



**HAL**  
open science

# Autocorrélation et stationnarité dans le processus autorégressif

Frédéric Proïa

► **To cite this version:**

Frédéric Proïa. Autocorrélation et stationnarité dans le processus autorégressif. Mathématiques générales [math.GM]. Université Sciences et Technologies - Bordeaux I, 2013. Français. NNT : 2013BOR14877. tel-01128258

**HAL Id: tel-01128258**

**<https://theses.hal.science/tel-01128258>**

Submitted on 9 Mar 2015

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

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



---

# THÈSE

Pour l'obtention du grade de

**Docteur de l'Université de Bordeaux I**

École Doctorale de Mathématiques et Informatique

Spécialité : **Mathématiques Appliquées**

Présentée par

**Frédéric PROÏA**

---

## *Autocorrélation et Stationnarité dans le Processus Autorégressif*

---

Directeur de thèse : **Pr. Bernard BERCU**

Soutenue le **04 Novembre 2013** à l'Institut de Mathématiques de Bordeaux

Devant la commission d'examen composée de

### JURY

Pr.	Bernard BERCU	Université de Bordeaux I	Directeur
Pr.	Steve LEYBOURNE	University of Nottingham	Rapporteur
Pr.	Alexander LINDNER	Technische Universität Braunschweig	Rapporteur
Pr.	Jean-Marc BARDET	Université de Paris I	Président
CR.	François CARON	University of Oxford	Examinateur
DR.	Pierre DEL MORAL	INRIA Bordeaux Sud-Ouest	Examinateur
Pr.	Arnaud GUILLIN	Université de Clermont-Ferrand	Examinateur
Pr.	Anne PHILIPPE	Université de Nantes	Examinatrice



Institut de Mathématiques de Bordeaux  
Université de Bordeaux I  
Bâtiment A33  
351, cours de la Libération  
33405 Talence Cedex

INRIA Bordeaux Sud-Ouest  
Équipe ALEA  
200, avenue de la Vieille Tour  
33405 Talence Cedex

École doctorale de Mathématiques et Informatique de Bordeaux  
Bâtiment A33  
351, cours de la Libération  
33405 Talence Cedex



« Winter is not here yet. There's a little flower, up yonder, the last bud from the multitude of bluebells that clouded those turf steps in July with a lilac mist. Will you clamber up and pluck it to show papa? »

– Emily Brontë, *Wuthering Heights*.



---

## Remerciements

---

À l'issue de ces trois années, je souhaite en tout premier lieu adresser mes remerciements les plus chaleureux à Bernard Bercu. Ses précieux conseils, sa vision critique de mes travaux de recherche et les compétences qu'il m'a permis d'acquérir au fil du temps furent certes appréciés à leur juste valeur, mais je tiens particulièrement à insister sur l'aspect relationnel, son côté humain, sa disponibilité et son écoute permanentes qui, plus que toute chose, ajoutent au travail la confiance et le plaisir de travailler. Et je considère cela comme très important dans une vie professionnelle.

Je me tourne désormais vers François Caron et Pierre Del Moral, à qui j'adresse également mes plus sincères remerciements. Je me sens obligé de rappeler ici que voilà quatre ans, je me retrouvai dans une équipe de recherche INRIA emmenée par Pierre, avec une mission d'ingénieur sur un contrat dirigé par Bernard et François. C'est à l'issue de ce contrat que, grâce à leur soutien permanent et à leur efficace collaboration administrative, j'ai pu entrer en thèse au sein du même laboratoire. Ils acceptent aujourd'hui d'examiner ce travail et de participer à mon jury de thèse, et je leur en sais gré une fois encore.

Je manifeste en outre toute ma reconnaissance à Alexander Lindner et à Steve Leybourne. Non content de leur enjoindre un manuscrit français à évaluer, je me sens coupable de leur imposer une soutenance française... Je leur adresse ainsi mes remerciements enthousiastes pour avoir accepté de rapporter cette thèse, et pour m'avoir en outre proposé une vision extérieure sur ces modestes contributions, dont je rappelle que le dernier chapitre est en grande partie inspirée des travaux fondateurs de Steve Leybourne. De les savoir dans le jury le jour de la soutenance est un honneur pour moi.

J'exprime enfin toute ma gratitude à Jean-Marc Bardet, Arnaud Guillin et Anne Philippe qui ont également accepté de participer à mon jury en qualité d'examineurs, et je me permets de les remercier pour l'intérêt qu'ils ont manifesté à l'égard de ces travaux. Je n'oublie bien sûr pas mes co-auteurs Valère Bitseki-Penda, Hacène Djellout, Sophie Bercu et Nicolas Savy pour leur aide précieuse.

Je terminerai par un petit clin d'œil à mes amis et à mes proches – qu'ils soient bordelais ou stéphanois\* – et à ma famille à laquelle j'accorde une pensée particulièrement émue pour mon oncle et mon père, à qui je dédie l'intégralité de ce travail.

---

\*. D'ailleurs, je remercie aussi l'AS Saint-Étienne d'avoir remporté un titre au cours de cette dernière saison, après 32 années blanches tristement longues.



---

## Résumé/Abstract

---

**Résumé.** Cette thèse est dévolue à l'étude de certaines propriétés asymptotiques du *processus autorégressif* d'ordre  $p$ . Ce dernier qualifie communément une suite aléatoire  $(Y_n)$  définie sur  $\mathbb{N}$  ou  $\mathbb{Z}$  et entièrement décrite par une combinaison linéaire de ses  $p$  valeurs passées, perturbée par un bruit blanc  $(\varepsilon_n)$ . Tout au long de ce mémoire, nous traitons deux problématiques majeures de l'étude de tels processus : l'*autocorrélation résiduelle* et la *stationnarité*. Nous proposons en guise d'introduction un survol nécessaire des propriétés usuelles du processus autorégressif. Les deux chapitres suivants sont consacrés aux conséquences inférentielles induites par la présence d'une autorégression significative dans la perturbation  $(\varepsilon_n)$  pour  $p = 1$  tout d'abord, puis pour une valeur quelconque de  $p$ , dans un cadre de stabilité. Ces résultats nous permettent d'apporter un regard nouveau et plus rigoureux sur certaines procédures statistiques bien connues sous la dénomination de *test de Durbin-Watson* et de *H-test*. Dans ce contexte de bruit autocorrélé, nous complétons cette étude par un ensemble de principes de déviations modérées liées à nos estimateurs. Nous abordons ensuite un équivalent en temps continu du processus autorégressif. Ce dernier est décrit par une équation différentielle stochastique et sa solution est plus connue sous le nom de *processus d'Ornstein-Uhlenbeck*. Lorsque le processus d'Ornstein-Uhlenbeck est lui-même engendré par une diffusion similaire, cela nous permet de traiter la problématique de l'autocorrélation résiduelle dans le processus à temps continu. Nous inférons dès lors quelques propriétés statistiques de tels modèles, gardant pour objectif le parallèle avec le cas discret étudié dans les chapitres précédents. Enfin, le dernier chapitre est entièrement dévolu à la problématique de la stationnarité. Nous nous plaçons dans le cadre très général où le processus autorégressif possède une tendance polynomiale d'ordre  $r$  tout en étant engendré par une marche aléatoire intégrée d'ordre  $d$ . Les résultats de convergence que nous obtenons dans un contexte d'instabilité généralisent le *test de Leybourne et McCabe* et certains aspects du *test KPSS*. De nombreux graphes obtenus en simulations viennent conforter les résultats que nous établissons tout au long de notre étude.

**Mots-clés.** Processus autorégressif, Autocorrélation résiduelle, Stabilité, Estimation paramétrique, Test de Durbin-Watson, H-test, Déviations modérées, Processus d'Ornstein-Uhlenbeck, Stationnarité, Instabilité, Racine unitaire, Test KPSS, Test de Leybourne et McCabe, Processus de Wiener, Principes d'invariance, Ponts browniens, Martingales.

**Abstract.** This thesis is devoted to the study of some asymptotic properties of the  $p$ -th order *autoregressive process*. The latter usually designates a random sequence  $(Y_n)$  defined on  $\mathbb{N}$  or  $\mathbb{Z}$  and completely described by a linear combination of its  $p$  last values and a white noise  $(\varepsilon_n)$ . All through this manuscript, one is concerned with two main issues related to the study of such processes : *serial correlation* and *stationarity*. We intend, by way of introduction, to give a necessary overview of the usual properties of the autoregressive process. The two following chapters are dedicated to inferential consequences coming from the presence of a significative autoregression in the disturbance  $(\varepsilon_n)$  for  $p = 1$  on the one hand, and then for any  $p$ , in the stable framework. These results enable us to give a new light on some statistical procedures such as the *Durbin-Watson test* and the *H-test*. In this autocorrelated noise framework, we complete the study by a set of moderate deviation principles on our estimates. Then, we tackle a continuous-time equivalent of the autoregressive process. The latter is described by a stochastic differential equation and its solution is the well-known *Ornstein-Uhlenbeck process*. In the case where the Ornstein-Uhlenbeck process is itself driven by an Ornstein-Uhlenbeck process, one deals with the serial correlation issue for the continuous-time process. Hence, we infer some statistical properties of such models, keeping the parallel with the discrete-time framework studied in the previous chapters as an objective. Finally, the last chapter is entirely devoted to the stationarity issue. We consider the general autoregressive process with a polynomial trend of order  $r$  driven by a random walk of order  $d$ . The convergence results in the unstable framework generalize the *Leybourne and McCabe test* and some angles of the *KPSS test*. Many graphs obtained by simulations come to strengthen the results established all along the study.

**Keywords.** Autoregressive process, Serial correlation, Stability, Parametric estimation, Durbin-Watson test, H-test, Moderate deviations, Ornstein-Uhlenbeck process, Stationarity, Instability, Unit root, KPSS test, Leybourne and McCabe test, Wiener process, Invariance principles, Brownian bridges, Martingales.

---

# Table des matières

---

<b>Remerciements</b> .....	7
<b>Résumé/Abstract</b> .....	9
<b>Introduction générale</b> .....	15
<b>Glossaire</b> .....	19
<b>Chapitre I. Sur les séries chronologiques linéaires</b> .....	21
I.1 Quelques concepts fondamentaux . . . . .	21
I.1.1 Le processus stochastique . . . . .	21
I.1.2 La stationnarité . . . . .	22
I.1.3 L'autocorrélation . . . . .	23
I.1.4 La densité spectrale . . . . .	25
I.2 Le processus ARMA stationnaire . . . . .	25
I.2.1 Le modèle et ses propriétés usuelles . . . . .	26
I.2.2 Inférence pour les processus autorégressifs . . . . .	32
I.3 Prise en compte de la non stationnarité . . . . .	34
I.3.1 La non stationnarité déterministe . . . . .	35
I.3.2 La non stationnarité stochastique . . . . .	36
<b>Chapitre II. Sur la statistique de Durbin-Watson</b> .....	41
II.1 Historique et motivation . . . . .	41
II.2 Le modèle et ses estimateurs . . . . .	42
II.2.1 Le paramètre de l'autorégression . . . . .	43
II.2.2 Le paramètre de l'autocorrélation résiduelle . . . . .	46
II.2.3 La variance résiduelle . . . . .	49
II.3 Application au test de Durbin-Watson . . . . .	50
II.3.1 Le cas pathologique . . . . .	50
II.3.2 Le cas général . . . . .	51

## Table des matières

---

II.3.3	Le cas standard . . . . .	54
II.4	Simulations et comparaisons . . . . .	54
II.4.1	Puissance empirique du test général . . . . .	55
II.4.2	Puissance empirique du test standard . . . . .	56
II.5	Conclusion et perspectives . . . . .	60
II.6	Annexe : démonstration des résultats . . . . .	61
II.6.1	Sur les estimateurs . . . . .	61
II.6.2	Sur la procédure statistique . . . . .	61
<b>Chapitre III.</b>	<b>Une généralisation vectorielle . . . . .</b>	<b>65</b>
III.1	Historique et motivation . . . . .	65
III.2	Le modèle et ses estimateurs . . . . .	66
III.2.1	Le paramètre de l'autorégression . . . . .	66
III.2.2	Le paramètre de l'autocorrélation résiduelle . . . . .	71
III.2.3	La variance résiduelle . . . . .	75
III.3	Application au test de Durbin-Watson . . . . .	76
III.3.1	Comparaisons empiriques . . . . .	77
III.3.2	Un équivalent du H-test . . . . .	79
III.4	Conclusion et perspectives . . . . .	79
III.5	Annexe : démonstration des résultats . . . . .	80
III.5.1	Un peu d'algèbre linéaire . . . . .	80
III.5.2	Sur les estimateurs . . . . .	84
III.5.3	Sur la procédure statistique . . . . .	100
<b>Chapitre IV.</b>	<b>Quelques principes de déviations modérées . . . . .</b>	<b>103</b>
IV.1	Introduction et motivation . . . . .	103
IV.2	Définitions et notations . . . . .	104
IV.2.1	La convergence (super-)exponentielle . . . . .	104
IV.2.2	Le principe de grandes déviations . . . . .	105
IV.2.3	Le principe de déviations modérées . . . . .	105
IV.3	Sur les déviations modérées dans le cas gaussien . . . . .	106
IV.3.1	Les hypothèses retenues . . . . .	106
IV.3.2	Déviations liées aux estimateurs . . . . .	106
IV.4	Sur les déviations modérées dans le cas Chen-Ledoux . . . . .	108
IV.4.1	Les hypothèses retenues . . . . .	108
IV.4.2	Déviations liées aux estimateurs . . . . .	109
IV.5	Conclusion et perspectives . . . . .	110
IV.6	Annexe : démonstration des résultats . . . . .	110

---

IV.6.1 Cas gaussien . . . . .	112
IV.6.2 Cas Chen-Ledoux . . . . .	120
<b>Chapitre V. Une extension en temps continu . . . . .</b>	<b>129</b>
V.1 Historique et motivation . . . . .	129
V.2 Une approche par maximum de vraisemblance . . . . .	131
V.2.1 Le paramètre de l'autorégression . . . . .	131
V.2.2 Le paramètre de l'autocorrélation résiduelle . . . . .	134
V.3 Application statistique . . . . .	137
V.3.1 Retour à Durbin-Watson . . . . .	137
V.3.2 Évidence empirique . . . . .	138
V.4 Conclusion et perspectives . . . . .	140
V.5 Annexe : démonstration des résultats . . . . .	141
V.5.1 L'ergodicité et ses implications . . . . .	141
V.5.2 Sur les estimateurs . . . . .	144
<b>Chapitre VI. Stationnarité ou multi-intégration ? . . . . .</b>	<b>151</b>
VI.1 Historique et motivation . . . . .	151
VI.1.1 Tester la présence d'une racine unitaire . . . . .	151
VI.1.2 Tester la stationnarité . . . . .	155
VI.2 Tester la stationnarité contre l'intégration . . . . .	157
VI.2.1 La procédure d'estimation . . . . .	158
VI.2.2 Focus sur les distributions asymptotiques . . . . .	162
VI.2.3 Application statistique . . . . .	163
VI.2.4 Sur le processus autorégressif à erreurs autocorrélées . . . . .	167
VI.3 Comportement en présence de multi-intégration . . . . .	169
VI.3.1 La procédure d'estimation . . . . .	170
VI.3.2 Focus sur les distributions asymptotiques . . . . .	172
VI.3.3 La variance résiduelle . . . . .	173
VI.4 Quelques processus stochastiques utiles . . . . .	174
VI.4.1 Le processus de Wiener intégré . . . . .	174
VI.4.2 Le pont brownien généralisé . . . . .	174
VI.4.3 Le pont brownien intégré . . . . .	175
VI.4.4 Le processus de Wiener recentré . . . . .	175
VI.5 Conclusion et perspectives . . . . .	177
VI.6 Annexe : démonstration des résultats . . . . .	178
VI.6.1 Un principe d'invariance en présence de multi-intégration . . . . .	178
VI.6.2 Sur la statistique de test . . . . .	180

<b>Annexe A. Une étude avec EDF R&amp;D sur la consommation individuelle d'électricité</b> .....	<b>193</b>
A.1 Le contexte de l'étude . . . . .	193
A.2 La stratégie de Box et Jenkins . . . . .	196
A.2.1 L'ensemble résiduel . . . . .	196
A.2.2 La saisonnalité . . . . .	197
A.2.3 La stationnarité . . . . .	198
A.2.4 L'autocorrélation . . . . .	198
A.2.5 La modélisation . . . . .	198
A.3 Application à la prédiction . . . . .	200
<b>Annexe B. Quelques outils techniques</b> .....	<b>203</b>
B.1 Lois fortes des grands nombres . . . . .	203
B.2 Théorèmes centraux limites . . . . .	204
B.3 Principes de déviations . . . . .	204
B.4 Principes d'invariance . . . . .	205
B.5 Théorème de continuité . . . . .	206

---

## Introduction générale

---

En statistique, on désigne par *série chronologique*, ou *série temporelle*, la modélisation d'une suite d'évènements aléatoires et séquentiellement observés, généralement sur une échelle temporelle. De nombreux flux de données réelles admettent une telle représentation, nous pouvons penser par exemple à la consommation d'électricité, l'évolution d'une action en bourse, le produit intérieur brut, le cycle des marées ou encore la progression d'un processus chimique. Bien qu'astronomes et météorologues puissent en revendiquer la paternité, les principales avancées dans la formalisation des séries chronologiques semblent devoir revenir aux statisticiens et aux économètres. Nous en retrouvons aujourd'hui l'héritage dans les domaines de l'économie et de la finance bien sûr, mais également dans l'ingénierie, le traitement du signal ou les sciences naturelles et sociales. C'est de la publication des travaux pionniers de Box et Jenkins [19]–[20] en 1970 qu'est née la théorie des séries chronologiques telle qu'on la connaît de nos jours. Celle-ci s'est trouvée motivée par la faible puissance prédictive des modèles structurels du keynésianisme en vigueur à l'époque et propose une approche tout à fait novatrice. Le *processus ARMA*, formant le cœur de l'ouvrage et sur lequel nous aurons l'occasion de revenir en détail par la suite, s'appuie exclusivement sur les observations passées d'une courbe pour en intuitiver un comportement linéaire global perturbé par un bruit aléatoire centré. La caractéristique principale d'une série temporelle est la dépendance forte et d'intérêt pratique considérable liant deux observations consécutives, à l'origine de la dynamique des modèles. Ces derniers sont dès lors d'autant plus aptes à refléter les flux réels et les évènements naturels au sein desquels l'évolution chronologique est manifeste. Parmi le panel d'applications issues de la théorie, ce sont principalement la modélisation et la prédiction qui concentrent toutes les attentions. Alors que l'on cherche d'un côté à bâtir le modèle le mieux adapté à un ensemble de données sans préjuger des évènements ayant cours en dehors de l'intervalle d'étude, on se focalise de l'autre sur l'apprentissage optimal des observations pour en inférer le comportement futur.

Du processus linéaire ARMA est née une série de modèles dérivés, de plus en plus généraux, avec pour ligne directrice la formalisation d'une classe élargie de phénomènes réels. Ces derniers tiennent compte de la *non stationnarité*, de l'*hétéroscédasticité*, voire de la *non linéarité*. Bien que cette thèse se veuille uniquement axée sur un sous-ensemble particulier de processus, il nous a semblé essentiel d'accorder au premier chapitre l'historique succinct, donc non exhaustif, des séries chronologiques aujourd'hui en vigueur. Il s'agira en outre d'explicitier toutes les notions que nous abordons

– en italiques – dans cette introduction, avant de présenter de manière descriptive le modèle de Box et Jenkins, ses propriétés puis ses évolutions majeures. Une fois le cadre de travail bien délimité, nous aborderons le *processus autorégressif*, modèle linéaire sur lequel toute notre étude sera focalisée. L'analyse rigoureuse des séries chronologiques linéaires a connu un vaste essor à la fin du siècle dernier. Alors que Harvey [62] en 1991 et Kitagawa et Gersch [71] en 1996 appréhendent les séries temporelles par une démarche bayésienne reposant principalement sur des modèles à espace d'états traités par un filtre de Kalman, c'est sur l'approche fréquentiste que nous concentrerons notre attention. Nous citerons à cet égard les travaux de Jenkins et Watts [68] en 1968, Hannan [60] en 1970, Bloomfield [17] en 1976, Priestley [111] et Brillinger [24] en 1981, ou encore Fuller [56] en 1995. Nous insisterons particulièrement sur les travaux de Brockwell et Davis [25]–[26] au cours des années 90 sur lesquels nous nous sommes largement appuyés, source de nos rappels théoriques. Ces derniers nous permettront de mettre en évidence les deux problématiques majeures au cœur de ce mémoire, la *stationnarité* et l'*autocorrélation résiduelle*, et d'ouvrir le second chapitre à l'origine de nos premiers résultats de recherche. Citons également Shumway et Stoffer [121] qui proposent en 2006 un vaste panorama des travaux existants ainsi que leur mise en pratique sur le logiciel R.

Les chocs aléatoires dans le processus autorégressif sont généralement assimilés à un *bruit blanc*, c'est en tout état de cause l'une des hypothèses de travail usuelles garantissant à la fois la consistance des estimateurs et leur normalité asymptotique. Dès le milieu du XX<sup>ème</sup> siècle, Durbin et Watson [49]–[50]–[51] ont adopté pour sujet d'étude le modèle de régression linéaire standard dont la perturbation aléatoire est issue d'un processus autorégressif du premier ordre, abandonnant de fait la blancheur résiduelle, mais aussi la consistance de l'estimation. Ils suggèrent alors une statistique aujourd'hui encore fort utilisée dans le domaine de l'économétrie, à l'origine d'une procédure de test relativement élémentaire permettant bien souvent de rejeter à raison une hypothèse d'absence de corrélation résiduelle. Nous verrons que l'extension au modèle chronologiquement dépendant présente encore certaines lacunes que nous nous proposons de combler dans les deuxième et troisième chapitres, respectivement dans le contexte simple du modèle autorégressif du premier ordre, puis dans le contexte plus technique du processus autorégressif d'ordre quelconque. En 1970, Durbin [47], confronté aux conséquences inférentielles de la présence d'une autocorrélation résiduelle et se plaçant dans le cadre chronologique, propose une subtile révision de la procédure plus connue aujourd'hui sous la dénomination de *H-test*. C'est ainsi que la fin du troisième chapitre est dédiée à la comparaison asymptotique entre ce dernier et une procédure synthétisant les améliorations que nous avons pu obtenir à l'aide d'outils *martingales* alors peu démocratisés dans les années 70. Ces outils sont essentiellement des théorèmes limites qui nous accompagneront également au cours des deux chapitres suivants, comme nous pourrons le constater. De ce fait, l'étude inférentielle des modèles autorégressifs à erreurs autocorrélées nous aura conduits, au cours des deuxième et troisième chapitres, à l'établissement de propriétés de convergence presque sûre et de normalité asymptotique des estimateurs des moindres carrés liés aux paramètres du processus, dans un contexte de *stabilité*.

Le prolongement naturel consistera dès lors à maintenir l'approximation gaussienne de ces estimateurs sur un domaine plus large de vitesses de convergence, à travers un ensemble de *principes de déviations modérées*. Cela fera l'objet du quatrième chapitre, dont la portée assez théorique orientera l'étude vers un renforcement des résultats de convergence et achèvera par là-même notre travail sur cette classe de processus.

Nous avons considéré jusqu'alors et de manière tacite des processus à temps discret, pour lesquels deux observations consécutives sont espacées d'un pas de temps unitaire. Dans un cadre stochastique et par l'intermédiaire d'équations différentielles, nous savons qu'il est généralement possible de bâtir des processus à temps continu dont les propriétés sont comparables à celles des processus autorégressifs. À titre d'exemple, c'est aux travaux fondateurs d'Ornstein et Uhlenbeck [129] dès 1930 que l'on doit le célèbre processus à temps continu du même nom, solution d'une équation différentielle stochastique, dont le parallèle avec le modèle autorégressif du premier ordre est assez intuitif. Les chocs aléatoires perturbant le *processus d'Ornstein-Uhlenbeck* sont issus d'un *mouvement brownien*, et même si diverses généralisations ont vu le jour, ces derniers restent généralement à accroissements indépendants. Le cinquième chapitre est ainsi dévolu au processus d'Ornstein-Uhlenbeck dirigé par un processus d'Ornstein-Uhlenbeck. Cette étude est motivée par l'intérêt en économétrie des processus d'Ornstein-Uhlenbeck dirigés par des processus de Lévy, et nous ferons également le parallèle avec le modèle autorégressif à erreurs autocorrélées. Nous serons amenés à constater que les estimateurs du maximum de vraisemblance d'un tel modèle présentent certaines similitudes asymptotiques avec leurs équivalents discrets que nous avons pu détailler au cours des chapitres précédents. D'une manière générale, cela nous permettra d'aborder la problématique de l'autocorrélation résiduelle dans un cadre de processus à temps continu et de nous familiariser avec la notion d'*instabilité*. Ce concept sera par ailleurs le centre d'intérêt du sixième et ultime chapitre de cette thèse.

La plupart des propriétés asymptotiques usuelles liées aux séries chronologiques linéaires reposent sur la nature du processus générateur sous-jacent. C'est ainsi qu'un modèle possédant des propriétés stabilisatrices – sommairement lorsque le comportement global de la série finit par se décorrélérer totalement de l'indice temporel – est de fait un sujet d'étude statistique beaucoup plus aisé à manipuler. On parlera dès lors de *série stationnaire*, par opposition avec une *série intégrée* dont le processus générateur possède une *racine unitaire*. Les procédures visant à tester l'existence d'une telle racine dans la structure d'une série observée, source d'instabilité, ont fait et font encore l'objet d'une abondante littérature, que nous nous attacherons à résumer aussi fidèlement que possible une fois le sixième chapitre abordé. Cependant, notons que les procédures évaluant l'hypothèse nulle de stationnarité sont de manière surprenante assez peu répandues. Nous nous sommes à cet égard largement inspirés des travaux de Kwiatkowski, Phillips, Schmidt et Shin [73] de 1992 d'une part, et de ceux de Leybourne et McCabe [82] de 1994 d'autre part. De nos jours, ces stratégies sont communément utilisées dans le prétraitement des séries chronologiques, particulièrement la procédure bien connue sous la dénomination de *test KPSS*. Nous nous proposons ainsi d'étendre la procédure de Ley-

bourne et McCabe à une classe plus large de processus, munis d'une tendance polynomiale d'ordre quelconque et modélisant des séries intégrées d'ordre supérieur. Il sera de plus aisé de constater que notre étude traite un aspect de la non stationnarité certes plus rare, mais qui faisait encore défaut aux procédures existantes. Si l'approche martingale constituait l'outil principal des chapitres précédents, nous ferons ici un usage intensif des *principes d'invariance* de type Donsker.

Concluons cet avant-propos en précisant que de nombreux graphes obtenus en simulations, que nous nous appliquerons à dûment commenter, viendront illustrer nos principaux résultats. Il est hélas délicat d'obtenir une évidence statistique de la bonne performance de nos procédures sur des données réelles dans la mesure où il est simplement impossible d'accéder au véritable processus générateur à l'origine de ces données, qui est et restera inconnu. Nous ajouterons cependant, en guise d'annexe, la description assez succincte d'une étude menée avec l'équipe ICAME d'EDF R&D. Le thème principal, cette fois très appliqué, traite de la prédiction de la consommation individuelle d'électricité. Cette annexe n'est certes pas directement lié au contenu théorique de ce mémoire, mais nous y abordons les problématiques usuelles des séries chronologiques linéaires du point de vue de l'ingénieur. Il s'agit, nous semble-t-il, d'un contrepois inéluctable et d'un exemple de mise en pratique. Une seconde annexe jouera le rôle d'aide-mémoire quant aux théorèmes principaux auxquels nous aurons recours. Leur utilisation récurrente nous a en effet incités à proposer quelques rappels explicites censés simplifier la lecture de cette thèse et centraliser les outils fondamentaux.

# Glossaire

Notation	Signification
$M'$	Transposée de la matrice $M$ .
$\det(M)$	Déterminant de la matrice carrée $M$ .
$\text{tr}(M)$	Trace de la matrice carrée $M$ .
$\rho(M)$	Rayon spectral de la matrice carrée $M$ .
$\lambda_{\min}(M) / \lambda_{\max}(M)$	Plus petite/grande valeurs propres de la matrice carrée $M$ .
$\ v\  / \ v\ _1 / \ v\ _\infty$	Normes euclidienne/1/infinie du vecteur $v$ .
$ s $	Module de $s$ .
$[s]$	Partie entière de $s$ .
$s \wedge t / s \vee t$	Minimum/Maximum entre $s$ et $t$ .
$\mathcal{N}(\mu, \sigma^2)$	Loi gaussienne d'espérance $\mu$ et de variance $\sigma^2$ .
$\mathcal{U}([a, b])$	Loi uniforme continue sur l'intervalle $[a, b]$ .
$\underset{\sim}{\text{iid}}$	Caractère indépendant et identiquement distribué d'une suite aléatoire.
$\longrightarrow$ p.s.	Convergence presque sûre.
$\xrightarrow{\mathcal{P}}$	Convergence en probabilité.
$\xrightarrow{\mathcal{L}}$	Convergence en loi.
$\xrightarrow{\text{exp}}$	Convergence exponentielle.
$\xrightarrow{\text{s-exp}}$	Convergence super-exponentielle.
$\stackrel{\mathcal{L}}{=}$	Égalité en loi.
$X_n = O(v_n)$ p.s.	$X_n$ est presque sûrement dominé par $v_n$ .
$X_n = o(v_n)$ p.s.	$X_n$ est presque sûrement négligeable devant $v_n$ .
$X_n = O_{\mathcal{P}}(v_n)$	$X_n$ est borné en probabilité par $v_n$ .
$X_n = o_{\mathcal{P}}(v_n)$	$X_n$ est négligeable en probabilité devant $v_n$ .
$\underset{\sim}{\sim}$ p.s.	Équivalence presque sûre.
$\underset{\sim}{\sim}^{\text{s-exp}}$	Équivalence super-exponentielle.
$\mathbb{I}_{\{e\}}$	Fonction indicatrice associée à la condition $\mathcal{C}$ .
$\Delta / \nabla$	Opérateurs de différenciation/sommation.
$L$	Opérateur retard.



---

## Sur les séries chronologiques linéaires

---

Nous avons dans l'introduction survolé les domaines d'application des séries chronologiques, il s'agit désormais de les formaliser. Ce chapitre est un bref panorama des principales définitions et propriétés des séries chronologiques, comme la notion de stationnarité qui est véritablement l'un des socles de l'étude dans son ensemble et d'une manière générale des séries chronologiques linéaires. Nous abordons ensuite la modélisation ARMA dont le processus autorégressif est, comme nous le verrons, un cas particulier. Nous rappelons ses propriétés usuelles ainsi que celles de ses estimateurs des moindres carrés. Nous refermons ce chapitre par l'historique non exhaustif des séries chronologiques plus complexes.

### I.1 Quelques concepts fondamentaux

Nous introduisons succinctement, en guise de préambule, certains concepts fondamentaux que nous considérerons comme convenus tout au long du manuscrit. Ces derniers sont d'une manière générale intimement liés à la modélisation chronologique.

#### I.1.1 Le processus stochastique

Considérons un espace de probabilité  $(\Omega, \mathcal{A}, \mathbb{P})$ , un ensemble d'indices  $\mathcal{T}$  et un espace métrique  $\mathcal{S}$  muni de la tribu borélienne  $\mathcal{B}(\mathcal{S})$ .

**Définition I.1.** On appelle « processus stochastique » une famille de variables aléatoires  $(Y_t)$  définies sur  $(\Omega, \mathcal{A}, \mathbb{P})$ , indexées par  $t \in \mathcal{T}$  et à valeurs dans  $\mathcal{S}$ . Pour toute réalisation  $\omega \in \Omega$ , la famille  $(y_t = Y_t(\omega))$  est une « trajectoire » du processus.

C'est à la trajectoire d'un processus observé sur un sous-ensemble de  $\mathcal{T}$  que l'on associera la notion de « série chronologique ». Cependant, pour alléger les notations, nous ne distinguerons généralement pas le processus  $(Y_t)$  d'une de ses réalisations  $(y_t = Y_t(\omega))$ , lorsqu'il n'y aura aucune ambiguïté quant à la grandeur que l'on manipule. Nous serons par la suite amenés à travailler sur des processus discrets indexés par  $\mathbb{N}$  ou  $\mathbb{Z}$  et à valeurs réelles. Nous ferons d'ailleurs implicitement référence à cette sous-classe à travers le terme générique de « processus ». Au cours des deux derniers chapitres, nous traiterons également des processus continus indexés par  $\mathbb{R}^+$  et à valeurs réelles. L'ensemble  $\mathcal{T}$  sera dans notre cadre de travail l'espace des temps,  $\mathbb{N}$  ou  $\mathbb{Z}$ .

### I.1.2 La stationnarité

Sommairement, la stationnarité traduit la capacité d'un processus à ne pas dépendre de l'indice temporel. Ce dernier est dès lors entièrement décrit par sa loi stationnaire qui, par définition, n'évolue plus au cours du temps. On comprend ainsi qu'une telle propriété est certes d'intérêt pratique considérable, mais possède également un fort impact théorique puisqu'on la retrouve comme hypothèse à la base de nombreux résultats. On distingue généralement la stationnarité au sens strict de la stationnarité au sens faible. Pour les définir, considérons un processus  $(Y_t)$  défini sur  $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ , l'ensemble des variables aléatoires de carré intégrable.

**Définition I.2.** On dit que le processus  $(Y_t)$  est « strictement stationnaire » si, pour tout  $k \in \mathbb{N}^*$  et tout décalage temporel  $h \in \mathbb{Z}$ , la loi du vecteur  $(Y_1, \dots, Y_{1+k})$  est la même que celle du vecteur  $(Y_{1+h}, \dots, Y_{k+h})$ .

La stationnarité stricte est une hypothèse de travail très forte, nécessairement délicate à vérifier en pratique lorsque le processus n'est pas gaussien. C'est pourquoi l'on a introduit une notion de stationnarité moins contraignante.

**Définition I.3.** On dit que le processus  $(Y_t)$  est « faiblement stationnaire » si, pour tout décalage temporel  $h \in \mathbb{Z}$ ,  $\mathbb{E}[Y_0]$  et  $\mathbb{V}(Y_0)$  sont constants, et  $\text{Cov}(Y_0, Y_h)$  ne dépend que de  $h$ .

On dit aussi que le processus est « stationnaire au second ordre », en relation avec la stabilisation de sa variance. C'est à cette propriété de stationnarité que nous ferons implicitement référence par la suite. Notons que la stationnarité stricte implique bien entendu la stationnarité faible. L'exemple le plus trivial de processus stationnaire est le bruit blanc.

**Définition I.4.** Un processus  $(\varepsilon_t)$  est qualifié de « bruit blanc » si, pour tous  $t_1, t_2 \in \mathcal{T}$ , on a simultanément  $\mathbb{E}[\varepsilon_{t_1}] = 0$ ,  $\mathbb{V}(\varepsilon_{t_1}) = \sigma^2 < +\infty$  et  $\text{Cov}(\varepsilon_{t_1}, \varepsilon_{t_2}) = 0$  dès que  $t_1 \neq t_2$ .

Si  $(\varepsilon_t)$  est constitué de variables aléatoires indépendantes et identiquement distribuées, on parlera de « bruit blanc fort ». Il est tout à fait usuel de considérer le processus des chocs aléatoires perturbant un modèle statistique comme formant un bruit blanc, par hypothèse. Un test de blancheur

résiduelle est ainsi considéré comme un indicateur de qualité de la modélisation, dans la mesure où l'absence de corrélation résiduelle signifie que toute l'information statistique a bien été prise en compte dans la modélisation et que, de fait, la perturbation inobservable est de nature purement aléatoire. Par ailleurs, il est aisé d'établir que toute combinaison linéaire de variables aléatoires issues d'un bruit blanc est, au sens de la Définition I.3, stationnaire. Cela nous permet d'introduire la notion de moyenne mobile.

**Définition I.5.** Un processus  $(Y_t)$  est qualifié de « moyenne mobile » d'ordre  $q \in \mathbb{N}^*$  si, pour tout  $t \in \mathcal{T}$ ,

$$Y_t = \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \dots + \varphi_q \varepsilon_{t-q}$$

où  $\varphi \in \mathbb{R}^q$  est un paramètre réel, et  $(\varepsilon_t)$  forme un bruit blanc.

Les valeurs initiales sont généralement choisies dans  $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ , lorsque  $\mathcal{T}$  est fermé à gauche. Nous prendrons l'habitude de faire référence au processus moyenne mobile d'ordre  $q$  par son abréviation usuelle  $MA(q)$ , provenant de l'anglais *moving average*. De surcroît, il serait préférable, pour être tout à fait rigoureux, de parler ici de « stationnarité asymptotique ». Espérance et variance du processus ne se stabilisent en effet qu'une fois l'influence des variables initiales estompée, à moins d'être elles-mêmes générées selon la loi stationnaire. Par analogie avec la Définition I.3, cela se traduit par le fait que  $\mathbb{E}[Y_t]$ ,  $\mathbb{V}(Y_t)$  et  $\text{Cov}(Y_t, Y_{t+h})$  convergent vers des limites finies lorsque  $t$  grandit indéfiniment.

### I.1.3 L'autocorrélation

L'une des principales motivations de la modélisation chronologique d'un événement aléatoire est sa structure de corrélation temporelle, en d'autres termes son niveau d'autocorrélation. Il existe, entre autres, deux outils permettant d'évaluer l'autocorrélation d'une série chronologique. Ces derniers sont pour nous d'une importance capitale puisque nous verrons dans la section suivante qu'ils possèdent des propriétés remarquables dans le cadre du processus ARMA stationnaire. Restons pour le moment dans le cadre plus général du processus  $(Y_t)$  stationnaire, indexé par  $\mathcal{T}$  et défini sur  $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ .

#### I.1.3.1 La fonction d'autocorrélation

Cette première fonction quantifie l'influence linéaire du décalage temporel entre deux observations du processus par un calcul classique de corrélation.

**Définition I.6.** On appelle « fonction d'autocorrélation » associée au processus stationnaire  $(Y_t)$ , la fonction

$\rho$  définie, pour tout décalage temporel  $h \in \mathbb{N}$ , par

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

où  $\gamma$  est la « fonction d'autocovariance » donnée par  $\gamma(h) = \text{Cov}(Y_h, Y_0)$ .

Elle est en conséquence à valeurs dans  $[-1, 1]$ , par construction. Il est possible de lui donner une interprétation pour  $h \in \mathbb{Z}$ , elle forme alors une fonction paire. Lorsque l'on représente graphiquement l'évolution de  $\rho$  avec  $h$ , on construit un « autocorrélogramme ».

### I.1.3.2 La fonction d'autocorrélation partielle

Supposons désormais que l'on souhaite connaître l'influence exacte d'une observation passée sur la valeur courante du processus en ôtant de l'étude toutes les observations intermédiaires. Cette mesure de corrélation est plus délicate à interpréter, mais nous pouvons malgré tout la formaliser. Nous commençons à cet égard par bâtir la suite  $(\phi_{h,j})$  où l'indice  $h \in \mathbb{N}^*$  joue le rôle de décalage temporel et l'indice  $j$  varie entre 1 et  $h$ . On fixe  $\phi_{1,1} = \rho(1)$  et, pour tout  $h \geq 2$ , on construit

$$\phi_{h,h} = \left( 1 - \sum_{k=1}^{h-1} \phi_{h-1,k} \rho(k) \right)^{-1} \left( \rho(h) - \sum_{k=1}^{h-1} \phi_{h-1,k} \rho(h-k) \right)$$

avec, lorsque  $j < h$ ,  $\phi_{h,j} = \phi_{h-1,j} - \phi_{h,h} \phi_{h-1,h-j}$ . La valeur de  $\phi_{h,h}$  correspond au coefficient de corrélation linéaire entre les résidus issus de la régression affine de  $Y_h$  et de  $Y_0$  sur les observations intermédiaires  $Y_1, \dots, Y_{h-1}$ , respectivement. Formellement, pour tout  $h \in \mathbb{N}^*$ ,

$$\phi_{h,h} = \text{Corr}(Y_h - \bar{P}_{\text{sp}\{1, Y_1, \dots, Y_{h-1}\}} Y_h, Y_0 - \bar{P}_{\text{sp}\{1, Y_1, \dots, Y_{h-1}\}} Y_0) \quad (\text{I.1})$$

où  $\bar{P}_{\text{sp}\{1, Y_1, \dots, Y_{h-1}\}}$  désigne le projecteur orthogonal de toute variable aléatoire de  $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$  sur le sous-espace fermé engendré par  $\{1, Y_1, \dots, Y_{h-1}\}$ . Cette construction récursive est aussi appelée « algorithme de Durbin-Levinson » et l'on trouvera plus de détails à son sujet dans la Proposition 5.2.1 de [25].

**Définition I.7.** On appelle « fonction d'autocorrélation partielle » associée au processus stationnaire  $(Y_t)$ , la fonction  $\alpha$  définie par  $\alpha(0) = 1$  et, pour tout décalage temporel  $h \in \mathbb{N}^*$ , par

$$\alpha(h) = \phi_{h,h}.$$

Cette fonction est également à valeurs dans  $[-1, 1]$ . Lorsque l'on représente graphiquement l'évolution de  $\alpha$  avec  $h$ , on construit un « autocorrélogramme partiel ».

### I.1.4 La densité spectrale

Par affinité avec la section précédente, il nous semble judicieux de conclure ces rapides rappels par une présentation succincte de la densité spectrale des processus stationnaires, en lien étroit avec leurs propriétés de corrélation. En effet, il est bien connu que les autocovariances d'un processus stationnaire  $(Y_t)$  coïncident avec les coefficients de Fourier d'une mesure positive, appelée « mesure spectrale », que l'on notera  $\mu_Y$  et dont on peut supposer qu'elle admet une densité par rapport à la mesure de Lebesgue sur le tore  $\mathbb{T} = [-\pi, \pi]$ . Par souci de simplification, nous considérons ici les processus « à courte mémoire » dont la fonction d'autocovariance est absolument sommable, car ce sont précisément ceux qui nous occuperont durant la majeure partie de ce travail. Cela implique par ailleurs que  $\gamma(h)$  tend vers 0 lorsque  $h$  grandit.

**Définition I.8.** Soit un processus  $(Y_t)$  stationnaire de fonction d'autocovariance  $\gamma$ . Alors, pour tout  $\lambda \in \mathbb{T}$ , on appelle « densité spectrale » du processus, la fonction

$$f_Y(\lambda) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma(h) e^{-i\lambda h}.$$

De manière réciproque par transformation de Fourier inverse, on a également, pour tout  $h \in \mathbb{Z}$ ,

$$\gamma(h) = \int_{\mathbb{T}} f_Y(\lambda) e^{i\lambda h} d\lambda.$$

Nous pourrions disserter longtemps sur les séries chronologiques – fussent-elles limitées aux processus stationnaires – si d'aventure nous devions expliciter l'état de l'art en la matière. Ce n'est bien heureusement pas notre objectif. Il s'agit désormais pour nous d'introduire le processus ARMA et, par corollaire, le processus autorégressif. Nous souhaitons ainsi effectuer un tour d'horizon des résultats classiques existants, en nous appuyant très largement sur les travaux de Brockwell et Davis [25]–[26] dans les années 90, afin d'aborder le second chapitre avec un cadre de travail bien défini et tous les outils nécessaires à disposition.

## I.2 Le processus ARMA stationnaire

Commençons cette section par réduire le champ d'étude du processus stochastique tel qu'il est introduit dans la Définition I.1. Nous nous restreignons ici aux processus définis sur  $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ , indexés par  $\mathcal{T} = \mathbb{Z}$  et à valeurs dans  $\mathcal{S} = \mathbb{R}$ . Proposée par Box et Jenkins [19]–[20] en 1970, ainsi que nous l'avons évoqué dans l'introduction, la modélisation ARMA doit son nom à l'abréviation de l'anglais *autoregressive moving average*. L'hypothèse fondatrice de ces travaux pionniers stipule que la valeur courante du processus peut être entièrement expliquée par une combinaison linéaire

de ses  $p$  dernières valeurs, aléatoirement perturbée par une moyenne mobile d'ordre  $q$ , au sens de la Définition I.5. Tout d'abord, pour alléger les notations, on a souvent recours à trois opérateurs chronologiques : le retard noté  $L$  comme *lag*, la différenciation notée  $\Delta$ , et la sommation notée  $\nabla$ .

**Définition I.9.** On définit l'opérateur « retard », sur la classe des processus  $(Y_t)$  indexés par  $\mathbb{Z}$ , comme associant à toute valeur courante du processus sa valeur précédente,

$$LY_t = Y_{t-1}.$$

**Définition I.10.** On définit les opérateurs « différenciation » et « sommation », sur la classe des processus  $(Y_t)$  indexés par  $\mathbb{Z}$ , comme associant à toute valeur courante du processus les combinaisons

$$\Delta Y_t = (1 - L)Y_t = Y_t - Y_{t-1} \quad \text{et} \quad \nabla Y_t = (1 + L)Y_t = Y_t + Y_{t-1}.$$

Nous généralisons facilement ces opérateurs à tout décalage temporel  $h \in \mathbb{N}$  par l'intermédiaire de la convention  $L^h Y_t = (L \circ \dots \circ L)Y_t = Y_{t-h}$ . Ainsi,  $\Delta^h Y_t = (1 - L)^h Y_t$ ,  $\Delta_h Y_t = (1 - L^h)Y_t = Y_t - Y_{t-h}$ ,  $\nabla^h Y_t = (1 + L)^h Y_t$  et  $\nabla_h Y_t = (1 + L^h)Y_t = Y_t + Y_{t-h}$ .

## I.2.1 Le modèle et ses propriétés usuelles

Soit  $(Y_t)$  un processus de  $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$  stationnaire indexé par  $\mathbb{Z}$ , et deux paramètres  $(p, q) \in \mathbb{N}$ .

**Définition I.11.** Le processus  $(Y_t)$  admet une représentation ARMA( $p, q$ ) si, pour tout  $t \in \mathbb{Z}$ , il est donné par la relation récursive

$$Y_t - \sum_{k=1}^p \theta_k Y_{t-k} = \mu + \sum_{k=1}^q \varphi_k \varepsilon_{t-k} + \varepsilon_t$$

où  $(\varepsilon_t)$  est un bruit blanc de variance  $\sigma^2 > 0$ ,  $\mu \in \mathbb{R}$ ,  $\theta \in \mathbb{R}^p$  et  $\varphi \in \mathbb{R}^q$ . Lorsque  $q = 0$ , le modèle admet la représentation AR( $p$ ) donnée, pour tout  $t \in \mathbb{Z}$ , par

$$Y_t = \mu + \sum_{k=1}^p \theta_k Y_{t-k} + \varepsilon_t$$

et l'on dit que  $(Y_t)$  est un « processus autorégressif » d'ordre  $p$ . Lorsque  $p = 0$ , le modèle admet la représentation MA( $q$ ) donnée, pour tout  $t \in \mathbb{Z}$ , par

$$Y_t = \mu + \sum_{k=1}^q \varphi_k \varepsilon_{t-k} + \varepsilon_t$$

et l'on dit que  $(Y_t)$  est une « moyenne mobile » d'ordre  $q$ .

Nous pouvons en outre, sans aucune perte de généralité, considérer que le processus  $(Y_t)$  est centré et donc que  $\mu = 0$ . En effet, si l'on appelle  $m$  l'espérance de la loi stationnaire, on a

$$m = \frac{\mu}{1 - \sum_{k=1}^p \theta_k}$$

et l'on voit alors immédiatement que le changement de variables  $Z_t = Y_t - m$  nous ramène par simple translation au modèle ARMA( $p, q$ ) centré, engendrant  $(Z_t)$ . Notons également que nous avons défini la moyenne mobile dans la section précédente, à travers la Définition I.5, dans un cadre plus général. De manière condensée, à l'aide de l'opérateur retard, le modèle s'écrit encore, pour tout  $t \in \mathbb{Z}$ ,

$$\mathcal{A}(L)Y_t = \mathcal{B}(L)\varepsilon_t \tag{I.2}$$

où l'on définit, pour tout  $z \in \mathbb{C}$ , les polynômes  $\mathcal{A}$  et  $\mathcal{B}$  par

$$\mathcal{A}(z) = 1 - \theta_1 z - \dots - \theta_p z^p \quad \text{et} \quad \mathcal{B}(z) = 1 + \varphi_1 z + \dots + \varphi_q z^q.$$

Dans le cas du modèle AR( $p$ ), on a  $\mathcal{B}(z) = 1$  alors que symétriquement, dans le cas du modèle MA( $q$ ), on a  $\mathcal{A}(z) = 1$ . La modélisation est qualifiée de « minimale » si  $\theta_p \neq 0$ ,  $\varphi_q \neq 0$  et si  $\mathcal{A}$  et  $\mathcal{B}$  n'ont pas de racine commune. Sans cela, il est toujours possible de trouver une formulation ARMA( $p', q'$ ) équivalente avec  $p' \leq p$  et  $q' \leq q$  engendrant  $(Y_t)$ . En guise d'exemples, la Figure I.1 propose deux trajectoires sur  $\{1, \dots, 500\}$  d'une série autorégressive avec  $p = 1$ ,  $\mu = 0$ ,  $\theta_1 = -0.1$  et  $(\varepsilon_t) \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$  d'une part, puis avec  $p = 2$ ,  $\mu = -2$ ,  $\theta_1 = 0.2$ ,  $\theta_2 = -0.5$  et  $(\varepsilon_t) \stackrel{\text{iid}}{\sim} \mathcal{U}([-2, 2])$  d'autre part. Nous observons clairement l'incapacité de telles trajectoires à s'éloigner significativement de leur valeur moyenne, propriété directement liée à la stationnarité asymptotique des processus.

### I.2.1.1 Causalité et inversibilité

Soit  $(\lambda_i)$  l'ensemble des racines complexes du polynôme  $\mathcal{A}$ . Si  $|\lambda_i| > 1$  pour tout  $1 \leq i \leq p$ , alors la modélisation est « causale » et il est bien connu que son polynôme autorégressif est inversible. En accord avec la Définition 3.1.3 de [25], on sait de plus que la suite des coefficients du polynôme inverse est absolument sommable. On a dès lors les relations

$$\mathcal{A}^{-1}(z) = \sum_{k=0}^{\infty} a_k z^k \quad \text{et} \quad \mathcal{A}^{-1}(z)\mathcal{A}(z) = \sum_{k=0}^{\infty} a_k z^k \prod_{i=1}^p \left(1 - \frac{z}{\lambda_i}\right) = 1,$$

ce qui permet d'identifier la suite  $(a_k)$ , avec en particulier  $a_0 = 1$ . Il en résulte qu'un processus ARMA( $p, q$ ) causal admet une représentation MA( $\infty$ ) définie, pour tout  $t \in \mathbb{Z}$ , par

$$Y_t = \mathcal{A}^{-1}(L)\mathcal{B}(L)\varepsilon_t. \tag{I.3}$$

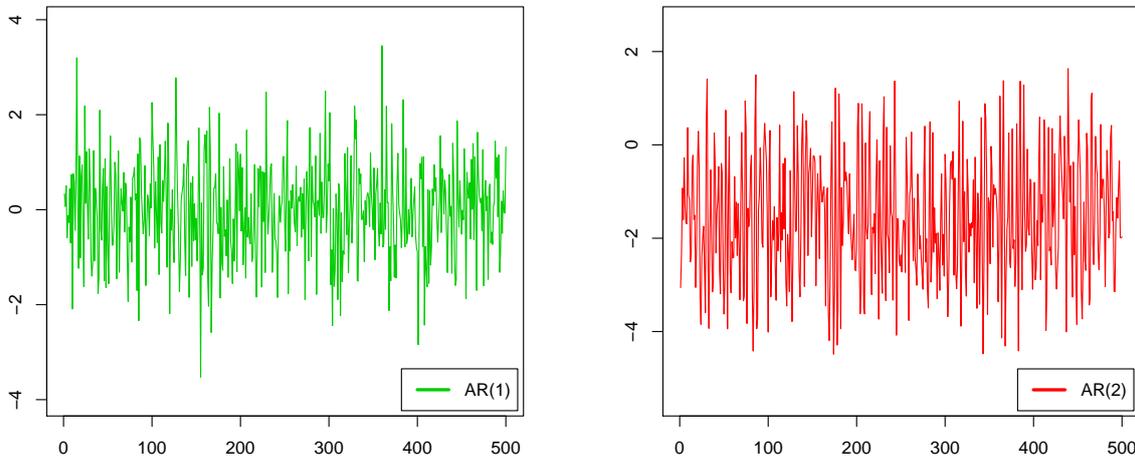


FIGURE I.1 – Exemples de trajectoires AR(1) stable (gauche), AR(2) stable (droite).

La causalité du polynôme autorégressif est ainsi une condition suffisante pour assurer la stationnarité du processus engendré par la modélisation ARMA, et nous avons de plus un résultat d'unicité de la solution (I.3) dont on trouvera la démonstration à la suite du Théorème 3.1.1 de [25] et de la Remarque 2 associée.

**Proposition I.1.** *Lorsque  $A(z) \neq 0$  pour tout  $z \in \mathbb{C}$  tel que  $|z| \leq 1$ , alors la représentation ARMA minimale  $A(L)Y_t = B(L)\varepsilon_t$  admet une unique solution stationnaire caractérisée par (I.3). De plus, les coefficients  $(\psi_k)$  du polynôme  $A^{-1}B$  vérifient*

$$\sum_{k=0}^{\infty} |\psi_k| < +\infty.$$

Par un raisonnement symétrique, on montre que la causalité du polynôme  $B$  implique que le modèle ARMA( $p, q$ ) admet une représentation AR( $\infty$ ) définie, pour tout  $t \in \mathbb{Z}$ , par

$$B^{-1}(L)A(L)Y_t = \varepsilon_t. \tag{I.4}$$

La modélisation est alors dite « inversible ». Cette dernière écriture peut trouver un intérêt pratique considérable dans la mesure où, comme nous le verrons, l'estimation des paramètres est toujours plus aisée dans un modèle purement autorégressif. Nous prendrons l'habitude de qualifier de « stable » un processus solution d'une modélisation ARMA dont le polynôme autorégressif est causal. Par opposition, le processus « instable » est issu d'une modélisation dont le polynôme autorégressif possède au moins une racine sur le cercle unité, et toutes les autres à l'extérieur. Lorsque l'une au moins des racines est située à l'intérieur du cercle unité, on parlera de processus « explosif », ce dernier sortant du cadre de notre étude.

### I.2.1.2 Autocorrélations

Évaluons désormais les fonctions d'autocorrélation simple et partielle, telles que décrites dans les Définitions I.6 et I.7, respectivement, dans le cas du processus stationnaire  $(Y_t)$  engendré par le modèle ARMA( $p, q$ ) causal. Nous considérons à cet égard l'écriture MA( $\infty$ ) donnée par (I.3) sous sa forme explicite, par l'intermédiaire d'une suite de coefficients réels  $(\psi_k)$ . On a

$$Y_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k} \quad (\text{I.5})$$

où les coefficients  $(\psi_k)$  sont identifiables à travers la relation polynomiale sur  $\mathbb{C}$ ,

$$\mathcal{A}^{-1}(z) \mathcal{B}(z) = \sum_{k=0}^{\infty} \psi_k z^k.$$

Cependant, à l'exception du cas trivial  $\psi_0 = 1$ , nous ne chercherons pas ici à les expliciter, dans la mesure où leur complexité analytique grandit très rapidement avec  $p$  et  $q$ . À titre d'exemple, pour le modèle ARMA(1, 1) dont la causalité est assurée par la condition  $|\theta_1| < 1$ , on a

$$\begin{aligned} (1 - \theta_1 z)^{-1} (1 + \varphi_1 z) &= (1 + \theta_1 z + \theta_1^2 z^2 + \theta_1^3 z^3 + \dots) (1 + \varphi_1 z) \\ &= \psi_0 + \psi_1 z + \psi_2 z^2 + \psi_3 z^3 + \dots, \end{aligned}$$

ce qui conduit après identification à  $\psi_0 = 1$ ,  $\psi_1 = \theta_1 + \varphi_1$ ,  $\psi_2 = \theta_1(\theta_1 + \varphi_1)$  et d'une manière générale à  $\psi_k = \theta_1^{k-1}(\theta_1 + \varphi_1)$  dès que  $k \geq 1$ . L'écriture (I.5) est dès lors particulièrement astucieuse dans l'évaluation de la fonction d'autocovariance  $\gamma$  du processus. En effet, puisque  $(\varepsilon_t)$  est un bruit blanc de variance  $\sigma^2$ , un calcul de covariance entraîne nombre de simplifications. Il est alors assez aisé d'établir que, pour tout décalage temporel  $h \in \mathbb{Z}$ ,

$$\gamma(h) = \sigma^2 \sum_{k=0}^{\infty} \psi_k \psi_{k+|h|} \quad \text{et} \quad \rho(h) = \frac{\sum_{k=0}^{\infty} \psi_k \psi_{k+|h|}}{\sum_{k=0}^{\infty} \psi_k^2}. \quad (\text{I.6})$$

Dans le cadre des processus MA( $q$ ) et AR( $p$ ), cette écriture se simplifie d'autant plus. Considérons tout d'abord la modélisation MA( $q$ ) pour laquelle  $\mathcal{A}(z) = 1$ , quel que soit  $z \in \mathbb{C}$ . Alors, on peut identifier  $\psi_0 = 1$ ,  $\psi_k = \varphi_k$  pour tout  $1 \leq k \leq q$  et  $\psi_k = 0$  sinon. Il s'ensuit que l'autocovariance et l'autocorrélation de la moyenne mobile d'ordre  $q$  sont données, pour  $|h| \leq q$ , par

$$\gamma(h) = \sigma^2 \sum_{k=0}^{q-|h|} \varphi_k \varphi_{k+|h|} \quad \text{et} \quad \rho(h) = \frac{\sum_{k=0}^{q-|h|} \varphi_k \varphi_{k+|h|}}{\sum_{k=0}^q \varphi_k^2} \quad (\text{I.7})$$

avec la convention que  $\varphi_0 = 1$ . En outre, dès que  $|h| > q$ , alors  $\gamma(h) = \rho(h) = 0$ . Cette propriété remarquable des processus MA( $q$ ) stationnaires se résume et se complète comme suit.

**Proposition I.2.** *Le processus stationnaire centré  $(Y_t)$  est engendré par une modélisation minimale MA( $q$ ) si et seulement si  $\rho(q) \neq 0$  et  $\rho(h) = 0$  pour tout  $|h| > q$ .*

Ce résultat est plus puissant que ce que nous avons rapidement vu ci-dessus dans la mesure où il fournit également l'aspect réciproque, à savoir qu'un processus stationnaire centré dont la fonction d'autocorrélation finit par s'annuler se comporte nécessairement comme une moyenne mobile. On se référera à la Proposition 3.2.1 de [25] pour la démonstration de ce résultat. Le processus AR( $p$ ), quant à lui, correspond à  $\mathcal{B}(z) = 1$  pour tout  $z \in \mathbb{C}$  et une écriture similaire à (I.6) peut être obtenue après identification, pas nécessairement plus simple puisqu'il s'agit toujours d'inverser le polynôme  $\mathcal{A}$ . Cependant, il existe une procédure plus astucieuse qui consiste à exploiter la relation linéaire entre les valeurs de  $\gamma(h)$  dans le cadre purement autorégressif. De fait, on a d'une part,

$$\begin{aligned}\gamma(0) &= \text{Cov}(\theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} + \varepsilon_t, Y_t) \\ &= \theta_1 \gamma(1) + \theta_2 \gamma(2) + \dots + \theta_p \gamma(p) + \sigma^2\end{aligned}$$

en vertu du caractère stationnaire de  $(Y_t)$  et de la parité de  $\gamma$ . D'autre part, pour tout  $1 \leq h \leq p$ , il est possible d'évaluer  $\gamma(h)$  en fonction de  $\gamma(0), \dots, \gamma(p)$  de la même manière. À titre d'exemple, cela donne, pour  $h = 1$ ,

$$\begin{aligned}\gamma(1) &= \text{Cov}(\theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} + \varepsilon_t, Y_{t-1}) \\ &= \theta_1 \gamma(0) + \theta_2 \gamma(1) + \dots + \theta_p \gamma(p-1).\end{aligned}$$

On construit ainsi un système de  $p + 1$  équations linéaires à  $p + 1$  inconnues se résumant à

$$\begin{pmatrix} 1 & -\theta_1 & -\theta_2 & \dots & \dots & -\theta_{p-1} & -\theta_p \\ -\theta_1 & 1 - \theta_2 & -\theta_3 & \dots & \dots & -\theta_p & 0 \\ \vdots & \vdots & \vdots & & & \vdots & \vdots \\ -\theta_p & -\theta_{p-1} & -\theta_{p-2} & \dots & \dots & -\theta_1 & 1 \end{pmatrix} \begin{pmatrix} \gamma(0) \\ \gamma(1) \\ \vdots \\ \gamma(p) \end{pmatrix} = \begin{pmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (\text{I.8})$$

Cette écriture matricielle est mieux connue comme étant les « équations de Yule-Walker » que l'on trouvera dans la Section 8.1 de [25], bien que sous une formulation duale. Nous savons par ailleurs, et nous le redémontrerons au cours du Chapitre III dans un cadre plus général, que ce système n'est pas singulier et qu'il admet une solution unique pour  $\gamma(0), \dots, \gamma(p)$  lorsque la condition de stabilité  $\|\theta\|_1 < 1$  est satisfaite, cette dernière entraînant de plus la causalité de la modélisation et donc la

stationnarité du processus. Le calcul de  $\gamma(p+1)$  est alors assuré par

$$\begin{aligned}\gamma(p+1) &= \text{Cov}(\theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} + \varepsilon_t, Y_{t-p-1}) \\ &= \theta_1 \gamma(p) + \theta_2 \gamma(p-1) + \dots + \theta_p \gamma(1),\end{aligned}$$

et ainsi de suite. La fonction d'autocorrélation est entièrement déterminée en renormalisant  $\gamma(h)$  par  $\gamma(0)$  pour tout  $h \in \mathbb{N}$ . L'espérance nulle et la fonction  $\gamma$  permettent de caractériser les moments d'ordre 2 de la loi stationnaire du processus. L'algorithme de Durbin-Levinson, résumé dans la Section I.1.3.2, décrit la construction pas à pas, à partir des valeurs de la fonction d'autocorrélation  $\rho$ , de celles de la fonction d'autocorrélation partielle  $\alpha$  qui ne possède pas d'écriture simplifiée dans le cas général, à l'exception du cadre autorégressif pur. On a alors le résultat suivant, par analogie avec la Proposition I.2, qui découle directement de l'expression (I.1) de  $\alpha(h)$  dans le cas d'un processus dont les coefficients de la représentation AR( $\infty$ ) finissent par s'annuler.

**Proposition I.3.** *Le processus stationnaire centré  $(Y_t)$  est engendré par une modélisation minimale AR( $p$ ) si et seulement si  $\alpha(p) \neq 0$  et  $\alpha(h) = 0$  pour tout  $h > p$ .*

Nous avons là encore l'aspect réciproque, stipulé par exemple dans le Théorème 3 de [115]. Les Propositions I.2 et I.3, à propos desquelles on pourra trouver plus d'information dans le Chapitre 3 de [25], sont communément utilisées en pratique pour évaluer l'ordre d'une moyenne mobile ou d'une autorégression, par l'étude visuelle des valeurs empiriques de  $\rho$  et de  $\alpha$ . Bien que nous ne puissions pas nous attarder plus sur le sujet, rappelons que de nombreux résultats existent en marge des fonctions d'autocorrélation (voir le Chapitre 7 de [25]) : estimateurs consistants, normalité asymptotique, intervalles de confiance, etc. On sait en outre que les conditions de stabilité  $\|\theta\|_1 < 1$  et  $\|\varphi\|_1 < 1$  sont suffisantes pour assurer la causalité des polynômes  $\mathcal{A}$  et  $\mathcal{B}$ , et qu'en conséquence,  $\alpha(h)$  et  $\rho(h)$  tendent exponentiellement vite vers 0 avec des motifs éventuellement sinusoïdaux, phénomène que les Chapitres 3 et 6 de [20] décrivent très bien. On connaît également leur comportement précis pour les petites valeurs de  $p$  et de  $q$ . Cela forme un outil visuel supplémentaire pour inférer la stationnarité du processus étudié et les ordres  $p$  et  $q$  d'une éventuelle modélisation ARMA.

### I.2.1.3 Densité spectrale

Un processus engendré par une modélisation ARMA( $p, q$ ), non nécessairement causale ni inversible, possède une densité spectrale telle que nous l'avons définie dans la Section I.1.4, quoique sous une forme simplifiée. Nous trouverons plus de détails ainsi que le résultat suivant tout au long du Chapitre 4 de [25].

**Proposition I.4.** *Soit le processus  $(Y_t)$  stationnaire engendré par la modélisation ARMA( $p, q$ ) minimale  $\mathcal{A}(L)Y_t = \mathcal{B}(L)\varepsilon_t$ , où  $(\varepsilon_t)$  est un bruit blanc de variance  $\sigma^2 > 0$ . Alors, pour  $\lambda \in \mathbb{T} = [-\pi, \pi]$ , sa densité*

spectrale est donnée par

$$f_Y(\lambda) = \frac{\sigma^2 |\mathcal{B}(e^{-i\lambda})|^2}{2\pi |\mathcal{A}(e^{-i\lambda})|^2}.$$

Le caractère stationnaire de  $(Y_t)$  est implicitement relié au fait que le polynôme  $\mathcal{A}$  ne s'annule pas sur le cercle unité, garantissant ainsi l'existence de  $f_Y(\lambda)$  sur tout le tore  $\mathbb{T}$ .

## I.2.2 Inférence pour les processus autorégressifs

Nous allons maintenant aborder l'aspect estimation des paramètres inconnus associés aux processus ARMA( $p, q$ ). Notons simplement qu'il est bien connu que l'estimateur du maximum de vraisemblance (sous l'approximation gaussienne) de  $\theta$  et de  $\rho$  d'un modèle ARMA causal et inversible est fortement consistant, en vertu par exemple du Théorème 10.8.1 de [25]. Il vérifie en outre une propriété de normalité asymptotique (voir le Théorème 10.8.2) dont la matrice de covariance peut être calculée explicitement, en relation étroite avec l'information de Fisher. La stratégie usuelle des moindres carrés n'est pas adaptée pour l'estimation du paramètre  $\varphi$  car ce dernier régit l'évolution d'un processus aléatoire inobservable. Elle est en revanche appropriée dans le cadre autorégressif pur pour lequel  $\mathcal{B}(z) = 1$ , et c'est ce à quoi nous allons désormais nous intéresser.

### I.2.2.1 L'estimateur des moindres carrés

Soit  $(Y_t)$  un processus de  $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$  tel que, pour tout  $t \in \mathbb{Z}$ , on a

$$Y_t = \theta_1 Y_{t-1} + \dots + \theta_p Y_{t-p} + \varepsilon_t \tag{I.9}$$

où  $(\varepsilon_t)$  est un bruit blanc de variance  $\sigma^2 > 0$ . On suppose de plus que  $\mathcal{A}(z) = 1 - \theta_1 z - \dots - \theta_p z^p$  est un polynôme causal pour  $z \in \mathbb{C}$ . En conséquence,  $(Y_t)$  est un processus autorégressif stationnaire d'ordre  $p$ , et l'on observe une trajectoire de ce processus sur l'intervalle de temps discret  $\{0, \dots, n\}$ . Celle-ci est donc entièrement décrite par la filtration  $\mathcal{F}_n = \sigma(Y_0, \dots, Y_{p-1}, \varepsilon_0, \dots, \varepsilon_n)$ , pour  $n \geq 0$ . On s'intéresse à la minimisation de la quantité

$$\Delta_n(\theta) = \sum_{t=0}^n \varepsilon_t^2 = \varepsilon_0^2 + \dots + \varepsilon_{p-1}^2 + \sum_{t=p}^n (Y_t - \theta_1 Y_{t-1} - \dots - \theta_p Y_{t-p})^2.$$

Dans ce cadre, l'estimateur des moindres carrés de  $\theta$  est donné, pour tout  $n \geq p$ , par

$$\widehat{\theta}_n = \left( \sum_{t=p}^n \Phi_{t-1}^p \Phi_{t-1}^{p'} \right)^{-1} \sum_{t=p}^n \Phi_{t-1}^p Y_t \quad (\text{I.10})$$

en notant  $\Phi_t^p = (Y_t \ Y_{t-1} \ \dots \ Y_{t-p+1})'$ , pour  $p-1 \leq t \leq n$ . Le caractère inversible de la matrice intervenant dans le calcul de  $\widehat{\theta}_n$  peut éventuellement être assuré par l'ajout de la matrice identité d'ordre  $p$ . L'estimateur des moindres carrés se comporte comme l'estimateur de Yule-Walker obtenu par l'intermédiaire des estimateurs consistants usuels des autocovariances  $\gamma$ , décrits dans la Section 7.2 de [25], puis par résolution du système (I.8). En fait, ils sont même égaux à l'exception des termes de bord qui, dans le cas stable, sont asymptotiquement négligeables. Plus précisément ils partagent dans ce cas la même normalité asymptotique.

### I.2.2.2 Consistance forte

Le résultat de consistance forte suivant est établi dans le Théorème 1 de [1], puis dans le Corollaire 1 de [75] dans un cadre plus général.

**Proposition I.5.** *Soit le processus  $(Y_t)$  stationnaire associé au modèle autorégressif (I.9) minimal causal, tel que  $\mathbb{E}[\varepsilon_0^2] = \sigma^2 < +\infty$ . Alors, on a la convergence presque sûre*

$$\lim_{n \rightarrow \infty} \widehat{\theta}_n = \theta \quad \text{p.s.}$$

Nous considérons ici que  $(\varepsilon_t)$  est un bruit blanc, mais le résultat ci-dessus est également valable lorsque  $(\varepsilon_t)$  est une différence de martingale possédant un moment conditionnel d'ordre strictement supérieur à 2 fini. On montre en outre facilement que la variance empirique est un estimateur fortement consistant de  $\sigma^2$ . Plus précisément, on a

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=p}^n (Y_t - \widehat{\theta}_n' \Phi_{t-1}^p)^2 = \sigma^2 \quad \text{p.s.} \quad (\text{I.11})$$

La Figure I.2 illustre la convergence presque sûre des estimateurs des moindres carrés de  $\theta$  et de  $\sigma^2$  pour deux trajectoires centrées observées sur  $\{1, \dots, 500\}$ , respectivement avec  $p = 1$ ,  $\theta_1 = -0.6$  et  $(\varepsilon_t) \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 0.5)$ , puis avec  $p = 2$ ,  $\theta_1 = -0.3$ ,  $\theta_2 = 0.4$  et  $(\varepsilon_t) \stackrel{\text{iid}}{\sim} \mathcal{U}([-2, 2])$ .

### I.2.2.3 Normalité asymptotique

Comme indiqué précédemment, les estimateurs des moindres carrés et de Yule-Walker partagent, dans le cas stable, la même normalité asymptotique. On trouvera dès lors le résultat suivant ainsi que sa preuve dans la Section 8.10 de [25].

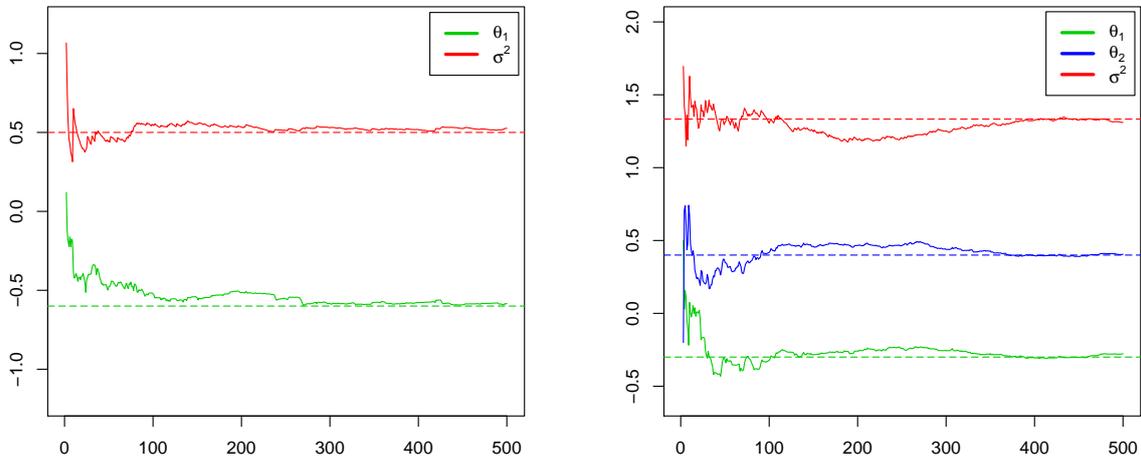


FIGURE I.2 – Convergence de  $\widehat{\theta}_n$  et de  $\widehat{\sigma}_n^2$  dans un AR(1) stable (gauche), dans un AR(2) stable (droite).

**Proposition I.6.** Soit le processus  $(Y_t)$  stationnaire associé au modèle autorégressif (I.9) minimal causal, tel que  $\mathbb{E}[\varepsilon_0^2] = \sigma^2 < +\infty$ . Alors, on a la normalité asymptotique

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2 \Gamma_p^{-1})$$

où  $\Gamma_p$  est la matrice de covariance d'ordre  $p$  du processus stationnaire.

Lorsque  $p = 1$ , on a  $\Gamma_1 = \gamma(0) = \sigma^2 / (1 - \theta_1^2)$  par simple résolution du système de Yule-Walker. Cela entraîne que la variance asymptotique dans le théorème central limite ci-dessus vaut  $1 - \theta_1^2$ . Nous reprenons ainsi l'exemple AR(1) de la Figure I.2 pour lequel  $\theta_1 = -0.6$ , et nous illustrons la normalité asymptotique de l'estimateur sur la Figure I.3 respectivement pour  $(\varepsilon_t) \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 0.5)$ , puis pour  $(\varepsilon_t) \stackrel{\text{iid}}{\sim} \mathcal{U}([-2, 2])$ , sur un échantillon de  $N = 1000$  réalisations.

Suite à cette présentation globale du processus ARMA stationnaire, qui nous sera précieuse dans les Chapitres II–IV, nous avons jugé utile de conclure en décrivant sommairement les évolutions majeures qu'a pu connaître ce dernier appliqué à la modélisation d'un panel plus vaste de phénomènes réels, en particulier leur capacité à tenir compte de la non stationnarité.

### I.3 Prise en compte de la non stationnarité

Par opposition à la Section I.1.2, un processus dont l'espérance et la variance ne se stabilisent pas simultanément au cours du temps est qualifié de « non stationnaire ». C'est une notion vaste et déli-

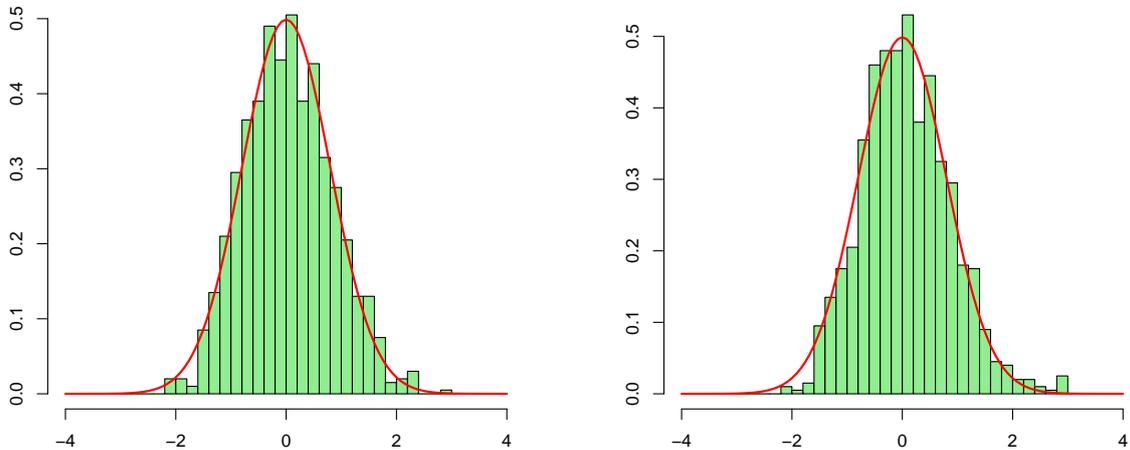


FIGURE I.3 – Exemples de normalité asymptotique de  $\widehat{\theta}_n$  dans un AR(1) stable.

cate à décrire de manière détaillée sans la réduire à « tout processus qui n'est pas stationnaire est non stationnaire ». Néanmoins, on peut distinguer deux phénomènes à l'origine de la non stationnarité d'un processus, selon sa nature déterministe ou stochastique.

### I.3.1 La non stationnarité déterministe

On regroupe sous cette dénomination les processus dont le comportement est impacté par un terme de nature déterministe. Un tel processus possède d'une manière générale, et bien que ce ne soit pas une condition nécessaire, une espérance qui dépend du temps. Nous citerons en exemple le processus indexé par  $\mathbb{N}$  à tendance linéaire donné, pour  $t \in \mathbb{N}$ , par

$$Y_t = a + bt + \varepsilon_t$$

où  $a \in \mathbb{R}$ ,  $b \in \mathbb{R}^*$  et  $(\varepsilon_t)$  forme un bruit blanc. Il est ici bien évident que  $\mathbb{E}[Y_t] = a + bt$  ne se stabilise pas dès que  $b \neq 0$ . On qualifiera ce processus de « stationnaire en tendance » dans la mesure où  $(Y_t)$  est effectivement stationnaire autour d'une tendance macroscopique que, dans la pratique, l'on s'appliquera à retirer. L'effet du choc aléatoire  $\varepsilon_t$  à l'instant  $t$  n'a de fait qu'une influence locale puisqu'il ne permet pas au processus de s'écarter significativement de sa tendance, à terme.

### I.3.2 La non stationnarité stochastique

Lorsque la perturbation aléatoire engendrant le processus empêche toute stabilisation de sa variance, on parle de non stationnarité de nature stochastique. L'exemple le plus répandu pour appréhender ce phénomène est le cas de la marche aléatoire.

**Définition I.12.** Un processus  $(Y_t)$  indexé par  $\mathbb{N}$  suit une « marche aléatoire » si, pour tout  $t \in \mathbb{N}^*$ , il vérifie

$$Y_t = Y_{t-1} + \varepsilon_t$$

où  $Y_0$  est arbitraire et  $(\varepsilon_t)$  forme un bruit blanc de variance  $\sigma^2 > 0$ .

Pour une valeur initiale  $Y_0$  choisie dans  $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$  indépendante de  $(\varepsilon_t)$ , il est facile de remonter jusqu'à l'expression explicite du processus. Ainsi,

$$Y_t = Y_0 + \sum_{k=1}^t \varepsilon_k$$

ce qui garantit certes une stationnarité en espérance, puisque  $\mathbb{E}[Y_t] = \mathbb{E}[Y_0]$ , mais le calcul direct nous montre que  $\mathbb{V}(Y_t) = \mathbb{V}(Y_0) + \sigma^2 t$ , source de non stationnarité en variance. Un tel processus est qualifié de « stationnaire en différence » car celle-ci est atteinte lorsque l'on considère le processus différencié  $(\Delta Y_t)$ . Ici, l'effet du choc aléatoire  $\varepsilon_t$  à l'instant  $t$  a une portée globale car ce dernier influence significativement le futur du processus, on le dit alors sensible à la persistance des chocs.

