

# CHAPITRE III

## Tests de Racine Unitaire

Michel LUBRANO

Septembre 2008

### Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Formalisation du Problème</b>	<b>3</b>
2.1	Composante déterministe et composante stochastique . . . . .	3
2.2	Un modèle général pour les tests . . . . .	5
<b>3</b>	<b>Tests de Dickey-Fuller</b>	<b>6</b>
3.1	Deux statistiques de test . . . . .	6
3.2	La distribution des tests est fonction de TD . . . . .	7
3.3	Régression en une ou deux étapes . . . . .	11
3.4	Tests séquentiels de la présence de plusieurs racines unitaires . . . . .	11
<b>4</b>	<b>Tests avec erreurs auto-corrélées</b>	<b>12</b>
4.1	Test de Dickey-Fuller augmenté . . . . .	13
4.2	Sélection des retards dans un test ADF . . . . .	13
4.3	Distribution de $\hat{\rho}_T$ dans le cas non IID . . . . .	14
4.4	Test non-paramétrique de Phillips-Perron . . . . .	15
<b>5</b>	<b>Tests de l'hypothèse de stationarité</b>	<b>17</b>
<b>6</b>	<b>Ruptures de Trend et Tests de Perron</b>	<b>19</b>
6.1	Modélisation d'une rupture de trend . . . . .	19
6.2	Motivation empirique . . . . .	21
6.3	Expérience de Monte Carlo . . . . .	23
6.4	Test avec trend segmenté . . . . .	24
6.5	Application sur données françaises . . . . .	25

<b>1</b>	<b>INTRODUCTION</b>	<b>2</b>
<b>7</b>	<b>Saisonnalité et Racines Unitaires</b>	<b>27</b>
7.1	Modélisation de la saisonnalité . . . . .	27
7.2	Test de la présence d'une seule racine . . . . .	28
7.3	Test de la présence de deux racines . . . . .	30
7.4	Décomposition de la racine saisonnière selon ses fréquences . . . . .	33
7.5	Influence de la périodicité sur les tests ADF . . . . .	36
<b>8</b>	<b>Questions Diverses</b>	<b>36</b>
8.1	Autres tests . . . . .	37
8.2	Construction des tables . . . . .	37
8.3	Intégration fractionnaire . . . . .	38
8.4	Tests Bayésiens . . . . .	39
<b>9</b>	<b>Conclusion</b>	<b>40</b>
<b>10</b>	<b>Lectures additionnelles</b>	<b>41</b>
<b>11</b>	<b>Exercices</b>	<b>41</b>
11.1	Processus I(1) . . . . .	41
11.2	Simulation d'un processus I(1) . . . . .	42
11.3	Processus ARIMA(0,1,1) . . . . .	42
11.4	Test ADF . . . . .	43
11.5	Les composantes déterministes . . . . .	43

# 1 Introduction

Il y a plusieurs raisons, tant statistiques qu'économiques pour s'intéresser à la présence d'une racine dans une série économique et nous allons essayer de les résumer dans cette introduction.

On a pu voir au cours du chapitre précédent que la présence d'une racine unitaire dans les données avait des conséquences très importantes sur le plan statistique. Premièrement les propriétés asymptotiques générales des estimateurs (vitesse de convergence, normalité asymptotique) ne tiennent plus. Il faut avoir recours à une théorie asymptotique spéciale. Ensuite la présence de régresseurs comportant une racine unitaire dans une régression peut conduire à estimer des régressions apparemment très bonnes entre des variables qui sont totalement indépendantes entre elles. C'est le problème des régressions factices. Enfin, une série trend stationnaire et une série stationnaire en différence se comportent de manière radicalement opposée dans le long terme. Une série trend stationnaire a tendance à se repositionner autour de son trend déterministe après un choc aléatoire. C'est ce que l'on appelle la propriété de "mean reversion". Une série stationnaire en différence ne revient pas autour de sa tendance après un choc, puisque le choc affecte aussi la tendance stochastique de la série.

La présence ou l'absence de la propriété de “mean reversion” a conduit une partie des macroéconomistes à s'intéresser de très près à la question des racines unitaires. On peut présenter trois domaines que nous allons ensuite un peu détailler:

- la théorie du cycle conjoncturel et la théorie de la croissance endogène avec la variable production
- la théorie du revenu permanent avec la variable consommation
- la théorie de l'hystérésis avec la variable chômage

La théorie du cycle conjoncturel implique une composante tendancielle déterministe dans l'évolution de la production. Les fluctuations cycliques n'affectent pas la tendance et la politique conjoncturelle parvient à stabiliser les fluctuations sans changer la tendance profonde. Au contraire avec la théorie de la croissance endogène, la tendance est stochastique et les chocs (réels ou monétaires) ont un impact permanent sur celle-ci. Voir par exemple King, Plosser, and Rebelo (1988).

La consommation peut être déterminée par le revenu courant comme chez Keynes, ou par le revenu permanent comme chez Friedman. Dans ce dernier cas, la consommation ne réagit pas à des changements transitoires du revenu, mais seulement par rapport à des changements du revenu anticipés de manière rationnelle. C'est le point de vue défendu et illustré dans Hall (1978). En conséquence, il n'est pas possible de prévoir les changements dans la consommation et celle-ci suit une marche aléatoire.

Blanchard and Summers (1986) ont introduit une théorie du marché du travail selon laquelle le salaire est déterminé de manière à conserver un même niveau d'emploi pour les insiders de la firme. Il en résulte un comportement particulier de la série d'emploi qui au temps  $t$  est égale à sa valeur au temps  $t - 1$  à un terme d'erreur près. L'emploi suit donc une marche aléatoire et si l'offre de travail est constante, le taux de chômage suit également une marche aléatoire. Il y a donc persistance du chômage d'où le terme d'hystérésis employé par les auteurs.

Il existe donc une ensemble de raisons variées, tant statistiques qu'économiques pour s'intéresser à la présence d'une racine unitaire dans une série. Pour mettre en évidence cette présence (ou son absence) les économètres se sont attachés à mettre au point différents tests que nous allons examiner dans ce chapitre. Ces tests ne sont pas très puissants et dépendent fort de conditions annexes qui tiennent à la forme de la composante déterministe de la série et éventuellement à la présence de saisonnalité.

## 2 Formalisation du Problème

### 2.1 Composante déterministe et composante stochastique

Pour poser le problème du test de la présence d'une racine unitaire, il est utile de décomposer une série en deux types de composantes: une composante déterministe  $TD_t$  et une composante stochastique  $u_t$ :

$$y_t = TD_t + u_t \quad (1)$$

L'hypothèse de racine unitaire concerne la partie stochastique  $u_t$ , mais la spécification correcte de la partie déterministe est cruciale pour l'établissement des tests. En fait ce qui importe principalement, c'est l'ordre de cette partie déterministe. On a l'habitude de distinguer trois cas:

- $TD_t = 0$  ou pas de partie déterministe (la partie déterministe est  $o(1)$ ).
- $TD_t = \mu$  ou seulement un terme constant. On peut également ajouter des constantes saisonnières qui sont de même ordre, c'est à dire  $O(1)$ .
- $TD_t = \mu + \delta t$  ou cette fois-ci une constante et un trend.  $TD_t$  est alors  $O(T)$ .

On peut compliquer la description du trend en considérant des trends non-linéaires qui vont changer dans le temps de manière à modéliser un changement structurel exogène. Perron (1989) montre que dans ce cas les résultats des tests précédents peuvent être inversés. L'étude de ce cas est donc extrêmement importante et la section 6 y sera entièrement consacrée.

C'est dans la partie stochastique  $u_t$  que va pouvoir se trouver la racine unitaire. On va modéliser cette partie stochastique au moyen d'un processus ARMA:

$$\tilde{A}(L) u_t = B(L) \epsilon_t \quad (2)$$

A cause de la présence du terme  $TD_t$ , on peut supposer que la moyenne de  $u_t$  est nulle, la moyenne du processus étant contenue dans  $TD_t$ . On va supposer que la partie moyenne mobile est inversible, c'est à dire que toutes les racines de  $B(L)$  sont en dehors du cercle unité. De plus dans les cas que l'on va envisager par la suite, on va supposer que le polynôme rationnel infini  $B^{-1}(L) \tilde{A}(L)$  peut s'approximer par un polynôme  $A(L)$  de degré fini et que donc il suffit de considérer un processus auto-régressif pour la partie stochastique donné par:

$$A(L) u_t = \epsilon_t \quad (3)$$

On peut maintenant distinguer deux modèles alternatifs pour  $y_t$ .

- dans le premier cas  $y_t$  est trend stationnaire, toutes les racines de l'équation caractéristique  $A(z) = 0$  sont en dehors du cercle unité. Le processus de  $u_t$  est donc stationnaire et celui de  $y_t$  est stationnaire autour d'un trend.
- dans le deuxième cas où  $y_t$  est stationnaire en différence,  $A(z) = 0$  comporte une racine sur le cercle unité et toutes les autres racines sont en dehors du cercle unité. On a donc la factorisation:

$$A(L) = (1 - L) A^*(L) \quad (4)$$

et  $(1 - L) u_t$  est stationnaire. Dans ce cas,  $(1 - L) y_t$  est stationnaire autour d'une moyenne fixe.

## 2.2 Un modèle général pour les tests

Le but des tests de racine unitaire est donc de tester la présence d'une racine unitaire dans la partie auto-régressive de la partie stochastique du processus de  $y_t$ . Pour faire ce test il est nécessaire de trouver un modèle qui contienne à la fois les deux modèles représentatifs que l'on a dégagé, trend stationnaire et stationnaire en différence de manière à ce qu'une hypothèse puisse apparaître comme une restriction paramétrique dans le modèle général. Le modèle que l'on a adopté peut se réécrire après substitution:

$$A(L) (y_t - TD_t) = \epsilon_t \quad (5)$$

Il est très intéressant de développer cette écriture dans un cas simple, celui où  $A(L) = (1 - \rho L)$ . On examinera par la suite le cas où  $A(L)$  est un polynôme de degré supérieur. Prenons ensuite le cas général  $TD_t = \mu + \delta t$ . Il vient:

$$\begin{aligned} (1 - \rho L) (y_t - \mu - \delta t) &= \epsilon_t \\ y_t &= \rho y_{t-1} + (1 - \rho) (\mu + \delta t) + \rho \delta + \epsilon_t \end{aligned}$$

Cette écriture conduit à formuler plusieurs observations.

- Elle montre tout d'abord comment les deux modèles sont emboîtés. Pour  $\rho = 1$  on retrouve le modèle de marche aléatoire. Par contre, on ne retrouve le modèle trend stationnaire que pour  $\rho = 0$ , ce qui fait que le test de  $\rho = 1$  contre  $\rho < 1$  n'est à proprement parler qu'un test de la présence d'une racine unitaire contre une alternative qui inclut l'autre modèle, mais pas que celui-ci.
- Ensuite cette écriture permet de voir les particularités du modèle de marche aléatoire. Pour  $\rho = 1$  on constate que le terme constant  $\mu$  n'est plus identifié. Une marche aléatoire n'a pas de moyenne constante au cours du temps. Deuxièmement le régresseur  $t$  disparaît, mais pas son coefficient  $\delta$  qui vient prendre la place du terme constant. Ce coefficient  $\delta$  conserve la même interprétation. Il représente la *dérive* du processus, ou *drift*. Si la série étudiée est en logarithme, alors le modèle  $\Delta y_t = \delta + \epsilon_t$  indique de façon claire que le taux de croissance de  $y_t$  est égal à  $\delta$  augmenté d'un terme aléatoire de moyenne nulle.
- Enfin il est habituel de considérer un modèle plus simple pour le test qui est:

$$y_t = \rho y_{t-1} + \beta_0 + \beta_1 t + \epsilon_t \quad (6)$$

avec:

$$\begin{aligned} \beta_0 &= (1 - \rho) \mu + \rho \delta \\ \beta_1 &= (1 - \rho) \delta \end{aligned} \quad (7)$$

Il est facile de voir que l'estimateur des moindres carrés de  $\rho$  a la même valeur et le même écart-type quelle que soit la paramétrisation employée. Toutefois ce modèle cache le fait qu'à  $\rho = 1$  correspond  $\beta_1 = 0$ . C'est pourtant cette écriture qui est généralement employée pour les tests car l'estimateur de  $\rho$  dans les deux cas est numériquement le même. Seuls changent les estimateurs des coefficients de  $TD_t$ .

### 3 Tests de Dickey-Fuller

Le modèle de régression que l'on a mis en avant:

$$y_t = \rho y_{t-1} + \beta_0 + \beta_1 t + \epsilon_t \quad (8)$$

est appelée régression de Dickey-Fuller à la suite de leur papier de 1981. L'idée d'un test de racine unitaire est très simple. Il suffit d'estimer cette régression par moindres carrés et de tester ensuite  $\rho = 1$  au moyen de la statistique de Student. Retirons maintenant  $y_{t-1}$  des deux côtés de la régression (8):

$$\Delta y_t = (\rho - 1) y_{t-1} + \beta_0 + \beta_1 t + \epsilon_t \quad (9)$$

L'interprétation de la statistique de Student devient directe étant donné qu'il suffit maintenant de tester la nullité du coefficient de  $y_{t-1}$ , en se souvenant que la distribution de cette statistique n'est pas asymptotiquement normale. Enfin pour certains calculs, il sera utile d'avoir présent à l'esprit la décomposition suivante de  $y_t$  en fonction des conditions initiales en se servant de la relation:

$$(1 - \rho L)^{-1} = \sum_{i=0}^{\infty} \rho^i L^i = \sum_{i=0}^{t-1} \rho^i L^i + \rho^t L^t \sum_{i=0}^{\infty} \rho^i L^i \quad (10)$$

ce qui donne:

$$\begin{aligned} y_t &= \sum_{i=0}^{\infty} \rho^i L^i (\beta_0 + \beta_1 t + \epsilon_t) \\ &= \rho^t y_0 + \beta_0 \sum_{i=0}^{t-1} \rho^i + \beta_1 \sum_{i=0}^{t-1} \rho^i (t - i) + \sum_{i=0}^{t-1} \rho^i \epsilon_{t-i} \end{aligned} \quad (11)$$

Quand  $\rho = 1$ , cette écriture permet de voir que le modèle comporte un trend au carré quand on n'impose pas que  $\beta_1 = 0$  sous l'hypothèse nulle :

$$y_t = y_0 + \beta_0 t + 0.5\beta_1(t + t^2) + \sum_{i=0}^{t-1} \epsilon_{t-i}$$

car  $\sum_{i=1}^{t-1} (t - i) = t(t + 1)/2$ .

#### 3.1 Deux statistiques de test

En partant de la régression de test (9), on va pouvoir employer deux statistiques de test. La première est la statistique de Student usuelle, que l'on va noter  $\tau$ , pour bien souligner le fait qu'elle a une distribution asymptotique qui n'est pas normale :

$$\tau = \frac{\hat{\rho}_T - 1}{\hat{\sigma}_{\rho_T}}. \quad (12)$$

La seconde statistique est basée sur le fait que  $T(\hat{\rho}_T - 1)$  converge en distribution vers une fonctionnelle de Browniens qui ne dépend pas du paramètre de nuisance  $\sigma^2$ . On a

déterminé cette distribution dans le chapitre précédent. On peut utiliser directement cette statistique pour tester  $H_0 : \rho = 1$ . Ces deux statistiques ont des distributions asymptotiques qui sont très liées.

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

$$z = T(\hat{\rho}_T - 1) = \frac{T^{-1} \sum y_{t-1} \epsilon_t}{T^{-2} \sum y_{t-1}^2} \xrightarrow{\mathcal{L}} \frac{0.5(W^2(1) - 1)}{\int_0^1 W^2(r) dr}, \quad (13)$$

ce qui a permis de montrer que  $\hat{\rho}_T$  est un estimateur super consistant de  $\rho$ . On sait également que  $\sqrt{T}(\hat{\rho}_T - 1)$  converge en probabilité vers zéro. En effet, la variance du numérateur est d'ordre  $T$ , alors que le dénominateur converge vers une quantité positive.

