

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA

RECHERCHE SCIENTIFIQUE

UNIVERSITE MOULOUD MAMMARI DE TIZI-OUZOU



FACULTE DES SCIENCES

DEPARTEMENT DE MATHÉMATIQUES

Mémoire de *master*

*Spécialité : mathématiques*

*Option : probabilités et statistique*

**Thème**

**LES TECHNIQUES DE BOOTSRAP  
APPLIQUEES AUX TESTS DE LA RACINE  
UNITE**

**présenté par :**

***M<sup>elle</sup> HADDADOU Kamilia***

**Devant le jury composé de**

***Président Mr : FELLAG Hocine (professeur)***

***Encadreur Mme : ATIL Lynda (MCA)***

***Examineur Mme: BELKACEM Cherifa (MCB)***

Soutenu le 23/ 09/ 2018

# Table des matières

0.1	Introduction . . . . .	3
<b>1</b>	<b>Notions préliminaires</b>	<b>6</b>
1.1	Introduction . . . . .	7
1.2	La fonction de répartition empirique . . . . .	7
1.2.1	Loi de probabilité empirique . . . . .	7
1.2.2	Convergence de la f.d.r. empirique vers la f.d.r. théorique . . . . .	8
1.2.3	Propriétés de la fonction de répartition empirique . . . . .	11
1.3	Bootstrap . . . . .	13
1.4	Histoire du bootstrap . . . . .	13
1.5	Principe du Bootstrap . . . . .	14
1.6	L'échantillon bootstrap . . . . .	15
1.6.1	Algorithme du bootstrap . . . . .	16
1.6.2	Moyenne bootstrap, écart type bootstrap . . . . .	17
1.6.3	Algorithme : Estimation bootstrap de l'écart-type . . . . .	19
1.6.4	Bootstrap d'une série temporelle . . . . .	19
1.6.5	L'estimateur bootstrap du biais . . . . .	21
1.6.6	Propriétés asymptotiques du bootstrap . . . . .	22
1.6.7	Intervalle de Confiance et Bootstrap . . . . .	22
1.6.8	Test bootstrap . . . . .	26
<b>2</b>	<b>Les tests de racine unité</b>	<b>27</b>
2.1	Introduction . . . . .	28
2.2	Stationnarité des processus . . . . .	28
2.2.1	Stationnarité stricte . . . . .	29
2.2.2	Stationnarité au second ordre . . . . .	29
2.2.3	Processus autorégressif : . . . . .	30
2.2.4	Éléments de théorie asymptotique pour les processus AR(1) . . . . .	34
2.3	Les tests de racine unité . . . . .	36
2.3.1	Tests de Dickey et Fuller . . . . .	36
2.3.2	Tests de Dickey et Fuller augmentés . . . . .	42
2.3.3	Le test de Phillips et Perron . . . . .	46

2.3.4	Test basé sur l'estimateur du maximum de vraisemblance conditionnel	47
2.3.5	Test de Kwiatkowski, Phillips, Schmidt et Shin . . . . .	48
2.3.6	Test basés sur l'estimateur symétrique simple . . . . .	50
<b>3</b>	<b>les technique de bootstrap appliquées aux tests de racine unité</b>	<b>54</b>
3.1	Introduction . . . . .	55
3.2	Procédure de bootstrap pour les tests de racine unité . . . . .	56
3.2.1	Le Bootstrap basé sur la série différenciée . . . . .	59
3.2.2	Bootstrap basé sur les résidus non restreints . . . . .	60
3.2.3	Comparaisons . . . . .	61
3.3	Application sur un modèle AR(1) sans intercept et AR(1) et tendance linéaire	71
3.3.1	Description du modèle . . . . .	71
3.3.2	Algorithme : calcule les seuil des test de racine unité pour AR(1) . .	71
3.3.3	Tables : . . . . .	72
3.4	Conclusion générale . . . . .	74
	Références . . . . .	76

## 0.1 Introduction

La qualité de l'analyse économétrique dépend en tout état de cause de la fiabilité des statistiques de tests employées. Lorsqu'on souhaite faire de l'inférence, y compris tester une hypothèse, il est nécessaire de calculer une statistique de test et d'en connaître la loi de probabilité. On peut alors calculer un seuil critique ou une P-value et accepter ou non l'hypothèse posée. Malheureusement, la distribution de probabilité d'une statistique de test est la plupart du temps inconnue, à moins de faire des hypothèses fortes et difficilement vérifiables sur le modèle. La théorie asymptotique joue alors un rôle déterminant puisqu'elle permet de remplacer ces hypothèses : Les développements asymptotiques au premier ordre déterminent la distribution de probabilité de la statistique lorsque la taille de l'échantillon est infiniment grande, appelée loi asymptotique. En pratique le nombre de données est fini et on utilise la loi asymptotique comme approximation de la vraie loi inconnue. Toutefois, pour que la loi asymptotique soit une bonne image de la vraie loi de la statistique, il faut que le nombre de données soit suffisamment important, sinon les tests peuvent être faussés. Une des caractéristiques majeure des méthodes du bootstrap est qu'elles permettent très souvent d'obtenir une meilleure approximation de la vraie loi de la statistique que celle donnée par la loi asymptotique. La fiabilité des tests en est accrue, notamment pour des échantillons dont le nombre d'observations n'est pas très grand. Ces progrès ont des conséquences profondes en sciences car le gain de précision apporté par le bootstrap peut avoir d'importants effets sur les conclusions scientifiques. Les applications du bootstrap sont diverses et variées, en économétrie leur apport principal concerne l'amélioration de

l'inférence dans les série chronologiques. Dans ce mémoire, nous présentons un aperçu de la façon dont un test de racine unité basé sur les techniques de bootstrap peut être construit. Ensuite, nous analysons les effets de ces options sur la performance (échantillon fini) des tests.

Ce mémoire est réparti en trois chapitres.

Nous présentons dans la première partie de ce chapitre les définitions, et les propriétés de bases en statistiques paramétriques et non paramétriques, ainsi que les théories permettant d'appliquer le bootstrap en général. Dans la deuxième partie, nous nous intéressons aux principes du bootstrap et les algorithmes généraux de cette méthode, (voir Efron et Tibshirani (1993)). Nous considérons également l'utilisation du bootstrap pour le calcul de l'erreur standard et du biais d'un estimateur. Nous décrivons les théorèmes qui sont utilisés pour arriver à une bonne approximation de la distribution bootstrappée. Ainsi que la détermination des intervalles de confiance d'un paramètre estimé .

Dans le chapitre 2, nous étudions la première étape de la démarche de modélisation d'une série temporelle qui consiste à vérifier la stationnarité du processus générateur des données. Généralement, on se limite à vérifier la stationnarité faible ou stationnarité du second ordre. Ensuite nous considérons de façon de plus précise ce qu'est un processus non stationnaire. Il existe en effet deux sortes de non stationnarité : déterministe et stochastique. Nous verrons que suivant l'origine de la non stationnarité, il convient d'adopter une méthode de stationnarisation particulière. La second partie de ce chapitre sera ensuite consacrée à la présentation des principaux tests de de la racine unité . Il s'agit alors de définir une stratégie empirique permettant de vérifier si les processus sont stationnaires ou au contraire si il est

nécessaire de les stationnariser et quelle est alors la méthode appropriée.

Au troisième chapitre, Nous nous basons sur l'article de Paparoditis et Politis (2005) intitulé les techniques de bootstrap appliquées aux tests de racine unité. Nous décrivons le modèle et les hypothèses considérées, nous énonçons brièvement deux procédures de bootstrap alternatives pour tester l'hypothèse nulle de racine unité. L'une repose sur la différenciation des séries et l'autre qui est une nouvelle procédure est basée sur de des résidus non restreintes. Nous étudions les propriétés asymptotiques des différentes techniques de bootstrap sous l'hypothèse nulle et sous l'hypothèse alternative et nous analysons leur performance de puissance relative. Enfin nous illustrons l'utilisation et les avantages du bootstrap à travers des applications empiriques.

# Chapitre 1

## Notions préliminaires

## 1.1 Introduction

Les problèmes d'inférence statistique impliquent souvent d'estimer certains paramètres d'une fonction de répartition  $F$  à partir d'un échantillon aléatoire tiré de  $F$ . La fonction de distribution empirique, que nous appellerons  $F_n(x)$ , est une estimation de fonction de répartition réelle noté  $F(x)$ . Estimer un paramètre d'un échantillon de la loi  $F$ , comme sa moyenne, sa médiane ou son coefficient de corrélation, consiste à utiliser l'estimateur correspondant de  $F_n(x)$ . C'est le "principe du plug-in". La méthode bootstrap est une application directe du principe du plug-in, comme nous le verrons au chapitre 1

## 1.2 La fonction de répartition empirique

**Définition 1.1** (14). *Pour tout  $x \in \mathbb{R}$ , on appelle valeur de la fonction de répartition empirique en  $x$ , la statistique, notée  $F_n(x)$ , définie par*

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{]-\infty, x]}(X_i)$$

*En d'autres termes  $F_n(x)$  est la variable aléatoire des  $n$  observations  $X_1, X_2, \dots, X_n$  iid (indépendantes et identiquement distribuées) prenant une valeur inférieure ou égale à  $x$ . Chaque  $X_i$  ayant une probabilité  $F(x)$  d'être inférieure ou égale à  $x$ .*

### 1.2.1 Loi de probabilité empirique

A partir d'un échantillon  $(X_1, X_2, \dots, X_n)$  de  $X$ , nous définirons la loi de probabilité empirique  $P_n$  qui sera la distribution sur l'ensemble fini des valeurs  $(X_1, X_2, \dots, X_n)$ , c'est-à-dire qui attribue la même masse de probabilité  $\frac{1}{n}$  à chacun des points  $X_i$ ,  $1 \leq i \leq n$ , et



qui peut être notée sous la forme :

$$P_n(I) = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$$

Où  $\delta_a$  est la masse de Dirac au point  $a$ . Ainsi, pour tout intervalle  $I$  de  $\mathbb{R}$ , sa probabilité est égale au pourcentage de points de l'échantillon qui appartiennent à cet intervalle :

$$P_n(I) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{(X_i \in I)}$$

Cette loi de probabilité admet une fonction de répartition, noté  $F_n(x)$ , appelée fonction de répartition empirique et définie pour tout  $X$  réel par :

$$F_n(x) = \int_{-\infty}^x dP_n = P_n(]-\infty, x]) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{(]-\infty, x])}(X_i)$$

qui représente le pourcentage de points observés qui sont situés avant  $x$ .

### 1.2.2 Convergence de la f.d.r. empirique vers la f.d.r. théorique

Dans ce qui suit nous citons les théorèmes fondamentaux qui sont liés à la fonction de répartition (f.d.r) empirique qui sont la justification de l'usage des échantillons en statistique.

#### Lois des grands nombres

Elles sont de deux types : lois faibles mettant en jeu la convergence en probabilité et lois fortes relatives à la convergence presque sûre. Nous considérons ici des suites de variables aléatoires  $(X_1, X_2, \dots, X_n)$  non nécessairement de même loi.

**Loi faible des grands nombres** Soit  $(X_1, X_2, \dots, X_n)$  une suite de variables aléatoires indépendantes d'espérance  $m_1, m_2, \dots, m_n$  finies et de variance  $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$  finies.

Si  $\frac{1}{n} \sum_{i=1}^n m_i \rightarrow m$  et si  $\frac{1}{n} \sum_{i=1}^n \sigma_i^2 \rightarrow 0$  Alors :

$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  converge en probabilité vers  $m$

$$\bar{X} \xrightarrow{p} m$$

**Loi forte des grands nombres** Soit  $(X_1, X_2, \dots, X_n)$  une suite de variables aléatoires indépendantes telles que

$\frac{1}{n} \sum_{i=1}^n m_i \rightarrow m$  et si  $\sum_{i=1}^n \frac{\sigma_i^2}{i^2}$  est convergente Alors :  $\bar{X}$  converge presque sûrement vers  $m$

$$\bar{X} \xrightarrow{ps} m$$

**Le théorème central limite** Le théorème central limite (aussi improprement appelé théorème de la limite centrale ou centrée) établit la convergence en loi de la somme d'une suite de variables aléatoires vers la loi normale. Intuitivement, ce résultat affirme que toute somme de variables aléatoires indépendantes tend dans certains cas vers une variable aléatoire gaussienne.

**Théorème 1.1.** Soit  $\{X_n\}$  une suite de v.a. indépendantes de même loi admettant une moyenne  $\mu$  et une variance  $\sigma^2$ . Alors la suite  $\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$  converge en loi vers la v.a. de loi  $N(0, 1)$ , ce que nous écrivons conventionnellement

$$\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow[n \rightarrow \infty]{L} N(0, 1)$$

En ce qui concerne la fonction de répartition empirique on a :

Soit  $F_n(x)$  la fonction de répartition empirique définie comme suit :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n Y_i, \text{ telle que } Y_i = \mathbb{I}_{(X_i \leq x)}$$

donc  $F_n(x) = \bar{Y}$

selon le théorème central limite

$$\frac{\bar{Y} - F(x)}{\sqrt{\frac{F(x)(1-F(x))}{n}}} \rightarrow N(0, 1)$$

Donc

$$\bar{Y} \rightarrow N\left(F(x), \sqrt{\frac{F(x)(1-F(x))}{n}}\right)$$

### **Théorème Glivenko Cantelli**

Le théorème suivant nous dit que presque-sûrement,  $F_n$  converge uniformément vers  $F$  sur  $\mathbb{R}$ . C'est en quelque sorte une loi forte des grands nombres fonctionnelle pour la suite des f.d.r. empiriques vues comme des fonctions aléatoires. La signification pratique est que si l'on a observé un échantillon de grande taille d'une loi inconnue de f.d.r.  $F$ , la fonction de répartition empirique peut être prise comme une approximation de  $F$ .

**Théorème 1.2.** *Soit  $F_n$  la f.d.r. empirique d'un échantillon  $(X_1, X_2, \dots, X_n)$  où les  $X_i$  ont pour f.d.r.  $F$ . Alors*

$$a) \forall x \in \mathbb{R}, F_n(x) \xrightarrow[n \rightarrow +\infty]{ps} F(x)$$

$$b) D_n = \|F_n - F\|_\infty = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow[n \rightarrow +\infty]{ps} 0$$

**Le théorème de Kolmogorov** Le théorème de Kolmogorov affirme que

$$\lim_{n \rightarrow \infty} P[\sqrt{n}D_n < y] = K(y) = \sum_{k=-\infty}^{+\infty} (-1)^k \exp(-2k^2 y^2)$$

Ce théorème signifie que la distribution asymptotique de la variable aléatoire  $D_n$  est connue et ne dépend pas de la variable de départ  $X$ , et permet de calculer des limites pour les valeurs de  $D_n$  ([20])

### 1.2.3 Propriétés de la fonction de répartition empirique

#### l'estimateur de la fonction de répartition

Soit  $X_1, X_2, \dots, X_n$  des variables aléatoires réelles iid de fonction de répartition (fdr)

$F : x \rightarrow F(x) = P(X \leq x)$ . L'estimateur naturel de la fdr  $F$  est la fonction de répartition

empirique  $F_n$  défini par  $F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{]-\infty, x]}(X_i)$

#### Biais de l'estimateur $F_n(x)$

$F_n(x)$  est-elle un estimateur sans biais de  $F(x)$  ?

Nous fixons  $x \in \mathbb{R}$  et nous estimons  $F(x)$ .

Soit la suite  $(F_n(x))$  la fonction de répartition empirique définie par :

$$nF_n(x) = \sum_{i=1}^n \mathbb{I}_{x_i \leq x}(X_i)$$
$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{]-\infty, x]}(X_i) = \frac{1}{n} \sum_{i=1}^n Y_i$$

où les  $Y_i$  sont iid selon une loi de Bernoulli de paramètre  $p$  tel que

$$p = P(Y_i = 1) = P(X_i \leq x) = F(x),$$

d'où la loi de la somme des  $Y_i$  est une binomiale de paramètre  $(n, p)$  ( $B(n, p)$ )

$$E(F_n(x)) = E\left(\frac{1}{n} \sum_{i=1}^n Y_i\right) = \frac{np}{n} = p = F(x)$$
$$E\left(\sum_{i=1}^n E\{\mathbb{I}_{(X_i \leq x)}(X_i)\}\right) = P(X \leq x) = F(x)$$

Donc, pour tout point  $x$ ,  $F_n(x)$  est un estimateur sans biais de  $F(x)$ .

#### Variance de l'estimateur $F_n(x)$

Il est facile de montrer que, pour tout  $x$ , la variance de l'estimateur  $F_n(x)$  est donnée

$$\text{par : } \text{Var}(F_n(x)) = \frac{1}{n^2} V\left(\sum_{i=1}^n Y_i\right) = \frac{np(1-p)}{n^2} = \frac{F(x)(1-F(x))}{n} \rightarrow 0 \text{ quand } n \rightarrow \infty$$

## Quantiles empiriques

La fonction quantile d'une loi de probabilité est l'inverse (généralisé) de sa fonction de répartition. Si  $F$  désigne la fonction de répartition, la fonction quantile  $Q$  est la fonction qui à  $u \in ]0, 1[$  associe :

$$Q(u) = \inf\{x, tq F(x) \geq u\}$$

**Définition 1.2** (9). Soit  $(X_1, X_2, \dots, X_n)$  un  $n$ -échantillon issu d'une loi  $F$  et  $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$  l'échantillon ordonné. Soit  $p \in ]0, 1[$ , la statistique d'ordre  $X_{([np] + 1)}$  (où  $[np]$  désigne la partie entière de  $np$ ) s'appelle le quantile empirique d'ordre  $p$  de l'échantillon. En particulier  $X_{([n/2] + 1)}$  est la médiane empirique de l'échantillon. La fonction quantile empirique d'un échantillon est la fonction quantile de sa distribution empirique.