#### I.3.2.1 La série intégrée : le modèle ARIMA

Considérons un processus  $(Y_t)$  de  $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ , indexé par  $\mathbb{Z}$  et engendré par une modélisation ARMA( $p+1, q$ ) dont le polynôme autorégressif  $\mathcal{A}$  possède une racine sur le cercle unité, et  $p$  racines en dehors. On parle alors de « série intégrée » dans la mesure où l'écriture condensée de ce processus peut se factoriser sous la forme

$$(1 - \theta_0 L) \mathcal{A}^*(L) Y_t = \mathcal{B}(L) \varepsilon_t$$

où le polynôme  $\mathcal{A}^*$  d'ordre  $p$  est causal et  $|\theta_0| = 1$ . Le polynôme défini par  $\mathcal{A}(z) = (1 - \theta_0 z) \mathcal{A}^*(z)$  n'étant pas inversible sur  $\mathbb{C}$ , il n'existe pas de solution stationnaire à la modélisation proposée ci-dessus. En revanche, le modèle correctement différencié admet une solution stationnaire, en vertu de la Proposition I.1. Ainsi, les processus  $(\Delta Y_t)$  et  $(\nabla Y_t)$  seront engendrés par une modélisation ARMA( $p, q$ ) stationnaire, respectivement pour  $\theta_0 = 1$  et  $\theta_0 = -1$ . Lorsque la racine 1 est de multiplicité  $d \in \mathbb{N}^*$ , on parle alors de « modélisation ARIMA( $p, d, q$ ) », I comme *integrated*. Littéralement, celle-ci signifie que le processus doit être différencié  $d$  fois pour être stationnarisé. Il est d'ailleurs facile de voir que la marche aléatoire, telle qu'elle est introduite dans la Définition I.12, est la restriction à  $\mathbb{N}$  d'un processus ARIMA(0, 1, 0).

Cependant, si la consistance forte de l'estimateur des moindres carrés du paramètre  $\theta$  dans un processus autorégressif intégré reste valable (voir par exemple [76]), en revanche ce n'est plus le cas de la normalité asymptotique. À titre d'exemple, pour la marche aléatoire introduite dans la Définition I.12, il est établi dans [39], puis d'une manière plus élégante dans [31] sous une condition d'existence de moments d'ordre strictement supérieur à 2 pour la perturbation  $(\varepsilon_t)$ , la convergence

$$n(\widehat{\theta}_n - 1) \xrightarrow{\mathcal{L}} \frac{\int_0^1 W_s dW_s}{\int_0^1 W_s^2 ds} \quad (\text{I.12})$$

où  $(W_t)$  est un mouvement brownien standard, que nous allons maintenant définir.

**Définition I.13.** On appelle « mouvement brownien standard » (ou « processus de Wiener ») un processus stochastique  $(W_t)$  indexé par  $t \in \mathbb{R}^+$ , à valeurs dans  $\mathbb{R}$ , caractérisé par  $W_0 = 0$ , dont les accroissements sont indépendants et tels que  $W_t - W_s$  suit la distribution  $\mathcal{N}(0, t - s)$  pour  $0 \leq s \leq t$ .

Par opposition avec la Proposition I.6, on voit donc qu'il existe une rupture dans la distribution asymptotique de l'estimateur des moindres carrés dès qu'un zéro du polynôme autorégressif se trouve sur le cercle unité, et ce au niveau de la distribution asymptotique – qui perd la propriété gaussienne – comme au niveau de la vitesse de convergence. Nous aurons l'occasion de revenir sur de telles convergences caractéristiques du cadre instable dans les Chapitres V–VI. Sur la Figure I.4, nous avons représenté quelques trajectoires d'une marche aléatoire gaussienne de variance 0.5 ainsi que la distribution de l'estimateur  $\widehat{\theta}_n$  convenablement renormalisé. Nous constatons l'asymétrie de la loi asymptotique dont la valeur théorique est donnée par (I.12), ainsi que la rupture manifeste avec les illustrations relatives au cas stable. Notons que la distribution (pseudo-)théorique superposée à l'histogramme est issue d'une large simulation de la décomposition de Karhunen-Loève de la loi limite (Corollaire 3.1.3 de [31]) que nous reverrons plus en détail dans le Chapitre V.

### I.3.2.2 La série saisonnière : le modèle SARIMA

L'analyse spectrale d'une série chronologique est susceptible de mettre à jour des fréquences privilégiées, source de périodicité. C'est en particulier le cas pour certains phénomènes naturels tels que les marées et la température de surface, ou des comportements humains tels que la consommation d'électricité, les embouteillages, etc. Si nous considérons l'exemple concret de la fréquence journalière sur un processus  $(Y_t)$  indexé par les heures, alors un modèle autorégressif permet effectivement de quantifier le poids de l'observation  $Y_{t-24}$  sur la valeur courante  $Y_t$ . Les limites du modèle AR(24) apparaissent dès que ce dernier est écrit dans son intégralité, avec un paramètre  $\theta$  de dimension 24 quand un paramètre de dimension 1 ou 2 suffirait à expliquer l'intégralité du phénomène. Le modèle SARIMA( $p, d, q$ )  $\times$  ( $P, D, Q$ ) $_s$  a été introduit dans le but de modéliser un comportement ARMA stationnaire sur une série éventuellement intégrée et munie d'une périodicité  $s \in \mathbb{N}^*$ , ce pourquoi

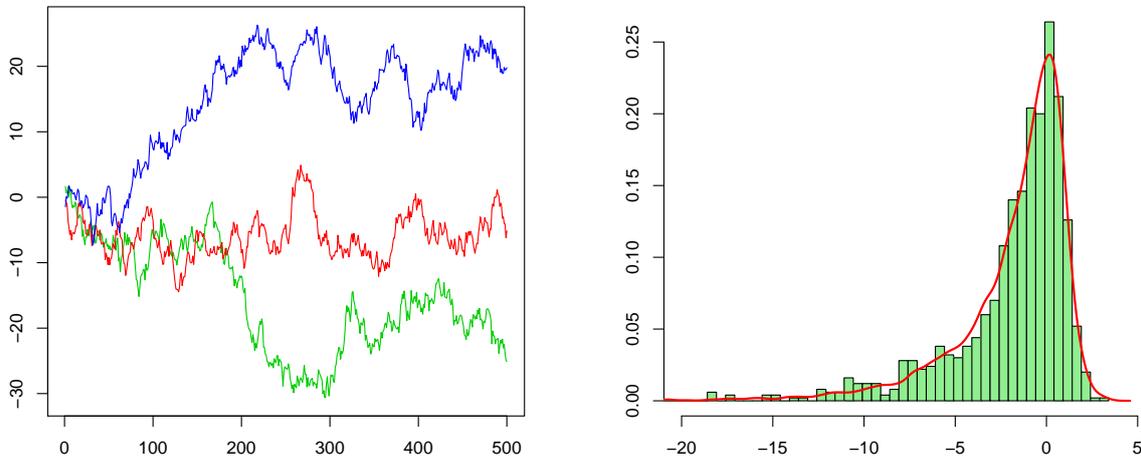


FIGURE I.4 – Exemples de marches aléatoires et distribution asymptotique de  $\widehat{\theta}_n$ .

l'on ajoute un S comme *seasonal*. Le processus est alors défini, pour tout  $t \in \mathbb{Z}$ , par

$$\mathcal{A}_s(L)\mathcal{A}(L)(1-L^s)^D(1-L)^d Y_t = \mathcal{B}_s(L)\mathcal{B}(L)\varepsilon_t$$

où, pour  $z \in \mathbb{C}$ ,  $\mathcal{A}_s(z) = 1 - \alpha_1 z^s - \dots - \alpha_p z^{Ps}$  et  $\mathcal{B}_s(z) = 1 + \beta_1 z^s + \dots + \beta_Q z^{Qs}$ , avec les paramètres  $\alpha \in \mathbb{R}^P$  et  $\beta \in \mathbb{R}^Q$ , et par analogie avec le modèle ARIMA, le triplet  $(P, D, Q) \in \mathbb{N}$ . On voit qu'une telle modélisation permet d'affecter un poids spécifique aux observations cycliques, tout en tenant compte d'une éventuelle multi-intégration du processus. On considère généralement que le polynôme produit  $\mathcal{A}_s \mathcal{A}$  est causal, ce qui se réduit à la causalité de chacune de ses composantes. De nombreux résultats existent également au sujet de la modélisation SARIMA, on pourra se référer à la Section 9.6 de [25], à la Section 3.9 de [121] ou au Chapitre 9 de [20]. Nous aurons quant à nous l'opportunité d'y revenir sommairement dans l'Annexe A.

### I.3.2.3 Diverses évolutions

Les modèles que nous avons présentés jusqu'alors sont purement chronologiques, ils s'appuient en effet sur le seul passé de la série. Il est possible d'y adjoindre un processus aléatoire jouant un rôle externe, on parle alors de modèles SARIMAX, X comme *exogenous*. Ces modélisations permettent de faire le lien entre les séries chronologiques et les modèles de régression linéaire standards. Nous pouvons également penser aux processus « hétéroscédastiques » pour lesquels la perturbation n'est plus de variance constante. Développé par Engle [53] en 1982, le modèle ARCH, pour *autoregressive conditional heteroskedasticity*, propose de tenir compte de la volatilité comme d'une fonction linéaire des

carrés des valeurs passées du processus. En 1986, Bollerslev [18] reprend la modélisation ARCH et ajoute un comportement autorégressif à la volatilité, on parle alors de modèles GARCH, pour *general ARCH*. Ces processus, qui atteignent rapidement un haut degré de complexité, sont communément utilisés en finance, au sein de laquelle l'hétéroscédasticité – et de fait la volatilité – jouent un rôle majeur. Nous pouvons également citer les modèles à seuil, les modèles fonctionnels, les modèles à bifurcation, les modèles fractionnaires, etc.

Au cours de la fin du siècle dernier, une très vaste littérature s'est développée sur les extensions du modèle de Box et Jenkins. Tous nos exemples ne sont bien entendu que des cas particuliers censés illustrer les deux types de non stationnarité usuels, qui peuvent par ailleurs se combiner – pensons à la marche aléatoire avec dérive linéaire. Les procédures statistiques ont bien souvent pour but de rechercher la stationnarité des processus par des transformations de données, et appliquer ainsi des résultats fidèles à la théorie : transformations stabilisatrices de variance (de type Box-Cox), désaisonnalisation (Fourier, ondelettes, splines), stationnarisation (différenciation, sommation), etc. Nous pensons avoir posé, de manière certes sommaire mais suffisante pour la compréhension de ce manuscrit, les bases de notre étude. Il est désormais temps pour nous d'aborder le Chapitre II.



---

# *Sur la statistique de Durbin-Watson*

---

Les résultats que nous allons présenter dans ce chapitre ont été obtenus avec la collaboration et l'aide précieuse de Bernard Bercu, ils sont également résumés dans [10] sous un point de vue très similaire. Nous y abordons la problématique de l'autocorrélation résiduelle.

## **II.1 Historique et motivation**

La statistique de Durbin-Watson est très répandue dans les domaines de l'économétrie et de la finance. Elle fut à l'origine introduite par les travaux pionniers de Durbin et Watson [49]–[50]–[51] dans les années 50, sous l'aspect d'un rapport de formes quadratiques inspiré de résultats antérieurs de Von Neumann [130] sur les petits échantillons. Sa motivation première était de tester l'absence de corrélation entre deux résidus successifs issus d'une régression linéaire classique, et de pouvoir ainsi garantir avec une certaine évidence statistique l'optimalité de la modélisation sous-jacente. Du reste, la procédure de test, dont la puissance ne sera étudiée en théorie comme en pratique par Tillman [128] que bien plus tard, en 1975, sous des hypothèses certes quelque peu restrictives, donnait et donne encore aujourd'hui de très bons résultats dans le cadre du modèle linéaire standard. En revanche, on savait depuis quelques années déjà que la présence de variables endogènes dans le modèle de régression était susceptible de biaiser les résultats. Ce phénomène a en effet été constaté par Malinvaud [89] tout d'abord, puis par Nerlove et Wallis [94], durant les années 60, et il met en garde contre l'utilisation abusive de la procédure de Durbin-Watson dans les modèles dynamiques, les conclusions étant manifestement faussées. Conscient de cette limitation, Durbin [47] propose en 1970 un ensemble d'améliorations de la procédure censées intégrer le caractère dépendant de la variable explicative, que Maddala et Rao [88] puis Park [103] s'appliqueront à mettre en pratique et dont ils constateront qu'elles surclassent effectivement la procédure dans sa version originale, et ce

également sur les petits échantillons. Les fondations pertinentes posées, c'est dans les années 80 que Inder [66]–[67], puis Durbin [48] lui-même, approfondissent l'étude des valeurs critiques des tests alternatifs, sous l'hypothèse nulle d'absence d'autocorrélation résiduelle. Par la suite, King et Wu [70], et très récemment Stocker [125] en 2007, apportent une pierre de plus à l'édifice en explorant respectivement l'influence de l'autocorrélation résiduelle dans les modèles dynamiques et le biais induit dans le comportement des estimateurs des moindres carrés.

Notre motivation repose principalement sur deux aspects de la problématique. La première est l'absence de réelle adaptation de la statistique de Durbin-Watson au cadre autorégressif. On a certes développé des procédures annexes, et l'on sait que la statistique se trouve avec grande probabilité autour de 2 sous l'hypothèse nulle alors qu'elle s'écarte vers 0 ou 4 en cas de forte corrélation résiduelle, mais sa distribution asymptotique, sous l'hypothèse nulle comme sous l'hypothèse alternative, n'a jamais été dûment établie. Les tables de Durbin-Watson reposent d'ailleurs sur des expériences de type Monte-Carlo et le test dans sa version actuelle manque encore de rigueur : toute une plage du spectre des valeurs possibles ne permettent aucune conclusion quant à la corrélation résiduelle. Nous souhaitons ainsi proposer une analyse plus fine et mieux structurée de la statistique de Durbin-Watson dans un cadre autorégressif. Notre seconde motivation provient de l'utilisation très (trop) répandue des procédures de Box-Pierce [22] et de Ljung-Box [21] par les statisticiens souhaitant évaluer l'hypothèse de corrélation résiduelle. Ces dernières présentent certes le net avantage d'autoriser un test de significativité d'un nombre arbitraire de corrélations dans les résidus, mais nous verrons que la distribution asymptotique ne tient aucunement compte de la dynamique du modèle engendrant ces mêmes résidus et qu'en conséquence, l'hypothèse nulle est surestimée. Il sera enfin utile de comparer en simulation la puissance empirique de notre procédure de test avec d'autres procédures couramment usitées. Nos théorèmes seront prouvés dans la dernière section du chapitre, mais dans la mesure où certains résultats se résument à des corollaires issus du chapitre suivant, nous ne dupliquerons pas inutilement les preuves associées.

## II.2 Le modèle et ses estimateurs

Nous considérons un modèle autorégressif du premier ordre, lui-même issu d'une perturbation autorégressive du premier ordre. Soit  $(Y_t)$  le processus engendré, indexé par  $\mathbb{Z}$ , dont on suppose disposer d'une trajectoire observée sur  $\{0, \dots, n\}$ . Pour tout  $1 \leq t \leq n$ , le modèle est alors donné par

$$\begin{cases} Y_t = \theta Y_{t-1} + \varepsilon_t \\ \varepsilon_t = \rho \varepsilon_{t-1} + V_t \end{cases} \quad (\text{II.1})$$

où les paramètres inconnus vérifient les conditions  $|\theta| < 1$  et  $|\rho| < 1$  dites « de stabilité », et où  $(V_t)$  est un bruit blanc de variance  $\sigma^2 > 0$ . La trajectoire sera donc entièrement décrite par la filtration

$\mathcal{F}_n = \sigma(Y_0, \varepsilon_0, V_1, \dots, V_n)$ , pour  $n \geq 1$ . Nous prendrons l'habitude de qualifier de AR(1)–AR(1) une telle modélisation, et il est important de ne pas la confondre avec un modèle ARMA(1,1) qui ne possède pas la même dynamique. Il est en effet immédiat de constater que le modèle (II.1) est en réalité un AR(2) avec pour paramètres  $\theta_1 = \theta + \rho$  et  $\theta_2 = -\theta\rho$ . Par ailleurs, les racines du polynôme autorégressif valent  $1/\theta$  et  $1/\rho$ , ce qui entraîne la causalité du modèle, et donc la stabilité du processus sous les hypothèses retenues. Deux trajectoires de taille 500 sont simulées sur la Figure II.1, respectivement pour  $(\theta, \rho) = (0.5, 0.3)$  et  $(V_t) \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$ , puis pour  $(\theta, \rho) = (-0.2, 0.7)$  et  $(V_t) \stackrel{\text{iid}}{\sim} \mathcal{U}([-2, 2])$ . Ces exemples nous accompagneront tout au long de ce chapitre.

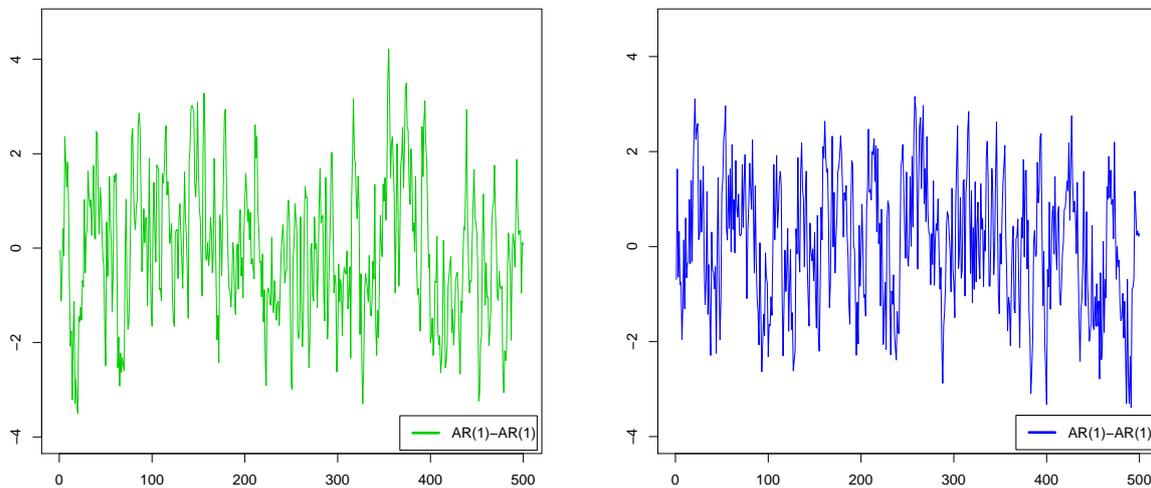


FIGURE II.1 – Exemples de trajectoires AR(1)–AR(1) stables.

Nous souhaitons estimer les paramètres  $\theta$  et  $\rho$  de manière non couplée. En effet, il serait possible de faire de l'inférence vectorielle sur le modèle (II.1) en travaillant directement sur sa formulation AR(2), mais ce n'est pas la philosophie de l'étude. Rappelons que l'objectif reste l'évaluation de l'hypothèse d'absence de corrélation résiduelle – soit  $\rho = 0$  – et qu'il s'agit ainsi pour nous de considérer l'estimation comme nous le ferions sous l'hypothèse nulle.

### II.2.1 Le paramètre de l'autorégression

L'estimateur des moindres carrés du paramètre  $\theta$  est obtenu par minimisation de la quantité

$$\Delta_n(\theta) = \sum_{t=0}^n \varepsilon_t^2 = \varepsilon_0^2 + \sum_{t=1}^n (Y_t - \theta Y_{t-1})^2$$

et l'on obtient naturellement, pour tout  $n \geq 1$ ,

$$\widehat{\theta}_n = \frac{\sum_{t=1}^n Y_{t-1} Y_t}{\sum_{t=1}^n Y_{t-1}^2}. \quad (\text{II.2})$$

### II.2.1.1 Convergence presque sûre

Avant d'aborder l'étude asymptotique de l'estimateur, nous devons rappeler qu'un résultat de convergence en probabilité de  $\widehat{\theta}_n$  vers la valeur limite

$$\theta^* = \frac{\theta + \rho}{1 + \theta\rho} \quad (\text{II.3})$$

est déjà établi dans [89] et [94]. Nous souhaitons tout d'abord renforcer cette convergence. Celle-ci sera illustrée par la suite sur la Figure II.3, relative aux deux exemples précités.

**Théorème II.1.** Soit  $\widehat{\theta}_n$  l'estimateur des moindres carrés donné par (II.2) dans le modèle (II.1) tel que  $\mathbb{E}[V_1^2] = \sigma^2 < +\infty$ . Alors, on a la convergence presque sûre

$$\lim_{n \rightarrow \infty} \widehat{\theta}_n = \theta^* \quad \text{p.s.}$$

où la valeur limite est donnée par (II.3).

### II.2.1.2 Normalité asymptotique

Étudions désormais la dispersion de l'estimateur autour de sa valeur limite. Pour cela, on commence par définir la variance

$$\sigma_{\theta}^2 = \frac{(1 - \theta^2)(1 - \theta\rho)(1 - \rho^2)}{(1 + \theta\rho)^3}. \quad (\text{II.4})$$

La normalité asymptotique sur un échantillon de  $N = 1000$  réalisations est également illustrée sur la Figure II.2 relative à nos exemples, pour  $n = 500$ .

**Théorème II.2.** Soit  $\widehat{\theta}_n$  l'estimateur des moindres carrés donné par (II.2) dans le modèle (II.1) tel que  $\mathbb{E}[V_1^4] = \tau^4 < +\infty$ . Alors, on a la normalité asymptotique

$$\sqrt{n} (\widehat{\theta}_n - \theta^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_{\theta}^2)$$

où la variance limite est donnée par (II.4).

On notera ici que les Théorèmes II.1 et II.2 sont bien compatibles avec les Propositions I.5 et I.6. En effet, lorsque  $\rho = 0$ , on a  $\theta^* = \theta$  et  $\sigma_{\theta}^2 = 1 - \theta^2$  ce qui correspond à la consistance forte et à la normalité asymptotique de l'estimateur dans un cadre autorégressif pur avec  $\rho = 1$ .

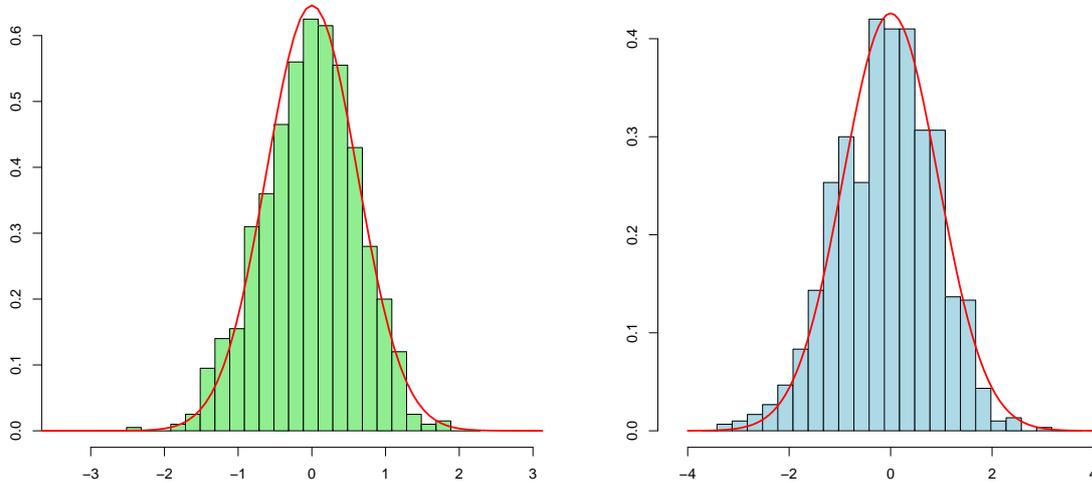


FIGURE II.2 – Normalité asymptotique de  $\widehat{\theta}_n$  pour deux AR(1)–AR(1) stables.

### II.2.1.3 Vitesse de convergence

Nous étudions enfin la vitesse de convergence presque sûre des bornes inférieure et supérieure de l'erreur  $\widehat{\theta}_n - \theta^*$ , ainsi que son taux d'accroissement.

**Théorème II.3.** Soit  $\widehat{\theta}_n$  l'estimateur des moindres carrés donné par (II.2) dans le modèle (II.1) tel que  $\mathbb{E}[V_1^4] = \tau^4 < +\infty$ . Alors, on a la loi forte quadratique

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{t=1}^n (\widehat{\theta}_t - \theta^*)^2 = \sigma_\theta^2 \quad \text{p.s.}$$

où la valeur limite est donnée par (II.4). On a de plus la loi du logarithme itéré

$$\limsup_{n \rightarrow \infty} \sqrt{\frac{n}{2 \log \log n}} (\widehat{\theta}_n - \theta^*) = -\liminf_{n \rightarrow \infty} \sqrt{\frac{n}{2 \log \log n}} (\widehat{\theta}_n - \theta^*) = \sigma_\theta \quad \text{p.s.}$$

Ce dernier résultat implique en particulier que

$$\limsup_{n \rightarrow \infty} \left( \frac{n}{2 \log \log n} \right) (\widehat{\theta}_n - \theta^*)^2 = \sigma_\theta^2 \quad \text{p.s.}$$

d'où l'on déduit la vitesse de convergence presque sûre

$$(\widehat{\theta}_n - \theta^*)^2 = O\left(\frac{\log \log n}{n}\right) \quad \text{p.s.} \tag{II.5}$$

*Démonstration.* Les Théorèmes II.1–II.2–II.3 sont des cas particuliers des Théorèmes III.1–III.2–III.3 respectivement, lorsque l'on choisit  $p = 1$ . ★

## II.2.2 Le paramètre de l'autocorrélation résiduelle

Il est tout d'abord nécessaire de bâtir un ensemble résiduel  $(\widehat{\varepsilon}_t)$ , sur lequel notre statistique de test pourra s'appuyer, censé refléter l'ampleur du biais occasionné par la présence d'une autocorrélation résiduelle sur la limite de  $\widehat{\theta}_n$ . Pour tout  $1 \leq t \leq n$ , on définit alors

$$\widehat{\varepsilon}_t = Y_t - \widehat{\theta}_n Y_{t-1} \quad (\text{II.6})$$

avec, de manière arbitraire,  $\widehat{\varepsilon}_0 = 0$ . Cela nous permet d'introduire un estimateur pour  $\rho$  donné, pour tout  $n \geq 1$ , par

$$\widehat{\rho}_n = \frac{\sum_{t=1}^n \widehat{\varepsilon}_{t-1} \widehat{\varepsilon}_t}{\sum_{t=1}^n \widehat{\varepsilon}_{t-1}^2}. \quad (\text{II.7})$$

Ce dernier a une interprétation cohérente au sens des moindres carrés, puisqu'il permet de minimiser

$$\Delta_n(\rho) = \sum_{t=1}^n (\widehat{\varepsilon}_t - \rho \widehat{\varepsilon}_{t-1})^2$$

tout en tenant compte du fait que  $(\widehat{\varepsilon}_t)$  sera une approximation d'autant meilleure de  $(\varepsilon_t)$  que  $\widehat{\theta}_n$  sera plus proche de  $\theta$ , ce qui se produit lorsque  $\rho$  est faible. On comprend dès lors que le comportement de  $\widehat{\rho}_n$  jouera un rôle capital dans le rejet de l'hypothèse nulle d'absence de corrélation résiduelle.

### II.2.2.1 Convergence presque sûre

Là encore, rappelons qu'un résultat de convergence en probabilité de  $\widehat{\rho}_n$  vers la valeur limite

$$\rho^* = \frac{\theta\rho(\theta + \rho)}{1 + \theta\rho} \quad (\text{II.8})$$

est déjà établi par exemple dans [89]. Nous souhaitons aussi renforcer cette convergence, puis l'illustrer sur la Figure II.3.

**Théorème II.4.** Soit  $\widehat{\rho}_n$  l'estimateur des moindres carrés donné par (II.7) dans le modèle (II.1) tel que  $\mathbb{E}[V_1^2] = \sigma^2 < +\infty$ . Alors, on a la convergence presque sûre

$$\lim_{n \rightarrow \infty} \widehat{\rho}_n = \rho^* \quad \text{p.s.}$$

où la valeur limite est donnée par (II.8).

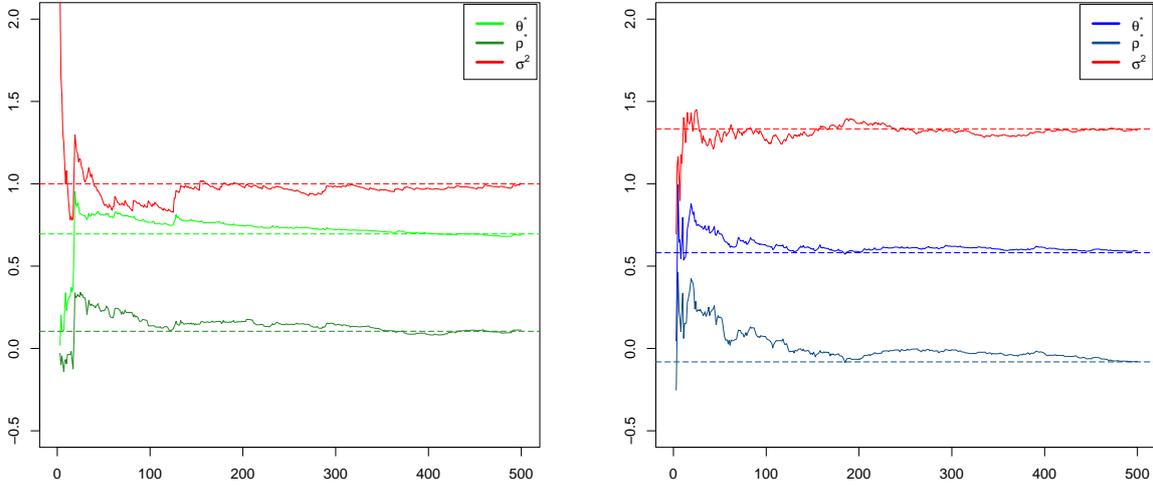


FIGURE II.3 – Convergence de  $\widehat{\theta}_n$ ,  $\widehat{\rho}_n$  et  $\widehat{\sigma}_n^2$  pour deux AR(1)–AR(1) stables.

### II.2.2.2 Normalité asymptotique

Pour obtenir la normalité asymptotique de  $\widehat{\rho}_n$ , nous allons commencer par établir la distribution asymptotique jointe de nos estimateurs. Ce résultat est de fait plus puissant, puisqu’il conduit également aux covariances asymptotiques. Soit la matrice de covariance

$$\Gamma = \begin{pmatrix} \sigma_\theta^2 & \theta\rho\sigma_\theta^2 \\ \theta\rho\sigma_\theta^2 & \sigma_\rho^2 \end{pmatrix} \quad (\text{II.9})$$

dans laquelle  $\sigma_\theta^2$  est donnée par (II.4), et

$$\sigma_\rho^2 = \frac{(1-\theta\rho)}{(1+\theta\rho)^3} \left( (\theta+\rho)^2(1+\theta\rho)^2 + (\theta\rho)^2(1-\theta^2)(1-\rho^2) \right). \quad (\text{II.10})$$

La Figure II.4 illustrera par la suite la normalité asymptotique de  $\widehat{\rho}_n$  sur un échantillon de  $N = 1000$  réalisations, toujours pour  $n = 500$ .

**Théorème II.5.** Soient  $\widehat{\theta}_n$  et  $\widehat{\rho}_n$  les estimateurs des moindres carrés donnés par (II.2) et (II.7) dans le modèle (II.1) tel que  $\mathbb{E}[V_1^4] = \tau^4 < +\infty$ . Alors, on a la normalité asymptotique jointe

$$\sqrt{n} \begin{pmatrix} \widehat{\theta}_n - \theta^* \\ \widehat{\rho}_n - \rho^* \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma)$$

où la matrice de covariance limite est donnée par (II.9). En particulier, on a la normalité asymptotique

$$\sqrt{n}(\widehat{\rho}_n - \rho^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_\rho^2)$$

où la variance limite est donnée par (II.10).

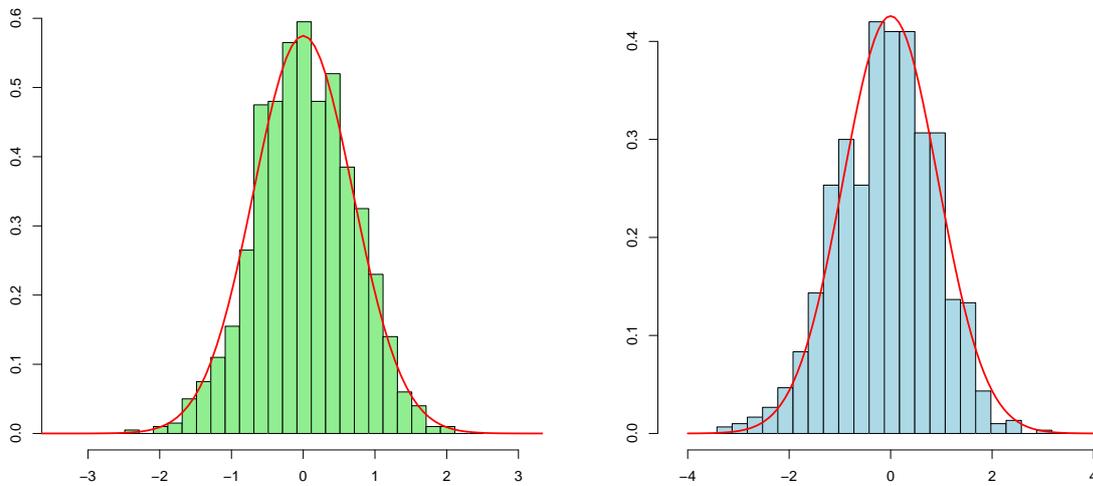


FIGURE II.4 – Normalité asymptotique de  $\widehat{\rho}_n$  pour deux AR(1)–AR(1) stables.

Par ailleurs, la matrice de covariance du Théorème II.5 est caractérisée par

$$\det(\Gamma) = \frac{(\theta + \rho)^2(1 - \theta\rho)\sigma_\theta^2}{1 + \theta\rho}$$

ce qui nous montre que lorsque  $\theta = -\rho$ , le théorème central limite, bien que toujours valable, est dégénéré. Il s’ensuit alors que  $\theta^* = 0$ ,  $\rho^* = 0$  et, par extension, que

$$\sqrt{n}\widehat{\theta}_n \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{1 + \theta^2}{1 - \theta^2}\right) \quad \text{et} \quad \sqrt{n}\widehat{\rho}_n \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\theta^4(1 + \theta^2)}{1 - \theta^2}\right).$$

Néanmoins, cette particularité restera un « cas pathologique » que nous serons amenés à traiter à part lors de l’étude empirique, et nous en expliquerons les raisons.

### II.2.2.3 Vitesse de convergence

Par analogie avec la Section II.2.1.3, nous concluons l’étude de l’estimateur de  $\rho$  par la vitesse de convergence presque sûre des bornes inférieure et supérieure de l’erreur  $\widehat{\rho}_n - \rho^*$ , ainsi que son taux d’accroissement.

**Théorème II.6.** Soit  $\widehat{\rho}_n$  l'estimateur des moindres carrés donné par (II.7) dans le modèle (II.1) tel que  $\mathbb{E}[V_1^4] = \tau^4 < +\infty$ . Alors, on a la loi forte quadratique

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{t=1}^n (\widehat{\rho}_t - \rho^*)^2 = \sigma_\rho^2 \quad \text{p.s.}$$

où la valeur limite est donnée par (II.10). On a de plus la loi du logarithme itéré

$$\limsup_{n \rightarrow \infty} \sqrt{\frac{n}{2 \log \log n}} (\widehat{\rho}_n - \rho^*) = -\liminf_{n \rightarrow \infty} \sqrt{\frac{n}{2 \log \log n}} (\widehat{\rho}_n - \rho^*) = \sigma_\rho \quad \text{p.s.}$$

Ce dernier résultat implique en particulier que

$$\limsup_{n \rightarrow \infty} \left( \frac{n}{2 \log \log n} \right) (\widehat{\rho}_n - \rho^*)^2 = \sigma_\rho^2 \quad \text{p.s.}$$

d'où l'on déduit la vitesse de convergence presque sûre

$$(\widehat{\rho}_n - \rho^*)^2 = O\left(\frac{\log \log n}{n}\right) \quad \text{p.s.} \quad (\text{II.11})$$

*Démonstration.* Les Théorèmes II.4–II.5–II.6 sont des cas particuliers des Théorèmes III.4–III.5–III.6 respectivement, lorsque l'on choisit  $p = 1$ . ★

### II.2.3 La variance résiduelle

Les estimateurs  $\widehat{\theta}_n$  et  $\widehat{\rho}_n$  sont autonormalisés, il s'ensuit que ni leur limite presque sûre ni leur variance asymptotique ne dépendent de  $\sigma^2$ . Il est pourtant d'intérêt statistique considérable d'estimer la variance résiduelle de façon consistante. Nous proposons alors d'utiliser l'estimateur donné, pour tout  $n \geq 1$ , par

$$\widehat{\sigma}_n^2 = \left(1 - \frac{\widehat{\rho}_n^2}{\widehat{\theta}_n^2}\right) \frac{1}{n} \sum_{t=1}^n \widehat{\varepsilon}_t^2. \quad (\text{II.12})$$

Nous avons illustré son comportement asymptotique sur la Figure II.3, dans la section précédente.

**Théorème II.7.** Soit  $\widehat{\sigma}_n^2$  l'estimateur donné par (II.12) dans le modèle (II.1) tel que  $\mathbb{E}[V_1^2] = \sigma^2 < +\infty$ . Alors, on a la convergence presque sûre

$$\lim_{n \rightarrow \infty} \widehat{\sigma}_n^2 = \sigma^2 \quad \text{p.s.}$$

*Démonstration.* Ce résultat est un cas particulier du Théorème III.7 lorsque l'on choisit  $p = 1$ . ★

## II.3 Application au test de Durbin-Watson

Lorsqu'elle fut introduite dans les années 50, la statistique de Durbin-Watson [49]–[50]–[51], appliquée à un ensemble résiduel  $(\widehat{\varepsilon}_t)$ , a été définie, pour tout  $n \geq 1$ , par

$$\widehat{D}_n = \frac{\sum_{t=1}^n (\Delta \widehat{\varepsilon}_t)^2}{\sum_{t=0}^n \widehat{\varepsilon}_t^2}. \quad (\text{II.13})$$

Tout d'abord, nous verrons dans la preuve associée au corollaire suivant que  $\widehat{D}_n$  est un équivalent asymptotique de  $2(1-\widehat{\rho}_n)$ , dans le cadre de stabilité qui nous intéresse ici. Il découle alors des résultats établis dans les Sections II.2.1 et II.2.2 le comportement asymptotique de  $\widehat{D}_n$ .

**Corollaire II.1.** Soit  $\widehat{D}_n$  la statistique de Durbin-Watson donnée par (II.13). Dès que  $\mathbb{E}[V_1^2] = \sigma^2 < +\infty$ , on a la convergence presque sûre

$$\lim_{n \rightarrow \infty} \widehat{D}_n = D^* \quad \text{p.s.}$$

où la valeur limite est définie par  $D^* = 2(1 - \rho^*)$ . De plus, dès que  $\mathbb{E}[V_1^4] = \tau^4 < +\infty$ , on a la normalité asymptotique

$$\sqrt{n} (\widehat{D}_n - D^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_D^2)$$

où la variance limite est définie par  $\sigma_D^2 = 4\sigma_\rho^2$ .

*Démonstration.* Le Corollaire II.1 est prouvé dans la Section II.6.2. ★

Il est également possible d'obtenir la vitesse de convergence presque sûre de  $\widehat{D}_n$  vers  $D^*$  par l'intermédiaire d'une loi forte quadratique et d'une loi du logarithme itéré similaires aux Théorèmes II.3 et II.6, mais notre intérêt ici repose exclusivement sur les implications du Corollaire II.1. C'est en particulier de la normalité asymptotique de  $\widehat{D}_n$  que nous tirons directement

$$\frac{n}{\sigma_D^2} (\widehat{D}_n - D^*)^2 \xrightarrow{\mathcal{L}} \chi_1^2 \quad (\text{II.14})$$

où  $\chi_1^2$  désigne une variable aléatoire distribuée selon une loi du khi-deux à un degré de liberté. Nous allons maintenant détailler une procédure statistique, appliquée au coefficient de corrélation  $\rho$ , généralisant et approfondissant le test bien connu de Durbin-Watson.

### II.3.1 Le cas pathologique

Tout d'abord, nous excluons de l'étude le cas où  $\theta = 0$ . En effet, nous admettons qu'un travail en amont a déjà fourni l'évidence statistique de la significativité d'une autocorrélation dans le processus observé. Nous avons par ailleurs mentionné l'existence d'un cas pathologique lorsque  $\theta = -\rho$  dans la

section précédente, il s'agit désormais d'en expliquer les raisons. Le caractère non inversible de  $\Gamma$  n'est en soi pas gênant puisque l'on a vu par la suite que les estimateurs, indépendamment l'un de l'autre, vérifient malgré cela un théorème central limite non dégénéré. Par contre, l'absence de corrélation résiduelle entraîne que  $D^* = 2$  et, par l'intermédiaire du Corollaire II.1, que  $\sqrt{n}(\widehat{D}_n - 2)$  possède une distribution asymptotique identifiable. Il s'ensuit qu'une procédure de test consistante doit être en mesure de garantir que  $D^* = 2$  si et seulement si  $\rho = 0$ , faute de quoi un ensemble d'interprétations alternatives existe. On vérifiera aisément que c'est hélas le cas lorsque  $\theta = -\rho$ , et c'est pourquoi nous considérons au préalable une procédure statistique permettant d'évaluer

$$\mathcal{H}_0 : " \theta = -\rho " \quad \text{vs} \quad \mathcal{H}_1 : " \theta \neq -\rho "$$

dans le modèle (II.1). Nous commençons par construire l'estimateur défini, pour tout  $n \geq 2$ , par

$$\widehat{\vartheta}_n = \frac{\sum_{t=2}^n Y_{t-2} Y_t}{\sum_{t=2}^n Y_{t-2}^2}, \tag{II.15}$$

ce qui nous conduit au résultat suivant.

**Théorème II.8.** *Supposons que  $\mathbb{E}[V_1^4] = \tau^4 < +\infty$ ,  $\theta \neq 0$  et  $\rho \neq 0$ . Alors, si l'on se place sous  $\mathcal{H}_0 : " \theta = -\rho "$ , on a la distribution asymptotique*

$$\frac{n(1 - \widehat{\vartheta}_n)}{4\widehat{\vartheta}_n^2(1 + \widehat{\vartheta}_n)} (\widehat{D}_n - 2)^2 \xrightarrow{\mathcal{L}} \chi_1^2$$

où  $\chi_1^2$  désigne une variable aléatoire distribuée selon une loi du khi-deux à un degré de liberté. De plus, si l'on se place sous  $\mathcal{H}_1 : " \theta \neq -\rho "$ , alors

$$\lim_{n \rightarrow \infty} \frac{n(1 - \widehat{\vartheta}_n)}{4\widehat{\vartheta}_n^2(1 + \widehat{\vartheta}_n)} (\widehat{D}_n - 2)^2 = +\infty \quad \text{p.s.}$$

Pour un niveau de significativité  $0 < \alpha < 1$ , on construit ainsi une zone d'acceptation  $\mathcal{A} = [0, z_\alpha]$  et une zone de rejet  $\mathcal{R} = ]z_\alpha, +\infty[$  à partir du quantile d'ordre  $1 - \alpha$  de la distribution du khi-deux considérée. Selon les conventions en vigueur associées aux tests d'hypothèse, on rejettera l'hypothèse nulle d'égalité entre  $\theta$  et  $-\rho$  dès que la statistique de test proposée sera supérieure à  $z_\alpha$ .

### II.3.2 Le cas général

Nous focalisons désormais notre attention sur le cas général. Nous souhaitons à cet égard tester la valeur de la corrélation résiduelle, et établir qu'avec une certaine marge d'erreur, il est possible de

considérer que  $\rho = \rho_0$ , pour un coefficient  $|\rho_0| < 1$ . De fait, nous évaluons

$$\mathcal{H}_0 : \text{“}\rho = \rho_0\text{”} \quad \text{vs} \quad \mathcal{H}_1 : \text{“}\rho \neq \rho_0\text{”}$$

dans le modèle (II.1). Supposons dans un premier temps qu’en vertu de la procédure statistique décrite par le Théorème II.8, nous soyons amenés à admettre que  $\theta = -\rho$ . Alors, et comme nous le verrons dans la preuve associée, tester  $\rho = \rho_0$  revient à tester que le coefficient de la modélisation AR(2) « à trou », c’est-à-dire avec un décalage temporel de 2, est significativement proche de  $\rho_0^2$ . Il s’ensuit que tester  $\mathcal{H}_0$  contre  $\mathcal{H}_1$  revient finalement à comparer  $\widehat{\vartheta}_n$ , défini en (II.15), avec la valeur qu’il estime presque sûrement sous  $\mathcal{H}_0$ , à savoir  $\rho_0^2$ . On utilise pour cela une légère variante, du reste très facile à établir, du théorème central limite rappelé dans la Proposition I.6. Cette dernière stipule que, sous  $\mathcal{H}_0$ , on a la convergence en loi

$$\frac{n}{1 - \rho_0^4} (\widehat{\vartheta}_n - \rho_0^2)^2 \xrightarrow{\mathcal{L}} \chi_1^2 \quad (\text{II.16})$$

alors que, sous  $\mathcal{H}_1$ ,  $\widehat{\vartheta}_n$  ne converge pas vers  $\rho_0^2$  et la statistique diverge nécessairement. Intéressons-nous maintenant au cas plus général dans lequel nous avons l’assurance statistique de considérer un processus engendré par la condition  $\theta \neq -\rho$ . On commence par noter, pour  $n \geq 1$ ,

$$\widetilde{\theta}_n = \widehat{\theta}_n + \widehat{\rho}_n - \rho_0, \quad \widetilde{\rho}_n = \frac{\widetilde{\theta}_n \rho_0}{1 + \widetilde{\theta}_n \rho_0} (\widetilde{\theta}_n + \rho_0) \quad \text{et} \quad \widetilde{D}_n = 2(1 - \widetilde{\rho}_n). \quad (\text{II.17})$$

Par application des Théorèmes II.1 et II.4, on obtient la convergence

$$\lim_{n \rightarrow \infty} \widetilde{\theta}_n = \theta^* + \rho^* - \rho_0 = \theta \quad \text{p.s.} \quad (\text{II.18})$$

valable sous  $\mathcal{H}_0$ . Nous disposons alors d’un estimateur consistant de  $\theta$  et de la valeur de  $\rho$  sous  $\mathcal{H}_0$ . On construit ainsi un estimateur de la matrice  $\Gamma$ , définie en (II.9), donné par

$$\widetilde{\Gamma}_n = \begin{pmatrix} \widetilde{\sigma}_{\theta,n}^2 & \widetilde{\theta}_n \rho_0 \widetilde{\sigma}_{\theta,n}^2 \\ \widetilde{\theta}_n \rho_0 \widetilde{\sigma}_{\theta,n}^2 & \widetilde{\sigma}_{\rho,n}^2 \end{pmatrix} \quad (\text{II.19})$$

à l’aide des estimateurs de  $\sigma_{\theta}^2$  et  $\sigma_{\rho}^2$  en (II.4) et (II.10),

$$\widetilde{\sigma}_{\theta,n}^2 = \frac{(1 - \widetilde{\theta}_n^2)(1 - \widetilde{\theta}_n \rho_0)(1 - \rho_0^2)}{(1 + \widetilde{\theta}_n \rho_0)^3} \quad (\text{II.20})$$

et

$$\tilde{\sigma}_{\rho,n}^2 = \frac{(1 - \tilde{\theta}_n \rho_0)}{(1 + \tilde{\theta}_n \rho_0)^3} \left( (\tilde{\theta}_n + \rho_0)^2 (1 + \tilde{\theta}_n \rho_0)^2 + (\tilde{\theta}_n \rho_0)^2 (1 - \tilde{\theta}_n^2) (1 - \rho_0^2) \right). \quad (\text{II.21})$$

La consistance forte de  $\tilde{\sigma}_{\theta,n}^2$  et  $\tilde{\sigma}_{\rho,n}^2$  sous  $\mathcal{H}_0$ , et par extension celle de  $\tilde{\Gamma}_n$ , est assurée par la convergence (II.18). On considère enfin les quantités

$$\tilde{\omega}_n = \left( -\rho_0 (\tilde{\theta}_n + \tilde{\theta}_n) \quad 1 - \tilde{\theta}_n \rho_0 \right)' \quad \text{et} \quad \tilde{\tau}_n^2 = \frac{4}{(1 + \tilde{\theta}_n \rho_0)^2} |\tilde{\omega}_n' \tilde{\Gamma}_n \tilde{\omega}_n|. \quad (\text{II.22})$$

Notons que la valeur absolue dans la définition de  $\tilde{\tau}_n^2$  est inutile sous  $\mathcal{H}_0$ , puisqu'alors  $\Gamma$  est semi-définie positive. Cependant, il existe des triplets  $(\theta, \rho, \rho_0)$  tels que la valeur limite de  $\tilde{\omega}_n' \tilde{\Gamma}_n \tilde{\omega}_n$  est négative sous  $\mathcal{H}_1$ , et donc la précaution devient nécessaire à la validité du théorème suivant. De même, si  $\tilde{\sigma}_{\theta,n}^2$  et  $\tilde{\sigma}_{\rho,n}^2$  sont des estimateurs consistants de  $\sigma_\theta^2$  et  $\sigma_\rho^2$  sous  $\mathcal{H}_0$ , il se crée en revanche une singularité sous  $\mathcal{H}_1$ , lorsque la limite de  $1 + \tilde{\theta}_n \rho_0$  s'annule. On note alors

$$\rho_1^* = \frac{(\theta + \rho) - \sqrt{(\theta + \rho)^2 + 4}}{2} \quad \text{et} \quad \rho_2^* = \frac{(\theta + \rho) + \sqrt{(\theta + \rho)^2 + 4}}{2} \quad (\text{II.23})$$

qui seront en quelque sorte les valeurs exclues de l'étude.

**Théorème II.9.** *Supposons que  $\mathbb{E}[V_1^4] = \tau^4 < +\infty$ ,  $\theta \neq 0$ ,  $\theta \neq -\rho$ ,  $\theta \neq \rho_0$ ,  $\rho_0 \neq \rho_1^*$  et  $\rho_0 \neq \rho_2^*$ . Alors, si l'on se place sous  $\mathcal{H}_0$  : " $\rho = \rho_0$ ", on a la distribution asymptotique*

$$\frac{n}{\tilde{\tau}_n^2} (\widehat{D}_n - \tilde{D}_n)^2 \xrightarrow{\mathcal{L}} \chi_1^2$$

où  $\chi_1^2$  désigne une variable aléatoire distribuée selon une loi du khi-deux à un degré de liberté. De plus, si l'on se place sous  $\mathcal{H}_1$  : " $\rho \neq \rho_0$ ", alors

$$\lim_{n \rightarrow \infty} \frac{n}{\tilde{\tau}_n^2} (\widehat{D}_n - \tilde{D}_n)^2 = +\infty \quad \text{p.s.}$$

Pour un niveau de significativité  $0 < \alpha < 1$ , les zones d'acceptation et de rejet sont également bâties comme nous l'avons décrit dans la section précédente, à partir du quantile d'ordre  $1 - \alpha$  de la distribution du khi-deux considérée. Malgré l'accumulation d'hypothèses dans l'énoncé du Théorème II.9, celles-ci ne sont pas toutes restrictives. En effet, outre la condition  $\theta \neq 0$ , faisant partie intégrante de l'intérêt de l'étude, et les conditions  $\rho_0 \neq \rho_1^*$  et  $\rho_0 \neq \rho_2^*$ , cas très particuliers, souvenons-nous que  $\theta \neq -\rho$  est censée avoir été éliminée suite à l'application du Théorème II.8. Il reste ainsi seulement la condition  $\theta \neq \rho_0$ . Mais là encore, nous pouvons voir que  $\theta$  et  $\rho$  jouent un rôle parfaitement symétrique dans le modèle AR(1)–AR(1) donné par (II.1) et que ses propriétés asymptotiques restent rigoureusement identiques si l'on choisit d'inverser  $\theta$  et  $\rho$ . Ainsi et comme nous le verrons

dans la preuve, nous sommes amenés à accepter éventuellement à tort  $\mathcal{H}_0$  lorsque  $\theta = \rho_0$ , mais le résultat reste satisfaisant, car c'est alors  $\theta$  qui peut être interprété comme un paramètre d'autocorrélation résiduelle.

### II.3.3 Le cas standard

Il nous reste à adapter le cas que nous avons qualifié de général au cas standard pour lequel nous souhaitons évaluer l'hypothèse d'absence de corrélation résiduelle, à savoir

$$\mathcal{H}_0 : \text{“}\rho = 0\text{”} \quad \text{vs} \quad \mathcal{H}_1 : \text{“}\rho \neq 0\text{”}$$

dans le modèle (II.1). Il suit alors immédiatement du Théorème II.9 avec  $\rho_0 = 0$  le résultat suivant.

**Corollaire II.2.** *Supposons que  $\mathbb{E}[V_1^4] = \tau^4 < +\infty$ ,  $\theta \neq 0$  et  $\theta \neq -\rho$ . Alors, si l'on se place sous  $\mathcal{H}_0 : \text{“}\rho = 0\text{”}$ , on a la distribution asymptotique*

$$\frac{n}{4\widehat{\theta}_n^2} (\widehat{D}_n - 2)^2 \xrightarrow{\mathcal{L}} \chi_1^2$$

où  $\chi_1^2$  désigne une variable aléatoire distribuée selon une loi du khi-deux à un degré de liberté. De plus, si l'on se place sous  $\mathcal{H}_1 : \text{“}\rho \neq 0\text{”}$ , alors

$$\lim_{n \rightarrow \infty} \frac{n}{4\widehat{\theta}_n^2} (\widehat{D}_n - 2)^2 = +\infty \quad \text{p.s.}$$

*Démonstration.* Les Théorèmes II.8–II.9 et le Corollaire II.2 sont prouvés dans la Section II.6.2. ★

## II.4 Simulations et comparaisons

Nous allons étudier empiriquement l'efficacité de notre procédure. Pour un grand nombre de réalisations, nous sommes en mesure de proposer un estimateur de la puissance du test, définie par

$$\mathbb{P}(\text{rejeter } \mathcal{H}_0 \mid \mathcal{H}_1 \text{ est vraie}).$$

Il s'agit simplement de la fréquence empirique de rejet. Suivant un raisonnement similaire à celui de Park [103], afin de minimiser l'impact des valeurs initiales du processus et d'ainsi se placer en régime stationnaire, les 50 premières valeurs simulées sont supprimées. Nous choisissons, pour évaluer la puissance empirique d'un paramétrage, de calculer la fréquence de rejet de  $\mathcal{H}_0$  sur  $N = 1000$  simulations de taille  $n = 300$ , munies d'une perturbation  $\mathcal{N}(0, 1)$ .

### II.4.1 Puissance empirique du test général

Pour rejeter ou non  $\mathcal{H}_0 : \rho = \rho_0$  sur une trajectoire, nous mettons en pratique la procédure décrite dans la Section II.3. Il s'agit, de façon algorithmique, de suivre le protocole suivant.

- Tester  $\mathcal{H}_0 : \theta = -\rho$  contre  $\mathcal{H}_1 : \theta \neq -\rho$  (Théorème II.8).
- Si  $\mathcal{H}_0$  est rejetée :
  - Tester  $\mathcal{H}'_0 : \rho = \rho_0$  contre  $\mathcal{H}'_1 : \rho \neq \rho_0$  (Théorème II.9 ou Corollaire II.2).
- Si  $\mathcal{H}_0$  n'est pas rejetée :
  - Tester  $\mathcal{H}'_0 : \rho = \rho_0$  contre  $\mathcal{H}'_1 : \rho \neq \rho_0$  (Convergence II.16).

Lorsque  $\rho$  est égal à 0, une limitation d'ordre empirique apparaît dans cette procédure. Il faut en effet supposer que  $\rho \neq 0$  pour tester " $\theta = -\rho$ ", alors qu'il faut supposer  $\theta \neq -\rho$  pour tester " $\rho = 0$ ". Si le cadre théorique est bien délimité, la mise en pratique est en revanche nécessairement incomplète de ce point de vue. Il s'agira donc de rester vigilant quant aux conclusions tirées de la procédure, et d'envisager certaines solutions alternatives lorsque les résultats semblent équivoques, comme par exemple tester la significativité du premier coefficient de la modélisation AR(2) du processus. On peut toutefois observer que " $\theta = -\rho$ " et " $\rho = 0$ " ne peuvent être simultanément vraies dans notre cadre de travail. On obtient alors, pour différentes valeurs de  $\theta$  et de  $\rho$ , des profils tels que ceux illustrés sur la Figure II.5. Les exemples présentés correspondent aux configurations  $\rho = -0.4$  et  $\theta = \{-0.2, 0.4, 0.8\}$  sur le graphique de gauche, puis  $\rho = 0.6$  et  $\theta = \{-0.9, -0.6, -0.2\}$  sur le graphique de droite.

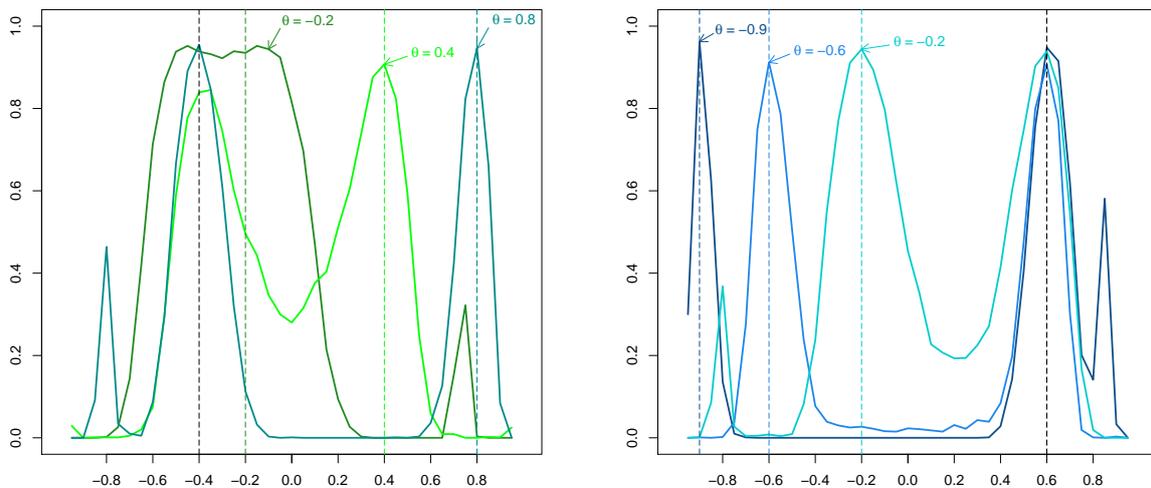


FIGURE II.5 – Fréquence de non rejet de  $\mathcal{H}_0$  en ordonnée et  $\rho_0$  variant de  $-0.95$  à  $0.95$  en abscisse, pour  $\rho = -0.4$  (gauche) et  $\rho = 0.6$  (droite).

Il ressort de ces graphes que la valeur de l'autocorrélation résiduelle  $\rho$  est d'une manière générale toujours retrouvée, avec une probabilité sous  $\mathcal{H}_0$  proche du niveau de significativité choisi, à savoir  $1 - \alpha = 0.95$ . Nous apercevons également un phénomène que nous avons décrit dans la conclusion de la Section II.3.2, relatif à la symétrie jouée par les variables  $\theta$  et  $\rho$  dans le modèle AR(1)–AR(1), que l'on peut résumer par le fait que la procédure de test reconnaît aussi comme corrélation résiduelle la valeur de  $\theta$ . Pour nous en persuader, écrivons la formulation AR(2) du modèle AR(1)–AR(1) donné par (II.1). Pour tout  $2 \leq t \leq n$ , on a

$$Y_t = (\theta + \rho)Y_{t-1} - \theta\rho Y_{t-2} + V_t$$

ce qui met immédiatement en lumière la symétrie jouée par  $\theta$  et  $\rho$ . En outre, quelques perturbations se produisent à l'approche de la frontière  $|\rho_0| = 1$ , et nous observons également que, d'une manière générale, les résultats sont moins convaincants lorsque  $\theta$  est proche de 0. Il est par ailleurs assez clair qu'une étude visuelle nous permet de cerner les valeurs de  $\theta$  et de  $\rho$ , à une interversion près. Sur des séries chronologiques de petite taille, de l'ordre de  $n = 30$ , on observe également les pics correspondant aux valeurs de  $\theta$  et  $\rho$ , mais naturellement de manière assez floue. Nous présenterons quelques exemples en petite dimension dans la section suivante.

## II.4.2 Puissance empirique du test standard

Nous allons comparer l'efficacité de notre procédure de test pour  $\rho_0 = 0$  avec les procédures que l'on rencontre couramment dans la pratique lorsque l'on souhaite évaluer l'hypothèse d'absence de corrélation dans les résidus issus d'un modèle de régression linéaire.

### II.4.2.1 Les tests de Box-Pierce et de Ljung-Box

Nous considérerons tout d'abord les tests que l'on appelle communément *du portemanteau*, regroupant la procédure de Box-Pierce [22] et celle de Ljung-Box [21]. Ces dernières permettent d'évaluer l'hypothèse d'absence de corrélation résiduelle à tout ordre  $q \in \mathbb{N}^*$ , mais nous nous limiterons bien entendu ici au cadre de l'étude,  $q = 1$ . Les statistiques de test utilisées sont données par

$$\widehat{Q}_n^{\text{BP}} = n\widehat{\rho}_n^2 \quad \text{et} \quad \widehat{Q}_n^{\text{LB}} = \frac{n(n+2)}{n-1}\widehat{\rho}_n^2 \quad (\text{II.24})$$

où  $\widehat{\rho}_n$  est défini en (II.7), BP et LB désignant respectivement *Box-Pierce* et *Ljung-Box*. Les procédures reposent sur la comparaison de  $\widehat{Q}_n^{\text{BP}}$  et  $\widehat{Q}_n^{\text{LB}}$  avec les quantiles d'une loi du khi-deux à un degré de liberté,  $\widehat{Q}_n^{\text{LB}}$  étant en outre muni d'un coefficient censé minimiser le biais en petite dimension. Or, nous avons montré à travers le Théorème II.5 que, dans le cadre du modèle AR(1)–AR(1), en l'absence de corrélation résiduelle, la variance asymptotique de  $n\widehat{\rho}_n^2$  ne vaut pas 1, mais  $\theta^2$ . Il suffit pour cela

de noter que les valeurs limites se simplifient en  $\rho^* = 0$  et  $\sigma_\rho^2 = \theta^2$  sous  $\mathcal{H}_0 : \rho = 0$ . Ainsi,

$$\frac{\widehat{Q}_n^{\text{BP}}}{\theta^2} \xrightarrow{\mathcal{L}} \chi_1^2$$

et, puisque  $|\theta| < 1$ , comparer  $\widehat{Q}_n^{\text{BP}}$  avec un khi-deux à un degré de liberté revient à surestimer  $\mathcal{H}_0$ , et ce d'autant plus que  $\theta$  se rapproche de 0. Il en va bien entendu de même pour  $\widehat{Q}_n^{\text{LB}}$ .

### II.4.2.2 Le H-test de Durbin

Dans [47], Durbin propose principalement deux procédures permettant de tenir compte de la dynamique du modèle de régression, ce que ne font pas les procédures de Box-Pierce et de Ljung-Box. Il s'agit du T-test et du H-test. En fait, nous ne considérerons ici que le H-test car, comme le remarque Park dans [103], le T-test se réduit à projeter  $(\widehat{\varepsilon}_t)$  sur  $(\widehat{\varepsilon}_{t-1})$  et à tester la significativité du coefficient de régression engendré. C'est finalement la stratégie utilisée au sein du test de Box-Pierce. Le H-test quant à lui s'appuie sur la statistique

$$\widehat{H}_n = \left(1 - \frac{\widehat{D}_n}{2}\right)^2 \left[ \frac{n}{1 - n\widehat{\mathbb{V}}(\widehat{\theta}_n)} \right] \quad (\text{II.25})$$

où  $\widehat{D}_n$  est défini en (II.13). Durbin suggère alors de comparer  $\widehat{H}_n$  avec les quantiles d'une loi du khi-deux à un degré de liberté, sous  $\mathcal{H}_0$ . Nous montrerons dans le Théorème III.9 du chapitre suivant qu'en réalité, il existe une équivalence asymptotique presque sûre entre  $\widehat{H}_n$  et la statistique de test que l'on utilise dans le Corollaire II.2, par ailleurs établie sous des hypothèses moins restrictives que celles du H-test. On s'attendra donc à obtenir des résultats de simulation asymptotiquement équivalents quant à la puissance empirique des procédures de test respectives.

### II.4.2.3 Le test de Breusch-Godfrey

La procédure de Breusch-Godfrey [23]–[57] s'applique à un vaste panel de modèles dynamiques, contenant variables exogènes comme endogènes. Il s'agit d'évaluer les coefficients d'une régression de  $(\widehat{\varepsilon}_t)$  certes sur  $(\widehat{\varepsilon}_{t-1})$ , mais également sur  $(\widehat{\varepsilon}_{t-2}, \dots, \widehat{\varepsilon}_{t-q})$  ainsi que sur l'ensemble des régresseurs dont ils sont eux-mêmes issus, à savoir  $(Y_{t-1}, \dots, Y_{t-p})$  sans oublier les variables exogènes. C'est une procédure fort complexe que nous limiterons au domaine d'étude de ce chapitre, qui reste le modèle autorégressif d'ordre  $p = 1$  et l'évaluation de la première autocorrélation résiduelle,  $q = 1$ . Breusch et Godfrey suggèrent alors dans ce cas de comparer la valeur de

$$\widehat{B}_n = nR^2 \quad (\text{II.26})$$

avec les quantiles d’une loi du khi-deux à un degré de liberté, où  $R^2$  est le coefficient de détermination usuel de la régression. Ce test est, à notre connaissance, le plus complet lorsque l’on souhaite évaluer une hypothèse de blancheur sur un ensemble résiduel issu d’une régression linéaire, et nous souhaitons modestement ici évaluer ses performances, dans un cadre certes très particulier, par une procédure simplifiée.

### II.4.2.4 En simulation

Reprenons le même protocole que précédemment (pour rappel, la puissance empirique a été définie comme la fréquence de rejet de  $\mathcal{H}_0$  : “ $\rho = 0$ ” sur  $N = 1000$  simulations, munies d’une perturbation  $\mathcal{N}(0,1)$ , les 50 premières observations étant supprimées). Sur les Figures II.6–II.7–II.8, nous avons représenté la fréquence de non rejet de  $\mathcal{H}_0$  pour les 5 procédures de test considérées,  $\rho$  variant de  $-0.95$  à  $0.95$  et  $\theta = -0.8$ ,  $\theta = 0.3$  et  $\theta = 0.6$  respectivement, pour  $n = 300$ . Nous y avons adjoint les résultats obtenus à partir des mêmes configurations sur des petits échantillons, pour  $n = 30$ . La légende précise simplement les abréviations explicites des procédures.

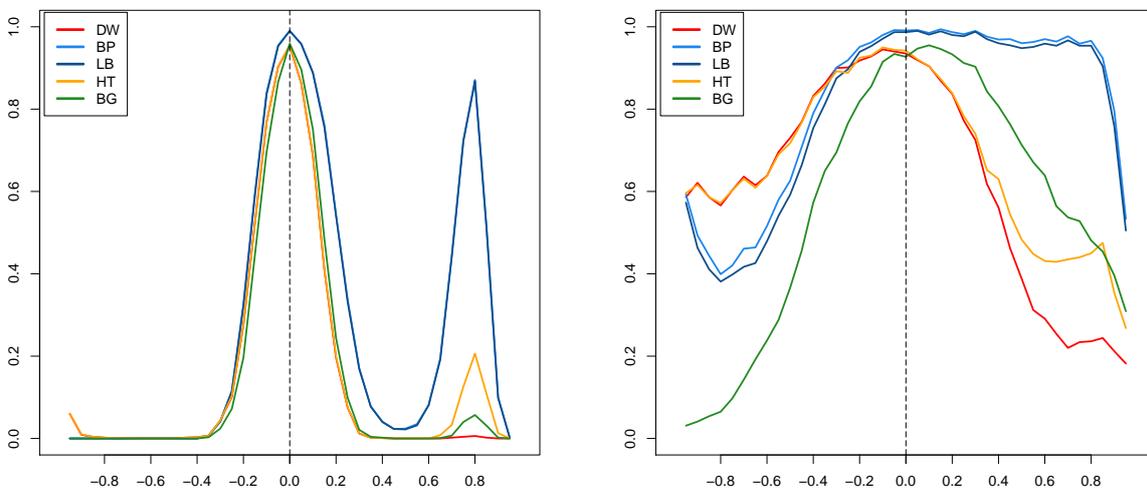


FIGURE II.6 – Fréquence de non rejet de  $\mathcal{H}_0$  en ordonnée pour  $\theta = -0.8$  et  $\rho$  variant de  $-0.95$  à  $0.95$  en abscisse, avec  $n = 300$  (gauche) et  $n = 30$  (droite).

Tout d’abord, il est très clair que les tests de Box-Pierce et de Ljung-Box surestiment  $\mathcal{H}_0$ , comme nous l’avons expliqué précédemment. Nous avons ainsi la confirmation empirique que ces procédures ne sont pas adaptées au cadre autorégressif (sauf bien sûr sous  $\mathcal{H}_0$ , mais cette supériorité n’est qu’artificielle puisqu’elle découle justement d’une mauvaise approximation). Nous constatons ensuite que sur de grands échantillons, notre procédure est équivalente au H-test, comme attendu, et au test de Breusch-Godfrey, comme espéré. D’une manière générale, l’absence d’autocorrélation ré-

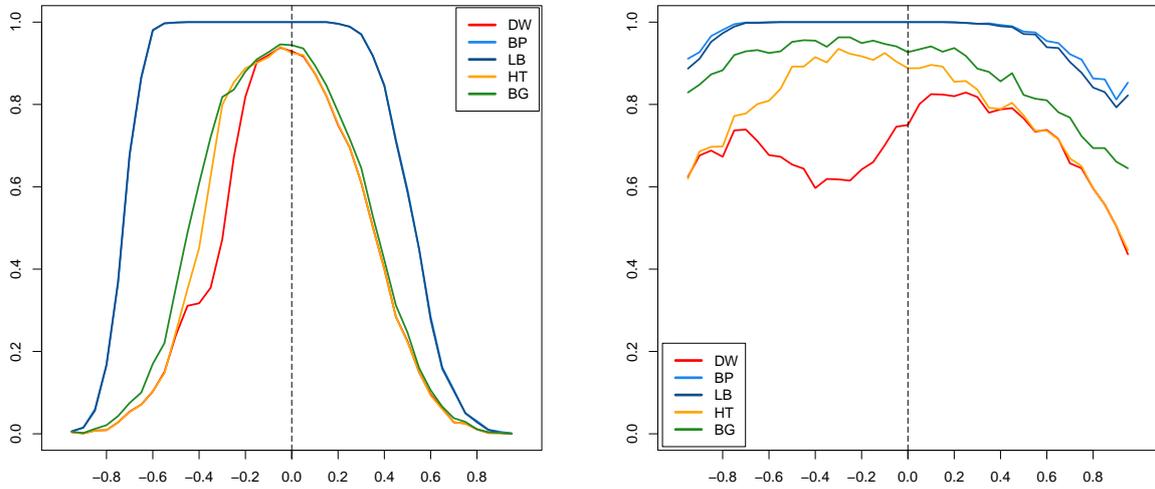


FIGURE II.7 – Fréquence de non rejet de  $\mathcal{H}_0$  en ordonnée pour  $\theta = 0.3$  et  $\rho$  variant de  $-0.95$  à  $0.95$  en abscisse, avec  $n = 300$  (gauche) et  $n = 30$  (droite).

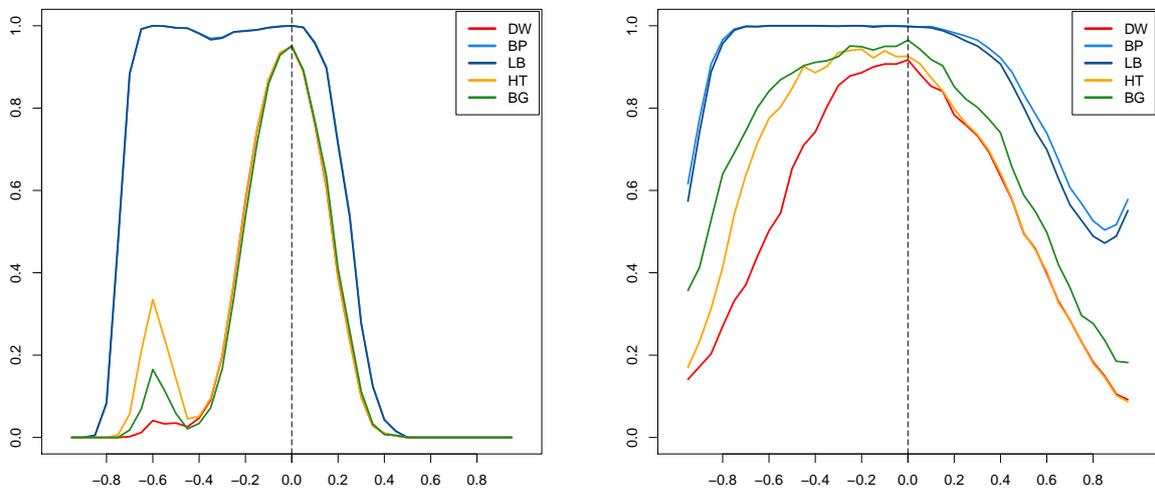


FIGURE II.8 – Fréquence de non rejet de  $\mathcal{H}_0$  en ordonnée pour  $\theta = 0.6$  et  $\rho$  variant de  $-0.95$  à  $0.95$  en abscisse, avec  $n = 300$  (gauche) et  $n = 30$  (droite).

siduelle est détectée avec une probabilité proche du niveau de sécurité choisi, à savoir  $1 - \alpha = 0.95$ , et ce d'autant plus distinctement que  $\theta$  est significatif. Notre procédure est du reste la plus robuste au cas pathologique  $\theta = -\rho$  qui, par l'intermédiaire du test préliminaire, est totalement intégré contrairement à ce que l'on peut visualiser avec les autres procédures. Sur les petits échantillons, il est bien évident que les conclusions sont à tempérer, et l'aspect des graphes parle de lui-même. Nous constatons

rons simplement, et sans en faire une généralité, que notre procédure de test semble légèrement plus puissante. Notons enfin que l'ensemble de ces procédures est asymptotiquement équivalent lorsque  $|\theta|$  est proche de 1.

## II.5 Conclusion et perspectives

Nous pensons avoir établi un ensemble de résultats apportant un regard nouveau sur la statistique de Durbin-Watson, à l'aide d'outils de martingales. C'est en effet à partir de sa distribution asymptotique que nous avons obtenu des tests d'hypothèse dont la puissance empirique concurrence celle des procédures couramment utilisées dans le cadre du modèle autorégressif du premier ordre. Bien que nous puissions en trouver l'esquisse dans [47], quoique dans le cadre restrictif de gaussianité des résidus, cette dernière n'était pas explicitement stipulée et le test de Durbin-Watson reposait jusqu'alors sur des bornes et des tabulations de Monte-Carlo. Nous espérons avoir ainsi approfondi de manière significative le comportement asymptotique de cette statistique, en sus de sa convergence presque sûre, et illustré le fait qu'elle peut être adaptée à d'autres fins statistiques. Rappelons pour conclure que les estimateurs du modèle AR(1)–AR(1) ne sont pas consistants, mais qu'il est possible de les rendre consistants à partir des combinaisons

$$\lim_{n \rightarrow \infty} \widehat{\theta}_n + \widehat{\rho}_n = \theta + \rho \quad \text{p.s.} \quad \text{et} \quad \lim_{n \rightarrow \infty} \frac{\widehat{\rho}_n}{\widehat{\theta}_n} = \theta \rho \quad \text{p.s.}$$

Ainsi, sous l'hypothèse  $\theta \leq \rho$ , les estimateurs  $\bar{\theta}_n$  et  $\bar{\rho}_n$  respectivement définis, pour tout  $n \geq 1$ , par

$$\bar{\theta}_n = \frac{1}{2} \left( (\widehat{\theta}_n + \widehat{\rho}_n) - \sqrt{(\widehat{\theta}_n + \widehat{\rho}_n)^2 - 4 \frac{\widehat{\rho}_n}{\widehat{\theta}_n}} \right) \quad \text{et} \quad \bar{\rho}_n = \frac{1}{2} \left( (\widehat{\theta}_n + \widehat{\rho}_n) + \sqrt{(\widehat{\theta}_n + \widehat{\rho}_n)^2 - 4 \frac{\widehat{\rho}_n}{\widehat{\theta}_n}} \right)$$

sont des estimateurs fortement consistants de  $\theta$  et de  $\rho$ . Quant au rôle symétrique joué par les paramètres, on le retrouve de nouveau ici puisqu'il nous suffit d'invertir les valeurs de  $\bar{\theta}_n$  et de  $\bar{\rho}_n$  pour conserver la consistance forte sous l'hypothèse  $\theta \geq \rho$ . Ce travail nous a inspiré plusieurs améliorations. En premier lieu, l'extension naturelle consiste à considérer le modèle AR( $p$ )–AR(1) pour tout  $p \geq 1$ . C'est de fait l'objectif du Chapitre III qu'il est désormais temps d'aborder. Par ailleurs, il est également possible de renforcer les convergences obtenues, d'entrer dans le domaine *super-exponentiel* et d'obtenir des *principes de déviations modérées*, ce sera l'objectif du Chapitre IV. Nous survolerons très sommairement dans le Chapitre VI les cas d'instabilité sur la bordure  $|\rho| = 1$ . Quant à l'étude du cas explosif, le problème reste ouvert...

## II.6 Annexe : démonstration des résultats

### II.6.1 Sur les estimateurs

Comme nous l'avons indiqué en début de chapitre, certains théorèmes sont des corollaires issus du chapitre suivant, en conséquence nous ne dupliquerons pas inutilement les preuves associées. Il s'agit des Théorèmes II.1–II.7 pour lesquels nous invitons le lecteur à se reporter à la Section III.5.2 et à ainsi accéder à la preuve de résultats plus généraux.

