255.

Nelson, C. R. and Plosser, C. I. (1982), « Trends and random walks in macroeconomic time series: Some evidence and implications », *in*: *J. Monet. Econ.* 10, pp. 139–162.

Ng, S. and Perron, P. (1995), « Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag », *in*: *J. Amer. Statist. Assoc.* 90, pp. 268–281.

Nielsen, H. B. and Rahbek, A. (2014), « Unit root vector autoregression with volatility induced stationarity », *in*: *J. Empir. Financ.* 29, pp. 144–167.

Ornstein, L. S. and Uhlenbeck, G. E. (1930), « On the Theory of the Brownian Motion », *in*: *Physical Review* 36, pp. 823–841.

Palma, W. (2007), *Long-Memory Time Series: Theory and Methods*, 1st ed., Wiley Series in Probability and Statistics, John Wiley & Sons: Hoboken, NJ.

Park, J. W. (2003), *Weak unit roots*, Department of Economics, Rice University.

Perron, P. (1988), « Trends and random walks in macroeconomic time series: Further evidence from a new approach », *in*: *J. Econom. Dynam. Control* 12, pp. 297–332.

Phillips, P. C. B. (1987a), « Time Series Regression with a Unit Root », *in*: *Econometrica* 55, pp. 277–301.

Phillips, P. C. B. (1987b), « Towards a unified asymptotic theory for autoregression », *in*: *Biometrika* 74, pp. 535–547.

Phillips, P. C. B. and Lee, J. H. (2015), « Limit theory for VARs with mixed roots near unity », *in*: *Econometric Review* 34, pp. 1034–1055.

Phillips, P. C. B. and Magdalinos, T. (2007), « Limit theory for moderate deviations from a unit root », *in*: *Journal of Econometrics* 136, pp. 115–130.

Phillips, P. C. B. and Perron, P. (1988), « Testing for a Unit Root in Time Series Regression », *in*: *Biometrika* 75, pp. 335–346.

Pollard, D. (1984), *Convergence of Stochastic Processes*, Springer: Berlin.

Proïa, F. (2018), « Stationarity against integration in the autoregressive process with polynomial trend », *in*: *Probability and Mathematical Statistics* 38, pp. 1–26.

Proïa, F. (2020), « Moderate deviations in a class of stable but nearly unstable processes », *in*: *Journal of Statistical Planning and Inference* 208, pp. 66–81.

Said, E. S. and Dickey, D. A. (1981), « Testing ARIMA(p,1,q) versus ARMA(p+1,q) », *in*: *Proc. Bus. Econ. Statist. Sect., Am. Statist. Assoc.*, pp. 318–322.

Said, E. S. and Dickey, D. A. (1984), « Testing for Unit Roots in Autoregressive Moving Average Models of Unknown Order », *in*: *Biometrika* 71(*3*), pp. 599–607.

Schmidt, P. and Phillips, P. (1992), « LM test for a unit root in the presence of deterministic trends », *in*: *Oxford Bull. Econ. Stat.* 54, pp. 257–287.

Schwarz, G. (1978), « Estimating the dimension of a model », *in*: *Annals of Statistics* 6, pp. 461–464.

Schwert, G. (1989), « Tests for unit roots: a Monte Carlo investigation », *in*: *J. Bus. Econom. Statist.* 7, pp. 147–160.

Shao, X. (2010), « Nonstationarity-extended Whittle estimation », *in*: *Econometric Theory* 26(*4*), pp. 1060–1087.

Shao, X. (2011), « Testing for white noise under unknown dependence and its applications to diagnostic checking for time series models », *in*: *Econometric Theory* 27(*2*), pp. 312–343.

Shao, X. (2012), « Parametric inference in stationary time series models with dependent errors », *in*: *Scandinavian Journal of Statistics* 39(*4*), pp. 772–783.