## Convergence des quantiles empiriques

Le théorème suivant montre que le quantile empirique est un estimateur du quantile théorique :

**Théorème 1.3.** Pour tout  $p \in ]0, 1[$ , si  $F$  possède un unique quantile d'ordre  $p$ , qui est alors égal à  $x_p$  (c'est-à-dire que  $F^{-1}$  (la fonction inverse de  $F$ ) est continue en  $p$ ), alors

$$X_{([np]+1)} \xrightarrow{n \rightarrow +\infty} x_p$$

Autrement dit  $X_{([np]+1)}$  est une estimation de  $x_p$  d'autant meilleure que la taille  $n$  de l'échantillon est grande.

## 1.3 Bootstrap

La motivation du bootstrap (Efron, 1982 ; Efron et Tibshirani, 1993) est d'approcher par simulation (Monte Carlo) la distribution d'un estimateur lorsque l'on ne connaît pas la loi de l'échantillon ou, plus souvent lorsque l'on ne peut pas supposer qu'elle est gaussienne. L'objectif est de remplacer des hypothèses probabilistes pas toujours vérifiées ou même invérifiables par des simulations .

Le principe fondamental de cette technique de ré-échantillonnage est de substituer à la distribution de probabilité inconnue  $F$ , dont est issu l'échantillon d'apprentissage, la distribution empirique  $F$  qui donne le même poids à chaque réalisation. Ainsi on obtient un échantillon de taille  $n$  dit échantillon bootstrap selon la distribution empirique  $F_n(X)$  par  $n$  tirages aléatoires avec remise parmi les  $n$  observations initiales.

Construire un grand nombre d'échantillons bootstrap sur lesquels on calcule l'estimateur concerné. La loi simulée de cet estimateur est une approximation asymptotiquement convergente sous des hypothèses raisonnables de la loi de l'estimateur. Cette approximation fournit ainsi des estimations du biais, de la variance, donc d'un risque quadratique, et même des intervalles de confiance de l' estimateur sans l'hypothèse (normalité) sur la vraie loi.

## 1.4 Histoire du bootstrap

La technique dite du bootstrap est une méthode d'inférence statistique introduite par Bradley Efron en 1979 qui a été utilisée initialement pour l'estimation de paramètres statistiques tels que la variance, la moyenne ou l'écart type d'un échantillon de population. Le nom de la méthode provient de l'expression anglaise "to pull oneself up by one's bootstraps" qui

se traduit littéralement par "se hisser en tirant sur les languettes de ses propres bottes" en référence au baron Von Münchhausen qui se serait sorti du marécage dans lequel il était embourbé en se tirant par les bottes. En effet, par bootstrap, on se réfère généralement à la technique de ré-échantillonnage, qui consiste à créer de nouveaux échantillons à partir des observations initiales.

Le ré-échantillonnage bootstrap est utilisé principalement pour accroître artificiellement les échantillons limités en données. À partir d'un échantillon initial issu d'une population dont on veut estimer certaines informations, on tire au sort, avec remise,  $n$  observations pour obtenir un nouvel échantillon. On répète l'opération plusieurs fois pour obtenir un grand nombre de nouveaux échantillons. Les nouvelles observations ainsi obtenues sont alors utilisées afin d'affiner l'estimation des paramètres faite sur les observations initiales.

## 1.5 Principe du Bootstrap

Le premier cadre du bootstrap est celui de la statistique classique : L'échantillon est constitué de  $n$  variables aléatoires  $(X_1, \dots, X_n)$  i.i.d de fonction de répartition  $F$  supposée inconnue. On cherche à estimer la loi de  $T_n = T(X_1, \dots, X_n)$ ,  $T_n$  étant un estimateur d'une grandeur  $\theta$  telle que :

$$\theta = t(F)$$

qui permet d'estimer la variance, le biais et les intervalles de confiance... L'idée pour cela est de substituer  $F$  par la fonction de répartition empirique  $F_n$  obtenue à partir de l'échantillon :

$$F_n(x) = \frac{1}{n} \sum_{k=1}^n \mathbb{I}_{(X_k \leq x)}$$

**Définition 1.3** (6). *On appelle estimateur plug-in d'un paramètre de  $F$ , l'estimateur obtenu en remplaçant la loi  $F$  par la loi empirique :*

$$\hat{\theta} = t(\hat{F})$$

*Dans le cas de l'estimation de la moyenne  $\hat{\mu} = E(\hat{F}) = \bar{X}$*

## 1.6 L'échantillon bootstrap

Un échantillon bootstrap est défini comme un échantillon aléatoire d'observations indépendantes et identiquement distribuées (i.i.d.) de taille  $m$  tiré avec remise de la fonction de distribution empirique  $\hat{F} = F_n$ , d'une population de  $n$  objets  $X = (X_1, \dots, X_n)$ , composé d'éléments de l'échantillon initial  $X_1, \dots, X_n$ . On dénote l'échantillon bootstrap par  $X^* = (X_1^*, \dots, X_m^*)$ . L'astérisque dans la notation indique que  $X^* = (X_1^*, \dots, X_m^*)$  n'est pas l'échantillon initial  $X = (X_1, \dots, X_n)$ . Notons que dans le bootstrap classique  $m = n$ . Cependant, Swanepoel (1986) a défini le procédé de bootstrap modifié ( $m \neq n$ ) et recommande cette méthode dans les cas où le bootstrap classique échoue.

Soit  $\theta$  le paramètre d'intérêt et  $\hat{\theta} = t(X)$  l'estimateur de  $\theta$  basé sur l'échantillon

$X = (X_1, \dots, X_n)$ .  $X^* = (X_1^*, X_2^*, \dots, X_n^*)$  est obtenu par ré échantillonnage avec remise  $n$  fois des données initiales  $X = (X_1, X_2, \dots, X_n)$ . Par exemple, considérons l'échantillon  $X = (X_1, X_2, X_3, X_4, X_5)$ , donc on aura :

$$X^*(1) = (X_2, X_4, X_2, X_5, X_3),$$



$$X^*(2) = (X_3, X_2, X_4, X_4, X_3),$$

$$X^*(3) = (X_4, X_3, X_4, X_4, X_2),$$

...

Avec chaque échantillon  $X^*(1), X^*(2), X^*(3), \dots, X^*(B)$ , nous pouvons calculer une réplique bootstrap de  $\hat{\theta}$ ,  $\hat{\theta}^*(b) = t(X^*(b))$ , (où  $b = 1, 2, \dots, B$ ), en utilisant le principe plug-in. Les problèmes de l'inférence statistique impliquent souvent l'estimation d'un certain aspect d'une distribution F basée sur un échantillon aléatoire tiré de F. La fonction de distribution empirique  $\hat{F}$  est une évaluation simple de la distribution F. La méthode de bootstrap est une application directe du principe plug-in.

### 1.6.1 Algorithme du bootstrap

L'algorithme du bootstrap peut être résumé comme suit :

**Etape 1.** Construire une distribution empirique de probabilité  $F_n$ , de l'échantillon en plaçant une probabilité de  $\frac{1}{n}$  à chaque point,  $X_1, \dots, X_n$  de l'échantillon. C'est la fonction de distribution empirique de l'échantillon, qui est l'estimation non paramétrique du maximum de vraisemblance de la distribution de la population.

**Etape 2.** De la fonction de distribution empirique  $F_n$ , tirer un échantillon aléatoire de taille n avec remise.

**Etape 3.** Calculer la statistique d'intérêt  $T_n$  pour cet échantillon, en remplaçant par  $T_n^*$ .

**Etape 4.** Répéter ( l'étape 2) et (l'étape 3) B fois, où B est un nombre assez grand, afin de créer B rééchantillons. La taille particulière de B dépend des tests à courir sur les données.

**Etape 5.** Construire l'histogramme de fréquence relative du B nombre de  $T_n^*$  en plaçant une probabilité de  $\frac{1}{B}$  à chaque point,  $T_n^{*1}, T_n^{*2}, \dots, T_n^{*B}$ . La distribution obtenue est l'estimation bootstrapée de la distribution d'échantillon  $T_n$ . Cette distribution peut être maintenant utilisée pour faire des inférences sur le paramètre  $\theta$ , qui doit être estimée par  $T_n$ .

### 1.6.2 Moyenne bootstrap, écart type bootstrap

On définit maintenant la moyenne bootstrap. Pour un ensemble d'estimateurs  $\hat{\theta}^*(b)$ ,  $b = 1, 2, 3, \dots, B$ , la moyenne est :

$$\hat{\theta}^*(.) = \frac{\sum_{b=1}^B \hat{\theta}^*(b)}{B}$$

L'écart type est aussi une caractéristique importante de chaque distribution. Pour un ensemble d'estimateurs  $\hat{\theta}^*(b)$ ,  $b = 1, 2, 3, \dots, B$ , l'écart type estimé est calculé par la formule :

$$\hat{\sigma}_B = \sqrt{\frac{\sum_{b=1}^B (\hat{\theta}^*(b) - \hat{\theta}^*(.))^2}{B - 1}}$$

**Exemple** ( l'erreur standard de la moyenne) :

Soit X un échantillon des variables aléatoires (i.i.d) avec n=9. Nous posons B=50

$$X = (52, 10, 40, 104, 50, 27, 146, 31, 46)$$

soit la moyenne  $\bar{X} = 56.22$  la statistique d'intérêt

$$X = (52, 10, 40, 104, 50, 27, 146, 31, 46)$$

↓

$$X^{*1} = (50, 10, 40, 50, 46, 10, 146, 40, 50)$$

$$X^{*2} = (10, 52, 104, 40, 104, 46, 50, 146, 27)$$

⋮

⋮

$$X^{*B} = (146, 31, 31, 10, 27, 40, 104, 46, 50)$$

Calcul de la moyenne de chaque échantillon bootstrap : Soit  $\hat{\theta}^*(1), \hat{\theta}^*(2), \dots, \hat{\theta}^*(B)$  les moyennes de  $X^{*1}, X^{*2}, \dots, X^{*B}$

$$\hat{\theta}^*(b) = \frac{1}{n} \sum_{i=1}^n X_i$$

$$\hat{\theta}^*(1) = 49.11$$

$$\hat{\theta}^*(2) = 64.33$$

⋮

$$\hat{\theta}^*(B) = 53.89$$

La moyenne de bootstrap estimée :

$$\hat{\theta}^*(.) = \frac{\sum_{k=1}^B \hat{\theta}^{*k}}{B} = 55.73$$

On calcule l'erreur standard de la moyenne :

$$SE(\theta) = \hat{\sigma}_B = \sqrt{\frac{\sum_{b=1}^B (\hat{\theta}^*(b) - \hat{\theta}^*(.))^2}{B-1}} = 13.32$$

**Remarque** La moyenne bootstrap estimée est proche de la moyenne de l'échantillon

### 1.6.3 Algorithme : Estimation bootstrap de l'écart-type

1. Tirer B échantillons bootstrap  $X^{*1}, X^{*2}, \dots, X^{*B}$  par tirage avec remise dans  $X$ .
2. Calculer la copie bootstrap  $\hat{\theta}^*(b) = t(X^*(b))$ ;  $b = 1, 2, \dots, B$ .
3. Calculer l'écart-type de l'échantillon ainsi construit :

$$\hat{\sigma}_B^2 = \frac{1}{1-B} \sum_{b=1}^B (\hat{\theta}^*(b) - \hat{\theta}^*(.))^2$$

avec

$$\hat{\theta}^*(.) = \frac{1}{B} \sum_{b=1}^B \hat{\theta}^*(b)$$

où  $\hat{\sigma}_B$  est l'approximation bootstrap de l'estimation plug-in recherchée de l'écart type de  $\hat{\theta}$ .

#### Nombre B de réplifications bootstrap nécessaires

- Même un petit nombre de réplifications fournit déjà des informations très utiles. B=50 est souvent suffisant pour une estimation fiable de l'erreur standard
- Il est rare que plus de 200 réplifications soient nécessaires pour estimer les erreurs standard

### 1.6.4 Bootstrap d'une série temporelle

Soit le modèle de régression linéaire suivant :

$$Y_t = \beta Y_{t-1} + \epsilon_t$$

#### Méthode 1 bootstrap des résidus

- calcul l'estimateur des moindres moindres carrés ordinaires (MCO) définit comme suite

$$\hat{\beta} = \frac{\sum_{t=1}^n Y_t Y_{t-1}}{\sum_{t=1}^n Y_{t-1}^2} \quad \text{et} \quad \hat{\epsilon}_t = Y_t - \hat{\beta} Y_{t-1}$$

- construire un modèle tel que les résidus soient non structurés.  $\hat{\epsilon}_t = Y_t - \hat{\beta}Y_{t-1}$
- créer plusieurs échantillons des résidus  $\hat{\epsilon}_t^*$  à partir de l'échantillon des résidus estimés.
- Reconstituer les données ou fonctions des résidus de chaque échantillon bootstrap et l'estimateur des moindres carrés comme suit  $\hat{\epsilon}_t^* = Y_t - \hat{\beta}Y_{t-1}$

- estimation de la statistique d'intérêt sur chaque série temporelle bootstrap reconstituée

**Méthode 2** Bootstrap par blocs.

- décomposition de la série en blocs indépendants
- reconstitution de séries bootstrap en joignant les blocs tirés aléatoirement avec remise
- estimation de la statistique d'intérêt sur chaque série temporelle bootstrap reconstituée

**Exemple** (bootstrap des résidus)

t	1	2	3	4	5	6	7	8	9	10	11	12
ct	2.4	2.4	2.4	2.2	2.1	1.5	2.3	2.3	2.5	2.0	1.9	1.7

Le modèle de régression linéaire est noté :

$$Y_t = \beta Y_{t-1} + \epsilon_t \text{ où } Y_t = c_t - \bar{c}_t$$

Les paramètres estimés par la méthode des moindres carrés ordinaires (MCO)  $\hat{\beta}$  et les résidus  $\hat{\epsilon}_t$  sont définis comme :

$$\hat{\beta} = \frac{\sum_{t=1}^n Y_t Y_{t-1}}{\sum_{t=1}^n Y_{t-1}^2} \text{ et } \hat{\epsilon}_t = Y_t - \hat{\beta} Y_{t-1}$$

On remplace  $Y_t$  par les valeurs de l'échantillon initial on obtient :

$$\hat{\epsilon}_t = (0.2, 0.4, -0.1, \dots, 0.2)$$

donc les échantillons bootstrap des résidus  $\hat{\epsilon}_t$  peuvent être comme suite :

$$\epsilon_t^{*1} = (0.2, 0.3, 0.2, \dots, -0.1)$$

$$\epsilon_t^{*2} = (-0.1, 0.6, -0.5, \dots, -0.3)$$

⋮

$$\epsilon_t^{*B} = (0.4, 0.4, -0.1, \dots, 0.2)$$

On calcule les  $Y_t^*(i) = \hat{\beta}Y_{t-1}^*(i) + \epsilon_t^*(i)$  pour chaque échantillons et leur estimateur des moindres carrées  $\hat{\beta}^*(i)$  pour chaque échantillon  $i$

On calcule l'erreur standart de  $\hat{\beta}$  leur définit comme suite

$$\hat{\sigma}_\beta^2 = \frac{1}{1-B} \sum_{b=1}^B (\hat{\beta} - \hat{\beta}^*(.))^2$$

avec

$$\hat{\beta}^*(.) = \frac{1}{B} \sum_{b=1}^B \hat{\beta}^*(b)$$

### 1.6.5 L'estimateur bootstrap du biais

Supposons que nous estimons le paramètre  $\theta = t(F)$  par la statistique

$$\hat{\theta} = t(\hat{F})$$

Le biais de l'estimateur  $\hat{\theta}$  est défini comme suit

$$bias(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta$$

En substituant la distribution empirique  $\hat{F}$  pour  $F$ , nous obtenons l'estimation bootstrap

du biais :  $\widehat{bias}_B = \hat{\theta}^*(.) - \hat{\theta}$

### 1.6.6 Propriétés asymptotiques du bootstrap

**Théorème 1.4** (13). Soit  $\kappa = (X_i), i = 1, \dots, n$ , une suite de v.a.r. indépendantes, de même loi  $P$  d'espérance  $m$  et de variance  $\sigma^2$ ; soit  $\kappa^* = (X_1^*, \dots, X_N^*)$  un échantillon bootstrap extrait de  $\kappa$ .

Conditionnellement à  $\kappa$ , pour  $N \rightarrow \infty$   $n \rightarrow \infty$  :

$$(i) \sqrt{N}(\bar{X}_n^* - \bar{X}_n) \xrightarrow[\text{loi}]{} N(0, \sigma^2)$$

**Théorème 1.5.** On suppose que la fonction de répartition  $F$  d'une v.a.  $X$  possède une médiane  $Md$  et une dérivée  $f$  positive et continue sur un voisinage de  $Md$ . Alors :

$$\sqrt{n}(Md_n^* - Md_n) \xrightarrow{L} N\left(0, \frac{1}{2f(Md)}\right), \quad n \rightarrow \infty$$

où  $Md_n$  est la médiane empirique et  $Md_n^*$  la médiane de l'échantillon bootstrap de taille  $n$ .

### 1.6.7 Intervalle de Confiance et Bootstrap

Il existe différentes méthodes utilisant le bootstrap et permettant d'obtenir l'intervalle de confiance, nous ne présenterons que celle qui apparaît comme étant la plus performante et la plus facile à implémenter, à savoir la méthode des quantiles et les différentes améliorations qui lui ont été apportées.