### II.6.2 Sur la procédure statistique

#### II.6.2.1 Preuve du Corollaire II.1

Il s'agit tout d'abord d'établir la relation quasi linéaire existant entre  $\widehat{D}_n$  et  $\widehat{\rho}_n$ . On rappelle que, de manière arbitraire, on a  $\widehat{\varepsilon}_0 = 0$ . Pour tout  $n \geq 1$ , commençons par noter

$$I_n = \sum_{t=1}^n \widehat{\varepsilon}_{t-1} \widehat{\varepsilon}_t, \quad J_n = \sum_{t=0}^n \widehat{\varepsilon}_t^2 \quad \text{et} \quad \widehat{f}_n = \frac{\widehat{\varepsilon}_n^2}{J_n}. \quad (\text{II.27})$$

Si nous considérons le numérateur de  $\widehat{D}_n$  défini en (II.13), il vient

$$(J_{n-1} + \widehat{\varepsilon}_n^2) \widehat{D}_n = J_n - 2I_n + J_{n-1} = 2(J_{n-1} - I_n) + \widehat{\varepsilon}_n^2.$$

En conséquence,

$$\widehat{D}_n = 2(1 - \widehat{f}_n)(1 - \widehat{\rho}_n) + \widehat{f}_n \quad (\text{II.28})$$

où  $\widehat{\rho}_n$  est donné par (II.7). Nous verrons dans le chapitre suivant que, par application des Lemmes III.3–III.4, nous avons les convergences presque sûres

$$\lim_{n \rightarrow \infty} \widehat{f}_n = 0 \quad \text{p.s.} \quad \text{et} \quad \lim_{n \rightarrow \infty} \sqrt{n} \widehat{f}_n = 0 \quad \text{p.s.}$$

dès que, respectivement,  $\mathbb{E}[V_1^2] = \sigma^2 < +\infty$  et  $\mathbb{E}[V_1^4] = \tau^4 < +\infty$ . Cela nous conduit d'une part à la convergence presque sûre de  $\widehat{D}_n$  vers  $D^* = 2(1 - \rho^*)$ , par l'intermédiaire du Théorème II.4. D'autre part, cela implique que la décomposition

$$\sqrt{n}(\widehat{D}_n - D^*) = -2\sqrt{n}(1 - \widehat{f}_n)(\widehat{\rho}_n - \rho^*) - \sqrt{n}(1 - 2\rho^*)\widehat{f}_n \quad (\text{II.29})$$

nous fournit, combinée au Théorème II.5 et au lemme de Slutsky, la normalité asymptotique de  $\sqrt{n}(\widehat{D}_n - D^*)$  munie de la variance  $\sigma_D^2 = 4\sigma_\rho^2$ , ce qui achève la preuve. ★

### II.6.2.2 Preuve du Théorème II.8

Plaçons-nous sous  $\mathcal{H}_0 : \theta = -\rho$ . Alors, il est facile d'exprimer le modèle (II.1) sous une forme AR(2) à trou. Pour tout  $2 \leq t \leq n$ , le processus est engendré par

$$Y_t = \theta^2 Y_{t-2} + V_t.$$

Du reste, il est bien connu (voir Proposition I.5) que

$$\lim_{n \rightarrow \infty} \widehat{\vartheta}_n = \theta^2 \quad \text{p.s.} \quad (\text{II.30})$$

où l'estimateur  $\widehat{\vartheta}_n$  est donné par (II.15). En outre, dès que  $\mathbb{E}[V_1^4] = \tau^4 < +\infty$ , il suit du Corollaire II.1, et par extension de (II.14), que l'on a la convergence

$$\frac{n(1-\theta^2)}{4\theta^4(1+\theta^2)} (\widehat{D}_n - 2)^2 \xrightarrow{\mathcal{L}} \chi_1^2$$

puisque  $D^* = 2$  sous  $\mathcal{H}_0$ , où  $\chi_1^2$  désigne une variable aléatoire distribuée selon une loi du khi-deux à un degré de liberté. La première partie de la preuve est alors achevée en utilisant la convergence (II.30) et le lemme de Slutsky. Sous  $\mathcal{H}_1 : \theta \neq -\rho$ , on voit immédiatement que  $D^* = 2$  si et seulement si  $\theta = 0$  ou  $\rho = 0$ . Ces deux cas étant exclus par hypothèse, on a nécessairement

$$\lim_{n \rightarrow \infty} n (\widehat{D}_n - 2)^2 = +\infty \quad \text{p.s.}$$

Par ailleurs, nous voyons à l'aide du Lemme III.4 que sous  $\mathcal{H}_1$  avec  $p = 1$ , il est possible d'établir la convergence de  $\widehat{\vartheta}_n$  vers  $\theta^*(\theta + \rho) - \theta\rho$ . En conséquence, après quelques simplifications,

$$\lim_{n \rightarrow \infty} \frac{1 - \widehat{\vartheta}_n}{4\widehat{\vartheta}_n^2 (1 + \widehat{\vartheta}_n)} = -\frac{(1 - \theta^2)(1 + \theta\rho)^2(1 - \rho^2)}{4((\theta + \rho)^2 - \theta\rho(1 + \theta\rho))^2(\theta^2\rho^2 - (\theta + \rho)^2 - 1)} \quad \text{p.s.}$$

La preuve est achevée en notant que le numérateur ne s'annule pas lorsque  $|\theta| < 1$  et  $|\rho| < 1$ . ★

### II.6.2.3 Preuve du Théorème II.9

À partir de leur définition respective en (II.7) et (II.17), on a, pour  $n \geq 1$ ,

$$\begin{aligned} \widehat{\rho}_n - \widetilde{\rho}_n &= \widehat{\rho}_n - \frac{\widetilde{\theta}_n \rho_0 (\widetilde{\theta}_n + \rho_0)}{1 + \widetilde{\theta}_n \rho_0} = \frac{\widehat{\rho}_n + \widetilde{\theta}_n \rho_0 (\widehat{\rho}_n - \widetilde{\theta}_n - \rho_0)}{1 + \widetilde{\theta}_n \rho_0} \\ &= \frac{\widehat{\rho}_n - \widetilde{\theta}_n \widetilde{\theta}_n \rho_0}{1 + \widetilde{\theta}_n \rho_0} = \frac{\widehat{\rho}_n - \theta \widetilde{\theta}_n \rho_0}{1 + \widetilde{\theta}_n \rho_0} - \frac{\widetilde{\theta}_n \rho_0 (\widetilde{\theta}_n - \theta)}{1 + \widetilde{\theta}_n \rho_0}. \end{aligned} \quad (\text{II.31})$$

Notons  $\theta_0^*$  et  $\rho_0^*$  les équivalents sous  $\mathcal{H}_0$  : “ $\rho = \rho_0$ ” de  $\theta^*$  et  $\rho^*$ ,

$$\theta_0^* = \frac{\theta + \rho_0}{1 + \theta\rho_0} \quad \text{et} \quad \rho_0^* = \frac{\theta\rho_0(\theta + \rho_0)}{1 + \theta\rho_0}.$$

On remarque que  $\theta_0^* + \rho_0^* = \theta + \rho_0$  et que  $\rho_0^* = \theta\rho_0\theta_0^*$ . Ainsi, en reprenant (II.31),

$$\begin{aligned} \widehat{\rho}_n - \widetilde{\rho}_n &= \frac{\widehat{\rho}_n - \theta\widehat{\theta}_n\rho_0 - \theta_0^*\rho_0(\widetilde{\theta}_n - \theta)}{1 + \widetilde{\theta}_n\rho_0} - \frac{\rho_0(\widetilde{\theta}_n - \theta)(\widehat{\theta}_n - \theta_0^*)}{1 + \widetilde{\theta}_n\rho_0} \\ &= \frac{(\widehat{\rho}_n - \rho_0^*) - \theta\rho_0(\widehat{\theta}_n - \theta_0^*) - \theta_0^*\rho_0(\widetilde{\theta}_n - \theta)}{1 + \widetilde{\theta}_n\rho_0} - \frac{\rho_0(\widetilde{\theta}_n - \theta)(\widehat{\theta}_n - \theta_0^*)}{1 + \widetilde{\theta}_n\rho_0} \\ &= \frac{(1 - \theta_0^*\rho_0)(\widehat{\rho}_n - \rho_0^*) - \rho_0(\theta + \theta_0^*)(\widehat{\theta}_n - \theta_0^*)}{1 + \widetilde{\theta}_n\rho_0} - \frac{\rho_0(\widetilde{\theta}_n - \theta)(\widehat{\theta}_n - \theta_0^*)}{1 + \widetilde{\theta}_n\rho_0} = \frac{\omega' \widehat{\pi}_n - \rho_0 r_n}{1 + \widetilde{\theta}_n\rho_0} \end{aligned}$$

où nous avons posé

$$\omega = \begin{pmatrix} -\rho_0(\theta + \theta_0^*) \\ 1 - \theta_0^*\rho_0 \end{pmatrix}, \quad \widehat{\pi}_n = \begin{pmatrix} \widehat{\theta}_n - \theta_0^* \\ \widehat{\rho}_n - \rho_0^* \end{pmatrix} \quad \text{et} \quad r_n = (\widetilde{\theta}_n - \theta)(\widehat{\theta}_n - \theta_0^*).$$

Quelques calculs supplémentaires nous conduisent à

$$\sqrt{n}(\widehat{D}_n - \widetilde{D}_n) = \frac{-2\sqrt{n}\omega' \widehat{\pi}_n}{1 + \widetilde{\theta}_n\rho_0} + \frac{2\rho_0\sqrt{n}r_n}{1 + \widetilde{\theta}_n\rho_0} - \sqrt{n}(1 - 2\widehat{\rho}_n)\widehat{f}_n \quad (\text{II.32})$$

où  $\widehat{f}_n$  est donné par (II.27), et l'on a vu que, dès que  $\mathbb{E}[V_1^4] = \tau^4 < +\infty$ ,  $\sqrt{n}\widehat{f}_n$  converge presque sûrement vers 0. Par ailleurs, les Théorèmes II.3 et II.6 nous montrent que

$$\sqrt{n}r_n = O\left(\frac{\log \log n}{\sqrt{n}}\right) \quad \text{p.s.} \quad (\text{II.33})$$

car  $|\widetilde{\theta}_n - \theta| = |\widehat{\theta}_n - \theta_0^* + \widehat{\rho}_n - \rho_0^*| \leq |\widehat{\theta}_n - \theta_0^*| + |\widehat{\rho}_n - \rho_0^*|$ , ce qui implique que  $\widetilde{\theta}_n$  converge à la même vitesse vers  $\theta$  que  $\widehat{\theta}_n$  et  $\widehat{\rho}_n$  vers leur limite respective. Il s'ensuit que  $\sqrt{n}r_n$  converge presque sûrement vers 0. Enfin, le Théorème II.5 entraîne que

$$\sqrt{n}\omega' \widehat{\pi}_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \omega' \Gamma_0 \omega) \quad (\text{II.34})$$

où  $\Gamma_0$  est la matrice de covariance  $\Gamma$  donnée par (II.9) évaluée avec  $\rho = \rho_0$ . Sous  $\mathcal{H}_0$ , on a vu que  $\widetilde{\theta}_n$  converge presque sûrement vers  $\theta$ . Par (II.34) et le lemme de Slutsky,

$$\frac{-2\sqrt{n}\omega' \widehat{\pi}_n}{1 + \widetilde{\theta}_n\rho_0} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \tau^2)$$

avec

$$\tau^2 = \frac{4}{(1 + \theta\rho_0)^2} \omega' \Gamma_0 \omega.$$

Combiné avec (II.33), ce résultat conclut la première partie de la preuve en remarquant que la quantité  $\tilde{\omega}_n$ , définie en (II.22), est un estimateur consistant de  $\omega$  sous  $\mathcal{H}_0$ . Sous  $\mathcal{H}_1$  : “ $\rho \neq \rho_0$ ”, nous avons les convergences presque sûres

$$\lim_{n \rightarrow \infty} \widehat{D}_n = 2 \left( 1 - \frac{\theta\rho(\theta + \rho)}{1 + \theta\rho} \right) \quad \text{p.s.} \quad \text{et} \quad \lim_{n \rightarrow \infty} \widetilde{D}_n = 2 \left( 1 - \frac{\rho_0(\theta + \rho - \rho_0)(\theta + \rho)}{1 + \rho_0(\theta + \rho - \rho_0)} \right) \quad \text{p.s.}$$

à partir du Corollaire II.1 et de la définition de  $\widetilde{D}_n$  en (II.17). De fait, après quelques simplifications,

$$\lim_{n \rightarrow \infty} (\widehat{D}_n - \widetilde{D}_n) = \frac{-2(\theta + \rho)(\theta - \rho_0)(\rho - \rho_0)}{(1 + \theta\rho)(1 + \rho_0(\theta + \rho - \rho_0))} \quad \text{p.s.} \quad (\text{II.35})$$

La valeur limite ne s’annule ainsi que lorsque  $\theta = -\rho$  ou  $\theta = \rho_0$ , situations exclues de l’énoncé par hypothèse, ou bien lorsque  $\rho = \rho_0$ , invalide sous  $\mathcal{H}_1$ . De plus, on montre que la valeur limite de  $\tilde{\tau}_n^2$  donnée par (II.22) prend une forme assez complexe sous  $\mathcal{H}_1$  – qu’il serait inutile de préciser ici, il suffit pour cela de remplacer  $\tilde{\theta}_n$  par  $\theta + \rho - \rho_0$  et  $\widehat{\theta}_n$  par  $\theta^*$  dans son expression – susceptible d’être infinie lorsque  $1 + (\theta + \rho - \rho_0)\rho_0 = 0$ . C’est pourquoi l’on est contraint de retirer de l’étude les valeurs  $\rho_1^*$  et  $\rho_2^*$  données par (II.23) et telles que  $|\rho_1^*| < 1$  ou  $|\rho_2^*| < 1$ , dont l’existence n’est d’ailleurs pas garantie pour tous les couples  $(\theta, \rho)$ . Ces valeurs mises à part,  $\tilde{\tau}_n^2$  converge presque sûrement vers une limite finie sous  $\mathcal{H}_1$ , ce qui achève la preuve puisque, par l’intermédiaire de (II.35),

$$\lim_{n \rightarrow \infty} \frac{n}{\tilde{\tau}_n^2} (\widehat{D}_n - \widetilde{D}_n)^2 = +\infty \quad \text{p.s.}$$

★

#### II.6.2.4 Preuve du Corollaire II.2

À l’aide des outils que nous venons d’établir, la preuve de ce Corollaire est assez immédiate. Sous  $\mathcal{H}_0$  : “ $\rho = 0$ ”, nous reprenons la preuve précédente pour  $\rho_0 = 0$ . Il vient alors  $\tilde{\rho}_n = 0$ ,  $\widetilde{D}_n = 2$  et

$$\sqrt{n} (\widehat{D}_n - 2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\theta^2).$$

Ce résultat peut aussi découler du Corollaire II.1 sous l’hypothèse  $\rho = 0$ . Ainsi, par la consistance forte de  $\widehat{\theta}_n$  sous  $\mathcal{H}_0$  et le lemme de Slutsky, la première partie de la preuve est terminée. Sous  $\mathcal{H}_1$  : “ $\rho \neq 0$ ”, on sait par le Théorème II.1 que  $\widehat{\theta}_n$  converge presque sûrement vers  $\theta^*$  qui ne s’annule pas lorsque  $\theta \neq -\rho$ , étant entendu que  $\theta \neq 0$ . On sait également que, si l’on a de plus  $\rho \neq 0$ , alors  $D^* \neq 2$ . Les hypothèses retenues dans l’énoncé sont suffisantes pour achever la preuve. ★

---

## *Une généralisation vectorielle*

---

Les résultats présentés dans ce chapitre ont été publiés, sous la forme d'une approche très similaire, dans [112]. Nous y abordons la problématique de l'autocorrélation résiduelle dans le processus autorégressif d'ordre  $p$  quelconque.

### **III.1 Historique et motivation**

La statistique de Durbin-Watson est une nouvelle fois l'axe directeur de cette étude, c'est pourquoi nous nous permettrons de limiter l'historique et la bibliographie à une simple référence à la Section II.1. Notre objectif est désormais l'extension des résultats du chapitre précédent (ou de [10]) au processus engendré par un modèle autorégressif d'ordre  $p \in \mathbb{N}^*$  quelconque. Nous serons ainsi amenés à constater que les propriétés des estimateurs sont similaires, mais que la généralisation est loin d'être triviale. Ce sera notre premier objectif majeur et cela nous permettra en outre de proposer la démonstration rigoureuse de nos résultats, dont la restriction à  $p = 1$  engendrera les théorèmes limites du chapitre précédent, stipulés mais non encore prouvés. C'est donc tout naturellement que nous suivrons le chemin balisé du Chapitre II, mais seulement jusqu'à un certain point. Nous ne développerons en effet pas de procédure statistique permettant d'évaluer  $\mathcal{H}_0 : \rho = \rho_0 \neq 0$  car, et comme nous le verrons très explicitement, le comportement précis d'une statistique de test ayant les mêmes contours que celle proposée dans le Théorème II.9 fait intervenir, sous  $\mathcal{H}_1 : \rho \neq \rho_0$ , un nombre décourageant de singularités. Le test d'absence de corrélation résiduelle ( $\rho_0 = 0$ ) concentrera en conséquence toute notre attention, et nous en profiterons pour faire digresser l'étude vers une comparaison asymptotique avec le H-test de Durbin, présenté dans la Section II.4.2.2. Il s'agit là de notre second objectif majeur que l'on pourrait résumer par le fait de montrer que notre procédure statistique simplifie, concurrence et améliore sous certains aspects le H-test. Une étude en simula-

tion sera menée en guise de conclusion, dans laquelle nous comparerons notre procédure avec celles couramment usitées afin de tester la blancheur résiduelle dans un modèle autorégressif, à l'image de ce que nous avons proposé dans la Section II.3.

Dans la suite de l'étude, nous utiliserons les notations ci-dessous pour respectivement désigner la matrice identité et la matrice d'échange d'ordre  $p$ , et le premier vecteur unitaire de  $\mathbb{R}^p$ .

$$I_p = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}, \quad J_p = \begin{pmatrix} 0 & \dots & 0 & 1 \\ 0 & \dots & 1 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 1 & \dots & 0 & 0 \end{pmatrix}, \quad e = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

## III.2 Le modèle et ses estimateurs

Nous considérons un modèle autorégressif d'ordre  $p \in \mathbb{N}^*$ , lui-même issu d'une perturbation autorégressive du premier ordre. Soit  $(Y_t)$  le processus engendré, indexé par  $\mathbb{Z}$ , dont on suppose disposer d'une trajectoire observée sur  $\{0, \dots, n\}$ . Pour tout  $p \leq t \leq n$ , le modèle est alors donné par

$$\begin{cases} Y_t = \theta_1 Y_{t-1} + \dots + \theta_p Y_{t-p} + \varepsilon_t \\ \varepsilon_t = \rho \varepsilon_{t-1} + V_t \end{cases} \quad (\text{III.1})$$

où les paramètres inconnus vérifient les conditions  $\|\theta\|_1 < 1$  et  $|\rho| < 1$  dites « de stabilité », et où  $(V_t)$  est un bruit blanc de variance  $\sigma^2 > 0$ . La trajectoire sera en conséquence entièrement décrite par la filtration  $\mathcal{F}_n = \sigma(Y_0, \dots, Y_{p-1}, \varepsilon_0, \dots, \varepsilon_{p-1}, V_1, \dots, V_n)$ , pour  $n \geq 1$ . Nous prendrons l'habitude de qualifier de AR( $p$ )-AR(1) une telle modélisation, et là encore il est important de ne pas la confondre avec un modèle ARMA( $p, 1$ ) qui ne possède pas la même dynamique. Nous montrerons que les conditions de stabilité retenues suffisent à établir que le modèle considéré peut être interprété comme un modèle AR( $p + 1$ ) dont le polynôme autorégressif est causal. Deux trajectoires de taille  $n = 500$  sont simulées sur la Figure III.1, respectivement pour  $p = 2$ ,  $(\theta, \rho) = ((0.5, 0.2)', -0.3)$  et  $(V_t) \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$ , puis pour  $p = 3$ ,  $(\theta, \rho) = ((-0.2, 0.1, -0.3)', 0.7)$  et  $(V_t) \stackrel{\text{iid}}{\sim} \mathcal{U}([-2, 2])$ . Comme nous l'avons fait dans le Chapitre II, nous souhaitons estimer les paramètres  $\theta$  et  $\rho$  de manière non couplée, sans considérer la formulation AR( $p + 1$ ) du modèle (III.1).

### III.2.1 Le paramètre de l'autorégression

Pour  $n \geq p$  et  $p - 1 \leq t \leq n$ , on note

$$\Phi_t^p = (Y_t \quad Y_{t-1} \quad \dots \quad Y_{t-p+1})' \quad (\text{III.2})$$

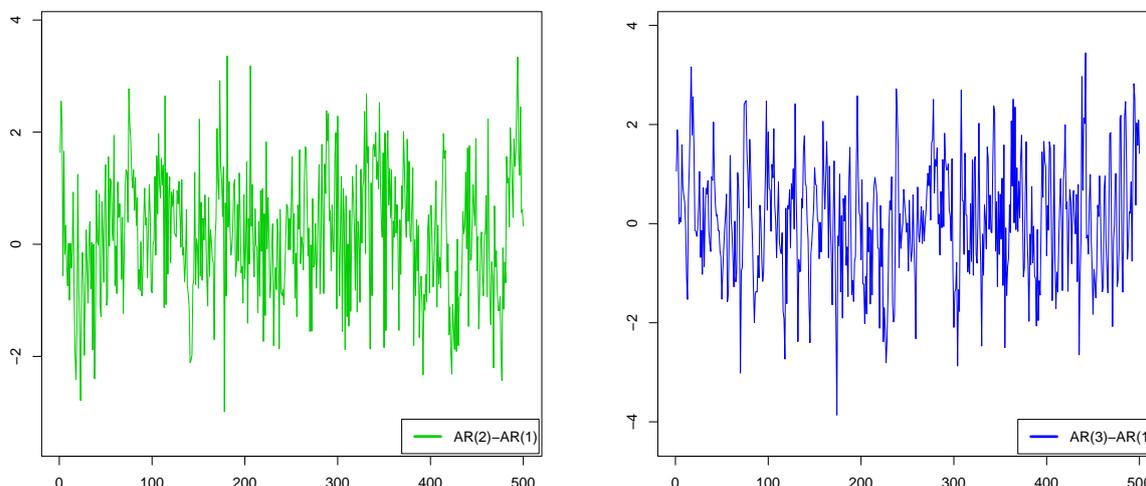


FIGURE III.1 – Exemples de trajectoires  $AR(p)$ – $AR(1)$  stables, pour  $p = 2$  (gauche) et  $p = 3$  (droite).

et l'on considère la matrice définie positive

$$S_n = \sum_{t=p-1}^n \Phi_t^p \Phi_t^{p'} + S \quad (\text{III.3})$$

où  $S$  est aussi une matrice définie positive ajoutée à  $S_n$  pour s'affranchir d'une hypothèse d'inversibilité. On pourra par exemple choisir  $S = I_p$ . Nous souhaitons alors minimiser la quantité

$$\Delta_n(\theta) = \sum_{t=0}^n \varepsilon_t^2 = \varepsilon_0^2 + \dots + \varepsilon_{p-1}^2 + \sum_{t=p}^n (Y_t - \theta' \Phi_{t-1}^p)^2$$

et l'on obtient l'estimateur des moindres carrés donné, pour tout  $n \geq p$ , par

$$\widehat{\theta}_n = (S_{n-1})^{-1} \sum_{t=p}^n \Phi_{t-1}^p Y_t. \quad (\text{III.4})$$

### III.2.1.1 Convergence presque sûre

Introduisons quelques notations supplémentaires, en particulier

$$\alpha = \frac{1}{(1 - \theta_p \rho)(1 + \theta_p \rho)} \quad \text{et} \quad \beta = (\theta_1 + \rho \quad \theta_2 - \theta_1 \rho \quad \dots \quad \theta_p - \theta_{p-1} \rho)' \quad (\text{III.5})$$

ainsi que la valeur limite

$$\theta^* = \alpha(I_p - \theta_p \rho J_p)\beta. \quad (\text{III.6})$$

**Théorème III.1.** Soit  $\widehat{\theta}_n$  l'estimateur des moindres carrés donné par (III.4) dans le modèle (III.1) tel que  $\mathbb{E}[V_1^2] = \sigma^2 < +\infty$ . Alors, on a la convergence presque sûre

$$\lim_{n \rightarrow \infty} \widehat{\theta}_n = \theta^* \quad \text{p.s.}$$

où la valeur limite est donnée par (III.6).

Nous illustrerons sur la Figure III.3 les convergences de  $\widehat{\theta}_n$  composante par composante, pour nos deux exemples. Lorsque  $\rho = 0$ , alors  $\alpha = 1$  et  $\beta = \theta$ . On retrouve bien le résultat de la Proposition I.5. Considérons maintenant le cas où  $p = 1$ . Alors, selon (III.5), nous avons  $\alpha = (1 - \theta\rho)^{-1}(1 + \theta\rho)^{-1}$  et  $\beta = \theta + \rho$ . Il est immédiat de voir que, par application de (III.6), la valeur limite de  $\widehat{\theta}_n$  se réduit à

$$\theta^* = \frac{\theta + \rho}{1 + \theta\rho}$$

ce qui justifie le Théorème II.1.

### III.2.1.2 Normalité asymptotique

Étudions désormais la dispersion de l'estimateur autour de sa valeur limite. Pour cela, on commence par définir la matrice carrée d'ordre  $p + 2$  suivante,

$$B = \begin{pmatrix} 1 & -\beta_1 & -\beta_2 & \dots & \dots & -\beta_{p-1} & -\beta_p & \theta_p \rho \\ -\beta_1 & 1 - \beta_2 & -\beta_3 & \dots & \dots & -\beta_p & \theta_p \rho & 0 \\ -\beta_2 & -\beta_1 - \beta_3 & 1 - \beta_4 & \dots & \dots & \theta_p \rho & 0 & 0 \\ \vdots & \vdots & \vdots & & & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & & & \vdots & \vdots & \vdots \\ -\beta_p & -\beta_{p-1} + \theta_p \rho & -\beta_{p-2} & \dots & \dots & -\beta_1 & 1 & 0 \\ \theta_p \rho & -\beta_p & -\beta_{p-1} & \dots & \dots & -\beta_2 & -\beta_1 & 1 \end{pmatrix} \quad (\text{III.7})$$

où  $\beta$  est donnée par (III.5).

**Lemme III.1.** Sous les conditions de stabilité  $\|\theta\|_1 < 1$  et  $|\rho| < 1$ , la matrice  $B$  d'ordre  $p + 2$  définie par (III.7) est inversible. De plus, la matrice  $C$  d'ordre  $p + 1$ , obtenue en supprimant la première ligne et la première colonne de  $B$ , est également inversible.

De l'inversibilité de  $B$ , il suit que le système linéaire de  $p + 2$  équations à  $p + 2$  inconnues, défini par  $B\Lambda = e$ , possède l'unique solution

$$\Lambda = B^{-1}e \quad (\text{III.8})$$

où le vecteur  $e = (1 \ 0 \ \dots \ 0)'$  de  $\mathbb{R}^{p+2}$  est le même que celui défini en préambule, mais en dimension supérieure. On note  $\lambda_0, \dots, \lambda_{p+1}$  les éléments de  $\Lambda$ , et l'on considère la matrice de Toeplitz d'ordre  $p$ ,

$$\Delta_p = \begin{pmatrix} \lambda_0 & \lambda_1 & \lambda_2 & \dots & \dots & \lambda_{p-1} \\ \lambda_1 & \lambda_0 & \lambda_1 & \dots & \dots & \lambda_{p-2} \\ \vdots & \vdots & \vdots & & & \vdots \\ \vdots & \vdots & \vdots & & & \vdots \\ \lambda_{p-1} & \lambda_{p-2} & \lambda_{p-3} & \dots & \dots & \lambda_0 \end{pmatrix}. \quad (\text{III.9})$$

**Lemme III.2.** *Sous les conditions de stabilité  $\|\theta\|_1 < 1$  et  $|\rho| < 1$ , la matrice  $\Delta_p$  d'ordre  $p$  donnée par (III.9) est définie positive, pour tout  $p \geq 1$ .*

Cette stratégie présente de nombreux points communs avec celle de Yule-Walker, que nous avons détaillée dans la Section I.2.1.2. La matrice  $\Delta_p$  est donc inversible, et nous pouvons alors stipuler la normalité asymptotique de  $\widehat{\theta}_n$  munie de la covariance

$$\Sigma_\theta = \alpha^2 (I_p - \theta_p \rho J_p) \Delta_p^{-1} (I_p - \theta_p \rho J_p). \quad (\text{III.10})$$

**Théorème III.2.** *Soit  $\widehat{\theta}_n$  l'estimateur des moindres carrés donné par (III.4) dans le modèle (III.1) tel que  $\mathbb{E}[V_1^4] = \tau^4 < +\infty$ . Alors, on a la normalité asymptotique*

$$\sqrt{n} (\widehat{\theta}_n - \theta^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_\theta)$$

où la covariance limite est donnée par (III.10).

Lorsque  $\rho = 0$ , il est immédiat de voir que la variance asymptotique se réduit à  $\Delta_p^{-1}$ , ce qui est bien cohérent avec la Proposition I.6 (le  $\sigma^2$  étant inclus dans notre calcul de  $\Delta_p$ ). Par ailleurs, pour  $p = 1$ , la résolution du système (III.8) conduit à

$$\lambda_0 = \frac{1 + \theta\rho}{(1 - \theta^2)(1 - \theta\rho)(1 - \rho^2)}$$

et l'on vérifie ainsi que l'on retrouve la variance asymptotique du Théorème II.2 (il en ira bien sûr de même pour le Théorème II.3 puisque les valeurs limites reposent sur la même quantité). Nous illustrons la normalité asymptotique sur la Figure III.2. Celle-ci provient d'un échantillon de  $N = 1000$  réalisations de l'exemple détaillé précédemment avec  $p = 3$ , pour les première et troisième composantes de  $\widehat{\theta}_n$ . On remarque enfin que la construction de la matrice  $\Sigma_\theta$  la rend de fait bisymétrique, ce qui explique la raison pour laquelle les courbes théoriques superposées sont identiques.

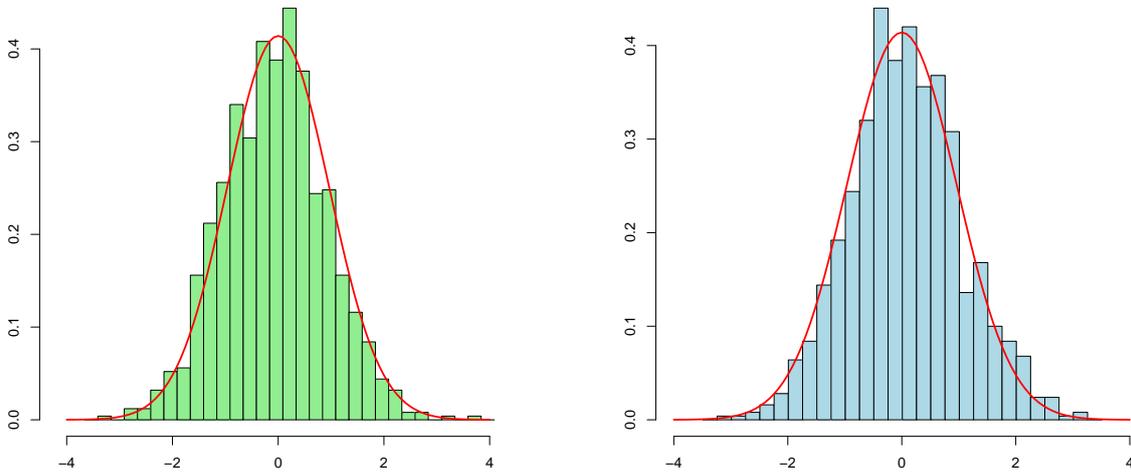


FIGURE III.2 – Normalité asymptotique de  $\widehat{\theta}_n$  dans un AR(3)–AR(1) stable (première composante à gauche, troisième composante à droite).

### III.2.1.3 Vitesse de convergence

Nous étudions enfin la vitesse de convergence presque sûre des bornes inférieure et supérieure de l'erreur  $\widehat{\theta}_n - \theta^*$ , ainsi que son taux d'accroissement.

**Théorème III.3.** Soit  $\widehat{\theta}_n$  l'estimateur des moindres carrés donné par (III.4) dans le modèle (III.1) tel que  $\mathbb{E}[V_1^4] = \tau^4 < +\infty$ . Alors, on a la loi forte quadratique

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{t=p}^n (\widehat{\theta}_t - \theta^*)(\widehat{\theta}_t - \theta^*)' = \Sigma_\theta \quad \text{p.s.}$$

où la valeur limite est donnée par (III.10). On a de plus, pour tout  $v \in \mathbb{R}^p$ , la loi du logarithme itéré

$$\limsup_{n \rightarrow \infty} \sqrt{\frac{n}{2 \log \log n}} v' (\widehat{\theta}_n - \theta^*) = -\liminf_{n \rightarrow \infty} \sqrt{\frac{n}{2 \log \log n}} v' (\widehat{\theta}_n - \theta^*) = \sqrt{v' \Sigma_\theta v} \quad \text{p.s.}$$

En conséquence,

$$\limsup_{n \rightarrow \infty} \sqrt{\frac{n}{2 \log \log n}} (\widehat{\theta}_n - \theta^*)(\widehat{\theta}_n - \theta^*)' = \Sigma_\theta \quad \text{p.s.}$$

Ce dernier résultat implique en particulier que l'on a la convergence

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{t=p}^n \|\widehat{\theta}_t - \theta^*\|^2 = \text{tr}(\Sigma_\theta) \quad \text{p.s.}$$

ainsi que la vitesse de convergence presque sûre

$$\|\widehat{\theta}_n - \theta^*\|^2 = O\left(\frac{\log \log n}{n}\right) \quad \text{p.s.} \quad (\text{III.11})$$

*Démonstration.* Les Lemmes III.1–III.2 sont prouvés dans la Section III.5.1. Quant aux Théorèmes III.1–III.2–III.3, ils sont prouvés dans la Section III.5.2. ★

## III.2.2 Le paramètre de l'autocorrélation résiduelle

Comme lors de la Section II.2.2, nous construisons un ensemble résiduel ( $\widehat{\varepsilon}_t$ ) censé refléter l'ampleur du biais occasionné par la présence d'une autocorrélation résiduelle sur la limite de  $\widehat{\theta}_n$ . Pour tout  $p \leq t \leq n$ , on définit alors

$$\widehat{\varepsilon}_t = Y_t - \widehat{\theta}_n' \Phi_{t-1}^p \quad (\text{III.12})$$

avec, de manière arbitraire,  $\widehat{\varepsilon}_0 = \dots = \widehat{\varepsilon}_{p-1} = 0$ . Cela nous permet d'introduire un estimateur pour  $\rho$  donné, pour tout  $n \geq p$ , par

$$\widehat{\rho}_n = \frac{\sum_{t=p}^n \widehat{\varepsilon}_{t-1} \widehat{\varepsilon}_t}{\sum_{t=p}^n \widehat{\varepsilon}_{t-1}^2}. \quad (\text{III.13})$$

Ce dernier a une interprétation cohérente au sens des moindres carrés, puisqu'il permet de minimiser

$$\Delta_n(\rho) = \sum_{t=p}^n (\widehat{\varepsilon}_t - \rho \widehat{\varepsilon}_{t-1})^2.$$

Il est important de noter que nous retrouvons ici un problème scalaire, par opposition avec la section précédente. Cependant, nous verrons que les résultats sont pénibles à obtenir, car très calculatoires.

### III.2.2.1 Convergence presque sûre

Notons pour commencer la valeur limite

$$\rho^* = \theta_p \rho \theta_p^*. \quad (\text{III.14})$$

**Théorème III.4.** Soit  $\widehat{\rho}_n$  l'estimateur des moindres carrés donné par (III.13) dans le modèle (III.1) tel que  $\mathbb{E}[V_1^2] = \sigma^2 < +\infty$ . Alors, on a la convergence presque sûre

$$\lim_{n \rightarrow \infty} \widehat{\rho}_n = \rho^* \quad \text{p.s.}$$

où la valeur limite est donnée par (III.14).

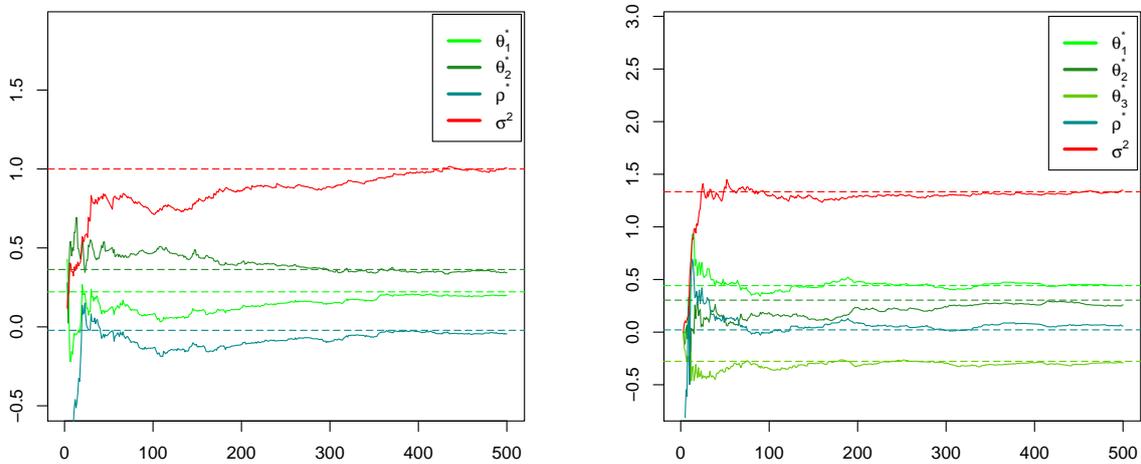


FIGURE III.3 – Convergence de  $\widehat{\theta}_n$ ,  $\widehat{\rho}_n$  et  $\widehat{\sigma}_n^2$  pour deux AR( $p$ )–AR(1) stables, avec  $p = 2$  (gauche) puis avec  $p = 3$  (droite).

### III.2.2.2 Normalité asymptotique

Nous voici arrivés au point capital de l'étude. La normalité asymptotique de  $\widehat{\rho}_n$  sera bien évidemment la clé de voûte de notre procédure de test – et du H-test – mais également la plus malaisée à établir. Nous allons pour cela commencer par stipuler la distribution asymptotique jointe de nos estimateurs, que nous illustrerons ensuite pour  $\widehat{\rho}_n$  sur la Figure III.4, relative à nos deux exemples. Soit la matrice carrée d'ordre  $p + 1$ ,

$$P = \begin{pmatrix} P_B & 0 \\ P_L' & \varphi \end{pmatrix} \quad (\text{III.15})$$

avec

$$\begin{aligned} P_B &= \alpha(I_p - \theta_p \rho J_p) \Delta_p^{-1}, \\ P_L &= J_p(I_p - \theta_p \rho J_p)(\alpha \theta_p \rho \Delta_p^{-1} e + \theta_p^* \beta), \\ \varphi &= -\alpha^{-1} \theta_p^*. \end{aligned}$$

Nous considérons ensuite la matrice de Toeplitz  $\Delta_{p+1}$ , définie comme  $\Delta_p$  en (III.9), mais en dimensions supérieure. Ainsi,

$$\Delta_{p+1} = \begin{pmatrix} \Delta_p & J_p \Lambda_p^1 \\ \Lambda_p^1 J_p & \lambda_0 \end{pmatrix}$$

où  $\Lambda_p^1 = (\lambda_1 \ \lambda_2 \ \dots \ \lambda_p)'$ . Nous introduisons alors la matrice de covariance semi-définie positive  $\Gamma = P\Delta_{p+1}P'$ , d'ordre  $p + 1$ , explicitée par

$$\Gamma = \begin{pmatrix} \Sigma_\theta & \theta_p \rho J_p \Sigma_\theta e \\ \theta_p \rho e' J_p \Sigma_\theta & \sigma_\rho^2 \end{pmatrix} \quad (\text{III.16})$$

où  $\Sigma_\theta$  est la matrice d'ordre  $p$  donnée par (III.10), et

$$\sigma_\rho^2 = P_L' \Delta_p P_L - 2\alpha^{-1} \theta_p^* \Lambda_p^1 J_p P_L + (\alpha^{-1} \theta_p^*)^2 \lambda_0. \quad (\text{III.17})$$

**Théorème III.5.** Soient  $\widehat{\theta}_n$  et  $\widehat{\rho}_n$  les estimateurs des moindres carrés donnés par (III.4) et (III.13) dans le modèle (III.1) tel que  $\mathbb{E}[V_1^4] = \tau^4 < +\infty$ . Alors, on a la normalité asymptotique jointe

$$\sqrt{n} \begin{pmatrix} \widehat{\theta}_n - \theta^* \\ \widehat{\rho}_n - \rho^* \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma)$$

où la matrice de covariance limite est donnée par (III.16). En particulier, on a la normalité asymptotique

$$\sqrt{n}(\widehat{\rho}_n - \rho^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_\rho^2)$$

où la variance limite est donnée par (III.17).

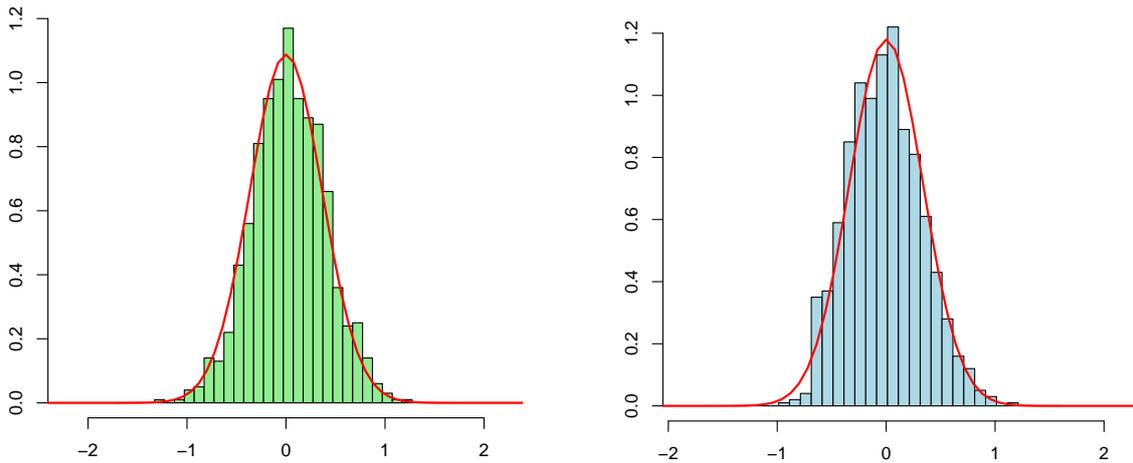


FIGURE III.4 – Normalité asymptotique de  $\widehat{\rho}_n$  pour deux  $\text{AR}(p)\text{-AR}(1)$  stables, avec  $p = 2$  (gauche) puis avec  $p = 3$  (droite).

Dans le Chapitre II, nous avons longuement commenté et mis en évidence l'existence d'un cas pathologique, correspondant à  $\theta = -\rho$ , pour lequel la matrice  $\Gamma$  n'était pas inversible. Il est naturel d'envisager un cas de figure similaire à l'ordre  $p$ . De fait, on a

$$\det(\Gamma) = \det(P)^2 \det(\Delta_{p+1}) = \alpha^{2(p-1)} (\theta_p^*)^2 \det(\Delta_{p+1}) \left( \frac{\det(I_p - \theta_p \rho J_p)}{\det(\Delta_p)} \right)^2$$

d'où il ressort que, puisque  $\det(\Delta_p) \neq 0$  et  $\det(\Delta_{p+1}) \neq 0$  (Lemme III.2), et que  $\det(I_p - \theta_p \rho J_p) \neq 0$  (la matrice étant à diagonale strictement dominante),  $\Gamma$  n'est pas inversible si et seulement si  $\theta_p^* = 0$ , c'est-à-dire lorsque  $\theta_p - \theta_{p-1} \rho = \theta_p \rho (\theta_1 + \rho)$ . D'une manière générale, englobant le chapitre précédent,  $\Gamma$  n'est pas inversible lorsque la dernière composante de  $\theta^*$  est nulle. Si  $p = 1$ , la valeur limite de  $\widehat{\rho}_n$  ne s'annule pas sous la condition  $\{\theta \neq -\rho\} \cap \{\theta \neq 0\} \cap \{\rho \neq 0\}$ , alors qu'en dimension quelconque, une condition nécessaire et suffisante est  $\{\theta_p - \theta_{p-1} \rho \neq \theta_p \rho (\theta_1 + \rho)\} \cap \{\theta_p \neq 0\} \cap \{\rho \neq 0\}$ . On voit par ailleurs que,  $\Delta_{p+1}$  étant définie positive, une condition suffisante (mais pas optimale) pour que  $\sigma_\rho^2$  ne s'annule pas, et donc que le théorème central limite associé à  $\widehat{\rho}_n$  ne soit pas dégénéré, est également que  $\theta_p^* \neq 0$ . Sous l'hypothèse nulle  $\mathcal{H}_0 : \rho = 0$  pour  $p = 1$ , nous avons alors considéré  $\{\theta \neq 0\} \cap \{\theta \neq -\rho\}$  comme cadre de validité de la procédure de test associée. Dans la suite, nous supposons donc que la condition  $\theta_p \neq 0$  est satisfaite, ce qui n'est pas choquant dans la mesure où l'on se propose d'étudier le processus autorégressif d'ordre  $p$ , et que  $\theta_p^* \neq 0$ , correspondant au cas pathologique en dimension quelconque. Revenons désormais à l'étude de l'estimateur  $\widehat{\rho}_n$ .

### III.2.2.3 Vitesse de convergence

Par analogie avec la Section III.2.1.3, nous concluons l'étude de l'estimateur de  $\rho$  par la vitesse de convergence presque sûre des bornes inférieure et supérieure de l'erreur  $\widehat{\rho}_n - \rho^*$ , ainsi que son taux d'accroissement.

**Théorème III.6.** *Soit  $\widehat{\rho}_n$  l'estimateur des moindres carrés donné par (III.13) dans le modèle (III.1) tel que  $\mathbb{E}[V_1^4] = \tau^4 < +\infty$ . Alors, on a la loi forte quadratique*

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{t=p}^n (\widehat{\rho}_t - \rho^*)^2 = \sigma_\rho^2 \quad \text{p.s.}$$

où la valeur limite est donnée par (III.17). On a de plus la loi du logarithme itéré

$$\limsup_{n \rightarrow \infty} \sqrt{\frac{n}{2 \log \log n}} (\widehat{\rho}_n - \rho^*) = -\liminf_{n \rightarrow \infty} \sqrt{\frac{n}{2 \log \log n}} (\widehat{\rho}_n - \rho^*) = \sigma_\rho \quad \text{p.s.}$$

Ce dernier résultat implique en particulier que

$$\limsup_{n \rightarrow \infty} \left( \frac{n}{2 \log \log n} \right) (\widehat{\rho}_n - \rho^*)^2 = \sigma_\rho^2 \quad \text{p.s.}$$

d'où l'on déduit la vitesse de convergence presque sûre

$$(\widehat{\rho}_n - \rho^*)^2 = O\left(\frac{\log \log n}{n}\right) \quad \text{p.s.} \quad (\text{III.18})$$

*Démonstration.* Les Théorèmes III.4–III.5–III.6 sont prouvés dans la Section III.5.2. ★

Là encore, nous pouvons vérifier que lorsque  $p = 1$ , les Théorèmes III.4–III.5–III.6 sont compatibles avec les théorèmes relatifs à la section précédente. Il s'agit d'une conclusion triviale quant à la valeur de  $\rho^*$ , mais très calculatoire quant à celle de  $\sigma_\rho^2$ . On obtient alors

$$P_L = \frac{(1 - \theta\rho)(\theta^3\rho^3 + 2\theta^2\rho^2 + (\theta + \rho)^2 + \theta\rho)}{(1 + \theta\rho)^2} \quad \text{et} \quad \varphi = -(1 - \theta\rho)(1 + \theta\rho)\theta^*.$$

On vérifie que la résolution des équations de Yule-Walker, et donc l'établissement des limites  $\lambda_0$ ,  $\lambda_1$  et  $\lambda_2$ , nous conduit bien à la valeur de  $\sigma_\rho^2$  donnée en (II.10). Il est à noter que la complexité des expressions explicites grandit drastiquement avec  $p$ . À titre d'exemple, pour  $p = 2$ ,

$$\begin{aligned} \sigma_\rho^2 = & (1 - \theta_2\rho)^{-3}(1 + \theta_2\rho)^{-3}(-2\theta_1^2\theta_2^5\rho^6 - 2\theta_1^2\theta_2^4\rho^6 - 5\theta_1^2\theta_2^4\rho^4 - 8\theta_1^2\theta_2^3\rho^4 - 3\theta_1^2\theta_2^2\rho^4 + \theta_1^2\theta_2^2\rho^2 \\ & + 2\theta_1^2\theta_2\rho^2 + \theta_1^2\rho^2 - 2\theta_1\theta_2^6\rho^7 + 2\theta_1\theta_2^6\rho^5 - 2\theta_1\theta_2^5\rho^7 + 2\theta_1\theta_2^5\rho^5 - 8\theta_1\theta_2^4\rho^5 + 8\theta_1\theta_2^4\rho^3 \\ & - 8\theta_1\theta_2^3\rho^5 + 8\theta_1\theta_2^3\rho^3 + 2\theta_1\theta_2^2\rho^3 - 2\theta_1\theta_2^2\rho + 2\theta_1\theta_2\rho^3 - 2\theta_1\theta_2\rho + \theta_2^8\rho^8 + 2\theta_2^7\rho^6 + \theta_2^6\rho^6 \\ & - 2\theta_2^6\rho^4 - 2\theta_2^5\rho^6 - 2\theta_2^5\rho^4 - 5\theta_2^4\rho^6 + 7\theta_2^4\rho^4 - 3\theta_2^4\rho^2 + 2\theta_2^3\rho^4 + \theta_2^2\rho^4 - \theta_2^2\rho^2 + \theta_2^2). \end{aligned}$$

On comprend donc qu'il est parfaitement inutile de chercher une formulation explicite de  $\sigma_\rho^2$  pour un  $p$  quelconque.

### III.2.3 La variance résiduelle

Les estimateurs  $\widehat{\theta}_n$  et  $\widehat{\rho}_n$  sont également autonormalisés, et ni leur limite presque sûre ni leur variance asymptotique ne dépendent de  $\sigma^2$ . Il est pourtant d'intérêt statistique considérable d'estimer la variance résiduelle de façon consistante. Nous proposons alors d'utiliser l'estimateur donné, pour tout  $n \geq p$ , par

$$\widehat{\sigma}_n^2 = \left(1 - \frac{\widehat{\rho}_n^2}{\widehat{\theta}_{p,n}^2}\right) \frac{1}{n} \sum_{t=p}^n \widehat{\varepsilon}_t^2. \quad (\text{III.19})$$

Nous avons illustré son comportement asymptotique sur la Figure III.3, dans la section précédente.

**Théorème III.7.** Soit  $\widehat{\sigma}_n^2$  l'estimateur donné par (III.19) dans le modèle (III.1) tel que  $\mathbb{E}[V_1^2] = \sigma^2 < +\infty$ . Alors, on a la convergence presque sûre

$$\lim_{n \rightarrow \infty} \widehat{\sigma}_n^2 = \sigma^2 \quad \text{p.s.}$$

*Démonstration.* Nous laissons la démonstration de ce résultat au soin du lecteur. Ce dernier découle en effet immédiatement des Théorèmes III.1 et III.4, et des preuves associées. ★

### III.3 Application au test de Durbin-Watson

Nous passerons rapidement sur les propriétés asymptotiques de  $\widehat{D}_n$  car la symétrie avec la Section II.3 est frappante, la seule différence résidant dans les valeurs limites  $D^*$  et  $\sigma_D^2$ . Nous proposons ainsi le résumé suivant.

**Corollaire III.1.** Soit  $\widehat{D}_n$  la statistique de Durbin-Watson donnée par (II.13). Dès que  $\mathbb{E}[V_1^2] = \sigma^2 < +\infty$ , on a la convergence presque sûre

$$\lim_{n \rightarrow \infty} \widehat{D}_n = D^* \quad \text{p.s.}$$

où la valeur limite est définie par  $D^* = 2(1 - \rho^*)$ . De plus, dès que  $\mathbb{E}[V_1^4] = \tau^4 < +\infty$ , on a la normalité asymptotique

$$\sqrt{n} (\widehat{D}_n - D^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_D^2)$$

où variance limite est définie par  $\sigma_D^2 = 4\sigma_\rho^2$ .

*Démonstration.* La preuve du Corollaire III.1 est quasiment identique à celle du Corollaire II.1, dans la Section II.6.2.1. ★

Là encore, il est possible d'obtenir la vitesse de convergence presque sûre de  $\widehat{D}_n$  vers  $D^*$  par l'intermédiaire d'une loi forte quadratique et d'une loi du logarithme itéré similaires aux Théorèmes III.3 et III.6. Notre intérêt repose ici exclusivement sur le test d'hypothèse

$$\mathcal{H}_0 : \text{“}\rho = 0\text{”} \quad \text{vs} \quad \mathcal{H}_1 : \text{“}\rho \neq 0\text{”}$$

pour les raisons évoquées précédemment. Tel que nous l'avions fait dans la Section II.4, nous souhaitons mettre à l'épreuve la normalité asymptotique de  $\widehat{D}_n$  en la comparant avec les tests usuels de corrélation résiduelle. Nous mènerons ensuite une étude approfondie afin de montrer que notre procédure de test simplifie le H-test de Durbin [47], et le généralise sous certains aspects.

### III.3.1 Comparaisons empiriques

Commençons par établir le socle théorique de la procédure de test.

**Théorème III.8.** *Supposons que  $\mathbb{E}[V_1^4] = \tau^4 < +\infty$ ,  $\theta_p \neq 0$  et  $\theta_p^* \neq 0$ . Alors, si l'on se place sous  $\mathcal{H}_0$  : “ $\rho = 0$ ”, on a la distribution asymptotique*

$$\frac{n}{4\widehat{\theta}_{p,n}^2} (\widehat{D}_n - 2)^2 \xrightarrow{\mathcal{L}} \chi_1^2$$

où  $\chi_1^2$  désigne une variable aléatoire distribuée selon une loi du khi-deux à un degré de liberté. De plus, si l'on se place sous  $\mathcal{H}_1$  : “ $\rho \neq 0$ ”, alors

$$\lim_{n \rightarrow \infty} \frac{n}{4\widehat{\theta}_{p,n}^2} (\widehat{D}_n - 2)^2 = +\infty \quad \text{p.s.}$$

*Démonstration.* Le Théorème III.8 sera prouvé dans la Section III.5.3. ★

Pour un niveau de significativité  $0 < \alpha < 1$ , on construit ainsi une zone d'acceptation  $\mathcal{A} = [0, z_\alpha]$  et une zone de rejet  $\mathcal{R} = ]z_\alpha, +\infty[$  à partir du quantile d'ordre  $1 - \alpha$  de la distribution du khi-deux considérée. Selon les conventions en vigueur associées aux tests d'hypothèse, on rejettera l'hypothèse nulle d'absence de corrélation résiduelle dès que la statistique de test proposée sera supérieure à  $z_\alpha$ . Il s'agit désormais, pour évaluer la puissance empirique d'un paramétrage, de calculer la fréquence de rejet de  $\mathcal{H}_0$  sur  $N = 1000$  simulations, pour les procédures que nous avons pu décrire dans la Section II.4.2. Sur les Figures III.5–III.6, nous avons représenté la fréquence de non rejet de  $\mathcal{H}_0$  pour les 5 procédures de test considérées,  $\rho$  variant de  $-0.95$  à  $0.95$  avec  $\theta = (0.5, 0.2)'$  et  $(V_t) \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$  dans un premier temps, avec  $\theta = (-0.2, 0.1, -0.3)'$  et  $(V_t) \stackrel{\text{iid}}{\sim} \mathcal{U}([-2, 2])$  dans un second temps, pour  $n = 300$ . Nous y avons adjoint les résultats obtenus à partir des mêmes configurations sur des petits échantillons, pour  $n = 30$ . La légende précise simplement les abréviations explicites des procédures.

Nos conclusions sont essentiellement les mêmes que celles de l'étude empirique du cas univarié. Notre procédure est asymptotiquement aussi puissante que les tests usuels, et supérieure aux tests *du portemanteau* pour les raisons évoquées précédemment, ces derniers n'étant pas adaptés (et trop souvent appliqués!) au cadre autorégressif. Le cas pathologique, correspondant respectivement à  $\rho \approx 0.30$  et à  $\rho \approx -0.77$  dans les exemples, est en outre assez bien géré par notre procédure, comme nous pouvons le voir. Sur les petits échantillons, la procédure se révèle également d'une manière générale plus sensible à la présence d'une autocorrélation résiduelle et clairement supérieure sous  $\mathcal{H}_1$  même si, disons-le, les conclusions sont à relativiser puisque l'on applique des résultats de temps long à des valeurs faibles de  $n$ . Nous allons conclure cette étude en montrant que la statistique de test que l'on utilise est en réalité un équivalent asymptotique de celle suggérée par le H-test, qu'elle

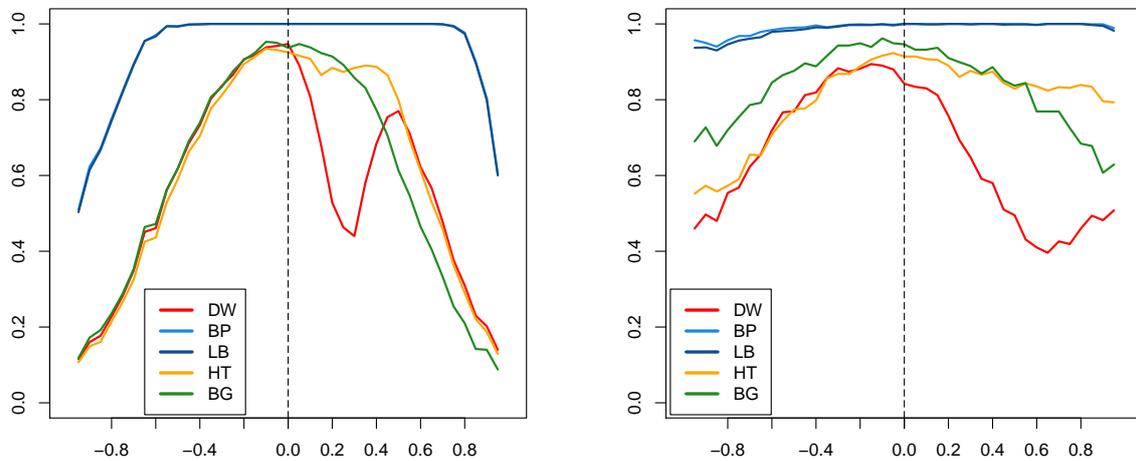


FIGURE III.5 – Fréquence de non rejet de  $\mathcal{H}_0$  en ordonnée pour  $\theta = (0.5, 0.2)'$  et  $\rho$  variant de  $-0.95$  à  $0.95$  en abscisse, avec  $n = 300$  (gauche) et  $n = 30$  (droite).

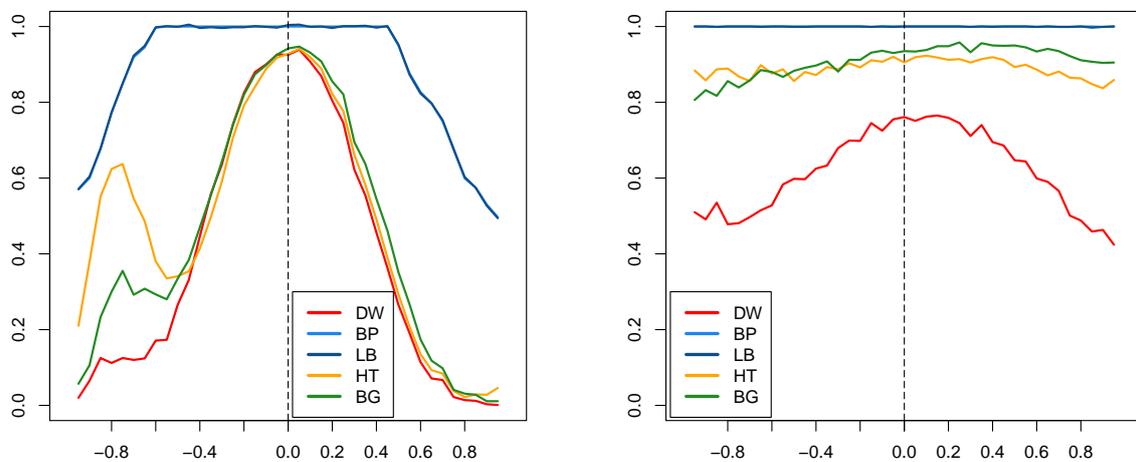


FIGURE III.6 – Fréquence de non rejet de  $\mathcal{H}_0$  en ordonnée pour  $\theta = (-0.2, 0.1, -0.3)'$  et  $\rho$  variant de  $-0.95$  à  $0.95$  en abscisse, avec  $n = 300$  (gauche) et  $n = 30$  (droite).

généralise et améliore de fait, s'appliquant à un panel plus vaste de processus et générant moins de perturbations sur les petits échantillons.

### III.3.2 Un équivalent du H-test

Le H-test, l'axe principal de [47], suggère la comparaison de la statistique

$$\widehat{H}_n = \widehat{\rho}_n \sqrt{\frac{n}{1 - n\widehat{\mathbb{V}}(\widehat{\theta}_{p,n})}} \quad (\text{III.20})$$

avec une déviation normale standard, pour tester l'hypothèse  $\mathcal{H}_0 : \rho = 0$ , ce résultat étant par ailleurs stipulé dans un cadre de normalité résiduelle sous une forme rappelant fermement un test de Student, et pour lequel la stratégie du maximum de vraisemblance est adaptée (puisque l'on connaît la loi conditionnelle de l'estimateur). En conséquence, il n'est pas prouvé dans le cas général que nous nous proposons ici de considérer. Il est entendu que  $\widehat{\mathbb{V}}(\widehat{\theta}_n)$  symbolise un estimateur consistant de la variance de l'estimateur  $\widehat{\theta}_n$ , au sens des moindres carrés.

**Théorème III.9.** *Supposons que  $\mathbb{E}[V_1^2] = \sigma^2 < +\infty$ . Alors, si l'on se place sous  $\mathcal{H}_0 : \rho = 0$ , on a l'équivalence asymptotique presque sûre*

$$\frac{n}{4\widehat{\theta}_{p,n}^2} (\widehat{D}_n - 2)^2 \sim \widehat{\rho}_n^2 \left[ \frac{n}{1 - n\widehat{\mathbb{V}}(\widehat{\theta}_{p,n}^2)} \right] \quad \text{p.s.}$$

*Démonstration.* Le Théorème III.9 est prouvé dans la Section III.5.3. ★

## III.4 Conclusion et perspectives

Tout comme lors des perspectives du chapitre précédent, il nous faut d'abord noter qu'il est aisé de produire des estimateurs consistants de  $\theta$  et de  $\rho$ . C'est un point que nous n'avons pas forcément développé car la finalité de l'étude ne s'y prêtait pas. Cependant, un estimateur consistant de  $\beta$  est donné par

$$\widehat{\beta}_n = \left( 1 - \frac{\widehat{\rho}_n^2}{\widehat{\theta}_{p,n}^2} \right) \left( I_p - \frac{\widehat{\rho}_n}{\widehat{\theta}_{p,n}} J_p \right)^{-1} \widehat{\theta}_n$$

et l'on montre que toute racine du polynôme  $R_n(z) = z^{p+1} - \widehat{\beta}_{1,n}z^p - \dots - \widehat{\beta}_{p,n}z + \widehat{\rho}_n/\widehat{\theta}_{p,n}$  est un estimateur consistant de  $\rho$ . On construit alors un estimateur de  $\theta$  à partir des relations le liant à  $\beta$  et à  $\rho$ , système qui génère  $p + 1$  solutions (rappelons-nous que nous avons obtenu 2 solutions pour  $p = 1$  dans le chapitre précédent). En conclusion, nous espérons avoir apporté un point de vue novateur sur la statistique de Durbin-Watson, établi sa normalité asymptotique dans un cadre général, et ainsi amélioré les procédures de test de corrélation résiduelle des modèles autorégressifs, par une large utilisation de techniques de martingales. Nous sommes en outre convaincus que, par l'intermédiaire de calculs certes pénibles, il serait possible de bâtir une statistique permettant l'évaluation de l'hy-

pothèse  $\mathcal{H}_0 : “\rho_1 = 0, \dots, \rho_q = 0”$  dans un modèle AR( $p$ )–AR( $q$ ), contre son alternative naturelle qu’il existe au moins une autocorrélation significative dans les résidus. L’efficacité des résultats que nous obtenons pour  $q = 1$  est en ce sens prometteur, et cela permettrait de combler l’évidente limitation de la procédure à la première corrélation dont, à titre d’exemple, le test de Breusch-Godfrey sait tenir compte. C’est un travail actuellement en cours. Durbin proposait d’ailleurs un aperçu d’une telle stratégie en conclusion de son article [47], hélas bien délicate à mettre en pratique car reposant sur des séries entières d’ordre infini, et cantonnée à la normalité résiduelle. Des approches récentes s’appuyant sur des méthodes de point-selle ont également vu le jour, par exemple dans [28], pour approximer la distribution d’un ratio de formes quadratiques, comme c’est le cas pour la statistique de Durbin-Watson. Les conditions de stabilité retenues semblent par ailleurs restrictives mais elles collent à l’aspect statistique de l’étude car, une fois stationnarisés, les processus que l’on se donne en pratique sont de fait stabilisés. Nous aborderons la problématique de l’instabilité dans le Chapitre VI, d’un point de vue plus théorique. Quant au Chapitre IV, qu’il nous faut maintenant aborder, il traite de *principes de déviations modérées* appliqués à nos estimateurs. Nous souhaitons ainsi proposer des bornes supérieures à la distribution de  $\widehat{D}_n$ , et montrer que l’on peut maintenir l’approximation gaussienne à un panel de vitesses beaucoup plus vaste que celle du théorème central limite.

## III.5 Annexe : démonstration des résultats

### III.5.1 Un peu d’algèbre linéaire

Nous regroupons dans cette section les preuves de l’inversibilité des matrices  $B$  et  $\Delta_p$ , respectivement données par (III.7) et (III.9), qui revêtent une importance capitale dans tout le chapitre.

#### III.5.1.1 Preuve du Lemme III.1

Considérons la décomposition  $B = B_1 + \rho B_2$ , avec

$$B_1 = \begin{pmatrix} 1 & -\theta_1 & -\theta_2 & \dots & \dots & -\theta_{p-1} & -\theta_p & 0 \\ -\theta_1 & 1 - \theta_2 & -\theta_3 & \dots & \dots & -\theta_p & 0 & 0 \\ -\theta_2 & -\theta_1 - \theta_3 & 1 - \theta_4 & \dots & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & & & \vdots & \vdots & \vdots \\ -\theta_p & -\theta_{p-1} & -\theta_{p-2} & \dots & \dots & -\theta_1 & 1 & 0 \\ 0 & -\theta_p & -\theta_{p-1} & \dots & \dots & -\theta_2 & -\theta_1 & 1 \end{pmatrix}$$

et

$$B_2 = \begin{pmatrix} 0 & -1 & \theta_1 & \dots & \dots & \theta_{p-2} & \theta_{p-1} & \theta_p \\ -1 & \theta_1 & \theta_2 & \dots & \dots & \theta_{p-1} & \theta_p & 0 \\ \theta_1 & -1 + \theta_2 & \theta_3 & \dots & \dots & \theta_p & 0 & 0 \\ \vdots & \vdots & \vdots & & & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & & & \vdots & \vdots & \vdots \\ \theta_{p-1} & \theta_{p-2} + \theta_p & \theta_{p-3} & \dots & \dots & -1 & 0 & 0 \\ \theta_p & \theta_{p-1} & \theta_{p-2} & \dots & \dots & \theta_1 & -1 & 0 \end{pmatrix}.$$

Des inégalités triangulaires  $|\theta_i + \theta_j| \leq |\theta_i| + |\theta_j|$  et  $1 - |\theta_i| \leq |1 - \theta_i|$ , pour tout  $1 \leq i, j \leq p$ , on tire que  $B_1$  est à diagonale strictement dominante dès que  $\|\theta\|_1 < 1$ . Il suit alors du théorème de Lévy-Desplanques [63] que  $B_1$  est inversible. Ainsi,  $B = (I_{p+2} + \rho B_2 B_1^{-1}) B_1$ . Cette nouvelle décomposition est intéressante, d'une part car la matrice  $B_2 B_1^{-1}$  engendrée est creuse, d'autre part car l'inversibilité de  $B$  ne dépend plus désormais que du rayon spectral de  $\rho B_2 B_1^{-1}$ . On remarque que chaque ligne de  $B_2$  est en fait la ligne précédente de  $B_1$  changée de signe, et que sa première ligne est la dernière ligne de  $B_1$  réagencée. Par l'intervention de la matrice de passage correspondante, on obtient

$$B_2 B_1^{-1} = \begin{pmatrix} -\theta_1 & -1 - \theta_2 & \theta_1 - \theta_3 & \dots & \theta_{p-2} - \theta_p & \theta_{p-1} & \theta_p \\ -1 & 0 & \dots & \dots & \dots & \dots & 0 \\ 0 & -1 & 0 & \dots & \dots & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ 0 & \dots & \dots & 0 & -1 & 0 & 0 \\ 0 & \dots & \dots & \dots & 0 & -1 & 0 \end{pmatrix}.$$

Cette forme simplifiée entraîne que  $-1$  et  $1$  sont des valeurs propres de  $B_2 B_1^{-1}$  associées aux vecteurs propres de dimension  $p + 2$  respectivement donnés par  $(1 \ 1 \ \dots \ 1)'$  et  $(1 \ -1 \ \dots \ (-1)^{p+1})'$ . Soit  $P(\lambda) = \det(B_2 B_1^{-1} - \lambda I_{p+2})$  son polynôme caractéristique, pour tout  $\lambda \in \mathbb{C}$ . Alors,

$$P(\lambda) = (-\lambda)^{p+2} + \sum_{k=1}^{p+2} b_k (-\lambda)^{p+2-k}$$

où les coefficients  $(b_k)$  forment la première ligne de  $B_2 B_1^{-1}$ . Puisque  $1$  et  $-1$  sont valeurs propres, il existe un polynôme  $Q$  de degré  $p$  tel que  $P(\lambda) = (\lambda^2 - 1)Q(\lambda)$ , et son calcul explicite conduit à

$$Q(\lambda) = (-\lambda)^p - \sum_{k=1}^p \theta_k (-\lambda)^{p-k}.$$

On définit alors le polynôme  $R$  de degré  $p$  par

$$R(\lambda) = \lambda^p - \sum_{k=1}^p |\theta_k| \lambda^{p-k}$$

et l'on note que, pour tout  $\lambda \in \mathbb{C}$ , on a  $R(|\lambda|) \leq |Q(\lambda)|$ . Supposons maintenant qu'il existe une valeur propre  $\lambda_0 \in \mathbb{C}$  de  $B_2 B_1^{-1}$  telle que  $|\lambda_0| > 1$ . On aurait alors

$$R(|\lambda_0|) = |\lambda_0|^p - \sum_{k=1}^p |\theta_k| |\lambda_0|^{p-k} = |\lambda_0|^p \left( 1 - \sum_{k=1}^p |\theta_k| |\lambda_0|^{-k} \right) \geq |\lambda_0|^p (1 - \|\theta\|_1) > 0$$

sous la condition  $\|\theta\|_1 < 1$ . Cela contredit clairement le fait que  $\lambda_0$  puisse être une valeur propre de  $B_2 B_1^{-1}$  annulant  $Q$ , puisque  $0 < R(|\lambda_0|) \leq |Q(\lambda_0)|$ . Cette stratégie ressemble d'ailleurs fortement aux travaux de Cauchy sur la localisation des racines des polynômes algébriques, nous pouvons citer en exemple le Théorème 2.1 de [90]. En conclusion, toutes les racines de  $Q$  sont à l'intérieur du cercle unité. Ainsi,  $\rho(B_2 B_1^{-1}) = 1$ , et  $\rho(\rho B_2 B_1^{-1}) = |\rho| < 1$  sous la seconde condition de stabilité, ce qui confirme l'inversibilité de  $B$ . Considérons maintenant la matrice  $C$  d'ordre  $p + 1$  extraite de  $B$  par suppression de sa première ligne et de sa première colonne. On appelle  $b$  le premier élément diagonal de  $B^{-1}$  et l'on note que, par la méthode des cofacteurs,

$$b = \frac{\det(C)}{\det(B)}.$$

Mais nous tirons aussi du système (III.8) que  $b = \lambda_0 > 0$ , par corollaire du Lemme III.2. Il s'ensuit que  $\det(C) \neq 0$  et donc que  $C$  est inversible, ce qui achève la preuve. ★

### III.5.1.2 Preuve du Lemme III.2

Commençons par montrer que le rayon spectral de la matrice compagne  $C_A$  d'ordre  $p + 1$  du modèle (III.1) est à l'intérieur strict du cercle unité. Notons pour cela  $W_n = (V_n \ 0 \ \dots \ 0)' \in \mathbb{R}^{p+1}$  et considérons l'écriture vectorielle, pour  $n \geq p$ ,

$$\Phi_n^{p+1} = C_A \Phi_{n-1}^{p+1} + W_n \tag{III.21}$$

où  $\Phi_n^{p+1} = (Y_n \ Y_{n-1} \ \dots \ Y_{n-p})'$  est l'extension naturelle de (III.2), et

$$C_A = \begin{pmatrix} \theta_1 + \rho & \theta_2 - \theta_1 \rho & \dots & \theta_p - \theta_{p-1} \rho & -\theta_p \rho \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}.$$

Son polynôme caractéristique s'écrit, pour tout  $\mu \in \mathbb{C}$ ,  $P_A(\mu) = \det(C_A - \mu I_{p+1})$ , et l'on tire du Lemme 4.1.1 de [45] l'expression simplifiée suivante,

$$\begin{aligned} P_A(\mu) &= (-1)^p \left( \mu^{p+1} - (\theta_1 + \rho) \mu^p - \sum_{k=2}^p (\theta_k - \theta_{k-1} \rho) \mu^{p+1-k} + \theta_p \rho \right) \\ &= (-1)^p (\mu - \rho) \left( \mu^p - \sum_{k=1}^p \theta_k \mu^{p-k} \right) = (-1)^p (\mu - \rho) P(\mu) \end{aligned} \quad (\text{III.22})$$

où le polynôme  $P$  de degré  $p$  est défini par identification. Raisonnons une nouvelle fois par l'absurde, et supposons qu'il existe une valeur propre  $\mu_0 \in \mathbb{C}$  de  $C_A$  telle que  $|\mu_0| \geq 1$ . Sachant que  $|\rho| < 1$  par hypothèse, on a de fait  $\mu_0 \neq \rho$ ,  $P(\mu_0) = 0$ , et ainsi

$$1 - \sum_{k=1}^p \theta_k \mu_0^{-k} = 0 \quad (\text{III.23})$$

puisque  $\mu_0 \neq 0$ . Cependant,

$$\left| \sum_{k=1}^p \theta_k \mu_0^{-k} \right| \leq \sum_{k=1}^p |\theta_k| |\mu_0^{-k}| \leq \|\theta\|_1 < 1$$

sous l'hypothèse de stabilité, ce qui contredit naturellement (III.23). La matrice  $C_A$  est par conséquent de rayon spectral strictement inférieur à 1. Pour  $n \geq p$ , définissons le processus

$$\Psi_n^{p+1} = C_A \Psi_{n-1}^{p+1} + W_n$$

avec  $\Psi_n^{p+1} = (Y_n^* \ Y_{n-1}^* \ \dots \ Y_{n-p}^*)'$ , où  $(Y_t^*)$  est le processus stationnaire engendré par (III.1). L'existence de  $(Y_t^*)$  est assurée par la Proposition I.1 combinée au caractère causal du polynôme autorégressif associé, dont  $\rho(C_A) < 1$  est une condition suffisante. Rappelons que  $(Y_t)$  est une trajectoire du processus stationnaire indexé par  $\mathbb{Z}$ , qui perd son caractère stationnaire dès que l'on fixe arbitrairement un ensemble de valeurs initiales, et devient de fait asymptotiquement stationnaire. Il s'agit dès lors de voir la trajectoire associée  $(Y_t^*)$  comme une proche approximation de  $(Y_t)$  – aux observations initiales près – qu'il sera plus aisé de manier dans la suite de cette démonstration. Ainsi, par le Théo-

rème 4.4.2 de [25] et par analogie avec la Section I.2.1.3, la densité spectrale de  $(Y_t^*)$  est donnée, pour tout  $x$  dans le tore  $\mathbb{T} = [-\pi, \pi]$ , par

$$f_Y(x) = \frac{\sigma^2}{2\pi |\mathcal{A}(e^{-ix})|^2} \quad (\text{III.24})$$

où le polynôme autorégressif  $\mathcal{A}$  s'écrit, pour  $\mu \in \mathbb{C}^*$ ,  $\mathcal{A}(\mu) = (-1)^p \mu^{p+1} P_A(\mu^{-1})$ , avec  $P_A$  défini en (III.22), et  $\mathcal{A}(0) = 1$ . Pour tout  $h \in \mathbb{Z}$ , on dénote par

$$\widehat{f}_h = \int_{\mathbb{T}} f_Y(x) e^{ixh} dx$$

le coefficient de Fourier d'ordre  $h$  associé à la densité  $f_Y$ . Comme nous l'avons stipulé dans la Définition I.8, il est bien connu que  $\widehat{f}_h$  coïncide avec la fonction d'autocovariance du processus stationnaire  $(Y_t^*)$ . Il suit que, pour tout  $p \geq 1$ , la matrice de covariance associée à  $\Psi_n^p$  s'écrit, à l'aide de l'opérateur de Toeplitz  $T$  d'ordre  $p$ ,

$$T_p(f_Y) = (\widehat{f}_{i-j})_{1 \leq i, j \leq p} = \sigma^2 \Delta_p.$$

Ce résultat vient en effet du comportement asymptotique de  $(Y_t)$ , convergent vers la loi stationnaire de  $(Y_t^*)$ , d'où l'on tire

$$\lim_{n \rightarrow \infty} \mathbb{E} [\Phi_n^p \Phi_n^{p'}] = \mathbb{E} [\Psi_p^p \Psi_p^{p'}] = \sigma^2 \Delta_p. \quad (\text{III.25})$$

En effet, le système linéaire (III.8) est bâti tel celui de Yule-Walker (voir Section I.2.1.2), afin de mettre les autocovariances de  $(Y_t^*)$  en évidence et assurer la convergence (III.25), ce qui est bien décrit par exemple dans [25]. Il nous reste alors à déduire de la Proposition 4.5.3 de [25], ou des propriétés des formes de Toeplitz dûment approfondies dans [58], que

$$2\pi \min_{x \in \mathbb{T}} f_Y(x) \leq \lambda_{\min}(T_p(f_Y)) \leq \lambda_{\max}(T_p(f_Y)) \leq 2\pi \max_{x \in \mathbb{T}} f_Y(x)$$

et, puisque  $0 < f_Y(x) < +\infty$  pour tout  $x \in \mathbb{T}$ , en vertu de (III.24) –  $\mathcal{A}$  n'ayant aucune racine sur le cercle unité, – la plus petite valeur propre de  $\Delta_p$  est strictement positive quel que soit  $p \geq 1$ , ce qui conclut la preuve. ★

### III.5.2 Sur les estimateurs

Rappelons-nous que les preuves proposées dans cette section servent également à démontrer les Théorèmes II.1–II.2–II.3, en choisissant  $p = 1$ . Nous utiliserons des outils de martingales pour prouver nos résultats, et nous pouvons citer en exemple [45] et [59] pour en trouver les fondements mathématiques. Commençons par proposer deux lemmes techniques.