Passons maintenant à la statistique  $\tau$ . Sa distribution sous  $H_0$  peut facilement se déduire de la distribution de la statistique  $z$ . En effet:

$$\begin{aligned} \tau &= \frac{T(\hat{\rho}_T - 1)}{(\hat{\sigma}_\epsilon^2)^{1/2} (T^{-2} \sum y_{t-1}^2)^{-1/2}} \\ &= \frac{T^{-1} \sum y_{t-1} \epsilon_t}{T^{-2} \sum y_{t-1}^2} \cdot \frac{(T^{-2} \sum y_{t-1}^2)^{1/2}}{(\hat{\sigma}_\epsilon^2)^{1/2}} \\ &= \frac{T^{-1} \sum y_{t-1} \epsilon_t}{(T^{-2} \sum y_{t-1}^2)^{1/2} (\hat{\sigma}_\epsilon^2)^{1/2}}. \end{aligned}$$

Comme on a montré que  $\hat{\rho}_T$  était un estimateur super consistant de  $\rho$ , alors  $\hat{\sigma}_\epsilon^2$  converge en probabilité vers  $\sigma_\epsilon^2$ . D'autre part, les résultats de convergence obtenus dans le chapitre 2 permettent d'écrire pour  $T \rightarrow \infty$  :

$$\tau \xrightarrow{\mathcal{L}} \frac{0.5 \sigma^2 (W^2(1) - 1)}{(\sigma^2 \int_0^1 W^2(r) dr)^{1/2} \sigma} = \frac{0.5 (W^2(1) - 1)}{(\int_0^1 W^2(r) dr)^{1/2}}. \quad (14)$$

La seule différence, c'est que le dénominateur de la distribution de  $\tau$  est égal la racine carrée du dénominateur de la distribution de  $z$ . Ces deux statistiques auront donc des comportements similaires, mais des valeurs critiques différentes. La statistique  $\tau$  est la plus utilisée. Mais la statistique  $z$  va servir de base au test de Philipps-Perron.

### 3.2 La distribution des tests est fonction de TD

Plusieurs modèles sont à distinguer en fonction de la partie déterministe  $TD_t$  introduite dans la régression de test.

- **Cas sans terme déterministe:** si  $TD_t = 0$  alors le modèle s'écrit:

$$\Delta y_t = (\rho - 1) y_{t-1} + \epsilon_t. \quad (15)$$

L'hypothèse nulle et l'hypothèse alternative sont:

$$\begin{cases} H_0 : \rho = 1 \\ H_1 : \rho < 1. \end{cases} \quad (16)$$

Il s'agit donc d'un test unilatéral puisque l'hypothèse alternative est  $\rho < 1$  et non  $\rho \neq 1$ . On vérifie au moyen de (11) que sous l'hypothèse alternative de stationnarité  $y_t = O_P(1)$  et qu'il en est bien sûr de même sous l'hypothèse nulle. Le test se fait soit en calculant la statistique de Student:

$$\tau = \frac{\hat{\rho}_T - 1}{\hat{\sigma}_{\rho_T}} \quad (17)$$

qui a la distribution limite (14), soit en calculant l'autre statistique

$$z = T(\hat{\rho}_T - 1)$$

dont on a donné la distribution limite en (13).

- **Cas avec constante:** si  $TD_t = \mu$  le modèle se complique un peu. On a:

$$\Delta y_t = (\rho - 1)y_{t-1} + \beta_0 + \epsilon_t \quad (18)$$

L'hypothèse nulle n'est plus la même que précédemment. On veut toujours tester  $\rho = 1$  contre  $\rho < 1$  mais l'autre paramètre doit bouger aussi. On a:

$$\begin{cases} H_0 : \rho = 1, & \beta_0 = 0 \\ H_1 : \rho < 1 \end{cases} \quad (19)$$

Ceci est bien visible si l'on remarque qu'en fait  $\beta_0 = (1 - \rho)\mu$ . Mais aussi si l'on veut que  $y_t$  ait le même ordre sous l'hypothèse nulle et sous l'alternative de stationnarité, il faut que sous l'hypothèse nulle la dérive de la marche aléatoire disparaisse, car sinon  $y_t$  serait  $O_P(T)$  alors que l'hypothèse alternative de stationnarité implique  $y_t = O_P(1)$ . Ceci peut aisément se vérifier au moyen de (11). Les deux tests sont basés sur les mêmes statistiques que précédemment. Il est cependant utile d'en modifier légèrement les notations avec  $\tau_\mu$  pour la statistique de Student et  $z_\mu$  pour l'autre statistique. En effet, les distributions asymptotiques sont légèrement différentes. On peut les trouver dans Phillips and Perron (1988) avec:

$$\tau_\mu \xrightarrow{\mathcal{L}} \frac{\int_0^1 W_*(r) dW(r)}{\left\{ \int_0^1 W_*^2(r) dr \right\}^{1/2}}, \quad (20)$$

et

$$z_\mu \xrightarrow{\mathcal{L}} \frac{\int_0^1 W_*(r) dW(r)}{\int_0^1 W_*^2(r) dr}. \quad (21)$$

Les formules sont identiques aux précédentes, sauf que le processus  $W_*(r)$  est en quelque sorte pris en déviation par rapport à sa "moyenne" et est défini par:

$$W_*(r) = W(r) - \int_0^1 W(r) dr. \quad (22)$$



- **Cas avec trend:** si enfin  $TD_t = \mu + \delta t$  le modèle est le suivant:

$$\Delta y_t = (\rho - 1)y_{t-1} + \beta_0 + \beta_1 t + \epsilon_t \quad (23)$$

L'hypothèse nulle a encore un peu changé. Cette fois-ci on a:

$$\begin{cases} H_0 : \rho = 1, & \beta_1 = 0 \\ H_1 : \rho < 1 \end{cases} \quad (24)$$

Ceci se justifie encore en explicitant la reparamétrisation car maintenant  $\beta_0 = (1 - \rho)\mu + \rho\delta$  et  $\beta_1 = (1 - \rho)\delta$ . Mais de même  $y_t$  doit être encore une fois du même ordre sous les deux hypothèses. Si  $H_1$  signifie que  $y_t$  est stationnaire autour d'un trend, cela signifie cette fois-ci que  $y_t = O_P(T)$ . Sous  $H_0$  le même résultat implique que l'on ait une dérive et rien d'autre. Le test de Student se note dans ce cas  $\tau_\tau$  et l'autre test  $z_\tau$ . les distributions sont encore un peu différentes que précédemment et sont données encore dans Phillips and Perron (1988):

$$\begin{aligned} \tau_\tau &\xrightarrow{\mathcal{L}} \frac{H + H_\tau}{(K + K_\tau)^{1/2}} \\ z_\tau &\xrightarrow{\mathcal{L}} \frac{H + H_\tau}{K + K_\tau} \end{aligned}$$

où  $H$  et  $K$  correspondent aux termes de la distribution du simple test  $\tau$  (14):

$$\begin{aligned} H &= 0.5 (W(1)^2 - 1) \\ K &= \int_0^1 W^2(r) dr \end{aligned}$$

et les termes supplémentaires répondent à la définition suivante:

$$\begin{aligned} H_\tau &= 12 \left[ \int_0^1 r W(r) dr - \frac{1}{2} \int_0^1 W(r) dr \right] \cdot \left[ \int_0^1 W(r) dr - \frac{1}{2} W(1) \right] - W(1) \int_0^1 W(r) dr \\ K_\tau &= -12 \left[ \int_0^1 r W(r) dr \right]^2 + 12 \int_0^1 W(r) dr \int_0^1 r W(r) dr - 4 \left[ \int_0^1 W(r) dr \right]^2. \end{aligned}$$

Il est alors intéressant de montrer au moyen d'une écriture utilisée par exemple dans Boswijk (1992) comment les distributions asymptotiques de ces tests s'emboîtent les unes dans les autres. Considérons la fonction particulière  $g(W, U)$  suivante:

$$g(W, U) = \frac{\int_0^1 U(r) dW(r)}{\left\{ \int_0^1 U^2(r) dr \right\}^{1/2}} \quad (25)$$

où  $W(r)$  et  $U(r)$  sont deux processus de Wiener sur  $[0, 1]$ . Définissons ensuite la procédure qui permet de prendre un processus de Wiener en déviation par rapport à sa "moyenne", quand celle-ci a une expression un peu générale. On connaît la formule standard qui au

moyen de la matrice de projection  $M_X$  permet de retirer de la variable  $Y$  l'influence de  $X$  en calculant  $M_X Y$  comme les résidus de la régression de  $Y$  sur  $X$ . Si  $X(r)$  et  $Y(r)$  sont deux processus stochastiques vectoriels en temps continu, on définit l'équivalent des résidus  $M_X Y$  par:

$$M_X Y(r) = Y(r) - \int_0^1 Y(r) X(r)' dr \left[ \int_0^1 X(r) X(r)' dr \right]^{-1} X(r). \quad (26)$$

Pour le cas avec constante, on définit  $M_{[1]} W(r)$  avec  $X(r) = 1$  et pour le cas avec constante et trend, on définit  $M_{[1,r]} W(r)$  avec  $X(r) = [1, r]$ . Ce qui fait que:

$$\tau \xrightarrow{\mathcal{L}} g(W, W) \quad (27)$$

$$\tau_\mu \xrightarrow{\mathcal{L}} g(W, M_{[1]} W) = g(W, W_*) \quad (28)$$

$$\tau_\tau \xrightarrow{\mathcal{L}} g(W, M_{[1,r]} W). \quad (29)$$

Les résultats analytiques sur la distribution asymptotique de la statistique de Student ne sont en général que de peu d'utilité pratique, car on ne connaît pas d'expression paramétrique pour un processus de Wiener. Tout ce que l'on puisse faire, c'est le simuler par une procédure de Monte Carlo pour donner des tables. On peut trouver des tables dans Fuller (1976), Dickey and Fuller (1981) ou MacKinnon (1991). Dans le cas asymptotique, les résultats s'accommodent de conditions très peu restrictives sur les erreurs qui n'ont besoin que d'être non auto-corrélées. Les erreurs peuvent en particulier être non normales et hétéroscédastiques. Les simulations des valeurs critiques en petit échantillon ne sont en revanche valides que sous l'hypothèse de normalité des erreurs. Une grande prudence est donc requise pour leur utilisation. La Table 1 donne les valeurs critiques du test en  $\tau$ . Elles

Table 1: Distribution de  $\tau$  sous  $H_0 : \Delta y_t \sim N(0, 1)$

statistique									
	$\tau$			$\tau_\mu$			$\tau_\tau$		
	probabilité dans la queue de gauche								
T	1%	5%	10%	1%	5%	10%	1%	5%	10%
25	-2.66	-1.96	-1.62	-3.72	-2.98	-2.63	-4.37	-3.60	-3.24
50	-2.61	-1.95	-1.62	-3.57	-2.92	-2.60	-4.15	-3.50	-3.18
100	-2.59	-1.94	-1.62	-3.50	-2.89	-2.58	-4.05	-3.45	-3.15
250	-2.57	-1.94	-1.62	-3.46	-2.87	-2.57	-4.00	-3.43	-3.14
500	-2.57	-1.94	-1.62	-3.45	-2.87	-2.57	-3.98	-3.42	-3.13
$\infty$	-2.57	-1.94	-1.62	-3.43	-2.86	-2.57	-3.96	-3.41	-3.13

sont calculées à partir de MacKinnon (1991). La Table 2 donne les valeurs critiques asymptotiques pour le test en  $z$ . Comme MacKinnon (1991) ne permet pas de calculer les valeurs critiques du test  $z$ , nous avons utilisé les résultats de Fuller (1976) en ne donnant que les

valeurs critiques asymptotiques. On peut tirer plusieurs conclusions de l'examen de ces tables<sup>1</sup>. Premièrement on voit bien que ces valeurs critiques du test en  $\tau$  sont très différentes de la Normale où les valeurs équivalentes seraient de 1.645 à 5% et 1.282 à 10%. Ensuite, la valeur critique dépend de  $TD_t$ . Elle croit en valeur absolue avec le nombre de régresseurs contenus dans  $TD_t$ . Enfin on peut montrer [voir Schwert (1989)] que la puissance de ces tests contre une alternative stationnaire diminue avec le nombre de composantes de  $TD_t$ .

Table 2: Distribution de  $z$  sous  $H_0 : \Delta y_t \sim N(0, 1)$ 

statistique									
	$z$			$z_\mu$			$z_\tau$		
	probabilité dans la queue de gauche								
T	1%	5%	10%	1%	5%	10%	1%	5%	10%
$\infty$	-13.7	-8.0	-5.7	-20.6	-14.1	-11.2	-29.4	-21.7	-18.2

### 3.3 Régression en une ou deux étapes

Jusqu'à présent on a calculé les tests de DF et ADF au moyen d'une régression en une seule étape. Mais on peut également imaginer, comme cela est fait par exemple dans Campbell and Perron (1991) de séparer la régression en deux. On commencera par estimer une régression de  $y_t$  sur les régresseurs déterministes contenus dans  $TD_t$  pour en tirer les résidus. Cette décomposition peut se noter:

$$\hat{u}_t = y_t - TD_t \quad (30)$$

où donc  $\hat{u}_t$  représente les résidus estimés. Puis ensuite on fera les tests habituels sur les résidus. Ceci revient à décomposer la série en sa partie déterministe et sa partie stochastique, seule la dernière nous intéressant pour effectuer les tests. On peut montrer que tant que le modèle complet est linéaire, les tests basés sur la régression en une étape et ceux basés sur la régression en deux étapes sont asymptotiquement équivalents.

### 3.4 Tests séquentiels de la présence de plusieurs racines unitaires

On a supposé jusqu'à présent que les séries examinées présentaient au maximum une racine unitaire, c'est à dire que le polynôme  $A(L)$  avait toutes ses racines endehors du cercle unité, sauf une qui était sur le cercle. Sous l'hypothèse alternative, la série est stationnaire, ou stationnaire autour d'un trend. Dans le cas de la présence de deux racines unitaires, on aura non-stationnarité à la fois sous l'hypothèse nulle, mais également sous l'hypothèse alternative, ce qui fait que le test précédent ne fonctionne plus. On va alors mettre en place, à la suite de Dickey and Pantula (1987), une procédure séquentielle. On va par exemple

<sup>1</sup>Les premières tables furent publiées dans Fuller (1976). Mais elles sont moins précises que celle-ci.

différencier deux fois la série et tester la présence d'une racine unitaire. Si cette présence est rejetée, on passe à une seule différentiation pour se retrouver dans le cadre usuel. Ce qui fait que l'on sera sûr à chaque étape que la série sera stationnaire sous l'hypothèse alternative. En d'autres termes, on va factoriser (ou imposer)  $k - 1$  racines unitaires pour tester la présence d'une  $k^{\text{ième}}$  racine. On se ramène donc toujours au test de la présence d'une seule racine.

Prenons l'exemple simple où l'on a factorisé deux racines unitaires dans  $A(L)$ , ce qui fait que le modèle se note maintenant:

$$(1 - \rho L)(1 - \phi L)(y_t - \mu) = \epsilon_t. \quad (31)$$

On ne modifie pas le produit en considérant  $((1 - L) - (\rho - 1)L)((1 - L) - (\phi - 1)L)$ , ce qui fait que la régression de test s'écrit:

$$\Delta^2 y_t = \beta_0 + (\rho + \phi - 2) \Delta y_{t-1} - (\rho - 1)(\phi - 1) y_{t-2} + \epsilon_t. \quad (32)$$

Alors pour  $\phi = 1$ , on voit que l'on a une première régression de test qui est :

$$\Delta^2 y_t = \beta_0 + \beta_1 \Delta y_{t-1} + \epsilon_t \quad (33)$$

Le test  $\tau_\mu$  sur  $\beta_1$  teste l'hypothèse nulle de deux racines unitaires (une que l'on a imposé et l'autre que l'on teste). Si cette hypothèse est rejetée en utilisant les tables standards données dans le texte, on passe à la régression augmentée suivante :

$$\Delta^2 y_t = \beta_0 + \beta_1 \Delta y_{t-1} + \beta_2 y_{t-1} + \epsilon_t. \quad (34)$$

Le test  $\tau_\mu$  sur  $\beta_2$  teste l'hypothèse nulle d'une seule racine unitaire. Dans cet exemple, nous n'avons pas détaillé la forme de  $\beta_0$ . Il faut en fait adapter le terme déterministe en fonction du nombre de racines unitaires que l'on teste. Si le modèle initial comporte une constante, la régression la plus contrainte n'en comporte pas. Plus précisément, développons l'expression générale

$$\begin{aligned} (1 - \rho L)(1 - \phi L)(\mu + \delta t) &= (1 - \rho)(1 - \phi)\mu \\ &+ (\phi + \rho - 2\rho\phi)\delta + (1 - \rho - \phi + \rho\phi)\delta t. \end{aligned}$$

Dans le premier cas où l'on impose  $\phi = 1$  pour tester  $\rho = 1$ , le trend disparaît et il ne reste que  $(1 - \rho)\delta$  comme terme constant.

## 4 Tests avec erreurs auto-corrélées

Les tests développés jusqu'à présent ne reposent pas sur l'hypothèse de Normalité des erreurs. Ils sont valides, du moins asymptotiquement, sous des hypothèses relativement générales concernant les erreurs. Mais ils ne sont plus valables dès que les erreurs sont autocorrélées. On a imaginé dans la littérature deux types de corrections. La première revient, dans une approche paramétrique, à modifier la régression de test. La seconde consiste à modifier la statistique de test proprement dite.