#### Intervalle de confiance standard

[6] Soit  $\hat{\theta}$  l'estimation habituelle du plug-in d'un paramètre  $\theta$  et  $\hat{\sigma}$  son écart type estimée. Considérons l'intervalle de confiance normal standard  $[\hat{\theta} - z^{(1-\alpha)} \cdot \hat{\sigma}, \hat{\theta} - z^{(\alpha)} \cdot \hat{\sigma}]$ . Les quantiles de cet intervalle peuvent être décrits d'une manière particulièrement pratique pour les calculs bootstrap. Soit  $\hat{\theta}^*$  une variable aléatoire tirée de la distribution  $N(\hat{\theta}, \hat{\sigma}^2)$ , alors  $\hat{\theta}_\alpha = \hat{\theta} -$

$z^{(1-\alpha)}\hat{\sigma}$  et  $\hat{\theta}_{1-\alpha} = \hat{\theta} - z^{(\alpha)}\hat{\sigma}$  sont les  $100\alpha$  ème et  $100(1 - \alpha)$  ème centiles de  $\hat{\theta}^*$ . En d'autres termes,  $\hat{\theta}_\alpha = \hat{\theta}^{*(\alpha)} = 100.\alpha$  ème percentile de la distribution de  $\hat{\theta}^*$ , et  $\hat{\theta}_{1-\alpha} = \hat{\theta}^{*(1-\alpha)} = 100.(1 - \alpha)$  ème percentile de la distribution de  $\hat{\theta}^*$ . Et  $z^{(\alpha)}$ ,  $z^{(1-\alpha)}$  sont les percentiles d'ordre  $\alpha$ , et  $1 - \alpha$  respectivement de la loi normale  $N(0, 1)$ .

### Intervalle de confiance par percentiles

Reposant moins sur le théorème central limite, il est fondé sur les percentiles empiriques de l'échantillon.

$$IC_{1-\alpha} = [W_{\frac{\alpha}{2}}^*; W_{1-\frac{\alpha}{2}}^*]$$

Où  $W_q^*$  représente le quantile d'ordre  $q$  de  $\hat{\theta}^*$ .

Afin d'obtenir une précision satisfaisante sur les percentiles, il faut au moins un millier d'échantillons de bootstrap pour un intervalle de confiance à 95%.

### Intervalle de confiance basique

On peut noter que l'intervalle des percentiles est inadapté aux situations d'asymétrie de la distribution de  $\hat{\theta}^*$ , ce qui arrive notamment en cas d'hétéroscédasticité, c'est-à-dire que la variance de  $\hat{\theta}$  est corrélée à  $\hat{\theta}$  lui même. Les fluctuations d'échantillonnage dans une direction se traduisant par une incertitude dans l'autre sens, l'intervalle de confiance basique renverse les bornes autour de  $\hat{\theta}$ .

$$IC_{1-\alpha} = [2\hat{\theta} - W_{1-\frac{\alpha}{2}}^*; 2\hat{\theta} - W_{\frac{\alpha}{2}}^*]$$

**Exemple** : Nous commençons avec un ensemble de données simulé suffisamment petit pour



montrer explicitement chaque étape. L'échantillon de données est :

$$E = \{30, 37, 36, 43, 42, 43, 43, 46, 41, 42\}$$

L'objectif est d'estimer la moyenne  $\theta$  de la distribution sous-jacente et donner un intervalle de confiance bootstrap de 80%

La moyenne de l'échantillon est  $\hat{\theta} = 40.3$ . Nous l'utilisons comme une estimation de la moyenne réelle  $\theta$  de la distribution sous-jacente. Pour déterminer l'intervalle de confiance, il faut savoir combien la distribution de  $\hat{\theta}$  varie autour de  $\theta$ . Nous aimerions connaître la distribution de

$$\gamma = \hat{\theta} - \theta$$

Si nous connaissions cette distribution, nous pourrions trouver  $\gamma_{.1}$  et  $\gamma_{.9}$ , les valeurs critiques de 0.1 et 0.9 de  $\gamma$ . Ensuite, nous aurions

$$P[\gamma_{.9} \leq \hat{\theta} - \theta \leq \gamma_{.1}/\theta] = 0.8 \Leftrightarrow P[\hat{\theta} - \gamma_{.9} \geq \theta \geq \hat{\theta} - \gamma_{.1}/\theta] = 0.8$$

ce qui donne un intervalle de confiance de 80% de

$$[\hat{\theta} - \gamma_{.1}; \hat{\theta} - \gamma_{.9}]$$

Comme toujours avec les intervalles de confiance, nous nous efforçons de souligner que les probabilités calculées ci-dessus sont des probabilités concernant la statistique  $\hat{\theta}$  étant donné que la vraie moyenne est  $\theta$ .

Le principe du bootstrap offre une approche pratique pour estimer la distribution de

$\gamma = \hat{\theta} - \theta$ . Il dit que nous pouvons l'approcher par la distribution de

$$\gamma^* = \hat{\theta}^* - \hat{\theta}$$

où  $\hat{\theta}^*$  est la moyenne d'un échantillon empirique de bootstrap.

Puisque  $\gamma^*$  est calculé en rééchantillonnant les données originales, nous pouvons faire en sorte que l'ordinateur simule  $\gamma^*$  autant de fois que nous le souhaitons. Ainsi, par la loi des grands nombres, on peut estimer la distribution de  $\gamma^*$  avec une grande précision.

Revenons maintenant aux données de l'échantillon avec 10 points. Nous avons utilisé R pour générer 20 échantillons bootstrap, chacun de taille 10. Chacune des 20 colonnes du tableau suivant est un échantillon bootstrap.

43	36	46	30	43	43	43	37	42	42	43	37	36	42	43	43	42	43	42	43
43	41	37	37	43	43	46	36	41	43	43	42	41	43	46	36	43	43	43	42
42	43	37	43	46	37	36	41	36	43	41	36	37	30	46	46	42	36	36	43
37	42	43	41	41	42	36	42	42	43	42	43	41	43	36	43	43	41	42	46
42	36	43	43	42	37	42	42	42	46	30	43	36	43	43	42	37	36	42	30
36	36	42	42	36	36	43	41	30	42	37	43	41	41	43	43	42	46	43	37
43	37	41	43	41	42	43	46	46	36	43	42	43	30	41	46	43	46	30	43
41	42	30	42	37	43	43	42	43	43	46	43	30	42	30	42	30	43	43	42
46	42	42	43	41	42	30	37	30	42	43	42	43	37	37	37	42	43	43	46
42	43	43	41	42	36	43	30	37	43	42	43	41	36	37	41	43	42	43	43

Ensuite, nous calculons  $\gamma^*(i) = \hat{\theta}^*(i) - \hat{\theta}$  ( $i = 1, \dots, 20$ ) pour chaque échantillon bootstrap (c'est-à-dire chaque colonne) et les trions du plus petit au plus grand :

$-1.6, -1.4, -1.4, -0.9, -0.5, -0.2, -0.1, 0.1, 0.2, 0.2, 0.4, 0.4, 0.7, 0.9, 1.1, 1.2, 1.2, 1.6, 1.6, 2.0$

Nous rapprocherons les valeurs critiques  $\gamma_{.1}$  et  $\gamma_{.9}$  par  $\gamma_{.1}^*$  et  $\gamma_{.9}^*$ . Puisque  $\gamma_{.1}^*$  est au 90 ème percentile nous choisissons le 18 ème élément de la liste, ie (1.6). De même , puisque  $\gamma_{.9}^*$  est au 10 ème percentile, nous choisissons le 2 ème élément de la liste, ie (-1.4). Par conséquent, notre intervalle de confiance à 80% pour  $\theta$  est

$$[\hat{\theta} - \gamma_{.1}^*, \hat{\theta} - \gamma_{.9}^*] = [40.3 - 1.6, 40.3 + 1.4] = [38.7, 41.7]$$

### 1.6.8 Test bootstrap

Soit  $X = (X_1, X_2, \dots, X_N)$   $\theta$  la loi de l'échantillon et  $\hat{\theta}$  la moyenne de l'échantillon observé

Dans ce paragraphe nous montrons la procédure à suivre pour tester si la moyenne de l'échantillon  $\hat{\theta}$  observé vaut-elle la vraie moyenne  $\theta$

Donc l'hypothèse nulle est  $H_0 : \hat{\theta} = \theta$

- Tirer B échantillons bootstrap  $X^*(b)$  de taille N à partir de x
- Pour chaque échantillon bootstrap, calculer :

$$t(X^*(b)) = \frac{\theta^*(b) - \hat{\theta}}{\sqrt{\sigma^{2*}(b)/N}}$$

avec

$$\theta^*(b) = \sum_{i=1}^N X_i^*(b)/N$$

$$\sigma^{2*}(b) = \sum_{i=1}^N (X_i^*(b) - \theta^*(b))^2 / (N - 1) \quad \text{et} \quad t(X) = \frac{\hat{\theta} - \theta}{\sqrt{\sigma^2/N}}$$

- Calculer le niveau de signification atteint(ASL) par :

$$ASL = \mathbb{I}_{(t(X^*(b)) \geq t(x))} / B$$

- Si  $ASL < \alpha$ . rejeter  $H_0$

## Chapitre 2

# Les tests de racine unit 

## 2.1 Introduction

Dans la première partie du chapitre, nous allons voir qu'une des première étape de l'approche de modélisation d'une série chronologique consiste à vérifier la stationnarité du processus générateur de données. nous donnons un aperçu sur la notion de stationnarité d'une série temporelle. Nous allons voir ensuite qu'il existe plusieurs sortes de non stationnarité : la non stationnarité déterministe et la non stationnarité stochastique. Nous verrons que suivant l'origine de la non stationnarité, il convient d'adopter une méthode de stationnarisation particulière. La seconde partie de ce chapitre sera ensuite consacrée à la présentation des principaux tests de non stationnarité. Il s'agit alors de définir une stratégie empirique permettant de vérifier si les processus sont stationnaires ou au contraire si il est nécessaire de les stationnariser et quelle est alors la méthode appropriée.

## 2.2 Stationnarité des processus

Dans ce paragraphe, nous nous proposons de donner un aperçu sur la notion de stationnarité d'une série chronologique que l'on appelle aussi série temporelle. Rappelons que celle-ci est un ensemble de valeurs représentant l'évolution d'un phénomène au cours du temps sur une période donnée.

Dans ce qui suit nous utiliserons le terme "séries chronologiques" pour désigner à la fois les données et le processus dont elles sont les réalisations.

La décomposition classique d'une série chronologique est :

$$Y_t = d_t + S_t + X_t$$

Où

- $d_t$  est le mouvement de longue durée appelée "tendance" (trend) qui est une fonction continue du temps.
- $S_t$  est la composante saisonnière qui correspond aux fluctuations saisonnières autour de la tendance.
- $\epsilon_t$  est la composante aléatoire.

Beaucoup de méthodes d'analyse de séries chronologiques reposent sur l'hypothèse de stationnarité de ces dernières. Nous donnons ci-après quelques éléments utiles à la compréhension de la notion de non stationnarité et l'intérêt des tests de racine unité.

### 2.2.1 Stationnarité stricte

Le processus  $(X_t)$  est dit strictement stationnaire si les vecteurs  $(X_1, X_2, \dots, X_k)'$  et  $(X_{1+h}, X_{2+h}, \dots, X_{k+h})'$  ont même loi jointe, pour tout entier  $k$  et tout entier relatif  $h$ .

### 2.2.2 Stationnarité au second ordre

Un processus  $\{X_t, t \in \mathbb{Z}\}$  est dit stationnaire au second ordre, ou stationnaire au sens faible, ou stationnaire d'ordre deux si les trois conditions suivantes sont satisfaites :

- $\forall t \in \mathbb{Z}, E(X_t) = m$  tel que  $m$  est une constante indépendante de  $t$
- $\forall t \in \mathbb{Z}, E(X_t^2) = m' < \infty$  tel que  $m'$  est une constante indépendante de  $t$

Donc la variance  $V(X_t) = m' - m^2 = \sigma^2$  est aussi une constante indépendante de  $t$

- $\forall (t, h) \in \mathbb{Z}^2, cov(X_t, X_{t+h}) = E[(X_{t+h} - m)(X_t - m)] = \gamma(h)$  est indépendante de  $t$

**Définition 2.1.** *Fonction d'autocovariance*

Soit  $\{X_t, t \in \mathbb{Z}\}$  un processus tel que  $V(X_t) < +\infty$  pour tout  $t \in \mathbb{Z}$ . La fonction d'autocovariance  $\gamma_X(\cdot, \cdot)$  de  $\{X_t\}$  est définie par :

$$\gamma_X(r, s) = \text{cov}(X_r, X_s) = E[(X_r - E(X_r))(X_s - E(X_s))] , r, s \in \mathbb{Z}$$

dans le cas d'un processus stationnaire, la fonction d'autocovariance sera une fonction d'une seule variable :

$$\gamma_X(h) = \gamma_X(h, 0) = \text{cov}(X_t, X_{t+h}), \forall t, h \in \mathbb{Z}$$

**Définition 2.2.** *Fonction d'autocorrélation* : La fonction d'autocorrélation d'un processus stationnaire  $\{X_t, t \in \mathbb{Z}\}$  est définie par :

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \text{corr}(X_t, X_{t+h}) = \frac{\text{cov}(X_t, X_{t+h})}{[V(X_t)V(X_{t+h})]^{\frac{1}{2}}}, \forall t, h \in \mathbb{Z}$$

### 2.2.3 Processus autorégressif :

[16] On dit que  $\{X_t, t \in \mathbb{Z}\}$  est un processus autoregressif d'ordre  $p$  noté AR( $p$ ) s'il satisfait l'équation :

$$X_t + a_1 X_{t-1} + \dots + a_p X_{t-p} = \epsilon_t \tag{2.1}$$

où  $a_1, a_2, \dots, a_p$  sont des constantes et  $\epsilon_t$  un processus purement aléatoire de moyenne  $\mu_\epsilon$  et de variance  $\sigma_\epsilon^2$ .

En utilisant l'opérateur de retard  $B$  défini par  $BX_t = X_{t-1}$ , l'équation (2.1) peut s'écrire :

$$\alpha(B)X_t = \epsilon_t \tag{2.2}$$

où

$$\alpha(z) = 1 - a_1z - \dots - a_pz^p$$

une solution de l'équation **(2.1)** est donnée par :

$$X_t = \frac{1}{\alpha(B)} \epsilon_t \tag{2.3}$$

En posant

$$\alpha(B) = \prod_{i=1}^p (1 - \mu_i B)$$

et en décomposant  $\alpha^{-1}(B)$  en éléments simples on obtient :

$$X_t = \alpha^{-1}(B)\epsilon_t = \left[ \sum_{i=1}^p \frac{k_i}{1 - \mu_i B} \right] \epsilon_t \tag{2.4}$$

Les  $\mu_i$  sont les racines du polynôme caractéristique  $g(z) = z^p - a_1z^{p-1} - \dots - a_p$ , et sont aussi l'inverse des racines du polynôme  $\alpha(z)$ , les  $k_i$  sont des constantes.

Si  $|\mu_i| < 1$ , pour tout  $i$ , chaque terme de cette somme peut être développé en une série convergente de  $\epsilon_t, \epsilon_{t-1}, \dots$ , et  $X_t$  sera une somme de séries convergentes qui représente la solution stationnaire de l'équation **(2.1)**.

Si les racines de  $g(z)$  sont toutes à l'extérieur du cercle unité, la solution **(2.4)** ne serait pas convergente. Mais il est toujours possible d'obtenir une solution stationnaire de **(2.1)**.

Seulement, elle sera en fonction de  $\epsilon_t, \epsilon_{t-1}, \dots$ , i.e  $X_t = \sum_{s=0}^{\infty} k_s \epsilon_{t+s}$

Si quelques racines de  $g(z)$  sont à l'intérieur et d'autres à l'extérieur du cercle unité, la solution de **(2.1)** sera en fonction de  $\dots, \epsilon_{t-2}, \epsilon_{t-1}, \epsilon_t, \epsilon_{t+1}, \epsilon_{t+2}, \dots$

Notons qu'un tel processus c'est-à-dire le processus dont les racines du polynôme  $g(z)$  sont à l'extérieur du cercle unité, admet une autre représentation autoregressive, tel que son polynôme caractéristique ait toutes ses racines à l'intérieur du cercle unité.



Dorénavant, en écrivant processus autorégressif (AR) stationnaire, on se réfère au cas où la solution est en fonction des valeurs passées de  $\epsilon_t$  seulement, i.e  $\{\epsilon_s, s < t\}$ .

Le cas le plus délicat à traiter est celui où  $g(z)$  a une racine ou plus sur le cercle unité (admet une racine unité) i.e il existe  $i$  tel que  $|\mu_i| = 1$ .

Dans ce cas, l'équation (2.1) n'admet pas de solution stationnaire.

Donc la condition de stationnarité d'un processus autorégressif est que toutes les racines de son polynôme caractéristique soient à l'intérieur du cercle unité.

### Cas particulier : Processus autorégressif d'ordre 1

On dit  $X_t, t \in \mathbb{Z}$  est un processus autorégressif d'ordre 1 qu'on note AR(1) s'il satisfait l'équation :

$$X_t = aX_{t-1} + \epsilon_t \quad (2.5)$$

Où  $a$  est le paramètre autorégressif (un nombre réel) et  $\epsilon_t$  un bruit blanc de moyenne  $\mu_\epsilon$  et de variance  $\sigma_\epsilon^2$

Le polynôme caractéristique de l'équation (2.5) est

$$g(z) = z - a$$

$$g(z) = 0 \Rightarrow z = a$$

La racine du polynôme caractéristique est le paramètre autorégressif.