**Lemme III.3.** *Soit  $(Y_t)$  le processus engendré, pour tout  $p \leq t \leq n$ , par le modèle (III.1) tel que  $\|\theta\|_1 < 1$  et  $|\rho| < 1$ , et dont les valeurs initiales  $Y_0, \dots, Y_{p-1}, \varepsilon_0, \dots, \varepsilon_{p-1}$  possèdent les mêmes propriétés de moments que*

( $V_t$ ). Alors, s'il existe  $a \geq 1$  tel que  $\mathbb{E}[|V_1|^a] < +\infty$ , on a

$$\sum_{t=0}^n |Y_t|^a = O(n) \quad \text{p.s.} \quad \text{et} \quad \sup_{0 \leq t \leq n} |Y_t| = o(n^{1/a}) \quad \text{p.s.}$$

*Démonstration.* Nous considérerons ici que toutes les valeurs initiales formant  $\mathcal{F}_{p-1}$  sont nulles, de manière arbitraire. En effet, cela allège considérablement les expressions et n'impacte en rien les résultats dans la mesure où  $(\varepsilon_t)$  et  $(Y_t)$  sont asymptotiquement stationnaires. On considère, pour tout  $z \in \mathbb{C}$ , le polynôme  $\mathcal{A}(z) = 1 - \theta_1 z - \dots - \theta_p z^p$ , et l'on a alors  $\mathcal{A}(L)Y_n = \varepsilon_n$ . Comme nous l'avons vu dans la démonstration du Lemme III.2, sous la condition  $\|\theta\|_1 < 1$ , le polynôme  $\mathcal{A}$  est causal et donc inversible. Il vient alors  $Y_n = \mathcal{A}^{-1}(L)\varepsilon_n$ , et

$$\sup_{0 \leq t \leq n} |Y_t| = \sup_{0 \leq t \leq n} \left| \sum_{k=0}^t \psi_k \varepsilon_{t-k} \right| \leq \|\psi\|_\infty \sup_{0 \leq t \leq n} |\varepsilon_t| \quad (\text{III.26})$$

où la suite de coefficients  $(\psi_k)$  est décrite dans la Proposition I.1. Or, comme l'on sait par cette même proposition que la suite  $(\psi_k)$  est absolument sommable lorsque le polynôme  $\mathcal{A}$  est causal, on tire aisément que  $\|\psi\|_\infty < +\infty$ . Par un raisonnement similaire sur  $(\varepsilon_t)$ , on trouve

$$\sup_{0 \leq t \leq n} |\varepsilon_t| = \sup_{1 \leq t \leq n} \left| \sum_{k=0}^{t-1} \psi'_k V_{t-k} \right| \leq \|\psi'\|_\infty \sup_{1 \leq t \leq n} |V_t| \leq \frac{1}{1-|\rho|} \sup_{1 \leq t \leq n} |V_t|$$

où les coefficients  $(\psi'_k)$  sont issus de l'inversion du polynôme  $\mathcal{A}'(z) = 1 - \rho z$ , et dans la mesure où l'on montre facilement que  $\|\psi'\|_\infty \leq \|\psi'\|_1 \leq (1-|\rho|)^{-1}$  dans le cadre d'un modèle AR(1). Ainsi,

$$\sup_{0 \leq t \leq n} |Y_t| \leq \frac{\|\psi\|_\infty}{1-|\rho|} \sup_{1 \leq t \leq n} |V_t| = o(n^{1/a}) \quad \text{p.s.} \quad (\text{III.27})$$

par application du Théorème B.1. De plus, si l'on reprend (III.26) et que l'on considère  $b \geq 1$ , l'exposant conjugué de  $a$ , c'est-à-dire tel que  $1/a + 1/b = 1$ , on a

$$|Y_n|^a \leq \left( \sum_{t=0}^n |\psi_t|^{1/a} |\psi_t|^{1/b} |\varepsilon_{n-t}| \right)^a \leq \|\psi\|_1^{a/b} \sum_{t=0}^n |\psi_t| |\varepsilon_{n-t}|^a$$

par l'inégalité de Hölder. On en déduit

$$\sum_{t=0}^n |Y_t|^a \leq \|\psi\|_1^{a/b} \sum_{t=0}^n \sum_{k=0}^t |\psi_k| |\varepsilon_{t-k}|^a = \|\psi\|_1^{a/b} \sum_{t=0}^n |\varepsilon_t|^a \sum_{k=t}^n |\psi_{n-k}| \leq \|\psi\|_1^a \sum_{t=0}^n |\varepsilon_t|^a,$$

et l'on obtient de la même manière

$$\sum_{t=0}^n |Y_t|^a \leq \frac{\|\psi\|_1^a}{(1-|\rho|)^a} \sum_{t=1}^n |V_t|^a. \quad (\text{III.28})$$

Selon la Proposition I.1,  $\|\psi\|_1 < +\infty$ . La preuve est donc achevée en considérant (III.27), (III.28) et le Théorème B.1. ★

**Lemme III.4.** *Soit la matrice  $S_n$  donnée par (III.3). Si  $\mathbb{E}[V_1^2] = \sigma^2 < +\infty$ , alors sous les conditions de stabilité  $\|\theta\|_1 < 1$  et  $|\rho| < 1$ , on a la convergence presque sûre*

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = \sigma^2 \Delta_p \quad \text{p.s.}$$

où la matrice  $\Delta_p$  est définie en (III.9).

*Démonstration.* Pour tout  $n \geq p+1$ , il est facile de voir que le modèle (III.1) satisfait l'équation autorégressive fondamentale donnée, pour  $p+1 \leq t \leq n$ , par

$$Y_t = \beta' \phi_{t-1}^p - \theta_p \rho Y_{t-p-1} + V_t \quad (\text{III.29})$$

où nous reprenons les notations de la Section III.2. On voit dès lors qu'un processus  $(Y_t)$  engendré par un modèle AR( $p$ )-AR(1) possède en réalité la dynamique d'un processus AR( $p+1$ ). Le polynôme autorégressif d'ordre  $p+1$  associé est causal sous les conditions de stabilité  $\|\theta\|_1 < 1$  et  $|\rho| < 1$ , comme nous l'avons montré dans la preuve du Lemme III.1. Tout d'abord, reprenons la filtration associée au processus  $(Y_t)$  définie, pour  $n \geq 1$ , par  $\mathcal{F}_n = \sigma(Y_0, \dots, Y_{p-1}, \varepsilon_0, \dots, \varepsilon_{p-1}, V_1, \dots, V_n)$ . Pour tout  $0 \leq h \leq p+1$ , soient  $M_0^h = \dots = M_{h-1}^h = 0$  et, pour  $n \geq h$ ,

$$M_n^h = \sum_{t=h}^n Y_{t-h} V_t. \quad (\text{III.30})$$

Dès que  $h \geq 1$ ,  $\mathbb{E}[M_{n+1}^h | \mathcal{F}_n] = M_n^h + \mathbb{E}[Y_{n+1-h} V_{n+1} | \mathcal{F}_n] = M_n^h$  p.s. La suite  $(M_n^h)$  est donc une  $(\mathcal{F}_n)$ -martingale, dont on montre facilement qu'elle est localement de carré intégrable. Sa variation quadratique prévisible est donnée, pour  $n \geq h$ , par

$$\langle M^h \rangle_n = \sum_{t=h}^n \mathbb{E}[(\Delta M_t^h)^2 | \mathcal{F}_{t-1}] = \sigma^2 \sum_{t=h}^n Y_{t-h}^2. \quad (\text{III.31})$$

Par le Lemme III.3, la condition  $\mathbb{E}[V_1^2] = \sigma^2 < +\infty$  permet d'établir que  $\langle M^h \rangle_n = O(n)$  p.s. Dans la mesure où l'on a également  $n = O(\langle M^h \rangle_n)$  p.s. par le Corollaire 1.3.25 de [45], il s'ensuit que

$$\lim_{n \rightarrow \infty} \langle M^h \rangle_n = +\infty \quad \text{p.s.}$$

On en déduit, à l'aide du Théorème B.2 pour  $p = 1$ , que dès que  $h \geq 1$ ,

$$\lim_{n \rightarrow \infty} \frac{M_n^h}{n} = 0 \quad \text{p.s.} \quad (\text{III.32})$$

De plus, on tire facilement de la formulation (III.29), de (III.32) et du Théorème B.1, que

$$\lim_{n \rightarrow \infty} \frac{M_n^0}{n} = \sigma^2 \quad \text{p.s.} \quad (\text{III.33})$$

Nous avons dès lors tous les outils à disposition pour achever la preuve. En effet, si l'on note  $\ell_h$  la valeur limite de l'estimateur de l'autocovariance d'ordre  $h$  de  $(Y_t)$ , on a

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=h}^n Y_{t-h} Y_t = \ell_h \quad \text{p.s.}$$

Cela se traduit, à travers (III.29), (III.32) et (III.33), par la relation

$$\ell_h = \sum_{k=1}^p \beta_k \ell_{h-k} - \theta_p \rho \ell_{h-p-1} + \sigma^2 \delta_h \quad (\text{III.34})$$

où  $\ell_{-k} = \ell_k$  pour tout  $1 \leq k \leq p+1$ , et  $\delta_h$  dénote la fonction *delta* de Kronecker renvoyant la valeur 1 sur  $\{h = 0\}$  et la valeur 0 sur  $\{h \neq 0\}$ . Faisant varier  $h$  de 0 à  $p+1$ , on construit un système linéaire de  $p+2$  équations à  $p+2$  inconnues, dont il est facile d'observer que les coefficients engendrés forment la matrice  $B$  donnée par (III.7), et que le membre de droite est le vecteur  $\sigma^2 e$  de dimension  $p+2$ . Il s'ensuit que  $BL = \sigma^2 e$ , où  $L$  est le vecteur formé de  $\ell_0, \dots, \ell_{p+1}$ . L'inversibilité de  $B$  (Lemme III.1) et la définition de  $\Delta_p$  en (III.9) achèvent la preuve. ★

### III.5.2.1 Preuve du Théorème III.1

Le travail effectué en amont dans la preuve du Lemme III.4 nous sera ici très précieux. Reprenons en effet la formulation  $\text{AR}(p+1)$  du modèle, décrite en (III.29). Il vient alors, par sommation,

$$\sum_{t=p}^n \Phi_{t-1}^p Y_t = \sum_{t=p}^n \Phi_{t-1}^p \beta' \Phi_{t-1}^p - \theta_p \rho \sum_{t=p}^n \Phi_{t-1}^p Y_{t-p-1} + \sum_{t=p}^n \Phi_{t-1}^p V_t \quad (\text{III.35})$$

où par convention,  $Y_{-1} = 0$ . On voit que (III.35) est la somme de 3 termes que nous allons décortiquer. Tout d'abord, on a par construction

$$\sum_{t=p}^n \Phi_{t-1}^p \beta' \Phi_{t-1}^p = (S_{n-1} - S) \beta \quad (\text{III.36})$$

où  $S_{n-1}$  et  $S$  sont données par (III.3). Le second terme semble plus délicat à traiter, et nous allons le développer composante par composante en utilisant (III.29).

$$\sum_{t=p}^n \Phi_{t-1}^p Y_{t-p-1} = S_{n-1} J_p \beta - \theta_p \rho \sum_{t=p}^n \Phi_{t-1}^p Y_t + J_p \sum_{t=p}^n \Phi_{t-1}^p V_t + \xi_n \quad (\text{III.37})$$

où l'on montre que le terme résiduel vérifie, en vertu du Lemme III.3,  $\|\xi_n\| = o(n)$  p.s. dès que  $\mathbb{E}[V_1^2] = \sigma^2 < +\infty$ . Enfin, le troisième terme est une  $(\mathcal{F}_n)$ -martingale vectorielle de dimension  $p$ , suivant le même raisonnement que précédemment, dans la preuve du Lemme III.4. Nous la notons, pour  $n \geq p$ ,

$$M_n = \sum_{t=p}^n \Phi_{t-1}^p V_t \quad (\text{III.38})$$

avec  $M_0 = \dots = M_{p-1} = 0$ . Quant à son processus croissant, pour tout  $n \geq p$ , il est donné par

$$\langle M \rangle_n = \sum_{t=p}^n \mathbb{E}[(\Delta M_t)(\Delta M_t)' | \mathcal{F}_{t-1}] = \sigma^2(S_{n-1} - S). \quad (\text{III.39})$$

Reprenons (III.35). On a alors

$$\sum_{t=p}^n \Phi_{t-1}^p Y_t = \alpha S_{n-1} (I_p - \theta_p \rho J_p) \beta + \alpha (I_p - \theta_p \rho J_p) M_n + r_n$$

où  $\alpha$  et  $\beta$  sont donnés par (III.5), et  $\|r_n\| = o(n)$  p.s. Selon sa définition en (III.4), on fait alors aisément apparaître  $\widehat{\theta}_n$  dans l'équation. Ainsi, pour  $n \geq p$ ,

$$\widehat{\theta}_n = \alpha (I_p - \theta_p \rho J_p) \beta + \alpha S_{n-1}^{-1} (I_p - \theta_p \rho J_p) M_n + S_{n-1}^{-1} r_n. \quad (\text{III.40})$$

En vertu du Lemme III.4 et du caractère défini positif de  $\Delta_p$  (Lemme III.2),  $\lambda_{\min}(\langle M \rangle_n)$  et  $\lambda_{\max}(\langle M \rangle_n)$  convergent vers des limites strictement positives, respectivement  $\sigma^4 \lambda_{\min}(\Delta_p)$  et  $\sigma^4 \lambda_{\max}(\Delta_p)$ , à la vitesse  $n$ . Cela suffit alors à vérifier les conditions du Théorème B.2, impliquant de fait

$$\lim_{n \rightarrow \infty} S_{n-1}^{-1} M_n = 0 \quad \text{p.s.} \quad (\text{III.41})$$

Puisque  $\|S_{n-1}^{-1} r_n\| = o(1)$  p.s. par les Lemmes III.3–III.4, on a également

$$\lim_{n \rightarrow \infty} S_{n-1}^{-1} r_n = 0 \quad \text{p.s.} \quad (\text{III.42})$$

Il nous reste à combiner (III.40), (III.41) et (III.42) pour achever la preuve du théorème. ★

### III.5.2.2 Preuve du Théorème III.2

Reprenons la décomposition (III.40). Pour  $n \geq p$ , il vient

$$\sqrt{n}(\widehat{\theta}_n - \theta^*) = \alpha \sqrt{n} S_{n-1}^{-1} (I_p - \theta_p \rho J_p) M_n + \sqrt{n} S_{n-1}^{-1} r_n. \quad (\text{III.43})$$

Le terme résiduel est facile à traiter. En effet, comme l'on a supposé que  $\mathbb{E}[V_1^4] = \tau^4 < +\infty$ , on a par les Lemmes III.3–III.4 que  $\|S_{n-1}^{-1} r_n\| = o(n^{-1/2})$  p.s. Ainsi,

$$\lim_{n \rightarrow \infty} \sqrt{n} S_{n-1}^{-1} r_n = 0 \quad \text{p.s.} \quad (\text{III.44})$$

Vérifions la condition de Lindeberg pour la  $(\mathcal{F}_n)$ -martingale vectorielle  $(M_n)$ . Tout d'abord, par le Lemme III.3, il est clair que

$$\sum_{t=p}^n \|\Phi_{t-1}^p\|^4 = O(n) \quad \text{p.s.}$$

en reprenant la définition de  $\Phi_t^p$  donnée par (III.2), et l'inégalité de Cauchy-Schwarz. Pour tout  $\varepsilon > 0$ ,

$$\begin{aligned} \frac{1}{n} \sum_{t=p}^n \mathbb{E} \left[ \|\Delta M_t\|^2 \mathbb{I}_{\{\|\Delta M_t\| \geq \varepsilon \sqrt{n}\}} \mid \mathcal{F}_{t-1} \right] &\leq \frac{1}{\varepsilon^2 n^2} \sum_{t=p}^n \mathbb{E} \left[ \|\Delta M_t\|^4 \mid \mathcal{F}_{t-1} \right] \\ &\leq \frac{\tau^4}{\varepsilon^2 n^2} \sum_{t=p}^n \|\Phi_{t-1}^p\|^4 = O(n^{-1}) \quad \text{p.s.} \end{aligned} \quad (\text{III.45})$$

Par ailleurs, la définition de  $\langle M \rangle_n$  en (III.39) et le Lemme III.4 assurent que

$$\lim_{n \rightarrow \infty} \frac{\langle M \rangle_n}{n} = \sigma^4 \Delta_p \quad \text{p.s.} \quad (\text{III.46})$$

La combinaison de (III.45) et de (III.46) permet d'appliquer le Théorème B.4 avec  $a_n = n$  et  $L = \sigma^4 \Delta_p$ . On a alors la normalité asymptotique

$$\sqrt{n} \langle M \rangle_n^{-1} M_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^{-4} \Delta_p^{-1})$$

puisque l'on sait, grâce au Lemme III.2, que  $\Delta_p$  est inversible. Par le lemme de Slutsky,

$$\alpha \sqrt{n} S_{n-1}^{-1} (I_p - \theta_p \rho J_p) M_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \alpha^2 (I_p - \theta_p \rho J_p) \Delta_p^{-1} (I_p - \theta_p \rho J_p)) \quad (\text{III.47})$$

ce qui, combiné à (III.43) et à la convergence (III.44), achève la preuve du théorème. ★

### III.5.2.3 Preuve du Théorème III.3

Définissons la suite de matrices de normalisation  $W_n = \sqrt{n}I_p$ , pour  $n \geq p$ . Nous pouvons alors reformuler la convergence (III.46) sous la forme

$$\lim_{n \rightarrow \infty} W_n^{-1} \langle M \rangle_n W_n^{-1} = \sigma^4 \Delta_p \quad \text{p.s.} \quad (\text{III.48})$$

Posons également  $T_0 = 0$  et, pour  $n \geq 1$ ,

$$T_n = \sum_{t=1}^n Y_t^4.$$

Sachant par hypothèse que  $\mathbb{E}[V_1^4] = \tau^4 < +\infty$ , il découle du Lemme III.3 que  $T_n = O(n)$  p.s. Par une transformation d'Abel,

$$\sum_{n=1}^{\infty} \frac{Y_n^4}{n^2} = \sum_{n=1}^{\infty} \frac{T_n - T_{n-1}}{n^2} = \sum_{n=1}^{\infty} \frac{2n+1}{n^2(n+1)^2} T_n = O\left(\sum_{n=1}^{\infty} \frac{T_n}{n^3}\right) = O\left(\sum_{n=1}^{\infty} \frac{1}{n^2}\right) < +\infty \quad \text{p.s.}$$

On en déduit

$$\sum_{n=p}^{\infty} \frac{\|\Phi_{n-1}^p\|^4}{n^2} < +\infty \quad \text{p.s.} \quad (\text{III.49})$$

De (III.48) et (III.49), il suit que la  $(\mathcal{F}_n)$ -martingale vectorielle  $(M_n)$  donnée par (III.38) vérifie la loi forte quadratique stipulée par le Théorème 2.1 de [29],

$$\lim_{n \rightarrow \infty} \frac{1}{\log n^p} \sum_{t=p}^n \left[1 - \frac{t^p}{(t+1)^p}\right] W_t^{-1} M_t M_t' W_t^{-1} = \sigma^4 \Delta_p \quad \text{p.s.}$$

d'où l'on tire immédiatement

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{t=p}^n \frac{M_t M_t'}{t^2} = \sigma^4 \Delta_p \quad \text{p.s.} \quad (\text{III.50})$$

Reprenons désormais (III.40). Pour  $n \geq p$ ,

$$\begin{aligned} (\widehat{\theta}_n - \theta^*)(\widehat{\theta}_n - \theta^*)' &= \alpha^2 S_{n-1}^{-1} \left[ (I_p - \theta_p \rho J_p) M_n + \alpha^{-1} r_n \right] \left[ M_n' (I_p - \theta_p \rho J_p) + \alpha^{-1} r_n' \right] S_{n-1}^{-1} \\ &= \alpha^2 S_{n-1}^{-1} (I_p - \theta_p \rho J_p) M_n M_n' (I_p - \theta_p \rho J_p) S_{n-1}^{-1} + \zeta_n \end{aligned} \quad (\text{III.51})$$

avec  $\zeta_n = \alpha S_{n-1}^{-1} (r_n M_n' (I_p - \theta_p \rho J_p) + (I_p - \theta_p \rho J_p) M_n r_n' + \alpha^{-1} r_n r_n') S_{n-1}^{-1}$ . On sait par le Lemme III.4 que  $n S_{n-1}^{-1}$  converge presque sûrement vers  $\sigma^{-2} \Delta_p^{-1}$ . Par la loi forte quadratique (III.54), il vient

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{t=p}^n S_{t-1}^{-1} (I_p - \theta_p \rho J_p) M_t M_t' (I_p - \theta_p \rho J_p) S_{t-1}^{-1} &= \Delta_p^{-1} (I_p - \theta_p \rho J_p) \Delta_p (I_p - \theta_p \rho J_p) \Delta_p^{-1} \\ &= (I_p - \theta_p \rho J_p) \Delta_p^{-1} (I_p - \theta_p \rho J_p) \end{aligned} \quad (\text{III.52})$$

puisque, de par leur caractère bisymétrique,  $\Delta_p^{-1}$  et  $(I_p - \theta_p \rho J_p)$  commutent. En outre, on sait par le Lemme III.3 que  $\|r_n\| = o(\sqrt{n})$  p.s. et qu'en conséquence,  $\|r_n\| = o(\|M_n\|)$  p.s. Ainsi,

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{t=p}^n \zeta_t = 0 \quad \text{p.s.} \quad (\text{III.53})$$

La combinaison de (III.51), (III.52) et (III.53) nous conduit à

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{t=p}^n (\widehat{\theta}_t - \theta^*) (\widehat{\theta}_t - \theta^*)' = \alpha^2 (I_p - \theta_p \rho J_p) \Delta_p^{-1} (I_p - \theta_p \rho J_p) \quad \text{p.s.} \quad (\text{III.54})$$

ce qui achève la première partie de la preuve. La loi du logarithme itéré pour les martingales vectorielles, que l'on trouvera par exemple dans le Lemme C.2 de [7], nous garantit que, sous la condition (III.49), on a pour tout  $v \in \mathbb{R}^p$ ,

$$\limsup_{n \rightarrow \infty} \sqrt{\frac{n}{2 \log \log n}} v' S_{n-1}^{-1} M_n = -\liminf_{n \rightarrow \infty} \sqrt{\frac{n}{2 \log \log n}} v' S_{n-1}^{-1} M_n = \sqrt{v' \Delta_p^{-1} v} \quad \text{p.s.}$$

Le résidu  $S_{n-1}^{-1} r_n$  étant négligeable devant  $1/\sqrt{n}$ , il l'est d'autant plus devant  $\sqrt{\log \log n}/\sqrt{n}$ . Il vient alors, par (III.40),

$$\begin{aligned} \limsup_{n \rightarrow \infty} \sqrt{\frac{n}{2 \log \log n}} v' (\widehat{\theta}_n - \theta^*) &= \alpha \limsup_{n \rightarrow \infty} \sqrt{\frac{n}{2 \log \log n}} v' S_{n-1}^{-1} (I_p - \theta_p \rho J_p) M_n \\ &= \alpha \sqrt{v' (I_p - \theta_p \rho J_p) \Delta_p^{-1} (I_p - \theta_p \rho J_p) v} \quad \text{p.s.} \end{aligned} \quad (\text{III.55})$$

La formulation matricielle associée à la forme quadratique (voir [45]) permet donc d'achever la seconde partie de la preuve. ★

### III.5.2.4 Preuve du Théorème III.4

Rappelons tout d'abord que, pour  $0 \leq h \leq p + 1$ , on a les convergences presque sûres

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=h}^n Y_{t-h} Y_t = \sigma^2 \lambda_h \quad \text{p.s.}$$

comme nous l'avons vu dans le Lemme III.4 et la preuve associée, où l'ensemble  $(\lambda_h)$  est défini en (III.8). On note alors  $\Lambda_p^0, \Lambda_p^1$  et  $\Lambda_p^2$  les vecteurs limites  $\Lambda_p^d = (\lambda_d \ \lambda_{d+1} \ \dots \ \lambda_{d+p-1})'$  pour  $d = \{0, 1, 2\}$ . Il s'ensuit que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=p+d}^n \Phi_{t-d}^p Y_t = \sigma^2 \Lambda_p^d \quad \text{p.s.}$$

Pour  $n \geq p$ , on note également

$$A_n = \sum_{t=p}^n \Phi_t^p \Phi_{t-1}^{p'}.$$

Si nous raisonnons comme dans la preuve du Théorème III.1, on obtient

$$\sum_{t=p}^n \Phi_t^p Y_t = A_n \theta^* + \alpha \sum_{t=p}^n \Phi_t^p V_t - \alpha \theta_p \rho J_p \sum_{t=p+1}^n \Phi_{t-2}^p V_t + \eta_n$$

où la condition  $\mathbb{E}[V_1^2] = \sigma^2 < +\infty$  assure, à travers le Lemme III.3, que  $\|\eta_n\| = o(n)$  p.s. Il s'agit là encore de l'influence négligeable des termes de bord. L'expression limite devient alors

$$\Lambda_p^0 = A_p \theta^* + \alpha e \tag{III.56}$$

où  $A_p$  est la matrice limite presque sûre de la quantité  $\sigma^{-2} A_n/n$ , donnée par

$$A_p = \begin{pmatrix} \lambda_1 & \lambda_2 & \lambda_3 & \dots & \dots & \lambda_p \\ \lambda_0 & \lambda_1 & \lambda_2 & \dots & \dots & \lambda_{p-1} \\ \vdots & \vdots & \vdots & & & \vdots \\ \vdots & \vdots & \vdots & & & \vdots \\ \lambda_{p-2} & \lambda_{p-3} & \lambda_{p-4} & \dots & \dots & \lambda_1 \end{pmatrix}.$$

D'une manière parfaitement similaire, nous obtenons

$$\Lambda_p^2 = A_p' \theta^* - \alpha \theta_p \rho J_p e \quad \text{et} \quad \Lambda_p^1 = \Delta_p \theta^*. \tag{III.57}$$

Les relations (III.56) et (III.57) nous seront très utiles dans la suite. En vertu de ce qui précède, l'étude

limite des numérateur et dénominateur de  $\widehat{\rho}_n$  conduit à

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=p}^n \widehat{\varepsilon}_{t-1} \widehat{\varepsilon}_t &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=p+1}^n (Y_{t-1} - \widehat{\theta}_n' \Phi_{t-2}^p) (Y_t - \widehat{\theta}_n' \Phi_{t-1}^p) \\ &= \sigma^2 (\lambda_1 - (\Lambda_p^0 + \Lambda_p^2)' \theta^* + \theta^{*'} A_p \theta^*) = \sigma^2 (\lambda_1 - \Lambda_p^2' \theta^* - \alpha \theta_1^*) \quad \text{p.s.} \end{aligned}$$

d'une part, et à

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=p}^n \widehat{\varepsilon}_{t-1}^2 &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=p+1}^n (Y_{t-1} - \widehat{\theta}_n' \Phi_{t-2}^p)^2 \\ &= \sigma^2 (\lambda_0 - 2\Lambda_p^1' \theta^* + \theta^{*'} \Delta_p \theta^*) = \sigma^2 (\lambda_0 - \Lambda_p^1' \theta^*) \quad \text{p.s.} \end{aligned}$$

d'autre part (on rappelle que  $\widehat{\varepsilon}_{p-1} = 0$ ). Cependant, le système linéaire (III.34) nous permet d'établir que  $\lambda_0 = \beta' \Lambda_p^1 - \theta_p \rho \lambda_{p+1} + 1$  pour  $d = 0$ , et que  $\lambda_{p+1} = \beta' J_p \Lambda_p^1 - \theta_p \rho \lambda_0$  pour  $d = p + 1$ . De la définition de  $\theta^*$  en (III.6), il vient

$$\begin{aligned} \lambda_1 - \Lambda_p^2' \theta^* - \alpha \theta_1^* &= \lambda_1 - \Lambda_p^2' \theta^* - \alpha \theta_1^* (\lambda_0 - \beta' \Lambda_p^1 + \theta_p \rho \lambda_{p+1}) \\ &= \lambda_1 - \Lambda_p^2' \theta^* - \alpha \theta_1^* (\lambda_0 - \beta' \Lambda_p^1 + \theta_p \rho (\beta' J_p \Lambda_p^1 - \theta_p \rho \lambda_0)) \\ &= \lambda_1 - \Lambda_p^2' \theta^* - \theta_1^* (\lambda_0 - \Lambda_p^1' \theta^*) \\ &= \lambda_1 - \Lambda_p^2' \theta^* - (\theta_1 + \rho) (\lambda_0 - \Lambda_p^1' \theta^*) + \theta_p \rho \theta_p^* (\lambda_0 - \Lambda_p^1' \theta^*) \quad \text{(III.58)} \end{aligned}$$

puisque'il faut également remarquer que  $\theta_1^* = \theta_1 + \rho - \theta_p \rho \theta_p^*$ . De (III.57), on tire  $\lambda_1 = \Lambda_p^0' \theta^*$ . Ainsi,

$$\begin{aligned} \lambda_1 - \Lambda_p^2' \theta^* &= \theta^{*'} (\Lambda_p^0 - \Lambda_p^2) \\ &= \theta^{*'} A_p' \theta^* - \theta^{*'} A_p \theta^* + \alpha (\theta_1 + \rho) \\ &= \alpha (\theta_1 + \rho) (\lambda_0 - \beta' \Lambda_p^1 + \theta_p \rho \lambda_{p+1}) \\ &= (\theta_1 + \rho) (\lambda_0 - \Lambda_p^1' \theta^*). \quad \text{(III.59)} \end{aligned}$$

En combinant (III.58) et (III.59), on obtient

$$\lambda_1 - \Lambda_p^2' \theta^* - \alpha \theta_1^* = \theta_p \rho \theta_p^* (\lambda_0 - \Lambda_p^1' \theta^*),$$

ce qui permet d'identifier  $\rho^* = \theta_p \rho \theta_p^*$ , et donc d'achever la preuve. ★

### III.5.2.5 Preuve du Théorème III.5

De la décomposition (III.40), on a pour  $n \geq p$ ,

$$S_{n-1}(\widehat{\theta}_n - \theta^*) = \alpha(I_p - \theta_p \rho J_p)M_n + r_n \quad (\text{III.60})$$

avec, grâce au Lemme III.3,  $\|r_n\| = o(\sqrt{n})$  p.s. lorsque l'on suppose que  $\mathbb{E}[V_1^4] = \tau^4 < +\infty$ . L'objectif est maintenant de trouver une décomposition similaire concernant  $\widehat{\rho}_n - \rho^*$ . Nous allons introduire deux quantités  $B_n$  et  $F_n$  (comme *backward* et *forward*), censées fluidifier les calculs, et agissant sur toute valeur du processus  $(Y_t)$  de telle sorte que, pour  $n \geq 1$ ,

$$B_n = (1 - \rho^* L)Y_n = Y_n - \rho^* Y_{n-1} \quad \text{et} \quad F_n = (L - \rho^*)Y_n = Y_{n-1} - \rho^* Y_n$$

où  $\rho^*$  est la valeur limite de  $\widehat{\rho}_n$  donnée par (III.14). Considérons également leur extension vectorielle

$$B_n^p = (B_n \ B_{n-1} \ \dots \ B_{n-p+1})' \quad \text{et} \quad F_n^p = (F_n \ F_{n-1} \ \dots \ F_{n-p+1})'$$

dès que  $n \geq p$ . Notons enfin  $G_n$  une expression récurrente de  $\mathbb{R}^p$ , définie par

$$G_n = \Phi_n^p \theta^{*'} F_n^p - (L F_n^p + B_n^p) Y_n = \Phi_n^p \theta^{*'} F_n^p - (F_{n-1}^p + B_n^p) Y_n.$$

En développant (III.12), il vient

$$J_{n-1}(\widehat{\rho}_n - \rho^*) = W_n + (\widehat{\theta}_n - \theta^*)' H_n \quad (\text{III.61})$$

avec, pour  $n \geq p$ , les expressions scalaires

$$J_n = \sum_{t=p}^n \widehat{\varepsilon}_t^2, \quad W_n = \sum_{t=p}^n F_t Y_t + \theta^{*'} \sum_{t=p}^n G_t + \nu_n$$

et l'expression vectorielle

$$H_n = \sum_{t=p}^n (F_t^p \theta^{*'} \Phi_t^p + G_t) + \sum_{t=p}^n \Phi_t^p (\widehat{\theta}_n - \theta^*)' F_t^p + \mu_n.$$

On montre que les termes de bord  $\nu_n$  et  $\mu_n$  ci-dessus vérifient  $\nu_n = o(\sqrt{n})$  p.s. et  $\|\mu_n\| = o(\sqrt{n})$  p.s. dès que  $\mathbb{E}[V_1^4] = \tau^4 < +\infty$ . Nous souhaitons établir dans un premier temps que  $W_n$  se décompose comme la somme d'une martingale et d'un résidu. Soit la matrice  $C$  d'ordre  $p + 1$  extraite de  $B$  en (III.7), définie par

$$C = \begin{pmatrix} 1 - \beta_2 & -\beta_3 & \dots & \dots & -\beta_p & \theta_p \rho & 0 \\ -\beta_1 - \beta_3 & 1 - \beta_4 & \dots & \dots & \theta_p \rho & 0 & 0 \\ \vdots & \vdots & & & \vdots & \vdots & \vdots \\ \vdots & \vdots & & & \vdots & \vdots & \vdots \\ -\beta_{p-1} + \theta_p \rho & -\beta_{p-2} & \dots & \dots & -\beta_1 & 1 & 0 \\ -\beta_p & -\beta_{p-1} & \dots & \dots & -\beta_2 & -\beta_1 & 1 \end{pmatrix} \quad (\text{III.62})$$

où  $\beta$  est donné par (III.5). Rappelons que cette matrice est inversible sous les conditions  $\|\theta\|_1 < 1$  et  $|\rho| < 1$ , en vertu du Lemme III.1. Soit encore la  $(\mathcal{F}_n)$ -martingale donnée, pour  $n \geq p + 1$ , par

$$N_n = \sum_{t=p}^n \Phi_{t-1}^{p+1} V_t \quad (\text{III.63})$$

avec  $N_0 = \dots = N_p = 0$ ,  $\Phi_t^{p+1}$  étant l'extension naturelle de  $\Phi_t^p$  en (III.2) à la dimension supérieure, avec  $Y_{-1} = 0$  par convention. Terminons cette accumulation de notations, nécessaires à la fluidité des calculs, en introduisant le vecteur  $T = (\beta_1 \ \beta_2 \ \dots \ \beta_p \ -\theta_p \rho)'$ . Une manipulation de la formulation AR( $p + 1$ ) du modèle, donnée par (III.29), nous conduit à

$$\sum_{t=p}^n \Phi_{t-1}^{p+1} Y_t = C^{-1} T \sum_{t=p}^n Y_t^2 + C^{-1} N_n. \quad (\text{III.64})$$

Quelques calculs supplémentaires entraînent alors que, pour  $n \geq p$ ,

$$W_n = \sum_{t=p}^n F_t Y_t - \theta^{*'} \sum_{t=p}^n F_{t-1}^p Y_t - \alpha \theta^{*'} \sum_{t=p}^n (\Phi_t^p - \theta_p \rho J_p \Phi_{t-2}^p) V_t + \alpha \rho^* \theta^{*'} (I_p - \theta_p \rho J_p) M_n + \nu_n$$

où  $M_n$  est donnée par (III.38) et, comme on l'a vu,  $\nu_n = o(\sqrt{n})$  p.s. Combiné à (III.64), cela nous donne

$$W_n = (G' C^{-1} T - \rho^* - \alpha \theta_1^*) \sum_{t=p}^n Y_t^2 + G' C^{-1} N_n + L_n + \nu_n \quad (\text{III.65})$$

où, pour  $n \geq p$ ,

$$L_n = \alpha \theta^{*'} \left( \rho^* (I_p - \theta_p \rho J_p) M_n - \sum_{t=p}^n (\Phi_t^p - \theta_p \rho J_p \Phi_{t-2}^p) V_t \right) + \alpha \theta_1^* \sum_{t=p}^n Y_t V_t \quad (\text{III.66})$$

avec  $G = \rho^* \vartheta^* + \alpha \theta_1^* T - \delta^*$ ,  $\vartheta^* = (\theta_1^* \ \dots \ \theta_p^* \ 0)'$  et  $\delta^* = (-1 \ \theta_1^* \ \dots \ \theta_p^*)'$ . En termes de limites presque sûres, par la même méthodologie que celle utilisée dans la preuve du Lemme III.4, on tire de (III.64)

la relation

$$\lambda_0 C^{-1} T = \Lambda_{p+1}^1$$

où  $\Lambda_{p+1}^1 = (\lambda_1 \ \lambda_2 \ \dots \ \lambda_{p+1})'$ . On a alors le résultat clé suivant,

$$\begin{aligned} \lambda_0(G'C^{-1}T - \rho^* - \alpha\theta_1^*) &= G'\Lambda_{p+1}^1 - \lambda_0(\rho^* + \alpha\theta_1^*) \\ &= \rho^*(\Lambda_p^{1'}\theta^* - \lambda_0) + \alpha\theta_1^*(T'\Lambda_{p+1}^1 - \lambda_0) + (\lambda_1 - \Lambda_p^{2'}\theta^*) \\ &= \theta_1^*(\alpha\Lambda_p^{1'}(I_p - \theta_p\rho J_p)\beta - \alpha(1 - (\theta_p\rho)^2)\lambda_0) + \rho^*(\Lambda_p^{1'}\theta^* - \lambda_0) + (\lambda_1 - \Lambda_p^{2'}\theta^*) \\ &= \theta_1^*(\Lambda_p^{1'}\theta^* - \lambda_0) + \rho^*(\Lambda_p^{1'}\theta^* - \lambda_0) + (\lambda_1 - \Lambda_p^{2'}\theta^*) \\ &= -\alpha(\rho^* + \theta_1^*) + \alpha(\rho^* + \theta_1^*) = 0 \end{aligned}$$

ce qui, puisque  $\lambda_0 > 0$  (Lemme III.2), implique nécessairement que

$$W_n = G'C^{-1}N_n + L_n + \nu_n. \quad (\text{III.67})$$

Dès lors, notons que  $(G'C^{-1}N_n + L_n)$  est une  $(\mathcal{F}_n)$ -martingale, localement de carré intégrable, ce qui se montre assez simplement. Combinons tout ce que l'on vient d'établir. Pour  $n \geq p$ , il vient à partir de (III.60) et (III.61), que

$$J_{n-1}(\widehat{\rho}_n - \rho^*) = G'C^{-1}N_n + L_n + \alpha M_n'(I_p - \theta_p\rho J_p)S_{n-1}^{-1}H_n + \xi_n$$

où le terme résiduel  $\xi_n = r_n'S_{n-1}^{-1}H_n + \nu_n$  vérifie  $\xi_n = o(\sqrt{n})$  p.s. Les systèmes d'équations (III.8) et (III.34) permettent également de simplifier l'expression ci-dessus, dans la mesure où

$$G'C^{-1} = \alpha \begin{pmatrix} 1 + \beta_2 & \beta_3 - \beta_1 & \dots & \beta_p - \beta_{p-2} & -\beta_{p-1} - \theta_p\rho & -\alpha^{-1}\theta_p^* - \theta_p\rho\theta_1^* \end{pmatrix}$$

d'où l'on extrait  $U_p = (1 + \beta_2 \ \beta_3 - \beta_1 \ \dots \ \beta_p - \beta_{p-2} \ -\beta_{p-1} - \theta_p\rho)'$ . Ainsi, de (III.66) et (III.67), la décomposition martingale de  $W_n$  que l'on s'était fixée comme objectif apparaît clairement. En effet,

$$W_n = \alpha \left( U_p + (I_p - \theta_p\rho J_p)(\rho^*\theta^* - \tau^*) \right)' M_n - \theta_p^* \sum_{t=p}^n Y_{t-p-1} V_t + \nu_n \quad (\text{III.68})$$

avec, on le rappelle,  $Y_{-1} = 0$ , et  $\tau^* = (\theta_2^* \ \dots \ \theta_p^* \ 0)'$ . Il s'ensuit que

$$\sqrt{n} \begin{pmatrix} \widehat{\theta}_n - \theta^* \\ \widehat{\rho}_n - \rho^* \end{pmatrix} = \frac{1}{\sqrt{n}} P_n N_n + R_n \quad (\text{III.69})$$

avec

$$P_n = \begin{pmatrix} P_n^{(1,1)} & 0 \\ P_n^{(2,1)} & P_n^{(2,2)} \end{pmatrix}, \quad R_n = \sqrt{n} \begin{pmatrix} S_{n-1}^{-1} r_n \\ J_{n-1}^{-1} \xi_n \end{pmatrix}$$

et

$$\begin{aligned} P_n^{(1,1)} &= nS_{n-1}^{-1}\alpha(I_p - \theta_p \rho J_p), \\ P_n^{(2,1)} &= nJ_{n-1}^{-1}\left(\alpha(U_p + (I_p - \theta_p \rho J_p)(\rho^* \theta^* - \tau^*))' + \alpha H_n' S_{n-1}^{-1}(I_p - \theta_p \rho J_p)\right), \\ P_n^{(2,2)} &= -nJ_{n-1}^{-1}\theta_p^*. \end{aligned}$$

Cette stratégie vectorielle rappelle le raisonnement utilisé dans [131]. Par les Lemmes III.3–III.4, on a immédiatement  $\|R_n\| = o(1)$  p.s. La relation (III.69) constitue dès lors l'axe principal de la fin de la preuve. En effet, on a vu que  $(N_n)$  était une martingale (vérifiant la condition de Lindeberg, nous l'avons montré dans la preuve du Théorème III.2, en (III.45)) dont on connaît la vitesse de croissance du crochet. Si l'on prouve que la matrice  $P_n$  converge, alors la normalité asymptotique jointe de nos estimateurs découlera du Théorème B.4 et du lemme de Slutsky. Mais nous avons tous les outils en main pour cela, il s'agit simplement d'établir que

$$\lim_{n \rightarrow \infty} \frac{H_n}{n} = -\alpha(I_p - \theta_p \rho J_p)e \quad \text{p.s.} \quad \text{et} \quad \lim_{n \rightarrow \infty} \frac{J_n}{n} = \alpha \sigma^2 \quad \text{p.s.}$$

à partir de leur définition respective et des techniques largement développées dans la preuve du Lemme III.4. Tout cela est très calculatoire, mais finalement relativement facile à établir (avec un bon logiciel de calcul formel!). Il vient alors

$$\lim_{n \rightarrow \infty} P_n = \sigma^{-2}P \quad \text{p.s.} \quad (\text{III.70})$$

où la matrice  $P$  est donnée par (III.15). Par ailleurs, on montre aisément, de la même manière que pour (III.46), que l'on a la convergence

$$\lim_{n \rightarrow \infty} \frac{\langle N \rangle_n}{n} = \sigma^4 \Delta_{p+1} \quad \text{p.s.} \quad (\text{III.71})$$

On tire alors du Théorème B.4 la normalité asymptotique de la martingale,

$$\frac{N_n}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^4 \Delta_{p+1})$$

ce qui, combiné à (III.69), à (III.70), à (III.71) et au lemme de Slutsky, permet de conclure le raisonnement et d'achever la preuve. ★

### III.5.2.6 Preuve du Théorème III.6

Cette preuve est assez rapide. Reprenons pour commencer l'expression (III.69) et les notations associées, et extrayons-en la dernière composante. Pour  $n \geq p$ ,

$$\widehat{\rho}_n - \rho^* = \frac{1}{n} \pi'_n N_n + J_{n-1}^{-1} \xi_n \quad (\text{III.72})$$

où  $\pi_n = (P_n^{(2,1)} \quad P_n^{(2,2)})' \in \mathbb{R}^{p+1}$ . Nous avons également établi que

$$\lim_{n \rightarrow \infty} \pi_n = \pi \quad \text{p.s.}$$

dans (III.70), avec  $\pi = \sigma^{-2}(P'_L \quad \varphi)'$ , en reprenant les notations de (III.15). En outre,

$$\widehat{\rho}_n - \rho^* = \frac{1}{n} \pi'_n N_n + \frac{1}{n} (\pi_n - \pi)' N_n + J_{n-1}^{-1} \xi_n. \quad (\text{III.73})$$

La suite  $(N_n)$  étant une  $(\mathcal{F}_n)$ -martingale vectorielle d'ordre  $p+1$ , il est évident que la suite  $(\pi'_n N_n)$  est aussi une  $(\mathcal{F}_n)$ -martingale, scalaire cette fois. Son processus croissant est donné, pour  $n \geq p$ , par

$$\langle \pi'_n N \rangle_n = \sigma^2 \pi' (T_{n-1} - T) \pi$$

où  $T_n$  est une matrice carrée définie positive d'ordre  $p+1$ , donnée par

$$T_n = \sum_{t=p}^n \Phi_t^{p+1} \Phi_t^{p+1'} + T$$

et  $T$  est aussi une matrice définie positive ajoutée à  $T_n$  pour s'affranchir d'une hypothèse d'inversibilité. Grâce au Lemme III.3, et puisque  $\mathbb{E}[V_1^4] = \tau^4 < +\infty$ , on a

$$\lim_{n \rightarrow \infty} \frac{\pi'_n \Phi_n^{p+1} \Phi_n^{p+1'} \pi}{\pi'_n T_n \pi} = 0 \quad \text{p.s.}$$

qui symbolise une condition de non explosivité de la martingale, et permet d'y affecter sa loi forte quadratique (voir le Théorème 3 de [8] ou [9]). Celle-ci nous dit que

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{t=p}^n \left( \frac{\pi'_t N_t}{\pi'_t T_{t-1} \pi} \right)^2 = \frac{1}{\pi'_t \Delta_{p+1} \pi} \quad \text{p.s.} \quad (\text{III.74})$$

où  $\Delta_{p+1}$ , décrite en (III.9), est la limite presque sûre de  $\sigma^{-2} T_n/n$ , comme nous l'avons établi dans les preuves précédentes. Les matrices  $T_n$  et  $\Delta_{p+1}$  étant définies positives, la convergence précédente ne présente aucune singularité dès que  $\pi \neq 0$ , et donc sous l'hypothèse fondamentale de l'étude,  $\theta_p \neq 0$ .

Quant aux termes résiduels de (III.73), on a

$$\begin{aligned} \sum_{t=p}^n \left( \frac{(\pi_t - \pi)' N_t}{t} + \frac{\xi_t}{J_{t-1}} \right)^2 &= O \left( \sum_{t=p}^n \left( \frac{(\pi_t - \pi)' N_t}{t} \right)^2 + \sum_{t=p}^n \left( \frac{\xi_t}{J_{t-1}} \right)^2 \right) \\ &= o(\log n) + o \left( \sum_{t=p}^n \frac{(\pi' N_t)^2}{t^2} \right) = o(\log n) \quad \text{p.s.} \end{aligned}$$

par comparaison avec (III.74), puisque l'on sait que  $\xi_n = o(\|N_n\|)$  p.s. par la vitesse dans la loi forte des grands nombres pour les martingales, sachant de plus que  $\xi_n = o(\sqrt{n})$  p.s. On a ainsi

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{t=p}^n (\widehat{\rho}_t - \rho^*)^2 &= \lim_{n \rightarrow \infty} \sum_{t=p}^n \frac{(\pi' N_t)^2}{t^2} = \lim_{n \rightarrow \infty} \sum_{t=p}^n \left( \frac{\pi' N_t}{\pi' T_{t-1} \pi} \right)^2 \left( \frac{\pi' T_{t-1} \pi}{t} \right)^2 \\ &= \frac{\sigma^4 (\pi' \Delta_{p+1} \pi)^2}{\pi' \Delta_{p+1} \pi} = \sigma^4 \pi' \Delta_{p+1} \pi \quad \text{p.s.} \end{aligned}$$

La définition de  $\pi$  en début de preuve achève la première partie de la démonstration. Appliquons désormais la loi du logarithme itéré à la martingale  $(\pi' N_n)$ , que nous trouverons dans [126]–[127] ou encore dans le Corollaire 6.4.25 de [45]. Nous devons pour cela vérifier que

$$\sum_{n=p}^{\infty} \frac{(\pi' \Phi_{n-1}^p)^4}{n^2} < +\infty \quad \text{p.s.} \quad (\text{III.75})$$

mais ce résultat a déjà été établi, sous une version équivalente, dans (III.49). Ainsi,

$$\limsup_{n \rightarrow \infty} \sqrt{\frac{n}{2 \log \log n}} \frac{\pi' N_n}{\langle \pi' N \rangle_n} = -\liminf_{n \rightarrow \infty} \sqrt{\frac{n}{2 \log \log n}} \frac{\pi' N_n}{\langle \pi' N \rangle_n} = \sigma^{-2} (\pi' \Delta_{p+1} \pi)^{-1/2} \quad \text{p.s.} \quad (\text{III.76})$$

On rappelle que  $\xi_n$  est négligeable devant  $\sqrt{n}$ , et donc *a fortiori* devant  $J_{n-1} \sqrt{\log \log n} / \sqrt{n}$ , presque sûrement, puisque  $J_n$  se comporte comme  $n$ . Nous combinons alors (III.73) et (III.76) pour obtenir, *via* la convergence presque sûre de  $\langle \pi' N \rangle_n / n$  vers  $\sigma^4 \pi' \Delta_{p+1} \pi$ ,

$$\limsup_{n \rightarrow \infty} \sqrt{\frac{n}{2 \log \log n}} (\widehat{\rho}_n - \rho^*) = -\liminf_{n \rightarrow \infty} \sqrt{\frac{n}{2 \log \log n}} (\widehat{\rho}_n - \rho^*) = \sigma^2 \sqrt{\pi' \Delta_{p+1} \pi} \quad \text{p.s.}$$

et la preuve est achevée. ★

### III.5.3 Sur la procédure statistique

#### III.5.3.1 Preuve du Théorème III.8

Sous  $\mathcal{H}_0 : \rho = 0$ , reprenant les notations de la Section III.2.2.2, on a  $P_L = \theta_p J_p \theta$ ,  $\varphi = \theta_p$ ,  $\alpha = 1$  et  $\theta^* = \theta$ . Ainsi,

$$\begin{aligned} \sigma_\rho^2 &= P_L' \Delta_p P_L - 2\alpha^{-1} \theta_p^* \Lambda_p^{1'} J_p P_L + (\alpha^{-1} \theta_p^*)^2 \lambda_0 \\ &= \theta_p^2 (\theta' \Delta_p \theta - 2\Lambda_p^{1'} \theta + \lambda_0) \\ &= \theta_p^2 ((\theta' \Delta_p - \Lambda_p^{1'}) \theta + (\lambda_0 - \Lambda_p^{1'} \theta)) = \theta_p^2 \end{aligned}$$

puisque, comme établi en (III.57), on peut voir que  $\Delta_p^{-1} \theta = \Lambda_p^1$  et que  $\lambda_0 - \Lambda_p^{1'} \theta = 1$ . La consistance forte de  $\widehat{\theta}_n$  sous  $\mathcal{H}_0$  et le Théorème III.5 permettent donc d'établir la convergence en loi stipulée. Il faut ensuite noter que  $D^* = 2$  si et seulement si  $\theta_p = 0$ ,  $\rho = 0$ , ou  $\theta_p^* = 0$ . Il s'ensuit que, sous  $\mathcal{H}_1 : \rho \neq 0$  et sous les hypothèses retenues, la statistique de test diverge nécessairement. ★

#### III.5.3.2 Preuve du Théorème III.9

L'estimateur des moindres carrés de la variance de l'estimateur  $\widehat{\theta}_n$  s'écrit, pour  $n \geq p$ ,

$$\widehat{\mathbb{V}}(\widehat{\theta}_n) = \widehat{\sigma}_n^2 S_{n-1}^{-1} \quad (\text{III.77})$$

au sens conditionnel, où  $S_n$  est donnée par (III.3), et

$$\widehat{\sigma}_n^2 = \frac{1}{n} \sum_{t=p}^n \widehat{\varepsilon}_t^2 \quad (\text{III.78})$$

est l'estimateur usuel de la variance résiduelle, sous l'hypothèse nulle  $\mathcal{H}_0 : \rho = 0$ . On rappelle que l'ensemble  $(\widehat{\varepsilon}_t)$  est défini en (III.12) et que, arbitrairement,  $\widehat{\varepsilon}_0 = \dots = \widehat{\varepsilon}_{p-1} = 0$ . Nous utiliserons dans cette preuve une version Toeplitz de  $S_n$ , donnée par

$$S_n^p = \begin{pmatrix} s_n^0 & s_n^1 & s_n^2 & \dots & s_n^{p-1} \\ s_n^1 & s_n^0 & s_n^1 & \dots & s_n^{p-2} \\ s_n^2 & s_n^1 & s_n^0 & \dots & s_n^{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_n^{p-1} & s_n^{p-2} & s_n^{p-3} & \dots & s_n^0 \end{pmatrix}$$

où, pour tout  $0 \leq h \leq p$ ,

$$s_n^h = \sum_{t=h}^n Y_t Y_{t-h}$$

et l'on note aisément, grâce au Lemme III.3, que  $S_n^p = S_n + o(n)$  p.s. Supposons, par souci de simplicité des calculs, que  $S_n^p$  est inversible (cela nous exempte d'y ajouter une matrice définie positive  $S$ , conduisant bien sûr aux mêmes résultats mais compliquant drastiquement les développements). On définit ainsi

$$\Pi_n^h = \begin{pmatrix} s_n^1 & s_n^2 & \dots & s_n^h \end{pmatrix}' \quad \text{et} \quad \widehat{\mathfrak{D}}_n^{p-1} = \begin{pmatrix} \widehat{\mathfrak{D}}_{1,n} & \widehat{\mathfrak{D}}_{2,n} & \dots & \widehat{\mathfrak{D}}_{p-1,n} \end{pmatrix}'$$

avec  $\Pi_n = \Pi_n^p$ ,  $\pi_n = \Pi_n^{p-1}$  et  $\widehat{\mathfrak{D}}_n = (S_n^p)^{-1} \Pi_n$  l'estimateur de Yule-Walker, qui est un équivalent asymptotique de l'estimateur des moindres carrés de  $\theta$  donné par (III.4), quoique plus agréable à manier. Tout d'abord, un simple calcul sur (III.78) nous montre que

$$n \widehat{\sigma}_n^2 = s_n^0 - \Pi_n' \widehat{\mathfrak{D}}_n \quad (\text{III.79})$$

où  $\widehat{\sigma}_n^2$  est construit sur  $\widehat{\mathfrak{D}}_n$ . De plus, le premier élément diagonal de  $(S_n^p)^{-1}$  est l'inverse du complément de Schur de  $S_n^{p-1}$  dans  $S_n^p$ , donné par

$$s_n^0 - \pi_n' (S_n^{p-1})^{-1} \pi_n. \quad (\text{III.80})$$

La combinaison de (III.79) et de (III.80) entraîne que

$$1 - n \widehat{\mathfrak{V}}(\widehat{\mathfrak{D}}_{1,n}) = \frac{\alpha_n - \beta_n}{\alpha_n} \quad (\text{III.81})$$

avec, dès que  $n \geq p$ ,

$$\alpha_n = s_n^0 - \pi_n' (S_n^{p-1})^{-1} \pi_n \quad \text{et} \quad \beta_n = s_n^0 - \Pi_n' (S_n^p)^{-1} \Pi_n.$$

Il est également clair que, *via* quelques manipulations de l'expression associée,

$$\pi_n = k_n (I_{p-1} + \widehat{\mathfrak{D}}_{p,n} J_{p-1}) S_n^{p-1} \widehat{\mathfrak{D}}_n^{p-1} \quad \text{avec} \quad k_n = \frac{1}{1 - \widehat{\mathfrak{D}}_{p,n}^2}. \quad (\text{III.82})$$

En effet, on tire de la définition de  $\widehat{\mathfrak{D}}_n$  que  $\pi_n = S_n^{p-1} \widehat{\mathfrak{D}}_n^{p-1} + \widehat{\mathfrak{D}}_{p,n} J_{p-1} \pi_n$  et, par calcul direct, que

$$(I_{p-1} - \widehat{\mathfrak{D}}_{p,n} J_{p-1})^{-1} = \frac{1}{1 - \widehat{\mathfrak{D}}_{p,n}^2} (I_{p-1} + \widehat{\mathfrak{D}}_{p,n} J_{p-1}).$$

Puisque  $S_n^{p-1}$  est bisymétrique et commute avec  $J_{p-1}$ , on tire de (III.82) que

$$\alpha_n = s_n^0 - k_n \pi_n' \widehat{\mathfrak{D}}_n^{p-1} - k_n \widehat{\mathfrak{D}}_{p,n} \pi_n' J_{p-1} \widehat{\mathfrak{D}}_n^{p-1} \quad \text{et} \quad \pi_n' J_{p-1} \widehat{\mathfrak{D}}_n^{p-1} = s_n^p - \widehat{\mathfrak{D}}_{p,n} s_n^0.$$

On en déduit

$$\begin{aligned} k_n^{-1} \alpha_n &= k_n^{-1} (s_n^0 - k_n \pi_n' \widehat{\vartheta}_n^{p-1} - k_n \widehat{\vartheta}_{p,n} s_n^p + k_n \widehat{\vartheta}_{p,n}^2 s_n^0) \\ &= s_n^0 - \pi_n' \widehat{\vartheta}_n^{p-1} - \widehat{\vartheta}_{p,n} s_n^p = s_n^0 - \Pi_n' \widehat{\vartheta}_n = \beta_n. \end{aligned} \quad (\text{III.83})$$

À partir de (III.81) et de (III.83), on a l'égalité non asymptotique

$$1 - n\widehat{\mathbb{V}}(\widehat{\vartheta}_{1,n}) = \widehat{\vartheta}_{p,n}^2.$$

Mais rappelons-nous que, presque sûrement,  $(S_n^p)^{-1} = S_n^{-1} + o(n^{-1})$  et considérons de nouveau l'estimateur des moindres carrés  $\widehat{\theta}_n$ . On obtient rapidement que  $\widehat{\theta}_n = \widehat{\vartheta}_n + o(1)$  p.s. et donc que

$$1 - n\widehat{\mathbb{V}}(\widehat{\theta}_{1,n}) = \widehat{\theta}_{p,n}^2 + o(1) \quad \text{p.s.}$$

ce qui conclut la preuve, puisqu'il est désormais évident que,  $n$  grandissant à l'infini,

$$\frac{n\widehat{\rho}_n^2}{1 - n\widehat{\mathbb{V}}(\widehat{\theta}_{1,n})} \sim \frac{n}{4\widehat{\theta}_{p,n}^2} (\widehat{D}_n - 2)^2 \quad \text{p.s.}$$

★

## Quelques principes de déviations modérées

Le travail que nous allons désormais présenter a été publié dans [16], il est le fruit d'une collaboration avec Valère Bitseki Penda et Hacène Djellout dont l'aide fut précieuse.

### IV.1 Introduction et motivation

Cette étude se veut avant tout une extension naturelle du Chapitre II. Nous souhaitons considérer à nouveau le modèle AR(1)–AR(1), défini en (II.1), ainsi que les résultats de convergence presque sûre et de normalité asymptotique obtenus dans les Théorèmes II.1–II.2–II.4–II.5 sur les estimateurs des moindres carrés  $\widehat{\theta}_n$  et  $\widehat{\rho}_n$  donnés par (II.2) et par (II.7), respectivement. Il s'agira de les traiter dans un cadre *super-exponentiel* et d'établir des *principes de déviations modérées*, notions que nous définirons en temps voulu, afin de transférer ces propriétés à la statistique de Durbin-Watson  $\widehat{D}_n$  donnée par (II.13). Les déviations modérées permettent en effet de maintenir l'approximation gaussienne bien au-delà du théorème central limite. Soit  $\varphi$  un paramètre d'intérêt que l'on peut estimer convenablement par un estimateur  $\widehat{\varphi}_n$ . Là où le théorème central limite nous fournit l'approximation asymptotique

$$\mathbb{P}\left(\sqrt{n}(\widehat{\varphi}_n - \varphi) \leq q_\alpha\right) \approx \alpha$$

où  $0 \leq \alpha \leq 1$  et  $q_\alpha$  est le quantile d'ordre  $\alpha$  de la loi gaussienne considérée, les déviations modérées permettent de donner un équivalent, *via* une fonction de taux à déterminer, du logarithme de la probabilité

$$\mathbb{P}\left(\frac{\sqrt{n}}{b_n}(\widehat{\varphi}_n - \varphi) \in A\right) \tag{IV.1}$$

où  $A$  est un domaine de déviations et  $(b_n)$  une suite déterministe. Lorsque  $b_n = 1$ , nous retrouvons la vitesse du théorème central limite. *A contrario*, lorsque  $b_n = \sqrt{n}$ , nous parlerons d'un *principe de*

*grandes déviations*. Nous focaliserons notre intérêt sur le panel de vitesses telles que  $1 \ll b_n \ll \sqrt{n}$ , à l'origine des déviations modérées, et nous montrerons que la fonction de taux associée à (IV.1) est, comme attendu, celle donnée par l'approximation gaussienne et qu'elle ne dépend pas nécessairement de  $\varphi$ , ce qui la rend de fait explicite.

Bien entendu, puisqu'il est manifeste que les résultats obtenus sont plus puissants que ceux proposés dans le Chapitre II, nous avons un prix à payer en échange. Et ce dernier se retrouve dans les hypothèses retenues sur le bruit blanc  $(V_t)$  associé au processus  $(Y_t)$ , donné par (II.1). L'étude se divisera alors en deux parties : la première sera attachée à une hypothèse de gaussianité des résidus, où les calculs sont assez aisés, tandis que la seconde concernera une condition de type de Chen-Ledoux [32]–[77], moins restrictive mais plus technique, que nous expliciterons. Ces parties seront rédigées de manière très symétrique, nous y établirons en effet les mêmes résultats de déviations modérées pour  $\widehat{\theta}_n$ ,  $\widehat{\rho}_n$  et  $\widehat{D}_n$ , successivement. Nous encadrerons l'étude par des définitions nécessaires en préambule, et nous aborderons succinctement le cas AR( $p$ )–AR(1), où le plus dur reste encore à faire, en guise de conclusion. Nos démonstrations s'appuieront majoritairement sur les travaux de Dembo [36] en 1996, de Dembo et Zeitouni [37] en 1998 et de Worms [135]–[136] dans les années 2000 d'une part, sur ceux de Puhalskii [114] en 1997 et de Djellout [42] en 2002 d'autre part, où l'on retrouvera des principes de déviations modérées dans un cadre de martingales.

## IV.2 Définitions et notations

### IV.2.1 La convergence (super-)exponentielle

Soit une suite de variables aléatoires  $(Z_n)$  à valeurs dans  $\mathbb{R}^{d \times p}$  et une suite  $(b_n^2)$  déterministe et croissante vers l'infini. Nous dirons que  $(Z_n)$  converge  $(b_n^2)$ -exponentiellement vite en probabilité vers la variable aléatoire  $Z$  si, pour tout  $\delta > 0$ , on a

$$\limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}(\|Z_n - Z\| > \delta) < 0, \quad (\text{IV.2})$$

et nous noterons  $Z_n \xrightarrow[b_n^2]{\text{exp}} Z$ . Si l'on a de plus, pour tout  $\delta > 0$ ,

$$\limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}(\|Z_n - Z\| > \delta) = -\infty, \quad (\text{IV.3})$$

alors nous dirons que  $(Z_n)$  converge  $(b_n^2)$ -super-exponentiellement vite en probabilité vers la variable aléatoire  $Z$ , et nous noterons  $Z_n \xrightarrow[b_n^2]{s\text{-exp}} Z$ . Ces modes de convergence vérifient donc immédiatement

$$Z_n \xrightarrow[b_n^2]{s\text{-exp}} Z \implies Z_n \xrightarrow[b_n^2]{\text{exp}} Z \implies Z_n \xrightarrow{\mathcal{P}} Z.$$

Lorsque deux suites aléatoires  $(Z_n)$  et  $(Z'_n)$  sont  $(b_n^2)$ -exponentiellement équivalentes, au sens de la convergence super-exponentielle (voir la Définition 4.2.10 de [37]), on notera  $Z_n \xrightarrow[b_n^2]{s\text{-exp}} Z'_n$ .

## IV.2.2 Le principe de grandes déviations

Nous introduisons tout d'abord la notion de *principe de grandes déviations*, abrégée en « PGD » par la suite.

**Définition IV.1.** Une suite de variables aléatoires  $(Z_n)$ , à valeurs dans un espace topologique  $(S, \mathcal{S})$ , satisfait un PGD de vitesse  $(b_n^2)$  et de fonction de taux  $I : S \rightarrow \mathbb{R}^+$  si  $b_n^2 \rightarrow +\infty$  et si, pour tout  $A \in \mathcal{S}$ ,

$$-\inf_{x \in A^o} I(x) \leq \liminf_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}(Z_n \in A) \leq \limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}(Z_n \in A) \leq -\inf_{x \in \bar{A}} I(x)$$

où  $A^o$  et  $\bar{A}$  désignent l'intérieur et la fermeture de  $A$ . La fonction de taux  $I$  est semi-continue inférieurement, c'est-à-dire que, pour tout  $c \geq 0$ , l'ensemble de niveau  $\{x \in S \mid I(x) \leq c\}$  est fermé dans  $S$ .

Nous nous limitons bien entendu ici au strict nécessaire pour la suite de notre étude, mais l'obtention de PGD est souvent d'une grande complexité et à l'origine de nombreux travaux, voir par l'exemple l'excellent livre de Dembo et Zeitouni [37].

## IV.2.3 Le principe de déviations modérées

Nous nous intéressons désormais à la notion de *principe de déviations modérées*, que nous abrégons « PDM » par la suite. Soit une suite  $(b_n^2)$  déterministe, mais cette fois limitée aux échelles

$$\lim_{n \rightarrow \infty} b_n = +\infty \quad \text{et} \quad \lim_{n \rightarrow \infty} \frac{b_n}{\sqrt{n}} = 0. \tag{IV.4}$$

**Définition IV.2.** Une suite de variables aléatoires  $(Z_n)$ , à valeurs dans un espace topologique  $(S, \mathcal{S})$ , satisfait un PDM de vitesse  $(b_n^2)$  vérifiant les conditions (IV.4), et de fonction de taux  $I : S \rightarrow \mathbb{R}^+$  si la suite  $(\sqrt{n}Z_n/b_n)$  satisfait un PGD de vitesse  $(b_n^2)$  et de fonction de taux  $I$ .

Formellement, nos résultats de PDM sur les estimateurs  $\widehat{\theta}_n$ ,  $\widehat{\rho}_n$  et  $\widehat{D}_n$  seront formulés sous la forme de PGD sur ces mêmes estimateurs correctement renormalisés.

## IV.3 Sur les déviations modérées dans le cas gaussien

### IV.3.1 Les hypothèses retenues

Reprenons donc le modèle AR(1)–AR(1) donné par (II.1), que nous rappelons être la trajectoire observée sur  $\{0, \dots, n\}$  d'un processus  $(Y_t)$  indexé par  $\mathbb{Z}$ . Pour tout  $1 \leq t \leq n$ ,

$$\begin{cases} Y_t = \theta Y_{t-1} + \varepsilon_t \\ \varepsilon_t = \rho \varepsilon_{t-1} + V_t \end{cases}$$

où les paramètres inconnus vérifient les conditions  $|\theta| < 1$  et  $|\rho| < 1$  dites « de stabilité », et où  $(V_t)$  est un bruit blanc de variance  $\sigma^2 > 0$ . Nous avons montré dans le Chapitre II que  $(Y_t)$  était stationnaire, et que la trajectoire observée était en conséquence asymptotiquement stationnaire, de valeurs initiales  $\varepsilon_0$  et  $Y_0$  arbitraires, possédant les mêmes propriétés de moments que  $(V_t)$ .

Nous considérons dans cette section que  $(V_t)$  est un bruit blanc gaussien. Nous supposons, et nous comprendrons pourquoi lorsque nous démontrerons nos résultats, qu'il existe  $t > 0$  tel que

**Hypothèse G1.**

$$\mathbb{E}[\exp(t \varepsilon_0^2)] < +\infty,$$

**Hypothèse G2.**

$$\mathbb{E}[\exp(t Y_0^2)] < +\infty.$$

### IV.3.2 Déviations liées aux estimateurs

Les estimateurs des moindres carrés  $\widehat{\theta}_n$ ,  $\widehat{\rho}_n$  et  $\widehat{D}_n$  sont définis en (II.2), (II.7) et (II.13), il semblerait donc superflu de les redéfinir ici. Rappelons simplement qu'ils convergent presque sûrement vers  $\theta^*$ ,  $\rho^*$  et  $D^* = 2(1 - \rho^*)$ , respectivement (voir (II.3) et (II.8)). Nous avons alors les résultats suivants.

**Théorème IV.1.** *Supposons qu'il existe  $t > 0$  tel que les hypothèses G1 et G2 sont vérifiées, et que  $(V_t)$  suit la distribution  $\mathcal{N}(0, \sigma^2)$ . Alors, pour  $n \geq 1$ , la suite*

$$\left( \frac{\sqrt{n}}{b_n} (\widehat{\theta}_n - \theta^*) \right)$$

satisfait un PGD sur  $\mathbb{R}$  de vitesse  $(b_n^2)$  et de fonction de taux

$$I_\theta(x) = \frac{x^2}{2\sigma_\theta^2} \tag{IV.5}$$

où la variance asymptotique  $\sigma_\theta^2$  est donnée par (II.4).

**Théorème IV.2.** Supposons qu'il existe  $t > 0$  tel que les hypothèses G1 et G2 sont vérifiées, et que  $(V_t)$  suit la distribution  $\mathcal{N}(0, \sigma^2)$ . Alors, pour  $n \geq 1$  et dès que  $\theta \neq -\rho$ , la suite

$$\begin{pmatrix} \sqrt{n} (\widehat{\theta}_n - \theta^*) \\ \frac{\sqrt{n}}{b_n} (\widehat{\rho}_n - \rho^*) \end{pmatrix}$$

satisfait un PGD sur  $\mathbb{R}^2$  de vitesse  $(b_n^2)$  et de fonction de taux

$$K(x) = \frac{1}{2} x' \Gamma^{-1} x \quad (\text{IV.6})$$

où la covariance asymptotique  $\Gamma$  est donnée par (II.9). En particulier, la suite

$$\left( \frac{\sqrt{n}}{b_n} (\widehat{\rho}_n - \rho^*) \right)$$

satisfait un PGD sur  $\mathbb{R}$  de vitesse  $(b_n^2)$  et de fonction de taux

$$I_\rho(x) = \frac{x^2}{2\sigma_\rho^2} \quad (\text{IV.7})$$

où la variance asymptotique  $\sigma_\rho^2$  est donnée par (II.10).

Nous avons vu dans la Section II.2.2.2 que la matrice de covariance en question,  $\Gamma$ , n'est inversible que lorsque  $\theta \neq -\rho$ , d'où la restriction dans le théorème ci-dessus. Lorsque  $\theta = -\rho$ , si le PDM n'est pas valable pour le couple, en revanche il est valable pour chacune de ses composantes. En particulier, pour  $n \geq 1$ , les suites

$$\left( \frac{\sqrt{n}}{b_n} (\widehat{\theta}_n - \theta^*) \right) \quad \text{et} \quad \left( \frac{\sqrt{n}}{b_n} (\widehat{\rho}_n - \rho^*) \right)$$

satisfont des PGD sur  $\mathbb{R}$  de vitesse  $(b_n^2)$  et de fonctions de taux respectives

$$I_\theta(x) = \frac{x^2(1-\theta^2)}{2(1+\theta^2)} \quad \text{et} \quad I_\rho(x) = \frac{x^2(1-\theta^2)}{2\theta^4(1+\theta^2)}.$$

**Corollaire IV.1.** Supposons qu'il existe  $t > 0$  tel que les hypothèses G1 et G2 sont vérifiées, et que  $(V_t)$  suit la distribution  $\mathcal{N}(0, \sigma^2)$ . Alors, pour  $n \geq 1$ , la suite

$$\left( \frac{\sqrt{n}}{b_n} (\widehat{D}_n - D^*) \right)$$

satisfait un PGD sur  $\mathbb{R}$  de vitesse  $(b_n^2)$  et de fonction de taux

$$I_D(x) = \frac{x^2}{2\sigma_D^2} \quad (\text{IV.8})$$

où la variance asymptotique est définie par  $\sigma_D^2 = 4\sigma_p^2$ .

*Démonstration.* Les Théorèmes IV.1–IV.2 et le Corollaire IV.1 sont prouvés dans la Section IV.6.1. ★

## IV.4 Sur les déviations modérées dans le cas Chen-Ledoux

### IV.4.1 Les hypothèses retenues

Il s'agit désormais d'oublier l'hypothèse de gaussianité, et de supposer que  $(V_t)$  satisfait une condition moins restrictive, dite « de Chen-Ledoux ». Outre les conditions générales de stabilité rappelées dans la Section IV.3.1, nous allons considérer que, pour un certain  $a > 0$ ,

**Hypothèse CL1( $a$ ) de Chen-Ledoux.**

$$\limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log n \mathbb{P}(|V_1|^a > b_n \sqrt{n}) = -\infty,$$

**Hypothèse CL2( $a$ ).**

$$\frac{|\varepsilon_0|^a}{b_n \sqrt{n}} \xrightarrow[b_n^2]{s\text{-exp}} 0,$$

**Hypothèse CL3( $a$ ).**

$$\frac{|Y_0|^a}{b_n \sqrt{n}} \xrightarrow[b_n^2]{s\text{-exp}} 0.$$

Il est à noter que la condition CL1 fournit de multiples interprétations quant aux moments, exponentiels ou non, de  $V_1$ . Supposons à titre d'exemple que  $V_1$  satisfait CL1(2). Alors,

$$\limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log n \mathbb{P}(|V_1^2 - \mathbb{E}[V_1^2]| > b_n \sqrt{n}) = -\infty, \quad (\text{IV.9})$$

ce qui implique en particulier que  $\mathbb{V}(V_1^2) < +\infty$ . De plus, s'il existe  $t > 0$  tel que  $\mathbb{E}[\exp(t V_1^2)] < +\infty$ , alors CL1(2) est vérifiée pour toute suite croissante  $(b_n)$ , par exemple dans le cas gaussien. De [2]–[3]–[52], on déduit également que la condition (IV.9) est équivalente au fait que, pour  $n \geq 1$ , la suite

$$\left( \frac{1}{b_n \sqrt{n}} \sum_{t=1}^n (V_t^2 - \mathbb{E}[V_t^2]) \right)$$

satisfait un PGD sur  $\mathbb{R}$  de vitesse  $(b_n^2)$  et de fonction de taux

$$I(x) = \frac{x^2}{2\mathbb{V}(V_1^2)}.$$

C'est ce que nous avons résumé dans le Théorème B.6, en annexe. Par ailleurs, pour les vitesses

usuelles du type  $b_n = n^\alpha$  avec  $0 < \alpha < 1/2$ , CL1(2) est immédiatement vérifiée s'il existe  $t > 0$  et  $0 < \beta < 1$  tels que

$$\mathbb{E}\left[\exp\left(t V_1^{2\beta}\right)\right] < +\infty,$$

ce qui est clairement moins restrictif que le fait d'imposer un comportement gaussien à  $(V_t)$ , ainsi que nous l'avons fait dans la section précédente. Cependant, nous voyons également que les vitesses de l'ordre de grandeur du logarithme itéré ne sont plus adaptées. Notons enfin que la condition CL1(4), la plus forte que nous aurons à supposer au cours de la preuve de nos résultats, semble à certains égards trop forte, justement. Si nous souhaitions être optimal, il nous suffirait de supposer l'existence d'une constante  $C$ , arbitrairement grande, telle que

$$\limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{1}{n} \sum_{t=1}^n V_t^4 > C\right) = -\infty.$$

Il n'est d'ailleurs pas évident *a priori* qu'il y a une inclusion des hypothèses CL1 lorsque le paramètre grandit, et nous allons commencer par le stipuler sous forme de lemme technique.

**Lemme IV.1.** *Si CL1(a) est vérifiée pour  $a > 0$ , alors CL1(b) est également vérifiée pour tout  $0 < b < a$ .*

*Démonstration.* Le Lemme IV.1 est prouvé dans la Section IV.6.2. ★

## IV.4.2 Déviations liées aux estimateurs

Les résultats proposés dans cette section étant identiques à ceux de la section précédente – aux hypothèses de départ près, – nous ne rappellerons pas le détail explicite des PGD.

**Théorème IV.3.** *Sous les hypothèses CL1(4), CL2(4), CL3(4) et pour  $n \geq 1$ , la suite*

$$\left(\frac{\sqrt{n}}{b_n} (\widehat{\theta}_n - \theta^*)\right)$$

*satisfait le PGD stipulé dans le Théorème IV.1.*

**Théorème IV.4.** *Sous les hypothèses CL1(4), CL2(4), CL3(4),  $\theta \neq -\rho$  et pour  $n \geq 1$ , les suites*

$$\left(\frac{\sqrt{n}}{b_n} \begin{pmatrix} \widehat{\theta}_n - \theta^* \\ \widehat{\rho}_n - \rho^* \end{pmatrix}\right) \quad \text{et} \quad \left(\frac{\sqrt{n}}{b_n} (\widehat{\rho}_n - \rho^*)\right)$$

*satisfont les PGD stipulés dans le Théorème IV.2.*

Bien évidemment, lorsque  $\theta = -\rho$ , les suites

$$\left(\frac{\sqrt{n}}{b_n} (\widehat{\theta}_n - \theta^*)\right) \quad \text{et} \quad \left(\frac{\sqrt{n}}{b_n} (\widehat{\rho}_n - \rho^*)\right)$$

satisfont également les PGD décrits à la suite du Théorème IV.2.

**Corollaire IV.2.** *Sous les hypothèses CL1(4), CL2(4), CL3(4) et pour  $n \geq 1$ , la suite*

$$\left( \frac{\sqrt{n}}{b_n} (\widehat{D}_n - D^*) \right)$$

*satisfait le PGD stipulé dans le Corollaire IV.1.*

*Démonstration.* Les Théorèmes IV.3–IV.4 et le Corollaire IV.2 sont prouvés dans la Section IV.6.2. ★

## IV.5 Conclusion et perspectives

Nous avons établi des principes de déviations modérées pour nos estimateurs  $\widehat{\theta}_n$ ,  $\widehat{\rho}_n$  et  $\widehat{D}_n$ . Il est maintenant temps de leur trouver une application pratique, si l'on ne souhaite pas cantonner ce travail au domaine purement théorique. Il nous faudrait pour cela établir des principes de déviations modérées précis, pour lesquels les inégalités seraient vérifiées pour toute valeur de  $n$ , et non plus seulement asymptotiquement. Nous pensons qu'un tel travail serait très calculatoire, mais il serait dès lors intéressant de comparer la puissance empirique d'un test construit sur cette méthodologie avec la procédure étudiée dans le Chapitre II. Par la suite, l'évolution naturelle consisterait à généraliser les PDM au modèle AR( $p$ )–AR(1) du chapitre précédent. Dans le cas gaussien, cela nous semble relativement aisé, seulement quelque peu calculatoire. Sous la condition de Chen-Ledoux, des difficultés sont susceptibles d'apparaître lors de l'établissement des convergences super-exponentielles associées au traitement des martingales. Enfin, le point d'orgue de l'étude serait bien sûr l'établissement de principes de grandes déviations sur nos estimateurs, ce qui paraît bien compliqué, même dans le cas gaussien. Le problème reste à ce jour totalement ouvert. Nous nous proposons désormais d'aborder l'autocorrélation résiduelle en temps continu.

## IV.6 Annexe : démonstration des résultats

Pour simplifier le suivi des démonstrations, nous commençons par introduire quelques notations. Rappelons que la filtration engendrant le processus  $(Y_t)$  est donnée par  $\mathcal{F}_n = \sigma(Y_0, \varepsilon_0, V_1, \dots, V_n)$ , pour  $n \geq 1$ . On définit alors

$$L_n = \sum_{t=1}^n V_t^2, \quad S_n = \sum_{t=0}^n Y_t^2 \quad \text{et} \quad P_n = \sum_{t=1}^n Y_{t-1} Y_t. \quad (\text{IV.10})$$

Nous définissons ensuite les  $(\mathcal{F}_n)$ -martingales (voir la preuve du Lemme III.4 en Section III.5.2) données, respectivement pour  $n \geq 1$  et  $n \geq 2$ , par

$$M_n = \sum_{t=1}^n Y_{t-1} V_t \quad \text{et} \quad N_n = \sum_{t=2}^n Y_{t-2} V_t \quad (\text{IV.11})$$

avec  $M_0 = N_0 = N_1 = 0$ . On a alors immédiatement

$$\langle M \rangle_n = \sigma^2 S_{n-1} \quad \text{et} \quad \langle N \rangle_n = \sigma^2 S_{n-2}. \quad (\text{IV.12})$$

Du Lemme III.4 pour  $p = 1$ , on tire également que

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = \ell = \frac{\sigma^2(1 + \theta\rho)}{(1 - \theta^2)(1 - \theta\rho)(1 - \rho^2)} \quad \text{p.s.} \quad (\text{IV.13})$$

par résolution du système de Yule-Walker associé. Voici enfin deux lemmes techniques dont l'objectif est de simplifier les calculs que l'on sera amené à traiter.