### 4.1 Test de Dickey-Fuller augmenté

Il est relativement facile de modifier la régression de test pour tenir compte de l'auto-corrélation. Le modèle autorégressif de base sur lequel on s'est basé était le suivant:

$$A(L)(y_t - TD_t) = \epsilon_t. \quad (35)$$

Nous nous sommes limités pour l'instant au cas où  $A(L)$  était un polynôme de degré un. On va maintenant simplement autoriser  $A(L)$  à être de degré  $p \geq 1$ . Factorisons ce polynôme  $A(L)$  selon la formule explicitée dans le chapitre 1:

$$A(L) = (1 - \rho L) - (1 - L) A^*(L)$$

où  $A^*(L)$  est un polynôme de degré  $p - 1$  sans terme constant:

$$A^*(L) = \alpha_1^* L + \alpha_2^* L^2 + \dots + \alpha_{p-1}^* L^{p-1}$$

et  $\rho = 1 - A(1)$ . On va supposer que le polynôme  $A^*(L)$  a toutes ses racines en dehors du cercle unité et s'intéresser à tester l'hypothèse nulle d'une seule racine unitaire  $\rho = 1$ . Effectuons la multiplication entre la factorisation de  $A(L)$  et  $(y_t - TD_t)$ :

$$(1 - \rho L)(y_t - TD_t) = A^*(L)(\Delta y_t - \Delta TD_t) + \epsilon_t \quad (36)$$

ce qui montre que l'on obtient le même type de modèle que précédemment, mais que cette fois-ci on a simplement rajouté les différences premières retardées de  $y_t$ . Pour le cas simple où  $p = 2$  et  $TD_t = (\mu + \delta t)$  on a  $A^*(L) = \alpha_1^* L$  et:

$$y_t = \rho y_{t-1} + (1 - \rho)(\mu + \delta t) + (\rho + \alpha_1^*)\delta + \alpha_1^* \Delta y_{t-1} + \epsilon_t. \quad (37)$$

Sous l'hypothèse nulle  $\rho = 1$ , on retrouve les mêmes propriétés de la régression à savoir que le trend disparaît, que le terme constant initial disparaît et est remplacé par une fonction du paramètre du trend.

Le test de l'hypothèse nulle s'effectue de la même manière que précédemment au moyen d'une statistique de Student que l'on va noter  $\tau'$ ,  $\tau'_\mu$ ,  $\tau'_\tau$  par analogie avec le cas précédent. Ce type de correction paramétrique a été proposée initialement par Dickey and Fuller (1979), d'où l'appellation de test de Dickey-Fuller augmenté ou ADF (augmentation de la régression initiale des retards de  $\Delta y_t$ ). Le résultat remarquable, c'est que  $\tau$  et  $\tau'$  ont la même distribution asymptotique sous des conditions très peu restrictives comme l'ont montré entre autres Phillips and Perron (1988) sans que l'on ait besoin de connaître la valeur de  $p$  comme l'avaient supposé Dickey and Fuller (1979).

### 4.2 Sélection des retards dans un test ADF

La taille et la puissance du test ADF dépendent fortement du nombre de retards. Si celui-ci est trop faible, il y aura une distortion de taille et si celui-ci est trop grand, il y aura une perte de puissance. On trouvera dans Ng and Perron (1995) une étude comparative des différentes méthodes de choix.

- Dans un modèle dynamique ajusté sur des séries stationnarisées, le nombre maximum de retards peut être choisi au moyen d'un critère d'information d'Akaike

$$I_{AIC} = \log \hat{\sigma}^2 + p \frac{2}{T},$$

ou de Schwarz

$$I_{SC} = \log \hat{\sigma}^2 + p \frac{\log T}{T}.$$

Ici on peut également utiliser ce type de critère car ce sont les retards d'une variable stationnarisée  $\Delta y_t$  que l'on doit sélectionner. En effet Sims, Stock, and Watson (1990) ont montré que l'on peut traiter de manière indépendante les régresseurs  $I(0)$  et les régresseurs  $I(1)$  et utiliser les tests usuels pour les coefficients des régresseurs  $I(0)$ . On part donc d'une régression de test avec suffisamment de retards et on sélectionne le modèle parcimonieux qui minimise le critère d'information. Ng and Perron (1995) montrent cependant dans une expérience de Monte Carlo que ces deux critères d'information ont tendance à choisir des valeurs trop petites de  $p$ , ce qui a pour conséquence de distordre la taille du test.

- Un autre façon de choisir les retards dans un modèle AR consiste aussi à partir d'un modèle avec un grand nombre de retards et à éliminer au fur et à mesure les retards qui ont un  $t$  de Student non significatif. Ng and Perron (1995) montrent toujours dans une expérience de Monte Carlo que cette procédure conduit à choisir un nombre de retards plus important que dans le cas précédent. Le test ADF perd en puissance, mais la distorsion de taille diminue.

La présence d'une composante MA dans le processus peut poser des problèmes particuliers. On sait que l'on peut approximer un processus MA par un processus AR d'ordre supérieur. Donc, il devrait suffire de choisir un nombre de retards suffisants pour résoudre le problème. Pourtant, Schwert (1989) a montré que les tests ADF présentaient de grandes distorsions de taille en présence de composante MA négative. Dans leur expérience de Monte Carlo, Ng and Perron (1995) montrent que l'ajout de retards supplémentaires a du mal à réduire la distorsion de taille dans le cas d'une forte composante MA négative.

### 4.3 Distribution de $\hat{\rho}_T$ dans le cas non IID

Nous avons donné en (13) la distribution asymptotique de  $T(\hat{\rho} - 1)$  pour le modèle  $y_t = \rho y_{t-1} + \epsilon_t$  en supposant que les erreurs  $\epsilon_t$  étaient IID. Nous allons maintenant relâcher cette hypothèse et voir comment se transforme cette distribution quand  $\rho$  est estimé dans un modèle de regression sans correction paramétrique, alors que celle-ci serait nécessaire pour tenir compte de l'auto-corrélation du terme d'erreur.

On a montré que dans le cas IID  $T^{-1} \sum y_{t-1} \epsilon_t \rightarrow (W^2(1) - 1)\sigma^2/2$ . Ce résultat ne tient plus dans le cas non IID. Dans le cas non IID, la variance d'une somme n'est plus égale à la somme des variances, il faut rajouter les covariances. On doit alors distinguer entre ce que l'on appelle la variance de long terme

$$\sigma_\infty^2 = \lim T^{-1} E[(\sum \epsilon_j)^2]$$

et  $\sigma_\epsilon^2$  la variance des erreurs

$$\sigma_\epsilon^2 = \lim T^{-1} \sum E[\epsilon_j^2]$$

Calculons la variance de long terme:

$$\begin{aligned} E[(\sum \epsilon_j)^2] &= E[\sum_{i=1}^T \epsilon_i^2 + 2 \sum_{i=1}^{T-1} \sum_{j=i+1}^T \epsilon_i \epsilon_j] \\ &= T\sigma_\epsilon^2 + 2 \sum_{i=1}^{T-1} \sum_{j=i+1}^T \gamma_\epsilon(j-i) \\ &= T\sigma_\epsilon^2 + 2T \sum_{i=1}^{T-1} (1-i/T) \gamma_\epsilon(i). \end{aligned}$$

Si les  $\gamma_\epsilon(i)$  sont nuls, alors  $\sigma_\infty^2 = \sigma_\epsilon^2$ .

Repartons maintenant de l'expression de l'estimateur des moindres carrés. On a toujours la décomposition suivante pour le numérateur:

$$\sum y_{t-1} e_t = 0.5(y_T^2 - \sum \epsilon_t^2).$$

Le comportement limite du premier terme

$$\frac{y_T^2}{T} = \frac{1}{T}(\epsilon_1 + \dots + \epsilon_T)^2 \xrightarrow{\mathcal{L}} \sigma_\infty^2 W^2(1).$$

Quant au deuxième terme

$$\text{plim } \frac{1}{T} \sum \left(\frac{\epsilon_t}{\sqrt{T}}\right)^2 = \sigma_\epsilon^2.$$

En regroupant ces résultats partiels, on obtient

$$\frac{1}{T} \sum y_{t-1} e_t \xrightarrow{\mathcal{L}} \frac{1}{2} \sigma_\infty^2 \left[ W(1)^2 - \frac{\sigma_\epsilon^2}{\sigma_\infty^2} \right]$$

qui est donc une généralisation du cas IID que l'on retrouve comme cas particulier quand  $\sigma_\epsilon^2 = \sigma_\infty^2$ . Il est facile de voir que la distribution du dénominateur de l'estimateur des moindres carrés est

$$\frac{1}{T^2} \sum y_{t-1}^2 \xrightarrow{\mathcal{L}} \sigma_\infty^2 \int_0^1 W(r)^2 dr.$$

En regroupant ces résultats, on arrive au résultat final

$$T(\hat{\rho} - 1) \xrightarrow{\mathcal{L}} \frac{1}{2} \frac{W^2(1) - \sigma_\epsilon^2/\sigma_\infty^2}{\int_0^1 W^2(r) dr}$$

#### 4.4 Test non-paramétrique de Phillips-Perron

Phillips (1987) et Phillips and Perron (1988) ont utilisé ce résultat pour proposer un test alternatif au test ADF. En réarrangeant les termes de la dernière expression, on trouve

$$T(\hat{\rho} - 1) \xrightarrow{\mathcal{L}} \frac{0.5(W^2(1) - 1) - \lambda/\sigma_\infty^2}{\int_0^1 W^2(r) dr}$$

avec  $\lambda = (\sigma_\infty^2 - \sigma_\epsilon^2)/2$ . On peut alors définir la statistique de test suivante

$$Z^{PP} = T(\hat{\rho} - 1) - \lambda T^2 / \sum y_{t-1}^2.$$

où l'on a remplacé  $\int_0^1 W^2(r) dr$  par son estimateur  $\sum y_{t-1}^2 / T^2$ . On peut facilement montrer que la statistique  $Z^{PP}$  a la même distribution asymptotique que statistique  $z$  de Dickey et Fuller. Il s'agit maintenant d'estimer les deux variances. Pour  $\sigma_\epsilon^2$ , les choses sont simples. L'estimateur

$$s_\epsilon^2 = T^{-1} \sum \epsilon_t^2$$

est consistant. Par contre, il est plus difficile de trouver un estimateur consistant de la variance de long terme. Phillips (1987) a montré que l'estimateur

$$s_\infty^2 = \frac{1}{T} \sum \epsilon_t^2 + \frac{2}{T} \sum_{j=1}^l \sum_{t=j+1}^T \epsilon_t \epsilon_{t-j}$$

était un estimateur consistant de  $\sigma_\infty^2$  où  $l$  est la troncature. Mais il est d'usage de préférer l'estimateur de Newey and West (1987) qui introduit une fenêtre de Barlett pour pondérer les autocovariances. On peut déterminer la valeur de  $l$  en regardant la graphique des autocorrélations de  $e_t = \Delta y_t$  ou bien en utilisant les techniques liées à l'estimation de la densité spectrale.

On a considéré jusqu'à présent le cas le plus simple. Quand le processus comporte un terme constant ou un trend, il faut modifier ces statistiques. On aura alors les statistiques  $Z_\mu^{PP}$  et  $Z_\tau^{PP}$  par analogie avec les statistiques de Dickey et Fuller

$$\begin{aligned} Z^{PP} &= T(\hat{\rho} - 1) - \frac{T^2}{2} \frac{s_\infty^2 - s_\epsilon^2}{\sum y_{t-1}^2} \\ Z_\mu^{PP} &= T(\hat{\rho} - 1) - \frac{T^2}{2} \frac{s_\infty^2 - s_\epsilon^2}{\sum (y_{t-1} - \bar{y})^2} \\ Z_\tau^{PP} &= T(\hat{\rho} - 1) - \frac{T^6}{24} \frac{s_\infty^2 - s_\epsilon^2}{|X'X|} \end{aligned}$$

où  $X$  est la matrice d'observations  $[1, t, y_{t-1}]$ . Ces tests sont plus puissants que le test ADF car ils ne nécessitent pas l'ajout de régresseurs supplémentaires en grand nombre quand les résidus ont une composante MA. Cependant, ils souffrent d'une distortion de taille si les résidus ont une composante MA négative. Perron and Ng (1996) ont suggéré de modifier ces tests en utilisant

$$MZ^{PP} = Z^{PP} + \frac{T}{2}(\hat{\rho} - 1)^2.$$

Ils montrent également que ces tests peuvent être améliorés en utilisant un estimateur paramétrique de la densité spectrale en zéro à la place de l'estimateur non-paramétrique de Newey and West (1987). Plus précisément, considérons la régression

$$\Delta y_t = \rho y_{t-1} + \sum_{j=1}^l b_j \Delta y_{t-j} + \epsilon_t$$



et extrayons les estimations des  $b_j$ . Un estimateur paramétrique de la densité spectrale en zéro est donnée par:

$$s_\infty^2 = \frac{s^2 \epsilon}{(1 - \sum \hat{b}_j)^2}.$$

## 5 Tests de l'hypothèse de stationarité

On a considéré jusqu'ici des tests où l'hypothèse nulle était la racine unitaire et la non-stationnarité. Dans le modèle  $y_t = \rho y_{t-1} + \epsilon_t$ , on testait  $H_0 : \rho = 1$ . L'alternative était la stationnarité. Supposons maintenant que l'on parte du modèle

$$y_t = \mu + u_t$$

où  $u_t$  est un processus stationnaire.  $y_t$  est donc  $I(0)$ . Si l'on différencie le processus, on obtiendra

$$\Delta y_t = u_t - \theta u_{t-1}$$

avec  $\theta = 1$ . Dans ce modèle ci, l'hypothèse nulle devient  $\theta = 1$  et elle est équivalente à la stationnarité. Par contre si  $\theta \neq 1$ , alors  $\epsilon_t = u_t - \theta u_{t-1}$  est stationnaire, et par conséquence le niveau de  $y_t$  s'obtient par une accumulation de bruits blancs stationnaires,  $y_t = \sum \epsilon_t$ , ce qui est la définition d'une série intégrée d'ordre 1. Le test de l'hypothèse nulle  $H_0 : \theta = 1$  contre l'alternative  $H_1 : \theta \neq 1$  permettra de tester cette fois-ci la stationnarité de la série. Il y a donc une symétrie entre d'un côté *racine unité dans la partie AR et non stationnarité* et de l'autre côté *racine unité dans la partie MA et stationnarité*.

Il n'est toutefois pas commode de tester directement la présence d'une racine unité dans un processus MA. En effet, même si la vraie valeur de  $\theta$  n'est pas 1, l'estimateur du maximum de vraisemblance aura tendance à être proche de 1 par un effet dit de *pile-up*. Ceci vient du fait que les deux MA(1) suivants ont les mêmes autocorrélation :  $y_t = u_t + \theta u_{t-1}$  et  $x_t = u_t + 1/\theta u_{t-1}$ . Il faut donc trouver une autre formulation du problème.

Considérons le modèle suivant:

$$\begin{aligned} y_t &= \mu + \delta t + z_t + u_t & \text{Var}(u_t) &= \sigma_u^2 \\ z_t &= x_{t-1} + v_t & \text{Var}(v_t) &= \sigma_v^2. \end{aligned}$$

La variable  $y_t$  y est décrite sous la forme de la somme d'une composante déterministe comportant un trend  $\delta t$  et d'une composante stochastique  $z_t$  qui est un trend stochastique avec comme valeur initiale  $z_0 = 0$ . Si dans ce modèle on pose  $\sigma_v^2 = 0$ , alors le trend stochastique se réduit à sa valeur initiale qui est zéro. Ce modèle admet comme forme réduite un processus MA(1). Effet, appliquons l'opérateur  $\Delta$  aux deux membres de la première équation et remplaçons  $z_t$  par sa valeur. Il vient:

$$\Delta y_t = \delta + u_t - u_{t-1} + v_t = \delta + \epsilon_t + \theta \epsilon_{t-1}$$

Tester que  $\theta = 1$  est équivalent à tester  $\sigma_v^2 = 0$ . On va donc chercher une statistique de test où sous  $H_0$  on aura  $\sigma_v^2 / \sigma_u^2$ . On va utiliser ce rapport pour éliminer le paramètre de nuisance  $\sigma_u^2$ .

Kwiatkowski, Phillips, Schmidt et Shin (1992) ont proposé une statistique de test particulière pour tester  $H_0 : \sigma_v^2 = 0$  contre  $H_1 : \sigma_v^2 > 0$ . C'est un test unilatéral. Considérons tout d'abord la régression auxiliaire

$$y_t = \mu + \delta t + e_t$$

dont on va tirer les résidus estimés  $\hat{e}_t$ . On retire donc les composantes déterministes du processus. Si  $y_t$  ne comporte pas de trend stochastique, les  $\hat{e}_t$  seront stationnaires. Par contre si  $y_t$  n'est pas stationnaire, il y aura une racine unité dans les  $\hat{e}_t$ .