Donc, la condition de stationnarité est :  $|a| < 1$

Dans ce cas

$$X_t = \sum_{s=0}^{\infty} a^s \epsilon_{t-s} \quad (2.6)$$

représente la solution stationnaire de l'équation (2.5).

Si  $|a| > 1$ , la série (2.6) ne converge pas, cependant on peut réécrire l'équation (2.4) sous la forme

$$X_t = a^{-1}X_{t+1} - a^{-1}\epsilon_{t+1}$$

Par substitution, on obtient :

$$\begin{aligned} X_t &= -a^{-1}\epsilon_{t+1} + a^{-2}\epsilon_{t+2} + a^{-2}X_{t+2} \\ &\quad \vdots \\ &= -a^{-1}\epsilon_{t+1} - a^{-2}\epsilon_{t+2} - \dots - a^{-k}\epsilon_{t+k} - a^{-k}X_{t+k} \\ X_t &= -\sum_{s=1}^{\infty} a^{-s}\epsilon_{t+s} \end{aligned} \tag{2.7}$$

est une solution stationnaire de (2.5)

La stationnarité de (2.7) est vue comme étant anormal puisque  $X_t$  dépend de  $\{\epsilon_s, s > t\}$ . Il est habituel dans la modélisation des séries chronologiques de se restreindre pour les AR(1) au cas  $|a| < 1$ , puisque tous les AR(1) avec  $|a| > 1$  peuvent être reformulés en un AR(1) avec  $|a| < 1$ , et une autre séquence de bruit blanc.

Si  $|a| = 1$ , l'équation (2.5) n'admet pas de solution stationnaire.

Dans ce cas, on suppose  $X_0 = 0$ ,  $X_t$  sera la somme de  $t$  variables aléatoires non corrélées de moyenne nulle et de variance  $\sigma_\epsilon^2$

$$X_t = \sum_{s=1}^t \epsilon_s$$

$E(X_t X_t) = t\sigma_\epsilon^2$  alors  $X_t$  n'est pas stationnaire. Si  $\epsilon_t$  sont indépendantes,  $X_t$  correspond au modèle de marche aléatoire.

## Non stationnarité stochastique

On dit que le processus  $X_t$  est caractérisé par une non stationnarité stochastique, ou encore que le processus  $X_t$  est DS (Difference stationary) si le processus différencié une fois  $(1 - B)X_t$  est stationnaire. Nous aborderons seulement les processus DS de première ordre, i.e les processus ne contenant qu'une seule racine unité.

## Non stationnarité déterministe

On dit que le processus  $X_t$  est caractérisé par une non stationnarité déterministe, ou encore que le processus  $X_t$  est TS (Trend stationary) s'il peut s'écrire :

$$X_t = f(t) + \epsilon_t$$

où  $f(t)$  est une fonction qui dépend du temps et  $\epsilon_t$  est un processus stationnaire.

Ainsi, ce processus est rendu stationnaire en lui enlevant sa tendance déterministe :

$$X_t - f(t) = \epsilon_t \text{ est stationnaire}$$

### 2.2.4 Eléments de théorie asymptotique pour les processus AR(1)

On cherche à connaître les propriétés asymptotiques d'un estimateur des moindres carrés dans des processus autorégressifs d'ordre 1 qui nous permettront de construire des tests de l'hypothèse de non stationnarité.

## Mouvement Brownien

**Définition 2.3.** *Un processus de Wiener ou mouvement Brownien est un processus stochastique en temps continu que l'on note  $W(t)$  et qui est défini par les trois propriétés suivantes :*

1) Le processus démarre en  $t = 0$  à la valeur  $0$ , c'est-à-dire  $P(W(0) = 0) = 1$

2) Les accroissements du processus sont stationnaires et indépendants. Pour

$0 \leq t_1 \leq t_2 \cdots \leq t_n$ , les accroissements  $W(t_i) - W(t_{i-1})$  sont des variables aléatoires indépendantes telles que :

$$E[W(t_i) - W(t_{i-1})] = 0$$

$$\text{Var}[W(t_i) - W(t_{i-1})] = \sigma^2(t_i - t_{i-1}).$$

3) Les accroissements  $W(t_i) - W(t_{i-1})$  ont une distribution normale de moyenne nulle et de variance  $\sigma^2(t_i - t_{i-1})$

Si  $\sigma^2 = 1$ , le processus est dit standard.

**Proposition 2.1.** On considère un processus  $X_t$  satisfaisant une représentation  $AR(1)$  non stationnaire telle que :

$$X_t = X_{t-1} + \epsilon_t$$

avec  $X_0 = 0$  et où  $\epsilon_t$  iid  $(0, \sigma^2)$ . Les distributions asymptotiques des principaux moments empiriques de  $X_t$  sont alors les suivantes :

$$T^{-1} \sum_{t=1}^T X_{t-1} \epsilon_t \xrightarrow[T \rightarrow \infty]{L} \sigma_\epsilon^2 [W(1)]^2 - 1$$

$$T^{-2} \sum_{t=1}^T X_{t-1}^2 \xrightarrow[T \rightarrow \infty]{L} \sigma_\epsilon^2 \int_0^1 [W(t)]^2 dt$$

où  $W(t)$  est processus de Wiener

## 2.3 Les tests de racine unité

Nous considérons la série temporelle  $\{X_t, t \in \mathbb{Z}\}$  de la forme générale suivante :

$$X_t = d_t + Y_t$$

avec  $d_t$  la tendance déterministe et  $Y_t$  la composante aléatoire. La tendance déterministe peut avoir trois formes :

- 1)  $d_t = 0$ , la série n'admet pas de tendance purement déterministe, ce qui correspond au modèle sans intercept (moyenne nulle).
- 2)  $d_t = \beta$ , la tendance déterministe se résume à une constante, ce qui correspond au modèle avec intercept (moyenne non nulle )
- 3)  $d_t = \beta + \alpha t$ , la tendance est une fonction linéaire du temps.

### 2.3.1 Tests de Dickey et Fuller

Dickey et Fuller (1979) sont les premiers à fournir un ensemble d'outils statistiques formels pour détecter la présence d'une racine unitaire dans un processus purement autorégressif du premier ordre AR(1). Cette procédure de test, maintenant bien connue, est fondée sur l'estimation par la méthode des moindres carrés, sous l'hypothèse alternative, de trois modèles autorégressifs du premier ordre dont les erreurs sont identiquement et indépendamment distribuées : le modèle sans constante, le modèle avec constante et le modèle avec constante et tendance.

#### Modèle 1 : AR(1) sans intercept

Soit le modèle

$$X_t = \rho X_{t-1} + \epsilon_t \tag{2.8}$$

où  $X_0 = 0$ ,  $\rho$  le paramètre autorégressif ( un nombre reel) et les  $\epsilon_t$  sont iid de loi  $N(0, 1)$ .

Sous  $H_0 : \rho = 1$ ,  $X_t = X_{t-1} + \epsilon_t$  est un modèle de marche aléatoire sans intercept.

Sous  $H_1 : |\rho| < 1$ ,  $X_t = \rho X_{t-1} + \epsilon_t$  est un AR(1) stationnaire de moyenne nulle.

L'estimateur des moindres carrés  $\hat{\rho}_{mc}$  de  $\rho$  est

$$\hat{\rho}_{mc} = \frac{\sum_{t=1}^n X_t X_{t-1}}{\sum_{t=1}^n X_{t-1}^2}$$

Pour exécuter ce test, on va pouvoir employer deux statistiques. La première est la statistique de Student usuelle, que l'on va noter  $\hat{\tau}$ , pour bien souligner le fait qu'elle a une distribution asymptotique qui n'est pas normale :

$$\hat{\tau} = \frac{\hat{\rho}_{mc} - 1}{\hat{\sigma}_\epsilon}$$

La seconde statistique est  $z = T(\hat{\rho}_{mc} - 1)$  basée sur le fait qu'il converge en distribution vers une fonctionnelle de Browniens qui ne dépend pas du paramètre de nuisance  $\sigma^2$ .

Commençons par la statistique  $z$  qui est la plus simple. On a vu que :

$$\begin{aligned} \hat{\rho}_{mc} &= \frac{\sum_{t=1}^n X_t X_{t-1}}{\sum_{t=1}^n X_{t-1}^2} = \frac{\sum_{t=1}^n (\rho X_{t-1} + \epsilon_t) X_{t-1}}{\sum_{t=1}^n X_{t-1}^2} \\ &= \frac{\sum_{t=1}^n \rho X_{t-1}^2 + \sum_{t=1}^n \epsilon_t X_{t-1}}{\sum_{t=1}^n X_{t-1}^2} = \rho + \frac{\sum_{t=1}^n \epsilon_t X_{t-1}}{\sum_{t=1}^n X_{t-1}^2} \\ T(\hat{\rho}_{mc} - \rho) &= \frac{T^{-1} \sum_{t=1}^n \epsilon_t X_{t-1}}{T^{-2} \sum_{t=1}^n X_{t-1}^2} \end{aligned}$$

Donc d'après la proposition **2.1**

$$z = T(\hat{\rho}_{mc} - 1) = \frac{T^{-1} \sum_{t=1}^n \epsilon_t X_{t-1}}{T^{-2} \sum_{t=1}^n X_{t-1}^2} \xrightarrow{L} \frac{\frac{1}{2}(W^2(1) - 1)}{\int_0^1 W^2(t) dt}$$

ce qui a permis de démontrer que  $\hat{\rho}_{mc}$  est un estimateur super consistant de  $\rho$ .

Nous nous intéressons maintenant à la statistique  $\hat{\tau}$ . Sa distribution sous  $H_0$  peut facilement se déduire de la distribution de la statistique  $z$ . En effet :

$$\hat{\tau} = \frac{T(\hat{\rho}_{mc} - 1)}{(\hat{\sigma}_\epsilon^2)^{\frac{1}{2}}(T^{-2} \sum X_{t-1}^2)^{-1}}$$

$$\hat{\tau} = \frac{T^{-1} \sum X_{t-1} \epsilon_t}{T^{-2} \sum X_{t-1}^2} \frac{(T^{-2} \sum X_{t-1}^2)^{\frac{1}{2}}}{(\hat{\sigma}_\epsilon^2)^{\frac{1}{2}}}$$

$$\hat{\tau} = \frac{T^{-1} \sum X_{t-1} \epsilon_t}{(T^{-2} \sum X_{t-1}^2)^{\frac{1}{2}} (\hat{\sigma}_\epsilon^2)^{\frac{1}{2}}}$$

Donc d'après la proposition **2.1**

$$\hat{\tau} \rightarrow \frac{\sigma_\epsilon^2 (W^2(1) - 1)}{2(\sigma_\epsilon^2 \int_0^1 W^2(t) dt)^{\frac{1}{2}} \sigma_\epsilon} = \frac{(W^2(1) - 1)}{2 \int_0^1 W^2(t) dt}$$

## Modèle 2 : AR(1) avec intercept

Soit le modèle

$$X_t = \beta + \rho X_{t-1} + \epsilon_t \tag{2.9}$$

Où  $X_0 = 0$ ,  $\rho$  le paramètre autorégressif,  $\beta = (1 - \rho)\mu$ ,  $\mu = E(X_t)$  et  $\epsilon_t$  sont iid de loi  $N(0, \sigma^2)$ .

Sous  $H_0 : \rho = 1$ ,  $X_t = X_{t-1} + \epsilon_t$  est un modèle de marche aléatoire sans intercept.

Sous  $H_1 : |\rho| < 1$ ,  $X_t = \beta + \rho X_{t-1} + \epsilon_t$  est un AR(1) stationnaire de moyenne non nulle.

Dans ce cas l'estimateur des moindres carrés  $\hat{\rho}_\mu$  de  $\rho$  est :

$$\hat{\rho}_\mu = \frac{\sum_{t=2}^n (X_t - \bar{X}_0)(X_{t-1} - \bar{X}_{-1})}{\sum_{t=2}^n (X_{t-1} - \bar{X}_{-1})^2}$$

Où  $\bar{X}_0 = \frac{1}{n-1} \sum_{t=2}^n X_t$  et  $\bar{X}_{-1} = \frac{1}{n-1} \sum_{t=2}^n X_{t-1}$

La statistique  $t$  associée à cet estimateur est :

$$\bar{\tau}_\mu = \frac{\hat{\rho}_\mu - 1}{\sqrt{\hat{V}(\hat{\rho}_\mu)}}$$

Où  $\hat{V}(\hat{\rho}_\mu) = \frac{s^2}{\sum_{t=2}^n (X_{t-1} - \bar{X}_{(-1)})^2}$  et  $s^2 = \frac{1}{n-3} \sum_{t=2}^n [X_t - \bar{X}_{(0)} - \hat{\rho}_\mu (X_{t-1} - \bar{X}_{(-1)})]^2$

**Théorème 2.1.** (Dickey et Fuller, 1979) Soit le modèle (2.9), sous l'hypothèse nulle de la racine unité  $H_0 : \rho = 1$ , on a

$$n(\hat{\rho}_\mu - 1) \xrightarrow{L} \frac{\frac{1}{2}(T^2 - 1) - TH}{G - H^2}$$

$$\hat{\tau}_\mu \xrightarrow{L} \frac{\frac{1}{2}(T^2 - 1) - TH}{\sqrt{G - H^2}}$$

Où  $H = \int_0^1 W(t)dt$  avec  $W(t)$  est le processus Wiener, standard.  $G = \int_0^1 W^2(t)dt$  et  $T =$

$W(1)$

### Modèle 3 : AR(1) avec tendance linéaire

Soit le modèle

$$X_t = \beta_1 + \alpha_1 t + Y_t \tag{2.10}$$

Où  $\alpha_1$  et  $\beta_1$  sont des paramètres réels non nuls et  $Y_t = \rho Y_{t-1} + \epsilon_t$

En remplaçant  $Y_t$  dans (2.10) on obtient

$$X_t = \beta + \alpha t + \rho X_{t-1} + \epsilon_t \tag{2.11}$$

avec  $\beta = (1 - \rho)\beta_1 + \rho\alpha_1$  et  $\alpha = (1 - \rho)\alpha_1$ ,  $X_0 = 0$  et  $\epsilon_t$  sont iid de loi  $N(0, \sigma^2)$ .



Sous  $H_0 : \rho = 1$ ,  $X_t = \beta + X_{t-1} + \epsilon_t$  est un modèle de marche aléatoire avec intercept.

Sous  $H_1 : |\rho| < 1$ ,  $X_t = \beta + \alpha t + \rho X_{t-1} + \epsilon_t$  est un AR(1) avec tendance linéaire qui devient stationnaire par écart à cette tendance.

L'estimateur des moindres carrés  $\hat{\rho}_\tau$  de  $\rho$  est :

$$\hat{\rho}_\tau = \frac{\sum_{t=2}^n X_t (X_{t-1} - \bar{X}_{(-1)})}{\sum_{t=2}^n (X_{t-1} - \bar{X}_{(-1)})^2}$$

Où  $\bar{X}_{(-1)} = \frac{1}{n-1} \sum_{t=2}^n X_{t-1}$

La statistique t associée à cet estimateur est :

$$\bar{\tau}_\tau = \frac{\hat{\rho}_\tau - 1}{\sqrt{\hat{V}(\hat{\rho}_\tau)}}$$

Où  $\hat{V}(\hat{\rho}_\tau) = \frac{s^2}{\sum_{t=2}^n (X_{t-1} - \bar{X}_{(-1)})^2}$  et  $s^2 = \frac{1}{n-3} (\sum_{t=2}^n (X_t - \hat{\rho}_\tau [X_{t-1} - \bar{X}_{(-1)}])^2)$

**Théorème 2.2.** (Dickey et Fuller, 1979) Soit le modèle (2.10), sous l'hypothèse nulle de la racine unité  $H_0 : \rho = 1$  on a

$$n(\hat{\rho}_\tau - 1) \xrightarrow{L} \frac{(T - 2H)(T - 6K) - 1}{2(G - H^2 - 3K^2)}$$

$$\hat{\tau}_\tau \xrightarrow{L} \frac{(T - 2H)(T - 6K) - 1}{2(\sqrt{G - H^2 - 3K^2})}$$

Où  $K = 2 \int_0^1 tW(t)dt - H$  et  $G$ ,  $T$  et  $H$  sont définis dans le théorème 2.1.  $W(t)$  est le processus de Wiener standard.

## Remarques

- Les distributions des statistiques ci-dessus sous  $H_0 : \rho = 1$  sont tabulées. On trouve les tables dans Fuller(1996)[6]. On les donne en annexe (table B1etB2).
- Les statistiques  $\hat{\tau}$ ,  $\hat{\tau}_\mu$  et  $\hat{\tau}_\tau$  gardent leur nom "statistique de student" ou "statistique t" bien qu'elle ne suivent plus une loi de student.
- Moyennant une lecture différente, on peut utiliser ces tables pour approximer les distributions quand  $\rho = -1$  puisque elles sont l'image opposée des distributions quand  $\rho = 1$ .

## Exemple ( Dickey et Fuller,1979)

Gould et Nelson (1974) étudient la structure stochastique de la vélocité de la monnaie, ils disposent de 92 observations de 1869 à 1960, Gould et Nelson ont conclu que le logarithme de cette série est bien approximé par les deux modèles :

$$X_t - X_1 = \rho(X_{t-1} - X_1) + \epsilon_t$$

$$X_t = \rho X_{t-1} + \epsilon_t$$

$$\text{Où } \epsilon_t \sim N(0, \sigma^2)$$

L'estimation des modèles par les moindres carrés donne :

$$\hat{X}_t - X_1 = \overset{(0.0094)}{1.0044}(X_{t-1} - X_1), \quad \sigma^2 = 0.0052$$

$$\hat{X}_t = \overset{(0.0176)}{0.0141} + \overset{(0.0199)}{0.9702}X_{t-1}, \quad \sigma^2 = 0.0050$$

Les nombres entre parenthèse sont les erreurs standards.

pour le premier modèle  $n(\hat{\rho}_{mc} - 1) = 91(0.0044) = 0.4004$

Et  $\hat{\tau} = (0.0094)^{-1}(0.0044) = 0.4681$

En utilisant les tables de Fuller (1996), l'hypothèse  $H_0 : \rho = 1$  est acceptée au seuil de 10%,

voire annexe première partie des tables B1 et B2).