**Lemme IV.2.** Pour tout  $\eta > 0$ , on a

$$\sum_{t=0}^n |Y_t|^\eta \leq (1 + \alpha(\eta)) |Y_0|^\eta + \alpha(\eta)\beta(\eta) |\varepsilon_0|^\eta + \alpha(\eta)\beta(\eta) \sum_{t=1}^n |V_t|^\eta$$

où  $\alpha(\eta) = (1 - |\theta|)^{-\eta}$  et  $\beta(\eta) = (1 - |\rho|)^{-\eta}$ . De plus,

$$\max_{1 \leq t \leq n} Y_t^2 \leq \alpha(1) Y_0^2 + \alpha(2)\beta(1) \varepsilon_0^2 + \alpha(2)\beta(2) \max_{1 \leq t \leq n} V_t^2.$$

*Démonstration.* La preuve est quasiment identique à celle du Lemme III.3 pour  $p = 1$ , en bornant de plus  $\|\psi\|_\infty$  par  $(1 - |\theta|)^{-1}$ . ★

**Lemme IV.3.** Pour  $n \geq 2$ , on a

$$\frac{S_n}{n} - \ell = \frac{\ell}{\sigma^2} \left[ \left( \frac{L_n}{n} - \sigma^2 \right) + 2\theta^* \frac{M_n}{n} - 2\theta\rho \frac{N_n}{n} + \frac{R_n}{n} \right] \quad (\text{IV.14})$$

où  $L_n, M_n, S_n$  et  $N_n$  sont donnés en (IV.10) et (IV.11),  $\ell$  est la valeur limite donnée en (IV.13),

$$R_n = (2(\theta + \rho)\rho^* - (\theta + \rho)^2 - (\theta\rho)^2) Y_n^2 - (\theta\rho)^2 Y_{n-1}^2 + 2\rho^* Y_n Y_{n-1} + \xi_1$$

et où le terme résiduel vaut précisément

$$\xi_1 = (1 - 2\theta\rho - \rho^2) Y_0^2 + \rho^2 \varepsilon_0^2 + 2\theta\rho Y_0 \varepsilon_0 - 2\rho\rho^* (\varepsilon_0 - Y_0) Y_0 + 2\rho (\varepsilon_0 - Y_0) V_1.$$

De plus, pour  $n \geq 1$ ,

$$\frac{P_n}{n} - \theta^* \frac{S_n}{n} = \frac{1}{1 + \theta\rho} \frac{M_n}{n} + \frac{1}{1 + \theta\rho} \frac{R_n(\theta)}{n} - \theta^* \frac{Y_n^2}{n} \quad (\text{IV.15})$$

avec

$$R_n(\theta) = \theta\rho Y_n Y_{n-1} + \rho Y_0 (\varepsilon_0 - Y_0).$$

*Démonstration.* Ces termes, à première vue compliqués, découlent simplement d'un calcul direct à partir des expressions (IV.10) et (IV.11), liées entre elles par le modèle (II.1). ★

Remarquons que nous pouvons ici nous permettre d'identifier les termes de bord dans la mesure où  $p = 1$ , alors que cela devenait rapidement mission impossible pour  $p > 1$  dans le Chapitre III.

## IV.6.1 Cas gaussien

Plaçons-nous désormais dans le cadre gaussien. Le premier théorème que nous sommes amenés à prouver dépend d'un certain nombre d'outils intermédiaires que nous avons choisi de stipuler en amont. Il s'agit de la convergence super-exponentielle des grandeurs définies en (IV.10) ainsi qu'un PDM liée aux martingales gaussiennes.

**Lemme IV.4.** *Sous les hypothèses du Théorème IV.1, on a la convergence super-exponentielle*

$$\frac{S_n}{n} \xrightarrow[b_n^2]{\text{s-exp}} \ell \quad (\text{IV.16})$$

où  $\ell$  est donnée par (IV.13).

*Démonstration.* Tout d'abord,  $(V_t)$  étant une suite indépendante, identiquement distribuée et de moyenne nulle, on a par le Théorème B.5 que, pour tout  $\delta > 0$ ,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \left| \frac{L_n}{n} - \sigma^2 \right| > \delta \right) < 0 \quad (\text{IV.17})$$

dans la mesure où  $L_n$  est une somme de khi-deux dont on connaît bien le comportement de la log-Laplace. Puisque  $b_n^2 = o(n)$ , on en tire

$$\frac{L_n}{n} \xrightarrow[b_n^2]{\text{s-exp}} \sigma^2 \quad (\text{IV.18})$$

par simple définition de la convergence super-exponentielle (Section IV.2.1). Il vient également de l'inégalité de Markov que, pour un  $t$  bien choisi (celui de l'hypothèse G2),

$$\mathbb{P}(Y_0^2 > n\delta) \leq \exp(-tn\delta) \mathbb{E}[\exp(tY_0^2)],$$

garantissant que

$$\frac{Y_0^2}{n} \xrightarrow[b_n^2]{s\text{-exp}} 0. \quad (\text{IV.19})$$

La même convergence est atteinte pour  $\varepsilon_0^2$  (par l'hypothèse G1), ainsi que pour  $V_1^2$  qui est un khi-deux (voir remarque ci-dessus). Reprenons désormais l'écriture vectorielle du modèle AR(1)–AR(1) donnée par (III.21) pour  $p = 1$ , et les notations associées. On voit que  $\rho(C_A) = |\theta| \vee |\rho| < 1$ , il suit donc de [134] que  $\|\Phi_n^2\|^2/n$  converge super-exponentiellement vite vers 0 à la vitesse  $(b_n^2)$ . Ainsi,

$$\frac{Y_n^2}{n} \xrightarrow[b_n^2]{s\text{-exp}} 0. \quad (\text{IV.20})$$

Il nous reste à étudier le comportement asymptotique exponentiel de  $M_n/n$  dont on ne sait *a priori* rien par les inégalités usuelles. Cependant, commençons par contrôler son crochet. On tire en effet du Lemme IV.2 et de (IV.12) que l'on peut trouver  $\alpha$  et  $\beta$  tels que, pour  $y > 0$  et  $t > 0$  bien choisis,

$$\begin{aligned} \mathbb{P}(\langle M \rangle_n > y) &\leq \mathbb{P}\left(Y_0^2 > \frac{y}{3\alpha\sigma^2}\right) + \mathbb{P}\left(\varepsilon_0^2 > \frac{y}{3\beta\sigma^2}\right) + \mathbb{P}\left(L_{n-1} > \frac{y}{3\beta\sigma^2}\right) \\ &\leq 3 \max\left(\exp\left(\frac{-yt}{3\alpha\sigma^2}\right) \mathbb{E}\left[\exp(tX_0^2)\right], \exp\left(\frac{-yt}{3\beta\sigma^2}\right) \mathbb{E}\left[\exp(t\varepsilon_0^2)\right], \mathbb{P}\left(L_{n-1} > \frac{y}{3\beta\sigma^2}\right)\right). \end{aligned}$$

Pour  $y = nx$  et  $x > 3\beta\sigma^4$ , on en déduit

$$\frac{1}{b_n^2} \log \mathbb{P}(\langle M \rangle_n > nx) \leq \frac{\log 3}{b_n^2} + \frac{1}{b_n^2} \max\left(\frac{-nxt}{3\alpha\sigma^2} + C_{X_0}, \frac{-nxt}{3\beta\sigma^2} + C_{\varepsilon_0}, \log \mathbb{P}\left(L_{n-1} > \frac{nx}{3\beta\sigma^2}\right)\right) \quad (\text{IV.21})$$

où  $C_{X_0} = \log \mathbb{E}[\exp(tX_0^2)] < +\infty$  et  $C_{\varepsilon_0} = \log \mathbb{E}[\exp(t\varepsilon_0^2)] < +\infty$ . Puisque  $x/3\beta\sigma^2 > \sigma^2$ , en combinant (IV.21) avec (IV.17), il vient

$$\limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}(\langle M \rangle_n > nx) = -\infty. \quad (\text{IV.22})$$

Par ailleurs, en décomposant l'évènement  $\{M_n > n\delta\}$ , on a pour tout  $\delta > 0$ ,

$$\begin{aligned} \mathbb{P}(M_n > n\delta) &= \mathbb{P}(M_n > n\delta, \langle M \rangle_n \leq nx) + \mathbb{P}(M_n > n\delta, \langle M \rangle_n > nx) \\ &\leq \exp\left(-\frac{n\delta^2}{2x}\right) + \mathbb{P}(\langle M \rangle_n > nx) \end{aligned} \quad (\text{IV.23})$$

par application du Théorème 4.1 de [12] dans le cas d'une martingale gaussienne. Il nous reste à appliquer (IV.22) au résultat ci-dessus pour établir que, puisque  $b_n^2 = o(n)$ ,

$$\limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}(M_n > n\delta) = -\infty. \quad (\text{IV.24})$$

La distribution de  $M_n$  étant symétrique, la convergence (IV.24) reste valable lorsque l'on remplace

$M_n$  par  $-M_n$ . En conclusion,

$$\frac{M_n}{n} \xrightarrow[b_n^2]{s\text{-exp}} 0. \quad (\text{IV.25})$$

Le raisonnement est rigoureusement identique dans le cas de la martingale  $(N_n)$  puisque son crochet est à un terme de bord près, que l'on sait traiter, le même que celui de  $(M_n)$ . Combinons dès lors les relations du Lemme IV.3 avec (IV.18), (IV.19), (IV.20) et (IV.25), et la preuve est achevée. ★

**Corollaire IV.3.** *Sous les hypothèses du Théorème IV.1, on a la convergence super-exponentielle*

$$\frac{P_n}{n} \xrightarrow[b_n^2]{s\text{-exp}} \theta^* \ell \quad (\text{IV.26})$$

où  $\theta^*$  est donné par (II.3) et  $\ell$  par (IV.13).

*Démonstration.* La preuve est immédiate par application des Lemmes IV.3–IV.4. ★

**Théorème IV.5.** *Soit  $(Y_n)$  une suite adaptée à valeurs dans  $\mathbb{R}^p$ , et  $(V_n)$  un bruit gaussien de variance  $\sigma^2 > 0$ . On suppose qu'il existe une matrice inversible  $C$  d'ordre  $p$  telle que  $(Y_n)$  satisfait, pour une vitesse  $b_n^2 = o(n)$  et tout  $\delta > 0$ , la convergence exponentielle*

$$\lim_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P} \left( \left\| \frac{1}{n} \sum_{t=0}^{n-1} Y_t Y_t' - C \right\| > \delta \right) = -\infty.$$

Alors, pour  $n \geq 1$ , la suite

$$\left( \frac{1}{b_n \sqrt{n}} \sum_{t=1}^n Y_{t-1} V_t \right)$$

satisfait un PGD sur  $\mathbb{R}^p$  de vitesse  $(b_n^2)$  et de fonction de taux

$$I(x) = \frac{1}{2\sigma^2} x' C^{-1} x.$$

*Démonstration.* Ce résultat, dû à Worms en 1999, est présent sous une version plus générale dans le Théorème 5 de [134]. Il suffit de se restreindre à  $d = 1$  dans la preuve associée. ★

#### IV.6.1.1 Preuve du Théorème IV.1

Considérons, pour  $n \geq 1$ , la décomposition

$$\frac{\sqrt{n}}{b_n} (\widehat{\theta}_n - \theta^*) = \frac{n}{\langle M \rangle_n} A_n + B_n \quad (\text{IV.27})$$

avec

$$A_n = \left( \frac{\sigma^2}{1 + \theta \rho} \right) \frac{M_n}{b_n \sqrt{n}} \quad \text{et} \quad B_n = \frac{\sqrt{n}}{b_n} \left( \frac{1}{1 + \theta \rho} \right) \frac{R_n(\theta)}{S_{n-1}},$$

obtenue par un simple développement, à partir des notations définies précédemment. Tout d'abord, en utilisant la même méthodologie que pour établir (IV.19), on obtient que, pour tout  $\delta > 0$  et un  $t > 0$  bien choisi (celui de l'hypothèse G2),

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P} \left( \frac{Y_0^2}{b_n \sqrt{n}} > \delta \right) &\leq \lim_{n \rightarrow \infty} \left( -t \delta \frac{\sqrt{n}}{b_n} \right) + \lim_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{E} \left[ \exp(t Y_0^2) \right] \\ &= -\infty, \end{aligned} \tag{IV.28}$$

puisque  $b_n = o(\sqrt{n})$ , et le même résultat suit pour tous les termes de bord de (IV.27) dont le numérateur ne dépend pas de  $n$ , tels que  $\varepsilon_0^2$  ou  $V_1^2$ . De plus, sous l'hypothèse gaussienne, on a

$$\begin{aligned} \mathbb{P} \left( \max_{1 \leq t \leq n} V_t^2 \geq \delta b_n \sqrt{n} \right) &= \mathbb{P} \left( \bigcup_{t=1}^n \{V_t^2 \geq \delta b_n \sqrt{n}\} \right) \leq \sum_{t=1}^n \mathbb{P} \left( V_t^2 \geq \delta b_n \sqrt{n} \right) \\ &\leq n \exp(-t \delta b_n \sqrt{n}) \mathbb{E} \left[ \exp(t V_1^2) \right] \end{aligned}$$

ce qui, dès que  $0 < t < 1/(2\sigma^2)$  et donc que la log-Laplace de  $V_1^2$  est finie, implique que

$$\frac{1}{b_n \sqrt{n}} \max_{1 \leq t \leq n} V_t^2 \xrightarrow[b_n^2]{s\text{-exp}} 0. \tag{IV.29}$$

On déduit alors de (IV.28), de (IV.29) et du Lemme IV.2, que

$$\frac{1}{b_n \sqrt{n}} \max_{1 \leq t \leq n} Y_t^2 \xrightarrow[b_n^2]{s\text{-exp}} 0. \tag{IV.30}$$

Le Lemme IV.4 ainsi que le Lemme 2 de [134] assurent que

$$\frac{n}{S_n} \xrightarrow[b_n^2]{s\text{-exp}} \frac{1}{\ell} \tag{IV.31}$$

ce qui est loin d'être trivial pour une convergence super-exponentielle, et le Lemme 2 de [134] nous accompagnera dans toute la suite de l'étude sans que nous ne le citions plus, lorsqu'il s'agira de multiplier entre elles ou d'inverser des expressions convergentes. Notons que l'on a bien  $\ell \neq 0$  sous les hypothèses  $|\theta| < 1$  et  $|\rho| < 1$ . Par (IV.30), (IV.31) et  $R_n(\theta)$  donné à la suite de (IV.15), on obtient

$$B_n \xrightarrow[b_n^2]{s\text{-exp}} 0 \tag{IV.32}$$

dans la décomposition (IV.27). Et l'on tire donc de (IV.31) qu'il en va de même pour

$$A_n \left( \frac{n}{\langle M \rangle_n} - \frac{1}{\sigma^2 \ell} \right) \xrightarrow[b_n^2]{s\text{-exp}} 0, \tag{IV.33}$$

puisque le Lemme IV.4 combiné au Théorème IV.5 pour  $p = 1$  nous montrent que la suite  $(M_n/(b_n\sqrt{n}))$  satisfait un PGD sur  $\mathbb{R}$  de vitesse  $(b_n^2)$  et de fonction de taux

$$J(x) = \frac{x^2}{2\ell\sigma^2}.$$

En conséquence,

$$\frac{\sqrt{n}}{b_n} (\widehat{\theta}_n - \theta^*) \underset{b_n^2}{\overset{\text{s-exp}}{\rightsquigarrow}} \frac{1}{\ell(1+\theta\rho)} \frac{M_n}{b_n\sqrt{n}} \quad (\text{IV.34})$$

et cela implique que chaque terme de l'équivalence satisfait le même PGD, en vertu du Théorème 4.2.13 de [37]. Le principe de contraction, établi dans le Théorème 4.2.1 du même ouvrage, nous permet alors d'identifier la fonction de taux associée au PGD de vitesse  $(b_n^2)$ . Ainsi,

$$I_\theta(x) = J(\ell(1+\theta\rho)x) = \frac{x^2}{2\sigma_\theta^2},$$

ce qui conclut la preuve, en reprenant les expressions de  $\ell$  en (IV.13) et de  $\sigma_\theta^2$  en (II.4). ★

Nous souhaitons, en guise de préambule à la preuve des théorèmes suivants, établir les convergences super-exponentielles de nouvelles quantités que nous allons définir, comme nous l'avons fait précédemment. Notons ainsi, pour  $n \geq 2$ ,

$$T_n = 1 + \theta^*\rho^* - \left(1 + \rho^*(\widehat{\theta}_n + \theta^*)\right) \frac{S_n}{S_{n-1}} + \left(2\rho^* + \widehat{\theta}_n + \theta^*\right) \frac{P_n}{S_{n-1}} - \frac{Q_n}{S_{n-1}} \quad (\text{IV.35})$$

avec les sommes  $S_n$  et  $P_n$  données en (IV.10), et

$$Q_n = \sum_{t=2}^n Y_{t-2} Y_t. \quad (\text{IV.36})$$

De plus, pour  $n \geq 1$ ,

$$J_n = \sum_{t=1}^n \widehat{\varepsilon}_t^2 \quad (\text{IV.37})$$

où l'ensemble  $(\widehat{\varepsilon}_t)$  est défini en (II.6).

**Corollaire IV.4.** *Sous les hypothèses du Théorème IV.1, on a la convergence super-exponentielle*

$$\frac{Q_n}{n} \underset{b_n^2}{\overset{\text{s-exp}}{\rightsquigarrow}} ((\theta + \rho)\theta^* - \theta\rho)\ell \quad (\text{IV.38})$$

où  $\theta^*$  est donné par (II.3) et  $\ell$  par (IV.13).

*Démonstration.* La preuve est assez immédiate lorsque l'on considère la décomposition

$$\frac{Q_n}{n} - ((\theta + \rho)\theta^* - \theta\rho)\frac{S_n}{n} = \theta^*\frac{M_n}{n} + \frac{N_n}{n} + \frac{\xi_n^Q}{n}$$

où l'on montre facilement, par des techniques déjà utilisées pour établir les convergences (IV.19) et (IV.20), que le résidu  $\xi_n^Q$ , composé de termes de bord (que l'on pourrait expliciter par calcul direct sans que cela nous paraisse indispensable ici), vérifie

$$\frac{\xi_n^Q}{n} \xrightarrow[b_n^2]{s\text{-exp}} 0.$$

La preuve est alors achevée en y associant la convergence (IV.25) et le Lemme IV.4. ★

**Corollaire IV.5.** *Sous les hypothèses du Théorème IV.1, on a la convergence super-exponentielle*

$$\frac{J_n}{n} \xrightarrow[b_n^2]{s\text{-exp}} (1 - \theta^*)(1 + \theta^*)\ell \quad (\text{IV.39})$$

où  $\theta^*$  est donné par (II.3) et  $\ell$  par (IV.13).

*Démonstration.* Là encore, il nous suffit de travailler l'expression de  $J_n$  en (IV.37) pour établir que, pour  $n \geq 1$ ,

$$J_n = S_n - Y_0^2 - 2\widehat{\theta}_n P_n + \widehat{\theta}_n^2 S_{n-1}.$$

On tire alors de l'écriture de  $\widehat{\theta}_n$  en (II.2) combinée au Lemme IV.4 et au Corollaire IV.3 que

$$\widehat{\theta}_n \xrightarrow[b_n^2]{s\text{-exp}} \theta^*. \quad (\text{IV.40})$$

Si l'on considère de plus la convergence (IV.19), la preuve est achevée. ★

**Corollaire IV.6.** *Sous les hypothèses du Théorème IV.1, on a la convergence super-exponentielle*

$$T_n \xrightarrow[b_n^2]{s\text{-exp}} (\theta^*)^2 + \theta\rho \quad (\text{IV.41})$$

où  $\theta^*$  est donné par (II.3).

*Démonstration.* Immédiate à partir de (IV.40), du Lemme IV.4 et des Corollaires IV.3–IV.4. ★

### IV.6.1.2 Preuve du Théorème IV.2

Selon le même principe que lors de la preuve du Théorème IV.1, nous commençons par considérer la décomposition, pour  $n \geq 2$ , donnée par

$$\frac{\sqrt{n}}{b_n} \begin{pmatrix} \widehat{\theta}_n - \theta^* \\ \widehat{\rho}_n - \rho^* \end{pmatrix} = \frac{1}{b_n \sqrt{n}} A_n Z_n + B_n \quad (\text{IV.42})$$

avec cette fois-ci

$$A_n = \frac{n}{1 + \theta \rho} \begin{pmatrix} \frac{1}{S_{n-1}} & 0 \\ \frac{T_n}{J_{n-1}} & -\frac{(\theta + \rho)}{J_{n-1}} \end{pmatrix} \quad \text{et} \quad B_n = \frac{1}{(1 + \theta \rho)} \frac{\sqrt{n}}{b_n} \begin{pmatrix} R_n(\theta) \\ \frac{S_{n-1}}{R_n(\rho)} \\ \frac{J_{n-1}}{J_{n-1}} \end{pmatrix}.$$

Le résidu  $R_n(\rho)$  est également composé de termes isolés, et nous pouvons d'ores et déjà, par l'intermédiaire de (IV.28) et de (IV.30), établir que

$$\frac{R_n(\theta)}{b_n \sqrt{n}} \xrightarrow[b_n^2]{s\text{-exp}} 0 \quad \text{et} \quad \frac{R_n(\rho)}{b_n \sqrt{n}} \xrightarrow[b_n^2]{s\text{-exp}} 0.$$

Si l'on y ajoute les résultats du Lemme IV.4 et du Corollaire IV.5, il vient facilement

$$B_n \xrightarrow[b_n^2]{s\text{-exp}} 0. \quad (\text{IV.43})$$

Le dernier terme non explicité dans (IV.42) reste la suite  $(Z_n)$ . Il s'agit de la  $(\mathcal{F}_n)$ -martingale vectorielle formée des suites  $(M_n)$  et  $(N_n)$  données en (IV.11), dont le crochet vérifie, par le Lemme IV.4 et le Corollaire IV.3,

$$\frac{\langle Z \rangle_n}{n} = \frac{\sigma^2}{n} \begin{pmatrix} S_{n-1} & P_{n-1} \\ P_{n-1} & S_{n-2} \end{pmatrix} \xrightarrow[b_n^2]{s\text{-exp}} \sigma^2 \ell \begin{pmatrix} 1 & \theta^* \\ \theta^* & 1 \end{pmatrix} = \sigma^4 \Delta_2$$

par analogie de notations avec la matrice  $\Delta_p$  en (III.9). Ainsi, le Théorème IV.5 pour  $p = 2$  nous montre que la suite  $(Z_n/(b_n \sqrt{n}))$  satisfait un PGD sur  $\mathbb{R}^2$  de vitesse  $(b_n^2)$  et de fonction de taux

$$J(x) = \frac{1}{2\sigma^4} x' \Delta_2^{-1} x.$$

On rappelle que  $\Delta_2$  est bien inversible sous les conditions  $|\theta| < 1$  et  $|\rho| < 1$ , nous l'avons en effet montré dans le Lemme III.2. Revenons à la matrice  $A_n$ . Du Lemme IV.4 et des Corollaires IV.5–IV.6, il vient, après quelques simplifications,

$$A_n \xrightarrow[b_n^2]{s\text{-exp}} A = \frac{1}{\ell(1 + \theta \rho)(1 - (\theta^*)^2)} \begin{pmatrix} 1 - (\theta^*)^2 & 0 \\ \theta \rho + (\theta^*)^2 & -(\theta + \rho) \end{pmatrix}$$

et ainsi

$$(A_n - A) \frac{Z_n}{b_n \sqrt{n}} \xrightarrow[b_n^2]{s\text{-exp}} 0 \quad (\text{IV.44})$$

en vertu du PGD satisfait par  $(Z_n/(b_n \sqrt{n}))$ . En conséquence, de la décomposition (IV.42), on tire

$$\frac{\sqrt{n}}{b_n} \begin{pmatrix} \widehat{\theta}_n - \theta^* \\ \widehat{\rho}_n - \rho^* \end{pmatrix} \xrightarrow[b_n^2]{s\text{-exp}} \frac{1}{b_n \sqrt{n}} A Z_n \quad (\text{IV.45})$$

et cela implique que chaque terme de l'équivalence satisfait le même PGD, par application du Théorème 4.2.13 de [37]. Le principe de contraction, établi dans le Théorème 4.2.1 du même ouvrage, nous permet alors d'identifier la fonction de taux associée au PGD de vitesse  $(b_n^2)$ . Ainsi,

$$K(x) = J(A^{-1}x) = \frac{1}{2} x' \Gamma^{-1} x$$

ce qui conclut la première partie de la preuve, en reprenant l'expression de  $\Gamma$  en (II.9). Notons quand même que  $A$  n'est pas inversible lorsque  $\theta = -\rho$ , par le calcul de son déterminant. C'est la raison pour laquelle cette hypothèse est exclue de l'étude jointe. Par ailleurs, on extrait du PGD vectoriel engendré par l'équivalence (IV.45), le PGD satisfait par sa seconde composante dont la fonction de taux est régie par le second élément diagonal de  $\Gamma$ . Il s'ensuit que

$$I_\rho(x) = \frac{x^2}{2\sigma_\rho^2},$$

ce qui termine la preuve, la variance  $\sigma_\rho^2$  étant définie en (II.10). ★

### IV.6.1.3 Preuve du Corollaire IV.1

Nous passerons très rapidement sur la preuve de ce corollaire, car les calculs ont déjà été effectués dans la Section II.6.2.1. En reprenant les notations associées et les raisonnements que nous venons de mener, on établit facilement que

$$\widehat{f}_n \xrightarrow[b_n^2]{s\text{-exp}} 0.$$

On en tire directement l'équivalence

$$\frac{\sqrt{n}}{b_n} (\widehat{D}_n - D^*) \xrightarrow[b_n^2]{s\text{-exp}} -2 \frac{\sqrt{n}}{b_n} (\widehat{\rho}_n - \rho^*) \quad (\text{IV.46})$$

et ainsi, de par le PGD satisfait par le membre de droite (voir la preuve précédente) et le principe de contraction (Théorème 4.2.1 de [37]), on en tire le PGD du membre de gauche sur  $\mathbb{R}$ , à vitesse  $(b_n^2)$  et dont la fonction de taux est donnée par

$$I_D(x) = I_\rho(-x/2) = \frac{x^2}{2\sigma_D^2}$$

où la variance asymptotique vaut  $\sigma_D^2 = 4\sigma_\rho^2$ . ★

## IV.6.2 Cas Chen-Ledoux

Plaçons-nous désormais dans le cadre Chen-Ledoux. Nous commençons par démontrer le lemme auquel nous avons eu recours lors de l'analyse des hypothèses retenues, puis nous aurons de nouveau besoin de certains résultats intermédiaires assez techniques. Il s'agira pour nous de montrer que toutes les convergences super-exponentielles établies dans le cas gaussien restent ici valables. Dès lors, la démonstration des théorèmes sera grandement simplifiée.

### IV.6.2.1 Preuve du Lemme IV.1

Soient  $a$  et  $b$  tels que  $0 < b < a$ . On a alors d'une part

$$\begin{aligned} \mathbb{P}(|V_1|^a > b_n \sqrt{n}) &= \mathbb{P}(|V_1|^a \mathbb{I}_{\{|V_1| < 1\}} + |V_1|^a \mathbb{I}_{\{|V_1| \geq 1\}} > b_n \sqrt{n}) \\ &\geq \mathbb{P}(|V_1|^a \mathbb{I}_{\{|V_1| \geq 1\}} > b_n \sqrt{n}) \\ &\geq \mathbb{P}(|V_1|^b \mathbb{I}_{\{|V_1| \geq 1\}} > b_n \sqrt{n}). \end{aligned} \tag{IV.47}$$

D'autre part,

$$\begin{aligned} \mathbb{P}(|V_1|^b > b_n \sqrt{n}) &= \mathbb{P}(|V_1|^b \mathbb{I}_{\{|V_1| < 1\}} + |V_1|^b \mathbb{I}_{\{|V_1| \geq 1\}} > b_n \sqrt{n}) \\ &\leq \mathbb{P}(|V_1|^b \mathbb{I}_{\{|V_1| \geq 1\}} > b_n \sqrt{n} - 1). \end{aligned} \tag{IV.48}$$

Ainsi, si l'on suppose que

$$\limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log n \mathbb{P}(|V_1|^a > b_n \sqrt{n}) = -\infty,$$

alors, par (IV.47), on a nécessairement

$$\limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log n \mathbb{P}(|V_1|^b \mathbb{I}_{\{|V_1| \geq 1\}} > b_n \sqrt{n}) = -\infty.$$

Cela implique encore, *via* (IV.48), que

$$\limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log n \mathbb{P}(|V_1|^b > b_n \sqrt{n}) = -\infty$$

et la preuve est achevée. ★

**Lemme IV.5.** *Sous les hypothèses CL1(4), CL2(4) et CL3(4), le Lemme IV.4 ainsi que les Corollaires IV.3–IV.4–IV.5–IV.6 restent valables.*

*Démonstration.* Reprenons les notations proposées en début de section. La suite  $(V_t)$  n'est plus gaussienne, mais elle reste indépendante et identiquement distribuée. À ce titre, on a toujours

$$\mathbb{P}\left(\max_{1 \leq t \leq n} V_t^2 \geq \delta b_n \sqrt{n}\right) \leq \sum_{t=1}^n \mathbb{P}(V_t^2 \geq \delta b_n \sqrt{n}) = n \mathbb{P}(V_1^2 \geq \delta b_n \sqrt{n}).$$

Il s'ensuit, *via* CL1(2), CL2(2), CL3(2) et le Lemme IV.2, que

$$\frac{Y_n^2}{b_n \sqrt{n}} \xrightarrow[b_n^2]{s\text{-exp}} 0 \quad \text{et (donc)} \quad \frac{Y_n^2}{n} \xrightarrow[b_n^2]{s\text{-exp}} 0 \quad (\text{IV.49})$$

et tous les termes de bord sont bien sûr traités de la même manière. Par ailleurs, n'ayant – contrairement à la section précédente – aucune information sur la log-Laplace de  $V_1$ , nous ne pouvons plus utiliser directement le Théorème B.5 pour établir un PGD sur la suite  $(L_n/n)$ . Cependant, grâce au Théorème B.6, nous avons un PDM sur  $(L_n/n)$  de vitesse  $(b_n^2)$ , dont on sait que la fonction de taux  $I$  ne s'annule pas, sauf en 0. On en déduit que, pour tout  $\delta > 0$ ,

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\left|\frac{L_n}{n} - \sigma^2\right| > \delta\right) &= \limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\left|\frac{1}{b_n \sqrt{n}} \sum_{t=1}^n (V_t^2 - \mathbb{E}[V_t]^2)\right| > \delta \frac{\sqrt{n}}{b_n}\right) \\ &= -\lim_{x \rightarrow +\infty} I(x) = -\infty. \end{aligned}$$

Ainsi,

$$\frac{L_n}{n} \xrightarrow[b_n^2]{s\text{-exp}} \sigma^2. \quad (\text{IV.50})$$

Il nous reste à étudier le comportement asymptotique exponentiel de  $M_n/n$ , maintenant que  $(M_n)$  n'est plus une martingale gaussienne. Par l'intermédiaire du Théorème 2.1 de [12], pour  $x, y > 0$  bien choisis, on a l'inégalité

$$\mathbb{P}\left(|M_n| > x, \langle M \rangle_n + [M]_n \leq y\right) \leq 2 \exp\left(-\frac{x^2}{2y}\right) \quad (\text{IV.51})$$

où, pour  $n \geq 1$ , le processus croissant de  $(M_n)$  vaut  $\langle M \rangle_n = \sigma^2 S_{n-1}$  et sa variation totale

$$[M]_n = \sum_{t=1}^n Y_{t-1}^2 V_t^2. \quad (\text{IV.52})$$

En appliquant (IV.51) à la décomposition de l'évènement  $\{|M_n| > n\delta\}$  avec  $x = n\delta$  et  $y = nd$  pour  $\delta, d > 0$ , il vient

$$\begin{aligned} \mathbb{P}(|M_n| > n\delta) &\leq \mathbb{P}(|M_n| > n\delta, \langle M \rangle_n + [M]_n \leq nd) + \mathbb{P}(\langle M \rangle_n + [M]_n > nd) \\ &\leq 2 \exp\left(-\frac{n\delta^2}{2d}\right) + \mathbb{P}(\langle M \rangle_n + [M]_n > nd). \end{aligned}$$

Nous obtenons en conséquence

$$\limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}(|M_n| > n\delta) \leq \limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}(\langle M \rangle_n + [M]_n > nd) \quad (\text{IV.53})$$

puisque  $b_n^2 = o(n)$ . Il nous reste à montrer que le terme de droite est contrôlable pour une valeur arbitraire de  $d$ . Définissons alors, pour  $n \geq 1$ ,

$$T_n = \sum_{t=0}^n Y_t^4 \quad \text{et} \quad \Gamma_n = \sum_{t=1}^n V_t^4.$$

Selon le Lemme IV.2 et pour une valeur de  $n$  suffisamment grande, on peut trouver  $\gamma > 0$  tel que

$$T_n < \gamma \Gamma_n$$

sous les conditions CL2(4) et CL3(4). Il suit alors du Théorème B.6, sous CL1(4), que l'on a la convergence super-exponentielle

$$\frac{\Gamma_n}{n} \xrightarrow[b_n^2]{\text{s-exp}} \tau^4 \quad (\text{IV.54})$$

de la même manière que celle permettant d'établir (IV.50), avec  $\tau^4 = \mathbb{E}[V_1^4]$ . Ainsi, par l'inégalité de Cauchy-Schwarz, on trouve que

$$\limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{[M]_n}{n} > \delta'\right) \leq \limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{\Gamma_n}{n} > \frac{\delta'}{\sqrt{\gamma}}\right) = -\infty$$

dès que  $\delta' > \tau^4 \sqrt{\gamma}$ , afin de satisfaire (IV.54). On montre de la même manière que

$$\limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{\langle M \rangle_n}{n} > \delta''\right) \leq \limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{L_n}{n} > \frac{\delta''}{\sigma^2 \gamma}\right) = -\infty$$

dès que  $\delta'' > \sigma^4 \gamma$ , afin de satisfaire (IV.50). On en déduit qu'il est possible d'établir que

$$\limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}(\langle M \rangle_n + [M]_n > nd) = -\infty \quad (\text{IV.55})$$

en choisissant par exemple  $d > \tau^4 \sqrt{\gamma} + \sigma^4 \gamma$ . Si l'on reprend (IV.53), on a donc bien

$$\frac{M_n}{n} \xrightarrow[b_n^2]{\text{s-exp}} 0. \quad (\text{IV.56})$$

Bien évidemment, il en va de même pour  $N_n/n$ . Par (IV.49), (IV.50) et (IV.56) et les relations du Lemme IV.3, nous obtenons

$$\frac{S_n}{n} \xrightarrow[b_n^2]{\text{s-exp}} \ell, \quad (\text{IV.57})$$

ce qui conclut la preuve, puisqu'il serait désormais largement redondant d'établir les convergences relatives aux Corollaires IV.3–IV.4–IV.5–IV.6. Elles découlent en effet toutes de (IV.57) et des outils que nous venons d'établir. On pourra consulter les preuves associées dans le cas gaussien pour s'en convaincre. ★

**Théorème IV.6.** Soit  $(m_t^n)$  un tableau triangulaire de différences de martingale, pour  $1 \leq t \leq n$ , à valeurs dans  $\mathbb{R}^p$ , adapté à la filtration  $\mathcal{F}_n$ . Soit  $(b_n)$  une suite réelle, positive, croissante et telle que  $b_n^2 = o(n)$ . Supposons qu'il existe une matrice symétrique semi-définie positive  $Q$  telle que

$$\frac{1}{n} \sum_{t=1}^n \mathbb{E}[m_t^n m_t^{n'} | \mathcal{F}_{t-1}] \xrightarrow[b_n^2]{\text{s-exp}} Q. \quad (\text{IV.58})$$

Supposons encore qu'il existe une constante  $c > 0$  telle que, pour tout  $1 \leq t \leq n$ ,

$$|m_t^n| \leq c \frac{\sqrt{n}}{b_n}. \quad (\text{IV.59})$$

Supposons enfin que, pour tout  $r > 0$ , la condition exponentielle de Lindeberg

$$\frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[ |m_t^n|^2 \mathbb{I}_{\{|m_t^n| \geq r \frac{\sqrt{n}}{b_n}\}} | \mathcal{F}_{t-1} \right] \xrightarrow[b_n^2]{\text{s-exp}} 0 \quad (\text{IV.60})$$

est vérifiée. Alors, pour  $n \geq 1$ , la suite

$$\left( \frac{1}{b_n \sqrt{n}} \sum_{t=1}^n m_t^n \right)$$

satisfait un PGD sur  $\mathbb{R}^d$  de vitesse  $(b_n^2)$  et de fonction de taux

$$\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}^p} \left( \lambda' x - \frac{1}{2} \lambda' Q \lambda \right).$$

En particulier, si  $Q$  est inversible,

$$\Lambda^*(x) = \frac{1}{2} x' Q^{-1} x. \quad (\text{IV.61})$$

*Démonstration.* Ce résultat, dû à Puhalskii en 1997, est présent sous une version plus générale dans le Théorème 3.1 de [114]. ★

**Lemme IV.6.** *Sous les hypothèses  $CL1(a)$ ,  $CL2(a)$  et  $CL3(a)$  dès que  $a > 2$ , nous avons, pour tout  $\delta > 0$ ,*

$$\lim_{R \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P} \left( \frac{1}{n} \sum_{t=0}^n Y_t^2 \mathbb{I}_{\{|Y_t| > R\}} > \delta \right) = -\infty.$$

*Démonstration.* Du Lemme IV.2, on voit que pour tout  $\eta > 0$  et une valeur de  $n$  suffisamment grande, il existe  $\gamma > 0$  tel que

$$\sum_{t=0}^n |Y_t|^{2+\eta} \leq \gamma \sum_{t=1}^n |V_t|^{2+\eta} \quad (\text{IV.62})$$

sous les hypothèses  $CL2(2 + \eta)$  et  $CL3(2 + \eta)$ . De plus, pour  $R > 0$ ,

$$R^\eta \sum_{t=0}^n Y_t^2 \mathbb{I}_{\{|Y_t| > R\}} \leq \sum_{t=0}^n |Y_t|^{2+\eta} \leq \gamma \sum_{t=1}^n |V_t|^{2+\eta}$$

ce qui nous conduit, pour tout  $\delta > 0$ , à

$$\frac{1}{b_n^2} \log \mathbb{P} \left( \frac{1}{n} \sum_{t=0}^n Y_t^2 \mathbb{I}_{\{|Y_t| > R\}} > \delta \right) \leq \frac{1}{b_n^2} \log \mathbb{P} \left( \frac{1}{n} \sum_{t=1}^n |V_t|^{2+\eta} > \frac{\delta}{\gamma} R^\eta \right).$$

Si l'on suppose que l'hypothèse  $CL1(2 + \eta)$  est également vérifiée, alors le résultat désiré est atteint par le Théorème B.6, en faisant de plus tendre  $R$  vers l'infini. Remarquons que ce lemme reste valable sous la condition moins restrictive  $CL1(2)$ , par l'intermédiaire d'une preuve très technique utilisant la mesure empirique associée à la chaîne de Markov géométriquement ergodique  $(Y_n)$ , pour  $n \geq 0$ . On trouvera un tel raisonnement par exemple dans [43]. Puisque nous retenons  $CL1(4)$  dans nos théorèmes, la condition  $CL1(2 + \eta)$  nous est suffisante (Lemme IV.1). ★

**Lemme IV.7.** *Sous les hypothèses  $CL1(4)$ ,  $CL2(4)$  et  $CL3(4)$ , pour  $n \geq 1$ , la suite*

$$\left( \frac{M_n}{b_n \sqrt{n}} \right)$$

satisfait un PGD sur  $\mathbb{R}$  de vitesse  $(b_n^2)$  et de fonction de taux

$$J(x) = \frac{x^2}{2\ell\sigma^2} \quad (\text{IV.63})$$

où  $\ell$  est donnée par (IV.13).

*Démonstration.* Le PDM sur la martingale non gaussienne  $(M_n)$  résultera de l'application du Théorème IV.6. Nous devons pour cela adopter une stratégie que nous conserverons jusqu'au bout de l'étude. Celle-ci consiste à tronquer les quantités considérées, à établir un PDM sur les quantités tronquées, et à montrer que les parties résiduelles sont exponentiellement négligeables. Pour  $r, R > 0$ , considérons

$$M_n^{(r,R)} = \sum_{t=1}^n Y_{t-1}^{(r)} V_t^{(R)} \quad (\text{IV.64})$$

où, pour tout  $1 \leq t \leq n$ ,

$$Y_t^{(r)} = Y_t \mathbb{I}_{\{|Y_t| \leq r \frac{\sqrt{n}}{b_n}\}} \quad \text{et} \quad V_t^{(R)} = V_t \mathbb{I}_{\{|V_t| \leq R\}} - \mathbb{E} \left[ V_t \mathbb{I}_{\{|V_t| \leq R\}} \right]. \quad (\text{IV.65})$$

Nous voyons donc que l'idée retenue est de tronquer  $(Y_t)$  à l'aide d'un paramètre  $r > 0$  muni d'une vitesse explosive, et de tronquer  $(V_t)$  à l'aide d'un paramètre  $R > 0$  que nous enverrons finalement à l'infini, tout en faisant en sorte de conserver un bruit centré. On sait, grâce au Lemme IV.5, que

$$\frac{\langle M \rangle_n}{n} \xrightarrow[b_n^2]{\text{s-exp}} \sigma^2 \ell.$$

De plus, par le Lemme IV.6, on sait que, pour tout  $r > 0$ ,

$$\frac{1}{n} \sum_{t=0}^n Y_t^2 \mathbb{I}_{\{|Y_t| \geq r \frac{\sqrt{n}}{b_n}\}} \xrightarrow[b_n^2]{\text{s-exp}} 0. \quad (\text{IV.66})$$

Notons encore, pour  $n \geq 1$ ,

$$\sigma_R^2 = \mathbb{E} \left[ V_1^{(R)2} \right] \quad \text{et} \quad S_n^{(r)} = \sum_{t=0}^n Y_t^{(r)2}.$$

Ainsi,  $(M_n^{(r,R)})$  conservant ses propriétés de  $(\mathcal{F}_n)$ -martingale, on a, pour  $r, R > 0$ ,

$$\frac{\langle M^{(r,R)} \rangle_n}{n} = \sigma_R^2 \frac{S_{n-1}^{(r)}}{n} = \sigma_R^2 \frac{S_{n-1}}{n} - \sigma_R^2 \left( \frac{S_{n-1}}{n} - \frac{S_{n-1}^{(r)}}{n} \right) \xrightarrow[b_n^2]{\text{s-exp}} \sigma_R^2 \ell$$

via (IV.66), garantissant de fait la validité de la première hypothèse (IV.58) du Théorème IV.6. Par ailleurs, il est immédiat de voir que le Lemme IV.6 s'applique tout aussi bien dans le cas du processus tronqué  $(Y_t^{(r)})$ , ce qui garantit que la troisième hypothèse (IV.60) du Théorème IV.6, la condition exponentielle de Lindeberg, est également satisfaite. La seconde hypothèse (IV.59), quant à elle, découle de la troncature (IV.65). Ainsi, par le Théorème IV.6, nous en déduisons que la suite

$$\left( \frac{M_n^{(r,R)}}{b_n \sqrt{n}} \right)$$

satisfait un PGD sur  $\mathbb{R}$  de vitesse  $(b_n^2)$  et de fonction de taux

$$J_R(x) = \frac{x^2}{2\sigma_R^2 \ell}. \quad (\text{IV.67})$$

Il nous reste à montrer que la différence  $M_n - M_n^{(r,R)}$  est négligeable au sens des grandes déviations, c'est-à-dire qu'une fois renormalisée par  $b_n \sqrt{n}$ , elle tend super-exponentiellement vite vers 0. Posons pour cela

$$M_n - M_n^{(r,R)} = L_n^{(r)} + F_n^{(r,R)}$$

avec, pour tout  $n \geq 1$ ,

$$L_n^{(r)} = \sum_{t=1}^n (Y_{t-1} - Y_{t-1}^{(r)}) V_t \quad \text{et} \quad F_n^{(r,R)} = \sum_{t=1}^n (V_t - V_t^{(R)}) Y_{t-1}^{(r)}.$$

Mais reprenons (IV.62). Pour une valeur suffisamment grande de  $n$ , on trouve

$$\begin{aligned} \frac{|L_n^{(r)}|}{b_n \sqrt{n}} &= \frac{1}{b_n \sqrt{n}} \left| \sum_{t=1}^n Y_{t-1} \mathbb{I}_{\{|Y_{t-1}| > r \frac{\sqrt{n}}{b_n}\}} V_t \right| \leq \frac{1}{b_n \sqrt{n}} \left( r \frac{\sqrt{n}}{b_n} \right)^{-\eta} \left( \sum_{t=1}^n |Y_{t-1}|^{2+\eta} \right)^{1/2} \left( \sum_{t=1}^n V_t^2 |Y_{t-1}|^\eta \right)^{1/2} \\ &\leq \lambda(r, \eta, \gamma) \left( \frac{b_n}{\sqrt{n}} \right)^{\eta-1} \frac{1}{n} \sum_{t=1}^n |V_t|^{2+\eta} \end{aligned} \quad (\text{IV.68})$$

par l'inégalité de Hölder, où  $\lambda(r, \eta, \gamma) > 0$  est une constante identifiable. Ainsi, pour tout  $\delta > 0$ ,

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P} \left( \frac{|L_n^{(r)}|}{b_n \sqrt{n}} > \delta \right) &\leq \limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P} \left( \frac{1}{n} \sum_{t=1}^n |V_t|^{2+\eta} > \frac{\delta}{\lambda(r, \eta, \gamma)} \left( \frac{\sqrt{n}}{b_n} \right)^{\eta-1} \right) \\ &= -\infty \end{aligned} \quad (\text{IV.69})$$

dès que  $\eta > 1$  par application du Théorème IV.6 sous  $\text{CL}1(2 + \eta)$ , puisqu'alors le membre de droite dans (IV.69) explose. On en tire que, sous  $\text{CL}1(2 + \eta)$ ,  $\text{CL}2(2 + \eta)$  et  $\text{CL}3(2 + \eta)$  avec  $\eta > 1$ ,

$$\frac{L_n^{(r)}}{b_n \sqrt{n}} \xrightarrow[b_n^2]{s\text{-exp}} 0. \quad (\text{IV.70})$$

Notons que cette hypothèse  $\text{CL}1(a)$  avec  $a > 3$  reste encore bien adaptée aux hypothèses de l'étude, en particulier à  $\text{CL}1(4)$ , par le Lemme IV.1. Remarquons désormais que  $(F_n^{(r,R)})$  est une  $(\mathcal{F}_n)$ -martingale dont le processus croissant est donné, pour  $n \geq 1$ , par

$$\langle F^{(r,R)} \rangle_n = Q_R S_{n-1}^{(r)} \quad \text{avec} \quad Q_R = \mathbb{E} \left[ (V_1 - V_1^{(R)})^2 \right].$$

Nous allons utiliser une méthodologie proche de celle du Théorème 1 de [42]. Pour  $R$  suffisamment grand et tout  $1 \leq t \leq n$ , on a

$$\mathbb{P}\left(\left|Y_{t-1}^{(r)}(V_t - V_t^{(R)})\right| > b_n \sqrt{n} \mid \mathcal{F}_{t-1}\right) \leq \mathbb{P}\left(\left|V_t - V_t^{(R)}\right| > \frac{b_n^2}{r}\right) = \mathbb{P}\left(\left|V_1 - V_1^{(R)}\right| > \frac{b_n^2}{r}\right) = 0.$$

Ainsi,

$$\limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \left( n \operatorname{ess\,sup}_{1 \leq t \leq n} \mathbb{P}\left(\left|Y_{t-1}^{(r)}(V_t - V_t^{(R)})\right| > b_n \sqrt{n} \mid \mathcal{F}_{t-1}\right) \right) = -\infty. \quad (\text{IV.71})$$

Par ailleurs, pour tout  $\nu > 0$  et  $\delta > 0$ , on tire du Lemme IV.6 que

$$\limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{1}{n} \sum_{t=1}^n Y_{t-1}^{(r)2} \mathbb{I}_{\left\{|Y_{t-1}^{(r)}| > \nu \frac{\sqrt{n}}{b_n}\right\}} > \delta\right) \leq \limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{1}{n} \sum_{t=1}^n Y_{t-1}^2 \mathbb{I}_{\left\{|Y_{t-1}| > \nu \frac{\sqrt{n}}{b_n}\right\}} > \delta\right) = -\infty.$$

Ainsi, s'en remettant une nouvelle fois au Lemme IV.5, il vient

$$\frac{\langle F^{(r,R)} \rangle_n}{n} = Q_R \frac{S_{n-1}^{(r)}}{n} = Q_R \frac{S_{n-1}}{n} - Q_R \left( \frac{S_{n-1}}{n} - \frac{S_{n-1}^{(r)}}{n} \right) \xrightarrow[b_n^2]{s\text{-exp}} Q_R \ell$$

où l'on rappelle que  $\ell$  est donnée par (IV.13). En outre, lorsque  $R$  tend vers l'infini, il est clair que  $Q_R$  tend en contrepartie vers 0. Il suit du Théorème 1 de [42] que, pour  $n \geq 1$ , la suite

$$\left( \frac{F_n^{(r,R)}}{b_n \sqrt{n}} \right)$$

satisfait un PGD sur  $\mathbb{R}$  de vitesse  $(b_n^2)$  et de fonction de taux

$$I_R(x) = \frac{x^2}{2Q_R \ell}.$$

Cela implique en particulier que, pour tout  $\delta > 0$ ,

$$\limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{|F_n^{(r,R)}|}{b_n \sqrt{n}} > \delta\right) = -\frac{\delta^2}{2Q_R \ell}$$

et donc que

$$\lim_{R \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{|F_n^{(r,R)}|}{b_n \sqrt{n}} > \delta\right) = -\infty. \quad (\text{IV.72})$$

La combinaison de (IV.70) et de (IV.72) entraîne que, pour tout  $r > 0$  et tout  $\delta > 0$ ,

$$\lim_{R \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{|M_n - M_n^{(r,R)}|}{b_n \sqrt{n}} > \delta\right) = -\infty,$$

ce qui nous montre que le résidu de la troncature est exponentiellement négligeable. Ainsi, par le Théorème 4.2.16 de [37], nous transférons le PDM associé à la martingale  $(M_n^{(r,R)})$  à la martingale  $(M_n)$ , la vitesse restant  $(b_n^2)$  et la fonction de taux étant donnée par

$$\tilde{J}(x) = \sup_{\delta > 0} \liminf_{R \rightarrow \infty} \inf_{z \in B_{x,\delta}} J_R(z)$$

où  $B_{x,\delta}$  désigne la boule  $\{z : |z - x| < \delta\}$ , et la fonction  $J_R$  est donnée en (IV.67). L'identification entre  $\tilde{J}$  et la fonction de taux  $J$  définie en (IV.63) conclut alors la preuve. ★

**Lemme IV.8.** *Sous les hypothèses CL1(4), CL2(4) et CL3(4), pour  $n \geq 2$ , la suite*

$$\left( \frac{1}{b_n \sqrt{n}} \begin{pmatrix} M_n \\ N_n \end{pmatrix} \right)$$

satisfait un PGD sur  $\mathbb{R}^2$  de vitesse  $(b_n^2)$  et de fonction de taux

$$J(x) = \frac{1}{2\sigma^4} x' \Delta_2^{-1} x \tag{IV.73}$$

où  $\Delta_2$  est donnée par (III.9), pour  $p = 2$ .

*Démonstration.* Nous ne développerons pas cette preuve, car elle repose sur la même stratégie que la preuve précédente. Il s'agit de tronquer la martingale vectorielle  $(M_n \ N_n)'$  et tous les calculs s'ensuivent de la même manière. Le résultat découle finalement du Théorème IV.6 avec  $p = 2$ . ★

#### IV.6.2.2 Preuve des Théorèmes IV.3–IV.4 et du Corollaire IV.2

Tout le travail a été fait en amont. Il nous suffit dès lors de reprendre les décompositions (IV.27) et (IV.42), d'y associer les convergences super-exponentielles à travers le Lemme IV.5 ainsi que les PDM par l'intermédiaire des Lemmes IV.7–IV.8. Les raisonnements sont alors identiques à ceux des preuves des théorèmes relatifs au cas gaussien. ★

---

## Une extension en temps continu

---

Ce travail est le fruit d'une collaboration avec Bernard Bercu et Nicolas Savy. Il est résumé dans [11], actuellement soumis et en attente de publication. Nous y abordons la problématique de la corrélation résiduelle dans un processus à temps continu.

### V.1 Historique et motivation

On considère le *processus d'Ornstein-Uhlenbeck*

$$dY_t = \theta Y_t dt + dW_t \tag{V.1}$$

où le paramètre  $\theta$  est inconnu et  $(W_t)$  est un mouvement brownien standard, encore appelé *processus de Wiener* (voir la Définition I.13). Soit maintenant la fonction  $f(Y_t, t) = Y_t e^{-\theta t}$  que nous allons différencier par le lemme d'Itô. Il vient alors

$$df(Y_t, t) = -\theta Y_t e^{-\theta t} dt + e^{-\theta t} dY_t = e^{-\theta t} dW_t.$$

Ainsi, en intégrant de 0 à  $T$ , on obtient la relation

$$Y_T = Y_0 e^{\theta T} + e^{\theta T} \int_0^T e^{-\theta t} dW_t \tag{V.2}$$

qui est une formulation simplifiée du processus d'Ornstein-Uhlenbeck. On a de plus immédiatement  $\mathbb{E}[Y_T] = e^{\theta T} \mathbb{E}[Y_0]$ . En outre, pour tous  $t, s \in \mathbb{R}^+$ , on a

$$\text{Cov}(Y_t, Y_s) = e^{\theta(t+s)} \mathbb{E} \left[ \int_0^t e^{-\theta u} dW_u \int_0^s e^{-\theta v} dW_v \right] = \frac{e^{\theta(t+s)}}{2\theta} (1 - e^{-2\theta t \wedge s}),$$

ce qui nous donne pour variance

$$\mathbb{V}(Y_T) = \frac{1}{2\theta} (e^{2\theta T} - 1).$$

Il en résulte que, si  $\theta < 0$ , le processus d'Ornstein-Uhlenbeck (V.2) est gaussien, à variance bornée et il admet une loi asymptotique stationnaire caractérisée par son espérance nulle et sa variance  $-1/(2\theta)$ . Lorsque  $\theta = 0$ , il s'agit d'un mouvement brownien de nature instable que l'on sera également amené à considérer par la suite. Enfin, lorsque  $\theta > 0$ , le processus (V.2) est explosif et nous sortons du cadre de l'étude. Depuis les travaux fondateurs [129] des probabilistes éponymes en 1930, une littérature considérable s'est développée sur le processus d'Ornstein-Uhlenbeck associé à un mouvement brownien ou un mouvement brownien fractionnaire (citons en exemple [72] ou [84]). Par la suite, les travaux se sont en particulier concentrés sur l'étude du processus d'Ornstein-Uhlenbeck dirigé par un processus de Lévy,

$$dY_t = \theta Y_t dt + dL_t \tag{V.3}$$

où  $(L_t)$  est donc un processus stochastique en temps continu à accroissements stationnaires et indépendants, mais non nécessairement brownien (voir [5] pour le socle mathématique). Par extension, on trouvera dans [4] l'étude du processus d'Ornstein-Uhlenbeck dirigé par un processus de Lévy fractionnaire. Il est toujours possible de complexifier le processus générateur des données, et l'on citera également l'étude dans [6] du processus d'Ornstein-Uhlenbeck dont la volatilité est elle-même engendrée par un processus d'Ornstein-Uhlenbeck, car la philosophie d'une telle modélisation se rapproche de la structure qui nous intéresse dans ce chapitre. Quant à la problématique de l'estimation dans le cas d'une perturbation de Lévy, elle est traitée de manière paramétrique dans [64] et de manière non paramétrique dans [69]. Nous pouvons aussi trouver une application en finance dans [100]. En bref, comme nous pouvons aisément le comprendre, une vaste littérature s'est développée autour du processus fondateur d'Ornstein-Uhlenbeck, axée sur une variété hétérogène de perturbations continues.

Nous nous proposons dès lors d'étudier le processus d'Ornstein-Uhlenbeck, lui-même dirigé par un processus d'Ornstein-Uhlenbeck, observés sur un intervalle de temps  $[0, T]$ ,

$$\begin{cases} dY_t = \theta Y_t dt + dV_t \\ dV_t = \rho V_t dt + dW_t \end{cases} \tag{V.4}$$

où les paramètres vérifient les conditions de stationnarité asymptotique évoquées précédemment, à savoir  $\theta < 0$  et  $\rho \leq 0$ . La perturbation  $(W_t)$  est quant à elle brownienne. De manière arbitraire et pour simplifier les calculs, nous choisirons  $Y_0 = 0$  et  $V_0 = 0$ . Venons-en aux motivations principales qui nous ont amenés à étudier ce modèle. D'une part, les modèles à volatilité stochastique trouvent nombre d'applications en finance [6]–[119]. D'autre part, alors que les recherches actuelles tendent à complexifier de plus en plus la dynamique des modèles et en particulier de leur volatilité, nous souhaitons translater tous les phénomènes de corrélation dans le processus résiduel afin de simplifier l'étude et de conserver l'homoscédasticité. Enfin, d'un point de vue plus personnel, le modèle d'Ornstein-Uhlenbeck couplé (V.4) n'est pas sans rappeler le modèle AR(1)–AR(1) étudié au Chapitre II. De par les propriétés asymptotiques que nous allons établir sur nos estimateurs, nous verrons qu'il peut même s'agir d'un prolongement naturel en temps continu, sous certains aspects, exactement comme le font Brockwell et Lindner en 2012 dans [27] sur les processus ARMA. À cet égard, nous y ferons référence à travers l'expression OU(1)–OU(1), par similitude, et nous axerons le contenu de l'étude sur le parallèle avec le Chapitre II.

## V.2 Une approche par maximum de vraisemblance

Nous allons dans cette section proposer des estimateurs  $\widehat{\theta}_T$  et  $\widehat{\rho}_T$  respectivement de  $\theta$  et de  $\rho$ , adaptés au modèle OU(1)–OU(1). Il s'agira d'étudier dans un premier temps leur comportement asymptotique, puis de tenter dans un second temps d'établir leur distribution asymptotique. Nous pourrions alors constater que la normalité asymptotique de  $\widehat{\rho}_T$  n'est pas atteinte sur tout le spectre  $-\infty < \rho \leq 0$ , et nous en expliquerons les raisons. Les deux exemples qui nous accompagneront tout au long de ce chapitre ont été générés avec  $\theta = -0.2$  et  $\rho = -0.3$  dans un premier temps, avec  $\theta = -1.5$  et  $\rho = -0.6$  dans un second temps. Les séries, représentées sur la Figure V.1, sont de taille  $T = 1000$  et nous avons mis en place un schéma de discrétisation basique (avec un pas de 1/100) pour approximer le calcul intégral et illustrer les convergences.

### V.2.1 Le paramètre de l'autorégression

Si l'on cherche à estimer  $\theta$  par maximum de vraisemblance dans le modèle (V.4) tel qu'on le fait dans un modèle d'Ornstein-Uhlenbeck standard, alors on obtient

$$\widehat{\theta}_T = \frac{\int_0^T Y_t dY_t}{\int_0^T Y_t^2 dt} = \frac{Y_T^2 - T}{2 \int_0^T Y_t^2 dt} \quad (\text{V.5})$$

par application du lemme d'Itô. Nous adopterons ainsi ce point de vue et cet estimateur, en relation avec la logique de l'étude voulant que l'on ne sache pas *a priori* si le coefficient de corrélation

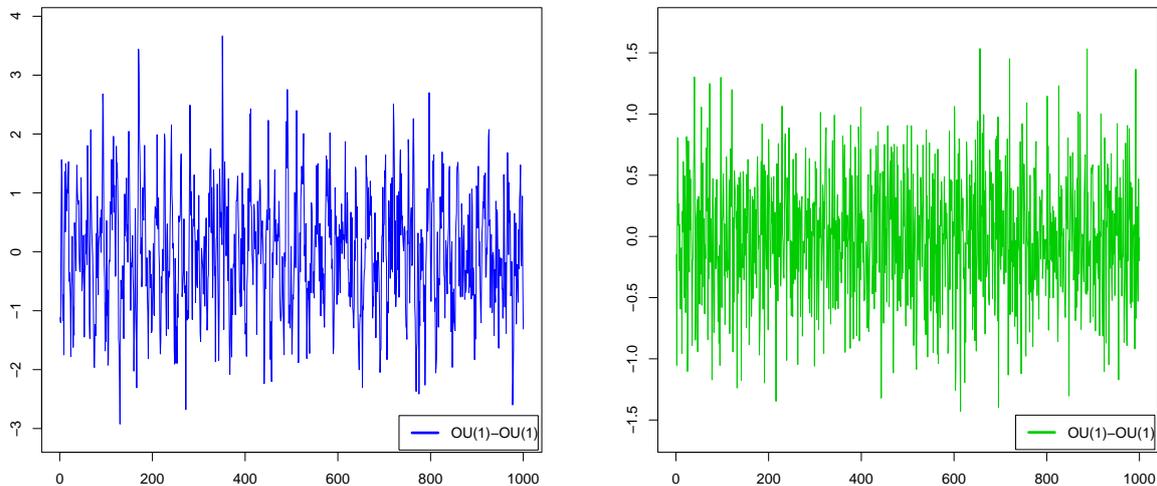


FIGURE V.1 – Exemples de trajectoires OU(1)–OU(1) stables.

résiduelle est significatif, ce que nous nous attacherons du reste à tester.

### V.2.1.1 Convergence presque sûre

Il est bien connu que l'estimateur  $\widehat{\theta}_T$  est fortement consistant lorsque  $\rho = 0$ , cependant nous devons ici définir la valeur

$$\theta^* = \theta + \rho. \tag{V.6}$$

La convergence présentée ci-dessous sera illustrée dans la section suivante, sur la Figure V.3.

**Théorème V.1.** Soit  $\widehat{\theta}_T$  l'estimateur du maximum de vraisemblance donné par (V.5) dans le modèle (V.4). Alors, on a la convergence presque sûre

$$\lim_{T \rightarrow \infty} \widehat{\theta}_T = \theta^* \quad \text{p.s.}$$

où la valeur limite est donnée par (V.6).

Ainsi, comme nous l'avons constaté dans le Chapitre II pour le modèle AR(1)–AR(1), la présence de corrélation résiduelle entraîne la perte de la consistance forte de l'estimateur du paramètre  $\theta$ . Et c'est de l'ampleur du biais occasionné que nous tirerons l'éventuelle évidence statistique de la présence de corrélation résiduelle.

### V.2.1.2 Normalité asymptotique

Étudions tout d'abord la variabilité de  $\widehat{\theta}_T$  autour de sa valeur limite. Pour cela, soit la variance

$$\sigma_\theta^2 = -2\theta^*. \quad (\text{V.7})$$

La normalité asymptotique stipulée dans le théorème suivant est illustrée sur la Figure V.2, correspondant à un échantillon de  $N = 1000$  répliques des deux exemples que nous avons décrits en préambule. Ces derniers sont moins précis que lors des chapitres précédents, la faute à un schéma de discrétisation loin d'être optimal.

**Théorème V.2.** Soit  $\widehat{\theta}_T$  l'estimateur du maximum de vraisemblance donné par (V.5) dans le modèle (V.4). Alors, on a la normalité asymptotique

$$\sqrt{T} (\widehat{\theta}_T - \theta^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_\theta^2)$$

où la variance limite est donnée par (V.7).

*Démonstration.* Les Théorèmes V.1–V.2 sont prouvés dans la Section V.5.2. ★

Là encore, remarquons la cohérence de l'étude avec les résultats bien connus relatifs au processus d'Ornstein-Uhlenbeck. On sait en effet que la normalité asymptotique vérifiée par  $\widehat{\theta}_T$  lorsque  $\rho = 0$  est de variance  $\sigma_\theta^2 = -2\theta$ .

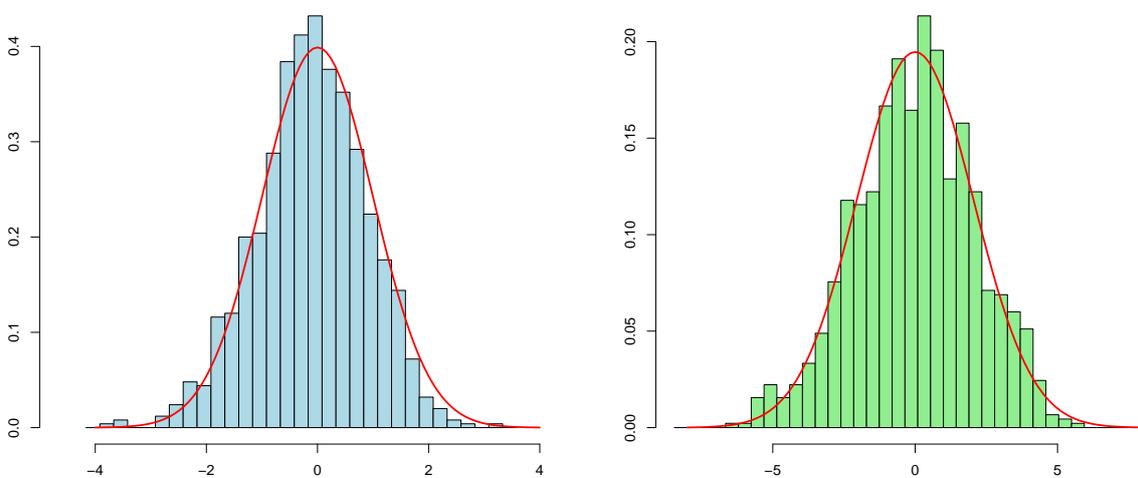


FIGURE V.2 – Normalité asymptotique de  $\widehat{\theta}_T$  pour deux OU(1)–OU(1) stables.

## V.2.2 Le paramètre de l'autocorrélation résiduelle

Comme nous l'avons fait dans le cas discret, nous devons commencer par bâtir un ensemble résiduel ( $\widehat{V}_t$ ) sur l'intervalle continu  $[0, T]$ , afin de proposer un estimateur de  $\rho$ . On définit ainsi, pour tout  $0 \leq t \leq T$ ,

$$\widehat{V}_t = Y_t - \widehat{\theta}_T \Sigma_t \quad \text{avec} \quad \Sigma_t = \int_0^t Y_s ds. \quad (\text{V.8})$$

Par analogie avec (V.5), nous considérons alors l'estimateur de  $\rho$  donné par

$$\widehat{\rho}_T = \frac{\widehat{V}_T^2 - T}{2 \int_0^T \widehat{V}_t^2 dt}. \quad (\text{V.9})$$

Ce dernier a une interprétation au sens du maximum de vraisemblance, relativement à la diffusion vérifiée par  $(V_t)$ . Notons que nous n'avons pas souhaité adopter pour numérateur l'expression attendue, à savoir  $\int_0^T \widehat{V}_t d\widehat{V}_t$ , d'une part car nous ne sommes pas certains de l'interprétation qu'il faille donner à  $d\widehat{V}_t$ , d'autre part car nous serions alors confrontés à des problèmes de  $(\mathcal{F}_t)$ -mesurabilité lors de la démonstration de nos résultats (puisque  $\widehat{\theta}_T$  est présent dans l'expression associée). Ainsi, nous avons choisi de considérer au numérateur la quantité que nous nous serions attendus à voir émerger du lemme d'Itô, si d'aventure l'intégrale en question était bien calculable.

### V.2.2.1 Convergence presque sûre

La convergence présentée ci-dessous est illustrée sur la Figure V.3, en compagnie de la convergence de  $\widehat{\theta}_T$ , pour les deux exemples définis en préambule. Définissons la limite

$$\rho^* = \frac{\theta\rho(\theta + \rho)}{(\theta + \rho)^2 + \theta\rho}. \quad (\text{V.10})$$

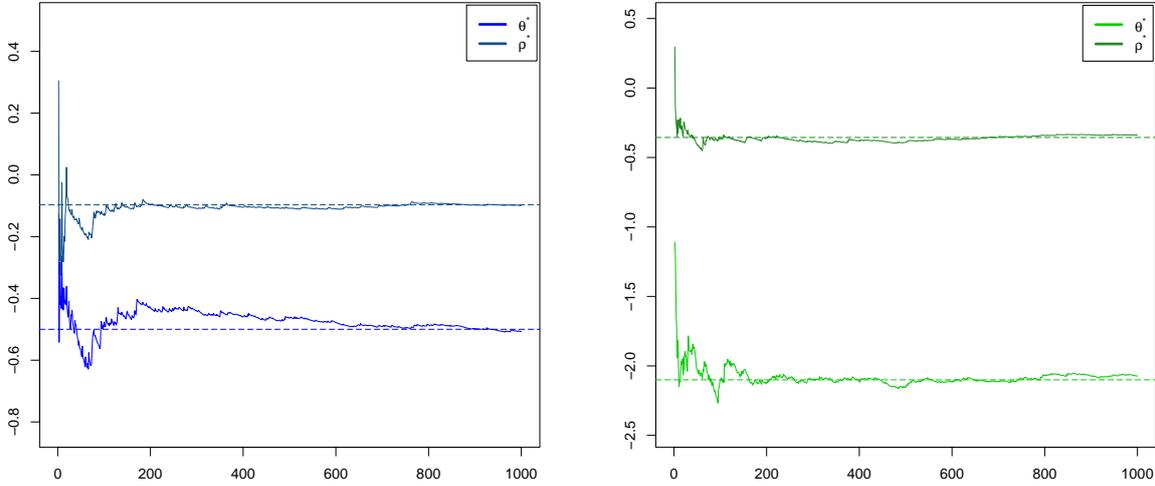
**Théorème V.3.** Soit  $\widehat{\rho}_T$  l'estimateur du maximum de vraisemblance donné par (V.9) dans le modèle (V.4). Alors, on a la convergence presque sûre

$$\lim_{T \rightarrow \infty} \widehat{\rho}_T = \rho^* \quad \text{p.s.}$$

où la valeur limite est donnée par (V.10).

### V.2.2.2 Normalité asymptotique

C'est précisément ici que se situe la rupture avec le cas discret AR(1)–AR(1). Nous allons voir que certes une normalité asymptotique existe pour  $\widehat{\rho}_T$  sous les hypothèses de stabilité  $\theta < 0$  et  $\rho < 0$ , mais que celle-ci est dégénérée sur la frontière  $\rho = 0$ , pourtant toujours source de stabilité du processus


 FIGURE V.3 – Convergence de  $\widehat{\theta}_T$  et  $\widehat{\rho}_T$  pour deux OU(1)–OU(1) stables.

( $Y_t$ ) grâce à  $\theta < 0$ , mais plus de ( $V_t$ ). Pour bien comprendre ce qui se passe sur la frontière, nous avons investigué ce cas particulier. Les distributions asymptotiques obtenues sont résumées dans le théorème ci-dessous. Notons pour commencer

$$\Gamma = \begin{pmatrix} \sigma_\theta^2 & \ell \\ \ell & \sigma_\rho^2 \end{pmatrix} \quad (\text{V.11})$$

où la variance asymptotique  $\sigma_\theta^2$  est donnée par (V.7), où

$$\sigma_\rho^2 = -\frac{2\rho^*((\theta^*)^6 + \theta\rho((\theta^*)^4 - \theta\rho(2(\theta^*)^2 - \theta\rho)))}{((\theta^*)^2 + \theta\rho)^3} \quad (\text{V.12})$$

et la covariance associée

$$\ell = \frac{2\rho^*((\theta^*)^2 - \theta\rho)}{(\theta^*)^2 + \theta\rho}. \quad (\text{V.13})$$

**Théorème V.4.** Soient  $\widehat{\theta}_T$  et  $\widehat{\rho}_T$  les estimateurs du maximum de vraisemblance donnés par (V.5) et (V.9) dans le modèle (V.4), et supposons que  $\rho < 0$ . Alors, on a la normalité asymptotique jointe

$$\sqrt{T} \begin{pmatrix} \widehat{\theta}_T - \theta^* \\ \widehat{\rho}_T - \rho^* \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma)$$

où la matrice de covariance limite est donnée par (V.11). En particulier, on a la normalité asymptotique

$$\sqrt{T}(\widehat{\rho}_T - \rho^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_\rho^2)$$

où la variance limite est donnée par (V.12). Supposons maintenant que  $\rho = 0$ . Alors, on a la distribution asymptotique

$$T\widehat{\rho}_T \xrightarrow{\mathcal{L}} \frac{\int_0^1 B_s dB_s}{\int_0^1 B_s^2 ds}$$

où  $(B_t)$  est un mouvement brownien standard.

*Démonstration.* Les Théorèmes V.3–V.4 sont prouvés dans la Section V.5.2. ★

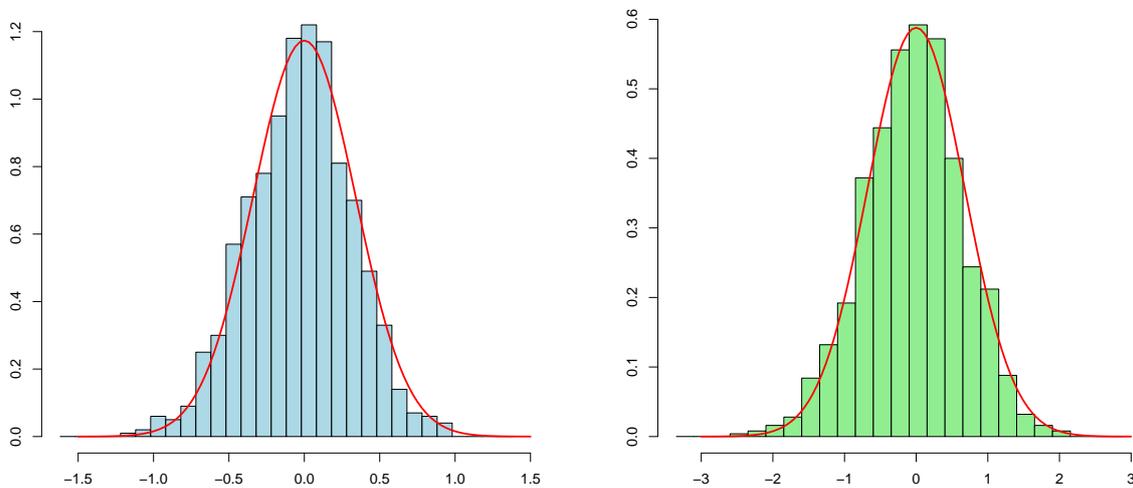


FIGURE V.4 – Normalité asymptotique de  $\widehat{\rho}_T$  pour deux OU(1)–OU(1) stables.

Sur la Figure V.4, un échantillon de  $N = 1000$  répliques des deux exemples que nous avons décrits permet d’illustrer la normalité asymptotique de  $\widehat{\rho}_T$  lorsque  $\rho < 0$ . Sur la Figure V.5, nous observons sa distribution asymptotique dans le cadre d’instabilité de  $(V_t)$  correspondant à  $\rho = 0$ , pour  $\theta = -0.4$  puis pour  $\theta = -1.5$ . Nous y avons superposé la distribution (pseudo-)théorique issue d’une large simulation de la décomposition de Karhunen-Loève que nous allons rappeler.

**Lemme V.1.** Soient les séries de Karhunen-Loève

$$\mathcal{T} = \sqrt{2} \sum_{n=1}^{\infty} \gamma_n Z_n \quad \text{et} \quad \mathcal{S} = \sum_{n=1}^{\infty} \gamma_n^2 Z_n^2$$

où  $\gamma_n = 2(-1)^n / ((2n - 1)\pi)$  et  $(Z_n)$  est une suite indépendante et identiquement distribuée de variables aléatoires  $\mathcal{N}(0, 1)$ . Alors,

$$\frac{\int_0^1 B_s dB_s}{\int_0^1 B_s^2 ds} \stackrel{\mathcal{L}}{=} \frac{\mathcal{T}^2 - 1}{2\mathcal{S}}.$$

*Démonstration.* Voir par exemple le Corollaire 3.1.3 de [31].

★

Nous avons quelque peu abordé ce sujet dans le Chapitre I, en Section I.3.2.1, et nous pouvons également consulter [132] ou [55] sur le même principe. Nous avons vu en particulier que cette distribution asymptotique est celle d'un estimateur correctement renormalisé d'une racine unitaire dans un processus autorégressif. Cette frontière, aussi bien au niveau de la vitesse de convergence (qui passe brusquement de  $\sqrt{T}$  à  $T$ ) qu'au niveau de la distribution asymptotique (qui perd son caractère gaussien et symétrique) est caractéristique du passage de la stabilité à l'instabilité. Cela se comprend tout à fait ici, puisque le fait de passer de  $\rho < 0$  à  $\rho = 0$  équivaut à bifurquer de l'estimation d'un paramètre (négatif) dans un modèle d'Ornstein-Uhlenbeck stable à l'estimation de ce même paramètre (nul) dans un mouvement brownien instable, dont la variance est bien sûr explosive. Ce sujet formera en outre la problématique majeure du chapitre suivant. L'avantage que cela nous procure dans le cadre de cette étude, c'est qu'il est dès lors aisé de bâtir un test statistique axé sur la valeur de  $\rho$ , ce que nous ferons dans la prochaine section.

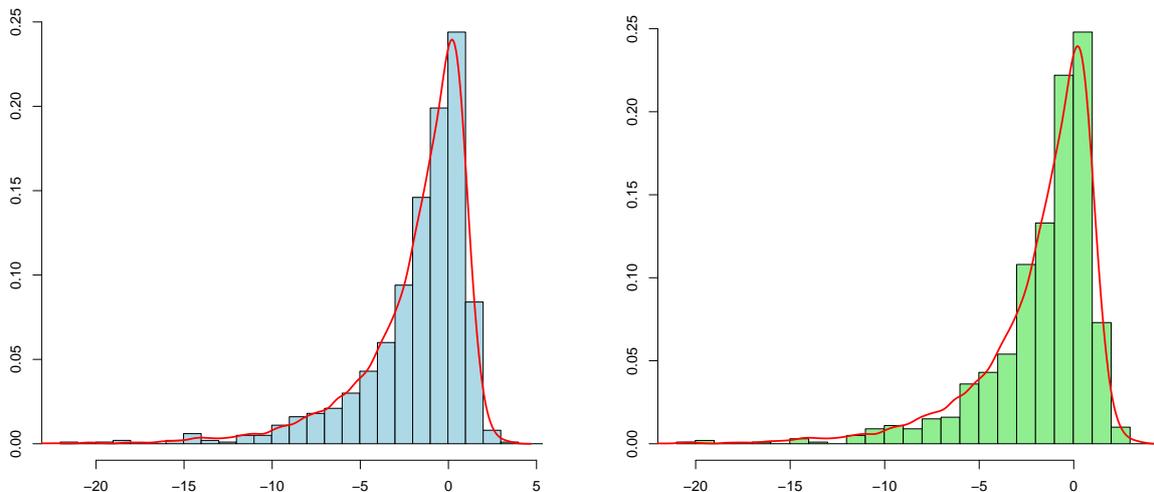


FIGURE V.5 – Distribution asymptotique de  $\hat{\rho}_T$  pour deux OU(1)–OU(1) dont la perturbation est instable.

## V.3 Application statistique

### V.3.1 Retour à Durbin-Watson

Nous avons établi, dans les Chapitres II–III, que la statistique de Durbin-Watson  $\hat{D}_n$  était, en temps discret, un équivalent asymptotique presque sûr de la quantité  $2(1 - \hat{\rho}_n)$ . L'estimateur  $\hat{\rho}_T$  nous

offre dès lors la possibilité d'une statistique de Durbin-Watson  $\widehat{D}_T$  en temps continu que nous pourrions en toute logique définir comme

$$\widehat{D}_T = 2(1 - \widehat{\rho}_T) = \frac{2 \int_0^T \widehat{V}_t^2 dt - \widehat{V}_T^2 + T}{\int_0^T \widehat{V}_t^2 dt}. \quad (\text{V.14})$$

À partir des résultats établis sur  $\widehat{\rho}_T$ , on en tire immédiatement le corollaire suivant.

**Corollaire V.1.** Soit  $\widehat{D}_T$  la statistique de Durbin-Watson en temps continu donnée par (V.14). Alors, on a la convergence presque sûre

$$\lim_{T \rightarrow \infty} \widehat{D}_T = D^* \quad \text{p.s.}$$

où la valeur limite est définie par  $D^* = 2(1 - \rho^*)$ . De plus, si l'on suppose que  $\rho < 0$ , on a la normalité asymptotique

$$\sqrt{T} (\widehat{D}_T - D^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_D^2)$$

où la variance limite est définie par  $\sigma_D^2 = 4\sigma_\rho^2$ . Si maintenant l'on suppose que  $\rho = 0$ , alors on a la distribution asymptotique

$$T (\widehat{D}_T - 2) \xrightarrow{\mathcal{L}} -\frac{2 \int_0^1 B_s dB_s}{\int_0^1 B_s^2 ds}$$

où  $(B_t)$  est un mouvement brownien standard.