Définissons les sommes partielles

$$S_t = \sum_{j=1}^t \hat{e}_j.$$

Si les  $\hat{e}_j$  sont stationnaires, alors  $S_t$  est par définition un processus  $I(1)$ . On sait du chapitre précédent que

$$\frac{1}{T^2} \sum S_t^2 \xrightarrow{\mathcal{L}} \sigma^2 \int_0^1 W^2(r) dr$$

dans le cas simplifié où  $y_t = e_t$ , c'est à dire sans terme déterministe. Par contre si les  $\hat{e}_j$  étaient  $I(1)$ , alors  $\frac{1}{T^2} \sum S_t^2$  divergerait. On a donc ici une statistique intéressante dans la mesure où son comportement diffère complètement sous  $H_0$  et  $H_1$ . Il suffit de se débarrasser du paramètre de nuisance  $\sigma^2$  en la normalisant correctement. Une première statistique de test est donc

$$\frac{1}{T^2} \frac{\sum S_t^2}{\hat{\sigma}^2}$$

où  $\hat{\sigma}^2$  est un estimateur de la variance des résidus  $e_t$ . La distribution asymptotique de ce test repose sur le fait que les  $e_t$  sont IID. S'il sont dépendants, il suffit de remplacer  $\hat{\sigma}^2$  par un estimateur de la variance de long terme et l'on a alors la statistique proposé par KPSS:

$$KPSS = \frac{1}{T^2} \frac{\sum S_t^2}{\hat{\sigma}_\infty^2}.$$

La variance de long terme peut être estimée en utilisant l'estimateur non-paramétrique de Newey-West:

$$s_\infty^2 = \frac{1}{T} \sum \hat{e}_t^2 + 2 \sum_{j=1}^l \left(1 - \frac{j}{l+1}\right) \frac{1}{T} \sum_{t=j+1}^T \hat{e}_t \hat{e}_{t-j}.$$

La distribution de la statistique KPSS sous  $H_0$  dépend de la présence de termes déterministes dans la régression initiale. Les valeurs critiques sont données dans la Table 3.

Il est intéressant de remarquer que le test KPSS est un cas particulier d'un test proposé antérieurement par Nabeya and Tanaka (1988). Ces auteurs cherchaient à déterminer si une régression linéaire a ses coefficients constants ( $H_0$ ) contre l'alternative que ces coefficients suivent une marche aléatoire ( $H_1$ ) dans le modèle

$$y_t = \beta_t x_t + u_t$$

$$\beta_t = \beta_{t-1} + v_t$$

Table 3: Valeurs critique pour le test KPPS

Niveau	0.10	0.05	0.01
Constante	0.347	0.463	0.739
Constante et trend	0.119	0.146	0.216

On a la constance de  $\beta_t$  si la variance de  $v_t$  est nulle. Le test de Nabeya and Tanaka (1988) était basé sur une estimation simple de la variance de  $u_t$  et donc ne tenait pas compte d'une possible auto-corrélation des erreurs.

## 6 Ruptures de Trend et Tests de Perron

La littérature empirique sur les tests de racine unitaire est devenue assez importante au fil des années. Elle est principalement d'origine américaine. On a vu la préoccupation de pouvoir séparer entre trend déterministe et trend stochastique et l'intérêt que cela pouvait avoir pour l'analyse conjoncturelle (business cycle). Le papier au retentissement important de Nelson and Plosser (1982) a montré que la plupart des séries macro-économiques américaines avaient une racine unitaire. L'hypothèse de racine unitaire disqualifie le point de vue selon lequel les fluctuations conjoncturelles sont de nature transitoire autour d'un tendance plus ou moins stable.

Le papier de Perron (1989) va à l'encontre de cette littérature dans la mesure où il montre que la plupart des séries macro-économiques américaines ne présentent pas de racine unitaire et que les fluctuations y sont de nature transitoire. Seuls deux événements ont une influence permanente sur les différentes séries et ce sont la crise de 1929 et le choc pétrolier de 1973. Le postulat sur lequel repose cette conclusion, c'est que ces deux chocs sont exogènes. Cette hypothèse sert à retirer l'influence de ces deux chocs de la partie aléatoire des séries. Encore une fois, nous verrons que la distribution des tests DF et ADF varie en fonction des éléments constituant le trend déterministe.

### 6.1 Modélisation d'une rupture de trend

Rappelons encore une fois que le modèle qu'il est très commode de considérer pour comprendre le fonctionnement des tests de racine unitaire est le suivant:

$$\begin{cases} y_t = \mu + \delta t + u_t \\ A(L)u_t = \epsilon_t \end{cases} \quad (38)$$

En factorisant le polynôme  $A(L)$  comme à l'acoutumée et en regroupant les deux équations du système on obtient une équation réduite dans laquelle il suffit de poser  $\rho = 1$  pour

obtenir une caractérisation de  $y_t$  sous l'hypothèse nulle.

On cherche à enrichir ce modèle de manière à pouvoir considérer des ruptures exogènes dans la partie déterministe de la série à partir d'une date inconnue que l'on appellera  $\zeta$ . Trois possibilités s'offrent à nous qui correspondent à trois cas empiriques précis. Une rupture dans le terme constant  $\mu$  va permettre de représenter un crash dans la série. On pense à la crise de 1929 au moins pour les Etat-Unis. Une rupture dans le trend va permettre de représenter un changement dans le taux de croissance de la série, ce qui est caractéristique de beaucoup de séries après le premier choc pétrolier de 1973. Enfin le dernier modèle possible combine ces deux possibilités. Introduisons donc pour commencer la fonction indicatrice  $ID(\zeta)$ :

$$\begin{cases} ID(\zeta) = 1 & \text{si } t > \zeta \\ ID(\zeta) = 0 & \text{autrement} \end{cases} \quad (39)$$

Les trois modèles que l'on veut considérer correspondent alors aux équations statiques suivantes:

$$y_t = \mu_1 + (\mu_2 - \mu_1)ID(\zeta) + \delta t + u_t \quad (40)$$

$$y_t = \mu + \delta_1 t + (\delta_2 - \delta_1)(t - \zeta)ID(\zeta) + u_t \quad (41)$$

$$y_t = \mu_1 + \delta_1 t + [(\mu_2 - \mu_1) + (\delta_2 - \delta_1)(t - \zeta)]ID(\zeta) + u_t \quad (42)$$

Il faut maintenant combiner ces trois modèles qui correspondent à l'hypothèse alternative avec le polynôme  $A(L)$  factorisé. Pour simplifier l'exposé, on prendra ce polynôme de degré 1 et égal à  $(1 - \rho L)$ . Nous allons effectuer ce calcul pour chacun des trois modèles. Mais tout d'abord, il est utile de remarquer quelques particularités de la fonction indicatrice  $ID(\zeta)$ . On a en premier lieu que

$$\begin{aligned} LID(\zeta) &= ID(\zeta + 1) \\ \Delta ID(\zeta) &= ID(\zeta) - ID(\zeta + 1) \end{aligned}$$

On remarque ensuite que  $\Delta ID(\zeta)$  correspond à une nouvelle fonction indicatrice

$$\begin{cases} \Delta ID(\zeta) = 1 & \text{si } t = \zeta + 1 \\ \Delta ID(\zeta) = 0 & \text{autrement} \end{cases} \quad (43)$$

qui modélise donc l'équivalent d'une variable dummy ponctuelle à la date  $t = \zeta + 1$ . Ce petit détour fait, on arrive au résultat  $(1 - \rho L)ID(\zeta) = (1 - \rho)\Delta ID(\zeta)$ , ce qui nous permet d'obtenir nos trois régressions de test:

$$\begin{aligned} y_t - \rho y_{t-1} &= (1 - \rho)(\mu_1 + (\mu_2 - \mu_1)ID(\zeta) + \delta t) \\ &\quad + \rho(\delta + (\mu_2 - \mu_1)\Delta ID(\zeta)) + \epsilon_t \end{aligned} \quad (44)$$

pour le modèle “crash”,

$$\begin{aligned} y_t - \rho y_{t-1} = & (1 - \rho)(\mu + \delta_1 t + (\delta_2 - \delta_1)(t - \zeta)ID(\zeta)) \\ & + \rho(\delta_1 + (\delta_2 - \delta_1)ID(\zeta)) + \epsilon_t \end{aligned} \quad (45)$$

pour le modèle à taux de croissance variable,

$$\begin{aligned} y_t - \rho y_{t-1} = & (1 - \rho)(\mu_1 + \delta_1 t + ((\mu_2 - \mu_1) + (\delta_2 - \delta_1)(t - \zeta))ID(\zeta)) \\ & + \rho(\delta_1 + (\delta_2 - \delta_1)ID(\zeta) + (\mu_2 - \mu_1)\Delta ID(\zeta)) + \epsilon_t \end{aligned} \quad (46)$$

pour le modèle combiné. En posant dans chacun de ces modèles  $\rho = 1$ , on trouve naturellement la forme de l’hypothèse nulle correspondante, qui autrement n’est pas évidente à expliquer:

$$\Delta y_t = \delta + (\mu_2 - \mu_1)\Delta ID(\zeta) + \epsilon_t \quad (47)$$

$$\Delta y_t = \delta_1 + (\delta_2 - \delta_1)ID(\zeta) + \epsilon_t \quad (48)$$

$$\Delta y_t = \delta_1 + (\delta_2 - \delta_1)ID(\zeta) + (\mu_2 - \mu_1)\Delta ID(\zeta) + \epsilon_t \quad (49)$$

Sous l’hypothèse nulle,  $y_t$  suit une marche aléatoire avec dérive. Cette dérive peut changer de trois façons différentes. Dans le premier modèle, la dérive fait un saut sur une seule observation correspondant à la date  $\zeta + 1$  et ce saut est mesuré par  $(\mu_2 - \mu_1)$ . Dans le deuxième modèle, la dérive change de façon permanente après la date  $\zeta$  et ce changement permanent est mesuré par  $(\delta_2 - \delta_1)$ . Dans le troisième modèle, nous avons un effet combiné de changement permanent dans la dérive après la date  $\zeta$  et d’une pointe transitoire à la date  $\zeta + 1$ . Quand le polynôme  $A(L)$  est de degré supérieur à 1, il suffit de rajouter à la régression de test un terme qui au maximum a la forme:

$$A^*(L)(\Delta y_t - \delta_1 - (\delta_2 - \delta_1)ID(\zeta)) \quad (50)$$

## 6.2 Motivation empirique

On ne s’intéressera ici qu’au deuxième modèle qui nous servira *in fine* à traiter une application sur données françaises trimestrielles. Perron (1989) motive son analyse de trend segmenté sur l’exemple du PIB réel américain sur la période 1947:1-1986:3. Celui-ci marque une rupture de tendance en 1973:1 (observation 105 de l’échantillon) quand on ajuste un modèle avec constante, trend et une variable dummy  $DT73$  qui vaut zéro avant cette date et ensuite  $(t - 105)$ . La régression est la suivante:

$$\begin{aligned} \log Y_{USA_t} = & \underset{[1149]}{6.98} + \underset{[96.7]}{0.0087t} - \underset{[-11.5]}{0.0030DT73} + u_t \\ R^2 = & 0.992 \quad DW = 0.12 \quad \hat{\sigma} = 0.0322 \end{aligned}$$

La série détrendée de cette façon présente une structure d’autocorrélation qui décroît très rapidement (voir la Table 1). Ce comportement n’est certainement pas celui associé avec

une marche aléatoire ou une marche aléatoire à laquelle on a retiré un trend. Il est par contre typique du comportement d'une série stationnaire. Mais d'autre part un test de Dickey-Fuller avec trend non segmenté ne permet pas de rejeter l'hypothèse nulle d'une racine unitaire, même au seuil de 10%, ce qui bien sûr pose un sérieux problème et motive l'article de Perron.

Nous allons refaire ce petit exercice sur le logarithme du PIB français pour la même période et les comparer aux calculs de Perron. On peut rapidement se rendre compte au

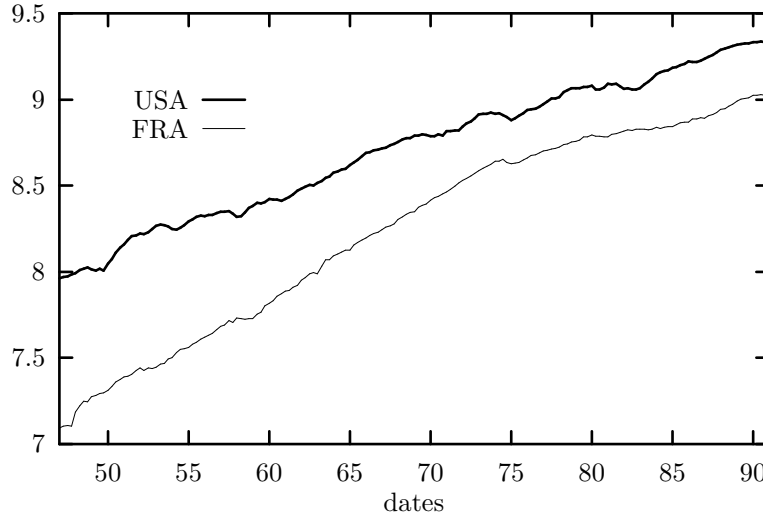


Figure 1: Logarithmes du PIB français et du GNP américain  
Les échelles ont été ajustées

moyen d'un graphique que le choc pétrolier de 1973 s'est traduit par une rupture de trend au premier trimestre de 1974. Par contre sur les données américaines la rupture s'est produite au premier trimestre de 1973, donc un an avant. Le modèle avec trend segmenté donne:

$$\log Y_{FR_t} = \underset{[2826.5]}{11.55} + \underset{[249.0]}{0.014t} - \underset{[-48.4]}{0.009DT'74} + u_t$$

$$R^2 = 0.999 \quad DW = 0.26 \quad \hat{\sigma} = 0.0210$$

On peut constater sur la Table 4 les similitudes entre les auto-corrélations des résidus de cette régression et les autocorrélation des résidus d'une régression effectuée sur le PIB américain. Un test de Dickey-Fuller augmenté traditionnel ne permet pas pourtant dans un cas comme dans l'autre de rejeter l'hypothèse nulle de racine unitaire:

$$\Delta \log Y_{FR_t} = \underset{[-0.308]}{-0.0386} - \underset{[-1.064]}{0.000138t} + \underset{[0.453]}{0.00487 \log Y_{FR_{t-1}}}$$

$$\hat{\sigma} = 0.0110 \quad \tau_\tau = 0.453$$

$$\Delta \log Y_{USA_t} = \underset{[2.90]}{0.386} + \underset{[2.71]}{0.0004t} - \underset{[-2.54]}{0.054 \log Y_{USA_{t-1}}}$$

$$\hat{\sigma} = 0.0100 \quad \tau_\tau = -2.54$$

Table 4: Autocorrélation des PIB stationnarisés

$\gamma(h)$	$\hat{u}_{FR}$	$\hat{u}_{USA}$
1	0.87	0.94
2	0.74	0.83
3	0.61	0.70
4	0.50	0.57
5	0.41	0.45
6	0.33	0.35

car les valeurs critiques asymptotiques pour le test ADF sont  $-3.41$  à  $5\%$  et  $-3.12$  à  $10\%$ . La prise en compte de la rupture du trend peut être déterminante.

### 6.3 Expérience de Monte Carlo

Pour bien voir ce qu'il se passe quand on autorise un changement dans la dérive d'une marche aléatoire, Perron fait une petite expérience de Monte Carlo qui consiste à générer 10.000 échantillons de 100 observations selon le modèle:

$$y_t = \mu + \beta_1 t + (\beta_2 - \beta_1)(t - \zeta)ID(\zeta) + \epsilon_t \quad (51)$$

Les paramètres ont été fixés à  $\mu = 0$ ,  $\beta_1 = 1$ ,  $\zeta = 50$  et  $\epsilon_t \sim N(0, 1)$ . On va ensuite ignorer la rupture de trend et estimer l'équation d'un test DF:

$$y_t = \tilde{\mu} + \tilde{\beta}t + \rho y_{t-1} + \tilde{\epsilon}_t \quad (52)$$

L'expérience de Monte Carlo sera conduite en faisant varier  $\beta_2$ , c'est à dire l'importance de la rupture dans le trend. Quand  $\beta_2 = 1$ , il n'y a pas de rupture de trend. La distribution empirique cumulée de  $\hat{\rho}$  est centrée autour de 0. Les données sont générées par un modèle trend stationnaire pur et les résultats empiriques reproduisent bien cela. Quand on baisse la valeur de  $\beta_2$  vers zéro, on augmente la différence de pente dans le trend entre les deux sous périodes. La distribution empirique cumulée de  $\hat{\rho}$  se concentre alors de plus en plus vers un, ce qui veut dire qu'empiriquement on se dirige vers un modèle de marche aléatoire. Ceci peut se voir dans la Table 5. Perron (1989) donne en outre le théorème suivant qui

Table 5: Comportement de  $\hat{\rho}$  en présence d'une rupture de trend

	$\beta_1 = 1, \mu = 0$				
	$\beta_2 = 1.0$	$\beta_2 = 0.9$	$\beta_2 = 0.7$	$\beta_2 = 0.4$	$\beta_2 = 0.0$
moyenne	-0.019	0.334	0.825	0.949	0.981
écart-type	0.0314	0.0969	0.0307	0.0095	0.0032

complète bien ces résultats empiriques :

**Théorème 1 :** *Soit un échantillon de taille  $T + 1$  de la variable  $y_t$  générée par un modèle auto-régressif stationnaire avec un trend comportant une rupture et des erreurs indépendantes normales. Soit  $\zeta = \lambda T$  avec  $\lambda \in [0, 1]$ . Alors  $\hat{\rho}$  estimé dans un modèle auto-régressif avec trend sans rupture converge en probabilité vers 1 et:*

$$T(\hat{\rho} - 1) \xrightarrow{p} \frac{3(-1 + 4\lambda - 5\lambda^2 + 2\lambda^3)}{2(-3 + 4\lambda - 3\lambda^2 + 3\lambda^3 - 4\lambda^4)}$$

Ce théorème est établi dans Perron (1989) sous des conditions relativement faibles de mélange portant sur les  $\epsilon_t$ . La normalité et l'indépendance sont des conditions plus fortes, mais plus simples à énoncer. Ce théorème montre que  $\hat{\rho}$  converge vers 1 asymptotiquement, et que cela ne dépend pas de  $\beta_2$ . Le second résultat de convergence montre que le biais normalisé  $T(\hat{\rho} - 1)$  converge vers une quantité finie qui varie entre zéro et 1/2 en fonction de  $\lambda$ .  $T(\hat{\rho} - 1)$  représente aussi la statistique de test  $z$ . Sa valeur critique asymptotique est de -21.8 à 5%, ce qui fait qu'avec une valeur limite comprise entre 0 et 1/2 ce test ne rejettera jamais l'hypothèse nulle de racine unitaire.