Pour le deuxième modèle  $n(\hat{\rho}_\mu - 1) = 92(0.9702 - 1) = -2.742$

Et  $\hat{\tau}_\mu = (0.0199)^{-1}(0.9702 - 1) = -1.50$

Là aussi, l'hypothèse nulle de la racine unité est acceptée au seuil de 10%, voire annexe deuxième partie des tables B1 et B2).

### 2.3.2 Tests de Dickey et Fuller augmentés

[1] Dans les modèles **1,2** et **3**, utilisés pour les tests de Dickey et Fuller simples, dans le processus, les erreurs utilisées sont bruit blanc. Or il n'y a aucune raison pour que, a priori, l'erreur soit non corrélée, on appelle test de Dickey Fuller augmentés (1981)[3] la prise en compte de l'hypothèse de erreur soit corrélée. Considérons par exemple le modèle (1) (cité dans le paragraphe 2.3.1), il peut s'écrire dans ce cas :

$$(1 - \phi_1 B)X_t = Z_t \text{ (où } Z_t \text{ est un processus } AR(p-1))$$

$$Z_t = \sum_{i=1}^{p-1} \theta_i Z_{t-i} + \epsilon_t \text{ où } \epsilon_t \rightarrow iid(0; \sigma_\epsilon^2) \quad (2.12)$$

$Z_t$  s'écrit en utilisant le polynôme d'opérateur :  $\theta_{p-1}(B)Z_t = \epsilon_t$ . En multipliant les deux membres du processus AR(1) avec erreur autocorrélé par  $\theta_{p-1}(B)$ , on a :

$\theta_{p-1}(B)(1 - \phi_1 B)X_t = \epsilon_t$  ou encore l'autoregressif d'ordre psuivant :

$$X_t = (\theta_1 + \phi_1)X_{t-1} + \dots + (\theta_{p-1} - \theta_{p-2}\phi_1)X_{t-p+1} - \theta_{p-1}\phi_1 X_{t-p} + \epsilon_t$$

#### Modèle 1 : AR(1) sans intercept avec erreurs autocorrélées

Le processus  $\{X_t\}$  obéissant à un AR(1) à erreurs autocorrélées d'ordre (p-1) est équivalent à un AR(p) à erreurs non autocorrélées, on dit que le processus a été "blanchi". Les tests

de Dickey Fuller simples peuvent donc lui être appliqués.

Cependant l'écriture du modèle  $\Delta X_t$  est plus complexe en raison de la présence des  $\theta_j$ .

Considérons le processus AR(p) :

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \epsilon_t$$

$$\Delta X_t = \left( \sum_{k=1}^p \alpha_k - 1 \right) X_{t-1} - \sum_{j=1}^{p-1} \left( \sum_{k=j+1}^p \alpha_k \right) \Delta X_{t-j} + \epsilon_t \quad (2.13)$$

$$\Delta X_t = \left( \sum_{k=1}^p \alpha_k \right) X_{t-1} - X_{t-1} - \sum_{j=1}^{p-1} \left( \sum_{k=j+1}^p \alpha_k \right) \Delta X_{t-j} + \epsilon_t$$

$$X_t = \psi_1 X_{t-1} + \sum_{j=2}^p \psi_j \Delta X_{t-j} + \epsilon_t$$

Avec  $\psi_1 = \sum_{k=1}^p \alpha_k$  et  $\psi_j = - \sum_{k=j+1}^p \alpha_k$

Sous l'hypothèse nulle de la racine unité  $H_0 : \psi_1 = 1$

on pose  $Y_t = \Delta X_t$

$$\Delta X_t = Y_t = \sum_{j=2}^p \psi_j \Delta X_{t-j} + \epsilon_t$$

$$= \sum_{j=2}^p \psi_j Y_{t-j+1} + \epsilon_t$$

Si on fait la régression de  $Y_t$  sur  $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p+1}$ , on obtient les estimateurs  $\tilde{\psi}_2, \tilde{\psi}_3, \dots, \tilde{\psi}_p$

de  $\psi_2, \psi_3, \dots, \psi_p$ .

Les estimateurs ainsi obtenus sont convergents et suivent asymptotiquement des lois normales, car le processus  $\Delta X_{t-j} = Y_{t-j}$ , tel que  $j = 1, \dots, p-1$  est stationnaire sous l'hypothèse nulle.

Si on fait la régression de  $X_t$  sur  $X_{t-1}, Y_{t-1}, Y_{t-2}, \dots, Y_{t-p+1}$ , on obtient l'estimateur  $\tilde{\psi}_1$  de  $\psi_1$  tel que la loi limite de  $nc(\tilde{\psi}_1 - 1)$  est la même que celle de  $n(\hat{\rho}_{mc} - 1)$  obtenue du modèle (1) (paragraphe **2.3.1**) avec  $c = \frac{1}{1 - \sum_2^p \psi_j}$ , c'est à dire :

$$nc(\tilde{\psi}_1 - 1) \xrightarrow{L} \frac{T^2 - 1}{2G}$$

La statistique t associée à cet estimateur est

$$\tilde{\tau} = \frac{\tilde{\psi}_1 - 1}{\sqrt{\hat{V}(\tilde{\psi}_1)}}$$

Où  $\hat{V}(\tilde{\psi}_1) = \frac{1}{s^2} \sum_{t=2}^n X_{t-1}^2$  et  $s^2 = \frac{1}{n-2} \sum_{t=2}^n (X_t - \tilde{\psi}_1 X_{t-1})^2$

$$\tilde{\tau} \xrightarrow{L} \frac{T^2 - 1}{2\sqrt{G}}$$

T et G sont définies dans le théorème **2.1**

**Remarque :**

La distribution asymptotique de  $\tilde{\tau}$  est la même que celle de la statistique t associée à l'estimateur des moindres carré  $\hat{\tau}$  obtenus dans le cas d'un AR(1) sans intercept.

**Modèle 2 : AR(1) avec tendance linéaire avec erreurs autocorrélées**

Maintenant, considérons le modèle :

$$X_t = \varphi_0 + \varphi_1 t + K_t \tag{2.14}$$

où  $K_t$  est un processus autorégressif d'ordre(p) qui satisfait l'équation :

$$K_t = \psi_1 K_{t-1} + \sum_{j=2}^p \psi_j (K_{t-j+1} - K_{t-j}) + \epsilon_t \tag{2.15}$$

en remplaçant  $K_t$  dans l'équation (2.14) on obtient :

$$X_t = \alpha_0 + \alpha_1 t + \psi_j X_{t-1} + \sum_{j=2}^p \lambda_j \Delta X_{t-j+1} + \epsilon_t \quad (2.16)$$

$$\alpha_0 = \varphi_0(1 - \psi_1) + \psi_1 \varphi_1 - \varphi_1 \sum_{j=2}^p \psi_j \text{ et } \alpha_1 = \varphi_1(1 - \psi_1)$$

$X_t$  est la somme d'un processus stationnaire et d'une tendance (fonction linéaire du temps).

Dans ce cas, la distribution de la statistique  $t$  associée à l'estimateur  $\tilde{\psi}_1$  diffère de celle obtenue dans le cas d'un AR(1) sans intercept avec erreurs autocorréllées.

**Théorème 2.3.** (*Dickey et Fuller, 1981*)

Soit  $\{X_t, t \in \mathbb{Z}\}$  un processus qui satisfait l'équation (2.16), avec  $\epsilon_t$  sont iid de moyenne nulle, et de variance  $\sigma^2$ . Soit

$$\tilde{\tau}_\tau = \frac{\tilde{\psi}_1 - 1}{\sqrt{\hat{V}(\tilde{\psi}_1)}}$$

la statistique  $t$  associée à  $\tilde{\psi}_1$ , alors

$$\tilde{\tau}_\tau \xrightarrow{L} \frac{(T - 2H)(T - 6K) - 1}{2\sqrt{G - H^2 - 3K^2}}$$

où  $K = 2 \int_0^1 W(t)dt - H$ ,

$H, T$  et  $G$  sont définies dans le théorème 2.1, et  $W(t)$  est processus de Wiener standard

### Remarques

- Pour le cas d'un AR(p) avec intercept (la tendance se résume à une constante), on suit la

même procédure, et on obtient le modèle :

$$X_t = \mu + \psi_1 X_{t-1} + \sum_{j=2}^p \psi_j \Delta X_{t-j+1} + \epsilon_t$$

- La distribution asymptotique de la statistique de student associé à l'estimateur des moindres carrés est la même que celle de  $\hat{\tau}_\mu$  donnée dans le théorème (2,2).
- On utilise les même tables que dans le cas d'un AR(1), on les trouve dans Fuller (1996), on les donne en annexe de ce mémoire (table B1 et B2).

### 2.3.3 Le test de Phillips et Perron

Le test de Phillips et Perron (1988)[19] est construit sur correction non paramétrique des statistiques de Dickey-Fuller pour prendre en compte des erreurs hétéroscédastiques et/ou autocorrélées. Les statistiques de test qu'ils proposent sont les transformées de celle de Dickey et Fuller. Nous les appellons statistique de Phillips et Perron. On les note  $Z(\rho_{mc})$ ,  $Z(\tilde{\tau})$  :

Soit le modèle  $X_t = \rho X_{t-1} + \epsilon_t$

Le test de phillips et Perron se déroule en quatre étapes :

- Estimation par les moindres carrés ordinaires des trois modèles de base des tests de Dickey-Fuller et calcul des statistiques associées, soit  $\epsilon_t$  le résidu estimé ;

- Estimation de la variance dite de court terme des résidus  $\hat{\sigma}^2 = \frac{1}{n} \sum_{j=2}^p \epsilon_t^2$

- Estimation d'un facteur correctif  $s_t^2$  (appelé variance de long terme) établi à partir de la structure des covariances des résidus des modèles précédemment estimés de telle sorte que

les transformations réalisées conduisent à des distributions identiques à celles de Dickey-Fuller standard :

$$s_t = \frac{1}{n} \sum_{t=1}^n \epsilon_t^2 + \sum_{i=1}^l \left(1 - \frac{i}{l+1}\right) \frac{1}{n} \sum_{t=i+1}^n \epsilon_t \epsilon_{t-i}$$

Pour estimer cette variance de long terme, il est nécessaire de définir un nombre de retards  $l$  estimé en fonction des  $n$

observations.

- Calcul des statistiques de Philips et Perron définis comme suit :

$$Z(\hat{\rho}_{mc}) = n(\hat{\rho}_{mc} - 1) - \frac{1}{2}(s_t^2 - \hat{\sigma}^2)(n^{-2} \sum_{t=1}^n X_{t-1}^2)^{-1}$$

$$Z(\hat{\tau}) = \frac{\hat{\sigma}}{s_t} \hat{\tau} - \frac{1}{2}(s_t^2 - \hat{\sigma}^2)(s_t^2 n^{-2} \sum_{t=1}^n (X_{t-1})^2)^{-\frac{1}{2}}$$

Phillips et Perron montrent que les distributions asymptotiques de  $Z(\hat{\rho}_{mc})$ , et  $Z(\hat{\tau})$  sont identiques à celles des statistiques  $n(\hat{\rho}_{mc} - 1)$  et  $\hat{\tau}$  respectivement.

### 2.3.4 Test basé sur l'estimateur du maximum de vraisemblance conditionnel

Soit le modèle autorégressif d'ordre 1

$$X_t = \rho X_{t-1} + \epsilon_t \tag{2.17}$$

où  $X_0 = 0$ ,  $\rho$  est le paramètre autorégressif et  $\epsilon_t$  sont iid de loi  $N(0, \sigma^2)$ .

L'estimateur du maximum de vraisemblance (MV) noté  $\hat{\rho}_{mv}$  de  $\rho$  est celui qui maximise la quantité

$$G(X, \rho, \sigma^2) = -2^{-1} \left[ \sigma^{-2} \left\{ \sum_{t=1}^n (X_t - X_{t-1})^2 \right\} - \log(1 - \rho^2) + (n+1) \log \sigma^2 - (n+1) \log 2\pi \right]$$

Dans le cas stationnaire, la distribution associée à l'estimateur  $\hat{\rho}_{mv}$  est la même que celle associée à  $\hat{\rho}_{mc}$  mais sous l'hypothèse de racine unité ceci n'est pas vrai. Gonzalez-Farias (1992) a donné la loi asymptotique de cet estimateur sous  $H_0 : \rho = 1$ .

**Théorème 2.4.** (Gonzalez-Farias, 1992)

Soit le modèle (2.17) et soit  $\hat{\rho}_{mv}$  l'estimateur MV de  $\rho$



Alors

$$n(\hat{\rho}_{mv} - 1) \xrightarrow{L} \frac{(T^2 - 1) - \sqrt{(T^2 - 1)^2 + 8G}}{4G}$$

$H$ ,  $T$  et  $G$  sont définies dans le théorème 2.1

Fuller (1996)[8] donne la distribution asymptotique de l'estimateur MV dans le cas d'un processus autorégressif d'ordre 1 avec intercept et avec tendance linéaire respectivement.

$\hat{\rho}_{m\mu}$  e  $t\hat{\rho}_{m\tau}$ .

**Théorème 2.5.** (Fuller, 1996 )

Soient  $\hat{\rho}_{m\mu}$  et  $\hat{\rho}_{m\tau}$  les estimateurs MV inconditionnels du paramètre autorégressif dans le processus autorégressif d'ordre 1 avec intercept et avec tendance linéaire respectivement.

Alors

$$n(\hat{\rho}_{m\mu} - 1) \xrightarrow{L} 0.5\zeta_{\mu} + 1 - [(\zeta_{\mu} - 1)^2 + 2\eta_{\mu} - 4]^{\frac{1}{2}}$$

$$n(\hat{\rho}_{m\tau} - 1) \xrightarrow{L} 0.5\zeta_{\tau} + 1 - [(\zeta_{\tau} - 1)^2 + 2\eta_{\tau} - 4]^{\frac{1}{2}}$$

Où

$\zeta_{\mu}$  est la distribution limite de  $n(\hat{\rho}_{m\mu} - 1)$  et  $\zeta_{\tau}$  est la distribution limite de  $n(\hat{\rho}_{m\tau} - 1)$

$\eta_{\mu} = (G - H^2)^{-1}[H^2 + (T - H)^2]$  et  $\eta_{\tau} = (G - H^2 - 3K^2)^{-1}[(H - 3K)^2(T - H - 3K)^2]$

$H$ ,  $T$  et  $G$  sont définies dans le théorème 2.1 et le théorème 2.2

### 2.3.5 Test de Kwiatkowski, Phillips, Schmidt et Shin

On a considéré jusqu'ici des tests où l'hypothèse nulle était la racine unitaire c'est-à-dire la non stationnarité. Dans le modèle  $X_t = \rho X_{t-1} + \epsilon_t$  on testait  $H_0 : \rho = 1$ . L'alternative était

$H_1 : |\rho| < 1$ .

Kwiatkowski, (1992) ont proposé une statistique de test particulière pour tester  $H_0 : \sigma_\mu^2 = 0$  contre  $H_1 : \sigma_\mu^2 \neq 0$ . C'est un test unilatéral.

Nous supposons que nous pouvons décomposer la série en somme d'une tendance déterministe, d'une marche aléatoire et d'une erreur stationnaire,

$$X_t = \xi t + r_t + \epsilon_t \tag{2.18}$$

$r_t$  est une marche aléatoire c'est-à-dire  $r_t = r_{t-1} + \mu_t$ . où les  $\mu_t$  sont iid  $(0, \sigma_\mu^2)$ . La valeur initiale  $r_0$  est considérée comme fixe et joue le rôle d'une interception. L'hypothèse de stationnarité est simplement  $\sigma_\mu^2 = 0$ . Puisque  $\epsilon_t$  est supposé stationnaire, sous l'hypothèse nulle.

Nous considérons aussi le cas particulier du modèle (2.18) dans lequel nous fixons  $\xi = 0$ , auquel cas sous l'hypothèse nulle,  $X_t$  est stationnaire autour d'un niveau  $r_0$  plutôt que d'une tendance.

La statistique que nous utilisons est à la fois la statistique du multiplicateur de Lagrange LM unilatérale et la statistique de test LBI( locally best invariant) pour l'hypothèse  $\sigma_\mu^2 = 0$  sous les hypothèses plus fortes que les  $\mu_t$  sont gaussiens et que les  $\epsilon_t$ , sont iid  $N(0, \sigma_\epsilon^2)$ .

[Puisque la valeur du paramètre spécifiée par l'hypothèse nulle est sur la limite de l'espace des paramètres, nous nous intéressons à un test LM unilatéral plutôt qu'à un test bilatéral]

Soit  $\epsilon_t$ ,  $t = 1, 2, \dots, T$  les résidus de la régression  $X$  sur une tendance d'interception et de temps. Soit  $\sigma_\epsilon^2$  la valeur de la variance d'erreur de cette régression (la somme des carrés

résiduels, divisé par T). On définit le processus de somme partielle des résidus comme suit :

$$s_t = \sum_{i=1}^t \epsilon_t \quad t = 1, 2, \dots, T$$

Alors la statistique LM (et LBI) est :

$$LM = \sum_{i=1}^T \frac{s_t^2}{\sigma_\epsilon^2}$$

### 2.3.6 Test basés sur l'estimateur symétrique simple

Soit  $\{X_t, t \in \mathbb{Z}\}$  un processus autorégressif d'ordre p :

$$X_t + \sum_{i=1}^p \alpha_i X_{t-i} = \epsilon_t$$

où les  $\alpha_i$  sont des paramètres réels et  $\{\epsilon_t\}$  un bruit blanc de moyenne nulle, et de variance  $\sigma^2$ .