*Démonstration.* La preuve est laissée au soin du lecteur dans la mesure où ce corollaire découle immédiatement des Théorèmes V.3–V.4. ★

### V.3.2 Évidence empirique

Il est clair que, sous les hypothèses  $\theta < 0$  et  $\rho \leq 0$ , la valeur limite  $\rho^*$  donnée par (V.10) ne s'annule pas dès que  $\rho < 0$ . Cette remarque est extrêmement importante en soi, puisqu'elle exclut de fait toute situation pathologique comme l'on avait pu en rencontrer dans le cas discret. Ainsi, il est facile d'étudier le comportement asymptotique d'une statistique adaptée au test

$$\mathcal{H}_0 : \text{"}\rho = 0\text{"} \quad \text{vs} \quad \mathcal{H}_1 : \text{"}\rho < 0\text{"}.$$

Nous proposons à cet égard la quantité suivante, vérifiant les propriétés asymptotiques, respectivement sous  $\mathcal{H}_0$  puis sous  $\mathcal{H}_1$ ,

$$T^2 (\widehat{D}_T - 2)^2 \xrightarrow{\mathcal{L}} \frac{4 \left( \int_0^1 B_s dB_s \right)^2}{\left( \int_0^1 B_s^2 ds \right)^2} \quad \text{et} \quad \lim_{T \rightarrow \infty} T^2 (\widehat{D}_T - 2)^2 = +\infty \quad \text{p.s.}$$

issues du Corollaire V.1. De telles procédures n'existant, à notre connaissance, pas encore, il nous est impossible de comparer sa puissance empirique avec un ensemble de tests de référence. Nous avons cependant simulé un ensemble de  $N = 1000$  trajectoires de taille  $T = 300$  dans un premier temps, de taille  $T = 30$  dans un second temps pour adjoindre les petits échantillons à l'étude, et ce pour  $\theta = \{-1, -0.5, -0.2\}$  et  $\rho$  variant sur le spectre  $[-2, 0]$ . Pour chaque trajectoire, nous avons mis en pratique la procédure de test définie ci-dessus en vue d'obtenir un estimateur de

$$\mathbb{P}(\text{rejeter } \mathcal{H}_0 \mid \mathcal{H}_1 \text{ est vraie}).$$

La fréquence empirique de non rejet de  $\mathcal{H}_0$  est résumée sur la Figure V.6.

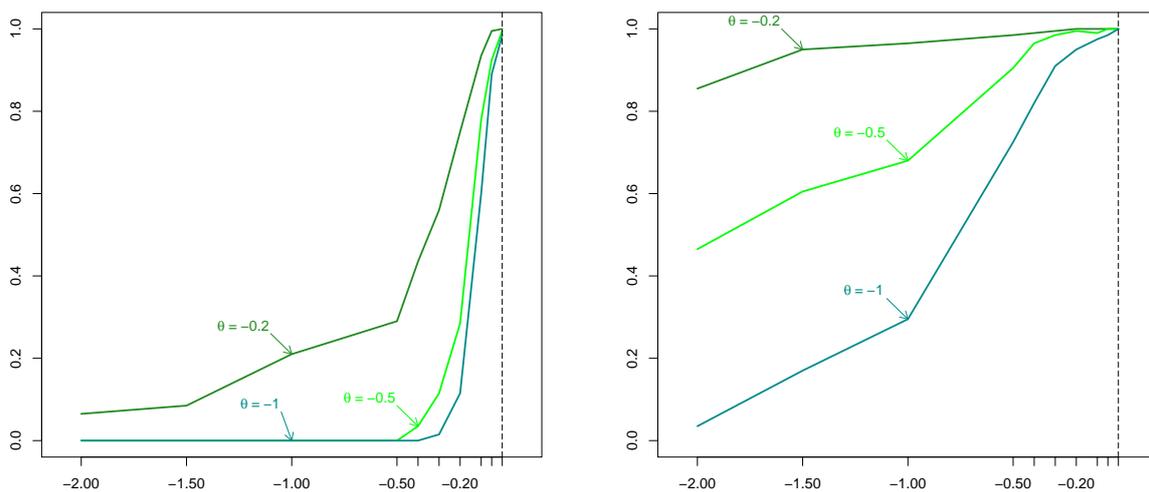


FIGURE V.6 – Fréquence de non rejet de  $\mathcal{H}_0$  en ordonnée et  $\rho$  variant de  $-2$  à  $0$  en abscisse, pour  $T = 300$  (gauche) et  $T = 30$  (droite).

Il est ainsi manifeste que la procédure suggérée donne d'excellents résultats dans le cas d'échantillons de taille raisonnable, et ce d'autant plus que la significativité du coefficient  $\theta$  augmente. Il ressort globalement, dans ce cas, que la procédure ne se trompe que très rarement dès que  $\rho < -0.20$ , ce qui, comparativement à la procédure suggérée dans le Chapitre II adaptée au cas discret, reste très satisfaisant. Cela s'explique en particulier par le fait que la frontière stabilité/instabilité est ici franchie, et qu'en conséquence la statistique de test est rejetée à l'infini sous  $\mathcal{H}_1$  avec une vitesse  $T^2$  alors qu'elle ne l'était qu'à la vitesse  $T$  dans le cas discret, d'où une facilité toute relative à mieux discriminer  $\mathcal{H}_0$  et  $\mathcal{H}_1$ . Bien évidemment, sur les petits échantillons, les résultats sont assez flous mais on repère malgré tout le comportement attendu et des résultats sous  $\mathcal{H}_0$  plutôt corrects pour les valeurs conséquentes de  $\theta$ .

## V.4 Conclusion et perspectives

Même si la pertinence empirique peut laisser sceptique de premier abord, nous pensons avoir d'une part modestement élargi l'étude théorique du processus d'Ornstein-Uhlenbeck, en le munissant d'une perturbation jamais encore considérée, et d'autre part entrouvert une porte quelque peu insolite, caractérisée par le test d'hypothèse en temps continu. Si nous souhaitions pousser le raisonnement encore un peu plus loin, nous pourrions conclure de la procédure statistique que nous avons entre les mains une trajectoire d'Ornstein-Uhlenbeck standard, ou une trajectoire continue autocorrélée au second ordre. Dans chacun des cas, l'inférence statistique est différente. Sous  $\mathcal{H}_0$  en effet, l'estimateur de  $\theta$  est bien connu et son comportement asymptotique déjà établi (comme nous l'avons rappelé tout au long de ce chapitre). Sous  $\mathcal{H}_1$  en revanche, le processus s'écrit, pour  $0 \leq t \leq T$ ,

$$Y_t = (\theta + \rho)\Sigma_t - \theta\rho\Pi_t + W_t$$

avec

$$\Sigma_t = \int_0^t Y_s ds \quad \text{et} \quad \Pi_t = \int_0^t \Sigma_s ds.$$

Il est alors aisé de construire un estimateur du maximum de vraisemblance fortement consistant des vrais paramètres du modèle, à savoir  $\theta + \rho$  et  $-\theta\rho$ , par l'intermédiaire de

$$\widehat{\vartheta}_T = \left( \int_0^T \Phi_t \Phi_t' dt \right)^{-1} \int_0^T \Phi_t dY_t$$

où  $\Phi_t = (Y_t \quad \Sigma_t)'$ . Il s'ensuit également que cet estimateur est asymptotiquement normal, de matrice de covariance

$$\Delta = \begin{pmatrix} -2\theta^* & 0 \\ 0 & -2\theta\rho\theta^* \end{pmatrix}$$

et qu'ainsi ses composantes sont asymptotiquement indépendantes. En résumé, nos objectifs étaient doubles : étendre l'étude du Chapitre II au temps continu, et travailler uniquement avec des estimateurs continus (sans passer par la discrétisation du processus autrement que pour le simuler). La suite logique de cette étude serait d'établir ces mêmes résultats dans le cas d'une perturbation de Lévy dans le processus  $(V_t)$ . Nous n'y avons pas encore réfléchi même si cela ne nous semble pas forcément inaccessible, avant bien sûr d'aborder les équivalents fractionnaires de ces perturbations. Il pourrait également être intéressant d'estimer la volatilité d'un tel processus par un estimateur continu, et de comparer l'efficacité d'un modèle OU-OU à erreurs autocorrélées avec les modèles à volatilité dynamique. Nous allons désormais aborder l'ultime chapitre de ce manuscrit, dédié à la problématique cruciale de la stationnarité dans le processus autorégressif.

## V.5 Annexe : démonstration des résultats

Il est indispensable, afin de bien structurer et de faciliter la démonstration de nos résultats, de commencer par stipuler un ensemble de lemmes techniques dont le premier est une clé de l'étude. Rappelons que l'on dispose d'une trajectoire continue de  $(Y_t)$  sur l'intervalle  $[0, T]$ , et notons tout d'abord

$$S_T = \int_0^T Y_t^2 dt, \quad P_T = \int_0^T Y_t V_t dt, \quad \Lambda_T = \int_0^T V_t^2 dt \quad \text{et} \quad \widehat{\Lambda}_T = \int_0^T \widehat{V}_t^2 dt. \quad (\text{V.15})$$

Définissons également les  $(\mathcal{F}_T)$ -martingales continues

$$M_T^Y = \int_0^T Y_t dW_t \quad \text{et} \quad M_T^V = \int_0^T V_t dW_t \quad (\text{V.16})$$

où  $\mathcal{F}_t$  désigne la filtration naturelle associée au mouvement brownien standard  $(W_t)$ .

### V.5.1 L'ergodicité et ses implications

**Lemme V.2.** *Sous les conditions de stabilité  $\theta < 0$  et  $\rho \leq 0$ , le processus  $(Y_t)$  solution du modèle (V.4) est géométriquement ergodique.*

*Démonstration.* Si  $\rho = 0$ , le processus est celui d'Ornstein-Uhlenbeck et le résultat d'ergodicité géométrique est bien connu dès que  $\theta < 0$ . Considérons donc que  $\rho < 0$ . De l'écriture du modèle OU(1)-OU(1), on tire

$$dY_t = (\theta + \rho)Y_t dt - \theta\rho\Sigma_t dt + dW_t \quad (\text{V.17})$$

où l'on dénote

$$\Sigma_t = \int_0^t Y_s ds.$$

Ainsi, avec  $\Phi_t = (Y_t \quad \Sigma_t)'$ , on a le processus d'Ornstein-Uhlenbeck vectoriel,

$$d\Phi_t = A\Phi_t dt + dB_t$$

avec

$$A = \begin{pmatrix} \theta + \rho & -\theta\rho \\ 1 & 0 \end{pmatrix} \quad \text{et} \quad B_t = \begin{pmatrix} W_t \\ 0 \end{pmatrix}$$

et l'on montre facilement, comme nous l'avons fait au Chapitre IV (dans la preuve du Lemme IV.4), que  $\lambda_{\max}(A) = \theta \vee \rho < 0$ . Ce résultat achève la preuve. ★

**Corollaire V.2.** Soit la quantité  $S_T$  donnée par (V.15). Alors, on a la convergence presque sûre

$$\lim_{T \rightarrow \infty} \frac{S_T}{T} = -\frac{1}{2(\theta + \rho)} \quad \text{p.s.}$$

*Démonstration.* En vertu de l'ergodicité du processus donnée par le Lemme V.2, il nous suffit de déterminer la valeur limite de  $\mathbb{E}[Y_t^2]$ . Reprenant les notations définies ci-dessus, on tire du lemme d'Itô que  $U_t$  vérifie le système différentiel

$$\frac{\partial U_t}{\partial t} = C U_t + e$$

avec

$$U_t = \begin{pmatrix} \mathbb{E}[Y_t^2] \\ \mathbb{E}[\Sigma_t^2] \\ \mathbb{E}[Y_t \Sigma_t] \end{pmatrix}, \quad C = \begin{pmatrix} 2(\theta + \rho) & 0 & -2\theta\rho \\ 0 & 0 & 2 \\ 1 & -\theta\rho & \theta + \rho \end{pmatrix} \quad \text{et} \quad e = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

Dès lors, on trouve  $\lambda_{\max}(C) = \theta + \rho \vee 2\theta \vee 2\rho < 0$  à partir du moment où  $\rho < 0$ . Nous avons dans ce cas  $\lim_{t \rightarrow \infty} U_t = -C^{-1}e$ . Ainsi, par calcul direct,

$$\lim_{t \rightarrow \infty} \mathbb{E}[Y_t^2] = -\frac{1}{2(\theta + \rho)}, \quad \lim_{t \rightarrow \infty} \mathbb{E}[\Sigma_t^2] = -\frac{1}{2\theta\rho(\theta + \rho)} \quad \text{et} \quad \lim_{t \rightarrow \infty} \mathbb{E}[Y_t \Sigma_t] = 0, \quad (\text{V.18})$$

ce qui achève la preuve. Si maintenant  $\rho = 0$ , alors  $(Y_t)$  est un processus d'Ornstein-Uhlenbeck ergodique et la valeur de sa variance asymptotique est  $-1/(2\theta)$  comme nous l'avons redémontré dans l'introduction. ★

**Corollaire V.3.** Soit la quantité  $\widehat{\Lambda}_T$  donnée par (V.15), et supposons que  $\rho < 0$ . Alors, on a la convergence presque sûre

$$\lim_{T \rightarrow \infty} \frac{\widehat{\Lambda}_T}{T} = -\frac{(\theta + \rho)^2 + \theta\rho}{2\theta\rho(\theta + \rho)} \quad \text{p.s.}$$

*Démonstration.* Sous l'hypothèse  $\rho < 0$ ,  $(V_t)$  est un processus d'Ornstein-Uhlenbeck ergodique, et ce dernier vérifie à ce titre

$$\lim_{T \rightarrow \infty} \frac{\Lambda_T}{T} = -\frac{1}{2\rho} \quad \text{p.s.} \quad (\text{V.19})$$

où la quantité  $\Lambda_T$  est également définie en (V.15). De plus, on trouve par intégration  $Y_t = \theta\Sigma_t + V_t$  pour  $0 \leq t \leq T$ , ce qui implique que

$$\int_0^T Y_t \Sigma_t dt = \frac{1}{\theta}(S_T - P_T) \quad \text{avec} \quad \Sigma_t = \int_0^t Y_s ds. \quad (\text{V.20})$$

Cependant, le Corollaire V.2 – et plus précisément (V.18) – est également à l'origine de la convergence

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T Y_t \Sigma_t dt = 0 \quad \text{p.s.}$$

ce qui, combiné à (V.20) et au Corollaire V.2, entraîne que

$$\lim_{T \rightarrow \infty} \frac{P_T}{T} = -\frac{1}{2(\theta + \rho)} \quad \text{p.s.} \quad (\text{V.21})$$

Par ailleurs, nous tirons de la définition de  $\widehat{V}_T$  en (V.8) la décomposition

$$\widehat{\Lambda}_T = I_T + (\widehat{\theta}_T - \theta^*) (J_T + (\widehat{\theta}_T - \theta^*) K_T) \quad (\text{V.22})$$

avec

$$\begin{aligned} \theta^2 I_T &= \rho^2 S_T + (\theta^*)^2 \Lambda_T - 2\theta^* \rho P_T, \\ \theta^2 J_T &= 2\rho S_T + 2\theta^* \Lambda_T - 2(\theta + 2\rho) P_T, \\ \theta^2 K_T &= S_T + \Lambda_T - 2P_T. \end{aligned}$$

Par l'intermédiaire du Théorème V.1, du Corollaire V.2 et des convergences (V.19) et (V.21), il vient

$$\lim_{T \rightarrow \infty} \frac{\widehat{\Lambda}_T}{T} = \lim_{T \rightarrow \infty} \frac{I_T}{T} = -\frac{(\theta + \rho)^2 + \theta\rho}{2\theta\rho(\theta + \rho)} \quad \text{p.s.}$$

après quelques manipulations supplémentaires, ce qui achève la preuve. ★

**Corollaire V.4.** *Supposons que  $\rho < 0$ . Alors, on a les normalités asymptotiques*

$$Y_T \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, -\frac{1}{2(\theta + \rho)}\right) \quad \text{et} \quad V_T \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, -\frac{1}{2\rho}\right).$$

La normalité asymptotique de  $Y_T$  reste valable lorsque  $\rho = 0$ .

*Démonstration.* La normalité asymptotique du processus d'Ornstein-Uhlenbeck ( $V_t$ ) est un résultat bien connu dès qu'il est ergodique (soit ici  $\rho < 0$ ). Le processus ( $Y_t$ ) est quant à lui gaussien, d'espérance nulle et de variance convergeant vers  $-1/(2(\theta + \rho))$  comme nous l'avons vu dans la preuve du Corollaire V.2. Cela suffit donc à établir la normalité asymptotique stipulée. ★

## V.5.2 Sur les estimateurs

### V.5.2.1 Preuve du Théorème V.1

Nous tirons de l'intégration du modèle (V.4), suivant les notations (V.15) et (V.16), la relation

$$\int_0^T Y_t dY_t = \theta S_T + \rho P_T + M_T^Y. \quad (\text{V.23})$$

Nous avons vu que  $(M_t^Y)$  est une  $(\mathcal{F}_t)$ -martingale continue, adaptée à la filtration naturelle du mouvement brownien standard  $(W_t)$ . Son processus croissant est alors donné par  $S_t$  qui, en vertu du Corollaire V.2, se comporte presque sûrement comme  $t$ . L'extension du Théorème B.2 aux martingales scalaires continues, que l'on trouvera par exemple dans [54] ou [78], nous certifie alors que  $M_T^Y = o(T)$  p.s. Il suit alors de (V.23), du Corollaire V.2 et de (V.21) que

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T Y_t dY_t = -\frac{\theta + \rho}{2(\theta + \rho)} = -\frac{1}{2} \quad \text{p.s.} \quad (\text{V.24})$$

Cela nous permet d'établir que

$$\lim_{T \rightarrow \infty} \frac{\int_0^T Y_t dY_t}{\int_0^T Y_t^2 dt} = \theta + \rho,$$

ce qui conclut la démonstration. ★

### V.5.2.2 Preuve du Théorème V.2

Quelques manipulations de l'expression de  $\widehat{\theta}_T$  en (V.5) nous conduisent à la décomposition

$$\widehat{\theta}_T - \theta^* = \frac{M_T^Y}{S_T} + \frac{R_T^Y}{S_T} \quad (\text{V.25})$$

où le terme résiduel se simplifie, grâce au lemme d'Itô, en

$$R_T^Y = \rho \int_0^T Y_t (V_t - Y_t) dt = -\theta \rho \int_0^T \Sigma_t d\Sigma_t = -\frac{\theta \rho}{2} \Sigma_T^2. \quad (\text{V.26})$$

De même que pour la loi des grands nombres dans la preuve précédente, nous trouverons dans [54] l'extension du théorème central limite aux martingales scalaires continues stipulant, en particulier ici, que

$$\sqrt{T} \frac{M_T^Y}{S_T} \xrightarrow{\mathcal{L}} \mathcal{N}(0, -2(\theta + \rho)) \quad (\text{V.27})$$

en vertu du Corollaire V.2. Par ailleurs, il suit de la normalité asymptotique de  $Y_T$  et de  $V_T$ , donnée par le Corollaire V.4, et de l'inégalité

$$\Sigma_T^2 = \frac{1}{\theta^2} (Y_T - V_T)^2 \leq \frac{2}{\theta^2} Y_T^2 + \frac{2}{\theta^2} V_T^2 \quad (\text{V.28})$$

obtenue par intégration de la première ligne du modèle (V.4), que  $\Sigma_T^2 = O_p(1)$ , et qu'en conséquence

$$\frac{R_T^Y}{\sqrt{T}} \xrightarrow{p} 0. \quad (\text{V.29})$$

Il nous reste à combiner (V.25), (V.27) et (V.29) pour achever la preuve de ce théorème. ★

### V.5.2.3 Preuve du Théorème V.3

Supposons tout d'abord que  $\rho < 0$ , et reprenons le raisonnement développé dans la preuve du Théorème V.1. On montre d'une manière parfaitement identique que l'on a la convergence

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T V_t dV_t = -\frac{1}{2} \quad \text{p.s.} \quad (\text{V.30})$$

Nous obtenons alors, à travers le lemme d'Itô, de (V.24) et de (V.30), les convergences

$$\lim_{T \rightarrow \infty} \frac{Y_T^2}{T} = 0 \quad \text{p.s.} \quad \text{et} \quad \lim_{T \rightarrow \infty} \frac{V_T^2}{T} = 0 \quad \text{p.s.}$$

Cela nous conduit bien sûr à la convergence du numérateur de  $\widehat{\rho}_T$ ,

$$\lim_{T \rightarrow \infty} \frac{\widehat{V}_T^2 - T}{2T} = -\frac{1}{2} \quad \text{p.s.} \quad (\text{V.31})$$

par l'intermédiaire du Théorème V.1, de la définition de  $\widehat{V}_T$  en (V.8) et de la relation (V.28). Ainsi, il ne reste plus qu'à appliquer le Corollaire V.3 pour achever la démonstration sous l'hypothèse  $\rho < 0$ . Si maintenant on a  $\rho = 0$ , alors  $(V_t)$  n'est plus un processus d'Ornstein-Uhlenbeck ergodique, mais un mouvement brownien dont la variance ne se stabilise pas. Il s'ensuit que

$$\widehat{V}_T^2 - T = 2M_T^W - 2W_T \Sigma_T (\widehat{\theta}_T - \theta) + \Sigma_T^2 (\widehat{\theta}_T - \theta)^2 \quad (\text{V.32})$$

où la  $(\mathcal{F}_T)$ -martingale continue  $(M_T^W)$  est donnée par

$$M_T^W = \int_0^T W_t dW_t$$

et  $\Sigma_T$  est donné en (V.8). Nous avons de même

$$\widehat{\Lambda}_T = \Lambda_T - 2(\widehat{\theta}_T - \theta) \int_0^T W_t \Sigma_t dt + (\widehat{\theta}_T - \theta)^2 \int_0^T \Sigma_t^2 dt, \quad (\text{V.33})$$

avec les notations de (V.15). Cependant, nous tirons de l'autosimilarité du mouvement brownien que

$$\Lambda_T = \int_0^T W_t^2 dt \stackrel{\mathcal{L}}{=} T \int_0^T W_{t/T}^2 dt = T^2 \int_0^1 W_s^2 ds = T^2 \Lambda_1. \quad (\text{V.34})$$

Il suit donc de (V.34) que, pour toute puissance  $0 < a < 2$ ,

$$\lim_{T \rightarrow \infty} \frac{\Lambda_T}{T^a} = +\infty \quad \text{p.s.} \quad (\text{V.35})$$

En effet, dans la mesure où  $\Lambda_1$  est presque sûrement positive, il nous suffit de montrer que

$$\lim_{T \rightarrow \infty} \mathbb{E} \left[ \exp \left( -\frac{\Lambda_T}{T^a} \right) \right] = 0,$$

ce qui est assuré par un calcul gaussien standard, comme nous pouvons en trouver l'exemple dans [84]. Ainsi, on obtient

$$\mathbb{E} \left[ \exp \left( -\frac{\Lambda_T}{T^a} \right) \right] = \mathbb{E} \left[ \exp \left( -\frac{T^2}{T^a} \Lambda_1 \right) \right] = \frac{1}{\sqrt{\cosh(v_T(a))}}$$

où la fonction  $v_T(a) = \sqrt{2T^{2-a}}$  diverge avec  $0 < a < 2$ , ce qui confirme (V.35). De plus, le processus croissant de la  $(\mathcal{F}_t)$ -martingale  $(M_t^W)$  étant donné par  $\Lambda_t$  qui, en vertu de (V.35), tend bien vers l'infini, on a  $M_T^W = o(\Lambda_T)$  p.s. De plus, par la décomposition (V.25) et la loi forte des grands nombres pour les martingales scalaires à temps continu, nous obtenons également

$$(\widehat{\theta}_T - \theta)^2 = O\left(\frac{\log T}{T}\right) \quad \text{p.s.} \quad \text{et} \quad \lim_{T \rightarrow \infty} \frac{W_T}{T} = 0 \quad \text{p.s.} \quad (\text{V.36})$$

Or, on a vu que  $\theta \Sigma_T = Y_T - W_T$  et que, par le Corollaire V.2,  $Y_T = o(T)$  p.s. On trouve alors que  $\Sigma_T^2 = o(T^2)$  p.s. et donc que

$$\Sigma_T^2 (\widehat{\theta}_T - \theta)^2 = o(T \log T) = o(\Lambda_T) \quad \text{p.s.} \quad \text{et} \quad W_T \Sigma_T (\widehat{\theta}_T - \theta) = o(T^{3/2} \log T) = o(\Lambda_T) \quad \text{p.s.}$$

On en déduit que, sous l'hypothèse  $\rho = 0$ , *via* (V.32),

$$\widehat{V}_T^2 - T = o(\Lambda_T) \quad \text{p.s.} \quad (\text{V.37})$$

Étudions désormais le comportement asymptotique de  $\widehat{\Lambda}_T$ . De l'inégalité

$$\int_0^T \Sigma_t^2 dt \leq \frac{2}{\theta^2} (S_T + \Lambda_T),$$

on tire facilement, par l'intermédiaire de (V.36), que

$$(\widehat{\theta}_T - \theta)^2 \int_0^T \Sigma_t^2 dt = o(\Lambda_T) \quad \text{p.s.} \quad \text{et} \quad (\widehat{\theta}_T - \theta) \int_0^T W_t \Sigma_t dt = o(\Lambda_T) \quad \text{p.s.}$$

Nous obtenons ainsi, pour le dénominateur de  $\widehat{\rho}_T$  développé en (V.33),

$$\widehat{\Lambda}_T = \Lambda_T (1 + o(1)) \quad \text{p.s.} \tag{V.38}$$

Ainsi, la combinaison de (V.37) et de (V.38) nous permet d'achever la preuve puisque lorsque  $\rho = 0$ , on a bien  $\rho^* = 0$ . ★

#### V.5.2.4 Preuve du Théorème V.4

Considérons dans un premier temps le cas stable, où  $\rho < 0$ . La preuve est alors très calculatoire, mais paradoxalement relativement facile. En reprenant les notations (V.15) et l'ensemble résiduel (V.8), il vient, pour tout  $0 \leq t \leq T$ ,

$$\begin{aligned} \widehat{V}_t &= Y_t - \widehat{\theta}_T \Sigma_t = V_t - (\widehat{\theta}_T - \theta) \Sigma_t = V_t - (\widehat{\theta}_T - \theta^*) \Sigma_t - \rho \Sigma_t \\ &= V_t - \frac{\rho}{\theta} (Y_t - V_t) - \frac{1}{\theta} (\widehat{\theta}_T - \theta^*) (Y_t - V_t) = \frac{\theta^*}{\theta} V_t - \frac{\rho}{\theta} Y_t - \frac{1}{\theta} (\widehat{\theta}_T - \theta^*) (Y_t - V_t), \end{aligned}$$

ce qui nous conduit précisément à l'écriture (V.22). En réutilisant les notations associées et *via* l'expression de  $\widehat{\rho}_T$  en (V.9), on en déduit l'écriture

$$2\widehat{\Lambda}_T (\widehat{\rho}_T - \rho^*) = I_T^V + (\widehat{\theta}_T - \theta^*) (J_T^V + (\widehat{\theta}_T - \theta^*) K_T^V)$$

avec  $I_T^V = \widehat{V}_T^2 - T - 2\rho^* I_T$ ,  $J_T^V = -2\rho^* J_T$  et  $K_T^V = -2\rho^* K_T$ . Appliquons une nouvelle fois le lemme d'Itô aux quantités  $\Lambda_T$ ,  $P_T$  et  $S_T$ . Ainsi, pour  $0 \leq t \leq T$ , on a

$$\begin{aligned} \Lambda_t &= \frac{1}{2\rho} V_t^2 - \frac{1}{\rho} M_t^V - \frac{t}{2\rho}, \\ P_t &= \frac{1}{\theta^*} Y_t V_t - \frac{1}{2\theta^*} V_t^2 - \frac{1}{\theta^*} M_t^Y - \frac{t}{2\theta^*}, \\ S_t &= \frac{1}{2\theta} Y_t^2 + \frac{\rho}{2\theta^*\theta} V_t^2 - \frac{\rho}{\theta^*\theta} Y_t V_t - \frac{1}{\theta^*} M_t^Y - \frac{t}{2\theta^*}. \end{aligned}$$

Il s'ensuit que

$$\widehat{\Lambda}_T (\widehat{\rho}_T - \rho^*) = C_Y M_T^Y + C_V M_T^V + \frac{J_T^V}{2} (\widehat{\theta}_T - \theta^*) + R_T^V \quad (\text{V.39})$$

où l'on a utilisé les constantes

$$C_V = \frac{(\theta^*)^2 \rho^*}{\theta^2 \rho} \quad \text{et} \quad C_Y = -\frac{\rho(2\theta + \rho)\rho^*}{\theta^2 \theta^*}.$$

Nous n'avons pas explicité le terme résiduel  $R_T^V$  par souci de clarté mais ce dernier, à l'instar de  $R_T^Y$  apparaissant dans la décomposition (V.25), est bien négligeable devant  $\sqrt{T}$ . En combinant cette décomposition avec celle donnée par (V.25), on obtient

$$\sqrt{T} \begin{pmatrix} \widehat{\theta}_T - \theta^* \\ \widehat{\rho}_T - \rho^* \end{pmatrix} = \frac{1}{\sqrt{T}} A_T Z_T + \sqrt{T} R_T \quad (\text{V.40})$$

où

$$A_T = \begin{pmatrix} T S_T^{-1} & 0 \\ T B_T \widehat{\Lambda}_T^{-1} & T C_V \widehat{\Lambda}_T^{-1} \end{pmatrix}, \quad Z_T = \begin{pmatrix} M_T^Y \\ M_T^V \end{pmatrix} \quad \text{et} \quad R_T = \begin{pmatrix} S_T^{-1} R_T^Y \\ \widehat{\Lambda}_T^{-1} D_T \end{pmatrix}$$

avec  $B_T = C_X + J_T^V (2S_T)^{-1}$  et  $D_T = R_T^V + J_T^V (2S_T)^{-1} R_T^Y$ . La suite  $(Z_t)$  forme bien entendu une  $(\mathcal{F}_t)$ -martingale vectorielle continue dont le processus croissant est donné par

$$\langle Z \rangle_t = \begin{pmatrix} S_t & P_t \\ P_t & \Lambda_t \end{pmatrix}. \quad (\text{V.41})$$

À partir des Corollaires V.2–V.3, on trouve immédiatement que

$$\lim_{T \rightarrow \infty} A_T = A = \begin{pmatrix} -2\theta^* & 0 \\ -2\rho^*(C_X - 2(\theta\rho)^{-1}\theta^*\rho^*) & -2\rho^*C_V \end{pmatrix} \quad \text{p.s.} \quad (\text{V.42})$$

de même que

$$\lim_{T \rightarrow \infty} \frac{\langle Z \rangle_T}{T} = \Delta = -\frac{1}{2\theta^*} \begin{pmatrix} 1 & 1 \\ 1 & \theta^*\rho^{-1} \end{pmatrix} \quad \text{p.s.} \quad (\text{V.43})$$

Enfin, par le Corollaire V.4,

$$\sqrt{T} R_T \xrightarrow{\mathbb{P}} 0. \quad (\text{V.44})$$

La combinaison de (V.40), (V.42), (V.43), (V.44), du lemme de Slutsky et du théorème central limite pour les martingales vectorielles continues [54] nous conduit finalement à

$$\sqrt{T} \begin{pmatrix} \widehat{\theta}_T - \theta^* \\ \widehat{\rho}_T - \rho^* \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0, A\Delta A'),$$

ce qui achève la première partie de la preuve, par calcul direct de la matrice de covariance. Considérons désormais le cadre instable où  $\rho = 0$ . Reprenons alors la stratégie utilisée dans la Section 2 de [55], axée sur l'autosimilarité du mouvement brownien. Ainsi,

$$\begin{aligned} \left( \int_0^T W_t^2 dt, \frac{1}{2}(W_T^2 - T) \right) &\stackrel{\mathcal{L}}{=} \left( T \int_0^T W_{t/T}^2 dt, \frac{T}{2}(W_1^2 - 1) \right), \\ &= \left( T^2 \int_0^1 W_s^2 ds, \frac{T}{2}(W_1^2 - 1) \right). \end{aligned} \quad (\text{V.45})$$

Or, nous avons vu avec (V.38) que, lorsque  $\rho = 0$ , alors  $\widehat{\Lambda}_T = \Lambda_T(1 + o(1))$  p.s. On remarque de plus que  $\widehat{V}_T^2/T$  et  $W_T^2/T$  partagent la même distribution asymptotique. La distribution jointe (V.45) ainsi que la remarque précédente permettent donc d'achever la preuve du théorème. ★



---

## *Stationnarité ou multi-intégration ?*

---

Voici la reformulation détaillée de l'article [113], soumis et en attente de publication. Nous abordons dans ce dernier chapitre la problématique de la stationnarité d'une réalisation d'un processus autorégressif, que l'on munit d'une tendance déterministe polynomiale.

### **VI.1 Historique et motivation**

Au tout début de notre étude, dans la Section I.3, nous avons présenté de manière succincte l'effet d'une non stationnarité de type déterministe ou stochastique sur l'évolution d'un processus autorégressif, ainsi que les techniques usuelles (principalement la différenciation) pour s'en affranchir et travailler ainsi avec un processus stationnaire. Seulement, comment détecter en amont la présence d'une non stationnarité de type stochastique ? Nous allons tout d'abord réaliser un historique rapide des deux approches complémentaires développées à cet égard.

#### **VI.1.1 Tester la présence d'une racine unitaire**

Considérons la série intégrée sur  $\mathbb{N}$  s'écrivant, sous forme condensée,

$$(1 - \theta_0 L) \mathcal{A}^*(L) Z_t = \mathcal{B}(L) \varepsilon_t \tag{VI.1}$$

où le polynôme  $\mathcal{A}^*$  d'ordre  $p$  est causal, le polynôme  $\mathcal{B}$  d'ordre  $q$  est causal et  $|\theta_0| = 1$ . Le polynôme défini par  $\mathcal{A}(z) = (1 - \theta_0 z) \mathcal{A}^*(z)$  n'est pas inversible sur  $\mathbb{C}$  et il n'existe en conséquence pas de solution stationnaire à cette modélisation. Si maintenant nous complexifions le modèle en y adjoignant une tendance déterministe  $(T_t)$  – puisque ce sera le cadre d'étude de ce chapitre – nous obtenons le

processus générateur

$$Y_t = T_t + Z_t \quad (\text{VI.2})$$

où  $(Z_t)$  est engendré par (VI.1). Nous retrouvons ici l'écriture traditionnelle de la série chronologique décomposée en la somme de sa composante déterministe et de sa composante stochastique. Nous voyons alors que la série  $((1 - \theta_0 L)Y_t)$  est stationnaire autour d'une tendance déterministe, dont les paramètres peuvent être estimés conjointement. Ils interviendront en conséquence dans la distribution asymptotique de l'estimateur de  $\theta_0$ . Voyons quelques exemples plus en détail.

### VI.1.1.1 Le test de Dickey-Fuller

Le cas particulier où  $p = 0, q = 0$  et où la tendance déterministe du modèle différencié est linéaire, fut le premier étudié par Dickey et Fuller [39] en 1979, en raison de sa simplicité et de sa capacité à permettre d'appréhender la problématique de la racine unitaire dans le processus autorégressif. Pour  $1 \leq t \leq T$ , le modèle est donné par

$$Y_t = \theta_0 Y_{t-1} + \alpha + \beta t + \varepsilon_t \quad (\text{VI.3})$$

avec  $(\varepsilon_t)$  formant un bruit blanc de variance finie. Il s'ensuit donc qu'un test de significativité du premier coefficient correctement recentré de la régression linéaire de  $(Y_t)$  sur le sous-espace engendré par  $\{Y_{t-1}, 1, t\}$  permet de tester l'hypothèse  $\mathcal{H}_0 : \theta_0 = 1$  contre son alternative  $\mathcal{H}_1 : \theta_0 \neq 1$ . Il est alors montré en particulier que, sous  $\mathcal{H}_0$  et lorsque les résidus sont gaussiens,

$$T(\widehat{\theta}_T - 1) \xrightarrow{\mathcal{L}} \frac{\int_0^1 \bar{W}(s) dW(s)}{\int_0^1 \bar{W}^2(s) ds} \quad (\text{VI.4})$$

où  $\widehat{\theta}_T$  est l'estimateur des moindres carrés du coefficient  $\theta_0$  sur la trajectoire  $(Y_t)$  restreinte à  $\{0, \dots, T\}$ , et  $\bar{W}(t)$  est un processus stochastique identifiable, dépendant de l'ordre de la tendance polynomiale, cette représentation n'étant dans un premier temps donnée que sous la forme d'une série infinie. À titre d'exemple,  $\bar{W}(t)$  est un mouvement brownien standard  $W(t)$  lorsque  $\alpha = \beta = 0$ . Ce résultat avait déjà été conjecturé par White [132] en 1958 (malgré une petite erreur sur le coefficient multiplicatif) dans un modèle gaussien sans tendance, et fut amélioré en 1988 par Chan et Wei [31] qui considèrent que  $(\varepsilon_t)$  est une différence de martingale dont le moment conditionnel d'ordre  $2 + \delta$  est fini, pour une valeur  $\delta > 0$ . Notons également qu'un cas particulier avait été étudié dès 1983 par Lai et Siegmund [74] alors que la première preuve complète de (VI.4) semble devoir revenir à Phillips [107] en 1987. Lorsque l'ordre augmente ( $\alpha \neq 0$  et/ou  $\beta \neq 0$ ),  $\bar{W}(t)$  décrit une famille de processus de Wiener recentrés que nous serons amenés à préciser dans la section suivante. La distribution asymptotique associée à la  $t$ -statistique a été tabulée par Dickey et Fuller [39] en 1979, ou encore par MacKinnon [86] en 1991 (il s'agit de (VI.4) avec une racine carrée au dénominateur). En cas de non rejet de  $\mathcal{H}_0$ , la

série différenciée ( $\Delta Y_t$ ) est stationnaire autour de sa tendance ( $T_t$ ), c'est ce que l'on espère tacitement lorsqu'en pratique, nous différencions les processus dans le but de les stationnariser. Nous voyons ici que ce test n'est applicable que lorsque  $\theta_0 = 1$ , mais ne tient pas compte de la non stationnarité stochastique associée au cas  $\theta_0 = -1$ . En outre,  $\alpha$  et  $\beta$  jouant un rôle décisif dans la distribution asymptotique, il importe de tester plusieurs modèles en faisant varier l'ordre de la tendance polynomiale, à travers ce que l'on nomme la *stratégie de test de Dickey-Fuller*. En guise qu'application pratique, nous citerons l'étude de Nelson et Plosser [93] de 1982 qui met en évidence la présence de racines unitaires dans nombre de séries macroéconomiques à l'aide du test de Dickey-Fuller.

### VI.1.1.2 Le test de Dickey-Fuller augmenté

Le test augmenté, que l'on note communément « test ADF » pour *augmented Dickey-Fuller*, est peut-être le test le plus utilisé pour évaluer l'hypothèse nulle de racine unitaire dans un processus ARMA( $p, q$ ). Ce dernier, originellement proposé par Dickey et Fuller [39]–[40] dans un cadre autorégressif, est finalement décrit par Dickey et Said [41] en 1981 dans le cas où  $p$  et  $q$  sont supposés connus, puis dans le cas plus général [116] en 1984 sous l'approximation AR( $k$ ) du modèle AR( $\infty$ ) associé, pour  $k$  grandissant à l'infini, la connaissance de  $p$  et  $q$  étant dès lors superflue. Il généralise la procédure de Dickey-Fuller au processus ARMA d'ordre quelconque. L'hypothèse fondamentale est que les polynômes  $\mathcal{A}^*$  et  $\mathcal{B}$  dans l'expression générale (VI.1) sont causaux et donc, en particulier, que l'écriture AR( $\infty$ ) existe. On utilise alors la représentation de Sims, Stock et Watson [123] qui stipule que l'on peut exprimer tout processus AR( $k$ ) de paramètre  $\theta = (\theta_1 \dots \theta_k)'$  comme

$$\Delta Y_t = (\theta_0 - 1)Y_{t-1} + \sum_{i=1}^{k-1} \delta_i \Delta Y_{t-i} + \alpha + \beta t + \varepsilon_t \quad (\text{VI.5})$$

où  $\delta_i = -(\theta_{i+1} + \dots + \theta_k)$ ,  $\theta_0 = \theta_1 + \dots + \theta_k$ , et avec  $\alpha$  et  $\beta$  éventuellement nuls. Dès lors, une régression linéaire de ( $\Delta Y_t$ ) sur le sous-espace engendré par  $\{Y_{t-1}, \Delta Y_{t-1}, \dots, \Delta Y_{t-k+1}, 1, t\}$  permet de tester l'hypothèse  $\mathcal{H}_0 : \theta_0 = 1$  contre son alternative  $\mathcal{H}_1 : \theta_0 \neq 1$ . Ce qu'il y a ici de remarquable, c'est que, indépendamment du nombre de retards ajoutés au modèle, la distribution limite de la  $t$ -statistique formée à partir de (VI.4) reste valable, adaptée bien sûr à la tendance polynomiale retenue, pour le processus AR( $p$ ) avec  $p = k$ , et pour le processus ARMA( $p, q$ ) dès que  $k = O(T^{1/3})$ , comme il est montré dans [116]. Cependant, la puissance du test s'en trouve impactée et l'on pourra consulter en 1989 avec Schwert [120], puis en 1995 avec Ng et Perron [97], certaines remarques pertinentes sur les distorsions susceptibles d'intervenir en cas de mauvais choix de troncature. Dès lors, une stratégie de test de Dickey-Fuller peut de nouveau être mise en œuvre, sans connaissance préalable du couple  $(p, q)$ , afin d'évaluer la présence d'une racine unitaire  $\theta_0 = 1$  dans le processus ainsi qu'une éventuelle tendance polynomiale. Là encore, notons que la localisation d'une racine à  $\theta_0 = -1$  n'est pas détectée par la procédure.

### VI.1.1.3 Le test de Phillips-Perron

Phillips [107] en 1987, puis Phillips et Perron [108] en 1988, étudient le modèle sans retard supplémentaire (VI.3), ainsi que la procédure de Dickey-Fuller, mais autorisent la perturbation à se comporter comme un processus très général vérifiant seulement certaines conditions de dépendance faible. Ces dernières, que l'on nomme communément *conditions de Phillips-Perron*, stipulent dans un premier temps que le processus perturbateur  $(\varepsilon_t)$  vérifie, pour tout  $1 \leq t \leq T$ ,  $\mathbb{E}[\varepsilon_t] = 0$  et  $\sup \mathbb{E}[|\varepsilon_t|^\beta] < +\infty$  pour au moins une valeur de  $\beta > 2$ . De plus, pour le *processus des sommes partielles*  $S_t = \varepsilon_1 + \dots + \varepsilon_t$ , la convergence

$$\lim_{T \rightarrow \infty} \frac{\mathbb{E}[S_T^2]}{T} = \omega^2 \quad (\text{VI.6})$$

est supposée vérifiée, la valeur  $\omega^2 > 0$  étant appelée *variance de long terme* du processus. Enfin, le processus  $(\varepsilon_t)$  doit être fortement mélangeant de coefficients  $(\alpha_k)$  satisfaisant

$$\sum_{k=1}^{\infty} \alpha_k^{1-2/\beta} < +\infty.$$

On pourra consulter [59] pour la définition du mélange fort et de ses coefficients. Comme le montre Withers [133] en 1981, ces conditions sont satisfaites en particulier dans le cas des perturbations ARMA( $p, q$ ) usuelles, ce qui est suffisant dans le cadre de notre étude. La motivation du test introduit par Phillips et Perron est que, dans le cas où la perturbation  $(\varepsilon_t)$  n'est plus indépendante et identiquement distribuée mais qu'elle possède une variance de long terme  $\omega^2$  et une variance locale  $\sigma^2$  (que l'on peut voir comme la variance du processus stationnaire), alors les statistiques de tests (A)DF dépendent de  $\omega^2$  et de  $\sigma^2$ . La correction proposée est de fait non paramétrique, comme nous allons le voir. Soit  $(\widehat{\varepsilon}_t)$  l'ensemble résiduel engendré par l'estimation du modèle (VI.3), à adapter si l'on ne souhaite retenir aucune tendance linéaire ou seulement constante. Alors, l'estimateur

$$\widehat{\omega}_T^2 = \frac{1}{T} \sum_{t=1}^T \widehat{\varepsilon}_t^2 + \frac{2}{T} \sum_{s=1}^{\ell} w(s, \ell) \sum_{t=s+1}^T \widehat{\varepsilon}_t \widehat{\varepsilon}_{t-s} \quad \text{avec} \quad w(s, \ell) = 1 - \frac{s}{1 + \ell} \quad (\text{VI.7})$$

a été proposé en 1987 par Newey et West [96], le noyau  $w$  est une fenêtre triangulaire (de Bartlett), et  $\ell$  est un paramètre de troncature. Phillips [107] montre en 1987 que  $\widehat{\omega}_T^2$  est faiblement consistant dès que  $\sup \mathbb{E}[|\varepsilon_t|^\beta] < +\infty$  pour au moins une valeur de  $\beta > 4$  (ce qui durcit quelque peu les conditions de Phillips-Perron), que  $\ell = \ell(T) \rightarrow +\infty$ , et que  $\ell^4(T) = o(T)$ , le tout pour  $w(s, t) = 1$ . L'utilisation de la fenêtre triangulaire ne change en rien ces résultats. Phillips et Perron [108] établissent alors que

$$T(\widehat{\theta}_T - 1) - \frac{\widehat{\omega}_T^2 - \widehat{\sigma}_T^2}{2M_{T,Y}} \xrightarrow{\mathcal{L}} \frac{\int_0^1 \bar{W}(s) dW(s)}{\int_0^1 \bar{W}^2(s) ds} \quad (\text{VI.8})$$

où  $\widehat{\sigma}_T^2$  est l'estimateur usuel de la variance de  $(\varepsilon_t)$  et  $M_{T,Y}$  est une fonction des observations  $(Y_t)$  différemment construite selon que l'on retient une tendance linéaire, constante ou nulle dans le processus générateur. Quant à la distribution limite (tabulée pour diverses tendances par Ouliaris et Phillips [102] en 1994), elle repose sur le processus stochastique  $\bar{W}(t)$  déjà introduit dans le paragraphe sur Dickey-Fuller, que nous définirons explicitement dans ce chapitre. Ce résultat de convergence fut au préalable établi dans [107] lorsqu'aucune tendance n'est retenue au cours de l'estimation. Diverses améliorations mineures ont pu être apportées concernant l'estimation de  $\omega^2$ , on pourra par exemple consulter Perron et Ng [106] en 1996. La correction non paramétrique apportée par Phillips et Perron est au test ADF ce que le test KPSS est à la procédure de Leybourne et McCabe, comme nous le comprendrons mieux dans la section suivante.

Nous avons brièvement résumé les trois procédures communément utilisées pour tester la présence d'une racine unitaire dans un processus autorégressif, même s'il existe une copieuse littérature et énormément d'évolutions apportées encore aujourd'hui aux tests de racine unitaire. On pourra se référer par exemple à Dickey, Bell et Miller [38] et Bhargava [14] en 1986, Perron [105] en 1988, Ouliaris, Park et Phillips [101] en 1989, Dolado, Jenkinson et Sosvilla-Rivero [44] en 1990, Schmidt et Phillips [118] en 1992, Leybourne, Kim et Newbold [79]–[80] en 2004–05, etc. Citons également les approches bayésiennes de Sims [122] en 1988 et de Lubrano [85] en 1995. Nous renvoyons enfin le lecteur au tour d'horizon très détaillé de Phillips et Xiao [109] de 1999. Ces outils nous permettent de rejeter une hypothèse nulle de non stationnarité d'un processus, mais, selon la stratégie usuelle des tests d'hypothèse, ils ne nous permettent pas d'accepter à proprement parler la non stationnarité, le cas échéant. C'est pourquoi il est d'intérêt pratique de disposer de procédures complémentaires pour lesquelles la stationnarité forme l'hypothèse nulle. Cela semble d'autant plus important que De Jong, Nankervis, Savin et Whiteman [33] constatent en 1992 que les tests de racine unitaire sont en pratique moins puissants que les tests usuels dans un cadre de stabilité où l'une des racines est censée se rapprocher de 1, point sur lequel insisteront beaucoup d'économètres arguant que la présence d'une racine unitaire à l'origine d'une trajectoire observée relève d'une probabilité nulle. Dans cette veine, Chan et Wei [30] en 1987 ont étudié la distribution asymptotique de l'estimateur du paramètre de l'autorégression dans un modèle AR(1) sans tendance, lorsque l'on suppose que ce dernier reste inférieur à 1 mais s'en approche rapidement. Ils établissent alors une distribution qui, effectivement, est légèrement différente.

## VI.1.2 Tester la stationnarité

Nous allons désormais faire intervenir la non stationnarité non plus dans l'autorégression du processus, mais dans le comportement résiduel. Soit alors, pour  $p \leq t \leq T$ , le processus autorégressif

$$\mathcal{A}(L)Y_t = T_t + S_t^\eta + \varepsilon_t \quad (\text{VI.9})$$

où le polynôme  $\mathcal{A}$  d'ordre  $p$  est causal,  $(T_t)$  est une tendance déterministe,  $(\varepsilon_t)$  un bruit blanc de variance  $\sigma_\varepsilon^2 > 0$  et  $(S_t^\eta)$  une marche aléatoire engendrée par un bruit blanc  $(\eta_t)$ , indépendant de  $(\varepsilon_t)$ , et de variance  $\sigma_\eta^2 \geq 0$ .

### VI.1.2.1 Le test de Leybourne-McCabe

En 1994, Leybourne et McCabe [82] montrent que, sous l'hypothèse nulle  $\mathcal{H}_0 : \sigma_\eta^2 = 0$ , le processus  $(Y_t)$  engendré par (VI.9) se comporte comme un AR( $p$ ) stationnaire autour de la tendance  $(T_t)$  alors que, sous l'alternative  $\mathcal{H}_1 : \sigma_\eta^2 > 0$ , il forme un ARIMA( $p, 1, 1$ ) inversible avec tendance, *a fortiori* non stationnaire. Reste alors à trouver une statistique dont le comportement asymptotique permette de discriminer  $\mathcal{H}_0$  et  $\mathcal{H}_1$ . La stratégie qu'ils proposent consiste à estimer  $\theta \in \mathbb{R}^p$  par maximum de vraisemblance après avoir correctement éliminé la tendance, puis à estimer les paramètres de cette tendance par moindres carrés sur le processus  $(\tilde{\mathcal{A}}(L)Y_t)$  obtenu. Nous disposons ainsi d'un ensemble résiduel  $(\hat{\varepsilon}_t)$  et il est établi que, sous  $\mathcal{H}_0 : \sigma_\eta^2 = 0$ ,

$$\frac{1}{T Q_T} \sum_{t=1}^T S_t^2 \xrightarrow{\mathcal{L}} \int_0^1 \bar{B}^2(s) ds \quad (\text{VI.10})$$

où  $(S_t)$  et  $(Q_t)$  sont les processus des sommes partielles respectivement associés à  $(\hat{\varepsilon}_t)$  et à  $(\hat{\varepsilon}_t^2)$ . Sous  $\mathcal{H}_1 : \sigma_\eta^2 > 0$ , la statistique diverge à la vitesse  $T$  et l'on peut obtenir sa distribution asymptotique convenablement renormalisée. Ici,  $\bar{B}(t)$  décrit une famille de ponts browniens dont l'ordre dépend de la tendance retenue, que l'on sera également amené à formaliser dans la suite de l'étude. Nous ferons désormais référence à cette procédure de test à travers la dénomination de « test LMC » et si nous restons volontairement évasifs quant au mécanisme sous-jacent, c'est qu'il est prévu que nous le reprenions en le généralisant dans les prochaines sections.

Sous l'hypothèse simplificatrice où  $p = 0$ , nous retrouvons les fondations de cette stratégie de test chez Nabeya et Tanaka [92] dès 1988, mais cette restriction semble trop éloignée de la réalité puisque tout phénomène d'autocorrélation a disparu. Antérieurement, Nyblom et Makelainen [99] en 1983, Nyblom [98] en 1986 et Leybourne et McCabe [81] en 1989 s'étaient déjà intéressés à de telles statistiques de test pour des modèles voisins. En 1993, Saikkonen et Luukkonen [117] adoptent un point de vue symétrique et préfèrent tester la présence d'une composante MA non inversible sous  $\mathcal{H}_0$  dans le processus différencié, toujours pour  $p = 0$ .

### VI.1.2.2 Le test de Kwiatkowski-Phillips-Schmidt-Shin

La procédure de Kwiatkowski, Phillips, Schmidt et Shin [73], datée de 1992 et plus communément abrégée en « test KPSS », repose également sur le modèle (VI.9) restreint à  $p = 0$ . Nous disions un peu plus tôt que *le test de Phillips-Perron est au test ADF ce que le test KPSS est à la procédure de*

Leybourne et McCabe, cette comparaison tient au fait que l'autocorrélation du processus n'est plus spécifiée dans le modèle, mais qu'elle est translatée dans le résidu ( $\varepsilon_t$ ) qui n'est plus désormais un bruit blanc, mais évolue librement dans les limites des conditions de Phillips-Perron (voir Section VI.1.1.3). Ils obtiennent alors le même type de convergence que (VI.10), à ceci près que la renormalisation se fait maintenant par l'intermédiaire de l'estimateur de la variance de long terme (VI.6). Il est alors montré que, sous  $\mathcal{H}_1 : \sigma_\eta^2 > 0$ , la statistique de test diverge à la vitesse  $T/\ell$ . Mais rappelons-nous que la consistance de  $\widehat{\omega}_T^2$  en (VI.7) implique que  $\ell = \ell(T) \rightarrow +\infty$  comme nous l'avons vu précédemment, et qu'ainsi  $T/\ell(T) = o(T)$ . Cette conclusion rend le test KPSS *a priori* moins puissant que le test LMC, puisque ce dernier rejette plus rapidement la statistique de test à l'infini le cas échéant. De telles constatations sont faites empiriquement dans [82].

La stationnarité dans les séries chronologiques linéaires étant un domaine de recherche contemporain et d'importance capitale, il n'est pas étonnant de trouver une littérature abondante quant aux études empiriques, aux détections d'anomalies ou aux améliorations apportées en particulier au test KPSS. De manière non exhaustive, citons simplement Leybourne et McCabe [83] en 1999, Newbold, Leybourne et Wohar [95] en 2001, Müller [91] en 2005, Harris, Leybourne et McCabe [61] ou encore De Jong, Amsler et Schmidt [34] en 2007, Pelagatti et Sen [104] en 2009, etc. De notre côté, nous nous tournons vers le test LMC, que nous nous proposons de généraliser sous plusieurs aspects. Nous montrerons dans un premier temps qu'il est possible d'obtenir des convergences similaires sous  $\mathcal{H}_0$  comme sous  $\mathcal{H}_1$ , et ce pour une tendance polynomiale d'ordre  $r$  quelconque et un bruit éventuellement multi-intégré d'ordre  $d$ . Cela nous donnera accès à de multiples procédures de test de stationnarité et de racine unitaire (grâce au comportement sous  $\mathcal{H}_1$ ). Nous expliquerons pourquoi, dans un second temps, le test LMC – et par extension le test KPSS – ne rejette pas l'hypothèse de stationnarité, à tort, lorsque la marche aléatoire est engendrée par une racine unitaire égale à  $-1$ . Nous nous inspirerons très largement des techniques de calculs utilisées dans [73] et [82], reposant elles-mêmes sur deux outils fondamentaux : le principe d'invariance de Donsker et le théorème de continuité de Mann-Wald (on se reportera aux Théorèmes B.7–B.9 pour plus de détails).

## VI.2 Tester la stationnarité contre l'intégration

Considérons une modélisation autorégressive d'ordre  $p$ , indexée par  $\mathbb{Z}$ , munie d'une tendance polynomiale d'ordre  $r$  et perturbée par une marche aléatoire et par un bruit additif. Supposons que l'on observe une trajectoire ( $Y_t$ ) engendrée par ce modèle, sur  $t \in \{0, \dots, T\}$ . Pour  $p \leq t \leq T$ , on a

$$\mathcal{A}(L)Y_t = (\alpha_0 + \alpha_1 t_T + \dots + \alpha_r t_T^r) \mathbb{I}_{\{\kappa \neq 0\}} + S_t^\eta + \varepsilon_t \quad (\text{VI.11})$$

où, pour tout  $z \in \mathbb{C}$ ,  $\mathcal{A}(z) = 1 - \theta_1 z - \dots - \theta_p z^p$  est un polynôme causal, où pour  $|\rho| = 1$  réel,

$$S_t^\eta = \rho S_{t-1}^\eta + \eta_t \quad (\text{VI.12})$$

est une marche aléatoire démarrant à  $S_0^\eta = 0$ , où enfin  $(\varepsilon_t)$  et  $(\eta_t)$  sont des bruits blancs mutuellement indépendants de variances respectives  $\sigma_\varepsilon^2 \geq 0$  et  $\sigma_\eta^2 \geq 0$ , non simultanément nulles. Pour simplifier le traitement des projections lors de la procédure d'estimation par moindres carrés, nous renormalisons la tendance et nous notons  $t_T = t/T$ , par convention. Nous retrouvons ainsi la tendance linéaire pour  $r = 1$ , la tendance constante pour  $r = 0$  et l'indicateur  $\kappa$  nous permettra de bifurquer d'une modélisation avec tendance déterministe ( $\kappa \neq 0$ ) à une modélisation sans tendance ( $\kappa = 0$ ), car alors les résultats de convergence sont différents, même lorsque  $r = 0$ . Notre objectif est de tester

$$\mathcal{H}_0 : \text{“}\sigma_\eta^2 = 0\text{”} \quad \text{vs} \quad \mathcal{H}_1 : \text{“}\sigma_\eta^2 > 0\text{”}$$

puis de renforcer l'étude par le test complémentaire. Sous  $\mathcal{H}_0$ , il est clair que la marche  $(S_t^\eta)$  est presque sûrement nulle, et qu'ainsi évaluer  $\mathcal{H}_0$  contre  $\mathcal{H}_1$  revient à tester la stationnarité du processus autour d'une tendance déterministe contre son alternative non stationnaire dans laquelle le processus est intégré. Pour  $r = \{0, 1\}$  et  $\rho = 1$ , nous retrouvons la procédure LMC et par un raisonnement équivalent, nous allons l'étendre au spectre  $r \geq 0$  et  $|\rho| = 1$ . Notons enfin que le domaine  $|\rho| < 1$  engendre la stationnarité en tendance du processus sous  $\mathcal{H}_0$  comme sous  $\mathcal{H}_1$ , cette alternative est en conséquence de peu d'intérêt dans le cadre de notre étude, et nous avons pu voir qu'une vaste littérature existe déjà sur le sujet.

### VI.2.1 La procédure d'estimation

Tout d'abord, il est important de noter que l'on considère ici une valeur connue de  $r$  d'une part, et que l'on a, d'autre part, de bonnes indications empiriques sur la valeur de  $\rho$  à retenir. En effet, nous avons représenté sur la Figure VI.1 des trajectoires simulées à partir du modèle (VI.11) muni de la configuration  $T = 1000$ ,  $p = 1$ ,  $\theta = 0.5$ ,  $r = 2$ ,  $\alpha = (-2, 100)$ ,  $(\varepsilon_t) \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 0.5)$ ,  $(\eta_t) \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$  et respectivement  $\rho = 1$  puis  $\rho = -1$ . On observe alors clairement que le comportement engendré est totalement différent selon la localisation de la racine unitaire, où l'on devine la présence d'une marche aléatoire avec dérive linéaire dans un cas, alors que dans l'autre il y a de plus une tendance à l'alternance prononcée. Cela explique pourquoi nous considérons que la connaissance préalable de la valeur de  $\rho$  à retenir sous  $\mathcal{H}_1$  n'est pas très restrictive, alors qu'elle est primordiale pour la bonne estimation de  $\theta$  comme nous allons le voir.

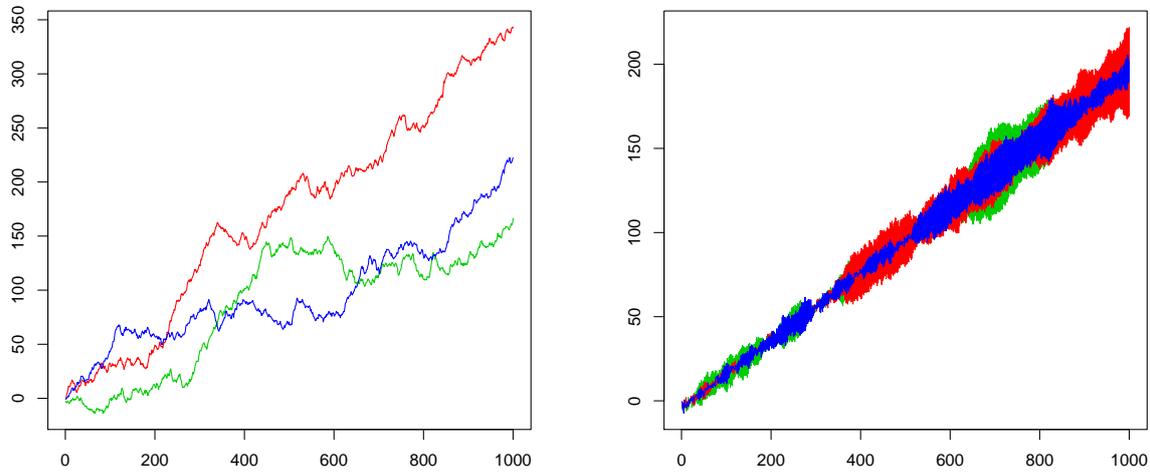


FIGURE VI.1 – Exemples de trajectoires non stationnaires simulées avec  $\rho = 1$  (gauche) puis avec  $\rho = -1$  (droite).

La non stationnarité du processus  $(Y_t)$  se situe dans le terme résiduel

$$S_t^\eta = \sum_{k=1}^t \rho^{t-k} \eta_k \quad (\text{VI.13})$$

qui forme le processus des sommes partielles associé à  $(\eta_t)$  lorsque  $\rho = 1$ . Ainsi, la stationnarité en tendance du processus est atteinte par l'intermédiaire de la différenciation

$$\begin{aligned} \mathcal{A}(L)(1 - \rho L)Y_t &= (1 - \rho L)(\alpha_0 + \alpha_1 t_T + \dots + \alpha_r t_T^r) \mathbb{I}_{\{r \neq 0\}} + (1 - \rho L)(S_t^\eta + \varepsilon_t) \\ &= (\gamma_0 + \gamma_1 t_T + \dots + \gamma_r t_T^r) \mathbb{I}_{\{r \neq 0\}} + \eta_t + \varepsilon_t - \rho \varepsilon_{t-1} \end{aligned}$$

où les constantes  $\gamma_0, \gamma_1, \dots, \gamma_r$  sont facilement identifiables ( $\gamma_r = 0$  pour  $\rho = 1$ ), et  $(\eta_t + \varepsilon_t - \rho \varepsilon_{t-1})$  est un résidu correspondant à un processus MA(1), ainsi qu'il est décrit dans [73]. En effet, il est immédiat de voir que

$$\mathbb{V}(\eta_t + \varepsilon_t - \rho \varepsilon_{t-1}) = \sigma_\eta^2 + (1 + \rho^2)\sigma_\varepsilon^2, \quad \text{Cov}(\eta_t + \varepsilon_t - \rho \varepsilon_{t-1}, \eta_{t-1} + \varepsilon_{t-1} - \rho \varepsilon_{t-2}) = -\rho \sigma_\varepsilon^2$$

et que la covariance est nulle pour tout décalage supérieur à 1. On construit dès lors un bruit blanc  $(\xi_t)$ , de variance  $\sigma_\xi^2 \geq 0$  tel que, pour tout  $p \leq t \leq T$ ,

$$\mathcal{A}(L)(1 - \rho L)Y_t = (\gamma_0 + \gamma_1 t_T + \dots + \gamma_r t_T^r) \mathbb{I}_{\{r \neq 0\}} + \xi_t + \beta \xi_{t-1} \quad (\text{VI.14})$$

et l'ACF de  $(\xi_t)$  nous conduit, après quelques identifications, à

$$\sigma_\xi^2 = \frac{-\rho\sigma_\varepsilon^2}{\beta} \quad \text{et} \quad \beta = \frac{-(\omega + 2) \pm \sqrt{\omega(\omega + 4)}}{2\rho}$$

où  $\omega = \sigma_\eta^2/\sigma_\varepsilon^2$  caractérise le ratio signal/bruit, et la valeur de  $\beta$  choisie est celle garantissant à la fois que  $\rho\beta < 0$  et que  $|\beta| \leq 1$ . Les solutions en question sont donc

$$\beta^- = \frac{(\omega + 2) - \sqrt{\omega(\omega + 4)}}{2} \quad \text{et} \quad \beta^+ = \frac{-(\omega + 2) + \sqrt{\omega(\omega + 4)}}{2}$$

où l'exposant de  $\beta$  symbolise le signe de  $\rho$ . Ainsi, tester  $\mathcal{H}_0 : " \sigma_\eta^2 = 0 "$  contre  $\mathcal{H}_1 : " \sigma_\eta^2 \neq 0 "$  revient, comme nous pouvons le voir, à choisir entre un processus AR( $p$ ) causal stationnaire en tendance que l'on aurait surdifférencié, et un processus ARIMA( $p, 1, 1$ ) causal inversible et non stationnaire (pour  $\rho = 1$ , sinon il ne s'agit pas d'un processus ARIMA au sens strict du terme, mais d'un équivalent avec une sommation en lieu et place de la différenciation). Soit alors  $(\tilde{\theta}_T, \tilde{\beta}_T)$  l'estimateur du maximum de vraisemblance de  $(\theta, \beta)$  dans le modèle (VI.14) dont la tendance a été correctement éliminée (par exemple par différenciation). Pour tout  $p \leq t \leq T$ , on considère le processus résiduel

$$\tilde{Y}_t = Y_t - \tilde{\theta}_1 Y_{t-1} - \dots - \tilde{\theta}_p Y_{t-p}. \quad (\text{VI.15})$$

Sous  $\mathcal{H}_1$ , on a  $|\beta| < 1$  ce qui justifie bien la propriété d'inversibilité de la modélisation ARIMA. En revanche, sous  $\mathcal{H}_0$ , on a  $\omega = 0$  et donc  $|\beta| = 1$ . Cependant, l'estimateur  $(\tilde{\theta}_T, \tilde{\beta}_T)$  reste fortement consistant, comme nous le verrons dans la preuve de nos résultats. Il est en conséquence tout à fait cohérent d'estimer  $\alpha$  par une stratégie des moindres carrés dans le modèle donné par

$$\tilde{Y}_t = (\alpha_0 + \alpha_1 t_T + \dots + \alpha_r t_T^r) \mathbb{I}_{\{\kappa \neq 0\}} + \varepsilon_t \quad (\text{VI.16})$$

et de bâtir ainsi l'ensemble résiduel, pour  $p \leq t \leq T$ ,

$$\hat{\varepsilon}_t = \tilde{Y}_t - (\hat{\alpha}_0 + \hat{\alpha}_1 t_T + \dots + \hat{\alpha}_r t_T^r) \mathbb{I}_{\{\kappa \neq 0\}} \quad (\text{VI.17})$$

où  $\hat{\alpha}_T$  est donc l'estimateur des moindres carrés du paramètre  $\alpha$  dans le modèle (VI.11) vu sous  $\mathcal{H}_0$ . Nous considérons arbitrairement que  $\hat{\varepsilon}_1 = \dots = \hat{\varepsilon}_{p-1} = 0$ . Il va de soi que  $\alpha$  n'a pas à être estimé lorsque  $\kappa = 0$ . Définissons alors  $(S_t)$  et  $(Q_t)$ , les processus des sommes partielles respectivement associés à  $(\hat{\varepsilon}_t)$  et à  $(\hat{\varepsilon}_t^2)$ . Pour  $1 \leq t \leq T$ ,

$$S_t = \sum_{k=1}^t \hat{\varepsilon}_k \quad \text{et} \quad Q_t = \sum_{k=1}^t \hat{\varepsilon}_k^2. \quad (\text{VI.18})$$

Finalement, nous construisons une statistique de test équivalente à celle utilisée dans les tests KPSS

et LMC. Cette dernière est donnée par

$$\widehat{K}_T = \frac{1}{TQ_T} \sum_{t=1}^T S_t^2. \quad (\text{VI.19})$$

Nous allons dans le théorème suivant établir le comportement asymptotique de  $\widehat{K}_T$  sous  $\mathcal{H}_0$ . Les distributions asymptotiques engendrées seront précisément décrites dans la Section VI.4.

**Théorème VI.1.** *Supposons que  $\sigma_\eta^2 = 0$ . Alors, pour  $\kappa \neq 0$ , on a la convergence en loi*

$$\widehat{K}_T \xrightarrow{\mathcal{L}} \int_0^1 B_r^2(s) ds$$

où  $B_r(t)$  est un pont brownien généralisé d'ordre  $r$ . De plus, pour  $\kappa = 0$ , on a la convergence en loi

$$\widehat{K}_T \xrightarrow{\mathcal{L}} \int_0^1 W^2(s) ds$$

où  $W(t)$  est le processus de Wiener.

Dans le théorème suivant, nous montrons que  $\widehat{K}_T$  diverge sous  $\mathcal{H}_1$  à la vitesse  $T$  lorsque  $\rho = 1$ , et nous étudions le comportement asymptotique de la statistique de test correctement renormalisée. Nous montrons également que cette dernière décroît vers 0 sous  $\mathcal{H}_1$  lorsque  $\rho = -1$ .

**Théorème VI.2.** *Supposons que  $\sigma_\eta^2 > 0$ . Alors, pour  $\kappa \neq 0$  et  $\rho = 1$ , on a la convergence en loi*

$$\frac{\widehat{K}_T}{T} \xrightarrow{\mathcal{L}} \frac{\int_0^1 C_{r,1}^2(s) ds}{\int_0^1 W_{r,0}^2(s) ds}$$

où  $C_{r,1}(t)$  est un pont brownien intégré d'ordre  $r \times 1$  et  $W_{r,0}(t)$  est le processus Wiener recentré d'ordre  $r \times 0$ . En outre, pour  $\kappa = 0$ , on a la convergence en loi

$$\frac{\widehat{K}_T}{T} \xrightarrow{\mathcal{L}} \frac{\int_0^1 W^{(1)2}(s) ds}{\int_0^1 W^2(s) ds}$$

où  $W(t)$  est le processus de Wiener et  $W^{(1)}(t)$  est le processus de Wiener intégré d'ordre 1. Enfin, lorsque  $\rho = -1$ , on a la convergence en probabilité

$$\widehat{K}_T \xrightarrow{p} 0.$$

Le cas  $\rho = -1$  engendre de nombreuses complications, comme nous pourrions le voir dans les démonstrations. C'est pour cela que nous nous limitons à stipuler que la statistique  $\widehat{K}_T$  tend vers 0

dans le cas général. Cependant, dans le cas très particulier où  $\kappa = 0$ , nous parvenons malgré tout au résultat suivant.

**Proposition VI.1.** *Supposons que  $\sigma_\eta^2 > 0$ . Alors, pour  $\kappa = 0$  et  $\rho = -1$ , on a la convergence en loi*

$$T \widehat{K}_T \xrightarrow{\mathcal{L}} \frac{2\sigma_\varepsilon^2 \int_0^1 W_\varepsilon^2(s) ds + \sigma_\eta^2 \int_0^1 W_\eta^2(s)}{\sigma_\eta^2 \int_0^1 W_\eta^2(s)}$$

où  $W_\varepsilon(t)$  et  $W_\eta(t)$  sont deux processus de Wiener indépendants.

Remarquons que l'on sépare ici  $W_\varepsilon(t)$  et  $W_\eta(t)$  alors que ce n'était pas le cas jusqu'alors. La racine unitaire située en  $-1$  représente en effet le seul cas de figure où  $(\varepsilon_t)$  et  $(\eta_t)$  jouent un rôle équivalent dans la distribution limite de  $\widehat{K}_T$ . Sous  $\mathcal{H}_1$  avec  $\rho = 1$ ,  $(\eta_t)$  domine  $(\varepsilon_t)$  tandis que sous  $\mathcal{H}_0$ , seul  $(\varepsilon_t)$  intervient. Nous conjecturons que, sous  $\mathcal{H}_1$  pour tout  $r \geq 0$  et lorsque  $\rho = -1$ , il existe une distribution asymptotique non dégénérée identifiable à la quantité  $T \widehat{K}_T$ . Nous ne sommes pas encore parvenus à ce résultat général, de par la complexification des calculs dans ce cadre.

## VI.2.2 Focus sur les distributions asymptotiques

Ainsi que nous l'avons établi dans le Théorème VI.1, la statistique de test vérifie sous  $\mathcal{H}_0$ , et lorsque  $\kappa \neq 0$ , la convergence en loi

$$\widehat{K}_T \xrightarrow{\mathcal{L}} \int_0^1 B_r^2(s) ds.$$

La distribution limite est donc paramétrée par la valeur de  $r$ , elle n'est en outre ni gaussienne, ni symétrique, encore moins usuelle, et admet en conséquence une densité de probabilité difficile à intuitier. Dans la Table 2 de [87], nous en trouvons les principaux quantiles pour  $r = \{0, \dots, 5\}$  et pour  $\kappa = 0$  (où  $B_r(t)$  est donc remplacé par  $W(t)$ ). Nous allons sur la Figure VI.2, par une large simulation du modèle (VI.11), tenter d'en représenter la densité. Nous choisissons à cet égard des valeurs triviales des paramètres, perturbant le moins l'estimation. Ainsi, la tendance déterministe sera nulle (bien qu'estimée lorsque  $r$  grandit),  $(\varepsilon_t) \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 0.5)$ ,  $p = 0$  puisque sa valeur est transparente,  $T = 1000$  et, bien entendu,  $\sigma_\eta^2 = 0$ . L'expérience est quant à elle répétée  $N = 10000$  fois.

Il est clair que la densité obtenue tend à se concentrer vers 0 lorsque  $r$  augmente. Ce phénomène est conforme à la théorie puisque, comme il est indiqué dans [87], on a  $\mathbb{E}[\int_0^1 W^2(s) ds] = 1/2$ ,  $\mathbb{E}[\int_0^1 B_0^2(s) ds] = 1/6$ ,  $\mathbb{E}[\int_0^1 B_1^2(s) ds] = 1/15$ ,  $\mathbb{E}[\int_0^1 B_2^2(s) ds] = 3/70$ ,  $\mathbb{E}[\int_0^1 B_3^2(s) ds] = 2/63$ ,  $\mathbb{E}[\int_0^1 B_4^2(s) ds] = 5/198$ , etc. Nous utiliserons par la suite les quantiles proposés dans cette même référence (Table 2). La distribution limite de  $\widehat{K}_T/T$  sous  $\mathcal{H}_1$  avec  $\rho = 1$  est, quant à elle, bien plus nébuleuse. Rappelons que, dans ce cas,

$$\frac{\widehat{K}_T}{T} \xrightarrow{\mathcal{L}} \frac{\int_0^1 C_{r,1}^2(s) ds}{\int_0^1 W_{r,0}^2(s) ds}$$

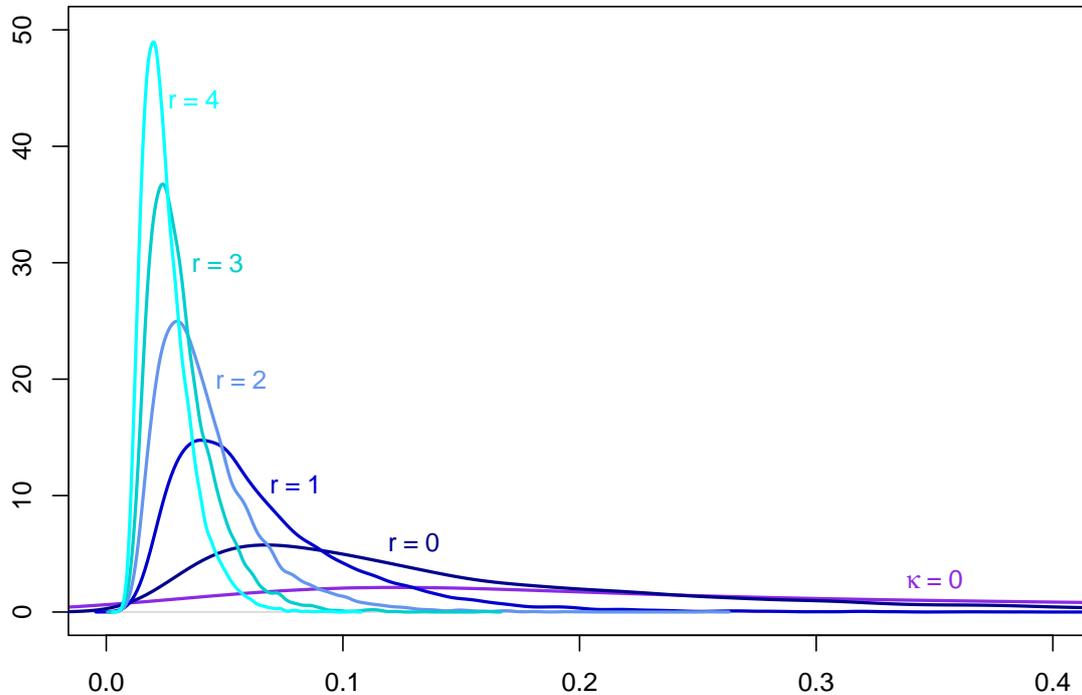


FIGURE VI.2 – Représentation de la distribution asymptotique de  $\widehat{K}_T$  sous  $\mathcal{H}_0$  pour  $\kappa = 0$ , puis pour  $\kappa \neq 0$  et des valeurs croissantes de  $r$ .

et représentons sur la Figure VI.3 la densité estimée issue de la simulation précédente, mais avec de plus  $\rho = 1$  et  $(\eta_t) \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 0.5)$ .

Nos expériences fournissent des approximations moins fluides, sans doute aurait-il fallu faire grimper les tailles d'échantillons pour mieux visualiser les distributions asymptotiques attendues. Nous remarquons cependant qu'à l'instar de leurs homologues sous  $\mathcal{H}_0$ , nous obtenons ici une famille de distributions qui tend à se concentrer vers 0 lorsque  $r$  augmente, et ce d'autant plus rapidement. On note de plus que ces distributions semblent être multimodales.