On peut donc conclure que le fait de négliger une rupture dans le trend va conduire à ne jamais rejeter l'hypothèse nulle de racine unitaire. Il faut donc proposer un nouveau type de test qui tienne compte de cette possibilité.

## 6.4 Test avec trend segmenté

On va conduire ce test un peu comme l'on avait conduit les précédents. Il suffit d'enrichir de manière adéquate la composante déterministe  $TD_t$ . On a vu qu'il existait deux façons équivalentes de calculer un test DF ou ADF. Soit on utilise une régression en une étape, soit on retire tout d'abord le trend et on fait le test sur les résidus estimés, utilisant de ce fait une régression en deux étapes. Ces deux procédures sont asymptotiquement équivalentes tant que le trend est linéaire. Ici ce n'est plus le cas. Perron (1989) ne détaille qu'une méthode en deux étapes. Il calcule donc:

$$\hat{u}_t = y_t - T\hat{D}_t \text{ avec } T\hat{D}_t = \hat{\mu} + \hat{\delta}_0 t + \hat{\delta}_1(t - \zeta)ID(\zeta) \quad (53)$$

estimé par moindres carrés. Sur ces résidus estimés, il calcule un test DF ou ADF selon la nécessité:

$$\Delta \hat{u}_t = (\rho - 1)\hat{u}_{t-1} + A^*(L)\Delta \hat{u}_t + \epsilon_t \quad (54)$$

Les deux tests DF et ADF ont encore une fois la même distribution asymptotique sous des conditions pas très restrictives. Notons  $\tau_\pi$  ce nouveau test (statistique de Student associée au coefficient de  $\hat{u}_{t-1}$ ). Sous l'hypothèse que les erreurs  $\epsilon_t$  soient IID (hypothèse plus restrictive que celles données dans l'article de Perron), on obtient la distribution asymptotique de ce nouveau test:

**Théorème 2 :** *Soit la série  $\{y_t\}$  générée sous l'hypothèse nulle d'une marche aléatoire d'innovations indépendantes identiquement distribuées, avec une dérive qui change au point  $\zeta = \lambda T$ . Alors pour  $T \rightarrow \infty$  la statistique  $\tau_\pi$  converge en distribution vers une fonctionnelle de processus de Wiener normalisés avec:*

$$\tau_\pi \xrightarrow{\mathcal{L}} H_\pi(\lambda) / [3\lambda^3 K_\pi(\lambda)]^{1/2}$$



Les expressions analytiques de  $H_\pi(\lambda)$  et  $K_\pi(\lambda)$  sont relativement complexes et données dans l'article de Perron. On remarque que cette distribution dépend de la valeur de  $\lambda$ . D'autre part on peut montrer que, pour les deux cas extrêmes où  $\lambda = 0$  ou  $\lambda = 1$ , l'on retombe sur la distribution de  $\tau_\tau$ . On a extrait de Perron (1989) la Table 6 qui donne les valeurs critiques asymptotiques de ce test. On peut remarquer tout d'abord que ces valeurs

Table 6: Quantiles de la distribution asymptotique de  $\tau_\pi$

	rupture relative en $\lambda$								
$\lambda =$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1%	-4.27	-4.41	-4.51	-4.55	-4.56	-4.57	-4.51	-4.38	-4.26
5%	-3.65	-3.80	-3.87	-3.94	-3.96	-3.95	-3.85	-3.82	-3.68
10%	-3.36	-3.49	-3.58	-3.66	-3.68	-3.66	-3.57	-3.50	-3.55

critiques sont en valeurs absolues plus grandes que celles du test  $\tau_\tau$  qui rappelons le sont respectivement de -3.96, -3.41, -3.13 pour les seuils de 1%, 5% et 10%. Ceci est une illustration de la perte en puissance de ces tests chaque fois que l'on doit estimer un trend plus complexe. Si les valeurs des paramètres qui rentrent dans  $TD_t$  étaient connues, alors la distribution asymptotique de ces tests serait la même et égale à celle du test  $\tau$  simple dans lequel  $TD_t = 0$ . Les valeurs critiques ensuite dépendent de  $\lambda$  mais ne sont pas très sensibles à sa valeur. On peut noter que la valeur critique est maximale en valeur absolue pour  $\lambda = 0.5$ . On se rapproche des valeurs critiques de  $\tau_\tau$  près des bornes.

## 6.5 Application sur données françaises

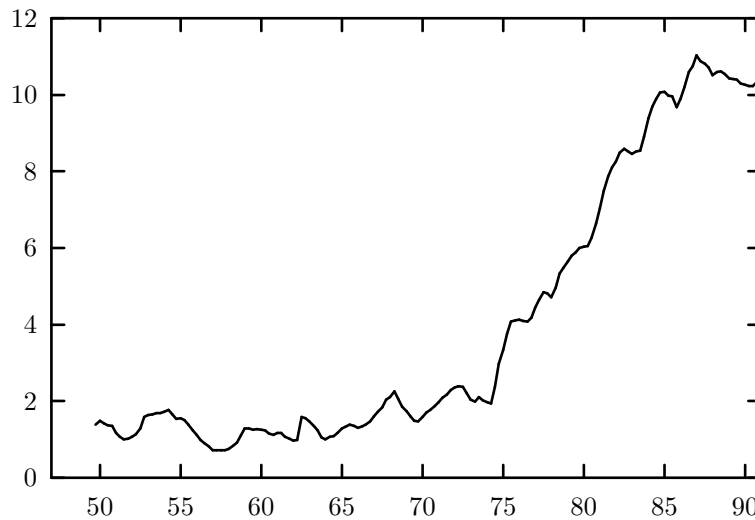


Figure 2: Taux de chômage en France

L'étude de la rupture du trend en 1974:1 est très instructive pour comprendre l'évolution de la croissance et du chômage en France et ses différences par rapport aux USA. La rupture de 1974 est très marquée sur beaucoup de variables contenues dans la banque de données trimestrielle de l'INSEE établie par Laroque, Ralle, Salanié, and Toujas-Bernate (1990) sur la période 1946:1-1989:4 (étendue ensuite jusqu'en 1991:4). On a choisit d'étudier les deux variables taux de chômage et PIB qui sont traditionnellement reliées par la "loi d'Okun" dont une formulation simplifiée est:

$$\Delta ur = \alpha_0 - \alpha_1 \Delta \log Y$$

Cette relation implique que l'on ait des caractéristiques communes entre ces variables. Or il est communément admis aux USA que le taux de chômage est stationnaire et  $Y$  réel intégré d'ordre un, ce qui, soit dit en passant, invalide dès le départ une telle relation comme on pourra le montrer dans les chapitres sur la cointégration. Perron (1989) suit la tradition américaine et dans son application sur les données de Nelson and Plosser (1982) et néglige de faire un test sur le taux de chômage, le prenant comme stationnaire. On ne rencontre pas du tout cette caractéristique sur le taux de chômage en Europe. On commencera donc par effectuer un test  $\tau_\pi$  sur les séries de taux de chômage. Commençons par estimer une régression reliant le logarithme du taux de chômage à un trend et sa rupture en 1974:1. C'est la première étape d'un test  $\tau_\pi$ . Le fait d'avoir pris la variable dépendante en logarithme nous permet d'interpréter les coefficients de la régression comme des taux de croissance:

$$\log ur_t = \underset{[-2.66]}{-0.156} + \underset{[11.80]}{0.0090}t + \underset{[11.71]}{0.022}DT74 + u_t$$

$$R^2 = 0.906 \quad DW = 0.08 \quad \hat{\sigma} = 0.264$$

On a 184 observations. La rupture de tendance est à l'observation 113, ce qui fait que  $DT74 = (t - 113)\mathbf{I}(t > 113)$ . Le test  $\tau_\pi$  sur les résidus montre qu'il existe une racine unitaire encore après avoir retiré le trend segmenté car:

$$\Delta \hat{u}_t = \underset{[-3.08]}{-0.058} \hat{u}_{t-1} + \underset{[7.82]}{0.518} \Delta \hat{u}_{t-1}$$

la valeur de  $\tau_\pi$  est en valeur absolue plus faible que celle de la table pour  $\lambda = 0.64$  et un seuil de 5% ou 10%.

Passons maintenant à la série du PIB en volume. La régression de première étape donne:

$$\log Y_t = \underset{[2881.5]}{11.54} + \underset{[255.5]}{0.0140}t - \underset{[-58.43]}{0.0087}DT74 + u_t$$

$$R^2 = 0.999 \quad DW = 0.26 \quad \hat{\sigma} = 0.0221$$

La régression de deuxième étape donne:

$$\Delta \hat{u}_t = \underset{[-3.86]}{-0.145} \hat{u}_{t-1} + \underset{[1.44]}{0.101} \Delta \hat{u}_{t-1}$$

ce qui fait que l'hypothèse de racine unitaire est rejetée au seuil de 10% mais pas à celui de 5%.

Ce qui est intéressant dans ces résultats c'est qu'il semble que l'on ait deux types de non stationnarité pour le PIB et le taux de chômage. Il y a une même rupture, au même moment, mais la croissance de la série n'est pas la même. On remarque aussi qu'avec un taux de croissance trimestriel de 1.4% on avait un taux de croissance correspondant du chômage de 0.8%. Le choc pétrolier a fait baisser le taux de croissance de 0.9% alors que dans le même temps le taux de croissance du chômage augmentait de 2.2%, ce qui correspond à une sur-réaction très nette. On peut donc dire que la relation entre le taux de chômage et la croissance a subi un changement structurel qu'il s'agirait d'expliquer. Une partie de la croissance du chômage semble autonome, car son trend est de nature différente et son changement de pente plus fort. Pour estimer une relation d'Okun, il faudrait donc inclure au moins une variable supplémentaire.

## 7 Saisonnalité et Racines Unitaires

Nous avons jusqu'à présent considéré le test de la présence d'une racine unité à l'ordre un quelle que soit la périodicité d'observation des séries. Mais dès lors que l'on considère des séries d'une périodicité autre qu'annuelle, il peut apparaître dans la série un cycle dit saisonnier vient perturber la simplicité des démarches précédentes et conduit à vouloir la généraliser. La première question consiste à s'interroger sur la façon dont on va tenir compte de la saisonnalité et comment celle-ci peut influencer les tests usuels. La deuxième question est très fort liée à la façon dont on peut modéliser la saisonnalité. On verra, en suivant la tradition initiée par Box et Jenkins, que la saisonnalité peut se traduire dans une série de périodicité  $s$  par une racine unitaire à l'ordre  $s$ , venant s'ajuster à une possible racine à l'ordre 1. La question d'un test concerne donc maintenant la présence non plus d'une , mais de plusieurs racines.

Cette section est organisée comme suit. On va tout d'abord indiquer comment se modélise la saisonnalité. Ensuite on regardera comment la saisonnalité influence la construction d'un test d'une seule racine unitaire. On traitera ensuite du cas de deux racines pour finir avec le test de racines aux différentes fréquences saisonnières.

### 7.1 Modélisation de la saisonnalité

Pour tenir compte de la possible saisonnalité d'une série de périodicité  $s$ , il faut se donner un modèle représentatif. La première façon d'introduire une composante saisonnière se situe au niveau déterministe. A la composante  $TD_t$ , on va rajouter  $SD_t$  ce qui donne:

$$\begin{cases} y_t &= TD_t + SD_t + u_t \\ A(L)u_t &= B(L)\epsilon_t \end{cases} \quad (55)$$

La composante saisonnière prend la plupart du temps une forme simple. Si les données sont trimestrielles, ce qui correspond à  $s = 4$ , on aura:

$$SD_t = \iota_1 d_1 + \iota_2 d_2 + \iota_3 d_3 + \iota_4 d_4 \quad (56)$$

$d_i$  est une variable muette saisonnière qui vaut un pour le trimestre  $i$  et zéro autrement. Notons alors que  $TD_t$  ne doit pas comprendre de terme constant car la somme des  $d_i$  est égale à un. D'autres solutions plus complexes existent pour modéliser de manière déterministe la saisonnalité en utilisant notamment des fonctions périodiques. On pourra regarder avec profit Granger and Newbold (1986) section 1.10. Ce type de modélisation implique que les mouvements saisonniers ne changent pas dans le temps. Ils se reproduisent à l'identique. Si la série comporte un trend, les constantes saisonnières vont simplement déplacer de manière régulière l'ordonnée à l'origine de ce trend. C'est à dire en fait comme si l'on avait quatre trends parallèles.

Le second type de saisonnalité que l'on peut considérer est une saisonnalité dont les mouvements évoluent dans le temps. On introduit alors la modélisation de la saisonnalité au niveau de  $u_t$  en suivant les procédures proposées par Box and Jenkins (1976). Cette démarche est complémentaire de la précédente. On peut supposer qu'il reste un mouvement saisonnier dans  $u_t$ . On va alors enlever cette saisonnalité au moyen d'un filtre. Si l'on appelle  $v_t$  la série résultante, le filtre dans le cas le plus simple dans le cas de séries trimestrielles s'écrit:

$$A_s^*(L^4)(1 - L^4)u_t = B_s(L^4)v_t \quad (57)$$

ce qui signifie que l'on a différencié la série à l'ordre quatre et utilisé des polynômes de retard d'ordre multiple de quatre. Une fois ce filtre appliqué, on peut utiliser une modélisation ARIMA sur  $v_t$  qui sera par exemple:

$$A_{ns}^*(L)(1 - L)v_t = B_{ns}(L)\epsilon_t \quad (58)$$

En combinant les deux modèles, on obtient pour  $u_t$ :

$$A_s^*(L^4)A_{ns}^*(L)(1 - L^4)(1 - L)u_t = B_s(L^4)B_{ns}(L)\epsilon_t \quad (59)$$

Cette écriture est connue dans la littérature comme modèle SARIMA pour ARMA saisonnier sur données intégrés. Pour notre propos on peut simplifier cette notation en confondant les polynômes saisonniers (indexés par  $s$ ) et non saisonniers (indexés par  $ns$ ):

$$A^*(L)(1 - L^4)(1 - L)u_t = B(L)\epsilon_t \quad (60)$$

On remarque que le filtre repose sur deux différenciations, une à l'ordre quatre pour la saisonnalité, l'autre à l'ordre un pour le trend. Cette écriture impose la présence de deux racines unitaires. Si au lieu d'imposer ces racines, on veut tester leur présence effective, le type de test à employé sera différent des tests usuels, à cause cette présence de multiples racines.

## 7.2 Test de la présence d'une seule racine

On a distingué entre une saisonnalité déterministe et une saisonnalité stochastique. Comment va se poser le test de racine unitaire dans ce contexte? Raisonnons dans le cas d'une saisonnalité stochastique évoluant dans le temps. Dans le cas extrême où le modèle complet se réduit à:

$$(1 - L^4)(y_t - \mu - \delta t) = \epsilon_t \quad (61)$$

ce qui implique:

$$\Delta_4 y_t = 4\delta + \epsilon_t \quad (62)$$

on a le mélange de quatre marches aléatoires (pour  $t, t-1, t-2, t-3$ ) avec la même dérive. Ceci implique que la variable  $y_t$  a un taux de croissance saisonnier égal en moyenne à  $4\delta$ . Cette volatilité au cours du temps dans le mouvement saisonnier n'est pas très réaliste. La période de Noël se traduit toujours par un boum de la consommation et jamais par une chute. Aussi on ne va pas désirer reproduire ici le débat précédent entre trend stationnaire et trend déterministe. D'autant que bien souvent les données dont on dispose sont désaisonnalisées, et qu'ensuite la théorie économique considère la saisonnalité comme un bruit dont il faut se débarrasser. On s'intéressera donc à deux choses: premièrement comment traiter le test de racine unitaire à l'ordre un en présence de saisonnalité déterministe et deuxièmement faut-il ou non différencier à l'ordre  $s$  une série saisonnière.