Avec une autre séquence de bruit blanc  $\{\nu_t\}$ , dans le cas stationnaire, on peut réécrire le modèle sous la forme

$$X_t + \sum_{i=1}^p \alpha_i X_{t+i} = \nu_t$$

Où  $\{\nu_t\}$  est un bruit blanc de moyenne nulle, de variance  $\sigma^2$ .

L'estimateur des moindres carrés ordinaires de  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_p)$  est la valeur de  $\alpha$  qui minimise la somme des carrés des  $\epsilon_t$ . On peut aussi construire un estimateur de  $\alpha$  qui minimise la somme des carrés des  $\nu_t$ . Une autre classe d'estimateurs a été suggérée, ceux qui minimisent la quantité :

$$Q(\alpha) = \sum_{t=p+1}^n \omega_t [X_t + \sum_{i=1}^p \alpha_i X_{t-i}]^2 + \sum_{t=1}^{n-p} (1 - \omega_{t+1}) [X_t + \sum_{i=1}^p \alpha_i X_{t+i}]^2$$

L'estimateur des moindres carrés ordinaire est obtenu en posant  $\omega_t = 1$ .

L'estimateur obtenu en posant  $\omega_t = \frac{1}{2}$  est étudié par Dickey, Hasza et Fuller (1984)[4]. On

l'appelle "estimateur symétrique simple".

### Cas d'un processus autorégressif d'ordre 1

Pour un processus autorégressif d'ordre 1

$$X_t = \rho X_{t-1} + \epsilon_t \tag{2.19}$$

Où  $\rho$  est le paramètre autorégressif et les  $\epsilon_t$  sont i.i.d de loi  $N(0, \sigma^2)$ .

Les estimateurs symétriques sont ceux qui minimisent la quantité :

$$Q(\rho) = \sum_{t=2}^n [X_t - \rho X_{t-1}]^2 + \sum_{t=1}^{n-1} (1 - \omega_{t+1}) [X_t + [X_t - \rho X_{t+1}]]^2$$

L'estimateur des moindres carrés  $\hat{\rho}_{mc}$  est obtenu en posant  $\omega_t = 1$ .

L'estimateur symétrique simple qu'on note  $\hat{\rho}_s$  est obtenu en posant  $\omega_t = \frac{1}{2}$

$$\hat{\rho}_s = \frac{\sum_{t=2}^n X_{t-1} X_t}{\frac{1}{2}(X_1^2 + X_n^2) + \sum_{t=2}^{n-1} X_t^2}$$

La statistique de student associée à l'estimateur symétrique simple est :

$$\hat{\tau}_s = \frac{\hat{\rho}_s - 1}{\hat{\sigma}_s} \sqrt{\frac{1}{2}(X_1^2 + X_n^2) + \sum_{t=2}^{n-1} X_t^2}$$

ou encore

$$\hat{\tau}_s = -\frac{\sqrt{(n-2)(1-\hat{\rho}_s)}}{1+\hat{\rho}_s}$$

où  $\hat{\sigma}_s^2$  est l'estimateur de  $\sigma^2$  associé à  $\hat{\rho}_s$

$$\hat{\sigma}_s^2 = \frac{1}{n-2} \left[ \sum_{t=2}^n (X_t - \hat{\rho}_s X_{t-1})^2 + \frac{1}{2}(1-\hat{\rho}_s^2)(X_1^2 - X_n^2) \right]$$

**Théorème 2.6.** (Fuller,1996)

Soit le processus autorégressif d'ordre 1 (2.19)

Les lois limites de l'estimateur  $\hat{\rho}_s$  et de la statistique  $\hat{\tau}_s$  sous l'hypothèse nulle de la racine unité  $H_0 : \rho = 1$  sont comme suit :

$$\begin{aligned} n(\hat{\rho}_s - 1) &\xrightarrow{L} -\frac{1}{2G} \\ \hat{\tau}_s &\xrightarrow{L} -\frac{1}{2\sqrt{G}} \end{aligned}$$

Où  $G$  est défini dans le théorème (2.1)

Fuller (1996) généralise les résultats ci-dessus aux processus autorégressif d'ordre 1 avec moyenne non nulle et ayant une tendance linéaire. Soit  $\hat{\rho}_{s\mu}$  l'estimateur symétrique simple et  $\hat{\tau}_{s\mu}$  la statistique t correspondant aux cas d'un processus autorégressif d'ordre 1 avec moyenne non nulle. Et  $\hat{\rho}_{s\tau}$  et  $\hat{s}_\tau$  sont l'estimateur symétrique et la statistique t correspondant respectivement au processus autorégressif d'ordre 1 avec tendance linéaire.

**Théorème 2.7.** (Fuller,1996)

Sous l'hypothèse nulle  $H_0 : \rho = 1$  on a :

$$n(\hat{\rho}_{s\mu} - 1) \xrightarrow{L} -\frac{1}{2(G-H^2)} \text{ et } \hat{\tau}_{s\mu} \xrightarrow{L} -\frac{1}{2\sqrt{G-H^2}}$$

$$n(\hat{\rho}_{s\tau} - 1) \xrightarrow{L} -\frac{1}{2(G - H^2 - 3K^2)} \text{ et } \hat{\tau}_{s\tau} \xrightarrow{L} -\frac{1}{2\sqrt{G - H^2 - 3K^2}}$$

Où  $G$ ,  $H$  et  $K$  sont définis dans le théorème (2.1) et le théorème (2.2)

## Chapitre 3

# les technique de bootstrap appliquées aux tests de racine unité

### 3.1 Introduction

Les techniques de bootstrap dans les tests de racine unitaire ont suscité un intérêt considérable dans la littérature. Un aspect important de la conception d'une procédure de bootstrap pour pouvoir effectuer un test, est que la procédure doit pouvoir reproduire le comportement d'échantillonnage de la statistique de test sous l'hypothèse nulle ( racine unitaire), que la série observée respecte ou non l'hypothèse nulle. Bien que cela semble être clair, différentes alternatives bootstrap pour tester l'hypothèse nulle d'intégration racine unitaire ont été proposées dans la littérature.

Une approche souvent suivie consiste à rééchantillonner les résidus obtenus en imposant l'hypothèse nulle. Dans un contexte autorégressif (AR), cela implique d'abord de différencier les séries observées, puis d'appliquer un schéma de bootstrap appliquées aux tests de racine unitaire pour générer les pseudo structures.

Une approche bootstrap différente dans les tests de racine unitaire consiste à rééchantillonner les résidus obtenus sans restriction, c'est-à-dire sans imposer l'hypothèse nulle. Pour les processus autorégressif de niveau supérieur, l'approche prédominante dans la littérature est celle basée sur les différences, alors que les approches basées sur des résidus non restreints ont été appliquées principalement au cas d'un processus autorégressif de premier ordre.

La principale contribution de ce chapitre est d'étudier le comportement des différentes approches bootstrap sous l'hypothèse nulle et l'hypothèse alternative et de comparer analytiquement leurs performances relatives. Une approche alternative à la mise en oeuvre de



la technique de bootstrap dans les tests de racine unité basée sur des résidus non restreints est proposée, et sa supériorité sur les méthodes existantes est établie. En particulier, les méthodes bootstrap basées sur la première différenciation des séries observées (c'est-à-dire sur les résidus restreints), bien que asymptotiquement cohérentes, sont moins puissantes que les méthodes bootstrap basées sur des résidus non restreints. Le chapitre est organisé comme suit. Le paragraphe 2 décrit le modèle et l'hypothèse considérées, il énonce brièvement les deux procédures de bootstrap alternatives pour tester l'hypothèse nulle de racine unité. L'une repose sur la différenciation des séries et l'autre qui est une nouvelle procédure est basée sur des résidus non restreints. Les propriétés asymptotiques des différentes techniques de bootstrap sous l'hypothèse nulle et sous les hypothèse alternative sont étudiées et leur performance de puissance relative est analysée. Un exemple numérique illustre ces résultats théoriques.

### 3.2 Procédure de bootstrap pour les tests de racine unité

Soit le processus autorégressif d'ordre  $p$ ,  $\{X_t, t \in \mathbb{N}\}$  avec

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t \quad (\mathbf{3.1})$$

où  $\phi_p \neq 0$ . Considérons les hypothèses suivantes :

#### Hypothèse (A1)

$\{\epsilon_t\}$  est une suite de variables aléatoires iid de moyenne 0 et tel que  $0 < E[\epsilon_t^2] = \sigma_\epsilon^2 < \infty$

Notre but est de tester l'hypothèse que  $\{X_t\}$  est obtenu en intégrant un processus autorégressif  $(p - 1)$  fois . Une représentation de **(3.1)** utile pour faire un tel test (voir Fuller 1996) est donnée par

$$X_t = \rho X_{t-1} + \sum_{i=1}^{p-1} a_i \Delta X_{t-i} + \epsilon_t \quad (3.2)$$

avec  $\Delta X_{t-i} = X_{t-i} - X_{t-i-1}$ ,  $\rho = \sum_{j=1}^p \phi_j$  et  $a_i = - \sum_{j=i+1}^p \phi_j$ .

### Hypothèse (A2)

Le polynôme  $a(Z) = 1 - a_1 Z - a_2 Z^2 \dots - a_{p-1} Z^{p-1}$  satisfait la condition  $a(Z) \neq 0$  pour  $|Z| \leq 1$ .

La condition précédente est imposée pour s'assurer que si  $\rho = 1$ , alors  $\phi(Z)$  n'a qu'une racine sur le disque de l'unité qui est la racine  $Z = 1$ .

Pour ceci, nous vérifions d'abord que  $\phi(Z) = a(Z)(1 - Z) - (\rho - 1)Z$ , nous concluons que  $\rho = 1$  est équivalent à  $\phi(1) = 0$ . De plus si  $\rho = 1$  alors nous obtenons  $\phi(Z) = a(Z)(1 - Z)$  qui par l'hypothèse (A2) ne s'annule que si  $Z = 1$ .

Soit  $\rho_{\min} = \inf \left\{ \sum_{j=1}^p \phi_j : \phi_1, \phi_2, \dots, \phi_p \text{ tel que } \phi(Z) \neq 0 \text{ pour } |Z| \leq 1 \right\}$

A l'aide de la représentation (3.2), le problème de test considéré peut être posé comme suit

$$H_0 : \rho = 1 \quad \text{contre} \quad H_1 : \rho \in (\rho_{\min}, 1)$$

Le fait que l'ensemble des valeurs de  $\rho$  correspondant à l'hypothèse alternative soit décrit comme dans l'intervalle précédent est une conséquence du fait que  $S = \{(\phi_1, \phi_2, \dots, \phi_p) :$

$\phi(Z) = 1 - \sum_{i=1}^p \phi_i Z^i \neq 0 \text{ pour } |Z| \leq 1\}$  est un ensemble convexe et  $\rho : S \rightarrow \mathbb{R}$ , avec

$\rho = \sum_{i=1}^p \phi_i$  est une fonction continue.

Soit,  $(\hat{\rho}_C, \hat{a}_1, \dots, \hat{a}_{p-1})$  l'estimateur des moindres carrés de  $(\rho, a_1, \dots, a_{p-1})$  de la régression linéaire

$$X_t = \delta + \rho X_{t-1} + \sum_{i=1}^{p-1} a_i \Delta X_{t-i} + \epsilon_t \quad (3.3)$$

de la série observée  $X_1, X_2, \dots, X_n$ .

Si le terme à tendance déterministe  $\delta$  n'est pas inclus dans cette régression, alors l'estimateur des moindres carrés correspondant de  $\rho$  est noté  $\hat{\rho}$ . Une statistique de test basique de l'hypothèse nulle est ensuite donnée par

$$T_n = n(\hat{\rho} - 1) \quad \text{ou bien} \quad T_{n,C} = n(\hat{\rho}_C - 1)$$

Il est bien connu que la distribution asymptotique de  $T_n$  et  $T_{n,C}$  sous l'hypothèse nulle n'est pas standard. Par exemple sous l'hypothèse nulle, nous avons

$$T_n \Rightarrow \left( 1 - \sum_{i=1}^{p-1} a_i \right) \frac{W^2(1) - 1}{2 \int_0^1 W^2(r) dr} \quad (3.4)$$

lorsque  $n \rightarrow +\infty$ , où  $W(\cdot)$  désigne le processus standard de Wiener sur  $[0, 1]$ , et " $\Rightarrow$ " la convergence faible (voir, Fuller 1996). Basé sur cette distribution limite, un test de niveau  $\alpha$  est alors obtenu en rejetant l'hypothèse nulle si  $T_n \leq C_\alpha$  où  $C_\alpha$  désigne le quantile d'ordre  $\alpha$  de la distribution du côté droit de (3.4). Notons que nous pouvons tenir compte d'un terme à tendance déterministe dans (3.1) et (3.2). Dans ce cas, la distribution de l'estimateur des moindres carrés de  $\rho$  sera également affectée et dépendra de la spécification particulière du terme déterministe dans l'équation (3.3) (voir Hamilton 1994 pour une discussion des différents cas). Dans notre étude, le cas avec un terme déterministe pour le modèle (3.1) ne sera pas considéré.

Comme alternative à l'approximation antérieure du grand échantillon, un test peut être construit en utilisant une procédure de bootstrap pour approcher la distribution de  $T_n$  et  $T_{n,C}$  sous l'hypothèse nulle. Pour ce faire, des pseudo-séries de bootstrap sont générées en intégrant des répliques d'un processus AR(p-1) stationnaire. Différentes approches peuvent être suivies pour générer des séries du processus AR(p-1).

### 3.2.1 Le Bootstrap basé sur la série différenciée

Une première approche est basée sur la différenciation des séries observées et l'ajustement d'un processus AR(p-1). Le test associé à l'hypothèse nulle peut alors être décrit par les trois étapes suivantes :

**Etape 1.** Soit  $\hat{\mu}_t$ ,  $t = p + 1, p + 2, \dots, n$  les résidus estimés obtenus en ajustant une autorégression d'ordre (p-1) à  $\Delta X_t$  c'est-à-dire

$$\hat{\mu}_t = \Delta X_t - \sum_{i=1}^{p-1} \hat{b}_i \Delta X_{t-i}$$

où  $\hat{b} = (\hat{b}_1, \hat{b}_2, \dots, \hat{b}_{p-1})'$  est l'estimateur de la régression de  $\Delta X_t$  sur  $\Delta X_{t-1}, \Delta X_{t-2}, \dots, \Delta X_{t-p+1}$ .

**Etape 2.** Générer  $X_1^*, X_2^*, \dots, X_n^*$  tel que  $X_t^* = \sum_{j=1}^t \Delta X_j^*$  où  $\Delta X_t^* = \Delta X_t$  pour  $t = 1, 2, \dots, p-1$  et

$$\Delta X_t^* = \sum_{i=1}^{p-1} \hat{b}_i \Delta X_{t-i}^* + \mu_t^* \quad \text{pour } t = 1, 2, \dots, n.$$

Ici  $\mu_t^*$ ,  $t = p, p + 1, \dots, n$  est une suite iid avec  $\mu_t^* \sim \hat{F}_\mu$  où  $\hat{F}_\mu$  est la fonction de répartition empirique de  $\hat{\mu}_t$  centré.

**Etape 3.** Soit  $\hat{\rho}^*$  le coefficient de  $X_{t-1}^*$  dans la régression linéaire des moindres carrés de  $X_t^*$  sur  $X_{t-1}^*$  et sur  $\Delta X_{t-1}^*, \Delta X_{t-2}^*, \dots, \Delta X_{t-p+1}^*$ .

Notons  $\hat{\rho}_C^*$  l'estimateur des moindres carrés dans la même régression si un terme déterministe est inclu et soit  $T_n^* = n(\hat{\rho}^* - 1)$  et  $T_{n,C}^* = n(\hat{\rho}_C^* - 1)$ . Utilisons la distribution de  $T_n^*$  (respectivement de  $T_{n,C}^*$ ) pour approximer la distribution de  $T_n$  (respectivement de  $T_{n,C}$ ) sous l'hypothèse nulle.

Notons que l'approche de bootstrap précédente est basée sur des résidus restreints c'est-à-dire les résidus sont obtenus de **(3.2)** sous la condition  $\rho = 1$ .

### 3.2.2 Bootstrap basé sur les résidus non restreints

L'approche alternative que nous proposons, consiste à utiliser les paramètres estimés et les résidus basés sur le modèle non restreint, c'est-à-dire sur le modèle **(3.2)** ajusté aux séries observées sans imposer l'hypothèse nulle  $\rho = 1$ . Une telle procédure de bootstrap sur un processus autorégressif basée sur les résidus comprend les trois étapes suivantes :

**Etape 1.** Calculer

$$\hat{\epsilon}_t = X_t - \hat{\rho}X_{t-1} - \sum_{i=1}^{p-1} \hat{a}_i \Delta X_{t-i} \quad \text{pour } t = p, p+1, \dots, n$$

**Etape 2.** Générer  $X_1^+, X_2^+, \dots, X_n^+$  par  $X_t^+ = \sum_{j=1}^t \Delta X_j^+$  où  $\Delta X_t^+ = \Delta X_t$

pour  $t = 1, 2, \dots, p-1$  et

$$\Delta X_t^+ = \sum_{i=1}^{p-1} \hat{a}_i \Delta X_{t-i}^+ + \epsilon_t^* \quad \text{pour } t = p, p+1, \dots, n.$$

ici  $\epsilon_t^*, t = p, p+1, \dots, n$  est une suite iid avec  $\epsilon_t^* \sim \hat{F}_\epsilon$  où  $\hat{F}_\epsilon$  la fonction de répartition empirique des  $\hat{\epsilon}_t$  centrés.