### VI.2.3 Application statistique

Nous souhaitons dans cette partie comparer notre procédure de test, issue des Théorèmes VI.1–VI.2, avec les tests KPSS et LMC. Nous nous attendons à obtenir des performances sensiblement équivalentes lorsque l'alternative  $\mathcal{H}_1$  repose sur  $\rho = 1$ . En revanche il est fort probable que notre procédure détecte également l'alternative  $\mathcal{H}_1$  axée sur  $\rho = -1$ , à la différence des tests KPSS et LMC. C'est

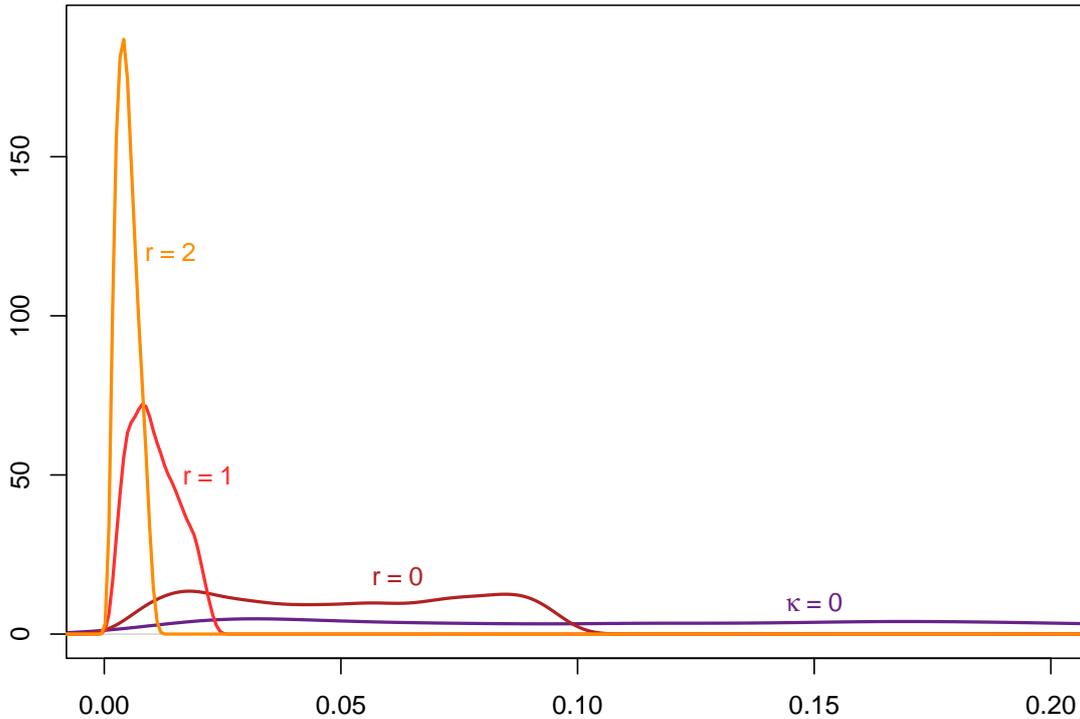


FIGURE VI.3 – Représentation de la distribution asymptotique de  $\widehat{K}_T/T$  sous  $\mathcal{H}_1$  avec  $d = 1$  ( $\rho = 1$ ) pour  $\kappa = 0$ , puis pour  $\kappa \neq 0$  et des valeurs croissantes de  $r$ .

en tout état de cause ce que la théorie nous suggère. Considérons en effet la distribution  $\int_0^1 W^2(s) ds$ , censée arbitrer le test de stationnarité lorsque  $\kappa = 0$ . Alors, les procédures KPSS et LMC rejettent  $\mathcal{H}_0$  dès que  $\widehat{K}_T > k_{1-\alpha}$  où, pour un niveau de risque  $0 < \alpha < 1$ ,  $k_{1-\alpha}$  est le quantile d'ordre  $1 - \alpha$  de la distribution en question (on se rappelle en outre que la statistique  $\widehat{K}_T$  n'est pas bâtie de la même façon pour le test KPSS). C'est ce que nous avons représenté sur la partie gauche de la Figure VI.4. Nous proposons de retirer une fraction  $\alpha/2$  de l'aire concentrée autour de 0, qui correspondrait, en vertu du Théorème VI.2, au cas non stationnaire où  $\rho = -1$ . En ajoutant cette même fraction de l'autre côté du spectre, nous conservons ainsi le niveau  $\mathbb{P}(\text{ne pas rejeter } \mathcal{H}_0 | \mathcal{H}_0) = 1 - \alpha$ , mais nous tenons compte de deux situations de rejet éventuel. Cette stratégie est résumée sur la partie droite de la Figure VI.4. Le rejet associé à  $\rho = 1$  peut quant à lui être vérifié par l'intermédiaire des quantiles de la distribution asymptotique établie sous  $\mathcal{H}_1$ .

Il nous reste maintenant à valider le bien-fondé de cette procédure en simulation. Nous testons donc  $\mathcal{H}_0 : \sigma_{\eta}^2 = 0$  contre son alternative  $\mathcal{H}_1 : \sigma_{\eta}^2 > 0$  en estimant la puissance empirique des

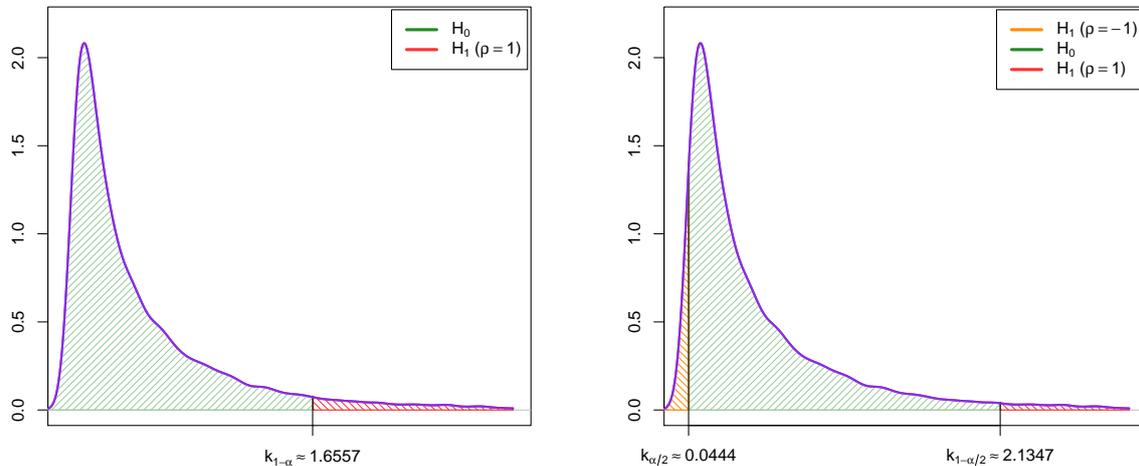


FIGURE VI.4 – Règle de rejet de  $\mathcal{H}_0$  au risque  $\alpha = 0.05$  des procédures KPSS et LMC (gauche), et de notre procédure (droite), dans le cas où  $\kappa = 0$ .

procédures KPSS, LMC ainsi que celle de la procédure générale (GEN) que nous avons proposée ci-dessus, selon le protocole largement détaillé tout au long du manuscrit, avec un risque de  $\alpha = 0.05$ . Nous choisissons à cette occasion les distributions résiduelles  $(\varepsilon_t) \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 0.5)$  et  $(\eta_t) \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\eta^2)$ . Nous générons de plus nos données avec  $\kappa = 0$  pour commencer, puis avec  $\kappa \neq 0$  et  $r \geq 0$ . En outre, nous faisons aussi varier  $p$  pour nous assurer que sa valeur est bien transparente relativement à la puissance des tests. Enfin, les courbes sont de taille  $T = 100$ , les échantillons de taille  $N = 1000$  et  $\sigma_\eta^2$  varie entre 0 et 1. Les résultats, reproduits pour  $\rho = 1$  puis pour  $\rho = -1$ , sont résumés sur la Figure VI.5 pour  $\kappa = 0$  et  $p = 0$ , sur la Figure VI.6 pour  $\kappa \neq 0$ ,  $r = 0$ ,  $p = 1$ ,  $\alpha_0 = -2$  et  $\theta_1 = 0.5$ , et sur la Figure VI.7 pour  $\kappa \neq 0$ ,  $r = 1$ ,  $p = 2$ ,  $\alpha_0 = -2$ ,  $\alpha_1 = 100$ ,  $\theta_1 = -0.2$  et  $\theta_2 = 0.3$ .

Nos expériences nous fournissent des résultats absolument conformes à la théorie, et ce pour des valeurs grandissantes de  $r$  et de  $p$ . La frontière brutale existant entre  $\sigma_\eta^2 = 0$  et  $\sigma_\eta^2 > 0$ , à l'origine de la présence d'un terme de variance explosive proportionnelle à  $T$ , permet de discriminer facilement la stationnarité et la non stationnarité. Les graphes pour  $\rho = 1$  présentent ainsi un palier radical dès que  $\sigma_\eta^2 = 0.01$  et les procédures tendent très rapidement à rejeter  $\mathcal{H}_0$ , à juste titre. Nous voyons de plus que  $p$  ne perturbe en rien, ou alors très peu en raison des instabilités numériques, la règle de décision et la distribution asymptotique. Lorsque l'on fait augmenter  $T$  (jusqu'à 1000 par exemple), nous observons les mêmes motifs quoiqu'en toute logique beaucoup plus marqués. Le niveau est en outre bien retrouvé sous  $\mathcal{H}_0$ , la procédure rejetant à tort dans 5% des cas, globalement. Le test KPSS semble donner de meilleurs résultats lorsque  $p$  augmente, il s'exonère en effet de l'estimation de  $\theta$ , source de perturbations numériques, et ce d'autant plus que l'horizon  $T$  n'est pas très élevé dans nos simulations. En revanche, lorsque  $p = 0$ , ce sont les tests GEN et LMC qui fournissent les

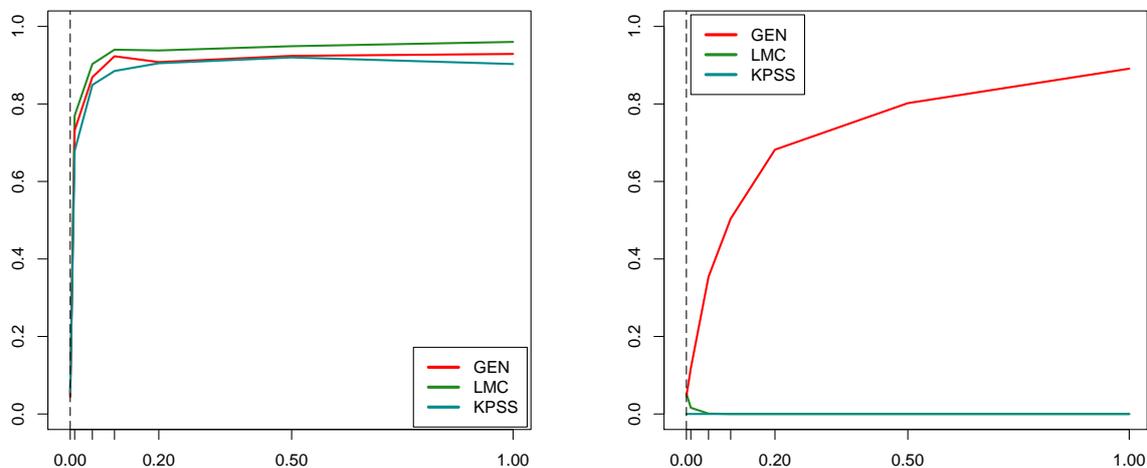


FIGURE VI.5 – Fréquence de rejet de  $\mathcal{H}_0$  en ordonnée et  $\sigma_\eta^2$  variant de 0 à 1 en abscisse, pour  $\kappa = 0$ ,  $\rho = 0$  et  $\rho = 1$  (gauche), puis  $\rho = -1$  (droite).

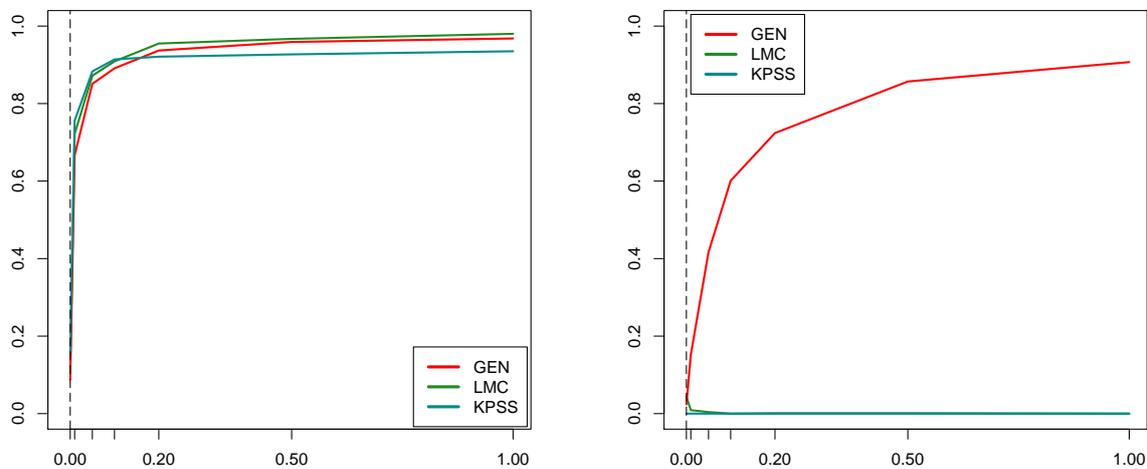


FIGURE VI.6 – Fréquence de rejet de  $\mathcal{H}_0$  en ordonnée et  $\sigma_\eta^2$  variant de 0 à 1 en abscisse, pour  $r = 0$ ,  $\rho = 1$  et  $\rho = 1$  (gauche), puis  $\rho = -1$  (droite).

meilleurs résultats, dans la mesure où la variance résiduelle est estimée de manière optimale, sans faire intervenir la troncature  $\ell(T)$  ralentissant la divergence sous  $\mathcal{H}_1$ . Enfin, puisque nous avons reculé la zone de rejet de  $k_{1-\alpha}$  à  $k_{1-\alpha/2}$ , il est normal que le test LMC tende à rejeter plus souvent  $\mathcal{H}_0$  lorsque  $\rho = 1$ , ce que nous constatons également. Considérons maintenant la situation où le paramètre  $\rho = -1$ . Les graphes parlent ainsi d'eux-mêmes, il est flagrant que notre procédure est la

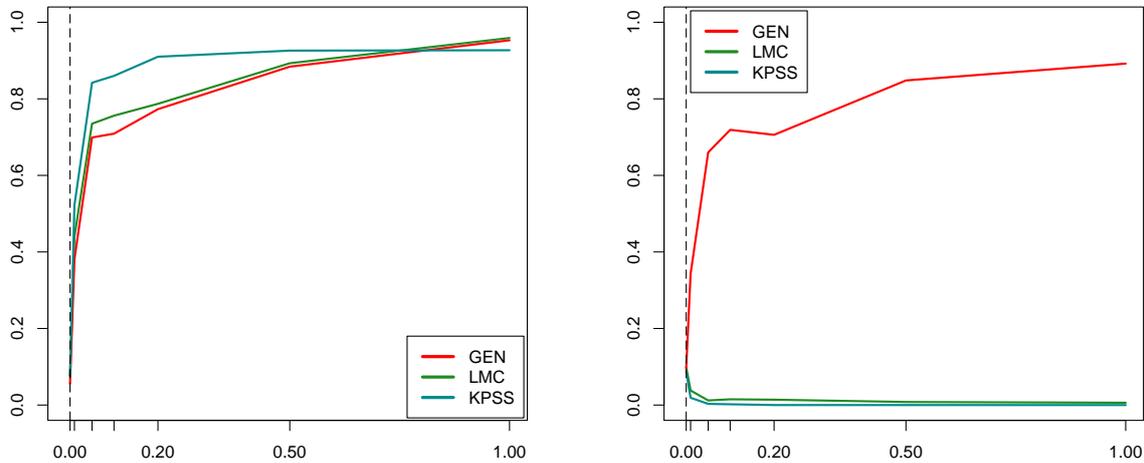


FIGURE VI.7 – Fréquence de rejet de  $\mathcal{H}_0$  en ordonnée et  $\sigma_\eta^2$  variant de 0 à 1 en abscisse, pour  $r = 1$ ,  $p = 2$  et  $\rho = 1$  (gauche), puis  $\rho = -1$  (droite).

seule à détecter la non stationnarité engendrée par une racine unitaire localisée à  $-1$ . Là encore, les observations sont conformes à la théorie. Il faut néanmoins bien insister ici sur la difficulté liée à l'estimation de  $\theta$ , qui nécessite de savoir au préalable si l'alternative la plus plausible est  $\rho = 1$  ou  $\rho = -1$  (car il peut alors s'agir de sommer le processus et non plus de le différencier), qui nécessite également de bien tenir compte du fait qu'une sommation ajoutée à une différenciation (par exemple pour tenir compte de  $\rho = -1$  et d'une tendance linéaire) se combinent en une composante MA(2) résiduelle, et non plus MA(1), car le résidu est alors  $((1-L)\eta_t + (1-L^2)\varepsilon_t)$ , et ce en plus de la présence d'une constante dans le modèle. De même, lorsque  $\kappa \neq 0$  et  $r = 0$ , le modèle contient une constante uniquement dans le cas où  $\rho = -1$ , sinon elle ne doit pas être estimée. En bref, un travail descriptif en amont est nécessaire à la bonne estimation de  $\theta$ , et c'est à ce titre que la procédure KPSS présente certains avantages dans le fait de translater à la perturbation tous les phénomènes d'autocorrélation.

## VI.2.4 Sur le processus autorégressif à erreurs autocorrélées

Ce travail nous permet d'envisager une extension des Chapitres II–III au cas instable où  $|\rho| = 1$ . En effet, considérons le modèle donné, pour tout  $p \leq t \leq T$ , par

$$\begin{cases} Y_t &= \theta_1 Y_{t-1} + \dots + \theta_p Y_{t-p} + S_t^\eta \\ S_t^\eta &= \rho S_{t-1}^\eta + \eta_t \end{cases} \quad (\text{VI.20})$$

où  $(\eta_t)$  est un bruit blanc de variance  $\sigma_\eta^2 > 0$ ,  $\|\theta\|_1 < 1$  et  $|\rho| \leq 1$ . Alors, il n'est pas difficile de constater que le processus est engendré par une racine (éventuellement unitaire) localisée en  $\rho$ , et que

$$\mathcal{A}(L)(1 - \rho L)Y_t = \eta_t \quad (\text{VI.21})$$

où, pour tout  $z \in \mathbb{C}$ ,  $\mathcal{A}(z) = 1 - \theta_1 z - \dots - \theta_p z^p$  est un polynôme causal. L'identification est faite avec le modèle (VI.11)–(VI.12) lorsque  $\sigma_\eta^2 > 0$ ,  $\sigma_\varepsilon^2 = 0$  et  $\kappa = 0$ . De même qu'au Chapitre V nous avons mis en évidence la présence d'une frontière stabilité/instabilité située entre  $\rho < 0$  et  $\rho = 0$  ayant des répercussions fortes sur la distribution asymptotique des estimateurs, de même ici nous allons constater ce phénomène entre  $|\rho| < 1$  et  $|\rho| = 1$ . Reprenons alors la statistique de Durbin-Watson définie en (II.13). Nous la dérivons ici sous deux versions légèrement différentes données, pour tout  $T \geq p$ , par

$$\widehat{D}_T^+ = \frac{\sum_{t=1}^T (\Delta \widehat{\varepsilon}_t)^2}{\sum_{t=0}^T \widehat{\varepsilon}_t^2} \quad \text{et} \quad \widehat{D}_T^- = \frac{\sum_{t=1}^T (\nabla \widehat{\varepsilon}_t)^2}{\sum_{t=0}^T \widehat{\varepsilon}_t^2} \quad (\text{VI.22})$$

où nous rappelons que, par convention,  $\widehat{\varepsilon}_0 = \dots = \widehat{\varepsilon}_{p-1} = 0$ . Nous avons établi, dans les Corollaires II.1–III.1 que, sous la condition de stabilité  $|\rho| < 1$ , on a les convergences

$$\lim_{T \rightarrow \infty} \widehat{D}_T^+ = D^* \quad \text{p.s.} \quad \text{et} \quad \frac{\sqrt{T}}{\sigma_D} (\widehat{D}_T^+ - D^*) \xrightarrow{\mathcal{L}} W(1)$$

où  $D^* = 2(1 - \rho^*)$  et  $\sigma_D^2 = 4\sigma_\rho^2$ , avec  $\rho^*$  et  $\sigma_\rho^2$  explicitement définis en (III.14) et (III.17), et où  $W(1)$  désigne la loi normale  $\mathcal{N}(0, 1)$ , pour conserver les notations inhérentes à ce chapitre. L'ensemble résiduel reposait alors sur une estimation biaisée de  $\theta$ , par moindres carrés. Si maintenant le paramètre  $\theta$  est estimé de manière consistante par maximum de vraisemblance, dans le modèle ARIMA( $p, 1, 0$ ) donné par (VI.21), alors on a le résultat suivant sur la frontière  $|\rho| = 1$ .

**Proposition VI.2.** *Supposons que  $\sigma_\varepsilon^2 = 0$ , que  $\sigma_\eta^2 > 0$  et que  $\kappa = 0$ . Alors, lorsque  $\rho = 1$ , nous avons les convergences*

$$\widehat{D}_T^+ \xrightarrow{\mathcal{P}} 0 \quad \text{et} \quad T \widehat{D}_T^+ \xrightarrow{\mathcal{L}} \frac{1}{\int_0^1 W^2(s) ds}$$

où  $W(t)$  est le processus de Wiener et la statistique  $\widehat{D}_T^-$  est donnée par (VI.22). De plus, lorsque  $\rho = -1$ , nous avons les convergences

$$\widehat{D}_T^- \xrightarrow{\mathcal{P}} 0 \quad \text{et} \quad T \widehat{D}_T^- \xrightarrow{\mathcal{L}} \frac{1}{\int_0^1 W^2(s) ds}$$

où  $W(t)$  est le processus de Wiener et la statistique  $\widehat{D}_T^+$  est donnée par (VI.22).

**Démonstration.** Les Théorèmes VI.1–VI.2 ainsi que les Propositions VI.1–VI.2 sont prouvés dans la Section VI.6.2. ★

Ces résultats nous permettent dès lors de faire une jonction importante entre la stationnarité de

la trajectoire autorégressive étudiée et son autocorrélation résiduelle. Nous pouvons en effet tester la présence d'une racine unitaire par l'intermédiaire des Théorèmes VI.1–VI.2 et, le cas échéant, affiner l'étude de la valeur de  $|\rho| < 1$  grâce aux outils développés dans les Chapitres II–III.

### VI.3 Comportement en présence de multi-intégration

Supposons désormais que le processus autorégressif  $(Y_t)$  est engendré par  $d = d^+ + d^-$  racines unitaires. Typiquement,  $z = 1$  est une racine du polynôme  $\mathcal{A}$  de multiplicité  $d^+$  et  $z = -1$  une racine de multiplicité  $d^-$ . Le modèle est ainsi donné, pour tout  $1 \leq t \leq T$  et reprenant les notations de la section précédente, par

$$\mathcal{A}(L)Y_t = (\alpha_0 + \alpha_1 t_T + \dots + \alpha_r t_T^r) \mathbb{I}_{\{\kappa \neq 0\}} + S_t^{(d)} + \varepsilon_t \quad (\text{VI.23})$$

où  $(S_t^{(d)})$  est une marche aléatoire d'ordre  $d$  que nous pouvons définir comme

$$\begin{cases} S_t^{(d)} = \rho_d S_{t-1}^{(d)} + S_t^{(d-1)} \\ \vdots \\ S_t^{(2)} = \rho_2 S_{t-1}^{(2)} + S_t^{(1)} \\ S_t^{(1)} = \rho_1 S_{t-1}^{(1)} + \eta_t \end{cases} \quad (\text{VI.24})$$

avec  $|\rho_1| = \dots = |\rho_d| = 1$  parmi lesquels  $d^+$  valent 1 et  $d^-$  valent  $-1$ . Nous supposons, pour alléger les calculs, que  $S_0^{(1)} = \dots = S_0^{(d)} = 0$ . L'étude de la statistique  $\widehat{K}_T$  nous conduira alors à une procédure de test permettant d'évaluer

$$\mathcal{H}_0 : \sigma_\eta^2 = 0 \quad \text{vs} \quad \mathcal{H}_1 : \sigma_\eta^2 > 0$$

où l'alternative  $\mathcal{H}_1$  forme un modèle ARIMA d'ordre  $d$  (avec éventuellement l'opérateur de sommation en lieu et place de la différenciation). En fait, nous ne tiendrons compte dans cette étude que des deux cas extrêmes  $(d^+, d^-) = (d, 0)$  et  $(d^+, d^-) = (0, d)$ . En effet, la combinaison d'une racine localisée à 1 et d'une autre à  $-1$  résulte en un processus intégré d'ordre 1, et non pas 2 comme l'on aurait pu s'y attendre. Pour illustrer ceci, considérons le résidu (VI.24) pour  $d = 2$ ,  $\rho_2 = -1$  et  $\rho_1 = 1$ ,

$$\begin{cases} S_t^{(2)} = -S_{t-1}^{(2)} + S_t^{(1)} \\ S_t^{(1)} = S_{t-1}^{(1)} + \eta_t. \end{cases}$$

Il suit que  $(1+L)S_t^{(2)} = S_t^{(1)}$  et que  $(1-L^2)S_t^{(2)} = \eta_t$ , ce qui signifie que le résidu se comporte comme un processus ARIMA(2, 1, 0) à trou. Ainsi, pour tout couple  $(d^+, d^-)$ , il est possible de trouver une représentation équivalente du processus ayant moins que  $d$  racines unitaires, à l'exception des extrêmes  $(d, 0)$  et  $(0, d)$ , ce qui explique la restriction que nous nous imposons.

### VI.3.1 La procédure d'estimation

Tout d'abord et comme précédemment, il est important de noter que l'on considère ici une valeur connue de  $r$  d'une part, et que l'on a d'autre part de bonnes indications empiriques sur la valeur de  $(d^+, d^-)$  à retenir. En effet, on observe visuellement que le comportement engendré est totalement différent selon la localisation des racines unitaires, avec une explosivité rapide d'un côté du spectre pour  $(d, 0)$  et une tendance très prononcée à l'alternance autour de la tendance pour  $(0, d)$ . Comme nous avons également pu le voir dans la section précédente, une surdifférenciation du processus engendre des racines unitaires dans la partie MA, mais ne gêne en rien la consistance de l'estimateur de  $\theta$ . Ainsi, une fois que nous sommes accordés sur le choix de  $(d^+, d^-)$  à retenir, la procédure découle de la même manière. Nous notons que la source de la non stationnarité stochastique de  $(Y_t)$  est un terme à première vue assez compliqué,

$$S_t^{(d)} = \sum_{t_d=1}^t \rho_d^{t-t_d} S_{t_d}^{(d-1)} = \dots = \sum_{t_d=1}^t \dots \sum_{t_1=1}^{t_2} \prod_{k=1}^d \rho_k^{t_{k+1}-t_k} \eta_{t_1} \quad (\text{VI.25})$$

avec la convention que  $t_{d+1} = t$ . Ainsi, le modèle (VI.23) possède l'expression intégrée donnée, pour  $1 \leq t \leq T$ , par

$$\mathcal{A}(L)(1-L)^{d^+}(1+L)^{d^-} Y_t = (\gamma_0 + \gamma_1 t_T + \dots + \gamma_r t_T^r) \mathbb{I}_{\{\kappa \neq 0\}} + \eta_t + (1-L)^{d^+}(1+L)^{d^-} \varepsilon_t$$

où  $\gamma_0, \gamma_1, \dots, \gamma_r$  sont identifiables et où le résidu  $(\eta_t + (1-L)^{d^+}(1+L)^{d^-} \varepsilon_t)$  correspond à un processus MA( $d$ ). Nous avons dès lors

$$\mathcal{A}(L)(1-L)^{d^+}(1+L)^{d^-} Y_t = (\gamma_0 + \gamma_1 t_T + \dots + \gamma_r t_T^r) \mathbb{I}_{\{\kappa \neq 0\}} + \mathcal{B}(L) \xi_t \quad (\text{VI.26})$$

où  $(\xi_t)$  est un bruit blanc de variance  $\sigma_\xi^2$  finie et  $\mathcal{B}$  est le polynôme associé d'ordre  $d$  donné, pour tout  $z \in \mathbb{C}$ , par  $\mathcal{B}(z) = 1 + \beta_1 z + \dots + \beta_d z^d$ . Sous  $\mathcal{H}_0 : \sigma_\eta^2 = 0$ ,  $\mathcal{B}$  possède uniquement des racines unitaires, et cela ne gêne en rien la consistance de l'estimation. Cependant, nous devons nous assurer de la causalité sous  $\mathcal{H}_1 : \sigma_\eta^2 > 0$ . Celle-ci découle de la Proposition 3.5.1 de [25] stipulant que, puisque  $\mathcal{A}$  est par hypothèse causal et que le processus est stationnarisé, il existe un bruit blanc  $(\xi_t)$  tel que toutes les racines de  $\mathcal{B}$  se situent en dehors du cercle unité. Ainsi, de même que dans la Section VI.2, soit  $(\tilde{\theta}_T, \tilde{\beta}_T)$  l'estimateur consistant du maximum de vraisemblance de  $(\theta, \beta)$  dans le modèle (VI.26) dont la tendance a été correctement éliminée, et soit le processus résiduel  $(\tilde{Y}_t)$  donné par (VI.15). Nous estimons de nouveau  $\alpha$  sous  $\mathcal{H}_0$  par l'intermédiaire d'une méthodologie des moindres carrés dans le modèle donné par (VI.16) et nous obtenons la même famille de résidus  $(\hat{\varepsilon}_t)$ . Explicitement, pour tout  $p \leq t \leq T$ ,

$$\hat{\varepsilon}_t = \tilde{Y}_t - (\hat{\alpha}_0 + \hat{\alpha}_1 t_T + \dots + \hat{\alpha}_r t_T^r) \mathbb{I}_{\{\kappa \neq 0\}}$$

où  $\widehat{\alpha}_T$  est donc l'estimateur des moindres carrés du paramètre  $\alpha$ . Nous considérons arbitrairement que  $\widehat{\varepsilon}_1 = \dots = \widehat{\varepsilon}_{p-1} = 0$ . La procédure repose sur la même statistique de test, à savoir

$$\widehat{K}_T = \frac{1}{TQ_T} \sum_{t=1}^T S_t^2$$

où les processus des sommes partielles  $(S_t)$  et  $(Q_t)$  sont définis en (VI.18). Bien sûr, le Théorème VI.1 reste valable puisque sous  $\mathcal{H}_0$ , la valeur de  $d$  n'intervient pas dans la génération des données. En revanche, nous proposons sous  $\mathcal{H}_1$  le résultat suivant.

**Théorème VI.3.** *Supposons que  $\sigma_\eta^2 > 0$ . Alors, pour  $\kappa \neq 0$  et  $d = d^+ \geq 1$ , on a la convergence en loi*

$$\frac{\widehat{K}_T}{T} \xrightarrow{\mathcal{L}} \frac{\int_0^1 C_{r,d}^2(s) ds}{\int_0^1 W_{r,d-1}^2(s) ds}$$

où  $C_{r,d}(t)$  est un pont brownien intégré d'ordre  $r \times d$  et  $W_{r,d-1}(t)$  est le processus Wiener recentré d'ordre  $r \times (d-1)$ . En outre, pour  $\kappa = 0$ , on a la convergence en loi

$$\frac{\widehat{K}_T}{T} \xrightarrow{\mathcal{L}} \frac{\int_0^1 W^{(d)2}(s) ds}{\int_0^1 W^{(d-1)2}(s) ds}$$

où  $W^{(d-1)}(t)$  et  $W^{(d)}(t)$  sont les processus de Wiener intégrés d'ordre  $d-1$  et  $d$ , respectivement. Enfin, lorsque  $d = d^- \geq 1$ , on a la convergence en probabilité

$$\widehat{K}_T \xrightarrow{\mathbb{P}} 0.$$

Nous pouvons observer que la vitesse de convergence de  $\widehat{K}_T$  sous  $\mathcal{H}_1$  ne dépend pas de  $d$  dès que  $d \geq 1$ , comme nous aurions pu l'espérer, et que la distribution asymptotique seule est impactée. Cela forme une restriction de la procédure puisqu'il est dès lors impossible d'estimer la valeur de  $d$  par l'intermédiaire du Théorème VI.3. Nous avons malgré tout le résultat suivant pour les racines positives.

**Proposition VI.3.** *Supposons que  $\sigma_\eta^2 > 0$  et que  $d = d^+ \geq 1$ . Alors, on a les convergences en loi*

$$\frac{Q_T}{T^{2d}} \xrightarrow{\mathcal{L}} \sigma_\eta^2 \int_0^1 W_{r,d-1}^2(s) ds \quad \text{et} \quad \frac{1}{T^{2(d+1)}} \sum_{t=1}^T S_t^2 \xrightarrow{\mathcal{L}} \sigma_\eta^2 \int_0^1 C_{r,d}^2(s) ds$$

où  $C_{r,d}(t)$  est un pont brownien intégré d'ordre  $r \times d$  et  $W_{r,d-1}(t)$  est le processus de Wiener recentré d'ordre  $r \times (d-1)$ .

Les convergences ci-dessus sont bien entendu inexploitable en pratique puisque la valeur de  $\sigma_\eta^2$  nous est inconnue, mais nous en tirons un critère visuel lié aux ordres de grandeur en puissance de  $d$ . Cela signifie que, si d'aventure nous devons hésiter entre un comportement ARIMA intégré à l'ordre 1 ou 2 pour une série chronologique, nous aurions la possibilité d'évaluer  $Q_T$  avec  $d = 2$  puis d'observer la chute de la valeur de  $Q_T$  renormalisée. C'est ce que nous avons simulé sur la Figure VI.8 où, pour une série chronologique engendrée avec  $T = 1000$ ,  $\kappa \neq 0$ ,  $p = 0$ ,  $r = 1$ ,  $\alpha_0 = -2$ ,  $(\varepsilon_t) \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 0.5)$ ,  $(\eta_t) \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$  et respectivement  $d = 1$  puis  $d = 2$ , les racines étant localisées à 1, nous avons représenté l'évolution de  $Q_T/T^{2\delta}$  pour  $\delta \in \{0, 1, 2, 3\}$ . C'est ainsi que l'étude visuelle des graphes obtenus nous permet de sélectionner sans ambiguïté, et à juste titre,  $\delta = 1$  à gauche puis  $\delta = 2$  à droite, à partir de la même estimation résiduelle.

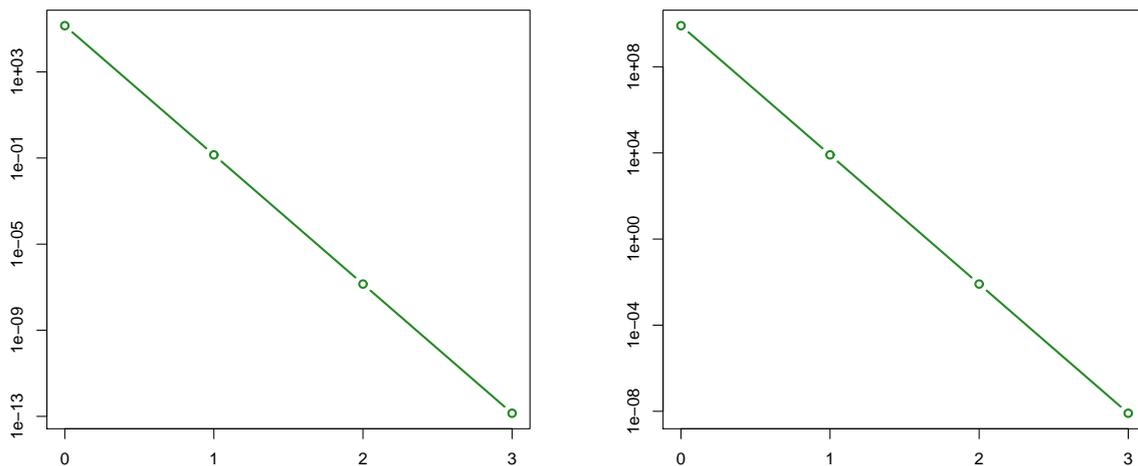


FIGURE VI.8 – Évolution de  $Q_T/T^{2\delta}$  en ordonnée pour  $\delta = \{0, 1, 2, 3\}$  en abscisse, avec  $d = 1$  (gauche) puis avec  $d = 2$  (droite), en échelle logarithmique.

### VI.3.2 Focus sur les distributions asymptotiques

Seule la famille de distributions impliquées dans le Théorème VI.3 nous intéresse ici. Nous avons en effet vu que sous  $\mathcal{H}_0$ , les distributions asymptotiques ne changeaient pas et qu'en conséquence, nous les avons déjà étudiées dans la Section VI.2.2. Nous représentons ainsi sur la Figure VI.9 les distributions empiriques observées pour  $\kappa = 0$ , puis pour  $\kappa \neq 0$  et  $r = \{0, 1, 2\}$ , avec  $d = 2$  et ce pour la même configuration que celle utilisée sur la Figure VI.3. Nous voyons par analogie que l'augmentation de  $d$  semble accentuer le caractère multimodal des distributions ainsi que l'intensité du pic principal mais que, d'une manière générale, la variation de  $d$  a moins d'influence sur la forme de la distribution que la variation de  $r$ , bien sûr dès que  $d \geq 1$ .

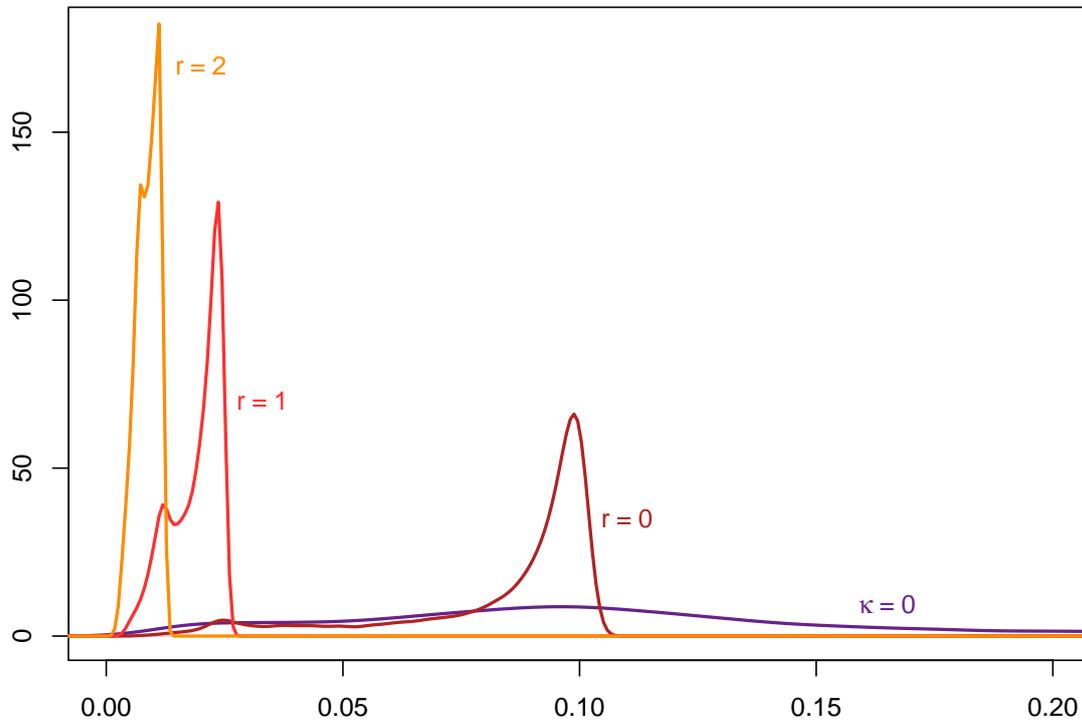


FIGURE VI.9 – Représentation de la distribution asymptotique de  $\widehat{K}_T/T$  sous  $\mathcal{H}_1$  avec  $d = 2$  ( $\rho_1 = \rho_2 = 1$ ) pour  $\kappa = 0$ , puis pour  $\kappa \neq 0$  et des valeurs croissantes de  $r$ .

### VI.3.3 La variance résiduelle

Nous allons conclure cette étude en nous intéressant à la variance résiduelle du processus stationnarisé. Sous  $\mathcal{H}_0$ , un estimateur consistant de  $\sigma_\varepsilon^2$  est donné par  $Q_T/T$ . Nous pouvons ainsi estimer cette variance et l'éliminer de la distribution asymptotique dans le Théorème VI.1. Cependant, sous  $\mathcal{H}_1$ , nous n'avons pas accès à l'estimation consistante de  $\sigma_\eta^2$ . Néanmoins, nous avons établi dans la section précédente que le processus admettait une modélisation ARMA( $p, d$ ) intégrée (si  $d > 0$ ) dont la partie MA( $d$ ) s'exprimait à travers le résidu  $(\eta_t + \Delta^d \varepsilon_t)$  pour  $d = d^+$ , et nous souhaitons estimer la variance de ce résidu stationnaire, donnée par

$$\omega^2 = \mathbb{V}(\eta_t + \Delta^d \varepsilon_t) = \sigma_\eta^2 + \sigma_\varepsilon^2 \sum_{k=0}^d \binom{d}{k}^2 = \sigma_\eta^2 + \binom{2d}{d} \sigma_\varepsilon^2. \quad (\text{VI.27})$$

**Proposition VI.4.** *Supposons que  $\sigma_\eta^2 > 0$  et que  $d = d^+ \geq 0$ . Alors, on a la convergence presque sûre*

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=d}^T (\Delta^d \widehat{\varepsilon}_t)^2 = \omega^2 \quad \text{p.s.}$$

où  $\omega^2$  est définie en (VI.27).

*Démonstration.* Le Théorème VI.3 et les Propositions VI.3–VI.4 sont prouvés dans la Section VI.6.2. ★

Notons qu'il est également possible d'obtenir le même résultat lorsque  $d = d^- \geq 0$  dans le cas où aucune tendance n'est retenue ( $\kappa = 0$ ). Il suffit alors de remplacer l'opérateur de différenciation  $\Delta^d$  par l'opérateur de sommation  $\nabla^d$  dans l'estimateur. La variance  $\omega^2$  reste quant à elle inchangée.

## VI.4 Quelques processus stochastiques utiles

Les distributions asymptotiques que nous avons obtenues, ainsi que la démonstration de nos résultats, font appel de manière récurrente à certaines familles de processus stochastiques paramétrés par les ordres  $r$  et  $d$  du modèle retenu, et construits sur le processus de Wiener  $W(t)$ . Nous allons dans cette section les expliciter.

### VI.4.1 Le processus de Wiener intégré

**Définition VI.1.** *Le processus stochastique défini, pour tout  $t \in [0, 1]$ , par*

$$W^{(d)}(t) = \int_0^t \int_0^{s_1} \dots \int_0^{s_{d-1}} W(s_d) ds_d \dots ds_1$$

où  $W(t)$  est le processus de Wiener, sera appelé « processus de Wiener intégré d'ordre  $d$  ». Par convention,  $W^{(0)}(t) \equiv W(t)$ .

Par exemple,

$$W^{(1)}(t) = \int_0^t W(s) ds \quad \text{et} \quad W^{(2)}(t) = \int_0^t \int_0^s W(u) du ds.$$

### VI.4.2 Le pont brownien généralisé

Nous commençons par introduire une application  $h_r$  de  $\mathcal{C}([0, 1])$  dans lui-même, l'ensemble des fonctions continues sur  $[0, 1]$ . Nous reprenons cette expression du travail de MacNeill [87] qui, en

1978, étudie les propriétés des résidus issus d'une régression polynomiale. Soit  $F$  une fonction continue sur  $[0, 1]$ . Alors, pour tout  $t \in [0, 1]$ ,

$$h_r(F(t)) = F(t) - \sum_{m=0}^r (2m+1) \left( \sum_{q=0}^{\lfloor m/2 \rfloor} K_{q,m} \int_0^t f_s^{m-2q} ds \right) \left( \sum_{q=0}^{\lfloor m/2 \rfloor} K_{q,m} \left[ \frac{F(1)}{2^{m-2q}} - (m-2q) \int_0^1 f_s^{m-2q-1} F(s) ds \right] \right)$$

avec

$$K_{q,m} = \frac{(-1)^q \binom{2m}{m, q, q, m-2q}}{2^{4q} \binom{m-\frac{1}{2}}{q}} \quad \text{et} \quad f_s = s - \frac{1}{2}$$

en empruntant les notations et conventions usuelles liées à l'expression des polynômes orthogonaux utilisées dans la référence en question. Ainsi,

$$h_0(F(t)) = F(t) - tF(1) \quad \text{et} \quad h_1(F(t)) = F(t) + t(2-3t)F(1) - 6t(1-t) \int_0^1 F(s) ds.$$

**Définition VI.2.** Le processus stochastique défini, pour tout  $t \in [0, 1]$ , par

$$B_r(t) = h_r(W(t))$$

où la fonction  $h_r$  est donnée ci-dessus et  $W(t)$  est le processus de Wiener, sera appelé « pont brownien généralisé d'ordre  $r$  ».

### VI.4.3 Le pont brownien intégré

**Définition VI.3.** Le processus stochastique défini, pour tout  $t \in [0, 1]$ , par

$$C_{r,d}(t) = h_r(W^{(d)}(t))$$

où la fonction  $h_r$  est donnée ci-dessus et  $W^{(d)}(t)$  est le processus de Wiener intégré d'ordre  $d$ , sera appelé « pont brownien intégré d'ordre  $r \times d$  ». Par convention,  $C_{r,0}(t) \equiv B_r(t)$ .

### VI.4.4 Le processus de Wiener recentré

Soit maintenant la matrice carrée non singulière  $M$  d'ordre  $r+1$  satisfaisant  $M_{ij} = 1/(i+j-1)$  pour tous  $1 \leq i, j \leq r+1$ , et  $\Lambda(t) = (1 \ t \ \dots \ t^r)'$ . Soit également le processus vectoriel défini, pour tout  $t \in [0, 1]$ , par

$$P'_d(t) = \left( W^{(d)}(t) \quad \int_0^t s W^{(d-1)}(s) ds \quad \dots \quad \int_0^t s^r W^{(d-1)}(s) ds \right) \quad (\text{VI.28})$$

où  $W^{(d-1)}(t)$  et  $W^{(d)}(t)$  sont les processus de Wiener intégrés d'ordre  $d - 1$  et  $d$ , respectivement.

**Définition VI.4.** Le processus stochastique défini, pour tout  $t \in [0, 1]$ , par

$$W_{r,d}(t) = \frac{dC_{r,d+1}(t)}{dt}$$

où  $C_{r,d+1}(t)$  est le pont brownien intégré d'ordre  $r \times (d + 1)$ , sera appelé « processus de Wiener recentré d'ordre  $r \times d$  ». Il est explicitement donné par

$$W_{r,d}(t) = W^{(d)}(t) - P'_d(1)M^{-1}\Lambda(t)$$

où le processus vectoriel  $P_d(t)$  est défini en (VI.28).

Nous allons illustrer ces définitions dans les cas particuliers souvent rencontrés en pratique, correspondant à  $r = \{0, 1\}$  et à  $d = 0$ . Selon la Définition VI.2, pour  $t \in [0, 1]$ ,

$$B_0(t) = h_0(W(t)) = W(t) - tW(1)$$

qui est le traditionnel pont brownien. De plus, à partir des Définitions VI.3 et VI.4, on a

$$C_{0,1}(t) = h_0(W^{(1)}(t)) = \int_0^t W(s) ds - t \int_0^1 W(s) ds$$

et

$$W_{0,0}(t) = \frac{dC_{0,1}(t)}{dt} = W(t) - \int_0^1 W(s) ds$$

qui est le processus de Wiener dont on a ôté la valeur moyenne. En outre, pour  $r = 1$ ,

$$B_1(t) = h_1(W(t)) = W(t) + t(2 - 3t)W(1) - 6t(1 - t) \int_0^1 W(s) ds$$

est un pont brownien du second ordre, dont la version intégrée donne, après simplifications,

$$C_{1,1}(t) = \int_0^t W(s) ds + t(3t - 4) \int_0^1 W(s) ds + 6t(1 - t) \int_0^1 sW(s) ds.$$

Enfin, le processus de Wiener linéairement recentré est donné par

$$W_{1,0}(t) = \frac{dC_{1,1}(t)}{dt} = W(t) + (6t - 4) \int_0^1 W(s) ds + (6 - 12t) \int_0^1 sW(s) ds.$$

## VI.5 Conclusion et perspectives

Nous avons ainsi, à travers cette étude, élargi quelque peu le champ d'applicabilité de la procédure LMC, en généralisant les résultats à la tendance polynomiale quelconque puis à la multi-intégration dans un premier temps, en mettant en évidence dans un second temps une région où l'on ne rejette pas  $\mathcal{H}_0$ , à tort. Nous souhaitons, en guise de conclusion, soulever une difficulté majeure de l'étude. Nous avons admis, en effet, que nous disposions systématiquement d'une valeur connue de  $r$ , mais surtout d'un estimateur consistant de  $\theta$ . Or, comme nous avons pu le voir, il faut pour cela disposer d'un protocole expérimental strict, permettant de savoir différencier ou sommer le processus lorsque c'est nécessaire, et de tenir compte ou non d'une constante dans le modèle. Supposons pour illustrer cela que nous choissions de différencier un processus AR( $p$ ) muni d'une tendance d'ordre  $r$ , engendré par une racine unitaire localisée à  $\rho = -1$ . Nous estimerions alors le paramètre  $\theta$  dans le modèle ARIMA( $p, 1, 1$ ) suivant,

$$\mathcal{A}(L)\Delta Y_t = \gamma_0 + \gamma_1 t_T + \dots + \gamma_{r-1} t_T^{r-1} + \xi_t + \beta \xi_{t-1},$$

alors qu'en réalité, le processus générateur des données se trouve être

$$\mathcal{A}(L)\Delta Y_t = \gamma_0 + \gamma_1 t_T + \dots + \gamma_{r-1} t_T^{r-1} + \Delta(S_t^\eta + \varepsilon_t)$$

où  $(S_t^\eta)$  est le processus des sommes partielles alternées du processus  $(\eta_t)$ . Sous l'hypothèse  $\mathcal{H}_1$  : " $\sigma_\eta^2 > 0$ ", le résidu se réduit à

$$\Delta(S_t^\eta + \varepsilon_t) = \sum_{k=1}^{t-1} \left( (-1)^{t-k} - (-1)^{t-k-1} \right) \eta_k + \eta_t + \varepsilon_t - \varepsilon_{t-1} = 2 \sum_{k=1}^{t-1} (-1)^{t-k} \eta_k + \eta_t + \varepsilon_t - \varepsilon_{t-1}.$$

Au contraire, si l'opérateur de sommation est appliqué à un processus engendré par une racine unitaire située à  $\rho = 1$ , nous générons le résidu

$$\nabla(S_t^\eta + \varepsilon_t) = 2 \sum_{k=1}^{t-1} \eta_k + \eta_t + \varepsilon_t + \varepsilon_{t-1}.$$

Dans un cas comme dans l'autre, le résidu est non stationnaire et il n'y a aucun argument pour justifier la consistance de l'estimateur de  $\theta$ . De même, pour  $r = 0$ , l'opérateur  $\Delta$  élimine la constante du modèle alors que l'opérateur  $\nabla$  la double, la procédure d'estimation de  $\theta$  est en conséquence différente. Ces remarques doivent donc nous inciter à tester plusieurs configurations jusqu'à satisfaire les Théorèmes VI.1–VI.2. En outre, si nous considérons qu'il existe une valeur limite  $\theta^*$  à l'estimateur du maximum de vraisemblance de  $\theta$  en présence de perturbation intégrée, alors nous pourrions

probablement, en approfondissant les démonstrations, établir des principes d'invariance évoluant à des vitesses identiques, mais dont les lois limites seraient déformées par l'ampleur du biais entre  $\theta^*$  et  $\theta$ . Ceci est également une piste que nous avons commencé à considérer dans le cadre d'un travail futur. La procédure LMC gagne en vitesse dans la divergence de  $\widehat{K}_T$  en présence de non stationnarité stochastique par rapport à la procédure KPSS, mais cette dernière s'exempte des difficultés cruciales dont nous venons de parler. Nous pensons enfin que la généralisation présentée dans ce chapitre s'applique également au test KPSS, sous réserve d'adapter quelque peu les démonstrations. Il pourrait être également intéressant d'adopter une stratégie des moindres carrés pour estimer  $\theta$  et  $\alpha$  conjointement, et d'intégrer à nos calculs le biais qui résulterait de la présence d'une racine unitaire, comme nous l'avons fait dans le Chapitre III concernant l'autocorrélation résiduelle. Nous n'aurions plus ainsi à nous préoccuper de transformations préalables de données, ainsi que des risques inhérents que nous venons d'exprimer.

## VI.6 Annexe : démonstration des résultats

Dans toute la suite, nous considérerons sans le préciser que  $0 < \tau \leq 1$ . De plus, pour alléger les notations, nous supprimerons l'indice implicite des variables pour désigner la formulation vectorielle associée. À titre d'exemple,  $\varepsilon = (\varepsilon_1 \dots \varepsilon_T)'$ . Nous définissons de plus la matrice de « design »  $X$  d'ordre  $(r+1) \times T$  comme

$$X = \begin{pmatrix} 1 & 1 & \dots & 1 & \dots & 1 \\ 1_T & 2_T & \dots & k_T & \dots & 1 \\ \vdots & \vdots & & \vdots & & \vdots \\ 1_T^r & 2_T^r & \dots & k_T^r & \dots & 1 \end{pmatrix} \quad \text{avec par convention} \quad k_T = \frac{k}{T}. \quad (\text{VI.29})$$

Pour faciliter la lecture de nos preuves, nous devons commencer par établir un principe d'invariance sur les résidus de l'estimation par moindre carrés de la tendance polynomiale d'un processus muni d'une perturbation multi-intégrée. Dans un second temps, nous démontrerons nos résultats.

### VI.6.1 Un principe d'invariance en présence de multi-intégration

Le résultat que nous développons ici est une extension du Theorème 1(d) de [124]. Nous en retrouvons également les fondations dans [65], avec une perturbation plus générale mais  $\kappa = 0$ .

**Lemme VI.1.** *Pour tout  $1 \leq t \leq T$ , soit le modèle*

$$Z_t = \alpha_0 + \alpha_1 t_T + \dots + \alpha_r t_T^r + S_t^{(d)} + \varepsilon_t$$

avec  $d = d^+ \geq 1$  et  $\kappa \neq 0$ . Soit  $(\widehat{\varepsilon}_t)$  l'ensemble résiduel engendré par une estimation par moindres carrés du paramètre  $\alpha$ . Alors, nous avons la convergence en loi

$$\frac{\widehat{\varepsilon}_{[T\tau]}}{\sigma_\eta T^{d-1/2}} \xrightarrow{\mathcal{L}} W_{r,d-1}(\tau)$$

où  $W_{r,d-1}(t)$  est le processus de Wiener recentré d'ordre  $r \times (d-1)$ .

*Démonstration.* L'estimateur des moindres carrés du paramètre  $\alpha$  est donné par

$$\widehat{\alpha}_T = \left( \sum_{t=1}^T x_t x_t' \right)^{-1} \sum_{t=1}^T x_t Z_t = R_T^{-1} \sum_{t=1}^T x_t Z_t \quad (\text{VI.30})$$

où  $x_t$  est la  $t$ -ème colonne de  $X$  défini en (VI.29). Il s'ensuit que

$$\widehat{\alpha}_T - \alpha = R_T^{-1} P_T \quad \text{avec} \quad P_T = \sum_{t=1}^T x_t w_t \quad (\text{VI.31})$$

où le résidu  $w_t = S_t^{(d)} + \varepsilon_t$ . Commençons par établir un principe d'invariance applicable à la suite  $(w_t)$ . Tout d'abord, nous avons

$$\frac{S_{[T\tau]}^{(1)}}{\sigma_\eta \sqrt{T}} = \frac{1}{\sigma_\eta \sqrt{T}} \sum_{t=1}^{[T\tau]} \eta_t \xrightarrow{\mathcal{L}} W(\tau) \quad (\text{VI.32})$$

par le Théorème B.7, puisque l'on rappelle que l'on a ici  $\rho_1 = \rho_2 = \dots = \rho_d = 1$ . Par extension,

$$\frac{S_{[T\tau]}^{(2)}}{\sigma_\eta T^{3/2}} = \frac{1}{\sigma_\eta T^{3/2}} \sum_{t=1}^{[T\tau]} S_t^{(1)} = \sum_{t=1}^{[T\tau]} \int_{\frac{t}{T}}^{\frac{t+1}{T}} \frac{S_{[Ts]}^{(1)}}{\sigma_\eta \sqrt{T}} ds \xrightarrow{\mathcal{L}} \int_0^\tau W(s) ds \equiv W^{(1)}(\tau) \quad (\text{VI.33})$$

via le Théorème B.9. En itérant le processus, on obtient, pour tout  $d \geq 2$ ,

$$\frac{S_{[T\tau]}^{(d)}}{\sigma_\eta T^{d-1/2}} \xrightarrow{\mathcal{L}} \int_0^\tau \int_0^{s_1} \dots \int_0^{s_{d-2}} W(s_{d-1}) ds_{d-1} \dots ds_1 \equiv W^{(d-1)}(\tau). \quad (\text{VI.34})$$

Puisque  $\varepsilon_{[T\tau]} = o(T^{d-1/2})$  p.s. par la loi forte des grands nombres (Théorème B.1), il suit que  $(w_t)$  satisfait également le principe d'invariance décrit par (VI.34), dès que  $d \geq 1$ . Lorsque  $d = 1$ , nous pouvons identifier la distribution limite  $W^{(d-1)}$  et  $\sigma_\eta$  à  $W$  et  $\sqrt{\omega}$  dans l'Hypothèse 1(a) de [124]. De plus, la  $k$ -ème ligne de  $P_T$  donnée en (VI.31) vaut

$$P_{k,T} = \sum_{t=1}^T t_T^{k-1} w_t = \frac{1}{T^{k-1}} \sum_{t=1}^T t^{k-1} w_t. \quad (\text{VI.35})$$

Il s'agit désormais d'étudier la vitesse de convergence de  $P_{k,T}$ . Pour tout  $1 \leq i \leq d$ , notons  $\delta_k(i) = i + k - 1/2$ . Par l'intermédiaire de (VI.34), on a

$$\frac{1}{\sigma_\eta T^{\delta_k(d)}} \sum_{t=1}^{[T\tau]} t^{k-1} w_t = \sum_{t=1}^{[T\tau]} \int_{\frac{t}{T}}^{\frac{t+1}{T}} \frac{[Ts]^{k-1} w_{[Ts]}}{\sigma_\eta T^{k-1} T^{\delta_0(d)}} ds \xrightarrow{\mathcal{L}} \int_0^\tau s^{k-1} W^{(d-1)}(s) ds. \quad (\text{VI.36})$$

La combinaison de (VI.35) et de (VI.36) nous conduit, pour tout  $d \geq 1$ , à

$$\frac{P_{[T\tau]}}{\sigma_\eta T^{d+1/2}} \xrightarrow{\mathcal{L}} P_d(\tau) \quad (\text{VI.37})$$

où la distribution limite est donnée par (VI.28). En outre, par un calcul direct,

$$\lim_{T \rightarrow \infty} \frac{R_T}{T} = M \quad \text{et} \quad \lim_{T \rightarrow \infty} T R_T^{-1} = M^{-1} \quad (\text{VI.38})$$

où  $R_T$  est donné en (VI.30) et la matrice non singulière  $M$  est construite avec  $M_{ij} = 1/(i + j - 1)$  pour tous  $1 \leq i, j \leq r + 1$ . Ces convergences découlent de la renormalisation des termes de la tendance polynomiale par  $T$ , et justifient le souhait exprimé dans l'introduction de *simplifier le traitement des projections lors de la procédure d'estimation par moindres carrés*. Ainsi, de (VI.31), (VI.37) et de (VI.38), il vient un équivalent du théorème central limite,

$$\frac{\widehat{\alpha}_T - \alpha}{\sigma_\eta T^{d-1/2}} \xrightarrow{\mathcal{L}} M^{-1} P_d(1). \quad (\text{VI.39})$$

Il nous reste simplement à noter que

$$\frac{\widehat{\varepsilon}_{[T\tau]}}{T^{d-1/2}} = \frac{w_{[T\tau]}}{T^{d-1/2}} - \frac{(\widehat{\alpha}_T - \alpha)' x_{[T\tau]}}{T^{d-1/2}} \quad (\text{VI.40})$$

et à combiner (VI.34) avec (VI.39) pour conclure que, dès que  $d \geq 1$ ,

$$\frac{\widehat{\varepsilon}_{[T\tau]}}{\sigma_\eta T^{d-1/2}} \xrightarrow{\mathcal{L}} W^{(d-1)}(\tau) - P_d'(1) M^{-1} \Lambda(\tau) \equiv W_{r,d-1}(\tau)$$

à partir du Théorème B.9, où  $\Lambda(\tau) = (1 \quad \tau \quad \dots \quad \tau^r)'$  est la valeur limite de  $x_{[T\tau]}$ . Lorsque  $d = 1$ , cette convergence est stipulée par le Théorème 1(d) de [124]. Cela achève la preuve du Lemme VI.1. ★

## VI.6.2 Sur la statistique de test

Avant de démontrer nos résultats, nous formulons quelques précisions quant à la consistance de l'estimateur du maximum de vraisemblance dans les modèles ARMA non nécessairement inversibles. Le lemme suivant est important dans le cadre de notre étude, puisqu'il nous garantit la consistance

de l'estimateur des moindres carrés associé à  $\theta$ , sous  $\mathcal{H}_0$  comme sous  $\mathcal{H}_1$ .

**Lemme VI.2.** *Soit un processus  $(Z_t)$  stationnaire engendré par une modélisation ARMA( $p, q$ ) causale satisfaisant*

$$\mathcal{A}(L)Z_t = \mu + \mathcal{B}(L)\xi_t$$

où  $(\xi_t)$  est un bruit blanc de variance finie,  $\mu \in \mathbb{R}$  et, pour tout  $z \in \mathbb{C}$ ,  $\mathcal{A}(z) = 1 - \theta_1 z - \dots - \theta_p z^p$  est un polynôme causal et  $\mathcal{B}(z) = 1 + \beta_1 z + \dots + \beta_q z^q$ . Si  $\mathcal{B}(z) \neq 0$  pour tout  $z \in \mathbb{C}$  tel que  $|z| < 1$ , alors l'estimateur du maximum de vraisemblance  $(\widetilde{\theta}_T, \widetilde{\beta}_T)$  des paramètres  $(\theta, \beta)$  est consistant.

**Démonstration.** Lorsque le polynôme  $\mathcal{B}$  n'a aucune racine à l'intérieur du cercle unité, la modélisation est causale et inversible, et le résultat est donné par le Théorème 10.8.1 de [26]. Lorsque  $\mathcal{B}$  possède au moins une racine unitaire, le résultat découle cette fois du Théorème 2.1 de [110]. ★

### VI.6.2.1 Preuve du Théorème VI.1

Notons  $P = X'(XX')^{-1}X$  la matrice de projection orthogonale sur le sous-espace vectoriel de  $\mathbb{R}^T$  engendré par les vecteurs colonnes de la matrice  $X$ , et  $I$  la matrice identité d'ordre  $T$ . Nous commençons par exprimer  $(\widehat{\varepsilon}_t)$  en termes de  $(\varepsilon_t)$  afin d'établir un principe d'invariance axé sur le Théorème B.7, applicable à la suite  $(S_t)$  donnée par (VI.18). Nous nous plaçons tout d'abord dans le cas où  $\kappa \neq 0$ . Des définitions (VI.15) et (VI.17), et puisque  $\widehat{\alpha}_T$  est l'estimateur des moindres carrés du paramètre  $\alpha$ , nous voyons que, pour tout  $1 \leq t \leq T$ ,

$$\widehat{\varepsilon}_t = \widetilde{Y}_t - \widehat{\alpha}_0 - \widehat{\alpha}_1 t_T - \dots - \widehat{\alpha}_r t_T^r = \sum_{i=1}^p (\theta_i - \widetilde{\theta}_i) u_{i,t} + u_t \quad (\text{VI.41})$$

où  $u_t$  est la  $t$ -ème composante de  $(I-P)\varepsilon$ , et, pour  $1 \leq i \leq p$ ,  $u_{i,t}$  est la  $t$ -ème composante de  $(I-P)Y_{-i}$  avec  $Y_{-i} = (Y_{1-i} \dots Y_{T-i})'$ . On tire alors du Théorème 1 de [87] que l'on a la convergence en loi

$$\frac{1}{\sigma_\varepsilon \sqrt{T}} \sum_{t=1}^{\lfloor T\tau \rfloor} u_t \xrightarrow{\mathcal{L}} B_r(\tau). \quad (\text{VI.42})$$

De plus, pour tout  $1 \leq i \leq p$  et puisque  $\mathcal{A}$  est causal (et donc inversible), l'expression du modèle (VI.11) nous conduit à

$$Y_{t-i} = \mathcal{A}^{-1}(L)(\alpha_0 + \alpha_1(t-i)_T + \dots + \alpha_r(t-i)_T^r) + \mu_{t-i} \quad (\text{VI.43})$$

où  $(t-i)_T = (t-i)/T$  et  $\mathcal{A}(L)\mu_{t-i} = \varepsilon_{t-i}$ . Les coefficients de la tendance déterministe sont facilement identifiables. Il suit que  $(\mu_t)$  est un processus AR( $p$ ) stationnaire qui vérifie également un principe

d'invariance, en vertu du Théorème B.8. Si nous définissons sa variance de long terme par

$$\sigma_\mu^2 = \mathbb{E}[\mu_0^2] + 2 \sum_{t=1}^{\infty} \mathbb{E}[\mu_0 \mu_t],$$

qui est de plus finie (voir le Chapitre 3 de [26]), alors, en utilisant de nouveau le Théorème 1 de [87], pour tout  $1 \leq i \leq p$ ,

$$\frac{1}{\sigma_\mu \sqrt{T}} \sum_{t=1}^{[T\tau]} u_{i,t} \xrightarrow{\mathcal{L}} B_r(\tau). \quad (\text{VI.44})$$

La combinaison de (VI.44) et du Lemme VI.2 implique que

$$\frac{1}{\sigma_\mu \sqrt{T}} \sum_{i=1}^p (\theta_i - \tilde{\theta}_i) \sum_{t=1}^{[T\tau]} u_{i,t} \xrightarrow{\mathcal{P}} 0. \quad (\text{VI.45})$$

En remarquant que  $(S_t)$  est le processus des sommes partielles associé à  $(\varepsilon_t)$ , nous en déduisons que nous avons la convergence en loi

$$\frac{S_{[T\tau]}}{\sigma_\varepsilon \sqrt{T}} \xrightarrow{\mathcal{L}} B_r(\tau). \quad (\text{VI.46})$$

En outre, il n'est pas compliqué de voir que

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T u_t^2 = \sigma_\varepsilon^2 \quad \text{p.s.}$$

puisque  $(u_t)$  peut être vu comme le résidu d'une régression de  $(\varepsilon_t)$  sur une tendance polynomiale dont les coefficients sont nuls. Le même type de convergence est atteinte pour  $(u_{i,t})$ , suivant une méthodologie similaire à celle utilisée dans [108], puisque  $(u_{i,t})$  peut également être vu comme le résidu de la régression du processus stationnaire  $(\mu_t)$  sur une tendance polynomiale dont l'estimation reste consistante. C'est pourquoi il est possible d'obtenir, par l'inégalité de Cauchy-Schwarz, que

$$\lim_{T \rightarrow \infty} \frac{Q_T}{T} = \sigma_\varepsilon^2 \quad \text{p.s.} \quad (\text{VI.47})$$

où  $Q_T$  est donné par (VI.18). Finalement,

$$\frac{1}{\sigma_\varepsilon^2 T^2} \sum_{t=1}^{[T\tau]} S_t^2 = \frac{1}{T} \sum_{t=1}^{[T\tau]} \left( \frac{S_t}{\sigma_\varepsilon \sqrt{T}} \right)^2 = \sum_{t=1}^{[T\tau]} \int_{\frac{t}{T}}^{\frac{t+1}{T}} \left( \frac{S_{[Ts]}}{\sigma_\varepsilon \sqrt{T}} \right)^2 ds \xrightarrow{\mathcal{L}} \int_0^\tau B_r^2(s) ds$$

par application du Théorème B.9. Cela achève la preuve du Théorème VI.1, en vertu de (VI.46), de (VI.47), du lemme de Slutsky et en choisissant  $\tau = 1$ , dans le cas où  $\kappa \neq 0$ . Considérons maintenant le cas plus trivial où  $\kappa = 0$ . Alors,  $P$  est la matrice nulle et nous obtenons simplement  $u_t = \varepsilon_t$  et  $u_{i,t} = Y_{-i}$

dans la relation (VI.41), pour tout  $1 \leq t \leq T$  et  $1 \leq i \leq p$ . Dès lors, la convergence (VI.47) se réduit à une application directe de la loi forte des grands nombres (Théorème B.1) et, *via* le Théorème B.7, le principe d'invariance (VI.46) devient

$$\frac{S_{[T\tau]}}{\sigma_\varepsilon \sqrt{T}} \xrightarrow{\mathcal{L}} W(\tau). \quad (\text{VI.48})$$

La fin du raisonnement suit alors de la même manière que pour  $\kappa \neq 0$ . ★

### VI.6.2.2 Preuve du Théorème VI.2

Supposons maintenant que  $\sigma_\eta^2 > 0$ , ce qui implique que le processus possède une non stationnarité de nature stochastique intervenant dans la marche aléatoire  $(S_t^\eta)$  donnée en (VI.13). Commençons tout d'abord par traiter le cas où  $\kappa \neq 0$ . De la même manière que pour l'expression résiduelle (VI.41) sous  $\mathcal{H}_0$ , nous obtenons ici

$$\widehat{\varepsilon}_t = \widetilde{Y}_t - \widehat{\alpha}_0 - \widehat{\alpha}_1 t_T - \dots - \widehat{\alpha}_r t_T^r = \sum_{i=1}^p (\theta_i - \widetilde{\theta}_i) u_{i,t} + u_{\eta,t} \quad (\text{VI.49})$$

où  $u_{\eta,t}$  est la  $t$ -ème composante de  $(I - P)(S^\eta + \varepsilon)$  et, pour tout  $1 \leq i \leq p$ ,  $u_{i,t}$  est la  $t$ -ème composante de  $(I - P)Y_{-i}$  avec  $Y_{-i}$  donné, pour tout  $1 \leq t \leq T$ , par

$$Y_{t-i} = \mathcal{A}^{-1}(L)(\alpha_0 + \alpha_1(t-i)_T + \dots + \alpha_r(t-i)_T^r) + T_{t-i}^\eta \quad (\text{VI.50})$$

et  $\mathcal{A}(L)T_{t-i}^\eta = S_{t-i}^\eta + \varepsilon_{t-i}$ , reprenant les notations de (VI.43). En conséquence,  $((1 - \rho L)T_{t-i}^\eta)$  correspond à un processus ARMA( $p, 1$ ) stationnaire, ce qui entraîne *de facto* que  $(T_{t-i}^\eta)$  satisfait un principe d'invariance, en vertu du Théorème B.8, dans lequel sa variance de long terme est impliquée. La vitesse associée est dès lors en  $\sqrt{T}$ . Ainsi, par un calcul simple et le Théorème B.9, nous voyons que les suites  $(u_{i,t})$  et  $(u_{\eta,t})$  se comportent de la même manière, et tout principe d'invariance applicable à l'une se transfère immédiatement à l'autre, à une constante multiplicative près. Il faut bien remarquer ici que la projection  $I - P$  annule de fait l'influence de la tendance polynomiale intervenant dans (VI.50). Cependant, grâce au Lemme VI.2, il suit que les résultats asymptotiques seront entièrement déterminés par le comportement de  $(u_{\eta,t})$ , de  $(u_{\eta,t}^2)$  ainsi que de leurs processus des sommes partielles. Tout d'abord, par le Théorème B.7 et lorsque  $\rho = 1$ , on a déjà vu en (VI.32) que nous avons le principe d'invariance

$$\frac{S_{[T\tau]}^\eta}{\sigma_\eta \sqrt{T}} \xrightarrow{\mathcal{L}} W(\tau). \quad (\text{VI.51})$$

Pour  $\rho = -1$ , nous ne pouvons pas appliquer directement le Théorème B.7 puisque la marche  $(S_t^\eta)$  ne repose plus sur une somme de variables aléatoires identiquement distribuées. Néanmoins, le Théorème B.8 s'applique en lieu et place, et la convergence (VI.51) reste valide. Selon la valeur de  $\rho$ ,

le raisonnement emprunte alors des voies totalement différentes. Dans un premier temps, si nous considérons que  $\rho = 1$ , le Lemme VI.1 appliqué avec  $d = 1$  nous conduit à la convergence

$$\frac{u_{\eta, [T\tau]} }{\sigma_{\eta} \sqrt{T}} \xrightarrow{\mathcal{L}} W_{r,0}(\tau). \quad (\text{VI.52})$$

Il s'ensuit que

$$\frac{1}{\sigma_{\eta} T^{3/2}} \sum_{t=1}^{[T\tau]} u_{\eta,t} = \sum_{t=1}^{[T\tau]} \int_{\frac{t}{T}}^{\frac{t+1}{T}} \frac{u_{\eta, [Ts]} }{\sigma_{\eta} \sqrt{T}} ds \xrightarrow{\mathcal{L}} \int_0^{\tau} W_{r,0}(s) ds \equiv C_{r,1}(\tau) \quad (\text{VI.53})$$

en vertu du Théorème B.9. Puisque le terme dominant de  $\widehat{\varepsilon}_t$  s'avère être  $u_{\eta,t}$ , ainsi que nous l'avons expliqué ci-dessus, et par l'intermédiaire de la convergence (VI.52), le processus des sommes partielles ( $S_t$ ) donné par (VI.18) satisfait le principe d'invariance décrit par la convergence

$$\frac{S_{[T\tau]} }{\sigma_{\eta} T^{3/2}} \xrightarrow{\mathcal{L}} C_{r,1}(\tau). \quad (\text{VI.54})$$

Remarquons qu'il est également possible d'obtenir (VI.54) en passant par le Théorème 1 de [87] que l'on combine à la convergence

$$\frac{1}{\sigma_{\eta} T^{3/2}} \sum_{t=1}^{[T\tau]} S_t^{\eta} = \sum_{t=1}^{[T\tau]} \int_{\frac{t}{T}}^{\frac{t+1}{T}} \frac{S_{[Ts]}^{\eta} }{\sigma_{\eta} \sqrt{T}} ds \xrightarrow{\mathcal{L}} \int_0^{\tau} W(s) ds \equiv W^{(1)}(\tau) \quad (\text{VI.55})$$

déjà établie en (VI.33). Naturellement, (VI.47) qui découlait directement sous  $\mathcal{H}_0$  de la loi forte des grands nombres, est nécessairement faussé sous  $\mathcal{H}_1$ . Cependant, on déduit de (VI.52) que

$$\frac{1}{\sigma_{\eta}^2 T^2} \sum_{t=1}^{[T\tau]} u_{\eta,t}^2 = \sum_{t=1}^{[T\tau]} \int_{\frac{t}{T}}^{\frac{t+1}{T}} \left( \frac{u_{\eta, [Ts]} }{\sigma_{\eta} \sqrt{T}} \right)^2 ds \xrightarrow{\mathcal{L}} \int_0^{\tau} W_{r,0}^2(s) ds$$

ce qui implique que

$$\frac{Q_{[T\tau]} }{\sigma_{\eta}^2 T^2} \xrightarrow{\mathcal{L}} \int_0^{\tau} W_{r,0}^2(s) ds. \quad (\text{VI.56})$$

En outre, il découle de (VI.54) que

$$\frac{1}{\sigma_{\eta}^2 T^4} \sum_{t=1}^{[T\tau]} S_t^2 = \frac{1}{T} \sum_{t=1}^{[T\tau]} \left( \frac{S_t}{\sigma_{\eta} T^{3/2}} \right)^2 = \sum_{t=1}^{[T\tau]} \int_{\frac{t}{T}}^{\frac{t+1}{T}} \left( \frac{S_{[Ts]} }{\sigma_{\eta} T^{3/2}} \right)^2 ds \xrightarrow{\mathcal{L}} \int_0^{\tau} C_{r,1}^2(s) ds.$$

Cette dernière convergence, combinée à (VI.56) et au Théorème B.9, permet d'achever la première partie de la preuve, en choisissant  $\tau = 1$ . Dans un second temps, si nous nous plaçons sous l'alternative  $\rho = -1$ , les choses se compliquent. Ceci est principalement dû à un phénomène de compensation dans la somme alternée ( $S_t^{\eta}$ ). En effet, nous voyons que, pour toute valeur de  $t$  respectivement paire

et impaire, nous avons

$$\sum_{k=1}^t S_k^\eta = \sum_{k=1}^{t/2} \eta_{2k} \quad \text{et} \quad \sum_{k=1}^t S_k^\eta = \sum_{k=1}^{(t+1)/2} \eta_{2k-1}.$$

Soit  $(\zeta_t)$  la suite définie pour  $T$  pair et  $1 \leq t \leq T/2$ , par

$$\zeta_t = \varepsilon_{2t-1} + \varepsilon_{2t} + \eta_{2t}$$

et, pour  $T$  impair et  $1 \leq t \leq (T+1)/2$ , par

$$\zeta_t = \varepsilon_{2t-1} + \varepsilon_{2(t-1)} + \eta_{2t-1}.$$

On a  $\mathbb{E}[\zeta_t] = 0$ ,  $\mathbb{E}[\zeta_t^2] = 2\sigma_\varepsilon^2 + \sigma_\eta^2$  et toutes les covariances sont nulles, puisque les suites  $(\varepsilon_t)$  et  $(\eta_t)$  sont mutuellement indépendantes. Il s'ensuit que  $(\zeta_t)$  est un bruit blanc, et qu'il vérifie en conséquence, à partir du Théorème B.7, le principe d'invariance

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[T\tau]} \zeta_t \xrightarrow{\mathcal{L}} \sqrt{2\sigma_\varepsilon^2 + \sigma_\eta^2} W(\tau). \quad (\text{VI.57})$$

Ainsi, nous obtenons les principes d'invariance

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[T\tau]} (S_t^\eta + \varepsilon_t) = \frac{1}{\sqrt{T}} \sum_{t=1}^{[T\tau/2]} \zeta_t \xrightarrow{\mathcal{L}} \sqrt{2\sigma_\varepsilon^2 + \sigma_\eta^2} W\left(\frac{\tau}{2}\right) \stackrel{\mathcal{L}}{=} \sqrt{\frac{2\sigma_\varepsilon^2 + \sigma_\eta^2}{2}} W(\tau)$$

et, par application du Théorème 1 de [87],

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[T\tau]} u_{\eta,t} \xrightarrow{\mathcal{L}} \sqrt{\frac{2\sigma_\varepsilon^2 + \sigma_\eta^2}{2}} B_r(\tau). \quad (\text{VI.58})$$

En exploitant cette dernière convergence et la domination de  $u_{\eta,t}$  dans  $\widehat{\varepsilon}_t$  (l'estimateur de  $\theta$  restant consistant), il advient que

$$\frac{1}{T^2} \sum_{t=1}^{[T\tau]} S_t^2 = \sum_{t=1}^{[T\tau]} \int_{\frac{t}{T}}^{\frac{t+1}{T}} \left( \frac{S_{[Ts]}}{\sqrt{T}} \right)^2 ds \xrightarrow{\mathcal{L}} \frac{2\sigma_\varepsilon^2 + \sigma_\eta^2}{2} \int_0^\tau B_r^2(s) ds. \quad (\text{VI.59})$$

Reprenons désormais le raisonnement développé dans le Lemme VI.1, mais pour  $d = 1$  et  $\rho = -1$ . Rappelons que, avec les notations associées à (VI.35), nous avons, pour  $1 \leq k \leq r+1$ ,

$$P_{k,T} = \sum_{t=1}^T t_T^{k-1} w_t = \frac{1}{T^{k-1}} \sum_{t=1}^T t^{k-1} (S_t^\eta + \varepsilon_t).$$

Tout d'abord, nous voyons que

$$M_T^k = \sum_{t=1}^T t^{k-1} \varepsilon_t$$

est une martingale adaptée à la filtration naturelle du processus  $(\varepsilon_t)$ , dont le crochet est tel que  $\langle M^k \rangle_T = O(T^{2k-1})$  p.s. Nous tirons donc de la loi forte des grands nombres pour les martingales (voir par exemple [45]) que  $M_T^k = o(T^k)$  p.s. Ainsi,

$$\frac{P_{k,T}}{T} = \frac{1}{T^k} \sum_{t=1}^T t^{k-1} S_t^\eta + o(1) \quad \text{p.s.} \quad (\text{VI.60})$$

Par ailleurs, soit  $(\Sigma_t^\eta)$  le processus des sommes partielles associé à  $(\eta_t)$  pour  $\rho = 1$ . Soient encore  $(\Lambda_t^\eta)$  et  $(\Pi_t^\eta)$  les processus des sommes partielles associés à  $(\eta_t)$ , respectivement pour les indices pairs et impairs. Il est clair que

$$\Lambda_{p_t}^\eta = \eta_2 + \eta_4 + \dots + \eta_{2p_t} = \sum_{\ell=1}^{p_t} \eta_{2\ell} \quad \text{et} \quad \Pi_{i_t}^\eta = \eta_1 + \eta_3 + \dots + \eta_{2i_t-1} = \sum_{\ell=1}^{i_t} \eta_{2\ell-1}$$

avec  $i_t = [(t+1)/2]$  et  $p_t = t - [(t+1)/2]$ . Un calcul direct nous montre alors que, lorsque  $\rho = -1$  et pour tout  $1 \leq k \leq r+1$ ,

$$\sum_{t=1}^T t^{k-1} S_t^\eta = \sum_{t=1}^T t^{k-1} \Sigma_t^\eta - 2 \sum_{t=1}^{p_T} (2t+1)^{k-1} \Lambda_t^\eta - 2 \sum_{t=1}^{i_T} (2t)^{k-1} \Pi_t^\eta + 2r_T \quad (\text{VI.61})$$

où l'on a respectivement  $r_T = (T+1)^{k-1} \Pi_{(T+1)/2}^\eta$  pour  $T$  impair et  $r_T = (T+1)^{k-1} \Lambda_{T/2}$  pour  $T$  pair. Il est possible, *via* le Théorème B.7, d'établir un principe d'invariance pour les processus  $(\Lambda_t^\eta)$  et  $(\Pi_t^\eta)$ . En effet,

$$\frac{\Lambda_{[p_T \tau]}^\eta}{\sigma_\eta \sqrt{p_T}} \xrightarrow{\mathcal{L}} W(\tau) \quad \text{et} \quad \frac{\Pi_{[i_T \tau]}^\eta}{\sigma_\eta \sqrt{i_T}} \xrightarrow{\mathcal{L}} W(\tau). \quad (\text{VI.62})$$

Il s'ensuit que, par le Théorème B.9,

$$\frac{1}{\sigma_\eta p_T^{k+1/2}} \sum_{t=1}^{[p_T \tau]} (2t+1)^{k-1} \Lambda_t^\eta = \sum_{t=1}^{[p_T \tau]} \int_{\frac{t}{p_T}}^{\frac{t+1}{p_T}} \frac{(2[p_T s] + 1)^{k-1} \Lambda_{[p_T s]}^\eta}{\sigma_\eta p_T^{k-1} \sqrt{p_T}} ds \xrightarrow{\mathcal{L}} \int_0^\tau (2s)^{k-1} W(s) ds \quad (\text{VI.63})$$

et que

$$\frac{1}{\sigma_\eta i_T^{k+1/2}} \sum_{t=1}^{[i_T \tau]} (2t)^{k-1} \Pi_t^\eta = \sum_{t=1}^{[i_T \tau]} \int_{\frac{t}{i_T}}^{\frac{t+1}{i_T}} \frac{(2[i_T s])^{k-1} \Pi_{[i_T s]}^\eta}{\sigma_\eta i_T^{k-1} \sqrt{i_T}} ds \xrightarrow{\mathcal{L}} \int_0^\tau (2s)^{k-1} W(s) ds \quad (\text{VI.64})$$

puisque'il est facile de voir que  $p_T$  et  $i_T$  se comportent comme  $T/2$ . Par ailleurs, les convergences (VI.62) ainsi que la définition de  $r_T$  nous donnent immédiatement

$$\frac{r_T}{T^{k+1/2}} \xrightarrow{\mathcal{P}} 0. \quad (\text{VI.65})$$

En outre, le principe d'invariance (VI.36) pour  $\rho = 1$  et  $d = 1$ , correspondant ici à celui associé à  $(\Sigma_t^\eta)$ , nous conduit directement, combiné avec (VI.61), (VI.63), (VI.64) et (VI.65), à

$$\frac{1}{T^{k+1/2}} \sum_{t=1}^T t^{k-1} S_t^\eta = O_{\mathcal{P}}(1)$$

et donc, en reprenant les notations du Lemme VI.1, pour tout  $1 \leq k \leq r + 1$ ,

$$\frac{P_{k,T}}{T^{3/2}} = O_{\mathcal{P}}(1) \quad \text{et} \quad \frac{u_{\eta,T}}{\sqrt{T}} = \frac{S_T^\eta + \varepsilon_T}{\sqrt{T}} + O_{\mathcal{P}}(1),$$

en utilisant successivement (VI.31) et (VI.40). Par les Théorèmes B.1–B.7 et de nouveau le Théorème B.9, nous en déduisons facilement, suivant le même raisonnement que précédemment, que  $Q_T$  grandit avec la vitesse  $T^2$  et cela achève la preuve pour  $\rho = -1$ . Finalement, dans le cas plus trivial où  $\kappa = 0$ , le principe d'invariance (VI.52) devient

$$\frac{u_{\eta,[T\tau]}}{\sigma_\eta \sqrt{T}} \xrightarrow{\mathcal{L}} W(\tau) \quad (\text{VI.66})$$

par l'intermédiaire du Théorème B.7, et la fin du raisonnement découle facilement de la même manière que pour  $\kappa \neq 0$ . ★

### VI.6.2.3 Preuve de la Proposition VI.1

Cette preuve sera très rapide puisque tous les outils nécessaires ont été développés dans la preuve précédente. En effet, pour  $\kappa = 0$  et  $\rho = -1$ , si nous reprenons la définition de la suite  $(\zeta_t)$  ainsi que son principe d'invariance en (VI.57), il vient

$$\frac{1}{T^2} \sum_{t=1}^{[T\tau]} S_t^2 \xrightarrow{\mathcal{L}} \sigma_\varepsilon^2 \int_0^\tau W_\varepsilon^2(s) ds + \frac{\sigma_\eta^2}{2} \int_0^\tau W_\eta^2(s) ds,$$

si nous décomposons la loi limite en deux composantes indépendantes, afin d'en faciliter le traitement par la suite. En l'absence d'estimation tendancielle, on a de plus  $u_{\eta,t} = S_t^\eta + \varepsilon_t$ , pour  $1 \leq t \leq T$ . Cela nous conduit, de la même manière que précédemment, à

$$\frac{Q_{[T\tau]} \xrightarrow{\mathcal{L}} \int_0^\tau W_\eta^2(s) ds.$$

Il nous reste à utiliser le Théorème B.9 puis à choisir  $\tau = 1$  pour achever la preuve. ★

#### VI.6.2.4 Preuve de la Proposition VI.2

Nous anticipons ici quelque peu, car nous avons besoin d'un résultat que nous prouverons par la suite. En effet, si nous appliquons la Proposition VI.4 à la configuration  $\sigma_\varepsilon^2 = 0$ ,  $\rho = 1$  et  $d = 1$ , alors on obtient la convergence presque sûre

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T (\Delta \widehat{\varepsilon}_t)^2 = \sigma_\eta^2 \quad \text{p.s.} \quad (\text{VI.67})$$

En suivant le même raisonnement, nous établissons aussi que, avec  $\rho = -1$ ,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T (\nabla \widehat{\varepsilon}_t)^2 = \sigma_\eta^2 \quad \text{p.s.} \quad (\text{VI.68})$$

Par ailleurs, nous avons déjà montré que, lorsque  $\kappa = 0$ , on a le principe d'invariance

$$\frac{Q_{[T\tau]} \xrightarrow{\mathcal{L}} \int_0^\tau W^2(s) ds \quad (\text{VI.69})$$

où le processus des sommes partielles  $(Q_t)$  est donné par (VI.18). La combinaison de (VI.69) avec (VI.67) pour  $\rho = 1$ , avec (VI.68) pour  $\rho = -1$ , nous permet d'achever la preuve, par le lemme de Slutsky, pour  $\tau = 1$ .