Shibata, R. (1976), « Selection of the order of an autoregressive model by Akaike's information criterion », *in*: *Biometrika* 63(*1*), pp. 117–126.

Shumway, R. H. and Stoffer, D. S. (2006), *Time Series Analysis and Its Applications: With R Examples*, 4th ed., Springer.

Sims, C. A. (1980), « Macroeconomics and Reality », *in*: *Econometrica* 48(*1*), pp. 1–48.

Sin, C.-Y. and Yu, S.-H. (2019), « Order selection for possibly infinite-order non-stationary time series », *in*: *AStA Advances in Statistical Analysis* 103(*2*), pp. 187–216.

Trapani, L. (2021), « Testing for strict stationarity in a Random Coefficient AutoRegression », *in*: *Econometric Review* 40, pp. 220–256.

White, J. S. (1958), « The limiting distribution of the serial correlation coefficient in the explosive case », *in*: *Ann. Math. Statist.* 29, pp. 1188–1197.

Whittle, P. (1953), « Estimation and information in stationary time series », *in*: *Arkiv för Matematik* 2, pp. 423–434.

Wold, H. (1938), *A Study in the Analysis of Stationary Time Series*, Almqvist & Wiksell: Uppsala.

Yule, G. U. (1927), « On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wölfer's Sunspot Numbers », *in*: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 226, pp. 267–298.

COLLEGE ED MATHEMATIQUES
DOCTORAL STIC
PAYS DE LA LOIRE

Le Mans Université

Titre : Statistique asymptotique pour les séries chronologiques quasi-instables et/ou à mémoire longue

Mot clés : Processus autorégressif quasi-instable ; Test de racine unitaire ; Estimateur des moindres carrés ; Processus à mémoire longue ; Processus non linaire ; Sélection de modèle

Résumé : Cette thèse explore deux problématiques statistiques liées aux séries chronologiques. Dans un premier temps, ce travail s'intéresse à l'inférence des processus autorégressifs stables mais dont les paramètres évoluent, menant vers un régime instable. L'objectif est notamment d'établir un pont entre stabilité et instabilité et de caractériser les comportements asymptotiques des estimateurs et leur vitesse de convergence. Des propriétés de consistance et normalité asymptotique sont obtenues pour l'estimateur des moindres carrés dans cette classe de processus autorégressifs quasi-instables. Ce cadre est particulièrement adapté à la problématique de racine unitaire, source de non stationnarité, et a permis le développement d'un test statistique visant à évaluer la proximité d'une série temporelle à un régime non stationnaire. Dans un second temps, les processus à mémoire longue sont considérés et l'étude se concentre sur la sélection de modèles dans une classe générale de processus FARIMA induits par un bruit blanc faible non corrélé mais susceptible de présenter des dépendances non linéaires très générales. La consistance des critères BIC et QBIC est établie pour les processus FARIMA faibles, y compris dans le cas non stationnaire.

Title: Asymptotic statistics for nearly unstable and/or long memory time series

Keywords: Nearly unstable autoregressive process; Unit root test; Least squares estimator; Long memory process; Nonlinear process; Order selection

Abstract: This thesis explores two statistical issues related to time series. In the first part, it focuses on the inference of stable autoregressive processes whose parameters evolve toward an unstable regime. The aim is to bridge the gap between stability and instability and in particular to characterize the asymptotic behaviors of the estimators and their rates of convergence. Consistency and asymptotic normality properties are established for the least squares estimator within this class of nearly-unstable autoregressive processes. This framework is particularly suitable for addressing the unit root problem, resulting in non-stationarity, and has enabled the development of a statistical test designed to assess whether a time series is close, or not, to a non-stationary regime. In the second part, long-memory processes are considered, with the study focusing on model selection within a general class of FARIMA processes driven by weakly white noise, which is uncorrelated but may exhibit very general nonlinear dependencies. The consistency of the BIC and QBIC criteria is established for weak FARIMA processes, including the non-stationary case.