**Etape 3.** Soit  $\hat{\rho}^+$  le coefficient de  $X_{t-1}^+$  dans la régression linéaire des moindres carrés de

$X_t^+$  sur  $X_{t-1}^+$  et sur  $\Delta X_{t-1}^+, \Delta X_{t-2}^+, \dots, \Delta X_{t-p+1}^+$ .

Notons  $\hat{\rho}_C^+$  l'estimateur des moindres carrés dans la même régression si un terme déterministe est inclu et soit  $T_n^+ = n(\hat{\rho}^+ - 1)$  et  $T_{n,C}^+ = n(\hat{\rho}_C^+ - 1)$ .

Utilisons la distribution de  $T_n^+$  ( respectivement de  $T_{n,C}^+$ ) pour approximer la distribution de  $T_n$  ( respectivement de  $T_{n,C}$ ) sous l'hypothèse nulle.

Notons que la proposition bootstrap précédente diffère pour  $p \geq 1$  de la procédure basée sur les résidus en deux étapes proposée par Feretti et Romo (1996). Selon la procédure de Feretti et Romo, les résidus  $\hat{U}_t$  de la régression par les moindres carrés  $X_t = \beta X_{t-1} + U_t$  sont calculés dans une première étape, tandis que dans un second temps un modèle AR (p-1) est ajusté à  $U_t$ . Feretti et Romo (1996)[7] ont étudié les propriétés de leur proposition seulement en supposant que l'hypothèse nulle est correcte (voir, par exemple, l'hypothèse  $\beta = 1$  dans Théorème.1 et 2). Les propriétés asymptotiques de leur procédure sous l'hypothèse alternative ne sont pas connues. Cependant, on voit facilement que sous l'hypothèse nulle,  $\hat{\beta} = 1 + O_p n^{-1}$ , alors que sous l'alternative,  $\hat{\beta} = r_1 + O_p n^{-\frac{1}{2}}$ , où  $r_1$  est l'autocorrélation d'ordre 1 de X.

Ainsi, si l'hypothèse alternative est vraie (et  $\rho < 1$ ), alors  $\hat{\beta}$  n'est pas un estimateur de  $\rho$  apparaissant en (3.2), alors que c'est le cas avec l'estimateur utilisé dans notre procédure.

### 3.2.3 Comparaisons

Le théorème suivant résume le comportement asymptotique des différentes statistiques de bootstrap que nous considérons.

**Théorème 3.1.** *(Paparoditis et Politis (2005)) Supposons que  $\{X_t\}$  suive (3.1) et que les*

hypothèse (A1) et (A2) soient vraies. Soit

$$C_D = \begin{cases} 1 - \sum_{i=1}^{p-1} a_i & \text{si } \rho = 1 \\ 1 - \sum_{i=1}^{p-1} b_i & \text{si } \rho_{\min} < \rho < 1 \end{cases}$$

et  $C_R = 1 - \sum_{i=1}^{p-1} a_i$ . Ici  $b_1, b_2, \dots, b_{p-1}$  sont les coefficients de  $\Delta X_{t-1}, \Delta X_{t-2}, \dots, \Delta X_{t-p+1}$ .

dans la projection (orthogonale) de  $\Delta X_t$  sur l'étendue fermée  $\bar{s}p\{\Delta X_{t-1}, \Delta X_{t-2}, \dots, \Delta X_{t-p+1}\}$ .

1) Si  $\delta = 0$  dans (3.3) alors, quand  $n \rightarrow \infty$ ,

$$\mathbf{a)} \mathcal{L}(T_n^*/X_1, X_2, \dots, X_n) \Rightarrow \frac{C_D(W^2(1)-1)}{2 \int_0^1 W^2(r)dr} \text{ en probabilité}$$

$$\mathbf{b)} \mathcal{L}(T_n^+/X_1, X_2, \dots, X_n) \Rightarrow \frac{C_R(W^2(1)-1)}{2 \int_0^1 W^2(r)dr} \text{ en probabilité}$$

Où  $\mathcal{L}(\cdot)$  désigne la loi d'une statistique.

2) Si le terme déterministe  $\delta$  est inclu dans (3.3) alors  $\mathcal{L}(T_{n,C}^*/X_1, X_2, \dots, X_n) \Rightarrow Y C_D$

et  $\mathcal{L}(T_{n,C}^+/X_1, X_2, \dots, X_n) \Rightarrow Y C_R$  en probabilité où

$$Y = \frac{2^{-1}(W^2(1) - 1) - W(1) \int_0^1 W(r)dr}{\int_0^1 W^2(r)dr - [\int_0^1 W(r)dr]^2}$$

Comme le montre le théorème précédent, une différence dans le comportement limite des deux propositions bootstrap apparaît seulement lorsque l'hypothèse alternative est vraie, c'est-à-dire si  $X$  est un processus stationnaire. Dans ce cas, les deux statistiques de bootstrap convergent pour  $p > 1$  vers des limites différentes. Notons que même pour l'approche bootstrap couramment utilisée basée sur les différences, le résultat précédent est nouveau, car le comportement asymptotique de  $T_n^*$  et  $T_{n,c}^*$  n'a été étudié dans la littérature que sous la condition que l'hypothèse nulle soit correcte, c'est-à-dire, sous l'hypothèse  $\rho = 1$  (voir,

par exemple, Park 2003 ). De plus, pour  $p = 1$ , nous avons  $C_D = C_R = 1$  et par conséquent, les distributions limites dans les parties (a) et (b) du théorème **3.1** sont identiques. Dans ce cas, les deux tests basés sur le bootstrap sont asymptotiquement équivalents. Ceci explique les résultats de Swensen (2003)[22] concernant le cas autorégressif de premier ordre simple, où aucune différence n'a été trouvée entre le comportement de la puissance du bootstrap basé sur la différence et celui basé sur le résidu.

Pour éviter la répétition des arguments, nous étudions les fonctions de puissance des deux tests bootstrap uniquement dans le cas où un terme déterministe n'est pas inclue dans **(3.3)**. Pour calculer la fonction de puissance pour des alternatives fixées, notons que si  $\{X_t\}$  est un processus AR (p) stationnaire, alors  $\sqrt{n}(\hat{\rho} - \rho) \implies N(0, \sigma_\rho^2)$  où  $0 < \sigma_\rho^2 = \lim_{n \rightarrow \infty} \text{Var}(\sqrt{n}\hat{\rho})$  (voir lemme A2 ). Maintenant, soit  $\beta_{D,n}(\rho) = P(T_n \leq C_\alpha^*/\rho \in ]\rho_{\min}, 1])$  et  $\beta_{R,n}(\rho) = P(T_n \leq C_\alpha^+/\rho \in ]\rho_{\min}, 1])$  indique les fonctions de puissance des tests bootstrap basés sur les valeurs critiques  $C_\alpha^*$  et  $C_\alpha^+$ . Et  $C_\alpha^*$  (respectivement  $C_\alpha^+$ ) désigne le quantile d'ordre  $\alpha$  de la distribution de  $T_n^*$  (respectivement de  $T_n^+$ )

Nous avons ensuite

$$\beta_{D,n}(\rho) = F_n((C_\alpha^*/\sqrt{n} - \sqrt{n}(\rho - 1))/\sigma_\rho)$$

et

$$\beta_{R,n}(\rho) = F_n((C_\alpha^+/\sqrt{n} - \sqrt{n}(\rho - 1))/\sigma_\rho)$$

où  $F_n$  désigne la fonction de répartition de  $\sqrt{n}(\hat{\rho} - \rho)/\sigma_\rho$ .

Notons que le lemme A2 et le théorème de Pólya impliquent que  $|F_n(x) - \phi(x)| \rightarrow 0$  quand  $n \rightarrow \infty$  où  $\phi(\cdot)$  désigne la fonction de répartition de la loi normale standard. De plus



par la convergence uniforme aux limites données dans le théorème 1 et la stricte monotonie des fonctions de répartition limites, on obtient  $C_\alpha^* \rightarrow C_{\alpha,\infty}^*$  et  $C_\alpha^+ \rightarrow C_{\alpha,\infty}^+$  en probabilité, où  $C_{\alpha,\infty}^*$  et  $C_{\alpha,\infty}^+$  désigne le quantile d'ordre  $\alpha$  des distributions limites données dans le théorème 1. Notons que  $C_{\alpha,\infty}^+ = C_\alpha$  pour tout  $\rho \in ]\rho_{\min}, 1]$ , alors que  $C_{\alpha,\infty}^* = C_\alpha$  seulement si  $\rho = 1$ , où  $C_\alpha$  est le quantile d'ordre  $\alpha$  de la distribution limite donnée en (3.4). En conséquence de l'expression précédente pour  $\beta_{D,n}(\cdot)$  et  $\beta_{R,n}(\cdot)$ , de l'uniformité de la convergence aux distributions limites, et de la convergence de  $C_\alpha^*$  et  $C_\alpha^+$  susmentionnée, le théorème suivant est facilement établi.

**Théorème 3.2.** (*Paparoditis et Politis (2005)*) *Supposons que  $\{X_t\}$  suive (3.2) et que les hypothèses (A1) et (A2) soient satisfaites. Alors*

(a)  $p - \lim_{n \rightarrow \infty} \beta_{D,n}(\rho) = p - \lim_{n \rightarrow \infty} \beta_{R,n}(\rho) = 1$  pour tout  $\rho \in ]\rho_{\min}, 1]$ , où  $p - \lim$  signifie la limite en probabilité

et

(b) Avec une probabilité tendant vers 1,  $\beta_{D,n}(\rho_1) \geq \beta_{D,n}(\rho_2)$  et  $\beta_{R,n}(\rho_1) \geq \beta_{R,n}(\rho_2)$

pour  $\rho_{\min} < \rho_1 < \rho_2 < 1$ .

Malgré la convergence des deux approches, une question intéressante est de savoir comment les différences dans le comportement limite des deux procédures bootstrap données dans le théorème 1 affectent leur performance relative sous l'hypothèse l'alternative. Pour faire face à cela, l'hypothèse de régularité suivante est imposée.

### Hypothèse (A3)

Si  $\rho \in (\rho_{\min}, 1)$ , alors la distribution de  $\sqrt{n}(\hat{\rho} - \rho)$  est absolument continue par rapport à la mesure de Lebesgue sur  $\mathbb{R}$  et a une densité bornée.

Le théorème suivant montre alors que la procédure de bootstrap sur un autorégressif basée sur des résidus non restreints est plus puissante.

**Théorème 3.3.** (*Paparoditis et Politis (2005)*) *Supposons que  $\{X_t\}$  satisfait (3.2) avec  $\rho \in ]\rho_{\min}, 1[$  et que  $\alpha \in ]0, 1[$  tel que  $\tilde{C}_\alpha < 0$ , où  $\tilde{C}_\alpha$  est le quantile d'ordre  $\alpha$  de la distribution de  $(W^2(1) - 1)/2 \int_0^1 W^2(r)dr$ .*

*En outre, considérons que les hypothèses (A1) - (A3) soient satisfaites. Puis, avec une probabilité qui tend vers 1*

$$\beta_{R,n}(\rho) \geq \beta_{D,n}(\rho)$$

*En particulier, nous avons*

$$\beta_{R,n}(\rho) = \beta_{D,n}(\rho)[1 + d_n + o_p(n^{-\frac{1}{2}})],$$

où

$$d_n = \frac{1}{\sqrt{n}} \tilde{C}_\alpha \sum_{i=1}^{p-1} (b_i - a_i) K_n(\rho) \tag{3.5}$$

et  $K_n(\rho)$  est une quantité non négative dépendant de  $F_n$  et sa dérivée. La suite  $d_n$  est non négative avec une probabilité 1 (i.e,  $d_n \geq 0$ ), et elle satisfait  $d_n \rightarrow 0$  en probabilité lorsque  $n \rightarrow \infty$ .

En outre, pour  $p > 1$ , nous avons  $d_n = 0$  si et seulement si  $K_n(\rho) = 0$ .

Notons que  $\tilde{C}_\alpha < 0$  est satisfait pour toutes les valeurs couramment utilisées pour  $\alpha \in (0, 1)$ . la quantité  $K_n(\rho)$  apparaissant en (3.5) est donnée par  $K_n(\rho) = F'_n(\xi_1)/F_n(\xi_2)$

où  $\xi_2 = [n^{-\frac{1}{2}}C_\alpha^* - n^{\frac{1}{2}}(\rho - 1)]/\sigma_\rho$  et  $\xi_1$  est une valeur entre  $\xi_2$  et  $[n^{-\frac{1}{2}}C_\alpha^+ - n^{\frac{1}{2}}(\rho - 1)]/\sigma_\rho$  (voir la preuve du Théorème **(3.3)** dans l'article). Ainsi avec une probabilité égale à 1,  $K_n(\rho) = 0$  si et seulement si  $F'_n(\xi_1) = 0$ . Maintenant, si la densité de  $\sqrt{n}(\hat{\rho} - \rho)$  est positive en  $\xi_1$ , alors  $d_n > 0$ ; c'est-à-dire que le test basé sur les résidus est strictement plus puissant que le test basé sur la différence.

Le théorème **(3.2)** montre que les deux tests sont convergents dans la mesure où leur puissance approche l'unité lorsque la taille de l'échantillon et / ou l'écart par rapport à la valeur de l'hypothèse nulle augmente. Cependant, le théorème **(3.3)** montre qu'avec une probabilité tendant vers 1, la puissance du test bootstrap basé sur des résidus non restreints est bornée inférieurement et uniformément en  $\rho$  par la puissance du test basé sur les différences. Notons que par le théorème **(3.3)**,  $d_n \rightarrow 0$  en probabilité quand  $n \rightarrow \infty$ . En outre,  $\beta_{D,n}(\rho)$  et  $\beta_{R,n}(\rho)$  augmentent lorsque  $\rho \rightarrow \rho_{\min}$ . En raison de cela et de la limite des fonctions de puissance, nous nous attendons à ce que les différences entre les deux méthodes soient plus apparentes quand la taille de l'échantillon et / ou l'écart par rapport à la valeur de l'hypothèse nulle augmente

Nous concluons cette section en considérant le comportement asymptotique des deux méthodes bootstrap dans le traitement de la statistique de Student  $t_n = (\hat{\rho} - 1)/\hat{\sigma}_\rho$  où  $\hat{\sigma}_\rho = \sqrt{\hat{\sigma}_\rho^2}$  et  $\hat{\sigma}_\rho^2 = \hat{\sigma}_\epsilon^2(\hat{\gamma}_0 - \hat{\mathbf{d}}_p' \hat{\mathbf{C}}_p^{-1} \hat{\mathbf{d}}_p)$  est l'estimateur de la variance de  $\hat{\rho}$  voir la preuve du lemme A.2 pour la définition de  $\mathbf{d}_p$  et  $\mathbf{C}_p$ . Ici, un "hat" sur  $\mathbf{d}_p$  et  $\mathbf{C}_p$  indique que les quantités théoriques correspondantes sont remplacées par leurs estimations .

Soit  $t_n^* = (\hat{\rho}^* - 1)/\hat{\sigma}_\rho^*$  et  $t_n^+ = (\hat{\rho}^+ - 1)/\hat{\sigma}_\rho^+$  les statistiques de bootstrap analogue à  $t_n$

où  $\hat{\sigma}_\rho^*$  (respectivement,  $\hat{\sigma}_\rho^+$ ) est l'estimateur correspondant de l'écart-type de  $\hat{\rho}^*$  (respecti-

vement de  $\hat{\rho}^+$ ). Le résultat suivant est une conséquence du théorème (3.1), du théorème du Slutsky, et le fait que les estimateurs bootstrap  $\hat{\sigma}_\rho^+$  et  $\hat{\sigma}_\rho^*$  convergent en probabilité vers des limites appropriées.

**Corollaire 3.2.1.** *Sous les hypothèses du théorème 3.1 et quand  $n \rightarrow \infty$ , les statistiques bootstrap de Student  $t_n^*$  et  $t_n^+$  convergent faiblement vers la même limite donnée par  $(W^2(1) - 1)/(2 \int_0^1 W^2(r) dr)$ .*

*Ainsi, pour la statistique de Student  $t_n$ , les deux méthodes bootstrap sont asymptotiquement équivalentes. Un résultat similaire peut être établi si l'on considère à la place de  $t_n$  la statistique étudiée  $t_{n,C} = (\hat{\rho}_C - 1)/\hat{\sigma}_{\rho,C}$ , où le "C" indique les quantités obtenues si un terme à tendance déterministe est inclu dans la régression ajustée aux séries observées. Cependant, dans cette section, nous nous sommes concentré sur le bootstrap des statistiques de test de base  $T_n$  et  $T_{n,C}$  au lieu de  $t_n$  et  $t_{n,C}$ , parce que cela a été proposé par plusieurs des auteurs et permet de mieux comprendre les méthodes alternatives de bootstrap. De plus, il n'est pas clair qu'utiliser les statistiques de Student dans le contexte du test de racine unitaire rende la statistique  $t_n$  ou  $t_{n,C}$ , plus puissante que  $T_n$  ou  $T_{n,C}$  (voir aussi Li et Maddala 1996). Ceci est également confirmé par les exemples numériques présentés dans la suite, où les tests bootstrap basés sur des résidus non restreints et les statistiques de test  $T_n$  ou  $T_{n,C}$  semblent avoir plus de puissance que les tests basés sur des valeurs critiques asymptotiques et les versions de Student  $t_n$  ou  $t_{n,C}$ .*

### Exemple 3.1

Considérons un modèle autorégressif AR d'ordre 5 :

$X_t = a(B) \times (1 - B)X_t - (\rho - 1)X_{t-1}$ , où B est un opérateur retard  $B^k X_t = X_{t-k}$  pour  $k \in \mathbb{Z}$ ,  $a(z) = (1 - 0.87z)(1 + 0.87z)(1 - z + 0.41z^2)$ . Les auteurs ont généré 2000 répliques de longueur  $n=50$  et  $n=100$  de ce modèle et ils appliquent 3 tests différents pour l'hypothèse de la racine unité pour différentes valeurs de  $\rho$  proche de la valeur  $\rho = 1$ . les tests appliqués sont les tests  $T_n$  et  $T_{n,C}$  basé sur les valeurs critiques obtenues en utilisant les deux méthodes différentes de Bootstrap et le t-test basé sur les statistiques  $t_n$  et  $t_{n,C}$ . Les résultats obtenus par les auteurs sont basés sur 1000 répliques bootstrap et sont reportés dans la table 1.