Considérons tout d'abord le cas simple d'une série trimestrielle dans laquelle on veut faire un test de racine unitaire à l'ordre 1. Il s'agit alors de tenir compte de la possible saisonnalité au moyen de composantes déterministes en supposant qu'il n'y a pas de saisonnalité stochastique. Comme  $SD_t$  est du même ordre que le terme constant dans  $TD_t$ , le test de Dickey et Fuller unitaire ne va pas être changé. En particulier les tables habituelles pourront continuer à être utilisées comme l'ont montré Dickey, Bell, and Miller (1986).

Considérons maintenant le cas plus spécifique de la présence d'une racine unitaire dans la partie saisonnière de la série. Le modèle le plus général examiné par Dickey, Hasza, and Fuller (1984) est le suivant:

$$A_s(L^s)(y_t - \sum \mu_i d_{it}) = \epsilon_t \quad (63)$$

Dans cette équation,  $d_{it}$  représente les constantes saisonnières correspondant à chaque trimestre. Pour effectuer un test de racine unitaire dans ce modèle, on commence donc par "purger" la série de ses composantes déterministes par une régression de première étape. Pour pouvoir effectuer la régression de deuxième étape, il est utile de factoriser le polynôme  $A_s(L^s)$  de manière à mettre en évidence une possible racine unitaire. Pour  $s = 4$ , cela donne:

$$A_s(L^4) = (1 - \phi L^4) - (1 - L^4)A_s^*(L^4) \quad (64)$$

Ce qui fait que la régression de test est la suivante:

$$\Delta_4 \hat{u}_t = (\phi - 1)\hat{u}_{t-4} + A_s^*(L^4)\Delta_4 \hat{u}_t + \epsilon_t \quad (65)$$

avec  $\Delta_4 = (1 - L^4)$ . Le test porte sur la nullité du coefficient de régression de  $u_{t-4}$ . Notons la structure particulière de cette régression de test qui ne comporte que des retard à l'ordre  $s$  ou un multiple de  $s$ .

On trouve les tables adéquates par exemple dans Dickey, Hasza, and Fuller (1984) pour le cas sans constante, le cas avec une constante et le cas avec constantes saisonnières (mais pas pour toutes les configurations possibles). On va désigner les statistiques de test correspondantes par  $\tau_s$ ,  $\tau_{s\mu}$  et  $\tau_{sd}$ . La Table 7 a été calculée au moyen d'une expérience de Monte

Carlo basée sur 10 000 tirages. Elle donne des valeurs comparables à celles contenues dans l'article de Dickey, Hasza, and Fuller (1984) à des variations d'échantillonnage près. Elle couvre les cas des données trimestrielles ( $s = 4$ ) et mensuelles ( $s = 12$ ) avec les trois types de termes déterministes. L'examen de cette table montre certaines caractéristiques du test

Table 7: Distribution de  $\tau_s$  sous  $H_0 : \Delta_s y_t \sim N(0, 1)$

statistique						
	$\tau_s$		$\tau_{s\mu}$		$\tau_{sd}$	
	probabilité dans la queue de gauche					
T	5%	10%	5%	10%	5%	10%
$d = 4$						
40	-1.87	-1.48	-2.36	-1.98	-4.21	-3.85
80	-1.85	-1.49	-2.38	-2.01	-4.14	-3.79
200	-1.88	-1.53	-2.35	-1.99	-4.09	-3.77
$d = 12$						
120	-1.76	-1.40	-1.99	-1.63	-5.88	-5.50
180	-1.75	-1.39	-2.00	-1.62	-5.83	-5.49
240	-1.74	-1.38	-1.99	-1.63	-5.80	-5.46

de la présence d'une racine unitaire à l'ordre  $s$  quand on la compare à une table de Dickey et Fuller usuelle. Tout d'abord un test à l'ordre  $s$  est plus puissant qu'un test à l'ordre 1 tant qu'il n'y a pas de constantes saisonnières. Cette puissance augmente avec la périodicité. Ensuite, le fait de rajouter des constantes saisonnières fait chuter la puissance du test de manière importante, à l'inverse de ce qui se passe avec un test DF à l'ordre 1. Ceci tient au fait que l'on met en concurrence deux méthodes de traitement de la saisonnalité, les constantes saisonnières ou la différenciation à l'ordre  $s$ .

### 7.3 Test de la présence de deux racines

Le test de Dickey, Hasza, and Fuller (1984) comporte une certaine ambiguïté dans la mesure où l'on a toujours:

$$\Delta_s y_t = \Delta y_t + \Delta y_{t-1} + \cdots + \Delta y_{t-s+1} \quad (66)$$

Ce test peut donc très bien tester la présence simplement d'une racine unitaire à l'ordre 1 pour peu que le nombre d'observations soit un multiple de  $s$ . On lève cette ambiguïté dès lors que l'on considère explicitement la présence possible de deux racines, une à l'ordre 1 et l'autre à l'ordre  $s$ . En outre le type de test joint que l'on va considérer est particulièrement commode quand on envisage la modélisation SARIMA d'une série saisonnière car il permet de répondre de manière satisfaisante à la question de l'ordre de différenciation à faire subir à la série. Cette problématique est exposée dans Barthelemy and Lubrano (1996). Considérons donc le modèle:

$$A(L)A_s(L^s)(y_t - TD_t - SD_t) = \epsilon_t \quad (67)$$

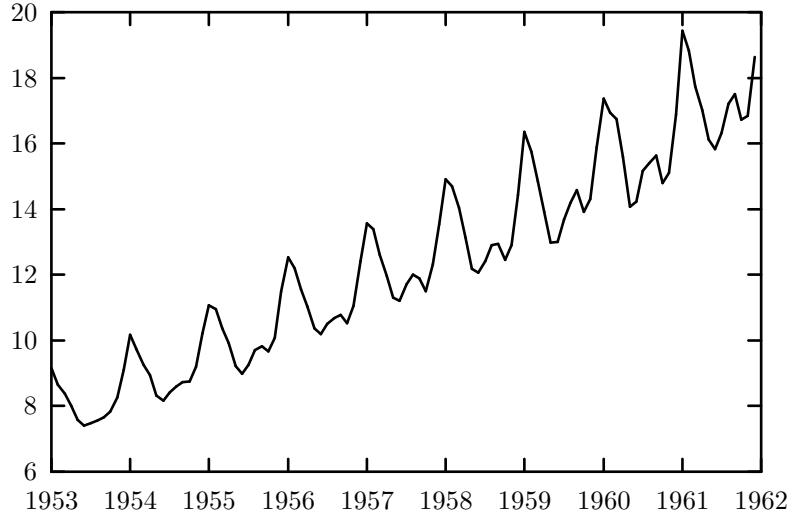


Figure 3: Consommation électrique aux USA

et les factorisations suivantes de ces deux polynômes:

$$\begin{aligned} A(L) &= (1 - \rho L) - (1 - L)A^*(L) \\ A_s(L^s) &= (1 - \phi L^s) - (1 - L^s)A_s^*(L^s) \end{aligned} \quad (68)$$

Le produit de ces deux factorisations donne:

$$\begin{aligned} A(L)A_s(L^s) &= (1 - \rho L)(1 - \phi L^s) - (1 - \rho L)(1 - L^s)A_s^*(L^s) \\ &\quad - (1 - \phi L^s)(1 - L)A^*(L) + (1 - L)(1 - L^s)A^*(L)A_s^*(L^s) \\ &= (1 - \rho L)(1 - \phi L^s) \\ &\quad + (1 - L)(1 - L^s) \left[ A^*(L)A_s^*(L^s) - \frac{1 - \phi L^s}{1 - L^s}A^*(L) - \frac{1 - \rho L}{1 - L}A_s^*(L^s) \right] \end{aligned}$$

Pour construire une régression de test, on a besoin d'approximer ce produit en:

$$A(L)A_s(L^s) \simeq (1 - \rho L)(1 - \phi L^s) + (1 - L)(1 - L^s)\bar{A}^*(L) \quad (69)$$

Cette approximation est exacte quand  $\rho$  et  $\phi$  sont proches de 1. Une fois que l'on a "purgé" la série de ses composantes déterministes par une régression de première étape, la régression de test se construit à partir de:

$$\hat{u}_t = \rho \hat{u}_{t-1} + \phi \hat{u}_{t-s} - \rho \phi \hat{u}_{t-s-1} + (1 - L)(1 - L^s)\psi^*(L)\hat{u}_t + \epsilon_t \quad (70)$$

En remarquant que  $(1 - L)(1 - L^s) = 1 - L - L^s + L^{s+1}$ , on peut transformer cette régression en:

$$\Delta_1 \Delta_4 \hat{u}_t = (\rho - 1)\hat{u}_{t-1} + (\phi - 1)\hat{u}_{t-4} - (\rho\phi - 1)\hat{u}_{t-5} + A_3^*(L)\Delta_1 \Delta_4 \hat{u}_t + \epsilon_t \quad (71)$$

La statistique de Student associée au coefficient de  $\hat{u}_{t-1}$  permet de tester une racine unitaire à l'ordre un; celle associée au coefficient de  $\hat{u}_{t-4}$  permet de tester une racine unitaire à l'ordre quatre. La double présence d'une racine unitaire à l'ordre un et à l'ordre quatre se teste au moyen d'un test en  $F$  sur la nullité jointe des coefficients associés à  $\hat{u}_{t-1}$  et  $\hat{u}_{t-4}$ . La Table 8 est extraite de Barthelemy and Lubrano (1996) et concerne le cas  $s = 4$ . On trouvera dans ce papier les tables pour le cas  $s = 12$ . Il est à noter que ces tables

Table 8: DGP sous  $H_0 : \Delta\Delta_4 y_t = \epsilon_t \sim N(0, 1)$ 

	Valeurs critiques			
	$t$ à 5%		$F$ à 95%	
	T	$t : \rho = 1$	$t : \phi = 1$	$F(2, T - k)$ $\rho = \phi = 1$
sans cste	60	-2.56	-1.94	4.60
sans var. sais.	100	-2.41	-1.94	4.20
sans trend	200	-2.26	-1.92	3.81
cste	60	-3.48	-2.06	7.39
sans var. sais.	100	-3.40	-2.00	6.99
sans trend	200	-3.26	-1.94	6.41
cste	60	-2.84	-3.99	12.48
var. sais.	100	-2.81	-3.86	11.41
sans trend	200	-2.67	-3.77	10.47
cste	60	-4.08	-2.63	9.95
sans var. sais.	100	-3.99	-2.33	9.41
trend	200	-3.89	-2.08	8.86
cste	60	-3.67	-4.25	15.83
var. sais.	100	-3.53	-4.01	14.18
trend	200	-3.40	-3.85	12.92

sont légèrement différentes de celles données dans Osborn, Chui, Smith, and Birchenhall (1988) qui pourtant ont une problématique similaire. Mais leur régression de test est plus restrictive car elle n'inclue par le régresseur  $\hat{u}_{t-s-1}$ .

Si l'on examine cette table en la comparant à celles de Dickey et Fuller pour l'hypothèse nulle  $\rho = 1$  et à celle de Dickey, Hasza et Fuller pour l'hypothèse nulle  $\phi = 1$ , on peut mettre en lumière les fait suivants. Le test de  $\rho = 1$  est moins puissant qu'un test de racine unitaire usuel à l'ordre 1, mais la présence de variables saisonnières rétabli la parité des puissances. Le test de  $\phi = 1$  est plus puissant que celui de Dickey, Hasza et Fuller.



Le papier de Barthélémy et Lubrano se termine par une application empirique qui va nous permettre de motiver d'une certaine façon le paragraphe suivant. Elle concerne une série mensuelle de consommation électrique aux USA portant sur la période 1953:1-1961:12. Cette série présente une forte saisonnalité (voir la figure 3) qui de plus évolue dans le temps à cause d'un équipement croissant en climatiseurs. La régression de test donne les résultats suivants après avoir enlevé un terme constant et onze variables saisonnières:

$$\Delta\Delta_{12}\hat{u}_t = -\underset{[-2.74]}{0.23}\hat{u}_{t-1} - \underset{[-4.46]}{0.45}\hat{u}_{t-12} + \underset{[4.82]}{0.56}\hat{u}_{t-13} - \underset{[-4.41]}{0.38}\Delta\Delta_{12}\hat{u}_{t-2}$$

avec un  $R^2 = 0.48$  et un écart-type de la régression de 0.206. Pour 120 observations, les valeurs critiques à 5% sont de -2.89 pour  $\rho$  et de -6.09 pour  $\phi$  (valeurs lues dans la Table 4 de Barthélemy et Lubrano). On peut donc accepter la présence d'une double racine unitaire à l'ordre 1 et à l'ordre 12. Le modèle SARIMA que l'on pourra ensuite ajuster comporte une différenciation à l'ordre 1 et à l'ordre 12. Une recherche de spécification conduit au modèle suivant:

$$(1 + \underset{[-4.92]}{0.59}L^{12})(1 - \underset{[5.31]}{0.50}L)\Delta\Delta_{12}y_t = (1 - \underset{[47.59]}{0.97}L)\epsilon_t + \underset{[2.76]}{0.0042}$$

$$R^2 = 0.39 \quad \hat{\sigma} = 0.222$$

Ce que l'on remarque tout de suite, c'est que la partie MA de ce modèle présente une racine proche de l'unité, ce qui laisse supposer une légère surdifférenciation à l'ordre 1, et ce malgré les tests.

## 7.4 Décomposition de la racine saisonnière selon ses fréquences

Hylleberg, Engle, Granger, and Yoo (1990) (HEGY) s'intéressent à un type de test de racine unitaire saisonnière assez particulier. Leur but consiste à pouvoir comparer les mouvements saisonniers présents dans deux ou plusieurs séries. A partir du moment où deux séries auront des mouvements saisonniers stochastiques similaires, on pourra alors chercher une relation de cointégration saisonnière entre ces deux séries. La notion de cointégration sera développée dans les deux chapitres suivants.

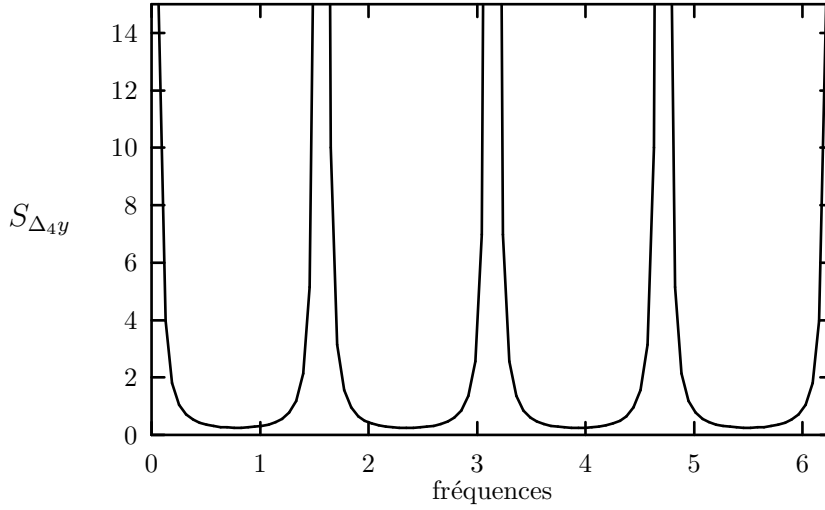
Considérant le cas  $s = 4$ , Hylleberg, Engle, Granger, and Yoo (1990) font remarquer que:

$$(1 - L^4) = (1 - L)(1 + L)(1 + L^2) \quad (72)$$

ce qui fait une racine unitaire usuelle qui représente le trend, et trois autres racines qui représentent la saisonnalité avec une racine égale à un et deux racines complexes conjuguées. Les racines sont donc  $(1, -1, i, -i)$ . Si maintenant on calcule le spectre associé à  $\Delta_4 y_t = \epsilon_t$ , on obtient:

$$S_{\Delta_4 y} = \frac{1}{2(1 - \cos(4\omega))} \sigma_\epsilon^2$$

Ce spectre comporte trois pointes en  $\omega = 0$ ,  $\omega = \pi$  et  $\omega = \pi/2$  comme on peut le constater sur la figure 4. Ces pointes correspondent aux racines 1 (le trend), -1 (cycle annuel) et aux

Figure 4: Spectre associé à  $\Delta_4 y_t$ 

deux racines complexes (cycle bi annuel). HYLLEBERG *et al* vont chercher la présence de racines unitaires à chacune de ces fréquences. C'est donc aussi un cas de tests de plusieurs racines de manière simultanée.