#### VI.6.2.5 Preuve du Théorème VI.3 et de la Proposition VI.3

Décomposons  $(\widehat{\varepsilon}_t)$  de la même manière que dans l'expression (VI.49) avec  $\kappa \neq 0$  dans un premier temps, pour obtenir

$$\widehat{\varepsilon}_t = \widetilde{Y}_t - \widehat{\alpha}_0 - \widehat{\alpha}_1 t_T - \dots - \widehat{\alpha}_r t_T^r = \sum_{i=1}^p (\theta_i - \widetilde{\theta}_i) u_{i,t} + u_{\eta,t} \quad (\text{VI.70})$$

où  $u_{\eta,t}$  est la  $t$ -ème composante de  $(I-P)(S^{(d)} + \varepsilon)$ , reprenant les notations de (VI.23). Ainsi, le Lemme VI.1 nous fournit le principe d'invariance

$$\frac{u_{\eta,[T\tau]}]}{\sigma_{\eta} T^{d-1/2}} \xrightarrow{\mathcal{L}} W_{r,d-1}(\tau) \quad (\text{VI.71})$$

dans le cas où  $\rho_1 = \rho_2 = \dots = \rho_d = 1$ , que nous considérons tout d'abord. À l'instar de nos raisonnements précédents et en vertu du Lemme VI.2, nous savons que le terme dominant dans (VI.70) est  $u_{\eta,t}$  et que le comportement asymptotique de la statistique de test sera déterminé par (VI.71). En conséquence, nous établissons la convergence en loi

$$\frac{1}{\sigma_{\eta} T^{d+1/2}} \sum_{t=1}^{[T\tau]} u_{\eta,t} = \sum_{t=1}^{[T\tau]} \int_{\frac{t}{T}}^{\frac{t+1}{T}} \frac{u_{\eta,[Ts]}}{\sigma_{\eta} T^{d-1/2}} ds \xrightarrow{\mathcal{L}} \int_0^{\tau} W_{r,d-1}(s) ds \equiv C_{r,d}(\tau) \quad (\text{VI.72})$$

par le Théorème B.9. Une fois encore, cette convergence peut également être obtenue *via* le Théorème 1 de [87] combiné à la convergence

$$\frac{1}{\sigma_{\eta} T^{d+1/2}} \sum_{t=1}^{[T\tau]} S_t^{(d)} = \sum_{t=1}^{[T\tau]} \int_{\frac{t}{T}}^{\frac{t+1}{T}} \frac{S_{[Ts]}^{(d)}}{\sigma_{\eta} T^{d-1/2}} ds \xrightarrow{\mathcal{L}} \int_0^{\tau} W^{(d-1)}(s) ds \equiv W^{(d)}(\tau) \quad (\text{VI.73})$$

établie à partir de (VI.34). En conséquence, nous avons un principe d'invariance associé au processus des sommes partielles ( $S_t$ ) donné par (VI.18). Ce dernier s'écrit

$$\frac{S_{[T\tau]}}{\sigma_{\eta} T^{d+1/2}} \xrightarrow{\mathcal{L}} C_{r,d}(\tau). \quad (\text{VI.74})$$

Le comportement asymptotique de ( $Q_t$ ) également donné par (VI.18) est relativement aisé à établir, en utilisant (VI.71). En effet,

$$\frac{1}{\sigma_{\eta}^2 T^{2d}} \sum_{t=1}^{[T\tau]} u_{\eta,t}^2 = \sum_{t=1}^{[T\tau]} \int_{\frac{t}{T}}^{\frac{t+1}{T}} \left( \frac{u_{\eta,[Ts]}}{\sigma_{\eta} T^{d-1/2}} \right)^2 ds \xrightarrow{\mathcal{L}} \int_0^{\tau} W_{r,d-1}^2(s) ds$$

ce qui implique que

$$\frac{Q_{[T\tau]}}{\sigma_{\eta}^2 T^{2d}} \xrightarrow{\mathcal{L}} \int_0^{\tau} W_{r,d-1}^2(s) ds. \quad (\text{VI.75})$$

D'autre part,

$$\frac{1}{\sigma_{\eta}^2 T^{2(d+1)}} \sum_{t=1}^{[T\tau]} S_t^2 = \sum_{t=1}^{[T\tau]} \int_{\frac{t}{T}}^{\frac{t+1}{T}} \left( \frac{S_{[Ts]}}{\sigma_{\eta} T^{d+1/2}} \right)^2 ds \xrightarrow{\mathcal{L}} \int_0^{\tau} C_{r,d}^2(s) ds \quad (\text{VI.76})$$

à partir de (VI.74). Cette convergence, combinée à (VI.75) et au Théorème B.9, achève la première partie de la démonstration, en choisissant  $\tau = 1$ . Par ailleurs, la Proposition VI.3 est quant à elle prou-

## Stationnarité ou multi-intégration ?

---

vée. Considérons maintenant le cas beaucoup plus complexe où  $\rho_1 = \rho_2 = \dots = \rho_d = -1$ . Le processus résiduel  $(S_t^{(d)})$ , une fois sommé, fait émerger un grand nombre de compensations. En effet, en écrivant explicitement la somme en question, nous voyons que, pour toute valeur de  $t$  respectivement paire et impaire, nous avons

$$\sum_{k=1}^t S_k^{(d)} = \sum_{k=1}^{t/2} S_{2k}^{(d-1)} \quad \text{et} \quad \sum_{k=1}^t S_k^{(d)} = \sum_{k=1}^{(t+1)/2} S_{2k-1}^{(d-1)}.$$

On montre tout d'abord, grâce aux Théorèmes B.8–B.9, que

$$\frac{S_{[T\tau]}^{(d)}}{\sigma_\eta T^{d-1/2}} \xrightarrow{\mathcal{L}} W^{(d-1)}(\tau). \quad (\text{VI.77})$$

Pour une valeur paire de  $T$ , considérons la suite  $(\zeta_t)$  définie, pour  $1 \leq t \leq T/2$ , par

$$\zeta_t = S_{2t}^{(d-2)} - S_{2t-1}^{(d-2)} = -2S_{2t-1}^{(d-2)} + S_{2t}^{(d-3)}$$

avec  $d \geq 3$ . Pour  $d = 1$ , le sujet est traité dans la preuve du Théorème VI.2 alors que, pour  $d = 2$ ,

$$\zeta_t = \eta_{2t} - \eta_{2t-1}$$

et, par convention,  $S_t^{(0)} = \eta_t$ , pour tout  $t$ . Nous tirons donc de (VI.77) qu'il existe une constante identifiable  $\delta \neq 0$  telle que, pour  $d \geq 3$ ,

$$\frac{\zeta_{[T\tau]}}{\sigma_\eta T^{d-5/2}} \xrightarrow{\mathcal{L}} \delta W^{(d-3)}(\tau).$$

Par un calcul direct et toujours lorsque  $T$  est pair, nous avons de plus

$$\sum_{t=1}^{[T\tau]} S_t^{(d)} = \sum_{t=1}^{[T\tau/2]} S_{2t}^{(d-1)} = \sum_{t=1}^{[T\tau/2]} \sum_{k=1}^t \zeta_k,$$

ce qui nous conduit à

$$\frac{1}{\sigma_\eta T^{d-3/2}} \sum_{t=1}^{[T\tau]} \zeta_t = \sum_{t=1}^{[T\tau]} \int_{\frac{t}{T}}^{\frac{t+1}{T}} \frac{\zeta_{[Ts]}}{\sigma_\eta T^{d-5/2}} ds \xrightarrow{\mathcal{L}} \delta W^{(d-2)}(\tau)$$

par le Théorème B.9. Finalement,

$$\frac{1}{\sigma_\eta T^{d-1/2}} \sum_{t=1}^{[T\tau]} S_t^{(d)} = \frac{1}{\sigma_\eta T^{d-1/2}} \sum_{t=1}^{[T\tau/2]} \sum_{k=1}^t \zeta_k = \sum_{t=1}^{[T\tau/2]} \int_{\frac{t}{T}}^{\frac{t+1}{T}} \frac{1}{\sigma_\eta T^{d-3/2}} \sum_{k=1}^{[Ts]} \zeta_k ds \xrightarrow{\mathcal{L}} \delta \int_0^{\tau/2} W^{(d-2)}(s) ds.$$

Si maintenant  $d = 2$ , nous obtenons le même principe d'invariance et il en va bien entendu de même lorsque  $T$  est impair, sous réserve d'une définition légèrement différente de  $(\zeta_t)$ . Sans avoir besoin d'identifier  $\delta$ , cela nous montre que le processus  $(S_t)$  se comporte comme  $T^{d-1/2}$  lorsque  $\rho_1 = \rho_2 = \dots = \rho_d = -1$ , et donc que le numérateur de  $\widehat{K}_T$  se comporte comme  $T^{2d}$  par les techniques de calcul maintes fois utilisées jusque là (sans oublier le Lemme VI.2). Nous allons traiter le dénominateur exactement comme nous l'avions fait dans la preuve du Théorème VI.2, pour  $\rho = -1$  et  $d = 1$ . Nous avons déjà établi que, pour  $1 \leq k \leq r + 1$ ,

$$\frac{P_{k,T}}{T} = \frac{1}{T^k} \sum_{t=1}^T t^{k-1} S_t^{(d)} + o(1) \quad \text{p.s.}$$

en reprenant les notations adéquates. Soit  $(\Sigma_t^\eta)$  le processus des sommes partielles associé à  $(S_t^{(d-1)})$  pour  $\rho_1 = \rho_2 = \dots = \rho_d = 1$ . Soient encore  $(\Lambda_t^\eta)$  et  $(\Pi_t^\eta)$  les processus des sommes partielles associés à  $(S_t^{(d-1)})$ , respectivement pour les indices pairs et impairs. Explicitement,

$$\Lambda_{p_t}^\eta = \Sigma_2^{(d-1)} + \Sigma_4^{(d-1)} + \dots + \Sigma_{2p_t}^{(d-1)} = \sum_{\ell=1}^{p_t} \Sigma_{2\ell}^{(d-1)} \quad \text{et} \quad \Pi_{i_t}^\eta = \Sigma_1^{(d-1)} + \Sigma_3^{(d-1)} + \dots + \Sigma_{2i_t-1}^{(d-1)} = \sum_{\ell=1}^{i_t} \Sigma_{2\ell-1}^{(d-1)}$$

avec  $i_t = [(t+1)/2]$  et  $p_t = t - [(t+1)/2]$ . Il est alors possible d'établir la même décomposition que (VI.61) à partir de ces nouvelles définitions. Ainsi, par l'intermédiaire du principe d'invariance (VI.34) et en suivant le raisonnement (VI.61)–(VI.65), nous obtenons en dimension supérieure

$$\frac{1}{T^{k+d-1/2}} \sum_{t=1}^T t^{k-1} S_t^{(d)} = O_{\mathcal{P}}(1)$$

et donc, par suite, pour tout  $1 \leq k \leq r + 1$ ,

$$\frac{P_{k,T}}{T^{d+1/2}} = O_{\mathcal{P}}(1) \quad \text{et} \quad \frac{u_{\eta,T}}{T^{d-1/2}} = \frac{S_T^{(d)} + \varepsilon_T}{T^{d-1/2}} + O_{\mathcal{P}}(1),$$

en utilisant successivement (VI.31) et (VI.40). Par les Théorèmes B.1–B.7 et de nouveau le Théorème B.9, nous en déduisons facilement, suivant le même raisonnement que précédemment, que  $Q_T$  grandit avec la vitesse  $T^{2d}$  et cela achève la preuve pour  $\rho_1 = \rho_2 = \dots = \rho_d = -1$ . Enfin, dans le cas plus trivial où  $\kappa = 0$ , le principe d'invariance (VI.71) devient

$$\frac{u_{\eta,[T\tau]}}{\sigma_\eta T^{d-1/2}} \xrightarrow{\mathcal{L}} W^{(d-1)}(\tau)$$

à partir des Théorèmes B.7–B.9, et le raisonnement se termine aisément de la même manière. ★

### VI.6.2.6 Preuve de la Proposition VI.4

Lorsque  $d = 0$ , le résultat est trivial à partir de (VI.47). Pour  $d > 0$ , reprenons la décomposition (VI.70), dans laquelle  $u_{\eta,t}$  est de nouveau le terme dominant, sous réserve d'une estimation correcte de  $\theta$  (Lemme VI.2). Nous avons dès lors, vectoriellement écrit,

$$\Delta^d u_{\eta} = (I - P) \Delta^d (S^{(d)} + \varepsilon) = (I - P)(\eta + \Delta^d \varepsilon).$$

Ainsi, la quantité  $(\Delta^d u_{\eta,t})$  peut être vue comme le résidu de la régression du processus stationnaire centré  $(\eta_t + \Delta^d \varepsilon_t)$  sur la tendance polynomiale à l'origine de la matrice de projection  $P$ , dont les coefficients sont nuls. On a alors la convergence

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=d}^T (\Delta^d u_{\eta,t})^2 = \mathbb{V}(\eta_t + \Delta^d \varepsilon_t) \quad \text{p.s.}$$

ce qui achève la démonstration.



---

## *Une étude avec EDF R&D sur la consommation individuelle d'électricité*

---

En guise d'annexe, nous proposons tout d'abord le résumé d'une étude menée sur deux années avec l'équipe ICAME d'EDF R&D sur la prévision de la consommation individuelle d'électricité. Cette étude, publiée dans [13] en collaboration avec Sophie Bercu, n'est pas fondamentalement liée au reste du manuscrit, mais elle permet d'entrevoir une facette des séries chronologiques que nous n'avons pas encore eu l'occasion d'aborder. Nous resterons cependant ici très évasifs quant au contenu, sans introduction ni bibliographie, car nous nous éloignons quelque peu des objectifs de la thèse. Le lecteur intéressé pourra consulter [13], pour plus de détails sur les enjeux, les objectifs, etc.

### **A.1 Le contexte de l'étude**

Pour résumer très brièvement l'un des axes d'intérêt d'EDF, il s'agit d'optimiser la quantité d'énergie produite et impossible à stocker, et cela passe dans notre cas par la prédiction de la consommation individuelle d'électricité. On se donne ainsi une courbe, fortement bruitée, représentant la consommation horaire d'un client sur plusieurs mois. Nous disposons également d'une chronique de température mesurée par la station météorologique la plus proche, sur la même période. Ces deux séries sont représentées sur la Figure A.1. Le client en question est dit « thermosensible » en cela que l'on note une différence significative de consommation d'une saison sur l'autre, ce qui peut s'expliquer en majorité par la présence d'un chauffage électrique, voire d'une climatisation.

Il est dès lors naturel de proposer une modélisation stochastique de la consommation individuelle d'électricité prenant en compte une certaine périodicité dans le temps. Par ailleurs, comme nous pouvons le constater sur la Figure A.2, il existe une relation quasi-linéaire entre le logarithme de la

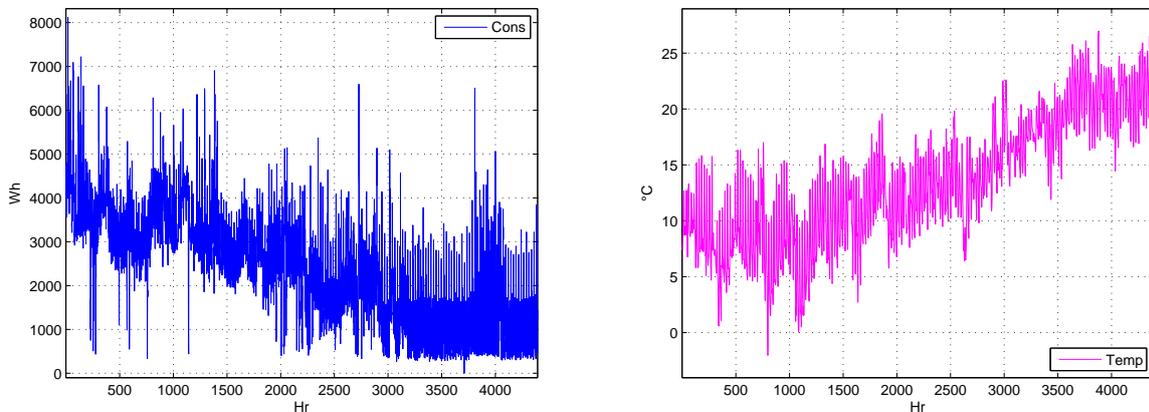


FIGURE A.1 – Consommation individuelle horaire d'un client thermosensible (gauche), température mesurée sur la même période (droite).

consommation et la température simultanément mesurée, c'est pourquoi il nous a semblé judicieux d'extraire l'information linéaire apportée par la température sur la consommation dans un premier temps, et de traiter les résidus de manière chronologique dans un second temps. Ces derniers sont également représentés sur la Figure A.2.

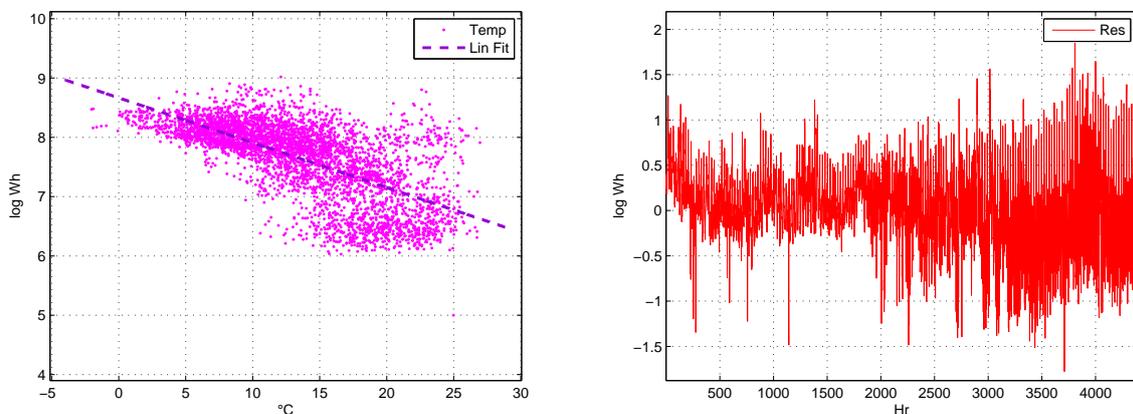


FIGURE A.2 – Relation entre le logarithme de la consommation et la température (gauche), résidus de la régression linéaire (droite).

La consommation sera notée  $(C_t)$  et la chronique de température  $(U_t)$ , pour  $1 \leq t \leq T$ . La transformation de Box-Cox stabilisatrice de la variance du processus est, comme indiqué précédemment, de nature logarithmique. On note ainsi, pour  $1 \leq t \leq T$ ,

$$Y_t = \log(C_t + e^\mu)$$

où la constante  $\mu > 0$  est arbitraire et introduite pour assurer que  $Y_t$  ne s'annule jamais, précaution utile lorsque nous évaluons nos modèles par l'intermédiaire de critères relatifs. Nous proposons alors

une modélisation SARIMAX couplée (voir la Section I.3.2.3), structure qui semble de prime abord le mieux correspondre à nos données, en vertu des graphes associés. On obtient alors, pour  $1 \leq t \leq T$ ,

$$Y_t = c_0 + \mathcal{C}(L)U_t + \varepsilon_t \quad (\text{A.1})$$

où  $c_0 \in \mathbb{R}$  est l'intercept du modèle et  $\mathcal{C}$  est un polynôme de degré  $r$  tel que, pour tout  $z \in \mathbb{C}$ ,

$$\mathcal{C}(z) = \sum_{k=1}^r c_k z^{k-1}.$$

Le paramètre vectoriel inconnu  $c \in \mathbb{R}^{r+1}$  est estimé par moindres carrés ordinaires. La corrélation résiduelle est susceptible d'engendrer un estimateur  $\widehat{c}_T$  loin d'être optimal, mais ce n'est pas ici un problème crucial. En effet, l'information chronologique est extraite dans un second temps de  $(\varepsilon_t)$  à travers le modèle SARIMA( $p, d, q$ )  $\times$  ( $P, D, Q$ ) $_s$  défini par

$$(1-L)^d(1-L^s)^D \mathcal{A}(L) \mathcal{A}_s(L) \varepsilon_t = \mathcal{B}(L) \mathcal{B}_s(L) V_t \quad (\text{A.2})$$

où  $(V_t)$  est un bruit blanc de variance  $\sigma^2 > 0$  et, pour tout  $z \in \mathbb{C}$ , les polynômes associés à (A.2) sont donnés par

$$\mathcal{A}(z) = 1 - \sum_{k=1}^p a_k z^k, \quad \mathcal{A}_s(z) = 1 - \sum_{k=1}^P \alpha_k z^{sk}, \quad \mathcal{B}(z) = 1 + \sum_{k=1}^q b_k z^k \quad \text{et} \quad \mathcal{B}_s(z) = 1 + \sum_{k=1}^Q \beta_k z^{sk}.$$

Pour éclaircir les idées, on a ici  $a \in \mathbb{R}^p$ ,  $b \in \mathbb{R}^q$ ,  $\alpha \in \mathbb{R}^P$  et  $\beta \in \mathbb{R}^Q$ , ils seront quant à eux estimés par des algorithmes optimisés reposant sur des principes de moindres carrés généralisés et de maximum de vraisemblance, fournis par l'environnement logiciel. Une fois correctement spécifié, nous voyons donc que le processus différencié  $(\Delta^d \Delta_s^D \varepsilon_t)$  est une solution stationnaire de la modélisation ARMA causale saisonnière cachée dans (A.2), c'est-à-dire que  $\mathcal{A}(z) \neq 0$  and  $\mathcal{A}_s(z) \neq 0$  pour tout  $z \in \mathbb{C}$  tel que  $|z| \leq 1$ . Cela nous amène à la définition du modèle SARIMAX couplé que nous proposons de mettre en pratique sur nos données.

**Définition A.1.** *Un processus stochastique  $(Y_t)$  satisfait une modélisation SARIMAX( $p, d, q, r$ )  $\times$  ( $P, D, Q$ ) $_s$  couplée si, pour tout  $1 \leq t \leq T$ , il est engendré par*

$$\begin{cases} Y_t = c_0 + \mathcal{C}(L)U_t + \varepsilon_t \\ (1-L)^d(1-L^s)^D \mathcal{A}(L) \mathcal{A}_s(L) \varepsilon_t = \mathcal{B}(L) \mathcal{B}_s(L) V_t \end{cases} \quad (\text{A.3})$$

où  $(V_t)$  est un bruit blanc de variance  $\sigma^2 > 0$ .

Notons que, par souci de simplification, nous autorisons les processus  $(Y_t)$  et  $(U_t)$  à être né-

gativement indexés. Toutes ces valeurs initiales sont par convention fixées à 0, à l'exception de  $U_{-r+2}, \dots, U_{-1}$  que nous supposons observées et qui nous seront utiles par la suite.

## A.2 La stratégie de Box et Jenkins

Avant d'estimer nos paramètres  $a, b, c, \alpha, \beta$  (voire  $\sigma^2$ ) liés à notre modélisation SARIMAX, il est nécessaire d'évaluer les ordres du modèle. Les ordres  $p, P, q$  et  $Q$  sont à l'origine de l'autorégression,  $d$  et  $D$  de la non stationnarité et  $s$  de la saisonnalité. Nous allons pour cela mettre en œuvre la stratégie de Box et Jenkins [19]–[20] sur les résidus  $(\hat{\varepsilon}_t)$  issus de la régression affine de  $(Y_t)$  sur  $r$  valeurs consécutives de la température. Remarquons tout d'abord que l'influence de  $c_0$  disparaît dès que le processus est au minimum une fois différencié. En effet, pour tout  $1 \leq t \leq T$ ,

$$(1 - L)^d (1 - L^s)^D \mathcal{A}(L) \mathcal{A}_s(L) (Y_t - \mathcal{C}(L)U_t) = \mathcal{B}(L) \mathcal{B}_s(L) V_t$$

dès que  $d + D > 0$ , ce qui sera, comme nous le verrons, une condition obligatoirement vérifiée.

### A.2.1 L'ensemble résiduel

On suppose dans un premier temps que  $r$  est connu. Soient  $I$  la matrice identité d'ordre  $T$ ,  $Y$  le vecteur d'ordre  $T$  des observations, et  $U$  la matrice « de design » d'ordre  $T \times (r + 1)$ , respectivement donnés par

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix} \quad \text{et} \quad U = \begin{pmatrix} 1 & U_T & U_{T-1} & \dots & U_{T-r+1} \\ 1 & U_{T-1} & U_{T-2} & \dots & U_{T-r} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & U_1 & U_0 & \dots & U_{-r+2} \end{pmatrix}.$$

**Théorème A.1.** *Supposons que  $U'U$  est inversible. Alors, le processus différencié  $(\Delta^d \Delta_s^D \varepsilon_t)$ , où  $(\varepsilon_t)$  est donné, pour tout  $1 \leq t \leq T$ , par la forme vectorielle*

$$\varepsilon = (I - U(U'U)^{-1}U')Y$$

*est une solution stationnaire du modèle couplé (A.3).*

**Démonstration.** Ce théorème est une conséquence directe du Théorème 3.1.1 de [25], combiné à quelques calculs de base de moindres carrés. ★

Il est par ailleurs facile de voir que, lorsque  $r = 0$ , l'ensemble résiduel est défini par  $\varepsilon = Y - \bar{Y}_T$ , où  $\bar{Y}_T$  est la moyenne empirique de  $(Y_t)$ . Le modèle se réduit alors à la formulation SARIMA( $p, d, q$ ) ×

$(P, D, Q)_s$  classique sur la courbe recentrée. Par ailleurs, soit  $\widehat{c}_T$  l'estimateur des moindres carrés dans le modèle linéaire liant  $(Y_t)$  à  $(U_t)$ . Pour  $1 \leq t \leq T$ , on définit

$$\widehat{\varepsilon}_t = Y_t - \widehat{c}_{0,T} - \sum_{k=1}^r \widehat{c}_{k,T} U_{t-k+1} \quad (\text{A.4})$$

où l'on suppose la variable exogène  $(U_t)$  connue sur l'intervalle  $-r + 2 \leq t \leq T + H$ , pour un horizon de prédiction  $H \geq 0$ . Pour une question de simplicité, nous choisirons  $r = 1$  lors de l'étude de Box et Jenkins. Il est à noter que ce choix est également pertinent d'un point de vue physique, car la température est un phénomène naturel hautement autocorrélé, et l'utilisation de  $U_t$  pour expliquer  $Y_t$  nous dispense généralement d'utiliser  $U_{t-1}, U_{t-2}, \dots$

## A.2.2 La saisonnalité

La Figure A.3 représente le spectrogramme de Fourier du signal  $(\widehat{\varepsilon}_t)$  et de sa version différenciée aux ordres 12 et 24, pour  $T = 17520$  (soit 2 ans de données), dont les très basses fréquences ont été tronquées (l'alternance été/hiver étant ici inexploitable, par manque d'observations). Nous voyons que  $(\widehat{\varepsilon}_t)$  est périodique, et sa principale fréquence indique que le cycle se reproduit 730 fois sur un signal de 730 jours. La saisonnalité journalière est ainsi clairement prononcée. En outre, le signal  $(\Delta_{12}\widehat{\varepsilon}_t)$  reste périodique alors que le signal  $(\Delta_{24}\widehat{\varepsilon}_t)$  est quasi-apériodique. Tout concorde donc pour que nous choissions  $s = 24$ .

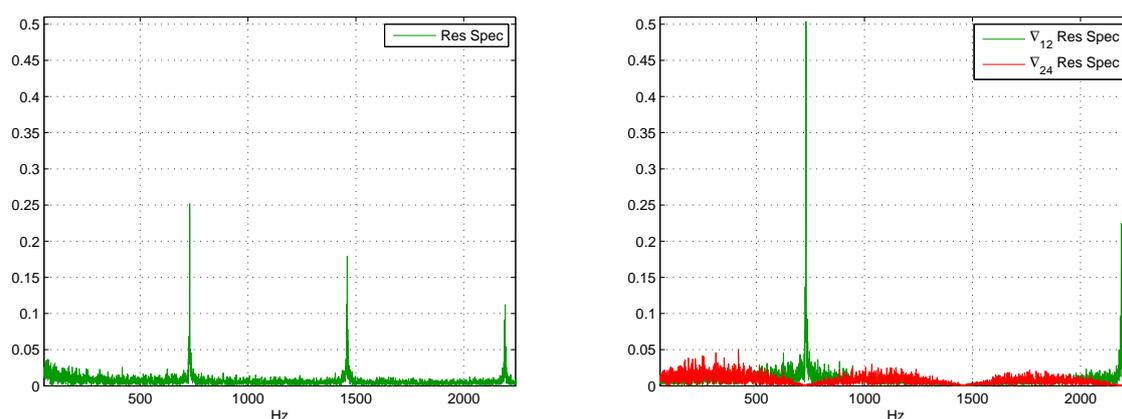


FIGURE A.3 – Spectrogramme de Fourier du signal  $(\widehat{\varepsilon}_t)$  résiduel (gauche) puis du signal résiduel différencié  $(\Delta_s \widehat{\varepsilon}_t)$  de période  $s = 12$  et  $s = 24$  (droite).

### A.2.3 La stationnarité

L'utilisation jointe de la procédure de KPSS et du test ADF (voir Chapitre VI) nous suggère que  $(\widehat{\varepsilon}_t)$  n'est pas stationnaire, alors que  $(\Delta\widehat{\varepsilon}_t)$ ,  $(\Delta_{24}\widehat{\varepsilon}_t)$  et  $(\Delta\Delta_{24}\widehat{\varepsilon}_t)$  sont stationnaires. En conséquence,  $(\widehat{\varepsilon}_t)$  est stationnaire en différence, et l'obtention d'une solution stationnaire au modèle ARMA causal passe nécessairement par une modélisation ARIMA avec  $d = 1$ , ou SARIMA avec  $d + D > 0$ .

### A.2.4 L'autocorrélation

Les Figures A.4 et A.5 illustrent l'ACF et la PACF des signaux  $(\Delta_{24}\widehat{\varepsilon}_t)$  et  $(\Delta\Delta_{24}\widehat{\varepsilon}_t)$ , respectivement. Nous observons sur la première les caractéristiques d'un modèle SARIMA( $p, 0, 0$ )  $\times$  ( $0, 1, Q$ )<sub>24</sub> avec  $p \leq 5$  et  $Q = 1$ , en vertu des principes de reconnaissance visuelle, largement décrits par exemple dans [20], liés aux chutes brutales ou aux décroissances exponentielles des fonctions d'autocorrélation. La seconde représentation, quant à elle, semble exhiber les propriétés d'une modélisation SARIMA( $p, 1, q$ )  $\times$  ( $0, 1, Q$ )<sub>24</sub>, avec  $p = 1$ ,  $q = 2$  et  $Q = 1$ .

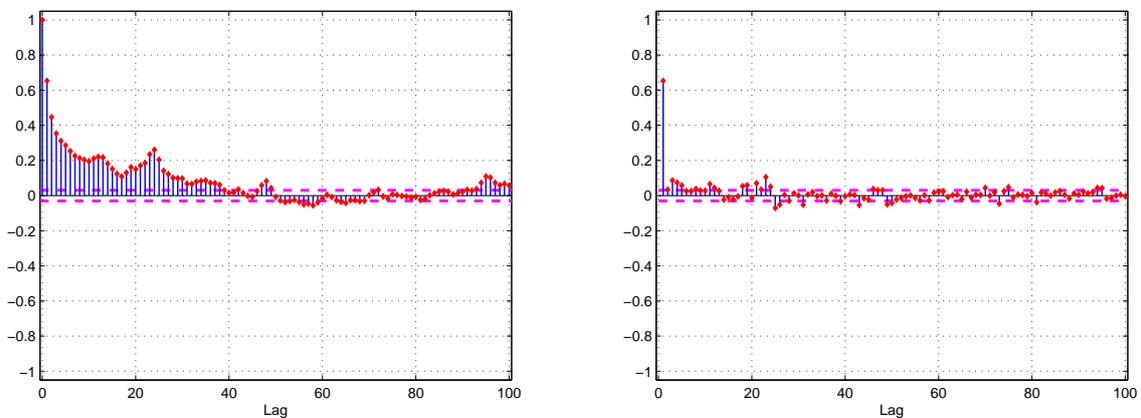


FIGURE A.4 – ACF (gauche) et PACF (droite) du signal  $(\Delta_{24}\widehat{\varepsilon}_t)$ .

### A.2.5 La modélisation

La stratégie de Box et Jenkins nous a permis d'identifier en première approximation les ordres de la modélisation. En pratique, il s'agira bien entendu de minimiser un certain critère tout en faisant varier ces valeurs dans leur voisinage. On a ainsi minimisé les critères bayésiens usuels (AIC, SBC, log-vraisemblance) sur une section de 6 mois de données, en vue d'obtenir la meilleure configuration au sens de la modélisation. Il en a résulté un SARIMAX( $3, 0, 2, 2$ )  $\times$  ( $0, 1, 1$ )<sub>24</sub>, dont l'expression

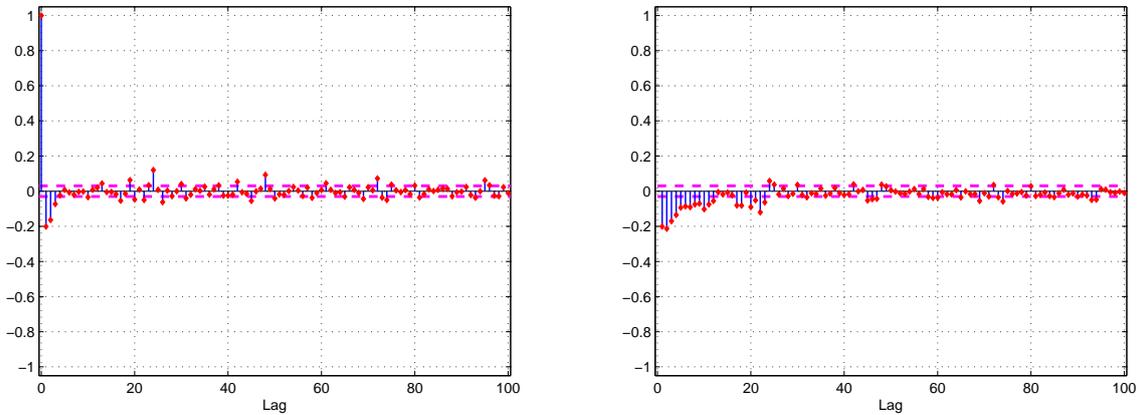


FIGURE A.5 – ACF (gauche) et PACF (droite) du signal  $(\Delta\Delta_{24}\widehat{\varepsilon}_t)$ .

explicite est donnée, pour tout  $28 \leq t \leq T = 4380$ , par

$$\begin{cases} Y_t = c_0 + c_1 U_t + c_2 U_{t-1} + \varepsilon_t \\ \varepsilon_t = \varepsilon_{t-24} + a_1(\varepsilon_{t-1} - \varepsilon_{t-25}) + a_2(\varepsilon_{t-2} - \varepsilon_{t-26}) + a_3(\varepsilon_{t-3} - \varepsilon_{t-27}) \\ \quad + (V_t - b_1 V_{t-1} - b_2 V_{t-2}) - \beta_1(V_{t-24} - b_1 V_{t-25} - b_2 V_{t-26}) \end{cases}$$

et la procédure d'estimation nous a fourni  $\widehat{c}_0 = 7.9871$ ,  $\widehat{c}_1 = 0.0166$ ,  $\widehat{c}_2 = -0.0420$ ,  $\widehat{a}_1 = 0.4776$ ,  $\widehat{a}_2 = 0.9030$ ,  $\widehat{a}_3 = -0.4305$ ,  $\widehat{b}_1 = 0.0801$ ,  $\widehat{b}_2 = -0.8524$ ,  $\widehat{\beta}_1 = -0.8125$  et  $\widehat{\sigma}^2 = 0.0522$ . Les tests de significativité associés nous ont confirmés dans cette voie, tout comme les tests de blancheur résiduelle. Le signal réel est ainsi reconstitué tel que nous le présentons sur la Figure A.6.

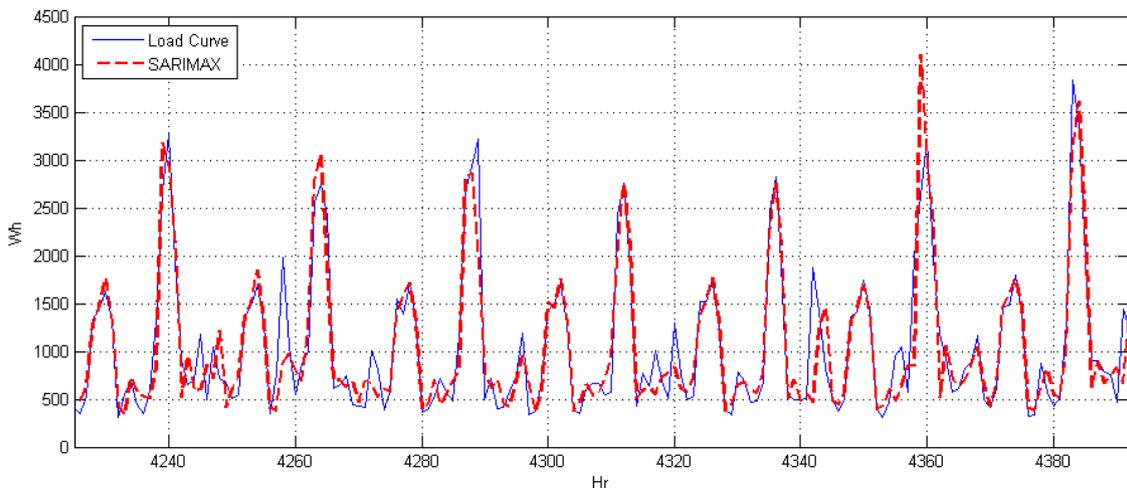


FIGURE A.6 – Reconstitution du signal horaire à partir de la modélisation SARIMAX(3,0,2,2) × (0,1,1)<sub>24</sub> en rouge, superposée au signal réel en bleu.

### A.3 Application à la prédiction

La problématique de la prédiction est tout autre, car il s'agit dès lors de minimiser non plus des critères bayésiens, mais des critères de prédiction. Il est d'ailleurs bien connu que, d'une manière générale, les modèles de prédiction sont bien plus parcimonieux que leurs homologues appliqués à la modélisation car il ne s'agit plus désormais d'expliquer, mais de prévoir, et en cela de diminuer au possible l'incertitude liée à l'estimation. C'est donc naturellement qu'il en a résulté sur nos données le modèle SARIMAX(1,0,0,2) × (0,1,1)<sub>24</sub> évalué sur un historique glissant de 0.75 mois de données. De manière explicite, pour tout 26 ≤ t ≤ T = 548, son expression est définie par

$$\begin{cases} Y_t = c_0 + c_1 U_t + c_2 U_{t-1} + \varepsilon_t \\ \varepsilon_t = \varepsilon_{t-24} + a_1(\varepsilon_{t-1} - \varepsilon_{t-25}) + V_t - \beta_1 V_{t-24} \end{cases}$$

et la procédure d'estimation fournit à cette occasion  $\widehat{c}_0 = 7.2494$ ,  $\widehat{c}_1 = 0.0497$ ,  $\widehat{c}_2 = -0.0629$ ,  $\widehat{a}_1 = 0.3540$ ,  $\widehat{\beta}_1 = -0.7086$  et  $\widehat{\sigma}^2 = 0.0708$ . Nous avons reproduit, sur la Figure A.7, 14 jours consécutifs du signal et chaque prédiction a été générée sur un horizon de  $H = 24$  heures, ce qui explique la présence d'intervalles de prédiction plutôt larges. Une prédiction à horizon  $H$  est du reste un abus de langage car cette dernière est constituée de  $H$  prédictions à horizon 1, chaque valeur intermédiaire étant par la suite considérée comme observée. De fait, seule la  $H$ -ème composante du vecteur  $(\widetilde{Y}_{T+1} \dots \widetilde{Y}_{T+H})$  est bien une prédiction à horizon  $H$ .

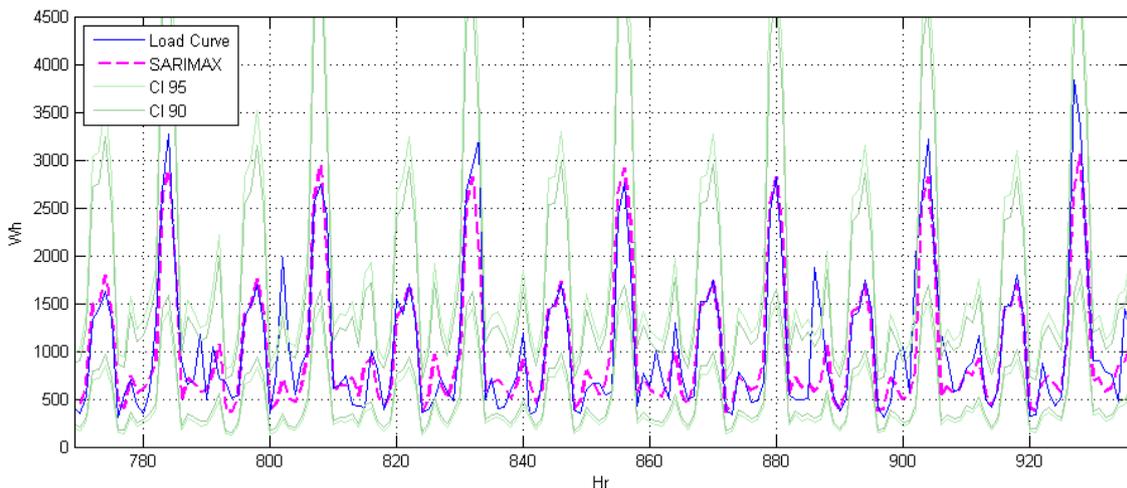


FIGURE A.7 – Prédiction de 14 jours du signal horaire à partir de la modélisation SARIMAX(1,0,0,2) × (0,1,1)<sub>24</sub> en magenta, superposée au signal réel en bleu.

En conclusion, nous voudrions insister sur la significativité parfois ténue de la variable exogène, et de la valeur de  $r$  associée. En effet, lorsque la modélisation est fortement chronologique et explore les données loin dans le passé, l'influence de la covariable est en quelque sorte déjà retrouvée car

elle s'inscrit dans la chronologie. Cela explique parfaitement pourquoi notre étude nous a suggéré d'utiliser de très faibles valeurs de  $r$ , parfois même  $r = 0$ . La chronique de température est de plus à manier avec précaution, et ce pour diverses raisons. Tout d'abord, les données utilisées en prédiction sont elles-mêmes des prédictions (en provenance de laboratoires spécialisés dans la météorologie). En outre – et surtout – la station météorologique la plus proche ne l'est que géographiquement parlant : il peut exister une différence d'altitude ou même de nébulosité susceptible de rendre une mesure locale parfaitement non adaptée à la réalité climatique du client considéré. D'un point de vue statistique, nous avons pu voir que, malgré la forte perturbation de la courbe de charge principalement due à son caractère individuel, il est possible d'obtenir d'excellents résultats de modélisation et de prédiction. Nous avons ainsi pu obtenir 3% seulement d'erreur relative entre le signal observé et 14 jours de prédictions horaires à horizon 24, par l'intermédiaire d'une utilisation scrupuleuse de la stratégie de Box et Jenkins, dont nous avons exploré diverses facettes théoriques tout au long de ce manuscrit.



## Quelques outils techniques

Nous centralisons dans cette annexe technique un ensemble de théorèmes qui, énumérés sans logique *a priori*, ont malgré tout un point commun : ce sont des résultats de probabilités et de statistique que nous utilisons de manière récurrente tout au long de ce manuscrit.

### B.1 Lois fortes des grands nombres

**Théorème B.1.** Soit  $(V_n)$  une suite de variables aléatoires indépendantes et identiquement distribuées. Supposons qu'il existe  $a \geq 1$  tel que  $\mathbb{E}[|V_0|^a] < +\infty$ . Alors,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=0}^n |V_t|^a = \mathbb{E}[|V_0|^a] \quad \text{p.s.} \quad \text{et} \quad \sup_{0 \leq t \leq n} |V_t| = o(n^{1/a}) \quad \text{p.s.}$$

*Démonstration.* Voir par exemple le Corollaire 1.3.21 de [45].

★

**Théorème B.2.** Soient  $(M_n)$  une martingale vectorielle de  $\mathbb{R}^p$  adaptée à une filtration  $\mathcal{F}_n$ , de carré intégrable, et  $\langle M \rangle_n$  son processus croissant. Si, pour une valeur de  $\gamma > 0$ ,

$$\lim_{n \rightarrow \infty} \lambda_{\min}(\langle M \rangle_n) = +\infty \quad \text{p.s.} \quad \text{et} \quad (\log \lambda_{\max}(\langle M \rangle_n))^{1+\gamma} = o(\lambda_{\min}(\langle M \rangle_n)) \quad \text{p.s.}$$

alors,

$$\lim_{n \rightarrow \infty} \langle M \rangle_n^{-1} M_n = 0 \quad \text{p.s.}$$

*Démonstration.* Voir par exemple [46], ou bien la Section 4.3.2 de [45].

★

## B.2 Théorèmes centraux limites

**Théorème B.3.** Soit  $(V_n)$  une suite de variables aléatoires indépendantes et identiquement distribuées. Supposons que  $\mathbb{E}[V_0] = \mu$  et que  $\mathbb{E}[V_0^2] = \sigma^2 < +\infty$ . Soit  $S_n = V_0 + \dots + V_n$ . Alors, on a la normalité asymptotique

$$\frac{S_n - n\mu}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

*Démonstration.* Voir par exemple [45], ou le Chapitre 7 de [25]. ★

**Théorème B.4.** Soient  $(M_n)$  une martingale vectorielle de  $\mathbb{R}^p$  adaptée à une filtration  $\mathcal{F}_n$ , de carré intégrable, et  $(\langle M \rangle_n)$  son processus croissant. S'il existe une suite  $(a_n)$ , positive et croissante vers l'infini ainsi qu'une matrice  $L$  telles que les conditions suivantes sont simultanément réalisées, pour tout  $\varepsilon > 0$ ,

$$\frac{\langle M \rangle_n}{a_n} \xrightarrow{\mathcal{P}} L \quad \text{et} \quad \frac{1}{a_n} \sum_{t=1}^n \mathbb{E} \left[ \|\Delta M_t\|^2 \mathbb{I}_{\{\|\Delta M_t\| \geq \varepsilon \sqrt{a_n}\}} \mid \mathcal{F}_{t-1} \right] \xrightarrow{\mathcal{P}} 0,$$

alors, on a la normalité asymptotique

$$\frac{M_n}{\sqrt{a_n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, L).$$

De plus, si  $L$  est inversible, on a la normalité asymptotique

$$\sqrt{a_n} \langle M \rangle_n^{-1} M_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, L^{-1}).$$

*Démonstration.* Voir par exemple le Corollaire 2.1.10 de [45]. Notons qu'on parlera de « condition de Lindeberg » pour qualifier la seconde hypothèse. ★

## B.3 Principes de déviations

**Théorème B.5.** Soit  $(V_n)$  une suite de variables aléatoires indépendantes et identiquement distribuées, telles que  $\mathbb{E}[\exp(tV_0)] < +\infty$  pour tout  $t$  dans un voisinage de l'origine. Soit  $S_n = V_0 + \dots + V_n$ . Alors, pour tout  $\delta > \mathbb{E}[V_0]$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(S_n > n\delta) = -I(\delta)$$

où  $I$  est la transformée de Fenchel-Legendre de la log-Laplace de  $V_0$ , donnée par

$$I(\delta) = \sup_{t \in \mathbb{R}} \left( \delta t - \log \mathbb{E} \left[ \exp(t V_0) \right] \right).$$

*Démonstration.* La démonstration de cette écriture du théorème de Cramér-Chernoff se trouve par exemple dans [37]. ★

**Théorème B.6.** Soit  $(V_n)$  une suite de variables aléatoires indépendantes et identiquement distribuées, d'espérance finie, et satisfaisant

$$\limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log n \mathbb{P}(|V_0| > b_n \sqrt{n}) = -\infty.$$

Soit  $(b_n)$  une suite déterministe, positive et croissante, telle que  $b_n^2 = o(n)$ . Alors, pour  $n \geq 1$ , la suite

$$\left( \frac{1}{b_n \sqrt{n}} \sum_{t=0}^n (V_t - \mathbb{E}[V_t]) \right)$$

satisfait un PGD sur  $\mathbb{R}$  de vitesse  $(b_n^2)$  et de fonction de taux  $I(x) \neq 0$  pour  $x \neq 0$ .

*Démonstration.* C'est une formulation du Théorème 2.2 de [52].

★

## B.4 Principes d'invariance

**Théorème B.7.** Soit  $(Z_T)$  une suite de variables aléatoires indépendantes et identiquement distribuées, d'espérance nulle et de variance  $\sigma^2 > 0$ . Soit  $S_0 = 0$  et  $S_T = Z_1 + \dots + Z_T$ . Alors, pour  $0 \leq \tau \leq 1$ , on a la convergence en loi

$$\frac{1}{\sigma \sqrt{T}} \left( S_{[T\tau]} + (T\tau - [T\tau]) Z_{[T\tau]+1} \right) \xrightarrow{\mathcal{L}} W(\tau)$$

où  $W(t)$  est un mouvement brownien standard.

*Démonstration.* Voir la Section 8 de [15] pour tout le bagage technique accompagnant le principe d'invariance de Donsker.

★

**Théorème B.8.** Soit  $(Z_T)$  un processus stationnaire indexé par  $\mathbb{N}$ . On note

$$\omega^2 = \mathbb{E}[Z_0^2] + 2 \sum_{t=1}^{\infty} \mathbb{E}[Z_0 Z_t]$$

sa variance de long terme, et  $S_T = Z_0 + \dots + Z_T$ . Alors, pour  $0 \leq \tau \leq 1$ , on a la convergence en loi

$$\frac{S_{[T\tau]}}{\omega \sqrt{T}} \xrightarrow{\mathcal{L}} W(\tau)$$

où  $W(t)$  est un mouvement brownien standard.

*Démonstration.* Voir le Théorème 1 de [35].

★

## B.5 Théorème de continuité

**Théorème B.9.** Soit  $(Z_T, Z)$  une suite de variables aléatoires définies sur un espace métrique  $\mathcal{S}$ . Supposons que l'application  $h : \mathcal{S} \rightarrow \mathcal{S}'$ , où  $\mathcal{S}'$  est aussi un espace métrique, possède un ensemble de discontinuités  $\mathcal{D}_h$  tel que  $\mathbb{P}(Z \in \mathcal{D}_h) = 0$ . Alors,

$$Z_T \longrightarrow Z \quad \implies \quad h(Z_T) \longrightarrow h(Z),$$

*l'implication restant valable pour la convergence en loi, la convergence en probabilité et la convergence presque sûre.*

*Démonstration.* Ce résultat, couramment appelé *continuous mapping theorem* dans la littérature, est par exemple introduit dans le Théorème 2.7 de [15], et prouvé par la suite. ★

---

## Bibliographie

---

- [1] T. W. ANDERSON et J. B. TAYLOR : Strong consistency of least squares estimates in dynamic models. *Ann. Statis.*, 7-3:484–489, 1979.
- [2] M. A. ARCONES : The large deviation principle for stochastic processes I. *Theor. Probab. Appl.*, 47:567–583, 2003.
- [3] M. A. ARCONES : The large deviation principle for stochastic processes II. *Theor. Probab. Appl.*, 48:19–44, 2003.
- [4] Ole E. BARNDORFF-NIELSEN et Andreas BASSE-O’CONNOR : Quasi Ornstein-Uhlenbeck processes. *Bernoulli.*, 17(3):916–941, 2011.
- [5] Ole E. BARNDORFF-NIELSEN et Neil SHEPHARD : Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 63(2):167–241, 2001.
- [6] Ole E. BARNDORFF-NIELSEN et A. VERAART : Stochastic volatility of volatility and variance risk premia. *J. Fin. Econ.*, 11(1):1–46, 2013.
- [7] B. BERCU : Central limit theorem and law of iterated logarithm for least squares algorithms in adaptive tracking. *SIAM J. Control. Optim.*, 36:910–928, 1998.
- [8] B. BERCU : On the convergence of moments in the almost sure central limit theorem for martingales with statistical applications. *Stochastic Process. Appl.*, 11:157–173, 2004.
- [9] B. BERCU, P. CÉNAC et G. FAYOLLE : On the almost sure central limit theorem for vector martingales : convergence of moments and statistical applications. *J. Appl. Probab.*, 46:151–169, 2009.
- [10] B. BERCU et F. PROĀA : A sharp analysis on the asymptotic behavior of the Durbin-Watson statistic for the first-order autoregressive process. *ESAIM Probab. Stat.*, 17-1:500–530, 2013.
- [11] B. BERCU, F. PROĀA et N. SAVY : On Ornstein-Uhlenbeck driven by Ornstein-Uhlenbeck processes. *En attente de publication.*, 2013.
- [12] B. BERCU et A. TOUATI : Exponential inequalities for self-normalized martingales with applications. *Ann. Appl. Probab.*, 18, no. 5:1848–1869, 2008.

- [13] S. BERCU et F. PROÏA : A SARIMAX coupled modelling applied to individual load curves intra-day forecasting. *J. Appl. Statist.*, 40, no. 6:1333–1348, 2013.
- [14] A. BHARGAVA : On the theory of testing for unit roots in observed time series. *Rev. Econ. Stud.*, 53:369–384, 1986.
- [15] P. BILLINGSLEY : *Convergence of probability measures*. Wiley Series in Probability and Statistics : Probability and Statistics. John Wiley & Sons Inc., New York, 1999.
- [16] V. BITSEKI PENDA, H. DJELLOUT et F. PROÏA : Moderate deviations for the Durbin-Watson statistic related to the first-order autoregressive process. *ESAIM Probab. Stat. À paraître.*, 2013.
- [17] P. BLOOMFIELD : *Fourier Analysis of Time Series : An Introduction*. Wiley, New-York, 1976.
- [18] T. BOLLERSLEV : General autoregressive conditional heteroscedasticity. *J. Econometrics.*, 31:307–327, 1986.
- [19] G. E. P. BOX et G. M. JENKINS : *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco, 1970.
- [20] G. E. P. BOX, G. M. JENKINS et G. C. REINSEL : *Time Series Analysis, Forecasting and Control*. Holden-Day, Third Edition, Series G, 1976.
- [21] G. E. P. BOX et G. M. LJUNG : On a measure of a lack of fit in time series models. *Biometrika.*, 65-2:297–303, 1978.
- [22] G. E. P. BOX et D. A. PIERCE : Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Am. Stat. Assoc.*, 65:1509–1526, 1970.
- [23] T. BREUSCH : Testing for autocorrelation in dynamic linear models. *Aust. Econ. Papers.*, 17-31:334–355, 1978.
- [24] D. R. BRILLINGER : *Time Series : Data Analysis and Theory*. Holden-Day, Second Edition, San Francisco, 1981.
- [25] P. J. BROCKWELL et R. A. DAVIS : *Time Series : Theory and Methods*. Springer-Verlag, Second Edition, New-York, 1991.
- [26] P. J. BROCKWELL et R. A. DAVIS : *Introduction to Time Series and Forecasting*. Springer-Verlag, New-York, 1996.
- [27] P. J. BROCKWELL et A. LINDNER : Ornstein-Uhlenbeck related models driven by Lévy processes. *In Statistical methods for stochastic differential equations*, volume 124 de *Monogr. Statist. Appl. Probab.*, pages 383–427. CRC Press, 2012.
- [28] R. W. BUTLER et M. S. PAOLELLA : Uniform saddlepoint approximations for ratios of quadratic forms. *Bernoulli.*, 14:140–154, 2008.
- [29] F. CHAABANE et F. MAAOUIA : Théorèmes limites avec poids pour les martingales vectorielles. *ESAIM Probab. Stat.*, 4:137–189, 2000.

- 
- [30] N. H. CHAN et C. Z. WEI : Asymptotic inference for nearly nonstationary AR(1) processes. *Ann. Statist.*, 15-3:1050–1063, 1987.
- [31] N. H. CHAN et C. Z. WEI : Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist.*, 16-1:367–401, 1988.
- [32] X. CHEN : Moderate deviations for  $m$ -dependent random variables with Banach space value. *Statis. and Probab. Letters.*, 35:123–134, 1998.
- [33] D. N. DE JONG, J. C. NANKERVIS, N. E. SAVIN et C. H. WHITEMAN : Integration versus trend stationarity in time series. *Econometrica.*, 60-2:423–433, 1992.
- [34] R. M. DE JONG, C. AMSLER et P. SCHMIDT : A robust version of the KPSS test, based on indicators. *J. Econometrics.*, 137-2:311–333, 2007.
- [35] J. DEDECKER et E. RIO : On the functional central limit theorem for stationary processes. *Ann. Inst. Henri Poincaré, B.*, 36 (1):1–34, 2000.
- [36] A. DEMBO : Moderate deviations for martingales with bounded jumps. *Electron. Comm. Probab.*, 1, no. 3:11–17, 1996.
- [37] A. DEMBO et O. ZEITOUNI : *Large deviations techniques and applications, second edition*, volume 38 de *Applications of Mathematics*. Springer, 1998.
- [38] D. A. DICKEY, W. R. BELL et R. B. MILLER : Unit roots in time series models : tests and implications. *Am. Stat.*, 40:12–26, 1986.
- [39] D. A. DICKEY et W. A. FULLER : Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Stat. Assoc.*, 74-366:427–431, 1979.
- [40] D. A. DICKEY et W. A. FULLER : Likelihood ratio tests for autoregressive time series with a unit root. *Econometrica.*, 49:1057–1072, 1981.
- [41] D. A. DICKEY et E. S. SAID : Testing ARIMA( $p, 1, q$ ) versus ARMA( $p + 1, q$ ). *Proc. Bus. Econ. Statist. Sect., Am. Statist. Assoc.*, pages 318–322, 1981.
- [42] H. DJELLOUT : Moderate deviations for martingale differences and applications to  $\phi$ -mixing sequences. *Stoch. Stoch. Rep.*, 73, 1-2:37–63, 2002.
- [43] H. DJELLOUT et A. GUILLIN : Moderate deviations for Markov chains with atom. *Stochastic Process. Appl.*, 95, no. 2:203–217, 2001.
- [44] J. J. DOLADO, T. JENKINSON et S. SOSVILLA-RIVERO : Cointegration and unit roots. *J. Econ. Surv.*, 4-3:249–273, 1990.
- [45] M. DUFLO : *Random iterative models*, volume 34 de *Applications of Mathematics*, New York. Springer-Verlag, Berlin, 1997.
- [46] M. DUFLO, R. SENOUSI et A. TOUATI : Sur la loi des grands nombres pour les martingales vectorielles et l'estimateur des moindres carrés d'un modèle de régression. *Ann. Inst. Henri Poincaré.*, 26:549–566, 1990.

- [47] J. DURBIN : Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables. *Econometrica.*, 38:410–421, 1970.
- [48] J. DURBIN : Approximate distributions of Student's  $t$ -statistics for autoregressive coefficients calculated from regression residuals. *J. Appl. Probab.*, Special Vol. 23A:173–185, 1986.
- [49] J. DURBIN et G. S. WATSON : Testing for serial correlation in least squares regression. I. *Biometrika.*, 37:409–428, 1950.
- [50] J. DURBIN et G. S. WATSON : Testing for serial correlation in least squares regression. II. *Biometrika.*, 38:159–178, 1951.
- [51] J. DURBIN et G. S. WATSON : Testing for serial correlation in least squares regression. III. *Biometrika.*, 58:1–19, 1971.
- [52] P. EICHELSBACHER et M. LÖWE : Moderate deviations for i.i.d. random variables. *ESAIM Probab. Stat.*, 7:209–218, 2003.
- [53] R. F. ENGLE : Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica.*, 50-4:987–1007, 1982.
- [54] P.D. FEIGIN : Maximum likelihood estimation for continuous-time stochastic processes. *Advances in Appl. Probability.*, 8(4):712–736, 1976.
- [55] P.D. FEIGIN : Some comments concerning a curious singularity. *J. Appl. Probab.*, 16(2):440–444, 1979.
- [56] W. A. FULLER : *Introduction to Statistical Time Series*. Wiley, Second Edition, New-York, 1995.
- [57] L. G. GODFREY : Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica.*, 46:1293–1302, 1978.
- [58] U. GRENANDER et G. SZEGÖ : *Toeplitz forms and their applications*. California Monographs in Mathematical Sciences. University of California Press, Berkeley, 1958.
- [59] P. HALL et C. C. HEYDE : *Martingale limit theory and its application*. Probability and Mathematical Statistics. Academic Press Inc., New York, 1980.
- [60] E. J. HANNAN : *Multiple Time Series*. Wiley, New-York, 1970.
- [61] D. HARRIS, S. J. LEYBOURNE et B. P. M. McCABE : Modified KPSS tests for near integration. *Economet. Theor.*, 23-2:355–363, 2007.
- [62] A. C. HARVEY : *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge, 1991.
- [63] R. A. HORN et C. R. JOHNSON : *Matrix Analysis*. Cambridge University Press, Cambridge, New-York, 1985.
- [64] Y. HU et H. LONG : Parameter estimation for Ornstein-Uhlenbeck processes driven by  $\alpha$ -stable Lévy motions. *Commun. Stoch. Anal.*, 1(2):175–192, 2007.

- [65] R. IBRAGIMOV et P. C. B. PHILLIPS : Regression asymptotics using martingale convergence methods. *Economet. Theor.*, 24-4:888–947, 2008.
- [66] B. A. INDER : Finite-sample power of tests for autocorrelation in models containing lagged dependent variables. *Economics Letters.*, 14:179–185, 1984.
- [67] B. A. INDER : An approximation to the null distribution of the Durbin-Watson statistic in models containing lagged dependent variables. *Economet. Theor.*, 2:413–428, 1986.
- [68] G. M. JENKINS et D. G. WATTS : *Spectral Analysis and Its Applications*. Holden-Day, San-Francisco, 1968.
- [69] G. JONGBLOED, F. H. VAN DER MEULEN et A. W. VAN DER VAART : Nonparametric inference for Lévy-driven Ornstein-Uhlenbeck processes. *Bernoulli.*, 11(5):759–791, 2005.
- [70] M. L. KING et P. X. WU : Small-disturbance asymptotics and the Durbin-Watson and related tests in the dynamic regression model. *J. Econometrics.*, 47:145–152, 1991.
- [71] G. KITAGAWA et W. GERSCH : *Smoothness Priors Analysis of Time Series*. Springer-Verlag, New-York, 1996.
- [72] Y.A. KUTOYANTS : *Statistical inference for ergodic diffusion processes*. Springer Series in Statistics. Springer-Verlag London Ltd., London, 2004.
- [73] D. KWIATKOWSKI, P. C. B. PHILLIPS, P. SCHMIDT et Y. SHIN : 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? *J. Econometrics.*, 54:159–178, 1992.
- [74] T. L. LAI et D. SIEGMUND : Fixed accuracy estimation of an autoregressive parameter. *Ann. Statist.*, 11:478–485, 1983.
- [75] T. L. LAI et C. Z. WEI : Asymptotic properties of projections with applications to stochastic regression problems. *J. Multivariate Anal.*, 12:346–370, 1982.
- [76] T. L. LAI et C. Z. WEI : Asymptotic properties of general autoregressive models and strong consistency of least-squares estimates of their parameters. *J. Multivariate Anal.*, 13:1–23, 1983.
- [77] M. LEDOUX : Sur les déviations modérées des sommes de variables aléatoires vectorielles indépendantes de même loi. *Ann. Inst. Henri-Poincaré.*, 35:123–134, 1992.
- [78] D. LÉPINGLE : Sur le comportement asymptotique des martingales locales. *In Séminaire de Probabilités, XII*, volume 649 de *Lecture Notes in Math.*, pages 148–161. Springer, Berlin, 1978.
- [79] S. J. LEYBOURNE, T. H. KIM et P. NEWBOLD : Behaviour of dickey-fuller unit-root tests under trend misspecification. *J. Time Ser. Anal.*, 25-5:755–764, 2004.
- [80] S. J. LEYBOURNE, T. H. KIM et P. NEWBOLD : Examination of some more powerful modifications of the Dickey-Fuller test. *J. Time Ser. Anal.*, 26-3:355–369, 2005.

- [81] S. J. LEYBOURNE et B. P. M. McCABE : On the distribution of some test statistics for parameter constancy. *Biometrika.*, 76:167–177, 1989.
- [82] S. J. LEYBOURNE et B. P. M. McCABE : A consistent test for a unit root. *J. Bus. Econ. Stat.*, 12-2:157–166, 1994.
- [83] S. J. LEYBOURNE et B. P. M. McCABE : Modified stationarity tests with data-dependent model-selection rules. *J. Bus. Econ. Stat.*, 17-2:264–270, 1999.
- [84] R.S. LIPTSER et A.N. SHIRYAEV : *Statistics of random processes. II*, volume 6 de *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 2001.
- [85] M. LUBRANO : Testing for unit roots in a bayesian framework. *J. Econometrics.*, 69-1:81–109, 1995.
- [86] J. G. MACKINNON : *Critical values for cointegration tests*. Long-Run Economic Relationships, ed. by R. F. Engle, and C. W. Granger, 266–276. Oxford University Press, Oxford, 1991.
- [87] I. B. MACNEILL : Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times. *Ann. Statist.*, 6-2:422–433, 1978.
- [88] G. S. MADDALA et A. S. RAO : Tests for serial correlation in regression models with lagged dependent variables and serially correlated errors. *Econometrica.*, 41:761–774, 1973.
- [89] E. MALINVAUD : Estimation et prévision dans les modèles économiques autorégressifs. *Review of the International Institute of Statistics.*, 29:1–32, 1961.
- [90] G. MILOVANOVIC et T. RASSIAS : Inequalities for polynomial zeros. *Math. Appl.*, 517, Kluwer, Dordrecht:165–202, 2000.
- [91] U. MÜLLER : Size and power of tests of stationarity in highly autocorrelated time series. *J. Econometrics.*, 128-2:195–213, 2005.
- [92] S. NABEYA et K. TANAKA : Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative. *Ann. Statist.*, 16-1:218–235, 1988.
- [93] C. R. NELSON et C. I. PLOSSER : Trends and random walks in macroeconomic time series : Some evidence and implications. *J. Monet. Econ.*, 10:139–162, 1982.
- [94] M. NERLOVE et K. F. WALLIS : Use of the Durbin-Watson statistic in inappropriate situations. *Econometrica.*, 34:235–238, 1966.
- [95] P. NEWBOLD, S. J. LEYBOURNE et M. E. WOHR : Trend-stationarity, difference-stationarity, or neither : further diagnostic tests with an application to u.s. real gnp, 1875-1993. *J. Econ. Bus.*, 53-1:85–102, 2001.
- [96] W. K. NEWEY et K. D. WEST : A simple, positive definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica.*, 55:703–708, 1987.

- [97] S. NG et P. PERRON : Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *J. Am. Stat. Assoc.*, 90:268–281, 1995.
- [98] J. NYBLOM : Testing for deterministic linear trend in time series. *J. Am. Stat. Assoc.*, 81:545–549, 1986.
- [99] J. NYBLOM et T. MAKELAINEN : Comparisons of tests for the presence of random walk coefficients in a simple linear model. *J. Am. Stat. Assoc.*, 78:856–864, 1983.
- [100] O. ONALAN : Financial modelling with Ornstein-Uhlenbeck processes driven by Lévy process. *Proceedings of the world congress engineering.*, 2:1–6, 2009.
- [101] S. OULIARIS, J. Y. PARK et P. C. B. PHILLIPS : *Testing for a Unit Root in the Presence of a Maintained Trend*, volume 15 de *Advanced Studies in Theoretical and Applied Econometrics, Advances in Econometrics and Modelling*, pp 7–28. Raj, Baldev, Springer Netherlands, 1989.
- [102] S. OULIARIS et P. C. B. PHILLIPS : *Coint 2.0*. Maple Valley. Washington : Aptech Systems, 1994.
- [103] S. B. PARK : On the small-sample power of Durbin's h test. *J. Am. Stat. Assoc.*, 70:60–63, 1975.
- [104] M. M. PELAGATTI et P. K. SEN : A robust version of the KPSS test based on ranks. *Working Papers from Università degli Studi di Milano-Bicocca, Dipartimento di Statistica.*, No 20090701, 2009.
- [105] P. PERRON : Trends and random walks in macroeconomic time series : Further evidence from a new approach. *J. Econ. Dyn. Control.*, 12:297–332, 1988.
- [106] P. PERRON et S. NG : Useful modifications to some unit root tests with dependent errors and their local asymptotic properties. *Rev. Econ. Stud.*, 63:435–463, 1996.
- [107] P. C. B. PHILLIPS : Time series regression with a unit root. *Econometrica.*, 55:277–302, 1987.
- [108] P. C. B. PHILLIPS et P. PERRON : Testing for a unit root in time series regression. *Biometrika.*, 75-2:335–346, 1988.
- [109] P. C. B. PHILLIPS et Z. XIAO : A primer on unit root testing. *J. Econ. Surv.*, Wiley Blackwell. 12-5:423–469, 1999.
- [110] B. M. PÖTSCHER : Noninvertibility and pseudo-maximum likelihood estimation of misspecified ARMA models. *Economet. Theor.*, 7:435–449, 1991.
- [111] M. B. PRIESTLEY : *Spectral Analysis and Time Series*, volume 1 : Univariate Series ; vol. 2 : Multivariate Series, Prediction and Control. Academic Press, New-York, 1981.
- [112] F. PROÏA : Further results on the H-Test of Durbin for stable autoregressive processes. *J. Multivariate Anal.*, 118:77–101, 2013.
- [113] F. PROÏA : Stationarity against multi-integration in the autoregressive process with polynomial trend. *En attente de publication.*, 2013.
- [114] A. PUHALSKII : Large deviations of semimartingales : a maxingale problem approach. I. Limits as solutions to a maxingale problem. *Stoch. Stoch. Rep.*, 61:141–243, 1997, no. 3-4.

## Bibliographie

---

- [115] F. L. RAMSEY : Characterization of the partial autocorrelation function. *Ann. Statist.*, 2-6:1296–1301, 1974.
- [116] E. S. SAID et D. A. DICKEY : Testing for unit roots in autoregressive moving average models of unknown order. *Biometrika.*, 71-3:599–607, 1984.
- [117] P. SAIKKONEN et R. LUUKKONEN : Testing for a moving average unit root in autoregressive integrated moving average models. *J. Am. Stat. Assoc.*, 88:596–601, 1993.
- [118] P. SCHMIDT et P.C.B. PHILLIPS : LM test for a unit root in the presence of deterministic trends. *Oxford B. Econ. Stat.*, 54-3:257–287, 1992.
- [119] W. SCHOUTENS : *Stochastic processes and orthogonal polynomials*, volume 146 de *Lecture Notes in Statistics*. Springer-Verlag, New York, 2000.
- [120] G. SCHWERT : Tests for unit roots : a Monte Carlo investigation. *J. Bus. Econ. Stat.*, 7:147–160, 1989.
- [121] R. H. SHUMWAY et D. S. STOFFER : *Time Series Analysis and Its Applications With R Examples*. Springer-Verlag, Second Edition, New-York, 2006.
- [122] C. A. SIMS : Bayesian skepticism on unit root econometrics. *J. Econ. Dyn. Control.*, 12:463–474, 1988.
- [123] C. A. SIMS, J. H. STOCK et M. W. WATSON : Inference in linear time series models with some unit roots. *Econometrica.*, 58-1:113–144, 1990.
- [124] J. STOCK : *A Class of Tests for Integration and Cointegration*. Cointegration, Causality and Forecasting : A Festschrift for Clive W.J. Granger. R. Engle and H. White, Oxford University Press, Oxford, 1999.
- [125] T. STOCKER : On the asymptotic bias of OLS in dynamic regression models with autocorrelated errors. *Statist. Papers.*, 48:81–93, 2007.
- [126] W. F. STOUT : A martingale analogue of Kolmogorov’s law of the iterated logarithm. *Z. Wahr. Verw. Geb.*, 15:279–290, 1970.
- [127] W. F. STOUT : *Almost sure convergence*, volume 24 de *Probability and Mathematical Statistics*. Academic Press, New York-London, 1974.
- [128] J. A. TILLMAN : The power of the Durbin-Watson test. *Econometrica.*, 43:959–974, 1975.
- [129] G. E. UHLENBECK et L. S. ORNSTEIN : On the theory of brownian motion. *Phys. Rev.*, 36:823–841, 1930.
- [130] J. VON NEUMANN : Distribution of the ratio of the mean square successive difference to the variance. *Ann. Math. Stat.*, 12:367–395, 1941.
- [131] C.Z. WEI et J. WINNICKI : Estimation on the means in the branching process with immigration. *Ann. Statist.*, 18:1757–1773, 1990.

- [132] John S. WHITE : The limiting distribution of the serial correlation coefficient in the explosive case. *Ann. Math. Statist.*, 29:1188–1197, 1958.
- [133] C. S. WITHERS : Conditions for linear processes to be strong mixing. *Z. Wahr. Verw. Geb.*, 57:477–480, 1981.
- [134] J. WORMS : Moderate deviations for stable Markov chains and regression models. *Electron. J. Probab.*, 4, no. 8:1–28, 1999.
- [135] J. WORMS : Moderate deviations of some dependent variables. I. Martingales. *Math. Methods Statist.*, 10, no. 1:38–72, 2001.
- [136] J. WORMS : Moderate deviations of some dependent variables. II. Some kernel estimators. *Math. Methods Statist.*, 10, no. 2:161–193, 2001.