**Tableau 3.1** Le seuil des tests de racine unité pour le modèle autorégressif d'ordre 5 de l'exemple 3.1 sous les différentes valeurs du paramètre  $\rho$

			$\rho$					
n	$\alpha$		1.00	0.98	0.96	0.94	0.92	0.90
50	0.05	$T_n^+$	0.072	0.408	0.675	0.819	0.905	0.944
		$T_n^*$	0.065	0.257	0.488	0.662	0.783	0.854
		$t_n$	0.073	0.395	0.618	0.754	0.841	0.884
		$T_{n,C}^+$	0.065	0.196	0.406	0.616	0.751	0.831
		$T_{n,C}^*$	0.012	0.060	0.138	0.251	0.361	0.454
		$t_{n,C}$	0.104	0.174	0.294	0.401	0.498	0.565
100	0.01	$T_n^+$	0.013	0.395	0.807	0.942	0.982	0.992
		$T_n^*$	0.008	0.231	0.606	0.832	0.932	0.970
		$t_n$	0.011	0.390	0.747	0.910	0.962	0.983
		$T_{n,C}^+$	0.009	0.187	0.586	0.832	0.931	0.973
		$T_{n,C}^*$	0.002	0.044	0.239	0.495	0.678	0.798
		$t_{n,C}$	0.013	0.114	0.367	0.585	0.720	0.821
	0.05	$T_n^+$	0.056	0.776	0.968	0.996	0.999	1.000
		$T_n^*$	0.045	0.675	0.938	0.990	0.996	0.999
		$t_n$	0.055	0.740	0.953	0.991	0.997	1.000
		$T_{n,C}^+$	0.046	0.498	0.866	0.965	0.991	0.997
		$T_{n,C}^*$	0.020	0.294	0.681	0.885	0.952	0.983
		$t_{n,C}$	0.059	0.371	0.709	0.871	0.939	0.968

NOTE : Les statistiques de test utilisées ici sont les tests  $T_n$  et  $T_{n,C}$  basés sur les valeurs critiques obtenues en utilisant les statistiques bootstrap basées sur les résidus  $T_n^+$  et  $T_{n,C}^+$  et les statistiques bootstrap basées sur la différence  $T_n^*$  et  $T_{n,C}^*$ .  $t_n$  et  $t_{n,C}$  désignent le test de Dickey Fuller basé sur des valeurs critiques asymptotiques.

**Remarques :**

Ces résultats confirment les résultats théoriques cités dans la table 1. Les tests basés sur les statistiques  $T_n$  et  $T_{n,C}$  et les valeurs critiques obtenues en utilisant les résidus bootstrappés sont clairement plus puissants que les tests correspondants en utilisant les valeurs

critiques obtenues par la procédure bootstrap basées sur les différences. En outre, l'approche bootstrap basée sur les résidus est plus puissante que le t-test basé sur les valeurs critiques asymptotique.

**Un exemple de données réelles.**(Paparoditis et Politis (2005))

Paparoditis et Politis (2005) ont considéré des séries trimestrielles, corrigées des variations saisonnières, des taux d'intérêt du papier commercial américain de 1953 à 1970 (voir tableau A.6 dans Reinsel 1993). Ils prennent un l'échantillon de taille  $n = 72$ , et il teste la présence d'une racine unité dans un processus autorégressif en utilisant la procédure de bootstrap basée sur le résidu. Cette série peut être modélisée de manière appropriée en utilisant un processus autoregressif d'ordre 4. L'ajustement de l'équation **(3.3)** avec terme déterministe donne les estimateurs des moindres carrés suivant :  $\rho_C = 0.973$ ,  $\hat{a}_1 = 0.793$ ,  $\hat{a}_2 = -0.457$ ,  $\hat{a}_3 = 0.121$ , et  $\hat{\delta} = 0.149$ . Ainsi, la statistique de test a la valeur  $T_{n,C} = -1.908$  et la valeur critique du bootstrap estimée basée sur  $B = 1000$  réplifications bootstrap et  $\alpha = 0.05$  est donnée par  $C_{0,05}^+ = -8,978$ .

Sur la base de ces résultats, l'hypothèse nulle d'une racine d'unité AR dans la série des taux d'intérêt du papier commercial n'est clairement pas rejetée.

### 3.3 Application sur un modèle AR(1) sans intercept et AR(1) et tendance linéaire

#### 3.3.1 Description du modèle

Le modèle utilisé est un autorégressif d'ordre un sans terme constant et sans tendance :

$$X_t = \rho * X_{t-1} + \epsilon_t \quad (\text{pour } t = 1, 2, \dots, T)(A)$$

où les  $\epsilon_t$  sont iid de moyenne 0 et de variance  $\sigma^2$

Le test considéré :  $H_0 : \rho = 1$  vs  $H_1 : \rho < 1$

Les statistique utiliser sont les statistique de Dickey Fuller simple  $Z$  et la statistique

symétrique  $\tau_s$  qui sont définit comme suit :  $Z = T(\rho_{mc} - 1)$  et  $\hat{\tau}_s = -\frac{\sqrt{(n-2)(1-\hat{\rho}_s)}}{1+\hat{\rho}_s}$  avec

$\hat{\rho}_{mc}$  est l'estimateur des moindres carrés ordinaire et  $\hat{\rho}_s$  l'estimateur symétrique

où

$$\hat{\rho}_s = \frac{\sum_{t=2}^n X_{t-1}X_t}{\frac{1}{2}(X_1^2 + X_n^2) + \sum_{t=2}^{n-1} X_t^2} \quad \text{et} \quad \hat{\rho}_{mc} = \frac{\sum_{t=1}^n X_t X_{t-1}}{\sum_{t=1}^n X_{t-1}^2}$$

#### 3.3.2 Algorithme : calcule les seuil des test de racine unité pour AR(1)

- Générer N échantillons  $(X_1, X_2, \dots, X_n)$  tel que

$$X_t = \rho X_{t-1} + \epsilon_t \quad (A)$$

(où  $\epsilon_t$  est un bruit blanc de moyenne nulle et de variance  $\sigma^2$ ).

- Calculer l'estimateur des moindres carrés  $\hat{\rho}_{mc}$  de  $\rho$  et la statistique de test de Dickey Fuller ( $DF$ ) associé a chaque série.



- Remplacer  $\rho_{mc}$  par  $\rho$  dans l'équation (A) ensuite calculer les résidus estimer  $\tilde{\epsilon}_t$  tel que.

$$\tilde{\epsilon}_t = X_t - \hat{\rho}_{mc}X_{t-1}$$

- Centrer les résidus estimer ensuite les ordonner tel que  $(\bar{\epsilon}_1, \bar{\epsilon}_2, \dots, \bar{\epsilon}_B)$  .

- Simuler B échantillons bootsrap  $(X_1^*, X_2^*, \dots, X_n^*)$  a partir des résidus centrer tel que

$$\bar{\epsilon}_t = X_t^* - X_{t-1}^*$$

- Calculer l'estimateur des moindres carrés  $\rho_{star}$  et la statistique Dekey Fuller associé

$DF_{star}$  pour chaque série bootstrappé

- Créer un compteur  $C_{df}$  qui calcule ( le nombre de fois où la statistique  $DF_{star} < DF$  ) /

B .

- Calculer le nombre des fois où la statistique  $DF < c_1$  ) / N

- Calculer (le nombre des fois où  $C_{df} > \alpha$  ) /N

### 3.3.3 Tables :

**table1. seuils des statistiques SYM et DF**

statistique de test	25	50	100	250
DF	0.045	0.045	0.058	0.046
DFstar	0.043	0.044	0.062	0.045
SYM	0.053	0.047	0.063	0.047
SYMstar	0.045	0.045	0.064	0.046

Les valeurs critiques pour la statistique  $DF$  sont

pour  $n = 25$  la valeur critique utilisée est  $c_1 = -7.3$ .

pour  $n = 50$  la valeur critique utilisée est  $c_1 = -7.7$ .

pour  $n = 100$  la valeur critique utilisée est  $c_1 = -7.9$ .

pour  $n = 250$  la valeur critique utilisée est  $c_1 = -8.0$ .

Les valeurs critiques pour la statistique  $SYM$  sont

pour  $n = 25$  la valeur critique utilisée est  $c_2 = -2.05$ .

pour  $n = 50$  la valeur critique utilisée est  $c_2 = -2.08$ .

pour  $n = 100$  la valeur critique utilisée est  $c_2 = -2.09$ .

pour  $n = 250$  la valeur critique utilisée est  $c_2 = -2.10$ .

**Remarque :** pour l'estimateur symétrique on suit le même algorithme pour calculer le seuil

### 3.4 Conclusion générale

Dans le premier chapitre, nous nous sommes limité au problème de l'estimation du biais, de l'erreur-standard d'un paramètre, et à la détermination des intervalles de confiance d'un paramètre et calcul de la p-value. Il ne s'agit cependant pas des seules applications des méthodes de rééchantillonnage. Celles-ci peuvent, en effet, aussi être utilisées pour la réalisation de différents tests d'hypothèses, pour le choix des variables et l'estimation de l'erreur de prédiction en régression. L'utilisation des techniques de bootstrap permet d'avoir une meilleure copie de l'information contenue dans l'échantillon initial pour les intervalles de prédiction.

Bien qu'elles puissent être utilisées dans des situations très variées, leur mise en oeuvre ne présente guère d'intérêt lorsque l'inférence statistique peut être réalisée par des méthodes analytiques classiques, pour lesquelles les conditions d'application sont remplies. Elles ne sont donc pas destinées à remplacer les méthodes d'inférence statistique classiques lorsque celles-ci sont applicables mais plutôt à fournir des réponses à des questions pour lesquelles les méthodes classiques sont inapplicables ou non disponibles.

Le second chapitre a permis d'exposer différents tests de la racine unité, L'étude critique des différentes procédures de test de la racine unitaire et des différentes stratégies nous a conduit à la réflexion suivante quant à l'application de ces procédures de test dans les travaux empiriques : il faut souligner d'abord l'importance cruciale que revêt la spécification de la composante déterministe dans les procédures de test de la racine unitaire. Les résultats des tests dépendent en effet de cette spécification. Le principe de base à retenir est le

suivant : la stratégie de test doit débiter par le modèle le plus général possible concernant la spécification de la composante déterministe. Par ailleurs, étant donné l'extrême sensibilité de ces procédures à la structure d'autocorrélation des erreurs. Un test racine unité peut être un instrument de modélisation utile pour spécifier une relation de cointégration.

Dans le dernier chapitre nous nous sommes basé sur l'article de Paparoditis et Politis (2005) qui fait l'étude de deux approches principales pour les tests de racine unitaire dans les processus autorégressif d'ordre  $p$  : l'approche traditionnelle basée sur la différenciation des séries observées (sur des résidus restreints obtenus en imposant l'hypothèse nulle de l'intégration racine unitaire) et l'approche basée sur les résidus non restreints. Les auteurs de cet article affirment que : pour les processus autorégressif d'ordre fini, les deux procédures convergent vers la même limite asymptotique si l'hypothèse nulle est vraie, alors que les deux limites sont différentes si l'hypothèse nulle n'est pas vraie. La puissance des deux tests basés sur l'approche bootstrap tend vers un quand la taille de l'échantillon et / ou l'écart par rapport à la valeur de l'hypothèse nulle augmente. Cependant, le test basé sur les résidus non restreints et la statistique de Student est plus puissant que le test basé sur les différences.

# Bibliographie

- [1] Bourbonnais, R. et Terraza, M. 2016 "*Analyse Des Séries Temporelles*". Dunod, Paris.
- [2] Dickey, D. A., and Fuller, W. A. (1979), "*Distribution Of The Estimators For Autoregressive Time Series With A Unit Root*", J. Amer. Statist. Assoc. 74, 427-431.
- [3] Dickey, D. A., and Fuller, W. A. (1981), "*Likelihood Ratio Statistics For Autoregressive Time Series With A Unit Root*", *Econometrica* 49, 1057-1072.
- [4] Dickey, D. A., Hasza, D.P and Fuller, W. A. (1984), "*Testing for unit roots in seasonal time series* ", J. Amer. Statist. Assoc. 79, 355-367.
- [5] Efron, B. (1979), "*Bootstrap methods : another look at the jackknife*", *Annals of Statistics* 7, 1-26
- [6] Efron, B. and Tibshirani R.J. (1993), "*An Introduction to the Bootstrap*". Chapman and Hall, New York
- [7] Ferretti, N., and Romo, J. (1996), "*Unit Root Bootstrap Tests for AR(1) Models*", *Biometrika* , 83, 849-860.
- [8] Fuller, W. A. (1996), "*Introduction to Statistical Time Series*". Wiley, New York

- [9] Gallardo, L.(2008), “*Cours de statistique Master 1, Chapitre 4* ”. Parc de Grandmont, 37200 Tours, France.
- [10] Gonzalez-Farias, G. M. (1992), *A new unit root test for autorepsive time series*. Ph.D, dissertation, North Carolina State University, Raleigh, North Carolina.
- [11] Gould.J.P,and Nelson.C.R *The Stochastic Structure of the Velocity of Money*. American Economic Review,64,405-417
- [12] Hamilton, J. D. (1994), *Time Series Analysis, Princeton, NJ : Princeton University Press*.
- [13] Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. et Shin, Y. (1992), “*Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root* ”, Journal of Econometrics, 54, 91-115.
- [14] Lecouttre, J.P.Tassi,P.(1987) , *Statistique Non Parametrique Et Robustesse*. Economica, Paris
- [15] Lejeune, M.(2010) , *Statistique La théorie Et Ses Applications*. Springer-Verlag France, Paris.
- [16] Li, H., and Maddala, G. S. (1996), “*Bootstrapping Time Series Models* ” , Econometric Reviews, 15, 115-158.
- [17] Maddala, G.S and Kim, I.M S. (1998), “*Unit roots cointegration and stuctural change*” ,Cambridge, New York.
- [18] Paparoditis. E and Politis, D. N. , J. D. (2005), “*Bootstrapping Unit Root Tests for Autoregressive Time Series*”. Journal of the American Statistical Association 100,

545-553.

- [19] Park, Y. J. (2003), "*Bootstrap Unit Root Tests*," *Econometrica*, 71, 1845-1895.
- [20] Phillips, P.C.B., and P. Perron (1988), "*Testing for a unit root in time series regression* ", *Biomètrika* 75, 335-346.
- [21] Reinsel, C. G. (1993) , "*Elements of Multivariate Time Series Analysis*," , New York : Springer-Verlag.
- [22] Saporta,G. (2006) , *Probabilités Analyse Des données Et Statistique*.Technip, Paris
- [23] Swensen, A. R. (2003), "*A Note on the Power of Bootstrap Unit Root Tests* ", *Econometric Theory*, 19, 32-48.

## Annexe

**Lemme A.1.** Soit  $X$  satisfait l'équation **(3.2)** avec  $\rho \in (\rho_{min}, 1)$ . alors

$$\sum_{i=1}^{p-1} (b_i - a_i) < 0$$

**Lemme A.2.** Supposons que  $X$  est un processus autorégressif satisfaisant l'équation **(3.2)** avec  $\rho \in (\rho_{min}, 1)$ . Ensuite, quand  $n \rightarrow \infty$ ,

$$\sqrt{n}(\hat{\rho} - \rho) \Rightarrow N(0, \sigma_\rho^2).$$

où  $0 < \sigma_\rho^2 = \sigma_\epsilon^2(\gamma_0 - \mathbf{d}'_{p-1} \mathbf{C}_{p-1}^{-1} \mathbf{d}_{p-1})$  et  $\mathbf{d}'_{p-1} = ((\gamma_{i-1} - \gamma_i), i = 1, 2, \dots, p-1)$ .

**Théorème 3.4.** (*Theorème de Slutsky*) Si  $(X_n)$  et  $(Y_n)$  sont deux suites de variables aléatoires convergeant respectivement en loi vers  $X$  et  $Y$  où  $Y$  est presque surement constante, alors la suite  $((X_n, Y_n))$  converge en loi vers  $(X, Y)$ .



## **Resumé**

Ce mémoire présente une étude théorique sur les différentes méthodes du Bootstrap et l'utilisation de cette technique de ré échantillonnage dans les séries chronologiques, y compris les tests de racine unité. Nous avons présenté plusieurs tests pour détecter la non stationnarité et nous avons appliqué les techniques de bootstrap sur les tests de Dickey Fuller.

Notre objectif est de montrer la procédure de bootstrap dans les tests de racine unité c'est-à-dire les techniques de bootstrap utilisées pour détecter la non stationnarité du processus autorégressif.

**Mots- clés :** Bootstrap, hypothèses, processus autorégressif, puissance du test, seuil du test, racine unité, test de stationnarité.