HYLLEBERG *et al* développent le polynôme  $A(L)$  autour de ses racines au moyen du théorème de Lagrange, ce qui donne tous calculs faits:

$$\begin{aligned}
 A(L) = & -\lambda_1 L(1 + L + L^2 + L^3) \\
 & -\lambda_2 L(1 - L + L^2 - L^3) \\
 & -(\lambda_3 + \lambda_4 L)L(1 - L^2) \\
 & +(1 - A^*(L))(1 - L^4)
 \end{aligned} \tag{73}$$

On définit les trois transformées suivantes de  $y_t$  que l'on appelle:

$$\begin{aligned}
 x_{1t} &= (1 + L + L^2 + L^3)y_t \\
 x_{2t} &= (1 - L + L^2 - L^3)y_t \\
 x_{3t} &= (1 - L^2)y_t
 \end{aligned} \tag{74}$$

qui vont permettre de capturer les trois pointes du spectre. Alors le modèle sans terme déterministe  $A(L)y_t = \epsilon_t$  peut s'écrire:

$$\Delta_4 y_t = \lambda_1 x_{1t-1} + \lambda_2 x_{2t-1} + \lambda_3 x_{3t-1} + \lambda_4 x_{3t-2} + A^*(L)\Delta_4 y_t + \epsilon_t \tag{75}$$

Cette régression ressemble à celle de Dickey, Hasza, and Fuller (1984), mais le terme  $y_{t-4}$  a été décomposé en quatre. Cette régression permet de tester la présence de quatre racines. La première racine, celle qui vient de  $(1 - L)$  se teste au moyen de  $\lambda_1 = 0$ .

Elle correspond au test ADF usuel et correspond à la tendance de long terme. Les racines saisonnières impliquent  $\lambda_2 = 0$  pour la première et le test joint  $\lambda_3 = \lambda_4 = 0$  pour les deux racines complexes conjuguées. Hylleberg, Engle, Granger, and Yoo (1990) donnent les tables adéquates en fonction des divers cas de trend dont on reproduit ici quelques extraits pour les valeurs à 5% pour les tests en  $t$  et à 95% pour le test en  $F$ . HYLLEBERG *et*

Table 9: Distribution des tests de saisonnalité de HEGY

	T	Valeurs critiques				
		$t$ à 5%		$F$ à 95%		
		$t : \lambda_1$	$t : \lambda_2$	$t : \lambda_3$	$t : \lambda_4$	$F : \lambda_3 \cap \lambda_4$
sans cste	48	-1.95	-1.95	-1.93	-1.76	3.26
sans dummies	100	-1.97	-1.92	-1.90	-1.68	3.12
sans trend	200	-1.94	-1.95	-1.92	-1.65	3.16
cste	48	-2.96	-1.95	-1.90	-1.72	3.04
sans dummies	100	-2.88	-1.95	-1.90	-1.68	3.08
sans trend	200	-2.87	-1.92	-1.90	-1.66	3.12
cste	48	-3.08	-3.04	-3.61	-1.98	6.60
dummies	100	-2.95	-2.94	-3.44	-1.96	6.57
sans trend	200	-2.91	-2.89	-3.38	-1.96	6.61
cste	48	-3.56	-1.91	-1.92	-1.70	2.95
sans dummies	100	-3.47	-1.94	-1.89	-1.65	2.98
trend	200	-3.44	-1.95	-1.92	-1.66	3.07
cste	48	-3.71	-3.08	-3.66	-1.91	6.55
dummies	100	-3.53	-2.94	-3.48	-1.94	6.60
trend	200	-3.49	-2.91	-3.41	-1.92	6.57

$$H_0 : \Delta_4 y_t \sim N(0, 1)$$

al font remarquer les points suivants sur la distribution asymptotique des tests, points qui se retrouvent dans la Table 9 tirée de leur article. En l'absence de terme déterministe, le test sur  $\lambda_1 = 0$  est un test ADF classique si  $\lambda_2 = \lambda_3 = \lambda_4 = 0$ . D'autre part, le test sur  $\lambda_3 = 0$  est équivalent au test de Dickey, Haza et Fuller si  $\lambda_4 = 0$ . Quand on ajoute un terme déterministe, la présence d'une constante et d'un trend influencent seulement le test sur  $\lambda_1 = 0$ . S'il y a une constante, la présence de trois variables saisonnières ne change pas la distribution du test pour  $\lambda_1 = 0$ , mais changent la distribution des tests pour  $\lambda_2 = 0$ ,  $\lambda_3 = 0$ ,  $\lambda_4 = 0$ .

Cette approche permet de caractériser plus finement le comportement saisonnier des séries et est particulièrement utile quand on veut trouver des similitudes entre deux séries et étudier leur possible cointégration. Elle permet aussi de se rendre compte qu'une différenciation

à l'ordre 4 impose la présence simultanée de plusieurs racines unitaires. Or il se peut très bien que certaines racines soient présentes dans une série, mais pas toutes. Ce cas intermédiaire ne peut être traité avec le filtre global  $(1 - L^4)$ , mais nécessite donc des filtres séparés correspondants à chacune des séries et données par (74).

## 7.5 Influence de la périodicité sur les tests ADF

Les modèles que nous avons considéré dans cette section s'appliquent sur des séries trimestrielles ou mensuelles qui n'ont subi aucune procédure de désaisonnalisation. Or il arrive bien souvent que de telles données ne soient disponibles que sous forme désaisonnalisée. L'exemple type sont les données des comptes trimestriels de la comptabilité nationale fournies par l'INSEE. Mais bien d'autres exemples existent sur les données étrangères. Quelle est alors l'influence des procédures de désaisonnalisation sur les tests de racine unitaire?

On sait déjà à la suite de l'article de Wallis (1974) que si des données sont désaisonnalisées avec des filtres différents, la régression basée sur ces données fournira des estimateurs biaisés. De plus Ghysels and Perron (1993) ont montré que dans un processus AR, l'utilisation de données désaisonnalisées par un filtre linéaire conduisait à un biais systématique vers le bas du coefficient d'auto-régression. Un test de racine unitaire étant basé sur un tel modèle, l'utilisation de données filtrées va conduire à rejeter trop souvent l'hypothèse de racine unitaire.

Un échantillon basé sur des données trimestrielles sera quatre fois plus long que le même échantillon basé sur des données annuelles. Pourtant ce qui vient d'être dit inciterait plutôt à utiliser des données annuelles pour faire des tests de racine unitaire. Quelle est donc l'information supplémentaire apportée par des données trimestrielles? Il semblerait en fait d'après Shiller and Perron (1985) que la puissance des tests soit beaucoup plus affectée par la longueur de la période couverte, que par le nombre d'observations. Donc on ne gagne pas beaucoup à utiliser des données trimestrielles et encore moins des données mensuelles. La caractérisation d'un trend stochastique se fait surtout sur la longue période. L'article de Nelson and Plosser (1982) est par exemple basé sur des séries annuelles dont certaines remontent à 1860.

Pourtant l'utilisation de séries historiques longues pose aussi certains problèmes. Premièrement les premières observations peuvent être de qualité douteuse. C'est le cas en France pour les années suivant l'immédiat après guerre et l'avant guerre aux États-Unis. Les procédures de collecte des données n'étaient pas encore parfaites. Certaines séries sont purement et simplement extrapolées à partir d'un trend, ce qui introduit un biais contre l'hypothèse de racine unitaire. Ensuite l'allongement de la période augmente les chances d'existence d'un changement structurel majeur qui affecterait le trend. Ces changements structurels sont alors interprétés de façon erronée comme la manifestation d'une racine unitaire. C'est ce que l'on a vu dans la section précédente.

## 8 Questions Diverses

Dans ce chapitre nous n'avons pas abordé un certain nombre de questions qui se rattachent aux tests de racine unitaire. Nous allons maintenant simplement les évoquer, renvoyant pour plus de détails aux surveys qui ont été faits sur le sujet et qu'il nous faut maintenant citer: Diebold and Nerlove (1990), Nerlove (1989), Dolado, Jenkinson, and Sosvilla-Rivero (1990), Campbell and Perron (1991), Davidson and MacKinnon (1993), chapitre 20, Hamilton (1994), chapitre 17.

### 8.1 Autres tests

Les tests les plus employés sont les tests de Dickey et Fuller notés  $\tau$ . Mais on a vu également la statistique  $z = T(\hat{\rho} - 1)$  qui mesure directement l'écart de  $\hat{\rho}$  par rapport à un. Il faut également citer les tests en  $F$  de Dickey et Fuller sur l'hypothèse jointe que  $\rho$  soit égal à un et que  $\beta_k$  soit nul.

Bhargava (1986) a proposé un type de tests basé sur une modification du test usuel de Durbin et Watson. Ces tests ont de bonnes propriétés de petit échantillon. Mais ce test ne peut détecter que des racines unitaires dans un processus  $AR(1)$ . Dans le modèle sans trend avec constante la statistique est:

$$R_1 = \frac{\sum_{t=2}^T (\Delta y_t)^2}{\sum_{t=2}^T (y_t - \bar{y})^2} \quad (76)$$

correspondant au modèle:

$$(1 - \rho L)(y_t - \mu) = \epsilon_t \quad (77)$$

Si l'on rajoute un trend à ce modèle, la statistique de test devient plus difficile à écrire avec:

$$R_2 = \frac{\sum_{t=2}^T (y_t - y_{t-1})^2 - \frac{1}{T-1} (y_t - y_1)^2}{\frac{1}{(T-1)^2} \sum_{t=1}^T \left[ (T-1)y_t - (t-1)y_T - (T-t)y_1 - (T-1)(\bar{y} - \frac{1}{2}(y_1 + y_T)) \right]^2} \quad (78)$$

Les valeurs critiques de ces deux tests à 5% sont données dans la Table 10.

Table 10: Valeurs critiques à 5% du test de Barghava

T	20	30	40	50	60	80	100
$R_1$	1.10	0.79	0.61	0.50	0.42	0.32	0.26
$R_2$	1.45	1.04	0.81	0.66	0.56	0.43	0.35

### 8.2 Construction des tables

Il existe deux manières de construire les tables que l'on a donné dans le texte. Toutes deux reposent sur des expériences de Monte Carlo.

- On peut tout d'abord essayer d'obtenir des résultats de petit échantillon en simulant pour une taille  $T$  donnée une marche aléatoire, calculer le test  $\tau$  correspondant et construire ensuite la distribution empirique de ce test. Au bout de  $N$  expérimentations, on aura une précision satisfaisante pour calculer les quantiles de cette distribution et donc les valeurs critiques du test. Dans la pratique  $N$  peut varier entre 10000 et 25000. On devra reproduire ce type d'expérience pour des valeurs différentes de  $T$  premièrement et ensuite pour les divers types de régression en rajoutant à la régression une constante, puis un trend. C'est selon ce principe qu'on a construit les tables de Dickey and Fuller (1981). Les tables les plus précises construites selon ce principe sont celles de MacKinnon (1991).

- On peut décider de simuler directement la distribution asymptotique des statistiques de test. On a vu que ces distributions étaient des fonctionnelles de processus de Wiener et on sait que si  $\epsilon_t \sim N(0, 1)$  alors on a les résultats de convergence suivants pour  $T \rightarrow \infty$  :

$$T^{-3/2} \sum_{t=1}^T \left( \sum_{j=1}^t \epsilon_j \right) \xrightarrow{\mathcal{L}} \int_0^1 W(r) dr \quad (79)$$

$$T^{-1} \sum_{t=1}^T \left( \sum_{j=1}^{t-1} \epsilon_j \right) \epsilon_t \xrightarrow{\mathcal{L}} \frac{1}{2} (W(1)^2 - 1) \quad (80)$$

$$T^{-2} \sum_{t=1}^T \left( \sum_{j=1}^t \epsilon_j \right)^2 \xrightarrow{\mathcal{L}} \int_0^1 W^2(r) dr \quad (81)$$

Ces résultats sont directement issus des résultats de convergence que l'on a discuté au chapitre 2. Ils peuvent être employés pour simuler les expressions asymptotiques correspondantes en se basant sur la génération d'un vecteur de  $IN(0,1)$  d'une longueur "suffisante". En disant "suffisante", on fait état du principal problème. Sera-t-on suffisamment près de la vraie distribution? Considérons l'exemple de la simulation de la distribution asymptotique de la statistique  $\tau$  pour fixer les idées. On a donné son expression en (14). L'algorithme qui permet de générer cette distribution est le suivant:

- 1) début de la boucle d'indice  $i$  avec  $nt = 10\,000$
- 2) générer un vecteur  $e \sim IN(0, 1)$  de longueur  $T = 500$
- 3) calculer  $a = T^{-1} \sum_{t=1}^T \left( \sum_{j=1}^{t-1} \epsilon_j \right) \epsilon_t$
- 4) calculer  $b = T^{-2} \sum_{t=1}^T \left( \sum_{j=1}^t \epsilon_j \right)^2$
- 5)  $f[i] = 0.5 * a / \sqrt{b}$
- 6)  $i = i + 1$ ; retour en 2 si  $i < nt$
- 7) ordonner le vecteur  $f$  de longueur  $nt$
- 9) imprimer les valeurs critiques à  $\alpha\%$  données par  $f[\alpha * nt]$

Cet algorithme requiert un nombre de tirages similaire à celui de la méthode précédente. C'est ce type d'algorithme qui a été employé par Perron (1989) pour le calcul de ses tables.

### 8.3 Intégration fractionnaire

L'analyse que l'on a menée jusqu'à présent reposait sur l'idée que l'on arrivait à stationnariser certaines séries en les différentiant. Si par exemple  $y_t \sim I(d)$  alors  $(1 - L)^d y_t \sim I(0)$  c'est à dire qu'en différentiant  $d$  fois la série, on arrive à la stationnariser. On a toujours fait l'hypothèse que  $d$  était entier et en général égal à l'unité. Une série  $I(1)$  a alors une représentation ARMA en différence première. Le modèle appliqué à la série brute est un modèle ARIMA. ? ont introduit une classe de modèle où  $d$  est un nombre fractionnaire. Le filtre  $(1 - L)^d$  est rendu opérationnel en considérant le développement en série:

$$(1 - L)^d \cong 1 - dL + \frac{d(d-1)}{2!}L^2 - \frac{d(d-1)(d-2)}{3!}L^3 + \dots \quad (82)$$

Si  $y_t \sim I(d)$  avec  $d$  fractionnaire, ce filtre permettra de la stationnariser.

Comme on l'a dit plus haut dans le texte, un des intérêts principaux des modèles intégrés est de modéliser la persistance de long terme dans les séries économiques. Les racines unitaires offrent une première possibilité. Mais elles introduisent des contraintes dans la modélisation de la persistance. On peut s'en rendre compte en examinant la variance des accroissements de  $y_t$ . Quand  $y_t$  est  $I(1)$  on a que:

$$\text{Var}(y_t - y_{t-k}) = O(k) \quad (83)$$

Ce résultat dans le cas où  $y_t$  est  $I(d)$  se transforme en:

$$\text{Var}(y_t - y_{t-k}) = O(k^{2d-1}) \quad (84)$$

ce qui montre que la variance devient indépendante de  $k$  pour  $d = 0.5$ . On peut montrer qu'alors le processus est stationnaire. Pour  $d < 0.5$  cette variance devient plate. Pour  $0.5 < d < 1$  elle croît à un taux décroissant et pour  $1 < d < 3/2$  à un taux croissant. Pour  $d > 3/2$  elle n'existe plus. Les modèles avec intégration fractionnaire apportent donc plus de souplesse. Par exemple certaines séries non stationnaires semblent présenter encore une racine unitaire après une différenciation. Mais les différencier deux fois est trop. Ces séries sont entre  $I(1)$  et  $I(2)$ . L'intégration fractionnaire peut apporter une solution à leur modélisation.

### 8.4 Tests Bayésiens

Le point central que l'on a voulu mettre en avant dans tout ce qui a été dit sur les tests de racine unitaire, c'est la discontinuité qu'il existe dans la théorie asymptotique entre le cas stationnaire et le cas non stationnaire<sup>2</sup>. Si ce n'était ce point, les tests de racine unitaire serait de simples tests de Student utilisant les tables usuelles. Quand on travaille dans une optique Bayésienne, les résultats d'inférence sont donnés conditionnellement à l'échantillon observé. Il n'y a donc dans ce cadre aucune différence entre le cas stationnaire

---

<sup>2</sup>Dans le modèle auto-régressif  $y_t = \rho y_{t-1} + \epsilon_t$ , l'estimateur OLS de  $\rho$  a une distribution asymptotique qui est Normale dans le cas stationnaire, qui est une fonctionnelle de processus de Wiener dans le cas  $\rho = 1$ , et qui est Cauchy dans le cas explosif.

et le cas non-stationnaire, pour la simple raison que l'on ne s'intéresse pas au comportement asymptotique de l'échantillon vu que l'on travaille conditionnellement à l'échantillon observé. Il faut noter de même qu'en analyse classique on obtient des résultats particuliers avec les tests de racine unitaire dans les données de panel. Un panel est un échantillon qui a deux dimensions d'observation: une dimension individuelle et une dimension temporelle. Quand on fait des tests de racine unitaire dans ces échantillons, il suffit de faire tendre vers l'infini, non le nombre d'observations temporelles, mais le nombre d'individus pour que l'on retrouve la normalité asymptotique.

Certains auteurs ont donc développé des tests Bayésiens de racine unitaire. Le premier fut Sims (1988), soulignant le caractère conditionnel de l'inférence Bayésienne. DeJong and Whiteman (1991) apportèrent les premières applications empiriques d'où il ressortait que dans une optique Bayésienne les séries macro-économiques américaines étaient en majorité trend stationnaires, contrairement aux résultats trouvés par Nelson and Plosser (1982) qui utilisaient une approche classique. Il s'en est suivi une polémique importante qui a démarré avec l'article de Phillips (1991) sur l'argument que si les tests Bayésien rejettent si souvent l'hypothèse de racine unitaire, c'est que l'a priori uniforme sur  $\rho$  qu'ils utilisaient favorisait trop l'hypothèse de stationnarité. Les résultats empiriques précédents furent donc déclarés non robustes par rapport à la spécification de l'a priori. En employant le modèle linéaire  $y_t = \rho y_{t-1} + \beta_0 + \beta_1 t + \epsilon_t$  et une a priori de Jeffreys sur  $\rho$ , Phillips est parvenu à reproduire plus ou moins les résultats classiques. La controverse a pourtant été nourrie par le fait que l'a priori de Jeffreys est très irréaliste et ne reflète certainement pas les opinions que l'on peut avoir sur la distribution de  $\rho$ .

Schotman and van Dijk (1991) et Schotman and van Dijk (1993) ont abordé le problème différemment en utilisant le modèle non-linéaire qui impose que  $y_t$  soit du même ordre sous l'hypothèse alternative et sous l'hypothèse nulle. Ils développent un test de l'hypothèse ponctuelle  $\rho = 1$  basés sur des posterior odds. Ils retrouvent des résultats finaux similaires à ceux de Phillips sur les données de Nelson et Plosser. Mais leurs résultats sont fragiles, car les posterior odds dépendent de l'a priori qui est choisie et qui dépend de la configuration de l'échantillon.

Lubrano (1995) utilise le même modèle type de modèle non linéaire, mais insiste sur le fait que les conditions initiales du processus doivent être prises en compte et modélisées. D'autre part, il s'intéresse au test de l'hypothèse  $|\rho| > 1$  en calculant la probabilité a posteriori de cette hypothèse, ce qui revient à trouver un intervalle de confiance pour l'hypothèse  $\rho = 1$ . Ce type de test est indépendant de la forme de l'a priori. Il conduit également à rejeter beaucoup moins souvent l'hypothèse de racine unitaire.

La conclusion que l'on peut aborder à ce débat, c'est que la spécification du modèle et des conditions initiales a une grande importance quand on veut discuter du comportement non-stationnaire d'un processus dynamique.



## 9 Conclusion

La littérature sur les racines unitaires est très volumineuse, tant sur le plan de la théorie statistique que sur celui des applications empiriques et de la théorie économique. On a tenté de donner un fil conducteur le plus clair possible sur la partie statistique. Un survey de la partie économique serait également intéressant à faire. On s'est attaché à montrer que l'hypothèse de racine unitaire entraînait un comportement très particulier des séries et une modélisation particulière de la permanence. Il est un point cependant que l'on a peu évoqué. Les tests de racine unitaire sont des tests très peu puissants. Le point qu'ont souligné Campbell and Perron (1991), c'est la presque équivalence empirique entre un modèle trend stationnaire et un modèle stationnaire en différence; la presque équivalence empirique est définie au sens où le corrélogramme des deux processus peuvent être rendus arbitrairement proches. Il suffit pour cela de considérer des résidus ARMA et non plus AR. Une racine unitaire dans la partie AR peut alors être annulée par une racine presque unitaire dans la partie MA. Si bien que certains auteurs comme Christiano and Eichenbaum (1990) ont émis l'idée que l'on devrait abandonner l'idée même de tester la présence d'une racine unitaire. On peut rester pragmatique en disant que ces tests doivent être pris avec un petit grain de sel. Il est illusoire de vouloir opposer deux théories économiques sur la base d'un simple test de racine unitaire. Par contre un tel test peut être un instrument de modélisation utile pour spécifier une relation de cointégration.

## 10 Lectures additionnelles

Il existe plusieurs survey généraux sur les tests de racine unitaires. On peut citer par exemple Stock (1994). L'ouvrage de Hamilton (1994) est intéressant pour la clarté des démonstrations. Celui de Maddala and Kim (1998) couvre plus de matière et met bien en perspective les différents acquis de la littérature. Il conduit des discussions parfois originales.

## 11 Exercices

### 11.1 Processus I(1)

Considérons le processus AR(1),  $y_t = \rho y_{t-1} + \epsilon_t$ . On ne précise pas pour l'instant la valeur de  $\rho$ .

- 1) Par substitutions successives, écrire  $y_t$  en fonction de la condition initiale  $y_0$  et de la somme des termes d'erreurs.
- 2) Calculez la moyenne et la variance de  $y_t$  pour  $\rho < 1$ . Montrez que l'estimateur des moindres carrés de  $\rho$  est consistant. Quelle hypothèse devez vous faire sur  $\rho$ ?
- 3) Calculez la moyenne et la variance de  $y_t$  pour  $\rho = 1$ , conditionnellement à  $y_0$ . Comparez avec le cas stationnaire. Que peut-on en conclure quand  $t \rightarrow \infty$ ?

- 4) Calculez de même les auto-covariances. Comparez les résultats pour  $\rho < 1$  et  $\rho = 1$ .
- 5) Différenciez la série et calculez alors la moyenne et la variance de  $\Delta y_t$ . Qu'en concluez vous?

• **Idées de solution** La décomposition MA infinie est

$$y_t = \sum_{i=0}^{\infty} \rho^i \epsilon_{t-i}$$

On calcule alors  $E(y_t) = 0$  et  $\text{Var}(y_t) = \sigma^2/(1 - \rho^2)$ . Les OLS sont consistants si  $\sum y_{t-1}^2/T$  converge vers  $\sigma^2/(1 - \rho^2)$ , ce qui ne peut se faire que si  $\rho < 1$ . Si  $\rho = 1$ , il faut faire le même développement, mais en le décomposant

$$y_t = \sum_{i=0}^{t-1} \rho^i \epsilon_{t-i} + \rho^t y_0$$

On a alors que l'espérance est toujours nulle, mais que la variance est  $\sigma^2 t$ . La série est non stationnaire et le moment d'ordre 2 tend vers l'infini. Si on différencie la série, on a que  $\Delta y_t = \epsilon_t$ . La moyenne est nulle et la variance est égale à  $\sigma^2$ .

## 11.2 Simulation d'un processus I(1)

On va maintenant simuler le même processus AR(1) que précédemment en prenant deux cas  $\rho = 0.5$  et  $\rho = 1$ . On utilisera bien sûr les mêmes valeurs pour le terme d'erreur  $\epsilon$ . Prenez  $T = 50$ .

- 1) Estimez l'auto-corrélation des deux processus et faites en le graphique en superposant à chaque fois la valeur empirique et la valeur théorique.
- 2) Refaites le même exercice pour  $T = 100$  et  $T = 500$ .

## 11.3 Processus ARIMA(0,1,1)

Considérons le processus ARIMA(0,1,1) que l'on note

$$\Delta y_t = \alpha + \epsilon_t + \beta \epsilon_{t-1}$$

- 1) Calculez par substitutions successives la représentation MA de  $y_t$ . Interpréter les termes trouvés de manière à dégager une tendance déterministe, une tendance stochastique et un cycle stationnaire.
- 2) Pour  $T = 100$ ,  $\alpha = 0.01$  et  $\beta = 0.4$ , tracer sur un même graphique le cycle, la tendance stochastique, la tendance déterministe et la série  $y_t$ . On prendra  $\epsilon \sim N(0, 0.001)$ .
- 3) Renouvelez l'exercice de simulation avec  $\epsilon \sim N(0, 0.01)$  et  $\epsilon \sim N(0, 0.1)$ .

• **Idées de solution**

## 11.4 Test ADF

Considérez la régression de test ADF

$$\Delta y_t = \rho y_{t-1} + \beta \Delta y_{t-1} + \epsilon_t$$

sans termes déterministes

- 1) Ecrire la matrice de variance-covariance de l'estimateur des moindres carrés de  $\rho$  et  $\beta$ .
- 2) Montrez que la distribution de  $\hat{\rho}$  et celle de  $\hat{\beta}$  sont indépendantes.
  - **Idées de solution** Écrire l'estimateur OLS en pensant à le normaliser correctement. Montrez que l'élément diagonal de la matrice des double produits tend asymptotiquement vers zéro.

## 11.5 Les composantes déterministes

Considérez la régression de test DF suivante

$$\Delta y_t = \rho y_{t-1} + \beta_0 + \beta_1 t + \epsilon_t$$

qui comporte donc deux termes déterministes.

- 1) Montrez que la distribution de  $\hat{\rho}$  et celle de  $\hat{\beta}_i$  ne sont pas indépendantes.
- 2) Qu'en concluez vous sur la distribution du test des hypothèses  $\beta_0 = 0$  et  $\beta_1 = 0$ ?
  - **Idées de solution** Il faut écrire l'estimateur OLS en pensant à le normaliser correctement. Ensuite, il faut montrer que l'élément diagonal de la matrice des double produits ne tend pas asymptotiquement vers zéro.

## References

- BARTHELEMY, F., AND M. LUBRANO (1996): "Unit Roots Tests and SARIMA Models," *Economic Letters*, 50(2), 147–154.
- BHARGAVA, A. (1986): "On the Theory of Testing for Unit Roots in Observed Time Series," *Review of Economic Studies*, 53, 369–384.
- BLANCHARD, O. J., AND L. H. SUMMERS (1986): "Hysteresis and the European unemployment problem," in *NBER Macroeconomics Annual*.
- BOSWIJK, P. H. (1992): *Cointegration, Identification and Exogeneity: Inference in Structural Error Correction Models*. Thesis Publisher, Tinbergen Institute Research Series, Amsterdam.

- BOX, G., AND G. JENKINS (1976): *Time Series Analysis Forecasting and Control*. San Francisco.
- CAMPBELL, J., AND P. PERRON (1991): "Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots," in *NBER macroeconomics annual 1991*, ed. by O. J. Blanchard, and S. Fischer, pp. 141–201. MIT Press, Cambridge and London.
- CHRISTIANO, L., AND M. EICHENBAUM (1990): "Unit Roots in Real GNP: do we Know and do we Care?," *Carnegie-Rochester Conference Series on Public Policy*, 32, 7–61.
- DAVIDSON, R., AND J. G. MACKINNON (1993): *Estimation and Inference in Econometrics*. Oxford University Press, Oxford.
- DEJONG, D., AND C. WHITEMAN (1991): "Reconsidering Trends and Random Walks in Macro economic Time Series," *Journal of Monetary Economics*, 28, 221–254.
- DICKEY, D., W. BELL, AND R. MILLER (1986): "Unit Roots in Time Series Models: Tests and Implications," *The American Statistician*, 40, 12–26.
- DICKEY, D., AND W. FULLER (1979): "Distribution of the Estimator for the autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427–431.
- (1981): "Likelihood Statistics for Autoregressive Time Series with a Unit Root," *Econometrica*.
- DICKEY, D., D. HASZA, AND W. FULLER (1984): "Testing for Unit Roots in Seasonal Time Series," *Journal of the American Statistical Association*, 79, 355–367.
- DICKEY, D., AND S. PANTULA (1987): "Determining the Order of Differencing in Autoregressive Processes," *Journal of Business and Economic Statistics*, 15, 455–461.
- DIEBOLD, F., AND M. NERLOVE (1990): "Unit Roots in Economic Time Series: a Selective Survey," in *Advances in Econometrics: Co-integration, Spurious Regression and Unit Roots*, ed. by T. Fomby, and G. Rhodes, vol. 8, pp. 3–69. JAI Press.
- DOLADO, J., T. JENKINSON, AND S. SOSVILLA-RIVERO (1990): "Cointegration and Unit Roots," *Journal of Economic Surveys*, 4, 249–273.
- FULLER, W. (1976): *Introduction to Statistical Time Series*. John Wiley, New York.
- GHYSELS, E., AND P. PERRON (1993): "The Effect of Seasonal Adjustment Filters on Tests for a Unit Root," *Journal of Econometrics*, 55, 57–98.
- GRANGER, C. W. J., AND P. NEWBOLD (1986): *Forecasting Economic Time Series*. San Diego.
- HALL, R. (1978): "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy*, 86, 971–987.

- HAMILTON, J. D. (1994): *Time Series Analysis*. Princeton University Press, Princeton.
- HYLLEBERG, S., R. F. ENGLE, C. W. GRANGER, AND S. YOO (1990): "Seasonal Integration and Cointegration," *Journal of Econometrics*, 44, 215–238.
- KING, R., C. PLOSSER, AND S. REBELO (1988): "Production, Growth and Business Cycles II. New Directions," *Journal of Monetary Economics*, 21, 309–341.
- LAROQUE, G., P. RALLE, B. SALANIÉ, AND J. TOUJAS-BERNATE (1990): "Description d'une Base de Données Trimestrielles Longues: 1946:1 à 1989: 4," Discussion paper.
- LUBRANO, M. (1995): "Testing for Unit Roots in a Bayesian Framework," *Journal of Econometrics*, 69, 81–109.
- MACKINNON, J. (1991): "Critical Values for Cointegration Tests," in *Long-Run Economic Relationships*, ed. by R. F. Engle, and C. W. Granger, pp. 266–276. Oxford University Press, Oxford.
- MADDALA, G., AND I.-M. KIM (1998): *Unit Roots, Cointegration and Structural Change*. Cambridge University Press, Cambridge, UK.
- NABEYA, S., AND K. TANAKA (1988): "Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative," *Annals of Statistics*, 16, 218–235.
- NELSON, C., AND C. PLOSSER (1982): "Trends and Random Walks in Macroeconomic Time Series," *Journal of Monetary Economics*, 10, 139–162.
- NERLOVE, M. (1989): "Unit Roots in Economic Time Series: an Introduction," Discussion paper.
- NEWBY, W., AND K. WEST (1987): "A simple positive semi-definite heteroskedasticity and autocorrelation-consistent covariance matrix," *Econometrica*, 55, 703–708.
- NG, S., AND P. PERRON (1995): "Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag," *Journal of the American Statistical Association*, 90(429), 268–281.
- OSBORN, D., A. CHUI, J. SMITH, AND C. BIRCHENHALL (1988): "Seasonality and the Order of Integration for Consumption," *Journal of Applied Econometrics*, 3, 255–266.
- PERRON, P. (1989): "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis," *Econometrica*, 57, 1361–1401.
- PERRON, P., AND S. NG (1996): "Useful modifications to some unit root tests with dependent errors and their local asymptotic properties," *Review of Economic Studies*, 63(3), 435–465.
- PHILLIPS, P. C. (1987): "Time Series Regression with a Unit Root," *Econometrica*, 55, 277–301.

- (1991): “To criticize the critics: an objective Bayesian analysis of stochastic trends,” *Journal of Applied Econometrics*.
- PHILLIPS, P. C., AND P. PERRON (1988): “Testing for a Unit Root in Time Series Regression,” *Biometrika*, 75, 335–346.
- SCHOTMAN, P., AND H. K. VAN DIJK (1991): “A Bayesian Analysis of the Unit Root in Real Exchange Rates,” *Journal of Econometrics*, 49, 195–238.
- (1993): “Posterior Analysis of Possibly Integrated Time Series with an Application to Real GNP,” in *New Directions in Time Series Analysis part II, IMA Volumes in Mathematics and its Applications 46*, ed. by P. Caines, J. Geweke, and M. Taqqu, pp. 341–361. Springer Verlag, Heidelberg.
- SCHWERT, G. (1989): “Tests for Unit Roots: a Monte Carlo Investigation,” *Journal of Business and Economics Statistics*, 7, 147–160.
- SHILLER, R., AND P. PERRON (1985): “Testing the Random Walk Hypothesis: Power versus Frequency of Observation,” *Economic Letters*, 18, 381–386.
- SIMS, C. A. (1988): “Bayesian Skepticism on Unit Root Econometrics,” *Journal of Economic Dynamic and Control*, 12, 463.
- SIMS, C. A., J. H. STOCK, AND M. W. WATSON (1990): “Inference in Linear Time Series with Some Unit Roots,” *Econometrica*, 58, 113–144.
- STOCK, J. (1994): “Unit roots, structural breaks and trends,” in *Handbook of Econometrics*, ed. by R. Engle, and D. MacFadden, vol. IV, pp. 2739–2841. Elsevier, Amsterdam.
- WALLIS, K. K. (1974): “Seasonal Adjustment and Relation between Variables,” *Journal of the American Statistical Association*, 69, 18–31.